


Article

A U-Statistic for Testing the Lack of Dependence in Functional Partially Linear Regression Model

Fanrong Zhao ¹  and Baoxue Zhang ^{2,*}¹ School of Mathematics and Statistics, Shanxi University, Taiyuan 030006, China; zhaofanrong@sxu.edu.cn² School of Statistics, Capital University of Economics and Business, Beijing 100070, China

* Correspondence: zhangbaoxue@cueb.edu.cn

Abstract: The functional partially linear regression model comprises a functional linear part and a non-parametric part. Testing the linear relationship between the response and the functional predictor is of fundamental importance. In cases where functional data cannot be approximated with a few principal components, we develop a second-order U-statistic using a pseudo-estimate for the unknown non-parametric component. Under some regularity conditions, the asymptotic normality of the proposed test statistic is established using the martingale central limit theorem. The proposed test is evaluated for finite sample properties through simulation studies and its application to real data.

Keywords: asymptotic normality; functional partially linear regression model; Nadaraya–Watson estimate; U-statistic

MSC: 62G10; 62G20; 62F03



Citation: Zhao, F.; Zhang, B. A U-Statistic for Testing the Lack of Dependence in Functional Partially Linear Regression Model. *Mathematics* **2024**, *12*, 2588. <https://doi.org/10.3390/math12162588>

Academic Editor: Jose Luis Vicente Villardon

Received: 17 July 2024

Revised: 14 August 2024

Accepted: 19 August 2024

Published: 21 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the past few decades, functional data analysis has been widely developed and applied in various fields, such as medicine, biology, economics, environmetrics, and chemistry (see [1–5]). An important model in functional data analysis is the partial functional linear model, which includes the parametric linear part and the functional linear part. To make the relationships between variables more flexible, the parametric linear part is usually replaced by the non-parametric part. This model is known as the functional partially linear regression model, which has been studied in [6–8]. The functional partially linear regression model is formulated as follows:

$$Y = g(u) + \int_0^1 \alpha(s)X(s)ds + \varepsilon, \quad (1)$$

where Y is the response variable. $X(\cdot)$ denotes the functional predictor, characterized by its mean function, $\mu_0(\cdot)$, and covariance operator, \mathcal{C} . The slope function $\alpha(\cdot)$ is an unknown function. $g(\cdot)$ is a general continuous function defined on a compact support Ω . The random error ε has a mean of zero and a finite variance σ^2 , and is statistically independent of the predictor $X(\cdot)$. When $g(\cdot)$ is a constant, model (1) reduces to a functional linear model. Refer to [9–11] for further details. With $g(\cdot)$ representing the parametric linear component, model (1) is identified as a partially functional linear model, an area explored in [12–14].

Hypothesis testing plays a critical role in statistical inference. For testing the linear relationship between the response and the functional predictor in the functional linear model, functional principal component analysis (FPCA) is a major idea in constructing test statistics. See [9,10,15]. Taking into account the flexibility of non-parametric functions, Ref. [6] introduced the functional partially linear model. Refs. [7,8] constructed the

estimators of the slope functions based on spline and FPCA respectively. They utilized B-spline for estimating non-parametric components. In the context of predictors with additive measurement error, ref. [16] investigated estimators for the slope function and non-parametric component using FPCA and kernel smoothing methods. Ref. [17] established estimators of the slope function, non-parametric component, and mean of the response variable in the presence of randomly missing responses.

However, testing the relationship between the response variable and functional predictor in the functional partially linear regression model has been rarely considered so far. In this paper, the following hypothesis testing for model (1) will be considered:

$$H_0 : \alpha(t) = \alpha_0(t) \quad v.s. \quad H_1 : \alpha(t) \neq \alpha_0(t), \tag{2}$$

where $\alpha_0(t)$ denotes an assigned function. Here we assume $\alpha_0(t) = 0$ without compromising generality. To test (2) within the framework of model (1), a chi-square test was devised by [18]. This test relies on estimators for the nonlinear and slope functions. The underlying assumption is that the functional data can be well-approximated by a small number of principal components.

In particular, we focus on functional data that cannot be approximated with a few principal components, such as the velocity and acceleration of changes in China’s Air Quality Index (AQI). If these changes are represented by some curves, the velocity and acceleration are equivalent to the first and second derivatives of the AQI, respectively. The number of principal components selected by FPCA may approach approximately 30. Only several research studies have considered this data structure in the functional data analysis. Ref. [19] constructed a FLUTE test based on order-four U-statistic for the testing in the functional linear model, which can be computationally very costly. In order to save calculation time, ref. [20] developed a faster test using a order-two U-statistic. Inspired by this, we introduce a non-parametric U-statistic that integrates functional data analysis with the traditional kernel method to test (2).

The structure of the paper is as follows. Section 2 details the development of a new test procedure for the functional partially linear regression model. Section 3 presents the theoretical properties of the proposed test statistic under some regularity conditions. Section 4 includes a simulation study to evaluate the finite sample performance of the proposed test. Section 5 presents the application of the test to spectrometric data. The proofs of the primary theoretical results are presented in Appendix A.

2. Test Statistic

Assume Y and U are random variables taking real values. $X(\cdot)$ is a stochastic process with sample paths in $L^2[0, 1]$, which is the set of all square-integrable functions defined on $[0, 1]$. Let $\langle \cdot, \cdot \rangle, \|\cdot\|$ represent the inner product and norm in $L^2[0, 1]$, respectively. $\{(Y_i, X_i(\cdot), U_i), i = 1, 2, \dots, n\}$ constitutes a random sample drawn from model (1),

$$Y_i = \int_0^1 \alpha(s) X_i(s) ds + g(U_i) + \varepsilon_i, \quad i = 1, 2, \dots, n. \tag{3}$$

For any given $\alpha(t) \in L^2[0, 1]$, we move $\alpha(t)$ to the left,

$$Y_i - \langle X_i, \alpha \rangle = g(U_i) + \varepsilon_i, \quad i = 1, 2, \dots, n. \tag{4}$$

Hence, model (4) simplifies to a classical non-parametric model. A pseudo-estimate for the non-parametric function employing Nadaraya–Watson method, can be formulated as follows:

$$\hat{g}(U_i) = \sum_{j \neq i}^n \frac{K_h(U_j - U_i)(Y_j - \langle X_j, \alpha \rangle)}{\sum_{k \neq i}^n K_h(U_k - U_i)}, \quad i = 1, 2, \dots, n, \tag{5}$$

where $K_h(\cdot) = K(\cdot/h)/h$ with $K(\cdot)$ being a preselected kernel function. A kernel function maps from the set of real numbers to the set of real numbers. It adheres to the following

properties: (i) Non-negativity: the kernel function $K(\cdot)$ must be non-negative. (ii) Normalization: The integral (or sum in the discrete case) of the kernel function over the entire real line must equal 1, which means it can be interpreted as a probability density function. The bandwidth h in (5) is typically selected through data-driven procedures, such as cross-validation techniques. Here, we estimate non-parametric $g(U_i)$ without the i th sample.

Let

$$\begin{aligned} \mathbf{W}_{-i} &= (W_{i1}, \dots, W_{i(i-1)}, W_{i(i+1)}, \dots, W_{in})^T, \\ \langle \mathbf{X}_{-i}, \alpha \rangle &= (\langle X_1, \alpha \rangle, \dots, \langle X_{i-1}, \alpha \rangle, \langle X_{i+1}, \alpha \rangle, \dots, \langle X_n, \alpha \rangle)^T, \\ \mathbf{Y}_{-i} &= (Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n)^T, \mathbf{X}_{-i} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)^T, \end{aligned}$$

where $W_{ij} = K_h(U_j - U_i) / \sum_{k \neq i} K_h(U_k - U_i)$. So the pseudo-estimate (5) of non-parametric function can be reformulated in matrix form as

$$\check{g}(U_i) = \mathbf{W}_{-i}^T (\mathbf{Y}_{-i} - \langle \mathbf{X}_{-i}, \alpha \rangle).$$

Substituting $\check{g}(U_i)$ for $g(U_i)$ in model (3), we have

$$\check{Y}_i = \langle \check{X}_i, \alpha \rangle + \varepsilon_i, \tag{6}$$

where $\check{X}_i(t) = X_i(t) - \mathbf{W}_{-i}^T \mathbf{X}_{-i}(t)$, $\check{Y}_i = Y_i - \mathbf{W}_{-i}^T \mathbf{Y}_{-i}$. If we denote $\mu_{it} \triangleq \mu(U_i, t) = E[X_1(t)|U_i]$, where “ \triangleq ” stands for “defined as”. Then $\hat{\mu}_{it} = \mathbf{W}_{-i}^T \mathbf{X}_{-i}(t)$ can be the estimator of the conditional expectation μ_{it} for any $t \in [0, 1]$.

Given an arbitrary orthonormal basis $\{\psi_j\}_{j=1}^\infty$ in $L^2[0, 1]$, the functional predictor $X(\cdot)$ and the slope function $\alpha(\cdot)$ admit the following series expansions: Let p represent the number of truncated basis functions, as follows:

$$X_i(t) = \sum_{j=1}^p \zeta_{ij} \psi_j(t) + \sum_{j=p+1}^\infty \zeta_{ij} \psi_j(t), \quad \alpha(t) = \sum_{j=1}^p \beta_j \psi_j(t) + \sum_{j=p+1}^\infty \beta_j \psi_j(t), \tag{7}$$

where $\zeta_{ij} = \langle X_i, \psi_j \rangle$, $\beta_j = \langle \alpha, \psi_j \rangle$. Let $\check{\zeta}_{ij} = \langle \check{X}_i, \psi_j \rangle$, then the model (6) can be rewritten as follows:

$$\check{Y}_i = \sum_{j=1}^\infty \check{\zeta}_{ij} \beta_j + \varepsilon_i = \sum_{j=1}^p \check{\zeta}_{ij} \beta_j + \sum_{j=p+1}^\infty \check{\zeta}_{ij} \beta_j + \varepsilon_i.$$

Denote $\check{\zeta}_i = (\check{\zeta}_{i1}, \check{\zeta}_{i2}, \dots, \check{\zeta}_{ip})^T$, which has mean μ and covariance matrix Σ . Let

$$\begin{aligned} \check{\zeta}_{-i} &= (\check{\zeta}_{11}, \dots, \check{\zeta}_{i-1,1}, \check{\zeta}_{i+1,1}, \dots, \check{\zeta}_{n,1})^T, \quad \check{\zeta}_i = (\check{\zeta}_{i1}, \check{\zeta}_{i2}, \dots, \check{\zeta}_{ip})^T, \\ \check{\zeta}_{-i} &= (\check{\zeta}_{11}, \dots, \check{\zeta}_{i-1,1}, \check{\zeta}_{i+1,1}, \dots, \check{\zeta}_{n,1})^T, \quad \beta = (\beta_1, \beta_2, \dots, \beta_p)^T. \end{aligned}$$

For model (3), the approximation error is defined as follows:

$$e_i = \int_0^1 \alpha(s) X_i(s) ds - \sum_{k=1}^p \check{\zeta}_{ik} \beta_k.$$

To investigate the influence of the approximation error, we impose the following conditions on the functional predictors and regression function:

(C1) The functional predictors $\{X_i(\cdot)\}_{i=1}^n$ and the regression function $\alpha(t)$ adhere to the following conditions:

(i) The functional predictors $\{X_i(\cdot)\}_{i=1}^n$ reside within a Sobolev ellipsoid of order two, then there exists a universal constant C , such that $\sum_{j=1}^\infty \check{\zeta}_{ij}^2 j^4 \leq C^2$, for $i = 1, \dots, n$.

(ii) The regression function satisfies $\int \alpha^2(t) dt \leq D$, where D is a constant.

By applying the Cauchy–Schwarz inequality, we obtain the following:

$$e_i^2 = \left(\sum_{j=p+1}^{\infty} \xi_{ij} \beta_j \right)^2 \leq \sum_{j=p+1}^{\infty} \xi_{ij}^2 j^4 \sum_{j=p+1}^{\infty} j^{-4} \alpha_j^2 \leq \frac{C^2 D}{p^4}.$$

Then the approximation error can be ignored as $p \rightarrow \infty$. Model (6) becomes as follows:

$$\check{Y}_i = \sum_{j=1}^p \check{\xi}_{ij} \beta_j + \varepsilon_i,$$

which is a high-dimensional partial linear model. Since

$$E \|(X_i - E[X_i|U_i])(Y_i - E(Y_i|U_i))\|^2 \tag{8}$$

can be an effective measure for assessing the distance between $\alpha(\cdot)$ and zero for test (2). Motivated by [21], we construct the following test statistic by estimating (8).

$$T_{np} = \left(1 - \frac{2}{n}\right)^{-2} \binom{n}{2}^{-1} \sum_{i=2}^n \sum_{j=1}^{i-1} \Delta_{ij}(\check{X}) \Delta_{ij}(\check{Y}), \tag{9}$$

where

$$\begin{aligned} \Delta_{ij}(\check{X}) &= \langle \check{X}_i - \bar{\check{X}}, \check{X}_j - \bar{\check{X}} \rangle + \frac{\langle \check{X}_i - \check{X}_j, \check{X}_i - \check{X}_j \rangle}{2n}, \\ \Delta_{ij}(\check{Y}) &= (\check{Y}_i - \bar{\check{Y}})(\check{Y}_j - \bar{\check{Y}}) + \frac{(\check{Y}_i - \check{Y}_j)^2}{2n}, \end{aligned}$$

where $\bar{\check{X}}(t)$ and $\bar{\check{Y}}$ denote the sample means of $\check{X}_i(t)$ and \check{Y}_i , respectively. By some calculations, we can obtain $E[\Delta_{ij}(\check{X})] = 0$, $E[\Delta_{ij}(\check{Y})] = 0$. The test statistic T_{np} quantifies the discrepancy between $\alpha(\cdot)$ and 0 under the null hypothesis. High values of the test statistic T_{np} suggest evidence in favor of the alternative hypothesis, prompting the rejection of the null hypothesis.

3. Asymptotic Theory

To achieve the asymptotic properties of the proposed test, we first suppose the following conditions based on [19,21]. We denote the following:

$$\mu(U_i) = (\mu_1(U_i), \mu_2(U_i), \dots, \mu_p(U_i))^T \triangleq E[\xi|U_i],$$

$$\Sigma_*(U_i) = E[\xi_i \xi_i^T | U_i] - \mu(U_i) \mu^T(U_i), \quad \Sigma_* = \Sigma - E[\mu(U_1) \mu^T(U_1)].$$

A condition on the dimensionality of matrix Σ_* is stipulated as follows:

(C2) As $n \rightarrow \infty, p \rightarrow \infty; \Sigma_* > 0, \text{tr}(\Sigma_*^4) = o(\text{tr}^2(\Sigma_*^2))$.

(C3) For a constant $m \geq p$, there exists an m -dimensional random vector $Z_i = (Z_{i1}, \dots, Z_{im})^T$ such that $\xi_i = E[\xi_i|U_i] + \Gamma(U_i)Z_i$. The vector Z_i is characterized by $E(Z_i) = 0, \text{var}(Z_i) = I_m$, and for any $U_i, \Gamma(U_i)$ is a $p \times m$ matrix with $\Gamma(U_i)\Gamma^T(U_i) = \Sigma_*(U_i)$. It is assumed that each random vector $\{Z_i, i = 1, \dots, n\}$ has finite fourth moments and $E(Z_{ij}^4) = 3 + \Delta$ for some constant Δ . Moreover, we assume the following:

$$E\left(Z_{ij_1}^{l_1} Z_{ij_2}^{l_2} \dots Z_{ij_d}^{l_d}\right) = E\left(Z_{ij_1}^{l_1}\right) E\left(Z_{ij_2}^{l_2}\right) \dots E\left(Z_{ij_d}^{l_d}\right)$$

for $\sum_{k=1}^d l_k \leq 4$ and $j_1 \neq j_2 \neq \dots \neq j_d$, where d is a positive integer.

(C4) $\beta^T \Sigma_* \beta = o(h^2)$, and $\beta^T \Sigma_*^3 \beta = o(\text{tr}(\Sigma_*^2)/n)$.

(C5) The error term satisfies $E[\varepsilon^4] < +\infty$.

(C6) The random variable U is confined to a compact domain Ω , and its density function f exhibits a continuously differentiable second derivative and bounded away from 0 on its support. The kernel $K(\cdot)$ is a symmetric probability density with compact support and is Lipschitz continuous.

(C7) $E(\xi_1|U_1)$ and $g(\cdot)$ are Lipschitz continuous and admit continuous second-order derivatives.

(C8) It is assumed that the sample size n and the smoothing parameter h satisfy the following: $\lim_{n \rightarrow \infty} h = 0, \lim_{n \rightarrow \infty} nh = \infty, \lim_{n \rightarrow \infty} nh^4 = 0$.

(C9) The truncated number p and the sample size n are assumed to satisfy $p = o(n^2h^2)$.

Condition (C2) is widely utilized in high-dimensional data research (see [21–23]). Condition (C3) resembles a factor model. To assess local power, we further impose condition (C4) on the coefficient vector β . In fact, (C4) can serve as the local alternative as its distance measurement between β and 0. This local alternative can be also found in [21]. (C5) is the typical assumption for the error term ε . Conditions (C6–C8) are very common in non-parametric smoothing. (C9) is a technical condition that is needed to derive the theorems.

In practical applications, the data must satisfy conditions (C1–C3) and (C7). Conditions (C1) and (C7) are generally met for most datasets. (C2) does not specify a relationship between p and n . The matrix’s positive definiteness ensures that the regression coefficients can be identified. $\text{tr}(\Sigma_*^4) = o(\text{tr}^2(\Sigma_*^2))$ holds if the eigenvalues of Σ_* are all bounded or the largest eigenvalue is of smaller order than $(p - b)^{1/2}b^{-1/4}$, where b is the number of unbounded eigenvalues. Condition (C3) essentially assumes that the functional predictor $X_i(t)$ is based on a latent factor model, where the factor loadings meet the pseudo-independence assumption. If $X(t)$ is a Gaussian process, it can be expanded as $X(t) = \sum_{j=1}^m \sqrt{\lambda_j} N_j u_j(t)$, with N_j being independent standard normal random variables. This expansion is a special case of (C3) when the (i, j) -th element of the transformation matrix Γ is $\sqrt{\lambda_j} \langle u_j, \phi_i \rangle$. These conditions are generally met for most data and do not affect the validity of the proposed test. Many datasets can be regarded as following a Gaussian process, such as changes in gene expression levels, logarithmic returns on financial asset prices, soil moisture, and temperature distribution.

We present the asymptotic theory for the proposed test statistic under the null hypothesis and local alternative (C4) in the subsequent two theorems:

Theorem 1. Under the assumptions of conditions (C1), (C3–C9), it follows that

$$(i) \quad E(T_{np}) = \|\mathcal{C}^*(\alpha)\|^2 + o\left(\sqrt{\text{tr}(\Sigma_*^2)/n}\right);$$

$$(ii) \quad T_{np} - \|\mathcal{C}^*(\alpha)\|^2 = \frac{1}{\binom{n}{2}} \sum_{i=2}^n \sum_{j=1}^{i-1} \langle X_i - \mu_{it}, X_j - \mu_{jt} \rangle \varepsilon_i \varepsilon_j + o_p\left(\sqrt{\text{tr}(\Sigma_*^2)/n}\right),$$

where $\mathcal{C}^*(\alpha) = E[\langle X_i - \mu_{it}, \alpha \rangle (X_i - \mu_{it})]$. It can be regarded as the covariance operator of a random variable $X_i - \mu_{it}$.

Theorem 2. Assume conditions (C1–C3) and (C5–C9) hold, we then have the following results under either the null hypothesis or the local alternative (C4):

$$\frac{n(T_{np} - \|\mathcal{C}^*(\alpha)\|^2)}{\sigma^2 \sqrt{2\text{tr}(\Sigma_*^2)}} \xrightarrow{D} N(0, 1), \text{ as } n \rightarrow \infty,$$

where \xrightarrow{D} represents convergence in distribution.

Theorem 2 demonstrates that, under the local alternative hypothesis (C4), the proposed test statistic possesses the following asymptotic local power at the nominal significance level α :

$$\Psi(\beta) = \Phi\left(-z_\alpha + \frac{n\|C^*(\alpha)\|^2}{\sigma^2\sqrt{2\text{tr}(\Sigma_*^2)}}\right),$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal, and z_α represents its $(1 - \alpha)$ th quantile. We define $\eta(\alpha) = C^*(\alpha)/\sigma^2\sqrt{2\text{tr}(\Sigma_*^2)}$, which represents the signal-to-noise ratio. When the term $\eta(\alpha) = o(1/n)$, the power converges to α , then the power converges to 1 if it has a high order of $1/n$. This implies that the proposed test is consistent. The power performance will be demonstrated through simulations in Section 4.

According to Theorem 2, the proposed test statistic leads to the rejection of H_0 at a significance level α when

$$nT_{np} \geq \sqrt{2\hat{\sigma}^2\widehat{\text{tr}(\Sigma_*^2)}}z_\alpha,$$

where $\hat{\sigma}^2$ and $\widehat{\text{tr}(\Sigma_*^2)}$ serve as consistent estimators for σ^2 and $\text{tr}(\Sigma_*^2)$, respectively. We use a similar method as in [24] to estimate the trace. That is,

$$\widehat{\text{tr}(\Sigma_*^2)} = Y_{1n} - 2Y_{2n} + Y_{3n},$$

where $Y_{1n} = \frac{1}{A_n^2} \sum_{i \neq j} \langle \check{X}_i^T, \check{X}_j \rangle^2$, $Y_{2n} = \frac{1}{A_n^3} \sum_{i \neq j \neq k} \langle \check{X}_i^T, \check{X}_j \rangle \langle \check{X}_j^T, \check{X}_k \rangle$, $Y_{3n} = \frac{1}{A_n^4} \sum_{i \neq j \neq k \neq l} \langle \check{X}_i^T, \check{X}_j \rangle \langle \check{X}_k^T, \check{X}_l \rangle$ with $A_n^m = n!/(n - m)!$. And the simple estimator $\hat{\sigma}^2 = (n - 1)^{-1} \sum_{i=1}^n (\check{Y}_i - \bar{\check{Y}})^2$ is used, which is consistent under the null hypothesis testing.

4. Simulation

This section evaluates the finite sample performance of the proposed test, including its size and power. The assessment is conducted through a series of simulation studies. Through numerical simulations, we will validate that the distribution of the proposed test statistic under the null hypothesis is consistent with the properties stated in Theorem 1. For each simulation, we create 1000 Monte Carlo samples. The basis expansion and FPCA are conducted using the R package `fa`.

To mitigate the probability of both Type I and Type II errors in the testing procedure, the sample size must be adequately large. However, to maintain computational efficiency during the numerical simulations, the sample size should not be excessively large. Consequently, the sample size n in this study has been set within a range of 50 to 200. To validate the effectiveness of our proposed test, the parameters are flexibly set.

Here we compare the proposed test T_{np} with the chi-square test T_n constructed by [18]. The cumulative percentage of total variance (CPV) method is used to estimate the number of principal components in T_n . Let CPV, explained by the first m empirical functional principal components, be defined as follows:

$$\text{CPV} = \frac{\sum_{i=1}^m \hat{\lambda}_i}{\sum_{i=1}^p \hat{\lambda}_i},$$

where $\{\hat{\lambda}_i\}_{i=1}^p$ is the estimate of the eigenvalue of the covariance operator. The smallest value of m for which CPV (m) surpasses the threshold of 95% is selected in this section. We denote p as the number of basis functions used to fit curves. The simulated data are produced according to the following model:

$$Y_i = \int_0^1 \alpha(s)X_i(s)ds + g(U_i) + \varepsilon_i, \quad i = 1, 2, \dots, n.$$

where $g(U_i) = 2U_i$ or $g(U_i) = 2 + \sin(2\pi U_i)$, and $\{U_i, i = 1, 2, \dots, n\}$ is independently drawn from the uniform distribution on $(0, 1)$. To analyze the impact of different error dis-

tributions, the following four distributions will be selected: (1) $\varepsilon_i \sim N(0, 1)$, (2) $\varepsilon_i \sim t(3)/\sqrt{3}$, (3) $\varepsilon_i \sim \Gamma(1, 1) - 1$, (4) $\varepsilon_i \sim (\ln \text{norm}(0, 1) - \sqrt{e})/\sqrt{e(e-1)}$. All results about $g(U_i) = 2U_i$ are presented in Supplementary Materials.

We next report the simulation results for two data structures of the predictor $X(t)$.

1. The predictor $X(t)$ is defined as $\sum_{j=1}^{50} \xi_j \phi_j(t)$, with ξ_j normally distributed with mean 0 and variance $\lambda_j = 10((j - 1/2)\pi)^{-2}$, $\phi_j(t) = \sqrt{2}(j - 1/2)\pi t$ for $j = 1, 2, \dots, 50$. The slope function $\alpha(t)$ is given by $c(\sqrt{2} \sin(\pi t/2) + 3\sqrt{2} \sin(3\pi t/2))$, where the coefficient c ranges from 0 to 0.2. $c = 0$ corresponds to the null hypothesis. The number of basis functions used to fit curves and the sample size are taken as follows: $p = 11, 49, n = 50, 100$. Under different error distributions, Tables 1 and 2 evaluate the empirical size and power of both tests for different non-parametric functions when the nominal level α is 0.05.

Table 1. When $g(u) = 2u$, the empirical size and power for two tests are evaluated.

(n, p)	c	N(0,1)		t(3)		$\Gamma(1, 1)$		lnorm(0,1)	
		T_n	T_{np}	T_n	T_{np}	T_n	T_{np}	T_n	T_{np}
(50,11)	0.00	0.069	0.060	0.074	0.077	0.071	0.071	0.084	0.053
	0.05	0.097	0.101	0.133	0.102	0.135	0.107	0.148	0.111
	0.10	0.250	0.211	0.292	0.244	0.269	0.225	0.337	0.283
	0.15	0.474	0.383	0.555	0.448	0.494	0.415	0.576	0.470
	0.20	0.713	0.584	0.761	0.631	0.738	0.602	0.755	0.656
(100,11)	0.00	0.049	0.052	0.055	0.052	0.058	0.059	0.049	0.052
	0.05	0.233	0.195	0.275	0.225	0.217	0.195	0.312	0.268
	0.10	0.689	0.603	0.746	0.660	0.715	0.618	0.743	0.652
	0.15	0.961	0.877	0.956	0.913	0.963	0.899	0.931	0.884
	0.20	0.998	0.984	0.986	0.975	0.995	0.975	0.978	0.965
(100,49)	0.00	0.057	0.060	0.050	0.061	0.051	0.049	0.055	0.047
	0.05	0.224	0.203	0.282	0.255	0.236	0.225	0.305	0.288
	0.10	0.741	0.607	0.757	0.665	0.718	0.615	0.747	0.659
	0.15	0.962	0.900	0.947	0.871	0.950	0.884	0.938	0.886
	0.20	0.998	0.981	0.987	0.977	0.997	0.978	0.988	0.969

Table 2. When $g(u) = 2 + \sin(2\pi u)$, the empirical size and power for two tests are evaluated.

(n, p)	c	N(0,1)		t(3)		$\Gamma(1, 1)$		lnorm(0,1)	
		T_n	T_{np}	T_n	T_{np}	T_n	T_{np}	T_n	T_{np}
(50,11)	0.00	0.062	0.064	0.072	0.069	0.071	0.072	0.060	0.054
	0.05	0.087	0.091	0.109	0.096	0.112	0.103	0.117	0.106
	0.10	0.235	0.197	0.250	0.219	0.241	0.208	0.293	0.249
	0.15	0.420	0.359	0.502	0.419	0.449	0.389	0.506	0.449
	0.20	0.658	0.545	0.724	0.605	0.694	0.573	0.735	0.638
(100,11)	0.00	0.062	0.062	0.050	0.046	0.063	0.060	0.054	0.060
	0.05	0.217	0.205	0.239	0.219	0.235	0.208	0.255	0.240
	0.10	0.668	0.568	0.714	0.617	0.696	0.605	0.748	0.644
	0.15	0.946	0.874	0.947	0.883	0.946	0.852	0.927	0.873
	0.20	0.996	0.980	0.993	0.972	1.000	0.979	0.987	0.964
(100,49)	0.00	0.050	0.062	0.056	0.063	0.047	0.067	0.060	0.056
	0.05	0.226	0.216	0.249	0.202	0.209	0.199	0.277	0.256
	0.10	0.674	0.562	0.734	0.611	0.692	0.582	0.733	0.639
	0.15	0.943	0.873	0.943	0.890	0.941	0.855	0.925	0.889
	0.20	0.997	0.976	0.994	0.981	0.998	0.980	0.982	0.955

From Tables 1 and 2, the following can be seen: (i) The performances of both tests remain consistent across various error distributions and non-parametric functions; (ii) Because T_{np} is intended for functional data beyond the reach of a few principal components, the power of the proposed test is somewhat less than that of T_n . (iii) The power of the test increases with the sample size n , but it is not significantly affected by increases in the parameter value p . In fact, for the functional data structure given in Simulation 1, the number of principal components selected is relatively small, regardless of the number of basis functions used to fit the functional data.

2. The functional predictor is constructed using the expansion in (7), with ϕ_k representing the Fourier basis function on $[0,1]$ defined as $\phi_1(t) = 1, \phi_2(t) = \sqrt{2}\sin(2\pi t), \phi_3(t) = \sqrt{2}\cos(2\pi t), \phi_4(t) = \sqrt{2}\sin(4\pi t), \phi_5(t) = \sqrt{2}\cos(4\pi t), \dots$. The first p of the basis functions will be used to generate the prediction function and slope function. Let $X_i(t) = \sum_{j=1}^p \xi_{ij}\phi_j(t), \alpha(t) = \sum_{j=1}^p \beta_j\phi_j(t)$, where $p = 11, 49, 201, 365, n = 50, 100, 200$, the coefficient of slope function $\{\beta_i = |\beta|/\sqrt{p}, i = 1, \dots, p\}$ with $|\beta|^2 = c * 10^{-2}$ and c varying from 0 to 1. $c = 0$ corresponds to the case in which H_0 is true. The coefficients of predictor ξ_{ij} follow the moving average model:

$$\xi_{ij} = \rho_1 Z_{ij} + \rho_2 Z_{i(j+2)} + \dots + \rho_T Z_{i(p+T-1)},$$

where the constant T adjusts the degree of dependence among the elements of the predictor. $\{Z_{ij}, Z_{i(j+1)}, \dots, Z_{i(p+T-1)}\}$ are drawn independently from the distribution $N(0, I_{p+T-1})$ with $T = 10$. The element at the (j, k) position of the covariance matrix Σ for coefficient vector ξ_i is $\sum_{l=1}^{T-|j-k|} \rho_l \rho_{l+|j-k|} I\{|j-k| < T\}$, where $\{\rho_k, k = 1, \dots, T\}$ is independently generated from the uniform distribution $U(0, 1)$.

The bandwidth is chosen using cross-validation (CV). At a significance level of $\alpha = 0.05$, Table 3 delineates the empirical size and power of the two tests when the function $g(\cdot)$ is linear. Table 4 presents the results for the case where $g(\cdot)$ is a trigonometric function.

From Tables 3 and 4, the number of basis functions used for fitting functions has a very important impact on the test. Specifically, (i) Across various error distributions, as p increases, the empirical size of test T_n significantly exceeds the nominal level, whereas our proposed test T_{np} maintains stable performance; (ii) The power of the test increases with the sample size n . Conversely, it decreases as the values of p increase. (iii) The proposed test demonstrates robustness across all scenarios presented in this simulation study. Actually, for the functional data structure given in Simulation 2, selecting too many principal components negates the effectiveness of FPCA-based test statistics. Instead, the proposed test has great advantages (see bold numbers in Tables 3 and 4).

Table 3. When $g(u) = 2u$, the empirical size and power for two tests are evaluated.

(n, p)	c	N(0,1)		t(3)		$\Gamma(1,1)$		Inorm(0,1)	
		T_n	T_{np}	T_n	T_{np}	T_n	T_{np}	T_n	T_{np}
(50,11)	0.00	0.097	0.066	0.088	0.076	0.096	0.074	0.104	0.059
	0.25	0.378	0.454	0.442	0.518	0.389	0.443	0.487	0.572
	0.50	0.589	0.695	0.669	0.749	0.610	0.702	0.725	0.783
	0.75	0.765	0.832	0.811	0.861	0.769	0.847	0.832	0.882
	1.00	0.865	0.917	0.883	0.925	0.867	0.912	0.889	0.922
(100,11)	0.00	0.060	0.067	0.086	0.064	0.072	0.058	0.062	0.049
	0.25	0.511	0.711	0.582	0.738	0.569	0.739	0.618	0.760
	0.50	0.820	0.932	0.857	0.932	0.846	0.922	0.841	0.924
	0.75	0.949	0.984	0.943	0.975	0.935	0.973	0.917	0.967
	1.00	0.982	0.998	0.971	0.986	0.971	0.989	0.958	0.984
(100,49)	0.00	0.245	0.064	0.224	0.058	0.236	0.058	0.202	0.048
	0.25	0.541	0.498	0.570	0.543	0.534	0.501	0.563	0.564
	0.50	0.754	0.771	0.804	0.807	0.767	0.776	0.769	0.798
	0.75	0.884	0.910	0.899	0.936	0.899	0.904	0.879	0.887
	1.00	0.957	0.966	0.949	0.968	0.956	0.964	0.928	0.934

Table 4. When $g(u) = 2 + \sin(2\pi u)$, the empirical size and power for two tests are evaluated.

(n, p)	c	N(0,1)		t(3)		$\Gamma(1, 1)$		lnorm(0,1)	
		T_n	T_{np}	T_n	T_{np}	T_n	T_{np}	T_n	T_{np}
(50,11)	0.00	0.099	0.069	0.087	0.075	0.086	0.063	0.101	0.064
	0.25	0.353	0.420	0.423	0.484	0.383	0.434	0.482	0.538
	0.50	0.574	0.661	0.640	0.714	0.602	0.677	0.695	0.758
	0.75	0.751	0.806	0.783	0.849	0.750	0.814	0.804	0.860
	1.00	0.841	0.893	0.869	0.913	0.842	0.897	0.874	0.913
(100,11)	0.00	0.067	0.058	0.060	0.054	0.065	0.055	0.072	0.059
	0.25	0.486	0.662	0.545	0.713	0.537	0.697	0.591	0.742
	0.50	0.783	0.901	0.819	0.906	0.814	0.907	0.832	0.916
	0.75	0.930	0.983	0.918	0.966	0.930	0.980	0.915	0.955
	1.00	0.976	0.996	0.959	0.983	0.977	0.994	0.954	0.972
(100,49)	0.00	0.236	0.066	0.212	0.066	0.218	0.061	0.243	0.065
	0.25	0.543	0.475	0.540	0.506	0.526	0.480	0.658	0.612
	0.50	0.744	0.745	0.781	0.791	0.747	0.750	0.835	0.815
	0.75	0.885	0.892	0.890	0.911	0.875	0.891	0.913	0.913
	1.00	0.948	0.964	0.937	0.956	0.938	0.955	0.947	0.956

To more effectively verify the accuracy of the asymptotic theory underlying our proposed test statistic, Table 5 provides the mean and standard deviation (sd) of the test statistic under different scenarios. From Table 5, it is observed that when $c = 0$, the mean of our proposed test statistic fluctuates around zero, and the standard deviation fluctuates around one. This aligns with the theoretical expectations. As c increases, the mean of the test statistic moves further away from zero, and the standard deviation moves further away from one, indicating a departure from the null hypothesis.

Table 5. Mean and standard deviations of our proposed test statistics across different scenarios.

(n, p)	c	N(0,1)		t(3)		$\Gamma(1, 1)$		lnorm(0,1)	
		mean	sd	mean	sd	mean	sd	mean	sd
(50,11)	0.00	-0.0333	0.9374	-0.0061	0.9418	-0.0043	0.9994	-0.0247	0.9825
	0.25	1.7738	2.2073	1.3510	1.9616	1.2082	1.8621	1.6662	2.2812
	0.50	3.3189	2.7381	2.6000	2.5333	2.3050	2.4013	3.1053	3.0144
	0.75	4.7067	3.0726	3.7388	2.9395	3.3175	2.7857	4.3626	3.5327
	1.00	5.9521	3.2898	4.7717	3.2447	4.2325	3.0603	5.4780	3.9246
(100,11)	0.00	-0.0200	0.9308	-0.0064	0.9198	-0.0638	0.9361	-0.0263	0.9369
	0.25	3.7480	3.0049	2.6939	2.5895	2.3069	2.4523	3.0643	3.0377
	0.50	7.0041	3.8159	5.1654	3.4800	4.5174	3.2602	5.7955	4.2112
	0.75	9.8990	4.2990	7.4175	4.1336	6.5733	3.8412	8.2343	5.0734
	1.00	12.5121	4.6155	9.4800	4.6283	8.4585	4.2903	10.4237	5.7426
(100,49)	0.00	-0.0344	0.9647	-0.0853	0.9221	-0.0634	0.9468	-0.0322	0.8838
	0.25	1.3974	1.6203	2.4578	2.0814	2.2854	2.0247	2.7763	2.3084
	0.50	2.6816	2.0302	4.6292	2.7638	4.2755	2.6032	5.0716	3.1856
	0.75	3.8472	2.3259	6.4930	3.2520	6.0145	2.9903	6.9775	3.7744
	1.00	4.9262	2.5550	8.1087	3.5979	7.5483	3.2715	8.6025	4.2086
(100,201)	0.00	-0.0284	0.9996	-0.0464	0.9366	-0.0799	0.9973	-0.1059	0.9523
	0.25	0.5965	1.1769	0.9410	1.2687	0.8517	1.2541	1.0221	1.3276
	0.50	1.1730	1.3154	1.8179	1.5234	1.6749	1.4624	1.9941	1.6616
	0.75	1.7107	1.4365	2.5847	1.7106	2.4135	1.6295	2.8149	1.9165
	1.00	2.2170	1.5368	3.2742	1.8666	3.0751	1.7569	3.5290	2.1234
(200,365)	0.00	-0.1197	0.9914	-0.1129	0.9450	-0.0892	0.9681	-0.1104	0.9420
	0.25	0.8032	1.2129	3.3462	1.8678	3.0628	1.6937	3.3950	1.9900
	0.50	1.6756	1.3740	5.9872	2.3996	5.5453	2.0681	5.9918	2.6471
	0.75	2.4867	1.5036	8.0726	2.7188	7.5441	2.3124	8.0182	3.0317
	1.00	3.2467	1.6108	9.7640	2.9315	9.1963	2.4703	9.6570	3.2855

Furthermore, to verify the asymptotic theory of our proposed test, we consider the case where $(n, p) = (200, 365)$. Figures 1 and 2 draw the null distributions and the q-q plots T_{np} , corresponding to $g(u) = 2u$ and $g(u) = 2 + \sin(2\pi u)$, respectively. The null distributions are represented by the dashed lines, while the solid lines are density function curves of standard normal distributions.

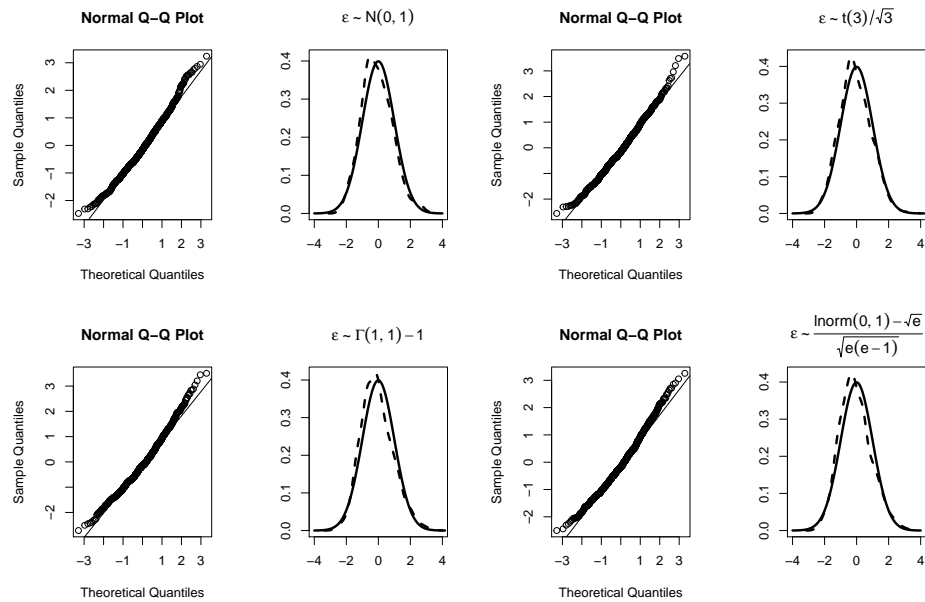


Figure 1. The null distributions and q-q plots of our proposed test when $g(u) = 2u$.

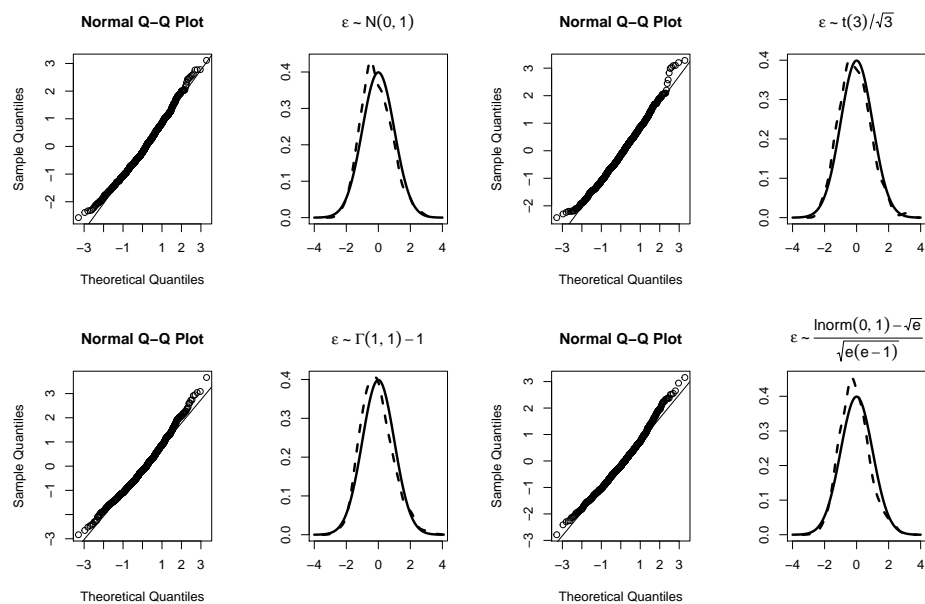


Figure 2. The null distributions and the q-q plots of our proposed test when $g(u) = 2 + \sin(2\pi u)$.

For different n, p , Figures 3 and 4, respectively, show the empirical power functions of the proposed test statistics. These figures are presented for four different error distribution functions. The function $g(\cdot)$ is linear in Figure 3 and trigonometric in Figure 4. When $(n, p) = (200, 201), (100, 201), (200, 365)$, the empirical power functions of the proposed test are represented by solid lines, dashed lines, and dotted lines, respectively. From

Figures 3 and 4, it can be seen that the power increases rapidly as long as c increases slightly. The test's power is positively related to the sample size n and inversely related to the magnitude of p . The proposed test is stable under different error distributions. These are consistent with the conclusions in Tables 3 and 4.

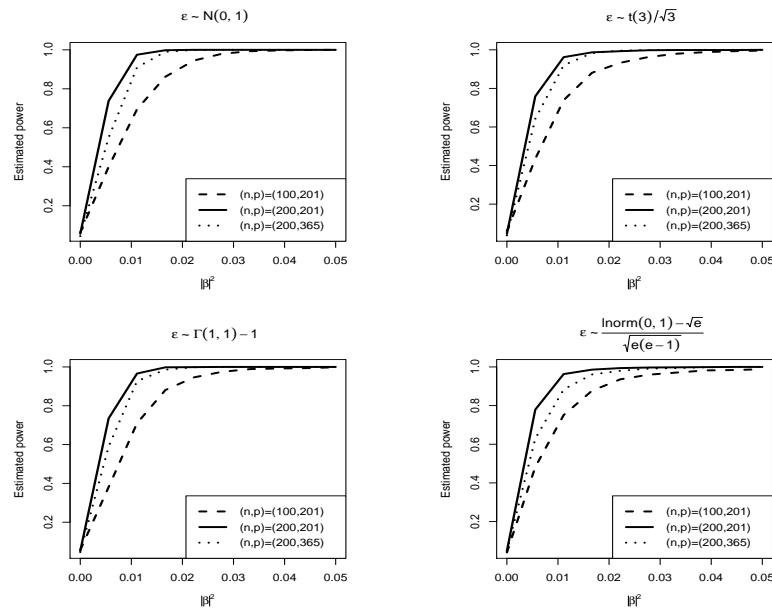


Figure 3. Empirical power functions of our proposed test when $g(u) = 2u$.

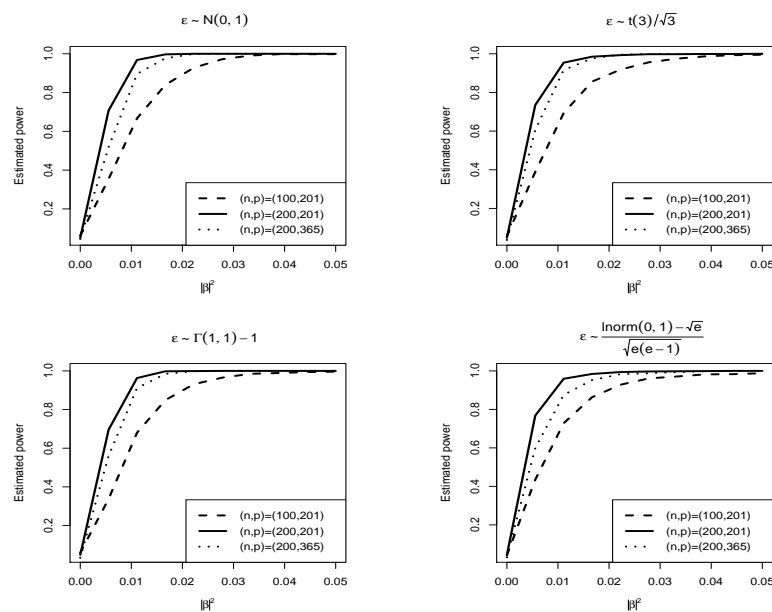


Figure 4. Empirical power functions of our proposed test when $g(u) = 2 + \sin(2\pi u)$.

It is worth noting that, theoretically, a kernel function $K(\cdot)$ is sufficient if it satisfies the conditions of symmetry and Lipschitz continuity. In practical applications, however, the choice of kernel function should be based on the characteristics and requirements of the data. For instance, the Epanechnikov kernel is more suitable for bounded data, while the Gaussian kernel is better suited for data with long tails. In this simulation study, according to the given data setting, the Epanechnikov kernel was chosen. To compare the effects of the two kernels, we replaced the Epanechnikov kernel used to generate Figure 4 with a

Gaussian kernel to produce Figure 5. From Figures 4 and 5, it can be observed that the impact of the two kernels on the test is relatively minor.

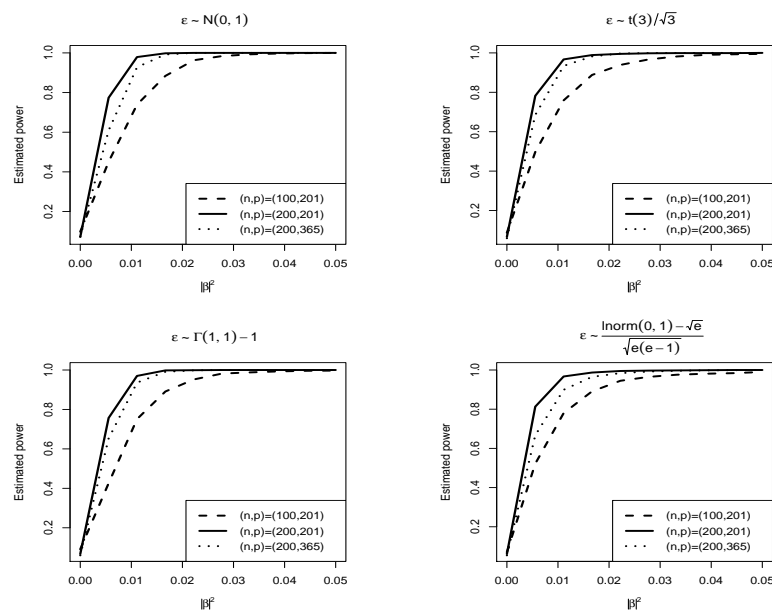


Figure 5. Empirical power functions of our proposed test with the Gaussian kernel when $g(u) = 2 + \sin(2\pi u)$.

The numerical simulations show that our proposed test performs well for the data types. However, with larger sample sizes, the numerical simulations in this paper require considerable computational time, which is a limitation of the proposed test statistic. Additionally, its performance on the datasets that violate the assumptions (C1–C3) and (C7), such as when the real data are not in a Sobolev ellipsoid of order two, remains to be seen.

5. Application

This section applies the proposed test to the spectral data, which has been described and analyzed in the literature (see [25,26]). This dataset can be obtained on the following platforms: <http://lib.stat.cmu.edu/datasets/tecat> (accessed on 16 July 2024). Each meat sample is characterized by a 100-channel spectrum of absorbance, along with the moisture (water), fat, and protein contents. The absorbance is calculated as the negative logarithm base 10 of the transmittance, as measured by the spectrometer. The three contents, measured in percent, are determined by analytic chemistry. The dataset comprises 240 samples, partitioned into 5 subsets for the validation of models and extrapolation studies. In this section, we utilize a total of 215 samples, which include both training and test samples drawn from the 5 subsets. The spectral measurement data consist of curves, represented by $X_i(\cdot)$, corresponding to absorbance values recorded at 100 equally spaced wavelengths from 850 nm to 1050 nm. Let Y_i represent the fat content as the response variable, Z_i represent the protein content, and U_i represent the moisture content. Similar to [27], the following two models will be used to assume the relationship between them:

$$Y_i = \int_{850}^{1050} \alpha(t)X_i(t)dt + g(Z_i) + \varepsilon_i, \tag{10}$$

$$Y_i = \int_{850}^{1050} \alpha(t)X_i(t)dt + g(U_i) + \varepsilon_i. \tag{11}$$

The present investigation primarily focuses on the test in models (10) and (11): $\alpha(t) = 0$. The number of basis functions used for fitting function curves p is selected as 129. Figure 6 shows the estimation of slope function $\alpha(t)$ in models (10) and (11).

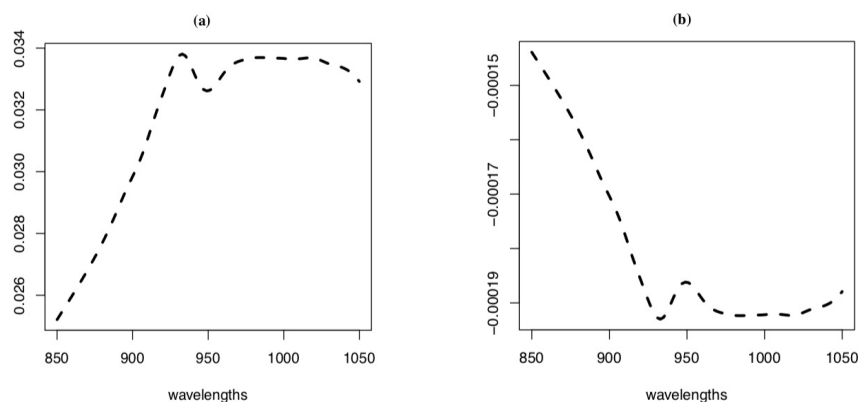


Figure 6. (a) The estimator of the slope function in model (10); (b) the estimator of the slope function in model (11).

The calculation results are as follows: (i) For model (10), the value of the statistic is $T_{np} = 31.186$; p -value is 0. (ii) For model (11), the value of the statistic is $T_{np} = -0.867$; p -values are 0.386. From this, we can see that the model test (10) is significant, while the model test (11) is not significant. This result can also be reflected in Figure 6. It is obvious that the estimated value of $\alpha(t)$ on the right side of Figure 6 is much smaller than that on the left side.

6. Conclusions

To test (2), this paper first provides a pseudo-estimate of the non-parametric function $g(u)$ using kernel methods with a fixed coefficient function $\beta(t)$. The pseudo-estimate is then substituted into the model, converting the original model into a linear one. This allows for the construction of the second-order U-test statistic employed in this paper, utilizing the corresponding testing methods from functional linear models. The proposed test does not require estimating the covariance operator of the predictor function. It follows a normal distribution asymptotically under both a null hypothesis and a local alternative. Moreover, numerical simulations show that our proposed test performs better than the test constructed in [18] when functional data cannot be approximated by a few principal components. Finally, the real data are applied to our proposed test to verify its feasibility.

Additionally, the proposed test is adaptable to cases where the response variable is functional, which is a focus for our upcoming research. In the real world, the proposed test requires real data to meet some technical conditions (C1–C3, C7). When data fail to satisfy these conditions, the viability of the test presented in this paper requires further investigation. Therefore, future work may focus on broadening the test's applicability. The calculation of the statistics in numerical simulations requires optimization for efficiency.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/math12162588/s1>.

Author Contributions: Conceptualization, F.Z. and B.Z.; methodology, B.Z.; Validation, B.Z.; formal analysis, F.Z.; investigation, F.Z.; data curation, F.Z.; writing—original draft preparation, F.Z.; writing—review and editing, B.Z.; visualization, B.Z.; supervision, B.Z.; project administration, B.Z.; funding acquisition, B.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (nos. 12271370, 12301349), the Natural Science Foundation of Shanxi Province, China (no. 202203021222009).

Data Availability Statement: Data are contained within the article..

Acknowledgments: The authors are grateful to the editor, associate editor, and referees for their valuable feedback, which has significantly enhanced the quality of this manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Several lemmas are established to facilitate the proofs of Theorems 1 and 2. Without loss of generality, we assume that $\mu_0(t) = 0$ and $E[g(U)] = 0$. Let $C_n = \sqrt{\log(1/h)/(nh)} + h^2$. With reference to the asymptotic theory of non-parametric estimation, the pseudo-estimation of the non-parametric function satisfies $\sup_{u \in \Omega} \|\check{g}(u) - g(u)\| = O_p(C_n)$. Denote

$$Dg(U_i) = g(U_i) - \check{g}(U_i), D\mu_{it} = \mu_{it} - \hat{\mu}_{it},$$

for $i = 1, 2, \dots, n$. Similarly to the lemmas in [21], it is easy to derive the following lemmas.

Lemma A1. *If (C1), (C3), and (C4) hold, it can be demonstrated that for any square matrix, M,*

- (i) $E[Z_1 Z_1^T M Z_1 Z_1^T] = M + M^T + \text{tr}(M) \mathbf{I}_p + \Delta \text{diag}(M);$
- (ii) $E[Z_1 Z_2^T M Z_2 Z_1^T] = \text{tr}(M) \mathbf{I}_p;$
- (iii) $E[\langle (X_1 - \mu_{1t}, \alpha) \langle X_1 - \mu_{1t}, X_2 - \mu_{2t} \rangle \langle X_2 - \mu_{2t}, \alpha \rangle \rangle^2] = o(\text{tr}(\Sigma_*^2)).$

Lemma A2. *Given that conditions (C1–C3) and (C5–C9) are satisfied, the following results are obtained.*

- (i) $E[\langle \check{X}_1, \check{X}_2 \rangle^4] = o(n \text{tr}^2(\Sigma_*^2));$
- (ii) $E[\langle \mathcal{C}^*(\check{X}_1), \check{X}_1 \rangle^2] = o(n \text{tr}^2(\Sigma_*^2)).$

Lemma A3. *If (C1–C9) hold, then we can obtain the following:*

- (i) $E[\langle \check{X}_1, \check{X}_1 \rangle] = O(\text{tr}(\Sigma_*));$
- (ii) $E[\check{Y}_1^2] = O(1);$
- (iii) $E[\langle \check{X}_{12}, \check{X}_{12} \rangle] = O(\text{tr}(\Sigma_*)/n);$
- (iv) $E[\check{Y}_{12}^2] = O(C_n^2);$
- (v) $E[\langle \check{X}_{12}, \check{X}_{12} \rangle \check{Y}_{12}^2] = O(\text{tr}(\Sigma_*)/n^2),$

where $\check{X}_{ij}(t), \check{Y}_{ij}$ represent the sample means of $\check{X}(t)$ and \check{Y} without i th and j th samples, for $i, j = 1, 2, \dots, n$. That is

$$\check{X}_{ij}(t) = \frac{1}{n-2} \sum_{k \neq i, j} \check{X}_k(t), \quad \check{Y}_{ij} = \frac{1}{n-2} \sum_{k \neq i, j} \check{Y}_k.$$

Proof of Theorem 1. Rewrite

$$\frac{n}{n-2}\Delta_{ij}(\check{X}) \triangleq P_{ij}^{(1)} + P_{ij}^{(2)} + P_{ij}^{(3)} + P_{ij}^{(4)}, \quad \frac{n}{n-2}\Delta_{ij}(\check{Y}) \triangleq L_{ij}^{(1)} + L_{ij}^{(2)} + L_{ij}^{(3)} + L_{ij}^{(4)},$$

where

$$\begin{aligned} P_{ij}^{(1)} &= \left(1 - \frac{1}{n}\right) \langle \check{X}_i, \check{X}_j \rangle, & P_{ij}^{(2)} &= -\frac{1}{2n} (\langle \check{X}_i, \check{X}_i \rangle + \langle \check{X}_j, \check{X}_j \rangle - 2E[\langle \check{X}_1, \check{X}_1 \rangle]), \\ P_{ij}^{(3)} &= -\left(1 - \frac{2}{n}\right) \langle \check{X}_i + \check{X}_j, \check{X}_{ij} \rangle, & P_{ij}^{(4)} &= \left(1 - \frac{2}{n}\right) \left(\langle \check{X}_{ij}, \check{X}_{ij} \rangle - \frac{E[\langle \check{X}_1, \check{X}_1 \rangle]}{n-2} \right), \\ L_{ij}^{(1)} &= \left(1 - \frac{1}{n}\right) \check{Y}_i \check{Y}_j, & L_{ij}^{(2)} &= -\frac{1}{2n} (\check{Y}_i^2 + \check{Y}_j^2 - 2E[\check{Y}_1^2]), \\ L_{ij}^{(3)} &= -\left(1 - \frac{2}{n}\right) (\check{Y}_i + \check{Y}_j) \check{Y}_{ij}, & L_{ij}^{(4)} &= \left(1 - \frac{2}{n}\right) \left(\check{Y}_{ij}^2 - \frac{E(\check{Y}_1^2)}{n-2} \right), \end{aligned}$$

then the expectation of test statistic Tnp is as follows:

$$E[Tnp] = \sum_{j < i} \sum_{k=1}^4 E[P_{ij}^{(l)} L_{ij}^{(k)}].$$

To prove the conclusion (i) in Theorem 1, it needs to be calculated one by one for $(l, k), l, k = 1, 2, 3, 4$. Because of the similarity to calculations in different cases of (l, k) , here, we mainly consider the case where $(l, k) = (1, 1)$,

$$E[P_{ij}^{(1)} L_{ij}^{(1)}] \triangleq G_1^{(1,1)} + G_2^{(1,1)} + G_3^{(1,1)} + G_4^{(1,1)} + G_5^{(1,1)} + G_6^{(1,1)},$$

where

$$\begin{aligned} G_1^{(1,1)} &= \frac{(n-1)^2}{n^2} E[\langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle Dg(U_1) Dg(U_2)], \\ G_2^{(1,1)} &= \frac{(n-1)^2}{n^2} E[\langle X_1 - \hat{\mu}_{1t}, \alpha \rangle \langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle \langle X_2 - \hat{\mu}_{2t}, \alpha \rangle], \\ G_3^{(1,1)} &= \frac{(n-1)^2}{n^2} E[\langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle \varepsilon_1 \varepsilon_2], \\ G_4^{(1,1)} &= \frac{2(n-1)^2}{n^2} E[\langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle \langle X_2 - \hat{\mu}_{2t}, \alpha \rangle Dg(U_1)], \\ G_5^{(1,1)} &= \frac{2(n-1)^2}{n^2} E[\langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle Dg(U_1) \varepsilon_2], \\ G_6^{(1,1)} &= \frac{2(n-1)^2}{n^2} E[\langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle \langle X_1 - \hat{\mu}_{1t}, \alpha \rangle \varepsilon_2]. \end{aligned}$$

For the above six items, we will analyze them one by one. Firstly, we consider the first term. We have the following:

$$\begin{aligned} &E[\langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle Dg(U_1) Dg(U_2)] \\ &= 2E[\langle D\mu_{1t}, X_2 - \mu_{2t} \rangle Dg(U_1) Dg(U_2)] + E[\langle D\mu_{1t}, D\mu_{2t} \rangle Dg(U_1) Dg(U_2)] \\ &= O\left(W_{12} W_{21} \left(\check{\xi}_2^T \check{\xi}_2 - \mu^T(U_2) \check{\xi}_2\right) g^2(U_1)\right) \\ &= O\left(\text{tr}(\Sigma_*) / n^2 h\right), \end{aligned}$$

then $G_1^{(1,1)} = o\left(\sqrt{\text{tr}(\Sigma_*)} / n\right)$ holds. For the second term, we have the following:

$$\begin{aligned}
 & E[\langle X_1 - \hat{\mu}_{1t}, \alpha \rangle \langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle \langle X_2 - \hat{\mu}_{2t}, \alpha \rangle] \\
 = & E[\langle X_1 - \mu_{1t}, \alpha \rangle \langle X_1 - \mu_{1t}, X_2 - \mu_{2t} \rangle \langle X_2 - \mu_{2t}, \alpha \rangle] \\
 & + 2E[\langle D\mu_{1t}, \alpha \rangle \langle D\mu_{1t}, D\mu_{2t} \rangle \langle D\mu_{2t}, \alpha \rangle] \\
 & + 2E[\langle X_1 - \mu_{1t}, \alpha \rangle \langle D\mu_{1t}, D\mu_{2t} \rangle \langle D\mu_{2t}, \alpha \rangle] \\
 & + 2E[\langle X_1 - \mu_{1t}, \alpha \rangle \langle X_2 - \mu_{2t}, D\mu_{1t} \rangle \langle D\mu_{2t}, \alpha \rangle] \\
 & + 2E[\langle X_1 - \mu_{1t}, \alpha \rangle \langle X_1 - \mu_{1t}, D\mu_{2t} \rangle \langle D\mu_{2t}, \alpha \rangle] \\
 & + 2E[\langle D\mu_{1t}, \alpha \rangle \langle X_1 - \mu_{1t}, D\mu_{2t} \rangle \langle D\mu_{2t}, \alpha \rangle] \\
 & + 2E[\langle X_1 - \mu_{1t}, \alpha \rangle \langle D\mu_{1t}, D\mu_{2t} \rangle \langle X_2 - \mu_{2t}, \alpha \rangle] \\
 = & \|C^*(\alpha)\|^2 + O\left(2\beta \Sigma_*^2 \beta / nh + \beta^T \Sigma_* \beta \text{tr}(\Sigma_*) / n^2\right).
 \end{aligned}$$

Combined with (C1), (C3), and (C9), $G_2^{(1,1)} = \|C^*(\alpha)\|^2 + o\left(\sqrt{\text{tr}(\Sigma_*^2)/n}\right)$ holds. The error term ε_i with a mean of zero is independent of the predictor; hence, it is easy to see that both the third term $G_3^{(1,1)}$ and the sixth term $G_6^{(1,1)}$ are zero. For the other two cross terms, $G_4^{(1,1)}$ and $G_5^{(1,1)}$, we need to prove that they are high-order infinitesimals of $\sqrt{\text{tr}(\Sigma_*^2)/n}$. In fact,

$$\begin{aligned}
 & E[\langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle \langle X_2 - \hat{\mu}_{2t}, \alpha \rangle Dg(U_1)] \\
 = & E[\langle X_1 - \mu_{1t}, \alpha \rangle \langle X_1 - \mu_{1t}, D\mu_{2t} \rangle Dg(U_2)] \\
 & + E[\langle D\mu_{1t}, \alpha \rangle \langle D\mu_{1t}, X_2 - \mu_{2t} \rangle Dg(U_2)] \\
 & + E[\langle D\mu_{1t}, \alpha \rangle \langle X_1 - \mu_{1t}, D\mu_{2t} \rangle Dg(U_2)] \\
 & + E[\langle X_1 - \mu_{1t}, \alpha \rangle \langle D\mu_{1t}, D\mu_{2t} \rangle Dg(U_2)] \\
 & + E[\langle D\mu_{1t}, \alpha \rangle \langle D\mu_{1t}, D\mu_{2t} \rangle Dg(U_2)] \\
 = & O\left(nE[W_{23}^2 \beta^T (\xi_1 \xi_1^T - \xi_1 \mu^T(U_1)) \xi_3 g(U_3)]\right) \\
 = & O\left(E[\beta^T \Sigma_* (U_1) \mu(U_2) g(U_2) f^{-1}(U_2)] / nh\right) \\
 = & o\left(\sqrt{\text{tr}(\Sigma_*^2)/n}\right).
 \end{aligned}$$

Finally, for $G_5^{(1,1)}$, we have the following:

$$\begin{aligned}
 & E[\langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle Dg(U_1) \varepsilon_2] \\
 = & E[\langle X_1 - \mu_{1t} + D\mu_{1t}, X_2 - \mu_{2t} + D\mu_{2t} \rangle (-W_{12} \varepsilon_2^2)] \\
 = & E[(\xi_1 - \mu(U_1))^T \xi_1 W_{21} W_{12} \sigma^2] + E[W_{12}^2 \xi_2^T (\xi_2 - \mu(U_1)) \sigma^2] \\
 = & O\left(\text{tr}(\Sigma_*) / n^2 h\right).
 \end{aligned}$$

Using (C3) and the following fact $\text{tr}^2(\Sigma_*) \leq p \text{tr}(\Sigma_*^2)$, we obtain $\text{tr}(\Sigma_*) / \sqrt{\text{tr}(\Sigma_*^2)} = o(nh)$, i.e., $A_5^{(1,1)} = o\left(\sqrt{\text{tr}(\Sigma_*^2)/n}\right)$. Then, the following can be seen:

$$E\left[P_{ij}^{(1)} L_{ij}^{(1)}\right] = \|C^*(\alpha)\|^2 + o\left(\sqrt{\text{tr}(\Sigma_*^2)/n}\right).$$

The conclusion (i) of Theorem 1 follows from the calculation of $E\left[P_{ij}^{(1)} L_{ij}^{(1)}\right]$ and the proof of Theorem 1 in [21]. Conclusion (ii) is addressed in Theorem 2's proof; here we omit it. \square

Proof of Theorem 2. By Theorem 1, we have the following:

$$\frac{n(E(T_{np}) - \|C^*(\alpha)\|^2)}{\sigma^2 \sqrt{2\text{tr}(\Sigma_*^2)}} = o(1),$$

then we only need to prove the following:

$$\frac{n(T_{np} - ET_{np})}{\sigma^2 \sqrt{2\text{tr}(\Sigma_*^2)}} \xrightarrow{D} N(0, 1). \tag{A1}$$

We denote $T_{np}^{(k,l)} = n \binom{n}{2}^{-1} \sum_{i>j} (P_{ij}^{(k)} L_{ij}^{(l)} - E(P_{ij}^{(k)} L_{ij}^{(l)}))$ with $k, l = 1, 2, 3, 4$. The subsequent result is established:

$$n(T_{np} - ET_{np}) = \sum_{k=1}^4 \sum_{l=1}^4 T_{np}^{(k,l)}.$$

In order to derive the asymptotic properties of the above equation, we will find the asymptotic order of all terms $T_{np}^{(k,l)}$. These items are divided into the following two groups according to the treatment methods.

- Group 1: $(k, l) = (1, 1), (1, 2), (1, 3), (1, 4), (2, 1), (2, 3), (3, 1), (3, 3), (4, 1), (4, 3)$.
- Group 2: $(k, l) = (2, 2), (2, 4), (3, 2), (3, 4), (4, 2), (4, 4)$.

Since the methods are similar, the cases of $(k, l) = (1, 1)$ and $(k, l) = (2, 2)$ will be considered, respectively, in detail in each group. Firstly, for $T_{np}^{(1,1)}$, we can rewrite the following:

$$T_{np}^{(1,1)} \triangleq T_{np,1}^{(1,1)} + T_{np,2}^{(1,1)} + T_{np,3}^{(1,1)} + T_{np,4}^{(1,1)} + T_{np,5}^{(1,1)} + T_{np,6}^{(1,1)} + T_{np,7}^{(1,1)} + T_{np,8}^{(1,1)} + T_{np,9}^{(1,1)}, \tag{A2}$$

where

$$\begin{aligned} T_{np,1}^{(1,1)} &= \frac{2(n-1)}{n^2} \sum_{j<i} \left\{ \langle X_i - \hat{\mu}_{it}, X_j - \hat{\mu}_{jt} \rangle Dg(U_i) Dg(U_j) - E[\langle X_i - \hat{\mu}_{it}, X_j - \hat{\mu}_{jt} \rangle Dg(U_i) Dg(U_j)] \right\}, \\ T_{np,2}^{(1,1)} &= \frac{2(n-1)}{n^2} \sum_{j<i} \left\{ \langle X_i - \hat{\mu}_{it}, X_j - \hat{\mu}_{jt} \rangle \langle X_j - \hat{\mu}_{jt}, \alpha \rangle Dg(U_i) - E[\langle X_i - \hat{\mu}_{it}, X_j - \hat{\mu}_{jt} \rangle \langle X_j - \hat{\mu}_{jt}, \alpha \rangle Dg(U_i)] \right\}, \\ T_{np,3}^{(1,1)} &= \frac{2(n-1)}{n^2} \sum_{j<i} \left\{ \langle X_i - \hat{\mu}_{it}, X_j - \hat{\mu}_{jt} \rangle \langle X_i - \hat{\mu}_{it}, \alpha \rangle Dg(U_j) - E[\langle X_i - \hat{\mu}_{it}, X_j - \hat{\mu}_{jt} \rangle \langle X_i - \hat{\mu}_{it}, \alpha \rangle Dg(U_j)] \right\}, \\ T_{np,4}^{(1,1)} &= \frac{2(n-1)}{n^2} \sum_{j<i} \left\{ \langle X_i - \hat{\mu}_{it}, X_j - \hat{\mu}_{jt} \rangle Dg(U_i) \varepsilon_j - E[\langle X_i - \hat{\mu}_{it}, X_j - \hat{\mu}_{jt} \rangle Dg(U_i) \varepsilon_j] \right\}, \end{aligned}$$

$$\begin{aligned}
 T_{np,5}^{(1,1)} &= \frac{2(n-1)}{n^2} \sum_{j<i} \left\{ \langle X_i - \hat{\mu}_{it}, X_j - \hat{\mu}_{jt} \rangle Dg(U_j) \varepsilon_i - E[\langle X_i - \hat{\mu}_{it}, X_j - \hat{\mu}_{jt} \rangle Dg(U_j) \varepsilon_i] \right\}, \\
 T_{np,6}^{(1,1)} &= \frac{2(n-1)}{n^2} \sum_{j<i} \left\{ \langle X_i - \hat{\mu}_{it}, X_j - \hat{\mu}_{jt} \rangle \langle X_j - \hat{\mu}_{jt}, \alpha \rangle \varepsilon_i \right\}, \\
 T_{np,7}^{(1,1)} &= \frac{2(n-1)}{n^2} \sum_{j<i} \left\{ \langle X_i - \hat{\mu}_{it}, X_j - \hat{\mu}_{jt} \rangle \langle X_i - \hat{\mu}_{it}, \alpha \rangle \varepsilon_j \right\}, \\
 T_{np,8}^{(1,1)} &= \frac{2(n-1)}{n^2} \sum_{j<i} \left\{ \langle X_i - \hat{\mu}_{it}, \alpha \rangle \langle X_i - \hat{\mu}_{it}, X_j - \hat{\mu}_{jt} \rangle \langle X_j - \hat{\mu}_{jt}, \alpha \rangle - E[\langle X_i - \hat{\mu}_{it}, \alpha \rangle \langle X_i - \hat{\mu}_{it}, X_j - \hat{\mu}_{jt} \rangle \langle X_j - \hat{\mu}_{jt}, \alpha \rangle] \right\}, \\
 T_{np,9}^{(1,1)} &= \frac{2(n-1)}{n^2} \sum_{j<i} \langle X_i - \hat{\mu}_{it}, X_j - \hat{\mu}_{jt} \rangle \varepsilon_i \varepsilon_j.
 \end{aligned}$$

To prove (A1), we shall prove the following:

$$\frac{T_{np}^{(1,1)} - ET_{np}^{(1,1)}}{\sigma^2 \sqrt{2\text{tr}(\Sigma_*^2)}} = \frac{T_{np,91}^{(1,1)}}{\sigma^2 \sqrt{2\text{tr}(\Sigma_*^2)}} + o_p(1), \tag{A3}$$

where $T_{np,91}^{(1,1)} = \frac{2(n-1)}{n^2} \sum_{j<i} \langle X_i - \mu_{it}, X_j - \mu_{jt} \rangle \varepsilon_i \varepsilon_j$.

It is easy to see that the means of nine items in the right equation of (A2) are all zero. To calculate their asymptotic order, it is necessary to prove their second moment. Due to the similarity in calculating the first eight items, we use the first item $T_{np,1}^{(1,1)}$ as an example to consider.

$$E\left[(T_{np,1}^{(1,1)})^2\right] = \frac{n^2}{\binom{n}{2}^2} \sum_{i=2}^n E\left[Q_{i,1}^{(1,1)} Q_{i,1}^{(1,1)}\right] + \frac{n^2}{\binom{n}{2}^2} \sum_{i=2}^n \sum_{j \neq i} E\left[Q_{i,1}^{(1,1)} Q_{j,1}^{(1,1)}\right] + o(\text{tr}(\Sigma_*^2)),$$

where

$$Q_{i,1}^{(1,1)} = \sum_{j=1}^{i-1} \langle X_i - \hat{\mu}_{it}, X_j - \hat{\mu}_{jt} \rangle Dg(U_i) Dg(U_j).$$

For $i \neq j$, let us calculate $E\left[Q_{i,1}^{(1,1)} Q_{i,1}^{(1,1)}\right]$ and $E\left[Q_{i,1}^{(1,1)} Q_{j,1}^{(1,1)}\right]$.

$$\begin{aligned}
 &E\left[Q_{i,1}^{(1,1)} Q_{i,1}^{(1,1)}\right] \\
 &= (i-1)E\left[\langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle \langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle Dg(U_1)^2 Dg(U_2)^2\right] \\
 &+ (i-1)(i-2)E\left[\langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle \langle X_1 - \hat{\mu}_{1t}, X_3 - \hat{\mu}_{3t} \rangle Dg(U_1)^2 Dg(U_2) Dg(U_3)\right] \\
 &\triangleq (i-1)B_{11}^{(1,1)} + (i-1)(i-2)B_{12}^{(1,1)}, \\
 &E\left[Q_{i,1}^{(1,1)} Q_{j,1}^{(1,1)}\right] \\
 &= \Xi_1 E\left[\langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle \langle X_1 - \hat{\mu}_{1t}, X_3 - \hat{\mu}_{3t} \rangle Dg(U_1)^2 Dg(U_2) Dg(U_3)\right]
 \end{aligned}$$

$$\begin{aligned}
 & + \Xi_2 E \left[\langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle \langle X_3 - \hat{\mu}_{3t}, X_4 - \hat{\mu}_{4t} \rangle Dg(U_1) Dg(U_2) Dg(U_3) Dg(U_4) \right] \\
 & \triangleq \Xi_1 B_{12}^{(1,1)} + \Xi_2 B_{13}^{(1,1)},
 \end{aligned}$$

where $\Xi_1 = (i - 1) \wedge (j - 1)$, $\Xi_2 = (i - 1)(j - 1) - (i - 1) \wedge (j - 1)$.

Using the Cauchy–Schwarz inequality and Lemma A.2, we can obtain

$$B_{11}^{(1,1)} = O(C_n^4 \sqrt{n} \text{tr}(\Sigma_*^2)).$$

For $B_{12}^{(1,1)}$ and $B_{13}^{(1,1)}$,

$$\begin{aligned}
 B_{12}^{(1,1)} &= E \left[\langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle \langle X_1 - \hat{\mu}_{1t}, X_3 - \hat{\mu}_{3t} \rangle Dg(U_1)^2 Dg(U_2) Dg(U_3) \right] \\
 &= O \left(E [W_{23} \text{tr}(\Sigma_*(U_1) \Sigma_*(U_3)) Dg(U_1)^2 Dg(U_2) Dg(U_3)] \right. \\
 &\quad + E [W_{13} W_{21} \text{tr}(\Sigma_*(U_1) \Sigma_*(U_3)) Dg(U_1)^2 Dg(U_2) Dg(U_3)] \\
 &\quad + E [W_{12} W_{13} \text{tr}(\Sigma_*(U_1)) \text{tr}(\Sigma_*(U_3)) Dg(U_1)^2 Dg(U_2) Dg(U_3)] \\
 &\quad \left. + E [\mu^T(U_1) \mu(U_2) \mu^T(U_3) \mu(U_1) Dg(U_1)^2 Dg(U_2) Dg(U_3)] \right) \\
 &= O \left(E [\text{tr}(\Sigma_*(U_1) \Sigma_*(U_2)) g^2(U_1) g(U_2)] / n^2 + E^2 [\text{tr}(\Sigma_*(U_1)) g^2(U_1)] / n^3 \right. \\
 &\quad + E [\mu^T(U_2) \mu(U_1) \mu^T(U_1) \mu(U_2) g(U_1)] / n^2 \\
 &\quad \left. + E [\mu^T(U_2) \mu(U_1) \mu^T(U_1) \mu(U_2) g^2(U_1) g^2(U_2)] / n^2 \right), \\
 B_{13}^{(1,1)} &= O \left(E [D\mu^T(U_1) D\mu(U_2) (\mu(U_3) - \hat{\mu}(U_3))^T (\mu(U_4) - \hat{\mu}(U_4)) \right. \\
 &\quad \left. Dg(U_1) Dg(U_2) Dg(U_3) Dg(U_4)] \right) \\
 &= O \left(\mu^T(U_1) \mu(U_2) \mu^T(U_3) \mu(U_4) g(U_1) g(U_2) g(U_3) g(U_4) / n^2 \right).
 \end{aligned}$$

So, we can have $T_{np,1}^{(1,1)} = o_p \left(\sqrt{\text{tr}(\Sigma_*^2)} \right)$. We apply similar methods to the $T_{np,1}^{(1,1)}$, the terms $\{T_{np,k}^{(1,1)}, k = 2, \dots, 7\}$ are all equal to $o_p \left(\sqrt{\text{tr}(\Sigma_*^2)} \right)$. For $T_{np,9}^{(1,1)}$, we rewrite

$$T_{np,9}^{(1,1)} \triangleq T_{np,91}^{(1,1)} + T_{np,92}^{(1,1)} + T_{np,93}^{(1,1)} + T_{np,94}^{(1,1)}$$

where

$$\begin{aligned}
 T_{np,91}^{(1,1)} &= \frac{2(n-1)}{n^2} \sum_{i=2}^n \sum_{j=1}^{i-1} \left\{ \langle X_i - \mu_{it}, X_j - \mu_{jt} \rangle \varepsilon_i \varepsilon_j \right\}, \\
 T_{np,92}^{(1,1)} &= \frac{2(n-1)}{n^2} \sum_{i=2}^n \sum_{j=1}^{i-1} \left\{ \langle X_i - \mu_{it}, D\mu_{jt} \rangle \varepsilon_i \varepsilon_j \right\}, \\
 T_{np,93}^{(1,1)} &= \frac{2(n-1)}{n^2} \sum_{i=2}^n \sum_{j=1}^{i-1} \left\{ \langle X_j - \mu_{jt}, D\mu_{it} \rangle \varepsilon_i \varepsilon_j \right\}, \\
 T_{np,94}^{(1,1)} &= \frac{2(n-1)}{n^2} \sum_{i=2}^n \sum_{j=1}^{i-1} \left\{ \langle D\mu_{it}, D\mu_{jt} \rangle \varepsilon_i \varepsilon_j \right\}.
 \end{aligned}$$

Since the means of the above four formulas are zero, in order to prove that (A3) is true, it is necessary to verify that the second moments of $T_{np,9k}^{(1,1)}$, $k = 2, 3, 4$ are higher-order infinitesimals of the quantity $\text{tr}(\Sigma_*^2)$. In fact,

$$\begin{aligned} E[T_{np,92}^{(1,1)}]^2 &= E[T_{np,93}^{(1,1)}]^2 \\ &= \frac{4(n-1)^2\sigma^4}{n^4} \sum_{i=2}^n \left\{ (i-1)E[\langle X_1 - \mu_{1t}, D\mu_{2t} \rangle]^2 \right\} \\ &= O(E[\langle X_1 - \mu_{1t}, D\mu_{2t} \rangle \langle X_1 - \mu_{1t}, D\mu_{2t} \rangle]) \\ &= O(nE[W_{23}^2 \xi_3^T \Sigma_* (U_1) \xi_3]) \\ &= o(\text{tr}(\Sigma_*^2)), \end{aligned}$$

$$\begin{aligned} E[T_{np,94}^{(1,1)}]^2 &= O(E[\langle D\mu_{it}, D\mu_{jt} \rangle \langle D\mu_{it}, D\mu_{jt} \rangle]) \\ &= O\left(E[\text{tr}(\Sigma_*(U_3)\mu(U_1)\mu^T(U_1)f^{-1}(U_3))]/nh\right) + O(\text{tr}^2(\Sigma_*)/n^2h) \\ &= o(\text{tr}(\Sigma_*^2)). \end{aligned}$$

Then Equation (A3) holds. Similarly, for Group 2, i.e., when $(k, l) = (2, 2), (2, 4), (3, 2), (3, 4), (4, 2), (4, 4)$, there is a similar proof process for the asymptotic behavior of each item in the group. Here, we only consider $T_{np}^{(2,2)}$. By careful calculation, we have the following:

$$\begin{aligned} &E[(\langle \check{X}_1, \check{X}_1 \rangle + \langle \check{X}_2, \check{X}_2 \rangle - 2E(\langle \check{X}_1, \check{X}_1 \rangle))^2] \\ &= O(E[\langle X_1 - \mu_{1t}, X_1 - \mu_{1t} \rangle^2] + E^2[\langle X_1 - \mu_{1t}, X_1 - \mu_{1t} \rangle] \\ &\quad + E[\langle X_1 - \mu_{1t}, X_1 - \mu_{1t} \rangle \langle X_2 - \mu_{2t}, X_2 - \mu_{2t} \rangle]) \\ &= O(E[(\xi_1 - \mu(U_1))^T (\xi_1 - \mu(U_1)) (\xi_1 - \mu(U_1))^T (\xi_1 - \mu(U_1))] \\ &\quad + E[(\xi_1 - \mu(U_1))^T (\xi_1 - \mu(U_1)) (\xi_2 - \mu(U_2))^T (\xi_2 - \mu(U_2))] \\ &\quad + E^2[(\xi_1 - \mu(U_1))^T (\xi_1 - \mu(U_1))]) \\ &= O(E[2\text{tr}(\Sigma_*^2(U_1)) + \text{tr}^2(\Sigma_*) + \Delta \text{tr}(\text{diag}(\Gamma^T(U_1)\Gamma(U_1))\Gamma^T(U_1)\Gamma(U_1))]). \end{aligned}$$

Using the fact that

$$E[\text{tr}(\text{diag}(\Gamma^T(U_1)\Gamma(U_1))\Gamma^T(U_1)\Gamma(U_1))] = E[\text{tr}(\Sigma_*^2(U_1))] = O(\text{tr}(\Sigma_*^2)),$$

we have the following:

$$E[(\langle \check{X}_1, \check{X}_1 \rangle + \langle \check{X}_2, \check{X}_2 \rangle - 2E(\langle \check{X}_1, \check{X}_1 \rangle))^2] = O(\text{tr}^2(\Sigma_*) + \text{tr}(\Sigma_*^2)). \tag{A4}$$

In addition, by a simple calculation, we have the following:

$$E[(\check{Y}_1^2 + \check{Y}_2^2 - 2E[\check{Y}_1^2])^2] = 2E[(\check{Y}_1^2)^2] + 2E[\check{Y}_1^2\check{Y}_2^2] - 4E^2[\check{Y}_1^2] = O(1). \tag{A5}$$

Combining (A4), (A5) with the Cauchy–Schwarz inequality, we have the following:

$$\begin{aligned} E|T_{np}^{(2,2)}| &\leq \frac{1}{4n} \sqrt{E[(\langle \check{X}_1, \check{X}_1 \rangle + \langle \check{X}_2, \check{X}_2 \rangle - 2E(\langle \check{X}_1, \check{X}_1 \rangle))^2]} \sqrt{E[\check{Y}_1^2 + \check{Y}_2^2 - 2E(\check{Y}_1^2)]^2} \\ &= o\left(\sqrt{\text{tr}(\Sigma_*^2)}\right). \end{aligned}$$

We denote the following: $\check{T}_{np1} \triangleq \binom{n}{2}^{-1/2} \sum_{i=2}^n \sum_{j=1}^{i-1} \langle X_i - \mu_{it}, X_j - \mu_{jt} \rangle \varepsilon_i \varepsilon_j$, by condition (C1), we only need to consider the following: $\check{T}_{np} = \binom{n}{2}^{-1/2} \sum_{i=2}^n \sum_{j=1}^{i-1} (\xi_i - \mu(U_i))^T (\xi_j - \mu(U_j)) \varepsilon_i \varepsilon_j$. Then, by Slutsky’s theorem, if the following conclusion can be obtained, Theorem 2 will be proved.

$$\frac{\check{T}_{np}}{\sqrt{\text{var}(\check{T}_{np})}} \xrightarrow{D} N(0, 1).$$

By some simple calculations, we have $\text{var}(\check{T}_{np}) = \sigma^4 \text{tr}(\Sigma_*^2)$. Let $Z_{ni} = \sum_{j=1}^{i-1} \langle X_i - \mu_{it}, X_j - \mu_{jt} \rangle \varepsilon_i \varepsilon_j / \sqrt{\binom{n}{2}}$, $v_{ni} = E[Z_{ni}^2 | \mathcal{F}_{i-1}]$, $Xu_i = (\xi_i, U_i, \varepsilon_i)^T$, where $\mathcal{F}_i = \{Xu_1, \dots, Xu_i\}$ is a σ -algebra produced by $\{Xu_i, k = 1, \dots, i\}$, $V_n = \sum_{i=2}^n v_{ni}$. The condition $E[Z_{ni} | \mathcal{F}_i] = 0$ is readily verifiable, and the sequence $\{\sum_{i=2}^j Z_{ni}, \mathcal{F}_j : 2 \leq j \leq n\}$ constitutes a mean-zero martingale. Adherence to the Martingale central limit theorem is ensured by verifying the following conditions:

$$\frac{V_n}{\text{var}(\check{T}_{np})} \xrightarrow{P} 1, \text{ as } n \rightarrow \infty; \tag{A6}$$

$$\sum_{i=2}^n \sigma^{-4} \text{tr}^{-1}(\Sigma_*^2) E\{Z_{ni}^2 I(|Z_{ni}| > \eta \sigma^2 \sqrt{\text{tr}(\Sigma_*^2)}) | \mathcal{F}_{i-1}\} \xrightarrow{P} 0 \text{ for } \forall \eta > 0. \tag{A7}$$

Note that

$$v_{ni} = \frac{\sigma^2}{\binom{n}{2}} \sum_{j=1}^{i-1} \left\{ \varepsilon_j^2 (\xi_j - \mu(U_j))^T \Sigma_* (\xi_j - \mu(U_j)) + 2 \sum_{k<l<i} \varepsilon_k \varepsilon_l (\xi_k - \mu(U_k))^T \Sigma_* (\xi_l - \mu(U_l)) \right\}.$$

Then, we define the following:

$$\frac{V_n}{\text{var}(\check{T}_{np})} \triangleq C_{n1} + C_{n2},$$

where

$$C_{n1} = \frac{1}{\binom{n}{2} \sigma^2 \text{tr}(\Sigma_*^2)} \sum_{j<i} \varepsilon_j^2 (\xi_j - \mu(U_j))^T \Sigma_* (\xi_j - \mu(U_j)),$$

$$C_{n2} = \frac{2}{\binom{n}{2} \sigma^2 \text{tr}(\Sigma_*^2)} \sum_{k<l<i} \varepsilon_k \varepsilon_l (\xi_k - \mu(U_k))^T \Sigma_* (\xi_l - \mu(U_l)).$$

The equality $E[C_{n1}] = 1$ can be readily confirmed, and we have the following:

$$\begin{aligned} \text{var}(C_{n1}) &= E[C_{n1}^2] - 1 \\ &= \frac{1}{n^4 (\sigma^4 \Sigma_*^2)^2} \sum_{i=2}^n \left((i-1) E \left[(\xi_1 - \mu(U_1))^T \Sigma_* (\xi_1 - \mu(U_1)) (\xi_1 - \mu(U_1))^T \Sigma_* (\xi_1 - \mu(U_1)) \sigma^4 \varepsilon_1^4 \right] \right. \\ &\quad \left. + (i-1)(i-2) E \left[(\xi_1 - \mu(U_1))^T \Sigma_* (\xi_1 - \mu(U_1)) (\xi_2 - \mu(U_2))^T \Sigma_* (\xi_2 - \mu(U_2)) \sigma^4 \varepsilon_1^2 \varepsilon_2^2 \right] \right) \\ &\quad + \frac{1}{n^4 (\sigma^4 \Sigma_*^2)^2} \sum_{i=2}^n \sum_{j \neq i} \left((i-1) \wedge (j-1) E \left[(\xi_1 - \mu(U_1))^T \Sigma_* (\xi_1 - \mu(U_1)) (\xi_1 - \mu(U_1))^T \Sigma_* (\xi_1 - \mu(U_1)) \sigma^4 \varepsilon_1^4 \right] \right. \\ &\quad \left. + ((i-1)(j-1) - (i-1) \wedge (j-1)) E \left[(\xi_1 - \mu(U_1))^T \Sigma_* (\xi_1 - \mu(U_1)) (\xi_2 - \mu(U_2))^T \Sigma_* (\xi_2 - \mu(U_2)) \sigma^4 \varepsilon_1^2 \varepsilon_2^2 \right] \right) - 1 \end{aligned}$$

$$\begin{aligned}
 &= \frac{E[\varepsilon_1^4]}{n\sigma^4\text{tr}^2(\Sigma_*^2)}O(E[(\xi_1 - \mu(U_1))^T \Sigma_* (\xi_1 - \mu(U_1))(\xi_1 - \mu(U_1))^T \Sigma_* (\xi_1 - \mu(U_1))]) \\
 &= \frac{E[\varepsilon_1^4]}{n\sigma^4\text{tr}^2(\Sigma_*^2)}O(E[\text{tr}(\Sigma_* (U_1) \Sigma_* \Sigma_* (U_1) \Sigma_*)] + \text{tr}^2(\Sigma_* (U_1) \Sigma_*) \\
 &\quad + \Delta \text{tr}(\text{diag}(\Gamma^T(U_1) \Sigma_* \Gamma(U_1)) \Gamma^T(U_1) \Sigma_* \Gamma(U_1))) \\
 &= \frac{E[\varepsilon_1^4]}{n\sigma^4\text{tr}^2(\Sigma_*^2)}O(\text{tr}(\Sigma_*^4) + \text{tr}^2(\Sigma_*^2)),
 \end{aligned}$$

By (C2), we have $C_{n1} \xrightarrow{P} 1$. Similarly, we can obtain $E[C_{n2}] = 0$, and we have the following:

$$\begin{aligned}
 \text{var}(C_{n2}) &= E[C_{n2}^2] \\
 &= O\left(\frac{2}{\binom{n}{2}^2 \text{tr}^2(\Sigma_*^2)} \sum_{i=2}^n ((i-1)(i-2) E[(\xi_1 - \mu(U_1))^T \Sigma_* (\xi_2 - \mu(U_2))(X_2 - \mu(U_2))^T \Sigma_* (\xi_1 - \mu(U_1))]) + \sum_{i=2}^n \sum_{j \neq i} (i-1) \wedge (j-1) ((i-1) \wedge (j-1) - 1) E[(\xi_1 - \mu(U_1))^T \Sigma_* (\xi_2 - \mu(U_2))(\xi_2 - \mu(U_2))^T \Sigma_* (\xi_1 - \mu(U_1))])\right) \\
 &= O\left(\frac{\text{tr}(\Sigma_*^4)}{\text{tr}^2(\Sigma_*^2)}\right).
 \end{aligned}$$

Combined with $\text{tr}(\Sigma_*^4) = o(\text{tr}^2(\Sigma_*^2))$, we have $C_{n2} \xrightarrow{P} 0$. Thus, Equation (A6) holds. Finally, we only need to prove (A7). Hence, leveraging the law of large numbers and the fact that $E[Z_{ni}^2 I(|Z_{ni}| > \eta \sigma^2 \text{tr}(\Sigma_*^2))] \leq E(Z_{ni}^4 | \mathcal{F}_{i-1}) / (\eta^2 \sigma^4 \text{tr}(\Sigma_*^2))$, it suffices to demonstrate that $\sum_{2 \leq i \leq n} E(Z_{ni}^4) = o(\text{tr}^2(\Sigma_*^2))$. Through straightforward computations, we obtain the following result:

$$\begin{aligned}
 &\sum_{i=2}^n E[Z_{ni}^4] \\
 &= \frac{1}{\binom{n}{2}^2} \sum_{j < i} E[((\xi_i - \mu(U_i))^T (\xi_j - \mu(U_j)))^4 \varepsilon_i^4 \varepsilon_j^4] + \sum_{i=2}^n E[Z_{ni}^4] \\
 &= \frac{1}{\binom{n}{2}^2} \sum_{j < i} \sum_{k \neq j} E[((\xi_i - \mu(U_i))^T (\xi_j - \mu(U_j)))^2 ((\xi_i - \mu(U_i))^T (\xi_k - \mu(U_k)))^2 \varepsilon_i^4 \varepsilon_j^2 \varepsilon_k^2] \\
 &= \frac{E^2(\varepsilon_1^4)}{\binom{n}{2}^2} \sum_{i=2}^n (i-1) E[(\xi_1 - \mu(U_1))^T (\xi_2 - \mu(U_2))]^4 + \frac{3\sigma^4 E(\varepsilon_1^4)}{\binom{n}{2}^2} \sum_{i=2}^n (i-1)(i-2) E[(\xi_1 - \mu(U_1))^T (\xi_2 - \mu(U_2))]^2 ((\xi_1 - \mu(U_1))^T (\xi_3 - \mu(U_3)))^2] \\
 &= O\left(\frac{E[(\xi_1 - \mu(U_1))^T (\xi_2 - \mu(U_2))]^4}{n^2}\right) + O\left(\frac{E[(\xi_1 - \mu(U_1))^T \Sigma_* (\xi_1 - \mu(U_1))]^2}{n}\right)
 \end{aligned}$$

Combining (C2) and Lemma 2, Equation (A7) holds. Thus, the proof of Theorem 2 is completed. □

References

1. Crainiceanu, C.M.; Staicu, A.M.; Di, C.Z. Generalized multilevel functional regression. *J. Am. Stat. Assoc.* **2009**, *104*, 1550–1561. [[CrossRef](#)] [[PubMed](#)]
2. Wang, J.; Zhou, F.; Li, C.; Yin, N.; Liu, H.; Zhuang, B.; Huang, Q.; Wen, Y. Gene Association Analysis of Quantitative Trait Based on Functional Linear Regression Model with Local Sparse Estimator. *Genes* **2023**, *14*, 834. [[CrossRef](#)] [[PubMed](#)]
3. Kokoszka, P.; Miao, H.; Zhang, X. Functional dynamic factor model for intraday price curves. *J. Financ. Econom.* **2014**, *13*, 456–477. [[CrossRef](#)]

4. Rigueira, X.; Araújo, M.; Martínez, J.; García-Nieto, P.J.; Ocarranza, I. Functional Data Analysis for the Detection of Outliers and Study of the Effects of the COVID-19 Pandemic on Air Quality: A Case Study in Gijón, Spain. *Mathematics* **2022**, *10*, 2374. [[CrossRef](#)]
5. Yao, F.; Müller, H.G. Functional quadratic regression. *Biometrika* **2010**, *97*, 49–64. [[CrossRef](#)]
6. Lian, H. Functional partial linear model. *J. Nonparametr. Stat.* **2011**, *23*, 115–128. [[CrossRef](#)]
7. Zhou, J.; Chen, M. Spline estimators for semi-functional linear model. *Stat. Probab. Lett.* **2012**, *82*, 505–513. [[CrossRef](#)]
8. Tang, Q. Estimation for semi-functional linear regression. *Statistics* **2015**, *49*, 1262–1278.
9. Zhang, Y.; Wu, Y. Robust hypothesis testing in functional linear models. *J. Stat. Comput. Simul.* **2023**, *93*, 2563–2581. [[CrossRef](#)]
10. Kokoszka, P.; Maslova, I.; Sojka, J.; Zhu, L. Testing for lack of dependence in the functional linear model. *Can. J. Stat.* **2008**, *36*, 207–222. [[CrossRef](#)]
11. James, G.M.; Wang, J.; Zhu, J. Functional linear regression that's interpretable. *Ann. Stat.* **2009**, *37*, 2083–2108. [[CrossRef](#)]
12. Shin, H. Partial functional linear regression. *J. Stat. Plan. Inference* **2009**, *139*, 3405–3418. [[CrossRef](#)]
13. Yu, P.; Zhang, Z.; Du, J. A test of linearity in partial functional linear regression. *Metrika* **2016**, *79*, 953–969. [[CrossRef](#)]
14. Hu, H.; Zhang, R.; Yu, Z.; Lian, H.; Liu, Y. Estimation and testing for partially functional linear errors-in-variables models. *J. Multivar. Anal.* **2019**, *170*, 296–314.
15. Smaga, Ł. General linear hypothesis testing in functional response model. *Commun. Stat.-Theory Methods* **2019**, *50*, 5068–5083.
16. Zhu, H.; Zhang, R.; Li, H. Estimation on semi-functional linear errors-in-variables models. *Commun. Stat.-Theory Methods* **2019**, *48*, 4380–4393. [[CrossRef](#)]
17. Zhou, J.; Peng, Q. Estimation for functional partial linear models with missing responses. *Stat. Probab. Lett.* **2020**, *156*, 108598. [[CrossRef](#)]
18. Zhao, F.; Zhang, B. Testing linearity in functional partially linear models. *Acta Math. Appl. Sin. Engl. Ser.* **2024**, *40*, 875–886. [[CrossRef](#)]
19. Hu, W.; Lin, N.; Zhang, B. Nonparametric testing of lack of dependence in functional linear models. *PLoS ONE* **2020**, *15*, e0234094. [[CrossRef](#)]
20. Zhao, F.; Lin, N.; Hu, W.; Zhang, B. A faster U-statistic for testing independence in the functional linear models. *J. Stat. Plan. Inference* **2022**, *217*, 188–203. [[CrossRef](#)]
21. Zhao, F.; Lin, N.; Zhang, B. A new test for high-dimensional regression coefficients in partially linear models. *Can. J. Stat.* **2023**, *51*, 5–18. [[CrossRef](#)]
22. Cui, H.; Guo, W.; Zhong, W. Test for high-dimensional regression coefficients using refitted cross-validation variance estimation. *Ann. Stat.* **2018**, *46*, 958–988. [[CrossRef](#)]
23. Zhong, P.; Chen, S. Tests for high-dimensional regression coefficients with factorial designs. *J. Am. Stat. Assoc.* **2011**, *106*, 260–274. [[CrossRef](#)]
24. Chen, S.; Zhang, L.; Zhong, P. Tests for high-dimensional covariance matrices. *J. Am. Stat. Assoc.* **2010**, *105*, 810–819. [[CrossRef](#)]
25. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis: Theory and Practice*; Springer: New York, NY, USA, 2006.
26. Shang, H.L. Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density. *Comput. Stat.* **2014**, *29*, 829–848. [[CrossRef](#)]
27. Yu, P.; Zhang, Z.; Du, J. Estimation in functional partial linear composite quantile regression model. *Chin. J. Appl. Probab. Stat.* **2017**, *33*, 170–190.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.