Journal of the Korean Statistical Society (2025) 54:920–931
https://doi.org/10.1007/s42952-024-00304-0

Online ISSN 2005-2863
Print ISSN 1226-3192

SHORT COMMUNICATION

Covariance projective resampling informative predictor subspace for multivariate regression

Minjee Kim[1] · Jae Keun Yoo[1]

Received: 1 October 2024 / Accepted: 23 December 2024 / Published online: 20 January 2025

Abstract

In this paper, we newly define a covariance projective resampling informative predictor subspace (covPRIPS) besides an existing projective resampling informative predictor subspace (Ko and Yoo in J Korean Stat Soc 51:1117–1131, 2022; PRIPS). To clarify the relation between covPRIPS and the central subspace, two mild conditions are assumed to hold. Under the conditions, covPRIPS becomes the smallest subspace to contain the central subspace up to date while being nested in PRIPS. Two possible benefits of covPRIPS over PRIPS are no necessity of slicing $t^{\mathrm{T}}\mathbf{Y}$ and intuitive interpretation of the role of $t^{\mathrm{T}}\mathbf{Y}$. Numerical studies show that the estimation method of covPRIPS is competitive with the inverse mean method of PRIPS.

Keywords Covariance method · Informative predictor subspace · Projective resampling · Sufficient dimension reduction

Mathematics Subject Classification 62G08 · 62H05

1 Introduction

Sufficient dimension reduction (SDR) for a multivariate regression of $\mathbf{Y} \in \mathbb{R}^r|\mathbf{X} \in \mathbb{R}^p$ with $r \geq 2$ pursues a lower-dimensional linear transformation $\mathbf{A}^{\mathrm{T}}\mathbf{X}$ without loss of information on $\mathbf{Y}|\mathbf{X}$, where $\mathbf{A} \in \mathbb{R}^{p\times q}$ and $q \leq p$. This can be equivalently stated as:

$$\mathbf{Y} \perp\!\!\!\perp \mathbf{X}|\mathbf{A}^{\mathrm{T}}\mathbf{X}, \tag{1}$$

where

✉ Jae Keun Yoo
peter.yoo@ewha.ac.kr

Minjee Kim
jee365@ewhain.net

[1] Department of Statistics, Ewha Womans University, Seoul 03760, Republic of Korea

$\perp\!\!\!\perp$

indicates statistical independence.

Then, the lower-dimensional predictors $\mathbf{A}^{\mathrm{T}}\mathbf{X}$ in (1) are called *sufficient predictors*. Within the framework of SDR, the primary objective is to estimate the minimal subspace among all possible subspaces spanned by the columns of $\mathbf{A}$ to satisfy (1), and it is called the *central subspace* $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$. Throughout the rest of the paper, a true basis matrix of $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ will be denoted as $\boldsymbol{\eta} \in \mathbb{R}^{p\times d}$.

Over two decades, various methodologies to estimate $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ have been developed. However, these methods often rely on the assumption of linearity and constant variance, which are essential but challenging to verify in practice. According to Li et al. (2004) and Li and Dong (2009), these conditions are difficult to assess through data in practice. Even more, the violation of these conditions can mislead the dimension reduction results.

To overcome this deficit, an informative predictor subspace (IPS) to contain $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ is proposed by Yoo (2016). According to Yoo (2022), however, the construction of the informative predictor subspace in the multi-dimensional responses faces the curse of dimensionality due to hierarchical categorization of both $\mathbf{X}$ and $\mathbf{Y}$.

To relieve this issue, Ko and Yoo (2022) newly define a projective resampling informative predictor subspace, which is, theoretically, to contain $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ but to be contained in the informative predictor subspace. Its estimation is to utilize one dimensional linear transformation $t^{\mathrm{T}}\mathbf{Y}$ instead of the r-dimensional $\mathbf{Y}$, so the curse of dimensionality is relieved, where t is a r-dimensional random vector. More detail of the projective resampling method will be given in the following section. According to Ko and Yoo (2022), their estimation methods show high accuracy to restore $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$.

In Ko and Yoo (2022), the estimation methods construct the inverse means of $E(\mathbf{X}|\mathbf{Y})$. However, it is not clear how the random vector t affects the estimation procedure, and the estimation accuracies quite depend on two tuning parameters of the number of categories of $\mathbf{X}$ and $\mathbf{Y}$. To investigate these issues, we adopt the covariance between $\mathbf{X}$ and $\mathbf{Y}$. By this, the function of the random vector t is clearly revealed, and the tuning parameter is simply reduced to the number of categories of $\mathbf{X}$.

The paper is organized as follows. In Sect. 2, the projective resampling informative predictor subspace and its estimation method are introduced. Section 3 is devoted to newly defining a covariance projective resampling subspace and to investigating the containment relation among the central subspace, the informative predictor subspace and the projective resampling informative predictor subspace. In the same section, the sample estimation method is proposed. Numerical studies and a real data example are presented in Sect. 3.1. Section 4 summarizes our work.

2 Review: projective resampling informative predictor subspace

For simplicity, it is assumed that $E(\mathbf{X}) = 0$ without loss of generality. In multivariate regression $\mathbf{Y} \in \mathbb{R}^r|\mathbf{X} \in \mathbb{R}^p$ with $r \geq 2$, the definition of an informative predictor subspace $\mathcal{S}^{\mathrm{IPS}}_{\mathbf{Y}|\mathbf{X}}$ (Yoo 2016; IPS) for $\boldsymbol{\eta}$ is as follows:

$$\mathcal{S}^{\mathrm{IPS}}_{\mathbf{Y}|\mathbf{X}} = \boldsymbol{\Sigma}^{-1}\mathcal{S}\{E(\mathbf{X}|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X})\},$$

where $\boldsymbol{\Sigma} = \mathrm{cov}(\mathbf{X})$ and $\mathcal{S}\{E(\mathbf{X}|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X})\}$ is a subspace spanned by $E(\mathbf{X}|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X})$ varying the values $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$. By its definition, $\mathcal{S}^{\mathrm{IPS}}_{\mathbf{Y}|\mathbf{X}}$ naturally contains $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$, and hence $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ can be restored through $\mathcal{S}^{\mathrm{IPS}}_{\mathbf{Y}|\mathbf{X}}$. Moreover, the estimation methods of $\mathcal{S}^{\mathrm{IPS}}_{\mathbf{Y}|\mathbf{X}}$ are not restricted by the conditions required in most SDR methods. This is the main usefulness of $\mathcal{S}^{\mathrm{IPS}}_{\mathbf{Y}|\mathbf{X}}$.

Recently, Ko and Yoo (2022) propose the so-called projective resampling informative predictor subspace, which is contained in $\mathcal{S}^{\mathrm{IPS}}_{\mathbf{Y}|\mathbf{X}}$ but still contains $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$. For a random vector $t \in \mathbb{R}^r$ with unit length, the central subspace $\mathcal{S}_{t^{\mathrm{T}}\mathbf{Y}|\mathbf{X}}$ of $t^{\mathrm{T}}\mathbf{Y}|\mathbf{X}$ is supposed to be spanned by a positive semi-definite $\mathbf{M}(t) \in \mathbb{R}^{p\times p}$. Also, suppose that a positive semi-definite matrix $\boldsymbol{\phi}(t) \in \mathbb{R}^{p\times p}$ spans an informative predictor subspace $\mathcal{S}^{\mathrm{IPS}}_{t^{\mathrm{T}}\mathbf{Y}|\mathbf{X}}$ for $\mathbf{M}(t)$. Then, Ko and Yoo (2022) show the following relation:

$$\mathcal{S}(\mathbf{M}(t)) = \mathcal{S}_{t^{\mathrm{T}}\mathbf{Y}|\mathbf{X}} \subseteq \mathcal{S}(E(\mathbf{M}(\mathbf{T}))) = \mathcal{S}_{\mathbf{Y}|\mathbf{X}} \subseteq \mathcal{S}(E(\boldsymbol{\phi}(\mathbf{T}))) \subseteq \mathcal{S}^{\mathrm{IPS}}_{\mathbf{Y}|\mathbf{X}},$$

where $E(\mathbf{M}(\mathbf{T}))$ and $E(\boldsymbol{\phi}(\mathbf{T}))$ are mean matrices of $\mathbf{M}(t)$ and $\boldsymbol{\phi}(t)$, respectively, and the notations emphasize that $E(\mathbf{M}(\mathbf{T}))$ and $E(\boldsymbol{\phi}(\mathbf{T}))$ do not depend on t.

The space $\mathcal{S}(E(\boldsymbol{\phi}(\mathbf{T})))$ is the definition of the *projective resampling informative predictor subspace* (PRIPS), whose population structure derived in Ko and Yoo (2022) is as follows:

$$\mathcal{S}(E(\boldsymbol{\phi}(\mathbf{T}))) = \boldsymbol{\Sigma}^{-1}\mathcal{S}\big(E_{\mathbf{T}}[\,\mathrm{cov}(E(E(\mathbf{X}|\mathbf{T}^{\mathrm{T}}\mathbf{Y}, C_{\mathbf{X}} = c)|C_{\mathbf{X}} = c))]\big),$$

where $C_{\mathbf{X}}$ stands for a cluster indicator of $\mathbf{X}$.

Let $\hat{\boldsymbol{\eta}}_{\mathrm{P}}^{\mathrm{T}}\mathbf{X}$ be a set of the first sufficient predictors computed from the ordinary least squares (Cook 1998) and the principal Hessian directions (Li 1992). Clustering of $\hat{\boldsymbol{\eta}}_{\mathrm{P}}^{\mathrm{T}}\mathbf{X}$ is called a partially informative clustering. In Yoo (2016) and Ko and Yoo (2022), $C_{\mathbf{X}}$ is constructed from the application of K-means algorithm to $\mathbf{X}$ or $\hat{\boldsymbol{\eta}}_{\mathrm{P}}^{\mathrm{T}}\mathbf{X}$.

The key to estimating $\mathcal{S}(E(\boldsymbol{\phi}(\mathbf{T})))$ is placed onto $E(\mathbf{X}|\mathbf{T}^{\mathrm{T}}\mathbf{Y}, C_{\mathbf{X}} = c)$. Then the random vector $\mathbf{T}$ is generated from a multivariate normal distribution, and then is transformed to have unit length. Next, the data is partitioned by the clusters of $C_{\mathbf{X}}$, and $\mathbf{T}^{\mathrm{T}}\mathbf{Y}$ is categorized, called slicing, within each partition. Finally, the quantity $E(\mathbf{X}|\mathbf{T}^{\mathrm{T}}\mathbf{Y}, C_{\mathbf{X}} = c)$ is estimated by the sample mean of $\mathbf{X}$ within the final slices.

Following this estimation scheme, Ko and Yoo (2022) developed three methods: projective resampling mean method, coordinate mean method, and coordinate projective resampling mean method. Among these, the coordinate projective resampling mean method along with partially informative clustering is recommended.

3 Covariance estimation of projective resampling informative predictor subspace

For further use, we start the section to define the following notations:

- the subscript c stands for the cluster indicator of $\mathbf{X}$ varying $c = 1, \ldots, k$.
- the subscript j represents the coordinate of $\mathbf{Y} = (Y_1, \ldots, Y_r)$ for $j = 1, \ldots, r$.
- $\mathbf{X}_c$ and $\mathbf{Y}_c$ represent subsets of $\mathbf{X}$ and $\mathbf{Y}$ corresponding to the cth cluster of $C_{\mathbf{X}} = c$, so that $E(\mathbf{X}|\mathbf{Y}, C_{\mathbf{X}} = c) = E(\mathbf{X}_c|\mathbf{Y}_c)$ and $\text{cov}\,(\mathbf{X}, \mathbf{Y}^{\mathrm{T}}|C_{\mathbf{X}} = c) = \text{cov}\,(\mathbf{X}_c, \mathbf{Y}_c^{\mathrm{T}})$.
- $\varphi_{cj} = \text{cov}\,(\mathbf{X}_c, Y_{cj}) \in \mathbb{R}^{p\times 1}$, which is the covariance between the jth coordinate Y_j of $\mathbf{Y}$ and $\mathbf{X}$ within the cth cluster.
- $\boldsymbol{\varphi}_{c\bullet} = \text{cov}\,(\mathbf{X}_c, \mathbf{Y}_c^{\mathrm{T}}) = (\varphi_{c1}, \ldots, \varphi_{cr}) \in \mathbb{R}^{p\times r}$, which is the covariance between $\mathbf{Y}$ and $\mathbf{X}$ within the cth cluster of $\mathbf{X}$ for $c = 1, \ldots, k$.

3.1 Informative predictor subspace and related containment

In many works, including Li et al. (2004), Yin and Bura (2006), Yoo and Cook (2007), Yoo (2008) and so on, the covariance between $\mathbf{X}$ and $\mathbf{Y}$ is utilized in estimating $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$.

To incorporate the covariance into the context of the informative predictor subspace, we consider the following lemma.

Lemma 3.1 *The following relation holds*: $\mathcal{S}\{E(\mathbf{X}_c - E(\mathbf{X}_c)|Y_{c_j})\} \subseteq \mathcal{S}\{E(\mathbf{X}_c|Y_{c_j})\}$.

Proof Since $E(\mathbf{X}_c) = E(E(\mathbf{X}_c|Y_{c_j}))$, it follows that $E(\mathbf{X}_c) \in \mathcal{S}\{E(\mathbf{X}_c|Y_{c_j})\}$. Also, it is true that $E(\mathbf{X}_c|Y_{c_j}) \in \mathcal{S}\{E(\mathbf{X}_c|Y_{c_j})\}$. This directly implies that $E(\mathbf{X}_c|Y_{c_j}) - E(\mathbf{X}_c) = E(\mathbf{X}_c - E(\mathbf{X}_c)|Y_{c_j}) \in \mathcal{S}\{E(\mathbf{X}_c|Y_{c_j})\}$. This completes the proof.□

Lemma 3.1 and Proposition 3.1 in Yoo (2016) lead to the following relations:

$$\boldsymbol{\varphi}_{cj} \in \mathcal{S}\{E(\mathbf{X}_c - E(\mathbf{X}_c)|Y_{cj})\} \subseteq \mathcal{S}\{E(\mathbf{X}_c|Y_{cj})\}. \tag{2}$$

The relation in (2) directly indicates the following relation:

$$\boldsymbol{\Sigma}^{-1}E(\boldsymbol{\varphi}_{c\bullet}) \in \boldsymbol{\Sigma}^{-1}\mathcal{S}\{E(E(\mathbf{X}_c|\mathbf{Y}_c))\}. \tag{3}$$

This confirms that $[\boldsymbol{\varphi}_{c\bullet}]_{c=1}^{k}$ can provide useful information for the informative predictor subspace. The relation in (3), however, does not guarantee that $[\boldsymbol{\varphi}_{c\bullet}]_{c=1}^{k}$ properly contains $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$. To address this issue, the following two conditions are considered:

C1. $\mathcal{S}\{E(\mathbf{X}_c|\mathbf{Y}_c)\} = \mathcal{S}\big(\boldsymbol{\varphi}_{c\bullet}\big)$ for $c = 1, \ldots, k$;
C2. $\mathcal{S}\{E(\mathbf{X}_c|\mathbf{Y}_c)\} \subseteq \mathcal{S}\big([\boldsymbol{\varphi}_{c\bullet}]_{c=1}^{k}\big)$ for $c = 1, \ldots, k$.

Proposition 3.2 *If one of Conditions* 1 *and* 2 *holds, it is guaranteed that* $\mathcal{S}_{\mathbf{Y}|\mathbf{X}} \subseteq \boldsymbol{\Sigma}^{-1}\mathcal{S}[\{\boldsymbol{\varphi}_{c\bullet}\}_{c=1}^{k}]$.

Proof It should be noted that $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}^{\mathrm{IPS}}$ is spanned by $\boldsymbol{\Sigma}^{-1}E(\mathbf{X}_c - E(\mathbf{X})|\mathbf{Y}_c)$ with varying c. Conditions 1 and 2 directly show that $E(\mathbf{X}_c - E(\mathbf{X})|\mathbf{Y}_c) \in \mathcal{S}\{[\boldsymbol{\varphi}_{c\bullet}]_{c=1}^{k}\}$ for $c = 1, \dots, k$. This directly implies that $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}^{\mathrm{IPS}} = \boldsymbol{\Sigma}^{-1}\mathcal{S}\{[\boldsymbol{\varphi}_{c\bullet}]_{c=1}^{k}\}$, and hence it contains $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$.□

Condition 1 can hold for a location regression such that

$$\mathbf{Y} \perp\!\!\!\perp \mathbf{X}|E(\mathbf{Y}|\mathbf{X})$$

, which indicates that the regression mean function has all information on the regression. According to Cook and Li (2002), Condition 1 is reasonable because, in many regression problems, the primary interest is in the mean function, and location regression encompasses a broad class of models, including single-index and multiple index models. Yoo and Cook (2007) discuss how Condition 2 is likely to hold for higher-dimensional $\mathbf{Y}$, as each coordinate regression captures more information about $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$. Additionally, Li et al. (2004) provide a thorough explanation of why Condition 2 would be highly plausible. Therefore, Conditions 1 and 2 should hold in many regression problems in practice, making them unlikely to pose an issue. If Condition 1 holds, then Condition 2 is satisfied.

3.2 Covariance projective resampling informative predictor subspace

Here, we develop the projective resampling approach for the covariance φ_{cj}. Without loss of generality, t is sampled from a random vector $\mathbf{T}$ with $E(\mathbf{T}) = 0$. Then, for $t^{\mathrm{T}}\mathbf{Y}|\mathbf{X}$, the population covariance within each cluster is as follows: cov $(\mathbf{X}_c, \mathbf{Y}_c^{\mathrm{T}})t = \boldsymbol{\varphi}_{c\bullet}t$.

Then, the collection of $\boldsymbol{\varphi}_{c\bullet}t$ is expressed as follows:

$$_{c=1}^{k} = [\boldsymbol{\varphi}_{1\bullet}, \dots, \boldsymbol{\varphi}_{k\bullet}]\mathbf{D}(t, \dots, t),$$

where, for a $r \times q$ matrix $\mathbf{A}_i$, $i = 1, \dots, d$, an operator $\mathbf{D}(\mathbf{A}_1, \dots, \mathbf{A}_d)$ denotes a $rd \times qd$ block diagonal matrix whose diagonal matrices are $\mathbf{A}_i$ and all off-diagonal matrices are zeros.

Then, we define a positive semi-definite matrix required in the covariance projective resampling approach:

$$\begin{aligned}\mathbf{M}_{\mathbf{C}}(t) =& [\boldsymbol{\varphi}_{1\bullet}, \dots, \boldsymbol{\varphi}_{k\bullet}]\mathbf{D}(t, \dots, t)\mathbf{D}(t^{\mathrm{T}}, \dots, t^{\mathrm{T}})[\boldsymbol{\varphi}_{1\bullet}, \dots, \boldsymbol{\varphi}_{k\bullet}]^{\mathrm{T}} \\ =& [\boldsymbol{\varphi}_{1\bullet}, \dots, \boldsymbol{\varphi}_{k\bullet}]\mathbf{D}(tt^{\mathrm{T}}, \dots, tt^{\mathrm{T}})[\boldsymbol{\varphi}_{1\bullet}, \dots, \boldsymbol{\varphi}_{k\bullet}]^{\mathrm{T}} \\ =& \sum_{c=1}^{k} \boldsymbol{\varphi}_{c\bullet}tt^{\mathrm{T}}\boldsymbol{\varphi}_{c\bullet}^{\mathrm{T}}\end{aligned} \tag{4}$$

In this case, we can closely look into $E(\mathbf{M}_{\mathbf{C}}(\mathbf{T}))$ from (4):

$$E(\mathbf{M_C}(\mathbf{T})) = E(\sum_{c=1}^{k} \boldsymbol{\varphi}_{c\bullet}\mathbf{T}\mathbf{T}^{\mathrm{T}}\boldsymbol{\varphi}_{c\bullet}^{\mathrm{T}}) = \sum_{c=1}^{k} \boldsymbol{\varphi}_{c\bullet}E(\mathbf{T}\mathbf{T}^{\mathrm{T}})\boldsymbol{\varphi}_{c\bullet}^{\mathrm{T}} = \sum_{c=1}^{k} \boldsymbol{\varphi}_{c\bullet}\operatorname{cov}(\mathbf{T})\boldsymbol{\varphi}_{c\bullet}^{\mathrm{T}}. \quad (5)$$

According to (5), the matrix $E(\mathbf{M_C}(\mathbf{T}))$ depends on $\mathbf{T}$ only through $\operatorname{cov}(\mathbf{T})$. This indicates that the asymptotic estimation behavior is primarily influenced by the choice of $\operatorname{cov}(\mathbf{T})$, rather than the distribution of $\mathbf{T}$. For this, the asymptotic distribution of the ordinary least squares estimator $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Sigma}}^{-1}\widehat{\operatorname{cov}}(\mathbf{X}, \mathbf{Y}^{\mathrm{T}})$, as derived in Yoo and Cook (2007), is noted. The matrix $\hat{\boldsymbol{\beta}}$ is used for the dimension reduction of $\mathbf{X}$ without loss of information on $E(\mathbf{Y}|\mathbf{X})$, which is a part of $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$. Its asymptotic distribution is as follows:

$$\sqrt{n}(\operatorname{vec}(\hat{\boldsymbol{\beta}}) - \operatorname{vec}(\boldsymbol{\beta})) \sim MN(0, E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}} \otimes \boldsymbol{\Sigma}^{-1}(\mathbf{X} - E(\mathbf{X}))(\mathbf{X} - E(\mathbf{X}))^{\mathrm{T}}\boldsymbol{\Sigma}^{-1})), \quad (6)$$

where $\boldsymbol{\varepsilon} = \mathbf{Y} - E(\mathbf{Y}) - \boldsymbol{\beta}^{\mathrm{T}}(\mathbf{X} - E(\mathbf{X}))$ is the ordinary least squares residual vector. The part of $\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}$ in (6) accounts for the relationship among the responses. Therefore, the inverse of $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}})$ would be a reasonable choice for $\operatorname{cov}(\mathbf{T})$, because the ordinary least squares estimator is nothing but the covariance.

If $\operatorname{cov}(\mathbf{T})$ is equal to $\mathbf{I}_r$ just like Ko and Yoo (2022), we have that $E_{\mathbf{C}}(\mathbf{M}(\mathbf{T})) = \sum_{c=1}^{k} \boldsymbol{\varphi}_{c\bullet}\boldsymbol{\varphi}_{c\bullet}^{\mathrm{T}}$, and it is a version of the covariance approach for the coordinate mean method of Ko and Yoo (2022).

Since we have that $\mathcal{S}\{\sum_{c=1}^{k} \boldsymbol{\varphi}_{c\bullet}\operatorname{cov}(\mathbf{T})\boldsymbol{\varphi}_{c\bullet}^{\mathrm{T}}\} \subseteq \mathcal{S}\{E(\boldsymbol{\phi}(\mathbf{T}))\}$ and Conditions 1–2 guarantee that

$$\mathcal{S}_{\mathbf{Y}|\mathbf{X}} \subseteq \boldsymbol{\Sigma}^{-1}\mathcal{S}\{\sum_{c=1}^{k} \boldsymbol{\varphi}_{c\bullet}\operatorname{cov}(\mathbf{T})\boldsymbol{\varphi}_{c\bullet}^{\mathrm{T}}\} \subseteq \boldsymbol{\Sigma}^{-1}\mathcal{S}\{E(\boldsymbol{\phi}(\mathbf{T}))\} \subseteq \mathcal{S}_{\mathbf{Y}|\mathbf{X}}^{\mathrm{IPS}},$$

the space of $\boldsymbol{\Sigma}^{-1}\mathcal{S}\{\sum_{c=1}^{k} \boldsymbol{\varphi}_{c\bullet}\operatorname{cov}(\mathbf{T})\boldsymbol{\varphi}_{c\bullet}^{\mathrm{T}}\}$ becomes a new upper bound of $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$.

Hereafter, the subspace of $\boldsymbol{\Sigma}^{-1}\mathcal{S}\{\sum_{c=1}^{k} \boldsymbol{\varphi}_{c\bullet}\operatorname{cov}(\mathbf{T})\boldsymbol{\varphi}_{c\bullet}^{\mathrm{T}}\}$ will be called *covariance projective resampling informative predictor subspace* (covPRIPS) to focus its key quantity $\operatorname{cov}(\mathbf{X}_c, \mathbf{Y}_c^{\mathrm{T}})$. So, the covPRIPS is proposed as the primary subspace to estimate for restoring $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$.

Two potential advantages of using covPRIPS are as follows. First, it eliminates a crucial tuning parameter in the PRIPS, namely the number of slices of $t^{\mathrm{T}}\mathbf{Y}$. Second, covPRIPS offers an intuitive understanding of how the linear transformation of $t^{\mathrm{T}}\mathbf{Y}$ influences $E(\mathbf{M}(\mathbf{T}))$, which ultimately depends on $\operatorname{cov}(\mathbf{T})$.

3.3 Estimation in sample

By Proposition 3.3 in Ko and Yoo (2022), the standardized predictors $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - E(\mathbf{X}))$ are used in the estimation, and it is back-transformed to the $\mathbf{X}$ -scale by pre-multiplying $\boldsymbol{\Sigma}^{-1/2}$. Following the guidance in Yoo (2016) and Ko and Yoo (2022), the partially informative clustering of $\mathbf{X}$ is used, instead of clustering $\mathbf{X}$.

Denoting $\hat{\mathbf{Z}}_i = \hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{X}_i - \bar{\mathbf{X}})$ and $\mathbf{Y}_i$ as the ith observations of $\mathbf{Z}$ and $\mathbf{Y}$ in the sample, $\hat{\boldsymbol{\varphi}}^z_{c\bullet}$ is as follows:

$$\hat{\boldsymbol{\varphi}}^z_{c\bullet} = \frac{1}{n_c} \sum_{\{i|C_{x,i}=c\}} \hat{\mathbf{Z}}_{c,i}\mathbf{Y}^{\mathrm{T}}_{c,i} - \bar{\mathbf{Z}}_c\bar{\mathbf{Y}}^{\mathrm{T}}_c.$$

Since t is sampled m_n times with unit length, $E(\boldsymbol{\varphi}^z(\mathbf{T}))$ is estimated by its usual moment estimator:

$$\hat{E}(\boldsymbol{\varphi}^z(\mathbf{T})) = \frac{1}{m_n}\sum_{i=1}^{m_n}\left(\frac{1}{k}\sum_{c=1}^{k}\hat{\boldsymbol{\varphi}}^z_{c\bullet}t_i t_i^{\mathrm{T}}\hat{\boldsymbol{\varphi}}^{z\mathrm{T}}_{c\bullet}\right) = \frac{1}{k}\sum_{c=1}^{k}\hat{\boldsymbol{\varphi}}^z_{c\bullet}\left(\sum_{i=1}^{m_n}\frac{1}{m_n}t_i t_i^{\mathrm{T}}\right)\hat{\boldsymbol{\varphi}}^{z\mathrm{T}}_{c\bullet}. \quad (7)$$

The sample estimation method in (7) will be called *projective resampling covariance method* (PRcovM). Once $\hat{E}(\boldsymbol{\varphi}^z(\mathbf{T}))$ is constructed, it is spectral-decomposed and its first d-largest eigenvectors are the estimate of the basis of $\mathcal{S}\{E(\boldsymbol{\varphi}^z(\mathbf{T}))\}$. Then, the d-eigenvectors pre-multiplied by $\hat{\boldsymbol{\Sigma}}^{-1/2}$ are the basis estimate of $\mathcal{S}\{E(\boldsymbol{\varphi}^z(\mathbf{T}))\}$.

4 Numerical studies and real data example

4.1 Numerical studies

We mimicked the numerical studies in Ko and Yoo (2022). The variables U_1, e, W_1, W_2 and W_3 were independently generated: $U_1 \sim U(0,1)$; $e \sim U(-0.5, 0.5)$; $U_2 = \log(U_1) + e$; $(W_1, W_2, W_3)\overset{iid}{\sim}N(0,1)$, where $U(a, b)$ represents a uniform distribution between a and b. With these variables, the following 10-dimensional predictors $(X_1, \ldots, X_{10})^{\mathrm{T}}$ were generated: $X_1 = U_1 + U_2$; $X_2 = U_2 + W_2 + W_3$; $X_3 = W_1 - W_2$; $X_4 = W_2$; $X_5 = W_3$.

Then, Models 1 and 2 were constructed with 10-dimensional predictors. For Model 1, the remaining five predictors $(X_6, \ldots, X_{10})$ were independently generated from $U(0, 1)$. For Model 2, these five predictors were randomly sampled from an exponential distribution with a scale parameter equal to one. In both models, the random errors $(\varepsilon_1, \ldots, \varepsilon_4)$ were independently sampled from $N(0, 1)$.

Model 1

$\boldsymbol{\eta}_1 = (1,-1,-1,0,0,0,0,0,0,0)^{\mathrm{T}}$; $\boldsymbol{\eta}_2 = (1,0,0,0,0,0,0,0,0,0)^{\mathrm{T}}$.

$Y_1 = \exp(0.5\boldsymbol{\eta}_1^{\mathrm{T}}\mathbf{X} + 1) + 0.1\varepsilon_1$; $Y_2 = \boldsymbol{\eta}_2^{\mathrm{T}}\mathbf{X} + (\boldsymbol{\eta}_2^{\mathrm{T}}\mathbf{X})^2 + 0.1\varepsilon_2$; $Y_3 = Y_1 + Y_2 + 0.1\varepsilon_3$; $Y_4 = |\boldsymbol{\eta}_1^{\mathrm{T}}\mathbf{X}| + 0.1\varepsilon_4$.

Model 2

$\boldsymbol{\eta}_1 = (1,-1,-1,0,0,0,0,0,0,0)^{\mathrm{T}}$; $\boldsymbol{\eta}_2 = (0,0,0,0,0,1,-1,0,0,0)^{\mathrm{T}}$.

$Y_1 = \log(\boldsymbol{\eta}_1^{\mathrm{T}}\mathbf{X}) + 0.1\varepsilon_1$; $Y_2 = \exp(\boldsymbol{\eta}_2^{\mathrm{T}}\mathbf{X}) + 0.1\varepsilon_2$; $Y_3 = (\boldsymbol{\eta}_2^{\mathrm{T}}\mathbf{X})^2 + 0.1\varepsilon_3$.

In Models 1 and 2, each coordinate regression of $Y_i|\mathbf{X}$ has various types of mean functions such as non-linear mean, symmetric mean, the second-order polynomial mean and so on. The models show the characteristic behaviors in the estimation of $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$ that are observed from the other simulation models in Ko and Yoo (2022).

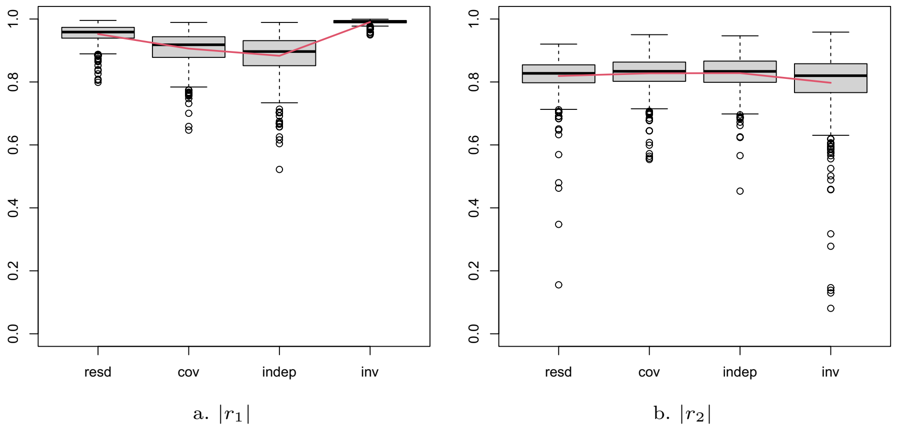


Fig. 1 Boxplots of $|r_1|$s and $|r_2|$s in Model 1 for comparison. Methods: resd, residual-PRcovM; cov, cov ($\mathbf{Y}$)-PRcovM; indep, indentity-PRcovM; inv, cPRmM of Ko and Yoo (2022)

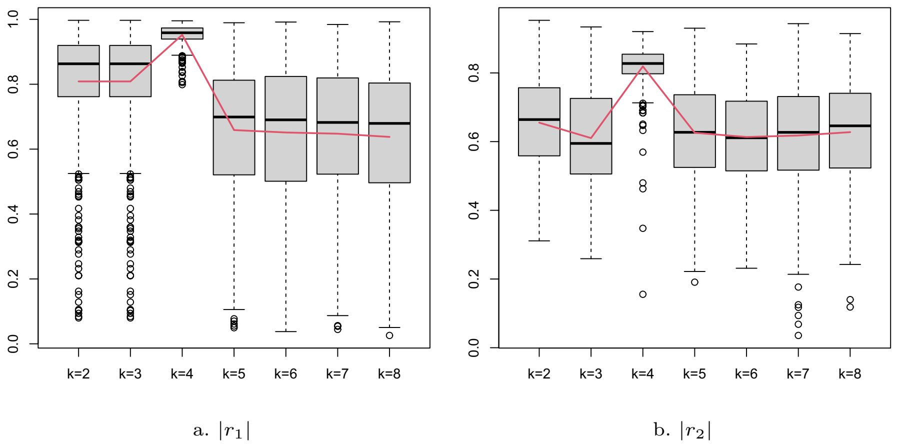


Fig. 2 Boxplots of $|r_1|$s and $|r_2|$s in Model 1, using residual-based PRcovM with varying cluster sizes

For cov ($\mathbf{T}$), the following three candidates were considered: the inverse of the sample covariance matrix of the ordinary least squared residual matrix (resd), the inverse of the sample covariance matrix of the response (cov), and the identity matrix (indep). For a methodological comparison, following the recommendation of Ko and Yoo (2022), the coordinate projective resampling mean method was considered. The number of resampling for PRcovM was 500, while it was 100 for the coordinate-projective resampling mean method (cPRmM).

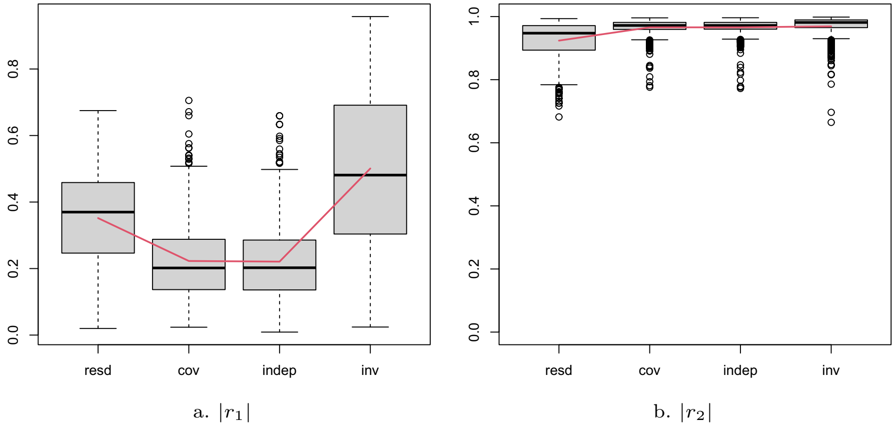


Fig. 3 Boxplots of $|r_1|$s and $|r_2|$s in Model 2 for comparison. Methods: resd, residual-PRcovM; cov, cov $(\mathbf{Y})$-PRcovM; indep, indentity-PRcovM; inv, cPRmM of Ko and Yoo (2022)

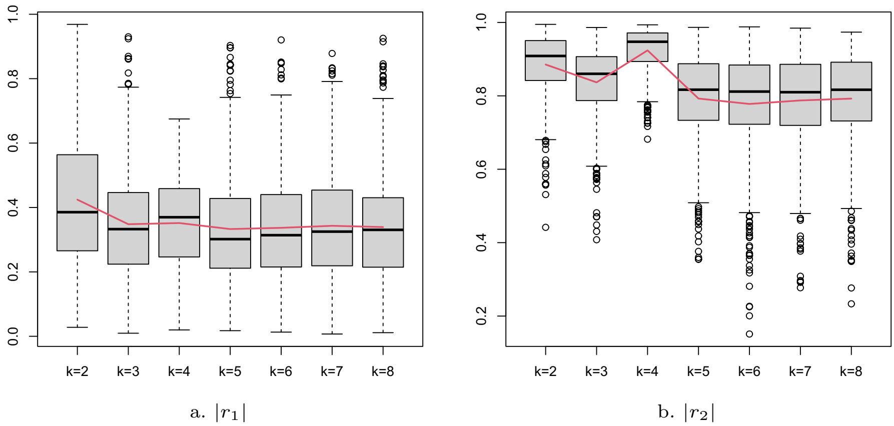


Fig. 4 Boxplots of $|r_1|$s and $|r_2|$s in Model 2, using residual-based PRcovM with varying cluster sizes

To measure how well $\boldsymbol{\eta}_i$ for each model is estimated, R_i^2 was calculated from a regression of $\boldsymbol{\eta}_i^{\mathrm{T}}\mathbf{X}|\hat{\boldsymbol{\eta}}^{\mathrm{T}}\mathbf{X}$ for $i = 1, 2$. Defining $|r_i| = |\sqrt{R_i^2}|$, higher values of $|r_i|$ indicates a better estimation of $\boldsymbol{\eta}_i$. As a graphical summary, $|r_i|$s were box-plotted.

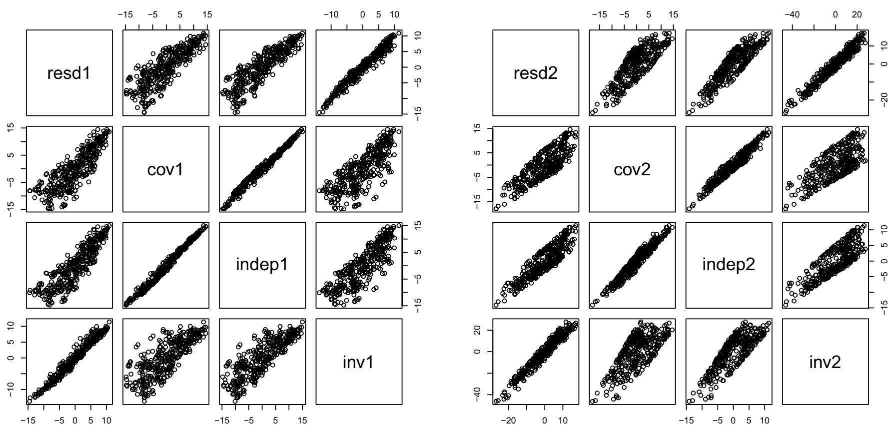
The results of Models 1 and 2 are summarized in Figs. 1, 2, 3 and 4, respectively. The solid lines in each boxplot in Figs. 1, 2, 3 and 4 represent the averages of $|r_1|$s and $|r_2|$s. First, to evaluate the estimation performances of $\boldsymbol{\eta}$, we set the number of clusters for $\mathbf{X}$ to four, Figs. 1 and 2 summarize the results for PRcovM with three different choices of cov $(\mathbf{T})$ and cPRmM of Ko and Yoo (2022). According to the figures, "resd" demonstrates superior or comparable performance to the other two choices across Models 1 and 2. For the first basis estimation, the Ko-Yoo method of cPRmM outperforms all three PRcovM, while PRcovM achieves slightly better or comparable accuracy in the second basis estimation compared to cPRmM. Next, the impact of the clusters were investigated with the residual-based PRcovM. Its results are reported in Figs. 2 and 4. The figures indicate that the four clusters are comparatively better than the other cluster sizes in Models 1 and 2.

The numerical studies highlight the importance of carefully selecting cov$(\mathbf{T})$ and the number of clusters for PRcovM in accurately estimating $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$. With these considerations, PRcovM can effectively compete with the existing cPRmM method proposed by Ko and Yoo (2022). Based on the results, we recommend to use the residual-based PRcovM with four clusters for $\mathbf{X}$.

4.2 Real data example

For illustration purposes, we revisit the Power Consumption of Tetuan City dataset from the UCI Machine Learning Repository, as previously demonstrated in Ko and Yoo (2022). The dataset is available at "https://archive.ics.uci.edu/ml/datasets/Power+consumption+of+Tetouan+city". This dataset captures the power consumption for three zones of Tetuan City in northern Morocco. The dataset includes measurements of Temperature, Humidity, Wind Speed, general diffuse flow, and diffuse flow as predictors. The three-dimensional responses correspond to the power consumption in Zones 1–3 of Tetuan City. The original dataset comprises 52,416 data points collected by the Supervisory Control and Data Acquisition (SCADA) system at 10-minute intervals between January 1, 2017, and December 31, 2017. The daily average for each variable was obtained and used for the analysis by consisting of 365 data.

For PRcovM, the three choices for cov$(\mathbf{T})$—namely, "resd", "cov" and "indep"—were adopted, consistent with the numerical studies. Again, the Ko-Yoo method of



a. Estimated first sufficient predictor

b. Estimated second sufficient predictor

Fig. 5 Scatterplot matrices of sufficient predictors for power consumption data. Method: resd, residual-PRcovM; cov, cov$(\mathbf{Y})$-PRcovM; indep, indentity-PRcovM; inv, cPRmM of Ko and Yoo (2022) cPRmM is denoted as “inv”. According to Ko and Yoo (2022), two-dimensional sufficient predictors should be enough to replace the original predictors. The scatterplot matrices of the estimated sufficient predictors obtained from each method are presented in Fig. 5. As shown in Fig. 5, PRcovM with “resd” and the Ko-Yoo method yield essentially equivalent estimations, while PRcovM with “cov” and “indep” produce highly similar results. These findings confirm that PRcovM with “resd” is a practical and reliable alternative to the Ko-Yoo method.

5 Discussion

In this paper, we propose a covariance approach to a projective resampling informative predictor subspace (Ko and Yoo 2022; PRIPS) for multivariate regression, and newly define a covariance projective resampling informative predictor subspace (covPRIPS). To clarify the relation between covPRIPS and the central subspace, two conditions are required, which are normally expected to hold. Under the conditions, covPRIPS becomes a smaller subspace to contain the central subspace but to be contained in the existing PRIPS of Ko and Yoo (2022).

Two potential advantages of covPRIPS over the PRIPS are as follows. One should be no necessity of the number of slices of $t^{\mathrm{T}}\mathbf{Y}$ in the PRIPS. Also, covPRIPS provides an intuitive understanding of the role of the linear transformation $t^{\mathrm{T}}\mathbf{Y}$, demonstrating how it affects the $E(\mathbf{M}(\mathbf{T}))$ through $\mathrm{cov}\,(\mathbf{T})$. According to numerical studies, the estimation method of covPRIPS can effectively compete with the existing inverse mean method of PRIPS.

Still, the projective resampling approach to recovering the central subspace requires iterative sampling of t, but its optimal sampling number is not theoretically derived. To overcome this limitation, a sufficient response dimension reduction is being integrated into PRIPS, and this work is currently in progress.

Acknowledgements For Jae Keun Yoo and Minjee Kim, this work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korean Ministry of Education (RS-2023-00240564 and RS-2023-00217022). The authors are grateful to the two referees and the Associate Editor for their many helpful comments and suggestions.

Declarations

Conflict of interest The authors declare no competing interests. Jae Keun Yoo is an Associate Editor of Journal of the Korean Statistical Society. Associate Editor status has no bearing on editorial consideration.

References

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

Cook, R. D. (1998). *Regression graphics: Ideas for studying regressions through graphics*. New York: Wiley.

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

Li, B., & Dong, Y. (2009). Dimension reduction for nonelliptically distributed predictors. *Annals of Statistics, 37*, 1272–1298.

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

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

Yin, X., & Bura, E. (2006). Moment-based dimension reduction for multivariate response regression. *Journal of Statistical Planning and Inference, 136*, 3675–3688.

Yoo, J. K. (2008). Sufficient dimension reduction for the conditional mean in multivariate regression with categorical predictors. *Journal of Multivariate Analysis, 99*, 1825–1839.

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

Yoo, J. K. (2022). Note on the estimation of informative predictor subspace and projective resampling informative predictor subspace. *The Korean Journal of Applied Statistics, 35*, 657–666.

Yoo, J. K., & Cook, R. D. (2007). Optimal sufficient dimension reduction for the conditional mean in multivariate regression. *Biometrika, 94*, 231–242.