2015

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

Xuejing Liu

Department: Mathematics and Statistics

ON TESTING COMMON INDICES FOR SEVERAL MULTI-INDEX MODELS: A LINK-FREE APPROACH

by

XUEJING LIU

A DISSERTATION

Presented to the Faculty of the Graduate School of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

MATHEMATICS

2015

Approved by:

Dr. Xuerong Meggie Wen, Advisor
Dr. Robert Paige, Co-adviser
Dr. V. A. Samaranayake
Dr. John Singler
Dr. Xiaoping Du

DEDICATION

I would like to dedicate this thesis to my father Qi Liu and my mother Sifen Guo, from whom I have gotten unconditional love and support. I hope this thesis is a good birthday gift for my mum.

ABSTRACT

To avoid the curse of dimensionality, and to help us better understand the structure of the high dimensional data, methods for dimension reduction are clearly called for. The common linear dimension reduction techniques for single population include principal component analysis (PCA) which is unsupervised in regression and supervised Partial Least Squares (PLS). Modern sufficient dimension reduction techniques, like the ones we consider, constitute a form of supervised linear dimension reduction which outperform PCA and PLS without the underlying model assumptions.

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 set of overall features. Common PCA and partial dimension reduction methods are extant methods for multiple groups. Note however that common PCA is unsupervised and doesn't take into account the information in Y and partial dimension reduction ignores the population-specific effects. Most importantly, these methods can not tell us if the same set of directions serve for all populations.

To determine these common directions, we first propose a link-free procedure for testing whether two multi-index models share identical indices via the sufficient dimension reduction approach. Then a general method is introduced for two or more models based on modified partial dimension reduction. We present our test statistics, with associated asymptotic distributions and simulation studies. Applications to the well-known AIS data and the beta-carotene data are also used to demonstrate our methods. Furthermore, simulation studies are used to compare the various methods we consider.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude towards my advisors Drs. Xuerong Wen and Robert Paige for their patient guidance in developing this dissertation and for their understanding and care. This work would not have been possible without the kind support and help from them. Their kindness and magnanimity have definitely influenced my life. I thank Dr. Yu Zhou for his generous help in research. I benefited much from the collaboration with him. I am also sincerely thankful to Drs. V. A. Samaranayake and Akim Adekpedjou for their teaching, advice and support in my job hunting process. I also greatly appreciate Drs. John Singler and Xiaoping Du's advice on my comprehensive exam and dissertation.

My experience with the Department of Mathematics and Statistics, Missouri University of Science and Technology has been very good and helped me to grow. I would like to thank all faculty members, staff and students here for their help and advice. In addition, my friends Renren Zhao and Fujian Tang's encouragements were very important to me.

Finally, I thank my parents Qi Liu and Sifen Guo for their faith and support in me. Whatever happens they will always provide me strong backing.

TABLE OF CONTENTS

Page

LIST OF ILLUSTRATIONS

LIST OF TABLES

1. INTRODUCTION

1.1. DIMENSION REDUCTION

Regression analysis investigates the dependence of a response Y on a vector $\mathbf{X} = (X_1, \ldots, X_p)^T$. The general object of interest is the conditional distribution of $Y|\mathbf{X}$, as a function of the value assumed by input $\mathbf{X}$. A parametric model is often used to guide the analysis. When there are no persuasive models available, nonparametric regression techniques can be used. If there are sufficiently many data points, local smoothing (Eubank 1988) is a successful approach. However, nonparametric regression is known to perform poorly in high dimensions due to the curse of dimensionality (Bellman, 1961). This phenomenon is manifest in practice due to the limited sample size of most real data sets. The curse of dimensionality refers to the fact that as the dimension of input $\mathbf{X}$ increases, the volume of the input space increases exponentially and as such the available data becomes increasingly sparse, see for instance Hastie et al. (2011, Section 2.5). Fortunately, it is very common for the input space $\mathbf{X}$ to essentially lie in a lower dimensional manifold or surface. This input space could be non-linear as is the case for manifold learning techniques (Izenman, 2013, Chapter 16) or linear as we shall assume here. Dimension reduction is a very common way to avoid the curse of dimensionality. Furthermore, many graphical tools can only help us to view low-dimension data directly. Even three-dimensional scatter plots must be constructed via a computer program, and the third dimension can only be visualized by rotating the coordinate axes. One tool for visualizing high dimensional data is the plots from section 6.3 of Hastie et al. (2011). These plots are of two variables conditional on the value of some other variables. However, when $\mathbf{X}$ is high-dimensional, it becomes increasingly challenging to construct and visually interpret conditional plots without reducing the dimension of data. Therefore, even in terms of data visualization, methods for dimension reduction are clearly called for.

The most common linear dimension reduction technique for $\mathbf{X}$ is principal component analysis (PCA) (Jolliffe, 2002), which performs a linear mapping of the data to a lower-dimensional space in such a way that the variance of the data in the low-dimensional

representation is maximized. In practice, the correlation matrix of the data is constructed and the eigenvectors on this matrix are computed. The eigenvectors that correspond to the largest eigenvalues (the principal components) can now be used to reconstruct a large fraction of the variance of the original data. The input space has been reduced to the space spanned by a few eigenvectors. 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, for a typical regression problem with a univariate response Y and a p-dimensional random vector predictor $\mathbf{X}$, PCA and other related methods such as independent components analysis often yield inferior results, when one aims to reduce the dimension of $\mathbf{X}$. This is because PCA, which is said to be an unsupervised dimension reduction technique, does not take into account the information in Y.

In contrast, a supervised dimension reduction technique takes into account Y when forming linear combination of the inputs (Hastie et al., 2011, Section 3.5.2). For example, Partial Least Squares (PLS) defines a relationship between Y and $\mathbf{X}$ which is determined by the values of both $\mathbf{X}$ and Y (Abdi, 2003, and Maitra and Yan, 2008). The PLS technique works by successively extracting factors from $\mathbf{X}$ and Y such that covariance between the extracted factors is maximized. However, in practice, it tends to yield results very similar to unsupervised principal component regression of Y onto principal components of $\mathbf{X}$.

Modern sufficient dimension reduction techniques to be discussed next, constitute a form of supervised linear dimension reduction which out perform PLS and PCA without the underlying model assumptions. It is a very active area of research.

1.2. SINGLE POPULATION SUFFICIENT DIMENSION REDUCTION

Li (1991) and Cook (1998) proposed sufficient dimension reduction that aims at reducing the dimension of $\mathbf{X}$ while preserving the regression relationship between Y and $\mathbf{X}$.

Sufficient dimension reduction could be applied on a lot of models. Consider the following generalized multi-index model

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

where $g(\cdot)$ is an unknown link function, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d)$ is a $p \times d$ matrix, $d \leq p$, and the random error ϵ is independent of $\mathbf{X}$. Model (1.1) is a very general semiparametric model which includes the multi-index model (Härdle and Stoker, 1989; Xia, 2008) and the single-index model (Härdle, Hall and Ichimura, 1993; Xue and Zhu, 2006) with $Y = g(\boldsymbol{\beta}^T\mathbf{X}) + \epsilon$ as special cases. Specifically, the scope of sufficient dimension reduction is to seek a set of linear combinations of $\mathbf{X}$, say $\boldsymbol{\beta}^T\mathbf{X}$, such that

$$Y \perp\!\!\!\perp \mathbf{X} | \boldsymbol{\beta}^T\mathbf{X}, \tag{1.2}$$

where $\perp\!\!\!\perp$ denotes stochastic independence. When this is the case, the projection of the p-dimensional explanatory variable $\mathbf{X}$ onto the d dimensional subspace $(\boldsymbol{\beta}_1^T\mathbf{X}, \cdots, \boldsymbol{\beta}_d^T\mathbf{X})^T$ contains all the information contained in the regression of Y on $\mathbf{X}$. Any linear combination of the $\boldsymbol{\beta}$'s is called an effective dimension-reduction (e.d.r.) direction, and the linear space generated by the $\boldsymbol{\beta}$'s is the e.d.r. space $\mathcal{S}_{edr}$ (Li, 1991) which is also known as the dimension reduction subspace (Cook, 1998).

Note that, the e.d.r. space is not unique since any space that contains the e.d.r. space is also an e.d.r. space. So one desires the smallest one which corresponds to the intersection of all the e.d.r. spaces satisfying $Y \perp\!\!\!\perp \mathbf{X} | \boldsymbol{\beta}^T\mathbf{X}$, which is defined as $\mathcal{S}_{Y|\mathbf{X}} = \cap\mathcal{S}_{edr}$, and is called the *central subspace* (Cook, 1998). The dimension of the central subspace:

$$\dim(\mathcal{S}_{Y|\mathbf{X}}) = d$$

is called the structural dimension of the regression. The goal of sufficient dimension reduction is to make inference about the central subspace and its dimension d.

The central subspace is well defined under very mild conditions (Cook, 1994, 1996, 1998b; Yin, Li and Cook, 2008). Specifically, if $\mathbf{X}$ has density $f(a) > 0$ for all $a \in \Omega$

and $f(a) = 0$ everywhere else, where Ω is the support of $\mathbf{X}$, $\Omega = \{\mathbf{x} \| \mathbf{f}(\mathbf{x}) > \mathbf{0}\}$, then the existence central subspace is guaranteed. We shall make the reasonable assumption that the central subspace exists throughout this dissertation.

A short example is given below which demonstrates the concepts of central subspace and the structural dimension. Suppose the true model is:

$$Y = exp((X_1 + X_2 + 1)(2X_3 - X_4) + 1) + \epsilon$$

where $\mathbf{X} = (X_1, X_2, \ldots, X_{15})^T$, and $\epsilon \perp\!\!\!\perp \mathbf{X}$. Then, the central subspace is the space spanned by the column vectors of $\boldsymbol{\beta}$ and

$$\boldsymbol{\beta} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 2 & -1 & 0 & \ldots & 0 \end{pmatrix}^T$$

and the structural dimension d is equal to 2. Sufficient dimension reduction tells us that we can replace the original 15-dimensional predictors $\mathbf{X}$ with the two linear combinations of the predictors $\boldsymbol{\beta}^T\mathbf{X}$ without the loss of information about the regression. In this example,

$$\boldsymbol{\beta}^T\mathbf{X} = \begin{pmatrix} X_1 + X_2 \\ 2X_3 - X_4 \end{pmatrix}.$$

Sufficient dimension reduction has received considerable interest in recent years due to the ubiquity of large high-dimension data sets which are now more readily available than in the past. Many methods have been developed, including sliced inverse regression (SIR; Li, 1991), sliced average variance estimator (SAVE; Cook and Weisberg, 1991), minimum average variance estimators (MAVE; Xia et al., 2002), directional regression (DR; Li and Wang, 2007), and likelihood acquired directions (LAD; Cook and Farzani, 2009). Recently, Li et al. (2011) and Lee et al. (2013) proposed a form of nonlinear sufficient dimension reduction which seeks an arbitrary function $\phi(.)$ from $\mathbb{R}^p$ to $\mathbb{R}^d$ satisfying

$Y \perp\!\!\!\perp \mathbf{X}|\phi(\mathbf{X})$, which greatly generalizes condition (1.2). Ma and Zhu (2012, 2013), and Luo et al. (2014) also investigate the dimension reduction problem in a semiparametric estimation framework and derive the associated estimating equations.

Let $\boldsymbol{\mu} = E(\mathbf{X})$, $\boldsymbol{\Sigma} = \mathrm{Var}(\mathbf{X})$, and $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ be the standardized predictor. As Yu et al. (2012) points out, many moment based sufficient dimension reduction methods can be formulated as an eigen-decomposition problem:

$$\mathcal{M}\boldsymbol{\eta}_i = \lambda_i \boldsymbol{\eta}_i, \quad i = 1, \cdots, p, \tag{1.5}$$

where $\mathcal{M}$ is the $\mathbf{Z}$ scale method-specific candidate matrix. Assuming the *linearity condition* (Li, 1991) holds, which is a mild condition imposed on the marginal distribution of the predictors alone, the eigenvectors $(\boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_d)$ 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, by invariance, and $\mathcal{S}_{Y|\mathbf{X}} = \boldsymbol{\Sigma}^{-1/2} S_{Y|\mathbf{Z}}$ as described by Cook (1998), $\boldsymbol{\beta} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_d)$ forms a basis of $\mathcal{S}_{Y|\mathbf{X}}$. The linearity condition, which basically requires that $E(\mathbf{X}|\boldsymbol{\beta}^T\mathbf{X})$ be a linear function of $\boldsymbol{\beta}^T\mathbf{X}$, is a common assumption in dimension reduction methods and holds for elliptically contoured predictors (Eaton, 1986). Additionally, Hall and Li (1993) showed that as the number of predictors p increases, the linearity condition holds to a reasonable degree of approximation in many problems.

Since most of the commonly used sufficient dimension reduction methods that target $\mathcal{S}_{Y|\mathbf{Z}}$, make use of candidate matrices satisfying the above eigen-decomposition, we list some as follows:

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

1.3. MULTIPLE POPULATION SUFFICIENT DIMENSION REDUCTION

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 set of overall features.

For example, consider the AIS dataset discussed by Weisberg (2005), which contains information on the lean body mass L and other physical and hematological measurements, from 102 male and 100 female elite Australian athletes who trained at the Australian Institute of Sport. In section 2, we investigate how the relationship between body fat and various predictors varies across gender. Various attempts have been made to develop valid analyses for multiple data sets, including common principal component analysis (Common PCA; Flury, 1984 and 1988).

Flury (1984, 1988) proposed a method called the common PCA, a type of simultaneous principal component analysis for several groups. It estimates principal components simultaneously in different groups and results in a joint dimension reducing transformation. The common principal components model has been employed in genetics, climatology, ontogeny and other fields (Biok, 2002). Flury (1987) extended the common PCA to partial common PCA. Other common space models also have been proposed (Krzanowski, 1979; Schott, 1988, 1991; Biok, 2002). The drawback of common PCA is that it is unsupervised and doesn't take into account the information in Y, as is the case for single population PCA.

Another method is the partial dimension reduction, as proposed by Chiaromonte et al. (2002), can also be adapted to perform analysis for multiple data sets. Partial dimension reduction was originally proposed to facilitate dimension reduction in regressions with both continuous predictors ($\mathbf{X} \in \mathbb{R}^p$) and a categorical predictor (W) with $W = 1, \ldots, G$ which can play the role of a group identifier. The partial central subspace is defined as the intersection of all subspaces $\mathcal{S}$ satisfying

$$Y \perp\!\!\!\perp \mathbf{X} \mid (P_{\mathcal{S}}\mathbf{X}, W), \tag{1.6}$$

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. This subspace is defined as

$$\mathcal{S}^{(W)}_{Y|\mathbf{X}} = \bigoplus\nolimits_{w=1}^{G} \mathcal{S}_{Y^w|\mathbf{X}^w} \tag{1.7}$$

where $(Y^w, \mathbf{X}^w)$ represents a generic pair of $(Y, \mathbf{X})$ when $W = w$ and $\mathcal{S}_{Y^w|\mathbf{X}^w}$ is the central subspace within each subpopulation. Here, $\bigoplus$ denotes the direct sum among subspaces.

There are several methods developed in the literature to infer about the partial central subspace, such as the partial SIR (Chiaromonte et al., 2002), partial SAVE (Shao et al., 2009) and partial IRE (partial inverse regression estimator; Wen and Cook, 2007). We will discuss these methods in Section 3. As we will see, the partial central subspace approach comprises the related directions for all populations which is a direct sum of all the marginal central subspaces, but there are still some 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.

1.4. TESTING COMMON INDICES FOR MULTI-INDEX MODELS

In this dissertation, we generalize the sufficient dimension reduction paradigm from a single population (dataset) to several populations. Specifically, we focus on testing the hypothesis that the central subspace of a particular group is the same as that of any other group:

$$\mathcal{S}_{Y^1|\mathbf{X}^1} = \mathcal{S}_{Y^2|\mathbf{X}^2} = \ldots = \mathcal{S}_{Y^G|\mathbf{X}^G}, \tag{1.8}$$

where $(Y^g, \mathbf{X}^g)$ is a generic pair of $(Y, \mathbf{X})$ for the gth group, $g = 1, \cdots, G$.

Let's consider the AIS dataset, as previously discussed. The goal was to investigate how the relationship between the body fat and various predictors varies across gender. Suppose that subject matter knowledge and prior modeling experiences suggest that a

d-dimensional multi-index model of the form $Y = g(\boldsymbol{\beta}_1^T\mathbf{X}, \cdots, \boldsymbol{\beta}_d^T\mathbf{X}; \epsilon)$ 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. If null hypothesis is true, then there is a common lower dimensional representation for the male and female groups. Informal comparisons such as graphical methods can of course be carried out. However, such comparisons might become unwieldy when d is greater than two, and the resulting conclusions could be overly subjective. Hence, a formal test is necessary here.

In Section 2, we will propose a link-free test statistic and its asymptotic distribution for testing (1.8) with $G = 2$ populations and apply our method to the numerical studies. Modified partial dimension reduction methods are introduced in Section 3. Based on the modified methods, we present our test statistics, which associated asymptotic distribution, simulation studies and an application which tests null hypothesis (1.8) for $G \geq 2$ in Section 4. Conclusions, future work and drawbacks of our proposed methods will be illustrated in Section 5.

2. TESTING COMMON INDICES FOR TWO MULTI-INDEX MODELS

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

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

where ϵ is the random error which is independent of $\mathbf{X}$. One is usually concerned with estimation of indices $\boldsymbol{\beta}$, the total number of indices d and the link function $g(\cdot)$ (Feng and Zhu, 2012). 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):

$$\begin{aligned} Y &= g_1(\boldsymbol{\beta}_1^T \mathbf{X}, \cdots, \boldsymbol{\beta}_d^T \mathbf{X}; \epsilon_1), \qquad & for\ group\ 1; \\ Y &= g_2(\boldsymbol{\xi}_1^T \mathbf{X}, \cdots, \boldsymbol{\xi}_d^T \mathbf{X}; \epsilon_2), \qquad & for\ group\ 2. \end{aligned} \tag{2.1}$$

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

$$H_0 : \quad \text{span}(\boldsymbol{\beta}) = \text{span}(\boldsymbol{\xi}), \tag{2.2}$$

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 Schott (1991).

A hypothesis test of type (2.2) 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.

Recall the AIS dataset from Section 1. It 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. It is of interest to determine how the relationship between the body fat and various predictors varies with gender. Naturally, we would like to know if the equivalent set of indices of the hematological measurements serve for both genders. This question is the motivation for our development of a test statistic for the null hypothesis in (2.2).

We propose a link-free test for testing hypothesis of (2.2) via a sufficient dimension reduction approach (Li, 1991; Cook, 1998). As we discussed in Section 1, there are numerous suffcient dimension reduction approaches considered in the literature.

The rest of this section is organized as follows. In Section 2.1, we give a brief review of sufficient dimension reduction methods. Specifically, we focus on those methods based upon spectral decomposition approach (Wen and Cook, 2007). In Section 2.2, we present our link-free test statistic for null hypothesis (2.2). The asymptotic distribution of our test statistic is also discussed. We illustrate the performance of our method with Monte Carlo studies in Section 2.3. We then apply our method to the AIS dataset in Section 2.4. Brief conclusions are given in Section 2.5. For ease of exposition, we defer some technical details in Section 2.6.

2.1. 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 (1.4). In particular, we consider three commonly used sufficient dimension reduction methods: SIR, SAVE and DR.

Recall from Section 1.2 that $\boldsymbol{\Sigma} = \mathrm{Var}(\mathbf{X})$, $\boldsymbol{\mu} = \mathrm{E}(\mathbf{X})$, $\mathbf{Z}$ is the standardized predictor $\boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ and that many moment based sufficient dimension reduction methods may be formulated as the solution to an eigen-decomposition problem:

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

where $\mathcal{M}_z$ is the $\mathbf{Z}$ scale method-specific candidate matrix.

For sufficient dimension reduction methods SIR, SAVE and DR, candidate matrices are

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

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 weighted inner-product of $< \mathbf{a}, \mathbf{b} >= \mathbf{a}^T\boldsymbol{\Sigma}\mathbf{b}$, and not the regular dot product, which induces unnecessary difficulty to the development of our test statistic. In this section, we work directly with the orginal predictor $\mathbf{X}$, and, as such, use the following symmetric candidate matrices $\mathcal{M}$.

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

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 $\mathcal{M}$ form a basis of $\mathcal{S}_{Y|\mathbf{X}}$, and are orthonormal with respect to the regular inner-product. The corresponding sample version of $\mathcal{M}$, $\widehat{\mathcal{M}}$ 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 Li and Dong (2009), and Li et al. (2010) for further details.

2.2. A LINK-FREE TEST FOR COMMON INDICES

Throughout this article, we assume that Model (2.1) 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 $\widehat{\mathbf{\Sigma}}_g = \frac{1}{n_g}\sum_{i=1}^{n_g}(\mathbf{X}_i^g - \bar{\mathbf{X}}_g)(\mathbf{X}_i^g - \bar{\mathbf{X}}_g)^T$, $g = 1, 2$. Let $\mathcal{M}_g$ denote the method-specific candidate matrix for the gth population, $\lambda_{g1} \geq \lambda_{g2} \ldots \geq \lambda_{gd} > \lambda_{g,(d+1)} = \ldots = \lambda_{gp} = 0$ be the eigenvalues of $\mathcal{M}_g$, and $\boldsymbol{\eta}_{gi}$ be the normalized eigenvector corresponding to λ_{gi}. Define eigenprojections $\mathbf{P}_{gd} = \boldsymbol{\eta}_{g1}\boldsymbol{\eta}_{g1}^T + \cdots + \boldsymbol{\eta}_{gd}\boldsymbol{\eta}_{gd}^T$, $\mathbf{Q}_{gd} = I_p - \mathbf{P}_{gd}$, 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 Moore-Penrose generalized inverse of matrix $\mathbf{A}$, from the perturbation theory (Kato, 1966) and Tyler (1981), we then have

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

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

Note that the approach we take is similar to that in Yu et al. (2012) but is motivated by a set of methodologies developed in Schott (1988, 1991, 1997) for making inference about common principal component subspaces.

Let $W = \text{trace}(\mathbf{P}_{1d}\mathbf{Q}_{2d}\mathbf{P}_{1d})$, then 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 (2.2) 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 $\text{span}(\mathbf{P}_{1d}) \subseteq \text{span}(\mathbf{P}_{2d})$. Because $\text{span}(\mathbf{P}_{1d})$ and $\text{span}(\mathbf{P}_{2d})$ have common dimension d, $\mathbf{P}_{1d} = \mathbf{P}_{2d}$ and hence (2.2) is true.

On the other hand, $\text{span}(\boldsymbol{\beta}) = \text{span}(\boldsymbol{\eta})$ if and only if $\mathbf{P}_{1d} = \mathbf{P}_{2d}$, which (2.2) implies that $\mathbf{P}_{1d}\mathbf{Q}_{2d} = 0$, and hence $W = 0$. □

We consider $\widehat{\mathbf{P}}_{1d}\widehat{\mathbf{Q}}_{2d}$, where $\widehat{\mathbf{Q}}_{2d} = I_p - \widehat{\mathbf{P}}_{2d} = \sum_{l=d+1}^{p} \hat{\boldsymbol{\eta}}_{gl}\hat{\boldsymbol{\eta}}_{gl}^T$, and let $T = n\widehat{W}$, where $\widehat{W} = trace(\widehat{\mathbf{P}}_{1d}\widehat{\mathbf{Q}}_{2d}\widehat{\mathbf{P}}_{1d})$. Let $n = n_1 + n_2$, $a_1 = \frac{n}{n_1}$, $a_2 = \frac{n}{n_2}$. As Yu et al. (2012) pointed out, $\mathbf{A}_g = \widehat{\mathcal{M}}_g - \mathcal{M}_g$ can be expressed via influence function approach as: $\mathbf{A}_g = \widehat{\mathcal{M}}_g - \mathcal{M}_g = \text{E}_{n_g}[\mathcal{M}_g^*(\mathbf{X}^g, Y^g)] + o_p(n^{-1/2})$, where $\text{E}_n\{\mathbf{.}\} = \frac{1}{n}\sum_{i=1}^{n}\{\mathbf{.}\}$ denotes the empirical expectation. This approach is key for the further development of our test statistic. The explicit formulas for $\mathcal{M}_g^*$ in the modified SIR, SAVE and DR are given in Section 2.6. Let $\text{vec}(\mathbf{A})$ denote the vec 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 2.1. *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* (2.2)*, we have*

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

where $\boldsymbol{\Psi} = a_1\boldsymbol{\Psi}_1 + a_2\boldsymbol{\Psi}_2$*,* $\boldsymbol{\Psi}_g = \mathbf{U}_g\boldsymbol{\Phi}_g\mathbf{U}_g^T$*,* $\boldsymbol{\Phi}_g = \text{E}\{\,\text{vec}(\mathcal{M}_g^*(\mathbf{X}^g, Y^g))\,\text{vec}(\mathcal{M}_g^*(\mathbf{X}^g, Y^g))^T\}$ *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 $\mathbf{R}_g = \sum_{i=1}^{d}[\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}^T\mathbf{A}_g(\lambda_{gi}I - \mathcal{M}_g)^+ + (\lambda_{gi}I - \mathcal{M}_g)^+\mathbf{A}_g\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}^T]$ for $g = 1, 2$.

Therefore

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

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

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

According to the null hypothesis (2.2), $\mathbf{P}_{1d}=\mathbf{P}_{2d}$, hence $\mathbf{P}_{1d}\mathbf{Q}_{2d}=0$. Then

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

Since $(\lambda_{gi}I-\mathcal{M}_g)^+=\sum_{k=1,\lambda_{gk}\neq\lambda_{gi}}^{p}(\lambda_{gi}-\lambda_{gk})^{-1}\boldsymbol{\eta}_{gk}\boldsymbol{\eta}_{gk}^T$ 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,

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

By the Multivariate Central Limit Theorem, we can conclude

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

so

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

that is

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

Since group 1 and group 2 are independent, we then have the following conclusion in the lemma:

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

□

This result leads to following result concerning the asymptotic distribution of our test statistic $T = n\widehat{W}$.

Theorem 2.2. *Assume the conditions of Proposition 1 hold, then under null hypothesis* (2.2)*, we have*

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

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

Proof:

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

By Lemma 2.1, under null hypothesis (2.2), the conclusion is obvious. □

A consistent estimate of $\mathbf{\Psi}$, $\widehat{\mathbf{\Psi}}$ can be obtained by substituting sample estimates for the unknown quantities. The weights ω_i's can be consistently estimated using the eigenvalues of $\widehat{\mathbf{\Psi}}$. 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 (2000) to approximate the tail probabilities.

2.3. SIMULATION STUDIES

Throughout our simulation studies, the random error ϵ is assumed to be standard normal and 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.

2.3.1. Estimated Test Levels. In this subsection, we evaluate the performance of our test statistic under three different models when the null hypothesis (2.2) holds.

2.3.1.1. Model I. We first consider the following model with one dimensional structure for both groups. The predictor vector $\mathbf{X} = (X_1, \cdots, X_p)$ 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 2.1 shows the estimated test levels for our three test statistics. The test levels are given in terms of percentages. As the group sizes n_1 and n_2 increase, the estimated levels are 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 deteriorate as p increases. All three dimension reduction methods perform reasonably well for all combinations of p and n.

Table 2.1. Estimated Test Levels (in percentages) for Model I

	Model I with $p = 4$				Model I with $p = 8$			
		Nominal Level (%)				Nominal Level (%)		
Sample Size	Test	1	5	10	Test	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

2.3.1.2. Model II. In this model, the predictor vector $\mathbf{X} = (X_1, \cdots, X_p)$ follows a multivariate normal distribution with mean 0, and the correlation between X_i and X_j is $0.5^{|i-j|}$, $i = 1, \cdots p$, $j = 1, \cdots p$. 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 2.2 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 $n_1 = n_2 = 600$, method DR performs the best.

Table 2.2. Estimated Test Levels (in percentages) for Model II

	Model II with $p = 4$				Model II with $p = 8$			
		Nominal Level (%)				Nominal Level (%)		
Sample Size	Test	1	5	10	Test	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

2.3.1.3. 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 independent and identically distributed standard normals. Several versions of this model were studied by Li (1991), Wen and Cook (2009) and others. This is a difficult test case for dimension reduction since some predictors are skewed or heavy tailed, and so are prone to outliers. As shown in Table 2.3, the performance of our test statistics for all of the three dimension reduction approaches is acceptable.

Table 2.3. Estimated Test Levels (in percentages) for Model III

	Model III with $p = 4$				Model III with $p = 8$			
		Nominal Level (%)				Nominal Level (%)		
Sample Size	Test	1	5	10	Test	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

2.3.1.4. Model IV.

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

This model considers a one-dimensional model with symmetric structure in $\mathbf{X}$ which is drawn from standard multivariate normal distribution. Table 2.4 presents the estimated significance levels for Model IV. Because SIR is known to fail when the response surface is symmetric about the origin, it comes as no surprise that the estimated test levels for this model using SIR are relatively far from nominal levels, while test methods using SAVE and DR candidate matrices both perform well.

Table 2.4. Estimated Test Levels (in percentages) for Model IV

	Model IV with $p = 4$				Model IV with $p = 8$			
		Nominal Level (%)				Nominal Level (%)		
Sample Size	Test	1	5	10	Test	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

2.3.2. Estimated Power. We examine the power of our test under the alternative hypotheses defined in this subsection. Two models are considered. The predictors $\mathbf{X}$ for both models follow the standard multivariate normal.

2.3.2.1. Model V.

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

The two populations in this model have the same structural dimension $d = 2$, however, they don't share the same set of indices. We can see from Table 2.5 that when d is correctly specified as 2, the power in percentages, 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.

Table 2.5. Estimated Power at 5% Nominal Levels for Model V

	SIR		SAVE		DR	
Sample Size	$p = 4$	$p = 8$	$p = 4$	$p = 8$	$p = 4$	$p = 8$
$n_1 = n_2 = 200$	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

2.3.2.2. Model VI.

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

Model VI is two-dimensional when $p > 4$. 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 2.6 shows that the power of our test with $d = 2$. As we can see, the power of our test increases as the sample size n increases which was also the case for Model **V**. Again, our tests perform reasonably well and are able to detect the different indices between the two populations for a very high percentage of simulated data sets.

Table 2.6. Estimated Power at 5% Nominal Levels for Model VI

	SIR		SAVE		DR	
Sample Size	$p = 4$	$p = 8$	$p = 4$	$p = 8$	$p = 4$	$p = 8$
$n_1 = n_2 = 200$	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

2.4. APPLICATION TO THE AIS DATA

We return to the AIS dataset discussed in Section 1. This data set was originally introduced by Nevill and Holder (1995), and studied by Cook and Weisberg (1999b), Chiaromonte, Cook and Li (2002) and others. As previous noted, 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 et al. (2002) show that there is only one relevant linear combination of predictors for both male and female groups, so a one-dimensional multi-index model of the form (1.4) 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 (2.2) at significance level 0.05 (p-values are 0.76, 0.54 and 0.65 for SIR, SAVE and DR, respectively).

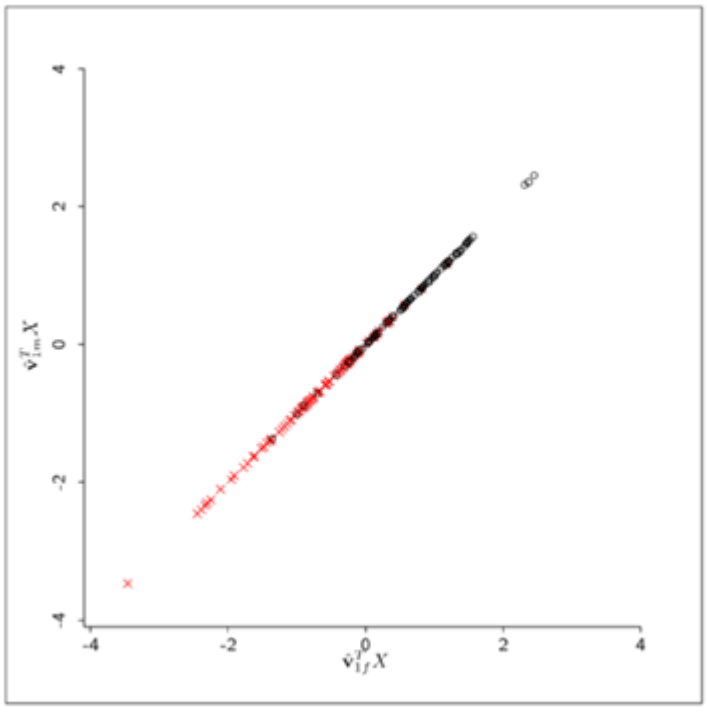

Figure 2.1. AIS Data: $\hat{\mathbf{v}}_{1m}^T \mathbf{X}$ vs. $\hat{\mathbf{v}}_{1f}^T \mathbf{X}$, $\times$ for females, $\circ$ for males.

Our result is consistent with that of Chiaromonte et al. (2002) where the same conclusion was drawn via an informal analysis. The authors also 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. Figure 2.1 shows the summary plot of the estimated SIR predictors for males and females.

2.5. SUMMARY

In this section, 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 from dependent variable Y, 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.

Finally, we develop a method of determining how similar the two groups are with respect to the span of the indices of predictors vectors. The method provides us a convenient way to compare two groups. The idea we presented here also opens the way for the comparison for more groups.

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 indices which is consistent with work in Chiaromonte et al. (2002).

2.6. EXPLICIT CANDIDATE MATRIX FORMULAS

The explicit formulas of candidate matrices for SIR, SAVE and DR are given in this section. It suffices to derive the expansion of $\widehat{\mathcal{M}}_g$ for the gth population and as such 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 Yu et al. (2012).

First we divide the range of Y into h slices $\{J_1, \cdots, J_h\}$. 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 $\widehat{\mathcal{M}}$.

Lemma 2.3. *Consider Frechet derivatives* $\boldsymbol{\Sigma}^* = (\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^T - \boldsymbol{\Sigma}$, $\widehat{\boldsymbol{\Sigma}}^{*-1} = -\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Sigma}^*\widehat{\boldsymbol{\Sigma}}^{-1}$, $\boldsymbol{\mu}^* = \mathbf{X} - \boldsymbol{\mu}$, $p_k^* = I(Y \in J_k) - p_k$, $\mathbf{U}_k^* = \mathbf{X}R_k - \mathbf{U}_k - \mathbf{X}p_k - \boldsymbol{\mu}R_k + \boldsymbol{\mu}p_k$, $\mathbf{V}_k^* = \mathbf{X}\mathbf{X}^T R_k - \mathrm{E}\left[\mathbf{X}\mathbf{X}^T R_k\right] - \mathrm{E}\left[\mathbf{X}R_k\right]\mathbf{X}^T - (\mathbf{X}R_k - 2\mathrm{E}\left[\mathbf{X}R_k\right])\boldsymbol{\mu}^T - \mathbf{X}\mathrm{E}\left[\mathbf{X}^T R_k\right] - \boldsymbol{\mu}\left(\mathbf{X}^T R_k - 2\mathrm{E}\left[\mathbf{X}^T R_k\right]\right) + (\mathbf{X} - \boldsymbol{\mu})\boldsymbol{\mu}^T \mathrm{E}\left[R_k\right] + \boldsymbol{\mu}\left(\mathbf{X} - \boldsymbol{\mu}\right)^T \mathrm{E}\left[R_k\right] + \boldsymbol{\mu}\boldsymbol{\mu}^T\left(R_k - \mathrm{E}\left[R_k\right]\right)$.

Then we have the following Von-mises 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 Horn and Johnson (1991), and Li and Wang (2007).

Following the latter, 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 $R_k = I\left(Y \in J_k\right)$, we have:

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

Also,

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

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,

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

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

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

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

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

All of this yields

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

□

2.6.1. Asymptotic Expansion of $\widehat{\mathcal{M}}_{SIR}$. 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 $\mathcal{M}_{SIR} = \boldsymbol{\Sigma}^{-1}\Lambda_{SIR}\boldsymbol{\Sigma}^{-1}$. 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 $\widehat{\mathcal{M}}_{SIR} = \widehat{\boldsymbol{\Sigma}}^{-1}\widehat{\Lambda}_{SIR}\widehat{\boldsymbol{\Sigma}}^{-1}$. The explicit expansion forms of $\widehat{\Lambda}_{SIR}$ and $\widehat{\mathcal{M}}_{SIR}$ follow,

Lemma 2.4. *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 2.5. $\widehat{\mathcal{M}}_{SIR}$ *can be expanded asymptotically as* $\widehat{\mathcal{M}}_{SIR} = \mathcal{M}_{SIR} + E_n(\mathcal{M}^*_{SIR}) + o_p(n^{-1/2})$, *where* $\mathcal{M}^*_{SIR} = \boldsymbol{\Sigma}^{*-1}\Lambda_{SIR}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\Lambda^*_{SIR}\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}^{-1}\Lambda_{SIR}\boldsymbol{\Sigma}^{*-1}$.

Proof of Theorem 2.5.

With the expansion given in Lemma 2.3, the conclusion can be easily derived by invoking Lemma 2.4. □

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

Lemma 2.6. $\Lambda_{SAVE} = \mathbf{\Sigma}\Lambda_{SIR} + \Lambda_{SIR}\mathbf{\Sigma} - \mathbf{\Sigma}^2 + \mathbf{\Gamma}$, *where* $\mathbf{\Gamma} = \sum_{l=1}^{h}(\mathbf{\Gamma}_l^1 - \mathbf{\Gamma}_l^2 - \mathbf{\Gamma}_l^3 + \mathbf{\Gamma}_l^4)$ *with* $\mathbf{\Gamma}_l^1 = \frac{\mathbf{V}_l{}^2}{p_l}$, $\mathbf{\Gamma}_l^2 = \frac{\mathbf{V}_l\mathbf{U}_l\mathbf{U}_l^T}{p_l^2}$, $\mathbf{\Gamma}_l^3 = \frac{\mathbf{U}_l\mathbf{U}_l^T\mathbf{V}_l}{p_l^2}$ *and* $\mathbf{\Gamma}_l^4 = \frac{\mathbf{U}_l\mathbf{U}_l^T\mathbf{U}_l\mathbf{U}_l^T}{p_l^3}$.

Proof of Lemma 2.6.

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

With more algebraic calculations, one can easily derive the stated result. □

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

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

Lemma 2.7. *Let* $\mathbf{\Gamma}^* = \sum_{l=1}^{h}\{(\mathbf{\Gamma}_l^1)^* - (\mathbf{\Gamma}_l^2)^* - (\mathbf{\Gamma}_l^3)^* + (\mathbf{\Gamma}_l^4)^*\}$ *and* $\Lambda_{SAVE}^* = \mathbf{\Sigma}\Lambda_{SIR}^* + \mathbf{\Sigma}^*\Lambda_{SIR} + \Lambda_{SIR}\mathbf{\Sigma}^* + \Lambda_{SIR}^*\mathbf{\Sigma} - \mathbf{\Sigma}\mathbf{\Sigma}^* - \mathbf{\Sigma}^*\mathbf{\Sigma} + \mathbf{\Gamma}^*$. *Then we have* $\widehat{\Lambda}_{SAVE} = \Lambda_{SAVE} + \mathrm{E}_n(\Lambda_{SAVE}^*) + o_p(n^{-1/2})$.

Proof of Lemma 2.7.

The conclusion can be derived by Lemmas 2.3, 2.4 and 2.6. Details are omitted.

□

Theorem 2.8. *$\widehat{\mathcal{M}}_{SAVE}$ can be expanded asymptotically as*

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

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

Proof of Theorem 2.8.

With the expansion in Lemma 2.3, the conclusion can be easily derived by invoking Lemma 2.7. □

2.6.3. Asymptotic Expansion of $\widehat{\mathcal{M}}_{DR}$. The candidate matrix of directional regression is

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

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

Lemma 2.9. *$\mathcal{M}_{DR}$ can be reformulated as $\mathcal{M}_{DR} = \mathbf{\Sigma}^{-1}\Lambda_{DR}\mathbf{\Sigma}^{-1}$, where*

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

Proof of Lemma 2.9. The conclusion can be derived by further algebraic calculations. We omit the details here. □

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

Lemma 2.10. *Define Frechet derivative*

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

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

Proof of Lemma 2.10. The conclusion can be derived by Lemmas 2.3, 2.4 and 2.9. Details are omitted. □

Theorem 2.11. *$\widehat{\mathcal{M}}_{DR}$ can be expanded asymptotically as*

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

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

Proof of Theorem 2.11. With the expansion given in Lemma 2.3, the conclusion can be easily derived by invoking Lemma 2.10. □

3. PARTIAL DIMENSION REDUCTION

For a random sample of size n_g from the gth population $(Y^g, \mathbf{X}^g)$, $g = 1, \ldots, G$, let $\bar{\mathbf{X}}_g = \frac{1}{n_g}\sum_{i=1}^{n_g}\mathbf{X}_i^g$, $\widehat{\mathbf{\Sigma}}_g = \frac{1}{n_g}\sum_{i=1}^{n_g}(\mathbf{X}_i^g - \bar{\mathbf{X}}_g)(\mathbf{X}_i^g - \bar{\mathbf{X}}_g)^T$, and $n = \sum_{g=1}^{G} n_g$. Standardize the predictor, $\mathbf{Z}_i^g = \widehat{\mathbf{\Sigma}}_g^{-\frac{1}{2}}(\mathbf{X}_i^g - \bar{\mathbf{X}}_g)$, $i = 1, \cdots, n_g$, $g = 1, \cdots, G$. Following the common practice in sufficient dimension reduction, we partition the range of Y^g into h_g slices, and calculate the intraslice mean vectors as

$$\bar{\mathbf{Z}}_{gs} = \frac{1}{n_{gs}}\sum_{j|s}\mathbf{Z}_j^g, \quad s = 1, \cdots, H_g, \quad g = 1, \cdots, G, \tag{3.1}$$

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 notation, in the following discussions, we use $\{Y^g = s\}$ as short for $\{Y^g \text{is in slice s}\}$.

3.1. MODIFIED PARTIAL SLICED INVERSE REGRESSION

The original partial SIR proposed by Chiaromonte et al. (2002) requires the following homogeneous predictor covariance condition across the populations:

$$\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \ldots = \mathbf{\Sigma}_G.$$

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

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

$$\widehat{\mathcal{M}}_g^{sir} = \sum_{s=1}^{h_g}\frac{n_{gs}}{n_g}\bar{\mathbf{Z}}_{gs}\bar{\mathbf{Z}}_{gs}^T, \quad g = 1, \cdots, G. \tag{3.2}$$

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 $\widehat{\mathbf{H}}_g = (\hat{f}_{g1}\bar{\mathbf{Z}}_{g1}, \ldots, \hat{f}_{gh_g}\bar{\mathbf{Z}}_{gh_g})$. Then $\widehat{\mathcal{M}}_g^{sir} = \widehat{\mathbf{H}}_g\widehat{\mathbf{H}}_g^T$. Averaging the sample candidate matrices over each population, we obtain

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

where $\widehat{\mathbf{H}} = (\alpha_1\widehat{\mathbf{H}}_1, \ldots, \alpha_G\widehat{\mathbf{H}}_G)$. 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 $\mathbf{H}$:

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

where $\begin{pmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_0 \end{pmatrix}$ is a $p \times p$ orthogonal matrix in which $\boldsymbol{\Gamma}_1$ and $\boldsymbol{\Gamma}_0$ have dimensions $p \times m$ and $p \times (p-m)$, $\begin{pmatrix} \boldsymbol{\Psi}_1 & \boldsymbol{\Psi}_0 \end{pmatrix}$ 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 (1994), under the null hypothesis $d = m$, $T_{sir}(m)$ has the same asymptotic distribution as

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

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

Lemma 3.1.

$$\sqrt{n}vec[\boldsymbol{\Gamma}_0^T(\widehat{\mathbf{H}} - \mathbf{H})\boldsymbol{\Psi}_0] \xrightarrow{\mathcal{D}} \mathrm{Normal}(0, \boldsymbol{\Omega}),$$

where $\otimes$ denotes the Kronecker product, $\boldsymbol{\Omega} = (\boldsymbol{\Psi}_0^T \otimes \boldsymbol{\Gamma}_0^T)\mathrm{diag}(\boldsymbol{\Delta}_1, \ldots, \boldsymbol{\Delta}_G)(\boldsymbol{\Psi}_0 \otimes \boldsymbol{\Gamma}_0)$, $\boldsymbol{\Delta}_g$, $g = 1, \ldots G$, are defined by Equation (8) in Bura and Cook (2001), and $\mathrm{diag}(\centerdot)$ *denotes a positive definite block diagonal matrix.*

Proof: By Equation (8) of Bura and Cook (2001), we have the following result:

$$\sqrt{n_g}\,\mathrm{vec}[(\widehat{\mathbf{H}}_g - \mathbf{H}_g)] \xrightarrow{\mathcal{D}} \mathrm{Normal}(0, \boldsymbol{\Delta}_g), \tag{3.3}$$

where $\boldsymbol{\Delta}_g$ is defined in Bura and Cook (2001). Hence:

$$\sqrt{n}vec[\boldsymbol{\Gamma}_0^T(\widehat{\mathbf{H}} - \mathbf{H})\boldsymbol{\Psi}_0] \xrightarrow{\mathcal{D}} \mathrm{Normal}(0, \boldsymbol{\Omega}).$$

□

In sufficient dimension reduction, estimation of d is often based on testing a sequence of hypotheses $H_0 : d = m$ versus $H_a : d > m$, 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 $H_0 : d = m$ versus $H_a : d > m$.

Theorem 3.2. *Assuming the linearity condition for $\mathbf{X}_g$, $g = 1, \ldots, G$, under the null hypothesis of $H_0 : d = m$, the limiting distribution of $T_{sir}(m)$ is the same as that of*

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

where $h = \sum_{g=1}^{G} h_g$ is the total number of slices, the $K_i's$ are iid χ^2_1, and $\omega_1 \geq \cdots \geq \omega_{(h-m)(p-m)}$ are the ordered eigenvalues of $\boldsymbol{\Omega}$.

The proof of Theorem 3.2 is straight forward following Lemma 3.1, hence is omitted here.

3.2. PARTIAL SLICED AVERAGE VARIANCE ESTIMATION

For partial SAVE, the sample version of $\mathcal{M}_g$ for population g is given by

$$\widehat{\mathcal{M}}_g^{save} = \sum_{s=1}^{h_g} \frac{n_{gs}}{n_g}[I_p - \widehat{cov}(\mathbf{Z}^g | Y^g = s)]^2, \tag{3.4}$$

where $\widehat{cov}(\mathbf{Z}^g | Y^g = s)$ is the sample variance in population g, with Y^g falling into the s slice, $s = 1, \ldots, h_g$.

Averaging the sample candidate matrices over each population, we obtain $\widehat{\mathcal{M}}^{save} = \sum_{g=1}^{G} \frac{n_g}{n} \widehat{\mathcal{M}}_g^{save}$, which can be spectrally decomposed as:

$$\widehat{\mathcal{M}}^{save} = \sum_{k=1}^{p} \hat{\lambda}_k \hat{\boldsymbol{\eta}}_k \hat{\boldsymbol{\eta}}_k^T,$$

where $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p \geq 0$ are the ordered eigenvalues of $\widehat{\mathcal{M}}^{save}$, and $\hat{\boldsymbol{\eta}}_k$, $k = 1, \ldots, p$ are the corresponding eigenvectors. Let $\hat{\boldsymbol{\eta}} = (\hat{\boldsymbol{\eta}}_{m+1}, \cdots, \hat{\boldsymbol{\eta}}_p)$, define

$$T_{save}(\hat{\boldsymbol{\eta}}) = \frac{1}{2} \sum_{g=1}^{G} \sum_{s=1}^{h_g} n_{gs} tr[(\hat{\boldsymbol{\eta}}^T (I_p - \widehat{cov}(\mathbf{Z}^g | Y^g = s)) \hat{\boldsymbol{\eta}})^2].$$

The following theorem (Shao et al. 2009) provides a test statistic for testing $H_0 : d = m$ versus $H_a : d > m$.

Theorem 3.3. *Assume that $\mathbf{X}_g's$ are normally distributed with the same covariance, then, under null hypothesis $H_0 : d = m$, $T_{save}(\hat{\boldsymbol{\eta}})$ follows chi-squared distribution with degrees of freedom of $(h - G)(p - d)(p - d + 1)$, where $h = \sum_{g=1}^{G} h_g$.*

3.3. PARTIAL DIRECTIONAL REGRESSION

Within the context of dimension reduction, Li and Wang (2007) proposed *directional regression* (DR) for a single population. In this subsection, we extend their result to multiple populations. Let $(\widetilde{\mathbf{Z}}^g, \widetilde{Y}^g)$ be an independent copy of $(\mathbf{Z}^g, Y^g)$, the sample version of $\mathcal{M}_g$ for population g is given by:

$$\begin{aligned}
\widehat{M}_g^{dr} &= 2 \sum_{s=1}^{h_g} E_n^2[\widehat{\mathbf{Z}}^g (\widehat{\mathbf{Z}}^g)^T - I_p | Y^g = s)] \hat{\phi}_{gs} \\
&\quad + 2[\sum_{s=1}^{h_g} E_n(\widehat{\mathbf{Z}}^g | Y^g = s) E_n((\widehat{\mathbf{Z}}^g)^T | Y^g = s) \hat{\phi}_{gs}]^2 \\
&\quad + 2 \sum_{s=1}^{h_g} E_n((\widehat{\mathbf{Z}}^g)^T | Y^g = s) E_n(\widehat{\mathbf{Z}}^g | Y^g = s) \hat{\phi}_{gs} \\
&\quad \times \sum_{s=1}^{h_g} E_n(\widehat{\mathbf{Z}}^g | Y^g = s) E_n((\widehat{\mathbf{Z}}^g)^T | Y^g = s) \hat{\phi}_{gs}, \qquad (3.5)
\end{aligned}$$

where notations such as $E_n(\widehat{\mathbf{Z}}^g|Y^g = s)$ stand for sample conditional moments, $\bar{\mathbf{Z}}_{gs}$.

We may rewrite $\widehat{\mathcal{M}}_g^{dr} = \widehat{\mathbf{H}}_g\widehat{\mathbf{H}}_g^T$, where $\widehat{\mathbf{H}}_g = (\widehat{\mathbf{H}}_{11}^g, \cdots, \widehat{\mathbf{H}}_{1h_g}^g; \widehat{\mathbf{H}}_2^g; \widehat{\mathbf{H}}_{31}^g, \cdots, \widehat{\mathbf{H}}_{3h_g}^g)$ is a $p \times (hp + h + p)$ matrix, see Li and Wang (2007) for details. Averaging the sample candidate matrices over each population, we obtain

$$\widehat{\mathcal{M}}^{dr} = \sum_{g=1}^{G} \frac{n_g}{n} \widehat{\mathcal{M}}_g^{dr} = \widehat{\mathbf{H}}\widehat{\mathbf{H}}^T,$$

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

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

Theorem 3.4. *Assuming the linearity condition for $\mathbf{X}_g$, $g = 1, \ldots, G$, the common covariance condition, and the coverage condition (See Li and Wang (2007) for details), under the null hypothesis of $H_0 : d = m$, the limiting distribution of $T_{dr}(m)$ is the same as that of*

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

where $h = \sum_{g=1}^{G} h_g$ is the total number of slices, the $K_i's$ are iid χ_1^2,
$\mathbf{\Omega} = (\mathbf{\Psi}_0^T \otimes \mathbf{\Gamma}_0^T)\text{diag}(\mathbf{\Delta}_1, \ldots, \mathbf{\Delta}_G)(\mathbf{\Psi}_0 \otimes \mathbf{\Gamma}_0)$, with $\mathbf{\Delta}_g$ defined by Equation (14) of Li and Wang (2007), and $\omega_1 \geq \cdots \geq \omega_{(hp+p+h-m)(p-m)}$ are the ordered eigenvalues of $\mathbf{\Omega}$.

The proof of Theorem 3.4 is straightforward following Theorem 5 of Li and Wang (2007), hence is omitted here.

4. TESTING COMMON INDICES FOR MULTI-INDEX MODELS

In this section, we focus on testing the hypothesis that the central subspace of a particular group is the same as that of any other group:

$$\mathcal{S}_{Y^1|\mathbf{X}^1} = \mathcal{S}_{Y^2|\mathbf{X}^2} = \ldots = \mathcal{S}_{Y^G|\mathbf{X}^G}, \tag{4.1}$$

where $(Y^g, \mathbf{X}^g)$ is a generic pair of $(Y, \mathbf{X})$ for the gth group, $g = 1, \cdots, G$, $G \geq 2$. Hence, we are interested in testing if the spaces spanned by the first d linear combinations of $\mathbf{X}$ are sufficient for each group. In Section 2, we considered an analogous testing procedure for two populations. The test statistics we propose in this section are, however, different from the previously discussed method even when $G = 2$.

Under the framework of the partial central subspace, our testing hypothesis (4.1) is equivalent to

$$\mathcal{S}_{Y^1|\mathbf{X}^1} = \mathcal{S}_{Y^2|\mathbf{X}^2} = \ldots = \mathcal{S}_{Y^G|\mathbf{X}^G} = \mathcal{S}^{(W)}_{Y|\mathbf{X}}. \tag{4.2}$$

Notice that although partial dimension reduction can be adapted to infer about multi-population dimension reduction problems, the partial central subspace it obtains is a direct sum of all the marginal central subspaces (Chiaromonte et al., 2002), which cannot deal with testing hypotheses such as (4.1).

The rest of this section is organized as follows. In Section 4.1, we give a quick review of sufficient dimension reduction methods for multiple populations. In Section 4.2 and Section 4.3, we present our test statistics for testing (4.1) for $G = 2$ and $G > 2$. The asymptotic distributions of these test statistics are also discussed. We illustrate the performance of our methods via simulation studies in Section 4.4. We then apply our method to the plasma retinol and beta-carotene data in Section 4.5. Brief conclusions are given in Section 4.6.

4.1. PARTIAL DIMENSION REDUCTION

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

$$\mathcal{S}^{(W)}_{Y|\mathbf{X}} = \bigoplus\nolimits_{g=1}^{G} \mathcal{S}_{Y^g|\mathbf{X}^g}, \tag{4.3}$$

where $\bigoplus$ indicates the direct sum between two subspaces. Chiaromonte et al. (2002) and Wen and Cook (2007) proposed estimation methods for the partial central subspace based on (4.3).

Under the null hypothesis (4.2), it is reasonable to assume that the dimensions of all the marginal central subspaces are equal to $d = \mathcal{S}_{Y^g|\mathbf{X}^g}$, $g = 1, \cdots, G$. We shall assume that d is known in what follows. An estimate of d can be easily obtained via any partial dimension reduction method including those discussed in Section 3 or the single population dimension reduction method discussed in Section 2.

For the multiple population setting, let $(Y^g_j, \mathbf{X}^g_j)$, $j = 1, \ldots, n_g$ be a simple random sample of size n_g from the gth population $(Y^g, \mathbf{X}^g)$ for $g = 1, \ldots, G$. Let $\bar{\mathbf{X}}_g = \frac{1}{n_g}\sum_{i=1}^{n_g} \mathbf{X}^g_i$, and $\widehat{\mathbf{\Sigma}}_g = \frac{1}{n_g}\sum_{i=1}^{n_g}(\mathbf{X}^g_i - \bar{\mathbf{X}}_g)(\mathbf{X}^g_i - \bar{\mathbf{X}}_g)^T$. In the following subsections, we propose test statistics for testing (4.1) with $G = 2$ populations and $G > 2$ populations.

4.2. TEST STATISTIC WITH TWO POPULATIONS

All the notation and symbols defined in Section 2 will be adopted in this section.

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 Chapter 3).

Following Section 2, we expand $\mathbf{A}_g = \widehat{\mathcal{M}}_g - \mathcal{M}_g$ via the influence function approach as:

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

where $E_n\{.\} = \frac{1}{n}\sum_{i=1}^{n}\{.\}$. The following lemma gives the asymptotic distribution of $\sqrt{n_g}\,\mathrm{vec}[\widehat{\mathbf{P}}(\widehat{\mathbf{P}}_{gd} - \mathbf{P}_{gd})]$:

Lemma 4.1. *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* (4.1)*, we have:*

$$\sqrt{n_g}\,\mathrm{vec}[\widehat{\mathbf{P}}(\widehat{\mathbf{P}}_{gd} - \mathbf{P}_{gd})] \xrightarrow{D} N(0, \mathbf{\Psi}_g),$$

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

Proof: Under the null hypothesis (4.1), the expected value of $\widehat{\mathbf{P}}$ is $\mathbf{P}_{gd}$, and $\lambda_{gk} = 0$, for $g = 1, \ldots, G$ and $k = d+1, \ldots, p$, so we have

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

Then the asymptotic distribution of $\sqrt{n_g}\,\mathrm{vec}[\widehat{\mathbf{P}}(\widehat{\mathbf{P}}_{gd} - \mathbf{P}_{gd})]$ is easily shown to be multivariate normal.

□

Let $t = \mathrm{vec}[\widehat{\mathbf{P}}(\widehat{\mathbf{P}}_{1d} - \widehat{\mathbf{P}}_{2d})]$, $c = \frac{n_1 n_2}{n_1+n_2}$, $c_1 = \frac{n_1}{n_1+n_2}$, $c_2 = \frac{n_2}{n_1+n_2}$, $\mathbf{\Psi} = c_2\mathbf{\Psi}_1 + c_1\mathbf{\Psi}_2$. Theorem 4.2 provides the asymptotic result concerning the test statistic $T = ct^T\widehat{\mathbf{\Psi}}^+t$, where $\widehat{\mathbf{\Psi}}$ is the sample estimate of $\mathbf{\Psi}$.

Theorem 4.2. *Assume that the data* $(\mathbf{X}_i^g, Y_i^g)$*, for* $g = 1, 2$*,* $i = 1, \ldots, n_g$*, are a simple random sample from* $(\mathbf{X}^g, Y^g)$ *with finite fourth order moments, then under null hypothesis* (4.1)*,* T *follows an asymptotically chi-squared distribution with degree of freedom of* $d(p-d)$*.*

4.3. TEST STATISTIC FOR $G > 2$

Let $\mathbf{t}_g = \mathrm{vec}[\widehat{\mathbf{P}}(\widehat{\mathbf{P}}_{gd} - \mathbf{P}_{gd})]$, $g = 1, \ldots, G$, $\bar{\mathbf{t}} = \sum_{i=1}^{G}\frac{n_i}{n}\mathbf{t}_i$ and $\widehat{\overline{\mathbf{P}}} = \sum_{i=1}^{G}\frac{n_i}{n}\widehat{\mathbf{P}}_{id}$, then

$$\mathbf{t}_g - \bar{\mathbf{t}} \;=\; \mathrm{vec}[\widehat{\mathbf{P}}(\widehat{\mathbf{P}}_{gd} - \widehat{\overline{\mathbf{P}}}) - \widehat{\mathbf{P}}(\mathbf{P}_{gd} - \sum_{i=1}^{G} \frac{n_i}{n}\mathbf{P}_{id})].$$

Also, let $\mathbf{m}_g \equiv \mathrm{vec}[\widehat{\mathbf{P}}(\widehat{\mathbf{P}}_{gd} - \widehat{\overline{\mathbf{P}}})]$, then, under null hypothesis (4.1), $\mathbf{m}_g$ has the same asymptotic distribution as $\mathbf{t}_g - \bar{\mathbf{t}}$.

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

$$\begin{aligned}
\sqrt{n}\mathbf{t} &= \sqrt{n}(\mathbf{I}_{p^2G} - \frac{1}{n}(n_1, \cdots, n_G) \otimes \mathbf{1}_G \otimes \mathbf{I}_{p^2})(\mathbf{t_1}^T, \cdots, \mathbf{t_G}^T)^T \\
&= \left(\mathrm{diag}\left(\sqrt{\frac{n}{n_1}}, \cdots, \sqrt{\frac{n}{n_G}}\right) - \frac{1}{\sqrt{n}}(\sqrt{n_1}, \cdots, \sqrt{n_G}) \otimes \mathbf{1}_G\right) \otimes \mathbf{I}_{p^2} \\
&\quad (\sqrt{n_1}(\mathbf{t}_1)^T, \cdots, \sqrt{n_G}(\mathbf{t}_G)^T)^T,
\end{aligned}$$

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

Let $r_g = \frac{n_g}{n}$, $g = 1, \ldots, G$. Assuming that as n goes to infinity, r_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 4.1, we could derive the asymptotic distribution of $\sqrt{n}\mathbf{m}$ which is the same as that of $\sqrt{n}\mathbf{t}$.

Theorem 4.3. *Assume that the data $(\mathbf{X}_i^g, Y_i^g)$, for $g = 1, 2$, $i = 1, \ldots, n_g$, are a simple random sample from $(\mathbf{X}^g, Y^g)$ with finite fourth order moments. Then under null hypothesis* (4.1)*, we have:*

$$\sqrt{n}\mathbf{m} \xrightarrow{D} N(0, \mathbf{BWB}^T),$$

where $\mathbf{B} = \left(diag\left(\sqrt{\frac{1}{r_1}}, \cdots, \sqrt{\frac{1}{r_G}}\right) - (\sqrt{r_1}, \cdots, \sqrt{r_G}) \otimes \mathbf{1}_G\right) \otimes \mathbf{I}_{p^2}$ *and* $\mathbf{W} = diag(\mathbf{\Psi}_1, \cdots, \mathbf{\Psi}_G)$.

Proof:

Lemma 4.1 tells us that under null hypothesis (4.1),

$$\sqrt{n_g}\mathbf{t_g} \xrightarrow{D} N(0, \mathbf{\Psi}_g).$$

And since $\widehat{\mathbf{P}}(\widehat{\mathbf{P}}_{gd} - \mathbf{P}_{gd}) = \sum_{i=1}^{d}\sum_{k=d+1}^{p} \lambda_{gi}^{-1}\boldsymbol{\eta}_{gi}\boldsymbol{\eta}_{gi}^T\mathbf{A}_g\boldsymbol{\eta}_{gk}\boldsymbol{\eta}_{gk}^T + o_p(n^{-\frac{1}{2}})$ is just related to the gth population, $\mathbf{t}_g, g = 1, \cdots, G$ are independent with each other. Then the asymptotic

covariance of $(\sqrt{n_1}\mathbf{t_1}^T, \cdots, \sqrt{n_G}\mathbf{t_G}^T)^T$ is $\mathbf{W} = \text{diag}(\mathbf{\Psi}_1, \cdots, \mathbf{\Psi}_G)$, that is

$$(\sqrt{n_1}\mathbf{t_1}^T, \cdots, \sqrt{n_G}\mathbf{t_G}^T)^T \xrightarrow{D} N(0, \mathbf{W}).$$

Hence,

$$\sqrt{n}\mathbf{t} \xrightarrow{D} N(0, \mathbf{BWB}^T).$$

Because $\sqrt{n}\mathbf{m}$ and $\sqrt{n}\mathbf{t}$ have the same asymptotic distribution, we could conclude that

$$\sqrt{n}\mathbf{m} \xrightarrow{D} N(0, \mathbf{BWB}^T).$$

Theorem 4.4 provides the asymptotic result concerning our test statistic $T = n\mathbf{m^T m}$.

Theorem 4.4. *Assume the conditions of Lemma 4.1 hold, then under null hypothesis* (4.1)*, we have*

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

where $\omega_1 \geq \cdots \geq \omega_{(G-1)d(p-d)}$ *are the eigenvalues of* $\mathbf{BWB}^T$*.*

Proof:

Since $\mathbf{B} = \left(\text{diag}(\sqrt{\frac{1}{r_1}}, \cdots, \sqrt{\frac{1}{r_G}}) - (\sqrt{r_1}, \cdots, \sqrt{r_G}) \otimes \mathbf{1}_G\right) \otimes \mathbf{I}_{p^2}$, then $\text{rank}(\mathbf{B}) = (G-1) \times p^2$.

Also, $\mathbf{W} = \text{diag}(\mathbf{\Psi}_1, \cdots, \mathbf{\Psi}_G)$, where $\mathbf{\Psi}_g = \mathbf{U}_g \mathbf{\Phi}_g \mathbf{U}_g^T$. Hence, $\text{rank}(\mathbf{\Psi}_g) = \text{rank}(\mathbf{U}_g) = d(p-d)$, and $\text{rank}(\mathbf{W}) = d(p-d)G$. By the elementary row and column operations, we may rewrite $\mathbf{B}$ as

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

where $\mathbf{Q}$ represents the composition of all elementary operations, $\mathbf{B}_i$ is a $p^2G \times p^2$ column full rank matrix, for $i = 1, \cdots, G-1$, and $\mathcal{O}$ is a $p^2G \times p^2$ matrix of zeros. Hence,

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

Because $(\mathbf{B}_1, \cdots, \mathbf{B}_{G-1})$ is a column full rank matrix, there exists an invertible matrix $\mathbf{D}$ such that $(\mathbf{B}_1, \cdots, \mathbf{B}_{G-1}) = \mathbf{D} \begin{pmatrix} \mathbf{I}_{p^2(G-1)} \\ \mathcal{O} \end{pmatrix}$.

So, we have

$$\begin{aligned} & \operatorname{rank}\left((\mathbf{B}_1, \cdots, \mathbf{B}_{G-1})\operatorname{diag}(\mathbf{\Psi}_1, \cdots, \mathbf{\Psi}_{G-1})(\mathbf{B}_1, \cdots, \mathbf{B}_{G-1})^T\right) \\ = \; & \operatorname{rank}(\mathbf{\Psi}_1, \cdots, \mathbf{\Psi}_{G-1}) = d(p-d)(G-1). \end{aligned}$$

Then, the conclusion just follows naturally. □

4.4. SIMULATION STUDIES

Throughout our simulation studies, the random error ϵ is assumed to be standard normal 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 compare the performance of our proposed tests among the three sufficient dimension reduction methods with different choices of n and p.

4.4.1. Estimated Test Levels. In this subsection, we evaluate the performance of our test statistics under different models when null hypothesis (4.1) holds.

4.4.1.1. Model I. We first consider the following model with one-dimensional structure for all three groups. The predictor vector $\mathbf{X} = (X_1, \cdots, X_p)$ 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 group size increases, the estimated levels approach nominal levels. For example, when $p = 5$ and the nominal level is 1%, the estimated levels for modified SIR are 1.8%, 1.4% and 1.1% respectively for sample sizes 400, 600 and 800. Also it comes as no surprise that the performance

of our tests slightly deteriorates as p increases. Although, SIR-based method generates simulation results greater than the true nominal levels, results at $p = 10$ are more far away from the true value. All three dimension reduction methods perform reasonably well for all combinations of p and n.

Table 4.1. Estimated Test Levels (in percentages) for Model I

	Model I with $p = 5$				Model I with $p = 10$			
		Nominal Level (%)				Nominal Level (%)		
Sample Size	Test	1	5	10	Test	1	5	10
$n_1 = n_2 = n_3 = 400$	SIR	1.80	6.80	13.30	SIR	2.20	7.10	14.10
	SAVE	0.50	4.50	6.90	SAVE	0.30	2.70	5.20
	DR	1.50	7.80	12.10	DR	0.80	3.80	7.20
$n_1 = n_2 = n_3 = 600$	SIR	1.40	6.20	10.40	SIR	2.00	6.60	11.90
	SAVE	0.60	4.10	7.80	SAVE	0.30	3.20	6.70
	DR	0.90	7.30	12.3	DR	0.70	4.20	7.80
$n_1 = n_2 = n_3 = 800$	SIR	1.10	5.40	10.4	SIR	0.90	6.20	10.80
	SAVE	0.70	5.10	8.90	SAVE	0.50	4.20	8.30
	DR	1.20	5.80	10.90	DR	0.80	4.20	7.90

4.4.1.2. Model II. We consider a two-dimensional model with $\mathbf{X}$ generated in the same way as those in Model **I**. All three groups have the same two-dimensional structure.

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

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

	Model II with $p = 5$				Model II with $p = 10$			
		Nominal Level (%)				Nominal Level (%)		
Sample Size	Test	1	5	10	Test	1	5	10
$n_1 = n_2 = n_3 = 400$	SIR	0.20	1.20	3.60	SIR	0.00	1.00	2.30
	SAVE	2.40	7.80	12.3	SAVE	3.20	6.50	14.00
	DR	0.70	4.10	11.60	DR	0.60	3.10	7.20
$n_1 = n_2 = n_3 = 600$	SIR	0.80	3.90	9.10	SIR	0.20	1.30	3.20
	SAVE	1.60	6.70	11.50	SAVE	1.80	6.80	12.60
	DR	0.90	6.30	8.80	DR	0.70	3.70	8.20
$n_1 = n_2 = n_3 = 800$	SIR	0.60	3.20	10.3	SIR	0.40	1.50	4.00
	SAVE	1.40	5.20	10.10	SAVE	1.50	5.80	11.8
	DR	0.90	5.40	9.00	DR	0.80	4.10	8.30

As shown in Table 4.2, our method based on the SAVE and DR candidate matrix performs better than those based on SIR, in this model. For example with sample size 800, $p = 5$ and the nominal level is 5%, the estimated test levels of methods SIR, SAVE and DR are $3.20\%, 5.20\%$ and 5.40%, respectively. Most of the simulation results for SIR-based methods are much less than the true nominal levels. It seems that SIR-based method is more sensitive to the choices of sample sizes and number of variables (p) here. For example, when $p = 10$ and sample size is 400, the estimation level under nominal level 1% is even equal to 0.

4.4.1.3. Model III. In this model, the predictor vector $X_i's$ are independent and t distributed with degrees of freedom 1.

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

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

	Model III with $p = 5$				Model III with $p = 10$			
		Nominal Level (%)				Nominal Level (%)		
Sample Size	Test	1	5	10	Test	1	5	10
$n_1 = n_2 = n_3 = 400$	SIR	3.1	7.9	15.6	SIR	0.30	3.40	6.60
	SAVE	2.20	7.40	15.0	SAVE	3.50	7.40	15.20
	DR	1.60	6.40	11.40	DR	1.80	6.40	11.70
$n_1 = n_2 = n_3 = 600$	SIR	2.80	7.50	13.20	SIR	0.50	3.70	6.90
	SAVE	1.80	6.60	12.90	SAVE	2.40	6.90	13.10
	DR	1.40	6.00	11.10	DR	1.50	5.90	10.90
$n_1 = n_2 = n_3 = 800$	SIR	1.60	6.40	11.70	SIR	0.60	3.50	7.10
	SAVE	1.40	5.50	11.5	SAVE	1.70	6.70	12.50
	DR	1.30	5.20	10.20	DR	1.30	5.70	10.60

From table 4.3, 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.

4.4.1.4. Model IV. We now consider a one-dimensional model 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 $\mathbf{X} = (X_1, \cdots, X_p)$ follows a multivariate normal distribution with mean 0, and the correlation between X_i and X_j as $0.5^{|i-j|}$, $i = 1, \cdots p$, $j = 1, \cdots p$. Different groups share common indices and $d = 1$. It seems that the corre-

lation among the predictors doesn't substantially affect the performance of our methods. The DR-based method performs best among our three test procedures.

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

	Model IV with $p = 5$				Model IV with $p = 10$			
		Nominal Level (%)				Nominal Level (%)		
Sample Size	Test	1	5	10	Test	1	5	10
$n_1 = n_2 = n_3 = 400$	SIR	2.00	7.00	16.20	SIR	2.30	7.20	17.20
	SAVE	1.60	6.20	12.4	SAVE	2.70	6.80	12.5
	DR	1.50	5.70	11.50	DR	0.70	5.90	11.90
$n_1 = n_2 = n_3 = 600$	SIR	1.70	6.70	13.70	SIR	1.90	7.00	14.90
	SAVE	1.60	5.90	12.00	SAVE	2.10	6.20	11.80
	DR	0.80	5.50	10.30	DR	0.80	5.70	11.30
$n_1 = n_2 = n_3 = 800$	SIR	1.30	5.80	12.2	SIR	1.60	6.20	13.5
	SAVE	1.50	5.30	11.50	SAVE	1.70	5.90	11.40
	DR	1.10	5.40	9.60	DR	1.10	5.70	10.40

4.4.1.5. Model V. In this model, the predictor vector $\mathbf{X} = (X_1, \cdots, X_p)$ is also generated from standard multivariate normal. We can see from Table 4.5 that SIR-based testing method fails for the symmetric model. Its estimated nominal levels are much higher than the true values. But methods based on SAVE and DR still work reasonably well, which is in line with our expectations.

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

Table 4.5. Estimated Test Levels (in percentages) for Model V

	Model V with $p = 5$				Model V with $p = 10$			
		Nominal Level (%)				Nominal Level (%)		
Sample Size	Test	1	5	10	Test	1	5	10
$n_1 = n_2 = n_3 = 400$	SIR	4.80	18.70	31.50	SIR	5.10	18.00	30.60
	SAVE	1.80	6.70	13.50	SAVE	1.70	7.30	13.50
	DR	0.60	5.70	7.90	DR	1.70	5.90	8.2
$n_1 = n_2 = n_3 = 600$	SIR	4.90	19.60	30.60	SIR	5.60	20.60	33.50
	SAVE	1.70	5.50	11.8	SAVE	1.60	6.40	11.40
	DR	1.40	5.60	8.60	DR	1.70	5.50	11.2
$n_1 = n_2 = n_3 = 800$	SIR	3.80	20.50	34.60	SIR	4.80	19.90	33.50
	SAVE	1.10	4.80	10.70	SAVE	1.40	4.30	11.20
	DR	1.30	4.80	9.40	DR	0.80	5.60	11.2

4.4.2. 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.

4.4.2.1. Model VI.

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

Here each group has a different direction, that is $\mathcal{S}_{Y^g|\mathbf{X}^g} = 1$ for $g = 1, 2, 3$. We first set $d = 1$. Here, $d = 1$ is the dimension of each group. As shown in Table 4.6, for all the sample sizes, p and choices of dimension reduction methods, SIR-based methods

performed extremely well with 100% of power. The power of SAVE and DR is around 90% and 70% which is also acceptable.

Table 4.6. Estimated Power at 5% Nominal Levels for Model VI at $d = 1$

Sample Size	SIR (power×100)		SAVE (power×100)		DR (power×100)	
	$p = 5$	$p = 10$	$p = 5$	$p = 10$	$p = 5$	$p = 10$
$n_1 = n_2 = n_3 = 400$	100	100	77.90	98.60	70.60	74.6
$n_1 = n_2 = n_3 = 600$	100	100	79.90	96.50	73.2	77.5
$n_1 = n_2 = n_3 = 800$	100	100	81.30	92.40	77.5	79.4

If we use a different structural dimension for Model VI, say $d = 3$, which is the dimension of partial central subspace. The power of our SIR-based method would greatly decrease to values less than 50%, but the power of SAVE and DR remains around 70%. It seems that the power of the SIR test method is quite sensitive to the choice of d. Based on our simulation studies including some results not presented here, an estimate of d using single population dimension reduction is recommended in practice.

4.4.3. Comparison of the Two Testing Methods for $G = 2$. In this section, we compare simulation results of methods introduced in Section 2 and this section for $G = 2$.

Recall that Model I from Section 2.3.1.1, where the predictor vector $\mathbf{X} = (X_1, \cdots, X_p)$, 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 4.7 and Table 4.8 compare the estimated test levels for the new method in this section and the two-sample method proposed in Section 2 with $p = 4$ and $p = 8$ respectively. The common sample size is taken to be 200, 400 and 600. We find that the performance

of method in Section 2 is more stable, its estimated levels are closer to the true nominal levels. Our new method also performs reasonably well here, though it seems that the testing method proposed in Section 2 works better for $G = 2$.

Table 4.7. Estimated Test Levels for Model I in Section 2 with $p = 4$

	New Method				Method in Section 2			
		Nominal Level (%)				Nominal Level (%)		
Sample Size	Test	1	5	10	Test	1	5	10
$n_1 = n_2 = 200$	SIR	1.70	7.60	16.4	SIR	1.50	5.50	9.30
	SAVE	0.50	3.50	7.50	SAVE	0.90	4.60	10.6
	DR	1.40	6.30	12.0	DR	1.40	5.30	10.8
$n_1 = n_2 = 400$	SIR	1.50	7.50	12.10	SIR	0.80	4.60	9.40
	SAVE	0.70	3.80	7.70	SAVE	0.80	4.70	10.4
	DR	1.40	5.90	10.80	DR	0.80	5.20	9.70
$n_1 = n_2 = 600$	SIR	1.20	6.40	10.70	SIR	1.10	4.90	10.3
	SAVE	0.90	4.50	8.60	SAVE	0.90	5.20	9.90
	DR	1.20	4.60	9.50	DR	1.10	5.00	9.80

Table 4.8. Estimated Test Levels for Model I in Section 2 with $p = 8$

	New Method				Method in Section 2			
		Nominal Level (%)				Nominal Level (%)		
Sample Size	Test	1	5	10	Test	1	5	10
$n_1 = n_2 = 200$	SIR	3.20	8.60	18.60	SIR	1.20	4.50	9.10
	SAVE	0.40	3.00	7.30	SAVE	1.40	5.30	9.60
	DR	1.70	7.20	12.60	DR	1.60	4.60	9.50
$n_1 = n_2 = 400$	SIR	1.80	7.50	13.5	SIR	1.30	4.50	10.5
	SAVE	0.70	3.50	7.60	SAVE	1.40	5.50	9.50
	DR	0.60	3.60	12.70	DR	0.70	4.70	10.3
$n_1 = n_2 = 600$	SIR	1.30	6.80	11.2	SIR	0.90	5.20	9.80
	SAVE	0.80	3.90	8.90	SAVE	1.10	4.90	10.1
	DR	0.80	6.10	8.90	DR	1.20	5.10	10.1

4.5. APPLICATION TO THE BETA-CATOTENE DATA

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. 1981). It has been of 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. (1989). Zhu et al. (2010b), Yoo (2008b, 2010), and Hilafu and Yin (2013) 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 datafiles 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 extremely high value of alcohol use is treated as outlier by Hilafu and Yin (2003), Zhu et al. (2010b) 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 9 variables are also continuous. Detailed descriptions of our variables are given in Table 4.9 as below.

Table 4.9. 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)

We take one of the categorical variables, smoking status (SMOKSTAT) as the group identifier and divide the study objects into 3 groups: nonsmoker, former smoker and current smoker. The remaining 9 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 $\mathcal{S}^{(W)}_{Y|\mathbf{X}}$, with W = SMOKSTAT. We apply the SIR-based test method since according to our simulation studies, it yields the most power among all three test procedures. The observed test statistic is $T = 259.15$, which is greater than 224.20, the $95th$ percentile 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 those of Hilafu and Yin (2013) and 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 other than lung cancer. Our conclusion combined with the observational studies conducted by Peto et al. (2008) could also help us better understand the relationships between smoking and the risk of these cancers.

4.6. SUMMARY

In this section, we developed a new test statistic, and its asymptotic distribution, for testing the common indices of more than two 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 data set, 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 chacratersitics and dietary factors influence the concentration of beta-carotene in human serums.

5. FUTURE WORK

In this dissertation, we proposed novel testing procedures for testing if several multi-index models share the same set of indices. When there are two populations (two multi-index models), both of the methods discussed in Section 2 and Section 4 are applicable. Based on our experiences with extensive simulation studies, we recommend the method proposed in Section 2 over that of Section 4 due to its better performance.

We also notice that the performance of our test procedures is somehow sensitive to the choice of d when there are more than two populations. It seems that an estimate of d based on the dimension reduction methods for a single population works the best. More research on how to choose d seems necessary.

When the multi-index models share only partial common indices as illustrated by the following example, in which all three groups share the same common index $X_1 + X_p$, and differ with respect to the second index.

$$Y = \begin{cases} \exp(X_1 + X_p) sign(X_2 + X_{p-1}) + 0.5\epsilon_1, & \text{for group 1;} \\ (X_1 + X_p)/(1 + X_3 - X_{p-1}) + 0.5\epsilon_2, & \text{for group 2;} \\ (X_1 + X_p)(X_2 - X_{p-1}) + 0.5\epsilon_3, & \text{for group 3.} \end{cases}$$

How to effectively detect this phenomenon is of special interest.

In our simulation studies, we also find that the choice of the number of slices h may affect the test results. Based on the analysis of Li (1991), the selection of h is not very crucial for SIR. SAVE, however, is pretty sensitive to the choice of h (Li and Zhu, 2007). How to select h optimally is an open question in the field.

BIBLIOGRAPHY

[1] Abdi, H. (2003). PLS-Regression; Multivariate Analysis. In M.Lewis-Beck, A. Bryman, T.Futing (Eds): *Encyclopedia for Research Methods for the Social Sciences.* Thousand Oaks: Sage.

[2] Bellman, R. E. (1961). *Adaptive Control Processes: a Guided Tour.* Princeton University Press.

[3] Bentler, P. and Xie, J. (2000). Corrections to Test Statistics in Principal Hessian Directions. *Statitics and Probability Letter*, **47**, 381–389.

[4] Biok R. J. (2002). Spectral Models for Covariance Matrices. *Biomatrika*, **89**, 159–182.

[5] Bura, E. and Cook, R. D. (2001). Extending Sliced Inverse Regression: The Weighted Chi-Squared Test. *Journal of the American Statistical Association*, **96**, 996–1003.

[6] Chiaromonte, F., Cook, R. D. and Li, Bing. (2002). Sufficient Dimension Reduction in Regressions with Categorical Predictors. *The Annals of Statistics*, **30**, 475–497.

[7] Cook, R. D.(1994). Using Dimension-reduction Subspaces to Identify Important Inputs in Models of Physical Systems. *American Statistical Association*, Washington.

[8] Cook, R. D.(1996). Graphics for Regression with a binary response. *Journal of the American Statistical Association*, **91**, 983–992.

[9] Cook, R. D. (1998). *Regression Graphics* Wiley, New York.

[10] Cook, R. D. and Forzani, B. (2009). Likelihood-based Sufficient Dimension Reduction. *Journal of the American Statistical Association*, **104**, 197–208.

[11] Cook, R. D. and Weisberg, S. (1991). Discussion of "Sliced Inverse Regression for Dimension Reduction" by Li. *Journal of the American Statistical Association*, **86**, 328–332.

[12] Cook, R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics.* Wiley, New York.

[13] Eaton, M. L. (1986). A Characterization of Spherical Distributions. *Journal of Multivariate Analysis*, **20**, 272–276.

[14] Eaton, M. L. and Tyler, D. E. (1994). The Asymptotic Distribution of Singular Values with Apllication to Canonical Correlations and Correpondence Analysis. *Journal of Multivariate Analysis*, **50**, 238–264.

[15] Eubank, R.L. (1988). *Spline Smoothing and Nonparametric Regression.* New York: Marcel Dekker.

[16] Feng, Z. and Zhu, L. X. (2012). An Alternating Determination Optimization Approach for an Additive Multi-index Model. *Computational Statistics and Data Analysis*, **56**, 1981–1993

[17] Fernholz, L. T. (1983). *Von Mises Calculus for Statistical Functionals*. New York: Springer.

[18] Flury, B. (1984). Common Principal Components in k Groups. *Journal of the American Statistical Association*, **79**, 892–898.

[19] Flury, B. (1987). Two Generalizations of the Common Principal Component Model. *Biometrika*, **74**, 59–69.

[20] Flury, B. and Riedwyl, H. (1988). *Multivariate Statistics, A Practical Approach*, New York: John Wiley and Sons.

[21] Hall, P. and Li, K. C. (1993). On Almost Linearity of Low Dimensional Projections from High Dimensional Data. *The Annals of Statistics*, **21**, 867–889.

[22] Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal Smoothing in Single-index Models. *The Annals of Statistics*, **21**, 157–178.

[23] Härdle, W. and Stoker, T. M. (1989). Investigating Smooth Multiple Regression by the Method of Average Derivatives. *Journal of the American Statistical Association*, **84**, 986–995.

[24] Hastie, T., Tibshirani, R. and Friedman, J. (2011). *The Elements of Statistical Learning*, 2nd ed. New York: Springer.

[25] Hilafu, H., Yin, X. (2013). Sufficient Dimension Reduction in Multivariate Regressions with Categorical Predictors *Computational Statistics Data Analysis*, **63**, 139–147.

[26] Horn, R. A. and Johnson, C. R. (1991). *Topics in Matrix Analysis*. Cambridge University Press

[27] Izenman, A.J. (2013). *Modern Multivariate Statistical Techniques*, Springer, New York.

[28] Jolliffe, I.T. (2002). *Principal Component Analysis*, Springer, New York.

[29] Kato, T. (1966). *Pertubation Theory for Linear Operators*, Springer, Berlin.

[30] Krzanowski, W. J. (1979). Between-Groups Comparison of Principal Components. *Journal of the American Statistical Association*, **74**, 703–707.

[31] Lee, K.Y., Li, B. and Chiaromonte F. (2013). A General Theory for Nonlinear Sufficient Dimension Reduction: Formulation and Estimation. *The Annals of Statistics*, **6**, 3182–3210.

[32] Li, B., Artemiou, A. and Li, L. (2011). Principal Support Vector Machines for Linear and Nonlinear Sufficient Dimension Reduction. *The Annals of Statistics*, **37**, 1272–1298.

[33] Li, B. and Dong, Y. (2009). Dimension Reduction for Nonelliptically Distributed Predictors. *The Annals of Statistics*, **37**, 1272–1298.

[34] Li, B., Kim, M. K. and Altman, N. (2010). On Dimension Folding of Matrix- or Array-Valued Statistical Objects. *The Annals of Statistics*, **38**, 1094–1121.

[35] Li, B. and Wang, S. (2007). On Directional Regression for Dimension Reduction. *Journal of the American Statistical Association*, **102**, 997–1008.

[36] Li, K. C. (1991). Sliced Inverse Regression for Dimension Reduction (with discussion). *Journal of the American Statistical Association*, **86**, 316–342.

[37] Li, Y. and Zhu, L. (2007). Asymptotics for Sliced Average Variance Estimation. *The Annals of Statistics*, **35**, 41–69.

[38] Luo, W., Li, B. and Yin, X. (2014). On Efficient Dimension Reduction with Respect to a Statistical Functional of Interest. *The Annals of Statistics*, **42**, 382–412.

[39] Ma, Y. and Zhu, L. (2012). A Semiparametric Approach to Dimension Reduction. *Journal of the American Statistical Association*, **107**, 168–179.

[40] Ma, Y. and Zhu, L. (2013). Efficient Estimation in Sufficient Dimension Reduction. *The Annals of Statistics*, **41**, 250–268.

[41] Maitra, S. and Yan, J. (2008) Principal Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression *Casualty Actuarial Society, 2008 Discussion Paper Program*, 79–90.

[42] Nevill, A.M. and Holder, R.L. (1995) Body Mass Index; A Measure of Fatness or Leanness? *The British Journal of Nutrition*, **73**, 507–516.

[43] Nierenberg, D.W., Stukel, T.A., Baron, J.A. (1989) Determinants of Plasma Levels of Beta-carotene and Retinol. *American Journal of Epidemiology*, **130**, 511–521.

[44] Peto, R., Doll, R., Buckley, JD., Sporn, MB.(2008) Can Dietary Beta-carotene Materially Reduce Human Cancer Rates? *Nature*, **290**, 201–208.

[45] Schott, J. R. (1988). Common Principal Component Subspaces in Two Groups. *Biometrika*, **75**, 229–236.

[46] Schott, J. R. (1991). Some Tests for Common Principal Component Subspaces in Several Groups. *Biometrika*, **78**, 771–777.

[47] Schott, J. R. (1997). Asymptotics of Eigenprojections of Correlation Matrices with Some Applications in Principal Components Analysis. *Biometrika*, **84**, 327–337.

[48] Tyler, D. E. (1981). Asymptotic Inferences for Eigenvectors. *The Annals of Statistics*, **9**, 725–736.

[49] Weisberg, S. (2005). *Applied Linear Regression*, 3rd ed. New York: Wiley.

[50] Wen, X and Cook, R. D. (2007). Optimal Sufficient Dimension Reduction for Regressions with Categorical Predictors. *Journal of Statistical Planning and Inference*, **137**, 1961-1978.

[51] Wen, X and Cook, R. D. (2009). New Approaches to Model-free Dimension Reduction for Bivariate Regression. *Journal of Statistical Planning and Inference*, **139**, 734-748.

[52] Xia, Y. (2008) A Multiple-index Model and Dimension Reduction. *Journal of the American Statistical Association*, **103**, 1631–1640.

[53] Xia, Y., Tong, H., Li, W. K. and Zhu, L. X. (2002). An Adaptive Estimation of Dimension Reduction Space. *Journal of the Royal Statistical Society, Ser. B*, **64**, 363–410.

[54] Xue, L. G. and Zhu, L. X. (2006). Empirical Likelihood for Single-index Model. *Journal of Multivariate Analysis*, **97**, 1295–1312.

[55] Yoo, J.K. (2008). A Novel Moment-based Sufficient Dimension Reduction Approach in Multivariate Regression. *Computational Statistics Data Analysis*, **52**, 3843–3851.

[56] Yoo, J.K. (2010). Integrated Partial Sufficient Dimension Reduction with Heavily Unbalanced Categorical Predictors *The Korean Journal of Applied Statistics*, **23**, 977–985.

[57] Yu, Z., Zhu, L. and Wen, X. (2012). On Model-free Conditional Coordinate Tests for Regressions. *Journal of Multivariate Analysis*, **109**, 61–72.

[58] Zhu, L.P., Zhu, L.X.(2007). On Kernel Method for Sliced Average Variance Estimation. *Journal of Multivariate Analysis*, **98**, 970–991.

[59] Zhu, L. P., Zhu, L. X., Wen, S.Q. (2010). On Dimension Reduction in Regressions with Multivariate Responses. *Statistica Sinica*, **20**, 1291–1307.

VITA

Xuejing Liu attended University of Science and Technology of China at 2004 and completed her bachelor's degree in Statistics in 2008. Xuejing entered Missouri University of Science and Technology for graduate studies in 2009. She received her MS degree in Applied Mathematics with Statistics emphasis in May 2011 and received her PhD degree in Statistics in August 2015 at Missouri University of Science and Technology.