Middle-East Journal of Scientific Research 14 (4): 570-579, 2013
ISSN 1990-9233

DOI: 10.5829/idosi.mejsr.2013.14.4.488

Comparison of Shrinkage Regression Methods for Remedy of Multicollinearity Problem

Muhammad Irfan, Maria Javed and Muhammad Ali Raza

Department of Statistics, Government College University, Faisalabad, Pakistan

Abstract: Biased regression methods provide better results as compared to the ordinary least squares (OLS) when the predictors are highly collinear. In the current study, we compared Partial Least Squares Regression (PLSR) with other prediction methods: Ordinary Least Squares (OLS), Ridge Regression (RR) and Principal Component Regression (PCR) to handle the problem of multicollinearity on Gross Domestic Product (GDP) data of Pakistan. All prediction methods have also been compared for efficiency through Root Mean Square Error (RMSE), Root Mean Square Error cross Validation (RMSECV), Cross Validation Parameter (CVP) and R-Square. Overall, it is found in our study that Partial Least Squares Regression (PLSR) provides better prediction as compared to the other prediction methods.

Key words: Multicollinearity · Ordinary Least Squares (OLS) · VIF · Ridge Regression (RR) · Principal Component Regression (PCR) · Partial Least Squares Regression (PLSR) · XLSTAT

INTRODUCTION

Multicollinearity refers to the situation where there is either an exact or approximately exact linear relationship among the explanatory variables Gujarati [1]. In multiple regressions BLUE property is no longer effective in the existence of multicollinearity. When multicollinearity is present in a set of predictors, the ordinary least squares (OLS) estimates of the individual regression coefficients tend to be unstable i,e. "t- ratios" of one or more coefficients tend to be statistically insignificant Chatterjee and Hadi [2] because of its large variance's and covariance's which means the estimates of the parameters tend to be less precise Adnan *et al*. [3] and can guide to wrong inferences. Therefore, the more the multicollinearity the less interpretable are the parameters. In such circumstances, three alternative estimation shrinkage or biased or prediction regression methods: Ridge Regression (RR), Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR) can be used to take more informative results of the data than the ordinary least squares (OLS) method. Although, all three shrinkage regression models are biased (with smaller variances) but tend to have more precision as measured by Mean Square Error (See Hoerl & Kennard [4] and Draper & Smith [5]). The use of all three shrinkage models to alleviate the problem of multicollinearity on real data are very frequently especially on time series data. Brief reviews of findings of some earlier research work are as under:

Yeniay and Goktas [6] compared three regularized regression methods by RMSE and RMSECV on real data. The data consists of the gross domestic product per capita (GDPPC) in Turkey. Ordinary least squares regression (OLS) and Ridge regression (RR) were found to be best because the value of RMSE is minimum while Partial least squares regression (PLSR) is best because RMSECV is minimum. Finally, they concluded in their study Partial least squares regression (PLSR) was the superior model in terms of the prediction ability as compared to the other regularized models.

Adnan *et al*. [3] compared prediction methods Ridge Regression (RR), Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR) using Monte Carlo simulation study taking predictors 2, 4, 6 and 50 and sample size 20, 30, 40, 60, 80 and 100. All simulations and calculations were done using statistical package S-Plus 2000 software. Authors concluded in their study on the basis of the simulation study the Ridge regression (RR) is found to be best for low number of regressors (When sample size lie between 20 to 100), Principal Component Regression (PCR) is best when number of observations are greater than number of regressors in the model and Partial Least Squares Regression (PLSR)

Corresponding Author: Muhammad Irfan, Department of Statistics, Government College University, Faisalabad, Pakistan.

perform better results as compared to the other two prediction methods when we have a large number of regressors.

Maitra and Yan [7] presented two techniques: PCA and PLSR for dimension reduction purpose when regressors are highly correlated. PCA technique is used without the consideration of the correlation while PLSR technique is applied based on the correlation using simulated data. They concluded that PLSR technique is more effective to the PCA technique for dimension reduction purpose.

Ahmad and Gilani [8] compared OLS, RR, PCR and PLSR to handle the problem of multicollinearity. Root Mean Square Error (RMSE), Root Mean Square Error cross Validation (RMSECV), Cross Validation Parameter (CVP) and R-Square techniques are used to compare the efficiency of all prediction methods. The authors used 34 district values for taking infant mortality rate as response variable in Punjab over 1000 live births on six predictor variables in Pakistan. They analyzed their results using UNSCRAMBLER software. They suggest that PLSR is the best model to tackle the problem of multicollinearity among three prediction methods. Kramer [9] discussed the shrinkage properties of PLSR.

Ahmad *et al*. [10] used only Ridge regression (RR) method instead of other prediction techniques to remove the problem of multicollinearity while ridge trace and L-curve methods are used to find out the appropriate value of ridge constant to develop a suitable ridge regression method.

Aldrin [11] described that the biased regression methods on the ordinary least square regression on noisy data when the multicollinearity exists in the explanatory variables. A new biased method was proposed that modified the ordinary least square estimate by adjusting each element of the estimated coefficient vector. The adjusting factors are founded by minimizing the prediction error. Ridge Regression is used as the principal method to find the preliminary estimate called length modified ridge regression. In addition, the length modified principal component regression method is considered, after that the results of this method is compared with the ridge regression, principal component regression and partial least square methods. Length modified method is more suitable for the prediction as compare to the other methods.

The rest of the article is prepared as follows. The second section reviews the methodology and data; the third section presents the observed results and discussion; and the fourth section concludes the study.

Methodology: In this section three prediction methods: Ridge Regression (RR), Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR) are described briefly which are used in our study to remove the problem of multicollinearity.

Ridge Regression (RR): The standardize regression model can be written as

$$Y = \beta_o + Z\beta + \epsilon \tag{2.1}$$

where Y is the $n\times 1$ vector of "n" observations, Z is the $n\times P$ matrix of "n" observations on P regressors, β is the $P\times 1$ vector of regression coefficients, ϵ is the $n\times 1$ vector of random errors with zero mean & variance $\sigma^2 I$.

Ordinary least squares estimators obtained by minimizing the sum of squared residuals as:

$$(Z^T Z)^{-1} Z^T Y = \hat{\beta}_{OLS} \tag{2.2}$$

$Z^T Z$ is ill- conditioned and the variances of the estimators is in Equation (2.2) becomes large Maitra and Yan [7]; multicollinearity will be appear in the data sets in this situation because predictors are highly correlated. To overcome the problem of multicollinearity, Hoerl [12] developed a new methodology which is known as Ridge regression (RR). Hoerl and Kennard [13] proposed ridge regression estimators by a parameter "$k\geq 0$", in the standardize form

$$(Z^T Z + kI)\hat{\beta}_{ridge} = Z^T y \tag{2.3}$$

Thus, the ridge estimate

$$(Z^T Z + kI)^{-1} Z^T y = \hat{\beta}_{ridge} \tag{2.4}$$

Where I is the $P\times P$ identity matrix, $Z^T Z$ is the correlation matrix of regressors and Ridge constant ("k") always lies between 0 and 1. When k = 0, $\hat{\beta}_{ridge}$ & $\hat{\beta}_{OLS}$ estimators are identical. Ridge estimators are referred to as shrinkage estimator because of ridge regression tend to shrink the estimate of the regression coefficients toward zero Chatterjee and Hadi [2]. There are many procedures in the literature for the choice of the ridge constant ("k"). These methods include: Fixed point, Iterative method and ridge trace.

Fixed Point: Hoerl *et al*. [14] proposed fixed point method to search out the best value of ridge constant. It is defined as:

$$k = \frac{P\hat{\sigma}^2(0)}{\sum_{j=1}^{P}\left[\hat{\beta}_j(0)\right]^2} \tag{2.5}$$

Where $\hat{\beta}_1(0), \hat{\beta}_2(0), \ldots \hat{\beta}_P(0)$ are the least squares estimators because ridge constant is zero and $\hat{\sigma}^2(0)$ is the standardize residual mean square.

Iterative Method: Hoerl and Kennard [4] offered iterative point method to find out the best value of ridge constant. Start with the value of k which has already been calculated by fixed point method then determine as:

$$k_1 = \frac{P\hat{\sigma}^2(0)}{\sum_{j=1}^{P}\left[\hat{\beta}_j(k_o)\right]^2} \tag{2.6}$$

Then, (2.6) to compute k_2 as:

$$k_2 = \frac{P\hat{\sigma}^2(0)}{\sum_{j=1}^{P}\left[\hat{\beta}_j(k_1)\right]^2} \tag{2.7}$$

When the difference of estimates is moderately small then stop the iterative procedure.

Ridge Trace: In Ridge regression to find the best value of the ridge constant is determined graphically by ridge trace plot. Ridge trace is a plot of the regression coefficients i,e. $\hat{\beta}_{ridge}$ against the ridge constant i,e. "k" which stars from the value of 0 and ending with 1. "k" is selected when all the coefficients of the estimates are stable Chatterjee and Hadi [2] and Ahmad *et al*. [10].

Principal Component Regression (PCR): Another method to remove the problem of multicollinearity among the predictors is principal component regression. PCR was developed by Massy in 1965. In PCR analysis, original variables are transformed into new orthogonal variables using Jordan decomposition of a matrix (known as principal components i,e. PC's). These new variables (PC's) are independent to each other Ahmad and Gilani [8].

The procedure of finding PCR according to Yeniay & Goktas [6] and Chatterjee & Hadi [2] runs a regression using the appropriate principal components as the regressors with the response variable under study. We can search out the appropriate number of PC's through Eigen values of the correlation matrix (i,e. select those PC's whose Eigen value is greater than one) and scree plot (the suitable number of PC's where the bend changes significantly). The regression model can be written in terms of standardize variables as:

$$\tilde{Y} = \beta_1\tilde{X}_1 + \beta_2\tilde{X}_2 +, \ldots + \beta_P\tilde{X}_P + \in \tag{2.8}$$

Assuming predictors are in standard form, "G" represents the orthogonal matrix. $GG^T = I$ Because "G" is orthogonal and $Z_m = XG$

The model in (2.1) may describe in the form as:

$$Y = \beta_0 + XGG^T\beta + \epsilon \tag{2.9}$$

The model in (2.9) can be written in terms of principal components as:

$$Y = \beta_o + Z_m\delta_m + \epsilon \tag{2.10}$$

Where $\delta_m = G^T\beta = (Z_m{}^T Z_m)^{-1} Z_m{}^T$ and "m" is the number of PC's retained in the model.

Thus, the Principal component estimate

$$\hat{\beta}_{PCR} = G\delta_m \tag{2.11}$$

where "G" is the Eigen vector of first "m" coefficients for principal components. $\hat{\beta}_{PCR}$ & $\hat{\beta}_{OLS}$ estimators are identical if all the PC's are used in the model instead of using first m PC's.

Partial Least Squares Regression (PLRS): PCA technique is used without the consideration of the correlation while PLSR technique is applied based on the correlation. Partial least squares regression (PLSR) is an alternative powerful approach rather than other Ridge regression (RR) and Principal component regression (PCA) techniques; it was originally developed by Swedish Statistician Herman Wold in 1966 (Father of PLS methodology) and Svante Wold (Son of Herman Wold) developed Partial least squares regression (PLSR) which is commonly used inchemo metric (Norgaard *et al*. [15]), chemical engineering (Matria & Yan [7]), Pharmacology (Lobaugh *et al*. [16]) and so many other fields where predictors consist of many different observations. PLSR is especially useful when regressors are highly correlated (i,e. colinearity exist; Car rascal *et al*. [17] even for the more than the number of observations Yeniay & Goktas [6]. Several statistical packages are available to perform PLSR which are: XL-STAT, SAS, SPSS 20 version, S-Plus 2000, MATLAB, Minitab 16 version and

UNSCRAMBLER. Partial least squares regression (PLSR) technique attempts to find a linear decomposition of X and Y such that

$$X = PM^T + A, \qquad Y = QN^T + B$$

where $P_{n\times r}$ and $Q_{n\times r}$ are the matrices of X & Y scores respectively, $M_{p\times r}$ & $N_{1\times r}$ are matrices of the X & Y loadings and $P_{p\times r}$ and $Q_{p\times r}$ are the X & Y residuals.

$$\hat{\beta}_{PLSR} = MN \qquad (2.12)$$

Sufficient literature is available on Partial least squares regression (PLSR) with its properties for example; Wold [18]), Adnan *et al*. [3], Yeniay & Goktas [6], Matria & Yan [7] and Ahmad & Gilani [8] etc.

Model fitting by all Prediction Methods: In this section, we will compare OLS, RR, PCR and PLSR on a real data. The study includes annual time series data of thirty nine years for the period of 1973 to 2011. The sample included total 39 observations consist of the Gross domestic product (real) and several variables affecting GDP in Pakistan (Table 1. for variable description). We estimated our results using STATA version 12 for (transformation purpose), XLSTAT for (Partial least squares regression) and NCCS version 2008 for (Ridge Regression and Principal Component Regression) software's. First of all, normality is checked through different transformations using the statistical package STATA version 12. Logarithm transformation Irfan *et al*. [19] is much better than the other transformations (Figures of the variables are not displayed). In Table 2. Regression coefficients have been calculated by the help of ordinary least squares (OLS) method and their other related information especially for diagnosing of multicollinearity using NCSS version 2008. Both the tolerance and VIF columns, three predictors (FDI, ER and INVGDP) are highly correlated which is the indication of multicollinearity and the INF variable is non collinear. The overall regression model is highly significant at 99.5% while the individual "t" statistic for INF variable is non-significant showing the occurrence of multicollinearity. Analysis of variance results (See in Table 3) also confirm that the overall model is highly significant with a probability of 95%. The results are in line with the findings of Yeniay & Gaoktos [6]. To cope with the problem of multicollinearity prediction methods: Ridge regression, Principal component regression and partial least squares regression are used in this study.

Table 1: Description of the variables

Variables	Description
Y (Response variable)	GDP: Gross domestic product (real)
X_1 (Predictor)	FDI: Foreign direct investment
X_2(Predictor)	ER: Exchange rate
X_3(Predictor)	INF: Inflation rate
X_4(Predictor)	INVGDP: Investment to GDP ratio

Sources: Pakistan Economic survey various issues

International Financial statistics

Hand book of statistics on Pakistan economy

Table 2: OLS regression coefficients of fitting Model (2.1) to the RGDP data and multicollinearity diagnostic

						Multicollinearity Diagnostic	
	Beta	S.E's	"t" statistic	Sig.	R-Square	Tolerance	VIF
Constant	5.555	0.054					
FDI	0.075	0.019	4.029	0.000	0.929	**0.070	*14.209
ER	0.388	0.052	7.398	0.000	0.938	**0.061	*16.263
INF	-0.016	0.019	-0.829	0.413	0.039	0.960	1.0410
INVGDP	0.131	0.059	2.219	0.033	0.974	**0.025	*39.385

*Multicollinearity is present because value of VIF is less than 10. ** Tolerance value is close to zero so predictors are collinear. $VIF = 1/1-R_i^2$ Where i= 1, 2, 3...P (P = no of regressors in the model)

Table 3: Analysis of variance results for RGDP data

Model	Sum of Squares	d.f	Mean Squares	F	Sig.
Regression	2.377	4	0.594	*831.094	0.000
Residual	0.024	34	0.001		
Total	2.401	38			

*Overall model is highly significant

Table 4: Analysis Standardized Ridge estimates $\hat{\beta}_i(k)$ for the RGDP data (1973-2011)

k	FDI	ER	INF	INVGDP
0.0000	0.2570	0.4925	-0.0115	0.2641
0.0010	0.2558	0.4888	-0.0116	0.2686
0.0030	0.2539	0.4821	-0.0117	0.2762
0.0070	0.2519	0.4712	-0.0121	0.2874
0.0080	0.2517	0.4689	-0.0122	0.2896
0.0090	0.2515	0.4666	-0.0122	0.2916
0.0100	0.2513	0.4645	-0.0123	0.2934
0.0200	0.2521	0.4473	-0.0131	0.3060
0.0300	0.2545	0.4346	-0.0139	0.3125
0.0400	0.2572	0.4245	-0.0146	0.3163
0.0500	0.2598	0.4160	-0.0152	0.3186
0.0600	0.2623	0.4087	-0.0158	0.3199
0.0700	0.2644	0.4023	-0.0163	0.3207
0.0800	0.2663	0.3966	-0.0169	0.3210
0.0900	0.2680	0.3915	-0.0174	0.3211
0.1000	0.2695	0.3868	-0.0178	0.3209
0.1443	0.2738	0.3699	-0.0196	0.3189
…	…	…	…	…
Continue up to the value of 1				
1.0000	0.2375	0.2590	-0.0289	0.2525

Table 5: Analysis Variance Inflation factor for different values of k

k	VF_1	VF_2	VF_3	VF_4
0.0000	14.2099	16.2634	1.0414	39.3857
0.0010	13.1169	14.9139	1.0388	35.0782
0.0030	11.3750	12.7735	1.0338	28.3460
0.0070	9.0198	9.9090	1.0240	19.6286
0.0080	8.5832	9.3833	1.0216	18.0822
0.0090	8.1897	8.9113	1.0192	16.7126
0.0100	7.8331	8.4853	1.0169	15.4939
0.0200	5.5110	5.7653	0.9942	8.2611
0.0300	4.2832	4.3816	0.9728	5.1491
0.0400	3.5038	3.5308	0.9523	3.5317
0.0500	2.9562	2.9475	0.9327	2.5848
0.0600	2.5469	2.5197	0.9138	1.9831
0.0700	2.2281	2.1913	0.8956	1.5769
0.0800	1.9726	1.9310	0.8781	1.2897
0.0900	1.7633	1.7198	0.8611	1.0792
0.1000	1.5889	1.5450	0.8447	0.9202
0.1443	1.0759	1.0374	0.7775	0.5337
…	…	…	…	…
1.0000	0.1101	0.1062	0.2484	0.0763

Estimation by Ridge Regression (RR) Method: Following Tables 4 & 5 contains the standardized Ridge estimates $\hat{\beta}_i(k)$ and variance inflation factor (VIF) for different values of k ranging from 0 to 1 for the RGDP data (1973-2011). After fitting regression for different values of the ridge constant "k", different methods as we discussed in the earlier section are used to select the most suitable value of the ridge parameter. A similar observation was made in the study of Chatterjee & Hadi [2].

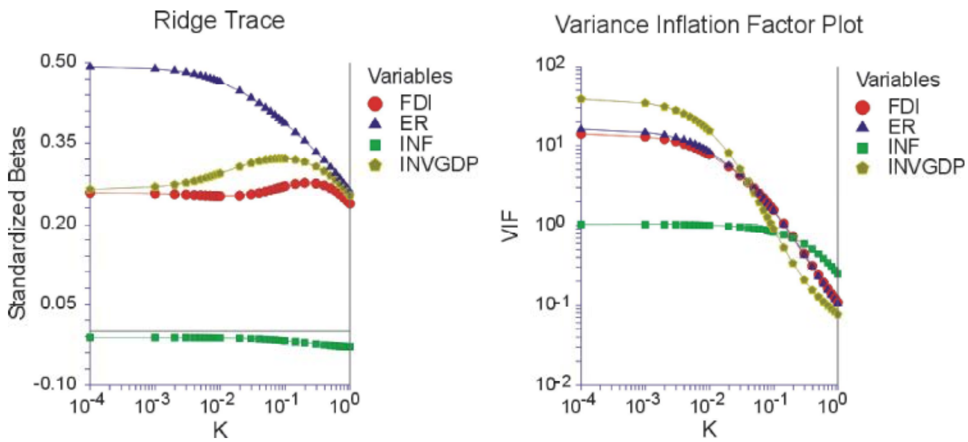


Fig. 1: Ridge trace and Variance inflation factor plot: RGDP (1973-2011)

For RGDP data (1973-2011), the fixed point formula in (2.5) gives

$$k = \frac{4 \times 0.0112}{(0.2570)^2 + (0.4925)^2 + (-0.0115)^2 + (0.2641)^2}, \quad k = 0.1183$$

The iterative method in (2.6) provides the resulting order:
k_o = 0.1183,k_1 = 01443 and k_2 = 01449. So, it converges after two iterations to k = 0.1443. Ridge trace and variance inflation are presented both numerically and graphically in Table 4 & Table 5 and Figure 1 respectively seems to stabilize for k around 0.020. Therefore, we have three estimates of k (0.1183, 0.1433 and 0.020). It is clear that values of k must occur in the interval (0.020 to 0.1183).

The resulting model in terms for the original variables fitted by ridge method using k = 0.020.

$$\text{RGDP} = 5.5956 + 0.0723*\text{FDI} + 0.3406*\text{ER} - 0.0143*\text{INF} + 0.1672*\text{INVGDP} \quad (2.13)$$

Estimation by Principal Component Regression: Second remedy to the problem of multicollinearity that has been used in this study is PCR. Correlation matrix (presented in Table 6) indicates that almost all variables are highly correlated and only INF variable is not correlated to the other variables i,e. produces insignificant results. To perform principal components regression, first of all we have to need the appropriate number of PC's. In Table 7 first two Eigen values covers maximum 97% of the information. It is also confirmed that first two PC's capture the maximum information by the scree plot (which is not shown). Individual contribution of the 3rd & 4th PC's is quite low to explain the variation for response variable (RGDP). If we include all PC's in the model then the results are equivalent to the OLS method as we discussed in the earlier section (2.2). Eigen vector of the correlation matrix is orthogonal to each other which are presented in Table 8. Therefore, two PC's are sufficient to estimate the coefficients of the variables which explain 97% variation for dependent variable as suggested by both Eigen values method and scree plot. Also, the estimates of the standardized and original regression coefficients using the three components models are shown in Table. 9, which clearly gives different results as increase the number of PC's in the model Chatterjee & Hadi [2]. The results of OLS estimates are inadequate when we used all PC's in the model. Two remaining principal component models, first one looks not properly fit because it covers low variation but in the second model R-square increase from 0.9528 to 0.9767, we may conclude strongly that our model is based on only two components.

Table 6: Correlation matrix of the predictive variables

	FDI	ER	INF	INVGDP	RGDP
FDI	1.00000	0.89528	-0.12690	0.95816	0.95244
ER	0.89528	1.00000	-0.17948	0.96370	0.97917
INF	-0.12690	-0.17940	1.00000	-0.16349	-0.17565
INVGDP	0.95816	0.96370	-0.16349	1.00000	0.98686
RGDP	0.95244	0.97917	-0.17565	0.98686	1.00000

Table 7: Eigen values of the correlation matrix

Eigen values	Percentage of Explained variation	Cumulative Percentage of Explained variation
2.916865	72.92	72.92
0.963175	24.08	97.00
0.103247	2.58	99.58
0.016713	0.42	100

Table 8: Eigen vector of the correlation matrix

Eigen vector 1	Eigen vector 2	Eigen vector 3	Eigen vector 4
-0.565249	-0.569010	0.140198	-0.580574
0.112753	0.052758	0.989199	0.077390
0.718222	-0.693886	-0.042552	-0.029474
-0.389792	-0.438140	0.004429	0.809985

Table 9: Comparison of estimated regression coefficients for standardize and original variables taking different number of PC's

	1st PC		1st& 2nd PC's		All PC's	
Variables	Standardize	Original	Standardize	Original	Standardize	Original
Constant	0	5.605741	0	5.595682	0	5.560681
FDI	0.3281	0.065482	0.3353	0.062341	0.2570	0.073763
ER	0.3303	0.394861	0.3337	0.340963	0.4925	0.375025
INF	-0.0814	-0.014863	-0.0180	-0.012041	-0.0115	-0.012533
INVGDP	0.3370	0.255694	0.3420	0.189524	0.2641	0.144332
R-Square	0.9528		0.9767		0.9897	

Now we use OLS method, the model in (2.8) can be written in terms of standardize variables as:

$$\tilde{Y} = 0.2570\tilde{X}_1 + 0.4925\tilde{X}_2 - 0.0115\tilde{X}_3 + 0.2541\tilde{X}_4$$

The principal components of the standardize predictor variables areas under:

$$PC_1 = -0.565249\tilde{X}_1 - 0.569010\tilde{X}_2 + 0.140198\tilde{X}_3 - 0.580574\tilde{X}_4$$

$$PC_2 = 0.112753\tilde{X}_1 + 0.0527558\tilde{X}_2 + 0.989199\tilde{X}_3 + 0.077390\tilde{X}_4$$

$$PC_3 = 0.718222\tilde{X}_1 - 0.693886\tilde{X}_2 - 0.042552\tilde{X}_3 - 0.029474\tilde{X}_4$$

$$PC_4 = -0.389792\tilde{X}_1 - 0.438140\tilde{X}_2 + 0.004429\tilde{X}_3 + 0.809985\tilde{X}_4$$

The final results of PCR using two PC's is presented in original units in Equation (2.14) is

$$GDP = 5.595682 + 0.06234164*FDI + 0.3409633*ER- 0.01204103*INF+0.189524*INVGDP \quad (2.14)$$

Estimation by Partial Least Squares Regression: Standardize variables are used to fit PLSR to the RGDP data; first of all we calculate the weights vectors, interpretation of the weights are very difficult (Table 10) because of the signs of the coefficients. In the second stage weights and standardize variables are used to calculate the loadings for the partial least square components (Table 11.). The results are in line with the findings of Ahmad and Gilani [8]. In third stage we used all four partial least squares components to fit a model. For the optimal number of latent variables in the fitting of PLSR, leave one out method for cross validation has been used on the basis of the PRESS criteria's (Yeniay and Gaoktos [6] and X-variance method (i,e. which decide the optimal number of components in the PLSR model. Where the value of X-variance is closer to one select those appropriate components to fit a PLSR model).Second component gives minimum value of PRESS which is 0.0386068 and using X-variance method it is also confirmed that the value of the X- variance for two component model is 0.995776; almost 100% variation is captured by the regressors.

The final results of PLSR using two components is presented in Equation (2.15) is

$$RGDP=5.631214+0.091467*FDI+0.268771*ER - 0.010459*INF+0.186622*INVGDP\times \quad (2.15)$$

Table 10: PLS weights vectors

	PLS_1	PLS_2	PLS_3	PLS_4
X_1	0.562145	-0.166145	-0.660561	0.469093
X_2	0.577919	0.299813	0.669872	0.356923
X_3	-0.103672	0.938940	-0.328051	-0.005156
X_4	0.582456	0.029997	-0.085517	-0.807794

Table 11: Loadings for PLS Components

	PLS_1	PLS_2	PLS_3	PLS_4
X_1	0.566506	0.06524	-0.702814	0.469093
X_2	0.570048	0.06517	0.637723	0.356923
X_3	-0.128321	1.05385	-0.327586	-0.005156
X_4	0.581669	0.05995	-0.012757	-0.807794

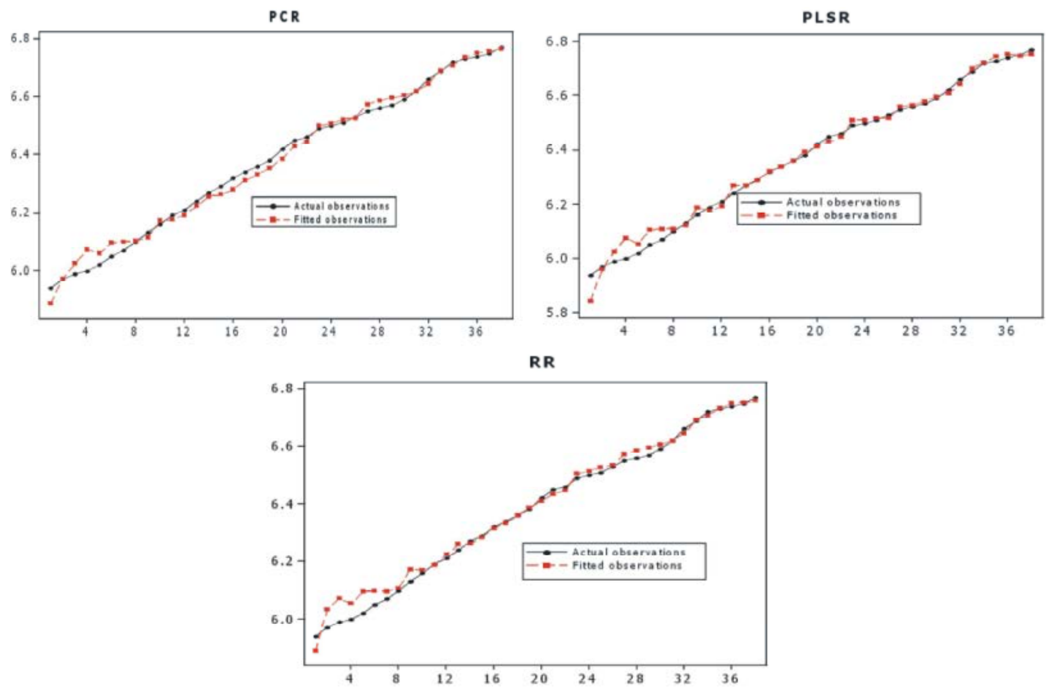
Table 12: Efficiency of prediction models to RGDP data

	RR	PCR	PLSR
*RMSE	0.0353	0.0272	0.0190
*RMSECV	0.0055	0.0042	0.0010
**CVP	0.9131	0.9336	0.9840
**R- Square	0.9824	0.9895	0.9951

Notes: RMSE= $\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}}$, CVP = $Q^2 = \frac{s^2 y - RMSECV}{s^2 y}$, RMSECV= $\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i,k)^2}{n}} = \sqrt{\frac{PRESS}{n}}$

Where S^2y is the un-biased variance of the response variable which is RGDP in our study, PRESS is the prediction error which is calculated through leaves one out cross validation method, *Smaller values of RMSE and RMSECV indicate the fitted model is good, ** Larger value of CVP & R-Square tells us the fitted model is best

Comparison of All Prediction Methods: In this section after estimating the coefficients of three prediction methods the next step is to check the efficiency of all these prediction techniques using different statistical tools like RMSE, RMSECV, CVP & R-Square (Table 12) and graphically (Figure 2 & 3). It can be easily observed from Table 12 that RMSE and RMSECV are smaller for PLSR as compared to the other prediction techniques similarly the largest value of CVP and R-Square clearly indicates that the PLSR model is good. The second smallest values of RMSE and RMSECV and the highest values of CVP and R-Square belong to the PCR model. Overall, we may conclude that PLSR as the best model amongst all other three models: OLS, RR and PCR. In general when there is no multicollinearity in the data sets than PLSR is much better than the others and after this PCR is better than the RR model respectively. It is also observed that after fitting the models successfully on the basis of the variance inflation factors of all the above models, values of VIF's are much smaller when we used PLSR model i,e. multicollinearity problem successfully remove in the RGDP data because the values of VIF's are smaller as compared to the OLS method (Table 2). VIF's values for all four regressors when we fit PCR model are: 5.1189, 4.7772, 1.0402 and 0.1302 and for the fitting of RR model: 5.5110, 7.7653, 0.9942 and 8.2611 which are quite smaller as compared to the OLS method. Statistical as well as graphical results indicate that the PLSR model is superior among others. The actual and fitted observations



*X-axis: Observations Numbers Y- axis: observed and fitted values

Fig. 2: Actual and fitted observations by RR, PCR and PLSR

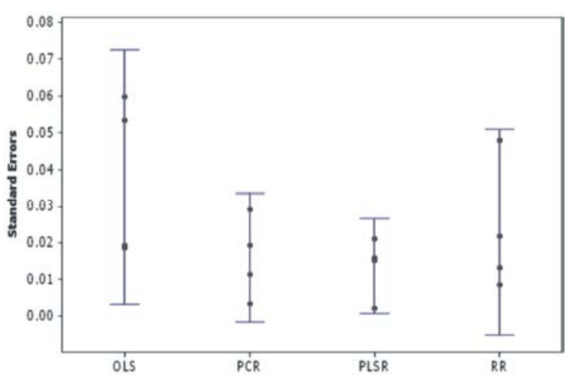


Fig. 3: Standard errors of the coefficients using all predictors using OLS, PCR, RR and PLSR

of all three prediction models RR, PCR and PLSR are plotted in Figure. 2 and the histogram of the residuals of all fitted models (not presented) evidently shows that the PLSR model is the best model to the other models. These results are consistent with the study of Yeniay & Goktas [6] and Ahmad & Gilani [8]. Figure. 3 show that the standard errors of the PLSR regression model are quite low as compared to the others shrinkage regression models and the second good shrinkage regression model is the PCR because its standard errors are small to the RR and OLS models and in the end RR is the best model as compared to the OLS model. Hence the best models in RGDP data are PLSR, PCR and RR respectively after removing the problem of multicollinearity.

CONCLUSIONS

All three shrinkage regression models, Ridge Regression (RR), Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR) provide more informative results as compared to the Ordinary Least Square (OLS) method to handle the problem of multicollinearity on real GDP data in Pakistan when predictors are highly correlated. We have seen that all three shrinkage regression methods provide biased regression coefficients but tend to have more precision as measured by Mean Square Error (MSE). It is found that the best shrinkage regression model is the Partial Least Squares Regression (PLSR) because it provides better results as compared to the other prediction methods on the basis of the Root Mean Square Error (RMSE), Root Mean Square Error cross Validation (RMSECV), Cross Validation Parameter (CVP) and R-Square. In general when there is no multicollinearity in the data set than Partial Least Squares Regression (PLSR) is superior to the others shrinkage models and after this Principal Component Regression (PCR) is better than the Ridge Regression (RR) model.

ACKNOWLEDGEMENTS

The authors are thankful to the anonymous referees whose valuable remarks led to a significant improvement of the paper.

REFERENCES

1. Gujarati, D.N., 2003. Basic Econometrics, 4^{th} ed., McGraw Hill, New York.
2. Chatterjee, S. and A.S. Hadi, 2006. Regression Analysis by Example, Fourth Edition.
3. Adnan, N., M.H. Ahmad and R. Adnan, 2006. A comparative study on some methods for handling multicollinearity problems. Matematika, 22(2): 109-119.
4. Hoerl, A.E. and R.W. Kennard, 1976. Ridge regression: iterative estimation of the biasing parameter", Commun. Statist. Theor. Meth., 5(1): 77-88.
5. Draper, N.R. and H. Smith, 1998. Applied regression analysis. NY: Wiley Inc.
6. Yeniay, O. and A. Goktas, 2002. A comparison of partial least squares regression with other prediction methods. Hacettepe journal of Mathematics and Statistics, 31: 99-111.
7. Maitra, S. and J. Yan, 2008. Principal component analysis and partial least squares: Two dimension reduction techniques for regression. Casualty actuarial society Discussion Paper Program, pp: 79-90.
8. Ahmad, M. and G.M. Gilani, 2010.Comparing Ridge regression (RR), Principal component regression (PCR) and Partial least square regression (PLSR) methods to tackle the multicollinearity problem.
9. Kramer, N., 2007. An overview on the shrinkage properties of partial least squares regression. PhysicaVerlag, An Imprint of Springer- Verlag GmbH, PhysicaVerlag.
10. Ahmad, M., G.M. Gilani and S. Kamal, 2010. Ridge regression model for near collinear factors of infant mortality rate in Punjab province of Pakistan: Evidence from multiple Indicator Cluster Survey, pp: 2003-04.
11. Aldrin, M., 1997. Length modified ridge regression.Computational statistics & Data Analysis, 25: 377-398.
12. Hoerl, A.E., 1962. Application of Ridge analysis to regression problems, Chemical engineering Preogress, 58: 54-59.
13. Hoerl, A.E. and R.W. Kennard, 1970. Ridge regression applications to non- orthogonal Problems, Technometrics, 12(1): 69-82.
14. Hoerl, A.E., R.W. Kennard and K.F. Baldwin, 1975. Ridge regression: some simulations Communications in Statistics, 4(2): 105-123.
15. Norgaard, A., L. Saudland, J. Wagner, J.P. Nielsen, L. Munck and S.B. Engelsen, 2000. Interval partial least squares regression (IPLSR): A comparative chemometric study with an example from infrared spectroscopy. Applied Spectroscopy, 54(3): 413-419.
16. Lobaygh, N.J., R. West and A.R. Mcintosh, 2001. Spatiotemporal analysis of experimental differences in event related potential data with partial least squares. Psychophysiol., 38: 517-530.
17. Carrascal, L.M., I. Galvan and O. Gordo, 2009. Partial least squares regression as an alternative to current regression methods used in ecology. Journal compilation Oikos, 118: 681-690.
18. Wold, H., 1966. Estimation of principal components and related models by iterative least squares. In multivariate analysis, Ed. Krishnaiah, P.R, New York: Academic Press, pp: 391-420.
19. Irfan, M., M. Irfan and M. Awais, 2010. Modeling volatility of short term interest rates by ARCH family models: Evidence from Pakistan and India. World Applied Sciences Journal, 9(10): 1089-1094.