Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 32(6): 3758–3776.
DOI: 10.3934/era.2024171
Received: 29 January 2024
Revised: 27 May 2024
Accepted: 27 May 2024
Published: 07 June 2024

Research article

Hybrid principal component regression estimation in linear regression

Jian-Ying Rong[1,*] **and Xu-Qing Liu**[2,*]

[1] Department of Quality Education, Jiangsu Vocational College of Electronics and Information, Huai'an 223003, China

[2] Faculty of Mathematics and Physics, Huaiyin Institute of Technology, Huai'an 223003, China

* **Correspondence:** Email: rjy98@163.com, liuxuqing@hyit.edu.cn.

Abstract: In this paper, the principal component regression (PCR) estimators for regression parameters were studied in a linear regression model. After discussing the advantages and disadvantages of the classical PCR, we put forward three versions of hybrid PCR estimators. For the first two versions, we obtained the corresponding optimal solutions under the prediction error sum of squares (PRESS) criterion, while for the last one we offered two methods for obtaining the solution. In order to examine their practicality and generalizability, we considered two real-world examples and conducted a simulation study, which took into account varying degrees of multicollinearity. The numerical experiment revealed that the new estimators could substantially improve the least squares (LS) and classical PCR estimators under the PRESS criterion.

Keywords: hybrid PCR; linear regression; PCR; weighted PCR (WPCR); WPCR with nonnegative weights

1. Introduction

Linear models, as one of the core methods in classical statistics and machine learning, hold significant theoretical and practical importance [1]. Theoretical research on linear models highlights their interpretability, solvability, and a solid mathematical foundation, enabling a deeper understanding of the patterns underlying model predictions and providing foundations for the development of more advanced models as well as algorithms [2]. In practical applications, linear models are intuitive, easily comprehensible, and applicable to various tasks. They have achieved significant outcomes in domains like financial risk control and medical diagnosis [3]. Additionally, linear models bring the advantages of low computational complexity, suitability for large-scale datasets and even online learning tasks, regularization techniques to improve generalizability, and inherent feature selection capabilities. Hence, linear models possess high practical value in real-world applications.

Consider a linear regression model

$$\boldsymbol{y} = \beta_0\mathbf{1} + \boldsymbol{X\beta} + \boldsymbol{e}, \tag{1.1}$$

where $\boldsymbol{y} = (y_1, \cdots, y_n)'$ is a random vector of responses, $\boldsymbol{e} = (e_1, \cdots, e_n)'$ is the vector of errors with mean $\mathscr{E}(\boldsymbol{e}) = \mathbf{0}$ and covariance matrix $\mathscr{D}(\boldsymbol{e}) = \sigma^2\boldsymbol{I}_n$, $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)'$ with $\boldsymbol{x}_i = (x_{i1}, \cdots, x_{ip})'$ for $i = 1, \cdots, n$ is the regressor matrix of full column rank, the constant β_0, the vector of regression parameters $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$, and the error variance σ^2 are assumed to be unknown, $\mathbf{1}$ is a vector of ones with suitable orders, $\mathbf{0}$ is a vector or matrix of zeros with suitable orders, and $\boldsymbol{I}_n$ denotes the identity matrix of order n. In addition, assume $\mathbf{1} \notin \mathscr{R}(\boldsymbol{X})$, in which $\mathscr{R}(\boldsymbol{X})$ denotes the (column) range space of $\boldsymbol{X}$.

It is well known that the ordinary LS estimators for β_0 and $\boldsymbol{\beta}$ (denoted by $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}$, respectively) play an important role in parametric estimation theory, which can be expressed as the solution of the following regular equation

$$\begin{bmatrix} n & \mathbf{1}'\boldsymbol{X} \\ \boldsymbol{X}'\mathbf{1} & \boldsymbol{X}'\boldsymbol{X} \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \mathbf{1}'\boldsymbol{y} \\ \boldsymbol{X}'\boldsymbol{y} \end{bmatrix}, \tag{1.2}$$

However, when severe multicollinearity is present in the model (1.1), the LS estimator usually performs poorly under the mean squared error (MSE) criterion. The problem of multicollinearity usually occurs in the case that there is potentially high approximate correlation among the regressors, which can lead to unstable parameter estimation, increased variance of explanatory variables, and decreased reliability and interpretability of the model.

To overcome the problem of multicollinearity, various biased estimators for different models were put forward in the literature, such as the ordinary and generalized ridge regression estimators [4–12], and very recently, the PCR estimator [13], the Liu and Liu-type estimators [14, 15] and their improved versions [16, 17], and the double-k class estimators [18]. These biased estimators can locally improve the LS estimator by appropriately choosing the biasing parameters involved. Among them, the PCR estimator is of particular interest to us because of its geometric meaning and interpretation in trying to capture the essence of the model and its effectiveness in addressing multicollinearity and enhancing model stability. However, it involves dimensionality reduction, which may lead to information loss. While the amount of information loss can be customized by the user, it can also give rise to subsequent issues and challenges. In this paper, we analyze a shortcoming of the PCR estimator in detail and then put forward an improvement from the perspective of overcoming the model instability and inaccurate estimation caused by multicollinearity, while minimizing or even avoiding the loss of information carried by the data as much as possible.

The remainder of the paper is organized as follows. Section 2 briefly analyzes the classical PCR estimator. In Section 3, we discuss the motivation by exemplifying the advantages and disadvantages of PCR. We then propose three versions of hybrid PCR estimators and provide the corresponding optimal solutions under the PRESS criterion. In Section 4, we apply the theoretical results to two real examples and conduct a simulation study. Section 5 provides concluding remarks and two suggestions for the estimators' use.

2. Classical PCR estimation

In this section, we concisely describe the classical PCR estimation and discuss a potential flaw of it when used in practice. Centralize $\boldsymbol{X}$ as $\boldsymbol{X}_c = \boldsymbol{X} - \frac{1}{n}\mathbf{1}\mathbf{1}'\boldsymbol{X}$ such that $\mathbf{1}'\boldsymbol{X}_c = \mathbf{0}$. Pre-multiplying the two sides of (1.2) with the nonsingular partitioned matrix $\begin{bmatrix} 1 & \mathbf{0} \\ -\frac{1}{n}\boldsymbol{X}'\mathbf{1} & \boldsymbol{I}_n \end{bmatrix}$, we have the following equivalent regular equation

$$\begin{bmatrix} 1 & \mathbf{0} \\ -\frac{1}{n}\boldsymbol{X}'\mathbf{1} & \boldsymbol{I}_n \end{bmatrix} \begin{bmatrix} n & \mathbf{1}'\boldsymbol{X} \\ \boldsymbol{X}'\mathbf{1} & \boldsymbol{X}'\boldsymbol{X} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{n}\mathbf{1}'\boldsymbol{X} \\ \mathbf{0} & \boldsymbol{I}_n \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{n}\mathbf{1}'\boldsymbol{X} \\ \mathbf{0} & \boldsymbol{I}_n \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0} \\ -\frac{1}{n}\boldsymbol{X}'\mathbf{1} & \boldsymbol{I}_n \end{bmatrix} \begin{bmatrix} \mathbf{1}'\boldsymbol{y} \\ \boldsymbol{X}'\boldsymbol{y} \end{bmatrix}.$$

By direct operations, the LS estimators are given as

$$\begin{cases} \hat{\boldsymbol{\beta}} = (\boldsymbol{X}_c'\boldsymbol{X}_c)^{-1}\boldsymbol{X}_c'\boldsymbol{y} \\ \hat{\beta}_0 = \bar{y} - \frac{1}{n}\mathbf{1}'\boldsymbol{X}\hat{\boldsymbol{\beta}} = \bar{y} - \bar{\boldsymbol{x}}'\hat{\boldsymbol{\beta}} \end{cases} \tag{2.1}$$

in which $\bar{y} = \frac{1}{n}\boldsymbol{y}'\mathbf{1} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{\boldsymbol{x}} = \frac{1}{n}\boldsymbol{X}'\mathbf{1} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i$ denote the sample mean of the responses and that of the regressors, respectively. Also, (2.1) can be derived from the centeralized model of (1.1), $\boldsymbol{y} = \alpha_0\mathbf{1} + \boldsymbol{X}_c\boldsymbol{\beta} + \boldsymbol{e}$ with $\alpha_0 = \beta_0 + \bar{\boldsymbol{x}}'\boldsymbol{\beta}$.

When multicollinearity is present in the centralized model, $\boldsymbol{X}_c$ is ill-conditioned. For this case, $\hat{\beta}_0$ and $\hat{\boldsymbol{\beta}}$ can be improved by PCR estimators [13]. Let $\lambda_1 \geqslant \cdots \geqslant \lambda_p$ (> 0) be the eigenvalues of $\boldsymbol{X}_c'\boldsymbol{X}_c$, and $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_p$ be the corresponding standardized eigenvectors.

We set $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \cdots, \lambda_p)$, $\boldsymbol{Q} = (\boldsymbol{q}_1, \cdots, \boldsymbol{q}_p)$, $\boldsymbol{Z} = \boldsymbol{X}_c\boldsymbol{Q} \triangleq (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_p)$, and $\boldsymbol{\gamma} = \boldsymbol{Q}'\boldsymbol{\beta} \triangleq (\gamma_1, \cdots, \gamma_p)'$. It follows that the centralized model can be written as

$$\boldsymbol{y} \;=\; \alpha_0\mathbf{1} + \boldsymbol{X}_c\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{\beta} + \boldsymbol{e} \;=\; \alpha_0\mathbf{1} + \boldsymbol{Z}\boldsymbol{\gamma} + \boldsymbol{e} \;=\; \alpha_0\mathbf{1} + \sum_{j=1}^{p}\gamma_j\boldsymbol{z}_j + \boldsymbol{e}, \tag{2.2}$$

considering that $\boldsymbol{Q}$ is an orthogonal matrix. If $\boldsymbol{z}_{r+1}'\boldsymbol{z}_{r+1} = \lambda_{r+1} \approx 0$ holds for some r $(1 \leqslant r < p)$, the value of $\sum_{j=r+1}^{p}\gamma_j\boldsymbol{z}_j$ is close to $\mathbf{0}$, and therefore can be omitted approximately or merged into the intercept term, $\alpha_0\mathbf{1}$. The number r can be commonly determined by letting the cumulative percent $(\lambda_1 + \cdots + \lambda_r)/(\lambda_1 + \cdots + \lambda_p)$ be as large as possible, specifically, not less than 85%. In this sense, the canonical model (2.2) reduces to

$$\boldsymbol{y} \;\approx\; \alpha_0\mathbf{1} + \sum_{j=1}^{r}\gamma_j\boldsymbol{z}_j + \boldsymbol{e} \;\triangleq\; \alpha_0\mathbf{1} + \boldsymbol{Z}_1\boldsymbol{\gamma}_1 + \boldsymbol{e}, \tag{2.3}$$

with $\boldsymbol{Z}_1 = (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_r)$ and $\boldsymbol{\gamma}_1 = (\gamma_1, \cdots, \gamma_r)'$. That is, $\gamma_{r+1}, \cdots, \gamma_p$ are regarded (or estimated) as zeros. This means $\boldsymbol{\beta} = \boldsymbol{Q}\boldsymbol{\gamma} \approx \boldsymbol{Q}_1\boldsymbol{\gamma}_1$, with $\boldsymbol{Q}_1 = (\boldsymbol{q}_1, \cdots, \boldsymbol{q}_r)$. Set $\boldsymbol{\Lambda}_1 = \mathrm{diag}(\lambda_1, \cdots, \lambda_r)$. Imposing the LS principle on the reduced model (2.3), it gives the PCR estimators as

$$\begin{cases} \tilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{Q}\boldsymbol{\gamma}} = \boldsymbol{Q}\hat{\boldsymbol{\gamma}} \approx \boldsymbol{Q}_1\hat{\boldsymbol{\gamma}}_1 = \boldsymbol{Q}_1\boldsymbol{\Lambda}_1^{-1}\boldsymbol{Z}_1'\boldsymbol{y} \\ \tilde{\beta}_0 = \bar{y} - \bar{\boldsymbol{x}}'\tilde{\boldsymbol{\beta}} \end{cases} \tag{2.4}$$

3. Hybrid PCR estimation

In this section, we briefly discuss the limitations of the classical PCR estimator, which motivates us to define three hybrid PCR estimators. We then employ the PRESS criterion to obtain the optimal hybrid PCR estimators.

3.1. Motivation and definition

The classical PCR estimators given in (2.4) can improve the LS estimator by discarding the redundant part of the centralized regressor matrix. However, as we can see from the previous procedure, there may be some potential problems for the PCR estimators: i) The cumulative percent (85% or other values) is subjective; ii) A small cumulative percent can lead to too much loss of useful information; and iii) A large cumulative percent will produce estimators performing badly.

This can also be illustrated by the following two toy examples. One is:

$$\lambda_1 = 30, \lambda_2 = 13.3, \lambda_3 = \cdots = \lambda_8 = 1.1, \lambda_9 = 0.1,$$

and the other is with $\lambda_1 = 30, \lambda_2 = 13.3, \lambda_3 = 6.6, \lambda_4 = 0.1$. Both of them suffer from multicollinearity, since they have the large (and identical) condition number, 300. Clearly, for the former, choosing the first two principal components to estimate the regression parameters are reasonable because $\lambda_1 + \lambda_2 \geqslant 0.85\,(\lambda_1 + \cdots + \lambda_9)$ and all of $\lambda_3, \cdots, \lambda_8$ are very small relative to λ_1, while for the latter, it is undesirable to discard the third principal component although $\lambda_1 + \lambda_2 \geqslant 0.85\,(\lambda_1 + \cdots + \lambda_4)$.

To overcome the problems (ii) and (iii), one can use different cumulative percents ($\geqslant$ 85%, or $\geqslant$ 90%, or $\geqslant$ 95%, etc.) in different problems. However, this may lead to much more subjectivity and thus intensifying (i). An alternative way is to combine all possible PCR estimators in a suitable way such that the contribution of each PCR estimator can be automatically computed. This will be studied in the next section.

As illustrated by the second example in Section 2, the third principal component with contribution percent $\lambda_3/(\lambda_1 + \cdots + \lambda_4) = 6.6/(30 + 13.3 + 6.6 + 0.1) = 13.2\%$ should not be discarded directly, but should be used with an appropriate proportion. This can be done by first weighting each principal component and then estimating the parameters. This will yield a nonlinear estimator with respect to the weights, and thus lead to new difficulties in determining the values of the weights.

An alternative method is to linearly weight all of the PCR estimators. This leads to the following concept of hybrid PCR (HPCR) estimation:

Definition 1. *Denote the PCR estimator of $\boldsymbol{\beta}$ based on the first k principal components by $\tilde{\boldsymbol{\beta}}^{(k)}$. For any p constants $w_1, \cdots, w_p \in \mathbb{R}$, we call $\boldsymbol{\beta}^*_{\boldsymbol{w}} \triangleq \sum_{k=1}^{p} w_k \tilde{\boldsymbol{\beta}}^{(k)}$ and $\beta^*_{0,\boldsymbol{w}} \triangleq \bar{y} - \bar{\boldsymbol{x}}'\boldsymbol{\beta}^*_{\boldsymbol{w}}$ to be the HPCR estimators for $\boldsymbol{\beta}$ and β_0, respectively, with respect to $\boldsymbol{w} = (w_1, \cdots, w_p)'$.*

Clearly, $\boldsymbol{\beta}^*_{\boldsymbol{w}}$ and $\beta^*_{0,\boldsymbol{w}}$ reduce to the classical PCR estimators presented in (2.4) if taking $w_r = 1$ and $w_i = 0$ for any $i \neq r$. When taking $w_1 = \cdots = w_{p-1} = 0$ and $w_p = 1$, the LS estimators given in (2.1) are derived. Hence, Definition 1 gives a set of estimators including classical PCR and LS estimations.

For Definition 1, the problem is to determine the values of $\boldsymbol{w} = (w_1, \cdots, w_p)'$. A feasible method is to simply take w_i as the contribution percent of the i^{th} principal component. This means that the first PCR estimator gets the largest w_1, the second PCR estimator gets the second largest w_2, and so on. However, this may not be suitable in some situations.

For example, consider the model (1.1) with $\lambda_1 = 23.3, \lambda_2 = 20, \lambda_3 = 6.6$, and $\lambda_4 = 0.1$. In this example, the first PCR estimator only uses all of the $23.3/(23.3 + 20 + 6.6 + 0.1) = 46.6\%$ information about the regressors, while the second PCR estimator uses $(23.3+20)/(23.3+20+6.6+0.1) = 86.6\%$ out of all information. Therefore, the first PCR estimator is quite bad relative to the second PCR estimator, and thus it should not be given the largest weight in the HPCR estimator. In this sense, the selection of $\boldsymbol{w}$ is a key procedure in getting a fine HPCR estimator. In what follows, we provide a selection under the PRESS criterion.

3.2. PRESS criterion

To find an optimal HPCR estimator, we use the PRESS put forward by [19, 20] to measure how $\boldsymbol{w}$ influences on the predictive performance of $\boldsymbol{\beta}^*_{\boldsymbol{w}}$ and $\beta^*_{0,\boldsymbol{w}}$. We do not consider how $\boldsymbol{\beta}^*_{\boldsymbol{w}}$ and $\beta^*_{0,\boldsymbol{w}}$ are different from $\boldsymbol{\beta}$ and β_0, because multicollinearity causes the differences between the true and estimated values to be no longer true. For example, for a model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e$ with $x_3 \approx 2x_1 - 3x_2$, it follows that $y \approx \beta_0 + (\beta_1 + 2\beta_3)x_1 + (\beta_2 - 3\beta_3)x_2 + 0 \cdot x_3 + e$. This means that the estimators of $\beta_1 + 2\beta_3$ and $\beta_2 - 3\beta_3$ can be good enough to estimate β_1 and β_2.

Observe now the PRESS criterion. Let $\hat{\boldsymbol{\alpha}}_{-i}$ be denoted as an estimator of $\boldsymbol{\alpha}$ based on all data points except the i^{th} one. With this notation (and some other similar ones), the PRESS statistic of the LS estimators, that of classical and hybrid PCR estimators, can be expressed as follows:

$$PRESS\left(\hat{\beta}_0, \hat{\boldsymbol{\beta}}; \beta_0, \boldsymbol{\beta}\right) = \sum_{i=1}^{n} \left[y_i - \left(\hat{\beta}_{0;-i} + \boldsymbol{x}_i' \hat{\boldsymbol{\beta}}_{-i}\right)\right]^2, \tag{3.1}$$

$$PRESS\left(\tilde{\beta}_0, \tilde{\boldsymbol{\beta}}; \beta_0, \boldsymbol{\beta}\right) = \sum_{i=1}^{n} \left[y_i - \left(\tilde{\beta}_{0;-i} + \boldsymbol{x}_i' \tilde{\boldsymbol{\beta}}_{-i}\right)\right]^2, \tag{3.2}$$

$$PRESS\left(\beta^*_{0,\boldsymbol{w}}, \boldsymbol{\beta}^*_{\boldsymbol{w}}; \beta_0, \boldsymbol{\beta}\right) = \sum_{i=1}^{n} \left[y_i - \left(\beta^*_{0,\boldsymbol{w};-i} + \boldsymbol{x}_i' \boldsymbol{\beta}^*_{\boldsymbol{w};-i}\right)\right]^2. \tag{3.3}$$

Note that the expression of $PRESS\left(\beta^*_{0,\boldsymbol{w}}, \boldsymbol{\beta}^*_{\boldsymbol{w}}; \beta_0, \boldsymbol{\beta}\right)$ contains $\boldsymbol{w}$, so the PRESS criterion imposed on hybrid PCR estimators is to find $\boldsymbol{w}$ such that $PRESS\left(\beta^*_{0,\boldsymbol{w}}, \boldsymbol{\beta}^*_{\boldsymbol{w}}; \beta_0, \boldsymbol{\beta}\right)$ is minimized.

The PRESS statistic is seemingly similar to the sum of the residuals, $\sum_{i=1}^{n} \left[y_i - \left(\hat{\beta}_0 + \boldsymbol{x}_i' \hat{\boldsymbol{\beta}}\right)\right]^2$, of the original LS principle. However, PRESS is essentially different from LS, because it avoids granting an observation (data point) to play a dual role in simultaneously fitting old observations and predicting new observations, and it can facilitate exploiting the predictive performance of estimation. This is why we consider using the PRESS criterion.

3.3. Optimal HPCR estimators under the PRESS criterion

To find the PRESS-optimal HPCR estimators, we rewrite (3.3) as follows:

$$\begin{aligned}
PRESS\left(\beta^*_{0,\boldsymbol{w}}, \boldsymbol{\beta}^*_{\boldsymbol{w}}; \beta_0, \boldsymbol{\beta}\right) &= \sum_{i=1}^{n} \left[y_i - \left(\bar{y}_{-i} - \bar{\boldsymbol{x}}_{-i}' \boldsymbol{\beta}^*_{\boldsymbol{w};-i} + \boldsymbol{x}_i' \boldsymbol{\beta}^*_{\boldsymbol{w};-i}\right)\right]^2 \\
&= \sum_{i=1}^{n} \left[\left(y_i - \bar{y}_{-i}\right) - \left(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_{-i}\right)' \sum_{k=1}^{p} w_k \tilde{\boldsymbol{\beta}}^{(k)}_{-i}\right]^2 \\
&= \sum_{i=1}^{n} \left\{\left(y_i - \bar{y}_{-i}\right) - \sum_{k=1}^{p} \left[\left(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_{-i}\right)' \tilde{\boldsymbol{\beta}}^{(k)}_{-i}\right] w_k\right\}^2 \\
&= \left(\frac{n}{n-1}\right)^2 \sum_{i=1}^{n} \left\{\left(y_i - \bar{y}\right) - \sum_{k=1}^{p} \left[\left(\boldsymbol{x}_i - \bar{\boldsymbol{x}}\right)' \tilde{\boldsymbol{\beta}}^{(k)}_{-i}\right] w_k\right\}^2,
\end{aligned}$$

in view of the algebraic facts that $y_i - \bar{y}_{-i} = \frac{n}{n-1}\left(y_i - \bar{y}\right)$ and $\boldsymbol{x}_i - \bar{\boldsymbol{x}}_{-i} = \frac{n}{n-1}\left(\boldsymbol{x}_i - \bar{\boldsymbol{x}}\right)$.

Denote $\boldsymbol{y}_c = \left(y_1 - \bar{y}, \cdots, y_n - \bar{y}\right)'$ and $\boldsymbol{A} = (a_{ik})_{n \times p}$, with $a_{ik} \triangleq \left(\boldsymbol{x}_i - \bar{\boldsymbol{x}}\right)' \tilde{\boldsymbol{\beta}}^{(k)}_{-i}$ for $i = 1, \cdots, n$ and $k = 1, \cdots, p$. With these notations, we have

$$PRESS\left(\beta^*_{0,\boldsymbol{w}}, \boldsymbol{\beta}^*_{\boldsymbol{w}}; \beta_0, \boldsymbol{\beta}\right) \propto \left(\boldsymbol{y}_c - \boldsymbol{A}\boldsymbol{w}\right)'\left(\boldsymbol{y}_c - \boldsymbol{A}\boldsymbol{w}\right) = \boldsymbol{y}_c'\boldsymbol{y}_c - 2\boldsymbol{y}_c'\boldsymbol{A}\boldsymbol{w} + \boldsymbol{w}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{w}. \tag{3.4}$$

If no constraints are imposed on $w_1, \cdots, w_p$, it is clear that minimizing $PRESS\left(\beta^*_{0,\boldsymbol{w}}, \boldsymbol{\beta}^*_{\boldsymbol{w}}; \beta_0, \boldsymbol{\beta}\right)$ gives

$$\boldsymbol{w}^* = \left(\boldsymbol{A}'\boldsymbol{A}\right)^- \boldsymbol{A}'\boldsymbol{y}_c \triangleq \left(w^*_1, \cdots, w^*_p\right)'. \tag{3.5}$$

This further implies

$$\begin{aligned} PRESS\left(\beta^*_{0,\boldsymbol{w}^*}, \boldsymbol{\beta}^*_{\boldsymbol{w}^*}; \beta_0, \boldsymbol{\beta}\right) &= \left(\frac{n}{n-1}\right)^2 \left[\boldsymbol{y}_c - \boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{A}\right)^- \boldsymbol{A}'\boldsymbol{y}\right]' \left[\boldsymbol{y}_c - \boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{A}\right)^- \boldsymbol{A}'\boldsymbol{y}\right] \\ &= \left(\frac{n}{n-1}\right)^2 \boldsymbol{y}'_c\left(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{A}}\right)\boldsymbol{y}_c, \end{aligned} \tag{3.6}$$

with $\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^-\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^+$ being the orthogonal projection matrix [1, p. 24] over the (column) range space, $\mathscr{R}(\boldsymbol{A})$, where $\boldsymbol{A}^-$ is any 1-inverse, and $\boldsymbol{A}^+$ is the unique Moore-Penrose inverse (Definition 2.2 of [1]) of $\boldsymbol{A}$.

According to the above derivations, we can present the following theorem:

Theorem 1. Let $\boldsymbol{w}^* = \left(w^*_1, \cdots, w^*_p\right)'$ be defined in (3.5). Then, $\boldsymbol{\beta}^*_{\boldsymbol{w}^*} = \sum_{k=1}^p w^*_k \tilde{\boldsymbol{\beta}}^{(k)}$ and $\beta^*_{0,\boldsymbol{w}^*} = \bar{y} - \bar{\boldsymbol{x}}'\boldsymbol{\beta}^*_{\boldsymbol{w}^*}$ have the minimal PRESS value presented in (3.6) in all HPCR estimators.

This theorem concludes how to choose $\boldsymbol{w}$ under the PRESS criterion to get a fine HPCR estimator. As seen, if the matrix $\boldsymbol{A}$ is of full column rank, $\boldsymbol{w}^*$ is unique; otherwise, $\boldsymbol{w}^*$ changes along with different selections of the generalized inverse of $\boldsymbol{A}$. For convenience, we will always use the Moore-Penrose inverse, $\boldsymbol{A}^+$, in the simulation study.

Computationally, in the case that both of the matrices $\boldsymbol{X}$ and $\boldsymbol{A}$ are of full column rank, $\boldsymbol{A}$ is usually more ill-conditioned than $\boldsymbol{X}$. Although we cannot prove this result theoretically, the simulation study will show this to us. The major reason may be that $\boldsymbol{A}$ derives from some PCR estimators consisting of too many minor principal components. A potential solution is to discard the last several PCR estimators, which contain one or more principal components with a too-small individual percentage (such as 5% and even smaller) of variance, when using HPCR estimators. Specifically, letting $K \in \{1, \cdots, p-1\}$ satisfy

$$\frac{\lambda_K}{\lambda_1 + \cdots + \lambda_p} \geqslant 5\% > \frac{\lambda_{K+1}}{\lambda_1 + \cdots + \lambda_p} \quad \text{or} \quad \frac{\lambda_1 + \cdots + \lambda_{K-1}}{\lambda_1 + \cdots + \lambda_p} < 85\% \leqslant \frac{\lambda_1 + \cdots + \lambda_K}{\lambda_1 + \cdots + \lambda_p},$$

the HPCR estimators that Definition 1 presents can be modified as $\boldsymbol{\beta}^{**}_{\boldsymbol{w}} \triangleq \sum_{k=1}^K w_k \tilde{\boldsymbol{\beta}}^{(k)}$ and $\beta^{**}_{0,\boldsymbol{w}} \triangleq \bar{y} - \bar{\boldsymbol{x}}'\boldsymbol{\beta}^{**}_{\boldsymbol{w}}$, with $\boldsymbol{w} = (w_1, \cdots, w_K)'$. That is, we use only the first K PCR estimators to get the hybrid version. Under this modification, the PRESS-optimal selection for $w_1, \cdots, w_K$ can be obtained in a similar fashion. The details are omitted here.

3.4. Optimal WPCR estimators

Now, we assume $w_1, \cdots, w_p$ are weights, satisfying $\sum_{k=1}^p w_k = \boldsymbol{1}'\boldsymbol{w} = 1$. In this case, we call $\boldsymbol{\beta}^*_{\boldsymbol{w}}$ and $\beta^*_{0,\boldsymbol{w}}$ the WPCR estimators. Here, we note that, similar to the ordinary HPCR estimators, WPCR estimators also do not require $w_1, \cdots, w_p$ to take nonnegative values, because a negative w_i implies the i^{th} PCR estimator may produce some opposite estimates for the corresponding parameters to other PCR estimators, and the negativity of w_i can offset such effects in a way. Then, the problem of finding optimal WPCR estimators under the PRESS criterion is equivalent to solving the optimization problem

$$\begin{cases} \min & \boldsymbol{y}'_c\boldsymbol{y}_c - 2\boldsymbol{y}'_c\boldsymbol{A}\boldsymbol{w} + \boldsymbol{w}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{w} \\ \text{s.t.} & \boldsymbol{1}'\boldsymbol{w} = 1 \end{cases} \tag{3.7}$$

To solve (3.7), we denote the Lagrange function by $\mathrm{L}(\boldsymbol{w},\ell)=\boldsymbol{y}_c'\boldsymbol{y}_c-2\boldsymbol{y}_c'\boldsymbol{A}\boldsymbol{w}+\boldsymbol{w}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{w}+2\ell(\boldsymbol{1}'\boldsymbol{w}-1)$, in which ℓ is the Lagrange multiplier. By the formulas for partial derivatives of matrix functions [1, pp. 38–47], we obtain the following matrix equations:

$$\begin{cases}\frac{\partial \mathrm{L}(\boldsymbol{w},\ell)}{\partial \boldsymbol{w}}=\boldsymbol{0}-2\boldsymbol{A}'\boldsymbol{y}_c+2\boldsymbol{A}'\boldsymbol{A}\boldsymbol{w}+2\ell\boldsymbol{1}\triangleq\boldsymbol{0}\\ \frac{\partial \mathrm{L}(\boldsymbol{w},\ell)}{\partial \ell}=2(\boldsymbol{1}'\boldsymbol{w}-1)\triangleq 0\end{cases}$$

Equivalently, we have the following constrained regular equation:

$$\begin{bmatrix}\boldsymbol{A}'\boldsymbol{A} & \boldsymbol{1}\\ \boldsymbol{1}' & 0\end{bmatrix}\begin{bmatrix}\boldsymbol{w}\\ \ell\end{bmatrix}=\begin{bmatrix}\boldsymbol{A}'\boldsymbol{y}_c\\ 1\end{bmatrix}. \tag{3.8}$$

Note here that $\boldsymbol{A}'\boldsymbol{A}$ is symmetric and nonnegative definite. In what follows, we show the Eq (3.8) is consistent. In fact, as proven by [21], it can be shown that

$$\mathscr{R}\begin{pmatrix}\boldsymbol{A}'\boldsymbol{A} & \boldsymbol{1}\\ \boldsymbol{1}' & 0\end{pmatrix}=\mathscr{R}\begin{pmatrix}\boldsymbol{A}' & \boldsymbol{1} & \boldsymbol{0}\\ \boldsymbol{0} & 0 & 1\end{pmatrix}.$$

Hence, we get

$$\begin{bmatrix}\boldsymbol{A}'\boldsymbol{y}_c\\ 1\end{bmatrix}=\begin{bmatrix}\boldsymbol{A}' & \boldsymbol{0}\\ \boldsymbol{0} & 1\end{bmatrix}\begin{bmatrix}\boldsymbol{y}_c\\ 1\end{bmatrix}\in\mathscr{R}\begin{pmatrix}\boldsymbol{A}' & \boldsymbol{0}\\ \boldsymbol{0} & 1\end{pmatrix}\subseteq\mathscr{R}\begin{pmatrix}\boldsymbol{A}' & \boldsymbol{1} & \boldsymbol{0}\\ \boldsymbol{0} & 0 & 1\end{pmatrix}=\mathscr{R}\begin{pmatrix}\boldsymbol{A}'\boldsymbol{A} & \boldsymbol{1}\\ \boldsymbol{1}' & 0\end{pmatrix}.$$

This shows the consistency of Eq (3.8). Using the formula for the generalized inverse (see Theorem 2.6 of [1]) of a partitioned matrix that

$$\begin{bmatrix}\boldsymbol{S} & \boldsymbol{L}\\ \boldsymbol{L}' & \boldsymbol{0}\end{bmatrix}^-=\begin{bmatrix}\boldsymbol{T}^--\boldsymbol{T}^-\boldsymbol{L}\boldsymbol{Q}^-\boldsymbol{L}'\boldsymbol{T}^- & \boldsymbol{T}^-\boldsymbol{L}\boldsymbol{Q}^-\\ \boldsymbol{Q}^-\boldsymbol{L}'\boldsymbol{T}^- & \boldsymbol{Q}^-\boldsymbol{Q}-\boldsymbol{Q}^-\end{bmatrix},$$

in which $\boldsymbol{T}=\boldsymbol{S}+\boldsymbol{L}\boldsymbol{L}'$ and $\boldsymbol{Q}=\boldsymbol{L}'\boldsymbol{T}^-\boldsymbol{L}$ with $\boldsymbol{S}$ being symmetric and nonnegative definite, we have

$$\begin{bmatrix}\boldsymbol{A}'\boldsymbol{A} & \boldsymbol{1}\\ \boldsymbol{1}' & 0\end{bmatrix}^-=\begin{bmatrix}\boldsymbol{T}^--\boldsymbol{T}^-\boldsymbol{1}(\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1})^{-1}\boldsymbol{1}'\boldsymbol{T}^- & \boldsymbol{T}^-\boldsymbol{1}(\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1})^{-1}\\ (\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1})^{-1}\boldsymbol{1}'\boldsymbol{T}^- & 1-(\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1})^{-1}\end{bmatrix},$$

with $\boldsymbol{T}=\boldsymbol{A}'\boldsymbol{A}+\boldsymbol{1}\boldsymbol{1}'$. Here, $\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1}\neq 0$, and this is an algebraic fact explained in what follows: First of all, we note $\boldsymbol{1}\in\mathscr{R}(\boldsymbol{1})\subseteq\mathscr{R}([\boldsymbol{A}',\boldsymbol{1}])=\mathscr{R}([\boldsymbol{A}',\boldsymbol{1}][\boldsymbol{A}',\boldsymbol{1}]')=\mathscr{R}(\boldsymbol{A}'\boldsymbol{A}+\boldsymbol{1}\boldsymbol{1}')=\mathscr{R}(\boldsymbol{T})$, in which $\mathscr{R}(\cdot)$ denotes the range space. This implies: i) the value of $\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1}$ is independent of the selection of $\boldsymbol{T}^-$, and therefore $\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1}=\boldsymbol{1}'\boldsymbol{T}^+\boldsymbol{1}$; and ii) $\boldsymbol{P}_{\boldsymbol{T}}\boldsymbol{1}=\boldsymbol{1}$.

Now, we prove $\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1}\neq 0$ holds. Suppose $\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1}=0$. Combined with the fact that $\boldsymbol{T}$ is symmetric and nonnegative definite, we obtain $\boldsymbol{1}'\boldsymbol{T}^+\boldsymbol{1}=\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1}=0\Rightarrow\boldsymbol{T}^+\boldsymbol{1}=\boldsymbol{0}\Rightarrow\boldsymbol{1}=\boldsymbol{P}_{\boldsymbol{T}}\boldsymbol{1}=\boldsymbol{T}\boldsymbol{T}^+\boldsymbol{1}=\boldsymbol{0}$. This contradicts with "$\boldsymbol{1}\neq\boldsymbol{0}$", so we must have $\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1}\neq 0$. Therefore,

$$\begin{bmatrix}\boldsymbol{w}\\ \ell\end{bmatrix}=\begin{bmatrix}\boldsymbol{A}'\boldsymbol{A} & \boldsymbol{1}\\ \boldsymbol{1}' & 0\end{bmatrix}^-\begin{bmatrix}\boldsymbol{A}'\boldsymbol{y}_c\\ 1\end{bmatrix}=\begin{bmatrix}\boldsymbol{T}^--\boldsymbol{T}^-\boldsymbol{1}(\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1})^{-1}\boldsymbol{1}'\boldsymbol{T}^- & \boldsymbol{T}^-\boldsymbol{1}(\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1})^{-1}\\ (\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1})^{-1}\boldsymbol{1}'\boldsymbol{T}^- & 1-(\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1})^{-1}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}'\boldsymbol{y}_c\\ 1\end{bmatrix}.$$

Further, the $\boldsymbol{w}$-solution of (3.8) can be expressed as

$$\boldsymbol{w}^{**}=\left(\boldsymbol{T}^--\frac{\boldsymbol{T}^-\boldsymbol{1}\boldsymbol{1}'\boldsymbol{T}^-}{\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1}}\right)\boldsymbol{A}'\boldsymbol{y}_c+\frac{\boldsymbol{T}^-\boldsymbol{1}}{\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1}}=\boldsymbol{T}^-\boldsymbol{A}'\boldsymbol{y}_c+\frac{1-\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{A}'\boldsymbol{y}_c}{\boldsymbol{1}'\boldsymbol{T}^-\boldsymbol{1}}\boldsymbol{T}^-\boldsymbol{1}\triangleq\left(w_1^{**},\cdots,w_p^{**}\right)'. \tag{3.9}$$

Note that both $\mathbf{1}'\boldsymbol{T}^{-}\mathbf{1}$ and $\mathbf{1}'\boldsymbol{T}^{-}\boldsymbol{A}'$ are invariant with respect to all generalized inverses of $\boldsymbol{T}$, since $\mathbf{1} \in \mathscr{R}(\boldsymbol{T})$ and $\mathscr{R}(\boldsymbol{A}') \subseteq \mathscr{R}(\boldsymbol{T})$. Clearly, $\sum_{k=1}^{p} w_k^{**} = \mathbf{1}'\boldsymbol{w}^{**} = 1$. Recalling that the objective function of (3.7) is quadratic with respect to $\boldsymbol{w}$, this gives the globally optimal WPCR estimators under the PRESS criterion. The result is summarized in the following theorem:

Theorem 2. Let $\boldsymbol{w}^{**} = \left(w_1^{**}, \cdots, w_p^{**}\right)'$ be defined in (3.9). Then,

$$\boldsymbol{\beta}^*_{\boldsymbol{w}^{**}} = \sum_{k=1}^{p} w_k^{**}\tilde{\boldsymbol{\beta}}^{(k)} \text{ and } \beta^*_{0,\boldsymbol{w}^{**}} = \bar{y} - \bar{\boldsymbol{x}}'\boldsymbol{\beta}^*_{\boldsymbol{w}^{**}}$$

have the minimal PRESS value in all of the WPCR estimators.

As Theorem 1 does, Theorem 2 also provides us with the method of choosing the weights to get the optimal WPCR estimators, $\boldsymbol{\beta}^*_{\boldsymbol{w}^{**}}$ and $\beta^*_{0,\boldsymbol{w}^{**}}$, under the PRESS criterion. Further, if the matrix $\boldsymbol{T}$ is of full column rank, $\boldsymbol{w}^{**}$ is unique; otherwise, $\boldsymbol{w}^{**}$ changes along with $\boldsymbol{T}$. In the simulation study, we will always use the Moore-Penrose inverse, $\boldsymbol{T}^{+}$, when considering $\boldsymbol{\beta}^*_{\boldsymbol{w}^{**}}$ and $\beta^*_{0,\boldsymbol{w}^{**}}$. Note that, in any case, the minimal PRESS value remains unchanged.

3.5. Optimal WPCR estimators with nonnegative weights

The above two subsections obtain optimal HPCR and WPCR estimators, respectively. Finally, we assume constants $w_1, \cdots, w_p$ are weights (that is, $\sum_{k=1}^{p} w_k = 1$) and each weighting constant is nonnegative. In this case, we call $\boldsymbol{\beta}^*_{\boldsymbol{w}}$ and $\beta^*_{0,\boldsymbol{w}}$ the WPCR estimators with nonnegative weights (WnnPCR estimators). That is, we need to solve the following quadratic programming (QP) problem

$$\begin{cases} \min & \boldsymbol{y}_c'\boldsymbol{y}_c - 2\boldsymbol{y}_c'\boldsymbol{A}\boldsymbol{w} + \boldsymbol{w}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{w} \\ \text{s.t.} & \mathbf{1}'\boldsymbol{w} = 1 \text{ and } \boldsymbol{w} \geqslant \mathbf{0} \end{cases} \tag{3.10}$$

Problem (3.10) can be solved by the commonly used procedure of quadratic programming in various mathematical softwares. To improve the performance, we take $\boldsymbol{w}_{+}^{**} \triangleq \left(w_{1;+}^{**}, \cdots, w_{p;+}^{**}\right)'$ as the initial value of the search, in which

$$w_{i;+}^{**} = \frac{u_i^{**}}{\sum_{j=1}^{p} u_j^{**}},$$

with $u_i^{**} = \max\{w_i^{**}, 0\}$, for $i = 1, \cdots, p$. Here, $w_1^{**}, \cdots, w_p^{**}$ are defined in (3.9).

In what follows, we give a procedure of getting an approximate solution of the QP problem (3.10). Let $\boldsymbol{I}$ be a subset of $\{1, \cdots, p\}$, and we denote the following QP problem as QP($\boldsymbol{I}$):

$$\begin{cases} \min & \boldsymbol{y}_c'\boldsymbol{y}_c - 2\boldsymbol{y}_c'\boldsymbol{A}\boldsymbol{w} + \boldsymbol{w}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{w} \\ \text{s.t.} & \mathbf{1}'\boldsymbol{w} = 1 \text{ and } w_i = 0 \ (\forall i \in \boldsymbol{I}) \end{cases} \tag{3.11}$$

We note here that this problem has the same structure as (3.7), because the constraint $w_i = 0$ with $i \in \boldsymbol{I}$ renders the reduction of matrix $\boldsymbol{A}$ in (3.11) to a sub-matrix consisting of the columns except those in $\boldsymbol{I}$. Then, the approximate solution of (3.10) can be obtained by the following steps:

Step 1: Initialize $\boldsymbol{I}^{(k)} = \emptyset$ and $k = 0$.

Step 2: Use (3.9) to get a solution of $\mathrm{QP}\big(\boldsymbol{I}^{(k)}\big)$, namely $\boldsymbol{w}^{(k)} \triangleq \big(w_1^{(k)}, \cdots, w_p^{(k)}\big)'$. We set

$$\boldsymbol{J}^{(k)} = \left\{ j \,\middle|\, w_j^{(k)} < 0, j = 1, \cdots, p \right\}.$$

Step 3: If $\boldsymbol{J}^{(k)} \neq \emptyset$, solve $\mathrm{QP}\big(\boldsymbol{I}^{(k)} \cup \{j\}\big)$ for every $j \in \boldsymbol{J}^{(k)}$, find $j_{\min}$ which minimizes the QP objectives, set

$$\boldsymbol{I}^{(k+1)} \leftarrow \boldsymbol{I}^{(k)} \cup \{j_{\min}\}$$

and $k \leftarrow k + 1$, and then go to Step 2. Otherwise, return to the approximate solution of the QP problem (3.10), $\boldsymbol{w}^{***} \triangleq \boldsymbol{w}^{(k)}$.

This procedure modifies negative weights as 0 stepwise. In the whole process, all calculations can be theoretically performed. Therefore, it is essentially different from the solution derived by any mathematical software, when nonnegativity is required for weights.

We mention here that, as explained after Definition 1, LS and PCR are two special cases of HPCR (as well as WPCR and WnnPCR), so the optimal HPCR/WPCR/WnnPCR estimators will always perform better than LS/PCR theoretically in the PRESS sense.

4. Numerical study

In this section, we first apply the theoretical results to two real examples, namely Hald data [22] and Acetylene data [23], to preliminarily observe their performance. To investigate the numerical performance of classical and hybrid PCR in detail, we then conduct a simulation study to observe the changes in the PRESS values of the estimators under various degrees of multicollinearity, and analyze the potential reasons behind the observations.

4.1. Real examples

The Hald dataset [22] uses the heat of hardening after 180 days as the response and four ingredients as regressors, while the Acetylene dataset [23] uses the reactor temperature, rate of H_2 to n-heptane, and contact time as regressors and conversion of n-heptane to acetylene (%) as the response. Under the model (1.1) with $p = 4$ for Hald and $p = 3$ for Acetylene, the condition numbers for the regressor matrices are 20.5846 and 36935.9119, so these two datasets represent moderate multicollinearity and severe multicollinearity, respectively.

By direct computations, we obtain the PRESS values and then list them into Tables 1 and 2. The results reveal that:

- Regardless of the severity of multicollinearity, the estimators (HPCR/WPCR/WnnPCR) consistently yield lower PRESS compared to LS and PCR. As a result, the new estimators can be considered as competitive biased estimators in practical applications.
- When the condition number of the regressor matrix is not excessively high, PCR tends to eliminate valuable information due to its inherent construction features, leading to a higher value of PRESS. However, when the condition number of the regression matrix is extremely high, PCR usually performs relatively well.

- For the Hald data, the PRESS value of the classic PCR estimator unexpectedly exceeds that of the LS estimator and three hybrid PCR estimators. After careful checking, we find that PCR uses only the first principal component (contributing 86.60%) to estimate parameters, while the information carried by the second and third principal components (contributing 11.29 and 2.07%, respectively) is directly discarded! Furthermore, we find that, as one of the hybrid PCR estimators, WnnPCR nearly gives a 100% (to be more specifically, it is 99.9999999995337%) proportion to the estimator based on the first three principal components. This means that the WnnPCR estimator for Hald data is very close to the one constructed by the first three principal components, retaining 86.60 + 11.29 + 2.07 = 99.96% of all information, rather than just retaining 85% of the information in the traditional sense.
- For the Acetylene data, the situation is different. As seen, the WnnPCR estimator is composed of the first two PCR estimators, with weights 51.45 and 48.55%, respectively. Note that the contribution rate of the first principal component is 99.53%, while that of the second is only 0.47%. Therefore, the WnnPCR assigns a weight slightly lower than 50% to the second PCR estimator. Maybe this is just an attempt to extract as much useful information as possible from the information carried by the second principal component, which contributes 0.47% only.

Table 1. Estimates of the parameters and the PRESS values with respect to the Hald data, in which the weights of HPCR, WPCR, and WnnPCR are (−0.0447, −0.0367, 2.3798, −1.2986), (−0.0445, −0.0366, 2.3803, −1.2991), and (0.0000, 0.0000, 1.0000, 0.0000), respectively.

Parameter	LS	PCR	HPCR	WPCR	WnnPCR
β_0	62.4054	89.3820	176.9994	177.0226	111.4759
β_1	1.5511	0.0375	0.4240	0.4238	1.0350
β_2	0.5102	0.3757	−0.6765	−0.6766	0.0073
β_3	0.1019	−0.0161	−1.1140	−1.1142	−0.4237
β_4	−0.1441	−0.4047	−1.3018	−1.3021	−0.6379
PRESS	110.3466	1095.7010	85.2456	85.2457	93.4901

Table 2. Estimates of the parameters and the PRESS values with respect to the Acetylene data, in which the weights of HPCR, WPCR, and WnnPCR are (0.3552, 11.7020, −11.0943), (0.4062, 11.5567, −10.9629), and (0.5145, 0.4855, 0.0000), respectively.

Parameter	LS	PCR	HPCR	WPCR	WnnPCR
β_0	−121.2696	−133.0231	−229.8512	−234.9098	−131.8903
β_1	0.1269	0.1395	0.2098	0.2141	0.1368
β_2	0.3482	0.0022	0.2463	0.2412	0.1716
β_3	−19.0217	−0.0000	211.0306	208.5313	−0.0000
PRESS	336.2955	311.4937	175.9425	178.8785	291.1463

4.2. Simulation

This subsection makes a short simulation study to examine the numerical performance of LS and classical/hybrid PCR estimators for the model (1.1). In this study, we take n = 30, 70, 100, 200 and $p = 3,\ 6,\ 9$. For each case of p, take σ from $\{0.75,\ 0.25\}$. The explanatory variables are generated by using the simulation procedure suggested by McDonald and Galarneau [24]:

$$x_{ij} = (1-\rho^2)^{1/2}\zeta_{ij} + \rho\zeta_{i0}, \quad i = 1, \cdots, n, \quad j = 1, \cdots, p,$$

where ζ_{ij}'s are independent standard normal pseudo-random numbers, and ρ^2 is the correlation between any two explanatory variables. To see how multicollinearity influences the performance, we take ρ as 0.5, 0.9, 0.999, and 0.99999, respectively, to get regressor matrices with different condition numbers, from small to large. In addition, for each case, we randomly generate β_0 and $\boldsymbol{\beta}$ from the interval $[-5, 5]$. After that, we create a pseudo observation to compute the PRESS values of the five estimates (including LS, PCR with cumulative percent not less than 85%, HPCR, WPCR, and weighted PCR with nonnegative weights (WnnPCR). 100 runs are then performed and averaged for each case.

The results are computed and presented in Tables A1–A4 in the Appendix. By the tables, it follows that: (i) PCR can improve LS only when explanatory variables are highly correlated, and the degree of improvement depends on the error variance σ^2. In particular, when $\sigma^2 = 0.75^2$, PCR has smaller PRESS values than LS if ρ takes either 0.999 or 0.99999; when $\sigma^2 = 0.25^2$, PCR cannot improve LS unless $\rho = 0.99999$. Especially, in the case of $\rho = 0.5$ or 0.9, PCR performs very badly. (ii) Each of HPCR, WPCR, and WnnPCR improves LS and PCR substantially because, in any case, these three estimators have far smaller PRESS values than LS/PCR estimators. Naturally, HPCR performs the best, since the values of $\boldsymbol{w}$ can be selected from a wider range. (iii) The degree of the improvement of HPCR/WPCR/WnnPCR over LS/PCR depends on p and ρ. Specifically, the larger p is, the higher the degree is; and the lager ρ is, the higher the degree is. (iv) All estimators can be computed efficiently.

In view of the fact that LS can be regarded as a special PCR with all principal components, PCR can perform the same theoretically as LS if taking the cumulative percent as 100%. However, in this case, PCR fails to deal with multicollinearity. Therefore, HPCR/WPCR/WnnPCR can be a desirable remedying procedure, because these estimators collect information carried by all possible PCR estimators in an efficient way.

4.3. Discussion

It is well known that when multicollinearity is severe, the LS estimator performs poorly under the MSE criterion. However, as shown by the two real examples and the simulation study, the LS estimator seems to be relatively robust under the PRESS criterion. Although it is only slightly worse than the three newly proposed HPCR estimators, it is not to the extent of being surprising.

Why is this?

In fact, this is directly related to the nature of the MSE and PRESS criteria. MSE measures the difference between the regression parameters and their estimates, while PRESS considers the contribution of each observation point, rather than the direct difference between parameters and estimates. Although the LS estimator appears to be only slightly worse than HPCR estimators in the sense of PRESS, this slight difference has indicated a substantial improvement of HPCR over LS.

On the other hand, we also note that under severe multicollinearity, the PRESS value of a classical PCR estimator is very large. The reason for the poor performance of classical PCR is different from the aforementioned reasons. Instead, this is mainly because the contribution rate, namely 85%, is chosen subjectively rather than being data-driven, which leads to the results of a PCR falling short of theoretical expectations.

To check how contribution rates influence the corresponding PRESS values, we reevaluate the classical PCR estimators in simulation studies, with the contribution rates taking 75, 85, and 95%, respectively. All of the results are presented in corresponding tables. The results indicate that:

- In any case of n, p, σ, and ρ, the value of PRESS of PCR decreases as the cumulative contribution rate of the principal components increases, although the decrease in PRESS may not be strict. For example, in the case of $n = 30$, $p = 6$, and $\sigma = 0.75$, the PRESS values of the three PCR estimators are 263.6573, 234.4363, and 108.8466 when ρ takes 0.9, while the values are equal to each other when ρ takes 0.999.
- For any case of n, p, σ, and a fixed cumulative contribution rate of the principal components, the value of PRESS of PCR strictly decreases as *large* ρ increases. Taking also the case of $n = 30$, $p = 6$, and $\sigma = 0.75$, the PRESS values of the 95% PCR estimators are 108.8466, 13.1987, and 9.8573 for ρ taking 0.9, 0.999, and 0.99999.
- In any case of p, σ, ρ, and a fixed cumulative contribution rate of the principal components, the *averaged* value of PRESS of PCR with respect to n, namely $\frac{1}{n}$PRESS, strictly decreases as n increases. For example, in the case of $p = 6$, $\sigma = 0.75$, and $\rho = 0.9$, the averaged PRESS values of the 95% PCR estimators are
$$\frac{35.5229}{30} = 1.1841, \quad \frac{72.2737}{70} = 1.0325,$$
$$\frac{83.8286}{100} = 0.8383, \quad \frac{91.9920}{200} = 0.4600,$$
respectively, for $n = 30,\ 70,\ 100$, and 200.

For the Hald and Acetylene data, we consider the performance of the ordinary ridge regression (ORR) [4] and the Liu estimator (LE) [14], since each of these two estimators involves only one biased parameter, which can be easily adjusted by linearly changing the values from 0 to 1 or to a smaller/larger scalar when computing the PRESS values for the associated estimates. By direct computations, the results are derived and presented in Figures 1 and 2. By the figures, it follows that

- For the Hald data, LE and ORR have the minimal PRESS values 97.6613 and 96.8488, respectively, when the Liu parameter d takes 0.1954 and the ridge parameter k takes 0.002153.
- For the Acetylene data, LE and ORR get the minimal PRESS values 330.8642 and 311.2461, respectively, when $d = 0.9345$ and the ridge parameter k takes 0.005355.

By Tables 1 and 2, all of the three new estimators (HPCR, WPCR, and WnnPCR) have much smaller PRESS values and therefore outperform LE and ORR under the PRESS criterion.

Additionally, note here that the smaller the PRESS value, the better the model's predictive ability. We employ a predicted version of R^2 to measure the predictive ability of the model. The *predicted* R^2 of an estimator, $(\vec{\beta}_0, \vec{\boldsymbol{\beta}})$, is defined as follows:

$$R^2_{\text{PRESS}}\left(\vec{\beta}_0, \vec{\boldsymbol{\beta}}; \beta_0, \boldsymbol{\beta}\right) \triangleq 1 - \frac{\text{PRESS}\left(\vec{\beta}_0, \vec{\boldsymbol{\beta}}; \beta_0, \boldsymbol{\beta}\right)}{\sum_{i=1}^{n}\left(y_i - \frac{1}{n-1}\sum_{j\neq i} y_j\right)^2}.$$

By direct computations, the R^2_{PRESS} values of the seven estimators (LS, PCR, HPCR, WPCR, WnnPCR, LE, and ORR) in the Hald and Acetylene data are

$$\begin{array}{rl} \text{Hald data}: & 0.9654,\ \ 0.6562,\ \ 0.9733,\ \ 0.9733,\ \ 0.9707,\ \ 0.9662,\ \ 0.9696; \\ \text{Acetylene data}: & 0.8608,\ \ 0.8711,\ \ 0.9272,\ \ 0.9260,\ \ 0.8795,\ \ 0.8631,\ \ 0.8712. \end{array}$$

The results indicate similar expected conclusions to that of Subsection 4.1.

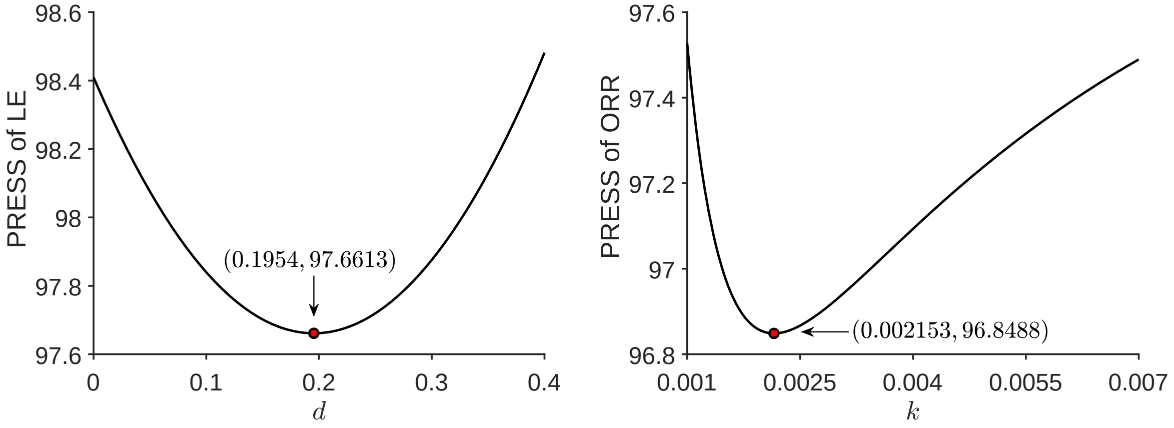

Figure 1. PRESS curves of LE (left) and ORR (right) versus the biased parameters, *d* and *k*, for the Hald data.

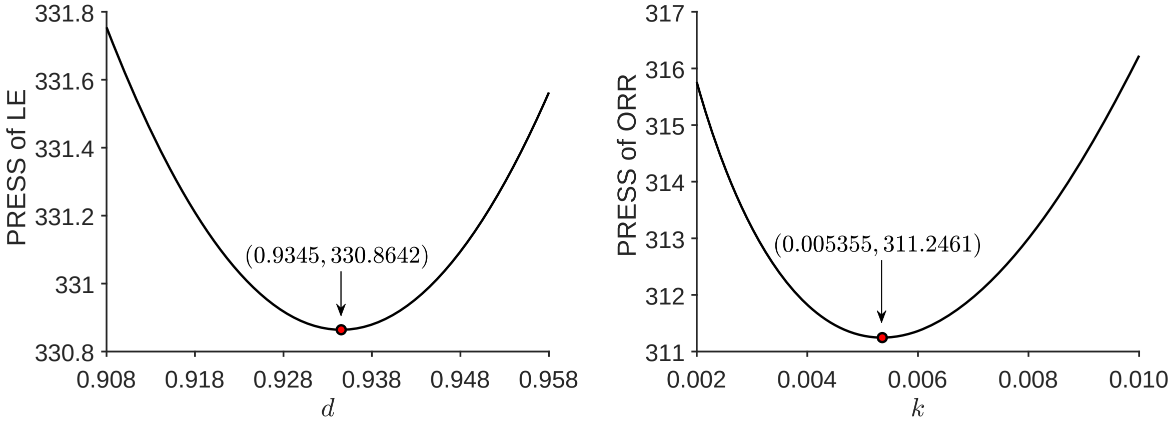

Figure 2. PRESS curves of LE (left) and ORR (right) versus the biased parameters, *d* and *k*, for the Acetylene data.

5. Conclusions and suggestion

This paper addresses the issues existing in the classic PCR estimation and proposes three hybrid PCR estimators. The two real examples and the simulation study demonstrate the desirable performance of the new methods. Also, the three hybrid PCR estimators could also be studied under the MSE criterion. However, since they are biased estimators, the determination of the weights in the MSE sense can only be iteratively solved from a numerical perspective. This also implies that the estimators will no longer be linear estimators after the first iteration, making it difficult to accurately represent the value of MSE and only approximate results can be obtained. In short, the study of hybrid PCR estimation under the MSE criterion is challenging. In what follows, we give two suggestions for the use of the new estimators.

Suggestion 1: Despite the issue of selecting the contribution rate, classic PCR estimation still yields decent estimators by automatically determining *which cumulative contribution rate to use* (in essence, this is equivalent to *how many principal components to use*). Therefore, in cases where data size is large and there are numerous regression variables, users can still employ the classic PCR method to estimate parameters. This can be seen from the aforementioned fact that the averaged PRESS value decreases as the data size increases.

Suggestion 2: We can determine which estimator to use by considering the degree of multicollinearity. *If multicollinearity is absent or weak*, we can use the LS estimator directly. *If multicollinearity is moderate*, we can combine the above Suggestion 1 to choose a classical PCR estimator with an appropriate cumulative contribution rate. *If multicollinearity is severe*, it is recommended to use the hybrid PCR estimator.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors are very grateful to the four anonymous reviewers for their valuable comments and constructive suggestions, which were helpful in improving the paper. They would also like to thank Miss Bing-Jie Li for her constructive comments during the drafting of this paper.

Conflict of interest

The authors declare that there are no conflicts of interest.

References

1. S. Wang, *Theory of Linear Models and its Applications*, Anhui Education Press, Hefei, 1987.
2. R. H. Myers, *Classical and Modern Regression with Application*, 2nd edition, Higher Education Press, Beijing, 2005.
3. C. R. Rao, H. Toutenburg, *Linear Models: Least Squares and Alternatives*, Springer, New York, 1995. https://doi.org/10.1007/978-1-4899-0024-1_2
4. A. E. Hoerl, R. W. Kennard, Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, **12** (1970), 55–67. https://doi.org/10.1080/00401706.1970.10488634
5. A. E. Hoerl, R. W. Kennard, Ridge regression: application for non-orthogonal problems, *Technometrics*, **12** (1970), 69–82. https://doi.org/10.1080/00401706.1970.10488635
6. X. Q. Liu, F. Gao, Linearized ridge regression estimator in linear regression, *Commun. Stat.-Theor. M.*, **40** (2011), 2182–2192. https://doi.org/10.1080/03610921003746693

7. X. Q. Liu, F. Gao, Z. F. Yu, Improved ridge estimators in a linear regression model, *J. Appl. Stat.*, **40** (2013), 209–220. https://doi.org/10.1080/02664763.2012.740623

8. X. Q. Liu, P. Hu, General ridge predictors in a mixed linear model, *Statistics*, **47** (2013), 363–378. https://doi.org/10.1080/02331888.2011.592190

9. X. Q. Liu, H. Y. Jiang, Optimal generalized ridge estimator under the generalized cross-validation criterion in linear regression, *Linear Algebra Appl.*, **436** (2012), 1238–1245. https://doi.org/10.1016/j.laa.2011.08.032

10. A. F. Lukman, K. Ayinde, B. M. G. Kibria, E. T. Adewuyi, Modified ridge-type estimator for the gamma regression model, *Commun. Stat.-Simul. C.*, **51** (2022), 5009–5023. https://doi.org/10.1080/03610918.2020.1752720

11. B. M. G. Kibria, More than hundred (100) estimators for estimating the shrinkage parameter in a linear and generalized linear ridge regression models, *J. Economet. Stat.*, **2** (2022), 233–252.

12. Z. Y. Algamal, A. F. Lukman, B. M. G. Kibria, A. O. Taofik, Modified jackknifed ridge estimator in bell regression model: theory, simulation and applications, *Iraqi J. Comput. Sci. Math.*, **4** (2023), 146–154. https://doi.org/10.52866/ijcsm.2023.01.01.0012

13. W. F. Massy, Principal components regression in exploratory statistical research, *J. Am. Stat. Assoc.*, **60** (1965), 234–256. https://doi.org/10.1080/01621459.1965.10480787

14. K. J. Liu, A new class of biased estimate in linear regression. *Commun. Stat.- Theor. M.*, **22** (1993), 393–402. https://doi.org/10.1080/03610929308831027

15. K. J. Liu, Using Liu-type estimator to combat collinearity. *Commun. Stat.- Theor. M.*, **32** (2003), 1009–1020. https://doi.org/10.1081/STA-120019959

16. X. Q. Liu, Improved Liu estimator in a linear regression model, *J. Stat. Plan. Infer.*, **141** (2011), 189–196. https://doi.org/10.1016/j.jspi.2010.05.030

17. J. Y. Rong, Adjustive Liu-type estimators in linear regression models, *Commun. Stat.-Simul. C.*, **39** (2010), 1162–1173. https://doi.org/10.1080/03610918.2010.484120

18. A. Ullah, S. Ullah, Double k-class estimators of coefficients in linear regression, *Econometrica*, **46** (1978), 705–722. https://doi.org/10.2307/1914242

19. D. M. Allen, Mean square error of prediction as a criterion for selection of variables, *Technometrics*, **13** (1971), 469–475. https://doi.org/10.1080/00401706.1971.10488811

20. D. M. Allen, The relationship between variable selection and data augmentation and a method for prediction., *Technometrics*, **16** (1974), 125–127. https://doi.org/10.1080/00401706.1974.10489157

21. X. Q. Liu, J. Y. Rong, X. Y. Liu, Best linear unbiased prediction for linear combinations in general mixed linear models. *J. Multivariate Anal.*, **99** (2008), 1503–1517. https://doi.org/10.1016/j.jmva.2008.01.004

22. H. Woods, H. H. Steinour, H. R. Starke, Effect of composition of Portland cement on heat evolved during hardening. *Ind. Eng. Chem.*, **24** (1932), 1207–1214. https://doi.org/10.1021/ie50275a002

23. T. Kunugi, T. Tamura, T. Naito , New acetylene process uses hydrogen dilution, *Chem. Eng. Prog.*, **57** (1961), 43–49.

24. G. C. McDonald, D. I. Galarneau, A Monte Carlo evaluation of some ridge-type estimators, *J. Am. Stat. Assoc.*, **70** (1975), 407–416. https://doi.org/10.1080/01621459.1975.10479882

Appendix

Table A1. PRESS values of the five estimates with respect to $n = 30$ and different p (the number of explanatory variables), σ (the model error standard deviation), and ρ (the correlation between regressors, measuring the degree of multicollinearity). In addition, the averaged time (AT) in seconds for every run is presented in the final column of each subtable.

$p = 3$ and $\sigma = 0.75$						$p = 3$ and $\sigma = 0.25$					
Estimation	ρ				**AT**	**Estimation**	ρ				**AT**
	0.5	0.9	0.999	0.99999			0.5	0.9	0.999	0.99999	
LS	11.0539	11.2180	10.6930	10.5669	5.20×10^{-6}	LS	0.1312	0.1394	0.1360	0.1354	5.19×10^{-6}
75% PCR	227.8098	91.9703	11.0269	9.8523	8.42×10^{-6}	75% PCR	178.7443	125.1205	1.1773	0.1356	8.15×10^{-6}
85% PCR	80.9064	85.9203	11.0269	9.8523	8.70×10^{-6}	85% PCR	36.3780	112.9114	1.1773	0.1356	8.60×10^{-6}
95% PCR	11.0539	35.5229	11.0269	9.8523	8.84×10^{-6}	95% PCR	0.1312	36.1212	1.1773	0.1356	8.55×10^{-6}
HPCR	10.8179	10.8774	8.8617	8.2479	4.49×10^{-4}	HPCR	0.1285	0.1370	0.1334	0.1138	4.38×10^{-4}
WPCR	10.8823	10.9076	8.9095	8.3367	4.51×10^{-4}	WPCR	0.1293	0.1374	0.1335	0.1144	4.36×10^{-4}
$\mathrm{W_{nn}PCR}$	10.9890	11.0487	9.8872	9.6319	1.07×10^{-3}	$\mathrm{W_{nn}PCR}$	0.1307	0.1388	0.1349	0.1251	1.03×10^{-3}
$p = 6$ and $\sigma = 0.75$						$p = 6$ and $\sigma = 0.25$					
Estimation	ρ				**AT**	**Estimation**	ρ				**AT**
	0.5	0.9	0.999	0.99999			0.5	0.9	0.999	0.99999	
LS	12.5214	13.4138	12.7414	12.2797	7.51×10^{-6}	LS	0.1523	0.1536	0.1457	0.1566	7.53×10^{-6}
75% PCR	567.5426	263.6573	13.1987	9.8573	1.26×10^{-5}	75% PCR	486.3268	266.7937	2.9079	0.1509	1.25×10^{-5}
85% PCR	395.4563	234.4363	13.1987	9.8573	1.31×10^{-5}	85% PCR	296.2013	238.4373	2.9079	0.1500	1.30×10^{-5}
95% PCR	99.7852	108.8466	13.1987	9.8573	1.35×10^{-5}	95% PCR	67.0567	105.7575	2.9079	0.1509	1.29×10^{-5}
HPCR	11.5709	12.2411	8.3825	6.7288	1.02×10^{-4}	HPCR	0.1382	0.1419	0.1357	0.1004	1.04×10^{-3}
WPCR	11.6566	12.2800	8.4516	6.7911	1.02×10^{-4}	WPCR	0.1393	0.1425	0.1360	0.1012	1.02×10^{-3}
$\mathrm{W_{nn}PCR}$	12.3923	13.0533	10.6819	9.5118	1.93×10^{-3}	$\mathrm{W_{nn}PCR}$	0.1510	0.1516	0.1430	0.1297	1.64×10^{-3}
$p = 9$ and $\sigma = 0.75$						$p = 9$ and $\sigma = 0.25$					
Estimation	ρ				**AT**	**Estimation**	ρ				**AT**
	0.5	0.9	0.999	0.99999			0.5	0.9	0.999	0.99999	
LS	14.7597	14.4471	14.6638	13.8221	1.12×10^{-5}	LS	0.1849	0.1789	0.1814	0.1783	1.09×10^{-5}
75% PCR	875.0244	420.0682	14.2038	9.8126	2.01×10^{-5}	75% PCR	826.4022	417.4707	4.6853	0.1645	1.93×10^{-5}
85% PCR	670.1962	376.9690	14.2038	9.8126	2.04×10^{-5}	85% PCR	595.5124	350.4838	4.6853	0.1645	2.00×10^{-5}
95% PCR	271.4669	209.5098	14.2038	9.8126	2.20×10^{-5}	95% PCR	257.5705	187.1392	4.6853	0.1645	2.11×10^{-5}
HPCR	12.3959	12.0719	7.8661	5.4684	2.07×10^{-3}	HPCR	0.1578	0.1539	0.1533	0.0924	2.04×10^{-3}
WPCR	12.5516	12.1457	7.9631	5.5474	2.01×10^{-3}	WPCR	0.1595	0.1545	0.1540	0.0935	2.02×10^{-3}
$\mathrm{W_{nn}PCR}$	14.3971	13.9815	10.8856	9.3178	2.78×10^{-3}	$\mathrm{W_{nn}PCR}$	0.1822	0.1765	0.1757	0.1318	2.69×10^{-3}

Table A2. PRESS values and the AT in seconds of the five estimates with respect to $n = 70$ and different p (the number of explanatory variables), σ (the model error standard deviation), and ρ (the correlation between regressors, measuring the degree of multicollinearity).

$p = 3$ and $\sigma = 0.75$						$p = 3$ and $\sigma = 0.25$					
Estimation	ρ				**AT**	**Estimation**	ρ				**AT**
	0.5	0.9	0.999	0.99999			0.5	0.9	0.999	0.99999	
LS	23.8596	23.2023	23.3577	22.6883	6.83×10^{-6}	LS	0.2953	0.2833	0.2816	0.2867	6.69×10^{-6}
75% PCR	411.9952	260.9086	25.1142	22.0457	1.06×10^{-5}	75% PCR	341.8998	239.6013	2.5855	0.3004	1.11×10^{-5}
85% PCR	37.2456	240.8555	25.1142	22.0457	1.02×10^{-5}	85% PCR	8.7850	219.9861	2.5855	0.3004	1.04×10^{-5}
95% PCR	23.8596	72.2737	25.1142	22.0457	1.04×10^{-5}	95% PCR	0.2953	29.7473	2.5855	0.3004	1.07×10^{-5}
HPCR	23.7499	23.0839	22.1389	19.7402	1.05×10^{-3}	HPCR	0.2938	0.2824	0.2790	0.2688	1.02×10^{-3}
WPCR	23.7774	23.1177	22.1569	19.7513	1.03×10^{-3}	WPCR	0.2941	0.2824	0.2791	0.2692	1.01×10^{-3}
$\mathrm{W_{nn}PCR}$	23.8379	23.1618	22.9429	21.8573	1.74×10^{-3}	$\mathrm{W_{nn}PCR}$	0.2951	0.2830	0.2804	0.2805	1.63×10^{-3}
$p = 6$ and $\sigma = 0.75$						$p = 6$ and $\sigma = 0.25$					
Estimation	ρ				**AT**	**Estimation**	ρ				**AT**
	0.5	0.9	0.999	0.99999			0.5	0.9	0.999	0.99999	
LS	25.0859	24.9995	23.8188	24.3008	9.65×10^{-6}	LS	0.3013	0.3014	0.3020	0.3018	9.67×10^{-6}
75% PCR	924.9686	647.3329	27.8634	22.39580	1.55×10^{-5}	75% PCR	1004.9610	578.6075	6.1000	0.3300	1.48×10^{-5}
85% PCR	469.1337	573.0600	27.8634	22.3958	1.56×10^{-5}	85% PCR	535.3425	497.5772	6.1000	0.3300	1.53×10^{-5}
95% PCR	25.0859	249.1194	27.8634	22.3958	1.64×10^{-5}	95% PCR	0.9748	222.4911	6.1000	0.3300	1.33×10^{-5}
HPCR	24.5928	24.5647	20.7465	17.4100	2.44×10^{-3}	HPCR	0.2959	0.2958	0.2963	0.2584	2.45×10^{-3}
WPCR	24.6228	24.5746	20.7771	17.5404	2.42×10^{-3}	WPCR	0.2960	0.2958	0.2964	0.2587	2.43×10^{-3}
$\mathrm{W_{nn}PCR}$	25.0408	24.9294	22.7655	22.0787	3.05×10^{-3}	$\mathrm{W_{nn}PCR}$	0.3006	0.3009	0.3010	0.2868	3.06×10^{-3}
$p = 9$ and $\sigma = 0.75$						$p = 9$ and $\sigma = 0.25$					
Estimation	ρ				**AT**	**Estimation**	ρ				**AT**
	0.5	0.9	0.999	0.99999			0.5	0.9	0.999	0.99999	
LS	26.5505	25.6543	25.6111	25.4260	2.09×10^{-5}	LS	0.3117	0.3226	0.3140	0.3199	2.07×10^{-5}
75% PCR	1832.3860	925.2033	31.6841	22.5608	2.23×10^{-5}	75% PCR	1752.8440	933.2482	9.6699	0.3885	2.22×10^{-5}
85% PCR	1067.9560	809.5815	31.6841	22.5608	2.90×10^{-5}	85% PCR	1056.16210	780.3915	9.6699	0.3885	2.87×10^{-5}
95% PCR	388.8009	377.3299	31.6841	22.5608	2.42×10^{-5}	95% PCR	485.5310	364.0007	9.6699	0.3885	2.69×10^{-5}
HPCR	25.5102	24.6260	20.9860	16.1435	6.45×10^{-3}	HPCR	0.2989	0.3100	0.3013	0.2577	6.36×10^{-3}
WPCR	25.5238	24.6415	21.0128	16.1996	6.61×10^{-3}	WPCR	0.2995	0.3101	0.3015	0.2580	6.52×10^{-3}
$\mathrm{W_{nn}PCR}$	26.4616	25.4983	23.9798	22.0273	7.71×10^{-3}	$\mathrm{W_{nn}PCR}$	0.3110	0.3218	0.3126	0.2975	7.49×10^{-3}

Table A3. PRESS values and the AT in seconds of the five estimates with respect to $n = 100$ and different p (the number of explanatory variables), σ (the model error standard deviation), and ρ (the correlation between regressors, measuring the degree of multicollinearity).

$p = 3$ and $\sigma = 0.75$						$p = 3$ and $\sigma = 0.25$					
Estimation	ρ				**AT**	**Estimation**	ρ				**AT**
	0.5	0.9	0.999	0.99999			0.5	0.9	0.999	0.99999	
LS	33.9000	33.0669	32.5858	32.6356	9.01×10^{-6}	LS	0.4054	0.4031	0.4175	0.4063	8.89×10^{-6}
75% PCR	576.7535	407.6325	35.1806	31.9583	1.20×10^{-5}	75% PCR	594.0972	323.1151	3.5926	0.4293	1.24×10^{-5}
85% PCR	33.9000	384.1914	35.1806	31.9583	1.26×10^{-5}	85% PCR	0.4445	292.3829	3.5926	0.4293	1.26×10^{-5}
95% PCR	33.9000	83.8286	35.18060	31.9583	1.12×10^{-5}	95% PCR	0.4054	28.7849	3.5926	0.4293	1.35×10^{-5}
HPCR	33.8552	32.9863	31.0655	28.7446	1.89×10^{-3}	HPCR	0.4044	0.4021	0.4159	0.3811	1.88×10^{-3}
WPCR	33.8610	33.0130	31.0746	28.7604	1.87×10^{-3}	WPCR	0.4046	0.4022	0.4159	0.3812	1.87×10^{-3}
$\mathrm{W_{nn}PCR}$	33.8837	33.0495	32.1851	31.7592	2.56×10^{-3}	$\mathrm{W_{nn}PCR}$	0.4052	0.4030	0.4168	0.4005	2.54×10^{-3}
$p = 6$ and $\sigma = 0.75$						$p = 6$ and $\sigma = 0.25$					
Estimation	ρ				**AT**	**Estimation**	ρ				**AT**
	0.5	0.9	0.999	0.99999			0.5	0.9	0.999	0.99999	
LS	34.0111	33.0242	34.2546	33.8952	1.25×10^{-5}	LS	0.4188	0.4164	0.4192	0.4185	1.24×10^{-5}
75% PCR	1366.4144	820.9073	40.7334	32.0514	1.88×10^{-5}	75% PCR	1229.5673	842.4245	9.5972	0.4833	1.88×10^{-5}
85% PCR	697.7698	738.3613	40.7334	32.0514	1.86×10^{-5}	85% PCR	571.2204	728.8761	9.5972	0.4833	1.84×10^{-5}
95% PCR	34.0111	332.7033	40.7334	32.0514	1.91×10^{-5}	95% PCR	0.4188	275.1532	9.5972	0.4833	2.02×10^{-5}
HPCR	33.5875	32.6632	31.8211	25.5422	4.16×10^{-3}	HPCR	0.4140	0.4124	0.4144	0.3811	4.14×10^{-3}
WPCR	33.6049	32.6659	31.8451	25.6628	4.17×10^{-3}	WPCR	0.4142	0.4124	0.4145	0.3813	4.15×10^{-3}
$\mathrm{W_{nn}PCR}$	33.9789	32.9370	33.5012	31.7416	4.86×10^{-3}	$\mathrm{W_{nn}PCR}$	0.4182	0.4159	0.4183	0.4077	4.83×10^{-3}
$p = 9$ and $\sigma = 0.75$						$p = 9$ and $\sigma = 0.25$					
Estimation	ρ				**AT**	**Estimation**	ρ				**AT**
	0.5	0.9	0.999	0.99999			0.5	0.9	0.999	0.99999	
LS	35.9664	35.3190	35.2747	34.8187	3.23×10^{-5}	LS	0.4213	0.4350	0.4327	0.4350	3.18×10^{-5}
75% PCR	2166.8784	1310.8235	46.3709	31.9581	4.16×10^{-5}	75% PCR	2229.7730	1263.3911	14.7056	0.5296	4.43×10^{-5}
85% PCR	1241.8423	1108.2647	46.3709	31.9581	4.13×10^{-5}	85% PCR	1162.7843	1145.1165	14.7056	0.5296	4.11×10^{-5}
95% PCR	439.4468	499.6003	46.3709	31.9581	4.20×10^{-5}	95% PCR	356.5309	466.6388	14.7056	0.5296	4.67×10^{-5}
HPCR	35.2201	34.4002	31.9476	23.9691	1.22×10^{-2}	HPCR	0.4114	0.4248	0.4222	0.3776	1.22×10^{-2}
WPCR	35.2739	34.4100	31.9613	24.0010	1.23×10^{-2}	WPCR	0.4115	0.4248	0.4223	0.3778	1.23×10^{-2}
$\mathrm{W_{nn}PCR}$	35.8971	35.2194	34.1270	31.4494	1.34×10^{-2}	$\mathrm{W_{nn}PCR}$	0.4207	0.4341	0.4317	0.4181	1.35×10^{-2}

Table A4. PRESS values and the AT in seconds of the five estimates with respect to $n = 200$ and different p (the number of explanatory variables), σ (the model error standard deviation), and ρ (the correlation between regressors, measuring the degree of multicollinearity).

$p = 3$ and $\sigma = 0.75$						$p = 3$ and $\sigma = 0.25$					
Estimation	ρ				**AT**	**Estimation**	ρ				**AT**
	0.5	0.9	0.999	0.99999			0.5	0.9	0.999	0.99999	
LS	65.8927	64.8278	65.3548	63.7734	3.19×10^{-5}	LS	0.7912	0.7970	0.7888	0.8047	3.06×10^{-5}
75% PCR	1101.2950	630.3620	72.6789	63.2837	4.03×10^{-5}	75% PCR	1006.2337	623.7216	7.2759	0.8634	3.90×10^{-5}
85% PCR	65.8927	621.7803	72.6789	63.2837	3.74×10^{-5}	85% PCR	0.7912	608.3858	7.2759	0.8634	3.55×10^{-5}
95% PCR	65.8927	91.9920	72.6789	63.2837	4.18×10^{-5}	95% PCR	0.7912	45.5516	7.2759	0.8634	3.83×10^{-5}
HPCR	65.8383	64.7109	63.7329	58.5412	4.05×10^{-3}	HPCR	0.7905	0.7963	0.7877	0.7821	4.01×10^{-3}
WPCR	65.8433	64.7160	63.9414	58.5526	3.86×10^{-3}	WPCR	0.7908	0.7963	0.7877	0.7821	3.84×10^{-3}
$\mathrm{W_{nn}PCR}$	65.8810	64.8086	65.0444	63.0243	4.61×10^{-3}	$\mathrm{W_{nn}PCR}$	0.7911	0.7969	0.7885	0.8008	4.57×10^{-3}
$p = 6$ and $\sigma = 0.75$						$p = 6$ and $\sigma = 0.25$					
Estimation	ρ				**AT**	**Estimation**	ρ				**AT**
	0.5	0.9	0.999	0.99999			0.5	0.9	0.999	0.99999	
LS	65.3578	66.0217	67.1194	64.6662	3.74×10^{-5}	LS	0.8292	0.8143	0.8016	0.8168	3.78×10^{-5}
75% PCR	2356.9438	1724.4008	81.6643	63.2941	4.81×10^{-5}	75% PCR	2493.9479	1601.8222	17.1199	0.9537	5.11×10^{-5}
85% PCR	1281.3240	1499.5096	81.6643	63.2941	4.53×10^{-5}	85% PCR	1232.6163	1343.0020	17.1199	0.9537	4.60×10^{-5}
95% PCR	65.3578	548.4416	81.6643	63.2941	4.06×10^{-5}	95% PCR	0.8292	423.7067	17.1199	0.9537	4.75×10^{-5}
HPCR	65.1159	65.7827	64.5475	54.7151	8.99×10^{-3}	HPCR	0.8262	0.8110	0.7984	0.7846	8.93×10^{-3}
WPCR	65.1209	65.7867	64.5600	54.7449	8.84×10^{-3}	WPCR	0.8262	0.8110	0.7984	0.7847	8.79×10^{-3}
$\mathrm{W_{nn}PCR}$	65.3326	65.9825	66.5801	62.7871	9.55×10^{-3}	$\mathrm{W_{nn}PCR}$	0.8290	0.8139	0.8012	0.8096	9.49×10^{-3}
$p = 9$ and $\sigma = 0.75$						$p = 9$ and $\sigma = 0.25$					
Estimation	ρ				**AT**	**Estimation**	ρ				**AT**
	0.5	0.9	0.999	0.99999			0.5	0.9	0.999	0.99999	
LS	66.9855	66.9353	65.4687	65.5627	5.19×10^{-5}	LS	0.8073	0.8207	0.8114	0.8317	5.17×10^{-5}
75% PCR	3941.4811	2510.8703	91.0613	62.9964	4.80×10^{-5}	75% PCR	3986.4039	2601.2332	27.1017	1.0544	7.01×10^{-5}
85% PCR	2587.3025	2073.7567	91.0613	62.9964	5.98×10^{-5}	85% PCR	2687.6931	2178.3749	27.1017	1.0544	5.97×10^{-5}
95% PCR	78.3062	884.2350	91.0613	62.9964	6.24×10^{-5}	95% PCR	7.1094	850.8616	27.1017	1.0544	6.34×10^{-5}
HPCR	66.4251	66.3840	62.3900	51.3104	2.49×10^{-3}	HPCR	0.8008	0.8149	0.8033	0.7884	2.49×10^{-3}
WPCR	66.4438	66.3861	62.4011	51.3373	2.52×10^{-3}	WPCR	0.8010	0.8149	0.8033	0.7885	2.54×10^{-3}
$\mathrm{W_{nn}PCR}$	66.9269	66.8833	64.8061	62.5617	2.64×10^{-3}	$\mathrm{W_{nn}PCR}$	0.8070	0.8203	0.8106	0.8222	2.72×10^{-3}