*Journal of the Nigerian Association of Mathematical Physics*
***Volume34,(March, 2016), pp141 – 150***


# Estimation of Nonorthogonal Problem Using Time Series Dataset

***Olewuezi N. P., Onoghojobi B. and Bartholomew D.C.***

**Department of Statistics, Federal University of Technology**
**Owerri, Nigeria.**

## *Abstract*

***Based on the presentation of the principles of nonorthogonal problem, we discuss the difference in some of the approaches. A simple procedure to include the R-squared and Root Mean Square Error (RMSE) is proposed and tested. The results showed that the Partial Least Square Regression provides better predictions due to a small RMSE value.***

**Key words:** Nonorthogonal, Mean Square, Partial Least Square, R Square

## 1.0 Introduction

Partial Least Square Regression (PLSR) was first proposed by Herman Wold around 1975 as a method of regressing complicated, collinear, many and even incomplete data sets in terms of chains of matrices [1]. It is becoming a powerful tool to explore the vague relationship between the independent variables and dependent variables. PLSR attempts to extract latent (non-observable) variables so that they collect most of the variation of the real $\theta$ (observable) variables in such a way that they may also be used to model the $\psi$ response (dependent) variable [2].

According to [3], multicollinearity refers to a situation in which two or more predictor variables in a multiple regression model are highly correlated. If multicollinearity is perfect, the regression coefficients are indeterminate and their standard errors are infinite but if it is less than perfect; then regression coefficients although determinate but possess large standard errors which means that the coefficients cannot be estimated with great accuracy [4]. We can define mutlicollinearity through the concept of orthogonality, when the predictors are orthogonal or uncorrelated, all eigenvalues of the design matrix are equal to one and the design matrix is of full rank. If at least one eigenvalue is different from one, especially when equal to zero or near zero, then nonorthogonality exists, meaning that multicollinearity is present [5]. The variables of multivariate partial least square was derived in [1].

The general recommendations for the use of Partial least squares regression, which include: data that is highly collinear, has large predictor variables compared to the number of observations and is not normal was listed in [6]. A comparison of covariance-based and variance-based structural equation models based on their respective variances was highlighted in [7]. Small survey sample size and skewed dataset with partial least square path modelling was handled in [8]. The use of partial least squares equation modelling in marketing data was assessed in [9]. Partial least squares regression with other prediction methods: Ordinary Least Squares (OLS), Ridge Regression (RR) and PCR to handle problem of multicollinearity on Gross Domestic Product (GDP) data of Pakistan was compared in [10]. The use of Partial least squares (PLS) structural equation modelling for building and testing behavioural causal theories was illustrated in [11]. The use and the misuse of structural equation modelling in management research was written in [12]. The inconsistency of Partial least squares path coefficient estimates in the case of reflective measurement can have adverse consequences for hypothesis testing as was shown in [13]. Thus consistent PLS provides correction for estimates when PLS is applied to reflective constructs. An alternative approach based on multitrait-multimethod matrix to assess discriminant validity since other approaches like Larcker criterion and examination of cross-loadings do not reliably detect the lack of discriminant validity in common research situations was proposed in [14].

The PLS methodology has also achieved an increasingly popular role in empirical research in international marketing, which may represent an appreciation of distinctive methodological features of PLS.

This paper focuses on using the R squared and Root mean square error for estimation of nonorthogonal problems and the performance is evaluated in comparison with the method involved. The paper proposes a simple method to incorporate times

Corresponding author: Olewuezi N. P., E-mail:ngolewe@yahoo.com, Tel.: +2348034933133

series dataset in Nonorthogonal problem.
Section 2 describes the definition and algorithms of Ordinary least square. Ridge, regression and partial least square regression. Section 3 introduces the applicability of the method involved in the Nonorthogonality with illustrative example. Discussions on the differences between the regression methods is hereby presented.

## 2.0 The Ordinary Least Square Model (OLS)

The OLS general model is defined as

$\psi = \theta$ B + E (matrix notation) (1)

where $\psi$ is n×1 vector of observations on the dependent variable.

$\theta$ is n×p matrix of predictors.
B is p×1 vector of regression parameters.
E is n×1 vector of errors.
When the matrix $\theta$ has a full rank of p, the OLS estimator $B_{OLS}$can be obtained by minimizing the sum of squared residuals, hence

$$B_{OLS} = (\theta'\theta)^{-1}\theta'\psi \quad (2)$$

where $B_{OLS}$ is pX1 vector of OLS estimated parameters.
Using unbiased linear estimation with minimum variance or maximum likelihood estimation when the random vector E is normal gives (2) as an estimate of B and this gives minimum sum of squares of the residuals.

## 2.1 Partial Least Square Regression (PLSR) for Nonorthogonal Problems

In general, according to [15] the variables of a multivariate PLS is derived as follows:

$$\theta = TP^t + \delta \quad (3)$$

$$\psi = UQ^t + \omega \quad (4)$$

where $\theta$ is n×m matrix of predictors
$\psi$ is n×p matrix of responses
T and U are n×l matrices which are respectively the projection of $\theta$ and $\psi$
P and Q are respectively m×1 and p×1 orthogonal loading matrices and
$\delta$ and $\omega$ are their respective error terms.

Formally, the pattern of relationships between the columns of $\theta$ and $\psi$ is stored in mxn cross-product matrix, denoted by $\xi$ (that is, usually a correlation matrix in that we compute it $\theta_0$ and $\psi_0$ instead of $\theta$ and $\psi$ ). $\xi$ is computed as:

$$\xi = \theta^T{}_o\psi_o \quad (5)$$

The main analytical tool for PLS is the singular value decomposition (SVD) of a matrix. The matrix of correlations (or covariance) between $\theta_o$ and $\psi_o$ (mean-centered and normalized variables) is computed as

$$\xi = W\Delta C^T \quad (6)$$

where

W is nxn eigenvectors of $RR^T$

C is mxm eigenvectors of $R^T R$ and

$\Delta$ is an nxm diagonal matrix made up of the square roots of the non-zero eigenvalues of both $RR^T$ and $R^T R$.
The first pair of singular vector (that is the first columns of W and C) are denoted by $w_1$ and $c_1$ and the first singular value (that is the first diagonal entry) is denoted by $\Delta_1$.
The first latent variable of $\theta$ is given by

$$t_1 = \theta_0{}^T w_1 \quad (7)$$

$$p_1 = \theta^T{}_0 t_1 \quad (8)$$

$$\hat{\theta}_1 = t_1 p^T \quad (9)$$

The first pseudo latent variable for $\psi$ is given by

$$U_1 = \psi_0 c_1$$

Reconstituting $\psi$ from its pseudo latent variable as

$$\hat{\psi}_1 = u_1 c_1^T \tag{10}$$

$$\psi = t_1 b_1 c_1^T \tag{11}$$

With $b_1 = t_1^t u_1$ (first regression coefficient) (12)

Matrices $\hat{\theta}_1$ and $\hat{\psi}_1$ are then subtracted from the original $\theta_o$ and $\psi_o$ respectively

That is $\theta_1 = \theta_0 - \hat{\theta}_1$ and $\psi_1 = \psi_0 - \hat{\psi}_1$

The iterative process continues until $\theta$ is completely decomposed into L components (that is, $\theta$ is now a null matrix and L is the rank of X). When this happens, we have succeeded in obtaining all the latent variables for the model.
The predicted $\psi$ scores are now given by

$$\psi = TBC^T = \theta B_{PLS} \tag{13}$$

where

$$B_{PLS} = (P^{T+})\,BC^T \tag{14}$$

$(P^{T+})$ is the Moore – penrosepseudoinverse of $P^T$

That is, $(P^{T+}) = (AA^T)^{-1}A$ (15)

## 2.2 Ridge Regression (RR)

The standard regression model can be written as

$$\psi = \theta\beta + \varepsilon \tag{16}$$

where $\psi$ is the n × 1 vector of "n" observations,

$\theta$ is the n × p matrix

$\beta$ is a p × 1 vector of regression coefficients

$\varepsilon$ is the n ×1 vector of random errors with zero mean and variance $\sigma^2 I$

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

$$\left(\theta^T\theta\right)^{-1}\theta^T\psi = \hat{\beta}_{OLS} \tag{17}$$

$$Var(\hat{\beta}) = \hat{\sigma}^2\left(\theta^T\theta\right)^{-1} \tag{18}$$

$$\text{M.S.E}\left(\hat{\beta}\right) = \hat{\sigma}^2 trace\left(\theta^T\theta\right)^{-1} = \hat{\sigma}^2\sum_{i=1}^{p}\lambda_i \tag{19}$$

where $\hat{\sigma}^2$ is the mean squares error. This estimator $\hat{\beta}$ is unbiased and has a minimum variance. However, if $\theta^T\theta$ is ill-conditioned (singular), the OLS estimate tends to become too erratic and some of the coefficients have wrong sign [16]. In order to prevent these difficulties of O.L.S, ridge regression as an alternative procedure to the OLS method in regression analysis was suggested in [17]. The ridge technique is based on adding a biasing constants IK's to the diagonal of Benson matrix before computing $\hat{\beta}'s$ by using method of [17]. Therefore, the ridge solution is given by:

$$\beta_{ridge} = \left(\theta^T\theta + KI\right)^{-1}\theta^T\psi\ ,\ K \geq 0 \tag{20}$$

**Iterative Method for estimating K**

Iterative point method to find out the best value of ridge constant was given in [17]. Start with the value of k which has already been calculated by fixed point method then determine as:

$$k_1 = \frac{p\hat{\sigma}^2(0)}{\sum_{j=1}^{p}\left[\hat{\beta}_j(k_0)\right]^2} \tag{21}$$

Then compute $k_2$ as

$$k_2 = \frac{p\hat{\sigma}^2(0)}{\sum_{j=1}^{p}\left[\hat{\beta}_j(k_1)\right]^2} \tag{22}$$

where p is the number of regressor variables

$\hat{\beta}_j$ is the jth regression parameter estimate.

$\hat{\sigma}^2(0)$ is the standardized residual mean square.

$k_o$ is the arbitrarily chosen K value.

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

## 3.0 Results and Discussions

### 3.1 Illustrative Example

Nigeria Insurance company expenditure dataset were used (N' Million). Let the independent variables be claims, Fire, Accident, Motor, Employers Liability, Marine and Miscellaneous while the dependent variable is the Total expenditure.

**Table 1:** Correlation Matrix of the Insurance Company dataset

| Variables | Claims | Fire | Accident | Motor | Employers | Marine | Misce | Total Expenditure |
|---|---|---|---|---|---|---|---|---|
| Claims | **1.000** | 0.801 | 0.973 | 0.984 | 0.931 | 0.954 | 0.818 | **0.968** |
| Fire | 0.801 | **1.000** | 0.664 | 0.802 | 0.560 | 0.623 | 0.330 | **0.863** |
| Accident | 0.973 | 0.664 | **1.000** | 0.957 | 0.965 | 0.970 | 0.888 | **0.918** |
| Motor | 0.984 | 0.802 | 0.957 | **1.000** | 0.901 | 0.917 | 0.764 | **0.984** |
| Employers | 0.931 | 0.560 | 0.965 | 0.901 | **1.000** | 0.991 | 0.949 | **0.832** |
| Marine | 0.954 | 0.623 | 0.970 | 0.917 | 0.991 | **1.000** | 0.930 | **0.863** |
| Misce- | 0.818 | 0.330 | 0.888 | 0.764 | 0.949 | 0.930 | **1.000** | **0.680** |
| **Total Expenditure** | **0.968** | **0.863** | **0.918** | **0.984** | **0.832** | **0.863** | **0.680** | **1.000** |

**Table 2:** Multicollinearity Statistics

| **Multicolinearity statistics:** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Statistic** | **Claims** | **Fire** | **Accident** | **Motor** | **Employers** | **Marine** | **Misce** |
| Tolerance | 0.000 | 0.002 | 0.004 | 0.002 | 0.004 | 0.005 | 0.002 |
| VIF | 4406.366 | 468.931 | 233.686 | 486.067 | 255.674 | 185.715 | 580.127 |

The correlation matrix showed that there is a perfect correlation between each variable and itself. The correlation values for between each variable and others are significantly closer to unity indicating a positive relationship and high collinearity amongst variables. Looking at both the tolerance and VIF rows in the multicollinearity diagnostic table, all the independent variables are significantly highly collinear since the VIF and tolerance values are greater than 10 and closer to zero respectively. The determinant of the correlation matrix above was calculated to be 0.0000000000696125 indicating the extreme dependency among variables.

**Ordinary Least Squares Result**

**Table 3:**Analysis of variance

| **Analysis of variance** | | | | | |
|---|---|---|---|---|---|
| **Source** | **DF** | **Sum of squares** | **Mean squares** | **F** | **Pr> F** |
| Model | 7 | 9166191209.744 | 1309455887.106 | 153.789 | < 0.0001 |
| Error | 8 | 68117032.630 | 8514629.079 | | |
| **Corrected Total** | **15** | **9234308242.374** | | | |

The Analysis of Variance for the O.L.S.R above yield a significant regression model since the probability value is less than 0.05 level of significance. The mean squares are very high due to the presence of multicollinearity.

**Table 4:**Model parameters

| Source | Value | Std error | t | Pr> \|t\| | Lower bound (95%) | Upper bound (95%) |
|---|---|---|---|---|---|---|
| Intercept | 1346.411 | 2425.234 | 0.555 | 0.594 | -4246.189 | 6939.012 |
| Claims | 2.868 | 2.355 | 1.218 | 0.258 | -2.562 | 8.298 |
| Fire | -3.789 | 3.977 | -0.953 | 0.369 | -12.961 | 5.383 |
| Accident | -2.211 | 2.951 | -0.749 | 0.475 | -9.016 | 4.595 |
| Motor | 0.779 | 2.751 | 0.283 | 0.784 | -5.565 | 7.122 |
| Employers | -17.477 | 57.365 | -0.305 | 0.768 | -149.762 | 114.807 |
| Marine | -1.367 | 3.986 | -0.343 | 0.741 | -10.559 | 7.826 |
| Misce- | -3.196 | 3.415 | -0.936 | 0.377 | -11.070 | 4.678 |

**Partial Least Square Regression Result**

**Table 5:**Model quality

| **Model quality** | | | |
|---|---|---|---|
| Index | Comp 1 | Comp 2 | Comp 3 |
| $Q^2$ cum | 0.859 | 0.941 | 0.976 |
| $R^2Y$ cum | 0.894 | 0.967 | 0.988 |
| $R^2X$ cum | 0.871 | 0.985 | 0.994 |

Consider as PLSR model quality. We should be interested in the quality of prediction we want to achieve. Perhaps we ought to use the latent variable that explains much information about the Y variable. Comp 1, Comp 2, Comp 3 and Comp 4 has the following $Q^2_l$ values as 0.859, 0.584, 0.590 and 0.03 respectively. This automatically makes Comp1 the chief latent variable and Comp 4 the least latent variable. Consequently, we are supposed to drop Comp 4 because it only explained 3% of information in Y variable and dropping it will not cause much harm as seen in the cumulative $Q^2_l$ value table. We still had a cumulative value of 0.976 with or without Comp 4.

**Table 6:** Model parameters

| Variable | Total Expenditure |
|---|---|
| Intercept | -20.656 |
| Claims | 0.339 |
| Fire | 0.527 |
| Accident | 2.115 |
| Motor | 2.756 |
| Employers | -16.801 |
| Marine | -1.517 |
| Misce- | -0.540 |

**Ridge Regression**

**Table 7:** Variance Inflation Factor Section

| **Variance Inflation Factor Section** | | | | | | | |
|---|---|---|---|---|---|---|---|
| **k** | **Claims** | **Fire** | **Accident** | **Motor** | **Employers** | **Marine** | **Miscellaneous** |
| 0.000 | 4406.366 | 468.9307 | 233.6862 | 486.0673 | 255.674 | 185.7148 | 580.1272 |
| 0.001 | 89.6244 | 18.5617 | 42.2834 | 35.1855 | 73.742 | 89.7348 | 34.5996 |
| 0.002 | 29.4692 | 10.783 | 35.4836 | 26.3248 | 54.8802 | 65.6983 | 24.5075 |
| 0.003 | 16.0242 | 8.3782 | 31.4218 | 22.7937 | 43.5515 | 50.7256 | 20.7886 |
| 0.004 | 10.7703 | 7.1223 | 28.2625 | 20.4624 | 35.8502 | 40.6236 | 18.4471 |
| 0.005 | 8.0919 | 6.3167 | 25.6452 | 18.6638 | 30.2876 | 33.448 | 16.6858 |
| 0.005 | 8.0919 | 6.3167 | 25.6452 | 18.6638 | 30.2876 | 33.448 | 16.6858 |
| 0.006 | 6.4929 | 5.7421 | 23.4198 | 17.1853 | 26.0965 | 28.1478 | 15.2579 |
| 0.007 | 5.4336 | 5.3045 | 21.4994 | 15.9302 | 22.8349 | 24.1087 | 14.0554 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0.008 | 4.6787 | 4.9562 | 19.8249 | 14.8435 | 20.2303 | 20.9509 | 13.0196 |
| 0.009 | 4.1113 | 4.6696 | 18.3528 | 13.89 | 18.1061 | 18.4287 | 12.1138 |
| 0.01 | 3.6675 | 4.4279 | 17.0499 | 13.045 | 16.3429 | 16.3775 | 11.3132 |
| 0.02 | 1.7164 | 3.0933 | 9.3801 | 7.9295 | 7.7244 | 7.0838 | 6.5282 |
| 0.03 | 1.0602 | 2.458 | 6.0461 | 5.521 | 4.6825 | 4.178 | 4.362 |
| 0.04 | 0.7357 | 2.06 | 4.268 | 4.1349 | 3.1929 | 2.8335 | 3.1747 |
| 0.05 | 0.5476 | 1.7823 | 3.1963 | 3.2436 | 2.338 | 2.0807 | 2.447 |
| 0.06 | 0.4279 | 1.5765 | 2.4952 | 2.6283 | 1.7975 | 1.6087 | 1.9656 |
| 0.07 | 0.3466 | 1.4179 | 2.009 | 2.1818 | 1.4322 | 1.2899 | 1.629 |
| 0.08 | 0.2888 | 1.2917 | 1.6566 | 1.8458 | 1.173 | 1.0627 | 1.3834 |
| 0.09 | 0.2461 | 1.1891 | 1.3924 | 1.5855 | 0.982 | 0.8943 | 1.198 |
| 0.1 | 0.2135 | 1.1038 | 1.1887 | 1.3793 | 0.8369 | 0.7655 | 1.0541 |
| 0.2 | 0.0913 | 0.6754 | 0.4008 | 0.5175 | 0.2957 | 0.2722 | 0.4752 |
| 0.3 | 0.0614 | 0.5023 | 0.2095 | 0.284 | 0.168 | 0.1515 | 0.3124 |
| 0.4 | 0.0485 | 0.4012 | 0.1339 | 0.1862 | 0.1163 | 0.1026 | 0.2352 |
| 0.5 | 0.0413 | 0.3326 | 0.0961 | 0.1353 | 0.0892 | 0.0774 | 0.189 |
| 0.6 | 0.0367 | 0.2822 | 0.0744 | 0.1051 | 0.0727 | 0.0625 | 0.1578 |
| 0.7 | 0.0333 | 0.2435 | 0.0606 | 0.0855 | 0.0617 | 0.0527 | 0.135 |
| 0.8 | 0.0307 | 0.2128 | 0.0512 | 0.0719 | 0.0538 | 0.0458 | 0.1176 |
| 0.9 | 0.0286 | 0.1879 | 0.0446 | 0.062 | 0.0478 | 0.0407 | 0.1039 |
| 1 | 0.0269 | 0.1674 | 0.0395 | 0.0545 | 0.0432 | 0.0368 | 0.0927 |

Table 7 shows the VIF values for the several independent variables at different k trial values. Observe that the multicollinearity reduces with increased k value since the VIF's kept reducing.

**Table 8:**Ridge vs. Least Squares Comparison Section for k = 0.005000

| Independent Variable | Regular Ridge Coeff's | Regular L.S. Coeff's | Standardized Ridge Coeff's | Standardized L.S. Coeff's | Ridge Standard Error | L.S. Standard Error |
|---|---|---|---|---|---|---|
| Intercept | 1493.466 | 1346.411 | | | | |
| Claims | 0.494533 | 2.867821 | 0.4233 | 2.4548 | 0.134311 | 2.354801 |
| Fire | 0.446711 | -3.78929 | 0.0739 | -0.6264 | 0.614432 | 3.977479 |
| Accident | 0.872502 | -2.21072 | 0.1372 | -0.3477 | 1.301242 | 2.9512 |
| Motor | 2.939932 | 0.778697 | 0.7155 | 0.1895 | 0.717422 | 2.750741 |
| Employers | -37.0777 | -17.4773 | -0.3138 | -0.1479 | 26.27917 | 57.36528 |
| Marine | -0.59686 | -1.36654 | -0.062 | -0.1419 | 2.251593 | 3.986157 |
| Miscellaneous | -0.0153 | -3.19616 | -0.0033 | -0.6846 | 0.770782 | 3.414654 |

At k value of 0.005, the Least square standard errors for parameter estimates were reduced as compared above since RR reduced the effect of multicollinearity.

**Table 9:** Ridge Regression Coefficient Section for k = 0.005000

| **Independent Variable** | **Regression Coefficient** | **Standard Error** | **Standardized Regression Coefficient** | **VIF** |
|---|---|---|---|---|
| Intercept | 1493.466 | | | |
| Claims | 0.494533 | 0.134311 | 0.4233 | 8.0919 |
| Fire | 0.4467107 | 0.614432 | 0.0739 | 6.3167 |
| Accident | 0.8725019 | 1.301242 | 0.1372 | 25.6452 |
| Motor | 2.939932 | 0.7174224 | 0.7155 | 18.6638 |
| Employers | -37.07767 | 26.27917 | -0.3138 | 30.2876 |
| Marine | -0.596858 | 2.251593 | -0.062 | 33.448 |
| Miscellaneous | -0.01530452 | 0.7707824 | -0.0033 | 16.6858 |

Notice that the VIF values are now smaller, which means that the multicollinearity has reduced due to Ridge regression parameter estimates.

**Table 10:** Analysis of Variance Section for k = 0.005000

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F-Ratio** | **Prob Level** |
|---|---|---|---|---|---|
| Intercept | 1 | 9.64E+09 | 9.64E+09 | | |
| Model | 7 | 9.11E+09 | 1.30E+09 | 86.314 | 0.000001 |
| Error | 8 | 1.21E+08 | 1.51E+07 | | |
| Total (Adjusted) | 15 | 9.23E+09 | 6.16E+08 | | |

**Table 11:** Comparison of the different estimation methods.

| | | **RR** | | **OLSR** | | **PLSR** | |
|---|---|---|---|---|---|---|---|
| **Year** | **Actual** | **Predicted** | **Residual** | **Predicted** | **Residual** | **Predicted** | **Residual** |
| 1996 | 5,916.14 | 2,662.29 | | | | | |
| | | | 3,253.85 | 3,276.22 | 2,639.92 | 2,020.39 | 3,895.75 |
| 1997 | 6,499.40 | 3,483.50 | 3,015.90 | 3,658.44 | 2,840.96 | 2,801.59 | 3,697.81 |
| 1998 | 7,174.28 | 3,874.73 | 3,299.56 | 3,844.41 | 3,329.87 | 3,024.78 | 4,149.50 |
| 1999 | 5,923.18 | 7,502.31 | -1,579.13 | 8,367.28 | - 2,444.10 | 7,560.00 | - 1,636.82 |
| 2000 | 5,629.52 | 6,330.17 | -700.65 | 6,009.15 | - 379.63 | 5,857.19 | - 227.67 |
| 2001 | 6,110.52 | 7,286.43 | -1,175.91 | 8,354.04 | - 2,243.52 | 7,241.56 | - 1,131.04 |
| 2002 | 6,856.15 | 9,433.87 | -2,577.73 | 9,370.65 | - 2,514.51 | 7,478.80 | - 622.65 |
| 2003 | 9,415.20 | 12,359.90 | -2,944.70 | 12,035.07 | - 2,619.87 | 12,643.05 | - 3,227.85 |
| 2004 | 12,084.04 | 13,535.59 | -1,451.55 | 12,176.93 | - 92.89 | 15,074.19 | - 2,990.15 |
| 2005 | 12,402.40 | 16,106.77 | -3,704.37 | 13,689.05 | - 1,286.65 | 17,336.87 | - 4,934.47 |
| 2006 | 76,276.11 | 76,001.85 | 274.25 | 75,950.58 | 325.53 | 75,438.15 | 837.96 |
| 2007 | 25,133.24 | 22,825.94 | 2,307.30 | 24,002.70 | 1,130.54 | 24,054.89 | 1,078.35 |
| 2008 | 37,412.55 | 36,345.20 | 1,067.35 | 35,164.29 | 2,248.26 | 37,686.21 | - 273.66 |
| 2009 | 61,969.15 | 61,528.85 | 440.30 | 61,511.82 | 457.33 | 60,609.04 | 1,360.11 |
| 2010 | 53,815.35 | 55,812.01 | -1,996.66 | 56,556.87 | - 2,741.52 | 5,942.62 | - 2,127.27 |
| 2011 | 60,204.76 | 57,732.58 | 2,472.18 | 58,854.50 | 1,350.26 | 58,052.66 | 2,152.10 |

**Table 12:** Prediction error comparison of the different methods.

| Year | RR | OLS | Difference | RR | PLSR | Difference |
|---|---|---|---|---|---|---|
| 1996 | 3253.846 | 2639.918 | - | 3253.846 | 3895.753 | - |
| 1997 | 3015.896 | 2840.962 | - | 3015.896 | 3697.814 | - |
| 1998 | 3299.555 | 3329.875 | + | 3299.555 | 4149.498 | - |
| 1999 | -1579.13 | -2444.100 | - | -1579.13 | -1636.815 | + |
| 2000 | -700.651 | -379.630 | + | -700.651 | -227.669 | - |
| 2001 | -1175.91 | -2243.516 | - | -1175.91 | -1131.035 | - |
| 2002 | -2577.73 | -2514.508 | + | -2577.73 | -622.652 | - |
| 2003 | -2944.7 | -2619.866 | + | -2944.7 | -3227.852 | + |
| 2004 | -1451.55 | -92.887 | + | -1451.55 | -2990.152 | + |
| 2005 | -3704.37 | -1286.655 | + | -3704.37 | -4934.471 | + |
| 2006 | 274.2549 | 325.534 | + | 274.2549 | 837.964 | - |
| 2007 | 2307.304 | 1130.544 | - | 2307.304 | 1078.349 | + |
| 2008 | 1067.349 | 2248.257 | + | 1067.349 | -273.661 | + |
| 2009 | 440.2982 | 457.328 | + | 440.2982 | 1360.109 | - |
| 2010 | -1996.66 | -2741.522 | - | -1996.66 | -2127.274 | + |
| 2011 | 2472.178 | 1350.265 | - | 2472.178 | 2152.097 | + |

Here, we try to compare the precision of the methods through their prediction residuals. A positive sign means that RR performed better than OLSR while a positive sign in the other end means that PLSR performed better than RR. The signs are as a result of the difference between errors due to prediction results by the methods being compared. Notice that RR performed better than OLSR but PLSR performed equally the same way as RR on the basis of the plus and minus signs. This leaves us with no doubt that PLSR performed better than OLSR.

**Table 13:** General Comparison

| | OLSR | RR | PLSR |
|---|---|---|---|
| **R-Squared** | 0.993 | 0.9869 | 0.988 |
| **M.S.E** | 8,514,629.079 | 15,083,863.6 | 6,695,165.045 |
| **R.M.S.E** | 2917.984 | 3883.795 | 2587.502 |

When model fitting is the aim and not prediction, we observe that O.L.S.R gave higher R-squared value than PLSR and RR This means that OLSR fits the data well even in the presence of multicollinearity but failed to predict better than PLSR since its RMSE value is higher compared to PLSR Note, not all high values of R-square indicates good model fit because R-squared values increases monotonously with the variables, whether important or irrelevant for the prediction of the dependent variable.

## 4.0 Conclusion

The PLSR model provided more précised prediction as compared with the OLSR and RR methods to handle the problem of multicollinearity on Nigeria Insurance company’s expenditure data when predictors are highly correlated.

## 5.0 Recommendation

It is highly recommended that a correlation matrix and collinearity diagnostics be computed on any data before embarking on regression analysis, this will enable one know the exact model that fits the data well. If the X variables are found to be collinear then the researcher should consider the following remedial measures to tackle the multicollinearity problem: Dropping some variables, transformation of variables, additional or new data and very importantly try using other shrinkage regression methods like Principal component regression (PCR), Ridge Regression (RR), Total least squares (TLS) and Partial Least Square regression (PLSR).

## 6.0 Reference

[1] Svante, W., Michael S., and Lennart E. (2001). PLS- regression: a basic tool of chemometrics. Chemometrics and intelligent laboratory systems (58) pp. 109-130.

[2] Herve, A., and Lynne J.W.(2013).Partial least squares methods: Partial least squares correlation and partial least squares regression. Computational toxicology: volume II. Pp. 549-578.

[3] El-dereny M., and Rashwan N.I., (2011). Solving multicollinearity problem using ridge regression models. Int. J. Contemp. Math. Science, Vol.6, no. 12, 585-600.

[4] Gujarati, D.N. Basic Econometrics, McGraw-Hill, New York 1995.

[5] Vinod, H.D. and A. Ullah (1981). Recent Advances in Regression Models, Marcel Dekker, New York.

[6] Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: a review of four recent studies. Strategic Management Journal, 20(2), 195–204.

[7] Reinartz, W.J., Haenlein, M. & Henseler, J. (2009). An empirical comparison of the efficacy of covariance-based and variance-based SEM," *International Journal of Market Research*, 26 (4), 332–344.

[8] Wong, K. K. (2010). Handling small survey sample size and skewed dataset with partial least square path modelling: The Magazine of the Marketing Research and Intelligence Association, November, 20-23.

[9] Hair, J.F., Sarstedt, M., Ringle, C.M. & Mena, J.A., (2012). An assessment of the use of partial least squares structural equation modelling in marketing research. *Journal of the Academy of Marketing Science*, 40(3), 414-433.

[10] Muhammad, I., Maria, J. and Muhammad, A.R., (2013). Comparison of Shrinkage Regression Methods for Remedy of Multicollinearity Problem. *Middle – East Journal of Scientific Research*. 14(4): 570-576.

[11] Paul, B.L., and James G., (2014). Partial least squares (PLS) Structural Equation Modelling (SEM) for building and testing Behavioural Causal Theory: When to choose it and How to use it. IEEE Transactions on Professional Communication, Vol.57, No.2, pp. 123-135.

[12] Nebojsa, S.D. (2014). "The use and misuse of structural equation modelling in management research". A review and critique, *journal of Advances in Management Research*, Vol.11 Issue 1 pp.47-81.

[13] Theo, K.D., and Jorg H., (2015). "Consistent Partial Least Squares Path Modelling". MIS quarterly Research essay Vol. 39.

[14] Jorg, H., Christian M.R., and Marko S. (2015). "A new criterion for assessing discriminant validity in variance-based structural equation modelling". *Journal of the Academic Marketing Science*. Vol.43 pp. 115-135.

[15] Wold, H., (1966). "Estimation of principal components. Comparing Ridge and related models by iterative least squares". In multivariate analysis, Ed. Krishnaiah, P.R, New York: Academic Press, pp: 391-420.

[16] Wethrill, H. (1986). Evaluation of ordinary Ridge Regression, Bulletin of Mathematical Statistics, 18 1 – 35.

[17] Hoerl, A.E. and R.W. Kennard (2000).Ridge Regression: Biased Estimation fornonorthogonal. Problems, Technometrics, 42, 80 – 86