

An adaptive-to-model test for partially parametric single-index models

Xuehu Zhu^{1,2} · Xu Guo³ · Lixing Zhu^{2,3}

Received: 29 January 2016 / Accepted: 20 June 2016 / Published online: 1 July 2016
© Springer Science+Business Media New York 2016

Abstract Residual marked empirical process-based tests are commonly used in regression models. However, they suffer from data sparseness in high-dimensional space when there are many covariates. This paper has three purposes. First, we suggest a partial dimension reduction adaptive-to-model testing procedure that can be omnibus against general global alternative models although it fully use the dimension reduction structure under the null hypothesis. This feature is because that the procedure can automatically adapt to the null and alternative models, and thus greatly overcomes the dimensionality problem. Second, to achieve the above goal, we propose a ridge-type eigenvalue ratio estimate to automatically determine the number of linear combinations of the covariates under the null and alternative hypotheses. Third, a Monte-Carlo approximation to the sampling null distribution is suggested. Unlike existing bootstrap approximation methods, this gives an approximation as close to the sampling null distribution as possible by fully utilising the dimension reduction model structure under the null model. Simulation

studies and real data analysis are then conducted to illustrate the performance of the new test and compare it with existing tests.

Keywords Model-adaptation · Model checking · Partial sufficient dimension reduction · Ridge-type eigenvalue ratio estimate

1 Introduction

Consider the partially parametric single-index model in the form:

$$Y = G(\beta^\top X, W, \theta) + \epsilon, \quad (1)$$

where Y is the response variable, (X, W) is the covariate vector in $\mathbb{R}^{p_1+p_2}$, $G(\cdot)$ is a known smooth function that depends not only on the covariate $\beta^\top X$ but also on the covariate W , β and θ are the unknown regression parameter vectors and the error ϵ follows a continuous distribution and is independent with the covariates (X, W) . The model (1) reduces to the parametric single-index model in the absence of the covariate W and to the general parametric model in the absence of the covariate $\beta^\top X$. This structure is often meaningful, as in many applications, p_1 is large while p_2 is not. See the relevant dimension reduction literature, such as Feng et al. (2013).

However, it is less clear whether a real data set fits the above statistical formalisation. It is worthwhile performing suitable and efficient model checking before any further statistical analysis. As we often have no idea about the model structure under the alternative hypothesis, the general alternative model is considered in the following form:

Electronic supplementary material The online version of this article (doi:10.1007/s11222-016-9680-z) contains supplementary material, which is available to authorized users.

✉ Lixing Zhu
lzhu@hkbu.edu.hk
Xuehu Zhu
zhuxuehu0320@163.com
Xu Guo
liushengjunyi@163.com

- ¹ School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China
- ² Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong
- ³ School of Statistics, Beijing Normal University, Beijing, China

$$Y = g(X, W) + \epsilon, \quad (2)$$

where $g(\cdot)$ donates an unknown smooth function.

Several methods for testing the parametric single-index model that removes the covariate W from the model (1), and the general nonlinear model in the absence of the covariate $\beta^\top X$ can be found in the literature. Two prevalent classes of method are locally and globally smoothing tests. A locally smoothing test involves a nonparametric smoothing technique in the estimation, whereas a globally smoothing test only requires a set of sample averages with respect to an index set to form an empirical process or an average over the set of sample averages. For examples, [Härdle and Mammen \(1993\)](#) suggested a locally smoothing test based on the L_2 distance between the parametric and nonparametric estimate of the conditional expectation of Y given (X, W) in our notation. [Zheng \(1996\)](#) and [Fan and Li \(1996\)](#) independently developed tests based on second order conditional moments. [Dette \(1999\)](#) proposed a consistent test that depended on the difference between the variance estimate under the null and alternative hypotheses. [Fan et al. \(2001\)](#) developed a generalised likelihood ratio test. For other developments, see the Neyman threshold test ([Fan and Huang 2001](#)), a class of minimum distance tests ([Koul and Ni 2004](#)) and the distribution distance test ([Keilegom et al. 2008](#)). [González-Manteiga and Crujeiras \(2013\)](#) is a comprehensive review. However, locally smoothing tests have two obvious shortcomings. First, those methodologies have the subjective constraint choice of tuning parameters such as bandwidth. Unlike estimation, finding an optimal bandwidth choice for hypothesis testing is still an open problem ([Stute and Zhu 2005](#)). Although practical evidence suggests that this issue is not critical when the number p of covariates is small, a proper choice is not easy at all when p is large, even moderate. This problem often results in poor performance on the type I error control. A more serious problem is the typical slow convergence rates of locally smoothing tests, that is $O(n^{-1/2}h^{-p/4})$ under the null hypothesis, where h is the bandwidth tending to zero. In the present setup, $p = p_1 + p_2$. In other words, locally smoothing tests suffer severely from the curse of dimensionality.

For globally smoothing tests, examples include [Bierens \(1990\)](#), [Stute \(1997\)](#) and [Khmaldze and Koul \(2004\)](#). [Stute et al. \(1998\)](#) used bootstrap approximation to determine the critical values of the residual-marked empirical process-based test. Resampling approximation is particularly required when p is larger than 2 as its limiting null distribution is intractable. [Escanciano \(2007\)](#) is also a relevant reference in this class of tests. The typical convergence rate of globally smoothing tests is $O(n^{-1/2})$. Thus, they have the theoretical advantages over locally smoothing tests. However, the data sparseness in high-dimensional space means that most globally smoothing tests suffer from the dimension-

ality problem, even for large sample sizes (see [Escanciano 2006](#)). Practical evidence shows that the power of globally smoothing tests deteriorates and maintaining the significance level becomes more difficult when the dimension p of X is large, or even moderate. This is particularly the case when the alternative model is high-frequency.

A direct way to alleviate this problem is to project the high-dimensional covariates onto one-dimensional spaces first, and a test can be an average of tests that are based on the projections. This is a typical method called the projection-pursuit. [Huber \(1985\)](#) is a comprehensive reference. [Zhu and Li \(1998\)](#) suggested using the projection pursuit technique to define a test based on an unweighted integral of expectations with respect to all one-dimensional directions. [Zhu and An \(1992\)](#) has already used this idea to deal with a relevant testing problem. [Lavergne and Patilea \(2008\)](#) adopted this idea and further developed a dimension-reduction nonparametric method by exploring an optimal direction. [Lavergne and Patilea \(2012\)](#) advised a smooth version of the integrated conditional moment test over all projection directions. All of these tests partly overcome the curse of dimensionality with use of one-dimensional projections. However, the computational burden is a serious issue. Computing the values of the test statistics is very time-consuming, and becomes even more serious if we further need to use re-sampling approximation such as the bootstrap to determine critical values. Based on our very limited numerical studies, which we do not report in this paper, the CPU time of computing such tests is more than 100 times of computing the test statistic developed in the present paper, even when p is only 4. [Wong et al. \(1995\)](#) discussed the relevant computational issue that involved the integral over all projection directions in a test statistic and suggested a number-theoretical method to reduce the computational workload. [Xia \(2009\)](#) also constructed a test that involved searching for an optimal direction, but the test had no way of controlling type I error.

[Stute and Zhu \(2002\)](#) considered a naive method to handle the curse of dimensionality when testing the parametric single-index model: $Y = G(\beta^\top X) + \epsilon$. Stemming from the fact that under the null hypothesis, $E\{[Y - G(\beta^\top X)]I(X \leq t)\} = 0$ for all $t \in R^p$ leads to $E\{[Y - G(\beta^\top X)]I(\beta^\top X \leq t)\} = 0$ for all $t \in R$, the test statistic is based on the empirical process:

$$R_n(x) = n^{-1/2} \sum_{i=1}^n \left\{ y_i - G(\hat{\beta}^\top x_i) \right\} I(\hat{\beta}^\top x_i \leq x),$$

where $\hat{\beta}$ is, under the null hypothesis, a root- n consistent estimate of β . It has been proven to be powerful in many cases. However, this test is a directional test rather than an omnibus test. Thus, the general alternative of (2) cannot be detected. This phenomenon can be easily illustrated by the

following alternative model: $Y = \beta_1^\top X + c \sin(\beta_2^\top X) + \epsilon$, where X is normally distributed $N(0, I_p)$ with a $p \times p$ identity matrix I_p , and β_1 and β_2 are two orthogonal vectors. The value $c = 0$ corresponds to the null hypothesis. However, for any c , $E(Y - \beta_1^\top X | \beta_1^\top X) = 0$. In other words, this conditional mean cannot distinguish between models under the null and alternative hypotheses.

However, the advantage of the SZ's test (Stute and Zhu 2002) under the null hypothesis is very important particularly in high-dimensional paradigms as under the null hypothesis it solely uses the dimension reduction structure and thus it helps well maintain the significance level. Guo et al. (2015) recently proposed an adaptive-to-model dimension-reduction test that can be used to test for the model $Y = G(\beta^\top X, \theta) + \epsilon$ against the general alternative model $Y = g(X) + \epsilon$. The main idea is to fully utilize the dimension reduction structure about X under the null hypothesis as Stute and Zhu (2002) did, but to adapt the alternative model such that the test is still omnibus. Their test is based on a locally smoothing technique. The improvement over existing locally smoothing tests is significant. The test has a much faster convergence rate of $O(n^{-1/2}h^{-1/4})$ than the typical rate of $O(n^{-1/2}h^{-p/4})$ and can detect local alternatives distinct from the null hypothesis at the rate of $O(n^{-1/2}h^{-1/4})$ that is also much faster than the typical rate of $O(n^{-1/2}h^{-p/4})$ that locally smoothing tests can achieve. In other words, asymptotically, the test works as if X was univariate. Thus, the test can significantly avoid the curse of dimensionality. The numerical studies in their paper also indicated its advantages in cases with moderate sample size.

To facilitate more general alternative models, a key idea in their methodology development is to treat the purely non-parametric regression model (2) as a special multi-index regression model as follows:

$$Y = g(B^\top X, W) + \epsilon, \quad (3)$$

where B is a $p_1 \times q$ matrix with q orthogonal columns for an unknown number q with $1 \leq q \leq p_1$ and $g(\cdot)$ is still an unknown smooth function. We assume that the matrix B satisfies $B^\top B = I_q$ for identifiability. This model covers many popular models in the literature, such as the single-index models with $B = \beta$, the multi-index models with the absence of W , and partial single-index models with the mean function $g_1(\beta^\top X) + g_2(W)$. Here β is considered to be a column of B . When $q = p_1$ and $B = I_{p_1}$, the model (3) is reduced to the usual nonparametric alternative model (2). This is still true even when B is not equal to I_{p_1} , but the rank is equal to p_1 . This is because when $q = p_1$, $g(X, W) = g(BB^\top X, W) \equiv: \tilde{g}(B^\top X, W)$, where B is any $p_1 \times p_1$ orthonormal matrix. This persuasively demonstrates that the model (2) can be treated as a special case of (3). Based on this, a test can be constructed by noticing that under the null

hypothesis, $E\{Y - g(\beta^\top X, W, \theta)I(B^\top X \leq t, W \leq \omega)\} = 0$ for all (t, ω) and under the alternative hypothesis, it is nonzero for some vector (t, ω) .

To define an empirical version of this function as the basis for constructing a test statistic, an adaptive estimate of B is crucial for ensuring the test to have the adaptive-to-model property. That is, we wish an estimate of B to be consistent to $\kappa\beta$ for a constant κ under the null and to B under the alternative. Then, under the null hypothesis, the test can only rely on the dimension-reduced covariates $(\beta^\top X, W)$, and is still omnibus to detect the general alternative model (3). As mentioned above, when W is absent, the test in Guo et al. (2015) has the adaptiveness property to the alternative model. To identify B and its structural dimension, various dimension reduction approaches such as minimum average variance estimation (MAVE, Xia et al. 2002) and discretization-expectation estimation (DEE, Zhu et al. 2010) have been suggested. However, when W is present, these methods fail to work. Furthermore, due to the existence of W , even when the dimension $p_1 = 1$, the corresponding locally smoothing test still has a slow convergence rate in the order of $O(n^{-1/2}h^{-(p_2+1)/4})$ where p_2 is the dimension of W .

In this paper, we consider a globally smoothing test that keeps the advantage of SZ's test, fully uses the dimension reduction structure and utilises an adaptive-to-model strategy to get the test omnibus. As mentioned above, the key is to identify B adaptive to the null and alternative hypothesis. To be precise, under the null hypothesis, B is identified to be a vector proportional to β to make the test dimension-reduced. While, under the alternative hypothesis, B is adaptively identified such that the constructed test is still omnibus. To this end, the partial sufficient dimension reduction approach (e.g. Chiaromonte et al. 2002; Feng et al. 2013) has to be applied. To achieve the above goal, we also need to identify and estimate the structural dimension q of B . Under the null hypothesis, $q = 1$ is automatically identified and estimated. We then suggest a ridge-type eigenvalue ratio estimate. The details are presented in the next section. Another issue is critical value determination. In the present setting, the limiting null distribution is intractable, as it is for all globally smoothing tests. A resampling approximation is required. We then propose a Monte Carlo approximation that also fully utilises the information in the hypothetical model so that the approximation can be as close to the sampling null distribution as possible.

The rest of the paper is organised as follows. In Sect. 2, a dimension-reduction method, the partial discretization-expectation estimation, is reviewed, and is then used to identify or estimate B . The ridge-type eigenvalue ratio is also defined and its asymptotic properties are investigated in this section. Based on these, a test is constructed in Sect. 3. The asymptotic properties under the null and local alternative

hypotheses are also presented in this section. As the limiting null distribution is intractable, the Monte Carlo test approximation is described in Sect. 4. In Sect. 5, the simulation results are reported and a real data analysis is conducted for illustration. Section 6 is a discussion section and also contains a remark about a limitation of the proposed test due to the inconsistency of the structure dimension estimation under the local alternatives. Regularity conditions and technical proofs are found in the online supplementary material.

2 Partial discretization-expectation estimation and structural dimension estimation

2.1 A brief review on partial discretization-expectation estimation

As discussed above, identifying or estimating B is important for constructing an adaptive test. To this end, sufficient dimension reduction techniques can be applied. According to the sufficient dimension reduction theories, we can identify the space spanned by the columns of the matrix B (see, Chiaromonte et al. 2002). Write \tilde{B} as the $p \times q$ matrix consisting of these q basis vectors. We call \tilde{B} the basis matrix. Note that B is also a basis matrix of the space. Thus it is easy to see that for a $q \times q$ nonsingular matrix C , $\tilde{B} = B \times C^\top$. When $q = 1$, C is a constant and thus \tilde{B} is a vector proportional to the vector β under the null hypothesis. In Sect. 3 we show that identifying \tilde{B} is enough for the testing problem described herein.

In this subsection, we focus on identifying a basis matrix \tilde{B} . Without confusion, we still write it as B . This is equivalent to identify the space spanned by the columns of the matrix B which is called the partial central subspace (see, Chiaromonte et al. 2002), and is written as $S_{Y|X}^{(W)}$. By their definition, it is the intersection of all subspaces S such that

$$Y \perp\!\!\!\perp X | (P_S X, W),$$

where $\perp\!\!\!\perp$ stands for ‘independent of’ and $P_{(\cdot)}$ indicates a projection operator with respect to the standard inner product. $\dim\{S_{Y|X}^{(W)}\}$ is called the structural dimension of $S_{Y|X}^{(W)}$. In our setup, the structural dimension is 1 under the null model and q under the alternative hypothesis. Chiaromonte et al. (2002) and Wen and Cook (2007) developed estimation methods for $S_{Y|X}^{(W)}$ when W is discrete. Li et al. (2010) proposed groupwise dimension reduction (GDR), which can also deal with this case. Feng et al. (2013) proposed partial discretization-expectation estimation (PDEE) by extending discretization-expectation estimation (DEE) in Zhu et al. (2010). All of those estimations use the root- n consistency with the partial central subspace. In this paper, we adopt PDEE because PDEE is computationally inexpensive, and

can be easily used to determine the structural dimension q . Also, when W is absent, PDEE can naturally reduce to DEE without any changes in the algorithm.

From Feng et al. (2013), the following are the basic estimation steps.

1. Discretise the covariate $W = (W_1, \dots, W_{p_2})$ into a set of binary variables by defining $W(\mathbf{t}) = (I\{W_1 \leq t_1\}, \dots, I\{W_{p_2} \leq t_{p_2}\})$ where $\mathbf{t} = (t_1, \dots, t_{p_2})$ and the indicator functions $I\{W_i \leq t_i\}$ take value 1 if $W_i \leq t_i$ and 0 otherwise, for $i = 1, \dots, p_2$.
2. Let $S_{Y|X}^{(W(\mathbf{t}))}$ denote the partial central subspace of $Y|X$, $W(\mathbf{t})$, and $M(\mathbf{t})$ be a $p_1 \times p_1$ positive semi-definite matrix satisfying $\text{Span}\{M(\mathbf{t})\} = S_{Y|X}^{(W(\mathbf{t}))}$.
3. Let $T = \tilde{W}$ where \tilde{W} is an independent copy of W . The target matrix is $M = E\{M(\tilde{W})\}$. B consists of the eigenvectors that are associated with the nonzero eigenvalues of $M = E\{M(\tilde{W})\}$.
4. Let w_1, \dots, w_n be the n observations of W . Define an estimate of M as

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

where $M_n(w_i)$ is the partial sliced inverse regression matrix estimate defined in Chiaromonte et al. (2002) with sliced inverse regression proposed by Li (1991). Then when q is given, an estimate $B_n(q)$ of B consists of the eigenvectors that are associated with the q largest eigenvalues of M_n . $B_n(q)$ can be root- n consistent for B . For more details, readers may refer to Feng et al. (2013).

2.2 Structural dimension estimation

The structural dimension q is unknown in general. Interestingly, even when it is given, we still want to estimate adaptively according to its values under the null and alternative hypotheses because of its importance for the adaptive-to-model construction for the test. To estimate q , Feng et al. (2013) advised the BIC-type criterion that is an extension of that in Zhu et al. (2006). However, all practical uses show that selecting a proper penalty is not easy. In this paper, we suggest a ridge-type eigenvalue ratio estimate to determine q as:

$$\hat{q} = \arg \min_{1 \leq j \leq p} \left\{ \frac{\hat{\lambda}_{j+1}^2 + c_n}{\hat{\lambda}_j^2 + c_n} \right\}, \quad (4)$$

where $\hat{\lambda}_p \leq \dots \leq \hat{\lambda}_1$ are the eigenvalues of the matrix M_n . This method is motivated by Xia et al. (2015). The basic idea is as follows. Let λ_j be the eigenvalues of the target matrix M . When $j \leq q$, the eigenvalue $\lambda_j > 0$ and thus, the ratio

$r_{j-1} = \lambda_j/\lambda_{j-1} > 0$; when $j > q$ $\lambda_j = 0$. Therefore, $r_q = \lambda_{q+1}/\lambda_q = 0$; and $\lambda_{j+1}/\lambda_j = 0/0$. To define all ratios well, we can add a ridge in the ratio as $r_j = (\lambda_{j+1} + c_n)/(\lambda_j + c_n)$ for $1 \leq j \leq p - 1$. As $\hat{\lambda}_j^2$ converges to λ_j^2 at the rate of order $1/\sqrt{n}$ for $1 \leq j \leq q$, and to 0 at the rate of order $1/n$ for $q + 1 \leq j \leq p$, then $c_n = \log n/n$ can be a good choice. Compared with the BIC criterion in Feng et al. (2013) that requires to choose a suitable penalty, we also need to choose a good ridge constant c_n . But we do find that the estimation can be much more stable against the ridge selection than the penalty selection in the BIC criterion. As this is beyond the score of this paper, we do not give the detail in this paper. The algorithm is very easy to implement and the estimation consistency can be guaranteed. The result is stated in the following.

Theorem 1 Under Conditions A1 and A2 in the Appendix, the estimate \hat{q} of (4) with $c_n = \log n/n$ has the following consistency:

- (i) under the null hypothesis (1), $P(\hat{q} = 1) \rightarrow 1$;
- (ii) under the alternative hypothesis (3), $P(\hat{q} = q) \rightarrow 1$.

From our justification presented in the Appendix, the choice of c_n can be in a relatively wide range to ensure consistency under the null and alternative hypotheses. However, to avoid the arbitrariness of its choice, we find that $c_n = \log n/n$ is a proper choice. The above identification of q is very important for ensuring that the test statistic is adaptive to the underlying models. Finally, an estimate of B is $B_n = B_n(\hat{q})$. This estimate is used in the following test statistic construction.

3 A partial dimension reduction adaptive-to-model test and its properties

3.1 Test statistic construction

The hypotheses of interest can now be restated. The null hypothesis is

$$H_0: E(Y|X, W) = G(\beta^\top X, W, \theta) \text{ for some } \beta \in \mathbb{R}^{p_1}, \theta \in \Theta \in \mathbb{R}^d,$$

against the alternative hypothesis: for any β and θ

$$H_1: E(Y|X, W) = g(B^\top X, W) \neq G(\beta^\top X, W, \theta).$$

In this subsection, let $\epsilon = Y - G(\beta^\top X, W, \theta)$ denote the error term under the null hypothesis. Under H_0 , $q = 1$, and $B = \kappa\beta$ for some constant κ , then we have:

$$E(\epsilon|X, W) = 0 \Leftrightarrow E(\epsilon|B^\top X, W) = 0 \\ \Leftrightarrow E[\epsilon I\{(B^\top X, W) \leq (u, \omega)\}] = 0$$

for all (u, ω) . Under H_1 , $E\{Y - G(\beta^\top X, W, \theta)|X, W\} = g(B^\top X, W) - G(\beta^\top X, W, \theta) \neq 0$, we then have:

$$E\{Y - G(\beta^\top X, W, \theta)|X, W\} \neq 0 \\ \Leftrightarrow E\{Y - G(\beta^\top X, W, \theta)|B^\top X, W\} \neq 0.$$

Before proceeding to the test statistic construction, recall that what we can identify is $\tilde{B} = B \times C$ for a $q \times q$ orthogonal matrix C . Thus, we need to make sure this non-identifiability does not affect the equivalence between $E\{Y - G(\beta^\top X, W, \theta)|\tilde{B}^\top X, W\} \neq 0$ and $E\{Y - G(\beta^\top X, W, \theta)|B^\top X, W\} \neq 0$. This is easy to check. Note that $\tilde{B} = B \times C^\top$ with C being a non-singular matrix and thus B and \tilde{B} map one-to-one. Then

$$E\{Y - G(\beta^\top X, W, \theta)|X, W\} = E\{g(B^\top X, W) - G(\beta^\top X, W, \theta)|X, W\} \\ = E\{\tilde{g}(\tilde{B}^\top X, W) - G(\beta^\top X, W, \theta)|X, W\},$$

where $\tilde{g}(\cdot, \cdot) = g(C^{-1}\cdot, \cdot)$. It is equivalent between $E\{Y - G(\beta^\top X, W, \theta)|B^\top X, W\} \neq 0$ and $E\{Y - G(\beta^\top X, W, \theta)|\tilde{B}^\top X, W\} \neq 0$. Therefore, identifying B itself is not necessary. As mentioned, we simply write \tilde{B} as B .

Now we are in the position to define a residual-marked empirical process. Let

$$V_n(u, \omega) = n^{-1/2} \sum_{i=1}^n \{y_i - G(\beta_n^\top x_i, w_i, \theta_n)\} \\ \times I\{(B_n(\hat{q}))^\top x_i, w_i\} \leq (u, \omega)\}, \tag{5}$$

where β_n and θ_n are the nonlinear least squares estimates respectively, and $B_n(\hat{q})$ was defined before.

Therefore, we use V_n as the basis for constructing a test statistic:

$$T_n = \int V_n^2(B_n(\hat{q}))^\top x, \omega) dF_n(B_n(\hat{q}))^\top x, \omega), \tag{6}$$

where $F_n(\cdot)$ denotes the empirical distribution based on the samples $\{B_n(\hat{q})^\top x_i, w_i\}_{i=1}^n$. Therefore, the null hypothesis is rejected for large values of T_n .

It is clear that this test statistic is not scale-invariant and thus usually a normalizing constant is required. This constant needs to be estimated which involves many unknowns. In this paper, a Monte Carlo test procedure is recommended which can automatically make the test scale-invariant so that normalisation is not necessary. Additionally, it can mimic

the sampling null distribution better than existing approximations such as that in [Stute et al. \(1998\)](#). The details can be found in [Sect. 4](#).

3.2 Limiting null distribution

To study the properties of the process $V_n(\cdot, \cdot)$ and the test statistic T_n , here we define a process for the purpose of theoretical investigation: for u and ω ,

$$V_n^0(u, \omega) = n^{-1/2} \sum_{i=1}^n \left\{ y_i - G(\beta^\top x_i, w_i, \theta) I\{(B^\top x_i, w_i) \leq (u, \omega)\} \right\}. \tag{7}$$

When $E(Y^2) < \infty$, take the conditional variance of Y given $B^\top X = u$ and $W = \omega$,

$$\sigma^2(u, \omega) = \text{Var}(Y | B^\top X = u, W = \omega),$$

and put

$$\psi(u, \omega) = \int_{-\infty}^{\omega} \int_{-\infty}^u \sigma^2(v_1, v_2) dF_{B^\top X, W}(v_1, v_2),$$

where $F_{B^\top X, W}(\cdot, \cdot)$ denotes the distribution function of $(B^\top X, W)$. It is easy to see that under H_0 ,

$$\text{Cov}\{V_n^0(u_1, \omega_1), V_n^0(u_2, \omega_2)\} = \psi(u_1 \wedge u_2, \omega_1 \wedge \omega_2).$$

By [Theorem 1.1](#) in [Stute \(1997\)](#), we can assert that under H_0 :

$$V_n^0 \longrightarrow V_\infty \text{ in distribution,} \tag{8}$$

where V_∞ is a continuous Gaussian process with mean zero and covariance kernel as follows:

$$K\{(u_1, \omega_1), (u_2, \omega_2)\} = \psi(u_1 \wedge u_2, \omega_1 \wedge \omega_2).$$

Theorem 2 *Under H_0 and the regularity conditions A1–A4 in the Appendix, we have the distribution*

$$V_n \longrightarrow V_\infty - G^\top V \equiv V_\infty^1,$$

where V_∞ is the Gaussian process defined in [\(8\)](#) and the vector-valued function $G^\top = (G_1, G_2, \dots, G_{p+d})$ is defined as

$$G_i(u, \omega) = E \left[m_i(X, W, \beta, \theta) I\{(B^\top X, W) \leq (u, \omega)\} \right],$$

where $B = \kappa\beta$ and V is a $(p_1 + d)$ -dimensional normal vector with mean zero and covariance matrix $L(\beta, \theta)$ which is defined in the Appendix.

Remark 1 From this theorem, we can see that the test statistic has the same convergence rate of order $n^{-1/2}$ to its limit as that of existing globally smoothing tests. In other words, in an asymptotic sense, there is no room for globally smoothing tests to improve their convergence rate. Locally and globally smoothing tests differ in this feature, as $n^{-1/2}h^{p/4}$ can be much improved ([Guo et al. 2015](#)). However, as in [Stute and Zhu \(2002\)](#), the new test can largely avoid the effect of dimensionality such that the test is powerful for relatively large p . The simulations in [Sect. 5](#) demonstrate this.

3.3 Power study

First, we present the asymptotic property under the global alternative hypothesis.

Theorem 3 *Under Conditions A1, A2, A3 and A4 and H_{1n} with $C_n = c$ a fixed constant, we have in probability*

$$n^{-1/2} V_n(u, \omega) \longrightarrow E[\{g(B^\top X, W) - G(\tilde{\beta}^\top X, W, \tilde{\theta})\} I\{(B^\top X, W) \leq (u, \omega)\}]$$

where $(\tilde{\beta}, \tilde{\theta})$ may be different from the true value (β, θ) under the null hypothesis. Then $T_n \rightarrow \infty$ in probability.

To study how sensitive our new method is to the alternative hypotheses, consider the following sequence of local alternative hypotheses:

$$H_{1n}: Y = G(\beta^\top X, W, \theta) + C_n g(B^\top X, W) + \varepsilon, \tag{9}$$

where C_n goes to zero.

Under the local alternatives with $C_n \rightarrow 0$, we also need to estimate the structural dimension q . Recall that under the global alternative hypothesis in [Sect. 2](#), the estimate $\hat{q} = q$ had a probability going to zero, which could be larger than 1 when B contains more than one basis vector. However, under the above local alternative models, when C_n goes to zero, the models converge to the hypothetical model that has one vector β . Thus, we anticipate that \hat{q} also converges to 1 under the local alternative hypotheses. The following lemma confirms this.

Lemma 1 *Under H_{1n} in [\(9\)](#), $C_n = n^{-1/2}$ and the regularity conditions in [Theorem 2](#), the estimate \hat{q} by [\(4\)](#) satisfies that as $n \rightarrow \infty$, $P(\hat{q} = 1) \rightarrow 1$.*

To further study the power performance of the test, assume an additional regularity Condition A5 in the Appendix.

Theorem 4 *Under H_{1n} and Conditions A1, A2, A4 and A5, when $C_n = n^{-1/2}$, we have in distribution*

$$V_n(u, \omega) \longrightarrow V_\infty(u, \omega) + E[g(B^\top X, W) I\{(\kappa\beta^\top X, W) \leq (u, \omega)\}] + G^\top(\eta - V)(u, \omega)$$

where V_∞ , G and V are defined as those in Theorem 2 and η is a $(p_1 + d)$ -dimensional constant vector, which are defined in Appendix. Then T_n has a finite limit.

Remark 2 This theorem shows that under the local alternative models, the test would also be directional, because \hat{q} is not a consistent estimate of q . This is caused by the difficulty of estimating q when the alternative models are too close to the null model. If the estimation of q could be improved, it is likely that the omnibus property would still hold under the local alternative hypotheses. We discuss this further in Sect. 6.

4 A Monte-Carlo test procedure

As the limiting null distribution of the test statistic T_n is not tractable, the nonparametric Monte Carlo test procedure is suggested to approximate the sampling null distribution, which is similar in spirit to the wild bootstrap, see Stute et al. (1998) and Zhu and Neuhaus (2000). However, to enhance the power of the test, we have a modified version that fully uses the model structure under the null model.

A magical algorithm is developed to determine the p values as follows:

Step 1 Generate a sequence of i.i.d variables $\mathbf{U} = \{U_i\}_{i=1}^n$ from the standard normal distribution $N(0, 1)$. Then construct the following process:

$$\Delta_n(u, \omega, \mathbf{U}) = n^{-1/2} \sum_{i=1}^n \hat{\varrho}(x_i, w_i, y_i, \beta, \theta) U_i,$$

where $\hat{\varrho}(x_i, w_i, y_i, \beta, \theta)$ is an estimate of $\varrho(x_i, w_i, y_i, \beta, \theta)$ with the following notations:

$$\varrho(x_i, w_i, y_i, \beta, \theta) = \epsilon_i I\{(B_1^\top x_i, w_i) \leq (u, \omega)\} - G^\top v_i,$$

$$\hat{\varrho}(x_i, w_i, y_i, \beta, \theta) = \hat{\epsilon}_i I\{(B_{1n}^\top x_i, w_i) \leq (u, \omega)\} - \hat{G}^\top \hat{v}_i,$$

$$G(u, \omega) = E \left[m(X, W, \beta, \theta) I\{(B_1^\top X, W) \leq (u, \omega)\} \right],$$

$$\hat{G} = n^{-1} \sum_{i=1}^n m(x_i, w_i, \beta_n, \theta_n) I\{(B_{1n}^\top X, W) \leq (u, \omega)\},$$

$$v_i = l(x_i, w_i, y_i, \beta, \theta), \quad \hat{v}_i = l(x_i, w_i, y_i, \beta_n, \theta_n),$$

$$\hat{\epsilon}_i = y_i - G(\beta_n^\top x_i, w_i, \theta_n),$$

with $l(\cdot)$ being a $(p_1 + d)$ -dimensional vector function defined in Appendix, and B_1 and B_{1n} being respectively the first column vectors of B and $B_n(\hat{q})$. The resulting Monte Carlo test statistic is

$$\tilde{T}_n(\mathbf{U}) = \int \Delta_n^2(B_{1n}^\top x, w, \mathbf{U}) dF_{B_{1n}}(x, w),$$

where $F_{B_{1n}}(\cdot)$ denotes the empirical distribution based on the samples $\{B_{1n}^\top x_i, w_i\}_{i=1}^n$.

Step 2 Generate m sets of $\mathbf{U}, \mathbf{U}_j, j = 1, \dots, m$, and get m values of $\tilde{T}_n(\mathbf{U})$, say $\tilde{T}_n(\mathbf{U}_j), j = 1, \dots, m$.

Step 3 The p value is estimated by

$$\hat{p} = m^{-1} \sum_{j=1}^m I\{\tilde{T}_n(\mathbf{U}_j) \geq T_n\},$$

where T_n is defined in (6). Whenever $\hat{p} \leq \alpha$, reject H_0 , for a given significance level α , or the critical value is determined as the $(1 - \alpha)100\%$ upper percentile of all \mathbf{U}_j 's.

As mentioned before, this test procedure is scale-invariant although T_n is not, because the resampling procedure does not need to involve test statistic normalisation and $\hat{p} = m^{-1} \sum_{j=1}^m I\{\tilde{T}_n(\mathbf{U}_j) \geq T_n\} = m^{-1} \sum_{j=1}^m I\{\tilde{T}_n(\mathbf{U}_j)/c \geq T_n/c\}$ for any $c > 0$.

Remark 3 It is worth pointing out that the algorithm is different from traditional wild bootstrap or nonparametric Monte Carlo test procedures that use the vector $B(\hat{q})_n^\top X$. More details can be found in Stute et al. (1998) and Zhu (2005). When we only use the vector B_{1n} , which is associated with the largest eigenvalue of the target matrix M_n defined in Sect. 2, we only use univariate covariate $B_{1n}^\top X$, which is $\beta^\top X$ under the null asymptotically. This makes the approximation as close to the sampling null distribution as possible.

The following theorem states the consistency of the conditional distribution approximation even under local alternative hypotheses.

Theorem 5 Under the conditions in Theorem 2 and the null hypothesis or the local alternative hypotheses with $C_n = n^{-1/2}$, we know that for almost all sequences $\{(y_1, x_1, w_1), \dots, (y_n, x_n, w_n), \dots\}$, the conditional distribution of $\tilde{T}_n(\mathbf{U})$ converges to the limiting null distribution of T_n .

5 Numerical studies

5.1 Simulations

In this subsection, we conduct simulations to examine the finite-sample performance of the proposed test. The simulations are based on 2000 Monte Carlo test replications to compute the critical values or p values. Each experiment is then repeated 1000 times to compute the empirical sizes and powers at the significance level $\alpha = 0.05$. To estimate the central subspace spanned by B , we use the SIR-based

PDEE/DEE procedure according to the cases with and without the variate W in the model. In these two cases, we call the test T_n^{PDEE} .

We choose the ZH’s test (Zheng 1996) and the SZ’s test (Stute and Zhu 2002) as the representatives of local and globally smoothing tests, respectively, to compare with our test. We choose these tests because (1) the ZH’s test has the explicitly and tractable limiting null distribution that can be used to determine the critical values; (2) like other locally smoothing tests, the re-sampling version helps improve its performance (we then also include the re-sampling version of the ZH’s test); and (3) the SZ’s test is asymptotically distribution-free and powerful in many situations, but is not an omnibus test. We will demonstrate this. Further, we also compare our test to The GWZ’s test (Guo et al. 2015), because it is based on the ZH’s test but also has the adaptive-to-model property, and can be much more powerful. We respectively write the proposed test, the ZH’s, the SZ’s and the GWZ’s tests as T_n^{PDEE} , T_n^{ZH} , T_n^{SZ} and T_n^{GWZ} .

In this section, we first design four examples to examine the performance in four scenarios without the random variable W . The first example has the same projection direction in both the hypothetical and alternative models. The second example is used to check the adaptiveness of our test to omnibus testing even when dimension reduction structure under the null hypothesis is fully adopted, showing that the SZ’s test is directional and thus has much less power. The third example is used to check the effect of dimensionality from X for locally smoothing tests, and to compare our test with the ZH’s and the GWZ’s test. The fourth example is used to assess the effect of correlations among the components of X . In the first three examples, the data (x_i, w_i) are generated from the multivariate standard normal distribution $N(0, I_p)$, independent of the standard normal errors ϵ_i .

Example 1 Consider the following regression model:

$$Y = \beta_0^\top X + a \times \cos(0.6\pi\beta_0^\top X) + 0.5 \times \epsilon \quad \text{and} \quad \beta_0 = (0, 0, 1, 1)/\sqrt{2}.$$

The values $a = 0, 0.2, 0.4, 0.6, 0.8, 1$ are used. The value $a = 0$ corresponds to the null hypothesis and $a \neq 0$ to the alternative hypothesis. The power function is plotted in Fig. 1. Some findings are as follows. The power increases reasonably with larger a . The proposed test T_n^{PDEE} is significantly and uniformly more powerful than T_n^{ZH} and T_n^{SZ} . When a is not large, T_n^{SZ} works better than T_n^{ZH} , and when a is large, T_n^{ZH} slightly outperforms T_n^{SZ} in power.

Example 2 To further check the omnibus property of the proposed test to detect general alternative models, a comparison with the SZ’s test and the ZH’s test is again carried out. In this

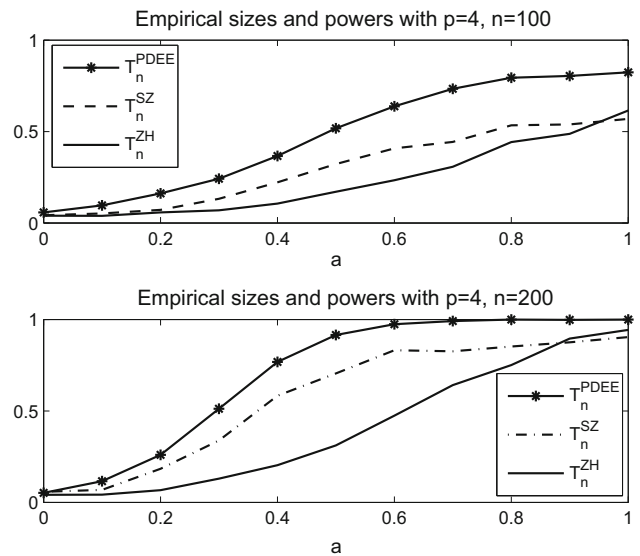


Fig. 1 The empirical size and power curves of T_n^{PDEE} , T_n^{SZ} and T_n^{ZH} in Example 1

example, we generate the data from the following regression model:

$$Y = \beta_0^\top X + a \times 0.125 \exp(0.3\beta_1^\top X) + 0.5 \times \epsilon;$$

where $\beta_0 = (1, 1, 0, 0)/\sqrt{2}$ and $\beta_1 = (0, 0, 1, 1)/\sqrt{2}$. The values $a = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$ are used. In this model, $B = (\beta_0^\top, \beta_1^\top)^\top$ and $\beta_0^\top X$ is orthogonal to the functions under the alternative hypotheses. We can see that the SZ’s test cannot detect such alternative hypotheses. The results are reported in Fig. 2. The results clearly show that the SZ’s test T_n^{SZ} and the ZH’s test T_n^{ZH} are not very sensitive to the alternative hypotheses. In particular, when the sample size is small ($n = 100$), the SZ’s test T_n^{SZ} has almost no power.

Example 3 To gain further insights into our test, we consider the effect of the dimensionality of X . When the number of dimensions is large, the ZH’s test does not maintain the significance level or power performance, due to slow convergence. Thus, the wild bootstrap is applied to approximate the sampling null distribution. The re-sampling time is 2000 in this simulation study. The bootstrap version is written as T_n^{ZHB} . the GWZ’s test is also compared.

Consider the models:

$$Y = \beta_0^\top X + a \times \{0.3(\beta_1^\top X)^3 + 0.3(\beta_1^\top X)^2\} + 0.5 \times \epsilon;$$

where $\beta_0 = (1, 1, 1, 1, 0, 0, 0, 0)/2$ and $\beta_1 = (0, 0, 0, 0, 1, 1, 1, 1)/2$. Then the dimension $p = 8$. The results are presented in Fig. 3. We first examine the significance level

maintenance of different methods. T_n^{ZH} does not maintain the significance level well, especially when $n = 50$, the empirical sizes of T_n^{ZH} is only 0.024 when the critical value is

determined by the limiting null distribution, but its bootstrap version T_n^{ZHB} can do. the test T_n^{GWZ} can also well maintain the significance level. T_n^{PDEE} performs uniformly the best among all competitors and the absolute differences between the significance level and empirical sizes are less than 0.005. From Fig. 3, we can clearly see that the test T_n^{GWZ} has advantage over its competitor T_n^{ZH} in gaining power, and T_n^{PDEE} uniformly works better than the others. The comparison between T_n^{GWZ} and T_n^{PDEE} further substantiates the theoretical advantage that the globally smoothing test is more sensitive to alternative hypotheses than the locally smoothing test when both are constructed via the dimension reduction technique. The results reported in Fig. 1 present that when the sample size is larger, the performance of all methods becomes better reasonably. Compared with the results in Figs. 1 and 2 with $p = 4$, the dimension p has little effect for T_n^{PDEE} . However, it has a very significant effect for T_n^{ZHB} and T_n^{ZH} . When the dimension of the covariates is higher, the performance of T_n^{ZH} and T_n^{ZHB} is worse. For space saving, We do not include the detail of the simulation results.

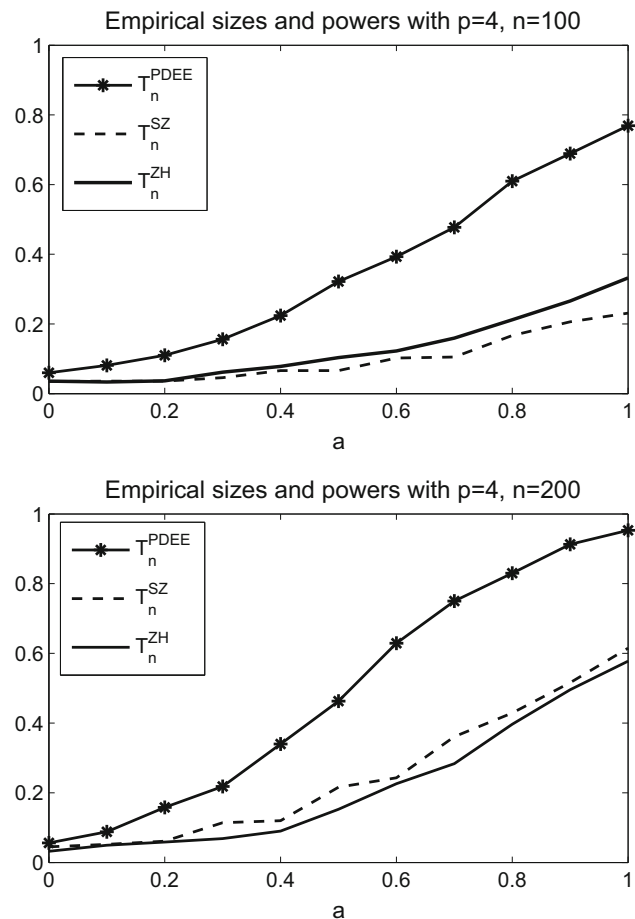


Fig. 2 The empirical size and power curves of T_n^{PDEE} , T_n^{SZ} and T_n^{ZH} in Example 2

Example 4 To further assess the performance of the test T_n^{PDEE} , we examine the effect of the correlated covariate X and the distribution of the error term ϵ . Consider the following model:

$$y = \beta_0^\top X + a \times \exp(-(\beta_0^\top X)^2/2)/2 + 0.5 \times \epsilon;$$

where X follows a normal distribution $N(0, \Sigma)$ with the covariance matrix $\Sigma_{ij} = I(i = j) + \rho^{|i-j|}I(i \neq j)$ for $\rho = 0.5, i, j = 1, 2, \dots, p, \beta_0 = (1, 1, -1, -1)/2$ and ϵ follows the student's t distribution with 4 degrees of freedom.

The results are listed in Table 1. Comparing the results in this table with those in Figs. 1 and 2, it is clear that with the correlated covariate X , we arrive at similar conclusions to

Fig. 3 The empirical size and power curves of T_n^{PDEE} , T_n^{ZHB} , T_n^{ZH} and T_n^{GWZ} in Example 3

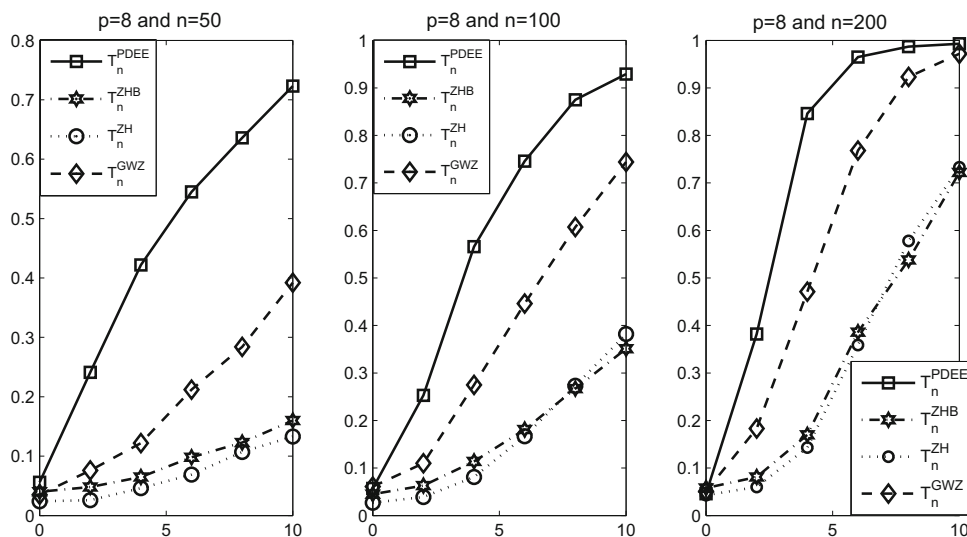


Table 1 Empirical sizes and powers of T_n^{PDEE} , T_n^{SZ} , T_n^{ZH} and T_n^{GWZ} for Example 4 with $p = 4$ and correlated covariates

n/a	T_n^{PDEE}			T_n^{SZ}			T_n^{ZH}			T_n^{GWZ}		
	50	100	200	50	100	200	50	100	200	50	100	200
0	0.062	0.057	0.052	0.031	0.039	0.050	0.038	0.039	0.041	0.051	0.053	0.045
0.2	0.102	0.167	0.207	0.052	0.060	0.139	0.055	0.069	0.082	0.081	0.111	0.167
0.4	0.235	0.416	0.587	0.092	0.156	0.406	0.088	0.149	0.272	0.153	0.298	0.522
0.6	0.431	0.660	0.885	0.156	0.378	0.726	0.215	0.374	0.595	0.298	0.611	0.852
0.8	0.582	0.854	0.978	0.268	0.542	0.918	0.369	0.586	0.885	0.541	0.828	0.969
1.0	0.696	0.951	0.996	0.292	0.730	0.980	0.530	0.783	0.964	0.704	0.961	0.999

those in Examples 1 and 2. T_n^{PDEE} easily maintains the significance level. We also find that when the structural dimension $q = 1$ under the alternative hypothesis, the power performance of T_n^{GWZ} is very similar to that of T_n^{PDEE} . Comparing Example 3 in Fig. 3 with Example 4 in Table 1, we can see that the lower structural dimension increases T_n^{GWZ} 's empirical power. This suggests that the structural dimension q still has a negative effect on T_n^{GWZ} although it is not very much. However, the power of T_n^{PDEE} does not deteriorate when the structural dimension is increased. Further, T_n^{PDEE} can control type I error very well and is significantly more powerful than the ZH's and the SZ's tests. It is evident that T_n^{PDEE} is robust to the error term.

In summary, the globally smoothing-based dimension reduction adaptive-to-model test inherits the advantages of classical globally smoothing tests and has the adaptive-to-model property when the dimension reduction structure is adopted.

Now we consider the parallel models in Examples 1, 2, 3 and 4 when the covariate W is included. However, we present only the results for T_n^{PDEE} because based on the results in the above examples and comparisons, the performance of the competitors is even worse when there are q_1 more dimensions in the model (meaning that q_1 more dimensions are added when W is q_1 -dimensional).

Example 5 The four models are:

Case 1 $Y = \beta_0^\top X + W + a \times \cos(0.6\pi\beta_0^\top X) + 0.5 \times \epsilon;$

Case 2 $Y = \beta_0^\top X + \sin(W) + a \times \{0.5(\beta_1^\top X)^2 + 2 \sin(W)\} + 0.5 \times \epsilon;$

Case 3 $Y = \beta_0^\top X + \cos(W) + a \times \{0.3(\beta_1^\top X)^3 + 0.3(\beta_1^\top X)^2\} + 0.5 \times \epsilon;$

Case 4 $y = \beta_0^\top X + \sin(W) + a \times \exp(-(\beta_0^\top X)^2/2) \times W + 0.5\epsilon.$

All of the settings are the same as the respective settings in Examples 1, 2, 3 and 4 except for the additional W following the normal distribution $N(0, 1)$. The results are reported in Table 2.

Table 2 Sizes and powers of T_n^{PDEE} for Example 5

a	$n = 50$	$n = 100$	$n = 200$	$n = 400$
Case 1 with $p = 4$ and $q = 1$				
0	0.068	0.059	0.056	0.050
0.2	0.172	0.172	0.294	0.642
0.4	0.262	0.512	0.874	1.000
0.6	0.508	0.895	1.000	1.000
0.8	0.652	0.964	1.000	1.000
1	0.702	0.984	1.000	1.000
Case 2 with $p = 4$ and $q = 2$				
0	0.058	0.055	0.053	0.048
0.2	0.069	0.096	0.139	0.376
0.4	0.110	0.246	0.691	0.997
0.6	0.212	0.573	0.985	1.000
0.8	0.304	0.866	1.000	1.000
1	0.421	0.938	1.000	1.000
Case 3 with $p = 8$ and $q = 2$				
0.2	0.133	0.182	0.296	0.550
0.4	0.245	0.415	0.719	0.952
0.6	0.357	0.597	0.911	0.998
0.8	0.421	0.724	0.951	1.000
1	0.501	0.802	0.978	1.000
Case 4 with $p = 4$ and $q = 1$				
0	0.060	0.045	0.047	0.051
0.2	0.102	0.154	0.270	0.496
0.4	0.217	0.427	0.744	0.982
0.6	0.405	0.724	0.973	1.000
0.8	0.541	0.897	0.999	1.000
1.0	0.646	0.956	0.999	1.000

The reported results clearly indicate that when W is presented, T_n^{PDEE} still works well in maintaining the significance level and detecting general alternative models.

5.2 Real data analysis

In this subsection, for illustration we perform the regression modelling of the well-known Boston Housing Data, initially studied by Harrison and Rubinfeld (1978). The

data set contains 506 observations and 14 variables, as follows: the median value of owner-occupied homes in \$1000's (*MEDV*), per capita crime rate by town (*CRIM*), proportion of residential land zoned for lots over 25,000 sq.ft. (*ZN*), proportion of non-retail business acres per town (*INDUS*), Charles River dummy variable (1 if tract bounds river; 0 otherwise) (*CHAS*), nitric oxides concentration (parts per 10 million) (*NOX*), average number of rooms per dwelling (*RM*), proportion of owner-occupied units built prior to 1940 (*AGE*), weighted distances to five Boston employment centres (*DIS*), index of accessibility to radial highways (*RAD*), full-value property-tax rate per 10,000 (*TAX*), pupil-teacher ratio by town (*PTRATIO*), the proportion of black people by town (*B*) and lower status of the population (*LSTAT*).

As suggested by Feng et al. (2013), we take the logarithm of (*MEDV*) as the response variable, the predictor *CRIM* as *W* and the other 11 predictors as *X*, except *CHAS*, because it has little influence on the housing price as advised by Wang et al. (2010), and is thus excluded from this data analysis. In this data analysis, we standardise the predictors for ease of explanation. Here a simple linear model is considered to be the hypothetical model. The SIR-based PDEE procedure is applied to determine the partial central subspace $S_{Y|X}^{(W)}$. A total of 2000 Monte Carlo test replications are implemented to compute the *p* value that is about 0. Hence, the null hypothesis is rejected. Moreover, \hat{q} is 2. Thus, a partial multi-index model would be plausible.

6 Discussions

In this paper, we propose an adaptive-to-model dimension reduction test based on a residual marked empirical process for partially parametric single-index models. The test can fully utilise the dimension reduction structure to reduce dimensionality problems, while remaining an omnibus test. Comparisons with existing local and globally smoothing tests suggest that (1) model-adaptation enhances the power performance, also maintaining the significance level; and (2) the globally smoothing-based adaptive-to-model test outperforms the locally smoothing-based adaptive-to-model test. Thus, a globally smoothing test is worthy of recommendation. This method can be readily applied to other models and problems when a dimension reduction structure is presented. The research is on-going.

In the hypothetical and alternative models, the independence between the error and the covariates is assumed. This condition is fairly strong. The condition can be weakened to handle the testing problem for the following hypothetical and alternative models:

$$Y = G(\beta^\top X, W, \theta) + \delta(\beta^\top X, W)\epsilon,$$

$$Y = g(B^\top X, W) + \delta(B^\top X, W)\epsilon.$$

Here, all of the settings are the same as those considered in the paper, except that the function $\delta(\cdot)$ is an unknown smooth function. $B_n(\hat{q})$ estimated by the SIR-based PDEE/DEE procedure is still a root-*n* consistent estimate of *B*. Thus, the proposed test can still be feasible.

From the asymptotic properties of the proposed test, we also find a limitation. Under the local alternatives that converge to the null hypothesis at a certain rate, the proposed test, unlike existing omnibus tests, cannot be powerful. This is because of the inconsistency of the estimator \hat{q} of *q* under the local alternatives with $C_n = 1/\sqrt{n}$. The method can only estimate *q* to be 1. Thus, the estimate \hat{B} converges to β , and when the other directions in *B* are orthogonal to β and the function has some special structure, our test may not have good power. However, this does not mean that our test cannot detect any local alternative models. When the convergence rate C_n becomes slower, the ridge-type eigenvalue ratio estimate can still well estimate *B* by choosing a suitable ridge value c_n and then local alternative hypotheses can be detected. Thus, to make the test omnibus under local alternatives, the key is to develop a good method to well estimate *q*. The research is ongoing.

Acknowledgements The authors are grateful to the editor, the associate editor and the anonymous referee for the constructive comments and suggestions that led a presentation improvement of an early manuscript. Lixing Zhu's research was supported by a grant from the University Grants Council of Hong Kong, Hong Kong, China. Xu Guo's research was supported by the Fundamental Research Funds for the Central Universities, No. NR2015001 and the Natural Science Foundation of Jiangsu Province, China, Grant No. BK20150732.

References

- Bierens, H.J.: A consistent conditional moment test of functional form. *Econometrica* **58**, 1443–1458 (1990)
- Chiaromonte, F., Cook, R.D., Li, B.: Sufficient dimension reduction in regressions with categorical predictors. *Ann. Stat.* **30**, 475–497 (2002)
- Dette, H.: A consistent test for the functional form of a regression based on a difference of variance estimates. *Ann. Stat.* **27**, 1012–1050 (1999)
- Escanciano, J.C.: A consistent diagnostic test for regression models using projections. *Econ. Theory* **22**, 1030–1051 (2006)
- Escanciano, J.C.: Model checks using residual marked empirical processes. *Stat. Sin.* **17**, 115–138 (2007)
- Fan, J., Zhang, C., Zhang, J.: Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Stat.* **29**, 153–193 (2001)
- Fan, J.Q., Huang, L.S.: Goodness-of-fit tests for parametric regression models. *J. Am. Stat. Assoc.* **96**, 640–652 (2001)
- Fan, Y., Li, Q.: Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica* **64**, 865–890 (1996)

- Feng, Z., Wen, X., Yu, Z., Zhu, L.X.: On partial sufficient dimension reduction with applications to partially linear multi-index models. *J. Am. Stat. Assoc.* **501**, 237–246 (2013)
- González-Manteiga, W., Crujeiras, R.M.: An updated review of Goodness-of-Fit tests for regression models. *Test* **22**, 361–411 (2013)
- Guo, X., Wang, T., Zhu, L. X.: Model checking for parametric single-index models: a dimension reduction model-adaptive approach. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* (2015). doi:[10.1111/rssb.12147](https://doi.org/10.1111/rssb.12147)
- Härdle, W., Mammen, E.: Comparing nonparametric versus parametric regression fits. *Ann. Stat.* **21**, 1926–1947 (1993)
- Harrison, D., Rubinfeld, D.L.: Hedonic prices and the demand for clean air. *J. Environ. Econ. Manag.* **5**, 81–102 (1978)
- Huber, P.J.: Projection pursuit. *Ann. Stat.* **13**, 435–475 (1985)
- Khmaladze, E.V., Koul, H.L.: Martingale transforms goodness-of-fit tests in regression models. *Ann. Stat.* **37**, 995–1034 (2004)
- Koul, H.L., Ni, P.P.: Minimum distance regression model checking. *J. Stat. Plan. Inference* **119**, 109–141 (2004)
- Lavergne, Q., Patilea, V.: Breaking the curse of dimensionality in non-parametric testing. *J. Econ.* **143**, 103–122 (2008)
- Lavergne, P., Patilea, V.: One for all and all for one: regression checks with many regressors. *J. Bus. Econ. Stat.* **30**, 41–52 (2012)
- Li, K.C.: Sliced inverse regression for dimension reduction. *J. Am. Stat. Assoc.* **86**, 316–327 (1991)
- Li, L., Li, B., Zhu, L.X.: Groupwise dimension reduction. *J. Am. Stat. Assoc.* **105**, 1188–1201 (2010)
- Stute, W.: Nonparametric model checks for regression. *Ann. Stat.* **25**, 613–641 (1997)
- Stute, W., González-Manteiga, W., Presedo-Quindimil, M.: Bootstrap approximation in model checks for regression. *J. Am. Stat. Assoc.* **93**, 141–149 (1998)
- Stute, W., Zhu, L.X.: Model checks for generalized linear models. *Scand. J. Stat.* **29**, 535–545 (2002)
- Stute, W., Zhu, L.X.: Nonparametric checks for single-index models. *Ann. Stat.* **33**, 1048–1083 (2005)
- Van Keilegom, I., González-Manteiga, W., Sellero, Sánchez: Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *Test* **17**, 401–415 (2008)
- Wang, J.L., Xue, L.G., Zhu, L.X., Chong, Y.S.: Estimation for a partial linear single-index model. *Ann. Stat.* **30**, 475–497 (2010)
- Wen, X., Cook, R.D.: Optimal sufficient dimension reduction in regressions with categorical predictors. *J. Stat. Plan. Inference* **137**, 1961–1978 (2007)
- Wong, H.L., Fang, K.T., Zhu, L.X.: A test for multivariate normality based on sample entropy and projection pursuit. *J. Stat. Plan. Inference* **45**, 373–385 (1995)
- Xia, Y.C., Tong, H., Li, W.K., Zhu, L.X.: An adaptive estimation of dimension reduction space. *J. R. Stat. Soc. B* **64**, 363–410 (2002)
- Xia, Y.C.: Model check for multiple regressions via dimension reduction. *Biometrika* **96**, 133–148 (2009)
- Xia, Q., Xu, W.L., Zhu, L.X.: Consistently determining the number of factors in multivariate volatility modelling. *Stat. Sin.* **25**, 1025–1044 (2015)
- Zheng, J.X.: A consistent test of functional form via nonparametric estimation techniques. *J. Econ.* **75**, 263–289 (1996)
- Zhu, L.X.: *Nonparametric Monte Carlo Tests and Their Applications*. Springer, New York (2005)
- Zhu, L.X., Miao, B.Q., Peng, H.: On sliced inverse regression with high dimensional covariates. *J. Am. Stat. Assoc.* **101**, 630–643 (2006)
- Zhu, L.P., Zhu, L.X., Ferré, L., Wang, T.: Sufficient dimension reduction through discretization-expectation estimation. *Biometrika* **97**, 295–304 (2010)
- Zhu, L.X., An, H.Z.A.: A test for nonlinearity in regression models. *J. Math.* **4**, 391–397 (1992). (**Chinese**)
- Zhu, L.X., Li, R.Z.: Dimension-reduction type test for linearity of a stochastic model. *Acta Math. Appl. Sin.* **14**, 165–175 (1998)
- Zhu, L.X., Neuhaus, G.: Nonparametric Monte Carlo tests for multivariate distributions. *Biometrika* **87**, 919–928 (2000)