J Syst Sci Complex (2021) 34: 1135–1155

Empirical Likelihood Test for Regression Coefficients in High Dimensional Partially Linear Models*

LIU Yan · REN Mingyang · ZHANG Sanguo

DOI: 10.1007/s11424-020-9260-3
Received: 17 September 2019 / Revised: 22 February 2020

Abstract This paper considers tests for regression coefficients in high dimensional partially linear Models. The authors first use the B-spline method to estimate the unknown smooth function so that it could be linearly expressed. Then, the authors propose an empirical likelihood method to test regression coefficients. The authors derive the asymptotic chi-squared distribution with two degrees of freedom of the proposed test statistics under the null hypothesis. In addition, the method is extended to test with nuisance parameters. Simulations show that the proposed method have a good performance in control of type-I error rate and power. The proposed method is also employed to analyze a data of Skin Cutaneous Melanoma (SKCM).

Keywords Empirical likelihood test, high dimensional analysis, partially linear models, regression coefficients.

1 Introduction

The Cancer Genome Atlas (TCGA) collects data on skin melanoma (SKCM). In the SKCM data, the sample size (n) is much smaller than the number of genes (p). This is the so-called large p, small n paradigm in statistics. Besides, the SKCM data also contains some environmental factors with nonlinear effects. In this situation, partially linear models, proposed by [1], are more flexible than linear models. Therefore, we consider high dimensional partially linear models. Partially linear models are widely used semiparametric models. Consider the following partially linear model

$$Y_i = X_i^{\mathrm{T}}\beta + g(T_i) + \varepsilon_i, \quad i = 1, 2, \cdots, n, \tag{1}$$

LIU Yan · REN Mingyang · ZHANG Sanguo (Corresponding author)
School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049*, China; Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing* 100190*, China.* Email: liuyan172@mails.ucas.ac.cn; renmingyang17@mails.ucas.ac.cn; sgzhang@ucas.ac.cn.
*This work is supported by the University of Chinese Academy of Sciences under Grant No. Y95401TXX2, Beijing Natural Science Foundation under Grant No. Z190004, and Key Program of Joint Funds of the National Natural Science Foundation of China under Grant No. U19B2040.
◇ *This paper was recommended for publication by Editor DONG Yuexiao.*

where Y_i is a response variable, $X_i = (x_{i1}, x_{i2}, \cdots, x_{ip})^{\mathrm{T}}$ is a p-dimensional random vector and $\Sigma = E(X_i X_i^{\mathrm{T}})$, T_i is another explanatory variable, and β is the regression coefficients vector. $g(T)$ is an arbitrary smooth function, the random variable ε_i is independent of X_i and T_i with $E(\varepsilon_i) = 0$ and $\mathrm{Var}(\varepsilon_i) = \sigma^2$. Our interest is to test

$$H_0 : \beta = \beta_0 \quad \text{vs} \quad H_1 : \beta \neq \beta_0 \tag{2}$$

for a specific $\beta_0 \in \mathbb{R}^p$ under $n < p$.

Hypothesis testing of high dimensional data brings a lot of difficulties to traditional statistics. Conventional methods, such as F-test and t-test, are unsuitable in the high dimensional cases. To estimate the regression coefficients β in linear models when $p \geq n$, [2] proposed a U statistic to estimates $\|\Sigma(\beta - \beta_0)\|_2^2$ and transform the original hypothesis problem into the question whether $\|\Sigma(\beta - \beta_0)\|_2^2$ equals to zero, where $\|\cdot\|_2$ denotes the L_2 norm. In high dimensional situation, asymptotic variance is difficult to estimate, while empirical likelihood method proposed by [3] could avoid the problem. Based on the empirical likelihood method, [4] proposed tests for high dimensional linear model. They define two U statistics to improve power when $\|\beta - \beta_0\|_1$ is small, where $\|\cdot\|_1$ denotes the L_1 norm. [5] extended the method to high dimensional generalized linear models. [6] and [7] applied jackknife empirical likelihood method to high dimensional inference. Besides, [8] and [9] applied the adjusted empirical likelihood methods whose asymptotic distributions are standard chi-squared.

Partially linear models could be transformed into linear models by nonparametric methods such as spline and kernel techniques. Nonparametric estimation and its convergence rate are significant part of the research of partially linear models. [10] studied the errors of least square spline and bias minimizing spline estimation. For the nonparametric regression problem, [11, 12] derived the optimal convergence rate of least squares estimation, and [13] studied the convergence rates of the series estimations. And for the regression of partially linear models, [14] and [15] researched the properties of second order approximation and m-estimation respectively, and [16] gave the optimal convergence rates of spline method. To test the regression coefficients of high dimensional partially linear models, [17] constructed a generalized F-test in the case of $p/n \to c \in (0, 1)$.

Our study has two main contributions. First, we provide a method for testing the regression coefficients in partially linear models in high dimensional cases where $p \geq n$ is available. By expanding the smooth function $g(T)$ by the B-spline method, the model (1) could be transformed into linear model. Then, motivated by [2] and [4], we proposed an empirical likelihood test. And to construct the confidence interval, we proposed an adjusted empirical likelihood test. Second, we show that, under the null hypothesis, the empirical likelihood test has an asymptotic scaled chi-square distribution, and the adjusted empirical likelihood test has an asymptotic chi-square distribution with two degrees of freedom. And, the asymptotic distribution of the proposed test statistic is independent of the dimension p and the number of the B-spline basis functions. Moreover, we consider the partial test with presence of nuisance parameters.

The paper is organized as follows. We introduce the estimation of the smooth function in the partially linear model and present a method bases on a empirical likelihood method in

Section 2. Section 3 considers the partial test with nuisance parameters. We show performances of the proposed method by simulations in Section 4 and employ the method to analyze a real dataset in Section 5. And, Section 6 provides some discussions. All technical details are given in the Appendix.

2 The Proposed Test

Consider the partially linear regression model (1), we have to approximate unknown smooth function $g(T)$ before handling the hypothesis (2). Suppose $g(T_i)$ can be approximated by basis expansion that $g(T_i) = \sum_{j=1}^{K} \gamma_j B_j(T_i) + e_i = B(T_i)^{\mathrm{T}}\gamma + e_i$, where K is the number of basis functions. $B(\cdot) = \{B_1(\cdot), B_2(\cdot), \cdots, B_K(\cdot)\}$ is the B-spline basis, and $\gamma = \{\gamma_1, \gamma_2, \cdots, \gamma_K\}^{\mathrm{T}}$ is the B-spline coefficient vector. The error of the expansion $e = \{e_1, e_2, \cdots, e_n\}^{\mathrm{T}}$. Through substituting β_0 into (1), we could gain the estimation of $g(T_i)$ (denoted as $\widehat{g}(T_i)$) by the ordinary least square method.

Inspired by the ideas of [4] and [6], we divide the sample into two parts: $\{(X_1, Y_1, T_1), (X_2, Y_2, T_2), \cdots, (X_m, Y_m, T_m)\}$ and $\{(X_{m+1}, Y_{m+1}, T_{m+1}), \cdots, (X_{2m}, Y_{2m}, T_{2m})\}$, where $m = \lfloor n/2 \rfloor$, which represents the maximal integer less then $n/2$. Then we consider two U-statistics:

$$U_i = (Y_i - X_i^{\mathrm{T}}\beta_0 - \widehat{g}(T_i))(Y_{i+m} - X_{i+m}^{\mathrm{T}}\beta_0 - \widehat{g}(T_{i+m}))X_i^{\mathrm{T}}X_{i+m},$$
$$V_i = (Y_i - X_i^{\mathrm{T}}\beta_0 - \widehat{g}(T_i))X_i^{\mathrm{T}}\alpha + (Y_{i+m} - X_{i+m}^{\mathrm{T}}\beta_0 - \widehat{g}(T_{i+m}))X_{i+m}^{\mathrm{T}}\alpha,$$

where $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_p)^{\mathrm{T}} \in \mathbb{R}^p$, and α could be set based on prior information. We usually choose α as $(1, 1, \cdots, 1) \in \mathbb{R}^p$, when the prior information is unknown. By applying the empirical likelihood method, we set $Z_i = (U_i, V_i)^{\mathrm{T}}$ and define the empirical likelihood test statistic $l_m = -2\log L_m$, where L_m is defined as

$$L_m = \sup\left\{\prod_{k=1}^{m}(mq_k)\,\middle|\, q_k \geq 0, \sum_{k=1}^{m} q_k = 1, \sum_{k=1}^{m} q_k Z_k = \mathbf{0}\right\}. \tag{3}$$

Using lagrange multiplier technique, we have

$$l_m = \sum_{k=1}^{m}\log(1 + \lambda^{\mathrm{T}}Z_k),$$

where λ satisfies

$$\frac{1}{m}\sum_{k=1}^{m}\frac{Z_k}{1 + \lambda^{\mathrm{T}}Z_k} = \mathbf{0}.$$

Let $\{\eta_1(t), \eta_2(t), \cdots, \eta_p(t)\}^{\mathrm{T}} = E(X_1|T_1 = t)$ and $d = (d_1, d_2, \cdots, d_n)^{\mathrm{T}}$, where $d_i = X_i - E(X_i|T_i)$ with $\mathrm{Var}(d_i) = \varSigma_1$, and $d_i = (d_{i1}, d_{i2}, \cdots, d_{ip})^{\mathrm{T}}$. To establish the asymptotic properties of the empirical likelihood test, we need the following assumptions:

A1 d_i is generated by an s-variate random vector $F_i = (F_{i1}, F_{i2}, \cdots, F_{is})^{\mathrm{T}}$ such that $d_i = \varGamma F_i$, where $\varGamma$ is a $p \times s$ matrix for some $p \geq s$ such that $\varGamma\varGamma^{\mathrm{T}} = \varSigma_1$, each F_{il} has finite 8-th moment, $E(F_i) = 0, \mathrm{Var}(F_i) = I_s, E(F_{ik}^4) = 3 + \Delta$ and for any $\sum_{v=1}^{d} l_v \leq 8, E(F_{1i_1}^{l_1}F_{2i_2}^{l_1}\cdots F_{1i_d}^{l_d}) = E(F_{1i_1}^{l_1})E(F_{1i_2}^{l_2})\cdots E(F_{1i_d}^{l_d})$, where Δ is some finite constant.

A2 $S(T)$ is a compact subset in the support of T, and without loss of generality, we set $S(T) = [0, 1]$. The distribution of T is absolutely continuous and its density function f satisfies that $0 < \inf_{t\in[0,1]} f(t) \leq \sup_{t\in[0,1]} f(t) < \infty$.

A3 Let $\mathcal{H}_h([0, 1])$ be the collection of functions defined on $[0, 1]$ which are h-times continuously differentiable, and $g(t)$ and $\eta_j(t) \in \mathcal{H}_h([0, 1])$. The order of the number of the knots is $O(n^{1/(2h+1)})$.

A4 $\mathrm{Var}(Y_i|X_i, T_i) < \infty$ and $\mathrm{Var}(X_i|T_i) < \infty$ for all i; $E(d_i d_i^{\mathrm{T}})$ is uniformly positive definite; $E(|d_{ij}|^{2+\delta}) < \infty$ for each j for some $\delta > 0$.

Remark 2.1 Assumption A1 resembles a factor model proposed by [18]. Assumption A3 indicates the smoothness of $g(T)$ and gives the optimal rate of the growth of the number of knots. And, under Assumption A3, the order of e_i could be constrained to be $O_p(m^{-h/(2h+1)})$. Under Assumptions A2–A4, we could obtain that $\|g(T_i) - \widehat{g}(T_i)\|^2 = O_p(m^{1/(2h+1)})$ (see [16]).

Theorem 2.1 in the following gives the asymptotic property of the proposed empirical likelihood test.

Theorem 2.1 *Suppose that* $\|\frac{\alpha}{\sqrt{\alpha^{\mathrm{T}}\Sigma_1\alpha}}\| = O(1)$, A1–A4 *hold, as* $n \to \infty$ *and* $p \to \infty$, *and there exists some* $\delta > 0$ *satisfies*

$$\frac{E|X_i^{\mathrm{T}} X_{i+m}|^{2+\delta}}{\{\mathrm{tr}(\Sigma^2)\}^{(2+\delta)/2}} = o(m^{\delta/2}), \tag{4}$$

and

$$\frac{E|X_i^{\mathrm{T}}\alpha + X_{i+m}^{\mathrm{T}}\alpha|^{2+\delta}}{\{\alpha^{\mathrm{T}}\Sigma_1\alpha\}^{(2+\delta)/2}} = o(m^{\delta/2}). \tag{5}$$

Then, under H_0,

$$l_m \xrightarrow{d} \chi^2_{1,1} + w\chi^2_{1,2}, \tag{6}$$

where $\chi^2_{1,1}$ *and* $\chi^2_{1,2}$ *are independent* χ^2_1 *variables,* $w = \alpha^{\mathrm{T}}\Sigma_1\alpha/\alpha^{\mathrm{T}}\Sigma\alpha$, *and* $\xrightarrow{d}$ *represents convergence in distribution.*

Remark 2.2 When X and T are independent, l_m converges in distribution to an asymptotic chi-square distribution with two degrees of freedom under H_0.

Theorem 2.1 indicates that l_m has an asymptotic scaled chi-square distribution. Hence, we consider an adjusted l_m:

$$l_m^* = \xi l_m, \tag{7}$$

where ξ is an adjustment factor, and we will discuss it in the appendix. Theorem 2.2 in the following gives the asymptotic property of the adjusted empirical likelihood test.

Theorem 2.2 *Under the assumptions of Theorem* 2.1, *as* $n \to \infty$ *and* $p \to \infty$, l_m^* *converges in distribution to an asymptotic chi-square distribution with two degrees of freedom under* H_0.

Based on Theorem 2.2, we reject H_0 when $l_m^* > \chi^2_{2,b}$, where $\chi^2_{2,b}$ represents the $(1 - b)$-quantile of a chi-square distribution with two degrees of freedom and b is the significant level. To elucidate the local power of the empirical likelihood test, we consider testing the null hypothesis H_0 against the local alternatives H_{10}:

$$\begin{pmatrix} (\beta-\beta_0)^{\mathrm{T}}\Sigma^2(\beta-\beta_0) \\ \alpha^{\mathrm{T}}\Sigma_1(\beta-\beta_0) \end{pmatrix} = \frac{C^{1/2}\gamma}{\sqrt{m}}, \tag{8}$$

where $\gamma=(\gamma_1,\gamma_2)^{\mathrm{T}}$, and γ is a finite constant vector,

$$C=\begin{pmatrix} \sigma^4\mathrm{tr}(\Sigma^2) & 0 \\ 0 & \sigma^2\alpha^{\mathrm{T}}\Sigma_1\alpha \end{pmatrix}.$$

According to [19] and [20], and under the local alternatives, l_m converges in distribution to a noncentral chi-square distribution with two degrees of freedom and the noncentrality parameter $\varrho=\gamma^{\mathrm{T}}\gamma$ as $n\to\infty$ and $p\to\infty$.

Remark 2.3 When $\gamma^{\mathrm{T}}\gamma=o(1)$, the test cannot distinguish H_0 from the local alternative.

3 The Partial Test with Nuisance Parameters

In this section, we extend the proposed method to partial test with the presence of nuisance parameters. Therefore, we consider the test for part of the regression coefficients. Without loss of generality, we split $X_i=(X_i^{(1)\mathrm{T}},X_i^{(2)\mathrm{T}})^{\mathrm{T}}$, where the dimensions of $X_i^{(1)}$ and $X_i^{(2)}$ are p_1 and p_2 respectively. And the regression coefficients of $X_i^{(1)}$ and $X_i^{(2)}$ are $\beta_i^{(1)}$ and $\beta_i^{(2)}$ respectively. Then we test

$$\widetilde{H}_0:\beta^{(2)}=\beta_0^{(2)} \quad \text{vs} \quad \widetilde{H}_1:\beta^{(2)}\neq\beta_0^{(2)} \tag{9}$$

with an existed $\beta^{(1)}\in\mathbb{R}^{p_1}$.

To test (9), we should estimate $\beta^{(1)}$ firstly. By substituting $\beta_0^{(2)}$ and $g(T_i)=B(T_i)^{\mathrm{T}}\gamma+e_i$ into the partially linear model (1), we can obtain the estimate of $\beta^{(1)}$ (denoted as $\widehat{\beta}^{(1)}$) and the estimate of $g(T_i)$ (denoted as $\widetilde{g}(T_i)$) with the ordinary least squares method. Set $\widehat{\beta}_0=(\widehat{\beta}^{(1)\mathrm{T}},\beta_0^{(2)\mathrm{T}})^{\mathrm{T}}$ and we consider two U-statistics:

$$\begin{aligned}
\widetilde{U}_i&=(Y_i-X_i^{\mathrm{T}}\widehat{\beta}_0-\widetilde{g}(T_i))(Y_{i+m}-X_{i+m}^{\mathrm{T}}\widehat{\beta}_0-\widetilde{g}(T_{i+m}))X_i^{(2)\mathrm{T}}X_{i+m}^{(2)},\\
\widetilde{V}_i&=(Y_i-X_i^{\mathrm{T}}\widehat{\beta}_0-\widetilde{g}(T_i))X_i^{\mathrm{T}}\widetilde{\alpha}+(Y_{i+m}-X_{i+m}^{\mathrm{T}}\widehat{\beta}_0-\widetilde{g}(T_{i+m}))X_{i+m}^{(2)\mathrm{T}}\widetilde{\alpha},
\end{aligned}$$

where $\widetilde{\alpha}=(\widetilde{\alpha}_1,\widetilde{\alpha}_2,\cdots,\widetilde{\alpha}_{p_2})^{\mathrm{T}}\in\mathbb{R}^{p_2}$. And the choice of $\widetilde{\alpha}$ is similar to that of α in Section 2. By applying the empirical likelihood method, we set $\widetilde{Z}_i=(\widetilde{U}_i,\widetilde{V}_i)^{\mathrm{T}}$ and define the empirical likelihood test statistic $\widetilde{l}_m=-2\log\widetilde{L}_m$, where $\widetilde{L}_m$ is defined as

$$\widetilde{L}_m=\sup\left\{\prod_{k=1}^m(m\widetilde{q}_k)\,\middle|\,\widetilde{q}_k\ge 0,\sum_{k=1}^m\widetilde{q}_k=1,\sum_{k=1}^m\widetilde{q}_k\widetilde{Z}_k=\mathbf{0}\right\}. \tag{10}$$

Using lagrange multiplier technique, we have

$$\widetilde{l}_m=\sum_{k=1}^m\log(1+\widetilde{\lambda}^{\mathrm{T}}\widetilde{Z}_k),$$

where $\widetilde{\lambda}$ satisfies

$$\frac{1}{m}\sum_{k=1}^{m}\frac{\widetilde{Z}_k}{1+\widetilde{\lambda}^{\mathrm{T}}\widetilde{Z}_k}=\mathbf{0}.$$

Let $d^{(2)}=(d_1^{(2)},d_2^{(2)},\cdots,d_n^{(2)})^{\mathrm{T}}$ and $d_i^{(2)}=X_i^{(2)}-E(X_i^{(2)}|T_i)-\zeta(X_i^{(1)}-E(X_i^{(1)}|T_i))$ with $\mathrm{Var}(d_i^{(2)})=\Sigma_{X^{(2)}1}$, where ζ is a $p_2\times p_1$ matrix such that $d_i^{(2)}$ and $X_i^{(1)}-E(X_i^{(1)}|T_i)$ are uncorrelated. In order to establish asymptotic properties of the proposed empirical likelihood test, besides Assumptions A2–A3, we need:

B1 $d_i^{(2)}$ is generated by an s-variate random vector $F_i^{(2)}=(F_{i1}^{(2)},F_{i2}^{(2)},\cdots,F_{is}^{(2)})$ such that $d_i^{(2)}=\Gamma_1F_i^{(2)}$, where Γ_1 is a $p_2\times s$ matrix for some $p_2\geq s$ such that $\Gamma_1\Gamma_1^{\mathrm{T}}=\Sigma_{X^{(2)}1}$, each $F_{il}^{(2)}$ has finite 8-th moment, $E(F_i^{(2)})=0,\mathrm{Var}(F_i^{(2)})=I_s,E({F^{(2)}}_{ik}^{4})=3+\Delta_1$ and for any $\sum_{v=1}^{d}l_v\leq 8,E({F^{(2)}}_{1i_1}^{l_1}{F^{(2)}}_{2i_2}^{l_1}\cdots{F^{(2)}}_{1i_d}^{l_d})=E({F^{(2)}}_{1i_1}^{l_1})E({F^{(2)}}_{1i_2}^{l_2})\cdots E({F^{(2)}}_{1i_d}^{l_d})$, where Δ_1 is some finite constant.

B2 $p_1=O(n^{1/(2h+1)})$.

B3 $\mathrm{Var}(Y_i|X_i,T_i)<\infty$ and $\mathrm{Var}(E(X_i^{(2)}|T_i)+\zeta(X_i^{(1)}-E(X_i^{(1)}|T_i)))<\infty$ for all i; $E(d_i^{(2)}d_i^{(2)\mathrm{T}})$ is uniformly positive definite; $E(|d_{ij}^{(2)}|^{2+\delta})<\infty$ for each j for some $\delta>0$.

Remark 3.1 Assumptions B1 and B3 are similar to Assumptions A1 and A4 respectively. Assumption B2 constrains the growth rate of p_1.

Theorem 3.1 in the following gives the asymptotic property of the proposed empirical likelihood test.

Theorem 3.1 *Suppose that* $\|\frac{\widetilde{\alpha}}{\sqrt{\widetilde{\alpha}^{\mathrm{T}}\Sigma_{X^{(2)}1}\widetilde{\alpha}}}\|=O(1)$, *Assumptions* A2–A3 *and* B1–B3 *hold, as $n\to\infty$ and $p\to\infty$, and there exists some $\delta>0$ satisfies*

$$\frac{E|{X_i^{(2)}}^{\mathrm{T}}X_{i+m}^{(2)}|^{2+\delta}}{\{\mathrm{tr}(\Sigma_{X^{(2)}}^2)\}^{(2+\delta)/2}}=o(m^{\delta/2}) \tag{11}$$

and

$$\frac{E|{X_i^{(2)}}^{\mathrm{T}}\widetilde{\alpha}+{X_{i+m}^{(2)}}^{\mathrm{T}}\widetilde{\alpha}|^{2+\delta}}{\{\widetilde{\alpha}^{\mathrm{T}}\Sigma_{X^{(2)}1}\widetilde{\alpha}\}^{(2+\delta)/2}}=o(m^{\delta/2}). \tag{12}$$

Then, under H_0,

$$\widetilde{l}_m\xrightarrow{d}\chi_{1,3}^2+\widetilde{w}\chi_{1,4}^2, \tag{13}$$

where $\chi_{1,3}^2$ and $\chi_{1,4}^2$ are independent χ_1^2 variables, $\widetilde{w}=\widetilde{\alpha}^{\mathrm{T}}\Sigma_{X^{(2)}1}\widetilde{\alpha}/\widetilde{\alpha}^{\mathrm{T}}\Sigma_{X^{(2)}}\widetilde{\alpha}$.

Theorem 3.1 shows that $\widetilde{l}_m$ has an asymptotic scaled chi-square distribution. Therefore, we consider an adjusted $\widetilde{l}_m^*$:

$$\widetilde{l}_m^*=\widetilde{\xi}\widetilde{l}_m, \tag{14}$$

where $\widetilde{\xi}$ is an adjustment factor, and we will discuss it in the appendix. Theorem 3.2 in the following gives the asymptotic property of the adjusted empirical likelihood test.

Theorem 3.2 *Under the assumptions of Theorem* 3.1, *as $n\to\infty$ and $p\to\infty$, $\widetilde{l}_m^*$ converges in distribution to an asymptotic Chi-square distribution with two degrees of freedom under H_0.*

Based on Theorem 3.2, we reject H_0 when $\widetilde{l}_m^* > \chi^2_{2,b}$, where $\chi^2_{2,b}$ represents the $(1-b)$-quantile of a chi-square distribution with two degrees of freedom and b is the significant level. Similarly, to illustrate the local power of the empirical likelihood test with the nuisance parameter, we consider testing the null hypothesis $\widetilde{H}_0 : \beta^{(2)} = \beta_0^{(2)}$ against the local alternatives $\widetilde{H}_{10}$:

$$\begin{pmatrix} \left(\beta^{(2)} - \beta_0^{(2)}\right)^{\mathrm{T}} \Sigma^2_{X^{(2)}} \left(\beta^{(2)} - \beta_0^{(2)}\right) \\ \widetilde{\alpha}^{\mathrm{T}} \Sigma_{X^{(2)}1} \left(\beta^{(2)} - \beta_0^{(2)}\right) \end{pmatrix} = \frac{\widetilde{C}^{1/2}\widetilde{\gamma}}{\sqrt{m}}, \tag{15}$$

where $\widetilde{\gamma} = (\widetilde{\gamma}_1, \widetilde{\gamma}_2)^{\mathrm{T}}$, and $\widetilde{\gamma}$ is a finite constant vector,

$$\widetilde{C} = \begin{pmatrix} \sigma^4 \mathrm{tr}(\Sigma^2_{X^{(2)}}) & 0 \\ 0 & \sigma^2 \widetilde{\alpha}^{\mathrm{T}} \Sigma_{X^{(2)}1} \widetilde{\alpha} \end{pmatrix}.$$

Therefore, under the local altenatives, $\widetilde{l}_m^*$ converges in distribution to a noncentral chi-square distribution with two degrees of freedom and the noncentrality parameter $\widetilde{\varrho} = \widetilde{\gamma}^{\mathrm{T}}\widetilde{\gamma}$ as $n \to \infty$ and $p \to \infty$.

4 Simulation Study

To evaluate our method (denoted as EL) in tests for regression coefficients in high dimensional partially linear models, we conduct following simulations and compare EL and U statistic method of [2] (denoted as U). In all the following simulations, we replicate 1000 times.

Simulation I. We generate the random data from the partially linear model

$$Y_i = X_i^{\mathrm{T}}\beta + g(T_i) + \varepsilon_i, \quad i = 1, 2, \cdots, n, \tag{16}$$

where $\varepsilon_i \sim N(0,1)$, $g(T_i) = \sin(2\pi T_i)$, and $T_i \sim U(0,1)$. And we choose the degree of the B-spline as 3. We consider two kinds of correlation structures of X_i. One is the banded (denoted as Band) correlation structure where the correlation between X_{ij} and X_{ik} is $\rho_{jk} = 0.6$ if $|j-k| = 1$, $\rho_{jk} = 0.33$ if $|j-k| = 2$ and $\rho_{jk} = 0$ otherwise, where X_{ij} and X_{ik} are the jth and kth components of X_i, respectively. Another one is the auto-regressive (denoted as AR) correlation where $\rho_{jk} = \rho^{|j-k|}$ with $\rho = 0.5$. Then, we generate X_i from p-variate normal distribution $N(0, \Sigma)$, where $\Sigma = (\rho_{ij})_{p\times p}$.

The coefficient vector is $\beta = (\beta_1, \beta_2, \cdots, \beta_p)^{\mathrm{T}}$, where the $\beta_i = c_1$ if the subscript $i \le [c_2 p]$, otherwise $\beta_i = 0$. Let the parameter $c_2 = 0.75$ and 0.25, and they stand for the non-sparse case and the sparse case respectively. We set $n = 500$, and choose $p \in \{100, 300, 500, 1500, 4000\}$. We choose $c_1 = 0.01$ to calculate the power, and choose $c_1 = 0$ to calculate the type-I error rates. And we set $\alpha = (1, 1, \cdots, 1)^{\mathrm{T}} \in \mathbb{R}^p$.

Table 1 and Table 2 stand for the performance of EL and U test in Simulation I with Band and AR correlation respectively. We can see that EL test is more powerful than U test. For example, with $(n, p) = (500, 300)$ and non-sparse case under the AR correlation, the power of the EL and U test are 0.993 and 0.821 respectively.

Table 1 The empirical type-I error rates and power of EL and U at the significance level of 5% under simulation I with the Band correlation structure

(n,p)	Non-sparse case				Sparse case			
	EL		U		EL		U	
	size	power	size	power	size	power	size	power
(500,100)	0.057	0.714	0.063	0.426	0.054	0.136	0.063	0.151
(500,300)	0.048	0.998	0.056	0.752	0.064	0.339	0.063	0.216
(500,500)	0.060	1.000	0.056	0.851	0.059	0.462	0.062	0.294
(500,1500)	0.056	1.000	0.052	0.993	0.056	0.887	0.049	0.547
(500,4000)	0.064	1.000	0.065	1.000	0.062	1.000	0.066	0.793

Table 2 The empirical type-I error rates and power of EL and U at the significance level of 5% under simulation I with the AR correlation structure

(n,p)	Non-sparse case				Sparse case			
	EL		U		EL		U	
	size	power	size	power	size	power	size	power
(500,100)	0.049	0.730	0.063	0.469	0.057	0.132	0.067	0.153
(500,300)	0.062	0.993	0.065	0.821	0.063	0.310	0.062	0.247
(500,500)	0.057	1.000	0.056	0.918	0.063	0.469	0.061	0.380
(500,1500)	0.063	1.000	0.054	0.995	0.065	0.905	0.051	0.643
(500,4000)	0.056	1.000	0.060	1.000	0.061	1.000	0.064	0.868

Simulation II. We consider different choices of α in this simulation. We consider two settings of α. One is that $\alpha_i = 1$ if the subscript $i \le \lfloor c_3 p \rfloor$, and another one is that $\alpha_i = 1$ if the subscript $i \ge p - \lfloor c_3 p \rfloor$. We choose c_3 from $\{0.25, 0.5, 0.75, 1\}$. Note that we choose $c_2 = 0.25$, which stands for the sparse structure, and choose the banded correlation. And the other settings are the same as those in Simulation I.

Table 3 shows that the performance of EL test with different choices of α in Simulation II. We can see that the power of EL test is high if the choice of α is appropriate. For example, when $c_3 = 0.25$ in the first setting, the sparse structure of α is the same as the data's, and the power of EL test is highest.

Simulation III. We conduct simulation for partial tests with the presence of nuisance parameter $\beta^{(1)}$ for the linear regression model. The nuisance parameter $\beta^{(1)}$ has $p_1 = 10$ and is generated randomly from $U(0,1)$. The other settings are the same as those in Simulation I.

Table 4 and Table 5 show the results in Simulation III. We can see that EL test also has great performance in power than U test with both the Band and AR correlation in the presence of nuisance parameters.

Table 3 The power of EL with different α at the significance level of 5% under simulation III with the Band correlation structure

(n,p)	$c_3=0.25$	$c_3=0.5$	$c_3=0.75$	$c_3=1$
$\alpha_i=1$, if $i\le\lfloor c_3p\rfloor$				
$(500,100)$	0.380	0.201	0.157	0.127
$(500,300)$	0.818	0.514	0.382	0.288
$(500,500)$	0.975	0.744	0.543	0.458
$(500,1500)$	1.000	0.994	0.947	0.877
$(500,4000)$	1.000	1.000	1.000	1.000
$\alpha_i=1$, if $i\ge p-\lfloor c_3p\rfloor$				
$(500,100)$	0.066	0.052	0.058	0.123
$(500,300)$	0.060	0.065	0.062	0.296
$(500,500)$	0.056	0.059	0.065	0.459
$(500,1500)$	0.059	0.060	0.057	0.863
$(500,4000)$	0.064	0.061	0.062	0.997

Table 4 The empirical type-I error rates and power of EL and U at the significance level of 5% under simulation III with the Band correlation structure.

(n,p)	Non-sparse case				Sparse case			
	EL		U		EL		U	
	size	power	size	power	size	power	size	power
(500,100)	0.063	0.716	0.062	0.397	0.060	0.143	0.057	0.125
(500,300)	0.061	0.995	0.055	0.708	0.064	0.282	0.054	0.211
(500,500)	0.063	1.000	0.046	0.827	0.055	0.455	0.057	0.294
(500,1500)	0.047	1.000	0.053	0.986	0.060	0.886	0.051	0.510
(500,4000)	0.043	1.000	0.059	1.000	0.059	0.998	0.064	0.768

Simulation IV. We conduct simulation to evaluate adjusted EL (denoted as AEL) when X and T are not independent. Similar to the method we used in Simulation I, we generate X^* with the Band and the AR correlation, and then let $X_i = X_i^* + T_i - E(T_i)$. The other setting is the same as that in Simulation I. We consider both global tests and partial tests with the presence of nuisance parameter, where the settings of partial tests are the same as those in Simulation III.

Table 6 and Table 7 stand for the results of AEL in Simulation IV. We can see that AEL have good performance in both global tests and partial tests with the presence of nuisance parameter.

Table 5 The empirical type-I error rates and power of EL and U at the significance level of 5% under simulation III with the AR correlation structure

	Non-sparse case				Sparse case			
	EL		U		EL		U	
(n, p)	size	power	size	power	size	power	size	power
(500, 100)	0.061	0.732	0.064	0.418	0.063	0.129	0.061	0.146
(500, 300)	0.060	0.993	0.054	0.719	0.059	0.325	0.058	0.252
(500, 500)	0.062	1.000	0.065	0.848	0.057	0.447	0.066	0.335
(500, 1500)	0.063	1.000	0.062	0.988	0.062	0.895	0.048	0.532
(500, 4000)	0.057	1.000	0.055	1.000	0.061	1.000	0.063	0.833

Table 6 The empirical type-I error rates and power of AEL in the global test at the significance level of 5% under simulation IV with the Band and the AR correlation structures

	Band correlation				AR correlation			
(n, p)	Non-sparse case		Sparse case		Non-sparse case		Sparse case	
	size	power	size	power	size	power	size	power
(500,100)	0.053	0.717	0.046	0.105	0.049	0.748	0.050	0.105
(500,300)	0.058	0.991	0.046	0.264	0.056	0.994	0.063	0.281
(500,500)	0.063	0.999	0.052	0.469	0.059	1.000	0.060	0.473
(500,1500)	0.056	1.000	0.044	0.899	0.060	1.000	0.054	0.896
(500,4000)	0.061	1.000	0.057	0.999	0.064	1.000	0.055	1.000

Table 7 The empirical type-I error rates and power of AEL in the partial test at the significance level of 5% under simulation IV with the Band and the AR correlation structures

	Band correlation				AR correlation			
(n, p)	Non-sparse case		Sparse case		Non-sparse case		Sparse case	
	size	power	size	power	size	power	size	power
(500, 100)	0.054	0.710	0.063	0.128	0.056	0.728	0.060	0.131
(500, 300)	0.062	0.985	0.058	0.298	0.058	0.994	0.064	0.330
(500, 500)	0.061	1.000	0.059	0.473	0.063	1.000	0.065	0.467
(500, 1500)	0.059	1.000	0.063	0.901	0.048	1.000	0.055	0.894
(500, 4000)	0.060	1.000	0.061	0.998	0.064	1.000	0.058	0.999

Moreover, we plot the cumulative distributions of EL and AEL test statistcs under null hypothesis with the AR correlation and compare them against the standard Chi-square distribution with two degrees of freedom. We set $(n, p) = (200, 60)$ and $(500, 100)$ stand for $p < n$ cases, and $(n, p) = (200, 1600)$ and $(500, 4000)$ stand for $p \gg n$ cases. Figures 1–4 show that, the null distribution of EL is close to chi-square distribution with two degree of freedom when X and T are independent but different when X and T are not independent. From Figures 5 and 6, we can see that, when X and T are not independent, the null distribution of AEL is close to the chi-square distribution with two degree of freedom. These figures confirm the asymptotic null distributions of the proposed test statistics given in the theorems.

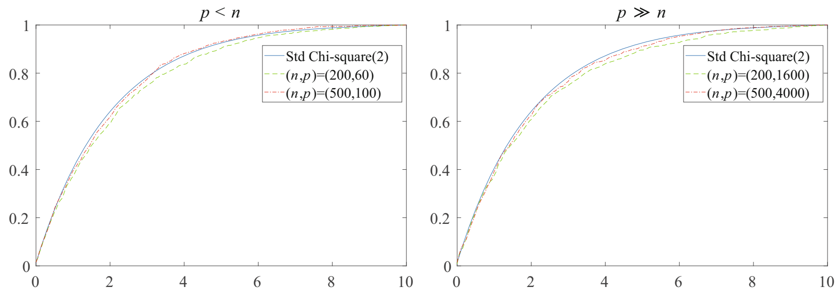

Figure 1 The EL test statistics under the null hypothesis without the presence of nuisance parameter when X and T are independent

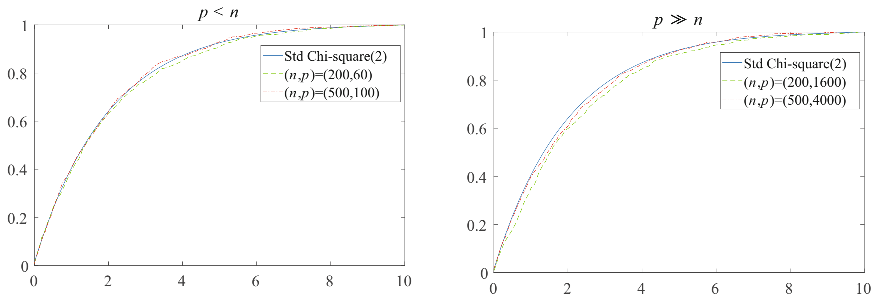

Figure 2 The EL test statistics under the null hypothesis with the presence of nuisance parameter when X and T are independent

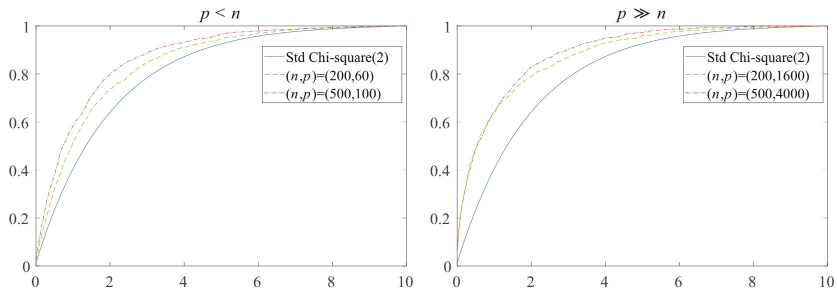

Figure 3 The EL test statistics under the null hypothesis without the presence of nuisance parameter when X and T are not independent

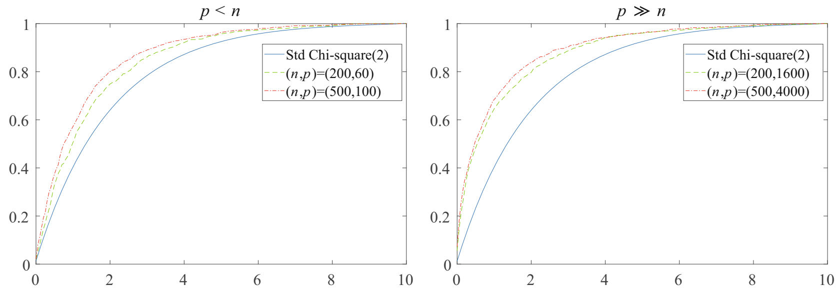

Figure 4 The EL test statistics under the null hypothesis with the presence of nuisance parameter when X and T are not independent

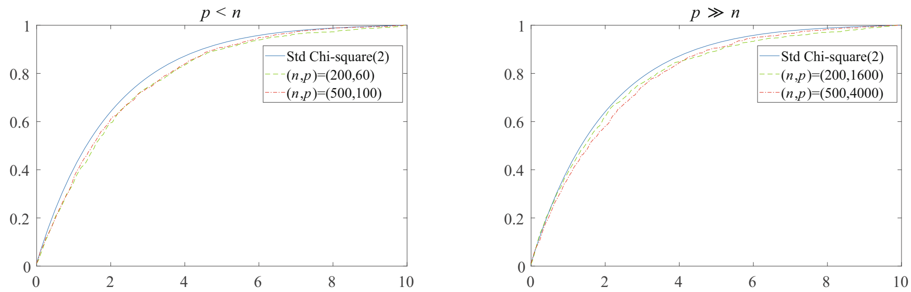

Figure 5 The adjusted EL test statistics under the null hypothesis without the presence of nuisance parameter when X and T are not independent

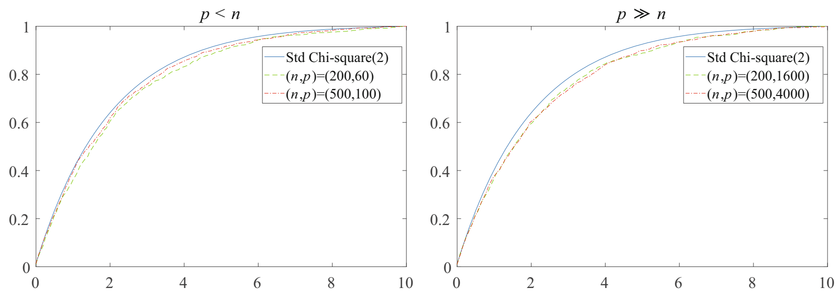

Figure 6 The adjusted EL test statistics under the null hypothesis with the presence of nuisance parameter when X and T are not independent

5 Real Data Application

The SKCM data provided by The Cancer Genome Atlas (TCGA) contains 342 cases of Breslow's depth, and the covariates of the data consist of 9 environmental variables and 20189 mRNA expression. Where Breslows depth, the response variable in the data, is considered to be a prognostic factor for cutaneous melanoma in medicine, while a larger Breslow's depth means lower survival. Nonlinear effects have been mentioned many times in biomedicine[21], and it is feasible to consider the nonlinear effects of environmental factors in genetic and environmental data. Therefore, we employ the proposed method to study whether Breslow's depth is associated with mRNA expression and environmental variables in the SKCM data.

Consider the partially linear model $Y_i = X_i^{\mathrm{T}}\beta + g(T_i) + \varepsilon_i$. Denote Y_i as the Breslow's depth (log-transformed) of ith-individual, X_i as the mRNA expression and discrete environmental variables of ith-individual, and T_i as the consecutive environmental variables of ith-individual. We remove the covariates with the missing value rate that is more than 0.15 and impute the sample mean into the missing values for the remaining covariates. We use cubic B-spline to approximate $g(T_i)$ and choose $\alpha = (1, 1, \cdots, 1)^{\mathrm{T}} \in \mathbb{R}^{p^*}$, where p^* is the number of the covariates with the missing rates less than 0.15. And then, we apply the proposed method to test $H_0 : \beta = 0$. We obtain a p-value of 0.0097, which suggests that these mRNA expression and environmental variables have a significant association with Breslow's depth.

For the partial test with nuisance parameter, we should first select the important covariates which associated with the response variable through variable selection methods such as sure independence screening[22] and iterative Lasso[23]. We randomly separate the samples into two parts, and then we use one part to select the important covariates and the other part to test. Let $\omega = (\omega_1, \omega_2, \cdots, \omega_{p^*})^{\mathrm{T}}$ be a vector of dimension p^*, where ω_i means the correlation coefficient between X_i and Y_i. Define $M = \{1 \leq i \leq p^* : |\omega_i|$ is among the first $\left\lfloor \frac{n}{\log(n)} \right\rfloor$ largest if all$\}$.

Further, we apply the iterative Lasso[23] method to select the significant covariates. Finally, we employ the proposed method to test whether the remaining covariates have effect on Breslow's depth. We repeat this process 100 times and plot a histogram of p-values. From Figure 7, we can see that most of the p-values are larger than 0.05, which suggests that the remaining covariates have no significant association with Breslow's depth.

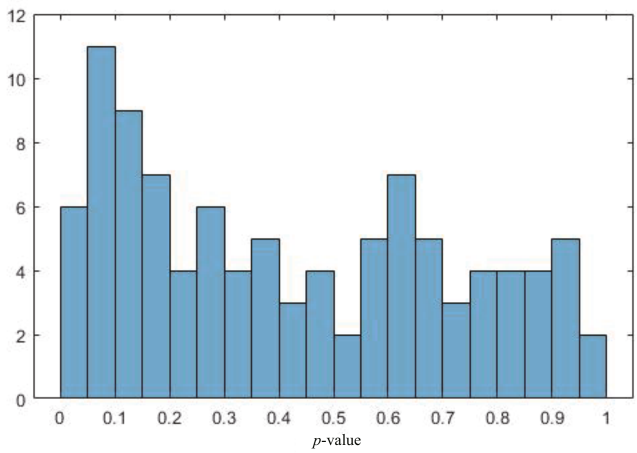

Figure 7 The p-values of the partial test with the SKCM data

6 Discussion

In this article, we consider tests for regression coefficients in high dimensional partially linear models. We approximate the unknow smooth function by the B-spline method and then propose an empirical likelihood method for testing regression coefficients. We find that, under null hypothesis, the empirical likelihood test has an asymptotic scaled chi-square distribution. Then we propose an adjusted empirical likelihood test, and establish its asymptotic properties. In addition, we extend our method to the partial test with presence of nuisance parameters. We conduct simulations which show that the proposed test has good performance both in control of type-I error rate and power. And the analysis of SKCM data indicates that the proposed method is available in global test and partial test.

Since we split the samples into two parts, we need more samples than [2] to control the type-I error rate. In simulations, we need a relatively large sample size n, while the dimension of the variables cannot be too large for computational feasibility, so that the ratio of p/n would not be too large. And, in the future work, it makes sense to improve the convergence rate of the proposed method and extend the proposed method to more complex models.

References

[1] Engle R F, Granger C W J, Rice J, et al., Semiparametric estimates of the relation between weather and electricity sales, *Journal of the American Statistical Association*, 1986, **81**(394): 310–320.

[2] Zhong P S and Chen S X, Tests for high-dimensional regression coefficients with factorial designs, *Journal of the American Statistical Association*, 2011, **106**(493): 260–274.

[3] Owen A B, Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, 1988, **75**(2): 237–249.

[4] Peng L, Qi Y C, and Wang R D, Empirical likelihood test for high dimensional linear models, *Statistics and Probability Letters*, 2014, **86**: 85–90.

[5] Zang Y G, Zhang Q Z, Zhang S G, et al., Empirical likelihood test for high dimensional generalized linear models, *Big and Complex Data Analysis*, Springer and Cham, 2017.

[6] Wang R D, Peng L, and Qi Y C, Jackknife empirical likelihood test for equality of two high dimensional means, *Statistica Sinica*, 2013, **23**(2): 667–690.

[7] Zang Y G, Zhang S G, Li Q Z, et al., Jackknife empirical likelihood test for high-dimensional regression coefficients, *Computational Statistics and Data Analysis*, 2016, **94**: 302–316.

[8] Wang Q H and Rao J N K, Empirical likelihood-based inference under imputation for missing response data, *Annals of Statistics*, 2002, **30**(3): 896–924.

[9] Xue L G and Zhu L, Empirical likelihood for single-index models, *Journal of Multivariate Analysis*, 2006, **97**(6): 1295–1312.

[10] Agarwal G G and Studden W J, Asymptotic integrated mean square error using least squares and bias minimizing splines, *The Annals of Statistics*, 1980, **8**(6): 1307–1325.

[11] Stone C J, Optimal rates of convergence for nonparametric estimators, *The Annals of Statistics*, 1980, **8**(6): 1348–1360.

[12] Stone C J, Optimal global rates of convergence for nonparametric regression, *The Annals of Statistics*, 1982, **10**(4): 1040–1053.

[13] Newey W K, Convergence rates and asymptotic normality for series estimators, *Journal of Econometrics*, 1997, **79**(1): 147–168.

[14] Linton O, Second order approximation in the partially linear regression model, *Econometrica*, 1995, **63**(6): 1079–1112.

[15] Shi P D and Li G Y, A note on the convergent rates of M-estimates for a partly linear model, *Statistics*, 1995, **26**(1): 27–47.

[16] Donald S G and Newey W K, Series estimation of semilinear models, *Journal of Multivariate Analysis*, 1994, **50**(1): 30–40.

[17] Wang S Y and Cui H J, Generalized F-test for high dimensional regression coefficients of partially linear models, *Journal of Systems Science and Complexity*, 2017, **30**(5): 1206–1226.

[18] Bai Z D and Saranadasa H, Effect of high dimension: By an example of a two sample problem, *Statistica Sinica*, 1996, **6**(2): 311–329.

[19] Owen A B, *Empirical Likelihood*, Chapman and Hall, 2001.

[20] Bravo F, Second-order power comparisons for a class of nonparametric likelihood based tests, *Biometrika*, 2003, **90**(4): 881–890.

[21] Stark A, Stahl M S, Kirchner H L, et al., Body mass index at the time of diagnosis and the risk of advanced stages and poorly differentiated cancers of the breast: Findings from a case-series study, *International Journal of Obesity*, 2010, **34**(9): 1381–1386.

[22] Fan J Q and Song R, Sure independence screening in generalized linear models with NP-dimensionality, *The Annals of Statistics*, 2010, **38**(6): 3567–3604.

[23] Huang J, Ma S G, and Zhang C H, The Iterated lasso for high-dimensional logistic regression, *The University of Iowa Department of Statistical and Actuarial Science Technical Report*, 2008.

Appendix

Without loss of generality, we assume $E(X_i) = 0$ and $n = 2m$. To prove Theorem 2.1, we need Lemma 6.1.

Lemma 6.1 *Under the assumptions of Theorem* 2.1 *and the null hypothesis, as* $n \to \infty$ *and* $p \to \infty$*, we have*

$$\left(\frac{\sum_{i=1}^m U_i}{\sqrt{m\sigma^4\mathrm{tr}(\Sigma^2)}}, \frac{\sum_{i=1}^m V_i}{\sqrt{2m\sigma^2\alpha^{\mathrm{T}}\Sigma_1\alpha}}\right)^{\mathrm{T}} \xrightarrow{d} N(0, I_2), \tag{17}$$

$$\frac{\sum_{i=1}^m U_i^2}{m\sigma^4\mathrm{tr}(\Sigma^2)} \xrightarrow{p} 1, \tag{18}$$

$$\frac{\sum_{i=1}^m V_i^2}{2m\sigma^2\alpha^{\mathrm{T}}\Sigma\alpha} \xrightarrow{p} 1, \tag{19}$$

$$\frac{\sum_{i=1}^m U_iV_i}{\sqrt{m\sigma^4\mathrm{tr}(\Sigma^2)}\sqrt{2m\sigma^2\alpha^{\mathrm{T}}\Sigma_1\alpha}} \xrightarrow{p} 0. \tag{20}$$

Proof Under null hypothesis, we could write $\sum_{i=1}^m U_i$ as

$$\begin{aligned}\sum_{i=1}^m U_i =&\sum_{i=1}^m (Y_i - X_i^{\mathrm{T}}\beta_0 - \widehat{g}(T_i))(Y_{i+m} - X_{i+m}^{\mathrm{T}}\beta_0 - \widehat{g}(T_{i+m}))X_i^{\mathrm{T}}X_{i+m}\\ =&\sum_{i=1}^m (\varepsilon_i + g(T_i) - \widehat{g}(T_i))(\varepsilon_{i+m} + g(T_{i+m}) - \widehat{g}(T_{i+m}))X_i^{\mathrm{T}}X_{i+m}\\ =&\sum_{i=1}^m \varepsilon_i\varepsilon_{i+m}X_i^{\mathrm{T}}X_{i+m} + \sum_{i=1}^m (g(T_i) - \widehat{g}(T_i))(g(T_{i+m}) - \widehat{g}(T_{i+m}))X_i^{\mathrm{T}}X_{i+m}\\ &+\sum_{i=1}^m \varepsilon_i(g(T_{i+m}) - \widehat{g}(T_{i+m}))X_i^{\mathrm{T}}X_{i+m} + \sum_{i=1}^m (g(T_i) - \widehat{g}(T_i))\varepsilon_{i+m}X_i^{\mathrm{T}}X_{i+m}\\ :=&U_1^* + U_2^* + U_3^* + U_4^*.\end{aligned}$$

It is easy to have $E(U_1^*) = 0$ and $\mathrm{Var}(U_1^*) = m\sigma^4\mathrm{tr}(\Sigma^2)$. Note that

$$\begin{aligned}E(|X_i^{\mathrm{T}}X_{i+m}|) \le&(E(X_i^{\mathrm{T}}X_{i+m})^2)^{1/2}\\ =&\mathrm{tr}^{1/2}(\Sigma^2),\end{aligned}$$

so we have

$$\begin{aligned}E(|U_2^*|) \le&\sum_{i=1}^m E\{|(g(T_i) - \widehat{g}(T_i))||(g_(T_{i+m}) - \widehat{g}(T_{i+m}))||X_i^{\mathrm{T}}X_{i+m}|\}\\ =&O(m)O(m^{-h/(2h+1)})O(m^{-h/(2h+1)})O(\mathrm{tr}^{1/2}(\Sigma^2))\\ =&o(m^{1/2}\mathrm{tr}^{1/2}(\Sigma^2)).\end{aligned}$$

Thus, $U_2^* = o_p(m^{1/2}\mathrm{tr}^{1/2}(\Sigma^2))$. Notice that

$$\begin{aligned}E(U_3^{*2}) =&E\left\{\sum_{i=1}^m\sum_{j=1}^m \varepsilon_i(g(T_{i+m}) - \widehat{g}(T_{i+m}))X_i^{\mathrm{T}}X_{i+m}\varepsilon_j(g(T_{j+m}) - \widehat{g}(T_{j+m}))X_j^{\mathrm{T}}X_{j+m}\right\}\\ =&\sigma^2 E\left\{\sum_{i=1}^m (g(T_{i+m}) - \widehat{g}(T_{i+m}))^2(X_i^{\mathrm{T}}X_{i+m})^2\right\}\\ =&o(m\mathrm{tr}(\Sigma^2)).\end{aligned}$$

Hence, $U_3^* = o_p(m^{1/2}\mathrm{tr}^{1/2}(\Sigma^2))$. And by the same technique, we could obtain $U_4^* = o_p(m^{1/2}\mathrm{tr}^{1/2}(\Sigma^2))$. Therefore, according to the central limit theorem, we have

$$\frac{\sum_{i=1}^m U_i}{\sqrt{m\sigma^4\mathrm{tr}(\Sigma^2)}} \xrightarrow{d} N(0,1).$$

And we could write $\sum_{i=1}^m V_i$ as

$$\begin{aligned}\sum_{i=1}^m V_i =&(Y - X\beta_0 - \widehat{g}(T))^{\mathrm{T}}X\alpha\\ =&\varepsilon^{\mathrm{T}}(I_{2m} - P_B)X\alpha + e^{\mathrm{T}}(I_{2m} - P_B)X\alpha\\ :=&V_1^* + V_2^*,\end{aligned}$$

where $\widehat{g}(T) = \{\widehat{g}(T_1), \widehat{g}(T_2), \cdots, \widehat{g}(T_n)\}^{\mathrm{T}}$, I_{2m} is $2m \times 2m$ identity matrix, $P_B = B(T)^{\mathrm{T}}(B(T) B(T)^{\mathrm{T}})^{-1}B(T)$ is the projection matrix of $B(T)$, and set $P_B = (b_{ij})_{2m\times 2m}$. Notice that the rank$(P_B) = O(m^{1/(2h+1)})$ under Assumption A3.

Set $\widetilde{X}_i = X_i^{\mathrm{T}}\alpha$ and $G(T_i) = E\Big(\frac{\widetilde{X}_i}{\sqrt{\alpha^{\mathrm{T}}\Sigma_1\alpha}}|T_i\Big)$. Notice that, $G(T_i) = \sum_{j=1}^p \frac{\eta_j(T_i)\alpha_i}{\sqrt{\alpha^{\mathrm{T}}\Sigma_1\alpha}}$, where α_j is the j-th component of α. With $\|\frac{\alpha}{\sqrt{\alpha^{\mathrm{T}}\Sigma_1\alpha}}\| = O(1)$, we have $G(t) \in \mathcal{H}_h$. Then, we could expand $G(T_i)$ as

$$E\bigg(\frac{\widetilde{X}_i}{\sqrt{\alpha^{\mathrm{T}}\Sigma_1\alpha}}|T_i\bigg) = B(T_i)^{\mathrm{T}}\theta + \widetilde{e}_i, \quad i = 1, 2, \cdots, n,$$

where $\theta = (\theta_1, \theta_2, \cdots, \theta_K)^{\mathrm{T}}$ is the B-spline coefficient vector, and $\widetilde{e} = (\widetilde{e}_1, \widetilde{e}_2, \cdots, \widetilde{e}_n)^{\mathrm{T}}$ with $\widetilde{e}_i = O_p(m^{-h/(2h+1)})$. So, we could decompose V_1 as

$$\begin{aligned} V_1^* =&\varepsilon^{\mathrm{T}}d\alpha - \varepsilon^{\mathrm{T}}P_Bd\alpha + (\alpha^{\mathrm{T}}\Sigma_1\alpha)^{1/2}\varepsilon^{\mathrm{T}}(I_{2m} - P_B)\widetilde{e} \\ :=&V_{1.1}^* - V_{1.2}^* + V_{1.3}^*. \end{aligned}$$

It is easy to show that $E(V_{1.1}^*) = 0$ and $\mathrm{Var}(V_{1.1}^*) = 2m\sigma^2\alpha^{\mathrm{T}}\Sigma_1\alpha$. Note that

$$\begin{aligned} E(V_{1.2}^{*2}) =&\sigma^2 E\bigg(\sum_{i,j}^{2m}\alpha^{\mathrm{T}}\Gamma F_i b_{ij} F_j^{\mathrm{T}}\Gamma^{\mathrm{T}}\alpha\bigg) \\ =&\sigma^2\alpha^{\mathrm{T}}\Sigma_1\alpha E\bigg(\sum_{i=1}^{2m} b_{ii}\bigg) \\ =&\sigma^2 E\mathrm{tr}(P_B)\alpha^{\mathrm{T}}\Sigma_1\alpha \\ =&o(m\alpha^{\mathrm{T}}\Sigma_1\alpha) \end{aligned}$$

and

$$\begin{aligned} E(V_{1.3}^{*2}) =&\sigma^2(\alpha^{\mathrm{T}}\Sigma_1\alpha)E(\widetilde{e}^{\mathrm{T}}(I_{2m} - P_B)\widetilde{e}) \\ \le&\sigma^2(\alpha^{\mathrm{T}}\Sigma_1\alpha)E(\widetilde{e}^{\mathrm{T}}\widetilde{e}) \\ =&o(m\alpha^{\mathrm{T}}\Sigma_1\alpha). \end{aligned}$$

Hence, we have $V_{1.2}^* = o_p(m^{1/2}(\alpha^{\mathrm{T}}\Sigma_1\alpha)^{1/2})$ and $V_{1.3}^* = o_p(m^{1/2}(\alpha^{\mathrm{T}}\Sigma_1\alpha)^{1/2})$.

And, we could decompose V_2 as

$$\begin{aligned} V_2^* =&e^{\mathrm{T}}(I_{2m} - P_B)d\alpha + (\alpha^{\mathrm{T}}\Sigma_1\alpha)^{1/2}e^{\mathrm{T}}(I_{2m} - P_B)\widetilde{e} \\ :=&V_{2.1}^* + V_{2.2}^*. \end{aligned}$$

Notice that

$$\begin{aligned} E(V_{2.1}^{*2}) =&\alpha^{\mathrm{T}}\Sigma_1\alpha E(e^{\mathrm{T}}(I_{2m} - P_B)e) \\ =&o(m\alpha^{\mathrm{T}}\Sigma_1\alpha). \end{aligned}$$

So, $V_{2.1}^* = o_p(m^{1/2}(\alpha^{\mathrm{T}}\Sigma_1\alpha)^{1/2})$. By the Cauchy-Schwartz inequality, we could obtain $V_{2.2}^* = o_p(m^{1/2}(\alpha^{\mathrm{T}}\Sigma_1\alpha)^{1/2})$. Therefore, we have

$$\frac{\sum_{i=1}^m V_i}{\sqrt{2m\sigma^2\alpha^{\mathrm{T}}\Sigma_1\alpha}} \xrightarrow{d} N(0, 1).$$

For any nonzero constants c_1 and c_2, it is easy to obtain that

$$E\left\{c_1\frac{\sum_{i=1}^m U_i}{\sqrt{m\sigma^4\mathrm{tr}(\Sigma^2)}}+c_2\frac{\sum_{i=1}^m V_i}{\sqrt{2m\sigma^2\alpha^{\mathrm{T}}\Sigma_1\alpha}}\right\}=0$$

and

$$\mathrm{Var}\left\{c_1\frac{\sum_{i=1}^m U_i}{\sqrt{m\sigma^4\mathrm{tr}(\Sigma^2)}}+c_2\frac{\sum_{i=1}^m V_i}{\sqrt{2m\sigma^2\alpha^{\mathrm{T}}\Sigma_1\alpha}}\right\}=c_1^2+c_2^2+o(1).$$

Therefore, (17) holds.

We could write $\frac{1}{m}\sum_{i=1}^m U_i^2$ as

$$\begin{aligned}\sum_{i=1}^m U_i^2=&\frac{1}{m}\sum_{i=1}^m(\varepsilon_i+g(T_i)-\widehat{g}(T_i))^2(\varepsilon_{i+m}+g(T_{i+m})-\widehat{g}(T_{i+m}))^2(X_i^{\mathrm{T}}X_{i+m})^2\\=&\sum_{i=1}^m\varepsilon_i^2\varepsilon_{i+m}^2(X_i^{\mathrm{T}}X_{i+m})^2+\sum_{i=1}^m(g(T_i)-\widehat{g}(T_i))^2(g(T_{i+m})-\widehat{g}(T_{i+m}))^2(X_i^{\mathrm{T}}X_{i+m})^2\\&+\sum_{i=1}^m\varepsilon_i^2(g(T_{i+m})-\widehat{g}(T_{i+m}))^2(X_i^{\mathrm{T}}X_{i+m})^2+\sum_{i=1}^m(g(T_i)-\widehat{g}(T_i))^2\varepsilon_{i+m}^2(X_i^{\mathrm{T}}X_{i+m})^2\\&+2\sum_{i=1}^m\varepsilon_i(g(T_i)-\widehat{g}(T_i))(\varepsilon_{i+m}+g(T_{i+m})-\widehat{g}(T_{i+m}))^2(X_i^{\mathrm{T}}X_{i+m})^2\\&+2\sum_{i=1}^m\varepsilon_{i+m}(g(T_{i+m})-\widehat{g}(T_{i+m}))(\varepsilon_i^2+(g(T_i)-\widehat{g}(T_i))^2)(X_i^{\mathrm{T}}X_{i+m})^2\\:=&W_1+W_2+W_3+W_4+W_5+W_6.\end{aligned}$$

We could show that

$$\frac{W_1}{m\sigma^4\mathrm{tr}(\Sigma^2)}\xrightarrow{p}1.$$

Note that

$$\begin{aligned}E(W_2)=&O(m)O(m^{-h/(2h+1)})O(m^{-h/(2h+1)})O(\mathrm{tr}(\Sigma^2))\\=&o(m\mathrm{tr}(\Sigma^2)).\end{aligned}$$

Hence, $W_2=o_p(m\mathrm{tr}(\Sigma^2))$. By the same method, we could obtain $W_3=o_p(m\mathrm{tr}(\Sigma^2))$ and $W_4=o_p(m\mathrm{tr}(\Sigma^2))$. It is easy to have $W_5=o_p(m\mathrm{tr}(\Sigma^2))$ and $W_6=o_p(m\mathrm{tr}(\Sigma^2))$. So we could show (18).

Then, we could write $\sum_{i=1}^m V_i^2$ as

$$\begin{aligned}\sum_{i=1}^m V_i^2=&\sum_{i=1}^{2m}(Y_i-X_i^{\mathrm{T}}\beta_0-\widehat{g}(T_i)^2(X_i^{\mathrm{T}}\alpha)^2\\=&\sum_{i=1}^{2m}\varepsilon_i^2(X_i^{\mathrm{T}}\alpha)^2+\sum_{i=1}^{2m}(g(T_i)-\widehat{g}(T_i))^2(X_i^{\mathrm{T}}\alpha)^2+2\sum_{i=1}^{2m}\varepsilon_i(g(T_i)-\widehat{g}(T_i))(X_i^{\mathrm{T}}\alpha)^2\\:=&W_1^*+W_2^*+W_3^*.\end{aligned}$$

It is easy to show
$$\frac{W_1^*}{m\sigma^2\alpha^{\mathrm{T}}\Sigma\alpha}\xrightarrow{p}1.$$
Note that
$$\begin{aligned}E(W_2^*)=&O(m)O(m^{-2h/(2h+1)})O(\alpha^{\mathrm{T}}\Sigma\alpha)\\=&o(m\alpha^{\mathrm{T}}\Sigma\alpha).\end{aligned}$$
Hence, $W_2^*=o_p(m\alpha^{\mathrm{T}}\Sigma\alpha)$. By the similar technique, we have $W_3^*=o_p(m\alpha^{\mathrm{T}}\Sigma\alpha)$. That is, (19) holds. Similarity, we can obtain and (20). ∎

Proof of Theorem 2.1 According to Theorem 3.2 in [19] and Lemma 6.1, we could have Theorem 2.1. ∎

Proof of Theorem 2.2 From Theorem 2.1, we have
$$l_m\xrightarrow{d}\frac{(\sum_{i=1}^m U_i)^2}{m\sigma^4\mathrm{tr}(\Sigma^2)}+\frac{(\sum_{i=1}^m V_i)^2}{2m\sigma^2\alpha^{\mathrm{T}}\Sigma\alpha},$$
and l_m has an asymptotic scaled chi-square distribution. Let
$$\xi=\left(\frac{(\sum_{i=1}^m U_i)^2}{m\sigma^4\mathrm{tr}(\Sigma^2)}+\frac{(\sum_{i=1}^m V_i)^2}{2m\sigma^2\alpha^{\mathrm{T}}\Sigma_1\alpha}\right)\Big/\left(\frac{(\sum_{i=1}^m U_i)^2}{m\sigma^4\mathrm{tr}(\Sigma^2)}+\frac{(\sum_{i=1}^m V_i)^2}{2m\sigma^2\alpha^{\mathrm{T}}\Sigma\alpha}\right).$$
And, we have
$$\begin{aligned}l_m^*=&\xi l_m\\ \xrightarrow{d}&\frac{(\sum_{i=1}^m U_i)^2}{m\sigma^4\mathrm{tr}(\Sigma^2)}+\frac{(\sum_{i=1}^m V_i)^2}{2m\sigma^2\alpha^{\mathrm{T}}\Sigma_1\alpha}\\ \xrightarrow{d}&\chi_2^2.\end{aligned}$$
∎

(18) and (19) indicate that we can use $\sum_{i=1}^m U_i^2$ and $\sum_{i=1}^m V_i^2$ as the estimators of $m\sigma^4\mathrm{tr}(\Sigma^2)$ and $2m\sigma^2\alpha^{\mathrm{T}}\Sigma\alpha$, respectively. However, the term $2m\sigma^2\alpha^{\mathrm{T}}\Sigma_1\alpha$ is unknown. We consider
$$\frac{1}{4m^2}(Y-X^{\mathrm{T}}\beta_0-\widehat{g}(T))^{\mathrm{T}}(Y-X^{\mathrm{T}}\beta_0-\widehat{g}(T))\alpha^{\mathrm{T}}X^{\mathrm{T}}(I_{2m}-P_B)X\alpha$$
as the estimator of $\sigma^2\alpha^{\mathrm{T}}\Sigma_1\alpha$. The following lemma shows that the estimator is consistent.

Lemma 6.2 *Suppose that* $\|\frac{\alpha}{\sqrt{\alpha^{\mathrm{T}}\Sigma_1\alpha}}\|=O(1)$, A1–A3 *hold, as* $n\to\infty$ *and* $p\to\infty$, *and under* H_0, *we have*
$$\frac{1}{4m^2}(Y-X^{\mathrm{T}}\beta_0-\widehat{g}(T))^{\mathrm{T}}(Y-X^{\mathrm{T}}\beta_0-\widehat{g}(T))\alpha^{\mathrm{T}}X^{\mathrm{T}}(I_{2m}-P_B)X\alpha\xrightarrow{p}\sigma^2\alpha^{\mathrm{T}}\Sigma_1\alpha. \tag{21}$$

Proof (21) will hold, if we can show
$$\frac{1}{2m}(Y-X^{\mathrm{T}}\beta_0-\widehat{g}(T))^{\mathrm{T}}(Y-X^{\mathrm{T}}\beta_0-\widehat{g}(T))\xrightarrow{p}\sigma^2 \tag{22}$$
and
$$\frac{1}{2m}\alpha^{\mathrm{T}}X^{\mathrm{T}}(I_{2m}-P_B)X\alpha\xrightarrow{p}\alpha^{\mathrm{T}}\Sigma_1\alpha. \tag{23}$$

Notice that

$$\begin{aligned}
&\frac{1}{2m}(Y - X^{\mathrm{T}}\beta_0 - \widehat{g}(T))^{\mathrm{T}}(Y - X^{\mathrm{T}}\beta_0 - \widehat{g}(T))\\
=&\frac{1}{2m}(\varepsilon + e)^{\mathrm{T}}(I_{2m} - P_B)(\varepsilon + e)\\
=&\frac{1}{2m}\varepsilon^{\mathrm{T}}(I_{2m} - P_B)\varepsilon + \frac{1}{2m}e^{\mathrm{T}}(I_n - P_B)e + \frac{1}{1m}\varepsilon^{\mathrm{T}}(I_{2m} - P_B)e\\
:=&Q_{11} + Q_{12} + Q_{13}.
\end{aligned}$$

Note that

$$\begin{aligned}
Q_{11} =&\frac{1}{2m}\varepsilon^{\mathrm{T}}\varepsilon - \frac{1}{2m}\varepsilon^{\mathrm{T}}P_B\varepsilon\\
:=&Q_{111} - Q_{112}.
\end{aligned}$$

It is easy to show that $Q_{111}/\sigma^2 \xrightarrow{p} 1$. And $E(|Q_{112}|) = \frac{1}{2m}E(\varepsilon^{\mathrm{T}}P_B\varepsilon) = \frac{\sigma^2}{n}\mathrm{tr}(P_B) = o(1)$, so $Q_{112} = o_p(1)$. Therefore, $Q_{11}/\sigma^2 \xrightarrow{p} 1$. With $e_i = O_p(m^{-h/(2h+1)})$, it is easy to have $Q_{12} = o_p(1)$ and $Q_{13} = o_p(1)$. Hence, (22) holds. And, similar to the derivation of V_1^* in Lemma 5.1, we could show (23). ▌

In the partial test. We need Lemma 6.3 to prove Theorem 3.1.

Lemma 6.3 *Under the assumptions of Theorem* 3.1 *and the null hypothesis, as $n \to \infty$ and $p \to \infty$, we have*

$$\left(\frac{\sum_{i=1}^m \widetilde{U}_i}{\sqrt{m\sigma^4\mathrm{tr}(\Sigma_{X^{(2)}}^2)}}, \frac{\sum_{i=1}^m \widetilde{V}_i}{\sqrt{2m\sigma^2\widetilde{\alpha}^{\mathrm{T}}\Sigma_{X^{(2)}1}\widetilde{\alpha}}}\right)^{\mathrm{T}} \xrightarrow{d} N(0, I_2), \tag{24}$$

$$\frac{\sum_{i=1}^m \widetilde{U}_i^2}{m\sigma^4\mathrm{tr}(\Sigma_{X^{(2)}}^2)} \xrightarrow{p} 1, \tag{25}$$

$$\frac{\sum_{i=1}^m \widetilde{V}_i^2}{2m\sigma^2\widetilde{\alpha}^{\mathrm{T}}\Sigma_{X^{(2)}}\widetilde{\alpha}} \xrightarrow{p} 1, \tag{26}$$

$$\frac{\sum_{i=1}^m \widetilde{U}_i\widetilde{V}_i}{\sqrt{m\sigma^4\mathrm{tr}(\Sigma_{X^{(2)}}^2)}\sqrt{2m\sigma^2\widetilde{\alpha}^{\mathrm{T}}\Sigma_{X^{(2)}1}\widetilde{\alpha}}} \xrightarrow{p} 0. \tag{27}$$

Proof The proof of Lemma 6.3 is similar to that of Lemma 6.1. ▌

Proof of Theorem 3.1 According to Theorem 3.2 in [19] and Lemma 6.2, we could have Theorem 3.1. ▌

Proof of Theorem 3.2 Similar to the derivation of Theorem 2.2, we set

$$\widetilde{\xi} = \left(\frac{(\sum_{i=1}^m U_i)^2}{m\sigma^4\mathrm{tr}(\Sigma_{X^{(2)}}^2)} + \frac{(\sum_{i=1}^m V_i)^2}{2m\sigma^2\widetilde{\alpha}^{\mathrm{T}}\Sigma_{X^{(2)}1}\widetilde{\alpha}}\right)\Big/\left(\frac{(\sum_{i=1}^m U_i)^2}{m\sigma^4\mathrm{tr}(\Sigma_{X^{(2)}}^2)} + \frac{(\sum_{i=1}^m V_i)^2}{2m\sigma^2\widetilde{\alpha}^{\mathrm{T}}\Sigma_{X^{(2)}}\widetilde{\alpha}}\right).$$

Then, we have

$$\begin{aligned}\widetilde{l}_m^* =& \widetilde{\xi}\widetilde{l}_m \\ \xrightarrow{d} & \frac{(\sum_{i=1}^m U_i)^2}{m\sigma^4 \mathrm{tr}(\Sigma_{X^{(2)}}^2)} + \frac{(\sum_{i=1}^m V_i)^2}{2m\sigma^2 \widetilde{\alpha}^{\mathrm{T}} \Sigma_{X^{(2)}1} \widetilde{\alpha}} \\ \xrightarrow{d} & \chi_2^2.\end{aligned}$$

∎

From (25) and (26), we can use $\sum_{i=1}^m \widetilde{U}_i^2$ and $\sum_{i=1}^m \widetilde{V}_i^2$ as the estimators of $m\sigma^4 \mathrm{tr}(\Sigma_{X^{(2)}}^2)$ and $2m\sigma^2 \alpha^{\mathrm{T}} \Sigma_{X^{(2)}} \alpha$, respectively. And we consider

$$\frac{1}{4m^2}(Y - X^{\mathrm{T}}\beta_0 - \widetilde{g}(T))^{\mathrm{T}}(Y - X^{\mathrm{T}}\beta_0 - \widetilde{g}(T))\widetilde{\alpha}^{\mathrm{T}} X^{(2)\mathrm{T}}(I_{2m} - P_B)X^{(2)}\widetilde{\alpha}$$

as the estimator of $\sigma^2 \widetilde{\alpha}^{\mathrm{T}} \Sigma_{X^{(2)}1} \widetilde{\alpha}$. The following lemma shows that the estimator is consistent.

Lemma 6.4 *Suppose that* $\|\frac{\widetilde{\alpha}}{\sqrt{\widetilde{\alpha}^{\mathrm{T}} \Sigma_{X^{(2)}1} \widetilde{\alpha}}}\| = O(1)$, *Assumptions* A2–A3 *and* B1–B2 *hold, as* $n \to \infty$ *and* $p \to \infty$, *and under* H_0, *we have*

$$\frac{1}{4m^2}(Y - X^{\mathrm{T}}\beta_0 - \widetilde{g}(T))^{\mathrm{T}}(Y - X^{\mathrm{T}}\beta_0 - \widetilde{g}(T))\widetilde{\alpha}^{\mathrm{T}} X^{(2)\mathrm{T}}(I_{2m} - P_B)X^{(2)}\widetilde{\alpha} \xrightarrow{p} \sigma^2 \alpha^{\mathrm{T}} \Sigma_1 \alpha. \quad (28)$$

Proof The proof of Lemma 6.4 is similar to that of Lemma 6.2. ∎