

Two-sample inference for sparse functional data

Qiyao Wang

*Department of Statistics, University of Pittsburgh
1826 Wesley W. Posvar Hall
Pittsburgh, PA 15260
e-mail: qiw31@pitt.edu*

Abstract: In this paper, we develop an asymptotic χ^2 test for detecting differences among mean functions when sparse and irregular observations are drawn from the underlying continuous stochastic processes for each subject. We provide theoretical arguments to justify the effectiveness of the proposed test procedure. Numerical experiments, including simulation studies and applications to a CD4 count data set and an eBay online auction data set, are presented to demonstrate the good performance of the developed test.

Keywords and phrases: Sparse and irregular design, longitudinal studies, shrinkage estimator, asymptotic distribution.

Received September 2019.

1. Introduction

Functional data refers to data drawn from continuous underlying processes. Examples include time series, image data, and tracing data such as hand-writings. Statistical methods for this type of data, called functional data analysis (FDA), have been under rapid development 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 common when automated instruments are utilized to record the data. Sparse functional data arise frequently in many real-world applications as well, such as longitudinal studies (e.g., the AIDS Clinical Trial Study in Section 4.2) and eBay online auctions in Section 4.3. These two types of functional data often require different sets of techniques to model. The book [33] offers a comprehensive perspective of FDA methods for densely-observed functional data, and the paper [26] provides a nice review for sparse functional data analysis.

There is substantial literature on modeling and estimation for functional data. In this paper, we focus on the mean function inference problem. Let $X_g(t) \in L^2(\mathcal{T})$, $g = 1, 2$, denote the g -th random process, with a mean function $\mu_g(t)$ and a covariance function $G_g(s, t)$, $s, t \in \mathcal{T}$. Both $\mu_g(t)$ and $G_g(s, t)$ are unknown. $X_1(t)$ and $X_2(t)$ are assumed to be independent. The two-sample mean function

testing problem can be stated as testing $\mu_1(t) = \mu_2(t)$ for all $t \in \mathcal{T}$ against the general alternative that the two mean functions are not equal at some time points within the time range \mathcal{T} . The statistical framework of one-way functional ANOVA is analogous.

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, including the pointwise t-test [33], the L^2 -norm-based test and the F-type test [41, 9, 42, 43], the Hotelling's T^2 test with permutation [31], tests involving basis representations [1, 16, 15, 17, 21, 36], and the pseudo-likelihood ratio test [37]. Extensions to multiple functional samples are discussed in [35, 8, 6, 27]. Besides, tests aiming at detecting differences in functional distributions [32, 13] are also applicable to functional mean testing problems under further assumptions. Recently, [12] developed a method for test of the equivalence for functional data. For a related purpose, [4] developed simultaneous confidence bands for mean functions. The point-wise F-test and a simultaneous test based on the permutation distribution of the supremum of the point-wise F-test are proposed to test the parameter functions in function-on-scalar regressions [33, 34]. When defining the scalar-valued predictor as a dummy variable that indicates the group label of functional response, the tests for the parameter function being 0 is equivalent to the two-sample mean function inference problem in the paper. It is noteworthy that the F-type tests in function-on-scalar regressions, under these circumstances, are the same as the F-type mean function tests in [41, 9, 42, 43].

When it comes to sparse functional data, where only a few irregularly-observed measurements are available for each subject, the testing problem is more challenging and complicated. Most of the aforementioned methods assume one can fully observe the individual functions, such that the regular sample mean and sample covariance functions are attainable. In practice, these tests require the individual curves from each subject being observed at the same dense regular grid. When observing times are not the same for all the subjects or observed data are seriously contaminated with measurement errors, pre-smoothing for each curve is used [13]. However, pre-smoothing techniques based on individual curves are no longer reliable for sparse functional data, considering the limited amount of data for each subject. In the simulation studies, we implement the PACE method proposed [40] to jointly use data from all curves to conduct curve recovery. But note that the curves recovered this way are not consistent to the true curves; therefore the tests either have inflated type I errors or very low powers. There is some recent work regarding one-sample inference for sparse functional data [25]. These methods cannot be easily generalized to two or more functional samples.

To our best knowledge, the pseudo likelihood ratio test [37] and the nonparametric distribution test [32] are the only two tests that are also applicable to sparsely sampled functional data. In [37], they proposed an extension to sparse data with some details, but the numerical studies focused on the experiments for dense data, and the most sparse case in simulations has 10 data points on each curve. We found that the test has inflated type I error in very sparse cases. [32] proposed to apply the Anderson-Darling test on FPC scores to test the equality

of two distributions. When the data are sparsely observed, they extended the FPC scores in [40] to pooled data of two groups and used the so-called marginal FPC scores in the distribution test, but they only briefly discussed that this replacement will still lead to “close” test results. All their numerical experiments are for dense data.

In this paper, we propose a test procedure specially designed for the sparse functional mean testing problem and theoretically derive its asymptotic null distribution and large sample consistency. We propose a Hotelling’s T^2 type statistic based on a shrinkage eigen-projection score vector. This is an alternative way of extending the sparse FPC in [40] to multiple groups of sparse functional data, in addition to the marginal FPC in [32]. Unlike the regular eigen-projection score vector, the distribution of the individual shrinkage score depends on its specific design, i.e., the observing time points. We notice that if we account for the randomness of designs, the distribution of the shrinkage score is still i.i.d across all the subjects. Based on the distribution of the estimated shrinkage score, we propose an asymptotic χ^2 test statistic. The development of the asymptotic null distribution is nontrivial since we need to consider the estimation error of the shrinkage score for each curve and control the overall error uniformly across all subjects.

The proposed statistic can be roughly viewed as a Hotelling’s T^2 type statistic applied on the estimated shrinkage scores, with a new variance estimator. [31] proposed a functional Hotelling’s T^2 statistic for fully observed functional data, which is based on the sample mean and sample covariance of the data. When the data are observed on a discrete but dense grid, one can use smoothing methods to recover the underlying function, and then the Hotelling’s T^2 statistic can be empirically computed by truncating at the first p eigen components, and they proposed to use a permutation distribution for the statistic. For sparsely observed functional data, one can consider the permutation test constructed on the recovered curve, which can still control type I error, but our numerical results show that it has low power.

The remainder of this paper is organized as follows. The main results for the two-sample mean testing problem for sparse functional data, including the problem settings, the proposed test procedure, the corresponding asymptotic results, and an extension to common principal component cases, are stated in Sections 2 and 3. The performance of our test procedure is illustrated by three simulation studies and applications to a CD4 count data set and an eBay online auction data set are described in Section 4. The paper is concluded in Section 5. All the proofs are in Appendix B. An extension to the one-way functional ANOVA problem is briefly described in Appendix C.

2. Proposed two-sample test procedure

2.1. Problem settings

The model we consider is

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

where Y_{gij} denotes the j -th observation of the i -th subject in group g , for $g = 1, 2; i = 1, \dots, n_g; j = 1, \dots, N_{gi}$. The random samples $X_{gi}(t)$ 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 $\mathcal{T} = [0, 1]$, and then X_{gi} are random samples from $X_g \sim SP(\mu_g, G)$. Here $\epsilon_{gij} \sim N(0, \sigma^2)$ are i.i.d measurement errors. The number of observations for the i -th subject in group g is denoted as N_{gi} , and we focus on sparse scenarios where N_{gi} is finite. Given N_{gi} , random observing times $T_{gi1}, \dots, T_{giN_{gi}}$ are i.i.d with a bounded density function within the time domain \mathcal{T} . It is also assumed that $X_{gi}(t)$, N_{gi} and ϵ_{gij} are mutually independent. The hypothesis testing of interest is

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

2.2. Proposed test procedure based on shrinkage estimator

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 $\mu_{pool}(t)$ from both sides of model in Eq. (2.1). Let $Y_{gij}^c = Y_{gij} - \mu_{pool}(T_{gij})$ and $X_{gi}^c(t) = X_{gi}(t) - \mu_{pool}(t)$, then model in Eq. (2.1) becomes

$$Y_{gij}^c = X_{gi}^c(T_{gij}) + \epsilon_{gij}, \quad (2.3)$$

where $X_{gi}^c \sim SP(\mu_g - \mu_{pool}, G)$.

Let $\{\phi_k(t), k = 1, 2, \dots, \infty\}$ be orthonormal eigenfunctions of $G(s, t)$, corresponding to the non-increasing eigenvalue sequence $\{\lambda_k, k = 1, 2, \dots, \infty\}$. We define the projection score of $\mu_g(t) - \mu_{pool}(t)$ onto the k -th eigenfunction $\phi_k(t)$ as θ_{gk}^c , which is the inner product of $\mu_g(t) - \mu_{pool}(t)$ and $\phi_k(t)$, i.e.,

$$\theta_{gk}^c = \langle \mu_g - \mu_{pool}, \phi_k \rangle, \quad g = 1, 2; k \geq 1. \quad (2.4)$$

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

$$H_0 : \{\theta_{1k}^c\}_{k=1}^{\infty} = \{\theta_{2k}^c\}_{k=1}^{\infty} \quad VS \quad H_a : \{\theta_{1k}^c\}_{k=1}^{\infty} \neq \{\theta_{2k}^c\}_{k=1}^{\infty}. \quad (2.5)$$

Given the following projection of $X_{gi}^c(t)$,

$$X_{gi}^c(t) = \sum_{k=1}^{\infty} \langle X_{gi}^c, \phi_k \rangle \phi_k(t) \stackrel{\text{def}}{=} \sum_{k=1}^{\infty} r_{gik}^c \phi_k(t), \quad (2.6)$$

and the fact that testing Eq. (2.5) is equivalent to testing

$$H_0 : \{E[r_{1ik}^c]\}_{k=1}^{\infty} = \{E[r_{2ik}^c]\}_{k=1}^{\infty} \quad VS \quad H_a : \{E[r_{1ik}^c]\}_{k=1}^{\infty} \neq \{E[r_{2ik}^c]\}_{k=1}^{\infty}, \quad (2.7)$$

under dense data settings, projection-based tests truncate at the first p dimensions and construct test statistics based on r_{gik}^c , $g = 1, 2; i = 1, \dots, n_g; k = 1, \dots, p$.

For dense functional data, to obtain a consistent estimate of the projection score r_{gik}^c , one can utilize the estimator $\int_{t \in \mathcal{T}} X_{gi}^c(t) \hat{\phi}_k(t) dt$. However, for sparse functional data, we do not have the functions $X_{gi}^c(t)$. We propose to consider the best linear unbiased predictor, $E[r_{gik}^c | \mathbf{Y}_{gi}^c]$, where $\mathbf{Y}_{gi}^c = (Y_{gi1}^c, \dots, Y_{giN_{gi}}^c)^T$ is the vector contains observations of the corresponding subject. $E[r_{gik}^c | \mathbf{Y}_{gi}^c]$ is a rational choice. This is because given

$$E[E[r_{gik}^c | \mathbf{Y}_{gi}^c]] = E[r_{gik}^c], \quad (2.8)$$

the expectation of random quantity $E[r_{gik}^c | \mathbf{Y}_{gi}^c]$ equals to the expectation of the projection score r_{gik}^c , when taking the randomness of design points under sparse data cases into account.

Under the null hypothesis of the testing problem in Eq. (2.2), we have $\mu_g(t) = \mu_{pool}(t)$ and $\boldsymbol{\theta}_g^c = (\theta_{g1}^c, \dots, \theta_{gp}^c)^T = \mathbf{0}$, so we propose to construct a test statistic based on the shrinked first p best linear predictors $E[r_{gi}^c | \mathbf{Y}_{gi}^c] \stackrel{\text{def}}{=} \tilde{\mathbf{r}}_{gi}^c$. When the projection scores \mathbf{r}_{gi}^c is Gaussian (i.e., the random processes $X_1(t)$ and $X_2(t)$ are Gaussian processes) and are jointly Gaussian distributed with random errors ϵ_{gij} , the explicit formula of $\tilde{\mathbf{r}}_{gi}^c$ under H_0 is

$$\tilde{\mathbf{r}}_{gi}^c \stackrel{H_0}{=} \text{diag}(\boldsymbol{\lambda}) \boldsymbol{\Phi}_{gi}^T \boldsymbol{\Sigma}_{\mathbf{Y}_{gi}^c}^{-1} (\mathbf{Y}_{gi} - \boldsymbol{\mu}_{pool,gi}), \quad (2.9)$$

where $\mathbf{r}_{gi}^c = (r_{gi1}^c, \dots, r_{gip}^c)^T$, $\mathbf{Y}_{gi} = (Y_{gi1}, \dots, Y_{giN_{gi}})^T$, $\boldsymbol{\mu}_{pool,gi} = (\mu_{pool}(T_{gi1}), \dots, \mu_{pool}(T_{giN_{gi}}))^T$, $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_p)^T$, $\boldsymbol{\Phi}_{gi} = (\phi_{gi1}, \dots, \phi_{gip})$ is a N_{gi} by p matrix with the k th column being $\phi_{gik} = (\phi_k(T_{gi1}), \dots, \phi_k(T_{giN_{gi}}))^T$, and $\boldsymbol{\Sigma}_{\mathbf{Y}_{gi}^c}$ is the covariance matrix of \mathbf{Y}_{gi}^c , whose (j, j') element equals to $G(T_{gij}, T_{gij'}) + \sigma^2 \mathbf{1}(j = j')$. This is referred as **shrinkage score** in the rest of the paper.

The shrinkage score in Eq. (2.9) is not the same as the original score \mathbf{r}_{gi}^c in Eq. (2.6), but it gets closer to \mathbf{r}_{gi}^c when the number of observations on each curve goes large and the measurement error gets small. Best linear predictors have been first proposed in [40] for principal component analysis for one-sample i.i.d sparse functional data. Our proposed shrinkage score for multiple groups of sparse data is different from the marginal FPC score predictor in [32]. The eigenfunctions and eigenvalues in our proposed shrinkage score correspond to the covariance matrix $G(s, t)$ instead of the mixture covariance matrix $G_{mix}(s, t) = G(s, t) + \nu(1 - \nu)(\mu_1(s) - \mu_2(s))(\mu_1(t) - \mu_2(t))$, where $\nu \in [0, 1]$ and $1 - \nu \in [0, 1]$ are respectively the probability of random samples of the mixture process coming from $X_1(t)$ and $X_2(t)$.

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

$$\begin{aligned} E[\tilde{\mathbf{r}}_{gi}^c | \mathbf{T}_{gi}] &= \text{diag}(\boldsymbol{\lambda}) \boldsymbol{\Phi}_{gi}^T \boldsymbol{\Sigma}_{\mathbf{Y}_{gi}^c}^{-1} (\boldsymbol{\mu}_{gi} - \boldsymbol{\mu}_{pool,gi}) \\ \text{cov}[\tilde{\mathbf{r}}_{gi}^c | \mathbf{T}_{gi}] &= \text{diag}(\boldsymbol{\lambda}) \boldsymbol{\Phi}_{gi}^T \boldsymbol{\Sigma}_{\mathbf{Y}_{gi}^c}^{-1} \boldsymbol{\Phi}_{gi} \text{diag}(\boldsymbol{\lambda}). \end{aligned} \quad (2.10)$$

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

$$\begin{aligned} E[\tilde{\mathbf{r}}_{gi}^c] &= E_t[\text{diag}(\boldsymbol{\lambda})\boldsymbol{\Phi}_{gi}^T\boldsymbol{\Sigma}_{\mathbf{Y}_{gi}^c}^{-1}(\boldsymbol{\mu}_{gi} - \boldsymbol{\mu}_{pool,gi})] \\ \text{cov}[\tilde{\mathbf{r}}_{gi}^c] &= E_t[\text{diag}(\boldsymbol{\lambda})\boldsymbol{\Phi}_{gi}^T\boldsymbol{\Sigma}_{\mathbf{Y}_{gi}^c}^{-1}\boldsymbol{\Phi}_{gi}\text{diag}(\boldsymbol{\lambda})] + \text{cov}_t(\text{diag}(\boldsymbol{\lambda})\boldsymbol{\Phi}_{gi}^T\boldsymbol{\Sigma}_{\mathbf{Y}_{gi}^c}^{-1}(\boldsymbol{\mu}_{gi} - \boldsymbol{\mu}_{pool,gi})). \end{aligned} \quad (2.11)$$

Here $\boldsymbol{\mu}_{gi} = (\mu_g(T_{gi1}), \dots, \mu_g(T_{giN_{gi}}))^T$. Note that the second term in $\text{cov}[\tilde{\mathbf{r}}_{gi}^c]$ is zero under the null hypothesis. All the quantities involved in estimating $\tilde{\mathbf{r}}_{gi}^c$ and $\text{cov}[\tilde{\mathbf{r}}_{gi}^c]$ can be obtained from data Y_{gij} , $g = 1, 2$; $i = 1, \dots, n_g$; $j = 1, \dots, N_{gi}$, with details included in the next subsection. Let's denote the estimated quantities as \hat{G} , $\hat{\lambda}_k$, $\hat{\phi}_k(t)$, $\hat{\mu}_{pool}$, and $\hat{\sigma}$. Under the null hypothesis, we have the following empirical estimators

$$\begin{aligned} \hat{\mathbf{r}}_{gi}^c &= \text{diag}(\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\Phi}}_{gi}^T[\hat{\mathbf{G}}_{gi} + \hat{\sigma}^2\mathbf{I}]^{-1}(\mathbf{Y}_{gi} - \hat{\boldsymbol{\mu}}_{pool;gi}) \\ \widehat{\text{cov}}[\tilde{\mathbf{r}}_{gi}^c] &= \frac{1}{n_1 + n_2} \sum_{g=1}^2 \sum_{i=1}^{n_g} \text{diag}(\hat{\boldsymbol{\lambda}})\hat{\boldsymbol{\Phi}}_{gi}^T[\hat{\mathbf{G}}_{gi} + \hat{\sigma}^2\mathbf{I}]^{-1}\hat{\boldsymbol{\Phi}}_{gi}\text{diag}(\hat{\boldsymbol{\lambda}}) \quad (2.12) \\ &\stackrel{\text{def}}{=} \hat{\mathbf{V}}, \end{aligned}$$

where $\hat{\mathbf{G}}_{gi}$ is the estimated covariance function evaluated at timestamps $\mathbf{T}_{gi} = (T_{gi1}, \dots, T_{giN_{gi}})^T$.

Given all these arguments, we propose the following test statistic for the hypothesis testing problem in Eq. (2.2),

$$T_{p,N} = [\bar{\tilde{\mathbf{r}}}_1^c - \bar{\tilde{\mathbf{r}}}_2^c]^T \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \hat{\mathbf{V}} \right]^{-1} [\bar{\tilde{\mathbf{r}}}_1^c - \bar{\tilde{\mathbf{r}}}_2^c], \quad (2.13)$$

where $\bar{\tilde{\mathbf{r}}}_g^c = \sum_{i=1}^{n_g} \hat{\mathbf{r}}_{gi}^c / n_g$, $g = 1, 2$.

It is noteworthy that, even though the proposed test statistic $T_{p,N}$ is motivated from joint Gaussian situations, we later derive that the asymptotic distribution of this statistic under H_0 is a χ_p^2 distribution, regardless of whether the random functions are Gaussian or not. For a specific significance level α , we reject H_0 , if $T_{p,N} > \chi_p^2(1 - \alpha)$, where $\chi_p^2(1 - \alpha)$ is the upper α quantile of χ_p^2 . The derivation of the asymptotic distribution is given in Section 3. Finite sample performances are demonstrated by three simulation studies and real data analyses in Section 4.

For modeling and testing procedures based on basis projections, one needs to specify the number p in practice. Popular approaches in functional data analysis literature include the cross-validation approach, the percentage of variance explained method, and other penalty criteria such as AIC and BIC. These approaches have been adopted in two-sample tests for dense functional data [1, 16, 15, 17, 21, 36, 31]. In the context of functional linear regression, [14] and [22] have proposed to combine P -values from $p \in (p_{min}, p_{max})$, where p_{min} and p_{max} grow slowly with n . Since we used the restricted maximum likelihood method [30] to estimate the eigenfunctions, the default method in our package is to choose p by cross-validation, which is described in the next subsection.

2.3. Estimation of model components

In this subsection, we discuss the estimation procedures for $\mu_{pool}(t)$, G , λ_k , $\phi_k(t)$, and σ . We propose an effective procedure to extend the restricted maximum likelihood covariance function estimation method [30] to multiple groups of sparse functional data scenarios.

Estimations of the pooled mean function $\mu_{pool}(t)$ are performed by a local linear smoothing technique, which has been used in various studies [7, 40, 30]. To be more specific, we define the local linear smoother of the pooled mean function $\mu_{pool}(t)$ by minimizing

$$\sum_{g=1}^2 \sum_{i=1}^{n_g} \sum_{j=1}^{N_{gi}} K\left(\frac{T_{gij} - t}{h_{\mu_{pool}}}\right) [Y_{gij} - \beta_0 - \beta_1(t - T_{gij})]^2 \quad (2.14)$$

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

To estimate the covariance function $G(s, t)$, we first estimate the group-wise mean functions $\mu_g(t)$, $g = 1, 2$. For each g , we define the local linear smoother of the mean function $\mu_g(t)$ by minimizing

$$\sum_{i=1}^{n_g} \sum_{j=1}^{N_{gi}} K\left(\frac{T_{gij} - t}{h_{\mu_g}}\right) [Y_{gij} - \beta_{g0}^* - \beta_{g1}^*(t - T_{gij})]^2 \quad (2.15)$$

with respect to β_{g0}^* and β_{g1}^* . Then $\hat{\mu}_g(t) = \hat{\beta}_{g0}^*(t)$. Given the mean function estimation $\hat{\mu}_g(t)$, $G(s, t)$ is estimated by modifying the restricted MLE method [30] as follows. 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 $\sum_{k=1}^r \lambda_k \phi_k(t) \phi_k(s)$ is small. Under some weak smoothness conditions on $X_g(\cdot)$, the first r eigenfunctions $\{\phi_1(t), \dots, \phi_r(t)\}$ can be modeled as,

$$\phi_k(t) = \sum_{l=1}^M d_{lk} \psi_l(t), k = 1, \dots, r, \quad (2.16)$$

where functions $\{\psi_1(t), \dots, \psi_M(t)\} \in L^2(\mathcal{T})$ are known. Let's put the coefficients in Eq. (2.16) into matrix $\mathbf{D} = (d_{lk})_{l=1, \dots, M; k=1, \dots, r}$. Based on the orthonormality of eigenfunctions, if $\{\psi_1(t), \dots, \psi_M(t)\}$ are also orthonormalized, then $\mathbf{D}^T \mathbf{D} = \mathbf{I}_r$ should hold. When the random samples are Gaussian processes and the measurement errors are Gaussian, conditional on time points, the negative log-likelihood of the data is given by

$$\begin{aligned} -\log L(\mathbf{D}, \text{diag}(\boldsymbol{\lambda}), \sigma^2) &\propto \frac{1}{2} \log |\sigma^2 \mathbf{I}_{N_{g_i}} + \boldsymbol{\Psi}_{g_i}^T \mathbf{D} \text{diag}(\boldsymbol{\lambda}) \mathbf{D}^T \boldsymbol{\Psi}_{g_i}| + \\ &\frac{1}{2} \sum_{g=1}^2 \sum_{i=1}^{n_g} \text{tr}[(\sigma^2 \mathbf{I}_{N_{g_i}} + \boldsymbol{\Psi}_{g_i}^T \mathbf{D} \text{diag}(\boldsymbol{\lambda}) \mathbf{D}^T \boldsymbol{\Psi}_{g_i})^{-1} (\mathbf{Y}_{g_i} - \hat{\boldsymbol{\mu}}_{g_i})(\mathbf{Y}_{g_i} - \hat{\boldsymbol{\mu}}_{g_i})^T], \end{aligned} \quad (2.17)$$

where $\Psi_{gi} = (\psi_{gi1}, \dots, \psi_{giM})$, with $\psi_{gil} = (\psi_l(T_{gi1}), \dots, \psi_l(T_{giN_{gi}}))^T$ for $l = 1, \dots, M$. $\hat{\mu}_g(t)$ is estimated in Eq. (2.15) and $\hat{\boldsymbol{\mu}}_{gi} = (\hat{\mu}_g(T_{gi1}), \dots, \hat{\mu}_g(T_{giN_{gi}}))^T$. 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 $\mathbf{D}^T \mathbf{D} = \mathbf{I}_r$. Then the corresponding eigenfunctions are estimated by $\hat{\phi}_k(t) = \sum_{l=1}^M \hat{d}_{lk} \psi_l(t)$.

Note that this method directly targets the eigen components. In our numerical studies, we found that the restricted MLE method produces more accurate estimates for the eigen components than the local linear smoothing method developed in [40], which first estimates the covariance and then does eigen decomposition on the discretized covariance.

To be able to obtain consistent estimates of $G(s, t)$, $\phi_k(t)$ and λ_k using restricted MLE, one has to have large enough M and r such that the true covariance is well-approximated by a member of the model space $\mathcal{M}(M, r)$. The model space consists of a class of covariance kernels \mathcal{C} , which have rank r , and whose eigenfunctions are represented in a known orthonormal basis $\{\psi_k\}_{k=1}^M$ [28]. In current implementations, the values of (M, r) are chosen by leave one curve out cross-validation [30]. Particularly, the likelihood-based cross-validation score $\text{CV}(M, r)$ to be maximized is

$$\text{CV}(M, r) = \sum_{g=1}^2 \sum_{i=1}^{n_g} \log L(\mathbf{Y}_{gi}, \mathbf{T}_{gi}, \hat{\mathbf{D}}^{-(g,i)}, \text{diag}(\hat{\boldsymbol{\lambda}}^{-(g,i)}, \hat{\sigma}^{2-(g,i)}), \quad (2.18)$$

where $L(\cdot)$ is the likelihood function of the sparse observations \mathbf{Y}_{gi} , and $\hat{\mathbf{D}}^{-(g,i)}$, $\text{diag}(\hat{\boldsymbol{\lambda}}^{-(g,i)}, \hat{\sigma}^{2-(g,i)})$ are the estimates of \mathbf{D} , $\boldsymbol{\lambda}$, σ^2 based on the data excluding the \mathbf{Y}_{gi} , i.e., the i -th curve in group g .

The next step is to select the dimension of truncation p . Intuitively, we want to identify the appropriate value such that the difference in the group mean functions is well approximated (i.e., p is sufficiently large) and non-essential difference that should be considered as random errors are excluded (i.e., p is not too large). In the literature, the dimension of truncation p for mean is often selected based on the covariance function. For instance, the percentage of variance explained method is used in [16, 15, 17, 21, 32]. Besides, the same number of projections for the mean and covariance are used in the mixed-effect model in [18]. In this paper, since we used restricted MLE to estimate the eigen components, the default method in the our package is to choose $p = r$, i.e., p is selected by the leave-one-curve cross-validation method in Eq. (2.18).

Now, by plugging in all these estimates, we achieve our test statistic $T_{p,N}$ as in Eq. (2.13).

2.4. An extension to common principal component cases

Our proposed test statistic $T_{p,N}$ is created based on the homogeneous covariance function assumption. In this section, we describe how to extend the test procedure to a certain type of heterogeneous covariance structure called the common

principal component (CPC) scenarios [11, 2, 3, 5]. Under CPC, the eigenvalues are different across different groups, while the eigenfunctions are the same. Mathematically, the g -th covariance function is $G_g(s, t) = \sum_{k=1}^{\infty} \lambda_{gk} \phi_k(s) \phi_k(t)$, with $\phi_k(t)$ being the common eigenfunctions and λ_{gk} be the group specific eigenvalues.

The specific modifications for CPC scenarios are summarized as follows. In the test statistic $T_{p,N}$, $\hat{\lambda}$ should be replaced with $\hat{\lambda}_g$. The estimation procedure also needs to be modified. In particular, $\hat{\phi}_k(t)$ is achieved by applying the restricted MLE method to the pooled data. $\hat{G}_g(s, t)$ is calculated by using data from group g alone. Then the group specific eigenvalues $\hat{\lambda}_{gk}$ is obtained by the following formula,

$$\hat{\lambda}_{gk} = \langle \hat{\phi}_k, \hat{\phi}_k^* \rangle, \text{ with } \hat{\phi}_k^*(t) = \int_{\mathcal{T}} \hat{G}_g(s, t) \hat{\phi}_k(s) ds. \quad (2.19)$$

Further extensions to cases with the general heterogeneous covariance functions (i.e., eigenvalues and eigenfunctions are both different) are not easy. This is a future topic of interest.

3. Asymptotic results

In this section, we develop the asymptotic theory of $T_{p,N}$ under both H_0 and H_a in test (2.2). The shrinkage score vector $\tilde{\mathbf{r}}_{gi}^c$ is not observable, needing to be estimated from data and this estimation error needs to be considered when deriving the asymptotic results. The projection-based method for dense functional data also needs to control errors in the estimated scores. What makes our arguments more complicated than dense data 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 the random designs.

To achieve the asymptotic results, we need the following assumptions.

1. The number of measurements N_{gi} is bounded. The observing time T_{gij} for $g = 1, 2, i = 1, \dots, n_g, j = 1, \dots, N_{gi}$ are random design points according to a density function f , with $f_l < f(t) < f_u, t \in \mathcal{T}$ for $f_l > 0$ and $f_u < \infty$.
2. $E[\Phi_{gik}^T \Sigma_{Y_{gi}}^{-1} \Phi_{gik}] < \infty$, for $k = 1, \dots, p$.
3. One can obtain consistency results,

$$\begin{aligned} \|\hat{\mu}_{pool} - \mu_{pool}\|_F &= o_p(1) \\ \|\hat{G} - G\|_F &= o_p(1) \\ |\hat{\lambda}_k - \lambda_k| &= o_p(1) \quad \text{for } k = 1, \dots, p \\ \|\hat{\phi}_k - \phi_k\|_F &= o_p(1) \quad \text{for } k = 1, \dots, p \\ |\hat{\sigma}^2 - \sigma^2| &= o_p(1) \end{aligned} \quad (3.1)$$

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

4. The leave-one-curve-out estimates of the quantities in Eq. (3.1) are close to the whole sample estimates in the sense that the difference is $o_p(1/\sqrt{n})$.

Assumption 1 is a common assumption for sparsely observed functional data. Assumption 2 requires that the variance of the shrinkage scores are bounded. Assumption 3 requires consistent estimates of model components. We estimate $\hat{\mu}_{pool}(t)$ using local linear smoothing techniques and the consistency result holds under some regular and mild conditions, which are discussed in [40, 23, 44]. The other quantities are estimated using the restricted MLE method [30]. The conditions needed for the restricted MLE to provide consistent estimates are given in Appendix A for completeness; details can be found in their original papers [30, 28]. We note that they proved the consistency of the estimated eigen components under the Gaussian assumption and discussed that their asymptotic results still hold under some relaxed conditions. Our simulation studies in Section 4.1 also show that restricted MLE has reasonable performances under non-Gaussian cases. One can use other methods to estimate the covariance, such as the local linear smoothing in [40, 23, 44], which do not require Gaussian assumptions. Therefore, Gaussian distribution is not an essential requirement for the asymptotic results developed for the proposed method. Assumption 4 is satisfied by the restricted MLE method and the local linear smoothing method, as the leave-one-curve-out estimates will be close to the whole sample estimates with difference $O_p(1/n)$. This assumption is needed to control the correlation between the estimation error of model components and observations from an individual curve.

Theorem 3.1. *Assuming assumptions 1-4 are true, assuming $\lim_{n_1, n_2 \rightarrow \infty} \frac{n_1}{n} = w$, with $w \in (0, 1)$ and $n = n_1 + n_2$. Under H_0 , we have*

$$T_{p,N} \xrightarrow{D} \chi_p^2. \quad (3.2)$$

Proof of Theorem 3.1 is postponed to Appendix B.

Theorem 3.2. *Assuming assumptions 1-4 are true, assuming $\lim_{n_1, n_2 \rightarrow \infty} \frac{n_1}{n} = w$, with $w \in (0, 1)$ and $n = n_1 + n_2$. Suppose that we have $E[\text{diag}(\boldsymbol{\lambda})\boldsymbol{\Phi}_{g_i}^T \boldsymbol{\Sigma}_{\mathbf{Y}_{g_i}}^{-1}(\boldsymbol{\mu}_{g_i} - \boldsymbol{\mu}_{pool;g_i})] \neq \mathbf{0}$ and $\text{var}(\text{diag}(\boldsymbol{\lambda})\boldsymbol{\Phi}_{g_i}^T \boldsymbol{\Sigma}_{\mathbf{Y}_{g_i}}^{-1}(\boldsymbol{\mu}_{g_i} - \boldsymbol{\mu}_{pool;g_i})) < \infty$, then*

$$T_{p,N} \xrightarrow{P} \infty. \quad (3.3)$$

$E[\text{diag}(\boldsymbol{\lambda})\boldsymbol{\Phi}_{g_i}^T \boldsymbol{\Sigma}_{\mathbf{Y}_{g_i}}^{-1}(\boldsymbol{\mu}_{g_i} - \boldsymbol{\mu}_{pool;g_i})]$ is basically some type of projection of $\mu_g(t) - \mu_{pool}(t)$. Note that the only randomness here is the random design points T_{g_i} . When $\mu_g(t) - \mu_{pool}(t) \neq 0$ and p is sufficiently large, it is unlikely that this projection scores equal to 0 on all the first p directions. Theorem 3.2 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 3.2 is in Appendix B.

Theorem 3.1 and Theorem 3.2 ensure that for a given $p < \infty$, the proposed test procedure exhibits the desired asymptotic behaviors. In particular, it controls the type I error within the pre-specified significance level and the power of the test converges in probability to 1. Due to the smoothness of functional data, the majority of information within data is captured by a finite number of projection scores. In this paper, we focus on scenarios $p < \infty$, following the convention in many functional data analysis literature.

4. Numerical experiments

4.1. Simulation studies

In this subsection, through three simulations we evaluate the performances of the proposed shrinkage-score-based Hotelling's T^2 type test, which has an asymptotic χ^2 distribution.

Simulation I-(1): In all the simulation studies, the data are always generated from model (2.1) $Y_{gij} = X_{gi}(T_{gij}) + \epsilon_{gij}$, with measurement error $\epsilon_{gij} \sim N(0, 1)$, and $X_{gi}(T_{gij}) = \mu_g(T_{gij}) + \sum_{k=1}^{\infty} \xi_{gik} \phi_k(T_{gij})$. The number of observations on each curves are $N_{gi} = N$ for $g = 1, 2; i = 1, \dots, n_g$. We let N takes two values, 4 and 8. The observing time points, $\{T_{gi1}, \dots, T_{giN}\}$, are uniformly distributed within the interval $[0, 1]$. In Simulation I, the eigenfunctions are $\phi_k(t) = \sqrt{2} \sin(k\pi t)$ for $k = 1, \dots, \infty$. The scores $\xi_{gik} \sim N(0, \lambda_k)$, with $\lambda_1 = 0.4, \lambda_2 = 0.22, \lambda_3 = 0.10, \lambda_4 = 0.04$, and $\lambda_k = 0$ for $k > 4$. The first mean function $\mu_1(t) = \sum_{k=1}^4 a_{1k} \phi_k(t)$, where $a_{1k} = 2\bar{a}_{1k} / \|\bar{\mathbf{a}}_1\|_2$, with $\bar{\mathbf{a}}_1 = (\bar{a}_{11}, \dots, \bar{a}_{14})^T$ and $\bar{a}_{1k} = (5-k)^3$. And the second mean function $\mu_2(t) = \sum_{k=1}^4 a_{2k} \phi_k(t)$, where $\bar{\mathbf{a}}_2 = \bar{\mathbf{a}}_1 - \varsigma \mathbf{d}$, with $\mathbf{d} = (1, 0.35, 0.125, 0.005)^T$ and ς quantifying the difference between the two mean functions. Note that $\varsigma = 0, 0.2, 0.3, 0.4$ correspond to the null hypothesis and three alternative hypotheses. Estimations of type I error and powers are calculated from 1000 Monte Carlo repetitions.

In this simulation study, we compare performances of the proposed asymptotic χ^2 test $T_{p,N}$ ('Shrink Hotelling'), with the distribution test in [32] ('Dist'), and the pseudo likelihood ratio test in the [37] ('pLRT-l' for linear and 'pLRT-c' for cubic). The distribution test in [32] aims to test whether $X_1(t)$ and $X_2(t)$ have the same distribution. It projects each individual curve to the eigenspace spanned by the eigenfunctions of the mixture process of $X_1(t)$ and $X_2(t)$, where a similar shrinkage is employed if the functional data is sparse. The equality of distributions of the first p projection scores is tested through the Anderson-Darling test with Bonferroni corrections. It is not a two-sample location test. Under Gaussian assumption with equal covariance settings, this test can be used to infer whether $X_1(t)$ and $X_2(t)$ have the same mean function. However, due to the non-parametric of the nature, the test is expected to have low powers. Besides, in [32], there are no theoretical results in terms of the size and the power of the proposed test. For the pLRT test, the original paper focuses on the one-sample mean function inference problem, and the generalization to two-sample mean function testing problems is described in their paper. We were able to modify

their package to perform the two-sample testing problem. We also compare the performance with several dense data methods, where the individual curves are first recovered at a dense regular grid using methods developed in [40], and then dense data tests are applied on the recovered curves. Three different dense data tests, including the projection-based normalized l_2 test [16] ('normalized l_2 '), the Globalized-F test [42] ('GF'), and the L^2 -norm-based test [41] (' l_2 '). We also consider the permutation version of the dense-recovered methods. The p-values of the permutation based dense-recovered tests ('normalized l_2 -pm', 'GF-pm', ' l_2 -pm') are calculated from 5000 Monte Carlo permutations. The Hotelling's T^2 permutation test proposed in [31] calculates a test statistic the same as the projection based normalized l_2 test and computes the p-value from the permuted samples of raw functional samples. This is the same as the permutation version of the projection-based normalized l_2 test, i.e., 'normalized l_2 -pm'.

We report the simulation results in Table 1 to 3, for different sample sizes $n_1 = n_2 = 50, 100, 200$, and different number of observations on each curve $N = 4, 8$. The significance level considered is $\alpha = 0.05$. The cross validation method mostly choose p to be 4 in the proposed method. Therefore, the true number of components $p = 4$ is used in the proposed test, the distribution test, and the dense-recovered projection method. We can see that our asymptotic χ^2 test, the distribution test and the permutation based dense-recovered tests effectively control the type I error at the pre-determined significance level. Among these tests, our proposed χ^2 test has the largest power than the other tests under all circumstances. The pLRT tests sometimes yield inflated type I errors, especially when the sample sizes and the number of observations on each curve are small. The original pLRT paper include some results for sparse data but with a relatively large number of observations on each curve, i.e., with $N = 10$. As expected, the dense-recovered methods based on asymptotic null distributions do not work, since the recovered curves from sparse functional data converge to $E[X_{gi}(t)|\mathbf{Y}_{gi}]$ instead of $X_{gi}(t)$ [40]. The permutation based versions of the dense data tests are capable of controlling the type I error, however, they suffer from lower power issues.

Simulation I-(2): In this simulation study, we assume that the mean functions are specified by a different set of basis functions other than the eigenfunctions $\{\phi_k(t) = \sqrt{2}\sin(k\pi t)\}_{k=1}^{\infty}$. In particular, the basis functions for the mean are $\{\phi'_k(t) = \sqrt{2}\cos(k\pi t)\}_{k=1}^{\infty}$. Similar to Simulation I, we also truncate at the first 4 dimensions. The mean functions are accordingly $\mu_1(t) = \sum_{k=1}^4 a_{1k}\phi'_k(t)$ and $\mu_2(t) = \sum_{k=1}^4 a_{2k}\phi'_k(t)$. The definition for a_{1k} and a_{2k} are the same as Simulation I.

The testing results for the proposed test are summarized Table 4. When implementing the proposed test procedure, we use the cross-validation approach to select the number of truncation, with p being mostly 4. As shown by Table 4, the proposed test is working reasonably in terms of size and power. The performance under the null is supported by Theorem 3.1. The test has reasonable powers is due to the fact the difference in the means on the first dimension (i.e., $\phi'_1(t)$) is captured by the second and fourth projections onto the eigenfunctions

TABLE 1

Simulation I-(1) ($n_1 = n_2 = 50$). The number of points per curve is $N = 4, 8$. Sample sizes are $n_1 = n_2 = 50$. The significance level is $\alpha = 0.05$. Consider nine tests: the proposed asymptotic χ^2 test ('Shrink Hotelling'), the distribution test ('Dist'), the linear and cubic pLRT test ('pLRT-l', 'pLRT-c'), the dense-recovered tests ('normalized l_2 ', 'GF', ' l_2 '), and the dense-recovered tests based on 5000 Monte Carlo permutations ('normalized l_2 -pm', 'GF-pm', ' l_2 -pm'). The dense-recovered tests and their permutation versions are conducted on the recovered curves from the sparsely observed data. Highest powers among 'Shrink Hotelling', 'Dist', 'normalized l_2 -pm', 'GF-pm', and ' l_2 -pm' are in bold font. Type I error ($\varsigma = 0$) and powers ($\varsigma = 0.2, 0.3, 0.4$) are from 1000 repetitions.

Test	N = 4				N = 8			
	$\varsigma = 0$	$\varsigma = 0.2$	$\varsigma = 0.3$	$\varsigma = 0.4$	$\varsigma = 0$	$\varsigma = 0.2$	$\varsigma = 0.3$	$\varsigma = 0.4$
Shrink Hotelling	0.053	0.135	0.267	0.489	0.045	0.198	0.384	0.656
Dist	0.054	0.131	0.238	0.419	0.051	0.170	0.334	0.583
pLRT-l	0.123	0.215	0.358	0.542	0.089	0.277	0.496	0.700
pLRT-c	0.116	0.226	0.357	0.556	0.104	0.306	0.530	0.737
normalized l_2	0.954	0.961	0.987	0.993	0.895	0.923	0.967	0.989
normalized l_2 -pm	0.053	0.062	0.122	0.159	0.052	0.073	0.087	0.203
GF	0.963	0.975	0.988	0.995	0.842	0.922	0.968	0.995
GF-pm	0.039	0.043	0.100	0.132	0.042	0.084	0.159	0.248
l_2	0.909	0.949	0.969	0.989	0.590	0.789	0.910	0.980
l_2 -pm	0.043	0.060	0.106	0.164	0.052	0.100	0.195	0.287

(i.e., $\phi_2(t)$ and $\phi_4(t)$), although the projection onto the first eigenfunction (i.e., $\phi_1(t)$) cannot capture it.

Simulation II: The second simulation focuses on exploring the performance of the proposed test under non-Gaussian circumstances, and we also include the results from the distribution test as a comparison. Simulation settings are the same as Simulation I, except that ξ_{gik} , $g = 1, 2$; $i = 1, \dots, n_g$; $k = 1, \dots, p$ 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 calculated from 1000 repetitions in Table 5, we can see that the proposed test and the distribution test still produce valid results for this non-Gaussian situation. The proposed χ^2 test has larger powers under all circumstances.

Simulation III: The data are generated from the same model as in simulation I, except that we have used the first 15 components to generate the data. The corresponding eigenvalues are 0.40, 0.22, 0.10, $4e-2$, $2e-2$, $8e-3$, $4e-3$, $1.6e-3$, $4e-4$, $2.4e-4$, $8e-5$, $3.2e-5$, $1.2e-5$, $2.8e-6$, $4e-7$. As for the mean functions, we set $\mu_g(t) = \sum_{k=1}^{15} a_{gk} \phi_k(t)$, where $a_{gk} = 2\bar{a}_{gk} / \|\bar{\mathbf{a}}_g\|_2$, with $\bar{\mathbf{a}}_g = (\bar{a}_{g1}, \dots, \bar{a}_{g15})^T$. Note that $\bar{a}_{1k} = (16 - k)^3$ and $\bar{\mathbf{a}}_2 = \bar{\mathbf{a}}_1 - \delta \mathbf{e}$ with $\delta = 0, 0.2, 0.3, 0.4$, and $\mathbf{e} = (1, 0.35, 0.125, 5e-3, 4e-3, 2e-3, 1e-3, 6e-4, 3e-4, 2e-4, 8e-5, 3e-5, 1e-5, 7e-6, 1e-6)^T$.

To evaluate the actual performance of the proposed method, we exploit the two-sample restricted MLE estimate and the leave-one-curve-out cross-validation based selection of p in Section 2.3 to construct the proposed asymptotic χ^2 test statistic. The size and power over 1000 Monte Carlo repetitions are summarized in Table 6. We can see our proposed procedure works well. It can

TABLE 2

Simulation I-(1) ($n_1 = n_2 = 100$). The number of points per curve is $N = 4, 8$. Sample sizes are $n_1 = n_2 = 100$. The significance level is $\alpha = 0.05$. Consider nine tests: the proposed asymptotic χ^2 test ('Shrink Hotelling'), the distribution test ('Dist'), the linear and cubic pLRT test ('pLRT-l', 'pLRT-c'), the dense-recovered tests ('normalized l_2 ', 'GF', ' l_2 '), and the dense-recovered tests based on 5000 Monte Carlo permutations ('normalized l_2 -pm', 'GF-pm', ' l_2 -pm'). The dense-recovered tests and their permutation versions are conducted on the recovered curves from the sparsely observed data. Highest powers among 'Shrink Hotelling', 'Dist', 'normalized l_2 -pm', 'GF-pm', and ' l_2 -pm' are in bold font. Type I error ($\zeta = 0$) and powers ($\zeta = 0.2, 0.3, 0.4$) are from 1000 repetitions.

Test	N = 4				N = 8			
	$\zeta = 0$	$\zeta = 0.2$	$\zeta = 0.3$	$\zeta = 0.4$	$\zeta = 0$	$\zeta = 0.2$	$\zeta = 0.3$	$\zeta = 0.4$
Shrink Hotelling	0.045	0.257	0.515	0.821	0.050	0.352	0.719	0.931
Dist	0.057	0.215	0.466	0.771	0.058	0.291	0.648	0.916
pLRT-l	0.095	0.347	0.564	0.803	0.081	0.449	0.758	0.944
pLRT-c	0.110	0.337	0.581	0.814	0.087	0.463	0.801	0.966
normalized l_2	0.986	0.979	0.995	1.000	0.913	0.978	0.993	1.000
normalized l_2 -pm	0.051	0.065	0.096	0.199	0.055	0.079	0.148	0.276
GF	0.984	0.992	0.998	1.000	0.872	0.978	0.995	1.000
GF-pm	0.046	0.075	0.149	0.204	0.039	0.083	0.212	0.349
l_2	0.918	0.968	0.993	0.999	0.602	0.907	0.988	0.999
l_2 -pm	0.053	0.090	0.164	0.242	0.047	0.116	0.250	0.475

control the type I error at the pre-determined significance level. The powers are increasing with larger sample sizes and the magnitude of difference between the two mean curves.

To study the sensitivity of the proposed test to the selection of p , we fix p at three values $p = 3, 4, 5$, which respectively captures 99.1%, 99.5%, and 99.7% of the difference among the mean functions. The results are summarized in Table 7. It can be seen that the size is well controlled at the pre-specified significance level $\alpha = 0.05$ for $p = 3, 4, 5$. This is expected as the asymptotic null distribution is valid for any p , as described in Theorem 3.1. Besides, for each p , the power of the proposed test increases with the increase of the sample size and the magnitude of the difference between the two mean curves. This observation adheres to the arguments in Theorem 3.2.

In Table 7, the power under different values of p is overall comparable, with some moderate variation across different p . These results numerically justify that the proposed test procedure is insensitive to the dimension of truncation when the dimension p is sufficiently large. In general, the power for $p = 3$ is slightly larger than the other two choices of p . Intuitively, unnecessarily large values of p tend to introduce the non-essential signal that should be considered as random noises into the test. The results in Table 6 is close to those with respect to $p = 3$ in Table 7. This is because the selected p from the cross-validation approach is mostly 3 in this simulation study.

4.2. Application to CD4 count data

Now we apply the proposed asymptotic χ^2 test to a CD4 count data set from the AIDS Clinical Trials Group (ACTG) 193 A study, which aims at comparing

TABLE 3

Simulation I-(1) ($n_1 = n_2 = 200$). The number of points per curve is $N = 4, 8$. Sample sizes are $n_1 = n_2 = 200$. The significance level is $\alpha = 0.05$. Consider nine tests: the proposed asymptotic χ^2 test ('Shrink Hotelling'), the distribution test ('Dist'), the linear and cubic pLRT test ('pLRT-l', 'pLRT-c'), the dense-recovered tests ('normalized l_2 ', 'GF', ' l_2 '), and the dense-recovered tests based on 5000 Monte Carlo permutations ('normalized l_2 -pm', 'GF-pm', ' l_2 -pm'). The dense-recovered tests and their permutation versions are conducted on the recovered curves from the sparsely observed data. Highest powers among 'Shrink Hotelling', 'Dist', 'normalized l_2 -pm', 'GF-pm', and ' l_2 -pm' are in bold font. Type I error ($\varsigma = 0$) and powers ($\varsigma = 0.2, 0.3, 0.4$) are from 1000 repetitions.

Test	N = 4				N = 8			
	$\varsigma = 0$	$\varsigma = 0.2$	$\varsigma = 0.3$	$\varsigma = 0.4$	$\varsigma = 0$	$\varsigma = 0.2$	$\varsigma = 0.3$	$\varsigma = 0.4$
Shrink Hotelling	0.045	0.513	0.877	0.992	0.051	0.664	0.940	0.999
Dist	0.049	0.455	0.839	0.982	0.056	0.617	0.918	0.998
pLRT-l	0.097	0.523	0.821	0.966	0.065	0.680	0.958	0.998
pLRT-c	0.101	0.515	0.843	0.980	0.069	0.734	0.978	0.999
normalized l_2	0.974	0.999	1.000	1.000	0.941	0.998	0.999	1.000
normalized l_2 -pm	0.048	0.079	0.178	0.351	0.073	0.115	0.270	0.584
GF	0.989	1.000	1.000	1.000	0.916	0.998	0.998	1.000
GF-pm	0.042	0.102	0.180	0.298	0.050	0.146	0.261	0.476
l_2	0.929	1.000	0.999	1.000	0.594	0.973	0.984	1.000
l_2 -pm	0.044	0.111	0.218	0.365	0.050	0.214	0.446	0.743

TABLE 4

Simulation I-(2). The number of points per curve is $N = 4, 8$. Sample sizes are $n_1 = n_2 = 50, 100, 200$. The significance level is $\alpha = 0.05$. The number of components p is selected by the cross-validation technique. Type I error ($\delta = 0$) and powers ($\delta = 0.2, 0.3, 0.4$) of our proposed test are calculated from 1000 repetitions.

n	N = 4				N = 8			
	$\delta = 0$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.4$	$\delta = 0$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.4$
50	0.049	0.126	0.255	0.445	0.053	0.204	0.397	0.651
100	0.051	0.243	0.549	0.809	0.047	0.369	0.711	0.924
200	0.053	0.522	0.888	0.985	0.051	0.669	0.953	1.000

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 the 40 weeks.

We model the data from the 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 significantly different within the 40-weeks period, we can understand whether the effectiveness of these two treatments is different over time. As suggested by previous literature [10, 39], we transform the original CD4 count using function $\log(x + 1)$. The Spaghetti plots of the

TABLE 5

Simulation II. The number of points per curve is $N = 4, 8$. Sample sizes are $n_1 = n_2 = 50, 100, 200$. The significance level is $\alpha = 0.05$. Consider two tests: our asymptotic χ^2 test ('Shrink Hotelling') and the distribution test ('Dist'). Type I error ($\zeta = 0$) and powers ($\zeta = 0.2, 0.3, 0.4$) are from 1000 repetitions.

n	Test	$N = 4$				$N = 8$			
		$\zeta = 0$	$\zeta = 0.2$	$\zeta = 0.3$	$\zeta = 0.4$	$\zeta = 0$	$\zeta = 0.2$	$\zeta = 0.3$	$\zeta = 0.4$
50	Shrink Hotelling	0.054	0.144	0.256	0.483	0.053	0.220	0.401	0.657
	Dist	0.049	0.120	0.203	0.421	0.042	0.171	0.315	0.569
100	Shrink Hotelling	0.048	0.254	0.531	0.806	0.051	0.353	0.701	0.933
	Dist	0.038	0.209	0.485	0.757	0.065	0.281	0.620	0.899
200	Shrink Hotelling	0.044	0.473	0.862	0.969	0.051	0.660	0.972	1.000
	Dist	0.049	0.431	0.808	0.970	0.059	0.580	0.929	1.000

TABLE 6

Simulation III (Choose p using cross-validation). The number of points per curve is $N = 4, 8$. Sample sizes are $n_1 = n_2 = 50, 100, 200$. The significance level is $\alpha = 0.05$. The number of components p is selected by the cross-validation technique. Type I error ($\delta = 0$) and powers ($\delta = 0.2, 0.3, 0.4$) of our proposed test are calculated from 1000 repetitions.

n	$N = 4$				$N = 8$			
	$\delta = 0$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.4$	$\delta = 0$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.4$
50	0.051	0.145	0.264	0.461	0.048	0.210	0.374	0.653
100	0.052	0.233	0.519	0.724	0.050	0.349	0.669	0.914
200	0.051	0.490	0.882	0.990	0.047	0.659	0.975	1.000

transformed data of groups A and B can be found in Figure 1. The first three estimated functional principal components for both groups are shown on the right panel of Figure 2. As shown by Figure 2, 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. 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 in Section 2.4, the corresponding p-values are $1.11e-61, 7.73e-61, 4.61e-92, 1.47e-97$. 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. Note that the selected p from cross-validation is 3, i.e., the result of the proposed test procedure based on the cross-validation method of choosing the truncation dimension is p -value = 0.0779 when assuming homogeneous covariance functions, and $7.73e-61$ under the CPC assumption. Given the estimated mean functions provided on the left panel of Figure 2, it can be seen that the therapy used by group B tends to be more effective in helping advanced AIDS patients to recover. We also implemented the permutation-based dense-recovered tests. These tests do not detect a significant difference between the two groups. We have seen that these tests have low power in simulations. We applied the proposed method on different sample sizes, i.e., 25%, 50%, 75% and 100% of subjects in each group. The proposed method can always detect the signal for 75% and 100%, and the power is lower for 25%, and 50%.

TABLE 7

Simulation III (Sensitivity study of p). The number of points per curve is $N = 4, 8$. Sample sizes are $n_1 = n_2 = 50, 100, 200$. The significance level is $\alpha = 0.05$. The number of components p is fixed as 3, 4 or 5. Type I error ($\delta = 0$) and powers ($\delta = 0.2, 0.3, 0.4$) of our proposed test are calculated from 1000 repetitions.

n	p	N = 4				N = 8			
		$\delta = 0$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.4$	$\delta = 0$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.4$
50	3	0.044	0.149	0.275	0.460	0.045	0.214	0.350	0.644
	4	0.041	0.142	0.260	0.441	0.050	0.172	0.337	0.633
	5	0.057	0.153	0.282	0.458	0.056	0.204	0.327	0.624
100	3	0.052	0.226	0.491	0.707	0.044	0.340	0.653	0.906
	4	0.049	0.219	0.507	0.703	0.044	0.306	0.637	0.892
	5	0.044	0.235	0.473	0.688	0.048	0.294	0.637	0.872
200	3	0.046	0.481	0.866	0.984	0.051	0.643	0.964	1.000
	4	0.054	0.459	0.863	0.978	0.048	0.599	0.948	1.000
	5	0.052	0.460	0.844	0.978	0.044	0.583	0.948	0.998

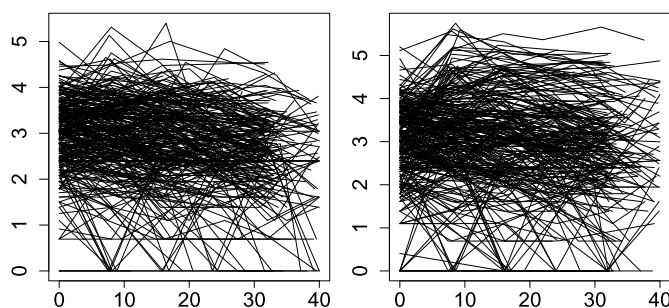


FIG 1. Plot of group A is on the left panel; Plot of group B is on the right panel

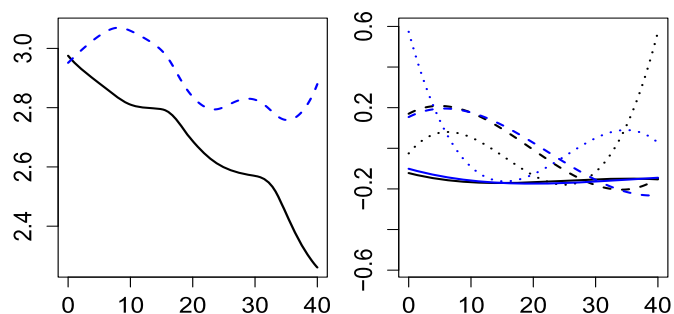


FIG 2. Left: Mean functions of two groups (group A: black; group B: blue); Right: Estimated eigenfunctions of two groups (group A: black; group B: blue)

4.3. Application to eBay online auction data

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 a 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. These real-time price trajectories come out of this system are often referred to 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 <http://www.ebay.com/>. eBay auction data for all items are appropriately stored by eBay 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 [19]. 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 item is given in Figure 3, which are not small. However, as shown in Figure 4, where 9 randomly selected auctions in the 7 days group are visualized, the bids are extremely irregular. There are big gaps among individual live bid trajectories.

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. [29, 24, 38, 20] also analyze ebay online auction data from functional

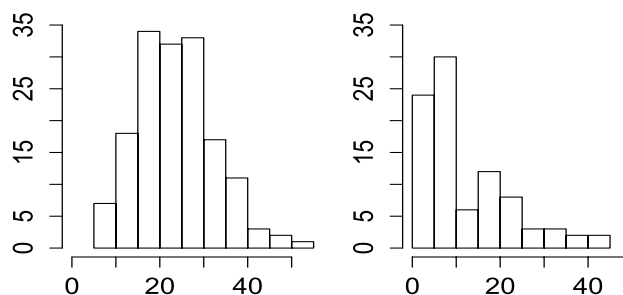


FIG 3. Left: histogram of number of observation within each item in the 3 days group; Right: histogram of number of observation within each item in the 7 days group

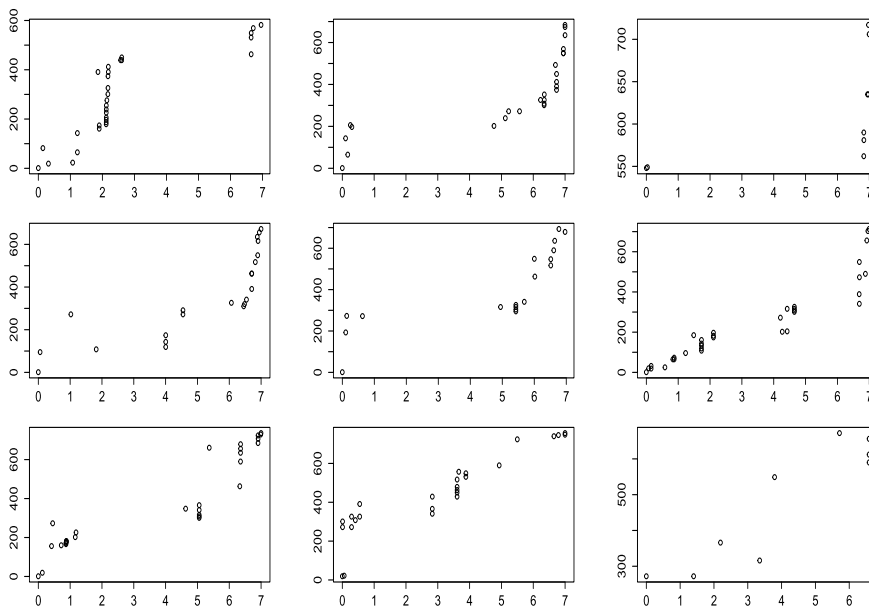


FIG 4. Bids for 9 randomly chosen Palm M515 in the 7 days group.

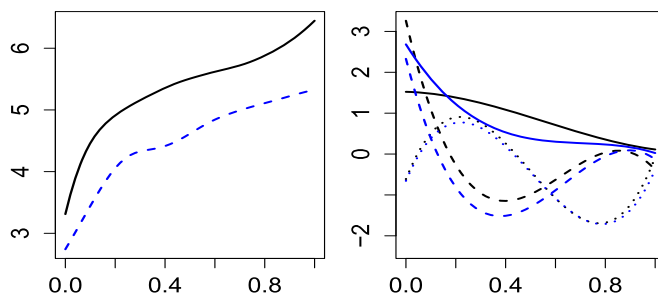


FIG 5. Left: Mean functions of two groups (3 days: blue; 7 days: black); Right: Estimated eigenfunctions of two groups (3 days: blue; 7days: black)

perspectives. Following previous papers [29, 20], we transform live bid data into a 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 of Figure 5, 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 of Figure 5. From the graph, we can see that the price of the 7 days group (black) is larger than that of the 3 days group (blue) over the entire time domain. The permutation based dense-recovered tests are also implemented but do not detect a significant difference between the two groups.

5. Conclusions and Discussions

In this paper, we developed an asymptotic χ^2 test for detecting differences among the mean functions of two independent stochastic processes with homogeneous covariance functions, when only a few irregularly spaced measurements are given for each subject. The proposed test statistic is constructed based on the shrinkage score vector, a counterpart of the one-sample PACE designed especially for multiple-sample mean testing problems. Note that as shown by the theoretical derivations, our proposed test does not require Gaussian random processes, although the test is motivated by Gaussian.

We described in detail the numerical implementation of the proposed test procedure. To benchmark the proposed test, we considered four baselines, including two existing tests that apply to sparse data settings (i.e., the non-parametric distribution test and the pseudo-likelihood ratio test), and two promising combinational approaches (i.e., dense tests equipped with PACE as a pre-processing step, and the permutation dense test with PACE). The numerical experiments demonstrated the superior performance of our proposed test, in comparison with these methods. In particular, the pLRT test yielded inflated type I errors when the sample size and the number of observations per curve are small. The dense data methods with the help of PACE to recover the individual curves did not work, since the recovered curves from sparse functional data do not converge to the actual random curves. The non-parametric distribution test performed reasonably, however, it had low powers due to the non-parametric nature.

We proposed two useful extensions of the proposed test. One extension aims to generalize the test to a certain type of heterogeneous covariance scenario called the common principal component structure, where the eigenvalues can be different across groups. The other extension that handles scenarios where more than two groups are considered in the hypothesis testing problem (see Appendix C).

The test procedure developed in the paper cannot handle cases with general heterogeneous covariance functions, i.e. when the eigenvalues and eigenfunctions are both different across groups. This is an important piece of future work to study.

Appendix A: Conditions for restricted MLE consistency

The estimators \hat{G} , $\hat{\phi}_k$, $\hat{\lambda}_k$, $k = 1, \dots, r$, obtained from restricted MLE utilizing model space $\mathcal{M}(M, r)$, are consistent when the following assumptions hold [28].

1. $X_1(t)$ and $X_2(t)$ are two independent Gaussian processes.
2. The r largest eigenvalues of $G(s, t)$ satisfy. (i) There exists a constant $a_1 < \infty$, such that $a_1 \geq \lambda_1 > \lambda_2 > \dots > \lambda_r > \lambda_{r+1}$; (ii) There exists a constant $a_2 < \infty$, such that $\max_{1 \leq k \leq r} (\lambda_k - \lambda_{k+1})^{-1} \leq a_2$.
3. The common eigenfunctions $\{\phi_k(t)\}_{k=1}^r$ are four times continuously differentiable and satisfy for some $0 < A_0 < \infty$

$$\max_{1 \leq k \leq r} \sup_{t \in \mathcal{T}} |\phi_k^{(4)}(t)| \leq A_0. \quad (\text{A.1})$$

4. For each g, i , $\{T_{gij}, j = 1, \dots, N_{gi}\}$ are i.i.d samples from a distribution f within the time domain \mathcal{T} , where f is a bounded function and it satisfies $f_l \leq f(x) \leq f_u$ for all $t \in \mathcal{T}$, where $0 < f_l \leq f_u < \infty$.
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 is the number of orthonormalized cubic B -spline basis used to represent the eigenfunctions $\phi_k(t)$.
7. Define the approximation error for an optimal estimator in model space $\mathcal{M}(M, r)$ as $\beta_n = \max_{1 \leq g \leq 2, 1 \leq i \leq N_{gi}} \frac{1}{N_{gi}} \|\Sigma_{gi} - \Sigma_{gi}^*\|_F$, where $\Sigma_{gi} = G_{gi} + \sigma^2 I$, and Σ_{gi}^* is the best estimator in the model space $\mathcal{M}(M, r)$. We have $\bar{N}\beta_n = O(\sqrt{\frac{M \log n}{n}})$.

Condition 1 is a working assumption. Note that the derivation of the asymptotic distribution of the proposed test statistic $T_{p,N}$ does not need Gaussian distribution as long as the estimates listed in Assumption 3 are consistent. Conditions 2-5 are common assumptions for sparse functional data. 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 Assumption 3 and have $\bar{N}^2 \alpha_n^2 = o(1)$ satisfied, then the current proofs for Theorem 3.1 and Theorem 3.2 can go through. Given that we use restricted MLE to estimate G , 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. Conditions 6-7 put conditions on (M, r) such that the true covariance can be approximated well by a member in the model space $\mathcal{M}(M, r)$. The restricted MLE method developed in [30] provides an estimate of σ^2 , which we call $\hat{\sigma}^2$. They do not directly provide a theorem for the consistency of $\hat{\sigma}^2$, but rather they note that all the consistency results for G , ϕ and λ hold when $\hat{\sigma}^2$ is used.

Appendix B: Proofs

Proof for Theorem 1:

The goal is to show that $T_{p,N} \rightarrow \chi_p^2$ in distribution, under H_0 . Denote $T_{p,N}^* = (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c)^T [(\frac{1}{n_1} + \frac{1}{n_2})\mathbf{V}]^{-1} (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c)$, where $\mathbf{V} = \frac{\sum_{g=1}^2 \sum_{i=1}^{n_g} \text{diag}(\boldsymbol{\lambda}) \boldsymbol{\Phi}_{gi}^T \boldsymbol{\Sigma}_{\mathbf{Y}_{gi}^c}^{-1} \boldsymbol{\Phi}_{gi} \text{diag}(\boldsymbol{\lambda})}{n_1 + n_2}$. According to the Central Limit Theorem and the Slutsky's theorem, $T_{p,N}^*$ is asymptotically χ_p^2 distributed under H_0 . Consequently, proving Theorem 3.1 is basically proving $T_{p,N} = T_{p,N}^* + o_p(1)$.

To simplify our notations, let's denote $\mathbf{S}_{gi} = \text{diag}(\boldsymbol{\lambda}) \boldsymbol{\Phi}_{gi}^T \boldsymbol{\Sigma}_{\mathbf{Y}_{gi}^c}^{-1} \boldsymbol{\Phi}_{gi} \text{diag}(\boldsymbol{\lambda})$ and $\hat{\mathbf{S}}_{gi} = \text{diag}(\hat{\boldsymbol{\lambda}}) \hat{\boldsymbol{\Phi}}_{gi}^T [\hat{G}_{gi} + \hat{\sigma}_g^2 \mathbf{I}]^{-1} \hat{\boldsymbol{\Phi}}_{gi} \text{diag}(\hat{\boldsymbol{\lambda}})$. Then we can see that $\hat{\mathbf{V}} = \frac{\hat{\mathbf{S}}_{11} + \dots + \hat{\mathbf{S}}_{1n_1} + \hat{\mathbf{S}}_{21} + \dots + \hat{\mathbf{S}}_{2n_2}}{n}$ and $\mathbf{V} = \frac{\mathbf{S}_{11} + \dots + \mathbf{S}_{1n_1} + \mathbf{S}_{21} + \dots + \mathbf{S}_{2n_2}}{n}$. The absolute value of $T_{p,N} - T_{p,N}^*$ can be decomposed as follows.

$$\begin{aligned} & |T_{p,N} - T_{p,N}^*| \\ &= \left\| (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c)^T [(\frac{1}{n_1} + \frac{1}{n_2})\hat{\mathbf{V}}]^{-1} (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c) - (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c)^T [(\frac{1}{n_1} + \frac{1}{n_2})\mathbf{V}]^{-1} (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c) \right\|_2 \\ &\leq \left\| (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c)^T [(\frac{1}{n_1} + \frac{1}{n_2})\hat{\mathbf{V}}]^{-1} - [(\frac{1}{n_1} + \frac{1}{n_2})\mathbf{V}]^{-1} \right\|_2 (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c) \\ &\quad + \left\| (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c - \bar{\mathbf{r}}_1^c + \bar{\mathbf{r}}_2^c)^T [(\frac{1}{n_1} + \frac{1}{n_2})\hat{\mathbf{V}}]^{-1} (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c) \right\|_2 \\ &\quad + \left\| (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c)^T [(\frac{1}{n_1} + \frac{1}{n_2})\hat{\mathbf{V}}]^{-1} (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c - \bar{\mathbf{r}}_1^c + \bar{\mathbf{r}}_2^c) \right\|_2 \\ &\stackrel{\text{def}}{=} A_1 + A_2 + A_3. \end{aligned} \tag{B.1}$$

Given Eq. (B.1), proving Theorem 3.1 can be accomplished by proving $A_1 = o_p(1)$, $A_2 = o_p(1)$, and $A_3 = o_p(1)$ hold.

We start with showing $A_1 = o_p(1)$. Under H_0 , i.e., $\mu_1(t) = \mu_2(t) = \mu_{pool}(t)$, we have

$$\left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c) \sim N_p(\mathbf{0}, \text{cov}(\bar{\mathbf{r}}_{gi}^c)). \tag{B.2}$$

So we have

$$\left\| \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c) \right\|_2 = O_p(1). \tag{B.3}$$

Under Assumption 3, we have

$$\begin{aligned} \|\hat{\mathbf{V}} - \mathbf{V}\|_2 &= O_p(\bar{N} \{ \sum_{k=1}^p |\hat{\lambda}_k - \lambda_k| + \sum_{k=1}^p \|\hat{\phi}_k - \phi_k\|_F + \|\hat{G} - G\|_F + |\hat{\sigma}^2 - \sigma^2| \}) \\ &= o_p(1). \end{aligned} \tag{B.4}$$

Combine Eq. (B.3) and (B.4) together, we have

$$\begin{aligned} A_1 &\leq \left\| \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c) \right\|_2^2 \left\| \hat{\mathbf{V}}^{-1} - \mathbf{V}^{-1} \right\|_2 \\ &\leq O_p(1) \times \frac{\|\mathbf{V}^{-1}\|_2}{[\|\mathbf{V}^{-1}\|_2 \|\hat{\mathbf{V}} - \mathbf{V}\|_2]^{-1} - 1} = o_p(1). \end{aligned} \tag{B.5}$$

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

$$\begin{aligned} A_2 &\leq \left\| \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-1/2} (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c) \right\|_2 \left\| \hat{\mathbf{V}}^{-1} \right\|_2 \left\| \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-1/2} (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c - \bar{\mathbf{r}}_1^c + \bar{\mathbf{r}}_2^c) \right\|_2 \\ &\leq O_p(1) O_p(1) \left\| \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-1/2} (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c - \bar{\mathbf{r}}_1^c + \bar{\mathbf{r}}_2^c) \right\|_2. \end{aligned} \quad (\text{B.6})$$

In order to prove $A_2 = o_p(1)$, we need to show $\left\| \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-1/2} (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c - \bar{\mathbf{r}}_1^c + \bar{\mathbf{r}}_2^c) \right\|_2 = o_p(1)$. We want to write out the explicit formula of $\bar{\mathbf{r}}_g^c - \bar{\mathbf{r}}_g^c$:

$$\begin{aligned} \hat{\mathbf{r}}_{gi}^c &= \text{diag}(\hat{\boldsymbol{\lambda}}) \hat{\boldsymbol{\Phi}}_{gi}^T \hat{\boldsymbol{\Sigma}}_{\mathbf{Y}_{gi}}^{-1} (\mathbf{Y}_{gi} - \hat{\boldsymbol{\mu}}_{pool;gi}) \\ &\stackrel{\text{def}}{=} \text{diag}(\hat{\boldsymbol{\lambda}}) \hat{\mathbf{Q}}_{gi} (\mathbf{Y}_{gi} - \hat{\boldsymbol{\mu}}_{pool;gi}) \\ &= \bar{\mathbf{r}}_{gi}^c + (\text{diag}(\hat{\boldsymbol{\lambda}}) - \text{diag}(\boldsymbol{\lambda})) \mathbf{Q}_{gik} (\mathbf{Y}_{gi} - \boldsymbol{\mu}_{pool;gi}) + \text{diag}(\hat{\boldsymbol{\lambda}}) (\hat{\mathbf{Q}}_{gik} - \mathbf{Q}_{gik}) (\mathbf{Y}_{gi} - \boldsymbol{\mu}_{pool;gi}) \\ &\quad + \text{diag}(\hat{\boldsymbol{\lambda}}) \hat{\mathbf{Q}}_{gik} (\boldsymbol{\mu}_{pool;gi} - \hat{\boldsymbol{\mu}}_{pool;gi}), \end{aligned} \quad (\text{B.7})$$

i.e.,

$$\begin{aligned} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-1/2} (\bar{\mathbf{r}}_g^c - \bar{\mathbf{r}}_g^c) &= \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-1/2} (\text{diag}(\hat{\boldsymbol{\lambda}}) - \text{diag}(\boldsymbol{\lambda})) \frac{1}{n_g} \sum_{i=1}^{n_g} \mathbf{Q}_{gi} (\mathbf{Y}_{gi} - \boldsymbol{\mu}_{pool;gi}) \\ &\quad + \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-1/2} \text{diag}(\hat{\boldsymbol{\lambda}}) \frac{1}{n_g} \sum_{i=1}^{n_g} (\hat{\mathbf{Q}}_{gi} - \mathbf{Q}_{gi}) (\mathbf{Y}_{gi} - \boldsymbol{\mu}_{pool;gi}) \\ &\quad + \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-1/2} \text{diag}(\hat{\boldsymbol{\lambda}}) \frac{1}{n_g} \sum_{i=1}^{n_g} \hat{\mathbf{Q}}_{gi} (\boldsymbol{\mu}_{pool;gi} - \hat{\boldsymbol{\mu}}_{pool;gi}) \\ &\stackrel{\text{def}}{=} B_{g1} + B_{g2} + B_{g3}. \end{aligned} \quad (\text{B.8})$$

According to Eq. (B.8), in order to ensure that $\left\| \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-1/2} (\bar{\mathbf{r}}_1^c - \bar{\mathbf{r}}_2^c - \bar{\mathbf{r}}_1^c + \bar{\mathbf{r}}_2^c) \right\|_2 = o_p(1)$, 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 ,

$$\begin{aligned} \|B_{g1}\|_2 &\leq \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-1/2} \left\| \text{diag}(\hat{\boldsymbol{\lambda}}) - \text{diag}(\boldsymbol{\lambda}) \right\|_2 \left\| \frac{1}{n_g} \sum_{i=1}^{n_g} \mathbf{Q}_{gi} (\mathbf{Y}_{gi} - \boldsymbol{\mu}_{pool;gi}) \right\|_2 \\ &= \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-1/2} o_p(1) \sqrt{\sum_{k=1}^p \left(\frac{1}{n_g} \sum_{i=1}^{n_g} \mathbf{Q}_{gik} (\mathbf{Y}_{gi} - \boldsymbol{\mu}_{pool;gi}) \right)^2} \\ &= \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{-1/2} o_p(1) \sqrt{O_P\left(\frac{1}{n_g}\right)} = o_P(1), \end{aligned} \quad (\text{B.9})$$

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

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

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 Eq. (B.9), we have

$$\|B_{11} - B_{21}\|_2 \leq \|B_{11}\|_2 + \|B_{21}\|_2 = o_p(1). \tag{B.10}$$

Now let's consider $B_{12} - B_{22}$. Using Assumption 4, we have

$$\begin{aligned} \|B_{g2}\|_2 &\leq \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} \|\text{diag}(\hat{\boldsymbol{\lambda}})\|_2 \left\| \frac{1}{n_g} \sum_{i=1}^{n_g} (\hat{\mathbf{Q}}_{gi}^{(-i)} - \mathbf{Q}_{gi})(\mathbf{Y}_{gi} - \boldsymbol{\mu}_{pool;gi}) + o_p(1/\sqrt{n_g}) \right\|_2 \\ &= \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} O_p(1) o_p(1/\sqrt{n_g}) = o_p(1), \end{aligned} \tag{B.11}$$

where $\hat{\mathbf{Q}}_{gi}^{(-i)}$ is the leave i -th curve out estimates. The reason is that $E[(\hat{\mathbf{Q}}_{gi}^{(-i)} - \mathbf{Q}_{gi})(\mathbf{Y}_{gi} - \boldsymbol{\mu}_{pool;gi})] = E[(\hat{\mathbf{Q}}_{gi}^{(-i)} - \mathbf{Q}_{gi})]E[(\mathbf{Y}_{gi} - \boldsymbol{\mu}_{pool;gi})] = \mathbf{0}$, and for each $k = 1, \dots, p$, $\text{Var}[(\hat{\mathbf{Q}}_{gik}^{(-i)} - \mathbf{Q}_{gik})(\mathbf{Y}_{gi} - \boldsymbol{\mu}_{pool;gi})] = o_p(1)$.

Based on Eq. (B.11), we have

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

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

For the last term in Eq. (B.8), $B_{13} - B_{23}$. 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 $E_t[\hat{\mathbf{Q}}_{1i}(\boldsymbol{\mu}_{pool;1i} - \hat{\boldsymbol{\mu}}_{pool;1i})] = E_t[\hat{\mathbf{Q}}_{2i}(\boldsymbol{\mu}_{pool;2i} - \hat{\boldsymbol{\mu}}_{pool;2i})]$. Let's denote $E_t[\hat{\mathbf{Q}}_{gi}(\boldsymbol{\mu}_{pool;gi} - \hat{\boldsymbol{\mu}}_{pool;gi})]$ as \mathbf{E}_g , and we know $\mathbf{E}_1 = \mathbf{E}_2$, then

$$\begin{aligned} \|B_{31} - B_{32}\|_2 &\leq \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} \|\text{diag}(\hat{\boldsymbol{\lambda}})\|_2 \left\| \frac{1}{n_1} \sum_{i=1}^{n_1} \hat{\mathbf{Q}}_{1i}(\boldsymbol{\mu}_{pool;1i} - \hat{\boldsymbol{\mu}}_{pool;1i}) - \frac{1}{n_2} \sum_{i=1}^{n_2} \hat{\mathbf{Q}}_{2i}(\boldsymbol{\mu}_{pool;2i} - \hat{\boldsymbol{\mu}}_{pool;2i}) \right\|_2 \\ &\leq \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} O_p(1) \left\| \frac{1}{n_1} \sum_{i=1}^{n_1} [\hat{\mathbf{Q}}_{1i}(\boldsymbol{\mu}_{pool;1i} - \hat{\boldsymbol{\mu}}_{pool;1i}) - \mathbf{E}_1] - \frac{1}{n_2} \sum_{i=1}^{n_2} [\hat{\mathbf{Q}}_{2i}(\boldsymbol{\mu}_{pool;2i} - \hat{\boldsymbol{\mu}}_{pool;2i}) - \mathbf{E}_2] \right\|_2 \\ &= \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} O_p(1) o_p(1/\sqrt{n_1} + 1/\sqrt{n_2}) = o_p(1). \end{aligned} \tag{B.13}$$

Combine Eq. (B.10), (B.12), and (B.13), we have

$$\left\| \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} (\bar{\tilde{\mathbf{r}}}_1^c - \bar{\tilde{\mathbf{r}}}_2^c - \bar{\tilde{\mathbf{r}}}_1^c + \bar{\tilde{\mathbf{r}}}_2^c) \right\|_2 \leq o_p(1) + o_p(1) + o_p(1) = o_p(1). \tag{B.14}$$

Therefore,

$$\begin{aligned} A_2 &\leq O_p(1) O_p(1) \left\| \left(\frac{1}{n_1} + \frac{1}{n_2}\right)^{-1/2} (\bar{\tilde{\mathbf{r}}}_1^c - \bar{\tilde{\mathbf{r}}}_2^c - \bar{\tilde{\mathbf{r}}}_1^c + \bar{\tilde{\mathbf{r}}}_2^c) \right\|_2 \\ &= O_p(1) O_p(1) o_p(1) = o_p(1). \end{aligned} \tag{B.15}$$

Similarly, we have

$$A_3 = o_p(1). \quad (\text{B.16})$$

Put Eq. (B.5), (B.15) and (B.16) together, we have

$$T_{p,N} \xrightarrow{D} \chi_p^2. \quad (\text{B.17})$$

Proof for Theorem 2:

When $E_t[\text{diag}(\boldsymbol{\lambda})\boldsymbol{\Phi}_{gi}^T \boldsymbol{\Sigma}_{Y_{gi}}^{-1}(\boldsymbol{\mu}_{gi} - \boldsymbol{\mu}_{pool;gi})] \neq \mathbf{0}$, without loss of generality, let's assume $E_t[\text{diag}(\boldsymbol{\lambda})\boldsymbol{\Phi}_{1i}^T \boldsymbol{\Sigma}_{Y_{1i}}^{-1}(\boldsymbol{\mu}_{1i} - \boldsymbol{\mu}_{pool;1i})] = c > 0$. Then $E_t[\text{diag}(\boldsymbol{\lambda})\boldsymbol{\Phi}_{2i}^T \boldsymbol{\Sigma}_{Y_{2i}}^{-1}(\boldsymbol{\mu}_{2i} - \boldsymbol{\mu}_{pool;2i})] = -\frac{\omega}{1-\omega}c < 0$.

Define $\tilde{\mathbf{r}}_{g\cdot}^* = \frac{1}{n_g} \sum_{i=1}^{n_g} \tilde{\mathbf{r}}_{gi}^*$, with $\tilde{\mathbf{r}}_{gi}^* = \text{diag}(\boldsymbol{\lambda})\boldsymbol{\Phi}_{gi}^T \boldsymbol{\Sigma}_{Y_{gi}}^{-1}(\boldsymbol{\mu}_{gi} - \boldsymbol{\mu}_{pool;gi})$. We have $\tilde{\mathbf{r}}_{1\cdot}^* - \tilde{\mathbf{r}}_{2\cdot}^* \xrightarrow{P} \frac{1}{1-\omega}c > 0$. Here we need $\text{Var}[\tilde{\mathbf{r}}_{gi}^*] < \infty$, i.e., the variance is bounded.

Next, we want to show that

$$(\tilde{\mathbf{r}}_{1\cdot}^c - \tilde{\mathbf{r}}_{2\cdot}^c)^T \hat{\mathbf{V}}^{-1}(\tilde{\mathbf{r}}_{1\cdot}^c - \tilde{\mathbf{r}}_{2\cdot}^c) \xrightarrow{P} (\tilde{\mathbf{r}}_{1\cdot}^* - \tilde{\mathbf{r}}_{2\cdot}^*)^T \mathbf{V}^{-1}(\tilde{\mathbf{r}}_{1\cdot}^* - \tilde{\mathbf{r}}_{2\cdot}^*). \quad (\text{B.18})$$

Note that we have the following decomposition

$$\begin{aligned} & |(\tilde{\mathbf{r}}_{1\cdot}^c - \tilde{\mathbf{r}}_{2\cdot}^c)^T \hat{\mathbf{V}}^{-1}(\tilde{\mathbf{r}}_{1\cdot}^c - \tilde{\mathbf{r}}_{2\cdot}^c) - (\tilde{\mathbf{r}}_{1\cdot}^* - \tilde{\mathbf{r}}_{2\cdot}^*)^T \mathbf{V}^{-1}(\tilde{\mathbf{r}}_{1\cdot}^* - \tilde{\mathbf{r}}_{2\cdot}^*)| \\ & \leq \left\| (\tilde{\mathbf{r}}_{1\cdot}^* - \tilde{\mathbf{r}}_{2\cdot}^*)^T [\hat{\mathbf{V}}^{-1} - \mathbf{V}^{-1}](\tilde{\mathbf{r}}_{1\cdot}^* - \tilde{\mathbf{r}}_{2\cdot}^*) \right\|_2 + \left\| (\tilde{\mathbf{r}}_{1\cdot}^c - \tilde{\mathbf{r}}_{2\cdot}^c - \tilde{\mathbf{r}}_{1\cdot}^* + \tilde{\mathbf{r}}_{2\cdot}^*)^T \hat{\mathbf{V}}^{-1}(\tilde{\mathbf{r}}_{1\cdot}^* - \tilde{\mathbf{r}}_{2\cdot}^*) \right\|_2 \\ & + \left\| (\tilde{\mathbf{r}}_{1\cdot}^c - \tilde{\mathbf{r}}_{2\cdot}^c)^T \hat{\mathbf{V}}^{-1}(\tilde{\mathbf{r}}_{1\cdot}^c - \tilde{\mathbf{r}}_{2\cdot}^c - \tilde{\mathbf{r}}_{1\cdot}^* + \tilde{\mathbf{r}}_{2\cdot}^*) \right\|_2 \\ & \stackrel{\text{def}}{=} C_1 + C_2 + C_3. \end{aligned} \quad (\text{B.19})$$

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

$$C_1 \leq \left\| (\tilde{\mathbf{r}}_{1\cdot}^* - \tilde{\mathbf{r}}_{2\cdot}^*) \right\|_2^2 o_p(1) = o_p(1). \quad (\text{B.20})$$

For C_2 , we have,

$$\tilde{\mathbf{r}}_{g\cdot}^c - \tilde{\mathbf{r}}_{g\cdot}^* = \tilde{\mathbf{r}}_{g\cdot}^c - \tilde{\mathbf{r}}_{g\cdot}^c + \text{diag}(\boldsymbol{\lambda}) \frac{1}{n_g} \sum_{i=1}^{n_g} \mathbf{Q}_{gi}(\mathbf{Y}_{gi} - \boldsymbol{\mu}_{gi}) = o_p(1). \quad (\text{B.21})$$

Then we have

$$\begin{aligned} C_2 & \leq \left\| \tilde{\mathbf{r}}_{1\cdot}^c - \tilde{\mathbf{r}}_{1\cdot}^* - \tilde{\mathbf{r}}_{2\cdot}^c + \tilde{\mathbf{r}}_{2\cdot}^* \right\|_2 \left\| \hat{\mathbf{V}}^{-1} \right\|_2 \left\| (\tilde{\mathbf{r}}_{1\cdot}^* - \tilde{\mathbf{r}}_{2\cdot}^*) \right\|_2 \\ & \leq o_p(1)O_p(1) \left\| \tilde{\mathbf{r}}_{1\cdot}^* - \tilde{\mathbf{r}}_{2\cdot}^* \right\|_2 = o_p(1). \end{aligned} \quad (\text{B.22})$$

Similarly, we have

$$C_3 = o_p(1). \quad (\text{B.23})$$

Combining Eq. (B.20), (B.22) and (B.23), and together with Eq. (B.19), we have

$$(\tilde{\mathbf{r}}_{1\cdot}^c - \tilde{\mathbf{r}}_{2\cdot}^c)^T \hat{\mathbf{V}}^{-1}(\tilde{\mathbf{r}}_{1\cdot}^c - \tilde{\mathbf{r}}_{2\cdot}^c) \xrightarrow{P} (\tilde{\mathbf{r}}_{1\cdot}^* - \tilde{\mathbf{r}}_{2\cdot}^*)^T \mathbf{V}^{-1}(\tilde{\mathbf{r}}_{1\cdot}^* - \tilde{\mathbf{r}}_{2\cdot}^*) \xrightarrow{P} c_1 > 0. \quad (\text{B.24})$$

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

Appendix C: An extension to the one-way functional ANOVA problem

Note that our proposed test statistic $T_{p,N}$ can be straightforwardly generalized to the one-way functional ANOVA problem. Suppose that the number of groups is $G > 2$. The hypothesis testing of interest is

$$H_0 : \mu_1(t) = \mu_2(t) = \dots = \mu_G(t), \quad t \in \mathcal{T} \quad (\text{C.1})$$

against the general alternative that at least two mean functions are not equal. Analogous to $T_{p,N}$, we propose to use

$$TG_{p,N} = \sum_{g=1}^G (\bar{\mathbf{r}}_{g\cdot}^c - \bar{\mathbf{r}}_{\cdot\cdot}^c)^T (\hat{\mathbf{V}}_2)^{-1} (\bar{\mathbf{r}}_{g\cdot}^c - \bar{\mathbf{r}}_{\cdot\cdot}^c) \xrightarrow{D} \chi_{(G-1)p}^2 \quad \text{under } H_0, \quad (\text{C.2})$$

with $\hat{\mathbf{V}}_2 = \frac{1}{n_1+n_2+\dots+n_G} \sum_{g=1}^G \sum_{i=1}^{n_g} \text{diag}(\hat{\boldsymbol{\lambda}}) \hat{\boldsymbol{\Phi}}_{gi}^T [\hat{\mathbf{G}}_{gi} + \hat{\sigma}^2 \mathbf{I}]^{-1} \hat{\boldsymbol{\Phi}}_{gi} \text{diag}(\hat{\boldsymbol{\lambda}})$.

References

- [1] 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. [MR2104869](#)
- [2] Michal Benko, Wolfgang Härdle, and Alois Kneip. Common functional principal components. *The Annals of Statistics*, 37(1):1–34, 2009. [MR2488343](#)
- [3] 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. [MR2564354](#)
- [4] Guanqun Cao, Lijian Yang, and David Todem. Simultaneous inference for the mean function based on dense functional data. *Journal of nonparametric statistics*, 24(2):359–377, 2012. [MR2921141](#)
- [5] Kehui Chen and Hans-Georg Müller. Modeling conditional distributions for functional responses, with application to traffic monitoring via gps-enabled mobile phones. *Technometrics*, 56(3):347–358, 2014. [MR3238072](#)
- [6] Antonio Cuevas, Manuel Febrero, and Ricardo Fraiman. An anova test for functional data. *Computational statistics & data analysis*, 47(1):111–122, 2004. [MR2087932](#)
- [7] Jianqing Fan and Irene Gijbels. *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press, 1996. [MR1383587](#)
- [8] Jianqing Fan and Sheng-Kuei Lin. Test of significance when data are curves. *Journal of the American Statistical Association*, 93(443):1007–1021, 1998. [MR1649196](#)
- [9] Julian J Faraway. Regression analysis for a functional response. *Technometrics*, 39(3):254–261, 1997. [MR1462586](#)
- [10] Garrett M Fitzmaurice, Nan M Laird, and James H Ware. *Applied longitudinal analysis*, volume 998. John Wiley & Sons, 2012. [MR2830137](#)

- [11] Bernhard N Flury. Common principal components in k groups. *Journal of the American Statistical Association*, 79(388):892–898, 1984. [MR0770284](#)
- [12] Colin B Fogarty and Dylan S Small. Equivalence testing for functional data with an application to comparing pulmonary function devices. *The Annals of Applied Statistics*, 8(4):2002–2026, 2014. [MR3292487](#)
- [13] Peter Hall and Ingrid Van Keilegom. Two-sample tests in functional data analysis starting from discrete data. *Statistica Sinica*, pages 1511–1531, 2007. [MR2413533](#)
- [14] Nadine Hilgert, André Mas, and Nicolas Verzelen. Minimax adaptive tests for the functional linear model. *The Annals of Statistics*, 41(2):838–869, 2013. [MR3099123](#)
- [15] Lajos Horváth, Piotr Kokoszka, and Ron Reeder. Estimation of the mean of functional time series and a two-sample problem. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(1):103–122, 2013. [MR3008273](#)
- [16] Lajos Horváth, Piotr Kokoszka, and Matthew Reimherr. Two sample inference in functional linear models. *Canadian Journal of Statistics*, 37(4):571–591, 2009. [MR2588950](#)
- [17] 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. [MR3300207](#)
- [18] Gareth M James, Trevor J Hastie, and Catherine A Sugar. Principal component models for sparse functional data. *Biometrika*, pages 587–602, 2000. [MR1789811](#)
- [19] Wolfgang Jank and Galit Shmueli. *Modeling Online Auctions*, volume 91. John Wiley & Sons, 2010. [MR2769017](#)
- [20] Wolfgang Jank, Galit Shmueli, Catherine Plaisant, and Ben Shneiderman. Visualizing functional data with an application to ebay’s online auctions. In *Handbook of Data Visualization*, pages 873–898. Springer, 2008.
- [21] Jong Soo Lee, Dennis D Cox, and Michele Follen. A two sample test for functional data. *Communications for Statistical Applications and Methods*, 22(2):121–135, 2015.
- [22] Jing Lei. Adaptive global testing for functional linear models. *Journal of the American Statistical Association*, 109(506):624–634, 2014. [MR3223738](#)
- [23] Yehua Li and Tailen Hsing. Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, 38(6):3321–3351, 2010. [MR2766854](#)
- [24] Bitao Liu and Hans-Georg Müller. Functional data analysis for sparse auction data. *Statistical Methods in eCommerce Research*, pages 269–290, 2008. [MR2414047](#)
- [25] Shujie Ma, Lijian Yang, and Raymond J Carroll. A simultaneous confidence band for sparse longitudinal regression. *Statistica Sinica*, 22:95, 2012. [MR2933169](#)
- [26] Hans-georg Müller. Functional modelling and classification of longitudinal data. *Scandinavian Journal of Statistics*, 32(2):223–240, 2005. [MR2188671](#)
- [27] E Paparoditis and T Sapatinas. Bootstrap-based testing of equality of

- mean functions or equality of covariance operators for functional data. *Biometrika*, 103(3):727–733, 2016. [MR3551795](#)
- [28] Debashis Paul and Jie Peng. Consistency of restricted maximum likelihood estimators of principal components. *The Annals of Statistics*, 37(3):1229–1271, 2009. [MR2509073](#)
- [29] 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. [MR2516804](#)
- [30] 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):995–1015, 2009. [MR2598035](#)
- [31] Alessia Pini, Aymeric Stamm, and Simone Vantini. Hotelling’s t_2 in separable hilbert spaces. *Journal of Multivariate Analysis*, 167:284–305, 2018. [MR3830647](#)
- [32] 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)*, 65(3):395–414, 2016. [MR3470583](#)
- [33] James O Ramsay. *Functional data analysis*. Wiley Online Library, 2006.
- [34] Philip T Reiss, Lei Huang, and Maarten Mennes. Fast function-on-scalar regression with penalized basis expansions. *The international journal of biostatistics*, 6(1), 2010. [MR2683940](#)
- [35] Qing Shen and Juliana Faraway. An f test for linear models with functional responses. *Statistica Sinica*, 14(4):1239–1258, 2004. [MR2126351](#)
- [36] Ana-Maria Staicu, Soumen N Lahiri, and Raymond J Carroll. Significance tests for functional data with complex dependence structure. *Journal of statistical planning and inference*, 156:1–13, 2015. [MR3267509](#)
- [37] 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. [MR3277031](#)
- [38] Shanshan Wang, Wolfgang Jank, and Galit Shmueli. Forecasting ebay’s online auction prices using functional data analysis. *University of Maryland, College Park, MD*, 20742, 2004.
- [39] 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.
- [40] 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. [MR2160561](#)
- [41] Jin-Ting Zhang and Jianwei Chen. Statistical inferences for functional data. *The Annals of Statistics*, 35(3):1052–1079, 2007. [MR2341698](#)
- [42] 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. [MR3181132](#)
- [43] 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. [MR2758746](#)
- [44] Xiaoke Zhang and Jane-Ling Wang. From sparse to dense functional data and beyond. *The Annals of Statistics*, 44(5):2281–2321, 2016. [MR3546451](#)