Journal of Multivariate Analysis 153 (2017) 236–245

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

A link-free approach for testing common indices for three or more multi-index models

Xuejing Liu, Lei Huo, Xuerong Meggie Wen *, Robert Paige

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

ARTICLE INFO

Article history:
Received 22 January 2016
Available online 8 October 2016

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

Keywords:
Sufficient dimension reduction
Multiple populations
Common principal component analysis
Multi-index model

ABSTRACT

Liu et al. (2015) proposed a novel link-free procedure for testing whether two multi-index models share identical indices via the sufficient dimension reduction approach. However, their method can only be applied to data with two populations. In practice, we often deal with situations where the same variables are being measured on objects from three or more groups, and we would like to know how similar these groups are with respect to some overall features. In this paper, we propose a link-free method which could test if three or more multi-index models share the same indices. The asymptotic properties of our test statistic are developed. Numerical studies and a real data analysis are conducted to illustrate the performance of our method.

1. Introduction

Principal component analysis (PCA) is a widely used multivariate statistical analysis technique. Typically, the first several of these principal components account for a large proportion of the total variance of the original p variables. As a result, one may achieve dimension reduction with little loss of information by simply working with those principal components. Such components frequently have interpretations, biological or otherwise, that provide valuable insights into the mechanisms generating the data. However, the standard PCA is normally a one-sample method. In practice, we often deal with situations where the same variables are being measured on objects from different groups, and we would like to know how similar the groups are with respect to some overall features. Various attempts have been made to develop valid analysis for multiple datasets. Flury [10,12] proposed a method called the common PCA, a type of simultaneous principal component analysis for several groups. The common principal components model has been employed in genetics, climatology, ontogeny and other fields; see, e.g., Biok [2]. Flury [11] extended the common PCA to partial common PCA. Other common space models also have been proposed; see, e.g., Krzanowski [15]; Schott [24,25], and Biok [2].

However, for a typical regression problem with a univariate response Y and a p-dimensional random vector predictor $\mathbf{X}$, (common) PCA and other related methods often yield inferior results, when one aims to reduce the dimension of $\mathbf{X}$, since PCA is an unsupervised dimension reduction technique that does not take into account the information in Y. To overcome this drawback, Li [17] and Cook [6] proposed *sufficient dimension reduction* that aims at reducing the dimension of $\mathbf{X}$ while preserving the regression relationship between Y and $\mathbf{X}$. Specifically, the scope of sufficient dimension reduction is to seek a minimal set of linear combinations of $\mathbf{X}$, say $\boldsymbol{\beta}^\top\mathbf{X}$, where $\boldsymbol{\beta}$ is a $p \times d$ matrix with $d \le p$, such that

$$Y \perp\!\!\!\perp \mathbf{X} | \boldsymbol{\beta}^\top\mathbf{X}. \tag{1.1}$$

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

http://dx.doi.org/10.1016/j.jmva.2016.10.002

The column space of $\boldsymbol{\beta}$ is then called a central subspace (Cook [6]), denoted by $\mathcal{S}_{Y|\mathbf{X}}$. Following Li [17] and Liu et al. [19], we can see that the notion of the central subspace is equivalent to assuming the following multi-index model:

$$Y = g(\boldsymbol{\beta}_1^\top \mathbf{X}, \ldots, \boldsymbol{\beta}_d^\top \mathbf{X}; \epsilon), \tag{1.2}$$

where g is an unknown link function, and the random error ϵ is independent with $\mathbf{X}$.

Sufficient dimension reduction has received considerable interests in recent years due to the ubiquity of large high-dimension datasets which are now more readily available than in the past. Many methods have been developed, including sliced inverse regression (SIR; Li [17]), sliced average variance estimation (SAVE; Cook and Weisberg [8]), directional regression (DR; Li and Wang [18]), likelihood acquired directions (LAD; Cook and Forzani [7]). Recently, Lee et al. [16] proposed a nonlinear sufficient dimension reduction which seeks an arbitrary function $\phi : \mathbb{R}^p \to \mathbb{R}^d$ satisfying $Y \perp\!\!\!\perp \mathbf{X} | \phi(\mathbf{X})$, which greatly generalizes condition (1.1). Ma and Zhu [20,21] investigated sufficient dimension reduction via a semiparametric approach.

Liu et al. [19] considered testing if the following two multi-index models share the same indices. Specifically, for the d-dimensional multi-index models for two populations (groups):

$$\begin{aligned} Y &= g_1(\boldsymbol{\beta}_1^\top \mathbf{X}, \ldots, \boldsymbol{\beta}_d^\top \mathbf{X}; \epsilon_1) \quad \text{for group 1},\\ Y &= g_2(\boldsymbol{\xi}_1^\top \mathbf{X}, \ldots, \boldsymbol{\xi}_d^\top \mathbf{X}; \epsilon_2) \quad \text{for group 2}, \end{aligned}$$

they proposed a link-free procedure for testing if

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

where both $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$ are $p \times d$ matrices. However, their method can only be applied to data with two populations. In practice, we often deal with datasets which naturally fall into three or more groups as illustrated by the plasma data example in Section 5.

In this article, we focus on testing the hypothesis that if three or more multi-index models share identical indices. For easy of exposition, we adopt the notion of the central subspace. Let $(Y^g, \mathbf{X}^g)$ be a generic pair of $(Y, \mathbf{X})$ for the gth group, where $g \in \{1, \ldots, G\}$ for some $G \geq 3$. Notice that the null hypothesis (1.3) is equivalent to

$$\mathcal{S}_{Y^1|\mathbf{X}^1} = \mathcal{S}_{Y^2|\mathbf{X}^2}.$$

Hence, to test if three or more multi-index models share the same indices is equivalent to testing if the central subspace of a particular group is the same as that of any other group:

$$\mathcal{H}_0 : \mathcal{S}_{Y^1|\mathbf{X}^1} = \mathcal{S}_{Y^2|\mathbf{X}^2} = \cdots = \mathcal{S}_{Y^G|\mathbf{X}^G}. \tag{1.4}$$

So in this article, we are interested in testing hypothesis of (1.4) against the alternative hypothesis $\mathcal{H}_1 : \neg \mathcal{H}_0$, when $G \geq 3$. If the null hypothesis (1.4) is true, we can then pool the data from all groups together for further inferences, and the resulted estimator will yield greater efficiency since a larger dataset is utilized during the estimation process. In the case that the null hypothesis (1.4) is rejected, it might be worthy to investigate which group differs from the rest. Further study can be conducted via pairwise comparison or graphical analysis.

The rest of this article is organized as follows. In Section 2, we give a quick review of sufficient dimension reduction methods for a single population. In Section 3, we present our test statistics for testing (1.4) for $G \geq 3$. The asymptotic distributions of our test statistics are also discussed. We illustrate the performances of our methods via simulation studies in Section 4. We then apply our method to a real dataset. Brief conclusions and a discussion on the future research directions are given in Section 5.

2. Sufficient dimension reduction for a single population

For the gth group, let $\boldsymbol{\mu}_g = \mathrm{E}(\mathbf{X}^g)$, $\boldsymbol{\Sigma}_g = \mathrm{var}(\mathbf{X}^g)$, and $\mathbf{Z}^g = \boldsymbol{\Sigma}_g^{-1/2}(\mathbf{X}^g - \boldsymbol{\mu}_g)$ be the standardized predictor. In this section, we will drop the subscript g for ease of exposition. As Yu et al. [30] pointed out, many moment based sufficient dimension reduction methods can be formulated as the following eigen-decomposition problem:

$$\forall_{i \in \{1,\ldots,p\}} \quad \mathcal{M} \boldsymbol{\eta}_i = \lambda_i \boldsymbol{\eta}_i, \tag{2.1}$$

where $\mathcal{M}$ is the $\mathbf{Z}$ scale method-specific candidate matrix. Assuming the *linearity condition* holds (Li [17]), 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 $\boldsymbol{\lambda}_1 \geq \cdots \geq \boldsymbol{\lambda}_d$ form a basis of the $\mathbf{Z}$ scale central subspace $\mathcal{S}_{Y|\mathbf{Z}}$. Then, according to the invariance property, $\mathcal{S}_{Y|\mathbf{X}} = \boldsymbol{\Sigma}^{-1/2} \mathcal{S}_{Y|\mathbf{Z}}$ as described by Cook [6], $\boldsymbol{\beta} = (\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\eta}_1, \ldots, \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\eta}_d)$ forms a basis of $\mathcal{S}_{Y|\mathbf{X}}$.

Although many sufficient dimension reduction methods, including SIR, which is used to implement our method in this paper due to its computational simplicity, took the above approach, i.e., obtain $\mathcal{S}_{Y|\mathbf{X}}$ via $\mathcal{S}_{Y|\mathbf{Z}}$, for testing hypothesis of (1.4), it is easier to work in terms of the original predictors $\mathbf{X}$ to construct our test statistics. Hence, we adopt the kernel matrix proposed by Liu et al. [19] as follows:

$$\mathrm{SIR}: \ \mathcal{M} = \boldsymbol{\Sigma}^{-1} \mathrm{var}\{\mathrm{E}(\mathbf{X}|Y)\} \boldsymbol{\Sigma}^{-1}.$$

We then spectrally decompose $\widehat{\mathcal{M}}$, the corresponding sample version of $\mathcal{M}$, to obtain an estimate of $\mathcal{S}_{Y|\mathbf{X}}$.

3. Sufficient dimension reduction for multiple populations

3.1. Partial dimension reduction

Partial dimension reduction is originally proposed to facilitate dimension reduction in regressions with both continuous predictors ($\mathbf{X} \in \mathbb{R}^p$) and a categorical predictor (W). The partial central subspace is defined as the intersection of all subspaces $\mathcal{S}$ satisfying

$$Y \perp\!\!\!\perp \mathbf{X} \mid (P_{\mathcal{S}}\mathbf{X}, W), \tag{3.1}$$

where $W \in \{1, \ldots, G\}$ is a categorical predictor and $P_{(\cdot)}$ stands for a projection operator with respect to the standard inner product. The partial central subspace, which is assumed to exist and is denoted as $\mathcal{S}^{(W)}_{Y|\mathbf{X}}$, allows for reduction of the vector $\mathbf{X}$ of continuous predictors simultaneously across all subpopulations determined by W. There are several methods developed in the literature to infer about the partial central subspace, such as the partial SIR (Chiaromonte et al. [5]), partial SAVE (Shao et al. [26]), and partial IRE (Wen and Cook [27]).

For the methods of the partial dimension reduction, the following key equation connects the conditional central subspaces $\mathcal{S}_{Y^g|\mathbf{X}^g}$ with the partial central subspace $\mathcal{S}^{(W)}_{Y|\mathbf{X}}$:

$$\mathcal{S}^{(W)}_{Y|\mathbf{X}} = \bigoplus_{g=1}^{G} \mathcal{S}_{Y^g|\mathbf{X}^g}, \tag{3.2}$$

where $\bigoplus$ indicates the direct sum between two subspaces.

Under the framework of the partial central subspace, our testing hypothesis (1.4) is equivalent to

$$\mathcal{S}_{Y^1|\mathbf{X}^1} = \mathcal{S}_{Y^2|\mathbf{X}^2} = \cdots = \mathcal{S}_{Y^G|\mathbf{X}^G} = \mathcal{S}^{(W)}_{Y|\mathbf{X}}. \tag{3.3}$$

As we can see, when there are multiple populations (groups), the partial dimension reduction can be adapted to conduct multi-group dimension reduction with $W = 1, \ldots, G$, setting as the group identifier. However, partial central subspace approach comprises the related directions for all populations which is a direct sum of all the marginal central subspaces, there are some inherent drawbacks of this method. First, the population-specific effects are ignored by this method. Second, this approach cannot test if the same set of directions serve for all populations. Hence, it cannot deal with testing hypothesis such as (1.4) directly.

Under the null hypothesis (1.4), it is reasonable to assume that the dimensions of all the marginal central subspaces are equal to $d = \dim(\mathcal{S}^{(W)}_{Y|\mathbf{X}})$. We will assume that d is known in our article. An estimate of d can be easily obtained via any partial dimension reduction or single population dimension reduction methods, since $\dim(\mathcal{S}_{Y^g|\mathbf{X}^g}) = \dim(\mathcal{S}^{(W)}_{Y|\mathbf{X}})$, for any $g \in \{1, \ldots, G\}$, under null hypothesis (1.4). Based on our experiences, when the ranks of the conditional central subspaces are all equal, a good choice is to set d to be that number. When the ranks of the conditional central subspaces are not equal, the rank of the partial central subspace would provide a good estimate for d.

We propose a modified partial SIR without the homogeneous covariance constraint which was required by the original partial SIR (Chiaromonte et al. [5]). Since the inferences on d and $\mathcal{S}^{(W)}_{Y|\mathbf{X}}$ are not the main focuses of our paper, detailed discussions on the related methods via the modified partial SIR are provided in Appendix.

3.2. Test statistic with three or more populations

For the multiple population setting with $G \geq 3$, let $(Y^g_j, \mathbf{X}^g_j)$ for all $j \in \{1, \ldots, n_g\}$ be a simple random sample of size n_g from the gth population $(Y^g, \mathbf{X}^g)$, where $g \in \{1, \ldots, G\}$. Let

$$\bar{\mathbf{X}}_g = \frac{1}{n_g}\sum_{i=1}^{n_g} \mathbf{X}^g_i, \qquad \widehat{\boldsymbol{\Sigma}}_g = \frac{1}{n_g}\sum_{i=1}^{n_g} (\mathbf{X}^g_i - \bar{\mathbf{X}}_g)(\mathbf{X}^g_i - \bar{\mathbf{X}}_g)^\top.$$

Assume the dimensions of $\mathcal{S}_{Y^g|\mathbf{X}^g}$ are all equal to $d = \dim(\mathcal{S}^{(W)}_{Y|\mathbf{X}})$, which holds trivially under the null hypothesis (1.4). Let $\lambda_{g1} \geq \cdots \geq \lambda_{gd} > \lambda_{g,(d+1)} \geq \cdots \geq \lambda_{gp}$ be the eigenvalues of $\mathcal{M}_g$, and $\boldsymbol{\eta}_{gi}$ be the normalized eigenvector corresponding to λ_{gi}. Denote $\mathbf{P}_{gd} = \boldsymbol{\eta}_{g1}\boldsymbol{\eta}_{g1}^\top + \cdots + \boldsymbol{\eta}_{gd}\boldsymbol{\eta}_{gd}^\top$, $\mathbf{Q}_{gd} = I_p - \mathbf{P}_{gd}$, and let $\hat{\boldsymbol{\eta}}_{gi}$ denote the corresponding sample version of $\boldsymbol{\eta}_{gi}$, then $\mathbf{P}_{gd}$ can be estimated by $\widehat{\mathbf{P}}_{gd} = \hat{\boldsymbol{\eta}}_{g1}\hat{\boldsymbol{\eta}}_{g1}^\top + \cdots + \hat{\boldsymbol{\eta}}_{gd}\hat{\boldsymbol{\eta}}_{gd}^\top$. Let $\mathbf{A}^+$ denote the generalized inverse of matrix $\mathbf{A}$, from the perturbation theory (Kato [14]) and Theorem 1 in Schott [25], we then have

$$\widehat{\mathbf{P}}_{gd} = \mathbf{P}_{gd} + \sum_{i=1}^{d}\{\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}^\top \mathbf{A}_g(\lambda_{gi}I - \mathcal{M}_g)^+ + (\lambda_{gi}I - \mathcal{M}_g)^+\mathbf{A}_g\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}^\top\} + o_p(n^{-1/2})$$

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

We consider $\widehat{\mathbf{P}}(\widehat{\mathbf{P}}_{gd} - \mathbf{P}_{gd})$, where $\widehat{\mathbf{P}}$ is the eigenprojection of $\widehat{\mathcal{M}}^{(W)}$ corresponding to its largest d eigenvalues (detailed expressions of $\widehat{\mathcal{M}}^{(W)}$ are given in Appendix). Under the null hypothesis (1.4), the expected value of $\widehat{\mathbf{P}}$ is $\mathbf{P}_{gd}$, and $\lambda_{gk} = 0$, for $g \in \{1, \ldots, G\}$ and $k \in \{d+1, \ldots, p\}$, so we have

$$\begin{aligned}
\widehat{\mathbf{P}}(\widehat{\mathbf{P}}_{gd} - \mathbf{P}_{gd}) &= \widehat{\mathbf{P}} \sum_{i=1}^{d} \{\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}^{\top}\mathbf{A}_g(\lambda_{gi}I - \mathcal{M}_g)^{+} + (\lambda_{gi}I - \mathcal{M}_g)^{+}\mathbf{A}_g\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}^{\top}\} + o_p(n^{-1/2}) \\
&= \sum_{i=1}^{d} \widehat{\mathbf{P}}\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}^{\top}\mathbf{A}_g(\lambda_{gi}I - \mathcal{M}_g)^{+} + \sum_{i=1}^{d} \widehat{\mathbf{P}}(\lambda_{gi}I - \mathcal{M}_g)^{+}\mathbf{A}_g\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}^{\top} + o_p(n^{-1/2}) \\
&= \sum_{i=1}^{d} \sum_{k=d+1}^{p} \lambda_{gi}^{-1}\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}^{\top}\mathbf{A}_g\boldsymbol{\eta}_{gk}\boldsymbol{\eta}_{gk}^{\top} + o_p(n^{-1/2}).
\end{aligned}$$

Following Liu et al. [19], we expand $\mathbf{A}_g = \widehat{\mathcal{M}}_g - \mathcal{M}_g$ via the influence function approach as:

$$\mathbf{A}_g = \widehat{\mathcal{M}}_g - \mathcal{M}_g = \mathrm{E}_{n_g}\{\mathcal{M}_g^{*}(\mathbf{X}^g, Y^g)\} + O_p(n^{-1}),$$

where $\mathrm{E}_n\{\cdot\} = \sum_{i=1}^{n}(\cdot)/n$. The explicit expressions of $\mathcal{M}_g^{*}$ for SIR, SAVE and DR can be found in Liu et al. [19]. Let $\mathrm{vec}(\mathbf{A})$ denote the operator which stacks the columns of matrix $\mathbf{A}$ to form a vector, the following lemma gives the asymptotic distribution of $\sqrt{n_g}\,\mathrm{vec}\{\widehat{\mathbf{P}}(\widehat{\mathbf{P}}_{gd} - \mathbf{P}_{gd})\}$.

Lemma 1. *Assume that the data $(\mathbf{X}_1^g, Y_1^g), \ldots, (\mathbf{X}_{n_g}^g, Y_{n_g}^g)$ form a simple random sample from $(\mathbf{X}^g, Y^g)$ with finite fourth order moments. Then under null hypothesis (1.4), we have:*

$$\sqrt{n_g}\,\mathrm{vec}\{\widehat{\mathbf{P}}(\widehat{\mathbf{P}}_{gd} - \mathbf{P}_{gd})\} \rightsquigarrow \mathcal{N}(0, \boldsymbol{\Psi}_g),$$

where $\boldsymbol{\Psi}_g = \mathbf{U}_g\boldsymbol{\Phi}_g\mathbf{U}_g^{\top}$, $\boldsymbol{\Phi}_g = \mathrm{E}[\mathrm{vec}\{\mathcal{M}_g^{*}(\mathbf{X}^g, Y^g)\}\,\mathrm{vec}\{\mathcal{M}_g^{*}(\mathbf{X}^g, Y^g)\}^{\top}]$ is the asymptotic covariance matrix of $\sqrt{n_g}\,\mathrm{vec}(\mathbf{A}_g)$, and

$$\mathbf{U}_g = \sum_{i=1}^{d} \sum_{k=d+1}^{p} \lambda_{gi}^{-1}(\boldsymbol{\eta}_{gk}\boldsymbol{\eta}_{gk}{}^{\top}) \otimes (\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}{}^{\top}).$$

For all $g \in \{1, \ldots, G\}$, let $\mathbf{t}_g = \mathrm{vec}\{\widehat{\mathbf{P}}(\widehat{\mathbf{P}}_{gd} - \mathbf{P}_{gd})\}$ and write

$$\bar{\mathbf{t}} = \sum_{i=1}^{G} \frac{n_i}{n}\mathbf{t}, \qquad \widehat{\widehat{\mathbf{P}}} = \sum_{i=1}^{G} \frac{n_i}{n}\widehat{\mathbf{P}}_{id}.$$

Then

$$\mathbf{t}_g - \bar{\mathbf{t}} = \mathrm{vec}\left\{\widehat{\mathbf{P}}(\widehat{\mathbf{P}}_{gd} - \widehat{\widehat{\mathbf{P}}}) - \widehat{\mathbf{P}}\left(\mathbf{P}_{gd} - \sum_{i=1}^{G} n_i/n\,\mathbf{P}_{id}\right)\right\}.$$

Also, let $\mathbf{m}_g = \mathrm{vec}\{\widehat{\mathbf{P}}(\widehat{\mathbf{P}}_{gd} - \widehat{\widehat{\mathbf{P}}})\}$. Then, under null hypothesis (1.4), $\mathbf{m}_g$ has the same asymptotic distribution as $\mathbf{t}_g - \bar{\mathbf{t}}$.

Define $\mathbf{t} = ((\mathbf{t}_1 - \bar{\mathbf{t}})^{\top}, \ldots, (\mathbf{t}_G - \bar{\mathbf{t}})^{\top})^{\top}$ and $\mathbf{m} = (\mathbf{m}_1^{\top}, \ldots, \mathbf{m}_G^{\top})^{\top}$. We have

$$\begin{aligned}
\sqrt{n}\,\mathbf{t} &= \sqrt{n}\left(\mathbf{I}_{p^2G} - \frac{1}{n}(n_1, \ldots, n_G) \otimes \mathbf{1}_G \otimes \mathbf{I}_{p^2}\right)(\mathbf{t}_1^{\top}, \ldots, \mathbf{t}_G^{\top})^{\top} \\
&= \left(\mathrm{diag}\left(\sqrt{\frac{n}{n_1}}, \ldots, \sqrt{\frac{n}{n_G}}\right) - \frac{1}{\sqrt{n}}(\sqrt{n_1}, \ldots, \sqrt{n_G}) \otimes \mathbf{1}_G\right) \otimes \mathbf{I}_{p^2}(\sqrt{n_1}(\mathbf{t}_1)^{\top}, \ldots, \sqrt{n_G}(\mathbf{t}_G)^{\top})^{\top},
\end{aligned}$$

where $\mathbf{1}_G$ is the G-dimensional vector of all ones.

Let $c_g = n_g/n$ for all $g \in \{1, \ldots, G\}$. Assuming that as $n \to \infty$, c_g is a constant. Based on the asymptotic distribution of $\sqrt{n_g}\,\mathrm{vec}\{\widehat{\mathbf{P}}(\widehat{\mathbf{P}}_{gd} - \mathbf{P}_{gd})\}$ obtained in Lemma 1, Theorem 1 gives the asymptotic distribution of $\sqrt{n}\,\mathbf{m}$, which is the same as that of $\sqrt{n}\,\mathbf{t}$.

Theorem 1. *Assume that for $g = 1, 2$, the data $(\mathbf{X}_1^g, Y_1^g), \ldots, (\mathbf{X}_{n_g}^g, Y_{n_g}^g)$ form a simple random sample from $(\mathbf{X}^g, Y^g)$ with finite fourth order moments. Then under null hypothesis (1.4), we have:*

$$\sqrt{n}\,\mathbf{m} \rightsquigarrow \mathcal{N}(0, \mathbf{B}\mathbf{W}\mathbf{B}^{\top}),$$

where $\mathbf{B} = (diag(\sqrt{1/c_1}, \ldots, \sqrt{1/c_G}) - (\sqrt{c_1}, \ldots, \sqrt{c_G}) \otimes \mathbf{1}_G) \otimes \mathbf{I}_{p^2}$ and $\mathbf{W} = diag(\boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_G)$.

Proof. Lemma 1 shows that under null hypothesis (1.4), $\sqrt{n_g}\,\mathbf{t_g} \rightsquigarrow \mathcal{N}(0, \boldsymbol{\Psi}_g)$. Since

$$\widehat{\mathbf{P}}(\widehat{\mathbf{P}}_{gd} - \mathbf{P}_{gd}) = \sum_{i=1}^{d}\sum_{k=d+1}^{p} \lambda_{gi}^{-1}\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}^{\top}\mathbf{A}_g\boldsymbol{\eta}_{gk}\boldsymbol{\eta}_{gk}^{\top} + o_p(n^{-1/2})$$

is just related to the gth population, $\mathbf{t}_1, \ldots, \mathbf{t}_G$ are asymptotically independent with each other. Then, we have

$$(\sqrt{n_1}\mathbf{t_1}^{\top}, \ldots, \sqrt{n_G}\mathbf{t_G}^{\top})^{\top} \rightsquigarrow \mathcal{N}(0, \mathbf{W}).$$

Hence, $\sqrt{n}\,\mathbf{t} \rightsquigarrow \mathcal{N}(0, \mathbf{BWB}^{\top})$. Because $\sqrt{n}\,\mathbf{m}$ and $\sqrt{n}\,\mathbf{t}$ have the same asymptotic distribution, we can conclude that $\sqrt{n}\,\mathbf{m} \rightsquigarrow \mathcal{N}(0, \mathbf{BWB}^{\top})$, as claimed. □

Define $T = n\mathbf{m}^{\top}\mathbf{m}$ as our test statistic, the following theorem provides its asymptotic distribution under the null hypothesis (1.4).

Theorem 2. *Assume the conditions of Theorem 1 hold, then under null hypothesis (1.4), we have*

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

where $\omega_1 \geq \cdots \geq \omega_{(G-1)d(p-d)}$ are the eigenvalues of $\mathbf{BWB}^{\top}$, and $\chi_1^2(1), \ldots, \chi_{(G-1)d(p-d)}^2(1)$ denote i.i.d. chi-square random variables with 1 degree of freedom.

Proof. Since $\mathbf{B} = (\mathrm{diag}(\sqrt{1/c_1}, \ldots, \sqrt{1/c_G}) - (\sqrt{c_1}, \ldots, \sqrt{c_G}) \otimes \mathbf{1}_G) \otimes \mathbf{I}_{p^2}$, one has $\mathrm{rank}(\mathbf{B}) = (G-1) \times p^2$. Also, $\mathbf{W} = \mathrm{diag}(\boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_G)$, where $\boldsymbol{\Psi}_g = \mathbf{U}_g\boldsymbol{\Phi}_g\mathbf{U}_g^{\top}$. Hence, $\mathrm{rank}(\boldsymbol{\Psi}_g) = \mathrm{rank}(\mathbf{U}_g) = d(p-d)$, and $\mathrm{rank}(\mathbf{W}) = d(p-d)G$.

By the elementary row and column operations, we may rewrite $\mathbf{B}$ as

$$\mathbf{B} = \mathbf{Q}(\mathbf{B}_1, \ldots, \mathbf{B}_{G-1}, \mathcal{O}),$$

where $\mathbf{Q}$ is the multiplication of all the row and column operations and $\mathbf{B}_i$ is a $p^2G \times p^2$ column full rank matrix, for each $i \in \{1, \ldots, G-1\}$. Therefore,

$$\begin{aligned}\mathrm{rank}(\mathbf{BWB}^{\top}) &= \mathrm{rank}((\mathbf{B}_1, \ldots, \mathbf{B}_{G-1}, \mathcal{O})\mathbf{W}(\mathbf{B}_1, \ldots, \mathbf{B}_{G-1}, \mathcal{O})^{\top}) \\ &= \mathrm{rank}((\mathbf{B}_1, \ldots, \mathbf{B}_{G-1})\mathrm{diag}(\boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_{G-1})(\mathbf{B}_1, \ldots, \mathbf{B}_{G-1})^{\top}).\end{aligned}$$

Because $(\mathbf{B}_1, \ldots, \mathbf{B}_{G-1})$ is a column full rank matrix, there exists an invertible matrix $\mathbf{D}$ such that

$$(\mathbf{B}_1, \ldots, \mathbf{B}_{G-1}) = \mathbf{D}\begin{pmatrix}\mathbf{I}_{p^2(G-1)} \\ \mathcal{O}\end{pmatrix}.$$

So, we have

$$\mathrm{rank}\big((\mathbf{B}_1, \ldots, \mathbf{B}_{G-1})\mathrm{diag}(\boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_{G-1})(\mathbf{B}_1, \ldots, \mathbf{B}_{G-1})^{\top}\big) = \mathrm{rank}(\boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_{G-1}) = d(p-d)(G-1).$$

Then, the conclusion just follows naturally. □

A consistent estimate $\widehat{\mathbf{W}}$ of $\mathbf{W}$ can be obtained by substituting sample estimates for the unknown quantities. The weights ω_i can be consistently estimated using the eigenvalues of $\mathbf{B}\widehat{\mathbf{W}}\mathbf{B}^{\top}$. In the following simulation studies, we compare the observed value of the test statistic T to the percentage points of

$$\sum_{i=1}^{d(p-d)(G-1)} \hat{\omega}_i \chi_i^2(1)$$

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.

4. Numerical studies

4.1. Simulation studies

Throughout our simulation studies, the random error ϵ is assumed to be standard normal, viz. $\mathcal{N}(0, 1)$, and independent of $\mathbf{X}$. The dimension of the predictor vector p is taken to be 5 or 10, the number of slices is $h = 5$. We summarize our results over 1000 replications for each simulation study. We studied the performance of our proposed tests via (modified partial) SIR with different choices of n and p.

Table 4.1
Estimated test levels (in percentages) for Model **I**.

	Model **I** with $p = 5$			Model **I** with $p = 10$		
Sample size	Nominal level (%)			Nominal level (%)		
	1	5	10	1	5	10
$n_1 = n_2 = n_3 = 400$	1.80	6.80	13.30	2.20	7.10	14.10
$n_1 = n_2 = n_3 = 600$	1.40	6.20	10.40	2.00	6.60	11.90
$n_1 = n_2 = n_3 = 800$	1.10	5.40	10.4	0.90	6.20	10.80

Table 4.2
Estimated test levels (in percentages) for Model **II**.

	Model **II** with $p = 5$			Model **II** with $p = 10$		
Sample size	Nominal level (%)			Nominal level (%)		
	1	5	10	1	5	10
$n_1 = n_2 = n_3 = 400$	3.10	7.90	15.60	0.30	3.40	6.60
$n_1 = n_2 = n_3 = 600$	2.80	7.50	13.20	0.50	3.70	6.90
$n_1 = n_2 = n_3 = 800$	1.60	6.40	11.70	0.60	3.50	7.10

4.1.1. Estimated test levels

In this subsection, we evaluate the performance of our test statistics under different models when null hypothesis (1.4) holds.

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

$$Y = \begin{cases} \exp(X_1 + X_2 + X_3) + \epsilon_1, & \text{for group 1;} \\ \sin(X_1 + X_2 + X_3) + \epsilon_2, & \text{for group 2;} \\ X_1 + X_2 + X_3 + \epsilon_3, & \text{for group 3.} \end{cases}$$

Table 4.1 shows the estimated test levels for our test statistics. As the sample size increases, the estimated levels are closer to the nominal levels. For example, when $p = 5$ at the nominal level 1%, the estimated levels are 1.8%, 1.4% and 1.1% respectively, for sample sizes of 400, 600 and 800. Also it comes as no surprise that the performance of our tests slightly deteriorates as p increases.

Model **II**. In this model, the predictor vector $X_i's$ are independent and t distributed with degrees of freedom 5.

$$Y = \begin{cases} (X_1 + 5)/(X_2 + 2) + \epsilon_1, & \text{for group 1;} \\ X_1 + 1/(X_2 + 3) + \epsilon_2, & \text{for group 2;} \\ X_1 + \exp(X_2 + 2) + \epsilon_3, & \text{for group 3.} \end{cases}$$

From Table 4.2, we could tell that SIR-based method is strongly affected by value of p. When p is 5, the estimation test levels based on SIR tend to be greater than the nominal levels, while they tilt to the other direction when p is 10. But generally speaking, the performance of our methods seems reasonable when the independent variables are not normally distributed.

Model **III**. We now consider a one-dimensional model with correlated predictors as follows:

$$Y = \begin{cases} \exp(X_1 + X_2) + \epsilon_1, & \text{for group 1;} \\ \sin(X_1 + X_2) + \epsilon_2, & \text{for group 2;} \\ X_1 + X_2 + \epsilon_3, & \text{for group 3.} \end{cases}$$

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 as $0.5^{|i-j|}$ for all $i \in \{1, \ldots, p\}$ and $j \in \{1, \ldots, p\}$. Different groups share common indices and $d = 1$. It seems that the correlation among the predictors does not substantially affect the performance of our methods (see Table 4.3).

4.1.2. Estimated power

We examine the power of our tests under the alternative hypothesis in this subsection. The predictors $\mathbf{X}$ for the model again follow the standard multivariate normal distribution.

Model **IV**.

$$Y = \begin{cases} \exp(X_1 + X_2) + \epsilon_1, & \text{for group 1;} \\ \sin(X_3 - X_2) + \epsilon_2, & \text{for group 2;} \\ X_4 + X_p + \epsilon_3, & \text{for group 3.} \end{cases}$$

Table 4.3
Estimated test levels (in percentages) for Model **III**.

Sample size	Model **IV** with $p = 5$			Model **IV** with $p = 10$		
	Nominal level (%)			Nominal level (%)		
	1	5	10	1	5	10
$n_1 = n_2 = n_3 = 400$	2.00	7.00	16.20	2.30	7.20	17.20
$n_1 = n_2 = n_3 = 600$	1.70	6.70	13.70	1.90	7.00	14.90
$n_1 = n_2 = n_3 = 800$	1.30	5.80	12.2	1.60	6.20	13.5

Table 4.4
Estimated power at 5% nominal levels for Model **IV** at $d = 1$.

Sample size	$p = 5$	$p = 10$
$n_1 = n_2 = n_3 = 400$	100	100
$n_1 = n_2 = n_3 = 600$	100	100
$n_1 = n_2 = n_3 = 800$	100	100

Table 4.5
Estimated power at 5% nominal levels for Model **IV** at $d = 3$.

Sample size	$p = 5$	$p = 10$
$n_1 = n_2 = n_3 = 400$	62.4	69.8
$n_1 = n_2 = n_3 = 600$	65.3	69.3
$n_1 = n_2 = n_3 = 800$	68.6	73.5

Table 4.6
Estimated power at 5% nominal levels for Model **V**.

Sample size	Our method (power × 100)		Liu et al. [19] (power × 100)	
	$p = 5$	$p = 10$	$p = 5$	$p = 10$
$n_1 = n_2 = 400$	79.8	58.0	0	0
$n_1 = n_2 = 600$	92.4	79.8	0	0
$n_1 = n_2 = 800$	98.4	91.0	0	0

Here each group has a different direction, that is $\mathcal{S}_{Y^g|\mathbf{X}^g} = 1$ for each $g \in \{1, 2, 3\}$. We first set $d = 1$. Here, $d = 1$ is the dimension of each group. As shown in Table 4.4, for all the sample sizes, p, our methods performed well with 100% of power.

If we use a different structural dimension for Model **IV**, say $d = 3$, which is shown in Table 4.5, the power of our test is generally smaller than those when $d = 1$. Here, $d = 1$ is the dimension of each group and $d = 3$ is the dimension of partial central subspace. It seems that the power of our procedures is pretty sensitive to the value we choose for d. Based on our simulation studies, an estimate of d using single population dimension reduction often yields greater power.

4.1.3. *Comparison with existing method for $G = 2$*

In this subsection, we compare our method with that of Liu et al. [19] when $G = 2$. We first consider the following model.

Model **V**.

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

The predictors $\mathbf{X}$ for this model again follow the standard multivariate normal distribution. Table 4.6 shows the power of the two tests with $d = 2$ at testing level of 0.05. It is clear that our test does a much better job than Liu et al. [19]. This is not surprising since for this particular model, we have $\mathbf{P}_1 \subset \mathbf{P}_2$, and the test statistic of Liu et al. [19] is constructed using the sample version of $\mathbf{P}_1\mathbf{Q}_2$. Hence, $\mathbf{P}_1\mathbf{Q}_2 = 0$ here, even though the null hypothesis (1.4) does not hold, the test proposed by Liu et al. [19] fails to reject it correctly. In contrast, our approach will not encounter this problem, since in a way, we utilize the difference between $\mathbf{P}_1$ and $\mathbf{P}_2$, which is clearly not zero under this model.

We also compared the performance of our method with that of Liu et al. [19] regarding the test levels. Simulation studies not reported here show that both methods yield very similar results.

4.2. *A real data analysis*

Numerous observational studies suggest that low dietary intake or low plasma concentrations of retinol, beta-carotene, or other carotenoids are associated with increased risk of developing certain types of cancer (Peto et al. [23]). It has been of

Table 4.7
Variable names.

Variable name	Brief description
AGE	Age (years)
SEX	Sex (1 = Male, 2 = Female)
SMOKSTAT	Smoking status (1 = Never, 2 = Former, 3 = Current Smoker)
QUETELET	Quetelet (weight/height^2)
VITUSE	Vitamin Use (1 = Yes, fairly often, 2 = Yes, not often, 3 = No)
CALORIES	Number of calories consumed per day
FAT	Grams of fat consumed per day
FIBER	Grams of fiber consumed per day
ALCOHOL	Number of alcoholic drinks consumed per week
CHOLESTEROL	Cholesterol consumed (mg per day)
BETADIET	Dietary beta-carotene consumed (mcg per day)
RETDIET	Dietary retinol consumed (mcg per day)
BETAPLASMA	Plasma beta-carotene (ng/ml)
RETPLASMA	Plasma Retinol (ng/ml)

interest to determine those factors that may affect these concentrations, and so several studies have been conducted in the past. For example, studies to investigate the effect of personal characteristics and dietary factors on plasma concentrations in human serum, and to build models using these variables to predict and evaluate plasma concentrations of retinol and beta-carotene accurately were carried out in Nierenberg et al. [22]. Zhu et al. [31], Yoo [28,29], and Hilafu and Yin [13] also considered such factors.

Here, we apply our method to a dataset to determine how smoking status affects the relationship between some personal characteristics, dietary factors and the concentration of beta-carotene. The data "plasma-retinol" is available at the online library of data files of Carnegie Mellon University (http://lib.stat.cmu.edu). Study objects containing 315 observations on 14 variables were patients who had an elective surgical procedure during a three-year period to biopsy or remove a lesion of the lung, colon, breast, skin, ovary or uterus that was found to be non-cancerous. Note that subject 62 with an extreme high value of alcohol use is treated as outlier by Hilafu and Yin [13], Zhu et al. [31] and was deleted ahead of time. We also remove subject 62 from our data analysis. Variables SEX, SMOKSTAT, VITUSE are categorical, BETAPLASMA and RETPLASMA are continuous response variables and the remaining nine variables are also continuous. Detailed descriptions of our data are given in Table 4.7 as below.

We take one of the categorical variables, smoking status (SMOKSTAT) as the group identifier and divide the study objects into three groups: nonsmoker, former smoker and current smoker. The remaining nine continuous variables (X_1 = AGE, X_2 = QUETELET, X_3 = CALORIES, X_4 = FAT, X_5 = FIBER, X_6 = ALCOHOL, X_7 = CHOLESTEROL, X_8 = BETADIET, X_9 = RETDIET) are the independent variables and plasma beta-carotene is the dependent variable. We take the number of slices h of our methods to be 5, and the number of directions within each group to be $d = 1$ which is the dimension of the partial central subspace $\mathcal{S}_{Y|\mathbf{X}}^{(W)}$, with W = SMOKSTAT. The observed test statistic is $T = 259.15$, which is greater than 224.20, the 95th percentiles of the simulated weighted chi-square distribution. Hence, we reject the null hypothesis which means that smoking status does affect how these dietary factors and personal characteristics considered in this study influence the concentration of beta-carotene in human serums.

The results of our analysis are consistent with that of Hilafu and Yin [13], which might shed light on the possible causal mechanisms between smoking and cancer risk. In fact, many studies have shown that smoking increases the risk of many cancers. Our conclusion combined with the observational studies conducted by Peto et al. [23] could also help us better understand the relationships between smoking and the risk of these cancers.

5. Summary

In this article, we developed a new test statistic, and its asymptotic distribution, for testing the common indices of three or more multi-index models. Simulation results show that our new method is able to detect if different groups share the same dimension reduction subspaces. In the real life, our method could also be used to check the significance of some categorical variable. Applying our method to the plasma beta-carotene dataset, we find that the dimension reduction subspaces of the three groups (nonsmoker, previous smoker and current smoker) are not the same. This conclusion means that the smoking status variable may significantly affect how those personal characteristics and dietary factors influence the concentration of beta-carotene in human serums. Recently, Chavent et al. [4] proposed an adaptive SIR method for data stream consisting of sequential blocks. Our method might also be used to assist the detection of aberrant blocks in that setting.

Our method utilizes the concept of partial central subspace. In simulation studies, we found that our new method tends to yield greater power comparing to the method proposed by Liu et al. [19].

Acknowledgments

We would like to thank the Editor-in-Chief, Associate Editor and the two referees for their constructive suggestions and thoughtful comments which greatly improved our paper.

Appendix. Modified partial SIR

For a random sample of size n_g from the gth population $(Y^g, \mathbf{X}^g)$with $g \in \{1, \dots, G\}$, let

$$\bar{\mathbf{X}}_g = \frac{1}{n_g}\sum_{i=1}^{n_g}\mathbf{X}_i^g, \qquad \widehat{\boldsymbol{\Sigma}}_g = \frac{1}{n_g}\sum_{i=1}^{n_g}(\mathbf{X}_i^g - \bar{\mathbf{X}}_g)(\mathbf{X}_i^g - \bar{\mathbf{X}}_g)^\top,$$

and $n = n_1 + \cdots + n_G$. Standardize the predictor, $\mathbf{Z}_i^g = \widehat{\boldsymbol{\Sigma}}_g^{-1/2}(\mathbf{X}_i^g - \bar{\mathbf{X}}_g)$ for each $i \in \{1, \dots, n_g\}$ and $g \in \{1, \dots, G\}$. Following the common practice in sufficient dimension reduction, partition the range of Y^g into H_g slices. For each $s \in \{1, \dots, H_g\}$ and $g \in \{1, \dots, G\}$, compute the intra-slice mean vector as

$$\bar{\mathbf{Z}}_{gs} = \frac{1}{n_{gs}}\sum_{j=1}^{n_{gs}}\mathbf{Z}_j^g, \tag{A.1}$$

where the sum is over indices j of response observations Y_j^g that fall into slice s, and n_{gs} is the number of observations in slice s, for population g. With a little abuse of notations, in the following discussions, we use $\{Y^g = s\}$ as short for $\{Y^g$ is in slice $s\}$.

The original partial SIR proposed by Chiaromonte et al. [5] requires the following homogeneous predictor covariance condition across the populations:

$$\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_G.$$

Experience has shown that this homogeneous covariance condition restricts application of partial SIR in practice, and that its failure can result in misleading conclusions. In this article, we propose a modified partial SIR without the homogeneous covariance constraint, which we still call partial SIR. Throughout this article, the partial SIR we used refers to the modified version.

For partial SIR, the sample version of $\mathcal{M}_g$ for population $g \in \{1, \dots, G\}$ is given by

$$\widehat{\mathcal{M}}_g^{sir} = \sum_{s=1}^{h_g}\frac{n_{gs}}{n_g}\bar{\mathbf{Z}}_{gs}\bar{\mathbf{Z}}_{gs}^\top. \tag{A.2}$$

Define $a_g = \Pr(G = g)$, $\alpha_g = \sqrt{a_g}$, $\hat{a}_g = n_g/n$, $\hat{\alpha}_g = \sqrt{\hat{a}_g}$, $\phi_{gs} = \Pr(Y^g = s)$, $f_{gs} = \sqrt{\phi_{gs}}$, $\hat{\phi}_{gs} = n_{gs}/n_g$, $\hat{f}_{gs} = \sqrt{\hat{\phi}_{gs}}$. Also, let $\widehat{\mathbf{H}}_g = (\hat{f}_{g1}\bar{\mathbf{Z}}_{g1}, \dots, \hat{f}_{gh_g}\bar{\mathbf{Z}}_{gh_g})$. Then $\widehat{\mathcal{M}}_g^{sir} = \widehat{\mathbf{H}}_g\widehat{\mathbf{H}}_g^\top$. Averaging the sample candidate matrices over each population, we obtain

$$\widehat{\mathcal{M}}^{(W)} = \sum_{g=1}^{G}\frac{n_g}{n}\widehat{\mathcal{M}}_g^{sir} = \widehat{\mathbf{H}}\widehat{\mathbf{H}}^\top,$$

where $\widehat{\mathbf{H}} = (\alpha_1\widehat{\mathbf{H}}_1, \dots, \alpha_G\widehat{\mathbf{H}}_G)$. Let $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p$ be the singular values of $\widehat{\mathbf{H}}$, and define

$$T_{sir}(m) = n\sum_{k=m+1}^{p}\hat{\lambda}_k^2.$$

Let $\mathbf{H}$ be the population version of $\widehat{\mathbf{H}}$. We first construct the singular value decomposition of $\mathbf{H}$, viz.

$$\mathbf{H} = \begin{pmatrix}\boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_0\end{pmatrix}\begin{pmatrix}\mathbf{D} & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\boldsymbol{\Psi}_1^\top\\ \boldsymbol{\Psi}_0^\top\end{pmatrix},$$

where $(\boldsymbol{\Gamma}_1 \;\; \boldsymbol{\Gamma}_0)$ is a $p \times p$ orthogonal matrix in which $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_0$ have dimensions $p \times m$ and $p \times (p - m)$, $(\boldsymbol{\Psi}_1 \;\; \boldsymbol{\Psi}_0)$ is an $h \times h$ orthogonal matrix, in which $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_0$ have dimensions $h \times m$ and $h \times (h - m)$, and $\mathbf{D}$ is an $m \times m$ diagonal matrix of positive diagonal elements. Following Eaton and Tyler [9], under the null hypothesis $d = m$, $T_{sir}(m)$ has the same asymptotic distribution as

$$\mathrm{vec}^\top\{\sqrt{n}\,\boldsymbol{\Gamma}_0^\top(\widehat{\mathbf{H}} - \mathbf{H})\boldsymbol{\Psi}_0\}\,\mathrm{vec}\{\sqrt{n}\,\boldsymbol{\Gamma}_0^\top(\widehat{\mathbf{H}} - \mathbf{H})\boldsymbol{\Psi}_0\}.$$

Thus we only need to derive the asymptotic distribution of $\sqrt{n}\,\boldsymbol{\Gamma}_0^\top(\widehat{\mathbf{H}} - \mathbf{H})\boldsymbol{\Psi}_0$, which is provided by the following lemma.

Lemma 2. *One has* $\sqrt{n}\,\mathrm{vec}\{\boldsymbol{\Gamma}_0^\top(\widehat{\mathbf{H}}-\mathbf{H})\boldsymbol{\Psi}_0\} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\boldsymbol{\Omega})$, *where* $\otimes$ *denotes the Kronecker product,* $\boldsymbol{\Omega} = (\boldsymbol{\Psi}_0^\top \otimes \boldsymbol{\Gamma}_0^\top)\,\mathrm{diag}(\boldsymbol{\Delta}_1,\ldots,\boldsymbol{\Delta}_G)(\boldsymbol{\Psi}_0 \otimes \boldsymbol{\Gamma}_0)$, *and for each* $g \in \{1,\ldots,G\}$, $\boldsymbol{\Delta}_g$ *is defined by Equation* (8) *in Bura and Cook* [3], *and* $\mathrm{diag}(\cdot)$ *denotes a positive definite block diagonal matrix.*

Proof. By Equation (8) of Bura and Cook [3], we have the following result:

$$\sqrt{n_g}\,\mathrm{vec}\{(\widehat{\mathbf{H}}_g-\mathbf{H}_g)\} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\boldsymbol{\Delta}_g), \tag{A.3}$$

where $\boldsymbol{\Delta}_g$ is defined in Bura and Cook [3]. The conclusion follows. □

In the area of sufficient dimension reduction, estimation of d is often based on testing a sequence of hypotheses $\mathcal{H}_0 : d = m$ versus $\mathcal{H}_a : d > m$, with m incremented by 1 until the hypothesis is not rejected. At which point $\hat{d}$ is the last value of m tested. For partial SIR, the following theorem provides a test statistic for testing $\mathcal{H}_0 : d = m$ versus $\mathcal{H}_a : d > m$.

Theorem 3. *Assuming the linearity condition for* $\mathbf{X}_1,\ldots,\mathbf{X}_G$ *under the null hypothesis of* $\mathcal{H}_0 : d = m$. *The limiting distribution of* $T_{sir}(m)$ *is then the same as that of*

$$\sum_{i=1}^{(h-m)(p-m)} \omega_i K_i$$

where $h = h_1 + \cdots + h_G$ *is the total number of slices, the* $K_1,\ldots,K_{(h-m)(p-m)}$ *are i.i.d.* χ^2 *random variables with degree of freedom one, and* $\omega_1 \ge \cdots \ge \omega_{(h-m)(p-m)}$ *are the ordered eigenvalues of* $\boldsymbol{\Omega}$.

The proof of Theorem 3 is straightforward from Lemma 2, hence it is omitted here.

References

[1] P.M. Bentler, J. Xie, Corrections to test statistics in principal Hessian directions, Statist. Probab. Lett. 47 (2000) 381–389.
[2] R.J. Biok, Spectral models for covariance matrices, Biometrika 89 (2002) 159–182.
[3] E. Bura, R.D. Cook, Extending sliced inverse regression: The weighted chi-squared test, J. Amer. Statist. Assoc. 96 (2001) 996–1003.
[4] M. Chavent, S. Girard, V. Kuentz-Simonet, B. Liquet, T. Nguyen, J. Saracco, A sliced inverse regression approach for data stream, Comput. Statist. 29 (2014) 1129–1152.
[5] F. Chiaromonte, R.D. Cook, B. Li, Sufficient dimension reduction in regressions with categorical predictors, Ann. Statist. 30 (2002) 475–497.
[6] R.D. Cook, Regression Graphics, Wiley, New York, 1998.
[7] R.D. Cook, B. Forzani, Likelihood-based sufficient dimension reduction, J. Amer. Statist. Assoc. 104 (2009) 197–208.
[8] R.D. Cook, S. Weisberg, Discussion of Sliced inverse regression for dimension reduction, J. Amer. Statist. Assoc. 86 (1991) 328–332.
[9] M.L. Eaton, D. Tyler, The asymptotic distribution of singular-values with applications to canonical correlations and correspondence analysis, J. Multivariate Anal. 50 (1994) 238–264.
[10] B. Flury, Common principal components in *k* groups, J. Amer. Statist. Assoc. 79 (1984) 892–898.
[11] B. Flury, Two generalizations of the common principal component model, Biometrika 74 (1987) 59–69.
[12] B. Flury, H. Riedwyl, Multivariate Statistics: A Practical Approach, Wiley, New York, 1988.
[13] H. Hilafu, X. Yin, Sufficient dimension reduction in multivariate regressions with categorical predictors, Comput. Statist. Data Anal. (2013) 139–147.
[14] T. Kato, Perturbation Theory for Linear Operators, Springer, Berlin, 1966.
[15] W.J. Krzanowski, Between-groups comparison of principal components, J. Amer. Statist. Assoc. 74 (1979) 703–707.
[16] K.Y. Lee, B. Li, F. Chiaromonte, A general theory for nonlinear sufficient dimension reduction: Formulation and estimation, Ann. Statist. 6 (2013) 3182–3210.
[17] K.C. Li, Sliced inverse regression for dimension reduction (with discussion), J. Amer. Statist. Assoc. 86 (1991) 316–342.
[18] B. Li, S. Wang, On directional regression for dimension reduction, J. Amer. Statist. Assoc. 102 (2007) 997–1008.
[19] X. Liu, Z. Yu, X. Wen, R. Paige, On testing common indices for two multi-index models: A link-free approach, J. Multivariate Anal. 136 (2015) 75–85.
[20] Y. Ma, L. Zhu, A semiparametric approach to dimension reduction, J. Amer. Statist. Assoc. 107 (2012) 168–179.
[21] Y. Ma, L. Zhu, Efficient estimation in sufficient dimension reduction, Ann. Statist. 41 (2013) 250–268.
[22] D.W. Nierenberg, T.A. Stukel, J.A. Baron, Determinants of plasma levels of beta-carotene and retinol, Am. J. Epidemiol. 130 (1989) 511–521.
[23] R. Peto, R. Doll, J.D. Buckley, M.B. Sporn, Can dietary beta-carotene materially reduce human cancer rates? Nature 290 (1981) 201–208.
[24] J.R. Schott, Common principal component subspaces in two groups, Biometrika 75 (1988) 229–236.
[25] J.R. Schott, Some tests for common principal component subspaces in several groups, Biometrika 78 (1991) 771–777.
[26] Y. Shao, R.D. Cook, S. Weisberg, Partial central subspace and sliced average variance estimation, J. Statist. Plann. Inference 139 (2009) 952–961.
[27] X. Wen, R.D. Cook, Optimal sufficient dimension reduction in regressions with categorical predictors, J. Statist. Plann. Inference 137 (2007) 1961–1978.
[28] J.K. Yoo, A novel moment-based sufficient dimension reduction approach in multivariate regression, Comput. Stat. Data Anal. 52 (2008) 3843–3851.
[29] J.K. Yoo, Integrated partial sufficient dimension reduction with heavily unbalanced categorical predictors, Korean J. Appl. Statist. 23 (2010) 977–985.
[30] Z. Yu, L.X. Zhu, X. Wen, On model-free conditional coordinate tests for regressions, J. Multivariate Anal. 109 (2012) 61–72.
[31] L.P. Zhu, L.X. Zhu, S.Q. Wen, On dimension reduction in regressions with multivariate responses, Statist. Sinica 20 (2010) 1291–1307.