Statistica Sinica **29** (2019), 1511-1534
doi:https://doi.org/10.5705/ss.202017.0248

AN ADAPTIVE-TO-MODEL TEST FOR PARAMETRIC SINGLE-INDEX ERRORS-IN-VARIABLES MODELS

Hira L. Koul[1], Chuanlong Xie[2,3] and Lixing Zhu[2,4]

[1]*Michigan State University,* [2]*Hong Kong Baptist University,*
[3]*Jinan University and* [4]*Beijing Normal University*

Abstract: This study provides a useful test for parametric single-index regression models when covariates are measured with errors and validation data are available. The proposed test is asymptotically unbiased, and its consistency rate does not depend on the dimension of the covariate vector. The proposed test behaves like a classical local smoothing test with only one covariate, and retains the omnibus property against general alternatives. This suggests that the proposed test can potentially alleviate the difficulty associated with the curse of dimensionality in this field. Furthermore, a systematic study is conducted to investigate the effect of the ratio between the sample size and the size of the validation data on the asymptotic behavior of these tests. Lastly, simulations are conducted to examine the performance in several finite sample scenarios.

Key words and phrases: Adaptive-to-model test, dimension reduction, errors-in-variables model.

1. Introduction

Consider a nonparametric regression model with measurement errors where the response variable Y, a p-dimensional unobservable predicting covariate X, and its observable cohort vector W are related to each other by the relations

$$Y = \mu(X) + \varepsilon, \quad W = X + U. \tag{1.1}$$

Here p is known, U is independent of ε and X, $E(\varepsilon|X) = 0$, and $E(U) = 0$. This is the so-called nonparametric errors-in-variables (EiV) regression model. The monographs of Fuller (2009), Cheng and Van Ness (1999), and Carroll et al. (2006) contain many real-data examples that employ this model.

Here, the problem of interest is to test a parametric single-index regression model for a regression function. That is, for a known real-valued link function g, we wish to test

$$H_0: \; P(\mu(X) = g(\beta_0^{\mathrm{T}} X, \gamma_0)) = 1, \quad \text{for } \theta_0 = (\beta_0; \gamma_0) \in \mathbb{R}^{p+d},$$

versus $H_1: H_0$ is not true.

A motivation for considering the above testing problem is that, in practice, model checking prevents incorrect conclusions being drawn as a result of an improper model being used. Hart (2013) described several tests for the lack-of-fit of a parametric regression model in the classical regression set up where X is observable. Since the mid 1990s, there has been a significant increase in research activities in this area, as summarized in the recent review by González-Manteiga and Crujeiras (2013).

It is well known that a naive application of inference procedures that are valid in a classical regression, where we replace X by W, often yields inefficient inference procedures for EiV models; see, for example, Fuller (2009) and Carroll et al. (2006). An alternative approach adopted in the literature is that of calibration, where the original regression relationship is transferred to the regression $E(Y|W)$ relationship between the response Y and the cohort W. Zhu, Cui and Ng (2004) established a sufficient and necessary condition for the linearity of $E[Y|W]$ with respect to W when $g(\beta_0^{\mathrm{T}}x, \gamma_0) = \beta_0^{\mathrm{T}}x$. They proposed a score-type lack-of-fitness test based on this fact. This testing procedure has been extended to polynomial EiV models by Cheng and Kukush (2004) and Zhu, Song and Cui (2003) independently. Hall and Ma (2007) proposed a test based on deconvolution methods, assuming that the distribution of the measurement error vector U is known. Zhu and Cui (2005) proposed a test for a general linear model $\beta_0^{\mathrm{T}}h(x)$, where h is a vector of known functions. Song (2008) proposed a test for $\beta_0^{\mathrm{T}}h(x)$ based on a deconvolution kernel density estimator. Koul and Song (2009) developed an analog of the minimum distance tests of Koul and Ni (2004) to fit a parametric form to the regression function for the Berkson measurement error models. Koul and Song (2010) developed tests to fit a parametric function to the nonparametric part of a partial linear regression Berkson measurement error model. All of these works, with the exception of Hall and Ma (2007) and Song (2008), employ the calibration methodology and test to fit the parameter form of the regression function $E[Y|W]$ implied by H_0.

However, there are no valid tests for a parametric model under general conditions in which the distributions of X and U may not be known. This is largely because of the difficulty of estimating the calibrated regression function and some of the other underlying functions involved in the construction of a test statistic. However, it is possible to circumvent some of these difficulties when validation data are available. Stute, Xue and Zhu (2007) used validation data and the empirical likelihood methodology to develop confidence regions for some underlying parameters. Song (2009) developed a test for general EiV models with the assistance of validation data, without assuming any knowledge of the distributions of X or U. Dai, Sun and Wang (2010) constructed a test using validation data for the same model as that of Zhu and Cui (2005). Xu and Zhu (2014) considered a nonparametric test for partial linear EiV models with validation data.

In the classical regression setup, it is known that a common property of lack-of-fit tests used to check a parametric regression model based on the nonparametric smoothing methodology is that the rate of consistency of the test statistics is $n^{-1/2}h^{-p/4}$. That is, the null distribution of a suitably centered and scaled test statistic multiplied by $n^{1/2}h^{p/4}$ has a weak limit. Furthermore these tests can detect local alternatives distinct from the null at the rate $n^{-1/2}h^{-p/4}$. When p is greater than or equal to two, this rate can be very slow. Consequently, for moderate sample sizes, local smoothing tests cannot maintain the significance level well and have low power, even for $p = 2$ or 3. See, Zheng (1996) and Koul and Ni (2004), among others, for further information. We expect this to hold for various local smoothing tests in the EiV setup as well.

The main goal of this study is to propose tests of a dimension-reduction nature when validation data are available and that do not suffer from the aforementioned slow rate of consistency. As such, we proceed as follows. First, we discuss the sufficient dimension-reduction (SDR) technique described in Cook (2009), Li and Yin (2007), and Carroll and Li (1992). The goal is to derive a technique that reduces the dimension of X to a one-dimensional projection $\beta_0^{\top} X$ under the null hypothesis, where β_0 denotes the projection direction in model (1.1), and to $B^{\top} X$ automatically under the alternative, where B is a $p \times q$ orthonormal matrix, with q to be specified. Second, based on the dimension reduction, we construct a test with a consistency rate of $n^{-1/2}h^{-1/4}$ when the size N of the validation data is proportional to or larger than the sample size n. When N is much smaller than n, the consistency rate can be slower. Therefore, the third goal is to investigate the relationship between the asymptotic behavior of the tests and the size of the validation data set. In Section 3, a systematic study is performed to analyze three scenarios: $N/n \to \lambda$ as $\min(n, N) \to \infty$, where $\lambda = 0$, ∞, or $0 < \lambda < \infty$. Furthermore, when validation data are used during the construction procedure to define the nonparametric kernel estimate of $E(Y|W)$ in order to derive the residuals, the resulting test has a bias term going to infinity as $n \to \infty$. Thus we also consider a bias correction.

To efficiently employ the SDR theory of Cook (2009) or Cook and Li (2002), we consider the following alternatives: for all $x \in \mathbb{R}^p$, and for some $p \times q$ orthonormal matrix B with an unknown $q \leq p$ and for an unknown function G,

$$\tilde{H}_1 : \mu(x) = G(B^{\top} x).$$

For the case of no measurement errors in the covariates, Guo, Wang and Zhu (2016) proposed a dimension-reduction model-adaptive approach to circumvent the dimensionality problem. To implement this methodology, we need to estimate the matrix B up to a $q \times q$ orthonormal matrix C. For simplicity, we assume ε is only related to $B^{\top} X$. A number of methods have been proposed in the literature for this purpose. Examples include the sliced inverse regression (SIR) of Li (1991), sliced average variance estimation (SAVE) of Cook and Weisberg (1991), minimum average variance estimation (MAVE) of Xia et al. (2002), contour regression (CR) of Li, Zha and Chiaromonte (2005), directional regression (DR) of Li and Wang (2007), discretization-expectation estimation (DEE) of Zhu et al. (2010), and the average partial mean estimation (APME) of Zhu, Zhu and Feng (2010). However, when measurement errors are present, the above methods need to be modified in order to consistently estimate B up to a $q \times q$ orthonormal matrix C. Here we extend the DEE to address this issue.

In this study, we construct an adaptive-to-model test. The proposed test is based on the test of Zheng (1996). We adapt the method as follows: 1) use the validation data to estimate the conditional expectation $E[g(\beta_0^{\top} X, \gamma_0) | b_0^{\top} W]$, where $b_0 = \beta_0 / \|\beta_0\|$; 2) estimate $\theta_0 = (\beta_0; \gamma_0)$ and b_0 under the null hypothesis; 3) derive an estimator of the target matrix B up to a $q \times q$ orthogonal matrix C. A key ingredient in our method is that the estimator of B can adaptively converge to the direction b_0 under the null hypothesis or to a matrix under the alternative hypothesis. As mentioned above, the test statistic is asymptotically biased. To reduce the bias, we propose a bias-correction method, which we use to construct another test.

The paper is organized as follows. Section 2 describes the construction of the test statistic and provides a brief review of a widely used dimension-reduction method. The asymptotic properties of the test statistic under the null and alternative hypotheses are described in Section 3. Section 4 discussed the findings of our simulation study that compares the proposed test with that of Song (2009). The assumptions are relegated to the Appendix and all proofs are contained in the Supplementary Material. Furthermore, the proposed test can be readily extended to handle a multi-index model, where β is a $p \times q_1$ matrix, without much difficulty. Thus, we focus only on the single-index case.

Before closing this section, we describe some of the notation used in this paper. The sample is denoted by $\{(y_i, w_i), i = 1, \ldots, n\}$ and the validation data are denoted by $\{(\tilde{w}_s, \tilde{x}_s), s = 1, \ldots, N\}$. The two data sets are assumed to be independent of each other. Throughout this paper, $\rightarrow_p$ denotes convergence in probability and $\rightarrow_D$ denotes convergence in distribution. All limits are taken as $n \wedge N \rightarrow \infty$, unless otherwise specified.

2. Methodology Development

2.1. Test construction: A model-adaptive strategy

In this subsection, we describe the construction of the test statistic, which consists of three components.

1) *Model adaptation.* Note that the primary data set does not have observable covariates X, but has W instead. The implementable hypotheses should relate to (Y, W). A natural idea is to use a calibrated regression function $E(Y|W)$. However, this function will not necessarily have the same dimension-reduction structure as that in the original null hypothesis. Note that the regression function $g(\beta_0^{\mathrm{T}} X, \gamma_0)$ depends on X only through a linear combination $\beta_0^{\mathrm{T}} X$. Thus, when W is used, we consider the conditional expectation $E(Y|\beta_0^{\mathrm{T}} W)$. However, we still call it a calibration regression function. Write $r(u, \theta) := E[g(\beta^{\mathrm{T}} X, \gamma)|b^{\mathrm{T}} W = u]$, where $\theta = (\beta; \gamma)$ and $b = \beta/\|\beta\|$. Then, the hypotheses change as follow: for θ_0 in the original null hypothesis, with $b_0 = \beta_0/\|\beta_0\|$, and the matrix B in the original alternative hypothesis $\tilde{H}_1$,

$$\mathcal{H}_0 : P\{E[Y|b_0^{\mathrm{T}} W] = r(b_0^{\mathrm{T}} W, \theta_0)\} = 1,$$

versus

$$\mathcal{H}_1 : P\{E[Y|B^{\mathrm{T}} W] = r(b^{\mathrm{T}} W, \theta)\} < 1, \quad \text{for all } \theta \in \mathbb{R}^{p+d} .$$

In general, the null hypothesis $\mathcal{H}_0$ is not equivalent to the original null hypothesis H_0. However, as in Song (2008), when the family of densities $f_{b_0^{\mathrm{T}} U}(b_0^{\mathrm{T}} w - \cdot)$ is a complete family over the parameter $b_0^{\mathrm{T}} w \in \mathbb{R}$, the equivalence can hold. Furthermore, for the dimension reduction, we need to determine the dimension-reduction structure of the original hypothetical and alternative models. Li and Yin (2007) investigated the equivalence of these two dimension-reduction structures with respect to X and $\Sigma_{XW} \Sigma_W^{-1} W$, respectively. Thus, b_0 and B under the null and alternative hypotheses can be identified by the SDR theory, leading to the model adaptation property of the test constructed below.

2). *Test statistic construction.* Let $e = Y - r(b_0^{\mathrm{T}} W, \theta_0)$. In the spirit of the conditional moment-based test (see Zheng (1996)), we have

$$E[eE[e|b_0^{\mathrm{T}}W]f_{b_0^{\mathrm{T}}W}(b_0^{\mathrm{T}}W)] = E[E^2(e|b_0^{\mathrm{T}}W)f_{b_0^{\mathrm{T}}W}(b_0^{\mathrm{T}}W)] = 0,$$

and under $\mathcal{H}_1$, $E[E^2(e|B^{\mathrm{T}}W)f_{B^{\mathrm{T}}W}(B^{\mathrm{T}}W)] > 0$. To obtain residuals for the construction of the test statistic, we assume the availability of validation data $(\tilde{w}_s, \tilde{x}_s), s = 1, \ldots, N$, which are used to estimate the function r. Let $M(\cdot)$ be a kernel function, let $v \equiv v_N$ be a bandwidth sequence, and set $M_v(\cdot) = v^{-1}M(\cdot/v)$. Then, an estimator of $r(b_0^{\mathrm{T}}W, \theta_0)$ is

$$\hat{r}(\hat{b}_0^{\mathrm{T}}w, \hat{\theta}_0) = \frac{\sum_{s=1}^{N} M_v(\hat{b}_0^{\mathrm{T}}w - \hat{b}_0^{\mathrm{T}}\tilde{w}_s)g(\hat{\beta}_0^{\mathrm{T}}\tilde{x}_s, \hat{\gamma}_0)}{\sum_{s=1}^{N} M_v(\hat{b}_0^{\mathrm{T}}w - \hat{b}_0^{\mathrm{T}}\tilde{w}_s)},$$

where $\hat{b}_0$ is $\hat{\beta}_0/\|\hat{\beta}_0\|$ and $\hat{\theta}_0 = (\hat{\beta}_0; \hat{\gamma}_0)$ is a consistent estimator of $\theta_0 = (\beta_0; \gamma_0)$, the construction of which is presented in Section 2.2. Define the residuals

$$e_i = y_i - r(b_0^{\mathrm{T}}w_i, \theta_0), \;\; \hat{e}_i = y_i - \hat{r}(\hat{b}_0^{\mathrm{T}}w_i, \hat{\theta}_0), \;\; i = 1, \ldots, n.$$

To estimate the conditional expectation of e, given $B^{\mathrm{T}}W$ or $b_0^{\mathrm{T}}W$, we need an estimator $\hat{B}(\hat{q})$ of B that is consistent for b_0 under the null and for B under the alternative. This model adaptation property of $\hat{B}(\hat{q})$ enables the test statistic to adapt to the model, thus alleviating the curse of dimensionality. This estimator is specified later.

To proceed further, let K be another kernel function and let $h \equiv h_n$ be another bandwidth. The analog of the Zheng (1996) test statistic in the current setup is based on an estimator of $E[eE[e|B^{\mathrm{T}}W]f_{B^{\mathrm{T}}W}(B^{\mathrm{T}}W)]$ (also $E[eE[e|b_0^{\mathrm{T}}W]$ $f_{b_0^{\mathrm{T}}W}(b_0^{\mathrm{T}}W)]$, automatically), given by

$$\tilde{V}_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{e}_i K_h(\hat{B}(\hat{q})^{\mathrm{T}}w_i - \hat{B}(\hat{q})^{\mathrm{T}}w_j)\hat{e}_j. \tag{2.1}$$

3). *Bias correction.* The technical details in Supplementary Material show that $\tilde{V}_n$ has a non-negligible asymptotic bias diverging to infinity. The main reason for this is the dependence between the residuals $\hat{e}_i$ and $\hat{e}_j$ for $i \neq j$ when all validation data are used to estimate the function r. There are two ways to correct for this bias. One is to center the test statistic using a suitable estimator of the bias. However, we propose a block-wise estimation approach to asymptotically eliminate the bias. Assume N is a positive even integer. We halve the overall validation data set and use the two halves to construct two estimators of the regression function r. This results in two sets of residuals, as follows. Let

$$\hat{r}_{(1)}(\hat{b}_0^{\mathrm{T}}w, \hat{\theta}_0) = \frac{\sum_{s=1}^{N/2} M_v(\hat{b}_0^{\mathrm{T}}w - \hat{b}_0^{\mathrm{T}}\tilde{w}_s)g(\hat{\beta}_0^{\mathrm{T}}\tilde{x}_s, \hat{\gamma}_0)}{\sum_{s=1}^{N/2} M_v(\hat{b}_0^{\mathrm{T}}w - \hat{b}_0^{\mathrm{T}}\tilde{w}_s)}, \tag{2.2}$$

$$\hat{r}_{(2)}(\hat{b}_0^{\mathrm{T}} w, \hat{\theta}_0) \;=\; \frac{\sum_{s=N/2+1}^{N} M_v(\hat{b}_0^{\mathrm{T}} w - \hat{b}_0^{\mathrm{T}} \tilde{w}_s) g(\hat{\beta}_0^{\mathrm{T}} \tilde{x}_s, \hat{\gamma}_0)}{\sum_{s=N/2+1}^{N} M_v(\hat{b}_0^{\mathrm{T}} w - \hat{b}_0^{\mathrm{T}} \tilde{w}_s)},$$

$$\hat{e}_{i(1)} := y_i - \hat{r}_{(1)}(\hat{b}_0^{\mathrm{T}} w_i, \hat{\theta}_0), \;\; \hat{e}_{i(2)} = y_i - \hat{r}_{(2)}(\hat{b}_0^{\mathrm{T}} w_i, \hat{\theta}_0), \;\; i = 1, \ldots, n.$$

Use these residuals to define the test statistic

$$V_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{e}_{i(1)} K_h(\hat{B}(\hat{q})^{\mathrm{T}} w_i - \hat{B}(\hat{q})^{\mathrm{T}} w_j) \hat{e}_{j(2)} \tag{2.3}$$

to perform the test. We prove that the asymptotic bias of V_n vanishes, but its asymptotic variance becomes larger than that of $\tilde{V}_n$. Note that $\tilde{V}_n$ and V_n are nonstandardized; the standardizing constants are specified in Section 5.

2.2. Parameter estimation

To adaptively estimate B and b_0, the key is to derive an estimator of B up to a $q \times q$ orthonormal matrix C, without depending on the assumed models under the null and alternative hypotheses. Using measurement errors, Carroll and Li (1992) extended the SIR method (Li (1991)) to an EiV regression model. Lue (2004) extended the principal Hessian directions (pHd, Li (1992)) method to the surrogate problem. Li and Yin (2007) established a general invariance law between the surrogate and the original dimension-reduction spaces when X and U are jointly multivariate normal. If X or U is not normally distributed, they suggested an approximation based on the results of Hall and Li (1993). See also Zhang, Zhu and Zhu (2014).

Guo, Wang and Zhu (2016) found that the DEE of Zhu et al. (2010) works well in cases without measurement errors. Unlike SIR or SAVE, the implementation of DEE does not require the selection of the number of slices. Hence, we adopt a DEE method for EiV models when SIR is used. Write $S_{Y|X}$ as the central subspace that is the intersection of all column spaces spanned by the columns of B that make Y conditionally independent of X, given $B^{\mathrm{T}} X$; that is, $Y \perp\!\!\!\perp X | B^{\mathrm{T}} X$. This means that identifying $S_{Y|X}$ is equivalent to identifying a base matrix $\tilde{B}$ that is equal to BC^{T} for a $q \times q$ orthogonal matrix C. Note that the function G is unknown in the alternative. We can rewrite $G(B^{\mathrm{T}} X)$ as $\tilde{G}(\tilde{B}^{\mathrm{T}} X)$. In other words, identifying $\tilde{B}$ is sufficient to identify the model. To prevent notational confusion, we write $\tilde{B} = B$ in the reminder of this paper.

To extend DEE to a setting with measurement errors, we first give a very brief review. Assume that $\mathrm{Cov}(X)$ is the identity matrix. SIR-based DEE uses the matrix $\Lambda = E\{\mathrm{Cov}(E(X|\tilde{Y}(T)))\}$ as the target matrix, where $\tilde{Y}(t) = I(Y \leq$

t), $t \in \mathbb{R}$ and T is an independent copy of Y. The eigenvectors associated with the nonzero eigenvalues of this matrix form the base matrix B. We use surrogate predictors $\mathrm{Cov}(X, W)\Sigma_W^{-1} W$ to replace X, where $\mathrm{Cov}(X, W)$ and Σ_W^{-1} are estimated from the validation data. Carroll and Li (1992) pointed out that SIR with surrogate predictors can produce consistent estimators of $S_{Y|X}$. In other words, all steps of the estimation are the same as those in the setup without measurement errors.

After constructing an estimate $\Lambda_{n,n}$ of Λ, we can obtain an estimate $\hat{B}(\hat{q})$ of B, which consists of the $\hat{q}$ eigenvectors of $\Lambda_{n,n}$ with nonzero eigenvalues, where $\hat{q}$ is defined as follows, using the BIC-type criterion proposed by Zhu, Miao and Peng (2006). Let $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p$ be the eigenvalues of the matrix $\Lambda_{n,n}$, in descending order. An estimate $\hat{q}$ of q is given by

$$\hat{q} = \arg \max_{l=1,\ldots,p} \left\{ \frac{n}{2} \times \frac{\sum_{i=1}^{l} \{\log(\hat{\lambda}_i + 1) - \hat{\lambda}_i\}}{\sum_{i=1}^{p} \{\log(\hat{\lambda}_i + 1) - \hat{\lambda}_i\}} - 2 \times D_n \times \frac{l(l+1)}{2p} \right\},$$

where D_n is a sequence of constants that do not depend on the data. Here, we take $D_n = n^{1/2}$. The following consistency results are obtained from Zhu et al. (2010).

Proposition 1. *Suppose the assumptions in* Zhu et al. (2010) *hold and* $N/n \to \lambda$. *Then, the following hold:*

(1). Under H_0, $P(\hat{q} = 1) \to 1$. *Moreover,*

$$\begin{aligned} \hat{B}(\hat{q}) - B_0 &= O_p(n^{-1/2}), \quad 0 < \lambda \leq \infty, \\ &= O_p(N^{-1/2}), \qquad \lambda = 0, \end{aligned} \tag{2.4}$$

where $B_0 = \pm b_0$ *is the dimension-reduction direction under* H_0.

(2). Under H_1, $P(\hat{q} = q) \to 1$, B *is a* $p \times q$ *orthonormal matrix and* $\hat{B}(\hat{q}) - B$ *satisfies* (2.4).

This proposition states the consistency and adaptation properties of the estimate $\hat{B}(\hat{q})$. Theoretically, the estimate $\hat{B}(\hat{q})$ makes the proposed test adaptive to the underlying model and helps to establish the omnibus property. We can see from the theoretical results that this estimation does not affect the properties of the test statistic asymptotically. However, for the finite sample performance, if $\hat{B}(\hat{q})$ is not sufficiently close to the true value, the empirical power is affected. This is because the underestimation of q by $\hat{q}$ results in $\hat{B}(\hat{q})$ having fewer columns, which may cause the test to not be omnibus, in practice. Obtaining an accurate estimate of q in finite samples is an important issue, but is beyond the scope of this study.

There are various estimators of θ_0 for EiV models available in the literature. Here, we focus on the estimator proposed by Lee and Sepanski (1995) for EiV regression models. Then, we have the following proposition.

Proposition 2. *Suppose the assumptions for Proposition* 2.2 *in* Lee and Sepanski (1995) *hold.*
(1). *Suppose H_0 holds and $N/n \to \lambda$. Then, for $0 < \lambda \le \infty$, $\sqrt{n}(\hat{\theta}_0 - \theta_0) = O_p(1)$, whereas for $\lambda = 0$, $\sqrt{N}(\hat{\theta}_0 - \theta_0) = O_p(1)$.*
(2). *Suppose the following sequence of local alternatives holds, where $C_n \to 0$:*

$$H_{1n} : \mu(x) = g(\beta_0^{\mathrm{T}} x, \gamma_0) + C_n G(B^{\mathrm{T}} x).$$

Then

$$\hat{\theta}_0 - \theta_0 = C_n H(\theta_0) + O_p(n^{-1/2}) + O_p(N^{-1/2}) + o_p(C_n),$$

where

$$H(\theta_0) = \left\{ E\left[\frac{\partial g(\beta_0^{\mathrm{T}} X, \gamma_0)}{\partial \theta} \bar{W}^{\mathrm{T}} \right] E^{-1}[\bar{W}\bar{W}^{\mathrm{T}}] E\left[\frac{\partial g(\beta_0^{\mathrm{T}} X, \gamma_0)}{\partial \theta^{\mathrm{T}}} \bar{W} \right] \right\}^{-1} \times \left\{ E\left[\frac{\partial g(\beta_0^{\mathrm{T}} X, \gamma_0)}{\partial \theta} \bar{W}^{\mathrm{T}} \right] E^{-1}[\bar{W}\bar{W}^{\mathrm{T}}] \right\} E[\bar{W} G(B^{\mathrm{T}} X)],$$

and $\bar{W}$ is a vector of polynomials of W. See the Supplementary Material for further details on $\bar{W}$.

3. Asymptotic Distributions

3.1. Limiting null distribution

In this section, we establish the asymptotic null distribution of the proposed test statistic $\tilde{V}_n$ in (2.1) and V_n in (2.3). Recall that $B_0 = \pm b_0$. Therefore, define

$$Z = B_0^{\mathrm{T}} W, \qquad \eta = g(\beta_0^{\mathrm{T}} X, \gamma_0) - r(b_0^{\mathrm{T}} W, \theta_0),$$
$$\sigma^2 = var(\varepsilon), \quad \xi^2(Z) = E[\eta^2 | Z]. \tag{3.1}$$

where $\theta_0 = (\beta_0; \gamma_0)$ is the true parameter under the null hypothesis and $b_0 = \beta_0 / \|\beta_0\|$. Write Z as $\tilde{Z}$, where W is replaced by the validation data $\tilde{W}$. To proceed further, we need to define

$$z_i = B_0^{\mathrm{T}} w_i, \quad g_i = g(\beta_0^{\mathrm{T}} x_i, \gamma_0), \quad r_i = r(b_0^{\mathrm{T}} w_i, \theta_0), \quad \eta_i = g_i - r_i. \tag{3.2}$$

Write $\tilde{z}_s$, $\tilde{g}_s$, $\tilde{r}_s$, and $\tilde{\eta}_s$ for the entities in (3.2), where w_i is replaced by the validation data $\tilde{w}_s$. When θ_0 and B_0 are replaced by their estimators $\hat{\theta}_0$ and $\hat{B}(\hat{q})$, respectively, in the above definitions, we write $\hat{z}_i$, $\hat{g}_i$, $\hat{r}_i$, and $\hat{\eta}_i$ for z_i, g_i, r_i, and η_i, respectively. Similarly, we write $\hat{\tilde{z}}_s$, $\hat{\tilde{g}}_s$, $\hat{\tilde{r}}_s$, and $\hat{\tilde{\eta}}_s$ for $\tilde{z}_s$, $\tilde{g}_s$, $\tilde{r}_s$, and

$\tilde{\eta}_s$, respectively. We also need to define

$$\mu = \frac{K(0)E[\xi^2(Z)]}{Nh},$$
$$\tau_1 = 2\int K^2(u)du \int (\sigma^2 + \xi^2(z))^2 f_Z^2(z)dz,$$
$$\tau_2 = \int K^2(u)du \int (\sigma^2 + \xi^2(z))\xi^2(z) f_Z^2(z)dz,$$
$$\tau_3 = 2\int K^2(u)du \int (\xi^2(z))^2 f_Z^2(z)dz, \tag{3.3}$$

where σ^2 and $\xi^2(\cdot)$ are defined in (3.1) and f_Z is the density of Z. Consistent estimates of τ_i, for $i = 1, 2, 3$, under H_0 are given by

$$\hat{\tau}_1 = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \frac{1}{h^{\hat{q}}} K^2\left(\frac{\hat{z}_i - \hat{z}_j}{h}\right) \hat{e}_i^2 \hat{e}_j^2,$$
$$\hat{\tau}_2 = \frac{1}{nN} \sum_{i=1}^{n} \sum_{s=1}^{N} \frac{1}{h^{\hat{q}}} K^2\left(\frac{\hat{z}_i - \hat{\tilde{z}}_s}{h}\right) \hat{e}_i^2 \hat{\tilde{\eta}}_s^2,$$
$$\hat{\tau}_3 = \frac{2}{N(N-1)} \sum_{s=1}^{N} \sum_{t\neq s}^{N} \frac{1}{h^{\hat{q}}} K^2\left(\frac{\hat{\tilde{z}}_s - \hat{\tilde{z}}_t}{h}\right) \hat{\tilde{\eta}}_s^2 \hat{\tilde{\eta}}_t^2. \tag{3.4}$$

We are now ready to state the first theorem.

Theorem 1. *Suppose H_0 and the conditions (f), (g), (r), (W), (e), (K), (M), (h1), and (h3) hold, and that $N/n \to \lambda$, for $0 < \lambda \leq \infty$. Then, $nh^{1/2}(\tilde{V}_n - \mu) \to_D N(0, \tilde{\tau})$, where*

$$\tilde{\tau} = \tau_1 + \frac{2}{\lambda}\tau_2 + \frac{1}{\lambda^2}\tau_3, \quad 0 < \lambda < \infty,$$
$$= \tau_1, \quad \lambda = \infty.$$

Here, consistent estimators of μ and τ under H_0 are given by

$$\hat{\mu} = \frac{1}{N^2 h} K(0) \sum_{s=1}^{N} \hat{\tilde{\eta}}_s^2,$$
$$\hat{\tilde{\tau}} = \hat{\tau}_1 + \frac{2}{\lambda}\hat{\tau}_2 + \frac{1}{\lambda^2}\hat{\tau}_3, \quad 0 < \lambda < \infty,$$

respectively, where $\hat{\tau}_i$ is defined as in (3.4). Let $\tilde{T}_n := nh^{1/2}\hat{\tilde{\tau}}^{-1/2}(\tilde{V}_n - \hat{\mu})$. Then, $\tilde{T}_n$ is asymptotically standard normal. The test $\tilde{T}_n$ rejects H_0 whenever $\tilde{T}_n > z_\alpha$, where z_α is the upper $100(1-\alpha)\%$ quantile of the standard normal distribution.

The next result gives the asymptotic null distribution of V_n in (2.3). The result shows that V_n does not have an asymptotic bias.

Theorem 2. *Under the conditions of Theorem* 1, $nh^{1/2}V_n \to_D N(0,\tau)$, *where*

$$\begin{aligned}\tau &= \tau_1 + \frac{4}{\lambda}\tau_2 + \frac{2}{\lambda^2}\tau_3, \quad 0<\lambda<\infty,\\ &= \tau_1, \qquad\qquad\qquad\qquad \lambda=\infty,\end{aligned}$$

where τ_i, *for* $i = 1,2,3$, *is defined as in* (3.3).

To standardize V_n, we use the following consistent estimate of τ for the case $0<\lambda<\infty$:

$$\begin{aligned}\hat{\tau} = &\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{h^{\hat{q}}}K^2\left(\frac{\hat{z}_i-\hat{z}_j}{h}\right)\hat{e}_{i(1)}^2\hat{e}_{j(2)}^2\\ &+\frac{4}{\lambda nN}\sum_{i=1}^{n}\sum_{s=N/2+1}^{N}\frac{1}{h^{\hat{q}}}K^2\left(\frac{\hat{z}_i-\hat{\hat{z}}_s}{h}\right)\hat{e}_{i(1)}^2\hat{\hat{\eta}}_s^2\\ &+\frac{4}{\lambda nN}\sum_{i=1}^{n}\sum_{t=1}^{N/2}\frac{1}{h^{\hat{q}}}K^2\left(\frac{\hat{z}_i-\hat{\hat{z}}_t}{h}\right)\hat{e}_{i(2)}^2\hat{\hat{\eta}}_t^2\\ &+\frac{16}{\lambda^2N^2}\sum_{t=1}^{N/2}\sum_{s=N/2+1}^{N}\frac{1}{h^{\hat{q}}}K^2\left(\frac{\hat{\hat{z}}_s-\hat{\hat{z}}_t}{h}\right)\hat{\hat{\eta}}_s^2\hat{\hat{\eta}}_t^2,\end{aligned}$$

where s and t are the indices of the two sets of validation data, and $\hat{\eta}_t$ or $\hat{\eta}_s$ is estimated by the other half of validation data. That is, $\hat{\hat{\eta}}_t = g(\hat{\beta}_0^{\mathrm{T}}\tilde{x}_t,\hat{\gamma}_0) - \hat{r}_{(2)}(\hat{b}_0^{\mathrm{T}}\tilde{w}_t,\hat{\theta}_0)$, $t=1,\ldots,N/2$ and $\hat{\hat{\eta}}_s = g(\hat{\beta}_0^{\mathrm{T}}\tilde{x}_s,\hat{\gamma}_0) - \hat{r}_{(1)}(\hat{b}_0^{\mathrm{T}}\tilde{w}_s,\hat{\theta}_0)$, $s=N/2+1,\ldots,N$, where $\hat{r}_{(1)}$ and $\hat{r}_{(2)}$ are defined as in (2.2). The standardized test statistic is

$$\begin{aligned}T_n &= \hat{\tau}^{-1/2}nh^{1/2}V_n, \quad 0<\lambda<\infty,\\ &= \hat{\tau}_1^{-1/2}nh^{1/2}V_n, \qquad \lambda=\infty,\end{aligned}$$

where $\hat{\tau}_1$ is defined as in (3.4). According to Slutsky's theorem, T_n is asymptotically standard normal. For the significance level α, the null hypothesis is rejected when $T_n > z_\alpha$.

Remark 1. A significant feature of this test is that we need only use the standardizing sequence $nh^{1/2}$, which is the same as that used in the classical local smoothing tests when X is one-dimensional. This shows that the test has a much faster convergence rate to its limit than those of some of the classical tests that have a rate of order $nh^{p/2}$. This helps to maintain the significance level in finite samples when its asymptotic null distribution is used to determine the critical values.

When $N/n \to \lambda = 0$, the standardizing constant will be different because of the estimate $\hat{r}$ of the function $r(\cdot)$.

Theorem 3. *Suppose H_0 and the conditions (f), (g), (r), (W), (e), (K), (M), (h2), and (h5) hold and $N/n \to 0$. Then, $Nv^{1/2}\{\tilde{V}_n - \tilde{\mu}\} \to_D N(0, \tilde{\tau})$, $Nv^{1/2}V_n \to_D N(0, \tau)$, where $\tilde{\mu} = (vN)^{-1} \int M^2(u)du\, E[\xi^2(Z)]$, $\tau := 2\tilde{\tau}$, and*

$$\tilde{\tau} = 2 \int \left(\int M(u)M(u+u')du \right)^2 du' \int (\xi^2(z))^2 f_Z^2(z)dz.$$

3.2. Asymptotic power

In this section, we assume $0 < \lambda \leq \infty$ and investigate the asymptotic properties of V_n under various alternatives. Consider a sequence of alternatives

$$H_{1n} : \mu(x) = g(\beta_0^{\mathrm{T}} x, \gamma_0) + C_n G(B^{\mathrm{T}} x), \quad x \in \mathbb{R}^p, \tag{3.5}$$

where $G(\cdot)$ satisfies $E(G^2(B^{\mathrm{T}} X)) < \infty$ and β_0 is a vector in *span*(B). When C_n is a nonzero constant, H_{1n} is a global alternative, and when $C_n = n^{-1/2}h^{-1/4}$ tends to zero, H_{1n} specifies the local alternatives of interest. Note that the asymptotic properties of $\hat{B}(\hat{q})$ and $\hat{\theta}_0 = (\hat{\beta}_0; \hat{\gamma}_0)$ affect the behavior of the test statistic V_n. The asymptotic results of $\hat{\theta}_0$ are illustrated in Proposition 2. Thus, we discuss the result for the consistency of $\hat{q}$ and $\hat{B}(\hat{q})$ here.

Theorem 4. *Suppose the conditions in* Zhu et al. (2010) *hold. Under H_{1n} of* (3.5) *with $C_n = n^{-1/2}h^{-1/4} \to 0$, $P(\hat{q} = 1) \to 1$ and $\hat{B}(\hat{q}) - B_0 = O_p(C_n)$, where $B_0 = \pm\beta_0/\|\beta_0\|$.*

This theorem indicates that $\hat{q}$ and $\hat{B}(\hat{q})$ are no longer consistent for q and B in (3.5).

Theorem 5. *Under the alternatives given in* (3.5)*, the following results hold:*
(i) Suppose (f), (g), (r), (G), (W), (e), (K), (M), (h1), and (h6) hold. Under the global alternative with fixed $C_n \equiv C$, $V_n/\hat{\tau}$ tends to a positive constant.
(ii) Suppose (f), (g), (r), (G), (W), (e), (K), (M), (h1), and (h4) hold. Under the local alternatives H_{1n} with $C_n = n^{-1/2}h^{-1/4}$, $nh^{1/2}V_n \to_D N(\Delta, \tau)$, where τ is given in Theorem 2 *and*

$$\Delta = E\left[\{E[G(B^{\mathrm{T}}X)|Z] - E\left[\frac{\partial g(\beta_0^{\mathrm{T}}\tilde{x}_s, \gamma_0)}{\partial \theta^{\mathrm{T}}} \middle| Z \right] H(\theta_0)\}^2 f_Z(Z) \right].$$

Remark 2. The result (i) implies the consistency of T_n against the global alternatives and (ii) shows that the test can detect local alternatives distinct from the null at a rate of order $n^{-1/2}h^{-1/4}$. In contrast, the classical methods can only detect the alternatives that converge to the null at a rate of order $n^{-1/2}h^{-p/4}$.

4. Numerical Studies

This section presents four simulation studies that examine the performance of the proposed test T_n. For comparison purposes, we consider the tests of Zheng (1996) (T_n^{Zh}) and Song (2009) (T_n^S). We adapt Zheng's test to the EiV settings, which are the same as those in our test, except that $B^{\top}W$ is replaced by the original W. This is a typical local smoothing test. The test proposed by Song (2009) is a score-type test and is designed for EiV models with validation data.

In the simulation **Study 1**, B is equal to b_0, and thus, the models are parametric single-index. In addition, we run simulation studies for the test $\tilde{T}_n$ of Theorem 1 when $0 < \lambda < \infty$ and report the results in the Supplementary Material. The purpose of **Study 2** is to confirm that the proposed test T_n is not a directional test. Here, we assume $B = (\beta_0, \beta_1)$, with $\beta_0 \perp \beta_1$, under the alternative hypothesis. **Study 3** examines the finite-sample performance when $N < n$ and $N > n$. **Study 4** considers four nonlinear models. All simulations are based on 2,000 replications.

Throughout the simulation studies, X is taken to be multi-normal with mean zero and covariance matrices $\Sigma_1 = I_{p\times p}$ and $\Sigma_2 = (0.3^{|i-j|})_{p\times p}$. The regression model error ε follows a standard normal distribution, whereas the measurement error $U \sim N(0, 0.5)$. The kernel function is $K(u) = (15/16)(1-u^2)^2 I(|u| \le 1)$, which is a second-order symmetric kernel, and $M(u) = K(u)$.

Bandwidth selection. Because the tests involve bandwidth selection in the kernel estimation, we run a simulation to empirically select the bandwidths for the three tests in the comparison. Furthermore, because maintaining the significance level is important, we select bandwidths such that the tests have empirical sizes close to the significance level and retain the use when the dimensionality changes. To this end, we use a simple model, $\mu(x) = \beta_0^{\top} x$, $\beta_0 = (1, 1, \ldots, 1)^{\top}/\sqrt{p}$, for $p = 2$, to select these bandwidths and to check whether they can be used if $p = 8$. In our test, there are two bandwidths. We adopt $h = c_1 n^{-1/(4+\hat{q})}$, which is the optimal rate for the kernel estimation, and determine the constant c_1. Similarly, for the kernel estimator of the function $r(b_0^{\top}W, \theta_0)$, we choose the bandwidth $v = c_2(N/2)^{-2/5}$, because we devide the validation data set of size N in two. For $\tilde{T}_n$, v is $c_2 N^{-2/5}$. In the selection process, we try different bandwidths to investigate their impact on the empirical size. In addition, to reduce the computational burden, we set $c_1 = c_2 = c$.

We compute the empirical size at every equal grid point $c = (i-1)/10$ for $i = 1, \ldots 21$. In Figure 1, we report the empirical sizes associated with different bandwidths when the regression model is $\mu(x) = \beta_0^{\mathrm{T}} x$ and $n = 100, 200$, $\lambda = 4$, and the covariance matrix of X is Σ_1. We can see that the test is not sensitive to the bandwidth and a value of $c = 1.6$ may be a good choice for both T_n and $\tilde{T}_n$. We also need to select two bandwidths for the Zheng's test. Here, We consider $h = cn^{-1/(4+p)}$ and $v = c(N/2)^{-2/(4+p)}$. We find that in order to maintain the significance level, the bandwidths must have a larger c. The results suggest that a good bandwidth within the equal grid points is $c = 2.5 + (i-1)/10$ for $i = 1, \ldots, 21$ (see Figure 1). For Song's score test, only one bandwidth is required. Here, we set the bandwidth to $v = cN^{-1/(4+p)}$ and search for c within the equal grid points using $c = 1 + (i-1)/10$ for $i = 1, \ldots, 21$. Here too, a large value of c is desirable. The reported curves are shown in Figure 1.

We find that the empirical sizes of T_n are not sensitive to the bandwidths selected. The curves for $p = 2$ and $p = 8$ are almost identical. Although the empirical size of $\tilde{T}_n$ is slightly effected by the dimensionality, it is still more robust than T_n^{Zh} and T_n^S. A value of $c = 1.6$ is recommended for both T_n and $\tilde{T}_n$. However, the empirical sizes of T_n^{Zh} and T_n^S associated with the bandwidths are not as robust as that of T_n. The empirical sizes show the efficient bandwidth changes as p increases. When p is small, a small h can maintain its theoretical size. As p increases, a larger h is necessary. This phenomenon is particularly serious for T_n^{Zh}. For the bandwidths of T_n^{Zh}, $c = 3.9$ is appropriate. Finally, $c = 2.2$ seems appropriate for T_n^S.

Study 1. The data are generated from the following model:

$$H_{11}: \ \mu(x) = \beta_0^{\mathrm{T}} x + a\,(\beta_0^{\mathrm{T}} x)^2,$$

$$H_{12}: \ \mu(x) = \beta_0^{\mathrm{T}} x + a\,\exp\left(\frac{-(\beta_0^{\mathrm{T}} x)^2}{2}\right),$$

$$H_{13}: \ \mu(x) = \beta_0^{\mathrm{T}} x + 2a\cos(0.6\pi\beta_0^{\mathrm{T}} x).$$

The case of $a = 0$ corresponds to the null hypothesis and $a \neq 0$ corresponds to the alternatives. In H_{11} and H_{12}, the alternative parts $(\beta_0^{\mathrm{T}} x)^2$ and $\exp(-(\beta_0^{\mathrm{T}} x)^2/2$, respectively, always exist for $a \neq 0$. However, the alternative part of H_{13}, $\cos(0.6\pi\beta_0^{\mathrm{T}} x)$, appears and disappears periodically for any nonzero a, which makes the bandwidth selection process even more challenging. This is because the large bandwidth selected to maintain the significance level may make the test obtuse to high-frequency alternatives. The dimension p is equal to 2 and 8, enabling us to check the impact from the dimensionality. Let $\beta_0 = (1, 1, \ldots, 1)^{\mathrm{T}}/\sqrt{p}$ and $\lambda = 4$. The simulation results are presented in Table 1.

From this table, we find that when $p = 2$, T_n^S performs very well. This is

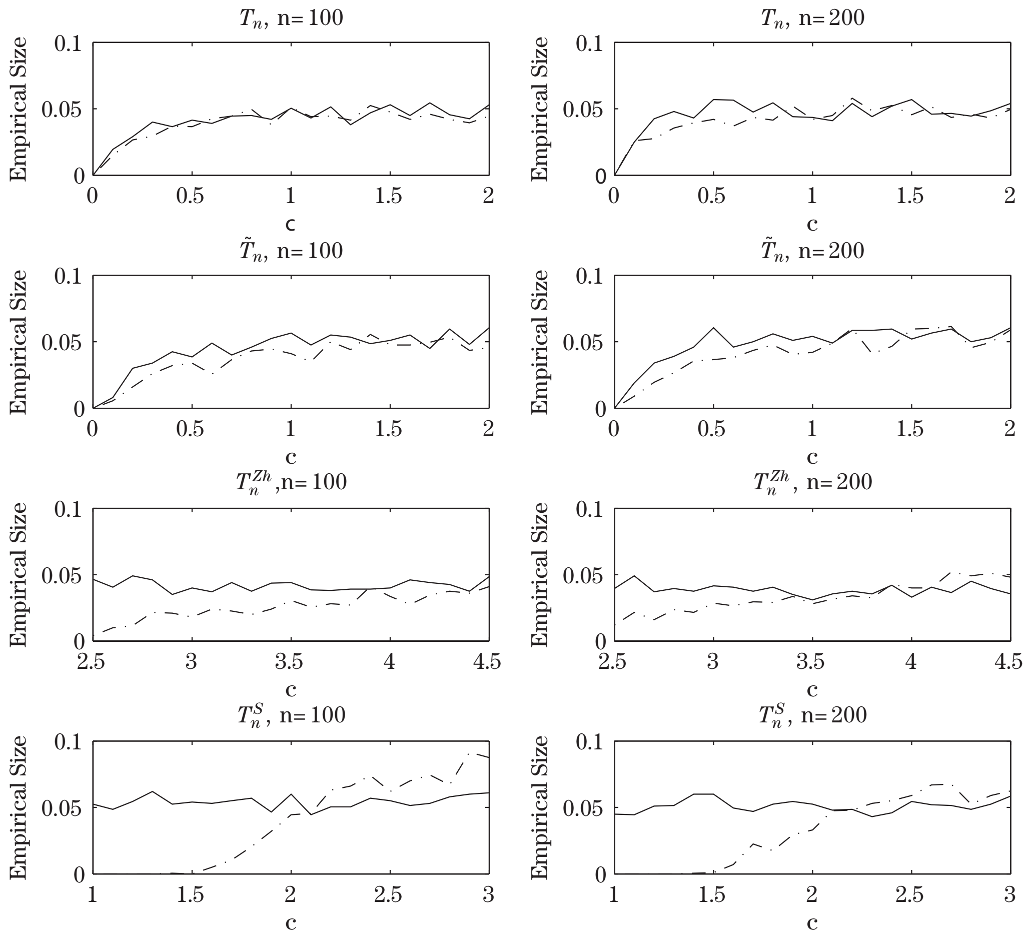

Figure 1. Plots of the empirical size curve against different values of c in the bandwidths. For model $Y = \beta_0^{\top} X + \varepsilon$, the solid lines show $p = 2$, and the dash-dotted lines show $p = 8$.

expected because the consistency rate of this test is $1/\sqrt{n}$. In addition, when p is small, T_n^{Zh} is comparable to T_n because both are local smoothing tests. However, when the dimension increases, T_n^{Zh} and T_n^S are severely impacted by the dimensionality. The test T_n^{Zh} behaves much worse. In particular, when $p = 8$, the test breaks down for $n = 100$ and regains its power as n increases. The test T_n^S is also affected by the dimensionality because the residuals contain nonparametric estimations based on the local smoothing technique. Its power decreases in both small and large samples. On the other hand, the test T_n does not suffer from the curse of dimensionality in the limited simulation studies presented here. When

Table 1. Empirical sizes and powers of T_n, T_n^{Zh}, and T_n^S for H_{1k}, $k = 1, 2, 3$, in **Study 1**.

	$\Sigma = \Sigma_1$	n=100		n=200		$\Sigma = \Sigma_2$	n=100		n=200	
H11	a	0	0.5	0	0.5	a	0	0.5	0	0.5
T_n	$p = 2$	0.0430	0.6960	0.0570	0.9625	$p = 2$	0.0505	0.8635	0.0525	0.9990
	$p = 8$	0.0515	0.6365	0.0340	0.9545	$p = 8$	0.0385	0.9390	0.0535	1.0000
T_n^{Zh}	$p = 2$	0.0360	0.7105	0.0335	0.9690	$p = 2$	0.0400	0.9065	0.0385	0.9985
	$p = 8$	0.0285	0.2875	0.0410	0.5500	$p = 8$	0.0350	0.6190	0.0405	0.9420
T_n^S	$p = 2$	0.0495	0.9550	0.0335	0.9995	$p = 2$	0.0655	0.9895	0.0595	1.0000
	$p = 8$	0.0440	0.7305	0.0410	0.9705	$p = 8$	0.0430	0.9715	0.0420	1.0000
H12	a	0	0.5	0	0.5	a	0	0.5	0	0.5
T_n	$p = 2$	0.0505	0.6435	0.0485	0.9505	$p = 2$	0.0430	0.6010	0.0520	0.8975
	$p = 8$	0.0450	0.6110	0.0465	0.9285	$p = 8$	0.0475	0.4470	0.0445	0.8090
T_n^{Zh}	$p = 2$	0.0330	0.7060	0.0425	0.9445	$p = 2$	0.0390	0.6620	0.0495	0.9115
	$p = 8$	0.0300	0.3060	0.0400	0.5900	$p = 8$	0.0420	0.2500	0.0405	0.4865
T_n^S	$p = 2$	0.0530	0.8375	0.0510	0.9825	$p = 2$	0.0505	0.7670	0.0475	0.9645
	$p = 8$	0.0460	0.6265	0.0365	0.9170	$p = 8$	0.0450	0.4890	0.0365	0.8050
H13	a	0	0.5	0	0.5	a	0	0.5	0	0.5
T_n	$p = 2$	0.0475	0.5395	0.0435	0.8775	$p = 2$	0.0440	0.4690	0.0410	0.7975
	$p = 8$	0.0460	0.4900	0.0470	0.8445	$p = 8$	0.0445	0.2510	0.0530	0.6075
T_n^{Zh}	$p = 2$	0.0350	0.5075	0.0450	0.8410	$p = 2$	0.0365	0.4010	0.0505	0.7330
	$p = 8$	0.0250	0.1610	0.0450	0.3225	$p = 8$	0.0355	0.0780	0.0415	0.1650
T_n^S	$p = 2$	0.0570	0.2540	0.0410	0.4225	$p = 2$	0.0560	0.1350	0.0565	0.1840
	$p = 8$	0.0405	0.1895	0.0420	0.3600	$p = 8$	0.0440	0.0660	0.0400	0.0600

p is large, T_n outperforms T_n^S. The finite-sample power of the T_n^S test is poor against the alternatives H_{13} for both $p = 2$ and $p = 8$.

Study 2. In this study, we design a simulation to confirm that the test T_n is not a directional test, but that Song's test T_n^S is. The data are generated from the following model:

$$H_{14} : \mu(x) = \beta_0^{\mathrm{T}} x + a(\beta_1^{\mathrm{T}} x)^2, \qquad H_{15} : \mu(x) = 2\beta_0^{\mathrm{T}} x + a(2\beta_1^{\mathrm{T}} x)^3.$$

Once again, $a = 0$ corresponds to the null hypothesis and $a \neq 0$ corresponds to the alternatives. The matrix $B = (\beta_0, \beta_1)$ and, thus, the structural dimension q under the alternative is 2. Let $\lambda = 4$, $p = 4$, $\beta_0 = (1, 1, 0, 0)^{\mathrm{T}}/2$, and $\beta_1 = (0, 0, 1, 1)^{\mathrm{T}}/2$. The simulation results are presented in Table 2. From these results, we first observe that T_n^S has good performance under H_{14}, which coincides with the findings in **Study 1**. However, the poor performance under H_{15} shows that T_n^S is a directional test, because this alternative is not detected at all. At the population level, we find that the conditional expectation of the residual is equal to zero under this alternative. In this case, T_n still works well. This supports the claim that T_n is an omnibus test.

Table 2. Empirical sizes and powers of T_n and T_n^S for H_{14} and H_{15} in **Study 2**.

	a	H_{14}				H_{15}			
$\lambda = 4$		$\Sigma = \Sigma_1$		$\Sigma = \Sigma_2$		$\Sigma = \Sigma_1$		$\Sigma = \Sigma_2$	
		n=100	n=200	n=100	n=200	n=100	n=200	n=100	n=200
T_n	0	0.0510	0.0500	0.0475	0.0485	0.0445	0.0445	0.0370	0.0480
	0.1	0.0545	0.0735	0.0650	0.0775	0.0495	0.0435	0.0525	0.0620
	0.3	0.1245	0.2615	0.2050	0.3525	0.1045	0.1510	0.2170	0.3725
	0.5	0.3115	0.5910	0.4510	0.8025	0.1915	0.3245	0.3270	0.6125
T_n^S	0	0.0525	0.0605	0.0605	0.0540	0.0450	0.0515	0.0540	0.0535
	0.1	0.0830	0.0970	0.0915	0.1155	0.0620	0.0545	0.0525	0.0490
	0.3	0.2310	0.4245	0.3575	0.6170	0.0485	0.0465	0.0590	0.0570
	0.5	0.5020	0.8040	0.6935	0.9410	0.0590	0.0515	0.0505	0.0410

Study 3. In this study, we explore the impact of the estimation of r on the performance of the proposed tests. Here, a small $\lambda = \lim(N/n)$ denotes a small amount of validation data and a large λ means the estimator $\hat{r}$ is very close to the true function r. Consider $N/n = 0.1, 0.5, 4, 8$. From Theorem 3, we know that when λ is small, we can have a test with a simpler limiting variance. Write the related test as $T_n^{(1)}$. From Theorem 2, when $\lambda = \infty$, we can also have a test for large N/n, which we write as $T_n^{(2)}$. To examine whether these two variants of the test T_n work, we generate data from the model H_{11} in **Study 1**. When the size of the validation data is such that $N/n = 0.1, 0.5$, $T_n^{(1)}$ is used, and when $N/n = 4, 8$, $T_n^{(2)}$ is applied. Because $T_n^{(1)}$ is a test with very different convergence rates, we also need to choose suitable bandwidths. Here, we search for the bandwidths at the rates $v = c(N/2)^{-1/3}$ and $h = cn^{-1/(2+\hat{q})}$ and find that $c = 2$ is a good choice. For $T_n^{(2)}$, only the asymptotic variance changes. Thus, we use the same bandwidths as before. When $\lambda = 0.1, 0.5$, we use a larger sample size for the validation data, $N = 100, 200$; otherwise, N is too small and the tests do not perform well. The simulation results are presented in Table 3.

From Table 3, we have the following two observations. First, for $\lambda = 0.1$, T_n is more conservative, with lower power than $T_n^{(1)}$. This seems to indicate that T_n is less sensitive to the alternative model than $T_n^{(1)}$. This phenomenon stems from the improper selection of bandwidths for T_n, because Conditions (h1) and (h2) ensure that the consistency of T_n and $T_n^{(1)}$ require different ratios of h and v. Thus, when N/n is very small, $T_n^{(1)}$ seems to be a better choice than T_n. However, when λ is closed to one, $T_n^{(1)}$ cannot maintain the significance level well. Second, $T_n^{(2)}$ has slightly higher empirical size and power than those of T_n. Overall, the performance of $T_n^{(2)}$ is very similar to that of T_n. Therefore, when

Table 3. Empirical sizes and powers of T_n, $T_n^{(1)}$, and $T_n^{(2)}$ for H_{11} in **Study 3**.

H_{11}		p=2		p=8		p=2		p=8	
		$\lambda=0.1$		$\lambda=0.1$		$\lambda=0.5$		$\lambda=0.5$	
	a	N=100	N=200	N=100	N=200	N=100	N=200	N=100	N=200
T_n	0	0.0160	0.0255	0.0080	0.0120	0.0330	0.0420	0.0235	0.0295
	0.1	0.0380	0.0865	0.0280	0.0535	0.0535	0.0725	0.0425	0.0685
	0.3	0.4695	0.8920	0.4465	0.8835	0.2720	0.6005	0.2370	0.5905
	0.5	0.9465	1.0000	0.9360	1.0000	0.7270	0.9860	0.6390	0.9805
$T_n^{(1)}$	0	0.0610	0.0555	0.0400	0.0475	0.1690	0.1720	0.1190	0.1490
	0.1	0.1135	0.1745	0.0885	0.1635	0.2175	0.2470	0.1745	0.2600
	0.3	0.7100	0.9680	0.6550	0.9595	0.5695	0.8410	0.5100	0.8165
	0.5	0.9865	1.0000	0.9860	1.0000	0.9115	0.9975	0.8625	0.9985
		$\lambda=4$		$\lambda=4$		$\lambda=8$		$\lambda=8$	
	a	n=100	n=200	n=100	n=200	n=100	n=200	n=100	n=200
T_n	0	0.0525	0.0545	0.0480	0.0405	0.0485	0.0385	0.0430	0.0545
	0.1	0.0590	0.0960	0.0530	0.0925	0.0705	0.0850	0.0615	0.0780
	0.3	0.2645	0.5715	0.2525	0.5445	0.3045	0.5815	0.2550	0.5605
	0.5	0.6705	0.9700	0.6295	0.9665	0.6885	0.9690	0.6620	0.9690
$T_n^{(2)}$	0	0.0610	0.0620	0.0575	0.0495	0.0530	0.0420	0.0445	0.0575
	0.1	0.0660	0.1075	0.0685	0.1085	0.0755	0.0890	0.0690	0.0840
	0.3	0.2910	0.5985	0.2845	0.5775	0.3145	0.5960	0.2735	0.5760
	0.5	0.6880	0.9735	0.6580	0.9720	0.6950	0.9700	0.6745	0.9715

the size of the validation data N is reasonably large and the ratio N/n is large, $T_n^{(2)}$ is appropriate. Furthermore, the simulations show that although $T_n^{(1)}$ can be used, it does not maintain the finite-sample significance level as well as the T_n test does. Thus, when the ratio N/n is not too small, we recommend using T_n rather than $T_n^{(1)}$ in practical situations.

Study 4. In this study, we consider a nonlinear single-index null model. We try four alternatives with different structural dimensions:

$$H_{16}: \ \mu(x) = (\beta_0^{\top} x)^3 + a|\beta_0^{\top} x|,$$
$$H_{17}: \ \mu(x) = (\beta_0^{\top} x)^3 + a x_3^2,$$
$$H_{18}: \ \mu(x) = (\beta_0^{\top} x)^3 + a\Big(\frac{x_2}{4} + x_3^2 + \cos(\pi x_4)\Big),$$
$$H_{19}: \ \mu(x) = (\beta_0^{\top} x)^3 + a\Big(\frac{x_2}{2} + x_3^2 + \cos(\pi x_4) + x_5 \exp\Big(\frac{x_6}{2}\Big) + x_8 x_7\Big).$$

Let $p = 4$ for H_{16}, H_{17}, and H_{18}, and let $p = 8$ for H_{19}. Then $\beta_0 = [1, 0, \ldots, 0]^{\top}$, $\Sigma = \Sigma_1$, $\sigma_u = 0.5$, and a is designed to be $0, 0.2, 0.4, 0.6, 0.8, 1.0$. In these cases, q is always one for the null, but is different for the alternatives. For H_{16}, $q = 1$ for any nonzero a. The structure dimension under H_{17} is 2, and under H_{18}, $p = q = 4$. For H_{19}, $p = q = 8$. The test T_n uses the same bandwidths as

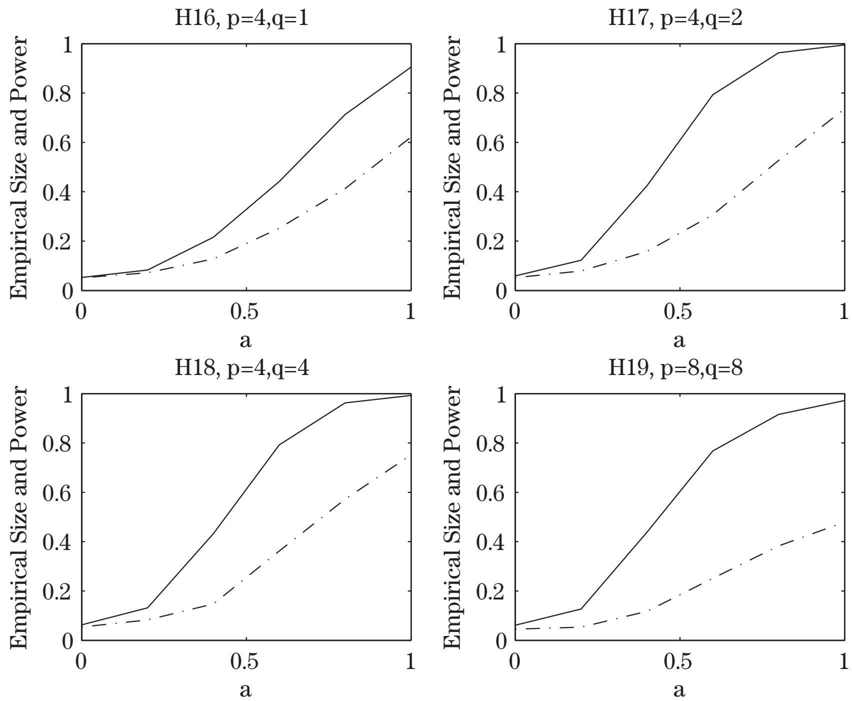

Figure 2. Plots of power curves over a for $H16 - H19$ in **Study 4**. The solid lines show T_n, and the dash-dotted lines show T_n^{Zh}.

those chosen for the above linear model. For T_n^{Zh}, we adjust the bandwidths to maintain the test's performance. Set $c = 2.7$ for H_{16}, H_{17}, and H_{18}, and let $c = 3$ for H_{19}. The results are presented in Figure 2.

We have the following observations. First, T_n has greater empirical power than T_n^{Zh} for all chosen alternatives. Under H_{18} and H_{19}, without the dimension-reduction structure, T_n is still more powerful than T_n^{Zh} because T_n is constructed from $nh^{1/2}V_n/\sqrt{\tau} = h^{(1-q/2)} \times nh^{q/2}V_n/\sqrt{\tau}$. Second, the power of T_n^{Zh} decreases quickly as p and q increase, but that of T_n does not.

5. A Real-Data Example

In this section, we apply the proposed tests to the NHANES-I Epidemiologic Study Cohort data (Jones et al. (1987)). The study interviewed 8,596 women about their nutrition habits and then looked for evidence of breast cancer. Part of the data are analyzed by Carroll and Li (1992) for the purpose of dimension reduction and by Carroll et al. (2006) for nonlinear regression modeling. In the data set, Y is an indicator of breast cancer. The relevant predictors include age, body mass index (bmi), alcohol (alch, yes/no), saturated fat intake (satfat), calorie intake (cals), vitamin A intake (vita), and vitamin C intake (vitc). Age, body mass index, and alcohol are measured without measurement errors. Because it is difficult to measure a person's long-term diet, the variables related to the nutrition intake are 24-hour recall, which are known to be measured with a large error (Beaton, Milner and Little (1979); Wu, Whittemore and Jung (1986)). Clearly, a single day's diet cannot serve as an adequate estimation of the average level of a day's nutrition intake in one year. This measurement error can be modeled using the Continuing Survey of Food Intakes by Individuals (CSFII) data, which includes a 24-hour recall and three follow-up interviews. Validation data are then obtained for this data set. Similarly to Carroll and Li (1992), we consider the part of the data set corresponding to ages between 25 and 50. The primary data set has size $n = 3{,}143$, and the size of the validation data set is $N = 1{,}847$. There are 59 cases in the primary data set of reported breast cancer. This setup suggests using a logistic model to fit the data set. Carroll and Li (1992) consider an estimation using dimension reduction that includes three predictors age, bmi, and satfat. We now use our tests to check the adequacy of the fit for this data set. The values of the test statistics are -0.9403 and -0.7785, and the p-values are 0.8265 and 0.7819, respectively. Note that these p-values are reasonable because the tests are one-sided, not two-sided. Because the other predictors of nutrition intake with measurement errors are modeled as a logistic model, we include all four nutrition intake predictors for checking purpose only. The values of the test statistics T_n and $\tilde{T}_n$ are -0.1533 and -0.1918, respectively. The p-values are 0.5609 and 0.5761, respectively. Again, a logistic model is plausible. Moreover, we use all seven predictors in the model checking for a logistic model, and find values of T_n and $\tilde{T}_n$ of 0.1154 and -0.2035, and p-values of 0.4541 and 0.5806, respectively. Therefore, a logistic model is feasible for fitting this data set.

6. Conclusion

In this study, we investigate an adaptive-to-model test for parametric single-index EiV models. The consistency rate of the proposed test does not depend on the dimension of the covariate vector. The simulation studies show that the proposed test can potentially alleviate the curse of dimensionality. In addition, the studies confirm that the proposed test is omnibus. However, selecting optimal bandwidths remains a problem because the optimal bandwidth in the estimation may not be optimal in the hypothesis testing. This problem requires further research in order to select bandwidths adaptively. A critical issue is consistently estimating the number of columns in the dimension-reduction structure under local alternative models. This is left to further research.

Appendix.Assumptions

(f). The density f of Z has bounded partial derivatives up to order 3 and satisfies

$$0 < \inf_z f_Z(z) \leq \sup_z f_Z(z) < \infty.$$

(g). $g(\beta^{\mathrm{T}}x, \gamma)$ is a measurable function of x for each $\theta = (\beta; \gamma)$ and is differentiable in θ up to order 3, and $E\big\|\partial g(\beta_0^{\mathrm{T}}X, \gamma_0)/\partial\theta\big\|^2 < \infty$.
(r). The function $r(b^{\mathrm{T}}w, \theta)$ has bounded partial derivatives with respect to $b^T w$ up to order 3, and $E[r^2(b^{\mathrm{T}}W, \theta)] < \infty$, for $\theta \in \mathbb{R}^{p+d}$.
(G). Let $\Delta(Z) = E[G(B^{\mathrm{T}}X)|Z]$. Then, $E[\Delta^2(Z)] < \infty$, $E[(G(B^{\mathrm{T}}X) - \Delta(Z))^4] < \infty$, and $\Delta(z)$ has bounded partial derivatives up to order 3.
(W). $\max_{1\leq k\leq p} E[W_{(k)}^2|Z] < \infty$, where $W_{(k)}$ represents the k-th coordinate of W, for $k = 1, \ldots, p$.
(e). $E[(\xi^2(Z))^2] < \infty$ and $\xi^2(z)$ are uniformly continuous functions with bounded partial derivatives up to order 3.
(K). K is a spherically symmetric and continuous kernel function with bounded support and is of order 2, with all derivatives bounded.
(M). M is a symmetric and continuous kernel function with bounded support and is of order 2, with all derivatives bounded.
(h1). $h \to 0$, $v \to 0$, and $v/h \to 0$.
(h2). $h \to 0$, $v \to 0$, and $h/v \to 0$.
(h3). $nh^2 \to \infty$, $Nv^2 \to \infty$, and $nv^4 \to 0$.
(h4). $nh^{5/2} \to \infty$, $Nv^2 \to \infty$, and $nv^4 \to 0$.
(h5). $Nh^2 \to \infty$, $Nv/(nh) \to 0$, and $Nv^4 \to 0$.
(h6). $nh^q \to \infty$ and $Nv \to \infty$.

Supplementary Material

The proofs are postponed to the online supplementary materials.

Acknowledgments

This research was supported by a grant from the University Grants Council of Hong Kong, Hong Kong, and an NSFC grant (NSFC11671042). The authors would like to thank the Associate Editor and two referees for their constructive comments, which helped to improve the paper. The authors are in alphabetical order and the second author plays the role of first authorship as this is a part of his PHD thesis.

References

Beaton, G. H., Milner, J. and Little, J. A. (1979). Sources of variation in 24-hour dietary recall data: implications for nutrition study design and interpretation. *Am. J. Clin. Nutr.* **32**, 2546–2559.

Carroll, R. J. and Li, K. C. (1992). Measurement error regression with unknown link: dimension reduction and data visualization. *J. Amer. Statist. Assoc.* **87**, 1040–1050.

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models.* Chapman and Hall/CRC.

Cheng C. L. and Kukush A. G. (2004). A goodness-of-fit test for a polynomial errors-in-variables model. *Ukrainian Math. J.* **56**, 527–543.

Cheng C. L. and Van Ness, J. W. (1999). *Statistical Regression with Measurement Error.* Arnold, London.

Cook, R. D. (2009). *Regression graphics: Ideas for studying regressions through graphics* (Vol. 482). John Wiley & Sons.

Cook, R. D. and Li, B. (2002). Dimension reduction for conditional mean in regression. *Ann. Statist.* **30**, 455–474.

Cook, R. D. and Weisberg, S. (1991). Sliced inverse regression for dimension reduction: comment. *J. Amer. Statist. Assoc.* **86**, 328–332.

Dai, P., Sun, Z. and Wang, P. (2010). Model checking for general linear error-in-covariables model with validation data. *J. Syst. Sci. Complex.* **23**, 1153–1166.

Fuller, W. A. (2009). *Measurement Error Models.* John Wiley & Sons.

González-Manteiga, W. and Crujeiras, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *TEST.* **22**, 361–411.

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

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

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

Hall, P. and Ma, Y. (2007). Testing the suitability of polynomial models in errors-in-variables problems. *Ann. Statist.* **35**, 2620–2638.

Hart, J. (2013). *Nonparametric smoothing and lack-of-fit tests.* Springer Science & Business Media.

Jones, D. Y., Schatzkin, A., Green, S. B., Block, G., Brinton, L. A., Ziegler, R. G., Hoover, R. and Taylor, P. R. (1987). Dietary fat and breast cancer in the national health and nutrition survey I: epidemiologic follow-up study. *J. Natl. Cancer Inst.* **79**, 465–471.

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

Koul, H. L. and Song, W. (2009). Minimum distance regression model checking with berkson measurement errors. *Ann. Statist.* **37**, 132–156.

Koul, H. L. and Song, W. (2010). Model checking in partial linear regression models with berkson measurement errors. *Statist. Sinica* **20**, 1551–1579.

Lee, L. F. and Sepanski, J. H. (1995). Estimation of linear and nonlinear errors-in-variables models using validation data. *J. Amer. Statist. Assoc.* **90**, 130–140.

Li, B. and Wang, S. L. (2007). On directional regression for dimension deduction. *J. Amer. Statist. Assoc.* **102**, 997–1008.

Li, B., Wen, S. and Zhu, L. X. (2008). On a projective resampling method for dimension reduction with multivariate responses. *J. Amer. Statist. Assoc.* **103**, 1177–1186.

Li, B. and Yin, X. R. (2007). On surrogate dimension reduction for measurement error regression: an invaraince law. *Ann. Statist.* **35**, 2143–2172.

Li, B., Zha, H. Y. and Chiaromonte, F. (2005). Contour regression: a general approach to dimension reduction. *Ann. Statist.* **33**, 1580–1616.

Li, K. C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86**, 316–342.

Li, K. C. (1992). On principal hessian directions for data visualization and dimension reduction: another application of stein's lemma. *J. Amer. Statist. Assoc.* **87**, 1025–1039.

Lue, H. H. (2004). Principal hessian directions for regression with measurement error. *Biometrika.* **91**, 409–423.

Song, W. X. (2008). Model checking in errors-in-variables regression. *J. Multivariate Anal.* **99**, 2406–2443.

Song, W. X. (2009). Lack-of-fit testing in errors-in-variables regression model with validation data. *Statist. Probab. Lett.* **79**, 765–773.

Stute, W., Xue, L. G. and Zhu, L. X. (2007). Empirical likelihood inference in nonlinear errors-in-covariables models with validation data. *J. Amer. Statist. Assoc.* **102**, 332–346.

Wu, M. L., Whittemore, A. S. and Jung, D. L. (1986). Errors in reported dietary intakes. *Am. J. Epidemiol.* **124**, 826–835.

Xia, Y., Tong, H., Li, W. K. and Zhu, L. X. (2002). An adaptive estimation of dimension reduction space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64**, 363–410.

Xu, W. and Zhu, L. X. (2014). Nonparametric check for partial linear errors-in-covariables models with validation data. *Ann. Inst. Stat. Math.* **67**, 1–23.

Zhang, J., Zhu, L. P. and Zhu, L. X. (2014). Surrogate dimension reduction in measurement error regressions. *Statist. Sinica* **24**, 1341–1363.

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation technique. *J. Econometrics.* **75**, 263–289.

Zhu, L. X., Cui, H. J. and Ng, K. W. (2004). Some properties of a lack-of-fit test for a linear errors in variables Model. *Acta Math. Appl. Sin.* **20**, 533–540.

Zhu, L. X. and Cui, H. J. (2005). Testing the adequacy for a general linear errors-in-variables model. *Statist. Sinica* **15**, 1049–1068.

Zhu, L. X., Miao, B. Q. and Peng, H. (2006). On sliced inverse regression with high-dimensional covariates. *J. Amer. Statist. Assoc.* **100**, 630–643.

Zhu, L. X., Song, W. X. and Cui, H. J. (2003). Testing lack-of-fit for a polynomial errors-in-variables model. *Acta Math. Appl. Sin.* **19**, 353–362.

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

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

Department of Statistics and Probability, 639 Red Cedar St., Michigan State University, East Lansing, Michigan 48824, USA.

E-mail: koul@msu.edu

Department of Statistics, Jinan University, No.601, Huangpu Avenue, Guangzhou, Guangdong, China.

E-mail: clxie510@outlook.com

Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong.

E-mail: lzhu@hkbu.edu.hk

(Received October 2016; accepted December 2017)