Reduced-Rank Mean Estimation for Projective-Resampling Informative Predictor Subspace

Jeesun Jang

The Graduate School

Yonsei University

Department of Applied Statistics

Reduced-Rank Mean Estimation for Projective-Resampling Informative Predictor Subspace

A Dissertation

Submitted to the Department of Applied Statistics

and the Graduate School of Yonsei University

in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy in Applied Statistics

Jeesun Jang

December 2023

This certifies that the dissertation of Jeesun Jang is approved.

Thesis Supervisor: Prof. Hakbae Lee

Committee Member: Prof. Sangwook Kang

Committee Member: Prof. Kyungdeock Park

Committee Member: Prof. Jae Keun Yoo

Committee Member: Dr. Kion Kim

The Graduate School

Yonsei University

December 2023

Acknowledgements

My deepest gratitude goes to my advisor, Prof. Hakbae Lee mentoring and guiding me throughout my entire time over the course of my graduate studies. He has showed me with an open mind into the world of a new academic discipline, and his passion for quantum, in particular, has been a great source of motivation and fascination for me. I owe particular gratitude to Prof. Jae Keun Yoo who is one of my dissertation committee chair. He has been a great mentor to me with meticulous and profound care, as well as his significant influence on the direction of the research ideas.

I would also like to thank my dissertation committee chair, Prof. Sangwook Kang, Prof. Kyung-deock Park, and Dr. Kion Kim for sharing brilliant comments and suggestions. I thank each of them for serving on my committee and for the time they spent reviewing this work.

I can't overlook mentioning Dr. Lee Eon-kyung. I got to know Dr. Lee at Korea Maritime Institute after briefly working there following my master's graduation. Dr. Lee guided me in writing various reports and papers within the company. Addressing Dr. Lee as 'Dr. Jang' in advance, Dr. Lee kindly waited for me until I obtained my doctoral degree. Currently serving diligently as a research committee member and head of the Logistics and Maritime Industry Research Institute, I aspire to share this honor with Dr. Lee.

My thanks and appreciation also goes to my graduate colleague, Dr. Myunghyun Park, who lives nearby, occasionally meets with me, serves as a conversation partner. Fortunately she inquired about the progress of my research and inspired me.

I would like to extend my sincere gratitude towards my parents and my parents-in-law for unwavering love and sacrifices which has allowed me to pursue my study. Without their backing I could have never completed this doctoral degree. Thank you also to my sister and sister-in law for their consistent understanding of my situation.

Most importantly, my heartfelt gratitude goes to my husband, Jinseouk, whose consistent weekend care of our two boys has made it possible for me to go forward during my difficult times. This thesis is especially dedicated to my sons Dohyun, who is kind-hearted yet adept at finding his own way, and Doyoon, who is fiercely independent with clear thoughts.

I am also thankful to the department staff and many others in the department who have supported me throughout my academic journey.

Last but not least, I am grateful to God, who, no matter the challenges that have come my way, fills my life with hope and light, inspiring me to live with an awakened heart full of love for the present.

Contents

List of Figures

List of Tables

ABSTRACT

The background of this research lies in the prerequisite conditions for estimating the central subspace or central mean subspace through sufficient dimension reduction (SDR). Traditionally, adherence to conditions such as linearity, constant variance, and coverage was imperative. However, in an attempt to transcend these restrictive conditions, this study aims to define an informative predictor subspace (IPS) as an alternative approach to estimating the central subspace. The approach involves clustering X and then identifying the space spanned by eigenvectors through categorizing or slicing Y. As all attention was historically directed towards dimension reduction of X, this study is focused on multivariate Y, emphasized the necessity of considering response dimension reduction in regression. Yoo and Cook introduced the method, referred to as YC, for response reduction. Additionally, they presented principal response reduction (PRR), principal fitted response reduction (PFRR), and extended beyond with unstructured PFRR (UPFRR) by employing semi-parametric constraints. To estimate the IPS space alongside these methods for multivariate Y, given that slicing Y after clustering X might not alleviate the curse of dimensionality, the approach incorporates projective resampling as proposed by Li et al.(2008) to obtain solutions. Upon employing the projective resampling method, it was observed that a space, referred to as Projective Resampling IPS (PRIPS), was identified. This space, while smaller than IPS, exhibited higher accuracy in encompassing the central subspace. The central subspace governed by IPS that we aim to estimate. Three methods were proposed to find IPS: the projective resampling mean method, coordinate mean method, and coordinate projective resampling method.

In this study, when dealing with multivariate Y, the aim was to conduct response dimension reduction while seeking to estimate the central subspace using PRIPS. However, it's important to note that the space derived from response dimension reduction isn't uniquely defined like the central subspace. Hence, a preference was established to adopt values extracted from the space composed of the mean of ψ values. Specifically, an approach named Reduced-rank response partial conditional mean (RRR-pcm) was proposed, wherein the mean of the previously suggested four methods in response dimension reduction research was utilized. Leveraging methods used in IPS estimation, inspiration was drawn from the third method, coordinate projective resampling. This led to the proposal of RRRcomb, integrating the coordinate mean method into RRR-pcm. The research findings indicate that, compared to performing response dimension reduction with a large number of random samplings, applying dimension reduction followed by RRR-pcm or RRRcomb with a smaller number of samplings leads to higher or sustained accuracy in estimating values.

Key words : Sufficient dimension reduction (SDR), Response dimension reduction, Informative predictor subspace (IPS), Projective resampling informative predictor subspace (PRIPS), Reduced-rank response regression

Chapter 1

Introduction

Sufficient dimension reduction (SDR) for a regression of $Y|\boldsymbol{X} \in \mathbb{R}^p$ replaces the original p-dimensional predictor $\boldsymbol{X}$ with a lower-dimensional linear projection $\boldsymbol{M^T X}$ without loss of any information on $Y|\boldsymbol{X}$, where $\boldsymbol{M} \in \mathbb{R}^{p\times q}$ and $q \leq p$. It can be equivalently stated that SDR is to find $\boldsymbol{M}$ or *S(*$\boldsymbol{M}$*)* satisfying the following conditional independence

$$Y \perp\!\!\!\perp \boldsymbol{X}|\boldsymbol{M^T X} \tag{1.1}$$

where $\perp\!\!\!\perp$ indicates statistical independence and *S(*$\boldsymbol{M}$*)* means a subspace spanned by the columns of $\boldsymbol{M}$. Then *S(*$\boldsymbol{M}$*)* and $\boldsymbol{M^T X}$ constructed from $\boldsymbol{M}$ to satisfy (1.1) are called a dimension reduction subspace and sufficient predictors, respectively.

The intersection of all possible dimension reduction subspace is called the central subspace $S_{Y|\boldsymbol{X}}$. The construction of $S_{Y|\boldsymbol{X}}$ guarantees that it is minimal and unique. The main goal of SDR is to restore $S_{Y|\boldsymbol{X}}$ from data. Hereafter, a true orthonormal basis matrix of $S_{Y|\boldsymbol{X}}$ and its structural dimension will be represented as $\boldsymbol{\eta}$ and d.

Popular SDR methodologies among many others should be sliced inverse regression (Li (1991)), sliced average variance estimation (Cook and Weisberg (1991)) , principal Hessian direction (Li (1992)) and ordinary least squares (Cook (2009)). The `dr`-package (Weisberg (2002)) in the statistical program R can implement these four methods.

Unfortunately, the methods commonly require one of both linearity and constant variance conditions. These conditions are restricted to the distribution of $\boldsymbol{X}$, not Y. Any of the four methods do not assure an exhaustive estimation of $S_{Y|\boldsymbol{X}}$. Therefore, a coverage condition to guarantee the exhaustiveness should be additionally assumed. According to Li and Dong (2009) these conditions are not easily diagnosed through data in practice, and , even worse, their violation can lead wrong dimension reduction results.

To overcome these deficits, Yoo (2016) develops a new dimension reduction subspace called an informative predictor subspace (IPS). The informative predictor subspace contains $S_{Y|\boldsymbol{X}}$ by its construction, and the estimation methodology proposed in Yoo (2016) , which is called a clustering mean method, does not require any of the conditions. A categorization of Y after clustering $\boldsymbol{X}$ is necessary in the implementation of the clustering mean method. Here, the categorization of Y is called slicing. We will discuss them in later section.

The informative predictor subspace and its estimation is restricted to a regression with one-dimensional response variable. The main purpose of the paper extends the informative predictor subspace to a multivariate regression of $\boldsymbol{Y} \in \mathbb{R}^r | \boldsymbol{X} \in \mathbb{R}^p$ where, $r \geq 2$. Since the construction of the informative predictor subspace requires a dimension reduction subspace, the dimensionality of $\boldsymbol{Y}$ is not an issue to define the informative predictor subspace for multivariate regression. However, the multi-dimensionality of $\boldsymbol{Y}$ should be a cause of concern in the estimation, because the slicing scheme faces the curse of dimensionality. In some slices of $\boldsymbol{Y}$, there will be few observations, and hence the estimation should be not reliable.

To relieve this issue, we will define a projective resampling informative predictor subspace (PRIPS), borrowing the idea of a projective resampling method by Li *et al.* (2008). The projective resampling method transforms the r-dimensional $\boldsymbol{Y}$ to one dimensional linearly transformed $t_j^T\boldsymbol{Y}$, where $t_j \in \mathbb{R}^{r \times 1}, j = 1, ..., m$, and t_j is generated from a distribution on the unit sphere $\mathbb{S}^r = \{t \in \mathbb{R}^r : \|t\| = 1\}$. More detail of the projective resampling method will be given in later section.

Then, the projective resampling IPS (PRIPS) turns out to be contained in the original IPS, but it still contains $S_{\boldsymbol{Y}|\boldsymbol{X}}$. That is, the newly defined projective resampling IPS will provide the upper bound of $S_{\boldsymbol{Y}|\boldsymbol{X}}$ for multivariate regression. To address the limitations of conducting random sampling among the three methods used for estimating PRIPS, this study was undertaken. It aims to compensate for the potential lack of information regarding the relationship between X and Y when relying solely on Y's coordinate base information.

The organization of the paper is as follows. In Chapter 2, the informative predictor subspace and the related estimation methods and the projective resampling method are briefly introduced. In Chapter 3, we forms a theoretical foundation of the projective resampling method to estimate the informative predictor subspace for multivariate regression, and an estimation approach is proposed. Numerical studies and two real data examples are presented in Section 3.3 and Section 3.4. We summarize our work in Chapter 4.

Chapter 2

Literature Review

2.1 Multivariate Central Mean Subspace

In the multivariate regression of $\boldsymbol{Y}$ on $\boldsymbol{X}$, a subspace $\mathcal{S}(\boldsymbol{M}) \in \mathbb{R}^p$ is called a *mean subspace* for $\boldsymbol{Y}|\boldsymbol{X}$ if

$$\boldsymbol{Y} \perp\!\!\!\perp E(\boldsymbol{Y}|\boldsymbol{X})|\boldsymbol{M}^T\boldsymbol{X} \tag{2.1}$$

where $\perp\!\!\!\perp$ stands for independence and the columns of the $p \times q$ matrix $\boldsymbol{M}$ forms a basis for $\mathcal{M}$. The statement says that $\boldsymbol{Y}$ is conditionally independent of $E(\boldsymbol{Y}|\boldsymbol{X})$ given any value for $\boldsymbol{M}^T\boldsymbol{X}$. It indicates that, for the purpose of characterizing $E(\boldsymbol{Y}|\boldsymbol{X})$ in the population, we can replace $\boldsymbol{X}$ with $\boldsymbol{M}^T\boldsymbol{X}$ without loss of information. There always exists a subspace $\mathcal{M}$ that satisfies (2.1) because the statement is trivially true when $\boldsymbol{M} = \boldsymbol{I}$. The following statement is equivalent to (2.1)

$$\text{cov}\left[(\boldsymbol{Y}, E(\boldsymbol{Y}^T|\boldsymbol{X}))|\boldsymbol{M}^T\boldsymbol{X}\right] = 0 \tag{2.2}$$

At the same time, it also intuitively says $E(Y|X)$ is a function of $\boldsymbol{M}^T\boldsymbol{X}$. (2.2) is a multivariate version of proposition 1 of Cook and Li (2002). These condition say that $\boldsymbol{Y}$ and $E(\boldsymbol{Y}|\boldsymbol{X})$ are uncorrelated given $\boldsymbol{M}^T\boldsymbol{X}$. Either condition (2.1) of (2.2) could be used in the definition of a mean subspace.

When the intersection of all mean subspaces is itself a mean subspace, it is called the central mean subspace (CMS) and denoted by $\mathcal{S}$, with dimension d=dim($\mathcal{S}$). The CMS does not always exist because the intersection of all mean subspaces, while always a subspace, is not necessarily a mean subspace. For example, there are mean subspaces spanned by $span(1,0)^T$ and $span(0,1)^T$. But their intersection is not a mean subspace, and consequently the CMS does not exist. In regressions where the CMS does not exist, it is not clear in general how to identify a unique parsimonious population construction to characterize the mean function, and this can cause substantial problems in theory and practice.

Nevertheless, the CMS does exist under reasonable regularity conditions, as shown in next two properties. Let $\boldsymbol{X}$ has marginal density $f(\mathbf{x}) > 0$ for $\mathbf{x} \in \Omega \subset \mathbb{R}^p$ and $f(\mathbf{x}) = 0$ otherwise.

First condition of existence of CMS is as follows: Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be mean subspaces for $\boldsymbol{Y}|\boldsymbol{X}$. If Ω is a convex set, then $\mathcal{S}_1 \cap \mathcal{S}_2$ is a mean subspace. It denotes that the CMS exists in any regression where the predictors have a density with convex support. Next, we call a regression that satisfy the relation $\boldsymbol{Y} \perp\!\!\!\perp \boldsymbol{X}|E(\boldsymbol{Y}|\boldsymbol{X})$ as *location regression* (Cook, 1998), because the mean function furnishes all of the information $\boldsymbol{X}$ about $\boldsymbol{Y}$. Although higher order conditional moment of $\boldsymbol{Y}|\boldsymbol{X}$ need not be constant within the class, they must all be a function of the mean, such as $var(\boldsymbol{Y}|\boldsymbol{X}) = var(\boldsymbol{Y}|E(\boldsymbol{Y}|\boldsymbol{X}))$.

Second condition of the existence of the CMS in the context of location regressions is as follows : Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be mean subspaces for a location regression. If $E(Y_k|\boldsymbol{X} = \mathbf{x}, k = 1, ..., r$ can be expressed as a convergent power series in the coordinates of $\mathbf{x} \in \Omega$ then $\mathcal{S}_1 \cap \mathcal{S}_2$ is a mean subspace. There is no requirement that the response be multivariate in the development of the CMS. In particular, there is also a CMS $\mathcal{S}_k$ for the univariate regression of the kth coordinate Y_k of $\boldsymbol{Y}$ on $\boldsymbol{X}, k = 1, ..., r$. Cook and Setodji (2003) shows the sum of the subspaces ($\mathcal{S}_k$) means the collection of all vectors of the univariate regression. That is, CMS $\mathcal{S}$ for $\boldsymbol{Y}|\boldsymbol{X}$ is summed up with CMSs $\mathcal{S}_k$ for the univariate regressions $Y_k|\boldsymbol{X}$.

2.2 Response dimension reduction

High-dimensional data is indeed abundant and widely distributed. Accordingly, multivariate regression $\boldsymbol{Y} \in \mathbb{R}^r|\boldsymbol{X} \in \mathbb{R}^p$, is quite common in many fields when analyzing repeated measures, longitudinal data and time series data, where $r \geq 2$ and $p \geq 2$. In regression, sufficient dimension reduction (SDR) replaces the p-original predictors with lower-dimensional linearly transformed predictors without loss of information with respect to selected aspects of $\boldsymbol{Y}|\boldsymbol{X}$. Unfortunately, most SDR methods in multivariate regression have focused on reducing the dimension of $\boldsymbol{X}$, not $\boldsymbol{Y}$. The only method to provide response dimension reduction was proposed by Li *et al.* (2003), but it does not perform well with correlated responses.

The multivariate response variables often causes difficulty in analysis. Therefore, a proper dimension reduction of the response variables are facilitate the statistical analysis so as that of the predictors can. If the reduction of multi-dimensional response variables is needed to be done, it is to follow the notion of the SDR, which is to replace the response with a lower-dimensional linearly transformed one, without loss of information. A paradigm for response dimension reduction was developed by Yoo and Cook (2008).

2.2.1 Model-free approach

Yoo and Cook (2008) defined two types of response dimension reduction for $\boldsymbol{Y}|\boldsymbol{X}$. The first one is referred to as linear response reduction, and its content is as follows. In multivariate regression of $\boldsymbol{Y} \in \mathbb{R}^r|X \in \mathbb{R}^p$, consider $\mathbf{L} \in \mathbb{R}^{q\times r}$ $(q \leq r)$ to have the smallest possible rank among matrices satisfying:

$$E(\boldsymbol{Y}|\boldsymbol{X}) = E(P^T_{\mathbf{L}(\Sigma_\mathrm{y})}\boldsymbol{Y}|\boldsymbol{X}) \tag{2.3}$$

where $P_{\mathbf{L}(\Sigma_{\mathrm{y}})} = \mathbf{L}(\mathbf{L}^T\Sigma_{\mathrm{y}}\mathbf{L})^{-1}\mathbf{L}^T\Sigma_{\mathrm{y}}$ is the orthogonal projection operator relative to the inner product $\langle v_1, v_2\rangle_{\Sigma_{\mathrm{y}}} = v_1^T\Sigma_{\mathrm{y}}v_2$. Equation (2.1) indicated that the predictors $\boldsymbol{X}$ influences the components of the conditional mean $E(\boldsymbol{Y}|\boldsymbol{X})$ only through $P_{\mathbf{L}(\Sigma_{\mathrm{y}})}$. This directly implies that lower-dimensional linear projections onto $S(\mathbf{L})$ can replace the original r-dimensional response $\boldsymbol{Y}$ without loss of information on $E(\boldsymbol{Y}|\boldsymbol{X})$.

The second one is called conditional response reduction, and its contents are as follows. Suppose that there exists a matrix $\mathbf{K} \in \mathbb{R}^{r\times k}$ with $k \leq r$ satisfying:

$$E(\boldsymbol{Y}|\boldsymbol{X}) = E\left\{E\left(\boldsymbol{Y}|\boldsymbol{X}, \mathbf{K}^T\boldsymbol{Y})|\boldsymbol{X}\right\} = E\left\{E\left(\boldsymbol{Y}|\mathbf{K}^T\boldsymbol{Y}\right)|\boldsymbol{X}\right\} \tag{2.4}$$

Then $E\left(\boldsymbol{Y}|\mathbf{K}^T\boldsymbol{Y}\right)$ is the function of $\mathbf{K}^T\boldsymbol{Y}$, so that (2.2) is equivalently expressed as $E\left\{f(\mathbf{K}^T\boldsymbol{Y})|\boldsymbol{X}\right\}$ for some function $f(\cdot)$. Hence we have that $E(\boldsymbol{Y}|\boldsymbol{X}) = E\left\{f(\mathbf{K}^T\boldsymbol{Y})|\boldsymbol{X}\right\}$, and another dimension reduction of Y is done if $k \leq r$. According to Yoo and Cook (2008), there is a relation $\mathbf{L}$ in (2,1) and $\mathbf{K}$ in (2,2) such that $S(\mathbf{K}) \subseteq S(\mathbf{L})$ .The equality holds under the linearity condition as follows:

$$E(\boldsymbol{Y}|\mathbf{K}^T\boldsymbol{Y} = a)$$

is linear in a. If linearity condition fails, $\boldsymbol{Y}$ can be one-to-one transformed for the normality. Under linearity condition, Yoo and Cook (2008) proposed $\Sigma_{\mathrm{y}}^{-1}cov(\boldsymbol{Y}, \boldsymbol{X})\Sigma_{\mathrm{x}}^{-1}$ to estimate $\mathbf{L}$ and $\mathbf{K}$.

2.2.2 Semi-parametric approach

Cook (2007) showed that a semi-parametric approach in SDR can outperform model free approach. Following the idea, Yoo (2018) proposed two versions of semi-parametric response dimension reduction approaches, called principal response reduction (PRR) and principal fitted response reduction (PFRR) in the context of Yoo and Cook (2008). Yoo (2018) conforms that the two semi-parametric approaches have potential advantages in the response dimension reduction over Yoo and Cook (2008). Moreover Yoo (2019) developed unstructured PFRR (UPFRR), which do not assume the structure of the covariance matrix of the random error vectors in Yoo (2018) in the estimation.Yoo (2019) provides a guidelines to choose either PRR or PFRR. In light of these three semi-parametric approaches, let us delve into each one individually for a more detailed examination.

A semi-parametric response reduction approach starts with the following multivariate regression with assuming $E(\boldsymbol{Y}) = 0$ and $E(\boldsymbol{X}) = 0$ without loss of generality:

$$\boldsymbol{Y} = \boldsymbol{\Gamma}\nu_x + \boldsymbol{\varepsilon} \tag{2.5}$$

where $\boldsymbol{\Gamma} \in \mathbb{R}^{r\times d}$ with $\boldsymbol{\Gamma}^T\boldsymbol{\Gamma} = \boldsymbol{I_d}$ and $d \leq r, \boldsymbol{\varepsilon} \sim N(0, \Sigma)$ and $\text{cov}(\nu_x, \boldsymbol{\varepsilon}) = 0$.

In model (2.5), ν_x is a d-dimensional random function of $\boldsymbol{X}$ with a positive definite sample covariance and $\sum_{X=x} \nu_x = 0$. The assumption of $\sum_{X=x} \nu_x = 0$ is for centering to have zero mean. It is not essentially required, so it may be unconstrained. In the estimation, ν_x is replaced with $\boldsymbol{Y}_i$ and $\sum_{i=1} \boldsymbol{Y}_i = 0$ by centering the observation of $\boldsymbol{Y}_i$. It is additionally assumed that $S(\boldsymbol{\Gamma})$ is reducing subspace of Σ under model (2.5). Also this can be equivalently stated that $\Sigma = \boldsymbol{\Gamma}\Omega\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\Omega_0\boldsymbol{\Gamma}_0^T$, where $\boldsymbol{\Gamma}_0 \in \mathbb{R}^{r\times(r-d)}$ with $\boldsymbol{\Gamma}_0^T\boldsymbol{\Gamma}_0 = I_{r-d}$ and $\boldsymbol{\Gamma}_0^T\boldsymbol{\Gamma} = 0, \Omega = \boldsymbol{\Gamma}^T\Sigma\boldsymbol{\Gamma}$ and $\Omega_0 = \boldsymbol{\Gamma}_0^T\Sigma\boldsymbol{\Gamma}_0$. The primary interest is placed onto the estimation of Γ. The maximum likelihood estimation (MLE) approach is a natural choice , because the normal distribution of ε is assumed. To construct the likelihood function, consider an orthogonal transformation of $\boldsymbol{Y}$ to $(\Gamma, \Gamma_0)^T\boldsymbol{Y}$.

$$\begin{aligned}\text{cov}(\boldsymbol{Y}) &= \boldsymbol{\Gamma}\text{cov}(\nu_x)\boldsymbol{\Gamma}^T + \Sigma + \boldsymbol{\Gamma}\text{cov}(\nu_x, \varepsilon) = \boldsymbol{\Gamma}\text{cov}(\nu_x)\boldsymbol{\Gamma}^T + \Sigma \\ &= \boldsymbol{\Gamma}\text{cov}(\nu_x)\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}\Omega\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\Omega_0\boldsymbol{\Gamma}_0^T\end{aligned}$$

The log likelihood for $\Gamma_0^T\boldsymbol{Y}$ without constant terms is as follows:

$$\mathcal{L}(\mathrm{B}_0) = -\frac{n}{2}\log|\Omega_0| - \frac{1}{2}trace\left(\mathrm{B}_0^T\mathbb{Y}^T\mathrm{B}_0\Omega_0^{-1}\right) = -\frac{n}{2}\log\left|\mathrm{B}_0^T\hat{\Sigma}_y\mathrm{B}_0\right|$$

where B_0 is the value for $\boldsymbol{\Gamma}_0$ and $\hat{\Sigma}_{\mathrm{y}} = \mathbb{Y}^T\mathbb{Y}/n$ is the sample covariance of $\boldsymbol{Y}$.

The log likelihood for $\mathbf{\Gamma}^T\boldsymbol{Y}$ is as follows:

$$\mathcal{L}(\mathrm{B}) = -\frac{n}{2}\mathrm{log}\,|\Omega| - \frac{1}{2}trace\left(\sum_{i=1}^{n}(\mathrm{B}^TY_i - \nu_x)(\mathrm{B}^TY_i - \nu_x)^T\Omega^{-1}\right)$$

where B is the value for $\mathbf{\Gamma}$. Since ν_x can be B^TY_i, Ω is not estimable. Therefore, this part does not contribute to the estimation of Γ. Let Υ be the value of Ω. With $\Upsilon \in \mathbb{S}^{u\times u}$ fixed, we have $\mathcal{L}(\Upsilon) = -(n/2)\mathrm{log}\,|\Upsilon|$. Finally, the full log-likelihood is given as follows:

$$\mathcal{L}(\mathrm{B}_0, \Upsilon) = -\frac{n}{2}\mathrm{log}\left|\mathrm{B}_0^T\hat{\Sigma}_{\mathrm{y}}\mathrm{B}_0\right| - (n/2)\mathrm{log}\,|\Upsilon|$$

For any fixed matrix of Υ, the likelihood is maximized, when B_0 becomes $(\hat{\gamma}_{(p-u)+1}, ..., \hat{\gamma}_p)$ where $\hat{\Sigma}_{\mathrm{y}} = \sum_{i=1}^{p}\hat{\lambda}_i\hat{\gamma}_i\hat{\gamma}_i^T$ with $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p$. Since $\mathcal{S}(\Gamma)$ is orthogonal complement of $\mathcal{S}(\Gamma_0)$, the MLE of Γ should be $\hat{\Gamma} = (\hat{\gamma}_1, ..., \hat{\gamma}_u)$, which is the eigenvectors corresponding to the first u largest eigenvalues of $\hat{\Sigma}_{\mathrm{y}}$. This dimension reduction under model (2.5) is called PRR.

Technically, model (2.5) is equivalent to the envelope model Cook *et al.* (2010) as follows :

$$\boldsymbol{Y} = \boldsymbol{\alpha} + \mathbf{\Gamma}\nu\boldsymbol{X} + \boldsymbol{\varepsilon} \tag{2.6}$$

where $\Sigma = \Sigma_1 + \Sigma_2$ with $\Sigma_1\Sigma_2 = 0$ and $\Sigma_1 = \Gamma\Omega\Gamma^T$ and $\Sigma_2 = \Gamma_0\Omega_0\Gamma_0^T$ where $r \times (r-u)$ matrix Γ_0 is the orthogonal complement of Γ, $\Omega = \Gamma^T\Sigma\Gamma$ and $\Omega_0 = \Gamma_0^T\Sigma\Gamma_0$. Using the envelope, Cook *et al.* (2010) developed an alternative model for classical multivariate linear regression :

$$\boldsymbol{Y} = \alpha + \boldsymbol{\beta}\boldsymbol{X} + \boldsymbol{\varepsilon} \tag{2.7}$$

where $\boldsymbol{Y} \in \mathbb{R}^r, \boldsymbol{X} \in \mathbb{R}^p$, $\alpha \in \mathbb{R}^{r\times 1}$ is an unknown intercept, $\boldsymbol{\beta} \in \mathbb{R}^{r\times p}$ is an unknown coefficient matrix, and $\boldsymbol{\varepsilon} \sim N(0, \Sigma)$ with $\boldsymbol{X} \perp\!\!\!\perp \boldsymbol{\varepsilon}$. Γ can fully explain $\boldsymbol{\beta}$ because of $\boldsymbol{\beta} = \mathbf{\Gamma}\boldsymbol{\nu}$ for an $u \times p$ matrix $\boldsymbol{\nu}$. Since $\boldsymbol{\beta} = \mathbf{\Gamma}\boldsymbol{\nu}$, $\boldsymbol{\beta}$ can estimated using likelihood functions involving the lower-dimensional matrix Γ. For efficient way in estimating $\boldsymbol{\beta}$, Cook *et al.* (2010) assume that there exists the Σ-envelope $\mathcal{E}_\Sigma(\mathcal{B})$ of $\mathcal{B}$, where $\mathcal{B} = \mathcal{S}(\beta)$. Letting dim$(\mathcal{B}) = d$ and dim$\{\mathcal{E}_\Sigma(\mathcal{B})\} = u$, the relation that $0 \leq d \leq u \leq r$ holds.

In PRR, $\mathbf{\Gamma}$ is estimated only through the marginal information of $\boldsymbol{Y}$ without utilizing information on $\boldsymbol{X}$. Basically, the purpose of the regression is to study of the association between $\boldsymbol{Y}$ and $\boldsymbol{X}$. The incorporation of $\boldsymbol{X}$ may potentially improve the estimation of $\mathbf{\Gamma}$. So, we set $\nu_x = \psi f_x$:

$$\boldsymbol{Y} = \mathbf{\Gamma}\psi f_x + \varepsilon \tag{2.8}$$

where ψ is unknown $d \times q$ matrix, and $f_x \in \mathbb{R}^q$ is a known vector-valued function of the predictor with $\Sigma_x f_x = 0$. Here are several notations.

$\mathbb{Y}$: the $n \times r$ data matrix for the responses

$\mathbb{X}$: the $n \times p$ data matrix for the predictors

$\mathbb{F}$: the $q \times n$ matrix constructed by stacking f_x^T and $P_{\mathbb{F}} = \mathbb{F}\left(\mathbb{F}^T\mathbb{F}\right)^{-1}\mathbb{F}^T$

Let $\hat{\Sigma}_{\text{fit}} = \mathbb{Y}^T P_{\mathbb{F}}\mathbb{Y}/n$ and $\hat{\Sigma}_{\text{res}} = \hat{\Sigma}_{\text{y}} - \hat{\Sigma}_{\text{fit}}$.It is noted that $P_{\mathbb{F}}\mathbb{Y}$ in $\hat{\Sigma}_{\text{fit}}$ is a moment estimator of $\text{cov}(Y, f_x)$. So, $\hat{\Sigma}_{\text{fit}}$ is a consistent estimator of $\text{cov}(Y, f_x)\text{cov}(Y, f_x)^T$. As candidates of f_x, Yoo (2018) suggested $\boldsymbol{X}, \boldsymbol{X}^2$, $\exp(\boldsymbol{X})$, their combinations and the cluster indicator of $\boldsymbol{X}$ acquired from the K-means clustering algorithm. Again, $\boldsymbol{Y}$ is transformed to $(\mathbf{\Gamma}^T\boldsymbol{Y}, \mathbf{\Gamma}_0^T\boldsymbol{Y})^T$.

Under model (2.8), the MLE of $\mathbf{\Gamma}$ does not have a close form. For $\Gamma^T\boldsymbol{Y}$ is as follows:

$$\begin{aligned}\mathcal{L}(\mathrm{B}) &= -\frac{n}{2}\log|\Omega| - \frac{1}{2}trace\left(\sum_{i=1}^{n}(\mathrm{B}^TY_i - \psi f_x)(\mathrm{B}^TY_i - \psi f_x)^T\Omega^{-1}\right)\\ &= -\frac{n}{2}\log|\Omega| - \frac{1}{2}trace\left((\mathbb{Y}\mathrm{B} - \mathbb{F}\psi^T)^T(\mathbb{Y}\mathrm{B} - \mathbb{F}\psi^T)\Omega^{-1}\right)\end{aligned}$$

With Ω fixed, ψ^T is maximized at $\psi^T = (\mathbb{F}^T\mathbb{F})^{-1}\mathbb{F}^T\mathbb{Y}\mathrm{B}$. Replacing it and updating the log-likelihood, then it goes to:

$$\begin{aligned}\mathcal{L}(\mathrm{B}) &= -\frac{n}{2}\log|\Omega| - \frac{1}{2}trace\left((\mathbb{Y}\mathrm{B} - \mathbb{F}(\mathbb{F}^T\mathbb{F})^{-1}\mathbb{F}^T\mathbb{Y}\mathrm{B})^T(\mathbb{Y}\mathrm{B} - \mathbb{F}(\mathbb{F}^T\mathbb{F})^{-1}\mathbb{F}^T\mathbb{Y}\mathrm{B})\Omega^{-1}\right)\\ &= -\frac{n}{2}\log|\Omega| - \frac{1}{2}trace\left((\mathbb{Y}\mathrm{B} - P_{\mathbb{F}}\mathbb{Y}\mathrm{B})^T(\mathbb{Y}\mathrm{B} - P_{\mathbb{F}}\mathbb{Y}\mathrm{B})\Omega^{-1}\right)\\ &= -\frac{n}{2}\log|\Omega| - \frac{1}{2}trace\left(\mathrm{B}^T(\mathbb{Y} - P_{\mathbb{F}}\mathbb{Y})^T(\mathbb{Y} - P_{\mathbb{F}}\mathbb{Y})\mathrm{B}\Omega^{-1}\right)\\ &= -\frac{n}{2}\log\left|\mathrm{B}^T\hat{\Sigma}_{\text{res}}\mathrm{B}\right|\end{aligned}$$

The full likelihood for $\mathbf{\Gamma}$ is as follows:

$$\mathcal{L}\left(\mathrm{B}_0, \mathrm{B}\right) = -\frac{n}{2}\log\left|\mathrm{B}_0^T\hat{\Sigma}_{\mathrm{y}}\mathrm{B}_0\right| - \frac{n}{2}\log\left|\mathrm{B}^T\hat{\Sigma}_{\mathrm{res}}\mathrm{B}\right| \tag{2.9}$$

It is because the log-likelihood for $\mathbf{\Gamma}^T\boldsymbol{Y}$ is the same as PRR, so $\mathcal{L}(\mathrm{B}_0)$ is as:

$$\mathcal{L}(\mathrm{B}_0) = -\frac{n}{2}\log\left|\mathrm{B}_0^T\hat{\Sigma}_{\mathrm{y}}\mathrm{B}_0\right|$$

Therefore, the MLE of $\mathbf{\Gamma}$ depend on both $\hat{\Sigma}_{\mathrm{y}}$ and $\hat{\Sigma}_{\mathrm{res}}$ unlike PRR which depends on only through $\hat{\Sigma}_{\mathrm{y}}$. Yoo (2018) recommends a sequential selection algorithm among a set of all the eigenvectors of $\hat{\Sigma}_{\mathrm{y}}, \hat{\Sigma}_{\mathrm{fit}}$, and $\hat{\Sigma}_{\mathrm{res}}$ following the suggestion in Cook (2007). This dimension reduction under model (2.8) is called PFRR.

Lastly, unstructured principal fitted response reduction(UPFRR) is based in model (2.5). The difference from model (2.5) is the structure of Σ along with $\mathbf{\Gamma}$. Yoo (2019) showed the relationship between Σ and Σ_{y} for the invariant condition so that $\mathcal{S}(\Sigma\Gamma) \subseteq \mathcal{S}(\Sigma)$ iff and only if $\mathcal{S}(\Sigma_{\mathrm{y}}\Gamma) \subseteq \mathcal{S}(\Sigma)$. That is the invariant condition for Σ is equivalent to that for Σ_{y}. Then $\mathcal{S}(\mathbf{\Gamma})$ is not required to be a reduction subspace of Σ and invariant subspace $\mathcal{S}(\mathbf{\Gamma})$ of Σ is equivalent to that of Σ_{y}. Also it holds that $E(\boldsymbol{Y}|\boldsymbol{X}) = E(P^T_{\mathbf{\Gamma}(\Sigma_{\mathrm{y}})}\boldsymbol{Y}|\boldsymbol{X})$. To utilize the information of predictors of $\mathbf{\Gamma}$, its fitted component model is constructed as:

$$\boldsymbol{Y} = \mathbf{\Gamma}\psi f_x + \varepsilon \tag{2.10}$$

Define E_d is to be the first d largest eigenvectors of a matrix E and $\mathcal{S}_d(E)$ is to be a subspace spanned by the columns of E_d. Let $\mathrm{B} = \hat{\Sigma}^{-1/2}\hat{\Sigma}_{\mathrm{fit}}\hat{\Sigma}^{-1/2}, B_{\mathrm{res}} = \hat{\Sigma}_{\mathrm{res}}^{-1/2}\hat{\Sigma}_{\mathrm{fit}}\hat{\Sigma}_{\mathrm{res}}^{-1/2}$, and $\mathrm{B}_{\mathrm{y}} = \hat{\Sigma}_{\mathrm{y}}^{-1/2}\hat{\Sigma}_{\mathrm{fit}}\hat{\Sigma}_{\mathrm{y}}^{-1/2}$. And let $\hat{\Lambda} = (\hat{\lambda}_1, ..., \hat{\lambda}_r)$ and $\hat{V} = (\hat{\gamma}_1, ..., \hat{\gamma}_r)$ be the ordered eigenvalues and corresponding eigenvectors of $\mathrm{B}_{\mathrm{res}}$. Define $\hat{K}_d = diag(0, ..., 0, \hat{\lambda}_{d+1}, ..., \hat{\lambda}_r)$. Then, under model (2.10), we have the following results:

(a) $\hat{\mathcal{S}}(\Gamma) = \hat{\Sigma}^{1/2}\mathcal{S}_d(\mathrm{B})$ or $\hat{\Gamma} = \hat{\Sigma}^{1/2}\mathrm{B}_d$

(b) $\hat{\Sigma} = \hat{\Sigma}_{\mathrm{res}} + \hat{\Sigma}_{\mathrm{res}}^{1/2}\hat{V}\hat{K}_d\hat{V}^T\hat{\Sigma}_{\mathrm{res}}^{1/2} = \hat{\Sigma}_{\mathrm{res}}^{1/2}(I_r + \hat{V}\hat{K}_d\hat{V}^T)\hat{\Sigma}_{\mathrm{res}}^{1/2}$

The MLE for $\mathbf{\Gamma}$ and Σ and the likelihood is as follows:

$$\mathcal{L} = -\frac{n}{2}\log\left|\hat{\Sigma}_{\text{res}}\right| + \frac{n}{2}\sum_{i=d+1}^{q}\log(1+\hat{\lambda}_i)$$

It is easily noted that Σ is estimated by $\hat{\Sigma}_{\text{res}}$, if $q = d$. Therefore, if selecting f_X to have smaller dimension, $\hat{\Sigma}$ can be replaced with $\hat{\Sigma}_{\text{res}}$, but its replacement is not recommended for relatively larger dimension than d. One of the popular standardization of a set of variables is to have its sample covariance matrix equal to the identity matrix. In this context, the standardization is equal to $\hat{\Sigma}_{\text{y}}^{-1/2}\boldsymbol{Y}$. Then the following relations implies that $\hat{\Sigma}_{\text{y}}^{-1/2}\boldsymbol{Y}$ yields the same reduction results as the original scale $\boldsymbol{Y}$.

$$\hat{\mathcal{S}}(\mathbf{\Gamma}) = \hat{\Sigma}^{1/2}\mathcal{S}_d(B) = \hat{\Sigma}_{\text{res}}^{1/2}\mathcal{S}_d(B_{\text{res}}) = \hat{\Sigma}_{\text{y}}^{1/2}\mathcal{S}_d(B_{\text{y}})$$

Therefore, the UPFRR is invariant under the standardization of $\boldsymbol{Y}$.

In summary, for PRR, the covariance matrix Σ cannot be restored because Ω is not estimable. In PFRR, Ω and Ω_0 can be estimated with $\hat{\Gamma}^T\hat{\Sigma}_{\text{res}}\hat{\Gamma}$ and $\hat{\Gamma}_0^T\hat{\Sigma}_{\text{y}}\hat{\Gamma}_0$, respectively. Sample version of Σ is possibly constructed as:

$$\Sigma = \hat{\Gamma}\hat{\Gamma}^T\hat{\Sigma}_{\text{res}}\hat{\Gamma}\hat{\Gamma}^T + \hat{\Gamma}_0\hat{\Gamma}_0^T\hat{\Sigma}_{\text{y}}\hat{\Gamma}_0\hat{\Gamma}_0^T$$

It should be noted that the different sample quantities for $\hat{\Sigma}_{\text{res}}$ and $\hat{\Sigma}_{\text{y}}$ are used for Γ and Γ_0. However, it does not coincide with its population structure. Normally the dimension d of Γ is unknown, but it is assumed to be known for PRR and PFRR. Since both use likelihood functions, a likelihood ratio test(LRT) should be a natural choice. Therefore, the dimension estimation by LRT can be done with $\chi^2_{q(r-m)}$ in case of PFRR, while it is not plausible in PRR because Σ is not estimable. According to Cook (2007) (Section 2.2), for a symmetric matrix A, any invariant subspace of A is a reducing subspace. Therefore, an invariant subspace of Σ becomes a reducing subspace. It indicate that PFRR and UPFRR are the same model. While UPFRR may have a disadvantage in terms of having $r(r-u)$ more parameters compared to PFRR, it is not a significant issue as long as the response reduction subspace is not high. Additionally, it has a significant advantage in Γ estimation due to having a closed form and obtain the equivalent transformation results for the responses. In this context, it appears to closely resemble the theory of an envelope and exhibits a strong connection. Envelope models also stem from sufficient dimension reduction and offer various methods in the context of multivariate regression, including ways to reduce predictors only, reduce response only, and simultaneously reduce both.

2.2.3 Connection to Envelope

Sufficient dimension reduction is originally from Fisher's foundations. In Fisher's paradigm, the statistical process of extracting the relevant information from a sample $\mathcal{D}$ begins by specifying the underlying model up to a parsimonious set of parameters $\theta \in \Theta$. Then a statistic $t(\mathcal{D})$ is said to be sufficient for θ if the distribution of $\mathcal{D}$ given t does not depend on $\theta : \mathcal{D}|(t, \theta = \theta_1) \sim \mathcal{D}|(t, \theta = \theta_2) \forall \theta_1, \theta_2 \in \Theta$, where $\sim$ means identically distributed. This statement is a formal expression of the idea that the reduced data t captures all the information about θ that is contained in $\mathcal{D}$. In particular, if we know t then we can replace $\mathcal{D}$ with $t(\mathcal{D})$ without loss of information on θ. This "reduction by sufficiency", which was referenced in Cox and Mayo (2010) was seen as a brilliant idea at the time and considerable effort was devoted to the study of sufficiency for the next 40 years. Unfortunately, the Fisher's sufficiency had generally fallen out of favor as a paradigm for guiding methodological studies because of its dependence on a known model and the increasingly complex nature of models.(Stigler (1973)). While parsimoniously parameterized models are still used widely, little weight is give to sufficiency. Nevertheless, the Fisher's paradigm is that parsimoniously parameterized models now as recipes for dimension reduction since an estimator $\hat{\theta}$ of θ serves as a reduction of $\mathcal{D}$, apart from any appeal to sufficiency.

The definition of a sufficient reduction from Cook (2007) was inspired by Fisher's paradigm. A reduction $R : \mathbb{R}^p \rightarrow \mathbb{R}^q, q \leq p$, of X is sufficient for Y if at least one of the following hold.

(1) $X|(Y = y_1, R(X)) \sim X|(Y = y_2, R(X)), \forall y_1, y_2$ in the sample space of Y

(2) $Y|X \sim Y|R(X)$,

(3) $Y \perp\!\!\!\perp X|R(X)$.

Statement (1) requires X to be random, but not Y. If we think of X as the total data and Y as a parameter then the statement reduces to the requirement for a sufficient statistic. That is, $R(X)$ is sufficient for Y. Statement (2) is the classical regression context where the predictors X are non-stochastic and only Y need be stochastic. In the standard linear regression context with $r = 1, Y = \alpha + \beta^T X + \varepsilon$, we have simply that $R(X) = \beta^T X$. Statement (3) requires (X, Y) be stochastic with a joint distribution. These three statements contain Fisher's fundamental notion of sufficiency and allow adaptation to a variety of contexts. Also, these statements serve to guide a substantial part of contemporary sufficient dimension reduction in Cook (2009) and it can be used to characterize many dimension reduction methods. The envelope model is a new construction introduced by Cook *et al.* (2010) in the context of multivariate linear regression, while Li (2018) described the contemporary state of the art.

Particularly, response envelope model is as follows:

$$\boldsymbol{Y} = \alpha + \beta \boldsymbol{X} + \varepsilon, \quad \boldsymbol{X} \in \mathbb{R}^p, \alpha \in \mathbb{R}^r, \beta \in \mathbb{R}^{r \times p}, \varepsilon \in \mathbb{R}^r \tag{2.11}$$

where $\boldsymbol{Y} \in \mathbb{R}^r$ is multivariate response vector, $\boldsymbol{X} \in \mathbb{R}^p$ is the predictor vector centered at 0 in the sample, the error vectors $\varepsilon \in \mathbb{R}^r$ with mean 0 and covariance matrix $\Sigma \in \mathbb{R}^{r \times r}$.

Let $(\Gamma, \Gamma_0) \in \mathbb{R}^{r \times r}$ which is an orthogonal matrix and let $Y \sim (\Gamma, \Gamma_0)Y$. Two conditions are required for envelope model.

(1) $\Gamma_0^T Y|X \sim \Gamma_0^T Y$

(2) $\Gamma_0^T Y \perp\!\!\!\perp \Gamma^T Y|X$

Condition (1) indicates that $\Gamma_0 Y$ carries no information on β and it presents the immaterial part of Y, while $\Gamma^T Y$ is the material part. And condition (2) says think of Y given X in two parts, $\Gamma^T Y$ and $\Gamma_0^T Y$ which are orthogonal to each other. Cook *et al.* (2010) showed the previous two conditions are equivalent to the following two conditions. We already say $Y \sim (\Gamma, \Gamma_0)Y, Y \in \mathbb{R}^r$, and β is in SDR subspace. Also we can represent $Y = PY + (I - P)Y = PY + QY$ where P is projection matrix and Q is orthogonal complement of P.

(1) $\mathcal{B} \subseteq span(\Gamma)$

(2) $\Sigma = P_\Gamma \Sigma P_\Gamma + Q_\Gamma \Sigma Q_\Gamma, \quad \Gamma$ reduces Σ

Condition (1) can be interpreted as the envelope is the smallest reducing subspace of Σ containing $\mathcal{B}$, denoted by $\mathcal{E}_\Sigma(\mathcal{B})$. Condition (2) is derived by taking the variance on both sides of the equation, $Y = PY + QY$.

In summary, $\Gamma \in \mathbb{R}^{r \times u}$ span the response envelope subspace with $\eta \in \mathbb{R}^{u \times p}$, cov$(\varepsilon) = \Sigma =$ cov(Y) if X is given. β can be represented as $\beta = \Gamma\eta$ then (2.11) can be switched to:

$$\boldsymbol{Y} = \alpha + \Gamma\eta\boldsymbol{X} + \varepsilon, \quad \Sigma = \text{var}(P_\Gamma Y) + \text{var}(Q_\Gamma Y) \tag{2.12}$$

Through this, it can be inferred that β and Σ are linked by Γ and it results in more efficient estimation of β. Especially when there is a significant difference in variation, such as when var$(P_\Gamma Y) \ll$ var$(Q_\Gamma Y)$, the envelope model can provide substantial efficiency gains. It would be a worthwhile endeavor to contemplate the similarities and differences between these approaches, namely the response envelope model and response dimension reduction.

2.3 Projective-resampling informative predictor subspace

Informative predictor subspace (IPS) emerged due to the potential advantages such as no requirements of linearity, constant variance and coverage conditions in methodological which are required in existing SDR methods. Also it is defined to contain the central subspace ($\mathcal{S}_{Y|X}$) and to develop methods for estimating the former subspace. One drawback is to overestimate $\mathcal{S}_{Y|X}$, but this will not be a cause of concern in practice, because the underestimation of $\mathcal{S}_{Y|X}$ should be more problematic. Let's briefly examine the three conditions mentioned above in this paper. Beforehand, let's use the relation of $\mathcal{S}_{Y|X} = \Sigma^{-1/2}\mathcal{S}_{Y|Z}$. The conditions are discussed with the regression of $Y|Z$ rather than of $Y|X$. Let $\nu_Z \in \mathbb{R}^{p \times d}$ orthonormal basis matrix of $\mathcal{S}_{Y|Z}$.

- *Linearity condition* is as follows:

$$E(Z|\eta_Z^T Z = \nu) \text{ is linear in } \nu$$

The main role of the linearity condition has been understood to force subspaces spanned by the columns of kernel matrices $M_\bullet \in \mathbb{R}^{p \times p}$ produced by most SDR methods to be a proper subspace of $\mathcal{S}_{Y|Z}$ such that $\mathcal{S}(M_\bullet) \subseteq \mathcal{S}_{Y|Z}$. Recent classical SDR methods of Cook and Zhang (2014), Hilafu and Yin (2013), Lu and Li (2011), Yoo (2013) and Yoo and Im (2014), including the classical SDR methods such as SIR and SAVE, require the condition. If the predictors $\boldsymbol{X}$ are elliptically distributed, the linearity condition is guaranteed to hold. Hall and Li (1993) show that with large p, the condition may hold to a reasonable approximation in many regressions. According to Li *et al.* (2004), nonlinearity among the predictors can degrade the performance of most estimation methods.

- *Constant variance condition* is as follows:

$$\text{cov}(Z|\eta_Z^T Z) = Q_{\eta_Z}$$

Some SDR methods such as SAVE require the constant variance condition in addition to the linearity condition. Q_{η_Z} is an orthonormal projection operator onto the orthogonal complement of $\mathcal{S}(\eta_Z)$. Suppose the constant variance condition holds, $\text{cov}(Z|Y)$ can be presented as follows:

$$\begin{aligned} cov(Z|Y) &= E\left\{cov(Z|\eta_Z^T)|Y\right\} + cov\left\{E(Z|\eta_Z^T)|Y\right\} \\ &= E\left[E\left\{(Z - E(Z|\eta_Z^T Z))^2|\eta_Z^T Z\right\}|Y\right] + cov\left(P_{\eta_Z} Z|Y\right) \\ &= E\left[E\left\{(Z - P_{\eta_Z} Z)^2|\eta_Z^T Z\right\}|Y\right] + P_{\eta_Z} cov\left(Z|Y\right) P_{\eta_Z} \\ &= E\left\{cov\left(Q_{\eta_Z} Z|\eta_Z^T Z\right)|Y\right\} + P_{\eta_Z} cov\left(Z|Y\right) P_{\eta_Z} \\ &= Q_{\eta_Z} + P_{\eta_Z} cov\left(Z|Y\right) P_{\eta_Z} \end{aligned}$$

Under both linearity and constant variance conditions, the kernel matrix to estimate $\mathcal{S}_{Y|Z}$ for SAVE goes to:

$$\begin{aligned} I_p - cov(Z|Y) &= P_{\eta_Z} - P_{\eta_Z} cov(Z|Y) P_{\eta_Z} \\ &= P_{\eta_Z}\left\{I_p - cov(Z|Y)\right\} P_{\eta_Z} \in \mathcal{S}_{Y|Z} \end{aligned}$$

If Z is normally distributed, the condition holds, or if Z is elliptically contoured, it approximately holds. (Dennis Cook (2000)).The condition can be inspected through a scatterplot matrix of the predictors.

- *Coverage condition* is as follows:

$$\mathcal{S}\left\{E(Z|\eta_Z^T Z)\right\} = \mathcal{S}_{E(Z|Y)}$$

where $\mathcal{S}_{E(Z|Y)}$ stands for a subspace spanned by $E(Z|Y)$ with Y varying in its marginal sample subspace. $E(Z|Y)$ is what SIR constructs as its kernel matrix. The condition allows SIR to contain $\mathcal{S}_{Y|Z}$, and the linearity condition forces that $\mathcal{S}_{E(Z|Y)} = \mathcal{S}_{Y|Z}$. In Cook and Ni (2005), coverage condition is first assumed, and the linearity condition is assumed later.

2.3.1 Informative predictor subspace

Informative predictor subspace is constructed technically based on the coverage condition. In Cook and Ni (2005), the purpose of $\mathcal{S}_\eta$ below was to establish the coverage condition.

$$\mathcal{S}_\eta = \Sigma^{-1}\mathcal{S}\left\{E(X|\eta^T X) - E(X)\right\}$$

This subspace is spanned by $E(X|\eta^T X)$ with $\eta^T X$ varying in its marginal sample subspace. It is noted that all columns of η can be expressed as an average of vectors in $\mathcal{S}_\eta$. It is the original scale version of $\mathcal{S}\left\{E(Z|\eta_Z^T Z)\right\}$ of the coverage condition. This directly indicates that $\mathcal{S}_{Y|X} \subseteq \mathcal{S}_\eta$. If $\mathcal{S}_\eta$ is constructed, it should be noted that $\mathcal{S}_{Y|X}$ can be exhaustively restored without the coverage condition. That is, an indirect estimation of $\mathcal{S}_{Y|X}$ through $\mathcal{S}_\eta$ gives us the benefit of not requiring the coverage condition.

Suppose that M is a $p \times r$ orthonormal basis matrix for a dimension reduction subspace of $Y \in \mathbb{R}^1|X \in \mathbb{R}^p$, where $r \leq p$. Then an informative predictor subspace $\mathcal{S}_M^{IPS}$ for the dimension reduction subspace $\mathcal{S}_M$ is defined as follows:

$$\mathcal{S}_M^{IPS} = \Sigma^{-1}\mathcal{S}\left\{E(\boldsymbol{X}|M^T\boldsymbol{X}) - E(\boldsymbol{X})\right\}$$

With reference to the aforementioned, it is possible to state $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}} \subseteq \mathcal{S}_M^{IPS}$. IPS is constructed as $M^T X$ which can replace the original predictor $\boldsymbol{X}$ without loss of information on $\boldsymbol{Y}|\boldsymbol{X}$, varies in its marginal sample space. This is why $\mathcal{S}_M^{IPS}$ can capture all information of $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}$. Since $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}$ is minimal dimension reduction subspace, we have that $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}} \subseteq \mathcal{S}(M) \subseteq \mathcal{S}_M^{IPS}$. Yoo (2016) shows the properties of $\mathcal{S}_M^{IPS}$ as follows:

(1) For a non-singular matrix A, $A^{-1}\mathcal{S}_M^{IPS}$ is an IPS for $Y|A^T X$.

(2) Suppose that $\phi \in \mathbb{R}^{p\times q}$ and $M \in \mathbb{R}^{p\times r}$ are orthonormal basis matrices of dimension reduction subspaces of $Y|X$ and that $\mathcal{S}_\phi^{IPS}$ and $\mathcal{S}_M^{IPS}$ denote their respective informative predictor subspaces. If $\mathcal{S}(M) \subseteq \mathcal{S}(\phi)$, it holds that $\mathcal{S}_M^{IPS} \subseteq \mathcal{S}_\phi^{IPS}$.

Suppose B_Z spans a dimension reduction subspace for $Y|Z$. Compatibly, property (1) implies that $\mathcal{S}_M^{IPS} = \Sigma^{-1/2}\mathcal{S}_{M_Z}^{IPS}$. So, usage of the standardized predictor Z can save numerical instability in its sample estimation. Property (2) indicates that the informative predictor subspace for η, which spans $\mathcal{S}_{Y|X}$, is minimal among all possible informative predictor subspaces. The informative predictor subspace for η is called the *central informative predictor subspace*, $\mathcal{S}_{Y|X}^{IPS}$, which is the primary interest to estimate. If we already know η, it would not be necessary to estimate $\mathcal{S}_{Y|X}^{IPS}$, because the goal of SDR is to know $\mathcal{S}_{Y|X}$. In practice, the estimation of $\mathcal{S}_{Y|X}^{IPS}$ is needed to be done without knowing η, so Yoo (2016) suggests to replace $\eta^T X$ by the K-means clusters of X. The response Y is involved in the estimation of $\mathcal{S}_{Y|X}^{IPS}$ by slicing it within each cluster of X.

Denote C_x and S_y as the K-means cluster and slice indicator of X and Y. With this, Yoo (2016) propose a method to estimate $E\left\{E\left(X|S_y, C_x\right)|C_x\right\}$ to recover $\mathcal{S}_{Y|X}$ through $\mathcal{S}_{Y|X}^{IPS}$. With sample data, the following matrix $\hat{\eta}_{h_c,h_s}^C \in \mathbb{R}^{p\times(h_c h_s)}$ is computed:

$$\hat{\eta}_{h_c,h_s}^C = \frac{n_{c,s}}{n}\left(\bar{X}_{c,s} - \bar{X}\right), \quad c = 1, ..., h_c \quad \text{and} \quad s = 1, ..., h_s$$

where $\bar{X}_{c,s}$ and $n_{c,s}$ stand for the mean vector of X and the sample size for a subsample corresponding to the sth slice of Y within the cth cluster of X, and $\bar{X}$ is the sample mean vector, which is equal to $(1/n)\sum_{i=1}^n X_i$. Then the matrix $\hat{\eta}_{h_c,h_s}^C$ is a kernel matrix to estimate $\mathcal{S}_{Y|X}$ through $\mathcal{S}_{Y|X}^{IPS}$. Therefore, $\mathcal{S}_{Y|X}$ is estimated by a subspace spanned by the eigenvectors corresponding to its non-zero eigenvalues. This approach is called *clustering conditional mean* (CCM).

The CCM is the method of clustering X, while new method is to cluster two-dimensional $\hat{\eta}_p^T$. $\hat{\eta}_p$ is as follows:

$$\hat{\eta}_p = (\hat{\eta}_{OLS}, \hat{\eta}_{rpHd}) \in \mathbb{R}^{p\times 2}$$

where $\hat{\eta}_{rpHd}$ is the largest eigenvector from residual-based principal Hessian directions (rpHd). The $p\times 2$ matrix $\hat{\eta}_p$ is composed of the ordinary least squares (OLS) and residual-based principal

Hessian directions (rpHd) (Cook (1998)). The reason to choose two methods among many SDR methods is thought to be simpler versions of SIR and SAVE, respectively. The two methods can compensate deficits of each other. OLS works poorly, if a linear trend in regression is weak, while rpHd is good in such situation. So, $\hat{\eta}_p$ can capture linear and non-linear information on the regression. If once the two-dimensional estimate $\hat{\eta}_p$ is constructed, then a categorization of $\hat{\eta}_p^T X$ is done by either clustering or double slicing with a 2×2 scheme. Double slicing with 2×2 means that pick up one of $\hat{\eta}_{OLS}^T X$ and $\hat{\eta}_{rpHd}^T X$ first and categorize it into two groups and then within each group, categorize the other into two groups. Let C_p denote the categorization result of $\hat{\eta}_p^T X$. The categorization of $\hat{\eta}_p^T X$ produces as good estimation as clustering X even failing the linearity condition or coverage condition or both. This new method is called *partially informative conditional mean* (PCM)

2.3.2 Projective resampling method

The definition of the central subspace for multivariate regression and its properties still remain the same as univariate regression. A projective resampling method proposed by Li *et al.* (2008) combines all univariate regression of $t_j^T \boldsymbol{Y}|\boldsymbol{X}$, where $t_j \in \mathbb{R}^r, j = 1, ..., m$ is a random vector generated from a distribution on the unit sphere $\mathbb{S}^r = \{t_j \in \mathbb{R}^r : ||t|| = 1\}$. That is, consider the projected sample $(X_1, t_j^T Y), ..., (X_n, t_j^T Y)$. For each $t \in \mathbb{R}^r$, we have available $p \times p$ positive semi-definite matrix, $M_n(t)$ that estimates $S_{t^T \boldsymbol{Y}|\boldsymbol{X}}$. It means $M(t)$ forms a basis matrix of $\mathcal{S}_{t^T Y|X}$ and $\mathcal{S}(E(M(t))) = \mathcal{S}_{Y|X}$ (Li *et al.* (2008)). For the distribution to generate t_j, the vector t_j should be sampled from $N(0, I_r)$ $n \log(n)$ or $n^{3/2}$ times with normalizing to have length 1. It directly implies that $S_{Y|X}$ can be fully recovered, once $M(t)$ is estimated for each value of t. Since $M(t)$ is for univariate regression of $t^T Y|X$, $M(t)$ can be estimated through usual SDR methods like SIR and SAVE estimation. Once again, here are details of the algorithm for projective resampling.

1. Choose a Monte Carlo sample size m_n that goes to ∞ faster than n, say $n\log(n)$ or $m_n = n^{3/2}$, and generate an iid sample $t_1, ..., t_m$ from the uniform distribution on the unit sphere $\mathbb{S}^r = \{t \in \mathbb{R}^r : \|t\| = 1\}$. Try to take $t_j = G_j/\, \|G_j\|$, where $G_1.,,, G_{m_n}$ are iid $N(0, I_r)$. From

2. Take standardized $X_1, ..., X_n$ as $\hat{Z}_1, ..., \hat{Z}_n$. For each $t_j, j = 1, ..., m_n$, compute $M_n(t_j)$ from $(\hat{Z}_1, t_j^T Y_1), ..., (\hat{Z}_n, t_j^T Y_n)$ using the basic method such as SIR and SAVE :

   $M_n^{SIR}(t_j) = \sum_{l=1}^{h} \hat{p}_l \hat{\mu}_l(t_j) \hat{\mu}_l^T(t_j)$,
   $M_n^{SAVE}(t_j) = \sum_{l=1}^{h} \hat{p}_l (I_p - \hat{Z}_l(t_j)\hat{Z}_l^T(t_j))^2$

   where $\hat{\mu}_l(t_j)$ is the sample mean of $\left\{ \hat{Z}_i : t_j^T Y_i \in J_l \right\}$.

3. Compute $M_{n,m_n} = m_n^{-1} \sum_{j=1}^{m_n} M_n(t_j)$. Let $\hat{v}_1, ..., \hat{v}_d$ be the d eigenvectors of M_{n,m_n} corresponding to its largest eigenvalues. Then span($\hat{v}_1, ..., \hat{v}_d$) will be used as the estimator of $\mathcal{S}_{Y|Z}$, and span($\hat{\Sigma}^{-1/2}\hat{v}_1, ..., \hat{\Sigma}^{-1/2}\hat{v}_d$) will be used as the estimator of $\mathcal{S}_{Y|X}$.

The determination d of $\mathcal{S}_{Y|X}$ is estimated by $G(k)$ as follows:

$$\hat{d} = argmax \{G(k) : k = 0, ..., p-1\}$$
$$G(k) = \frac{n}{2} \sum_{l=1+k}^{p} (log\hat{\lambda}_l + 1 - \hat{\lambda}_l) - C_n k(2p - k + 1)/2$$

where C_n is any sequence of constants satisfying the conditions of theorem 2 and the consistency of the selection of d is demonstrated in Zhu *et al.* (2006). Let $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \hat{\lambda}_p$ be the eigenvalues of the matrix $M_{n,m_n} + I_p$. Because M_{n,m_n} is assumed to be positive semi-definite, $\hat{\lambda}_i \geq 0$ for all $i = 1, ..., p$. Therefore, $M_{n,m_n} + I_p$ can be assumed to be positive semi-definite, so we see that $\hat{\lambda}_i \geq 1$ for all $i = 1, ..., p$.

2.3.3 Projective resampling informative predictor subspace

Let $M(t)$ be a $p \times p$ positive semi-definite matrix such that $\mathcal{S}(M(t)) = \mathcal{S}_{t^TY|X}$. Define $\phi(t)$ as $p \times p$ positive semi-definite matrix such that $\mathcal{S}(\phi(t)) = \mathcal{S}^{IPS}_{M(t)}$, where $\mathcal{S}^{IPS}_{M(t)}$ is an informative predictor subspace for $M(t)$ in the regression of $t^TY|X$. Let η and $\mathcal{S}^{IPS}_{Y|X}$ be a true basis matrix for $\mathcal{S}_{Y|X}$ and the informative predictor subspace for η in a regression of $Y|X$, respectively.

The following relations directly derived from Li *et al.* (2008), and Yoo (2016) will be used without proof:

(a) $\mathcal{S}(M(t)) = \mathcal{S}_{t^TY|X} \subseteq \mathcal{S}_{Y|X}$

(b) $\mathcal{S}(M(t)) \subseteq \mathcal{S}_{Y|X} \subseteq \mathcal{S}(\phi(t))$

(c) $\mathcal{S}(E(M(T))) = \mathcal{S}_{Y|X}$

Along with these relations, new proper relation among $\mathcal{S}_{t^TY|X}, \mathcal{S}_{Y|X}, \mathcal{S}(E(\phi(t)))$, and $\mathcal{S}^{IPS}_{Y|X}$ is as follows:

$$\mathcal{S}(\phi(t)) = \mathcal{S}^{IPS}_{M(t)} \subseteq \mathcal{S}^{IPS}_{Y|X}, \forall t. \tag{2.13}$$

The relation (a) implies that η forms a dimension reduction subspace for $t^TY|X$. Yoo (2016) already showed that $\mathcal{S}^{IPS}_{M(t)} \subseteq \mathcal{S}^{IPS}_{\eta} = \mathcal{S}^{IPS}_{Y|X}$. Also relation (b) indicate that $v \perp M(t)$ for any v orthogonal to $\phi(T)$ for all t. Therefore if $v^T\phi(t)v = 0, v^TM(t)v = 0$ for all t and $E(v^TM(T)v) = v^TE(M(T))v = 0$. In the forthcoming sections, it shall be designated that the lowercase 't' will be represented as 'T' in the central mean subspace. This directly implies that $v \perp E(M(T))$ and v is orthogonal to $S_{Y|X}$. Based on this, we can say:

$$\mathcal{S}(E(M(T))) \subseteq \mathcal{S}(E(\phi(T))) \tag{2.14}$$

Suppose $v \perp \mathcal{S}^{IPS}_{Y|X}$. Then v is orthogonal to $\phi(t)$ for all t by (2.13). So, we have $E(\phi(T))v = E(\phi(T)v) = 0$, hence $v \perp E(\phi(T))$. Based on this, we can say:

$$\mathcal{S}(E(\phi(T))) \subseteq \mathcal{S}^{IPS}_{Y|X} \tag{2.15}$$

The relation (c) with (2.15) directly implies as follows:

$$\mathcal{S}_{t^TY|X} \subseteq \mathcal{S}_{Y|X} \subseteq \mathcal{S}(E(\phi(T))) \subseteq \mathcal{S}^{IPS}_{Y|X} \tag{2.16}$$

(2.16) shows that $\mathcal{S}^{IPS}_{Y|X}$ is unnecessarily big to recover $S_{Y|X}$ exhaustively in multivariate regression. Technically, the restoration of $E(\phi(T))$ is the main target, and $\mathcal{S}(E(\phi(T)))$ will be called a projective resampling informative predictor subspace(PRIPS). Ko and Yoo (2022) shows that if $\mathcal{S}(E(\phi(T)))$ is PRIPS for a regression of $Y|X$, $A^{-1}\mathcal{S}(E(\phi(T)))$ is a PRIPS for $Y|A^TX$. So, the standardized predictor Z can be used for stability in practice and can be back-transformed to the original scale of X like the central subspace and IPS.

There are three methods to estimate $E(\phi(T))$. First approach is called projective resampling mean method. We define $\phi(\hat{T_j})$ as an estimator from CCM and PCM in 2.3.1 for $T_j^T\boldsymbol{Y}|\boldsymbol{X}$. Then an usual moment estimator of $E(\phi(T_j))$ is the sample mean of $\phi(\hat{T_j})$ for $j = 1, ..., m_n$, where T_j is generated from $N(0, I_q)$ and then $T_j = T_j/\, \|T_j\|$ to have unit length and $m_n = n\log(n)$ or $n^{3/2}$. Then we estimate $E(\phi(T))$ as follows:

$$\hat{E}(\phi(T)) = \frac{1}{m_n}\sum_{j=1}^{m_n}\hat{\phi}(T_j)$$

Second approach is called coordinate mean method. Although it is fully informative as seen in section 2.3.2, the collection of the coordinate regression of $Y_k|\boldsymbol{X}, k = 1, ..., r$, should be very informative to $\mathcal{S}_{Y|X}$. The coordinate regression is a special case of T_j. If $T_j^T\boldsymbol{Y} = Y_j$, T_j is the jth canonical basis e_j, whose jth element alone is equal to one, and zeros elsewhere. Instead of random sampling T_jm_n times, simply averaging $\hat{\phi}(e_j), j = 1, ..., r$ from the all coordinate regressions. It would recover $E(\phi(T))$, which is the primary focus of PRIPS.

Third approach is called coordinate-projective resampling mean method. Literally, this combines the two approaches in the above. Although the coordinate mean method is simple and relatively faster than projective resample mean method, it lacks in recovering the information of the dependency of $\boldsymbol{X}$ in cov$(\boldsymbol{Y})$. To overcome this deficit, the projective resampling and coordinate mean methods are combined. Then the number of resampling is needed to be neither $n \log(n)$ nor $n^{3/2}$, and rather it will be significantly smaller than those values. It is because the coordinate means are very informative to $\mathcal{S}_{Y|X}^{IPS}$. In the coordinate mean method, the number of T_j is equal to the dimension of responses, which does not depend on the sample sizes. Also the choice of the number of the resampling is not theoretically derived, but the estimation performance of the projective resampling and coordinate mean methods are to be shown with various numerical studies.

It can be stated that PRIPS is newly defined as a primary target subspace, which is a projective resampling informative predictor subspace. The PRIPS is smaller than the original informative predictor subspace(IPS) but still contain the central subspace, which is a clear advantage to use it.

Chapter 3

Informative Predictor Subspace in Reduced-Rank Responses

3.1 Reduced-rank response regression

From this chapter, we consider the reduced-rank regression over original regression. Reduced rank regression is an extended multivariate linear regression model with the function of dimension reduction. As mentioned earlier, using dimension reduction methods in multivariate regression focus on the reduction of predictors, not responses.

Proposition 3.1.1 *Suppose that $E(\boldsymbol{Y}|\boldsymbol{X}) = E(\boldsymbol{P}^T_{\psi(\Sigma_y)}\boldsymbol{Y}|\boldsymbol{X})$, and the two regressions of $\boldsymbol{Y}|\boldsymbol{X}$ and $\boldsymbol{P}^T_{\psi(\Sigma_y)}\boldsymbol{Y}|\boldsymbol{X}$ have the their own central mean subspaces, which are spanned by the columns of θ and ξ, respectively. Then, $\mathcal{S}(\theta) = \mathcal{S}(\xi)$.*

We can show this by (2.1) in Cook and Setodji (2003). Proof is as follows:

$$\text{cov}\big(P^TY, E(Y^T|X)P|\theta^TX\big) = 0 \Leftrightarrow P^T\,\text{cov}\big(Y, E(Y^T|X)|\theta^TX\big)\,P = 0$$

This implies that $\text{cov}\big(Y, E(Y^T|X)|\theta^TX\big) \in \mathcal{S}(\psi)^{\perp}$.

Then $Q^T\text{cov}\ \big(Y, E(Y^T|X)|\theta^T X\big)\ Q = Q^T\text{cov}\big(Y, E(Y^T|X)PQ\big)\ \theta^T X$ =$Q^T\ (Y, 0) = 0$

This indicates $\text{cov}\big(Y, E(Y^T|X|\theta^T X)\big) \in \mathcal{S}(\psi)$. This statement is contradictory.

Therefore, $\text{cov}\big(Y, E(Y^T|X)|\theta^T X\big) \notin \mathcal{S}(\psi)^{\perp}$. Then $\text{cov}\big(Y, E(Y^T|X)|\theta^T X\big)$ must be equal to zero.Since θ spans mean subspace of $\boldsymbol{Y}|\boldsymbol{X}$, we can say $\mathcal{S}(\xi) \in \mathcal{S}(\theta)$

$E(\boldsymbol{Y}|\boldsymbol{X}) = E(Y|\xi^T X) \Leftrightarrow P^T E(Y|X) = P^T E(Y|\xi^T X) \Leftrightarrow E(P^T Y|X) = E(P^T Y|\xi^T X)$

It means ξ spans mean subspace to $P^T Y|X$, and we can say $\mathcal{S}(\xi) \in \mathcal{S}(\theta)$.

This implies that the regression of $\psi^T\boldsymbol{Y}|\boldsymbol{X}$ can exhaustively restore $\mathcal{S}_{E(\boldsymbol{Y}|\boldsymbol{X})}$. If we assume that $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}} = \mathcal{S}_{E(\boldsymbol{Y}|\boldsymbol{X})}$, the regression of $\psi^T\boldsymbol{Y}|\boldsymbol{X}$ have the same amount of information about $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}$ to the original regression of $\boldsymbol{Y}|\boldsymbol{X}$. We will call the regression $\psi^T\boldsymbol{Y}|\boldsymbol{X}$ as *reduced-rank response regression.*

Lemma 3.1.2 *The following relation holds.*

$$\mathcal{S}_{E(\boldsymbol{Y}|\boldsymbol{X})} = \mathcal{S}_{E(\psi^T\boldsymbol{Y}|\boldsymbol{X})} \subseteq \mathcal{S}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}} \subseteq \mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}$$

If the location regression holds for $\boldsymbol{Y}|\boldsymbol{X}$, then the four subspaces are equal to each other.

The lemma above directly implies that the location regression holds for $\psi^T\boldsymbol{Y}|\boldsymbol{X}$, as long as it holds for $\boldsymbol{Y}|\boldsymbol{X}$. Also, the reduced-rank response regression does not lose any information on $E(\boldsymbol{Y}|\boldsymbol{X})$, and always have useful information on $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}$. Hereafter, our primary interest is to recover $\mathcal{S}^{\phi(T)}_{\boldsymbol{Y}|\boldsymbol{X}}$ through $\psi^T\boldsymbol{Y}|\boldsymbol{X}$. In other words, our research is based on estimating informative predictor subspace through reduced-rank regression. It's the difference from the former study, Ko and Yoo (2022).

3.1.1 Theoretical relations

Our primary focus is given to $\psi^T\boldsymbol{Y}|\boldsymbol{X}$, not the original regression $\boldsymbol{Y}|\boldsymbol{X}$. By Lemma 3.1.2, we have that $\mathcal{S}_{\psi^T Y|X} \subseteq \mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}$, and hence that $\mathcal{S}^{IPS}_{\psi^T Y|X} \subseteq \mathcal{S}^{IPS}_{\boldsymbol{Y}|\boldsymbol{X}}$, regardless of the location regression. Also, $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}} \subseteq \mathcal{S}^{IPS}_{\boldsymbol{Y}|\boldsymbol{X}}$ is certain by section 2.3.1. However, generally, it does not hold that $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}} \subseteq \mathcal{S}^{IPS}_{\psi^T Y|X}$.

The projective-resampling informative predictor subspace $\mathcal{S}^{\phi(T)}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}$ for $\psi^T\boldsymbol{Y}|\boldsymbol{X}$ is guaranteed to be contained in $\mathcal{S}^{\phi(T)}_{\boldsymbol{Y}|\boldsymbol{X}}$. In constructing $\mathcal{S}^{\phi(T)}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}$, the regression $a^T\psi^T\boldsymbol{Y}|\boldsymbol{X}$ has to be considered for a random vector $a \in \mathbb{R}^{d_y \times 1}$, and the vector $\psi a \in \mathbb{R}^{r\times 1}$ should be a possible value of t for the construction of $\mathcal{S}^{\phi(T)}_{\boldsymbol{Y}|\boldsymbol{X}}$. This directly implies that $\mathcal{S}^{\phi(T)}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}} \subseteq \mathcal{S}^{\phi(T)}_{\boldsymbol{Y}|\boldsymbol{X}}$. However, this relation does not theoretically indicates that $\mathcal{S}^{IPS}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}} \subseteq \mathcal{S}^{\phi(T)}_{\boldsymbol{Y}|\boldsymbol{X}}$. That is to say that we cannot guarantee the relation between $\mathcal{S}^{\phi(T)}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}$ and $\mathcal{S}^{IPS}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}$. This relation is summarized the following lemma.

Lemma 3.1.3 *The following two containments hold.*

$$\mathcal{S}^{IPS}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}} \subseteq \mathcal{S}^{IPS}_{\boldsymbol{Y}|\boldsymbol{X}} \quad \textit{and} \quad \mathcal{S}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}} \subseteq \mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}$$

The following two containments are not guaranteed to hold.

$$\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}} \subseteq \mathcal{S}^{IPS}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}} \quad \textit{and} \quad \mathcal{S}^{IPS}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}} \subseteq \mathcal{S}^{\phi(T)}_{\boldsymbol{Y}|\boldsymbol{X}}$$

Therefore, to estimate $\mathcal{S}^{\phi(T)}_{\boldsymbol{Y}|\boldsymbol{X}}$ properly through the reduced-rank canonical informative predictor subspace without assuming the location regression, the kernel matrix for the coordinate mean method of Ko and Yoo (2022) is additionally considered:

$$M_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}^{COMB} = \frac{1}{kd_y + r}\left[\sum_{b=1}^{k}\sum_{i=1}^{d_y} M_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}^{IPS(b[i]])} + \sum_{i=1}^{r} M_{y_i|\boldsymbol{X}}^{IPS}\right] \tag{3.1}$$

where $M_{y_i|\boldsymbol{X}}^{IPS}$ is the kernel matrix for the informative predictor subspace of the coordinate regression $y_i|\boldsymbol{X}$ for $i = 1, ..., r$. As mentioned before in section 2.3.3, the coordinate mean method to estimate PRIPS is simple and faster than projective resample mean method, but it lacks in recovering the information of the dependency of $\boldsymbol{X}$ in cov$(\boldsymbol{Y})$. In response to this limitation, the authors in Ko and Yoo (2022) proposed a combination of the projective resampling and coordinate mean methods. This is the reason why the new kernel matrix also combined two methods.

Since $\psi_i^{(b)}$ and the canonical basis vector for y_i in the construction of (3.1) are the possible values of t for $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}^{\phi(T)}$, the matrix $M_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}^{COMB}$ is a special form of the coordinate-projective resampling mean method in Ko and Yoo(2022). Here, instead of sampling t uninformatively and randomly, the random vector t is informatively selected without losing information on $E(\boldsymbol{Y}|\boldsymbol{X})$. Therefore, the suggested numbers $n\log(n)/r$ or $n^{3/2}/r$ for sampling t can be reduced to kd_y+r. For n=100 and $r = 4$ along with $k = 4$ and $d_y = 2$, the suggested numbers of sampling t should be around 115 or 250, but in the proposed $M_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}^{COMB}$, it just needs 8 times. Finally, the matrix $M_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}^{COMB}$ is newly proposed to estimate $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}^{\phi(T)}$. Let's examine how much we can reduce the sampling size for n=100 in the following table.

Table 3.1: The change in sampling size

		Existing method	New method
$r = 2$	$d_y = 1$	230 or 500	6
	$d_y = 2$		10
	$d_y = 3$		14
$r = 4$	$d_y = 1$	116 or 250	8
	$d_y = 2$		12
	$d_y = 3$		16

3.1.2 Related research

Reduced-rank response regression can be traced back to its origins in reduced-rank regression, without overstating its significance. Reduced-rank regression (Anderson (1951), Izenman (1975), Reinsel *et al.* (2022)). imposes a rank constraint on the regression coefficients in the multivariate linear regression. It is a popular method for reducing the dimensionalities of $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^r$ for better estimation of β at (2.7). Rewriting as follows:

$$Y = \alpha + \beta X + \varepsilon$$

By restricting the rank of the regression coefficient matrix $rank(\beta) = d <$ min(r, p), the total number of parameters is reduced and efficiency in estimation is improved. Mentioned before in section 2.2.3, Cook *et al.* (2015) introduced the envelope structure to the reduced rank regression, which removes the immaterial variation in the response vector and further improves the efficiency gains. Reduced-rank regression allows that $rank(\beta) = d <$ min(p, r), so that we can write the model parameterization as:

$$\beta = AB, A \in \mathbb{R}^{r\times d}, B \in \mathbb{R}^{d\times p}, rank(A) = rank(B) = d. \tag{3.2}$$

where no additional constraints are imposed on A or B. The maximum likelihood estimators for the reduced-rank regression parameters were derived by Anderson (1999), Reinsel *et al.* (2022), and Stoica and Viberg (1996), under various constraints on A and B for identifiability, such as $BB^T = I_d$ or $A^T A = I_d$.

Envelope and reduced-rank regressions have difference perspectives on dimension reduction. Reduced-rank envelope model combines the strengths of both, which mitigates the burden of selecting between them. When one of the two methods behaves poorly, the reduced-rank envelope estimator automatically degenerates towards the other one. The reduced-rank envelope model is as follows:

$$Y = \alpha + \Gamma\eta BX + \varepsilon, \quad \Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T, \quad i = 1, ..., n \tag{3.3}$$

where error, ε follows a normal distribution. It allows that error covariance to be non-constant. This model fits to the responses and predictors using the maximum likelihood estimation. Let u be the dimension of the envelope. When $d < u = r$, then the model is equivalent to a reduced rank regression model. When $d = u$ or $d = p < r$, then B can be taken as the identity matrix and the model reduces to a response envelope model. When the dimension is $d = u = r$, then the envelope model degenerates to the standard multivariate linear regression. When the $u = 0$, it means that X and Y are uncorrelated, and the fitting is different.

In terms of the envelope, $u = r$, there is no immaterial information to be reduced by the envelope method. Then the reduced-rank envelope model degenerates to the reduced-rank regression (3.2) with $\Gamma = I_r$. When the regression coefficient matrix is full rank, reduced-rank regression is equivalent to ordinary least squares and the reduced-rank envelope degenerates to the ordinary envelope model. Two extreme situation are then: if $p > r = 1$ then both methods degenerate to the standard method, which produces no reduction; if $r > p = 1$ then reduced-rank regression cannot provide any response reduction, while reduced-rank envelopes can still gain efficiency.

Moreover, Cook and Zhang (2015) derived the simultaneous envelope model which performs dimension reduction on both X and Y to achieve further efficiency gains than the response envelope model or predictor envelope model. Our study has the advantage of performing dimension reduction through Response Dimension Reduction while estimating the informative predictor subspace by considering both X and Y, thereby reducing the need for extensive sampling. However, it takes a different visual approach from the simultaneous envelope method. We plan to conduct further research to explore the differences between our approach and the simultaneous envelope method.

3.2 Methological development for reduced-rank mean estimation

3.2.1 Construction in population

According to Yoo and Cook (2008), such ψ is not uniquely defined, because the response dimension reduction is not constructed by intersecting all possible response dimension reduction subspaces like $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}$ and $\mathcal{S}_{E(\boldsymbol{Y}|\boldsymbol{X})}$, we had better consider as many candidates of ψ as possible. Let $\psi^{(b)}$ be the candidate matrices for $b = 1, ..., k$. Since it holds that $E(\boldsymbol{Y}|\boldsymbol{X}) = E(P^T_{\psi(\Sigma_y)}\boldsymbol{Y}|\boldsymbol{X})$ for all candidate matrices $\psi^{(b)}$s, the central mean subspace for $\psi^{(b)T}\boldsymbol{Y}|\boldsymbol{X}$ are equal to each other, and their central subspaces are contained in $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}$.

Letting $\mathcal{S}^{IPS(b)}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}$ stand for the informative predictor subspace for $\mathcal{S}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}$. In Section 2.3.1, the properties of $\mathcal{S}^{IPS}_M$ is as $\mathcal{S}^{IPS}_M \subseteq \mathcal{S}^{IPS}_\phi$ if $\mathcal{S}(M) \subseteq \mathcal{S}(\phi)$. Also with Lemma 3.1.2, we can directly force that $\mathcal{S}^{IPS(b)}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}} \subseteq \mathcal{S}^{IPS}_{\boldsymbol{Y}|\boldsymbol{X}}$.

Define $M^{IPS(b)}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}$ be the kernel matrix to span $\mathcal{S}^{IPS(b)}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}$. Let $\psi^{(b)} = (\psi^{(b)}_1, ..., \psi^{(b)}_{d_y})$ and $\mathcal{S}^{IPS(b[i])}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}$ be an informative predictor subspace for a coordinate regression of $\psi_1^{(b)T}\boldsymbol{Y}|\boldsymbol{X}$. Also, define $M^{IPS(b[i])}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}$ as the kernel matrix of $\mathcal{S}^{IPS(b[i])}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}$. Instead of directly focusing on $\mathcal{S}^{IPS(b)}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}$, we adopt projective-resampling informative predictor subspace by Ko and Yoo (2022). According to them, combining the informative predictor subspaces from the coordinate regressions is normally expected to be equal to $\mathcal{S}^{IPS(b)}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}}$ such that

$$\mathcal{S}^{\phi(T)(b)}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}} = \bigcup_{i=1}^{d_y} \mathcal{S}^{IPS(b[i])}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}} \tag{3.4}$$

Then, the following kernel matrix is a natural choice to recover $\mathcal{S}^{\phi(T)(b)}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}}$:

$$M^{\phi(T)(b)}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}} = \frac{1}{d_y} \sum_{i=1}^{d_y} M^{IPS(b[i])}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}} \tag{3.5}$$

If considering all possible $\psi^{(b)}$s, the population kernel matrix $M^{\phi(T)}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}}$ to restore $\mathcal{S}^{\phi(T)}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}}$ can be constructed as follows:

$$M^{\phi(T)}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}} = \frac{1}{k} \sum_{b=1}^{k} M^{\phi(T)(b)}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}} = \frac{1}{k} \sum_{b=1}^{k} \left[\frac{1}{d_y} \sum_{i=1}^{d_y} M^{IPS(b[i])}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}} \right] = \frac{1}{kd_y} \sum_{b=1}^{k} \sum_{i=1}^{d_y} M^{IPS(b[i])}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}} \tag{3.6}$$

Each subspace $\mathcal{S}^{\phi(T)(b)}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}}$ is large enough to contain $\mathcal{S}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}}$, so is $\mathcal{S}^{\phi(T)}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}}$. Lemma 3.1.2 directly indicates that $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}} \subseteq \mathcal{S}^{\phi(T)}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}}$ under the location regression. In the context of Lemma 3.1.2 , we can say that as below.

Theorem 3.2.1 *The following containment holds*

$$\mathcal{S}_{E(\boldsymbol{Y}|\boldsymbol{X})} = \mathcal{S}_{E(\psi^T \boldsymbol{Y}|\boldsymbol{X})} \subseteq \mathcal{S}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}} \subseteq \mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}} \subseteq \mathcal{S}^{\phi(T)}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}} \subseteq \mathcal{S}^{\phi(T)}_{\boldsymbol{Y}|\boldsymbol{X}}$$

We also can conjecture $\mathcal{S}^{\phi(T)}_{E(\psi^T \boldsymbol{Y}|\boldsymbol{X})} \subseteq \mathcal{S}^{\phi(T)}_{\psi^T \boldsymbol{Y}|\boldsymbol{X}} \subseteq \mathcal{S}^{\phi(T)}_{\boldsymbol{Y}|\boldsymbol{X}}$. It shows that it focus the smaller upper bound $\mathcal{S}(E(\phi(T)))$ under $\psi^T \boldsymbol{Y}|\boldsymbol{X}$ to the exhaustive estimation of $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}$. The restoration of $E(\phi(T))$ is the main target, and $\mathcal{S}(E(\phi(T)))$ had been previously referred to as a projective resampling informative predictor subspace (PRIPS).

Before conducting a numerical study, let's discuss the following property for numerical stability. If there is a non-singular $p \times p$ matrix A, $A^{-1}\mathcal{S}(E(\phi(T)))$ is a PRIPS for $\boldsymbol{Y}|\boldsymbol{A^T X}$.

In Section 2.3.3, the relation (b) indicates that

$$\mathcal{S}(\theta_t) = \mathcal{S}^{IPS}_{M_A(t)} = A^{-1}\mathcal{S}^{IPS}_{M(t)} = A^{-1}\mathcal{S}(\phi(T)) = \mathcal{S}(A^{-1}\phi(t)), \forall t.$$

We need to prove that $\mathcal{S}(E((T))) = \mathcal{S}(A^{-1}E(\phi(t)))$ to both sides. First, let v be orthogonal to $\mathcal{S}(E(\theta(T)))$. This implies $v^T E(\theta(T))v = 0$ and also it can be expressed as $E(v^T \theta(T)v) = 0$. Since θ is positive semi-definite, we have $v^T\theta(t)v = 0$ for all t, and the following relation is established:

$$v^T\theta(t)v = 0 \Leftrightarrow v^T A^{-1}\phi(t)v = 0, \forall t$$

$$E(v^T A^{-1}\phi(t)v) = 0 \Leftrightarrow v^T E(A^{-1}\phi(T))v = 0$$

It says that v is orthogonal to $E(A^{-1}\phi(T))$. Therefore, we have shown one side, $\mathcal{S}(A^{-1}E(\phi(t))) \subseteq \mathcal{S}(E((T)))$. Following the same argument above, it is shown the other side, $\mathcal{S}(E((T))) \subseteq \mathcal{S}(A^{-1}E(\phi(t)))$. It completes the proof. So, the standardized predictor $\boldsymbol{Z}$ can be used instead of $\boldsymbol{X}$ for numerical stability and can be back transformed to the original scale $\boldsymbol{X}$ on the central subspace or the informative predictor subspace.

3.2.2 Sample estimation

In reduced-rank response regression, $\psi^T\boldsymbol{Y}|\boldsymbol{X}$, our primary goal is to estimate ψ. For the estimation of a projective resampling informative predictor subspace, Theorem 3.2.1 plays the key role. Say that again, the reduced-rank response regression does not lose any information on $E(\boldsymbol{Y}|\boldsymbol{X})$. It could be the difference from the previous study, Ko and Yoo (2022).

Following the semi-parametric approach in SDR, we extended its application to response dimension reduction using Principal Response Reduction (PRR), Principal Fitted Response Reduction (PFRR), and the UPFRR method based on any assumption of error covariance matrix. Additionally we compare with the initial method (YC) in Yoo and Cook (2008) , too. We will estimate ψ using a sample-based approach. That is, we estimate ψ from $\hat{\psi}_i^{(k)}$, where k= $\{$1: PRR, 2: PFRR, 3: UPFRR, 4: YC$\}$ and $i = 1, ..., d_y$. The pre-selected values for d_y are given by $d_y = 1, ..., r - 1$.

We define $\hat{\psi}_i$ as an estimator from PCM for $\hat{\psi}_i^{(k)T}\boldsymbol{Y}|\boldsymbol{X}$. Then, the central subspace for $\hat{\psi}_i^{(k)T}\boldsymbol{Y}|\boldsymbol{X}$ are equal to each other, and their central subspaces are contained in $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}$. The sample mean of $\hat{\psi}$ where $\psi_i^k = \psi_i^k / \left\|\psi_i^k\right\|$ to have unit length for kernel matrix is as below.

$$\hat{M}_{RRR} = \frac{1}{4d_y}\sum_{k=1}^{4}\sum_{i=1}^{d_y}\hat{M}_{\hat{\psi}_i}^{(k)} \tag{3.7}$$

We call this reduced-rank response (RRR) kernel matrix. It is the sample version of (3.6). As mentioned in Ko and Yoo (2022), to estimate $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}^{IPS}$ through $\mathcal{S}_{\psi^T\boldsymbol{Y}|\boldsymbol{X}}^{IPS}$, the kernel matrix for the coordinate mean method is considered as (3.1). Building on this idea, we propose new kernel matrix called RRRcomb as follows:

$$\hat{M}_{RRRcomb} = \frac{1}{2}\left(\hat{M}_{RRR} + \hat{M}_{CM}\right) \tag{3.8}$$

Let's summarize the above process to estimate ψ in an algorithm as follows.

```
Algorithm
1. Use four methods for the candidate matrices $\hat{\psi}_i^{(k)}$, k=1,2,3,4
    $\psi^{(1)} = \text{PRR}, \psi^{(2)} = \text{PFRR}, \psi^{(3)} = \text{UPFRR}, \psi^{(4)} = \text{YC}$
2. Make unit length for each method, $\psi_i^k = \psi_i^k / \left\|\psi_i^k\right\|$
3. Get the sample mean of $\psi$ from reduced-rank response(RRR) kernel matrix
    $\hat{M}_{RRR} = \frac{1}{4d_y} \sum_{k=1}^{4} \sum_{i=1}^{d_y} \hat{M}_{\hat{\psi}_i}^{(k)}$
4. Get the sample mean of $\psi$ by combining reduced-rank response(RRR) kernel matrix
and kernel from coordinate mean method
    $\hat{M}_{RRRcomb} = \frac{1}{2}\left(\hat{M}_{RRR} + \hat{M}_{CM}\right)$
```

3.3 Simulation

We considered three artificial models for numerical studies where either the linearity condition or the coverage condition is not satisfied. The first model was mimicked from Li *et al.* (2004). For numerical studies, first the variables U_1, e, W_1, W_2 and W_3 were independently generated:

$$U_1 \sim U(0,1), e \sim U(-0.5, 0.5),$$

$$U_2 \sim \log(U_1) + e$$

$$(W_1, W_2, W_3) \overset{iid}{\sim} N(0,1)$$

where $U(a,b)$ represent a uniform distribution between a and b. With these variables, the following 10-dimensional predictors $(X_1, ..., X_{10})^T$ were generated:

$$X_1 = U_1 + W_1, X_2 = U_2 + W_2 + W_3, X_3 = W_1 - W_2, X_4 = W_2, X_5 = W_3$$

$$(X_6, ..., X_{10}) \overset{iid}{\sim} N(0,1)$$

Then the following three multivariate regressions were constructed with the predictors. In each model, the random errors $(\varepsilon_1, ..., \varepsilon_4)$ were independently sampled from $N(0,1)$.

Model 1:

$$\boldsymbol{\eta} = (1, -1, -1, 0, ..., 0)^T$$

$$Y_1 = \exp(0.5\eta^T \boldsymbol{X} + 1) + 0.1\varepsilon_1, Y_2 = \left(\boldsymbol{\eta^T X}\right)^2 + 0.1\varepsilon_2,$$

$$Y_3 = Y_1 + Y_2 + 0.1\varepsilon_3, Y_4 = \left|\eta^T \boldsymbol{X}\right| + 0.1\varepsilon_4$$

Model 2:

$$\boldsymbol{\eta_1} = (1, -1, -1, 0, ..., 0)^T, \boldsymbol{\eta_2} = (0, 1, 0, ..., 0)^T$$

$$Y_1 = exp(0.5^T \boldsymbol{\eta_1^T X} + 1) + 0.1\varepsilon_1, Y_2 = \boldsymbol{\eta_2^T X} + \left(\boldsymbol{\eta_2^T X}\right)^2 + 0.1\varepsilon_2$$

Model 3:

$$\boldsymbol{\eta_1} = (1, -1, -1, 0, ..., 0)^T, \boldsymbol{\eta_2} = (1, 0, 0, ..., 0)^T$$

$$Y_1 = \exp(0.5\boldsymbol{\eta_1^T X} + 1) + 0.1\varepsilon_1, Y_2 = \boldsymbol{\eta_2^T X} + \left(\boldsymbol{\eta_2^T X}\right)^2 + 0.1\varepsilon_2,$$

$$Y_3 = Y_1 + Y_2 + 0.1\varepsilon_3, Y_4 = \left|\boldsymbol{\eta_1^T X}\right| + 0.1\varepsilon_4$$

Model 1-3 have similar coordinate regressions. Coordinate regressions for each model have various type of mean functions such as non-linear mean, symmetric mean, the second-order polynomial mean and so on. The structural dimension of Model 2 and 3 is equal to two, while Model 1 is one. For the comparison of Model 2 and Model 3, we can study how the number of responses affect the estimation of η.

The coordinate and coordinate-projective resampling mean methods are two special variation of the projective mean method. The reason to consider these two methods are more effectively to absorb good information on the central subspace with avoiding heavy numbers of resampling. This indicates that larger dimension of the response should be important to the success of the two methods. This became the ultimate starting point of this research. That is, Model 2 and Model 3 are designed to investigate this. Based on this methodological perspective, before discussing the simulation results, it would be expected that there would be no notable difference in the estimation of η by the three proposed methods, because the responses are 4-dimensional with the structural dimension being one. However, the coordinate mean method should yield worse estimation of η than the other two projective methods.

Each model was generated 500 times, and the projected resampling previous mean method (PRp-pcm), reduced-rank response with pcm (RRR-pcm) and combining RRR and coordinate mean method (RRRcomb) based on PCM were applied to the models. This study aims to compare the variations in estimating ψ between the conventional random projective resampling method, such as PRR and the process of obtaining ψ through reduced rank response regression, such as RRR and RRRcomb. For the projective resampling method, the number of resampling was 500 and it was 100 for the coordinate projective resampling mean method.

To measure how well the $\boldsymbol{\eta_i}$ for each model is estimated, R_i^2 as calculated from a regression of $\boldsymbol{\eta_i^T X}|\boldsymbol{\hat{\eta}^T X}$ for $i = 1, ..., d$, where d represents the true structural dimension of each model. Defining $|r_i| = \left|\sqrt{R_i^2}\right|$, higher values of r_i indicates a better estimation of η_i. As a graphical summary, $|r_i|$ was box-plotted. In the plot, the horizontal axis represents the number of slices of $\boldsymbol{X}$ or $\boldsymbol{\hat{\eta}_p^T X}$ and response. For example, (4,2) means that the number of clusters of $\boldsymbol{X}$ or $\boldsymbol{\hat{\eta}_p^T X}$ is equal to four, and then the response is sliced into two groups within each cluster.

In Ko and Yoo (2022), numerical study showed that

- In model 1 to have one-structural dimension with 4-dimensional responses, there is no difference among three methods; PR,C, CPR
- In model 2, CPR-PCM yields better estimation results than PR-PCM for the second direc-

tion, $(\boldsymbol{\eta}_2^T \boldsymbol{X})$.

- In model 3, CPR-PCM is especially better than C-PCM.

- The PCM-based method is better estimation performance than CCM-based ones.

- Although the estimation of the second directions for model 2 and model 3 are sensitive to it, PCM-based methods are robust to the choice of the number of slices for $\boldsymbol{X}$ or $\hat{\boldsymbol{\eta}}_p^T \boldsymbol{X}$ and Y.

- The choice of $(2, 2)$ provides more robust estimation results than the other choices.

Based on the findings of previous research, this study used PCM as the foundational method for experimentation. First, it should be noted that our expectation above is confirmed. By comparing model 2 and model 3, higher-dimensional responses yield better estimation results partially due to accumulating the information of the central subspace. In model 1 to have one structural dimension with 4-dimensional responses, it's not easy to discern a distinct difference. While the conventional method, PR required 100 resamplings for sufficient accuracy, it can be observed that the RRR and RRRcomb methods achieved comparable results with fewer iterations. Figure 3.1,3.3,3.4,3.5 and 3.7 are performed with when $d_y = 2$. Saying $d_y = 2$ means performing the extraction of the principal eigenvector direction for each of the four methods, and then calculating the mean of these directions 100 times.

There are boxplots and summary statistics of estimated η values for Model 1, 2, and 3 using both the conventional method, PRp-pcm, and two proposed methods. The boxplots represent how closely each method approximated the actual η value, and one could assert that a larger value is indicative of a better estimation. While observing the median (Q2), and Average (Ave) values in the table's summary, it is attempted to indicate the highest value. Upon reviewing tables 3.2 through 3.6, it is evident that the maximum values for Q2 and Ave are consistently highlighted in bold for specific methods. This trend is particularly notable for cases (2,4), (3,4), and (4,3).

Let me examine whether there are differences between the conventional method and the proposed method across different models. Therefore, an investigation will be conducted on three models to observe how much difference is evident in estimating eta, specifically when using the PRp-pcm method based on 100 random samples and estimating eta through Response dimension reduction for cases where $d_y = 2$. This involves employing PRR, PFRR, UPFRR, and YC to determine the mean within the Central mean subspace. Also I will further investigate whether there are differences among the proposed methods for the previously mentioned cases (2,4), (3,4), and (4,3) with respect to changes in the values of d_y, $d_y = 1, 2, 3$ in this study.

When examining Model 1, compared to the conventional method, the boxplots of the two proposed methods exhibit a slightly larger trend. However, considering an error range of ±0.01 for median or average values, and taking into account that $d_y = 2$, it can be inferred that there is an effect of reduced rank response regression.

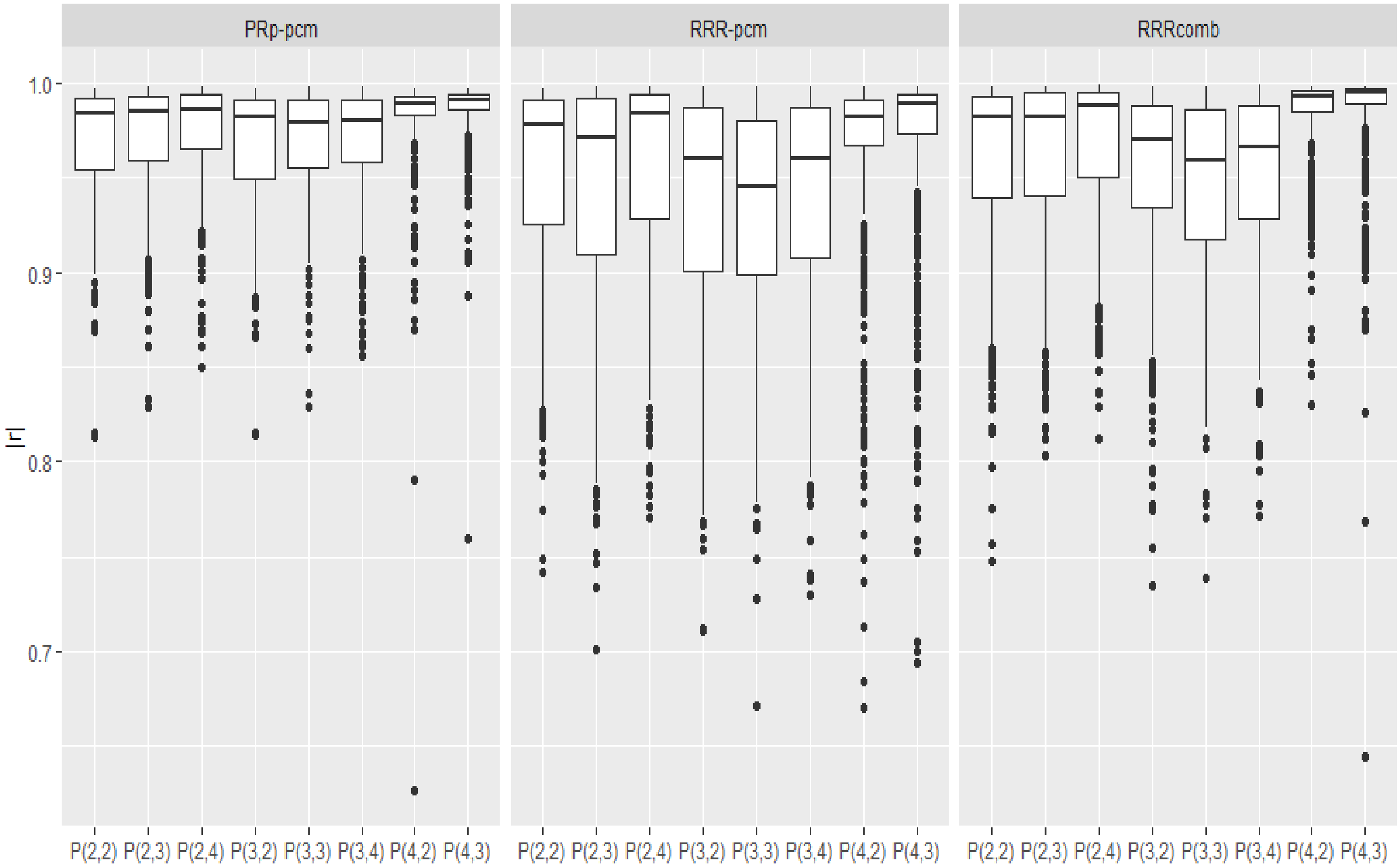

Figure 3.1: Model 1 of |r|s based on PCM

Table 3.2: Summary of |r| depends on three methods in Model 1

	PRp-pcm							
Summary	(2,2)	(2,3)	(2,4)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)
Min	0.8135	0.8296	0.8504	0.8145	0.8286	0.8562	0.6261	0.7592
Q1	0.9542	0.9592	0.9649	0.9497	0.9559	0.9583	0.9833	0.9858
Q2	0.9845	0.9851	0.9859	0.9819	0.9788	0.9799	0.9896	**0.9911**
Ave	0.9703	0.9734	0.9757	0.9682	0.9699	0.9714	0.9838	0.9870
Q3	0.9920	0.9931	0.9937	0.9909	0.9909	0.9911	0.9934	0.9944
Max	0.9978	0.9988	0.9989	0.9985	0.9988	0.9989	0.9981	0.9986
Sd	0.0300	0.0274	0.0256	0.0306	0.0278	0.0265	0.0244	0.0170

	RRR-pcm							
Summary	(2,2)	(2,3)	(2,4)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)
Min	0.7416	0.7006	0.7705	0.7113	0.6712	0.7299	0.6699	0.6938
Q1	0.9255	0.9094	0.9283	0.9004	0.8989	0.9079	0.9669	0.9736
Q2	0.9780	0.9712	0.9839	0.9602	0.9452	0.9607	0.9821	**0.9888**
Ave	0.9510	0.9435	0.9556	0.9373	0.9309	0.9410	0.9645	0.9692
Q3	0.9908	0.9919	0.9945	0.9874	0.9804	0.9870	0.9915	0.9941
Max	0.9986	0.9990	0.9994	0.9992	0.9990	0.9993	0.9990	0.9992
Sd	0.0551	0.0618	0.0534	0.0600	0.0584	0.0568	0.0506	0.0498

	RRRcomb							
Summary	(2,2)	(2,3)	(2,4)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)
Min	0.7471	0.8035	0.8121	0.7342	0.7384	0.7717	0.8302	0.6447
Q1	0.9400	0.9402	0.9500	0.9342	0.9172	0.9285	0.9856	0.9889
Q2	0.9820	0.9823	0.9884	0.9701	0.9591	0.9666	0.9928	**0.9949**
Ave	0.9623	0.9681	0.9692	0.9517	0.9537	0.9560	0.9856	0.9890
Q3	0.9941	0.9953	0.9959	0.9875	0.9879	0.9869	0.9955	0.9970
Max	0.9993	0.9996	0.9996	0.9992	0.9993	0.9991	0.9996	0.9993
Sd	0.0471	0.0420	0.0378	0.0470	0.0453	0.0411	0.0211	0.0195

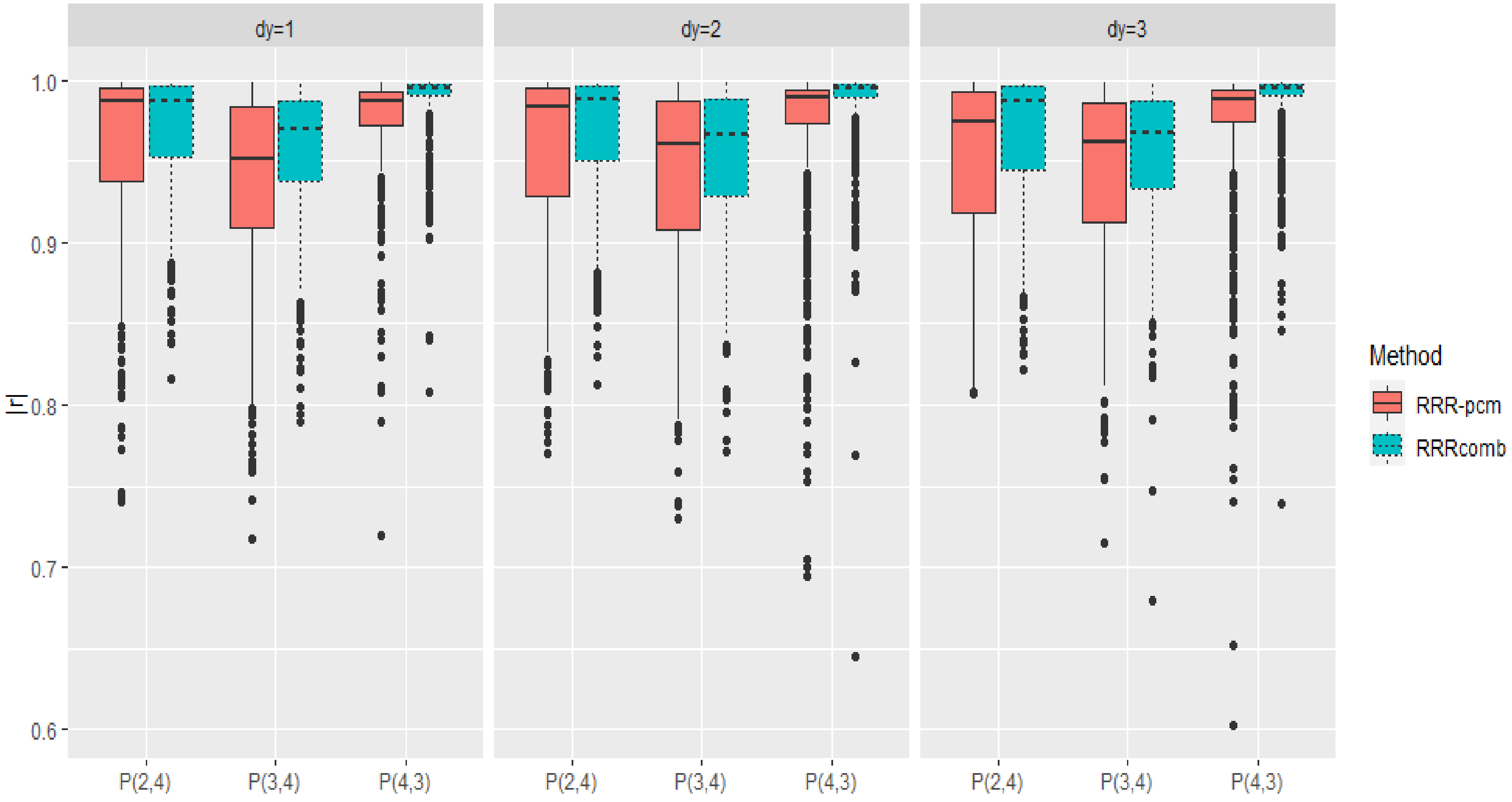

Figure 3.2: The variation of |r| with respect to d_y=1,2,3 in Model 1

From the above graphs Figure 3.2, that doesn't seem to result in such a significant change as d_y varies. This leads us to the conclusion that d_y is no big deal to estimate just η in model 1.

Considering Model 2, which includes both η_1 and η_2, let's first examine the estimation of η_1. Figure 3.3 shows that there is a tendency for the box sizes of the proposed methods to be slightly smaller than those of the conventional method except (3,2), (3,3), and (3,4). For cases such as (2,4), it is noteworthy that all three methods consistently exhibited the highest estimation, as confirmed through Table 3.2. Moreover, the similar trends observed in the two proposed methods suggest that for Model 2, reduced-rank response regression provides a more stable estimation.

Next, turning our attention to the estimation of η_2, Figure 3.4 show that it is evident that the proposed method provides a more stable and reliable estimation compared to the conventional method. Additionally, the RRRcomb values estimated using the average of RRR-pcm and Coordinate mean(CM) methods exhibit a slightly higher and evenly distributed pattern, in contrast to RRR-pcm. For PRp-pcm and RRR-pcm, the highest estimation was observed at (3,4), while for RRRcomb, it was at (2,4). However, this difference is not significantly larger compared to other cases. Furthermore, since Model 2 involves a 2-dimensional y, the analysis was restricted to cases only under the condition of $d_y = 1$.

Model 3, similar to Model 2, estimates both η_1 and η_2. However, unlike Model 2, which has a 2-dimensional Y, Model 3, like Model 1, has a 4-dimensional Y. Therefore, we also examine the changes with respect to d_y, as we did for Model 1. For Model 3, with the conventional method PRp-pcm for η_1, it exhibited consistently high estimation accuracy regardless of the number of clusters for X or the slicing number for Y. On the other hand, using the proposed methods, while the boxplot sizes are slightly larger, the RRRcomb method generally shows higher values compared to the other proposed method for various cluster numbers for X and slicing numbers for Y. Furthermore, as these results are obtained under the condition of $d_y = 2$, it can be interpreted that there is a significant effect.

When looking at the Figure 3.5, although PRp-pcm seems to have estimated the best, considering $d_y = 2$, the proposed method utilizing reduced-rank response could be seen as effective. Looking at Table 3.5, it can be observed that the highest values were found at (2,4). Referring to Figure 3.6, it can be observed that there is little significant impact on the variation of d_y concerning η_1 in Model 3. However, η_2 displayed a distinct pattern. It is evident that the proposed method shows a clear effectiveness compared to the conventional approach specifically concerning η_2. Even though $d_y = 2$, it is evident that the proposed method outperforms the estimation, showcasing its effectiveness in extracting from the central mean subspace. In many cases, good results are achieved regardless of the value for d_y; however, As we can see in Figure 3.8, occasionally, when $d_y = 1$, the estimation performance might not be optimal. Therefore, a default value of 2 is recommended for d_y. We have confirmed that while the conventional method takes approximately 113 minutes of computing time, the proposed RRR-pcm takes about 2.65 minutes, and RRRcomb takes about 3.9 minutes to run the program.

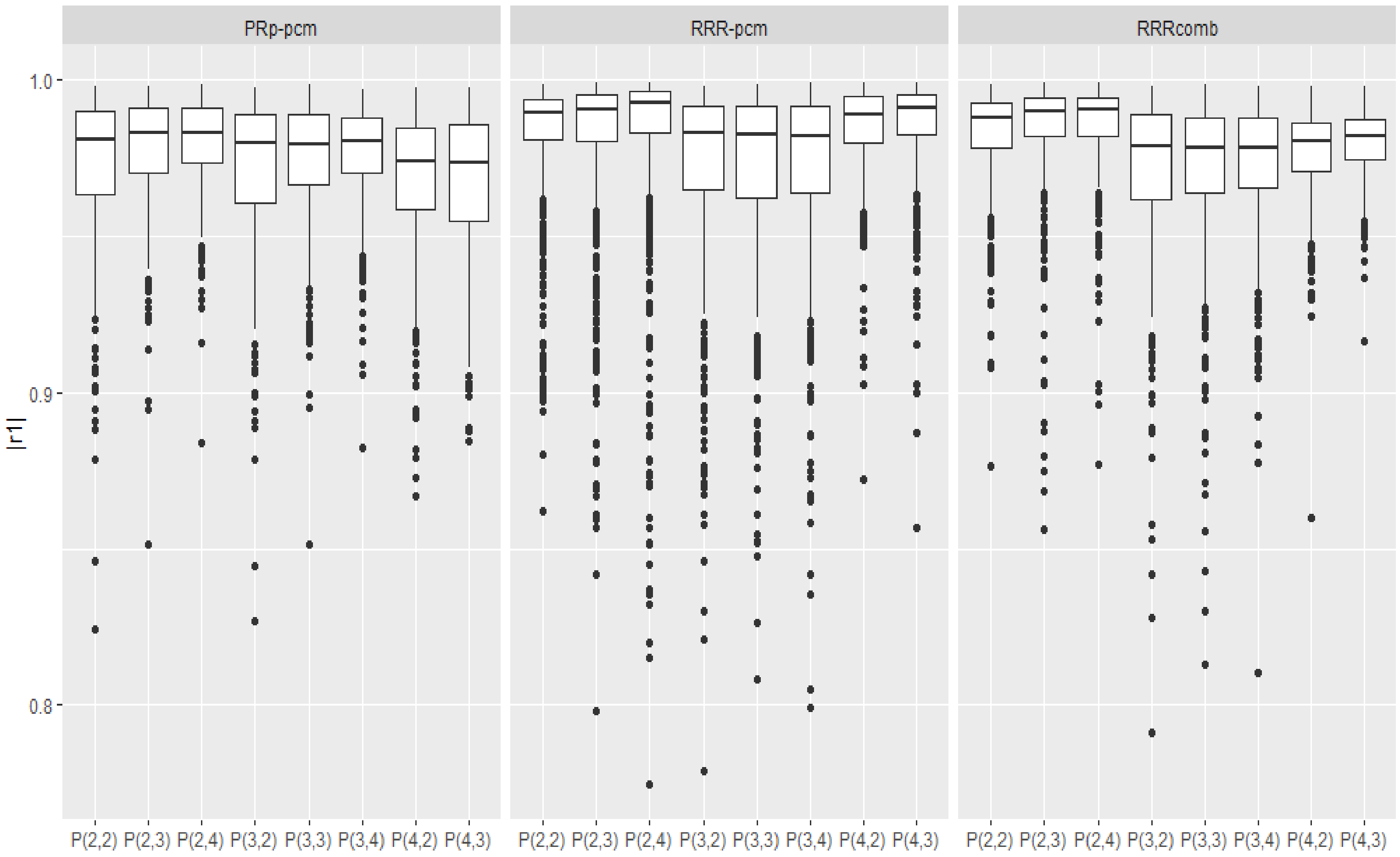

Figure 3.3: Model 2 of $|\mathrm{r}_1|$s based on PCM

Table 3.3: Summary of $|r_1|$ depends on three methods in Model 2

				PRp-pcm				
Summary	(2,2)	(2,3)	(2,4)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)
Min	0.8240	0.8515	0.8840	0.8269	0.8514	0.8823	0.8669	0.8846
Q1	0.9635	0.9703	0.9734	0.9605	0.9663	0.9700	0.9586	0.9547
Q2	0.9809	0.9829	**0.9830**	0.9797	0.9791	0.9803	0.9740	0.9734
Ave	0.9735	0.9777	0.9800	0.9721	0.9749	0.9763	0.9685	0.9676
Q3	0.9899	0.9909	0.9909	0.9887	0.9886	0.9876	0.9847	0.9858
Max	0.9985	0.9983	0.9990	0.9981	0.9992	0.9975	0.9981	0.9978
Sd	0.0229	0.0178	0.0145	0.0229	0.0180	0.0160	0.0220	0.0227

				RRR-pcm				
Summary	(2,2)	(2,3)	(2,4)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)
Min	0.8619	0.7981	0.7749	0.7791	0.8084	0.7994	0.8722	0.8565
Q1	0.9810	0.9803	0.9828	0.9648	0.9621	0.9640	0.9796	0.9824
Q2	0.9893	0.9904	**0.9927**	0.9831	0.9823	0.9819	0.9889	0.9909
Ave	0.9816	0.9793	0.9802	0.9714	0.9717	0.9720	0.9847	0.9855
Q3	0.9938	0.9952	0.9961	0.9916	0.9914	0.9915	0.9944	0.9951
Max	0.9988	0.9993	0.9997	0.9985	0.9989	0.9993	0.9994	0.9992
Sd	0.0218	0.0289	0.0326	0.0315	0.0293	0.0295	0.0146	0.0163

				RRRcomb				
Summary	(2,2)	(2,3)	(2,4)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)
Min	0.8766	0.8561	0.8769	0.7914	0.8130	0.8106	0.8600	0.9162
Q1	0.9782	0.9820	0.9821	0.9616	0.9636	0.9656	0.9709	0.9744
Q2	0.9877	0.9900	**0.9904**	0.9788	0.9782	0.9783	0.9805	0.9816
Ave	0.9827	0.9844	0.9856	0.9715	0.9707	0.9735	0.9773	0.9795
Q3	0.9925	0.9943	0.9942	0.9888	0.9877	0.9879	0.9864	0.9871
Max	0.9989	0.9992	0.9996	0.9986	0.9991	0.9985	0.9984	0.9982
Sd	0.0155	0.0182	0.0148	0.0256	0.0261	0.0208	0.0138	0.0106

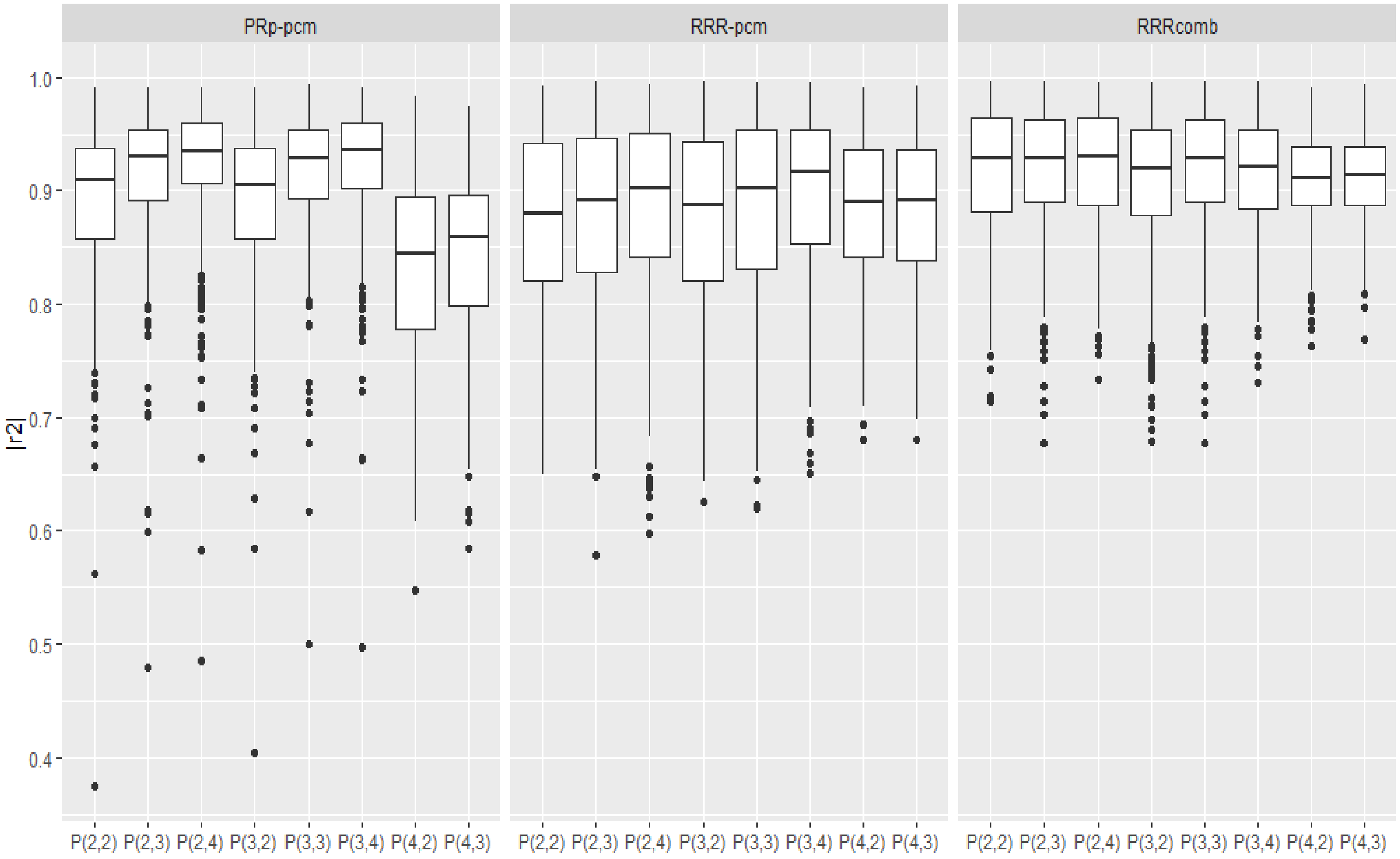

Figure 3.4: Model 2 of $|r_2|$s based on PCM

Table 3.4: Summary of $|r_2|$ depends on three methods in Model 2

				PRp-pcm				
Summary	(2,2)	(2,3)	(2,4)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)
Min	0.3753	0.4797	0.4854	0.4043	0.4996	0.4977	0.5476	0.5842
Q1	0.8584	0.8924	0.9063	0.8575	0.8935	0.9023	0.7778	0.7991
Q2	0.9091	0.9297	0.9349	0.9051	0.9282	**0.9365**	0.8439	0.8594
Ave	0.8927	0.9165	0.9248	0.8921	0.9176	0.9254	0.8306	0.8443
Q3	0.9373	0.9540	0.9600	0.9370	0.9536	0.9602	0.8947	0.8963
Max	0.9929	0.9921	0.9922	0.9924	0.9956	0.9927	0.9854	0.9757
Sd	0.0662	0.0571	0.0534	0.0652	0.0532	0.0511	0.0834	0.0736

				RRR-pcm				
Summary	(2,2)	(2,3)	(2,4)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)
Min	0.6487	0.5785	0.5970	0.6257	0.6203	0.6514	0.6808	0.6804
Q1	0.8206	0.8289	0.8409	0.8213	0.8319	0.8536	0.8414	0.8385
Q2	0.8807	0.8918	0.9016	0.8879	0.9026	**0.9166**	0.8903	0.8915
Ave	0.8758	0.8818	0.8857	0.8756	0.8859	0.8973	0.8873	0.8852
Q3	0.9424	0.9467	0.9506	0.9441	0.9537	0.9540	0.9356	0.9357
Max	0.9938	0.9975	0.9956	0.9981	0.9963	0.9962	0.9924	0.9935
Sd	0.0763	0.0763	0.0791	0.0799	0.0799	0.0726	0.0617	0.0663

				RRRcomb				
Summary	(2,2)	(2,3)	(2,4)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)
Min	0.7144	0.6773	0.7343	0.6785	0.6773	0.7307	0.7627	0.7689
Q1	0.8811	0.8899	0.8879	0.8785	0.8899	0.8839	0.8876	0.8877
Q2	0.9290	0.9286	**0.9304**	0.9199	0.9286	0.9207	0.9117	0.9137
Ave	0.9180	0.9206	0.9215	0.9106	0.9206	0.9158	0.9103	0.9124
Q3	0.9648	0.9633	0.9636	0.9538	0.9633	0.9540	0.9397	0.9394
Max	0.9982	0.9988	0.9973	0.9960	0.9988	0.9980	0.9927	0.9955
Sd	0.0562	0.0553	0.0524	0.0586	0.0553	0.0503	0.0410	0.0371

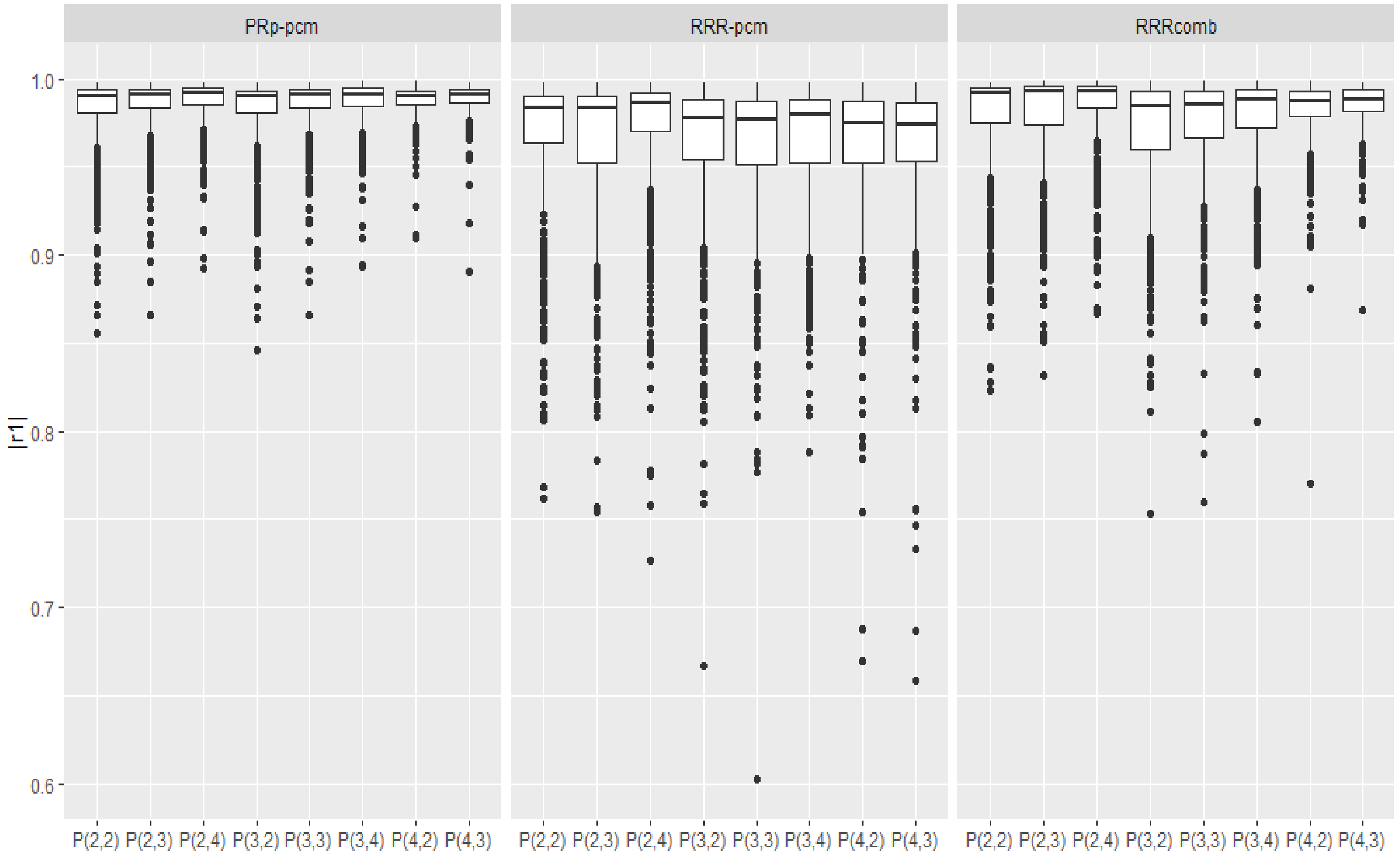

Figure 3.5: Model 3 of $|r_1|$s based on PCM

Table 3.5: Summary of $|\mathrm{r}_1|$ depends on three methods in Model 3

	PRp-pcm							
Summary	(2,2)	(2,3)	(2,4)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)
Min	0.8554	0.8661	0.8927	0.8462	0.8661	0.8933	0.9097	0.8907
Q1	0.9806	0.9840	0.9854	0.9807	0.9836	0.9847	0.9852	0.9867
Q2	0.9898	0.9907	**0.9918**	0.9899	0.9908	0.9910	0.9899	0.9909
Ave	0.9815	0.9852	0.9877	0.9814	0.9851	0.9874	0.9879	0.9891
Q3	0.9936	0.9943	0.9946	0.9931	0.9939	0.9945	0.9929	0.9936
Max	0.9984	0.9991	0.9989	0.9988	0.9991	0.9994	0.9997	0.9987
Sd	0.0214	0.0164	0.0127	0.0219	0.0162	0.0125	0.0091	0.0085

	RRR-pcm							
Summary	(2,2)	(2,3)	(2,4)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)
Min	0.7619	0.7543	0.7270	0.6671	0.6028	0.7884	0.6692	0.6582
Q1	0.9635	0.9518	0.9701	0.9544	0.9510	0.9525	0.9518	0.9530
Q2	0.9832	0.9837	**0.9866**	0.9776	0.9769	0.9793	0.9747	0.9743
Ave	0.9659	0.9639	0.9702	0.9616	0.9600	0.9646	0.9626	0.9613
Q3	0.9904	0.9905	0.9922	0.9880	0.9877	0.9885	0.9869	0.9865
Max	0.9987	0.9984	0.9989	0.9985	0.9984	0.9983	0.9990	0.9988
Sd	0.0425	0.0435	0.0397	0.0432	0.0444	0.0365	0.0397	0.0435

	RRRcomb							
Summary	(2,2)	(2,3)	(2,4)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)
Min	0.8229	0.8321	0.8668	0.7534	0.0218	0.8057	0.7700	0.8691
Q1	0.9751	0.9741	0.9836	0.9599	0.9394	0.9720	0.9787	0.9814
Q2	0.9916	0.9927	**0.9933**	0.9846	0.9818	0.9881	0.9872	0.9887
Ave	0.9798	0.9808	0.9832	0.9729	0.9751	0.9749	0.9798	0.9827
Q3	0.9939	0.9947	0.9949	0.9921	0.9915	0.9918	0.9900	0.9920
Max	0.9990	0.9996	0.9995	0.9988	0.9988	0.9989	0.9984	0.9991
Sd	0.0260	0.0241	0.0198	0.0293	0.0243	0.0260	0.0145	0.0128

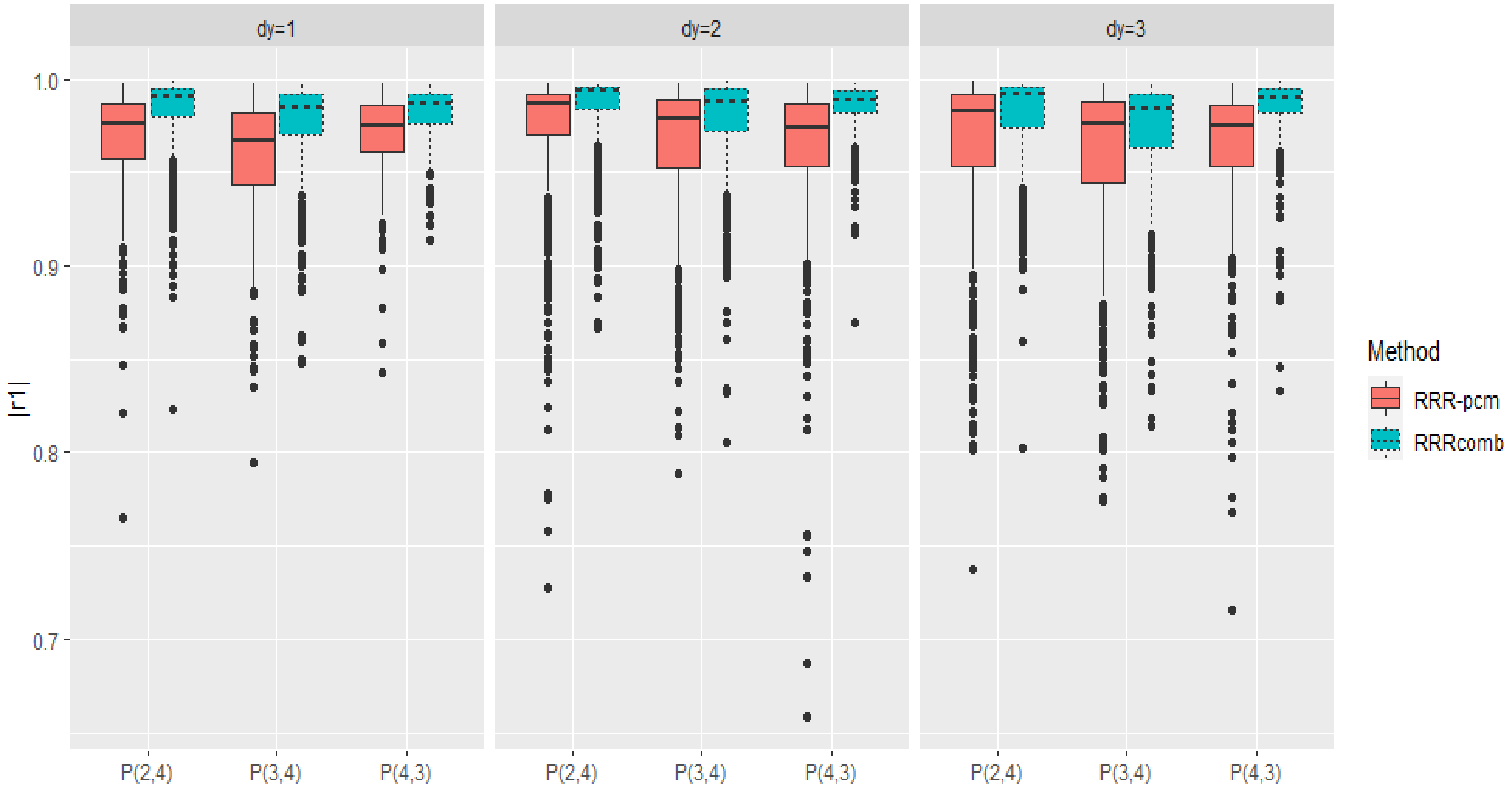

Figure 3.6: The variation of $|\mathrm{r}_1|$s with respect to d_y=1,2,3 in Model 3

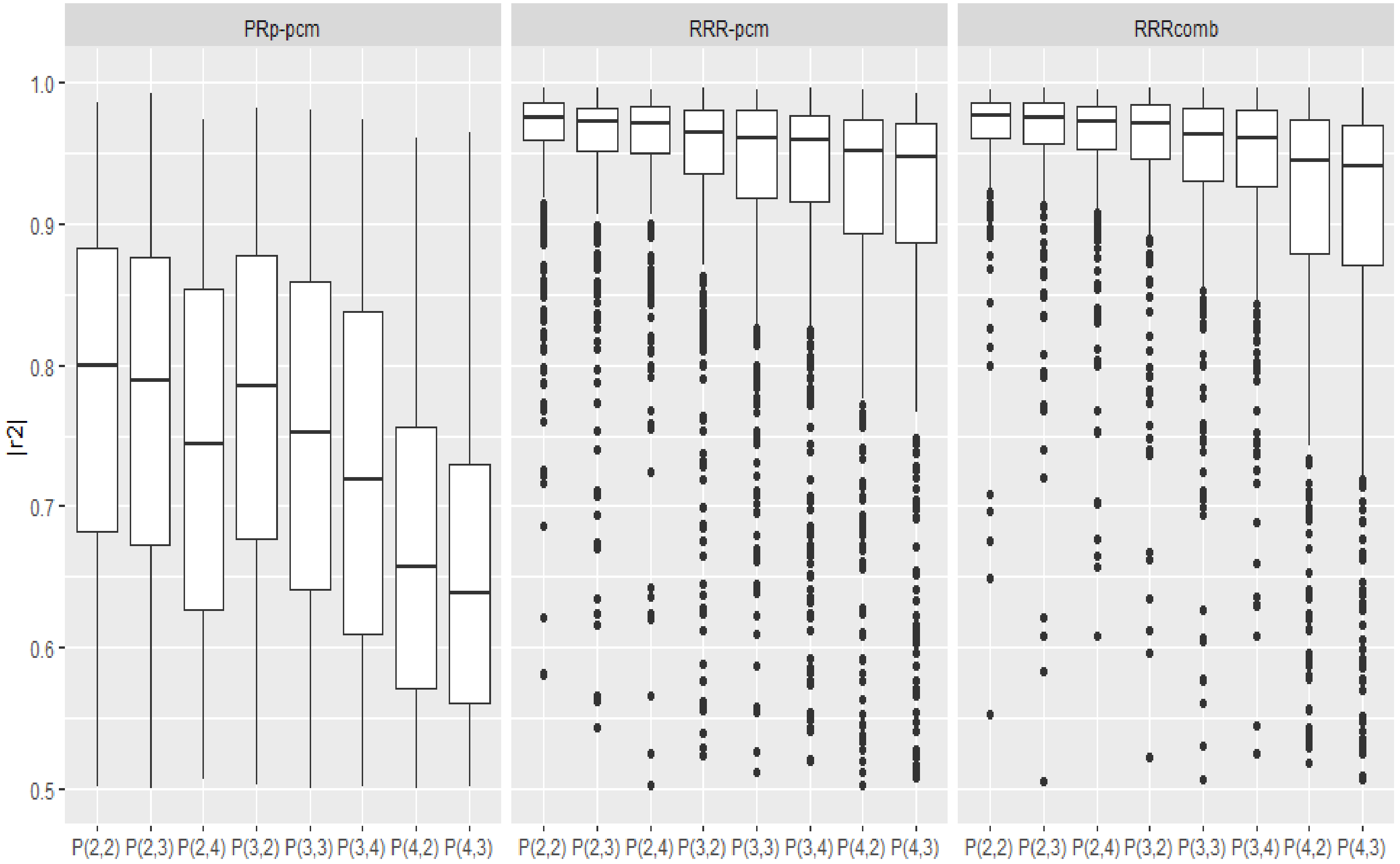

Figure 3.7: Model 3 of $|r_2|$s based on PCM

Table 3.6: Summary of $|\mathrm{r}_2|$ depends on three methods in Model 3

	PRp-pcm							
Summary	(2,2)	(2,3)	(2,4)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)
Min	0.3323	0.2875	0.2557	0.1863	0.2118	0.2252	0.2177	0.2138
Q1	0.6454	0.6350	0.6042	0.6291	0.5980	0.5593	0.4852	0.4892
Q2	**0.7791**	0.7701	0.7267	0.7594	0.7264	0.6889	0.5906	0.5779
Ave	0.7527	0.7411	0.7164	0.7359	0.7099	0.6800	0.5987	0.5893
Q3	0.8790	0.8715	0.8479	0.8703	0.8479	0.8145	0.7158	0.7039
Max	0.9877	0.9934	0.9758	0.9829	0.9815	0.9761	0.9623	0.9661
Sd	0.1531	0.1546	0.1500	0.1612	0.1583	0.1610	0.1523	0.1434

	RRR-pcm							
Summary	(2,2)	(2,3)	(2,4)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)
Min	0.1964	0.0237	0.1257	0.1223	0.0688	0.0761	0.0497	0.0297
Q1	0.9581	0.9490	0.9477	0.9289	0.8974	0.9003	0.8226	0.8139
Q2	**0.9752**	0.9726	0.9709	0.9628	0.9566	0.9550	0.9451	0.9342
Ave	0.9531	0.9386	0.9347	0.9145	0.8939	0.8904	0.8326	0.8290
Q3	0.9858	0.9818	0.9832	0.9808	0.9788	0.9766	0.9710	0.9691
Max	0.9978	0.9974	0.9972	0.9977	0.9962	0.9979	0.9972	0.9941
Sd	0.0797	0.1185	0.1245	0.1417	0.1631	0.1638	0.2286	0.2262

	RRRcomb							
Summary	(2,2)	(2,3)	(2,4)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)
Min	0.2329	0.3725	0.2440	0.0769	0.0038	0.1652	0.0851	0.0136
Q1	0.9611	0.9565	0.9528	0.9442	0.9135	0.9187	0.8595	0.8049
Q2	**0.9770**	0.9752	0.9732	0.9708	0.9598	0.9603	0.9410	0.9284
Ave	0.9644	0.9595	0.9565	0.9407	0.8826	0.9216	0.8754	0.8293
Q3	0.9862	0.9855	0.9827	0.9840	0.9811	0.9802	0.9734	0.9668
Max	0.9972	0.9980	0.9966	0.9978	0.9982	0.9976	0.9973	0.9974
Sd	0.0578	0.0646	0.0613	0.1096	0.2201	0.1250	0.1615	0.2204

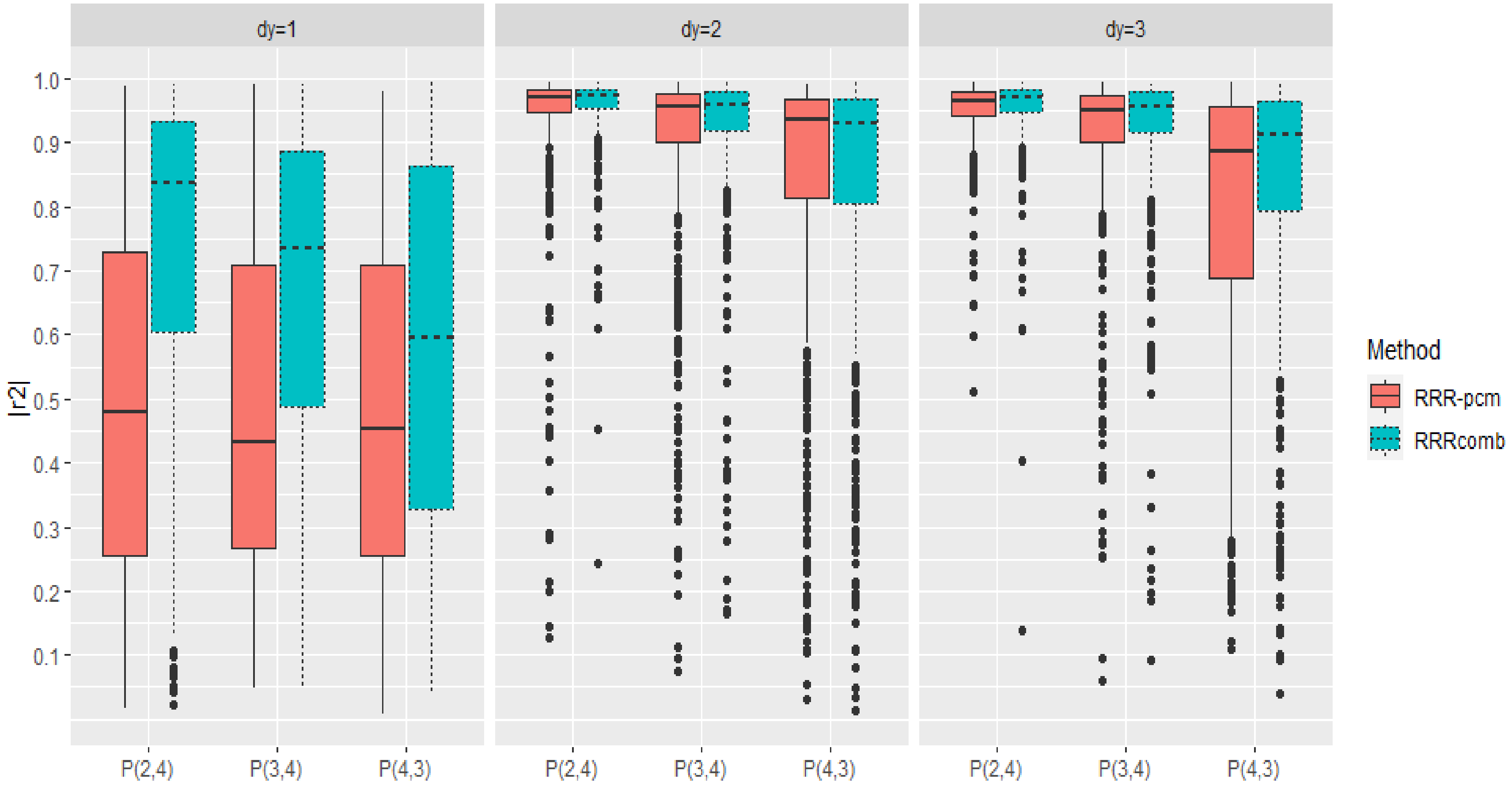

Figure 3.8: The variation of $|\mathrm{r}_2|$s with respect to d_y=1,2,3 in Model 3

3.4 Real data application

Here are the UCI Machine Learning Repository's Power consumption of Tetuan City Data Set . This dataset is the power consumption for the three zones of Tetuan City in northern Morocco. The data consist of measurements of the composition of Temperature, Humidity, Wind Speed, general diffuse flow, diffuse flow, and the power consumption for the three zones of Tetuan City. Predictors(X) are Temperature, Humidity, Wind Speed, general diffuse flow, diffuse flow, and Response variables(Y) are Zone 1 Power Consumption, Zone 2 Power Consumption, and Zone 3 Power Consumption. A total of 52,416 data were taken from the Supervisor Control and Data Acquisition System (SCADA) every 10 minutes from January 1, 2017 to December 31, 2017. The daily average for each variable was obtained and used for the analysis by consisting of 364 data.

We applied the conventional approach, PRp-pcm, along with the newly proposed methods, RRR-pcm and RRRcomb to the data with $(4, 2)$ combinations. Their first three sufficient predictors are reported in Figure 3.9, Figure 3.10, and Figure 3.11 respectively. According to the Figures, the first two sufficient predictors are essentially the same, while the strong relation does not hold for the third one. The scatterplot of the three responses and the dimension reduced predictors by RRR are reported in Figure 3.12. The first sufficient predictor is related to the mean functions with a negative tendency. On the other hand, the second sufficient predictor does not contribute to the mean function, which explains heteroscedasticity in Zone 2 and Zone 3.

The results obtained by applying the existing three PCM-based methods showed little difference from the Ko and Yoo (2022) results. However, the proposed RRR-pcm and RRRcomb demonstrated similarity despite being derived from a smaller number of samplings, indicating an improvement in efficiency.

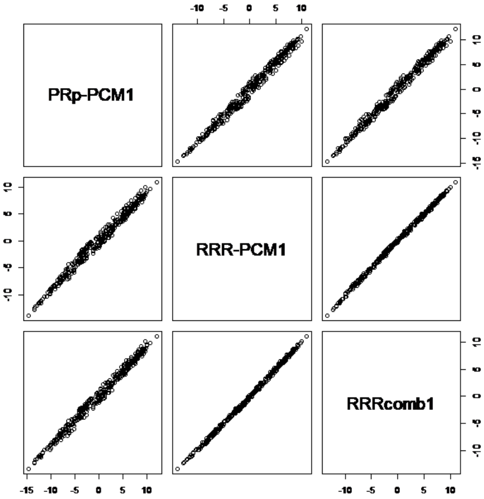

Figure 3.9: First sufficient predictor

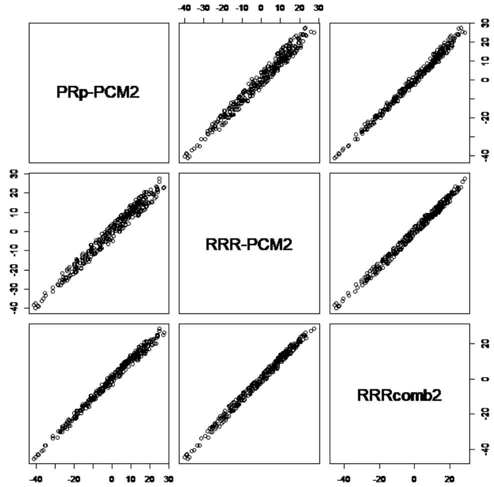

Figure 3.10: Second sufficient predictor

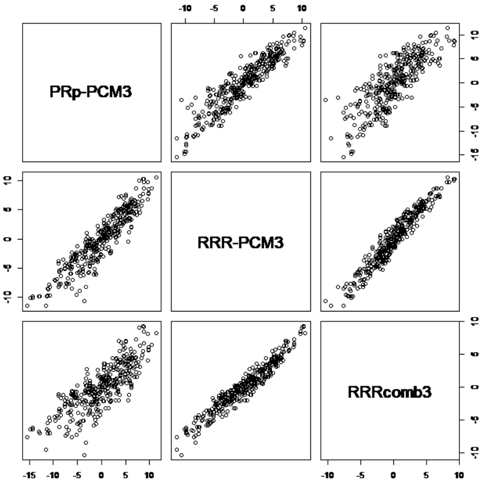

Figure 3.11: Third sufficient predictor

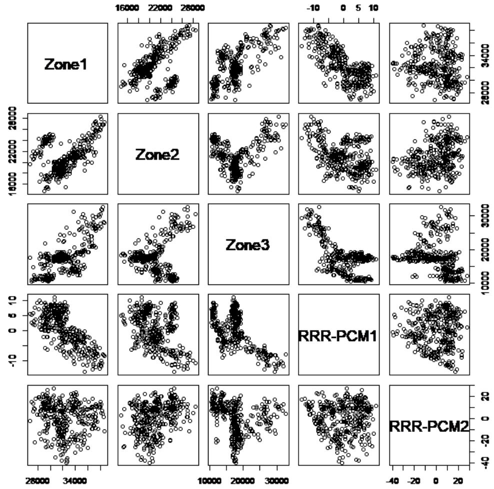

Figure 3.12: Summary plot

Chapter 4

Conclusion

In Chapter 2, seeking an alternative to sufficient dimension reduction (SDR) in previous research, we aimed to define the informative predictor subspace (IPS) and estimate the central subspace. This eliminates the prerequisite of employing the classic central subspace or central mean subspace estimation, replacing dimension reduction via K-means clustering on X's clusters. However, IPS had the limitation of analyzing y as a one-dimensional response variable. Recognizing the focus predominantly on dimension reduction regarding X, Yoo and Cook (2008) highlighted the need for attention toward dimension reduction of y when y is high-dimensional. They introduced the concept of Response Dimension Reduction, proposing methods so called YC that retain information regarding the conditional expectation of Y given X, thereby avoiding information loss. It has been expanded to include Principal Response Reduction (PRR), Principal Fitted Response Reduction (PFRR) proposed by Yoo (2018), and Unstructured PFRR (UPFRR) presented by Yoo (2019).

At this juncture, aiming to estimate the IPS space for cases involving multivariate Y, an attempt was made using the projective resampling method proposed by Li *et al.* (2008) to estimate Gamma. Surprisingly, this approach yielded a space, termed Projective Resampling IPS (PRIPS), that is smaller than IPS yet more accurate, encompassing the central subspace. PRIPS achieved higher accuracy by estimating a space that is more precise and inclusive of the central subspace compared to IPS. However, there are two disadvantages: it requires numerous attempts in the Projective Resampling mean method, and it does not consider the joint information from X and Y in the coordinate mean method. To address this, there is a third method that iteratively computes the coordinate mean, allowing for better reflection of how Xs are dependent on Y variations.

The space derived using response dimension reduction isn't uniquely defined, prompting an approach using Projective Resampling Inner Product Space (PRIPS) to obtain candidates. The new method called RRR-pcm includes using the average of kernel matrices that reduce the dimensions of Y to 1 or 2 using previously proposed techniques like PRR, PFRR, UPFRR, and YC. Additionally, leveraging the idea from the preceding paragraph, two proposed methods, RRR-pcm and RRRcomb, were compared with existing research. Results generally showed higher estimations around the cases of (2,4), (3,4), and (4,3) in the (X_{clust}, Y_{slice}) variations for the three models, prompting an exploration into whether this trend persists with d_y variations. In certain instances where $d_y = 1$, the estimation performance might not reach its peak due to a limited number of ψs. As a result, a default value of 2 for d_y is suggested. Finally, one last point to mention is the computing time. This study attempted to approach the construction of a central mean subspace by leveraging the non-uniqueness of response dimension reduction. However, compared to the conventional method, which relied on random sampling, our approach, while different, demonstrated significant performance due to a sampling time difference of over 100 minutes.

Bibliography

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics*, pages 327–351.

Anderson, T. W. (1999). Asymptotic distribution of the reduced rank regression estimator under general conditions. *The Annals of Statistics*, **27**(4), 1141–1154.

Cook, R. D. (1998). Principal hessian directions revisited. *Journal of the American Statistical Association*, **93**(441), 84–94.

Cook, R. D. (2007). Fisher lecture: Dimension reduction in regression.

Cook, R. D. (2009). *Regression graphics: Ideas for studying regressions through graphics*. John Wiley & Sons.

Cook, R. D. and Li, B. (2002). Dimension reduction for conditional mean in regression. *The Annals of Statistics*, **30**(2), 455–474.

Cook, R. D. and Ni, L. (2005). Sufficient dimension reduction via inverse regression: A minimum discrepancy approach. *Journal of the American Statistical Association*, **100**(470), 410–428.

Cook, R. D. and Setodji, C. M. (2003). A model-free test for reduced rank in multivariate regression. *Journal of the American Statistical Association*, **98**(462), 340–351.

Cook, R. D. and Weisberg, S. (1991). Discussion of sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, **86**(414), 328–332.

Cook, R. D. and Zhang, X. (2014). Fused estimators of the central subspace in sufficient dimension reduction. *Journal of the American Statistical Association*, **109**(506), 815–827.

Cook, R. D. and Zhang, X. (2015). Simultaneous envelopes for multivariate linear regression. *Technometrics*, **57**(1), 11–25.

Cook, R. D., Li, B., and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statistica Sinica*, pages 927–960.

Cook, R. D., Forzani, L., and Zhang, X. (2015). Envelopes and reduced-rank regression. *Biometrika*, **102**(2), 439–456.

Cox, D. R. and Mayo, D. G. (2010). Ii objectivity and conditionality in frequentist inference. *Error and inference: Recent exchanges on experimental reasoning, reliability, and the objectivity and rationality of science*, pages 276–304.

Dennis Cook, R. (2000). Save: a method for dimension reduction and graphics in regression. *Communications in statistics-Theory and methods*, **29**(9-10), 2109–2121.

Hall, P. and Li, K.-C. (1993). On almost linearity of low dimensional projections from high dimensional data. *The annals of Statistics*, pages 867–889.

Hilafu, H. and Yin, X. (2013). Sufficient dimension reduction in multivariate regressions with categorical predictors. *Computational Statistics & Data Analysis*, **63**, 139–147.

Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *Journal of multivariate analysis*, **5**(2), 248–264.

Ko, S. and Yoo, J. K. (2022). Projective resampling estimation of informative predictor subspace for multivariate regression. *Journal of the Korean Statistical Society*, **51**(4), 1117–1131.

Li, B. (2018). *Sufficient dimension reduction: Methods and applications with R*. CRC Press.

Li, B. and Dong, Y. (2009). Dimension reduction for nonelliptically distributed predictors.

Li, B., Wen, S., and Zhu, L. (2008). On a projective resampling method for dimension reduction with multivariate responses. *Journal of the American Statistical Association*, **103**(483), 1177–1186.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, **86**(414), 316–327.

Li, K.-C. (1992). On principal hessian directions for data visualization and dimension reduction: Another application of stein's lemma. *Journal of the American Statistical Association*, **87**(420), 1025–1039.

Li, K.-C., Aragon, Y., Shedden, K., and Thomas Agnan, C. (2003). Dimension reduction for multivariate response data. *Journal of the American Statistical Association*, **98**(461), 99–109.

Li, L., Cook, R. D., and Nachtsheim, C. J. (2004). Cluster-based estimation for sufficient dimension reduction. *Computational Statistics & Data Analysis*, **47**(1), 175–193.

Lu, W. and Li, L. (2011). Sufficient dimension reduction for censored regressions. *Biometrics*, **67**(2), 513–523.

Reinsel, G. C., Velu, R. P., and Chen, K. (2022). *Multivariate reduced-rank regression: theory, methods and applications*, volume 225. Springer Nature.

Stigler, S. M. (1973). Studies in the history of probability and statistics. xxxii: Laplace, fisher, and the discovery of the concept of sufficiency. *Biometrika*, **60**(3), 439–445.

Stoica, P. and Viberg, M. (1996). Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions. *IEEE Transactions on signal processing*, **44**(12), 3069–3078.

Weisberg, S. (2002). Dimension reduction regression in r. *Journal of Statistical Software*, **7**, 1–22.

Yoo, J. K. (2013). Chi-squared tests in k th-moment sufficient dimension reduction. *Journal of Statistical Computation and Simulation*, **83**(1), 191–201.

Yoo, J. K. (2016). Sufficient dimension reduction through informative predictor subspace. *Statistics*, **50**(5), 1086–1099.

Yoo, J. K. (2018). Response dimension reduction: model-based approach. *Statistics*, **52**(2), 409–425.

Yoo, J. K. (2019). Unstructured principal fitted response reduction in multivariate regression. *Journal of the Korean Statistical Society*, **48**(4), 561–567.

Yoo, J. K. and Cook, R. D. (2008). Response dimension reduction for the conditional mean in multivariate regression. *Computational statistics & data analysis*, **53**(2), 334–343.

Yoo, J. K. and Im, Y. (2014). Multivariate seeded dimension reduction. *Journal of the Korean Statistical Society*, **43**, 559–566.

Zhu, L., Miao, B., and Peng, H. (2006). On sliced inverse regression with high-dimensional covariates. *Journal of the American Statistical Association*, **101**(474), 630–643.

국 문 요 약

투영 재표본 정보적 설명변수 공간을 통한 축소된 평균추정

이 연구의 배경은 우선 충분 차원축소(sufficient dimension reduction, SDR)을 통해서 중심공간(central subspace) 또는 중심 평균 공간(central mean subspace)를 추정하는데 있어 전제 조건이 있다는 점에 있다. 이런 방법들은 소위 선형성(linearity) 조건, 등분산성(constant variance) 조건 그리고 범위(coverage) 조건이 만족되어야 한다. 반응 변수가 1차원인 일변량 회귀에서 이런 조건적 한계를 극복하고자 Yoo(2016)에서 정보적 설명 변수 공간(informative predictor subspace, IPS)를 정의하여 중심 부분공간을 추정하고자 하였다. X를 k평균 군집화 방법을 이 용하여 자료를 범주화 한 후, 이 군집에 해당되는 자료에 대한 반응 변수 Y를 다시 범주화 한 후 최종범주에서 X의 평균을 추정하여 중심 부분 공간을 추정하는 방법이 다. 이는 모두 X의 차원축소에 초점이 맞춰졌기에 다변량 Y일 경우 회귀분석에 대해 반응변수의 차원축소에도 관심을 가져야 한다는 초기 연구가 기반이 되어 본 연구가 뿌리내려졌다 할 수 있다. Yoo and Cook에서 제시한 방법을 간단히 YC라 부를 것 이고, 이밖에 반응변수 차원축소 방법으로 principal response reduction(PRR), principal fitted response reduction (PFRR), 그리고 준모수적 제약으로도 더 벗어 난 unstructured PFRR(UPFRR) 방법을 소개한바 있다. 그러나 다차원 반응변수 Y 인 상황에서 정보적 설명 변수 공간을 추정을 한다면 X를 군집화 한 후, Y를 범주화 할 때 고차원의 저주를 피하지 못할 것이다. 이에 Li et al.(2008)에서 나온 투영 재 표본(projective resampling) 방법을 차용한다. 반응 변수의 차원이 상대적으로 높다 면 T의 재표본 수를 증가시켜 정확도를 유지할 수 있는 가능성에 그 이유가 있다.

투영 재표본방법을 써보니 정보적 설명 변수 공간보다는 작은 공간이면서 중심 부분공간을 포함하는 정확도가 높은 공간을 추정하게 되었는데 이 공간을 투영-재표본 설명 변수 공간(Projective resampling IPS)이라고 하였다. 정보적 설명 변수 공간을 찾기 위한 방법 3가지 방법, 투영-재표본 평균 방법(projective resampling mean method), 반응변수 요소별 평균 방법(coordinate mean method), 그리고 요소별 투영-재표본 평균 (coordinate projective resampling) 방법을 제안했었다.

본 연구에서는 다차원 반응변수 Y일 경우 반응변수의 차원축소를 실시하되 투영재표본 설명변수 공간을 통해 중심 부분공간을 추정하고자 하였다. 다만 여기서 언급할 점은 반응변수 차원축소를 통한 공간은 중심 부분공간처럼 유일하게 정의가 되지 않기에 되도록이면 phi(T)들의 평균값으로 이뤄진 공간에서 추출된 값을 채택하도록 하였다. 즉 반응 변수 차원 축소 연구로 제안되었던 앞 4가지 방법들(YC, PRR,PFRR,UPFRR)의 평균값을 취하는 축소된 차원의 반응변수의 부분 조건적 평균(Reduced-rank response partial conditional mean, RRR-pcm) 방법을 제안하였다. 추가적으로 RRR-pcm에 요소별 평균 방법을 추가하여 평균값을 취한 RRRcomb 방법도 제안하였다. 연구 결과 기존 연구와 같이 임의의 많은 수의 랜덤 샘플링으로 반응변수 차원 축소를 했을 때에 비해 차원 축소를 한뒤, RRR-pcm 또는 RRRcomb를 했을 때에는 적은 수의 샘플링에도 추정치의 근사도가 더 높거나 유지됨을 알 수 있었다.

핵심되는 말: 충분 차원 축소(SDR), 반응변수 차원 축소(Response dimension reduction), 정보적 설명 변수 공간(IPS), 투영-재표본 정보적 설명 변수 공간(PRIPS), 축소된 반응변수 차원의 회귀(Reduced-rank response regression)