Journal of Multivariate Analysis 136 (2015) 75–85

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

On testing common indices for two multi-index models: A link-free approach

Xuejing Liu [a], Zhou Yu [b], Xuerong Meggie Wen [a,*], Robert Paige [a]

[a] Department of Mathematics and Statistics, Missouri University of Science and Technology, MO 65409, USA

[b] School of Finance and Statistics, East China Normal University, Shanghai, China

ARTICLE INFO

Article history:
Received 28 May 2014
Available online 21 January 2015

AMS subject classifications:
62H86
62H15
62H25
62G08

Keywords:
Multi-index models
Sufficient dimension reduction
Sliced inverse regression
Sliced average variance estimation
Directional regression

ABSTRACT

We propose a link-free procedure for testing whether two multi-index models share identical indices via the sufficient dimension reduction approach. Test statistics are developed based upon three different sufficient dimension reduction methods: (i) sliced inverse regression, (ii) sliced average variance estimation and (iii) directional regression. The asymptotic null distributions of our test statistics are derived. Monte Carlo studies are performed to investigate the efficacy of our proposed methods. A real-world application is also considered.

1. Introduction

For a regression problem with univariate response Y and p-dimensional predictors $\mathbf{X} = (X_1, \ldots, X_p)^T$, we consider the following generalized multi-index model

$$Y = g(\boldsymbol{\beta}_1^T\mathbf{X}, \ldots, \boldsymbol{\beta}_d^T\mathbf{X}; \epsilon), \tag{1.1}$$

where $g(\cdot)$ is an unknown link function, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d)$ is a $p \times d$ matrix, $d \le p$, and the random error ϵ is independent with $\mathbf{X}$. Model (1.1) is a very general semiparametric model which includes the multi-index model [12,27] and the single-index model [11,29] with $Y = g(\boldsymbol{\beta}^T\mathbf{X}) + \epsilon$ as special cases. One is usually concerned with estimation of indices $\boldsymbol{\beta}$, the total number of indices d and the link function $g(\cdot)$ [9]. We, however, focus on testing if two multi-index models share identical indices (subspaces). Specifically, consider two d-dimensional multi-index models for two populations (groups):

$$Y = g_1(\boldsymbol{\beta}_1^T\mathbf{X}, \ldots, \boldsymbol{\beta}_d^T\mathbf{X}; \epsilon_1), \quad \textit{for group } 1;$$

$$Y = g_2(\boldsymbol{\xi}_1^T\mathbf{X}, \ldots, \boldsymbol{\xi}_d^T\mathbf{X}; \epsilon_2), \quad \textit{for group } 2. \tag{1.2}$$

Since the identifiable parameters here are the subspaces spanned by the columns of $\boldsymbol{\beta}$ and $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_d)$, rather than $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$ themselves, we develop a test of null hypothesis

$$\text{span}(\boldsymbol{\beta}) = \text{span}(\boldsymbol{\xi}), \tag{1.3}$$

* Corresponding author.
E-mail address: wenx@mst.edu (X.M. Wen).

http://dx.doi.org/10.1016/j.jmva.2015.01.009
0047-259X/

where both $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$ are $p \times d$ matrices. This hypothesis is similar in nature to the null hypothesis of common principal component subspaces for Common PCA considered in [20].

A hypothesis test of this type might be of special interest in many applications involving two datasets, where the same variables are being measured on objects from two different groups, and for which it is of interest to determine how similar the two groups are with respect to the span of the indices of predictor vectors regardless of the unknown link functions.

Consider the AIS dataset discussed by Weisberg [24], which contains information on the lean body mass L and other physical and hematological measurements ($\mathbf{X}$), from 102 male and 100 female elite Australian athletes who trained at the Australian Institute of Sport. We investigate how the relationship between the body fat and various predictors varies with gender. Suppose that subject matter knowledge and prior modeling experience suggest that a d-dimensional multi-index model of the form (1.1) applies to both female and male groups, naturally, we would like to know if the equivalent set of indices of the hematological measurements serve for both genders. Informal comparisons such as those based upon graphical methods can be carried out. However, such comparisons might become unwieldy when d is greater than 2, and the resulting conclusions could be overly subjective. Hence, a formal test seems necessary here. This is the motivation for our development of a test statistic for the null hypothesis in (1.3).

We propose a link-free test for testing hypothesis of (1.3) via a sufficient dimension reduction approach [15,3]. For a regression problem, the scope of sufficient dimension reduction is to seek a minimal set of indices of $\mathbf{X}$, say $\boldsymbol{\beta}^T\mathbf{X} = (\boldsymbol{\beta}_1^T\mathbf{X}, \ldots, \boldsymbol{\beta}_d^T\mathbf{X})$, for which the distribution of $Y|\boldsymbol{\beta}^T\mathbf{X}$ is the same as the original regression $Y|\mathbf{X}$, without assuming a parametric model. Numerous approaches are available in the literature including sliced inverse regression (SIR; [15]), sliced average variance estimation (SAVE; [5]), minimum average variance estimation (MAVE; [28]), directional regression (DR; [18]), likelihood acquired directions (LAD; [4]), and dimension reduction via central solution space [16].

The rest of this article is organized as follows. In Section 2, we give a brief review of sufficient dimension reduction methods. Specifically, we focus on those methods based upon a spectral decomposition [25]. In Section 3, we present our link-free test statistic for null hypothesis (1.3). The asymptotic distribution of our test statistic is also discussed. We illustrate the performance of our method with Monte Carlo studies in Section 4. We then apply our method to the AIS dataset. Brief conclusions and a discussion on the future research directions are given in Section 5. For ease of exposition, we defer some technical details to Appendix.

2. A brief review on sufficient dimension reduction: the spectral decomposition approach

In this section, we give a brief review on how to use sufficient dimension reduction to make inference about span($\boldsymbol{\beta}$) in model (1.1). In particular, we consider three commonly used sufficient dimension reduction methods: SIR, SAVE and DR.

Let $\boldsymbol{\Sigma} = \mathrm{Var}(\mathbf{X})$, $\boldsymbol{\mu} = \mathrm{E}(\mathbf{X})$, and $\mathbf{Z}$ be the standardized predictor $\boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$. Many moment based sufficient dimension reduction methods may be formulated as the solution to the following eigendecomposition problem:

$$\mathbf{M}_z\boldsymbol{\eta}_i = \lambda_i\boldsymbol{\eta}_i, \quad i = 1, \ldots, p,$$

where $\mathbf{M}_z$ is the $\mathbf{Z}$ scale method-specific candidate matrix. Assuming the *linearity condition* [15] holds, which is a mild condition imposed on the marginal distribution of the predictors alone, the eigenvectors $(\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_d)$ corresponding to the non-zero eigenvalues $\lambda_1 \geq \cdots \geq \lambda_d$ form a basis of the $\mathbf{Z}$ scale central subspace $\mathcal{S}_{Y|\mathbf{Z}}$. Then by the invariance property $\mathcal{S}_{Y|\mathbf{X}} = \boldsymbol{\Sigma}^{-1/2}\mathcal{S}_{Y|\mathbf{Z}}$, as described by Cook [3], $\boldsymbol{\beta} = (\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\eta}_1, \ldots, \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\eta}_d)$ forms a basis for $\mathcal{S}_{Y|\mathbf{X}}$. The linearity condition, which basically requires that $\mathrm{E}(\mathbf{X}|\boldsymbol{\beta}^T\mathbf{X})$ is a linear function of $\boldsymbol{\beta}^T\mathbf{X}$, is a common assumption in dimension reduction methods and holds for elliptically contoured predictors [8]. Additionally, Hall and Li [10] showed that as the number of predictors p increases, the linearity condition holds to a reasonable approximation in many problems.

For the three sufficient dimension reduction methods that target $\mathcal{S}_{Y|\mathbf{Z}}$, the corresponding candidate matrices are summarized below:

$$\begin{aligned}
\text{Sliced Inverse Regression:}\quad & \mathbf{M}_z = \mathrm{Var}\{\mathrm{E}(\mathbf{Z}|Y)\};\\
\text{Sliced Average Variance Estimation:}\quad & \mathbf{M}_z = \mathrm{E}\{I_p - \mathrm{Var}(\mathbf{Z}|Y)\}^2;\\
\text{Directional Regression:}\quad & \mathbf{M}_z = 2\mathrm{E}\{\mathrm{E}^2(\mathbf{Z}\mathbf{Z}^T)\} + 2\mathrm{E}^2\{\mathrm{E}(\mathbf{Z}|Y)\mathrm{E}(\mathbf{Z}^T|Y)\}\\
& \quad + 2\mathrm{E}\{\mathrm{E}(\mathbf{Z}^T|Y)E(\mathbf{Z}|Y)\}\mathrm{E}\{\mathrm{E}(\mathbf{Z}|Y)\mathrm{E}(\mathbf{Z}^T|Y)\} - 2I_p.
\end{aligned}$$

Although in the literature, people tend to work with standardized predictors, for our purpose, it is easier to describe the candidate matrices in terms of the original predictor $\mathbf{X}$. Since we will make use of the eigenprojection corresponding to the non-zero eigenvalues, the $\boldsymbol{\beta}_i = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\eta}_i$ provided by the above approach are orthonormal under the inner-product of $\langle\mathbf{a}, \mathbf{b}\rangle = \mathbf{a}^T\boldsymbol{\Sigma}\mathbf{b}$, but not the regular dot product, which induces unnecessary difficulty to the development of our test statistic. In this paper, we work directly with the original predictor $\mathbf{X}$, and, as such, use the following symmetric candidate matrices $\mathbf{M}$:

$$\begin{aligned}
\text{SIR:}\quad & \mathbf{M} = \boldsymbol{\Sigma}^{-1}\mathrm{Var}\{\mathrm{E}(\mathbf{X}|Y)\}\boldsymbol{\Sigma}^{-1};\\
\text{SAVE:}\quad & \mathbf{M} = \boldsymbol{\Sigma}^{-1}\mathrm{E}\{\boldsymbol{\Sigma} - \mathrm{Var}(\mathbf{X}|Y)\}^2\boldsymbol{\Sigma}^{-1};\\
\text{DR:}\quad & \mathbf{M} = \boldsymbol{\Sigma}^{-1}\mathrm{E}\{2\boldsymbol{\Sigma} - \mathrm{E}\big((\widetilde{\mathbf{X}} - \mathbf{X})(\widetilde{\mathbf{X}} - \mathbf{X})^T|Y, \widetilde{Y}\big)\}^2\boldsymbol{\Sigma}^{-1},
\end{aligned}$$

where $(\widetilde{Y}, \widetilde{\mathbf{X}})$ is an independent copy of $(Y, \mathbf{X})$. The eigenvectors $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d$ corresponding to the first d nonzero eigenvalues of $\mathbf{M}$ form a basis of $\mathcal{S}_{Y|\mathbf{X}}$, and are orthonormal with respect to the standard inner-product. The corresponding sample version of $\mathbf{M}$, $\widehat{\mathbf{M}}$ can then be spectrally decomposed to obtain an estimate of span$(\boldsymbol{\beta})$. Notice that these symmetric candidate matrices are not exactly the same as those traditionally used in the sufficient dimension reduction literature. Their symmetry facilitates the derivation of the asymptotic distribution of our test statistic. Interested readers may refer to [16,17] for further details.

3. A link-free test for common indices

Throughout this article, we assume that Model (1.2) holds for the two populations under consideration. Let $(Y_j^g, \mathbf{X}_j^g)$, $j = 1, \ldots, n_g$ be a simple random sample of size n_g from the gth population $(Y^g, \mathbf{X}^g)$, $g = 1, 2$. Also, let $\bar{\mathbf{X}}_g = \frac{1}{n_g}\sum_{i=1}^{n_g} \mathbf{X}_i^g$, and $\widehat{\boldsymbol{\Sigma}}_g = \frac{1}{n_g}\sum_{i=1}^{n_g}(\mathbf{X}_i^g - \bar{\mathbf{X}}_g)(\mathbf{X}_i^g - \bar{\mathbf{X}}_g)^T$, $g = 1, 2$. Let $\mathbf{M}_g$ denote the method-specific candidate matrix for the gth population, $\lambda_{g1} \geq \lambda_{g2} \cdots \geq \lambda_{gd} > \lambda_{g,(d+1)} = \cdots = \lambda_{gp} = 0$ be the eigenvalues of $\mathbf{M}_g$, and $\boldsymbol{\eta}_{gi}$ be the normalized eigenvector corresponding to λ_{gi}. Define eigenprojections $\mathbf{P}_{gd} = \boldsymbol{\eta}_{g1}\boldsymbol{\eta}_{g1}^T + \cdots + \boldsymbol{\eta}_{gd}\boldsymbol{\eta}_{gd}^T$, $\mathbf{Q}_{gd} = \mathbf{I}_p - \mathbf{P}_{gd}$, and let $\hat{\boldsymbol{\eta}}_{gi}$ denote the corresponding sample version of $\boldsymbol{\eta}_{gi}$, then $\mathbf{P}_{gd}$ can be estimated by $\widehat{\mathbf{P}}_{gd} = \hat{\boldsymbol{\eta}}_{g1}\hat{\boldsymbol{\eta}}_{g1}^T + \cdots + \hat{\boldsymbol{\eta}}_{gd}\hat{\boldsymbol{\eta}}_{gd}^T$. Let $\mathbf{A}^+$ denote the generalized inverse of matrix $\mathbf{A}$, from the perturbation theory [14,23], we then have

$$\widehat{\mathbf{P}}_{gd} = \mathbf{P}_{gd} + \sum_{i=1}^{d}\left[\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}^T\mathbf{A}_g(\lambda_{gi}\mathbf{I} - \mathbf{M}_g)^+ + (\lambda_{gi}\mathbf{I} - \mathbf{M}_g)^+\mathbf{A}_g\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}^T\right] + o_p\left(n_g^{-\frac{1}{2}}\right)$$

where $\mathbf{A}_g = \widehat{\mathbf{M}}_g - \mathbf{M}_g$.

Note that the approach we take is similar to that in Yu, Zhu and Wen [30] but is motivated by a set of methodologies developed in Schott [22,20,21] for making inference about common principal component subspaces.

Let $W = \text{trace}(\mathbf{P}_{1d}\mathbf{Q}_{2d}\mathbf{P}_{1d})$, we have the following proposition:

Proposition 1. *Assume that the data $(\mathbf{X}_j^g, Y_j^g)$, for $j = 1, \ldots, n_g$, $g = 1, 2$, are a simple random sample from $(\mathbf{X}^g, Y^g)$ with finite fourth order moments, then the null hypothesis (1.3) is true if and only if $W = 0$.*

Proof.

$$\begin{aligned} W = 0 &\iff \text{trace}(\mathbf{P}_{1d}\mathbf{Q}_{2d}\mathbf{Q}_{2d}\mathbf{P}_{1d}) = 0 \\ &\iff \text{trace}((\mathbf{P}_{1d}\mathbf{Q}_{2d})(\mathbf{P}_{1d}\mathbf{Q}_{2d})^T) = 0 \\ &\iff \mathbf{P}_{1d}\mathbf{Q}_{2d} = 0 \end{aligned}$$

$\mathbf{P}_{1d}\mathbf{Q}_{2d} = 0$ implies that span$(\mathbf{P}_{1d}) \subseteq$ span$(\mathbf{P}_{2d})$. Because span$(\mathbf{P}_{1d})$ and span$(\mathbf{P}_{2d})$ have common dimension d, $\mathbf{P}_{1d} = \mathbf{P}_{2d}$ and hence (1.3) is true.

On the other hand, span$(\boldsymbol{\beta})$ = span$(\boldsymbol{\eta})$ if and only if $\mathbf{P}_{1d} = \mathbf{P}_{2d}$, which (1.3) implies that $\mathbf{P}_{1d}\mathbf{Q}_{2d} = 0$, and hence $W = 0$. □

We consider $\widehat{\mathbf{P}}_{1d}\widehat{\mathbf{Q}}_{2d}$, where $\widehat{\mathbf{Q}}_{2d} = \mathbf{I}_p - \widehat{\mathbf{P}}_{2d} = \sum_{l=d+1}^{p}\hat{\boldsymbol{\eta}}_{gl}\hat{\boldsymbol{\eta}}_{gl}^T$, and let $T = n\widehat{W}$, where $\widehat{W} = \text{trace}(\widehat{\mathbf{P}}_{1d}\widehat{\mathbf{Q}}_{2d}\widehat{\mathbf{P}}_{1d})$. Let $n = n_1 + n_2$, $a_1 = \frac{n}{n_1}$, $a_2 = \frac{n}{n_2}$. As Yu et al. [30] pointed out, $\mathbf{A}_g = \widehat{\mathbf{M}}_g - \mathbf{M}_g$ can be expressed via the influence function approach as: $\mathbf{A}_g = \widehat{\mathbf{M}}_g - \mathbf{M}_g = \text{E}_{n_g}[\mathbf{M}_g^*(\mathbf{X}^g, Y^g)] + o_p(n_g^{-1/2})$, where $\text{E}_n\{.\} = \frac{1}{n}\sum_{i=1}^{n}\{.\}$. This approach is key to the further development of our test statistic. The explicit formulas for $\mathbf{M}_g^*$ in the modified SIR, SAVE and DR are given in the Appendix. Let vec$(\mathbf{A})$ denote the operator which stacks the columns of matrix $\mathbf{A}$ to form a vector, the following lemma gives the asymptotic distribution of $\sqrt{n}\,\text{vec}(\widehat{\mathbf{P}}_{1d}\widehat{\mathbf{Q}}_{2d})$.

Lemma 1. *Assume that the data $(\mathbf{X}_i^g, Y_i^g)$, for $i = 1, \ldots, n_g$, are a simple random sample from $(\mathbf{X}^g, Y^g)$ with finite fourth order moments, then under null hypothesis (1.3), we have*

$$\sqrt{n}\,\text{vec}(\widehat{\mathbf{P}}_{1d}\widehat{\mathbf{Q}}_{2d}) \xrightarrow{D} N(0, \boldsymbol{\Psi}),$$

where $\boldsymbol{\Psi} = a_1\boldsymbol{\Psi}_1 + a_2\boldsymbol{\Psi}_2$, $\boldsymbol{\Psi}_g = \mathbf{U}_g\boldsymbol{\Phi}_g\mathbf{U}_g^T$, $\boldsymbol{\Phi}_g = \text{E}\{\text{vec}(\mathbf{M}_g^*(\mathbf{X}^g, Y^g))\,\text{vec}(\mathbf{M}_g^*(\mathbf{X}^g, Y^g))^T\}$ is the asymptotic covariance matrix of $\sqrt{n_g}\,\text{vec}(\mathbf{A}_g)$, and $\mathbf{U}_g = \sum_{i=1}^{d}\sum_{k=d+1}^{p}\lambda_{gi}^{-1}(\boldsymbol{\eta}_{gk}\boldsymbol{\eta}_{gk}{}^T) \otimes (\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}{}^T)$.

Proof. Let $\mathbf{R}_g = \sum_{i=1}^{d}[\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}^T\mathbf{A}_g(\lambda_{gi}\mathbf{I} - \mathbf{M}_g)^+ + (\lambda_{gi}\mathbf{I} - \mathbf{M}_g)^+\mathbf{A}_g\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}^T]$ for $g = 1, 2$.

Therefore

$$\widehat{\mathbf{P}}_{1d} = \mathbf{P}_{1d} + \mathbf{R}_1 + o_p\left(n^{-\frac{1}{2}}\right)$$

$$\widehat{\mathbf{P}}_{2d} = \mathbf{P}_{2d} + \mathbf{R}_2 + o_p\left(n^{-\frac{1}{2}}\right)$$

and

$$\begin{aligned}\mathrm{vec}(\widehat{\mathbf{P}}_{1d}\widehat{\mathbf{Q}}_{2d}) &= \mathrm{vec}((\mathbf{P}_{1d}+\mathbf{R}_1)(\mathbf{Q}_{2d}-\mathbf{R}_2)) + o_p\left(n^{-\frac{1}{2}}\right)\\ &= \mathrm{vec}(\mathbf{P}_{1d}\mathbf{Q}_{2d} - \mathbf{P}_{1d}\mathbf{R}_2 + \mathbf{R}_1\mathbf{Q}_{2d}) + o_p\left(n^{-\frac{1}{2}}\right).\end{aligned}$$

According to the null hypothesis (1.3), $\mathbf{P}_{1d} = \mathbf{P}_{2d}$, hence $\mathbf{P}_{1d}\mathbf{Q}_{2d} = 0$. Then

$$\mathrm{vec}(\widehat{\mathbf{P}}_{1d}\widehat{\mathbf{Q}}_{2d}) = \mathrm{vec}(-\mathbf{P}_{1d}\mathbf{R}_2 + \mathbf{R}_1\mathbf{Q}_{2d}) + o_p\left(n^{-\frac{1}{2}}\right).$$

Since $(\lambda_{gi}\mathbf{I} - \mathbf{M}_g)^+ = \sum_{k=1,\lambda_{gk}\neq\lambda_{gi}}^{p}(\lambda_{gi}-\lambda_{gk})^{-1}\boldsymbol{\eta}_{gk}\boldsymbol{\eta}_{gk}^T$ and $\lambda_{gk} = 0$ when $k \geq d+1$, we have

$$\mathbf{R}_g = \sum_{i=1}^{d}\sum_{k=d+1}^{p}\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}^T\mathbf{A}_g\lambda_{gi}^{-1}\boldsymbol{\eta}_{gk}\boldsymbol{\eta}_{gk}^T + \sum_{i=1}^{d}\sum_{k=d+1}^{p}\boldsymbol{\eta}_{gk}\boldsymbol{\eta}_{gk}^T\mathbf{A}_g\lambda_{gi}^{-1}\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}^T.$$

Hence,

$$\begin{aligned}\mathrm{vec}\big(\widehat{\mathbf{P}}_{1d}\widehat{\mathbf{Q}}_{2d}\big) &= \mathrm{vec}\Big(-\sum_{i=1}^{d}\sum_{k=d+1}^{p}\boldsymbol{\eta}_{2i}{\boldsymbol{\eta}_{2i}}^T\mathbf{A}_2\lambda_{2i}^{-1}\boldsymbol{\eta}_{2k}{\boldsymbol{\eta}_{2k}}^T + \sum_{i=1}^{d}\sum_{k=d+1}^{p}\boldsymbol{\eta}_{1i}{\boldsymbol{\eta}_{1i}}^T\mathbf{A}_1\lambda_{1i}^{-1}\boldsymbol{\eta}_{1k}{\boldsymbol{\eta}_{1k}}^T\Big)\\ &= \sum_{i=1}^{d}\sum_{k=d+1}^{p}\lambda_{1i}^{-1}(\boldsymbol{\eta}_{1k}{\boldsymbol{\eta}_{1k}}^T\otimes\boldsymbol{\eta}_{1i}{\boldsymbol{\eta}_{1i}}^T)\,\mathrm{vec}(\mathbf{A}_1) - \sum_{i=1}^{d}\sum_{k=d+1}^{p}\lambda_{2i}^{-1}(\boldsymbol{\eta}_{2k}{\boldsymbol{\eta}_{2k}}^T\otimes\boldsymbol{\eta}_{2i}{\boldsymbol{\eta}_{2i}}^T)\,\mathrm{vec}(\mathbf{A}_2).\end{aligned}$$

By the Central Limit Theorem, we can conclude

$$\sqrt{n_g}\left(\sum_{i=1}^{d}\sum_{k=d+1}^{p}\lambda_{gi}^{-1}(\boldsymbol{\eta}_{gk}{\boldsymbol{\eta}_{gk}}^T\otimes\boldsymbol{\eta}_{gi}{\boldsymbol{\eta}_{gi}}^T)\,\mathrm{vec}(\mathbf{A}_g)\right)\xrightarrow{D}N(0,\boldsymbol{\Psi}_g),$$

so

$$\sqrt{a_g}\sqrt{n_g}\left(\sum_{i=1}^{d}\sum_{k=d+1}^{p}\lambda_{gi}^{-1}(\boldsymbol{\eta}_{gk}{\boldsymbol{\eta}_{gk}}^T\otimes\boldsymbol{\eta}_{gi}{\boldsymbol{\eta}_{gi}}^T)\,\mathrm{vec}(\mathbf{A}_g)\right)\xrightarrow{D}N(0,a_g\boldsymbol{\Psi}_g)$$

that is

$$\sqrt{n}\left(\sum_{i=1}^{d}\sum_{k=d+1}^{p}\lambda_{gi}^{-1}(\boldsymbol{\eta}_{gk}{\boldsymbol{\eta}_{gk}}^T\otimes\boldsymbol{\eta}_{gi}{\boldsymbol{\eta}_{gi}}^T)\,\mathrm{vec}(\mathbf{A}_g)\right)\xrightarrow{D}N(0,a_g\boldsymbol{\Psi}_g).$$

Since group 1 and group 2 are independent, we then have:

$$\sqrt{n}\,\mathrm{vec}(\widehat{\mathbf{P}}_{1d}\widehat{\mathbf{Q}}_{2d})\xrightarrow{D}N(0,\boldsymbol{\Psi}). \quad \square$$

Theorem 1 provides the asymptotic result concerning our test statistic $T = n\widehat{W}$.

Theorem 1. *Assume the conditions of Proposition 1 hold, then under null hypothesis (1.3), we have*

$$T \longrightarrow \sum_{i=1}^{d(p-d)}\omega_i\chi_i^2(1),$$

where $\omega_1 \geq \cdots \geq \omega_{d(p-d)}$ are the eigenvalues of $\boldsymbol{\Psi}$.

Proof.

$$\begin{aligned}T &= n\mathrm{trace}\left(\left(\widehat{\mathbf{P}}_{1d}\widehat{\mathbf{Q}}_{2d}\right)\left(\widehat{\mathbf{P}}_{1d}\widehat{\mathbf{Q}}_{2d}\right)^T\right) = n\,\mathrm{vec}\left(\widehat{\mathbf{P}}_{1d}\widehat{\mathbf{Q}}_{2d}\right)^T\,\mathrm{vec}\left(\widehat{\mathbf{P}}_{1d}\widehat{\mathbf{Q}}_{2d}\right)\\ &= \left(\sqrt{n}\,\mathrm{vec}\left(\widehat{\mathbf{P}}_{1d}\widehat{\mathbf{Q}}_{2d}\right)^T\right)\left(\sqrt{n}\,\mathrm{vec}\left(\widehat{\mathbf{P}}_{1d}\widehat{\mathbf{Q}}_{2d}\right)\right).\end{aligned}$$

By Lemma 1, under null hypothesis (1.3), the conclusion is obvious. $\square$

A consistent estimate of $\boldsymbol{\Psi}$, $\widehat{\boldsymbol{\Psi}}$ can be obtained by substituting sample estimates for the unknown quantities. The weights ω_i's can be consistently estimated using the eigenvalues of $\widehat{\boldsymbol{\Psi}}$. In the simulation studies which follow we compare the observed value of the test statistic T to the percentage points of $\sum_{i=1}^{d(p-d)}\hat{\omega}_i\chi_i^2(1)$ to approximate the p-value of our test. We may also use the modified test statistics proposed by Bentler and Xie [1] to approximate the tail probabilities.

Table 4.1
Estimated Test Levels (in percentages) for Model I with $d = 1$.

Sample size	Model **I** with $p = 4$				Model **I** with $p = 8$			
	Test	Nominal Level (%)			Test	Nominal Level (%)		
		1	5	10		1	5	10
$n_1 = n_2 = 200$	SIR	1.50	5.50	9.30	SIR	1.20	4.50	9.10
	SAVE	0.90	4.60	10.6	SAVE	1.40	5.30	9.60
	DR	1.40	5.30	10.8	DR	1.60	4.60	9.50
$n_1 = n_2 = 400$	SIR	0.80	4.60	9.40	SIR	1.30	4.50	10.5
	SAVE	0.80	4.70	10.4	SAVE	1.40	5.50	9.50
	DR	0.80	5.20	9.70	DR	0.70	4.70	10.3
$n_1 = n_2 = 600$	SIR	1.10	4.90	10.3	SIR	0.90	5.20	9.80
	SAVE	0.90	5.20	9.90	SAVE	1.10	4.90	10.1
	DR	1.10	5.00	9.80	DR	1.20	5.10	10.1

Table 4.2
Estimated Test Levels (in percentages) for Model II.

Sample size	Model **II** with $p = 4$				Model **II** with $p = 8$			
	Test	Nominal Level (%)			Test	Nominal Level (%)		
		1	5	10		1	5	10
$n_1 = n_2 = 200$	SIR	1.20	4.70	10.4	SIR	0.80	5.40	9.60
	SAVE	0.80	4.50	9.50	SAVE	1.30	5.50	10.9
	DR	0.90	5.40	9.50	DR	0.80	4.70	10.3
$n_1 = n_2 = 400$	SIR	1.20	4.60	10.3	SIR	1.40	4.50	9.70
	SAVE	1.30	5.20	9.70	SAVE	0.80	4.70	9.70
	DR	0.80	4.70	10.3	DR	1.30	4.60	9.70
$n_1 = n_2 = 600$	SIR	1.10	5.20	9.90	SIR	0.90	4.90	10.1
	SAVE	1.00	4.90	9.80	SAVE	1.20	4.80	9.80
	DR	1.00	5.00	9.80	DR	0.90	5.10	10.1

4. Numerical studies

4.1. Simulation studies

Throughout our simulation studies, the random error ϵ is assumed to be standard normal and is independent of $\mathbf{X}$. The dimension of the predictor vector p is taken to be 4 and 8, the number of slices $h = 4$. We summarize the results over 1000 replications for each simulation study. We compare the performance of our proposed tests among the three sufficient dimension reduction methods with different choices of n and p.

4.1.1. Estimated test levels

In this subsection, we evaluate the performance of our test statistic under three different models when the null hypothesis (1.3) holds.

Model I: We first consider the following model with one dimensional structure for both groups. The predictor vector $\mathbf{X} = (X_1, \ldots, X_p)$ is generated from standard multivariate normal.

$$Y = \begin{cases} \exp(X_1 + X_2 + X_3) + \epsilon_1, & \text{for group 1;} \\ 10\sin(X_1 + X_2 + X_3) + \epsilon_2, & \text{for group 2.} \end{cases}$$

Table 4.1 shows the estimated test levels via our test statistics. As the group sizes n_1 and n_2 increase, the estimated levels are getting closer to the nominal levels. For example, when $p = 4$ and the nominal level is 1%, the estimated levels for modified SIR are 1.5%, 0.8% and 1.1% respectively for sample sizes 200, 400 and 600. Also it is not a surprise that the performance of our tests slightly deteriorates as p increases. All three dimension reduction methods perform reasonably well for all combinations of p and n.

Model II: In this model, the predictor vector $\mathbf{X} = (X_1, \ldots, X_p)$ follows a multivariate normal distribution with mean 0, and the correlation between X_i and X_j is $0.5^{|i-j|}$, $i = 1, \ldots, p; j = 1, \ldots, p$. The two groups share common indices and $d = 1$.

$$Y = \begin{cases} \exp(2X_1 + X_2) + \epsilon_1, & \text{for group 1;} \\ X_1 + 0.5X_2 + \epsilon_2, & \text{for group 2.} \end{cases}$$

Table 4.2 presents the estimated significance levels for Model II. Even though the components of independent variables are correlated, significance levels are still close to nominal levels which means our methods work well for models with correlated predictors. When $n_1 = n_2 = 600$, the test method based on modified DR performs the best.

Table 4.3
Estimated Test Levels (in percentages) for Model III.

Sample size	Model **III** with $p = 4$				Model **III** with $p = 8$			
	Test	Nominal Level (%)			Test	Nominal Level (%)		
		1	5	10		1	5	10
$n_1 = n_2 = 200$	SIR	1.20	5.40	10.3	SIR	0.80	4.60	10.5
	SAVE	0.80	4.50	9.80	SAVE	0.80	5.30	9.60
	DR	0.70	4.70	10.5	DR	1.30	4.60	10.8
$n_1 = n_2 = 400$	SIR	1.20	5.20	10.2	SIR	0.80	5.40	10.2
	SAVE	0.80	5.40	10.3	SAVE	1.40	4.70	9.60
	DR	0.70	4.90	9.70	DR	1.20	4.70	9.50
$n_1 = n_2 = 600$	SIR	1.10	4.90	9.90	SIR	0.90	4.90	10.3
	SAVE	0.90	5.10	10.0	SAVE	1.10	4.80	9.70
	DR	1.10	4.90	9.80	DR	1.00	5.20	9.70

Table 4.4
Estimated Test Levels (in percentages) for Model IV.

Sample size	Model **IV** with $p = 4$				Model **IV** with $p = 8$			
	Test	Nominal Level (%)			Test	Nominal Level (%)		
		1	5	10		1	5	10
$n_1 = n_2 = 200$	SIR	0.90	7.50	17.2	SIR	0.80	6.80	18.2
	SAVE	1.40	4.50	9.50	SAVE	0.60	4.20	9.40
	DR	0.70	5.60	9.60	DR	1.60	5.60	11.2
$n_1 = n_2 = 400$	SIR	1.30	6.00	15.4	SIR	0.70	7.80	16.0
	SAVE	0.90	5.30	9.60	SAVE	1.30	5.60	9.60
	DR	0.90	5.30	10.2	DR	0.80	4.70	9.70
$n_1 = n_2 = 600$	SIR	0.90	7.80	14.4	SIR	0.80	8.20	15.8
	SAVE	1.10	4.90	9.80	SAVE	0.90	5.10	10.2
	DR	1.00	4.90	9.90	DR	1.00	5.20	10.1

Model III: We now consider a two-dimensional model as follows:

$$Y = \begin{cases} 1.5(5 + X_1)(2 + X_2) + 0.5\epsilon_1, & \text{for group 1;} \\ 2(1 + X_1)(3 + X_2) + 0.5\epsilon_2, & \text{for group 2.} \end{cases}$$

$X_1 = W, X_2 = V + 0.5W$ where W and V are independent with V drawn from a $t_{(5)}$ distribution and W from a standard exponential distribution. The rest of predictors are iid standard normals. Different versions of this model were also studied by Li [15], Wen and Cook [26] and others. This is a difficult test case for dimension reduction since some predictors are skewed or heavy tailed, and are prone to outliers. As shown in Table 4.3, the performance of our test statistics for all of the three dimension reduction approaches is acceptable.

Model IV: This model considers a one-dimensional model with symmetric structure in **X** which is drawn from standard multivariate normal distribution. Table 4.4 presents the estimated significance levels for Model IV. Because SIR is known to fail when the response surface is symmetric about the origin, the estimated test levels for this model using SIR are relatively far from nominal levels, while test methods with SAVE and DR candidate matrices both perform well.

$$Y = \begin{cases} {X_1}^2 + 1 + 0.5\epsilon_1, & \text{for group 1;} \\ 2{X_1}^2 + \epsilon_2, & \text{for group 2.} \end{cases}$$

4.1.2. Estimated power

We examine the power of our test under the alternative hypothesis in this subsection. Two models are considered. The predictors **X** for both models follow the standard multivariate normal.

Model V:

$$Y = \begin{cases} \exp(X_1 + X_p)\text{sign}(X_2 + X_{p-1}) + 0.5\epsilon_1, & \text{for group 1;} \\ (2X_1 - 3X_p)/(0.5 + (1 + 3X_2 - X_{p-1})^2) + 0.5\epsilon_2, & \text{for group 2.} \end{cases}$$

The two populations in this model have the same structural dimension $d = 2$, however, they do not share the same set of indices. We can see from Table 4.5 that when d is correctly specified as 2, the power of different settings of n and p for our test statistic using the three dimension reduction methods, are all reasonably good. When d is underspecified, unreported simulation studies show that the power of our test is also good.

Model VI:

$$Y = \begin{cases} \exp(X_1 + X_p)\text{sign}(X_2 + X_{p-1}) + 0.5\epsilon_1, & \text{for group 1;} \\ (X_1 + X_p)/(0.5 + (1 + X_3 - X_{p-1})^2) + 0.5\epsilon_2, & \text{for group 2.} \end{cases}$$

Table 4.5
Estimated Power at 5% Nominal Levels for Model V at $d = 2$.

Sample size	SIR		SAVE		DR	
	$p = 4$	$p = 8$	$p = 4$	$p = 8$	$p = 4$	$p = 8$
$n_1 = n_2 = 200$	0.887	0.901	0.910	0.920	0.935	0.960
$n_1 = n_2 = 400$	0.965	0.935	0.980	0.956	0.980	0.996
$n_1 = n_2 = 600$	0.998	0.995	0.993	0.978	0.994	0.998

Table 4.6
Estimated Power at 5% Nominal Levels for Model VI at $d = 2$.

Sample size	SIR		SAVE		DR	
	$p = 4$	$p = 8$	$p = 4$	$p = 8$	$p = 4$	$p = 8$
$n_1 = n_2 = 200$	0.889	0.915	0.932	0.935	0.905	0.925
$n_1 = n_2 = 400$	0.905	0.940	0.935	0.960	0.920	0.950
$n_1 = n_2 = 600$	0.925	0.956	0.968	0.965	0.954	0.940

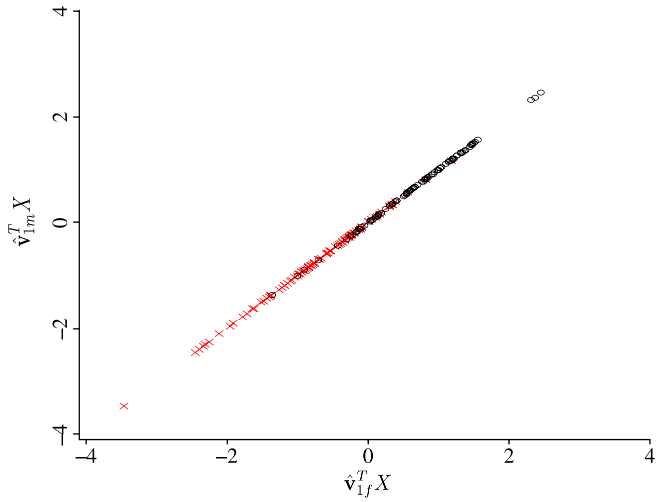

Fig. 4.1. AIS Data: $\hat{\mathbf{v}}_{1m}^T\mathbf{X}$ vs. $\hat{\mathbf{v}}_{1f}^T\mathbf{X}$, $\times$ for females, $\circ$ for males.

Model VI is also two-dimensional. However, in this model, the two populations share one set of common index $X_1 + X_p$, and only differ with respect to the second set of index. Table 4.6 showed the power of our test with $d = 2$. As we can see, the power of our test increases as the sample size n increases. Again, our tests perform reasonably well and are able to detect the different indices between the two populations for most of the time.

4.2. AIS data revisited

We return to the AIS dataset discussed in Section 1. This dataset was originally introduced by Nevill and Holder [19], and studied by Cook and Weisberg [6,7], Chiaromonte, Cook and Li [2] and others. We are interested in investigating how the relationship between the lean body mass L and various predictors including the logarithms of height, weight, red cell count, white cell count and hemoglobin vary across gender. Studies in Chiaromonte, Cook and Li [2] show that there is only one relevant linear combination of predictors for both male and female groups, so a one-dimensional single-index model of the form (1.1) can be applied to both groups. We then conduct our link-free tests via SIR, SAVE and DR to AIS data. All three methods suggest that we cannot reject the null hypothesis (1.3) at significance level 0.05 (p-values are 0.76, 0.54 and 0.65 for SIR, SAVE and DR, respectively).

Our result is consistent with that of Chiaromonte, Cook and Li [2] where the same conclusion was drawn via an informal analysis. Chiaromonte, Cook and Li [2] applied SIR to the conditional regression of L on $\mathbf{X}$ for males and females separately, identifying only one relevant predictor in each group, $\hat{\mathbf{v}}_{1m}^T\mathbf{X}$ and $\hat{\mathbf{v}}_{1f}^T\mathbf{X}$. The sample correlation between the two estimated SIR predictors is 0.96, suggesting that relevant linear combinations for males and females are the same. Fig. 4.1 shows the summary plot of the estimated SIR predictors for males and females.

5. Summary

In this article, we consider a test for common indices in multi-index models with unknown link functions via sufficient dimension reduction approach. Specifically, we focus on testing if two different multi-index models share identical indices. This hypothesis is of particular interest in practice where data from both populations share a common set of explanatory variables. Although Common PCA and partial dimension reduction methods can be adopted to make inference in multi-population dimension reduction problems, they both have drawbacks. Common PCA does not take into account the information of dependent variables, and partial dimension reduction methods focus on obtaining the direct sum of all the conditional central subspaces which could not deal with testing for a set of common indices across the populations.

We propose a link free test via SIR, SAVE and DR. The asymptotic distribution of our test statistic is also derived. Numerical studies indicate that our method works well in practice, both in terms of test level and power, and in particular works best when applied with the DR candidate matrix. Furthermore, we applied our method to the AIS data and found that men and female populations share the same set of common index which is consistent with the work in Chiaromonte, Cook and Li [2]. Research on developing methods for testing problems similar to (1.3) but for more than two groups are under way.

Acknowledgments

The authors would like to thank Dr. James Schott for sending them the technical supplement to his paper [21]. Zhou Yu is supported by NSFC grants (No. 11201151), Program of Shanghai Subject Chief Scientist (14XD1401600) and the 111 Project (B14019).

Appendix

The explicit formulas of candidate matrices for SIR, SAVE and DR are given in this section. It suffices to derive the expansion of $\widehat{\mathbf{M}}_g$ for the gth population. For ease of exposition, we drop the subscript g in the discussion which follows. Also notice that our kernel matrices are different from those used in [30].

Divide the range of Y into h slices $\{J_1, \ldots, J_h\}$. Let $p_k = \mathrm{E}\{I(Y \in J_k)\}$, $\boldsymbol{\mu} = \mathrm{E}(\mathbf{X})$, $\mathbf{U}_k = \mathrm{E}\{(\mathbf{X} - \boldsymbol{\mu})I(Y \in J_k)\}$ and $\mathbf{V}_k = \mathrm{E}\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T I(Y \in J_k)\}$. Denote $\hat{p}_k = \mathrm{E}_n\{I(Y \in J_k)\}$, $\hat{\boldsymbol{\mu}} = E_n(\mathbf{X})$, $\widehat{\mathbf{U}}_k = \mathrm{E}_n\{(\mathbf{X} - \hat{\boldsymbol{\mu}})I(Y \in J_k)\}$ and $\widehat{\mathbf{V}}_k = \mathrm{E}_n\{(\mathbf{X} - \hat{\boldsymbol{\mu}})(\mathbf{X} - \hat{\boldsymbol{\mu}})^T I(Y \in J_k)\}$ be the corresponding sample estimators.

The following lemma is useful for deriving the asymptotic expansion of $\widehat{\mathbf{M}}$.

Lemma 2. *Let* $\boldsymbol{\Sigma}^* = (\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T - \boldsymbol{\Sigma}$, $\widehat{\boldsymbol{\Sigma}}^{*-1} = -\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}^*\widehat{\boldsymbol{\Sigma}}^{-1}$, $\boldsymbol{\mu}^* = \mathbf{X} - \boldsymbol{\mu}$, $p_k^* = I(Y \in J_k) - p_k$, $\mathbf{U}_k^* = \mathbf{X}I(Y \in J_k) - \mathbf{U}_k - \mathbf{X}p_k - I(Y \in J_k) + p_k$, $\mathbf{V}_k^* = \mathbf{X}\mathbf{X}^T I(Y \in J_k) - \mathrm{E}\left[\mathbf{X}\mathbf{X}^T I(Y \in J_k)\right] - \mathrm{E}\left[\mathbf{X}I(Y \in J_k)\right]\mathbf{X}^T - (\mathbf{X}I(Y \in J_k) - 2\mathrm{E}\left[\mathbf{X}I(Y \in J_k)\right])\boldsymbol{\mu}^T - \mathbf{X}\mathrm{E}\left[\mathbf{X}^T I(Y \in J_k)\right] - \boldsymbol{\mu}\left(\mathbf{X}^T I(Y \in J_k) - 2\mathrm{E}\left[\mathbf{X}^T I(Y \in J_k)\right]\right) + (\mathbf{X} - \boldsymbol{\mu})\boldsymbol{\mu}^T\mathrm{E}\left[I(Y \in J_k)\right] + \boldsymbol{\mu}(\mathbf{X} - \boldsymbol{\mu})^T\mathrm{E}\left[I(Y \in J_k)\right] + \boldsymbol{\mu}\boldsymbol{\mu}^T(I(Y \in J_k) - \mathrm{E}\left[I(Y \in J_k)\right])$.

Then we have the following expansions:

$$\widehat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \mathrm{E}_n\{\boldsymbol{\Sigma}^*\} + o_p(n^{-\frac{1}{2}});$$
$$\widehat{\boldsymbol{\Sigma}}^{-1} = \boldsymbol{\Sigma}^{-1} + \mathrm{E}_n\{\boldsymbol{\Sigma}^{*-1}\} + o_p(n^{-\frac{1}{2}});$$
$$\hat{\boldsymbol{\mu}} = \boldsymbol{\mu} + \mathrm{E}_n(\boldsymbol{\mu}^*) + o_p(n^{-1/2}); \quad \hat{p}_k = p_k + \mathrm{E}_n(p_k^*) + o_p(n^{-1/2});$$
$$\widehat{\mathbf{U}}_k = \mathbf{U}_k + \mathrm{E}_n(\mathbf{U}_k^*) + o_p(n^{-1/2}); \quad \widehat{\mathbf{V}}_k = \mathbf{V}_k + \mathrm{E}_n(\mathbf{V}_k^*) + o_p(n^{-1/2});$$
$$\hat{p}_k^{-1} = p_k^{-1} - \mathrm{E}_n(p_k^2 p_k^*) + o_p(n^{-1/2}); \quad \hat{p}_k^{-2} = p_k^{-2} - \mathrm{E}_n(2p_k^3 p_k^*) + o_p(n^{-1/2});$$
$$\hat{p}_k^{-3} = p_k^{-3} - \mathrm{E}_n(3p_k^4 p_k^*) + o_p(n^{-1/2}).$$

Proof. Most of these asymptotic expansions can be derived by the Von Mises expansion in combination with Theorem 6.6.30 in [13,18]. Following [18], we use $S^*(F)$ to indicate the Frechet derivative; for example, $\mathrm{E}^*g(X, F)$ denotes the Frechet derivative of $\int g(X, F)dF$. Here we just validate expressions for $\mathbf{U}_k^*$ and $\mathbf{V}_k^*$. Let $R_k = I(Y \in J_k)$, we have:

$$\begin{aligned}
\mathbf{U}_k^* &= \mathrm{E}^*\left[(\mathbf{X} - \boldsymbol{\mu})R_k\right] \\
&= \mathbf{X}R_k - \mathrm{E}\left[\mathbf{X}R_k\right] - \mathbf{X}p_k - \mathrm{E}\left[\mathbf{X}\right](R_k - 2p_k) \\
&= \mathbf{X}R_k - \mathrm{E}\left[(\mathbf{X} - \boldsymbol{\mu} + \boldsymbol{\mu})R_k\right] - \mathbf{X}p_k - \mathrm{E}\left[\mathbf{X}\right](R_k - 2p_k) \\
&= \mathbf{X}R_k - \mathrm{E}\left[(\mathbf{X} - \boldsymbol{\mu})R_k\right] - \boldsymbol{\mu}p_k - \mathbf{X}p_k - \boldsymbol{\mu}R_k + 2\boldsymbol{\mu}p_k \\
&= \mathbf{X}R_k - \mathbf{U}_k - \mathbf{X}p_k - \boldsymbol{\mu}R_k + \boldsymbol{\mu}p_k.
\end{aligned}$$

Also,

$$\mathbf{V}_k^* = \mathrm{E}^*\left[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T R_k\right] = \mathrm{E}^*\left[\mathbf{X}\mathbf{X}^T R_k\right] - \mathrm{E}^*\left[\mathbf{X}\boldsymbol{\mu}^T R_k\right] - \mathrm{E}^*\left[\boldsymbol{\mu}\mathbf{X}^T R_k\right] + \mathrm{E}^*\left[\boldsymbol{\mu}\boldsymbol{\mu}^T R_k\right].$$

Note that

$$\mathrm{E}^*\left[\mathbf{X}\mathbf{X}^T R_k\right] = \mathbf{X}\mathbf{X}^T R_k - \mathrm{E}\left[\mathbf{X}\mathbf{X}^T R_k\right],$$

and $\mathrm{E}^*\left[\mathbf{X}\boldsymbol{\mu}^T R_k\right]$ is the Frechet derivative of

$$\mathrm{E}\left[\mathbf{X}\boldsymbol{\mu}^T R_k\right] = \mathrm{E}\left[\mathbf{X}R_k\right]\boldsymbol{\mu}^T.$$

So,

$$\begin{aligned}\mathrm{E}^*\left[\mathbf{X}\boldsymbol{\mu}^T R_k\right] &= \mathrm{E}\left[\mathbf{X}R_k\right](\mathbf{X}-\boldsymbol{\mu})^T + (\mathbf{X}R_k - \mathrm{E}\left[\mathbf{X}R_k\right])\boldsymbol{\mu}^T \\ &= \mathrm{E}\left[\mathbf{X}R_k\right]\mathbf{X}^T + (\mathbf{X}R_k - 2\mathrm{E}\left[\mathbf{X}R_k\right])\boldsymbol{\mu}^T.\end{aligned}$$

$\mathrm{E}^*\left[\boldsymbol{\mu}\mathbf{X}^T R_k\right]$ is the Frechet derivative of

$$\mathrm{E}\left[\boldsymbol{\mu}\mathbf{X}^T R_k\right] = \boldsymbol{\mu}\mathrm{E}\left[\mathbf{X}^T R_k\right].$$

So it follows that

$$\begin{aligned}\mathrm{E}^*\left[\boldsymbol{\mu}X^T R_k\right] &= (\mathbf{X}-\boldsymbol{\mu})\,\mathrm{E}\left[\mathbf{X}^T R_k\right] + \boldsymbol{\mu}\left(\mathbf{X}^T R_k - \mathrm{E}\left[\mathbf{X}^T R_k\right]\right) \\ &= \mathbf{X}\mathrm{E}\left[\mathbf{X}^T R_k\right] + \boldsymbol{\mu}\left(\mathbf{X}^T R_k - 2\mathrm{E}\left[\mathbf{X}^T R_k\right]\right).\end{aligned}$$

$\mathrm{E}^*\left[\boldsymbol{\mu}\boldsymbol{\mu}^T R_k\right]$ is the Frechet derivative of

$$\mathrm{E}\left[\boldsymbol{\mu}\boldsymbol{\mu}^T R_k\right] = \boldsymbol{\mu}\boldsymbol{\mu}^T\mathrm{E}\left[R_k\right],$$

so that

$$\mathrm{E}^*\left[\boldsymbol{\mu}\boldsymbol{\mu}^T R_k\right] = (\mathbf{X}-\mu)\,\boldsymbol{\mu}^T\mathrm{E}\left[R_k\right] + \boldsymbol{\mu}(\mathbf{X}-\boldsymbol{\mu})^T\mathrm{E}\left[R_k\right] + \boldsymbol{\mu}\boldsymbol{\mu}^T\left(R_k - \mathrm{E}\left[R_k\right]\right).$$

All of this yields

$$\begin{aligned}\mathrm{E}^*\left[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T R_k\right] = {} & \mathbf{X}\mathbf{X}^T R_k - \mathrm{E}\left[\mathbf{X}\mathbf{X}^T R_k\right] - \mathrm{E}\left[\mathbf{X}R_k\right]\mathbf{X}^T - (\mathbf{X}R_k - 2\mathrm{E}\left[\mathbf{X}R_k\right])\boldsymbol{\mu}^T \\ & - \mathbf{X}\mathrm{E}\left[\mathbf{X}^T R_k\right] - \boldsymbol{\mu}\left(\mathbf{X}^T R_k - 2\mathrm{E}\left[\mathbf{X}^T R_k\right]\right) \\ & + (\mathbf{X}-\boldsymbol{\mu})\,\boldsymbol{\mu}^T\mathrm{E}\left[R_k\right] + \boldsymbol{\mu}(\mathbf{X}-\boldsymbol{\mu})^T\mathrm{E}\left[R_k\right] + \boldsymbol{\mu}\boldsymbol{\mu}^T\left(R_k - \mathrm{E}\left[R_k\right]\right). \quad \square\end{aligned}$$

Asymptotic expansion of $\widehat{\mathbf{M}}_{SIR}$

Define $\Lambda_{SIR} = \sum_{l=1}^h p_l \mathrm{E}(\mathbf{X}-\boldsymbol{\mu}|Y \in J_l)\{\mathrm{E}(\mathbf{X}-\boldsymbol{\mu}|Y \in J_l)\}^T = \sum_{l=1}^h p_l^{-1}\mathbf{U}_l\mathbf{U}_l^T$. Then $\mathbf{M}_{SIR} = \boldsymbol{\Sigma}^{-1}\Lambda_{SIR}\boldsymbol{\Sigma}^{-1}$. The corresponding sample estimators are $\widehat{\Lambda}_{SIR} = \sum_{l=1}^h \hat{p}_l^{-1}\widehat{\mathbf{U}}_l\widehat{\mathbf{U}}_l^T$ and $\widehat{\mathbf{M}}_{SIR} = \widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\Lambda}_{SIR}\widehat{\boldsymbol{\Sigma}}^{-1}$. The explicit expansion forms of $\widehat{\Lambda}_{SIR}$ and $\widehat{\mathbf{M}}_{SIR}$ are given in the following.

Lemma 3. *Let* $\Lambda_{SIR}^* = \sum_{l=1}^h\left(-\frac{p_l^*\mathbf{U}_l\mathbf{U}_l^T}{p_l^2} + \frac{\mathbf{U}_l^*\mathbf{U}_l^T}{p_l} + \frac{\mathbf{U}_l\mathbf{U}_l^{*T}}{p_l}\right)$, *then we have the expansion* $\widehat{\Lambda}_{SIR} = \Lambda_{SIR} + E_n(\Lambda_{SIR}^*) + o_p(n^{-1/2})$.

Theorem 2. $\widehat{\mathbf{M}}_{SIR}$ *can be expanded asymptotically as* $\widehat{\mathbf{M}}_{SIR} = \mathbf{M}_{SIR} + E_n(\mathbf{M}_{SIR}^*) + o_p(n^{-1/2})$, *where* $\mathbf{M}_{SIR}^* = \boldsymbol{\Sigma}^{*-1}\Lambda_{SIR}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\Lambda_{SIR}^*\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\Lambda_{SIR}\boldsymbol{\Sigma}^{*-1}$.

Proof of Theorem 2. With the expansion given in Lemma 2, the conclusion can be easily derived by invoking Lemma 3. $\square$

Asymptotic expansion of $\widehat{\mathbf{M}}_{SAVE}$

Let $\Lambda_{SAVE} = \mathrm{E}\{\boldsymbol{\Sigma} - \mathrm{Var}(\mathbf{X}|\delta(Y))\}^2$, where $\delta(Y) = \sum_{l=1}^h lI(Y \in J_l)$. Then $\mathbf{M}_{SAVE} = \boldsymbol{\Sigma}^{-1}\Lambda_{SAVE}\boldsymbol{\Sigma}^{-1}$.

Lemma 4. $\Lambda_{SAVE} = \boldsymbol{\Sigma}\Lambda_{SIR} + \Lambda_{SIR}\boldsymbol{\Sigma} - \boldsymbol{\Sigma}^2 + \boldsymbol{\Gamma}$, *where* $\boldsymbol{\Gamma} = \sum_{l=1}^h(\boldsymbol{\Gamma}_l^1 - \boldsymbol{\Gamma}_l^2 - \boldsymbol{\Gamma}_l^3 + \boldsymbol{\Gamma}_l^4)$ *with* $\boldsymbol{\Gamma}_l^1 = \frac{\mathbf{V}_l^2}{p_l}$, $\boldsymbol{\Gamma}_l^2 = \frac{\mathbf{V}_l\mathbf{U}_l\mathbf{U}_l^T}{p_l^2}$, $\boldsymbol{\Gamma}_l^3 = \frac{\mathbf{U}_l\mathbf{U}_l^T\mathbf{V}_l}{p_l^2}$ *and* $\boldsymbol{\Gamma}_l^4 = \frac{\mathbf{U}_l\mathbf{U}_l^T\mathbf{U}_l\mathbf{U}_l^T}{p_l^3}$.

Proof of Lemma 4.

$$\begin{aligned}\Lambda_{SAVE} &= \boldsymbol{\Sigma}^2 - \boldsymbol{\Sigma}\mathrm{E}[\mathrm{Var}(\mathbf{X}|Y)] - \mathrm{E}[\mathrm{Var}(\mathbf{X}|Y)]\boldsymbol{\Sigma} + \mathrm{E}[\mathrm{Var}(\mathbf{X}|Y)^2] \\ &= \boldsymbol{\Sigma}^2 - \boldsymbol{\Sigma}(\boldsymbol{\Sigma} - \Lambda_{SIR}) - (\boldsymbol{\Sigma} - \Lambda_{SIR})\boldsymbol{\Sigma} + \sum_{l=1}^h p_l\left(\frac{\mathbf{V}_l}{p_l} - \frac{\mathbf{U}_l\mathbf{U}_l^T}{{p_l}^2}\right)^2.\end{aligned}$$

With more algebraic calculations, one can easily derive the stated result. $\square$

Let $\widehat{\boldsymbol{\Gamma}}_l^1$, $\widehat{\boldsymbol{\Gamma}}_l^1$, $\widehat{\boldsymbol{\Gamma}}_l^1$, $\widehat{\boldsymbol{\Gamma}}_l^4$ and $\widehat{\Lambda}_{SAVE}$ be the sample estimators of $\boldsymbol{\Gamma}_l^1$, $\boldsymbol{\Gamma}_l^2$, $\boldsymbol{\Gamma}_l^3$, $\boldsymbol{\Gamma}_l^4$ and Λ_{SAVE}, respectively. The associated Frechet derivatives are

$$\begin{aligned}
(\boldsymbol{\Gamma}_l^1)^* &= -\frac{p_l^*\mathbf{V}_l^2}{p_l^2} + \frac{\mathbf{V}_l^*\mathbf{V}_l}{p_l} + \frac{\mathbf{V}_l\mathbf{V}_l^*}{p_l}, \\
(\boldsymbol{\Gamma}_l^2)^* &= -2\frac{p_l^*\mathbf{V}_l\mathbf{U}_l\mathbf{U}_l^T}{p_l^3} + \frac{\mathbf{V}_l^*\mathbf{U}_l\mathbf{U}_l^T}{p_l^2} + \frac{\mathbf{V}_l\mathbf{U}_l^*\mathbf{U}_l^T}{p_l^2} + \frac{\mathbf{V}_l\mathbf{U}_l\mathbf{U}_l^{*T}}{p_l^2}, \\
(\boldsymbol{\Gamma}_l^3)^* &= -2\frac{p_l^*\mathbf{U}_l\mathbf{U}_l^T\mathbf{V}_l}{p_l^3} + \frac{\mathbf{U}_l^*\mathbf{U}_l^T\mathbf{V}_l}{p_l^2} + \frac{\mathbf{U}_l\mathbf{U}_l^{*T}\mathbf{V}_l}{p_l^2} + \frac{\mathbf{U}_l\mathbf{U}_l^T\mathbf{V}_l^*}{p_l^2}, \\
(\boldsymbol{\Gamma}_l^4)^* &= -3\frac{p_l^*\mathbf{U}_l\mathbf{U}_l^T\mathbf{U}_l\mathbf{U}_l^T}{p_l^4} + \frac{\mathbf{U}_l^*\mathbf{U}_l^T\mathbf{U}_l\mathbf{U}_l^T}{p_l^3} + \frac{\mathbf{U}_l\mathbf{U}_l^{*T}\mathbf{U}_l\mathbf{U}_l^T}{p_l^3} + \frac{\mathbf{U}_l\mathbf{U}_l^T\mathbf{U}_l^*\mathbf{U}_l^T}{p_l^3} + \frac{\mathbf{U}_l\mathbf{U}_l^T\mathbf{U}_l\mathbf{U}_l^{*T}}{p_l^3}.
\end{aligned}$$

Lemma 5. *Let* $\boldsymbol{\Gamma}^* = \sum_{l=1}^h\{(\boldsymbol{\Gamma}_l^1)^* - (\boldsymbol{\Gamma}_l^2)^* - (\boldsymbol{\Gamma}_l^3)^* + (\boldsymbol{\Gamma}_l^4)^*\}$ *and* $\Lambda_{SAVE}^* = \boldsymbol{\Sigma}\Lambda_{SIR}^* + \boldsymbol{\Sigma}^*\Lambda_{SIR} + \Lambda_{SIR}\boldsymbol{\Sigma}^* + \Lambda_{SIR}^*\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\boldsymbol{\Sigma}^* - \boldsymbol{\Sigma}^*\boldsymbol{\Sigma} + \boldsymbol{\Gamma}^*$. *Then we have* $\widehat{\Lambda}_{SAVE} = \Lambda_{SAVE} + \mathrm{E}_n(\Lambda_{SAVE}^*) + o_p(n^{-1/2})$.

Proof of Lemma 5. The conclusion can be derived by Lemmas 2–4. Details are omitted. □

Theorem 3. $\widehat{\mathbf{M}}_{SAVE}$ *can be expanded asymptotically as*

$$\widehat{\mathbf{M}}_{SAVE} = \mathbf{M}_{SAVE} + \mathrm{E}_n(\mathbf{M}_{SAVE}^*) + o_p(n^{-1/2}),$$

where $\mathbf{M}_{SAVE}^* = \boldsymbol{\Sigma}^{*-1}\Lambda_{SAVE}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\Lambda_{SAVE}^*\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\Lambda_{SAVE}\boldsymbol{\Sigma}^{*-1}$.

Proof of Theorem 3. With the expansion in Lemma 2, the conclusion can be easily derived by invoking Lemma 5. □

Asymptotic expansion of $\widehat{\mathbf{M}}_{DR}$

The candidate matrix of directional regression is

$$\mathbf{M}_{DR} = \boldsymbol{\Sigma}^{-1}\left\{2\sum_{l=1}^h p_l\left(\frac{\mathbf{V}_l}{p_l} - \boldsymbol{\Sigma}\right)^2 + 2\left(\sum_{l=1}^h \frac{\mathbf{U}_l\mathbf{U}_l^T}{p_l}\right)^2 + 2\left(\sum_{l=1}^h \frac{\mathbf{U}_l^T\mathbf{U}_l}{p_l}\right)\left(\sum_{l=1}^h \frac{\mathbf{U}_l\mathbf{U}_l^T}{p_l}\right)\right\}\boldsymbol{\Sigma}^{-1}.$$

We first rewrite $\mathbf{M}_{DR}$ as given in the following lemma.

Lemma 6. $\mathbf{M}_{DR}$ *can be reformulated as* $\mathbf{M}_{DR} = \boldsymbol{\Sigma}^{-1}\Lambda_{DR}\boldsymbol{\Sigma}^{-1}$, *where*

$$\Lambda_{DR} = 2\sum_{l=1}^h \boldsymbol{\Gamma}_l^1 - 2\boldsymbol{\Sigma}^2 + 2\left(\sum_{l=1}^h \frac{\mathbf{U}_l\mathbf{U}_l^T}{p_l}\right)^2 + 2\left(\sum_{l=1}^h \frac{\mathbf{U}_l^T\mathbf{U}_l}{p_l}\right)\left(\sum_{l=1}^h \frac{\mathbf{U}_l\mathbf{U}_l^T}{p_l}\right).$$

Proof of Lemma 6. The conclusion can be derived by further algebraic calculations. We omit the details here. □

Let $\widehat{\Lambda}_{DR}$ and $\widehat{\mathbf{M}}_{DR}$ be the sample estimators of Λ_{DR} and $\mathbf{M}_{DR}$ respectively.

Lemma 7. *Define*

$$\begin{aligned}
\Lambda_{DR}^* = {} & 2\sum_{l=1}^h \boldsymbol{\Gamma}_l^{1*} - 2\boldsymbol{\Sigma}\boldsymbol{\Sigma}^* - 2\boldsymbol{\Sigma}^*\boldsymbol{\Sigma} - 2\sum_{l=1}^h\sum_{k=1}^h \frac{p_l^*\mathbf{U}_l\mathbf{U}_l^T\mathbf{U}_k\mathbf{U}_k^T}{p_l^2 p_k} - 2\sum_{l=1}^h\sum_{k=1}^h \frac{p_k^*\mathbf{U}_l\mathbf{U}_l^T\mathbf{U}_k\mathbf{U}_k^T}{p_l p_k^2} \\
& + 2\sum_{l=1}^h\sum_{k=1}^h \frac{\mathbf{U}_l^*\mathbf{U}_l^T\mathbf{U}_k\mathbf{U}_k^T}{p_l p_k} + 2\sum_{l=1}^h\sum_{k=1}^h \frac{\mathbf{U}_l\mathbf{U}_l^{*T}\mathbf{U}_k\mathbf{U}_k^T}{p_l p_k} + 2\sum_{l=1}^h\sum_{k=1}^h \frac{\mathbf{U}_l\mathbf{U}_l^T\mathbf{U}_k^*\mathbf{U}_k^T}{p_l p_k} \\
& + 2\sum_{l=1}^h\sum_{k=1}^h \frac{\mathbf{U}_l\mathbf{U}_l^T\mathbf{U}_k\mathbf{U}_k^{*T}}{p_l p_k} - 2\sum_{l=1}^h\sum_{k=1}^h \frac{p_l^*\mathbf{U}_l^T\mathbf{U}_l\mathbf{U}_k\mathbf{U}_k^T}{p_l^2 p_k} - 2\sum_{l=1}^h\sum_{k=1}^h \frac{p_k^*\mathbf{U}_l^T\mathbf{U}_l\mathbf{U}_k\mathbf{U}_k^T}{p_l p_k^2} \\
& + 2\sum_{l=1}^h\sum_{k=1}^h \frac{\mathbf{U}_l^{*T}\mathbf{U}_l\mathbf{U}_k\mathbf{U}_k^T}{p_l p_k} + 2\sum_{l=1}^h\sum_{k=1}^h \frac{\mathbf{U}_l^T\mathbf{U}_l^*\mathbf{U}_k\mathbf{U}_k^T}{p_l p_k} + 2\sum_{l=1}^h\sum_{k=1}^h \frac{\mathbf{U}_l^T\mathbf{U}_l\mathbf{U}_k^*\mathbf{U}_k^T}{p_l p_k} \\
& + 2\sum_{l=1}^h\sum_{k=1}^h \frac{\mathbf{U}_l^T\mathbf{U}_l\mathbf{U}_k\mathbf{U}_k^{*T}}{p_l p_k}.
\end{aligned}$$

Then we have the expansion $\widehat{\Lambda}_{DR} = \Lambda_{DR} + \mathrm{E}_n(\Lambda_{DR}^*) + o_p(n^{-1/2})$.

Proof of Lemma 7. The conclusion can be derived by Lemmas 2, 3 and 6. Details are omitted. □

Theorem 4. *$\widehat{\mathbf{M}}_{DR}$ can be expanded asymptotically as*

$$\widehat{\mathbf{M}}_{DR} = \mathbf{M}_{DR} + \mathrm{E}_n(\mathbf{M}^*_{DR}) + o_p(n^{-1/2}),$$

where $\mathbf{M}^*_{DR} = \boldsymbol{\Sigma}^{*-1}\Lambda_{DR}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\Lambda^*_{DR}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\Lambda_{DR}\boldsymbol{\Sigma}^{*-1}$.

Proof of Theorem 4. With the expansion given in Lemma 2, the conclusion can be easily derived by invoking Lemma 7. □

References

[1] P. Bentler, J. Xie, Corrections to test statistics in principal hessian directions, Statist. Probab. Lett. 47 (2000) 381–389.
[2] F. Chiaromonte, R.D. Cook, Bing. Li, Sufficient dimension reduction in regressions with categorical predictors, Ann. Statist. 30 (2002) 475–497.
[3] R.D. Cook, Regression Graphics, Wiley, New York, 1998.
[4] R.D. Cook, B. Forzani, Likelihood-based sufficient dimension reduction, J. Amer. Statist. Assoc. 104 (2009) 197–208.
[5] R.D. Cook, S. Weisberg, Discussion of sliced inverse regression for dimension reduction, J. Amer. Statist. Assoc. 86 (1991) 328–332.
[6] R.D. Cook, S. Weisberg, An Introduction to Regression Graphics, Wiley, New York, 1994.
[7] R.D. Cook, S. Weisberg, Applied Regression including Computing and Graphics, Wiley, New York, 1999.
[8] M.L. Eaton, A characterization of spherical distributions, J. Multivariate Anal. 20 (1986) 272–276.
[9] Z. Feng, L.X. Zhu, An alternating determination optimization approach for an additive multi-index model, Comput. Statist. Data Anal. 56 (2012) 1981–1993.
[10] P. Hall, K.C. Li, On almost linearity of low dimensional projections from high dimensional data, Ann. Statist. 21 (1993) 867–889.
[11] W. Härdle, P. Hall, H. Ichimura, Optimal smoothing in single-index models, Ann. Statist. 21 (1993) 157–178.
[12] W. Härdle, T.M. Stoker, Investigating smooth multiple regression by the method of average derivatives, J. Amer. Statist. Assoc. 84 (1989) 986–995.
[13] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, 1991.
[14] T. Kato, Pertubation Theory for Linear Operators, Springer, Berlin, 1966.
[15] K.C. Li, Sliced inverse regression for dimension reduction (with discussion), J. Amer. Statist. Assoc. 86 (1991) 316–342.
[16] B. Li, Y. Dong, Dimension reduction for nonelliptically distributed predictors, Ann. Statist. 37 (2009) 1272–1298.
[17] B. Li, M.K. Kim, N. Altman, On dimension folding of matrix- or array-valued statistical objects, Ann. Statist. 38 (2010) 1094–1121.
[18] B. Li, S. Wang, On directional regression for dimension reduction, J. Amer. Statist. Assoc. 102 (2007) 997–1008.
[19] A.M. Nevill, R.L. Holder, Body Mass Index; a measure of fatness or leanness? Br. J. Nutr. 73 (1995) 507–516.
[20] J.R. Schott, Some tests for common principal component subspaces in several groups, Biometrika 78 (1991) 771–777.
[21] J.R. Schott, Asymptotics of eigenprojections of correlation Matrices with some applications in principal components analysis, Biometrika 84 (1997) 327–337.
[22] J.R. Schott, Common principal component subspaces in two groups, Biometrika 75 (1988) 229–236.
[23] D.E. Tyler, Asymptotic inferences for eigenvectors, Ann. Statist. 9 (1981) 725–736.
[24] S. Weisberg, Applied Linear Regression, third ed., Wiley, New York, 2005.
[25] X. Wen, R.D. Cook, Optimal sufficient dimension reduction for regressions with categorical predictors, J. Statist. Plann. Inference 137 (2007) 1961–1978.
[26] X. Wen, R.D. Cook, New approaches to model-free dimension reduction for bivariate regression, J. Statist. Plann. Inference 139 (2009) 734–748.
[27] Y. Xia, A multiple-index model and dimension reduction, J. Amer. Statist. Assoc. 103 (2008) 1631–1640.
[28] Y. Xia, H. Tong, W.K. Li, L.X. Zhu, An adaptive estimation of dimension reduction space, J. R. Stat. Soc. Ser. B 64 (2002) 363–410.
[29] L.G. Xue, L.X. Zhu, Empirical likelihood for single-index model, J. Multivariate Anal. 97 (2006) 1295–1312.
[30] Z. Yu, L.X. Zhu, X. Wen, On model-free conditional coordinate tests for regressions, J. Multivariate Anal. 109 (2012) 61–72.