Journal of the Nigerian Association of Mathematical Physics
Volume34,(March, 2016), pp151 – 156

Comparison of Linear Prediction Methodsin Time Series

Onoghojobi B., Olewuezi N. P. and Obite C. P.

Department of Statistics,Federal University of Technology, Akure, Nigeria.

Abstract

In this paper, the relationship between linear prediction methods are being investigated. The main tool for this analysis are the mean squares error and the Normalised root mean square error criteria. The results show that principal component regression and ridge regression methods performed better than the ordinary least square method in the presence of multicollinearity.

Key words: Linear prediction, Normalised root, Mean square,Multicollinearity

1.0 Introduction

In multiple linear regression model, we usually assume that the explanatory variables are independent. However, in practice, there maybe strong or near to strong linear relationships among the explanatory variables. In that case, the independent assumptions are no longer valid, which cause the problem of multicollinearity. In the presence of multicollinearity, it is impossible to estimate the unique effects of individual variables in the regression equation. The variances and covariances of the least squares (LS) estimates becometoo large though still BLUE (best linear unbiased estimator). Multicollinearity becomes one of the serious problems in linear regression analysis [1].

Many attempts have been made to improve the LS estimation procedure, some of which are Ridge Regression, latent root regression, Partial least square, etc. which have smaller mean square error (MSE) than the LS estimators. Principal Component Regression (PCR) deals straightforward with the dependency nature of the explanatory variables. Principal component estimators are obtained by using less than the full set of principal components to explain the variations in the dependent variable [2,3].

PCR has been found to be a procedure that can be used to help circumvent many of the difficulties associated with the usual least squares estimates. This method can be used to obtain a point estimate with a smaller mean square error.

The use of principal components to replace the original regressor variables by their principal components, thus orthogonalizing the regression problem and making computations easier and more stable was suggested in [4]. A technique called supervised principal components that can be applied to solve the problem of multicollinearity has been described in [5]. Supervised principal componentsis similar to conventional principal components analysis except that it uses a subset of the predictors selected based on their association with the outcome. This same method of principal components regression as a solution of multicollinearity was examined in [6]. Using this technique, some fairly precise estimates of the coefficients were obtained. This special property of the principal components regression made it superior to the method of ordinary least squares when multicollinearity is present in the data. Two biased regression methods (Ridge regression and Principal component regression) were proposed in [7,1] when the assumption of general linear regression model – there is no correlation (or no multicollinearity) between the explanatory variables – is violated. In their works, they discussed the theoretical relationship between ridge and principal component regression and a practical comparison through the parameter estimation of land price function. They applied ridge regression and principal component regression to solve one of the most difficult problems of transport and regional analysis with regression model. The performances of both methods were compared using the MSE criterion.

Two of the most widely employed multivariate calibration methods, principal components regression and partial least squares regression (PLS) were comparedin[8,9]. The performances of these methods were compared using simulation data and it was seen that PCR performs better for the first type of data and PCovR performs better for the second type of data.

Instead of comparing two linear models [10] compared the performance of principal component regression (a linear model) with the artificial neural network (ANN) based on visible and shortwave near infrared (VIS-SWNIR) (400-1000 nm) spectra in the non-destructive soluble solids content measurement of an apple.

Corresponding author: Onoghojobi B., E-mail:ngolewe@yahoo.com, Tel.: +2348034933133

A discussion of the collinearity problem in regression and discriminant analysis was presented in[11]. They described the reasons why collinearity is a problem for the prediction ability and classification ability of the classical methods. Some typical ways of handling the collinearity problems based on principal component analysis (PCA) was described and empirical illustrations were given.Different selection rules were provided and compared in [12] and it was shown that it significantly influence the regression results. They also provided a complete derivation of the method used to estimate standard errors of the principal component estimators and the appropriate test statistic, does not depend on the selection rule.
Aregression model to detect apnea sleep disorder by using PCR was presented in[13]. To deal with the problem of multicollinearity, they transformed their 11 non orthogonal independent variables to orthogonal variables (principal component) and used only 10 of the PC so as to obtain a suitable model to detect apnea. The inner egg quality characteristics albumen height, albumen width, albumen length, yolk diameter and yolk height and their estimates were determined and compared with the ordinary least square method [14].Principal component regression model was developed in [15], by combining multiple linear regression and principal component analysis to forecast future water demand in the blue mountains water supply and also the performance of the developed PCR model was compared with multiple linear regression model by adopting several model evaluation statistics such as relative errors, bias, Nash-sutcliffe efficiency and accuracy factor. It was shown that the developed PCR model was able to predict future water demand with a higher degree of accuracy with an average relative error, bias, Nash-Sutcliffe efficiency and accuracy factor values of 3.47%,2.92%, 0.44 and 1.04, respectively than the multiple linear regression models and could be used to eliminate the multicollinearity problem. Bertrand et al [16] showed that the PCR estimator outperforms other regression models and that it fits a significant proportion (10% to 25%) of the between subject variability.
The main contribution of this paper is to establish ridge regression from ordinary least square and to show the comparison between ordinary least square regression, ridge regression and principal component regression in the presence of multicollinearity.

2.0 Ordinary Least Squares Method

Consider the standard model for multiple regression

$$Y = XB + \varepsilon \qquad (1)$$

where
X is (n x p) matrix of predictive variables (henceforth referred to as data-matrix), of rank p, where each row denotes an observation on p different predictive variables, $X_1, X_2, ..., X_p$.
B is (p x 1) matrix of the regression coefficients
ε is the error term.
The usual estimation procedure for the unknown β is Gauss-Markov linear functions of Y that are unbiased and have minimum variance. This estimation procedure is a good one if X^TX, when in the form of a correlation matrix (the data matrix is standardized), is nearly a unit matrix i.e. the column of the data matrix X are not correlated. However, if X^TX is not nearly a unit matrix, the least square estimates are sensitive to a number of "errors" [17].
The OLS coefficient estimators are those formulas or expressions for $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_p$ that minimize thesum of squared residuals (RSS)for any given sample of size N.
The OLSestimation criterionis therefore:

$$\text{Minimize RSS } (\hat{\beta}_0, \hat{\beta}_1, \ldots\ldots, \hat{\beta}_p) = \sum_{i=1}^{N} \hat{\boldsymbol{\varepsilon}}^2 = \sum_{i=1}^{N} (Y - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 \qquad (2)$$

In order to find $\hat{\beta}_i$ we minimise the squared error term $\varepsilon^T\varepsilon$.
Hence the problem is one of minimising:

$$\varepsilon^T\varepsilon = (y - X\beta)^T(y - X\beta) \qquad (3)$$

This term needs to be differentiated with respect to β and set equal to 0 to obtain an estimate of β provided the inverse of X^TX exist. We therefore have:

$$\hat{\beta}_{OLS} = (X^TX)^{-1}X^TY$$

$$Var(\hat{\beta}_{OLS}) = \sigma^2 (X^TX)^{-1}$$

Let $L_1 = \hat{\beta}_{OLS} - \beta$

$$E(L_1^2) = E[(\hat{\beta}_{OLS} - \beta)^T(\hat{\beta}_{OLS} - \beta)]$$

Let $\hat{\beta}_{OLS} - \beta = Z$

$$E(Z) = E(\hat{\beta}_{OLS}) - \beta \text{ hence}$$

$$E(Z) = \beta - \beta = 0$$

$$Var(Z) = Var(\hat{\beta}_{OLS} - \beta)$$

$$= Var(\hat{\beta}_{OLS}) = Var[(X^TX)^{-1}X^TY]$$

Let $(X^TX)^{-1}X^T = K$

$$Var(\hat{\beta}_{OLS}) = Var[KY]$$

$= \sigma^2 (X^TX)^{-1}$

$\therefore Var(\hat{\beta}_{OLS}) = \sigma^2 (X^TX)^{-1}$

$E(L_1^2) = E(Z^TIZ) = Trace[I * Var(Z)] + E(Z^T)IE(Z)$

$= \sigma^2 Trace(X^TX)^{-1}$

2.1 Principal Component Regression

Principal component regression (PCR) is a regression analysis technique that is based on principal component analysis (PCA). Typically, it considers regressing the response variable on a set of covariates.

In PCR, instead of regressing the dependent variable on the explanatory variables directly, the principal components of the explanatory variables are used as regressors. One typically uses only a subset of all the principal components for the regression, thus making PCR some kind of a regularised procedure. Often, the principal components with higher variances (the ones based on eigenvectors corresponding to the higher eigenvalues of the sample variance-covariance matrix of the explanatory variables) are selected as regressors. PCR is one way to deal with the problem of ill conditioned matrices.It is used in overcoming the multicollinearity problem which arises when two or more of the explanatory variables are close to being collinear. PCR can aptly deal with such situations by excluding some of the low-variance principal components in the regression step. In addition, by usually regressing on only a subset of the principal components, PCR can result in dimension reduction through substantially lowering the effective number of parameters characterising the underlying model. This can be particularly useful in settings with high-dimensional covariates. Also, through appropriate selection of the principal components to be used for regression, PCR can lead to efficient prediction of the outcome based on the assumed model.

PCR looks for a few linear combinations of the variables that can be used to summarise the data without losing too much information in the process. It looks for where there is a sharp drop in the component variance.The components with small variance will be omitted. This method reduces the variance when compared to ordinary least square method (OLS) but introduces more bias i.e. it is no longer unbiased.

Decomposing the standardised data matrix X (k X m) using singular value decomposition gives,

$$X = U\sum V^T$$

where

U is a k X m matrix which contains the first m orthonormalised eigenvectors of XX^T and $U^TU = 1$. They are also known as left singular vectors of X.

$\sum$ is a diagonal matrix m X m matrix with positive non-increasing elements. The values of $\sum$ are also known as singular values of X. They are the positive square roots of the eigenvalues of X^TX.

Vis a m X m matrix which contains the orthonormalised eigenvectors of X^TX. These vectors are also known as the right singular vectors of X or just singular vectors of X.

Spectral decomposition of X^TX gives

$$X^TX = (U\sum V^T)^T(U\sum V^T) \quad (4)$$

and

$$X^TX = VDV^T \quad (5)$$

where $D = \sum^T\sum$ is a diagonal m X m matrix of the non-negative eigenvalues of X^TX

$$D = diag(\lambda_1,..., \lambda_m)$$

where $\lambda_1,..., \lambda_m$ are the eigenvalues of X^TX.

Multiplying matrix X with matrix V gives us the principal component

$$XV_m = [XV_1 \; XV_2 \ldots\ldots\ldots\ldots XV_m]$$

2.2 Ridge Regression

Ridge regression (RR) was introduced in[18] as a solution to the problem of unstable ordinary least squares (OLS) estimates under multicollinearity in multiple linear regression. Considerlng the standard linear regression in (1),

the OLS estimator is $\hat{\beta}_{OLS} = (X^TX)^{-1}X^TY$.

Under multicollinearity, when some of the regressors can be expressed as linear combinations of the other variables, the OLS assumption of full rank is not fulfilled, because X^TX approaches singularity and the existence of an inverse to X^TX is not supported. This creates imprecise parameter estimates, with large variances, and accordingly some of the variables might be insignificant under the presence of other covariates, although they do explain variation in the dependent variable in the population. The idea in RR is to adjust the OLS estimator by adding increments to the quantity X^TX, forcing it to be non-singular. By doing so, bias is actually added to the estimator but it becomes more precise in terms of MSE. That is, the RR estimator reduces variance to the cost of increased bias. For the linear regression model defined in (1), the RR estimator is

$$\hat{\beta}_{RR} = (X^TX + KI_p)^{-1}X^TY, \; k \geq 0$$

where k is referred to as the ridge parameter.

There exists $k \geq 0$ such that the MSE of the RR estimator is less than the MSE of the OLS estimator [18].
The relationship between the ridge estimate and the ordinary estimate is given below.

$$(X^TX)\hat{\beta}_{OLS} = X^TY \quad (6)$$

$$(X^TX + KI_p)\hat{\beta}_{RR} = X^TY \quad (7)$$

Combining (10) and (11)

$$(X^TX)\hat{\beta}_{OLS} = (X^TX + KI_p)\hat{\beta}_{RR}$$

$$\hat{\beta}_{OLS} = \frac{(X^TX + KI_p)\hat{\beta}_{RR}}{X^TX}$$

$$\hat{\beta}_{OLS} = (I_p + K(X^TX)^{-1})\hat{\beta}_{RR}$$

$$\hat{\beta}_{RR} = (I_p + K(X^TX)^{-1})^{-1}\,\hat{\beta}_{OLS}$$

$$\hat{\beta}_{RR} = M\hat{\beta}_{OLS}$$

where $M = (I_p + K(X^TX)^{-1})^{-1}$

3.0 Numerical Illustration

The data used in this study is a secondary data for the months of July – September 2014 of Nigerian Ports Authority. Length in meters, max. draught, NRT, GRT and time taken to carry out the operation were used as the independent variables while total tonnages discharged/loaded was used as the dependent variable.
The model involved is defined as

$$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_4X_4 + B_5X_5 + \varepsilon \quad (8)$$

Firstly, the data would be tested if multicollinearity is present or not before OLS, RR and PCR would be used to obtain an estimate of the regression coefficients.

Multicollinearity

Table 1: Multicollinearity Section

Independent Variable	Variance Inflation Factor	R2 Versus Other I.V.'s	Tolerance	Diagonal of X'X Inverse
GRT	8.9392	0.8881	0.1119	5.193565E-09
Length	15.6969	0.9363	0.0637	0.0007587841
Max_Draught	6.6053	0.8486	0.1514	0.2107602
NRT	12.8541	0.9222	0.0778	1.779623E-08
Time	4.1646	0.7599	0.2401	0.004064382

Since some VIF's are greater than 10, multicollinearity is a problem.

Eigenvalues of Centered Correlations

Table 2 shows all the eigenvalues of the correlation matrix, the percentage of the variation explained by each of the eigenvalues and their respective condition number.

Table 2: Eigenvalues of the correlation matrix

No.	Eigenvalue	Incremental Percent	Cumulative Percent	Condition Number
1.	4.4233	88.465	88.465	1.000
2.	0.2502	5.005	93.470	17.676
3.	0.2183	4.366	97.836	20.260
4.	0.0677	1.353	99.190	65.376
5.	0.0405	0.810	100.000	109.155

Since some condition numbers greater than 100,multicollinearity is a mild problem.

Conclusion: The independent variables in the model are collinear since some of the variance inflation factors (VIF) are greater than 10, some of the condition numbers are greater than 100 and most of the variables are correlated with each other from the correlation matrix.

Since multicollinearity is a severe problem in this data, OLS estimate are unstable, all of the regression coefficients are not significant and any prediction made with this model would be unsatisfactory. PCR and RR would be used to correct this problem of multicollinearity.

Ordinary Least Square

The regression model using the OLS estimation method is

Total Tonnages = -19942.087 + .04409 * GRT + 89.4696 * Length + 1432.950 * Max. Draught + .03584 * NRT + 672.095 * Time

Table 3: Principal Component Regression

No.	Eigenvalue	Incremental Percent	Cumulative Percent
1.	4.4233	88.465	88.465
2.	0.2502	5.005	93.470
3.	0.2183	4.366	97.836
4.	0.0677	1.353	99.190
5.	0.0405	0.810	100.000

Only the first two principal components were chosen as the new independent variables to regress against the original dependent variable Y as they explain about 93.47% of the variation in the original data.
The regression model using the Principal Component Regression estimation method is
Total Tonnages = 27363.1282 - 19923.0199 * PC1 - 15524.2462 * PC2

Ridge Regression

The formula proposed in[18] was used to determine the ridge constant to add to the X^TX matrix.

$K = \frac{\hat{\sigma}^2}{B'B} = 0.022238722$

The regression model using the OLS estimation method is
Total Tonnages = -18008.4856 + 76.22959 * GRT + 1409.2204 * Length + 0.0846 * Max. Draught + 0.07751 * NRT + 644.05474 * Time

Table 4: Comparison of the two prediction methods

	OLS Method	PCR Method	RR Method
Mean Square Error	9587952.984	8030347	9535249
Root Mean Square Error	3096.44	2833.787	3087.92

From the table, the mean square error and the root mean square error of the PCR and RR methods are smaller than that of OLSmethod, which means that the biased estimation method performs better when multicollinearity is present and all the other regression assumptions are met.

4.0 Conclusion

In this research work, PCR model was developed by combining multiple linear regression (MLR) and PCA to obtain the relationship between total expenditure and the total breakdown of income. Firstly, the data matrix was standardised and Singular Value Decomposition (SVD) was used to decompose the data matrix X into three different matrices USV^T. The standardised matrix was then multiplied with the matrix of eigenvector V to obtain six principal components; PC1, PC2, PC3, PC4, PC5 and PC6 but only the first five of the principal component was regressed on the dependent variable since they capture most of the variation in the data.
Also, Ridge regression was used to obtain an estimate of the regression coefficients by adding a constant k to the X^TX matrix. The results from the two methods were compared with the result obtained from OLS method of which PCR and RR method performed better based on their mean square error and root mean square error.
PCR and RR, though biased estimation methods are solutions to non orthogonal problems. Using the actual values of the explanatory variable when multicollinearity exists would make the regression coefficient to be unstable, insignificant (small t-ratios), have large variances and covariances, high probability of committing both types 1 and 2 errors and any slight change will totally change the parameters of the data and prediction based on this coefficients would be unsatisfactory, but using the transformed explanatory variable – their corresponding principal component – and adding a constant k to the X^TX matrix would solve most of this problem if not all.

5.0 References

[1] Yazid, M. A. and Mowafaq, M. A. (2009). “a monte carlo comparison between ridge and principal components regression methods”, *Applied Mathematical Sciences*, Vol. 3, Vol. 42, pp. 2085 – 2098.

[2] Kugiumtzis, D., Lingjærde, O. C. and Christophersen, N. (1998). “Regularised local linear prediction of chaotic time series”, *Physica D,* (112) 344-360.

[3] Jolliffe, I. T. (2002). *Principal component analysis,* second edition,springer-Verlag New York, Inc., USA.

[4] Hotelling, H. (1957). “The relations of the newer multivariate statistical methods to factor analysis”, *Brit. J . Stat. Psychol*., Vol. 10, pp. 69-79.

[5] Eric, B., Trevor, H., Debashis, P. and Robert, T. (2006). “Prediction by Supervised Principal Components”, *Journal of the American Statistical Association*, Vol. 101, pp. 119-137.

[6] Saikia, B. and Singh, R. (2012). “Estimation of Principal Components Regression Coefficients”, *International Journal of Science and Research (IJSR)*. Vol. 1, pp. 92-115.

[7] Tsutsumi, M., Shimizu, E. and Matsuba, Y. (1997). "a comparative study on counter measures for multicollinearity in regression analysis", *Journal of the Eastem Asia Society for Transportation Studies*, Vol. 2, pp. 1891-1904.

[8] Peter, D. W. and Lorenzo, V. M. (2003). "Comparison of principal components regression and partial least squares regression through generic simulations of complex mixtures". *Chemometrics and intelligent laboratory system*, Vol. 5, pp. 257-279.

[9] Rocio, M., Miguel, A. P., Cristima, D. T. and Carlos, E. C. (2009). "comparison of principal component regression (pcr) and partial least square (pls) methods in prediction of raw milk composition by vis-nir spectrometry. application to development of on-line sensors for fat, protein and lactose contents", *Fundamental and Applied Metrology*, Vol. 1, pp. 2564-2568.

[10] Kim-seng, C., Herlina, A. R. and Ruzairi, A. R. (2011). "Neural network and principal component regression in non-destructive soluble solids content assessment: a comparison". *Journal of Zhejiang University-Science B (biomedicine & Biotechnology)*, Vol. 4, pp. 218-238.

[11] Tormod, N. and Mervik, B. (2001). "Understanding the collinearity problem in regression and discriminant Analysis", *Journal of Chemometrics*,Vol.15, pp. 413–426.

[12] Fekedulegn, B. D., Colbert, J. J., Hicks, R. R. and Schuckers, M. E. (2002). "Coping with Multicollinearity: An Example on Application of Principal Components Regression in Dendroecology", *usda forest service*, Vol. 1, pp 1-40.

[13] Sidik, M., Sani, M. I., Mohamad, I. F. and Wisnu, J. (2010)." Principal Component Regression Analysis on Automatic Sleep Apnea Detection from ECG Data", *Competence Based Research Grant University of Indonesia*, Vol. 1, pp 193-198.

[14] Tulin, C.R., Fatih, U., Dogan, N. and Tulin, A. (2011). "comparison of principalcomponent regression with the least square method in prediction of internal egg quality characteristics in Japanese quails",*Kafkas Univ Vet Fak Derg*, Vol. 17 (5):pp. 687-692.

[15] Haque, M. M., Rahman, A., Hagare, D. and Kibria, G. (2013). "principal component regression analysis in water demand forecasting: an application to the blue mountains, nsw, Australia". *Journal of Hydrology and Environment Research*, Vol. 1, pp. 49-59.

[16] Bertrand, T., Gael, V., Olivier, G. and Cyril, P. (2014). "Principal Component Regression predicts functional responses across individuals". *Miccai*, Vol. 4, pp. 145-172.

[17] Arthur, E. H. and Robert, W. K. (1970). "Ridge Regression: Biased Estimation for Nonorthogonal Problems" *technometrics*, Vol. 12, pp. 55-67.

[18] Hoerl, A.E. and Kennard R.W. (1970) " Ridge regression: Biased estimation forNonorthogonal problems" Technometrics. Vol. 12, Issue 1, pp. 57-65