Testing independence and goodness-of-fit jointly for functional linear models

Tingyu Lai, Zhongzhan Zhang & Yafei Wang

Journal of the Korean Statistical Society

ISSN 1226-3192

J. Korean Stat. Soc.
DOI 10.1007/s42952-020-00083-4

ONLINE FIRST

Springer

Journal of the Korean Statistical Society
https://doi.org/10.1007/s42952-020-00083-4
Online ISSN 2005-2863
Print ISSN 1226-3192

RESEARCH ARTICLE

Testing independence and goodness-of-fit jointly for functional linear models

Tingyu Lai[1] · Zhongzhan Zhang[1] · Yafei Wang[2]

Received: 15 February 2020 / Accepted: 10 August 2020

Abstract

A conventional regression model for functional data involves expressing a response variable in terms of the predictor function. Two assumptions, that (i) the predictor function and the error are independent and (ii) the relationship between the response variable and the predictor function takes functional linear model, are usually added to the model. Checking the validation of these two assumptions is fundamental to statistic inference and practical applications. We develop a test procedure to check these assumptions simultaneously based on generalized distance covariance. We establish the asymptotic theory for the proposed test under null and alternative hypotheses, and provide a bootstrap procedure to obtain the critical value of the test. The proposed test is consistent against all alternatives provided that the semimetrics related to the generalized distance are strong negative, and can be readily generalized to other functional regression models. We explore the finite sample performance of the proposed test by using both simulations and real data examples. The results illustrate that the proposed method has favorable performance compared with the competing method.

Keywords Functional linear model · Generalized distance covariance · Goodness-of-fit · Independence test · Strong negative type space

1 Introduction

With technology development, increasingly complex and high dimensional data have been produced. Within these data, a large fraction can be characterized as functional data. Usually, these functions are defined on a 1-dimensional Euclidean domain, but functions defined on higher dimensional domains such as 2d and 3d image data

✉ Zhongzhan Zhang
zzhang@bjut.edu.cn

[1] Faculty of Science, Beijing University of Technology, Beijing 100124, China

[2] Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, AB T6G 2R3, Canada

Published online: 14 August 2020

or time-space data and functions observed on manifolds are also included in this paradigm. There is practical need to study the relationship between two variables including functional data (see Ramsay and Silverman 2005). An important tool for describing the relationship is regression model

$$Y = m(X) + \varepsilon, \tag{1}$$

where X is a random element in a functional space, Y is a scalar response, m is the regression function, and ε is a random error with mean zero. For convenience, we assume that the X and Y are centered.

In statistical inference on (1), two additional assumptions are crucial: (i) $m(x)$ takes some certain parametric form, for example, the functional linear model $m(x) = \langle X, \beta \rangle$ and (ii) the error ε is independent of X. Choosing the correct parametric model has a major impact on consistent parameter estimation, accurate prediction and valid policy recommendation (Cai and Hall 2006; Cardot et al. 2007; Hall and Horowitz 2007). On the other hand, the independence between X and ε plays a key role in certain estimation and testing procedures, see Cardot et al. (2003), Horváth and Reeder (2013), Cuesta-Albertos et al. (2019). Among the literature mentioned above, most of them suppose that the two assumptions hold simultaneously.

Several goodness-of-fit tests have been proposed for different parametric models of functional data. The simple null hypothesis $H_0 : m(X) = c$, where c is a fixed constant, has been studied by Delsol et al. (2011). For functional linear model (FLM) $m(X) = \langle X, \beta \rangle$, where $\langle \cdot, \cdot \rangle$ stands for the inner product in a Hilbert space, $H_0 : \beta = \beta_0$ has been considered by Cardot et al. (2003), Kokoszka et al. (2008) and Hilgert et al. (2013). For the omnibus hypothesis $H_0 : m(X) \in \mathcal{M}_\beta = \{\langle X, \beta \rangle : \beta \in \mathcal{H}\}$, where $\mathcal{H}$ is a Hilbert space, see Patilea et al. (2012), García-Portugués et al. (2014), Patilea et al. (2016) and Cuesta-Albertos et al. (2019). Some of these tests assume that the error is independent of the predictor X and some others involve tuning parameters or random projections. As far as we know, there is no method in the literature to test the independence between the predictor X and the error ε in the functional data setting.

As stated in Sen and Sen (2014), it can be hard to test the goodness-of-fit of the parametric model and the independence of X and ε separately since they depend on each other. To test the independence of X and ε, one must estimate ε precisely enough and this only can be done with assuming that the model is correctly specified. On the other hand, many goodness-of-fit tests in functional setting crucially employ the assumption that X and ε are independent. In this paper we aim to check the hypothesis that X and ε are independent (denoted by $X \perp \varepsilon$), and the model is functional linear model , that is,

$$H_0 : X \perp \varepsilon, \quad m \in \mathcal{M}_\mathcal{H} := \{\langle \cdot, \beta \rangle : \beta \in \mathcal{H}\} \quad \text{v.s.} \quad H_1 : \text{otherwise}. \tag{2}$$

The literature for testing the independence and the goodness-of-fit simultaneously is rare even in Euclidean settings. For the case of X being real random vector, Sen and Sen (2014) proposed a method serving this purpose when $m(x)$ is a linear model. Their main idea was based on the test to the independence of the predictor and the regression residual obtained from the linear parametric fit, using the

Hilbert-Schmidt independence criterion (Gretton et al. 2008). For functional linear model, their method and results cannot be used directly, since the dimension of the functional data is infinite. Hence, a more sophisticated method for testing the independence of error and predictor and goodness-of-fit of the functional linear model simultaneously is needed.

In this paper, we propose a nonparametric test to check the independence and goodness-of-fit jointly for functional linear model by choosing the semimetrics defined on the spaces of the predictors and the responses (residuals). Firstly the residuals are obtained by fitting the functional linear model, then the test statistic is constructed by predictor observations and the residuals, using the generalized distance covariance (Sejdinovic et al. 2013) with the chosen semimetrics. We derive the asymptotic behaviors of the proposed test statistic under the null and alternative hypothesis respectively, and provide a bootstrap procedure to obtain the critical value of the test. The proposed test enjoys the following properties: (i) The method is straightforward and easy to compute, since when the estimator of the slope function β is given, we only need to compute the distances between points in $\{X_i\}_{i=1}^n$ and in residuals $\{\hat{\varepsilon}_i\}_{i=1}^n$; (ii) With the virtue of the generalized distance covariance, the proposed test is consistent against all alternatives; (iii) Compared to the random projection test of goodness-of-fit given by Cuesta-Albertos et al. (2019), this method appears more powerful in most simulated examples in Sect. 3; (iv) The proposed test procedure can apply to a large collection of estimators of the slope function β and is easily generalized to other functional regression models.

The rest of the paper is organized as follows. In Sect. 2, we establish the test procedure, present its asymptotic properties, provide a bootstrap algorithm to determine the critical value of the test and give some suggestions for choosing suitable semimetrics. We explore the finite sample performance by simulations in Sect. 3 and illustrate real data applications in Sect. 4. In Sect. 5, further discussions are included. All technical proofs are presented in the Appendix.

2 Test procedure and asymptotic properties

In this section, we first define the test statistic for hypothesis (2) using the generalized distance covariance, and then give the asymptotic properties. A bootstrap method for obtaining the critical value of the test and some suggestions for choosing suitable semimetrics are also presented. We first introduce briefly the generalized distance covariance for easy reference.

2.1 Generalized distance covariance

The following definition of the generalized distance covariance, which has been used for testing the independence of two random elements, is according to Sejdinovic et al. (2013); see Székely et al. (2007), Székely and Rizzo (2009) and Lyons (2013) for more details.

Let $\mathcal{Z}$ be a nonempty set, then a bivariate function $\rho : \mathcal{Z} \times \mathcal{Z} \mapsto [0, \infty]$ is called a semimetric if $\rho(z, z') = \rho(z', z)$, and $\rho(z, z') = 0$ if and only if $z = z'$, for any $z, z' \in \mathcal{Z}$. A semimetric space $(\mathcal{Z}, \rho)$ is said to have negative type if $\forall z_1, \cdots, z_n \in \mathcal{Z}, a_1, \cdots, a_n \in \mathbb{R}$ and $n \geq 2$, with $\sum_{i=1}^n a_i = 0$,

$$\sum_{i,j=1}^{n} a_i a_j \rho(z_i, z_j) \leq 0.$$

Furthermore, we say $(\mathcal{Z}, \rho)$ has strong negative type if it is of negative type and satisfies

$$\int \rho(z, z') d([P - Q] \times [P - Q])(z, z') < 0$$

for any two different probability measures P, Q.

Let $(\mathcal{X}, \rho_{\mathcal{X}})$ and $\left(\mathcal{Y}, \rho_{\mathcal{Y}}\right)$ be semimetric spaces of negative type, where $\rho_{\mathcal{X}}$ and $\rho_{\mathcal{Y}}$ are semimetrics defined on $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let $(\Omega, \mathcal{F}, P)$ be a probability space, $(X, Y) : \Omega \mapsto \mathcal{X} \times \mathcal{Y}$ be a random element with joint distribution P_{XY} and marginal distributions P_X and P_Y. Further assume that

$$\mathrm{E}\{\rho_{\mathcal{X}}^2(X, X)\} < \infty, \quad \mathrm{E}\{\rho_{\mathcal{Y}}^2(Y, Y)\} < \infty.$$

Then the generalized distance covariance of X and Y is defined as

$$\begin{aligned} \theta(X, Y) \;&= \mathrm{E}_{XY}\mathrm{E}_{X'Y'}\rho_{\mathcal{X}}\left(X, X'\right)\rho_{\mathcal{Y}}\left(Y, Y'\right) + \mathrm{E}_X\mathrm{E}_{X'}\rho_{\mathcal{X}}\left(X, X'\right)\mathrm{E}_Y\mathrm{E}_{Y'}\rho_{\mathcal{Y}}\left(Y, Y'\right) \\ &\quad - 2\mathrm{E}_{XY}\left[\mathrm{E}_{X'}\rho_{\mathcal{X}}\left(X, X'\right)\mathrm{E}_{Y'}\rho_{\mathcal{Y}}\left(Y, Y'\right)\right], \end{aligned} \tag{3}$$

where (X', Y') is an i.i.d. copy of (X, Y). Equivalently, $\theta(X, Y)$ can be expressed as

$$\theta(X, Y) = \int \rho_{\mathcal{X}}(x, x')\rho_{\mathcal{Y}}(y, y') d([P_{XY} - P_X P_Y] \times [P_{XY} - P_X P_Y])((x, y), (x', y')),$$

where $\rho_{\mathcal{X}}\rho_{\mathcal{Y}}$ is viewed as a function on $(\mathcal{X} \times \mathcal{Y}) \times (\mathcal{X} \times \mathcal{Y})$. It is not hard to see that $\theta(X, Y) \geq 0$, and if the semimetrics $\rho_{\mathcal{X}}$ and $\rho_{\mathcal{Y}}$ are of strong negative type, then $\theta(X, Y) = 0$ if and only if X and Y are independent (Lyons 2013).

Given an independent and identically distributed sample $(X_1, Y_1), \cdots, (X_n, Y_n)$ from P_{XY}, a straightforward estimator of $\theta(X, Y)$ according to Definition (3) is

$$\theta_n(X, Y) = \frac{1}{n^2}\sum_{i,j}^{n} k_{ij} l_{ij} + \frac{1}{n^4}\sum_{i,j,q,r}^{n} k_{ij} l_{qr} - \frac{2}{n^3}\sum_{i,j,q}^{n} k_{ij} l_{iq}, \tag{4}$$

where $k_{ij} = \rho_{\mathcal{X}}(X_i, X_j)$ and $l_{ij} = \rho_{\mathcal{Y}}(Y_i, Y_j)$. Since the predictor and the error are independent under the null hypothesis, the generalized distance covariance can be employed to detect the discrepancy from the null hypothesis. This leads to the proposed test in the next subsection.

2.2 Test statistic and asymptotic properties

Suppose that we observe independent and identically distributed data $(X_1, Y_1), \cdots, (X_n, Y_n)$ from model (1). Our goal is to test hypothesis

$$H_0 : X \perp \varepsilon, \quad m \in \mathcal{M}_{L^2([0,1])} = \{\langle \cdot, \beta\rangle : \beta \in L^2([0,1])\},$$

where $L^2([0,1])$ denotes the space consisting of square integrable functions on interval $[0, 1]$. Given an estimate $\hat{\beta}$ of β under H_0, the residuals are

$$\hat{\varepsilon}_i = Y_i - \langle X_i, \hat{\beta}\rangle, \quad i = 1, \cdots, n.$$

Denote

$$k_{ij} = k(X_i, X_j), \quad l_{ij} = l(\hat{\varepsilon}_i, \hat{\varepsilon}_j),$$

where k and l are semimetrics defined on $L^2([0,1]) \times L^2([0,1])$ and $\mathbb{R} \times \mathbb{R}$ respectively. We consider test statistic

$$T_n = \frac{1}{n^2}\sum_{i,j}^{n} k_{ij}l_{ij} + \frac{1}{n^4}\sum_{i,j,q,r}^{n} k_{ij}l_{qr} - \frac{2}{n^3}\sum_{i,j,q}^{n} k_{ij}l_{iq}. \tag{5}$$

There are many estimators of β that can be chosen for the above construction. For example, Hall and Horowitz (2007) proposed the least square estimator based on functional principal components and obtained the optimal convergence rate of the estimator. Crambes et al. (2009) studied smoothing spline estimator for functional linear models. Yuan et al. (2010); Cai and Yuan (2012) obtained the optimal convergence rate of the estimator for slope function in the framework of reproducing kernel Hilbert space.

The convergence rate of the estimator $\hat{\beta}$ has a major impact on the asymptotic behaviors of T_n. To investigate the asymptotic behaviors of the test statistic with different estimators, we consider two scenarios: (i) $\sqrt{n}(\hat{\beta} - \beta)$ converges in distribution to an random element in $L^2([0,1])$; (ii) $\|\hat{\beta} - \beta\|$ converges in probability to zero, where $\|\cdot\|$ denotes the norm in $L^2([0,1])$. Scenario (i) usually happens in the setting that β or X is of finite dimension, while scenario (ii) is just declaring the L^2 consistency of the estimator, which is satisfied by most useful estimators of β. As is well known, for general functional linear models, it is impossible for $\hat{\beta} - \beta$ to converge in distribution to a non-degenerate random element in the norm topology of $L^2([0,1])$ (see Theorem 1 in Cardot et al. 2007), thus we turn to the consistency of the test in scenario (ii), rather than the null asymptotic distribution of T_n.

We first explore the scenario (i). Let X_s , $s = 1, \cdots, 4$ and ε_q, $q = 1, 2$ be independent copies of X and ε, respectively. Denote the partial derivatives of $l(x, y)$ as $l_x(x, y) = \partial_x l(x, y)$, $l_{xy} = \partial_x\partial_y l(x, y)$, etc. Assume the following conditions for H_0.

C1: $\sqrt{n}(\hat{\beta} - \beta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(Z_i) + o_p(1)$, where $Z_i = (X_i, \varepsilon_i)$ and ψ is a function from $L^2([0,1]) \times \mathbb{R}$ to $L^2([0,1])$ satisfying $\mathrm{E}[\|\psi(Z)\|^2] < \infty$.

C2: (a) $l(\cdot,\cdot)$ is second order differentiable and l_{xx}, l_{yy}, l_{xy} satisfy the Lipschitz continuous condition, that is, using l_{xx} as an example,

$$|l_{xx}(x',y') - l_{xx}(x,y)| \le L\|(x',y') - (x,y)\|_\infty,$$

where $\|(x,y)\|_\infty = \max\{|x|,|y|\}$, $x, y \in \mathbb{R}$ and L is a positive constant.

(b) $\mathrm{E}[|k(X_q,X_r)|(\|X_s\| + \|X_t\|)^3] < \infty, \quad (1 \le q,r,s,t \le 4)$;

(c) $\mathrm{E}[g^2(\varepsilon_q,\varepsilon_r)] < \infty$, for $1 \le q,r \le 2$, $g = l, l_x, l_y, l_{xx}, l_{xy}, l_{yy}$.

(d) $\mathrm{E}[k^2(X_q,X_r)] < \infty$, for $1 \le q,r \le 2$;

(e) $\mathrm{E}[k^2(X_q,X_r)\|X_s\|\|X_t\|] < \infty$, for $1 \le q,r,s,t \le 4$.

Theorem 1 *Suppose that conditions* C1 *and* C2 *hold. Then, under* H_0,

$$nT_n \xrightarrow{d} 6\sum_{j=1}^{\infty}\gamma_j\mathcal{Z}_j^2 + \langle\mathcal{G},\mathcal{N}\rangle + \langle\Lambda(\mathcal{G}),\mathcal{G}\rangle$$

where $\mathcal{Z}_j$ *are i.i.d.* $N(0,1)$ *random variables,* $\mathcal{G} \in L^2([0,1])$ *is the limiting element of* $\sqrt{n}(\hat\beta - \beta)$, *that is* $\sqrt{n}(\hat\beta-\beta)\xrightarrow{d}\mathcal{G}$; $\mathcal{N}$ *is a random element in* $L^2([0,1])$, Λ *is a bounded linear operator from* $L^2([0,1])$ *to* $L^2([0,1])$, γ_i*'s are eigenvalues of an operator, and the definitions of the last three quantities are given in the proof of Theorem* 1.

Remark 1 Condition C1 usually holds for most estimating methods when the dimension of β is finite, since in this case we can convert the functional linear model into an ordinary linear model, and most estimating methods for the later model satisfy C1. Condition C2(a), (b) are required when dealing with the negligibility of the remainder term in the proof of Theorem 1, and condition C2(c),(d),(e) are basic assumptions for the convergence of the V-statistics which are the components of the test statistic, see the proof of Theorem 1.

Remark 2 Theorem 1 has a similar assertion with Theorem 3.1 of Sen and Sen (2014) which was for the test of analogy of hypothesis (2) in ordinary linear models.

To investigate the behaviour of T_n in scenario (*ii*), assume the following conditions for H_0.

C3: $\|\hat\beta - \beta\| = o_p(1)$.

C4: (a) $l(\cdot,\cdot)$ is first order differentiable almost surely and l_x, l_y satisfy the Lipschitz continuous condition.

(b) $\mathrm{E}[|k(X_q,X_r)|(\|X_s\| + \|X_t\|)^2] < \infty, \quad (1 \le q,r,s,t \le 4)$.

(c) $\mathrm{E}[|g(\varepsilon_q,\varepsilon_r)|] < \infty$, for $1 \le q,r \le 2$, $g = l, l_x, l_y$.

(d) $\mathrm{E}[|k(X_q,X_r)|] < \infty$, for $1 \le q,r \le 2$.

(e) $\mathrm{E}[|k(X_q,X_r)| \cdot \|X_s\|] < \infty$, for $1 \le q,r,s \le 4$.

Theorem 2 *Suppose that conditions C3–C4 hold. Then, under* H_0,

$$T_n \xrightarrow{\mathrm{p}} 0.$$

Remark 3 Note that the condition C4 is less strict than condition C2, this is because in the proof of Theorem 2 we only need the weak law of large number for V-statistics while in the proof of Theorem 1 we need the central limiting theorem for V-statistics.

To study the behaviors of T_n under alternative hypothesis, we divide the alternative into three sub-hypotheses:

$$H_{1,1}: X \not\perp \varepsilon, m \in \mathcal{M}_{L^2([0,1])}, \quad H_{1,2}: X \perp \varepsilon, m \notin \mathcal{M}_{L^2([0,1])}, \quad H_{1,3}: X \not\perp \varepsilon, m \notin \mathcal{M}_{L^2([0,1])}.$$

Suppose that $\tilde{\beta} = \arg\inf_\beta \mathrm{E}[(Y - \langle X, \beta \rangle)^2]$ exists uniquely. The estimator $\hat{\beta}$ considered here is the corresponding estimator based on least square and functional principal components, see Hall and Hosseini-Nasab (2006); Hall and Horowitz (2007). Denote $\epsilon = m(X) - \langle X, \tilde{\beta} \rangle + \varepsilon$. If $m \in \mathcal{M}_{L^2([0,1])}$, then $\epsilon = \varepsilon$. We assume the following conditions for $H_{1,1}$, $H_{1,2}$ and $H_{1,3}$.

C5: Let C be some constant (can be distinct in different conditions), $\{\lambda_j\}_{j=1}^\infty$ and $\{\mathbf{e}_j\}_{j=1}^\infty$ be the eigenvalues and eigenfunctions of the covariance operator of X, and $\chi_j = \langle X, \mathbf{e}_j \rangle$.

(a) $\mathrm{E}\left[\|X\|^4\right] < \infty$, $\mathrm{E}\left[\chi_j^4\right] \leq C\lambda_j^2$, for all j, and ε is a continuous random variable with $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{var}(\varepsilon) \leq C$.
(b) The eigenvalues $\{\lambda_j\}_{j=1}^\infty$ satisfy, for all j, $C^{-1}j^{-a} \leq \lambda_j \leq Cj^{-a}$ and $\lambda_j - \lambda_{j+1} \geq C^{-1}j^{-a-1}$, where $a > 1$.
(c) $|\langle \beta, \mathbf{e}_j \rangle| < Cj^{-b}$, $b > \frac{a}{2} + 1$.
(d) There exists $C_1, C_2 > 0$, $C_1 n^{\frac{1}{a+2b}} < k_n < C_2 n^{\frac{1}{a+2b}}$, where k_n is the number of functional principal components to obtain the estimator $\hat{\beta}$.

C6: (a) Both semimetrics k, l have strong negative type.
(b) $\mathrm{E}[|m(X)|^2] < \infty$.
(c) Under $H_{1,2}$ and $H_{1,3}$, $m(X) - \langle X, \tilde{\beta} \rangle$ is a non-constant function of X.

Theorem 3 *Suppose that conditions C4, C5, C6 and one of* $H_{1,1}$, $H_{1,2}$ *and* $H_{1,3}$ *hold. Then*

$$T_n \xrightarrow{\mathrm{p}} \tau,$$

where $\tau = \theta(X, \epsilon) > 0$ *is a constant.*

Remark 4 Conditions in C5 are parallel to those in Hall and Horowitz (2007), which are required to obtain the proper convergent rate of $\|\hat{\beta} - \tilde{\beta}\|$. Condition C6(a) guarantees that the generalized distance covariance can detect the dependence between X and ϵ. Condition C6(c) means that when the regression function is not the form of $\langle X, \beta \rangle$, the difference $m(X) - \langle X, \tilde{\beta} \rangle$ and X are dependent. Condition C6(b) is a moment condition required for the consistency of $\hat{\beta}$. When the model is not linear, the estimator $\hat{\beta}$ can still approximate $\tilde{\beta}$ well. We can interpret $\langle X, \tilde{\beta} \rangle$ as the closest function to m in $\mathcal{M}_{L^2([0,1])}$.

Remark 5 Note that, by the result of Hall and Horowitz (2007), the estimator $\hat{\beta}$ satisfies the condition C3 under H_0. Although Theorem 3 is established based on the classical estimator of functional linear model using least squares and functional principal components, the result can be extended to other estimators, for example, the m-estimator proposed in Huang et al. (2014) as well.

Now we give the test procedure based on T_n. Denote the null distribution function of T_n by F_n. Then, we reject H_0 if

$$T_n > c_{n,\alpha} := F_n^{-1}(1 - \alpha), \tag{6}$$

where α is the significance level of the test. By the results of Theorems 2 and 3 and the Lemma 14.15 of van der Vaart (2000), this test is consistent against any alternative hypothesis. Since the distribution function F_n is unknown, we introduce a bootstrap method to approximate the critical value $c_{n,\alpha}$ in the next subsection.

2.3 Bootstrap procedure and the choice of semimetrics

In this subsection, we first provide a bootstrap procedure to find the critical value of the proposed test, and then give some suggestions to find suitable semimetics k and l.

Algorithm (bootstrap for the critical value)

1. Let $\mathbb{P}_{n,\varepsilon^0}$ be the empirical distribution of centered residuals, i.e.,

$$\varepsilon_i^o = \hat{\varepsilon}_i - \overline{\hat{\varepsilon}} \quad (i = 1, \dots, n)$$

where $\hat{\varepsilon}_i$ is residual described in Section 2.2 and $\overline{\hat{\varepsilon}} = n^{-1} \sum_{i=1}^n \hat{\varepsilon}_i$.

2. Generate an independent and identically distributed bootstrap sample $\left\{X_{in}^*, \varepsilon_{in}^*\right\}_{1 \le i \le n}$ of size n from the measure $\mathbb{P}_n = \mathbb{P}_{n,X} \times \mathbb{P}_{n,\varepsilon^o}$ where $\mathbb{P}_{n,X}$ is the empirical distribution of the observed X_i's.

3. Define

$$Y_{in}^* = \langle X_{in}^*, \hat{\beta}_n \rangle + \varepsilon_{in}^* \quad (i = 1, \dots, n)$$

where $\hat{\beta}_n$ is a given estimator. Compute the bootstrapped estimator $\hat{\beta}_n^*$ by the same method of $\hat{\beta}_n$ with the bootstrap sample $\left(Y_{in}^*, X_{in}^*\right)$, and the bootstrap residuals

$$\hat{\varepsilon}^*_{in} = Y^*_{in} - \langle X^*_{in}, \hat{\beta}^*_n \rangle \quad (i = 1, \ldots, n).$$

4. Compute the bootstrap test statistic T^*_n by (5) using X^*_{in} and $\hat{\varepsilon}^*_{in}$, $i = 1, \cdots, n$.

5. Repeat steps $1 - 4$ for B times and collect the T^*_n values as $\{T^{*b}_n\}_{b=1}^B$. Then take the $(1-\alpha)$th quantile $c^*_{n,\alpha}$ of $\{T^{*b}_n\}_{b=1}^B$ as the critical value of the test with a significant level α. Thus we reject the null hypothesis when $T_n > c^*_{n,\alpha}$. Or alternatively, when the corresponding p-value, defined as $\frac{1}{B}\sum_{b=1}^B 1_{\{T^{*b}_n \geq T_n\}}$, is less than the given significant level, we reject the null hypothesis.

Now we discuss the choice of the semimetrics. It is obvious that T_n depends on the semimetrics k and l, and the choice of them influences the practical performance of the test. From Theorem 3, the semimetrics of strong negative type should be considered to keep the test's consistency. The concept of strong negative type for semimetrics is related to the term of characteristic kernel. A kernel $h(\cdot,\cdot)$ on $\mathcal{Z} \times \mathcal{Z}$ is characteristic if and only if the map

$$\nu \mapsto \int h(\cdot, z) d\nu(z)$$

is injective on the space of all finite signed Borel measures on $\mathcal{Z}$ for which $\int |h(z,z)| d|\nu|(z) < \infty$. By the result of Sejdinovic et al. (2013), if h is a characteristic kernel defined on $L^2([0,1]) \times L^2([0,1])$ or $\mathbb{R} \times \mathbb{R}$, then

$$\rho(z,z') = h(z,z) + h(z',z') - 2h(z,z'), z, z' \in L^2([0,1]) \text{ or } \mathbb{R} \tag{7}$$

is a strong negative type semimetric. There are many characteristic kernels that have been studied, such as Gaussian kernel $h(z,z') = \exp(-\sigma^{-1}\|z-z'\|^2)$, Laplace kernel $h(z,z') = \exp(-\sigma^{-1}\|z-z'\|)$, $\sigma > 0$, resulting the corresponding semimetrics $\rho(z,z') = 2 - 2\exp(-\sigma^{-1}\|z-z'\|^2)$ and $\rho(z,z') = 2 - 2\exp(-\sigma^{-1}\|z-z'\|)$, where the parameter σ can be determined by the median of $\|Z_i - Z_j\|^2$ or $\|Z_i - Z_j\|$, $1 \leq i < j \leq n$, respectively. For more details on characteristic kernels, see Fukumizu et al. (2009), Sriperumbudur et al. (2008) and Sriperumbudur et al. (2010). We can also find more strong negative type semimetrics from the results of Lyons (2013), for example $\rho(z,z') = \|z-z'\|$. We use $\rho(z,z') = \|z-z'\|$ and $\rho(z,z') = 2 - 2\exp(-\sigma^{-1}\|z-z'\|^2)$ for our simulations.

Remark 6 If we choose the semimetrics constructed by Gaussian kernel, i.e. $k(x,x') = 2 - 2\exp(-\sigma^{-1}\|x-x'\|^2)$ and $l(\epsilon,\epsilon') = 2 - 2\exp(-\sigma^{-1}\|\epsilon-\epsilon'\|^2)$, then k and l satisfy conditions C2(a),(c),(d) and C4(a),(c),(d) since k, l and l's partial derivatives are bounded. And conditions C2(b),(e) and C4(b),(e) are implied by the simpler condition $\mathrm{E}\|X\|^3 < \infty$. If we choose $k(x,x') = \|x-x'\|$ and $l(\epsilon,\epsilon') = |\epsilon-\epsilon'|$, then conditions C2(a) and C4(a) hold since l has second order Taylor expansion almost surely (the exceptional set is $\{(\varepsilon,\varepsilon') : \varepsilon = \varepsilon'\}$). Conditions C2(b),(d),(e) are implied by condition $\mathrm{E}\|X\|^4 < \infty$; Conditions C4(b)(d)(e) are implied by condition $\mathrm{E}\|X\|^3 < \infty$; Condition C2(c) is implied by condition $\mathrm{E}|\varepsilon|^2 < \infty$; Condition C4(c) is implied by condition $\mathrm{E}|\varepsilon| < \infty$.

3 Simulation studies

In this section we explore the finite sample performance of the proposed test in three examples. The first is to test the independence between predictor X and error ε, provided that the regression model is functional linear model; the second is to test the goodness-of-fit of the functional linear model assuming that X and ε are independent; and the last is concerned with the situation that both independence of predictor and error and the linearity of the regression functional are invalid. The estimator of β used in this part is based on least squares and functional principal components (Hall and Horowitz 2007), and the tuning parameter k_n (see Condition C5(d)) is derived by minimizing a Schwarz-type information criterion, which is implemented by the *fregre.pc.cv* function in *fda.usc* R package.

To examine the performance of different semimetrics, we consider two couples of semimetrics, $k(x,x') = \|x - x'\|$, $l(\varepsilon,\varepsilon') = |\varepsilon - \varepsilon'|$ and $k(x,x') = 2 - 2e^{-\sigma^{-1}\|x-x'\|^2}$, $l(\varepsilon,\varepsilon') = 2 - 2e^{-\sigma^{-1}|\varepsilon-\varepsilon'|^2}$ for the proposed method, where σ is determined by the median criterion described in Section 2.3. The corresponding tests are denoted as T_1 and T_2 respectively. We will compare our tests with CvM, the test of Cuesta-Albertos et al. (2019) using 3 projections as the authors suggested. The CvM test is implemented by *rp.flm.test* function in *fda.usc* R package. We consider the following model, which is used in Cuesta-Albertos et al. (2019) for other purpose,

$$Y = \langle X, \beta \rangle + \delta \Delta(X) + \varepsilon,$$

where $\delta\Delta(X)$ presents a deviation from the linear model, and the values of them considered in the simulations are presented in Table 1, with $\Delta_1(X) := \|X\|$, $\Delta_2(X) := 25 \int_0^1 \int_0^1 \sin(2\pi ts)s(1-s)t(1-t)X(s)X(t)\mathrm{d}s\mathrm{d}t$, and $\Delta_3(X) := \left\langle e^{-X}, X^2 \right\rangle$. In order to examine the impact of the underlying processes of predictor and the different slope functions, nine scenarios are designed, which are also presented in Table 1. The underlying processes of X, indexed in [0, 1], include 5 types, which are described in the following:

Table 1 Simulation scenarios

Scenario	Coefficient $\beta(t)$	Process X	Deviation
S1	$(2\psi_1(t) + 4\psi_2(t) + 5\psi_3(t))/\sqrt{2}$	**BM**	$\Delta = \Delta_1, \delta = 0,\ 0.50,\ 0.75$
S2	$(2\tilde{\psi}_1(t) + 4\tilde{\psi}_2(t) + 5\tilde{\psi}_3(t))/\sqrt{2}$	**BB**	$\Delta = \Delta_2, \delta = 0, -2.00, -7.50$
S3	$(2\psi_2(t) + 4\psi_3(t) + 5\psi_7(t))/\sqrt{2}$	**BM**	$\Delta = \Delta_1, \delta = 0, -0.30, -0.50$
S4	$\sum_{j=1}^{20} 2^{3/2}(-1)^j j^{-2}\phi_j(t)$	**HHN**($l = 1$)	$\Delta = \Delta_2, \delta = 0, -3.00, -5.00$
S5	$\sum_{j=1}^{20} 2^{3/2}(-1)^j j^{-2}\phi_j(t)$	**HHN**($l = 2$)	$\Delta = \Delta_2, \delta = 0, -1.50, -3.00$
S6	$\log(15t^2 + 10) + \cos(4\pi t)$	**BM**	$\Delta = \Delta_1, \delta = 0,\ 0.50,\ 1.00$
S7	$\sin(2\pi t) - \cos(2\pi t)$	**OU**	$\Delta = \Delta_2, \delta = 0, -0.05, -0.10$
S8	$t - (t - \frac{3}{4})^2$	**OU**	$\Delta = \Delta_3, \delta = 0, -0.01, -0.10$
S9	$\pi^2(t^2 - \frac{1}{3})$	**GBM**	$\Delta = \Delta_3, \delta = 0,\ 2.50,\ 5.00$

BM. The Brownian motion, denoted as $\mathbf{B}$ in the following, with eigenfunctions $\psi_j(t) := \sqrt{2}\sin((j-\frac{1}{2})\pi t),\ j \geq 1$.

HHN. The functional process from Hall and Hosseini-Nasab (2006), $X(t) = \sum_{j=1}^{20} \xi_j \phi_j(t)$, where $\phi_j(t) := \sqrt{2}\cos(j\pi t)$ and $\xi_j \sim N(0, j^{-2l})$ $(l = 1, 2)$ are independent random variables.

BB. Brownian bridge, $X(t) = \mathbf{B}(t) - t\mathbf{B}(1)$, whose eigenfunctions are $\tilde{\psi}_j(t) := \psi_{j+\frac{1}{2}}(t),\ j \geq 1$.

OU. Ornstein-Uhlenbeck process, that is, Gaussian process with zero mean and covariance $\mathrm{Cov}[X(t), X(s)] = \frac{\sigma^2}{2\alpha} e^{-\alpha(s+t)}(e^{2\alpha \min(s,t)} - 1)$. Here we choose $\alpha = \frac{1}{3}, \sigma = 1$, and $X(0) \sim N(0, \frac{\sigma^2}{2\alpha})$.

GBM. Geometric Brownian motion, $X(t) = s_0 \exp\{(\mu - \frac{\sigma^2}{2})t + \sigma\mathbf{B}(t)\}$. Here we let $\sigma = 1$, $\mu = \frac{1}{2}$, and $s_0 = 2$.

To examine the performance of the tests under full observation and sparse observation, we consider 101 and 17 equidistant observation points for every sample curve of X. We conduct 1000 Monte Carlo trials and 200 bootstrap replicates at the significance level 0.05 for all examples.

Example 1 Consider the following data generating mechanism. Generate data from model $Y = \langle X, \beta \rangle + \varepsilon$ $(\delta = 0)$, where X and β take values in Table 1 and $\varepsilon | X \sim N(0, \sigma^2 + \eta \|X\|^2)$ with $\eta = 0, 0.01, 0.05$, where σ^2 is defined by $\frac{\mathrm{Var}[\langle X, \beta \rangle]}{\mathrm{Var}[\langle X, \beta \rangle] + \sigma^2} = 0.95$.

The results are summarized in Tables 2 and 3. For the full observation design (101 points), when $\eta = 0$, which indicates the independence of predictor and the error (H_0 holds), the empirical sizes of T_1 and T_2 are close to the significance level, which performs better than that of CvM. When $\eta \neq 0$, the predictor and the error are dependent. From Table 2, the powers of T_1 and T_2 increase as the parameter η increases or the sample size gets larger, while the CvM totally loses the efficiency. This is not surprising as the CvM is designed for testing the linearity of the regression model, not for testing independence of predictor X and error ε. For the sparse design (17 points), from Table 3, the performance of all the tests is similar to that in the full observation, except that the powers of T_1 and T_2 slightly decrease. When the observation points of X are sparse, the accuracy of $\hat{\beta}$ decreases, which would impact the powers of T_1 and T_2. In this example T_1 and T_2 perform very similar, and they are dominated in different scenarios, respectively.

Example 2 In this example, we test the goodness-of-fit of the functional linear model, provided that the predictor and the error are independent. Generate data from model

$$Y = \langle X, \beta \rangle + \delta\Delta(X) + \varepsilon,$$

where X and ε are independent, β, X, δ, Δ take values in Table 1 and the error ε is a normal distribution $N(0, \sigma^2)$ with σ^2 satisfying $R^2 = \frac{\mathrm{Var}[\langle X, \beta \rangle]}{\mathrm{Var}[\langle X, \beta \rangle] + \sigma^2} = 0.95$.

Table 2 Empirical sizes and powers of the tests in Example 1 with 101 equidistant observation points for every sample curve $X(t)$

	Scenario	$\eta = 0$			$\eta = 0.01$			$\eta = 0.05$		
		T_1	T_2	CvM	T_1	T_2	CvM	T_1	T_2	CvM
$n = 30$	S_1	0.043	0.045	0.064	0.301	0.249	0.085	0.264	0.267	0.081
	S_2	0.027	0.038	0.079	0.170	0.135	0.079	0.150	0.171	0.090
	S_3	0.025	0.044	0.091	0.264	0.251	0.097	0.285	0.312	0.053
	S_4	0.039	0.029	0.070	0.122	0.138	0.064	0.179	0.193	0.062
	S_5	0.047	0.041	0.042	0.248	0.226	0.056	0.354	0.425	0.056
	S_6	0.037	0.030	0.041	0.222	0.217	0.047	0.282	0.310	0.051
	S_7	0.059	0.043	0.069	0.363	0.433	0.060	0.365	0.455	0.050
	S_8	0.056	0.039	0.033	0.358	0.403	0.058	0.369	0.440	0.070
	S_9	0.031	0.045	0.082	0.227	0.250	0.070	0.324	0.340	0.080
$n = 50$	S_1	0.047	0.046	0.054	0.363	0.415	0.080	0.397	0.500	0.080
	S_2	0.027	0.045	0.060	0.220	0.229	0.057	0.195	0.296	0.080
	S_3	0.022	0.039	0.071	0.392	0.457	0.064	0.408	0.513	0.081
	S_4	0.032	0.033	0.070	0.193	0.244	0.059	0.232	0.311	0.042
	S_5	0.048	0.037	0.040	0.346	0.368	0.065	0.522	0.677	0.046
	S_6	0.040	0.052	0.041	0.257	0.310	0.045	0.382	0.530	0.035
	S_7	0.049	0.047	0.059	0.542	0.727	0.068	0.516	0.712	0.064
	S_8	0.053	0.042	0.046	0.506	0.702	0.058	0.527	0.711	0.070
	S_9	0.043	0.044	0.100	0.557	0.568	0.098	0.633	0.641	0.087

The results are summarized in Tables 4 and 5, where Delta=1 means that δ takes the second value in every scenario in Table 1 and Delta=2 is defined similarly, for example, in scenario S_1, Delta=1 means $\delta = 0.5$ and Delta=2 means $\delta = 0.75$. From Tables 4 and 5, the tests T_1 and T_2, based on our test procedure, have higher powers than the tests based on CvM in almost all scenarios, especially in scenarios 2, 4 and 6. In most scenarios, T_1 performs slightly better than T_2. The powers of all tests increase when the sample size gets larger or the deviation increases. It is worth noting that when the observation points of X become sparse, the powers of the tests are almost the same as those in full observation.

Example 3 In this example, we study the power performance when both independence and linear model are invalid. The data are generated by the same setting of Example 2 except that the predictor X and the error are dependent, satisfying $\varepsilon | X \sim N(0, \sigma^2 + 0.03\|X\|^2)$.

Tables 6 and 7 summary the powers of the tests. The tests T_1 and T_2 have higher powers than the test CvM in all scenarios. Comparing to the results in Example 2, the tests based on CvM suffer a dramatic loss of power when the predictor and

Table 3 Empirical sizes and powers of the tests in Example 1 with 17 equidistant observation points for every sample curve $X(t)$

	Scenario	$\eta = 0$			$\eta = 0.01$			$\eta = 0.05$		
		T_1	T_2	CvM	T_1	T_2	CvM	T_1	T_2	CvM
$n = 30$	S_1	0.046	0.054	0.046	0.181	0.124	0.078	0.257	0.232	0.076
	S_2	0.023	0.034	0.075	0.096	0.077	0.097	0.146	0.122	0.080
	S_3	0.031	0.045	0.084	0.209	0.161	0.078	0.272	0.253	0.090
	S_4	0.027	0.032	0.060	0.068	0.063	0.056	0.138	0.141	0.044
	S_5	0.036	0.038	0.033	0.086	0.071	0.049	0.230	0.219	0.057
	S_6	0.040	0.043	0.045	0.098	0.072	0.046	0.202	0.170	0.040
	S_7	0.046	0.047	0.052	0.367	0.387	0.090	0.396	0.426	0.067
	S_8	0.037	0.041	0.047	0.362	0.384	0.064	0.361	0.420	0.048
	S_9	0.033	0.040	0.063	0.189	0.162	0.085	0.284	0.279	0.092
$n = 50$	S_1	0.049	0.045	0.057	0.201	0.171	0.066	0.354	0.392	0.077
	S_2	0.026	0.045	0.074	0.096	0.097	0.077	0.210	0.206	0.058
	S_3	0.028	0.046	0.082	0.262	0.288	0.070	0.323	0.424	0.069
	S_4	0.033	0.037	0.072	0.067	0.080	0.059	0.135	0.186	0.047
	S_5	0.057	0.049	0.045	0.125	0.102	0.039	0.334	0.348	0.056
	S_6	0.040	0.042	0.060	0.126	0.126	0.051	0.275	0.306	0.050
	S_7	0.048	0.046	0.057	0.486	0.646	0.063	0.498	0.678	0.060
	S_8	0.046	0.045	0.042	0.503	0.634	0.045	0.530	0.688	0.059
	S_9	0.032	0.050	0.080	0.433	0.343	0.087	0.573	0.560	0.089

the error are dependent. For our tests, things are complicated. It may be anticipated that the powers of T_1 and T_2 should increase when both assumptions in null hypothesis are violated. However, in our simulations, the powers in some scenarios decrease. This phenomenon might be raised by the masking effect that the heteroscedasticity of the error might weaken the identifiability of the some kinds of nonlinearity hidden in the observations. This also happens in the diagnosis of ordinary linear models. Furthermore, the results in Tables 6 and 7 show the similarity in power performance of T_1 and T_2 in full observation and sparse observation designs.

4 Data applications

In this section, the proposed test is applied to several real datasets. The semimetric used for k and l is $\rho(z, z') = 2 - 2e^{-\|z-z'\|^2}$, where $z, z' \in L^2([0, 1])$ or $\mathbb{R}$. The first data set is from Ramsay et al. (2009) and provided in R package *fda*. It consists of daily temperature and precipitation at 35 different locations in Canada for the period 1960 to 1994. The goal is to check whether there is a linear relationship between the logarithm of annual precipitation for 35 Canadian weather stations and their temperature profiles. The p-value of our test is 0.248, while the p-value of CvM is 0.310. We assert that the functional linear model is a

Table 4 Powers of the tests in Example 2 with 101 equidistant observation points for every sample curve $X(t)$

	Scenario	Delta = 1			Delta = 2		
		T_1	T_2	CvM	T_1	T_2	CvM
$n = 30$	S_1	0.470	0.312	0.242	0.776	0.600	0.351
	S_2	0.504	0.399	0.155	0.928	0.850	0.179
	S_3	0.423	0.297	0.303	0.765	0.532	0.415
	S_4	0.395	0.370	0.081	0.649	0.596	0.104
	S_5	0.292	0.178	0.229	0.681	0.480	0.613
	S_6	0.266	0.193	0.126	0.794	0.662	0.279
	S_7	0.585	0.365	0.490	0.906	0.723	0.783
	S_8	0.674	0.550	0.331	0.990	0.984	0.595
	S_9	0.209	0.173	0.216	0.670	0.572	0.379
$n = 50$	S_1	0.792	0.621	0.337	0.975	0.908	0.546
	S_2	0.858	0.702	0.154	0.993	0.987	0.251
	S_3	0.790	0.644	0.529	0.975	0.869	0.736
	S_4	0.694	0.588	0.137	0.908	0.848	0.193
	S_5	0.485	0.297	0.436	0.917	0.745	0.841
	S_6	0.548	0.422	0.193	0.987	0.908	0.486
	S_7	0.854	0.635	0.782	0.994	0.937	0.963
	S_8	0.893	0.808	0.531	0.999	0.999	0.686
	S_9	0.383	0.325	0.284	0.910	0.846	0.522

Table 5 Powers of the tests in Example 2 with 17 equidistant observation points for every sample curve $X(t)$

	Scenario	Delta = 1			Delta = 2		
		T_1	T_2	CvM	T_1	T_2	CvM
$n = 30$	S_1	0.458	0.325	0.246	0.763	0.601	0.388
	S_2	0.513	0.426	0.155	0.929	0.864	0.192
	S_3	0.374	0.274	0.291	0.762	0.537	0.409
	S_4	0.382	0.369	0.122	0.621	0.578	0.142
	S_5	0.246	0.148	0.210	0.686	0.470	0.606
	S_6	0.282	0.217	0.139	0.773	0.635	0.330
	S_7	0.571	0.386	0.503	0.901	0.715	0.794
	S_8	0.671	0.551	0.372	0.987	0.982	0.566
	S_9	0.203	0.166	0.197	0.679	0.566	0.427
$n = 50$	S_1	0.770	0.594	0.367	0.965	0.893	0.528
	S_2	0.834	0.683	0.172	0.996	0.984	0.271
	S_3	0.738	0.616	0.511	0.960	0.853	0.629
	S_4	0.673	0.586	0.161	0.893	0.845	0.232
	S_5	0.462	0.295	0.414	0.918	0.748	0.837
	S_6	0.519	0.389	0.181	0.980	0.909	0.458
	S_7	0.873	0.638	0.774	0.995	0.942	0.952
	S_8	0.891	0.792	0.524	1.000	1.000	0.688
	S_9	0.383	0.312	0.309	0.900	0.855	0.551

Table 6 Powers of the tests in Example 3 with 101 equidistant observation points for every sample curve $X(t)$

	Scenario	Delta = 1			Delta = 2		
		T_1	T_2	CvM	T_1	T_2	CvM
$n = 30$	S_1	0.303	0.289	0.086	0.324	0.335	0.098
	S_2	0.180	0.177	0.089	0.541	0.480	0.131
	S_3	0.282	0.287	0.067	0.312	0.320	0.079
	S_4	0.195	0.193	0.064	0.279	0.270	0.061
	S_5	0.319	0.346	0.081	0.404	0.380	0.144
	S_6	0.284	0.283	0.046	0.337	0.340	0.081
	S_7	0.363	0.426	0.058	0.383	0.433	0.068
	S_8	0.401	0.435	0.055	0.603	0.611	0.194
	S_9	0.333	0.345	0.068	0.406	0.399	0.080
$n = 50$	S_1	0.420	0.531	0.084	0.509	0.551	0.129
	S_2	0.283	0.332	0.084	0.800	0.719	0.120
	S_3	0.403	0.497	0.082	0.413	0.507	0.107
	S_4	0.306	0.347	0.051	0.405	0.433	0.072
	S_5	0.505	0.607	0.096	0.571	0.636	0.194
	S_6	0.402	0.504	0.063	0.496	0.551	0.091
	S_7	0.518	0.719	0.068	0.529	0.705	0.058
	S_8	0.519	0.694	0.062	0.795	0.832	0.292
	S_9	0.664	0.664	0.079	0.704	0.677	0.108

Table 7 Powers of the tests in Example 3 with 17 equidistant observation points for every sample curve $X(t)$

	Scenario	Delta = 1			Delta = 2		
		T_1	T_2	CvM	T_1	T_2	CvM
$n = 30$	S_1	0.369	0.268	0.113	0.426	0.323	0.179
	S_2	0.335	0.247	0.106	0.870	0.765	0.170
	S_3	0.304	0.256	0.096	0.365	0.296	0.114
	S_4	0.259	0.216	0.081	0.437	0.393	0.095
	S_5	0.289	0.212	0.114	0.468	0.349	0.320
	S_6	0.251	0.197	0.078	0.475	0.391	0.171
	S_7	0.357	0.408	0.073	0.370	0.424	0.079
	S_8	0.405	0.433	0.061	0.826	0.793	0.375
	S_9	0.362	0.328	0.087	0.546	0.492	0.134
$n = 50$	S_1	0.504	0.443	0.137	0.675	0.556	0.218
	S_2	0.533	0.427	0.106	0.979	0.939	0.223
	S_3	0.417	0.418	0.117	0.544	0.512	0.150
	S_4	0.446	0.419	0.105	0.718	0.646	0.163
	S_5	0.397	0.326	0.174	0.718	0.575	0.538
	S_6	0.380	0.354	0.108	0.741	0.621	0.275
	S_7	0.527	0.685	0.073	0.505	0.676	0.083
	S_8	0.596	0.721	0.094	0.948	0.928	0.536
	S_9	0.676	0.605	0.115	0.814	0.734	0.182

reasonable candidate model for this data. By the property of our test, we can say more about the data: no significant evidence for the dependence of temperature and the model error.

The second data set considered here is about the series of daily summaries of 73 Spanish weather stations selected for the period 1980-2009, provided in the data set *aemet* in the R *fda.usc* package. The dataset contains geographic information of each station and the average of daily temperature, daily precipitation and daily wind speed for the period 1980–2009. The goal is to check whether there is a linear relationship between logarithm of annual precipitation and the daily temperature curve, and daily temperature curve and the error are independent. Using the proposed test, the p-value is 0.010. The CvM method has a close p-value 0.012. Hence we reject the null hypothesis. This suggests that the functional linear model is not enough to describe the relationship between logarithm of daily precipitation and the daily temperature.

The third data set is *gasoline* in the *refund* package in *R*. This data set consists of 60 near infrared spectral curves and octane numbers of 60 gasoline samples. The goal is to determine the form of dependence between the octane and the spectral curve. The p-value of our test is 0.010, hence we reject the null hypothesis. For the method of CvM, the p-value is 0.650. Consequently, based on results of our test and CvM, we may speculate that the octane and the spectral curve satisfy functional linear model, but the spectral curve and the error are dependent. For further analysis, we should pay more attention when modeling this data by traditional methods with the assumption that the predictor and error are independent. This example also shows how to combine our test with other goodness-of-fit tests to find out more about the relationship among data.

5 Discussion

We have presented a method for testing the independence of the predictor and the error and goodness-of-fit for functional linear models simultaneously. The methodology is based on the generalized distance covariance between the predictor X and the residual obtained by fitting the functional linear model. From the construction of the test procedure, one can easily extend this procedure by replacing generalized distance covariance with other measures of independence, such as ball covariance (Pan et al. 2019). On the other hand, instead of the linear model considered here, one can consider other functional regression models such as functional partially linear model (Aneiros-Pérez and Vieu 2006), functional quadratic regression model (Horváth and Reeder 2013). Using similar arguments, one can construct a similar test procedure for the corresponding model. The implementation is easy but the related theory might require more efforts.

Acknowledgements We would like to thank the editor and the anonymous reviewers for thoughtful comments that led to a substantial improvement of the paper. This work is supported by NSFC with grant No.11771032.

Appendix: technical proofs

We prove the theorems in the Appendix with the asymptotic theory of V-statistics, which could be found in Koroljuk and Borovskich (1994).

Proof of Theorem 1 We divide the proof into three parts as follows.

Step 1: Decomposition of T_n

Observe that

$$\hat{\varepsilon}_i = \varepsilon_i - \langle X_i, \hat{\beta} - \beta \rangle$$

Based on this and by Taylor expansion it holds

$$l_{ij} = l_{ij}^{(0)} + \langle \hat{\beta} - \beta, l_{ij}^{(1)} \rangle + \frac{1}{2} \langle \mathcal{V}_{ij}(\hat{\beta} - \beta), \hat{\beta} - \beta \rangle \tag{8}$$

where

$$l_{ij}^{(0)} = l(\varepsilon_i, \varepsilon_j), \quad l_{ij}^{(1)} = -\{l_x(\varepsilon_i, \varepsilon_j)X_i + l_y(\varepsilon_i, \varepsilon_j)X_j\} \in L^2([0,1]),$$
$$\mathcal{V}_{ij} = \{l_{xx}(\varsigma_{ij}, \tau_{ij})X_i \otimes X_i + l_{yy}(\varsigma_{ij}, \tau_{ij})X_j \otimes X_j + 2l_{xy}(\varsigma_{ij}, \tau_{ij})X_i \otimes X_j\} \in L^{2*}([0,1]),$$

for some point $(\varsigma_{ij}, \tau_{ij})$ on the straight line connecting the two points $(\hat{\varepsilon}_i, \hat{\varepsilon}_j)$ and $(\varepsilon_i, \varepsilon_j)$ on $\mathbb{R}^2$, where $L^{2*}([0,1])$ denotes the space of linear operators from $L^2([0,1])$ to $L^2([0,1])$. By (8), T_n can be decomposed in the following way

$$T_n = T_n^{(0)} + \langle \hat{\beta} - \beta, T_n^{(1)} \rangle + \frac{1}{2} \langle T_n^{(2)}(\hat{\beta} - \beta), \hat{\beta} - \beta \rangle + R_n, \tag{9}$$

where

$$T_n^{(p)} = \frac{1}{n^2} \sum_{i,j}^{n} k_{ij} l_{ij}^{(p)} + \frac{1}{n^4} \sum_{i,j,q,r}^{n} k_{ij} l_{qr}^{(p)} - \frac{2}{n^3} \sum_{i,j,q}^{n} k_{ij} l_{iq}^{(p)} \quad (p = 0, 1, 2),$$
$$l_{ij}^{(2)} = \{l_{xx}(\varepsilon_i, \varepsilon_j)X_i \otimes X_i + l_{yy}(\varepsilon_i, \varepsilon_j)X_j \otimes X_j + 2l_{xy}(\varepsilon_i, \varepsilon_j)X_i \otimes X_j\},$$

and R_n is the reminder term.

We can express $T_n^{(p)}$, $p \in \{0, 1, 2\}$, as V-statistics of the form

$$T_n^{(p)} = \frac{1}{n^4} \sum_{i,j,q,r}^{n} h^{(p)}(Z_i, Z_j, Z_q, Z_r), \tag{10}$$

for the symmetrical kernel $h^{(p)}$ defined as

$$h^{(p)}(Z_i, Z_j, Z_q, Z_r) = \frac{1}{4!} \sum_{(t,u,v,w)}^{(i,j,q,r)} (k_{tu} l_{tu}^{(p)} + k_{tu} l_{vw}^{(p)} - 2k_{tu} l_{tv}^{(p)}), \tag{11}$$

where $Z_i = (X_i, \varepsilon_i)$ and the sum is taken over all $4!$ permutations of (i, j, q, r).

Step 2: Negligibility of the remainder term R_n

In this part we will show that

$$nR_n \xrightarrow{\text{p}} 0. \tag{12}$$

Denote

$$\begin{aligned} Q_n =& \frac{1}{n^2} \sum_{i,j}^{n} k_{ij}(\mathcal{V}_{ij} - l_{ij}^{(2)}) + \frac{1}{n^4} \sum_{i,j,q,r}^{n} k_{ij}(\mathcal{V}_{qr} - l_{qr}^{(2)}) \\ & - \frac{2}{n^3} \sum_{i,j,q}^{n} k_{ij}(\mathcal{V}_{iq} - l_{iq}^{(2)}) \in L^{2*}([0,1]), \end{aligned}$$

then

$$\begin{aligned} |nR_n| &= \frac{1}{2} |\langle Q_n(\sqrt{n}(\hat{\beta} - \beta)), \sqrt{n}(\hat{\beta} - \beta)\rangle| \\ &\le \frac{1}{2} \|Q_n\| \|\sqrt{n}(\hat{\beta} - \beta)\|^2. \end{aligned}$$

Since $\|\sqrt{n}(\hat{\beta} - \beta)\| = O_p(1)$, we only need to show that $\|Q_n\| = o_p(1)$. Note that Q_n is the sum of three terms and each of these terms can be shown to converge to zero in probability. We will only show the first term, the other two terms can be done in a similar way. Using the condition of Lipschitz continuity of l_{xx}, l_{yy} and l_{xy}, we have

$$\begin{aligned} \|\mathcal{V}_{ij} - l_{ij}^{(2)}\| \le& L|(\hat{\varepsilon}_i, \hat{\varepsilon}_j) - (\varepsilon_i, \varepsilon_j)|_\infty (\|X_i\| + \|X_j\|)^2 \\ \le& L\|\hat{\beta} - \beta\| (\|X_i\| + \|X_j\|)^3. \end{aligned}$$

Therefore, $\frac{1}{n^2} \sum_{i,j}^n k_{ij}(\mathcal{V}_{ij} - l_{ij}^{(2)})$ is bounded by

$$L\|\hat{\beta} - \beta\| n^{-2} \sum_{i,j=1}^{n} |k_{ij}| (\|X_i\| + \|X_j\|)^3.$$

By condition C2(b) and the week law of large number for V-statistics,

$$n^{-2} \sum_{i,j=1}^{n} |k_{ij}| (\|X_i\| + \|X_j\|)^3 = O_{\text{p}}(1),$$

hence $\frac{1}{n^2} \sum_{i,j}^n k_{ij}(\mathcal{V}_{ij} - l_{ij}^{(2)}) = o_{\text{p}}(1)$. With similar techniques for the other two terms, we obtain $\|Q_n\| = o_p(1)$.

Step 3: Finding the limiting distribution

By (9) and (12), it is enough to show that the following term

$$nT_n^{(0)} + n\langle \hat{\beta} - \beta, T_n^{(1)} \rangle + \frac{1}{2} n \langle T_n^{(2)}(\hat{\beta} - \beta), \hat{\beta} - \beta \rangle \tag{13}$$

converges in distribution. By conditions C2(c)-(e), $\text{E}\{|h^{(p)}(Z_i, Z_j, Z_q, Z_r)|^2\} < \infty$ for $1 \le i, j, q, r \le 4, p = 0, 1$. With calculation, $\text{E}\{h^{(0)}(z_1, Z_2, Z_3, Z_4)\} = 0$ almost surely, so the kernel $h^{(0)}$ is degenerate. Denote

$$h_2^{(0)}(z_1, z_2) = \mathrm{E}\{h^{(0)}(z_1, z_2, Z_3, Z_4)\}$$

and define the V-statistic $\mathcal{S}_n^{(0)}$ with kernel $h_2^{(0)}$, that is,

$$\mathcal{S}_n^{(0)} = \frac{6}{n^2} \sum_{i,j=1}^{n} h_2^{(0)}(Z_i, Z_j).$$

By the standard results of V-statistics, we have

$$n(T_n^{(0)} - \mathcal{S}_n^{(0)}) \xrightarrow{\mathrm{p}} 0.$$

Define the linear operator $(Af)(s) = \int h_2^{(0)}(s,t)f(t)dP_{X\varepsilon}(t)$ for $f \in L^2(L^2([0,1]) \times \mathbb{R}, P_{X\varepsilon})$, where $L^2(L^2([0,1]) \times \mathbb{R}, P_{X\varepsilon})$ denotes the space consisting of all square integrable functions defined on $L^2([0,1]) \times \mathbb{R}$, and $P_{X\varepsilon}$ is the joint probability measure of X and ε. Then the symmetric function $h_2^{(0)}$ admits an eigenvalue decomposition

$$h_2^{(0)}(z_1, z_2) = \sum_{r=1}^{\infty} \gamma_r \phi_r(z_1)\phi_r(z_2),$$

where $\{\gamma_r\}_{r=1}^{\infty}$ and $\{\phi_r\}_{r=1}^{\infty}$ are the eigenvalues and eigenfunctions of A, respectively, satisfying $\mathrm{E}[\phi_i(Z)\phi_j(Z)] = \delta_{ij}$. Clearly, we have $\mathrm{E}[h_2^{(0)}(Z_1, Z_1)] = \sum_{r=1}^{\infty} \gamma_r$ and $\mathrm{E}[h_2^{(0)}(Z_1, Z_2)]^2 = \sum_{r=1}^{\infty} \gamma_r^2$. Since $\mathrm{E}\{|h^{(0)}(Z_1, Z_2, Z_3, Z_4)|^2\} < \infty$, by the results in page 182 of Serfling (1980), $\mathrm{E}[h_2^{(0)}(Z_1, Z_2)]^2 < \infty$. Similarly, we also have $\mathrm{E}|h^{(0)}(Z_1, Z_1)| < \infty$. Hence, $|\sum_{r=1}^{\infty} \gamma_r| < \infty$ and $\sum_{r=1}^{\infty} \gamma_r^2 < \infty$. Note that

$$\{n^{-\frac{1}{2}} \sum_{i=1}^{n} \phi_r(Z_i)\}^2 = n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} \phi_r(Z_i)\phi_r(Z_j).$$

In view of this, $n\mathcal{S}_n^{(0)}$ can be expressed as

$$n\mathcal{S}_n^{(0)} = 6 \sum_{r=1}^{\infty} \gamma_r \{n^{-\frac{1}{2}} \sum_{i=1}^{n} \phi_r(Z_i)\}^2.$$

Now let us turn to the terms $T_n^{(1)}$ and $T_n^{(2)}$ in (13). It can be shown that $\mathrm{E}\{h^{(1)}(Z_1, Z_2, Z_3, Z_4)\} = 0$. Define

$$h_1^{(1)}(z_1) = \mathrm{E}\{h^{(1)}(z_1, Z_2, Z_3, Z_4)\},$$

then, by the standard theory of V-statistics,

$$n^{1/2}T_n^{(1)} - 4n^{-1/2} \sum_{i=1}^{n} h_1^{(1)}(Z_i) \xrightarrow{\mathrm{p}} 0.$$

Meanwhile, by the weak law of large numbers for V-statistics,

$$T_n^{(2)} \xrightarrow{\text{p}} \mathrm{E}\{h^{(2)}(Z_1, Z_2, Z_3, Z_4)\} := \Lambda.$$

Recall that $\sqrt{n}(\hat{\beta} - \beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(Z_i) + o_p(1)$. According to the multivariate central limit theorem and the Theorem 1.4.8 in Vaart and Wellner (1996), the countable random sequence

$$\left\{ n^{-1/2} \sum_{i=1}^{n} \phi_r(Z_i) \right\}_{r \geq 1}, \quad \left\{ n^{-1/2} \sum_{i=1}^{n} 4h_1^{(1)}(Z_i) \right\}, \quad \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(Z_i) \right\}$$

convergences in distribution to the joint Gaussian random sequence

$$\mathcal{Z} = (\mathcal{Z}_r)_{r \geq 1}, \quad \mathcal{N}, \quad \mathcal{G},$$

where $\mathcal{Z}_r$ are i.i.d. $N(0, 1)$ random variables, $\mathcal{N}$, $\mathcal{G}$ are Gaussian random functions in $L^2([0,1])$ with mean zero and covariance functions $\mathrm{cov}(\mathcal{N}(s), \mathcal{N}(t)) = \mathrm{E}\{16h_1^{(1)}(X(s), \varepsilon)h_1^{(1)}(X(t), \varepsilon)\}$ and $\mathrm{cov}(\mathcal{G}(s), \mathcal{G}(t)) = \mathrm{E}\{\psi(X(s), \varepsilon)\psi(X(t), \varepsilon)\}$, respectively. Then, by the continuous mapping theorem, we have

$$\begin{aligned} nT_n =& nT_n^{(0)} + n\langle \hat{\beta} - \beta, T_n^{(1)} \rangle + \frac{1}{2} n \langle T_n^{(2)}(\hat{\beta} - \beta), \hat{\beta} - \beta \rangle + o_{\text{p}}(1) \\ & \xrightarrow{d} 6 \sum_{j=1}^{\infty} \gamma_j \mathcal{Z}_j^2 + \langle \mathcal{G}, \mathcal{N} \rangle + \langle \Lambda(\mathcal{G}), \mathcal{G} \rangle. \end{aligned}$$

□

Proof of Theorem 2 Recall that

$$\hat{\varepsilon}_i = \varepsilon_i - \langle X_i, \hat{\beta} - \beta \rangle.$$

Using Taylor expansion of l_{ij} to order 1, we have, almost surely,

$$l_{ij} = l_{ij}^{(0)} + \langle \hat{\beta} - \beta, \mathcal{V}_{ij} \rangle \tag{14}$$

where $l_{ij}^{(0)} = l(\varepsilon_i, \varepsilon_j)$ and $\mathcal{V}_{ij} = -\{l_x(\varsigma_{ij}, \tau_{ij})X_i + l_y(\varsigma_{ij}, \tau_{ij})X_j\}$ for some point $(\varsigma_{ij}, \tau_{ij})$ on the straight line connecting the two points $(\hat{\varepsilon}_i, \hat{\varepsilon}_j)$ and $(\varepsilon_i, \varepsilon_j)$ on $\mathbb{R}^2$. By (14), T_n can be decomposed in three terms

$$T_n = T_n^{(0)} + \langle \hat{\beta} - \beta, T_n^{(1)} \rangle + \langle \hat{\beta} - \beta, R_n \rangle \tag{15}$$

where $T_n^{(p)}$, $p = 0, 1$ are defined the same as in the proof of Theorem 1 and

$$R_n = \frac{1}{n^2} \sum_{i,j}^{n} k_{ij}(\mathcal{V}_{ij} - l_{ij}^{(1)}) + \frac{1}{n^4} \sum_{i,j,q,r}^{n} k_{ij}(\mathcal{V}_{qr} - l_{qr}^{(1)}) - \frac{2}{n^3} \sum_{i,j,q}^{n} k_{ij}(\mathcal{V}_{iq} - l_{iq}^{(1)}).$$

Under H_0, the predictor X and the error ε are independent, therefore the generalized distance $\theta(X, \varepsilon)$ between X and ε is zero. By the condition C4(c)-(e),

$\mathrm{E}\{|h^{(p)}(Z_i, Z_j, Z_q, Z_r)|\} < \infty$ for $1 \le i, j, q, r \le 4, p = 0, 1$. By the law of large number for V-statistics,

$$T_n^{(0)} \xrightarrow{\mathrm{p}} \theta(X, \varepsilon) = 0,$$

$$T_n^{(1)} \xrightarrow{\mathrm{p}} \mathrm{E}\{|h^{(1)}(Z_i, Z_j, Z_q, Z_r)|\}.$$

Therefore $T_n^{(1)} = O_p(1)$. Observe that

$$\langle \hat{\beta} - \beta, T_n^{(1)} \rangle \le \|\hat{\beta} - \beta\| \|T_n^{(1)}\|,$$
$$\langle \hat{\beta} - \beta, R_n \rangle \le \|\hat{\beta} - \beta\| \|R_n\|.$$

By condition C3, $\|\hat{\beta} - \beta\| = o_p(1)$. We only need to show that $\|R_n\| = O_p(1)$. Note that R_n is a sum of three terms and each of these terms can be shown to converge to zero in probability. We will only show the first term, the other two terms can be done in a similar way. Using the condition of Lipschitz continuity of l_x and l_y, we have

$$\begin{aligned}\|\mathcal{V}_{ij} - l_{ij}^{(1)}\| &\le L|(\hat{\varepsilon}_i, \hat{\varepsilon}_j) - (\varepsilon_i, \varepsilon_j)|_\infty (\|X_i\| + \|X_j\|) \\ &\le L\|\hat{\beta} - \beta\|(\|X_i\| + \|X_j\|)^2.\end{aligned}$$

Therefore, $\frac{1}{n^2}\sum_{i,j}^n k_{ij}(\mathcal{V}_{ij} - l_{ij}^{(1)})$ is bounded by

$$L\|\hat{\beta} - \beta\| n^{-2} \sum_{i,j=1}^n |k_{ij}|(\|X_i\| + \|X_j\|)^2.$$

By condition C4(b) and the week law of V-statistics,

$$n^{-2} \sum_{i,j=1}^n |k_{ij}|(\|X_i\| + \|X_j\|)^2 = O_{\mathrm{p}}(1),$$

hence $\frac{1}{n^2}\sum_{i,j}^n k_{ij}(\mathcal{V}_{ij} - l_{ij}^{(1)}) = O_{\mathrm{p}}(1)$. With similar techniques for the other two terms, we obtain $\|R_n\| = O_p(1)$. □

Proof of Theorem 3 Let $\epsilon_i = m(X_i) - \langle X_i, \tilde{\beta} \rangle + \varepsilon_i$. Even though $m(x)$ might not be linear, $\langle X, \tilde{\beta} \rangle$ is the closest function to $m(x)$ in $\mathcal{M}_{L^2([0,1])}$ in the sense of square loss. By the consistency of M-estimator (see Corollary 3.2.3 in Vaart and Wellner 1996), the estimator $\hat{\beta}$ convergence in probability to $\tilde{\beta}$, that is, $\|\hat{\beta} - \tilde{\beta}\| = o_{\mathrm{p}}(1)$. Using Taylor expansion, we have, almost surely,

$$l(\hat{\varepsilon}_i, \hat{\varepsilon}_j) = l(\epsilon_i, \epsilon_j) + \{(\hat{\varepsilon}_i - \epsilon_i) l_x(\varsigma_{ij}, \tau_{ij}) + (\hat{\varepsilon}_j - \epsilon_j) l_y(\varsigma_{ij}, \tau_{ij})\},$$

where $(\varsigma_{ij}, \tau_{ij})$ is some point on the line connecting the two points $(\hat{\varepsilon}_i, \hat{\varepsilon}_j)$ and (ϵ_i, ϵ_j). Note that

$$\hat{\varepsilon}_i - \epsilon_i = -\langle X_i, \hat{\beta} - \tilde{\beta} \rangle,$$

we can decompose

$$T_n = T_n^{(0)} + \langle \hat{\beta} - \tilde{\beta}, T_n^{(1)} \rangle + R_n, \tag{16}$$

where

$$\begin{aligned}
T_n^{(p)} &= \frac{1}{n^2}\sum_{i,j}^{n} k_{ij} l_{ij}^{(p)} + \frac{1}{n^4}\sum_{i,j,q,r}^{n} k_{ij} l_{qr}^{(p)} - \frac{2}{n^3}\sum_{i,j,q}^{n} k_{ij} l_{i,q}^{(p)} \quad (p = 0, 1),\\
l_{ij}^{(0)} &= l_{ij} = l(\epsilon_i, \epsilon_j), \quad l_{ij}^{(1)} = -\{l_x(\epsilon_i, \epsilon_j)X_i + l_y(\epsilon_i, \epsilon_j)X_j\} \in \mathcal{H},\\
R_n &= \langle \hat{\beta} - \tilde{\beta}, \frac{1}{n^2}\sum_{i,j}^{n} k_{ij} l_{ij}^{*} + \frac{1}{n^4}\sum_{i,j,q,r}^{n} k_{ij} l_{qr}^{*} - \frac{2}{n^3}\sum_{i,j,q}^{n} k_{ij} l_{i,q}^{*} \rangle,
\end{aligned}$$

and

$$l_{ij}^{*} = -\{(l_x(\varsigma_{ij}, \tau_{ij}) - l_x(\epsilon_i, \epsilon_j))X_i + (l_y(\varsigma_{ij}, \tau_{ij}) - l_y(\epsilon_i, \epsilon_j))X_j\}.$$

We will show that $T_n^{(0)} \xrightarrow{\mathrm{p}} \tau$, $\tau > 0$, $\langle \hat{\beta} - \tilde{\beta}, T_n^{(1)} \rangle = o_{\mathrm{p}}(1)$ and $R_n = o_{\mathrm{p}}(1)$. By the results in the proof of Theorem 2, $\langle \hat{\beta} - \tilde{\beta}, T_n^{(1)} \rangle = o_{\mathrm{p}}(1)$ and $R_n = o_{\mathrm{p}}(1)$.

Now we show that $T_n^{(0)} \xrightarrow{\mathrm{p}} \tau$, $\tau > 0$. Using the same arguments of proof of Theorem 1, $T_n^{(0)}$ is a V-statistic. By the weak law of V-statistics, $T_n^{(0)}$ convergence in probability to generalized distance covariance of X and ϵ , $\theta(X, \epsilon)$. Under $H_{1,1}$, $\epsilon = \varepsilon$, hence X and ϵ are dependent. Under scenarios $H_{1,2}$ or $H_{1,3}$, $m(X) \neq \langle X, \tilde{\beta} \rangle$ with positive probability. And the conditional mean of ϵ given X is

$$\mathrm{E}(\epsilon | X) = m(X) - \langle X, \tilde{\beta} \rangle + \mathrm{E}(\varepsilon | X) = m(X) - \langle X, \tilde{\beta} \rangle.$$

With the condition that $m(X) - \langle X, \tilde{\beta} \rangle$ is a non-constant function of X, $\mathrm{E}(\epsilon | X)$ depends on X, and hence X and ϵ are dependent. Since k and l are strong negative type, $\tau = \theta(X, \epsilon) > 0$. □

References

Aneiros-Pérez, G., & Vieu, P. (2006). Semi-functional partial linear regression. *Statistics and Probability Letters*, *76*(11), 1102–1110.

Cai, T. T., & Hall, P. (2006). Prediction in functional linear regression. *The Annals of Statistics*, *34*(5), 2159–2179.

Cai, T. T., & Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. *Journal of the American Statistical Association*, *107*(499), 1201–1216.

Cardot, H., Ferraty, F., Mas, A., & Sarda, P. (2003). Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics*, *30*(1), 241–255.

Cardot, H., Mas, A., & Sarda, P. (2007). Clt in functional linear regression models. *Probability Theory and Related Fields*, *138*(3–4), 325–361.

Crambes, C., Kneip, A., Sarda, P., et al. (2009). Smoothing splines estimators for functional linear regression. *The Annals of Statistics*, *37*(1), 35–72.

Cuesta-Albertos, J. A., García-Portugués, E., Febrero-Bande, M., González-Manteiga, W., et al. (2019). Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes. *The Annals of Statistics*, *47*(1), 439–467.

Delsol, L., Ferraty, F., & Vieu, P. (2011). Structural test in regression on functional variables. *Journal of Multivariate Analysis*, *102*(3), 422–447.

Fukumizu, K., Gretton, A., Schölkopf, B., & Sriperumbudur, B. K. (2009). Characteristic kernels on groups and semigroups. *Advances in Neural Information Processing Systems*, 473–480.

García-Portugués, E., González-Manteiga, W., & Febrero-Bande, M. (2014). A goodness-of-fit test for the functional linear model with scalar response. *Journal of Computational and Graphical Statistics*, *23*(3), 761–778.

Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B., & Smola, A. J. (2008). A kernel statistical test of independence. *Advances in neural information processing systems*, 585–592.

Hall, P., & Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics*, *35*(1), 70–91.

Hall, P., & Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *68*(1), 109–126.

Hilgert, N., Mas, A., Verzelen, N., et al. (2013). Minimax adaptive tests for the functional linear model. *The Annals of Statistics*, *41*(2), 838–869.

Horváth, L., & Reeder, R. (2013). A test of significance in functional quadratic regression. *Bernoulli Official Journal of the Bernoulli Society for Mathematical Statistics and Probability*, *19*(5A), 225–232.

Huang, L., Wang, H., & Zheng, A. (2014). The m-estimator for functional linear regression model. *Statistics and Probability Letters*, *88*, 165–173.

Kokoszka, P., Maslova, I., Sojka, J., & Zhu, L. (2008). Testing for lack of dependence in the functional linear model. *Canadian Journal of Statistics*, *36*(2), 207–222.

Koroljuk, V., & Borovskich, Y. (1994). *Theory of U-statistics*. Springer Science+Business Media Dordrecht.

Lyons, R. (2013). Distance covariance in metric spaces. *The Annals of Probability*, *41*(5), 3284–3305.

Pan, W., Wang, X., Zhang, H., Zhu, H., & Zhu, J. (2019). Ball covariance: A generic measure of dependence in banach space. *Journal of the American Statistical Association*, 1–24.

Patilea, V., Sanchez-Sellero, C., & Saumard, M. (2012). Projection-based nonparametric goodness-of-fit testing with functional covariates. arXiv preprint arXiv:1205.5578.

Patilea, V., Sánchez-Sellero, C., & Saumard, M. (2016). Testing the predictor effect on a functional response. *Journal of the American Statistical Association*, *111*(516), 1684–1695.

Ramsay, J., Hooker, G., & Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. New York: Springer.

Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis*. New York: Springer.

Sejdinovic, D., Sriperumbudur, B., Gretton, A., Fukumizu, K., et al. (2013). Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *The Annals of Statistics*, *41*(5), 2263–2291.

Sen, A., & Sen, B. (2014). Testing independence and goodness-of-fit in linear models. *Biometrika*, *101*(4), 927–942.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. Hoboken: Wiley.

Sriperumbudur, B. K., Gretton, A., Fukumizu, K., Lanckriet, G., & Schölkopf, B. (2008). Injective hilbert space embeddings of probability measures. In *21st Annual Conference on Learning Theory (COLT 2008)*, pages 111–122. Omnipress.

Sriperumbudur, B. K., Gretton, A., Fukumizu, K., Schölkopf, B., & Lanckriet, G. R. (2010). Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research*, 11(Apr):1517–1561.

Székely, G. J., & Rizzo, M. L. (2009). Brownian distance covariance. *The Annals of Applied Statistics*, *3*(4), 1236–1265.

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

Vaart, A. W. V. D., & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes*. New York: Springer.

van der Vaart, A. W. (2000). *Asymptotic statistics*. Cambridge: Cambridge University Press.

Yuan, M., Cai, T. T., et al. (2010). A reproducing kernel hilbert space approach to functional linear regression. *The Annals of Statistics*, *38*(6), 3412–3444.