TWO-SAMPLE INFERENCE AND CHANGE POINT DETECTION FOR SPARSE FUNCTIONAL DATA

by

Qiyao Wang

B.S. in Statistics, Wuhan University, China, 2012

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences in partial

fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

KENNETH P. DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Qiyao Wang

It was defended on

May 24, 2017

and approved by

Kehui Chen, PhD, Departments of Statistics and Psychiatry

Satish Iyengar, PhD, Department of Statistics

Sungkyu Jung, PhD, Department of Statistics

Ying Ding, PhD, Department of Biostatistics

Dissertation Director: Kehui Chen, PhD, Departments of Statistics and Psychiatry

TWO-SAMPLE INFERENCE AND CHANGE POINT DETECTION FOR SPARSE FUNCTIONAL DATA

Qiyao Wang, PhD

University of Pittsburgh, 2017

With recent advances in technology, functional-type data is arising fast in a number of fields, including finance, physics, meteorology, public health, and information technology. Driven by explosive needs in real practice, statistical methods for functional data have been developed quickly in recent decades. There are two typical types of functional data which possess different features and require different sets of techniques to model. One is called dense functional data, where for each random subject there are a large number of regularly-spaced observations. Dense functional data has been relatively well studied in terms of modeling, estimation and inference. The second one is called sparse functional data, where only a few irregularly-spaced observations are attainable for each subject. Statistical methods for sparse functional data are of less development, despite the importance and demand for these methods. In this thesis, we focus on three topics within the field of sparse functional data inference: two-sample inference of mean functions of two independent groups of functional data, one-way functional ANOVA (FANOVA), and change point detection in mean functions for sparse functional time series.

For each of the three topics mentioned above, methods for dense functional data are firstly reviewed. It helps us to understand why or why not each of these methods is applicable to sparse functional data situations. For the two-sample mean function testing problem and one-way functional ANOVA, we develop asymptotic chi-square tests for detecting differences among mean functions when sparse and irregular observations are drawn from the underlying stochastic processes for each subject. For the change point detection in a sequence of functional samples, we create two test procedures whose asymptotic distributions are related to a summation of independent

Brownian Bridge squares. We provide theoretical arguments to justify the validity of the proposed tests. Numerical experiments, including simulation studies and applications to a CD4 count data set and two eBay online auction data sets, are presented to demonstrate the good performances of the proposed test procedures.

Keywords: mean function, shrinkage estimator, sparse design, functional time series, asymptotic distribution, CD4 count, eBay auction

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

1.1 DENSE AND SPARSE FUNCTIONAL DATA

Functional data refers to data drawn from continuous underlying random processes. Put differently, each subject in the sample is a random function (a curve), instead of a random number or a random vector. With recent advances in technology, functional data is arising fast in a number of scientific fields, including finance, psychiatry, physics and public health. It has become one of the most commonly encountered types of data. Practically functional data is consist of discretized realizations of the underlying random function. For each subject, there are a number of repeated observations over time or any other continuum. Even though both functional data and vector data contain multiple observations for each subject, function-type data do not require the same number of observations for all the subjects. Functional data also account for the order of observations within each subject. Examples of functional data include repeated measurements of stock prices over time for a collection of stocks and repeated measurements of brain signals over space for a sample of patients.

Driven by the increasing needs in various fields, statistical methods for functional data, called functional data analysis (FDA), have been expanded quickly in recent decades. In FDA, there are two typical types of data: dense functional data, where a large number of regularly-observed measurements for each subject are attainable; and sparse functional data, where only a few irregularly-spaced measurements are available for each subject. Dense functional data is very common in fields where automated instruments are used to record data. A Canadian Weather study mentioned in (Ramsay, 2006) uses automated sensors to monitor the daily precipitation near 35 weather stations for a year. The left panel in Figure 1.1 displays the precipitation trajectory for a randomly selected weather station within the study period. For dense functional data, smoothing techniques

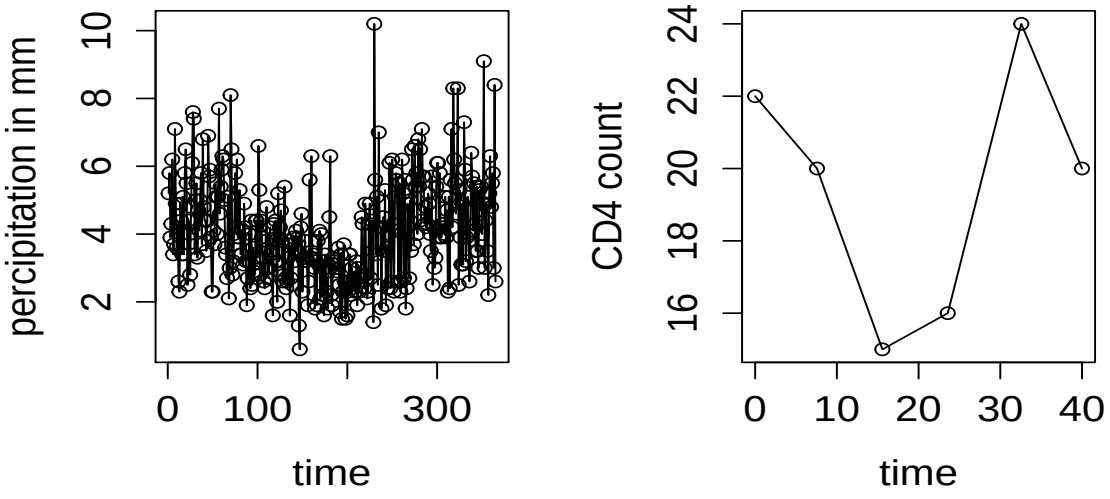

Figure 1.1: The observed daily average precipitation of a randomly selected weather station in the Canadian Weather study are on the left panel; The observed CD4 counts of a randomly selected subject in the AIDS Clinical Trial Group 193A study are on the right panel.

can be used to recover the underlying individual trajectories. Sparse functional data arises frequently in longitudinal studies. For example, in the AIDS Clinical Trial Group 193A study, 2 to 10 irregularly distributed measurements of CD4 count data are available for each patients over the 40 week study period. The right panel in Figure 1.1 visualizes the observations of a randomly selected subject in the study. Another typical example of sparse functional data comes from eBay online auctions where live bids are irregularly spaced for the pre-specified listing duration (3,5, or 7 days). For sparse functional data, pre-smoothing of each individual function is unreliable.

The reason for distinguishing between these two types of functional data is that they often exhibit different features and require different modeling techniques. Dense and sparse functional data are usually represented by two distinct formats. For dense functional data, a common practice is to first recover each underlying curve through pre-smoothing techniques, such as kernels, splines, and local polynomial. The resulting individual curves are consistent, given certain dense regular design assumptions. Then people can evaluate all recovered curves at a pre-specified common dense regular grid. Consequently, the dense functional data can be represented by an $n \times M$

matrix, where each row records M equally-spaced and noiseless observations from an individual curve. For sparse functional data, the pre-smoothing techniques no longer yield consistent recovery for the individual curves, because of the limited number of observations on each curve. The data is usually recorded in the form of (T_{ij}, Y_{ij}), for $i = 1, \ldots, n$ and $j = 1, \ldots, N_i$, where Y_{ij} is the jth observation of the ith subject, and T_{ij} is the corresponding observing time. In addition to the format, dense and sparse functional data have different routines in terms of estimation and inference. For example, when estimating the population mean function for dense functional data, a regular mean sample function can be used. At each time point, the estimate equals the sample mean value at this time over all the subjects. For sparse functional data, it is usually assumed that the pooled data across different subjects are dense. Local linear smoothing on the pooled sample is used, and the achieved smoother estimates the population mean function. The book (Ramsay, 2006) offers a comprehensive perspective of FDA methods for densely observed functional data, and the paper (Müller, 2005) provides a nice review for sparse functional data analysis.

1.2 MOTIVATION AND STRUCTURE OF THIS THESIS

Dense functional data is relatively well studied in terms of both modeling and inference. For modeling, some topics covered are functional principal component analysis (Cardot et al., 2003; Silverman, 1996), regression with functional response, predictor or both (Cardot et al., 2003; Cai et al., 2006; Yao et al., 2005), functional classification and clustering (Müller, 2005), and functional quantile analysis (Cardot et al., 2005; Ferraty et al., 2005; Chen and Müller, 2012a). Recently, there are increasing interests in modeling multivariate functional data and repeatedly observed functional data (Morris and Carroll, 2006; Chen and Müller, 2012b; Chen et al., 2017). As for statistical inferences for dense functional data, the two-sample inference for the population means of independent groups is studied by (Ramsay, 2006; Zhang and Chen, 2007; Faraway, 1997; Zhang and Liang, 2014; Zhang et al., 2010; Abramovich et al., 2004; Lillo et al., 2015; Staicu et al., 2014) from several different perspectives. Extension to multiple independent group cases (one-way FANOVA) is justified by (Shen and Faraway, 2004; Fan and Lin, 1998; Cuevas et al., 2004; Paparoditis and Sapatinas, 2014). Inferences for the covariance functions of multiple groups are

presented in (Gaines et al., 2011; Fremdt et al., 2013). Hypothesis testings regarding the functional distributions across groups are given by (Pomann et al., 2016). For repeated dense functional data, i.e., functional time series, testing whether the mean function changes at some unknown time point is discussed in (Aston and Kirch, 2012; Berkes et al., 2009; Aue et al., 2009; Račkauskas and Suquet, 2006). This topic is known as the change point detection problem in functional time series.

For sparse functional data, there are some works regarding modeling, including principal component analysis (Yao et al., 2005; James et al., 2000), and clustering analysis (James and Sugar, 2003). Statistical methods for inferencing on sparse functional data are less developed, despite their importances. In this thesis, we focus on three topics within the field of sparse functional data inference: two-sample inference of mean functions of two independent groups of functional data, one-way functional ANOVA (FANOVA), and change point detection in mean functions for sparse functional time series.

1.2.1 Motivation for two-sample mean function inference for sparse functional data

The two-sample inference for the population mean functions problem is one of the most fundamental research topics in FDA. The statistical framework of the specific two-sample inference problem that we consider can be stated as follows. Let $L^2(\mathcal{T})$ be the space that is consist of all squared-integrable functions within the closed interval $\mathcal{T}$ (mathematically, $L^2(\mathcal{T}) = \{f(t), t \in \mathcal{T} : \int_{\mathcal{T}} f^2(t)dt < \infty\}$) and let $X_g(t) \in L^2(\mathcal{T})$, $g = 1, 2$, denote the gth random process, with a continuous unknown mean function $\mu_g(t)$. Based on random samples of $X_g(t)$, we concentrate on testing $\mu_1(t) = \mu_2(t)$, against the general alternative that the two mean functions are not equal at some time points within the time range $\mathcal{T}$. Testing the equality of mean evolution curves of a variable across two groups is of interest in numerous fields. For example, in a historical study of African slave trade from 1650 to 1890, one of the most important questions that the researchers want to answer is whether the slave population of west Africa is the same as those of east Africa over the 250 years. Similarly, in an AIDS clinical trial, the researchers want to compare the timely effectivenesses of two medical treatments by determining whether the CD4 count curves of the two treatment groups are the same over the entire study period.

For dense functional data, where a large number of regularly-observed measurements for each subject are attainable, some well-developed methods for the two-sample mean function testing problem exist. The pointwise t-test (Ramsay, 2006) simplifies the problem by testing the null hypothesis over the alternative hypothesis at each time point separately. The L^2-norm-based test and the F-type test (Zhang and Chen, 2007; Faraway, 1997; Zhang and Liang, 2014; Zhang et al., 2010) quantify the overall difference between the two mean functions by integrating the standardized sample mean function difference over the entire time range. They prove that the defined random quantity follows an approximate χ^2 distribution. Tests involving basis representations (Abramovich et al., 2004; Lillo et al., 2015) project the sample mean function difference into a common orthonormal space. They argue that the original test problem is equivalent to testing whether the mean of the achieved random score vector is zero. The pseudo likelihood ratio test (Staicu et al., 2014) first represents the random samples in the $L^2(\mathcal{T})$ by penalized spline bases, transforming the test problem into simultaneously testing a vector of polynomial parameters and a vector of random spline coefficients. Then they propose a pseudo likelihood ratio test (pseudo LRT) for this equivalent problem. In addition, tests aimed at detecting differences in functional distributions (Pomann et al., 2016; Hall and Van Keilegom, 2007) are also applicable to functional mean testing problem under further assumptions. Before introducing our proposed method, we first review these test procedures, compare them through numerical studies, and discuss their advantages and disadvantages. These results can be found in Section 2.2.

When it comes to sparse functional data, where only a few irregularly-observed measurements are available for each subject, the two-sample mean testing problem is much more challenging. Most of the previous methods designed for dense functional data cannot be used with sparse data. The L^2-norm-based test and the F-type test involve pre-smoothing for each individual curve such that regular sample mean and covariance functions are attainable. However, pre-smoothing is not reliable considering the limited amount of data for each subject. The projection-based tests require consistent scores obtained from basis expansions for individual curves. However these scores are no longer consistent due to the small number of observations on each curve. To the best of our knowledge, (Staicu et al., 2014) and (Pomann et al., 2016) are the only two dense data tests that are adaptive to sparse data. Their limitations will be detailed later in Section 2.3.

In Section 2.4, we propose a test procedure specially designed for the sparse functional mean

testing problem. The test statistic is based on a shrinkage estimator for the projection score of each subject, conditional on the observing locations. Under the null, the proposed test statistic is proven to follow a χ^2 distribution asymptotically. We conduct several simulation studies and also apply the proposed method to a CD4 count data set and an eBay online auction data set to investigate the numerical performance of the proposed method. We also compare with several dense data methods which are adaptive to sparse functional data. The proposed method is demonstrated to have more satisfactory performances for sparse functional data.

1.2.2 Motivation for one-way functional ANOVA for sparse functional data

In practice, there are oftentimes more than two independent groups being compared. For instance, in the CD4 data set in Chapter 2, the original data set contains four different therapy groups. The research goal is to see which of the four medical treatments is the most effective one at certain time within the study period. When there are multiple groups involved, the mean function testing problem is referred to as one-way functional ANOVA. It tries to determine whether there exist any statistically significant differences among the multiple mean functions. Mathematically, let $L^2(\mathcal{T}) = \{f(t), t \in \mathcal{T} : \int_{\mathcal{T}} f^2(t)dt < \infty\}$, and $X_g(t) \in L^2(\mathcal{T})$, $g = 1, ..., G (G > 2)$ denotes the gth random process, with a continuous unknown mean function $\mu_g(t)$. Based on random samples of $X_g(t)$, we concentrate on testing $\mu_1(t) = ... = \mu_G(t)$, against the general alternative that there exist $i \neq j$ such that $\mu_i(t)$ and $\mu_j(t)$ are not equal at some time points within the time range $\mathcal{T}$.

We generalize the asymptotic χ^2 test proposed in Chapter 2 to multiple group cases, proposing a test statistic which is the summation of multiple quadratic forms. This new test statistic is still asymptotically χ^2 distributed, and its degree of freedom depends on the number of groups in addition to the dimension reduction result from orthogonal projections. Its performance is confirmed by simulation studies and an application to an eBay online auction data set. All of these results can be found in Chapter 3.

1.2.3 Motivation for change point detection for sparse functional time series

In the domain of univariate time series analysis, change point detection is a fundamental problem that is widely of interest. In general, the change point detection problem tries to solve two

tasks: determining whether certain statistical properties of a stochastic process changes significantly at some time points in the series, and identifying the times when such changes happen. Change point detection techniques have been found to be powerful in analyzing time series data in numerous fields, such as climate (Reeves et al., 2007), environment (Siris and Papagalou, 2004), and information technology (Tartakovsky et al., 2006; Weng and Lee, 2011). Driven by different needs, the change point detection problem can be represented in some form or other. There are several types of change point detection problems including the at most one change point problem (AMOC) (Hawkins, 1977; Worsley, 1986) and epidemic changes (Yao, 1993; Ramanayake and Gupta, 2003), where the statistical property changes and then returns to its original level. Most of the time researchers are concerned with detecting possible changes in terms of mean levels, while detecting possible changes in distributions is considered by (Pollak, 1985; Zhou et al., 2017). Traditional change point detection problems assume that random variables in the stochastic process are independently normal distributed. Exponential distribution situations is considered by (Ramanayake and Gupta, 2003; Worsley, 1986). Change point detection in non-independent sequences is studied by (Henderson, 1986; Kim, 1996; Joseph et al., 1996). Among all these variations, the simplest and most classical one is detecting a sequence of independent variables for a shift in mean level (the AMOC problem). Mathematically, X_i is the random variable at time i, with mean $\mu_i = E[X_i]$ and variance σ^2. The hypothesis testing of interest is $\mu_i = \mu$ for all $i = 1, ..., n$ against the alternative that $\mu_i = \mu$ for $i = 1, ..., \theta$ (θ is unknown) and $\mu_i = \mu'$ for $i = \theta + 1, ..., n$. Note that the two means μ and μ' are not equal. Methods regarding this problem are reviewed in Section 4.1.

In this thesis, we consider a functional version of the aforementioned AMOC problem. For the rest of this thesis, we refer to the functional AMOC change point problem when we say functional change point detection. Basically, let $X_1(t), ..., X_i(t), ..., X_n(t)$ be a sequence of independent stochastic processes with mean function $\mu_i(t)$ and covariance function $G(s, t)$, with $s, t \in \mathcal{T} = [0, 1]$. The functional change point detection problem tries to test $\mu_i(t) = \mu(t)$ for all $i = 1, ..., n$ against the alternative that $\mu_i(t) = \mu(t)$ for $i = 1, ..., \theta$ (θ is unknown) and $\mu_i(t) = \mu'(t)$ for $i = \theta + 1, ..., n$. Let's assume that we have the daily precipitation in Pittsburgh from 1951 to 2000. The climate researchers hope to utilize these data to determine whether the precipitation starts to differ since some year within the fifty years. To use the univariate time series change

point methods, one has to summarize the yearly profile by taking an average within a year. This method might lose valuable information and the different choices of pre-aggregation can lead to different testing results. Functional change point detection methods view the yearly profile as a function $X_i(t)$, for $i = 1, ..., 50$. The original problem can be answered by testing whether the mean function changes at some unknown year θ.

For dense functional data, there are several recent literature talking about the AMOC problem, such as, (Berkes et al., 2009) and (Aue et al., 2009). They both utilize functional principal component analysis to approximate the infinite dimensional inference problem by an finite approximate in the eigenspace. As for sparse functional data, the change point detection problem has not been specifically explored. The aforementioned dense data methods no longer work for sparse data. This is because they require consistent individual projection scores, which are no longer achievable when only a few irregularly-spaced observations are available for each curve. In Chapter 4, we propose a testing method specifically designed for the sparse functional AMOC problem. The test statistic is formulated from shrinkage estimators of the individual projection scores, which are conditional on the subject-dependent observing locations. Under the null hypothesis, the proposed test statistic follows a finite summation of independent standard Brownian bridge squares. Two simulation studies are conducted to justify its finite sample performance. And it is also applied to an eBay online auction data set.

The rest of the thesis is organized as follows. Chapter 2 first reviews the two-sample inference problem for dense functional data. It then thoroughly presents the proposed method for sparse functional data cases, including the statistical framework, proposed test statistics, estimation procedures, asymptotic theorems, and numerical experiments. Extension to multiple sample cases, i.e., one-way FANOVA, is investigated in Chapter 3. Chapter 4 contains discussion for the functional change point detection problem. The univariate change point detection problem in univariate time series is briefly reviewed, followed by the proposed method for its counterpart under sparse functional data cases. The proposed sparse functional method is justified by simulation studies and application to an eBay online auction data set. Chapter 5 summarizes the entire thesis with some further discussions. Proofs for the theorems are provided in the Appendix.

2.0 TWO-SAMPLE MEAN FUNCTION INFERENCE FOR SPARSE FUNCTIONAL DATA

For the two-sample mean function testing problem, methodologies for dense regular designs have been relatively well developed. There exist several popular test procedures, including the pointwise t-test (Ramsay, 2006), the L^2-norm-based test and the F-type test (Zhang and Chen, 2007; Faraway, 1997; Zhang and Liang, 2014; Zhang et al., 2010), tests involving basis representations (Abramovich et al., 2004; Horváth and Rice, 2015; Lillo et al., 2015), and the pseudo LRT test (Staicu et al., 2014). In addition, tests aiming at detecting differences in functional distributions (Pomann et al., 2016; Hall and Van Keilegom, 2007) are also applicable to functional mean testing problems under further assumptions. These dense functional data methods are briefly reviewed in Section 2.2. They are invented by different people at different times, there is no universal conclusion to say which one is better than the others. We conduct several simulation studies to provide some insights for their advantages and disadvantages. Reviewing the literature for dense functional data not only helps us to have a thorough understanding about the two-sample mean function inference problem, but also allows us to figure out why or why not each of these methods are applicable to sparse functional data.

Sparse functional data is also very common in real world applications. Inventing two-sample mean function testing procedures for sparse functional data will be beneficial to researchers from all the fields where sparsely and repeatedly observed data are provided for answering the question of general interest: are the average curves of a variable the same across the groups? For instance, it can assist financial analysts to determine whether client's transaction behaviors in stock markets have changed from 2014 to 2015, even though for each client, only a few irregularly distributed transaction data are available in each year. Similarly, it can help researchers to figure out which of the two medical treatments is more effective for advanced AIDS patients, even though there are

only 2 to 10 irregularly-observed measurements for each patient.

Constructing appropriate test procedures for sparsely observed functional samples is rather challenging. Most of the aforementioned methods designed for dense functional data are unfeasible now. The L^2-norm-based test and the F-type test require a dense regular design on each curve, such that the regular sample mean and sample covariance functions are attainable. When observing times are not the same for all the subjects, pre-smoothing for each individual curve is involved. However, pre-smoothing techniques are no longer reliable for sparse functional data, considering the limited amount of data for each subject. The projection-based tests are not applicable to sparse functional data because it is well known that the estimated scores obtained from basis expansions for individual curves are no longer consistent. There are some recent works regarding one-sample inference for sparse functional data (Ma et al., 2012). These methods cannot be easily generalized to two or more functional samples. To the best of our knowledge, the pLRT test in (Staicu et al., 2014) and and the distribution test in (Pomann et al., 2016) are the only two dense data tests that are adaptive to sparse data cases. How to extend these two methods to sparse data cases are briefly summarized in Section 2.3, and their performances are compared with those of our proposed procedure through simulation studies in Section 2.8.1.

In this chapter, we propose a test procedure specially designed for the sparse functional mean testing problem. We propose to construct a test statistic based on a shrinkage eigen-projection score vector. Unlike the regular eigen-projection score vector, the distribution of the individual shrinkage score depends on its specific design, i.e., the observing times. However, we notice that if we take the randomness of designs into consideration, the distribution of the shrinkage score is still i.i.d across all the subjects. Based on the distribution of the estimator of the shrinkage score, we propose a χ^2 type test statistic. We rigorously derive its asymptotic distribution under the null. The development of the asymptotic null distribution is nontrivial since we need to consider the estimation error of the shrinkage estimator for each individual subject, which is itself a random quantity. The overall error needs to be controlled uniformly across all subjects.

The remaining of this chapter is organized as follows. In Section 2.1, we state the statistical framework for the two-sample inference problem. In Section 2.2, we review and compare available tests for dense functional data. In Section 2.3, we concisely explain how to modify the adaptable dense data tests to sparse functional data. In Section 2.4 through Section 2.8, all the details about

the proposed test for sparse functional data are presented, including the derivation, estimation procedures, asymptotic results, simulation studies, and applications to both the CD4 count data set and the eBay online auction data set mentioned before.

2.1 STATISTICAL FRAMEWORK

The model we consider is

$$Y_{gij} = X_{gi}(T_{gij}) + \epsilon_{gij} \tag{2.1}$$

where Y_{gij}, $g = 1, 2, i = 1, .., n_g, j = 1, .., N_{gi}$ denotes the jth observation of the ith subject in group g. The random samples X_{gi} are realizations of two independent stochastic processes with homogeneous covariance structures in a bounded time domain $\mathcal{T}$. Without loss of generality, let's assume $\mathcal{T} = [0, 1]$. Basically, $X_{gi}(t)$ are random samples from $X_g(t) \sim SP(\mu_g(t), G(s, t))$, $s, t \in \mathcal{T} = [0, 1]$. Here $\epsilon_{gij} \sim N(0, \sigma^2)$ are i.i.d measurement errors. The number of observations for the ith subject in group g is denoted as N_{gi}. The corresponding observing times are $\{T_{gi1}, ..., T_{giN_{gi}}\}$.

For dense functional data, the number of observations N_{gi} is relatively large which goes to infinity with sample size n_g with a high rate, and the observing times $\{T_{gi1}, ..., T_{giN_{gi}}\}$ are usually regularly-spaced. When observing times are not the same for all the subjects, one can use pre-smoothing techniques to get a consistent estimate for the entire underlying curve $X_{gi}(t)$, then evaluate all subjects at a common dense regular grid. Consequently, for dense functional data, the number of observations are often assumed to be the same across subjects, i.e., $N_{gi} = N$. The common observing times $\{T_1, ..., T_N\}$ are usually assumed to be fixed and equally spaced without loss of generality.

For sparse functional data, the number of observations N_{gi} is assumed to be finite or grow slowly with sample size n_g. Given N_{gi}, random observing times $\{T_{gi1}, ..., T_{giN_{gi}}\}$ are i.i.d with a bounded density function (specified in Section 2.7) within the time domain $\mathcal{T}$. It is also assumed that X_{gi}, N_{gi} and ϵ_{gij} are mutually independent.

Based on the observed data Y_{gij}, $g = 1, 2$, $i = 1, ..., n_g$, $j = 1, ..., N_{gi}$, the two-sample mean function testing problem is interested in testing

$$H_0 : \mu_1(t) = \mu_2(t),\ t \in \mathcal{T} \quad VS \quad H_a : \exists\ t \in \mathcal{T},\ \mu_1(t) \neq \mu_2(t). \tag{2.2}$$

2.2 REVIEWS FOR DENSE FUNCTIONAL DATA

2.2.1 Pointwise t-test

The pointwise test is introduced in (Ramsay, 2006). Its key idea is to test the null hypothesis against the alternative hypothesis at each time point $t_0 \in \mathcal{T}$. In practice, one actually only can observe data on a grid of time points, so they need to first apply some pre-smoothing techniques, such as kernel, spline, and local polynomial, to each of the observed individual curve $(Y_{gi1}, ..., Y_{giN})$ to consistently recover the underlying curve in the entire domain $t \in \mathcal{T}$. Let's denote the reconstructed curves as $\check{X}_{gi}(t)$. These recovered curves are $\sqrt{n}$ consistent in supnorm to the true curves, therefore they are usually assumed to have the same properties as the true curve $X_{gi}(t)$. Under Gaussian processes, for each time point $t_0 \in \mathcal{T}$, the pointwise t-test use $t_n(t_0) = \frac{\bar{\check{X}}_{1\cdot}(t_0) - \bar{\check{X}}_{2\cdot}(t_0)}{\sqrt{(1/n_1+1/n_2)\hat{G}(t_0,t_0)}} \sim t(n_1+n_2-2)$, where $\hat{G}(t_0,t_0) = \frac{1}{n_1+n_2-1}\sum_{g=1}^{2}\sum_{i=1}^{n_g}[\check{X}_{gi}(t_0) - \bar{\check{X}}_{g\cdot}(t_0)][\check{X}_{gi}(t_0) - \bar{\check{X}}_{g\cdot}(t_0)]$, to test

$$H_0 : \mu_1(t_0) = \mu_2(t_0) \quad VS \quad H_a : \mu_1(t_0) \neq \mu_2(t_0). \tag{2.3}$$

And they propose to reject the H_0 in (2.2) whenever there exist some time point t_0 such that $|t_n(t_0)| > t_{n_1+n_2-2}(1-\alpha/2)$, for any predetermined significant level α.

It is unrealistic to test (2.3) over all points in the entire domain. Empirically, people pre-specific a finite set of grids on which they want to test. And the H_0 in (2.2) is rejected, if (2.3) is rejected for at least one time points in the pre-specific grids set.

There are four comments that we want to make. First, Gaussian processes are not required by the pointwise t-test. Under non-Gaussian cases, the test can be replaced by a pointwise z-test, as long as both n_1 and n_2 are large enough. Second, the pointwise t-test is user-dependent. A set of grids needs to be pre-specified by users. Third, even though our simulation studies show that pre-smoothing is not necessary when the subjects share a common observing grid and the measurement errors are not too large, it is required when the subjects do not share a common observing grid. Pre-smoothing for each individual subject could be very time consuming. Finally, the pointwise test cannot guarantee the overal significance level α. All the other global tests in Section 2.2.2 to Section 2.2.6 can approximately control the type I errors. This point is clearly

illustrated by simulation studies in Section 2.2.7, where the type I errors of the pointwise t-test are much larger than the significance level $\alpha = 0.05$, while the type I errors of the L^2-norm-based test, the globalized F-test, and the eigen-space projection method are all successfully controlled.

2.2.2 L^2-norm-based test

The L^2-norm-based test is described in detail by (Zhang, 2013). Similar to the pointwise t-test, it is created based on the assumption that the entire curves $X_{gi}(t)$ are obtainable. So the first step is again to achieve consistent curve estimates $\check{X}_{gi}(t)$. The L^2-norm-based test is motivated by the fact that the integration of the squared difference between the two group's sample mean functions is a reasonable statistical quantity to measure the overall difference between their population mean functions. The explicit formula of their test statistic is

$$L_n = \frac{n_1 n_2}{n} \int_{\mathcal{T}} [\bar{\check{X}}_{1\cdot}(t) - \bar{\check{X}}_{2\cdot}(t)]^2 dt. \tag{2.4}$$

where $n = n_1 + n_2$, $\bar{\check{X}}_{g\cdot}(t) = \frac{1}{n_g} \sum_{i=1}^{n_g} \check{X}_{gi}(t)$. Under Gaussian processes cases, they prove that

$$L_n \overset{D}{=} \sum_{k=1}^{p} \lambda_k A_k, \text{ with } A_k \sim \text{i.i.d } \chi^2(1) \tag{2.5}$$

where '$\overset{D}{=}$' denotes equal in distribution, $\lambda_1, ..., \lambda_p$ are the positive eigenvalues of the common covariance function $G(s,t)$. They propose to use the Welch-Scatterthwaite χ^2-approximation to get the following approximate for L_n,

$$L_n \sim \beta \chi^2(d) \text{ approximately , where } \beta = \frac{\text{tr}(G^{\bigotimes 2})}{\text{tr}(G)}, \; d = \frac{\text{tr}^2(G)}{\text{tr}(G^{\bigotimes 2})} \tag{2.6}$$

Note that $\text{tr}(G) = \sum_{k=1}^{p} \lambda_k$, and $\text{tr}(G^{\bigotimes 2}) = \sum_{k=1}^{p} \lambda_k^2$.

In terms of the numerical implementation of the L^2-norm-based test, they propose to evaluate values of the test statistics L_n by

$$L_n = \frac{n_1 n_2}{n} \left\| \bar{\check{X}}_{1\cdot}(t) - \bar{\check{X}}_{2\cdot}(t) \right\|_2^2 \approx \frac{n_1 n_2}{n} \frac{1}{N} \sum_{j=1}^{N} (\bar{\check{X}}_{1\cdot}(T_j) - \bar{\check{X}}_{2\cdot}(T_j))^2, \tag{2.7}$$

and $\text{tr}(G)$ and $\text{tr}(G^{\bigotimes 2})$ are estimated by

$$\text{tr}(\hat{G}) = \int_{\mathcal{T}} \hat{G}(t,t) dt \approx \frac{1}{N} \sum_{j=1}^{N} \hat{G}(T_j, T_j) \tag{2.8}$$

$$\text{tr}(\hat{G}^{\bigotimes 2}) = \int_{\mathcal{T}^2} \hat{G}^2(s,t)dsdt \approx \frac{1}{N^2}\sum_{j=1}^{N}\sum_{j'=1}^{N}\hat{G}^2(T_j,T_{j'}) \tag{2.9}$$

where $\hat{G}(s,t) = \frac{1}{n_1+n_2-1}\sum_{g=1}^{2}\sum_{i=1}^{n_g}[\check{X}_{gi}(s) - \bar{\check{X}}_{g\cdot}(s)][\check{X}_{gi}(t) - \bar{\check{X}}_{g\cdot}(t)]$.

Under the Gaussian process assumption, the L^2-norm-based test statistic is shown to be a mixture χ^2 distribution. Empirically, it is approximated by a single χ^2 distribution, based on the Welch-Scatterthwaite method. So it is not an exact test, and as shown by the simulation studies in Section 2.2.7, performances of this approximation works well with relative large sample sizes. Under non-Gaussian cases, as long as n_1 and n_2 are both large enough, the distribution of the test statistic is approximately a mixture χ^2 distribution according to the functional Central Limit Theorem.

2.2.3 Globalized F-test

The globalized F-test is studied in (Zhang and Liang, 2014). The basic idea is to come up with a globalized version of the pointwise F-test by using its integral over $\mathcal{T}$,

$$F_n = \int_{\mathcal{T}} \frac{\sum_{g=1}^{2} n_g[\bar{\check{X}}_{g\cdot}(t) - \bar{\check{X}}_{\cdot\cdot}(t)]^2}{\sum_{g=1}^{2}\sum_{i=1}^{n_g}[\check{X}_{gi}(t) - \bar{\check{X}}_{g\cdot}(t)]^2/(n-2)}dt. \tag{2.10}$$

Under Gaussian processes and several other regular assumptions, they prove that

$$F_n \xrightarrow{D} \sum_{k=1}^{\infty}\lambda'_k A_k, \;\; A_k \sim \text{i.i.d } \chi^2(1) \tag{2.11}$$

where '$\xrightarrow{D}$' denotes converge in distribution, $\lambda'_k, k = 1, ..., \infty$ are the non-increasing eigenvalues of $G_\omega(s,t) = G(s,t)/\sqrt{G(s,s)G(t,t)}$. Similar to the L^2-norm-based test in Section 2.2.2, the asymptotic distribution of F_n can be approximated by the Welch-Scatterthwaite χ^2-approximation. The resulting approximation is

$$F_n \xrightarrow{D} \beta_\omega \chi^2(d_\omega) \text{ approximately}, \tag{2.12}$$

where $\beta_\omega = \text{tr}\,(G_\omega^{\bigotimes 2})$ and $d_\omega = \frac{1}{\text{tr}\,(G_\omega^{\bigotimes 2})}$. Note that $\text{tr}(G_\omega^{\bigotimes 2}) = \sum_{k=1}^{p}\lambda'^2_k$.

The numerical implement for the Globalized F-test is very similar to the L^2-norm-based test in Section 2.2.3. The empirical F_n is evaluated by

$$F_n = \int_{\mathcal{T}} \frac{\sum_{g=1}^{2} n_g [\bar{\check{X}}_{g\cdot}(t) - \bar{\check{X}}_{\cdot\cdot}(t)]^2}{\sum_{g=1}^{2} \sum_{i=1}^{n_g} [\check{X}_{gi}(t) - \bar{\check{X}}_{g\cdot}(t)]^2/(n-2)} dt \approx \frac{1}{N} \sum_{j=1}^{N} \frac{\sum_{g=1}^{2} n_g [\bar{\check{X}}_{g\cdot}(T_j) - \bar{\check{X}}_{\cdot\cdot}(T_j)]^2}{\sum_{g=1}^{2} \sum_{i=1}^{n_g} [\check{X}_{gi}(T_j) - \bar{\check{X}}_{g\cdot}(T_j)]^2/(n-2)}, \tag{2.13}$$

and $\text{tr}(G_\omega)$ and $\text{tr}(G_\omega^{\bigotimes 2})$ are estimated by

$$\text{tr}(\hat{G}_\omega) = \int_{\mathcal{T}} \hat{G}_\omega(t,t)dt \approx \frac{1}{N} \sum_{j=1}^{N} \hat{G}_\omega(T_j, T_j) \tag{2.14}$$

$$\text{tr}(\hat{G}^{\bigotimes 2}) = \int_{\mathcal{T}^2} \hat{G}_\omega^2(s,t)dsdt \approx \frac{1}{N^2} \sum_{j=1}^{N} \sum_{j'=1}^{N} \hat{G}_\omega^2(T_j, T_{j'}) \tag{2.15}$$

where $\hat{G}_\omega(s,t) = \frac{\hat{G}(s,t)}{\sqrt{\hat{G}(s,s)\hat{G}(t,t)}}$, with $\hat{G}(s,t) = \frac{1}{n_1+n_2-1} \sum_{g=1}^{2} \sum_{i=1}^{n_g} [\check{X}_{gi}(s) - \bar{\check{X}}_{g\cdot}(s)][\check{X}_{gi}(t) - \bar{\check{X}}_{g\cdot}(t)]$.

The globalized F-test improves upon the pointwise t-test and can control the overall type I error. Similar to the L^2-norm-based test in Section 2.2.2, the functional Central Limit Theorem ensures that F_n is still approximately mixture χ^2 distributed as long as n_1 and n_2 are both large enough.

2.2.4 Projection-based distance test

The eigen-space projection based test is considered by (Horváth and Kokoszka, 2012; Horváth and Rice, 2015; Lillo et al., 2015). This method is based on the eigen-decomposition of the common covariance function $G(s,t)$, i.e., $G(s,t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t)$, where $\{\phi_k(t), k = 1, 2, ..., \infty\}$ are the orthonormal eigenfunctions of $G(s,t)$, corresponding to the non-increasing eigenvalue sequence $\{\lambda_k, k = 1, 2, ..., \infty\}$. They first project the difference between the sample mean functions to the eigen-space spanned by the common eigenfunctions $\{\phi_k(t), k = 1, ..., \infty\}$. Mathematically, the projection score is $\upsilon_k =< \bar{\check{X}}_{1\cdot} - \bar{\check{X}}_{2\cdot}, \phi_k >, k = 1, ..., \infty$, where $\bar{\check{X}}_{g\cdot}(t) = \frac{1}{n_g} \sum_{i=1}^{n_g} \check{X}_{gi}(t)$.

Based on some regularization assumptions on the mean functions $\mu_g(t)$ and the eigenvalues $\{\lambda_k, k = 1, 2, ..., \infty\}$, they are able to argue that the first several dominant scores capture the

majority of the signals of $\bar{\breve{X}}_{1\cdot} - \bar{\breve{X}}_{2\cdot}$. Let p denote the number of random scores to keep. Under H_0, it is easy to get the following property

$$\sum_{k=1}^{p} \frac{\upsilon_k^2}{\lambda_k} \sim \chi^2(p) \tag{2.16}$$

In practice, $\{\lambda_k, \phi_k(t)\}_{k=1}^p$ can be estimated by computing the eigenvalues and eigenfunctions for the $N \times N$ sample covariance matrix of random vectors $(\breve{X}_{gi}(T_1), ..., \breve{X}_{gi}(T_N))$, $g = 1, 2; i = 1, ..., n_g$. The corresponding estimated eigenvalues and eigenfunctions of covariance matrix of $(\breve{X}_{gi}(T_1), ..., \breve{X}_{gi}(T_N))$ are denoted as $\{\hat{\breve{\lambda}}_k, \hat{\breve{\phi}}_k(t)\}_{k=1}^p$. Then the estimate for eigenvalues and eigenfunctions of $G(s,t)$ are

$$\hat{\lambda}_k = \hat{\breve{\lambda}}_k \tag{2.17}$$

$$\hat{\phi}_k(t) = \sqrt{N}\hat{\breve{\phi}}_k(t) \tag{2.18}$$

Let $\hat{\upsilon}_k =< \bar{\breve{X}}_{1\cdot} - \bar{\breve{X}}_{2\cdot}, \hat{\phi}_k >\approx \frac{1}{N}\sum_{j=1}^{N}(\bar{\breve{X}}_{1\cdot}(T_j) - \bar{\breve{X}}_{2\cdot}(T_j))\hat{\phi}_k(T_j)$, then they propose the following test statistic

$$D_n = \sum_{k=1}^{p} \frac{\hat{\upsilon}_k^2}{\hat{\lambda}_k} \xrightarrow{D} \chi^2(p). \tag{2.19}$$

They comment on p by arguing that p can be chosen by the classical fraction of variance explained (FVE) approach. For instance, if the threshold is 0.80, then the FVE method takes the first p leading principal components such that they can explain at least 80% of the total variance.

The eigen-space projection based test needs to estimate the eigenvalue and eigenfunctions, which usually can be efficiently and consistently estimated. It also requires additional regularization assumptions to ensure that the signals of the mean functions are be mostly captured by the projection scores of the first several eigen-directions, and some addition techniques are involved to determine the appropriate dimensions of the random score vectors. The test statistic is based on a quadratic form of the estimated projection scores. The derivation of the null distribution relies on consistency results for $\hat{\upsilon}_k$. As we will discuss later, this only works for densely observed functional data with moderate noise levels. Gaussian processes are not required by eigen-space projection based test. The Central Limit Theorem ensures that the test statistic converges to a χ^2 distribution regardless of whether the random functions are Gaussian or not.

2.2.5 Pseudo likelihood ratio test

The Pseudo Likelihood Ratio Test (pLRT) is investigated in (Staicu et al., 2014). Their model is a delicate version of (2.1). According to Mercer's theorem, they have the spectral decomposition, $G(s,t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s)\phi_k(t)$, where $\{\phi_k(t), k = 1, 2, ..., \infty\}$ are the orthonormal eigenfunctions of $G(s,t)$, corresponding to the non-increasing eigenvalue sequence $\{\lambda_k, k = 1, 2, ..., \infty\}$. Then based on the Karhunen-Loève expansion, they have $X_{gi}(t) = \mu_g(t) + \sum_{k=1}^{\infty} \xi_{gk}\phi_k(t)$, with $\xi_{gk} = \int_{\mathcal{T}}(X_{gi}(t) - \mu_g(t))\phi_k(t)dt$. Under the additional assumption that the overall mean function of the two groups is $\mu_{pool}(t)$, they have the following model

$$Y_{gij} = \mu_{pool}(T_j) + [\mu_g(T_j) - \mu_{pool}(T_j)] + \sum_{k=1}^{\infty} \xi_{gk}\phi_k((T_j)) + \epsilon_{gij}. \tag{2.20}$$

By several regularization assumptions, they have an estimated $\mu_{pool}(t)$, $\hat{\mu}_{pool}(t) = \frac{\hat{\mu}_1(t)+\hat{\mu}_2(t)}{2}$, where $\hat{\mu}_g(t)$ is the mean functional estimate based on the gth sample alone. Let $\tilde{Y}_{gij} = Y_{gij} - \hat{\mu}_{pool}(T_{gij})$. And they assume that μ_{pool} is estimated well enough, such that model (2.20) becomes

$$\tilde{Y}_{gij} = [\mu_g(T_j) - \mu_{pool}(T_j)] + \sum_{k=1}^{\infty} \xi_{gk}\phi_k((T_j)) + \epsilon_{gij}. \tag{2.21}$$

Observe that testing (2.2) is the same as testing

$$H_0 : \mu_1(t) - \mu_{pool}(t) = 0,\ t \in \mathcal{T} \quad VS \quad H_a : \exists\ t \in \mathcal{T}, \mu_1(t) - \mu_{pool}(t) \neq 0, \tag{2.22}$$

the two-sample test problem (2.2) is transformed into an one-sample inference problem.

In order to test (2.22), they suppose that $\mu_1(t) - \mu_{pool}(t)$ can be represented by the penalized spline class of functions,

$$\mu_1(t) - \mu_{pool}(t) = \beta_0 + \beta_1 t + ... + \beta_p t^q + \sum_{s=1}^{S} b_s(t - \kappa_s)_+^q \tag{2.23}$$

where $x_+^q = \max(0, x)^q$, $\kappa_1, ..., \kappa_S$ are the knots placed at equally spaced quantiles, and $\mathbf{b} = (b_1, ..., b_S)^T \sim N(\mathbf{0}, \sigma_b^2\mathbf{I})$. Then model (2.21) becomes

$$\tilde{\mathbf{Y}}_i = \tilde{\mathbf{X}}_i\boldsymbol{\beta} + \tilde{\mathbf{Z}}_i\mathbf{b} + \mathbf{e}_i. \tag{2.24}$$

where $\tilde{\mathbf{Y}}_i = (\tilde{\mathbf{Y}}_{1i}^T, \tilde{\mathbf{Y}}_{2i}^T)^T = (Y_{1i1}, ..., Y_{i1N}, Y_{2i1}, ..., Y_{2iN})^T$, $\tilde{\mathbf{X}}_i = [\tilde{\mathbf{X}}_{1i}^T | -\tilde{\mathbf{X}}_{2i}^T]^T$ ($\tilde{\mathbf{X}}_{gij} = (1, T_j, ..., T_j^q)$), $\tilde{\mathbf{Z}}_i = [\tilde{\mathbf{Z}}_{1i}^T | -\tilde{\mathbf{Z}}_{2i}^T]^T$ ($\tilde{\mathbf{Z}}_{gij} = ((T_j - \kappa_1)_+^q, ..., (T_j - \kappa_S)_+^q)$), and the covariance matrix of $\mathbf{e}_i$ is

$\boldsymbol{\Sigma}_i = diag(\boldsymbol{\Sigma}_{1i}, \boldsymbol{\Sigma}_{2i})$, the (j, j') element of $\boldsymbol{\Sigma}_{gi}$ equals to $G(T_j, T_{j'}) + \sigma^2 1(j = j')$. And the test problem (2.22) becomes

$$H_0 : \boldsymbol{\beta} = \mathbf{0} \text{ and } \sigma_b^2 = 0 \quad VS \quad H_a : \text{not } H_0, \tag{2.25}$$

Let $N = \sum_{g=1}^{2} \sum_{i=1}^{n_g} N_{gi}$. They propose to use the following pseudo LRT statistics

$$pLRT_N = \sup_{H_0 \bigcup H_a} 2 \log L_{\tilde{\mathbf{Y}}}(\boldsymbol{\beta}, \sigma_b^2) - \sup_{H_0} 2 \log L_{\tilde{\mathbf{Y}}}(\boldsymbol{\beta}, \sigma_b^2) \tag{2.26}$$

where $L_{\tilde{\mathbf{Y}}}(\boldsymbol{\beta}, \sigma_b^2)$ is likelihood function. Under Gaussian cases, the asymptotic null distribution of $pLRT_N$ is proven to be the same as

$$\sup_{\eta \geq 0} [\sum_{s=1}^{S} \frac{\eta}{1 + \eta \zeta_s} - \sum_{s=1}^{S} \log(1 + \eta \gamma_s)] + \sum_{j=1}^{q+1} \nu_j^2 \tag{2.27}$$

where $\nu_j \sim N(0, 1)$, and let γ_{sN} and ζ_{sN} are eigenvalues of $N^{-\rho} \tilde{\mathbf{Z}}^T \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{Z}}$ and $N^{-\rho} \{\tilde{\mathbf{Z}}^T \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{Z}} - \tilde{\mathbf{Z}}^T \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{X}} (\tilde{\mathbf{X}}^T \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{Z}}\}$ respectively, and $\gamma_{sN} \xrightarrow{p} \gamma_s$, $\zeta_{sN} \xrightarrow{p} \zeta_s$.

The numerical implementation of the pLRT test can be done by using the R function 'exactLRT' in the package 'RLRsim' (Scheipl et al., 2008), which not only calculates the value of the pLRT test statistic but also determines the p-value by approximating its asymptotic null distribution.

The pLRT test solves the problem in the traditional likelihood ratio test prospective. However, the pLRT test suffers from several drawbacks. First, it is based on the fact that the mean functions can be represented by the penalized spline class of functions. This is not always the case in real practice, and then the pLRT test may be biased. Second, as pointed by the author, the pLRT test require a relatively high level of accuracy of $\hat{\mu}_g(t)$ and $\hat{G}(s, t)$. This is a rather strict requirement, especially for sparse functional data. Third, the asymptotic null distribution of the test statistic depends on the limiting eigenvalues of two complex sequences of matrices, which is inconvenient and unstable in practice. This shortcoming is demonstrated in our simulation studies in Section 2.8.1. Finally, unlike the tests discussed in Section 2.2.1 to Section 2.2.4, the Gaussian assumption is required for the pLRT test.

2.2.6 Distribution test

This distribution test is proposed in (Pomann et al., 2016). It is originally invented to test the null hypothesis,

$$H_0 : X_1(t) \overset{D}{=} X_2(t) \quad VS \quad H_a : X_1(t) \overset{D}{\neq} X_2(t), \tag{2.28}$$

where '$\overset{D}{=}$' denotes equal in distribution as before.

In order to test (2.28), they denote $X(t)$ as the mixture process of $X_1(t)$ and $X_2(t)$, with mean function $\mu_{pool}(t)$ and pooled covariance function $\tilde{G}(s,t)$. Similar to the pLRT test, based on the Karhunen-Loève expansion, they have $X_g(t) = \mu_{pool}(t) + \sum_{k=1}^{\infty} \omega_{gk}\eta_k(t)$, with $\omega_{gk} = \int_{\mathcal{T}}(X_g(t) - \mu_{pool}(t))\eta_k(t)dt$, where $\eta_k(t)$ are the eigenfunctions of the pooled covariance $\tilde{G}(s,t)$. Then it is straightforward that testing (2.28) is equivalent to testing

$$H_0 : \{\omega_{1k}\}_{k=1}^{\infty} \overset{D}{=} \{\omega_{2k}\}_{k=1}^{\infty} \quad VS \quad H_a : \{\omega_{1k}\}_{k=1}^{\infty} \overset{D}{\neq} \{\omega_{2k}\}_{k=1}^{\infty}. \tag{2.29}$$

Similar to the argument in Section 2.2.4, they argue that they can consider only the first p projection scores, i.e., $\{\omega_{g1}, ..., \omega_{gp}\}$. For each of the p dimensions, they propose to utilize the Anderson-Darling test, a distribution test to test the quality of distributions. Note that multiple comparison adjustments, such as the Bonferroni correction, should be used to ensure the overall significance.

Under additional Gaussian assumptions and common covariance assumptions, i.e., $X_g(t) \sim GP(\mu_g(t), G(s,t))$, we notice that testing (2.28) is equivalent to testing (2.2). However, the distribution test tends to yield lower powers for the mean function testing problem, because of the nonparametric nature of the AD test and the multiple testing correction. In order to apply the distribution test to the mean function testing problem, the Gaussian assumption is required.

2.2.7 Comparisons through simulation studies

Simulation I

In this simulation, we compare the performances of the pointwise t-test, the L^2- norm-based test, the globalized F-test and the projection-based test. We compute the type I error and the powers of these tests through 1000 repetitions and summarize the results in Table 2.1. The mean function in this simulation is generated as linear combinations of eigenfunctions, therefore we

know the true value of p in the projection based method. A more complex mean function setting is considered in simulation II.

The data is generated based on model (2.1), and the Karhunen-Loève expansion $X_{gi}(T_j) = \mu_g(T_j) + \sum_{k=1}^{\infty} \xi_{gik}\phi_k(T_j)$, with $\xi_{gik} \sim N(0, \lambda_k)$. The observing time points, $\{T_1, ..., T_N\}$ are equally spaced within the interval [0,1]. We assume that the eigenvalues are $\lambda_k = (k+1)^{-2}$ for $k = 1, ..., 4$, and $\lambda_k = 0$ for $k > 4$. The eigenfunctions are $\phi_k(t) = \sqrt{2}\cos[(k-1)\pi t]$ for $k \geq 2$ and $\phi_1(t) = 1$. The first four eigenfunctions are visualized on the left panel in Figure 2.3. The mean functions are $\mu_1(t) = \sum_{k=1}^{4} a_{1k}\phi_k(t)$ and $\mu_2(t) = \sum_{k=1}^{4} a_{2k}\phi_k(t)$, where $a_{gk} = \frac{3\bar{a}_{gk}}{\|\bar{a}_g\|_2}$, with $\bar{a}_g = (\bar{a}_{g1}, ..., \bar{a}_{g4})^T$ and $\bar{a}_{1k} = (2)^{5-k}(5-k)^6$ and $\bar{a}_{2k} = (2-\varsigma)^{5-k}(5-k)^6$. Note that $\varsigma = 0, 0.4, 0.8, 1.2$ correspond to the null hypothesis and three alternative hypotheses for (2.2). The variance of the measurement errors takes two values, $\sigma = 0$ and $\sigma = 0.3$. The sample sizes under consideration are $n_1 = n_2 = 100$, and $n_1 = n_2 = 200$. The number of observations on each curve is chosen to be $N = 100$ or $N = 50$. Figure 2.1 and Figure 2.2 visualize 9 randomly selected subjects for situations where $(N = 100, \sigma = 0)$ and $(N = 100, \sigma = 0.3)$.

All the simulation results are summarized in Table 2.1, with all numbers representing the percentages of rejection (at significance level $\alpha = 0.05$) calculated based on 1000 repetitions. We use the true value of p, i.e., $p = 4$, to perform the 'projection-based' test. First, indicated by all the three sections in the table, the pointwise t-test is incapable of controlling the pre-specified significance level. The performance of all the other three tests ('L^2-norm-based', 'Globalized', 'Projection-based-test') are reasonable: they not only successfully control the type I error, but also have increased powers with the increase of either sample sizes or discrepancies between the two mean functions. The powers of 'L^2-norm-based' and 'Globalized' are comparable. The 'Projection-based test' is much more powerful than 'L^2-norm-based' and 'Globalized'. The projection-based distance-test is essentially utilizing the same quantity as the L^2-norm-based statistic. However, the L^2-norm-based statistic considers all the positive eigenvalues and the projection-based method truncate at the first few large eigenvalues. The truncation, if employed appropriately, can serve as a de-noising step and therefore results in better performance. Second, by comparing the first part and the second part in Table 2.1, we can see that performances of all the three tests are not affected by moderate random measurement errors. Finally, the first part and the third part of Table 2.1 together confirm that these three tests are not sensitive to the number of observations on each individual

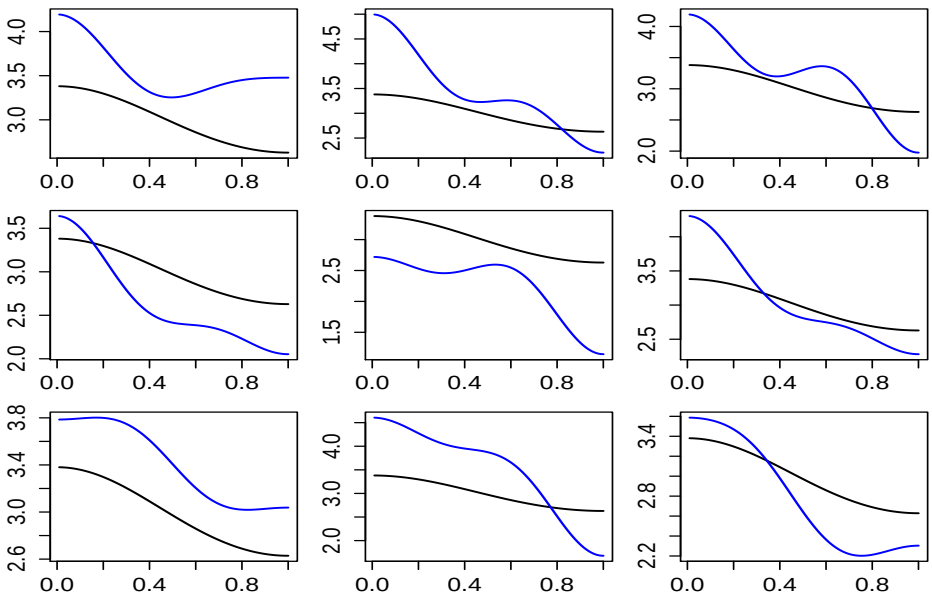

Figure 2.1: $(N = 100, \sigma = 0)$. Simulated data for nine randomly chosen subjects in group 1. Note that in each plot, the black curve is the mean curve for this subject, and the blue curve represent the observed curve.

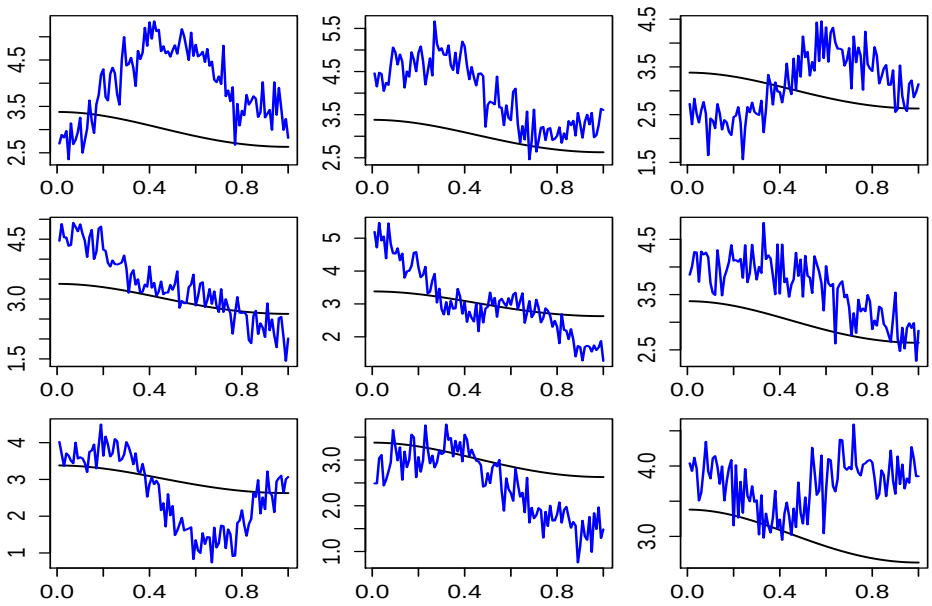

Figure 2.2: ($N = 100$, $\sigma = 0.3$). Simulated data for nine randomly chosen subjects in group 1. Note that in each plot, the black curve is the mean curve for this subject, and the blue curve represent the observed curve.

curve, as long as the observations are dense enough.

Table 2.1: Results for simulation study I (p=4). We presented results for $(N, \sigma) = (100, 0)$, $(100, 0.3)$, $(50, 0)$, where N is the number of grids on each curve and σ is the standard deviation of random error. Sample sizes are $n_1 = n_2 = 100$ or 200. The type I error ($\varsigma = 0$) and powers ($\varsigma = 0.4, 0.8, 1.2$) for four different methods: the pointwise t-test ('pointwise'), the L^2-norm-based test, the Globalized F-test, the projection-based test are calculated based on 1000 repetitions.

$(\mathbf{N}, \boldsymbol{\sigma})$	**sample size**	**Test**	$\boldsymbol{\varsigma = 0}$	$\boldsymbol{\varsigma = 0.4}$	$\boldsymbol{\varsigma = 0.8}$	$\boldsymbol{\varsigma = 1.2}$
$(N, \sigma) = (100, 0)$	$(n_1, n_2) = (100, 100)$	Pointwise	*0.156*	0.252	0.800	1.000
		L^2-norm-based	0.046	0.059	0.357	**1.000**
		Globalized	0.047	0.062	0.331	**1.000**
		Projection-based	0.053	**0.168**	**0.862**	**1.000**
	$(n_1, n_2) = (200, 200)$	Pointwise	*0.150*	0.404	0.997	1.000
		L^2-norm-based	0.045	0.102	0.816	**1.000**
		Globalized	0.044	0.099	0.785	**1.000**
		Projection-based	0.062	**0.307**	**0.993**	**1.000**
$(N, \sigma) = (100, 0.3)$	$(n_1, n_2) = (100, 100)$	Pointwise	*0.392*	0.552	0.951	1.000
		L^2-norm-based	0.049	0.062	0.349	**1.000**
		Globalized	0.050	0.063	0.332	**1.000**
		Projection-based	0.060	**0.191**	**0.860**	**1.000**
	$(n_1, n_2) = (200, 200)$	Pointwise	*0.381*	0.673	0.996	1.000
		L^2-norm-based	0.055	0.116	0.797	**1.000**
		Globalized	0.054	0.108	0.770	**1.000**
		Projection-based	0.059	**0.333**	**0.992**	**1.000**
$(N, \sigma) = (50, 0)$	$(n_1, n_2) = (100, 100)$	Pointwise	*0.159*	0.272	0.807	1.000
		L^2-norm-based	0.046	0.068	0.367	**1.000**
		Globalized	0.045	0.067	0.341	**1.000**
		Projection-based	0.051	**0.193**	**0.869**	**1.000**
	$(n_1, n_2) = (200, 200)$	Pointwise	*0.153*	0.600	0.999	1.000
		L^2-norm-based	0.045	0.105	0.800	**1.000**
		Globalized	0.048	0.100	0.770	**1.000**
		Projection-based	0.047	**0.314**	**0.994**	**1.000**

Simulation II

For the second simulation study, we consider a more complex setting. We want to explore the performance of the four test procedures when the mean functions are not linear combinations of the eigenfunctions. That is to say, we do not know the value of the true p, the number of dominant random scores after projecting the mean functions into the common eigen-space. We adopt the the classical fraction of variance explained (FVE) approach. The specific threshold levels that we use is 80% ('FVE80') and 90% ('FVE90'). The mean functions are $\mu_1(t) = 1 + 2.3t + 3.4t^2 + 1.5t^3$

and $\mu_2(t) = \mu_1(t) + (\frac{1}{30} + \frac{2}{30}t + \frac{3}{30}t^2 + \frac{4}{30}t^3)\eta$, where $\eta = 0, 0.13, 0.26, 0.4$, correspond to the null hypothesis and three alternative hypotheses. The mean functions are visualized on the right panel in Figure 2.3. All the other settings are almost the same as Simulation I. The number of observations on each curve equal 100, i.e, $N = 100$. The observing time points, $\{T_1, ..., T_N\}$ are equally spaced within the interval [0,1], i.e., $\{T_1, ..., T_N\} = \{\frac{1}{100}, \frac{2}{100}, ..., 1\}$. We assume that the eigenvalues are $\lambda_k = (k+1)^{-2}$ for $k = 1, ..., 4$, and $\lambda_k = 0$ for $k > 4$. The eigenfunctions are $\phi_k(t) = \sqrt{2}\cos[(k-1)\pi t]$ for $k \geq 2$ and $\phi_1(t) = 1$. The first four eigenfunctions are visualized on the left panel in Figure 2.3. The variance of the measurement errors is $\sigma = 0.3$. The sample sizes are $n_1 = n_2 = 100$, and $n_1 = n_2 = 200$.

All the simulation results are summarized in Table 2.2, with all numbers representing the percentages of rejection (at significance level $\alpha = 0.05$) calculated based on 1000 repetitions. Shown by all Table 2.2, the pointwise t-test is incapable of controlling the pre-specified significance level. The performance of all the other tests ('L^2-norm-based', 'Globalized', 'Projection-based (FVE80)', and 'Projection-based (FVE90)') are reasonable: they successfully control the type I error, but have increased powers with the increment of either sample sizes or discrepancies between the two mean functions. The powers of 'L^2-norm-based' and 'Globalized' are comparable. The 'Projection-based test' is more powerful, even though the true value of p is no longer obtainable.

2.3 ADAPTATIONS OF DENSE DATA METHODS TO SPARSE CASES

In this section, we briefly introduce how the 'pLRT test' and the 'Distribution' test can be adaptive to sparse data.

Conceptually, the 'pLRT' test can be straightforwardly extended to the sparse data, which requires essentially no change in the test statistic. Different assumptions are required to make the null distribution result holds. More details can be found in their paper (Staicu et al., 2014). The numerical implementation of the 'pLRT' method involves estimation of the mean function and covariance function, which is simply the sample mean and sample covariance estimates in dense functional data case. For sparse functional data, one could use methods developed in (Yao et al.,

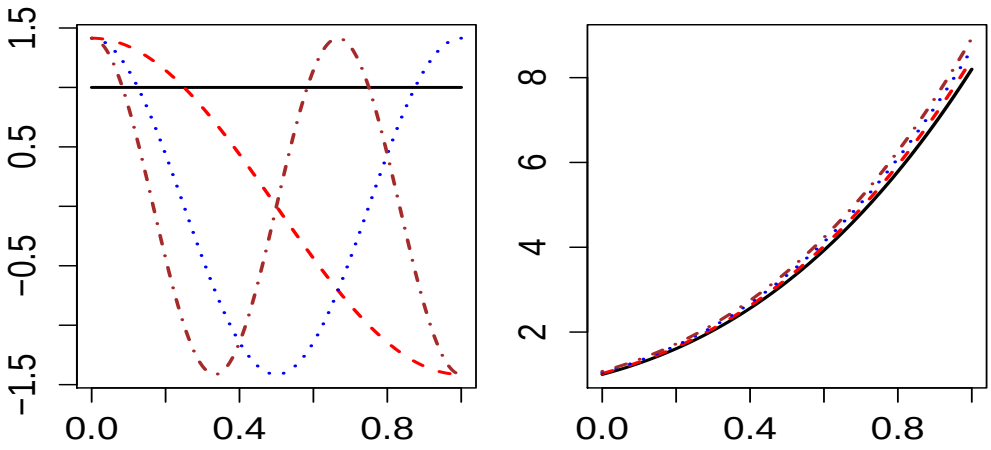

Figure 2.3: Visualization for Simulation II. Eigenfunctions $\{\phi_1(t), \phi_2(t), \phi_3(t), \phi_4(t)\}$ are on the left panel; Mean functions under the null and three alternative hypotheses are on the right panel.

2005) or (Peng and Paul, 2009). The original paper focuses on the one-sample mean function inference problem. And it is mentioned that the method can be generalized to two-sample mean function testing problems. However the asymptotic null distribution of the test statistic depends on the limiting eigenvalues of two complex sequences of matrices, and the implementation involves quite a few tuning parameters. In addition, as mentioned by the authors, their method requires a high accuracy level for the mean and covariance function estimates. We notice that the performance of the pLRT test is unstable, being sensitive to simulation settings. This shortcoming is illustrated by our simulation studies in Section 2.8.1.

The distribution test is applicable to sparse functional data as follows. Under the joint Gaussian assumption and the common covariance assumption, they claim that testing (2.29) is the same as testing

$$H_0 : \{\tilde{\omega}_{1k}\}_{k=1}^{\infty} \stackrel{D}{=} \{\tilde{\omega}_{2k}\}_{k=1}^{\infty} \quad VS \quad H_a : \{\tilde{\omega}_{1k}\}_{k=1}^{\infty} \stackrel{D}{\neq} \{\tilde{\omega}_{2k}\}_{k=1}^{\infty}, \tag{2.30}$$

where $\tilde{\omega}_{gk} = E[\omega_{gk}|\mathbf{Y}_{gi}]$. The specific formula of $\tilde{\omega}_{gk}$ has been studied in (Yao et al., 2005; Peng and Paul, 2009). So for sparse functional data, $\tilde{\omega}_{gk}$ instead of ω_{gk} should be used to perform the AD test with multiple testing corrections. We note that the K-L expansion used in the distribution test is based on the pooled mean μ_{pool} and pooled covariance $\tilde{G}(s,t)$.

Table 2.2: Results for simulation study II (p unknown). We presented results $N = 100$, $\sigma = 0$ and $n_1 = n_2 = 100$ or 200. Fraction of variance explained methods ('FVE80' and 'FVE90') are utilized to choose p for the projection-based test. We compare four different methods: the pointwise t-test ('pointwise'), the L^2-norm-based test, the Globalized F-test, the Projection-based test. The type I error ($\eta = 0$) and powers ($\eta = 0.4, 0.8, 1.2$) are calculated based on 1000 repetitions.

sample size	**Test**	$\boldsymbol{\eta = 0}$	$\boldsymbol{\eta = 0.13}$	$\boldsymbol{\eta = 0.26}$	$\boldsymbol{\eta = 0.4}$
$(n_1, n_2) = (100, 100)$	Pointwise	*0.399*	0.644	0.964	1.000
	L^2-norm-based	0.066	0.157	0.569	**1.000**
	Globalized	0.069	0.154	0.537	**1.000**
	Projection-based (FVE80)	0.059	**0.235**	**0.748**	**1.000**
	Projection-based (FVE90)	0.059	**0.235**	**0.748**	**1.000**
$(n_1, n_2) = (200, 200)$	Pointwise	*0.308*	0.807	0.993	1.000
	L^2-norm-based	0.051	0.265	0.832	**1.000**
	Globalized	0.050	0.256	0.812	**1.000**
	Projection-based (FVE80)	0.051	**0.436**	**0.950**	**1.000**
	Projection-based (FVE90)	0.051	**0.436**	**0.950**	**1.000**

2.4 PROPOSED TEST PROCEDURE FOR SPARSE FUNCTIONAL DATA

Let $\mu_{pool}(t)$ denote the mean function of the mixture process of $X_1(t)$ and $X_2(t)$. We first achieve a centered model by deducting this overall mean function from both sides of (2.1). Let $Y^c_{gij} = Y_{gij} - \mu_{pool}(T_{gij})$ and $X^c_{gi}(t) = X_{gi}(t) - \mu_{pool}(t)$, then model (2.1) becomes

$$Y^c_{gij} = X^c_{gi}(T_{gij}) + \epsilon_{gij} \tag{2.31}$$

where $X^c_{gi}(t) \sim SP(\mu_g(t) - \mu_{pool}(t), G(s,t))$.

Let $\{\phi_k(t), k = 1, 2, ..., \infty\}$ be orthonormal eigenfunctions of $G(s,t)$, corresponding to the non-increasing eigenvalue sequence $\{\lambda_k, k = 1, 2, ..., \infty\}$. We define the projection score of

$\mu_g(t) - \mu_{pool}(t)$ onto the kth eigenfunction $\phi_k(t)$ as θ^c_{gk}, which is calculated by

$$\theta^c_{gk} = \langle \mu_g - \mu_{pool}, \phi_k \rangle, \quad g = 1, 2; k \geq 1. \tag{2.32}$$

Given the fact that $\mu_g(t) - \mu_{pool}(t) = \sum_{k=1}^{\infty} \theta^c_{gk}\phi_k(t)$, it can be seen that testing (2.2) is equivalent to testing

$$H_0 : \{\theta^c_{1k}\}_{k=1}^{\infty} = \{\theta^c_{2k}\}_{k=1}^{\infty} \quad VS \quad H_a : \{\theta^c_{1k}\}_{k=1}^{\infty} \neq \{\theta^c_{2k}\}_{k=1}^{\infty}. \tag{2.33}$$

Observe that

$$X^c_{gi}(t) = \sum_{k=1}^{\infty} \langle X^c_{gi}, \phi_k \rangle \phi_k(t) \stackrel{\text{def}}{=} \sum_{k=1}^{\infty} r^c_{gik}\phi_k(t), \tag{2.34}$$

and the fact that testing (2.33) is equivalent to testing

$$H_0 : \{E[r^c_{1ik}]\}_{k=1}^{\infty} = \{E[r^c_{2ik}]\}_{k=1}^{\infty} \quad VS \quad H_a : \{E[r^c_{1ik}]\}_{k=1}^{\infty} \neq \{E[r^c_{2ik}]\}_{k=1}^{\infty}, \tag{2.35}$$

in dense data setting, projection-based tests truncate at the first p dimensions and construct test statistics based on r^c_{gik}, $g = 1, 2, i = 1, .., n_g, k = 1, .., p$.

For extremely dense functional data, one can first recover the underlying true curve and estimate the scores using $\int_{t\in\mathcal{T}} X^c_{gi}(t)\hat{\phi}_k(t)\, dt$. However, in sparse functional data settings, this is impossible.

We propose to consider the best linear predictor, $E[r^c_{gik}|\boldsymbol{Y}^c_{gi}]$, which is a rational choice because

$$E_t[E[r^c_{gik}|\boldsymbol{Y}^c_{gi}]] = E[r^c_{gik}]. \tag{2.36}$$

That is to say, by taking the randomness of design points under sparse data cases into account, the expectation of random quantity $E[r^c_{gik}|\boldsymbol{Y}^c_{gi}]$ also equals to θ^c_{gk} defined in (2.32). Similar ideas have been proposed in (Yao et al., 2005) for principal component analysis for sparse functional data. Let $\boldsymbol{Y}^c_{gi} = (Y^c_{gi1}, ..., Y^c_{giN_{gi}})^T$ be the vector contains observations of the corresponding subject. Denote $\boldsymbol{\mu}_{gi} = (\mu_g(T_{gi1}), ..., \mu_g(T_{giN_{gi}}))^T$ and $\boldsymbol{\mu}_{pool,gi} = (\mu_{pool}(T_{gi1}), ..., \mu_{pool}(T_{giN_{gi}}))^T$. Motivated from a special case where the projection scores r^c_{gik} and the random errors ϵ_{gij} are jointly Gaussian distributed, the best linear predictor $E[r^c_{gik}|\boldsymbol{Y}^c_{gi}]$ has the following explicit formula

$$E[r^c_{gik}|\boldsymbol{Y}^c_{gi}] = \theta^c_{gk} + \lambda_k \boldsymbol{\phi}^T_{gik}\boldsymbol{\Sigma}^{-1}_{\mathbf{Y}^c_{gi}}(\boldsymbol{Y}^c_{gi} - \boldsymbol{\mu}_{gi} + \boldsymbol{\mu}_{pool,gi}), \tag{2.37}$$

where $\mathbf{\Sigma}_{\mathbf{Y}^c_{gi}}$ is the covariance matrix of $\boldsymbol{Y}^c_{gi}$, and the (j, j') element of $\mathbf{\Sigma}_{\mathbf{Y}^c_{gi}}$ equals to $G(T_{gij}, T_{gij'})+\sigma^2 1(j = j')$.

Let's truncate at the first p directions and denote the projection score vector as $\boldsymbol{r}^c_{gi} = (r^c_{gi1}, ..., r^c_{gip})^T$. The best linear predictor of $\boldsymbol{r}^c_{gi}$ is

$$E[\boldsymbol{r}^c_{gi}|\boldsymbol{Y}^c_{gi}] = \boldsymbol{\theta}^c_g + diag(\boldsymbol{\lambda})\mathbf{\Phi}^T_{gi}\mathbf{\Sigma}^{-1}_{\mathbf{Y}^c_{gi}}(\boldsymbol{Y}^c_{gi} - \boldsymbol{\mu}_{gi} + \boldsymbol{\mu}_{pool,gi}), \tag{2.38}$$

where $\boldsymbol{\theta}^c_g = (\theta^c_{g1}, ..., \theta^c_{gp})^T$, $\mathbf{\Phi}_{gi} = (\boldsymbol{\phi}_{gi1}, ..., \boldsymbol{\phi}_{gip})$, with $\boldsymbol{\phi}_{gik} = (\phi_k(T_{gi1}), ..., \phi_k(T_{giN_{gi}}))^T$, $\boldsymbol{\lambda} = (\lambda_1, ..., \lambda_p)^T$.

Under the null hypothesis of the testing problem (2.2), we have $\mu_g(t) = \mu_{pool}(t)$ and $\boldsymbol{\theta}^c_g = \mathbf{0}$, so we propose to construct a test statistic based on the following quantity,

$$\tilde{\boldsymbol{r}}^c_{gi} = diag(\boldsymbol{\lambda})\mathbf{\Phi}^T_{gi}\mathbf{\Sigma}^{-1}_{\mathbf{Y}^c_{gi}}\boldsymbol{Y}^c_{gi} = diag(\boldsymbol{\lambda})\mathbf{\Phi}^T_{gi}\mathbf{\Sigma}^{-1}_{\mathbf{Y}^c_{gi}}(\boldsymbol{Y}_{gi} - \boldsymbol{\mu}_{pool,gi}), \tag{2.39}$$

which is referred as ***shrinkage score*** in later parts. The shrinkage score introduced in (2.39) is not the same as the original score $\boldsymbol{r}^c_{gi}$ in (2.34), but it gets closer to $\boldsymbol{r}^c_{gi}$ when the number of observations on each curve goes large and the measurement error gets small.

To construct a test statistic based on the shrinkage score vector $\tilde{\boldsymbol{r}}^c_{gi}$, let's first calculate its mean vector and covariance matrix through the following two steps. First, assuming the observing times $\boldsymbol{T}_{gi} = (T_{gi1}, ..., T_{giN_{gi}})^T$ are fixed, the conditional mean and variance of $\tilde{\boldsymbol{r}}^c_{gi}$ are

$$\begin{aligned} E[\tilde{\boldsymbol{r}}^c_{gi}|\boldsymbol{T}_{gi}] &= diag(\boldsymbol{\lambda})\mathbf{\Phi}^T_{gi}\mathbf{\Sigma}^{-1}_{\mathbf{Y}^c_{gi}}(\boldsymbol{\mu}_{gi} - \boldsymbol{\mu}_{pool,gi}) \\ Cov[\tilde{\boldsymbol{r}}^c_{gi}|\boldsymbol{T}_{gi}] &= diag(\boldsymbol{\lambda})\mathbf{\Phi}^T_{gi}\mathbf{\Sigma}^{-1}_{\mathbf{Y}^c_{gi}}\mathbf{\Phi}_{gi}diag(\boldsymbol{\lambda}) \end{aligned} \tag{2.40}$$

Next, by taking the randomness of $\boldsymbol{T}_{gi}$ into account, we have the mean and variance of $\tilde{\boldsymbol{r}}^c_{gi}$ are

$$\begin{aligned} E[\tilde{\boldsymbol{r}}^c_{gi}] &= E_t[diag(\boldsymbol{\lambda})\mathbf{\Phi}^T_{gi}\mathbf{\Sigma}^{-1}_{\mathbf{Y}^c_{gi}}(\boldsymbol{\mu}_{gi} - \boldsymbol{\mu}_{pool,gi})] \\ Cov[\tilde{\boldsymbol{r}}^c_{gi}] &= E_t[diag(\boldsymbol{\lambda})\mathbf{\Phi}^T_{gi}\mathbf{\Sigma}^{-1}_{\mathbf{Y}^c_{gi}}\mathbf{\Phi}_{gi}diag(\boldsymbol{\lambda})] \end{aligned} \tag{2.41}$$

All the quantities involved in estimating $\tilde{\boldsymbol{r}}_{gi}^c$, $E[\tilde{\boldsymbol{r}}_{gi}^c]$ and $Cov[\tilde{\boldsymbol{r}}_{gi}^c]$ can be obtained from data Y_{gij}, $g = 1, 2, i = 1, .., n_g, j = 1, .., N_{gi}$, with details included in the next section. Let's denote the estimated quantities as $\hat{\boldsymbol{\lambda}}$, $\hat{\phi}_k(t)$, $\hat{\mu}_{pool}(t)$, and $\hat{\sigma}$. We have the following empirical estimators

$$\begin{aligned}
\hat{\tilde{\boldsymbol{r}}}_{gi}^c &= diag(\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\Phi}}_{gi}^T[\hat{\boldsymbol{\Phi}}_{gi}diag(\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\Phi}}_{gi}^T + \hat{\sigma}^2\mathbf{I}]^{-1}(\mathbf{Y}_{gi} - \hat{\boldsymbol{\mu}}_{pool;gi}) \\
\hat{Cov}[\tilde{\boldsymbol{r}}_{gi}^c] &= \frac{1}{n_1+n_2}\sum_{g=1}^{2}\sum_{i=1}^{n_g} diag(\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\Phi}}_{gi}^T[\hat{\boldsymbol{\Phi}}_{gi}diag(\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\Phi}}_{gi}^T + \hat{\sigma}^2\mathbf{I}]^{-1}\hat{\boldsymbol{\Phi}}_{gi}diag(\hat{\boldsymbol{\lambda}}) \\
&\stackrel{\text{def}}{=} \hat{\boldsymbol{V}}
\end{aligned} \tag{2.42}$$

Given all these arguments, we propose to use the following test statistic,

$$T_{p,N} = [\bar{\hat{\tilde{\boldsymbol{r}}}}_{1\cdot}^c - \bar{\hat{\tilde{\boldsymbol{r}}}}_{2\cdot}^c]^T[(\frac{1}{n_1} + \frac{1}{n_2})\hat{\boldsymbol{V}}]^{-1}[\bar{\hat{\tilde{\boldsymbol{r}}}}_{1\cdot}^c - \bar{\hat{\tilde{\boldsymbol{r}}}}_{2\cdot}^c] \tag{2.43}$$

where $\bar{\hat{\tilde{\boldsymbol{r}}}}_{g\cdot}^c = \sum_{i=1}^{n_g}\hat{\tilde{\boldsymbol{r}}}_{gi}^c/n_g$.

Even though the proposed test statistic $T_{p,N}$ is motivated from joint Gaussian situations, we later proved that it has an asymptotic $\chi^2(p)$ null distribution regardless of whether the joint Gaussian assumption is true or not. See Theorem 1 for more details. The proof of this theoretical result is rather challenging. The shrinkage score vector $\tilde{\boldsymbol{r}}_{gi}^c$ is not observable, needing to be estimated from data. This additional estimation error is required to be controlled. The projection-based method for dense functional data also involves controlling errors in the estimated scores. What makes our argument more complicated than dense cases is that we need to consider each estimation error conditional on the individual design points and ensure that the overall error is uniformly controlled over random designs.

2.5 ESTIMATION PROCEDURES

In this section, we discuss the estimation procedure for all elements in $T_{p,N}$. We utilize the restricted Maximum Likelihood Estimate (rMLE) in (Peng and Paul, 2009). To implement the rMLE procedure, we modified their publicly available R package called 'fpca'. Note that Gaussian distributions are assumed in the following description. However, as stated in their paper, the Gaussian processes assumption is only a working condition and their asymptotic results still hold under

some relaxed conditions. Our simulation studies in Section 2.8.1 also show that rMLE has reasonable performance under non-Gaussian cases. Our test procedure itself does not require Gaussian distributions. One can always use other estimating methods, such as the local linear smoothing in (Yao et al., 2005; Li and Hsing, 2010; Zhang and Wang, 2016) and the EM algorithm in (James et al., 2000), as long as they produce consistent estimates.

Estimations of the mean function $\mu_g(t)$ and the pooled mean function $\mu_{pool}(t)$ are performed by a local linear smoothing technique. This method has been used in various studies, including (Fan and Gijbels, 1996; Yao et al., 2005; Peng and Paul, 2009). To be more specific, we define the local linear smoother of the pooled mean function $\mu_{pool}(t)$ by minimizing

$$\sum_{g=1}^{2}\sum_{i=1}^{n_g}\sum_{j=1}^{N_{gi}} K(\frac{T_{gij}-t}{h_{\mu_{pool}}})[Y_{gij}-\beta_0-\beta_1(t-T_{gij})] \tag{2.44}$$

with respect to β_0 and β_1, where $K(\cdot)$ is a smoothing kernel and $h_{\mu_{pool}}$ is the bandwidth. Then $\hat{\mu}_{pool}(t) = \hat{\beta}_0(t)$. Similarly, for each g, we define the local linear smoother of the mean function $\mu_g(t)$ by minimizing

$$\sum_{i=1}^{n_g}\sum_{j=1}^{N_{gi}} K(\frac{T_{gij}-t}{h_{\mu_g}})[Y_{gij}-\beta^*_{g0}-\beta^*_{g1}(t-T_{gij})] \tag{2.45}$$

with respect to β^*_{g0} and β^*_{g1}. Then $\hat{\mu}_g(t) = \hat{\beta}^*_{g0}(t)$.

As for the eigenfunctions $\phi_k(t)$, the eigenvalues λ_k of the common covariance function $G(s,t)$, and the measurement errors σ, estimates are achieved by modifying the restricted maximum likelihood method in (Peng and Paul, 2009), such that two independent samples instead of one can be dealt with. This method is based on a reduced rank model, i.e., it is assumed that the eigenvalues of $G(s,t)$ decay to zero efficiently fast such that the difference between $G(s,t) = \sum_{k=1}^{\infty}\lambda_k\phi_k(t)\phi_k(s)$ and $\sum_{k=1}^{p}\lambda_k\phi_k(t)\phi_k(s)$ are very small. Further more, under some weak smoothness conditions on $X_g(\cdot)$, the first p eigenfunctions $\{\phi_1(t),...,\phi_p(t)\}$ can be modeled as,

$$\phi_k(t) = \sum_{l=1}^{M} d_{lk}\psi_l(t), k = 1,,,.,p, \tag{2.46}$$

where functions $\{\psi_1(t),...,\psi_M(t)\} \in L^2(\mathcal{T})$ are known. Let's define $\mathbf{D} = (d_{lk})_{l=1,..,M;k=1,..,p}$. Based on the orthonormality of eigenfunctions, we have $\mathbf{D}^T\mathbf{D} = \mathbf{I}_p$, if $\{\psi_1(t),...,\psi_M(t)\}$ are also

orthonormalized. Under the Gaussian process and normal error assumption, conditional on time points, the negative log-likelihood of the data is given by

$$-\log L(\mathbf{D}, diag(\boldsymbol{\lambda}), \sigma^2) \propto \frac{1}{2}\sum_{g=1}^{2}\sum_{i=1}^{n_g} Tr[(\sigma^2\mathbf{I}_{N_{gi}} + \boldsymbol{\Psi}_{gi}^T\mathbf{D}diag(\boldsymbol{\lambda})\mathbf{D}^T\boldsymbol{\Psi}_{gi})(\mathbf{Y}_{gi} - \hat{\boldsymbol{\mu}}_{gi})(\mathbf{Y}_{gi} - \hat{\boldsymbol{\mu}}_{gi})^T] + \frac{1}{2}\log|\sigma^2\mathbf{I}_{N_{gi}} + \boldsymbol{\Psi}_{gi}^T\mathbf{D}diag(\boldsymbol{\lambda})\mathbf{D}^T\boldsymbol{\Psi}_{gi}| \tag{2.47}$$

where $\boldsymbol{\Psi}_{gi} = (\boldsymbol{\psi}_{gi1}, ..., \boldsymbol{\psi}_{giM})$, with $\boldsymbol{\psi}_{gil} = (\psi_l(T_{gi1}), ..., \psi_l(T_{giN_{gi}}))^T$ for $l = 1, .., M$. And $\hat{\mu}_g(t)$ is estimated above in (2.45) and $\hat{\boldsymbol{\mu}}_{gi} = (\hat{\mu}_g(T_{gi1}), ..., \hat{\mu}_g(T_{giN_{gi}}))^T$. A Newton-Raphson algorithm is utilized to achieve $\hat{\mathbf{D}}$, $\hat{\boldsymbol{\lambda}}$ and $\hat{\sigma}^2$ such that they minimize the negative log-likelihood, subject to the constraint that $\mathbf{D}^T\mathbf{D} = \mathbf{I}_p$. Then we can estimate the corresponding eigenfunctions by $\hat{\phi}_k(t) = \sum_{l=1}^{M}\hat{d}_{lk}\psi_l(t)$.

Combining all these estimates, we achieve estimation of the covariance $Cov[\tilde{\boldsymbol{r}}_{gi}^c]$. We raise the following estimator

$$\hat{\boldsymbol{V}} = \frac{1}{n_1 + n_2}\sum_{g=1}^{2}\sum_{i=1}^{n_g} diag(\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\Phi}}_{gi}^T[\hat{\boldsymbol{\Phi}}_{gi}diag(\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\Phi}}_{gi}^T + \hat{\sigma}^2\mathbf{I}]^{-1}\hat{\boldsymbol{\Phi}}_{gi}diag(\hat{\boldsymbol{\lambda}}), \tag{2.48}$$

and the empirical estimate for $\tilde{\boldsymbol{r}}_{gi}^c$ is $\hat{\tilde{\boldsymbol{r}}}_{gi}^c = (\hat{\tilde{r}}_{gi1}^c, ..., \hat{\tilde{r}}_{gip}^c)^T$, where $\hat{\tilde{r}}_{gik}^c = \hat{\lambda}_k\hat{\boldsymbol{\phi}}_{gik}^T\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{gi}}^{-1}(\mathbf{Y}_{gi} - \hat{\boldsymbol{\mu}}_{pool;gi})$. Then our test statistic $T_{p,N}$ is

$$T_{p,N} = [\bar{\hat{\tilde{\boldsymbol{r}}}}_{1\cdot}^c - \bar{\hat{\tilde{\boldsymbol{r}}}}_{2\cdot}^c]^T[(\frac{1}{n_1} + \frac{1}{n_2})\hat{\boldsymbol{V}}]^{-1}[\bar{\hat{\tilde{\boldsymbol{r}}}}_{1\cdot}^c - \bar{\hat{\tilde{\boldsymbol{r}}}}_{2\cdot}^c] \tag{2.49}$$

2.6 EXTENSION TO COMMON PRINCIPAL COMPONENT CASES

Even though our proposed test statistic $T_{p,N}$ is created based on the homogeneous covariance function assumption, we notice that it only requires that the eigenfunctions of the covariance functions are the same. To be more specific, the gth covariance function can be $G_g(s,t) = \sum_{k=1}^{\infty} \lambda_{gk}\phi_k(s)\phi_k(t)$, with $\phi_k(t)$ being the same across different groups. This is called the common principal component structure (CPC) (Flury, 1984; Benko et al., 2009; Boente et al., 2010; Chen and Müller, 2013). Under CPC, the test statistic has a formula that is similar to $T_{p,N}$, but $\hat{\boldsymbol{\lambda}}$ should be replaced with $\hat{\boldsymbol{\lambda}}_g$. The estimation procedure also needs to be modified. $\hat{\phi}_k(t)$ is achieved by applying the restricted maximum likelihood estimate to the pooled data. $\hat{G}_g(s,t)$ is calculated by using data from group g alone. Then the group-dependent eigenvalues $\hat{\lambda}_{gk}$ is obtained by the following formula,

$$\hat{\lambda}_{gk} =< \hat{\phi}_k, \hat{\phi}'_k >, \text{ with } \hat{\phi}'_k(t) = \int_{\mathcal{T}} \hat{G}_g(s,t)\hat{\phi}_k(s)dt. \tag{2.50}$$

2.7 ASYMPTOTIC RESULTS

In this section, we develop the asymptotic theory of $T_{p,N}$ under both the H_0 and the H_a in test (2.2). To achieve these asymptotic arguments, we need the following consistency results

$$\begin{aligned}
&\left\|\hat{\mu}_{pool} - \mu_{pool}\right\|_F = o_p(1)\\
&\left\|diag(\hat{\boldsymbol{\lambda}}) - diag(\boldsymbol{\lambda})\right\|_2 = o_p(1)\\
&\left\|\hat{\phi}_k - \phi_k\right\|_F = o_p(1)\\
&\left\|\hat{\sigma}^2 - \sigma^2\right\|_2 = o_p(1)
\end{aligned} \tag{2.51}$$

where $\|f\|_F$ is defined as $\{\int_{t\in\mathcal{T}} f(t)^2 dt\}^{1/2}$.

We estimate $\hat{\mu}_{pool}$ using local linear smoothing techniques and the consistency result holds under some regular and mild conditions, which are discussed in (Yao et al., 2005; Li and Hsing, 2010; Zhang and Wang, 2016). The other quantities are estimated using rMLE, and they are consistent when the following assumptions hold (Paul and Peng, 2009).

1. $X_1(t)$ and $X_2(t)$ are two independent Gaussian processes.
2. The p largest eigenvalues of $G(s,t)$ satisfy. (i) There exists a constant $a_1 < \infty$, such that $a_1 \geq \lambda_1 > \lambda_2 > ... > \lambda_p > \lambda_{p+1}$; (ii) There exists a constant $a_2 < \infty$, such that $\max_{1\leq k\leq p}(\lambda_k - \lambda_{k+1})^{-1} \leq a_2$.
3. The common eigenfunctions $\{\phi_k\}_{k=1}^p$ are four times continuously differentiable and satisfy for some $0 < A_0 < \infty$
$$\max_{1\leq k\leq p} \sup_{t\in\mathcal{T}} |\phi_k^{(4)}(t)| \leq A_0 \tag{2.52}$$
4. For each g, i, $\{T_{gij}, j = 1, ..., N_{gi}\}$ are i.i.d samples from a distribution g within the time domain $\mathcal{T}$, where g is a bounded function and it satisfies $c_{g,0} \leq g(x) \leq c_{g,1}$ for all $t \in \mathcal{T}$, where $0 < c_{g,0} \leq c_{g,1} < \infty$.
5. The number of measurements N_{gi} satisfies $\underline{N} \leq N_{gi} \leq \bar{N}$ with $\underline{N} \geq 4$ and $\bar{N} < \infty$.
6. The following two assumptions are correct: $M^{-1}(n/\log n)^{1/9} = O(1)$, and $M = o(\sqrt{n/\log n})$, where $n = n_1 + n_2$ and M as defined in Section 2.5 is the number of orthonormalized cubic B-spline basis used to represent the eigenfunctions $\phi_k(t)$.

The Gaussian process requirement is merely a working assumption for their rMLE method, which can be replaced by a weaker condition on the tail distributions. More details can be found in (Paul and Peng, 2009). Note that the derivation of the asymptotic distribution of the proposed test statistic $T_{p,N}$ does not need Gaussian distribution. We can always use other estimation methods such as the local linear smoothing, and then we need other conditions to make sure that the estimates are consistent (Yao et al., 2005; Li and Hsing, 2010). We focus on the sparse scenario that the number of observations on each curve is bounded; while the proposed method also works when $\bar{N}$ grows with the sample size n. If one can achieve rate α_n consistency for all the estimated quantities in (2.51) and have $\bar{N}^2\alpha_n^2 = o(1)$ satisfied, then the current proofs for Theorem 1 and 2 can go through. Given that we use rMLE to estimate λ, ϕ and σ and use local linear smoothing for μ_{pool}, we can at least allow $\bar{N} = O(n^{1/5})$ with $M \asymp (n\bar{N}^2/\log n)^{1/9}$ and an appropriately chosen bandwidth in local linear smoothing.

Theorem 1. *Under H_0 and regularization assumptions 1-6, assuming $\lim_{n_1.n_2\to\infty} \frac{n_1}{n} = w$, with $w \in (0,1)$ and $n = n_1 + n_2$, if $E_t[\mathbf{\Phi}_{gik}^T \mathbf{\Sigma}_{Y_{gi}}^{-1} \mathbf{\Phi}_{gik}] < \infty$ is satisfied. For any fixed p, we have*
$$T_{p,N} \xrightarrow{D} \chi^2(p) \tag{2.53}$$

Given Theorem 1, we propose the following test procedure. For a specific significance level α, reject H_0 in (2.2), if $T_{p,N} > \chi^2(1-\alpha; p)$, the upper α quantile of $\chi^2(p)$; otherwise, we will not have enough evidence to reject H_0. Proof of Theorem 1 is postponed to Appendix. The theorem guarantees the performance when the total sample size goes to infinity, finite sample performance is justified in the Section 4 by three simulation studies and real data analyses.

In the following, Theorem 2 provides theoretical justifications for the power of our test procedure under certain type of alternatives.

Theorem 2. *Under regularization assumptions 1-6, assuming* $\lim_{n_1.n_2\to\infty} \frac{n_1}{n} = w$, *with* $w \in (0,1)$ *and* $n = n_1 + n_2$. *Suppose that we have* $E_t(diag(\boldsymbol{\lambda})\boldsymbol{\Phi}_{gi}^T \boldsymbol{\Sigma}_{Y_{gi}}^{-1}(\boldsymbol{\mu}_{gi} - \boldsymbol{\mu}_{pool;gi})) \neq \mathbf{0}$, *then*

$$T_{p,N} \xrightarrow{p} \infty \tag{2.54}$$

Under random design t, $E_t(diag(\boldsymbol{\lambda})\boldsymbol{\Phi}_{gi}^T \boldsymbol{\Sigma}_{Y_{gi}}^{-1}(\boldsymbol{\mu}_{gi} - \boldsymbol{\mu}_{pool;gi}))$ is basically some type of projection of $\mu_g(t) - \mu_{pool}(t)$. When $\mu_g(t) - \mu_{pool}(t) \neq 0$, it is unlikely that this projection scores equal to 0 on all the first p directions. Theorem 2 ensures that, under any alternative H_a when the difference between $\mu_g(t)$ and $\mu_{pool}(t)$ is captured by some of the first p directions of such projection, the power of our test procedure goes to 1. Proof of Theorem 2 is also postponed to Appendix.

2.8 NUMERICAL EXPERIMENTS

2.8.1 Simulation studies

In this subsection, through three simulations we evaluate the performances of the proposed test $T_{p,N}$ ('Shrink'). First, we investigate the performance of the proposed χ^2 test in comparison with three other methods. Second, we examine the performance of our proposed χ^2 test under a non-Gaussian circumstance. Third, we investigate different methods of choosing number of p.

Simulation I

The data is generated based on model (2.1), and the Karhunen-Loève expansion $X_{gi}(T_{gij}) = \mu_g(T_{gij}) + \sum_{k=1}^{\infty} \xi_{gik}\phi_k(T_{gij})$, with $\xi_{gik} \sim N(0, \lambda_k)$. The number of observations are the same

across different curves, i.e., $N_{gi} = N$ for $g = 1, 2; i = 1, ..., n_g$. N takes two values, 4 and 8. The observing time points, $\{T_{gi1}, ..., T_{giN}\}$, are uniformly distributed within the interval [0,1]. We assume that the eigenvalues are $\lambda_k = 2(k+1)^{-2}$ for $k = 1, ..., 4$, and $\lambda_k = 0$ for $k > 4$. The eigenfunctions are $\phi_1(t) = \sqrt{2}\cos(\pi t)$, $\phi_2(t) = \sqrt{2}\sin(\pi t)$, $\phi_3(t) = \sqrt{2}\cos(2\pi t)$, $\phi_4(t) = \sqrt{2}\sin(2\pi t)$. The mean functions are $\mu_1 = \sum_{k=1}^{4} a_{1k}\phi_k(t)$, where $a_{1k} = 3\bar{a}_{1k}/\|\bar{a}_1\|_2$, with $\bar{a}_1 = (\bar{a}_{11}, ..., \bar{a}_{14})^T$ and $\bar{a}_{1k} = (2)^{5-k}(5-k)^6$. And $\mu_2 = \sum_{k=1}^{4} a_{2k}\phi_k(t)$, where $a_{2k} = 3\bar{a}_{2k}/\|\bar{a}_2\|_2$, with $\bar{a}_2 = (\bar{a}_{21}, ..., \bar{a}_{24})^T$ and $\bar{a}_{2k} = (2-\varsigma)^{5-k}(5-k)^6$. Note that $\varsigma = 0, 0.8, 1.2, 1.6$ correspond to the null hypothesis and three alternative hypotheses for (2.2). The standard deviation of the measurement errors takes three values, $\sigma = 0$, $\sigma = 0.6$, $\sigma = 1.2$. The sample sizes are $n_1 = n_2 = 100$, $n_1 = n_2 = 300$, and $n_1 = n_2 = 600$. Figure 2.4 and Figure 2.5 visualize 9 randomly selected subjects for $(N = 4, \sigma = 0.6)$ and $(N = 4, \sigma = 1.2)$.

In this simulation study, we compare performances of the proposed asymptotic χ^2 test $T_{p,N}$ ('Shrink'), the distribution test in (Pomann et al., 2016) ('Distribution'), the pLRT test in the (Staicu et al., 2014) ('pLRT-linear' and 'pLRT-cubic'), and a naive adaption from the dense functional data projection-based test in (Horváth and Rice, 2015; Lillo et al., 2015) ('Naive dense'). The distribution test in (Pomann et al., 2016) aims to test whether $X_1(t)$ and $X_2(t)$ have the same distribution. They projected each individual curve to the eigenspace spanned by the covariance of the mixture process of $X_1(t)$ and $X_2(t)$. A similar shinkage score was used for sparse functional data. Then test the equality of distributions of the first p projection scores through the Anderson-Darling test with Bonferroni corrections. It is a nonparametric test which does not require the derivation of the score distributions. Using Gaussian assumption with equal covariance settings, this test can be used to test whether $X_1(t)$ and $X_2(t)$ have the same mean function. For the pLRT test, the original paper focuses on the one-sample mean function inference problem. And it is mentioned that the method can be generalized to two-sample mean function testing problems. We were able to modify their package to perform the two-sample testing problem. However in our simulation studies, the pLRT test yield inflated type I errors on the average of 0.4. The original pLRT paper includes some results for sparse data with $N = 10$, and the code involves quite a few tuning parameters. We were not sure if the inflated type I errors were due to the rather sparse settings with $N = 4$ and $N = 8$ or due to parameter tunings that have not reach the optimal. For the naive adaption from the dense functional data, we first use the B-spline interpolation to recover the

underlying curve for each individual subject. Then apply the dense data eigen-projection test to the recovered data. Though we do not compare with other dense functional tests using the recovered data, we expect them to work similarly as the one considered here, since the pre-smoothing step is problematic for sparse functional data.

All the tests except for the pLRT test need to choose p, we use the true $p = 4$ for this simulation. All the simulation results are summarized in Table 2.3-Table 2.5. The distribution test performs reasonably good, with incidentally inflated type I errors. Our proposed χ^2 test performs the best under almost all circumstances. As expected, the naive adaptation from the dense data test does not work, since the null distribution of the test statistic is incorrect under sparse designs. The pLRT test cannot control type I error at pre-specified level.

Simulation II

The second part focus on exploring the performance of the proposed test under non-Gaussian circumstances. The number of observation on each curve is $N = 4$. Most of the other simulation settings are the same as Simulation I, except that ξ_{gik}, $g = 1, 2; i = 1, .., n_g; k = 1, ..., p$ are now generated from a mixture of two normal distributions, i.e., they are distributed as $N(\sqrt{\lambda_k/2}, \lambda_k/2)$ with probability 1/2 and $N(-\sqrt{\lambda_k/2}, \lambda_k/2)$ with probability 1/2. In this way, we get samples from mixture Gaussian processes instead of Gaussian processes. According to the results summarized in Table 2.6, we can see that the proposed test still produces valid results for this non-Gaussian situation. It can control the type I error at the predetermined significance level and the powers are reasonable.

Simulation III

Now we adopt two more complex settings. Under the first setting, the number of observations $N_{gi} \sim U[2, ..., 6]$, and then conditional on the value of N_{gi}, $(T_{gi1}, ..., T_{giN_{gi}})$ are i.i.d uniformly distributed. Eigenfunctions are $\phi_{2k-1}(t) = \sqrt{2}\cos((2k-1)\pi t)$ for $k = 1, ..., 15$, $\phi_{2k}(t) = \sqrt{2}\sin(2k\pi t)$ for $k = 1, ..., 15$. As for the mean functions, we used $\mu_g(t) = \sum_{k=1}^{30} a_{gk}\phi_k(t)$, where $a_{gk} = 3\bar{a}_{gk}/\left\|\bar{a}_g\right\|_2$, with $\bar{a}_g = (\bar{a}_{g1}, ..., \bar{a}_{g30})^T$, $\bar{a}_{1k} = (2)^{31-k}(31-k)^6$ and $\bar{a}_{2k} = (2+\delta_1)^{31-k}(31-k)^6$ with $\delta_1 = 0, 1.2, 2.4, 3.6$, $\sigma = 1$. This setting represents situations where p is rather large. Under the second setting, the mean functions are $\mu_1(t) = 1.08 - 3.2t$ and $\mu_2(t) = \mu_1(t) + (1-t)\delta_2$, with $\delta_2 = 0, 0.3, 0.6, 0.9$ for the null and three alternative hypotheses. This setting represents situations where p is unknown. All the other settings are the same.

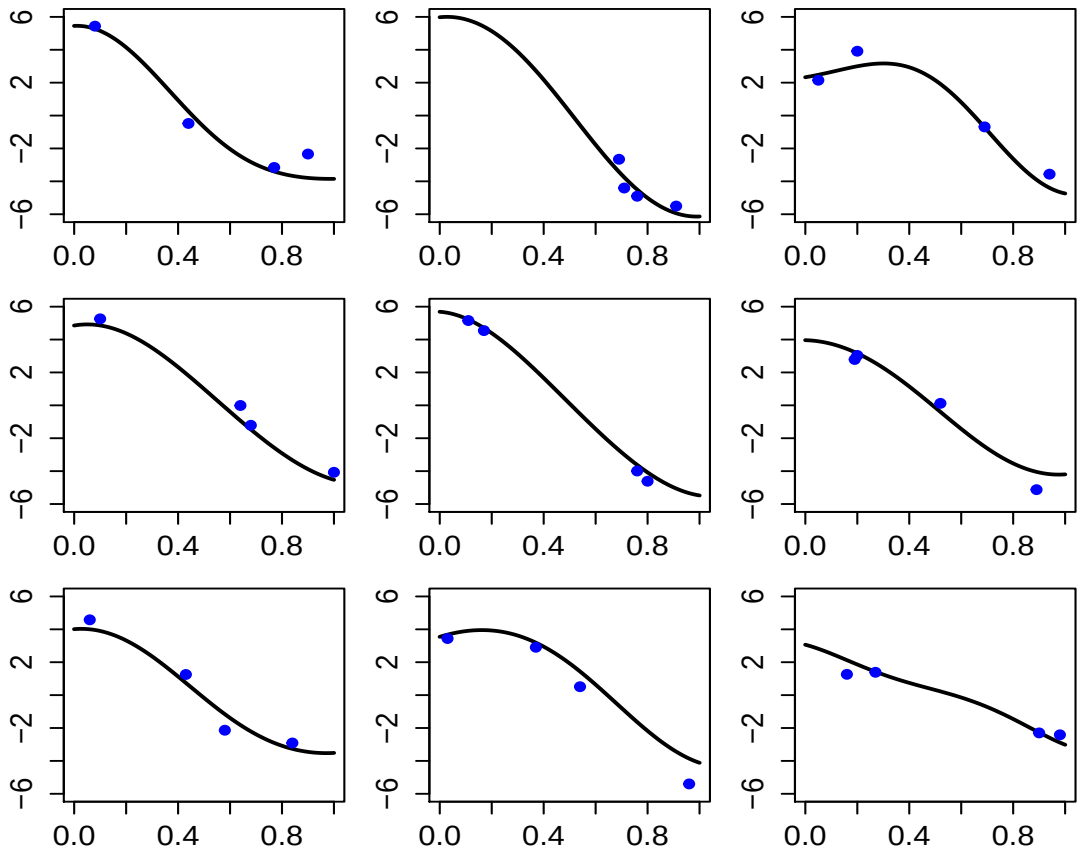

Figure 2.4: ($N = 4$, $\sigma = 0.6$) Simulated data for nine randomly chosen subjects in group 1. Note that in each plot, the black curve is the underlying curve for this subject, and the blue dots represent the observations contaminated with measurement errors.

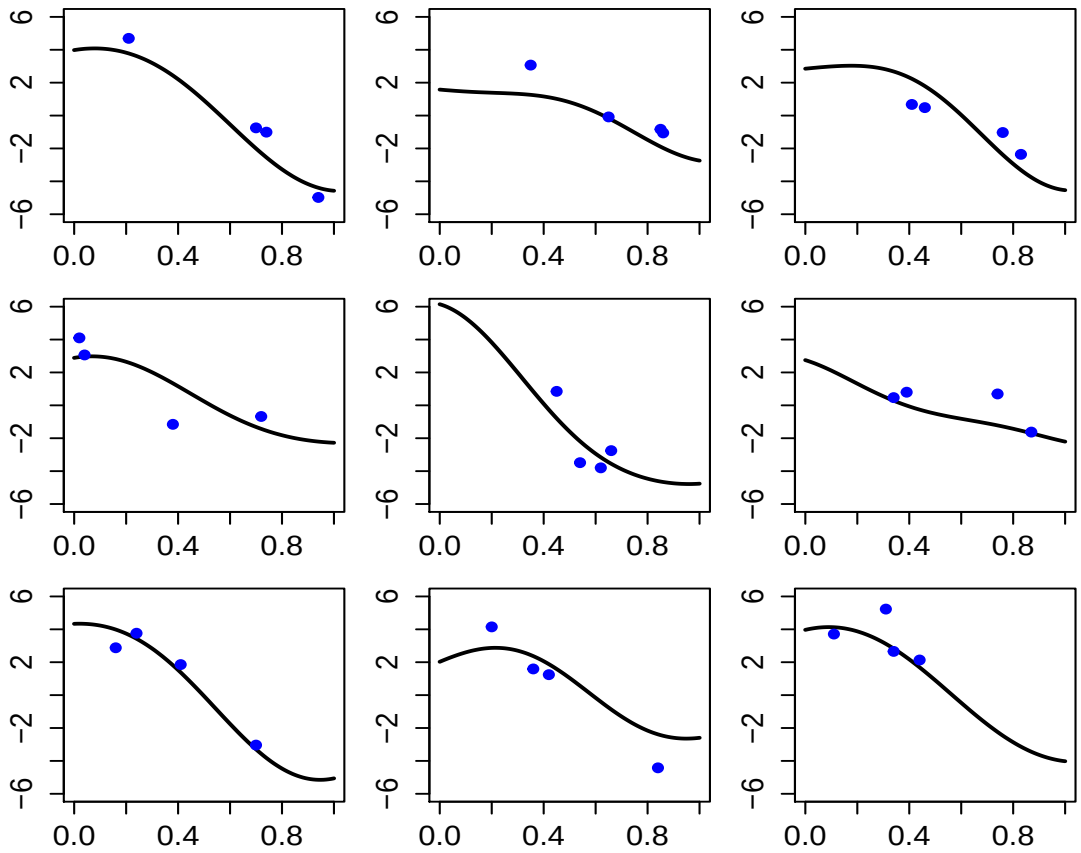

Figure 2.5: ($N = 4$, $\sigma = 1.2$) Simulated data for nine randomly chosen subjects in group 1. Note that in each plot, the black curve is the underlying curve for this subject, and the blue dots represent the observations contaminated with measurement errors.

We exploit two methods of choosing p: the leave one curve out cross-validation approach described in (Peng and Paul, 2009) ('CV'), and the fraction of variance explained method ('FVE'), with thresholds 80% ('FVE80') and 85% ('FVE85'). Combining the results in Table 2.7 and Table 2.8, we can see that performance of our proposed χ^2 test is not sensitive to the method of choosing p. Both 'CV' and 'FVE' have reasonable performances, while the cross-validation method tends to have better performances in terms of power.

2.8.2 Application to an AIDS clinical trial study

Now we apply the proposed χ^2 test to a CD4 count data from the AIDS Clinical Trials Group (ACTG) 193 A study, which aims at comparing the effectiveness of two different therapies, 600mg of zidovudine alternating monthly with 400mg didanosine (group A) and 600mg of zidovudine plus 400mg didanosine (group B), for advanced AIDS patients with CD4 counts less than or equal to 50 cells per cubic millimeter. Totally we have 655 advanced AIDS patients, with 325 subjects in group A and 330 in group B. These patients were followed for 40 weeks after they started to receive either of the two treatments mentioned above. The number of observations available for each subject ranges from 1 to 9, and the observing times are randomly distributed within 40 weeks.

We model the data from sparse functional data perspective. The 325 CD4 count trajectories in group A are assumed to be i.i.d random samples of an unknown stochastic process. The 330 CD4 count trajectories in group B are i.i.d samples of another unknown random process. Through inferring whether the two population mean functions are significant different within the 40-weeks period, we can understand whether the effectiveness of these two treatments are different over time. As suggested by previous literature (Fitzmaurice et al., 2012; Weinfurt et al., 2000), we transform the original CD4 count using function $log(x+1)$. The Spaghetti plots of transformed data of group A and B can be found in Figure 2.8. The first three estimated functional principal components for both groups respectively are shown on the right panel in Figure 2.9. As shown by Figure 2.9, the estimated eigenfunctions of the two groups, especially the first two components, are rather close to each other. The corresponding eigenvalues are $(24.868, 1.763, 0.564)^T$ for group A and $(38.759, 1.805, 0.404)^T$ for group B. We implement our proposed χ^2 test under both the homogeneous covariance (Section 2.4) and the common principal component structures (Section 2.6).

The p-values using the homogeneous covariance structure under $p = 2, 3, 4, 5$, are 0.0333, 0.0779, 0.0133 and 0.0270. Under the CPC structure mentioned in Section 2.6, the corresponding p-values are smaller, they are 1.111e-61, 7.727e-61, 4.608e-92, 1.472e-97. We can see that under all different values of p, the proposed test is able to detect the discrepancy between the effectiveness of group A and group B. Given the estimated mean functions provided on the left panel in Figure 2.9, it can seen that the therapy used by group B tends to be more effective on helping advanced AIDS patient to recover.

2.8.3 Application to an eBay online auction data set

EBay.com is one of the largest online auction markets. The most common auctions on eBay are single-item auctions, even though multiple-item auctions are also feasible. The problem that we consider belongs to the single-item auction category. For this kind of auctions, eBay adopts the second-price rule to decide the winner. To be more specific, within the pre-selected bidding period (3 days, 5 days, or 7 days), bidders can submit the maximum amounts that they are willing to pay (WTP). The first bidder to provide the second largest WTP within the bidding period is the winner. The WTP's are hidden from the public, instead, eBay's proxy bidding system automatically increases each bidder's bid by a minimum increment determined by the current bidding price and certain rules set by eBay. And it displays these prices lively on the item page. This real-time price trajectories come out of this system are often referred as live bid. The auction data in our problem are live bid data. More details about the mechanism of eBay online auctions are provided by their official website (Ebay.com, 1995). eBay auction data for all items are appropriately stored by eBay on (Ebay.com, 1995) and are completely accessible to all registered users for up to 90 days.

As mentioned in the previous paragraph, the auctions can last for 3 days, 5 days, or 7 days. We consider the problem that whether the price trends for a certain type of item are different for different bidding length choices. We use the live bid for Palm M515 Personal Digital Assistant as an example. The data set can be found at (Jank and Shmueli, 2010). The two bidding lengths that we compare are 3 days and 7 days. In the 3 days group, there are 90 auctions of Palm M515 that happened between March and May, 2003. In the 7 days group, there are 158 auctions of Palm M515 happened within the same period of time. The number of bids for each items are given in

Figure 2.10, which are not small. However, as shown in Figure 2.11, where 9 randomly selected auctions in the 7 days group are visualized, the bids are extremely irregular. There are big gaps among the data.

We use sparse functional data approaches to model this ebay online live bid data. We assume that the 90 auctions with duration 3 days are i.i.d random samples from an unknown population distribution, and the 158 auctions with duration 7 days are i.i.d random sample from another unknown stochastic process. (Peng and Müller, 2008; Liu and Müller, 2008; Wang et al., 2004; Jank et al., 2008) also analyze ebay online auction data from sparse functional perspectives. Following previous papers (Peng and Müller, 2008; Jank et al., 2008), we transform live bid data into log-scale. We also scale the bidding time variables of both the 3 days group and the 7 days group to $[0, 1]$. Now the problem of detecting differences in average price trends becomes testing whether the population mean curves are significantly different within some parts of $[0, 1]$.

As indicated by the right panel in Figure 2.12, the estimated first three eigenfunctions of the 7 days group (black) and the 3 days group (blue) are very close. The corresponding eigenvalues are $(0.344, 0.101, 0.020)^T$ for the 7 days group and $(0.416, 0.162, 0.011)^T$ for the 3 days group. It is reasonable to assume that they share the same covariance function. The proposed χ^2 test is implemented to this mean function testing problem. We use the cross-validation method to choose $p = 5$ and the p-value of our proposed test is less than 0.000001. This means that we have enough confidence to conclude that the mean price evolution curves are different when different bidding time periods are selected. The estimated mean functions by local linear smoothing are included on the left panel in Figure 2.12. From the graph, we can see that the price of the 7 days group (solid black) is larger than that of the 3 days group (dash blue) over the entire time domain.

Table 2.3: Results for simulation I. Results for the number of observations on each curve is 4 ($N = 4$) and 8 ($N = 8$), when the standard deviation of the random measurement error is 0 ($\sigma = 0$). Three different sample sizes $(n_1, n_2) = (100, 100)$,$(300, 300)$, and $(600, 600)$ are considered. We compare the performances of four different methods: the proposed χ^2 test $T_{p,N}$ ('shrink'), the distribution test in (Pomann et al., 2016)('distribution'), the naive adaption of dense eigen-projection based test in (Horváth and Rice, 2015; Lillo et al., 2015) ('Naive dense'), and the pLRT test in (Staicu et al., 2014)('pLRT-linear' and 'pLRT-cubic'). The type I error ($\varsigma = 0$) and powers ($\varsigma = 0.8, 1.2, 1.6$) are calculated based on 1000 repetitions.

$(\mathbf{N}, \boldsymbol{\sigma})$	**Sample size**	**Test**	$\boldsymbol{\varsigma = 0}$	$\boldsymbol{\varsigma = 0.8}$	$\boldsymbol{\varsigma = 1.2}$	$\boldsymbol{\varsigma = 1.6}$
$N = 4, \sigma = 0$	$(n_1, n_2) = (100, 100)$	Shrink	0.043	**0.435**	**0.990**	0.999
		Naive dense	0.048	0.053	0.048	0.046
		Distribution	*0.166*	0.491	0.985	**1.000**
		pLRT-linear	*0.671*	0.653	0.643	0.718
		pLRT-cubic	*0.302*	0.310	0.328	0.348
	$(n_1, n_2) = (300, 300)$	Shrink	0.044	**0.904**	**1.000**	**1.000**
		Naive dense	0.054	0.062	0.076	0.069
		pLRT-linear	*0.564*	0.512	0.592	0.679
		pLRT-cubic	*0.398*	0.377	0.403	0.537
	$(n_1, n_2) = (600, 600)$	Shrink	0.034	**0.999**	**1.000**	**0.998**
		Naive dense	0.136	0.152	0.145	0.137
		Distribution	*0.110*	0.999	1.000	0.999
		pLRT-linear	*0.389*	0.387	0.427	0.360
		pLRT-cubic	0.464	0.467	0.467	0.465
$N = 8, \sigma = 0$	$(n_1, n_2) = (100, 100)$	Shrink	0.027	**0.484**	**0.995**	**1.000**
		Naive dense	*0.617*	0.637	0.607	0.633
		Distribution	0.041	0.433	0.992	**1.000**
		pLRT-linear	*0.270*	0.276	0.266	0.312
		pLRT-cubic	*0.107*	0.105	0.111	0.139
	$(n_1, n_2) = (300, 300)$	Shrink	0.035	**0.962**	**1.000**	**1.000**
		Naive dense	*0.922*	0.921	0.927	0.895
		Distribution	0.060	0.938	**1.000**	**1.000**
		pLRT-linear	*0.046*	0.054	0.096	0.325
		pLRT-cubic	*0.165*	0.174	0.178	0.374
	$(n_1, n_2) = (600, 600)$	Shrink	0.036	**1.000**	**1.000**	**1.000**
		Naive dense	*0.937*	0.947	0.931	0.944
		Distribution	0.050	**1.000**	**1.000**	0.997
		pLRT-linear	*0.106*	0.110	0.198	0.110
		pLRT-cubic	*0.314*	0.331	0.397	0.306

Table 2.4: Results for simulation I. Results for the number of observations on each curve is 4 ($N = 4$) and 8 ($N = 8$), when the standard deviation of the random measurement error is 0.6 ($\sigma = 0.6$). Three different sample sizes $(n_1, n_2) = (100, 100)$,$(300, 300)$, and $(600, 600)$ are considered. We compare the performances of four different methods: the proposed χ^2 test $T_{p,N}$ ('shrink'), the distribution test in (Pomann et al., 2016)('distribution'), the naive adaption of dense eigen-projection based test in (Horváth and Rice, 2015; Lillo et al., 2015) ('Naive dense'), and the pLRT test in (Staicu et al., 2014)('pLRT-linear' and 'pLRT-cubic'). The type I error ($\varsigma = 0$) and powers ($\varsigma = 0.8, 1.2, 1.6$) are calculated based on 1000 repetitions.

$(\mathbf{N}, \boldsymbol{\sigma})$	**Sample size**	**Test**	$\varsigma = 0$	$\varsigma = 0.8$	$\varsigma = 1.2$	$\varsigma = 1.6$
$N = 4, \sigma = 0.6$	$(n_1, n_2) = (100, 100)$	Shrink	0.049	**0.274**	**0.927**	**1.000**
		Naive dense	0.032	0.037	0.036	0.308
		Distribution	0.065	0.268	0.891	**1.000**
		pLRT-linear	*0.399*	0.407	0.471	0.714
		pLRT-cubic	*0.356*	0.383	0.422	0.713
	$(n_1, n_2) = (300, 300)$	Shrink	0.040	**0.776**	**1.000**	**1.000**
		Naive dense	*0.091*	0.056	0.846	0.808
		Distribution	0.058	0.734	**1.000**	**1.000**
		pLRT-linear	*0.662*	0.648	0.513	0.841
		pLRT-cubic	*0.667*	0.718	0.721	0.904
	$(n_1, n_2) = (600, 600)$	Shrink	0.039	**0.973**	**1.000**	**1.000**
		Naive dense	*0.134*	0.128	0.132	0.965
		Distribution	0.058	0.960	**1.000**	**1.000**
		pLRT-linear	*0.637*	0.647	0.684	0.896
		pLRT-cubic	*0.830*	0.827	0.859	0.951
$N = 8, \sigma = 0.6$	$(n_1, n_2) = (100, 100)$	Shrink	0.041	**0.418**	**0.988**	**1.000**
		Naive dense	*0.349*	0.331	0.353	0.336
		Distribution	0.035	0.340	0.960	**1.000**
		pLRT-linear	*0.349*	0.383	0.465	0.714
		pLRT-cubic	*0.391*	0.388	0.465	0.714
	$(n_1, n_2) = (300, 300)$	Shrink	0.047	**0.881**	**1.000**	**1.000**
		Naive dense	*0.804*	0.813	0.824	0.827
		Distribution	0.060	0.849	**1.000**	**1.000**
		pLRT-linear	*0.386*	0.410	0.519	0.836
		pLRT-cubic	*0.653*	0.671	0.728	0.896
	$(n_1, n_2) = (600, 600)$	Shrink	0.049	**0.994**	**1.000**	**1.000**
		Naive dense	*0.958*	0.955	0.953	0.955
		Distribution	0.037	**0.994**	**1.000**	**1.000**
		pLRT-linear	*0.564*	0.537	0.638	0.907
		pLRT-cubic	*0.792*	0.790	0.816	0.959

Table 2.5: Results for simulation I. Results for the number of observation on each curve is 4 ($N = 4$) and 8 ($N = 8$), when the standard deviation of the random measurement error is 1.2 ($\sigma = 1.2$). Three different sample sizes $(n_1, n_2) = (100, 100)$,$(300, 300)$, and $(600, 600)$ are considered. We compare the performances of four different methods: the proposed χ^2 test $T_{p,N}$ ('shrink'), the distribution test in (Pomann et al., 2016)('distribution'), the naive adaption of dense eigen-projection based test in (Horváth and Rice, 2015; Lillo et al., 2015) ('Naive dense'), and the pLRT test in (Staicu et al., 2014)('pLRT-linear' and 'pLRT-cubic'). The type I error ($\varsigma = 0$) and powers ($\varsigma = 0.8, 1.2, 1.6$) are calculated based on 1000 repetitions.

$(\mathbf{N}, \boldsymbol{\sigma})$	**Sample size**	**Test**	$\varsigma = 0$	$\varsigma = 0.8$	$\varsigma = 1.2$	$\varsigma = 1.6$
$N = 4, \sigma = 1.2$	$(n_1, n_2) = (100, 100)$	Shrink	0.038	0.121	0.628	**1.000**
		Naive dense	0.033	0.025	0.032	0.022
		Distribution	0.049	0.162	0.639	**1.000**
		pLRT-linear	*0.372*	0.380	0.446	0.700
		pLRT-cubic	*0.445*	0.435	0.505	0.728
	$(n_1, n_2) = (300, 300)$	Shrink	0.031	0.455	0.991	**1.000**
		Naive dense	0.046	0.039	0.045	0.041
		Distribution	0.051	0.455	**0.994**	**1.000**
		pLRT-linear	*0.605*	0.634	0.685	0.883
		pLRT-cubic	*0.729*	0.771	0.787	0.921
	$(n_1, n_2) = (600, 600)$	Shrink	0.036	**0.784**	**1.000**	**1.000**
		Naive dense	*0.134*	0.123	0.128	0.132
		Distribution	0.059	0.782	**1.000**	**1.000**
		pLRT-linear	*0.578*	0.615	0.661	0.908
		pLRT-cubic	*0.804*	0.831	0.860	0.958
$N = 8, \sigma = 1.2$	$(n_1, n_2) = (100, 100)$	Shrink	0.045	**0.224**	**0.880**	**1.000**
		Naive dense	*0.143*	0.156	0.143	0.148
		Distribution	0.065	0.213	0.835	**1.000**
		pLRT-linear	*0.421*	0.428	0.539	0.777
		pLRT-cubic	*0.552*	0.537	0.616	0.815
	$(n_1, n_2) = (300, 300)$	Shrink	0.046	**0.658**	**1.000**	**1.000**
		Naive dense	*0.508*	0.502	0.484	0.512
		Distribution	0.042	0.622	**1.000**	**1.000**
		pLRT-linear	*0.494*	0.526	0.613	0.857
		pLRT-cubic	*0.763*	0.746	0.792	0.927
	$(n_1, n_2) = (600, 600)$	Shrink	0.047	**0.955**	**1.000**	**1.000**
		Naive dense	*0.830*	0.846	0.845	0.856
		Distribution	0.051	0.940	**1.000**	**1.000**
		pLRT-linear	*0.649*	0.607	0.702	0.649
		pLRT-cubic	*0.847*	0.808	0.877	0.847

Table 2.6: Results for simulation study II. Performances of our χ^2 test is examined under a non-Gaussian case. We presented results for the standard deviation of the random measurement error is 0 ($\sigma = 0$), 0.6 ($\sigma = 0.6$), and 1.2 ($\sigma = 1.2$), when the number of observation on each curve is 4 ($N = 4$). Three different sample sizes $(n_1, n_2) = (100, 100), (300, 300),$ or $(600, 600)$ are considered. The type I error ($\varsigma = 0$) and powers ($\varsigma = 0.8, 1.2, 1.6$) are calculated based on 1000 repetitions.

$(\mathbf{N}, \boldsymbol{\sigma})$	**sample size**	$\boldsymbol{\varsigma = 0}$	$\boldsymbol{\varsigma = 0.8}$	$\boldsymbol{\varsigma = 1.2}$	$\boldsymbol{\varsigma = 1.6}$
$N = 4, \sigma = 0$	$(n_1, n_2) = (100, 100)$	0.068	0.462	0.987	1.000
	$(n_1, n_2) = (300, 300)$	0.044	0.946	1.000	1.000
	$(n_1, n_2) = (600, 600)$	0.061	1.000	1.000	1.000
$N = 4, \sigma = 0.6$	$(n_1, n_2) = (100, 100)$	0.051	0.273	0.930	1.000
	$(n_1, n_2) = (300, 300)$	0.038	0.736	1.000	1.000
	$(n_1, n_2) = (600, 600)$	0.046	0.975	1.000	1.000
$N = 4, \sigma = 1.2$	$(n_1, n_2) = (100, 100)$	0.029	0.125	0.637	0.998
	$(n_1, n_2) = (300, 300)$	0.030	0.391	0.995	1.000
	$(n_1, n_2) = (600, 600)$	0.042	0.776	1.000	1.000

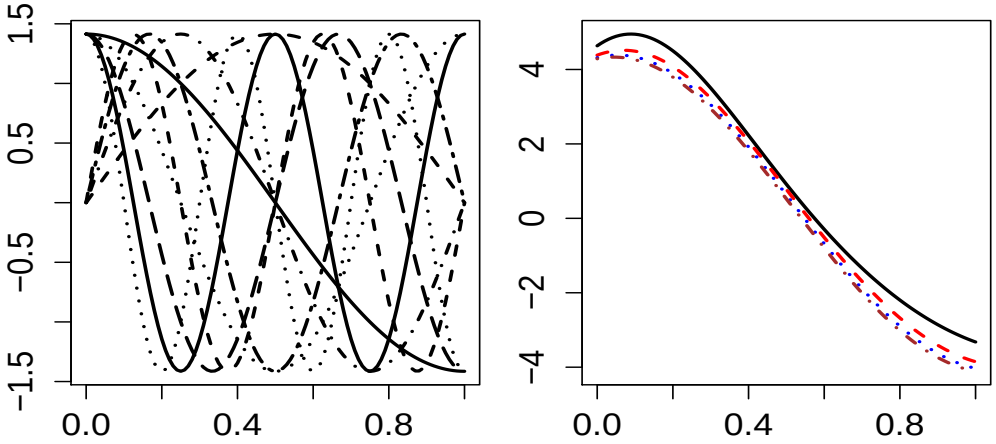

Figure 2.6: Visualizations of the first setting in Simulation III: first 10 eigenfunctions are on the left panel; mean functions are on the right panel.

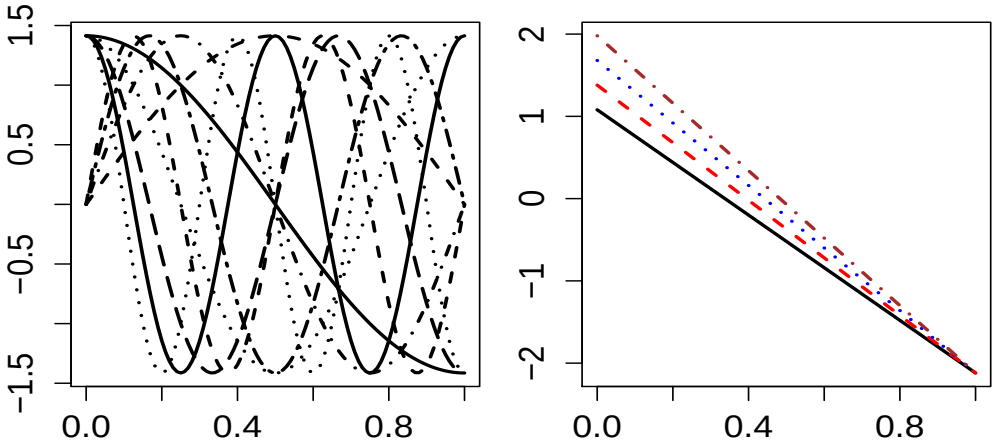

Figure 2.7: Visualizations of the second setting in Simulation III: first 10 eigenfunctions are on the left panel; mean functions are on the right panel.

Table 2.7: Results for simulation III. True value of p is 30, and the leave one curve out cross-validation method ("CV") (Peng and Paul, 2009) as well as the fraction of variance explained rule ("FVE80" and "FVE85") are implemented to choose an appropriate p. Performances of our shrink test are evaluated in terms of type I error ($\delta_1 = 0$) and powers ($\delta_1 = 1.2, 2.4, 3.6$) over 1000 repetitions.

sample size	**Testing**	δ_1**=0**	δ_1**=1.2**	δ_1**=2.4**	δ_1**=3.6**
$(n_1, n_2) = (100, 100)$	Shrink (CV)	0.051	**0.250**	**0.736**	**0.909**
	Shrink (FVE80)	0.043	0.080	0.248	0.525
	Shrink (FVE85)	0.048	0.079	0.252	0.531
$(n_1, n_2) = (300, 300)$	Shrink (CV)	0.038	**0.736**	**0.990**	**0.996**
	Shrink (FVE80)	0.040	0.125	0.542	0.907
	Shrink(FVE85)	0.036	0.162	0.555	0.915
$(n_1, n_2) = (600, 600)$	Shrink (CV)	0.046	**0.966**	**1.000**	0.998
	Shrink (FVE80)	0.044	0.219	0.851	**1.000**
	Shrink (FVE85)	0.042	0.289	0.851	**1.000**

Table 2.8: Results for simulation III. True value of p is unknown, and the leave one curve out cross-validation method ("CV") (Peng and Paul, 2009) as well as the fraction of variance explained rule ("FVE80" and "FVE85") are implemented to choose an appropriate p. Performances of our shrink test are evaluated in terms of type I error ($\delta_2 = 0$) and powers ($\delta_2 = 0.3, 0.6, 0.9$) over 1000 repetitions.

sample size	**Testing**	δ_2**=0**	δ_2**=0.3**	δ_2**=0.6**	δ_2**=0.9**
$(n_1, n_2) = (100, 100)$	Shrink (CV)	0.051	0.158	**0.533**	**0.882**
	Shrink (FVE80)	0.066	0.156	0.451	0.810
	Shrink (FVE85)	0.066	**0.159**	0.471	0.831
$(n_1, n_2) = (300, 300)$	Shrink (CV)	0.041	**0.393**	**0.966**	0.998
	Shrink (FVE80)	0.062	0.360	0.896	**0.999**
	Shrink (FVE85)	0.061	0.383	0.924	**0.999**
$(n_1, n_2) = (600, 600)$	Shrink (CV)	0.042	**0.763**	0.997	0.998
	Shrink (FVE80)	0.061	0.678	0.998	**1.000**
	Shrink (FVE85)	0.056	0.722	**0.999**	**1.000**

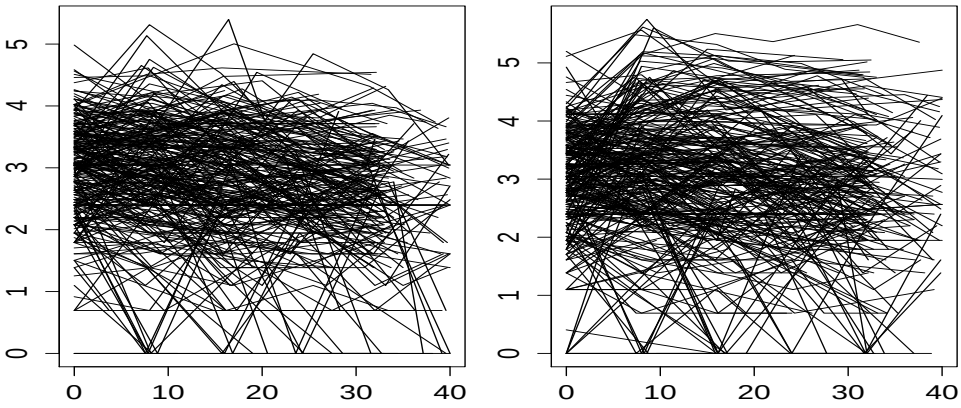

Figure 2.8: Plots of CD4 trajectories in group A are on the left panel; plots of CD4 trajectories in group A are on the left panel

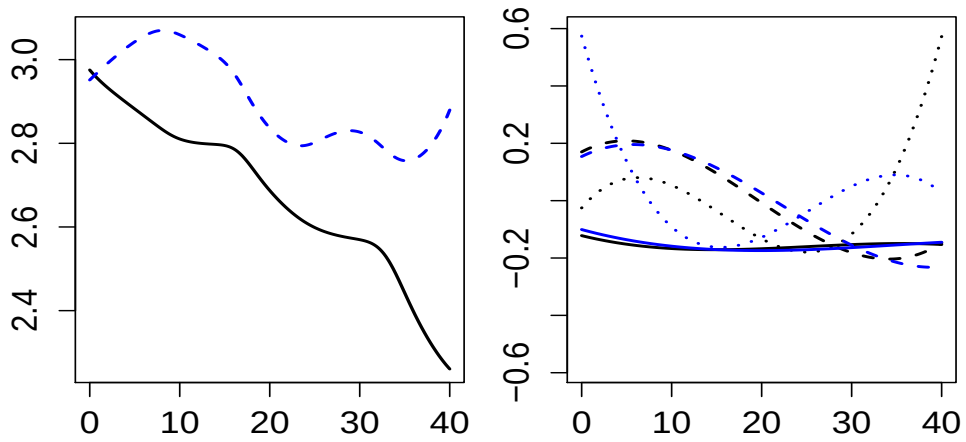

Figure 2.9: Left: estimated mean CD4 curves of two treatment groups (group A: solid black; group B: dash blue); Right: estimated first three eigenfunctions of each group (group A: solid black; group B: dash blue)

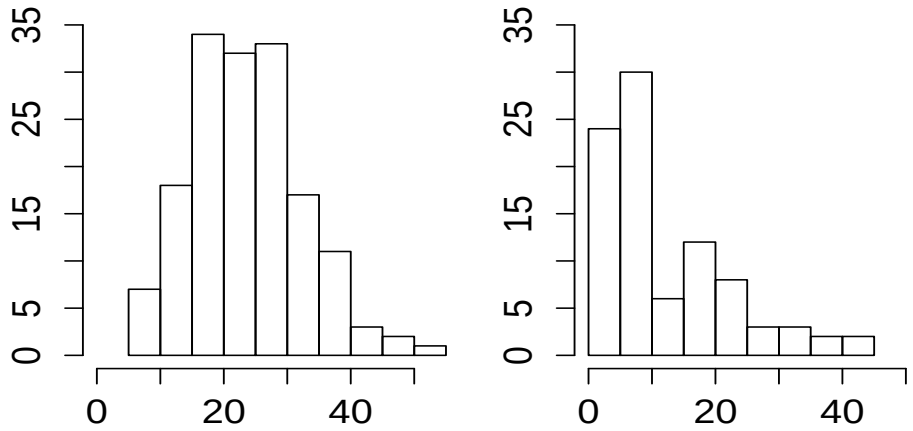

Figure 2.10: Left: histogram of number of observation within each item in the 3 days group; Right: histogram of number of observation within each item in the 7 days group

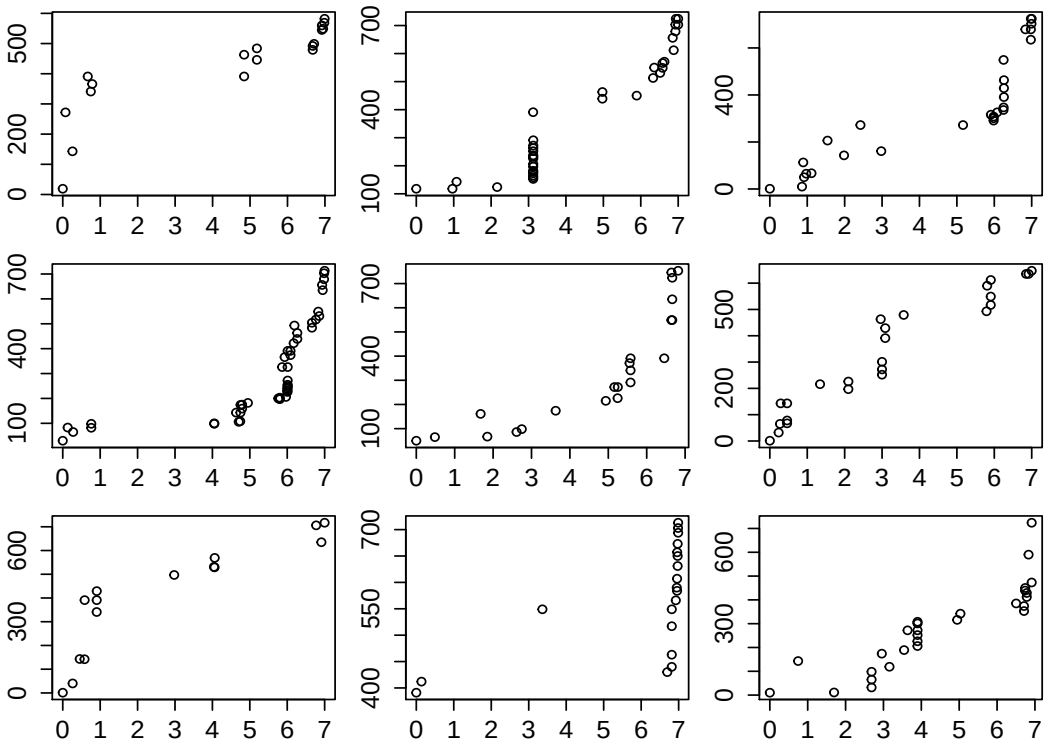

Figure 2.11: Live bids for 9 randomly chosen Palm M515 in the 7 days group.

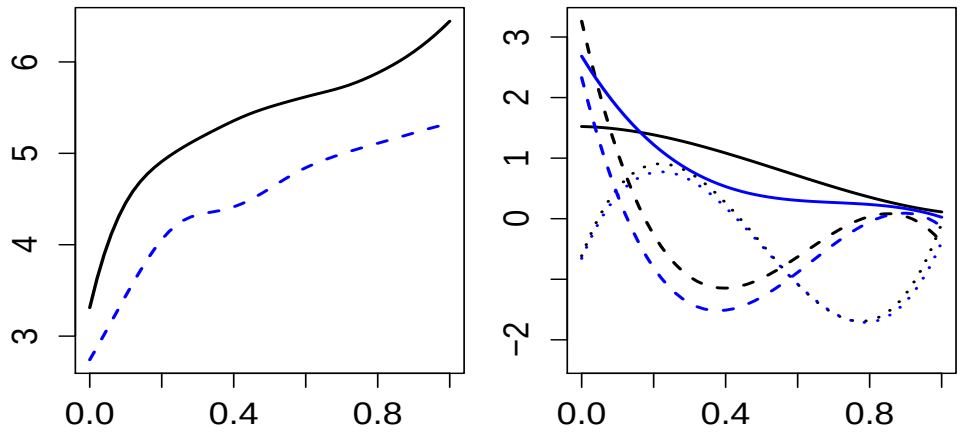

Figure 2.12: Left: estimated mean log(price) curves of two bidding groups (3 days: dashed blue; 7 days: solid black); Right: estimated first three eigenfunctions of each group (3 days: blue; 7days: black)

3.0 ONE-WAY FUNCTIONAL ANOVA FOR SPARSE FUNCTIONAL DATA

For dense functional data, some of the six methods mentioned in Chapter 2 can be easily generalized to multiple sample circumstances, while some are only applicable to the two-sample inference problem. In Section 3.2, we briefly talk about the possibility of extending them one by one.

For sparse functional data, there are no methods specifically designed for the one-way functional ANOVA problem. We propose an asymptotic χ^2 test based on the shrinkage score estimator introduced in Chapter 2. The test statistic is a summation of multiple square forms.

This chapter is organized as follows. The general statistical framework of one-way functional ANOVA is given in Section 3.1. Reviews for dense functional data is in Section 3.2. Proposed method for sparse functional data is deliberately discussed in Section 3.3 to Section 3.4, where the derivation, estimation methods, asymptotic results, and numerical experiments are presented, including an application to an eBay online auction data set.

3.1 STATISTICAL FRAMEWORK

For one-way functional ANOVA, the model we consider is

$$Y_{gij} = X_{gi}(T_{gij}) + \epsilon_{gij} \tag{3.1}$$

where $X_{gi} \sim SP(\mu_g(t), G(s,t))$, $g = 1, .., G(G > 2)$, $i = 1, ..., n_g$, $j = 1, ..., N_{gi}$. Y_{gij} denotes the jth observation of the ith subject in group g, and it is observed at time T_{gij}. The random samples X_{gi} are realizations of G independent Gaussian processes with homogeneous covariance structures in a bounded time domain $\mathcal{T} = [0, 1]$. Namely, X_{gi} are random samples from $X_g \sim$

$GP(\mu_g(t), G(s,t))$, $s,t \in \mathcal{T}$. The number of observations for the ith subject in group g is denoted as N_{gi}. The corresponding observing times are $\{T_{gi1}, ..., T_{giN_{gi}}\}$.

For dense functional data, the number of observations within each subject, i.e., N_{gi}, goes to infinity with a high rates and the observing times are usually regular. The entire underlying curve X_{gi} can be consistently recovered from $(Y_{gi1}, ..., Y_{giN_{gi}})$. Consequently, the observing times are only assumed to be common across subjects and are assumed to be equally spaced without loss of generality. Let's denote the common observing times as $\{T_1, ..., T_N\}$.

For sparse functional data, the number of observations within each subject N_{gi} is assumed to be bounded or grows slowly with n_g. random observing times $\{T_{gi1}, ..., T_{giN_{gi}}\}$ are i.i.d with a bounded density function (specified in Section 2.7) within the time domain $\mathcal{T}$. It is also assumed that X_{gi}, N_{gi} and ϵ_{gij} are mutually independent.

Based on the observed data Y_{gij}, $g = 1, .., G(G > 2)$, $i = 1, ..., n_g$, $j = 1, ..., N_{gi}$, one-way functional ANOVA is interested in testing

$$H_0 : \mu_1(t) = \mu_2(t) = ... = \mu_G(t),\ t \in \mathcal{T} \quad VS \quad H_a : \text{not } H_0. \tag{3.2}$$

3.2 REVIEWS FOR DENSE FUNCTIONAL DATA

Six dense test methods are reviewed in Section 2.2 for the two-sample mean function inference problem. For each of these tests, we briefly examine whether it can be extended to multiple sample cases (3.2) in this section.

Extension of the pointwise t-test is simple, as there is a multiple sample version of the pointwise t-test, called pointwise F-test. Basically, a F-test, instead of a t-test, is utilized at each local time point t_0.

The L^2-norm-based test can also be generalized to functional ANOVA problems by using a statistical quantity that quantifies the difference among all the G mean functions. To be more specific, the new pivotal quantity is $\sum_{g=1}^{G} n_g[\bar{\check{X}}_{g\cdot}(t) - \bar{\check{X}}_{\cdot\cdot}(t)]^2$, with $\bar{\check{X}}_{g\cdot}(t) = \frac{1}{n_g}\sum_{i=1}^{n_g}\check{X}_{gi}(t)$ and $\bar{\check{X}}_{\cdot\cdot}(t) = \frac{1}{n_1+....+n_G}\sum_{g=1}^{G}\sum_{i=1}^{n_g}\check{X}_{gi}(t)$. Then the test statistic is the integral of this pivotal over the

entire time domain $\mathcal{T}$.

$$\tilde{L}_n = \sum_{g=1}^{G} n_g \int_{\mathcal{T}} [\bar{\check{X}}_{g\cdot}(t) - \bar{\check{X}}_{\cdot\cdot}(t)]^2 dt \sim \sum_{k=1}^{\infty} \lambda_k B_k, \text{ with } B_k \sim \text{i.i.d } \chi^2(G-1) \tag{3.3}$$

where λ_k, $k = 1, ..., \infty$ are the eigenvalues of the common covariance function $G(s,t)$.

Generalizing the globalized F-test is originally designed for the one-way functional ANOVA problem. Compared with the test statistic defined in (2.11), the degree of freedom of the independent χ^2 distribution is now $G-1$ instead of 1.

For the eigen-space projection based test, it can also be generalized to multiple cases. The problem basically becomes testing the equality of multiple mean vectors, whose test statistic is the summation of multiple square forms. Mathematically, let $\hat{v}_{gk} =< \bar{\check{X}}_{g\cdot} - \bar{\check{X}}_{\cdot\cdot}, \hat{\phi}_k >$, $k = 1, ..., \infty$, the test statistic is

$$\tilde{D}_n = \sum_{g=1}^{G} \sum_{k=1}^{p} \frac{n_g \hat{v}_{gk}^2}{\hat{\lambda}_k}, \tag{3.4}$$

because under the H_0 of (3.2) and Gaussian processes, $\tilde{D}_n \sim \chi^2(p(G-1))$.

The pLRT test can not be easily utilized to the one-way functional ANOVA problem. All their theory is designed for one-sample inference problem. When there are more than two samples being considered, the original problem can no longer be easily transformed into a one-sample inference problem. This is one of the shortcomings of the pLRT test.

The distribution test can be used to test (3.2) as long as there is a multiple version of the nonparametric AD test, which is capable of testing the equality of distributions of certain univariate variables across multiple groups.

3.3 PROPOSED METHOD FOR SPARSE FUNCTIONAL DATA

Under sparse functional data cases, by arguments similar to two sample situations, we can see that test (3.2) is equivalent to

$$H_0 : \theta_{1k}^c = \theta_{2k}^c = ... = \theta_{Gk}^c \quad k \geq 1 \tag{3.5}$$

against the general alternative that there exist $k \geq 1$ such that at least two of $\theta^c_{1k}, ..., \theta^c_{Gk}$ are not equal, where $\theta^c_{gk} =< \mu_g - \mu_{pool}, \phi_k >, g = 1, ..., G; k = 1, ..., p$. In order to create test procedures for test (3.5), given the expansion

$$X^c_{gi}(t) = \sum_{k=1}^{\infty} \langle X^c_{gi}, \phi_k \rangle \phi_k(t) \stackrel{\text{def}}{=} \sum_{k=1}^{\infty} r^c_{gik} \phi_k(t), \tag{3.6}$$

a straightforward thinking is to form a test statistic based on r^c_{gik}, $g = 1, G; i = 1, ..., n_g; k = 1, ..., \infty$.

Similar to the two-sample situation in Chapter 2, we propose to use the best linear unbiased predictor of the projection score vector $\boldsymbol{r}^c_{gi}$. Under joint Gaussian assumptions, the predictor $E[\boldsymbol{r}^c_{gi}|\boldsymbol{Y}^c_{gi}]$ is

$$E[\boldsymbol{r}^c_{gi}|\boldsymbol{Y}^c_{gi}] = \boldsymbol{\theta}^c_g + diag(\boldsymbol{\lambda})\boldsymbol{\Phi}^T_{gi}\boldsymbol{\Sigma}^{-1}_{\mathbf{Y}^c_{gi}}(\boldsymbol{Y}^c_{gi} - \boldsymbol{\mu}_{gi} + \boldsymbol{\mu}_{pool,gi}), \tag{3.7}$$

Under the null hypothesis, $E[\boldsymbol{r}^c_{gi}|\boldsymbol{Y}^c_{gi}]$ in (3.7) becomes the following ***shrinkage score***

$$\tilde{\boldsymbol{r}}^c_{gi} = diag(\boldsymbol{\lambda})\boldsymbol{\Phi}^T_{gi}\boldsymbol{\Sigma}^{-1}_{\mathbf{Y}^c_{gi}}\mathbf{Y}^c_{gi} = diag(\boldsymbol{\lambda})\boldsymbol{\Phi}^T_{gi}\boldsymbol{\Sigma}^{-1}_{\mathbf{Y}^c_{gi}}(\mathbf{Y}_{gi} - \boldsymbol{\mu}_{pool,gi}). \tag{3.8}$$

The mean and covariance of $\tilde{\boldsymbol{r}}^c_{gi}$ are

$$\begin{aligned} E[\tilde{\boldsymbol{r}}^c_{gi}] &= E_t[diag(\boldsymbol{\lambda})\boldsymbol{\Phi}^T_{gi}\boldsymbol{\Sigma}^{-1}_{\mathbf{Y}^c_{gi}}(\boldsymbol{\mu}_{gi} - \boldsymbol{\mu}_{pool,gi})] \\ Cov[\tilde{\boldsymbol{r}}^c_{gi}] &= E_t[diag(\boldsymbol{\lambda})\boldsymbol{\Phi}^T_{gi}\boldsymbol{\Sigma}^{-1}_{\mathbf{Y}^c_{gi}}\boldsymbol{\Phi}_{gi}diag(\boldsymbol{\lambda})] \end{aligned} \tag{3.9}$$

All the components in $\tilde{\boldsymbol{r}}^c_{gi}$ and $Cov[\tilde{\boldsymbol{r}}^c_{gi}]$ can be estimated by slightly modifying procedures described in Section 2.5. By plugging int the estimated quantities, $\hat{\boldsymbol{\lambda}}$, $\hat{\phi}_k(t)$, $\hat{\mu}_{pool}(t)$, and $\hat{\sigma}$, we have the following empirical estimators

$$\begin{aligned} \hat{\tilde{\boldsymbol{r}}}^c_{gi} &= diag(\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\Phi}}^T_{gi}[\hat{\boldsymbol{\Phi}}_{gi}diag(\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\Phi}}^T_{gi} + \hat{\sigma}^2\mathbf{I}]^{-1}(\mathbf{Y}_{gi} - \hat{\boldsymbol{\mu}}_{pool;gi}) \\ \hat{Cov}[\tilde{\boldsymbol{r}}^c_{gi}] &= \frac{1}{n_1 + ... + n_G}\sum_{g=1}^{G}\sum_{i=1}^{n_g} diag(\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\Phi}}^T_{gi}[\hat{\boldsymbol{\Phi}}_{gi}diag(\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\Phi}}^T_{gi} + \hat{\sigma}^2\mathbf{I}]^{-1}\hat{\boldsymbol{\Phi}}_{gi}diag(\hat{\boldsymbol{\lambda}}) \\ &\stackrel{\text{def}}{=} \hat{\boldsymbol{V}}_2 \end{aligned} \tag{3.10}$$

Consequently, we propose to use

$$TG_{p,N} = \sum_{g=1}^{G}(\bar{\hat{\tilde{\boldsymbol{r}}}}^c_{g\cdot} - \bar{\hat{\tilde{\boldsymbol{r}}}}^c_{\cdot\cdot})^T(\frac{\hat{\boldsymbol{V}}_2}{n_g})^{-1}(\bar{\hat{\tilde{\boldsymbol{r}}}}^c_{g\cdot} - \bar{\hat{\tilde{\boldsymbol{r}}}}^c_{\cdot\cdot}) \sim \chi^2((G-1)p) \tag{3.11}$$

where $\overline{\hat{\hat{\boldsymbol{r}}}}^{c}_{..} = \frac{n_1 \overline{\hat{\hat{\boldsymbol{r}}}}^{c}_{1.} + ... + n_G \overline{\hat{\hat{\boldsymbol{r}}}}^{c}_{G.}}{n}$, with $n = n_1 + n_2 + ... + n_G$. Even though the proposed test statistic $TG_{p,N}$ is motivated from joint Gaussian situations, it has an asymptotic $\chi^2((G-1)p)$ null distribution regardless of whether the joint Gaussian assumption is true or not.

The theoretical results for $TG_{p,N}$ is analogous to those for the two-sample inference problem.

Theorem 3. *Under H_0 and regularization assumptions, assuming $\lim_{n_g \to \infty} \frac{n_g}{n} = w_g$, with $w_g \in (0,1)$ and $n = n_1 + ... + n_G$, if $E[\boldsymbol{\Phi}_{gik}^T \boldsymbol{\Sigma}_{Y_{gi}}^{-1} \boldsymbol{\Phi}_{gik}] < \infty$ is satisfied, we have*

$$TG_{p,N} \xrightarrow{D} \chi^2((G-1)p) \tag{3.12}$$

Given Theorem 3, it is reasonable for us to reach our conclusion based on the following rule: for a specific significance level α, reject H_0 in (3.2), if $TG_{p,N} > \chi^2(1-\alpha; (G-1)p)$, the upper α quantile of $\chi^2((G-1)p)$; otherwise, we will not have enough evidence to reject H_0. Theorem 3 ensures good performance for the proposed test statistic $TG_{p,N}$ when the sample sizes n_g go to infinity proportionally. Finite sample performances are justified through numerical experiments in Section 3.4.

Next, we want to examine the performance of our test statistic under certain types of alternatives in (3.2).

Theorem 4. *Under H_a and regularization assumptions, assuming $\lim_{n_1.n_2 \to \infty} \frac{n_g}{n} = w_g$, with $w_g \in (0,1)$ and $n = n_1 + ... + n_G$, and $\exists\, g$, $E_t(diag(\boldsymbol{\lambda}) \boldsymbol{\Phi}_{gi}^T \boldsymbol{\Sigma}_{Y_{gi}}^{-1} (\boldsymbol{\mu}_{gi} - \boldsymbol{\mu}_{pool;gi})) \neq \mathbf{0}$, for some $1 \leq k \leq p$, we have*

$$TG_{p,N} \xrightarrow{p} \infty \tag{3.13}$$

$E_t(diag(\boldsymbol{\lambda}) \boldsymbol{\Phi}_{gi}^T \boldsymbol{\Sigma}_{Y_{gi}}^{-1} (\boldsymbol{\mu}_{gi} - \boldsymbol{\mu}_{pool;gi}))$ is actually a type of projection for the function $\mu_g(t) - \mu_{pool}(t)$. Under the alternative in (3.2), it is really rare that these projection score of all G groups equals to 0 for all the first p directions. Theorem 4 ensures that, when the mean function differences are captured by these projection scores, the power of our test procedure goes to 1, when sample sizes n_g go to infinity proportionally. Proof of the above theorems are similar to the proof of Theorem 1 and Theorem 2, and thus we omit the details.

3.4 NUMERICAL EXPERIMENTS

3.4.1 Simulation studies

In this subsection, we examine the finite sample performance of the proposed test procedure $TG_{p,N}$ through two simulations studies. For both simulation studies, we consider comparing the mean functions of four independent groups ($G = 4$).

Simulation I

The data is generated based on model (3.1), and the Karhunen-Loève expansion $X_{gi}(T_{gij}) = \mu_g(T_{gij}) + \sum_{k=1}^{\infty} \xi_{gik}\phi_k(T_{gij})$, with $\xi_{gik} \sim N(0, \lambda_k)$. The number of observations on each curve is 4, i.e., $N_{gi} = N = 4$ for $g = 1, ..., G; i = 1, ..., n_g$. The observing time points, $\{T_{gi1}, ..., T_{giN}\}$, are uniformly distributed within the interval [0, 1]. We assume that the eigenvalues are $\lambda_k = 2(k+1)^{-2}$ for $k = 1, ..., 4$, and $\lambda_k = 0$ for $k > 4$. The eigenfunctions are $\phi_1(t) = \sqrt{2}\cos(\pi t)$, $\phi_2(t) = \sqrt{2}\sin(\pi t)$, $\phi_3(t) = \sqrt{2}\cos(2\pi t)$, $\phi_4(t) = \sqrt{2}\sin(2\pi t)$. All the mean functions are assumed to be linear combinations of the first four eigenfunctions. That is to say $\mu_1 = \sum_{k=1}^{4} a_{1k}\phi_k(t)$, where $a_{1k} = \frac{3\bar{a}_{1k}}{\|\bar{a}_1\|_2}$, with $\bar{a}_1 = (\bar{a}_{11}, ..., \bar{a}_{14})^T$ and $\bar{a}_{1k} = (2)^{5-k}(5-k)^6$. And $\mu_2 = \sum_{k=1}^{4} a_{2k}\phi_k(t)$, where $a_{2k} = \frac{3\bar{a}_{2k}}{\|\bar{a}_2\|_2}$, with $\bar{a}_2 = (\bar{a}_{21}, ..., \bar{a}_{24})^T$ and $\bar{a}_{2k} = (2-\varsigma_1)^{5-k}(5-k)^6$. And $\mu_3 = \sum_{k=1}^{4} a_{3k}\phi_k(t)$, where $a_{3k} = \frac{3\bar{a}_{3k}}{\|\bar{a}_3\|_2}$, with $\bar{a}_3 = (\bar{a}_{31}, ..., \bar{a}_{34})^T$ and $\bar{a}_{3k} = (2-\varsigma_2)^{5-k}(5-k)^6$. And $\mu_4 = \sum_{k=1}^{4} a_{4k}\phi_k(t)$, where $a_{4k} = \frac{3\bar{a}_{4k}}{\|\bar{a}_4\|_2}$, with $\bar{a}_4 = (\bar{a}_{41}, ..., \bar{a}_{44})^T$ and $\bar{a}_{4k} = (2-\varsigma_3)^{5-k}(5-k)^6$. ς_1, ς_2 and ς_3 quantifies the magnitude of the distinctions among the mean functions. Note that $(\varsigma_1, \varsigma_2, \varsigma_3)^T = (0, 0, 0)^T, (0.6, 0.8, 1.0)^T, (1.0, 1.2, 1.4)^T$, and $(1.4, 1.6, 1.8)^T$ correspond to the null hypothesis and three alternative hypotheses for (3.2). The standard deviation of the measurement errors is $\sigma = 1.2$. The sample sizes are $n_1 = n_2 = n_3 = n_4 = 100$, $n_1 = n_2 = n_3 = n_4 = 300$, and $n_1 = n_2 = n_3 = n_4 = 600$. The eigenfunctions used here are exactly the same as as Simulation I in Chapter 2. Visualizations for the mean functions are given in Figure 3.2.

The true p is utilized in all the calculation in this simulation study. Based on Table 3.1, it can be seen that the proposed test (3.8) is valid, even though only finite samples are provided. Under all sample size situations, $TG_{p,N}$ (3.8) succeeds in controlling the pre-specified significance level $\alpha = 0.05$. For a given alternative, its power increases with the increasing of sample sizes. For a given sample size setting, the power is larger when the discrepancy between H_0 and H_a is larger.

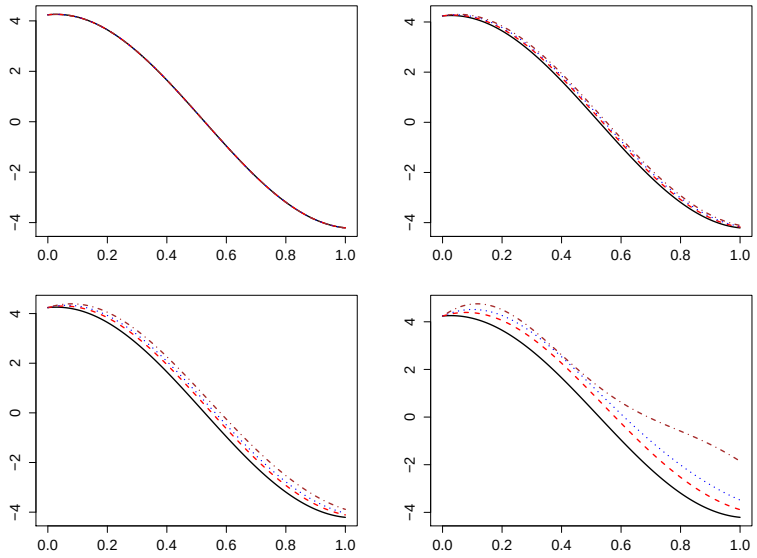

Figure 3.1: Mean functions under the null and three alternative hypotheses of Simulation I.

Table 3.1: Results for simulation I. We presented results for the number of observation on each random curve is 4 ($N = 4$) and the standard deviation of the random measurement error is 1.2 ($\sigma = 1.2$), different sample sizes ($n_g = 100, 300,$ or 600). The type I error ($(\varsigma_1, \varsigma_2, \varsigma_3)^T = (0, 0, 0)^T$) and powers ($H_{a1} : (\varsigma_1, \varsigma_2, \varsigma_3)^T = (0.6, 0.8, 1.0)^T; H_{a2} : (\varsigma_1, \varsigma_2, \varsigma_3)^T = (1.0, 1.2, 1.4)^T; H_{a3} : (\varsigma_1, \varsigma_2, \varsigma_3)^T = (1.4, 1.6, 1.8)^T$) are calculated based on 1000 repetitions. The significance level α is 0.05.

sample size	$\boldsymbol{H_0}$	$\boldsymbol{H_{a1}}$	$\boldsymbol{H_{a2}}$	$\boldsymbol{H_{a3}}$
$(n_1, n_2, n_3, n_4) = (100, 100, 100, 100)$	0.034	0.195	0.884	1.000
$(n_1, n_2, n_3, n_4) = (300, 300, 300, 300)$	0.031	0.649	1.000	1.000
$(n_1, n_2, n_3, n_4) = (600, 600, 600, 600)$	0.036	0.955	1.000	1.000

Simulation II

Now we adopt two more complex settings. Under the first setting, the number of observations varies across subjects, $N_{gi} \sim U[2, 3, 4, 5, 6]$. In the meanwhile, the true value of p is no longer given. Eigenfunctions are $\phi_{2k-1}(t) = \sqrt{2}\cos((2k-1)\pi t)$ for $k = 1, ..., 15$, $\phi_{2k}(t) = \sqrt{2}\sin(2k\pi t)$

for $k = 1, ..., 15$. The mean functions are $\mu_g(t) = \sum_{k=1}^{30} a_{gk}\phi_k(t)$, where $a_{gk} = 3 * \frac{\bar{a}_{gk}}{\|\bar{a}\|_2}$, with $\bar{a} = (\bar{a}_{g1}, ..., \bar{a}_{g30})^T$, $\bar{a}_{1k} = (2)^{31-k}(31-k)^6$, $\bar{a}_{2k} = (2+\delta_1)^{31-k}(31-k)^6$, $\bar{a}_{3k} = (2+\delta_2)^{31-k}(31-k)^6$, and $\bar{a}_{4k} = (2+\delta_3)^{31-k}(31-k)^6$, with $(\delta_1, \delta_2, \delta_3)^T = (0,0,0)^T$, $(0.2, 0.4, 0.6)^T$, $(0.8, 1.2, 1.6)^T$, $(1.2, 1.6, 2.0)^T$. The corresponding eigenvalues are $\lambda_k = 2(k+1)^{-2}$ for $k = 1, ..., 30$. The standard deviation of random measurement errors is $\sigma = 1$. The mean functions under the null and three different alternatives are illustrated by Figure 3.2. Data for 9 randomly selected subjects are provided in Figure 3.3. This setting represents situations where p is rather large. Under the second setting, the mean functions are $\mu_1(t) = 1.08 - 3.2t$ and $\mu_g(t) = \mu_1(t) + (1-t)\nu_g$, $g = 2, 3, 4$. Values of $(\nu_1, \nu_2, \nu_3)^T$ are $(0,0,0)^T$, $(0.2, 0.3, 0.4)^T$, $(0.4, 0.6, 0.8)^T$, and $(0.6, 0.9, 1.2)^T$ (Figure 3.4). The eigenfunctions, eigenvalues and the random error standard deviation share the same values as the first setting.

For both settings, we exploit two methods to choose p: the cross-validation approach described in (Peng and Paul, 2009) and the fraction of variance explained method (FVE), with threshold 80% (FVE80) and 90% (FVE90). Combining the results in Table 3.2 and Table 3.3, we can see that our proposed test $TG_{p,N}$ has reasonable performance regardless which method of choosing p is utilized. The cross-validation method in (Peng and Paul, 2009) tends to provide higher powers.

3.4.2 Application to an eBay auction data set

In Chapter 2, we use the live bid for Palm M515 Personal Digital Assistants as an example to analyze eBay online auction data through our proposed method $T_{p,n}$. In that part, only two independent groups are considered, i.e., auctions with a 3-days duration and auctions with a 7-days duration. There is also a 5-days duration option for eBay sellers. Now we want to consider all these three groups at the same time to see whether the average bidding curves are different across different duration groups.

The data set can be found at (Jank and Shmueli, 2010). In the 3-days group, there are 90 auctions of Palm M515 that happened between March and May, 2003. In the 5-days and 7-days group, there are respectively 49 and 158 auctions of the same item happened within the same period of time. As shown in Figure 2.11, the bids are extremely irregular. There are big gaps among the data. Similar to Section 2.8.3, we model the data from sparse functional data perspective. We

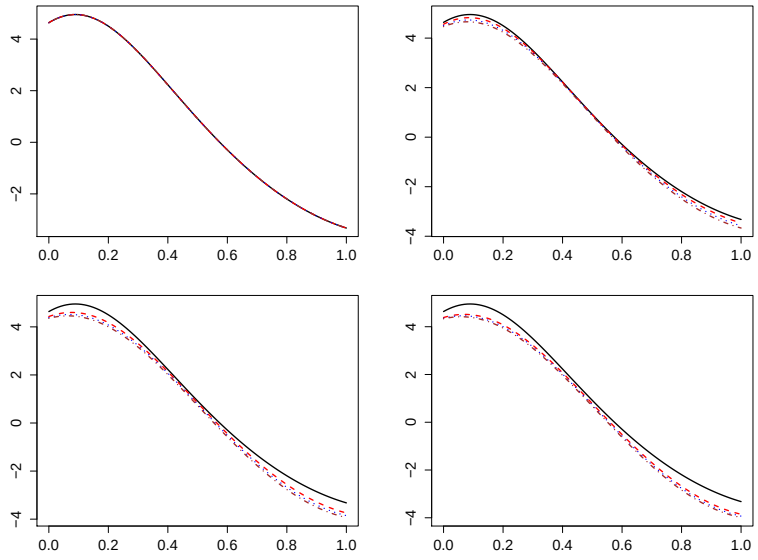

Figure 3.2: Mean functions under the null and three alternative hypotheses of the first setting in Simulation II

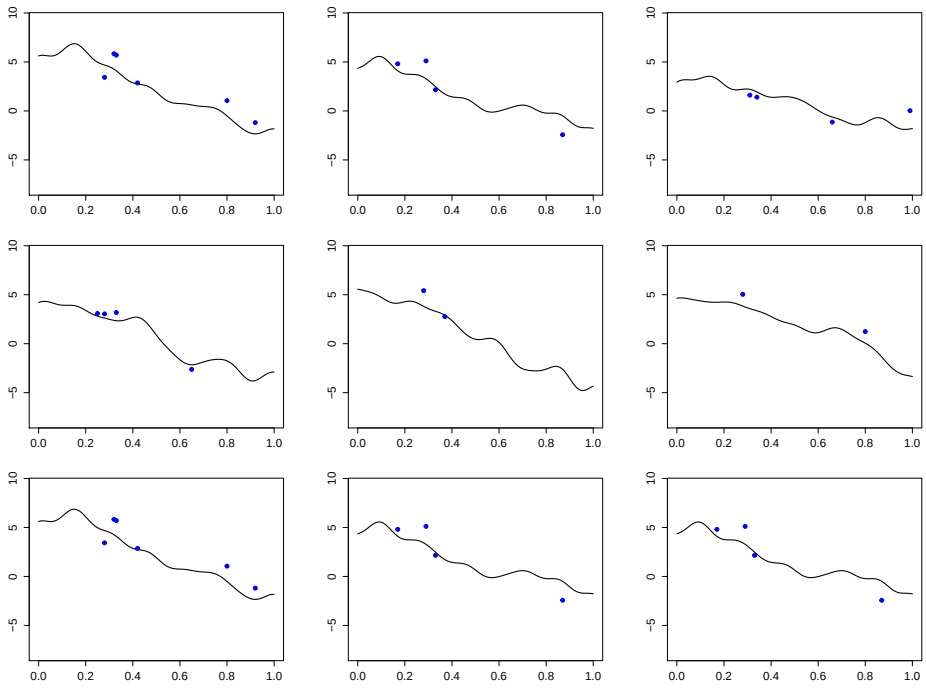

Figure 3.3: Data for nine randomly chosen subjects under the first setting of Simulation II.

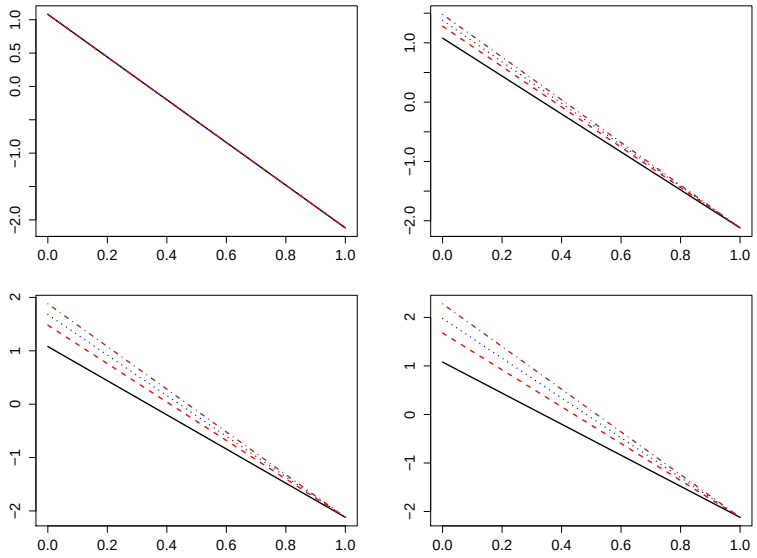

Figure 3.4: Mean functions under the null and three alternative hypotheses of the second setting in Simulation II.

assume that each group of data are i.i.d realizations of an underlying continuous stochastic process, with its own mean function and covariance function. we transform live bid data into log-scale. We also scale the bidding time variables of all the 3 days group , the 5 day group and the 7 days group to $[0, 1]$. Now the problem of detecting differences in average price trends becomes testing whether the three population mean curves are significantly different within some parts of $[0, 1]$.

Before applying our proposed method, we first explore the covariance functions of the three different groups. As indicated by the right panel in Figure 3.5, the estimated first three eigenfunctions of the 7 days group (solid black), the 5 days group (dash blue) and the 3 days group (dash red) are very close. The corresponding eigenvalues are $(0.344, 0.101, 0.020)^T$ for the 7 days group, $(0.602, 0.093, 0.023)^T$ for the 7 days group and $(0.416, 0.162, 0.011)^T$ for the 3 days group. It is reasonable to assume that they share the same covariance function. The proposed χ^2 test is implemented to this mean function testing problem. We use the cross-validation method to choose $p = 5$ and the p-value of our proposed test is less than 0.000001. This means that we have enough confidence to conclude that the mean price evolution curves are different when different bidding time periods are selected. The estimated mean functions by local linear smoothing are included on

Table 3.2: Results for simulation II (the first setting). True value of p is 30, and the leave one curve out cross-validation method ("CV") (Peng and Paul, 2009) as well as the fraction of variance explained rule ("FVE80" and "FVE90") are implemented to choose an appropriate p. Performances of our shrink test are evaluated in terms of type I error ($(\delta_1, \delta_2, \delta_3)^T = (0, 0, 0)^T$) and powers ($H_{a1} : (\delta_1, \delta_2, \delta_3)^T = (0.2, 0.3, 0.4)^T; H_{a2} : (\delta_1, \delta_2, \delta_3)^T = (0.4, 0.6, 0.8)^T; H_{a3} : (\delta_1, \delta_2, \delta_3)^T = (0.6, 0.9, 1.2)^T$) over 1000 repetitions.

sample size	**Testing**	$\boldsymbol{H_0}$	$\boldsymbol{H_{a1}}$	$\boldsymbol{H_{a2}}$	$\boldsymbol{H_{a3}}$
$(n_1, n_2, n_3, n_4) = (100, 100, 100, 100)$	shrink(CV)	0.047	**0.192**	**0.697**	**0.962**
	shrink(FVE80)	0.032	0.052	0.092	0.554
	shrink (FVE90)	0.037	0.066	0.109	0.562
$(n_1, n_2, n_3, n_4) = (300, 300, 300, 300)$	shrink (CV)	0.036	**0.556**	**0.989**	**0.986**
	shrink (FVE80)	0.051	0.099	0.263	0.678
	shrink (FVE90)	0.033	0.141	0.260	0.681
$(n_1, n_2, n_3, n_4) = (600, 600, 600, 600)$	shrink (CV)	0.030	**0.923**	**0.993**	**0.998**
	shrink (FVE80)	0.043	0.159	0.599	0.830
	shrink (FVE90)	0.037	0.264	0.581	0.786

the left panel in Figure 3.5. From the graph, we can see that the price of the 7 days group (solid black) is larger than that of the 5 days group (dash blue) and that of the 3 days group (dash red) over the entire time domain.

Table 3.3: Results for simulation II (the second setting). True value of p is unknown, and the leave one curve out cross-validation method ("CV") (Peng and Paul, 2009) as well as the fraction of variance explained rule ("FVE80" and "FVE90") are implemented to choose an appropriate p. Performances of our shrink test are evaluated in terms of type I error ($(\nu_1, \nu_2, \nu_3)^T = (0, 0, 0)^T$) and powers ($(\nu_1, \nu_2, \nu_3)^T = (0.2, 0.3, 0.4)^T$; $(0.4, 0.6, 0.8)^T$, $(0.6, 0.9, 1.2)^T$) over 1000 repetitions.

sample size	**Testing**	$\boldsymbol{H_0}$	$\boldsymbol{H_{a1}}$	$\boldsymbol{H_{a2}}$	$\boldsymbol{H_{a3}}$
$(n_1, n_2, n_3, n_4) = (100, 100, 100, 100)$	shrink (CV)	0.060	**0.181**	**0.662**	**0.976**
	shrink (FVE80)	0.056	0.153	0.497	0.873
	shrink (FVE90)	0.070	0.163	0.530	0.906
$(n_1, n_2, n_3, n_4) = (300, 300, 300, 300)$	shrink (CV)	0.043	**0.503**	**0.988**	0.998
	shrink (FVE80)	0.068	0.437	0.968	0.999
	shrink (FVE90)	0.059	0.477	0.984	**1.000**
$(n_1, n_2, n_3, n_4) = (600, 600, 600, 600)$	shrink (CV)	0.029	**0.878**	**0.998**	**1.000**
	shrink (FVE80)	0.065	0.787	0.998	**1.000**
	shrink (FVE90)	0.057	0.860	**1.000**	**1.000**

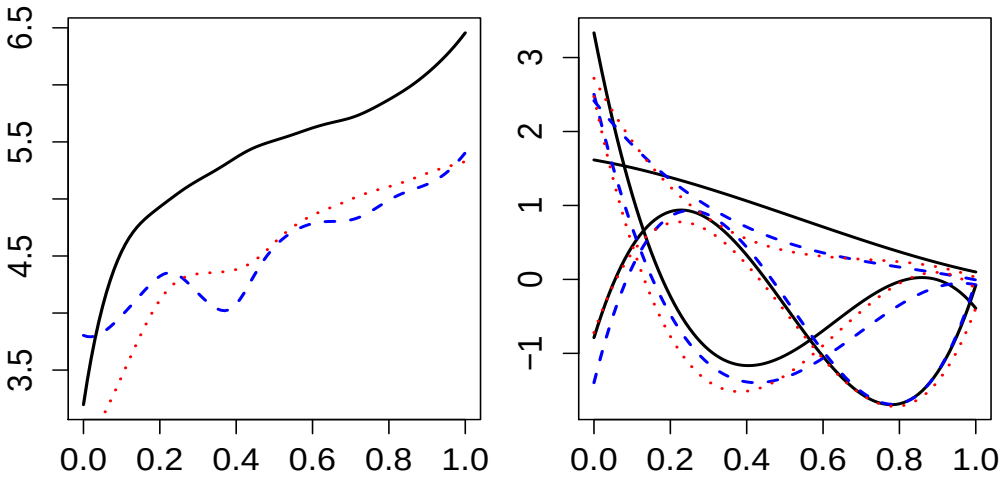

Figure 3.5: Left: estimated mean log(price) curves of three bidding groups (3 days: dash red; 5 days: dash blue; 7 days: solid black); Right: estimated first three eigenfunctions of each groups (3 days: dash red; 5 days: dash blue; 7 days: solid black)

4.0 CHANGE POINT DETECTION FOR SPARSE FUNCTIONAL TIME SERIES

In this chapter, we consider the change point detection problem in functional time series, whose objective is to detect possible changes among the mean functions of a sequence of independent stochastic processes. This problem is usually modeled from hypothesis testing perspectives. The null hypothesis is that all the random functions in the sequence share the same mean function. As for the alternative, there are two typical types: at most one change point (AMOC), where the mean function changes at most once; and epidemic changes, where the mean function changes at some time point and then it returns to its original level. In this thesis, we focus on the AMOC problem under sparse functional data settings. Similar ideas can be used to generalize epidemic change point problems, but we do not study the details here. For the rest of this chapter, the two terms functional AMOC and change point detection are used interchangeably. They both refer to the functional AMOC problem. The mathematical model of the AMOC problem can be formalized as follows. Let $L^2(\mathcal{T})$ be the space that is consists of all squared-integrable functions within the closed interval $\mathcal{T}$. And let $X_i(t) \in L^2(\mathcal{T})$, $i = 1, ..., n$, denote the independent random process at time i, with mean function $E[X_i(t)]$. The functional AMOC change point detection problem concentrates on testing $E[X_i(t)] = \mu(t)$ for all i, against the alternative that the mean function $E[X_i(t)]$ equals to $\mu(t)$ before some unknown time point θ and equals to $\mu'(t) \neq \mu(t)$ for $i = \theta + 1, ..., n$. The unknown change point θ can be any value within the range of 1 and $n - 1$, and we want to test whether the change exists.

For dense functional data, the change point detection problem has been considered by (Berkes et al., 2009) and (Aue et al., 2009). They both utilize functional principal component analysis to reduce the infinite dimensional inference problem to a finite approximate in the eigenspace. They solve this problem in two steps. First, they propose consistent testing procedures to infer whether there is a shift in terms of the mean function among the functional time series $X_i(t) \in L^2(\mathcal{T})$,

$i = 1, ..., n$. Next, if a significant shift is detected by their test procedures, they propose estimates for θ, the location of the change point.

For sparse functional data, the change point detection problem has not yet been addressed. In this chapter, we propose a test procedure specifically designed for the sparse functional change point detection problem. A subject-dependent shrinkage estimator similar to Chapter 2 is utilized to construct the current test statistic. Details can be found in Section 4.3. In Section 4.3, an estimate for the location of the change point is also proposed.

This chapter is organized as follows. In Section 4.1, a brief review for the change point detection problem in univariate time series is presented. In Section 4.2 to Section 4.4, the proposed method for sparse functional data is discussed in detail, including the statistical framework, the proposed test statistic, estimation procedures, and numerical experiments.

4.1 REVIEWS FOR THE CHANGE POINT DETECTION PROBLEM IN UNIVARIATE TIME SERIES

The statistical framework for the traditional AMOC change point detection problem can be summarized as follows. Suppose that we have a sequence of independent random variables with homogeneous variance. Let's denote the random series as $\{X_1, ..., X_i, ..., X_n\}$, where X_i denotes the random variable at time i, and n is the length of the random series. The population mean at the ith location is denoted as $E[X_i]$, and the common variance is σ^2. The change point detection problem focuses on testing

$$H_0 : E[X_i] = \mu, i = 1, ..., n \quad VS \quad H_a : E[X_i] = \begin{cases} \mu, & i = 1, ..., \theta \\ \mu', & i = \theta + 1, ..., n \end{cases} \tag{4.1}$$

where θ is the unknown location of the change point, which can take any value within $\{1, ..., n - 1\}$. μ and μ' are the unknown mean values, which are not equal.

According to (Hawkins, 1977), when the random variables are normal distributed, test statistics for the change point detection problem in (4.1) can be constructed from the likelihood ratio test (LRT) point of view. To be more specific, under H_0 and H_a respectively, one should first achieve

the the maximum likelihood estimates (MLE) for all the unknown parameters, including μ, μ', σ, and θ. Then the log likelihood ratio function can be calculated by plugging these estimates back. To simplify the notation, it is assumed that σ is known. Under the null, the only unknown parameter is μ, and its MLE estimate is $\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Under H_a, for any fixed $\theta = k$, the MLE of μ and μ' are

$$\begin{aligned} \hat{\mu} &= \bar{X}_k = \frac{1}{k}\sum_{i=1}^{k} X_i \\ \hat{\mu}' &= \bar{X}_{k'} = \frac{1}{n-k}\sum_{i=k+1}^{n} X_i \end{aligned} \tag{4.2}$$

According to the probability density function of Gaussian distributions, the likelihood function under H_a is monotonically decreasing in $S_k = \sum_{i=1}^{k}(X_i - \bar{X}_k)^2 + \sum_{i=k+1}^{n}(X_i - \bar{X}_{k'})^2$. So the MLE of θ is

$$\hat{\theta} = \arg\min_k S_k \tag{4.3}$$

Let's denote $S = \sum_{i=1}^{n}(X_i - \bar{X})^2$. It is then straightforward to show that

$$\begin{aligned} -2\sigma^2 \log(\text{likelihood ratio}) &= S - S_{\hat{\theta}} \\ &= \frac{n}{\hat{\theta}(n-\hat{\theta})}[\sum_{i=1}^{\hat{\theta}}(X_i - \bar{X})]^2 \\ &= \max_k \left| \frac{n}{k(n-k)} \sum_{i=1}^{k}(X_i - \bar{X}) \right|^2 \\ &\stackrel{\text{def}}{=} U_n \end{aligned} \tag{4.4}$$

where $k = 1, ..., n-1$. It is easy to show that, for any fixed k, $\frac{n}{k(n-k)}\sum_{i=1}^{k}(X_i - \bar{X}) \sim N(0,1)$. The distribution of the LRT test statistic U_n is therefore the maximum absolute value of a discrete Gaussian process, whose mean is zero, variance is 1, and correlation between any position l and m ($l < m$) is

$$\rho_{lm} = \sqrt{\frac{l(n-m)}{m(n-l)}}. \tag{4.5}$$

Specifically, as shown by (Hawkins, 1977), the formula of the probability density function of the test statistic U_n is

$$\begin{aligned} f_{U_n}(x) &= 2z(x,0,1)\sum_{k=1}^{n-1} g_k(x,x)g_{n-k}(x,x), \\ g_k(x,s) &= \int_{y=0}^{s} g_{k-1}(y,s)[z(y,\rho x,\sqrt{1-\rho^2}) + z(y,-\rho x,\sqrt{1-\rho^2})]\,dy. \end{aligned} \tag{4.6}$$

where $\rho = \rho_{k-1,k} = \sqrt{(k-1)(n-k)/k(n-k+1)}$, and $z(x,a,b)$ is the probability density function of $N(a,b^2)$ evaluated at x. This results is derived based on the Markovian property of the discrete Gaussian process and several other probabilistic basics. More details can be found in (Hawkins, 1977). A fractile table is provided in (Hawkins, 1977), which contains simulated quantiles for U_n at different combinations of sequence length n and several commonly-used significance levels.

The LRT test statistic defined in (4.4) intuitively makes sense. For any fixed k, the random quantity $\frac{n}{k(n-k)}\sum_{i=1}^{k}(X_i - \bar{X})$ measures the magnitude of the distinction between the former part $\{X_1, ..., X_k\}$ and the latter part $\{X_{k+1}, ..., X_n\}$. The test statistic uses the maximum distinction among all dividing methods, i.e., $k = 1, ..., n-1$. The Gaussian distribution assumption is required for the LRT test discussed here. Non-Gaussian circumstances are investigated by (Cardot et al., 2013).

4.2 STATISTICAL FRAMEWORK FOR FUNCTIONAL CHANGE POINT DETECTION

Analogous to univariate cases reviewed in Section 4.1, for functional data, it is assumed that we have a sequence of independent stochastic processes with homogeneous covariance structure. That is to say, $X_1(t), ..., X_i(t), ..., X_n(t) \sim SP(E(X_i(t)), G(s,t))$, $t \in \mathcal{T}$. For each stochastic process $X_i(t)$, it is only observed at a finite set of points. And all the observed values are contaminated with random measurement errors. Mathematically, the model we consider is

$$Y_{ij} = X_i(T_{ij}) + \epsilon_{ij}, \tag{4.7}$$

where Y_{ij}, $i = 1, .., n, j = 1, .., N_i$ denotes the jth observation of the random function at the ith time cycle $X_i(t)$. Here ϵ_{ij} are i.i.d Gaussian measurement errors with standard deviation σ. The number of observations for the ith time cycle is denoted as N_i. The corresponding observing times are $\{T_{i1}, ..., T_{iN_i}\}$.

For dense functional data, the number of observations N_i is relatively large which can goes to infinity with a high rate. Similar to the argument in Section 2.1, the observing times $T_{i1}, ..., T_{iN_i}$ are usually common for all i and denoted as $\{T_1, ..., T_N\}$.

For sparse functional data, the number of observations N_i is assumed to be finite or grow slowly with sample size n. Given N_i, random observing times $\{T_{i1}, ..., T_{iN_i}\}$ are i.i.d with a bounded density function (specified in Section 4.3.3) within the time domain $\mathcal{T}$. It is also assumed that X_i, N_i and ϵ_{ij} are mutually independent.

For the functional change point detection problem, the hypothesis of interest is

$$H_0 : E[X_i(t)] = \mu(t), i = 1, ..., n \quad VS \quad H_a : E[X_i(t)] = \begin{cases} \mu(t), & i = 1, ..., \theta \\ \mu'(t), & i = \theta + 1, ..., n \end{cases} \tag{4.8}$$

where $\mu(t) \neq \mu'(t)$ are the corresponding mean functions. Note that $\mu(t)$, $\mu'(t)$, θ and $G(s, t)$ are all unknown.

4.3 PROPOSED METHOD FOR SPARSE FUNCTIONAL TIME SERIES

4.3.1 Derivation of test statistics

To construct a test statistic for (4.8), we exploit a direct extension of the LRT test statistic U_n in (4.4). First, analogous to (4.2), let's introduce several notations. For any fixed $k = 1, ..., n - 1$,

$$\begin{aligned} \bar{X}_k(t) &= \frac{1}{k} \sum_{i=1}^{k} X_i(t) \\ \bar{X}'_k(t) &= \frac{1}{n-k} \sum_{i=k+1}^{n} X_i(t). \end{aligned} \tag{4.9}$$

Then the following stochastic process quantifies the difference in the mean functions between the first k and the latter $n-k$ random functions

$$\begin{aligned} P_k(t) &= \frac{k(n-k)}{n}[\bar{X}_k(t) - \bar{X}'_k(t)] \\ &= \sum_{i=1}^{k} X_i(t) - \frac{k}{n}\sum_{i=1}^{n} X_i(t) \\ &= \sum_{i=1}^{k}[X_i(t) - \mu_{pool}(t)] - \frac{k}{n}\sum_{i=1}^{n}[X_i(t) - \mu_{pool}(t)], \end{aligned} \tag{4.10}$$

where μ_{pool} is the mean function of the mixture stochastic process of $\{X_i(t), i = 1, ..., k\}$ and $\{X_i(t), i = k+1, ..., n\}$. Mathematically,

$$\mu_{pool}(t) = \zeta\mu(t) + (1-\zeta)\mu'(t), \tag{4.11}$$

where $\zeta = \theta/n$.

If the mean function changes, then $P_k(t)$ should be large for some k and some t. The quantity is infinite dimensional, we project it on to the orthonormal eigen space of the covariance function $\tilde{G}(s,t) = G(s,t) + \zeta(1-\zeta)[\mu(t) - \mu'(t)][\mu(s) - \mu'(s)]$. Let $\eta_l(t)$ being the eigenfunction corresponding to the lth largest eigenvalue ω_l of $\tilde{G}(s,t)$. Let's the define eigen projection scores of $X_i(t) - \mu_{pool}(t)$ as $\xi_{il} = \int[X_i(t) - \mu_{pool}(t)]\eta_l(t)\,dt$ for $i = 1, ..., n; l = 1, ..., \infty$. Then the projection score of $P_k(t)$ onto $\eta_l(t)$

$$\int_{t\in\mathcal{T}} P_k(t)\eta_l(t)\,dt = \sum_{i=1}^{k}\xi_{il} - \frac{k}{n}\sum_{i=1}^{n}\xi_{il} \;\; l = 1, ..., \infty. \tag{4.12}$$

If the mean function changes, then $\int_{t\in\mathcal{T}} P_k(t)\eta_l(t)\,dt$ should be large for some k and some l.

For dense functional data, one can truncate at the first p dimensions and construct the following test statistic

$$DC_{p,n} = \frac{1}{n}[\sum_{i=1}^{k}\hat{\boldsymbol{\xi}}_i - \frac{k}{n}\sum_{i=1}^{n}\hat{\boldsymbol{\xi}}_i]^T \text{diag}[\hat{\omega}_1, ..., \hat{\omega}_p]^{-1}[\sum_{i=1}^{k}\hat{\boldsymbol{\xi}}_i - \frac{k}{n}\sum_{i=1}^{n}\hat{\boldsymbol{\xi}}_i] \tag{4.13}$$

where $\boldsymbol{\xi}_i = (\xi_{i1}, ..., \xi_{ip})^T$, $\hat{\boldsymbol{\xi}}_i$ and $\hat{\omega}_l$ are consistent estimates from data Y_{ij}, $i = 1, ..., n; j = 1, ..., N_i$. To obtain the consistent estimate of the projection score ξ_{il}, one can utilize numerical approximations of $\int[X_i(t) - \mu_{pool}(t)]\eta_l(t)\,dt$ based on the discrete observations Y_{ij}, $i = 1, ..., n; j = 1, ..., N_i$.

For sparse functional data, we can no longer achieve consistent estimates for $\boldsymbol{\xi}_i$ through numerical integration due to the limited number of observations on each curve. We propose to use the best linear predictor, $E[\xi_{il}|\boldsymbol{Y}_i]$, with $\boldsymbol{Y}_i = (Y_{i1}, ..., Y_{iN_i})^T$ being the vector contains all the observation at the ith time cycle. This is a reasonable choice due to the fact that

$$E_t[E[\xi_{il}|\boldsymbol{Y}_i]] = E[\xi_{il}], \tag{4.14}$$

when we take the randomness of observing times into account. Under a special case where the projection scores ξ_{il} and the random errors ϵ_{ij} are jointly Gaussian distributed, the best linear predictor $E[\xi_{il}|\boldsymbol{Y}_i]$ has the following explicit formula

$$\begin{aligned} E[\xi_{il}|\boldsymbol{Y}_i] &= \int (\mu_i(t) - \mu_{pool}(t))\eta_l(t)\, dt + \omega_l \boldsymbol{\eta}_{il}^T \boldsymbol{\Sigma}_{\mathbf{Y}_i}^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_{pool,i} - \boldsymbol{\mu}_i + \boldsymbol{\mu}_{pool,i}) \\ &\stackrel{\text{def}}{=} \kappa_l + \omega_l \boldsymbol{\eta}_{il}^T \boldsymbol{\Sigma}_{\mathbf{Y}_i}^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_{pool,i} - \boldsymbol{\mu}_i + \boldsymbol{\mu}_{pool,i}), \end{aligned} \tag{4.15}$$

where $\mu_i(t) = E[X_i(t)]$, $\boldsymbol{\mu}_i = (\mu_i(T_{i1}), ..., \mu_i(T_{iN_i}))^T$, $\boldsymbol{\mu}_{pool,i} = (\mu_{pool}(T_{i1}), ..., \mu_{pool}(T_{iN_i}))^T$, $\boldsymbol{\eta}_{il} = (\eta_l(T_{i1}), ..., \eta_l(T_{iN_i}))^T$, and $\boldsymbol{\Sigma}_{\mathbf{Y}_i}$ is the covariance matrix of $\boldsymbol{Y}_i$, and the (j, j') element of $\boldsymbol{\Sigma}_{\mathbf{Y}_i}$ equals to $\tilde{G}(T_{ij}, T_{ij'}) + \sigma^2 1(j = j')$.

Let's truncate at the first p directions. The best linear predictor of $\boldsymbol{\xi}_i$ is

$$E[\boldsymbol{\xi}_i|\boldsymbol{Y}_i] = \boldsymbol{\kappa} + diag(\boldsymbol{\omega})\boldsymbol{E}_i^T \boldsymbol{\Sigma}_{\mathbf{Y}_i}^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_{pool,i} - \boldsymbol{\mu}_i + \boldsymbol{\mu}_{pool,i}), \tag{4.16}$$

where $\boldsymbol{\kappa} = (\kappa_1, ..., \kappa_p)^T$, $\boldsymbol{\omega} = (\omega_1, ..., \omega_p)^T$, $\boldsymbol{E}_i = (\boldsymbol{\eta}_{i1}, ..., \boldsymbol{\eta}_{ip})$ with $\boldsymbol{\eta}_{il} = (\eta_l(T_{i1}), ..., \eta_l(T_{iN_i}))^T$.

Under the null hypothesis, we have $E[X_i(t)] = \mu_i(t) = \mu_{pool}(t)$, so we propose to construct a test statistic based on the following quantity,

$$\tilde{\boldsymbol{\xi}}_i = diag(\boldsymbol{\omega})\boldsymbol{E}_i^T \boldsymbol{\Sigma}_{\mathbf{Y}_i}^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_{pool,i}), \tag{4.17}$$

which is referred as ***shrinkage score*** in later parts.

To construct a test statistic based on the shrinkage score vector $\tilde{\boldsymbol{\xi}}_i$, let's first calculate its mean vector and covariance matrix through the following two steps. First, assuming the observing times $\boldsymbol{T}_i = (T_{i1}, ..., T_{iN_i})^T$ are fixed, the conditional mean and variance of $\tilde{\boldsymbol{\xi}}_i$ are

$$\begin{aligned} E[\tilde{\boldsymbol{\xi}}_i|\boldsymbol{T}_i] &= diag(\boldsymbol{\omega})\boldsymbol{E}_i^T \boldsymbol{\Sigma}_{\mathbf{Y}_i}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{pool,i}) \\ Cov[\tilde{\boldsymbol{\xi}}_i|\boldsymbol{T}_i] &= diag(\boldsymbol{\omega})\boldsymbol{E}_i^T \boldsymbol{\Sigma}_{\mathbf{Y}_i}^{-1}\boldsymbol{E}_i diag(\boldsymbol{\omega}). \end{aligned} \tag{4.18}$$

Next, by taking the randomness of $\boldsymbol{T}_i$ into account, we have the mean and variance of $\tilde{\boldsymbol{\xi}}_i$ are

$$\begin{aligned} E[\tilde{\boldsymbol{\xi}}_i] &= E_t[diag(\boldsymbol{\omega})\boldsymbol{E}_i^T\boldsymbol{\Sigma}_{\mathbf{Y}_i}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{pool,i})] \\ Cov[\tilde{\boldsymbol{\xi}}_i] &= E_t[diag(\boldsymbol{\omega})\boldsymbol{E}_i^T\boldsymbol{\Sigma}_{\mathbf{Y}_i}^{-1}\boldsymbol{E}_i diag(\boldsymbol{\omega})]. \end{aligned} \tag{4.19}$$

All the quantities involved in estimating $\tilde{\boldsymbol{\xi}}_i$, $E[\tilde{\boldsymbol{\xi}}_i]$ and $Cov[\tilde{\boldsymbol{\xi}}_i]$ can be obtained from data Y_{ij}, $i = 1, .., n, j = 1, .., N_i$, with details included in the next subsection. Let's denote the estimated quantities as $\hat{\boldsymbol{\omega}}$, $\hat{\eta}_k(t)$, $\hat{\mu}_{pool}(t)$, and $\hat{\sigma}$. We have the following empirical estimators

$$\begin{aligned} \hat{\tilde{\xi}}_i &= diag(\hat{\boldsymbol{\omega}})\hat{\boldsymbol{E}}_i^T[\hat{\boldsymbol{E}}_i diag(\hat{\boldsymbol{\omega}})\hat{\boldsymbol{E}}_i^T + \hat{\sigma}^2\mathbf{I}]^{-1}(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_{pool;i}) \\ \hat{Cov}[\tilde{\boldsymbol{\xi}}_i] &= \frac{1}{n}\sum_{i=1}^{n} diag(\hat{\boldsymbol{\omega}})\hat{\boldsymbol{E}}_{gi}^T[\hat{\boldsymbol{E}}_{gi} diag(\hat{\boldsymbol{\omega}})\hat{\boldsymbol{E}}_{gi}^T + \hat{\sigma}^2\mathbf{I}]^{-1}\hat{\boldsymbol{E}}_{gi} diag(\hat{\boldsymbol{\omega}}) \\ &\stackrel{\text{def}}{=} \hat{\boldsymbol{V}}_3 \end{aligned} \tag{4.20}$$

Combining all these results, for the change point $\theta = k$, we propose to use the following test statistic

$$S_{p,n,k}^{(1)} = \frac{1}{n}[\sum_{i=1}^{k}\hat{\tilde{\boldsymbol{\xi}}}_i - \frac{k}{n}\sum_{i=1}^{n}\hat{\tilde{\boldsymbol{\xi}}}_i]^T\hat{\boldsymbol{V}}_3^{-1}[\sum_{i=1}^{k}\hat{\tilde{\boldsymbol{\xi}}}_i - \frac{k}{n}\sum_{i=1}^{n}\hat{\tilde{\boldsymbol{\xi}}}_i] \tag{4.21}$$

Remember that the quantity $S_{p,n,k}^{(1)}$ defined in (4.21) should be small for all k to make H_0 favorable. There are two ways to take k into account. The first one is exactly the same as (4.4). It takes the maximum across k and yields the following test statistic

$$C_{p,n}^{(1)} = \max_{1\le k<n} S_{p,n,k}^{(1)} \tag{4.22}$$

The second one continualizes k by defining $k = [nz]$, $z \in (0, 1)$, then integrates over z,

$$\begin{aligned} C_{p,n}^{(2)} &= \int_0^1 \frac{1}{n}[\sum_{i=1}^{[nz]}\hat{\tilde{\xi}}_i - z\sum_{i=1}^{n}\hat{\tilde{\xi}}_i]^T\hat{\boldsymbol{V}}_3^{-1}[\sum_{i=1}^{[nz]}\hat{\tilde{\xi}}_i - z\sum_{i=1}^{n}\hat{\tilde{\xi}}_i]\,dz \\ &\stackrel{\text{def}}{=} \int_0^1 S_{p,n,z}^{(2)}\,dz \end{aligned} \tag{4.23}$$

Even though the proposed test statistics $C_{p,n}^{(1)}$ and $C_{p,n}^{(2)}$ are motivated from joint Gaussian situations, we later prove that the asymptotic null distribution is valid regardless of whether the joint Gaussian assumption is true or not.

4.3.2 Estimation procedures

In the proposed test statistics $C_{p,n}^{(1)}$ and $C_{p,n}^{(2)}$, the unknown parameters include eigen-values $\boldsymbol{\omega}$ of $\tilde{G}(s,t)$, eigen-functions $\eta_l(t)$, the variance of random measurement errors σ^2, and the overall mean function $\mu_{pool}(t)$. We adopt the pLRT method in (Peng and Müller, 2008) to achieve the estimates for these parameters. In the following description, Gaussian processes are utilized to calculate the likelihood function. However, as stated in their paper, the Gaussian processes assumption is only a working condition and their asymptotic results still hold under some relaxed conditions. To implement our test procedures, one can always utilize other estimating methods, such as the local linear smoothing in (Yao et al., 2005; Li and Hsing, 2010; Zhang and Wang, 2016), as long as they produce consistent estimates. As our test procedure itself does not require Gaussian conditions.

First, in order to the overall mean function $\mu_{pool}(t)$ defined in (4.11), we use the local linear smoothing procedure. To be more specific, we define the local linear smoother of pooled mean function $\mu_{pool}(t)$ by minimizing

$$\sum_{i=1}^{n}\sum_{j=1}^{N_i} K(\frac{T_{ij}-t}{h_{\mu_{pool}}})[Y_{ij}-\beta_0-\beta_1(t-T_{ij})] \tag{4.24}$$

with respect to β_0 and β_1, where $h_{\mu_{pool}}$ is the bandwidth, and $K(\cdot)$ is the smoothing kernel. Then $\hat{\mu}_{pool}(t)=\hat{\beta}_0(t)$.

As for the eigenfunctions $\eta_l(t)$ and eigenvalues ω_l of the covariance function $\tilde{G}(s,t)$, estimates are achieved through the restricted maximum likelihood method in (Peng and Paul, 2009). It's assumed that, under some weak smoothness conditions on the population stochastic process, the first p eigenfunctions $\{\eta_1(t),...,\eta_p(t)\}$ can be modeled as,

$$\eta_l(t)=\sum_{m=1}^{M} d_{ml}\psi_l(t) \tag{4.25}$$

where functions $\{\psi_1(t),...,\psi_M(t)\} \in L^2(\mathcal{T})$ and are known. Note that $\{\psi_1(t),...,\psi_M(t)\}$ are usually orthonormalized. Based on the orthonormality of eigenfunctions $\{\eta_1(t),...,\eta_p(t)\}$, we

have the constraint for matrix $\mathbf{D}$, $\mathbf{D}^T\mathbf{D} = \mathbf{I}_p$. Note that under Gaussian Process and normal error assumption, conditional on time points, the negative log-likelihood of the data is given by

$$
\begin{aligned}
-\log L(\mathbf{D}, diag(\boldsymbol{\omega}, \sigma^2)) \propto & \frac{1}{2}\sum_{i=1}^{n} Tr[(\sigma^2\mathbf{I}_{N_i} + \boldsymbol{\Psi}_i^T\mathbf{D}diag(\boldsymbol{\omega})\mathbf{D}^T\boldsymbol{\Psi}_i)(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_{pool;i})(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_{pool;i})^T] \\
& + \frac{1}{2}\log|\sigma^2\mathbf{I}_{N_i} + \boldsymbol{\Psi}_i^T\mathbf{D}diag(\boldsymbol{\omega})\mathbf{D}^T\boldsymbol{\Psi}_i|
\end{aligned}
\tag{4.26}
$$

where $\boldsymbol{\Psi}_i = (\boldsymbol{\psi}_{i1}, ..., \boldsymbol{\psi}_{iM})^T$, with $\boldsymbol{\psi}_{il} = (\psi_l(t_{i1}), ..., \psi_l(t_{iN_i}))^T$ for $l = 1, .., M$. Estimate for the pooled mean function $\hat{\mu}_{pool}(t)$ is given in (4.24) and $\hat{\boldsymbol{\mu}}_{pool;i} = (\hat{\mu}_{pool}(t_{i1}), ..., \hat{\mu}_{pool}(t_{iN_i}))^T$. A Newton-Raphson algorithm is utilized to achieve $\hat{\mathbf{D}}$, $\hat{\boldsymbol{\omega}}$ and $\hat{\sigma}^2$ such that they minimize the negative log-likelihood, subject to the constraint that $\mathbf{D}^T\mathbf{D} = \mathbf{I}_p$. Then we can estimate the corresponding eigenfunctions by $\hat{\eta}_l(t) = \sum_{m=1}^{M}\hat{d}_{ml}\psi_l(t)$.

Combing all the estimation above, we achieve estimation for the covariance $Cov[\tilde{\boldsymbol{\xi}}_i]$

$$
\begin{aligned}
\hat{Cov}[\tilde{\boldsymbol{\xi}}_i] &= \hat{E}_t[diag(\boldsymbol{\omega})\boldsymbol{E}_i^T\boldsymbol{\Sigma}_{\mathbf{Y}_i}^{-1}\boldsymbol{E}_i diag(\boldsymbol{\omega})] \\
&= \frac{1}{n}\sum_{i=1}^{n} diag(\hat{\boldsymbol{\omega}})\hat{\boldsymbol{E}}_i^T[\hat{\boldsymbol{E}}_i diag(\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{E}}_i^T + \hat{\sigma}^2\mathbf{I}]^{-1}\hat{\boldsymbol{E}}_i diag(\hat{\boldsymbol{\omega}}) \\
&\stackrel{\text{def}}{=} \hat{\boldsymbol{V}}_3
\end{aligned}
\tag{4.27}
$$

And the empirical estimate for $\tilde{\boldsymbol{\xi}}_i$ is $\hat{\tilde{\boldsymbol{\xi}}}_i = (\hat{\tilde{\xi}}_{i1}, ..., \hat{\tilde{\xi}}_{ip})^T$, where $\hat{\tilde{\xi}}_{il} = \hat{\omega}_l\hat{\boldsymbol{\eta}}_{il}^T\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}_i}^{-1}(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_{pool;i})$.

By plugging all the estimates into equation (4.22) and equation (4.23), we have the following two test statistics,

$$
\begin{aligned}
C_{p,n}^{(1)} &= \max_{1\le k<n}\frac{1}{n}[\sum_{i=1}^{k}\hat{\tilde{\boldsymbol{\xi}}}_i - \frac{k}{n}\sum_{i=1}^{n}\hat{\tilde{\boldsymbol{\xi}}}_i]^T\hat{\boldsymbol{V}}_3^{-1}[\sum_{i=1}^{k}\hat{\tilde{\boldsymbol{\xi}}}_i - \frac{k}{n}\hat{\tilde{\boldsymbol{\xi}}}_i] \\
C_{p,n}^{(2)} &= \int_0^1 \frac{1}{n}[\sum_{i=1}^{[nz]}\hat{\tilde{\boldsymbol{\xi}}}_i - z\sum_{i=1}^{n}\hat{\tilde{\boldsymbol{\xi}}}_i]^T\hat{\boldsymbol{V}}_3^{-1}[\sum_{i=1}^{[nz]}\hat{\tilde{\boldsymbol{\xi}}}_i - z\sum_{i=1}^{n}\hat{\tilde{\boldsymbol{\xi}}}_i]\,dz
\end{aligned}
\tag{4.28}
$$

4.3.3 Asymptotic results

In this section, we develop the asymptotic theory of $C_{p,n}^{(1)}$ and $C_{p,n}^{(2)}$ under both the H_0 and the H_a. Given that we use the rMLE method and local linear smoothing, we need the following assumptions.

1. $X_1(t)$,....,$X_n(t)$ are independent Gaussian processes.
2. The p largest eigenvalues of $\tilde{G}(s,t)$ satisfy. (i) There exists a constant $a_1 < \infty$, such that $a_1 \geq \omega_1 > \omega_2 > ... > \omega_p > \omega_{p+1}$; (ii) There exists a constant $a_2 < \infty$, such that $\max_{1\leq l\leq p}(\omega_l - \omega_{l+1})^{-1} \leq a_2$.
3. The common eigenfunctions $\{\eta_k\}_{k=1}^p$ are four times continuously differentiable and satisfy for some $0 < A_0 < \infty$
$$\max_{1\leq l\leq p} \sup_{t\in\mathcal{T}} |\eta_l^{(4)}(t)| \leq A_0 \tag{4.29}$$
4. For each i, $\{T_{ij}, j = 1, ..., N_i\}$ are i.i.d samples from a distribution g within the time domain $\mathcal{T}$, where g is a bounded function and it satisfies $c_{g,0} \leq g(x) \leq c_{g,1}$ for all $t \in \mathcal{T}$, where $0 < c_{g,0} \leq c_{g,1} < \infty$.
5. The number of measurements N_i satisfies $\underline{N} \leq N_i \leq \bar{N}$ with $\underline{N} \geq 4$ and $\bar{N} < \infty$.
6. The following two assumptions are correct: $M^{-1}(n/\log n)^{1/9} = O(1)$, and $M = o(\sqrt{n/\log n})$, where M as defined in Section 4.3.2 is the number of orthonormalized cubic B-spline basis used to represent the eigenfunctions $\eta_l(t)$.

   Similar conditions are needed and discussed in Chapter 2.

Theorem 5. *Under H_0 and regularization assumptions 1-6, if $E_t[\boldsymbol{E}_{il}^T \boldsymbol{\Sigma}_{Y_i}^{-1} \boldsymbol{E}_{il}] < \infty$ is satisfied. For any fixed p, we have*

$$\begin{aligned} C_{p,n}^{(1)} &\xrightarrow{D} \sup_{0\leq z\leq 1} \sum_{l=1}^{p} B_l^2(z) \\ C_{p,n}^{(2)} &\xrightarrow{D} \sum_{l=1}^{p} \int_0^1 B_l^2(z)\, dz \end{aligned} \tag{4.30}$$

where $B_l(t)$, $l = 1, ..., p$, are independent standard Brownian Bridges.

The above asymptotic results are the basis of making decisions in the testing problem (4.8). The distribution of $\sum_{l=1}^{p} \int_0^1 B_l^2(z)\, dz$ is derived by (Kiefer, 1959). However, they only provide results for a small number of p, i.e., $p < 5$. In practice, people can always get the $1-\alpha$ percentiles of $\sup_{0\le z\le 1} \sum_{l=1}^{p} B_l^2(z)$ or $\sum_{l=1}^{p} \int_0^1 B_l^2(z)\, dz$ by repeatedly simulating p independent standard Brownian Bridges. And the decision rule is to reject H_0 if $C_{p,n}^{(1)}$ or $C_{p,n}^{(2)}$ is greater than the corresponding $1-\alpha$ percentile. For $C_{p,n}^{(2)}$, its $1-\alpha$ quantile calculated from 1000 points over 100,000 replications are given by (Aue et al., 2009). We copy their results for $p = 1, ..., 15$ in Table 4.1. Following the same schema, we simulated a quantile table for $C_{p,n}^{(1)}$, as shown in Table 4.2. Proof of Theorem 5 is postponed to Appendix.

Table 4.1: Simulated critical values ($\alpha = 0.01, 0.05, 0.10$) of the distribution of $\sum_{l=1}^{p} \int_0^1 B_l^2(z)\, dz$. The value of $p = 1, .., 15$.

α	$\mathbf{p=1}$	$\mathbf{p=2}$	$\mathbf{p=3}$	$\mathbf{p=4}$	$\mathbf{p=5}$
0.10	0.345165	0.606783	0.842567	1.065349	1.279713
0.05	0.460496	0.748785	1.001390	1.239675	1.469008
0.01	0.740138	1.072101	1.352099	1.626695	1.866702
α	$\mathbf{p=6}$	$\mathbf{p=7}$	$\mathbf{p=8}$	$\mathbf{p=9}$	$\mathbf{p=10}$
0.10	1.485200	1.690773	1.897365	2.096615	2.288572
0.05	1.684729	1.895557	2.124153	2.322674	2.526781
0.01	2.125950	2.342252	2.589244	2.809778	3.033944
α	$\mathbf{p=11}$	$\mathbf{p=12}$	$\mathbf{p=13}$	$\mathbf{p=14}$	$\mathbf{p=15}$
0.10	2.496635	2.686238	2.884214	3.066906	3.268958
0.05	2.744438	2.949004	3.147604	3.336262	3.544633
0.01	3.268031	3.491102	3.708033	3.903995	4.116829

In the following, Theorem 6 provides theoretical justifications for the power of our test procedures under certain type of alternatives.

Theorem 6. *Under regularization assumptions 1-6, suppose that we have* $E_t(diag(\boldsymbol{\omega})\boldsymbol{E}_i^T\boldsymbol{\Sigma}_{Y_i}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{pool;i})) \neq \mathbf{0}$*, then*

$$\begin{aligned} C_{p,n}^{(1)} &\xrightarrow{p} \infty \\ C_{p,n}^{(2)} &\xrightarrow{p} \infty \end{aligned} \tag{4.31}$$

Under random design t, $E_t(diag(\boldsymbol{\omega})\boldsymbol{E}_i^T\boldsymbol{\Sigma}_{Y_i}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{pool;i}))$ is basically some type of projection of $\mu(t) - \mu_{pool}(t)$. When $\mu(t) - \mu_{pool}(t) \neq 0$, it is unlikely that this projection scores equal to 0

Table 4.2: Simulated critical values ($\alpha = 0.01, 0.05, 0.10$) of the distribution of $\sup \sum_{l=1}^{p} B_l^2(z)$. The value of $p = 1, .., 10$.

α	$\mathbf{p = 1}$	$\mathbf{p = 2}$	$\mathbf{p = 3}$	$\mathbf{p = 4}$	$\mathbf{p = 5}$
0.10	1.451416	2.067668	2.563871	3.022649	3.45234
0.05	1.796405	2.458266	2.987638	3.479112	3.942844
0.01	2.605361	3.318181	3.91891	4.467059	4.98733
α	$\mathbf{p = 6}$	$\mathbf{p = 7}$	$\mathbf{p = 8}$	$\mathbf{p = 9}$	$\mathbf{p = 10}$
0.10	3.840457	4.241502	4.634173	5.002395	5.371901
0.05	4.348183	4.770868	5.183215	5.579701	5.954123
0.01	5.413516	5.897421	6.343253	6.792156	7.183547

on all the first p directions. Theorem 6 ensures that, under any alternative H_a when the difference between $\mu_g(t)$ and $\mu_{pool}(t)$ is captured by some of the first p directions of such projection, the power of our test procedure goes to 1. Proof of Theorem 6 is also postponed to Appendix.

4.3.4 Estimation of change point

When a significant shift in mean function is detected, we propose to estimate its location by

$$\begin{aligned} \hat{k^*}^{(1)} &= \inf\{k : S^{(1)}_{p,n,k} = \max_{1 \le s < n-1} S^{(1)}_{p,n,s}\} \\ \hat{k^*}^{(2)} &= \inf\{z : S^{(2)}_{p,n,z} = \sup_{0 \le y \le 1} S^{(2)}_{p,n,y}\} \end{aligned} \tag{4.32}$$

which is the first location that yields the largest difference in the mean function estimates of the former sub-sequence and the latter sub-sequence. The performances of the both estimators are evaluated by simulation studies in Section 4.4.1.

4.4 NUMERICAL EXPERIMENTS

4.4.1 Simulation studies

In this subsection, we examine the finite sample performance of the proposed test procedures $C^{(1)}_{p,n}$ (4.22) and $C^{(2)}_{p,n}$ (4.23) through two simulations studies, both of which only have one change

point. In the first one, we eliminate the effect of choosing p by using its true value. In the second study, we consider a more complex situation, where the true value of p is unknown. In both studies, we examine the performances of the proposed test procedures and the estimates for the change point location defined in (4.32).

Simulation I

The data is generated based on model (4.7), and the Karhunen-Loève expansion $X_i(T_{ij}) = \mu_i(T_{ij}) + \sum_{l=1}^{\infty} \xi_{il}\phi_l(T_{ij})$, with $\xi_{il} \sim N(0, \lambda_l)$. The number of observations on each curve is 4, i.e., $N_i = N = 4$ for $i = 1, ..., n$. The observing time points, $\{T_{i1}, ..., T_{iN}\}$, are uniformly distributed within the interval $[0, 1]$. We assume that the eigenvalues are $\lambda_l = 2(l+1)^{-2}$ for $l = 1, ..., 4$, and $\lambda_l = 0$ for $l > 4$. The eigenfunctions are $\phi_1(t) = \sqrt{2}\cos(\pi t)$, $\phi_2(t) = \sqrt{2}\sin(\pi t)$, $\phi_3(t) = \sqrt{2}\cos(2\pi t)$, $\phi_4(t) = \sqrt{2}\sin(2\pi t)$. The standard deviation of the measurement errors is $\sigma = 0.6$. The lengths of the random sequence are $n = 50$, $n = 100$, and $n = 200$. For each length situation, there are three different change point considered, $\theta = [\frac{n}{4}]$, $[\frac{n}{2}]$ and $[\frac{3n}{4}]$. The mean functions $\mu_i(t)$ are $\mu_i(t) = \mu(t)$, for $i = 1, ..., \theta$, and $\mu_i(t) = \mu'(t)$, for $i = \theta + 1, ..., n$. And $\mu(t) = \sum_{l=1}^{4} a_{1l}\phi_l(t)$, where $a_{1l} = \frac{3\bar{a}_{1l}}{\|\bar{a}_1\|_2}$, with $\bar{a}_1 = (\bar{a}_{11}, ..., \bar{a}_{14})^T$ and $\bar{a}_{1l} = (2)^{5-l}(5-l)^6$. And $\mu'(t) = \sum_{l=1}^{4} a_{2l}\phi_l(t)$, where $a_{2l} = \frac{3\bar{a}_{2l}}{\|\bar{a}_2\|_2}$, with $\bar{a}_2 = (\bar{a}_{21}, ..., \bar{a}_{24})^T$ and $\bar{a}_{2l} = (2-\varsigma)^{5-l}(5-l)^6$. ς quantifies the magnitude of the distinctions among the two mean functions, and $\varsigma = 0, 1.2, 1.35, 1.5$ correspond to the null hypothesis and three alternative hypotheses for (4.8).

In the calculation of all the results, we use the true $p = 4$, such that the results are free of the effect of choosing p. Given the results shown in Table 4.3, the performances of both $C_{p,n}^{(1)}$ and $C_{p,n}^{(2)}$ are quite reasonable. Under the null in (4.8) ($\varsigma = 0$), the tests are capable of controlling the pre-specified significance level $\alpha = 0.05$, even though they tends to be relatively biased when sample size is small. For a given change point situation θ and a fixed sample size n, the powers of the proposed tests increases with the increasing the distinction between $\mu(t)$ and $\mu'(t)$, i.e., ς. For a specific combination of sample size n and alternative ς, the powers are larger if the change happens in the middle ($\theta = [\frac{n}{2}]$), while the powers at the first quarter and the third quarter are very close. According to Figure 4.1, which visualizes $\hat{\theta}$ from (4.32) under different n and θ, the point estimate for θ is pretty close to its true value.

Simulation II

Now we adopt a more complex setting. The number of observations $N_i \sim U[2, ..., 6]$, and

Table 4.3: Results for simulation I. We present results for both $C_{p,n}^{(1)}$ and $C_{p,n}^{(2)}$. The length of the time series is $n = 50, 100, 200$, and the change point is $\theta = [\frac{n}{4}]$, $[\frac{n}{2}]$ and $[\frac{3n}{4}]$. The number of observation on each random curve is 4 ($N = 4$) and the standard deviation of the random measurement error is 0.6 ($\sigma = 0.6$). The type I error ($\varsigma = 0$) and powers ($\varsigma = 1.2, 1.35, 1.5$) are calculated based on 1000 repetitions. The significance level α is 0.05.

Test statistic	n	**Location**	$\varsigma = 0$	$\varsigma = 1.2$	$\varsigma = 1.35$	$\varsigma = 1.5$
$C_{p,n}^{(1)}$	$n = 50$	$\theta = 13$	0.009	0.045	0.109	0.361
		$\theta = 25$	0.006	0.120	0.366	0.767
		$\theta = 38$	0.045	0.083	0.108	0.261
	$n = 100$	$\theta = 25$	0.014	0.211	0.484	0.934
		$\theta = 50$	0.013	0.496	0.877	1.000
		$\theta = 75$	0.011	0.195	0.469	0.927
	$n = 200$	$\theta = 50$	0.030	0.621	0.942	0.999
		$\theta = 100$	0.040	0.934	0.998	1.000
		$\theta = 150$	0.023	0.581	0.953	1.000
$C_{p,n}^{(2)}$	$n = 50$	$\theta = 13$	0.016	0.078	0.179	0.392
		$\theta = 25$	0.009	0.148	0.342	0.742
		$\theta = 38$	0.017	0.070	0.127	0.321
	$n = 100$	$\theta = 25$	0.020	0.237	0.504	0.882
		$\theta = 50$	0.016	0.260	0.508	0.992
		$\theta = 75$	0.022	0.255	0.490	0.873
	$n = 200$	$\theta = 50$	0.051	0.603	0.928	0.999
		$\theta = 100$	0.033	0.883	0.997	1.000
		$\theta = 150$	0.037	0.581	0.930	0.999

then conditional on the value of N_i, $(T_{i1}, ..., T_{iN_i})$ are i.i.d uniformly distributed with [0,1]. The lengths of the random sequence are $n = 50$, $n = 100$, and $n = 200$. For each length situation, there are three different change point considered, $[\frac{n}{2}]$ and $[\frac{3n}{4}]$. The mean functions $\mu_i(t)$ are $\mu_i(t) = \mu(t)$, for $i = 1, ..., \theta$, and $\mu_i(t) = \mu'(t)$, for $i = \theta+1, ..., n$. $\mu(t) = 20t^{\varrho}(1-t)$, and $\mu'(t) = 20t(1-t)^{\varrho}$, where $\varrho = 1, 1.12, 1.24, 1.36$ correspond to the null and three alternative hypotheses. Eigenfunctions are $\phi_{2l-1}(t) = \sqrt{2}\cos((2l-1)\pi t)$ for $l = 1, ..., 15$, $\phi_{2l}(t) = \sqrt{2}\sin(2l\pi t)$ for $l = 1, ..., 15$. The standard deviation of random errors is $\sigma = 0.8$.

Different from simulation I, the true value of p is unknown. We use the leave-one-out cross-validation method (CV) in (Peng and Paul, 2009) and the fraction of variance explained (FVE80,

FVE85) to estimate p. Then use this $\hat{p}$ to conduct the proposed test procedures. Given the numbers in Table 4.4, the cross-validation method is more reliable all the time. The FVE methods fails to appropriately control the type I errors when the length of the independent random sequence is not long enough. In Figure 4.2, the estimated $\hat{\theta}$ based on $\hat{p}$ selected from the CV method are visualized. It can be seen that the point estimate for θ is pretty close to its true value.

4.4.2 Application to an eBay online auction data set

It is all known that iPhone 7 and iPhone 7 Plus was announced on September 7th, 2016 and was released on September 16th, 2016. We are interested in checking whether the release of the new versions has any effects on the sales of the previous versions on eBay markets.

To investigate this question, we scraped the live bidding price trajectories of used 64GB iPhone 6 Plus with AT&T network. The auctions that we consider are all with starting dates between August 18th, 2016 and November 13rd, 2016. Within all the auctions, we consider only the 3-day auctions. Totally we include 233 individual auctions in the functional time series, after deleting two outliers. Similar to the data analysis in Chapter 2, we transform live bid data into log-scale. Now the problem of checking whether the release of new iPhones have any effect on the previous versions becomes testing whether there exists any change points in mean bidding curves between August 18th, 2016 and November 13rd, 2016.

The proposed tests is implemented to this mean function change point detection problem. We use the cross-validation method to choose $\hat{p} = 3$. For test statistic $C^{(2)}_{p,n}$, we have its value 5.14, which is greater than the 95% quantile given in Table 4.1. This means that we have enough confidence to say the mean functions are not consistent over time. The estimated change point is September 9th, 2016. This date is two days after the new iPhone was displayed to the public. The estimated mean functions by local linear smoothing are included in Section 4.4.2. From the graph, we can see that the price of the before group (solid black) is larger than that of the after group (solid red) over the entire time domain. For test statistic $C^{(1)}_{p,n}$, we have its value 7.28, which is also greater than the 95% quantile given in Table 4.2. This means that we do have enough confidence to say the mean functions are not consistent over time.

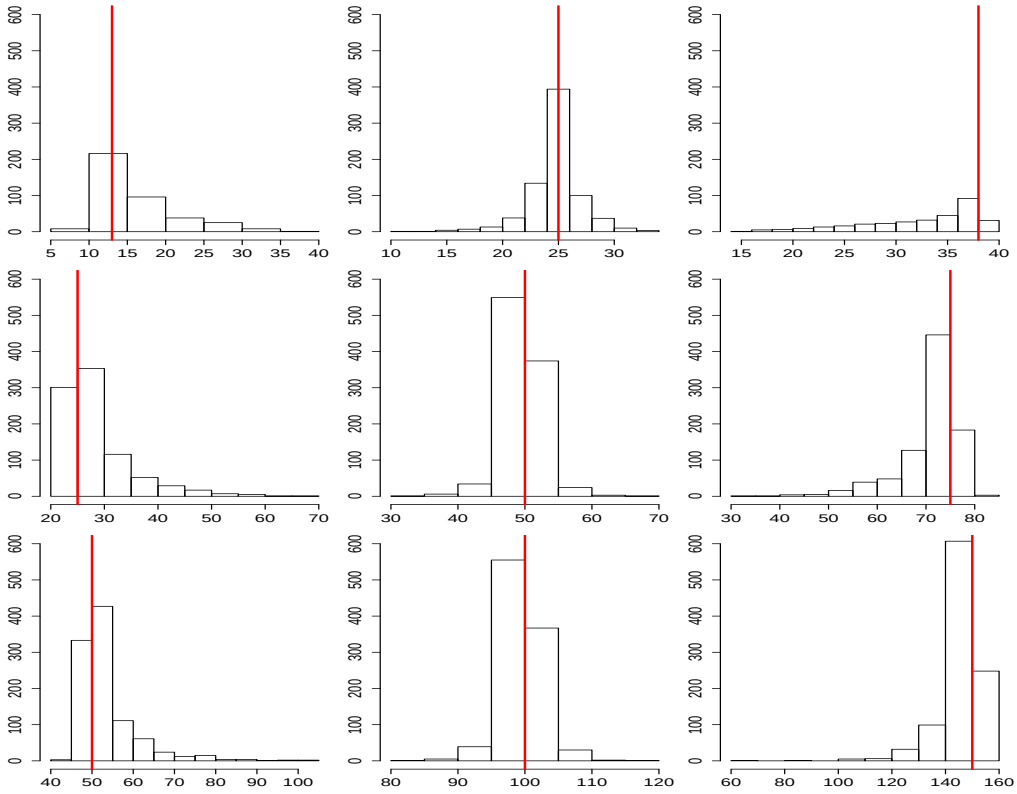

Figure 4.1: On each row of the figure matrix, histograms of estimated change point θ is presented for length $n = 50, 100, 200$. On each column of the plot matrix, histograms is provided for $\theta = [\frac{n}{4}]$, $[\frac{n}{2}]$ and $[\frac{3n}{4}]$. For all the situations, the majority of the estimates are very close to the true value θ. For example, let's take a look at the histogram right in the middle ($n = 100, \theta = 50$), majority of $\hat{\theta}$ fall within the range from 45 to 55.

Table 4.4: Results for simulation II. We present results for both $C_{p,n}^{(1)}$ and $C_{p,n}^{(2)}$. The length of the time series is $n = 50, 100, 200$, and the change point location is $[\frac{n}{2}]$ and $[\frac{3n}{4}]$. The number of observation on each random curve is 4 ($N = 4$) and the standard deviation of the random measurement error is 0.8 ($\sigma = 0.8$), different sample sizes ($n = 50, 100,$ or 200). The type I error ($\varrho = 0$) and powers ($\varsigma = 1.12, 1.24, 1.36$) are calculated based on 1000 repetitions.

Test statistic	n	**Location**	**Test**	$\boldsymbol{\varrho = 0}$	$\boldsymbol{\varrho = 1.12}$	$\boldsymbol{\varrho = 1.24}$	$\boldsymbol{\varrho = 1.36}$
$C_{p,n}^{(1)}$	$n = 50$	$\theta = 25$	CV	0.014	**0.156**	**0.658**	**0.943**
			FVE80	*0.296*	0.504	0.752	0.867
			FVE85	*0.292*	0.507	0.756	0.869
		$\theta = 38$	CV	0.020	**0.058**	**0.256**	**0.546**
			FVE80	*0.278*	0.431	0.637	0.799
			FVE85	*0.283*	0.436	0.642	0.802
	$n = 100$	$\theta = 50$	CV	0.021	**0.489**	**0.978**	**0.999**
			FVE80	*0.108*	0.534	0.843	0.955
			FVE85	*0.109*	0.537	0.849	0.958
		$\theta = 75$	CV	0.023	**0.212**	**0.813**	**0.985**
			FVE80	*0.119*	0.357	0.743	0.891
			FVE85	*0.120*	0.358	0.745	0.893
	$n = 200$	$\theta = 100$	CV	0.024	**0.874**	**1.000**	**1.000**
			FVE80	0.060	0.756	0.946	0.992
			FVE85	0.058	0.762	0.946	0.998
		$\theta = 150$	CV	0.030	**0.609**	**0.995**	**1.000**
			FVE80	*0.070*	0.528	0.893	0.981
			FVE85	*0.070*	0.528	0.899	0.983
$C_{p,n}^{(2)}$	$n = 50$	$\theta = 25$	CV	0.032	**0.217**	**0.690**	**0.924**
			FVE80	*0.274*	0.476	0.741	0.863
			FVE85	*0.279*	0.484	0.746	0.862
		$\theta = 38$	CV	0.028	**0.092**	**0.324**	**0.608**
			FVE80	*0.242*	0.398	0.638	0.752
			FVE85	*0.250*	0.406	0.640	0.758
	$n = 100$	$\theta = 50$	CV	0.035	**0.517**	**0.969**	**0.999**
			FVE80	*0.092*	0.517	0.852	0.945
			FVE85	*0.098*	0.518	0.857	0.947
		$\theta = 75$	CV	0.047	**0.359**	**0.638**	**1.000**
			FVE80	0.060	0.537	0.752	0.968
			FVE85	0.061	0.543	0.756	0.969
	$n = 200$	$\theta = 100$	CV	0.040	**0.873**	**0.999**	**1.000**
			FVE80	0.056	0.713	0.940	0.996
			FVE85	0.056	0.724	0.941	0.997
		$\theta = 150$	CV	0.038	**0.515**	**0.970**	**1.000**
			FVE80	*0.118*	0.515	0.857	0.950
			FVE85	*0.122*	0.520	0.865	0.953

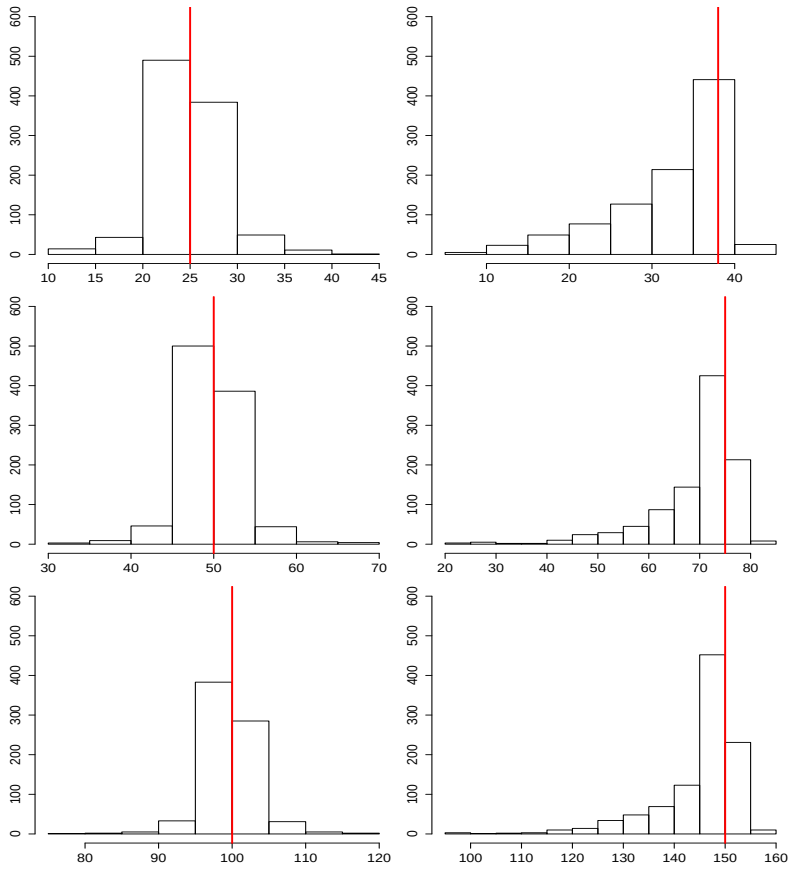

Figure 4.2: On each row of the figure matrix, histograms of estimated change point θ is presented for length $n = 50, 100, 200$. On each column of the plot matrix, histograms is provided for $[\frac{n}{2}]$ and $[\frac{3n}{4}]$. For all the situations, the majority of the estimates are very close to the true value θ. For instance, let's take a look at the last histogram on the left column in Figure 4.2 ($n = 200, \theta = 100$), majority of $\hat{\theta}$ fall within the range from 105 to 110.

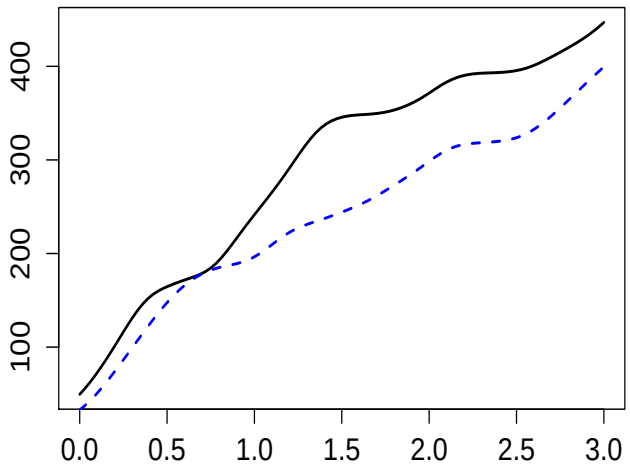

Figure 4.3: Estimated mean function of iPhone 6 plus live biddings before September 9th, 2016 (solid black); estimated mean function of iPhone 6 plus live biddings after September 9th, 2016 (dash blue)

5.0 CONCLUDING REMARKS

In this thesis, we studied several related inference problems for sparsely observed functional data, including two-sample test, one-way functional ANOVA problem and change point detection in functional time series. We reviewed and discussed the existing methods for dense functional data, which at the same time explained and motivated the need of methodological development for sparse functional data. The proposed test statistics are based on the construction of a shrinkage score and a careful derivation of the asymptotic distribution. Numerical experiments demonstrated the good performance of the proposed testing procedures. All the methods were illustrated by applications to the CD4 count data and the eBay online auction data.

This thesis focused on the inference problem for mean functions, assuming the covariance functions are the same. We discussed the extensions to some relaxed assumptions on covariance functions such as common covariance structures. The inference and formal checking of covariance structures are interesting and challenging problems on their own, which we think would be nice topics for further research.

APPENDIX

PROOF OF THEOREMS

Proof of Theorem 1

The goal is to show that $T_{p,N} \rightarrow \chi^2(p)$ in distribution, under H_0. Denote $T^*_{p,N} = (\bar{\tilde{\boldsymbol{r}}}^c_{1\cdot} - \bar{\tilde{\boldsymbol{r}}}^c_{2\cdot})^T[(\frac{1}{n_1}+\frac{1}{n_2})\boldsymbol{V}]^{-1}(\bar{\tilde{\boldsymbol{r}}}^c_{1\cdot} - \bar{\tilde{\boldsymbol{r}}}^c_{2\cdot})$, where $\boldsymbol{V} = \frac{\sum_{g=1}^{2}\sum_{i=1}^{n_g} diag(\boldsymbol{\lambda})\boldsymbol{\Phi}^T_{gi}\boldsymbol{\Sigma}^{-1}_{\mathbf{Y}^c_{gi}}\boldsymbol{\Phi}_{gi}diag(\boldsymbol{\lambda})}{n_1+n_2}$. According to the Central Limit Theorem and the Slutsky's theorem, $T^*_{p,N}$ is asymptotically $\chi^2(p)$ distributed under H_0. Consequently, proving Theorem 1 is basically proving $T_{p,N} = T^*_{p,N} + o_p(1)$.

Let's denote $\mathbf{S}_{gi} = diag(\boldsymbol{\lambda})\boldsymbol{\Phi}^T_{gi}\boldsymbol{\Sigma}^{-1}_{\mathbf{Y}^c_{gi}}\boldsymbol{\Phi}_{gi}diag(\boldsymbol{\lambda})$ and $\hat{\mathbf{S}}_{gi} = diag(\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\Phi}}^T_{gi}[\hat{\boldsymbol{\Phi}}_{gi}diag(\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\Phi}}^T_{gi} + \hat{\sigma}^2_g\mathbf{I}]^{-1}\hat{\boldsymbol{\Phi}}_{gi}diag(\hat{\boldsymbol{\lambda}})$, then $\hat{\boldsymbol{V}} = \frac{\hat{\mathbf{S}}_{11}+\ldots+\hat{\mathbf{S}}_{1n_1}+\hat{\mathbf{S}}_{21}+\ldots+\hat{\mathbf{S}}_{2n_2}}{n}$ and $\boldsymbol{V} = \frac{\mathbf{S}_{11}+\ldots+\mathbf{S}_{1n_1}+\mathbf{S}_{21}+\ldots+\mathbf{S}_{2n_2}}{n}$. The absolute value of $T_{p,N} - T^*_{p,N}$ can be decomposed as follows.

$$
\begin{aligned}
&|T_{p,N} - T^*_{p,N}| \\
=&\left\|(\bar{\hat{\tilde{\boldsymbol{r}}}}^c_{1\cdot} - \bar{\hat{\tilde{\boldsymbol{r}}}}^c_{2\cdot})^T[(\frac{1}{n_1}+\frac{1}{n_2})\hat{\boldsymbol{V}}]^{-1}(\bar{\hat{\tilde{\boldsymbol{r}}}}^c_{1\cdot} - \bar{\hat{\tilde{\boldsymbol{r}}}}^c_{2\cdot}) - (\bar{\tilde{\boldsymbol{r}}}^c_{1\cdot} - \bar{\tilde{\boldsymbol{r}}}^c_{2\cdot})^T[(\frac{1}{n_1}+\frac{1}{n_2})\boldsymbol{V}]^{-1}(\bar{\tilde{\boldsymbol{r}}}^c_{1\cdot} - \bar{\tilde{\boldsymbol{r}}}^c_{2\cdot})\right\|_2 \\
\leq&\left\|(\bar{\tilde{\boldsymbol{r}}}^c_{1\cdot} - \bar{\tilde{\boldsymbol{r}}}^c_{2\cdot})^T[[(\frac{1}{n_1}+\frac{1}{n_2})\hat{\boldsymbol{V}}]^{-1} - [(\frac{1}{n_1}+\frac{1}{n_2})\boldsymbol{V}]^{-1}](\bar{\tilde{\boldsymbol{r}}}^c_{1\cdot} - \bar{\tilde{\boldsymbol{r}}}^c_{2\cdot})\right\|_2 \\
+&\left\|(\bar{\hat{\tilde{\boldsymbol{r}}}}^c_{1\cdot} - \bar{\hat{\tilde{\boldsymbol{r}}}}^c_{2\cdot} - \bar{\tilde{\boldsymbol{r}}}^c_{1\cdot} + \bar{\tilde{\boldsymbol{r}}}^c_{2\cdot})^T[(\frac{1}{n_1}+\frac{1}{n_2})\hat{\boldsymbol{V}}]^{-1}(\bar{\tilde{\boldsymbol{r}}}^c_{1\cdot} - \bar{\tilde{\boldsymbol{r}}}^c_{2\cdot})\right\|_2 \\
+&\left\|(\bar{\hat{\tilde{\boldsymbol{r}}}}^c_{1\cdot} - \bar{\hat{\tilde{\boldsymbol{r}}}}^c_{2\cdot})^T[(\frac{1}{n_1}+\frac{1}{n_2})\hat{\boldsymbol{V}}]^{-1}(\bar{\hat{\tilde{\boldsymbol{r}}}}^c_{1\cdot} - \bar{\hat{\tilde{\boldsymbol{r}}}}^c_{2\cdot} - \bar{\tilde{\boldsymbol{r}}}^c_{1\cdot} + \bar{\tilde{\boldsymbol{r}}}^c_{2\cdot})\right\|_2 \\
\stackrel{\text{def}}{=}& A_1 + A_2 + A_3
\end{aligned}
\tag{.1}
$$

Given (.1), proving Theorem 1 can be accomplished by proving $A_1 = o_p(1)$, $A_2 = o_p(1)$, and $A_3 = o_p(1)$ hold.

First, let's start to show $A_1 = o_p(1)$. Under H_0, i.e., $\mu_1(t) = \mu_2(t) = \mu_{pool}(t)$, we have

$$(\frac{1}{n_1} + \frac{1}{n_2})^{-1/2}(\bar{\tilde{\boldsymbol{r}}}^c_{1\cdot} - \bar{\tilde{\boldsymbol{r}}}^c_{2\cdot}) \sim N_p(\mathbf{0}, Cov(\tilde{\boldsymbol{r}}^c_{gi})) \tag{.2}$$

So we have

$$\left\|(\frac{1}{n_1} + \frac{1}{n_2})^{-1/2}(\bar{\tilde{\boldsymbol{r}}}^c_{1\cdot} - \bar{\tilde{\boldsymbol{r}}}^c_{2\cdot})\right\|_2 = O_p(1) \tag{.3}$$

According to (Peng and Paul, 2009), under condition 1 to 6 in Section 2.7 and several other mild conditions for local linear smoothing (Zhang and Wang, 2016), we have

$$\begin{aligned}
&\left\|\hat{\mu}_{pool} - \mu_{pool}\right\|_F = o_p(1)\\
&\left\|diag(\hat{\boldsymbol{\lambda}}) - diag(\boldsymbol{\lambda})\right\|_2 = o_p(1)\\
&\left\|\hat{\phi}_k - \phi_k\right\|_F = o_p(1)\\
&\left\|\hat{\sigma}^2 - \sigma^2\right\|_2 = o_p(1)
\end{aligned} \tag{.4}$$

we have

$$\begin{aligned}
\left\|\hat{\mathbf{V}} - \mathbf{V}\right\|_2 &\leq \max_{1\leq g\leq 2; 1\leq i\leq n_g}\left\|diag(\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\Phi}}^T_{gi}\hat{\boldsymbol{\Sigma}}^{-1}_{gi}\hat{\boldsymbol{\Phi}}_{gi}diag(\hat{\boldsymbol{\lambda}}) - diag(\boldsymbol{\lambda})\boldsymbol{\Phi}^T_{gi}\boldsymbol{\Sigma}^{-1}_{gi}\boldsymbol{\Phi}_{gi}diag(\boldsymbol{\lambda})\right\|_2\\
&\overset{\text{def}}{=} \max_{1\leq g\leq 2; 1\leq i\leq n_g}\left\|diag(\hat{\boldsymbol{\lambda}})\hat{\mathbf{M}}_{gi}diag(\hat{\boldsymbol{\lambda}}) - diag(\boldsymbol{\lambda})\mathbf{M}_{gi}diag(\boldsymbol{\lambda})\right\|_2\\
&\leq \max_{1\leq g\leq 2; 1\leq i\leq n_g}\left\|(diag(\hat{\boldsymbol{\lambda}}) - diag(\boldsymbol{\lambda}))\mathbf{M}_{gi}diag(\boldsymbol{\lambda})\right\|_2\\
&+ \max_{1\leq g\leq 2; 1\leq i\leq n_g}\left\|diag(\hat{\boldsymbol{\lambda}})\mathbf{M}_{gi}(diag(\hat{\boldsymbol{\lambda}}) - diag(\boldsymbol{\lambda}))\right\|_2\\
&+ \max_{1\leq g\leq 2; 1\leq i\leq n_g}\left\|diag(\hat{\boldsymbol{\lambda}})(\hat{\mathbf{M}}_{gi} - \mathbf{M}_{gi})diag(\hat{\boldsymbol{\lambda}})\right\|_2\\
&\leq o_p(1) + o_p(1) + o_p(1) = o_p(1)
\end{aligned} \tag{.5}$$

The last inequality requires more arguments. We need to show for any g and i, we have

$$\begin{aligned}
\left\|\hat{\mathbf{M}}_{gi} - \mathbf{M}_{gi}\right\|_2 \leq &\left\|(\hat{\boldsymbol{\Phi}}_{gi} - \boldsymbol{\Phi}_{gi})^T\boldsymbol{\Sigma}^{-1}_{gi}\boldsymbol{\Phi}_{gi}\right\|_2 + \left\|\hat{\boldsymbol{\Phi}}_{gi}\boldsymbol{\Sigma}^{-1}_{gi}(\hat{\boldsymbol{\Phi}}_{gi} - \boldsymbol{\Phi}_{gi})\right\|_2\\
&+\left\|\hat{\boldsymbol{\Phi}}_{gi}(\hat{\boldsymbol{\Sigma}}^{-1}_{gi} - \boldsymbol{\Sigma}^{-1}_{gi})\hat{\boldsymbol{\Phi}}_{gi}\right\|_2\\
&\leq o_p(1) + o_p(1) + \frac{\left\|\boldsymbol{\Sigma}^{-1}_{gi}\right\|_2}{[\left\|\boldsymbol{\Sigma}^{-1}_{gi}\right\|_2\left\|\hat{\boldsymbol{\Sigma}}_{gi} - \boldsymbol{\Sigma}_{gi}\right\|_2]^{-1} - 1}\\
&\leq o_p(1) + o_p(1) + o_p(1) = o_p(1)
\end{aligned} \tag{.6}$$

Combine (.3) and (.5) together, we have

$$\begin{aligned} A_1 \leq & \left\| (\frac{1}{n_1}+\frac{1}{n_2})^{-1/2}(\bar{\tilde{\boldsymbol{r}}}^c_{1\cdot}-\bar{\tilde{\boldsymbol{r}}}^c_{2\cdot}) \right\|_2^2 \left\| \hat{\boldsymbol{V}}^{-1}-\boldsymbol{V}^{-1} \right\|_2 \\ \leq & O_p(1) \times \frac{\left\| \boldsymbol{V}^{-1} \right\|_2}{[\left\| \boldsymbol{V}^{-1} \right\|_2 \left\| \hat{\boldsymbol{V}}-\boldsymbol{V} \right\|_2]^{-1}-1} = o_p(1) \end{aligned} \tag{.7}$$

Next, we start to show that $A_2 = o_p(1)$. Note that

$$\begin{aligned} A_2 \leq & \left\| (\frac{1}{n_1}+\frac{1}{n_2})^{-1/2}(\bar{\tilde{\boldsymbol{r}}}^c_{1\cdot}-\bar{\tilde{\boldsymbol{r}}}^c_{2\cdot}) \right\|_2 \left\| \hat{\boldsymbol{V}}^{-1} \right\|_2 \left\| (\frac{1}{n_1}+\frac{1}{n_2})^{-1/2}(\bar{\hat{\tilde{\boldsymbol{r}}}}^c_{1\cdot}-\bar{\hat{\tilde{\boldsymbol{r}}}}^c_{2\cdot}-\bar{\tilde{\boldsymbol{r}}}^c_{1\cdot}+\bar{\tilde{\boldsymbol{r}}}^c_{2\cdot}) \right\|_2 \\ \leq & O_p(1)O_p(1) \left\| (\frac{1}{n_1}+\frac{1}{n_2})^{-1/2}(\bar{\hat{\tilde{\boldsymbol{r}}}}^c_{1\cdot}-\bar{\hat{\tilde{\boldsymbol{r}}}}^c_{2\cdot}-\bar{\tilde{\boldsymbol{r}}}^c_{1\cdot}+\bar{\tilde{\boldsymbol{r}}}^c_{2\cdot}) \right\|_2 \end{aligned} \tag{.8}$$

In order to prove $A_2 = o_p(1)$, we need to show $\left\| (\frac{1}{n_1}+\frac{1}{n_2})^{-1/2}(\bar{\hat{\tilde{\boldsymbol{r}}}}^c_{1\cdot}-\bar{\hat{\tilde{\boldsymbol{r}}}}^c_{2\cdot}-\bar{\tilde{\boldsymbol{r}}}^c_{1\cdot}+\bar{\tilde{\boldsymbol{r}}}^c_{2\cdot}) \right\|_2 = o_p(1)$. We want to write out the explicit formula of $\bar{\hat{\tilde{\boldsymbol{r}}}}^c_{g\cdot}-\bar{\tilde{\boldsymbol{r}}}^c_{g\cdot}$:

$$\begin{aligned} \hat{\tilde{\boldsymbol{r}}}^c_{gi} &= diag(\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\Phi}}^T_{gi}\hat{\boldsymbol{\Sigma}}^{-1}_{\mathbf{Y}_{gi}}(\mathbf{Y}_{gi}-\hat{\boldsymbol{\mu}}_{pool;gi}) \\ &\overset{\text{def}}{=} diag(\hat{\boldsymbol{\lambda}})\hat{\mathbf{Q}}_{gi}(\mathbf{Y}_{gi}-\hat{\boldsymbol{\mu}}_{pool;gi}) \\ &= \tilde{\boldsymbol{r}}^c_{gi} + (diag(\hat{\boldsymbol{\lambda}})-diag(\boldsymbol{\lambda}))\mathbf{Q}_{gik}(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi}) + diag(\hat{\boldsymbol{\lambda}})(\hat{\mathbf{Q}}_{gik}-\mathbf{Q}_{gik})(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi}) \\ &+ diag(\hat{\boldsymbol{\lambda}})\hat{\mathbf{Q}}_{gik}(\boldsymbol{\mu}_{pool;gi}-\hat{\boldsymbol{\mu}}_{pool;gi}) \end{aligned} \tag{.9}$$

i.e.,

$$\begin{aligned} (\frac{1}{n_1}+\frac{1}{n_2})^{-1/2}(\bar{\hat{\tilde{\boldsymbol{r}}}}^c_{g\cdot}-\bar{\tilde{\boldsymbol{r}}}^c_{g\cdot}) &= (\frac{1}{n_1}+\frac{1}{n_2})^{-1/2}(diag(\hat{\boldsymbol{\lambda}})-diag(\boldsymbol{\lambda}))\frac{1}{n_g}\sum_{i=1}^{n_g}\mathbf{Q}_{gi}(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi}) \\ &+ (\frac{1}{n_1}+\frac{1}{n_2})^{-1/2}diag(\hat{\boldsymbol{\lambda}})\frac{1}{n_g}\sum_{i=1}^{n_g}(\hat{\mathbf{Q}}_{gi}-\mathbf{Q}_{gi})(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi}) \\ &+ (\frac{1}{n_1}+\frac{1}{n_2})^{-1/2}diag(\hat{\boldsymbol{\lambda}})\frac{1}{n_g}\sum_{i=1}^{n_g}\hat{\mathbf{Q}}_{gi}(\boldsymbol{\mu}_{pool;gi}-\hat{\boldsymbol{\mu}}_{pool;gi}) \\ &\overset{\text{def}}{=} B_{g1}+B_{g2}+B_{g3} \end{aligned} \tag{.10}$$

According to (.10), in order to ensure that $\left\|(\frac{1}{n_1}+\frac{1}{n_2})^{-1/2}(\bar{\hat{\boldsymbol{r}}}^c_{1\cdot}-\bar{\hat{\boldsymbol{r}}}^c_{2\cdot}-\bar{\boldsymbol{r}}^c_{1\cdot}+\bar{\boldsymbol{r}}^c_{2\cdot})\right\|_2 = o_p(1)$, we need to examine $B_{11}-B_{21}$, $B_{12}-B_{22}$, and $B_{13}-B_{23}$. Let's start with $B_{11}-B_{21}$, for each g,

$$\begin{aligned}\left\|B_{g1}\right\|_2 &\leq (\frac{1}{n_1}+\frac{1}{n_2})^{-1/2}\left\|diag(\hat{\boldsymbol{\lambda}})-diag(\boldsymbol{\lambda})\right\|_2\left\|\frac{1}{n_g}\sum_{i=1}^{n_g}\mathbf{Q}_{gi}(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi})\right\|_2 \\ &= (\frac{1}{n_1}+\frac{1}{n_2})^{-1/2}o_p(1)\sqrt{\sum_{k=1}^{p}(\frac{1}{n_g}\sum_{i=1}^{n_g}\mathbf{Q}_{gik}(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi}))^2} \\ &= (\frac{1}{n_1}+\frac{1}{n_2})^{-1/2}o_p(1)\sqrt{O_P(\frac{1}{n_g})} = o_P(1)\end{aligned} \tag{.11}$$

where Central Limit theorem is used to bound the second norm term. Under H_0, we have $E[\mathbf{Q}_{gik}(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi})|\mathbf{t}]=0$, $Var[\mathbf{Q}_{gik}(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi})|\mathbf{t}]=\mathbf{Q}_{gik}\boldsymbol{\Sigma}_{Y_{gi}}\mathbf{Q}_{gik}^T=\boldsymbol{\Phi}_{gik}^T\boldsymbol{\Sigma}_{Y_{gi}}^{-1}\boldsymbol{\Phi}_{gik}$. As a result, we have

$$\begin{aligned}E[\mathbf{Q}_{gik}(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi})] &= E[E[\mathbf{Q}_{gik}(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi})|\mathbf{t}]]=0 \\ Var[\mathbf{Q}_{gik}(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi})] &= Var[E[\mathbf{Q}_{gik}(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi})|\mathbf{t}]]+E[Var[\mathbf{Q}_{gik}(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi})|\mathbf{t}]] \\ &= 0+E[\boldsymbol{\Phi}_{gik}^T\boldsymbol{\Sigma}_{Y_{gi}}^{-1}\boldsymbol{\Phi}_{gik}]\end{aligned}$$

So as long as we have $E[\boldsymbol{\Phi}_{gik}^T\boldsymbol{\Sigma}_{Y_{gi}}^{-1}\boldsymbol{\Phi}_{gik}]<\infty$, based on the Central Limit Theorem, we have $\frac{1}{n_g}\sum_{i=1}^{n_g}\mathbf{Q}_{gik}(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi})=O_P(\frac{1}{\sqrt{n_g}})$. From (.11), we have

$$\|B_{11}-B_{21}\|_2 \leq \|B_{11}\|_2+\|B_{21}\|_2 = o_p(1) \tag{.12}$$

Now let's consider $B_{12}-B_{22}$. To simplify our arguments, we assume that we split the entire sample into two parts. We use the first part to estimate all the components needed in $T_{p,N}$, and use the second part to conduct the proposed test. Thus, we have independence between $\hat{\mathbf{Q}}_{gi}-\mathbf{Q}_{gi}$ and $\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi}$. Then we have

$$\begin{aligned}\left\|B_{g2}\right\|_2 &\leq (\frac{1}{n_1}+\frac{1}{n_2})^{-1/2}\left\|diag(\hat{\boldsymbol{\lambda}})\right\|_2\left\|\frac{1}{n_g}\sum_{i=1}^{n_g}(\hat{\mathbf{Q}}_{gi}-\mathbf{Q}_{gi})(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi})\right\|_2 \\ &= (\frac{1}{n_1}+\frac{1}{n_2})^{-1/2}O_p(1)o_p(1/\sqrt{n_g}) = o_p(1),\end{aligned} \tag{.13}$$

because $E[(\hat{\mathbf{Q}}_{gi}-\mathbf{Q}_{gi})(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi})]=E[(\hat{\mathbf{Q}}_{gi}-\mathbf{Q}_{gi})]E[(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi})]=\mathbf{0}$, and for each $k=1,...,p$, $Var[(\hat{\mathbf{Q}}_{gik}-\mathbf{Q}_{gik})(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{pool;gi})]=o_p(1)$. Based on (.13), we have

$$\|B_{12}-B_{22}\|_2 \leq \|B_{12}\|_2+\|B_{22}\|_2 = o_p(1) \tag{.14}$$

Note that our simulation study demonstrates good performances of our proposed test statistic, without sample splitting.

For the last term in (.10), $B_{13} - B_{23}$. Note that the randomness of $\hat{\mathbf{Q}}_{gi}(\boldsymbol{\mu}_{pool;gi} - \hat{\boldsymbol{\mu}}_{pool;gi})$ come from random design, which are the same for group 1 and group 2. Consequently, we know that $E_t[\hat{\mathbf{Q}}_{1i}(\boldsymbol{\mu}_{pool;1i} - \hat{\boldsymbol{\mu}}_{pool;1i})] = E_t[\hat{\mathbf{Q}}_{2i}(\boldsymbol{\mu}_{pool;2i} - \hat{\boldsymbol{\mu}}_{pool;g2})]$. Let's denote $E_t[\hat{\mathbf{Q}}_{gi}(\boldsymbol{\mu}_{pool;gi} - \hat{\boldsymbol{\mu}}_{pool;gi})]$ as $\boldsymbol{E}_g$, and we know $\boldsymbol{E}_1 = \boldsymbol{E}_2$, then

$$
\begin{aligned}
&\|B_{31} - B_{32}\|_2 \\
&\leq (\frac{1}{n_1} + \frac{1}{n_2})^{-1/2} \left\| diag(\hat{\boldsymbol{\lambda}}) \right\|_2 \left\| \frac{1}{n_1} \sum_{i=1}^{n_1} \hat{\mathbf{Q}}_{1i}(\boldsymbol{\mu}_{pool;1i} - \hat{\boldsymbol{\mu}}_{pool;1i}) - \frac{1}{n_2} \sum_{i=1}^{n_2} \hat{\mathbf{Q}}_{2i}(\boldsymbol{\mu}_{pool;2i} - \hat{\boldsymbol{\mu}}_{pool;2i}) \right\|_2 \\
&\leq (\frac{1}{n_1} + \frac{1}{n_2})^{-1/2} O_p(1) \left\| \frac{1}{n_1} \sum_{i=1}^{n_1} [\hat{\mathbf{Q}}_{1i}(\boldsymbol{\mu}_{pool;1i} - \hat{\boldsymbol{\mu}}_{pool;1i}) - \boldsymbol{E}_1] - \frac{1}{n_2} \sum_{i=1}^{n_2} [\hat{\mathbf{Q}}_{2i}(\boldsymbol{\mu}_{pool;2i} - \hat{\boldsymbol{\mu}}_{pool;2i}) - \boldsymbol{E}_2] \right\|_2 \\
&= (\frac{1}{n_1} + \frac{1}{n_2})^{-1/2} O_p(1) o_p(1/\sqrt{n_1} + 1/\sqrt{n_2}) = o_p(1)
\end{aligned}
\tag{.15}
$$

Combine (.12), (.14) and (.15), we have

$$
\left\| (\frac{1}{n_1} + \frac{1}{n_2})^{-1/2} (\bar{\hat{\boldsymbol{r}}}^c_{1\cdot} - \bar{\hat{\boldsymbol{r}}}^c_{2\cdot} - \bar{\boldsymbol{r}}^c_{1\cdot} + \bar{\boldsymbol{r}}^c_{2\cdot}) \right\|_2 \leq o_p(1) + o_p(1) + o_p(1) = o_p(1) \tag{.16}
$$

Therefore,

$$
\begin{aligned}
A_2 &\leq O_p(1) O_p(1) \left\| (\frac{1}{n_1} + \frac{1}{n_2})^{-1/2} (\bar{\hat{\boldsymbol{r}}}^c_{1\cdot} - \bar{\hat{\boldsymbol{r}}}^c_{2\cdot} - \bar{\boldsymbol{r}}^c_{1\cdot} + \bar{\boldsymbol{r}}^c_{2\cdot}) \right\|_2 \\
&= O_p(1) O_p(1) o_p(1) = o_p(1)
\end{aligned}
\tag{.17}
$$

Similarly, we have

$$
A_3 = o_p(1) \tag{.18}
$$

Put (.7), (.17) and (.18) together, we have

$$
T_{p,N} \xrightarrow{D} \chi^2(p) \tag{.19}
$$

Proof of Theorem 2

Under H_a when $E_t(diag(\boldsymbol{\lambda})\boldsymbol{\Phi}_{gi}^T\boldsymbol{\Sigma}_{Y_{gi}}^{-1}(\boldsymbol{\mu}_{gi}-\boldsymbol{\mu}_{pool;gi}))\neq \mathbf{0}$, then with probability 1

$$\bar{\tilde{\boldsymbol{r}}}_{1\cdot}^{*}-\bar{\tilde{\boldsymbol{r}}}_{2\cdot}^{*}\neq \mathbf{0}, \tag{.20}$$

where $\bar{\tilde{\boldsymbol{r}}}_{g\cdot}^{*}=\frac{1}{n_g}\sum_{i=1}^{n_g}\tilde{\boldsymbol{r}}_{gi}^{*}$, with $\tilde{\boldsymbol{r}}_{gi}^{*}=diag(\boldsymbol{\lambda})\boldsymbol{\Phi}_{gi}^T\boldsymbol{\Sigma}_{Y_{gi}}^{-1}(\boldsymbol{\mu}_{gi}-\boldsymbol{\mu}_{pool;gi})$

Next, we want to show that

$$(\bar{\hat{\tilde{\boldsymbol{r}}}}_{1\cdot}^{c}-\bar{\hat{\tilde{\boldsymbol{r}}}}_{2\cdot}^{c})^T\hat{\boldsymbol{V}}^{-1}(\bar{\hat{\tilde{\boldsymbol{r}}}}_{1\cdot}^{c}-\bar{\hat{\tilde{\boldsymbol{r}}}}_{2\cdot}^{c})\xrightarrow{p}(\bar{\tilde{\boldsymbol{r}}}_{1\cdot}^{*}-\bar{\tilde{\boldsymbol{r}}}_{2\cdot}^{*})^T\boldsymbol{V}^{-1}(\bar{\tilde{\boldsymbol{r}}}_{1\cdot}^{*}-\bar{\tilde{\boldsymbol{r}}}_{2\cdot}^{*}) \tag{.21}$$

Note that we have the following decomposition

$$\begin{aligned}
&|(\bar{\hat{\tilde{\boldsymbol{r}}}}_{1\cdot}^{c}-\bar{\hat{\tilde{\boldsymbol{r}}}}_{2\cdot}^{c})^T\hat{\boldsymbol{V}}^{-1}(\bar{\hat{\tilde{\boldsymbol{r}}}}_{1\cdot}^{c}-\bar{\hat{\tilde{\boldsymbol{r}}}}_{2\cdot}^{c})-(\bar{\tilde{\boldsymbol{r}}}_{1\cdot}^{*}-\bar{\tilde{\boldsymbol{r}}}_{2\cdot}^{*})^T\boldsymbol{V}^{-1}(\bar{\tilde{\boldsymbol{r}}}_{1\cdot}^{*}-\bar{\tilde{\boldsymbol{r}}}_{2\cdot}^{*})|\\
\leq&\left\|(\bar{\tilde{\boldsymbol{r}}}_{1\cdot}^{*}-\bar{\tilde{\boldsymbol{r}}}_{2\cdot}^{*})^T[\hat{\boldsymbol{V}}^{-1}-\boldsymbol{V}^{-1}](\bar{\tilde{\boldsymbol{r}}}_{1\cdot}^{*}-\bar{\tilde{\boldsymbol{r}}}_{2\cdot}^{*})\right\|_2+\left\|(\bar{\hat{\tilde{\boldsymbol{r}}}}_{1\cdot}^{c}-\bar{\hat{\tilde{\boldsymbol{r}}}}_{2\cdot}^{c}-\bar{\tilde{\boldsymbol{r}}}_{1\cdot}^{*}+\bar{\tilde{\boldsymbol{r}}}_{2\cdot}^{*})^T\hat{\boldsymbol{V}}^{-1}(\bar{\tilde{\boldsymbol{r}}}_{1\cdot}^{*}-\bar{\tilde{\boldsymbol{r}}}_{2\cdot}^{*})\right\|_2\\
+&\left\|(\bar{\hat{\tilde{\boldsymbol{r}}}}_{1\cdot}^{c}-\bar{\hat{\tilde{\boldsymbol{r}}}}_{2\cdot}^{c})^T\hat{\boldsymbol{V}}^{-1}(\bar{\hat{\tilde{\boldsymbol{r}}}}_{1\cdot}^{c}-\bar{\hat{\tilde{\boldsymbol{r}}}}_{2\cdot}^{c}-\bar{\tilde{\boldsymbol{r}}}_{1\cdot}^{*}+\bar{\tilde{\boldsymbol{r}}}_{2\cdot}^{*})\right\|_2\\
\stackrel{\text{def}}{=}&C_1+C_2+C_3
\end{aligned} \tag{.22}$$

As we have $\hat{\boldsymbol{V}}-\boldsymbol{V}=o_p(1)$, then

$$C_1\leq\left\|(\bar{\tilde{\boldsymbol{r}}}_{1\cdot}^{*}-\bar{\tilde{\boldsymbol{r}}}_{2\cdot}^{*})\right\|_2^2 o_p(1)=o_p(1) \tag{.23}$$

For C_2, we have,

$$\bar{\hat{\tilde{\boldsymbol{r}}}}_{g\cdot}^{c}-\bar{\tilde{\boldsymbol{r}}}_{g\cdot}^{*}=\bar{\hat{\tilde{\boldsymbol{r}}}}_{g\cdot}^{c}-\bar{\tilde{\boldsymbol{r}}}_{g\cdot}^{c}+diag(\boldsymbol{\lambda})\frac{1}{n_g}\sum_{i=1}^{n_g}\mathbf{H}_{gi}(\mathbf{Y}_{gi}-\boldsymbol{\mu}_{gi})=o_p(1) \tag{.24}$$

Then we have

$$\begin{aligned}
C_2\leq&\left\|\bar{\hat{\tilde{\boldsymbol{r}}}}_{1\cdot}^{c}-\bar{\tilde{\boldsymbol{r}}}_{1\cdot}^{*}-\bar{\hat{\tilde{\boldsymbol{r}}}}_{2\cdot}^{c}+\bar{\tilde{\boldsymbol{r}}}_{2\cdot}^{*}\right\|_2\left\|\hat{\boldsymbol{V}}^{-1}\right\|_2\left\|(\bar{\tilde{\boldsymbol{r}}}_{1\cdot}^{*}-\bar{\tilde{\boldsymbol{r}}}_{2\cdot}^{*})\right\|_2\\
\leq& o_p(1)O_p(1)\left\|\bar{\tilde{\boldsymbol{r}}}_{1\cdot}^{*}-\bar{\tilde{\boldsymbol{r}}}_{2\cdot}^{*}\right\|_2=o_p(1)
\end{aligned} \tag{.25}$$

Similarly, we have

$$C_3=o_p(1) \tag{.26}$$

Combining (.23), (.25) and (.26), and together with (.51), we have

$$(\bar{\hat{\tilde{\boldsymbol{r}}}}_{1\cdot}^{c}-\bar{\hat{\tilde{\boldsymbol{r}}}}_{2\cdot}^{c})^T\hat{\boldsymbol{V}}^{-1}(\bar{\hat{\tilde{\boldsymbol{r}}}}_{1\cdot}^{c}-\bar{\hat{\tilde{\boldsymbol{r}}}}_{2\cdot}^{c})\xrightarrow{p}(\bar{\tilde{\boldsymbol{r}}}_{1\cdot}^{*}-\bar{\tilde{\boldsymbol{r}}}_{2\cdot}^{*})^T\boldsymbol{V}^{-1}(\bar{\tilde{\boldsymbol{r}}}_{1\cdot}^{*}-\bar{\tilde{\boldsymbol{r}}}_{2\cdot}^{*})>0 \tag{.27}$$

Thus, $T_{p,N}\to\infty$

Proof of Theorem 5

Under the null hypothesis, for each of the independent shrinkage projection score vector $\tilde{\boldsymbol{\xi}}_i$, we know that it has mean zero vectors and a covariance matrix

$$Cov[\tilde{\boldsymbol{\xi}}_i] = E_t[diag(\boldsymbol{\omega})\boldsymbol{E}_i^T\boldsymbol{\Sigma}_{\mathbf{Y}_i}^{-1}\boldsymbol{\Psi}_i diag(\boldsymbol{\omega})]. \tag{.28}$$

The functional central limit theorem thus implies that

$$n^{-1/2}\sum_{i=1}^{nz}\tilde{\boldsymbol{\xi}}_i \xrightarrow{D} \boldsymbol{\Delta}_p(z),\ 0 \le z \le 1, \tag{.29}$$

where the process $\boldsymbol{\Delta}_p(z)$ takes value in R^p, and it has zero mean and covariance matrix $Cov[\tilde{\boldsymbol{\xi}}_i]$. Given the convergence in (.29), we have

$$\frac{1}{n}[\sum_{i=1}^{k}\tilde{\boldsymbol{\xi}}_i - z\sum_{i=1}^{n}\tilde{\boldsymbol{\xi}}_i]^T Cov[\tilde{\boldsymbol{\xi}}_i]^{-1}[\sum_{i=1}^{k}\tilde{\boldsymbol{\xi}}_i - z\sum_{i=1}^{n}\tilde{\boldsymbol{\xi}}_i] \xrightarrow{D} \sum_{l=1}^{p}B_l^2(z). \tag{.30}$$

According to the Slutsky's Theorem, let's denote $\boldsymbol{V}_3 = \frac{\sum_{i=1}^{n} diag(\boldsymbol{\omega})\boldsymbol{E}_i^T\boldsymbol{\Sigma}_{\mathbf{Y}_i}^{-1}\boldsymbol{E}_i diag(\boldsymbol{\omega})}{n}$, we also have

$$S^*_{p,n,z} \overset{\text{def}}{=} \frac{1}{n}[\sum_{i=1}^{k}\tilde{\boldsymbol{\xi}}_i - z\sum_{i=1}^{n}\tilde{\boldsymbol{\xi}}_i]^T\boldsymbol{V}_3^{-1}[\sum_{i=1}^{k}\tilde{\boldsymbol{\xi}}_i - z\sum_{i=1}^{n}\tilde{\boldsymbol{\xi}}_i] \xrightarrow{D} \sum_{l=1}^{p}B_l^2(z). \tag{.31}$$

Then in order to prove Theorem 5, we should prove that

$$S_{p,n,z} = S^*_{p,n,z} + o_p(1), \tag{.32}$$

where $S_{p,n,z} = \frac{1}{n}[\sum_{i=1}^{k}\hat{\tilde{\boldsymbol{\xi}}}_i - z\sum_{i=1}^{n}\hat{\tilde{\boldsymbol{\xi}}}_i]^T\hat{\boldsymbol{V}}_3^{-1}[\sum_{i=1}^{k}\hat{\tilde{\boldsymbol{\xi}}}_i - z\sum_{i=1}^{n}\hat{\tilde{\boldsymbol{\xi}}}_i]$, with $\hat{\boldsymbol{V}}_3 = \frac{1}{n}\sum_{i=1}^{n} diag(\hat{\boldsymbol{\omega}})\hat{\boldsymbol{E}}_i^T\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}_i}^{-1}\hat{\boldsymbol{E}}_i diag(\hat{\boldsymbol{\omega}})$. The absolute value of $S_{p,n,z} - S^*_{p,n,z}$ can be decomposed as follows.

$$\begin{aligned}
&|S_{p,n,z} - S^*_{p,n,z}| \\
=&\left\|\frac{1}{n}[\sum_{i=1}^{nz}\hat{\tilde{\boldsymbol{\xi}}}_i - z\sum_{i=1}^{n}\hat{\tilde{\boldsymbol{\xi}}}_i]^T\hat{\boldsymbol{V}}_3^{-1}[\sum_{i=1}^{nz}\hat{\tilde{\boldsymbol{\xi}}}_i - z\sum_{i=1}^{n}\hat{\tilde{\boldsymbol{\xi}}}_i] - \frac{1}{n}[\sum_{i=1}^{nz}\tilde{\boldsymbol{\xi}}_i - z\sum_{i=1}^{n}\tilde{\boldsymbol{\xi}}_i]^T\boldsymbol{V}_3^{-1}[\sum_{i=1}^{nz}\tilde{\boldsymbol{\xi}}_i - z\sum_{i=1}^{n}\tilde{\boldsymbol{\xi}}_i]\right\|_2 \\
\le&\left\|\frac{1}{n}[\sum_{i=1}^{nz}\tilde{\boldsymbol{\xi}}_i - z\sum_{i=1}^{n}\tilde{\boldsymbol{\xi}}_i]^T[\hat{\boldsymbol{V}}_3^{-1} - \boldsymbol{V}_3^{-1}][\sum_{i=1}^{nz}\tilde{\boldsymbol{\xi}}_i - z\sum_{i=1}^{n}\tilde{\boldsymbol{\xi}}_i]\right\|_2 \\
+&\left\|\frac{1}{n}(\sum_{i=1}^{nz}\hat{\tilde{\boldsymbol{\xi}}}_i - z\sum_{i=1}^{n}\hat{\tilde{\boldsymbol{\xi}}}_i - \sum_{i=1}^{nz}\tilde{\boldsymbol{\xi}}_i + z\sum_{i=1}^{n}\tilde{\boldsymbol{\xi}}_i)^T\hat{\boldsymbol{V}}_3^{-1}[\sum_{i=1}^{nz}\tilde{\boldsymbol{\xi}}_i - z\sum_{i=1}^{n}\tilde{\boldsymbol{\xi}}_i]\right\|_2 \\
+&\left\|\frac{1}{n}(\sum_{i=1}^{nz}\hat{\tilde{\boldsymbol{\xi}}}_i - z\sum_{i=1}^{n}\hat{\tilde{\boldsymbol{\xi}}}_i)^T\hat{\boldsymbol{V}}_3^{-1}(\sum_{i=1}^{nz}\hat{\tilde{\boldsymbol{\xi}}}_i - z\sum_{i=1}^{n}\hat{\tilde{\boldsymbol{\xi}}}_i - \sum_{i=1}^{nz}\tilde{\boldsymbol{\xi}}_i + z\sum_{i=1}^{n}\tilde{\boldsymbol{\xi}}_i)\right\|_2 \\
\overset{\text{def}}{=}& D_1 + D_2 + D_3
\end{aligned} \tag{.33}$$

Given (.33), proving Theorem 5 can be accomplished by proving $D_1 = o_p(1)$, $D_2 = o_p(1)$, and $D_3 = o_p(1)$ hold.

First, let's start to show $D_1 = o_p(1)$. Under H_0, for any fixed z, we have

$$\begin{aligned} & \frac{1}{n}^{1/2}(\sum_{i=1}^{nz}\tilde{\boldsymbol{\xi}}_i - z\sum_{i=1}^{n}\tilde{\boldsymbol{\xi}}_i) \\ =& \frac{1}{n}^{1/2}\frac{nz(n-nz)}{n}(\bar{\tilde{\boldsymbol{\xi}}}_{nz} - \bar{\tilde{\boldsymbol{\xi}}}_{n-nz}) \\ \xrightarrow{D} & N_p(\mathbf{0}, Cov[\tilde{\boldsymbol{\xi}}_i]) \end{aligned} \tag{.34}$$

Consequently,

$$\left\| \frac{1}{n}^{1/2}(\sum_{i=1}^{nz}\tilde{\boldsymbol{\xi}}_i - z\sum_{i=1}^{n}\tilde{\boldsymbol{\xi}}_i) \right\|_2 = O_p(1). \tag{.35}$$

And similar to (.5), we have $\left\|\hat{\mathbf{V}}_3 - \mathbf{V}_3\right\|_2 = o_p(1)$, then

$$\begin{aligned} D_1 \leq & \left\| \frac{1}{n}^{1/2}(\sum_{i=1}^{nz}\tilde{\boldsymbol{\xi}}_i - z\sum_{i=1}^{n}\tilde{\boldsymbol{\xi}}_i) \right\|_2^2 \left\|\hat{\boldsymbol{V}}_3^{-1} - \boldsymbol{V}_3^{-1}\right\|_2 \\ \leq & O_p(1) \times \frac{\left\|\boldsymbol{V}_3^{-1}\right\|_2}{[\left\|\boldsymbol{V}_3^{-1}\right\|_2\left\|\hat{\boldsymbol{V}}_3 - \boldsymbol{V}_3\right\|_2]^{-1} - 1} = o_p(1) \end{aligned} \tag{.36}$$

Next, we consider D_2. Note that

$$\begin{aligned} D_2 \leq & \left\| \frac{1}{n}^{1/2}(\sum_{i=1}^{nz}\hat{\tilde{\boldsymbol{\xi}}}_i - z\sum_{i=1}^{n}\hat{\tilde{\boldsymbol{\xi}}}_i - \sum_{i=1}^{nz}\tilde{\boldsymbol{\xi}}_i + z\sum_{i=1}^{n}\tilde{\boldsymbol{\xi}}_i) \right\|_2 \left\|\hat{\boldsymbol{V}}_3^{-1}\right\|_2 \left\| \frac{1}{n}^{1/2}\sum_{i=1}^{nz}(\tilde{\boldsymbol{\xi}}_i - z\sum_{i=1}^{n}\tilde{\boldsymbol{\xi}}_i) \right\|_2 \\ \leq & O_p(1)O_p(1)\left\| \frac{1}{n}^{1/2}(\sum_{i=1}^{nz}\hat{\tilde{\boldsymbol{\xi}}}_i - z\sum_{i=1}^{n}\hat{\tilde{\boldsymbol{\xi}}}_i - \sum_{i=1}^{nz}\tilde{\boldsymbol{\xi}}_i + z\sum_{i=1}^{n}\tilde{\boldsymbol{\xi}}_i) \right\|_2 \\ = & O_p(1)O_p(1)\left\| \frac{1}{n}^{1/2}\frac{nz(n-nz)}{n}[(\bar{\hat{\tilde{\boldsymbol{\xi}}}}_{nz} - \bar{\hat{\tilde{\boldsymbol{\xi}}}}_{n-nz}) - (\bar{\tilde{\boldsymbol{\xi}}}_{nz} - \bar{\tilde{\boldsymbol{\xi}}}_{n-nz})] \right\|_2 \\ = & O_p(1)O_p(1)\left\| n^{1/2}[(\bar{\hat{\tilde{\boldsymbol{\xi}}}}_{nz} - \bar{\hat{\tilde{\boldsymbol{\xi}}}}_{n-nz}) - (\bar{\tilde{\boldsymbol{\xi}}}_{nz} - \bar{\tilde{\boldsymbol{\xi}}}_{n-nz})] \right\|_2 \end{aligned} \tag{.37}$$

In order to prove $D_2 = o_p(1)$, we need to show $\left\| n^{1/2}[(\bar{\hat{\boldsymbol{\xi}}}_{nz} - \bar{\hat{\boldsymbol{\xi}}}_{n-nz}) - (\bar{\tilde{\boldsymbol{\xi}}}_{nz} - \bar{\tilde{\boldsymbol{\xi}}}_{n-nz})] \right\|_2 = o_p(1)$. We want to write out the explicit formula of $\bar{\hat{\boldsymbol{\xi}}}_{nz} - \bar{\tilde{\boldsymbol{\xi}}}_{nz}$:

$$\begin{aligned}
\hat{\tilde{\boldsymbol{\xi}}}_i &= diag(\hat{\boldsymbol{\omega}})\hat{\boldsymbol{E}}_i^T \hat{\boldsymbol{\Sigma}}_{\mathbf{Y}_i}^{-1}(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_{pool;i}) \\
&\overset{\text{def}}{=} diag(\hat{\boldsymbol{\omega}})\hat{\mathbf{U}}_i(\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_{pool;i}) \\
&= \tilde{\boldsymbol{\xi}}_i + (diag(\hat{\boldsymbol{\omega}}) - diag(\boldsymbol{\omega}))\mathbf{U}_i(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i}) + diag(\hat{\boldsymbol{\omega}})(\hat{\mathbf{U}}_i - \mathbf{U}_i)(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i}) \\
&+ diag(\hat{\boldsymbol{\omega}})\hat{\mathbf{U}}_i(\boldsymbol{\mu}_{pool;i} - \hat{\boldsymbol{\mu}}_{pool;i})
\end{aligned} \tag{.38}$$

i.e.,

$$\begin{aligned}
n^{1/2}(\bar{\hat{\boldsymbol{\xi}}}_{nz} - \bar{\tilde{\boldsymbol{\xi}}}_{nz}) &= n^{1/2}(diag(\hat{\boldsymbol{\omega}}) - diag(\boldsymbol{\omega}))\frac{1}{nz}\sum_{i=1}^{nz}\mathbf{U}_i(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i}) \\
&+ n^{1/2}diag(\hat{\boldsymbol{\omega}})\frac{1}{nz}\sum_{i=1}^{nz}(\hat{\mathbf{U}}_i - \mathbf{U}_i)(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i}) \\
&+ n^{1/2}diag(\hat{\boldsymbol{\omega}})\frac{1}{nz}\sum_{i=1}^{nz}\hat{\mathbf{U}}_i(\boldsymbol{\mu}_{pool;i} - \hat{\boldsymbol{\mu}}_{pool;i}) \\
&\overset{\text{def}}{=} E_1 + E_2 + E_3
\end{aligned} \tag{.39}$$

$$\begin{aligned}
n^{1/2}(\bar{\hat{\boldsymbol{\xi}}}_{n-nz} - \bar{\tilde{\boldsymbol{\xi}}}_{n-nz}) &= n^{1/2}(diag(\hat{\boldsymbol{\omega}}) - diag(\boldsymbol{\omega}))\frac{1}{n-nz}\sum_{i=nz+1}^{n}\mathbf{U}_i(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i}) \\
&+ n^{1/2}diag(\hat{\boldsymbol{\omega}})\frac{1}{n-nz}\sum_{i=nz+1}^{n}(\hat{\mathbf{U}}_i - \mathbf{U}_i)(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i}) \\
&+ n^{1/2}diag(\hat{\boldsymbol{\omega}})\frac{1}{n-nz}\sum_{i=nz+1}^{n}\hat{\mathbf{U}}_i(\boldsymbol{\mu}_{pool;i} - \hat{\boldsymbol{\mu}}_{pool;i}) \\
&\overset{\text{def}}{=} F_1 + F_2 + F_3
\end{aligned} \tag{.40}$$

According to (.39) and (.40), in order to ensure that $\left\| n^{-1/2}[(\bar{\hat{\boldsymbol{\xi}}}_{nz} - \bar{\hat{\boldsymbol{\xi}}}_{n-nz}) - (\bar{\tilde{\boldsymbol{\xi}}}_{nz} - \bar{\tilde{\boldsymbol{\xi}}}_{n-nz})] \right\|_2 = o_p(1)$, we need to examine $E_1 - F_1$, $E_2 - F_2$, and $E_3 - F_3$. Let's start with $E_1 - F_1$, we have

$$\begin{aligned}
\|E_1\|_2 &\le n^{1/2}\|diag(\hat{\boldsymbol{\omega}}) - diag(\boldsymbol{\omega})\|_2 \left\| \frac{1}{nz}\sum_{i=1}^{nz}\mathbf{U}_i(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i}) \right\|_2 \\
&= n^{1/2}o_p(1)\sqrt{\sum_{k=1}^{p}(\frac{1}{nz}\sum_{i=1}^{nz}\mathbf{U}_i(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i}))^2} \\
&= n^{1/2}o_p(1)\sqrt{O_P(\frac{1}{nz})} = o_P(1)
\end{aligned} \tag{.41}$$

where Central Limit theorem is used to bound the second norm term. Under H_0, we have $E[\mathbf{U}_i(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i})|\mathbf{t}] = 0$, $Var[\mathbf{Q}_i(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i})|\mathbf{t}] = \mathbf{U}_i\boldsymbol{\Sigma}_{Y_i}\mathbf{U}_i^T = \boldsymbol{E}_i^T\boldsymbol{\Sigma}_{Y_i}^{-1}\boldsymbol{E}_i$. As a result, we have

$$\begin{aligned} E[\mathbf{U}_i(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i})] &= E[E[\mathbf{U}_i(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i})|\mathbf{t}]] = 0 \\ Var[\mathbf{U}_i(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i})] &= Var[E[\mathbf{U}_i(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i})|\mathbf{t}]] + E[Var[\mathbf{U}_i(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i})|\mathbf{t}]] \\ &= 0 + E[\boldsymbol{E}_i^T\boldsymbol{\Sigma}_{Y_i}^{-1}\boldsymbol{E}_i] \end{aligned}$$

So as long as we have $E[\boldsymbol{E}_i^T\boldsymbol{\Sigma}_{Y_i}^{-1}\boldsymbol{E}_i] < \infty$, based on the Central Limit Theorem, we have $\frac{1}{nz}\sum_{i=1}^{nz}\mathbf{U}_i(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i}) = O_P(\frac{1}{\sqrt{nz}})$. Similarly, we have

$$\|F_1\|_2 = o_P(1) \tag{.42}$$

Combining (.41) and (.42), we have

$$\|E_1 - F_1\|_2 \leq \|E_1\|_2 + \|F_1\|_2 = o_p(1) \tag{.43}$$

For the second term in (.39) and (.40), $E_3 - F_3$. We use sample splitting techniques. We randomly select a proportion among the n random subjects, then with probability 1 that we will have data from both the former and the latter parts of the series. We utilize these data to estimate all the components, and use the remaining data to calculate the test statistics. Then we have that $(\hat{\mathbf{U}}_i - \mathbf{U}_i)(\mathbf{Y}_i$ and $\boldsymbol{\mu}_{pool;i})$ are independent. Then we have

$$\begin{aligned} \|E_2\|_2 &\leq n^{1/2}\|diag(\hat{\boldsymbol{\omega}})\|_2\left\|\frac{1}{nz}\sum_{i=1}^{nz}(\hat{\mathbf{U}}_i - \mathbf{U}_i)(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i}) - \frac{1}{n-nz}\sum_{i=nz+1}^{n}(\hat{\mathbf{U}}_i - \mathbf{U}_i)(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i})\right\|_2 \\ &= n^{1/2}O_p(1)o_p(1/\sqrt{n}) = o_p(1), \end{aligned} \tag{.44}$$

because $E[(\hat{\mathbf{U}}_i - \mathbf{U}_i)(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i})] = E[(\hat{\mathbf{U}}_i - \mathbf{U}_i)]E[(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i})] = \mathbf{0}$, and for each $k = 1, ..., p$, $Var[(\hat{\mathbf{U}}_i - \mathbf{U}_i)(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i})] = o_p(1)$. Based on (.45), we have

$$\|E_2 - F_2\|_2 \leq \|E_2\|_2 + \|F_2\|_2 = o_p(1) \tag{.45}$$

Note that our simulation study demonstrates good performances of our proposed test statistic, without sample splitting.

For the last term in (.39) and (.40), $E_3 - F_3$, note that the randomness of $\hat{\mathbf{U}}_i(\mathbf{Y}_i - \boldsymbol{\mu}_{pool;i})$ come from random design, which are the same for the former part and latter part. Let's denote the common expectation as $\boldsymbol{E}_0$, then

$$\begin{aligned}
&\|E_3 - F_3\|_2 \\
&\leq n^{1/2}\|diag(\hat{\boldsymbol{\omega}})\|_2 \left\| \frac{1}{nz}\sum_{i=1}^{nz} \hat{\mathbf{U}}_i(\boldsymbol{\mu}_{pool;i} - \hat{\boldsymbol{\mu}}_{pool;i}) - \frac{1}{n-nz}\sum_{i=nz+1}^{n-nz} \hat{\mathbf{U}}_i(\boldsymbol{\mu}_{pool;i} - \hat{\boldsymbol{\mu}}_{pool;i}) \right\|_2 \\
&\leq n^{1/2}O_p(1) \left\| \frac{1}{nz}\sum_{i=1}^{nz} [\hat{\mathbf{U}}_i(\boldsymbol{\mu}_{pool;i} - \hat{\boldsymbol{\mu}}_{pool;i}) - \boldsymbol{E}_0] - \frac{1}{n-nz}\sum_{i=nz+1}^{n} [\hat{\mathbf{U}}_i(\boldsymbol{\mu}_{pool;i} - \hat{\boldsymbol{\mu}}_{pool;i}) - \boldsymbol{E}_0] \right\|_2 \\
&= n^{1/2}O_p(1)o_p(1/\sqrt{nz} + 1/\sqrt{n-nz}) = o_p(1)
\end{aligned} \tag{.46}$$

Combining (.43), (.45) and (.46), we have

$$\left\| n^{1/2}[(\bar{\hat{\boldsymbol{\xi}}}_{nz} - \bar{\hat{\boldsymbol{\xi}}}_{n-nz}) - (\bar{\tilde{\boldsymbol{\xi}}}_{nz} - \bar{\tilde{\boldsymbol{\xi}}}_{n})] \right\|_2 \leq o_p(1) + o_p(1) + o_p(1) = o_p(1) \tag{.47}$$

Therefore,

$$\begin{aligned}
D_2 &\leq O_p(1)O_p(1) \left\| n^{1/2}[(\bar{\hat{\boldsymbol{\xi}}}_{nz} - \bar{\hat{\boldsymbol{\xi}}}_{n-nz}) - (\bar{\tilde{\boldsymbol{\xi}}}_{nz} - \bar{\tilde{\boldsymbol{\xi}}}_{n-nz}) \right\|_2 \\
&= O_p(1)O_p(1)o_p(1) = o_p(1)
\end{aligned} \tag{.48}$$

Similarly, we have

$$D_3 = o_p(1) \tag{.49}$$

Put (.36), (.48) and (.49) together, we have

$$S_{p,n,z} \xrightarrow{D} \sum_{l=1}^{p} B_l^2(z) \tag{.50}$$

Proof of Theorem 6

Under H_a when $E_t(diag(\boldsymbol{\omega})\boldsymbol{E}_i^T\boldsymbol{\Sigma}_{Y_i}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{pool;i})) \neq \mathbf{0}$, then with probability 1 that there exists z such that

$$\bar{\tilde{\boldsymbol{\xi}}}^*_{nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz} \neq \mathbf{0}, \tag{.51}$$

where $\bar{\tilde{\boldsymbol{\xi}}}^*_{nz} = \frac{1}{nz}\sum_{i=1}^{nz}\tilde{\boldsymbol{\xi}}_i^*$ and $\bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz} = \frac{1}{n-nz}\sum_{i=nz+1}^{n}\tilde{\boldsymbol{\xi}}_i^*$, with $\tilde{\boldsymbol{\xi}}_i^* = diag(\boldsymbol{\omega})\boldsymbol{E}_i^T\boldsymbol{\Sigma}_{Y_i}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{pool;i})$

Next, we want to show that

$$(\bar{\hat{\tilde{\boldsymbol{\xi}}}}_{nz} - \bar{\hat{\tilde{\boldsymbol{\xi}}}}_{n-nz})^T\hat{\boldsymbol{V}}_3^{-1}(\bar{\hat{\tilde{\boldsymbol{\xi}}}}_{nz} - \bar{\hat{\tilde{\boldsymbol{\xi}}}}_{n-nz}) \xrightarrow{p} (\bar{\tilde{\boldsymbol{\xi}}}^*_{nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz})^T\boldsymbol{V}_3^{-1}(\bar{\tilde{\boldsymbol{\xi}}}^*_{nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz}) \tag{.52}$$

Note that we have the following decomposition

$$\begin{aligned}
&|(\bar{\hat{\tilde{\boldsymbol{\xi}}}}_{nz} - \bar{\hat{\tilde{\boldsymbol{\xi}}}}_{n-nz})^T\hat{\boldsymbol{V}}_3^{-1}(\bar{\hat{\tilde{\boldsymbol{\xi}}}}_{nz} - \bar{\hat{\tilde{\boldsymbol{\xi}}}}_{n-nz}) - (\bar{\tilde{\boldsymbol{\xi}}}^*_{nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz})^T\boldsymbol{V}_3^{-1}(\bar{\tilde{\boldsymbol{\xi}}}^*_{nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz})| \\
\leq &\left\|(\bar{\tilde{\boldsymbol{\xi}}}^*_{nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz})^T[\hat{\boldsymbol{V}}_3^{-1} - \boldsymbol{V}_3^{-1}](\bar{\tilde{\boldsymbol{\xi}}}^*_{nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz})\right\|_2 + \left\|(\bar{\hat{\tilde{\boldsymbol{\xi}}}}_{nz} - \bar{\hat{\tilde{\boldsymbol{\xi}}}}_{n-nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{nz} + \bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz})^T\hat{\boldsymbol{V}}_3^{-1}(\bar{\tilde{\boldsymbol{\xi}}}^*_{nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz})\right\|_2 \\
+ &\left\|(\bar{\hat{\tilde{\boldsymbol{\xi}}}}_{nz} - \bar{\hat{\tilde{\boldsymbol{\xi}}}}_{n-nz})^T\hat{\boldsymbol{V}}_3^{-1}(\bar{\hat{\tilde{\boldsymbol{\xi}}}}_{nz} - \bar{\hat{\tilde{\boldsymbol{\xi}}}}_{n-nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{nz} + \bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz})\right\|_2 \\
\stackrel{\text{def}}{=} &G_1 + G_2 + G_3
\end{aligned} \tag{.53}$$

Note that $\hat{\boldsymbol{V}}_3 - \boldsymbol{V}_3 = o_p(1)$, then

$$G_1 \leq \left\|\bar{\tilde{\boldsymbol{\xi}}}^*_{nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz}\right\|_2^2 o_p(1) = o_p(1) \tag{.54}$$

For G_2, we have,

$$\begin{aligned}
\bar{\hat{\tilde{\boldsymbol{\xi}}}}_{nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{nz} &= \bar{\hat{\tilde{\boldsymbol{\xi}}}}_{nz} - \bar{\tilde{\boldsymbol{\xi}}}_{nz} + diag(\boldsymbol{\omega})\frac{1}{n}\sum_{i=1}^{nz}\mathbf{U}_i(\mathbf{Y}_i - \boldsymbol{\mu}_i) = o_p(1) \\
\bar{\hat{\tilde{\boldsymbol{\xi}}}}_{n-nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz} &= \bar{\hat{\tilde{\boldsymbol{\xi}}}}_{n-nz} - \bar{\tilde{\boldsymbol{\xi}}}_{n-nz} + diag(\boldsymbol{\omega})\frac{1}{n}\sum_{i=nz+1}^{n}\mathbf{U}_i(\mathbf{Y}_i - \boldsymbol{\mu}_i) = o_p(1)
\end{aligned} \tag{.55}$$

Then we have

$$\begin{aligned}
G_2 &\leq \left\|\bar{\hat{\tilde{\boldsymbol{\xi}}}}_{nz} - \bar{\hat{\tilde{\boldsymbol{\xi}}}}_{n-nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{nz} + \bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz}\right\|_2\left\|\hat{\boldsymbol{V}}_3^{-1}\right\|_2\left\|\bar{\tilde{\boldsymbol{\xi}}}^*_{nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz}\right\|_2 \\
&\leq o_p(1)O_p(1)\left\|\hat{\boldsymbol{V}}_3^{-1}\right\|_2\left\|\bar{\tilde{\boldsymbol{\xi}}}^*_{nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz}\right\|_2 = o_p(1)
\end{aligned} \tag{.56}$$

Similarly, we have

$$G_3 = o_p(1) \tag{.57}$$

Combining (.54), (.56) and (.57), we have

$$(\bar{\tilde{\boldsymbol{\xi}}}_{nz} - \bar{\tilde{\boldsymbol{\xi}}}_{n-nz})^T \hat{\boldsymbol{V}}_3^{-1} (\bar{\tilde{\boldsymbol{\xi}}}_{nz} - \bar{\tilde{\boldsymbol{\xi}}}_{n-nz}) \xrightarrow{p} (\bar{\tilde{\boldsymbol{\xi}}}^*_{nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz})^T \boldsymbol{V}_3^{-1} (\bar{\tilde{\boldsymbol{\xi}}}^*_{nz} - \bar{\tilde{\boldsymbol{\xi}}}^*_{n-nz}) > 0 \tag{.58}$$

Thus

$$\begin{aligned} C^{(1)}_{p,n} &\xrightarrow{p} \infty \\ C^{(2)}_{p,n} &\xrightarrow{p} \infty \end{aligned} \tag{.59}$$

BIBLIOGRAPHY

Felix Abramovich, Anestis Antoniadis, Theofanis Sapatinas, and Brani Vidakovic. Optimal testing in a fixed-effects functional analysis of variance model. *International Journal of Wavelets, Multiresolution and Information Processing*, 2(04):323–349, 2004.

John AD Aston and Claudia Kirch. Detecting and estimating changes in dependent functional data. *Journal of Multivariate Analysis*, 109:204–220, 2012.

Alexander Aue, Robertas Gabrys, Lajos Horváth, and Piotr Kokoszka. Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis*, 100(10):2254–2269, 2009.

Michal Benko, Wolfgang Härdle, and Alois Kneip. Common functional principal components. *The Annals of Statistics*, 37(1):1–34, 2009.

István Berkes, Robertas Gabrys, Lajos Horváth, and Piotr Kokoszka. Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(5):927–946, 2009.

Graciela Boente, Daniela Rodriguez, and Mariela Sued. Inference under functional proportional and common principal component models. *Journal of Multivariate Analysis*, 101(2):464–475, 2010.

T Tony Cai, Peter Hall, et al. Prediction in functional linear regression. *The Annals of Statistics*, 34(5):2159–2179, 2006.

Hervé Cardot, Frédéric Ferraty, and Pascal Sarda. Spline estimators for the functional linear model. *Statistica Sinica*, pages 571–591, 2003.

Hervé Cardot, Christophe Crambes, and Pascal Sarda. Quantile regression when the covariates are functions. *Nonparametric Statistics*, 17(7):841–856, 2005.

Hervé Cardot, Camelia Goga, and Pauline Lardin. Uniform convergence and asymptotic confidence bands for model-assisted estimators of the mean of sampled functional data. *Electronic journal of statistics*, 7:562–596, 2013.

K. Chen and H.-G Müller. Modeling conditional distributions for functional responses, with application to traffic monitoring via gps-enabled mobile phones. *Technometrics*, 2013. http://www.tandfonline.com/doi/abs/10.1080/00401706.2013.842933.

Kehui Chen and Hans-Georg Müller. Conditional quantile analysis when covariates are functions, with application to growth data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(1):67–89, 2012a.

Kehui Chen and Hans-Georg Müller. Modeling repeated functional observations. *Journal of the American Statistical Association*, 107(500):1599–1609, 2012b.

Kehui Chen, Pedro Delicado, and Hans-Georg Müller. Modelling function-valued stochastic processes, with applications to fertility dynamics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(1):177–196, 2017.

Antonio Cuevas, Manuel Febrero, and Ricardo Fraiman. An anova test for functional data. *Computational statistics & data analysis*, 47(1):111–122, 2004.

Ebay.com. *Ebay website*, 1995. http://www.ebay.com/.

Jianqing Fan and Irene Gijbels. *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press, 1996.

Jianqing Fan and Sheng-Kuei Lin. Test of significance when data are curves. *Journal of the American Statistical Association*, 93(443):1007–1021, 1998.

Julian J Faraway. Regression analysis for a functional response. *Technometrics*, 39(3):254–261, 1997.

Frédéric Ferraty, Abbes Rabhi, and Philippe Vieu. Conditional quantiles for dependent functional data with application to the climatic" el niño" phenomenon. *Sankhyā: The Indian Journal of Statistics*, pages 378–398, 2005.

Garrett M Fitzmaurice, Nan M Laird, and James H Ware. *Applied longitudinal analysis*, volume 998. John Wiley & Sons, 2012.

Bernhard N Flury. Common principal components in k groups. *Journal of the American Statistical Association*, 79(388):892–898, 1984.

Stefan Fremdt, Josef G Steinebach, Lajos Horváth, and Piotr Kokoszka. Testing the equality of covariance operators in functional samples. *Scandinavian Journal of Statistics*, 40(1):138–152, 2013.

G Gaines, K Kaphle, and F Ruymgaart. Application of a delta-method for random operators to testing equality of two covariance operators. *Mathematical Methods of Statistics*, 20(3):232–245, 2011.

Peter Hall and Ingrid Van Keilegom. Two-sample tests in functional data analysis starting from discrete data. *Statistica Sinica*, 17(4):1511, 2007.

Douglas M Hawkins. Testing a sequence of observations for a shift in location. *Journal of the American Statistical Association*, 72(357):180–186, 1977.

Robin Henderson. Change-point problem with correlated observations, with an application in material accountancy. *Technometrics*, 28(4):381–389, 1986.

Lajos Horváth and Piotr Kokoszka. *Inference for functional data with applications*, volume 200. Springer Science & Business Media, 2012.

Lajos Horváth and Gregory Rice. Testing equality of means when the observations are from functional time series. *Journal of Time Series Analysis*, 36(1):84–108, 2015.

Gareth M James and Catherine A Sugar. Clustering for sparsely sampled functional data. *Journal of the American Statistical Association*, 98(462):397–408, 2003.

Gareth M James, Trevor J Hastie, and Catherine A Sugar. Principal component models for sparse functional data. *Biometrika*, 87(3):587–602, 2000.

Jank and Shmueli. *Modeling Online Auctions*, 2010. http://www.modelingonlineauctions.com/datasets.

Wolfgang Jank, Galit Shmueli, Catherine Plaisant, and Ben Shneiderman. Visualizing functional data with an application to ebayõs online auctions. In *Handbook of Data Visualization*, pages 873–898. Springer, 2008.

Lawrence Joseph, Alain C Vandal, and David B Wolfson. Estimation in the multipath change point problem for correlated data. *Canadian Journal of Statistics*, 24(1):37–53, 1996.

J Kiefer. K-sample analogues of the kolmogorov-smirnov and cramér-v. mises tests. *The Annals of Mathematical Statistics*, pages 420–447, 1959.

Hyune-Ju Kim. Change-point detection for correlated observations. *Statistica Sinica*, pages 275–287, 1996.

Yehua Li and Tailen Hsing. Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, pages 3321–3351, 2010.

Rosa E Lillo, Pedro Galeano, and Esdras Joseph. Two-sample hotelling's t^2 statistics based on the functional mahalanobis semi-distance. Technical report, Universidad Carlos III de Madrid. Departamento de Estadística, 2015.

Bitao Liu and Hans-Georg Müller. Functional data analysis for sparse auction data. *Statistical Methods in eCommerce Research*, pages 269–290, 2008.

Shujie Ma, Lijian Yang, and Raymond J Carroll. A simultaneous confidence band for sparse longitudinal regression. *Statistica Sinica*, 22:95, 2012.

Jeffrey S Morris and Raymond J Carroll. Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(2):179–199, 2006.

H.-G. Müller. Functional modelling and classification of longitudinal data*. *Scandinavian Journal of Statistics*, 32(2):223–240, 2005.

Efstathios Paparoditis and Theofanis Sapatinas. Bootstrap-based testing for functional data. *arXiv preprint arXiv:1409.4317*, 2014.

Debashis Paul and Jie Peng. Consistency of restricted maximum likelihood estimators of principal components. *The Annals of Statistics*, 37(3):1229–1271, 2009.

Jie Peng and Hans-Georg Müller. Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. *The Annals of Applied Statistics*, pages 1056–1077, 2008.

Jie Peng and Debashis Paul. A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data. *Journal of Computational and Graphical Statistics*, 18(4), 2009.

Moshe Pollak. Optimal detection of a change in distribution. *The Annals of Statistics*, pages 206–227, 1985.

Gina-Maria Pomann, Ana-Maria Staicu, and Sujit Ghosh. A two-sample distribution-free test for functional data with application to a diffusion tensor imaging study of multiple sclerosis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 2016.

Alfredas Račkauskas and Charles Suquet. Testing epidemic changes of infinite dimensional parameters. *Statistical Inference for Stochastic Processes*, 9(2):111–134, 2006.

Asoka Ramanayake and Arjun K Gupta. Tests for an epidemic change in a sequence of exponentially distributed random variables. *Biometrical journal*, 45(8):946–958, 2003.

James O Ramsay. *Functional data analysis*. Wiley Online Library, 2006.

Jaxk Reeves, Jien Chen, Xiaolan L Wang, Robert Lund, and Qi Qi Lu. A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, 46(6):900–915, 2007.

Fabian Scheipl, Sonja Greven, and Helmut Küchenhoff. Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Computational statistics & data analysis*, 52(7):3283–3299, 2008.

Qing Shen and Juliana Faraway. An f test for linear models with functional responses. *Statistica Sinica*, 14(4):1239–1258, 2004.

Bernard W Silverman. Smoothed functional principal components analysis by choice of norm. *The Annals of Statistics*, 24(1):1–24, 1996.

Vasilios A Siris and Fotini Papagalou. Application of anomaly detection algorithms for detecting syn flooding attacks. In *Global Telecommunications Conference, 2004. GLOBECOM'04. IEEE*, volume 4, pages 2050–2054. IEEE, 2004.

Ana-Maria Staicu, Yingxing Li, Ciprian M Crainiceanu, and David Ruppert. Likelihood ratio tests for dependent data with applications to longitudinal and functional data analysis. *Scandinavian Journal of Statistics*, 41(4):932–949, 2014.

Alexander G Tartakovsky, Boris L Rozovskii, Rudolf B Blažek, and Hongjoong Kim. Detection of intrusions in information systems by sequential change-point methods. *Statistical Methodology*, 3(3):252–293, 2006.

Shanshan Wang, Wolfgang Jank, and Galit Shmueli. Forecasting ebayõs online auction prices using functional data analysis. *University of Maryland, College Park, MD*, 20742, 2004.

Kevin P Weinfurt, Richard J Willke, Henry A Glick, William W Freimuth, and Kevin A Schulman. Relationship between cd4 count, viral burden, and quality of life over time in hiv-1-infected patients. *Medical Care*, 38(4):404–410, 2000.

Jianshu Weng and Bu-Sung Lee. Event detection in twitter. *ICWSM*, 11:401–408, 2011.

KJ Worsley. Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika*, 73(1):91–104, 1986.

Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100(470):577–590, 2005.

Qiwei Yao. Tests for change-points with epidemic alternatives. *Biometrika*, 80(1):179–191, 1993.

Jin-Ting Zhang. *Analysis of variance for functional data*. CRC Press, 2013.

Jin-Ting Zhang and Jianwei Chen. Statistical inferences for functional data. *The Annals of Statistics*, 35(3):1052–1079, 2007.

Jin-Ting Zhang and Xuehua Liang. One-way anova for functional data via globalizing the pointwise f-test. *Scandinavian Journal of Statistics*, 41(1):51–71, 2014.

Jin-Ting Zhang, Xuehua Liang, and Shengning Xiao. On the two-sample behrens-fisher problem for functional data. *Journal of Statistical Theory and Practice*, 4(4):571–587, 2010.

Xiaoke Zhang and Jane-Ling Wang. From sparse to dense functional data and beyond. *The Annals of Statistics*, 44(5):2281–2321, 2016.

Yan Zhou, Liya Fu, and Baoxue Zhang. Two non parametric methods for change-point detection in distribution. *Communications in Statistics-Theory and Methods*, 46(6):2801–2815, 2017.