Electronic Journal of Statistics
Vol. 11 (2017) 1491–1526
ISSN: 1935-7524
DOI: 10.1214/17-EJS1257

An adaptive-to-model test for parametric single-index models with missing responses*

Cuizhen Niu†

School of Statistics, Beijing Normal University, Beijing
e-mail: nczlbc_890@126.com

and

Lixing Zhu‡

School of Statistics, Beijing Normal University, Beijing
Department of Mathematics, Hong Kong Baptist University, Hong Kong
e-mail: lzhu@hkbu.edu.hk

Abstract: This paper is devoted to implementing model checking for parametric single-index models with missing responses at random. Two dimension reduction adaptive-to-model tests applying to the missing responses situation are proposed. Unlike the existing smoothing tests, our methods can greatly alleviate the curse of dimensionality in the sense that the tests behave like a test with only one covariate. It results in better significance level maintenance and higher power than the classical tests. The finite sample performance is evaluated through several simulation studies and a comparison with other popularly used tests. A real data analysis is conducted for illustration.

MSC 2010 subject classifications: Primary 62G10, 62H15; secondary 62H10.
Keywords and phrases: Adaptive-to-model test, dimension reduction, missing responses at random.

Received November 2016.

Contents

*Corresponding author: Lixing Zhu. The authors thank the Editor, the Associate Editor and referees for their constructive comments and suggestions, which led to the substantial improvement of an early manuscript.

†Cuizhen Niu's research was supported by the Fundamental Research Funds for the Central Universities and China Postdoctoral Science Foundation (2016M600951).

‡Lixing Zhu's research was supported by Natural Science Foundation of China (11671042) and a grant from the University Grants Council of Hong Kong.

1. Introduction

Consider the following parametric single-index model

$$Y = g(\beta^\top X, \theta) + \varepsilon, \tag{1.1}$$

where Y is the scalar response variable, X represents the covariate vector of dimension p, β and θ are the parameter vectors of dimensions p and d, respectively and $E(\varepsilon|X) = 0$. Besides, $g(\cdot)$ is a known link function and the superscript $\top$ denotes transposition. Throughout this paper, we focus on the situation where some of the response measurements are missing and all of the covariates are observable completely. For this model, it is assumed that Y is missing at random (MAR) and δ is the missing indicator for the individual whether Y is observed ($\delta = 1$) or not ($\delta = 0$). Thus, an incomplete sample $\{(y_i, \delta_i, x_i), 1 \le i \le n\}$ is obtained. The MAR assumption implies that δ and Y are conditionally independent given X, that is,

$$P(\delta = 1|Y, X) = P(\delta = 1|X) = \pi(X),$$

which is often presumed for statistical analysis in the presence of missing data and is reasonable in many practical situations, see Little and Rubin [22] for detail. For the recent literature on the analysis of missing data, see for instance, Kim and Shao [13].

In practice, it is not uncommon that some of response measurements are unobservable due to various reasons. The readers can refer to Little and Rubin [22]. In order to prevent wrong conclusion and to avoid model misspecification, performing an effective model checking before further statistical analysis is necessary. This is to ascertain whether the real dataset marries the hypothetical parametric model or not. In recent decade, there are a number of proposals available in the literature. Among others, Sun et al [28] built two empirical process-based tests, that is, an imputation-based test and a marginal inverse probability weighted test, for examining the adequacy of partial linearity of the model with missing responses at random. With two completed data sets based on imputation and marginal inverse probability weighted methods, respectively, Xu and Zhu [32] constructed two empirical process-based tests to check the adequacy of varying coefficient models with missing responses. All of the above tests are empirical process-based tests, which rely on the sample averages of residuals with respect to an index set to construct empirical processes. This kind of test is also called global smoothing test because it is usually the average of weighted averages of residuals and average is a kind of overall smoothing. The superiority of these tests is that the convergence rate of empirical process-based tests to their weak limits is $n^{-1/2}$, which is the fastest possible rate the existing methods can achieve, where n is the sample size. The tests in this class are usually not asymptotically distribution-free when the dimension of X is larger than 1, and Monte Carlo approximation such as the bootstrap is required for critical value determination. Further, such kinds of tests are not very sensitive to high frequency alternatives. Also, the dimension of X still has a significant negative impact for their significance level maintenance and power performance (e.g. Guo et al [6]).

Another class of methods to carry out model checking are local smoothing tests, which involve nonparametric smoothing methods in estimation. The main advantage of local smoothing tests is that they are sensitive to high frequency alternative models. For example, González-Manteiga and Pérez-González [4] constructed a test that is based on the L_2 distance between the nonparametric and parametric fits with missing response data. Guo et al [7] extended Zheng's [33] test to carry out model checking for parametric regressions with missing responses. Niu et al [25] suggested a misspecification test of the nonparametric component in a partially linear model with responses missing at random. Li [21] proposed a lack-of-fit test, which is based on minimum integrated square distances between the nonparametric and parametric fits and can be regarded as an extension of the minimum distance test suggested by Koul and Ni [14] to handle missing responses. Nevertheless, a serious problem causing them not to be powerful against the alternatives is the typical slow convergence rate $O(n^{-1/2}h^{-p/4})$ of the test statistics to their weak limits when p is large. Also, only the local alternatives that are distinct from the null models at the same rate $O(n^{-1/2}h^{-p/4})$ are detectable by these tests. Here p is the dimension of covariate vector X and h is the bandwidth in kernel estimation tending to zero at a certain rate. This class of tests inevitably suffer from the curse of dimensionality since the multivariate nonparametric estimation must be applied, which leads to the significance level is hard to maintain and less powerful to detect alternatives even in the situations with moderate dimensions.

The aforementioned difficulties drive researchers to make various modifications of the existing testing techniques in order to solve the dimensionality problem. A commonly used method to construct tests is based on projection pursuit, which was first proposed by Friedman and Stuetzle [3]. The method is essentially to find one or a few directions along which the departure can be easily detected. When all of the responses are available, the amendment for local smoothing tests include the work by Lavergne and Patilea [16, 17]. For single-index models, Maistre and Patilea [24] proposed a new kernel smoothing-based approach to conduct the goodness-of-fit test, where a convergence rate of $nh^{1/2}$ is achieved. Lavergne et al [15] considered a kernel smoothing-based test that smoothes only over the covariates appearing the null hypothesis, which mitigated the curse of dimensionality. Maistre et al [23] addressed the issue of lack-of-fit testing for a parametric quantile regression and they proposed a test involving one-dimensional kernel smoothing, whose convergence rate is free of the number of covariates. As for empirical process-based tests when projection approach is applied, examples include Zhu and An [38], Stute et al [27], Zhu [37], Escanciano [2], Stute et al [26]. All of the above tests must resort to Monte Carlo techniques to determine critical values even though some of them have the limiting null distributions. This way is computationally intensive and time consuming when all of projections are involved. Xia [31] also proposed a projection-based test that however has no way to control type I error. Recently, a dimension reduction adaptive-to-model test was proposed by Guo et al [6]. It is also a local smoothing test, not only enjoying the fastest possible convergence rate $O(n^{-1/2}h^{-1/4})$ that the existing local smoothing tests can achieve but also being able to detect the local alternatives at this rate $O(n^{-1/2}h^{-1/4})$ as if X were univariate. Further their test can utilize the limiting null distribution to determine the critical values without the assistance from Monte Carlo approximation, which significantly reduces the computational burden.

To the best of our knowledge, few research works could be effectively to handle the curse of dimensionality in goodness-of-fit testing with missing responses. In the present paper, we intend to incorporate the advantages of Guo et al's [6] test into the test we will propose. It is not a trivial combination as we will see later, the model-adaptation step requires a different approach than that used in Guo et al [6]. More specifically, in the construction of test statistic, we must take the missingness into consideration. Such a combination thus makes the test be adaptive-to-model as well as handle the missingness. For this purpose, we in Section 2.1 incorporate the missing indicator δ and the function $\pi(X)$ to construct the test statistic. To make every step of construction consistently reduce the curse of dimensionality, we use a semiparametric approach to estimate the parameters of interest and the function in $\delta = h(\gamma_0^\top X) + \varsigma$ and $\pi(X) = E(\delta|X) = E(\delta|\gamma_0^\top X)$, where $E(\varsigma|X) = 0$, $h(\cdot)$ is an unknown link function and γ_0 is an unknown p-dimensional parameter vector. Further to identify and estimate the matrix B, the so-called complete case assisted recovery (CCAR) is adopted in Section 2.2. The model (1.1) is adopted as the hypothetical model and the alternative model is as follows

$$Y = m(B^\top X) + \varepsilon, \tag{1.2}$$

where $E(\varepsilon|X) = 0$, the link function $m(\cdot)$ is unknown and the response Y is missing at random. B is a $p \times q$ orthonormal matrix with q orthogonal columns where $B^\top B = I_q$, $1 \leq q \leq p$ and q is the structural dimension of the matrix B. When $q = p$, the above model becomes a purely nonparametric model. This shows the generality of the model.

Two test statistics will be constructed. Under the null hypothesis, the tests solely utilize the dimension reduction structure such that the dimensionality is largely avoided and then the significance level can be well maintained, and under the alternative hypothesis, they can automatically adapt to the related model structure to enjoy the omnibus property. A key step for this is adaptively estimating the matrix B and its number of columns q such that under the null, the estimate is consistent to $c\beta$ for a constant s and under the alternative to the general matrix BC for a $q \times q$ orthonormal matrix C. To achieve this goal, the complete case assisted recovery (CCAR) method proposed by Guo et al [5] and a BIC-type criterion are respectively applied to estimate these two vector and matrix.

The rest materials of the paper are organized as follows. In Section 2, two tests are constructed. The CCAR method to estimate the matrix B and the BIC-type criterion to identify the structural dimension q are also presented in this section. The asymptotic properties under the null hypothesis, local alternatives and global alternatives are stated in Section 3. The simulation results are reported in Section 4 and a real data analysis is reported in Section 5. Some discussions are given in Section 6. Technical proofs are postponed to the Appendix.

2. Test Statistics Construction

In this section, the following hypotheses of interest are:

$$\begin{aligned} &H_0 : P(E(Y|X) = g(\beta_0^\top X, \theta_0)) = 1 \text{ for some } \beta_0 \in R^p, \theta_0 \in R^d; \\ &H_1 : P(E(Y|X) = g(\beta^\top X, \theta)) < 1 \text{ for any } \beta \in R^p, \theta \in R^d, \end{aligned} \tag{2.1}$$

where $g(\cdot)$ is a known link function whereas $m(\cdot)$ is an unknown link function.

2.1. Test statistics

Denote $e = Y - g(\beta_0^\top X, \theta_0)$. Thus under the MAR assumption, we have

$$E(\delta e|B^\top X) \quad = \quad E(eE(\delta|X, Y)|B^\top X) = \pi(X)E(e|B^\top X) \tag{2.2}$$

and

$$E\Big\{\frac{\delta}{\pi(X)}e|B^\top X\Big\} = E\Big[eE\Big\{\frac{\delta}{\pi(X)}|X, Y\Big\}|B^\top X\Big] = E(e|B^\top X). \tag{2.3}$$

When the null hypothesis H_0 holds, the model (1.2) reduces to the model (1.1) with $q = 1$ and $B = \tilde{\beta} = c\beta_0$ for a constant c. Therefore, under H_0, we have

$$E(e|X) = 0 \Rightarrow E(e|\beta_0^\top X) = E(e|B^\top X) = 0.$$

Further, combining (2.2) with (2.3), the following two formulae can be derived

$$E\{\delta eE(\delta e|B^\top X)f(B^\top X)\} = E\{E^2(\delta e|B^\top X)f(B^\top X)\} = 0 \tag{2.4}$$

and

$$E\Big\{\frac{\delta}{\pi(X)}eE\Big\{\frac{\delta}{\pi(X)}e|B^\top X\Big\}f(B^\top X)\Big\} = E\Big\{E^2\Big\{\frac{\delta}{\pi(X)}e|B^\top X\Big\}f(B^\top X)\Big\} = 0, \tag{2.5}$$

where $f(B^\top X)$ is the probability density function of $B^\top X$. Under the alternative hypothesis H_1, we have

$$E(e|B^\top X) = E(Y|B^\top X) - g(\beta_0^\top X, \theta_0) = m(B^\top X) - g(\beta_0^\top X, \theta_0) \neq 0.$$

Consequently, when H_1 holds,

$$E\{\delta e E(\delta e|B^\top X) f(B^\top X)\} > 0$$

and

$$E\Big\{\frac{\delta}{\pi(X)} e E\Big\{\frac{\delta}{\pi(X)} e|B^\top X\Big\} f(B^\top X)\Big\} > 0.$$

From the above analysis, the null hypothesis H_0 holds if and only if any one of the two equations (2.4) and (2.5) is zero. When the alternative hypothesis H_1 holds, they are both positive. Based on their different performance, the sample analogues of the left hand sides in these two formulae can then be applied to construct test statistics.

Let (y_i, x_i, δ_i), $i = 1, \ldots, n$ be an i.i.d. sample from (Y, X, δ). In order to construct test statistics, the estimates of $E(\delta e|B^\top X)$ and $E\{\frac{\delta}{\pi(X)} e|B^\top X\}$ are needed. Once an estimate of $\hat{B}(\hat{q})$ is available, the kernel estimates of these two quantities can be obtained as follows:

$$\hat{E}(\delta e|\hat{B}(\hat{q})^\top x_i) = \frac{\frac{1}{n-1}\sum_{j\neq i}^n \delta_j \hat{e}_j \mathcal{K}_h\{\hat{B}(\hat{q})^\top (x_i - x_j)\}}{\hat{f}(\hat{B}(\hat{q})^\top x_i)},$$

$$\hat{E}\Big\{\frac{\delta}{\pi(X)} e|\hat{B}(\hat{q})^\top x_i\Big\} = \frac{\frac{1}{n-1}\sum_{j\neq i}^n \frac{\delta_j}{\hat{\pi}(x_j)} \hat{e}_j \mathcal{K}_h\{\hat{B}(\hat{q})^\top (x_i - x_j)\}}{\hat{f}(\hat{B}(\hat{q})^\top x_i)},$$

where $\mathcal{K}_h(\cdot) = \mathcal{K}(\cdot/h)/h^{\hat{q}}$ with $\mathcal{K}(\cdot)$ being the $\hat{q}$-dimensional kernel function, h being the bandwidth, and $\hat{e}_j = y_j - g(\hat{\beta}^\top x_j, \hat{\theta})$ with $\hat{\alpha} = (\hat{\beta}, \hat{\theta})$ being the least square estimators of $\alpha_0 = (\beta_0, \theta_0)$. The formula is

$$(\hat{\beta}^\top, \hat{\theta}^\top)^\top = \arg\min_{\beta,\theta} \sum_{i=1}^n \frac{\delta_i}{\hat{\pi}(x_i)} \Big\{y_i - g(\beta^\top x, \theta)\Big\}^2.$$

The estimate $\hat{B}(\hat{q})$ is an sufficient dimension reduction (SDR) estimate of the matrix B with an estimated structural dimension $\hat{q}$ of q. Since the estimates of B and q are critical for our tests, we will specify them in a separate subsection. The probability density function $f(B^\top X)$ of $B^\top X$ can be estimated by

$$\hat{f}(\hat{B}(\hat{q})^\top x_i) = \frac{1}{n-1}\sum_{j\neq i}^n \mathcal{K}_h\{\hat{B}(\hat{q})^\top (x_i - x_j)\}.$$

Further, two quadratic conditional moment test statistics are constructed as follows:

$$T_n = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \delta_i \delta_j \mathcal{K}_h\{\hat{B}(\hat{q})^\top (x_i - x_j)\} \hat{e}_i \hat{e}_j, \qquad (2.6)$$

$$V_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \frac{\delta_i}{\hat{\pi}(x_i)} \frac{\delta_j}{\hat{\pi}(x_j)} \mathcal{K}_h\{\hat{B}(\hat{q})^\top (x_i - x_j)\} \hat{e}_i \hat{e}_j. \tag{2.7}$$

In general, the function $\pi(X)$ is unknown and needs to be estimated. Throughout our paper, we assume $\pi(X)$ follows a semi-parametric structure, that is, $\delta = h(\gamma_0^\top X) + \varsigma$ and $\pi(X) = E(\delta|X) = E(\delta|\gamma_0^\top X)$, where $E(\varsigma|X) = 0$, $h(\cdot)$ is an unknown link function and γ_0 is an unknown p-dimensional parameter vector. Here, we assume that the structural dimension of the central subspace $\mathcal{S}_{\delta|X}$ equals 1 as assumed in Guo et al [5]. It is a semiparametric assumption that can avoid greatly the curse of dimensionality and is weaker than the assumption applied in Li and Lu [20] and Zhu et al [34], in which they employed parametric models for the selection probability $P(\delta = 1|X)$.

When a sample $\{(y_i, x_i, \delta_i),\ i = 1, \ldots, n\}$ is available, noticing that δ is a binary variable, we consider obtaining a consistent estimate $\hat{\gamma}$ of γ_0 via sliced inverse regression (SIR) as it is very easy to implement. Li [19], Hsing and Carroll [10], Zhu and Ng [41] showed the $\sqrt{n}$-consistency of the SIR estimate. A commonly used condition for SIR to work is the linearity condition, that is, $E(X|\gamma_0^\top X = x)$ is linear in x, here γ_0 is a base vector of $S_{\delta|X}$. Given this condition, it can be obtained that $\mathrm{Span}(\Sigma_X^{-1}\Sigma_{E(X|\delta)}) \subseteq S_{\delta|X}$, where $\Sigma_X = \mathrm{Cov}(X) \in R^{p\times p}$ and $\Sigma_{E(X|\delta)} = \mathrm{Cov}\{E(X|\delta)\} \in R^{p\times p}$. In general case, when the range of $\delta_i,\ i = 1, \ldots, n$ is divided into H slices, $I_1, \ldots, I_H$, the SIR estimate is a weighted sum of local covariances of x_i whose concomitant δ_i lies in each slice. In binary variable case the number of slices H naturally equals 2, $\mathrm{Cov}\{E(X|\delta)\}$ can be estimated by

$$\widehat{\mathrm{Cov}}\{E(X|\delta)\} = \sum_{h=1}^{2} \hat{p}_h (\hat{m}_h - \bar{x})(\hat{m}_h - \bar{x})^\top,$$

where $\bar{x} = \sum_{i=1}^n x_i/n$, $\hat{p}_h$ is the proportion of $\delta_i,\ i = 1, \ldots, n$ that fall in the hth slice and $\hat{m}_h,\ h = 1, 2$ is the sample mean of x_i within each slice. Thus, the eigenvector that is associated with the largest eigenvalues of $\hat{\Sigma}_X^{-1}\widehat{\mathrm{Cov}}\{E(X|\delta)\}$ can be an estimate $\hat{\gamma}$, where $\hat{\Sigma}_X = n^{-1}\sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^\top$. To be precise,

$$\hat{\gamma} = \hat{\Sigma}_X^{-1}(\bar{x}_{obs} - \bar{x}_{miss})^\top,$$

here $\bar{x}_{obs}$ and $\bar{x}_{miss}$ are the sample means of X corresponding to $\delta = 1$ and $\delta = 0$, respectively. the function $\pi(x_i)$ can be estimated by

$$\hat{\pi}(x_i) = \hat{E}(\delta|x_i) = \frac{\sum_{j\neq i}^n \delta_j \tilde{\mathcal{K}}_h\{\hat{\gamma}^\top (x_i - x_j)\}}{\sum_{j\neq i}^n \tilde{\mathcal{K}}_h\{\hat{\gamma}^\top (x_i - x_j)\}}, \tag{2.8}$$

where $\tilde{\mathcal{K}}_h(\cdot) = \tilde{\mathcal{K}}(\cdot/h)/h$ with $\tilde{\mathcal{K}}(\cdot)$ being the one-dimensional kernel function and $\tilde{\mathcal{K}}(\cdot)$ can be different from $\mathcal{K}(\cdot)$ in (2.6).

Remark 1. *When the dimension reduction structure is not fully used, Guo et al. [7] constructed two tests as*

$$T_n^\star = \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}^n \delta_i\delta_j\mathcal{K}_h^\star(x_i-x_j)\hat{e}_i\hat{e}_j,$$
$$V_n^\star = \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}^n \frac{\delta_i}{\hat{\pi}^\star(x_i)}\frac{\delta_j}{\hat{\pi}^\star(x_j)}\mathcal{K}_h^\star(x_i-x_j)\hat{e}_i\hat{e}_j, \tag{2.9}$$

where $\mathcal{K}_h^\star(\cdot) = \mathcal{K}^\star(\cdot/h)/h^p$ with $\mathcal{K}^\star(\cdot)$ being a p-dimensional multivariate kernel function and

$$\hat{\pi}^\star(x_i) = \frac{\sum_{l=1}^n \delta_l\mathcal{K}_h^\star(x_i-x_l)}{\sum_{l=1}^n \mathcal{K}_h^\star(x_i-x_l)}.$$

Compared our tests in (2.6) and (2.7) with the above, the difference is between the used covariates x_i and $\hat{B}(\hat{q})^\top x_i$. Note that the dimension of x_i is always p whereas the dimension of $\hat{B}(\hat{q})^\top x_i$ automatically adapts to the null and alternative hypothesis as 1 and q respectively. In other words, the tests under the null hypothesis only involves one-dimensional covariate and suffers much less from the curse of dimensionality when we need to estimate relevant functions. While under the alternative hypothesis, q-dimensional covariate is used so that the tests can be sensitive to the alternative models. Therefore, to derive the asymptotic properties of the test statistics under the null hypothesis, the standardizing constant can be $n^{1/2}h^{1/4}$ rather than $n^{1/2}h^{p/4}$ that is used for the classical tests in (2.9). This also helps enhancing the power performance of the tests in (2.6) and (2.7) because $n^{1/2}h^{1/4}$ diverges to infinity much faster than $n^{1/2}h^{p/4}$ such that our tests can diverge to infinity faster under the alternative hypothesis.

2.2. Identification and estimation of B

In this section, we first specify the estimate of the matrix B under given q and then study how to identify q consistently. As pointed out by Li [19], even with complete data, the matrix B is not identifiable since $m(B^\top X)$ can be rewritten as $\tilde{m}(C^\top B^\top X)$ for any $q\times q$ orthonormal matrix C. However, identifying BC for a $q\times q$ orthonormal matrix C is sufficient for the problem studied herewith. See Guo et al. [6] for details. Thus, we will still use B in the lieu of BC without notational confusion throughout the rest of this paper. Therefore, we can identify the mean central subspace $\mathcal{S}_{E(Y|X)}$ spanned by B through sufficient dimension reduction (SDR) technique and q is the structural dimension of $\mathcal{S}_{E(Y|X)}$. When there exist responses missing at random, we adopt the methods proposed by Guo et al. [5], including selection probability assisted recovery (SPAR) and complete case assisted recovery (CCAR). Since the semiparametric dimension reduction proposal with CCAR is easy to implement and possesses good performance in general, we only focus on this with a brief review below.

The semiparametric dimension reduction method with CCAR is a two-stage method. The main idea of this procedure is: First, obtain an estimate $\hat{\psi}$ for a base matrix $\psi\in R^{p\times q}$ of $\mathcal{S}_{Y|X}$ from the CC analysis with, say, sliced inverse regression (SIR) proposed by Li [19]. Second, impute missing responses through the conditional distributions of Y given $\hat{\psi}^\top X$; Finally, based on the imputed data to define a more accurate SIR estimate of ψ, denoted as $\hat{\psi}^{CCAR}$. To ensure that $\mathcal{S}_{Y|X} = \mathcal{S}_{E(Y|X)}$ for ε in the model (1.2), we need to assume that $\varepsilon = m_1(B^\top X)\tilde{\varepsilon}$, where $m_1(\cdot)$ is an unknown smooth function and $\tilde{\varepsilon} \perp\!\!\!\perp X$. A final estimate of B is $\hat{B} = \hat{\psi}^{CCAR}$.

The detailed steps for obtaining an estimate of the base matrix $\hat{B} = \hat{\psi}^{CCAR}$ is listed as follows:

Step 1: Obtain an initial base matrix by the SIR estimate $\hat{\psi}$ for $\mathcal{S}_{Y|X}$ with the CC sample.

Step 2: Compute $\hat{p}_{ij}^{CC}$ as

$$\hat{p}_{ij}^{CC} = \frac{\delta_j \mathcal{K}_h\{\hat{\psi}^\top (x_j - x_i)\}}{\sum_{j=1}^n \delta_j \mathcal{K}_h\{\hat{\psi}^\top (x_j - x_i)\}}.$$

Step 3: Divide the range of Y into S slices, $I_1, I_2 \dots I_s \dots, I_S$ and denote $\hat{p}_s^{CCAR}$, $s = 1, \dots, S$ as the proportion of y_i falling in the sth slice as

$$\hat{p}_s^{CCAR} = \frac{1}{n} \sum_{i,j=1}^n \Big\{ [(1-\delta_i)\hat{p}_{ij}^{CC} + \delta_i I(i=j)] I(y_j \in I_s) \Big\}.$$

Further, estimate the sample mean of X within each slice, $\hat{m}_s^{CCAR}$, as

$$\hat{m}_s^{CCAR} = \frac{1}{n} \sum_{i,j=1}^n \Big\{ [(1-\delta_i)\hat{p}_{ij}^{CC} + \delta_i I(i=j)] I(y_j \in I_s) \Big\} \times x_i \times \frac{1}{\hat{p}_s^{CCAR}}.$$

Step 4: Estimate $\mathrm{Cov}\{E(X|Y)\}$ by

$$\widehat{\mathrm{Cov}}\{E(X|Y)\} = \sum_{s=1}^S \hat{p}_s^{CCAR} (\hat{m}_s^{CCAR} - \bar{x})(\hat{m}_s^{CCAR} - \bar{x})^\top.$$

Thus, the q eigenvectors corresponding to q largest eigenvalues of the matrix $\Sigma_X^{-1}\mathrm{Cov}\{E(X|Y)\}$ can be regarded as $\hat{B}$, an estimate of the matrix B.

Under the conditions presented in the Appendix, Guo et al. [5] showed the $\sqrt{n}$ consistency of $\hat{B}$ when the structural dimension q is given in advance. See Guo et al. [5] for more discussions on this method.

2.3. Structural dimension determination

To accommodate the alternative hypothesis, we need to estimate q consistently. For this purpose, Zhu et al. [40] first introduced an approach of Bayesian information criterion (BIC) type. Zhu et al. [35, 36] further considered a modified BIC-type criterion. Based on their theory, the structural dimension q can be determined by

$$\hat{q} = \arg\max_{l=1,\dots,p} \left[\frac{n}{2} \times \frac{\sum_{i=1}^l \{\log(\hat{\lambda}_i + 1) - \hat{\lambda}_i\}}{\sum_{i=1}^p \{\log(\hat{\lambda}_i + 1) - \hat{\lambda}_i\}} - 2 \times \sqrt{n} \times \frac{l(l+1)}{2p} \right], \quad (2.10)$$

where $\hat\lambda_1 \geq \hat\lambda_2 \geq \ldots \geq \hat\lambda_p \geq 0$ are the eigenvalues of $\Sigma_X^{-1}\mathrm{Cov}\{E(X|Y)\}$. It is clear that the first term in the bracket can be regarded as the likelihood ratio in the classical BIC and the second term is the penalty term with $l(l+1)/2$ free parameters.

Proposition 1. *Assume the estimate $\hat B$ based on semiparametric dimension reduction with CCAR is $\sqrt{n}$ consistent to the matrix B and under the conditions in the Appendix, the BIC-type estimate $\hat q = q$ as $n \to \infty$ with a probability going to one. Therefore, for a $q \times q$ orthogonal matrix C, $\hat B(\hat q)$ is a consistent estimate of B.*

Proposition 1 shows that under the null hypothesis H_0 and global alternative H_1, $\hat q = 1$ or q respectively.

3. Asymptotic Properties

In this section, the limiting null distributions of the statistics T_n in (2.6) and V_n in (2.7) are derived and the relevant asymptotic properties under local alternative and global alternative hypothesis are also investigated.

3.1. Limiting null distribution

To state the theorems, we first introduce some notations that are useful to define the limiting variances of the proposed test statistics. Denote

$$\Sigma^T = 2\int \mathcal{K}^2(u)du \cdot \int \pi^2(x)\{\sigma^2(\beta_0^\top x)\}^2 p^2(\beta_0^\top x)dx,$$
$$\Sigma^V = 2\int \mathcal{K}^2(u)du \cdot \int \frac{\{\sigma^2(\beta_0^\top x)\}^2 p^2(\beta_0^\top x)}{\pi^2(x)}dx,$$

with $\sigma^2(\beta_0^\top x) = E(e^2|\beta_0^\top X = \beta_0^\top x)$. The function $p(\cdot)$ in Σ^V denotes the probability density function of $\beta_0^\top X$. In general, Σ^T and Σ^V are unknown, two consistent estimates can be defined as

$$\hat\Sigma^T = \frac{2}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}^n \frac{1}{h^{\hat q}}\delta_i\delta_j\mathcal{K}^2\Big(\frac{\hat B(\hat q)^\top(x_i - x_j)}{h}\Big)\hat e_i^2\hat e_j^2,$$
$$\hat\Sigma^V = \frac{2}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}^n \frac{1}{h^{\hat q}}\frac{\delta_i\delta_j}{\hat\pi^2(x_i)\hat\pi^2(x_j)}\mathcal{K}^2\Big(\frac{\hat B(\hat q)^\top(x_i - x_j)}{h}\Big)\hat e_i^2\hat e_j^2.$$

The following theorem states the asymptotic properties of the test statistics T_n and V_n under the null hypothesis H_0.

Theorem 1. *Suppose that conditions $(C1)-(C7)$ in the Appendix hold. Under the null hypothesis H_0, we have*

$$nh^{1/2}T_n \Rightarrow N(0,\Sigma^T) \quad \textit{and} \quad nh^{1/2}V_n \Rightarrow N(0,\Sigma^V),$$

where Σ^T and Σ^V can be respectively estimated by $\hat\Sigma^T$ and $\hat\Sigma^V$ defined above.

Further, define the scale-invariant standardized version of statistics T_n and V_n as

$$S_n = \sqrt{\frac{n-1}{n}} \frac{nh^{1/2}T_n}{\sqrt{\hat{\Sigma}^T}} \quad \text{and} \quad U_n = \sqrt{\frac{n-1}{n}} \frac{nh^{1/2}V_n}{\sqrt{\hat{\Sigma}^V}}, \tag{3.1}$$

respectively.

Corollary 1. *Under the null hypothesis H_0 and conditions (C1)-(C7) in the Appendix, we have*

$$S_n \Rightarrow N(0,1) \quad \textit{and} \quad U_n \Rightarrow N(0,1).$$

Theorem 1 and Corollary 1 depict the asymptotic normality of the proposed test statistics under H_0. From Theorem 1, it is clear that our tests own the much faster converging rates, $O(n^{-1}h^{-1/2})$ in the sense that T_n and V_n multiplying $O(n^1h^{1/2})$ have weak limits whereas the classical tests $T_n^\star$ and $V_n^\star$ in (2.9) need to multiply $O(n^1h^{p/2})$ to get weak limits. Corollary 1 can be applied to determine the critical value of our tests. Note that the tests are one-sided due to its nature at the population level, the null hypothesis H_0 can be rejected when $S_n > Z_{1-\alpha}$(or $U_n > Z_{1-\alpha}$), where $Z_{1-\alpha}$ is the $1-\alpha$ quantile of the standard normal distribution with $P(Z \geq Z_{1-\alpha}) = \alpha$.

3.2. Power study

We now investigate the power behaviors of our tests under the local and global alternatives. The following sequence of alternatives H_{1n} is considered:

$$H_{1n} : Y = g(\beta_0^\top X, \theta_0) + C_n m(B^\top X) + \varepsilon \tag{3.2}$$

where $E(\varepsilon|X) = 0$, $E[m^2(B^\top X)] < \infty$ and $\{C_n\}$ is a sequence of constants. When C_n is a fixed constant, it is the global alternative (1.2); When C_n goes to zero, it becomes a sequence of local alternatives. Denote $\tilde{\alpha} = (\tilde{\beta}, \tilde{\theta})^\top$ as the minimizer of the following

$$\tilde{\alpha} = \arg \min_{(\beta,\theta)^\top} E\{g(\beta^\top X, \theta) - m(X)\}^2,$$

where $m(X) = E(Y|X)$. For the least squares estimate $\hat{\alpha}$, we always have $\hat{\alpha} - \tilde{\alpha} = O_p(1/\sqrt{n})$.

We first present the asymptotic property of $\hat{q}$ determined by (2.10) under the above local alternatives.

Lemma 1. *Assume Conditions (C1)-(C7) in the Appendix hold. Under the local alternatives (3.2) with $C_n = n^{-1/2}h^{-1/4}$, we have $\hat{q} = 1$ as $n \to \infty$ with a probability going to one.*

Denote $\alpha = (\beta, \theta)^\top$. In order to state the following theorem, we first give some notations:

$$\dot{g}(X, \beta_0, \theta_0) = \partial g(\alpha, x_j)/\partial \alpha^\top|_{\alpha=\alpha_0}, \quad \Sigma_1 = E\{\dot{g}(X, \beta_0, \theta_0)\dot{g}(X, \beta_0, \theta_0)^\top\},$$

$$\mu^T = E\Big[\pi^2(X)f(\beta_0^\top X)\Big(m(B^\top X) - \dot{g}(X,\beta_0,\theta_0)^\top \Sigma_1^{-1}E\{\dot{g}(X,\beta_0,\theta_0)m(B^\top X)\}\Big)^2\Big],$$

$$\mu^V = E\Big[f(\beta_0^\top X)\Big(m(B^\top X) - \dot{g}(X,\beta_0,\theta_0)^\top \Sigma_1^{-1}E\{\dot{g}(X,\beta_0,\theta_0)m(B^\top X)\}\Big)^2\Big],$$

where $f(\cdot)$ denotes the probability density function of $\beta_0^\top X$.

Theorem 2. *Given Conditions (C1)-(C7) in the Appendix, we have the following conclusions.*

(i) Under the global alternative (1.2) with C_n being fixed, S_n and V_n go to infinity at the rate of $nh^{1/2}$:

$$S_n/(nh^{1/2}) \Rightarrow l_1 \quad \textit{and} \quad U_n/(nh^{1/2}) \Rightarrow l_2,$$

where l_1 and l_2 are both positive constants.

(ii) Under the local alternatives (3.2) with $C_n = n^{-1/2}h^{-1/4}$, we have

$$nh^{1/2}T_n \Rightarrow N(\mu^T, \Sigma^T) \quad \textit{and} \quad nh^{1/2}V_n \Rightarrow N(\mu^V, \Sigma^V),$$
$$S_n \Rightarrow N(\mu^T/\sqrt{\Sigma^T}, 1) \quad \textit{and} \quad U_n \Rightarrow N(\mu^V/\sqrt{\Sigma^V}, 1).$$

Theorem 2 suggests that under the global alternative hypothesis, the proposed tests are consistent with the asymptotic power 1. Under the local alternatives, the tests own better power performance than those proposed by Guo et al. [7] when the dimension p of covariates is large, even moderate. That can be ascribed to the convergence rates of the test statistics T_n and V_n to infinity $O(nh^{1/2})$ rather than $O(nh^{p/2})$ in their tests.

It is worth noticing that when an alternative is nonparametric, there are several ways to characterize the power of tests. In this paper, we focus on investigating the Pitman-like local alternatives. Theorem 2 indicates that the proposed tests have asymptotic power 1 for the local alternatives that are distinct from the null hypothesis at the rate slower than $n^{-1/2}h^{-1/4}$. Also the tests can still detect the alternatives converging to the null hypothesis at the rate $n^{-1/2}h^{-1/4}$. Further, as the local smoothing tests, the model adaptation properties allow us to use the normalizing constant that has been used for existing local smoothing tests with one-dimensional covariate. As the result, the power comes from both the non-centrality of the residuals and the convergence rate of the test statistics under the alternative. This enhances the power performance. In this sense, the rate of convergence partly summarizes the power properties of the tests.

Another issue is about bandwidth selection. It seems natural to us that different bandwidths are used when the underlying working dimensions are different. In fact, we use $h = O(n^{-1/(4+\hat{q})})$ in our numerical studies, where $\hat{q}$ is equal to the true dimension under either the null or alternative in a probability sense. This is different, but adaptive to the models under the null and alternative respectively. Although, in practice, we do not know whether the null is true or not, due to this model adaptation, its automatic selection adapts to the unknown value of q under the null and alternative. Therefore, for practical use, bandwidth selection does not have problem. For studying the theoretical properties, it is also an issue to discuss which bandwidth is required under the null and alternative as the convergence rate would be different under these two scenarios. Note that again due to the model adaptation, the bandwidth is also adaptive to the underlying model, we then need not to use a unique bandwidth, while need to use different ones to derive the theoretical results.

Remark 2. *We consider a theoretical comparison between S_n and U_n. This comparison is to examine the sensitivity of tests to local alternative hypothesis. To this end, we compare T_n with V_n that requires an estimate of the probability π. Theorem 1 indicates that the limiting null variance Σ^T of T_n is smaller than that $< \Sigma^V$ of V_n since $0 < \pi(x) < 1$ and thus suggests that T_n is more sensitive to alternatives. On the other hand, Theorem 2 shows that under the alternatives, the test statistic T_n owns a smaller drift term μ^T than μ^V that V_n has. Let $D^T = \mu^T/\sqrt{\Sigma^T}$ and $D^V = \mu^V/\sqrt{\Sigma^V}$. Together with the above two results we can see that the asymptotic powers of T_n and V_n are $1-\Phi(z_{1-\alpha/2}-D^T)$ and $1-\Phi(z_{1-\alpha/2}-D^V)$, respectively when we consider the local alternatives with $C_n = n^{-1/2}h^{-1/4}$. Here, $\Phi(\cdot)$ is the cumulative probability function of the standard normal distribution and $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. When the response is missing at random, since $\mu^T < \mu^V$ and $\Sigma^T < \Sigma^V$, we cannot obtain a one-side inequality between D^T and D^T for all local alternative models. Thus, theoretically, no one can be superior to the other in general, while different alternatives would be in favor of different test. From the power study in the simulations, S_n seems to work better.*

Remark 3. *The preliminarily unreported simulation results based on S_n and U_n in (3.1) show that the empirical sizes tend to be slightly larger than the significance level. Empirically, we implement the size-adjusted versions as follows:*

$$\tilde{S}_n = \frac{S_n}{1+4n^{-4/5}} \quad \text{and} \quad \tilde{U}_n = \frac{U_n}{1+3n^{-4/5}}. \tag{3.3}$$

The size-adjustment constants are carefully selected through intensive Monte Carlo simulations and the above values are recommended. It is worth noticing that the size-adjustment values are both asymptotically negligible when $n \to \infty$ since $\tilde{S}_n - S_n$ and $\tilde{U}_n - U_n$ tend to zero at the rate of order $n^{-4/5}$ as $n \to \infty$. After such an adjustment, the tests can much better control type I errors and enhance powers than those without size-adjustment.

4. Simulation studies

In this section, two simulation studies are carried out to check the theory and investigate the finite-sample performance of the proposed test statistics. Throughout this section, the adjusted test statistics $\tilde{S}_n$ and $\tilde{U}_n$ in (3.3) are applied to conduct the simulations. The objective of the first study is to examine and compare the performance of the two proposed tests. The effects of different distributions of the error, the correlation between the components of X and nonlinearity under the null hypothesis on the performance of our new tests are considered in this study. The power comparison is presented when the test denoted as Q_n^{CC} proposed by Guo et al (2016) [6] is applied by using the complete case data in the datasets. The second study is used to check the usefulness of the dimension reduction technique to overcome the curse of dimensionality. We intend to see how much improvement the proposed tests can achieve compared with the tests $T_n^\star, V_n^\star$ in (2.9) proposed by Guo et al [7]. The $q = p$ alternative models without dimension reduction structures are also discussed in this study to examine the effectiveness of model-adaptation, although an extra estimation of $q = p$ is needed.

Study 1: The data are generated from the following models:

$$\begin{aligned}
&H_{11} : Y = \beta^\top X + a(\beta^\top X)^2 + \varepsilon,\\
&H_{12} : Y = \beta^\top X + a\exp(\beta^\top X - 0.8) + \varepsilon,\\
&H_{13} : Y = 0.25(\beta^\top X)^2 - 3a\cos(0.3\pi\beta^\top X + 1) + \varepsilon,
\end{aligned}$$

where $p = 8, q = 1, \beta = (1, \dots, 1)^\top/\sqrt{p}$. The covariate vector $X = (X_1, \dots, X_p)^\top$ and the error ε are independent. Two types of the covariates and two kinds of errors are considered. The covariate vectors $X_i = (X_{1i}, \dots, X_{pi})^\top$ are generated from a multivariate normal distribution $N(0, \Sigma_1^X)$ or $N(0, \Sigma_2^X)$ with $\Sigma_1^X = I_{p\times p}$ and $\Sigma_2^X = (0.2^{|j-l|})_{p\times p}$, where I_p is a $p \times p$ identity matrix. The ε_i's come from standard normal distribution $N(0, 1)$ and double exponential distribution $DE(0, \sqrt{3}/2)$ with probability density function $f(x) = \sqrt{3}/3\exp(-2\sqrt{3}|x|/3)$. Assume that the covariate vectors X_i are completely observed and some of the responses Y_i are missing at random. The response indicator variable δ_i follow the Bernoulli distribution with the probability $\pi(x)$. The following two missing mechanisms $\pi_k(x),\ k = 1, 2$ are adopted:

- Case 1: $\pi_1(x) = P(\delta = 1|X = x) = 1/\Big\{1 + \exp\big[-(0.2\gamma_1^\top x + 1.2)\big]\Big\}$;
- Case 2: $\pi_2(x) = P(\delta = 1|X = x) = 0.7 + 0.25(|\gamma_1^\top x - 1| + |\gamma_1^\top x - 2|)$ if $|\gamma_1^\top x - 1| + |\gamma_1^\top x - 2| \le 1.5$ and 0.7 otherwise,

where $\gamma_1 = (\underbrace{1, \dots, 1}_{p/2}, 0, \dots, 0)^\top/\sqrt{p/2}$. For these two probability selection functions, the non-missing proportions are roughly 78% and 76%, respectively. Here, we set $a = 0, 0.2, \dots, 1.0$ where $a = 0$ corresponds to the null hypotheses and $a \neq 0$ to the alternative hypotheses. To be specific, for H_{11} and H_{12}, the corresponding null hypotheses are both $H_{10} : Y = \beta^\top X + \varepsilon$; For H_{13}, the corresponding null hypothesis is $\tilde{H}_{10} : Y = 0.25(\beta^\top X)^2 + \varepsilon$. For H_{11} and H_{12}, the models under the null hypothesis are linear and the alternative models are both single-index. The hypothetical model against H_{13} is nonlinear and under H_{13}, the alternative model is a high frequency alternative. We will check whether the newly proposed tests are effective for all kinds of models or not.

To compute the proposed tests, we need to choose the kernel function $\mathcal{K}(\cdot)$ in (2.6) and (2.7) and the kernel function $\tilde{K}(\cdot)$ in (2.8) for estimating $\pi(x)$. Also we need to select the corresponding bandwidth h and $\tilde{h}$. Throughout the simulations, unless otherwise specified, the $\hat{q}$-dimensional kernel function $K(\cdot)$ is taken to be $\mathcal{K}(u) = \mathcal{K}(u_1, \dots, u_{\hat{q}}) = \mathcal{K}_1(u_1) \cdot \dots \cdot \mathcal{K}_{\hat{q}}(u_{\hat{q}})$ with $\mathcal{K}_l(u_l) = 15/16(1-u_l^2)^2,\ l = 1, \dots, \hat{q}$ if $|u_l| \leq 1$ and 0, otherwise and the bandwidth $h = 1.2n^{-1/(\hat{q}+4)}$ is recommended through intensive computation. The one-dimensional kernel function $\tilde{K}(\cdot)$ is selected as $\tilde{K}(\nu) = 15/16(1-\nu^2)^2$ if $|\nu| \leq 1$ and 0, otherwise and $\tilde{h} = 1.2n^{-1/5}$. Our limited empirical experience shows that the proposed tests are not very sensitive to the choices of kernel function and bandwidth. In the CCAR method to estimate the matrix B, the number of slices S is set to be 10. The significance level is set to be $\alpha = 0.05$ and two sample sizes $n = 100$ and 200 are conducted. the replication time of experiment is 5000.

Table 1 tabulates the empirical sizes and powers of the proposed tests $\tilde{S}_n$ and $\tilde{U}_n$ against the alternatives H_{11}, H_{12} with different values of a. Here, $\varepsilon \sim N(0,1)$ and two kinds of covariate vectors X are considered. Based on this table, we can have the following observations. First, for every combination of the covariate X and the sample size we conduct, the tests $\tilde{S}_n$ and $\tilde{U}_n$ have comparable empirical sizes, which are all very close to the pre-specified significance level $\alpha = 0.05$. Second, the simulated powers of the two tests become higher with increasing of the parameter a and the tests are more powerful against the alternatives with larger sample size. Compared with the two tests $\tilde{S}_n$ and $\tilde{U}_n$, we can see that in most cases, $\tilde{S}_n$ is superior to $\tilde{U}_n$: $\tilde{S}_n$ has higher simulated powers than $\tilde{U}_n$. Third, under the same alternative hypothesis, the effect of the correlation between the components of X on the empirical sizes and simulated powers is not very significant and the simulation results are all acceptable.

Table 2 displays the simulation results under models H_{11} and H_{12} for $a = 0, \dots, 1$ at the significance level $\alpha = 0.05$. Unlike the above table, here $\varepsilon \sim DE(0, \sqrt{3}/2)$ is set. Combining Table 1 and Table 2, we can see that the distribution of random error makes no significant influence, which indicates the robustness of the proposed tests against error distributions. The other similar conclusions to those in Table 1 can be made based on Table 2 and thus we omit them.

To see the gain of efficiency the proposed tests have compared with the test that only uses complete case data in the datasets, we report the results when the test Q_n^{CC} proposed by Guo et al [6] is applied. The powers comparison of the three tests $\tilde{S}_n$, $\tilde{U}_n$, Q_n^{CC} for the null hypothesis $H_{10} : Y = \beta^\top X + \varepsilon$ against the alternative H_{12} with $n = 100, \varepsilon \sim N(0,1)$ are presented in Figure 1. From this figure, we can see that the powers of these three tests are all reasonable and acceptable. To be specific, $\tilde{S}_n$ shows the the best performance among the three tests and has uniformly higher powers than Q_n^{CC}. The powers of $\tilde{U}_n$ and Q_n^{CC} are comparable: when a is small, $\tilde{U}_n$ performs better; whereas a is large, Q_n^{CC} is the winner.

Figure 2 reports the empirical sizes and powers for the nonlinear null model under $\tilde{H}_{10} : Y = 0.25(\beta^\top X)^2 + \varepsilon$ against the high frequency alternative H_{13} with $n = 100$ and $\varepsilon \sim N(0,1)$. Under this nonlinear null hypothesis, the tests

Table 1
Empirical sizes and powers of $\tilde{S}_n$ and $\tilde{U}_n$ for H_{10} v.s. H_{11} and H_{12} at the significance level $\alpha = 0.05$ with $p = 8, \varepsilon \sim N(0, 1)$.

	X	a	Case 1				Case 2			
			$n = 100$		$n = 200$		$n = 100$		$n = 200$	
			$\tilde{S}_n$	$\tilde{U}_n$	$\tilde{S}_n$	$\tilde{U}_n$	$\tilde{S}_n$	$\tilde{U}_n$	$\tilde{S}_n$	$\tilde{U}_n$
H_{11}	$X \sim N(0, \Sigma_1^X)$	0	0.052	0.050	0.047	0.045	0.055	0.051	0.050	0.048
		0.2	0.142	0.125	0.296	0.283	0.126	0.113	0.292	0.286
		0.4	0.434	0.353	0.884	0.789	0.407	0.350	0.885	0.802
		0.6	0.678	0.558	0.983	0.888	0.664	0.552	0.981	0.903
		0.8	0.826	0.664	0.994	0.911	0.797	0.670	0.994	0.927
		1.0	0.886	0.731	0.997	0.917	0.866	0.717	0.998	0.933
	$X \sim N(0, \Sigma_2^X)$	0	0.050	0.047	0.054	0.050	0.050	0.048	0.054	0.053
		0.2	0.228	0.195	0.556	0.499	0.223	0.196	0.532	0.502
		0.4	0.660	0.527	0.984	0.866	0.637	0.532	0.981	0.909
		0.6	0.834	0.664	0.998	0.904	0.804	0.677	0.997	0.928
		0.8	0.902	0.728	0.999	0.917	0.879	0.742	0.998	0.939
		1.0	0.932	0.755	1.000	0.928	0.916	0.766	0.999	0.947
H_{12}	$X \sim N(0, \Sigma_1^X)$	0	0.047	0.045	0.054	0.053	0.048	0.046	0.055	0.054
		0.2	0.102	0.093	0.173	0.172	0.100	0.093	0.170	0.165
		0.4	0.290	0.252	0.624	0.563	0.284	0.249	0.618	0.578
		0.6	0.620	0.520	0.959	0.865	0.599	0.508	0.954	0.875
		0.8	0.857	0.707	0.999	0.923	0.853	0.733	0.997	0.942
		1.0	0.963	0.811	1.000	0.943	0.953	0.827	1.000	0.953
	$X \sim N(0, \Sigma_2^X)$	0	0.053	0.047	0.051	0.047	0.051	0.049	0.050	0.053
		0.2	0.120	0.110	0.226	0.221	0.120	0.110	0.233	0.223
		0.4	0.402	0.330	0.791	0.704	0.390	0.334	0.779	0.718
		0.6	0.735	0.594	0.991	0.885	0.726	0.621	0.987	0.917
		0.8	0.922	0.761	0.998	0.922	0.908	0.775	0.999	0.949
		1.0	0.980	0.815	1.000	0.940	0.974	0.845	1.000	0.955

$\tilde{S}_n$ and $\tilde{U}_n$ can maintain the empirical sizes reasonably well. From this figure, we can see that for the cosine alternative H_{13}, the proposed tests show good power performance and the simulated power curves for them have both popular sigmoidal shapes. The test $\tilde{S}_n$ outperforms $\tilde{U}_n$. It seems that having an estimate of π in the test $\tilde{U}_n$ slightly deteriorates its power performance although in theory, it is not affected.

Study 2: Consider the following models:

$$
\begin{aligned}
H_{21} &: Y = \beta_1^\top X + a\cos(0.3\pi\beta_2^\top X + 0.5) + \varepsilon, \\
H_{22} &: Y = \beta_1^\top X + a\Big\{0.2\log(\beta_2^\top X) + 1\Big\} + \varepsilon, \\
H_{23} &: Y = X_1 + a\Big\{(0.1X_2)^2 - \cos(0.25\pi X_3) + |0.1X_4|\Big\} + \varepsilon, \\
H_{24} &: Y = X_1 + a\Big\{0.8\exp(0.2X_2) + X_3X_4\Big\} + \varepsilon, \\
H_{25} &: Y = \beta_1^\top X + 2.5a(0.3\beta_2^\top X + 0.5)^3 + \varepsilon,
\end{aligned}
$$

where $p = 4$, $\beta_1 = (\underbrace{1, \ldots, 1}_{p/2}, 0, \ldots, 0)^\top/\sqrt{p/2}$ and $\beta_2 = (\underbrace{0, \ldots, 0}_{p/2}, 1, \ldots, 1)^\top/\sqrt{p/2}$. Here, we set $a = 0, 0.2, \ldots, 1.0$ where $a = 0$ corresponds to the null

TABLE 2

Empirical sizes and powers of $\tilde{S}_n$ and $\tilde{U}_n$ for H_{10} v.s. H_{11} and H_{12} at the significance level $\alpha = 0.05$ with $p = 8, \varepsilon \sim DE(0, \frac{\sqrt{3}}{2})$.

	X	a	Case 1				Case 2			
			$n=100$		$n=200$		$n=100$		$n=200$	
			$\tilde{S}_n$	$\tilde{U}_n$	$\tilde{S}_n$	$\tilde{U}_n$	$\tilde{S}_n$	$\tilde{U}_n$	$\tilde{S}_n$	$\tilde{U}_n$
H_{11}	$X \sim N(0, \Sigma_1^X)$	0	0.050	0.047	0.054	0.053	0.045	0.045	0.050	0.053
		0.2	0.097	0.089	0.195	0.200	0.097	0.096	0.183	0.185
		0.4	0.303	0.273	0.708	0.645	0.292	0.257	0.686	0.633
		0.6	0.541	0.440	0.927	0.842	0.517	0.433	0.916	0.843
		0.8	0.689	0.569	0.976	0.891	0.671	0.558	0.972	0.901
		1.0	0.795	0.652	0.990	0.892	0.777	0.649	0.986	0.921
	$X \sim N(0, \Sigma_2^X)$	0	0.049	0.046	0.054	0.050	0.053	0.054	0.052	0.053
		0.2	0.164	0.144	0.377	0.357	0.148	0.137	0.363	0.353
		0.4	0.506	0.413	0.927	0.817	0.478	0.418	0.922	0.840
		0.6	0.722	0.578	0.982	0.876	0.700	0.588	0.981	0.916
		0.8	0.824	0.643	0.996	0.901	0.805	0.686	0.993	0.924
		1.0	0.879	0.712	1.000	0.905	0.863	0.725	0.996	0.934
H_{12}	$X \sim N(0, \Sigma_1^X)$	0	0.049	0.047	0.051	0.046	0.047	0.047	0.054	0.054
		0.2	0.080	0.073	0.130	0.126	0.081	0.076	0.129	0.128
		0.4	0.199	0.185	0.446	0.421	0.199	0.189	0.438	0.411
		0.6	0.451	0.387	0.832	0.761	0.420	0.368	0.815	0.748
		0.8	0.678	0.582	0.976	0.894	0.657	0.566	0.973	0.905
		1.0	0.849	0.719	0.997	0.937	0.850	0.731	0.998	0.945
	$X \sim N(0, \Sigma_2^X)$	0	0.051	0.046	0.053	0.053	0.047	0.046	0.052	0.054
		0.2	0.103	0.093	0.167	0.171	0.093	0.090	0.160	0.160
		0.4	0.287	0.256	0.603	0.554	0.278	0.250	0.578	0.547
		0.6	0.563	0.470	0.926	0.832	0.551	0.484	0.923	0.851
		0.8	0.795	0.658	0.997	0.904	0.773	0.669	0.995	0.929
		1.0	0.917	0.759	1.000	0.926	0.908	0.776	1.000	0.946

hypothesis and $a \neq 0$ to the alternative hypotheses. That is, for H_{21}, H_{22} and H_{25}, the corresponding null hypotheses are all $H_{20} : Y = \beta_1^\top X + \varepsilon$; For H_{23} and H_{24}, the corresponding null hypotheses are both $\tilde{H}_{20} : Y = X_1 + \varepsilon$. Thus, for the alternative models H_{21}, H_{22} and H_{25} with $a \neq 0$, $q = 2$ and $B = (\beta_1, \beta_2)$. Denote $\beta_i^\star,\ i = 1, \ldots, p$ to be the unit vector, in which the i-th element is 1. As for the alternative H_{23} and H_{24} with $a \neq 0$, $B = (\beta_1^\star, \beta_2^\star, \beta_3^\star, \beta_4^\star)$ and $q = p = 4$. In other words, H_{23} and H_{24} are alternative models without dimension reduction structure and are used to check the performance of the model-adaptive tests even when an extra estimate $\hat{q}$ for the dimension p is involved. The model H_{21} is high frequent whereas the others are not. In this simulation study, the covariate vector X and the random error ε are independent. X comes from multivariate normal distribution $N(0, I_p)$ where I_p is the p-dimensional identity matrix. Two types of the random errors are conducted: one is $\varepsilon \sim N(0, 1)$ and the other is $\varepsilon \sim \exp(1) - 1$, where $\exp(1)$ denotes the exponential distribution with the parameter $\lambda = 1$. The second simulation study is used to examine the impact from dimensionality on both the proposed tests $\tilde{S}_n, \tilde{V}_n$ and the tests $T_n^\star, V_n^\star$ in (2.9) proposed by Guo et al [7]. The p-dimensional kernel function $K^\star(\cdot)$ in $T_n^\star, V_n^\star$ is chosen as $\mathcal{K}^\star(u) = \mathcal{K}^\star(u_1, \ldots, u_p) = \mathcal{K}_1^\star(u_1) \cdot \ldots \cdot \mathcal{K}_p^\star(u_p)$ with $\mathcal{K}_l^\star(u_l) = 15/16(1 - u_l^2)^2,\ l = 1, \ldots, p$ if $|u_l| \leq 1$ and 0, otherwise. The

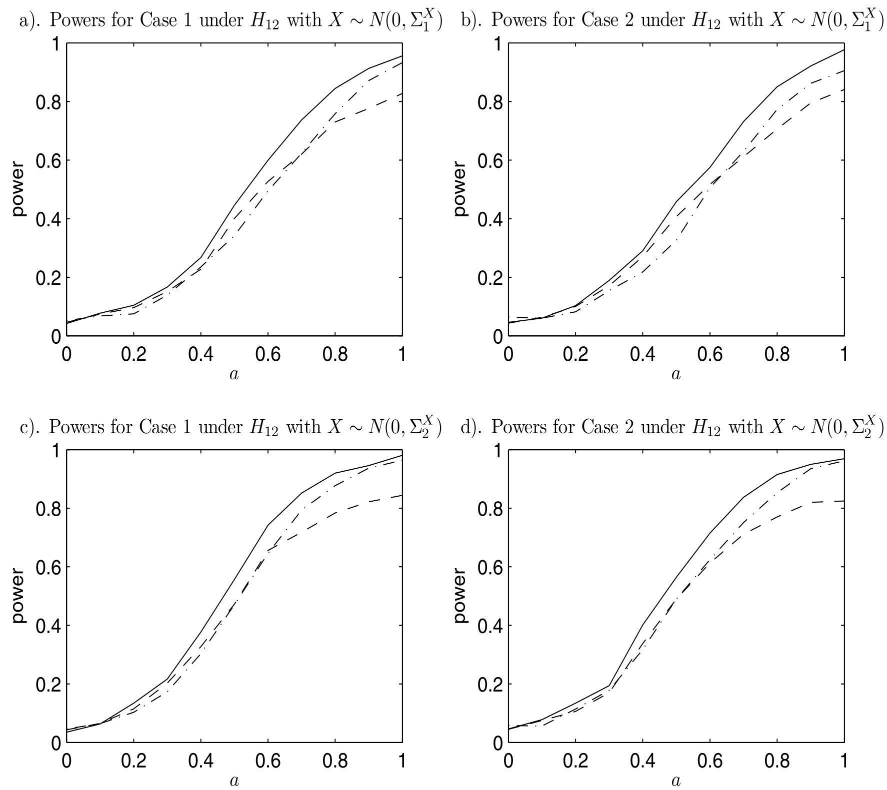

FIG 1. *Empirical sizes and powers of $\tilde{S}_n$, $\tilde{U}_n$ and Q_n^{CC} for H_{10} v.s. H_{13} at the significance level $\alpha = 0.05$ with $n = 100$, $\varepsilon \sim N(0,1)$. In four plots, the solid line, the dash line and the dotted-dash line are for $\tilde{S}_n$, $\tilde{U}_n$ and Q_n^{CC} respectively.*

corresponding bandwidth $h^\star = 1.25n^{-1/6}$ is applied. Here, the covariate vector X is completely observed and the response Y is missing at random. We consider the following two missing mechanisms similarly as before:

- Case 3: $\pi_3(x) = P(\delta = 1|X = x) = 1.5 - 0.12(|\gamma_2^\top x - 1| + |\gamma_2^\top x - 0.5|)$ if $|\gamma_2^\top x - 1| + |\gamma_2^\top x - 0.5| \leq 1.5$ and 0.75 otherwise;
- Case 4: $\pi_4(x) = P(\delta = 1|X = x) = 0.78$ for all of $X = x$,

where $\gamma_2 = (1, \ldots, 1)^\top/\sqrt{p}$. The average non-missing rates corresponding to the above two cases are approximately 84% and 78%. The significance level $\alpha = 0.05$ and the sample size $n = 100$ are adopted. Every simulation result is the average of 5000 replications.

The effects of missing mechanisms and the distributions of random errors on simulation results are considered. The corresponding empirical sizes and simulated powers are presented in Table 3 for $n = 100, p = 4, q = 2$ at the significance level $\alpha = 0.05$.

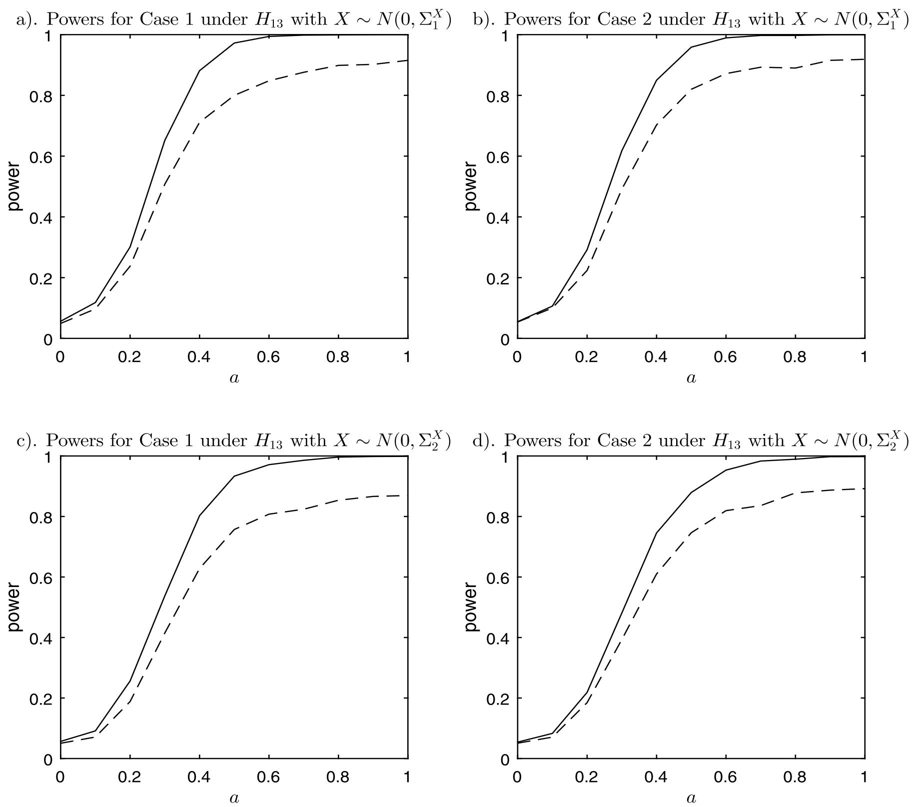

FIG 2. *Empirical sizes and powers of $\tilde{S}_n$ and $\tilde{U}_n$ for $\tilde{H}_{10}$ v.s. H_{13} at the significance level $\alpha = 0.05$ with $n = 100$, $\varepsilon \sim N(0, 1)$. In four plots, the solid line and the dash line are for $\tilde{S}_n$ and $\tilde{U}_n$, respectively.*

Based on Table 3, we can observe that for all the cases we conduct, the empirical type I errors of the proposed tests $\tilde{S}_n, \tilde{U}_n$ can be under control which are close to the significance level. However, the tests $T_n^\star$ and $V_n^\star$ tend to be conservative although the sizes are still acceptable. All of the tests are generally sensitive to the alternatives in the sense that, as a increases, the simulated powers would also increase. Also, they are robust to various distributions of random errors. For Case 3 and Case 4, when there is more than one direction under the alternative hypothesis, the test $\tilde{S}_n$ is still more powerful than $\tilde{U}_n$, which conforms to the results in Study 1. Compared the proposed tests with $T_n^\star$ and $V_n^\star$, it is obvious that the proposed tests are more powerful to detect the alternative models even in the cases with $p = 4$ although the proposed tests have a seemingly unnecessary estimate of $q = p$. These results indicate that the normalizing constant in the proposed tests $\tilde{S}_n, \tilde{U}_n$ play an useful role to enhance the power performance as we claimed in Section 1. Also the tests are not significantly affected by the dimension of the covariate vector X and the tests $T_n^\star, V_n^\star$ have relatively poor powers caused by the dimensionality problem.

Table 3

Empirical sizes and powers of $\tilde{S}_n$, $\tilde{U}_n$, $T_n^\star$ and $V_n^\star$ for H_{20} v.s. H_{21} and H_{22} at the significance level $\alpha = 0.05$ with $X \sim N(0, I_4), p = 4, q = 2, n = 100$.

	ε	a	Case 3				Case 4			
			$\tilde{S}_n$	$\tilde{U}_n$	$T_n^\star$	$V_n^\star$	$\tilde{S}_n$	$\tilde{U}_n$	$T_n^\star$	$V_n^\star$
H_{21}	$\varepsilon \sim N(0,1)$	0	0.046	0.049	0.042	0.041	0.043	0.046	0.042	0.042
		0.2	0.097	0.093	0.055	0.054	0.095	0.091	0.053	0.053
		0.4	0.269	0.239	0.065	0.064	0.266	0.252	0.058	0.057
		0.6	0.555	0.496	0.116	0.114	0.537	0.500	0.101	0.099
		0.8	0.821	0.725	0.191	0.190	0.810	0.745	0.156	0.155
		1.0	0.949	0.848	0.296	0.296	0.935	0.860	0.257	0.256
	$\varepsilon \sim \exp(1) - 1$	0	0.052	0.053	0.040	0.042	0.053	0.054	0.042	0.043
		0.2	0.076	0.069	0.054	0.053	0.066	0.065	0.052	0.053
		0.4	0.236	0.213	0.055	0.055	0.229	0.212	0.054	0.053
		0.6	0.610	0.537	0.072	0.071	0.571	0.531	0.063	0.066
		0.8	0.897	0.795	0.185	0.183	0.877	0.802	0.142	0.144
		1.0	0.988	0.885	0.401	0.402	0.980	0.909	0.336	0.335
H_{22}	$\varepsilon \sim N(0,1)$	0	0.049	0.051	0.042	0.041	0.049	0.048	0.044	0.043
		0.2	0.113	0.112	0.058	0.057	0.125	0.122	0.055	0.054
		0.4	0.417	0.393	0.071	0.072	0.415	0.387	0.073	0.074
		0.6	0.794	0.733	0.150	0.150	0.784	0.740	0.135	0.135
		0.8	0.960	0.885	0.275	0.274	0.956	0.905	0.248	0.247
		1.0	0.997	0.934	0.412	0.415	0.995	0.949	0.377	0.383
	$\varepsilon \sim \exp(1) - 1$	0	0.053	0.054	0.047	0.049	0.055	0.054	0.044	0.045
		0.2	0.086	0.085	0.053	0.052	0.087	0.088	0.051	0.052
		0.4	0.430	0.403	0.054	0.054	0.427	0.413	0.054	0.055
		0.6	0.875	0.808	0.117	0.115	0.876	0.828	0.107	0.108
		0.8	0.994	0.928	0.323	0.325	0.993	0.944	0.296	0.297
		1.0	1.000	0.948	0.565	0.561	1.000	0.960	0.524	0.523

Table 4 shows the empirical sizes and powers under $\tilde{H}_{20} : Y = X_1 + \varepsilon$ against the alternative models H_{23} and H_{24} without dimension reduction structures. In this case, $p = q = 4$. We consider one case with $n = 100$. From this table, it can be seen that the proposed tests can maintain the significance level and the empirical sizes of $T_n^\star$ and $V_n^\star$ are also reasonable. As to the power performance, it is clear that the proposed tests $\tilde{S}_n, \tilde{U}_n$ are the winners compared with $T_n^\star, V_n^\star$ and even for the purely nonparametric regression model with $p = q$, the power is still higher because of using the standardizing constant $nh^{1/2}$ rather than the $nh^{p/2}$ that diverges to infinity slower.

In order to see the impact from dimensionality increasing in the covariate vector X, a more in-depth analysis with the alternative H_{25} is considered. Figure 3 presents the simulated power curves with p increasing from 2 to 4. Here, $n = 100, q = 2, \varepsilon \sim N(0,1)$ and the significance level is $\alpha = 0.05$. From this figure, we can see that the proposed tests $\tilde{S}_n, \tilde{U}_n$ are almost not affected by the increase of dimension p whereas the powers of $T_n^\star, V_n^\star$ drop rapidly when the dimension p increases from 2 to 4, even this increase is not very significant. Whatever $p = 4$ or $p = 2$, the proposed tests have advantage.

In summary, the above simulation studies indicate that the adjusted tests $\tilde{S}_n$ and $\tilde{U}_n$ can control the type I error very well and is more powerful than the tests $T_n^\star$ and $V_n^\star$. Meanwhile, the proposed tests can avoid the curse of dimensionality to a great extent. The test $\tilde{S}_n$ is more recommendable.

TABLE 4
Empirical sizes and powers of $\tilde{S}_n$, $\tilde{U}_n$ $T_n^\star$ and $V_n^\star$ for $\tilde{H}_{20}$ v.s. H_{23} and H_{24} at the significance level $\alpha = 0.05$ with $X \sim N(0, I_4), p = q = 4, n = 100$.

	ε	a	Case 3				Case 4			
			$\tilde{S}_n$	$\tilde{U}_n$	$T_n^\star$	$V_n^\star$	$\tilde{S}_n$	$\tilde{U}_n$	$T_n^\star$	$V_n^\star$
H_{23}	$\varepsilon \sim N(0,1)$	0	0.048	0.049	0.044	0.043	0.044	0.049	0.045	0.044
		0.2	0.128	0.118	0.053	0.052	0.113	0.109	0.051	0.051
		0.4	0.363	0.329	0.073	0.072	0.360	0.336	0.064	0.063
		0.6	0.718	0.632	0.141	0.139	0.684	0.621	0.119	0.118
		0.8	0.929	0.814	0.235	0.227	0.913	0.833	0.201	0.202
		1.0	0.989	0.893	0.354	0.353	0.985	0.911	0.317	0.318
	$\varepsilon \sim \exp(1) - 1$	0	0.052	0.053	0.045	0.043	0.054	0.055	0.043	0.043
		0.2	0.159	0.149	0.080	0.080	0.152	0.150	0.081	0.081
		0.4	0.399	0.371	0.138	0.138	0.382	0.364	0.128	0.129
		0.6	0.674	0.610	0.227	0.226	0.665	0.614	0.210	0.208
		0.8	0.869	0.790	0.322	0.321	0.849	0.794	0.309	0.307
		1.0	0.960	0.879	0.427	0.426	0.948	0.884	0.403	0.403
H_{24}	$\varepsilon \sim N(0,1)$	0	0.048	0.045	0.044	0.045	0.055	0.054	0.045	0.045
		0.2	0.161	0.147	0.052	0.051	0.152	0.149	0.052	0.052
		0.4	0.498	0.444	0.081	0.082	0.478	0.438	0.084	0.083
		0.6	0.791	0.702	0.166	0.168	0.785	0.716	0.147	0.146
		0.8	0.929	0.830	0.268	0.268	0.916	0.844	0.251	0.254
		1.0	0.975	0.877	0.396	0.395	0.968	0.900	0.357	0.357
	$\varepsilon \sim \exp(1) - 1$	0	0.050	0.052	0.046	0.043	0.054	0.053	0.046	0.045
		0.2	0.125	0.117	0.052	0.051	0.121	0.112	0.052	0.053
		0.4	0.514	0.460	0.054	0.053	0.488	0.448	0.055	0.054
		0.6	0.857	0.759	0.143	0.143	0.843	0.764	0.129	0.129
		0.8	0.961	0.870	0.318	0.319	0.955	0.874	0.294	0.294
		1.0	0.982	0.891	0.529	0.531	0.983	0.907	0.492	0.491

5. A real data example

We now apply the proposed tests $\tilde{S}_n$ and $\tilde{U}_n$ to a real data set collected from an AIDS clinical trial to compare two treatment effects. The HIV positive patients in this study were randomly divided into four groups to receive antiretroviral regimen: (i) 600mg of zidovudine; (ii) 400mg of didanosine; (iii) 600mg of zidovudine plus 400 mg of didanosine and (iv) 600 mg of zidovudine plus 2.25 mg of zalcitabine. A more detailed description of this dataset can be found in Hammer et al. [9]. Several researchers have made use of this data set to illustrate their dimension reduction estimation methods with missing responses and further to compare the treatment effects of the first therapy (say (i)) and the other three therapies (say (ii)-(iv)), including Ding and Wang [1], Guo et al [5] and Hu et al [11]. Recently, Niu et al [25] analyzed this data set to check nonparametric component for partial linear regression model with missing responses at random and they made a conclusion that it is proper to fit this data set with a linear regression model.

In this dataset, there are 746 male patients who had not received antiretroviral therapy before the clinical trial. Further, based on the way of therapy, we divide our dataset into two subsets: the first dataset is 179 male patients receiving the first therapy and the second dataset is 567 patients receiving the other three therapies. Due to death and dropout, 199 patients in the first subset and

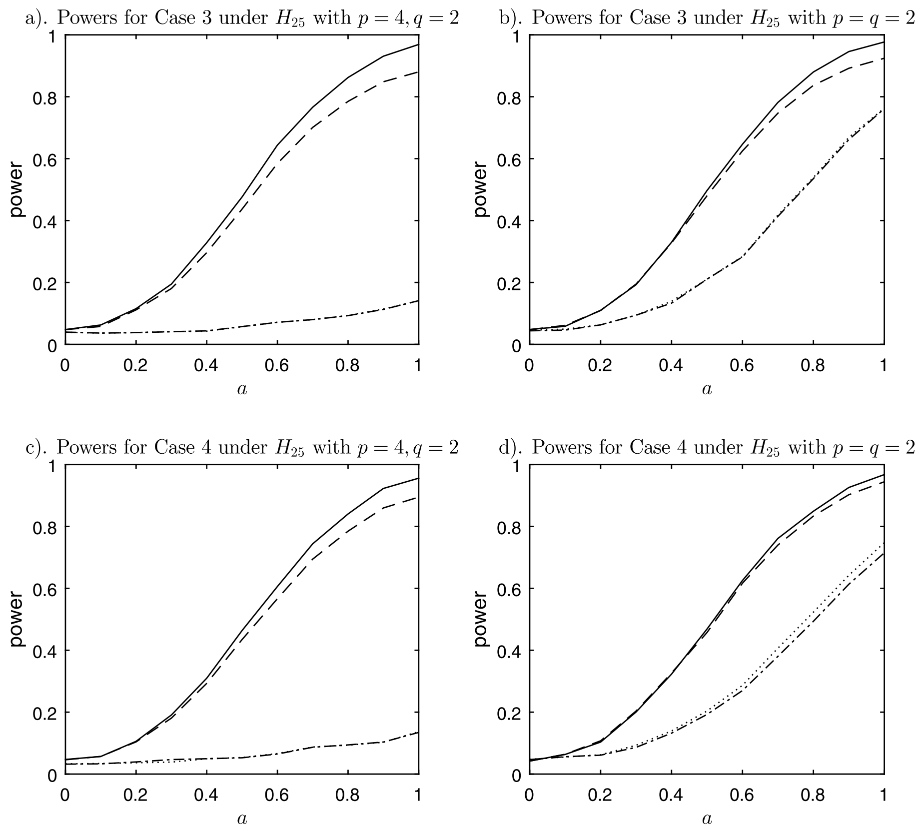

FIG 3. *Empirical sizes and powers of $\tilde{S}_n$, $\tilde{U}_n$, $T_n^\star$ and $V_n^\star$ for H_{20} v.s. H_{25} at the significance level $\alpha = 0.05$ with $n = 100$, $\varepsilon \sim N(0,1)$. In four plots, the solid line, the dash line, the dash-dotted line and the dotted line are for $\tilde{S}_n$, $\tilde{U}_n$, $T_n^\star$ and $V_n^\star$ respectively.*

74 patients in the second subset have missing response values. For each dataset, the response variable Y is CD4 cell counts at 96 ± 5 weeks post therapy and the six covariates $X = (X_1, \ldots, X_6)^\top$ are age (X_1), weight (X_2), CD4 cell counts at baseline (X_3), CD4 cell counts at 20 ± 5 weeks (X_4), CD8 cell counts at baseline (X_5) and CD8 cell counts at 20 ± 5 weeks (X_6). Let the indicator variable T denote the types of receiving therapies, with $T = 0$ for the first therapy and $T = 1$ otherwise. All of the indicators and covariates are observed.

Of primary interest is to test whether the data can be fitted by linear regression models: for the first dataset with $T = 0$

$$\begin{aligned} &H_{10} : E(Y|X) = \beta_{10} + \beta_1^\top X \text{ for some } \beta_{10} \in R \text{ and } \beta_1 \in R^6, \\ &H_{11} : E(Y|X) = m(B_1^\top X) \neq \beta_{10} + \beta_1^\top X \text{ for any } \beta_{10} \in R \text{ and } \beta_1 \in R^6 \end{aligned} \tag{5.1}$$

and for the second dataset with $T = 1$

$$H_{20} : E(Y|X) = \beta_{20} + \beta_2^\top X \text{ for some } \beta_{20} \in R \text{ and } \beta_2 \in R^6,$$

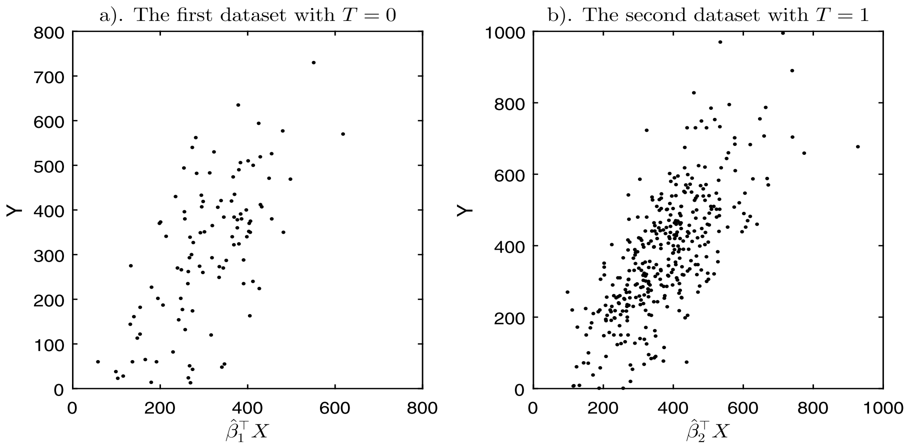

FIG 4. *Scatter plots for CD4 counts at 96 ± 5 weeks (Y) versus a). $\hat{\beta}_1 X$ for the first dataset $T = 0$ with complete observations; b). $\hat{\beta}_2 X$ for the second dataset $T = 1$ with complete observations.*

$$H_{21} : E(Y|X) = m(B_2^\top X) \neq \beta_{20} + \beta_2^\top X \text{ for any } \beta_{20} \in R \text{ and } \beta_2 \in R^6. \quad (5.2)$$

The same kernel function and bandwidth as those in the simulation section are adopted. With the proposed tests, we can obtain that for the first dataset, $\hat{\beta}_{10} = 47.900$ and $\hat{\beta}_1 = (-0.901, -0.845, 0.531, 0.483, -0.003, -0.020)^\top$. The adjusted tests $\tilde{S}_n$ and $\tilde{U}_n$ take the values of -0.484 and -0.492 respectively and the corresponding p-values are 0.686 and 0.689. For the second dataset with $T = 1$, $\hat{\beta}_{20} = -80.076$ and $\hat{\beta}_2 = (0.329, 0.938, 0.180, 0.771, 0.031, -0.057)^\top$. The test statistic $\tilde{S}_n$ and $\tilde{U}_n$ take values of 1.339 and 1.348 respectively with the corresponding p-values 0.090 and 0.089. All of these results indicate that the two datasets can be fitted by linear regression models.

To further compare the treatment effects for the two kinds of therapies, we first draw the scatter plots of the response Y versus $\hat{\beta}_1 X$ with complete observations for the first dataset and $\hat{\beta}_2 X$ with complete cases for the second dataset respectively. Figure 4 contains these two plots a) and (b) showing the linear patterns. Let the difference of the two subsets as $\Delta = E(Y|T = 1) - E(Y|T = 0)$. Based on the linear regression models, the estimator of Δ is 81.428, and the standard error obtained by the bootstrap approximation is 21.719.

6. Discussions

In this paper, two dimension reduction adaptive-to-model tests are proposed to implement model checking for parametric single-index models with missing responses at random. A significant advantage of the two tests is that the tests multiplying a normalizing constant $O(nh^{1/2})$ can converges to its weak limit, which is much faster than the typical convergence rate $O(nh^{p/2})$ for local smoothing tests, especially when the dimension p is large. Therefore, they can greatly avoid the curse of dimensionality, control type-I errors very well and have higher powers. Also, they can adapt to models under the null and alternative hypothesis automatically via estimating the structural dimension q that is a model selection step. This shows that the performance of the proposed procedure relies on a first-step model selection followed by a specification test. A relevant comment was given in Leeb and Pötscher [18]. A thorough study on hypothesis testing is deserved in the future.

Appendix. Proofs of theorems

The following conditions are asssumed for proving the theorems in Section 3.

(C1) $g(\beta^\top x, \theta)$ is a Borel measurable function on R^p for each θ and a twice continuously differentiable real function on a compact subset of $\theta \in \Theta$ for each $x \in R^p$. The matrix $\Sigma_1 = E\{\dot{g}(X, \beta_0, \theta_0)\dot{g}(X, \beta_0, \theta_0)^\top\}$ is nonsingular where $\dot{g}(\cdot)$ denotes the gradient of the function $g(\cdot)$ over $(\beta, \theta)^\top$.

(C2) The density function $f_\beta(z)$ of $\beta^\top X$ on support of $\mathcal{C}$ exists and has two bounded derivatives for all $\beta : |\tilde{\beta} - \beta| < \delta$ where $\delta > 0$ and satisfies

$$0 < \inf_{\beta^\top x \in \mathcal{C}} f(z) < \sup_{\beta^\top x \in \mathcal{C}} f(z) < 1.$$

(C3) $\pi(x)$ has bounded partial derivatives up to order 2 almost surely and $\inf_x \pi(x) > 0$. Assume $\pi(X)$ follows a semi-parametric structure, that is, $\delta = h(\gamma_0^\top X) + \varsigma$ and $\pi(X) = E(\delta|X) = E(\delta|\gamma_0^\top X)$, where $E(\varsigma|X) = 0$, $h(\cdot)$ is an unknown link function and γ_0 is an unknown p-dimensional parameter vector.

(C4) Under the null hypothesis and local alternative hypothesis, $nh^2 \to \infty$; Under the global alternative hypothesis, $nh^q \to \infty$.

(C5) The kernel function $\mathcal{K}(\cdot)$ is a spherically symmetric density function with a bounded derivative and all of the moments of $\mathcal{K}(\cdot)$ exist and $\int UU^\top K(U)dU = I$. The bandwidth satisfies $1/nh^2 \to 0$ and $h \to 0$.

(C6) $\sup E(\varepsilon^4|X = x) < \infty$, $E|X|^4 < \infty$ and $E|Y|^4 < \infty$.

(C7) Let $l(\psi^\top x)$ is the density of $\psi^\top x$, $l(\psi^\top x) = P(\delta = 1|\psi^\top X = \psi^\top x)$, which is assumed to be bounded away from zero and above. The three functions $l_1(\psi^\top x) = E(I(Y \in I_s)|\psi^\top X)$, $l_2(\psi^\top x) = \pi(\psi^\top x)l(\psi^\top x)$, $l_3(\psi^\top x) = l_1(\psi^\top x)l_2(\psi^\top x)l_2(\psi^\top x)$ are defined on a compact support and $\inf_{\psi^\top x} l_2(\psi^\top x) > 0$. $\pi(\psi^\top x)$, $l_1(\psi^\top x)$, $l_2(\psi^\top x)$ and $l_3(\psi^\top x)$ have bounded partial derivatives to order 2.

(C8) The estimate $\hat{B}$ based on semiparametric dimension reduction with CCAR is $\sqrt{n}$ consistent to the matrix B.

Remark 4. *Conditions* $(C1)$ *and* $(C6)$ *are necessary for the asymptotic normality of the least-squares estimator. Condition* $(C2)$ *and* $(C8)$ *are applied for guarantee the asymptotic normality of our test statistic. Condition* $(C3)$ *is a common assumption in missing data analysis, just as Guo et al [7] mentioned. Conditions* $(C4)$ *and* $(C5)$ *are the common requisite for the kernel density estimation problem. Condition* $(C7)$ *is wildly assumed in many papers, which is* *related to the smoothness of the response density function and regression curves. See e.g. Zhu and Fang [39] and Guo et al. [5].*

The following lemmas are used to prove the theorems in Section 3. We first give the proof of Lemma 1.

Proof of Lemma 1. We first consider the base estimate $\hat{\psi}$ in the CC analysis. Following Ding and Wang [1], in the following proof, without loss of generality, we assume that X has mean 0 and identity convariance matrix.

For the CC analysis, define $\Lambda = \sum_{s=1}^{S} p_s m_s m_s^\top$, where $p_s = P(\delta Y \in I_s)$, $m_s = E(\delta X|\delta Y \in I_s)$. Further denote $\hat{\Lambda} = \sum_{s=1}^{S} \hat{p}_s \hat{m}_s \hat{m}_s^T$, where $\hat{p}_s = n^{-1}\sum_{i=1}^{n} I(\delta_i y_i \in I_s)$, $\hat{m}_s = n^{-1}\sum_{i=1}^{n} I(\delta_i y_i \in I_s)\delta_i x_i/\hat{p}_s$. Note that

$$\hat{\Lambda} - \Lambda = \sum_{s=1}^{S}\left[(\hat{p}_s - p_s)m_s m_s^\top + p_s(\hat{m}_s - m_s)m_s^\top + p_s m_s(\hat{m}_s - m_s)^\top\right] + o_p(H_n),$$

here, $H_n = \sum_{s=1}^{S}[(\hat{p}_s - p_s)m_s m_s^\top + p_s(\hat{m}_s - m_s)m_s^\top + p_s m_s(\hat{m}_s - m_s)^\top]$. Since $\hat{m}_s = n^{-1}\sum_{i=1}^{n} I(\delta_i y_i \in I_s)\delta_i X_i/\hat{p}_s$, it is sufficient to study the convergence rate of $\hat{p}_s$ to p_s and $n^{-1}\sum_{i=1}^{n} I(\delta_i y_i \in I_s)\delta_i x_i$ to $E(\delta X I(\delta Y \in I_s))$. In the following, we only study the convergence rate of $\hat{p}_s$ to p_s. The other one can be proven similarly.

To check the distance of $\hat{p}_s$ and $P(\delta Y \in I_s)$ under the local alternative model and null model, respectively, denote the corresponding response under the null hypothesis (1.1) and the local alternative (3.2) as Y and Y_n. Note that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} I(\delta_i y_{in} \in I_s) - P(\delta Y \in I_s) \\
= \quad & \frac{1}{n}\sum_{i=1}^{n} I(\delta_i y_{in} \in I_s) - P(\delta Y_n \in I_s) + P(\delta Y_n \in I_s) - P(\delta Y \in I_s) \\
= \quad & O_p(\frac{1}{\sqrt{n}}) + P((\delta Y + C_n \delta m(B^\top X)) \in I_s) - P(\delta Y \in I_s) \\
= \quad & O_p(\frac{1}{\sqrt{n}}) + O_p(C_n) = O_p(C_n).
\end{aligned}$$

Similarly, we have $\hat{m}_s - m_s = O_p(C_n)$. Thus, we can finally get $\hat{\Lambda} - \Lambda = O_p(C_n)$, which implies that $\hat{\psi} - \psi = O_p(C_n)$. Under the local alternative (3.2), denote the relative notations as follows:

$$\begin{aligned}
\hat{h}(\hat{\psi}^\top x) &= \frac{1}{n}\sum_{j=1}^{n} \delta_j \mathcal{K}_h(\hat{\psi}^\top(x_j - x)), \\
\hat{H}(\hat{\psi}^\top x) &= \frac{1}{n}\sum_{j=1}^{n} \delta_j \mathcal{K}_h(\hat{\psi}^\top(x_j - x)) I(y_{jn} \in I_s), \\
\hat{m}^{CC}(\hat{\psi}^\top x) &= \frac{\sum_{j=1}^{n} \delta_j \mathcal{K}_h(\hat{\psi}^\top(x_j - x)) I(y_{jn} \in I_s)}{\sum_{j=1}^{n} \delta_j \mathcal{K}_h(\hat{\psi}^\top(x_j - x))},
\end{aligned}$$

here $\hat{\psi}$ is the base of $\mathcal{S}_{Y|X}$ in CC analysis.

Recall that

$$\hat{p}_s^{CCAR} = \frac{1}{n}\sum_{i,j=1}^{n}\Big\{[(1-\delta_i)\hat{p}_{ij}^{CC} + \delta_i I(i=j)]I(y_{jn}\in I_s)\Big\},$$

where

$$\hat{p}_{ij}^{CC} = \frac{\delta_j \mathcal{K}_h\{\hat{\psi}^\top(x_j - x_i)\}}{\sum_{j=1}^n \delta_j \mathcal{K}_h\{\hat{\psi}^\top(x_j - x_i)\}}.$$

Thus,

$$\begin{aligned}
\hat{p}_s^{CCAR} &= \frac{1}{n}\sum_{i=1}^{n}\Big\{\delta_i I(y_{in}\in I_s) + (1-\delta_i)\hat{m}^{CC}(\hat{\psi}^\top x_i)\Big\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big\{\delta_i I(y_{in}\in I_s) + (1-\delta_i)\frac{\hat{H}(\hat{\psi}^\top x_i)}{\hat{h}(\hat{\psi}^\top x_i)}\Big\} \\
&= \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{H(\hat{\psi}^\top x_i) - H(\psi^\top x_i)}{h(\psi^\top x_i)} \\
&\quad -\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{H(\hat{\psi}^\top x_i)[h(\hat{\psi}^\top x_i) - h(\psi^\top x_i)]}{h^2(\psi^\top x_i)} \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{\hat{H}(\psi^\top x_i) - H(\psi^\top x_i)}{h(\psi^\top x_i)} \\
&\quad -\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{H(\hat{\psi}^\top x_i)[\hat{h}(\psi^\top x_i) - h(\psi^\top x_i)]}{h^2(\psi^\top x_i)} \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}[\delta_i I(y_{in}\in I_s) + (1-\delta_i)m^{CC}(\psi^\top x_i)] \\
&=: D_1 - D_2 + D_3 - D_4 + D_5. \qquad \text{(A.1)}
\end{aligned}$$

As for the term D_1, we have

$$D_1 = \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{\dot{H}(\tilde{\psi}^\top x_i)(\hat{\psi}-\psi)^\top x_i}{h(\psi^\top x_i)} = O_p(C_n), \qquad \text{(A.2)}$$

where $\dot{H}(t) = \partial H/\partial t$. Similarly, we can obtain that $D_2 = O_p(C_n)$, $D_3 = O_p(1/\sqrt{n})$ and $D_4 = O_p(1/\sqrt{n})$. Further note that

$$\begin{aligned}
D_5 - P(Y\in I_s) &= D_5 - P(Y_n \in I_s) + P(Y_n\in I_s) - P(Y\in I_s) \\
&= O_p(\frac{1}{\sqrt{n}}) + P((Y + C_n m(B^\top X))\in I_s) - P(Y\in I_s) \\
&= O_p(\frac{1}{\sqrt{n}}) + O_p(C_n). \qquad \text{(A.3)}
\end{aligned}$$

Based on the above analysis for the formulae (A.1), (A.2) and (A.3), when $C_n = n^{-1/2}h^{-1/4}$, we have

$$\hat{p}_s^{CCAR} - P(Y \in I_s) = O_p(C_n).$$

Similarly, we can obtain that $\hat{m}_s^{CCAR} - m_s^{CCAR} = O_p(C_n)$, where $m_s^{CCAR} = E(X|Y_n \in I_s)$. Further, $\widehat{\mathrm{Cov}}\{E(X|Y_n)\} - \mathrm{Cov}\{E(X|Y)\} = O_p(C_n)$. Consequently, the convergence rate for eigenvalues of $\Sigma_X^{-1}\mathrm{Cov}\{E(X|Y_n)\}$ satisfies $\hat{\lambda}_i - \lambda_i = O_p(C_n)$, where $\lambda_i, i = 1, \dots, p$ are eigenvalues for $\Sigma_X^{-1}\mathrm{Cov}\{E(X|Y)\}$ under the null hypothesis.

Now we turn to prove the consistency of BIC criterion-based estimate when CCAR is adopted. Invoking the definition in (2.10), denote

$$G(l) = \frac{n}{2} \times \frac{\sum_{i=1}^{l}\{\log(\hat{\lambda}_i + 1) - \hat{\lambda}_i\}}{\sum_{i=1}^{p}\{\log(\hat{\lambda}_i + 1) - \hat{\lambda}_i\}} - 2 \times \sqrt{n} \times \frac{l(l+1)}{2p}.$$

When $l > 1$,

$$\begin{aligned} G(l) - G(1) &= \sqrt{n}\frac{l(l+1)-2}{p} - \frac{n\sum_{i=2}^{l}\{\log(\hat{\lambda}_i + 1) - \hat{\lambda}_i\}}{2\sum_{i=1}^{p}\{\log(\hat{\lambda}_i + 1) - \hat{\lambda}_i\}} \\ &= \sqrt{n}\left[\frac{l(l+1)-2}{p} - \frac{\sqrt{n}\sum_{i=2}^{l}\{\log(\hat{\lambda}_i + 1) - \hat{\lambda}_i\}}{2\sum_{i=1}^{p}\{\log(\hat{\lambda}_i + 1) - \hat{\lambda}_i\}}\right]. \end{aligned}$$

Note that $\log(\hat{\lambda}_i + 1) - \hat{\lambda}_i = -\hat{\lambda}_i^2/2 + o_p(\hat{\lambda}_i^2)$ and $\lambda_i = 0$ for any $l > 1$. We can further get that $\sum_{i=2}^{l}\{\log(\hat{\lambda}_i+1)-\hat{\lambda}_i\} = O_p(C_n^2)$ and $\sum_{i=1}^{p}\{\log(\hat{\lambda}_i+1)-\hat{\lambda}_i \to d$ in probability where d is a negative constant. When $C_n = n^{-1/2}h^{-1/4}$, it is not difficult to see that

$$\sqrt{n}\sum_{i=2}^{l}\{\log(\hat{\lambda}_i + 1) - \hat{\lambda}_i\} = (nh)^{-1/2} \to 0.$$

For any $l > 1$, $l(l+1) > 2$ holds. Therefore $P(G(1) > G(l)) \to 1$, which completes our proof. $\square$

Lemma 2. *Under the null hypothesis and conditions (C1)-(C7) in the Appendix, we have*

$$W_{n1} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{\delta_i}{\pi(x_i)}\frac{\delta_j}{\pi(x_j)}\mathcal{K}_h\{\hat{B}(\hat{q})^\top(x_i - x_j)\}e_i M(x_j) = O_p(1/\sqrt{n}),$$

$$\begin{aligned} W_{n2} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{\delta_i}{\pi(x_i)}\frac{\delta_j}{\pi(x_j)}\mathcal{K}_h\{\hat{B}(\hat{q})^\top(x_i - x_j)\}M(x_i)M(x_j)^\top \\ &= E\Big\{M(X)M(X)^\top f(\beta_0^\top X)\Big\} + o_p(1), \end{aligned}$$

$$W_{n3} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\delta_i\delta_j\mathcal{K}_h\{\hat{B}(\hat{q})^\top(x_i - x_j)\}e_i M(x_j) = O_p(1/\sqrt{n}),$$

$$W_{n4} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \delta_i \delta_j \mathcal{K}_h\{\hat{B}(\hat{q})^\top (x_i - x_j)\} M(x_i) M(x_j)^\top$$
$$= E\Big\{\pi^2(X) M(X) M(X)^\top f(\beta_0^\top X)\Big\} + o_p(1).$$

where $M(\cdot)$ is continuously differentiable and $E(M^2(X_l)|B^\top X) \le b(B^\top X)$ for $X_l \in R^p$ and $E[b(B^\top X)] < \infty$.

This can be obtained following the same argument as Guo et al [6], so we omit the details.

Lemma 3. *Suppose the conditions (C1)-(C7) in the Appendix hold and under the local alternatives (3.2), we have*

$$\begin{aligned}\sqrt{n}(\hat{\alpha} - \alpha_0) \quad &= \quad C_n \sqrt{n} \Sigma_1^{-1} E\{\dot{g}(X, \beta_0, \theta_0) m(B^\top X)\} \\ &\quad + \frac{\Sigma_1^{-1}}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i \dot{g}(X, \tilde{\beta}, \tilde{\theta})(y_i - g(\tilde{\beta}^\top x_i, \tilde{\theta}))}{\pi(x_i)} + o_p(1),\end{aligned}$$

where $\dot{g}(X, \beta_0, \theta_0) = \partial g(\alpha, x_j)/\partial \alpha^\top|_{\alpha=\alpha_0}$ and $\Sigma_1 = E\{\dot{g}(X, \beta_0, \theta_0)\dot{g}(X, \beta_0, \theta_0)^\top\}$.

The above Lemma can be similarly obtained based on Lemma 4.2 in Van Keilegom et al [29] and Lemma 3 in Guo et al [7], so we omit the detailed proof here.

In the following, we give the proof of Theorem 1.

Proof of Theorem 1. We first prove the asymptotic properties for V_n in (2.7). Denote $\mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\} = \mathcal{K}_h\{\hat{B}(\hat{q})^\top (x_i - x_j)\}$. For V_n, noting the symmetry of $\mathcal{K}_h(\cdot)$, it can be decomposed as three parts,

$$\begin{aligned} V_n \quad = \quad & \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{\delta_i}{\pi(x_i)} \frac{\delta_j}{\pi(x_j)} \mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\} \hat{e}_i \hat{e}_j \\ & - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{\delta_i \delta_j (\hat{\pi}(x_i) - \pi(x_i))}{\hat{\pi}(x_j)\hat{\pi}(x_i)\pi(x_i)} \mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\} \hat{e}_i \hat{e}_j \\ & - \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{\delta_i \delta_j (\hat{\pi}(x_i) - \pi(x_i))(\hat{\pi}(x_j) - \pi(x_j))}{\hat{\pi}(x_j)\hat{\pi}(x_i)\pi(x_i)\pi(x_j)} \mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\} \hat{e}_i \hat{e}_j \\ =: \quad & V_{n1} - 2V_{n2} - V_{n3}. \end{aligned} \tag{A.4}$$

Below we analyze the term V_{n1} first. Let $\alpha = (\beta, \theta)^\top$ and it can be divided as

$$\begin{aligned} V_{n1} = & \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{\delta_i}{\pi(x_i)} \frac{\delta_j}{\pi(x_j)} \mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\} e_i e_j \\ & - \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{\delta_i}{\pi(x_i)} \frac{\delta_j}{\pi(x_j)} \mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\} e_i \frac{\partial g(\alpha, x_j)}{\partial \alpha^\top} (\hat{\alpha} - \alpha_0) \end{aligned}$$

$$
\begin{aligned}
&+ (\hat{\alpha}-\alpha_0)^\top \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \frac{\delta_i}{\pi(x_i)} \frac{\delta_j}{\pi(x_j)} \\
&\times \mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\} \frac{\partial g(\alpha, x_i)}{\partial \alpha} \frac{\partial g(\alpha, x_j)}{\partial \alpha^\top} (\hat{\alpha}-\alpha_0) + o_p(V_{n1}^\star) \\
=: &\, V_{n11} - 2V_{n12} + V_{n13} + o_p(V_{n1}^\star), \qquad \text{(A.5)}
\end{aligned}
$$

where $V_{n1}^\star$ denotes the term $V_{n11} - 2V_{n12} + V_{n13}$. As for the term V_{n11}, we can make the decomposition as follows:

$$
\begin{aligned}
V_{n11} \quad = \quad & \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \frac{\delta_i}{\pi(x_i)} \frac{\delta_j}{\pi(x_j)} \mathcal{K}_h(B_{ij}) e_i e_j \\
& + \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \frac{\delta_i}{\pi(x_i)} \frac{\delta_j}{\pi(x_j)} \Big(\mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\} - \mathcal{K}_h(B_{ij})\Big) e_i e_j \\
=: \quad & V_{n11,1} + V_{n11,2}. \qquad \text{(A.6)}
\end{aligned}
$$

Since the dimension of $B^\top X$ is assumed to be fixed, the term $V_{n11,1}$ is an U-statistic. Note that under the null hypothesis H_0, we have β_0 is the true parameter, $q = 1$ and $\hat{q} \to 1$. By some tedious calculations and according to Theorem 1 of Hall [8], we can easy to derive the asymptotic normality:

$$
nh^{1/2} V_{n11,1} \to N(0, \Sigma^V), \qquad \text{(A.7)}
$$

here $\Sigma^V = 2\int \mathcal{K}^2(u)du \cdot \int \pi^{-2}(x) p^2(\beta_0^\top x)\{\sigma^2(\beta_0^\top x)\}^2 dx$ with $\sigma^2(\beta_0^\top x) = E(e^2|\beta_0^\top X = \beta_0^\top x)$. The function $p(\cdot)$ in Σ^V denotes the probability density function. Turn to the term $V_{n11,2}$ in (A.6) and an application of Taylor expansion yields

$$
V_{n11,2} = V_{n11,2}^\star + o_p(V_{n11,2}^\star),
$$

where

$$
V_{n11,2}^\star = \frac{h}{h^{\hat{q}}} \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \frac{\delta_i}{\pi(x_i)} \frac{\delta_j}{\pi(x_j)} \frac{1}{h} \mathcal{K}'(\frac{B_{ij}}{h})(x_i - x_j)^\top e_i e_j \cdot \frac{\hat{B}(\hat{q}) - B}{h}.
$$

Notice that the kernel function $\mathcal{K}(\cdot)$ is spherical symmetric, the following term

$$
\frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \frac{\delta_i}{\pi(x_i)} \frac{\delta_j}{\pi(x_j)} \frac{1}{h} \mathcal{K}'(\frac{B_{ij}}{h})(x_i - x_j)^\top e_i e_j
$$

can be regarded as an U-statistic. Further,

$$
\begin{aligned}
& E\Big\{\frac{\delta_i}{\pi(x_i)} \frac{\delta_j}{\pi(x_j)} \mathcal{K}'(\frac{B_{ij}}{h})(x_i - x_j)^\top e_i e_j | x_i, y_i\Big\} \\
= \quad & E\left\{E\Big[\frac{\delta_i}{\pi(x_i)} \frac{\delta_j}{\pi(x_j)} \mathcal{K}'(\frac{B_{ij}}{h})(x_i - x_j)^\top e_i e_j | x_i, y_i, x_j\Big] | x_i, y_i\right\}
\end{aligned}
$$

$$= \quad E\Big\{\mathcal{K}'(\frac{B_{ij}}{h})(x_i - x_j)^\top e_i E[e_j|x_j]|x_i, y_i\Big\} = 0.$$

Therefore, the above U-statistic is degenerate. Together with $\|\hat{B}(\hat{q}) - B\|_2 = O_p(1/\sqrt{n})$ and $nh^2 \to 0$, similarly as the proof of the term $V_{n11,1}$ above, we can derive

$$nh^{1/2}V^{\star}_{n11,2} = o_p(1). \tag{A.8}$$

Based on the formulae (A.6), (A.7) and (A.8), we have

$$nh^{1/2}V_{n11} \to N(0, \Sigma^V).$$

Turn to the term V_{n12} in (A.5), together with $\hat{\alpha} - \alpha = O_p(1/\sqrt{n})$ and based on the Lemma 2, we can obtain that $V_{n12} = O_p(1/n)$. Further, $nh^{1/2}V_{n12} = o_p(1)$. Now we consider the term V_{n13} in (A.5). Recalling the Lemma 2, we have

$$\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{\delta_i}{\pi(x_i)}\frac{\delta_j}{\pi(x_j)}\mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\}\frac{\partial g(\alpha, x_i)}{\partial \alpha}\frac{\partial g(\alpha, x_j)}{\partial \alpha^\top} = O_p(1).$$

Further,

$$nh^{1/2}V_{13} = nh^{1/2}\cdot O_p(\frac{1}{\sqrt{n}})\cdot O_p(1)\cdot O_p(\frac{1}{\sqrt{n}}) = O_p(h^{1/2}) = o_p(1).$$

Through the above analysis, we can conclude that

$$nh^{1/2}V_{n1} \to N(0, \Sigma^V). \tag{A.9}$$

As for the term V_{n2} in (A.4), similarly to the proof for V_{n1}, the key part in V_{n2} is $V_{n2,1}$, which is

$$V_{n2,1} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{\delta_i\delta_j\{\hat{\pi}(x_i) - \pi(x_i)\}}{\hat{\pi}(x_j)\hat{\pi}(x_i)\pi(x_i)}\mathcal{K}_h(B_{ij})e_ie_j.$$

In order to prove $nh^{1/2}V_{n2} = o_p(1)$, it is enough to prove that $nh^{1/2}V_{n2,1} = o_p(1)$. Since $E(V_{n2,1}) = 0$ and under H_0, $\hat{q} \to 1$, the variance of $V_{n2,1}$ can be computed as

$$\begin{aligned}\mathrm{Var}(V_{n2,1}) \quad &= \quad \frac{1}{n^2(n-1)^2h^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{k=1}^{n}\sum_{l\neq k}^{n} \\ &\quad \times E\Big\{\frac{\delta_i\delta_j\delta_k\delta_l(\hat{\pi}(x_j) - \pi(x_j))(\hat{\pi}(x_l) - \pi(x_l))}{\hat{\pi}(x_i)\hat{\pi}(x_j)\pi(x_j)\hat{\pi}(x_k)\hat{\pi}(x_l)\pi(x_l)} \\ &\quad \times\mathcal{K}\big(\frac{B^\top(x_i - x_j)}{h}\big)\mathcal{K}\big(\frac{B^\top(x_k - x_l)}{h}\big)e_ie_je_ke_l\Big\}\end{aligned}$$

As to the above formula, only the terms with $i = k, j = l$ and $i = l, j = k$ are not zero. When $i = k, j = l$, we have the following term in $\mathrm{Var}(V_{n2,1})$ as follows:

$\mathrm{Var}(V_{n2,1})_1$

$$= \frac{1}{n^2(n-1)^2h^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} E\Big[\frac{\delta_i\delta_j(\hat{\pi}(x_j)-\pi(x_j))^2}{\hat{\pi}^2(x_i)\hat{\pi}^2(x_j)\pi^2(x_j)}\mathcal{K}^2\big(\frac{B^\top(x_i-x_j)}{h}\big)e_i^2e_j^2\Big].$$

Further notice that

$$\begin{aligned}
& E\Big[\frac{\delta_i\delta_j(\hat{\pi}(x_j)-\pi(x_j))^2}{\hat{\pi}^2(x_i)\hat{\pi}^2(x_j)\pi^2(x_j)}\mathcal{K}^2\big(\frac{B^\top(x_i-x_j)}{h}\big)e_i^2e_j^2)\Big] \\
= \;& E\Big[\frac{(\hat{\pi}(x_j)-\pi(x_j))^2}{\pi(x_i)\pi^3(x_j)}\mathcal{K}^2\big(\frac{B^\top(x_i-x_j)}{h}\big)\sigma^2(x_i)\sigma^2(x_j)\Big]+o(1) \\
= \;& E\Big[\frac{(\hat{\pi}(x_j)-\pi(x_j))^2}{\pi(x_i)\pi^3(x_j)}\mathcal{K}^2\big(\frac{B^\top(x_i-x_j)}{h}\big)\times\sup_x(\hat{\pi}(x)-\pi(x))^2\Big] \\
\leq \;& E^{1/2}\Big[\frac{\sigma^4(x_i)\sigma^4(x_j)}{\pi^2(x_i)\pi^6(x_j)}\mathcal{K}^4\big(\frac{B^\top(x_i-x_j)}{h}\big)\Big]\times E^{1/2}[\sup_x(\hat{\pi}(x)-\pi(x))^4] \\
= \;& \Big[h\int\mathcal{K}^4(u)du\cdot\int(\sigma^4(x))^2p^2(x)\pi^{-8}(x)dx\Big]^{1/2}\times O(\frac{\ln(n)}{nh})=o(h)
\end{aligned}$$

The last formula applies the fact $E[\sup_x(\hat{\pi}(x)-\pi(x))]=O(\sqrt{\ln(n)/nh^{\hat{q}}})$ with $\hat{q}\to 1$ and it needs the condition that $nh^{3/2}\to\infty$. Therefore, we have $\mathrm{Var}(V_{n2,1})_1=o(n^{-2}h^{-1})$. Similarly, denote the term $i=l, j=k$ in $\mathrm{Var}(V_{n2,1})$ as $\mathrm{Var}(V_{n2,1})_2$ and we can also derive that $\mathrm{Var}(V_{n2,1})_2=o(n^{-2}h^{-1})$. At last, $V_{n2,1}=o(n^{-1}h^{-1/2})$ and it is also true for the term V_{n2}. Consequently,

$$nh^{1/2}V_{n2}=nh^{1/2}\cdot o(n^{-1}h^{-1/2})=o_p(1). \tag{A.10}$$

Similarly, we can get that

$$nh^{1/2}V_{n3}=o_p(1). \tag{A.11}$$

Combining the formulae (A.4), (A.9), (A.10) and (A.11) together, we have

$$nh^{1/2}V_n\to N(0,\Sigma^V).$$

In general, Σ^V is unknown, an estimate for it can be defined as

$$\hat{\Sigma}^V=\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{h^{\hat{q}}}\frac{\delta_i\delta_j}{\hat{\pi}^2(x_i)\hat{\pi}^2(x_j)}\mathcal{K}^2\Big(\frac{\hat{B}(\hat{q})^\top(x_i-x_j)}{h}\Big)\hat{e}_i^2\hat{e}_j^2.$$

The consistency of $\hat{\Sigma}^V$ can be proved similarly as the proof for Theorem 1 in Guo et al [6] via U-statistic theory. We omit it here.

Similarly, the asymptotic properties for T_n can be concluded. Thus, the proof for Theorem 1 is finished. □

Proof of Theorem 2. We first consider the asymptotic properties of T_n and V_n under the global alternative (1.2). Under this alternative, from White [30], we can obtain that $\hat{\alpha}-\tilde{\alpha}=O_p(1/\sqrt{n})$. However, $\hat{\alpha}$ is different from the true value $\alpha_0=(\beta_0,\theta_0)$ under the null hypothesis H_0. Denote $\Delta(x_i)=m(B^\top x_i)-g(\tilde{\beta}^\top x_i,\tilde{\theta})$, thus $\hat{e}_i=\varepsilon_i+\Delta(x_i)-[g(\hat{\beta}^\top x_i,\hat{\theta})-g(\tilde{\beta}^\top x_i,\tilde{\theta})]$. According to the U-statistic theory, it is not difficult to obtain that $T_n\Rightarrow E\{\pi^2(X)\Delta^2(X)f(B^\top X)\}$

and $V_n \Rightarrow E\{\Delta^2(X)f(B^\top X)\}$. Similarly, we can also prove that in probability $\hat{\Sigma}^T$ and $\hat{\Sigma}^V$ converge to positive values which are different from Σ^T and Σ^V, respectively. Therefore, it is obvious that $nh^{1/2}T_n \Rightarrow \infty$, $nh^{1/2}V_n \Rightarrow \infty$ and $S_n/(nh^{1/2}) \to \text{Constant} > 0$, $U_n/(nh^{1/2}) \to \text{Constant} > 0$, which completes the proof of the global alternative situation.

First, denote

$$\bar{V}_n = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \frac{\delta_i}{\pi(x_i)} \frac{\delta_j}{\pi(x_j)} \mathcal{K}_h\{\hat{B}(\hat{q})^\top (x_i - x_j)\}\hat{e}_i\hat{e}_j.$$

Similar as the proof in Theorem 1, it is not difficult to show

$$V_n = \bar{V}_n + o_p(\bar{V}_n). \tag{A.12}$$

Under the local alternative (3.2), we have $\hat{e}_i = [C_n m(B^\top x_i)+\varepsilon_i] - [g(\hat{\beta}^\top x_i, \hat{\theta}) - g(\beta_0^\top x_i, \theta_0)]$. Let $\mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\} = \mathcal{K}_h\{\hat{B}(\hat{q})^\top (x_i - x_j)\}$. As for the term $\bar{V}_n$, we have the following expansion

$$\begin{aligned}
\bar{V}_n &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \frac{\delta_i}{\pi(x_i)} \frac{\delta_j}{\pi(x_j)} \\
&\quad \times \mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\}[C_n m(B^\top x_i) + \varepsilon_i][C_n m(B^\top x_j) + \varepsilon_j] \\
&\quad - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \frac{\delta_i}{\pi(x_i)} \frac{\delta_j}{\pi(x_j)} \\
&\quad \times \mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\}[C_n m(B^\top x_i) + \varepsilon_i] \frac{\partial g(\alpha, x_j)}{\partial \alpha^\top}(\hat{\alpha} - \alpha_0) \\
&\quad + (\hat{\alpha} - \alpha_0)^\top \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \frac{\delta_i}{\pi(x_i)} \frac{\delta_j}{\pi(x_j)} \\
&\quad \times \mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\} \frac{\partial g(\alpha, x_i)}{\partial \alpha} \frac{\partial g(\alpha, x_j)}{\partial \alpha^\top}(\hat{\alpha} - \alpha_0) + o_p(\bar{V}_n^\star) \\
&=: \bar{V}_{n1} - \bar{V}_{n2} + \bar{V}_{n3} + o_p(\bar{V}_n^\star),
\end{aligned} \tag{A.13}$$

where $\bar{V}_n^\star = \bar{V}_{n1} - \bar{V}_{n2} + \bar{V}_{n3}$.

As for the term $\bar{V}_{n1}$ in (A.13),

$$\begin{aligned}
\bar{V}_{n1} \quad &= \quad \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \frac{\delta_i}{\pi(x_i)} \frac{\delta_j}{\pi(x_j)} \mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\}\varepsilon_i\varepsilon_j \\
&\quad + \frac{2C_n}{n(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \frac{\delta_i}{\pi(x_i)} \frac{\delta_j}{\pi(x_j)} \mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\}\varepsilon_i m(B^\top x_j) \\
&\quad + \frac{C_n^2}{n(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \frac{\delta_i}{\pi(x_i)} \frac{\delta_j}{\pi(x_j)} \mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\} m(B^\top x_i) m(B^\top x_j) \\
&=: \quad \bar{V}_{n1,1} + C_n \bar{V}_{n1,2} + C_n^2 \bar{V}_{n1,3}.
\end{aligned}$$

Similarly as the proof for V_{n11} in (A.5), we can obtain that

$$nh^{1/2}\bar{V}_{n1,1} \Rightarrow N(0, \Sigma^V).$$

Based on the Lemma 2, we have

$$\begin{aligned}\bar{V}_{n1,2} &= O_p(1/\sqrt{n})\\ \bar{V}_{n1,3} &= E[m^2(B^\top X)f(\beta_0^\top X)] + o_p(1).\end{aligned}$$

Consequently, when $C_n = n^{-1/2}h^{-1/4}$,

$$nh^{1/2}\bar{V}_{n1} \Rightarrow N(E[m^2(B^\top X)f(\beta_0^\top X)], \Sigma^V). \tag{A.14}$$

For the term $\bar{V}_{n2}$ in (A.13), we have

$$\begin{aligned}\bar{V}_{n2} &= \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{\delta_i}{\pi(x_i)}\frac{\delta_j}{\pi(x_j)}\mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\}\varepsilon_i\frac{\partial g(\alpha, x_j)}{\partial\alpha^\top}(\hat{\alpha}-\alpha_0)\\ &\quad + \frac{2C_n}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{\delta_i}{\pi(x_i)}\frac{\delta_j}{\pi(x_j)}\mathcal{K}_h\{\hat{B}(\hat{q})_{ij}\}m(B^\top x_i)\frac{\partial g(\alpha, x_j)}{\partial\alpha^\top}(\hat{\alpha}-\alpha_0)\\ &=: \bar{V}_{n2,1}(\hat{\alpha}-\alpha_0) + C_n\bar{V}_{n2,2}^\top(\hat{\alpha}-\alpha_0).\end{aligned}$$

Based on the Lemma 2, we have $\bar{V}_{n2,1} = O_p(1/\sqrt{n})$ and

$$\bar{V}_{n2,2} = 2E\{m(B^\top X)\dot{g}(X, \beta_0, \theta_0)f(\beta^\top X)\} + o_p(1),$$

where $\dot{g}(X, \beta_0, \theta_0) = \partial g(\alpha, x_j)/\partial\alpha^\top|_{\alpha=\alpha_0}$ and $f(\cdot)$ denotes the probability density function of $\beta_0^\top X$. Recall Lemma 3 and when $C_n = n^{-1/2}h^{-1/4}$, we have

$$\begin{aligned}&nh^{1/2}\bar{V}_{n2}\\ &\quad= nh^{1/2}\Big[O_p(n^{-1/2})O_p(C_n)\\ &\qquad + 2C_n^2E^\top\{m(B^\top X)\dot{g}(X, \beta_0, \theta_0)f(\beta^\top X)\}\Sigma_1^{-1}E\{\dot{g}(X, \beta_0, \theta_0)m(B^\top X)\}\Big]\\ &\quad= 2E^\top\{m(B^\top X)\dot{g}(X, \beta_0, \theta_0)f(\beta^\top X)\}\Sigma_1^{-1}E\{\dot{g}(X, \beta_0, \theta_0)m(B^\top X)\} + o_p(1).\end{aligned} \tag{A.15}$$

Turn to the term $\bar{V}_{n3}$ in (A.13), it is not difficult to conclude that

$$\begin{aligned}\bar{V}_{n3} &= (\hat{\alpha}-\alpha_0)^\top E\{\dot{g}(X, \beta_0, \theta_0)\dot{g}(X, \beta_0, \theta_0)^\top f(\beta^\top X)\}(\hat{\alpha}-\alpha_0) + o_p(C_n^2)\\ &= C_n^2E^\top\{\dot{g}(X, \beta_0, \theta_0)m(B^\top X)\}\Sigma_1^{-1}E\{\dot{g}(X, \beta_0, \theta_0)\dot{g}(X, \beta_0, \theta_0)^\top f(\beta^\top X)\}\\ &\quad\times \Sigma_1^{-1}E^\top\{\dot{g}(X, \beta_0, \theta_0)m(B^\top X)\} + o_p(C_n^2).\end{aligned}$$

Further, when $C_n = n^{-1/2}h^{-1/4}$,

$$nh^{1/2}\bar{V}_{n3} = E^\top\{\dot{g}(X, \beta_0, \theta_0)m(B^\top X)\}\Sigma_1^{-1}E\{\dot{g}(X, \beta_0, \theta_0)\dot{g}(X, \beta_0, \theta_0)^\top f(\beta^\top X)\}$$

$$\times\, \Sigma_1^{-1} E^\top\{\dot g(X,\beta_0,\theta_0)m(B^\top X)\} + o_p(1). \tag{A.16}$$

Combining the formulae (6), (A.14), (A.15), (A.12) and (A.16), we have

$$nh^{1/2}V_n \Rightarrow N(\mu^V, \Sigma^V),$$

where

$$\begin{aligned}
\mu^V &= E[m^2(B^\top X)f(\beta_0^\top X)] \\
&\quad - 2E^\top\{m(B^\top X)\dot g(X,\beta_0,\theta_0)f(\beta^\top X)\}\Sigma_1^{-1}E\{\dot g(X,\beta_0,\theta_0)m(B^\top X)\} \\
&\quad + E^\top\{\dot g(X,\beta_0,\theta_0)m(B^\top X)\}\Sigma_1^{-1}E\{\dot g(X,\beta_0,\theta_0)\dot g(X,\beta_0,\theta_0)^\top f(\beta^\top X)\} \\
&\quad \times \Sigma_1^{-1}E^\top\{\dot g(X,\beta_0,\theta_0)m(B^\top X)\} \\
&= E\Big[\Big(m(B^\top X) - \dot g(X,\beta_0,\theta_0)^\top\Sigma_1^{-1}E\{\dot g(X,\beta_0,\theta_0)m(B^\top X)\}\Big)^2 f(\beta_0^\top X)\Big].
\end{aligned}$$

At the same time, $\hat\Sigma^V$ is still a consistent estimate of Σ^V. Similarly, the asymptotic properties for the test statistic T_n can be also proved and we omit them here. The proof for Theorem 2 is finished. □

References

[1] Ding, X. B. and Wang, Q. H. (2011). Fusion-refinement procedure for dimension reduction with missing response at random. *J. Amer. Statist. Assoc.* **106** 1193–1207. MR2894774

[2] Escanciano, J. C. (2006). A consistent diagnostic test for regression models using projections. *Economet. Theor.* **22** 1030–1051. MR2328527

[3] Friedman, J. H. and Stuetzle, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823. MR0650892

[4] González-Manteiga, W. and Pérez-González, A. (2006). Goodness-of-fit tests for linear regression models with missing response data. *Can. J. Stat.* **34** 149–170. MR2280923

[5] Guo, X., Wang, T., Xu, W. L. and Zhu, L. X. (2014). Dimension reduction with missing response at random. *Comput. Statist. Data Anal.* **69** 228–242. MR3146891

[6] Guo, X., Wang, T. and Zhu, L. X. (2016). Model checking for parametric single index models: A dimension-reduction model-adaptive approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 1013–1035. MR3557187

[7] Guo, X., Xu, W. L. and Zhu, L. X. (2015). Model checking for paramatric regressions with response missing at random. *Ann. Inst. Statist. Math.* **67** 229–259. MR3315259

[8] Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.* **14** 1–16. MR0734096

[9] Hammer, S. M., et al. (1996). A trial comparing nucleotide monotherapy with combined therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter. *New Eng. J. Med.* **335** 1081–1090.

[10] Hsing, T. and Carroll, R. J. (1992). An asymptotic theory for sliced inverse regression. *Ann. Statist.* **20** 1040–1061. MR1165605

[11] Hu, Z. H., Follmann, D. A. and Qin J. (2010). Semiparametric dimension reduction estimation for mean response with missing data. *Biometrika* **97** 305–319. MR2650740

[12] Ingster Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives (I II and III). *Math. Methods Statist.* **2** 85–114, 171–189, 249–268. MR1259686, 1259687, 1259685

[13] Kim, J. K. and Shao, J. (2014). *Statistical Methods for Handling Incomplete Data*, Chapman & Hall/CRC. MR3307946

[14] Koul, H. L. and Ni, P. P. (2004). Minimum distance regression model checking. *J. Statist. Plann. Inference* **119** 109–141. MR2018453

[15] Lavergne, P., Maistre, S. and Patilea, V. (2015). A significance test for covariates in nonparametric regression. *Electron. J. Stat.* **9** 643–678. MR3331853

[16] Lavergne, P. and Patilea, V. (2008). Breaking the curse of dimensionality in nonparametric testing. *J. Econometrics* **143** 103–122. MR2384435

[17] Lavergne, P. and Patilea, V. (2012). One for all and all for one: regression checks with many regressors. *J. Bus. Econom. Statist.* **30** 41–52. MR2899183

[18] Leeb, H. and Pötscher, B. M. (2005). Model selection and inference facts and fiction. *Economet. Theor.* **1** 21–59. MR2153856

[19] Li, K. C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86** 316–327. MR1137117

[20] Li, L. and Lu, W. (2008). Sufficient dimension reduction with missing predictors. *J. Amer. Statist. Assoc.* **103** 822–831. MR2524333

[21] Li, X. Y. (2012). Lack-of-fit testing of regression model with response missing at random. *J. Statist. Plann. Inference* **142** 155–170. MR2827138

[22] Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*, New York, Wiley. MR0890519

[23] Maistre, S., Lavergne, P. and Patilea, V. (2017). Powerful nonparametric checks for quantile regression. *J. Statist. Plann. Inference* **180** 13–19. MR3554191

[24] Maistre, S. and Patilea, V. (2014). Nonparametric model checks of single-index assumptions. *arXiv: 1410.4934*.

[25] Niu, C. Z., Guo, X., Xu, W. L. and Zhu, L. X. (2016). Checking nonparametric component for partial linear regression model with missing response. *J. Statist. Plann. Inference* **168** 1–19. MR3412218

[26] Stute, W., Xu, W. L. and Zhu, L. X. (2008). Model Diagnosis for Parametric Regression in High Dimensional Spaces. *Biometrika* **95** 1–17. MR2521592

[27] Stute, W., Thies, S. and Zhu, L. X. (1998). Model checks for regression: an innovation process approach. *Ann. Statist.* **26** 1916–1934. MR1673284

[28] Sun, Z. H., Wang, Q. H. and Dai, P. J. (2009). Model checking for partially linear models with missing responses at random. *J. Multivariate Anal.* **100** 636–651. MR2478187

[29] Van Keilegom, I., Gonzáles-Manteiga, W. and Sánchez Sellero, C. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *Test* **17** 401–415. MR2434335

[30] White, H. (1981). Consequences and detection of misspecified nonlinear regression models. *J. Amer. Statist. Assoc.* **76** 419–433. MR0624344

[31] Xia, Y. C. (2009). Model checking in regression via dimension reduction. *Biometrika* **96** 133–148. MR2482140

[32] Xu, W. L. and Zhu, L. X. (2013). Testing the adequacy of varying coefficient models with missing responses at random. *Metrika* **76** 53–69. MR3018357

[33] Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75** 263–289. MR1413644

[34] Zhu, L. P., Wang, T. and Zhu, L. X. (2012). Sufficient dimension reduction in regression with missing predictors. *Statist. Sinica* **22** 1611–1637. MR3027101

[35] Zhu, L. P., Wang, T., Zhu, L. X. and Ferré, L. (2010a). Sufficient dimension reduction through discretization-expectation estimation. *Biometrika* **97** 295–304. MR2650739

[36] Zhu, L. P., Zhu, L. X. and Feng, Z. H. (2010b). Dimension reduction in regressions through cumulative slicing estimation. *J. Amer. Statist. Assoc.* **105** 1455–1466. MR2796563

[37] Zhu, L. X. (2003). Model checking of dimension-reduction type for regression. *Statist. Sinica* **13** 283–296. MR1977726

[38] Zhu, L. X. and An, H. Z. (1992). A nonlinearity test in regression models. *J. Mathe.* **12** 391–397. (in Chinese)

[39] Zhu, L. X. and Fang, K. T. (1996). Asymptotics for kernel estimate of sliced inverse regression. *Ann. Statist.* **24** 1053–1068. MR1401836

[40] Zhu, L. X., Miao, B. Q. and Peng, H. (2006). On sliced inverse regression with high-dimensional covariates. *J. Amer. Statist. Assoc.* **101** 630–643. MR2281245

[41] Zhu, L. X. and Ng, K. W. (1995). Asymptotics of sliced inverse regression. *Statist. Sinica* **5** 727–736. MR1347616