

## RESEARCH ARTICLE

## Nonparametric testing of lack of dependence in functional linear models

Wenjuan Hu<sup>1,2</sup>, Nan Lin<sup>3</sup>, Baoxue Zhang<sup>1\*</sup>

**1** School of Statistics, Capital University of Economics and Business, Beijing, China, **2** School of Mathematics and Statistics, Northeast Normal University, Changchun, Jilin Province, China, **3** Department of Mathematics, College of Arts and Sciences, Washington University in St. Louis, St. Louis, Missouri, United States of America

\* [zhangbaoxue@cueb.edu.cn](mailto:zhangbaoxue@cueb.edu.cn)

## Abstract

An important inferential task in functional linear models is to test the dependence between the response and the functional predictor. The traditional testing theory was constructed based on the functional principle component analysis which requires estimating the covariance operator of the functional predictor. Due to the intrinsic high-dimensionality of functional data, the sample is often not large enough to allow accurate estimation of the covariance operator and hence causes the follow-up test underpowered. To avoid the expensive estimation of the covariance operator, we propose a nonparametric method called Functional Linear models with U-statistics TEsting (FLUTE) to test the dependence assumption. We show that the FLUTE test is more powerful than the current benchmark method (Kokoszka P, 2008; Patilea V, 2016) in the small or moderate sample case. We further prove the asymptotic normality of our test statistic under both the null hypothesis and a local alternative hypothesis. The merit of our method is demonstrated by both simulation studies and real examples.

## OPEN ACCESS

**Citation:** Hu W, Lin N, Zhang B (2020) Nonparametric testing of lack of dependence in functional linear models. PLoS ONE 15(6): e0234094. <https://doi.org/10.1371/journal.pone.0234094>

**Editor:** Xiaofeng Wang, Cleveland Clinic Lerner Research Institute, UNITED STATES

**Received:** December 31, 2019

**Accepted:** May 18, 2020

**Published:** June 26, 2020

**Copyright:** © 2020 Hu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data underlying this study are from the ‘CanadianWeather’ dataset, which is publicly available. The authors used the ‘CanadianWeather’ dataset by loading the R package ‘fda’ directly in the R program. The link of the R package ‘fda’ is available here: <https://cran.r-project.org/web/packages/fda/> and the description of this dataset is here: <https://www.rdocumentation.org/packages/fda/versions/5.1.4/topics/CanadianWeather>. Another link to the dataset is here: <http://www.psych.mcgill.ca/misc/fda/ex-weather-a1.html>. Ramsay, J. O. and

## Introduction

Functional regression studies how a response variable  $Y$  varies with a functional predictor  $X(s)$ , where  $Y$  can be scalar ( $Y \in \mathcal{R}$ ) or functional ( $Y(t) \in L^2([0, 1])$ ). The space  $L^2([0, 1])$  denotes the Hilbert space for square integrable functions. Without loss of generality, we define the index  $s$  and  $t$  on the closed interval  $[0, 1]$ . In case the raw support of  $s$  and  $t$  is a closed interval  $[a, b]$ , one can simply rescale it to the interval  $[0, 1]$ . In this paper, we assume data following the widely used functional linear model (FLM) [1–4]. For a functional response, the FLM is defined as

$$Y(t) = \alpha(t) + \int_0^1 \beta(t, s)X(s)ds + \epsilon(t), \quad (1)$$

where both the intercept  $\alpha(t)$  and the random error  $\epsilon(t)$  are square integrable and independent of  $X(s)$ , the regression coefficient  $\beta(t, s)$  is in  $L^2[0, 1] \times L^2[0, 1]$ . Denote that

Silverman, B. W. (2005). *Functional Data Analysis*. The authors did not have special access privileges.

**Funding:** Research was supported by the National Natural Science Foundation of China (Grant No.11671268). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

$\mathcal{B}(X) = \int_0^1 \beta(t, s)X(s)ds$ , where  $\mathcal{B}$  is the regression operator  $\mathcal{B} : L^2[0, 1] \rightarrow L^2[0, 1]$ . For a scalar response variable  $Y$ , the FLM has a simpler form  $Y = \alpha + \int_0^1 X(s)\beta(s)ds + \epsilon$ , where both the intercept  $\alpha$  and the random error  $\epsilon$  are real valued, and the regression parameter  $\beta(s) \in L^2[0, 1]$ . Hereafter, we mainly focus on FLMs with functional responses, but the general methodology also applies to scalar responses.

In this paper, we consider testing whether the regression operator  $\mathcal{B}$  has an assigned structure  $\mathcal{B}_0$ , that is, to test

$$\mathcal{H}_0 : \mathcal{B} = \mathcal{B}_0 \quad \text{versus} \quad \mathcal{H}_1 : \mathcal{B} \neq \mathcal{B}_0. \quad (2)$$

In practice, people often focus on the special case with  $\mathcal{B}_0 = 0$ , i.e. to test the dependency between the response variable and the predictor. Existing tests in the literature for this problem can be categorized into parametric and nonparametric tests. In parametric tests, the test statistics are usually established by first estimating the functional regression coefficient through dimension reduction, such as functional PCA [5, 6–8]. Methods for real-valued responses include [6], [7] and [8]. [6] used a test statistic based on the  $L_2$  norm of the empirical cross-covariance operator of  $(X, Y)$ . [8] proposed a Wald-type test with varying thresholds in selecting the number of principal components. [7] developed four test statistics based on the functional principal component (FPC) scores. They assume normality on the error distribution due to the need of the likelihood function. For functional responses, the test statistic proposed by [5] is constructed based on the eigenvalues and eigenfunctions decomposed from the functional PCA of the response variable  $Y(t)$  and the predictor  $X(s)$ . Such parametric methods require the costly estimation of the covariance operator of the predictor. Due to the intrinsic high dimensionality of functional data, the inaccuracy and numerical instability in the covariance operator estimation may render the parametric tests invalid especially for small or moderate size samples. The same issue also occurs in high dimensional problems in multivariate statistics [9, 10]. Another limitation of FPC is that the principal component scores are computed independently from the predictor. Then the directions which explain  $X(t)$  best may not be the best predictors for the response which may lead to disparate test results for the regression problems. On the other hand, nonparametric tests utilized a different idea to avoid estimating the covariance operator [11, 12, 15]. For real-valued responses, [11] used the Nadaraya-Watson technique [13, 14] to estimate the conditional mean of  $Y - \mathcal{B}_0(X)$  given  $X = x$ . [15] also proposed a nonparametric test based on a kernel function for real responses. However, this method still requires estimating the covariance operator to calculate the semi-metric. Furthermore, this test needs to split the sample into three groups, one of the three groups is used to estimate the kernel function, another to estimate the sample mean of responses variables, and the last group contributes to statistic, which is suitable for large sample data. The test proposed by [12] is for functional responses, and its test statistic is a weighted U-statistic with weights obtained from nearest neighbor smoothing. While this test possesses the correct Type-I error rate, identification of the neighbors requires defining distances between the functional predictors in the least favorable direction, which tends to result in lower power in general.

Motivated by [10], we propose a novel nonparametric test, called FLUTE, based on a U-statistic that measures the  $L_2$  distance in the induced space after transforming the original space of the functional predictor by the covariance operator. Our approach avoids explicit estimation of the covariance operator as it is based on the distance in the induced space. The FLUTE test can be applied to both real-valued and functional responses.

The paper is organized as follows. In Section Methodology, we introduce basic notations about functional linear regression model and the FLUTE test statistic. After presenting the

theory in Section Asymptotic theory for functional responses, we further discuss the FLUTE test for FLMs with a scalar response in the next Section. Section Simulation and real data reports results from simulation studies and real data. The last section is the conclusion section.

## Methodology

### Notation and assumptions

Let  $\langle \cdot, \cdot \rangle$  denote the inner product in  $L^2[0, 1]$ , that is, for any  $f_1, f_2 \in L^2[0, 1]$   $\langle f_1, f_2 \rangle = \int_0^1 f_1(t)f_2(t)dt$ . The  $L_2$  norm  $\|\cdot\|$  is defined by  $\|f\|^2 = \langle f, f \rangle$ . We assume in the FLM (1) that both the predictor  $X(s)$  and response variable  $Y(t)$  are random elements of  $L^2[0, 1]$  and integrable. The sample functions  $X_i(s), i = 1, \dots, n$ , are independently and identically distributed (i.i.d.) with  $E[X(s)] = \mu_x(s)$  and  $E\|X(s)\|^4 < \infty$ . We also assume that the random trajectories  $\epsilon_i(t)$  are i.i.d with  $E[\epsilon(t)] = 0, E[\epsilon^2(t)] = \sigma^2(t)$  and  $E\|\epsilon(t)\|^4 < \infty$ .

Suppose  $\{\phi_k, k \geq 1\}$  and  $\{\eta_\ell, \ell \geq 1\}$  are some orthonormal bases of the Hilbert space  $\mathcal{H}_1$  and  $\mathcal{H}_2$ , respectively. To simplify the representation, hereafter we focus on the case where the Hilbert spaces  $\mathcal{H}_1$  and  $\mathcal{H}_2$  are  $L^2[0, 1]$ . Then we represent the predictor  $X_i(s)$  and the regression coefficient  $\beta(t, s)$  via the Karhunen-Loève decomposition [16]. For any  $s \in [0, 1]$ , we have

$$X_i(s) = \mu_x(s) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(s), \tag{3}$$

where  $\mu_x(s)$  is the mean function of the predictor  $X_i(s)$ , and the expansion coefficient  $\xi_{ik} = \langle X_i, \phi_k \rangle$ , with  $E[\xi_{ik}] = 0$ . For any  $t, s \in [0, 1]$ , the regression coefficient  $\beta(t, s)$  is represented as

$$\beta(t, s) = \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \beta_{\ell k} \eta_\ell(t) \phi_k(s), \tag{4}$$

where  $\beta_{\ell k} = \iint \beta(t, s) \eta_\ell(t) \phi_k(s) dt ds$ .

Next we introduce the covariance operator  $\mathcal{C}$  of the predictor  $X(s)$  and its empirical counterpart  $\mathcal{C}_n$ . For any element  $f \in L^2([0, 1])$ , we define

$$\mathcal{C}(f) = E[\langle X - \mu_x, f \rangle (X - \mu_x)] \quad \text{and} \quad \mathcal{C}_n(f) = \frac{1}{n} \sum_{i=1}^n \langle X_i - \bar{X}, f \rangle (X_i - \bar{X}),$$

where  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ . And denote the corresponding eigenlements by  $\mathcal{C}(v_k) = \lambda_k v_k$  with the eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots$  and  $v_k$  the eigenfunction corresponding to  $\lambda_k$ .

If both the predictor  $X(s)$  and response variable  $Y(t)$  are fully observed, hereafter we assume that  $\alpha(t) = 0$  and  $\mu_x(s) = 0$  which will be explained in the next section, then the FLM with a functional response (1) can be represented as,

$$Y(t) = \int_0^1 \sum_{k=1}^{\infty} \sum_{\ell=1}^{\infty} \beta_{\ell k} \eta_\ell(t) \phi_k(s) \sum_{k'=1}^{\infty} \xi_{k'} \phi_{k'}(s) ds + \epsilon(t). \tag{5}$$

In practice, the infinite expansion (3), (4) and (5) above is usually approximated by a truncated basis expansion (e.g. B-spline basis and Fourier basis) [4, 11, 16, 17]. If the functional variables are densely observed, then recovering each trajectory of the functional variables based on the least square method is straightforward [18, 19]. If the functional variables are sparsely observed, [20] and [21] proposed to estimate the FPC scores through local linear surface smoother for the covariance operator, and then approximate each trajectory using the first  $K$  eigenfunctions. For the sparse observation, the error can not be ignored. Due to the

certified complexity of the asymptotic normality of the statistic, we leave this area for future investigation.

We will represent the FLM with a functional response (1) using basis expansion when the approximation error is controlled. Denote  $e_i(t)$  as the approximation error produced through  $\sum_{k=1}^K \xi_{ik} \phi_k(s)$  and  $\sum_{k=1}^K \sum_{\ell=1}^L \beta_{\ell k} \eta_\ell(t) \phi_k(s)$  approximating  $\mathcal{B}(X)$ , that is,

$$e_i(t) = \int_0^1 \beta(t, s) x_i(s) ds - \sum_{k=1}^K \sum_{\ell=1}^L \xi_{ik} \beta_{\ell k} \eta_\ell(t) \tag{6}$$

Similar with Condition 1 in Appendix B in [22], then by the Cauchy-Schwarz inequality, uniformly across all  $i = 1, \dots, n$ , we have  $\langle e_i, e_i \rangle \leq 6\tilde{C}C/K^4 + 3C^2/L^4$  (Please see lemma 1), where  $C$  and  $\tilde{C}$  are two positive constants.

As  $K, L \rightarrow \infty$ , the approximation error should be more precise and become ignored. Hence the FLM with a functional response (1) can be rewritten as,

$$Y_i(t) = \int_0^1 \sum_{k=1}^K \sum_{\ell=1}^L \beta_{\ell k} \eta_\ell(t) \phi_k(s) \sum_{k=1}^K \xi_{ik} \phi_k(s) ds + \epsilon_i(t). \tag{7}$$

In this paper, we assume that both the predictor  $X(s)$  and response variable  $Y(t)$  are fully observed or the approximation error is controlled.

### The FLUTE test

In this section, we introduce the FLUTE test whose test statistic is a U-statistic. The theory of U-statistics for fixed-dimensional data, pioneered by [23], has been well documented; see [24] and [25] for summaries. Recently, [10] developed the theory for high-dimensional multivariate data.

Under the functional linear model (1), if  $\alpha(t) = 0$  and  $E[X(s)] = 0$ , we can see that  $\|Y(t) - \mathcal{B}_0(X)\|^2 = \langle \mathcal{B}(X) - \mathcal{B}_0(X) + \epsilon(t), \mathcal{B}(X) - \mathcal{B}_0(X) + \epsilon(t) \rangle$ , which is then the perturbed  $L^2$  norm for measuring the distance between  $\mathcal{B}(X)$  and  $\mathcal{B}_0(X)$ . Further, it is easy to see that

$$E[\|Y(t) - \mathcal{B}_0(X)\|^2] = \int \int \int \{ \beta(t, s) - \beta_0(t, s) \} c(s, e) \{ \beta(t, e) - \beta_0(t, e) \} ds de dt + \int \sigma^2(t) dt, \tag{8}$$

where  $c(s, e) = E[X(s)X(e)]$ . Note that  $\mathcal{B} = \mathcal{B}_0$  is equivalent to the first term on the right-hand side of Eq (8) being zero. Thus we may consider testing the hypothesis (2) by a U-statistic with  $\int (Y_i(t) - \mathcal{B}_0(X_i)) \langle X_i, X_j \rangle (Y_j(t) - \mathcal{B}_0(X_j)) dt$  as the kernel, whose expectation is  $\int \int \mathcal{C}^2(\beta(t, s) - \beta_0(t, s))(t, e) de dt = \|\mathcal{C}(\beta - \beta_0)\|^2$ , where  $\mathcal{C}(\beta) = \int \beta(t, s) c(s, e) ds$ .

For the general case where  $\alpha(t) \neq 0$  and  $E[X(s)] \neq 0$ , we consider the U-statistic  $T_n$ ,

$$T_n = \frac{1}{\binom{n}{4}} \sum_{\mathcal{A}} \psi(i_1, i_2, i_3, i_4), \tag{9}$$

where  $\psi(i_1, i_2, i_3, i_4) = \frac{1}{3} \{ \varphi(i_1, i_2, i_3, i_4) + \varphi(i_1, i_3, i_2, i_4) + \varphi(i_1, i_4, i_2, i_3) \}$ , and  $\varphi(i_1, i_2, i_3, i_4) = \frac{1}{4} \langle X_{i_1} - X_{i_2}, X_{i_3} - X_{i_4} \rangle \int \pi_{i_1, i_2} \pi_{i_3, i_4} dt$ , with  $\pi_{i_j} = Y_i(t) - Y_j(t) - \mathcal{B}_0(X_i - X_j)$ , and  $\mathcal{A} = \{ (i_1, \dots, i_4) : 1 \leq i_1 < \dots < i_4 \leq n \}$  denotes combinations over all subscripts  $(i_1, \dots, i_4)$ . As the statistic  $T_n$  is invariant to location shifts in both  $X_i(s)$  and  $Y_i(t)$ , without loss

of generality, we assume that  $\alpha(t) = 0$  and  $\mu_x(s) = 0$  in the rest of the paper. Define  $\theta(F) = E[\psi(i_1, i_2, i_3, i_4)]$ , then  $E[T_n] = \theta(F)$ .

As the statistic  $T_n$  measures the distance between the regression operator  $\mathcal{B}$  and the assumed structure  $\mathcal{B}_0$  under the null hypothesis, large values of the statistic  $T_n$  are in favor of the alternative hypothesis and leads to rejection of the null hypothesis.

For the representation of the predictor  $X(s)$ , we have

$$X(s) = \xi' \Phi(s), \tag{10}$$

where  $\Phi(s) = (\phi_1(s), \dots, \phi_K(s))'$ ,  $\xi = (\xi_1, \dots, \xi_K)'$ , and  $\text{var}[\xi] = \Sigma$ . We next follow the general condition in [9] and assume that the loadings  $\xi$  of the predictor  $X(s)$  have a factor design structure.

**Assumption 1** There exists a  $m$ -variate random vector

$\mathbf{N} = (N_1, \dots, N_m)'$  for some  $m < \infty$  so that  $\xi = \Gamma \mathbf{N}$ . Here  $\Gamma$  is a  $K \times m$  matrix such that  $\Gamma \Gamma' = \Sigma$ , and  $E[\mathbf{N}] = \mathbf{0}$ ,  $\text{var}[\mathbf{N}] = I_m$ , where  $I_m$  is the  $m \times m$  identity matrix. Each random variable  $N_\ell$ ,  $\ell = 1, \dots, m$ , is assumed to have finite 8th moment and  $E(N_\ell^4) = 3 + \rho$  for some constant  $\rho \in [0, 1)$ . Further, for any  $\sum_{d=1}^D \ell_d \leq 8$  and  $1 \leq m_1 < m_2 < \dots < m_d \leq m$ , we assume  $E[N_{m_1}^{\ell_1} N_{m_2}^{\ell_2} \dots N_{m_d}^{\ell_d}] = E[N_{m_1}^{\ell_1}] E[N_{m_2}^{\ell_2}] \dots E[N_{m_d}^{\ell_d}]$ .

Assumption 1 allows factors  $\mathbf{N}$  to have a weak correlation. If the predictor  $X(s)$  follows a Gaussian process, [16] pointed out that  $X(s)$  admits the following expansion

$$X(s) \stackrel{d}{=} \sum_{k=1}^m \sqrt{\lambda_k} N_k v_k(s),$$

with independent standard normal random variables  $N_k$ 's. It is easy to see that this is a special case of the factor design structure in Assumption 1, where the  $(a, b)$ th element of the transformation matrix  $\Gamma_{K \times m}$  is  $\sqrt{\lambda_b} \langle v_b, \phi_a \rangle$ .

Let  $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{iL})'$  which is the expansion coefficients of  $\epsilon(t)$ , and  $\Lambda = \text{var}[\epsilon]$ . We assume the following assumption.

**Assumption 2** For  $i \neq j$ ,  $E[(\epsilon'_i \epsilon_j)^4] = O(\text{tr}^2(\Lambda^2))$  and  $E[(\epsilon'_i \Lambda \epsilon_i)^2] = O(\text{tr}^2(\Lambda^2))$ .

### Asymptotic theory

In this section, we derive the asymptotic unbiasedness of the FLUTE test and the asymptotic normality of its test statistic under both the null and a local alternative hypothesis through the Hoeffding decomposition.

Let  $W_i = (X_i(s), \epsilon_i(t))$ , where  $\epsilon_i(t) = Y_i(t) - \mathcal{B}_0(X_i)(t)$ . Thus,  $\psi(i_1, i_2, i_3, i_4)$  in Eq (9) can be represented as  $\psi(W_{i_1}, W_{i_2}, W_{i_3}, W_{i_4})$ . And  $\psi_c(w_1, \dots, w_c) = E[\psi(w_1, \dots, w_c, W_{c+1}, \dots, W_4)]$ , be the projections of  $\psi$  to lower-dimensional sample spaces for  $c = 1, \dots, 4$ , where  $w_1, \dots, w_c$  are fixed variables (e.g.  $\psi_1(w_1) = E(\psi(w_1, W_{i_2}, W_{i_3}, W_{i_4}))$ ,  $\psi_2(w_1, w_2) = E(\psi(w_1, w_2, W_{i_3}, W_{i_4}))$ ,  $\psi_3(w_1, w_2, w_3) = E(\psi(w_1, w_2, w_3, W_{i_4}))$ ). The specific forms have been given in the appendix of Proof of Theorem 1. Let  $v_c = \text{var}[\psi_c]$  be the variance. Let  $\tilde{\psi}_c = \psi_c - \theta(F)$ , then we have the Hoeffding decompositions for  $T_n$  is  $T_n - \theta(F) = \sum_{c=1}^4 \binom{4}{c} \binom{n-4}{4-c}^{-1} V_{nc}$ , where  $V_{nc} = \sum_{\mathcal{A}_c} g_c(w_{i_1}, \dots, w_{i_c})$  and  $g_c(w_1, \dots, w_c) = \tilde{\psi}_c - \sum_{j=1}^{c-1} \sum_{\mathcal{A}_j} g_j(w_{i_1}, \dots, w_{i_j})$  with  $\mathcal{A}_c = \{(i_1, \dots, i_c) : 1 \leq i_1 < \dots < i_c \leq n\}$ . The decomposition for the variance of  $T_n$  is  $\text{var}[T_n] = n4^{-1} \sum_{c=1}^4 \binom{4}{c} \binom{n-4}{4-c} v_c$ . We assume that  $E[\psi^2(W_1, \dots, W_4)]$  exists. The proofs of the Hoeffding decompositions can be found in [23] and also [24]. [10] recently showed that the decomposition also holds when the dimension of the predictor  $K$  increases to infinity. Based

on Proposition 1 in [10], if we find the minimum  $c'$  such that  $v'_c, c' = 1, 2, \text{ or } 3$ , is of the same order as  $v_4$ , then  $T_n$  will be dominated by the first  $c'$  terms, so that

$$T_n - \theta(F) = \sum_{c=1}^{c'} \binom{c'}{c} \binom{n}{c}^{-1} V_{nc} \{1 + o_p(1)\}. \tag{11}$$

**Theorem 1** Under the FLM (1) and assuming Assumption (1),  $K, L \rightarrow \infty$  as  $n \rightarrow \infty$ , we have

- (i).  $E[T_n] = \|C(\beta - \beta_0)\|^2$  and  $\text{var}[T_n] = \{\frac{16}{n} v_1 + \frac{72}{n(n-1)} v_2\} \{1 + o(1)\}$ ;
- (ii).  $T_n - \|C(\beta - \beta_0)\|^2 = \{\frac{16}{n} V_{n1} + \frac{72}{n(n-1)} V_{n2}\} \{1 + o_p(1)\}$ ,

where  $E[V_{n1}^2] = v_1$  and  $E[V_{n2}^2] = v_2 - 2v_1$ .

Please see the Proof of Theorem 1 in Appendix.

Let  $\Delta = (\beta_{\ell k} - \beta_{0,\ell k})_{\ell,k}$ , where  $\beta_{0,\ell k}$  define the loadings of  $\beta_0(t, s)$ . And let  $M_a = \Delta \Sigma^a \Delta'$ ,  $a = 0, 1, 2, 3$  (e.g.  $\Sigma^0 = I_K, \Sigma^2 = \Sigma \Sigma$ ),  $Q_0 = \Gamma' \Gamma, Q_1 = \Gamma' \Delta' \Delta \Gamma, Q_2 = \Gamma' \Sigma \Delta' \Delta \Gamma, Q_3 = \Gamma' \Sigma \Gamma, Q_4 = \Gamma' \Delta' \Delta \Sigma \Delta' \Delta \Gamma$ . Under  $H_0 : \mathcal{B} = \mathcal{B}_0$ , we have  $\Delta = 0$ , and hence  $Q_1 = Q_2 = Q_3 = M_i = 0$  for  $i = 0, 1, 2, 3$ . So it is obvious that  $v_1 = 0$ , and  $T_n$  is then a degenerate U-statistic. Under this case, we have

$$\text{var}[T_n] = \frac{2}{n(n-1)} \text{tr}(\Lambda^2) \text{tr}(\Sigma^2) \{1 + o(1)\}.$$

Next we show that the form of the variance for  $T_n$  also holds under a subclass of local alternative hypothesis  $\mathcal{H}_1$  specified by the following condition,

$$\text{tr}(M_1) = o(1) \quad \text{and} \quad \text{tr}(M_3) = o\{n^{-1} \text{tr}(\Sigma^2)\}. \tag{12}$$

Under the null hypothesis, the equation  $v_1 = o(n^{-1} v_2)$  holds with  $v_1 = 0$ . Under the local hypothesis, the equation  $v_1 = o(n^{-1} v_2)$  still holds (see Appendix). The following theorem then states the asymptotic normality of our test statistic under this local alternative.

**Theorem 2.** Under the FLM (1), assuming Assumptions (1) and (2), under either the null hypothesis  $\mathcal{H}_0$  or the local alternatives  $\mathcal{H}_1$ , as  $n \rightarrow \infty$ , we have

$$\frac{n}{\sqrt{2 \text{tr}(\Lambda^2) \text{tr}(\Sigma^2)}} (T_n - \|C(\beta - \beta_0)\|^2) \xrightarrow{d} N(0, 1).$$

Please also see the Proof of Theorem 2 in Appendix.

For real data, the trace  $\text{tr}(\Sigma^2)$  and  $\text{tr}(\Lambda^2)$  need to be estimated. We use the estimator given in Chen and Qin [26], which was shown to be unbiased and ratio consistent, i.e.  $\widehat{\text{tr}(\Sigma^2)} / \text{tr}(\Sigma^2) \xrightarrow{p} 1$ , under the null hypothesis or the local alternatives. Specifically, the estimator is given as

$$\widehat{\text{tr}(\Sigma^2)} = R_{1n} - 2R_{2n} + R_{3n}, \tag{13}$$

where  $R_{1n} = \frac{1}{K_n^2} \sum_{\mathcal{A}} \langle X_{i_1}, X_{i_2} \rangle^2, R_{2n} = \frac{1}{K_n^3} \sum_{\mathcal{A}} \langle X_{i_1}, X_{i_2} \rangle \langle X_{i_2}, X_{i_3} \rangle$ , and  $R_{3n} = \frac{1}{K_n^4} \sum_{\mathcal{A}} \langle X_{i_1}, X_{i_2} \rangle \times \langle X_{i_3}, X_{i_4} \rangle$  with  $K_n^m = n! / (n - m)!$ . Following the same idea, we can also construct a consistent estimator of  $\text{tr}(\Lambda^2)$  under  $H_0$ .

Following Theorem 2, the FLUTE test rejects  $\mathcal{H}_0$  at significant level  $\alpha$  if

$$nT_n \geq \sqrt{2\text{tr}(\widehat{\Lambda}^2)\text{tr}(\widehat{\Sigma}^2)}z_\alpha,$$

where  $z_\alpha$  is the upper  $\alpha$ -quantile of  $N(0, 1)$ .

Theorem 2 also implies that the asymptotic power of the proposed statistic under the local alternative is

$$\Upsilon(\|\beta - \beta_0\|) = \Phi\left(-z_\alpha + \frac{n\|C(\beta - \beta_0)\|^2}{\sqrt{2\text{tr}(\Lambda^2)\text{tr}(\Sigma^2)}}\right).$$

The quantity  $r_n(\beta - \beta_0, \Sigma, \Lambda) = n\|C(\beta - \beta_0)\|^2 / \sqrt{2\text{tr}(\Lambda^2)\text{tr}(\Sigma^2)}$  can be viewed as a signal to noise ratio (SNR). If  $r_n(\beta - \beta_0, \Sigma, \Lambda) = o(1)$ , it is obvious that the power converges to  $\alpha$ . If  $r_n(\beta - \beta_0, \Sigma, \Lambda)$  is in the order of  $O(1)$ , the power converges to 1.

### FLUTE for scalar responses

In the FLM with a scalar response,

$$Y = \alpha + \int_0^1 X(t)\beta(t)dt + \epsilon, \tag{14}$$

where  $Y \in R$ . The null hypothesis for the scalar response is defined as

$$\mathcal{H}_0 : \beta(t) = \beta_0(t) \quad \text{versus} \quad \mathcal{H}_1 : \beta(t) \neq \beta_0(t).$$

The idea of the FLUTE method in Section Asymptotic theory directly applies and only requires slight modification toward the dimension of the response and functional regression coefficients. For example, the kernel of the FLUTE statistic  $\check{T}_n$  is  $(Y_i - \langle X_i, \beta_0 \rangle)\langle X_i(s), X_j(s) \rangle (Y_j - \langle X_j, \beta_0 \rangle)$  with expectation  $E[\check{T}_n] = \check{\theta}(F) = \|C(\beta - \beta_0)\|^2$ , where  $C(\beta) = \int_0^1 c(t, s)\beta(s)ds$ . The expansion of parameter  $\beta(t)$  is  $\beta(t) = \sum_{k=1}^K \beta_k \phi_k(t)$ . The theory can be developed using the same idea as in Section Asymptotic theory. We distinguish by denoting the counterpart to notations in Section Asymptotic theory with a check mark. For example, the kernel of the FLUTE statistic for scalar response model (14) is denoted by  $\check{\psi}_c$ , and its variance is  $\check{v}_c$ . The following theorems show that the same asymptotic null distribution in Theorems 1 and 2 hold for the scalar response case.

**Theorem 3.** *Under the FLM with scalar response (14), assuming Assumption (1), when  $K \rightarrow \infty$  as  $n \rightarrow \infty$ , we have*

- (i).  $E[\check{T}_n] = \|C(\beta - \beta_0)\|^2$  and  $\text{var}[\check{T}_n] = \{\frac{16}{n}\check{v}_1 + \frac{72}{n(n-1)}\check{v}_2\}\{1 + o(1)\}$ ,
- (ii).  $\check{T}_n - \|C(\beta - \beta_0)\|^2 = \{\frac{16}{n}\check{V}_{n1} + \frac{72}{n(n-1)}\check{V}_{n2}\}\{1 + o_p(1)\}$ .

We consider the local alternative hypothesis as follows.

$$\iint (\beta(t) - \beta_0(t))c(t, s)(\beta(s) - \beta_0(s))dtds = o(1),$$

and

$$\iint (\beta(t) - \beta_0(t))c^3(t, s)(\beta(s) - \beta_0(s))dtds = o\{n^{-1}\text{tr}(\Sigma^2)\},$$

where  $c^3(t, s) = \sum_{i=1}^\infty \lambda_i^3 v_i(t)v_i(s)$ .

**Theorem 4.** *In the FLM with a scalar response ( 14 ), assume Assumption (1) and  $E[\epsilon^4]$  is finite. Under either the null hypothesis  $\mathcal{H}_0$  or the local alternatives  $\mathcal{H}_1$ , as  $n \rightarrow \infty$ , we have*

$$\frac{n}{\sigma^2 \sqrt{2\text{tr}(\Sigma^2)}} (\check{T}_n - \|C(\beta - \beta_0)\|^2) \xrightarrow{d} N(0, 1),$$

where  $\sigma^2 = \text{var}[\epsilon]$ .

The proofs of Theorems 3 and 4 are omitted because they can be proved in the same way as Theorems 1 and 2 except with slight modification to the notations to reflect the difference in dimensionality.

### Simulation and real data

In this section, we demonstrate the performance of the FLUTE method by simulation studies and an application to a real data example. For cases with functional responses, we compare the FLUTE method with the method in [1], which is constructed based on the functional PCA, we call KMSZ, and the nonparametric test in [12] is constructed by a weighted U-statistic and we name as NP. The KMSZ method depends on the functional principal components and is more suitable for large sample case which could estimate the covariance operator well. The test statistic of the NP method depends on so-called the least-favorable direction  $\gamma$  which is more suitable for the low dimension case. Under the simulation setup, this direction  $\gamma$  can be decided in three different ways: 1) Pre-estimate  $\gamma$  based on a super large simulated data set and then use it for all simulated data sets; 2) pre-estimate  $\gamma$  based on the data set generated at each level of  $|\beta|^2$  and then use it for simulated data sets generated at the same level of  $|\beta|^2$ ; and 3) estimate  $\gamma$  based on each simulated data set. The simulation results please see Table 1 and more details can be found in the Supplementary Material A in S1 File. Results reported in this section are based on the second way, which is consistent with applications of the NP method to real data.

For FLMs with a scalar response, neither the KMSZ nor NP method is applicable because the former involves functional PCA on the response and the latter requires computing the  $L^2$  norm between two functional response values. The nonparametric test proposed by [15] which we name as NETRF is for the scalar response. However, the NETRF test still requires estimating the covariance operator to calculate the semetric. Furthermore, this test needs sufficiently large sample data to provide accurate estimations of each group. Therefore, we do not directly compare DelSol’s method with our FLUTE test, but we conduct simulation studies for small/moderate sample cases to demonstrate the incapability of DelSol’s method under

**Table 1.** Size and power for NP test with different searching methods.

$ \beta ^2$	case 1		case 2		case 3	
	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$
0.00	0.059	0.100	0.059	0.100	0.020	0.021
0.02	0.062	0.117	0.082	0.128	0.038	0.039
0.04	0.088	0.140	0.393	0.497	0.042	0.042
0.06	0.090	0.151	0.604	0.708	0.119	0.119
0.08	0.115	0.185	0.755	0.831	0.062	0.062
0.10	0.150	0.227	0.858	0.905	0.185	0.185
0.20	0.246	0.359	0.989	0.994	0.731	0.731
0.30	0.268	0.384	0.998	0.999	0.204	0.204
0.40	0.347	0.467	0.999	0.999	0.922	0.922
0.50	0.359	0.463	0.995	0.998	0.988	0.988

<https://doi.org/10.1371/journal.pone.0234094.t001>

these scenarios. Here we choose the current comparison benchmark as the F-test proposed by [7]. [7] actually proposed four asymptotically equivalent tests which also depends on the functional components, and can be more suitable for large sample case. We use the F-test because it behaves the best of the four tests for small to moderate samples.

### Simulation results

We next conduct a simulation study to evaluate the empirical size and power of our FLUTE test for small to moderate samples with sample size  $n$  varying between 40 and 100. In each simulation, we generate 1,000 Monte Carlo samples. Our computer codes are written in R. For basis expansion and functional PCA, we use the implementation in the R package `fda`.

**Functional response.** First we present the case of the FLM with functional responses. This simulation design follows the FLM (1), where we set  $\beta(t, s) = |\beta|^2 \exp\{(t^2 + s^2)/2\}$ . Here  $|\beta|^2$  indicates the  $L_2$  norm of  $\beta(t, s)$  and is used to control the SNR. We generate the functional predictor  $X_t(s)$  according to Eq (10), where the bases  $\{\phi_k(s)\}_{k=1}^K$  are chosen as Fourier bases. For instance, the first five orthonormal Fourier basis functions are  $\phi_1(s) = 1, \phi_2(s) = \sqrt{2}\sin(2\pi s), \phi_3(s) = \sqrt{2}\cos(2\pi s), \phi_4(s) = \sqrt{2}\sin(4\pi s)$ , and  $\phi_5(s) = \sqrt{2}\cos(4\pi s)$ . Without loss of generality, we set the mean  $\mu_x(s) = 0$ . According to the factor design (Assumption 1), the loadings  $\xi_1, \dots, \xi_n$  are independently generated from the following moving average model,

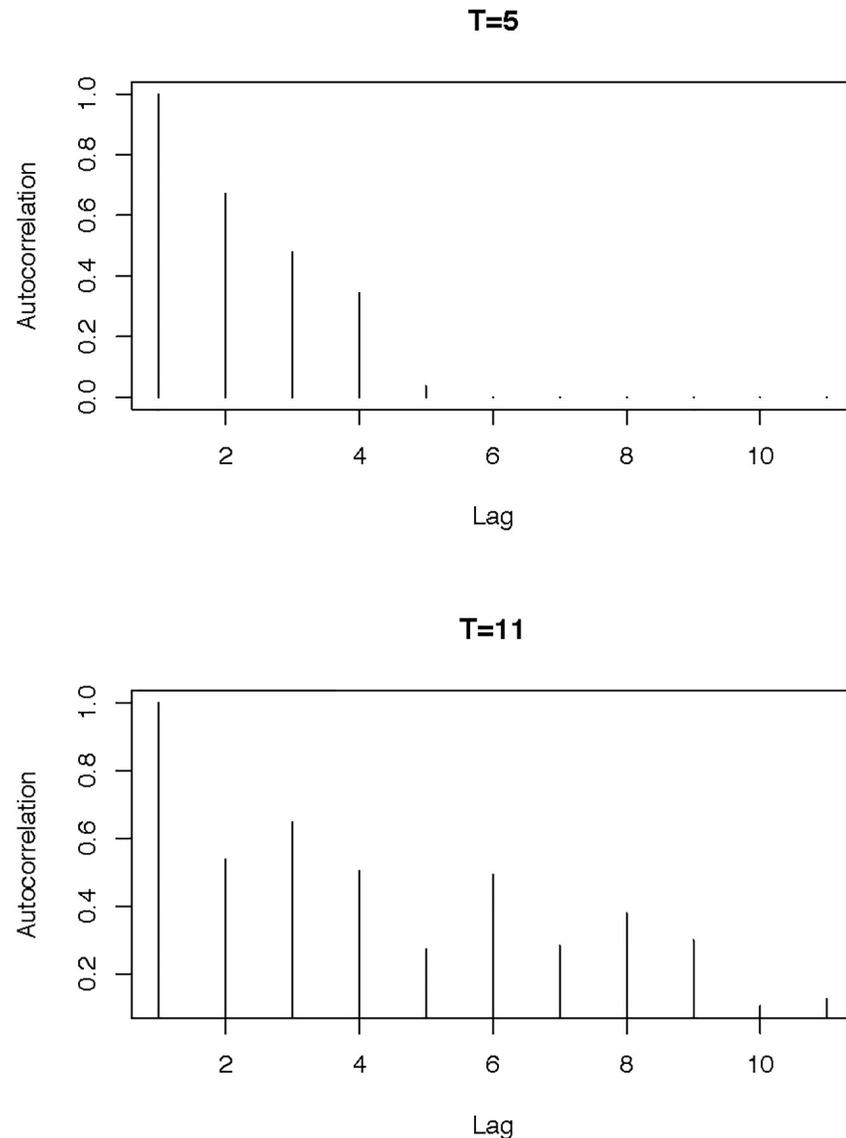
$$\xi_{ik} = \rho_1 N_{ik} + \rho_2 N_{i(k+1)} + \dots + \rho_T N_{i(k+T-1)}, \quad k = 1, \dots, K, \tag{15}$$

where the constant  $T$  controls range of dependency(see Fig 1). The coefficients  $\{\rho_t\}_{t=1}^T$  are randomly generated from  $U(0, 1)$ , where  $U(a, b)$  denotes the Uniform distribution on the interval  $(a, b)$ . And the random vectors  $\mathbf{N}_i = (N_{i1}, \dots, N_{i(K+T-1)})'$  are independently generated from the  $N(0, I_{K+T-1})$  distribution. It then follows that the  $(k, \ell)$ th element of the covariance matrix  $\text{var}(\xi)$  is  $\sum_{t=1}^{T-|k-\ell|} \rho_t \rho_{t+|k-\ell|} I\{|k-\ell| < T\}$ , which shows that the correlation between  $\xi_{ik}$  and  $\xi_{i\ell}$  is controlled by  $|k-\ell|$  and  $T$ . The random error function  $\varepsilon_i(t)$  is generated according to the decomposition in Eq (10). We also set bases  $\{\eta_\ell(t)\}_{\ell=1}^L$  in the same way as  $\{\phi_k(s)\}$ . And the loadings  $\varepsilon_{i1}, \dots, \varepsilon_{iL}$  are independent identical distribution and generate from  $N(0, \Sigma_\varepsilon)$ .

To evaluate the impact of dimensionality and sample size, we carry out simulations under four different settings, varying in dimensionality and sample size  $K = L = 5$  (low-dimensional) and  $K = L = 11$  (high-dimensional),  $n = 40$  and 100. When generating  $X(t)$ , we set  $T = 5$  in Eq (15). The variances of the loadings  $\varepsilon_{i1}, \dots, \varepsilon_{iL}$  are the same, we set  $\Sigma_\varepsilon = I_L$ . Under each setting, we vary the  $|\beta|^2$  at 10 levels from 0 to 0.5 (see Tables 2 and 3). When  $|\beta|^2$  is 0, the result provides the empirical size of all tests, and results at the other 9 levels give the power. Each testing method is evaluated at two nominal significance levels  $\alpha = 0.05$  and 0.1.

Table 2 shows the empirical sizes and power obtained for different dimensionality and sample size under the nominal significance level  $\alpha = 0.05$ . Under the same sample size, the power of all three tests decrease as the dimensions  $K$  and  $L$  increase. When the dimensionality is the same, the power of all three methods improves as the sample size increases. Table 2 also shows that the FLUTE method performs stably in both the low dimensional cases and the high dimensional cases. The KMSZ and NP tests are conservative and their power decreases significantly as the dimension increases.

Further the NP method has almost no power in the case of high dimension and small sample size ( $K = L = 11, n = 40$ ). It is apparent that in Table 2, power of the FLUTE method is consistently higher than that of the KMSZ and NP methods, especially in high dimensional cases. The simulation results also show that the FLUTE method respects the nominal levels under high dimensionality at both sample size  $n = 40$  and 100.



**Fig 1. The autocorrelation functions for loadings  $\{\zeta_{ik}\}_{k=1}^K$ .**

<https://doi.org/10.1371/journal.pone.0234094.g001>

Table 3 shows the results under the nominal significance level  $\alpha = 0.1$ , and provides the same conclusion as Table 2.

To evaluate the impact of the correlation structure, we carry out simulations under two different settings,  $T = 5$  (weakly correlated) and 11 (strongly correlated) (see Fig 1). Under each setting, we vary  $|\beta|^2$  at 6 levels from 0 to 0.5 (see Table 4). When  $T$  is 5, the correlation is weak. Table 4 shows the empirical sizes and power obtained for the case of  $K = L = 11$  and  $n = 40$ . The FLUTE method is stable for different  $T$ , both of the KMSZ and the NP method are more sensitive to the correlation structure. On the other hand, the power of the NP statistic decreases significantly when  $T$  increases, since this method needs to search the least-favorable direction. While the power of the KMSZ statistic decreases significantly when  $T$  reduces, since this method depends on functional PCA. When the correlation is weak, the number of

Table 2. Size and power for different tests at  $\alpha = 0.05$ .

$ \beta ^2$	$K = L = 5, n = 40$			$K = L = 11, n = 40$			$K = L = 5, n = 100$			$K = L = 11, n = 100$		
	NP	KMSZ	FLUTE	NP	KMSZ	FLUTE	NP	KMSZ	FLUTE	NP	KMSZ	FLUTE
0.00	0.043	0.042	<b>0.072</b>	0.043	0.025	<b>0.055</b>	0.048	0.051	<b>0.065</b>	0.035	0.037	<b>0.057</b>
0.02	0.096	0.178	<b>0.482</b>	0.040	0.107	<b>0.445</b>	0.408	0.699	<b>0.848</b>	0.104	0.317	<b>0.620</b>
0.04	0.132	0.361	<b>0.799</b>	0.034	0.174	<b>0.781</b>	0.788	0.973	<b>0.997</b>	0.217	0.762	<b>0.873</b>
0.06	0.198	0.600	<b>0.932</b>	0.046	0.256	<b>0.907</b>	0.939	0.990	<b>1.000</b>	0.335	0.896	<b>0.969</b>
0.08	0.267	0.711	<b>0.973</b>	0.041	0.341	<b>0.968</b>	0.986	0.999	<b>1.000</b>	0.543	0.954	<b>0.998</b>
0.1	0.300	0.801	<b>0.992</b>	0.034	0.380	<b>0.979</b>	0.995	1.000	<b>1.000</b>	0.747	0.996	<b>1.000</b>
0.2	0.514	0.980	<b>1.000</b>	0.033	0.619	<b>0.998</b>	1.000	1.000	<b>1.000</b>	0.853	0.999	<b>1.000</b>
0.3	0.667	1.000	<b>1.000</b>	0.048	0.707	<b>0.999</b>	1.000	1.000	<b>1.000</b>	0.910	1.000	<b>1.000</b>
0.4	0.751	0.998	<b>1.000</b>	0.045	0.745	<b>1.000</b>	1.000	1.000	<b>1.000</b>	0.995	1.000	<b>1.000</b>
0.5	0.789	1.000	<b>1.000</b>	0.040	0.772	<b>1.000</b>	1.000	1.000	<b>1.000</b>	1.000	1.000	<b>1.000</b>

<https://doi.org/10.1371/journal.pone.0234094.t002>

Table 3. Size and power for different tests at  $\alpha = 0.1$ .

$ \beta ^2$	$K = L = 5, n = 40$			$K = L = 11, n = 40$			$K = L = 5, n = 100$			$K = L = 11, n = 100$		
	NP	KMSZ	FLUTE	NP	KMSZ	FLUTE	NP	KMSZ	FLUTE	NP	KMSZ	FLUTE
0.00	0.079	0.093	<b>0.128</b>	0.080	0.068	<b>0.111</b>	0.077	0.085	<b>0.103</b>	0.078	0.079	<b>0.097</b>
0.02	0.165	0.322	<b>0.557</b>	0.084	0.208	<b>0.536</b>	0.520	0.803	<b>0.896</b>	0.126	0.342	<b>0.665</b>
0.04	0.193	0.535	<b>0.851</b>	0.067	0.326	<b>0.836</b>	0.862	0.990	<b>0.999</b>	0.248	0.784	<b>0.889</b>
0.06	0.288	0.735	<b>0.948</b>	0.082	0.460	<b>0.942</b>	0.964	0.996	<b>1.000</b>	0.326	0.899	<b>0.983</b>
0.08	0.353	0.843	<b>0.980</b>	0.073	0.525	<b>0.984</b>	0.991	1.000	<b>1.000</b>	0.533	0.905	<b>0.999</b>
0.1	0.392	0.909	<b>0.996</b>	0.066	0.605	<b>0.988</b>	0.998	1.000	<b>1.000</b>	0.738	0.998	<b>1.000</b>
0.2	0.635	0.997	<b>1.000</b>	0.062	0.805	<b>1.000</b>	1.000	1.000	<b>1.000</b>	0.866	1.000	<b>1.000</b>
0.3	0.768	1.000	<b>1.000</b>	0.085	0.885	<b>0.999</b>	1.000	1.000	<b>1.000</b>	0.930	1.000	<b>1.000</b>
0.4	0.821	1.000	<b>1.000</b>	0.076	0.903	<b>1.000</b>	1.000	1.000	<b>1.000</b>	0.997	1.000	<b>1.000</b>
0.5	0.860	1.000	<b>1.000</b>	0.074	0.911	<b>1.000</b>	1.000	1.000	<b>1.000</b>	1.000	1.000	<b>1.000</b>

<https://doi.org/10.1371/journal.pone.0234094.t003>

functional PCs would increase to achieve the same percentage of variance explanation, hence the number of  $p$  also increase which results in lower power.

To evaluate the performance of the FLUTE method with heteroscedastic variance, we carry out simulations under the following settings, the designed variances of the expansion coefficients of  $\varepsilon_i(t)$  are  $\text{Var}(\varepsilon_{i\ell}) = 1/\ell$ , for  $i = 1, \dots, n, \ell = 1, \dots, L$ . We set  $T = 5, n = 40$ . And we vary  $|\beta|^2$  at 6 levels from 0 to 0.5. The significance levels are  $\alpha = 0.05$  and  $\alpha = 0.1$  respectively. Table 5 shows the empirical sizes and power for the cases of  $K = L = 5$  and  $K = L = 11$ , and

Table 4. Size and power for different correlation when  $K = L = 11, n = 40$ .

$ \beta ^2$	$T = 5, \alpha = 0.05$			$T = 11, \alpha = 0.05$			$T = 5, \alpha = 0.1$			$T = 11, \alpha = 0.1$		
	NP	KMSZ	FLUTE	NP	KMSZ	FLUTE	NP	KMSZ	FLUTE	NP	KMSZ	FLUTE
0.00	0.050	0.041	<b>0.064</b>	0.043	0.034	<b>0.066</b>	0.086	0.082	<b>0.098</b>	0.086	0.100	<b>0.107</b>
0.10	0.580	0.452	<b>0.940</b>	0.238	0.849	<b>1.000</b>	0.675	0.652	<b>0.959</b>	0.348	0.954	<b>1.000</b>
0.20	0.794	0.740	<b>0.999</b>	0.340	0.946	<b>1.000</b>	0.824	0.894	<b>0.999</b>	0.443	0.991	<b>1.000</b>
0.30	0.860	0.863	<b>1.000</b>	0.404	0.971	<b>1.000</b>	0.884	0.959	<b>1.000</b>	0.523	0.998	<b>1.000</b>
0.40	0.883	0.883	<b>1.000</b>	0.433	0.983	<b>1.000</b>	0.903	0.975	<b>1.000</b>	0.537	0.998	<b>1.000</b>
0.50	0.900	0.929	<b>1.000</b>	0.443	0.982	<b>1.000</b>	0.911	0.986	<b>1.000</b>	0.562	0.998	<b>1.000</b>

<https://doi.org/10.1371/journal.pone.0234094.t004>

Table 5. Size and power for heteroscedastic variance when  $K = L = 11, n = 40$ .

$ \beta ^2$	$K = L = 5, \alpha = 0.05$			$K = L = 5, \alpha = 0.1$			$K = L = 11, \alpha = 0.05$			$K = L = 11, \alpha = 0.1$		
	NP	KMSZ	FLUTE	NP	KMSZ	FLUTE	NP	KMSZ	FLUTE	NP	KMSZ	FLUTE
0.00	0.055	0.042	<b>0.061</b>	0.088	0.102	<b>0.095</b>	0.057	0.033	<b>0.055</b>	0.092	0.092	<b>0.087</b>
0.10	0.972	1.000	<b>1.000</b>	0.981	1.000	<b>1.000</b>	0.116	0.404	<b>0.771</b>	0.188	0.628	<b>0.873</b>
0.20	0.998	1.000	<b>1.000</b>	0.999	1.000	<b>1.000</b>	0.176	0.647	<b>0.972</b>	0.286	0.813	<b>0.986</b>
0.30	0.999	1.000	<b>1.000</b>	1.000	1.000	<b>1.000</b>	0.308	0.793	<b>0.993</b>	0.427	0.918	<b>0.997</b>
0.40	1.000	1.000	<b>1.000</b>	1.000	1.000	<b>1.000</b>	0.408	0.808	<b>0.997</b>	0.518	0.929	<b>0.998</b>
0.50	0.998	1.000	<b>1.000</b>	0.999	1.000	<b>1.000</b>	0.474	0.829	<b>1.000</b>	0.589	0.951	<b>1.000</b>

<https://doi.org/10.1371/journal.pone.0234094.t005>

provides a similar conclusion as Tables 2 and 3 when  $n = 40$ . The power of all three tests decreases as the dimensions  $K$  and  $L$  increase. However, the FLUTE method performs stably in both low dimensional cases and the high dimensional cases, the power of the KMSZ and NP tests decreases significantly as the dimension increases.

Fig 2 shows the histograms of the FLUTE statistic for different dimensionality and sample size under the null hypothesis, which matches nicely with the imposed standard normal density. This is consistent with our results in Theorem 2.

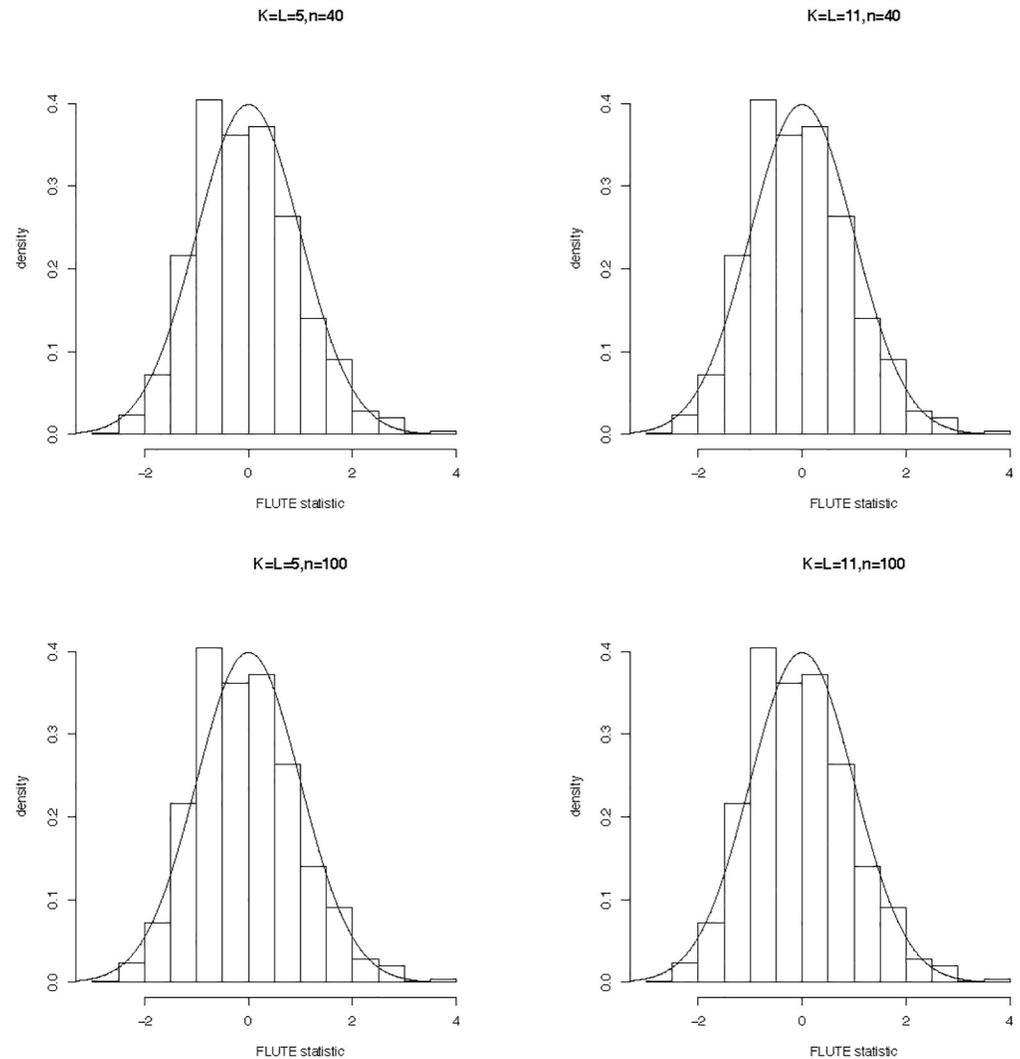
Fig 3 shows the power curves of the FLUTE statistic under four different cases with varying dimensionality and sample size when the nominal significance level  $\alpha = 0.05$  and level  $\alpha = 0.1$ . Under all the four cases, power curves have effective size, and when  $|\beta|^2$  is 0.2, the four power curves almost reached 1.

**Scalar response.** This section presents the results for FLMs with scalar responses. This simulation design follows the model (14). We set the coefficient of regression parameter as  $\beta_k = \frac{1}{\sqrt{k}}|\beta|^2, k = 1, 2, \dots, K$ . The functional predictor  $X(t)$  is generated in the same way as in Section Functional response. And the random errors  $\varepsilon_i$  are independently generated from  $N(0, 1)$ .

Same with FLMs with functional responses in Section Functional response, we carry out simulations under four different settings,  $K = 5$  (low-dimensional) and  $K = 11$  (high-dimensional),  $n = 40$  and 100. Under each setting, we vary  $|\beta|^2$  at 10 levels from 0 to 0.5 (see Tables 5 and 6). Each testing method is evaluated at two nominal significance levels  $\alpha = 0.05$  and 0.1.

Table 6 shows the empirical sizes and powers obtained for different dimensionality and sample size under the nominal significance level  $\alpha = 0.05$ . The power of the two tests, FLUTE and the F-test in [7], is similar in these four cases. The results show that the FLUTE test is more powerful than the F-test. Table 7 shows the same conclusions at nominal significance level  $\alpha = 0.1$ .

Table 8 shows the comparison between FLUTE and Delsol’s method at significant level  $\alpha = 0.1$ , and  $K = 11$ . NETRF1, NETRF2 and NETRF3 stand for three bootstrap methods in Delsol’s paper. Due to the three test statistics are nonparametric tests that are constructed based on a kernel function, the estimation of bias and variance terms seems difficult. Further, it is usually irrelevant to use the quantiles of the asymptotic law to estimate the threshold directly. Thus, the bootstrap procedure is needed. For all three methods, we choose the semi-metric induced by functional principal components, and split the samples into three groups as 20, 10 and 10, when  $n = 40, 40, 30$  and 30, when  $n = 100$ . Under each setting, the empirical significance level are calculated by 1000 bootstrap iterations. FLUTE stands for our method. It is obviously that the sizes of Delsol’s methods can not be well controlled at the nominal level for small/moderate samples.



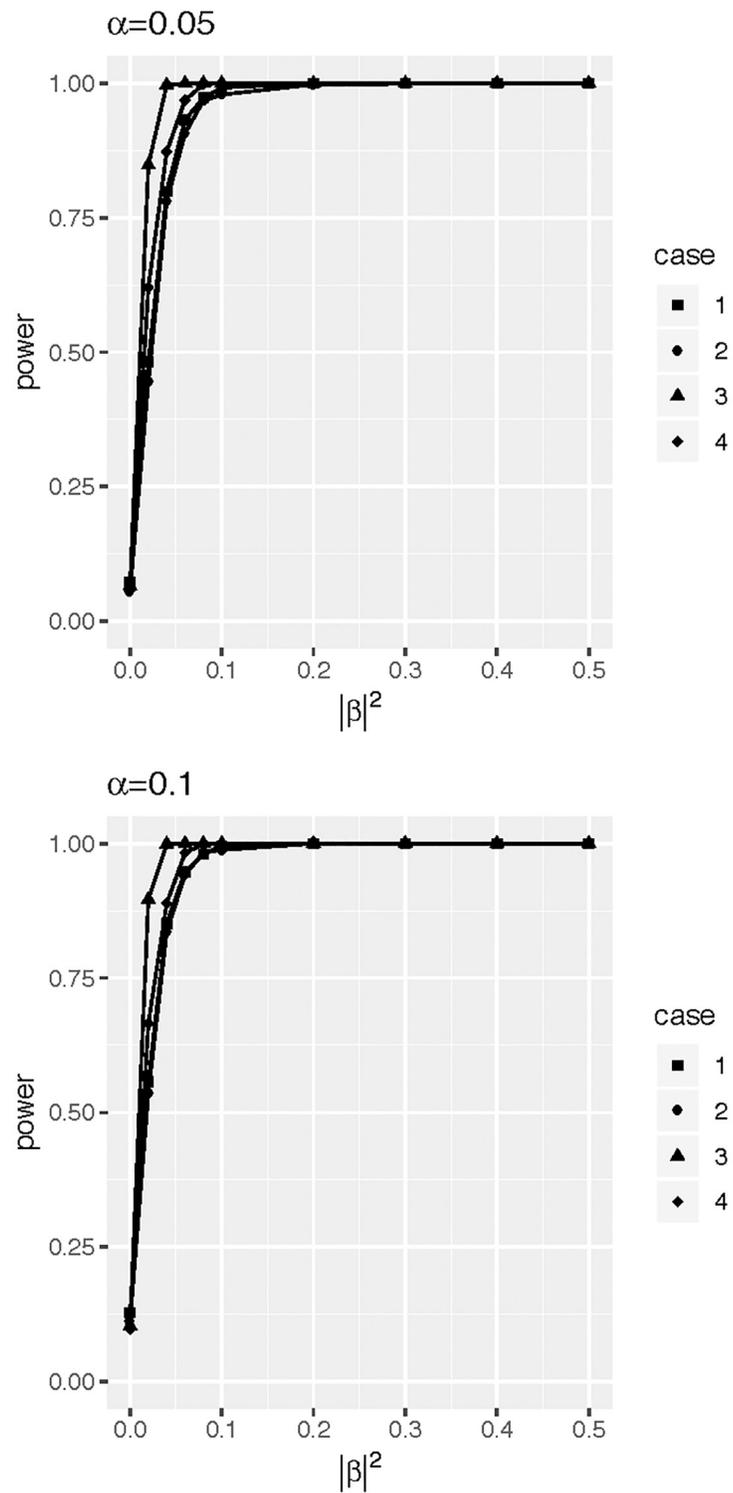
**Fig 2. The null distribution of the FLUTE statistic in FLMs with functional responses.** The solid line indicates the density of the standard normal distribution.

<https://doi.org/10.1371/journal.pone.0234094.g002>

### Application to Canadian Weather data

The Canadian Weather data is available from the R package *fda* (<http://www.r-project.org>) which named *CanadianWeather*. The data consists of the daily temperature and rainfall registered in 35 weather stations in Canada averaged over 1960 to 1994, hence the sample size is 35. We view the daily temperature as the predictor and the rainfall as the response variable. Both the predictor and the response variable are functional. We use the FLUTE test to check the dependency between the daily temperature and the rainfall. Following [3], we choose 11 Fourier bases to fit the temperature curve and rainfall curve for each station separately.

Let  $Y_i(t)$  represent the logarithm of the rainfall at the station  $i$  at time  $t$  and  $x_i(t)$  be the temperature of the same station at time  $t$  of the year. The value of FLUTE statistic is 12.17159 based on the whole 35 stations, hence we reject the null hypothesis. To illustrate the efficacy of the test, we repeat the test on 1000 bootstrap samples. Each bootstrap sample consists of data at 35 randomly selected stations with replacement from the total 35 stations. Fig 4 shows that



**Fig 3. Power curves of the FLUTE method.** Case 1:  $K = L = 5$  and  $n = 40$ ; Case 2:  $K = L = 11$  and  $n = 40$ ; Case 3:  $K = L = 5$  and  $n = 100$ ; Case 4:  $K = L = 11$  and  $n = 100$ . The left figure is for  $\alpha = 0.05$ , and the right is for  $\alpha = 0.1$ .

<https://doi.org/10.1371/journal.pone.0234094.g003>

Table 6. Size and power for normal residual at significant level  $\alpha = 0.05$ .

$ \beta ^2$	$K = 5, n = 40$		$K = 11, n = 40$		$K = 5, n = 100$		$K = 11, n = 100$	
	F-test	FLUTE	F-test	FLUTE	F-test	FLUTE	F-test	FLUTE
0.00	0.046	<b>0.047</b>	0.044	<b>0.040</b>	0.048	<b>0.055</b>	0.043	<b>0.057</b>
0.02	0.271	<b>0.379</b>	0.267	<b>0.403</b>	0.603	<b>0.771</b>	0.584	<b>0.791</b>
0.04	0.480	<b>0.664</b>	0.452	<b>0.680</b>	0.934	<b>0.968</b>	0.924	<b>0.973</b>
0.06	0.676	<b>0.822</b>	0.671	<b>0.820</b>	0.979	<b>0.993</b>	0.982	<b>0.991</b>
0.08	0.778	<b>0.882</b>	0.760	<b>0.891</b>	0.995	<b>0.999</b>	0.997	<b>0.999</b>
0.10	0.848	<b>0.931</b>	0.881	<b>0.957</b>	0.998	<b>1.000</b>	1.000	<b>1.000</b>
0.20	0.980	<b>0.995</b>	0.989	<b>0.995</b>	1.000	<b>1.000</b>	1.000	<b>1.000</b>
0.30	0.999	<b>1.000</b>	1.000	<b>1.000</b>	1.000	<b>1.000</b>	1.000	<b>1.000</b>
0.40	0.999	<b>1.000</b>	1.000	<b>1.000</b>	1.000	<b>1.000</b>	1.000	<b>1.000</b>
0.50	1.000	<b>1.000</b>	1.000	<b>1.000</b>	1.000	<b>1.000</b>	1.000	<b>1.000</b>

<https://doi.org/10.1371/journal.pone.0234094.t006>

Table 7. Size and power for normal residual at significant level  $\alpha = 0.1$ .

$ \beta ^2$	$K = 5, n = 40$		$K = 11, n = 40$		$K = 5, n = 100$		$K = 11, n = 100$	
	F-test	FLUTE	F-test	FLUTE	F-test	FLUTE	F-test	FLUTE
0.00	0.097	<b>0.098</b>	0.098	<b>0.104</b>	0.104	<b>0.104</b>	0.091	<b>0.092</b>
0.02	0.375	<b>0.449</b>	0.381	<b>0.472</b>	0.718	<b>0.818</b>	0.728	<b>0.829</b>
0.04	0.632	<b>0.712</b>	0.584	<b>0.739</b>	0.962	<b>0.975</b>	0.953	<b>0.982</b>
0.06	0.782	<b>0.857</b>	0.769	<b>0.858</b>	0.990	<b>0.997</b>	0.953	<b>0.982</b>
0.08	0.854	<b>0.912</b>	0.851	<b>0.914</b>	0.998	<b>1.000</b>	0.990	<b>0.995</b>
0.10	0.905	<b>0.947</b>	0.948	<b>0.974</b>	1.000	<b>1.000</b>	0.999	<b>0.999</b>
0.20	0.992	<b>0.995</b>	0.997	<b>0.996</b>	1.000	<b>1.000</b>	1.000	<b>1.000</b>
0.30	0.999	<b>1.000</b>	1.000	<b>1.000</b>	1.000	<b>1.000</b>	1.000	<b>1.000</b>
0.40	0.999	<b>1.000</b>	1.000	<b>1.000</b>	1.000	<b>1.000</b>	1.000	<b>1.000</b>
0.50	1.000	<b>1.000</b>	1.000	<b>1.000</b>	1.000	<b>1.000</b>	1.000	<b>1.000</b>

<https://doi.org/10.1371/journal.pone.0234094.t007>

Table 8. Size for normal residual at significant level  $\alpha = 0.1$ .

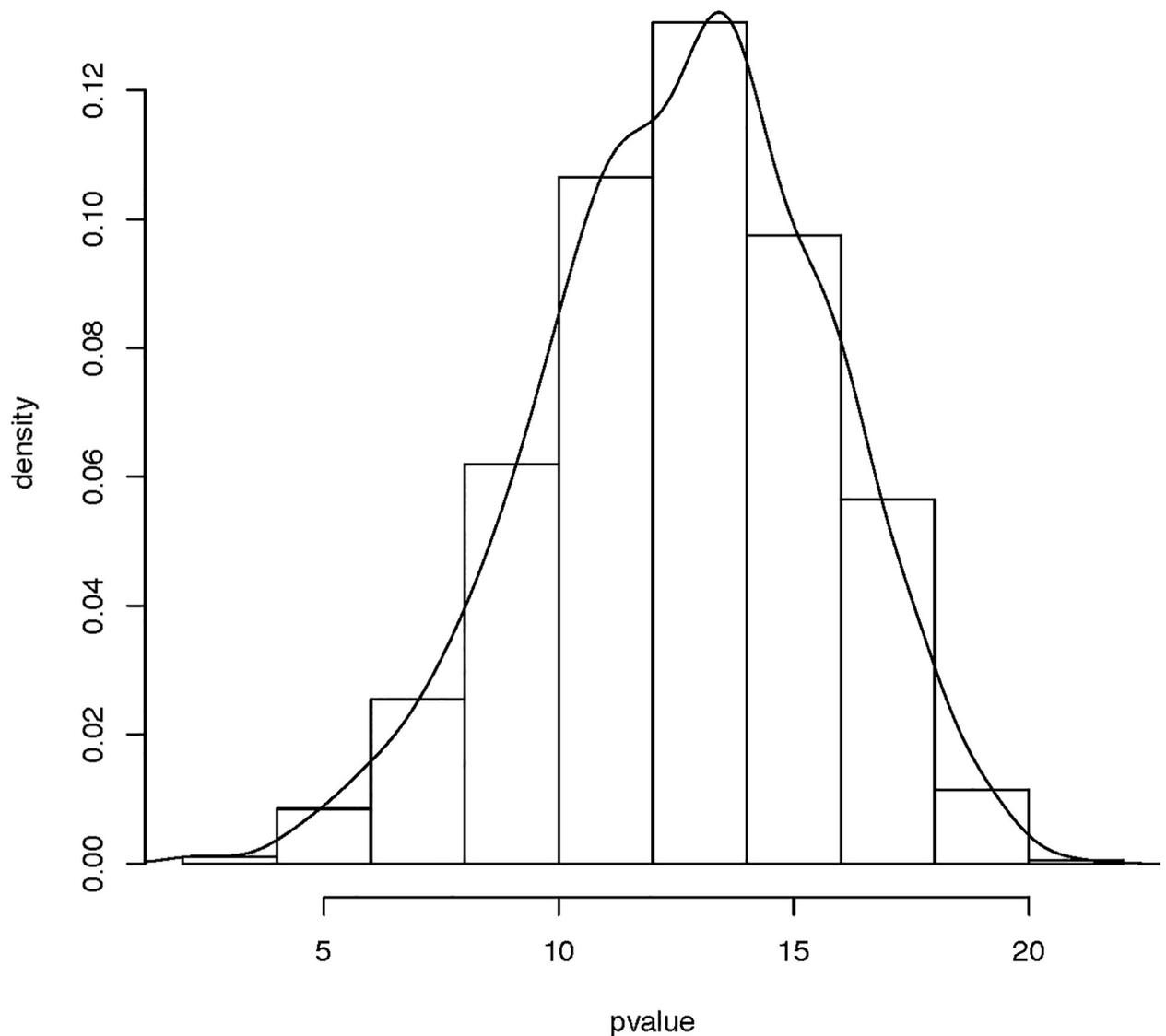
$Y = \varepsilon$	NETRF1	NETRF2	NETRF3	FLUTE
n = 40	0.138	0.142	0.146	0.099
n = 100	0.121	0.127	0.123	0.096

<https://doi.org/10.1371/journal.pone.0234094.t008>

the density of the FLUTE statistic is far away from the standard normal distribution, hence we prefer to reject the null hypothesis.

### Conclusion

We proposed the FLUTE test for testing dependence between the response and functional predictor in FLMs with either a real or functional response. By constructing a U-statistic that measures the  $L_2$  distance in an induced space, the FLUTE statistic avoids estimating the covariance operator of the predictor. The parametric test in [1] requires estimation of the covariance



**Fig 4. Empirical distribution of the FLUTE statistic based on 1000 bootstrap samples of size 35 drawn from the Canadian Weather dataset.**

<https://doi.org/10.1371/journal.pone.0234094.g004>

operator and demands large samples. The nonparametric test in [12], although avoids explicitly estimating the covariance operator, requires estimating the least-favorable direction  $\gamma$ . In general, using the least-favorable direction leads to lower power. Meanwhile, our experience suggests the estimation of  $\gamma$  can be numerically unstable across different simulated data sets, which results in poor test performance.

Our FLUTE test does not suffer from these problems. It requires minimum effort in estimating model parameters, hence achieves higher power, especially for high dimensional cases. One potential weakness of the FLUTE test is its high computational cost in evaluating a U-statistic in large samples. However, estimating covariance operator is less a concern in large samples, one can switch to using parametric methods. We recommend the best context of using the FLUTE test is small to moderate sample problems.

### Appendix

#### Proof of Theorems.

**Lemma 1.** *Suppose the functional predictors  $\{X_i, i = 1, \dots, n\}$  and the regression function  $\beta(t, s)$  satisfy the following two conditions,*

- (A). *Functional predictors,  $\{X_i, i = 1, \dots, n\}$ , belongs to a Sobolev ellipsoid of order two: there exists a universal constant  $C$ , such that  $\sum_{k=1}^{\infty} \xi_{ik}^2 k^4 \leq C^2$  for all  $i = 1, \dots, n$ .*
- (B). *The regression functions satisfy  $\iint \beta^2(t, s) dt ds \leq \tilde{C}$  with some constant  $\tilde{C}$ . Further as  $L \rightarrow \infty$ , the summation of coefficients  $\sum_{\ell=L+1}^{\infty} \beta_{k,\ell} \leq 1/L^4$  for  $k = 1, 2, \dots$*

*then we have the approximation error  $\langle e_i, e_i \rangle \leq 6\tilde{C}C/K^4 + 3C^2/L^4$ .*

*Proof.* Recall that

$$\begin{aligned} e_i(t) &= \int_0^1 \beta(t, s)x_i(s)ds - \sum_{k=1}^K \sum_{\ell=1}^L \xi_{ik} \beta_{\ell k} \eta_{\ell}(t) \\ &= \sum_{\ell=1}^L \sum_{k=K+1}^{\infty} \xi_k \beta_{\ell k} \eta_{\ell}(t) + \sum_{\ell=L+1}^{\infty} \sum_{k=1}^K \xi_k \beta_{\ell k} \eta_{\ell}(t) + \sum_{\ell=L+1}^{\infty} \sum_{k=K+1}^{\infty} \xi_k \beta_{\ell k} \eta_{\ell}(t) \\ &\stackrel{def}{=} A_1 + A_2 + A_3. \end{aligned}$$

Then by the Cauchy-Schwarz inequality, we have

$$\langle e_i, e_i \rangle = \langle A_1 + A_2 + A_3, A_1 + A_2 + A_3 \rangle \leq 3(\langle A_1, A_1 \rangle + \langle A_2, A_2 \rangle + \langle A_3, A_3 \rangle).$$

Next we show the three parts are controlled separately. According to the Holder inequality and Condition (A), we have

$$\begin{aligned} \langle A_1, A_1 \rangle &= \left\langle \sum_{\ell=1}^L \sum_{k=K+1}^{\infty} \xi_{ik} \beta_{\ell k} \eta_{\ell}(t), \sum_{\ell=1}^L \sum_{k=K+1}^{\infty} \xi_{ik} \beta_{\ell k} \eta_{\ell}(t) \right\rangle \\ &= \sum_{k,k'=K+1}^{\infty} \sum_{\ell=1}^L \xi_{ik} \xi_{ik'} \beta_{\ell k} \beta_{\ell k'} \\ &\leq \sum_{k,k'=K+1}^{\infty} \xi_{ik} \xi_{ik'} \left( \sum_{\ell=1}^L \beta_{\ell k}^2 \sum_{\ell'=1}^L \beta_{\ell' k'}^2 \right)^{\frac{1}{2}} \\ &= \left( \sum_{k=K+1}^{\infty} \xi_{ik} \left( \sum_{\ell=1}^L \beta_{\ell k}^2 \right)^{\frac{1}{2}} \right)^2 \\ &\leq \sum_{k=K+1}^{\infty} \xi_{ik}^2 k^4 \sum_{k=K+1}^{\infty} \sum_{\ell=1}^L \beta_{\ell k}^2 k^{-4} \\ &\leq \frac{C^2 \tilde{C}}{K^4}. \end{aligned} \tag{16}$$

And we have

$$\begin{aligned}
 \langle A_2, A_2 \rangle &= \left\langle \sum_{\ell=L+1}^{\infty} \sum_{k=1}^K \xi_{ik} \beta_{\ell k} \eta_{\ell}(t), \sum_{\ell=L+1}^{\infty} \sum_{k=1}^K \xi_{ik} \beta_{\ell k} \eta_{\ell}(t) \right\rangle \\
 &= \sum_{k,k'=1}^K \sum_{\ell=L+1}^{\infty} \xi_{ik} \xi_{ik'} \beta_{\ell k} \beta_{\ell k'} \\
 &\leq \sum_{k,k'=1}^K \xi_{ik} \xi_{ik'} \left( \sum_{\ell=L+1}^{\infty} \beta_{\ell k}^2 \sum_{\ell'=L+1}^{\infty} \beta_{\ell' k'}^2 \right)^{\frac{1}{2}} \\
 &= \left( \sum_{k=1}^K \xi_{ik} \left( \sum_{\ell=L+1}^{\infty} \beta_{\ell k}^2 \right)^{\frac{1}{2}} \right)^2 \\
 &\leq \frac{1}{L^4} \sum_{k=1}^K \xi_{ik}^2 k^4 \sum_{k=1}^K k^{-4} \\
 &\leq \frac{C^2}{L^4}.
 \end{aligned} \tag{17}$$

Similar with the proof of Eqs (16) and (17), we get

$$\begin{aligned}
 \langle A_3, A_3 \rangle &= \left\langle \sum_{\ell=L+1}^{\infty} \sum_{k=K+1}^{\infty} \xi_{ik} \beta_{\ell k} \eta_{\ell}(t), \sum_{\ell=L+1}^{\infty} \sum_{k=K+1}^{\infty} \xi_{ik} \beta_{\ell k} \eta_{\ell}(t) \right\rangle \\
 &\leq \frac{C^2 \tilde{C}}{K^4}.
 \end{aligned} \tag{18}$$

Hence we complete the proof by combining the bounds on each of the three parts.

Next, to prove Theorems 1 and 2, we first introduce some lemmas.

**Lemma 2.** Suppose random vector  $Z_i = (z_{i1}, \dots, z_{ip})' \in \mathbb{R}^p, i = 1, 2$ , satisfy  $E[Z_i] = 0, \text{var}[Z_i] = I_p, E[Z_{ik}^4] = 3 + \rho, k = 1, \dots, p$ , where  $\rho$  is a constant in  $(0, 1)$ . If the two random variables  $Z_1$  and  $Z_2$  are independent, for any square matrix  $M = (m_{k\ell})_{p \times p}$ , we have

- (1).  $E[Z_1 Z_1' M Z_1 Z_1'] = M + M' + \text{tr}(M)I_p + \rho \text{diag}(M)$ ;
- (2).  $E[Z_1 Z_2' M Z_2 Z_1'] = \text{tr}(M)I_p$ ;
- (3).  $E[Z_1 Z_2' M Z_1 Z_2'] = M'$ .

*Proof.*

- (1). Let  $W_1 = Z_1 Z_1' M Z_1 Z_1'$ , where  $W_1(k, \ell)$  indicates the  $(k, \ell)$ th element of  $W_1$ . With direct computation, we have  $W_1(k, \ell) = Z_{1k} Z_{1\ell} \sum_{i,j} m_{ij} Z_{1i} Z_{1j}$ . If  $k = \ell$ ,

$$E[W_1(k, \ell)] = m_{kk} E[Z_{1k}^4] + \sum_{i \neq k} m_{ii} E[Z_k^2] E[Z_i^2] = \sum_i m_{ii} + (2 + \rho)m_{kk};$$

If  $k \neq \ell, E[W_1(k, \ell)] = m_{k\ell} + m_{\ell k}$ . Then  $E[W_1] = M + M' + \text{tr}(M)I_p + \rho \text{diag}(M)$ .

- (2). Since  $E[Z_1 Z_2' M Z_2 Z_1'] = E[Z_1 Z_1'] E[Z_2' M Z_2]$ , and  $E[Z_2' M Z_2] = \sum_i m_{ii}$ , then we have  $E[Z_1 Z_2' M Z_2 Z_1'] = \text{tr}(M)I_p$ .
- (3). It's simple to show that  $E[Z_1 Z_2' M Z_1 Z_2'] = E[Z_1 Z_1'] M' E[Z_2 Z_2'] = M'$ .

**Lemma 3.** Consider symmetrical and semi-positive definite matrices  $A$  and  $B$ , [27] has improved some inequalities:

- (1).  $tr(AB)^2 \leq tr(A^2)tr(B^2)$ ;
- (2).  $tr^2(AB) \leq tr(A^2)tr(B^2)$ .

**Lemma 4.** For matrices  $M_a$ ,  $a = 1, 2, 3$  defined as  $M_a = \Delta \Sigma^a \Delta'$ , we have  $tr^2(M_2) \leq tr(M_1)tr(M_3)$ .  
*Proof.*

$$tr(M_2) = tr(\Delta \Sigma^{\frac{1}{2}} \cdot \Sigma^{\frac{3}{2}} \Delta') \leq \sqrt{tr(\Delta \Sigma^{\frac{1}{2}} \Sigma^{\frac{1}{2}} \Delta') tr(\Delta \Sigma^{\frac{3}{2}} \Sigma^{\frac{3}{2}} \Delta')} = \sqrt{tr(M_1)tr(M_3)}.$$

**Proof of Theorem 1.**

Recall that the definition of the statistic in Eq (9), it is straightforward to show that  $E[T_n] = \|C(\beta - \beta_0)\|^2$ .

To find the dominating terms, we need to calculate the following projections,

$$\psi_1(w_1) = \frac{1}{2} \left\{ tr(\Delta(\xi_1 \xi_1' + \Sigma)\Sigma \Delta') + tr(\epsilon_1 \xi_1' \Sigma \Delta') \right\},$$

$$\begin{aligned} \psi_2(w_1, w_2) &= \frac{1}{6} \left\{ tr(\Delta(\xi_1 - \xi_2)(\xi_1 - \xi_2)' \Sigma \Delta') + tr((\epsilon_1 - \epsilon_2)(\xi_1 - \xi_2)' \Sigma \Delta') \right. \\ &+ tr(\Delta(\xi_1 \xi_1' + \Sigma)(\xi_2 \xi_2' + \Sigma) \Delta') + tr(\epsilon_1 \xi_1' (\xi_2 \xi_2' + \Sigma) \Delta') \\ &\left. + tr(\Delta(\xi_1 \xi_1' + \Sigma) \xi_2 \epsilon_2') + tr(\epsilon_1 \xi_1' \xi_2 \epsilon_2') \right\}, \end{aligned}$$

$$\begin{aligned} \psi_3(w_1, w_2, w_3) &= \frac{1}{12} \left\{ tr(\Delta(\xi_1 - \xi_2)(\xi_1 - \xi_2)' (\xi_3 \xi_3' + \Sigma) \Delta') \right. \\ &+ tr((\epsilon_1 - \epsilon_2)(\xi_1 - \xi_2)' (\xi_3 \xi_3' + \Sigma) \Delta') + tr(\Delta(\xi_1 - \xi_2)(\xi_1 - \xi_2)' \xi_3 \epsilon_3') \\ &+ tr((\epsilon_1 - \epsilon_2)(\xi_1 - \xi_2)' \xi_3 \epsilon_3') + tr(\Delta(\xi_1 - \xi_3)(\xi_1 - \xi_3)' (\xi_2 \xi_2' + \Sigma) \Delta') \\ &+ tr((\epsilon_1 - \epsilon_3)(\xi_1 - \xi_3)' (\xi_2 \xi_2' + \Sigma) \Delta') + tr(\Delta(\xi_1 - \xi_3)(\xi_1 - \xi_3)' \xi_2 \epsilon_2') \\ &+ tr((\epsilon_1 - \epsilon_3)(\xi_1 - \xi_3)' \xi_2 \epsilon_2') + tr(\Delta(\xi_1 \xi_1' + \Sigma)(\xi_2 - \xi_3)(\xi_2 - \xi_3)' \Delta') \\ &+ tr(\epsilon_1 \xi_1' (\xi_2 - \xi_3)(\xi_2 - \xi_3)' \Delta') + tr(\Delta(\xi_1 \xi_1' + \Sigma)(\xi_2 - \xi_3)(\epsilon_2 - \epsilon_3)') \\ &\left. + tr(\epsilon_1 \xi_1' (\xi_2 - \xi_3)(\epsilon_2 - \epsilon_3)') \right\}. \end{aligned}$$

Based on the expansion of  $X_i(t)$  and the orthogonality of the bases, we can derive the variance of the projections  $v_c$ .

With straightforward calculations, we get

$$\begin{aligned}
 v_1 &= \frac{1}{4} \left\{ \text{tr}(M_1 M_3) + \text{tr}(M_2^2) + \rho \text{tr}(Q_2 \circ Q_2) + \text{tr}(\Lambda M_3) \right\}, \\
 v_2 &= \frac{1}{36} \left\{ 15 \text{tr}(M_2^2) + 28 \text{tr}^2(M_2) + 17 \text{tr}(M_1 M_3) + 18 \text{tr}(\Lambda M_3) + \text{tr}(M_1^2) \text{tr}(\Sigma^2) \right. \\
 &\quad \left. + 2 \text{tr}(\Lambda M_1) \text{tr}(\Sigma^2) + 15 \rho \text{tr}(Q_2 \circ Q_2) + \rho^2 \text{tr}(Q_0 \circ Q_1)^2 \right. \\
 &\quad \left. + 2 \rho \text{tr}(\Gamma' \Delta' \Lambda \Delta \Gamma \circ Q_3) + 2 \rho \text{tr}(Q_3 \circ Q_4) + \text{tr}(\Lambda^2) \text{tr}(\Sigma^2) \right\} - \text{tr}^2(M_2), \\
 v_4 &= \frac{1}{48} \left\{ (62 + 12\rho) \text{tr}(M_2^2) + (70 + 8\rho) \text{tr}^2(M_2) + 130 \text{tr}(M_1 M_3) \right. \\
 &\quad \left. + (46\rho + 2) \text{tr}(Q_2 \circ Q_2) + (34 + 4\rho) \text{tr}(M_1^2) \text{tr}(\Sigma^2) + 14 \rho \text{tr}(Q_3 \circ Q_4) \right. \\
 &\quad \left. + 128 \text{tr}(\Lambda M_3) + 48 \text{tr}(\Lambda M_1) \text{tr}(\Sigma^2) + 24 \rho \text{tr}(\Gamma' \Delta' \Lambda \Delta \Gamma \circ Q_3) \right. \\
 &\quad \left. + 8 \rho^2 \text{tr}(Q_0 \circ Q_1)^2 + 24 \text{tr}(\Lambda^2) \text{tr}(\Sigma^2) \right\} - \text{tr}^2(M_2).
 \end{aligned}$$

Here, the Hadamard product is defined as  $A \circ B = (a_{ij} b_{ij})$  for matrices  $A = (a_{ij})$  and  $B = (b_{ij})$ . Since both variances  $v_2$  and  $v_4$  are linear combinations of  $\text{tr}(M_2^2)$ ,  $\text{tr}^2(M_2)$ ,  $\text{tr}(M_1 M_3)$ ,  $\text{tr}(\Lambda M_3)$ ,  $\text{tr}(M_1^2)$ ,  $\text{tr}(\Lambda M_1) \text{tr}(\Sigma^2)$ ,  $\text{tr}(Q_2 \circ Q_2)$ ,  $\text{tr}(Q_0 \circ Q_1)^2$ ,  $\text{tr}(\Gamma' \Delta' \Lambda \Delta \Gamma \circ Q_3)$ ,  $\text{tr}(Q_3 \circ Q_4)$ , and  $\text{tr}(\Sigma^2)$ , they are of the same asymptotic order. This means that the statistic  $T_n$  is dominated by the first two terms corresponding to  $V_{n1}$  and  $V_{n2}$ . Hence we can get the Hoeffding decomposition (11) of  $T_n$ ,  $T_n - \theta(F) = \sum_{c=1}^2 \binom{4}{c} \binom{n}{c}^{-1} \{V_{nc} + o_p(1)\}$ , and  $\text{var}[T_n] = \left\{ \frac{16}{n} v_1 + \frac{72}{n(n-1)} v_2 \right\} \times \{1 + o(1)\}$ . Then we complete the proof.

**Proof of Theorem 2.**

Using the inequalities in Lemma 3, under either the null hypothesis or the local alternative, we have

$$\begin{aligned}
 v_1 &\leq \frac{1}{4} \left\{ \text{tr}(M_1) \text{tr}(M_3) + \text{tr}(M_1) \text{tr}(M_3) + \rho \text{tr}(M_1) \text{tr}(M_3) + \text{tr}(\Lambda) \text{tr}(M_3) \right\} \\
 &= \left\{ \left( \frac{1}{2} + \frac{1}{4} \rho \right) \text{tr}(M_1) + \text{tr}(\Lambda) \right\} \text{tr}(M_3) \\
 &= \left\{ \left( \frac{1}{2} + \frac{1}{4} \rho \right) o(1) + \text{tr}(\Lambda) \right\} o(n^{-1} \text{tr}(\Sigma^2)) \\
 &= o(n^{-1} \text{tr}(\Sigma^2) \text{tr}(\Lambda)).
 \end{aligned} \tag{19}$$

$$\begin{aligned}
 v_2 &\leq \frac{1}{36} \left\{ 15 \text{tr}^2(M_2) + 28 \text{tr}^2(M_2) - 36 \text{tr}^2(M_2) + 17 \text{tr}(M_1 M_3) \right. \\
 &\quad \left. + 18 \text{tr}(\Lambda) \text{tr}(M_3) + \text{tr}^2(M_1) \text{tr}(\Sigma^2) + 2 \text{tr}(\Lambda) \text{tr}(M_1) \text{tr}(\Sigma^2) \right. \\
 &\quad \left. + 15 \rho \text{tr}(M_1) \text{tr}(M_3) + \rho^2 \text{tr}(\Sigma^2) \text{tr}^2(M_1) + 2 \text{tr}(M_3) \text{tr}(M_1) \text{tr}(\Lambda) \right. \\
 &\quad \left. + 2 \rho \text{tr}(\Sigma^2) \text{tr}^2(M_1) + \text{tr}(\Lambda^2) \text{tr}(\Sigma^2) \right\} \\
 &\leq \frac{1}{36} \left\{ (24 \text{tr}(M_1) + 18 \text{tr}(\Lambda) + 15 \rho \text{tr}(M_1) + 2 \text{tr}(M_1) \text{tr}(\Lambda)) \text{tr}(M_3) \right. \\
 &\quad \left. + (\text{tr}^2(M_1) + 2 \text{tr}(\Lambda) \text{tr}(M_1) + \rho^2 \text{tr}^2(M_1)) \right. \\
 &\quad \left. + 2 \rho \text{tr}^2(M_1) + \text{tr}(\Lambda^2) \text{tr}(\Sigma^2) \right\} \\
 &= o\left( \frac{1}{n} \text{tr}(\Sigma^2) \text{tr}(\Lambda) \right) + O(\text{tr}(\Lambda^2) \text{tr}(\Sigma^2)) \\
 &= O(\text{tr}(\Lambda^2) \text{tr}(\Sigma^2)).
 \end{aligned} \tag{20}$$

Thus  $v_1 = o(n^{-1} v_2)$ .

Define

$$\hat{T}_n = \|\mathcal{C}(\delta_\beta)\|^2 + \frac{12}{n(n-1)} \sum_{1 \leq i_1 < i_2 \leq n} \tilde{\psi}_2(w_{i_1}, w_{i_2}), \tag{21}$$

where  $\delta_\beta = \beta - \beta_0$ . Then we can get  $T_n - \|\mathcal{C}\delta_\beta\|^2 = (T_n - \hat{T}_n) + (\hat{T}_n - \|\mathcal{C}\delta_\beta\|^2)$ . We have

$$T_n - \hat{T}_n = \binom{n}{4}^{-1} \sum_{\mathcal{A}} \left\{ \tilde{\psi}_4(w_{i_1}, w_{i_2}, w_{i_3}, w_{i_4}) - \sum_{1 \leq k_1 < k_2 \leq 4} \tilde{\psi}_2(w_{i_{k_1}}, w_{i_{k_2}}) \right\},$$

which can be regarded as a U-statistic with the kernel

$$\Psi(w_1, w_2, w_3, w_4) = \tilde{\psi}_4(w_1, w_2, w_3, w_4) - \sum_{1 \leq i_1 < i_2 \leq 4} \tilde{\psi}_2(w_{i_1}, w_{i_2}).$$

Through direct calculation, we can get the projections of  $\Psi$ ,  $\Psi_1(w_1) = -2\tilde{\psi}_1(w_1)$ ,  $\Psi_2(w_1, w_2) = -2 \sum_{i=1}^2 \tilde{\psi}_1(w_i)$ , and  $\Psi_3 = \tilde{\psi}_3(w_1, w_2, w_3) - \sum_{i=1}^3 \tilde{\psi}_1(w_i) - \sum_{1 \leq i < j \leq 3} \tilde{\psi}_2(w_i, w_j)$ . By Hoeffding’s variance decomposition, we have  $\text{var}[\hat{T}_n] = O(n^{-2}v_2)$ ,  $\text{var}[T_n - \hat{T}_n] = o(n^{-2}v_2)$  (see Supplementary Material B in [S1 File](#)).

Because

$$\frac{T_n - \|\mathcal{C}\delta_\beta\|^2}{\sqrt{\text{var}(\hat{T}_n)}} = \frac{T_n - \hat{T}_n}{\sqrt{\text{var}(\hat{T}_n)}} + \frac{\hat{T}_n - \|\mathcal{C}\delta_\beta\|^2}{\sqrt{\text{var}(\hat{T}_n)}} = \frac{\hat{T}_n - \|\mathcal{C}\delta_\beta\|^2}{\sqrt{\text{var}(\hat{T}_n)}} + o_p(1),$$

we only need to show that

$$\frac{\hat{T}_n - \|\mathcal{C}\delta_\beta\|^2}{\sqrt{\text{var}(\hat{T}_n)}} \xrightarrow{P} N(0, 1).$$

From [Eq \(21\)](#) and the form of  $\tilde{\psi}_2$ , let  $\hat{T}_n - \|\mathcal{C}\delta_\beta\|^2 = \hat{T}_n^{(1)} + \hat{T}_n^{(2)}$ , where

$$\begin{aligned} \hat{T}_n^{(1)} &= \binom{n}{2}^{-1} \sum_{1 \leq i_1 < i_2 \leq n} \left\{ \text{tr}(\Delta(\xi_{i_1} - \xi_{i_2})(\xi_{i_1} - \xi_{i_2})' \Sigma \Delta') \right. \\ &\quad + \text{tr}((\epsilon_{i_1} - \epsilon_{i_2})(\xi_{i_1} - \xi_{i_2})' \Sigma \Delta') + \text{tr}(\Delta(\xi_{i_1} \xi_{i_1}' + \Sigma)(\xi_{i_2} \xi_{i_2}' + \Sigma) \Delta') \\ &\quad \left. + \text{tr}(\epsilon_{i_1} \xi_{i_1}' (\xi_{i_2} \xi_{i_2}' + \Sigma) \Delta') + \text{tr}(\Delta(\xi_{i_1} \xi_{i_1}' + \Sigma) \xi_{i_2} \epsilon_{i_2}') \right\}, \end{aligned}$$

and  $\hat{T}_n^{(2)} = \binom{n}{2}^{-1} \sum_{1 \leq i_1 < i_2 \leq n} \text{tr}(\epsilon_{i_1} \xi_{i_1}' \xi_{i_2} \epsilon_{i_2}')$ . Under the assumptions of this theorem and following [Eqs \(19\)](#) and [\(20\)](#) we have  $\text{var}[\hat{T}_n] = \text{var}[\hat{T}_n^{(2)}](1 + o(1))$  and  $\hat{T}_n^{(1)} / \sqrt{\text{var}(\hat{T}_n)} = o_p(1)$ .

To complete the proof, we now need to show

$$\hat{T}_n^{(2)} / \sqrt{\text{var}(\hat{T}_n^{(2)})} = \sqrt{n} \hat{T}_n^{(2)} / \sqrt{\text{tr}(\Lambda^2) \text{tr}(\Sigma^2)} \xrightarrow{P} N(0, 1).$$

Define  $Z_{ni} = \sum_{j=1}^{i-1} \epsilon_j' \xi_j \xi_i' / \sqrt{n}$  and  $\tilde{T}_{nk} = \sum_{i=2}^k Z_{ni}$ , thus  $\tilde{T}_{nm} = \sqrt{n} \hat{T}_n^{(2)}$ , which we define as  $\tilde{T}_n$ . Let  $\mathcal{F}_i = \sigma\{\xi_j, \epsilon_j', j \leq i\}$  be a  $\sigma$ -field generated by  $\{\xi_j, \epsilon_j', j \leq i\}$ . It is obvious to see that  $E[Z_{ni} | \mathcal{F}_{i-1}] = 0$ ,  $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$ . Then it shows that  $\{\tilde{T}_{nk}, \mathcal{F}_k : 2 \leq k \leq n\}$  is a zero mean martingale. Let  $V_{ni}^* = E[Z_{ni}^2 | \mathcal{F}_{i-1}]$ ,  $2 \leq i \leq n$ ,  $V_n^* = \sum_{i=2}^n V_{ni}^*$ . The central limit theorem will

hold Hall 28 if we can show  $V_n^*$  satisfies the following two conditions:

$$\frac{V_n^*}{\text{var}(\tilde{T}_n)} \xrightarrow{P} 1, \tag{22}$$

and for  $\forall \tau > 0$

$$\sum_{i=1}^n \text{tr}^{-1}(\Lambda^2) \text{tr}^{-1}(\Sigma^2) \mathbb{E} \left\{ Z_{ni}^2 I(|Z_{ni}| > \tau \sqrt{\text{tr}(\Lambda^2) \text{tr}(\Sigma^2)}) | \mathcal{F}_{i-1} \right\} \xrightarrow{P} 0.$$

We have  $\text{var}[\tilde{T}_n] = \text{tr}(\Lambda^2) \text{tr}(\Sigma^2)$  and

$$V_n^* = \binom{n}{2}^{-1} \left( \sum_{j=1}^{i-1} \boldsymbol{\varepsilon}_j' \Lambda \boldsymbol{\varepsilon}_j \boldsymbol{\zeta}_j' \Sigma \boldsymbol{\zeta}_j + 2 \sum_{1 \leq j < \ell \leq i} \boldsymbol{\varepsilon}_j' \Lambda \boldsymbol{\varepsilon}_\ell \boldsymbol{\zeta}_j' \Sigma \boldsymbol{\zeta}_\ell \right).$$

Hence we can define  $V_n^*/\text{var}(\tilde{T}_n) = C_{n1} + C_{n2}$ , where

$$C_{n1} = \frac{1}{\binom{n}{2} \text{tr}(\Sigma^2) \text{tr}(\Lambda^2)} \sum_{j=1}^{n-1} j \boldsymbol{\varepsilon}_j' \Lambda \boldsymbol{\varepsilon}_j \boldsymbol{\zeta}_j' \Sigma \boldsymbol{\zeta}_j,$$

$$C_{n2} = \frac{1}{\binom{n}{2} \text{tr}(\Sigma^2) \text{tr}(\Lambda^2)} 2 \sum_{1 \leq j < \ell \leq i} \boldsymbol{\varepsilon}_j' \Lambda \boldsymbol{\varepsilon}_\ell \boldsymbol{\zeta}_j' \Sigma \boldsymbol{\zeta}_\ell.$$

It can be shown that  $\mathbb{E}[C_{n1}] = 1$ , and

$$\text{var}[C_{n1}] = \frac{1}{\binom{n}{2}^2 \text{tr}^2(\Sigma^2) \text{tr}^2(\Lambda^2)} \left\{ \mathbb{E} \left[ \sum_{j=1}^{n-1} j^2 \boldsymbol{\varepsilon}_j' \Lambda \boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_j' \Lambda \boldsymbol{\varepsilon}_j \boldsymbol{\zeta}_j' \Sigma \boldsymbol{\zeta}_j \boldsymbol{\zeta}_j' \Sigma \boldsymbol{\zeta}_j \right] - \sum_{j=1}^{n-1} j^2 \text{tr}^2(\Sigma^2) \text{tr}^2(\Lambda^2) \right\}.$$

As  $\text{tr}(\Sigma^4) = O(\text{tr}^2(\Sigma^2))$ , and  $\text{var}[C_{n1}] \xrightarrow{P} 0$ . Then  $C_{n1} \xrightarrow{P} 1$  (see Supplementary Material B in [S1 File](#)). Similarly,  $\mathbb{E}[C_{n2}] = 0$ ,  $\text{var}[C_{n2}] = \frac{4}{\binom{n}{2}^2 \text{tr}^2(\Sigma^2) \text{tr}^2(\Lambda^2)} \text{tr}(\Lambda^4) \text{tr}(\Sigma^4)$ , then  $C_{n2} \xrightarrow{P} 0$ . In summary, Eq (22) holds.

Since  $\mathbb{E}[Z_{ni}^2 I\{|Z_{ni}| > \tau \sqrt{\text{tr}(\Lambda^2) \text{tr}(\Sigma^2)}\} | \mathcal{F}_{i-1}] \leq \mathbb{E}[Z_{ni}^4 | \mathcal{F}_{i-1}] / \{\tau^2 \text{tr}(\Lambda^2) \text{tr}(\Sigma^2)\}$ , by the law of large numbers, the last step is to prove

$$\sum_{i=1}^n \mathbb{E}[Z_{ni}^4] = o(\text{tr}^2(\Lambda^2) \text{tr}^2(\Sigma^2)). \tag{23}$$

We have  $\mathbb{E}[Z_{ni}^4] = \binom{n}{2}^{-2} \mathbb{E}[\sum_{j=1}^{i-1} (i-2) \boldsymbol{\varepsilon}_j' \Lambda \boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_j' \Lambda \boldsymbol{\varepsilon}_j \boldsymbol{\zeta}_j' \Sigma \boldsymbol{\zeta}_j \boldsymbol{\zeta}_j' \Sigma \boldsymbol{\zeta}_j + \sum_{j=1}^{i-1} (\boldsymbol{\varepsilon}_i' \boldsymbol{\varepsilon}_j)^4 (\boldsymbol{\zeta}_i' \boldsymbol{\zeta}_j)^4]$ , and  $\mathbb{E}[(\boldsymbol{\zeta}_i' \boldsymbol{\zeta}_j)^4] = 6\text{tr}(\Sigma^4) + (6\rho + \rho^2 + 4)\text{tr}(Q_3 \circ Q_3) + 2\text{tr}^2(\Sigma^2)$ , thus under Assumption (16), we

have (see details in [S1 File](#))

$$\begin{aligned} \sum_{i=1}^n E[Z_{ni}^4] &\leq \binom{n}{2}^{-2} \sum_{i=1}^n \left\{ 2 \binom{i-1}{2} + (i-1) \right\} O(\text{tr}^2(\Lambda^2) \text{tr}^2(\Sigma^2)) \\ &= O(n^{-1} \text{tr}^2(\Lambda^2) \text{tr}^2(\Sigma^2)) \\ &= o(\text{tr}^2(\Lambda^2) \text{tr}^2(\Sigma^2)). \end{aligned}$$

Hence we prove that [Eq \(23\)](#) holds. And this completes the proof.

## Supporting information

### S1 File.

(PDF)

## Author Contributions

**Funding acquisition:** Baoxue Zhang.

**Investigation:** Nan Lin.

**Methodology:** Wenjuan Hu, Baoxue Zhang.

**Software:** Wenjuan Hu.

**Writing – original draft:** Wenjuan Hu.

**Writing – review & editing:** Nan Lin.

## References

1. Ramsay JO and Silverman BW. *Functional Data Analysis*. New York: Springer, 2005.
2. Yao F, Müller HG, and Wang J. Functional linear regression analysis for longitudinal data. *The Annals of Statistics*. 2005; 33(6):2873–2903.
3. Malfait N and Ramsay JO. The historical functional linear model. *The Canadian Journal of Statistics*. 2008 Jun; 31(2):115–128.
4. Chiou JM, Yang Y, and Chen Y. Multivariate functional linear regression and prediction. *Journal of Multivariate Analysis*. 2016; 146:301–312, April.
5. Kokoszka P, Maslova I, Sojka J, Zhu L. Testing for lack of dependence in the functional linear model. *Canadian Journal of Statistics*. 2008 Jun; 36(2):207–222.
6. Cardo H, Ferraty F, Mas A, and Sarda P. Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics*. 2016 Mar; 30(1):241–255.
7. Kong D, Staicu AM, Maity A. Classical testing in functional linear models. *Journal of Nonparametric Statistics*. 2016; 28(4):813–838. <https://doi.org/10.1080/10485252.2016.1231806> PMID: 28955155
8. Su Y, Di C, and Hsu L. Hypothesis testing in functional linear models. *Biometrics*. 2017 Jun; 73(2):551–561. <https://doi.org/10.1111/biom.12624> PMID: 28295175
9. Bai Z, Saranadasa H. Effect of high dimension: by an example of a two sample problem. *Statistica Sinica*. 1996; 6:311–329.
10. Zhong P and Chen S. Tests for high-dimensional regression coefficients with factorial designs. *Journal of the American Statistical Association*. 2011 Jan; 106(493):260–274.
11. Delsol L, Ferraty F, and Vieu P. Structural test in regression on functional variables. *Journal of Multivariate Analysis*. 2011 Mar; 102(3):422–447.
12. Patilea V, Sellero CS, and Saumard M. Testing the predictor effect on a functional response. *Journal of the American Statistical Association*. 2016 Jul; 111(516):1684–1695.
13. Nadaraya EA. On estimating regression. *Theory of Probability & Its Applications*. 1964; 9(1):141–142.
14. Watson GS. Smooth regression analysis. *Sankhya: The Indian Journal of Statistics, Series A(1961-2002)*. 1964; 26:359–372.

15. Delsol L. No effect tests in regression on functional variable and some applications to spectrometric studies. *Computational Statistics*. 2013; 28(4):1775–1811.
16. Horváth L and Kokoszka P. *Inference for Functional Data with Applications*. New York: Springer, April 2012.
17. Shin H and Lee MH. On prediction rate in partial functional linear regression. *Journal of Multivariate Analysis*. 2012 Jan; 103(1)93–106.
18. Ramsay JO and Dalzell CJ. Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1991; 53(3):539–572.
19. Hervé C, Ferraty F, and Sarda P. Spline estimators for the functional linear model. *Statistica Sinica*. 2003 Jul; 13:571–591.
20. Yao F, Müller HG, and Wang JL. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*. 2005; 100(470):577–590.
21. Yao F, Lei E, Wu Y. Effective dimension reduction for sparse functional data. *Biometrika*. 2015 Jun; 102(2):421–437. <https://doi.org/10.1093/biomet/asv006> PMID: 26566293
22. Fan Y, James GM, and Radchenko P. Functional additive regression. *Ann. Statist.* 2015 Oct; 43(5):2296–2325.
23. Hoeffding W. A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics*. 1948 Sep; 19:293–325.
24. Serfling RJ. *Approximation Theorems of Mathematical Statistics*. New York: Wiley.;1980.
25. Lee AJ. *U-statistics: Theory and Practice*. New York: Marcel Dekker, 1990.
26. Chen S. X. and Qin Y.-L. A two-sample test for high-dimensional data with applications to gene-set testing. *Annals of Statistics* 2010; 38(2):808–835.
27. Bellman R. *Some Inequalities for Positive Definite Matrices*. Springer, 1980.
28. Hall P and Heyde CC. *Martingale Limit Theory and Its Application*. New York: Academic Press, 1980.