Appl. Math. J. Chinese Univ.
2025, 40(3): 666-686

Testing independence in Hilbert spaces using random projection

HU Zhi-ming[1,2] JIANG Tao[4,1,*] XU Jin-feng[3]

Abstract. As data becomes increasingly complex, measuring dependence among variables is of great interest. However, most existing measures of dependence are limited to the Euclidean setting and cannot effectively characterize the complex relationships. In this paper, we propose a novel method for constructing independence tests for random elements in Hilbert spaces, which includes functional data as a special case. Our approach is using distance covariance of random projections to build a test statistic that is computationally efficient and exhibits strong power performance. We prove the equivalence between testing for independence expressed on the original and the projected covariates, bridging the gap between measures of testing independence in Euclidean spaces and Hilbert spaces. Implementation of the test involves calibration by permutation and combining several p-values from different projections using the false discovery rate method. Simulation studies and real data examples illustrate the finite sample properties of the proposed method under a variety of scenarios.

§1 Introduction

Recent technological advancements in science and engineering have resulted in an abundance of complex data structures, such as high-dimensional, nonlinear, and infinite-dimensional data including functional data. Measuring and testing dependence among such complex data is crucial for statistics, machine learning, and scientific discovery. Testing for independence has long been a fundamental problem in statistics, and several desirable methods have been proposed, including kernel-based criteria [7, 11, 12, 25, 31], distance correlation [26, 27], maximal information coefficient [22], copula based measures [23, 24], projection correlation [13, 32], Ball

Received: 2024-02-20. Revised: 2024-06-10.
MR Subject Classification: 62G10, 62H20.
Keywords: computational efficiency, distance covariance, false discovery rate, functional data, Hilbert space, permutation calibration, random projection.
Digital Object Identifier(DOI): https://doi.org/10.1007/s11766-025-5162-4.
Supported by the Grant of National Science Foundation of China (11971433), Zhejiang Gongshang University "Digital+" Disciplinary Construction Management Project (SZJ2022B004), Institute for International People-to-People Exchange in Artificial Intelligence and Advanced Manufacturing (CCIPERGZN202439), and the Development Fund for Zhejiang College of Shanghai University of Finance and Economics (2023FZJJ15).
*Corresponding author.

covariance proposed recently [20], and others [17, 18, 28]. However, many of these techniques were developed in the Euclidean setting and may not be directly applicable to complex data, particularly functional data with infinite dimension. Therefore, it is crucial to establish an efficient procedure for testing independence in these complex datasets.

Distance correlation [26, 27] is a widely used method for testing independence due to its desirable theoretical properties and powerful performance in numerical studies. It has also been extended to metric spaces [19], making it a useful tool for detecting dependence among complex data including functional data, which can be viewed as random elements in Hilbert space. Despite these advantages, the direct computation of distance correlation takes $O(n^2)$ time, where n is the sample size, and the complexity of data often makes the computation of pairwise distances between sample points prohibitively time-consuming. This issue is particularly problematic for complex data, where the computation cost of distance covariance could be substantial and limit its applicability in practice. Therefore, achieving a favorable trade-off between computational efficiency and test performance for distance covariance in complex data is crucial. While it may be difficult to simultaneously achieve both goals, sacrificing some power performance to increase computational efficiency may be necessary. Finding such a balance would have a direct impact on the practical applicability of distance covariance for complex data.

The aim of this paper is two-fold: to bridge the gap between measures of testing independence in Euclidean spaces and Hilbert spaces [6], and to establish a method that reduces computational complexity while maintain desirable power performance. We consider testing the independence of two random elements X and Y in Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively,

$$H_0 : X \text{ and } Y \text{are independent} \qquad vs. \qquad H_1 : \text{otherwise} \tag{1}$$

by randomly projecting them onto univariate variables X^f and Y^g. We show in Section 2 that testing independence of X and Y is equivalent to testing independence of their projections. As a result, we can measure dependence between random elements in Hilbert spaces by measuring their projections, and apply methods suitable for testing independence in Euclidean settings to Hilbert spaces including functional data.

To reduce computational complexity, we combine the projection procedure with distance covariance. Randomly projecting onto a number L of different directions and merging the resulting p-values using the False Discovery Rate (FDR) [2], we account for a higher power while reducing the influence of individual directions. This method has a time complexity of $O(n\log(n))$ using efficient numerical algorithms [16], and simulation studies show competitive performance with small values of L. This approach builds upon the idea of combining random projection and distance covariance used in [15] for random vectors, although their method includes constant dependent on vector dimension, making it unsuitable for functional data with infinite dimension. We generalize this idea for functional data using a different strategy to overcome the dimensionality issue. While distance covariance is a special choice within our framework, other methods for testing independence can be used instead.

The paper is organized as follows. Section 2.1 provides a brief review of distance covariance,

its relevant properties, and the fast algorithm for its computation. Section 2.2 shows how the independence of two random elements is equivalent to the independence of their projections and establishes a testing procedure for independence based on distance covariance and random projections in Section 2.3. Section 3 describes the implementation of the test and other practical considerations. In Section 4, we conduct simulations to illustrate the finite sample properties of the test and apply it to some real datasets. Finally, Section 5 provides some concluding remarks. All technical proofs are provided in the Appendix.

§2 Theoretical Framework and Methodology

In functional data analysis, functions or curves can be viewed as elements in $L^2([0,1])$, a space consisting of all square integrable functions defined on the interval $[0,1]$, or as realizations of stochastic processes. However, these perspectives do not always coincide, unless under certain joint measurability assumptions [14]. In this article, we adopt the random element perspective, which links functional data to ordinary data in Euclidean spaces, since Euclidean spaces are special cases of Hilbert spaces. We will establish a random projection distance covariance in this setting.

2.1 Distance Covariance Estimation

We provide a brief introduction to distance covariance, as proposed by [27]. Let Z and W be two random vectors with dimensions p and q, respectively, and with characteristic functions $f_Z(t)$ and $f_W(s)$. The joint characteristic function is denoted by $f_{Z,W}(t,s)$. If the first moments of Z and W are finite, then the squared distance covariance between Z and W is defined as the weighted L_2 distance between $f_{Z,W}(t,s)$ and $f_Z(t)f_W(s)$:

$$\mathcal{V}^2(Z,W) = \frac{1}{c_p c_q}\int_{\mathbb{R}^{p+q}} \frac{|\, f_{Z,W}(t,s) - f_Z(t)f_W(s)\,|^2}{|\, t\,|_p^{1+p}|\, s\,|_q^{1+q}}dtds,$$

where $|\cdot|_p$ is the Euclidean norm in $\mathbb{R}^p$, $c_p = \frac{\sqrt{\pi}\Gamma((p+1)/2)}{\Gamma(p/2)}$, $c_q = \frac{\sqrt{\pi}\Gamma((q+1)/2)}{\Gamma(q/2)}$, and $\Gamma(\cdot)$ is the complete gamma function. The nonnegative value $\mathcal{V}(Z,W)$ is known as the distance covariance (dCov).

The distance covariance has the following important property, established in Theorem 3 of [27]: If $\mathrm{E}(|Z|_p + |W|_q) < \infty$, then Z and W are independent if and only if $\mathcal{V}^2(Z,W) = 0$.

If $\mathrm{E}|Z|_p^2 < \infty$ and $\mathrm{E}|W|_q^2 < \infty$, then the distance covariance between Z and W can be expressed as

$$\begin{aligned}\mathcal{V}^2(Z,W) =& \mathrm{E}[|Z-Z'|_p|W-W'|_q] + \mathrm{E}|Z-Z'|_p\mathrm{E}|W-W'|_q \\ &- 2\mathrm{E}[|Z-Z'|_p|W-W''|_q],\end{aligned}$$

where (Z,W), $(Z^{'},W^{'})$, and $(Z^{''},W^{''})$ are i.i.d. Thus, an empirical distance covariance $\mathcal{V}_n^2(Z,W)$ can be defined based on a sample $(Z,W) = \{(Z_k,W_k) : k = 1,\cdots,n\}$ from the joint distribution

of random vectors $Z \in \mathbb{R}^p$ and $W \in \mathbb{R}^q$, where

$$a_{ij} = |Z_i - Z_j|_p, \qquad a_{i\cdot} = \frac{1}{n}\sum_{l=1}^{n} a_{il}, \qquad a_{\cdot j} = \frac{1}{n}\sum_{l=1}^{n} a_{lj},$$

$$a_{\cdot\cdot} = \frac{1}{n^2}\sum_{k,l=1}^{n} a_{kl}, \qquad A_{ij} = a_{ij} - a_{i\cdot} - a_{\cdot j} + a_{\cdot\cdot},$$

$i, j = 1, \cdots, n$, and similarly for W. The squared empirical distance covariance is then given by

$$\mathcal{V}_n^2(Z, W) = \frac{1}{n^2}\sum_{i,j=1}^{n} A_{ij}B_{ij}, \tag{2}$$

where $b_{ij} = |W_i - W_j|_q$ and $B_{ij} = b_{ij} - b_{i\cdot} - b_{\cdot j} + b_{\cdot\cdot}$, $i, j = 1, \cdots, n$. Computing the sample distance covariance typically requires $O(n^2)$ pairwise distance calculations and $O(n^2)$ memory units for storing them. With the large datasets commonly encountered in the era of big data, it is often impractical to implement an $O(n^2)$ algorithm on a personal computer.

Although computing the distance covariance for $Z \in \mathbb{R}$ and $W \in \mathbb{R}$ typically requires $O(n^2)$ pairwise distance calculations, a fast algorithm is available for this case in [16]. Using the same notation a_{ij} and b_{ij} as before, but now for $(Z_i, W_i) \in \mathbb{R} \times \mathbb{R}$, it has been shown (Szekely and Rizzo, 2014; Huo and Szekely, 2016) that the unbiased estimator of $\mathcal{V}^2(Z, W)$ is given by

$$\begin{aligned}\Lambda_n(Z, W) \quad = \quad & \frac{1}{n(n-3)}\sum_{i\neq j} a_{ij}b_{ij} \\ & - \frac{2}{n(n-2)(n-3)}\sum_{i=1}^{n} a_{i\cdot}b_{i\cdot} + \frac{a_{\cdot\cdot}b_{\cdot\cdot}}{n(n-1)(n-2)(n-3)}.\end{aligned} \tag{3}$$

This estimator can be computed with the fast algorithm proposed in [16], which has a computational complexity of $O(n \log n)$ and storage requirements of $O(n)$. The algorithm is implemented in the 'dcov2d' function in the **energy** package.

2.2 Projection Representation of Independent Random Elements in Hilbert Space

In this subsection, we explore the connection between independence of random elements in Hilbert spaces and the independence of their projections. Throughout this paper, we consider a separable Hilbert space $\mathcal{H}$ with associated norms $\|\cdot\|$ and inner products $\langle\cdot,\cdot\rangle$. Let $\mathscr{B}(\mathcal{H})$ denote the Borel σ-algebra of $\mathcal{H}$, generated by the open sets, and let $(\Omega, \mathscr{E}, \mathrm{P})$ be a probability space. We define a $\mathcal{H}$-valued random element X as a mapping from Ω to $\mathcal{H}$ that is measurable with respect to the σ-algebras $\mathscr{E}$ and $\mathscr{B}(\mathcal{H})$. This definition is analogous to the definition of independence in Banach spaces [29].

For ease of reading and convenience of proofing we review the definition of independence of two random elements X and Y, taking values in Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$. Random elements X and Y are independent if $\mathrm{P}(X \in A, Y \in B) = \mathrm{P}(X \in A)\mathrm{P}(Y \in B)$ for any $A \in \mathscr{B}(\mathcal{H}_1)$ and $B \in \mathscr{B}(\mathcal{H}_2)$. Based on this definition, we present the following theorem, which establishes a relationship between the random elements and their projections. The Theorem 1 (4) is the

result of Lemma 1 in [18]. For easy reference, we restate this conclusion and give a different proof in Appendix.

Theorem 1. *Suppose that X and Y are random elements in Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively, then the following statements are equivalent*

(1) X and Y are independent;

(2) $\langle X, f\rangle$ and Y are independent for all $f \in \mathcal{H}_1$;

(3) X and $\langle Y, g\rangle$ are independent for all $g \in \mathcal{H}_2$;

(4) $\langle X, f\rangle$ and $\langle Y, g\rangle$ are independent for all $f \in \mathcal{H}_1$ and $g \in \mathcal{H}_2$.

According to the property of distance covariance and Theorem 1, we can readily obtain the following result, which is analogous to Lemma 4.1 in [15].

Corollary 1. *Suppose that X and Y are random elements in Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively. If their first order moments exist, then X and Y are independent if and only if $\mathcal{V}^2(\langle X, f\rangle, \langle Y, g\rangle) = 0$ for all $f \in \mathcal{H}_1$ and $g \in \mathcal{H}_2$.*

2.3 Distance Covariance of Random Projections

We provide a detailed procedure for testing the hypothesis (1) for two functional random variables based on the results in Section 2.2. For convenience, we use X^f and Y^g to denote $\langle X, f\rangle$ and $\langle Y, g\rangle$, respectively. Based on Theorem 2, the null hypothesis H_0 can be stated as follows

$$H_0^{'}\text{: } X^f \text{ and } Y^g \text{ are independent for all } f \in \mathcal{H}_1 \text{ and } g \in \mathcal{H}_2.$$

This hypothesis includes a family of sub-hypotheses of independence of real-valued random variables. We can then define a total measure for the dependence as the integrated distance covariance

$$T(X, Y) = \int\int \mathcal{V}^2(X^f, Y^g)\mu_1(df)\mu_2(dg), \tag{4}$$

where μ_1 and μ_2 are nondegenerate Gaussian measures on $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. According to Corollary 1, if $H_0^{'}$ holds and the conditions are satisfied, we have $T(X, Y) = 0$.

Suppose that we have a sample $(X_1, Y_1), \cdots, (X_n, Y_n) \in \mathcal{H}_1 \times \mathcal{H}_2$, then the empirical version of $T(X, Y)$ is given by

$$T_n(X, Y) = \int\int \mathcal{V}_n^2(X^f, Y^g)\mu_1(df)\mu_2(dg),$$

where $\mathcal{V}_n^2(X^f, Y^g)$ is defined by (2). Although $T_n(X, Y)$ can be used as the test statistic for (1), it is difficult to compute directly. One possible approach is to use Monte Carlo simulation to approximate the integral. Specifically, we can compute the empirical estimate as

$$T_n^{MC}(X, Y) = \frac{1}{m_1 m_2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\mathcal{V}_n^2(X^{f_i}, Y^{g_j}),$$

where $f_1, \cdots, f_{m_1}$ and $g_1, \cdots, g_{m_2}$ are directions randomly generated by μ_1 and μ_2, respectively. This estimate approximates $T_n(X, Y)$ well when the numbers m_1 and m_2 are large enough, although it requires a significant amount of computation.

Fortunately, the following theorem provides a feasible solution.

We define $\mathcal{M}_1$ as the set of all $f \in \mathcal{H}_1$ such that $\langle X, f\rangle$ and Y are independent, $\mathcal{M}_2$ as the set of all $g \in \mathcal{H}_2$ such that X and $\langle Y, g\rangle$ are independent, and $\mathcal{M}$ as the set of all $(f, g) \in \mathcal{H}_1 \times \mathcal{H}_2$ such that $\langle X, f\rangle$ and $\langle Y, g\rangle$ are independent. We denote the product measure on $\mathcal{H}_1 \times \mathcal{H}_2$ as $\mu_1 \times \mu_2$.

The following theorem establishes the relationship between the independence of random elements in Hilbert spaces and the measures of the sets $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}$.

Theorem 2. *Let μ_1 and μ_2 be non-degenerate Gaussian measures on $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. Then,*

(1) X and Y are independent if and only if $\mu_1(\mathcal{M}_1) = 1$;

(2) X and Y are independent if and only if $\mu_2(\mathcal{M}_2) = 1$;

(3) X and Y are independent if and only if $\mu_1 \times \mu_2(\mathcal{M}) = 1$.

Remark 1. *In [9], Theorem 4 presents a similar result for random elements in Banach spaces. Their theorem is based on Theorem 1 in [5]. However, their conditions are much stronger than ours. They require that all absolute values of the random element's moments are finite, i.e., $m_n = \int ||x||^n \mathrm{P}(dx) < \infty$ for $n = 1, 2, \cdots$, and the Carleman's condition must be satisfied, $\sum_{n\geq 1} m_n^{-1/n} = \infty$.*

According to Theorem 2, to test the null hypothesis H_0: X and Y are independent, we can randomly select $f \in \mathcal{H}_1$ and $g \in \mathcal{H}_2$ using μ_1 and μ_2, respectively, and then test the projected null hypothesis $H_0^{fg} : X^f$ and Y^g are independent.

Remark 2. *If X is a random function and Y is a variable, we can test the independence of X and Y by randomly selecting a projection of X and leaving Y unchanged. In other words, we test the hypothesis H_0^f, where f is randomly selected from $\mathcal{H}_1$ using μ_1.*

When both X and Y are random functions, we test the hypothesis H_0^{fg} instead of H_0. To do so, we randomly select $f \in \mathcal{H}_1$ and $g \in \mathcal{H}_2$ using μ_1 and μ_2, respectively, and then test the hypothesis $H_0^{fg} : X^f$ and Y^g are independent.

To test hypothesis H_0^{fg}, the test statistic we consider is

$$\Pi_n = n \cdot \Lambda_n(X^f, Y^g), \tag{5}$$

where Λ_n is defined in (3). Since Π_n is the squared distance covariance of X^f and Y^g, and X^f and Y^g are univariate, similar to Corollary 2 in [27], we have the following property for the test statistic Π_n.

Corollary 2. *If X and Y are independent, the asymptotic distribution of Π_n is given by*

$$\Pi_n \xrightarrow{d} \sum_{i=1}^{\infty} \lambda_i Z_i^2,$$

where $Z_i^2 \sim \chi_1^2$ are i.i.d. random variables and λ_i are non-negative constants that depend on the distribution of (X, Y).

If X^f and Y^g are dependent, then Π_n tends to infinity almost surely as $n \to \infty$.

§3 Testing Procedure for Independence

Testing H_0^{fg} instead of H_0 has the advantage that the random variables are real, but it may result in a loss of power and vary for different projections. To address these issues, we follow [4] and sample several directions $f_1, \cdots, f_L$ and $g_1, \cdots, g_L$ from $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. We test the projected hypotheses $H_0^{f_1 g_1}, \cdots, H_0^{f_L g_L}$ and combine the resulting p-values to control the final rejection rate to be at most α under H_0. We use the FDR method [2, 3] for this purpose.

The testing procedure is described in the following algorithm

(1) Let Π_n denote a test for checking H_0^{fg} with f chosen by a non-degenerate Gaussian measure μ_1 on $\mathcal{H}_1$ and g chosen by a non-degenerate Gaussian measure μ_2 on $\mathcal{H}_2$.

(2) Calibrate the test statistic for H_0^{fg} by randomly permuting the index of the Y sample.

(3) Sample L directions $f_1, \cdots, f_L$ and $g_1, \cdots, g_L$ from $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively.

(4) For each $i = 1, \ldots, L$, test $H_0^{f_i g_i}$ using Π_n and record the corresponding p-value.

(5) Apply the FDR method to the p-values to control the final rejection rate to be at most α under H_0.

Algorithm 3.1 (Permutation Calibration for H_0^{fg})

(1) Compute the test statistic $\Pi_n = \Lambda_n(X^f, Y^g)$ using the formula (3) for the sample $\{(X_i^f, Y_i^g)\}_{i=1}^n$.

(2) For each ℓ, generate a random permutation $Y^{*g,\ell} = (Y_1^{(l)g}, \cdots, Y_n^{(l)g})$ of the vector $\boldsymbol{Y}^g = (Y_1^g, \ldots, Y_n^g)$.

(3) Compute the test statistic $V_\ell = \Lambda_n(X^f, Y^{*g,\ell})$ using the formula (3) for X^f and $Y^{*g,\ell}$.

(4) Repeat steps (2) and (3) for all $\ell = 1, \cdots, B$. Reject H_0 if the p-value is less than a prescribed level α, where the p-value is given by

$$p\text{-value} = \frac{\sum_{\ell=1}^B I(\Pi_n < V_\ell)}{1+B}.$$

Note that under the hypothesis H_0^{fg} and random permutation, X^f and $Y^{*g,\ell}$ are independent.

We use the FDR method [2] to control the rejection rate to be at most α under H_0, where α is the significance level. The following algorithm is used to choose the final p-value.

Algorithm 3.2 (Testing Procedure for H_0)

(1) For $i = 1, \cdots, L$, compute the p-value p_i of $H_0^{f_i g_i}$ using Algorithm 3.1.

(2) Set the final p-value of H_0 as $\min_{i=1,\cdots,L} \frac{L}{i} p_{(i)}$, where $p_{(1)} \le \cdots \le p_{(L)}$.

The choice of random projection directions is critical since it can affect the power of the test. Drawing directions that are almost orthogonal to the sample data can lead to unreliable results. To avoid this issue, we use a data-driven method proposed by [4]. We need some preparation before the statement. For a random function $X(t)$, $x \in T$, where T is an interval of $\mathbb{R}$, the mean and covariance function of $X(t)$ are defined

$$\mu(t) = EX(t), \quad c(t,s) = E[(X(t) - \mu(t))(X(s) - \mu(s))].$$

The covariance operator C of covariance function $c(t,s)$ is

$$C(f)=\int c(t,s)f(s)ds.f\in L^2(T),$$

where $L^2(T)$ means the space of square integral functions. The pairs $(\lambda_1,e_1),(\lambda_2,e_2),\dots$ are the eigenelements of X (or of the covariance function c). The empirical eigenpairs $\left\{\left(\hat{\lambda}_j,\hat{e}_j\right)\right\}$ are estimates of $(\lambda_1,e_1),(\lambda_2,e_2),\dots$, which can be computed using the sample $X_1,\cdots,X_n$(See Chapter 3 of [1] for details). The procedure of choosing random directions is

(1) Compute the empirical eigenpairs $\{(\hat{\lambda}_j,\hat{e}_j)\}$ using the sample $X_1,\cdots,X_n$.

(2) Choose a tuning parameter $j_n:=\min\{k:(\sum_{j=1}^{k}\hat{\lambda}_j^2)/(\sum_{j=1}^{n-1}\hat{\lambda}_j^2)\geq r,1\leq k\leq n-1\}$ for a given threshold r, such as $r=0.95$.

(3) The random directions are generated by the data-driven Gaussian process $\boldsymbol{h_1}:=\sum_{j=1}^{j_n}\varepsilon_j\times\hat{e}_j$, where $\varepsilon_j\sim\mathcal{N}(0,s_j^2)$ and s_j^2 is the sample variance of the scores in the jth functional principal component.

Note that the Gaussian measure μ associated with $\boldsymbol{h_1}$ is degenerate and does not satisfy the assumptions in Theorem 2. To obtain a non-degenerate Gaussian process, we add a Gaussian process $\mathcal{G}$ that is tightly concentrated around zero, i.e., $\boldsymbol{h_1}+\mathcal{G}$. However, the choice of using $\boldsymbol{h_1}$ or $\boldsymbol{h_1}+\mathcal{G}$ has negligible influence in practice. We use this data-driven process to draw projection directions f and g in the preceding algorithms.

For $Y_1,\cdots,Y_n$, we use the same procedure to generate another Gaussian process $\boldsymbol{h_2}$ and draw random directions g from it. This ensures that the directions are not orthogonal to the sample data.

§4 Numerical Studies

We evaluate the finite sample performance of the proposed test using Algorithms 3.1 and 3.2 through numerical studies. In Section 4.1, we investigate the impact of the number of random projections L on the test results. In Section 4.2, we compare our method with the generalization of distance covariance proposed by [19]. Section 4.3 demonstrates the application of our test to real data sets. We refer to our test as RPdcov and [19]'s test as Fdcov.

To examine the test's performance on different spaces, we consider two scenarios in the simulations. In the first scenario, both X and Y are functional, while in the second scenario, X is functional and Y is scalar.

We also investigate the potential impact of different underlying processes and distributions. We use three random processes for functional data: the Wiener process with covariance function $\Sigma(s,t)=\min(s,t)$, the fractional Brownian process with covariance function $\Sigma(s,t)=\frac{1}{2}\left(|t|^2+|s|^2-|t-s|^2\right)$, and the Ornstein-Uhlenbeck process with covariance function $\Sigma(s,t)=3\exp(-\frac{1}{3}(s+t))$. All processes are generated by the rproc2fdata function in the fda.usc package on the interval $[0,1]$. We also consider the Gaussian distribution for scalar data. Figure 1 displays the realizations of the considered processes.

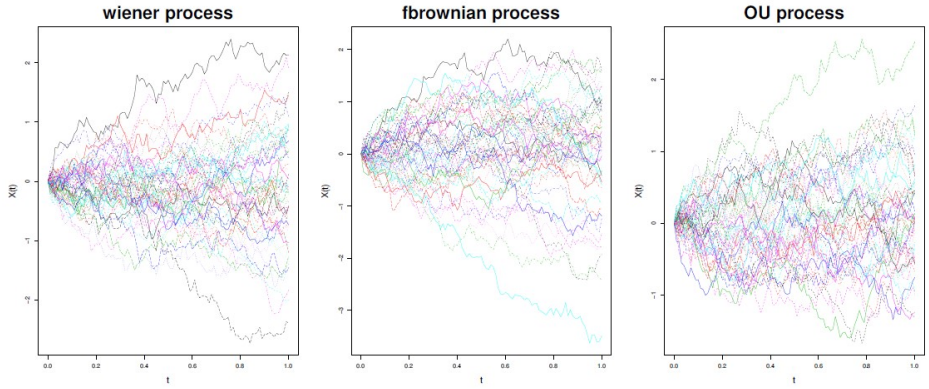

Figure 1. Realizations of Wiener, fractional Brownian, and Ornstein-Uhlenbeck processes from left to right.

4.1 Effect of the Number of Random Projections

This subsection aims to investigate the effect of the number of random projections L on the proposed test's performance through simulation studies.

To consider various potential factors that could influence the test's performance, we design four scenarios for the simulations, which we present in Example 1.

Example 1.

(1a) Draw X independently from the Wiener, fractional Brownian, and Ornstein-Uhlenbeck processes, respectively, and draw Y independently from a standard normal distribution. In this case, X and Y are independent.

(1b) This scenario is identical to scenario (1a), except that Y is drawn independently from a Wiener process.

(1c) X is a random process, and Y is a random variable. We consider two models: $Y = a\langle X, \beta\rangle + \varepsilon$ and $Y = ae^{\langle X,\beta\rangle} + \varepsilon$, where ε is a standard normal random variable independent of X, and β is a function. X is generated by the Wiener, fractional Brownian, and Ornstein-Uhlenbeck processes, respectively, and $a = 0.1$ and 0.3. In this scenario, X and Y are dependent.

(1d) X and Y are random processes. We consider models $Y = aX + \varepsilon$ and $Y = aX^2 + \varepsilon$, where $a \neq 0$ is a real constant and ε is a Wiener process independent of X. In this scenario, X and Y are dependent. X is generated by the Wiener, fractional Brownian, and Ornstein-Uhlenbeck processes, respectively.

Note that to demonstrate the proposed method's performance for different relationships between X and Y, we consider both linear and nonlinear relationships in scenarios (1c) and (1d).

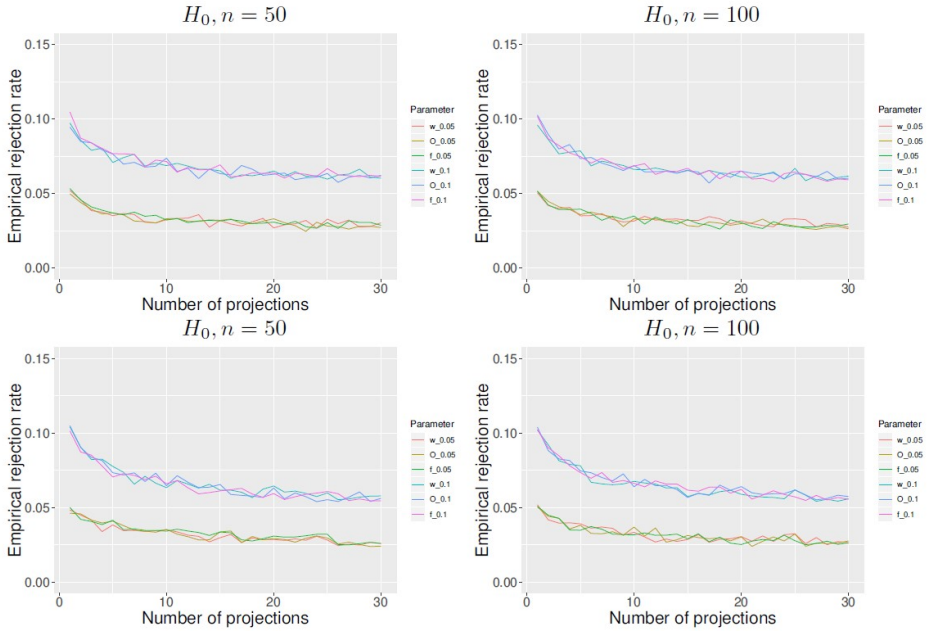

Figure 2. Empirical sizes of RPdcov tests for scenarios (1a) and (1b) with varying number of projections and sample Sizes.

In the simulations, we vary the number of random projections L from 1 to 30, and consider sample sizes of 50 and 100, respectively. We generate the random projection directions using the data-driven method described in Section 3 and perform 300 permutations, following [27]. We use two significant levels, $\alpha = 0.05$ and $\alpha = 0.1$, and carry out 10,000 Monte Carlo repetitions for each simulated case.

Figure 2 shows the empirical sizes for scenarios (1a) and (1b) when the null hypothesis H_0 is true. The empirical rejection rate curves exhibit an L-shape pattern, which is due to the conservative correction of the false discovery rate. Under H_0, the test method ensures that the rejection rate is at most α. For small L (around 3), the tests calibrate the two levels for different sample sizes reasonably well. For moderate to large L values, the empirical rejection rates decrease and stabilize below α.

Figures 3 and 4 show the empirical rejection rates for scenarios (1c) and (1d), respectively. The figures illustrate that the empirical powers with respect to L are almost constant or exhibit mild decrements, except for lower values of L, where increasing values of L can provide a significant power gain. These findings suggest that choosing a relatively small number of projections, such as $L \in \{1, 2, 3, 4, 5\}$, and particularly $L = 3$, can make a reasonable compromise between correct calibration and power.

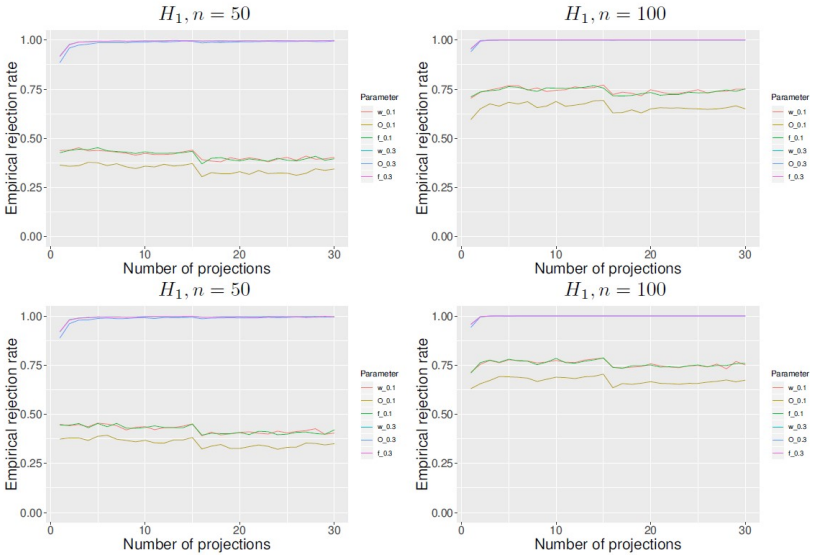

Figure 3. Empirical power of RPdcov tests for scenario (1c) with respect to the number of projections and sample sizes.

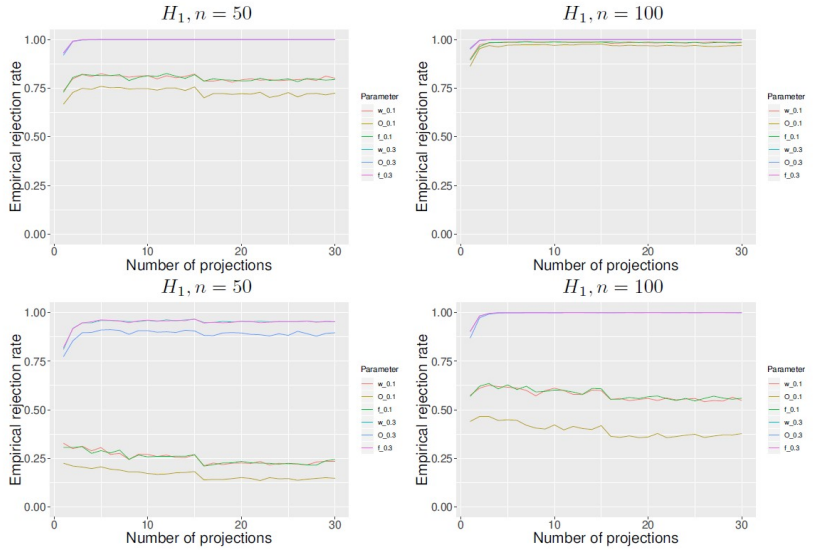

Figure 4. Empirical power of RPdcov tests for scenario (1c) with respect to the number of projections and sample sizes.

4.2 Comparison between RPdcov and Fdcov Tests

This subsection compares the proposed RPdcov method with Fdcov, an approach presented in [19], in terms of power performance and computational time consumption. We consider the following scenarios in the simulations

Example 2.

(2a) X and Y are independent. X is generated by a Wiener or Ornstein-Uhlenbeck process, and Y is generated by a standard normal distribution or a Wiener process.

(2b) X is a random process and Y is a random variable. We employ two models: $Y = a\langle X, \beta\rangle + \varepsilon$ and $Y = ae^{\langle X,\beta\rangle} + \varepsilon$, where ε is a normal random variable independent of X and $\beta(t) = \sin(2\pi t)$. X is generated by Wiener or Ornstein-Uhlenbeck process, respectively.

(2c) Both X and Y are random processes. We consider models $Y = aX + \varepsilon$ and $Y = aX^2 + \varepsilon$, where a is a real constant and ε is a Wiener random process independent of X. X is generated by Wiener or Ornstein-Uhlenbeck process, respectively.

We choose values of a equal to 0.1, 0.5, and 0.8 to indicate the closeness between X and Y. We evaluate the performance of RPdcov and Fdcov using 28 settings denoted by $H_{k,\delta}$, where $k = 0, 1, 2$ and $\delta = 1, \cdots, 12$. Table 1 explains the meaning of each setting, with $H_{0,1}$-$H_{0,4}$, $H_{1,1}$-$H_{1,12}$, and $H_{2,1}$-$H_{2,12}$ corresponding to Example 2 (2a), (2b), and (2c), respectively. We perform 2,000 Monte Carlo repetitions for each setting.

Table 2 presents the empirical rejection rates of different simulation settings with $L = 1, 3, 5$ (indexed by the subscript of RPdcov), $\alpha = 0.05$, and $n = 100, 200$. The results show a consistent pattern. In the independent scenarios $(H_{0,1}, \cdots, H_{0,4})$, the empirical sizes are close to the significance level. For most situations, Fdcov tends to have a larger power than the proposed RPdcov test. This drop in performance for RPdcov compared to Fdcov is expected due to the construction of RPdcov, which only measures dependence in a few directions. However, the loss of power relative to Fdcov is acceptable, especially considering the significantly shorter running times of RPdcov, particularly for large n. For example, when $n = 100$ and $L = 3$, the average relative loss of power for RPdcov relative to Fdcov is 8.5%, while Fdcov takes 58 times longer to run than RPdcov. This demonstrates one of the merits of RPdcov, which is its relatively short running times.

The RPdcov statistic can be computed in $O(n\log(n))$ time, which represents a significant improvement over the $O(n^2)$ required by Fdcov. This reduction in computational complexity is supported by Figure 5, and the analysis presented in Figure 6 confirms that the computational order of RPdcov is indeed $O(n\log(n))$. This favorable trade-off between computational efficiency and test performance makes RPdcov a promising tool for analyzing large datasets.

Table 1. Simulation scenarios.

$H_{k,\delta}$	Y	X	model	a
$H_{0,1}$	$N(0,1)$	wiener	independent	
$H_{0,2}$	$N(0,1)$	OU	independent	
$H_{0,3}$	Wiener	wiener	independent	
$H_{0,4}$	Wiener	OU	independent	
$H_{1,1}$	scalar	wiener	$Y = a\langle X, \beta\rangle + \varepsilon$	0.1
$H_{1,2}$	scalar	wiener	$Y = a\langle X, \beta\rangle + \varepsilon$	0.5
$H_{1,3}$	scalar	wiener	$Y = a\langle X, \beta\rangle + \varepsilon$	0.8
$H_{1,4}$	scalar	OU	$Y = a\langle X, \beta\rangle + \varepsilon$	0.1
$H_{1,5}$	scalar	OU	$Y = a\langle X, \beta\rangle + \varepsilon$	0.5
$H_{1,6}$	scalar	OU	$Y = a\langle X, \beta\rangle + \varepsilon$	0.8
$H_{1,7}$	scalar	wiener	$Y = ae^{\langle X,\beta\rangle} + \varepsilon$	0.1
$H_{1,8}$	scalar	wiener	$Y = ae^{\langle X,\beta\rangle} + \varepsilon$	0.5
$H_{1,9}$	scalar	wiener	$Y = ae^{\langle X,\beta\rangle} + \varepsilon$	0.8
$H_{1,10}$	scalar	OU	$Y = ae^{\langle X,\beta\rangle} + \varepsilon$	0.1
$H_{1,11}$	scalar	OU	$Y = ae^{\langle X,\beta\rangle} + \varepsilon$	0.5
$H_{1,12}$	scalar	OU	$Y = ae^{\langle X,\beta\rangle} + \varepsilon$	0.8
$H_{2,1}$	functional	wiener	$Y = aX + \varepsilon$	0.1
$H_{2,2}$	functional	wiener	$Y = aX + \varepsilon$	0.5
$H_{2,3}$	functional	wiener	$Y = aX + \varepsilon$	0.8
$H_{2,4}$	functional	OU	$Y = aX + \varepsilon$	0.1
$H_{2,5}$	functional	OU	$Y = aX + \varepsilon$	0.5
$H_{2,6}$	functional	OU	$Y = aX + \varepsilon$	0.8
$H_{2,7}$	functional	wiener	$Y = aX^2 + \varepsilon$	0.1
$H_{2,8}$	functional	wiener	$Y = aX^2 + \varepsilon$	0.5
$H_{2,9}$	functional	wiener	$Y = aX^2 + \varepsilon$	0.8
$H_{2,10}$	functional	OU	$Y = aX^2 + \varepsilon$	0.1
$H_{2,11}$	functional	OU	$Y = aX^2 + \varepsilon$	0.5
$H_{2,12}$	functional	OU	$Y = aX^2 + \varepsilon$	0.8

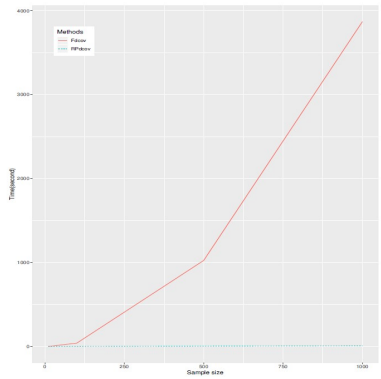

Figure 5. Computational time comparison of RPdcov and Fdcov tests for sample sizes $n = (10, 100, 500, 1000)$, with $L = 3$ projections. Measurements obtained using a 2.3 GHz Intel Core i5 MacBook Pro.

Table 2. Comparison of empirical sizes and powers for Fdcov and RPdcov tests with $1, 3$, and 5 projections, using significance level $\alpha = 0.05$ and sample sizes $n = 100, 200$.

	$n = 100$				$n = 200$			
$H_{k,\delta}$	$RPdcov_1$	$RPdcov_3$	$RPdcov_5$	$Fdcov$	$RPdcov_1$	$RPdcov_3$	$RPdcov_5$	$Fdcov$
$H_{0,1}$	0.057	0.045	0.035	0.050	0.054	0.032	0.038	0.059
$H_{0,2}$	0.043	0.050	0.035	0.048	0.051	0.048	0.047	0.045
$H_{0,3}$	0.040	0.043	0.036	0.046	0.058	0.040	0.046	0.057
$H_{0,4}$	0.047	0.044	0.039	0.058	0.055	0.047	0.036	0.041
$H_{1,1}$	0.679	0.774	0.763	0.804	0.895	0.970	0.969	0.989
$H_{1,2}$	0.965	1.000	1.000	1.000	0.983	1.000	1.000	1.000
$H_{1,3}$	0.972	1.000	1.000	1.000	0.978	1.000	1.000	1.000
$H_{1,4}$	0.595	0.639	0.703	0.738	0.819	0.918	0.937	0.977
$H_{1,5}$	0.957	0.999	1.000	1.000	0.962	1.000	1.000	1.000
$H_{1,6}$	0.965	1.000	1.000	1.000	0.972	1.000	1.000	1.000
$H_{1,7}$	0.706	0.771	0.769	0.836	0.900	0.982	0.981	0.990
$H_{1,8}$	0.964	1.000	1.000	1.000	0.970	1.000	1.000	1.000
$H_{1,9}$	0.971	1.000	1.000	1.000	0.983	1.000	1.000	1.000
$H_{1,10}$	0.634	0.704	0.702	0.776	0.847	0.929	0.949	0.971
$H_{1,11}$	0.963	1.000	1.000	1.000	0.971	1.000	1.000	1.000
$H_{1,12}$	0.945	1.000	1.000	1.000	0.981	1.000	1.000	1.000
$H_{2,1}$	0.141	0.126	0.099	0.153	0.202	0.209	0.187	0.256
$H_{2,2}$	0.881	0.985	0.986	0.995	0.943	1.000	1.000	1.000
$H_{2,3}$	0.941	1.000	1.000	1.000	0.966	1.000	1.000	1.000
$H_{2,4}$	0.127	0.101	0.099	0.135	0.195	0.157	0.179	0.245
$H_{2,5}$	0.850	0.968	0.975	0.996	0.933	0.999	1.000	1.000
$H_{2,6}$	0.933	0.997	1.000	1.000	0.949	1.000	1.000	1.000
$H_{2,7}$	0.057	0.048	0.053	0.075	0.077	0.052	0.063	0.079
$H_{2,8}$	0.573	0.628	0.635	0.768	0.815	0.937	0.966	0.991
$H_{2,9}$	0.799	0.947	0.942	0.991	0.906	0.996	0.999	1.000
$H_{2,10}$	0.046	0.042	0.056	0.059	0.078	0.061	0.051	0.065
$H_{2,11}$	0.454	0.450	0.457	0.613	0.741	0.851	0.845	0.942
$H_{2,12}$	0.744	0.877	0.866	0.967	0.861	0.992	0.998	1.000

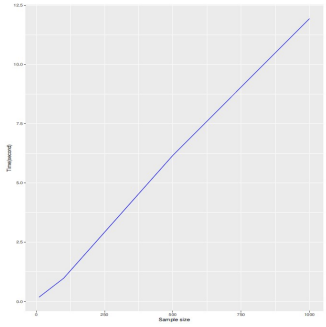

Figure 6. Computational time for RPdcov with sample sizes $n = (10, 100, 500, 1000)$ and $L = 3$ projections. Measurements obtained using a 2.3 GHz Intel Core i5 MacBook Pro.

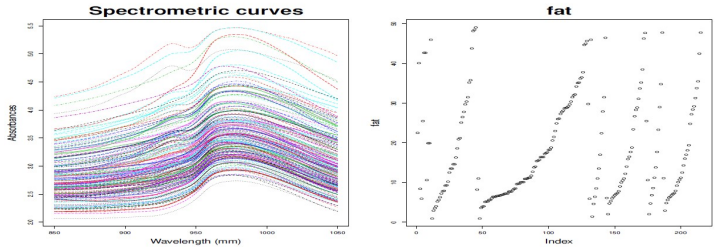

Figure 7. Spectrum of Absorbance (Left) and Fat Content (Right) for Tecator dataset.

4.3 Real Data Applications

We demonstrate the application of the new test to two datasets described in [10], which are publicly available in the *fda.usc* library. Our goal is not to provide a comprehensive case study, but rather to illustrate the potential utility of the test in assessing the dependence between variables of interest before modeling the dataset.

Our analysis begins with the classical Tecator dataset, which has been studied in [8] and [1]. This dataset consists of finely chopped pure meat samples with varying levels of protein, fat, and moisture content, as well as a spectrum of absorbances measured at wavelengths between 850 and 1050 using the near infrared transmission (NIT) principle. Figure 7 illustrates some units of the original fat and absorbance data. Typically, the objective of analyzing this dataset is to predict the fat content of a given meat sample using the spectrometric curve or one of its derivatives. However, before establishing a regression model, it is necessary to determine whether there is a relationship between the variables. Therefore, testing for dependence is essential and should be considered as a first step.

We apply our proposed method to test for dependence between the fat content and spectrometric data in the Tecator dataset. Using $L = 3$ projections, we obtain a p-value of 0.003. Consequently, we conclude that, at a significance level of $\alpha = 0.05$, there is a significant relationship between the fat content and the spectrometric curve.

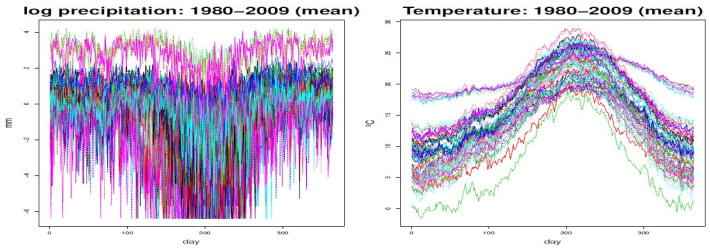

Figure 8. Daily temperature (right) and log precipitation (left) data averaged over the period from 1960 to 1994, recorded at 35 locations across Canada. Different colors mean different locations.

We turn our attention to the Canadian weather dataset, a well-known benchmark dataset in functional data analysis [21]. Figure 8 depicts the daily temperature and log precipitation profiles for a geographic location in Canada. The primary objective is to investigate the presence of a dependence between the two variables. The dataset comprises daily temperature and precipitation data averaged over the period from 1960 to 1994, recorded at 35 locations across Canada. Using our proposed method, we test for dependence in the dataset and obtain a p-value of 0.03 with $L = 3$ projections. Based on this result, we conclude that, at a significance level of $\alpha = 0.05$, there is a significant correlation between the daily precipitation profile and the daily temperature profile.

§5 Discussion

We have developed a test procedure for assessing independence in complex data represented in Hilbert spaces, with functional data being a special case. Our procedure involves applying random projections to the random elements in Hilbert spaces, thereby converting the testing of independence for random elements into the testing of independence for real variables. This approach enables the use of more traditional techniques to analyze complex data. We calibrate the test using permutation and apply the false discovery rate (FDR) method to combine p-values from L projections for increased power. Our simulation analysis suggests that a choice of $L \in \{1, \cdots, 5\}$, particularly $L = 3$, strikes a reasonable balance between maintaining size and improving power. If the collected data differs from the simulated data, a data-driven method for selecting the projection number is possible. Based on the empirical power performance in Section 4.1, it appears that the power becomes invariant after the projection number exceeds a certain threshold. In practice, one can calculate the p-value using increasing projection numbers, plot the corresponding $p - L$ curve, and identify a point at which the curve begins to flatten. Although the proposed test may sacrifice some power compared to the functional distance covariance test, the significant reduction in computational complexity is noteworthy.

In conclusion, we outline several promising applications of our methodology for testing the independence of functional data and other forms of complex data. The equivalence between

testing for the null hypothesis with the original and projected variables provides more options for assessing independence in functional data, as there are numerous methods available for testing independence in scalar data. For instance, although distance covariance is a reliable measure of dependence, it requires a finite first moment; when this condition is not met, the performance of distance covariance can be less efficient [32]. In such cases, we can use the projection correlation method proposed by [32] instead. If robustness is a primary concern, a rank-based method would be a prudent choice. Additionally, for other types of complex data, if we can construct an appropriate Hilbert space, then the method presented in this paper could be applied. For example, the color picture data, which are recorded as array with dimension $n \times n \times 3$, can be seen as elements in a Hilbert space since we can easily define a inner operation for array.

Appendix: Technical Proofs

We present several useful lemmas before proving the theoretical results. Let $\mathcal{H}$ be a separable Hilbert space with norm $\|\cdot\|$ and inner product $\langle\cdot,\cdot\rangle$. The space $\mathcal{H}$ can be viewed as a metric space with the metric

$$d(f,g) = \|f-g\| = \langle f-g, f-g\rangle^{1/2}.$$

The Borel σ-field of $\mathcal{H}$ is the smallest σ-field containing all open subsets (relative to the norm-based metric) of $\mathcal{H}$ and is denoted by $\mathscr{B}(\mathcal{H})$. The σ-field generated by the inverse images of sets in $\mathscr{B}(\mathcal{H})$ is denoted by $\sigma(X)$, and the smallest σ-field containing a class $\mathscr{C}$ of sets is denoted by $\sigma(\mathscr{C})$. Let $\mathscr{M}$ be the class of all sets of the form $\{x \in \mathcal{H} : \langle x, f\rangle \in C\}$, where $f \in \mathcal{H}$ and C is an open subset of $\mathbb{R}$. We restate Theorems 7.1.1 and 7.1.2 in [14] as Lemmas 1 and 2, respectively.

Lemma 1. *The σ-field $\sigma(\mathscr{M})$ is identical to $\mathscr{B}(\mathcal{H})$.*

Lemma 2. *Let X be a mapping from a probability space $(\Omega, \mathcal{F}, \mathbb{B})$ into $(\mathcal{H}, \mathscr{B}(\mathcal{H}))$. Then,*
(1) X is measurable if $\langle X, f\rangle$ is measurable for all $f \in \mathcal{H}$, and
(2) if X is measurable, its distribution is uniquely determined by the (marginal) distributions of $\langle X, f\rangle$ over $f \in \mathcal{H}$.

We also state a useful result on page 251 of [29] as Lemma 3.

Lemma 3. *If $\lim_{n\to\infty} X_n = X$ and $\lim_{n\to\infty} Y_n = Y$ in probability, and each X_n is independent of Y_n, then X and Y are independent.*

The support S_μ of a probability measure μ in a Hilbert space is defined as the smallest closed (measurable) set with μ-measure 1. The following lemma is derived from results in [30].

Lemma 4. *Assuming μ is a non-degenerate Gaussian measure on a separable Hilbert space $\mathcal{H}$, the support S_μ of μ is $\mathcal{H}$.*

Proof of Theorem 1. We prove the following equivalences
$1 \Rightarrow 2$: Since X and Y are independent, and $\langle X, f\rangle$ is a measurable function of X, it follows

that $\langle X, f\rangle$ and Y are independent.

$2 \Rightarrow 1$: For any $A \in \mathscr{B}(\mathcal{H}_1)$ and $B \in \mathscr{B}(\mathcal{H}_2)$, by Lemma 1, we have $A \in \sigma(\mathscr{M}_X)$, where $\mathscr{M}_X$ is the class of all sets of the form $\{x \in \mathcal{H}_1 : \langle x, f\rangle \in C\}$ for $f \in \mathcal{H}_1$ and C is an open subset of $\mathbb{R}$. For any $A' \in \mathscr{M}_X$, A' has the form $A = \{x \in \mathcal{H}_1 : \langle x, f\rangle \in C\}$ for some $f \in \mathcal{H}_1$ and some open subset C of $\mathbb{R}$. Since $\langle X, f\rangle$ and Y are independent for any $f \in \mathcal{H}_1$, we have $\mathrm{P}(X \in A', Y \in B) = \mathrm{P}(\langle X, f\rangle \in C, Y \in B) = \mathrm{P}(\langle X, f\rangle \in C)\mathrm{P}(Y \in B) = \mathrm{P}(X \in A')\mathrm{P}(Y \in B)$. Since A is an element of the sigma field generated by $\mathscr{M}_X$, it also hold that $\mathrm{P}(X \in A, Y \in B) = \mathrm{P}(X \in A)\mathrm{P}(Y \in B)$. Thus X and Y are independent.

$1 \Leftrightarrow 3$: The proof is the same as $1 \Leftrightarrow 2$.

$1 \Rightarrow 4$: Since X and Y are independent, and $\langle X, f\rangle$ and $\langle Y, g\rangle$ are measurable functions of X and Y, respectively, it follows that $\langle X, f\rangle$ and $\langle Y, g\rangle$ are independent for any $f \in \mathcal{H}_1$ and $g \in \mathcal{H}_2$.

$4 \Rightarrow 1$: For any $A \in \mathscr{B}(\mathcal{H}_1)$ and $B \in \mathscr{B}(\mathcal{H}_2)$, by Lemma 1, we have $A \in \sigma(\mathscr{M}_X)$ and $B \in \sigma(\mathscr{M}_Y)$, where $\mathscr{M}_X$ is the class of all sets of the form $\{x \in \mathcal{H}_1 : \langle x, f\rangle \in C\}$ for $f \in \mathcal{H}_1$ and C is an open subset of $\mathbb{R}$, and $\mathscr{M}_Y$ is defined similarly. For any $A' \in \mathscr{M}_X$ and $B' \in \mathscr{M}_Y$, since $\langle X, f\rangle$ and $\langle Y, g\rangle$ are independent for any $f \in \mathcal{H}_1$ and $g \in \mathcal{H}_2$, we have $\mathrm{P}(X \in A', Y \in B') = \mathrm{P}(\langle X, f\rangle \in C_1, \langle Y, f\rangle \in C_2) = \mathrm{P}(\langle X, f\rangle \in C_1)\mathrm{P}(\langle Y, f\rangle \in C_2) = \mathrm{P}(X \in A')\mathrm{P}(Y \in B')$, where C_1 and C_2 are open sets in $\mathbb{R}$. Since A is an element of the sigma field generated by $\mathscr{M}_X$ and B is an element of the sigma field generated by $\mathscr{M}_Y$, it also hold that $\mathrm{P}(X \in A, Y \in B) = \mathrm{P}(X \in A)\mathrm{P}(Y \in B)$. Thus, X and Y are independent. □

Proof of Theorem 2. We prove the following

(1) If X and Y are independent, then by Theorem 1 (2), $\langle X, f\rangle$ and Y are independent for any $f \in \mathcal{H}_1$. Thus, $\mu_1(\mathcal{M}_1) = 1$. Now if $\mu_1(\mathcal{M}_1) = 1$, then by Lemma 4 and the definition of support, the closure $\bar{\mathcal{M}}_1$ of $\mathcal{M}_1$ is $\mathcal{H}_1$. For any $f \in \mathcal{H}_1$, since $\mathcal{M}_1$ is dense in $\mathcal{H}_1$, there exists a sequence $\{f_n\}$, where $f_n \in \mathcal{M}_1$, such that $f_n \to f$ as $n \to \infty$. This yields $\langle f_n, X\rangle \to \langle f, X\rangle$ as $n \to \infty$. By the definition of $\mathcal{M}_1$, $\langle f_n, X\rangle$ and Y are independent. Making use of Lemma 3, we obtain that $\langle X, f\rangle$ and Y are independent for any $f \in \mathcal{H}_1$. According to Theorem 1 (2), X and Y are independent.

(2) See proof in (1).

(3) If X and Y are independent, then by Theorem 1 (4), $\langle X, f\rangle$ and $\langle Y, g\rangle$ are independent for any $f \in \mathcal{H}_1$ and $g \in \mathcal{H}_2$. Thus, $\mu_1 \times \mu_2(\mathcal{M}) = 1$. Now if $\mu_1 \times \mu_2(\mathcal{M}) = 1$, then by Lemma 4 and the definition of support, the closure $\bar{\mathcal{M}}$ of $\mathcal{M}$ is $\mathcal{H}_1 \times \mathcal{H}_2$. For any $f \in \mathcal{H}_1$ and $g \in \mathcal{H}_2$, there exist sequences $\{f_k\}$ and $\{g_k\}$ such that $f_k \in \mathcal{M}_1$, $g_k \in \mathcal{M}_2$, $f_k \to f$, and $g_k \to g$ as $k \to \infty$. This yields $\langle f_n, X\rangle \to \langle f, X\rangle$ and $\langle g_k, Y\rangle \to \langle g, Y\rangle$ as $k \to \infty$. By the definition of $\mathcal{M}$, $\langle f_k, X\rangle$ and $\langle g_k, Y\rangle$ are independent. By Lemma 3, we obtain that $\langle X, f\rangle$ and $\langle Y, g\rangle$ are independent for any $f \in \mathcal{H}_1$ and $g \in \mathcal{H}_2$. According to Theorem 1 (4), X and Y are independent.

□

Proof of Corollary 2. After applying random projection, the random functions X and Y are transformed into real variables $\langle X, f\rangle$ and $\langle Y, g\rangle$. The test statistic is then the same as the one used in [27]. For the proof of Theorem 4, please refer to [27] or [19].

□

Declarations

Conflict of interest The authors declare no conflict of interest.

References

[1] G Aneiros, P Vieu. *Semi-functional partial linear regression*, Statistics & Probability Letters, 2006, 76(11): 1102-1110.

[2] Y Benjamini, D Yekutieli. *The control of the false discovery rate in multiple testing under dependency*, The Annals of Statistics, 2001, 29(4): 1165-1188.

[3] J A Cuesta-Albertos, M Febrero-Bande. *A simple multiway anova for functional data*, Test, 2010, 19(3): 537-557.

[4] J A Cuesta-Albertos, E Garcia-Portugues, M Febrero-Bande, et al. *Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes*, The Annals of Statistics, 2019, 47(1): 439-467.

[5] J A Cuesta-Albertos, R Fraiman, T Ransford. *A sharp form of the cramer-wold theorem*, Journal of Theoretical Probability, 2007, 20(2): 201-209.

[6] T K M Djonguet, G M Nkiet, A M Mbina. *Testing independence of functional variables by an hilbert-schmidt independence criterion estimator*, Statistics & Probability Letters, 2024, 207: 110016.

[7] D Edelmann, J Goeman. *A regression perspective on generalized distance covariance and the hilbert-schmidt independence criterion*, Statistical Science, 2022, 37(4): 562-579.

[8] F Ferraty, P Vieu. *Nonparametric functional data analysis: theory and practice*, New York: Springer Science and Business Media, 2006.

[9] R Fraiman, L Moreno, S Vallejo. *Some hypothesis tests based on random projection*, Computational Statistics, 2017, 32(3): 1165-1189.

[10] E Garcia-Portugues, W Gonzalez-Manteiga, M Febrero-Bande. *A goodness-of-fit test for the functional linear model with scalar response*, Journal of Computational and Graphical Statistics, 2014, 23(3): 761-778.

[11] A Gretton, O Bousquet, A Smola, et al. *Measuring statistical dependence with hilbert-schmidt norms*, International conference on algorithmic learning theory, Berlin: Springer, 2005.

[12] A Gretton, K Fukumizu, C H Teo, et al. *A kernel statistical test of independence*, Advances in neural information processing systems, New York: Curran Associates Inc, 2008.

[13] Y He, R Zhao, W X Zhou. *An efficient iterative least squares algorithm for large-dimensional matrix factor model via random projection*, 2023, arXiv:2301.00360.

[14] T Hsing, R Eubank. *Theoretical foundations of functional data analysis, with an introduction to linear operators*, Hoboken: John Wiley & Sons, 2015.

[15] C Huang, X Huo. *A statistically and numerically efficient independence test based on random projections and distance covariance*, Frontiers in Applied Mathematics and Statistics, 2022, 7: 779841.

[16] X Huo, G J Székely. *Fast computing for distance covariance*, Technometrics, 2016, 58(4): 435-447.

[17] T Lai, Z Zhang, Y Wang. *Testing independence and goodness-of-fit jointly for functional linear models*, Journal of the Korean Statistical Society, 2021, 50: 380-402.

[18] T Lai, Z Zhang, Y Wang, et al. *Testing independence of functional variables by angle covariance*, Journal of Multivariate Analysis, 2021, 182: 104711.

[19] R Lyons. *Distance covariance in metric spaces*, The Annals of Probability, 2013, 41(5): 3284-3305.

[20] W Pan, X Wang, H Zhang, et al. *Ball covariance: A generic measure of dependence in banach space*, Journal of the American Statistical Association, 2020, 115(529): 307-317.

[21] J Ramsay. *Functional data analysis*, Encyclopedia of Statistics in Behavioral Science, New York: John Wiley & Sons, 2005.

[22] D N Reshef, Y A Reshef, H K Finucane, et al. *Detecting novel associations in large data sets*, Science, 2011, 334(6062): 1518-1524.

[23] B Schweizer, E F Wolff. *On nonparametric measures of dependence for random variables*, The annals of statistics, 1981, 9(4): 879-885.

[24] K F Siburg, P A Stoimenov. *A measure of mutual complete dependence*, Metrika, 2010, 71(2): 239-251.

[25] A Smola, A Gretton, L Song, et al. *A hilbert space embedding for distributions*, International Conference on Algorithmic Learning Theory, Berlin: Springer, 2007.

[26] G J Székely, M L Rizzo. *Brownian distance covariance*, The Annals of Applied Statistics, 2009, 3(4): 1236-1265.

[27] G J Székely, M L Rizzo, N K Bakirov. *Measuring and testing dependence by correlation of distances*, The annals of statistics, 2007, 35(6): 2769-2794.

[28] D Tjøstheim, H Otneim, B Støve. *Statistical dependence: Beyond pearson's* ρ, Statistical Science, 2022, 37(1): 90-109.

[29] N N Vakhania, V I Tarieladze, S A Chobanyan. *Probability distributions on Banach spaces*, New York: Springer Science & Business Media, 1987.

[30] N Vakhania. *The topological support of gaussian measure in banach space*, Nagoya Mathematical Journal, 1975, 57: 59-63.

[31] K Zhang, J Peters, D Janzing, et al. *Kernel-based conditional independence test and application in causal discovery*, Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence, Corvallis: AUAI Press, 2011, 804-813.

[32] L Zhu, K Xu, R Li, et al. *Projection correlation between two random vectors*, Biometrika, 2017, 104(4): 829-843.

[1]School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou 310018, China.
[2]Zhejiang College, Shanghai University of Finance and Economics, Jinhua 321013, China.
[3]Department of Biostatistics, City University of Hong Kong, Hong Kong, China.
[4]Department of Statistics, Zhejiang Gongshang University Hangzhou College of Commerce, Hangzhou 311508, China.
Emails: z2011159@shufe-zj.edu.cn, jtao@263.net, jinfengxu@gmail.com