

# Principal Components Regression by using Generalized Principal Components Analysis

(Last Modified: June 6, 2012)

Masakazu FUJIWARA<sup>1</sup>, Tomohiro MINAMIDANI<sup>1</sup>, Isamu NAGAI<sup>1\*</sup> and Hirofumi WAKAKI<sup>1</sup>

<sup>1</sup>*Department of Mathematics, Graduate School of Science, Hiroshima University  
1-3-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8526, Japan*

## Abstract

Principal components analysis (PCA) is one method for reducing the dimension of the explanatory variables, although the principal components are derived by using all the explanatory variables. Several authors have proposed a modified PCA (MPCA), which is based on using only selected explanatory variables in order to obtain the principal components (see e.g., Jolliffe, 1972, 1986; Robert and Escoufier, 1976; Tanaka and Mori, 1997). However, MPCA uses all of the selected explanatory variables to obtain the principal components. There may, therefore, be extra variables for some of the principal components. Hence, in the present paper, we propose a generalized PCA (GPCA) by extending the partitioning of the explanatory variables. In this paper, we estimate the unknown vector in the linear regression model based on the result of a GPCA. We also propose some improvements in the method to reduce the computational cost.

*Key words:* Cross validation; MPCA; Linear regression model; Principal components analysis; Step-up procedure; Variable selection.

## 1. Introduction

In the present paper, we work with a linear regression model with  $n$  observations of the response variables and a  $p$ -dimensional vector of the regressors. Let  $\mathbf{y} = (y_1, \dots, y_n)$  be the  $n$ -dimensional response vector,  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$  be the  $n \times p$  matrix of nonstochastic explanatory variables of rank( $\mathbf{X}$ ) =  $p$  ( $< n$ ), and let  $\boldsymbol{\varepsilon}$  be the  $n$ -dimensional error vector, where  $n$  is the sample size. Then, the form of the linear regression model is expressed as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where  $\boldsymbol{\beta}$  is a  $p$ -dimensional unknown vector. The least-square estimator of  $\boldsymbol{\beta}$  is derived as  $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ . When  $n$  becomes small or  $p$  becomes large, the multicollinearity problem tends to occur. When that happens, the estimator of  $\boldsymbol{\beta}$  becomes unstable.

---

\*Corresponding author, E-mail: [inagai@hiroshima-u.ac.jp](mailto:inagai@hiroshima-u.ac.jp)

In order to avoid this problem, several methods for reducing the dimension of  $\mathbf{X}$  have been proposed and are often used. One of these methods is to use principal components analysis (PCA) (see, e.g., Jolliffe (1986), Hasie, Tibshirani, and Friedman (2009), Srivastava (2002)). When we use PCA, we combine the explanatory variables based on the eigenvalues and eigenvectors of the sample covariance matrix  $\mathbf{S} = \sum_{i=1}^n (\mathbf{x}_i - \sum_{j=1}^n \mathbf{x}_j/n)(\mathbf{x}_i - \sum_{j=1}^n \mathbf{x}_j/n)'/(n-1)$ . For estimating  $\boldsymbol{\beta}$ , we solve the following eigenvalue problem:

$$\mathbf{S}\mathbf{a} = \lambda\mathbf{a}. \quad (1.1)$$

Let the solutions of the above equation be  $\lambda_1 \geq \dots \geq \lambda_p \geq 0$  where  $\lambda_i$  is an eigenvalue of  $\mathbf{S}$ , and let  $\mathbf{a}_i$  be the eigenvector corresponding with  $\lambda_i$ . When we use the first through  $r_0$ th eigenvectors  $\mathbf{a}_1, \dots, \mathbf{a}_{r_0}$  ( $1 \leq r_0 \leq p$ ), we estimate  $\boldsymbol{\beta}$  and derive the predictor of  $\mathbf{y}$  as follows:

$$\hat{\boldsymbol{\beta}}_{\text{PCR}} = (\mathbf{A}'\mathbf{X}'\mathbf{X}\mathbf{A})^{-1}\mathbf{A}'\mathbf{X}'\mathbf{y} \text{ and } \hat{\mathbf{y}}_{\text{PCR}} = \mathbf{X}\mathbf{A}\hat{\boldsymbol{\beta}}_{\text{PCR}},$$

where  $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_{r_0})$ , which is a  $p \times r_0$  matrix. This method of estimation is called principal components regression (PCR). One method to decide  $r_0$  is based on the cross-validation (CV) method for minimizing the predicted mean squared error (PMSE). In Section 3, we illustrate the algorithm for selecting  $r_0$  in detail.

The principal components in PCR are derived by using all of the explanatory variables  $\mathbf{X}$ . However, occasionally,  $\hat{\boldsymbol{\beta}}_{\text{PCR}}$  or  $\hat{\mathbf{y}}_{\text{PCR}}$  becomes unstable. In order to stabilize the estimation method, we combine the variable selection method with PCA. Tanaka and Mori (1997) proposed a modified PCA (MPCA), which is obtained by using PCA after selecting the explanatory variables. In MPCA, we partition  $\mathbf{X}$  as  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ , where  $\mathbf{X}_i$  is an  $n \times q_i$  matrix,  $q_1 + q_2 = p$ , and  $\mathbf{X}_1$  corresponds with the variables selected by some method. Then, we use PCA only for  $\mathbf{X}_1$ , and then we can estimate  $\boldsymbol{\beta}$  and derive the predictor of  $\mathbf{y}$  with the same method as the PCR. We refer to this method as MPCR and will illustrate it in more detail in Section 2.

In MPCA, the principal components are derived from  $\mathbf{X}_1$ , which contains the selected variables. If some principal components depend only on part of  $\mathbf{X}_1$ , we can further partition  $\mathbf{X}_1$  into dependent and independent variables. By thus partitioning the selected explanatory variables of  $\mathbf{X}_1$ , we obtain more accurate principal components. We refer to this method as generalized PCA (GPCA) since we generalize the partitioning of the explanatory variables, and more details are presented in Section 2. In Section, 2 we propose that we can estimate  $\boldsymbol{\beta}$  and derive the predictor of  $\mathbf{y}$  by using the result of this method, which we call generalized PCR (GPCR). On the other hand, some group of the explanatory variables may depend on several of the principal components. If this is the case, then we change the partitioning of  $\mathbf{X}_1$  and obtain the estimator and predictor by using almost the same method as in GPCR. We call this method the more generalized PCR (MGPCR), and we illustrate this method in the appendix.

The remainder of the present paper is organized as follows: In Section 2, we illustrate MPCA and MPCR and propose our methods of GPCA and GPCR. In Section 3, we show the algorithm for optimizing several parameters in each of the estimating methods for  $\beta$ . In Section 4, we compare these methods by conducting numerical studies and show some techniques for reducing the computational cost. In Section 5, we present our conclusions.

## 2. MPCA, MPCR, and GPCR

In this section, firstly, we illustrate the MPCA method that was proposed by Tanaka and Mori (1997), and propose a new method, MPCR, which is based on MPCA. These methods are obtained by partitioning  $\mathbf{X}$  into two parts, the selected variables and the non-selected variables. Then, we obtain the results of MPCA by using PCA for the selected variables, and then we are able to perform MPCR by using the results of MPCA. In MPCA, we use the principal components that are obtained from all of the selected variables. Some of the principal components may depend on only some of the selected variables. Then, for obtaining these principal components, we can partition the selected variables into those that are dependent or independent for these components. Based on this idea, we propose a new method, which is referred as GPCA.

### 2.1. MPCA and MPCR

In this subsection, we illustrate MPCA, which was proposed by Tanaka and Mori (1997), and also MPCR, which is based on the results of MPCA. MPCA uses PCA for the selected variables in order to stabilize the estimator of  $\beta$ . That is, MPCA is based on  $\mathbf{X}_1$  which contains the selected variables, where  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ ,  $\mathbf{X}_i$  is an  $n \times q_i$  matrix, and  $q_1 + q_2 = p$ . For introducing MPCA, we partition the sample covariance matrix  $\mathbf{S}$  as follows:

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11}^{[1]} & \mathbf{S}_{12}^{[1]} \\ \mathbf{S}_{21}^{[1]} & \mathbf{S}_{22}^{[1]} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_1^{[1]} \\ \mathbf{S}_2^{[1]} \end{pmatrix},$$

where  $\mathbf{S}_{ij}^{[1]}$  is a  $q_i \times q_j$  matrix. This partitioning corresponds with the partitioning of  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ . Based on the results in Rao (1964), we wish to minimize  $\text{tr}\{\mathbf{S} - \mathbf{S}_1^{[1]'} \mathbf{W} (\mathbf{W}' \mathbf{S}_1^{[1]} \mathbf{W})^{-1} \mathbf{W} \mathbf{S}_1^{[1]}\}$ , which is the trace of the residual covariance matrix between  $\mathbf{X}$  and the best linear prediction for  $\mathbf{X}$  by using  $\mathbf{W}' \mathbf{X}_1'$ . This minimization problem boils down to the following generalized eigenvalue problem:

$$\mathbf{S}_1^{[1]} \mathbf{S}_1^{[1]'} \mathbf{b} = \nu \mathbf{S}_{11}^{[1]} \mathbf{b}. \quad (2.1)$$

By solving the above problem, we derive the generalized eigenvalues  $\nu_1 \geq \dots \geq \nu_{q_1}$  and  $\mathbf{b}_i$ , which is the eigenvector that corresponds with  $\nu_i$ . When we use the first through  $r_1$ th eigenvectors, we obtain the estimator of  $\beta$  and predictor of  $\mathbf{y}$  as follows:

$$\hat{\beta}_{\text{MPCR}} = (\mathbf{B}' \mathbf{X}_1' \mathbf{X}_1 \mathbf{B})^{-1} \mathbf{B}' \mathbf{X}_1' \mathbf{y} \text{ and } \hat{\mathbf{y}}_{\text{MPCR}} = \mathbf{X}_1 \mathbf{B} \hat{\beta}_{\text{MPCR}},$$

where  $\mathbf{B} = (\mathbf{b}_1, \dots, \mathbf{b}_{r_1})$  is a  $q_1 \times r_1$  matrix. This estimation method is referred as MPCR. Note that MPCR only depends on  $\mathbf{X}_1$ , which contains the selected explanatory variables. In MPCA, we need to decide the selected variables  $\mathbf{X}_1$ , the dimension  $q_1$  of  $\mathbf{X}_1$ , and the number  $r_1$  of principal components. We determine these parameters by using the CV method. Details of the selection algorithm are shown in Section 3. Extensions of this method for qualitative data were proposed by Mori, Tanaka, and Tarumi (1997).

## 2.2. GPCA and GPCR

In this subsection, we propose our methods of GPCA and GPCR. In MPCA, the principal components  $\mathbf{b}_1, \dots, \mathbf{b}_{r_1}$  are obtained from  $\mathbf{X}_1$ . For example, there is some group of variables that are meaningful to the principal component  $\mathbf{b}_i$  but meaningless to the other principal components  $\mathbf{b}_j$  ( $i \neq j$ ); there is another group of variables that are meaningful to  $\mathbf{b}_j$  but meaningless to  $\mathbf{b}_i$ . Hence we consider extending the partitioning of  $\mathbf{X}$  with and without overlapping partitions in order to derive more accurate principal components. We present GPCA, which is based on partitioning without overlapping partitions, in this subsection. MGPCR, which is based on partitioning with overlapping partitions, is presented in the appendix.

For GPCA, let  $\mathbf{X}$  be partitioned as  $(\mathbf{X}_1, \dots, \mathbf{X}_{s+1})$  without overlapping, where  $\mathbf{X}_i$  is an  $n \times q'_i$  matrix,  $q'_{s+1} = q_2$ , and  $q'_1 + \dots + q'_{s+1} = p$ . First, we derive a simple extension method in which the  $i$ th principal component is obtained from  $\mathbf{X}_i$ . As we did for MPCA in the above subsection, we partition the sample covariance matrix  $\mathbf{S}$  as follows:

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{11}^{[2]} & \cdots & \mathbf{S}_{1(s+1)}^{[2]} \\ \vdots & \ddots & \vdots \\ \mathbf{S}_{(s+1)1}^{[2]} & \cdots & \mathbf{S}_{(s+1)(s+1)}^{[2]} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_1^{[2]} \\ \vdots \\ \mathbf{S}_{(s+1)}^{[2]} \end{pmatrix},$$

where  $\mathbf{S}_{ij}^{[2]}$  is a  $q'_i \times q'_j$  matrix. This partitioning also corresponds with the partitioning of the explanatory variables  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_{s+1})$ . From Rao (1964), as was done for MPCA, we wish to minimize  $\text{tr}\{\mathbf{S} - \mathbf{S}_1^{[2]'} \mathbf{c} (\mathbf{c}' \mathbf{S}_{11}^{[2]} \mathbf{c})^{-1} \mathbf{c}' \mathbf{S}_1^{[2]}\}$ , which is the trace of the residual covariance matrix between  $\mathbf{X}$  and the best linear prediction for  $\mathbf{X}$  by using  $\mathbf{c}' \mathbf{X}_1$ . Without loss of generality, we assume  $\mathbf{c}' \mathbf{S}_{11}^{[2]} \mathbf{c} = 1$ . By using the Lagrange multiplier and by differentiating it with respect to  $\mathbf{c}$ , the following generalized eigenvalue problem is derived:

$$\mathbf{S}_1^{[2]} \mathbf{S}_1^{[2]'} \mathbf{c} = \eta \mathbf{S}_{11}^{[2]} \mathbf{c},$$

. A simpler form is obtained by letting  $\boldsymbol{\gamma} = (\mathbf{S}_{11}^{[2]})^{1/2} \mathbf{c}$ , and the eigenvalue problem is expressed as follows:

$$(\mathbf{S}_{11}^{[2]})^{-1/2} \mathbf{S}_1^{[2]} \mathbf{S}_1^{[2]'} (\mathbf{S}_{11}^{[2]})^{-1/2} \boldsymbol{\gamma} = \eta \boldsymbol{\gamma}. \quad (2.2)$$

Hence we obtain the maximum eigenvalue  $\eta_1$  and the corresponding eigenvector  $\boldsymbol{\gamma}_1$ . Then the first principal component is derived by  $\mathbf{c}_1 = (\mathbf{S}_{11}^{[2]})^{-1/2}\boldsymbol{\gamma}_1$ . The  $k$ th principal component is obtained by maximizing  $\mathbf{c}'\mathbf{S}_k^{[2]}\mathbf{S}_k^{[2]'}\mathbf{c}$  under  $\mathbf{c}$ , and satisfies  $\mathbf{c}'\mathbf{S}_{kk}^{[2]}\mathbf{c} = 1$  and  $\mathbf{c}'\mathbf{S}_{kj}^{[2]}\mathbf{c}_j = 0$  ( $j = 1, \dots, k-1$ ). Let  $\boldsymbol{\gamma} = (\mathbf{S}_{kk}^{[2]})^{1/2}\mathbf{c}$  and  $\boldsymbol{\gamma}_i = (\mathbf{S}_{ii}^{[2]})^{1/2}\mathbf{c}_i$  ( $i = 1, \dots, k-1$ ). Then these conditions can be represented as follows:

$$\boldsymbol{\gamma}'\boldsymbol{\gamma} = 1 \text{ and } \boldsymbol{\gamma} \in \left\langle (\mathbf{S}_{kk}^{[2]})^{-1/2}\mathbf{S}_{k1}^{[2]}(\mathbf{S}_{11}^{[2]})^{-1/2}\boldsymbol{\gamma}_1, \dots, (\mathbf{S}_{kk}^{[2]})^{-1/2}\mathbf{S}_{k(k-1)}(\mathbf{S}_{(k-1)(k-1)}^{[2]})^{-1/2}\boldsymbol{\gamma}_{k-1} \right\rangle^\perp,$$

where  $\langle \boldsymbol{\delta}_1, \dots, \boldsymbol{\delta}_\kappa \rangle$  is the space that is based on  $\boldsymbol{\delta}_1, \dots, \boldsymbol{\delta}_\kappa$ , and  $\langle \cdot \rangle^\perp$  is the orthogonal complement space. Let  $\boldsymbol{\Pi}_k$  be the orthogonal projection matrix (i.e.,  $\boldsymbol{\Pi}_k^2 = \boldsymbol{\Pi}_k' = \boldsymbol{\Pi}_k$ , and  $\boldsymbol{\Pi}_k\boldsymbol{\gamma} = \boldsymbol{\gamma}$ ) for the space  $\left\langle (\mathbf{S}_{kk}^{[2]})^{-1/2}\mathbf{S}_{k1}^{[2]}(\mathbf{S}_{11}^{[2]})^{-1/2}\boldsymbol{\gamma}_1, \dots, (\mathbf{S}_{kk}^{[2]})^{-1/2}\mathbf{S}_{k(k-1)}(\mathbf{S}_{(k-1)(k-1)}^{[2]})^{-1/2}\boldsymbol{\gamma}_{k-1} \right\rangle^\perp$ . Then maximizing  $\mathbf{c}'\mathbf{S}_k^{[2]}\mathbf{S}_k^{[2]'}\mathbf{c}$  under the conditions of  $\mathbf{c}'\mathbf{S}_{kk}^{[2]}\mathbf{c} = 1$  and  $\mathbf{c}'\mathbf{S}_{kj}^{[2]}\mathbf{c}_j = 0$  ( $j = 1, \dots, k-1$ ) is equivalent to maximizing  $\boldsymbol{\gamma}'\boldsymbol{\Pi}_k(\mathbf{S}_{kk}^{[2]})^{-1/2}\mathbf{S}_k^{[2]'}(\mathbf{S}_{kk}^{[2]})^{-1/2}\boldsymbol{\Pi}_k\boldsymbol{\gamma}$  under the condition of  $\boldsymbol{\gamma}'\boldsymbol{\gamma} = 1$ . By using the Lagrange multiplier and differentiating with respect to  $\boldsymbol{\gamma}$ , the following eigenvalue problem is obtained:

$$\boldsymbol{\Pi}_k(\mathbf{S}_{kk}^{[2]})^{-1/2}\mathbf{S}_k^{[2]}\mathbf{S}_k^{[2]'}(\mathbf{S}_{kk}^{[2]})^{-1/2}\boldsymbol{\Pi}_k\boldsymbol{\gamma} = \zeta\boldsymbol{\gamma}. \quad (2.3)$$

When we solve this equation, we obtain the maximum eigenvalue and the corresponding eigenvector  $\boldsymbol{\gamma}_k$ . We obtain the  $k$ th principal component as  $\mathbf{c}_k = (\mathbf{S}_{kk}^{[2]})^{-1/2}\boldsymbol{\gamma}_k$ .

When we use this simple extended method, we obtain only the  $i$ th principal component in the  $i$ th group  $\mathbf{X}_i$  ( $i = 1, \dots, s$ ). Next, we consider obtaining more than one principal component from each group  $\mathbf{X}_i$ . The  $j_1$ th through  $j_{r'_j}$ th ( $r'_j \leq q'_j$ ) principal components are derived from the  $j$ th group  $\mathbf{X}_j$  ( $j = 1, \dots, s$ ). When all  $j_{r'_j}$  are 1, the following method corresponds with the above simple extension. Letting  $\boldsymbol{\gamma} = (\mathbf{S}_{11})^{1/2}\mathbf{c}$  and considering the same problem as in the above simple extension, the  $1_1$ th through  $1_{r'_1}$ th principal components are obtained by solving equation (2.2). When we solve equation (2.2), we obtain the eigenvalues  $\eta_1 \geq \dots \geq \eta_{q'_1}$  and the corresponding eigenvectors  $\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_{q'_1}$ . By using these eigenvectors  $\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_{r'_1}$  ( $r'_1 \leq q'_1$ ), we obtain the  $1_j$ th principal components as  $\mathbf{c}_{1_j} = (\mathbf{S}_{11}^{[2]})^{-1/2}\boldsymbol{\gamma}_j$  ( $j = 1, \dots, r'_1, r'_1 \leq q'_1$ ). By using these principal components, we obtain the matrix  $\mathbf{C}_1 = (\mathbf{c}_{1_1}, \dots, \mathbf{c}_{1_{r'_1}})$ , which is a  $q'_1 \times r'_1$  matrix. As in the simple extension, we solve equation (2.3) to obtain the  $k_1$ th through  $k_{r'_k}$ th principal components. Let the solutions of equation (2.3) be  $\zeta_1 \geq \dots \geq \zeta_{q'_k}$  that are eigenvalues, and the corresponding eigenvectors be  $\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_{q'_k}$ . Since we need the first through  $k_{r'_k}$ th principal components, we use  $\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_{r'_k}$  and refer to them as  $\boldsymbol{\gamma}_{k_1}, \dots, \boldsymbol{\gamma}_{k_{r'_k}}$ . Then we obtain  $\mathbf{C}_k = (\mathbf{c}_{k_1}, \dots, \mathbf{c}_{k_{r'_k}})$ , which is a  $q'_k \times r'_k$  matrix where  $\mathbf{c}_{k_j} = (\mathbf{S}_{kk}^{[2]})^{-1/2}\boldsymbol{\gamma}_{k_j}$  ( $j = 1, \dots, r'_k$ ). Let  $\mathbf{Z} = (\mathbf{X}_1, \dots, \mathbf{X}_s)$  and

$$\boldsymbol{\Psi} = \begin{pmatrix} \mathbf{C}_1 & \mathbf{O}_{q'_1 \times r'_2} & \cdots & \mathbf{O}_{q'_1 \times r'_s} \\ \mathbf{O}_{q'_2 \times r'_1} & \mathbf{C}_2 & \cdots & \mathbf{O}_{q'_2 \times r'_s} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O}_{q'_s \times r'_1} & \mathbf{O}_{q'_s \times r'_2} & \cdots & \mathbf{C}_s \end{pmatrix},$$

where  $\mathbf{O}_{l \times m}$  is a  $l \times m$  matrix where all the elements are zero. Then we obtain the estimator of  $\boldsymbol{\beta}$  and the predictor of  $\mathbf{y}$  as follows:

$$\hat{\boldsymbol{\beta}}_{\text{GPCR}} = (\boldsymbol{\Psi}' \mathbf{Z}' \mathbf{Z} \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}' \mathbf{Z}' \mathbf{y} \text{ and } \hat{\mathbf{y}}_{\text{GPCR}} = \mathbf{Z} \boldsymbol{\Psi} \hat{\boldsymbol{\beta}}_{\text{MMPCR}}.$$

This estimation method is referred as GPCR. To use this method, we need to decide the selected variables  $\mathbf{Z}$ , the dimension  $q'_i$  ( $i = 1, \dots, s$ ) of each matrix  $\mathbf{X}_i$ , and the number  $r'_i$  ( $i = 1, \dots, s$ ) of principal components in each  $\mathbf{X}_i$ . In this paper, by using the CV method, we can determine these parameters. We present the details of the algorithm for deciding these parameters in Section 3.

### 3. The algorithms for each method

In this section, we present the algorithms for PCR, MPCR, and GPCR. Many computational tasks are required when we consider determining several parameters in each method by using the CV method. Hence we propose the modified CV (MCV), which determines several parameters and thus reduces the computational cost.

#### 3.1. The algorithm for PCR

In this subsection, we illustrate the algorithm for PCR. In PCR, we need to determine the number of principal components. The algorithm for deciding the parameters for PCR by using the CV method is as follows:

1. Delete the  $i$ th row in  $\mathbf{y}$  and  $\mathbf{X}$  to obtain  $\mathbf{y}_{(-i)}$  and  $\mathbf{X}_{(-i)}$ .
2. Solve equation (1.1) using  $\mathbf{X}_{(-i)}$ , that is, use  $\mathbf{S}_{(-i)} = \sum_{j \neq i} \{\mathbf{x}_j - \sum_{l \neq i} \mathbf{x}_l / (n - 1)\} \{\mathbf{x}_j - \sum_{l \neq i} \mathbf{x}_l / (n - 1)\}' / (n - 2)$  instead of  $\mathbf{S}$ .
3. Obtain the solutions  $\lambda_{1(-i)} \geq \dots \geq \lambda_{p(-i)}$ , which are eigenvalues, and the corresponding eigenvectors  $\mathbf{a}_{1(-i)}, \dots, \mathbf{a}_{p(-i)}$  for  $\mathbf{S}_{(-i)}$ , and make  $\mathbf{A}_{(-i, r_0)} = (\mathbf{a}_{1(-i)}, \dots, \mathbf{a}_{r_0(-i)})$ .
4. Obtain the estimator and predictor,  $\hat{\boldsymbol{\beta}}_{\text{PCR}(-i, r_0)} = (\mathbf{A}'_{(-i, r_0)} \mathbf{X}'_{(-i)} \mathbf{X}_{(-i)} \mathbf{A}_{(-i, r_0)})^{-1} \mathbf{A}'_{(-i, r_0)} \mathbf{X}'_{(-i)} \mathbf{y}_{(-i)}$  and  $\hat{\mathbf{y}}_{(-i, r_0)} = \mathbf{x}'_i \mathbf{A}_{(-i, r_0)} \hat{\boldsymbol{\beta}}_{\text{PCR}(-i, r_0)}$ , where  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$ .
5. Calculate  $s_{i, r_0} = (y_i - \hat{y}_{(-i, r_0)})^2$  for  $i = 1, \dots, n$ .
6. Obtain  $E_{r_0} = \sum_{i=1}^n s_{i, r_0} / n$  after calculation for the fixed  $r_0$ .
7. Calculate  $E_{r_0}$  for  $r_0 = 1, \dots, p - 1$ , and minimize  $E_{r_0}$  to obtain the optimal  $r_0^*$ .
8. Obtain the eigenvalues  $\lambda_1 \geq \dots \geq \lambda_{r_0^*} \geq \dots \geq \lambda_p$  and the corresponding eigenvectors  $\mathbf{a}_1, \dots, \mathbf{a}_{r_0^*}, \dots, \mathbf{a}_p$  by solving equation (1.1).

9. Obtain the estimator and predictor,  $\hat{\boldsymbol{\beta}}_{\text{PCR}} = (\mathbf{A}^* \mathbf{X}' \mathbf{X} \mathbf{A}^*)^{-1} \mathbf{A}^* \mathbf{X}' \mathbf{y}$  and  $\hat{\mathbf{y}}_{\text{PCR}} = \mathbf{X} \mathbf{A}^* \hat{\boldsymbol{\beta}}_{\text{PCR}}$ , where  $\mathbf{A}^* = (\mathbf{a}_1, \dots, \mathbf{a}_{r_0})$ .

Note that for this method it is necessary to optimize only  $r_0$ .

Deciding  $r_0$  is one of the important problems in PCA. We use the CV method in this paper, and thus we can determine  $r_0$  with the same method when we use PCA.

### 3.2. The algorithm for MPCR

In this subsection, we illustrate the algorithm for MPCR which is based on the result of MPCA (Tanaka and Mori, 1997). For MPCA and MPCR, we need to select the variables for  $\mathbf{X}_1$ , the dimension  $q_1$  of  $\mathbf{X}_1$ , and the number  $r_1$  of the principal components. Note that  $q_1$  is decided when we choose  $\mathbf{X}_1$ . There are several methods for selecting variables in order to derive  $\mathbf{X}_1$  from  $\mathbf{X} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(p)})$ . We use a step-up procedure, since many computational tasks would otherwise be required if we were to consider all combinations of the explanatory variables. The algorithm for optimizing parameters and variable selection with a step-up procedure in MPCR is as follows:

1. Let  $\mathbf{y}_{(-i)}$  be obtained by deleting the  $i$ th row from  $\mathbf{y}$ , and let  $\mathbf{X}_{(\cdot, -j)}$  be obtained by deleting  $\mathbf{x}^{(j)}$ , which is the  $j$ th column of  $\mathbf{X}$ , and set  $t = 1$ . More generally, we express  $\mathbf{x}^{(j)}$  as  $\mathbf{X}_{(\cdot, j)}$ .
2. Let  $\mathbf{x}'_{(i, j)}$  and  $\mathbf{x}'_{(i, -j)}$  be the  $i$ th row vectors in  $\mathbf{X}_{(\cdot, j)}$  and  $\mathbf{X}_{(\cdot, -j)}$ , respectively. Let  $\mathbf{X}_{(-i, j)}$  and  $\mathbf{X}_{(-i, -j)}$  be obtained by deleting  $\mathbf{x}'_{(i, j)}$  and  $\mathbf{x}'_{(i, -j)}$  from  $\mathbf{X}_{(\cdot, j)}$  and  $\mathbf{X}_{(\cdot, -j)}$ , respectively.
3. Calculate  $\mathbf{S}_{11}^{[1]}$  and  $\mathbf{S}_{12}^{[1]}$ , which are the sample covariance matrices in  $\mathbf{X}_{(-i, j)}$  and between  $\mathbf{X}_{(-i, j)}$  and  $\mathbf{X}_{(-i, -j)}$ , respectively. Note that  $\mathbf{S}_{11}^{[1]}$  is a  $t \times t$  matrix, and  $\mathbf{S}_{12}^{[1]}$  is a  $t \times (p - t)$  matrix.
4. Obtain the eigenvalues  $\nu_1^{(-i, j)} \geq \dots \geq \nu_t^{(-i, j)}$  and the corresponding eigenvectors  $\mathbf{b}_1^{(-i, j)}, \dots, \mathbf{b}_t^{(-i, j)}$  by solving the generalized eigenvalue problem in equation (2.1) where  $\mathbf{S}_1^{[1]} = (\mathbf{S}_{11}^{[1]}, \mathbf{S}_{12}^{[1]})$ .
5. Obtain the estimator and predictor as  $\hat{\boldsymbol{\beta}}_{\text{MPCR}(-i, j, r_1)} = (\mathbf{B}'_{(-i, j, r_1)} \mathbf{X}'_{(-i, j)} \mathbf{X}_{(-i, j)} \mathbf{B}_{(-i, j, r_1)})^{-1} \mathbf{B}'_{(-i, j, r_1)} \mathbf{X}'_{(-i, j)} \mathbf{y}$  and  $\hat{\mathbf{y}}_{\text{MPCR}(-i, j, r_1)} = \mathbf{x}'_{(i, j)} \mathbf{B}_{(-i, j, r_1)} \hat{\boldsymbol{\beta}}_{\text{MPCR}(-i, j, r_1)}$ , where  $\mathbf{B}_{(-i, j, r_1)} = (\mathbf{b}_1^{(-i, j)}, \dots, \mathbf{b}_{r_1}^{(-i, j)})$  ( $r_1 \leq t$ ) under a given  $r_1$ .
6. Calculate  $s_{i, j, r_1} = (y_i - \hat{\mathbf{y}}_{\text{MPCR}(-i, j, r_1)})$  for  $i = 1, \dots, n$ ;  $j = 1, \dots, p - t + 1$ .
7. Calculate  $E_1(j, r_1) = \sum_{i=1}^n s_{i, j, r_1}^2$ .
8. Obtain  $j_t^*(r_1)$ , which minimizes  $E_1(j, r_1)$  under fixed  $r_1$ , and let  $\boldsymbol{\ell}_t = \mathbf{x}^{(j_t^*(r_1))}$ .
9. Renew  $t$  as  $t = t + 1$ . Set  $\mathbf{X}_{(\cdot, -j)} = \mathbf{X}^{[1]}$  and  $\mathbf{X}_{(\cdot, j)} = \mathbf{X}^{[2]}$ , where  $\mathbf{X}^{[1]}$  is  $\mathbf{X}$  with  $\mathbf{x}^{(j_t^*(r_1))}$  deleted, and  $\mathbf{X}^{[2]} = (\boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_{t-1}, \mathbf{x}^{(j_t^*(r_1))})$ .

10. Return to step 2 until  $t = p$ . Note that this method coincides with the method of PCR when  $t = p$ .
11. Obtain the optimized  $t$  as  $t^*$ ,  $r_1$  as  $r_1^*$ , and  $\mathbf{X}_1$  as  $\mathbf{X}_1^* = (\mathbf{x}^{(j_1^*(r_1^*))}, \dots, \mathbf{x}^{(j_{t^*}^*(r_1^*))})$  after iteration over  $i, j$ , and  $r_1$ .
12. Obtain the eigenvalues  $\nu_1 \geq \dots \geq \nu_{r_1^*} \geq \dots \geq \nu_{q_1}$  and the corresponding eigenvectors  $\mathbf{b}_1^*, \dots, \mathbf{b}_{r_1^*}^*, \dots, \mathbf{b}_{q_1}^*$  by using  $\mathbf{X}_1^*$ , which is an  $n \times q_1$  matrix. Partition  $\mathbf{S}$  corresponding to the dimension of  $\mathbf{X}_1^*$ , and solve equation (2.1).
13. Obtain the estimator and predictor as  $\hat{\boldsymbol{\beta}}_{\text{MPCR}} = (\mathbf{B}^{*\prime} \mathbf{X}_1^{*\prime} \mathbf{X}_1^* \mathbf{B}^*)^{-1} \mathbf{B}^{*\prime} \mathbf{X}_1^{*\prime} \mathbf{y}$  and  $\hat{\mathbf{y}}_{\text{MPCR}} = \mathbf{X}_1^* \mathbf{B}^* \hat{\boldsymbol{\beta}}_{\text{MPCR}}$ , where  $\mathbf{B}^* = (\mathbf{b}_1^*, \dots, \mathbf{b}_{r_1^*}^*)$ .

Note that this method needs to obtain  $\mathbf{X}_1^*$ ,  $q_1$ , and  $r_1^*$ . Hence, it is easy to predict that this method has more computations than PCR. Since GPCR needs to decide a greater number of parameters than are needed for MPCR, we can predict that more computations are required. Thus, when we use GPCR with the same methods as for PCR in Section 3.1 and MPCR in Section 3.2, it becomes impractical. Hence we consider improving the CV method in order to reduce the number of computations.

### 3.3. Improvements to the CV method

In this subsection, we improve the CV method in order to reduce the required computations so that it becomes a practical system. When we use the CV method with MPCR or GPCR, we must make many computations in order to optimize several parameters. In order to reduce the number of computations, we will improve the CV method, as follows. Let  $\mathbf{P}$  be the principal components matrix for each method, that is,  $\mathbf{P} = \mathbf{A}$  when we use PCR, and  $\mathbf{P} = \mathbf{B}$  when we use MPCR. First, we consider improving the method for  $s_{i,r_0}$  or  $s_{i,j,r_1}$  in each method. Using matrices, we can calculate this as follows:

$$\sum_{i=1}^n (y_i - \mathbf{x}_i' \mathbf{P} \hat{\boldsymbol{\beta}})^2 = \mathbf{y}' (\mathbf{I}_n - \mathbf{H}) (\mathbf{I}_n - \mathbf{G})^{-2} (\mathbf{I}_n - \mathbf{H}) \mathbf{y},$$

where  $\mathbf{H} = \mathbf{X} \mathbf{P} (\mathbf{P}' \mathbf{X}' \mathbf{X} \mathbf{P})^{-1} \mathbf{P}' \mathbf{X}'$ ,  $\mathbf{G} = \text{diag}(h_{11}, \dots, h_{nn})$ ,  $h_{ij}$  is the  $(i, j)$ th element of  $\mathbf{H}$ , and  $\hat{\boldsymbol{\beta}}$  is the estimator of each method. Second, we consider reducing the computations for obtaining the eigenvalues and eigenvectors, which are in  $\mathbf{P}$  for each method. Using the ordinary CV method, we delete the  $i$ th row vector in  $\mathbf{y}$  and  $\mathbf{X}$ . Then we obtain  $\mathbf{P}$  by using the eigenvectors in each method and calculate  $s_{i,r_0}$  or  $s_{i,j,r_1}$ . For each  $i$ , we calculate  $\mathbf{P}_{(-i)}$ , which contains the eigenvectors based on deleting the  $i$ th row vector in  $\mathbf{X}$ . We consider using  $\mathbf{P}$ , which is made up of the eigenvectors obtained by  $\mathbf{X}$ , instead of using  $\mathbf{P}_{(-i)}$ . By using  $\mathbf{P}$ , we can obtain  $\mathbf{Q} = (\mathbf{P}' \mathbf{x}_1, \dots, \mathbf{P}' \mathbf{x}_n)'$ . Then we consider deleting the  $i$ th row vector in  $\mathbf{Q}$  and using it to obtain  $\hat{\boldsymbol{\beta}}_{(-i)}$  for each method. Then we



calculate  $s_{i,r_0}$  or  $s_{i,j,r_1}$ . When we use this method, the number of computations is reduced since we obtain the eigenvector only one time for several fixed parameters. We call this improved method the modified CV (MCV) method.

When we use these improvements, especially the second improvement methods, the eigenvectors are changed. But, by conducting numerical studies with MPCR, we will show below that the precision of the CV method is not harmed with these improvements. First, we compare the CV and MCV methods based on the PMSE. Let  $\mathbf{x}_i$  be obtained independently from  $N_p(\mathbf{0}_p, \mathbf{I}_p)$  ( $i = 1, \dots, n+1$ ), and then  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$ , where  $\mathbf{0}_p$  is a  $p$ -dimensional vector all of whose elements are zero. Next, we obtain  $\mathbf{y}$  from  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ , where  $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}_n, \mathbf{I}_n)$  and  $\boldsymbol{\beta} = (\mathbf{1}_m, \mathbf{0}_{p-m})'$ , and  $\mathbf{1}_m$  is an  $m$ -dimensional vector all of whose elements are ones, and  $y_{n+1} = \mathbf{x}'_{n+1}\boldsymbol{\beta} + e$ , where  $e \sim N(0, 1)$ . We consider the following two estimators for  $R = E[(y_{n+1} - \mathbf{x}'_{n+1}\mathbf{B}\hat{\boldsymbol{\beta}}_{\text{MPCR}})^2]$ :

$$R_1 = \frac{1}{n} \sum_{i=1}^n (y_i - \mathbf{x}'_i \mathbf{B} \hat{\boldsymbol{\beta}}_{\text{MPCR}(-i)})^2 \text{ and } R_2 = \frac{1}{n} \sum_{i=1}^n (y_i - \mathbf{x}'_i \mathbf{B}_{(-i)} \hat{\boldsymbol{\beta}}_{\text{MPCR}(-i)})^2.$$

In these estimators,  $R_1$  is the risk corresponding with the MCV method, and  $R_2$  is the risk corresponding with the ordinary CV method. In order to compare these estimators, we calculate the averages of  $R - R_1$ ,  $R - R_2$ ,  $(R - R_1)^2$ , and  $(R - R_2)^2$ , all across 1,000 repetitions. By the following method, we compared the  $R_1$  and  $R_2$  obtained from using MPCR:

1. Let  $\mathbf{X}_0$  be obtained by the first through  $w$ th columns in  $\mathbf{X}$ , and let  $\mathbf{X}_1$  be obtained by deleting  $\mathbf{x}^{(w)}$  from  $\mathbf{X}$ .
2. Let  $r'$  be the number of principal components,  $\mathbf{B}$  be a  $w \times r'$  matrix, and  $\mathbf{B}_{(-i)}$  be a  $(w-1) \times r_1$  matrix.
3. Obtain  $\hat{\boldsymbol{\beta}}_{\text{MPCR}}$  and  $\hat{\boldsymbol{\beta}}_{\text{MPCR}(-i)}$ .
4. Let  $\mathbf{x}'_{i,0}$  be the  $i$ th row vector in  $\mathbf{X}_0$ , and then calculate  $R$ ,  $R_1$ , and  $R_2$  by using  $\mathbf{x}'_{n+1}$  and  $\mathbf{x}'_{i,0}$ .
5. Calculate the average of  $R - R_i$  and  $(R - R_i)^2$  for  $i = 1, 2$  across 1,000 repetitions.

The results of several situations are provided in Tables 1 and 2.

In Tables 1 and 2,  $R_1$  was obtained by using the MCV method, and  $R_2$  was obtained by the CV method. Based on the results, although  $R_1$  is biased, the mean squared error (MSE) of  $R_1$  is smaller than that of  $R_2$ . The values of the MSE obtained from  $R_1$ ,  $R_2$ , and  $R$  are nearly equal. Hence, we use  $R_1$  instead of using  $R_2$  on the sense of the MSE

Second, we conducted numerical studies to compare the PMSEs for the CV and MCV methods. As in the above numerical studies, we obtained the  $\mathbf{x}_i$  independently from  $N_p(\mathbf{0}_p, \mathbf{I}_p)$  for  $i = 1, \dots, n+1$ , and let  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$ ,  $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}_n, \mathbf{I}_n)$ , and  $\boldsymbol{\beta} = (\mathbf{1}_m, \mathbf{0}_{p-m})'$ . Then we obtained  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

**Table 1. The results of comparison when  $(n, p) = (30, 10)$** 

$m$	$w$	$r'$	$E[R_1]$	$E[R_2]$	$E[R]$	$E[R - R_1]$	$E[R - R_2]$	$E[(R - R_1)^2]$	$E[(R - R_2)^2]$
3	5	3	1.9762	2.2412	2.1579	0.1817	-0.0834	0.7017	0.9081
8	5	3	5.8131	6.4084	6.2043	0.3911	-0.2042	3.2592	4.5862

**Table 2. The results of comparison when  $(n, p) = (40, 15)$** 

$m$	$w$	$r'$	$E[R_1]$	$E[R_2]$	$E[R]$	$E[R - R_1]$	$E[R - R_2]$	$E[(R - R_1)^2]$	$E[(R - R_2)^2]$
3	12	9	1.7933	1.9922	1.9847	0.1914	-0.0076	0.3692	0.4970
10	9	7	3.6245	4.3016	4.5416	0.9170	0.2399	2.5817	2.9288
12	13	12	1.8991	2.2638	2.2950	0.3960	0.0312	0.7219	1.2654

**Table 3. The results of comparison based on the PMSE when  $(n, p) = (30, 15)$** 

$m$	MPCR with CV	MPCR with MCV	$p$ -value	$E_{\text{MPCR}}$	95% confidence interval
3	1.7837	1.7630	0.5372	0.0206	(-0.0445, 0.0863)
8	1.9468	1.8920	0.2645	0.0548	(-0.0415, 0.1511)
13	2.5616	2.6992	0.1759	0.1759	(-0.3370, 0.0618)

and  $y_{n+1} = \mathbf{x}'_{n+1}\boldsymbol{\beta} + e$ , where  $e \sim N(0, 1)$ . By using the algorithm for MPCR in Section 3.2, with the CV and MCV methods, we determined  $\mathbf{X}_1$ , the dimension of  $\mathbf{X}_1$ , and the number of principal components. Based on  $\mathbf{X}_1$ , we let  $\mathbf{x}_{i,1}$  be the corresponding variables in  $\mathbf{x}_{n+1}$ . Then, by using the number of principal components and  $\mathbf{X}_1$ , we obtained  $\mathbf{B}$  and  $\hat{\boldsymbol{\beta}}_{\text{MPCR}}$ . We calculated  $E_{\text{MPCR}} = (y_{n+1} - \mathbf{x}'_{i,1}\mathbf{B}\hat{\boldsymbol{\beta}}_{\text{MPCR}})^2$  and averaged  $E_{\text{MPCR}}$  across 1,000 repetitions. We fixed  $(n, p) = (30, 15)$ , and  $m = 3, 5, 8$ . We used the  $t$  test, which is a method to test if the difference is 0.

As shown in Table 3, the results of the  $t$  test show that there are not significant differences between the methods. Hence, based on the PMSE, the results of using the CV and MCV methods are nearly equal, and so, in order to reduce the number of computations, we can safely use the MCV method instead of the CV method.

### 3.4. The improved algorithms for PCR and MPCR

By using the MCV method, we improve the algorithms for both the PCR and MPCR methods, and the number of computations may be reduced to only a tenth of those needed for ordinary PCR and MPCR.

First, we present this improved method for PCR. We have to decide the number of the principal components  $r_0$  for PCR. We propose the following improved algorithm:

1. Solve the eigenvalue problem (1.1), and form  $\mathbf{A}_{(r_0)}$  from  $r_0$ th eigenvectors ( $1 \leq r_0 \leq p$ ).
2. Calculate  $\mathbf{y}'(\mathbf{I}_n - \mathbf{H}_{(r_0)})(\mathbf{I}_n - \mathbf{G}_{(r_0)})^{-2}(\mathbf{I}_n - \mathbf{H}_{(r_0)})\mathbf{y}$ , where  $\mathbf{G}_{(r_0)} = \text{diag}(h_{11}^{(r_0)}, \dots, h_{nn}^{(r_0)})$ ,  $h_{ij}^{(r_0)}$  is the  $(i, j)$ th element of  $\mathbf{H}_{(r_0)}$ , and  $\mathbf{H}_{(r_0)} = \mathbf{X}\mathbf{A}_{(r_0)}(\mathbf{A}'_{(r_0)}\mathbf{X}'\mathbf{X}\mathbf{A}_{(r_0)})^{-1}\mathbf{A}'_{(r_0)}\mathbf{X}'$ .
3. Calculate the above value for  $r_0 = 1, \dots, p$ , and then obtain the  $r_0^*$  which minimizes it.

4. Obtain the estimator and predictor by the same method as for ordinary PCR.

By using this algorithm, we can obtain the estimator and predictor with fewer computations.

Second, we propose an improved method for MPCR. We need to decide the dimension of  $\mathbf{X}_1$ , the selected variables  $\mathbf{X}_1$ , and the number of principal components  $r_1$ . The proposed improved algorithm is as follows:

1. Let  $t = 1$ , let  $\mathbf{X}_{(\cdot, -j)}$  be obtained by deleting  $\mathbf{x}^{(j)}$ , and let  $\mathbf{X}_{(\cdot, j)}$  be  $\mathbf{x}^{(j)}$ .
2. Obtain  $\mathbf{S}_{11}^{[1]}$  and  $\mathbf{S}_{12}^{[1]}$ , which are calculated by the sample covariance matrix of  $\mathbf{X}_{(\cdot, j)}$  and the sample covariance matrix between  $\mathbf{X}_{(\cdot, j)}$  and  $\mathbf{X}_{(\cdot, -j)}$ , respectively, then we obtain  $\mathbf{S}_1^{[1]} = (\mathbf{S}_{11}^{[1]}, \mathbf{S}_{12}^{[1]})$ .
3. Obtain the eigenvalues  $\nu_1^{(j)} \geq \dots \geq \nu_t^{(j)}$  and the corresponding eigenvectors  $\mathbf{b}_1^{(j)}, \dots, \mathbf{b}_t^{(j)}$  by solving the generalized eigenvalue problem in equation (2.1).
4. When we use  $\mathbf{B}_{(j, r_1)} = (\mathbf{b}_1^{(j)}, \dots, \mathbf{b}_{r_1}^{(j)})$ , we obtain  $\mathbf{y}'(\mathbf{I}_n - \mathbf{H}_{(j, r_1)})(\mathbf{I}_n - \mathbf{G}_{(j, r_1)})^{-2}(\mathbf{I}_n - \mathbf{H}_{(j, r_1)})\mathbf{y}$ , where  $\mathbf{G}_{(j, r_1)} = \text{diag}(h_{11}^{(j, r_1)}, \dots, h_{nn}^{(j, r_1)})$ ,  $h_{st}^{(j, r_1)}$  is the  $(s, t)$ th element of  $\mathbf{H}_{(j, r_1)}$ , and  $\mathbf{H}_{(j, r_1)} = \mathbf{X}\mathbf{B}_{(j, r_1)}(\mathbf{B}'_{(j, r_1)}\mathbf{X}'\mathbf{X}\mathbf{B}_{(j, r_1)})^{-1}\mathbf{B}'_{(j, r_1)}\mathbf{X}'$ .
5. Calculate the above value for  $1 \leq r_1 \leq t$  and obtain  $j_t^*(r_1)$  that minimizes the above value, and let  $\ell_t = \mathbf{x}^{(j_t^*(r_1))}$ .
6. Return to step 2 until  $t = p$ , the same as in step 9 in the algorithm for MPCR in Section 3.2.
7. Obtain the optimized  $t$  as  $t^*$ ,  $r_1$  as  $r_1^*$ , and  $\mathbf{X}_1^* = (\mathbf{x}^{(j_1^*(r_1^*))}, \dots, \mathbf{x}^{(j_{t^*}^*(r_1^*))})$ , the same as in the algorithm for MPCR, and the estimator and predictor are as obtained by the same method as in step 13 in the algorithm for the MPCR in Section 3.2.

By using this algorithm, we can reduce the number of computations for obtaining the estimator and predictor.

### 3.5. The algorithm for GPCR

Since we can reduce the number of computations, we propose an algorithm for GPCR that uses the MCV method. When we perform GPCR, the number of computations increases with the number of partitions of  $\mathbf{X}$ . Hence we consider partitioning  $\mathbf{X}$  into three groups,  $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)$ . For selecting the explanatory variables and obtaining  $\mathbf{X}_1$  and  $\mathbf{X}_2$ , we use the step-up procedure. We fix one explanatory variable as  $\mathbf{X}_1$ , one as  $\mathbf{X}_2$ , and then the rest of  $\mathbf{X}$  is  $\mathbf{X}_3$ . Then we use the same step-up procedure method for  $\mathbf{X}_2$  as we used for MPCR until the dimension of  $\mathbf{X}_3$  becomes zero, and we obtain the optimal  $\mathbf{X}_2^*$  which minimizes  $\mathbf{y}'(\mathbf{I}_n - \mathbf{H})(\mathbf{I}_n - \mathbf{G})^{-2}(\mathbf{I}_n - \mathbf{H})\mathbf{y}$ . Then we use the same method for selecting  $\mathbf{X}_1$  under fixed  $\mathbf{X}_2^*$ . Although we selected the explanatory variable for obtaining

$\mathbf{X}_1$  until the dimension of  $\mathbf{X}_1$  becomes  $p$  in MPCR, since GPCR requires many computations, we stop the loop for selection when the minimized value by using the added explanatory variables is no larger than that by using the originally explanatory variables. In GPCR, we need to determine  $\mathbf{X}_1$ ,  $\mathbf{X}_2$ , the dimensions of each matrix, and the number of principal components  $r_1$  and  $r_2$ . The proposed algorithm for this method is as follows:

1. Let  $\mathbf{X}_{(\cdot,j)} = \mathbf{X}_1$ ;  $\mathbf{X}_{(\cdot,-j)} = \mathbf{L}$ , which is obtained by deleting  $\mathbf{x}^{(j)}$  from  $\mathbf{X}$ ;  $\mathbf{L}_q = \mathbf{X}_2$ ;  $\mathbf{L}_{-q} = \mathbf{X}_3$ ;  $t = 1$ ; and  $l = 1$ .
2. Calculate  $\mathbf{S}_{11}^{[2]}$ ,  $\mathbf{S}_{12}^{[2]}$ ,  $\mathbf{S}_{13}^{[2]}$ ,  $\mathbf{S}_{22}^{[2]}$ ,  $\mathbf{S}_{23}^{[2]}$ , and  $\mathbf{S}_{33}^{[2]}$ , where  $\mathbf{S}_{ij}^{[2]}$  is the sample covariance matrix between  $\mathbf{X}_i$  and  $\mathbf{X}_j$ .
3. Obtain the 1<sub>1</sub>th through 1 <sub>$r'_1$</sub> th principal components and  $\mathbf{C}_1^{(j,r'_1)}$  based on these principal components by solving equation (2.2), where  $\mathbf{S}_1^{[2]} = (\mathbf{S}_{11}^{[2]}, \mathbf{S}_{12}^{[2]}, \mathbf{S}_{13}^{[2]})$  and  $\mathbf{S}_2 = (\mathbf{S}_{12}^{[2]'}, \mathbf{S}_{22}^{[2]}, \mathbf{S}_{23}^{[2]})$ .
4. Make  $\mathbf{\Pi}_2$  and obtain the 2<sub>1</sub>th through 2 <sub>$r'_2$</sub> th principal components by solving (2.3), and then we obtain  $\mathbf{C}_2^{(j,r'_2)}$  based on these principal components and  $\mathbf{\Psi}_{(j,r'_1,r'_2)}$ .
5. Calculate  $\mathbf{y}'(\mathbf{I}_n - \mathbf{H}_{(j,r'_1,r'_2)})(\mathbf{I}_n - \mathbf{G}_{(j,r'_1,r'_2)})^{-2}(\mathbf{I}_n - \mathbf{H}_{(j,r'_1,r'_2)})\mathbf{y}$ , where  $\mathbf{Z} = (\mathbf{X}_1, \mathbf{X}_2)$ ,  $\mathbf{H}_{(j,r'_1,r'_2)} = \mathbf{Z}\mathbf{\Psi}_{(j,r'_1,r'_2)}(\mathbf{\Psi}'_{(j,r'_1,r'_2)}\mathbf{Z}'\mathbf{Z}\mathbf{\Psi}_{(j,r'_1,r'_2)})^{-1}\mathbf{\Psi}'_{(j,r'_1,r'_2)}\mathbf{Z}'$ ,  $h_{ii'}^{(j,r'_1,r'_2)}$  is the  $(i, i')$ th element of  $\mathbf{H}_{(j,r'_1,r'_2)}$ , and  $\mathbf{G}_{(j,r'_1,r'_2)} = \text{diag}(h_{11}^{(j,r'_1,r'_2)}, \dots, h_{nn}^{(j,r'_1,r'_2)})$ .
6. Obtain  $q^*$  which minimizes the above value after changing  $q$ .
7. Renew  $\mathbf{L}_l = \mathbf{L}_{q^*}$ ,  $l = l+1$ , and  $\mathbf{L}$  by deleting the  $q^*$ th column of  $\mathbf{L}$ . Then  $\mathbf{X}_2 = (\mathbf{L}_1, \dots, \mathbf{L}_{l-1}, \mathbf{L}_q)$  and  $\mathbf{L}_{-q} = \mathbf{X}_3$ . We return to step 3 until the dimension of  $(\mathbf{X}_1, \mathbf{X}_2)$  is equal to the dimension of  $\mathbf{X}$ . When  $l \geq 2$ , we must select  $r'_2$ .
8. Obtain  $r_2^*$  and  $\mathbf{X}_2$  that minimize  $\mathbf{y}'(\mathbf{I}_n - \mathbf{H}_{(j,r'_1,r'_2)})(\mathbf{I}_n - \mathbf{G}_{(j,r'_1,r'_2)})^{-2}(\mathbf{I}_n - \mathbf{H}_{(j,r'_1,r'_2)})\mathbf{y}$ , change  $j$  to 1, and return to step 2.
9. Obtain  $d$ ,  $r_2^*$ , and  $\mathbf{X}_2$  that minimize  $\mathbf{y}'(\mathbf{I}_n - \mathbf{H}_{(j,r'_1,r'_2)})(\mathbf{I}_n - \mathbf{G}_{(j,r'_1,r'_2)})^{-2}(\mathbf{I}_n - \mathbf{H}_{(j,r'_1,r'_2)})\mathbf{y}$ .
10. Let  $\mathbf{X}_d = \mathbf{K}_t$ , obtain  $\mathbf{X}$  by deleting the  $d$ th column of  $\mathbf{X}$ , and increase  $t$  to  $t + 1$ . Also, let  $(\mathbf{K}_1, \dots, \mathbf{K}_{t-1}, \mathbf{X}_{(\cdot,j)}) = \mathbf{X}_1$ ,  $\mathbf{X}_{(\cdot,-j)} = \mathbf{L}$ ,  $\mathbf{L}_q = \mathbf{X}_2$ , and  $\mathbf{L}_{-q} = \mathbf{X}_3$ . Return to step 2. When  $t \geq 2$ , we have to select the number  $r'_1$  of principal components.
11. Stop the loop when the minimized value in  $t = t + 1$  is greater than the minimized value in  $t$ . If we stop the loop,  $\mathbf{X}_1^*$ ,  $\mathbf{X}_2^*$ ,  $r_1^*$ , and  $r_2^*$  are obtained by minimizing  $\mathbf{y}'(\mathbf{I}_n - \mathbf{H}_{(j,r'_1,r'_2)})(\mathbf{I}_n - \mathbf{G}_{(j,r'_1,r'_2)})^{-2}(\mathbf{I}_n - \mathbf{H}_{(j,r'_1,r'_2)})\mathbf{y}$ .

12. Obtain  $\mathbf{X}_1^*$  based on  $d$ ,  $\mathbf{X}_2^*$  based on  $q^*$ , then  $\mathbf{X}_3$  contains the remainder of  $\mathbf{X}$ . Derive  $\mathbf{C}_1$  and  $\mathbf{C}_2$  based on  $r_1'^*$  and  $r_2'^*$ , respectively, and  $\Psi$ .
13. Obtain the estimator and predictor as  $\hat{\beta}_{\text{GPCR}}$  and  $\hat{y}_{\text{GPCR}}$ .

## 4. Numerical studies

In this section, we compare PCR, MPCR, and GPCR by conducting numerical studies. We then propose a new method for improving the GPCR, which is referred to as GPCR-h and is based on the selected variables in MPCR. After proposing GPCR-h, we compare PCR, MPCR, GPCR, and GPCR-h by conducting numerical studies.

### 4.1. Numerical study 1

By using the algorithm proposed in Section 3.4, we compare these methods as follows:

1. Obtain  $\mathbf{x}_i \sim N_p(\mathbf{0}_p, \mathbf{I}_p)$  ( $i = 1, \dots, n+1$ ) independently and set  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$ .
2. Obtain  $\mathbf{y} = \mathbf{X}\beta + \varepsilon$  and  $y_{n+1} = \mathbf{x}'_{n+1}\beta + e$ , where  $\varepsilon \sim N_n(\mathbf{0}_n, \mathbf{I}_n)$ ,  $e \sim N(0, 1)$ , and  $\beta = (\mathbf{1}_m, \mathbf{0}_{p-m})'$ .
3. By using the appropriate algorithms, decide  $r_0^*$  for PCR;  $r_1^*$  and  $\mathbf{X}_1^*$  for MPCR; and  $r_1'^*$ ,  $r_2'^*$ ,  $\mathbf{X}_1^*$ , and  $\mathbf{X}_2^*$  for GPCR .
4. Use  $\mathbf{X}$  in PCR,  $\mathbf{X}_1^*$  in MPCR, and  $\mathbf{X}^\# = (\mathbf{X}_1^*, \mathbf{X}_2^*)$  in GPCR, and obtain the estimator for  $\beta$  in each method.
5. Calculate  $E_{\text{PCR}} = (y_{n+1} - \mathbf{x}'_{n+1}\mathbf{A}\hat{\beta}_{\text{PCR}})^2$  for PCR;  $E_{\text{MPCR}} = (y_{n+1} - \mathbf{x}'_{[1]}\mathbf{B}\hat{\beta}_{\text{MPCR}})^2$  for MPCR, where  $\mathbf{x}_{[1]}$  is the corresponding column with  $\mathbf{X}_1^*$  in  $\mathbf{x}_{n+1}$ ; and  $E_{\text{GPCR}} = (y_{n+1} - \mathbf{x}'_{[2]}\Psi\hat{\beta}_{\text{GPCR}})^2$  for GPCR, where  $\mathbf{x}_{[2]}$  is the corresponding column with  $\mathbf{X}^\#$  in  $\mathbf{x}_{n+1}$ .
6. Calculate the averages of  $E_{\text{PCR}}$ ,  $E_{\text{MPCR}}$ , and  $E_{\text{GPCR}}$  across 1,000 repetitions. Using each value of  $E_{\text{PCR}}$ ,  $E_{\text{MPCR}}$ , and  $E_{\text{GPCR}}$  in each repetition, perform the  $t$  test for the expected predicted error.

The results when  $(n, p) = (30, 15)$  are presented in Tables 4 and 5.

Based on these results, when  $p$  is nearly equal to  $m$ , PCR is the best and MPCR is the worst method. On the other hand, MPCR is the best method when  $m$  is small with respect to  $p$ . This result means that MPCR is the best method when we need to select the explanatory variables. GPCR is the second best method when  $m$  is small. However, the results of GPCR become worse as  $m$  becomes large. The estimator or predictor may become unstable, that is become very small or

**Table 4. The results of each method when  $(n, p) = (30, 15)$** 

$m$	$E_{\text{PCR}}$	$E_{\text{MPCR}}$	$E_{\text{GPCR}}$
3	2.2540	1.5719	1.8663
8	2.1247	1.9694	2.0181
13	2.0682	2.3362	2.3709

**Table 5. The test between each method when  $(n, p) = (30, 15)$** 

$m$	testing methods	$p$ -value	the difference	95% confidence interval
3	GPCR and PCR	$2.30 \times 10^{-6}$	-0.3878	(-0.5479, -0.2277)
	GPCR and MPCR	$7.58 \times 10^{-6}$	0.2944	(0.1660, 0.4227)
	PCR and MPCR	$2.20 \times 10^{-6}$	0.6821	(0.5352, 0.8291)
8	GPCR and PCR	0.2645	-0.1066	(-0.2940, 0.0808)
	GPCR and MPCR	0.5624	0.0487	(-0.1162, 0.2136)
	PCR and MPCR	0.1188	0.1553	(-0.0399, 0.3506)
13	GPCR and PCR	0.0020	0.3027	(0.1110, 0.4943)
	GPCR and MPCR	0.7092	0.0347	(-0.1479, 0.2174)
	PCR and MPCR	0.000	-0.2679	(-0.4034, -0.1325)

very large values because of this reason. For this reason, we stopped the selection algorithm even when we did not compare the all of adding explanatory variables. Thus, we propose a new method (GPCR-h) for stabilizing the result of GPCR.

## 4.2. Proposal of GPCR-h

Since the algorithm for GPCR is complex, occasionally the estimator or predictor may become unstable and the GPCR's results become worse. We improved the complexity of GPCR by combining it with MPCR. In this improvement, we used the step-up procedure based on the result of the selected variables in MPCR. That is, we lead the  $1_1$ th through  $1_{r'_1}$ th principal components based on MPCR, then we use the step-up procedure and lead  $2_1$ th through  $2_{r'_2}$ th principal components based on the added variables. By using this method, we not only combine the merits of MPCR and that of GPCR, but also reduce the number of computations since the algorithm can be simplified. We refer to this method as GPCR-h, and the algorithm we used for the GPCR-h numerical studies is as follows:

1. From MPCR, derive  $\mathbf{X}_1^*$ , which contains the selected variables and is an  $n \times q'_1$  matrix, and the number  $r_0^*$  of principal components. Obtain  $\mathbf{L}$  by omitting  $\mathbf{X}_1^*$  from  $\mathbf{X}$ .
2. Set  $t = q'_1$ ,  $\mathbf{L}_j = \mathbf{X}_2$ ,  $\mathbf{L}_{-j} = \mathbf{X}_3$ , and  $l = 1$  in the algorithm for GPCR. Then return to step 2 in the algorithm for GPCR in Section 3.5.
3. Obtain  $r_1'^*$  that minimizes  $\mathbf{y}'(\mathbf{I}_n - \mathbf{H}_{(r'_1, r'_2)})(\mathbf{I}_n - \mathbf{G}_{(r'_1, r'_2)})^{-2}(\mathbf{I}_n - \mathbf{H}_{(r'_1, r'_2)})\mathbf{y}$  for fixed  $r'_2$  by using the GPCR algorithm for each  $r'_1$ .

**Table 6. The results of each method when  $(n, p) = (30, 15)$** 

$m$	$E_{\text{PCR}}$	$E_{\text{MPCR}}$	$E_{\text{GPCR}}$	$E_{\text{GPCR-h}}$
3	2.1580	1.6307	1.9550	1.7381
8	2.1386	1.8213	1.9230	1.8340
13	2.3924	2.7941	2.6455	2.4176

**Table 7. The tests between each method when  $(n, p) = (30, 15)$** 

$m$	testing methods	$p$ -value
3	GPCR-h and GPCR	$3.17 \times 10^{-4}$
	GPCR-h and PCR	$4.00 \times 10^{-9}$
	GPCR-h and MPCR	$4.16 \times 10^{-5}$
8	GPCR-h and GPCR	0.2831
	GPCR-h and PCR	0.000
	GPCR-h and MPCR	0.7009
13	GPCR-h and GPCR	0.0237
	GPCR-h and PCR	0.7902
	GPCR-h and MPCR	0.0069

4. Instead of comparing the values when  $t$  and  $t + 1$  to select the variables in  $\mathbf{X}_2$ , use the step-up procedure until the number of selected variables becomes  $p - q'_1$ . Then obtain  $\mathbf{X}_2^*$  and  $r_2^*$  to minimize the evaluation for GPCR under fixed  $X_1^*$ ,  $r_1^*$  and  $q'_1$ .
5. Decide  $\mathbf{X}_1^*$  by using MPCR, and decide  $\mathbf{X}_2^*$ ,  $r_1^*$ , and  $r_2^*$  by using GPCR. Then we make  $\mathbf{X}_3$  and  $\Psi$ . After making them, we use the same method as we used to obtain  $\hat{\beta}_{\text{GPCR}}$  to obtain the estimator for  $\beta$ , which is referred as  $\hat{\beta}_{\text{GPCR-h}}$ .
6. Calculate the average of  $E_{\text{GPCR-h}} = (y_{n+1} - \mathbf{x}'_{[3]} \Psi \hat{\beta}_{\text{GPCR-h}})^2$  across 1,000 repetitions, where  $\mathbf{x}_{[3]}$  contains the explanatory variables that correspond with  $\mathbf{X}_1^*$  and  $\mathbf{X}_2^*$ .

By using this method, the number of computations may be reduced to a third of those required for ordinary GPCR. We compared the improved PCR and MPCR methods, and the GPCR and GPCR-h methods by conducting numerical studies in the next subsection.

### 4.3. Numerical study 2

By conducting numerical studies, we compared these four methods. The setting is the same as in numerical study 1 in Section 4.1, and the algorithm for GPCR-h is in the above subsection. We present the results in Tables 6 through 11.

From these results, it is obvious that the number of computations and the MSE of GPCR-h are better than those of GPCR. Thus we recommend using GPCR-h instead of GPCR. Hence, we compare only PCR, MPCR, and GPCR-h.

**Table 8. The results of each method when  $(n, p) = (60, 20)$** 

$m$	$E_{\text{PCR}}$	$E_{\text{MPCR}}$	$E_{\text{GPCR-h}}$
5	1.5595	1.2581	1.3376
10	1.5314	1.3843	1.4230
15	1.4525	1.3550	1.3801
18	1.4878	1.4564	1.4703

**Table 9. The tests between each method when  $(n, p) = (60, 20)$** 

$m$	testing methods	$p$ -value	the difference	95% confidence interval
5	GPCR-h and PCR	$1.09 \times 10^{-9}$	-0.2219	(-0.2887,-0.1552)
	GPCR-h and MPCR	$6.56 \times 10^{-5}$	0.0795	(0.0406,0.1184)
	PCR and MPCR	$3.40 \times 10^{-14}$	0.3014	(0.2245,0.3783)
10	GPCR-h and PCR	$9.17 \times 10^{-5}$	-0.1084	(-0.1626,-0.0543)
	GPCR-h and MPCR	0.0386	0.0386	(0.0075,0.0698)
	PCR and MPCR	$4.83 \times 10^{-6}$	0.1471	(0.0843,0.2099)
15	GPCR-h and PCR	0.0015	-0.0724	(-0.1170,-0.0278)
	GPCR-h and MPCR	0.0118	0.0251	(0.0056,0.0447)
	PCR and MPCR	0.000	0.0975	(0.0462,0.1488)
18	GPCR-h and PCR	0.3965	-0.0175	(-0.0580,0.0230)
	GPCR-h and MPCR	0.2577	0.0138	(-0.0101,0.0378)
	PCR and MPCR	0.1009	0.0314	(-0.061,0.0688)

**Table 10. The results of each method when  $(n, p) = (30, 25)$** 

$m$	$E_{\text{PCR}}$	$E_{\text{MPCR}}$	$E_{\text{GPCR-h}}$
5	3.3668	2.4664	3.0130
13	5.1312	5.6361	5.6260
20	5.60311	8.9809	7.554

**Table 11. The tests between each method when  $(n, p) = (30, 25)$** 

$m$	testing methods	$p$ -value	the difference	95% confidence interval
5	GPCR-h and PCR	0.3251	-0.3538	(-0.6780,-0.0295)
	GPCR-h and MPCR	$1.41 \times 10^{-10}$	0.5466	(0.3812,0.7121)
	PCR and MPCR	$1.12 \times 10^{-9}$	0.9004	(0.6131,1.1877)
13	GPCR-h and PCR	0.0768	0.4948	(-0.0533,1.0430)
	GPCR-h and MPCR	0.9539	-0.0101	(-0.3535,0.3333)
	PCR and MPCR	0.0586	-0.5049	(-1.0283,0.0184)
20	GPCR-h and PCR	$5.76 \times 10^{-6}$	1.9510	(1.1113,2.7906)
	GPCR-h and MPCR	$6.97 \times 10^{-7}$	-1.4268	(-1.9874,-0.8662)
	PCR and MPCR	$1.60 \times 10^{-12}$	-3.3777	(-4.3039,-2.4515)



**Table 12. The summarized results of our numerical studies**

$n \backslash m$	1	$\dots$	$p$
$n \gg p$	MPCR*	MPCR*	MPCR
$n > p$	MPCR*	GPCR-h	PCR*
$n \approx p$	MPCR*	PCR*	PCR*

We note the following results. First, we discuss the results when  $(n, p) = (60, 20)$ , which are presented in Tables 8 and 9. For all  $m$ , MPCR is the best method when it is easy to estimate the unknown vector, since the difference between  $n$  and  $p$  becomes large. We consider that this is the reason that MPCR includes the PCR, and that the results of MPCR are stable. We note that sometimes GPCR-h derived the same values as did MPCR, though GPCR-h derived significantly different values when  $m$  was small. When  $m$  was nearly equal to  $p$ , the methods were did not produce significantly different results. We can see that GPCR-h is always the second-best method and derives stable values. Second, we discuss the results when  $(n, p) = (30, 25)$ , which are presented in Tables 10 and 11. For this setting,  $n$  is nearly equal to  $p$ . When  $m$  is small, MPCR is the best method. The differences between the methods become small when  $m$  becomes large. Moreover, for large  $m$ , PCR is the best method and MPCR is the worst method, and the difference is significant. The results of MPCR have a large variance when  $n$  is nearly equal to  $p$  and  $m$  is large. A stabilized estimator is derived by using GPCR-h, and GPCR-h is always the second-best method. We summarize these results in Table 12. The name of the method in each cell is the method which best minimizes the PMSE, and the name with \* indicates the significantly best method.

When  $n \gg p$ , such as  $(n, p) = (80, 20)$  or  $(n, p) = (60, 10)$ , MPCR is the best method for all  $m$ . The significant differences between the methods disappear when  $m$  becomes nearly equal to  $p$ . When  $n > p$ , such as  $(n, p) = (40, 20)$  or  $(n, p) = (50, 25)$ , the significantly best method is MPCR when  $m$  is small. On the other hand, when  $m$  is large, the order is reversed, and the significantly best method is PCR. When  $n \approx p$ , such as  $(n, p) = (30, 25)$  or  $(n, p) = (40, 30)$ , the significantly best method also is MPCR when  $m$  is small. MPCR becomes the worst method, and PCR becomes the best method, when  $m$  becomes large. In this table, GPCR-h appeared only one time. However, the values of the PMSE are as small or smaller than with the other methods. The most stable method is GPCR-h since the GPCR-h has a significant difference between the best and the worst methods. When we consider using these methods for actual data, the parameter  $m$  is the unknown variable. Since GPCR-h is the most stable method for all  $m$ , we recommend using this method for analyzing actual data.

#### 4.4. Numerical study 3

**Table 13.** The percentages of variables selected correctly in each method when  $(n, p) = (50, 25)$

$m$	Method	PMSE	PC-1	PC-2
3	GPCR-h	1.5311	100.0	23.50
	MPCR	1.4028	100.0	38.25
13	GPCR-h	1.8275	99.98	67.17
	MPCR	1.7541	99.92	79.17
23	GPCR-h	2.0577	99.87	95.78
	MPCR	2.1066	99.64	95.96

In the present paper, we select the explanatory variables by minimizing the PMSE when using the MCV method. In this subsection, we evaluate the true explanatory variables selected by the MCV method, based on the percentage of those that are selected correctly. We consider two ways of evaluating the percentage that are selected correctly,  $100 \times \#(T \cap S)/\#(T)$  (PC-1) and  $100 \times \#(T \cap S)/\#(S)$  (PC-2), where  $\#(U)$  means the number of elements in the set  $U$ ,  $T$  is the set of true explanatory variables, and  $S$  is the set of selected variables in each method. As an example, we show the results when  $m = 4$ . We obtain  $\mathbf{y} = \mathbf{x}^{(1)}\beta_1 + \mathbf{x}^{(2)}\beta_2 + \mathbf{x}^{(3)}\beta_3 + \mathbf{x}^{(4)}\beta_4 + \boldsymbol{\varepsilon}$ , where  $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \mathbf{0}'_{p-4})'$ . Then  $T = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}\}$ . By performing MPCR, the selected variables  $\mathbf{X}_1^*$  are  $\mathbf{X}_1^* = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(5)}, \mathbf{x}^{(6)})$ . Then  $S = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(5)}, \mathbf{x}^{(6)}\}$ . Thus we obtain PC-1 and PC-2 as 75 and 60. When we perform GPCR-h, the selected variables can be obtained as  $(\mathbf{X}_1^*, \mathbf{X}_2^*)$ . We compute these values in each repetition, and calculate the average of the values across 1,000 repetitions. The setting of data is the same as in the previous numerical studies. The results are in Table 13.

Based on these results, the values of PC-1 for each method are nearly equal to 100. This result means that the true explanatory variables are almost selected in each method. On the other hand, when  $m$  is small, the values of PC-2 become small for each method. This means that, in each method, there are unnecessary variables in the selected variables. By comparing each method, the values in PC-1 are nearly equal to each other in several situations. The values of PC-1 from GPCR-h are a little larger than those from MPCR. In the values of PC-2, those from MPCR are larger than those from GPCR-h. This is the reason for the differences between the PMSEs.

## 5. Conclusions

In the ordinary linear model, when the dimension of the explanatory variables becomes large, the ordinary estimation methods tend to become unstable. Then, the selection of a subset of explanatory variables is often used to reduce the dimension of the explanatory variables. One method of reducing the dimension of the explanatory variables is PCA. By using PCA, we shrink the original explanatory

variables to some synthesis variables, and this method means that we can reduce the dimension of explanatory variables by using several synthesis variables. Hence, based on PCA, an estimation method that is referred to as PCR has been proposed. MPCR, which is based on the selecting variables and PCA, proposed in order to stabilize the result of PCR. In the present paper, we also propose GPCR and GPCR-h, which are based on MPCR by partitioning the selected variables into dependent and independent variables for each principal component.

By solving the generalized eigenvalue problem, we can obtain the principal components for MPCR. In GPCR, we only solve the ordinary eigenvalue problems by using the orthogonal projection matrix for the fixed space. Then we showed the algorithms for PCR and MPCR with each parameter optimized by using the CV method. When we use the same optimizing method for GPCR, many computations are required since GPCR is obtained by optimizing a greater number of parameters than for MPCR. Hence we modified the CV method and proposed the MCV method. Then we compared the difference between the results of the MPCR by using the CV and MCV methods. By conducting numerical studies, we noted this modification worked very well. By using the MCV method, we improved the algorithms for PCR, MPCR, and GPCR.

After conducting several numerical studies, we noted that the result of GPCR is occasionally unstable. Hence we proposed GPCR-h based on MPCR and GPCR. We determined that GPCR-h is better than GPCR, by conducting numerical studies. Furthermore, several methods have advantages or disadvantages that depend on the structure of the data. For obtaining stable results, we recommend using GPCR-h for actual data analysis.

## Appendix

### GPCA and GPCR with overlapping partitioning

In this appendix, we propose a more generalized PCR, which is based on partitioning with overlap and is referred to as MGPCR. In Section 2.2, we partitioned the selected explanatory variables  $\mathbf{X}_1$  without overlapping. This partitioning is based on the idea that some principal components only relate to a subset of the explanatory variables. However, there are some principal components which depend on several groups of explanatory variables, that is, the  $i$ th principal components depend on  $\mathbf{X}_{j_1}$  and  $\mathbf{X}_{j_2}$  for some  $i$ ,  $j_1$ , and  $j_2$  ( $j_1 \neq j_2$ ). Hence we consider generalizing the partitioning for the explanatory variables with overlapping situations.

Let  $\mathbf{X}$  be partitioned  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_s, \mathbf{X}_{s+1})$ , where  $\mathbf{X}_{s+1}$  does not overlap but the other  $\mathbf{X}_i$  ( $i = 1, \dots, s$ ) may have overlapping parts. To simplify the discussion, we consider some simple overlapping in the partition. Let  $\mathbf{X}_i$  be made up of  $\mathbf{M}_0$  and  $\mathbf{M}_i$  for  $i = 1, \dots, s$ ,  $\mathbf{M}_0, \dots, \mathbf{M}_s$  do not have any overlapping parts, and let  $q_i''$  be the dimension of  $\mathbf{M}_i$ . By using  $\mathbf{M}_i$  ( $i = 0, 1, \dots, s$ ), the partitioning for  $\mathbf{X}$  can be expressed as  $(\mathbf{M}_0, \dots, \mathbf{M}_s, \mathbf{M}_{s+1})$ , where  $\mathbf{M}_{s+1}$  corresponds with  $\mathbf{X}_{s+1}$ ,

$q''_{s+1}$  is the dimension of  $\mathbf{M}_{s+1}$ , and  $q''_0 + \dots + q''_{s+1} = q$ .

Based on this partitioning of  $(\mathbf{M}_0, \dots, \mathbf{M}_{s+1})$ , we can partition the sample covariance matrix  $\mathbf{S}$  as follows:

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{00}^{[3]} & \cdots & \mathbf{S}_{0(s+1)}^{[3]} \\ \vdots & \ddots & \vdots \\ \mathbf{S}_{(s+1)0}^{[3]} & \cdots & \mathbf{S}_{(s+1)(s+1)}^{[3]} \end{pmatrix},$$

where  $\mathbf{S}_{ij}^{[3]}$  is a  $q''_i \times q''_j$  matrix. Let  $\mathbf{T}_{11}$  and  $\mathbf{T}_1$  be

$$\mathbf{T}_{11} = \begin{pmatrix} \mathbf{S}_{00}^{[3]} & \mathbf{S}_{01}^{[3]} \\ \mathbf{S}_{10}^{[3]} & \mathbf{S}_{11}^{[3]} \end{pmatrix}, \quad \mathbf{T}_1 = \begin{pmatrix} \mathbf{S}_{00}^{[3]} & \cdots & \mathbf{S}_{0(s+1)}^{[3]} \\ \mathbf{S}_{10}^{[3]} & \cdots & \mathbf{S}_{1(s+1)}^{[3]} \end{pmatrix},$$

which correspond with the  $\mathbf{S}_{11}^{[2]}$  and  $\mathbf{S}_1^{[2]}$ , respectively, in the ordinary GPCR method. The 1<sub>1</sub>th through 1 <sub>$r''_1$</sub> th principal components are found by solving the following eigenvalue problem:

$$\mathbf{T}_{11}^{-1/2} \mathbf{T}_1 \mathbf{T}'_1 \mathbf{T}_{11}^{-1/2} \mathbf{m} = \theta \mathbf{m}.$$

We obtain the eigenvalues  $\theta_1 \geq \dots \geq \theta_{r''_1}$  and the corresponding eigenvectors  $\mathbf{m}_1, \dots, \mathbf{m}_{r''_1}$ . Then  $\mathbf{D}_1 = (\mathbf{m}_1, \dots, \mathbf{m}_{r''_1})$  is obtained which is a  $(q''_0 + q''_1) \times r''_1$  matrix. The  $k_1$ th through  $k_{r''_k}$ th principal components are obtained from  $(\mathbf{M}_0, \mathbf{M}_k)$  ( $k = 2, \dots, s$ ). Let  $\mathbf{T}_{kj}$  and  $\mathbf{T}_k$  be

$$\mathbf{T}_{kj} = \begin{pmatrix} \mathbf{S}_{00}^{[3]} & \mathbf{S}_{0j}^{[3]} \\ \mathbf{S}_{k0}^{[3]} & \mathbf{S}_{kj}^{[3]} \end{pmatrix} \quad (j = 2, \dots, k), \quad \mathbf{T}_k = \begin{pmatrix} \mathbf{S}_{00}^{[3]} & \cdots & \mathbf{S}_{0(s+1)}^{[3]} \\ \mathbf{S}_{k0}^{[3]} & \cdots & \mathbf{S}_{k(s+1)}^{[3]} \end{pmatrix}.$$

We note that  $\mathbf{T}_{kk}$  and  $\mathbf{T}_k$  correspond with the  $\mathbf{S}_{kk}^{[2]}$  and  $\mathbf{S}_k^{[2]}$ , respectively, in the ordinary GPCR method. Using the same process as in the ordinary GPCR method, the  $k_1$ th through  $k_{r''_k}$ th principal components are obtained by maximizing  $\mathbf{d}' \mathbf{\Phi}_k \mathbf{T}_k \mathbf{T}'_k \mathbf{\Phi}_k \mathbf{d}$  under the conditions that  $\mathbf{d}' \mathbf{d} = 1$  and  $\mathbf{d} \in \mathbf{V}^\perp$ , where  $\mathbf{V} = \langle \mathbf{U}_1 \mathbf{d}_1, \dots, \mathbf{U}_1 \mathbf{d}_{1r_1}, \dots, \mathbf{U}_{k-1} \mathbf{d}_{(k-1)1}, \dots, \mathbf{U}_{k-1} \mathbf{d}_{(k-1)r_{(k-1)}} \rangle$ ,  $\mathbf{U}_j = \mathbf{T}_{kk}^{-1/2} \mathbf{T}_{kj} \mathbf{T}_{jj}^{-1/2}$ , and  $\mathbf{\Phi}_k$  is the projection matrix for  $\mathbf{V}^\perp$ . The maximization problem comes down to the following eigenvalue problem:

$$\mathbf{\Phi}_k \mathbf{T}_{kk}^{-1/2} \mathbf{T}_k \mathbf{T}'_k \mathbf{T}_{kk}^{-1/2} \mathbf{\Phi}_k \mathbf{m} = \theta \mathbf{m}.$$

By solving this eigenvalue problems, we obtain the eigenvalues  $\theta_{k_1} \geq \dots \geq \theta_{k_{r''_k}}$  and the corresponding eigenvectors  $\mathbf{m}_{k_1}, \dots, \mathbf{m}_{k_{r''_k}}$ . Then  $\mathbf{D}_k = (\mathbf{m}_{k_1}, \dots, \mathbf{m}_{k_{r''_k}})$ , which is a  $(q''_0 + q''_k) \times r''_k$  matrix. When we obtain  $\mathbf{D}_1, \dots, \mathbf{D}_s$  by using the same method as above, we obtain the estimation and prediction,  $\boldsymbol{\beta}$  and  $\mathbf{y}$ . Let  $\mathbf{W} = (\mathbf{M}_0, \dots, \mathbf{M}_s)$  and  $\mathbf{D} = (\mathbf{D}'_0, \mathbf{D}'_{1,\dots,s})'$ , where  $\mathbf{D}_0 = (\mathbf{D}_{01}, \dots, \mathbf{D}_{0s})$ ,

$$\mathbf{D}_{1,\dots,s} = \begin{pmatrix} \mathbf{D}_{11} & \mathbf{O}_{q''_1 \times r''_2} & \cdots & \mathbf{O}_{q''_1 \times r''_s} \\ \mathbf{O}_{q''_2 \times r''_1} & \mathbf{D}_{22} & \cdots & \mathbf{O}_{q''_2 \times r''_s} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{O}_{q''_s \times r''_1} & \mathbf{O}_{q''_s \times r''_2} & \cdots & \mathbf{D}_{ss} \end{pmatrix},$$

and  $\mathbf{D}_i = (\mathbf{D}_{0i}, \mathbf{D}_{ii})'$  for  $i = 1 \dots, s$ . By using these matrices, the estimator and predictor are obtained by  $\hat{\boldsymbol{\beta}}_{\text{MGPCR}} = (\mathbf{D}'\mathbf{W}'\mathbf{W}\mathbf{D})^{-1}\mathbf{D}'\mathbf{W}\mathbf{y}$  and  $\hat{\mathbf{y}}_{\text{MGPCR}} = \mathbf{W}\mathbf{D}\hat{\boldsymbol{\beta}}_{\text{MGPCR}}$ . In this method, we need to decide  $\mathbf{M}_0, \dots, \mathbf{M}_{s+1}$ ;  $q''_0, \dots, q''_{s+1}$  and the numbers of principal components  $r''_1, \dots, r''_s$ . When we use the CV or MCV method to decide these parameters, we need more computations than for the GPCR. In the present paper, we do not compare the MGPCR with other methods since we are interested in a practical system, and the MGPCR needs more processing time to determine each parameter.

## References

- [1] Hastie, T., Tibshirani, R. & Friedman, J., (2009). *The Elements of Statistical Learning* (2nd ed.), Springer Science+Business Media.
- [2] Jolliffe, I. T., (1972). Discarding variables in a principal components analysis I: Artificial data. *Journal of the Royal Statistical Society. Series C*, **21**, 160–173.
- [3] Jolliffe, I. T., (1986). *Principal Components Analysis*, Springer-Verlag, New York.
- [4] Mori, Y., Tanaka, Y. & Tarumi, T., (1997). Principal component analysis based on a subset of variables for qualitative data. *Data Science, Classification, and Related Methods* (Proceedings of IFCS-96), 547-554, Springer-Verlag.
- [5] Rao, C. R., (1964). The used and interpretation of principal component analysis in applied research. *Sankhya Series A*, **26**, 329–358.
- [6] Robert, P. & Escoufier, Y., (1976). A unifying tool for linear multivariate statistical methods: The RV-coefficient. *Journal of the Royal Statistical Society. Series C*, **25**, 257–265.
- [7] Srivastava, M. S., (2002). *Methods of Multivariate Statistics*, John Wiley & Sons.
- [8] Tanaka, Y. & Mori, Y., (1997). Principal component analysis based on a subset of variables: Variable selection and sensitivity analysis. *American Journal of Mathematical and Management Sciences.*, **17**, 61–89.