

# Integrated conditional moment test for partially linear single index models incorporating dimension-reduction

Shujie Ma\*

*Department of Statistics  
University of California  
Riverside, CA 92521, USA  
e-mail: [shujie.ma@ucr.edu](mailto:shujie.ma@ucr.edu)*

Jun Zhang†

*Shen Zhen-Hong Kong Joint Research Center for Applied Statistical Sciences  
College of Mathematics and Computational Science  
Institute of Statistical Sciences at Shenzhen University, Shenzhen, 518060, China  
e-mail: [zhangjunstat@gmail.com](mailto:zhangjunstat@gmail.com)*

Zihua Sun‡

*Department of Mathematics  
University of Chinese Academy of Sciences  
Beijing, China  
e-mail: [sunzh@amss.ac.cn](mailto:sunzh@amss.ac.cn)*

and

Hua Liang§

*Department of Statistics  
George Washington University  
801 22nd St. NW  
Washington, D.C. 20052, USA  
e-mail: [hliang@gwu.edu](mailto:hliang@gwu.edu)*

**Abstract:** Studying model checking problems for partially linear single-index models, we propose a variant of the integrated conditional moment test using a linear projection weighting function, which gains dimension reduction and makes the proposed method act as if there exists only one covariate even in the presence of multiple dimensional regressors. We derive asymptotic distributions of the proposed test; i.e., an integral of a

---

\*Ma's research was supported by NSF grant DMS 1306972.

†Zhang's research was supported by Natural Science Foundation of SZU (801, 00036112), the NSFC (Tianyuan fund for Mathematics, NO.11326179) of China, and NSFC grant 11101157 of China.

‡Sun's research was supported by the President Fund of University of Chinese Academy of Sciences.

§Corresponding author. Liang's research was partially supported by NSF grants DMS-1007167 and DMS-1207444, and by Award Number 11228103, made by National Natural Science Foundation of China.

centered Gaussian process under the null hypothesis and an integral of a non-centered one under Pitman local alternatives. We also suggest a consistent bootstrap procedure for calculating the critical values. Simulation studies are conducted to demonstrate the performance of the proposed procedure and a real example is analyzed for an illustration.

**AMS 2000 subject classifications:** Primary 62G08; secondary 62G20, 62J02, 62F12.

**Keywords and phrases:** Conditional moment test, curse of dimensionality, empirical process, dimension reduction, estimating function method, linear projection, projection direction, smoothing-based tests.

Received March 2013.

## 1. Introduction

One common task of model regression is to estimate a conditional mean function. Single index models, a generalization of multiple linear regression models with an unknown link function, have been widely used to estimate a conditional mean function because they relax restrictive assumptions imposed on parametric models of conditional mean functions such as linear or generalized linear models (Ichimura; 1993; Härdle et al.; 1993), and therefore gain more flexibility. There are various estimation procedures for single-index models. See Horowitz (2009) for a comprehensive survey and various applications of single-index models. To further combine the interpretability of multiple linear models with the flexibility of single-index models, their hybrid, the partially linear single-index models (PLSiM), have been studied and applied for analyzing various complex data generated from biological and economic studies in the literature (Xia and Härdle; 2006; Yu and Ruppert; 2002; Wang et al.; 2010; Liang et al.; 2010). The first remarkable work on PLSiM can be traced back to Carroll et al. (1997), in which a backfitting algorithm was proposed to estimate parameters of interest in a more general case. Yu and Ruppert (2002) suggested a penalized spline estimation procedure. Xia and Härdle (2006) applied the minimum average variance estimation (MAVE, Xia et al.; 2002) to PLSiM and developed an effective algorithm. More recently, Wang et al. (2010) studied estimation in PLSiM with additional assumptions imposed on the model structure. Liang et al. (2010) proposed a profile least squares estimation procedure. Using PLSiM for data analysis, a natural question is that, given a set of covariates, how one feels confident that the model fits data well. In this paper, we propose a method for model checking in PLSiM integrating the dimension reduction principle.

Correct model specification is fundamental and critical in data analysis. There is a vast literature on model checking problems. Most results were established by evaluating the difference of the conditional expectation under the null and alternative hypotheses, or the expectation of the residual under the null hypothesis. Generally speaking, there are two classes of methods for model checking. The first class focuses on the procedure based on smoothing residuals and is called “local approach”, which includes Dette (1999); Härdle et al. (1998); Horowitz and Härdle (1994); Eubank and Hart (1992); Härdle and Mammen

(1993). Hart (1997) gave a comprehensive survey on nonparametric lack-of-fit tests. It is worth mentioning that the null asymptotic distributions of these tests are generally free of the data generating process and therefore are convenient to calculate critical values. However, these methods mostly suffer from “curse of dimensionality”, detect the alternative hypothesis with only nonparametric rate, and need careful selection of bandwidth. Lavergne and Patilea (2008) incorporated a dimension-reduction idea in smoothing the residuals and constructed a test by using the linear projection approach. The resulting test can avoid “curse of dimensionality” and has higher power as they advocated. But they also introduced a penalty function and needed an initial value, which need to be determined in advance. Furthermore, no guideline on how to choose the penalty function is available.

An alternative to the local approach is the conditional moment (CM) test (Newey; 1985; White; 1984), which was inspired by a fact that certain conditional moments can be reexpressed as a finite number of unconditional moment restrictions. As a consequence, the sample versions of these unconditional moments result in various CM tests. Unfortunately, these CM tests are generally not consistent as pointed out by Bierens (1982), because for any tests based on a finite number of moment restrictions, one can always construct a data generating process which guarantees that the null hypothesis is false but the moment restrictions involved hold; that is, such tests are not consistent. As a remedy, Bierens (1982) proposed an integrated conditional moment (ICM) test, which differs from the CM tests such that ICM is still a CM test, but based on uncountable many moments. The idea works as follows.

Let  $\mathbf{X}$  be the covariates and  $Y$  be the response variable,  $\mathbf{x}$  be a vector with the same length as  $\mathbf{X}$  and  $e(\boldsymbol{\lambda}, \mathbf{X}, Y)$  be the model error with an unknown parameter vector  $\boldsymbol{\lambda}$ , that can be infinitely dimensional. The integrated conditional moment (ICM) tests transform the conditional expectation of the null hypothesis  $E(e(\boldsymbol{\lambda}, \mathbf{X}, Y)|\mathbf{X}) = 0$  into uncountable many moments  $E\{e(\boldsymbol{\lambda}, \mathbf{X}, Y)w(\mathbf{X}, \mathbf{x})\} = 0$  for any  $\mathbf{x}$  with the weighting function satisfying the equivalence of the conditional expectation and infinite unconditional moments. There is a large amount of literature on the ICM test and the following five weight functions have been proposed in the literature: exponential weighting function  $\exp(\sqrt{-1}\mathbf{x}^\top \mathbf{X})$  by Escanciano (2006); Bierens (1982); linear indicator weighting function  $I(\mathbf{X}^\top \boldsymbol{\theta} \leq u)$  for  $u \in \mathbb{R}^1$  by Stute and Zhu (2002); Escanciano (2006); logistic weighting function  $\{1 + \exp(\mathbf{x}^\top \mathbf{X})\}^{-1}$  by Lee et al. (2001); simple indicator weighting function  $I(\mathbf{X} \leq \mathbf{x})$  by Stute (1997); Lin et al. (2002); and trigonometric weighting function  $\cos(\mathbf{x}^\top \mathbf{X}) + \sin(\mathbf{x}^\top \mathbf{X})$  by Bierens and Ploberger (1997). But some weighting functions lead to inconsistent model checking methods and different choices of weighting function lead to different power properties. Furthermore, no best weighting function in terms of power is available since Bierens and Ploberger (1997) showed that all these weighting functions lead to asymptotic admissible tests. It is worth pointing out that the tests based on the linear indicator weighting and simple indicator weighting functions are similar to the idea of using integrated moments over half-spaces for goodness of fit, which can be traced back to Beran and Millar (1989).

Among these weighting functions, the linear indicator weighting function is the most attractive because its associated tests avoid choosing the integrating function which is necessary for the logistic, exponential and trigonometric weighting functions. In fact, the introduction of the linear indicator weighting function initially aims to prevent the high-dimensional problem of data which may harass the testing methods based on the simple indicator weighting function. The linear indicator weighing function was first studied by Stute and Zhu (2005, 2002) and Xia et al. (2004) for checking generalized linear models and single-index models, where the projection direction is chosen to be the regression parameter. The rationality of the linear indicator weighting function was justified in Escanciano (2006) on a basis of an important proposition proved by Jones (1987) that the departure can be detected by a projection of the function along a certain direction; that is, for random variables  $\varepsilon$  and  $\mathbf{x}$  with  $E\|\mathbf{x}\| < \infty$ ,  $E(\varepsilon|\mathbf{x}) = 0$  if and only if  $E(\varepsilon|\mathbf{x}^\top\boldsymbol{\theta}) = 0$  almost sure for any unit vector  $\boldsymbol{\theta}$ .

The linear indicator weighing function involves a nuisance parameter, i.e. the projection direction. The choice of the projecting direction is very critical to three pivotal concerns. The first one is that the choice of the projection direction must ensure the equivalence of the null hypothesis and the weighted infinite unconditional moments. Stute and Zhu (2002) and Xia et al. (2004) chose the projection directions as regression parameters and then weakened the null hypothesis into the independence of the residual and the regressors (Escanciano; 2006). The second one is concerned with the power performance of the test. As it has been known that the projection direction should distinguish the projected residual of the alternative hypothesis from that of the null hypothesis along this direction. Otherwise, the corresponding test will lose power completely and thus have a bad performance. Moreover, to improve power performance, the projection direction should detect the difference as possible. The third concern caused by the nuisance projection direction is the calculability of the critical value. The null distribution of the ICM test is case-dependent, and thus a bootstrap method is usually required to define the critical value. The complexity of the projection direction makes the testing statistic difficult to compute and the bootstrap method hard to carry through.

Aiming to avoid the deficiency of the most smoothing based tests and the existing ICM tests, we propose a variant of the ICM test using a linear projection weighting function and choose the projection direction by fitting a single-index model from the estimated squared model errors against all the covariates. Such a method of choosing the projection direction prevents the three problems aforementioned. It will be shown that the proposed method can avoid the problem of data sparsity, is free from the smoothing parameter, is easy to compute and has satisfactory power performance.

The rest of this paper is organized as follows. In Section 2, we describe the proposed test and study its asymptotic properties. We further suggest a bootstrap method for calculating the critical value, and investigate the power performance. Simulation studies and a real data analysis are respectively conducted to evaluate the proposed tests in Sections 3 and 4. The technical details for the

proofs of the main results and the estimation of the parameters under the null hypothesis are presented in the [Appendix](#).

## 2. The test procedures and their theoretical properties

Consider the partially linear single-index model (PLSiM) of the form:

$$E(Y|\mathbf{Z}) = g(\boldsymbol{\beta}^\top \mathbf{X}) + \boldsymbol{\alpha}^\top \mathbf{T}, \tag{2.1}$$

where  $\mathbf{Z}=(\mathbf{X}^\top, \mathbf{T}^\top)^\top$ , and  $\mathbf{X}$  and  $\mathbf{T}$  are the  $p$ -th dimensional and  $q$ -th dimensional covariate vectors, respectively,  $\boldsymbol{\beta}$  is an unknown index vector which belongs to the parameter space  $\Theta = \{\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top : \|\boldsymbol{\beta}\| = 1, \beta_1 > 0, \boldsymbol{\beta} \in \mathbf{R}^p\}$ , where  $\|\boldsymbol{\beta}\| = (\beta_1^2 + \dots + \beta_p^2)^{1/2}$  is the Euclidean norm of  $\boldsymbol{\beta}$ ,  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_q)^\top$ , and  $g(\cdot)$  is an unknown differentiable function. We are interested in checking the specification of the PLSiM:

$$H_0 : E(Y|\mathbf{X} = \mathbf{x}, \mathbf{T} = \mathbf{t}) = g(\boldsymbol{\beta}^\top \mathbf{x}) + \boldsymbol{\alpha}^\top \mathbf{t} \text{ for some } \boldsymbol{\beta}, \boldsymbol{\alpha}, \text{ and } g(\cdot). \tag{2.2}$$

Let  $e(\boldsymbol{\beta}, \boldsymbol{\alpha}, g) = Y - \{g(\boldsymbol{\beta}^\top \mathbf{X}) + \boldsymbol{\alpha}^\top \mathbf{T}\}$ . Note that the null hypothesis is equivalent to  $E\{e(\boldsymbol{\beta}, \boldsymbol{\alpha}, g)|\mathbf{Z} = \mathbf{z}\} = 0$  for any  $\mathbf{z} \in \mathbb{R}^{p+q}$ . Assume that  $E\{e^2(\boldsymbol{\beta}, \boldsymbol{\alpha}, g)|\mathbf{Z} = \mathbf{z}\} = \sigma^2(\mathbf{z}) < \infty$ . Inspired by Jones (1987), we consider  $E\{e(\boldsymbol{\beta}, \boldsymbol{\alpha}, g)|\boldsymbol{\theta}^\top \mathbf{Z}\} = 0$  for any unit vector  $\boldsymbol{\theta}$ . Since  $\boldsymbol{\theta}^\top \mathbf{Z}$  is scalar, this transformation overcomes the dimensional problem; However, such a testing method requires the nonparametric smoothing on the residual and belongs to the category of “local method”. Though it prevents the deficiency of the dimensional problem, it suffers from all other shortages of the local method. We further target the equivalent form  $E\{e(\boldsymbol{\beta}, \boldsymbol{\alpha}, g)I(\boldsymbol{\theta}^\top \mathbf{Z} \leq u)\} = 0$  for any  $u \in \mathbb{R}^1$  and any unit  $(p+q)$ -vector  $\boldsymbol{\theta}$ . This operation can avoid estimating  $E\{e(\boldsymbol{\beta}, \boldsymbol{\alpha}, g)|\boldsymbol{\theta}^\top \mathbf{Z}\}$  nonparametrically and therefore no bandwidth selection is required. By treating  $u$  as a random variable, we formulate our null hypothesis as:  $E[E^2\{e(\boldsymbol{\beta}, \boldsymbol{\alpha}, g)I(\boldsymbol{\theta}^\top \mathbf{Z} \leq u)|\boldsymbol{\theta}, u\}] = 0$ . Based on this equality, we construct a Crámer-von Mises type testing statistic. By considering that the nuisance parameter  $\boldsymbol{\theta}$  follows a degenerated distribution, the Crámer-von Mises type testing statistic can be simplified as a weighted summand. It is worth mentioning that this simplification leads to higher statistical power at the cost of inconsistency. Further, the resulting statistic works as satisfactory under the null hypothesis. The value of  $\boldsymbol{\theta}$  is determined by fitting a single-index model with the squared model error of PLSiM as the response and all the regressors as covariates. The obtained value of  $\boldsymbol{\theta}$  can tell the projected residual of the alternative models from that of the null hypothesis.

Suppose that  $\{(Y_i, \mathbf{X}_i, \mathbf{T}_i), i = 1, \dots, n\}$  is the random sample from  $(Y, \mathbf{X}, \mathbf{T})$ . Based on this sample, we estimate the parameters  $\boldsymbol{\beta}$  and  $\boldsymbol{\alpha}$  and the unknown function  $g$  under the null hypothetical model (2.2). The estimation details and the associated asymptotic properties of the proposed estimator are presented in [Appendix A.2](#). Let  $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}$  and  $\hat{g}$  be the estimators of  $\boldsymbol{\beta}, \boldsymbol{\alpha}$  and  $g$ . Then our

sample version of  $E\{e(\beta, \alpha, g)I(\theta^\top \mathbf{Z} \leq u)\}$  is defined as

$$R_{nw}(u) = \frac{1}{n} \sum_{i=1}^n \left[ Y_i - \left\{ \hat{g} \left( \hat{\beta}^\top \mathbf{X}_i \right) + \hat{\alpha}^\top \mathbf{T}_i \right\} \right] I(\theta^\top \mathbf{Z}_i \leq u). \tag{2.3}$$

To study the asymptotic properties of  $R_{nw}(u)$ , we first introduce the following notation. Let  $\mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^\top$  for a matrix  $\mathbf{A}$ . Let  $V(\beta) = \beta^\top \mathbf{X}$ , and for any random variable (or vector)  $\xi$ ,  $\tilde{\xi}(\beta) = \xi - E\{\xi|V(\beta)\}$ . For example,  $\tilde{\mathbf{X}}(\beta) = \mathbf{X} - E\{\mathbf{X}|V(\beta)\}$ . Let  $\mathbf{J}(\beta) = \frac{\partial \beta}{\partial \beta^{(1)}}$  be the Jacobian matrix of size  $p \times (p-1)$  with

$$\mathbf{J}(\beta) = \begin{pmatrix} -\beta^{(1)\top} / \sqrt{1 - \|\beta^{(1)}\|^2} \\ \mathbf{I}_{p-1} \end{pmatrix}, \beta^{(1)} = (\beta_2, \dots, \beta_p)^\top.$$

Let  $\beta^0$  and  $\alpha^0$  be the true parameters in the null hypothetical model (2.2). For notation simplicity, we denote  $V = V(\beta^0)$ ,  $\tilde{\mathbf{X}} = \tilde{\mathbf{X}}(\beta^0)$ ,  $\mathbf{T} = \mathbf{T}(\beta^0)$  and  $\mathbf{J} = \mathbf{J}(\beta^0)$ . Define

$$\begin{aligned} \Pi_u(\mathbf{Z}) &= I(\theta^\top \mathbf{Z} \leq u) - E\{I(\theta^\top \mathbf{Z} \leq u) | V\} \\ &- E\left[ I(\theta^\top \mathbf{Z} \leq u) \left\{ g'(V) \tilde{\mathbf{X}}^\top \mathbf{J}, \tilde{\mathbf{T}}^\top \right\} \right] \Sigma^{-1} \left\{ g'(V) \tilde{\mathbf{X}}^\top \mathbf{J}, \tilde{\mathbf{T}}^\top \right\}^\top, \end{aligned} \tag{2.4}$$

in which

$$\Sigma = E \left[ \left\{ g'(V) \tilde{\mathbf{X}}^\top \mathbf{J}, \tilde{\mathbf{T}}^\top \right\}^\top \right]^{\otimes 2}. \tag{2.5}$$

**Proposition 2.1.** *Under Assumptions (C1)–(C5) in the Appendix and the null hypothesis  $H_0$ ,  $\sqrt{n}R_{nw}(u)$  converges to  $R_w(u)$  in the Skorohod space  $D[-\infty, \infty]^{p+q}$ , where  $R_w(u)$  is a centered Gaussian process with covariance function*

$$\text{Cov}\{R_w(u_1), R_w(u_2)\} = E\{\Pi_{u_1}(\mathbf{Z})\Pi_{u_2}(\mathbf{Z})\sigma^2(\mathbf{Z})\}.$$

For any fixed  $u$ , Proposition 2.1 shows that  $\sqrt{n}R_{nw}(u)$  is asymptotically normal. A simple test may be constructed by considering  $\{\sqrt{n}R_{nw}(u)\}^2$ , which follows an asymptotically central  $\chi^2$  distribution. But this score-type test may not be consistent since it is based on only one moment condition.

### 2.1. The proposed test

Note that  $E\{e(\beta, \alpha, g)I(\theta^\top \mathbf{Z} \leq u)\}$  is equivalent to

$$E_{\theta, u} \left[ E_{\mathbf{Z}, Y}^2 \{ e(\beta, \alpha, g)I(\theta^\top \mathbf{Z} \leq u) \} | \theta, u \right] = 0$$

for any  $u$  and unit vector  $\theta$ . If we take  $\theta$  as a random variable and denote the distribution of  $\theta$  by  $f(\theta)d\theta$ , then our test statistic is defined as

$$\mathcal{T}_n = \int \{ \sqrt{n}R_{nw}(u) \}^2 f(\theta) F_{n\theta}(du) d\theta, \tag{2.6}$$

in which  $F_{n\theta}(du)$  is the empirical distribution of  $(\theta^\top \mathbf{Z}_1, \dots, \theta^\top \mathbf{Z}_n)$ . Accordingly we have the following result for  $\mathcal{T}_n$ .

**Theorem 1.** *Under Assumptions (C1)–(C5) in the Appendix and the null hypothesis  $H_0$ , one has*

$$\mathcal{T}_n \stackrel{\text{def}}{=} \int \{R_w(u)\}^2 f(\theta) F_\theta(du) d\theta,$$

in which  $R_w(u)$  is defined in Proposition 2.1 and  $F_\theta(du)$  is the distribution of  $\theta^\top \mathbf{Z}$ .

Theorem 1 follows directly from the result of Proposition 2.1 and the continuous mapping theorem.

### 2.2. Calculation of the critical value

Since the asymptotic covariance of the process  $\sqrt{n}R_{nw}(u)$  depends on the variables  $Y$  and  $\mathbf{Z}$ , the test statistic  $\mathcal{T}_n$  defined in (2.6) is case-dependent and the critical value cannot be obtained directly based on this distribution. We use a bootstrap procedure to mimic the null distribution of the test statistic.

The procedure for calculating the critical value based on the bootstrap test statistic is as follows:

- Step 1: Compute the estimated projection direction  $\hat{\theta}$  by fitting a single-index model with “synthesis” data  $\{e_i^2(\hat{\alpha}, \hat{\beta}, \hat{g}), \mathbf{Z}_i\}_{i=1}^n$ , where  $e_i(\hat{\beta}, \hat{\alpha}, \hat{g}) = Y_i - \hat{\alpha}^\top \mathbf{T}_i - \hat{g}(\hat{\beta}^\top \mathbf{X}_i)$ ,  $1 \leq i \leq n$ .
- Step 2: Compute the test statistic  $\mathcal{T}_n = \frac{1}{n^2} \sum_{r=1}^n \left\{ \sum_{i=1}^n e_i(\hat{\beta}, \hat{\alpha}, \hat{g}) I(\hat{\theta}^\top \mathbf{Z}_i \leq \hat{\theta}^\top \mathbf{Z}_r) \right\}^2$ .
- Step 3: Generate the random variable sequence  $\{\epsilon_{ib}\}_{i=1}^n$ ,  $b = 1, \dots, B$  from the two-point distribution which respectively takes values  $\frac{1 \mp \sqrt{5}}{2}$  with probability  $\frac{5 \pm \sqrt{5}}{10}$ , so that the variance equals to 1, and compute the following arguments for each  $b$ :

$$\left\{ e_1(\hat{\beta}, \hat{\alpha}, \hat{g})\epsilon_{1b}, e_2(\hat{\beta}, \hat{\alpha}, \hat{g})\epsilon_{2b}, \dots, e_n(\hat{\beta}, \hat{\alpha}, \hat{g})\epsilon_{nb} \right\},$$

$$Y_i^b = \hat{g}(\hat{\beta}^\top \mathbf{X}_i) + \hat{\alpha}^\top \mathbf{T}_i + e_i(\hat{\beta}, \hat{\alpha}, \hat{g})\epsilon_{ib}, \quad i = 1, \dots, n.$$

- Step 4: For each  $b$ , we re-calculate the bootstrap estimators  $\hat{\alpha}_b$  and  $\hat{\beta}_b$ . Then we calculate the bootstrap fitted value and residuals, and we further define the bootstrap test statistic

$$\begin{aligned} \mathcal{T}_n^b &= \int \{ \sqrt{n}R_{nw}^b(u) \}^2 f(\theta) F_{n\theta}(du) d\theta \\ &= \frac{1}{n^2} \sum_{r=1}^n \left\{ \sum_{i=1}^n e_i^b(\hat{\beta}_b, \hat{\alpha}_b, \hat{g}_b) I(\hat{\theta}^\top \mathbf{Z}_i \leq \hat{\theta}^\top \mathbf{Z}_r) \right\}^2, \end{aligned}$$

where

$$R_{nw}^b(u) = \frac{1}{n} \sum_{i=1}^n \left\{ e_i^b(\widehat{\boldsymbol{\beta}}_b, \widehat{\boldsymbol{\alpha}}_b, \widehat{g}_b) \right\} I(\widehat{\boldsymbol{\beta}}^\top \mathbf{Z}_i \leq u), \quad (2.7)$$

$e_i^b(\widehat{\boldsymbol{\beta}}_b, \widehat{\boldsymbol{\alpha}}_b, \widehat{g}_b)$  is defined as  $Y_i^b - \widehat{g}_b(\widehat{\boldsymbol{\beta}}_b^\top \mathbf{X}_i) - \widehat{\boldsymbol{\alpha}}_b^\top \mathbf{T}_i$  and  $\widehat{g}_b(\cdot)$  is obtained as nonparametric smoothing from  $Y_i^b - \widehat{\boldsymbol{\alpha}}_b^\top \mathbf{T}_i$  against  $\widehat{\boldsymbol{\beta}}_b^\top \mathbf{X}_i$ .

Step 5: We calculate the  $1 - \kappa$  quantile of the bootstrap test statistic as the  $\kappa$ -level critical value.

**Theorem 2.** Under Assumptions (C1)–(C5) in the Appendix and the null hypothesis  $H_0$ , one has for all  $b = 1, \dots, B$ ,

$$\mathcal{T}_n^b \stackrel{\text{def}}{=} \int \{R_w(u)\}^2 f(\boldsymbol{\theta}) F_{\boldsymbol{\theta}}(du) d\boldsymbol{\theta},$$

where  $R_w(u)$  is defined in Proposition 2.1.

### 2.3. Local power properties

To investigate the sensitivity of the proposed test, we consider the alternative hypothetical models

$$H_{1n} : Y_i = g(\boldsymbol{\beta}^\top \mathbf{X}_i) + \boldsymbol{\alpha}^\top \mathbf{T}_i + \delta_n D(\mathbf{X}_i, \mathbf{T}_i) + \eta_i, i = 1, \dots, n \quad (2.8)$$

with  $E(\eta_i | \mathbf{X}_i, \mathbf{T}_i) = 0$  and some arbitrary bounded measurable function  $D(\cdot, \cdot)$ .

**Theorem 3.** Under Assumptions (C1)–(C5) and the alternatives (2.8), if  $n^\gamma \delta_n \rightarrow a$  for  $a \neq 0$  and  $0 \leq \gamma < 1/2$ ,  $\mathcal{T}_n$  converges to  $\infty$ ; if  $n^{1/2} \delta_n \rightarrow 1$ , then  $\mathcal{T}_n \stackrel{\text{def}}{=} \int \{R_w(u) + \Omega\}^2 f(\boldsymbol{\theta}) F_{\boldsymbol{\theta}}(du) d\boldsymbol{\theta}$ , and  $E(\mathcal{T}_n) \rightarrow \Omega^2$  with

$$\begin{aligned} \Omega &= E \left\{ I(\boldsymbol{\theta}^\top \mathbf{Z} \leq u) D(\mathbf{X}, \mathbf{T}) \right\} - E \left\{ I(\boldsymbol{\theta}^\top \mathbf{Z} \leq u) \left( g'(V) \widetilde{\mathbf{X}}^\top \mathbf{J}, \widetilde{\mathbf{T}}^\top \right)^\top \right\} \\ &\quad \times \boldsymbol{\Sigma}^{-1} E \left[ \left( g'(V) \widetilde{\mathbf{X}}^\top \mathbf{J}, \widetilde{\mathbf{T}}^\top \right)^\top D(\mathbf{X}, \mathbf{T}) \right]. \end{aligned}$$

**Theorem 4.** Under Assumptions (C1)–(C5) and the alternatives (2.8) with  $n^{1/2} \delta_n \rightarrow 1$ , the conditional distributions of  $\mathcal{T}_n^b$  converge to the limiting null distribution of  $\mathcal{T}_n$  for almost all sample sequences  $\{(Y_1, \delta_1, \mathbf{X}_1, \mathbf{T}_1), \dots, (Y_n, \delta_n, \mathbf{X}_n, \mathbf{T}_n), \dots\}$ .

## 3. A simulation study

In this section, we report simulation results to evaluate the performance of the proposed testing procedure.

**Example 1.** We consider the data generating process (DGP) given as follows:

$$Y_i = 0.5 \exp\left(0.5\boldsymbol{\beta}^\top \mathbf{X}_i\right) + \boldsymbol{\alpha}^\top \mathbf{T}_i + C_o \left(\sum_{r=1}^p |X_{ri}| + |\boldsymbol{\alpha}^\top \mathbf{T}_i|\right) + e_i, \quad (3.1)$$

where  $i = 1, \dots, n$ . We consider two different sample sizes:  $n = 200$  and  $400$ . The number of replications in all the simulations is 500. Clearly, the null hypothesis given in (2.2) is true if and only if  $C_o = 0$  for model (3.1). In this example, we let  $\boldsymbol{\alpha} = (2, 2)^\top$ ,  $\boldsymbol{\beta} = (2, 3, 4)^\top / \sqrt{29}$ .  $e_i$  is generated from  $N(0, 0.1^2)$ , independently with  $(\mathbf{X}_i^\top, \mathbf{T}_i^\top)^\top$ . We generate the covariates  $(\mathbf{X}_i^\top, \mathbf{T}_i^\top)^\top$  and select a sequence of values for  $C_o$  in two cases:

- Case 1.  $(\mathbf{X}_i^\top, \mathbf{T}_i^\top)^\top$  is generated from the multivariate normal distribution  $N_5(0, \Sigma)$ , where  $\Sigma = (\sigma_{jj'})$  with  $\sigma_{jj'} = 0.1^{|j-j'|}$ ,  $C_o$  is set to be 0.00, 0.06, 0.12, 0.18, 0.24, respectively.
- Case 2. We let  $\mathbf{T}_i = (T_{i1}, T_{i2})^\top = \Sigma_0^{1/2}(S_{i1}, S_{i2})^\top$ , where  $S_{i1}$  is generated from the binomial distribution with  $P(S_{i1} = 0) = 0.5$  and  $P(S_{i1} = 1) = 0.5$ , and  $S_{i2}$  is generated from the normal distribution  $N(0, 1)$ , independently with  $S_{i1}$ . We let  $\Sigma_0 = (\sigma_{0,jj'})$  with  $\sigma_{0,jj'} = 0.5^{|j-j'|}$ .  $\mathbf{X}_i$  is generated from the multivariate normal distribution  $N_3(0, \Sigma_{\mathbf{X}})$  with  $\Sigma_{\mathbf{X}} = (\sigma_{\mathbf{x},jj'})$ ,  $\sigma_{\mathbf{x},jj'} = 0.1^{|j-j'|}$ , independently with  $\mathbf{T}_i$ .  $C_o$  is set to be 0.00, 0.04, 0.08, 0.12, 0.16, respectively.

We examine the proposed test procedure for the PLSiM under a sequence of alternative models with different values for  $C_o$  given in Cases 1 and 2. When  $C_o = 0$ , the model is the null model. In this example, we use the estimated projection direction  $\hat{\boldsymbol{\theta}}$  by fitting a single-index model with the squared model error  $e_i^2(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}, \hat{g})$  of (3.1) as the response and  $\mathbf{Z}_i$  as covariates.

The powers of the tests are calculated at the four significant levels: 0.01, 0.025, 0.05, and 0.10. For the bandwidth selection, the conditions of Lemma A.4 in the Appendix include the optimal bandwidth for  $h$  in the process of estimating  $(\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top$ . Thus, the standard bandwidth selection methods, such as  $K$ -fold cross validation, validation (CV), generalized cross validation (GCV) or the rule of thumb, can be employed. In this example, we use 5-fold cross validation as it is not too computationally intensive suggested by Cui et al. (2011). After obtaining these estimators  $(\hat{\boldsymbol{\alpha}}^\top, \hat{\boldsymbol{\beta}}^\top)^\top$ ,  $\hat{g}(\cdot)$  is further calculated to obtain the residual  $e(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}, \hat{g})$ . From Condition (C3), we know that bandwidth  $h$  needs undersmoothing. Thus, we can use the bandwidth  $\tilde{h} = \hat{h} \times n^{-2/15}$ , where  $\hat{h}$  is the one estimated by the 5-fold cross validation, for obtaining  $(\hat{\boldsymbol{\alpha}}^\top, \hat{\boldsymbol{\beta}}^\top)^\top$ . The power pattern in this simulation study is depicted in Figure 1 with dotted and solid lines for  $n = 200$  and  $n = 400$  by using  $\tilde{h}$ , from which it is easy to see that when the null hypothesis  $H_0$  is true, that is,  $C_o = 0$ , the percentages of  $H_0$  being rejected are close to the corresponding nominal level for all four nominal levels, and they are closer to the significance levels for larger sample size  $n$ . When  $H_0$  is not true, that is,  $C_o \neq 0$ , as the value of  $C_o \neq 0$  increases or the sample size  $n$  increases, the empirical percentages of rejecting  $H_0$  approach

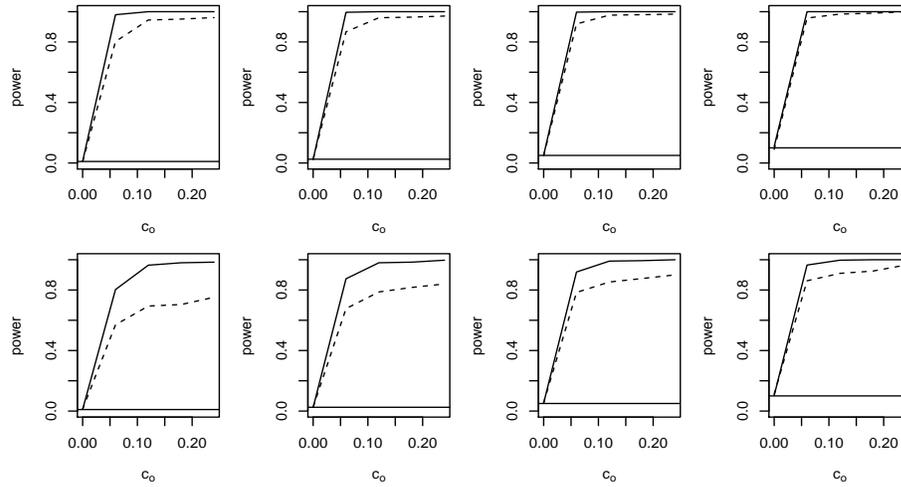


FIG 1. Powers calculation for Case 1(upper panel) and Case 2 (lower panel) with  $n = 200$  (dotted lines) and  $n = 400$  (solid lines). Columns from left to right correspond to nominal levels 0.01, 0.025, 0.05, and 0.1 represented by horizontal lines, respectively.

to 1. Clearly, as the value of  $C_o$  increases, the power increases rapidly. Moreover, the powers for  $n = 400$  are larger than the powers for  $n = 200$  for the same  $C_o$  value and significance level. For  $n = 400$  and large  $C_o$ , the powers are close to 1. The results demonstrate that our proposed testing procedure is powerful.

#### 4. Real data analysis

In this section, we study a dataset from an automobile data set (Johnson; 2003) to discover possible factors that influence the price. The suggested retail price (the manufacturer's assessment of the vehicle's value, including adequate profit for the automaker and the dealer) to serve as the response variable,  $Y$ . Four binary variables are used to indicate the type of a vehicle including: sports car,  $T_1$  (1=yes, 0=no); sport utility vehicle,  $T_2$  (1=yes, 0=no); wagon,  $T_3$  (1=yes, 0=no); and minivan,  $T_4$  (1=yes, 0=no). In addition, two other binary variables  $T_5$  (1=yes, 0=no) and  $T_6$  (1=yes, 0=no) are used to indicate whether the car/truck is all-wheel drive and rear-wheel drive, respectively. Since the number of cylinders takes the values 3, 4, 5, 6, 8 and 12, we created 6 binary variables  $T_7$ ,  $T_8$ ,  $T_9$ ,  $T_{10}$ ,  $T_{11}$  and  $T_{12}$  to indicate whether the number of cylinders is 3, 4, 5, 6, 8 and 12, respectively. All these binary variables mentioned above are treated as the explanatory variables in the linear part of the PLSiM. Moreover, 6 additional measurements are considered including engine size,  $X_1$ ; horsepower,  $X_2$ ; weight in pounds,  $X_3$ ; wheel base in inches,  $X_4$ ; length in inches,  $X_5$ ; and width in inches,  $X_6$ . These continuous variables are considered as the index variables in the PLSiM. We have a total of 18 explanatory variables and 386 observations by removing 41 missing values and one large outlier from the original data set with 428 observations. The PLSiM (2.1) investigated in this paper is applied

TABLE 1  
The Results for real data example

Variable	Coefficient	Standard Error	Absolute value*	<i>p</i> -value
$T_1$	0.3085	0.1016	3.0370	$2.3890 \times 10^{-3}$
$T_2$	-0.2661	0.0979	2.7174	$6.5787 \times 10^{-3}$
$T_3$	-0.0971	0.0860	1.1296	$2.5865 \times 10^{-1}$
$T_4$	0.0934	0.1158	0.8062	$4.2008 \times 10^{-1}$
$T_5$	0.1399	0.0704	1.9848	$4.7166 \times 10^{-2}$
$T_6$	0.3983	0.0711	5.5991	$2.1551 \times 10^{-8}$
$T_7$	0.0773	0.9784	0.0790	$9.3704 \times 10^{-1}$
$T_8$	0.3441	0.1787	1.9256	$5.4149 \times 10^{-2}$
$T_9$	0.1692	0.0833	2.0313	$2.2227 \times 10^{-2}$
$T_{10}$	0.4006	0.1701	2.3549	$1.8526 \times 10^{-2}$
$T_{11}$	1.3867	0.5308	2.6123	$8.9936 \times 10^{-3}$
$T_{12}$	0.3253	0.3065	1.0612	$2.8860 \times 10^{-1}$
$X_1$	-0.1358	0.0716	1.8954	$5.8034 \times 10^{-2}$
$X_2$	0.6837	0.0489	13.9714	$< \times 10^{-12}$
$X_3$	0.5338	0.0529	10.0810	$< \times 10^{-12}$
$X_4$	-0.2366	0.0586	4.0391	$5.3655 \times 10^{-5}$
$X_5$	0.2494	0.0697	3.3438	$8.2629 \times 10^{-4}$
$X_6$	-0.3384	0.0498	6.9174	$4.5988 \times 10^{-12}$

\* The absolute value is calculated by the absolute value of coefficient/standard error

to fit this dataset and the estimated coefficients, the associated standard errors and *p*-values are given in Table 1.

We first applied the proposed testing procedure to check whether the assumed semiparametric PLSiM fits the data adequately. We treated “wheel base in inches ( $X_3$ )” as the leave-one-component for implementing the EFM procedure to estimate  $\theta$ . We generated 500 wild bootstrap replications, and obtained the associated *p* value 0.56. This indicates that the PLSiM is appropriate for fitting this dataset. After we use the EFM method to estimate  $(\hat{\alpha}^\top, \hat{\beta}^\top)^\top$ , we can estimate  $g(\cdot)$ . Figure 2 shows the scatter plots of  $Y - \hat{\alpha}^\top \mathbf{T}$  against the estimated single-index variants  $\hat{\beta}^\top \mathbf{X}$ , along with the estimated curve  $\hat{g}(\cdot)$  by local linear smoothing. It indicates that a nonlinear curve  $g(v)$  may be considered for this dataset. To support our conjecture, we also linearly fit  $Y - \hat{\alpha}^\top \mathbf{T}$  against  $\hat{\beta}^\top \mathbf{X}$ , and draw the straight line in Figure 2, which is not entirely encapsulated in the 95% pointwise confidence bands of the nonlinear curve estimate  $\hat{g}(\cdot)$ . This again suggests that a linear regression is not enough. Figure 2 also shows that a larger value of the estimated index variable composed from the car’s efficiency and power yields a higher retail price. As a result, the PLSiM is sensible and useful for the manufacturer’s suggested retail price.

### 5. Discussions

We have proposed a convenient testing procedure to check whether PLSiM fit complex datasets properly, and studied asymptotic properties of the proposed test statistic. Our numerical experiments indicate that the proposed testing procedure is a promising tool.

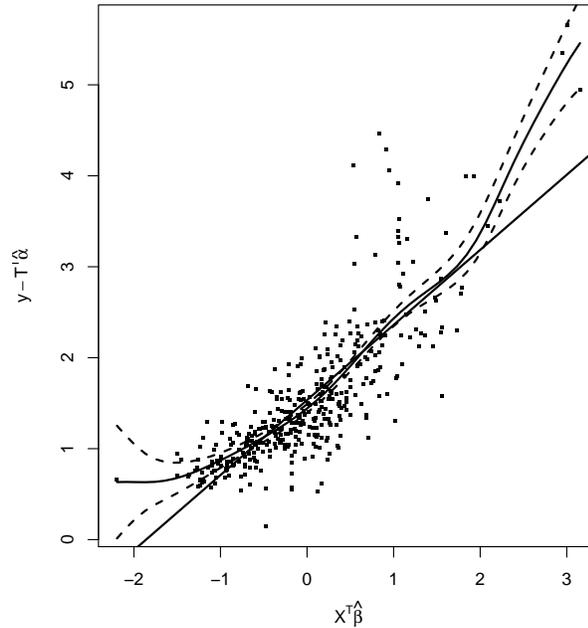


FIG 2. Real data example: Scatterplot of  $Y - \hat{\alpha}^\top \mathbf{T}$  versus the estimated single-index  $\hat{\beta}^\top \mathbf{X}$  (squares); local linear estimated curve of  $g(\cdot)$  (solid line) with the associated 95% pointwise confidence intervals (broken lines) along with a linear regression (solid straight line).

For testing nonparametric models or semi-parametric models, the conventional CV, GCV and 5-fold procedures of bandwidth selection are generally suggested under the null hypothesis (Härdle et al.; 2004). However, how to select bandwidth under the alternative hypothesis is not trivial and deserves a further investigation for true model is generally unavailable. In our numerical experiments, we have simply used the same bandwidth selection strategy as under the null hypothesis.

Some further studies from this line of work include: (i) the dimension of covariates is diverging with the sample size, and (ii) the response variable is not continuous.

### Acknowledgements

The authors thank the Editor, an Associate Editor and two referees for their insightful comments and suggestions that lead to substantially improve the paper.

### Appendix

#### A.1. Assumptions

We begin this section by listing the conditions needed in the proofs of the main results.

- (C1)  $g(\cdot)$  has two bounded and continuous derivatives.
- (C2) The density function  $f_{V(\beta)}(\cdot)$  of the random variable  $V(\beta) = \beta^\top \mathbf{X}$  is bounded away from 0 on  $S_\beta$  for  $\beta$  in a neighborhood of  $\beta^0$  and satisfies the Lipschitz condition of order 1 on  $S_\beta$ , where  $S_\beta = \{\beta^\top \mathbf{x}, \mathbf{x} \in S\}$  and  $S$  is a compact support set of  $\mathbf{X}$ .
- (C3) The kernel  $K$  is a bounded and symmetric density function with a bounded derivative, and satisfies

$$\int v^2 K(v) dt \neq 0, \int |v|^j K(v) dv < \infty, j = 1, 2, \dots$$

- (C4)  $nh^3 \rightarrow \infty, nh^4 \rightarrow 0$ , and  $h \rightarrow 0$ , as  $n \rightarrow \infty$ , in which  $h$  is the kernel bandwidth for estimating  $g(\cdot)$ .
- (C5)  $E\{\mathbf{J}^\top g'(V)\tilde{\mathbf{X}}\}^{\otimes 2}$  and  $E(\mathbf{T})^{\otimes 2}$  are positive definite.

**A.2. Estimation of  $\beta, \alpha$  and  $g$  under the null hypothesis**

For any vector  $\zeta = (\zeta_1, \dots, \zeta_s) \in \mathbb{R}^s$ , denote the Euclidean norm by  $\|\zeta\| = (\sum_{k=1}^s |\zeta_k|^2)^{1/2}$ . For positive numbers  $a_n$  and  $b_n, n \geq 1$ , let  $a_n \sim b_n$  denote that  $\lim_{n \rightarrow \infty} a_n/b_n = 1$ . Let  $\mathbf{I}_d$  be an identity matrix of dimension  $d \times d$ . In this article, we estimate the unknown parameters  $\beta, \alpha$  and the unknown function  $g$  using the estimating function method (EFM) proposed by Cui et al. (2011). We first choose an identifiable parameterization by eliminating  $\beta_1$ . The parameter space  $\Theta$  can be rearranged to a form  $\{(1 - \sum_{r=2}^p \beta_r^2)^{1/2}, \beta_2, \dots, \beta_p\}^\top : \sum_{r=2}^p \beta_r^2 < 1\}$ .

*A.2.1. The kernel estimating functions for the nonparametric function  $g$*

For a given  $\beta$  and  $\alpha$ , we can estimate  $g(\cdot)$  and  $g'(\cdot)$  by using the local linear estimating functions. Let  $h$  be the bandwidth, and  $K(\cdot)$  be the kernel density function satisfying  $K_h(\cdot) = h^{-1}K(\cdot/h)$ . Denote by  $a$  and  $b$  the values of  $g$  and  $g'$  evaluating at  $v$ , respectively. The local linear approximation for  $g(\beta^\top \mathbf{x})$  in a neighborhood of  $v$  is  $\tilde{g}(\beta^\top \mathbf{x}) = a + b(\beta^\top \mathbf{x} - v)$ . Let  $\varpi = (\beta^\top, \alpha^\top)^\top$ . The estimators  $\hat{g}(v, \varpi)$  and  $\hat{g}'(v, \varpi)$  are obtained by solving the local estimating equations with respect to  $a$  and  $b$  given as:

$$\begin{aligned} \sum_{i=1}^n K_h(\beta^\top \mathbf{X}_i - v) [Y_i - \{\tilde{g}(\beta^\top \mathbf{X}_i) + \alpha^\top \mathbf{T}_i\}] &= 0, \\ \sum_{i=1}^n (\beta^\top \mathbf{X}_i - v) K_h(\beta^\top \mathbf{X}_i - v) [Y_i - \{\tilde{g}(\beta^\top \mathbf{X}_i) + \alpha^\top \mathbf{T}_i\}] &= 0. \end{aligned} \tag{A.1}$$

Let  $\hat{a}(v, \varpi)$  and  $\hat{b}(v, \varpi)$  be the estimators of  $a$  and  $b$  at  $v$ , the local linear estimators of  $g(v)$  and  $g'(v)$  are  $\hat{g}(v, \varpi) = \hat{a}(v, \varpi)$  and  $\hat{g}'(v, \varpi) = \hat{b}(v, \varpi)$ , respectively. Then

$$\hat{g}(v, \varpi) = \frac{K_{20}(v, \varpi) K_{01}(v, \varpi) - K_{10}(v, \varpi) K_{11}(v, \varpi)}{K_{00}(v, \varpi) K_{20}(v, \varpi) - K_{10}^2(v, \varpi)},$$

where  $K_{jl}(v, \varpi) = \sum_{i=1}^n K_h(\beta^\top \mathbf{X}_i - v)(\beta^\top \mathbf{X}_i - v)^j (Y_i - \alpha^\top \mathbf{T}_i)^l$  for  $j = 0, 1, 2$  and  $l = 0, 1$ .

### A.2.2. The estimating equations for the estimators of $\beta$ and $\alpha$

The estimators  $\widehat{\beta}^{(1)} = (\widehat{\beta}_2, \dots, \widehat{\beta}_p)^\top$  and  $\widehat{\alpha}$  of  $\beta^{(1)}$  and  $\alpha$  are obtained by solving the following equations:

$$\sum_{i=1}^n \left( \begin{array}{c} \frac{\partial \widehat{g}(\beta^\top \mathbf{X}_i, \varpi)}{\partial \beta^{(1)}} \\ \frac{\partial \widehat{g}(\beta^\top \mathbf{X}_i, \varpi)}{\partial \alpha} + \mathbf{T}_i \end{array} \right) \left[ Y_i - \left\{ \widehat{g}(\beta^\top \mathbf{X}_i, \varpi) + \alpha^\top \mathbf{T}_i \right\} \right] = 0. \quad (\text{A.2})$$

For any random variable (or vector)  $\xi$ , let  $\widehat{\xi}(\beta) = \xi - \widehat{E}\{\xi|V(\beta)\}$ , where  $\widehat{E}\{\xi|V(\beta)\}$  is the local linear estimator of  $E\{\xi|V(\beta)\}$ . Following the same reasoning as given in Cui et al. (2011), it can be shown that the partial derivatives

$$\frac{\partial \widehat{g}(\beta^\top \mathbf{X}_i, \varpi)}{\partial \beta^{(1)}} = \mathbf{J}(\beta)^\top \widehat{g}'(\beta^\top \mathbf{X}_i, \varpi) \widehat{\mathbf{X}}_i(\beta) \quad \text{and} \quad \frac{\partial \widehat{g}(\beta^\top \mathbf{X}_i, \varpi)}{\partial \alpha} = -\widehat{\mathbf{T}}_i(\beta).$$

Then (A.2) can be written as

$$\widehat{\mathbf{G}}(\beta, \alpha) \equiv \sum_{i=1}^n \left\{ \begin{array}{c} \mathbf{J}(\beta)^\top \widehat{g}'(\beta^\top \mathbf{X}_i, \varpi) \widehat{\mathbf{X}}_i(\beta) \\ \widehat{\mathbf{T}}_i(\beta) \end{array} \right\} \left[ Y_i - \left\{ \widehat{g}(\beta^\top \mathbf{X}_i, \varpi) + \alpha^\top \mathbf{T}_i \right\} \right] = 0. \quad (\text{A.3})$$

We use (A.3) to estimate  $\beta^{(1)}$ , and then use the fact that  $\beta_1 = \sqrt{1 - \|\beta^{(1)}\|^2}$  to obtain  $\widehat{\beta}_1$ . Define

$$\gamma_j = \int u^j K(u) du, \nu_j = \int u^j K^2(u) du, j = 1, 2, \dots$$

Denote  $\varpi^0 = \{(\beta^0)^\top, (\alpha^0)^\top\}^\top$ . Fan (1992) and Ruppert and Wand (1994) have shown the bias and noise terms of  $\widehat{g}(v, \varpi^0)$  satisfying

$$\begin{aligned} E\{\widehat{g}(v, \varpi^0) | \mathbf{Z}\} - g(v) &= 1/2h^2 g''(v) \gamma_2 + o_p(h^2), \\ \widehat{g}(v, \varpi^0) - E\{\widehat{g}(v, \varpi^0) | \mathbf{Z}\} &= f_V^{-1}(v) \\ n^{-1} \sum_{i=1}^n K_h(\beta^\top \mathbf{X}_i - v) e_i(\beta^0, \alpha^0, g) &+ o_p\left(\frac{1}{\sqrt{nh}}\right). \end{aligned} \quad (\text{A.4})$$

### A.3. Preliminary lemmas

**Lemma A.1.** Under Assumptions (C1)–(C3) and the null hypothesis  $H_0$ , if  $h \rightarrow 0$  and  $nh \rightarrow \infty$  as  $n \rightarrow \infty$ , we have, for  $\forall \beta \in \Theta$ , the asymptotic conditional mean squared error of  $\widehat{g}(\beta^\top \mathbf{x}, \varpi^0)$  given as

$$E\left[\left\{\widehat{g}(\beta^\top \mathbf{x}, \varpi^0) - g(\beta^\top \mathbf{x})\right\}^2 | \mathbf{Z}\right] = \left\{1/2h^2 g''(\beta^\top \mathbf{x}) \gamma_2\right\}^2$$

$$+ \nu_0 \sigma^2 / \left\{ n h f_V \left( \boldsymbol{\beta}^\top \mathbf{x} \right) \right\} + o_p \left( h^4 + n^{-1} h^{-1} \right).$$

Lemma A.1 follows from (A.4) directly.

**Lemma A.2.** Under Assumptions (C1)–(C3) and the null hypothesis  $H_0$ , if  $h \rightarrow 0$  and  $nh^3 \rightarrow \infty$  as  $n \rightarrow \infty$ , we have, for  $\forall \boldsymbol{\beta} \in \Theta$ , the asymptotic mean squared error of  $\widehat{g}'(\boldsymbol{\beta}^\top \mathbf{x}, \boldsymbol{\varpi}^0)$  given as

$$\begin{aligned} E \left[ \left\{ \widehat{g}' \left( \boldsymbol{\beta}^\top \mathbf{x}, \boldsymbol{\varpi}^0 \right) - g' \left( \boldsymbol{\beta}^\top \mathbf{x} \right) \right\}^2 \mid \mathbf{Z} \right] &= \left\{ 1/6 h^2 g''' \left( \boldsymbol{\beta}^\top \mathbf{x} \right) \gamma_4 \gamma_2^{-1} \right. \\ &\quad \left. + 1/2 \left( \gamma_4 \gamma_2^{-1} - \gamma_2 \right) h^2 g'' \left( \boldsymbol{\beta}^\top \mathbf{x} \right) f'_V \left( \boldsymbol{\beta}^\top \mathbf{x} \right) / f_V \left( \boldsymbol{\beta}^\top \mathbf{x} \right) \right\}^2 \\ &\quad + \nu_2 \gamma_2^{-2} \sigma^2 / \left\{ n h^3 f_V \left( \boldsymbol{\beta}^\top \mathbf{x} \right) \right\} + o_p \left( h^4 + n^{-1} h^{-3} \right). \end{aligned}$$

Lemma A.2 is a consequence of the derivation of bias and variance for  $\widehat{g}'(\boldsymbol{\beta}^\top \mathbf{x}, \boldsymbol{\varpi}^0)$ , which is similar to that of Carroll et al. (1998). Following the same reasoning as the proofs for (2.4) in Cui et al. (2011), one can obtain the following lemma.

**Lemma A.3.** Under Assumptions (C1)–(C3) and the null hypothesis  $H_0$ , if  $h \rightarrow 0$  and  $nh^3 \rightarrow \infty$  as  $n \rightarrow \infty$ , then

$$E \left[ \left\| \frac{\partial \widehat{g} \left( \boldsymbol{\beta}^{0\top} \mathbf{x}, \boldsymbol{\varpi}^0 \right)}{\partial \boldsymbol{\beta}^{(1)}} - g' \left( \boldsymbol{\beta}^{0\top} \mathbf{x} \right) \mathbf{J}^\top \left\{ \mathbf{x} - E \left( \mathbf{x} \mid \boldsymbol{\beta}^{0\top} \mathbf{x} \right) \right\} \right\|^2 \mid \mathbf{Z} \right] = O_p \left( h^4 + n^{-1} h^{-3} \right).$$

**Lemma A.4.** Assume that the estimating equations given in (A.3) have a unique solution, say  $\widehat{\boldsymbol{\varpi}}^{(1)} = \left( \widehat{\boldsymbol{\beta}}^{(1)\top}, \widehat{\boldsymbol{\alpha}}^\top \right)^\top$ . Under Assumptions (C1)–(C3) and (C5) and the null hypothesis  $H_0$ , the following results hold:

- (i) If  $h \rightarrow 0$  and  $(nh)^{-1} \log(1/h) \rightarrow 0$  as  $n \rightarrow \infty$ ,  $\widehat{\boldsymbol{\varpi}}^{(1)}$  converges in probability to the true parameter  $\boldsymbol{\varpi}^{0(1)} = \left( \boldsymbol{\beta}^{0(1)\top}, \boldsymbol{\alpha}^{0\top} \right)^\top$ , where  $\boldsymbol{\beta}^{0(1)} = \left( \beta_2^0, \dots, \beta_p^0 \right)^\top$ ;
- (ii) If  $nh^6 \rightarrow 0$  and  $nh^3 \rightarrow \infty$ ,  $\sqrt{n} \left( \widehat{\boldsymbol{\varpi}}^{(1)} - \boldsymbol{\varpi}^{0(1)} \right) \rightarrow N_{p+q-1} \left( \mathbf{0}, \boldsymbol{\Sigma}^{-1} \boldsymbol{\Upsilon} \boldsymbol{\Sigma}^{-1} \right)$ , where  $\boldsymbol{\Sigma}$  is defined in (2.5) and  $\boldsymbol{\Upsilon} = E \left[ \left\{ g'(V) \widetilde{\mathbf{X}}^\top \mathbf{J}, \widetilde{\mathbf{T}}^\top \right\}^\top \sigma(\mathbf{Z}) \right]^{\otimes 2}$ .

*Proof.* Following the same argument as in the Appendix of Cui et al. (2011), it can be seen that the population version of  $\widehat{\mathbf{G}}(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0)$  in (A.3) is

$$\mathbf{G}(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0) \equiv \sum_{i=1}^n \left\{ \begin{array}{c} \mathbf{J}^\top g' \left( \boldsymbol{\beta}^{0\top} \mathbf{X}_i \right) \widetilde{\mathbf{X}}_i \\ \widetilde{\mathbf{T}}_i \end{array} \right\} \left[ Y_i - \left\{ g \left( \boldsymbol{\beta}^{0\top} \mathbf{X}_i \right) + \boldsymbol{\alpha}^{0\top} \mathbf{T}_i \right\} \right].$$

By Taylor expansion, as  $nh^6 \rightarrow 0$  and  $nh^3 \rightarrow \infty$ , we know

$$\sqrt{n} \left( \widehat{\boldsymbol{\varpi}}^{(1)} - \boldsymbol{\varpi}^{0(1)} \right) = n^{-1/2} \boldsymbol{\Sigma}^{-1} \mathbf{G}(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0) + o_p(1),$$

$$= n^{-1/2} \boldsymbol{\Sigma}^{-1} \sum_{i=1}^n \left\{ \begin{array}{c} \mathbf{J}^\top g'(V_i) \tilde{\mathbf{X}}_i \\ \tilde{\mathbf{T}}_i \end{array} \right\} e_i(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, g) + o_p(1). \quad (\text{A.5})$$

Lemma A.4 follows from (A.5) and the fact that  $E\{\mathbf{G}(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0)\}^{\otimes 2} = n\boldsymbol{\Upsilon}$ .  $\square$

#### A.4. Proofs of the main results

*Proof of Proposition 2.1.* Note that

$$\begin{aligned} \sqrt{n}R_{nw}(u) &= n^{-1/2} \sum_{i=1}^n \left[ Y_i - \left\{ \hat{g}(\hat{\boldsymbol{\beta}}^\top \mathbf{X}_i) + \hat{\boldsymbol{\alpha}}^\top \mathbf{T}_i \right\} \right] I(\boldsymbol{\theta}^\top \mathbf{Z}_i \leq u) \\ &= n^{-1/2} \sum_{i=1}^n \left[ Y_i - \left\{ g(\boldsymbol{\beta}^{0\top} \mathbf{X}_i) + \boldsymbol{\alpha}^{0\top} \mathbf{T}_i \right\} \right] I(\boldsymbol{\theta}^\top \mathbf{Z}_i \leq u) + I_n \\ &= n^{-1/2} \sum_{i=1}^n e_i(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, g) I(\boldsymbol{\theta}^\top \mathbf{Z}_i \leq u) + I_n, \end{aligned}$$

where

$$I_n = n^{-1/2} \sum_{i=1}^n \left[ \left\{ g(\boldsymbol{\beta}^{0\top} \mathbf{X}_i) + \boldsymbol{\alpha}^{0\top} \mathbf{T}_i \right\} - \left\{ \hat{g}(\hat{\boldsymbol{\beta}}^\top \mathbf{X}_i, \hat{\boldsymbol{\beta}}) + \hat{\boldsymbol{\alpha}}^\top \mathbf{T}_i \right\} \right] I(\boldsymbol{\theta}^\top \mathbf{Z}_i \leq u),$$

which can be further decomposed as  $I_{n,1} + I_{n,2}$  with

$$\begin{aligned} I_{n,1} &= n^{-1/2} \sum_{i=1}^n \left\{ g(\boldsymbol{\beta}^{0\top} \mathbf{X}_i, \boldsymbol{\beta}^{0\top}) - \hat{g}(\boldsymbol{\beta}^{0\top} \mathbf{X}_i) \right\} I(\boldsymbol{\theta}^\top \mathbf{Z}_i \leq u), \\ I_{n,2} &= n^{-1/2} \sum_{i=1}^n \left[ \hat{g}(\boldsymbol{\beta}^{0\top} \mathbf{X}_i, \boldsymbol{\beta}^{0\top}) + \mathbf{T}_i^\top \boldsymbol{\alpha}^{0\top} \right. \\ &\quad \left. - \left\{ \hat{g}(\hat{\boldsymbol{\beta}}^\top \mathbf{X}_i, \hat{\boldsymbol{\beta}}) + \mathbf{T}_i^\top \hat{\boldsymbol{\alpha}} \right\} \right] I(\boldsymbol{\theta}^\top \mathbf{Z}_i \leq u). \end{aligned}$$

Recall (A.4) and given  $nh^4 \rightarrow 0$ , we have the following simplification.

$$\begin{aligned} I_{n,1} &= -n^{-1/2} \sum_{i=1}^n \left\{ \frac{1}{2} h^2 g''(\boldsymbol{\beta}^{0\top} \mathbf{X}_i) \gamma_2 + o_p(h^2) + \right. \\ &\quad \left. f_V^{-1}(\boldsymbol{\beta}^{0\top} \mathbf{X}_i) \frac{1}{n} \sum_{j=1}^n K_h(\boldsymbol{\beta}^{0\top} \mathbf{X}_i - \boldsymbol{\beta}^{0\top} \mathbf{X}_j) e_j(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, g) \right\} I(\boldsymbol{\theta}^\top \mathbf{Z}_i \leq u) \\ &= -n^{-1/2} \sum_{i=1}^n f_V^{-1}(\boldsymbol{\beta}^{0\top} \mathbf{X}_i) I(\boldsymbol{\theta}^\top \mathbf{Z}_i \leq u) \frac{1}{n} \sum_{j=1}^n K_h(\boldsymbol{\beta}^{0\top} \mathbf{X}_i - \boldsymbol{\beta}^{0\top} \mathbf{X}_j) \\ &\quad e_j(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, g) + o_p(1) \\ &= -E \left\{ I(\boldsymbol{\theta}^\top \mathbf{Z} \leq u) | V \right\} \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, g) + o_p(1). \end{aligned}$$

By using Lemmas A.3 and A.4 and (A.5), we have that

$$\begin{aligned}
 I_{n,2} &= n^{-1/2} \sum_{i=1}^n \left\{ \begin{array}{c} \mathbf{J}^\top \widehat{g}' \left( \boldsymbol{\beta}^{0\top} \mathbf{X}_i, \boldsymbol{\varpi}^0 \right) \widehat{\mathbf{X}}_i \left( \boldsymbol{\beta}^0 \right) \\ \widehat{\mathbf{T}}_i \left( \boldsymbol{\beta}^0 \right) \end{array} \right\} \\
 &\quad \times \left( \boldsymbol{\varpi}^{(1)} - \widehat{\boldsymbol{\varpi}}^{(1)} \right) I(\boldsymbol{\theta}^\top \mathbf{Z}_i \leq u) + o_p(1) \\
 &= -E \left\{ I(\boldsymbol{\theta}^\top \mathbf{Z}_i \leq u) \left( g'(V) \widetilde{\mathbf{X}}^\top \mathbf{J}, \widetilde{\mathbf{T}}^\top \right)^\top \right\} \sqrt{n} \left( \widehat{\boldsymbol{\varpi}}^{(1)} - \boldsymbol{\varpi}^{(1)} \right) + o_p(1) \\
 &= -E \left\{ I(\boldsymbol{\theta}^\top \mathbf{Z} \leq u) \left( g'(V) \widetilde{\mathbf{X}}^\top \mathbf{J}, \widetilde{\mathbf{T}}^\top \right)^\top \right\} n^{-1/2} \\
 &\quad \times \boldsymbol{\Sigma}^{-1} \sum_{i=1}^n \left\{ \begin{array}{c} \mathbf{J}^\top g' \left( \boldsymbol{\beta}^{0\top} \mathbf{X}_i \right) \widetilde{\mathbf{X}}_i \\ \widetilde{\mathbf{T}}_i \end{array} \right\} \times e_i(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, g) + o_p(1).
 \end{aligned}$$

As a result,

$$\begin{aligned}
 \sqrt{n} R_{nw}(u) &= n^{-1/2} \sum_{i=1}^n e_i(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, g) \left[ I(\boldsymbol{\theta}^\top \mathbf{Z}_i \leq u) - E \left\{ I(\boldsymbol{\theta}^\top \mathbf{Z} \leq u) \mid \boldsymbol{\beta}^{0\top} \mathbf{X} \right\} \right] \\
 &\quad - E \left[ I(\boldsymbol{\theta}^\top \mathbf{Z} \leq u) \left( g'(V) \widetilde{\mathbf{X}}^\top \mathbf{J}, \widetilde{\mathbf{T}}^\top \right)^\top \right] n^{-1/2} \boldsymbol{\Sigma}^{-1} \\
 &\quad \times \sum_{i=1}^n \left\{ \begin{array}{c} \mathbf{J}^\top g' \left( \boldsymbol{\beta}^{0\top} \mathbf{X}_i \right) \widetilde{\mathbf{X}}_i \\ \widetilde{\mathbf{T}}_i \end{array} \right\} e_i(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, g) + o_p(1).
 \end{aligned}$$

Proposition 2.1 follows from the central limit theorem.  $\square$

*Proof of Theorem 2.* Because the moment estimation is consistent, by simple calculation, one has, for  $R_{nw}^b(u)$  defined in (2.7),

$$\sqrt{n} R_{nw}^b(u) = n^{-1/2} \sum_{i=1}^n \epsilon_i e_i(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, g) \Pi_u(\mathbf{Z}_i) + o_p(1).$$

The tightness of  $\epsilon e(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, g) \Pi_u(\mathbf{Z})$  can be obtained by the tightness of  $e_i(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, g) \Pi_u(\mathbf{Z}_i)$ . Theorem 2 can be proved by applying Theorems 2.5.2 and 2.6.7 of van der Vaart and Wellner (1996).  $\square$

*Proof of Theorem 3.* We prove the case that  $\sqrt{n} \delta_n \rightarrow 1$ . The case that  $n^\gamma \delta_n \rightarrow a \neq 0$  ( $0 \leq \gamma \leq 1/2$ ) can be proved in a similar way. Note that under the alternative (2.8) with  $\delta_n = 1/\sqrt{n} + o(1/\sqrt{n})$ , one can verify that (A.4) and Lemmas A.1, A.2 and A.3 still hold for the associated estimator of  $g(\cdot)$ . Let  $\widehat{\boldsymbol{\varpi}}^{(1)}$  be the estimator of  $\boldsymbol{\varpi}^{(1)}$  under (2.8). We know

$$\begin{aligned}
 \sqrt{n} \left( \widehat{\boldsymbol{\varpi}}^{(1)} - \boldsymbol{\varpi}^{(1)} \right) &= n^{-1/2} \boldsymbol{\Sigma}^{-1} \sum_{i=1}^n \left( g'(V_i) \widetilde{\mathbf{X}}_i^\top \mathbf{J}, \widetilde{\mathbf{T}}_i^\top \right)^\top e_i(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, g) \\
 &\quad + \boldsymbol{\Sigma}^{-1} E \left[ \left( g'(V) \widetilde{\mathbf{X}}^\top \mathbf{J}, \widetilde{\mathbf{T}}^\top \right)^\top D(\mathbf{X}, \mathbf{T}) \right] + o_p(1).
 \end{aligned}$$

It follows the same procedures as the proof of Proposition 2.1 that

$$\begin{aligned}
\sqrt{n}R_{nw}(u) &= n^{-1/2} \sum_{i=1}^n e_i(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, g) \left[ I(\boldsymbol{\theta}^\top \mathbf{Z}_i \leq u) - E \left\{ I(\boldsymbol{\theta}^\top \mathbf{Z} \leq u) \mid \boldsymbol{\beta}^{0\top} \mathbf{X} \right\} \right] \\
&\quad - E \left[ I(\boldsymbol{\theta}^\top \mathbf{Z} \leq u) \left( g'(V) \tilde{\mathbf{X}}^\top \mathbf{J}, \tilde{\mathbf{T}}^\top \right)^\top \right] n^{-1/2} \boldsymbol{\Sigma}^{-1} \\
&\quad \times \sum_{i=1}^n \left\{ \begin{array}{c} \mathbf{J}^\top g'(\boldsymbol{\beta}^{0\top} \mathbf{X}_i) \tilde{\mathbf{X}}_i \\ \tilde{\mathbf{T}}_i \end{array} \right\} e_i(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, g) \\
&\quad - E \left[ I(\boldsymbol{\theta}^\top \mathbf{Z} \leq u) \left( g'(V) \tilde{\mathbf{X}}^\top \mathbf{J}, \tilde{\mathbf{T}}^\top \right)^\top \right] \boldsymbol{\Sigma}^{-1} \\
&\quad \times E \left[ \left( g'(V) \tilde{\mathbf{X}}^\top \mathbf{J}, \tilde{\mathbf{T}}^\top \right)^\top D(\mathbf{X}, \mathbf{T}) \right] \\
&\quad + E \left\{ D(\mathbf{X}, \mathbf{T}) I(\boldsymbol{\theta}^\top \mathbf{Z} \leq u) \right\} + o_p(1).
\end{aligned}$$

Then Theorem 3 can be proved by the arguments similar to the proofs of Proposition 2.1 and Theorem 1.  $\square$

*Proof of Theorem 4.* Under the alternative (2.8) with  $n^{1/2}\delta_n \rightarrow 1$ , the asymptotic expansion of  $\sqrt{n}R_{nw}^b(u)$  has a non-random shift  $\Omega$ . It is easy to validate that the random symmetrization variable  $V_i$  in  $\sqrt{n}R_{nw}^b(u)$  makes the effect of such non-random shift vanish. As a result, under the alternative (2.8) with  $n^{1/2}\delta_n \rightarrow 1$ , for  $R_{nw}^b(u)$  defined in (2.7), we have

$$\sqrt{n}R_{nw}^b(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_i e_i(\boldsymbol{\beta}^0, \boldsymbol{\alpha}^0, g) \Pi_u(\mathbf{Z}_i) + o_p(1).$$

Theorem 4 can be proved using the arguments similar as for proving Proposition 2.1 and Theorem 1.  $\square$

## References

- BERAN, R. and MILLAR, P. W. (1989). A stochastic minimum distance test for multivariate parametric models, *The Annals of Statistics* **17**: 125–140. [MR0981440](#)
- BIERENS, H. J. (1982). Consistent model specification tests, *Journal of Econometrics* **20**: 105–134. [MR0685673](#)
- BIERENS, H. J. and PLOBERGER, W. (1997). Asymptotic theory of integrated conditional moment tests, *Econometrica* **65**: 1129–1151. [MR1475077](#)
- CARROLL, R. J., FAN, J., GIJBELS, I. and WAND, M. P. (1997). Generalized partially linear single-index models, *Journal of the American Statistical Association* **92**: 477–489. [MR1467842](#)
- CARROLL, R., RUPPERT, D. and WELSH, A. (1998). Local estimating equations, *Journal of the American Statistical Association* **93**: 214–227. [MR1614624](#)

- CUI, X., HÄRDLE, W. and ZHU, L.-X. (2011). The EFM approach for single-index models, *The Annals of Statistics* **39**: 1658–1688. [MR2850216](#)
- DETTE, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators, *The Annals of Statistics* **27**: 1012–1040. [MR1724039](#)
- ESCANCIANO, J. C. (2006). A consistent diagnostic test for regression models using projections, *Econometric Theory* **22**: 1030–1051. [MR2328527](#)
- EUBANK, R. L. and HART, J. D. (1992). Testing goodness-of-fit in regression via order selection criteria, *The Annals of Statistics* **20**: 1412–1425. [MR1186256](#)
- FAN, J. (1992). Design-adaptive nonparametric regression, *Journal of the American Statistical Association* **87**: 998–1004. [MR1209561](#)
- HÄRDLE, W., HALL, P. and ICHIMURA, H. (1993). Optimal smoothing in single-index models, *The Annals of Statistics* **21**: 157–178. [MR1212171](#)
- HÄRDLE, W. and MAMMEN, E. (1993). Testing parametric versus nonparametric regression, *The Annals of Statistics* **21**: 1926–1947. [MR1245774](#)
- HÄRDLE, W., MAMMEN, E. and MÜLLER, M. (1998). Testing parametric versus semiparametric modeling in generalized linear models, *Journal of the American Statistical Association* **93**: 1461–1474. [MR1666641](#)
- HÄRDLE, W., MÜLLER, M., SPERLICH, S. and WERWATZ, A. (2004). *Nonparametric and Semiparametric Models*, Springer-Verlag, New York. [MR2061786](#)
- HART, J. (1997). *Nonparametric Smoothing and Lack of Fit Tests*, Springer, New York. [MR1461272](#)
- HOROWITZ, J. L. (2009). *Semiparametric and Nonparametric Methods in Econometrics*, Springer, New York. [MR2535631](#)
- HOROWITZ, J. L. and HÄRDLE, W. (1994). Testing a parametric model against a semiparametric alternative, *Econometric Theory* **10**: 821–848. [MR1323277](#)
- ICHIMURA, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *Journal of Econometrics* **58**: 71–120. [MR1230981](#)
- JOHNSON, R. W. (2003). Kiplinger’s personal finance, *Journal of Statistics Education* **57**: 104–123.
- JONES, L. K. (1987). On a conjecture of Huber concerning the convergence of projection pursuit regression, *The Annals of Statistics* **15**: 880–882. [MR0888447](#)
- LAVERGNE, P. and PATILEA, V. (2008). Breaking the curse of dimensionality in nonparametric testing, *Journal of Econometrics* **143**: 103–122. [MR2384435](#)
- LEE, T.-H., WHITE, H. and GRANGER, C. W. J. (2001). Testing for neglected nonlinearity in time series models: a comparison of neural network methods and alternative tests, *Journal of Econometrics* **56**: 208–229.
- LIANG, H., LIU, X., LI, R. and TSAI, C. L. (2010). Estimation and testing for partially linear single-index models, *The Annals of Statistics* **38**: 3811–3836. [MR2766869](#)
- LIN, D. Y., WEI, L. J. and YING, Z. (2002). Model-checking techniques based on cumulative residuals, *Biometrics* **58**: 1–12. [MR1891037](#)
- NEWBY, W. K. (1985). Maximum likelihood specification testing and conditional moment tests, *Econometrica* **53**: 1047–1070. [MR0809903](#)

- RUPPERT, D. and WAND, M. P. (1994). Multivariate locally weighted least squares regression, *The Annals of Statistics* **22**(3): 1346–1370. [MR1311979](#)
- STUTE, W. (1997). Nonparametric model checks for regression, *The Annals of Statistics* **25**: 613–641. [MR1439316](#)
- STUTE, W. and ZHU, L.-X. (2002). Model checks for generalized linear models, *Scandinavian Journal of Statistics* **29**: 535–545. [MR1925573](#)
- STUTE, W. and ZHU, L.-X. (2005). Nonparametric checks for single-index models, *The Annals of Statistics* **33**: 1048–1083. [MR2195628](#)
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence And Empirical Processes*, Springer Series in Statistics, Springer-Verlag, New York. [MR1385671](#)
- WANG, J.-L., XUE, L., ZHU, L. and CHONG, Y. S. (2010). Estimation for a partial-linear single-index model, *The Annals of Statistics* **38**: 246–274. [MR2589322](#)
- WHITE, H. (1984). *Asymptotic Theory for Econometricians*, Academic Press, New York.
- XIA, Y. C. and HÄRDLE, W. (2006). Semi-parametric estimation of partially linear single-index models, *Journal of Multivariate Analysis* **97**: 1162–1184. [MR2276153](#)
- XIA, Y., LI, W. K., TONG, H. and ZHANG, D. (2004). A goodness-of-fit test for single-index models, *Statistica Sinica* **14**: 1–39. [MR2036761](#)
- XIA, Y., TONG, H., LI, W. K. and ZHU, L.-X. (2002). An adaptive estimation of dimension reduction space, *Journal of the Royal Statistical Society, Series B* **64**: 363–410. [MR1924297](#)
- YU, Y. and RUPPERT, D. (2002). Penalized spline estimation for partially linear single-index models, *Journal of the American Statistical Association* **97**: 1042–1054. [MR1951258](#)