ORIGINAL PAPER

Specification testing of partially linear single-index models: a groupwise dimension reduction-based adaptive-to-model approach

Junmin Liu[1] · Deli Zhu[1] · Luoyao Yu[1] · Xuehu Zhu[1]

Abstract

This paper develops a groupwise dimension reduction-based adaptive-to-model test for partially linear single-index models. The test behaves as a local smoothing test would if the model were bivariate. The test statistic under the null hypothesis is asymptotically normally distributed. The test can detect local alternatives distinct from the null hypothesis at the rate that existing local smoothing tests can achieve when the regression model contains bivariate covariates. Therefore, the curse of dimensionality is largely alleviated. Numerical studies, including two real data examples, are conducted to examine the finite sample performance of the proposed test.

Keywords Model checking · Partially linear single-index model · Group sufficient dimension reduction · Adaptive-to-model

1 Introduction

Consider the following partially linear single-index model (PLSIM):

$$Y = \beta^{\top} X + g(\alpha^{\top} Z) + \epsilon, \tag{1.1}$$

where Y is the scale response, $W = (X^{\top}, Z^{\top})^{\top} \in \mathbb{R}^{p_1 + p_2}$ is the $(p_1 + p_2)-$dimensional covariate, $g(\cdot)$ an unknown smooth functions, β and α are unknown vectors of parameters, and ϵ is the error term such that $E(\epsilon | X, Z) = 0$. For identifiability, α satisfies $||\alpha|| = 1$. Typically, model (1.1) is a reasonable compromise between

✉ Xuehu Zhu
  zhuxuehu@xjtu.edu.cn

[1] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China

fully parametric and fully nonparametric modelling. The literature on PLSIM estimation is enormous. For example, Carroll et al. (1997) first proposed a maximum quasi-likelihood method to estimate the generalised PLSIM, Xia and Härdle (2006) extended the minimum average variance estimation (MAVE) approach developed by Xia et al. (2002), Zhu and Xue (2006) proposed empirical-likelihood-based inference for the PLSIM, Wang et al. (2010) developed a two-step method to estimate the model, Liang et al. (2010) proposed the semi-parametrically efficient profile least-squares estimators of coefficients, and Lu et al. (2019) devised a method to consistently estimate the biased PLSIM. Zhao et al. (2020) considered PLSIMs of panel data with errors correlated in space and time and proposed a generalised F-type test method to check index parameters. It is not all so easy to determine whether a real data set corresponds to the given statistical formalisation. Therefore, it is crucial to perform a suitable and efficient model checking before further statistical analysis.

Many efforts have been devoted to checking parametric models since the 1980s. The two most popular methodologies in the literature are local and global smoothing methods. Local smoothing methods are sensitive to high-frequency/oscillating alternative models in low-dimensional cases. However, they suffer from slow convergence rates due to nonparametric estimation and thus are greatly affected by the curse of dimensionality. For examples of local smoothing methods based on nonparametric estimation, see Härdle and Mammen (1993), Zheng (1996), Fan and Li (1996), Dette (1999), Fan et al. (2001), Fan and Huang (2001), Koul and Ni (2004) and Van Keilegom et al. (2008).

Globally smoothing methods involve empirical process-based tests that are typically functions of the averages of weighted sums of residuals. These tests converge to their weak limits at the rate of $n^{-1/2}$, which is the fastest possible rate at which they can detect local alternatives at the fastest possible rate of $n^{-1/2}$. Thus, they have theoretical advantages over local smoothing tests. However, when the dimension is greater than 1, the intractability of the limiting null distributions requires a resampling approximation such as the wild bootstrap to determine critical values. Practical evidence shows that these type of tests are less sensitive to oscillating alternatives models. For examples of these methods, see Stute (1997), Stute et al. (1998), Zhu (2003), Khmaladze and Koul (2004), and Stute et al. (2008). González Manteiga and Crujeiras (2013) provided a comprehensive review of the studies in this domain.

A direct way to alleviate the curse of dimensionality is to project the high-dimensional covariates onto one-dimensional spaces. Escanciano (2006) and Lavergne and Patilea (2008, 2012) proposed tests based on projected covariates. Zhu (2003) and Stute et al. (2008) used residual processes to construct tests that can also be considered dimension reduction types. These tests typically require Monte Carlo approximations to determine critical values (e.g. Escanciano 2006; Lavergne and Patilea 2008), although some of them, such as that of Lavergne and Patilea (2012), are asymptotically distribution-free. All these tests use one-dimensional projections to overcome the curse of dimensionality. However, the computational of the test statistics is a serious burden. The computations become even more elaborate if we further need to use a resampling approximation such as the bootstrap to determine critical values. Recently, Guo et al. (2016) developed the innovative model-adaptive local smoothing methodology to test the specification of parametric single-index models, thus alleviating the

dimensionality problem. Zhu et al. (2017), Tan et al. (2018), Tan and Zhu (2019), Zhu et al. (2021) and Zhu et al. (2022) extended this strategy to use other parametric dimension reduction models. Li et al. (2021) proposed an adaptive-to-model hybrid of tests for parametric regression models, which fully inherits the merits of nonparametric estimation-based tests and empirical process-based tests and while avoiding their shortcomings. However, few studies have considered model checking for PLSIMs based on dimension reduction. In this paper, we construct a groupwise dimension reduction-based adaptive-to-model test for PLSIMs, to mitigate the curse of dimensionality. First, the groupwise dimension reduction method ensures that the proposed test statistic is automatically adaptive to the underlying model under the respective null and alternative hypotheses. It can thus alleviate the dimensionality problem, and simultaneously achieve the omnibus property under the alternative hypothesis. Under the null hypothesis, in probability, the test statistic only involves the bivariate covariates $\beta^\top X$ and $\alpha^\top Z$. The null distribution of the proposed test statistic converges to the limiting null distribution at a faster rate $O_p(nh)$, where h is being a bandwidth converging to zero at a certain rate. Moreover, by fully utilising the information of the low-dimensional null model, the proposed test can detect the local alternatives distinct from the null hypothesis at a faster rate $O_p(n^{-1/2}h^{-1/2})$ than the convergence rate $O_p(n^{-1/2}h^{-(p_1+p_2)/4})$ of existing local smoothing tests for parametric models, where $p_1 + p_2$ is the dimension of the complete set (X, Z), for example, as in the tests developed by Fan and Li (1996). Therefore, when the dimensions of X and Z are large, the new test is superior to existing local smoothing tests in terms of significance level maintenance and power enhancement.

The rest of the paper is organised as follows. In Sect. 2, we construct the test statistic. Because growpwise sufficient dimension reduction techniques play a crucial role in the proposed test, we also review groupwise least-squares (GLS) estimation. Section 3 discusses some asymptotic properties of the new test. Section 4 includes simulation studies and the analyses for two real data sets. The regularity conditions and all of the proofs of the theoretical results are presented in the Appendix.

2 Model-adaptive test construction

2.1 Basic construction

As we often have no knowledge of the model structure under the alternative hypothesis, the general alternative hypothesis takes the following form:

$$Y = m(X, Z) + \epsilon, \tag{2.1}$$

where $m(X, Z) = E(Y|X, Z)$ with $m(\cdot, \cdot)$ being an unknown smoothing function. For any $p_1 \times p_1$ orthogonal matrix B_1 and $p_2 \times p_2$ orthogonal matrix B_2, we have $m(X, Z) = m(B_1 B_1^\top X, B_2 B_2^\top Z) \equiv: \tilde{m}(B_1^\top X, B_2^\top Z)$ with $\tilde{m}(\cdot, \cdot) = m(B_1\cdot, B_2\cdot)$ as the function m is unknown. Therefore, any purely nonparametric regression model (2.1) can be reformulated as the groupwise dimension reduction model:

$$Y = m(B_1^\top X, B_2^\top Z) + \epsilon, \qquad (2.2)$$

where B_1 is a $p_1 \times q_1$ matrix with q_1 orthogonal columns and B_2 is a $p_2 \times q_2$ matrix with q_1 orthogonal columns, q_1 and q_2 are unknown numbers such that $1 \leq q_1 \leq p_1$ and $1 \leq q_2 \leq p_2$. For identifiability, we assume that the matrixes B_1 and B_2 satisfy $B_1^\top B_1 = I_{q_1}$ and $B_2^\top B_2 = I_{q_2}$. Based on this observation, we consider an alternative model (2.2) that covers more model structures and is widely used in the sufficient dimension reduction field. We reformulate the hypotheses as

$$H_0 : E(Y|X, Z) = \beta^\top X + g(\alpha^\top Z) \text{ for some } \beta \in \mathbb{R}^{p_1}, \alpha \in \mathbb{R}^{p_2},$$
$$H_1 : E(Y|X, Z) = m(B_1^\top X, B_2^\top Z) \neq \beta^\top X + g(\alpha^\top Z) \text{ for any } \beta \in \mathbb{R}^{p_1}, \alpha \in \mathbb{R}^{p_2}.$$

The null and alternative models can then be unified. Under the null hypothesis, $q_1 = 1$, $q_2 = 1$, $B_1 = \beta/||\beta||_2$ and $B_2 = \alpha$. $||A||_2$ denotes the L_2-norm of a vector A throughout this paper. Under the alternative hypothesis, $q_1 \geq 1$ and $q_2 \geq 1$. Therefore, we can construct a test that is automatically adaptive to the null and alternative models by consistently estimating q_1, q_2, B_1 and B_2 under the null and the alternative models, respectively.

Let $\epsilon = Y - \beta^\top X - g(\alpha^\top Z)$. Under the null hypothesis H_0, $B_1 = \kappa\beta$ with $\kappa = 1/||\beta||_2$, and $B_2 = \alpha$. Thus, we have

$$E(\epsilon|X, Z) = 0 \Rightarrow E(\epsilon|X, Z) = E(\epsilon|B_1^\top X, B_2^\top Z) = 0.$$

such that

$$E(\epsilon E(\epsilon|B_1^\top X, B_2^\top Z) f(B_1^\top X, B_2^\top Z)) = E(E^2(\epsilon|B_1^\top X, B_2^\top Z) f(B_1^\top X, B_2^\top Z)) = 0,$$

where $f(\cdot)$ denotes the density function of $(B_1^\top X, B_2^\top Z)$. Under the alternative hypothesis H_1, as

$$E(Y|X, Z) = E(\epsilon|B_1^\top X, B_2^\top Z) = m(B_1^\top X, B_2^\top Z) - \beta^\top X - g(\alpha^\top Z) \neq 0,$$

we have

$$E\left[\epsilon E(\epsilon|B_1^\top X, B_2^\top Z) W(B_1^\top X, B_2^\top Z)\right] = E\left[E^2(\epsilon|B_1^\top X, B_2^\top Z)\right] > 0. \quad (2.3)$$

The above argument implies that we can construct a consistent test based on the left term in (2.3). The null hypothesis H_0 is rejected for large values of the test statistic.

Let $\{(x_1, z_1, y_1), \cdots, (x_n, z_n, y_n)\}$ denote independent and identically distributed (i.i.d.) samples. We estimate $E(\epsilon|B_1^\top X, B_2^\top Z)$ using the kernel estimate:

$$\hat{E}(\hat{\epsilon}_i|B_{1n}^\top x_i, B_{2n}^\top z_i) = \frac{\frac{1}{n-1}\sum_{j\neq i}^n K_h(B_{1n}^\top x_j - B_{1n}^\top x_i, B_{2n}^\top z_j - B_{2n}^\top z_i)\hat{\epsilon}_j}{\frac{1}{n-1}\sum_{j\neq i}^n K_h(B_{1n}^\top x_j - B_{1n}^\top x_i, B_{2n}^\top z_j - B_{2n}^\top z_i)},$$

where $K_h = K(\cdot/h)/h^{\hat{q}_1+\hat{q}_2}$ with $K(\cdot)$ being a $(\hat{q}_1 + \hat{q}_2)$-dimensional product kernel function with univariate kernel function $k(\cdot)$ and h being a bandwidth, B_{1n} and B_{2n} are estimates of B_1 and B_2 with estimated structural dimensions $\hat{q}_1$ and $\hat{q}_2$, respectively, which we discuss in the next subsection. Here, $\hat{\epsilon}_i$ denotes the estimate of the residual term ϵ_i, and $\hat{\epsilon}_i = y_i - \hat{\beta}_1^\top x_i - \hat{g}(\hat{\alpha}^\top z_i)$, where $\hat{\beta}_1^\top$, $\hat{\alpha}$ and $\hat{g}(\cdot)$ denote the estimators of β, θ and $g(\cdot)$, respectively. More details are presented in the Appendix.

The density function $f(B_{1n}^\top x_i, B_{2n}^\top z_i)$ for $i = 1, 2, ..., n$, can be estimated by the following kernel form:

$$\hat{f}_i = \frac{1}{n-1} \sum_{j \neq i}^{n} K_h(B_{1n}^\top x_j - B_{1n}^\top x_i, B_{2n}^\top z_j - B_{2n}^\top z_i).$$

Therefore, a non-standardised test statistic is defined by

$$S_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{\epsilon}_i \hat{\epsilon}_j K_h(B_{1n}^\top x_j - B_{1n}^\top x_i, B_{2n}^\top z_j - B_{2n}^\top z_i). \qquad (2.4)$$

Remark 2.1 The nonparametric kernel-based test in Zheng (1996) can also be extended to check the PLSIM, and its test statistic can be defined as:

$$S_{ZHn} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{1}{h^{p_1+p_2}} \tilde{K}\left(\frac{x_i - x_j}{h}, \frac{z_i - z_j}{h}\right) \hat{\epsilon}_i \hat{\epsilon}_j, \qquad (2.5)$$

where $\tilde{K}_h = \tilde{K}(\cdot/h, \cdot/h)/h^{p_1+p_2}$ with $\tilde{K}(\cdot)$ being a $(p_1 + p_2)-$dimensional kernel function. There are two differences between the formulae (2.4) and (2.5). First, we use $K_h(\cdot)$ in S_n instead of $\tilde{K}_h(\cdot, \cdot)$ in T_{ZH}. We prove that under H_0, $\hat{q}_1 = \hat{q}_2 = 1$, $B_{1n} \to \beta/||\beta||_2$ and $B_{2n} \to \alpha/||\alpha||_2$ in probability, implying that our test reduces the dimension $p_1 + p_2$ to 2. Under the alternative hypothesis, we show that B_{1n} and B_{2n} are automatically consistently estimated. Second, it can be easily shown that for the test statistic S_{ZHn} to have a finite limiting distribution under the null hypothesis, the standardising constant must be $nh^{(p_1+p_2)/2}$ to obtain $nh^{(p_1+p_2)/2}S_{ZHn}$. We will see that the standardising form nhS_n has a finite limit under H_0 and diverges to infinity much faster than the typical rate $nh^{(p_1+p_2)/2}$ of the test in Zheng (1996) under H_1. The results are presented in Sect. 3.

2.2 A review of groupwise least squares estimation

In general, matrices B_1 and B_2 are not identifiable. For any $q_1 \times q_1$ orthogonal matrix C_1 and $q_2 \times q_2$ orthogonal matrix C_2, as the σ-fields generated by the random variable $(B_1^\top X, B_2^\top Z)$ are equivalent to those generated by $(\tilde{B}_1^\top X, \tilde{B}_2^\top Z)$ with $\tilde{B}_1 = B_1 \times C_1$ and $\tilde{B}_2 = B_2 \times C_2$, we have

$$E(\epsilon|B_1^\top X, B_2^\top Z) = E(\epsilon|\tilde{B}_1^\top X, \tilde{B}_2^\top Z).$$

However, it is sufficient to identify the spaces spanned by B_1 and B_2 when we construct an adaptive-to-model test in Sect. 2.1. Groupwise dimension reduction (Li 2009) can be used to identify the subspaces spanned by the column vectors of matrices B_1 and B_2.

There are several groupwise dimension reduction approaches available in the literature. Li (2009) proposed a framework for grouped sufficient dimension reduction. Motivated by the MAVE method in Xia et al. (2002), Li et al. (2010) proposed an estimator to incorporate group information into dimension reduction. Guo et al. (2015) developed groupwise dimension reduction based on the "direct sum envelope". Zhou et al. (2016) discussed overlapped groupwise dimension reduction. Generally, these methods are computationally demanding because the resulting estimators need to be solved by an iterative procedure. This procedure involves iteratively estimating the nonparametric and the parametric component. Zhu et al. (2021) proposed a GLS estimation method for groupwise dimension reduction to avoid the iterative procedure. Thus, we recommend the GLS method in Zhu et al. (2021) to estimate B_1 and B_2.

According to Zhu et al. (2021), GLS estimation proceeds as follows:

1. Consider some transformation functions of the response variable, $f_1(Y), \ldots, f_t(Y)$ satisfying $E(f_k(Y)) = 0$, where t is a prespecified number, and the least-squares estimates of regressing $f_k(Y)$ on $W = (X^\top, Z^\top)^\top$ is

$$\beta_k = \arg\min_{\beta_k} E\{f_k(Y) - W^\top \beta_k\}^2,$$

for $k = 1, \cdots, t$. Then the target matrix M can be constructed as follows:

$$M = (M_1^\top, M_2^\top)^\top = \left(\frac{\beta_1}{1 + ||\beta_1||}, \ldots, \frac{\beta_t}{1 + ||\beta_t||} \right), \qquad (2.6)$$

where M_1 and M_2 are $p_1 \times t$ and $p_2 \times t$ matrixes, respectively.
2. Then $M_i M_i^\top$ is a $p_i \times p_i$ positive semi-definite matrix satisfying $\mathrm{Span}(M_i M_i^\top) = \mathrm{Span}(B_i)$, for $i = 1, 2$.
3. When the observations $\{w_i, y_i\}_{i=1}^n$ are available, β_k can be estimated by the least squares estimates as:

$$\beta_{kn} = \arg\min_{\beta_k} \sum_{i=1}^{n} \left\{ f_k(y_i) - w_i^\top \beta_k \right\}^2.$$

The target matrix M in (2.6) can be estimated by:

$$M_n = (M_{1n}^\top, M_{2n}^\top)^\top = \left(\frac{\beta_{1n}}{1 + ||\beta_{1n}||}, \ldots, \frac{\beta_{tn}}{1 + ||\beta_{tn}||} \right).$$

When q_i is given, an estimate B_{in} of B_i consists of the eigenvectors associated with the q_i largest eigenvalues of $M_i M_i^\top$, for $i = 1, 2$. For more details, refer to Zhu et al. (2021).

2.3 Dimensionality estimation

As illustrated above, the two structure dimensions q_1 and q_2 are essential to estimating matrices B_1 and B_2. We adopt the thresholding double ridge ratio criterion proposed by Zhu et al. (2020). To address the special case $q_i = p_i$, we define some artificial eigenvalues $\hat{\lambda}_{i(p_i+1)} = \hat{\lambda}_{i(p_i+2)} = 0$. Let

$$\hat{r}_{ij} = \frac{\hat{s}^*_{i(j+1)} + c_{2n}}{\hat{s}^*_{ij} + c_{2n}} \quad \text{with} \quad \hat{s}^*_{ij} = \frac{\hat{\lambda}_{i(j+1)} + c_{1n}}{\hat{\lambda}_{ij} + c_{1n}} - 1, \qquad (2.7)$$

where c_{1n} and c_{2n} are the two ridges converging to 0, and $0 \leq \hat{\lambda}_{ip_i} \leq \ldots \leq \hat{\lambda}_{i1}$ are the eigenvalues of matrix $M_{in}M_{in}^\top/t$. The dimension q_i can be estimated by:

$$\hat{q}_i := \begin{cases} 0, & \text{if } \hat{r}_{ij} > \tau, \text{ for all } 1 \leq j \leq p_i, \\ \arg\max_{1 \leq j \leq p_i} \{j : \hat{r}_{ij} \leq \tau\}, & \text{otherwise,} \end{cases} \qquad (2.8)$$

where $0 < \tau < 1$. Based on the rule of thumb in Zhu et al. (2020), we set $\tau = 0.5$ to avoid overestimation with a large τ or underestimation with small τ. As the target matrix here differs from that in Zhu et al. (2020), we recommend the ridge values $c_{1n} = 0.4\log(n)/\sqrt{n}$ and $c_{2n} = 0.8\log(n)/\sqrt{n}$, based on some numerical studies.

The following proposition presents the consistency of the estimators $\hat{q}_k$, for $k = 1, 2$.

Proposition 2.1 *Under conditions A1 and A2 in the Appendix, assume that $c_{1n} \to 0$, $c_{2n} \to 0$ and $c_{1n}c_{2n}n \to \infty$. Then we have*

$$\lim_{n \to \infty} P\left(\hat{q}_1 = q_1, \hat{q}_2 = q_2\right) = 1.$$

3 Asymptotic properties

3.1 Limiting null distribution

First, we provide some notation. Let

$$s^2 = 2 \int K^2(u, v)dudv E\{[Var(\varepsilon^2|B_1^\top X, B_2^\top Z)]^2 p(B_1^\top X, B_2^\top Z)\}, \qquad (3.1)$$

and

$$\hat{s}^2 = \frac{2h^2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} K_h^2(B_{1n}^\top x_j - B_{1n}^\top x_i, B_{2n}^\top z_j - B_{2n}^\top z_i)\hat{\epsilon}_i^2\hat{\epsilon}_j^2. \qquad (3.2)$$

Next, we state the result for the null limiting distribution.

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

$$nh\,S_n \xrightarrow{d} N(0, s^2).$$

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

From Theorem 3.1, we have the standardised test statistic T_n:

$$T_n = nh\,S_n/\hat{s} \xrightarrow{d} N(0, 1).$$

3.2 Power study

To study how sensitive the new test statistic is against the alternative hypothesis, we consider the following sequence of local alternative models:

$$H_{1n} : Y = \beta^\top X + g(\alpha^\top Z) + C_n f(B_1^\top X, B_2^\top Z) + \eta, \tag{3.3}$$

where the function $f(\cdot, \cdot)$ is continuous and differentiable and satisfies $E[f^2(B_1^\top X, B_2^\top Z)] < \infty$, $E(\eta|X, Z) = 0$, and β and θ are the linear combination of the columns of B_1 and B_2, respectively.

Lemma 3.1 *Under the local alternative H_{1n} in (3.3) with $C_n = n^{-1/2}h^{-1/2}$ and the same conditions in Proposition 2.1 except that $C_n^2 \log n \le c_n \to 0$, we have*

$$\lim_{n\to\infty} P\left(\hat{q}_1 = q_1, \hat{q}_2 = q_2\right) = 1.$$

Next, we state the results of the power performance of the test statistic.

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

(i) Under H_{1n} with a fixed $C_n > 0$

$$T_n/(nh) \xrightarrow{P} Constant > 0.$$

(ii) Under the sequence of local alternative hypotheses H_{1n} in (3.3), T_n has different asymptotic properties based on the rates of C_n as follows.

(a) If $q_1 = q_2 = 1$ and $C_n = n^{-1/2}h^{-1/2}$,

$$T_n \xrightarrow{d} N(u, 1),$$

where

$$u = E\{f^2(B_1^\top X, B_2^\top Z) p_{B_1 B_2}(B_1^\top X, B_2^\top Z)\}/s$$

with s^2 defined by (3.1);

(b) *if $q_1 + q_2 > 2$, $C_n n^{1/2} h^{1/2} \to c_0 > 0$ for some constant c_0 or $C_n n^{1/2} h^{1/2} \to \infty$ and $C_n n^{1/2} h^{(q_1+q_2)/4} \to 0$, we have*

$$T_n / h^{(q_1+q_2-2)/2} \xrightarrow{d} N(0, 1);$$

(c) *if $q_1 + q_2 > 2$ and, $C_n = n^{-1/2} h^{-(q_1+q_2)/4}$, we have*

$$T_n / h^{(q_1+q_2-2)/2} \xrightarrow{d} N(u, 1);$$

(d) *if $q_1 = q_2 = 1$ and $C_n n^{1/2} h^{1/2} \to \infty$, or $q_1 + q_2 > 2$ and $C_n n^{1/2} h^{(q_1+q_2)/4} \to \infty$, we have*

$$T_n / (C_n^2 n h) \xrightarrow{P} u > 0.$$

4 Numerical studies

4.1 Simulations

In this subsection, we examine the finite-sample performance of the proposed test using several numerical examples. Each experiment is repeated 1000 times to compute the empirical sizes and powers at the significance level $\alpha = 0.05$. The parametrical vectors β and α are estimated using the MAVE method in Xia and Härdle (2006) (see the Appendix). In some applications, the asymptotic normal approximation does not work well in finite sample settings. Thus, re-sampling techniques are often used in finite samples. We apply the wild bootstrap algorithm adopted by Guo et al. (2016).

Consider the bootstrap observations $y_i^* = \hat{\beta}^T x_i + g(\hat{\alpha}^T z_i) + \epsilon_i^*$, where $\epsilon_i^* = \hat{\epsilon}_i \times U_i$. $\{U_i\}_{i=1}^n$ can be chosen to be i.i.d. Bernoulli variates with

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

Let T_n^* be the bootstrap version of T_n, based on the bootstrap samples $\{(x_1, z_1, y_1^*), \cdots, (x_n, z_n, y_n^*)\}$. The null hypothesis is rejected if T_n is larger than the corresponding quantile of the bootstrap distribution of T_n^*.

For our test, we select the kernel function $K(\cdot)$ to be Gaussian, and set the bandwidths as $h_1 = 0.8 n^{-1/(1+4)}$ and $h = 0.8 n^{-1/(\hat{q}_1 + \hat{q}_2 + 4)}$ with the estimated dimensions $\hat{q}_1$ and $\hat{q}_2$ of $B_{1n}^\top X$ and $B_{2n}^\top Z$, where h_1 is used to estimate β, α and the function $g(\cdot)$ and h is used to construct the test statistics.

As in Remark 2.1, we extend the test in Zheng (1996) to check the PLSIMs for comparison and write the test statistic as T_n^{ZH}. For the test in Zheng (1996), we set the bandwidth as $h = 1.5 n^{-1/(4+p_1+p_2)}$ and select the quartic kernel function $K(u) = \frac{15}{16}\left(1 - u^2\right)^2$, if $|u| \le 1$ and 0, otherwise. We compare the results of these two test statistics computed from the 500 wild bootstrap samples.

The observations $\{x_i\}_{i=1}^n$ and $\{z_i\}_{i=1}^n$ are generated i.i.d. from multivariate normal distributions $N(0_{p_1}, \Sigma_1)$ and $N(0_{p_2}, \Sigma_2)$, respectively, and independent of the standard normal errors. Here, $\Sigma_1 = I_{p_1 \times p_1}$ and $\Sigma_2 = I_{p_2 \times p_2}$. The sample sizes are set as $n = 200, 400$ and the dimensions are $p_1 = p_2 = 4, 8$.

Example 1 Consider the model

$$Y = \beta^\top X + (\alpha^\top Z + 1)^2 + a(0.4\beta^\top X + 1)^2 + \epsilon,$$

where β and α are set using the following two cases:

- Case 1: $\beta = \underbrace{(1, \ldots, 1, 0, \ldots, 0)}_{p_1/2}^\top / \sqrt{p_1/2}$, and $\alpha = \underbrace{(0, \ldots, 0, 1, \ldots, 1)}_{p_2/2}^\top / \sqrt{p_2/2}$;
- Case 2: $\beta = \underbrace{(1, \ldots, 1)}_{p_1} / \sqrt{p_1}$, and $\alpha = \underbrace{(1, \ldots, 1)}_{p_2} / \sqrt{p_2}$.

The error term ϵ follows the standard normal distribution. In this example, $a = 0$ and $a \neq 0$ correspond to the null and the alternative hypothesis, respectively. The results of the empirical sizes and powers are displayed in Table 1. From Table 1, we make the following observations:

First, the proposed T_n can control the empirical sizes well. T_n^{ZH} tends to be conservative, especially when $p_1 = p_2 = 4$, but when $p_1 = p_2 = 8$, the significance level can be better maintained. Second, the empirical powers of the proposed test T_n increase reasonably with larger a, and thus the test is significantly and uniformly more powerful than T_n^{ZH}. T_n^{ZH} is invalid when detecting the alternative hypothesis with a large dimension $p_1 = p_2 = 8$. For $p_1 = p_2 = 4$ and $p_1 = p_2 = 8$, the dimensions of X and Z have less influence for T_n than they do for T_n^{ZH}. T_n performs uniformly better than T_n^{ZH} in the above two cases.

Example 2 In this example, we examine the finite sample performance of the proposed method under the following model:

$$Y = \beta^\top X + exp(\alpha^\top Z) + 0.6(\alpha^\top Z)^2 + a(\beta^\top X)^2 + \epsilon,$$

where all parameter values are set to be the same as those in Example 1. $a = 0$ corresponds to the null hypothetical model. The empirical sizes and powers are presented in Table 2.

We observed that the power performances of T_n in Example 1 and Example 2 are similar by comparing Tables 1, 2 and 3. The proposed test can maintain the significance level as well as in the previous example, and its power performs significantly and uniformly better than that of the test in Zheng (1996).

Example 3 In this example, we examine the finite sample performance of the proposed method under the model:

$$Y = \beta^\top X + 0.9(\alpha^\top Z)^3 + a(X_1 + 0.5X_2^2) + \epsilon,$$

Table 1 Empirical sizes and powers of the tests for Example 1

| | | T_n | | | | T_n^{ZH} | | | |
| | | $p_1 = p_2 = 4$ | | $p_1 = p_2 = 8$ | | $p_1 = p_2 = 4$ | | $p_1 = p_2 = 8$ | |
	a	$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
Case 1	0	0.041	0.054	0.051	0.057	0.099	0.081	0.061	0.053
	0.2	0.214	0.495	0.263	0.522	0.101	0.080	0.060	0.050
	0.4	0.770	0.979	0.751	0.983	0.095	0.118	0.068	0.051
	0.6	0.987	1.000	0.978	1.000	0.149	0.192	0.065	0.069
	0.8	1.000	1.000	0.999	1.000	0.191	0.255	0.045	0.057
	1	1.000	1.000	1.000	1.000	0.237	0.315	0.072	0.080
Case 2	0	0.039	0.046	0.044	0.052	0.095	0.079	0.057	0.067
	0.2	0.200	0.490	0.238	0.495	0.100	0.086	0.058	0.069
	0.4	0.747	0.985	0.768	0.984	0.097	0.110	0.055	0.069
	0.6	0.984	1.000	0.977	1.000	0.152	0.190	0.060	0.062
	0.8	1.000	1.000	1.000	1.000	0.206	0.270	0.064	0.058
	1	1.000	1.000	1.000	1.000	0.205	0.331	0.074	0.080

Table 2 Empirical sizes and powers of the tests for Example 2

| | | T_n | | | | T_n^{ZH} | | | |
| | | $p_1 = p_2 = 4$ | | $p_1 = p_2 = 8$ | | $p_1 = p_2 = 4$ | | $p_1 = p_2 = 8$ | |
	a	$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
Case 1	0	0.035	0.045	0.040	0.047	0.098	0.077	0.055	0.046
	0.2	0.194	0.633	0.121	0.513	0.093	0.086	0.061	0.048
	0.4	0.756	0.989	0.514	0.955	0.121	0.160	0.056	0.036
	0.6	0.909	1.000	0.758	0.987	0.195	0.261	0.058	0.058
	0.8	0.933	0.998	0.799	0.993	0.271	0.434	0.063	0.051
	1	1.000	1.000	0.824	1.000	0.370	0.584	0.066	0.045
Case 2	0	0.031	0.044	0.041	0.048	0.078	0.068	0.059	0.050
	0.2	0.199	0.604	0.117	0.464	0.084	0.099	0.048	0.054
	0.4	0.699	0.994	0.461	0.952	0.137	0.175	0.045	0.041
	0.6	0.900	1.000	0.752	0.993	0.196	0.295	0.056	0.045
	0.8	0.938	0.999	0.792	0.997	0.283	0.441	0.045	0.044
	1	0.942	0.997	0.839	0.995	0.358	0.583	0.063	0.065

where X_i and Z_i denote the i−th components of X and Z, respectively. All parameters values are the same as those used in Example 1. The related results are displayed in Table 3.

From Table 3, both tests still control the size well. When $p_1 = p_2 = 4$, the empirical powers of T_n increase faster than those of T_n^{ZH} as a increase. For T_n^{ZH}, the empirical sizes and powers are close to the significance level $\alpha = 0.05$ when $p_1 = p_2 = 8$.

Table 3 Empirical sizes and powers of the tests for Example 3

| | | T_n | | | | T_n^{ZH} | | | |
| | | $p_1 = p_2 = 4$ | | $p_1 = p_2 = 8$ | | $p_1 = p_2 = 4$ | | $p_1 = p_2 = 8$ | |
	a	$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
Case 1	0	0.036	0.046	0.037	0.055	0.077	0.073	0.027	0.030
	0.2	0.157	0.470	0.100	0.254	0.080	0.092	0.039	0.027
	0.4	0.699	0.975	0.444	0.874	0.095	0.132	0.041	0.041
	0.6	0.962	1.000	0.837	0.996	0.138	0.226	0.033	0.039
	0.8	0.996	1.000	0.983	1.000	0.222	0.366	0.028	0.037
	1	1.000	1.000	0.996	1.000	0.270	0.464	0.048	0.045
Case 2	0	0.041	0.054	0.043	0.055	0.074	0.072	0.033	0.032
	0.2	0.086	0.298	0.080	0.470	0.082	0.090	0.035	0.037
	0.4	0.435	0.856	0.292	0.666	0.080	0.117	0.037	0.038
	0.6	0.854	1.000	0.667	0.968	0.134	0.189	0.039	0.049
	0.8	0.987	0.998	0.938	0.999	0.179	0.306	0.036	0.038
	1	0.998	1.000	0.986	1.000	0.222	0.411	0.043	0.047

T_n is significantly better than T_n^{ZH} at detecting the alternative hypothesis because dimension reduction mitigates the curse of dimensionality.

Example 4 To examine the stability of powers against the structure dimension, we consider a more complex model with higher structure dimensions q_1 and q_2:

$$Y = \beta^\top X + \sin(\alpha^\top Z) + 2\alpha^\top Z + a(X_1^3 + 0.3X_2^2 + \exp(X_3) + |X_4|$$
$$+ 0.2Z_1^2 + Z_2^2 + \cos(Z_3) + Z_4^3)/4 + \epsilon,$$

where X_i and Z_i denote the $i-$th components of X and Z, respectively. All parameters values are the same as those used in Example 1. $a = 0$ corresponds to the null hypothetical model. Under the alternative hypotheses, the structural dimensions q_1 and q_2 of the alternative models are equal to 4. The results are presented in Table 4.

We observe that the performances of empirical powers and sizes in Example 4 and Examples 1–3 are similar by comparing Table 4 and Tables 1, 2 and 3. The empirical sizes of the two tests are close to the significance level. However, the proposed test has significantly higher power than T_n^{ZH}, especially when $p_1 = p_2 = 8$. We also find that the structural dimensions have a negligible effect on the empirical sizes and powers of our test.

Table 4 Empirical sizes and powers of the tests for Example 4

| | | T_n | | | | T_n^{ZH} | | | |
| | | $p_1 = p_2 = 4$ | | $p_1 = p_2 = 8$ | | $p_1 = p_2 = 4$ | | $p_1 = p_2 = 8$ | |
	a	$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$	$n = 200$	$n = 400$
Case 1	0	0.041	0.046	0.044	0.054	0.109	0.108	0.082	0.071
	0.2	0.087	0.191	0.133	0.267	0.110	0.102	0.091	0.095
	0.4	0.308	0.573	0.399	0.741	0.107	0.116	0.065	0.086
	0.6	0.617	0.946	0.712	0.965	0.132	0.148	0.077	0.072
	0.8	0.832	0.990	0.888	0.998	0.163	0.214	0.060	0.087
	1	0.930	1.000	0.965	1.000	0.166	0.259	0.061	0.096
Case 2	0	0.044	0.057	0.042	0.053	0.115	0.079	0.073	0.057
	0.2	0.127	0.256	0.111	0.188	0.118	0.087	0.078	0.069
	0.4	0.430	0.759	0.288	0.567	0.106	0.117	0.067	0.082
	0.6	0.727	0.976	0.566	0.881	0.123	0.164	0.082	0.067
	0.8	0.918	1.000	0.794	0.981	0.155	0.187	0.068	0.068
	1	0.971	1.000	0.899	0.998	0.157	0.280	0.064	0.080

4.2 Real data analysis

4.2.1 Body fat data

We now apply the proposed method to the analysis of the body fat data, which can be found at http://lib.stat.cmu.edu/datasets/bodyfat. Chen et al. (2016) analysed this dataset, which provides estimates of the percentage of body fat determined by underwater weighing and various body circumference measurements. The dataset contains 251 samples with 14 attributes. Consistent with Chen et al. (2016), we select the following 12 attributes: $height^4/weight^2$, age, and circumferences of 10 body parts, namely, neck, chest, abdomen, hip, thigh, knee, ankle, biceps, forearm, and wrist. We set the knee, ankle, and forearm circumferences as $X = (X_1, X_2, X_3)$. The other predictors are set as $Z = (Z_1, Z_2, \cdots, Z_9)$. The response variable is the logarithm of the percentage of body fat (Y). We delete one observations in which the percentage of body fat is equal to 0.

The value of the test statistic is $T_n = 18.2261$ and the $p-$value is approximately 0. Therefore, there is enough evidence to reject the null hypothesis at the significance level $\alpha = 0.05$. The residual plot of the PLSIM fitted against $\hat{\beta}^\top X$ and $\hat{\alpha}^\top Z$ in Fig. 1 shows a linear relationship between $\hat{\beta}^\top X$ and the residuals.

4.2.2 Auto MPG

This data set is obtained from the Machine Learning Repository at the University of California-Irvine (http://archive.ics.uci.edu/ml/datasets/Auto+MPG). Xia (2007) and Guo et al. (2016) analysed this data set. It has 406 samples and 8 attributes. The predictors are cylinders, engine displacement, horsepower, vehicle weight, time to
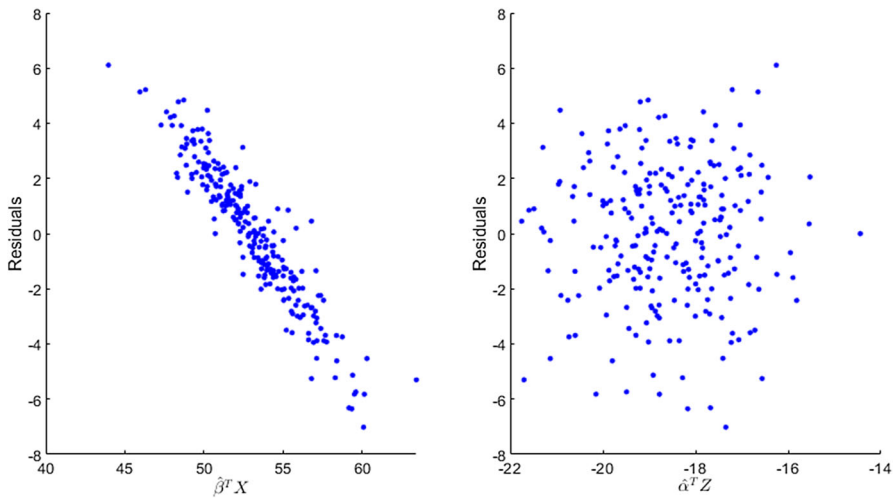
Fig. 1 Plot of the residuals against the linear part and single-indexing direction in the body fat data
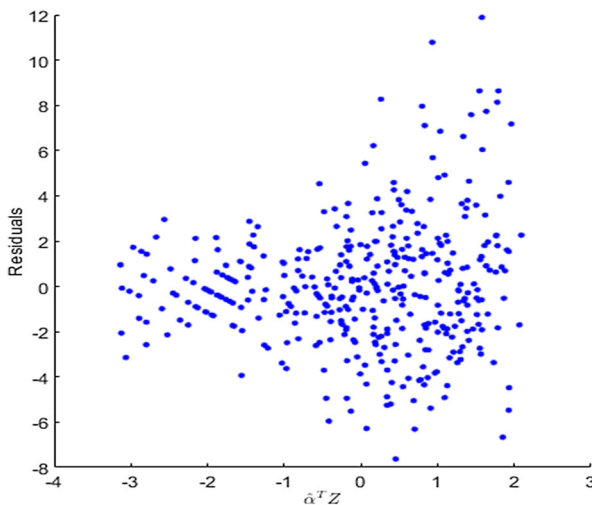
Fig. 2 Plot of the residuals against the single-index direction in the Auto MPG data

accelerate from 0 to 60 mph, model year, and the origin of the car. Miles per gallon (Y) is the response variable. Consistent with Xia (2007), we set cylinders, engine displacement, horsepower, vehicle weight, time to accelerate from 0 to 60 mph, and model year as $Z_1, Z_2, \cdots, Z_6$. $X_1 = 1$ if a car is from America, and 0 otherwise. $X_2 = 1$ if a car is from Europe and 0 otherwise.

The value of the proposed test statistic is $T_n = -1.1447$, and the p-value is 0.2523. Hence, the null hypothesis should be accepted at the significance level $\alpha = 0.05$. We plot the residual plot against the single-indexing direction in Fig. 2, and the plot shows no pattern. Thus, it is reasonable to fit this dataset based on the PLSIM.

5 Appendix

5.1 Brief review of the minimum average variance estimation of the PLSIM

As stated previously, the proposed test procedure needs to estimate the parameter vectors β and α and the function g. Therefore, we briefly review the MAVE approach developed by Xia and Härdle (2006) to estimate the PLSIM. The basic algorithm is based on the minimum average variance:

$$(\beta, \alpha) = \arg\min_{\beta, \alpha} E\left[Y - \beta^\top X - g\left(\alpha^\top Z\right)\right]^2$$

subject to $\|\alpha\| = 1$. A two-step iterative algorithm to estimate the PLSIM. Given (β, θ), we have:

$$\begin{pmatrix} a_j \\ d_j \end{pmatrix} = \left\{ \sum_{i=1}^{n} w_{ij} \begin{pmatrix} 1 \\ z_{ij}^\top \alpha \end{pmatrix} \begin{pmatrix} 1 \\ z_{ij}^\top \alpha \end{pmatrix}^T \right\}^{-1} \sum_{i=1}^{n} w_{ij} \begin{pmatrix} 1 \\ z_{ij}^\top \alpha \end{pmatrix} \left(y_i - \beta^\top z_i\right).$$

Given $\left(a_j, d_j\right)$, we have:

$$\begin{pmatrix} \hat{\beta} \\ \hat{\alpha} \end{pmatrix} = \left\{ \sum_{j=1}^{n} G\left(\alpha^\top z_j\right) I_n\left(z_j\right) \sum_{i=1}^{n} w_{ij} \begin{pmatrix} X_i \\ d_j z_{ij} \end{pmatrix} \begin{pmatrix} x_i \\ d_j z_{ij} \end{pmatrix}^T \right\}^{-1}$$

$$\times \sum_{j=1}^{n} G\left(\alpha^\top z_j\right) I_n\left(Z_j\right) \sum_{i=1}^{n} w_{ij} \begin{pmatrix} x_i \\ d_j Z_{ij} \end{pmatrix} \left(y_i - a_j\right),$$

where $z_{ij} = z_i - z_j$, w_{ij}, for $i, j = 1, 2, \cdots, n$, are some weights with $\sum_{i=1}^{n} w_{ij} = 1$, and $G(\cdot)$ is another weight function that controls the contribution of (X_j, Z_j, Y_j) to the estimation of β and θ. The function g can be estimated using the following kernel estimate:

$$\hat{g}(B_{2n}^{\top} z_i) = \frac{\frac{1}{n-1} \sum_{j \neq i}^{n} Q_{h_1}(B_{2n}^\top z_j - B_{2n}^\top z_i)(y_i - \hat{\beta}^\top x_i)}{\frac{1}{n-1} \sum_{j \neq i}^{n} Q_{h_1}(B_{2n}^\top z_j - B_{2n}^\top z_i)},$$

where $Q_{h_1} = K(\cdot/h_1)/h_1$ with $Q(\cdot)$ being some univariate kernel function.

In our numerical studies, the MAVE method is implemented by using the first column vector of B_{1n} and B_{2n} as the initial estimators for β and α and setting the maximum number of iterations to 100. More details can be found in Xia and Härdle (2006).

5.2 Regularity conditions

A1 Let $W = (X, Z)$. Assume that $E(W|B^\top W)$ is a linear function of $B^\top W$ with the columns of $B \in \mathbb{R}^{p \times q}$ being any basis of the central space $\mathcal{S}_{Y|W}$, where $p = p_1 + p_2$ and q denotes the dimension of $\mathcal{S}_{Y|W}$.

A2 $f_i(Y)$'s satisfy $E[f_i(Y)] = 0$ and $E[f_i^2(Y)] < \infty$ for $i = 1, \cdots, t$ and the coverage condition holds, namely, the target matrix M defined in (2.6) is such that $\text{Span}(M) = \mathcal{S}_{Y|W}$. Additionally, the second moment of W exists.

A3 The density function $p_{B_1, B_2}(\cdot, \cdot)$ of $(B_1^\top X, B_2^\top Z)$ is continuous with a bounded first-order derivative and satisfies, on its support $\mathbb{C}_1$,

$$0 < \inf_{(B_1^\top x, B_2^\top z) \in \mathbb{C}_2} p_{B_1, B_2}(B_1^\top x, B_2^\top z) \leq \sup_{(B_1^\top x, B_2^\top z) \in \mathbb{C}_2} p_{B_1, B_2}(B_1^\top x, B_2^\top z) < \infty.$$

A4 The function $g(\cdot)$ is $\eta-$order partially differentiable for some positive integer η, and the ηth partially derivative of $g(\cdot)$ is bounded.

A5 The kernel function $Q(\cdot)$ is symmetric and second-order continuously differentiable, and satisfies

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

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

where δ_{ij} is the Kronecker delta with η given in Condition A4.

A6 The kernel function $K(\cdot)$ is a bounded, symmetric kernel function. It is first order continuously differentiable and satisfies $\int K(u) du = 1$.

A7 $n \to \infty, h_1 \to 0, h \to 0$,

1) under the null or local alternative hypotheses with $C_n = n^{-1/2} h^{-1/2}$, $n h_1 \to \infty$, $n h^2 \to \infty$ and $n h_1^{2\eta} h \to 0$;
2) under the global alternative hypothesis, $n h_1 \to \infty$, $n h^{q_1+q_2} \to \infty$ and $n h_1^{2\eta} h^{(q_1+q_2)/2} \to 0$,

where η is given in Condition A6.

Remark 5.1 Conditions A1 and A2 are necessary for obtaining consistent estimators of matrixes B_1 and B_2. The linearity condition A1 holds when W follows an elliptically contoured distribution, see Li (1991), and it is mild in high-dimensional scenarios (see Hall and Li 1993). The coverage condition $\text{Span}(M) = \mathcal{S}_{Y|W}$ in Condition A2 in the literature is widely adopted to overcome technical issues, see Wu and Li (2011). Conditions A3, A4, A5 and A6 are widely used in the literature for nonparametric estimation. The four conditions ensure that the test is well-behaved, see Fan and Li (1996) and Zheng (1996). It is worth pointing out that Condition A5 pertaining to the higher-order kernel plays a critical role in bias reduction, see Robinson (1988) and Fan and Li (1996). We use different bandwidths h_1 and h to estimate the function $\hat{g}(\cdot)$ and construct the test statistic T_n, because they involve the various covariates under the alternative hypothesis. This phenomenon has been discussed in several studies. For

example Stute and Zhu (2005) stated that the optimal bandwidth for estimation should be different from that for test statistic construction. For more details on Conditions A3-A7 refer to Fan and Li (1996), Zhu and Zhu (2018) and Zhu et al. (2021).

5.3 The proofs of the theoretical results

Proof of Proposition 2.1. Employing the same justification procedure to that for Theorem 2.1 in Zhu et al. (2020), we can get the results. Then we omit the detail here.

Proof of Theorem 3.1. Define the events $A_{1n} = \{T_n \leq c\}$ for any constant c and $A_{2n} = \{\hat{q}_1 = 1, \hat{q}_2 = 1\}$. Proposition 2.1 shows that under the null hypothesis, $\lim_{n\to\infty} P(A_{2n}) = 1$, where $P(A)$ denotes the probability that event A happens. Then we have $\lim_{n\to\infty} P(A_{2n}) = \lim_{n\to\infty} P(A_{1n} \cap A_{2n})$. This result ensures that under the null hypothesis, in an asymptotic sense we only need to consider the properties of the test statistic on the event that $\hat{q}_1 = 1$ and $\hat{q}_2 = 1$.

For notation simplicity, define $K_{B_{1n}B_{2n}ij} = K((B_1^\top x_i - B_1^\top x_j)/h, (B_2^\top z_i - B_2^\top z_j)/h)$, $K_{B_1B_2ij}^1 = \frac{\partial K(B_1^\top(x_i-x_j)/h, B_1^\top(z_i-z_j)/h)}{\partial(B_1^\top(x_i-x_j)/h)}$, $K_{B_1B_2ij}^1 = \frac{\partial K(B_1^\top(x_i-x_j)/h, B_1^\top(z_i-z_j)/h)}{\partial(B_2^\top(z_i-z_j)/h)}$, $g_i = g(\theta^\top z_i)$, $\hat{g}_i = \hat{g}(\hat{\theta}^\top z_i)$, $\epsilon_i = y_i - \beta_1^\top x_j - g_i$, $\hat{\epsilon}_i = y_i - \hat{\beta}_1^\top x_j - \hat{g}_i$, $p_i = p_{B_2}(B_2^\top z_i)$ and $\hat{p}_i = \hat{p}_{B_{2n}}(B_{2n}^\top z_i)$, $\tilde{y}_i = y_i - \beta^\top x_i$. Throughout the proof of this theorem, $E_i(\cdot) = E(\cdot|B_1^\top x_i, B_2^\top w_i)$.

Noted that $\hat{\epsilon}_i = y_i - \hat{\beta}_1^\top x_i - \hat{g}(\hat{\alpha}^\top z_i) = \epsilon_i + [\beta^\top x_i - \hat{\beta}^\top x_i] + [g(\alpha^\top z_i) - \hat{g}(\hat{\alpha}^\top z_i)]$. We then decompose the term S_n as

$$S_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{i\neq j}^{n} \frac{1}{h^2} K_{B_{1n}B_{2n}ij} \epsilon_i \epsilon_j$$

$$+ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{i\neq j}^{n} \frac{1}{h^2} K_{B_{1n}B_{2n}ij} [\beta_1^\top x_i - \hat{\beta}_1^\top x_i][\beta_1^\top x_j - \hat{\beta}_1^\top x_j]$$

$$+ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{i\neq j}^{n} \frac{1}{h^2} K_{B_{1n}B_{2n}ij} [g_i - \hat{g}_i][g_j - \hat{g}_j]$$

$$+ \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{i\neq j}^{n} \frac{1}{h^2} K_{B_{1n}B_{2n}ij} [\beta^\top x_i - \hat{\beta}^\top x_i] \epsilon_j$$

$$+ \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{i\neq j}^{n} \frac{1}{h^2} K_{B_{1n}B_{2n}ij} [g_i - \hat{g}_i] \epsilon_j$$

$$+ \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{i\neq j}^{n} \frac{1}{h^2} K_{B_{1n}B_{2n}ij} [\beta^\top x_i - \hat{\beta}^\top x_i][g_j - \hat{g}_j]$$

$$\equiv: S_{1n} + S_{2n} + S_{3n} + 2S_{4n} + 2S_{5n} + 2S_{6n}. \tag{5.1}$$

We now prove that both the terms nhS_n and nhS_{1n} have the same limiting null distribution, $nhS_{2n} = o_p(1)$, $nhS_{3n} = o_p(1)$, $nhS_{4n} = o_p(1)$, $nhS_{5n} = o_p(1)$ and $nhS_{6n} = o_p(1)$.

Consider the term S_{1n}. The first order Taylor expansion for S_{1n} with respect to B_1 and B_2 yields

$$S_{1n} =: S_{11n} + S_{12n} + S_{13n},$$

where S_{11n} and S_{12n} have the following forms:

$$S_{11n} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{i \neq j}^{n} \frac{1}{h^2} K_{B_1 B_2 ij} \epsilon_i \epsilon_j,$$

$$S_{12n} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{i \neq j}^{n} \frac{1}{h^3} K^1_{\tilde{B}_1 \tilde{B}_2 ij} (B_{1n} - B_1)^{\top} (x_i - x_j) \epsilon_i \epsilon_j,$$

$$S_{13n} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{i \neq j}^{n} \frac{1}{h^3} K^1_{\tilde{B}_1 \tilde{B}_2 ij} (B_{2n} - B_2)^{\top} (z_i - z_j) \epsilon_i \epsilon_j,$$

with $\tilde{B}_1 = \{\tilde{B}_{1ij}\}_{p_1 \times q_1}$, $\tilde{B}_{1ij} \in [\min\{B_{1ij}, B_{1nij}\}, \max\{B_{1ij}, B_{1nij}\}]$, $B_1 = \{B_{1ij}\}_{p_1 \times q_1}$ and $B_{1n} = \{B_{1nij}\}_{p_1 \times q_1}$. Due to the facts that $||B_{1n} - B_1|| = O_p(1/\sqrt{n})$ and the kernel function $K(\cdot)$ has a continuous and bounded first-order derivative, we infer that replacing $\tilde{B}_1$ by B_1 does not influence the convergence rates of S_{12n} and S_{13n}. Similarly, we can infer that replacing $\tilde{B}_2$ by B_2 does not influence the convergence rates of S_{12n} and S_{13n}.

It is obvious that S_{11n} is an U−statistic with the kernel as:

$$H(v_i, v_j) = \frac{1}{h^2} K_{B_1 B_2 ij} \epsilon_i \epsilon_j,$$

where $v_i = (x_i, z_i, y_i)$. Under null hypothesis, as $E[H(v_i, v_j)|v_j] = 0$, S_{11n} is a degenerate U-statistic. Under Conditions A3–A7, adapting the similar arguments as that for Lemma 3.3 in Zheng (1996), we can obtain

$$nhS_{11n} \xrightarrow{d} N(0, s^2).$$

We now prove that the second term nhS_{12n} and nhS_{13n} tend to zero. As $K(\cdot)$ is spherically symmetric, the term S_{12n} can be rewritten as an U-statistic with the kernel:

$$H(v_i, v_j) = \frac{1}{h^3} K^1_{B_1 B_2 ij} (B_{1n} - B_1)^{\top} (x_i - x_j) \epsilon_i \epsilon_j$$

$$+ \frac{1}{h^3} K^1_{B_1 B_2 ij} (B_{1n} - B_1)^{\top} (x_j - x_i) \epsilon_i \epsilon_j.$$

It is obvious that $E[H(v_i, v_j)|v_i] = 0$. Thus, the term S_{12n} is a degenerate $U-$statistic. Applying the arguments used for handling the term S_{11n}, together with $||B_{1n} - B_1|| = O_p(1/\sqrt{n})$, it yields that $nhS_{12n} = o_p(1)$. Similarly, we can get that $nhS_{13n} = o_p(1)$.

Therefore, together with the results about S_{11n}, S_{12n} and S_{13n}, we have

$$nhS_{1n} \xrightarrow{d} N(0, s^2).$$

Now we turn to prove that $nhS_{2n} = o_p(1)$. Note that

$$S_{2n} = (\beta_1 - \hat{\beta}_1)^\top \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{i \neq j} \frac{1}{h^2} K_{B_{1n}B_{2n}ij}(x_i - x_i)(x_j - x_j)^\top (\beta_1 - \hat{\beta}_1)$$

$$\equiv: (\beta_1 - \hat{\beta}_1)^\top \tilde{S}_{2n}(\beta_1 - \hat{\beta}_1).$$

Due to the fact $\tilde{S}_{2n}$ is a $U-$statistic, it is easy to conclude that $\tilde{S}_{2n} = O_p(1)$. Because $\beta_1 - \hat{\beta}_1 = O_p(1/\sqrt{n})$ and $h \to 0$, we get $nhS_{2n} = O_p(h) = o_p(1)$.

Consider the term S_{3n} that can be written as:

$$S_{3n} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^2} K_{B_{1n}B_{2n}ij} \frac{\hat{p}_i}{p_i} \frac{\hat{p}_j}{p_j} [\hat{g}_i - g_i(u)][\hat{g}_j - g_j]$$

$$+ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^2} K_{B_{1n}B_{2n}ij} \left[\frac{\hat{p}_i - p_i}{p_i} \frac{\hat{p}_j - p_j}{p_j} - 2\frac{(\hat{p}_i - p_i)\hat{p}_j}{p_i p_j} \right]$$

$$[\hat{g}_i - g_i(u)][\hat{g}_j - g_j]$$

$$=: \tilde{S}_{3n} + o_p(\tilde{S}_{3n}).$$

Substituting the kernel estimators $\hat{g}_i$ and $\hat{p}_i$ into $\tilde{S}_{3n}$, we have

$$\tilde{S}_{3n} = \frac{1}{n(n-1)^3} \sum_{i=1}^{n} \sum_{j \neq i} \sum_{l \neq i} \sum_{k \neq j} \frac{1}{h^2} \frac{1}{h_1^2} \frac{1}{p_i p_j} K_{B_{1n}B_{2n}ij} Q_{B_{2n}il} Q_{B_{2n}jk}$$

$$[y_l - \hat{\beta}^\top x_l - g_i][y_k - \hat{\beta}^\top x_k - g_j].$$

Due to the facts $B_{1n} - B_1 = O_p(1/\sqrt{n})$, $B_{2n} - B_2 = O_p(1/\sqrt{n})$ and $\beta_n - \beta = O_p(1/\sqrt{n})$, adapting the similar statement for dealing with the term S_{1n}, we can infer that replacing B_{1n}, B_{2n} and $\hat{\beta}$ by B_1, B_2 and β respectively does not influence the convergence rates of $\tilde{S}_{3n}$, namely,

$$\tilde{S}_{3n} = \tilde{S}_{31n} + o_p(\tilde{S}_{31n}),$$

where

$$\tilde{S}_{31n} = \frac{1}{n(n-1)^3} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{l \neq i}^{n} \sum_{k \neq j}^{n} \frac{1}{h^2} \frac{1}{h_1^2} \frac{1}{p_i p_j} K_{B_1 B_2 ij} Q_{B_2 il} Q_{B_2 jk}$$
$$[\tilde{y}_l - g_i][\tilde{y}_k - g_j].$$

Following the similar idea of justifying Proposition A.1 in Fan and Li (1996), we will show $nh\tilde{S}_{31n} = o_p(1)$ by proving that $E(\tilde{S}_{31n}^2) = o(n^{-2}h^{-2})$. However, it is difficult and tedious to directly calculate $E(\tilde{S}_{31n}^2)$ since it contains eight summations. We first show $E(\tilde{S}_{31n}) = o(n^{-1}h^{-1})$. Then we use this result and a symmetry argument to show $E(\tilde{S}_{31n}^2) = o(n^{-2}h^{-2})$.

Decompose $E(\tilde{S}_{31n})$ with two terms with two subsets of subscripts as:

$\mathcal{A}_1 = \{i, j, l, k$ are all different from each other$\}$;
$\mathcal{A}_2 = \{i, j, l, k$ take no more than three different values$\}$.

Then $\tilde{S}_{31n}$ can be decomposed as $\tilde{S}_{31n} = \tilde{S}_{311n} + \tilde{S}_{312n}$, where the summation indices of $\tilde{S}_{311n}$ and $\tilde{S}_{312n}$ are associated with $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively.

Under the assumption that $nh_1^{2\eta}h = o(1)$, by applying Lemma B.1, Lemmas 2 and 3 in Robinson (1988), we have

$$E(\tilde{S}_{311n}) = \frac{1}{hh_1^2} E\left(K_{B_1 B_2 13} E_1 \left\{ \frac{1}{p_1} Q_{B_1 12}[g_2 - g_1] \right\} \right.$$
$$\left. E_3 \left\{ \frac{1}{p_3} Q_{B_1 34}[g_4 - g_3] \right\} \right)$$
$$\leq C \frac{h_1^{2\eta}}{h^{q_1}} E\left[D_F(B_2^\top z_1, u) D_F(B_2^\top z_3, u) K_{B_1 B_2 13} \right]$$
$$= O(h_1^{2\eta}) = o(n^{-1}h^{-1}).$$

Next we consider subset $\mathcal{A}_2$. It can be divided into three groups: case (I) $l = k$; case (II) $l = j$; case (III) $k = i$. Then $\tilde{S}_{312n}$ can be further decomposed as $\tilde{S}_{312n} = \tilde{S}_{3121n} + \tilde{S}_{3122n} + \tilde{S}_{3123n}$ associated with the above three sub-events. For $\tilde{S}_{3121n}$ with the sub-event (I),

$$E(\tilde{S}_{3121n}) = \frac{1}{n(n-1)^3} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{k \neq i, k \neq j}^{n} \frac{1}{hh_1^2} E\left(\frac{1}{p_i p_j} K_{B_1 B_2 ij} Q_{B_2 ik} Q_{B_2 jk} \right.$$
$$\left. \times [g_k - g_i][g_k - g_j] \right).$$

By applying Lemma B.1, Lemmas 2 and 3 in Robinson (1988) and Fubini's theorem, we have

$$E(\tilde{S}_{3121n}) = \frac{n(n-1)(n-2)}{n(n-1)^3 hh_1^2} E\left(\frac{1}{p_1 p_2} K_{B_1 B_2 12} Q_{B_2 13} Q_{B_2 23}[g_3 - g_1][g_3 - g_2]\right)$$

$$= \frac{(n-2)}{(n-1)^2 hh_1^2} E\left(\frac{1}{p_1 p_2} K_{B_1 B_2 12} E_1\{[g_3 - g_1]Q_{B_2 13}\} Q_{B_2 23}[g_3 - g_2]\right)$$

$$= O(n^{-1} h_1^\eta h^{-1}) = o(n^{-1} h^{-1}).$$

The same argument can be applied to the terms $E(\tilde{S}_{3122n})$ and $E(\tilde{S}_{3123n})$ to obtain the upper bound $o(n^{-1}h^{-1})$. Hence, altogether we have $E(\tilde{S}_{31n}) = E(\tilde{S}_{311n}) + E(\tilde{S}_{312n}) = o(n^{-1}h^{-1})$.

Now we consider $E(\tilde{S}_{31n}^2)$ by using the similar decomposition of $E(\tilde{S}_{31n})$ although it is much more complicated. Note that

$$E(\tilde{S}_{31n}^2) = E\left\{\frac{1}{n^3(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{l \neq j}^{n} \sum_{k \neq j}^{n} \frac{1}{hh_1^2} \frac{1}{p_i p_j} K_{B_1 B_2 ij} Q_{B_2 il} Q_{B_2 jk} \right.$$

$$\left. [g_l - g_i][g_k - g_j]\right\}^2$$

$$= \frac{1}{n^6(n-1)^2} \frac{1}{h^2 h_1^4} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{l \neq i}^{n} \sum_{k \neq j}^{n} \sum_{i'=1}^{n} \sum_{j' \neq i'}^{n} \sum_{l' \neq i'}^{n} \sum_{k' \neq j'}^{n}$$

$$E\left(\left\{\frac{1}{p_i p_j} K_{B_1 B_2 ij} Q_{B_2 il} Q_{B_2 jk}[g_l - g_i][g_k - g_j]\right\}\right.$$

$$\left.\left\{\frac{1}{p_{i'} p_{j'}} K_{B_1 B_2 i'j'} Q_{B_2 i'l'} Q_{B_2 j'k'}[g_{l'} - g_{i'}][g_{k'} - g_{j'}]\right\}\right)$$

$$=: LA_1 + LA_2 + LA3,$$

where the summation indices of LA_1, LA_2 and LA_3 respectively associated with three subsets of subscripts $\mathcal{B}_1$, $\mathcal{B}_2$ and $\mathcal{B}_3$ as:

$\mathcal{B}_1 = \{i, j, l, k$ are all different from $i', j', l', k'\}$;
$\mathcal{B}_2 = \{$exactly one index from i, j, l, k equals one of subscripts $i', j', l', k'\}$;
$\mathcal{B}_3 = \{$the eight summation indices $i, j, l, k, i', j', l', k'$ take no more than six different values$\}$.

With $\mathcal{B}_1$, the sums in LA_1 with i, j, l, k and i', j', l', k' are independent of each others. Thus LA_1 is equal to the square of $E(\tilde{S}_{21n})$. Hence we can obtain that $LA_1 = o(n^{-2}h^{-q_1})$. Next we consider LA_2 with the subset $\mathcal{B}_2$. By symmetry, we only need to compute three cases: case (I) $i = i'$; case (II) $i = l'$; case (III) $l = l'$, and LA_2 can be further decomposed as $LA_2 = LA_{21} + LA_{22} + LA_{23}$ associated with the above three subsets.

Under case (I),

$$LA_{21}$$
$$= \frac{1}{n^2(n-1)^6 h^2 h_1^4} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{l \neq i}^{n} \sum_{k \neq j}^{n} E\left[\frac{1}{p_i p_j} K_{B_1 B_2 ij} Q_{B_2 il} Q_{B_2 jk} [g_l - g_i][g_k - g_j] \right.$$
$$\left. \sum_{j' \neq i}^{n} \sum_{l' \neq i}^{n} \sum_{k' \neq j}^{n} \frac{1}{p_i p_{j'}} K_{B_1 B_2 ij'} Q_{B_2 il'} Q_{B_2 j'k'} [g_{l'} - g_i][g_{k'} - g_{j'}] \right].$$

Via an application of the Fubini's theorem and Lemma B.1, Lemmas 2 and 3 in Robinson (1988), we have

$$LA_{21}$$
$$= \frac{1}{nh^2 h_1^4} E\left[\frac{1}{p_1 p_2} K_{B_1 B_2 12} Q_{B_2 13} Q_{B_2 24} [g_3 - g_1][g_4 - g_2] \right.$$
$$\left. \frac{1}{p_1 p_5} K_{B_1 B_2 15} Q_{B_2 16} Q_{B_2 57} [g_6 - g_1][g_7 - g_5] \right]$$
$$= \frac{1}{nh^2 h_1^4} E\left(\frac{1}{p_1 p_2} E_1\{[g_3 - g_1] Q_{B_2 13}\} E_2\{[g_4 - g_2] Q_{B_2 24}\} \right.$$
$$\left. \frac{1}{p_1 p_5} E_1\{[g_6 - g_1] Q_{B_2 16}\} E_5\{[g_7 - g_5] Q_{B_2 57}\} K_{B_1 B_2 12} K_{B_1 B_2 15} \right)$$
$$\leq \frac{h_1^{4\eta}}{nh^2} E\left[\frac{1}{p_1 p_2 p_1 p_5} K_{B_1 B_2 12} D_F(B_2^\top z_1) D_F(B_2^\top z_2) D(B_2^\top z_1) D(B_2^\top z_5) K_{B_1 B_2 15} \right]$$
$$= O(h_1^{4\eta} n^{-1}) = o(n^{-2} h^{-2}).$$

In case (II),

$$LA_{22}$$
$$= \frac{1}{n^2(n-1)^6 h^2 h_1^4} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{l \neq i}^{n} \sum_{k \neq j}^{n} E\left(\frac{1}{p_i p_j} K_{B_1 B_2 ij} Q_{B_2 il} Q_{B_2 jk} [g_l - g_i][g_k - g_j] \right.$$
$$\left. \sum_{i' \neq i}^{n} \sum_{j' \neq i'}^{n} \sum_{k' \neq j'}^{n} \frac{1}{p_{i'} p_{j'}} K_{B_1 B_2 i'j'} Q_{B_2 i'i} Q_{B_2 j'k'} [g_i - g_{i'}][g_{k'} - g_{j'}] \right).$$

The application of Fubini's theorem and Lemma B.1, Lemmas 2 and 3 in Robinson (1988) again yields

$$LA_{22}$$
$$= \frac{1}{nh^2 h_1^4} E\left(\frac{1}{p_1 p_2} K_{B_1 B_2 12} Q_{B_2 13} Q_{B_2 24} [g_3 - g_1][g_4 - g_2] \right.$$
$$\left. \frac{1}{p_5 p_6} K_{B_1 B_2 56} Q_{B_2 15} Q_{B_2 67} [g_1 - g_5][g_7 - g_6] \right)$$
$$= \frac{1}{nh^2 h_1^4} E\left(K_{B_1 B_2 12} \frac{1}{p_1 p_2} E_1\{[g_3 - g_1(u_1)] Q_{B_2 13}\} E_2\{[g_4 - g_2 Q_{B_2 24}\} \right.$$

$$\frac{1}{p_5 p_6} K_{B_1 B_2 56}[g_1 - g_5] Q_{B_2 15} E_6 \big\{ [g_7 - g_6] Q_{B_2 67} \big\} \Big)$$

$$\leq \frac{h_1^{3\eta}}{nh^2} E\left[\frac{1}{p_1 p_2 p_1 p_5} K_{B_1 B_2 12} D_F(B_2^\top z_1) D_F(B_2^\top z_2) \right.$$

$$\left. [g_1 - g_5] Q_{B_2 15} D_F(B_2^\top z_6) K_{B_1 B_2 56} \right]$$

$$= O(h_1^{3\eta} n^{-1}) = o(n^{-2} h^{-2}).$$

For case (III), the similar argument as above for case (II) can be used to justify $LA_{23} = o(n^{-2} h^{-q_1})$. Altogether, we have $LA_2 = o(n^{-2} h^{-q_1})$.

Last, consider the sum LA_3 with the subset $\mathcal{B}_3$ in which the eight summation indices $i, j, l, k, i', j', l', k'$ take no more than six different values. As it is easy to show that $LA_3 = o(n^{-2} h^{-2})$ by the similar arguments used above, we then omit the detail. Altogether, we conclude that $E(\tilde{S}_{31n}^2) = o(n^{-2} h^{-2})$. The application of Chebyshiev's inequality implies $\tilde{S}_{31n} = o_p(n^{-1} h^{-1})$.

Using the similar process as those for the term $\tilde{S}_{31n}$, under Conditions A3−A7 in Appendix, we can justify that the terms $\tilde{S}_{321n}$, $\tilde{S}_{322n}$ and $\tilde{S}_{323n}$ have the following converging rates:

$$E(||\tilde{S}_{321n}||^2) = O(\max\{h_1^{2\eta}, h_1^{2\eta} n^{-1}, h_1^\eta n^{-1}\}) = o(n^{-1} h^{-1});$$

$$E(||\tilde{S}_{322n}^2||^2) = O(\max\{h_1^{2\eta}, h_1^{2\eta} n^{-1}, h_1^\eta n^{-1}\}) = o(n^{-1} h^{-1});$$

$$E(||\tilde{S}_{323n}^2||^2) = O(\max\{h_1^{4\eta}, h_1^{4\eta} n^{-1}, h_1^{3\eta} n^{-1}\}) = o(n^{-1} h^{-1}).$$

To save the space, we omit some more detailed deductions. The Chebyshiev's inequality yields that

$$\tilde{S}_{321n} = o_p(n^{-1/2} h^{-1}); \;\; \tilde{S}_{322n} = o_p(n^{-1/2} h^{-1}); \;\; \tilde{S}_{323n} = o_p(n^{-1/2} h^{-1}).$$

As $||B_{1n} - B_1|| = O_p(n^{-1/2})$, together with the above convergence rates, we have $\tilde{S}_{32n} = o_p(n^{-1} h^{-1})$. Combining the convergence rates of $\tilde{S}_{31n}$ and $\tilde{S}_{32n}$, we conclude $\tilde{S}_{3n} = o_p(n^{-1} h^{-q_1/2})$.

Now we discuss the convergence rate of the term S_{5n}. Note that

$$S_{5n} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{i \neq j}^{n} \frac{1}{h} K_{B_{1n} B_{2n} ij} \epsilon_i [\hat{g}_j - g_j] \frac{\hat{p}_j}{p_j}$$

$$+ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{i \neq j}^{n} \frac{1}{h} K_{B_{1n} B_{2n} ij} \epsilon_i [\hat{g}_j - g_j] \left(\frac{\hat{p}_j - p_j}{p_j} \right)$$

$$=: \tilde{S}_{5n} + o_p(\tilde{S}_{5n}).$$

Substituting the kernel estimators $\hat{g}_j$ and $\hat{p}_j$ into $\tilde{S}_{5n}$, we have

$$\tilde{S}_{5n} = \frac{1}{n(n-1)^2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{k \neq j}^{n} \frac{1}{hh_1} \frac{1}{p_j} K_{B_{1n}B_{2n}ij} \epsilon_i Q_{B_{2n}jk}[g_k - g_j].$$

Adopting the similar process for dealing with the term S_{1n} and S_{3n}, we can infer that replacing B_{1n}, B_{2n} and $\hat{\beta}$ by B_1, B_2 and β respectively does not influence the convergence rates of $\tilde{S}_{3n}$, namely,

$$\tilde{S}_{5n} = \tilde{S}_{51n} + o_p(\tilde{S}_{51n}),$$

where

$$\tilde{S}_{51n} = \frac{1}{n(n-1)^2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{k \neq j}^{n} \frac{1}{hh_1} \frac{1}{p_j} K_{B_1 B_2 ij} \epsilon_i Q_{B_2 jk}[\tilde{y}_k - g_j].$$

As $E[\epsilon_i | x_i, z_i] = 0$, Fubini's theorem and the properties of conditional expectation yield $E(\tilde{S}_{51n}) = 0$. We then consider the second order moment of $\tilde{S}_{51n}$ as:

$$E(\tilde{S}_{51n}^2)$$
$$= E\left\{\frac{1}{n(n-1)^2} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{k \neq j}^{n} \frac{1}{hh_1} \frac{1}{p_j} K_{B_1 B_2 ij} \epsilon_i Q_{B_2 jk}[\tilde{y}_k - g_j]\right\}^2$$
$$= E\left\{\frac{1}{n^2(n-1)^4} \frac{1}{h^2 h_1^2} \sum_{i=1}^{n} \sum_{i \neq j}^{n} \sum_{k \neq j}^{n} \sum_{i'=1}^{n} \sum_{i' \neq j'}^{n} \sum_{k' \neq j'}^{n} \right.$$
$$\left. \frac{1}{p_j p_{j'}} K_{B_1 B_2 ij} K_{B_1 B_2 i'j'} Q_{B_2 jk} Q_{B_2 j'k'} \epsilon_i [\tilde{y}_k - g_j] \epsilon_{i'} [\tilde{y}_{k'} - g_{j'}]\right\}$$
$$= E\left\{\frac{1}{n^2(n-1)^4} \frac{1}{h^2 h_1^2} \sum_{i=1}^{n} \sum_{i \neq j}^{n} \sum_{k \neq j}^{n} \sum_{i'=1}^{n} \sum_{i' \neq j'}^{n} \sum_{k' \neq j'}^{n} \right.$$
$$\left. \frac{1}{p_j p_{j'}} K_{B_1 B_2 ij} Q_{B_2 jk} K_{B_1 B_2 i'j'} Q_{B_2 j'k'} \epsilon_i [g_k - g_j] \epsilon_{i'} [g_{k'} - g_{j'}]\right\} + o(n^{-2}h^{-2}).$$

Note that $E[\epsilon_i \epsilon_{i'} | w_i, w_{i'}] \neq 0$ if and only if $i = i'$. Fubini's theorem and the Lemma B.1, Lemmas 2 and 3 in Robinson (1988) altogether yields:

$$E(\tilde{S}_{51n}^2)$$
$$= \frac{1}{nh^2 h_1^2} E\left\{\frac{1}{p_2 p_4} K_{B_1 B_2 12} Q_{B_2 23} K_{B_1 B_2 14} Q_{B_2 45} E[\epsilon_1^2 | w_1][g_3 - g_2][g_5 - g_4]\right\}$$
$$= \frac{1}{nh^2 h_1^2} E\left(\frac{1}{p_2 p_4} K_{B_1 B_2 12} K_{B_1 B_2 14} E[\epsilon_1^2 | w_1]\right.$$

$$E_2\{Q_{B_123}[g_3 - g_2]\}E_4\{Q_{B_145}[g_5 - g_4]\}\Big)$$

$$\leq \frac{h_1^{2\eta}}{nh^2}E\left[\frac{1}{p_2p_4}K_{B_1B_212}K_{B_1B_214}D_F(B_2^\top z_2)D_F(B_2^\top z_4)\right]$$

$$= O(h_1^{2\eta}n^{-1}) = o(n^{-2}h^{-1}).$$

The Chebyshiev's inequality implies that $\tilde{S}_{51n} = o_p(n^{-1}h^{-1})$. Thus $S_{5n} = o_p(n^{-1}h^{-1})$.

Using the similar statement to deal with the terms S_{1n}, S_{2n} and S_{5n}, we can conclude that $S_{4n} = o_p(n^{-1}h^{-1})$ and $S_{6n} = o_p(n^{-1}h^{-1})$.

To sum up, together with all the results about the terms S_{in} for $i = 1, \cdots, 6$, we conclude that

$$nhS_n \xrightarrow{\text{d}} N(0, s^2).$$

To complete the proof of this theorem, we justify the consistency of s_n^2 to s^2, where s_n^2 is

$$s_n^2 = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{h^2}K_{B_{1n}B_{2n}ij}^2\hat{\epsilon}_i^2\hat{\epsilon}_j^2.$$

Under the null hypothesis, since B_{1n}, B_{2n} and $\hat{g}$ are uniformly consistent to B_1, B_2 and g, respectively, some elementary computations yield an asymptotic presentation:

$$s_n^2 = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{h^2}K_{B_1B_2ij}^2\epsilon_i^2\epsilon_j^2 + o_p(1) =: \tilde{s}_n^2 + o_p(1).$$

It is clear that $\tilde{s}_n^2$ is an U−statistic with kernel:

$$H(w_i, y_i, w_j, y_j) = \frac{1}{h^2}K_{B_1B_2ij}^2\epsilon_i^2\epsilon_j^2.$$

We can easily show that the condition of lemma 3.1 of Zheng (1996) is satisfied. Based on U-statistic theory, it is easy to justify $\tilde{s}_n^2 = E(\tilde{s}_n^2) + o(1)$, where

$$E(\tilde{s}_n^2) = s^2 + o(1)$$
$$= 2\int K^2(u, v)dudvE\{[Var(\varepsilon^2|B_1^\top X, B_2^\top Z)]^2 p(B_1^\top X, B_2^\top Z)\} + o(1).$$

The more details can be found in Fan and Li (1996) and Zheng (1996). □

Proof of Theorem 3.2. Here we use the similar notations as those in Theorem 3.1 of the main body. Define $K_{B_{1n}B_{2n}ij} = K((B_1^\top x_i - B_1^\top x_j)/h, (B_2^\top z_i - B_2^\top z_j)/h)$, $g_i = g(\theta^\top z_i)$, $\hat{g}_i = \hat{g}(\hat{\theta}^\top z_i)$, $\epsilon_i = y_i - \beta_1^\top x_j - g_i$, $\hat{\epsilon}_i = y_i - \hat{\beta}_1^\top x_j - \hat{g}_i$, $p_i = p_{B_2}(B_2^\top z_i)$

and $\hat{p}_i = \hat{p}_{B_{2n}}(B_{2n}^\top z_i)$, $\tilde{y}_i = y_i - \beta^\top x_i$. Throughout the proof of this theorem, $E_i(\cdot) = E(\cdot | B_1^\top x_i, B_2^\top z_i)$. Define the events $A_{1n} = \{T_n \le c\}$ for any constant c and $A_{2n} = \{\hat{q}_1 = q_1, \hat{q}_2 = q_2\}$. Proposition 2.1 and Lemma 3.1 purport that under the global and local alternative hypothesis, we have $\lim_{n\to\infty} P(A_{2n}) = \lim_{n\to\infty} P(A_{1n} \cap A_{2n})$. This result ensures that under the global and local alternative hypotheses, in an asymptotic sense it is only needed to consider the events $\hat{q}_1 = q_1$ and $\hat{q}_2 = q_2$.

Proof of Part (I). Under Conditions A1 and A2 in Appendix, due to the facts $||B_{1n} - B_1|| = O_p(1/\sqrt{n})$ and $||B_{2n} - B_2|| = O_p(1/\sqrt{n})$, $\hat{g}$ is an uniformly consistent estimator of g, see Fan and Gijbels (1996). It is easy to prove that under the global alternative hypothesis, we have

$$S_n = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \ne i} \frac{1}{h^{q_1+q_2}} K_{B_{1n}B_{2n}ij} \hat{\epsilon}_i \hat{\epsilon}_j$$

$$= \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{i \ne j} \frac{1}{h^{q_1+q_2}} K_{B_1 B_2 ij} \epsilon_i \epsilon_j + o_p(1),$$

where $\epsilon_i = y_i - \beta_0^\top x_i - g(\alpha_0^\top z_i)$ with $(\beta_0, \alpha_0) = \arg\min_{\beta,\alpha} E[Y - \{\beta^\top X + g(\alpha^\top Z)\}]^2$. It is obvious that the term

$$\frac{1}{n(n-1)} \sum_{i=1}^n \sum_{i \ne j} \frac{1}{h}^{q_1+q_2} K_{B_1 B_2 ij} \epsilon_i \epsilon_j$$

is an U−statistic with the kernel $\frac{1}{h^{q_1+q_2}} K_{B_1 B_2 ij} \epsilon_i \epsilon_j$.

Using the element properties of U−statistic and Fubini's theorem, we have

$$S_n = E\left\{ \frac{1}{h^{q_1+q_2}} K_{B_1 B_2 12} \epsilon_1 \epsilon_2 \right\} + o_p(1)$$

$$= E\{f^2(B_1^\top X, B_2^\top Z) p_{B_1 B_2}(B_1^\top X, B_2^\top Z)\} + o_p(1) > 0,$$

where $p_{B_1 B_2}$ stands for the density function of $(B_1^\top X, B_2^\top Z)$.

Additionally, applying the same argument as that in the justification of Theorem 3.1, we can prove that in probability s_n^2 converges to some positive value which may be different from s^2 in Theorem 3.1. Thus, altogether, we have

$$T_n/(nh) \xrightarrow{\text{p}} Constant > 0.$$

Proof Part (II): As the description as the proof of Part (I) in this theorem, here it is also only needed to consider the events $\hat{q}_1 = q_1$ and $\hat{q}_2 = q_2$ in an asymptotic sense. Under the local alternative hypotheses H_{1n}, using the similar statement as that used to justify Theorem 3.1, we can conclude that:

$$S_n = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \ne i} \frac{1}{h^{q_1+q_2}} K_{B_{1n}B_{2n}ij} \hat{\epsilon}_i \hat{\epsilon}_j$$

$$=: Q_n + o_p(Q_n),$$

where

$$Q_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{i \neq j}^{n} \frac{1}{h^{q_1+q_2}} K_{B_1 B_2 ij} \epsilon_i \epsilon_j,$$

with $\epsilon_i = y_i - \beta_0^\top x_i + g(\alpha_0^\top z_i)$. Let $\varepsilon_i = y_i - \beta_0^\top x_i - g(\alpha_0^\top z_i) - C_n f(B_1^\top x_i, B_2^\top z_i)$. Then we have $\epsilon_i = \varepsilon_i + C_n f(B_1^\top x_i, B_2^\top z_i)$ and $E[\varepsilon_i | w_i] = 0$. Furthermore, Q_n is decomposed as:

$$Q_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{1}{h^{q_1+q_2}} K_{B_1 B_2 ij} [\varepsilon_i + C_n f(B_1^\top x_i, B_2^\top z_i)]$$
$$[\varepsilon_j + C_n f(B_1^\top x_j, B_2^\top z_j)]$$
$$=: Q_{1n} + C_n Q_{2n} + C_n^2 Q_{3n},$$

where Q_{1n}, Q_{2n} and Q_{3n} have the following forms as:

$$Q_{1n} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{1}{h^{q_1+q_2}} K_{B_1 B_2 ij} \varepsilon_i \varepsilon_j;$$

$$Q_{2n} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{1}{h^{q_1+q_2}} K_{B_1 B_2 ij} f(B_1^\top x_i, B_2^\top z_i) \varepsilon_j;$$

$$Q_{3n} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{1}{h^{q_1+q_2}} K_{B_1 B_2 ij} f(B_1^\top x_i, B_2^\top z_i) f(B_1^\top x_j, B_2^\top z_j).$$

Again following the similar argument to that in the justification of Theorem 3.1, we can easily obtain that $nh^{\frac{q_1+q_2}{2}} Q_{1n} \xrightarrow{d} N(0, s^2)$ and thus $Q_{1n} = O_p(n^{-1} h^{-\frac{q_1+q_2}{2}})$.

Then we consider the term Q_{2n}. In fact, Q_{2n} can be written as $U-$statistic with the kernel:

$$H_n(t_i, t_j) = \frac{1}{h^{q_1+q_2}} K_{B_1 B_2 ij} \{ f(B_1^\top x_i, B_2^\top z_i) \varepsilon_j + f(B_1^\top x_j, B_2^\top z_j) \} \varepsilon_i \},$$

where $t_i = (x_i, z_i, y_i)$. Since $E[\varepsilon_j | x_i, z_i] = 0$, then $E[H(t_i, t_j)] = 0$. To use the properties of a non-degenerate U-statistic (Serfling 1980), it is essential to prove $E[H^2(t_i, t_j)] = o(n)$. Note that

$$E[H_n^2(t_i, t_j)] \leq 4E[\frac{1}{h^{q_1+q_2}} K_{B_1 B_2 ij} f(B_1^\top x_i, B_2^\top z_i) \varepsilon_j]^2$$
$$= \frac{4}{h^{q_1+q_2}} E[\frac{1}{h^{q_1+q_2}} K_{B_1 B_2 12} f^2(B_1^\top x_1, B_2^\top z_1) E(\varepsilon_2^2 | B_1^\top x_2, B_2^\top z_2)].$$

By the application of Fubini's theorem and the change of the original variables to be $v_1 = (B_1^\top x_2 - B_1^\top x_1)/h$ and $v_2 = (B_2^\top z_2 - B_2^\top z_1)/h$ yield

$$E[H_n^2(t_i, t_j)] = \frac{4}{h^{q_1+q_2}} \int_{\mathbb{R}^{q_1+q_2}} K(v)dv E[f^2(B_1^\top x_1, B_2^\top z_1)$$

$$E(\varepsilon_2^2|B_1^\top x_2, B_2^\top z_2)] + o\left(\frac{1}{h^{q_1+q_2}}\right)$$

$$= O\left(\frac{1}{h^{q_1+q_2}}\right) = o(n).$$

We now turn to discuss the conditional expectation of $H_n(t_i, t_j)$. By Fubini's theorem, it is easy to calculate $r_n(t_i) = E\{H_n(t_i, t_j)|t_i\}$ to be

$$\begin{aligned}
r_n(t_i) &= \frac{1}{h^{q_1+q_2}} E\{K_{B_1 B_2 ij} f(B_1^\top x_j, B_2^\top z_j)\varepsilon_i|t_i\} \\
&= \varepsilon_i \int_{\mathbb{R}^{q_1}} \int_{\mathbb{R}^{q_2}} K(v_1, v_2) f(B_1^\top x_i - hv_1, B_2^\top z_i - hv_2) \\
&\quad p_{B_1 B_2}(B_1^\top x_i - hv_1, B_2^\top z_i - hv_2)dv_1 dv_2 du \\
&= \varepsilon_i f(B_1^\top x_i, B_2^\top z_i) p_{B_1 B_2}(B_1^\top x_i, B_2^\top z_i) + l_n(t_i) \\
&=: m(t_i) + l_n(t_i).
\end{aligned}$$

Let $\tilde{Q}_{2n}$ denote the "projection" of the statistic Q_{2n} as:

$$\begin{aligned}
\tilde{Q}_{2n} &= \frac{1}{n}\sum_{i=1}^{n} r_n(t_i) = \frac{1}{n}\sum_{i=1}^{n} m(t_i) + \frac{1}{n}\sum_{i=1}^{n} l_n(t_i) \\
&=: Q_{21n} + Q_{22n}.
\end{aligned}$$

Central-limit theorem yields that $\sqrt{n}Q_{21n} = O_p(1)$. As the functions $g(\cdot, \cdot)$ and $p_{B_1 B_2}(\cdot, \cdot)$ satisfy the Lipschitz condition, we have $E\{l_n^2(t_i)\} = O(h^2) \to 0$. Note that $E\{l_n(t_i)\} = 0$. We can conclude that $\sqrt{n}Q_{22n} = o_p(1)$. Therefore, altogether, $Q_{2n} = O_p(1/\sqrt{n})$. Under the local alternative hypothesis, $C_n Q_{2n} = O_p(C_n/\sqrt{n})$.

Finally consider the term $C_n^2 Q_{3n}$. It is obvious that Q_{3n} is also an $U-$statistic with the kernel:

$$H_n(t_i, t_j) = \frac{1}{h^{q_1+q_2}} K_{B_1 B_2 ij} f(B_1^\top x_i, B_2^\top z_i) f(B_1^\top x_j, B_2^\top z_j)$$

with $t_i = (x_i, z_i, y_i)$. Using the element characteristic of $U-$statistic, we have

$$\begin{aligned}
Q_{3n} &= E(H_n(t_i, t_j)) + o_p(1) \\
&= \frac{1}{h^{q_1+q_2}} E[K_{B_1 B_2 ij} f(B_1^\top x_i, B_2^\top z_i) f(B_1^\top x_j, B_2^\top z_j)] + o_p(1).
\end{aligned}$$

Again, we can derive that $E\{H_n(t_i, t_j)\} = \mu_1 + o(1)$ with

$$\mu_1 = E\{f^2(B_1^\top X, B_2^\top Z) p_{B_1 B_2}(B_1^\top X, B_2^\top Z)\}.$$

Therefore, we have $Q_{3n} = \mu_1 + o_p(1)$. The above equation implies $C_n^2 Q_{3n} = O_p(C_n^2)$. Altogether, we have

$$Q_{1n} = O_p(n^{-1} h^{-\frac{q_1+q_2}{2}}); \quad C_n Q_{2n} = O_p(C_n/\sqrt{n}); \quad C_n^2 Q_{3n} = O_p(C_n^2).$$

Additionally, following the similar arguments for proving Theorem 3.1, $s_n^2 \xrightarrow{\text{p}} s^2$. Therefore, we have the following conclusions.

If $q_1 = q_2 = 1$ and $C_n = n^{-1/2} h^{-1/2}$, Q_{1n} and Q_{3n} are the leading terms of Q_n, which yields that

$$T_n \xrightarrow{d} N(u, 1),$$

where $u = E\{f^2(B_1^\top X, B_2^\top Z) p_{B_1 B_2}(B_1^\top X, B_2^\top Z)\}/s$.

If $q_1 = q_2 = 1$ and $C_n n^{1/2} h^{1/2} \to \infty$, Q_{3n} is the leading term of Q_n. This implies that

$$T_n/(C_n^2 nh) \xrightarrow{\text{P}} u > 0.$$

If $q_1 + q_2 > 2$, if $C_n n^{1/2} h^{1/2} \to c_0 > 0$ for some constant c_0 or $C_n n^{1/2} h^{1/2} \to \infty$ and $C_n n^{1/2} h^{(q_1+q_2)/4} \to 0$, Q_{1n} is the leading term of Q_n, then we have

$$T_n/h^{(q_1+q_2-2)/2} \xrightarrow{d} N(0, 1).$$

If $q_1 + q_2 > 2$, if $C_n = n^{-1/2} h^{-(q_1+q_2)/4}$, Q_{1n} and Q_{3n} are the leading terms of Q_n, thus, we have

$$T_n/h^{(q_1+q_2-2)/2} \xrightarrow{d} N(u, 1).$$

If $q_1 + q_2 > 2$ and $C_n n^{1/2} h^{(q_1+q_2)/4} \to \infty$, Q_{3n} is the leading term of Q_n. This implies that

$$T_n/(C_n^2 nh) \xrightarrow{\text{P}} u > 0.$$

The proof is finished. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

References

Carroll R, Fan J, Gijbels I, Wand M (1997) Generalized partially linear single-index models. J Am Stat Assoc 92:477–489

Chen K, Lin J, Wang Z (2016) Least product relative error estimation. J Multivar Anal 144:91–98

Dette H (1999) A consistent test for the functional form of a regression based on a difference of variance estimates. Ann Stat 27:1012–1050

Escanciano JC (2006) A consistent diagnostic test for regression models using projections. Economet Theor 22:1030–1051

Fan JQ, Huang LS (2001) Goodness-of-fit tests for parametric regression models. J Am Stat Assoc 96:640–652

Fan Y, Li Q (1996) Consistent model specification tests: omitted variables and semiparametric functional forms. Econometrica 64:865–890

Fan J, Zhang C, Zhang J (2001) Generalized likelihood ratio statistics and Wilks phenomenon. Ann Stat 29:153–193

González Manteiga W, Crujeiras RM (2013) An updated review of goodness of fit tests for regression models. TEST 22:361–411

Guo Z, Li L, Lu W, Li B (2015) Groupwise dimension reduction via envelope method. J Am Stat Assoc 110:1515–1527

Guo X, Wang T, Zhu LX (2016) Model checking for generalized linear models: a dimension-reduction model-adaptive approach. J Roy Stat Soc B 78:1013–1035

Hall P, Li K (1993) On almost linearity of low dimensional projections from high dimensional data. Ann Stat 21:867–889

Härdle W, Mammen E (1993) Comparing nonparametric versus parametric regression fits. Ann Stat 21:1926–1947

Khmadladze EV, Koul HL (2004) Martingale transforms goodness-of-fit tests in regression models. Ann Stat 37:995–1034

Koul HL, Ni PP (2004) Minimum distance regression model checking. J Stat Plan Inference 119:109–141

Lavergne P, Patilea V (2008) Breaking the curse of dimensionality in nonparametric testing. J Econ 143:103–122

Lavergne P, Patilea V (2012) One for all and all for one: regression checks with many regressors. J Bus Econ Stat 30:41–52

Li LZ, Zhu XH, Zhu LX (2021) Adaptive-to-model hybrid of tests for regressions. J Am Stat Assoc 1–10

Li KC (1991) Sliced inverse regression for dimension reduction. J Am Stat Assoc 86:316–327

Li LX (2009) Exploiting predictor domain information in sufficient dimension reduction. Comput Stat Data Anal 94:603–613

Li LX, Li B, Zhu LX (2010) Groupwise dimension reduction. J Am Stat Assoc 105:1188–1201

Liang H, Liu X, Li RZ, Tsai CL (2010) Estimation and testing for partially linear single-index models. Ann Stat 38:3811–3836

Lu J, Zhu XH, Lin L, Zhu LX (2019) Estimation for biased partial linear single index models. Comput Stat Data Anal 139:1–13

Robinson PM (1988) Root-n-consistent semiparametric regression. Econometrica 56:931–954

Serfling RJ (1980) Approximation theorems of mathematical statistics. Wiley, New York

Stute W (1997) Nonparametric model checks for regression. Ann Stat 25:613–641

Stute W, Zhu LX (2005) Nonparametric checks for single-index models. Ann Stat 33:1048–1083

Stute W, Manteiga WG, Quindimil MP (1998) Bootstrap approximation in model checks for regression. J Am Stat Assoc 93:141–149

Stute W, Xu WL, Zhu LX (2008) Model diagnosis for parametric regression in high dimensional spaces. Biometrika 95:1–17

Tan FL, Zhu LX (2019) Adaptive-to-model checking for regressions with diverging number of predictors. Ann Stat 47:1960–1994

Tan FL, Zhu XH, Zhu LX (2018) A projection-based adaptive-to-model test for regressions. Stat Sin 28:157–188

Van Keilegom I, Gonzáles-Manteiga W, Sánchez Sellero C (2008) Goodness of fit tests in parametric regression based on the estimation of the error distribution. TEST 17:401–415

Wang JL, Xue LG, Zhu LX, Chong YS (2010) Estimation for a partial-linear single-index model. Ann Stat 38:246–274

Wu CF (1986) Jackknife, bootstrap and other resampling methods in regression analysis. Ann Stat 14:1261–1295

Wu Y, Li L (2011) Asymptotic properties of sufficient dimension reduction with a diverging number of predictors. Stat Sin 21:707–730

Xia YC (2007) A constructive approach to the estimation of dimension reduction directions. Ann Stat 35:2654–2690

Xia Y, Härdle W (2006) Semi-parametric estimation of partially linear single-index model. J Multivar Anal 97:1162–1184

Xia Y, Tong H, Li WK, Zhu LX (2002) An adaptive estimation of dimension reduction space. J Roy Stat Soc B 64(3):363–388

Zhao J, Zhao Y, Lin J, Miao Z, Khaled W (2020) Estimation and testing for panel data partially linear single-index models with errors correlated in space and time. Random Matrices Theory Appl 9:2150005

Zheng JX (1996) A consistent test of functional form via nonparametric estimation techniques. J Econ 75:263–289

Zhou JK, Wu JR, Zhu LX (2016) Overlapped groupwise dimension reduction. Sci China Math 59:2543–2560

Zhu XH, Zhang QM, Zhu LX, Zhang J, Yu LY (2022) Specification testing of regression models with mixed discrete and continuous predictors. J Bus Econ Stat 1–39 (just-accepted)

Zhu LX (2003) Model checking of dimension-reduction type for regression. Stat Sin 13:283–296

Zhu LX, Xue LG (2006) Empirical likelihood confidence regions in a partially linear single-index model. J Roy Stat Soc B 68:549–570

Zhu XH, Zhu LX (2018) Significance testing in nonparametric regression based on dimension reduction. Electron J Stat 12:1468–1506

Zhu XH, Guo X, Zhu LX (2017) An adaptive-to-model test for partially parametric single-index models. Stat Comput 27:1193–1204

Zhu XH, Guo X, Wang T, Zhu LX (2020) Dimensionality determination: a thresholding double ridge ratio criterion. Comput Stat Data Anal 146:106910

Zhu XH, Lu J, Zhang J, Zhu LX (2021) A groupwise dimension reduction adaptive-to-model test for conditional independence. Scand J Stat 48:549–576