

A projection-based adaptive-to-model test for regressions *

Falong Tan^{1,2}, Xuehu Zhu³ and Lixing Zhu²

¹ School of Mathematics and Statistics, Shen Zhen University, Shen Zhen, China

² Department of Mathematics, Hong Kong Baptist University, Hong Kong

³ School of Mathematics, Xi'an Jiaotong University, Xi'an, China

Abstract

A longstanding problem of existing empirical process-based tests for regressions is that when the number of covariates is greater than one, they either have no tractable limiting null distributions or are not omnibus. To attack this problem, we in this paper propose a projection-based adaptive-to-model approach. When the hypothetical model is parametric single-index, the method can fully utilize the dimension reduction model structure under the null hypothesis as if the covariate were one-dimensional such that the martingale transformation-based test can be asymptotically distribution-free. Further, the test can automatically adapt to the underlying model structure such that the test can be omnibus and thus detect alternative models distinct from the hypothetical model at the fastest possible rate in hypothesis testing. The method is examined through simulation studied and is illustrated by a real data analysis.

Key words: Adaptive-to-model test, martingale transformation, model checking, projection pursuit.

*Lixing Zhu is a Chair professor of Department of Mathematics at Hong Kong Baptist University, Hong Kong, China. He was supported by a grant from the University Grants Council of Hong Kong, Hong Kong, China.

1 Introduction

Even when the dimension of covariates is moderate, dimensionality still causes data structure not to be visualized and thus makes regression modelling difficult. Therefore, in regression analysis, dimension reduction model structure is often used to approximate underlying models. A typical example is the parametric single-index regression model:

$$Y = g(\beta_0^\top X, \theta_0) + \varepsilon, \quad (1.1)$$

where Y is the response variable with the covariates $X \in \mathbb{R}^p$, $g(\cdot)$ is a known smooth function, $\beta_0 \in \mathbb{R}^p$ and $\theta_0 \in \mathbb{R}^d$ are the unknown regression parameter vectors, ε is the error term with $E(\varepsilon|X) = 0$ and the notation \top denotes transposition.

It is necessary to check the mis-specification of the regression function such that further regression analysis can be proceeded. Thus, the saturated alternative model is considered:

$$Y = G(X) + \varepsilon, \quad (1.2)$$

where $G(\cdot)$ denotes an unknown smooth function. There are several methods available to test the null hypothesis of model (1.1), which can be used for more general hypothetical parametric models. As this paper focuses on dimension-reduction issue, we only briefly mention existing locally and globally smoothing tests and then give a more detailed comment on existing methods that are used to handle the curse of dimensionality. Locally smoothing tests include Härdle and Mammen (1993), Zheng (1996), Fan and Li (1996), Dette (1999), Fan and Huang (2001), Koul and Ni (2004) and Van Keilegom et al.(2008). In low-dimensional cases, this type of tests can be sensitive to high-frequent alternative models. However, these tests rely on nonparametric regression estimation and thus suffer severely from the curse of dimensionality. This is because nonparametric regression estimation is very inefficient in high-dimensional scenarios. Guo et al. (2015) had detailed comments. Globally smoothing tests are nonparametric estimation free and particularly

sensitive to low frequency alternative models and have better asymptotic behaviours. This is because they are the averages over empirical processes. Examples include Stute (1997), Stute et. al. (1998a), Stute et. al. (1998b), Zhu (2003), Khmadladze and Koul (2004), Stute, Xu and Zhu (2008). For more references, see the review paper by González-Manteiga and Crujeiras (2013). However, when the dimension is greater than 1, they are usually not asymptotically distribution-free and thus require Monte Carlo approximations such as the wild bootstrap to determine critical values. Stute et al. (1998a) is a typical reference for this type of tests.

To attack this longstanding problem, there are several efforts in the literature to alleviate the curse of dimensionality. Guo, et al. (2015), as a first attempt in this field, suggested a model adaptive test that can avoid the dimensionality problem largely, but still requires nonparametric estimation. Thus, the test has slower convergence rate than $1/\sqrt{n}$ and theoretically, cannot detect the alternatives only distinct from the null at this fastest possible rate in hypothesis testing. A commonly used and efficient idea is to construct tests that are based on projected covariates in lower dimensional space. Most of existing methods are inspired by the projection pursuit technique that was first proposed by Friedman and Stuetzle (1981), since it is essential to find one or a few directions along which the departures from hypothetical models can be easily detected. Escanciano (2006) and Lavergne and Patilea (2008, 2012) proposed tests that are based on projected covariates. Two earlier and relevant references are Zhu and An (1992) and Zhu and Li (1998). Zhu (2003) and Stute, Xu and Zhu (2008) used residual processes to construct tests that can also be regarded as dimension reduction type. These tests usually need to resort to Monte Carlo approximations to determine critical values (e.g. Escanciano 2006; and Lavergne and Patilea 2008) though some of them are even asymptotically distribution-free such as Lavergne and Patilea (2012). This is either because of intractability of the null distribution or because of computational instability and complexity caused by the computation over all projected covariates at all directions. A relevant reference about the

computation issue is Wong et al. (1995). Xia (2009) also proposed a projection-based test that however has no way to control type I error.

More specifically, existing projection-based tests that involve residual-marked empirical processes are either the supremum or integral over all projected covariates $\{a^\top X : a \in \mathbb{R}^p \text{ with } \|a\| = 1\}$ to form Kolmogorov-Smirnov type or Cr amer-von Mises type statistics. It is worthwhile to note that test statistics naturally involve all projections under both the null and alternative hypothesis. Although it is reasonable and the omnibus property can also be guaranteed, the limiting null distributions are often intractable. In contrast, Stute and Zhu (2002) simply used one projection $\beta_0^\top X$ and thus the test behaves like the one with one-dimensional covariate. For model (1.1) letting $\epsilon = Y - g(\beta_0^\top X, \theta_0)$, we have that under the null hypothesis,

$$E(\epsilon|X) = 0 \Rightarrow E[Y - g(\beta_0^\top X, \theta_0)]I(\beta_0^\top X \leq u) = 0 \quad \text{for all } u \in \mathbb{R},$$

the residual marked empirical process defined by Stute and Zhu (2002) is

$$R_n(u) = n^{-1/2} \sum_{i=1}^n [Y_i - g(\beta_n^\top X_i, \theta_n)] I(\beta_n^\top X_i \leq u), \quad (1.3)$$

where $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ denote an *i.i.d.* sample from the distribution of (X, Y) , β_n and θ_n are, under the null hypothesis, root- n consistent estimators of β and θ , respectively. The martingale transformation can lead to an asymptotically distribution-free test (Stute et al. 1998a). However, *the test obviously fails to be omnibus* (see the comment in Escanciano 2006) because the construction only uses the model structure under the null hypothesis. Guo et al. (2015) gave an example to explicitly illustrate this phenomenon.

The purpose of this paper is to construct a globally smoothing test that inherits the asymptotically distribution-free and dimension reduction properties of Stute and Zhu's (2002) test under the null hypothesis and the omnibus property of general projection-based tests under the alternative hypothesis such as Escanciano (2006) and Lavergne and

Patilea (2008). To simultaneously achieve these two goals, we suggest an adaptive-to-model martingale transformation approach that can make the test automatically adapt to the underlying model structure under the respective null and alternative hypothesis.

To accommodate more general alternatives, we consider the following model:

$$Y = G(B^\top X) + \epsilon, \tag{1.4}$$

where G is an unknown smooth function, B is a $p \times q$ matrix with q orthogonal columns for an unknown q with $1 \leq q \leq p$ and $E(\epsilon|X) = 0$. When $q = 1$ and $B = \kappa\beta_0$ for some constant κ , model (1.4) becomes a semiparametric single-index model with the same index parameter as that in model (1.1). When $q = p$, model (1.4) reduces to model (1.2) since $G(\cdot)$ is unknown and $G(X) = G(BB^\top X) \equiv: \tilde{G}(B^\top X)$.

This paper is organised as follows. Basic test construction is described in Section 2. As sufficient dimension reduction technique is crucial to implement the adaptive-to-model strategy for test construction, we also give a short review in this section. In Section 3, we first present the asymptotic properties of the residual marked empirical process under the null hypothesis. The martingale transformation-based innovative process is then discussed. After that, we investigate the properties of the process and its innovative process under the alternative hypothesis. In Section 4, the test statistic is presented and simulation results for small to moderate sample size are reported and a real data analysis is used as an illustration of application. Appendix contains technical proofs of the theoretical results.

2 Projection-based adaptive-to-model empirical process

2.1 Basic construction

The null hypothesis now can be restated as

$$H_0 : E(Y|X) = g(\beta_0^\top X, \theta_0) \quad \text{for some } \beta_0 \in \mathbb{R}^p, \theta_0 \in \Theta \subset \mathbb{R}^d$$

and the alternative hypothesis is that for any $\beta \in \mathbb{R}^p$, $\theta \in \mathbb{R}^d$ and a $p \times q$ matrix B

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

where G is unknown. Without loss of generality, assume β_0 is a linear combination of the columns of B . Recall $\epsilon = Y - g(\beta_0^\top X, \theta_0)$. Under the null hypothesis, $q = 1$ and $B = \kappa\beta_0$ for some constant κ , then we have $E(\epsilon|\beta_0^\top X) = E(\epsilon|B^\top X) = 0$. Under the alternative hypothesis $E(\epsilon|B^\top X) = G(B^\top X) - g(\beta_0^\top X, \theta_0) \neq 0$. Thus, under the null hypothesis,

$$E[(Y - g(\beta_0^\top X, \theta_0))I(\beta_0^\top X \leq u)] = 0. \quad (2.1)$$

According to Lemma 1 of Escanciano (2006), Lemma 2.1 of Lavergne and Patilea (2008) or a similar result in Zhu and Li (1998) that can be traced back to Zhu and An (1992), we have that, under the alternative hypothesis, for an $\alpha \in \mathbb{R}^q$ whose first element is positive and $\|\alpha\| = 1$

$$E[(Y - g(\beta_0^\top X, \theta_0))I(\alpha^\top B^\top X \leq u)] \neq 0. \quad (2.2)$$

Note that under the null and alternative hypothesis, we use respective $\beta_0^\top X$ and $\alpha^\top B^\top X$. It is clear that we cannot define two estimates separately according to null and alternative hypothesis as we do not know the underlying model, while need an estimate \hat{B}_n of B that can adapt the underlying model: under the null \hat{B}_n converges to a vector proportional to β_0 and under the alternative, to B . If this can be achieved, we can use the empirical version of the left hand side of (2.2) to be the basis of a test statistic. Let $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ be

a sample with the same distribution as (X, Y) . Thereby we propose an adaptive-to-model residual marked empirical process for checking model (1.1) as follows:

$$V_n(u, \hat{\alpha}) = n^{-1/2} \sum_{i=1}^n [Y_i - g(\beta_n^\top X_i, \theta_n)] I(\hat{\alpha}^\top \hat{B}_n^\top X_i \leq u), \quad (2.3)$$

$$V_n(u) = \sup_{\|\hat{\alpha}\|=1, a_1 \geq 0} V_n(u, \hat{\alpha}) \quad (2.4)$$

where $\hat{\alpha}^\top = (a_1, \dots, a_{\hat{q}}) \in \mathbb{R}^{\hat{q}}$, \hat{B}_n is a sufficient dimension reduction estimator of B with an estimated structural dimension \hat{q} of q , β_n and θ_n are respectively ordinary least squares estimators of β_0 and θ .

It is clear that in order to have the model adaptation property of the process such that under the null hypothesis, $\sup_{\|\hat{\alpha}\|=1, a_1 \geq 0} V_n(u, \hat{\alpha})$ is equal to $n^{-1/2} \sum_{i=1}^n [Y_i - g(\beta_n^\top X_i, \theta_n)] I(\beta_n^\top X_i \leq u)$ Stute and Zhu (2002) defined, we must have that under the null hypothesis, \hat{q} and \hat{B} converge to 1 and $\kappa\beta_0$, respectively. Thus, we discuss their estimations.

2.2 A review on discretization-expectation estimation

To identify and estimate the number q and the matrix B , we use a method of sufficient dimension reduction (SDR). There are several proposals available in the literature. Examples include sliced inverse regression (SIR, Li 1991), sliced average variance estimation (SAVE, Cook and Weisberg 1991), minimum average variance estimation (MAVE, Xia et al. 2002), directional regression (DR, Li and Wang, 2007), likelihood acquired directions (LAD, Cook and Forzani, 2009) and average partial mean estimation (APME, Zhu et al. 2010b). In this section we briefly review discretization-expectation estimation (DEE, Zhu, et al. 2010b). Since G is unknown, for any $q \times q$ orthogonal matrix C , $G(B^\top X)$ can be rewritten as $\tilde{G}(C^\top B^\top X)$. This means B is not identifiable in model (1.4). Thus, SDR methodologies show that we can only identify q base vectors in the central mean subspace $\mathcal{S}_{E(Y|X)}$ spanned by B (see Cook (1998)), or precisely BC for a $q \times q$ orthonormal matrix

C. This can be achieved through identifying $\mathcal{S}_{E(Y|X)}$. In the SDR theory, $\mathcal{S}_{E(Y|X)}$ is defined to be the intersection of all subspaces $\text{span}(A)$ such that $Y \perp\!\!\!\perp E(Y|X) | A^\top X$ where $\perp\!\!\!\perp$ means statistical independence and $\text{span}(A)$ means the subspace spanned by the columns of A . The dimension of $\mathcal{S}_{E(Y|X)}$ is called the structural dimension, denoted as $d_{E(Y|X)}$. Therefore, under the null hypothesis (1.1), $\mathcal{S}_{E(Y|X)} = \text{span}(\beta_0 / \|\beta_0\|)$ and $d_{E(Y|X)} = 1$; while under the alternative (1.4), $\mathcal{S}_{E(Y|X)} = \text{span}(B)$ and $d_{E(Y|X)} = q$. Similarly, the central subspace (Cook (1998)) denoted by $\mathcal{S}_{Y|X}$ is defined to be the intersection of all subspaces $\text{span}(A)$ such that $Y \perp\!\!\!\perp X | A^\top X$. It is easy to see that $\mathcal{S}_{E(Y|X)} \subset \mathcal{S}_{Y|X}$. For simplicity, we assume $\mathcal{S}_{E(Y|X)} = \mathcal{S}_{Y|X}$. A special case is that $\eta \perp\!\!\!\perp X$ in model (1.4).

The basic procedure of DEE is given below.

1. Define the discrete response variable $Z(t) = I\{Y \leq t\}$ where the indicator function $I\{Y \leq t\}$ take value 1 if $Y \leq t$ and 0 otherwise.
2. Let $\mathcal{S}_{Z(t)|X}$ denote the central subspace of $Z(t)|X$ and $M(t)$ be a $p \times p$ positive semi-definite matrix such that $\text{Span}\{M(t)\} = \mathcal{S}_{Z(t)|X}$.
3. Let \tilde{Y} be an independent copy of Y and $M = E\{M(\tilde{Y})\}$. Theorem 1 in Zhu et al. (2010) asserts that $\text{Span}(M) = \mathcal{S}_{Y|X}$ and B consists of the eigenvectors corresponding to the nonzero eigenvalues of M .
4. The estimator of the target matrix M is given by

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

where $M_n(y_i)$ is the estimator of the matrix $M(y_i)$ obtained by a chosen sufficient dimension reduction method such as SIR (Li 1991). Then an estimator $B_n(q)$ of B consists of the eigenvectors associated with the largest q eigenvalues of M_n when q is given.

According to Theorems 2 and 3 in Zhu et al. (2010), $B_n(q)$ could achieve root- n consistence to B . For more details the readers can refer to Zhu et al. (2010).

2.3 Structural dimension estimation

As argued in Section 2.2, a consistency estimator of the structure dimension q is required. Zhu et al.(2010a) suggested the BIC-type criterion to determine q , which is a modification of the method in Zhu et al. (2006). Here, we suggest a minimum ridge-type eigenvalue ratio estimate (MRER) to determine the structure dimension q . Let $\hat{\lambda}_p \leq \dots \leq \hat{\lambda}_1$ denote the eigenvalues of the estimated matrix M_n of M . Note that $q = \dim \mathcal{S}_{Y|X} = \text{rank}(M)$. Then the true structure dimension q can be estimated by

$$\hat{q} = \arg \min_{1 \leq i \leq p} \left\{ i : \frac{\hat{\lambda}_{i+1}^2 + c}{\hat{\lambda}_i^2 + c} \right\}. \quad (2.5)$$

The algorithm of MRER is easy to implement. The following theorem shows that the consistency of MRER is adaptive to the underlying model.

Lemma 1 *Under the regularity conditions assumed by Zhu et al. (2010a), the estimator \hat{q} of (2.5) with $c = \log n/n$ satisfies that, as $n \rightarrow \infty$,*

(i) *under H_0 , $Pr(\hat{q} = 1) \rightarrow 1$,*

(ii) *under H_1 , $Pr(\hat{q} = q) \rightarrow 1$.*

A justification of this lemma can be found in the Appendix.

3 Main results

3.1 Basic properties of the process

First, we consider the following process

$$V_n^0(u, \alpha) = n^{-1/2} \sum_{i=1}^n [Y_i - g(\beta_0^\top X_i, \theta_0)] I(\alpha^\top B^\top X_i \leq u),$$

where $\alpha = (a_1, a_2, \dots, a_q)^\top$, $a_1 \geq 0$ and $\|\alpha\| = 1$. Denote $\sigma^2(v, \alpha) = \text{Var}(Y | \alpha^\top B^\top X = v)$ and $\psi(u, \alpha) = E[\text{Var}(Y | \alpha^\top B^\top X) I(\alpha^\top B^\top X \leq u)]$. Under the null hypothesis, $q = 1$, $\alpha = 1$ and $B = \kappa\beta_0$. Thus, we rewrite it as

$$\begin{aligned} \sigma^2(v) &\equiv: \sigma^2(v, 1) = \text{Var}(Y | \kappa\beta_0^\top X = v), \\ \psi(u) &\equiv: \psi(u, 1) = E[\text{Var}(Y | \kappa\beta_0^\top X) I(\kappa\beta_0^\top X \leq u)], \\ V_n^0(u) &\equiv: V_n^0(u, 1) = n^{-1/2} \sum_{i=1}^n [Y_i - g(\beta_0^\top X_i, \theta_0)] I(\kappa\beta_0^\top X_i \leq u). \end{aligned}$$

Obviously, $\psi(u) = \int_{-\infty}^u \sigma^2(v) dF_{\kappa\beta_0}(dv)$ where $F_{\kappa\beta_0}$ denotes the distribution of $\kappa\beta_0^\top X$.

Under the null hypothesis, Theorem 1.1 in Stute (1997) implies that

$$V_n^0(u) \longrightarrow V_\infty(u) \quad \text{in distribution} \tag{3.1}$$

in the Skorohod space $D[-\infty, \infty)$ where V_∞ is a continuous Gaussian process with mean zero and covariance kernel $K(u_1, u_2) = \psi(u_1 \wedge u_2)$.

To study the process $V_n(u, \hat{\alpha})$ proposed in (2.3), we give some regularity conditions on the function $g(\beta^\top X, \theta)$ and the parameters in the following.

A1 Under H_0 , we suppose that (β_n, θ_n) has a linear expansion

$$\sqrt{n} \begin{pmatrix} \beta_n - \beta_0 \\ \theta_n - \theta_0 \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^n l(x_i, y_i, \beta_0, \theta_0) + o_p(1),$$

where l is a vector-valued function satisfying that

I $E(l(X, Y, \beta_0, \theta_0)) = 0;$

II $L(\beta_0, \theta_0) = E(l(X, Y, \beta_0, \theta_0)l^\top(X, Y, \beta_0, \theta_0))$ is positive definite.

It is easy to see that the ordinary least squares estimator satisfies condition (A1).

A2 The function $g(\beta^\top x, \theta)$ is continuously differentiable with respect to (β, θ) in some neighbourhood of (β_0, θ_0) . The first-order partial derivatives

$$m(x, \beta, \theta) = \frac{\partial g(\beta^\top x, \theta)}{\partial(\beta, \theta)} = (m_1(x, \beta, \theta), \dots, m_{p+d}(x, \beta, \theta))^\top$$

satisfies that there exists a μ -integrable function $K_0(x)$ such that

$$|m_j(x, \beta, \theta)| \leq K_0(x) \quad \text{for all } (\beta, \theta) \quad \text{and} \quad 1 \leq j \leq p + d.$$

where μ denotes the distribution of X .

A3 Let $H(u, \beta) = E[\text{Var}(Y|X)I(\kappa\beta^\top X \leq u)]$ and we assume that $H(u, \beta)$ is uniformly continuous in u at β_0 .

Theorem 3.1. *Under H_0 and Conditions A1-A2, we have in distribution*

$$V_n(u) \longrightarrow V_\infty(u) - M(u)^\top V \equiv: V_\infty^1(u)$$

where $V_\infty(u)$ has the same Gaussian process as given by (3.1), the vector-valued function $M^\top = (M_1, M_2, \dots, M_{p+d})$ is defined as

$$M_i(u) = E(m_i(X, \beta_0, \theta_0)I(\kappa\beta_0^\top X \leq u))$$

and V is a $p + d$ -dimensional normal vector with zero means and covariance matrix $L(\beta_0, \theta_0)$.

Theorem 3.1 is almost the same as Theorem 1 in Stute and Zhu(2002) except for the definition of M . In other words, under the null hypothesis, our process can have almost the same limiting property as that in Stute and Zhu(2002). Thus, a martingale transformation

can be implemented. However, we also need to check whether it can be done under the alternative hypothesis. We then discuss the adaptive-to-model martingale transformation below.

3.2 Adaptive-to-model martingale transformation

To have the model adaptation property of the process, our idea is that under the null hypothesis, the process is only with $\beta_0^\top X$ (or proportional to it as $\kappa\beta_0^\top X$), and under the alternatives, the process is automatically with $\alpha^\top B^\top X$ such that the transformation is still implementable and the transformed process can capture the information of alternative models.

First, motivate our method from the transformation under the null hypothesis. Recall M a vector-valued function on \mathbb{R} and

$$\psi(u) = \int_{-\infty}^u \sigma^2(v) dF_{\kappa\beta_0}(dv)$$

a nonnegative increasing function with $\psi(-\infty) = 0$. Let $a = \frac{\partial M}{\partial \psi}$ stand for the Radon-Nikodym derivative of G w.r.t. ψ , assuming that it exists. Let

$$A(u) = \int_u^\infty a(v)a^\top(v)\psi(dv) = \int_u^\infty a(v)a^\top(v)\sigma^2(v)dF_{\kappa\beta_0}(dv)$$

be a $(d+p) \times (d+p)$ matrix. Finally, we define the innovation process transformation as

$$(Tf)(z) = f(z) - \int_{-\infty}^z a^\top(u)A^{-1}(u) \left[\int_u^\infty a(v)f(dv) \right] \psi(du). \quad (3.2)$$

Here we suppose that $A(u)$ is non-singular and the process f should be either bounded variation or Brownian motion.

Using the same arguments in the proofs of lemma 3.1 and 3.2 in Nikabadzé and S-tute(1997), we have the following two vital facts:

(i) $T(G^\top V) \equiv 0$,

(ii) $TV_\infty = V_\infty$ in distribution.

Thus TV_∞ is a centered Gaussian process with a covariance kernel $K(u_1, u_2) = \psi(u_1 \wedge u_2)$ which can be considered as the martingale part in the Doob-Meyer decomposition of V_∞^1 . See Stute et al. (1998b).

Note that T relies on some unknown quantities and thus needs to be replaced by its empirical version. To this end, let $g_1(t, \theta) = \frac{\partial g(t, \theta)}{\partial t}$ and $g_2(t, \theta) = \frac{\partial g(t, \theta)}{\partial \theta}$, then

$$m(x, \beta_0, \theta_0) = (g_1(\beta_0^\top X, \theta_0)X^\top, g_2(\beta_0^\top X, \theta_0))^\top.$$

Recall that $M(u) = E(m(X, \beta_0, \theta_0)I(\kappa\beta_0^\top X \leq u))$, then we obtain

$$M(u) = \begin{pmatrix} E[g_1(\beta_0^\top X, \theta_0)X I(\kappa\beta_0^\top X \leq u)] \\ E[g_2(\beta_0^\top X, \theta_0)^\top I(\kappa\beta_0^\top X \leq u)] \end{pmatrix} = \begin{pmatrix} \int_{-\infty}^u g_1(v/\kappa, \theta_0)r(v)F_{\kappa\beta_0}(dv) \\ \int_{-\infty}^u g_2(v/\kappa, \theta_0)^\top F_{\kappa\beta_0}(dv) \end{pmatrix},$$

where $r(v) = E(X|\kappa\beta_0^\top X = v)$. It is easy to see that

$$a = \frac{\partial M}{\partial \psi} = \begin{pmatrix} g_1(v/\kappa, \theta_0)r(v)/\sigma^2(v) \\ g_2(v/\kappa, \theta_0)^\top/\sigma^2(v) \end{pmatrix},$$

and

$$A(u) = \int_u^{-\infty} a(v)M^\top(dv) = \begin{pmatrix} \int_u^{-\infty} g_1(v/\kappa, \theta_0)r(v)/\sigma^2(v)M^\top(dv) \\ \int_u^{-\infty} g_2(v/\kappa, \theta_0)^\top/\sigma^2(v)M^\top(dv) \end{pmatrix}.$$

In a general nonparametric framework, there are no assumptions on r and σ except for smoothness, thus both the functions need to be estimated by some curve estimators. Here we adopt a standard Nadaraya-Watson estimator for r

$$r_n(v) = \frac{\sum_{i=1}^n X_i K(\frac{v - \hat{\alpha}^\top \hat{B}_n^\top X_i}{h})}{\sum_{i=1}^n K(\frac{v - \hat{\alpha}^\top \hat{B}_n^\top X_i}{h})}.$$

For σ^2 , note that $\sigma^2(u) = E(\epsilon^2|\kappa\beta_0 X = u)$ which can be replaced by

$$\sigma_n^2(u) = \frac{\sum_{i=1}^n \hat{\epsilon}_i^2 K(\frac{u - \hat{\alpha}^\top \hat{B}_n^\top X_i}{h})}{\sum_{i=1}^n K(\frac{u - \hat{\alpha}^\top \hat{B}_n^\top X_i}{h})},$$

where $K(\cdot)$ is a univariate kernel function and h is a bandwidth. It is worth mentioning that we use $\hat{\alpha}^\top \hat{B}_n^\top X$ rather than $\beta_n^\top X$ where β_n is the nonlinear least squares estimate of β_0 that was used by Stute and Zhu (2002). As \hat{q} and \hat{B} have the model adaptation property, in the following, we can derive the model adaptation property of the transformed process.

Now we can respectively obtain the empirical versions a_n, M_n and A_n of a, M and A :

$$\begin{aligned} a_n(v) &= \begin{pmatrix} g_1(v/\kappa_n, \theta_n) r_n(v) / \sigma_n^2(v) \\ g_2(v/\kappa_n, \theta_n)^\top / \sigma_n^2(v) \end{pmatrix}, \\ M_n(u) &= \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n g_1(\beta_n^\top X_i, \theta_n) X_i I(\hat{\alpha}^\top \hat{B}_n^\top X_i \leq u) \\ \frac{1}{n} \sum_{i=1}^n g_2(\beta_n^\top X_i, \theta_n)^\top I(\hat{\alpha}^\top \hat{B}_n^\top X_i \leq u) \end{pmatrix}, \\ A_n(u) &= \frac{1}{n} \sum_{i=1}^n I(\hat{\alpha}^\top \hat{B}_n^\top X_i \geq u) \begin{pmatrix} g_1(\hat{\alpha}^\top \hat{B}_n^\top X_i / \kappa_n, \theta_n) r_n(\hat{\alpha}^\top \hat{B}_n^\top X_i) / \sigma_n^2(\hat{\alpha}^\top \hat{B}_n^\top X_i) \\ g_2(\hat{\alpha}^\top \hat{B}_n^\top X_i / \kappa_n, \theta_n)^\top / \sigma_n^2(\hat{\alpha}^\top \hat{B}_n^\top X_i) \end{pmatrix} \\ &\quad \times (g_1(\beta_n^\top X_i, \theta_n) X_i^\top, g_2(\beta_n^\top X_i, \theta_n)). \end{aligned}$$

Replace a, M and A in (3.2) by their empirical versions, we obtain the empirical version of TV :

$$\begin{aligned} (T_n V_n)(u, \hat{\alpha}) &= V_n(u, \hat{\alpha}) - \int_{-\infty}^u a_n(v)^\top A_n^{-1}(v) \int_v^\infty a_n(z) V_n(dz) \sigma_n^2(v) F_{\hat{\alpha}}(dv) \\ &= \frac{1}{n^{1/2}} \sum_{i=1}^n [Y_i - g(\beta_n^\top X_i, \theta_n)] I(\hat{\alpha}^\top \hat{B}_n^\top X_i \leq u) - \frac{1}{n^{3/2}} \sum_{i,j=1}^n I(\hat{\alpha}^\top \hat{B}_n^\top X_i \leq u) \\ &\quad \times [g_1(\hat{\alpha}^\top \hat{B}_n^\top X_i / \kappa_n, \theta_n) r_n(\hat{\alpha}^\top \hat{B}_n^\top X_i)^\top, g_2(\hat{\alpha}^\top \hat{B}_n^\top X_i / \kappa_n, \theta_n)] \\ &\quad \times A_n^{-1}(\hat{\alpha}^\top \hat{B}_n^\top X_i) I(\hat{\alpha}^\top \hat{B}_n^\top X_j \geq \hat{\alpha}^\top \hat{B}_n^\top X_i) (Y_j - g(\beta_n^\top X_j, \theta_n)) \\ &\quad \times \begin{pmatrix} g_1(\hat{\alpha}^\top \hat{B}_n^\top X_i / \kappa_n, \theta_n) r_n(\hat{\alpha}^\top \hat{B}_n^\top X_i) / \sigma_n^2(\hat{\alpha}^\top \hat{B}_n^\top X_i) \\ g_2(\hat{\alpha}^\top \hat{B}_n^\top X_i / \kappa_n, \theta_n)^\top / \sigma_n^2(\hat{\alpha}^\top \hat{B}_n^\top X_i) \end{pmatrix} \end{aligned}$$

where κ_n is the estimator of κ and $F_{\hat{\alpha}}$ is the empirical distribution function of $\hat{\alpha}^\top \hat{B}_n^\top X_i, 1 \leq i \leq n$.

Theorems 3.2 and 3.3 below show that the resulting transformed process is adaptive to the underlying model.

Theorem 3.2. *Let $\sigma_n^2(u)$ be a consistent estimator of σ^2 which is bounded away from zero. Under the regularity conditions of Theorem (3.1) and H_0 , we have in distribution*

$$\sup_{\|\hat{\alpha}\|=1, a_1 \geq 0} T_n V_n(u, \hat{\alpha}) \longrightarrow V_\infty(u) \quad \text{in distribution in the space } D[-\infty, \infty).$$

The result meets our expectation. Because of model adaptation, the supremum under the null is actually over only one direction and the transformed process is a standard Gaussian process as proved by Stute and Zhu (2002).

When the distribution of X is elliptically contoured, particularly spherically contoured such as normal distributions, the formulation of the transformation can be much simpler and thus the computation can be much easier. Without loss of generality, consider spherically contoured distributions. Suppose the regression function g does not rely on the parameter θ . Let $g'(\beta_0^\top x)$ be the derivative of $g(\cdot)$ about $\beta_0^\top x$. Thus we have $m(x, \beta_0) = g'(\beta_0^\top x)x$ and

$$M(u) = E[g'(\beta_0^\top X)XI(\kappa\beta_0^\top X \leq u)] = \int_{-\infty}^u g'(v/\kappa)r(v)dF_{\kappa\beta_0}(dv).$$

Therefore $a = \frac{\partial M}{\partial \psi} = \frac{g'(v/\kappa)r(v)}{\sigma^2(v)}$ and $A(u) = \int_u^{-\infty} a(v)M^\top(dv) = \int_u^{-\infty} \frac{g'(v/\kappa)r(v)}{\sigma^2(v)}M^\top(dv)$.

Let Γ be any orthogonal matrix with the first row $\beta_0^\top / \|\beta_0\|$, thus the first component of ΓX is $\beta_0^\top X / \|\beta_0\|$. Since the conditional distribution of the other components of ΓX given the first is still spherical, these conditional expectations are zero. Therefore,

$$\begin{aligned} M(u) &= \Gamma^\top E g'(\beta_0^\top X) \Gamma X I(\kappa\beta_0^\top X \leq u) \\ &= \frac{\beta_0}{\|\beta_0\|^2} E g'(\beta_0^\top X) \beta_0^\top X I(\kappa\beta_0^\top X \leq u) \\ &= \frac{\beta_0}{\|\beta_0\|^2} \int_{-\infty}^u [g'(v/\kappa)v/\kappa] F_{\kappa\beta_0}(dv). \end{aligned}$$

Thus we obtain

$$\begin{aligned} a(v) &= \frac{\beta_0}{\|\beta_0\|^2} \frac{g'(v/\kappa)v/\kappa}{\sigma^2(v)}, \\ A(u) &= \frac{\beta_0\beta_0^\top}{\|\beta_0\|^4} \int_u^{-\infty} \frac{[g'(v/\kappa)v/\kappa]^2}{\sigma^2(v)} F_{\kappa\beta_0}(dv). \end{aligned}$$

The matrix $A(x)$ is singular with rank 1. To derive relevant asymptotic results as those in Theorem 3.2, denote

$$h(v) = Eg'(\beta_0^\top X)\beta_0^\top XI(\kappa\beta_0^\top X \leq v) \quad \text{and} \quad V_1 = \frac{\beta_0^\top V}{\|\beta_0\|^2}.$$

Therefore $M^\top(u)V = h(u)V_1$ and the conclusion of Theorem 3.1 would be rewritten as

$$V_n \longrightarrow V_\infty - hV_1 \quad \text{in distribution}$$

The new a and A now become real-valued:

$$a = \frac{\partial h}{\partial \psi} \quad \text{and} \quad A(u) = \int_u^\infty a^2(v)\sigma^2(v)F_{\kappa\beta_0}(dv).$$

Similarly, we can also obtain the empirical analogues A_n and T_n of A and T respectively. Therefore Theorem 3.2 can be applied with these new functions in the results.

It is noteworthy to point out that convergence in $D[-\infty, \infty)$ means convergence in $D[-\infty, u]$ for any finite u . Since the transformation involves the inverses of $A(u)$, Our test statistic would yield instabilities in the distributional behaviors for large values of u . Thus all the processes should be constrained in proper subsets of the real line. In practice, we would consider $T_n V_n$ on a given quantile of $\kappa_n \beta_n^\top X_i$, $1 \leq i \leq n$. See Stute and Zhu (2002). Now we give the result to show that the transformed process can automatically adapt to the alternative models such that a constructed test can detect them.

3.3 The properties under the alternative hypothesis

Consider the following sequence of alternatives converging to the null hypothesis

$$H_{1n} : Y_n = g(\beta_0^\top X, \theta_0) + C_n G(B^\top X) + \eta \tag{3.3}$$

where $E(\eta|X) = 0$, and β_0 is the linear combination of the columns of B . When C_n is a fixed constant, the model is under global alternative equivalent to model (1.4). If it

tends to zero, the models are under local alternatives. We give the asymptotic property of the estimator \hat{q} of the structure dimension q under the local alternatives. The following lemma shows that when C_n goes to zero quickly, \hat{q} is not equal to q , while to 1. In other words, \hat{q} is an inconsistent estimate of q .

Lemma 2 *Under the local alternative H_{1n} and the same conditions in Lemma 1 with $C_n = 1/\sqrt{n}$, the estimator \hat{q} of (2.5) with $c = \log n/n$ satisfies that $P(\hat{q} = 1) \rightarrow 1$, as $n \rightarrow \infty$.*

To derive the asymptotic properties of $\sup_{\|\hat{\alpha}\|=1, a_1 \geq 0} T_n V_n(u, \hat{\alpha})$, under the local alternative H_{1n} , we need an additional condition:

A3

$$\sqrt{n} \begin{pmatrix} \beta_n - \beta_0 \\ \theta_n - \theta_0 \end{pmatrix} = \gamma + \frac{1}{\sqrt{n}} \sum_{i=1}^n l(x_i, y_i, \beta_0, \theta_0) + o_p(1),$$

where γ is some constant vector and the vector-valued function l is the same as given in (A1). Under H_{1n} with $C_n = \frac{1}{\sqrt{n}}$, condition (A3) is satisfied for the nonlinear least squares estimate. See Lemma 3 in Guo et al.(2015).

Now we derive the limits of $\sup_{\|\hat{\alpha}\|=1, a_1 \geq 0} T_n V_n(x, \hat{\alpha})$ under the global alternative and local alternatives.

Theorem 3.3. *Under the regularity conditions A1-A3, we have*

(i) *under the global alternative H_1 that is equivalent to H_{1n} with fixed C_n ,*

$$\frac{1}{\sqrt{n}} \sup_{\|\hat{\alpha}\|=1, a_1 \geq 0} T_n V_n(u, \hat{\alpha}) \rightarrow L(u);$$

where $L(u)$ is some nonzero process. That is, $\sup_{\|\hat{\alpha}\|=1, a_1 \geq 0} T_n V_n(u, \hat{\alpha})$ diverges to infinity at the rate of $1/\sqrt{n}$.

(ii) *under the local alternatives H_{1n} with $C_n = 1/\sqrt{n}$,*

$$\sup_{\|\hat{\alpha}\|=1, a_1 \geq 0} T_n V_n(u, \hat{\alpha}) \rightarrow V_\infty(u) + \int_{-\infty}^u \{H_0(v) - a(v)^\top A^{-1}(v)[W_1(v) + W_2(v)]\sigma^2(v)\} F_{\kappa, \beta_0}(dv)$$

where $H_0(v) = E[G(B^\top X)|\kappa\beta_0^\top X = v]$, $W_1(v) = \int_v^\infty a(z)H_0(z)F_{\kappa\beta_0}(dz)$ and $W_2(v)$ have a normal distribution with mean zero and variance

$$\sigma_2^2(v) = \int_v^\infty \sigma^2(u)a(u)a(u)^\top F_{\kappa\beta_0}(du).$$

Based on Theorems 3.2 and 3.3, we can derive the asymptotic properties of functionals of $\sup_{\|\hat{\alpha}\|=1, a_1 \geq 0} T_n V_n(u, \hat{\alpha})$ over all u . The resulting test statistic is defined in the next section.

4 Numerical studies

4.1 Test statistics in practical use

The test statistic is a functional of $T_n V_n$. In this paper, we consider the Cramer–von Mises statistic of the form

$$CW_n^2 = \int_{-\infty}^{x_0} \sup_{\|\hat{\alpha}\|=1, a_1 \geq 0} (T_n V_n(u, \hat{\alpha}))^2 F_n(du), \quad (4.1)$$

where F_n is the empirical distribution function of $\frac{\beta_n}{\|\beta_n\|} X_i$, $1 \leq i \leq n$. Note that we take $\kappa_n = 1/\|\beta_n\|$ here. Using Theorem 3.2 and the continuous mapping theorem we obtain that, under the null hypothesis

$$CW_n^2 \longrightarrow \int_{-\infty}^{x_0} B^2(\psi(u)) F_{\kappa\beta_0}(du) \quad \text{in disitribution.}$$

where $B(u)$ is a standard Brownian motion. To obtain a distribution-free limit for our test, note that

$$\frac{1}{\psi(x_0)^2} \int_{-\infty}^{x_0} B^2(\psi(u)) \sigma^2(u) F_{\kappa\beta_0}(du) = \int_0^1 B(u)^2 du \quad \text{in disitribution,}$$

thus we consider

$$W_n^2 = \frac{1}{\psi_n(x_0)^2} \int_{-\infty}^{x_0} \sup_{\|\hat{\alpha}\|=1, a_1 \geq 0} (T_n V_n(u, \hat{\alpha}))^2 \sigma_n^2 F_n(du),$$

where $\psi_n(u) = \frac{1}{n} \sum_{i=1}^n (Y_i - g(\beta_n^\top X_i))^2 I(\frac{\beta_n}{\|\beta_n\|} X_i \leq u)$ is the estimator of $\psi(u)$ and σ_n^2 can be any consistent estimator of the conditional variance σ^2 defined in subsection 3.1.

Therefore we have

$$W_n^2 \longrightarrow \int_0^1 B^2(u) du \quad \text{in disitribution.}$$

If the regression model is homoscedastic, then σ^2 is a constant and we can estimate it by

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n [Y_i - g(\beta_n^\top X_i, \theta_n)]^2.$$

Under the null hypothesis $\psi(x_0) = \sigma^2 F_{\kappa\beta}(x_0)$ and it can be estimated by $\sigma_n^2 F_n(x_0)$, thus W_n^2 becomes

$$W_n^2 = \frac{1}{\sigma_n^2 F_n^2(x_0)} \int_{-\infty}^{x_0} \sup_{\|\hat{\alpha}\|=1, a_1 \geq 0} (T_n V_n(u, \hat{\alpha}))^2 F_n(du).$$

For ease of comparison, we give four examples in the following. For x_0 , as Stute and Zhu (2002) did, we choose the 99% quantile of F_n in the numerical examples.

4.2 Numerical examples

In this subsection we conduct some simulations to show the performance of the distributional approximations for small to moderate sample size. We make a comparison with Guo et al.(2015)'s dimension reduction model-adaptive test T_n^{GWZ} that is based on SIR-based DEE, Stute et al.(1998a)'s test T_n^{SGP} that determines critical values by the wild bootstrap. We design four representative examples. The first is to confirm that the proposed test that can be regarded as an extension of Stue and Zhu's (2002) test is omnibus. The second includes both high-frequency and low-frequency model such as we can compare with an adaptive-to-model test that is based on locally smoothing approach.

The third includes models with higher structural dimension and the fourth is also used to check the influence of dimensionality. The significant level is set to be $\alpha = 0.05$, and the reported results are the average of 2000 replications. In all models, the value of $a = 0$ corresponds to the null hypothesis and $a \neq 0$ to the alternatives.

Example 1. The data are generated from the model

$$Y = \frac{1}{4} \exp(2\beta_1^\top X) + a\beta_2^\top X + \varepsilon;$$

Here we consider two cases: $p = 3, \beta_1 = (1, 0, 0)^\top, \beta_2 = (0, 1, 0)^\top$ and $p = 4, \beta_1 = (1, 1, 0, 0)^\top/\sqrt{2}, \beta_2 = (0, 0, 1, 1)^\top/\sqrt{2}$. In both cases, $n = 50, 100$, X follows the standard multivariate normal distribution $N(0, I_p)$ and ε is from $N(0, 1)$. Note that under the alternatives, we have $E(Y - \exp(\frac{5}{4}\beta_1^\top X)|\beta_1^\top X) = 0$. The results in Figure 1 obviously show that T_n^{SZ} fails to work while W_n^2 performs very well.

Figure 1 about here

Now we consider the comparison between local and global smoothing tests and between the adaptive-to-model method and the classical method.

Example 2. Consider

$$\begin{aligned} H_{11} : Y &= \beta_0^\top X + a \cos\left(\frac{\pi}{2}\beta_0^\top X\right) + \varepsilon; \\ H_{12} : Y &= \beta_0^\top X + \frac{1}{4}a \exp(\beta_0^\top X) + \varepsilon; \\ H_{13} : Y &= \beta_0^\top X + \frac{1}{2}a(\beta_0^\top X)^2 + \varepsilon. \end{aligned}$$

where $p = 8, \beta_0 = (1, 1, \dots, 1)/\sqrt{p}$, $X = (X_1, X_2, \dots, X_p)^\top$ which is independent with ε . The central mean subspaces $\mathcal{S}_{E(Y|X)}$ has the structural dimension 1 which means $B = \beta_0$ under both the null and alternative hypothesis. The predictors $x_i, i = 1, \dots, n$ are i.i.d. from two distributions: multivariate normal distributions $N(0, I_p)$ and $N(0, \Sigma)$ with

$\Sigma = (1/2^{|i-j|})_{p \times p}$ so as to check the influence of correlation between the covariates. The errors ε_i 's are drawn independently from $N(0, 1)$. The first is a high-frequency model and the others are low-frequent.

The empirical sizes and powers of the three tests are presented in Tables 1 and 2. We can see that both T_n^{GWZ} and W_n^2 control the size very well even for the small size of $n = 50$. T_n^{SGP} seems slightly more conservative with higher empirical size than 0.05. For the first model, T_n^{GWZ} has relatively higher power than W_n^2 and T_n^{SGP} have, especially in the correlated case with covariance matrix Σ . For the models H_{12} and H_{13} , both W_n^2 and T_n^{SGP} are more powerful. These results further confirm that locally smoothing method performs better for high frequency models and globally smoothing method works better for low frequency models. The comparison also shows that W_n^2 is more robust against the underlying correlation structure of the predictors than T_n^{SGP} in both significance level maintenance and power performance.

Table 1 – 2 about here

To further investigate the performance of the proposed test, we consider the following model whose structural dimension q is greater than 1 under the alternatives. In this simulation, we show the the advantage of our test in alleviating the dimensionality problem when compared to a locally smoothing test proposed by Zheng (1996) which can be regarded as one of the representative locally smoothing methods.

Example 3. The data are generated from the model

$$\begin{aligned} H_{31} : Y &= \beta_1^\top X + a(\beta_2^\top X)^2 + \varepsilon; \\ H_{32} : Y &= \beta_1^\top X + a \exp\{-(\beta_2^\top X)^2\} + \varepsilon; \end{aligned}$$

where $\beta_1 = (\underbrace{1, \dots, 1}_{p/2}, 0, \dots, 0)^\top / \sqrt{p/2}$ and $\beta_2 = (0, \dots, 0, \underbrace{1, \dots, 1}_{p/2}) / \sqrt{p/2}$. In this example, the predictors $x_i, i = 1, \dots, n$ are i.i.d from multivariate normal distribution $N(0, I_p)$ and $N(0, \Sigma)$ and ε_i 's are from the standard univariate normal distribution $N(0, 1)$. In each case we have two cases with $p = 2$ and $p = 8$ respectively.

The simulation results are presented in Tables 3 and 4. When $p = 2$, we can see that Zheng(1996)'s test T_n^{ZH} can maintain the significance level occasionally, but usually, the empirical sizes are lower than 0.05. In contrast, W_n^2 works much better even for $n = 50$. For the empirical power, both T_n^{ZH} and W_n^2 have high power. But the power of W_n^2 grows slightly faster as a increases. When the dimension p is 8, the situation becomes very different. The empirical size of T_n^{ZH} is far away from the significance level and its power becomes much lower than that in the $p = 2$ case. Nevertheless, our test W_n^2 is much less unaffected by the dimensionality increasing than T_n^{ZH} . These phenomena validate the theoretical results that locally smoothing tests suffer from the dimensionality curse that causes slower convergence rate to their limits under the null and slower divergence rate to infinity under the alternative than globally smoothing tests.

Table 3 – 4 about here

In the following example, we consider a nonlinear null model against alternative models with higher structural dimensions. A more comprehensive comparison is made with Zheng(1996)'s test T_n^{ZH} , Guo et al.(2015)'s test T_n^{GWZ} and our test W_n^2 .

Example 4. Consider the following models

$$\begin{aligned} H_{41} : Y &= \exp\left(\frac{1}{2}X_1\right) + aX_2^3 + \varepsilon; \\ H_{42} : Y &= \exp\left(\frac{1}{2}X_1\right) + a\{X_2^3 + \cos(\pi X_3) + X_4\} + \varepsilon; \\ H_{43} : Y &= \exp\left(\frac{1}{2}X_1\right) + a\{X_2^3 + \cos(\pi X_3) + X_4 - |X_5| + X_6^2 + X_7 \times X_8\} + \varepsilon; \end{aligned}$$

where (X_1, \dots, X_p) is independent of ε and follows the standard multivariate normal

distribution $N(0, I_p)$ with $p = 4$ or 8 . Let β_i be the unit vector with the i -th component 1, $i = 1, \dots, p$ and $a = 0, 0.2, 0.4, \dots, 1$. When $a \neq 0$, the structural dimension $q = 2, B = (\beta_1, \beta_2)$ for H_{41} ; $q = 4, B = (\beta_1, \beta_2, \beta_3, \beta_4)$ for H_{42} and $q = 8, B = (\beta_1, \beta_2, \dots, \beta_8)$ for H_{43} . Note that under the alternatives, the models H_{42} and H_{43} do not have dimension reduction structure for $p = 4$ and $p = 8$ respectively. Thus, this can be used to further check the usefulness of the model adaptation method. The simulation results are reported in Figure 2.

Figure 2 about here

From this figure, we can see that when $p = 4$, the performance of our test is slightly better than the other two competitors. However, when $p = 8$, Zheng's test T_n^{ZH} behaves much worse than W_n^2 and T_n^{GWZ} . This again indicates that the dimensionality is a severe issue for the locally smoothing test without model adaptation as the adaptive-to-model test T_n^{GWZ} can also work well though it is also locally smoothing-based. For model H_{42} with $p = 4$ and model H_{43} with $p = 8$, W_n^2 and T_n^{GWZ} still work well in the power performance, even though the model has no dimension reduction structure when $a \neq 0$. Further, globally smoothing-based test procedure shows its advantage as our test W_n^2 can outperform T_n^{GWZ} when both are of the model adaptation property.

4.3 Real data analysis

This data set is studied to understand the various self-noise mechanisms. The data set is available at UCI Machine Learning Repository <https://archive.ics.uci.edu/ml/datasets/Airfoil+Self-Noise>. There are 1503 observations on one output variable: Scaled sound pressure level Y (in decibels) and five input variables: Frequency X_1 (in Hertz), Angle of attack X_2 (in degrees), Chord length X_3 (in meters), Free-stream velocity X_4 (in meters per second) and Suction side displacement thickness X_5 (in meters). For

easy interpretation, all variables are standardized separately. To establish a regression relationship between Y and 5 covariates $X = (X_1, \dots, X_5)$, we try simple model first. When the dimension reduction is applied, we find that Y may be conditionally independent of X given a projected covariate $\beta_1^\top X$ in which the direction β_1 is searched by DEE. The scatter plot in Figure 3 shows a seemly linear relationship.

Figure 3 about here

To further explore the exhaustive search of projected covariables, we use the second projected covariable searched by DEE, and then scatter plot of Y against $(\beta_1^\top X, \beta_2^\top X)$ is presented in Figure 4.

Figure 4 about here

We can see clearly that the second direction β_2 is not necessary to use as the curves along the second direction could be almost identical. In other words, the projection of the data onto the space $\beta_1^\top X$ contains almost all information of model structure. Thus, we use a linear model to fit the data where the direction $\hat{\beta}_1^\top = (-0.6323, -0.4339, -0.5339, 0.2386, -0.2644)$. To test whether the linear model is adequate, we use our test. The value of the test statistic $W_n^2 = 7.5322$ and the p -value is about 0. Hence we need to further explore a possible model. When we use a polynomial to fit the data, a cubic polynomial of $\hat{\beta}_1^\top X$ may be appropriate. The fitted curve is added into the scatter plot. See Figure 5.

Figure 5 about here

The model is as follows:

$$Y = \theta_1 + \theta_2(\beta^\top X) + \theta_3(\beta^\top X)^2 + \theta_4(\beta^\top X)^3 + \varepsilon.$$

The value of the test statistic $W_n^2 = 0.1596$ and the p -value is 0.70. Thus this model is plausible.

5 Discussions

In this paper, we propose a projection-based test that is based on residual marked empirical process and an adaptive-to-model martingale transformation. Compared to existing projection-based tests, the new test have the asymptotically distribution-free property under the null hypothesis and the omnibus property under the alternative hypothesis. This method is now for hypothetical models with dimension reduction structure. It is of great interest to investigate the application or extension of the method to hypothetical models without such kind of structure. The research is ongoing.

6 Appendix

Proof of Lemma 1. Under the regularity conditions given by Zhu et. al. (2010), Theorem 2 therein asserts that $M_n - M = O_p(n^{-1/2})$. Therefore, following the analogous argument of Zhu and Ng (1995) or Zhu and Fang (1996), we have $\hat{\lambda}_i - \lambda_i = O_p(n^{-1/2})$ for $i = 1, \dots, p$.

(I) Under H_0 , since $\dim(\mathcal{S}_{Y|X}) = 1$, we have $\lambda_1 > 0, \lambda_l = 0$ for any $l > 1$. Therefore, $\hat{\lambda}_1^2 = \lambda_1^2 + O_p(n^{-1/2})$ and $\hat{\lambda}_l^2 = O_p(n^{-1}), l = 2, \dots, p$. Hence, for any $l > 1$,

$$\begin{aligned} \frac{\hat{\lambda}_2^2 + c_n}{\hat{\lambda}_1^2 + c_n} &= \frac{c_n + O_p(1/n)}{\lambda_1^2 + c_n + O_p(1/\sqrt{n})} \rightarrow 0, \\ \frac{\hat{\lambda}_{l+1}^2 + c_n}{\hat{\lambda}_l^2 + c_n} &= \frac{c_n + O_p(1/n)}{c_n + O_p(1/n)} \rightarrow 1. \end{aligned}$$

Therefore, the minimizer $\hat{q} = 1$ with a probability going to 1.

(II) Under the alternative H_1 , $\dim(\mathcal{S}_{Y|X}) = q$, we have $\lambda_l > 0$ and $\hat{\lambda}_l^2 = \lambda_l^2 + O_p(1/\sqrt{n})$

for $l = 1, \dots, q$ and $\hat{\lambda}_l^2 = O_p(1/n)$ for $l = q + 1, \dots, p$. Hence, for $l < q$

$$\begin{aligned} \frac{\hat{\lambda}_{q+1}^2 + c_n}{\hat{\lambda}_q^2 + c_n} - \frac{\hat{\lambda}_{l+1}^2 + c_n}{\hat{\lambda}_l^2 + c_n} &= \frac{c_n + O_p(1/n)}{\lambda_q^2 + c_n + O_p(1/\sqrt{n})} - \frac{\lambda_{l+1}^2 + c_n + O_p(1/\sqrt{n})}{\lambda_l^2 + c_n + O_p(1/\sqrt{n})} \\ &\rightarrow -\frac{\lambda_{l+1}^2}{\lambda_l^2} < 0. \end{aligned}$$

For $l > q$

$$\frac{\hat{\lambda}_{q+1}^2 + c_n}{\hat{\lambda}_q^2 + c_n} - \frac{\hat{\lambda}_{l+1}^2 + c_n}{\hat{\lambda}_l^2 + c_n} = \frac{c_n + O_p(1/n)}{\lambda_q^2 + c_n + O_p(1/\sqrt{n})} - \frac{c_n + O_p(1/n)}{c_n + O_p(1/n)} \rightarrow -1 < 0.$$

Therefore, we can conclude that $Pr(\hat{q} = q) \rightarrow 1$. \square

Proof of Theorem 3.1. Under the null hypothesis, $Pr(\hat{q} = 1) \rightarrow 1$. Thus we can only work on the event $\hat{q} = 1$ as the probability of the event $\hat{q} \neq 1$ tends to 0. Therefore $\hat{\alpha} = 1$ and $V_n(u) = V_n(u, \hat{\alpha})$. Decompose the term $V_n(u)$ as follows

$$\begin{aligned} V_n(u) = V_n(u, \hat{\alpha}) &= n^{-1/2} \sum_{i=1}^n \{Y_i - g(\beta_n^\top X_i, \theta_n)\} I(\hat{B}_n^\top X_i \leq u) \\ &= n^{-1/2} \sum_{i=1}^n \{Y_i - g(\beta_n^\top X_i, \theta_n)\} I(\kappa \beta_0^\top X_i \leq u) + \\ &\quad n^{-1/2} \sum_{i=1}^n \{Y_i - g(\beta_n^\top X_i, \theta_n)\} [I(\hat{B}_n^\top X_i \leq u) - I(\kappa \beta_0^\top X_i \leq u)] \\ &\equiv: V_{1n} + V_{2n} \end{aligned}$$

where $\kappa = 1/\|\beta_0\|$. Following the analogous argument of Theorem 1 in Stute and Zhu (2002), we obtain $V_{1n} \rightarrow V_\infty - M(u)^\top V \equiv V_\infty^1$ and V_{2n} tends to zero uniformly in u . \square

Proof of Lemma 2. Using the same notations as in the proof of Lemma 1. Following the analogous argument for proving Theorem 2 in Guo et al.(2015), we obtain $M_n - M = O_p(n^{-1/2})$, therefore $\hat{\lambda}_i - \lambda_i = O_p(n^{-1/2})$ for $i = 1, \dots, p$. Note that $\lambda_l = 0$ for any $l > 1$. The proof is concluded from the exact arguments for proving Lemma 1. \square

Proof of Theorem 3.2. Again, we only work on the event $\hat{q} = 1$ as $q = 1$ under the null hypothesis. Denote $V_n^2(u) = n^{-1/2} \sum_{i=1}^n [Y_i - g(\beta_n^\top X_i, \theta_n)] I(\hat{B}_n^\top X_i \leq u)$. Under the

null, $\hat{\alpha} = 1$ and $\sup_{\|\hat{\alpha}\|=1, a_1 \geq 0} T_n V_n(u, \hat{\alpha}) = T_n V_n^2(u)$. More explicitly,

$$T_n V_n^2(u) = V_n^2(u) - \int_{-\infty}^u a_n(v)^\top A_n^{-1}(v) \left[\int_v^\infty a_n(z) V_n^2(dz) \right] \sigma_n^2(v) F_{1n}(dv).$$

Let $V_n^1(u) = n^{-1/2} \sum_{i=1}^n [Y_i - g(\beta_0^\top X_i, \theta_0)] I(\hat{B}_n^\top X_i \leq u)$. Then

$$T_n V_n^1(u) = V_n^1(u) - \int_{-\infty}^u a_n(v)^\top A_n^{-1}(v) \left[\int_v^\infty a_n(z) V_n^1(dz) \right] \sigma_n^2(v) F_{1n}(dv).$$

Here F_{1n} is the empirical distribution function of $\hat{B}_n^\top X_i, 1 \leq i \leq n$. Similarly as the arguments for proving Lemma 3.2 and Theorem 1.3 in Stute et al.(1998b), we obtain that

$$T_n V_n^2(u) = T_n V_n^1(u) + o_p(1),$$

$$T_n V_n^1(u) = TV_n^1(u) + o_p(1).$$

Recalling that under H_0 we have $V_n^0(u) = n^{-1/2} \sum_{i=1}^n [Y_i - g(\beta_0^\top X_i, \theta_0)] I(\kappa \beta_0^\top X_i \leq u)$, then

$$\begin{aligned} TV_n^0(u) - TV_n^1(u) &= V_n^0(u) - V_n^1(u) \\ &\quad - \int_{-\infty}^u a(v)^\top A^{-1}(v) \int_v^\infty a(z) V_n^0(dz) \sigma^2(v) F_{\kappa \beta_0}(dv) \\ &\quad + \int_{-\infty}^u a(v)^\top A^{-1}(v) \int_v^\infty a(z) V_n^1(dz) \sigma^2(v) F_{\kappa \beta_0}(dv). \end{aligned}$$

Using the same proof for Theorem 1 in Stute and Zhu(2002), we obtain that $TV_n^0 - TV_n^1 = o_p(1)$ uniformly in u . Therefore Lemma 3.3 in Stute et al.(1998b) gives our result. \square

Proof of Theorem 3.3. (I) First we consider the global alternative hypothesis. Under the alternative H_1 , Lemma 1 asserts that $Pr(\hat{q} = q) \rightarrow 1$, thus we work on the event $\hat{q} = q, \hat{\alpha} = \alpha = (a_1, \dots, a_q)^\top$. Therefore on this event, $\sup_{\|\hat{\alpha}\|=1, a_1 \geq 0} T_n V_n(u, \hat{\alpha}) = \sup_{\|\alpha\|=1, a_1 \geq 0} T_n V_n(u, \alpha)$. Denote $\delta_n = (\beta_n^\top, \theta_n^\top)^\top$ and $\delta_0 = (\beta_0^\top, \theta_0^\top)^\top$. According to White (1981), we have $\sqrt{n}(\delta_n - \tilde{\delta}_0) = O_p(1)$ where $\tilde{\delta}_0$ may not be equal to the true value δ_0 under the null hypothesis.

Denote

$$V_n^1(u, \alpha) = n^{-1/2} \sum_{i=1}^n [Y_i - g(\tilde{\beta}_0^\top X_i, \tilde{\theta}_0)] I(\alpha^\top \hat{B}_n^\top X_i \leq u),$$

$$V_n^2(u, \alpha) = n^{-1/2} \sum_{i=1}^n [Y_i - g(\tilde{\beta}_0^\top X_i, \tilde{\theta}_0)] I(\alpha^\top B^\top X_i \leq u).$$

Similarly as the proof for Theorem 3.2, we obtain that

$$n^{-1/2} [T_n V_n(u, \alpha) - T_n V_n^1(u, \alpha)] = o_p(1),$$

$$n^{-1/2} [T_n V_n^1(u, \alpha) - T V_n^2(u, \alpha)] = o_p(1).$$

Here

$$\begin{aligned} n^{-1/2} T V_n^2(u, \alpha) &= \frac{1}{n} \sum_{i=1}^n [Y_i - g(\tilde{\beta}_0^\top X_i, \tilde{\theta}_0)] I(\alpha^\top B^\top X_i \leq u) \\ &\quad - \frac{1}{\sqrt{n}} \int_{-\infty}^u a_1(v, \alpha)^\top A_1^{-1}(v, \alpha) \int_v^\infty a_1(z, \alpha) V_n^2(dz) \sigma_1^2(v, \alpha) F_\alpha(dv) \end{aligned}$$

Note that

$$\frac{1}{\sqrt{n}} \int_v^\infty a_1(z, \alpha) V_n^2(dz) = \frac{1}{n} \sum_{i=1}^n I(\alpha^\top B^\top X_i \geq v) a_1(\alpha^\top B^\top X_i, \alpha) [Y_i - g(\tilde{\beta}_0^\top X_i, \tilde{\theta}_0)].$$

Then we derive that under H_1 ,

$$n^{-1/2} T V_n^2(u, \alpha) \rightarrow \int_{-\infty}^u H(v, \alpha) F_\alpha(dv) - \int_{-\infty}^u a_1(v, \alpha)^\top A_1^{-1}(v, \alpha) a_2(v, \alpha) \sigma_1^2(v, \alpha) F_\alpha(dv),$$

where F_α is the distribution function of $\alpha^\top B^\top X$ and

$$\begin{aligned} H(v, \alpha) &= E(G(B^\top X) - g(\tilde{\beta}_0^\top X, \tilde{\theta}_0) | \alpha^\top B^\top X = v), \\ \sigma_1^2(v, \alpha) &= E[(G(B^\top X) - g(\tilde{\beta}_0^\top X, \tilde{\theta}_0))^2 + \varepsilon^2 | \alpha^\top B^\top X = v], \\ a_1(v, \alpha) &= \{g_1(v/\kappa_1, \tilde{\theta}_0) E(X^\top | \alpha^\top B^\top X = v) / \sigma_1^2(v, \alpha), g_2(v/\kappa_1, \tilde{\theta}_0) / \sigma_1^2(v, \alpha)\}^\top, \\ A_1(v, \alpha) &= E\{I(\alpha^\top B^\top X \geq v) a_1(\alpha^\top B^\top X, \alpha) (g_1(\tilde{\beta}_0^\top X, \tilde{\theta}_0) X^\top, g_2(\tilde{\beta}_0^\top X, \tilde{\theta}_0))\}, \\ a_2(v, \alpha) &= E\{I(\alpha^\top B^\top X \geq v) a_1(\alpha^\top B^\top X, \alpha) (G(B^\top X) - g(\tilde{\beta}_0^\top X, \tilde{\theta}_0))\}. \end{aligned}$$

Hence we conclude that

$$n^{-1/2}T_nV_n(u, \alpha) \rightarrow \int_{-\infty}^u H(v, \alpha)F_\alpha(dv) - \int_{-\infty}^u a_1(v, \alpha)^\top A_1^{-1}(v, \alpha)a_2(v, \alpha)\sigma_1^2(v, \alpha)F_\alpha(dv).$$

Therefore

$$\frac{1}{\sqrt{n}} \sup_{\|\hat{\alpha}\|=1, a_1 \geq 0} T_nV_n(u, \hat{\alpha}) \rightarrow \text{some nonzero process.}$$

The resulting test statistic converges to infinity at the rate of $O(1/n)$.

(II) Under the local alternatives H_{1n} , Lemma 2 asserts that $P(\hat{q} = 1) \rightarrow 1$ as $n \rightarrow \infty$, thus we also consider the event $\hat{q} = 1$. Denote

$$V_n^2(u) = n^{-1/2} \sum_{i=1}^n [Y_i - g(\beta_n^\top X_i, \theta_n)] I(\hat{B}_n^\top X_i \leq u).$$

Therefore $\hat{\alpha} = 1$, \hat{B}_n is a vector and $\sup_{\|\hat{\alpha}\|=1, a_1 \geq 0} T_nV_n(u, \hat{\alpha}) = T_nV_n^2(u)$. Let

$$V_n^1(u) = n^{-1/2} \sum_{i=1}^n [Y_i - g(\beta_0^\top X_i, \theta_n)] I(\hat{B}_n^\top X_i \leq u).$$

Following the analogous argument for proving Theorem 3.3 we obtain that

$$T_nV_n^2(u) = T_nV_n^1(u) + o_p(1),$$

$$T_nV_n^1(u) = TV_n^1(u) + o_p(1),$$

$$TV_n^0(u) = TV_n^1(u) + o_p(1).$$

To finish the proof, it remains to derive the limit of $TV_n^0(u)$. Recall that $V_n^0(u) = n^{-1/2} \sum_{i=1}^n [Y_i - g(\beta_0^\top X_i, \theta_0)] I(\kappa\beta_0^\top X_i \leq u)$ and

$$TV_n^0(u) = V_n^0(u) - \int_{-\infty}^u a(v)^\top A^{-1}(v) \int_v^\infty a(z)V_n^0(dz)\sigma^2(v)F_{\kappa\beta_0}(dv).$$

Under the local alternative H_{1n} ,

$$V_n^0(u) = \frac{1}{n} \sum_{i=1}^n G(B^\top X_i) I(\kappa\beta_0^\top X_i \leq u) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \varepsilon_i I(\kappa\beta_0^\top X_i \leq u).$$

Similarly as the proof for Theorem 1.1 in Stute(1997), we have

$$V_n^0(u) \rightarrow V_\infty(u) + \int_{-\infty}^u H_0(v) F_{\kappa\beta_0}(dv),$$

where $H_0(v) = E[G(B^\top X) | \kappa\beta_0^\top X = v]$.

For the second term in $TV_n^0(u)$, note that

$$\begin{aligned} \int_v^\infty a(z) V_n^0(dz) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n I(\kappa\beta_0^\top X_i \geq v) a(\kappa\beta_0^\top X_i) [Y_i - g(\beta_0^\top X_i, \theta_0)] \\ &= \frac{1}{n} \sum_{i=1}^n I(\kappa\beta_0^\top X_i \geq v) a(\kappa\beta_0^\top X_i) G(B^\top X_i) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n I(\kappa\beta_0^\top X_i \geq v) a(\kappa\beta_0^\top X_i) \varepsilon_i. \end{aligned}$$

Therefore,

$$\int_v^\infty a(z) V_n^0(dz) \rightarrow W_1(v) + W_2(v) \quad \text{in distribution}$$

where $W_1(v) = \int_v^\infty a(z) H_0(z) F_{\kappa\beta_0}(dz)$ and $W_2(v)$ have a normal distribution with mean zero and variance

$$\sigma_2^2(v) = \int_v^\infty \sigma^2(u) a(u) a(u)^\top F_{\kappa\beta_0}(du).$$

Here $\sigma^2(u) = E(\varepsilon^2 | \kappa\beta_0^\top X = u)$. Hence we can conclude that

$$TV_n^0(u) \rightarrow V_\infty(u) + \int_{-\infty}^u \{H_0(v) - a(v)^\top A^{-1}(v) [W_1(v) + W_2(v)] \sigma^2(v)\} F_{\kappa\beta_0}(dv).$$

□

References

- Cook, R. D. (1998). *Regression Graphics: Ideas for Studying Regressions Through Graphics*. New York: Wiley.
- Cook, R. D. and Weisberg, S. (1991). Discussion of Sliced inverse regression for dimension reduction, by K. C. Li. *Journal of the American Statistical Association*, **86**, 316-342.
- Cook, R. D. and Forzani, L. (2009). Likelihood-based sufficient dimension reduction. *Journal of the American Statistical Association*, **104**, 197-208.
- Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimates. *The Annals of Statistics*. **27**, 1012-1050.
- Escanciano, J. C.(2006). A consistent diagnostic test for regression models using projections. *Econometric Theory*, **22**, 1030-1051.
- Eubank, R. L., Li, C. S. and Wang, S. (2005). Testing lack-of-fit of parametric regression models using nonparametric regression techniques. *Statistica Sinica*, **15**, 135-152.
- Fan, J. Q. and Huang, L. S. (2001). Goodness-of-fit tests for parametric regression models, *Journal of the American Statistical Association*, **96**, 640-652.
- Fan, Y. and Li, Q. (1996). Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica*, **64**, 865-890.
- Fan, J., Zhang, C. and Zhang, J. (2001) Generalized likelihood ratio statistics and Wilks phenomenon. *The Annals of Statistics*, **29**, 153-193.
- Friedman, J. H. and Stuetzle, W. (1981) Projection pursuit regression. *Journal of the American Statistical Association*, **76**, 817-823.

- Guo, X., Wang, T. and Zhu, L. X. (2015). Model checking for generalized linear models: a dimension-reduction model-adaptive approach. *Journal of the Royal Statistical Society: Series B*,
- González-Manteiga, W. and Crujeiras, R. M. (2013). An updated review of Goodness-of-Fit tests for regression models. *TEST*, **22**, 361-411.
- Khmaladze, E. V. and Koul, H. L. (2004) Martingale transforms goodness-of-fit tests in regression models. *The Annals of Statistics*, **37**, 995-1034
- Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, **21**, 1926-1947.
- Hart, J. (1997) *Nonparametric smoothing and lack-of-fit tests*. Springer, Berlin.
- Khmaladze, E V. (1982). Martingale Approach in the Theory of Goodness-of-fit Tests. *Theory of Probability & Its Applications*, **26**, 240-257.
- Khmaladze, E V. and Koul, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *The Annals of Statistics*, **32**, 995-1034.
- Koul, H. L. and Ni, P. P. (2004). Minimum distance regression model checking. *Journal of Statistical Planning and Inference*, **119**, 109-141.
- Lavergne, P. and Patilea, V. (2008). Breaking the curse of dimensionality in non parametric testing. *Journal of Econometrics*, **143**, 103-122.
- Lavergne, P. and Patilea, V. (2012). One for all and all for one: regression checks with many regressors. *Journal of Business & Economic Statistics*, **30**, 41-52.
- Li, K. C. (1991). Sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, **86**, 316-327.

- Li, B. and Wang, S. (2007). On directional regression for dimension reduction. *Journal of the American Statistical Association*, **102**, 997-1008.
- Li, B., Wen, S. Q. and Zhu, L. X. (2008). On a Projective Resampling method for dimension reduction with multivariate responses. *Journal of the American Statistical Association*. **103**, 1177-1186.
- Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*. **25**, 613-641.
- Stute, W., Gonzales-Manteiga, W. and Presedo-Quindimil, M. (1998a). Bootstrap approximation in model checks for regression. *Journal of the American Statistical Association*., **93**, 141-149.
- Stute, W., Thies, S. and Zhu, L. X. (1998b). Model checks for regression: An innovation process approach. *The Annals of Statistics*. **26**, 1916-1934.
- Stute, W., Xu, W. L. and Zhu, L. X. (2008). Model diagnosis for parametric regression in high dimensional spaces. *Biometrika*. 95. 1-17.
- Stute, W. and Zhu, L. X. (2002). Model checks for generalized linear models. *Scandinavian Journal of Statistics*. **29**, 535-545.
- Stute, W. and Zhu, L. X. (2005). Nonparametric checks for single-index models, *The Annals of Statistics*, **33**, 1048-1083.
- 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.
- Wong, H.L., Fang, K.T. and Zhu, Lixing (1995). A test for multivariate normality based on sample entropy and projection pursuit. *J. of Statistical planning and inference*. 45, 373-385.

- Xia, Y. C. (2006). Asymptotic distributions for two estimators of the single index model. *Econometric Theory*, **22**, 1112-1137.
- Xia, Y. C. (2009). Model checking in regression 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 (with discussion). *Journal of the Royal Statistical Society: Series B*, **64**, 363-410.
- Xia, Q., Xu, W. L. and Zhu, L. X. (2014). Consistently determining the number of factors in multivariate volatility modelling. *Statistica Sinica*, accepted.
- Zhang, C. and Dette, H. (2004). A power comparison between nonparametric regression tests. *Statistics & Probability Letters*, **66**, 289-301.
- Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, **75**, 263-289.
- Zhu, L. X. (2003). Model checking of dimension-reduction type for regression. *Statistica Sinica*, **13**, 283-296.
- Zhu, L. X. and An, H. Z. (1992). A nonlinearity test in regression models. *Journal of Mathematics*, **12**, 391-397. (in Chinese)
- Zhu, L. X. and Fang, K. T. (1996). Asymptotics for the kernel estimates of sliced inverse regression. *Annals of Statistics*, **24**, 1053-1067.
- Zhu, L.X., and Li R. (1998). Dimension-reduction type test for linearity of a stochastic model. *Acta Math. Appli. Sinica*, **14**, 165 - 175.
- Zhu, L. X. and Ng, K. W. (1995). Asymptotics for sliced inverse regression. *Statistica Sinica*, **5**, 727-736.

- Zhu, L. P., Zhu, L.X. , Ferré, L. and Wang, T. (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. *Journal of the American Statistical Association* , **105**, 1455-1466.
- Zhu, L. X. and Ng, K. W. (1995). Asymptotics for sliced inverse regression. *Statistica Sinica*, **5**, 727-736.
- Zhu, X. H., Guo, X. and Zhu, L. X. (2014). Model checking for generalized partially linear models: a dimension reduction approach. *Working paper*.

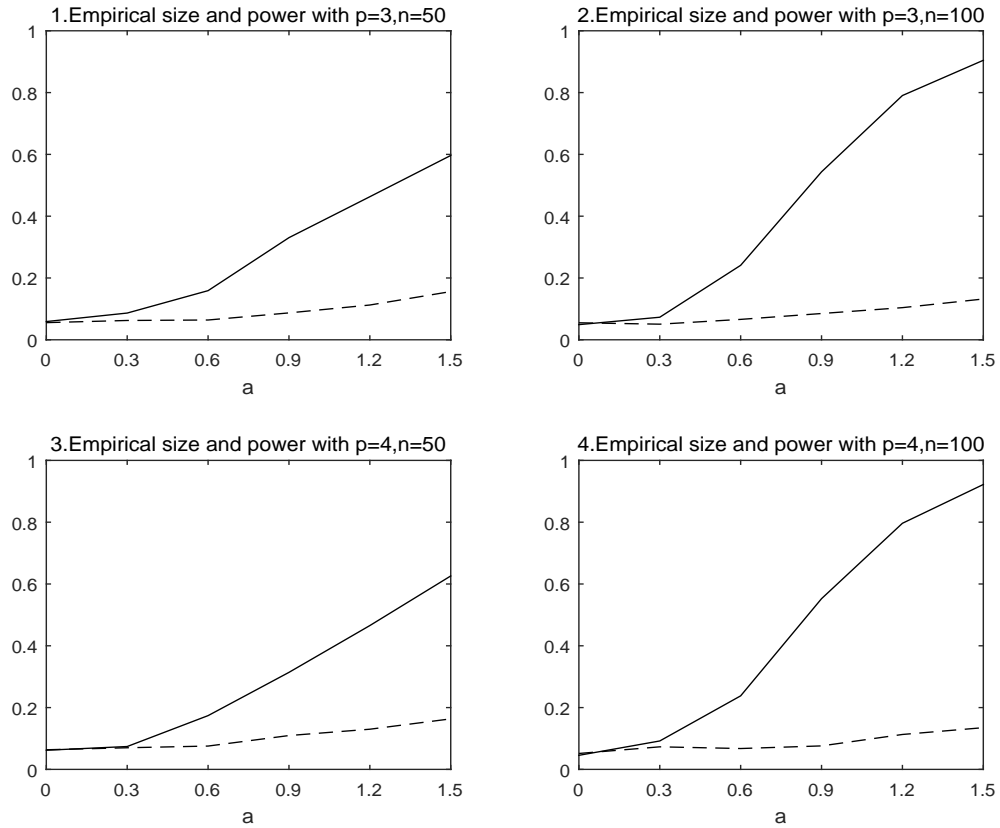


Figure 1: The empirical sizes and powers of T_n^{SZ} and W_n^2 in Example 1. The dash and solid line denote the results of T_n^{SZ} and W_n^2 respectively.

Table 1: Empirical sizes and powers of T_n^{GWZ} , W_n^2 and T_n^{SGP} for H_0 vs. H_{11}, H_{12} and H_{13} in Example 2.

	a	T_n^{GWZ}		W_n^2		T_n^{SGP}	
		n=50	n=100	n=50	n=100	n=50	n=100
$H_{11}, X \sim N(0, I_p)$	0.0	0.0445	0.0480	0.0470	0.0525	0.0640	0.0650
	0.2	0.0830	0.1490	0.0635	0.0950	0.1050	0.1320
	0.4	0.1915	0.4595	0.1180	0.2155	0.1710	0.3260
	0.6	0.4245	0.8115	0.2005	0.4245	0.2930	0.5680
	0.8	0.6025	0.9590	0.3090	0.6480	0.4600	0.8010
	1.0	0.7590	0.9915	0.4170	0.8285	0.5830	0.9060
$H_{11}, X \sim N(0, \Sigma)$	0.0	0.0470	0.0460	0.0465	0.0505	0.0760	0.0680
	0.2	0.0655	0.1090	0.0500	0.0495	0.0910	0.0650
	0.4	0.1350	0.3595	0.0485	0.0590	0.0790	0.0960
	0.6	0.2645	0.6870	0.0645	0.0935	0.0970	0.1370
	0.8	0.4065	0.8905	0.0640	0.1035	0.1100	0.1700
	1.0	0.5580	0.9780	0.0840	0.1650	0.1300	0.2280
$H_{12}, X \sim N(0, I_p)$	0.0	0.0450	0.0515	0.0495	0.0535	0.0590	0.0590
	0.2	0.0530	0.0680	0.0750	0.1220	0.0930	0.1190
	0.4	0.0965	0.1550	0.1865	0.3430	0.2180	0.3260
	0.6	0.1670	0.3145	0.3505	0.6565	0.3610	0.6020
	0.8	0.2595	0.5400	0.5320	0.8705	0.5550	0.8420
	1.0	0.3685	0.7535	0.7085	0.9655	0.7170	0.9570
$H_{12}, X \sim N(0, \Sigma)$	0.0	0.0520	0.0540	0.0525	0.0510	0.0760	0.0680
	0.2	0.0955	0.1675	0.1705	0.4230	0.2020	0.3420
	0.4	0.2465	0.5385	0.5050	0.8770	0.4370	0.7460
	0.6	0.4510	0.8520	0.7330	0.9900	0.6670	0.9130
	0.8	0.6455	0.9670	0.8780	0.9995	0.7980	0.9510
	1.0	0.7940	0.9935	0.9550	1.0000	0.8980	0.9600

Table 2: Empirical sizes and powers of T_n^{GWZ} , W_n^2 and T_n^{SGP} for H_0 vs. H_{11}, H_{12} and H_{13} in Example 2.

	a	T_n^{GWZ}		W_n^2		T_n^{SGP}	
		n=50	n=100	n=50	n=100	n=50	n=100
$H_{13}, X \sim N(0, I_p)$	0.0	0.0450	0.0500	0.0490	0.0500	0.0790	0.0650
	0.2	0.0540	0.0735	0.1075	0.1610	0.1280	0.1640
	0.4	0.0990	0.2030	0.2605	0.5250	0.2100	0.3870
	0.6	0.1905	0.4590	0.4610	0.8350	0.3970	0.6900
	0.8	0.3365	0.7550	0.6625	0.9575	0.5520	0.8660
	1.0	0.4830	0.9120	0.7925	0.9940	0.7120	0.9620
$H_{13}, X \sim N(0, \Sigma)$	0.0	0.0495	0.0480	0.0500	0.0470	0.0710	0.0740
	0.2	0.1385	0.2640	0.3565	0.7110	0.3050	0.5380
	0.4	0.4490	0.8575	0.7930	0.9920	0.6910	0.9610
	0.6	0.7750	0.9935	0.9455	0.9995	0.8970	0.9970
	0.8	0.9005	0.9995	0.9790	1.0000	0.9700	1.0000
	1.0	0.9525	1.0000	0.9925	1.0000	0.9860	1.0000

Table 3: Empirical sizes and powers of T_n^{ZH} and W_n^2 for H_0 vs. H_{31} in Example 3.

	a	T_n^{ZH}		W_n^2	
		n=50	n=100	n=50	n=100
$H_{31}, X \sim N(0, I_p), p = 2$	0.0	0.0345	0.0430	0.0465	0.0500
	0.2	0.0820	0.1505	0.2095	0.4375
	0.4	0.3020	0.6170	0.6240	0.9210
	0.6	0.6180	0.9440	0.8615	0.9925
	0.8	0.8410	0.9930	0.9445	1.0000
	1.0	0.9345	0.9995	0.9885	1.0000
$H_{31}, X \sim N(0, I_p), p = 8$	0.0	0.0265	0.0295	0.0500	0.0450
	0.2	0.0260	0.0475	0.2095	0.4190
	0.4	0.0360	0.0850	0.5770	0.9020
	0.6	0.0765	0.1640	0.8100	0.9935
	0.8	0.1145	0.2600	0.9260	0.9980
	1.0	0.1635	0.3805	0.9560	1.0000
$H_{31}, X \sim N(0, \Sigma), p = 2$	0.0	0.0315	0.0390	0.0520	0.0475
	0.2	0.0930	0.1565	0.2335	0.4660
	0.4	0.3250	0.6530	0.6275	0.9305
	0.6	0.6515	0.9550	0.8690	0.9955
	0.8	0.8740	0.9985	0.9510	1.0000
	1.0	0.9550	1.0000	0.9775	1.0000
$H_{31}, X \sim N(0, \Sigma), p = 8$	0.0	0.0185	0.0350	0.0465	0.0530
	0.2	0.0695	0.1495	0.5565	0.9055
	0.4	0.2045	0.4365	0.9330	1.0000
	0.6	0.3370	0.7320	0.9835	1.0000
	0.8	0.4740	0.8580	0.9930	1.0000
	1.0	0.5545	0.9200	0.9970	1.0000

Table 4: Empirical sizes and powers of T_n^{ZH} and W_n^2 for H_0 vs. H_{32} in Example 3.

	a	T_n^{ZH}		W_n^2	
		n=50	n=100	n=50	n=100
$H_{32}, X \sim N(0, I_p), p = 2$	0.0	0.034	0.0455	0.0530	0.0500
	0.2	0.083	0.1155	0.1215	0.2050
	0.4	0.250	0.4730	0.3430	0.6100
	0.6	0.524	0.8480	0.6200	0.9195
	0.8	0.782	0.9785	0.8540	0.9920
	1.0	0.935	0.9985	0.9575	0.9995
$H_{32}, X \sim N(0, I_p), p = 8$	0.0	0.0215	0.0265	0.0550	0.0480
	0.2	0.0285	0.0375	0.1185	0.1970
	0.4	0.0475	0.0760	0.3215	0.5830
	0.6	0.0650	0.1550	0.5690	0.8910
	0.8	0.1280	0.2930	0.7965	0.9900
	1.0	0.1765	0.4210	0.9230	1.0000
$H_{32}, X \sim N(0, \Sigma), p = 2$	0.0	0.0310	0.0430	0.0510	0.0515
	0.2	0.0745	0.1410	0.1205	0.1880
	0.4	0.2545	0.4900	0.3190	0.5955
	0.6	0.5420	0.8540	0.6160	0.9150
	0.8	0.8105	0.9835	0.8400	0.9880
	1.0	0.9420	0.9990	0.9480	0.9995
$H_{32}, X \sim N(0, \Sigma), p = 8$	0.0	0.0270	0.0295	0.0520	0.0470
	0.2	0.0235	0.0395	0.0885	0.1430
	0.4	0.0485	0.0770	0.1895	0.3720
	0.6	0.0725	0.1480	0.3750	0.7000
	0.8	0.1145	0.2845	0.5705	0.9015
	1.0	0.1900	0.4605	0.7465	0.9715

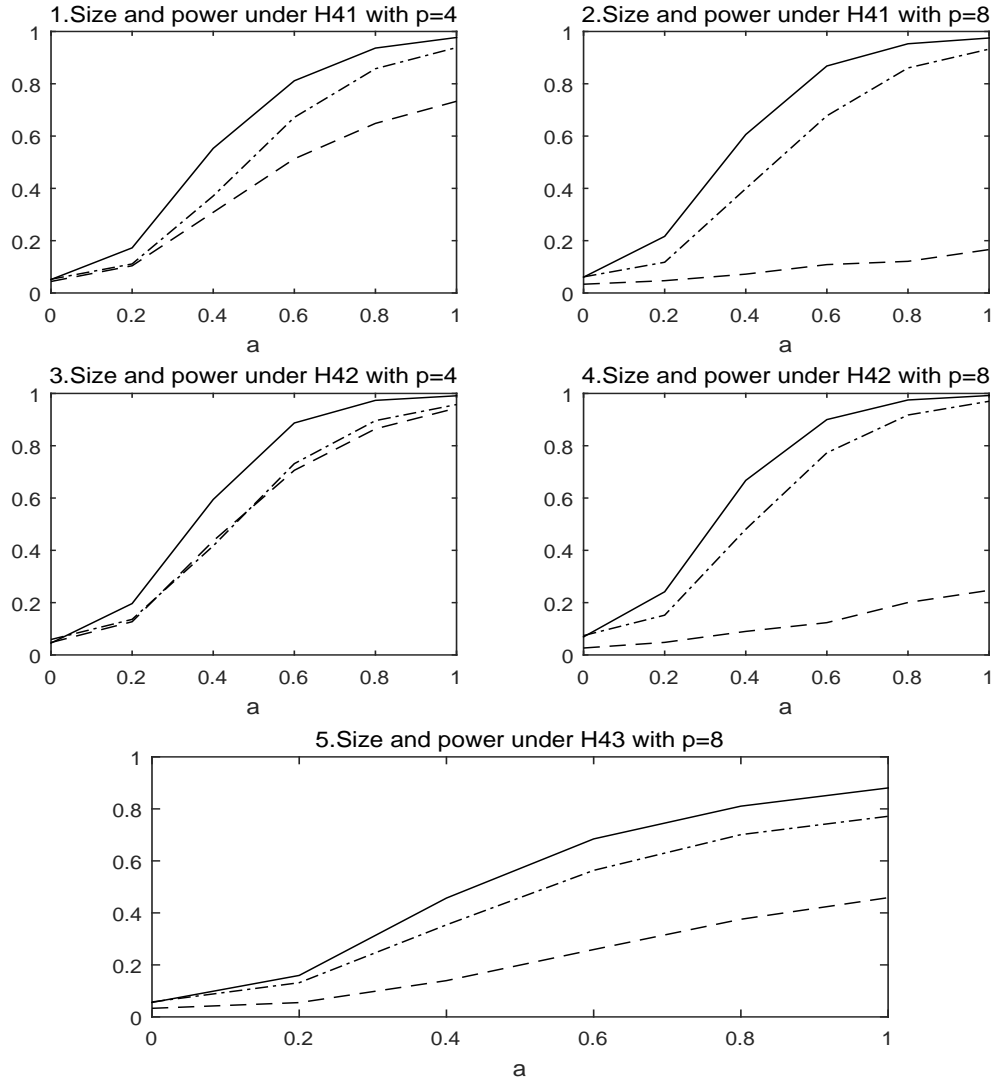


Figure 2: The empirical sizes and powers of T_n^{ZH} , T_n^{GWZ} and W_n^2 in Example 4. The dash, dash-dotted and solid line denote the results of T_n^{ZH} , T_n^{GWZ} and W_n^2 respectively.

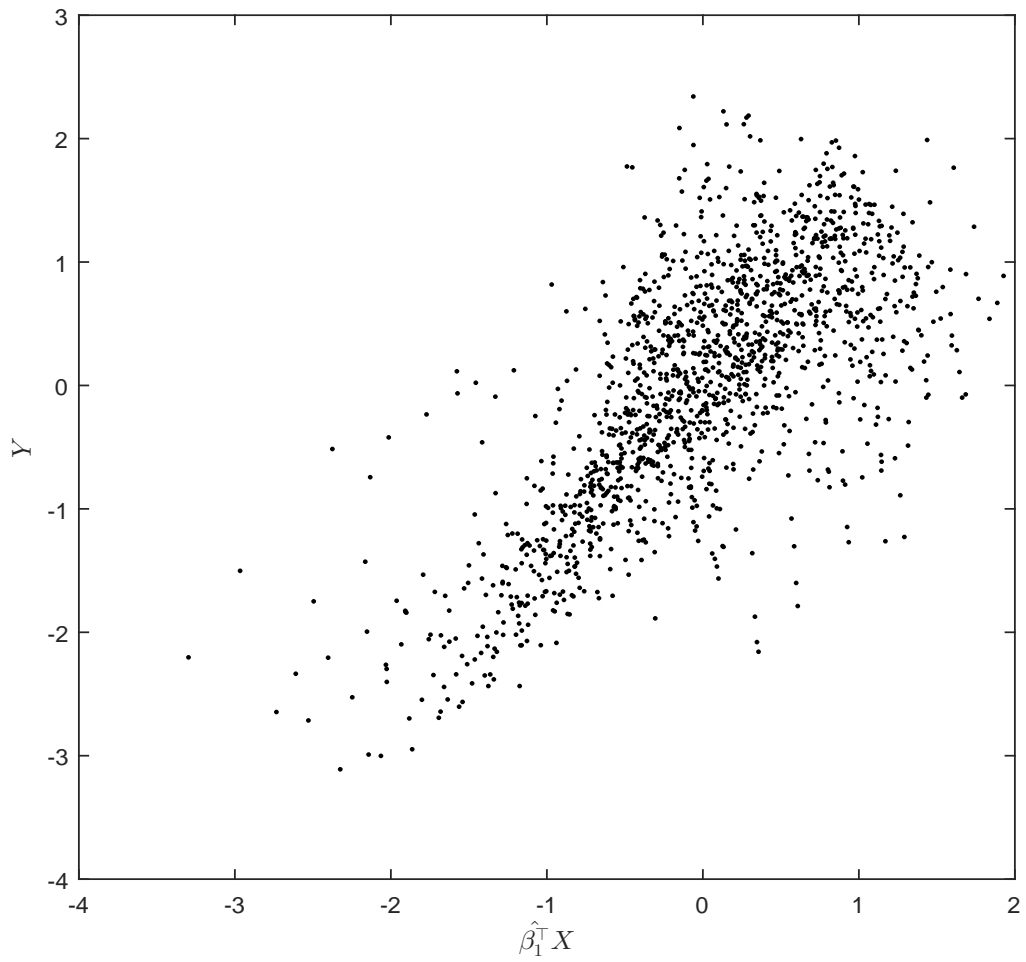


Figure 3: Scatter plot of the response against the $\hat{\beta}_1^T X$ in which the direction $\hat{\beta}_1$ is obtained by DEE.

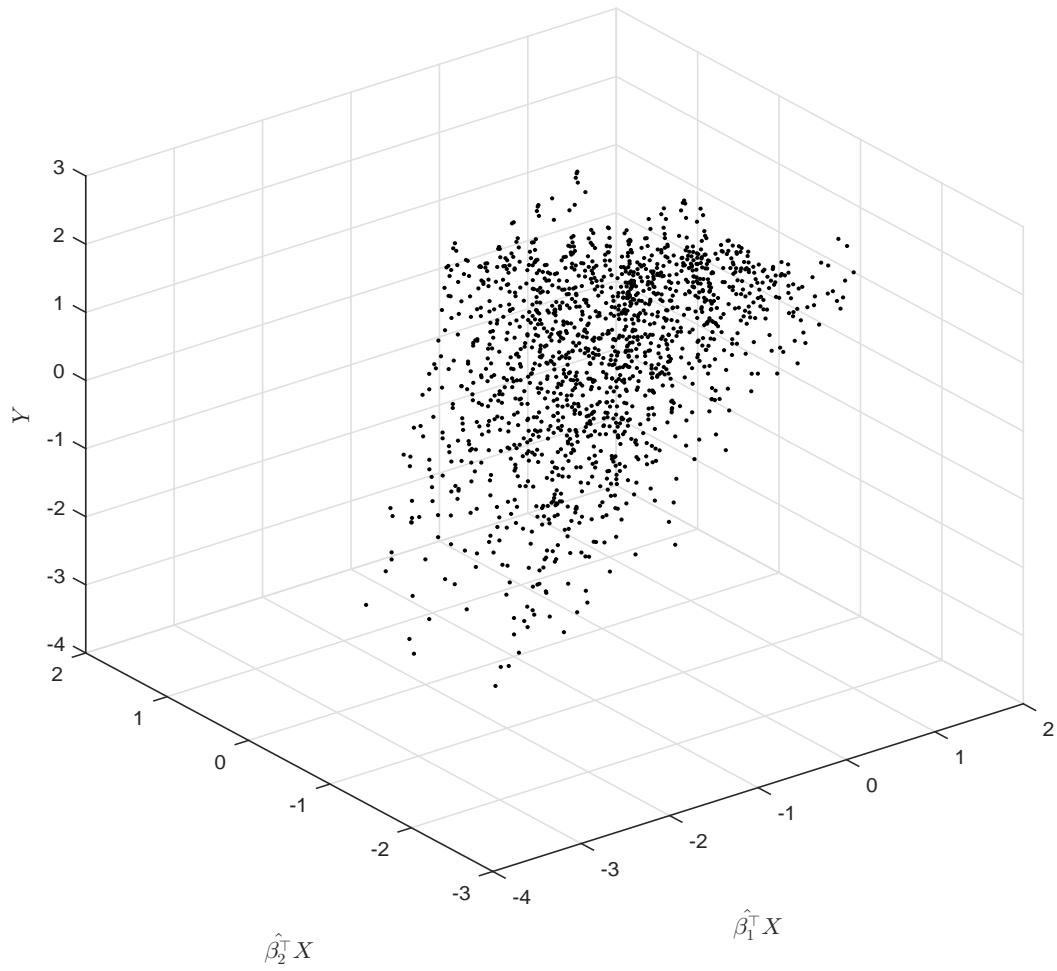


Figure 4: Scatter plot of the response against the $(\hat{\beta}_1^\top X, \hat{\beta}_2^\top X)$ in which the directions $\hat{\beta}_1$ and $\hat{\beta}_2$ are obtained by DEE.

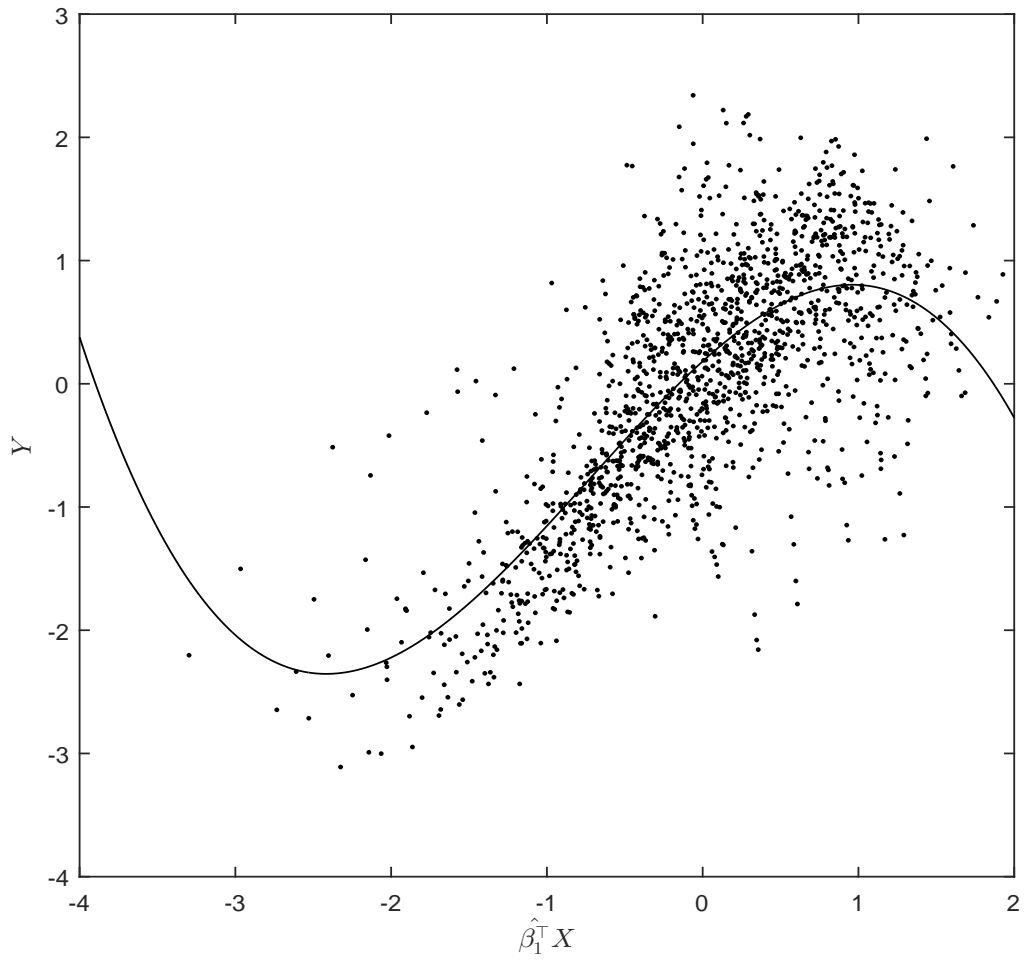


Figure 5: Plot of Y against $\hat{\beta}_1^T X$ obtained by DEE and the fitted cubic polynomial curve.