MATEMATIKA, 2006, Volume 22, Number 2, pp. 109–119

A Comparative Study On Some Methods For Handling Multicollinearity Problems

Norliza Adnan, [1]Maizah Hura Ahmad, Robiah Adnan
Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia
81310 UTM Skudai, Johor, Malaysia.
e-mail: [1]maizah@utm.my

Abstract In regression, the objective is to explain the variation in one or more response variables, by associating this variation with proportional variation in one or more explanatory variables. A frequent obstacle is that several of the explanatory variables will vary in rather similar ways. As a result, their collective power of explanation is considerably less than the sum of their individual powers. This phenomenon called multicollinearity, is a common problem in regression analysis. Handling multicollinearity problem in regression analysis is important because least squares estimations assume that predictor variables are not correlated with each other. The performances of ridge regression (RR), principal component regression (PCR) and partial least squares regression (PLSR) in handling multicollinearity problem in simulated data sets are compared to help and give future researchers a comprehensive view about the best procedure to handle multicollinearity problems. PCR is a combination of principal component analysis (PCA) and ordinary least squares regression (OLS) while PLSR is an approach similar to PCR because a component that can be used to reduce the number of variables need to be constructed. RR on the other hand is the modified least square method that allows a biased but more precise estimator. The algorithm is described and for the purpose of comparing the three methods, simulated data sets where the number of cases were less than the number of observations used. The goal was to develop a linear equation that relates all the predictor variables to a response variable. For comparison purposes, mean square errors (MSE) were calculated. A Monte Carlo simulation study was used to evaluate the effectiveness of these three procedures. The analysis including all simulations and calculations were done using statistical package S-Plus 2000 software.

Keywords Partial least squares regression, principal component regression, ridge regression, multicollinearity.

1 The Problem of Multicollinearity

In the applications of regression analysis, multicollinearity is a problem that always occur when two or more predictor variables are correlated with each other. This problem can cause the value of the least squares estimated regression coefficients to be conditional upon the correlated predictor variables in the model. As defined by Bowerman and O'Connell [1], multicollinearity is a problem in regression analysis when the independent variables in a regression model are intercorrelated or are dependent on each other.

Multicollinearity is a condition in a set of regression data that have two or more regressors that are redundant and have the same information. The linear dependencies among the regressors can effect the model ability to estimate regression coefficients.

Redundant information means, what one variable explains about Y is exactly what the other variable explains. In this case, the two or more redundant predictor variables would be completely unreliable since the b_i would measure the same effect of those x_i and the same goes for the other b. Furthermore $\boldsymbol{(X'X)^{-1}}$ would not exist because the denominators, $1 - r_{ik}^2$ is zero. As a result, the values for b_i cannot be found since the elements of the inverse matrix and coefficients become quite large (Younger [2]).

The presence of multicollinearity in least squares regression can cause larger variances of parameter estimates which means that the estimates of the parameters tend to be less precise. As a result, the model will have insignificant test and wide confidence interval. Therefore, the more the multicollinearity, the less interpretable are the parameters.

The problems of multicollinearity in regression analysis can have effects on least squares estimated regression coefficients, computational accuracy, estimated standard deviation of least squares estimated regression coefficients, $t-$test, extra sum of squares, fitted values and predictions, and coefficients of partial determination.

There are a variety of informal and formal methods that have been developed for detecting the presence of serious multicollinearity. One of the most commonly used method is the variance inflation factor that measures how much the variances of the estimated regression coefficients are inflated compared to when the independent variables are not linearly related (Neter et al. [3]).

2 Methods for Handling Multicollinearity

Various methods have been developed to cope with multicollinearity problems. Among such methods are Ridge Regression, Principal Component Regression, Partial Least Squares Regression and Continuum Regression.

2.1 Ridge Regression

Ridge Regression is developed by Hoerl and Kennard [4] and this method is the modification of the least squares method that allow biased estimators of the regression coefficients. Although the estimators are biased, the biases are small enough for these estimators to be substantially more precise than unbiased estimators. Therefore, these biased estimators are prefered over unbiased ones since they will have a larger probability of being close to the true parameter values.

The ridge regression estimator of the coefficient β is found by solving for b_R in the equation

$$\boldsymbol{(X'X + \delta\ I)}b_R = \boldsymbol{X'y}$$

where $\delta \geq 0$ is often referred to as a shrinkage parameter. Thus, the solution for ridge estimator is given by

$$b_R = \boldsymbol{(X'X + \delta\ I)}^{-1}\ \boldsymbol{X'y}$$

The matrix $(\boldsymbol{X^{*\prime}X^{*}})$ considered is replaced by $(\boldsymbol{X^{*\prime}X^{*}} + \delta\ \boldsymbol{I})$, where δ is a small positive quantity. Since the $\boldsymbol{V}$ matrix diagonalizes $(\boldsymbol{X^{*\prime}X^{*}})$, it also diagonalizes $(\boldsymbol{X^{*\prime}X^{*}} + \delta\ \boldsymbol{I})$. Thus

$$\boldsymbol{V'(X^{*\prime}X^{*} + \delta\, I)V} = \begin{bmatrix} \lambda_1 + \delta & 0 & \dots & 0 \\ 0 & \lambda_2 + \delta & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k + \delta \end{bmatrix}$$

The eigenvalues of the new matrix $\boldsymbol{(X^{*\prime}X^{*} + \delta I)}$ are $\lambda_i + \delta$ for $i = 1, 2, \dots, k$ where adding δ to the main diagonal effectively replaces λ_i by $\lambda_i + \delta$.

From the properties of the ridge estimator, the role of δ are revealed in moderating the variance of the estimators. The impact of eigenvalues on the variances of the ridge regression coefficients can be illustrated as

$$\sum_i \frac{\mathrm{Var}(b_{i,R})}{\sigma^2} = \sum_i \frac{\lambda_i}{(\lambda_i + \delta)^2}$$

Therefore, the δ in ridge regression moderates the damaging impact of the small eigenvalues that result from collinearity (Myers [5]).

There are various procedures for choosing the shrinkage parameter δ. According to [5], the ridge trace is a very pragmatic procedure for choosing the shrinkage parameter where it allows δ to increase until stability is indicated in all coefficients. A plot of the coefficients against δ that pictorially displays the trace often helps the analyst to make a decision regarding the appropriate value of δ. However, stability does not imply that the regression coefficients have converged. As δ grows, variances reduce and the coefficients become more stable. Therefore, the value of δ is chosen at the point for which the coefficients no longer change rapidly.

The C_p-like statistic that is based on the same variance-bias type trade-off is one of the proposed procedures. The use of C_δ statistic is by a simple plotting of C_δ against δ, with the use of δ-value for which C_δ is minimized. The statistic is given as

$$C_\delta = \frac{SS_{\mathrm{Re}\ s,\delta}}{\hat{\sigma}^2} - n + 2 + 2tr\left[H_\delta\right]$$

where $\boldsymbol{H_\delta} = \left[\boldsymbol{X^{*}(X^{*\prime}X^{*} + \delta I)^{-1}X^{*\prime}}\right]$, $SS_{\mathrm{Re}\ s,\delta}$ is the residual sum of squares using ridge regression and $tr\left[\boldsymbol{H_\delta}\right]$ is the trace of $\boldsymbol{H_\delta}$. Notice that $\boldsymbol{H_\delta}$ plays the same role as the HAT matrix in ordinary least squares. In ordinary least squares, residuals are helpful in identifying outliers which do not appear to be consistent with the rest of the data while the HAT matrix is used to identify "high leverage" points which are outliers among the independent variables. The HAT matrix $\boldsymbol{H}$ is given by $\boldsymbol{H = X(X'X)^{-1}X'}$. The trace is $tr\{\boldsymbol{H}\} = p$, where p is the m vector of adjustable model parameters to be estimated from the available data set and for all diagonal elements, $0 < h_i < 1$. The statistic $\hat{\sigma}^2$ comes from the residual mean square from ordinary least squares estimation.

The other criterion that represents a prediction approach is the generalized cross

validation (GCV) that is given by

$$\begin{aligned} \text{GCV} &= \frac{\sum_{i=1}^{n} e_{i,\delta}^2}{\{n - [1 + tr(\boldsymbol{H}_\delta)]\}^2} \\ &= \frac{SS_{\text{Re } s,k}}{\{n - [1 + tr(\boldsymbol{H}_\delta)]\}^2} \end{aligned}$$

where the value of 1 in $1 + tr(\boldsymbol{H}_\delta)$ accounts for the fact that the role of the constant is not involved in $\boldsymbol{H}_\delta$. The use of this procedure is to choose δ so as to minimize GCV by a simple plotting of GCV against δ.

2.2 Principal Component Regression

The development of Principal Component Regression was done by Massy [6] to handle the problem of multicollinearity by eliminating model instability and reducing the variances of the regression coefficients. Farebrother [7] stated that the Principal Component Regression performs the Principal Component Analysis on the explanatory variables and then runs a regression using the principal component scores as the explanatory variables with the response variable of interest.

In the first step, one computes principal components which are linear combinations of the explanatory variables while in the second step, the response variable is regressed to the selected principal components. Combining both steps in a single method will maximize the relation to the response variable (Filzmoser and Croux [8]).

Principal components are orthogonal to each other, so that it becomes quite easy to attribute a specific amount of variance to each. Assuming the predictors are in a standard form, $\boldsymbol{V}$ denotes the orthogonal matrix of eigenvectors of the correlation matrix, and $\boldsymbol{Z} = \boldsymbol{XV}$, where the variables are defined by the columns of $\boldsymbol{Z}$, that are the principal components (Hocking [9]).

The matrix of normalized eigenvectors considered are associated with the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_k$ of $\boldsymbol{X^{*\prime}X^*}$ (correlation form). $\boldsymbol{V\,V'} = \boldsymbol{I}$ since $\boldsymbol{V}$ is an orthogonal matrix. Hence, the original regression model are written in the form

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{\beta_0 1} + \boldsymbol{X^* V V' \beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{y} &= \boldsymbol{\beta_0 1} + \boldsymbol{Z\alpha} + \boldsymbol{\varepsilon} \end{aligned} \tag{1}$$

where $\boldsymbol{Z} = \boldsymbol{X^* V}$ with $\boldsymbol{Z}$ an $n \times k$ matrix, and $\boldsymbol{\alpha} = \boldsymbol{V'\beta}$ with $\boldsymbol{\alpha}$ a $k \times 1$ vector of new coefficients $\alpha_1, \alpha_2, \ldots, \alpha_k$.

The columns of $\boldsymbol{Z}$ (typical element z_{ij}) can be visualized as representing readings on k new variables, called principal components. The components are orthogonal to each other as follows

$$\begin{aligned} \boldsymbol{Z'Z} &= \boldsymbol{(X^*V)'(X^*V)} \\ &= \boldsymbol{V'X^{*\prime}X^*V} \\ &= \boldsymbol{diag(\lambda_1, \lambda_2, ..., \lambda_k)} \end{aligned}$$

By performing the regression on the Z's via the model in Eq. (1) the variables of the coefficients, that is the diagonal elements of $\boldsymbol{(Z'Z)^{-1}}$ apart from σ^2 are the reciprocals of eigenvalues. Thus,

$$\frac{Var(\hat{\alpha}_j)}{\sigma^2} = \frac{1}{\lambda_j} \qquad j = 1, 2, ..., k$$

Note that, the $\hat{\alpha}$'s are least squares estimators. An elimination of at least one principal component, that is associated with the small eigenvalue, may substantially reduce the total variance in the model and thus produce an appreciably improved prediction equation [5].

According to Hwang and Nettleton [10], there are a few methods for using the principal components. One popular method is to use the principal components corresponding to the k largest eigenvalues. Xie and Kalivas [11] agreed that principal component regression is widely used for analytical calibration and in most application of PCR, the principal components are included in regression models in sequence according to respective variances. The problem of this approach is that the magnitude of the eigenvalue depends on X only and has nothing to do with the response variable.

2.3 Partial Least Squares Regression

Partial least squares (PLS) is a method for modeling relationships between a Y variable and other explanatory variables (Garthwaite [12]). This method was first developed by Wold [13]. An extension of this method is the SIMPLS method which was proposed by De Jong [14]. The SIMPLS algorithm is the leading PLSR algorithm because of its speed and efficiency. SIMPLS is based on the empirical cross-variance matrix between the response variables and the regressors. This is based on linear least squares regression.

The goal of PLS regression is to predict Y from X and to describe their common structure when X is likely to be singular and the regression approach is no longer possible to be used because of multicollinearity problem. PLSR is also used to calculate the relationship between two matrices (Boeris et al. [15]).

This method is similar to Principal Component Regression because components are first extracted before being regressed to predict Y. However in contrast, PLS regression searches for a set of components called latent vectors, factors or components from X that are also relevant for Y that performs a simultaneous decomposition of X and Y with the constraint that these components explain as much as possible of the covariance between X and Y (Abdi [16]).

This step generalizes Principal Component Analysis (PCA) and it is followed by a regression step where the decompositions of X is used to predict Y. In other words, it combines features from PCA and multiple linear regression (MLR), and the sample correlation for any pair of components is 0.

The SIMPLS method assumes that the x and y variables are related through a bilinear model

$$x_i = \bar{x} + P\tilde{t}_i + g_i$$
$$y_i = \bar{y} + A'\tilde{t}_i + f_i$$

where, $\bar{x}$ and $\bar{y}$ denote the means of the x and y variables. The $\tilde{t}_i$ are called the scores which are k-dimensional, with $k \leq p$, where $\boldsymbol{P} = \boldsymbol{P}_{p,k}$ is the matrix of x-loadings. The residuals

of each equation are represented by the g_i and f_i respectively. The matrix $\boldsymbol{A} = \boldsymbol{A}_{k,q}$ represents the slope matrix in the regression of y_i on $\tilde{t}_i$.

The elements of the scores $\tilde{t}_i$ are defined as a linear combination of the mean-centered data, $\tilde{t}_{ia} = \tilde{x}_i'\mathbf{r_a}$ or equivalently $\tilde{T}_{n,k} = \tilde{X}_{n,p}R_{p,k}$ with $\boldsymbol{R}_{p,k} = (\mathbf{r}_1, ..., \mathbf{r}_k)$. The SIMPLS weight vectors are in the pair of $(\mathbf{r_a}, \mathbf{q}_a)$. The first pair of $(\mathbf{r}_1, \mathbf{q}_1)$ is thus obtained as the first left and right singular vector of S_{xy}. This implies that $\mathbf{q}_1$ is the dominant eigenvector of $\boldsymbol{S_{yx}S_{xy}}$ and is the dominant eigenvector of $S_{xy}S_{yx}$ (with $\boldsymbol{S_{yx}} = \boldsymbol{S_{yx}'}$). The following pairs of SIMPLS weight vectors $(\mathbf{r_a}, \mathbf{q}_a)$ where $2 \leq a \leq k$ are obtained as the dominant eigenvectors of $\boldsymbol{S_{xy}^a S_{yx}^a}$ and $\boldsymbol{S_{yx}^a S_{xy}^a}$ respectively where $\boldsymbol{S_{xy}^a}$ is the deflated cross-covariance matrix

$$\boldsymbol{S_{xy}^a} = \boldsymbol{S_{xy}^{a-1}} - \boldsymbol{v_a}(\boldsymbol{v_a'S_{xy}^{a-1}}) = (\boldsymbol{I_p} - \boldsymbol{v_a v_a'})\boldsymbol{S_{xy}^{a-1}}$$

and $\boldsymbol{S_{xy}^1} = \boldsymbol{S_{xy}}$. The $\boldsymbol{\{v_1, \ldots, v_{a-1}\}}$ represents an orthonormal basis of the x-loadings $\{p_1, ..., p_{a-1}\}$. Thus, the iterative algorithm starts with $\boldsymbol{S_{xy}} = \boldsymbol{S_{xy}^1}$ and the process is repeated until k components are obtained.

The optimal number of components is often selected as that k for which this RMSE value is minimal. The Root Mean Squared Error (RMSE) defined as

$$RMSE_k = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_{i,k})^2}$$

with, $\hat{y}_{i,k}$ the predicted y-value of observation i from the test set when the regression parameter estimates are based on the training set (X, Y) of size n and k scores are retained. For the second stage of the algorithm, the responses are regressed onto these k components and the following formal regression model is considered:

$$y_i = \alpha_0 + A_{q,k}'\tilde{t}_i + f_i$$

where $E(f_i) = 0$ and $Cov(f_i) = \Sigma_f$ where multiple linear regression (MLR) is performed on the extracted components $T_1, T_2, ..., T_k$ of the original y-variables. Multiple linear regression provides estimates

$$\begin{aligned}
\boldsymbol{\hat{A}_{k,q}} &= \boldsymbol{(S_t)^{-1}S_{ty}} = \boldsymbol{(R_{k,p}'S_xR_{p,k}')^{-1}R_{k,p}'S_{xy}} \\
\hat{\alpha}_0 &= \bar{y} - \hat{A}_{q,k}'\bar{\tilde{t}} \\
\boldsymbol{S_f} &= \boldsymbol{S_y - \hat{A}_{q,k}'S_t\hat{A}_{k,q}}
\end{aligned}$$

where $\boldsymbol{S_y}$ and $\boldsymbol{S_t}$ stand for the empirical covariance matrix of the y-variables and t-variables. Because $\bar{\tilde{t}} = 0$, the intercept α_0 is thus estimated by $\bar{y}$. By plugging in $\boldsymbol{\tilde{t}_i = R_{k,p}'(x_i - \bar{x})}$ in the equation $\boldsymbol{y_i = \bar{y} + A_{q,k}'\tilde{t}_i + f}$, the estimates for the parameters in the original model $y_i = \beta_0 + B_{q,p}'x_i + e_i$ are obtained. These are

$$\begin{aligned}
\boldsymbol{\hat{B}_{p,q}} &= \boldsymbol{R_{p,k}\hat{A}_{k,q}} \\
\boldsymbol{\hat{\beta}_0} &= \boldsymbol{\bar{y} - \hat{B}_{q,p}'\bar{x}}
\end{aligned}$$

Note that for a univariate response variable ($q = 1$), the parameter estimate $\hat{B}_{p,1}$ can be rewritten as the vector $\hat{\beta}$ ([12], [14] and [17]).

3 Experimental Study

We will compare the efficiency of RR, PCR and PLSR by performing a simulation study on simulated data sets. However, emphasis is on the parameter estimation and not on the predictive performance of the methods. The experiments described in this section consider only univariate responses ($q = 1$). Consider the regression model

$$y_i = \beta_0 + B'_{1,p} x_i + e_i$$

with $B'_{1,p} = (\beta_1, \ldots, \beta_p)'$. The regression vector including the intercept is denoted as $\beta = (\beta_0, \beta_1, \ldots, \beta_p)'$.

3.1 Comparing the Performances

The efficiency of the considered methods is evaluated by means of the MSE of the estimated regression paramaters $\hat{\beta}$. It is defined by

$$MSE(\hat{\beta}) = \frac{1}{m} \sum_{l=1}^{m} \left\| \hat{\beta}^{(l)} - \beta \right\|^2$$

where $\hat{\beta}^{(l)}$ denotes the estimated parameter in the l-th simulation. The MSE indicates to what extent the slope and intercept are correctly estimated. Therefore, the goal is to obtain an MSE value close to zero.

3.2 Simulation Settings

For the purpose of comparing the three methods for handling multicollinearity problem, simulated data sets were used in this study. These data consist of $p = 2, 4, 6$ and 50 predictors variables for $n = 20, 30, 40, 60, 80$ and 100. The goal was to develop a linear equation that relates all the predictor variables to a response variables. The data was analyzed using RR, PCR and SIMPLS and the analysis was done using S-Plus software.

The data were constructed as follows

$$\begin{aligned} x_1 &= \mathrm{N}(0, 1) \\ x_{p-1} &= \mathrm{N}(0, 0.1) + x_1 \\ Y &= x_1 + ... + x_p + \mathrm{N}(0, 1) \end{aligned}$$

where $p = 2, 4, 6$ and 50 that represent low, medium (4 and 6) and high number of cases. For each situation, $m = 100$ data sets were generated.

3.3 Simulation Results

To determine whether multicollinearity exists or not, variance inflation factor (VIF) for each predictor for all given cases are computed. VIF is the measure of the speed with which variances and covariances increase and it is the most commonly used method for detecting multicollinearity problem. VIF is a measure of multicollinearity in a regression design matrix (that is, independent variables) in a scaled version of the multiple correlation coefficient between the independent variable, and the rest of the dependent variables. The

measure shows the number of times the variances of the corresponding parameter estimate is increased due to multicollinearity as compared to as what it would be if there were no multicollinearity. Therefore, this diagnostic is designed to indicate the strength of the linear dependencies and how much the variances of each regression coefficient is inflated above ideal [5]. The formula is

$$(VIF)_j = \frac{1}{1 - R_j^2}$$

where R_j^2 is the multiple correlation coefficient and measures the coefficient of correlation between two variables with $-1 < R_j < 1$.

There is no formal cutoff value to use with the VIF for determining the presence of multicollinearity but Neter et al. [3] recommended looking at the largest VIF value. A value greater than 10 is often used as an indication of potential multicollinearity problem.

The VIF values for each predictor for all given cases in this study are greater than 50. This shows that all the regression coefficients $b_1, \ldots, b_p$ appear to be affected by collinearity. The efficiency test of the considered methods is evaluated by means of the estimated regression parameters $\hat{\beta}$. These values indicates to what extent the slope and intercept are correctly estimated. According to Engelen et al. [18], an MSE value close to zero indicates that the slope and intercept are correctly estimated. The results of the simulations are listed in Tables 1 – 4.

Table 1: The efficiency tests for low-number of regressors data sets, $p = 2$.

$\boldsymbol{n}$	20	30	40	60	80	100
RR	4.5158	4.0621	2.2457	1.3620	0.9296	0.7523
PCR	15.3422	8.6804	5.4349	4.5474	2.2499	2.2356
PLSR	15.2835	8.6362	5.4018	4.5244	2.2354	2.2250

Table 2: The efficiency tests for medium-number of regressors data sets, $p = 4$.

$\boldsymbol{n}$	20	30	40	60	80	100
RR	24.06891	13.46989	9.30244	6.27415	5.36279	3.50251
PCR	25.14823	17.46042	12.11719	8.89374	6.15083	4.86143
PLSR	2.62392	1.39834	0.91491	0.49154	0.17670	0.12939

From Table 1 where $p = 2$ and the specified n observations, Ridge Regression performed best compared to the other two methods which gives MSE= 4.52 for $n = 20$ and MSE= 0.75 for $n = 100$, followed by PLS regression with MSE= 15.28 for $n = 20$ and MSE= 2.23 for $n = 100$, and PC regression with MSE= 15.34 for $n = 20$ and MSE= 2.24 for $n = 100$, respectively. The Ridge Regression method is considered the best since it has the lowest MSE values for all specified n observations and the differences from the other two methods are quite big. On the other hand, there is a slight difference in the MSE for PLS and PC regressions which are chosen at $k_{opt} = 1$. The results are consistent for each n specified

Table 3: The efficiency tests for medium-number of regressors data sets, $p = 6$.

$\boldsymbol{n}$	20	30	40	60	80	100
RR	53.36859	26.6182	21.92014	13.13541	9.20576	6.37893
PCR	65.70394	28.93769	21.59502	13.69208	10.31598	7.130633
PLSR	5.35992	2.11927	0.68219	0.37051	0.17128	0.13095

Table 4: The efficiency tests for high-number of regressors data sets, $p = 50$.

$\boldsymbol{n}$	60	80	100
RR	13.13541	9.20576	6.37893
PCR	13.69208	10.31598	7.130633
PLSR	0.37051	0.17128	0.13095

cases. These results showed that for a low number of regressor, $p = 2$, MSE decreases as the number of observations increases.

PLS regression performed better than Ridge Regression and PC regression when the components are chosen at the optimal $k_{opt} = 3$ for $p = 4$ and the specified n observations (Table 2). The MSE values for PLS regression differ a lot from the other two methods where MSE= 2.62 for $n = 20$ and MSE= 0.13 for $n = 100$, MSE= 24.07 for $n = 20$ and MSE= 3.50 for $n = 100$ for RR and MSE= 25.15 for $n = 20$ and MSE= 4.86 for $n = 100$ for PC regression, respectively. The optimal number of components for PLS and PC regression should be chosen at the smallest MSE [18]. These show that both methods performed well with optimal number of components in handling multicolinearity for $p = 4$ regressors.

PLS performed best followed by Ridge Regression and PC regression when the components for both PLS and PC regressions are chosen at the optimal $k_{opt} = 4$ for $p = 6$ and the specified n observations (Table 3). The MSE values for PLS are 5.36 for $n = 20$ and 0.13 for $n = 100$, while for RR, the MSE values for the same n values are 53.37 and 6.38, respectively, and for PC, the MSE values are 65.70 and 7.13. The results also show that MSE values decrease as n increases from 20 to 100. This shows that, as the number of observations becomes higher, the MSE values become smaller compared to a small number of observations. The results are also consistent where PLS performed better than RR followed by PC regression for every specified number of observations.

Ridge Regression performs best followed by PLS and PC respectively when the number of regressors is high, $p = 50$ (Table 4). RR gives MSE values of 1.41, 0.94 and 0.74 for $n = 60, 80$ and 100, respectively, followed by PLS which gives MSE values of 1.92, 1.19 and 0.81 and PC which gives MSE values of 9.35, 9.211 and 9.32.

4 Conclusions

In this paper, the number of dimensions is less than the number of cases. From the tables, it appears that RR and PLSR methods are generally effective in handling multicollinearity

problems in the specified cases with $p = 2, 4, 6$ (for low and moderate number of regressors) and 50 (for high number of regressors). The performances of RR are most efficient than others when $p = 2$ and $p = 50$, while PLSR is most efficient when $p = 4$ and $p = 6$.

The results also show that both PLSR and RR performed better than PCR in all the cases. However, the differences between the PCR performance from PLSR and RR are only slight. These confirmed Rougoor et. al.'s [19] findings that there is no one method that dominates the other, and that the difference between the methods is typically small when the number of observations is large.

In all the cases, it shows that the superior method performed well when the number of observations, n are larger than the number of regressors. It also shows that the results are consistent for every specified number of observations, n that were included in the analysis. Generally, RR is approximately effective and efficient for a small and high number of regressors.

References

[1] B. L. Bowerman and R. T. O' Connell, *Linear Statistical Models an Applied Approach*, 2nd Ed., Boston: PWS-KENT Publishing Co, 1990.

[2] M. S. Younger, *A Handbook for Linear Regression*, USA: DUXBURY Press, 1979.

[3] J. Neter, W. Wasserman and M. H. Kutner, *Applied Linear Regression Models*, 3rd Ed. USA: IRWIN Book Team, 1990.

[4] A. E. Hoerl and R. W. Kennard, *Ridge Regression: Biased Estimation to Nonorthogonal Problems*, Technometrics, 12 (1970), 56 – 67.

[5] R. H. Myers, *Classical and Modern Regression With Applications*, 2nd Ed. USA: PWS-KENT Publishing Company, 1986.

[6] W. F. Massy, *Principal Components Regression in Exploratory Statistical Research*, Journal of the American Statistical Association, 60 (1965), 234 – 246.

[7] R. W. Farebrother, *A Class of Statistical Estimators Related to Principal Components. Linear Algebra and its Aplications*, 289 (1999), 121 – 126.

[8] P. Filzmoser and C. Croux, *A Projection Algorithm for Regression with Collinearity*, In K. Jajuga, A. Sokolowski, H. Bock (ed.): Classification, Clustering and Data Analysis, Berlin: Springer, 227 – 234, 2002.

[9] R. R. Hocking, *Methods and Applications of Linear Models: Regression and the Analysis of Variance*, USA: John Wiley & Sons, 1996.

[10] J. T. Hwang and D. Nettleton, *Principal Components Regression With Data Chosen Components and Related Methods*, Technometrics, 45 (2000), 70 – 79.

[11] Y. L. Xie and J. H. Kalivas, *Local Prediction Models by Principal Component Regression*, Analytica Chimica Acta, 348 (1997), 29 – 38.

[12] P. H. Garthwaite, *An Interpretation of Partial Least Squares*, Journal of the American Statistical Association, 89 (1994), 122 – 127.

[13] H. Wold, *Estimation of Principal Components and Related Models by Iterative Least Squares*, In P. R. Krishnaiah, (ed.): Multivariate Analysis, New York: Academic Press, 391 – 420, 1966.

[14] S. De Jong, *SIMPLS: An Alternative Approach to Partial Least Squares Regression; Chemometrics and Intelligent Laboratory Systems,* 18 (1993), 251 – 263.

[15] M. S. Boeris, J. M. Luco and R. A. Olsina, *Simultaneous Spectrophotometric Determination of Phenobarbital, Phenytoin and Methylphenobarbital in Pharmaceutical Preparations by Using Partial Least Squares and Principal Component Regression Multivariate Calibration*, Journal of Phamaceutical and Biomedical Analysis, 24 (2000), 259 – 271.

[16] H. Abdi, *Partial Least Squares Regression (PLS regression)*, In M. Lewis-Beck, A. Bryman, T. Futing (Eds): Encyclopedia for Research Methods for the Social Sciences. Thousand Oaks (CA): Sage, 2003.

[17] K. V. Branden and M. Hubert, *Robust Methods for Partial Least Squares Regression*, Journal of Chemometrics, 17 (2003), 537 – 549.

[18] S. Engelen, M. Hubert, B. K. Vanden, and S. Verboven, *Robust PCR and Robust PLSR: A Comparative Study. Theory and Applications of Recent Robust Methods*, Edited by M. Hubert, G. Pison, A. Struyf and Van Aelst, Series: Statistics for Industry and Technology, Birkhauser, Basel, 2003.

[19] C. W. Rougoor, R. Sundaram and J. A. M. Van Arendonk, *The Relation Between Breeding Management and 305-day Milk Production, Determined via Principal Components Regression and Partial Least Squares*, Livestock Product Science, 66 (2000), 71 – 83.