

Model checking for parametric single-index models: A dimension-reduction model-adaptive approach

Xu Guo

Department of Mathematics, Hong Kong Baptist University, Hong Kong and College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing

Tao Wang

Department of Biostatistics, Yale School of Public Health, New Haven

and Lixing Zhu[†]

School of Statistics, Beijing Normal University, Beijing and Department of Mathematics, Hong Kong Baptist University, Hong Kong

Summary. Local smoothing testing based on multivariate nonparametric regression estimation is one of the main model checking methodologies in the literature. However, the relevant tests suffer from typical curse of dimensionality, resulting in slow convergence rates to their limits under the null hypothesis and less deviation from the null hypothesis under alternative hypotheses. This problem prevents tests from maintaining the significance level well and makes tests less sensitive to alternative hypotheses. In this paper, a **model-adaptation** concept in lack-of-fit testing is introduced and a dimension-reduction model-adaptive test procedure is proposed for parametric single-index models. The test behaves like a local smoothing test, as if the model was univariate. It is consistent against any global alternative hypothesis and can detect local alternative hypotheses distinct from the null hypothesis at a fast rate that existing local smoothing tests can achieve only when the model is univariate. Simulations are conducted to examine the performance of our methodology. An analysis of real data is shown for illustration. The method can be readily extended to global smoothing methodology and other testing problems.

Keywords: Dimension reduction; parametric single index models; **model-adaptation**; model checking.

1. Introduction

Consider the following parametric single-index regression model:

$$Y = g(\beta_0^T \mathbf{X}, \theta_0) + \epsilon, \quad (1)$$

[†]*Address for correspondence:* Lixing Zhu, lzhu@hkbu.edu.hk.

where Y is a scalar response, \mathbf{X} is a predictor vector in p dimensions, $g(\cdot)$ is a known square integrable continuous function, β_0 is a p -dimensional unknown index vector, θ_0 is a d -vector of parameters and $E(\epsilon|\mathbf{X}) = 0$. Throughout, the ‘T’ superscript denotes the transpose operator.

As is well known, statistical analysis based on this model can be easily and accurately performed under the assumption that this parametric structure is correct. Therefore, in practice, parametric models are frequently used. Compared with nonparametric regression models that do not specify the function form of $E(Y|\mathbf{X})$, the parametric single index model can lead to more accurate estimates, especially when the dimension p of \mathbf{X} is high, provided that the model is correctly built. However, if the parametric single index model is not correctly fitted for the data, we must apply some other more flexible regression models, otherwise, further statistical analysis would give the wrong conclusions. A model check must therefore be performed. A practical example is production theory, in which the Cobb-Douglas function is commonly used to describe the linear relationship between the log-inputs, such as labor and capital, and the log-output. However, this function may not describe the relationship well. To avoid model mis-specification, Kumbhakar et al. (2007) developed semi-parametric regression models (see also Simar et al. 2014). Financial markets are another example, in which linear error correction model is used to describe the dynamics of spot and futures prices. This model implicitly assumes that pricing errors are reduced at a speed that is independent of the magnitude of the price deviation between spot and futures. However, this may be not the case. To relax this assumption, Gaul and Theissen (2015) developed a partial linear regression model and proposed a test for the adequacy of the linear error correction model. Zhang and Wu (2011), Lin et al. (2014) and Lahaye and Shaw (2014) worked on tests of parametric functional form in nonstationary time series, fixed effects panel data models and heterogenous autoregressive models, respectively. In summary, when describing the regression relationship between responses and predictors, we must make a choice between the simple but fragile parametric model and the flexible but complicated nonparametric model.

To make reliable statistical inferences for model (1), we should carry out suitable and efficient model checking. The literature contains a number of options for testing mode (1) against a general alternative model:

$$Y = G(\mathbf{X}) + \eta, \quad (2)$$

where $G(\cdot)$ is an unknown smooth function and $E(\eta|\mathbf{X}) = 0$. There are two general classes of methods, local and global smoothing methods. In local smoothing methods, Härdle and Mammen (1993) considered the L_2 distance between the null parametric regression and the alternative nonparametric regression as the base of their test statistic construction. Zheng (1996) proposed a quadratic form of the conditional moment test that was also independently developed by Fan and Li (1996). Dette (1999) developed a test based on the difference between variance estimates under the null and alternative models. See also Fan et al. (2001), Koul and Ni (2004), Zhang and Dette (2004) and Van

Keilegom et al. (2008). In global smoothing approaches based on empirical processes, Stute (1997) introduced nonparametric principal component decomposition based on a residual marked empirical process. Inspired by the Khmaladze transformation used in goodness-of-fitting for distributions, Stute et al. (1998b) first developed the innovation martingale approach to obtain distribution-free tests. Stute and Zhu (2002) was a relevant reference for parametric single-index models. Khmaladze and Koul (2004) studied the goodness-of-fit problem for errors in nonparametric regression. González-Manteiga and Crujeiras (2013) provided a comprehensive review of the research in this area.

The empirical studies in the literature have shown that the existing local smoothing methods are sensitive to high-frequency regression models and thus can have high power in detecting these alternative models. However, a very obvious and serious shortcoming is that these methods suffer severely from dimensionality due to the inevitable use of multivariate nonparametric function estimation. Under the corresponding null hypotheses, existing local smoothing test statistics converge to their limits at the rate $O(n^{-1/2}h^{-p/4})$ (or $O(n^{-1}h^{-p/2})$ if the tests are in quadratic forms), which is very slow when p is large. Therefore, the significance level very often cannot be maintained when used with a moderate sample size. Härdle and Mammen (1993) provided a good summary of this method. This problem has been acknowledged in the literature and there are a number of local smoothing tests that apply re-sampling or Monte Carlo approximation to help determine critical values (or p values), for example as described by Härdle and Mammen (1993), Delgado and González-Manteiga (2001), Härdle et al. (2004), Dette et al. (2007) and Neumeyer and Van Keilegom (2010). These tests can also only detect alternative hypotheses distinct from the null hypothesis at the rate of order $O(n^{-1/2}h^{-p/4})$ (see, e.g., Zheng 1996). Asymptotically, existing local smoothing tests are therefore less powerful for detecting alternative models. In contrast, although their rate is of order $1/\sqrt{n}$, most of the existing global smoothing methods depend on high-dimensional stochastic processes (see, e.g., Stute et al. 1998a). Their power performance often drops significantly as p increases due to the data sparseness in high-dimensional space.

It is thus of critical importance to investigate how to make local smoothing methods get rid of the curse of dimensionality when model (1) is the hypothetical model with a dimension-reduction structure. To motivate our method, we very briefly review the basic idea of local smoothing approaches. Under the null hypothesis, $E\{Y - g(\beta_0^T \mathbf{X}, \theta_0) | \mathbf{X}\} = E(\epsilon | \mathbf{X}) = 0$. Under the alternative model (2), $E\{Y - g(\beta_0^T \mathbf{X}, \theta_0) | \mathbf{X}\} \neq 0$. Thus, the empirical version with root- n consistent estimates of β_0 and θ_0 can be used as a base to construct test statistics. Note that under the null hypothesis, $E(Y | \mathbf{X}) - g(\beta_0^T \mathbf{X}, \theta_0) = 0$. The distance between a nonparametric estimate of $E(Y | \mathbf{X})$ and a parametric estimate of $g(\beta_0^T \mathbf{X}, \theta_0)$ is then a base for test statistic construction (see, e.g., Härdle and Mammen 1993). However, these methods inevitably involve high-dimensional nonparametric estimation of $E(Y | \mathbf{X})$ or $E(\epsilon | \mathbf{X})$. This is the main cause of inefficiency in hypothesis testing with aforementioned slow rate of order $O(n^{-1/2}h^{-p/4})$.

To address this problem, we note that under the null hypothesis, $E\{Y - g(\beta_0^T \mathbf{X}, \theta_0) | \mathbf{X}\} = E(\epsilon | \beta_0^T \mathbf{X}) = 0$. Thus, a naive idea is to construct a test statistic based on $E\{Y - g(\beta_0^T \mathbf{X}, \theta_0) | \beta_0^T \mathbf{X}\}$. Intuitively, this can sufficiently explore the information provided in the hypothetical model. From the technical development in the present paper, it is easy to see that a relevant test can have a rate of order $O(n^{-1/2}h^{-1/4})$ as if the dimension of \mathbf{X} was 1. Stute and Zhu (2002) conducted related work, using a global smoothing method. However, this idea leads to another very obvious shortcoming; specifically, as the test statistic construction is based completely on the hypothetical model, any test will be directional rather than omnibus. The general alternative model (2) cannot be handled. For instance, when the alternative model is $E(Y | \mathbf{X}) = g(\beta_0^T \mathbf{X}, \theta_0) + \tilde{g}(\beta_1^T \mathbf{X})$, where β_1 is orthogonal to β_0 and \mathbf{X} follows the standard multivariate normal distribution $N(0, I_p)$. Then $\beta_0^T \mathbf{X}$ is independent of $\beta_1^T \mathbf{X}$. When $E\{\tilde{g}(\beta_1^T \mathbf{X})\} = 0$, it is clear that under this alternative model, $E\{Y - g(\beta_0^T \mathbf{X}, \theta_0) | \beta_0^T \mathbf{X}\} = 0$ still holds. Thus, a test statistic based on $E\{Y - g(\beta_0^T \mathbf{X}, \theta_0) | \beta_0^T \mathbf{X}\}$ cannot detect the above alternative. Xia (2009) proposed a test statistic comparing the empirical cross-validation counterparts of the minimum of $E^2\{\epsilon - E(\epsilon | \alpha^T \mathbf{X})\}$ over all of the unit vectors α with the centered residual sum of squares, where ϵ is the error term of the model. However, this procedure cannot provide the corresponding limiting distributions under the null and alternative hypotheses and thus cannot test significance at a nominal level. As was pointed out by Xia, under the null hypothesis, the rejection frequency tends to 0 as $n \rightarrow \infty$. In other words, this method cannot control the significance level asymptotically. Cross validation also involves intensive computation. Xia also provided a single-indexing bootstrap F test, but the consistency of this method has not been established. Thus, in certain sense it is hard for users to recognize type I and type II errors.

The above discussions suggest that a good test should use the information under the null model (1) to avoid the dimensionality problem, as if the dimension of \mathbf{X} was 1, and adapt to the alternative model (2) so that it can detect general alternative models. We use a sufficient dimension reduction (SDR) technique (Cook 1998) to achieve this goal. Although SDR has been studied intensively over the past two decades (see Subsections 2.2-2.5), this is the first time that it has been used as an efficient tool in lack-of-fit testing. The theoretical and numerical results obtained in this paper suggest that SDR has potential for the development of lack-of-fit testing for other problems. The basic idea is as follows. Note that for any orthogonal $p \times p$ matrix B , $G(\mathbf{X}) = G(BB^T \mathbf{X}) := \tilde{G}(B^T \mathbf{X})$. Thus, $E\{G(\mathbf{X}) - g(\beta_0^T \mathbf{X}, \theta_0) | \mathbf{X}\} \neq 0$ is equivalent to $E\{\tilde{G}(B^T \mathbf{X}) - g(\beta_0^T \mathbf{X}, \theta_0) | B^T \mathbf{X}\} \neq 0$. It is also clear that $E(\epsilon | \mathbf{X}) = 0$ is equivalent to $E(\epsilon | B^T \mathbf{X}) = 0$. Based on this observation, we consider a more parsimonious alternative model that is widely used in SDR:

$$Y = G(B^T \mathbf{X}) + \eta, \quad (3)$$

where B is a $p \times q$ matrix with q orthogonal columns for an unknown number q with $1 \leq q \leq p$, G is an unknown smooth function, and $E(\eta | \mathbf{X}) = 0$. When $q = p$, this model is identical to model (2). When $q = 1$ and $B = \beta_0 / \|\beta_0\|_2$ is a column vector, model (3) reduces to a single-index model with the

same index as that in the null model (1). Here $\|\cdot\|_2$ denotes the L_2 norm. Thus, it offers us a way to construct a test that is automatically adaptive to the null and alternative models through consistently determining q and estimating B (or BC for a $q \times q$ orthogonal matrix C) under both the null and alternative models. A local smoothing test will be constructed using this idea in Section 2, giving two nice and somewhat surprising features: the test sufficiently uses the dimension-reduction structure under the null model and is an omnibus test for detecting general alternative models as existing local smoothing tests try to do. More precisely, the test statistic under the null model converges to its limit at the faster rate of order $O(n^{-1/2}h^{-1/4})$ (or $O(n^{-1}h^{-1/2})$ if the test is in a quadratic form), is consistent against any global alternative model and can detect local alternative models distinct from the null model at the rate of order $O(n^{-1/2}h^{-1/4})$. This is a significant improvement, particularly when p is large, because the new test behaves like a local smoothing test, as if \mathbf{X} was one-dimensional. Thus, the test is expected to maintain the significance level well and have better power performance than existing local smoothing tests.

The paper is organized as follows. As SDR plays a crucial role, we briefly review it in the next section. In Section 2, a dimension-reduction model-adaptive (DRMA) test is constructed. The asymptotic properties under the null and alternative models are investigated in Section 3. In Section 4, the simulation results are reported and a data analysis example using real data is conducted for illustration. The basic idea can be readily applied to other test procedures, the details of which we leave to Section 5. The conditions are described in the Appendix. The proofs of the theoretical results are given in the Supplementary Material. The data set can be obtained from the link at <http://archive.ics.uci.edu/ml/datasets/Auto+MPG> and the computer codes for simulations can be obtained from the link at <http://www.math.hkbu.edu.hk/~lzhu/publications.html>.

2. DRMA test procedure

2.1. Basic test construction

We formulate the null hypothesis as follows:

$$H_0 : \quad \exists \beta_0 \in \mathbb{R}^p, \theta_0 \in \mathbb{R}^d, \text{ such that, } E(Y|\mathbf{X}) = g(\beta_0^T \mathbf{X}, \theta_0).$$

The alternative hypothesis is for any $\beta \in \mathbb{R}^p$, $\theta \in \mathbb{R}^d$ and a $p \times q$ matrix B with $1 \leq q \leq p$:

$$H_1 : \quad E(Y|\mathbf{X}) = E(Y|B^T \mathbf{X}) \neq g(\beta^T \mathbf{X}, \theta). \quad (4)$$

The null and alternative models can then be unified. Under the null hypothesis, $q = 1$. Then $B = \tilde{\beta} = \beta_0 / \|\beta_0\|_2$. Under the alternative hypothesis, $q \geq 1$. In this subsection, let $\epsilon = Y - g(\beta_0^T \mathbf{X}, \theta_0)$ denote the random error under the null hypothesis. Thus, under H_0 ,

$$E(\epsilon|\mathbf{X}) = 0 \implies E(\epsilon|\beta_0^T \mathbf{X}) = E(\epsilon|B^T \mathbf{X}) = 0.$$

Then,

$$E\{\epsilon E(\epsilon|B^T \mathbf{X})W(B^T \mathbf{X})\} = E\{E^2(\epsilon|B^T \mathbf{X})W(B^T \mathbf{X})\} = 0, \quad (5)$$

where $W(B^T \mathbf{X})$ is some positive weight function, discussed below.

Under H_1 , we have $E(\epsilon|B^T \mathbf{X}) = E(Y|B^T \mathbf{X}) - g(\beta_0^T \mathbf{X}, \theta_0) \neq 0$. Thus

$$E\{\epsilon E(\epsilon|B^T \mathbf{X})W(B^T \mathbf{X})\} = E\{E^2(\epsilon|B^T \mathbf{X})W(B^T \mathbf{X})\} > 0. \quad (6)$$

Here is a toy example to illustrate the above point. Consider the model:

$$Y = X_1 + aX_2^3 + \eta,$$

where $\mathbf{X} = (X_1, X_2, X_3, X_4)^T$ and η is the error term independent of \mathbf{X} . $a = 0$ corresponds to the null hypothesis with $\beta_0 = (1, 0, 0, 0)^T$ and $a \neq 0$ to the alternative hypotheses with $B = (\beta_0, \beta_1)$, where $\beta_1 = (0, 1, 0, 0)^T$. Thus, under H_0 , $q = 1, B = \beta_0, \epsilon = \eta$, and under H_1 , $q = 2, \epsilon = aX_2^3 + \eta$. As a result, $E(\epsilon|B^T \mathbf{X}) = E(aX_2^3 + \eta|X_1, X_2) = aX_2^3 \neq 0$ when $a \neq 0$. Clearly, q and B are according to the underlying regression model.

The empirical version of the left hand side in (5) can be used as a test statistic. H_0 will be rejected for large values of the test statistic. As B is generally unknown with unknown dimension q , we estimate $E(\epsilon|B^T \mathbf{X})$ as follows: when a sample $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$ is available,

$$\hat{E}\{\epsilon_i|\hat{B}(\hat{q})^T \mathbf{x}_i\} = \frac{\frac{1}{n-1} \sum_{j \neq i}^n \hat{\epsilon}_j K_h\{\hat{B}(\hat{q})^T \mathbf{x}_i - \hat{B}(\hat{q})^T \mathbf{x}_j\}}{\frac{1}{n-1} \sum_{j \neq i}^n K_h\{\hat{B}(\hat{q})^T \mathbf{x}_i - \hat{B}(\hat{q})^T \mathbf{x}_j\}}.$$

Here, $\hat{\epsilon}_j = y_j - g(\hat{\beta}^T \mathbf{x}_j, \hat{\theta})$, $\hat{\beta}$ and $\hat{\theta}$ are the commonly used least squares estimates of β_0 and θ_0 , $\hat{B}(\hat{q})$ is an SDR estimate with an estimated structural dimension \hat{q} of q and $K_h(\cdot) = K(\cdot/h)/h^{\hat{q}}$, with $K(\cdot)$ a \hat{q} -dimensional kernel function and h a bandwidth. As the estimates of B and q are crucial for the DRMA test, we will specify them later. The weight $W(\cdot)$ is chosen to be the density function $f(B^T \mathbf{X})$ of $B^T \mathbf{X}$, which is estimated by

$$\hat{f}\{\hat{B}(\hat{q})^T \mathbf{x}_i\} = \frac{1}{n-1} \sum_{j \neq i}^n K_h\{\hat{B}(\hat{q})^T \mathbf{x}_i - \hat{B}(\hat{q})^T \mathbf{x}_j\}.$$

A non-standardized test statistic is defined by

$$V_n = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \hat{\epsilon}_i \hat{\epsilon}_j K_h\{\hat{B}(\hat{q})^T (\mathbf{x}_i - \mathbf{x}_j)\}. \quad (7)$$

REMARK 1. *The test statistic suggested by Zheng (1996) is*

$$\tilde{V}_n = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \hat{\epsilon}_i \hat{\epsilon}_j \tilde{K}_h(\mathbf{x}_i - \mathbf{x}_j). \quad (8)$$

Here, $\tilde{K}_h(\cdot) = \tilde{K}(\cdot/h)/h^p$, with $\tilde{K}(\cdot)$ a p -dimensional kernel function. There are two main differences between equations (7) and (8). First, our test uses $\hat{B}(\hat{q})^T \mathbf{X}$ in lieu of \mathbf{X} in Zheng (1996)'s test and

applies $K_h(\cdot)$ instead of $\tilde{K}_h(\cdot)$. This reduces the dimension p down to \hat{q} . Second, under H_0 , we will show that $\hat{q} \rightarrow 1$, $\hat{B}(\hat{q}) \rightarrow \beta_0/\|\beta_0\|_2$ and $nh^{1/2}V_n$ has a finite limit. Under the alternative model (3), we will show that $\hat{q} \rightarrow q \geq 1$ and $\hat{B}(\hat{q}) \rightarrow BC$ for a $q \times q$ orthogonal matrix C .

REMARK 2. If we assume that there is an oracle who knows the true B under the alternative (3), Zheng's oracle test statistic should be

$$V_n^{ZHO} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \hat{\epsilon}_i \hat{\epsilon}_j K_h^O\{B^T(\mathbf{x}_i - \mathbf{x}_j)\}. \quad (9)$$

Here, $K_h^O(\cdot) = K^O(\cdot/h)/h^q$, with $K^O(\cdot)$ a q -dimensional kernel function. In the toy example, $B = \{(1, 0, 0, 0)^T, (0, 1, 0, 0)^T\}$. Although the above test statistic uses a low-dimensional kernel estimate, the dimension q is fixed according to the alternative model. In other words, V_n^{ZHO} does not adapt to the dimension under both the null and alternative models. It can be easily shown that to make the test statistic have finite limiting distribution under the null hypothesis, the standardizing constant should be $nh^{q/2}$ to get $nh^{q/2}V_n^{ZHO}$. As discussed in Remark 1, our test statistic V_n exploits the dimension-reduction structure under the null hypothesis, giving $nh^{1/2}V_n$ a finite limit. The standardizing constant $nh^{1/2}$ has a beneficial effect for alternatives with $q > 1$, because it makes $nh^{1/2}V_n$ larger than $nh^{q/2}V_n^{ZHO}$. Under global alternatives, the main difference between V_n and V_n^{ZHO} is the difference in the two standardizing constants. $nh^{1/2}V_n$ is approximately related to $nh^{q/2}V_n^{ZHO}$ as follows:

$$nh^{1/2}V_n = h^{(1-q)/2}nh^{q/2}V_n^{ZHO}.$$

With $q > 1$, $h^{(1-q)/2}$ is divergent to infinity as $n \rightarrow \infty$ and $h \rightarrow 0$, and hence $nh^{1/2}V_n$ can be more powerful. In Section 4, we will compare the performance of our test, Zheng's test and the oracle Zheng's test.

2.2. Identification and estimation of B and q

In general, B is not identifiable because for any $q \times q$ orthogonal matrix C , $G(B^T\mathbf{X})$ can also be written $\tilde{G}(C^TB^T\mathbf{X})$. It is thus sufficient to identify BC for a $q \times q$ orthogonal matrix C . To achieve this, we use the SDR methodology. Define $\mathcal{S}_{E(Y|\mathbf{X})}$ as the intersection of all subspaces \mathcal{S}_A , spanned by matrices A , such that $Y \perp\!\!\!\perp E(Y|\mathbf{X})|A^T\mathbf{X}$, where $\perp\!\!\!\perp$ means 'is independent of'. In SDR, $\mathcal{S}_{E(Y|\mathbf{X})}$ is called the central mean subspace, and its dimension, denoted by $d_{E(Y|\mathbf{X})}$, is called the structural dimension (Cook and Li 2002). Similarly, the central subspace (Cook 1998), denoted by $\mathcal{S}_{Y|\mathbf{X}}$, is defined as the intersection of all subspaces \mathcal{S}_A of minimal dimension, such that $Y \perp\!\!\!\perp \mathbf{X}|A^T\mathbf{X}$. In model (3), we have $\mathcal{S}_{E(Y|\mathbf{X})} = \text{span}(B)$ and $d_{E(Y|\mathbf{X})} = q$. In other words, we can identify q basis vectors of $\mathcal{S}_{E(Y|\mathbf{X})}$. The literature contains several proposals, including sliced inverse regression (SIR, Li 1991), sliced average variance estimation (SAVE, Cook and Weisberg 1991), contour regression (Li et al. 2005), directional regression (Li and Wang 2007), likelihood acquired directions (Cook and

Forzani 2009), discretization-expectation estimation (DEE, Zhu et al. 2010a), and average partial mean estimation (Zhu et al. 2010b). The minimum average variance estimation (MAVE, Xia et al, 2002) can identify and estimate the relevant space with fewer regularity conditions on \mathbf{X} , but requires nonparametric smoothing. As DEE and MAVE have good performance in general, we review these two methods below.

2.3. A review of DEE

Our test procedure involves the estimation of the $p \times q$ matrix B and the dimension q . In this subsection, we first assume that the dimension q is known and then discuss how to determine it consistently. SIR and SAVE are two popular SDR methods involving the partition of the range of Y into several slices. However, as documented by many authors (Li 1991, Zhu and Ng 1995, and Li and Zhu 2007), the choice of the number of slices can affect the efficiency and can even yield inconsistent estimates. To avoid this, Zhu et al. (2010a) introduced DEE. The basic idea is simple. We first define the new response variable $Z(t) = I(Y \leq t)$, which takes the value 1 if $Y \leq t$ and 0 otherwise. Let $\mathcal{S}_{Z(t)|\mathbf{X}}$ be the central subspace and $\mathcal{M}(t)$ be a $p \times p$ positive semi-definite matrix, such that $\text{span}\{\mathcal{M}(t)\} = \mathcal{S}_{Z(t)|\mathbf{X}}$. Define $\mathcal{M} = E\{\mathcal{M}(T)\}$. Under certain mild conditions, $\mathcal{M} = \mathcal{S}_{Y|\mathbf{X}}$. To ensure that $\mathcal{S}_{Y|\mathbf{X}} = \mathcal{S}_{E(Y|\mathbf{X})}$ for η in model (3), we assume that $\eta = G_1(B^T \mathbf{X})\tilde{\eta}$, where $G_1(\cdot)$ is an unknown smooth function and $\tilde{\eta} \perp \perp \mathbf{X}$. $\eta \perp \perp \mathbf{X}$ is a special case.

In the discretization step, we construct a new sample $\{\mathbf{x}_i, z_i(y_j)\}$ with $z_i(y_j) = I(y_i \leq y_j)$. For each fixed y_j , we estimate $\mathcal{M}(y_j)$ using SIR or SAVE. Let $\mathcal{M}_n(y_j)$ denote the candidate matrix obtained from the chosen method. In the expectation step, we estimate \mathcal{M} by $\mathcal{M}_{n,n} = n^{-1} \sum_{j=1}^n \mathcal{M}_n(y_j)$. The q eigenvectors of $\mathcal{M}_{n,n}$ corresponding to its q largest eigenvalues are then used to form an estimate of B . Denote the DEE procedure based on SIR and SAVE to be DEE_{SIR} and DEE_{SAVE} , respectively. To save space, we focus on these two basic methods. Zhu et al. (2010a) proved that $\hat{B}(q)$ is a root- n consistent estimator of BC for a $q \times q$ non-singular matrix C when q is given. For DEE_{SIR} , a mild linearity condition is often assumed: $E(\mathbf{X}|B^T \mathbf{X} = u)$ is linear in u . For DEE_{SAVE} , we require the linearity condition and a constant covariance condition: $\text{Var}(\mathbf{X}|B^T \mathbf{X})$ is a constant matrix.

2.4. A review of MAVE

In contrast, MAVE requires fewer regularity conditions, but needs local smoothing in high-dimensional space. In the following, we use MAVE to estimate the basis vectors of $\mathcal{S}_{E(Y|\mathbf{X})}$.

From the population, MAVE minimizes the objective function

$$E\{Y - E(Y|B^T \mathbf{X})\}^2 \quad \text{subject to } B^T B = I_q.$$

It is equivalent to minimize the following problem:

$$\min_{B \in \mathbb{R}^{p \times q}} E\{\sigma_B^2(B^T \mathbf{X})\} \quad \text{subject to } B^T B = I_q,$$

where $\sigma_B^2(B^T \mathbf{X}) = E[\{Y - E(Y|B^T \mathbf{X})\}^2 | B^T \mathbf{X}]$ is the conditional variance of Y given $B^T \mathbf{X}$.

When a sample $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$ is available, the corresponding estimate \hat{B} is the minimizer of

$$\sum_{j=1}^n \sum_{i=1}^n (y_i - a_j - \mathbf{d}_j^T B^T \mathbf{x}_{ij})^2 K_h(B^T \mathbf{x}_{ij}),$$

over all B satisfying $B^T B = I_q$, a_j and \mathbf{d}_j , where $\mathbf{x}_{ij} = \mathbf{x}_i - \mathbf{x}_j$. The details of the algorithm are given in Xia et al. (2002). Compared with DEE, MAVE only requires $E(\eta|\mathbf{X}) = 0$ for η in model (3), which is a very mild condition. Again, $\hat{B}(q)$ has been shown to be consistent for BC for a $q \times q$ orthogonal matrix C when q is given.

2.5. Estimation of the structural dimension q

For DEE, according to Zhu et al. (2010a), we can determine q by

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

where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_p \geq 0$ are the eigenvalues of $\mathcal{M}_{n,n}$. We note that the first term in the bracket can be considered as a likelihood ratio and that the second term is a penalty term. Zhu et al. (2010a) proved that under some regularity conditions, \hat{q} is a consistent estimate of q .

For MAVE, we suggest a BIC-type criterion that is a modified version of that proposed by Wang and Yin (2008) of the following form:

$$BIC_k = \log \left(\frac{RSS_k}{n} \right) + \frac{\log(n)k}{\min(nh^k, \sqrt{n})},$$

where RSS_k is the residual sum of squares and k is the estimate of the dimension. Let $B(k)$ denote the matrix B when the dimension is k . We then have

$$RSS_k = \sum_{j=1}^n \sum_{i=1}^n \{y_i - \hat{a}_j - \hat{\mathbf{d}}_j^T \hat{B}(k)^T \mathbf{x}_{ij}\}^2 K_h\{\hat{B}(k)^T \mathbf{x}_{ij}\}.$$

The estimated dimension is then

$$\hat{q} = \arg \min_{1 \leq k \leq p} (BIC_k).$$

Wang and Yin (2008) showed that under some mild conditions, \hat{q} is also a consistent estimate of q .

PROPOSITION 1. *Under conditions 8, 9 in the Appendix, the DEE-based estimate $\hat{q} \rightarrow q$ as $n \rightarrow \infty$. Under conditions 3, 4 and 7 in the Appendix, the MAVE-based estimate $\hat{q} \rightarrow q$ as $n \rightarrow \infty$. Consequently, the estimate $\hat{B}(\hat{q})$ is a consistent estimate of BC for a $q \times q$ orthogonal matrix C .*

These consistencies are established under the null and global alternative models. Under the local alternative models, specified in Section 3, the results are different and we will show the consistency of \hat{q} to 1.

3. Asymptotic properties

3.1. Limiting null distribution

Notice that β_0 is the true parameter value under H_0 . Let $\mathbf{Z} = \beta_0^T \mathbf{X}$, $\sigma^2(\mathbf{z}) = E(\epsilon^2 | \mathbf{Z} = \mathbf{z})$, and

$$\begin{aligned} Var &= 2 \int K^2(u) du \cdot \int \{\sigma^2(\mathbf{z})\}^2 f^2(\mathbf{z}) d\mathbf{z}, \\ \widehat{Var} &= \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{1}{h^q} K^2 \left\{ \frac{\hat{B}(\hat{q})^T (\mathbf{x}_i - \mathbf{x}_j)}{h} \right\} \hat{\epsilon}_i^2 \hat{\epsilon}_j^2. \end{aligned}$$

We first state the asymptotic property of the test statistic under the null hypothesis.

THEOREM 1. *Under H_0 and the conditions in the Appendix, we have*

$$nh^{1/2}V_n \Rightarrow N(0, Var).$$

Further, Var can be consistently estimated by \widehat{Var} .

We now standardize V_n to obtain a scale-invariant statistic:

$$T_n = \sqrt{\frac{n-1}{n}} \frac{nh^{1/2}V_n}{\sqrt{\widehat{Var}}}.$$

COROLLARY 1. *Under H_0 and the conditions in the Appendix, we have*

$$T_n^2 \Rightarrow \chi_1^2,$$

where χ_1^2 is the chi-square distribution with one degree of freedom.

From this corollary, we can calculate the p values easily using its limiting null distribution. A Monte Carlo simulation can also be used. We will discuss this in Section 4.

3.2. Power study

We now examine the power performance of the proposed test statistic under a sequence of alternative models with the form

$$H_{1n} : Y = g(\beta_0^T \mathbf{X}, \theta_0) + C_n G(B^T \mathbf{X}) + \eta, \quad (10)$$

where $E(\eta | \mathbf{X}) = 0$ and the function $G(\cdot)$ satisfies $E\{G^2(B^T \mathbf{X})\} < \infty$. When C_n is a fixed constant, it reduces to the global alternative model (3), whereas when $C_n \rightarrow 0$, the models are local alternative models. In this sequence of models, β_0 is one of the columns in B .

Let $m(\mathbf{X}, \beta_0, \theta_0) = \text{grad}_{\beta, \theta} \{g(\beta_0^T \mathbf{X}, \theta_0)\}^T$ be the gradient of $g(\beta_0^T \mathbf{X}, \theta_0)$, $H(\mathbf{X}) = G(B^T \mathbf{X})m(\mathbf{X}, \beta_0, \theta_0)$ and $\Sigma_x = E\{m(\mathbf{X}, \beta_0, \theta_0)m(\mathbf{X}, \beta_0, \theta_0)^T\}$.

We first give the asymptotic behavior of the estimate \hat{q} under the local alternative models.

THEOREM 2. *Under the local alternative models of (10) with $C_n = n^{-1/2}h^{-1/4}$ and the conditions in the Appendix, we have $\hat{q} \rightarrow 1$. Here, \hat{q} is either the DEE-based or the MAVE-based estimate.*

Note that the central subspace of this alternative model is $\text{span}(B)$. That is, at the population level under the local alternatives, the structural dimension should be q . However, the above theorem shows that at the sample level, the estimate we defined before is consistent to 1, rather than q . In this case, by sufficient dimension reduction theory, with a probability going to 1 the corresponding estimated matrix \hat{B} with $\hat{q} = 1$ asymptotically estimates a vector proportional to β_0 at a certain rate.

We are now ready to present the result on power performance.

THEOREM 3. *Under the conditions in the Appendix, we have:*

(i) *under the global alternative of (3),*

$$T_n/(nh^{1/2}) \Rightarrow \text{Constant} > 0;$$

(ii) *under the local alternatives of (10) with $C_n = n^{-1/2}h^{-1/4}$, $nh^{1/2}V_n \Rightarrow N(\mu, \text{Var})$ and $T_n^2 \Rightarrow \chi_1^2(\mu^2/\text{Var})$, where*

$$\mu = E \left(\left[G(B^T \mathbf{X}) - m(\mathbf{X}, \beta_0, \theta_0)^T \Sigma_x^{-1} E\{H(\mathbf{X})\} \right]^2 f(\beta_0^T \mathbf{X}) \right),$$

and $\chi_1^2(\mu^2/\text{Var})$ is a noncentral chi-squared random variable with one degree of freedom and the non-centrality parameter μ^2/Var .

REMARK 3. *Together with the discussions in Remark 2, we further discuss the difference between our test and Zheng's test and give an intuitive explanation of why our test is more powerful than Zheng's test. For the local alternative models (10), from (6), we have $\epsilon = Y - g(\beta_0^T \mathbf{X}, \theta_0) = C_n G(B^T \mathbf{X}) + \eta$, $E(\epsilon|B^T \mathbf{X}) = C_n G(B^T \mathbf{X})$. In other words, $C_n G(B^T \mathbf{X}) = E(Y|B^T \mathbf{X}) - g(\beta_0^T \mathbf{X}, \theta_0)$ is the difference between the functions under the null and alternative models. Thus,*

$$V := E\{\epsilon E(\epsilon|B^T \mathbf{X})W(B^T \mathbf{X})\} = C_n^2 E\{G(B^T \mathbf{X})^2 W(B^T \mathbf{X})\}.$$

Both V_n and \tilde{V}_n are empirical versions of V . The standardizing constants must be $nh^{1/2}$ and $nh^{p/2}$, respectively, such that our test statistic T_n and Zheng's test \tilde{T}_n have finite limiting distributions under the null hypothesis as we discussed in Remark 1. As shown in Theorem 3, the rate $nh^{1/2}$ then determines that under the global alternative model with fixed C_n , T_n goes to infinity at the rate $nh^{1/2}$. Under the local alternative models, when C_n tends to zero at the rate $n^{-1/2}h^{-1/4}$, T_n still has a constant drift. However, Zheng's test statistic \tilde{T}_n diverges to infinity only at the rate $nh^{p/2}$ under the global alternative model and C_n can only have the rate $n^{-1/2}h^{-p/4}$ under the local alternative models. Our test can detect local alternative models distinct from the hypothetical model at a rate as if the dimension of \mathbf{X} was one. Without this model-adaptiveness property, we would not have this power enhancement. **Model-adaptation** is a generic methodology that can be used to construct other tests.

REMARK 4. *Note that $\mu = 0$ is equivalent to $G(B^T \mathbf{X}) = m(\mathbf{X}, \beta_0, \theta_0)^T \Sigma_x^{-1} E\{H(\mathbf{X})\}$ almost surely. Moreover $\Sigma_x^{-1} E\{H(\mathbf{X})\}$ is a $(p+d)$ -dimensional vector $l(\beta_0, \theta_0)$. Further recall that $m(\mathbf{X}, \beta_0, \theta_0) =$*

$\{\mathbf{X}\partial g(\beta_0^T \mathbf{X}, \theta_0)/\partial(\beta_0^T \mathbf{X}), \partial g(\beta_0^T \mathbf{X}, \theta_0)/\partial\theta_0\}$, and $C_n G(B^T \mathbf{X}) = E(Y|B^T \mathbf{X}) - g(\beta_0^T \mathbf{X}, \theta_0)$ is the difference between the functions under the null and alternative. Thus generally $G(B^T \mathbf{X})$ should not be equal to $m(\mathbf{X}, \beta_0, \theta_0)^T \Sigma_x^{-1} E\{H(\mathbf{X})\}$. In other words, $\mu > 0$ holds. For example, assume that $g(\beta_0^T \mathbf{X}, \theta_0) = \beta_0^T \mathbf{X}$, then $m(\mathbf{X}, \beta_0, \theta_0) = \mathbf{X}$. Thus to make $\mu = 0$, $G(B^T \mathbf{X})$ must be equal to $\mathbf{X}^T l(\beta_0, \theta_0)$, a linear function of \mathbf{X} . However, this is not the case.

4. Numerical analysis

4.1. Simulations

We now carry out simulations to examine the performance of the proposed test. To save space, we concentrate on SIR-based DEE and MAVE. Four simulation studies are considered. In Study 1, $B = \beta$ under both the null and alternative hypotheses. The aim of this study is to compare the DEE-based and MAVE-based tests. We also make a comparison with Stute and Zhu (2002)'s test, because it works well in this scenario. The comparison can be used to examine whether and how much our test loses in power. In Study 2, we compare our test with Stute and Zhu (2002)'s test to show that our test is omnibus whereas Stute and Zhu's test is directional. The purpose of Study 3 is to examine the effect of dimensionality on both our test and a local smoothing test (Zheng 1996). We choose Zheng's test for comparison because it has a tractable limiting null distribution, which can be used to calculate the critical values (or p values). The literature shows that local smoothing tests often do not maintain the significance level well when limiting null distributions are used. We also use the re-sampling version of Zheng's test to determine the critical values (or p values). We conducted a comparison with Härdle and Mammen (1993)'s test in a small-scale simulation. As the conclusions were very similar, the results are not reported here. In Study 4, we compare our test with the oracle version of Zheng's test, which was defined in Remark 2 in Subsection 2.1. This comparison examines the effect of estimating q on our test and the effect of dimensionality on Zheng's test. In the comparison, we use a model with no dimension reduction structure under the alternative hypothesis, so that we can show the performance of our adaptive test when the dimension $q = p$ must be estimated, although it is not necessary in this case. We consider linear models in the first three studies. In the fourth study, the null model is nonlinear.

Study 1. Consider

$$H_{11} : Y = \beta^T \mathbf{X} + a \cos(0.6\pi\beta^T \mathbf{X}) + \eta;$$

$$H_{12} : Y = \beta^T \mathbf{X} + a \exp\{-(\beta^T \mathbf{X})^2\} + \eta;$$

$$H_{13} : Y = \beta^T \mathbf{X} + a(\beta^T \mathbf{X})^2 + \eta.$$

Here, $a = 0$ corresponds to the null hypothesis and $a \neq 0$ the alternative hypotheses. The null model is a linear model and the alternative models are all single-index models. Let $p = 8$, $\beta = (1, 1, \dots, 1)^T / \sqrt{p}$. $\mathbf{X} = (X_1, X_2, \dots, X_p)^T$ and η are independent. The observations $\mathbf{x}_i, i = 1, 2, \dots, n$,

are generated i.i.d. from the multivariate normal distribution $N(0, \Sigma_1)$ or $N(0, \Sigma_2)$ with $\Sigma_1 = I_{p \times p}$ and $\Sigma_2 = (0.5^{|j-l|})_{p \times p}$. The η_i 's are drawn independently from $N(0, 1)$. In this simulation study, the replication time is 2,000. The first model is a high-frequency model that is in favor of local smoothing tests. The other two models are low-frequency models that are in favor of global smoothing tests.

In the nonparametric regression estimation, we use the kernel function $K(u) = 15/16(1 - u^2)^2$ if $|u| \leq 1$, and 0 otherwise. The bandwidth is taken to be $h = 1.5n^{-1/(4+\hat{q})}$ with separately standardized predictors for simplicity. Our limited empirical experience suggests that it works well.

Tables 1-3 show the empirical sizes and powers of our proposed test against the alternatives $H_{1i}, i = 1, 2, 3$ with the nominal size $\alpha = 0.05$. Let T_n^{DEE} and T_n^{MAVE} denote two versions of test T_n based on DEE and MAVE, respectively. For all of the cases considered here, T_n^{DEE} controls the size very well, even at sample size $n = 50$. We can therefore rely on the limiting null distribution for determining critical values (or p values). For T_n^{MAVE} , unreported results show that the empirical size tends to be slightly larger than 0.05. Empirically, we recommend using an adjusted version,

$$\tilde{T}_n^{MAVE} = \frac{T_n^{MAVE}}{1 + 4n^{-4/5}},$$

which is asymptotically equivalent to T_n^{MAVE} . Tables 1-3 show that it performs better with small to moderate sample sizes.

As is well known, local smoothing tests suffer from the dimensionality problem; they cannot maintain the significance level and have good power performance at the same time. Thus, re-sampling techniques are often used in finite samples. A typical technique is the wild bootstrap first suggested by Wu (1986) and developed further by others (see, e.g., Härdle and Mammen 1993). Consider the bootstrap observations $y_i^* = \hat{\beta}^T \mathbf{x}_i + \hat{\epsilon}_i \times U_i$. Here, $\{U_i\}_{i=1}^n$ can be chosen to be i.i.d. Bernoulli variates with

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

Let T_n^* be the bootstrap version of T_n , based on the bootstrap sample $(\mathbf{x}_1, y_1^*), \dots, (\mathbf{x}_n, y_n^*)$. The null hypothesis is rejected if T_n is bigger than the corresponding quantile of the bootstrap distribution of T_n^* . The bootstrap versions of T_n based on DEE and MAVE are denoted by T_n^{DEE*} and T_n^{MAVE*} , respectively.

Tables 1-3 show that T_n^{DEE} , T_n^{DEE*} , T_n^{MAVE*} and \tilde{T}_n^{MAVE} have comparable significance levels. Thus, the re-sampling technique is unnecessary for our DRMA tests, although T_n^{MAVE} needs some adjustment to get \tilde{T}_n^{MAVE} . For power performance, when $\mathbf{X} \sim N(0, \Sigma_1)$, \tilde{T}_n^{MAVE} generally has a higher power than T_n^{DEE} . However, when \mathbf{X} follows $N(0, \Sigma_2)$, T_n^{DEE} becomes the winner. T_n^{MAVE*} has a slightly higher power than T_n^{DEE*} . For the alternatives H_{11} and H_{12} , T_n^{MAVE*} has a relatively higher power than \tilde{T}_n^{MAVE} , whereas for the alternative H_{13} , \tilde{T}_n^{MAVE} is more powerful. In the DEE-based tests, in almost all of the cases, T_n^{DEE} has a higher power than the bootstrapped version T_n^{DEE*} . All of the tests are generally sensitive to the alternatives in the sense that, as a increases, the power

increases quickly. In summary, MAVE-based tests tend to be more conservative than DEE-based tests. T_n^{DEE} controls the size satisfactorily and has high power.

Tables 1-3 about here

Stute and Zhu's (2002) test is a dimension reduction test. They developed an innovation process transformation of the empirical process $n^{-1/2} \sum_{i=1}^n \{y_i - g(\hat{\beta}^T \mathbf{x}_i, \hat{\theta})\} I(\hat{\beta}^T \mathbf{x}_i \leq u)$. This test has been shown to be powerful in many scenarios (see, e.g., Stute and Zhu 2002, Mora and Moro-Egido 2008). Stute and Zhu (2002)'s test is denoted T_n^{SZ} . The results are presented in Tables 1 and 3. For H_{11} , T_n^{DEE} is found to be more powerful, particularly when $\mathbf{X} \sim N(0, \Sigma_2)$, but T_n^{SZ} does not perform well. H_{11} corresponds to a high-frequency model, for which existing numerical studies in the literature have already suggested that local smoothing based methods work better, as shown by the results here. For the alternatives H_{12} and H_{13} , T_n^{SZ} becomes the winner. However, our tests are robust to the underlying models. We use Study 2 below to demonstrate that Stute and Zhu (2002)'s test is not an omnibus test.

Study 2. Data are generated from the model

$$Y = \beta_1^T \mathbf{X} + a(\beta_2^T \mathbf{X})^3 + \eta. \quad (11)$$

Consider two cases, $p = 3$, $\beta_1 = (1, 0, 0)^T$, $\beta_2 = (0, 1, 0)^T$ and $p = 4$, $\beta_1 = (1, 1, 0, 0)^T / \sqrt{2}$, $\beta_2 = (0, 0, 1, 1)^T / \sqrt{2}$. When $p = 3$, $n = 50, 100$. When $p = 4$, $n = 100$. In both cases, \mathbf{X} and η are independently generated from the multivariate and univariate standard normal distributions, respectively. We use $a = 0.0, 0.3, \dots, 1.5$. To save space, we only present the results of T_n^{DEE} in Figure 1. It is obvious that T_n^{DEE} performs much better than T_n^{SZ} , as T_n^{SZ} has very low power and T_n^{DEE} efficiently detects the alternative models. However, for a fair comparison of global and local smoothing tests, we should incorporate our model-adaptive technique into a global smoothing test, which is possible in principle. This is an ongoing project.

Figure 1 about here

To investigate the performance of our test when there is more than one direction under the alternative hypothesis and to investigate the effect of dimensionality on local smoothing tests, we conduct the following simulation study.

Study 3. The data are generated from the model

$$Y = \beta_1^T \mathbf{X} + a(\beta_2^T \mathbf{X})^2 + \eta, \quad (12)$$

where $\beta_1 = (\underbrace{1, \dots, 1}_{p/2}, 0, \dots, 0)^T / \sqrt{p/2}$ and $\beta_2 = (0, \dots, 0, \underbrace{1, \dots, 1}_{p/2})^T / \sqrt{p/2}$. Under the null model, we have $B = \beta_1$. Under the alternative models, $B = (\beta_1, \beta_2)$. We take $p = 2$ and $p = 8$.

The observations $\mathbf{x}_i, i = 1, 2, \dots, n$ are generated i.i.d. from the multivariate normal distribution $N(0, \Sigma_1)$ or $N(0, \Sigma_2)$. The η_i 's are generated from $N(0, 1)$ or the double exponential distribution $DE(0, \sqrt{2}/2)$ with density $f(x) = \sqrt{2}/2 \exp\{-\sqrt{2}|x|\}$. To save space, we only consider T_n^{DEE} and T_n^{DEE*} due to their good performance in size control and easy computation. In theory, the dimension reduction method DEE cannot identify the direction in the quadratic term of this model. However, we can see that even in this case, our test works well.

When $p = 2$, $\beta_1 = (1, 0)^T$ and $\beta_2 = (0, 1)^T$. The results are reported in Table 4 and show that Zheng (1996)'s test T_n^{ZH} maintains the significance level reasonably in some cases, but the empirical size is generally lower than it. The bootstrap version T_n^{ZH*} performs better in this aspect. In contrast, both T_n^{DEE} and T_n^{DEE*} maintain the significance level very well. We can find that T_n^{ZH} and T_n^{DEE} generally have higher empirical powers than their bootstrap versions. However, the differences are negligible in our tests. In other words, our DRMA test does not need assistance from the bootstrap approach, whereas Zheng's test does. Both T_n^{DEE} and T_n^{DEE*} are uniformly and significantly more powerful than T_n^{ZH} and T_n^{ZH*} .

Table 4 about here

We now consider the $p = 8$ case. The results are reported in Table 5 and show that when p increases from 2 to 8, the performance of Zheng's test deteriorates significantly. Table 5 indicates that the empirical size of T_n^{ZH} is far away from the significance level. The empirical size of T_n^{ZH*} is much better, but is still not very close to the nominal level. Again, T_n^{DEE} maintains the significance level very well and T_n^{DEE*} is not better than T_n^{DEE} . T_n^{ZH} performs much worse than in the $p = 2$ case and its bootstrap version does not enhance its power. T_n^{DEE} is not significantly affected by the increase in dimension.

Table 5 about here

These findings coincide with the prior theoretical results: existing local smoothing tests have much slower convergence rate (of order $n^{-1/2}h^{-p/4}$ or $n^{-1}h^{-p/2}$ if the tests are quadratic forms) to their limits under the null model. They are much less sensitive under the alternative models because, in theory, the tests diverge to infinity at a much slower rate of order $n^{1/2}h^{p/4}$ than the DRMA test. The simulation results show that we can simply use the DMRA tests limiting null distribution to determine critical values without a heavy computational burden.

As we comment above, the null models in Studies 1-3 are linear. We now consider a nonlinear null model. To determine whether Zheng's test can also benefit the dimension reduction structure, we compare our test, the oracle Zheng test V_n^{ZHO} and the original Zheng test. When $p = q$, the oracle Zheng test V_n^{ZHO} is equivalent to the original Zheng test.

Study 4. Consider three models:

$$H_{41} : Y = 0.25 \exp(2X_1) + a(X_2^3) + \eta;$$

$$H_{42} : Y = 0.25 \exp(2X_1) + a\{X_2^3 + \cos(\pi X_3) + X_4\} + \eta;$$

$$H_{43} : Y = 0.25 \exp(2X_1) + a\{X_2^3 + \cos(\pi X_3) + X_4 + |X_5| + |X_6|^{1.5} + X_7 * X_8\} + \eta;$$

where $X_i, i = 1, \dots, p$ are independent of η and both follow $N(0, 1)$ and $N(0, 0.5)$, respectively. For the null model, $\theta = 0.25$ and $\beta_1 = (2, 0, \dots, 0)^T$. Define β_i to be the unit vector, in which the i -th element is 1, $i = 2, \dots, 8$. When $a \neq 0$, $q = 2, B = (\beta_1/2, \beta_2)$ for H_{41} , $q = 4, B = (\beta_1/2, \beta_2, \beta_3, \beta_4)$ for H_{42} and $q = 8, B = (\beta_1/2, \beta_2, \dots, \beta_8)$ for H_{43} . Note that the third model under the alternative is regarded as a full model without dimension reduction structure. It is used to examine the usefulness of **model-adaptation**, although an extra and unnecessary estimation of $q = p$ is required. Recall that T_n^{ZHO} is the oracle Zheng test statistic. Consider $a = 0, 0.1, \dots, 0.5, p = 4, 8$ and $n = 100$. T_n^{ZHO} is equivalent to T_n^{ZH} under H_{42} with $p = 4$. They are also equivalent under H_{43} with $p = 8$. The bootstrap versions T_n^{ZHO*} and T_n^{ZH*} , of T_n^{ZHO} and T_n^{ZH} , respectively, are used to maintain the significance level. The bootstrap versions are not required for our test. T_n^{DEE} can control the significance level very well for all three models, as all of the sizes are around 0.053. The results are presented in Figure 2.

Figure 2 about here

We make the following observations. First, T_n^{ZHO*} has greater power than T_n^{DEE} under H_{41} ($q = 2$), although in theory, under the alternative models our test can diverge to infinity faster than the oracle Zheng's test. This seems to indicate the negative effect of the estimation of q . For H_{42} ($q = 4$), T_n^{DEE} has the best performance. For H_{43} ($q = 8$), T_n^{DEE} also performs much better than T_n^{ZHO*} . These results imply that even for the purely nonparametric regression model ($p = q = 4, H_{42}$ or $p = q = 8, H_{43}$), our **model-adaptation** procedure can have gains in the power performance. As explained, this improvement mainly comes from the **model-adaptation**. The standardizing constant $nh^{1/2}$ can be used, which has a faster convergence rate to infinity. Further, comparing sub-figures 1 with 3 or sub-figures 2 with 4, it is clear that the power of T_n^{ZH*} decreases quickly as p increases. These results suggest that estimating q has a negative effect on our test and that dimensionality is a more serious issue for Zheng's test.

4.2. Real data analysis

A data set is obtained from the Machine Learning Repository at the University of California-Irvine (<http://archive.ics.uci.edu/ml/datasets/Auto+MPG>). The data set was first analysed by Quinlan (1993). Recently, Xia (2007) analysed this data set with their method. There are 406 observations in the original data set. To illustrate our method, we first clear the units with missing response

and/or predictor values, leaving 392 sample points. The response variable is miles per gallon (Y). There are seven predictors, the number of cylinders (X_1), engine displacement (X_2), horsepower (X_3), vehicle weight (X_4), time to accelerate from 0 to 60 mph (X_5), model year (X_6) and origin of the car (1 = American, 2 = European, 3 = Japanese). As the origin of the car contains more than two categories, we follow Xia (2007)'s suggestions to define two indicator variables. Let $X_7 = 1$ if a car is from America and 0 otherwise. Let $X_8 = 1$ if a car is from Europe and 0 otherwise. All of the predictors are standardized separately. We aim to predict the response in terms of the eight predictors $\mathbf{X} = (X_1, \dots, X_8)^T$. Quinlan (1993) used a simple linear regression model. However, we need to check its adequacy to avoid model mis-specification. The value of T_n^{DEE} is 9.7040 and the p value is 0. The value of \tilde{T}_n^{MAVE} is 16.2309 and the p value is also 0. A linear regression model is therefore not plausible for predicting the response. Figure 3 suggests that a nonlinear model should be used. With DEE, $\hat{q} = 1$, thus a single-index model may be appropriate.

Figure 3 about here

5. Discussions

In this paper, we propose a DRMA test procedure and use Zheng's test as an example to construct a test statistic. The method is readily extendable to other local smoothing methods, as discussed in Section 1. The same principle can also be applied to global smoothing methods. As discussed in Section 2, under the null hypothesis $Y = g(\beta_0^T \mathbf{X}, \theta_0) + \epsilon$ with $E(\epsilon|\mathbf{X}) = 0$, we can have $E(\epsilon|\beta_0^T \mathbf{X}) = E(\epsilon|B^T \mathbf{X}) = 0$. Under the alternative model H_1 , $E(\epsilon|B^T \mathbf{X}) = E\{Y - g(\beta_0^T \mathbf{X}, \theta_0)|B^T \mathbf{X}\} = G(B^T \mathbf{X}) - g(\beta_0^T \mathbf{X}, \theta_0) \neq 0$. We should note that the above B can be different under the null and the alternative models, as illustrated in the toy example. This motivates us to define the following test statistic:

$$R_n(\mathbf{z}) = n^{-1/2} \sum_{i=1}^n \left\{ y_i - g(\hat{\beta}^T \mathbf{x}_i, \hat{\theta}) \right\} I\{\hat{B}(\hat{q})^T \mathbf{x}_i \leq \mathbf{z}\}.$$

This statistic is different from that in Stute and Zhu (2002), in which $\hat{B}(\hat{q}) = \hat{\beta}$. As shown in the simulations, Stute and Zhu's (2002) test is a directional test and is inconsistent under general alternative models. Work on this problem is ongoing.

Extensions of our methodology to missing, censored or dependent data can also be considered. For example, let δ_i be a missing indicator: $\delta_i = 1$ if y_i is observed, otherwise it is equal to zero. Assume that the response is missing at random. Therefore $P(\delta = 1|\mathbf{X}, Y) = P(\delta = 1|\mathbf{X}) := \pi(\mathbf{X})$. For more details, see Little and Rubin(1987). Again, we test whether the following regression model holds or not, $H_0 : Y = g(\beta_0^T \mathbf{X}, \theta_0) + \epsilon$ with $E(\epsilon|\mathbf{X}) = 0$ and Y missing at random. Note that under the null hypothesis, $E\{\delta\epsilon/\pi(\mathbf{X})|\beta^T \mathbf{X}\} = E\{\delta\epsilon/\pi(\mathbf{X})|B^T \mathbf{X}\} = 0$, whereas under the alternative model, $E\{\delta\epsilon/\pi(\mathbf{X})|B^T \mathbf{X}\} \neq 0$. Similarly, we can construct a consistent test statistic of the following form:

$$V_{n1} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \frac{\delta_i}{\hat{\pi}(\mathbf{x}_i)} \frac{\delta_j}{\hat{\pi}(\mathbf{x}_j)} \hat{\epsilon}_i \hat{\epsilon}_j K_h\{\hat{B}(\hat{q})^T (\mathbf{x}_i - \mathbf{x}_j)\},$$

where $\hat{\pi}(\mathbf{x}_i)$ is a nonparametric or parametric estimate of $\pi(\mathbf{x}_i)$, $\hat{\epsilon}_i = y_i - g(\hat{\beta}^T \mathbf{x}_i, \hat{\theta})$ and $\hat{\beta}, \hat{\theta}$ and $\hat{B}(\hat{q})$ are obtained from the completely observed units. Another possible test statistic is

$$V_{n2} = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i}^n \delta_i \delta_j \hat{\epsilon}_i \hat{\epsilon}_j K_h \{ \hat{B}(\hat{q})^T (\mathbf{x}_i - \mathbf{x}_j) \}.$$

This corresponds to the test statistic obtained from the complete case.

We can consider applying the methodology to other testing problems, such as testing for homoscedasticity, parametric quantile regression model and conditional parametric density function of Y given \mathbf{X} . This research is ongoing.

Alternatively, we may rely on a multiple testing procedure for all of the marginal functions $E(\epsilon|X_i)$ for $i = 1, \dots, p$. However, note that $E(\epsilon|\mathbf{X}) = 0 \Rightarrow E(\epsilon|X_i) = 0$ for $i = 1, \dots, p$, but they are not equivalent. This approach is therefore still not necessary and sufficient, unless the function $E(\epsilon|\mathbf{X}) = 0$ is equivalent to there being a X_i such that $E(\epsilon|X_i) = 0$. Note that H_0 can be rejected if

$$\hat{q} > 1 \text{ or } \left(\hat{q} = 1 \text{ but } |\hat{E}\{\epsilon \hat{E}(\epsilon|\hat{B}^T \mathbf{X})\}| \text{ is large} \right).$$

A two-stage testing procedure would be feasible. However, a big challenge is that this testing procedure cannot determine both type I error and type II error when we use $\hat{q} > 1$ in the first stage in the case where the alternative model is with $q > 1$. Thus, in our testing procedure, we do not separate the above two hypotheses and the identification of the structural dimension q is adaptive to the underlying models. When $\hat{q} > 1$, the alternative can be detected, and when $\hat{q} = 1$, the alternative function can also be detected. In both cases, the type I and type II error can be computed. On the other hand, it is worthwhile to point out that a two-stage testing procedure deserves a further study because in the case where $\hat{q} > 1$, it seems that by using the estimation of q would offer a possible way to handle the testing problem.

In summary, the proposed methodology is a general one that can be readily applied to many other testing problems.

Acknowledgements

The authors thank the Editor, the Associate Editor and three referees for their constructive comments and suggestions, which helped us clarify more precisely the construction and theoretical results. The research described here was supported by a grant from the University Grants Council of Hong Kong, Hong Kong and the Fundamental Research Funds for the Central Universities (NR2015001). The manuscript was begun when the first author was at Hong Kong Baptist University and was finished at Nanjing University of Aeronautics and Astronautics.

Appendix. Conditions of the theorems

The following conditions are assumed for the theorems in Section 3.

- 1) There exists an integrable function $L(x)$ such that $|m_i(x, \beta, \theta)| \leq L(x)$ for all (β, θ) and $1 \leq i \leq d + p$. $g(\mathbf{X}^T \beta, \theta)$ is a Borel measurable function on \mathbb{R}^p for each β, θ and a twice continuously differentiable real function on a compact subset of \mathbb{R}^p and \mathbb{R}^d , Λ and Θ for each $x \in \mathbb{R}^p$. Here, $m(\mathbf{X}, \beta_0, \theta_0) = \text{grad}_{\beta, \theta} \{g(\beta_0^T \mathbf{X}, \theta_0)\}^T$.
- 2) Let $\gamma = (\beta, \theta)^T$ and $\tilde{\gamma}_0$ be the value of γ that minimizes $\tilde{S}_{0n}(\gamma) = E[\{E(Y|\mathbf{X}) - g(\mathbf{X}^T \beta, \theta)\}^2]$. $\tilde{\gamma}_0$ is an interior point and is the unique minimizer of the function \tilde{S}_{0n} . $\Sigma_x = E\{m(\mathbf{X}, \beta_0, \theta_0)m(\mathbf{X}, \beta_0, \theta_0)^T\}$ is positive definite.
- 3) $E|Y|^k < \infty$, $E\|\mathbf{X}\|_2^k < \infty$ for all $k > 0$. $\sup E(X_l^2|B^T \mathbf{X}) < \infty, l = 1, \dots, p$; $E(\eta^2|B^T \mathbf{X}) < \infty$, $G(\cdot)$ has bounded, continuous third-order derivatives. $E(\mathbf{X}|Y)$ and $E(\mathbf{X}\mathbf{X}^T|Y)$ have bounded, continuous third-order derivatives. Here, $\eta = Y - E(Y|B^T \mathbf{X})$.
- 4) The density function $f(\mathbf{X})$ of \mathbf{X} has bounded second derivatives and is bounded away from 0 in a neighborhood around 0. The density function $f(Y)$ of Y has bounded derivative and is bounded away from 0 on a compact support. The conditional densities $f_{\mathbf{X}|Y}(\cdot)$ of \mathbf{X} given Y and $f_{(\mathbf{X}_0, \mathbf{X}_l)|(Y_0, Y_l)}(\cdot)$ of $(\mathbf{X}_0, \mathbf{X}_l)$ given (Y_0, Y_l) are bounded for all $l \geq 1$.
- 5) The density $f(B^T \mathbf{X})$ of $B^T \mathbf{X}$ on support \mathcal{C} exists, has two bounded derivatives and satisfies

$$0 < \inf_{B^T \mathbf{X} \in \mathcal{C}} f(B^T \mathbf{X}) \leq \sup_{B^T \mathbf{X} \in \mathcal{C}} f(B^T \mathbf{X}) < \infty.$$

- 6) $nh^2 \rightarrow \infty$ under the null (1) and local alternative hypothesis (10); $nh^q \rightarrow \infty$ under the global alternative hypothesis (3).
- 7) $K(\cdot)$ is a spherically symmetric density function with a bounded derivative and support, all of the moments of $K(\cdot)$ exist and $\int UU^T K(U)dU = I$.
- 8) $\mathcal{M}_n(t)$ has the following expansion:

$$\mathcal{M}_n(t) = \mathcal{M}(t) + E_n\{\psi(\mathbf{X}, Y, t)\} + R_n(t),$$

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

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

REMARK 5. *Conditions 1) and 2) are necessary for the asymptotic properties of the least squares estimates $\hat{\beta}$ and $\hat{\theta}$. Conditions 3), 4) and 7) are required for MAVE. Condition 5) is needed for the asymptotic normality of our statistic. In Condition 6), $nh^2 \rightarrow \infty$ is a usual assumption in nonparametric estimation. Conditions 8) and 9) are assumed for DEE. Under the linearity condition and the constant conditional variance condition, DEE_{SIR} and DEE_{SAVE} satisfy the conditions 8) and 9).*

Table 1. Empirical sizes and powers of \tilde{T}_n^{MAVE} , T_n^{DEE} and T_n^{SZ} for H_0 vs. H_{11} and H_{12} in *Study 1*, with $X \sim N(0, \Sigma_i)$, $i = 1, 2$ and $\epsilon \sim N(0, 1)$.

| | a | \tilde{T}_n^{MAVE} | | T_n^{DEE} | | T_n^{SZ} | |
|---------------------------------|-----|----------------------|-----------|-------------|-----------|------------|-----------|
| | | $n = 50$ | $n = 100$ | $n = 50$ | $n = 100$ | $n = 50$ | $n = 100$ |
| $H_{11}, X \sim N(0, \Sigma_1)$ | 0 | 0.0630 | 0.0565 | 0.0470 | 0.0500 | 0.0730 | 0.0680 |
| | 0.2 | 0.0890 | 0.1535 | 0.0730 | 0.1263 | 0.0940 | 0.0850 |
| | 0.4 | 0.2175 | 0.4735 | 0.1623 | 0.3857 | 0.1070 | 0.1870 |
| | 0.6 | 0.4290 | 0.8125 | 0.3207 | 0.7227 | 0.2030 | 0.3200 |
| | 0.8 | 0.6470 | 0.9630 | 0.4910 | 0.9207 | 0.2930 | 0.5010 |
| | 1.0 | 0.8135 | 0.9990 | 0.6347 | 0.9803 | 0.3410 | 0.6390 |
| $X \sim N(0, \Sigma_2)$ | 0 | 0.0460 | 0.0545 | 0.0480 | 0.0563 | 0.0640 | 0.0580 |
| | 0.2 | 0.0570 | 0.1050 | 0.0767 | 0.1173 | 0.0720 | 0.0780 |
| | 0.4 | 0.1165 | 0.3255 | 0.1647 | 0.3667 | 0.0790 | 0.0910 |
| | 0.6 | 0.2275 | 0.6530 | 0.3243 | 0.7203 | 0.0910 | 0.0900 |
| | 0.8 | 0.4100 | 0.8885 | 0.4953 | 0.9293 | 0.0870 | 0.0900 |
| | 1.0 | 0.5230 | 0.9690 | 0.6787 | 0.9887 | 0.0920 | 0.1250 |
| $H_{12}, X \sim N(0, \Sigma_1)$ | 0 | 0.0535 | 0.0610 | 0.0490 | 0.0470 | 0.0670 | 0.0620 |
| | 0.2 | 0.1275 | 0.1845 | 0.0990 | 0.1687 | 0.1460 | 0.2390 |
| | 0.4 | 0.2950 | 0.5695 | 0.2657 | 0.5510 | 0.3960 | 0.6500 |
| | 0.6 | 0.5715 | 0.8895 | 0.5383 | 0.8980 | 0.7030 | 0.9480 |
| | 0.8 | 0.8100 | 0.9945 | 0.7763 | 0.9890 | 0.8940 | 0.9950 |
| | 1.0 | 0.9410 | 1.0000 | 0.9267 | 0.9993 | 0.9630 | 1.0000 |
| $X \sim N(0, \Sigma_2)$ | 0 | 0.0395 | 0.0380 | 0.0460 | 0.0523 | 0.0700 | 0.0540 |
| | 0.2 | 0.0700 | 0.1160 | 0.0773 | 0.1140 | 0.1260 | 0.1480 |
| | 0.4 | 0.1545 | 0.3690 | 0.1820 | 0.3717 | 0.2590 | 0.4220 |
| | 0.6 | 0.3285 | 0.6920 | 0.3660 | 0.6953 | 0.4650 | 0.7390 |
| | 0.8 | 0.5490 | 0.9025 | 0.5723 | 0.9130 | 0.6280 | 0.9030 |
| | 1.0 | 0.7340 | 0.9805 | 0.7587 | 0.9857 | 0.8060 | 0.9830 |

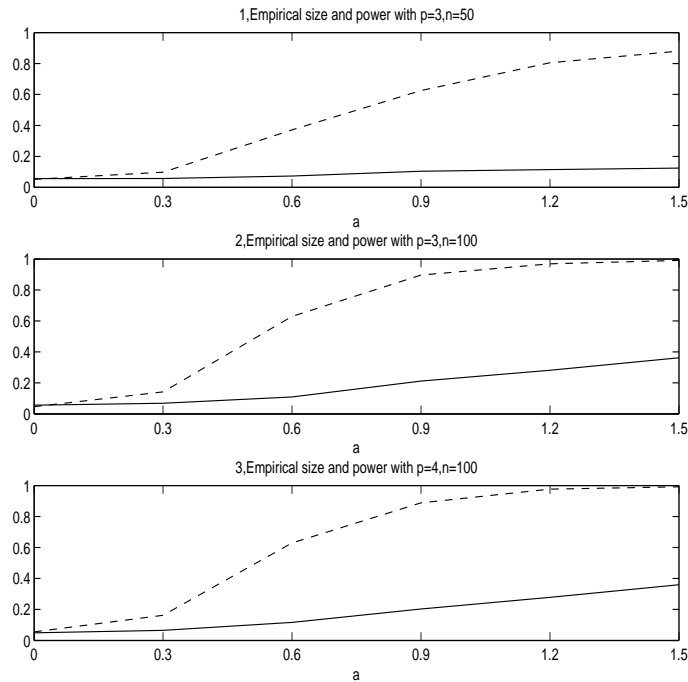


Fig. 1. The empirical size and power curves of T_n^{SZ} and T_n^{DEE} in *Study 2*. The solid and dashed line represent the results of T_n^{SZ} and T_n^{DEE} , respectively.

Table 2. Empirical sizes and powers of T_n^{MAVE*} and T_n^{DEE*} for H_0 vs. H_{11} and H_{12} in Study 1, with $X \sim N(0, \Sigma_i)$, $i = 1, 2$ and $\epsilon \sim N(0, 1)$.

| | a | T_n^{MAVE*} | | T_n^{DEE*} | |
|---------------------------------|-----|---------------|-----------|--------------|-----------|
| | | $n = 50$ | $n = 100$ | $n = 50$ | $n = 100$ |
| $H_{11}, X \sim N(0, \Sigma_1)$ | 0 | 0.0697 | 0.0580 | 0.0470 | 0.0500 |
| | 0.2 | 0.1160 | 0.1890 | 0.0840 | 0.1425 |
| | 0.4 | 0.2740 | 0.5260 | 0.1635 | 0.3900 |
| | 0.6 | 0.4670 | 0.8510 | 0.3255 | 0.7235 |
| | 0.8 | 0.6750 | 0.9750 | 0.5115 | 0.9185 |
| | 1.0 | 0.8320 | 0.9960 | 0.6160 | 0.9780 |
| $X \sim N(0, \Sigma_2)$ | 0 | 0.0490 | 0.0550 | 0.0500 | 0.0475 |
| | 0.2 | 0.0770 | 0.1160 | 0.0680 | 0.1285 |
| | 0.4 | 0.1600 | 0.3460 | 0.1515 | 0.3480 |
| | 0.6 | 0.2850 | 0.7170 | 0.3080 | 0.7155 |
| | 0.8 | 0.4520 | 0.8930 | 0.4835 | 0.9255 |
| | 1.0 | 0.5800 | 0.9740 | 0.6660 | 0.9820 |
| $H_{12}, X \sim N(0, \Sigma_1)$ | 0 | 0.0770 | 0.0740 | 0.0435 | 0.0555 |
| | 0.2 | 0.1520 | 0.2070 | 0.1095 | 0.1680 |
| | 0.4 | 0.3320 | 0.5990 | 0.2545 | 0.5510 |
| | 0.6 | 0.6170 | 0.9200 | 0.5600 | 0.8975 |
| | 0.8 | 0.8410 | 0.9990 | 0.7690 | 0.9925 |
| | 1.0 | 0.9430 | 1.0000 | 0.9215 | 1.0000 |
| $X \sim N(0, \Sigma_2)$ | 0 | 0.0570 | 0.058 | 0.0420 | 0.0540 |
| | 0.2 | 0.0990 | 0.1250 | 0.0705 | 0.1275 |
| | 0.4 | 0.2070 | 0.3560 | 0.1770 | 0.3495 |
| | 0.6 | 0.3710 | 0.7070 | 0.3380 | 0.6870 |
| | 0.8 | 0.5920 | 0.9150 | 0.5390 | 0.9190 |
| | 1.0 | 0.7600 | 0.9810 | 0.7445 | 0.9805 |

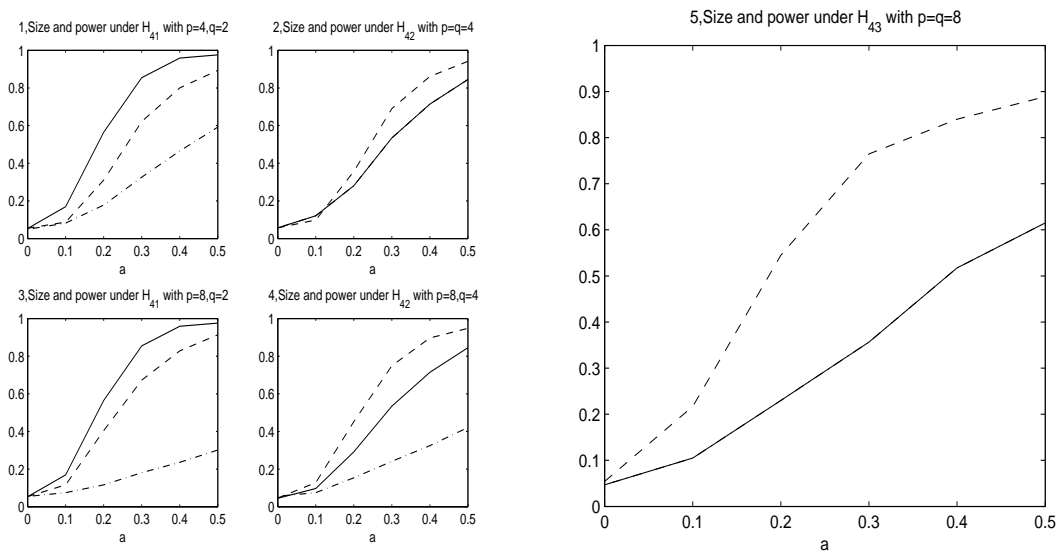


Fig. 2. The empirical size and power curves of T_n^{ZHO*} , T_n^{DEE} and T_n^{ZH*} in Study 4 with $p = 4$ and $p = 8$. The solid, dashed and dashed-dotted line represent the results of T_n^{ZHO*} , T_n^{DEE} and T_n^{ZH*} , respectively.

Table 3. Empirical sizes and powers for H_0 vs. H_{13} in *Study 1*, with $X \sim N(0, \Sigma_i)$, $i = 1, 2$ and $\epsilon \sim N(0, 1)$.

| | a | \tilde{T}_n^{MAVE} | | T_n^{DEE} | | T_n^{SZ} | |
|--|-----|----------------------|-----------|--------------|-----------|------------|-----------|
| | | $n = 50$ | $n = 100$ | $n = 50$ | $n = 100$ | $n = 50$ | $n = 100$ |
| $X \sim N(0, \Sigma_1), \epsilon \sim N(0, 1)$ | 0 | 0.0515 | 0.0625 | 0.0507 | 0.0563 | 0.0690 | 0.0570 |
| | 0.2 | 0.1255 | 0.2245 | 0.1067 | 0.2000 | 0.2430 | 0.4250 |
| | 0.4 | 0.3465 | 0.7520 | 0.3330 | 0.7127 | 0.6020 | 0.9060 |
| | 0.6 | 0.6390 | 0.9790 | 0.6170 | 0.9580 | 0.8180 | 0.9940 |
| | 0.8 | 0.8335 | 0.9980 | 0.8023 | 0.9960 | 0.9370 | 1.0000 |
| | 1.0 | 0.9240 | 1.0000 | 0.8897 | 0.9993 | 0.9840 | 1.0000 |
| $X \sim N(0, \Sigma_2), \epsilon \sim N(0, 1)$ | 0 | 0.0485 | 0.0505 | 0.0477 | 0.0480 | 0.0670 | 0.0650 |
| | 0.2 | 0.4160 | 0.8745 | 0.4163 | 0.8523 | 0.7200 | 0.9580 |
| | 0.4 | 0.8760 | 0.9995 | 0.8933 | 0.9993 | 0.9790 | 1.0000 |
| | 0.6 | 0.9515 | 1.0000 | 0.9680 | 0.9997 | 0.9980 | 1.0000 |
| | 0.8 | 0.9750 | 1.0000 | 0.9860 | 1.0000 | 1.0000 | 1.0000 |
| | 1.0 | 0.9765 | 1.0000 | 0.9907 | 1.0000 | 1.0000 | 1.0000 |
| | a | T_n^{MAVE*} | | T_n^{DEE*} | | | |
| | | $n = 50$ | $n = 100$ | $n = 50$ | $n = 100$ | | |
| $X \sim N(0, \Sigma_1), \epsilon \sim N(0, 1)$ | 0 | 0.0767 | 0.0640 | 0.0490 | 0.0490 | | |
| | 0.2 | 0.1580 | 0.2570 | 0.1075 | 0.1835 | | |
| | 0.4 | 0.3860 | 0.7190 | 0.3120 | 0.6800 | | |
| | 0.6 | 0.6170 | 0.9560 | 0.5655 | 0.9410 | | |
| | 0.8 | 0.7940 | 0.9920 | 0.7400 | 0.9885 | | |
| | 1.0 | 0.8660 | 0.9930 | 0.8475 | 0.9930 | | |
| $X \sim N(0, \Sigma_2), \epsilon \sim N(0, 1)$ | 0 | 0.0580 | 0.0435 | 0.0400 | 0.0475 | | |
| | 0.2 | 0.4170 | 0.8310 | 0.3925 | 0.7855 | | |
| | 0.4 | 0.7950 | 0.9880 | 0.7755 | 0.9870 | | |
| | 0.6 | 0.8770 | 0.9920 | 0.8985 | 0.9955 | | |
| | 0.8 | 0.8870 | 1.0000 | 0.9390 | 1.0000 | | |
| | 1.0 | 0.8880 | 1.0000 | 0.9445 | 1.0000 | | |

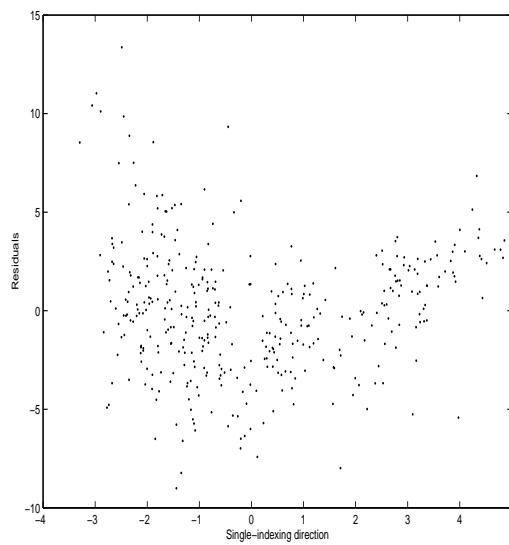


Fig. 3. Plot of the residuals from the linear regression model against the single-indexing direction obtained from DEE in the real data analysis.

Table 4. Empirical sizes and powers in *Study 3*, with $p = 2$. Here cases 1-4 represent situations with $X \sim N(0, \Sigma_1)$, $\epsilon \sim N(0, 1)$ (case 1) or $DE(0, \sqrt{2}/2)$ (case 2) and $X \sim N(0, \Sigma_2)$, $\epsilon \sim N(0, 1)$ (case 3) or $DE(0, \sqrt{2}/2)$ (case 4).

| | a | T_n^{ZH} | | T_n^{ZH*} | | T_n^{DEE} | | T_n^{DEE*} | |
|--------|-----|------------|-----------|-------------|-----------|-------------|-----------|--------------|-----------|
| | | $n = 50$ | $n = 100$ | $n = 50$ | $n = 100$ | $n = 50$ | $n = 100$ | $n = 50$ | $n = 100$ |
| Case 1 | 0 | 0.0470 | 0.0375 | 0.0500 | 0.0490 | 0.0483 | 0.0453 | 0.0463 | 0.0527 |
| | 0.2 | 0.0820 | 0.1390 | 0.0905 | 0.1335 | 0.1523 | 0.2653 | 0.1390 | 0.2683 |
| | 0.4 | 0.2430 | 0.5065 | 0.2560 | 0.4885 | 0.4127 | 0.7610 | 0.4053 | 0.7530 |
| | 0.6 | 0.5175 | 0.8635 | 0.4245 | 0.8320 | 0.6747 | 0.9650 | 0.6573 | 0.9557 |
| | 0.8 | 0.7335 | 0.9825 | 0.6350 | 0.9465 | 0.8453 | 0.9943 | 0.8237 | 0.9933 |
| | 1.0 | 0.8875 | 0.9990 | 0.7175 | 0.9750 | 0.9227 | 1.0000 | 0.9010 | 1.0000 |
| Case 2 | 0 | 0.0410 | 0.0375 | 0.0370 | 0.0585 | 0.0497 | 0.0487 | 0.0453 | 0.0530 |
| | 0.2 | 0.0915 | 0.1260 | 0.1005 | 0.1485 | 0.1653 | 0.2867 | 0.1540 | 0.2867 |
| | 0.4 | 0.2725 | 0.5300 | 0.2630 | 0.5035 | 0.4293 | 0.7540 | 0.4213 | 0.7683 |
| | 0.6 | 0.5460 | 0.8635 | 0.4840 | 0.8335 | 0.6757 | 0.9593 | 0.6643 | 0.9557 |
| | 0.8 | 0.7565 | 0.9790 | 0.6115 | 0.9520 | 0.8370 | 0.9927 | 0.8287 | 0.9947 |
| | 1.0 | 0.8745 | 0.9990 | 0.7220 | 0.9675 | 0.9183 | 0.9978 | 0.9000 | 0.9965 |
| Case 3 | 0 | 0.0370 | 0.0405 | 0.0435 | 0.0590 | 0.0473 | 0.0527 | 0.0597 | 0.0513 |
| | 0.2 | 0.0830 | 0.1130 | 0.0945 | 0.1760 | 0.1237 | 0.2107 | 0.1087 | 0.2130 |
| | 0.4 | 0.2630 | 0.5340 | 0.2705 | 0.5220 | 0.3667 | 0.6417 | 0.3320 | 0.6410 |
| | 0.6 | 0.5295 | 0.8955 | 0.4910 | 0.8505 | 0.6220 | 0.9363 | 0.5913 | 0.9250 |
| | 0.8 | 0.7915 | 0.9855 | 0.6385 | 0.9580 | 0.8170 | 0.9923 | 0.7797 | 0.9870 |
| | 1.0 | 0.9115 | 0.9995 | 0.7300 | 0.9775 | 0.9183 | 0.9990 | 0.8757 | 0.9987 |
| Case 4 | 0 | 0.0370 | 0.0410 | 0.0455 | 0.0420 | 0.0463 | 0.0513 | 0.0565 | 0.0565 |
| | 0.2 | 0.0890 | 0.1495 | 0.0995 | 0.1540 | 0.1367 | 0.2140 | 0.1305 | 0.2105 |
| | 0.4 | 0.2960 | 0.5490 | 0.2680 | 0.5285 | 0.3677 | 0.6623 | 0.3690 | 0.6695 |
| | 0.6 | 0.5750 | 0.8955 | 0.5060 | 0.8500 | 0.6450 | 0.9333 | 0.5960 | 0.9175 |
| | 0.8 | 0.7885 | 0.9895 | 0.6600 | 0.9510 | 0.8147 | 0.9897 | 0.7725 | 0.9825 |
| | 1.0 | 0.9005 | 0.9980 | 0.7380 | 0.9790 | 0.9110 | 0.9997 | 0.8855 | 0.9980 |

Table 5. Empirical sizes and powers in *Study 3*, with $p = 8$. Here cases 1-4 represent the situations with $X \sim N(0, \Sigma_1)$, $\epsilon \sim N(0, 1)$ (case 1) or $DE(0, \sqrt{2}/2)$ (case 2) and $X \sim N(0, \Sigma_2)$, $\epsilon \sim N(0, 1)$ (case 3) or $DE(0, \sqrt{2}/2)$ (case 4).

| | a | T_n^{ZH} | | T_n^{ZH*} | | T_n^{DEE} | | T_n^{DEE*} | |
|--------|-----|------------|-----------|-------------|-----------|-------------|-----------|--------------|-----------|
| | | $n = 50$ | $n = 100$ | $n = 50$ | $n = 100$ | $n = 50$ | $n = 100$ | $n = 50$ | $n = 100$ |
| Case 1 | 0 | 0.0182 | 0.0297 | 0.0450 | 0.0415 | 0.0605 | 0.0460 | 0.0465 | 0.0495 |
| | 0.2 | 0.0280 | 0.0475 | 0.0500 | 0.0705 | 0.1400 | 0.2645 | 0.1345 | 0.2610 |
| | 0.4 | 0.0442 | 0.0795 | 0.0785 | 0.0930 | 0.3485 | 0.6990 | 0.3555 | 0.7145 |
| | 0.6 | 0.0742 | 0.1573 | 0.1035 | 0.1895 | 0.5905 | 0.9420 | 0.5555 | 0.9280 |
| | 0.8 | 0.1022 | 0.2627 | 0.1330 | 0.2770 | 0.7510 | 0.9855 | 0.7275 | 0.9890 |
| | 1.0 | 0.1422 | 0.3715 | 0.1630 | 0.3500 | 0.8475 | 0.9960 | 0.8170 | 0.9935 |
| Case 2 | 0 | 0.0175 | 0.0265 | 0.0480 | 0.0430 | 0.0543 | 0.0517 | 0.0400 | 0.0540 |
| | 0.2 | 0.0283 | 0.0508 | 0.0655 | 0.0590 | 0.1463 | 0.2843 | 0.1395 | 0.2760 |
| | 0.4 | 0.0522 | 0.0953 | 0.0865 | 0.1240 | 0.3740 | 0.7240 | 0.3610 | 0.7365 |
| | 0.6 | 0.0885 | 0.1955 | 0.1295 | 0.2070 | 0.6073 | 0.9323 | 0.5990 | 0.9260 |
| | 0.8 | 0.1323 | 0.2953 | 0.1560 | 0.2925 | 0.7470 | 0.9873 | 0.7355 | 0.9860 |
| | 1.0 | 0.1675 | 0.4070 | 0.1880 | 0.3955 | 0.8510 | 0.9980 | 0.8170 | 0.9935 |
| Case 3 | 0 | 0.0213 | 0.0280 | 0.0480 | 0.0485 | 0.0463 | 0.0483 | 0.0450 | 0.0590 |
| | 0.2 | 0.0600 | 0.1335 | 0.0765 | 0.1660 | 0.3237 | 0.6443 | 0.2905 | 0.6595 |
| | 0.4 | 0.1935 | 0.4572 | 0.2245 | 0.4265 | 0.6970 | 0.9797 | 0.6780 | 0.9780 |
| | 0.6 | 0.3445 | 0.7280 | 0.3220 | 0.6570 | 0.8773 | 0.9993 | 0.8340 | 0.9980 |
| | 0.8 | 0.4758 | 0.8588 | 0.4290 | 0.7535 | 0.9237 | 0.9993 | 0.8985 | 0.9990 |
| | 1.0 | 0.5480 | 0.9205 | 0.4750 | 0.8020 | 0.9527 | 1.0000 | 0.9335 | 0.9990 |
| Case 4 | 0 | 0.0190 | 0.0275 | 0.0460 | 0.0540 | 0.0507 | 0.0533 | 0.0475 | 0.0500 |
| | 0.2 | 0.0742 | 0.1465 | 0.1095 | 0.1915 | 0.3280 | 0.6583 | 0.3140 | 0.6600 |
| | 0.4 | 0.2350 | 0.4950 | 0.2370 | 0.4420 | 0.6977 | 0.9783 | 0.6975 | 0.9760 |
| | 0.6 | 0.3757 | 0.7445 | 0.3440 | 0.6495 | 0.8660 | 0.9993 | 0.8355 | 0.9985 |
| | 0.8 | 0.4765 | 0.8585 | 0.4285 | 0.7425 | 0.9250 | 0.9997 | 0.9110 | 1.0000 |
| | 1.0 | 0.5537 | 0.9147 | 0.4865 | 0.8095 | 0.9550 | 1.0000 | 0.9355 | 1.0000 |

References

- Cook, R. D. (1998). *Regression Graphics: Ideas for Studying Regressions Through Graphics*. New York: Wiley.
- Cook, R. D. and Forzani, L. (2009). Likelihood-based sufficient dimension reduction. *J. Amer. Statist. Assoc.*, **104**, 197-208.
- Cook, R. D. and Li, B. (2002). Dimension reduction for conditional mean in regression. *Ann. of Statist.*, **30**, 455-474.
- Cook, R. D. and Weisberg, S. (1991). Discussion of "Sliced inverse regression for dimension reduction," by K. C. Li, *J. Amer. Statist. Assoc.*, **86**, 316-342.
- Delgado, M. A. and González-Manteiga, W. (2001). Significance testing in nonparametric regression based on the bootstrap. *Ann. of Statist.*, **29**, 1469-1507.
- Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimates. *Ann. of Statist.*, **27**, 1012-1050.
- Dette, H., Neumeyer, N. and Van Keilegom, I. (2007). A new test for the parametric form of the variance function in nonparametric regression. *J. Roy. Statist. Soc. Ser. B.*, **69**, 903-917.
- Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. of Statist.*, **29**, 153-193.
- Fan, Y. and Li, Q. (1996). Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica*, **64**, 865-890.
- Gaul, J. and Theissen, E. (2015). A partially linear approach to modelling the dynamics of spot and futures prices. *J. of Futu. Mark.*, **35**, 371-384.
- González-Manteiga, W. and Crujeiras, R. M. (2013). An updated review of Goodness-of-Fit tests for regression models. *Test*, **22**, 361-411.
- Härdle, W., Huet, S., Mammen, E. and Sperlich, S. (2004). Bootstrap inference in semiparametric generalized additive models. *Economet. Theo.*, **20**, 265-300.
- Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. of Statist.*, **21**, 1926-1947.
- Khmaladze, E. V. and Koul, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *Ann. of Statist.*, **37**, 995-1034.
- Koul, H. L. and Ni, P. P. (2004). Minimum distance regression model checking. *J. Stat. Plan. Infer.*, **119**, 109-141.

- Kumbhakar, S. C., Park, B. U., Simar, L. and Tsionas, E. G. (2007). Nonparametric stochastic frontiers: a local likelihood approach. *J. Econometrics*, **137**, 1-27.
- Lahaye, J. and Shaw, P. (2014). Can we reject linearity in an HAR-RV model for the S&P 500? Insights from a nonparametric HAR-RV. *Econom. Lett.*, **125**, 43-46.
- Li, B. and Wang, S. (2007). On directional regression for dimension reduction. *J. Amer. Statist. Assoc.*, **102**, 997-1008.
- Li, B., Zha, H. and Chiaromonte, F. (2005). Contour regression: A general approach to dimension reduction. *Ann. of Statist.*, **33**, 1580-1616.
- Li, K. C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.*, **86**, 86, 316-327.
- Li, Y. X. and Zhu, L. X. (2007). Asymptotics for sliced average variance estimation. *Ann. of Statist.*, **35**, 41-69.
- Lin, Z. J., Li, Q. and Sun, Y. G. (2014). A consistent nonparametric test of parametric regression functional form in fixed effects panel data models. *J. Econometrics*, **178(1)**, 167-179.
- Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley, New York.
- Mora, J. and Moro-Egido, A. I. (2008). On specification testing of ordered discrete choice models. *J. Econometrics*, **143**, 191 - 205.
- Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multivariate Anal.*, **101**, 1067-1078.
- Quinlan, R. (1993). Combining instance-based and model-based Learning. *In Proceedings on the Tenth International Conference of Machine Learning*, 236-243, University of Massachusetts, Amherst. Morgan Kaufmann.
- Simar, L., Van Keilegom, I. and Zelenyuk, V. (2014). Nonparametric least squares methods for stochastic frontier models. Working paper.
- Stute, W. (1997). Nonparametric model checks for regression. *Ann. of Statist.*, **25**, 613-641.
- Stute, W., González-Manteiga, W. and Presedo-Quindimil, M. (1998a). Bootstrap approximation in model checks for regression. *J. Amer. Statist. Assoc.*, **93**, 141-149.
- Stute, W., Thies, S. and Zhu, L. X. (1998b). Model checks for regression: An innovation process approach. *Ann. of Statist.*, **26**, 1916-1934.
- Stute, W. and Zhu, L. X. (2002). Model checks for generalized linear models. *Scand. J. Stat.*, **29**, 535-546.

- Van Keilegom, I., González-Manteiga, W. and Sánchez Sellero, C. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *Test*, **17**, 401-415.
- Wang, Q. and Yin, X. R. (2008). A nonlinear multi-dimensional variable selection method for high dimensional data: Sparse MAVE. *Comput. Stat. Data Anal.*, **52**, 4512-4520.
- Wu, C. F. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. of Statist.*, **14**, 1261-1295.
- Xia, Y. (2007). A constructive approach to the estimation of dimension reduction directions. *Ann. of Statist.*, **35**, 2654-2690.
- Xia, Y. (2009). Model check for multiple regressions via dimension reduction. *Biometrika*, **96**, 133-148.
- Xia, Y. C., Tong, H., Li, W. K. and Zhu, L. X. (2002). An adaptive estimation of dimension reduction space. *J. Roy. Statist. Soc. Ser. B.*, **64**, 363-410.
- Zhang, C. and Dette, H. (2004). A power comparison between nonparametric regression tests. *Stat. Probab. Lett.*, **66**, 289-301.
- Zhang, T. and Wu, W. B. (2011). Testing parametric assumptions of trends of a nonstationary time series. *Biometrika*, **98**, 599-614.
- Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics*, **75**, 263-289.
- Zhu, L. P., Wang, T., Zhu, L.X. and Ferré, 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. Am. Statist. Assoc.*, **105**, 1455-1466.
- Zhu, L. X. and Ng, K. (1995). Asymptotics for sliced inverse regression. *Statist. Sini.*, **5**, 727-736.