Testing homogeneity: the trouble with sparse functional data

Changbo Zhu[1] and Jane-Ling Wang[2]

[1]Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, United States
[2]Department of Statistics, University of California, Davis, Davis, United States

Address for correspondence: Changbo Zhu, Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN 46556, USA. Email: czhu4@nd.edu

Abstract

Testing the homogeneity between two samples of functional data is an important task. While this is feasible for intensely measured functional data, we explain why it is challenging for sparsely measured functional data and show what can be done for such data. In particular, we show that testing the marginal homogeneity based on point-wise distributions is feasible under some mild constraints and propose a new two-sample statistic that works well with both intensively and sparsely measured functional data. The proposed test statistic is formulated upon energy distance, and the convergence rate of the test statistic to its population version is derived along with the consistency of the associated permutation test. The aptness of our method is demonstrated on both synthetic and real data sets.

Keywords: convergence rate, energy distance, longitudinal data, measurement errors, sparse functional data, two-sample test

1 Introduction

Two-sample testing of equality of distributions, which is the homogeneity hypothesis, is fundamental in statistics and has a long history that dates back to Cramér (1928), Von Mises (1928), Kolmogorov (1933), and Smirnov (1939). The literature on this topic can be categorized by the data type considered. Classical tests designed for low-dimensional data include Bickel (1969), Friedman and Rafsky (1979), Bickel and Breiman (1983), Schilling (1986), Henze (1988) among others. For recent developments that are applicable to data of arbitrary dimension, we refer to energy distance (ED) (Székely & Rizzo, 2004) and maximum mean discrepancy (Gretton et al., 2012; Sejdinovic et al., 2013). To suit high-dimensional regimes (Aoshima et al., 2018; Hall et al., 2005; Zhong et al., 2021), extensions of ED and maximum mean discrepancy were studied in Chakraborty and Zhang (2021), Gao and Shao (2021), and Zhu and Shao (2021). Some other interesting developments of ED for data residing in a metric space include Klebanov (2006) and Lyons (2013). In this paper, we focus on functional data that are random samples of functions on a real interval, e.g., [0, 1] (Davidian et al., 2004; Hsing & Eubank, 2015; Ramsay & Silverman, 2005; Wang et al., 2016).

Two-sample inference for functional data is gaining more attention due to the explosion of data that can be represented as functions. A substantial literature has been devoted to comparing the mean and covariance functions between two groups of curves or testing the nullity of model coefficients, see Fan and Lin (1998), Cuevas et al. (2004), Cox and Lee (2008), Panaretos et al. (2010), Zhang et al. (2010), Horváth and Kokoszka (2012), Zhang and Liang (2014), Staicu

et al. (2015), Pini and Vantini (2016), Paparoditis and Sapatinas (2016), Wang et al. (2018), Guo et al. (2019), Zhang et al. (2019), He et al. (2020), Yuan et al. (2020), and Wang (2021). However, the underlying distributions of two random functions can have the same mean and covariance function, but differ in other aspects. Testing homogeneity, which refers to a hypothesis testing procedure to determine the equality of the underlying distributions of any two random objects, is thus of particular interest and practical importance. The literature on testing the homogeneity for functional data is much smaller and restricted to the two-sample case based on either fully observed (Benko et al., 2009; Cabaña et al., 2017; Krzyśko & Smaga, 2021; Wynne & Duncan, 2022) or intensively measured functional data (Hall & Keilegom, 2007; Jiang et al., 2019).

For intensively measured functional data, a presmoothing step is often adopted to each individual curve in order to construct a smooth curve before carrying out subsequent analysis, which may reduce mean square error (Ferraty et al., 2012) or, with luck, remove the noise, a.k.a. measurement error, in the observed discrete data. This results in a two-stage procedure, smoothing first and then testing homogeneity based on the presmoothed curves. For example, Hall and Keilegom (2007) and Jiang et al. (2019) adopted such an approach. In addition, the ED in Székely and Rizzo (2004) could be extended to the space of L^2 functions to characterize the distribution differences of random functions, provided the functional data are fully observed without errors (Klebanov, 2006).

Specifically, the ED between two random functions X and Y is defined as

$$\text{ED}(X, Y) = 2E[\|X - Y\|_{L^2}] - E[\|X - X'\|_{L^2}] - E[\|Y - Y'\|_{L^2}] \tag{1}$$

where X', Y' are i.i.d copies of X, Y, respectively, and $\|X - Y\|_{L^2} = (\int_0^1 (X(t) - Y(t))^2 \, dt)^{1/2}$. According to the results of Klebanov (2006) and Lyons (2013), $\text{ED}(X, Y)$ fully characterizes the distributions of X and Y in the sense that $\text{ED}(X, Y) \geq 0$ and $\text{ED}(X, Y) = 0 \Leftrightarrow X =^d Y$, where we use $X =^d Y$ to indicate that X, Y are identically distributed. Given two samples of functional data $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=n+1}^{n+m}$, which are intensively measured at some discrete time points, the reconstructed functions, denoted as $\{\widehat{X}_i\}_{i=1}^n$ and $\{\widehat{Y}_i\}_{i=n+1}^{n+m}$, can be obtained using the aforementioned presmoothing procedure. Then, $\text{ED}(X, Y)$ can be estimated by the U-type statistic

$$
\begin{aligned}
\text{ED}_n(X, Y) = {} & \frac{2}{mn} \sum_{i_1=1}^n \sum_{i_2=n+1}^{n+m} \|\widehat{X}_{i_1} - \widehat{Y}_{i_2}\|_{L^2} \\
& - \frac{2}{n(n-1)} \sum_{1 \leq i_1 < i_2 \leq n} \|\widehat{X}_{i_1} - \widehat{X}_{i_2}\|_{L^2} - \frac{2}{m(m-1)} \sum_{n+1 \leq i_1 < i_2 \leq n+m} \|\widehat{Y}_{i_1} - \widehat{Y}_{i_2}\|_{L^2}
\end{aligned}
\tag{2}
$$

While the above presmoothing procedure to reconstruct the original curves may be promising for intensely measured functional data, it has not yet been utilized to our knowledge, perhaps due to the technical and practical challenges to implement this approach as the level of intensity in the measurement schedule and the proper amount of smoothing are both critical. First, the distance between the reconstructed functional data and its target (the true curve) needs to be tracked and reflected in the subsequent calculations. Second, such a distance would depend on the intensity of the measurements and the bandwidth used in the presmoothing stage. Neither is easy to nail down in practice.

Furthermore, in real world applications, such as in longitudinal studies, each subject often may only have a few measurements, leading to sparse functional data (Yao et al., 2005a). Here, presmoothing individual data no longer works and one must borrow information from all subjects to reconstruct the trajectory of an individual subject. The principal analysis through conditional estimation (PACE) approach in Yao et al. (2005a) offers such an imputation method, yet it does not lead to consistent estimates of the true curve for sparsely observed functional data as there are not enough data available for each individual. Consequently, the quantities $E[\|X - Y\|_{L^2}]$, $E[\|X - X'\|_{L^2}]$, and $E[\|Y - Y'\|_{L^2}]$ in equation (1) are not consistently estimable as these expectations are outside the corresponding L^2 norms. Pomann et al. (2016) reduced the problem to testing the homogeneity of the scores of the two processes by assuming that the random functions are finite dimensional. Such an approach would not be consistent either as the scores still cannot be consistently estimated for sparse functional data. To our knowledge, there exists no

consistent test of homogeneity for sparse functional data. In fact, it is not feasible to test full homogeneity based on sparsely observed functional data as there are simply not enough data for such an ambitious goal.

This seems disappointing, since although it has long been recognized that sparse functional data are much more challenging to handle than intensively measured functional data, much progress has been made to resolve this challenge. For instance, both the mean and covariance function can be estimated consistently at a certain rate (Li & Hsing, 2010; Yao et al., 2005a; Zhang & Wang, 2016) for sparsely observed functional data. Moreover, the regression coefficient function in a functional linear model can also be estimated consistently with rates (Yao et al., 2005b). This motivated us to explore a less stringent concept of homogeneity that can be tested consistently for sparse functional data. In this paper, we provide the answer by proposing a test of marginal homogeneity for two independent samples of functional data. For ease of presentation, we assume that the random functions are defined on the unit interval [0, 1].

Definition of marginal homogeneity. Two random functions X and Y defined on [0, 1] are marginal homogeneous if

$$X(t) =^d Y(t) \text{ for almost all } t \in [0, 1]$$

From this definition we can see that unlike testing homogeneity that involves testing the entire distribution of functional data, testing marginal homogeneity only involves simultaneously testing the marginal distributions at all time points. This is a much more manageable task that works for all sampling designs, be it intensively or sparsely observed functional data, and it is often adequate in many applications. Testing marginal homogeneity is not new in the literature and has been investigated by Chakraborty and Zhang (2021) and Zhu and Shao (2021) for high-dimensional data. They show that the marginal tests can be more powerful than their joint counterparts under the high-dimensional regime. In a larger context, the idea of aggregating marginal information originates from Zhu et al. (2020), where they consider a related problem of testing the independence between two high-dimensional random vectors.

For real applications of testing marginal homogeneity, taking the analysis of biomarkers over time in clinical research as an example, comparing differences between marginal distributions of the treatment and control groups may be sufficient to establish the treatment effect. To contrast stocks in two different sectors, the differences between marginal distributions might be more important than the differences between joint distributions. In addition, differences between marginal distributions can be seen as the main effect of differences between distributions. Thus, it makes good sense to test marginal homogeneity, especially in situations where joint distribution testing is not feasible or inefficient.

Let λ be the Lebesgue measure on $\mathbb{R}$. The focus of this paper is to test

$$
\begin{aligned}
&H_0 : X(t) =^d Y(t) \text{ for almost all } t \in [0, 1] \\
&\qquad\qquad\qquad \text{versus} \\
&H_A : \text{there exists a set } \mathbb{T} \subseteq [0, 1] \text{ such that } \lambda(\mathbb{T}) > 0 \quad \text{and} \quad X(t) \neq^d Y(t) \quad \text{if } t \in \mathbb{T}
\end{aligned}
\tag{3}
$$

This can be accomplished through the marginal energy distance (MED) defined as

$$\text{MED}(X, Y) = \int 2E[|X(t) - Y(t)|] - E[|X(t) - X'(t)|] - E[|Y(t) - Y'(t)|] \, dt \tag{4}$$

where X' and Y' are independent copies of X and Y, respectively. Indeed, MED is a metric for marginal distributions in the sense that $\text{MED}(X, Y) \geq 0$ and $\text{MED}(X, Y) = 0 \Leftrightarrow X(t) =^d Y(t)$ for almost all $t \in [0, 1]$. A key feature of our approach is that it can consistently estimate $\text{MED}(X, Y)$ for all types of sampling plans. Moreover, $E[|X(t) - Y(t)|]$, $E[|X(t) - X'(t)|]$, and $E[|Y(t) - Y'(t)|]$ can all be reconstructed consistently for both intensively and sparsely observed functional data. Such a unified procedure for all kinds of sampling schemes may be more practical as the separation between intensively and sparsely observed functional data is usually unclear in practical applications. Moreover, it could happen that while some of the subjects are intensively

observed, others are sparsely observed. In the extremely sparse case, our method can still work if each subject only has one measurement.

Measurement errors (or noise) are common for functional data, so it is important to accommodate them. If noise is left unattended, there will be bias in the estimates of MED as the observed distributions are no longer the true distributions of X and Y. One might hope that the measurement errors can be averaged out during the estimation of the function $E[|X(t) - Y(t)|]$ in MED(X, Y) in analogy to estimating the mean function $\mu(t) = E[X(t)]$ or covariance function $C(s, t) = E[(X(t) - \mu(t))(X(s) - \mu(s))]$ (Yao et al., 2005a). However, this is not the case. To see why, let e_1, e_1', e_2, and e_2' be independent white noise. When estimating mean or covariance function at any fixed time $t, s \in [0, 1]$, it holds that $\mu(t) = E[X(t) + e_1]$ and $C(s, t) = E[(X(t) - \mu(t) + e_1)(X(s) - \mu(s) + e_1')]$ for $t \neq s$. But $E[|X(t) - Y(s) + e_1 - e_2|] \neq E[|X(t) - Y(s)|]$. Likewise, we can see that the ED in (1) would have the same challenge to handle measurement errors unless these errors were removed in a presmoothing step before carrying out the test. So the challenges with measurement errors is not triggered by the use of the L^1 norm in MED. The L^2 norm in ED will face the same challenge.

For intensely measured functional data, a presmoothing step is often used to handle measurement errors in the observed data in the hope that smoothing will remove the error. However, this is a delicate issue, as it is difficult to know the amount of smoothing needed in order for the subsequent analysis to retain the same convergence rate as if the true functional data were fully observed without errors. For instance, Zhang and Chen (2007) study the effects of smoothing to obtained reconstructed curves and show that in order to retain the same convergence rate of mean estimation for fully observed functional data, the number of measurements per subject that generates the curves must be of higher order than the number of independent subjects. This requires functional data that are intensively sampled well beyond ultra-dense (or dense) functional data that have been studied in the literature (Zhang & Wang, 2016).

In this paper, we propose a new way of handling measurement errors so that the MED-based testing procedure is still consistent in the presence of measurement errors. The key idea is to show that when the measurement errors e_1 and e_2 of X and Y, respectively, are identically distributed, i.e., $e_1 =^d e_2$, the MED(X, Y)-based approach can still be applied to the contaminated data with consistency guaranteed under mild assumptions (cf. Corollary 3). When $e_1 \neq^d e_2$, we propose an error-augmentation approach, which can be applied jointly with our unified estimation procedure.

To conduct hypothesis testing, MED(X, Y)-based approaches can be implemented as permutation tests. We refer the book by Lehmann and Romano (2005) for an extensive introduction to permutation test. The use of permutation test for distance/kernel-based tests is not new in the literature. The consistency of permutation test for distance covariance (Székely et al., 2007) or Hilbert–Schmidt independence criterion (Gretton et al., 2007) has been investigated by Pfister et al. (2018), Rindt et al. (2021), and Kim et al. (2022) for low-dimensional data, where distance covariance and Hilbert–Schmidt independence criterion are distance-based and kernel-based independence measures, respectively. Under the high dimension, low sample size setting (Hall et al., 2005; Zhu & Shao, 2021) studied the size and power behaviours of permutation test for ED. The difference in the proof of consistency of permutation test between longitudinal data and vector-valued data is substantial. For instance, MED estimated from longitudinal data is an integration of local weighted averages, where the measurement time points in the weights are also random and hence requires special handling in the proof.

The rest of the paper is organized as follows. Section 2 contains the main methodology and supporting theory about testing marginal homogeneity. Numerical studies are presented in Section 3. The conclusion is in Section 4. All technical details are postponed to Section 5.

2 Testing marginal homogeneity

We first consider the case where there are no measurement errors and postpone the discussion of measurement errors to the end of this section. Let $\{X_i\}_{i=1}^{n}$ and $\{Y_i\}_{i=n+1}^{n+m}$ be i.i.d copies of X and Y, respectively. In practice, the functions are only observed at some discrete points:

$$x_{ij} = X_i(T_{ij}) \quad \text{if } i = 1, 2, \ldots, n, \ j = 1, 2, \ldots, N_i$$
$$y_{ij} = Y_i(T_{ij}) \quad \text{if } i = n + 1, \ldots, n + m, \ j = 1, 2, \ldots, N_i$$

This sampling plan allows the two samples to be measured at different schedules and additionally each subject within the sample has its own measurement schedule. This is a realistic assumption but the consequence is that two-dimensional smoothers will be needed to estimate the targets. Fortunately, the convergence rate of our estimator attains the same convergence rate as that of a one-dimensional smoothing method. This intriguing phenomenon will be explained later.

For notational convenience, denote $Z_i = X_i$, if $i = 1, 2, \ldots, n$ and $Z_i = Y_i$, if $i = n + 1, \ldots, n + m$. Let $\mathbf{Z} = (\mathbf{z}_1; \ldots; \mathbf{z}_{n+m})$ be the combined observations, where for $i = 1, 2, \ldots, n + m$, $\mathbf{z}_i$ is a vector of length N_i,

$$\mathbf{z}_i = (z_{i1}, z_{i2}, \ldots, z_{i,N_i})^T = \begin{cases} (x_{i1}, x_{i2}, \ldots, x_{i,N_i})^T & \text{if } 1 \leq i \leq n \\ (y_{i1}, y_{i2}, \ldots, y_{i,N_i})^T & \text{if } n + 1 \leq i \leq n + m \end{cases}$$

The observations corresponding to X and Y are defined as $\mathbf{X} = (\mathbf{z}_1; \ldots; \mathbf{z}_n)$ and $\mathbf{Y} = (\mathbf{z}_{n+1}; \ldots; \mathbf{z}_{n+m})$, respectively.

To estimate MED(X, Y) in (4), note that we actually have no observations for the one-dimensional functions $E[|X(t) - Y(t)|]$, $E[|X(t) - X'(t)|]$, and $E[|Y(t) - Y'(t)|]$, due to the longitudinal design where X and Y are observed at different time points. Thus, the sampling schedule for X and Y are not synchronized. A consequence of such asynchronized functional/longitudinal data is that a one-dimensional smoothing method that has typically been employed to estimate a one-dimensional target function, e.g., $E[|X(t) - Y(t)|]$, does not work here because $X(t) - Y(t)$ cannot be observed at any point t. However, a workaround is to estimate the following two-dimensional functions first:

$$G_1(t_1, t_2) := E[|X(t_1) - Y(t_2)|]$$
$$G_2(t_1, t_2) := E[|X(t_1) - X'(t_2)|]$$
$$G_3(t_1, t_2) := E[|Y(t_1) - Y'(t_2)|]$$

then set $t_1 = t_2 = t$ in all three estimators. Since G_1, G_2, G_3 can all be recovered by some local linear smoother, MED(X, Y) admits consistent estimates for both intensively and sparsely observed functional data. For instance, $G_1(t_1, t_2)$ can be estimated by $\widehat{G}_1(t_1, t_2) = \hat{\beta}_0$, where

$$
\begin{aligned}
(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \underset{\beta_0, \beta_1, \beta_2}{argmin} \frac{1}{nm} \sum_{1 \leq i_1 \leq n} \sum_{n+1 \leq i_2 \leq n+m} \frac{1}{N_{i_1}} \frac{1}{N_{i_2}} \sum_{j_1=1}^{N_{i_1}} \sum_{j_2=1}^{N_{i_2}} K_{h_x}(T_{i_1 j_1} - t_1) \\
\times K_{h_y}(T_{i_2 j_2} - t_2)\left[|z_{i_1 j_1} - z_{i_2 j_2}| - \beta_0 - \beta_1(T_{i_1 j_1} - t_1) - \beta_2(T_{i_2 j_2} - t_2) \right]^2
\end{aligned}
\tag{5}
$$

and $G_2(t_1, t_2)$ can be estimated by $\widehat{G}_2(t_1, t_2) = \hat{\alpha}_0$, where

$$
\begin{aligned}
(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2) = \underset{\alpha_0, \alpha_1, \alpha_2}{argmin} \frac{2}{n(n-1)} \sum_{1 \leq i_1 < i_2 \leq n} \frac{1}{N_{i_1}} \frac{1}{N_{i_2}} \sum_{j_1=1}^{N_{i_1}} \sum_{j_2=1}^{N_{i_2}} K_{h_x}(T_{i_1 j_1} - t_1) K_{h_x}(T_{i_2 j_2} - t_2) \\
\left[|z_{i_1 j_1} - z_{i_2 j_2}| - \alpha_0 - \alpha_1(T_{i_1 j_1} - t_1) - \alpha_2(T_{i_2 j_2} - t_2) \right]^2
\end{aligned}
\tag{6}
$$

$G_3(t_1, t_2)$ can be estimated similarly as $G_2(t_1, t_2)$ by an estimator $\widehat{G}_3(t_1, t_2)$. In the above, N_{i1} and N_{i2} should be understood as the respective length of the vector $\mathbf{z}_{i1}$ and $\mathbf{z}_{i2}$ and $K_h(\cdot) = K(\cdot / h)/h$ is a one-dimensional kernel with bandwidth h. The sample estimate of MED(X, Y) can then be constructed as

$$\text{MED}_n(\mathbf{Z}) := \int_0^1 2\widehat{G}_1(t, t) - \widehat{G}_2(t, t) - \widehat{G}_3(t, t) \, dt \tag{7}$$

For hypothesis testing, the critical value or p-value can be determined by permutations (Lehmann & Romano, 2005). To be more specific, let $\pi : \{1, 2, \ldots, n + m\} \to \{1, 2, \ldots, n + m\}$ be a permutation. There are $(n + m)!$ number of permutations in total and we denote the set of permutations as $\mathbb{P}_{n+m} = \{\pi_l : l = 1, 2, \ldots, (n + m)!\}$. For $l = 1, 2, \ldots, (n + m)!$, define the permutation of π_l on $\mathbf{Z}$ as

$$\pi_l \cdot \mathbf{Z} = (\mathbf{z}_{\pi_l(1)}; \mathbf{z}_{\pi_l(2)}; \ldots; \mathbf{z}_{\pi_l(n+m)}) \tag{8}$$

Write the statistic that is based on the permuted sample $\pi_l \cdot \mathbf{Z}$ as $\mathrm{MED}_n(\pi_l \cdot \mathbf{Z})$ and let $\Pi_1, \ldots, \Pi_{S-1}$ be i.i.d and uniformly sampled from $\mathbb{P}_{n+m}$, we define the permutation-based p-value as

$$\widehat{p} = \frac{1}{S} \left\{ 1 + \sum_{l=1}^{S-1} \mathbb{I}\{\mathrm{MED}_n(\Pi_l \cdot \mathbf{Z}) \geq \mathrm{MED}_n(\mathbf{Z})\} \right\}$$

Then, the level-α permutation test w.r.t. $\mathrm{MED}_n(\mathbf{Z})$ can be defined as

$$\text{Reject } H_0 \quad \text{if } \widehat{p} \leq \alpha$$

2.1 Convergence theory

In this subsection, we show that $\mathrm{MED}_n(\mathbf{Z})$ is a consistent estimator and develop its convergence rate.

Assumption 1   (A.1) The kernel function $K(\cdot) \geq 0$ is symmetric, Lipschitz continuous, supported on $[-1, 1]$, and satisfies

$$\int K(u)\, du = 1, \quad \int_0^1 u^2 K(u)\, du < \infty, \quad \text{and} \quad \int_0^1 K(u)^2\, du < \infty$$

   (A.2) Let $\{T_{ij} : 1 \leq i \leq n, 1 \leq j \leq N_i\} \overset{i.i.d}{\sim} T_x$, $\{T_{ij} : n + 1 \leq i \leq n + m, 1 \leq j \leq N_i\} \overset{i.i.d}{\sim} T_y$ and denote the density functions of T_x, T_y by g_x, g_y, respectively. There exists constants c and C such that $0 < c \leq g_x(s), g_y(t) \leq C < \infty$ for any $s, t \in [0, 1]$.

   (A.3) $\{X_{i_1}, Y_{i_2}, T_{ij} : 1 \leq i_1 \leq n, n + 1 \leq i_2 \leq n + m, 1 \leq i \leq n + m, 1 \leq j \leq N_i\}$ are mutually independent.

   (A.4) The second-order partial derivatives of G_1, G_2, G_3 are bounded on $[0, 1]$.

   (A.5) $\sup_t E|X(t)|^2 < \infty$ and $\sup_t E|Y(t)|^2 < \infty$.

Remark 1   Conditions (A.1)–(A.3) and (A.5) are fairly standard and also used in Li and Hsing (2010). Condition (A.4) may seem a bit problematic at first, as the absolute value function $|\cdot|$ is not differentiable at 0. However, its expectation can easily be differentiable. For instance, if the density functions of $X(t)$, $Y(t)$ are $f_x(\cdot|t)$, $f_y(\cdot|t)$, respectively, then we have $G_1(s, t) = \iint |u - v| f_x(u|s) f_y(v|t)\, du\, dv$. Assuming the conditions of the Leibniz integral rule, we can interchange the partial derivatives and integration, i.e.,

$$\frac{\partial^2}{\partial s \partial t} G_1(s, t) = \iint |u - v| \frac{\partial}{\partial s} f_x(u|s) \frac{\partial}{\partial t} f_y(v|t)\, du\, dv$$

Thus, the partial derivatives of $G_1(s, t)$ are bounded if the second-order partial derivatives of $f_x(u|s)$, $f_y(v|t)$ w.r.t. s, t exist for all u, v and

$$\sup_u \left| \frac{\partial^2}{\partial s \partial s} f_x(u|s) \right| < \infty \quad \text{and} \quad \sup_v \left| \frac{\partial^2}{\partial t \partial t} f_y(v|t) \right| < \infty \tag{9}$$

A more specific example is when $X(t)$ is Gaussian and $Y(t)$ is a mixture of Gaussians with density functions

$$f_x(u|t) = \frac{1}{\sigma_1(t)\sqrt{2\pi}} e^{-(1/2)\left((u-\mu_1(t))/\sigma_1(t)\right)^2}$$

$$f_y(u|t) = \frac{1}{2}\frac{1}{\sigma_2(t)\sqrt{2\pi}} e^{-(1/2)\left((u-\mu_2(t))/\sigma_2(t)\right)^2} + \frac{1}{2}\frac{1}{\sigma_2(t)\sqrt{2\pi}} e^{-(1/2)\left((u+\mu_2(t))/\sigma_2(t)\right)^2}$$

Then (A.4) holds if we assume $\sigma_1(s)$, $\sigma_2(t)$ are bounded from below by a positive constant, $\mu_1(s)$, $\mu_2(t)$, $\sigma_1(s)$, $\sigma_2(t)$ are bounded and have bounded second-order derivatives. Similar conclusions can be drawn for G_2 and G_3. Therefore, Condition (A.4) is not restrictive as it is customary to assume that the mean and covariance functions for functional data are differentiable.

The next assumption specifies the relationship of the number of observations per subject and the decay rate of the bandwidth parameters h_x, h_y.

Assumption 2 Suppose $h_x := h_x(n)$, $h_y := h_y(m) \to 0$ and

$$\log\left(\frac{n}{\sum_{i=1}^n N_i^{-1}/n}\right) \frac{\max_{1\le i\le n} N_i^{-1}}{h_x} \frac{\max_{1\le i\le n} N_i^{-1}}{\sum_{i=1}^n N_i^{-1}/n} \frac{1}{n} \to 0$$

$$\log\left(\frac{m}{\sum_{i=n+1}^{n+m} N_i^{-1}/m}\right) \frac{\max_{n+1\le i\le n+m} N_i^{-1}}{h_y} \frac{\max_{n+1\le i\le n+m} N_i^{-1}}{\sum_{i=n+1}^{n+m} N_i^{-1}/m} \frac{1}{m} \to 0$$

The following theorem states that we can consistently estimate $\mathrm{MED}(X, Y)$ with sparse observations.

Theorem 1 Under Assumptions 1 and 2,

$$\left|\mathrm{MED}_n(\mathbf{Z}) - \mathrm{MED}(X, Y)\right| = O_p\left(h_x^2 + \sqrt{\frac{1}{n^2}\sum_{i=1}^n \phi_i} + h_y^2 + \sqrt{\frac{1}{m^2}\sum_{i=n+1}^{n+m} \phi_i}\right)$$

where $\{\phi_i : i = 1, 2, \ldots, n+m\}$ are defined as

$$\phi_i = \begin{cases} \dfrac{N_i h_x + N_i(N_i - 1)h_x^2}{N_i^2 h_x^2}, & 1 \le i \le n \\[3mm] \dfrac{N_i h_y + N_i(N_i - 1)h_y^2}{N_i^2 h_y^2}, & n+1 \le i \le n+m \end{cases}$$

Remark 2 For any two-dimensional function $F \in L^2([0, 1]^2)$, define the L^2-norm as $\|F\|_2 := (\int_{t_1}\int_{t_2} [F(t_1, t_2)]^2 \, dt_1 \, dt_2)^{1/2}$. The above theorem is a consequence of

$$\left\| G_I(t_1, t_2) - \widehat{G}_I(t_1, t_2) \right\|_2 = O_p\left(h_x^2 + \sqrt{\frac{1}{n^2}\sum_{i=1}^n \phi_i} + h_y^2 + \sqrt{\frac{1}{m^2}\sum_{i=n+1}^{n+m} \phi_i}\right)$$

where $I = 1, 2, 3$. Compared with the mean function $\mu(t) = E[X(t)]$ and the covariance function $C_X(s, t) = E[(X(s) - \mu(s))(X(t) - \mu(t))]$, G_1, G_2, G_3 are functions involving two independent stochastic processes. An intriguing phenomenon is that even though G_1, G_2, G_3 are two-dimensional functions, the convergence rate of their linearly smoothed estimates is the same as for a

one-dimensional function, such as the mean function. This is because the expectation $G_1(s, t) = E[|X(s) - Y(t)|]$ involves two independent stochastic processes and $n \times m$ pairs $\{(X_{i_1}, Y_{i_2}) : i_1 = 1, \ldots, n, i_2 = n + 1, \ldots, n + m\}$ are used in the linear smoother, leading to a faster convergence rate. Therefore, this situation is different from the standard estimation of a bivariate function. For instance, if the goal is to estimate $E[|X(s) - X(t)|]$, the convergence rate would be slower and would be the same as that for a two-dimensional smoother. We further point out that even though a two-dimensional smoothing method is used to estimate $G_I(s, t)$, we only need to evaluate its values at the diagonal where $s = t$. Therefore, the computational effort, which is equivalent to a one-dimensional smoother, is manageable.

Given two sequences of positive real numbers a_n and b_n, we say that a_n and b_n are of the same order as $n \to \infty$ (denoted as $a_n \asymp b_n$) if there exists constants $0 < c_1 < c_2 < \infty$ such that $c_1 \leq \lim_{n \to \infty} a_n/b_n \leq c_2$ and $c_1 \leq \lim_{n \to \infty} b_n/a_n \leq c_2$. The convergence rates of $\mathrm{MED}_n(\mathbf{Z})$ for different sampling plans are provided in the following corollary.

Corollary 1 Under Assumptions 1 and 2, and further assume $m(n) \asymp n$.

(i) When $N_i \asymp C$ for all $i = 1, 2, \ldots, n + m$, where $0 < C < \infty$ is a constant, and $h_x \asymp h_y \asymp n^{-1/5}$, we have

$$\left| \mathrm{MED}_n(\mathbf{Z}) - \mathrm{MED}(X, Y) \right| = O_p\left(\frac{1}{n^{2/5}}\right)$$

(ii) When $N_i \asymp n^{1/4}$ for all $i = 1, 2, \ldots, n + m$ and $h_x \asymp h_y \asymp n^{-1/4}$, we have

$$\left| \mathrm{MED}_n(\mathbf{Z}) - \mathrm{MED}(X, Y) \right| = O_p\left(\frac{1}{\sqrt{n}}\right)$$

2.2 Validity of the permutation test and power analysis

We now justify the permutation-based test for sparsely observed functional data. Under the null hypothesis and the mild assumption that $\{N_i\}$ are i.i.d across subjects, the size of the test can be guaranteed by the fact that the distribution of the sample is invariant under permutation. For a rigorous argument, see Theorem 15.2.1 in Lehmann and Romano (2005). Thus, the permutation test based on the test statistic (7) produces a legitimate size of the test. The power analysis is much more challenging and will be presented below.

Let $\pi \in \mathbb{P}_{n+m}$ be a fixed permutation and $\widehat{G}_{\pi,I}(t_1, t_2)$, $I = 1, 2, 3$, be the estimated functions from algorithms (5) and (6) using permuted samples $\pi \cdot \mathbf{Z}$. The conditions on the decay rate of bandwidth parameters h_x, h_y that ensure the convergence of $\widehat{G}_{\pi,I}$ for any fixed permutation π are summarized below.

Assumption 3 Suppose $h_x := h_x(n)$, $h_y := h_y(m) \to 0$ and

$$\sup_{\pi \in \mathbb{P}_{n+m}} \log\left(\frac{n^2}{\sum_{i=1}^{n} N_{\pi(i)}^{-1}}\right) \frac{\max_{1 \leq i \leq n} N_{\pi(i)}^{-1}}{\min\{h_x, h_y\}} \frac{\max_{1 \leq i \leq n} N_{\pi(i)}^{-1}}{\sum_{i=1}^{n} N_{\pi(i)}^{-1}/n} \frac{1}{n} \to 0$$

$$\sup_{\pi \in \mathbb{P}_{n+m}} \log\left(\frac{m^2}{\sum_{i=n+1}^{n+m} N_{\pi(i)}^{-1}}\right) \frac{\max_{n+1 \leq i \leq n+m} N_{\pi(i)}^{-1}}{\min\{h_x, h_y\}} \frac{\max_{n+1 \leq i \leq n+m} N_{\pi(i)}^{-1}}{\sum_{i=n+1}^{n+m} N_{\pi(i)}^{-1}/m} \frac{1}{m} \to 0$$

Let Π be a random permutation uniformly sampled from $\mathbb{P}_{n+m}$. If the sample is randomly shuffled, it holds that $Z_{\Pi(i)} =^d Z_{\Pi(j)}$ and $\mathrm{MED}(Z_{\Pi(i)}, Z_{\Pi(j)}) = 0$ for any $i, j = 1, 2, \ldots, n + m$. For the sample estimate $\mathrm{MED}_n(\Pi \cdot \mathbf{Z})$ based on the permuted sparse observations, we show that $\mathrm{MED}_n(\Pi \cdot \mathbf{Z})$ converges to 0 in probability.

Theorem 2 Under Assumptions 1 and 3,

$$\left|\text{MED}_n(\Pi \cdot \mathbf{Z})\right| = O_p\left(\sup_{\pi \in \mathbb{P}_{n+m}} \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}\phi_{\pi(i)}} + \sup_{\pi \in \mathbb{P}_{n+m}} \sqrt{\frac{1}{m^2}\sum_{i=n+1}^{n+m}\phi_{\pi(i)}}\right)$$

where $\Pi \sim \text{Uniform}(\mathbb{P}_{n+m})$ and is independent of the data, and $\{\phi_i\}_{i=1}^n$ are defined in Theorem 1.

For the original data that have not been permuted, $\text{MED}_n(\mathbf{Z}) \to^p \text{MED}(X, Y)$, which is strictly positive under the alternative hypothesis. On the other hand, we know from Theorem 2 that the permuted statistics converges to 0 in probability. This suggests that under mild assumptions, the probability of rejecting the null approaches 1 as $n, m \to \infty$. We make this idea rigorous in the following theorem.

Theorem 3 Under Assumptions 1 and 3, for any fixed $S > 1/\alpha$, we have

$$P_{H_A}\left(\widehat{p} \le \alpha\right) \to 1$$

Remark 3 Since Assumption 3 implies Assumption 2, Theorem 1 holds under the assumption of Theorem 3 and it facilitates the proof of Theorem 3.

2.3 Handling of measurement errors

With the presence of measurement errors, the actual observed data are

$$\widetilde{x}_{ij} = X_i(T_{ij}) + e_{ij}, \quad i = 1, 2, \ldots, n, \ j = 1, 2, \ldots, N_i$$
$$\widetilde{y}_{ij} = Y_i(T_{ij}) + e_{ij}, \quad i = n+1, \ldots, n+m, \ j = 1, 2, \ldots, N_i$$

where $\{e_{ij} : i = 1, 2, \ldots, n, j = 1, 2, \ldots, N_i\} \overset{i.i.d}{\sim} e_1$, $\{e_{ij} : i = n+1, \ldots, n+m, j = 1, 2, \ldots, N_i\} \overset{i.i.d}{\sim} e_2$, and e_1, e_2 are mean 0 independent univariate random variables. Denote the combined noisy observations by $\widetilde{\mathbf{Z}} = (\widetilde{\mathbf{z}}_1; \ldots; \widetilde{\mathbf{z}}_{n+m})$, where

$$\widetilde{\mathbf{z}}_i = (\widetilde{z}_{i1}, \widetilde{z}_{i2}, \ldots, \widetilde{z}_{i,N_i})^T = \begin{cases} (\widetilde{x}_{i1}, \widetilde{x}_{i2}, \ldots, \widetilde{x}_{i,N_i})^T & \text{if } 1 \le i \le n \\ (\widetilde{y}_{i1}, \widetilde{y}_{i2}, \ldots, \widetilde{y}_{i,N_i})^T & \text{if } n+1 \le i \le n+m \end{cases}$$

The local linear smoothers described in equations (5) and (6) are then applied with the input data $\{z_{i_1 j_1}\}$ and $\{z_{i_2 j_2}\}$ replaced, respectively, by $\{\widetilde{z}_{i_1 j_1}\}$ and $\{\widetilde{z}_{i_2 j_2}\}$. The resulting outputs are denoted as $\widehat{H}_1, \widehat{H}_2, \widehat{H}_3$, leading to the estimator

$$\text{MED}_n(\widetilde{\mathbf{Z}}) = \int 2\widehat{H}_1(t, t) - \widehat{H}_2(t, t) - \widehat{H}_3(t, t)\, dt \tag{10}$$

Correspondingly, the proposed test with contaminated data $\widetilde{\mathbf{Z}}$ is

$$\text{Reject } H_0 \quad \text{if } \widetilde{p} \le \alpha$$

where $\widetilde{p} = (1/S)\{1 + \sum_{l=1}^{S-1} \mathbb{I}_{\{\text{MED}_n(\Pi_l \cdot \widetilde{\mathbf{Z}}) \ge \text{MED}_n(\widetilde{\mathbf{Z}})\}}\}$. To study the convergence of $\text{MED}_n(\widetilde{\mathbf{Z}})$, define the two-dimensional functions $H_1(s, t), H_2(s, t)$, and $H_3(s, t)$ as

$$H_1(s, t) = E[|X(s) + e_1 - Y(t) - e_2|]$$
$$H_2(s, t) = E[|X(s) + e_1 - X'(t) - e_1'|] \tag{11}$$
$$H_3(s, t) = E[|Y(s) + e_2 - Y'(t) - e_2'|]$$

where e_1' and e_2' are independent and identical copies of e_1 and e_2, respectively. The target of $\mathrm{MED}_n(\widetilde{\mathbf{Z}})$ is shown to be

$$\widetilde{\mathrm{MED}}(X, Y) = \int 2H_1(t, t) - H_2(t, t) - H_3(t, t)\, \mathrm{d}t \tag{12}$$

Remark 4 An unpleasant fact is that $\widetilde{\mathrm{MED}}(X, Y)$ involves errors, which cannot be easily removed due to the presence of the absolute error function in (11). The handling of measurement errors in both method and theory is thus very different here from conventional approaches for functional data, where one does not deal with the absolute function. The ED with L^2-norm in equation (2) also has this issue. Thus, measurement errors would also be a challenge for the full homogeneity test even if we can approximate the L^2 norm in the ED well.

To show the approximation error of $\mathrm{MED}_n(\widetilde{\mathbf{Z}})$, we need the following assumptions.

Assumption 4 (D.1) $E[e_1^2] < \infty$ and $E[e_2^2] < \infty$.
(D.2) $\{X_{i_1}, Y_{i_2}, T_{ij}, e_{ij} : 1 \le i_1 \le n, n+1 \le i_2 \le n+m, 1 \le i \le n+m, 1 \le j \le N_i\}$ are mutually independent.
(D.3) The second-order partial derivatives of H_1, H_2, H_3 are bounded on $[0, 1]$.

Remark 5 Using the notations $f_x(\cdot|t)$ and $f_y(\cdot|t)$ in Remark 1, let the density functions of e_1 and e_2 be $\eta_1(\cdot)$ and $\eta_2(\cdot)$, respectively. Under the conditions of the Leibniz integral rule

$$\frac{\partial^2}{\partial s \partial t} H_1(s, t) = \int |u - v + a - b| \frac{\partial}{\partial s} f_x(u|s) \frac{\partial}{\partial t} f_y(v|t) \eta_1(a) \eta_2(b)\, \mathrm{d}u\, \mathrm{d}v\, \mathrm{d}a\, \mathrm{d}b$$

which admits bounded second-order partial derivatives if (9) holds. Similar conclusions can be drawn for H_2 and H_3. Therefore, Assumption (D.3) is mild.

Corollary 2 Under Assumptions 1, 2, and 4, we have

$$\left| \mathrm{MED}_n(\widetilde{\mathbf{Z}}) - \widetilde{\mathrm{MED}}(X, Y) \right| = O_p\left(h_x^2 + \sqrt{\frac{1}{n^2} \sum_{i=1}^{n} \phi_i} + h_y^2 + \sqrt{\frac{1}{m^2} \sum_{i=n+1}^{n+m} \phi_i} \right)$$

By the property of ED, it holds that $\widetilde{\mathrm{MED}}(X, Y) \ge 0$ and $\widetilde{\mathrm{MED}}(X, Y) = 0 \Leftrightarrow X(t) + e_1 =^d Y(t) + e_2$ for almost all $t \in [0, 1]$. Then, we show that under the following assumptions, the condition that $X(t) + e_1 =^d Y(t) + e_2$ would imply the homogeneity of $X(t)$ and $Y(t)$.

Assumption 5 Suppose that for any $t \in [0, 1]$

(E.1) $X(t), Y(t)$ are continuous random variables with density functions $f_x(\cdot|t), f_y(\cdot|t)$, respectively.
(E.2) e_1, e_2 are i.i.d continuous random variables with characteristic function $\phi(\cdot)$ and the real zeros of $\phi(\cdot)$ have Lebesgue measure 0.
(E.3) $\{X(t), Y(t), e_1, e_2\}$ are mutually independent.

For common distributions, such as Gaussian and Cauchy, their characteristic functions are of exponential form and have no real zeros. Some other random variables, such as exponential, Chi-square, and gamma, have characteristic functions of the form $\psi(t) = (1 - it\theta)^{-k}$ with only a finite number of real zeros. Since it is common to assume Gaussian measurement errors, the restriction on the real zeros of the characteristic function in Assumption 5(E.2) is very mild.

Theorem 4 Under Assumption 5, for any $t \in [0, 1]$,

$$X(t) + e_1 =^d Y(t) + e_2 \Leftrightarrow X(t) =^d Y(t)$$

Based on the above theorem, we have the important property that $\widetilde{\mathrm{MED}}(X, Y) = 0 \Leftrightarrow X(t) =^d Y(t)$ for almost all $t \in [0, 1]$. As discussed before, $\widetilde{\mathrm{MED}}(X, Y)$ can be consistently estimated by $\mathrm{MED}_n(\widetilde{\mathbf{Z}})$ and the test can be conducted via permutations. Consequently, the data contaminated with measurement errors can still be used to test marginal homogeneity as long as $e_1 =^d e_2$. We make this statement rigorous below.

Corollary 3 Under Assumptions 1, 3–5, for any fixed $S > 1/\alpha$,

$$P_{H_A}(\widetilde{p} \leq \alpha) \to 1$$

The circumstance of identically distributed errors among the two samples is a strong assumption that nevertheless can be satisfied in many real situations, for example, when the curves $\{X_i\}$ and $\{Y_j\}$ are measured by the same instrument. The primary biliary cirrhosis (PBC) data in Section 3.2 underscore this phenomenon.

When $e_1 \neq^d e_2$, not all is lost and we show that some workarounds exist. In particular, we propose an error-augmentation method that raises the noise of one sample to the same level as that of the other sample.

For instance, suppose that $e_1 \sim N(0, \sigma_1^2)$ and $e_2 \sim N(0, \sigma_2^2)$. Then the variances σ_1^2 and σ_2^2 can be estimated consistently using the R package 'fdapace' (Carroll et al., 2021) under both intensive and sparse designs with estimates $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$, respectively. Yao et al. (2005a) showed that

$$|\widehat{\sigma}_1 - \sigma_1| = O_p\left(\frac{1}{\sqrt{n}}\left(\frac{1}{h_{G_x}^2} + \frac{1}{h_{V_x}}\right)\right)$$

where h_{G_x} and h_{V_x} are the bandwidth parameters for estimating the covariance function $\mathrm{cov}(X(s), X(t))$ and the diagonal function $\mathrm{cov}(X(t), X(t)) + \sigma_1^2$, respectively. A different estimator for $\widehat{\sigma}_1$ that has a better convergence rate is provided in Lin and Wang (2022). An analogous result holds for $\widehat{\sigma}_2$. Then, by adding additional Gaussian white noise, we obtain the error-augmented data $\{\check{x}_{ij}\}, \{\check{y}_{ij}\}$ as follows:

$$\begin{cases} \begin{array}{l} \check{x}_{ij} = \widetilde{x}_{ij} + \epsilon_{ij} \quad \text{for } i = 1, 2, \ldots, n, j = 1, \ldots, N_i \\ \check{y}_{ij} = \widetilde{y}_{ij} \quad\quad \text{for } i = n + 1, \ldots, n + m, j = 1, \ldots, N_i \end{array} \left.\right\} \quad \text{if } \widehat{\sigma}_1^2 < \widehat{\sigma}_2^2 \\ \begin{array}{l} \check{x}_{ij} = \widetilde{x}_{ij} \quad\quad \text{for } i = 1, 2, \ldots, n, j = 1, \ldots, N_i \\ \check{y}_{ij} = \widetilde{y}_{ij} + \epsilon_{ij} \quad \text{for } i = n + 1, \ldots, n + m, j = 1, \ldots, N_i \end{array} \left.\right\} \quad \text{if } \widehat{\sigma}_1^2 > \widehat{\sigma}_2^2 \end{cases}$$

where $\{\epsilon_{ij}\} \sim^{\text{i.i.d}} N(0, |\widehat{\sigma}_2^2 - \widehat{\sigma}_1^2|)$. With $\check{\mathbf{Z}}$ being the combined error-augmented data, the proposed test is

$$\text{Reject } H_0 \quad \text{if } \check{p} \leq \alpha$$

where $\check{p} = (1/S)\{1 + \sum_{l=1}^{S-1} \mathbb{I}_{\{\mathrm{MED}_n(\Pi_l \cdot \check{\mathbf{Z}}) \geq \mathrm{MED}_n(\check{\mathbf{Z}})\}}\}$.

The normal error assumption is common in practice. The variance augmentation approach also works for any parametric family of distributions that is close under convolution (i.e., the sum of two independent distributions from this family is also a member of the family) and that has the property that the first two moments of a distribution determine the distribution.

3 Numerical studies

In this section, we examine the proposed testing procedure for both synthetic and real data sets.

3.1 Performance on simulated data

For simulations, we set $\alpha = 0.05$ and perform 500 Monte Carlo replications with 200 permutations for each test. We select the bandwidth parameters as $h_x = h_y = 0.1$ in estimating MED. Commonly used methods in the literature such as cross-validation and generalized cross-validation could be used. But they require more computing resources and their success is not guaranteed. Therefore, we adopted a simple default bandwidth, which uses 10% of the data range as specified in the R package fdapace (Carroll et al., 2021). This choice performed satisfactorily. For real data, the best practice is to explore various bandwidths, including subjective ones based on visual inspection of the data, and decide which one fits the data best without evidence of under or over fitting. The following example is used to examine the size of our test.

Example 1   The stochastic processes $\{X_i\}_{i=1}^n$ are i.i.d copies of X and $\{Y_i\}_{i=n+1}^{n+m}$ are i.i.d copies of Y, where for $t \in [0, 1]$,

$$X(t) = \xi_1\big(-\cos(2\pi t)\big) + \xi_2\big(\sin(2\pi t)\big)$$
$$Y(t) = \varsigma_1\big(-\cos(2\pi t)\big) + \varsigma_2\big(\sin(2\pi t)\big)$$

and $\xi_1, \xi_2, \varsigma_1, \varsigma_2, \sim^{\text{i.i.d}} N(0, 1)$. These curves are observed at discrete time points

$$\widetilde{x}_{ij} = X_i(T_{ij}) + e_{ij}, \quad i = 1, 2, \ldots, n, \ j = 1, 2, \ldots, N_i$$
$$\widetilde{y}_{ij} = Y_i(T_{ij}) + e_{ij}, \quad i = n + 1, \ldots, n + m, \ j = 1, 2, \ldots, N_i$$

where $\{T_{ij} : i = 1, 2, \ldots, n + m, j = 1, 2, \ldots, N_i\} \sim^{\text{i.i.d}}$ Uniform[0, 1] and the measurement errors are sampled from Gaussian distributions or scaled Student's t-distributions

$$\{e_{ij} : i = 1, 2, \ldots, n, j = 1, 2, \ldots, N_i\} \sim^{\text{i.i.d}} N(0, \sigma_1^2) \quad \text{or} \quad c_1 t_4$$

$$\{e_{ij} : i = n + 1, \ldots, n + m, j = 1, 2, \ldots, N_i\} \sim^{\text{i.i.d}} N(0, \sigma_2^2) \quad \text{or} \quad c_2 t_4$$

where t_4 follows Student's t-distribution with four degrees-of-freedom and c_1, c_2 are scaling constants such that $\text{var}(c_1 t_4) = \sigma_1^2$ and $\text{var}(c_2 t_4) = \sigma_2^2$.

The quantities $\{N_i\}$ are used to control the sparsity level. We consider three designs with different sparsity levels: N_i are uniformly selected from $\{1, 2\}$ (extremely sparse), $\{4, 5, 6\}$ (medium sparse), or $\{8, 9, 10\}$ (not so sparse) for all $i = 1, 2, \ldots, n + m$. The variances σ_1^2 and σ_2^2 quantify the magnitude of the measurement errors. We set $\sigma_1^2 = \sigma_2^2 = 0.2$ when $\sigma_1 = \sigma_2$ and $\sigma_1^2 = 0.05, \sigma_2^2 = 0.25$ when $\sigma_1 \neq \sigma_2$. If $\sigma_1 \neq \sigma^2$, the error augmentation method described in Section 2.3 is used. To examine whether our test works for noise-contaminated situations, we add Gaussian white noises when applying the error augmentation method regardless of whether or not the measurement errors follow a Gaussian distribution.

Table 1 contains the size comparison results under the sparse designs. As a baseline method for comparison, the functional principal component analysis (FPCA) approach is also included, where we first impute the principal scores and then apply the ED on the imputed scores. To be more specific, the first step is to reconstruct the principal scores using the R package 'fdapace' (Carroll et al., 2021). Then, a two-sample tests for multivariate data are applied on the recovered scores. For this, we choose the ED-based procedure and conduct the hypothesis testing via the R package 'energy' (Rizzo & Szekely, 2021). The FPCA approach has two drawbacks. First, the scores cannot be estimated consistently; second, the infinite dimensional vector of scores has to be truncated for computational purposes, which causes information loss. When the Gaussian errors have different variances, the same error-augmentation method is applied to the FPCA approach. From Table 1, we see that the MED-based methods have satisfactory size for all cases, while the FPCA approach leads to sizes larger than the nominal level for the very sparse case and sizes smaller than the nominal

Table 1. Comparison of size

Error	$n = m$	N_i	Without augmentation		With augmentation	
			FPCA	MED	FPCA	MED
Normal	150	1,2	0.464	0.044	0.422	0.062
		4,5,6	0.120	0.042	0.050	0.056
		8,9,10	0.002	0.037	0	0.048
	200	1,2	0.560	0.034	0.452	0.056
		4,5,6	0.136	0.055	0.028	0.060
		8,9,10	0	0.057	0	0.058
	300	1,2	0.626	0.052	0.452	0.074
		4,5,6	0.096	0.059	0.018	0.060
		8,9,10	0	0.037	0	0.052
Student t	150	1,2	0.580	0.058	0.488	0.077
		4,5,6	0.016	0.036	0.010	0.044
		8, 9,10	0	0.052	0	0.048
	200	1,2	0.606	0.054	0.526	0.060
		4,5,6	0.012	0.048	0.002	0.060
		8,9,10	0.002	0.042	0	0.042
	300	1,2	0.638	0.058	0.518	0.099
		4,5,6	0.012	0.060	0	0.066
		8,9,10	0	0.044	0	0.038

Note. MED = marginal energy distance.

level for the less sparse case, meaning that the size of the FPCA approach is often different from the nominal level. This may be due to the inherent inaccuracy in the imputed scores. When $\sigma_1 \neq \sigma_2$, we apply the error-augmentation approach and add additional Gaussian noises, even when the measurement errors follow scaled Student t-distributions. The closeness of the size and the nominal level 5% reveals the robustness of our test. The following example is used to examine the power of our test.

Example 2 The stochastic processes $\{X_i\}_{i=1}^{n}$ are i.i.d copies of X, $\{Y_i\}_{i=n+1}^{n+m}$ are i.i.d copies of Y, where for $t \in [0, 1]$,

$$X(t) = \xi_1\left(-\cos\left(2\pi t\right)\right) + \xi_2\left(\sin\left(2\pi t\right)\right)$$

$$Y(t) = \varsigma_1 t^2 + \varsigma_2\left(\sqrt{1 - t^4}\right)$$

$\xi_1, \xi_2 \sim^{\text{i.i.d}} N(0, 1)$, and ς_1, ς_2 are independently sampled from the following mixture of Gaussian distributions:

$$P(\varsigma \leq a) = \frac{1}{2}P(N(\mu_\varsigma, \sigma_\varsigma^2) \leq a) + \frac{1}{2}P(N(-\mu_\varsigma, \sigma_\varsigma^2) \leq a) \text{ for any } a \in \mathbb{R} \nsubseteq$$

The sampling plan is similar to Example 1.

By selecting μ_ς and σ_ς^2 such that $\mu_\varsigma^2 + \sigma_\varsigma^2 = 1$, we have $\text{var}(X(t)) = \text{var}(Y(t))$. Under this scenario, $X(s)$ and $Y(t)$ have the same marginal mean and variance, but different marginal distributions. In this example, we set $\mu_\varsigma = 0.98$, $\sigma_\varsigma = 0.199$, $\sigma_1^2 = \sigma_2^2 = 0.2$ when $\sigma_1 = \sigma_2$ and $\sigma_1^2 = 0.05$, $\sigma_2^2 = 0.25$ when $\sigma_1 \neq \sigma_2$. The power comparison results are provided in Table 2. Since the FPCA approach is often different from the nominal level, which makes the power comparison not so meaningful as a fair

Table 2. Comparison of power

		Without augmentation		With augmentation	
Error	$n = m$	MED (N_i) dense	MED (N_i) sparse	MED (N_i) dense	MED (N_i) sparse
Normal	150	1 (100)	0.269(1, 2)	1(100)	0.184(1, 2)
			0.894(4, 5, 6)		0.694(4, 5, 6)
			0.976(8, 9, 10)		0.904(8, 9, 10)
	200	1(100)	0.424(1, 2)	1(100)	0.290(1, 2)
			0.994(4, 5, 6)		0.876(4, 5, 6)
			0.998(8, 9, 10)		0.984(8, 9, 10)
	300	1(100)	0.836(1, 2)	1(100)	0.556(1, 2)
			1.000(4, 5, 6)		0.992(4, 5, 6)
			1.000(8, 9, 10)		1.000(8, 9, 10)
Student t	150	0.852(100)	0.130(1, 2)	0.906(100)	0.134(1, 2)
			0.372(4, 5, 6)		0.314(4, 5, 6)
			0.496(8, 9, 10)		0.426(8, 9, 10)
	200	0.986(100)	0.122(1, 2)	0.978(100)	0.148(1, 2)
			0.586(4, 5, 6)		0.512(4, 5, 6)
			0.760(8, 9, 10)		0.674(8, 9, 10)
	300	1 (100)	0.295(1, 2)	1 (100)	0.259(1, 2)
			0.892(4, 5, 6)		0.795(4, 5, 6)
			0.990(8, 9, 10)		0.948(8, 9, 10)

Note. MED = marginal energy distance.

comparison requires similar levels of actual sizes. We have thus not included FPCA in the power analysis. Instead, we compare the power of MED-based tests between sparse and intensively sampled data, where the sparse data are uniformly sampled from intensively sampled data. The same error-augmentation approach is applied to MED when $\sigma_1 \neq \sigma_2$, regardless of design intensities. When the measurement errors follow scaled Student t-distributions, MED-based tests have moderate power loss compared to intensively sampled data when $N_i = 150, 200$, but the powers catch up quickly when $N_i = 300$. We note that when the measurement errors follow scaled t-distributions with $\sigma_1 \neq \sigma_2$, the error-augmentation approach is applied by adding additional Gaussian noise to the observations. The high powers for medium and not so sparse cases demonstrate the robustness of our tests against additional noises. For the extremely sparse case ($N_i = 1, 2$) with Gaussian measurement errors, and as the sample sizes increase to 300, the power grows to 0.836 when $\sigma_1 = \sigma_2$ and to 0.556 when $\sigma_1 \neq \sigma_2$.

Example 3 The stochastic processes $\{X_i\}_{i=1}^n$ are i.i.d copies of X, $\{Y_i\}_{i=n+1}^{n+m}$ are i.i.d copies of Y, where for $t \in [0, 1]$, X is a Gaussian process with mean 0 and covariance structure $\text{cov}(X(s), X(t)) = \min\{s, t\}$, and Y is a Gaussian process with mean $\alpha(t + \sin(2\pi t))$ and covariance structure $\text{cov}(Y(s), Y(t)) = (1 - \beta)\min\{s, t\} + \beta \min\{1 - s, 1 - t\}$. When $\alpha = 0$ and $\beta > 0$, X and Y have the same mean, but difference covariance structure. When $\alpha > 0$ and $\beta = 0$, X and Y have the same covariance structure but different means. In the two-sample context, the magnitudes of α and β determine the level of deviation from the null hypothesis. Larger values of α, β would lead to easier testing problem so we can explore the power performance under various alternative hypotheses. We consider similar sampling designs to Example 1 with Gaussian measurement errors.

For this example, we consider the case $n = m = 200$ with noise level $\sigma_1^2 = \sigma_2^2 = 0.2$. We compare the results for both sparsely ($N_i \in \{4, 5, 6, 7, 8\}$) and intensively sampled data ($N_i = 100$), from
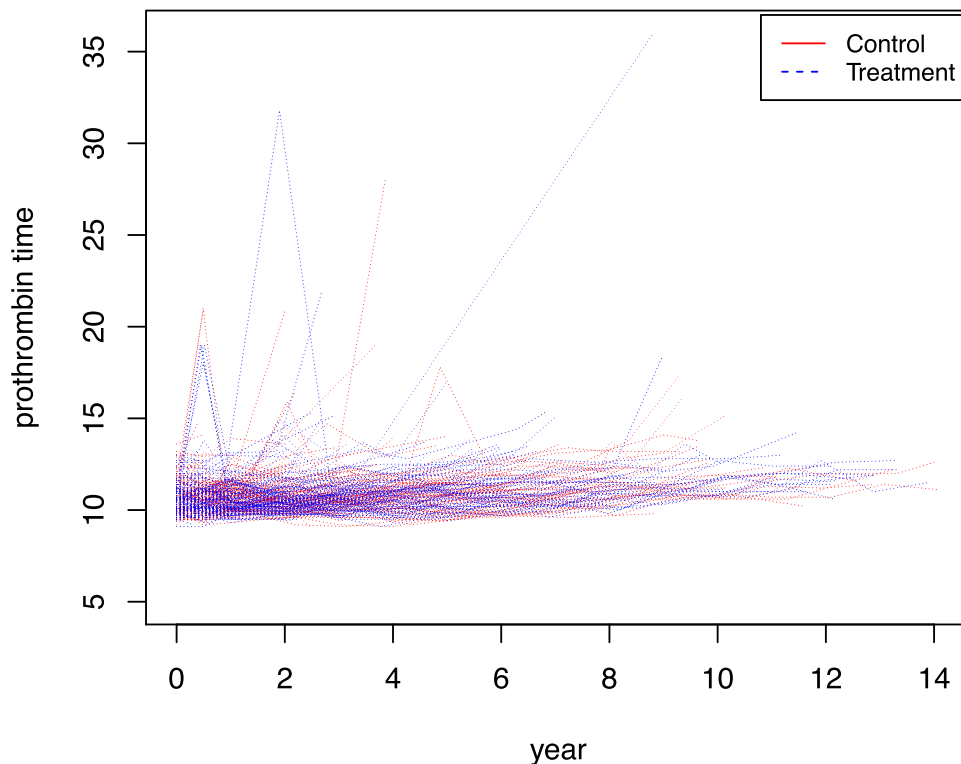
Figure 1. Power study w.r.t. Example 3.

which the sparse data are sampled. The simulation results are shown in Figure 1, which reveals that MED-based approach enjoys strong power growth for intensively sampled data, and comparable power growth for sparsely sample data (see the left panel of Figure 1) when testing the equal-mean hypothesis. The result in the right panel of Figure 1 suggests that the sampling frequency has a more prominent role on the power of testing equal-covariance.

3.2 Applications to real data

In this subsection, we apply the proposed MED-based tests to two real data sets. *PBC data*: The first data set is the PBC data from Mayo Clinic (Fleming & Harrington, 2005). This data set is from a clinical trial studying PBC of the liver. There were 312 patients assigned to either the treatment or control group. The drug D-penicillamine is given to the treatment group. Here, we are interested in testing the equality of the marginal distributions of prothrombin time, which is a blood test that measures how long it takes blood to clot. The trajectories of prothrombin time for different subjects are plotted in Figure 2, and there are on average six measurements per subject. For our tests, the bandwidth is set to be 2. Here, the equal distribution assumption for errors seems to work (the estimated variances for treatment and control group are 0.96 and 1.009, respectively). By using 200 permutations, the *p*-value of the MED-based test is 0.54, which means that there is not enough evidence to conclude that the marginal distributions of prothrombin time are different between the two groups. This conclusion matches with existing knowledge that D-penicillamine is ineffective to treat PBC of liver.

Strawberry data: In the food industry, there is a continuing interest in distinguishing the pure fruit purees from the adulterated ones (Holland et al., 1998). One practical way to detect adulteration is by looking at the spectra of the fruit purees. Here, we are interested in testing the marginal distribution between the spectra of strawberry purees (authentic samples) and non-strawberry purees (adulterated strawberries and other fruits). The strawberry data can be downloaded from the UCR Time Series Classification Archive (Dau et al., 2018; https://www.cs.ucr.edu/~eamonn/time_series_data_2018/). The single-beam spectra of the purees were normalized to
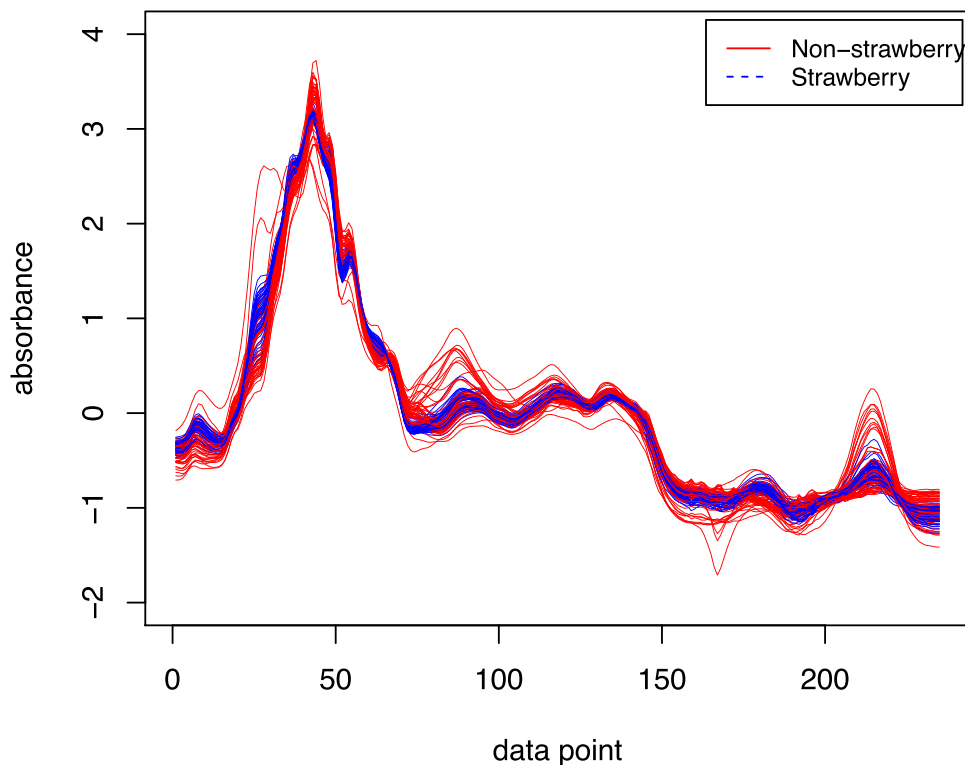
Figure 2. Trajectories of real data.

back-ground spectra of water and then transformed into absorbance units. The spectral range was truncated to 899–1802 cm^{-1} (235 data points). The two samples of spectra are plotted in Figure 2 and more information about this data set can be found at Holland et al. (1998). The estimated variances of the measurement errors are 0.000279 and 0.00031 for the two samples, which indicates that there are practically no measurement errors. To check the performance of our method, we analyse the data using all 235 measurements as well as sparse subsamples that contain 2–10 observations per subject. The R package 'energy' is applied for the complete data. Both tests are conducted with 200 permutations and have p-value 0.005. Thus, we have strong evidence to conclude that the marginal distributions between the spectra of strawberry and non-strawberry purees are significantly different and our test produced similar results regardless of the sampling plan.

4 Conclusion

The literature on testing homogeneity for functional data is scarce probably because most approaches rely on intensive measurement schedules and the hope that measurement errors could be addressed by presmoothing the data. Since reconstruction of noise-free functional data is not feasible for sparsely observed functional data, a test of homogeneity is infeasible. In this work, we show what is feasible for sparse functional data, a.k.a. longitudinal data, and propose a test of marginal homogeneity that adapts to the sampling plan and provides the corresponding convergence rate. Our test is based on ED with a focus on testing the marginal homogeneity. To the best of our knowledge, this is the only nonparametric test with theoretical guarantees under sparse designs, which are ubiquitous.

There are several twists in our approach, including the handling of asynchronized longitudinal data and the unconventional way that measurement errors affect the method and theory. The asynchronization of the data can be overcome completely as we demonstrated in Section 2.1, but the handling of measurement errors requires some compromise when the distributions of the measurement errors are different for the two samples. This is the price one pays for lack of data and is not due to the use of the L_1 norm associated with testing the marginal homogeneity, as an L^2 norm for testing full homogeneity would also face the same challenge with measurement errors unless a presmoothing step has been

employed to eliminate the measurement errors. As we mentioned in Section 1, this would require a super intensive sampling plan well beyond the usual requirement for dense or ultra-dense functional data (Zhang & Wang, 2016). While the new approach may involve error augmentation, numerical results show that the efficiency loss is minimal. Moreover, such an augmentation strategy is not uncommon. For instance, an error augmentation method has also been adopted in the SIMEX approach (Cook & Stefanski, 1994) to deal with measurement errors for vector data.

While testing marginal homogeneity has its own merits and advantages over a full-fledged test of homogeneity, our intention is not to particularly endorse it. Rather, we point out what is feasible and infeasible for sparsely or intensively measured functional data and develop theoretical support for the proposed test. To the best of our knowledge, we are the first to provide the convergence rate for the permuted statistics for sparse functional data. This proof and the proof of consistency for the proposed permutation test are non-conventional and different from the multivariate/high-dimensional case.

5 Technical details

5.1 Proof of Theorem 1

Here, we show that

$$\|\widehat{G}_1 - G_1\|_2 = O_p\left(h_x^2 + \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}\phi_i} + h_y^2 + \sqrt{\frac{1}{n^2}\sum_{i=n+1}^{n+m}\phi_i} \right)$$

$$\|\widehat{G}_2 - G_2\|_2 = O_p\left(h_x^2 + \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}\phi_i} \right)$$

and $\|\widehat{G}_3 - G_3\|_2$ can be bounded similarly. For $p_1, p_2 = 0, 1, 2$, set

$$\mathcal{T}_{i_1 i_2}^{p_1 p_2}(t_1, t_2) = \frac{1}{N_{i_1}}\frac{1}{N_{i_2}}\sum_{j_1=1}^{N_{i_1}}\sum_{j_2=1}^{N_{i_2}} K_{h_{i_1}}\left(T_{i_1 j_1} - t_1 \right) K_{i_2}\left(T_{i_2 j_2} - t_2 \right)$$

$$\left(\frac{T_{i_1 j_1} - t_1}{h_{i_1}} \right)^{p_1} \left(\frac{T_{i_2 j_2} - t_2}{h_{i_2}} \right)^{p_2}$$

where $h_i = h_x$, if $1 \le i \le n$; otherwise $h_i = h_y$. The weighted raw data are denoted as

$$\mathcal{Z}_{i_1 i_2}^{p_1 p_2}(t_1, t_2) = \frac{1}{N_{i_1}}\frac{1}{N_{i_2}}\sum_{j_1=1}^{N_{i_1}}\sum_{j_2=1}^{N_{i_2}} K_{h_{i_1}}\left(T_{i_1 j_1} - t_1 \right) K_{h_{i_2}}\left(T_{i_2 j_2} - t_2 \right)$$

$$\left(\frac{T_{i_1 j_1} - t_1}{h_{i_1}} \right)^{p_1} \left(\frac{T_{i_2 j_2} - t_2}{h_{i_2}} \right)^{p_2} |z_{i_1 j_1} - z_{i_2 j_2}|$$

If there is no confusion, we will only write $\mathcal{T}_{i_1 i_2}, \mathcal{Z}_{i_1 i_2}$ instead of $\mathcal{T}_{i_1 i_2}^{p_1 p_2}(t_1, t_2), \mathcal{Z}_{i_1 i_2}^{p_1 p_2}(t_1, t_2)$ for simplicity. Both (5) and (6) admit closed form solutions and some algebra show that for $I = 1, 2$

$$\widehat{G}_I(t_1, t_2) = \frac{W_{I,1}(t_1, t_2)V_I^{0,0}(t_1, t_2) - W_{I,2}(t_1, t_2)V_I^{1,0}(t_1, t_2) + W_{I,3}(t_1, t_2)V_I^{0,1}(t_1, t_2)}{W_{I,1}(t_1, t_2)U_I^{0,0}(t_1, t_2) - W_{I,2}(t_1, t_2)U_I^{1,0}(t_1, t_2) + W_{I,3}(t_1, t_2)U_I^{0,1}(t_1, t_2)}$$

Here, $\{W_{I,J} : I = 1, 2, J = 1, 2, 3\}$ are two-dimensional functions defined as

$$W_{I,1}(t_1, t_2) = (U_I^{2,0}(t_1, t_2)U_I^{0,2}(t_1, t_2)) - (U_I^{1,1}(t_1, t_2))^2$$

$$W_{I,2}(t_1, t_2) = (U_I^{1,0}(t_1, t_2)U_I^{0,2}(t_1, t_2)) - U_I^{0,1}(t_1, t_2)U_I^{1,1}(t_1, t_2)$$

$$W_{I,3}(t_1, t_2) = (U_I^{1,0}(t_1, t_2)U_I^{1,1}(t_1, t_2)) - U_I(t_1, t_2; 0, 1)U_I^{2,0}(t_1, t_2)$$

where for $p_1, p_2 = 0, 1, 2$, $\{U_I, V_I : I = 1, 2\}$ have the following expressions:

$$U_1^{p_1 p_2}(t_1, t_2) = \frac{1}{nm} \sum_{1 \leq i_1 \leq n} \sum_{n+1 \leq i_2 \leq n+m} \mathcal{T}_{i_1 i_2}^{p_1 p_2}(t_1, t_2)$$

$$U_2^{p_1 p_2}(t_1, t_2) = \frac{2}{n(n-1)} \sum_{1 \leq i_1 < i_2 \leq n} \mathcal{T}_{i_1 i_2}^{p_1 p_2}(t_1, t_2)$$

$$V_1^{p_1 p_2}(t_1, t_2) = \frac{1}{nm} \sum_{1 \leq i_1 \leq n} \sum_{n+1 \leq i_2 \leq n+m} \mathcal{Z}_{i_1 i_2}^{p_1 p_2}(t_1, t_2)$$

$$V_2^{p_1 p_2}(t_1, t_2) = \frac{2}{n(n-1)} \sum_{1 \leq i_1 < i_2 \leq n} \mathcal{Z}_{i_1 i_2}^{p_1 p_2}(t_1, t_2)$$

By some straight-forward calculations, for $I = 1, 2$

$$
\begin{aligned}
&\widehat{G}_I(t_1, t_2) - G_I(t_1, t_2) \\
&= \frac{W_{I,1}(t_1, t_2)F_I^{0,0}(t_1, t_2) - W_{I,2}(t_1, t_2)F_I^{1,0}(t_1, t_2) + W_{I,3}(t_1, t_2)F_I^{0,1}(t_1, t_2)}{W_{I,1}(t_1, t_2)U_I^{0,0}(t_1, t_2) - W_{I,2}(t_1, t_2)U_I^{1,0}(t_1, t_2) + W_{I,3}(t_1, t_2)U_I^{0,1}(t_1, t_2)}
\end{aligned}
\tag{13}
$$

where for $p_1, p_2 = 0, 1, 2$,

$$
\begin{aligned}
F_1^{p_1 p_2}(t_1, t_2) = {}& V_1^{p_1 p_2}(t_1, t_2) - G_1(t_1, t_2)U_1^{p_1 p_2}(t_1, t_2) \\
& - h_x \frac{\partial G_1(t_1, t_2)}{\partial t_1} U_1^{p_1+1, p_2}(t_1, t_2) - h_y \frac{\partial G_1(t_1, t_2)}{\partial t_2} U_1^{p_1, p_2+1}(t_1, t_2)
\end{aligned}
$$

and

$$
\begin{aligned}
F_2^{p_1 p_2}(t_1, t_2) = {}& V_2^{p_1 p_2}(t_1, t_2) - G_2(t_1, t_2)U_2^{p_1 p_2}(t_1, t_2) \\
& - h_x \frac{\partial G_2(t_1, t_2)}{\partial t_1} U_2^{p_1+1, p_2}(t_1, t_2) - h_x \frac{\partial G_2(t_1, t_2)}{\partial t_2} U_2^{p_1, p_2+1}(t_1, t_2)
\end{aligned}
$$

Lemma 3 entails that the denominator in (13) is bounded away from 0 with high probability. Thus, for $I = 1, 2$, we have

$$\|\widehat{G}_I - G_I\|_2 = O_p\left(\|F_I^{0,0}\|_2 + \|F_I^{1,0}\|_2 + \|F_I^{0,1}\|_2\right)$$

It can be easily seen that $E[|z_{i_1 j_1} - z_{i_2 j_2}| \mid T_{i_1 j_1}, T_{i_2 j_2}] = G_1(T_{i_1 j_1}, T_{i_2 j_2})$ if $1 \leq i_1 \leq n, n+1 \leq i_2 \leq n + m$ and $E[|z_{i_1 j_1} - z_{i_2 j_2}| \mid T_{i_1 j_1}, T_{i_2 j_2}] = G_2(T_{i_1 j_1}, T_{i_2 j_2})$ if $1 \leq i_1 < i_2 \leq n$. By Taylor expansion

$$\|F_I^{p_1 p_2}\|_2 = O(\|L_I^{p_1 p_2}\|_2 + h_x^2 + h_y^2 + h_x h_y)$$

where

$$L_1^{p_1 p_2}(t_1, t_2) = \frac{1}{nm} \sum_{1 \leq i_1 \leq n} \sum_{n+1 \leq i_2 \leq n+m} \mathcal{Z}_{i_1 i_2}^{p_1 p_2} - E\left[\mathcal{Z}_{i_1 i_2}^{p_1 p_2} \mid \{T_{i_1 j_1}\}_{j_1=1}^{N_{i_1}}, \{T_{i_2 j_2}\}_{j_2=1}^{N_{i_2}}\right]$$

$$L_2^{p_1 p_1}(t_1, t_2) = \frac{2}{n(n-1)} \sum_{1 \leq i_1 < i_2 \leq n} \mathcal{Z}_{i_1 i_2}^{p_1 p_2} - E\left[\mathcal{Z}_{i_1 i_2}^{p_1 p_2} \mid \{T_{i_1 j_1}\}_{j_1=1}^{N_{i_1}}, \{T_{i_2 j_2}\}_{j_2=1}^{N_{i_2}}\right]$$

Lemma 1 Under Assumptions 1 and 2, for $I = 1, 2$ and $p_1, p_2 = 0, 1, 2$, it holds that

$$E\left[\|L_I^{p_1 p_2}\|_2^2\right] \lesssim \frac{1}{n^2}\sum_{i=1}^{n}\phi_i + \frac{1}{m^2}\sum_{i=n+1}^{n+m}\phi_i$$

Proof. By setting

$$w_{i_1 i_2, j_1 j_2}^{p_1 p_2}(t_1, t_2) = K_{b_{i_1}}(T_{i_1 j_1} - t_1)K_{b_{i_2}}(T_{i_2 j_2} - t_2)\left(\frac{T_{i_1 j_1} - t_1}{b_{i_1}}\right)^{p_1}\left(\frac{T_{i_2 j_2} - t_2}{b_{i_2}}\right)^{p_2}$$

$$\left\{\left|z_{i_1 j_1} - z_{i_2 j_2}\right| - E\left[\left|z_{i_1 j_1} - z_{i_2 j_2}\right|\big|T_{i_1 j_1}, T_{i_2 j_2}\right]\right\}$$

we can write

$$\overline{\mathcal{Z}}_{i_1 i_2}^{p_1 p_2} = \mathcal{Z}_{i_1 i_2}^{p_1 p_2} - E\left[\mathcal{Z}_{i_1 i_2}^{p_1 p_2} \mid \{T_{i_1 j_1}\}_{j_1=1}^{N_{i_1}}, \{T_{i_2 j_2}\}_{j_2=1}^{N_{i_2}}\right]$$

$$= \frac{1}{N_{i_1}}\frac{1}{N_{i_2}}\sum_{j_1=1}^{N_{i_1}}\sum_{j_2=1}^{N_{i_2}}w_{i_1 i_2, j_1 j_2}^{p_1 p_2}(t_1, t_2)$$

For any $p_1, p_2 = 0, 1, 2$,

$$E\left[w_{i_1 i_2, j_1 j_2}^{p_1 p_2}(t_1, t_2)\right] = E\left[E\left[w_{i_1 i_2}(t_1, t_2)\big|T_{i_1 j_1}, T_{i_2 j_2}\right]\right] = 0$$

which implies that $E[\overline{\mathcal{Z}}_{i_1 i_2}^{p_1 p_2}] = 0$ and $E[\overline{\mathcal{Z}}_{i_1 i_2}^{p_1 p_2}\overline{\mathcal{Z}}_{i_1' i_2'}^{p_1 p_2}] \neq 0$, only if $\{i_1, i_2\} \cap \{i_1', i_2'\} \neq \emptyset$. If $1 \leq i_1, i_1' \leq n$, $n+1 \leq i_2, i_2' \leq n+m$, $i_1 = i_1'$, we have $b_{i_1} = b_{i_1'}$, $b_{i_2} = b_{i_2'}$ and

$$h_{i_1}^2 h_{i_2}^2\left|\iint E\left[w_{i_1 i_2, j_1 j_2}(t_1, t_2)w_{i_1' i_2', j_1' j_2'}(t_1, t_2)\right]dt_1\,dt_2\right|$$

$$\leq C \times \begin{cases} h_{i_1}h_{i_2}^2, & j_1 = j_1', \ i_2 \neq i_2' \\ h_{i_1}h_{i_2}^2, & j_1 = j_1', \ i_2 = i_2', \ j_2 \neq j_2' \\ h_{i_1}h_{i_2}, & j_1 = j_1', \ i_2 = i_2', \ j_2 = j_2' \\ h_{i_1}^2 h_{i_2}^2, & j_1 \neq j_1', \ i_2 \neq i_2' \\ h_{i_1}^2 h_{i_2}^2, & j_1 \neq j_1', \ i_2 = i_2', \ j_2 \neq j_2' \\ h_{i_1}^2 h_{i_2}, & j_1 \neq j_1', \ i_2 = i_2', \ j_2 = j_2' \end{cases}$$

where C is a constant that depends on T_x, T_y, K, X, Y. Using the above inequality, if $i_1 = i_1'$ and $i_2 = i_2'$, we have

$$\left|\iint E\left[\overline{\mathcal{Z}}_{i_1 i_2}(t_1, t_2)\overline{\mathcal{Z}}_{i_1' i_2'}(t_1, t_2)\right]dt_1\,dt_2\right|$$

$$\leq \frac{1}{N_{i_1}}\frac{1}{N_{i_2}}\sum_{j_1=1}^{N_{i_1}}\sum_{j_2=1}^{N_{i_2}}\frac{1}{N_{i_1'}}\frac{1}{N_{i_2'}}\sum_{j_1'=1}^{N_{i_1'}}\sum_{j_2'=1}^{N_{i_2'}}\left|\iint E\left[w_{i_1 i_2, j_1 j_2}(t_1, t_2)w_{i_1' i_2', j_1' j_2'}(t_1, t_2)\right]dt_1\,dt_2\right|$$

$$\leq C\left\{\frac{N_{i_1}h_{i_1} + N_{i_1}(N_{i_1} - 1)h_{i_1}^2}{N_{i_1}^2 h_{i_1}^2}\frac{N_{i_2}h_{i_2} + N_{i_2}(N_{i_2} - 1)h_{i_2}^2}{N_{i_2}^2 h_{i_2}^2}\right\}$$

$$= C\phi_{i_1}\phi_{i_2}$$

$$(14)$$

If $i_1 = i'_1$, $i_2 \neq i'_2$, we have

$$\left| \iint E\left[\overline{\mathcal{Z}}_{i_1 i_2}(t_1, t_2) \overline{\mathcal{Z}}_{i'_1 i'_2}(t_1, t_2) \right] dt_1 \, dt_2 \right|$$
$$\leq C \left\{ \frac{N_{i_1} b_1 + N_{i_1}(N_{i_1} - 1) b_1^2}{N_{i_1}^2 b_1^2} \right\} \frac{N_{i_2} N_{i'_2} b_2^2}{N_{i_2} N_{i'_2} b_2^2} = C\phi_{i_1} \tag{15}$$

Similarly, for the case $i_1 \neq i'_1$ and $i_2 = i'_2$, it holds that

$$\left| \iint E\left[\overline{\mathcal{Z}}_{i_1 i_2}(t_1, t_2) \overline{\mathcal{Z}}_{i'_1 i'_2}(t_1, t_2) \right] dt_1 \, dt_2 \right| \leq C\phi_{i_2} \tag{16}$$

Then, it follows from the above inequalities that

$$E\left[\int_0^1\int_0^1 L_1(t_1, t_2)^2 \, dt_1 \, dt_2 \right]$$

$$\lesssim \frac{1}{n^2 m^2} \sum_{1 \leq i_1 \leq n} \sum_{n+1 \leq i_2 \leq n+m} \sum_{1 \leq i'_1 \leq n} \sum_{n+1 \leq i'_2 \leq n+m} \left| \iint E\left[\overline{\mathcal{Z}}_{i_1 i_2}(t_1, t_2) \overline{\mathcal{Z}}_{i'_1 i'_2}(t_1, t_2) \right] dt_1 \, dt_2 \right|$$

$$\lesssim \frac{1}{n^2} \sum_{1 \leq i_1 \leq n} \phi_{i_1} + \frac{1}{m^2} \sum_{n+1 \leq i_2 \leq n+m} \phi_{i_2} + \frac{1}{n^2 m^2} \sum_{1 \leq i_1 \leq n} \sum_{n+1 \leq i_2 \leq n+m} \phi_{i_1} \phi_{i_2}$$

$$\lesssim \frac{1}{n^2} \sum_{1 \leq i_1 \leq n} \phi_{i_1} + \frac{1}{m^2} \sum_{n+1 \leq i_2 \leq n+m} \phi_{i_2}$$

The bound for $E[\int_0^1\int_0^1 L_2(t_1, t_2)^2 \, dt_1 \, dt_2]$ can be derived using the same tactics. $\square$

5.2 Proof of Theorem 2

For any permutation π, let $\widehat{G}_{\pi,I}$ be the estimated function based on the permuted sample

$$\pi \cdot \mathbf{Z} = (\mathbf{z}_{\pi(1)}; \mathbf{z}_{\pi(2)}; \ldots; \mathbf{z}_{\pi(n+m)})$$

Correspondingly, the explicit form of $\widehat{G}_{\pi,I}$ depends on the following quantities:

$$U_{\pi,1}^{p_1 p_2}(t_1, t_2) = \frac{1}{nm} \sum_{1 \leq i_1 \leq n} \sum_{n+1 \leq i_2 \leq n+m} \mathcal{T}_{\pi(i_1)\pi(i_2)}$$

$$U_{\pi,2}^{p_1 p_2}(t_1, t_2) = \frac{2}{n(n-1)} \sum_{1 \leq i_1 < i_2 \leq n} \mathcal{T}_{\pi(i_1)\pi(i_2)}$$

$$V_{\pi,1}^{p_1 p_2}(t_1, t_2) = \frac{1}{nm} \sum_{1 \leq i_1 \leq n} \sum_{n+1 \leq i_2 \leq n+m} \mathcal{Z}_{\pi(i_1)\pi(i_2)}$$

$$V_{\pi,2}^{p_1 p_2}(t_1, t_2) = \frac{2}{n(n-1)} \sum_{1 \leq i_1 < i_2 \leq n} \mathcal{Z}_{\pi(i_1)\pi(i_2)}$$

For $I = 1, 2$, $J = 1, 2, 3$, let $W_{\pi,I,J}^{p_1 p_2}$ be defined similarly with $W_{I,J}^{p_1 p_2}$ with U_I, V_I replaced by $U_{\pi,I}, V_{\pi,I}$, respectively. Then it can be shown that $\widehat{G}_{\pi,I}(t_1, t_2)$ admits a similar form as $\widehat{G}_I(t_1, t_2)$ with W_I, U_I, V_I replaced by $W_{\pi,I}, U_{\pi,I}, V_{\pi,I}$.

The following lemma shows that $U_{\pi,I}, V_{\pi,I}$ as well as $U_{\Pi,I}, V_{\Pi,I}$ converge to their mean functions, where Π is a random permutation sampled uniformly from $\mathbb{P}_{n+m}$.

Lemma 2   Under the assumptions of Theorem 2, for any $p_1, p_2 = 0, 1, 2$, we have

(i) for any fixed permutation $\pi \in \mathbb{P}_{n+m}$,

$$\iint E\left(V_{\pi,I}^{p_1 p_2}(t_1, t_2) - E\left[V_{\pi,I}^{p_1 p_2}(t_1, t_2)\right]\right)^2 dt_1\, dt_2$$

$$\leq C\left\{\frac{1}{n^2}\sum_{1\leq i_1 \leq n}\phi_{\pi(i_1)} + \frac{1}{m^2}\sum_{n+1\leq i_2 \leq n+m}\phi_{\pi(i_2)}\right\}$$

where C is a constant that depends on T_x, T_y, K, X, Y. Moreover, $\iint E(U_{\pi,I}^{p_1 p_2}(t_1, t_2) - E[U_{\pi,I}^{p_1 p_2}(t_1, t_2)])^2 dt_1\, dt_2$ satisfies the same bound.

(ii) For any random permutation Π sampled from $\mathbb{P}_{n+m}$ uniformly,

$$\iint E\left(V_{\Pi,I}^{p_1 p_2}(t_1, t_2) - E\left[V_{\Pi,I}^{p_1 p_2}(t_1, t_2)\right]\right)^2 dt_1\, dt_2$$

$$\leq C\left\{\sup_{\pi \in \mathbb{P}_{n+m}}\frac{1}{n^2}\sum_{1\leq i_1 \leq n}\phi_{\pi(i_1)} + \sup_{\pi \in \mathbb{P}_{n+m}}\frac{1}{m^2}\sum_{n+1\leq i_2 \leq n+m}\phi_{\pi(i_2)}\right\}$$

where C is a constant that depends on T_x, T_y, K, X, Y. Moreover, $\iint E(U_{\Pi,I}^{p_1 p_2}(t_1, t_2) - E[U_{\Pi,I}^{p_1 p_2}(t_1, t_2)])^2 dt_1\, dt_2$ can attain the same rate as above.

Proof. (i)   Here, we only show the result for $V_{\pi,1}$, as the proof for $V_{\pi,2}$ is similar. Set $\widetilde{V}_{\pi,1} = (1/nm)\sum_{1\leq i_1 \leq n}\sum_{n+1\leq i_2 \leq n+m}\widetilde{\mathcal{Z}}_{\pi(i_1)\pi(i_2)}$, where

$$\widetilde{\mathcal{Z}}_{\pi(i_1)\pi(i_2)} = \mathcal{Z}_{\pi(i_1)\pi(i_2)} - E[\mathcal{Z}_{\pi(i_1)\pi(i_2)}]$$

Since $E[\widetilde{\mathcal{Z}}_{\pi(i_1)\pi(i_2)}\widetilde{\mathcal{Z}}_{\pi(i_1')\pi(i_2')}] \neq 0$ only if $\{\pi(i_1), \pi(i_2)\} \cap \{\pi(i_1'), \pi(i_2')\} \neq \emptyset$, by similar arguments with equations (14), (15), and (16), we can show that

$$\left|\iint E\left[\widetilde{\mathcal{Z}}_{i_1 i_2}(t_1, t_2)\widetilde{\mathcal{Z}}_{i_1' i_2'}(t_1, t_2)\right] dt_1\, dt_2\right| \leq C\begin{cases}\phi_{\pi(i_1)}\phi_{\pi(i_2)} & \text{if } i_1 = i_1', i_2 = i_2' \\ \phi_{\pi(i_1)} & \text{if } i_1 = i_1', i_2 \neq i_2' \\ \phi_{\pi(i_2)} & \text{if } i_1 \neq i_1', i_2 = i_2'\end{cases}$$

which implies

$$\iint E\left(V_{\pi,I}^{p_1 p_2}(t_1, t_2) - E\left[V_{\pi,I}^{p_1 p_2}(t_1, t_2)\right]\right)^2 dt_1\, dt_2$$

$$\leq 2C\left\{\frac{1}{n^2}\sum_{1\leq i_1 \leq n}\phi_{\pi(i_1)} + \frac{1}{m^2}\sum_{n+1\leq i_2 \leq n+m}\phi_{\pi(i_2)}\right\}$$

The bound on $\iint E(U_{\pi,I}^{p_1 p_2}(t_1, t_2) - E[U_{\pi,I}^{p_1 p_2}(t_1, t_2)])^2 dt_1\, dt_2$ can be shown similarly.

(ii) The result follows from the inequality:

$$\iint E\left(V_{\pi,I}^{p_1 p_2}(t_1, t_2) - E\left[V_{\pi,I}^{p_1 p_2}(t_1, t_2)\right]\right)^2 dt_1\, dt_2$$

$$\leq 2C\left\{\sup_{\pi \in \mathbb{P}_{n+m}}\frac{1}{n^2}\sum_{1\leq i_1 \leq n}\phi_{\pi(i_1)} + \sup_{\pi \in \mathbb{P}_{n+m}}\frac{1}{m^2}\sum_{n+1\leq i_2 \leq n+m}\phi_{\pi(i_2)}\right\}$$

and the fact that C is a constant independent of π.   □

Lemma 3 Under the assumptions of Theorem 2, for any $p_1, p_2 = 0, 1, 2$, we have

    (i) for any fixed permutation $\pi \in \mathbb{P}_{n+m}$,

$$\sup_{t_1, t_2} \left| U_{\pi,I}^{p_1 p_2}(t_1, t_2) - E\left[U_{\pi,I}^{p_1 p_2}(t_1, t_2) \right] \right| = o_p(1)$$

    (ii) For any random permutation Π sampled from $\mathbb{P}_{n+m}$ uniformly,

$$\sup_{t_1, t_2} \left| U_{\Pi,I}^{p_1 p_2}(t_1, t_2) - E\left[U_{\Pi,I}^{p_1 p_2}(t_1, t_2) \right] \right| = o_p(1)$$

Proof. (i) We can write $\mathcal{T}_{i_1 i_2}^{p_1 p_2} = \mathcal{S}_{i_1}^{p_1} \mathcal{S}_{i_2}^{p_2}$ and

$$U_{\pi,1}^{p_1 p_2}(t_1, t_2) = R_{\pi,1}^{p_1}(t_1) R_{\pi,2}^{p_2}(t_2)$$

$$U_{\pi,2}^{p_1 p_2}(t_1, t_2) = R_{\pi,1}^{p_1}(t_1) R_{\pi,1}^{p_1}(t_1) - \frac{1}{n^2} \sum_{i_1=1}^{n} (\mathcal{S}_{\pi(i_1)}^{p_1}(t_1))^2$$

where $R_{\pi,1}^{p_1}(t_1) = (1/n) \sum_{i_1=1}^{n} \mathcal{S}_{\pi(i_1)}^{p_1}(t_1)$ and $R_{\pi,2}^{p_2}(t_2) = (1/m) \sum_{i_2=n+1}^{n+m} \mathcal{S}_{\pi(i_2)}^{p_2}(t_2)$ and

$$\mathcal{S}_{i_1}^{p_1}(t_1) = \frac{1}{N_{i_1}} \sum_{j_1=1}^{N_{i_1}} K_{b_{i_1}}(T_{i_1 j_1} - t_1) \left(\frac{T_{i_1 j_1} - t_1}{b_{i_1}} \right)^{p_1}$$

$$\mathcal{S}_{i_2}^{p_2}(t_2) = \frac{1}{N_{i_2}} \sum_{j_2=1}^{N_{i_2}} K_{b_{i_2}}(T_{i_2 j_2} - t_2) \left(\frac{T_{i_2 j_2} - t_2}{b_{i_2}} \right)^{p_2}$$

It is sufficient to show that $\sup_{t_1} |R_{\pi,I}(t_1) - E[R_{\pi,I}(t_1)]| = o_p(1)$. Here, we only provide details for the case $I = 1$ and the rest can be shown similarly. The superscripts p_1, p_2 will be dropped if there is no confusion. Set

$$\widetilde{R}_{\pi,1} = \frac{1}{n} \sum_{1 \le i_1 \le n} \widetilde{\mathcal{S}}_{\pi(i_1)}$$

where $\widetilde{\mathcal{S}}_{\pi(i_1)} = \mathcal{S}_{\pi(i_1)} - E[\mathcal{S}_{\pi(i_1)}]$ and

$$a_n = \left\{ \log\left(\frac{n}{\sum_{i=1}^{n} N_{\pi(i)}^{-1}/n} \right) \frac{\sum_{i=1}^{n} N_{\pi(i)}^{-1}/n}{\min\{h_x, h_y\}} \frac{1}{n} \right\}^{1/2}, \qquad b_n = \frac{1}{(n^{-2} \sum_{i=1}^{n} N_{\pi(i)}^{-1})^2}$$

Let $\chi(b_n) \subset [0, 1]$ be the set of equally spaced grid points of size b_n, then

$$\sup_{t_1} |\widetilde{R}_{\pi,1}(t_1)| \lesssim \sup_{t_1 \in \chi(b_n)} |\widetilde{R}_{\pi,1}(t_1)| + \sup_{|t_1 - t_1'| < 1/b_n} |\widetilde{R}_{\pi,1}(t_1) - \widetilde{R}_{\pi,1}(t_1')|$$

To bound the second term, note that

$$|\widetilde{R}_{\pi,1}(t_1) - \widetilde{R}_{\pi,1}(t_1')|$$

$$\le \frac{1}{n} \sum_{1 \le i_1 \le n} \left\{ \left| \mathcal{S}_{\pi(i_1)}(t_1) - \mathcal{S}_{\pi(i_1)}(t_1') \right| + \left| E\left[\mathcal{S}_{\pi(i_1)}(t_1) \right] - E\left[\mathcal{S}_{\pi(i_1)}(t_1') \right] \right| \right\}$$

For any t_1, t_1' such that $|t_1 - t_1'| < 1/b_n$,

$$\sup_{|t_1-t_1'|<1/b_n} \frac{1}{n} \sum_{1 \le i_1 \le n} |\mathcal{S}_{\pi(i_1)}(t_1) - \mathcal{S}_{\pi(i_1)}(t_1')|$$

$$\le C_K \sup_{|t_1-t_1'|<1/b_n} \frac{1}{n} \sum_{1 \le i_1 \le n} \frac{1}{N_{\pi(i_1)}} \sum_{j_1=1}^{N_{\pi(i_1)}} \left\{ \left| K_{b_{\pi(i_1)}}(T_{\pi(i_1)j_1} - t_1) - K_{b_{\pi(i_1)}}(T_{\pi(i_1)j_1} - t_1') \right| \right.$$

$$\left. + \left| \left(\frac{T_{\pi(i_1)j_1} - t_1}{b_{\pi(i_1)}} \right)^{p_1} - \left(\frac{T_{\pi(i_1)j_1} - t_1'}{b_{\pi(i_1)}} \right)^{p_1} \right| \right\} \tag{17}$$

$$\le C_K' \frac{1}{b_n} \frac{1}{\min\{h_x^2, h_y^2\}} \to 0$$

where C_K, C_K' are constants that depend on K. Similarly, we can show that

$$\sup_{|t_1-t_1'|<1/b_n} \frac{1}{n} \sum_{1 \le i_1 \le n} |E[\mathcal{S}_{\pi(i_1)}(t_1)] - E[\mathcal{S}_{\pi(i_1)}(t_1')]| \le C_K' \frac{1}{b_n} \frac{1}{\min\{h_x^2, h_y^2\}} \to 0$$

By similarly arguments as in the proof of Lemma 2 (Zhang & Wang, 2016), we can show that if M is large enough

$$P\left(\sup_{t_1 \in \chi(b_n)} \left| \widetilde{R}_{\pi,1}(t_1) \right| > Ma_n \right) \le 2 \left(\frac{\max_{1 \le i \le n+m} N_i^{-1}}{n} \right)^{M-C_{K,T}-2} \to 0 \tag{18}$$

where $C_{K,T}$ is a constant that depends on K and T_x, T_y.

(ii) The result follows from the fact that the upper bounds in (17) and (18) hold uniformly for all permutations in $\mathbb{P}_{n+m}$. $\qquad\square$

Next, we resume the proof of Theorem 2. For any $1 \le i_1 \ne i_2 \le n+m$, we set

$$\widetilde{T}(t_1, t_2) = E\left[\mathcal{T}_{\Pi(i_1),\Pi(i_2)} \right] \quad \text{and} \quad \widetilde{G}(t_1, t_2) = E\left[\mathcal{Z}_{\Pi(i_1),\Pi(i_2)} \right]$$

By symmetry of the kernel K, $E[U_{\Pi,I}^{p_1 p_2}] = 0$ if $p_1 = 1$ or $p_2 = 1$. Consequently, by Lemmas 2 and 3, if $p_1 = 1$ or $p_2 = 1$,

$$\left\| U_{\Pi,I}^{p_1 p_2}(t_1, t_2) \right\|_2 = o_p(1) \quad \text{and} \quad \sup_{t_1, t_2} \left| U_{\Pi,I}^{p_1 p_2}(t_1, t_2) \right| = o_p(1)$$

Otherwise, $\sup_{t_1, t_2} |U_{\Pi,I}^{p_1 p_2}(t_1, t_2)|$ would converge to a positive constant in probability. This also implies that the denominator of $\widehat{G}_{\Pi,I}(t_1, t_2)$ is bounded away from 0 with high probability. Thus,

$$\iint \left| \widehat{G}_{\Pi,I}(t_1, t_2) - \frac{\widetilde{G}(t_1, t_2)}{\widetilde{T}(t_1, t_2)} \right| dt_1 \, dt_2$$

$$= O_p\left(\iint |U_{\Pi,I}^{0,0}(t_1, t_2) - \widetilde{T}(t_1, t_2)| \, dt_1 \, dt_2 + \iint \left| V_{\Pi,I}^{0,0}(t_1, t_2) - \widetilde{G}(t_1, t_2) \right| dt_1 \, dt_2 \right.$$

$$+ \iint |U_{\Pi,I}^{1,0}(t_1, t_2) V_{\Pi,I}^{1,0}(t_1, t_2)| \, dt_1 \, dt_2 + \iint |U_{\Pi,I}^{0,1}(t_1, t_2) V_{\Pi,I}^{1,0}(t_1, t_2)| \, dt_1 \, dt_2$$

$$\left. + \iint |U_{\Pi,I}^{1,0}(t_1, t_2) V_{\Pi,I}^{0,1}(t_1, t_2)| \, dt_1 \, dt_2 + \iint |U_{\Pi,I}^{0,1}(t_1, t_2) V_{\Pi,I}^{0,1}(t_1, t_2)| \, dt_1 \, dt_2 \right)$$

$$= O_p\left(\left\| U_{\Pi,I}^{0,0} - \widetilde{T} \right\|_2 + \left\| V_{\Pi,I}^{0,0} - \widetilde{G} \right\|_2 + \left\| U_{\Pi,I}^{1,0} \right\|_2 + \left\| U_{\Pi,I}^{0,1} \right\|_2 \right)$$

$$= O_p\left(\sup_{\pi \in \mathbb{P}_{n+m}} \sqrt{\frac{1}{n^2} \sum_{i_1=1}^{n} \phi_{\pi(i_1)}} + \sup_{\pi \in \mathbb{P}_{n+m}} \sqrt{\frac{1}{m^2} \sum_{i_2=n+1}^{n+m} \phi_{\pi(i_2)}} \right)$$

which entails that

$$
\begin{aligned}
\left|\text{MED}(\Pi \cdot \mathbf{Z})\right| &\leq \int_0^1 2\left|\widetilde{G}_{\Pi,1}(t, t) - \frac{\widetilde{G}(t, t)}{\widetilde{T}(t, t)}\right| \\
&\quad + \left|\widetilde{G}_{\Pi,2}(t, t) - \frac{\widetilde{G}(t, t)}{\widetilde{T}(t, t)}\right| + \left|\widetilde{G}_{\Pi,3}(t, t) - \frac{\widetilde{G}(t, t)}{\widetilde{T}(t, t)}\right| \, dt \\
&= O_p\left(\sup_{\pi \in \mathbb{P}_{n+m}} \sqrt{\frac{1}{n^2}\sum_{i_1=1}^{n} \phi_{\pi(i_1)}} + \sup_{\pi \in \mathbb{P}_{n+m}} \sqrt{\frac{1}{m^2}\sum_{i_2=n+1}^{n+m} \phi_{\pi(i_2)}}\right)
\end{aligned}
$$

5.3 Proof of Theorem 3

Under H_A, set $c = \text{MED}(X, Y) \geq 0$. By Theorems 1 and 2,

$$
\begin{aligned}
P_{H_A}\left(\widehat{p} \leq \alpha\right) &\geq P_{H_A}\left(|\text{MED}_n(\mathbf{Z}) - c| < c/2, |\text{MED}_n(\Pi_l \cdot \mathbf{Z})| < c/2, l = 1, \ldots, S - 1\right) \\
&\geq 1 - \left\{P_{H_A}\left(|\text{MED}_n(\mathbf{Z}) - c| \geq c/2\right) + (S - 1)P_{H_A}\left(|\text{MED}_n(\Pi_l \cdot \mathbf{Z})| \geq c/2\right)\right\} \\
&\to 1
\end{aligned}
$$

5.4 Proof of Corollary 2

Under Assumptions 1, 2, and 4, the result can be shown by adopting similar arguments as in the proof of Theorem 1 by replacing $z_{i_1j_1}$, $z_{i_2j_2}$, G_1, G_2, $\widehat{G}_1$, $\widehat{G}_2$ with $\widetilde{z}_{i_1j_1}$, $\widetilde{z}_{i_2j_2}$, H_1, H_2, $\widehat{H}_1$, $\widehat{H}_2$, respectively.

5.5 Proof of Theorem 4

Proof.   Denote the density functions of $X(t) + e_1$, $Y(t) + e_2$, $X(t)$, $Y(t)$, e_1, e_2 as $\widetilde{f}_x(\cdot\,|t)$, $\widetilde{f}_y(\cdot\,|t)$, $f_x(\cdot\,|t)$, $f_y(\cdot\,|t)$, $\eta_1(\cdot)$, and $\eta_2(\cdot)$, respectively. Under Assumption 5, we have $\eta_1 = \eta_2$. Suppose $X(t) + e_1 =^d Y(t) + e_2$, then for all $a \in \mathbb{R} \not\subseteq$

$$
\int_{-\infty}^{\infty} f_x(u|t)\eta_1(a - u)\,du = \widetilde{f}_x(a|t) = \widetilde{f}_y(a|t) = \int_{-\infty}^{\infty} f_y(u|t)\eta_2(a - u)\,du
$$

which entails that $f_x(u|t) - f_y(u|t)$ is orthogonal to $l_a(u) = \eta_1(a - u)$ for any $a \in \mathbb{R}$. By Wiener's Tauberian theorem, the span of $\{l_a(u) = \eta_1(a - u) \mid a \in \mathbb{R}\}$ forms a dense subset of $L^2(\mathbb{R})$. Thus, we can conclude that $f_x(u|t) - f_y(u|t) = 0$.   □

5.6 Proof of Corollary 3

Proof.   Under Assumptions 1, 3, and 4, we can adopt similar arguments in Section 5.2 and show that

$$
\left|\text{MED}_n(\Pi \cdot \widetilde{\mathbf{Z}})\right| = O_p\left(\sup_{\pi \in \mathbb{P}_{n+m}} \sqrt{\frac{1}{n^2}\sum_{i=1}^{n} \phi_{\pi(i)}} + \sup_{\pi \in \mathbb{P}_{n+m}} \sqrt{\frac{1}{m^2}\sum_{i=n+1}^{n+m} \phi_{\pi(i)}}\right)
$$

By Corollary 2 and Theorem 4, the result follows similarly from Theorem 3.   □

Conflict of interest: We have no conflicts of interest to disclose.

Funding

Data availability

The data are available in the book 'Counting Process and Survival Analysis' by Thomas R. Fleming and David P. Harrington (2005) [https://onlinelibrary.wiley.com/doi/book/10.1002/9781118150672] and from the UCR Time Series Classification Archive [https://www.cs.ucr.edu/eamonn/time_series_data_2018/].

References

Aoshima M., Shen D., Shen H., Yata K., Zhou Y.-H., & Marron J. S. (2018). A survey of high dimension low sample size asymptotics. *Australian & New Zealand Journal of Statistics*, *60*(1), 4–19. https://doi.org/10.1111/anzs.12212

Benko M., Härdle W., & Kneip A. (2009). Common functional principal components. *The Annals of Statistics*, *37*(1), 1–34. https://doi.org/10.1214/07-AOS516

Bickel P. J. (1969). A distribution free version of the Smirnov two sample test in the *p*-variate case. *The Annals of Mathematical Statistics*, *40*(1), 1–23. https://doi.org/10.1214/aoms/1177697800

Bickel P. J., & Breiman L. (1983). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. *The Annals of Probability*, *11*(1), 185–214. https://doi.org/10.1214/aop/1176993668

Cabaña A., Estrada A. M., Peña J., & Quiroz A. J. (2017). Permutation tests in the two-sample problem for functional data. In G. Aneiros, G. E. Bongiorno, R. Cao, & P. Vieu, (Eds) *Functional statistics and related fields. contributions to statistics*. Cham: Springer. https://doi.org/10.1007/978-3-319-55846-2_11

Carroll C., Gajardo A., Chen Y., Dai X., Fan J., Hadjipantelis P. Z., Han K., Ji H., Mueller H.-G., & Wang J.-L. (2021). *fdapace: Functional data analysis and empirical dynamics*. R package version 0.5.6. https://github.com/functionaldata/tPACE

Chakraborty S., & Zhang X. (2021). A new framework for distance and kernel-based metrics in high dimensions. *Electronic Journal of Statistics*, *15*(2), 5455–5522. https://doi.org/10.1214/21-EJS1889

Cook J. R., & Stefanski L. A. (1994). Simulation–extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, *89*(428), 1314–1328. https://doi.org/10.1080/01621459.1994.10476871

Cox D. D., & Lee J. S. (2008). Pointwise testing with functional data using the Westfall–Young randomization method. *Biometrika*, *95*(3), 621–634. https://doi.org/10.1093/biomet/asn021

Cramér H. (1928). On the composition of elementary errors. *Scandinavian Actuarial Journal*, *1928*(1), 13–74. https://doi.org/10.1080/03461238.1928.10416862

Cuevas A., Febrero M., & Fraiman R. (2004). An ANOVA test for functional data. *Computational Statistics & Data Analysis*, *47*(1), 111–122. https://doi.org/10.1016/j.csda.2003.10.021

Dau H. A., Keogh E., Kamgar K., Yeh C.-C. M., Zhu Y., Gharghabi S., Ratanamahatana C. A., Yanping H., Bagnall A., Mueen A., & Batista G. (2018). *The UCR time series classification archive*. https://www.cs.ucr.edu/eamonn/time_series_data_2018/

Davidian M., Lin X., & Wang J. (2004). Introduction: Emerging issues in longitudinal and functional data analysis. *Statistica Sinica*, *14*(3), 613–614.

Fan J., & Lin S.-K. (1998). Test of significance when data are curves. *Journal of the American Statistical Association*, *93*(443), 1007–1021. https://doi.org/10.1080/01621459.1998.10473763

Ferraty F., González-Manteiga W., Martínez-Calvo A., & Vieu P. (2012). Presmoothing in functional linear regression. *Statistica Sinica*, *22*(1), 69–94. https://doi.org/10.5705/ss.2010.085

Fleming T. R., & Harrington D. P. (2005). *Counting processes and survival analysis*. John Wiley & Sons.

Friedman J. H., & Rafsky L. C. (1979). Multivariate generalizations of the Wald–Wolfowitz and Smirnov two-sample tests. *The Annals of Statistics*, *7*(4), 697–717. https://doi.org/10.1214/aos/1176344722

Gao H., & Shao X. (2021). *Two sample testing in high dimension via maximum mean discrepancy*, arXiv, arXiv:2109.14913, preprint: not peer reviewed.

Gretton A., Borgwardt K. M., Rasch M. J., Schölkopf B., & Smola A. (2012). A kernel two-sample test. *Journal of Machine Learning Research*, *13*(1), 723–773.

Gretton A., Fukumizu K., Teo C., Song L., Schölkopf B., & Smola A. (2007). A kernel statistical test of independence. In *Advances in neural information processing systems 20* (pp. 585–592).

Guo J., Zhou B., & Zhang J.-T. (2019). New tests for equality of several covariance functions for functional data. *Journal of the American Statistical Association*, *114*(527), 1251–1263. https://doi.org/10.1080/01621459.2018.1483827

Hall P., & Keilegom I. V. (2007). Two-sample tests in functional data analysis starting from discrete data. *Statistica Sinica*, *17*(4), 1511–1531.

Hall P., Marron J. S., & Neeman A. (2005). Geometric representation of high dimension, low sample size data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*(3), 427–444. https://doi.org/10.1111/j.1467-9868.2005.00510.x

He T., Zhong P.-S., Cui Y., & Mandrekar V. (2023). Unified tests for nonparametric functions in RKHS with kernel selection and regularization. *Statistica Sinica* (forthcoming), https://doi.org/10.5705/SS.202020.0339

Henze N. (1988). A multivariate two-sample test based on the number of nearest neighbor type coincidences. *The Annals of Statistics*, *16*(2), 772–783. https://doi.org/10.1214/aos/1176350835

Holland J. K., Kemsley E. K., & Wilson R. H. (1998). Use of Fourier transform infrared spectroscopy and partial least squares regression for the detection of adulteration of strawberry purées. *Journal of the Science of Food and Agriculture*, *76*(2), 263–269. https://doi.org/10.1002/(SICI)1097-0010(199802)76:2<263::AID-JSFA943>3.0.CO;2-F

Horváth L., & Kokoszka P. (2012). *Inference for functional data with applications*. (Vol. 200). Springer-Verlag.

Hsing T., & Eubank R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons.

Jiang Q., Hušková M., Meintanis S. G., & Zhu L. (2019). Asymptotics, finite-sample comparisons and applications for two-sample tests with functional data. *Journal of Multivariate Analysis*, *170*, 202–220. https://doi.org/10.1016/j.jmva.2018.09.002

Kim I., Balakrishnan S., & Wasserman L. (2022). Minimax optimality of permutation tests. *The Annals of Statistics*, *50*(1), 225–251. https://doi.org/10.1214/21-AOS2103

Klebanov L. (2006). *N-distances and their applications*. University of Chicago Press.

Kolmogorov A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giorn Dell'inst Ital Degli Att*, *4*, 89–91.

Krzyśko M., & Smaga Ł. (2021). Two-sample tests for functional data using characteristic functions. *Austrian Journal of Statistics*, *50*(4), 53–64. https://doi.org/10.17713/ajs.v50i4.1099

Lehmann E. L., & Romano J. P. (2005). *Testing statistical hypotheses* (3rd ed.). Springer-Verlag.

Li Y., & Hsing T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, *38*(6), 3321–3351. https://doi.org/10.1214/10-AOS813

Lin Z., & Wang J.-L. (2022). Mean and covariance estimation for functional snippets. *Journal of the American Statistical Association*, *117*(537), 348–360. https://doi.org/10.1080/01621459.2020.1777138

Lyons R. (2013). Distance covariance in metric spaces. *The Annals of Probability*, *41*(5), 3284–3305. https://doi.org/10.1214/12-AOP803

Panaretos V. M., Kraus D., & Maddocks J. H. (2010). Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *Journal of the American Statistical Association*, *105*(490), 670–682. https://doi.org/10.1198/jasa.2010.tm09239

Paparoditis E., & Sapatinas T. (2016). Bootstrap-based testing of equality of mean functions or equality of covariance operators for functional data. *Biometrika*, *103*(3), 727–733. https://doi.org/10.1093/biomet/asw033

Pfister N., Bühlmann P., Schölkopf B., & Peters J. (2018). Kernel-based tests for joint independence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *80*(1), 5–31. https://doi.org/10.1111/rssb.12235

Pini A., & Vantini S. (2016). The interval testing procedure: A general framework for inference in functional data analysis. *Biometrics*, *72*(3), 835–845. https://doi.org/10.1111/biom.12476

Pomann G.-M., Staicu A.-M., & Ghosh S. (2016). 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. https://doi.org/10.1111/rssc.12130

Ramsay J., & Silverman B. W. (2005). *Functional data analysis* (2nd ed.). Springer-Verlag.

Rindt D., Sejdinovic D., & Steinsaltz D. (2021). Consistency of permutation tests of independence using distance covariance, HSIC and dHSIC. *Stat*, *10*(1), e364. https://doi.org/10.1002/sta4.364

Rizzo M., & Szekely G. (2021). *energy: E-statistics: Multivariate inference via the energy of data*. R package version 1.7-8. https://CRAN.R-project.org/package=energy

Schilling M. F. (1986). Multivariate two-sample tests based on nearest neighbors. *Journal of the American Statistical Association*, *81*(395), 799–806. https://doi.org/10.1080/01621459.1986.10478337

Sejdinovic D., Gretton A., & Bergsma W. (2013). A kernel test for three-variable interactions. In *Advances in neural information processing systems 26* (pp. 1124–1132).

Smirnov N. V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Moscow University Mathematics Bulletin*, *2*(2), 3–14.

Staicu A.-M., Lahiri S. N., & Carroll R. J. (2015). Significance tests for functional data with complex dependence structure. *Journal of Statistical Planning and Inference*, *156*, 1–13. https://doi.org/10.1016/j.jspi.2014.08.006

Székely G. J., & Rizzo M. L. (2004). Testing for equal distributions in high dimension. *InterStat*, *5*, 1–6.

Székely G. J., Rizzo M. L., & Bakirov N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, *35*(6), 2769–2794. https://doi.org/10.1214/009053607000000505

Von Mises R. (1928). *Statistik und wahrheit*. Julius Springer.

Wang H., Zhong P.-S., Cui Y., & Li Y. (2018). Unified empirical likelihood ratio tests for functional concurrent linear models and the phase transition from sparse to dense functional data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *80*(2), 343–364. https://doi.org/10.1111/rssb.12246

Wang J.-L., Chiou J.-M., & Müller H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, *3*(1), 257–295. https://doi.org/10.1146/annurev-statistics-041715-033624

Wang Q. (2021). Two-sample inference for sparse functional data. *Electronic Journal of Statistics*, *15*, 1395–1423. https://doi.org/10.1214/21-EJS1802

Wynne G., & Duncan A. B. (2022). A kernel two-sample test for functional data. *Journal of Machine Learning Research*, *23*, 1–51.

Yao F., Müller H.-G., & Wang J.-L. (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, *100*(470), 577–590. https://doi.org/10.1198/016214504000001745

Yao F., Müller H.-G., & Wang J.-L. (2005b). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, *33*(6), 2873–2903. https://doi.org/10.1214/009053605000000660

Yuan A., Fang H.-B., Li H., Wu C. O., & Tan M. T. (2020). Hypothesis testing for multiple mean and correlation curves with functional data. *Statistica Sinica*, *30*(2), 1095–1116. https://doi.org/10.5705/ss.202017.0262

Zhang J.-T., & Chen J. (2007). Statistical inferences for functional data. *The Annals of Statistics*, *35*(3), 1052–1079. https://doi.org/10.1214/009053606000001505

Zhang J.-T., Cheng M.-Y., Wu H.-T., & Zhou B. (2019). A new test for functional one-way ANOVA with applications to ischemic heart screening. *Computational Statistics & Data Analysis*, *132*, 3–17. https://doi.org/10.1016/j.csda.2018.05.004

Zhang J.-T., & Liang X. (2014). One-way ANOVA for functional data via globalizing the pointwise F-test. *Scandinavian Journal of Statistics*, *41*(1), 51–71. https://doi.org/10.1111/sjos.12025

Zhang J.-T., Liang X., & Xiao S. (2010). On the two-sample Behrens–Fisher problem for functional data. *Journal of Statistical Theory and Practice*, *4*(4), 571–587. https://doi.org/10.1080/15598608.2010.10412005

Zhang X., & Wang J.-L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics*, *44*(5), 2281–2321. https://doi.org/10.1214/16-AOS1446

Zhong P.-S., Li J., & Kokoszka P. (2021). Multivariate analysis of variance and change points estimation for high-dimensional longitudinal data. *Scandinavian Journal of Statistics*, *48*(2), 375–405. https://doi.org/10.1111/sjos.12460

Zhu C., & Shao X. (2021). Interpoint distance based two sample tests in high dimension. *Bernoulli*, *27*(2), 1189–1211. https://doi.org/10.3150/20-BEJ1270

Zhu C., Zhang X., Yao S., & Shao X. (2020). Distance-based and RKHS-based dependence metrics in high dimension. *The Annals of Statistics*, *48*(6), 3366–3394. https://doi.org/10.1214/19-AOS1934