

# Dimension reduction-based significance testing in nonparametric regression\*

Xuehu Zhu<sup>†</sup>

*School of Mathematics and Statistics, Xi'an Jiaotong University, China  
Department of Mathematics, Hong Kong Baptist University, Hong Kong  
e-mail: zhuxuehu@xjtu.edu.cn*

Lixing Zhu<sup>‡</sup>

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

**Abstract:** A dimension reduction-based adaptive-to-model test is proposed for significance of a subset of covariates in the context of a nonparametric regression model. Unlike existing locally smoothing significance tests, the new test behaves like a locally smoothing test as if the number of covariates was just that under the null hypothesis and it can detect local alternative hypotheses distinct from the null hypothesis at the rate that is only related to the number of covariates under the null hypothesis. Thus, the curse of dimensionality is largely alleviated when nonparametric estimation is inevitably required. In the cases where there are many insignificant covariates, the improvement of the new test is very significant over existing locally smoothing tests on the significance level maintenance and power enhancement. Simulation studies and a real data analysis are conducted to examine the finite sample performance of the proposed test.

**MSC 2010 subject classifications:** Primary 62G10, 62H15; secondary 62H10.

**Keywords and phrases:** Significance testing, sufficient dimension reduction, ridge-type eigenvalue ratio.

Received March 2017.

## Contents

|   |  |      |
|---|--|------|
| 1 | Introduction . . . . .                                     | 1469 |
| 2 | Dimension reduction-based adaptive-to-model test . . . . . | 1473 |

---

\*Corresponding author: Lixing Zhu. The authors are grateful to the editor, the associate editor and the anonymous referees for the constructive comments and suggestions that led to the improvement of an early manuscript.

<sup>†</sup> Xuehu Zhu's research was supported by National Natural Science Foundation of China (No. 11601415, No. 11571204), China Postdoctoral Science Foundation (No. 2016M590934, No.2017T100731) and was supported by "the Fundamental Research Funds for the Central Universities".

<sup>‡</sup>Lixing Zhu's research was supported by a grant from the University Grants Council of Hong Kong, Hong Kong, China and by National Natural Science Foundation of China (No. 11671042).

2.1 Basic test statistic construction . . . . . 1473  
 2.2 A brief review on discretization-expectation estimation . . . . . 1475  
 2.3 Structural dimension estimation . . . . . 1477  
 3 Asymptotic properties . . . . . 1478  
 3.1 Limiting null distribution . . . . . 1478  
 3.2 Power study . . . . . 1478  
 4 Numerical studies . . . . . 1480  
 4.1 Simulations . . . . . 1480  
 4.2 Baseball hitters' salary data . . . . . 1487  
 5 Conclusions . . . . . 1488  
 Appendix . . . . . 1488  
 A.1 Regularity conditions . . . . . 1488  
 A.2 Proof of the theorems . . . . . 1489  
 References . . . . . 1504

**1. Introduction**

Consider the nonparametric regression model:

$$Y = m(Z) + \epsilon, \tag{1.1}$$

where  $Y$  is a scale dependent variable with the covariates  $Z = (X^\top, W^\top)^\top$ ,  $X = (X_1, \dots, X_{p_1})^\top \in \mathbb{R}^{p_1}$ ,  $W = (W_1, \dots, W_{p_2})^\top \in \mathbb{R}^{p_2}$  and  $p_1 + p_2 = d$ , the regression function  $m(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$  is unknown in its form and  $\epsilon$  is the error term with zero conditional expectation when  $Z$  is given:  $E(\epsilon|Z) = 0$ . As well known, the success of any further statistical analysis hinges on the correction of the working model. Note that part of the covariates are often redundant in regression models. The subset of the covariates  $W$  is said to be insignificant for the response variable  $Y$  given  $X$  if

$$E(Y|X, W) = E(Y|X). \tag{1.2}$$

The equality (1.2) means that  $W$  does not provide more information to predict  $Y$ .  $W$  should be removed from the regression model (1.1), otherwise, such redundant variables cause statistical analysis more complicated and less accurate and efficient. This step, particularly in the first stage of regression analysis, is necessary. In the literature, relevant testing problem called significance testing has attracted much attention. There exist several proposals that are based on prevalent locally and globally smoothing methodologies in the literature. For the former, Lavergne and Vuong (2000)[10] extended the idea introduced by Fan and Li (1996)[6], proposed a test based on a second conditional moment to check the significance of a subset of covariates. Further, Li (1999)[15] developed a nonparametric significance test that is based on the idea of Fan and Li (1996)[6] for nonparametric and semiparametric time-series models. Racine et al. (2006)[18] suggested a test for the significance of categorical variables in

fully nonparametric regression models. It is noted that these locally smoothing-based test statistics converge to the respective limiting null distributions at the typical rate  $O_p(n^{-1/2}h^{-d/4})$ , where  $d$  is the number of all the covariates and  $h$  is the bandwidth in kernel estimation. Further, these tests can also only detect the local alternative hypotheses distinct from the null hypothesis at the rate  $O_p(n^{-1/2}h^{-d/4})$ . When  $d = p_1 + p_2$  is large, the convergence rate is very slow because  $h$  converges to zero at a certain rate. This implies that these locally smoothing methodologies severely suffer from the curse of dimensionality. This problem is caused by using nonparametric estimation for the models under both the null and alternative hypothesis that assume the significance of all covariates. However, this clearly has not yet fully used the model structure under the null hypothesis. For globally smoothing tests, Racine (1997)[17] advised a test based on nonparametric estimation of partial derivative. Delgado and González Manteiga (2001)[4] introduced a consistent test based on a stochastic process. Both the tests have a fast rate of order  $\sqrt{n}$ . Racine's (1997)[17] test involves the nonparametric estimation of partial derivative  $\partial E(Y|X, W)/\partial W$  and thus, estimation inefficiency can seriously deteriorate the performance of the test. Delgado and González-Manteiga's (2001)[4] test involves the multivariate nonparametric estimation of  $E(Y|X)$  and all the covariates in the process. Thus, the data sparseness in high-dimensional space still causes negative impact for the power performance of the resulting test. We can see from the following example in Section 4 that the power drops down very quickly as  $d$  increases. Finally, when the dimension  $d$  of the covariates is large, the computational burden is also an issue because the Monte Carlo approximations to their sampling null distributions are computationally intensive.

We now present some simulation results of Example 1 in Section 4 to illustrate those disadvantageous phenomena. Here the sample size is  $n = 200$ , the dimension of  $X$  is set to be  $p_1 = 2, 4$  and the dimension  $p_2$  of  $W$  is chosen to be 2 to 10 in this numerical simulation. The empirical powers are simulated across through 2000 replications for each experiment at the significance level  $\alpha = 0.05$ . For the presentation purpose in the figure, the power is set to be 1 when  $p_2 = 0$ . We use this example to examine how the powers of existing tests are negatively affected by increasing the dimension  $p_2$ .

Figure 1 depicts the empirical power curves of Fan and Li's (1996)[6] test and Delgado and González Manteiga's test (2001)[4] that are regarded as the representatives of local smoothing tests and global smoothing tests. More details about the bandwidth selection and other details for this simulation can be found in Section 4. Obviously, the empirical powers of these two tests rapidly decrease as  $p_1$  and  $p_2$  increase. This indicates that both the tests severely suffer from the curse of dimensionality. Relevant discussions on the significance level maintenance will be discussed below. To this end, how to overcome the aforementioned problems caused by dimensionality is of great importance.

Recently, Lavergne et al. (2015)[11] devised a new kernel-based test that is based on a suitable equivalent Fourier transformation. Their test is based on an idea that combines locally smoothing and globally smoothing test in construction. Namely, they locally smooth the regression function with covariates under

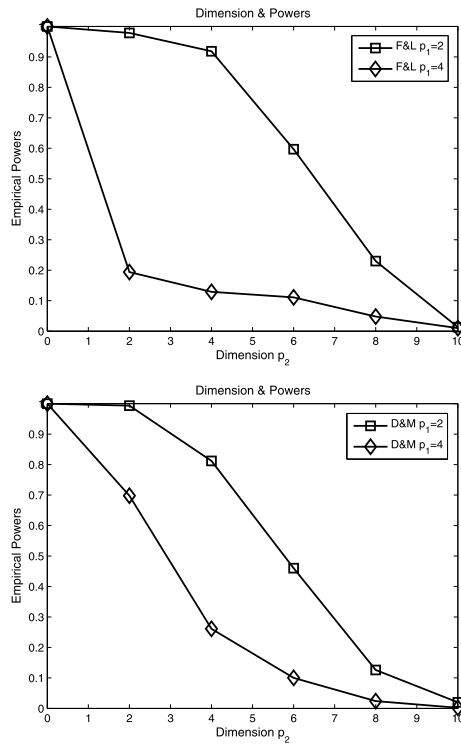


FIG 1. The empirical power curves of Fan and Li's (1996) test and Delgado and González Manteiga's (2001) test against the dimensions of  $X$  and  $W$ , where  $X \sim N(0, \Sigma_1)$  and  $W \sim N(0, \Sigma_1)$  with sample size  $n = 200$  and  $a = 2$  at the significance level  $\alpha = 0.05$  for Example 1.

the null hypothesis, and globally smooth by using the average over the regression function with the other covariates. The nice feature is that it inherits the merit of locally smoothing tests that can be sensitive to high-frequency alternative models and is still omnibus. Theoretically, the test statistic has a faster rate of order  $O(n^{-1/2}h^{-p_1/4})$  to its weak limit rather than the rate  $O(n^{-1/2}h^{-d/4})$  that classical locally smoothing tests achieve. But even under the null hypothesis, it still uses all covariates. Thus, to maintain the significance level, the test still encounters the data sparseness issue even under the null hypothesis.

How to construct a test that involves all covariates only under the alternative hypothesis is of interest. Guo et al. (2016)[8] advocated an adaptive-to-model strategy to construct test statistics. They first proposed a dimension-reduction locally smoothing test which used to test generalized linear regression models. Under the null hypothesis, the test solely uses the information provided by the hypothetical model such that it can well maintain the significance level, and under the alternative hypothesis it can automatically adapt to the alternative model such that the test is omnibus. Zhu et al. (2017a)[30] followed the similar idea to develop a dimension reduction globally smoothing test for more general

regression models. It is of interest to apply this strategy in test statistic construction in significance testing. However, such an application is far from trivial. As the key, identifying the projected covariates in the hypothetical and alternative models plays a crucial role in this method. In Guo et al. (2016)[8], the projected covariates in the hypothetical models are contained in the alternative models. In the present setup, we cannot impose this assumption and thus, the identification procedure will be different from the model adaptation step of Guo et al. (2016)[8]. Therefore, we need a new method to tackle this issue. The details will be seen from the following model and the test statistic construction in the next section.

As is known, the objective of significance testing focuses on choosing the significant covariates  $X$  in the nonparametric regression setting. Let  $g(x) = E(Y|X = x)$ . Then the significance testing becomes the following hypothesis:

$$H_0 : E(Y|X, W) = g(X) \quad \text{versus} \quad H_1 : E(Y|X, W) \neq g(X). \quad (1.3)$$

Let  $U = Y - g(X)$ . Recall  $Z = (X^\top, W^\top)^\top$ . Then we have under the null hypothesis  $H_0$ ,  $E(U|Z) = m(Z) - g(X) = 0$ , but under the alternative hypothesis  $H_1$ ,  $E(U|Z) \neq 0$ . To facilitate a more general model structure we want to test, consider the following reformulation of the above model. Note that  $g(\cdot)$  is an unknown function, the nonparametric regression model  $Y = g(X) + U$  can be reformulated as:

$$Y = g(X) + U = g(B_1 B_1^\top X) + U \equiv: \tilde{g}(B_1^\top X) + U,$$

where  $B_1$  is an orthonormal  $p_1 \times p_1$  matrix. This means that the above regression model can be viewed as a special multi-index regression model with  $p_1$  indices corresponding to the covariates  $X$ . Thus, in the present paper, we consider a general null hypothesis as:

$$H_0 : E(Y|Z) = E(Y|X) = g(B_1^\top X) \quad \text{a.s.}, \quad (1.4)$$

where  $B_1$  is a  $p_1 \times q_1$  orthonormal matrix for a given  $q_1$  with  $1 \leq q_1 \leq p_1$ . The hypothetical regression model covers many popularly used models in the literature, including the single-index model, the multi-index model and the partially linear single-index model. When the above regression model is a single-index model or partially linear single index model, the corresponding number  $q_1$  of the indices becomes one or two, respectively. Particularly, when  $q_1 = p_1$ , the hypothetical model becomes the classical model (1.2). Consider a general alternative model. Let  $B$  be a  $d \times q_2$  orthonormal matrix with  $B^\top B = I_{q_2 \times q_2}$  and  $q_2$  being an unknown number with  $d \geq q_2 > q_1$ . The alternative hypothesis is

$$H_1 : E(Y|Z) = m(B^\top Z) \neq g(B_1^\top X) = E(Y|X). \quad (1.5)$$

Note that model (1.2) can also be written in this format. That is, under  $H_0$ , we have a matrix  $(B_1^\top, 0_{p_2 \times q_1}^\top)$  and  $B_1^\top X = (B_1^\top, 0_{p_2 \times q_1}^\top)Z$ .

We develop a test that solely depends on the linear combinations of the significant covariates  $X$  under the null hypothesis  $H_0$ , and automatically uses

the linear combinations of all the covariates under the alternative hypothesis  $H_1$  such that the test is omnibus. We will call it a Dimension Reduction Adaptive-to-Model test (DREAM). Compared with existing locally smoothing tests, we will show that, rather than at the rate of order  $O(n^{-1/2}h^{-d/4})$ , DREAM converges to their weak limit at the rate of order  $O(n^{-1/2}h^{-q_1/4})$  and may detect the local alternative models distinct from the null model also at this rate. Further, it is worthwhile to mention that almost all existing locally smoothing tests require the assistance of Monte Carlo approximation to determine critical values, although the limiting null distribution is tractable, otherwise, the significance level is difficult to maintain. In contrast, the dimension reduction structure of our test alleviates this difficulty, the critical values of the new test can be determined by its limiting null distribution even when the sample size is moderate. Similarly as the conventionally local smoothing tests, the proposed test has the subjective constraint choice of bandwidth, which is done by the rule of thumb in this paper. It naturally hopes that an optimal bandwidth be selected. However, as we know that choosing optimal bandwidths for local smoothing tests is still an open problem in theory (Stute and Zhu, 2005)[20]. To see the impact from the bandwidth selection, we implement several simulation studies and find that this issue is not critical for the proposed test when the rule of thumb is applied to select the bandwidth. More details will be discussed in the next sections.

This paper is organized as follows. In Section 2, the test statistic construction is described. Because dimension reduction technique plays a very important role, we first briefly review a promising method: the discretization-expectation estimation. To make the test have the model-adaptation property, a ridge-type eigenvalue ratio estimate (RERE) for the dimension  $q$  of  $B$  is recommended, which can consistently estimate  $q_1$  and  $q_2$  under the null and alternative hypotheses accordingly. The asymptotic properties of the test statistic are presented in Section 3. Further, the test statistic tends to infinity at the certain rate under the global alternative hypothesis in Section 3. In Section 4, we examine the finite sample performance of our test and also apply it to a real data analysis for illustration. All the technical conditions and the proofs of the theoretical results are postponed to the Appendix.

## 2. Dimension reduction-based adaptive-to-model test

### 2.1. Basic test statistic construction

It is worth noticing that in the above formulations,  $B$  and  $B_1$  are usually not identifiable. Actually, under  $H_1$ , for any  $q_2 \times q_2$  orthogonal matrix  $C$ ,  $E(Y|Z) = E(Y|B^\top Z) = E(Y|CB^\top Z)$ . And under  $H_0$ , for any  $q_1 \times q_1$  orthogonal matrix  $C_1$ ,  $E(Y|X) = E(Y|B_1^\top X) = E(Y|C_1B_1^\top X)$ . Hence, what we can identify is  $BC^\top$  for a matrix  $C$  and  $B_1C_1^\top$  for a matrix  $C_1$ . Note that it is enough to have such a weaker identification because under the null hypothesis,  $C_1B_1^\top X = C_1(B_1^\top, 0_{p_2 \times q_1}^\top)Z$  does not involve  $W$  for any  $q_1 \times q_1$  orthogonal matrix  $C_1$  and  $E(Y|Z) = E(Y|C_1B_1^\top X)$ . In the following subsection, we will briefly introduce

a method to identify  $BC^\top$  for a matrix  $C$  and  $B_1C_1^\top$  for a matrix  $C_1$ . Without notational confusion, we use  $B$  and  $B_1$  to write  $BC^\top$  and  $B_1C_1^\top$ , respectively, throughout the rest of the present paper.

Recall  $U = Y - g(X)$ . Under the null hypothesis  $H_0$ , we have

$$E(U|Z) = E(\epsilon|B^\top Z) = E(\epsilon|B_1^\top X) = 0.$$

Then

$$E(U E(U|B^\top Z) W(B^\top Z)) = E(E^2(\epsilon|B^\top Z) W(B^\top Z)) = 0, \quad (2.1)$$

where  $W(B^\top Z)$  is some positive weight function that will be discussed below. Under the alternative hypothesis  $H_1$ , since

$$E(U|B^\top Z) = m(B^\top Z) - g(B_1^\top X) \neq 0,$$

we have

$$E(U E(U|B^\top Z) W(B^\top Z)) = E(E^2(U|B^\top Z) W(B^\top Z)) > 0. \quad (2.2)$$

The above argument implies that the empirical version of the left hand side in (2.1) can be used as a base to construct a test statistic. Further, the null hypothesis  $H_0$  is rejected for large values of the test statistic. This motivates a naive construction as any one in the literature. However, we also note that under the null hypothesis, for a selected weight function  $\tilde{W}$ ,

$$E(U E(U|B^\top Z) W(B^\top Z)) = E(E^2(\epsilon|B_1^\top X) \tilde{W}(B_1^\top X)) = 0. \quad (2.3)$$

This means that under the null hypothesis, the matrix  $B$  of dimension  $q_2$  is reduced to the matrix of dimension  $q_1$ . To have a uniform version, we define  $\tilde{B} = (B_1^\top, 0_{q_1 \times p_2})^\top$  such that  $B_1^\top X = \tilde{B}^\top Z$ .

The key is how to construct a test statistic that fully uses this piece of information and can automatically adapt the model structure under the alternative hypothesis such that the test is still omnibus. We will present our idea in the following construction. When a sample  $\{(z_1, y_1), \dots, (z_n, y_n)\}$  is available, the residual term  $u_i$  is estimated as  $\hat{u}_i = y_i - \hat{g}(\hat{B}_1^\top x_i)$ , where  $\hat{g}(\hat{B}_1^\top x_i)$  is a kernel estimate of  $g(B_1^\top x_i)$  as follows:

$$\hat{g}(\hat{B}_1^\top x_i) = \frac{\frac{1}{n-1} \sum_{j \neq i}^n Q_{h_1}(\hat{B}_1^\top x_j - \hat{B}_1^\top x_i) y_j}{\frac{1}{n-1} \sum_{j \neq i}^n Q_{h_1}(\hat{B}_1^\top x_j - \hat{B}_1^\top x_i)},$$

and  $Q_{h_1} = Q(\cdot/h_1)/h_1^{q_1}$  with  $Q(\cdot)$  being a  $q_1$ -dimensional product kernel function from the univariate kernel  $\tilde{Q}(\cdot)$ ,  $h_1$  being a bandwidth and  $\hat{B}_1$  being an estimate of  $B_1$ . Then we obtain the following kernel estimate of  $E(U|B^\top Z)$  (or  $E(U|\tilde{B}^\top Z)$ ) as:

$$\hat{E}(\hat{u}_i | \hat{B}_1^\top z_i) = \frac{\frac{1}{n-1} \sum_{j \neq i}^n K_{\hat{q}h}(\hat{B}_1^\top z_j - \hat{B}_1^\top z_i) \hat{u}_j}{\frac{1}{n-1} \sum_{j \neq i}^n K_{\hat{q}h}(\hat{B}_1^\top z_j - \hat{B}_1^\top z_i)}.$$

In this formula,  $\hat{B}$  is an estimate of  $B$  (or  $\tilde{B}$ ) with an estimated structural dimension  $\hat{q}$  in a certain sense that will be specified in the next subsection, where  $K_{\hat{q}h} = K(\cdot/h)/h^{\hat{q}}$  with  $K(\cdot)$  being a  $\hat{q}$ -dimensional kernel function and  $h$  being a bandwidth. If we choose the weight  $W(\cdot)$  (or  $\tilde{W}(\cdot)$ ) to be the density function  $p_B(\cdot)$  (or  $p_{\tilde{B}}(\cdot)$ ) of  $B^\top Z$  (or  $\tilde{B}^\top Z$ ), a non-standardized test statistic can be constructed as  $V_n$  by:

$$V_n = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \hat{u}_i \hat{u}_j K_{\hat{q}h}(\hat{B}^\top z_j - \hat{B}^\top z_i). \quad (2.4)$$

**Remark 2.1.** Note that the test statistic developed by Fan and Li (1996) is:

$$\tilde{V}_n = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \hat{u}_i \hat{f}_{1i} \hat{u}_j \hat{f}_{1j} \tilde{K}_h(z_i - z_j), \quad (2.5)$$

where  $\tilde{K}_h(\cdot) = K(\cdot/h)/h^d$  with  $K(\cdot)$  being a  $(p_1 + p_2)$ -dimensional kernel function and  $\hat{f}_1$  is an estimate of the density function  $f_1(\cdot)$  of  $X$  no matter whether the underlying model is under the null or alternative hypothesis. Compared the formula (2.4) with (2.5), the difference is that our test uses  $\hat{B}^\top Z$  in lieu of  $Z$  and applies  $K_h(\cdot)$  in  $V_n$  instead of  $\tilde{K}_h(\cdot)$ . At first glance, this difference seems not fundamental because it just follows the test statistic construction in the classical way without use of the model structures under the null and alternative hypothesis. Note that under the null, the dimension  $q_1$  is smaller than the dimension  $q_2$ . We then want to construct an estimate  $\hat{B}_2$ , which has the following model adaptation properties: under  $H_0$ , it estimates  $\tilde{B}$  and under  $H_1$  it automatically estimates  $B$ . This model adaptation property can allow us to fully use the dimension reduction structure under the null hypothesis. Thus, how to construct such an estimate plays a crucial role for this model-adaptation. To achieve this goal, we also need an estimate  $\hat{q}$  that can be consistent to  $q_1$  under  $H_0$  and to  $q_2$  under  $H_1$ . We will see that a standardizing version  $nh^{q_1/2}V_n$  can be used such that it has a finite limit under  $H_0$  and diverges to infinity much faster than  $nh^{d/2}\tilde{V}_n$  in Fan and Li's (1996) test statistic. The results are reported in Section 3.

## 2.2. A brief review on discretization-expectation estimation

As we commented above, we need to identify  $BC^\top$  and  $B_1C_1^\top$ . To this end, a method is discussed in this subsection. The method is to identify the spaces spanned by  $B$  and  $B_1$  automatically under the null and alternative hypothesis. In other words, the method is to identify the basis vectors in the respective subspaces. This is an estimation issue for the central mean subspaces in sufficient dimension reduction (e.g. Cook, 1998)[1]. The respective central mean subspaces are respectively denoted as  $S_{E(Y|X)}$  and  $S_{E(Y|Z)}$  that is related to the conditional independence between  $Y$  and  $X$  given  $B_1^\top X$  and between  $Y$  and  $Z$  given  $B^\top Z$ . Also,  $q_1 = \dim(S_{E(Y|X)})$  and  $q_2 = \dim(S_{E(Y|Z)})$  are respectively called the structural dimensions of  $S_{E(Y|X)}$  and  $S_{E(Y|Z)}$ . Here we assume that  $q_1$  is given, but  $q_2$  is unknown.



There exist several dimension reduction proposals available in the literature. For example, Li (1991)[14] proposed sliced inverse regression (SIR), Cook and Weisberg (1991)[3] advised sliced average variance estimation (SAVE), Xia et al. (2002)[22] discussed minimum average variance estimation (MAVE), Li and Wang (2007)[12] presented directional regression (DR). Cook and Forzani (2009)[2] developed likelihood acquired directions (LAD), Zhu et al. (2010a)[24] suggested discretization-expectation estimation (DEE) and Zhu et al. (2010b)[25] provided average partial mean estimation (APME). In this paper, we adopt DEE because it is computationally inexpensive without any tuning parameter selection that is required by SIR, SAVE or DR, and can be easily used to construct a criterion for determining  $q$ .

From Zhu et al. (2010a)[24], to identify and estimate  $B$  (and  $\tilde{B}$ ), the DEE estimation procedure can be summarized as the following steps. Without notational confusion, we consider identify  $B$  first and later we clarify that such an identification can automatically be applied to that for  $\tilde{B}$ .

1. Define the set of binary variables  $\Upsilon(t) = I\{Y \leq t\}$  by discretizing the response variable  $Y$ , where the indicator function  $I\{Y \leq t\} = 1$  if  $Y \leq t$  and 0, otherwise.
2. Let  $S_{\Upsilon(t)|Z}$  denote the central subspace of  $\Upsilon(t)|Z$ , and  $M(t)$  be an  $d \times d$  positive semi-definite matrix satisfying  $\text{Span}\{M(t)\} = S_{\Upsilon(t)|Z}$ .
3. Let  $\tilde{Y}$  denote an independent copy of  $Y$ . Taking the expectation over the random variable  $\tilde{Y}$ , the target matrix becomes  $M = E\{M(\tilde{Y})\}$ .  $B$  consists of the eigenvectors associated with the  $q_2$  nonzero eigenvalues of  $M$ .
4. Get an estimation of the target matrix  $M$  as:

$$M_n = \frac{1}{n} \sum_{i=1}^n M_n(y_i),$$

where  $M_n(y_i)$  is the estimating matrix of  $M(y_i)$  by some certain sufficient dimension reduction method such that SIR. Then the estimate  $\hat{B}$  consists of the eigenvectors associated with the largest  $q_2$  eigenvalues of  $M_n$ .  $\hat{B}$  is root- $n$  consistent to the matrix  $B$  when  $q_2$  is given.

Note that under the null hypothesis,  $S_{E(Y|Z)}$  is reduced to  $S_{E(Y|X)}$  because the conditional independence between  $Y$  and  $Z$  is equivalent to this independence between  $Y$  and  $X$ . In Step 2 above, sufficient dimension reduction theory tells us that the central subspace is reduced to  $S_{\Upsilon(t)|X}$  and the matrix in Step 3 above. The matrix  $M = E\{M(\tilde{Y})\}$  could have  $q_1$  nonzero eigenvalues, and the estimate  $\hat{q}$  could automatically be consistent to  $q_1$ . The above estimation procedure is called discretization-expectation estimation (DEE) in Zhu et al. (2010a)[24], more details are referred to this paper.

The following proposition states the consistency of the estimated matrix  $\hat{B}$  under  $H_0$ .

**Proposition 2.1.** *Under  $H_0$  and Conditions A1 and A2 in the Appendix, the DEE-based estimate  $\hat{B}$  is consistent to  $(C_1 B_1^\top, 0_{q_1 \times p_2})^\top$  with some  $q_1 \times q_1$  orthonormal matrix  $C_1$ .*

Proposition 2.1 indicates that under  $H_0$ , in probability the test statistic only uses those variables that are significant. The curse of dimensionality can then be largely alleviated when nonparametric estimation is inevitably required.

However, as  $q_2$  is unknown, to accommodate the alternative hypothesis, we should estimate it and  $B$  consistently. Let  $q$  be  $q_1$  under the null and  $q_2$  under the alternative. Thus, we want to define an estimate of  $q$ , which could be adaptive to the underlying model and thus be consistent to  $q_1$  and  $q_2$  under the respective hypotheses. The following subsection provides an estimate and its consistency.

### 2.3. Structural dimension estimation

We define a criterion to estimate  $q$  in an automatic manner. Although the BIC-type criterion in Zhu et al. (2010a)[24] that was motivated from Zhu et al. (2006)[27] is a candidate, choosing an appropriate penalty is a difficult issue.

In this paper, we recommend a ridge-type eigenvalue ratio estimate (RERE). Based on our experience in practice, it is not very sensitive to the ridge choice. Let  $\hat{\lambda}_d \leq \hat{\lambda}_{d-1} \leq \dots \leq \hat{\lambda}_1$  be the eigenvalues of the estimating matrix  $M_n$ . Define

$$\lambda_i^* = \frac{\hat{\lambda}_i - \frac{1}{\sqrt{n}}}{\hat{\lambda}_i - \frac{1}{\sqrt{n}} + 1}, \text{ for } 1 \leq i \leq d. \tag{2.6}$$

Theoretically, we can use the original  $\hat{\lambda}_i$  to define the following ratio-based criterion. However, we found that a couple of largest eigenvalues tend to be much larger than the other non-zero eigenvalues. Then some ratios of estimated non-zero eigenvalues could be smaller than the minimizer, and then the structure dimension  $q$  is often underestimated. Thus, we ‘standardize’ the eigenvalues  $\lambda_i^*$  to define a criterion and estimator:

$$\hat{q} = \arg \min_{1 \leq j \leq d} \left\{ j : \frac{(\lambda_{j+1}^*)^2 + c_n}{(\lambda_j^*)^2 + c_n} \right\}. \tag{2.7}$$

This method is motivated by Xia et al. (2015)[21], Zhu et al. (2016)[29] and Zhu et al. (2017b)[31]. This algorithm is easily implemented.

The following proposition states the estimation consistency.

**Proposition 2.2.** *In addition to Conditions A1, A2 and A3 in the Appendix, assume  $c \times \log n/n \leq c_n \rightarrow 0$  with some fixed  $c > 0$ . Then the estimate  $\hat{q}$  by (2.7) has the following consistency.*

- (i) Under  $H_0$ ,  $P(\hat{q} = q_1) \rightarrow 1$ .
- (ii) Under  $H_1$ ,  $P(\hat{q} = q_2) \rightarrow 1$ .

From Proposition 2.2, the choice of  $c_n$  can be in a relatively wide range to guarantee the estimation consistency under the null and global alternative hypotheses.

Altogether, the final estimate  $\hat{B}$  can have the model adaptive property in the sense that under  $H_0$ , it is consistent to  $\tilde{B} = (C_1 B_1^\top, 0_{q_1 \times p_2})^\top$  for a  $q_1 \times q_1$  orthogonal matrix  $C_1$  and under  $H_1$ , to  $B$ .

As  $\hat{q}$  is equal to  $q_1$  under the null and larger than  $q_1$  under the alternatives, it seems that we could simply use it to perform a test. However, based on it, neither type I error nor power can be evaluated. Thus, it can only be an estimate rather than a test statistic.

### 3. Asymptotic properties

#### 3.1. Limiting null distribution

First, define some notations. Let

$$s^2 = 2 \int K^2(u) du E\{[\text{Var}(\epsilon^2 | \tilde{B}^\top Z)]^2 p(\tilde{B}^\top Z)\}, \quad (3.1)$$

and

$$\hat{s}^2 = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n K_h^2(\hat{B}^\top z_i - \hat{B}^\top z_j) \hat{u}_i^2 \hat{u}_j^2. \quad (3.2)$$

where  $p(\cdot)$  denotes the density function of  $B_1^\top X$ .

**Theorem 3.1.** *Under  $H_0$  and the regularity conditions in the Appendix, we have*

$$nh^{q_1/2} V_n \Rightarrow N(0, s^2).$$

Further,  $s^2$  can be consistently estimated by  $\hat{s}^2$ .

Therefore, according to Theorem 3.1, we can get the standardized test statistic as:

$$T_n = nh^{q_1/2} V_n / \hat{s}.$$

Further, applying the Slutsky's Theorem yields that under  $H_0$ , the test statistic  $T_n$  is asymptotically normal:

$$T_n \Rightarrow N(0, 1).$$

#### 3.2. Power study

We now study the power performance of the test statistic  $T_n$ . Consider the following sequence of local alternative hypotheses as:

$$H_{1n} : E(Y|Z) = g(B_1^\top X) + C_n G(B^\top Z). \quad (3.3)$$

Fixed  $C_n$  corresponds to the global alternative model and when  $C_n$  goes to zero, the sequences are the local alternative hypotheses.

To obtain the main results about the power performance under  $H_{1n}$  of (3.3), we first present the asymptotic behavior of the estimate  $\hat{q}$  when  $C_n \rightarrow 0$ . It is

noted that the estimate  $\hat{q}$  may be not consistently converge to  $q_2$ . This is caused by the convergence of the local alternative models to the hypothetical model as  $n \rightarrow \infty$ , the following lemma states the result.

**Proposition 3.1.** *Under  $H_{1n}$  of (3.3) and the regularity conditions in Proposition 2.2 except that  $c \times n^{-1}h^{-q_1/2} \log n \leq c_n \rightarrow 0$  for some fixed  $c > 0$ , as  $n \rightarrow 0$ ,  $\hat{q}$  determined by (2.7) has the following properties.*

- (I) If  $C_n = O(n^{-1/2}h^{-q_1/4})$ ,  $P(\hat{q} = q_1) \rightarrow 1$ ;
- (II) If  $C_n \times n^{1/2}h^{q_1/4} \rightarrow \infty$ ,  $P(\hat{q} \geq q_1) \rightarrow 1$ .

Comparing this proposition with Proposition 2.2, we can see the main difference is that when  $C_n$  has slower rate than  $O(n^{-1/2}h^{-q_1/4})$ , the criterion we are using cannot determine to which value the estimate is consistent. From the proof, it seems that with this rate  $C_n$  converges to zero, the estimation consistency of  $\hat{q}$  is difficult to satisfy. This is an ongoing research topic beyond the scope of this paper. As we do not know whether  $\hat{q}$  converges, and furthermore, even it could be convergent, whether it converges to a constant. It affects the power study described here. The following theorem states how sensitive the test is to the alternative models in other cases.

**Theorem 3.2.** *Under the regularity conditions in the Appendix, we have the following results.*

- (I) Under  $H_{1n}$  with a fixed  $C_n > 0$ ,

$$T_n/(nh^{q_1/2}) \Rightarrow \text{Constant} > 0.$$

- (II) Under  $H_{1n}$  with  $C_n = n^{-1/2}h^{-q_1/4}$ ,

$$nh^{q_1/2}V_n \Rightarrow N(u, s^2),$$

where  $u = E([E\{G(B^\top Z)|B_1^\top X\}]^2 p_{B_1}(B_1^\top X))$  with the density function  $p_{B_1}(\cdot)$  of  $B_1^\top X$ . and  $s^2$  is given by (3.1). Further,  $s^2$  can be consistently estimated by  $\hat{s}^2$ .

**Remark 3.1.** *The results in this theorem confirm our claim in the first section. The convergence rate of the test statistic is  $nh^{q_1/2}$  and the test can detect the local alternative models converging to the hypothetical model also at the rate of order  $C_n = n^{-1/2}h^{-q_1/4}$  under certain constraint on the alternative model structure. Fan and Li's (1996)[6] test, which is also the case for existing locally smoothing tests, can have the respective rates where  $q_1$  is replaced by  $d$ , which causes a much slower rate. However, as we mentioned right before the theorem, when  $C_n$  has slower rate to zero than the rate of order  $O(n^{-1/2}h^{-q_1/4})$ , the property of the estimate  $\hat{q}$  is not clear and thus the asymptotic property of the test statistic is not clear although it would reasonably guess that the power could be higher or be equal to 1 asymptotically. In the numerical studies, we will consider several values in a range of  $C_n$  to see whether the power could be higher with increasing the value of  $C_n$ . When the estimator  $\hat{q}$  is less than the true structural dimension  $q$ , we can also check whether the test statistic  $V_n$  based on  $\hat{q}$  can detect the corresponding alternatives with positive probability.*

## 4. Numerical studies

### 4.1. Simulations

In this subsection, we conduct the simulations to investigate the finite sample performance of our proposed test. The empirical sizes and powers are computed via 2000 replications of the experiments at the significance level  $\alpha = 0.05$ . Write the DREAM test as  $T_n^{DEE}$ . For comparison, we use Fan and Li's (1996) test and Delgado and González Manteiga's (2001)[4] test as the representatives of existing local and global tests. Write them as  $T_n^{FL}$  and  $T_n^{DM}$ .

Delgado and González Manteiga's (2001)[4] test is defined as:

$$\tilde{V}_n = \frac{1}{n(n-1)} \sum_{i=1}^n \left\{ \sum_{j \neq i}^n \hat{u}_j \hat{f}_{1j} I(x_j < x_i) I(w_j < w_i) \right\}^2.$$

As its limiting null distribution is intractable, the critical values are determined by the wild bootstrap. The bootstrap observations are from:  $y_i^* = \hat{g}(x_i) + \hat{u}_i \times v_i$ , where  $\{v_i\}_{i=1}^n$  is a sequence of i.i.d. random variables from the two-point distribution as:

$$P\left(v_i = \frac{1 - \sqrt{5}}{2}\right) = \frac{1 + \sqrt{5}}{2\sqrt{5}}, \quad P\left(v_i = \frac{1 + \sqrt{5}}{2}\right) = 1 - \frac{1 + \sqrt{5}}{2\sqrt{5}}.$$

The bootstrap critical values are computed by 1000 bootstrap replications. The Gaussian-based kernel of order 4,  $Q(u) = (u^4 - 7u^2 + 6)\phi(u)/2$ , is used to estimate the nonparametric function  $g(\cdot)$ , where  $\phi(\cdot)$  denotes the standard normal density, see Fan and Hu (1992)[5]. For both DREAM and Fan and Li's (1996)[6] test, we use the Quartic kernel function as  $K(u) = \frac{15}{16}(1 - u^2)^2$ , if  $|u| \leq 1$  and 0, otherwise, in constructing the test statistic such as that in (2.4). To determine the structural dimension  $q$ ,  $c_n = 0.1 \log n/nh^{\frac{q_1}{2}}$  is used.

The observations  $\{x_i\}_{i=1}^n$  and  $\{w_i\}_{i=1}^n$  are i.i.d., respectively, from multivariate normal distribution  $N(0, \Sigma_1)$ ,  $N(0, \Sigma_2)$ , or  $N(0, \Sigma_3)$  and independent of the standard normal errors, in which  $\Sigma_1 = (\sigma_{ij}^{(1)})$ ,  $\Sigma_2 = (\sigma_{ij}^{(2)})$  and  $\Sigma_3 = (\sigma_{ij}^{(3)})$  with  $\sigma_{ij}^{(1)} = I(i = j) + 0I(i \neq j)$ ,  $\sigma_{ij}^{(2)} = I(i = j) + 0.5^{|i-j|}I(i \neq j)$  and  $\sigma_{ij}^{(3)} = I(i = j) + 0.2I(i \neq j)$ . In the following examples,  $\epsilon$  is from standard normal distribution.

In this section, we design 4 examples. The first example is to show that when the dimensions  $p_1$  and  $p_2$  are small and the model is low-frequent, how the performances of the three competitors are. In Example 2, the dimensions grow up to higher under a high-frequent model, we then check the impact from the dimensionality. Example 3 is to examine whether DREAM is still omnibus even when the test statistic fully uses the information of low dimensionality under the null hypothesis. The model in Example 4 is with higher dimension of  $B^\top Z$  and then we can see whether, like existing locally smoothing tests, DREAM also fails to work. The details are in the examples.

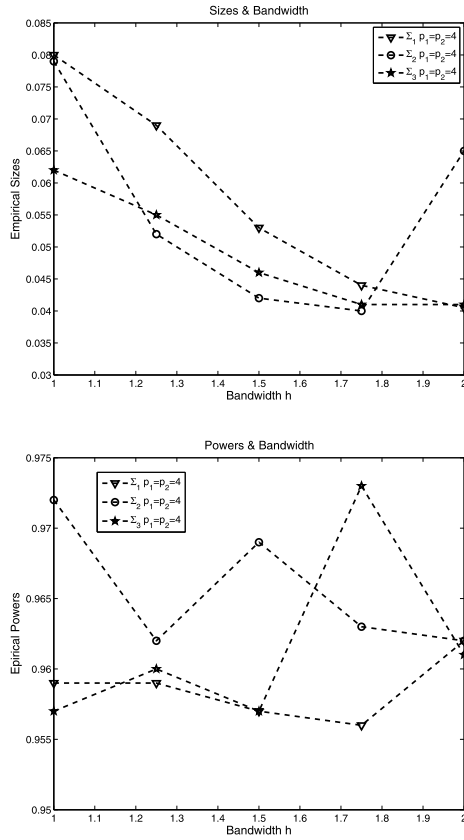


FIG 2. The empirical sizes and powers curves of  $T_n^{DEE}$  against the bandwidth  $h$  with sample size  $n = 200$  at the significance level  $\alpha = 0.05$  for Example 1.

**Example 1.** Consider the linear regression model:

$$\bullet Y = 2\beta_1^\top X + 2a \times \beta_2^\top W + 0.5 \times \epsilon,$$

where  $\beta_1 = (\underbrace{1, \dots, 1}_{p_1/2}, 0, \dots, 0)^\top / \sqrt{p_1/2}$ ,  $\beta_2 = (0, \dots, 0, \underbrace{1, \dots, 1}_{p_2/2})^\top / \sqrt{p_2/2}$ . In

this example,  $p_1 = p_2 = 2$  and  $p_1 = p_2 = 4$  are considered, where the hypothetical and alternative models respectively respond to  $a = 0$  and  $a \neq 0$ . To check the sensitivity of the bandwidth selection, we choose the different bandwidths  $h = c \times n^{-1/4+\hat{q}}$  for  $c \in \{1, 1.25, 1.5, 1.75, 2\}$ .

Figure 2 reports the empirical sizes and powers with the above bandwidths when  $n = 200$ . The empirical power is relatively robust against the different bandwidths that we use. The empirical size is not very sensitive to the bandwidth. Thus, the bandwidth  $h = 1.75 \times n^{-1/(4+\hat{q})}$  is recommended throughout the simulations.

TABLE 1  
Empirical sizes and powers of  $T_n^{DEE}$ ,  $T_n^{FL}$  and  $T_n^{DM}$  with  $p_1 = 2$  and  $p_2 = 2$  in Example 1.

|                         | $a/n$ | $T_n^{DEE}$ |        |        | $T_n^{FL}$ |        |        | $T_n^{DM}$ |        |        |
|-------------------------|-------|-------------|--------|--------|------------|--------|--------|------------|--------|--------|
|                         |       | 50          | 100    | 200    | 50         | 100    | 200    | 50         | 100    | 200    |
| $X \sim N(0, \Sigma_1)$ | 0     | 0.0515      | 0.0485 | 0.0515 | 0.0570     | 0.0595 | 0.0550 | 0.0490     | 0.0520 | 0.0545 |
|                         | 0.4   | 0.1060      | 0.1535 | 0.4530 | 0.1560     | 0.2465 | 0.4560 | 0.3665     | 0.8490 | 0.9710 |
| $W \sim N(0, \Sigma_1)$ | 0.8   | 0.4060      | 0.7320 | 0.9230 | 0.3075     | 0.5880 | 0.9335 | 0.4420     | 0.8935 | 0.9860 |
|                         | 1.2   | 0.6020      | 0.8770 | 0.9570 | 0.4550     | 0.7750 | 0.9655 | 0.4485     | 0.9075 | 0.9805 |
|                         | 1.6   | 0.6680      | 0.9045 | 0.9695 | 0.5115     | 0.8700 | 0.9780 | 0.4590     | 0.9050 | 0.9890 |
|                         | 2.0   | 0.7320      | 0.9195 | 0.9785 | 0.5415     | 0.8945 | 0.9890 | 0.4375     | 0.9135 | 0.9935 |
| $X \sim N(0, \Sigma_2)$ | 0     | 0.0400      | 0.0635 | 0.0515 | 0.0615     | 0.0790 | 0.0580 | 0.0475     | 0.0470 | 0.0460 |
|                         | 0.4   | 0.0935      | 0.1450 | 0.3600 | 0.1610     | 0.2345 | 0.4230 | 0.3670     | 0.8520 | 0.9770 |
| $W \sim N(0, \Sigma_2)$ | 0.8   | 0.3435      | 0.6770 | 0.9015 | 0.3780     | 0.5430 | 0.9295 | 0.4400     | 0.9055 | 0.9880 |
|                         | 1.2   | 0.5655      | 0.8365 | 0.9420 | 0.5265     | 0.7085 | 0.9510 | 0.4410     | 0.9005 | 0.9930 |
|                         | 1.6   | 0.6650      | 0.8985 | 0.9520 | 0.6020     | 0.8340 | 0.9795 | 0.4480     | 0.9015 | 0.9890 |
|                         | 2.0   | 0.7110      | 0.9015 | 0.9675 | 0.6145     | 0.8980 | 0.9910 | 0.4645     | 0.9035 | 0.9960 |
| $X \sim N(0, \Sigma_3)$ | 0     | 0.0515      | 0.0525 | 0.0520 | 0.0590     | 0.0640 | 0.0555 | 0.0455     | 0.0535 | 0.0465 |
|                         | 0.4   | 0.0985      | 0.1155 | 0.3825 | 0.1225     | 0.2440 | 0.3270 | 0.3660     | 0.8495 | 0.9750 |
| $W \sim N(0, \Sigma_3)$ | 0.8   | 0.3640      | 0.7455 | 0.9380 | 0.3035     | 0.5905 | 0.8015 | 0.4275     | 0.8935 | 0.9870 |
|                         | 1.2   | 0.5930      | 0.8675 | 0.9470 | 0.4150     | 0.8190 | 0.9355 | 0.4455     | 0.9095 | 0.9865 |
|                         | 1.6   | 0.6830      | 0.9070 | 0.9640 | 0.5185     | 0.8685 | 0.9660 | 0.4605     | 0.8950 | 0.9875 |
|                         | 2.0   | 0.7495      | 0.9175 | 0.9700 | 0.5495     | 0.9055 | 0.9765 | 0.4925     | 0.9045 | 0.9880 |

TABLE 2  
Empirical sizes and powers of  $T_n^{DEE}$ ,  $T_n^{FL}$  and  $T_n^{DM}$  with  $p_1 = 4$  and  $p_2 = 4$  in Example 1.

|                         | $a/n$ | $T_n^{DEE}$ |        |        | $T_n^{FL}$ |        |         | $T_n^{DM}$ |        |        |
|-------------------------|-------|-------------|--------|--------|------------|--------|---------|------------|--------|--------|
|                         |       | 50          | 100    | 200    | 50         | 100    | 200     | 50         | 100    | 200    |
| $X \sim N(0, \Sigma_1)$ | 0     | 0.0545      | 0.0435 | 0.0520 | 0.0310     | 0.0620 | 0.0670  | 0.0020     | 0.0045 | 0.0240 |
|                         | 0.4   | 0.0730      | 0.1690 | 0.3770 | 0.0520     | 0.0615 | 0.0805  | 0.0060     | 0.0325 | 0.2205 |
| $W \sim N(0, \Sigma_1)$ | 0.8   | 0.3420      | 0.7355 | 0.9190 | 0.0550     | 0.0695 | 0.0955  | 0.0055     | 0.0480 | 0.2645 |
|                         | 1.2   | 0.5635      | 0.8660 | 0.9465 | 0.0545     | 0.0700 | 0.0950  | 0.0090     | 0.0425 | 0.2485 |
|                         | 1.6   | 0.6290      | 0.8860 | 0.9610 | 0.0530     | 0.0790 | 0.0960  | 0.0065     | 0.0520 | 0.2545 |
|                         | 2.0   | 0.6960      | 0.9115 | 0.9825 | 0.0550     | 0.0870 | 0.1290  | 0.0025     | 0.0485 | 0.2610 |
| $X \sim N(0, \Sigma_2)$ | 0     | 0.0620      | 0.0550 | 0.0525 | 0.0330     | 0.0600 | 0.0565  | 0.0250     | 0.0440 | 0.0800 |
|                         | 0.4   | 0.0815      | 0.1550 | 0.3320 | 0.0685     | 0.0820 | 0.0930  | 0.1260     | 0.2005 | 0.5100 |
| $W \sim N(0, \Sigma_2)$ | 0.8   | 0.3555      | 0.6950 | 0.9165 | 0.0810     | 0.1005 | 0.1205  | 0.1410     | 0.3930 | 0.7040 |
|                         | 1.2   | 0.5390      | 0.8295 | 0.9540 | 0.0900     | 0.1065 | 0.1415  | 0.1270     | 0.4260 | 0.7420 |
|                         | 1.6   | 0.6330      | 0.8795 | 0.9615 | 0.1000     | 0.1215 | 0.1430  | 0.1400     | 0.4285 | 0.7600 |
|                         | 2.0   | 0.6965      | 0.9090 | 0.9795 | 0.1070     | 0.1360 | 0.1595  | 0.1410     | 0.4180 | 0.7720 |
| $X \sim N(0, \Sigma_3)$ | 0     | 0.0400      | 0.0420 | 0.0555 | 0.0320     | 0.0705 | 0.06255 | 0.0070     | 0.0280 | 0.0715 |
|                         | 0.4   | 0.0680      | 0.1325 | 0.3900 | 0.0660     | 0.0720 | 0.07500 | 0.0410     | 0.1265 | 0.6300 |
| $W \sim N(0, \Sigma_3)$ | 0.8   | 0.3325      | 0.6960 | 0.9170 | 0.0815     | 0.0775 | 0.09505 | 0.0520     | 0.2320 | 0.6640 |
|                         | 1.2   | 0.5500      | 0.8345 | 0.9490 | 0.0830     | 0.0915 | 0.10005 | 0.0530     | 0.2315 | 0.6545 |
|                         | 1.6   | 0.6245      | 0.8845 | 0.9535 | 0.0865     | 0.1040 | 0.11850 | 0.0630     | 0.2320 | 0.6520 |
|                         | 2.0   | 0.6740      | 0.8890 | 0.9775 | 0.0875     | 0.1030 | 0.11655 | 0.0560     | 0.2500 | 0.6430 |

The results of the three tests under different combinations of sample sizes, dimensions of covariates  $X$  and  $W$  and covariance matrices  $\Sigma$  are reported in Tables 1 and 2.

From Table 1, we can clearly observe that when the dimensions  $p_1$  and  $p_2$  are lower, all the tests have similar empirical powers and can control the empirical sizes well. The power performances of the competitors are very good. However, from Table 2, we can see that with increasing the dimensions  $p_1$  and  $p_2$ ,  $T_n^{DEE}$

is significantly and uniformly more powerful than  $T_n^{FL}$  and  $T_n^{DM}$ . Meanwhile  $T_n^{DEE}$  can still well maintain the significance level. Comparing Table 1 with Table 2, we can see that the dimensions of  $X$  and  $W$  have less influence for  $T_n^{DEE}$  than they do for  $T_n^{FL}$  and  $T_n^{DM}$ . When  $p_1 = p_2 = 4$ , Fan and Li's (1996)[6] test completely fails to detect the alternative hypothesis with a power similar to the significance level even when  $n = 200$ . Further, DREAM is robust against the correlation structure of  $(X, Z)$  whereas it significantly influences the power performance of Delgado and González Manteiga's (2001)[4] test, particularly when  $p_1 = p_2 = 4$ .

**Example 2.** In this example, consider a nonlinear high-frequency regression model as:

- $Y = 2 \sin(\beta_1^\top X) + 2a \times \sin(\beta_2^\top W) + 0.5 \times \epsilon,$

where  $\beta_1 = (\underbrace{1, \dots, 1}_{p_1/2}, 0, \dots, 0)^\top / \sqrt{p_1/2}$  and  $\beta_2 = (0, \dots, 0, \underbrace{1, \dots, 1}_{p_2/2})^\top / \sqrt{p_2/2}$ .

We also consider two cases of dimensions:  $p_1 = p_2 = 4$  and  $p_1 = p_2 = 6$ . Again  $a = 0$  responds to the hypothetical model.

TABLE 3  
Empirical sizes and powers of  $T_n^{DEE}$ ,  $T_n^{FL}$  and  $T_n^{DM}$  with  $p_1 = 4$  and  $p_2 = 4$  in Example 2.

|                         | $a/n$ | $T_n^{DEE}$ |        |        | $T_n^{FL}$ |        |        | $T_n^{DM}$ |        |        |
|-------------------------|-------|-------------|--------|--------|------------|--------|--------|------------|--------|--------|
|                         |       | 50          | 100    | 200    | 50         | 100    | 200    | 50         | 100    | 200    |
| $X \sim N(0, \Sigma_1)$ | 0     | 0.0575      | 0.0520 | 0.0495 | 0.0360     | 0.0320 | 0.0695 | 0.0060     | 0.0100 | 0.0180 |
|                         | 0.4   | 0.0805      | 0.2225 | 0.6120 | 0.0665     | 0.0780 | 0.0910 | 0.0080     | 0.0520 | 0.1225 |
| $W \sim N(0, \Sigma_1)$ | 0.8   | 0.2910      | 0.7200 | 0.9190 | 0.0775     | 0.0875 | 0.0915 | 0.0040     | 0.0560 | 0.1790 |
|                         | 1.2   | 0.5465      | 0.8295 | 0.9405 | 0.0885     | 0.0820 | 0.1075 | 0.0200     | 0.0655 | 0.2290 |
|                         | 1.6   | 0.6215      | 0.8445 | 0.9685 | 0.0775     | 0.0915 | 0.1110 | 0.0160     | 0.0805 | 0.2415 |
|                         | 2.0   | 0.6565      | 0.8695 | 0.9730 | 0.0830     | 0.0960 | 0.1220 | 0.0260     | 0.0940 | 0.2615 |
| $X \sim N(0, \Sigma_2)$ | 0     | 0.0435      | 0.0435 | 0.0470 | 0.0640     | 0.0695 | 0.0760 | 0.0240     | 0.0380 | 0.0470 |
|                         | 0.4   | 0.1145      | 0.3035 | 0.6545 | 0.0745     | 0.0860 | 0.1005 | 0.0620     | 0.1255 | 0.5870 |
| $W \sim N(0, \Sigma_2)$ | 0.8   | 0.3735      | 0.6780 | 0.8755 | 0.0820     | 0.1065 | 0.1230 | 0.0845     | 0.2380 | 0.7190 |
|                         | 1.2   | 0.5570      | 0.7850 | 0.8830 | 0.0780     | 0.1000 | 0.1420 | 0.1070     | 0.2565 | 0.7660 |
|                         | 1.6   | 0.6445      | 0.8130 | 0.9070 | 0.0905     | 0.1090 | 0.1545 | 0.1150     | 0.3070 | 0.7610 |
|                         | 2.0   | 0.6745      | 0.8455 | 0.9300 | 0.0890     | 0.1265 | 0.1620 | 0.1115     | 0.3330 | 0.7900 |
| $X \sim N(0, \Sigma_3)$ | 0     | 0.0475      | 0.0490 | 0.0525 | 0.0360     | 0.0820 | 0.0705 | 0.0130     | 0.0230 | 0.0405 |
|                         | 0.4   | 0.0970      | 0.2465 | 0.6405 | 0.0590     | 0.0845 | 0.0920 | 0.0370     | 0.1140 | 0.4645 |
| $W \sim N(0, \Sigma_3)$ | 0.8   | 0.3360      | 0.7120 | 0.9120 | 0.0615     | 0.0850 | 0.1045 | 0.0580     | 0.1835 | 0.5715 |
|                         | 1.2   | 0.5600      | 0.7995 | 0.9265 | 0.0520     | 0.0925 | 0.1180 | 0.0690     | 0.2000 | 0.6370 |
|                         | 1.6   | 0.6625      | 0.8405 | 0.9370 | 0.0680     | 0.0990 | 0.1260 | 0.0700     | 0.2020 | 0.6125 |
|                         | 2.0   | 0.6760      | 0.8605 | 0.9500 | 0.0715     | 0.0950 | 0.1215 | 0.0730     | 0.2225 | 0.6240 |

The results are presented in Tables 3 and 4. Comparing Table 2 with Table 3-4, we find that when the dimensions of  $X$  and  $W$  grow up to  $p_1 = p_2 = 6$ , the empirical powers of  $T_n^{FL}$  and  $T_n^{DM}$  are close to 0. This result means that both the competitors completely fail to detect the alternative hypothesis. We can also find that the empirical power of DREAM is similar to that when  $p_1 = p_2 = 4$  in Table 2. This again suggests that the dimensions of  $X$  and  $W$  have much less impact for  $T_n^{DEE}$  than they do for  $T_n^{FL}$  and  $T_n^{DM}$ .

The next example is to confirm that DREAM is still omnibus rather than



TABLE 4  
Empirical sizes and powers of  $T_n^{DEE}$ ,  $T_n^{FL}$  and  $T_n^{DM}$  with  $p_1 = 6$  and  $p_2 = 6$  in Example 2.

|                         | $a/n$ | $T_n^{DEE}$ |        |        | $T_n^{FL}$ |        |        | $T_n^{DM}$ |        |        |
|-------------------------|-------|-------------|--------|--------|------------|--------|--------|------------|--------|--------|
|                         |       | 50          | 100    | 200    | 50         | 100    | 200    | 50         | 100    | 200    |
| $X \sim N(0, \Sigma_1)$ | 0     | 0.0405      | 0.0455 | 0.0465 | 0.0005     | 0.0005 | 0.0055 | 0          | 0      | 0      |
|                         | 0.4   | 0.0720      | 0.1860 | 0.5515 | 0.0005     | 0.0025 | 0.0035 | 0          | 0      | 0      |
| $W \sim N(0, \Sigma_1)$ | 0.8   | 0.2445      | 0.6240 | 0.9045 | 0          | 0.0010 | 0.0030 | 0          | 0      | 0      |
|                         | 1.2   | 0.4575      | 0.7925 | 0.9190 | 0          | 0.0015 | 0.0035 | 0          | 0      | 0      |
|                         | 1.6   | 0.5595      | 0.8260 | 0.9240 | 0          | 0.0005 | 0.0035 | 0          | 0      | 0      |
|                         | 2.0   | 0.6255      | 0.8455 | 0.9445 | 0          | 0.0015 | 0.0040 | 0          | 0      | 0      |
| $X \sim N(0, \Sigma_2)$ | 0     | 0.0425      | 0.0470 | 0.0485 | 0.0035     | 0.0135 | 0.0215 | 0.0020     | 0.0110 | 0.0195 |
|                         | 0.4   | 0.1335      | 0.2490 | 0.5550 | 0.0020     | 0.0080 | 0.0270 | 0.0020     | 0.0250 | 0.1040 |
| $W \sim N(0, \Sigma_2)$ | 0.8   | 0.3160      | 0.6215 | 0.8120 | 0.0020     | 0.0185 | 0.0275 | 0.0030     | 0.0270 | 0.1550 |
|                         | 1.2   | 0.5025      | 0.7645 | 0.8580 | 0.0030     | 0.0100 | 0.0250 | 0.0030     | 0.0290 | 0.1695 |
|                         | 1.6   | 0.5770      | 0.7950 | 0.8850 | 0.0050     | 0.0105 | 0.0290 | 0.0020     | 0.0280 | 0.1725 |
|                         | 2.0   | 0.6375      | 0.8290 | 0.9105 | 0.0070     | 0.0140 | 0.0205 | 0.0020     | 0.0320 | 0.1685 |
| $X \sim N(0, \Sigma_3)$ | 0     | 0.0455      | 0.0460 | 0.0475 | 0          | 0.0025 | 0.0100 | 0          | 0.0035 | 0.0115 |
|                         | 0.4   | 0.0985      | 0.2430 | 0.5830 | 0.0005     | 0.0035 | 0.0090 | 0          | 0.0040 | 0.0520 |
| $W \sim N(0, \Sigma_3)$ | 0.8   | 0.3225      | 0.6515 | 0.8750 | 0.0030     | 0.0020 | 0.0095 | 0.0010     | 0.0075 | 0.0605 |
|                         | 1.2   | 0.4915      | 0.7695 | 0.8940 | 0.0015     | 0.0015 | 0.0080 | 0.0040     | 0.0080 | 0.0780 |
|                         | 1.6   | 0.5880      | 0.8030 | 0.9135 | 0.0010     | 0.0005 | 0.0060 | 0.0020     | 0.0120 | 0.0720 |
|                         | 2.0   | 0.6265      | 0.8380 | 0.9395 | 0.0010     | 0.0035 | 0.0085 | 0.0050     | 0.0135 | 0.0895 |

directional even when DREAM fully uses the information under the null hypothesis.

**Example 3.** The data are generated from the following model:

$$\bullet Y = 2 \sin(\beta_1^\top X) + \exp(\beta_2^\top X/2) + 2a \times \sin(\beta_2^\top W) + 0.5 \times \epsilon,$$

where  $p_1 = 4$ ,  $p_2 = 4$ ,  $\beta_1 = (1, 1, 0, 0)^\top / \sqrt{2}$  and  $\beta_2 = (0, 0, 1, 1)^\top / \sqrt{2}$ . Thus,  $q_1 = 2$ , and  $q_2 = 3$ . In this model, the conditional expectation  $E(Y|B_1^\top X)$  is the same under the null and alternative hypotheses. If we simply use this function to define a test, the alternative hypothesis cannot be detected at all from the theoretical point of view. However, DREAM can automatically adapt to the alternative model with the matrix  $B^\top Z$ , thus it still works under the alternative hypothesis.

By the comparison between Tables 2, 3 and 5, we observe that the power performances of DREAM in Example 3 are similar to those in Examples 1 and 2. This means that the test can have the advantage of dimension reduction and is still omnibus.

**Example 4.** The model is adopted from Lavergne et al. (2015)[11] under different combinations of dimensions of the covariate vector  $W$  and the distribution of the error  $\epsilon$  as:

$$\bullet Y = (\beta_1^\top X)^3 + \beta_1^\top X + a * (\beta_2^\top W - 1)^3 + \epsilon,$$

where  $\beta_1 = (1, 1)^\top$  and  $\beta_2 = \underbrace{(1, \dots, 1)}_{p_2}^\top / \sqrt{p_2}$  with two cases of dimensions:

$p_1 = p_2 = 2$  and  $p_1 = 2, p_2 = 4$ .  $X \sim N(0, \Sigma_1)$  and  $X \sim N(0, \Sigma_1)$ . Again the value  $a = 0$  responds to the hypothetical model, otherwise to the alternative

TABLE 5  
Empirical sizes and powers of  $T_n^{DEE}$ ,  $T_n^{FL}$  and  $T_n^{DM}$  with  $p_1 = 4$  and  $p_2 = 4$  in Example 3.

|                         | $a/n$ | $T_n^{DEE}$ |        |        | $T_n^{FL}$ |        |        | $T_n^{DM}$ |        |        |
|-------------------------|-------|-------------|--------|--------|------------|--------|--------|------------|--------|--------|
|                         |       | 50          | 100    | 200    | 50         | 100    | 200    | 50         | 100    | 200    |
| $X \sim N(0, \Sigma_1)$ | 0     | 0.0420      | 0.0530 | 0.0525 | 0.0575     | 0.0770 | 0.0760 | 0.0015     | 0.0080 | 0.0190 |
|                         | 0.4   | 0.0655      | 0.0985 | 0.3755 | 0.0510     | 0.0740 | 0.0765 | 0.0015     | 0.0280 | 0.1495 |
| $W \sim N(0, \Sigma_1)$ | 0.8   | 0.1470      | 0.5635 | 0.8930 | 0.0600     | 0.0845 | 0.1070 | 0.0040     | 0.0350 | 0.2085 |
|                         | 1.2   | 0.3690      | 0.8045 | 0.9390 | 0.0510     | 0.0840 | 0.0980 | 0.0035     | 0.0420 | 0.2525 |
|                         | 1.6   | 0.5730      | 0.8765 | 0.9485 | 0.0555     | 0.0840 | 0.1080 | 0.0090     | 0.0330 | 0.2510 |
|                         | 2.0   | 0.6480      | 0.8960 | 0.9645 | 0.0565     | 0.1000 | 0.1225 | 0.0030     | 0.0280 | 0.2595 |
| $X \sim N(0, \Sigma_2)$ | 0     | 0.0510      | 0.0545 | 0.0510 | 0.0720     | 0.0745 | 0.0810 | 0.0180     | 0.0440 | 0.0580 |
|                         | 0.4   | 0.0755      | 0.1165 | 0.2820 | 0.0685     | 0.0915 | 0.0945 | 0.0715     | 0.2460 | 0.3295 |
| $W \sim N(0, \Sigma_2)$ | 0.8   | 0.2075      | 0.5265 | 0.8490 | 0.0800     | 0.1175 | 0.1315 | 0.0925     | 0.3525 | 0.7430 |
|                         | 1.2   | 0.4245      | 0.7610 | 0.9365 | 0.0860     | 0.1040 | 0.1575 | 0.0950     | 0.3740 | 0.7615 |
|                         | 1.6   | 0.5595      | 0.8600 | 0.9480 | 0.0810     | 0.1095 | 0.1470 | 0.1145     | 0.3885 | 0.7630 |
|                         | 2.0   | 0.6315      | 0.8730 | 0.9560 | 0.0885     | 0.1205 | 0.1610 | 0.1140     | 0.3910 | 0.8185 |
| $X \sim N(0, \Sigma_3)$ | 0     | 0.0513      | 0.0550 | 0.0495 | 0.0515     | 0.0405 | 0.0790 | 0.0110     | 0.0220 | 0.0340 |
|                         | 0.4   | 0.0750      | 0.1255 | 0.3135 | 0.0625     | 0.0720 | 0.0915 | 0.0200     | 0.1320 | 0.4925 |
| $W \sim N(0, \Sigma_3)$ | 0.8   | 0.1945      | 0.5490 | 0.8840 | 0.0625     | 0.0875 | 0.1060 | 0.0340     | 0.1870 | 0.6035 |
|                         | 1.2   | 0.4330      | 0.7795 | 0.9300 | 0.0655     | 0.0945 | 0.1270 | 0.0230     | 0.1990 | 0.6100 |
|                         | 1.6   | 0.5955      | 0.8370 | 0.9460 | 0.0580     | 0.0960 | 0.1225 | 0.0470     | 0.1960 | 0.6035 |
|                         | 2.0   | 0.6405      | 0.8610 | 0.9535 | 0.0595     | 0.0990 | 0.1290 | 0.0460     | 0.2200 | 0.6280 |

models. For this model, we consider three different cases about the distributions of the error  $\epsilon$  as

- Case 1.  $\epsilon \sim N(0, 1)$ ;
- Case 2.  $\epsilon \sim N(0, 4)$ ;
- Case 3.  $\epsilon \sim t(4)$ ,

where  $t(4)$  represents Student's t-distribution with degrees 4 of freedom.

We have the following observations from the results of Table 6. First,  $T_n^{DEE}$  works better than its competitor. The comparison between  $T_n^{DEE}$  and  $T_n^{FL}$  further substantiates the theoretical advantage that our test has the faster convergence rate than that of Fan and Li's (1996)[6] test. Second, in the limited simulations, the heavy tail (Cases 3) does not have significant impact for the performance of all the three tests. Third, when the dimension  $p_2$  becomes larger, the performance of  $T_n^{DEE}$  is affected slightly, whereas  $T_n^{FL}$  and  $T_n^{DM}$  is significantly negatively affected. Therefore, the dimension is a key factor for power deterioration of  $T_n^{FL}$  and  $T_n^{DM}$ , while it is not for the proposed test  $T_n^{DEE}$ .

The following example considers higher dimensional  $B^T Z$  in a model.

**Example 5.** The data are generated from the following model:

- $Y = X_1 + 0.2 \exp(X_2) + a \times \left( \frac{1.5(W_1+W_2)}{0.5+(1.5W_3+0.5)^{1.5}} + 0.75 \sin(W_4 + 1) \right) + 0.2 \times \epsilon$ ,

where  $p_1 = 4$ ,  $p_2 = 4$ ,  $X \sim N(0, 4\Sigma_1)$  and  $W \sim N(0, 4\Sigma_1)$  and  $\epsilon \sim N(0, 1)$ . Thus,  $B_1$  is a  $4 \times 2$  matrix with  $\beta_1 = (1, 0, 0, 0)^T$  and  $\beta_2 = (0, 1, 0, 0)^T$  and  $B$  is a  $8 \times 5$  matrix with low-right block in which the columns are  $b_1 = (1, 1, 0, 0)^T$ ,  $b_2 = (0, 0, 1, 0)^T$ , and  $b_3 = (0, 0, 0, 1)^T$ . The results are summarized in Table 7.

Comparing them with the results in Examples 1–4, we can see that  $T_n^{DEE}$  can maintains the significance level well and its power reasonably becomes lower,

TABLE 6  
 Empirical sizes and powers of  $T_n^{DEE}$ ,  $T_n^{FL}$  and  $T_n^{DM}$  with  $X \sim N(0, \Sigma_1)$  and  $W \sim N(0, \Sigma_1)$  in Example 4.

|                         | $a/n$     | $T_n^{DEE}$ |        |        | $T_n^{FL}$ |        |        | $T_n^{DM}$ |        |        |        |
|-------------------------|-----------|-------------|--------|--------|------------|--------|--------|------------|--------|--------|--------|
|                         |           | 50          | 100    | 200    | 50         | 100    | 200    | 50         | 100    | 200    |        |
| $\epsilon \sim N(0, 1)$ | 0         | 0.0380      | 0.0415 | 0.0535 | 0.0405     | 0.0380 | 0.0520 | 0.0815     | 0.0735 | 0.0420 |        |
|                         | 0.3       | 0.1935      | 0.2830 | 0.5115 | 0.1595     | 0.3010 | 0.6025 | 0.1495     | 0.3585 | 0.5965 |        |
|                         | $p_1 = 2$ | 0.6         | 0.3830 | 0.6215 | 0.8675     | 0.2315 | 0.5395 | 0.8850     | 0.1595 | 0.4230 | 0.7040 |
|                         |           | 0.9         | 0.4970 | 0.7605 | 0.9095     | 0.2895 | 0.6080 | 0.9410     | 0.1815 | 0.4210 | 0.7610 |
|                         | $p_2 = 2$ | 1.2         | 0.5405 | 0.8010 | 0.9385     | 0.3140 | 0.6490 | 0.9505     | 0.1820 | 0.4350 | 0.7710 |
| 1.5                     |           | 0.5920      | 0.8425 | 0.9450 | 0.3250     | 0.6615 | 0.9565 | 0.1940     | 0.4495 | 0.7845 |        |
| $\epsilon \sim N(0, 4)$ | 0         | 0.0545      | 0.0575 | 0.0495 | 0.0615     | 0.0575 | 0.0540 | 0.0795     | 0.0690 | 0.0570 |        |
|                         | 0.3       | 0.1305      | 0.1875 | 0.3205 | 0.1600     | 0.2600 | 0.4500 | 0.1220     | 0.2760 | 0.4375 |        |
|                         | $p_1 = 2$ | 0.6         | 0.3575 | 0.5385 | 0.8240     | 0.2590 | 0.5600 | 0.8265     | 0.1655 | 0.3715 | 0.6800 |
|                         |           | 0.9         | 0.5065 | 0.7145 | 0.9045     | 0.3355 | 0.6915 | 0.9260     | 0.1650 | 0.4055 | 0.7265 |
|                         | $p_2 = 2$ | 1.2         | 0.6010 | 0.8245 | 0.9330     | 0.3785 | 0.7670 | 0.9455     | 0.1815 | 0.4090 | 0.7560 |
| 1.5                     |           | 0.6550      | 0.8670 | 0.9485 | 0.4005     | 0.7785 | 0.9545 | 0.1975     | 0.4190 | 0.7735 |        |
| $\epsilon \sim t(4)$    | 0         | 0.0385      | 0.0560 | 0.0565 | 0.0715     | 0.0675 | 0.0470 | 0.0890     | 0.0725 | 0.0400 |        |
|                         | 0.3       | 0.1570      | 0.2350 | 0.4140 | 0.2230     | 0.4240 | 0.7545 | 0.1310     | 0.2645 | 0.5190 |        |
|                         | $p_1 = 2$ | 0.6         | 0.3905 | 0.6025 | 0.8180     | 0.3475 | 0.6815 | 0.9625     | 0.1485 | 0.3690 | 0.6735 |
|                         |           | 0.9         | 0.5025 | 0.7525 | 0.9160     | 0.3875 | 0.7735 | 0.9850     | 0.1760 | 0.4035 | 0.7535 |
|                         | $p_2 = 2$ | 1.2         | 0.5870 | 0.8330 | 0.9445     | 0.4170 | 0.7820 | 0.9855     | 0.1940 | 0.4080 | 0.7680 |
| 1.5                     |           | 0.6400      | 0.8390 | 0.9475 | 0.4410     | 0.7965 | 0.9895 | 0.1895     | 0.4165 | 0.7935 |        |
| $\epsilon \sim N(0, 1)$ | 0         | 0.0570      | 0.0555 | 0.0415 | 0.0900     | 0.0885 | 0.0850 | 0.0415     | 0.0630 | 0.0330 |        |
|                         | 0.3       | 0.1995      | 0.2730 | 0.4905 | 0.1320     | 0.2270 | 0.4240 | 0.0255     | 0.0290 | 0.0735 |        |
|                         | $p_1 = 2$ | 0.6         | 0.4180 | 0.6390 | 0.8315     | 0.1525 | 0.2910 | 0.5525     | 0.0200 | 0.0315 | 0.0920 |
|                         |           | 0.9         | 0.4755 | 0.7525 | 0.9185     | 0.1645 | 0.3165 | 0.5825     | 0.0155 | 0.0300 | 0.0950 |
|                         | $p_2 = 4$ | 1.2         | 0.5945 | 0.8220 | 0.9390     | 0.1665 | 0.3230 | 0.6055     | 0.0105 | 0.0360 | 0.1010 |
| 1.5                     |           | 0.6255      | 0.8455 | 0.9390 | 0.1815     | 0.3135 | 0.6195 | 0.0135     | 0.0235 | 0.0965 |        |
| $\epsilon \sim N(0, 4)$ | 0         | 0.0405      | 0.0570 | 0.0455 | 0.0860     | 0.0940 | 0.0875 | 0.0400     | 0.0590 | 0.0280 |        |
|                         | 0.3       | 0.1195      | 0.1775 | 0.2755 | 0.1270     | 0.1665 | 0.2445 | 0.0280     | 0.0475 | 0.0815 |        |
|                         | $p_1 = 2$ | 0.6         | 0.3550 | 0.5395 | 0.8015     | 0.1430 | 0.2275 | 0.4145     | 0.0265 | 0.0385 | 0.0920 |
|                         |           | 0.9         | 0.5095 | 0.7315 | 0.8985     | 0.1465 | 0.2680 | 0.5045     | 0.0190 | 0.0345 | 0.0920 |
|                         | $p_1 = 4$ | 1.2         | 0.6045 | 0.8220 | 0.9300     | 0.1580 | 0.2810 | 0.5665     | 0.0165 | 0.0285 | 0.1040 |
| 1.5                     |           | 0.6360      | 0.8490 | 0.9465 | 0.1695     | 0.3075 | 0.5770 | 0.0155     | 0.0450 | 0.0965 |        |
| $\epsilon \sim t(4)$    | 0         | 0.0410      | 0.0510 | 0.0565 | 0.0975     | 0.0955 | 0.0890 | 0.0335     | 0.0655 | 0.0420 |        |
|                         | 0.3       | 0.1680      | 0.2540 | 0.4180 | 0.1250     | 0.1920 | 0.3320 | 0.0260     | 0.0290 | 0.0700 |        |
|                         | $p_1 = 2$ | 0.6         | 0.3805 | 0.6050 | 0.8410     | 0.1520 | 0.2680 | 0.5125     | 0.0195 | 0.0305 | 0.0850 |
|                         |           | 0.9         | 0.4995 | 0.7525 | 0.9100     | 0.1625 | 0.2940 | 0.5675     | 0.0155 | 0.0375 | 0.0900 |
|                         | $p_2 = 4$ | 1.2         | 0.5935 | 0.8275 | 0.9305     | 0.1635 | 0.2980 | 0.5920     | 0.0115 | 0.0280 | 0.0870 |
| 1.5                     |           | 0.6205      | 0.8435 | 0.9380 | 0.1700     | 0.2920 | 0.6100 | 0.0125     | 0.0245 | 0.1005 |        |

TABLE 7  
 Empirical sizes and powers of  $T_n^{DEE}$ ,  $T_n^{FL}$  and  $T_n^{DM}$  with the dimensions  $p_1 = p_2 = 4$  and structure dimensions  $q_1 = 2$  and  $q_2 = 5$  in Example 5.

| $n$       | $a$ | $T_n^{DEE}$ | $T_n^{FL}$ | $T_n^{DM}$ |
|-----------|-----|-------------|------------|------------|
| $n = 200$ | 0   | 0.0555      | 0.0600     | 0.0290     |
|           | 1   | 0.3315      | 0.1455     | 0.1870     |
| $n = 400$ | 0   | 0.0535      | 0.0645     | 0.0465     |
|           | 1   | 0.4910      | 0.1710     | 0.3205     |

but is still higher than those of  $T_n^{FL}$  and  $T_n^{DM}$ . Based on our very limited numerical studies, which are not reported, we find that it is difficult to estimate the structural dimension  $q = 5$ , but our test can still detect the alternative models with positive probability.

In summary, the above simulations sustain the aforementioned theoretical properties that the proposed test is significantly superior to existing tests among which Fan and Li's (1996)[6] test and Delgado and González Manteiga's (2001)[4] test are regarded as representatives of existing tests.

#### 4.2. Baseball hitters' salary data

We now analyze the well-known Baseball hitters' salary data set, which was originally published for the 1988 ASA Statistical Graphics and Computing Data Exposition and is available at <http://euclid.psych.yorku.ca/ftp/sas/sssg/data/baseball.sas>. The data set consists of information on salary and 16 performance measures of 263 major league baseball hitters. As always, the question of main interest is whether salary reflects performance. As displayed by Friendly (2002)[7], the 16 measures naturally belong to three performance categories: the season hitting statistics, which include the numbers of times at bat ( $X_1$ ), hits ( $X_2$ ), home runs ( $X_3$ ), runs ( $X_4$ ), runs batted in ( $X_5$ ), and walks ( $X_6$ ) in 1986; the career hitting statistics, which include the numbers of years in the major leagues ( $X_7$ ), times at bat ( $X_8$ ), hits ( $X_9$ ), home runs ( $X_{10}$ ), runs ( $X_{11}$ ), runs batted in ( $X_{12}$ ) and walks ( $X_{13}$ ) during the players' entire career up to 1986; and the fielding variables, which include the numbers of putouts ( $X_{14}$ ), assists ( $X_{15}$ ) and errors ( $X_{16}$ ) in 1986.

Further, the covariates from different groups have weak correlations. The logarithm of annual salary in 1987 is used to be the response variable ( $Y$ ) and the new covariates from the career totals by dividing totals by years in the major leagues are constructed. Let  $X_j^* = X_j/X_7$  for  $j = 8, \dots, 13$ . As remarked by Hoaglin and Velleman (1995)[9], the analyses working with  $\ln(\text{salary})$  and with the annual rate covariates fared better than those worked with the raw forms of these covariates. Below, we use  $X_j^*$ 's instead. All the covariates are standardized to have mean zero and unit length. We write  $V_1 = (X_1, \dots, X_6)$ ,  $V_2 = (X_7, X_8^*, \dots, X_{13}^*)$  and  $V_3 = (X_{14}, X_{15}, X_{16})$ . In this application, we consider two cases:

- Case (I):  $X = (V_1, V_2)$  and  $W = V_3$ ;  
 Case (II):  $X = V_1$  and  $W = (V_2, V_3)$ ;

Under the two cases, the values of the test statistics are respectively 1.1171 and 6.5831 and the corresponding  $p$ -values are 0.1320 and 0.0000.

From these results under Cases (I) and (II), we can conclude that the career hitting statistic of the group  $V_2$  has positive impact for the annual salary. The results are consistent with those advised by Xia et al. (2002)[22] who found that the variables  $X_7$ ,  $X_9$  and  $X_{13}$  in the group  $V_2$  are prominently to affect the annual salary. The coefficients of the fielding covariates in the group  $V_3$  are closed to 0 in the estimated directions suggested by Xia et al. (2002)[22]. Therefore, for the annual salary, the group  $V_2$  contains the significance covariates while the group  $V_3$  does not.

## 5. Conclusions

In this paper, we develop a dimension reduction adaptive-to-model test to determine significant covariates under the nonparametric regression framework. The approach employs a dimension reduction technique to reduce the dimension such that the constructed test can well maintain the significance level and be more powerful than existing tests in the literature. This methodology can be applicable to check other semi-parametric regression models, for example partially linear models, single-index models and partially linear single-index models. The research is on-going. Further, as the test involves nonparametric estimation under the null hypothesis, when the dimension of  $X$  (or  $B_1^\top X$ ) is high, none of tests could work well. Thus, it deserves a further study. Another issue is about the estimation inconsistency for the structural dimension  $q_2$  under the local alternative hypotheses, which leads to the destroy of the omnibus property. As the problem is not caused by the model adaptation idea, while is caused by the estimation accuracy, defining a good estimate of  $q$  under the local alternative hypotheses is a very interesting topic in a further study.

## Appendix

### A.1. Regularity conditions

To prove the asymptotic properties in Sections 2 and 3, we provide the following regularity conditions:

A1  $M_n(t)$  has the following expansion:

$$M_n(t) = M(t) + E_n\{\psi(Z, Y, t)\} + R_n(t),$$

where  $E_n(\cdot)$  denotes the average over all sample points,  $E\{\psi(Z, Y, t)\} = 0$  and  $E\{\psi^2(Z, Y, t)\} < \infty$ .

A2  $\sup_t \|R_n(t)\|_F = o_p(n^{-1/2})$ , where  $\|\cdot\|_F$  denotes the Frobenius norm of a matrix.

A3 The estimate  $\tilde{M}_n(t)$  has the following expansion:

$$\tilde{M}_n(t) = \tilde{M}(t) + E_n\{\tilde{\psi}(X, Y, t)\} + \tilde{R}_n(t),$$

where  $\tilde{M}_n(y_i)$  is an estimate of the  $p_1 \times p_1$  positive semi-definite matrix  $\tilde{M}(t)$  satisfying  $\text{Span}\{\tilde{M}(t)\} = S_{Y(t)|X}$ ,  $E\{\tilde{\psi}(X, Y, t)\} = 0$ ,  $E\{\tilde{\psi}^2(X, Y, t)\} < \infty$  and  $\sup_t \|\tilde{R}_n(t)\|_F = o_p(n^{-1/2})$ . Corresponding,  $\tilde{M}_n = \frac{1}{n} \sum_{i=1}^n \tilde{M}_n(y_i)$  is an estimate of the target matrix  $\tilde{M}$  satisfying  $\text{Span}\{\tilde{M}\} = S_{Y|E(Y|X)}$ .

A4  $(B^\top z_i, y_i)_{i=1}^n$  is from the probability distribution  $F(B^\top z, y)$  on  $\mathbb{R}^q \times \mathbb{R}$ . The error  $\epsilon = Y - m(B^\top Z)$  satisfies that  $E(\epsilon^8 | B^\top Z = B^\top z)$  is continuous and  $E(\epsilon^8 | B^\top Z = B^\top z) \leq b(B^\top z)$  almost surely, where  $b(B^\top z)$  is a measurable function such that  $E\{b^2(B^\top Z)\} < \infty$ .

A5 The density function  $p_{B_1}(\cdot)$  of  $B_1^\top X$  exists with support  $\mathbb{C}$  and has a continuous and bounded first-order derivative on the support  $\mathbb{C}$ . The density  $p_{B_1}(\cdot)$  satisfies

$$0 < \inf_{B_1^\top X \in \mathbb{C}} p_{B_1}(B_1^\top X) \leq \sup_{B_1^\top X \in \mathbb{C}} p_{B_1}(B_1^\top X) < \infty.$$

A6 The function  $g(\cdot)$  is  $\eta$ -order partially differentiable for some positive integer  $\eta$ , and the  $\eta$ th partially derivative of  $g(\cdot)$  is bounded.

A7  $\tilde{Q}(\cdot)$  is a symmetric and twice order continuously differentiable kernel function satisfying

$$\int u^i \tilde{Q}(u) du = \delta_{i0}, \quad \text{for } i = 0, 1, \dots, \eta - 1,$$

$$\tilde{Q}(u) = O\{(1 + |u|^{\eta+1+\beta})^{-1}\}, \quad \text{some } \beta > 0,$$

where  $\delta_{ij}$  is the Kronecker's delta and  $\eta$  is given in Condition A6.

A8  $K(\cdot)$  is a bounded, symmetric kernel function and it is a first order continuously differentiable kernel function satisfying  $\int K(u) du = 1$ .

A9  $n \rightarrow \infty$ ,  $h_1 \rightarrow 0$ ,  $h \rightarrow 0$ ,

- 1) under the null or local alternative hypotheses with  $C_n = n^{-1/2} h^{-q_1/4}$ ,  $nh_1^{q_1} \rightarrow \infty$ ,  $nh^{q_1} \rightarrow \infty$  and  $nh_1^{2\eta} h^{q_1/2} \rightarrow 0$ ;
- 2) under the global alternative hypothesis,  $nh_1^{q_1} \rightarrow \infty$ ,  $nh^{q_2} \rightarrow \infty$  and  $nh_1^{2\eta} h^{q_2/2} \rightarrow 0$ ,

where  $\eta$  is given in Condition A6.

**Remark A.1.** Conditions A1, A2 and A3 are necessary for DEE to estimate the matrixes  $B$  and  $B_1$ . Under the linearity condition and constant conditional variance condition,  $DEE_{SIR}$  satisfies the Conditions A1, A2 and A3. Conditions A4, A5, A6 and A7 are widely used for nonparametric estimation in the literature. It is worth pointing out that Condition A6 about the higher-order kernel plays an important roles in bias reduction, see Fan and Li (1996)[6]. Conditions A5 and A8 guarantee the asymptotic normality of DREAM statistic and make the test well-behaved. Condition A9 about the choice of bandwidth  $h$  is reasonable because the estimation  $\hat{q}$  is different under the null and alternative hypotheses.

### A.2. Proof of the theorems

*Proof of Proposition 2.1.* Note that under the null hypothesis,  $E(Y|Z) = g(B_1^\top X)$ . Then, we have

$$Y \perp\!\!\!\perp E(Y|Z) | B_1^\top X,$$

where the notation  $\perp\!\!\!\perp$  stands for independence. This is equivalent to

$$Y \perp\!\!\!\perp E(Y|Z) | \tilde{B}^\top Z,$$

where  $\tilde{B} = (B_1^\top, 0_{q_1 \times p_2})^\top$ . Recall, without notational confusion,  $q$  is a number that takes value  $q_1$  under the null and  $q_2$  under the alternatives. From the definition of central mean subspace,  $S_{E(Y|Z)}$  is the intersection of all the linear spaces spanned respectively by the columns of any  $d \times q$  orthogonal matrix  $\Gamma$  with  $1 \leq q \leq d$  such that the above conditional independence holds. Thus, under the null,  $S_{E(Y|Z)} \subseteq \text{Span}(\tilde{B})$  where  $\text{Span}(\tilde{B})$  is the linear space spanned by the columns of  $\tilde{B}$ . Let  $\{\gamma_1, \dots, \gamma_{q_1}\}$  the eigenvectors associated with the nonzero eigenvalue eigenvalues of  $M$ . As Zhu et al. (2010a)[24] argued that  $\{\gamma_1, \dots, \gamma_{q_1}\} \in S_{E(Y|Z)}$ , we have  $\{\gamma_1, \dots, \gamma_{q_1}\} \in \text{Span}(\tilde{B})$ . This implies that  $\gamma_j$  for  $j = 1, \dots, q_1$  can be denoted as a linear combination of the columns of  $\tilde{B}$ . Thus, for  $j = 1, \dots, q_1$ ,  $\gamma_j$  has the similar form as  $\gamma_j = (\tilde{\gamma}_j^\top, 0_{1 \times p_2})^\top$  with  $\tilde{\gamma}_j$  being a  $p_1 \times 1$  vector. This implies that any element in  $S_{E(Y|Z)}$  can also be written as  $\gamma_j = (\tilde{\gamma}_j^\top, 0_{1 \times p_2})^\top$ . Further, the structural dimension of  $S_{E(Y|Z)}$  is smaller than or equal to  $q_1$ . Further, we note that under  $H_0$ ,  $E(Y|Z) = E(Y|B^\top Z) = E(Y|\tilde{B}^\top Z)$  and  $E(Y|Z) = E(Y|X)$ . Thus,  $q = q_1 = \dim\{S_{E(Y|X)}\}$ .

Under Conditions A1 and A2, Theorem 2 in Zhu et al. (2010a)[24] shows that  $M_n - M = O_p(n^{-1/2})$ . From the arguments in Zhu and Fang (1996)[26] and Zhu and Ng (1995)[28], under some regularity conditions,  $\hat{\lambda}_i - \lambda_i = O_p(n^{-1/2})$ , where  $\hat{\lambda}_d \leq \hat{\lambda}_{d-1} \leq \dots \leq \hat{\lambda}_1$  are the eigenvalues of the matrix  $M_n$  and  $\lambda_i$  are the eigenvalues of the matrix  $M$ . The estimate  $\tilde{B}$  that consists of the eigenvectors associated with the largest  $q_1$  eigenvalues of  $M_n$  is consistent for  $(C_1 B_1^\top, 0_{q_1 \times p_2})^\top$  for a  $q_1 \times q_1$  orthogonal matrix  $C_1$ .  $\square$

*Proof of Proposition 2.2.* From Proposition 2.1,  $\hat{\lambda}_i - \lambda_i = O_p(n^{-1/2})$ , where  $\hat{\lambda}_d \leq \hat{\lambda}_{d-1} \leq \dots \leq \hat{\lambda}_1$  are the eigenvalues of the matrix  $M_n$ .

Define

$$\tilde{\lambda}_i = \frac{\lambda_i}{\lambda_i + 1}, \text{ for } 1 \leq i \leq d, \tag{A.1}$$

where  $\lambda_d = \dots = \lambda_{d-q} = 0$  and  $0 < \lambda_q \leq \dots \leq \lambda_1$  are the eigenvalues of the target matrix  $M$ .

Recall the definition of  $\lambda_l^*$  (2.6) in Subsection 2.3. For any  $1 < l \leq q$ , since  $\lambda_l > 0$  and  $\hat{\lambda}_l^2 = \lambda_l^2 + O_p(1/\sqrt{n})$ , we have  $(\lambda_l^*)^2 = (\tilde{\lambda}_l)^2 + O_p(1/\sqrt{n})$ . On the other hand, for any  $q < l \leq d$ , as  $\lambda_l = 0$  and  $\hat{\lambda}_l^2 = \lambda_l^2 + O_p(1/n) = O_p(1/n)$ , and then  $(\lambda_l^*)^2 = O_p(1/n)$ .

For any  $l < q$ , because  $\tilde{\lambda}_l^2 > 0$  and  $\tilde{\lambda}_{l+1}^2 > 0$ , we have

$$\begin{aligned} \frac{(\lambda_{q+1}^*)^2 + c_n}{(\lambda_q^*)^2} - \frac{(\lambda_{l+1}^*)^2 + c_n}{(\lambda_l^*)^2} &= \frac{\tilde{\lambda}_{q+1}^2 + c_n + O_p(1/n)}{\tilde{\lambda}_q^2 + c_n + O_p(1/\sqrt{n})} - \frac{\tilde{\lambda}_{l+1}^2 + c_n + O_p(1/\sqrt{n})}{\tilde{\lambda}_l^2 + c_n + O_p(1/\sqrt{n})} \\ &= \frac{c_n + O_p(1/n)}{\tilde{\lambda}_q^2 + c_n + O_p(1/\sqrt{n})} - \frac{\tilde{\lambda}_{l+1}^2 + c_n + O_p(1/\sqrt{n})}{\tilde{\lambda}_l^2 + c_n + O_p(1/\sqrt{n})}. \end{aligned}$$

Since  $c \times \log n/n \leq c_n \rightarrow 0$  with some fixed  $c > 0$ , we get

$$\frac{(\lambda_{q+1}^*)^2 + c_n}{(\lambda_q^*)^2} - \frac{(\lambda_{l+1}^*)^2 + c_n}{(\lambda_l^*)^2} \rightarrow \frac{0}{\tilde{\lambda}_{q_1}^2} - \frac{\tilde{\lambda}_{(l+1)}^2}{\tilde{\lambda}_l^2} = -\frac{\tilde{\lambda}_{(l+1)}^2}{\tilde{\lambda}_l^2} < 0.$$

For any  $l > q$ ,  $\tilde{\lambda}_l = 0$  and  $\tilde{\lambda}_q^2 > 0$ , then we have

$$\begin{aligned} \frac{(\lambda_{q+1}^*)^2 + c_n}{(\lambda_q^*)^2} - \frac{(\lambda_{l+1}^*)^2 + c_n}{(\lambda_l^*)^2} &= \frac{\tilde{\lambda}_{q+1}^2 + c_n + O_p(1/n)}{\tilde{\lambda}_q^2 + c_n + O_p(1/\sqrt{n})} - \frac{\tilde{\lambda}_{l+1}^2 + c_n + O_p(1/n)}{\tilde{\lambda}_l^2 + c_n + O_p(1/n)} \\ &= \frac{c_n + o_p(c_n)}{\tilde{\lambda}_q^2 + c_n + O_p(1/\sqrt{n})} - \frac{c_n + o_p(c_n)}{c_n + o_p(c_n)} \\ &\rightarrow -1 < 0. \end{aligned}$$

Therefore, altogether, it is concluded that  $\hat{q} = q$  in probability. This implies that under  $H_0$ ,  $P(\hat{q} = q_1) = 1$  and under  $H_1$ ,  $P(\hat{q} = q_2) = 1$ .  $\square$

*Proof of Theorem 3.1.* Define the events  $A_n = \{T_n \geq c\}$  for any constant  $c$ , and  $B_n = \{\hat{q} = q_1\}$ . According to Proposition 2.2, under the null hypothesis,  $\hat{q} = q_1$  with a probability going to one, namely  $\lim_{n \rightarrow \infty} P(B_n) = 1$ , where  $P(A)$  denotes the probability that the event  $A$  happens. The facts that if  $P(A_n)$  converges,  $\lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} P(A_n \cap B_n)$ , implies that under the null hypothesis, in an asymptotic sense it is only needed to consider the case  $\hat{q} = q_1$  for the asymptotic properties of the test statistic  $T_n$  throughout the proof of Theorem 3.1.

For notational convenience, denote  $z_i = (x_i, w_i)$ ,  $g_i = g(B_1^\top x_i)$ ,  $\hat{g}_i = \hat{g}(\hat{B}_1^\top x_i)$ ,  $u_i = y_i - g_i$ ,  $\hat{u}_i = y_i - \hat{g}_i$ ,  $K_{Bij} = K(B^\top(z_i - z_j)/h)$ ,  $p_i = p_{B_1 i}$  and  $\hat{p}_i = \hat{p}_{\hat{B}_1 i}$ , where  $\hat{p}_{\hat{B}_1 i}$  is a kernel estimate of the density function  $p_{B_1}(\cdot)$  of  $B_1^\top x_i$  given by

$$\hat{p}_{\hat{B}_1 i} = \frac{1}{n-1} \sum_{j \neq i}^n Q_{h_1}(\hat{B}_1^\top x_j - \hat{B}_1^\top x_i).$$

Throughout this appendix,  $E_i(\cdot) = E(\cdot | z_i)$ . To have a uniform version in this proof, under the null hypothesis, define  $B = \tilde{B} = (B_1^\top, 0_{q_1 \times p_2})^\top$ .

Note that  $\hat{u}_i \equiv: y_i - \hat{g}_i = u_i - (\hat{g}_i - g_i)$ . We then decompose the term  $V_n$  as:

$$\begin{aligned} V_n &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{q_1}} K_{\hat{B}ij} u_i u_j + \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{q_1}} K_{\hat{B}ij} (\hat{g}_i - g_i)(\hat{g}_j - g_j) \\ &\quad - 2 \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{q_1}} K_{\hat{B}ij} u_i (\hat{g}_j - g_j) \\ &\equiv: Q_{1n} + Q_{2n} - 2Q_{3n}. \end{aligned} \tag{A.2}$$

First, consider the term  $Q_{1n}$ . By taking a Taylor expansion for  $Q_{1n}$  with respect to  $B$ , we have

$$Q_{1n} \equiv: Q_{11n} + Q_{12n},$$



where

$$Q_{11n} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{q_1}} K_{B_{ij}} u_i u_j, \tag{A.3}$$

$$Q_{12n} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{2q_1}} K'_{B^*_{ij}} u_i u_j (\hat{B} - B)^\top (z_i - z_j). \tag{A.4}$$

Here we assert  $B^* = \{B^*_{ij}\}_{d \times q_1}$  with  $B^*_{ij} \in [\min\{\hat{B}_{ij}, B_{ij}\}, \max\{\hat{B}_{ij}, B_{ij}\}]$ . Let  $V(t) = B + t(\hat{B} - B)$  for a value  $t \in (-\infty, \infty)$ . Define a function as

$$f(t) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{q_1}} K_{V(t)_{ij}} u_i u_j.$$

We have  $f(1) = Q_{1n}$  and  $f(0) = Q_{11n}$ . By an application of mean value theorem, we have  $f(1) - f(0) = f'(\tilde{t}) = Q_{12n}$  with  $\tilde{t} \in (0, 1)$ . Thus, we can conclude that  $B^* = B + \tilde{t}(\hat{B} - B)$  with  $\tilde{t} \in (0, 1)$ . Corresponding, we get  $B^*_{ij} = B_{ij} + \tilde{t}(\hat{B}_{ij} - B_{ij})$  with  $\tilde{t} \in (0, 1)$ . This affirms the assertion that  $B^* = \{B^*_{ij}\}_{d \times q_1}$  with  $B^*_{ij} \in [\min\{\hat{B}_{ij}, B_{ij}\}, \max\{\hat{B}_{ij}, B_{ij}\}]$ .

Because  $\|\hat{B} - B\| = O_p(1/\sqrt{n})$  and the first derivative of  $K_B(\cdot)$  with respect to  $B$  is a bounded continuity function of  $B$ , we conclude that replacing  $B^*$  by  $B$  does not affect the convergence rate of  $Q_{12n}$ .

In the present paper, as the dimension of  $B^\top Z$  is fixed, the term  $Q_{11n}$  is an  $U$ -statistic. Since under  $H_0$ ,  $q = q_1$ , following a similar argument as that for Lemma 3.3 in Zheng (1996)[23], it is easy to obtain:

$$nh^{q_1/2} Q_{11n} \Rightarrow N(0, s^2),$$

where

$$s^2 = 2 \int K_B^2(u) du \cdot \int \{\sigma^2(B^\top Z)\}^2 p^2(B^\top Z) dB^\top Z,$$

with  $\sigma^2(B^\top z) = E(u^2 | B^\top Z = B^\top z)$  and  $B = (B_1^\top, 0_{q_1 \times p_2})^\top$ . We then omit the details.

We turn to discuss the term  $Q_{12n}$  in (A.4). Since  $E(u_i | z_i) = 0$ , we have  $E(Q_{12n}) = 0$ . We then calculate the second order moment of  $Q_{12n}$  as follows:

$$\begin{aligned} E(Q_{12n}^2) &= E \left[ \left\{ \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{2q_1}} K'_{B_{ij}} u_i u_j (\hat{B} - B)^\top (z_i - z_j) \right\}^2 \right] \\ &= E \left\{ \frac{1}{n^2(n-1)^2} \frac{1}{h^{4q_1}} \sum_{i=1}^n \sum_{i' \neq j'}^n \sum_{i=1}^n \sum_{i' \neq j'}^n K'_{B_{ij}} K'_{B_{i'j'}} \right. \\ &\quad \left. u_i u_j u_{i'} u_{j'} (\hat{B} - B)^\top (z_i - z_j) (z_{i'} - z_{j'})^\top (\hat{B} - B) \right\}. \end{aligned}$$

Since  $E(u_i u_j u_{i'} u_{j'}) \neq 0$  if and only if  $i = i', j = j'$  or  $i = j', j = i'$ , we have

$$\begin{aligned} E(Q_{12n}^2) &= \frac{2n(n-1)}{n^2(n-1)^2} \frac{1}{h^{4q_1}} E\{(K'_{B12})^2 u_1^2 u_2^2 (\hat{B} - B)^\top (z_1 - z_2)(z_1 - z_2)^\top (\hat{B} - B)\} \\ &= \frac{2}{n(n-1)} \frac{1}{h^{4q_1}} E\{(K'_{B12})^2 u_1^2 u_2^2 (\hat{B} - B)^\top (z_1 - z_2)(z_1 - z_2)^\top (\hat{B} - B)\} \\ &= \frac{2}{n(n-1)} \frac{1}{h^{4q_1}} \{E(u_1^2)\}^2 (\hat{B} - B)^\top E\{(K'_{B12})^2 (z_1 - z_2)(z_1 - z_2)^\top\} (\hat{B} - B). \end{aligned}$$

By changing variables as  $v_1 = (z_1 - z_2)/h$ , a further computation yields

$$\begin{aligned} E(Q_{12n}^2) &= \frac{1}{n(n-1)} \frac{1}{h^{4q_1}} \{E(u_1^2)\}^2 \int \int (K'(u))^2 (\hat{B} - B)^\top (z_1 - z_2) \\ &\quad (z_1 - z_2)^\top (\hat{B} - B) p(B^\top z_1) p(B^\top z_2) dz_1 dz_2 \\ &= \frac{1}{n(n-1)} \frac{1}{h^{q_1}} \{E(u_1^2)\}^2 \int \int (K'(u))^2 (\hat{B} - B)^\top uu^\top (\hat{B} - B) \\ &\quad p(B^\top z_1) p(B^\top (z_1 - hu)) dz_1 du. \end{aligned}$$

By taking Taylor expansion of  $p(B^\top (z_1 - hu))$  around  $z_1$  and using Conditions A4, A7, A8 and A9, we have

$$\begin{aligned} E(Q_{12n}^2) &= \frac{1}{n(n-1)} \frac{1}{h^{q_1}} \{E(u_1^2)\}^2 \int \int \{K'(u)\}^2 (\hat{B} - B)^\top uu^\top (\hat{B} - B) \\ &\quad p(B^\top z_1) p(B^\top (z_1 - hu)) dz_1 du \\ &= \frac{1}{n(n-1)} \frac{1}{h^{q_1}} \{E(u_1^2)\}^2 \int \int \{K'(u)\}^2 (\hat{B} - B)^\top uu^\top (\hat{B} - B) \\ &\quad [p^2(B^\top z_1) + p(B^\top z_1) p'(B^\top z_1) h^p u] dz_1 du + o_p \left\{ \frac{1}{n(n-1)} \right\} \\ &= \frac{1}{n(n-1)} \frac{1}{h^{q_1}} \{E(u_1^2)\}^2 \int \int \{K'(u)\}^2 (\hat{B} - B)^\top uu^\top (\hat{B} - B) p^2(B^\top z_1) dz_1 du \\ &\quad + \frac{1}{n(n-1)} \frac{1}{h^{q_1}} \{E(u_1^2)\}^2 \int \int \{K'(u)\}^2 (\hat{B} - B)^\top uu^\top (\hat{B} - B) \\ &\quad p(B^\top z_1) p'(B^\top z_1) h^{q_1} B^\top u dz_1 du + o_p \left\{ \frac{1}{n(n-1)} \right\} O\left(\frac{1}{n}\right) \\ &= O_p \left\{ \frac{1}{n^2(n-1)h^{q_1}} \right\} = o_p\left(\frac{1}{n^2}\right). \end{aligned}$$

The application of Chebyshev's inequality leads to  $|Q_{12n}| = o_p(n^{-1}h^{-q_1/2})$ . Using the above results for the terms  $Q_{11n}$  and  $Q_{12n}$ , it is deduced that  $nh^{q_1/2}Q_{1n} \Rightarrow N(0, s^2)$ .

Now we consider the term  $Q_{2n}$  in (A.2). We can derive that

$$\begin{aligned} Q_{2n} &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{q_1}} K_{\hat{B}_q^{ij}}(\hat{g}_i - g_i)(\hat{g}_j - g_j) \frac{\hat{p}_i \hat{p}_j}{p_i p_j} \\ &\quad + \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{q_1}} K_{\hat{B}_q^{ij}}(\hat{g}_i - g_i)(\hat{g}_j - g_j) \\ &\quad \times \left\{ \frac{\hat{p}_i - p_i}{p_i} \frac{\hat{p}_j - p_j}{p_j} - 2 \frac{(\hat{p}_i - p_i) \hat{p}_j}{p_i p_j} \right\} \\ &\equiv: \tilde{Q}_{2n} + o_p(\tilde{Q}_{2n}). \end{aligned} \tag{A.5}$$

Substituting the kernel estimates  $\hat{g}$  and  $\hat{p}$  into  $\tilde{Q}_{2n}$ , we have

$$\begin{aligned} \tilde{Q}_{2n} &= \frac{1}{n(n-1)^3} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{l \neq i}^n \sum_{k \neq j}^n \frac{1}{h^{q_1} h_1^{2q_1}} \frac{1}{p_i p_j} K_{\hat{B}_1^{ij}} Q_{\hat{B}_1^{il}} Q_{\hat{B}_1^{jk}} \\ &\quad \times \{y_l - g(B_1^\top x_i)\} \{y_k - g(B_1^\top x_j)\}. \end{aligned}$$

By an application of Taylor expansion for  $\tilde{Q}_{2n}$  with respect to  $B$  and  $B_1$ , we can have

$$\tilde{Q}_{2n} \equiv Q_{21n} + Q_{22n},$$

where  $Q_{21n}$  and  $Q_{22n}$  have the following forms:

$$\begin{aligned} Q_{21n} &= \frac{1}{n(n-1)^3} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{l \neq i}^n \sum_{k \neq j}^n \frac{1}{h^{q_1} h_1^{2q_1}} \frac{1}{p_i p_j} K_{B_1^{ij}} Q_{B_1^{il}} Q_{B_1^{jk}} \\ &\quad \{y_l - g(B_1^\top x_i)\} \{y_k - g(B_1^\top x_j)\}, \end{aligned} \tag{A.6}$$

$$Q_{22n} \equiv (\hat{B}_1 - B_1)^\top Q_{221n} + (\hat{B}_1 - B_1)^\top Q_{222n} + (\hat{B} - B)^\top Q_{223n} \tag{A.7}$$

with  $Q_{221n}$ ,  $Q_{222n}$  and  $Q_{223n}$  being the following forms:

$$\begin{aligned} Q_{221n} &= \frac{1}{n(n-1)^3} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{l \neq i}^n \sum_{k \neq j}^n \frac{1}{h^{q_1} h_1^{3q_1}} \frac{1}{p_i p_j} K_{B^*{}^{ij}} Q'_{B_1^*{}^{il}} Q_{B_1^*{}^{jk}} \\ &\quad \{y_l - g(B_1^\top x_i)\} \{y_k - g(B_1^\top x_j)\} (x_i - x_l), \end{aligned} \tag{A.8}$$

$$\begin{aligned} Q_{222n} &= \frac{1}{n(n-1)^3} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{l \neq i}^n \sum_{k \neq j}^n \frac{1}{h^{q_1} h_1^{3q_1}} \frac{1}{p_i p_j} K_{B^*{}^{ij}} Q_{B_1^*{}^{il}} Q'_{B_1^*{}^{jk}} \\ &\quad \{y_l - g(B_1^\top x_i)\} \{y_k - g(B_1^\top x_j)\} (x_j - x_k), \end{aligned} \tag{A.9}$$

and

$$Q_{223n} = \frac{1}{n(n-1)^3} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{l \neq i}^n \sum_{k \neq j}^n \frac{1}{h^{2q_1} h_1^{2q_1}} \frac{1}{p_i p_j} K'_{B^*{}^{ij}} Q_{B_1^*{}^{il}} Q_{B_1^*{}^{jk}}$$

$$\{y_l - g(B_1^\top x_i)\}\{y_k - g(B_1^\top x_j)\}(z_i - z_j). \tag{A.10}$$

Here, using the similar argument as the justification for the term  $Q_{1n}$ , we also conclude that  $B^* = \{B_{ij}^*\}_{d \times q_1}$  with  $B_{ij}^* \in [\min\{\hat{B}_{ij}, B_{ij}\}, \max\{\hat{B}_{ij}, B_{ij}\}]$  and  $B_1^* = \{B_{1ij}^*\}_{p \times q_1}$  with  $B_{1ij}^* \in [\min\{\hat{B}_{1ij}, B_{1ij}\}, \max\{\hat{B}_{1ij}, B_{1ij}\}]$ . As proved for the term  $Q_{12n}$ , we also assert that replacing  $B^*$  and  $B_1^*$  by  $B$  and  $B_1$ , respectively, do not influence the convergence rate of the term  $Q_{22n}$ .

Similarly as the proof of Proposition A.1 in Fan and Li (1996)[6], when we want to finish the proof of this theorem, what we need to prove is that  $E(Q_{21n}^2) = o_p(n^{-2}h^{-q_1})$ . It is obvious that the calculation of  $E(Q_{21n}^2)$  would be very tedious. We first prove that  $E(Q_{21n}) = o_p(n^{-1}h^{-q_1/2})$ .

Consider two cases with different combinations of indices  $i, j, l, k$ .

Case I:  $\mathcal{A}_1 = \{i, j, l, k \text{ are all different from each other}\}$ . Denote the resulting expression as  $Q_{211n}$ . Under the assumption that  $nh_1^{2\eta}h^{q_1/2} = o_p(1)$ , by applying Lemma B.1, Lemmas 2 and 3 in Robinson (1988)[19], we have

$$\begin{aligned} E(Q_{211n}) &= \frac{1}{h^{q_1}h_1^{2q_1}}E\left(E_1\left[\frac{1}{p_1}Q_{B_1 12}\{g(B_1^\top x_2) - g(B_1^\top x_1)\}\right]\right. \\ &\quad \left.E_3\left[\frac{1}{p_3}Q_{B_1 34}\{g(B_1^\top x_4) - g(B_1^\top x_3)\}\right]K_{B13}\right); \\ &\leq C\frac{h_1^{2\eta}}{h^{q_1}}E\{D_g(B_1^\top x_1)D_g(B_1^\top x_3)K_{B13}\} \\ &= O_p(h_1^{2\eta}) = o_p(n^{-1}h^{-q_1/2}). \end{aligned}$$

Case II:  $\mathcal{A}_2 = \{i, j, l, k \text{ take no more than three different values}\}$ . Denote the term as  $Q_{212n}$ . It is easy to derive that  $E(Q_{212n}) = o_p(n^{-1}h^{-q_1/2})$ .

Hence, altogether, we have  $E(Q_{21n}) = E(Q_{211n}) + E(Q_{212n}) = o_p(n^{-1}h^{-q_1/2})$ .

Now we turn to compute  $E(Q_{21n}^2)$ . It can be decomposed as:

$$\begin{aligned} &E(Q_{21n}^2) \\ &= E\left[\frac{1}{n(n-1)^3} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{l \neq i}^n \sum_{k \neq j}^n \frac{1}{h^{q_1}h_1^{2q_1}} \frac{1}{p_i p_j} K_{Bij} Q_{B_1 il} Q_{B_1 jk} \right. \\ &\quad \left. \{y_l - g(B_1^\top x_i)\}\{y_k - g(B_1^\top x_j)\}\right]^2 \\ &= \frac{1}{n^2(n-1)^6} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{l \neq i}^n \sum_{k \neq j}^n \sum_{i'=1}^n \sum_{j' \neq i'}^n \sum_{l' \neq i'}^n \sum_{k' \neq j'}^n \frac{1}{h^{2q_1}h_1^{4q_1}} \\ &\quad E\left(\left[\frac{1}{p_i p_j} K_{Bij} Q_{B_1 il} Q_{B_1 jk} \{g(B_1^\top x_l) - g(B_1^\top x_i)\}\{g(B_1^\top x_k) - g(B_1^\top x_j)\}\right] \right. \\ &\quad \left. \left[\frac{1}{p_{i'} p_{j'}} K_{B_{i'j'}} Q_{B_1 i'l'} Q_{B_1 j'k'} \{g(B_1^\top x_{l'}) - g(B_1^\top x_{i'})\}\{g(B_1^\top x_{k'}) - g(B_1^\top x_{j'})\}\right] \right) \\ &:= LA. \end{aligned}$$

Firstly, when the indices  $i, j, l, k$  are all different from  $i', j', l', k'$ , the two parts in two different braces are independent of each other. Define the related sum as  $LA1$ . Applying the same argument as that for proving  $E(Q_{21n}) = o_p(n^{-1}h^{-q_1/2})$ , we derive  $LA1 = o_p(n^{-2}h^{-q_1})$ .

Secondly, consider the case where exactly one index from  $i, j, l, k$  equals one of subscripts  $i', j', l', k'$ . By symmetry, we only need to compute Case (i):  $i = i'$ ; Case (ii):  $i = l'$  and Case (iii)  $l = l'$ . The three cases respond the related sums defined as  $LA2$ ,  $LA3$  and  $LA4$ , respectively.

Under Case (i), we have

$$\begin{aligned}
LA2 &= \frac{1}{n^2(n-1)^6 h^{2q_1} h_1^{4q_1}} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{l \neq i}^n \sum_{k \neq j}^n E \left( \left[ \frac{1}{p_i p_j} K_{B_{ij}} Q_{B_{1i}l} Q_{B_{1jk}} \right. \right. \\
&\quad \left. \left. \{g(B_1^\top x_l) - g(B_1^\top x_i)\} \{g(B_1^\top x_k) - g(B_1^\top x_j)\} \right] \right. \\
&\quad \left. \times \sum_{j' \neq i'}^n \sum_{l' \neq i'}^n \sum_{k' \neq j'}^n \left[ \frac{1}{p_{i'} p_{j'}} K_{B_{i'j'}} Q_{B_{1i'l'}} Q_{B_{1j'k'}} \right. \right. \\
&\quad \left. \left. \{g(B_1^\top x_{l'}) - g(B_1^\top x_{i'})\} \{g(B_1^\top x_{k'}) - g(B_1^\top x_{j'})\} \right] \right) \\
&= \frac{1}{nh^{2q_1} h_1^{4q_1}} E \left[ \frac{1}{p_1 p_2} K_{B_{12}} Q_{B_{113}} Q_{B_{124}} \right. \\
&\quad \left. \times \{g(B_1^\top x_3) - g(B_1^\top x_1)\} \{g(B_1^\top x_4) - g(B_1^\top x_2)\} \right. \\
&\quad \left. \frac{1}{p_1 p_5} K_{B_{15}} Q_{B_{116}} Q_{B_{157}} \{g(B_1^\top x_6) - g(B_1^\top x_1)\} \{g(B_1^\top x_7) - g(B_1^\top x_5)\} \right] \\
&= \frac{1}{nh^{2q_1} h_1^{4q_1}} E \left( \left[ E_1(g(B_1^\top x_3) - g(B_1^\top x_1)) Q_{B_{113}} \right] \right. \\
&\quad \times \left[ \frac{1}{p_1 p_2} E_2(g(B_1^\top x_4) - g(B_1^\top x_2)) Q_{B_{124}} \right] \\
&\quad \times \frac{1}{p_1 p_5} \left[ E_1 \{g(B_1^\top x_6) - g(B_1^\top x_1)\} Q_{B_{116}} \right] \\
&\quad \left. \times \left[ E_5 \{g(B_1^\top x_7) - g(B_1^\top x_5)\} Q_{B_{157}} \right] \right) \\
&\leq \frac{(n-1)^4 h_1^{4q_1}}{n^5 h^{2q_1}} E \left\{ \frac{1}{p_1 p_2 p_1 p_5} D_g(B_1^\top x_1) D_g(B_1^\top x_2) \right. \\
&\quad \left. \times K_{B_{12}} D_g(B_1^\top x_1) D_g(B_1^\top x_5) K_{B_{15}} \right\} \\
&= O_p(h_1^{4q_1} n^{-1}) = o_p(n^{-2} h^{-q_1}).
\end{aligned}$$

For Case (ii), we have

$$\begin{aligned}
LA3 &= \frac{1}{n^2(n-1)^6 h^{2q_1} h_1^{4q_1}} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{l \neq i}^n \sum_{k \neq j}^n E \left( \left[ \frac{1}{p_i p_j} K_{B_{ij}} Q_{B_{1i}l} Q_{B_{1jk}} \right. \right. \\
&\quad \left. \left. \{g(B_1^\top x_l) - g(B_1^\top x_i)\} \{g(B_1^\top x_k) - g(B_1^\top x_j)\} \right] \right)
\end{aligned}$$

$$\begin{aligned}
 & \times \sum_{i' \neq i}^n \sum_{j' \neq i'}^n \sum_{k' \neq j'}^n \left[ \frac{1}{p_{i'} p_{j'}} K_{B_{i'j'}} Q_{B_{1i'i}} Q_{B_{1j'k'}} \right. \\
 & \left. \{g(B_1^\top x_i) - g(B_1^\top x_{i'})\} \{g(B_1^\top x_{k'}) - g(B_1^\top x_{j'})\} \right] \\
 & = \frac{1}{nh^{2q_1} h_1^{4q_1}} E \left[ \frac{1}{p_1 p_2} K_{B_{12}} Q_{B_{113}} Q_{B_{124}} \right. \\
 & \quad \times \{g(B_1^\top x_3) - g(B_1^\top x_1)\} \{g(B_1^\top x_4) - g(B_1^\top x_2)\} \\
 & \quad \left. \frac{1}{p_5 p_6} K_{B_{56}} Q_{B_{115}} Q_{B_{167}} \{g(B_1^\top x_1) - g(B_1^\top x_5)\} \{g(B_1^\top x_7) - g(B_1^\top x_6)\} \right] \\
 & = \frac{1}{nh^{2q_1} h_1^{4q_1}} E \left\{ K_{B_{12}} [E_1 \{g(B_1^\top x_3) - g(B_1^\top x_1)\} Q_{B_{113}}] \right. \\
 & \quad \times \left[ \frac{1}{p_1 p_2} E_2 \{g(B_1^\top x_4) - g(B_1^\top x_2)\} Q_{B_{124}} \right] \\
 & \quad \times \frac{1}{p_5 p_6} K_{B_{56}} E_1 [\{g(B_1^\top x_1) - g(B_1^\top x_5)\} Q_{B_{115}}] \\
 & \quad \left. \times [E_6 \{g(B_1^\top x_7) - g(B_1^\top x_6)\} Q_{B_{167}}] \right\} \\
 & \leq \frac{(n-1)^4 h_1^{4\eta}}{n^5 h^{2q_1}} E \left\{ \frac{1}{p_1 p_2 p_1 p_5} D_g(B_1^\top x_1) D_g(B_1^\top x_2) \right. \\
 & \quad \left. \times K_{B_{12}} D_g(B_1^\top x_1) D_g(B_1^\top x_6) K_{B_{56}} \right\} \\
 & = O_p(h_1^{3\eta} n^{-1}) = o_p(n^{-2} h^{-q_1}).
 \end{aligned}$$

Similarly, it is easy to prove that for Case (iii),  $LA4 = o_p(n^{-2} h^{-q_1})$ .

Finally, when the indices  $i, j, l, k, i', j', l', k'$  take no more than six different values whose sum is defined as  $LA5$ , it is easy to see that  $LA5 = o_p(n^{-2} h^{-q_1})$ .

Combining all cases, we get  $E(Q_{21n}^2) = LA1 + LA2 + 9LA3 + 6LA4 + LA5 = o_p(n^{-2} h^{-q_1})$ . The application of Chebyshev's inequality yields  $Q_{21n} = o_p(n^{-1} h^{-q_1/2})$ .

The similar arguments can be applied to handle the first and second moment of  $Q_{22n}$  defined in (A.7). The first moment is very similar to that for  $Q_{21n}$ . Note that  $Q_{22n}$  is the weighted sum of  $Q_{221n}$ ,  $Q_{222n}$  and  $Q_{223n}$  in (A.8)–(A.10). Thus, we only need to respectively handle  $Q_{221n}$ ,  $Q_{222n}$  and  $Q_{223n}$ . Also, they can be treated similarly as those for  $Q_{21n}$ . When Conditions A4–A9 hold, it is easy to derive the following convergence rates:

$$E(Q_{221n}^2) = O_p(\max\{h_1^{2\eta}, h_1^{2\eta} n^{-1}, h_1^\eta n^{-1}\}) = o_p(n^{-1} h^{-q_1/2}), \quad (\text{A.11})$$

$$E(Q_{222n}^2) = O_p(\max\{h_1^{2\eta}, h_1^{2\eta} n^{-1}, h_1^\eta n^{-1}\}) = o_p(n^{-1} h^{-q_1/2}), \quad (\text{A.12})$$

$$E(Q_{223n}^2) = O_p(\max\{h_1^{4\eta}, h_1^{4\eta} n^{-1}, h_1^{3\eta} n^{-1}\}) = o_p(n^{-1} h^{-q_1/2}). \quad (\text{A.13})$$

Due to the facts  $\|\hat{B} - B\| = O_p(1/\sqrt{n})$  and  $\|\hat{B}_1 - B_1\| = O_p(1/\sqrt{n})$ , together with (A.11)–(A.13), we derive  $Q_{22n} = o_p(n^{-1} h^{-q_1/2})$  by an application of Chebyshev's inequality. Therefore, altogether, from the definition of  $Q_{2n}$  in (A.5), we have that  $Q_{2n} = o_p(n^{-1} h^{-q_1/2})$ .

Lastly, we consider the terms  $Q_{3n}$  in (A.2). Also it is easy to see that

$$\begin{aligned} Q_{3n} &= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{q_1}} K_{\hat{B}_{ij}} u_i (\hat{g}_j - g_j) \frac{\hat{p}_j}{p_j} \\ &\quad + \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{q_1}} K_{\hat{B}_{ij}} u_i (\hat{g}_j - g_j) \left( \frac{\hat{p}_j - p_j}{p_j} \right) \\ &\equiv: \tilde{Q}_{3n} + o_p(\tilde{Q}_{3n}). \end{aligned}$$

Substituting the kernel estimates  $\hat{g}$  and  $\hat{p}$  into  $\tilde{Q}_{3n}$ , we have

$$\tilde{Q}_{3n} = \frac{1}{n^2(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{k \neq j}^n \frac{1}{h^{q_1} h_1^{q_1}} \frac{1}{p_j} K_{\hat{B}_{ij}} u_i Q_{\hat{B}_{1jk}} \{y_k - g(B_1^\top x_j)\}.$$

By using the Taylor expansion for  $\tilde{Q}_{3n}$  with respect to  $B$  and  $B_1$ , we can have

$$\tilde{Q}_{3n} \equiv: Q_{31n} + Q_{32n},$$

where  $Q_{31n}$  and  $Q_{32n}$  have following forms:

$$Q_{31n} = \frac{1}{n(n-1)^2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{k \neq j}^n \frac{1}{h^{q_1} h_1^{q_1}} \frac{1}{p_j} K_{B_{ij}} u_i Q_{B_{1jk}} \{y_k - g(B_1^\top x_j)\}$$

and

$$\begin{aligned} Q_{32n} &= \frac{1}{n(n-1)^2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{k \neq j}^n \frac{1}{h^{q_1} h_1^{q_1}} \frac{1}{p_j} K_{B^*_{ij}} u_i Q'_{B^*_{1jk}} \\ &\quad \times \{y_k - g(B_1^\top x_j)\} (\hat{B}_1 - B_1)^\top (x_i - x_l) \\ &\quad + \frac{1}{n(n-1)^2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{k \neq j}^n \frac{1}{h^{q_1} h_1^{q_1}} \frac{1}{p_j} K'_{B^*_{ij}} u_i Q_{B^*_{1jk}} \\ &\quad \times \{y_k - g(B_1^\top x_j)\} (\hat{B} - B)^\top (z_i - z_j) \\ &\equiv (\hat{B}_1 - B_1)^\top Q_{321n} + (\hat{B} - B)^\top Q_{322n}. \end{aligned}$$

Here, applying the similar justification for the term  $Q_{1n}$ , we also conclude  $B^* = \{B^*_{ij}\}_{d \times q_1}$  with  $B^*_{ij} \in [\min\{\hat{B}_{ij}, B_{ij}\}, \max\{\hat{B}_{ij}, B_{ij}\}]$  and  $B^*_1 = \{B^*_{1ij}\}_{p \times q_1}$  with  $B^*_{1ij} \in [\min\{\hat{B}_{1ij}, B_{1ij}\}, \max\{\hat{B}_{1ij}, B_{1ij}\}]$ . Similarly, replacing  $B^*$  and  $B^*_1$  by  $B$  and  $B_1$ , respectively, does not effect the convergence rate of the term  $Q_{32n}$ .

Because  $E(u_i | z_i) = 0$ , we have  $E(Q_{31n}) = 0$ . Then we compute the second moment of  $Q_{31n}$  as follows:

$$E(Q_{31n}^2) = E \left[ \frac{1}{n(n-1)^2} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{k \neq j}^n \frac{1}{h^{q_1} h_1^{q_1}} \frac{1}{p_j} K_{B_{ij}} u_i Q_{B_{1jk}} \{y_k - g(B_1^\top x_j)\} \right]^2$$

$$\begin{aligned}
 &= E \left[ \frac{1}{n^2(n-1)^4} \frac{1}{h^{2q_1} h_1^{2q_1}} \right. \\
 &\quad \times \sum_{i=1}^n \sum_{j \neq i}^n \sum_{k \neq j}^n \sum_{i'=1}^n \sum_{j' \neq i'}^n \sum_{k' \neq j'}^n \frac{1}{p_j p_{j'}} K_{B_{ij}} Q_{B_{1jk}} K_{B_{i'j'}} Q_{B_{1j'k'}} \\
 &\quad \left. u_i u_{i'} \{y_k - g(B_1^\top x_j)\} \{y_{k'} - g(B_1^\top x_{j'})\} \right] \\
 &= E \left[ \frac{1}{n^2(n-1)^4} \frac{1}{h^{2q_1} h_1^{2q_1}} \right. \\
 &\quad \times \sum_{i=1}^n \sum_{j \neq i}^n \sum_{k \neq j}^n \sum_{i'=1}^n \sum_{j' \neq i'}^n \sum_{k' \neq j'}^n \frac{1}{p_j p_{j'}} K_{B_{ij}} Q_{B_{1jk}} K_{B_{i'j'}} Q_{B_{1j'k'}} \\
 &\quad \left. u_i u_{i'} \{g(B_1^\top x_k) - g(B_1^\top x_j)\} \{g(B_1^\top x_{k'}) - g(B_1^\top x_{j'})\} \right] + o_p(n^{-2}h^{-q_1}).
 \end{aligned}$$

Since  $E(u_i u_{i'}) \neq 0$  if and only if  $i = i'$ , we have

$$\begin{aligned}
 &E(Q_{31n}^2) \\
 &= \frac{1}{n} \frac{1}{h^{2q_1} h_1^{2q_1}} E(u_1^2) E \left( \frac{1}{p_2 p_4} K_{B_{12}} Q_{B_{123}} K_{B_{14}} Q_{B_{145}} \right. \\
 &\quad \left. \{g(B_1^\top x_3) - g(B_1^\top x_2)\} \{g(B_1^\top x_5) - g(B_1^\top x_4)\} \right) \\
 &= \frac{1}{n} \frac{1}{h^{2q_1} h_1^{2q_1}} E(u_1^2) E \left( \frac{1}{p_2 p_4} K_{B_{12}} K_{B_{14}} E_2 [Q_{B_{123}} \{g(B_1^\top x_3) - g(B_1^\top x_2)\}] \right. \\
 &\quad \left. \times E_4 [Q_{B_{145}} \{g(B_1^\top x_5) - g(B_1^\top x_4)\}] \right) \\
 &\leq \frac{1}{n} \frac{h_1^{2\eta}}{h^{2q_1}} E \left\{ \frac{1}{p_2 p_4} K_{B_{12}} K_{B_{14}} D_g(B_1^\top x_2) D_g(B_1^\top x_4) \right\} \\
 &= O_p(h_1^{2\eta} n^{-1}) = o_p(n^{-2}h^{-q_1}),
 \end{aligned}$$

Employing Lemma B.1, Lemmas 2 and 3 in Robinson (1988)[19] again, we have  $Q_{31n} = o_p(n^{-1}h^{-q_1/2})$ . Also, following the similar arguments used for proving  $Q_{22n}$  we get  $Q_{32n} = o_p(n^{-1}h^{-q_1/2})$  and then  $Q_{3n} = o_p(n^{-1}h^{-q_1/2})$ .

In summary, we conclude that:

$$nh^{q_1/2} V_n \Rightarrow N(0, s^2).$$

The variance  $s^2$  can be estimated by

$$\hat{s}^2 = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n K_h^2 \left( \hat{B}^\top z_i - \hat{B}^\top z_j \right) \hat{u}_i^2 \hat{u}_j^2.$$

Since the proof is straightforward, we only give a very brief outline. Under the null hypothesis, the estimates  $\hat{B}$  and  $\hat{g}$  are consistent to  $B$  and  $g$ , and some elementary calculations result in an asymptotic presentation as:

$$\hat{s}^2 = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n K_h^2 \left( \hat{B}^\top z_i - \hat{B}^\top z_j \right) u_i^2 u_j^2 + o_p(1).$$



Applying a similar argument used for proving Lemma 2 in Guo et al. (2016)[8], one can derive

$$\begin{aligned}\hat{s}^2 &= \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n K_h^2(B^\top z_i - B^\top z_j) u_i^2 u_j^2 + o_p(1) \\ &\equiv: \tilde{s}^2 + o_p(1),\end{aligned}$$

with  $B = (B_1^\top, 0_{q_1 \times p_2})^\top$ . As  $\tilde{s}^2$  is an  $U$ -statistic and it is easy to prove  $\tilde{s}^2 \rightarrow s^2$  in probability. The more details can be referred to Zheng (1996)[23]. The proof of Theorem 3.1 is finished.  $\square$

*Proof of Proposition 3.1.* Consider RERE that is based on  $DEE_{SIR}$  under the local alternative hypotheses. From the proof of Theorem 3.2 in Li et al. (2008)[13], we see that to detain  $M_n - M = O_p(C_n)$ , it is only needed to prove  $M_n(t) - M(t) = O_p(C_n)$  uniformly, where  $M(t) = \Sigma^{-1} \text{Var}(E\{Z|I(Y \leq t)\}) = \Sigma^{-1}(\nu_1 - \nu_0)(\nu_1 - \nu_0)^\top p_t(1 - p_t)$ ,  $\Sigma$  is the covariance matrix of  $Z$ ,  $\nu_0 = E\{Z|I(Y \leq t) = 0\}$ ,  $\nu_1 = E\{Z|I(Y \leq t) = 1\}$  and  $p_t = E(I(Y \leq t))$ .

Further, we note that

$$\begin{aligned}\nu_1 - \nu_0 &= \frac{E\{ZI(Y \leq t)\}}{p_t} - \frac{E\{ZI(Y > t)\}}{1 - p_t} \\ &= \frac{E\{ZI(Y \leq t) - E(Z)\}E\{I(Y \leq t)\}}{p_t(1 - p_t)}.\end{aligned}$$

Therefore, the matrix  $M(t)$  can also be reformulated as

$$\begin{aligned}M(t) &= \Sigma^{-1}(E\{\{Z - E(Z)\}I(Y \leq t)\})(E\{\{Z - E(Z)\}I(Y \leq t)\})^\top \\ &=: \Sigma^{-1}\tilde{m}(t)\tilde{m}(t)^\top,\end{aligned}$$

where  $\tilde{m}(t) = E[\{Z - E(Z)\}I(Y \leq t)]$ . Correspondingly,  $\tilde{m}(t)$  can be simply estimated by

$$\tilde{m}_n(t) = n^{-1} \sum_{i=1}^n (z_i - \bar{z})I(y_i \leq t).$$

Then  $M(t)$  can be estimated by

$$M_n(t) = \hat{\Sigma}^{-1}L_n(t),$$

where  $\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i$ ,  $L_n(t) = \tilde{m}_n(t)\tilde{m}_n(t)^\top$  and  $\hat{\Sigma}$  is the estimate of  $\Sigma$ . Denote respectively the response under the null and local alternative hypotheses as  $Y$  and  $Y_n$  to show the dependence of the response under the local alternative hypotheses. Then under  $H_{1n}$ ,

$$E\{ZI(Y_n \leq t)\} - E\{ZI(Y \leq t)\} = E\{ZP(Y_n \leq t|Z)\} - E\{ZP(Y \leq t|Z)\},$$

where  $Y_n = g(B_1^\top X) + C_n G(B^\top Z) + \epsilon \equiv: Y + C_n G(B^\top Z)$ . Thus, for all  $t$ , we have

$$P(Y_n \leq t|Z) - P(Y \leq t|Z) = F_{Y|Z}\{t - C_n G(B^\top Z)\} - F_{Y|Z}(t)$$

$$= -C_n G(B^\top Z) f_{Y|Z}(t) + O_p(C_n).$$

Under Condition A2, we can conclude that  $n^{-1} \sum_{i=1}^n z_i I(y_{ni} \leq t) - E\{ZI(Y \leq t)\} = O_p(\max\{C_n, n^{-1/2}\}) = O_p(C_n)$ . By the parallel argument for justifying Theorem 3.2 of Li et al. (2008)[13],  $M_n(t) - M(t) = O_p(C_n)$  uniformly and then  $M_n - M = O_p(C_n)$ .

Similarly as those in Zhu and Fang (1996)[26] and Zhu and Ng (1995)[28], we get that  $\hat{\lambda}_i - \lambda_i = O_p(C_n)$ , where  $\hat{\lambda}_d \leq \hat{\lambda}_{d-1} \leq \dots \leq \hat{\lambda}_1$  are the eigenvalues of the matrix  $M_n$ . Note that  $M$  corresponds to the null hypothesis and then satisfies that  $\text{Span}\{M\} = S_{Y|E(Y|Z)}$  and  $\text{rank}(M) = q_1$ . This implies that  $\lambda_d = \dots = \lambda_{q_1+1} = 0$  and  $0 < \lambda_{q_1} \leq \dots \leq \lambda_1$  that are the eigenvalues of the matrix  $M$ . It is clear that for any  $l > q_1$ ,  $\lambda_l = 0$ , so we get  $(\lambda_l^*)^2 = O_p(C_n^2)$ , where  $\lambda^*$  is defined in (2.6). Recall the definition of  $\tilde{\lambda}$  (A.1) in the beginning of the proof of Proposition 2.2. For any  $1 \leq l \leq q_1$ , we have  $(\lambda_l^*)^2 = (\tilde{\lambda}_l)^2 + O_p(C_n)$ .

*Prove part (I).* We consider the case  $C_n = O(n^{-1/2}h^{-q_1/4})$ . Since  $c \times n^{-1}h^{-q_1/2} \log n \leq c_n \rightarrow 0$  with some fixed  $c > 0$  where  $c_n$  is the ridge value in the criterion RERE. we have  $C_n^2 = o_p(c_n)$ .

When  $l > q_1$ , as  $c \times n^{-1}h^{-q_1/2} \log n \leq c_n \rightarrow 0$ , we have,

$$\begin{aligned} \frac{(\lambda_{q_1+1}^*)^2 + c_n}{(\lambda_{q_1}^*)^2} - \frac{(\lambda_{l+1}^*)^2 + c_n}{(\lambda_l^*)^2} &= \frac{\tilde{\lambda}_{q_1+1}^2 + c_n + O_p(C_n^2)}{\tilde{\lambda}_{q_1}^2 + c_n + O_p(C_n)} - \frac{\tilde{\lambda}_{l+1}^2 + c_n + O_p(C_n^2)}{\tilde{\lambda}_l^2 + c_n + O_p(C_n^2)} \\ &= \frac{\tilde{\lambda}_{q_1+1}^2 + c_n + o_p(c_n)}{\tilde{\lambda}_{q_1}^2 + c_n + O_p(C_n^2)} - \frac{\tilde{\lambda}_{l+1}^2 + c_n + o_p(c_n)}{\tilde{\lambda}_l^2 + c_n + o_p(c_n)} \\ &= \frac{c_n + o_p(c_n)}{\tilde{\lambda}_{q_1}^2 + c_n + O_p(C_n^2)} - \frac{c_n + o_p(c_n)}{c_n + o_p(c_n)}. \end{aligned}$$

Therefore in probability

$$\frac{(\lambda_{q_1+1}^*)^2 + c_n}{(\lambda_{q_1}^*)^2} - \frac{(\lambda_{l+1}^*)^2 + c_n}{(\lambda_l^*)^2} \rightarrow -1 < 0.$$

When  $1 \leq l < q_1$ , we derive that:

$$\begin{aligned} \frac{(\lambda_{q_1+1}^*)^2 + c_n}{(\lambda_{q_1}^*)^2} - \frac{(\lambda_{l+1}^*)^2 + c_n}{(\lambda_l^*)^2} &= \frac{\tilde{\lambda}_{q_1+1}^2 + c_n + O_p(C_n^2)}{\tilde{\lambda}_{q_1}^2 + c_n + O_p(C_n)} - \frac{\tilde{\lambda}_{l+1}^2 + c_n + O_p(C_n)}{\tilde{\lambda}_l^2 + c_n + O_p(C_n)} \\ &= \frac{c_n + o_p(c_n)}{\tilde{\lambda}_{q_1}^2 + c_n + O_p(c_n)} - \frac{\tilde{\lambda}_{l+1}^2 + c_n + o_p(c_n)}{\tilde{\lambda}_l^2 + c_n + o_p(c_n)}. \end{aligned}$$

Then we have in probability

$$\frac{(\lambda_{q_1+1}^*)^2 + c_n}{(\lambda_{q_1}^*)^2} - \frac{(\lambda_{l+1}^*)^2 + c_n}{(\lambda_l^*)^2} \rightarrow -\frac{\tilde{\lambda}_{l+1}^2}{\tilde{\lambda}_l^2} < 0.$$

Therefore, altogether, we can conclude that  $\hat{q} = q_1$  with a probability approaching 1.

Prove part (II). Now we consider the case  $C_n \times n^{1/2}h^{q_1/4} \rightarrow \infty$ . When  $1 \leq l < q_1$ , if  $c \times n^{-1}h^{-q_1/2} \log n \leq c_n \rightarrow 0$ , we derive that:

$$\begin{aligned} & \frac{(\lambda_{q_1+1}^*)^2 + c_n}{(\lambda_{q_1}^*)^2} - \frac{(\lambda_{l+1}^*)^2 + c_n}{(\lambda_l^*)^2} \\ &= \frac{\tilde{\lambda}_{q_1+1}^2 + c_n + O_p(C_n^2)}{\tilde{\lambda}_{q_1}^2 + c_n + O_p(C_n)} - \frac{\tilde{\lambda}_{l+1}^2 + c_n + O_p(C_n)}{\tilde{\lambda}_l^2 + c_n + O_p(C_n)} \\ &= \frac{O_p(\max\{c_n, C_n^2\})}{\tilde{\lambda}_{q_1}^2 + O_p(\max\{c_n, C_n\})} - \frac{\tilde{\lambda}_{l+1}^2 + O_p(\max\{c_n, C_n\})}{\tilde{\lambda}_l^2 + O_p(\max\{c_n, C_n\})}. \end{aligned}$$

Then we have in probability

$$\frac{(\lambda_{q_1+1}^*)^2 + c_n}{(\lambda_{q_1}^*)^2} - \frac{(\lambda_{l+1}^*)^2 + c_n}{(\lambda_l^*)^2} \rightarrow -\frac{\tilde{\lambda}_{l+1}^2}{\tilde{\lambda}_l^2} < 0.$$

This implies that  $\hat{q}$  determined by (2.7) is greater than  $q_1$  or equals to  $q_1$  with a probability approaching 1, namely  $P(\hat{q} \geq q_1) = 1$ . □

*Proof of Theorem 3.2.* Prove Part (I). Since the details of the proof are similar to those for proving Theorem 3.1, we only sketch it. First, according to Proposition 2.2, under the global alternative hypothesis,  $P(\hat{q} = q_2) \rightarrow 1$ , using the similar argument as that in Theorem 3.1, we can get that for any fixed constant  $c$ ,  $\lim_{n \rightarrow \infty} P(T_n \geq c) = \lim_{n \rightarrow \infty} P(T_n \geq c, \hat{q} = q_2)$ . Therefore, under the global alternative hypothesis, in an asymptotic sense we can only need to consider the conditional probability when the event  $\hat{q} = q_2$  is given.

By using  $\|\hat{B}_1 - B_1\| = O_p(1/\sqrt{n})$  and Conditions A6 and A8 in the Appendix,  $\hat{g}(\hat{B}_1^\top x)$  is a uniformly consistent estimate of  $g(B_1^\top x) = E(Y|B_1^\top X = B_1^\top x)$ , see Powell et al. (1989)[16] or Robinson (1988)[19]. We then have

$$V_n = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} K_{\hat{B}_{ij}} \hat{u}_i \hat{u}_j = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{q_2}} K_{B_{ij}} u_i u_j + o_p(1),$$

where  $u_i = y_i - g(B_1^\top x_i)$  with  $g(B_1^\top x_i) = E(y_i|x_i)$ . Let  $\Delta(z_i) = m(B^\top z_i) - g(B_1^\top x_i)$ . Therefore, by using the  $U$ -statistics theory, we get that

$$V_n = E(K_{B_{12}} u_1 u_2) + o_p(1) = E\{\Delta^2(Z)p(B^\top Z)\} + o_p(1).$$

Similarly, we can also prove that in probability  $\hat{s}^2$  converges to a positive value which may be different from  $s^2$  defined by (3.1). Therefore, we can obtain  $T_n/(nh^{q_1/2}) \Rightarrow Constant > 0$  in probability.

Consider Part (II). Under the local alternative hypotheses with  $C_n = n^{-1/2}h^{-q_1/4}$ , according to Lemma 3.1, we have  $\hat{q} = q_1$  with a probability tending to one. By applying the same argument as that in Theorem 3.1, we can get that for any constant  $c$ , we have  $\lim_{n \rightarrow \infty} P(T_n \geq c) = \lim_{n \rightarrow \infty} P(T_n \geq c, \hat{q} = q_1)$ . This implies that under the local alternative hypotheses with  $C_n = n^{-1/2}h^{-q_1/4}$ ,

in an asymptotic sense we still only need to consider the case  $\hat{q} = q_1$  in the rest proof of this theorem.

Following the similar arguments used to prove Theorem 3.1 or the similar decomposition by Taylor expansion as the proof of Theorem 3 in Guo et al. (2016)[8], we can show that:

$$V_n = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{\hat{q}}} K_{\tilde{B}ij} \hat{u}_i \hat{u}_j =: Q_n + o_p(Q_n)$$

where  $Q_n = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{q_1}} K_{\tilde{B}ij} u_i u_j$  with  $\tilde{B} = (B_1^\top, 0_{q_1 \times p_2})^\top$  and  $u_i = y_i - g(B_1^\top x_i) = C_n G(B^\top z_i) + \epsilon_i$  with  $E(\epsilon_i | z_i) = 0$ . Here we omit the technical details. Then we prove  $nh^{q_1/2}Q_n$  is asymptotically normal.  $Q_n$  is further decomposed as:

$$\begin{aligned} Q_n &= \left[ \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{q_1}} K_{\tilde{B}ij} \{C_n G(B^\top z_i) + \epsilon_i\} \{C_n G(B^\top z_j) + \epsilon_j\} \right] \\ &= \left\{ \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{q_1}} K_{\tilde{B}ij} \epsilon_i \epsilon_j \right\} \\ &\quad + C_n \left\{ \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{q_1}} K_{\tilde{B}ij} G(B^\top z_i) \epsilon_j \right\} \\ &\quad + C_n^2 \left\{ \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^{q_1}} K_{\tilde{B}ij} G(B^\top z_i) G(B^\top z_j) \right\} \\ &\equiv W_{1n} + C_n W_{2n} + C_n^2 W_{3n}. \end{aligned}$$

Again following the similar argument as that for Lemma 3.3 in Zheng (1996)[23], we can easily derive that  $nh^{q_1/2}W_{1n} \xrightarrow{d} N(0, s^2)$ . By Lemma 3.1 of Zheng (1996)[23], we get that  $\sqrt{n}W_{2n} = O_p(1)$ . Since  $C_n = n^{-1/2}h^{-q_1/4}$ , it is deduced that  $nh^{q_1/2}C_n W_{2n} = o_p(1)$ . Lastly, we consider the term  $W_{3n}$ . It is obvious that the term  $W_{3n}$  is  $U$ -statistic with the kernel as:

$$H(z_i, z_j) = \frac{1}{h^{q_1}} K_{\tilde{B}ij} G(B^\top z_i) G(B^\top z_j).$$

We firstly calculate the expectation of  $H(z_i, z_j)$  as

$$\begin{aligned} E\{H(z_i, z_j)\} &= \frac{1}{h^{q_1}} E\{K_{\tilde{B}ij} G(B^\top z_i) G(B^\top z_j)\} \\ &= \frac{1}{h^{q_1}} E\left[ K_{\tilde{B}ij} E\{G(B^\top z_i) | \tilde{B}^\top z_i\} E\{G(B^\top z_j) | \tilde{B}^\top z_j\} \right]. \end{aligned}$$

For notational simplicity, write  $M(\tilde{B}^\top z_i) = E\{G(B^\top z_i) | \tilde{B}^\top z_i\}$  and  $t_i = \tilde{B}^\top z_i$ . The expectation of  $H(z_i, z_j)$  can be further calculated as

$$E\{H(z_i, z_j)\} = \frac{1}{h^{q_1}} E\left\{ K\left(\frac{t_i - t_j}{h}\right) M(t_i) M(t_j) \right\}$$

$$= \int \int \frac{1}{h^{q_1}} K\left(\frac{t_i - t_j}{h}\right) M(t_i) M(t_j) p_1(t_i) p_1(t_j) dt_i dt_j.$$

Further, apply the changing variables  $u = \frac{t_i - t_j}{h}$  to get that:

$$\begin{aligned} E\{H(z_i, z_j)\} &= \int \int K(u) M(t_i) M(t_i - hu) p_1(t_i) p_1(t_i - hu) du dt_i \\ &= \int K(u) du \int M^2(t_i) p_1^2(t_i) dt_i + o_p(h) \\ &= E([E\{G(B^\top Z) | \tilde{B}^\top Z\}]^2 p_{\tilde{B}}(\tilde{B}^\top Z)) + o_p(1) \\ &= E([E\{G(B^\top Z) | B_1^\top X\}]^2 p_{B_1}(B_1^\top X)) + o_p(1). \end{aligned}$$

Again using the element characteristics of  $U$ -statistic, we derive that:

$$W_{3n} \Rightarrow E([E\{G(B^\top Z) | B_1^\top X\}]^2 p_{B_1}(B_1^\top X)).$$

Thus, we can deduce that

$$V_n \Rightarrow N(E([E\{G(B^\top Z) | B_1^\top X\}]^2 p_{B_1}(B_1^\top X)), s^2).$$

As the variance  $s^2$  can be consistently estimated and thus, the expected result can be proved.

In summary, invoking Slutsky's theorem, it is concluded that

$$T_n \Rightarrow N(E([E\{G(B^\top Z) | B_1^\top X\}]^2 p_{B_1}(B_1^\top X))/s, 1).$$

The proof of Theorem 3.2 is concluded.  $\square$

## References

- [1] COOK, R. D. (1998). *Regression graphics: ideas for studying regressions through graphics*. Wiley, New York. [MR1645673](#)
- [2] COOK, R. D. and FORZANI, L. (2009). Likelihood-based sufficient dimension reduction. *Journal of the American Statistical Association*. **104** 197–208. [MR2504373](#)
- [3] COOK, R. D. and WEISBERG, S. (1991). Discussion of “Sliced inverse regression for dimension reduction,” by K. C. Li. *Journal of the American Statistical Association*. **86** 316–342. [MR1137117](#)
- [4] DELGADO, M. A. and GONZÁLEZ MANTEIGA, W. (2001). Significance testing in nonparametric regression based on the bootstrap. *The Annals of Statistics*. **29** 1469–1507. [MR1873339](#)
- [5] FAN, J. Q. and HU, T. C. (1992). Bias correction and higher order kernel functions. *Statistics & probability letters*. **13** 235–243. [MR1158864](#)
- [6] FAN, Y. Q. and LI, Q. (1996). Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica*. **64** 865–890. [MR1399221](#)

- [7] FRIENDLY, M. (2002). Corrgrams: Exploratory displays for correlation matrices. *The American Statistician*. **56** 316–324. [MR1963277](#)
- [8] GUO, X., WANG, T. and ZHU, L. X. (2016). Model checking for generalized linear models: a dimension-reduction model-adaptive approach. *Journal of the Royal Statistical Society: Series B*. **78** 1013–1035. [MR3557187](#)
- [9] HOAGLIN, D. C. and VELLEMAN, P. F. (1995). A critical look at some analyses of major league baseball salaries. *The American Statistician*. **49** 277–285.
- [10] LAVERGNE, P. and VUONG, Q. (2000). Nonparametric significance testing. *Econometric Theory*. **16** 576–601. [MR1790292](#)
- [11] LAVERGNE, P., MAISTRE, S. and PATILEA, V. (2015). A significance test for covariates in nonparametric regression. *Electronic Journal of Statistics*. **9** 643–678. [MR3331853](#)
- [12] LI, B. and WANG, S. (2007). On directional regression for dimension reduction. *Journal of the American Statistical Association*. **102** 997–1008. [MR2354409](#)
- [13] LI, B., WEN, S. Q. and ZHU, L. X. (2008). On a projective resampling method for dimension reduction with multivariate responses. *Journal of the American Statistical Association*. **103** 1177–1186. [MR2462891](#)
- [14] LI, K. C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*. **86** 316–327. [MR1137117](#)
- [15] LI, Q. (1999). Consistent model specification tests for time series econometric models. *Journal of Econometrics*. **92** 101–147. [MR1706996](#)
- [16] POWELL, J. L., STOCK, J. H. and STOKER, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica*. **57** 1403–1430. [MR1035117](#)
- [17] RACINE, J. (1997). Consistent significance testing for nonparametric regression. *Journal of Business & Economic Statistics*. **15** 369–378.
- [18] RACINE, J. S., HART, J. and LI, Q. (2006). Testing the significance of categorical predictor variables in nonparametric regression models. *Econometric Reviews*. **25** 523–544. [MR2283340](#)
- [19] ROBINSON, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica*. **56** 931–954. [MR0951762](#)
- [20] STUTE, W. and ZHU, L. X. (2005). Nonparametric checks for single-index models. *The Annals of Statistics*. **33** 1048–1083. [MR2195628](#)
- [21] XIA, Q., XU, W. L. and ZHU, L. X. (2015). Consistently determining the number of factors in multivariate volatility modelling. *The Statistica Sinica*. **25** 1025–1044. [MR3409736](#)
- [22] XIA, Y. C., TONG, H., LI, W. K. and ZHU, L. X. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society: Series B*. **64** 363–410. [MR1924297](#)
- [23] ZHENG, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*. **75** 263–289. [MR1413644](#)
- [24] ZHU, L. P., ZHU, L. X., FERRÉ, L. and WANG, T. (2010a). Sufficient dimension reduction through discretization-expectation estimation. *Biometrika*. **97** 295–304. [MR2650739](#)

- [25] ZHU, L. P., ZHU, L. X. and FENG, Z. H. (2010b). Dimension reduction in regressions through cumulative slicing estimation. *Journal of the American Statistical Association*. **105** 1455–1466. [MR2796563](#)
- [26] ZHU, L. X. and FANG, K. T. (1996). Asymptotics for the kernel estimates of sliced inverse regression. *The Annals of Statistics*. **24** 1053–1067. [MR1401836](#)
- [27] ZHU, L. X., MIAO, B. Q. and PENG, H. (2006). On sliced inverse regression with high dimensional covariates. *Journal of the American Statistical Association*. **101** 630–643. [MR2281245](#)
- [28] ZHU, L. X. and NG, K. W. (1995). Asymptotics for sliced inverse regression. *Statistica Sinica*. **5** 727–736. [MR1347616](#)
- [29] ZHU, X. H., CHEN, F., GUO, X. and ZHU, L. X. (2016). Heteroscedasticity testing for regression models: A dimension reduction-based model adaptive approach. *Computational Statistics & Data Analysis*. **103** 263–283. [MR3522632](#)
- [30] ZHU, X. H., GUO, X. and ZHU, L. X. (2017a). An adaptive-to-model test for partially parametric single-index models. *Statistics and Computing*. **27** 1193–1204. [MR3647092](#)
- [31] ZHU, X. H., WANG, T., ZHAO, J. L. and ZHU, L. X. (2017b). Inference for biased transformation models. *Computational Statistics & Data Analysis*. **109** 105–120. [MR3603644](#)