

Article

Not peer-reviewed version

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

[Fanrong Zhao](#) and [Baoxue Zhang](#) *

Posted Date: 18 July 2024

doi: 10.20944/preprints202407.1547.v1

Keywords: Asymptotic normality; Functional partially linear regression model; Nadaraya-Watson estimate; U-statistic



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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.

Article

A U-Statistic for Testing 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;
² School of Statistics, Capital University of Economics and Business, Beijing 100070, China
* Correspondence: zhangbaoxue@cueb.edu.cn

Abstract: Functional partially linear regression model, contains a functional linear part and a non-parametric part, in which testing the linear relationship between the response and the functional predictor is of fundamental importance. When functional data cannot be approximated with a few principal components, based on a pseudo estimate for the unknown non-parametric component, we develop a U-statistic of order-two in this paper. Under some regularity conditions, we use the martingale central limit theorem to prove that the proposed test statistic is asymptotically normal. The finite sample performance with simulation studies and real data application are assessed to verify the proposed test procedure.

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

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, chemistry (see [1–4]). An important model in functional data analysis is 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 functional partially model, which has been studied in [5]. The functional partially linear regression model is defined as follows:

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

where Y is response variable, $X(\cdot)$ is functional predictor with mean $\mu_0(\cdot)$ 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 . ε is a random error with mean zero and finite variance σ^2 , and is independent of the predictor $X(\cdot)$. When the $g(\cdot)$ is constant, the model (1) becomes a functional linear model. See [6]. When $g(\cdot)$ is parametric linear part, the model (1) becomes a partially functional linear model, which has been studied in [7].

Hypothesis testing plays a critical role in statistical inference. For testing the linear relationship between the response and the functional predictor in functional linear model, functional principal component analysis (FPCA) of the functional predictor $X(\cdot)$ in the literature is a major idea to construct test statistic. See [8]. Taking into account the flexibility of non-parametric functions, [9] introduced the functional partially linear regression model. [10] constructed the estimators of the slope functions based on spline and FPCA respectively, and the estimate of non-parametric components in their papers adopted B-spline. When the predictors are measured with additive error, [11] studied the estimators of slope function and non-parametric function by FPCA and kernel smoothing techniques. [12] established the estimators of slope function, non-parametric component and mean of response variable when existing missing responses at random.

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

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

where $\alpha_0(t)$ is an assigned function. Without loss of generality, let $\alpha_0(t) = 0$. For testing (??) constructed a chi-square test using FPCA when the functional data can be approximated with a few principal components. In particular, when the functional data cannot be approximated with a few principal components, only several researches have considered this situation in functional data analysis. [?] constructed a FLUTE test based on order-four U-statistic for the testing in functional linear model, which can be computationally very costly. In order to save calculation time, [?] developed a faster test using a order-two U-statistic. Motivated by this, we propose a non-parametric U-statistic that combines the functional data analysis with the classical kernel method for testing (??).

The paper is organized as follows. In Section 2, we construct a new test procedure for testing in functional partially linear regression model. The theoretical properties of the proposed test statistic under some regularity conditions will be considered in Section 3. Simulation study is conducted in Section 4 to assess the finite sample performance of our proposed test procedure. Section 5 reports the testing result for spectrometric data. All the proofs of main theoretical results are presented in Appendix.

2. Test Statistic

Suppose that Y and U are real-valued random variables. $\{X(\cdot)\}$ 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. $\{(Y_i, X_i(\cdot), U_i), i = 1, 2, \dots, n\}$ is a random sample from model (??),

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

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

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

Then model (??) becomes classical non-parametric model. A pseudo kernel estimate of the non-parametric function using Nadaraya-Watson regression method can be constructed as follows:

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

where $K_h(\cdot) = K(\cdot/h)/h$ with $K(\cdot)$ being a preselected kernel function and h being the bandwidth whose optimal value can be determined by some data-driven methods such as the cross-validation methods. Here we estimate non-parametric $g(U_i)$ without i th sample.

Let

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

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

$$\hat{g}(U_i) = \mathbf{W}_{-i}^T (\mathbf{Y}_{-i} - \langle \mathbf{X}_{-i}, \alpha \rangle). \quad (6)$$

If we replace $g(U_i)$ by $\check{g}(U_i)$ in the model (??), we have

$$\check{Y}_i = \langle \check{X}_i, \alpha \rangle + \varepsilon_i, \quad (7)$$

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 ” stand for “defined as”. Then $\hat{\mu}_{it} = \mathbf{W}_{-i}^T \mathbf{X}_{-i}(t)$ can be estimator of the conditional expectation μ_{it} for any $t \in [0, 1]$.

For an arbitrary orthonormal basis $\{\phi_j\}_{j=1}^\infty$ in $L^2[0, 1]$, the predictor $X(\cdot)$ and the slope function $\alpha(\cdot)$ have following expansions. Denote the number of truncated basis function by p ,

$$X_i(t) = \sum_{j=1}^p \xi_{ij} \phi_j(t) + \sum_{j=p+1}^\infty \xi_{ij} \phi_j(t), \quad \alpha(t) = \sum_{j=1}^p \beta_j \phi_j(t) + \sum_{j=p+1}^\infty \beta_j \phi_j(t), \quad (8)$$

where $\xi_{ij} = \langle X_i, \phi_j \rangle$, $\beta_j = \langle \alpha, \phi_j \rangle$. Let $\check{\xi}_{ij} = \langle \check{X}_i, \phi_j \rangle$, the model (??) can be rewritten as follows:

$$\check{Y}_i = \sum_{j=1}^\infty \check{\xi}_{ij} \beta_j + \varepsilon_i = \sum_{j=1}^p \check{\xi}_{ij} \beta_j + \sum_{j=p+1}^\infty \check{\xi}_{ij} \beta_j + \varepsilon_i. \quad (9)$$

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

$$\check{\xi}_{-i} = (\check{\xi}_1, \dots, \check{\xi}_{i-1}, \check{\xi}_{i+1}, \dots, \check{\xi}_n)^T, \quad \check{\xi}_i = (\check{\xi}_{i1}, \check{\xi}_{i2}, \dots, \check{\xi}_{ip})^T, \\ \check{\xi}_{-i} = (\check{\xi}_1, \dots, \check{\xi}_{i-1}, \check{\xi}_{i+1}, \dots, \check{\xi}_n)^T, \quad \beta = (\beta_1, \beta_2, \dots, \beta_p)^T.$$

For model (??), we define the approximation error as

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

To analyze the effect of the approximation error, we impose following condition on functional predictors and regression function.

(C1) Functional predictors $\{X_i, i = 1, \dots, n\}$ and regression function $\alpha(t)$ satisfy:

(i) Functional predictors $\{X_i, i = 1, \dots, n\}$ belongs to a Sobolev ellipsoid of order-two, then there exists a universal constant C , such that $\sum_{j=1}^\infty \xi_{ij}^2 j^4 \leq C^2$, for $i = 1, \dots, n$.

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

Using Cauchy-Schwarz inequality, we have

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

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

$$\check{Y}_i = \sum_{j=1}^p \check{\xi}_{ij} \beta_j + \varepsilon_i, \quad (10)$$

which is a high-dimensional partial linear model. Since

$$E\|(X_i - E[X_i|U_i])(Y_i - E[Y_i|U_i])\|^2 \quad (11)$$

can be used as an effective measure of the distance between $\alpha(\cdot)$ and 0 when testing (??). Motivated by [?], we construct following test statistic by estimating (??).

$$T_{np} = \left(1 - \frac{2}{n}\right)^{-2} \binom{n}{2}^{-1} \sum_{i=2}^n \sum_{j=1}^{i-1} \Delta_{ij}(\check{X}) \Delta_{ij}(\check{Y}), \quad (12)$$

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 $\bar{\check{Y}}$ are sample means of $\check{X}_i(t)$ and \check{Y}_i . By some calculations, we can get $E[\Delta_{ij}(\check{X})] = 0$, $E[\Delta_{ij}(\check{Y})] = 0$. The test statistic T_{np} can served as a measurement of distance between $\alpha(\cdot)$ and 0 under null hypothesis. Large values of test statistic T_{np} are in favor of alternative hypothesis and leads to rejection of null hypothesis.

3. Asymptotic Theory

To achieve the asymptotic properties of the proposed test, we first suppose following conditions based on [?] and [?]. Denote

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

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

A condition on the dimension of matrix Σ_* is:

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

(C3) There exists a m -dimensional random vector $Z_i = (Z_{i1}, \dots, Z_{im})^T$ for a constant $m \geq p$ so that $\zeta_i = E[\zeta_i | U_i] + \Gamma(U_i) Z_i$. Here Z_i satisfies $E(Z_i) = 0$, $\text{var}(Z_i) = I_m$, and for any U_i , $\Gamma(U_i)$ is a $p \times m$ matrix such that $\Gamma(U_i) \Gamma^T(U_i) = \Sigma_*(U_i)$. Each random vector $\{Z_i, i = 1, \dots, n\}$ is assumed to have finite 4th moments and $E(Z_{ij}^4) = 3 + \Delta$ for some constant Δ . Furthermore, we assume

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

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 has a compact support Ω , and its density function $f(\cdot)$ has continuous second order derivative and bounded away from 0 on its support. The kernel function $K(\cdot)$ is a symmetric probability density function with a compact support. Also, it is Lipschitz continuous.

(C7) $E(\zeta_1 | U_1)$ and $g(\cdot)$ are Lipschitz continuous, and have continuous second order derivatives.

(C8) The sample size n and the smoothing parameter h are assumed to satisfy $\lim_{n \rightarrow \infty} h = 0$, $\lim_{n \rightarrow \infty} nh = \infty$, $\lim_{n \rightarrow \infty} nh^4 = 0$.

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

Condition (C2) has been adopted in many studies about high-dimensional data (see [? ? ?]). Condition (C3) resembles a factor model. To analyze the local power, we also impose the condition (C4) on the coefficient vector β . In fact, (C4) can be served as the local alternatives as its distance measurement between β and 0. This local alternative can be also found in [?]. (C5) is the typical assumptions for the error term ε . Conditions (C6-C8) are very common in non-parametric smoothing. (C9) is a technical condition which is needed to derive the theorems.

We will show the asymptotic theory of our proposed test statistic under the null hypothesis and the local alternative (C4) in the following two theorems.

Theorem 1. Suppose that the conditions (C1), and (C3-C9) hold, then we have

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

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

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

Theorem 2. Suppose that the conditions (C1-C3) and (C5-C9) hold, then under the null hypothesis or the local alternative (C4), we have

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

where \xrightarrow{D} represents convergence in distribution.

Theorem 2 indicates that under the local alternative hypothesis (C4), the proposed test statistic has the following asymptotic local power for the nominal level α ,

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

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and z_α denotes its upper α quantile. Let $\eta(\alpha) = C^*(\alpha)/\sigma^2 \sqrt{2\text{tr}(\Sigma_*^2)}$, and this quantity can be viewed as a signal-to-noise ratio. When the term $\eta(\alpha) = o(1/n)$, the power converges to α , then the power converge to 1 if it has a high order of $1/n$. This implies that the proposed test is consistent. Our proposed test statistic also process the identical asymptotic local power. The performance of the power will be shown by simulation in Section 4.

By Theorem 2, the proposed test statistic rejects H_0 at a signification level α if

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

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

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

where $Y_{1n} = \frac{1}{A_n^2} \sum_{i \neq j} \langle \check{X}_i^T, \check{X}_j \rangle^2$, $Y_{2n} = \frac{1}{A_n^3} \sum_{i \neq j \neq k} \langle \check{X}_i^T, \check{X}_j \rangle \langle \check{X}_j^T, \check{X}_k \rangle$, $Y_{3n} = \frac{1}{A_n^4} \sum_{i \neq j \neq k \neq l} \langle \check{X}_i^T, \check{X}_j \rangle \langle \check{X}_k, \check{X}_l \rangle$ 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 Study

In this section, to evaluate the finite sample performance of the proposed test, some simulation studies are conducted. We generate 1,000 Monte Carlo samples in each simulation. For basis expansion and FPCA, we use the implementation in the R package *fda*.

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

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

where $\{\hat{\lambda}_i\}_{i=1}^p$ is the estimate of the eigenvalue of covariance operator. The minimal m for which $\text{CPV}(m)$ exceeds a desired level, 95% is chosen in this section. We denote p as the number of basis functions used to fit curves. The simulated data is generated from the following model:

$$Y_i = \int_0^1 \alpha(s) X_i(s) ds + g(U_i) + \varepsilon_i, \quad i = 1, 2, \dots, n. \quad (13)$$

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 generated from the uniform distribution $U(0, 1)$. To analyze the impact of different error distributions, 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) $\varepsilon_i \sim (\text{lnorm}(0, 1) - \sqrt{e})/\sqrt{e(e-1)}$. 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 $X(t)$. Because fitting curves with several basis functions are not reasonable for functional data that cannot be approximated by a few principal components, the performance of T_n and T_{np} will be compared based on the change in the number of basis functions used in the fitting curves, and the performance of the proposed test T_{np} when the number of basis functions used to fit curves is large enough (at this point, the value of T_n cannot be calculated) will be presented.

1. The predictor $X(t) = \sum_{j=1}^{50} \xi_j \phi_j(t)$, where ξ_j follows a normal distribution with mean zero and variance $\lambda_j = 10((j-0.5)\pi)^{-2}$, $\phi_j(t) = \sqrt{2}(j-0.5)\pi t$ for $j = 1, 2, \dots, 50$. The slope function $\alpha(t) = c(\sqrt{2}\sin(\pi t/2) + 3\sqrt{2}\sin(3\pi t/2))$ with c varying 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: $p = 11, 49, n = 50, 100$. 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. Empirical size and power when $g(u) = 2u$ for two tests.

(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	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
	0.1	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
	0.2	0.713	0.584	0.761	0.631	0.738	0.602	0.755	0.656
(100,11)	0	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
	0.1	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
	0.2	0.998	0.984	0.986	0.975	0.995	0.975	0.978	0.965
(100,49)	0	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
	0.1	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
	0.2	0.998	0.981	0.987	0.977	0.997	0.978	0.988	0.969

Table 2. Empirical size and power when $g(u) = 2 + \sin(2\pi u)$ for two tests.

(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	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
	0.1	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
	0.2	0.658	0.545	0.724	0.605	0.694	0.573	0.735	0.638
(100,11)	0	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
	0.1	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
	0.2	0.996	0.980	0.993	0.972	1.000	0.979	0.987	0.964
(100,49)	0	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
	0.1	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
	0.2	0.997	0.976	0.994	0.981	0.998	0.980	0.982	0.955

From Tables 1 and 2, it can be seen that, (i) For different error distributions and different non-parametric functions, the performance of both tests is stable; (ii) Due to the asymptotic distribution of T_{np} is for functional data that cannot be approximated by a few principal components, the power of proposed test is slightly lower than that of T_n . (iii) As the sample size n increases, the power increases, while the power does not increase or decrease significantly as the value of p increases. In fact, for this data structure of the predictor, no matter how many basis functions are selected to fit the function curves, the number of principal components we finally select is very small.

2. The functional predictor is generated based on the expansion (??), where ϕ_k 's are Fourier basis functions on $[0,1]$ defined as $\phi_1(t) = 1$, $\phi_2(t) = \sqrt{2}\sin(2\pi t)$, $\phi_3(t) = \sqrt{2}\cos(2\pi t)$, $\phi_4(t) = \sqrt{2}\sin(4\pi t)$, $\phi_5(t) = \sqrt{2}\cos(4\pi t)$, ... The first p of these 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 controls the range of dependence between the components of predictor $X(t)$. $\{Z_{ij}, Z_{i(j+1)}, \dots, Z_{i(p+T-1)}\}$ are independently generated from the normal distribution $N(0, I_{p+T-1})$ with $T = 10$. The element at (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 on $[0,1]$.

The Epanechnikov kernel is adopted in estimating non-parametric part $g(\cdot)$. we select the bandwidth with cross validation (CV). When the significant level α is 0.05, Tables 3 and 4 show the empirical size and power of both tests for different non-parametric functions with different error distributions.

Table 3. Empirical size and power when $g(u) = 2u$ for two tests.

(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	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
	0.5	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	0.865	0.917	0.883	0.925	0.867	0.912	0.889	0.922
(100,11)	0	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
	0.5	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	0.982	0.998	0.971	0.986	0.971	0.989	0.958	0.984
(100,49)	0	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
	0.5	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	0.957	0.966	0.949	0.968	0.956	0.964	0.928	0.934

Table 4. Empirical size and power when $g(u) = 2 + \sin(2\pi u)$ for two tests.

(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	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
	0.5	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	0.841	0.893	0.869	0.913	0.842	0.897	0.874	0.913
(100,11)	0	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
	0.5	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	0.976	0.996	0.959	0.983	0.977	0.994	0.954	0.972
(100,49)	0	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
	0.5	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	0.948	0.964	0.937	0.956	0.938	0.955	0.947	0.956

From Tables 3 and 4, the number of basis functions used for fitting functions has a very important impact on the test. Specifically, (i) Under different error distributions, with the increase of p , the empirical size of test T_n is much greater than the nominal level, while our proposed test T_{np} has a stable performance; (ii) As the sample size n increases, the power increases, while the power decreases as the value of p increase. In fact, for this data structure of the predictor, the number of selected principal components is too large to make the test statistics based on FPCA perform well. Instead, the proposed test has great advantages(see bold numbers in Tables 3-4).

Furthermore, to verify the asymptotic theory of our proposed test, when $(n, p) = (200, 365)$, Figures 1 and 2 draw the null distributions and the q-q plots of T_{np} for $g(u) = 2u$ 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. For different n, p , Figures 3 and 4 respectively show the empirical power functions of the proposed test statistics under four different error distribution functions when the non-parametric function is a linear function and a

trigonometric function. 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. It can be seen that from Figures 3 and 4, the power increases rapidly as long as c increases slightly. As the sample size n increases, the power increases, while the power decreases as p increases. The proposed test is stable under different error distributions. These are consistent with the conclusions in Tables 3 and 4 (i.e. $p < n$).

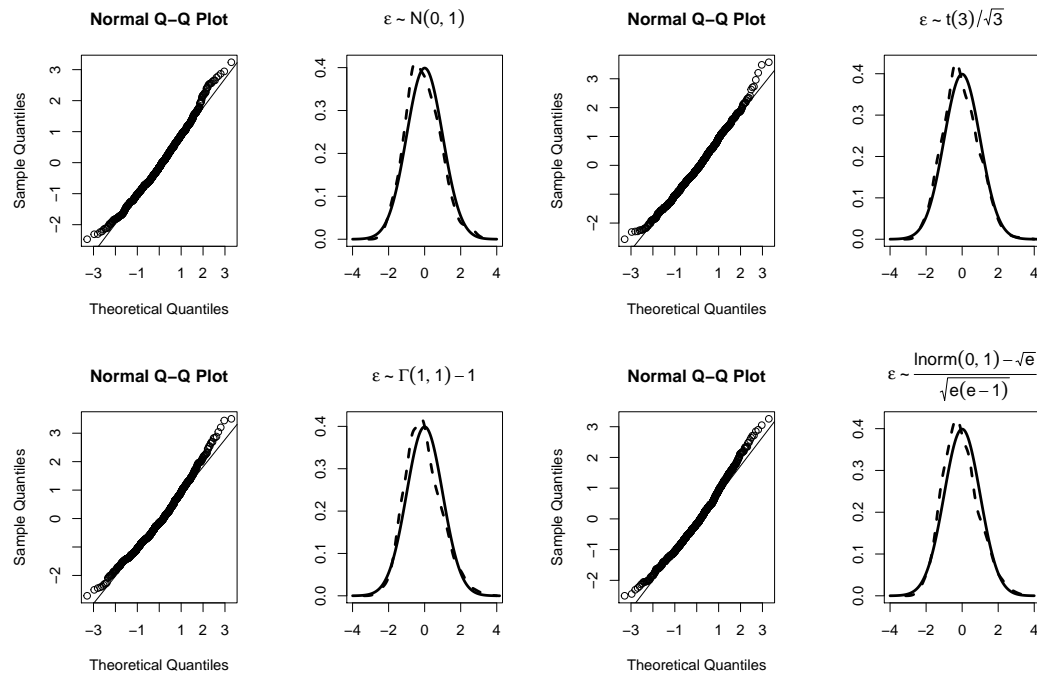


Figure 1. The null distributions and the q-q plots of our proposed test when $g(u) = 2u$.

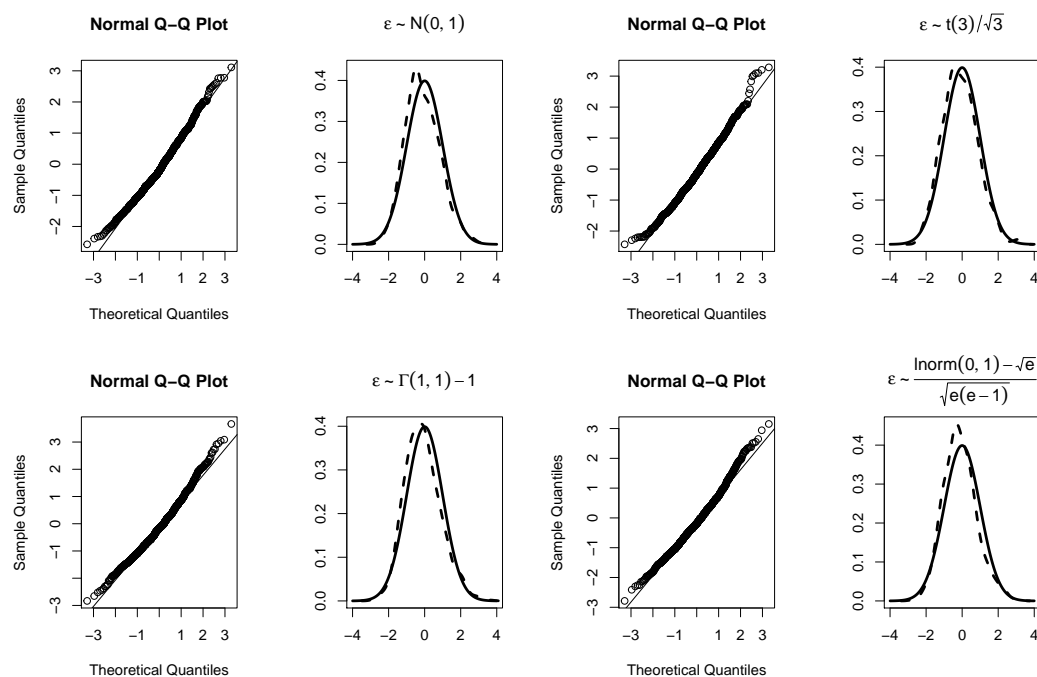


Figure 2. The null distributions and the q-q plots of our proposed test when $g(u) = 2 + \sin(2\pi u)$.

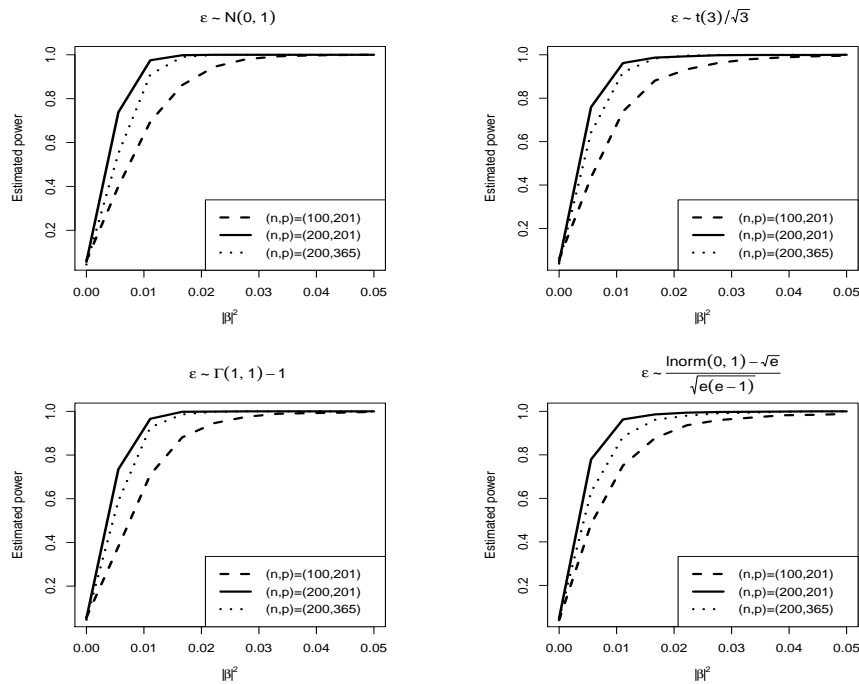


Figure 3. Empirical power functions of our proposed test when $g(u) = 2u$.

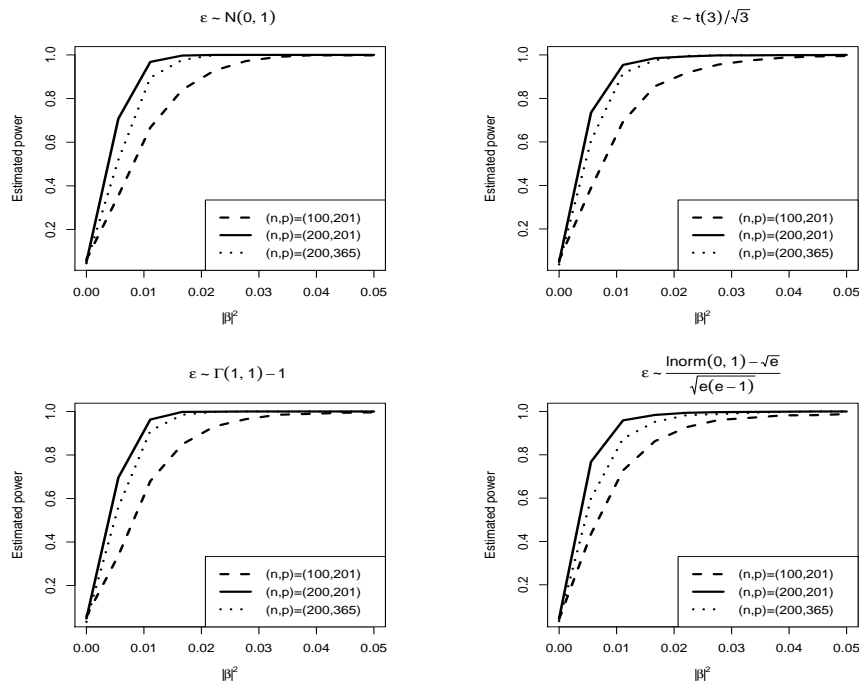


Figure 4. Empirical power functions of our proposed test when $g(u) = 2 + \sin(2\pi u)$.

5. Application

This section applies the proposed test to the spectral data, which has been described and analyzed in the literature (see [? ?]). This dataset can be obtained on the following platforms: <http://lib.stat.cmu.edu/datasets/tecator>. There are 215 meat samples. Each sample contains chopped pure meat, which contains different absorption spectra and fat, protein and water contents. The

observations of the spectral measurement are some curves, denoted by $X_i(\cdot)$, which corresponds to the absorbance measured on the grid with 100 wavelengths ranging from 850nm to 1050nm in step. Fat, protein and water content (in percentage) are measured by chemical analysis method. Denote the fat contents as response variable Y_i , the protein contents as Z_i , and the moisture content as U_i . Similar to [?], the following two models will be used to assume the relationship between them:

$$Y_i = \int_{850}^{1050} \alpha(t) X_i(t) dt + g(Z_i) + \varepsilon_i, \quad (14)$$

$$Y_i = \int_{850}^{1050} \alpha(t) X_i(t) dt + g(U_i) + \varepsilon_i. \quad (15)$$

Here we mainly study the test in models (??) and (??): $\alpha(t) = 0$. The number of basis functions used for fitting function curves p is selected as 129. Figure 5 shows the estimation of slope function $\alpha(t)$ in models (??) and (??).

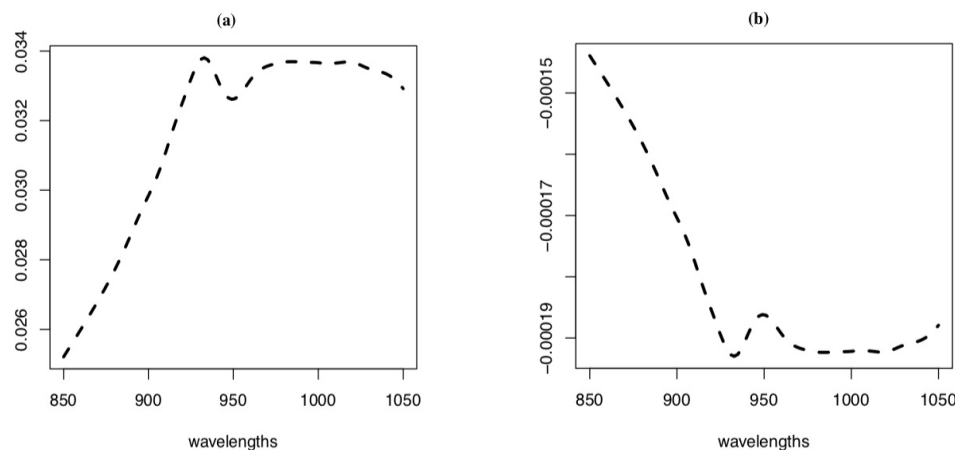


Figure 5. (a) The estimator of the slope function in model (??); (b) The estimator of the slope function in model (??)

The calculation results are as follows: (i) For model (??), the value of statistic is $T_{np} = 31.186$; p-value is 0. (ii) For model (??), the value of statistic is $T_{np} = -0.867$; p-values are 0.386.

From this we can see that the conclusions. The test on model (??) is significant, while the test on model (??) is not significant. This result can also be reflected from Figure 5. It is obvious that the estimated value of $\alpha(t)$ on the right side of Figure 5 is much smaller than that on the left side.

6. Conclusions

In this paper, we constructed a U-statistic of order-two for testing linearity between the functional predictor and the response variable in functional partially linear regression model. The proposed test procedure didn't depend on the estimate of covariance operator of predictor function. The asymptotic distribution of proposed test statistic is normal under a null hypothesis and a local alternative assumption. Furthermore, numerical simulations show that our proposed test performs well when functional data cannot be approximated by a few principal components. Finally, the real data has applied to our proposed test to verify its feasibility.

Author Contributions: Conceptualization, F.Z. and B.Z.; methodology, B.Z.; software, F.Z.; validation, B.Z.; formal analysis, F.Z.; investigation, F.Z.; resources, 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, F.Z. and 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: Not applicable.

Acknowledgments: The authors are grateful to the editor, associate editor, and referees for reviewing this manuscript, and hope to receive constructive comments and suggestions.

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

Appendix A

We present some lemmas in order to complete the proofs of Theorem 1 and Theorem 2. Without loss of generality, we assume that $\mu_0(t) = 0$ and $E[g(U)] = 0$ in the sequel. Let $C_n = \sqrt{\log(1/h)/(nh)} + h^2$. With reference to the asymptotic theory of non-parametric estimation, the pseudo estimation of 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 as the lemmas in [?], it is easy to derive the following lemmas.

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

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

Lemma A2. *If (C1-C3) and (C5-C9) hold, then we have*

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

Lemma A3. *If (C1-C9) hold, we can get*

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

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. Proof of Theorem 1. Write

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

where

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

then the expectation of test statistic Tnp is:

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

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 of calculation in different cases of (l, k) , here we mainly consider the case where $(l, k) = (1, 1)$,

$$E[P_{ij}^{(1)} L_{ij}^{(1)}] \triangleq G_1^{(1,1)} + G_2^{(1,1)} + G_3^{(1,1)} + G_4^{(1,1)} + G_5^{(1,1)} + G_6^{(1,1)}, \quad (A1)$$

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. Since

$$\begin{aligned} &E[\langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle Dg(U_1) Dg(U_2)] \\ &= 2E[\langle D\mu_{1t}, X_2 - \mu_{2t} \rangle Dg(U_1) Dg(U_2)] + E[\langle D\mu_{1t}, D\mu_{2t} \rangle Dg(U_1) Dg(U_2)] \\ &= O\left(W_{12} W_{21} \left(\xi_2^T \xi_2 - \mu^T(U_2) \xi_2\right) g^2(U_1)\right) \\ &= O\left(\text{tr}(\Sigma_*) / n^2 h\right), \end{aligned}$$

then $G_1^{(1,1)} = o\left(\sqrt{\text{tr}(\Sigma_*^2)/n}\right)$ holds. For the second term, we have

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

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 mean 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,

$$\begin{aligned} & E[\langle X_1 - \hat{\mu}_{1t}, X_2 - \hat{\mu}_{2t} \rangle \langle X_2 - \hat{\mu}_{2t}, \alpha \rangle Dg(U_1)] \\ &= E[\langle X_1 - \mu_{1t}, \alpha \rangle \langle X_1 - \mu_{1t}, D\mu_{2t} \rangle Dg(U_2)] \\ &+ E[\langle D\mu_{1t}, \alpha \rangle \langle D\mu_{1t}, X_2 - \mu_{2t} \rangle Dg(U_2)] \\ &+ E[\langle D\mu_{1t}, \alpha \rangle \langle X_1 - \mu_{1t}, D\mu_{2t} \rangle Dg(U_2)] \\ &+ E[\langle X_1 - \mu_{1t}, \alpha \rangle \langle D\mu_{1t}, D\mu_{2t} \rangle Dg(U_2)] \\ &+ E[\langle D\mu_{1t}, \alpha \rangle \langle D\mu_{1t}, D\mu_{2t} \rangle Dg(U_2)] \\ &= O\left(nE[W_{23}^2\beta^T(\xi_1\xi_1^T - \xi_1\mu^T(U_1))\xi_3g(U_3)]\right) \\ &= O\left(E[\beta^T\Sigma_*(U_1)\mu(U_2)g(U_2)f^{-1}(U_2)]/nh\right) \\ &= o\left(\sqrt{\text{tr}(\Sigma_*^2)/n}\right). \end{aligned}$$

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

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

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

$$E\left[P_{ij}^{(1)} L_{ij}^{(1)}\right] = \|C^*(\alpha)\|^2 + o\left(\sqrt{\text{tr}(\Sigma_*^2)/n}\right). \quad (\text{A2})$$

For the rest, refer to calculation of the above mean $E \left[P_{ij}^{(1)} L_{ij}^{(1)} \right]$ and the proof of Theorem 1 in [?]. The proof of the conclusion (i) in Theorem 1 is completed. The conclusion (ii) of Theorem 1 can be found in the proof of Theorem 2, here we omit it. \square

Proof. Proof of Theorem 2

By the Theorem 1, we have the fact

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

then we only need to prove that

$$\frac{n(T_{np} - ET_{np})}{\sigma^2 \sqrt{2\text{tr}(\Sigma_*^2)}} \xrightarrow{D} N(0,1). \quad (\text{A3})$$

Denote $T_{np}^{(k,l)} = n \binom{n}{2}^{-1} \sum_{i>j} \left(P_{ij}^{(k)} L_{ij}^{(l)} - E \left(P_{ij}^{(k)} L_{ij}^{(l)} \right) \right)$, where $k, l = 1, 2, 3, 4$. Then we have

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

In order to obtain the asymptotic properties of 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

$$T_{np}^{(1,1)} \triangleq T_{np,1}^{(1,1)} + T_{np,2}^{(1,1)} + T_{np,3}^{(1,1)} + T_{np,4}^{(1,1)} + T_{np,5}^{(1,1)} + T_{np,6}^{(1,1)} + T_{np,7}^{(1,1)} + T_{np,8}^{(1,1)} + T_{np,9}^{(1,1)}, \quad (\text{A4})$$

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\}, \\
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 (??), We shall prove

$$\frac{T_{np}^{(1,1)} - ET_{np}^{(1,1)}}{\sigma^2 \sqrt{2\text{tr}(\Sigma_*^2)}} = \frac{T_{np,91}^{(1,1)}}{\sigma^2 \sqrt{2\text{tr}(\Sigma_*^2)}} + o_p(1), \quad (\text{A5})$$

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 9 items in the right equation of (??) are all zero. Then in order to calculate their asymptotic order, it is necessary to prove their second moment. Due to the similarity of calculation of the first 8 items, we use the first item $T_{np,1}^{(1,1)}$ as an example to consider.

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

where

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

Let's calculate $E[Q_{i,1}^{(1,1)} Q_{i,1}^{(1,1)}]$ and $E[Q_{i,1}^{(1,1)} Q_{j,1}^{(1,1)}]$ ($i \neq j$).

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

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

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 get

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

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

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

$$\begin{aligned} B_{13}^{(1,1)} &= O\left(E[D\mu^T(U_1) D\mu(U_2) (\mu(U_3) - \hat{\mu}(U_3))^T (\mu(U_4) - \hat{\mu}(U_4)) \right. \\ &\quad \left. Dg(U_1) Dg(U_2) Dg(U_3) Dg(U_4)]\right) \\ &= O\left(\mu^T(U_1) \mu(U_2) \mu^T(U_3) \mu(U_4) g(U_1) g(U_2) g(U_3) g(U_4) / n^2\right). \end{aligned}$$

So we can have $T_{np,1}^{(1,1)} = o_p\left(\sqrt{\text{tr}(\Sigma_*^2)}\right)$. 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 $T_{np,9}^{(1,1)}$, 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 above four formulas are zero, in order to prove that (??) is true, it is necessary to verify the second moments of $T_{np,9k}^{(1,1)}, k = 2, 3, 4$ are the high-order infinitesimal of quantity $\text{tr}(\Sigma_*^2)$. In fact,

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

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

Then the equation (??) holds. Similarly, for Group 2, that is, 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,

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

Using the fact $E[\text{tr}(\text{diag}(\mathbf{\Gamma}^T(U_1)\mathbf{\Gamma}(U_1))\mathbf{\Gamma}^T(U_1)\mathbf{\Gamma}(U_1))] = E[\text{tr}(\mathbf{\Sigma}_*^2(U_1))] = O(\text{tr}(\mathbf{\Sigma}_*^2))$, we have

$$E[(\langle \check{X}_1, \check{X}_1 \rangle + \langle \check{X}_2, \check{X}_2 \rangle - 2E(\langle \check{X}_1, \check{X}_1 \rangle))^2] = O(\text{tr}^2(\mathbf{\Sigma}_*) + \text{tr}(\mathbf{\Sigma}_*^2)). \quad (\text{A6})$$

In addition, by a simple calculation,

$$E[(\check{Y}_1^2 + \check{Y}_2^2 - 2E[\check{Y}_1^2])^2] = 2E[(\check{Y}_1^2)^2] + 2E[\check{Y}_1^2\check{Y}_2^2] - 4E^2[\check{Y}_1^2] = O(1). \quad (\text{A7})$$

Combined (??), (??) with Cauchy-Schwarz inequality, we have

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

Denote $\check{T}_{np1} \triangleq \binom{n}{2}^{-1/2} \sum_{i=2}^n \sum_{j=1}^{i-1} \langle X_i - \mu_{it}, X_j - \mu_{jt} \rangle \varepsilon_i \varepsilon_j$, by condition (C1), we only need to consider $\check{T}_{np} = \binom{n}{2}^{-1/2} \sum_{i=2}^n \sum_{j=1}^{i-1} (\xi_i - \mu(U_i))^T (\xi_j - \mu(U_j)) \varepsilon_i \varepsilon_j$. 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 $\text{var}(\check{T}_{np}) = \sigma^4 \text{tr}(\mathbf{\Sigma}_*^2)$. 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 $\{Xu_i, k = 1, \dots, i\}$, $V_n = \sum_{i=2}^n v_{ni}$. It is easy to check $E[Z_{ni} | \mathcal{F}_i] = 0$, and $\{Z_{ni}, \mathcal{F}_j : 2 \leq j \leq n\}$ is a martingale with mean 0. The martingale central limit theorem follows if we can check

$$\frac{V_n}{\text{var}(\check{T}_{np})} \xrightarrow{P} 1, \text{ as } n \rightarrow \infty; \quad (\text{A8})$$

$$\sum_{i=2}^n \sigma^{-4} \text{tr}^{-1}(\mathbf{\Sigma}_*^2) E\{Z_{ni}^2 I(|Z_{ni}| > \eta \sigma^2 \sqrt{\text{tr}(\mathbf{\Sigma}_*^2)}) | \mathcal{F}_{i-1}\} \xrightarrow{P} 0 \text{ for } \forall \eta > 0. \quad (\text{A9})$$

Note that

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

Then we define

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

where

$$\begin{aligned} C_{n1} &= \frac{1}{\binom{n}{2} \sigma^2 \text{tr}(\mathbf{\Sigma}_*^2)} \sum_{j < i} \varepsilon_j^2 (\xi_j - \mu(U_j))^T \mathbf{\Sigma}_* (\xi_j - \mu(U_j)), \\ C_{n2} &= \frac{2}{\binom{n}{2} \sigma^2 \text{tr}(\mathbf{\Sigma}_*^2)} \sum_{k < l < i} \varepsilon_k \varepsilon_l (\xi_k - \mu(U_k))^T \mathbf{\Sigma}_* (\xi_l - \mu(U_l)). \end{aligned}$$

It is easy to check $E[C_{n1}] = 1$, and

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

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

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

Combined with $\text{tr}(\Sigma_*^4) = o(\text{tr}^2(\Sigma_*^2))$, then we have $C_{n2} \xrightarrow{P} 0$. Thus, equation (??) holds. Finally, we only need to prove (??). Therefore, using the law of large numbers, and the fact $E[Z_{ni}^2 I(|Z_{ni}| >$

$\eta\sigma^2\text{tr}(\Sigma_*^2)] \leq E(Z_{ni}^4|\mathcal{F}_{i-1})/(\eta^2\sigma^4\text{tr}(\Sigma_*^2))$, we only need to show that $\sum_{2 \leq i \leq n} E(Z_{ni}^4) = o(\text{tr}^2(\Sigma_*^2))$. By simple calculation, we can get

$$\begin{aligned} & \sum_{i=2}^n E[Z_{ni}^4] \\ &= \frac{1}{\binom{n}{2}^2} \sum_{j < i} E[(\xi_i - \mu(U_i))^T (\xi_j - \mu(U_j))]^4 \epsilon_i^4 \epsilon_j^4] + \sum_{i=2}^n E[Z_{ni}^4] \\ &= \frac{1}{\binom{n}{2}^2} \sum_{j < i} \sum_{k \neq j} E[(\xi_i - \mu(U_i))^T (\xi_j - \mu(U_j))]^2 (\xi_i - \mu(U_i))^T (\xi_k - \mu(U_k))^2 \epsilon_i^4 \epsilon_j^2 \epsilon_k^2] \\ &= \frac{E^2(\epsilon_1^4)}{\binom{n}{2}^2} \sum_{i=2}^n (i-1) E[(\xi_1 - \mu(U_1))^T (\xi_2 - \mu(U_2))]^4 + \frac{3\sigma^4 E(\epsilon_1^4)}{\binom{n}{2}^2} \sum_{i=2}^n (i-1)(i-2) \\ & \quad E[(\xi_1 - \mu(U_1))^T (\xi_2 - \mu(U_2))]^2 (\xi_1 - \mu(U_1))^T (\xi_3 - \mu(U_3))^2] \\ &= O\left(\frac{E[(\xi_1 - \mu(U_1))^T (\xi_2 - \mu(U_2))]^4}{n^2}\right) + O\left(\frac{E[(\xi_1 - \mu(U_1))^T \Sigma_* (\xi_1 - \mu(U_1))]^2}{n}\right) \end{aligned}$$

Combined (C2) and Lemma 2, equation (??) holds. Thus, the proof of Theorem 2 completed. \square

References

- . Crainiceanu, C.M.; Staicu, A.M.; Di, C.Z. Generalized multilevel functional regression. *Journal of the American Statistical Association* **2009**, *104*, 1550–1561.
- . Leng, X. Müller, H.G. Time ordering of gene co-expression. *Biostatistics* **2006**, *7*, 569–584.
- . Kokoszka, P.; Miao, H.; Zhang, X. Functional Dynamic Factor Model for Intraday Price Curves. *Journal of Financial Econometrics* **2014**, *13*, 456–477.
- . Ignaccolo, R. and Ghigo, S. and Giovenali, E. Analysis of air quality monitoring networks by functional clustering. *Environmetrics* **2008**, *19*, 672–686.
- . Yao, F.; Müller, H.G. Functional quadratic regression. *Biometrika* **2010**, *97*, 49–64.
- . Lian, H. Functional partial linear model. *Journal of Nonparametric Statistics* **2011**, *23*, 115–128.
- . Zhou, J.; Chen, M. Spline estimators for semi-functional linear model. *Statistics & Probability Letters* **2012**, *82*, 505–513.
- . Tang, Q. Estimation for semi-functional linear regression. *Statistics* **2015**, *49*, 1262–1278.
- . Cardot, H.; Ferraty, F.; Mas, A.; Sarda, P. Testing hypotheses in the functional linear model. *Scandinavian Journal of Statistics* **2003**, *30*, 241–255.
- . Kokoszka, P.; Maslova, I.; Sojka, J.; Zhu, L. Testing for lack of dependence in the functional linear model. *Canadian Journal of Statistics* **2008**, *36*, 207–222.
- . James, G.M.; Wang, J.; and Zhu, J. Functional linear regression that's interpretable. *The Annals of Statistics* **2009**, *37*, 2083–2108.
- . Shin, H. Partial functional linear regression. *Journal of Statistical Planning and Inference* **2009**, *139*, 3405–3418.
- . Yu, P.; Zhang, Z.; Du, J. A test of linearity in partial functional linear regression. *Metrika* **2016**, *79*, 953–969.
- . hu, H.; Zhang, R.; Yu, Z.; Lian, H.; Liu, Y. Estimation and testing for partially functional linear errors-in-variables models. *Journal of Multivariate Analysis* **2019**, *170*, 296–314.
- . Cardot, H.; Goia, A.; Sarda, P. Testing for no effect in functional linear regression models, some computational approaches. *Communications in Statistics-Simulation and Computation* **2004**, *33*, 179–199.
- . Zhu, H.; Zhang, R.; Li, H. Estimation on semi-functional linear errors-in-variables models. *Communications in Statistics-Theory and Methods* **2019**, *48*, 4380–4393.
- . Zhou, J.; Peng, Q. Estimation for functional partial linear models with missing responses. *Statistics & Probability Letters* **2020**, *156*, 108598.
- . Zhao, F.; Zhang, B. Testing Linearity in Functional Partially Linear Models. *Acta Mathematicae Applicatae Sinica, English Series* **2024**, *40*, 875–886.
- . Hu, W.; and Lin, N.; Zhang, B. Nonparametric testing of lack of dependence in functional linear models. *PLoS ONE* **2020**, *15*, e0234094.

- . Zhao, F.; Lin, N.; Hu, W.; Zhang, B. A faster U-statistic for testing independence in the functional linear models. *Journal of Statistical Planning and Inference* **2022**, *217*, 188–203.
- . Zhao, F.; Lin, N.; Zhang, B. A new test for high-dimensional regression coefficients in partially linear models. *Canadian Journal of Statistics* **2023**, *51*, 5–18.
- . Cui, H.; Guo, W.; Zhong, W. Test for high-dimensional regression coefficients using refitted cross-validation variance estimation. *The Annals of Statistics* **2018**, *46*, 958–988.
- . Zhong, P.; Chen, S. Tests for High-Dimensional Regression Coefficients With Factorial Designs. *Journal of the American Statistical Association* **2011**, *106*, 260–274.
- . Chen, S.; Zhang, L.; Zhong, P. Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association* **2010**, *105*, 810–819.
- . Ferraty, F.; Vieu, P. *Nonparametric functional data analysis: theory and practice*; Springer: New York, USA, 2006.
- . Shang, H.L. Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density. *Computational Statistics* **2014**, *29*, 829–848.
- . Yu, P.; Zhang, Z.; Du, J. Estimation in functional partial linear composite quantile regression model. *Chinese Journal of Applied Probability and Statistics* **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.