

Solving Multicollinearity Problem in Linear Regression Model: The Review Suggests New Idea of Partitioning and Extraction of the Explanatory Variables

Kayode Ayinde, Olusegun O. Alabi ¹, Ugochinyere I. Nwosu ²

¹Department of Statistics, Federal University of Technology Akure, Ondo State, Nigeria

²Department of Statistics, Federal University of Technology Owerri, P.M.B. 1526, Owerri, Nigeria

✉ **Corresponding Author:** Ugochinyere I. Nwosu, E-mail: nwosuugoichinyere@yahoo.com

ARTICLE INFORMATION

Received: December 08, 2020

Accepted: February 11, 2021

Volume: 2

Issue: 1

DOI: 10.32996/jmss.2021.2.1.2

KEYWORDS

Sustainably, Regressors, Extracting, Sample size, Model parameters

ABSTRACT

Multicollinearity has remained a major problem in regression analysis and should be sustainably addressed. Problems associated with multicollinearity are worse when it occurs at high level among regressors. This review revealed that studies on the subject have focused on developing estimators regardless of effect of differences in levels of multicollinearity among regressors. Studies have considered single-estimator and combined-estimator approaches without sustainable solution to multicollinearity problems. The possible influence of partitioning the regressors according to multicollinearity levels and extracting from each group to develop estimators that will estimate the parameters of a linear regression model when multicollinearity occurs is a new econometrics idea and therefore requires attention. The results of new studies should be compared with existing methods namely principal components estimator, partial least squares estimator, ridge regression estimator and the ordinary least square estimators using wide range of criteria by ranking their performances at each level of multicollinearity parameter and sample size. Based on a recent clue in literature, it is possible to develop innovative estimator that will sustainably solve the problem of multicollinearity through partitioning and extraction of explanatory variables approaches and identify situations where the innovative estimator will produce most efficient result of the model parameters. The new estimator should be applied to real data and popularized for use.

1. Introduction

Multicollinearity is an important econometric problem that has received increased attention in recent times (Ayinde *et al.*, 2015; Lukman *et al.*, 2015; Ismail and Manjula, 2016; Olanrewaju *et al.*, 2017; Tyagi and Chandra, 2017). The findings of researchers (particularly what they have done to address the problems associated with this phenomenon) and the gaps left uncovered are the clear objectives of this review. It is obvious (from literature reviewed) that multicollinearity till date is recognized as a very serious problem in linear regression model. Multicollinearity is the term used to describe cases in which the explanatory variables or regressors are correlated (Lukman *et al.*, 2015). According to this report, in multicollinearity, the regression coefficients are characterized by large standard errors and some possess the wrong sign. This culminates in wrong inferences. Efforts are ongoing, both at the national and international levels to tackle the problems of multicollinearity and the result is that an array of estimation methods (though with one limitation or the other) have been designed to mitigate the effects of this econometric problem. As recently reported by Tyagi and Chandra (2017), such techniques in chronological order available for overcoming this problem include stein estimator proposed by Stein (1956), partial least squares originated by Wold in 1966, principal component regression estimator introduced by Massy (1965) and ordinary ridge regression estimator originated by Hoerl and Kennard (1970). This review will inform new studies by helping to identify important gaps.

2. Single Estimator Approach to Multicollinearity Problem

2.1 The ridge regression estimator

The ridge regression estimator is a single-estimator approach to solving multicollinearity problems. It was first developed by Hoerl and Kennard (1970). Ridge estimator characteristically possesses a smaller mean square error (MSE) than the ordinary least square estimator (Vinod and Ullah, 1981). It is defined as

$$\hat{\beta} = (X'X + KI)^{-1} X'Y \quad (1)$$

Where k is a non-negative constant called the biasing or ridge parameter. It is observed that when $k = 0$, (1) returns to ordinary least square estimates,

$$K_i = \frac{\sigma^2}{\alpha_i^2}, i = 1, 2, \dots, p \quad (2)$$

Since σ^2 and α_i^2 are generally unknown, they were suggested to be replaced by their corresponding unbiased estimates

$$\hat{\sigma}^2 \text{ and } \hat{\alpha}_i^2 \text{ where } \hat{\sigma}^2 = \frac{\sum_{i=1}^n \ell_i^2}{n-p}, \alpha_i \text{ is the } i^{\text{th}} \text{ element of the vector } \hat{\alpha} = Q'\hat{\beta} \text{ and } Q \text{ is an orthogonal matrix. It is}$$

known that as k increases, the ridge regression estimators are biased but however yields more precise estimates than ordinary least square estimator (Mardikyan and Cetin, 2008). It has earlier been suggested that the value of k should be chosen small enough such that the mean squared error of ridge estimator is less than the mean squared error of ordinary least square estimator. Different estimation techniques had been proposed by many researchers. A graphical method called ridge trace was

proposed by Hoerl and Kennard (1970) to select the value of the ridge parameter k . This is a plot of the values of individual $\hat{\beta}_k$

against range of values of k ($0 < k < 1$). The minimum value for which $\hat{\beta}_k$ becomes stable and the wrong signs in the regression coefficient corrected, is used. Hoerl *et al.* (1975) proposed a different estimator of k by taking the harmonic mean of the ridge parameter. Furthermore, Kibria (2003) proposed some new estimators of k by taking the geometric mean, arithmetic mean and median of the ridge parameter. Khalaf and Shukur (2005) proposed k in the form of fixed maximum of the ridge parameter while Alkhamisi *et al.* (2006) suggested another ridge parameter as the arithmetic mean and median of the ridge parameter. Muniz *et al.* (2012) proposed the estimator of the ridge parameter k as the varying maximum and arithmetic mean of the ridge parameter, varying maximum and its reciprocal, its square root and reciprocal of its square root and the geometric mean of k , its square root and reciprocal of its square root. Lukman and Ayinde (2017) recently stated that the value of k for which the residual sum of squares is not too large can be selected. According to this literature, the ridge parameter can take different forms such as fixed maximum, varying maximum, arithmetic mean, harmonic mean, geometric mean and median and various types as original form (O), reciprocal form (R), square root form (SR) and reciprocal of square root (RSR). Combining the knowledge from these works and comparing with simulated data, results showed that best estimator of the ridge parameter technique are of the different forms and types and include fixed maximum original, varying maximum original, harmonic mean original and arithmetic mean square root. It has got a smaller mean square error than the ordinary least square estimator, a feature which gives it an edge over the ordinary least square estimator in the presence of multicollinearity (Lukman *et al.*, 2014). The ridge regression estimator reduces multicollinearity by adding the ridge parameter, K , to the main diagonal elements of the correlation matrix, $X'X$. Furthermore, Liu (1993) introduced an estimator (new class of biased estimate in linear regression) with similarity in form but however differs from the ridge regression estimator of Hoerl and Kennard (1970).

2.2 Principle component

Principle component is a traditional multivariate statistical method mainly employed to decrease the number of predictive variables and provide remedy for the multicollinearity problem (Bair *et al.*, 2006). As revealed by Rosipal and Nicole (2006), the aim of principle component is to identify a few linear combinations of the variables that can be used to summarize the data without loss of too much information in the process. This approach is based on the fact that the p set of variables are transformed to new p set of orthogonal variables denoted C_1, \dots, C_p where each variable C_j is a linear combination of the variable $\bar{X}_1, \dots, \bar{X}_p$ that is,

$$C_j = v_{1j}\bar{X}_1 + v_{2j}\bar{X}_2 + \dots + v_{pj}\bar{X}_p, \quad j = 1, 2, 3, \dots, p \quad \dots\dots\dots (3)$$

The linear combinations are variables C_1, \dots, C_p , such that they are orthogonal and the variance covariance matrix of the principal component is;

$$\begin{matrix}
 & C_1 & C_2 & \dots & C_p \\
 C_1 & \left(\begin{matrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \lambda_p \end{matrix} \right) & & & \\
 C_2 & & & & \\
 \cdot & & & & \\
 \cdot & & & & \\
 \cdot & & & & \\
 C_p & & & &
 \end{matrix}$$

Because the PCs are orthogonal, the off diagonal elements are zero. The diagonal elements $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ (eigenvalues) arranged in decreasing order of magnitude, are the variances of the C_j with corresponding eigenvectors

$$\begin{pmatrix} v_{1j} \\ v_{2j} \\ \cdot \\ \cdot \\ \cdot \\ v_{pj} \end{pmatrix}, \text{ the first few of the PCs are used in the regression, satisfying the following properties;}$$

- (i) $E(C_i) = 0$
- (ii) $\text{var}(C_i) = \lambda_i$
- (iii) $\text{cov}(C_i, C_j) = 0$
- (iv) $\text{var}(C_1) \geq \text{var}(C_2) \geq \dots \geq \text{var}(C_p)$
- (v) the first principal component, has the highest variance compared with the variance of any standardized linear combination of x.

(vi) if $z = \sum \alpha_i x_i$ be a standardized linear combination of X , uncorrelated with the first k PCs, then the variance of $z = (k + 1)$ th PCs. Item (iii) justifies why the use of principle component will remove the problem of multicollinearity. Items iv, v and vi show two things. First, principal component successively captures the maximum of the variance of x and secondly, there is no standardized linear combination that can capture maximum variances without serving as one of the components. From the literature review, this method has been adopted by numerous workers to solve multicollinearity problem, either as a single estimator (Massy, 1965; Marquardt, 1970; Bock *et al.*, 1973; Naes and Martens, 1988; Ayinde *et al.*, 2012) or as a combined estimator (Ayinde and Lukman, 2014; Ayinde *et al.*, 2015).

2.3 Partial least squares

Partial least squares is a recent technique that generalizes and combines attributes from principal component analysis and multiple regression (Abdi, 2003). It originated in the social sciences especially economics (Wold, 1966) and now, it is popular and particularly very useful in areas such as chemical engineering, where predictive variables often consist of several different measurements in the experiment and where the relationship between these variables are not well-understood (Geladi and Kowalski, 1986). Partial least squares is becoming a multivariate technique of choice for both non-experimental and experimental data in social sciences (McIntosh *et al.*, 1996). Notably, the partial least square was first presented as an algorithm (Tenenhaus, 1998) and its statistical properties have been investigated by several workers (Stone and Brooks, 1990; Naes and Helland, 1993; Garthwaite, 1994; Helland and Almoy, 1994) with an impressive result. One of the desirable properties of partial least squares is that it has got a closed form which is given in equation (4) below. Assume the variables in the set $(x'_i y_i)'$ are independent for $i = 1, 2, \dots, n$. Then, the PLS estimator of β is given by

$$\hat{\beta}_{PLS} = \hat{R}(\hat{R}' S_{xx} \hat{R})^{-1} \hat{R}' S_{xy} \quad (4)$$

Where

$\hat{R} = (S_{xy}, S_{xx} S_{xy}, \dots, S_{xx}^{q-1} S_{xy})$ is the $p \times q$ matrix of the krylov sequence,

$$S_{xx} = \frac{X'(I - 11'/n)X}{n-1} \text{ is the } p \times p \text{ matrix,}$$

$$S_{xy} = \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{n-1} \text{ is the } p \times 1 \text{ vector,}$$

$$Y = (Y_1, Y_2, \dots, Y_n)'$$

$$X = (X_1, X_2, \dots, X_n)', I \text{ is the identity matrix}$$

$$1 = (1, 1, \dots, 1)'$$

The NIPLAS algorithm of Wold (1975) was ultimately modified so as to take into account the responses which culminated in the partial least square (PLS) regression algorithm on orthogonal scores presented in Wold *et al.* (1983). Thus, the general NIPLAS algorithm for PLS is given in algorithm below

$$Y = TQ + E \quad (5)$$

$$X = t_1 p'_1 + \dots + t_k p'_k + F \quad (6)$$

$$Y = U_1 q'_1 + \dots + U_k q'_k + E \quad (7)$$

with E and F corresponding to residual terms. The decomposition in equations (6) and (7) above is analyzed to a great extent in Martens and Naes (1989). In Burnham *et al.* (2001), the procedure for extraction of the latent variables as well as for the estimation of the coefficient of the regression mode has been presented. Considering a univariate partial least square, attempt is made to find model of the form

$$\hat{Y} = \hat{\beta}_1 T_1 + \dots + \hat{\beta}_p T_p \quad (8)$$

Where T_p is the linear combination of the X 's and β_p , the parameter variables are assumed to be centered to have mean zero and this implies that the intercept terms will always be zero. Thus the algorithm for determining the T_s is

1 Regress $(Y - \bar{Y})$ on each $X_i - \bar{X}$ in term to get b_{1i}

$$1 \quad \text{Form } T_1 = \sum_{i=1}^m w_{1i} b_{1i} X_{1i}$$

Where,

b_{1i} is the parameter, X_{1i} explanatory variables where the weights w_{1i} sum to one.

3 Regress $(Y - \bar{Y})$ on T_1 and on each $(X_i - \bar{X})$ on T_1 . The residual from these regressions have the effect of T_i removed

4 Replace $(Y - \bar{Y})$ and each $(X_i - \bar{X})$ by the residual of each corresponding regression.

5 Go back to step one updating the index

6 Continue in this manner until all the components or latent variables are extracted.

Valera (1978) and Gujarati and Porter (2009) stated that the two commonly used weighting scheme are equal weight of each predictor and weight directly proportional to variance of X_j and then proposed the use of

1 The standard errors ($w_{ij} \propto SE(\beta_j)$) such that

$$w_{ij} = \frac{SE(\beta_j)}{\sum SE(\beta_j)}$$

Where,

$SE(\beta_j)$ = standard error of the jth regression coefficient and

2 The ranking of their standard errors ($w_{ij} \propto \text{rank } SE(\beta_j)$), such that

$$w_{ij} = \frac{SE_R(\beta_j)}{\sum SE_R(\beta_j)}$$

Where,

$SE_R(\beta_j)$ = ranking of standard error of the jth regression coefficient.

How to choose among all extracted latent variables, precisely the ones that provide the best fit to the data is another task. According to Burnham and Anderson (2004), this consideration calls for a balance between under and over-fitted models. Furthermore, Akaike (1974) proposed a statistical model fit measure known as the Akaike information criterion (AIC) defined as

$$AIC = -2Lm + 2m \quad (9)$$

Where

Lm is the maximized log-likelihood and for linear regression models; $-2Lm = -2\log\text{likelihood}$ (known as the deviance) is $n\log(\text{RSS}/n)$ and m stands for the number in the model. It is interesting to note that this index takes into consideration both the statistical goodness of fit and the number of parameters to be estimated and therefore imposing a penalty for increasing the number of parameters. Therefore, the index with lowest values indicates the preferred model. Schwarz (1978) developed the Bayesian information criterion (BIC) for selection of models among a finite set of models. Partly, it is based on the likelihood function and as a matter of fact it is closely related to the Akaike information criterion. However, to prevent over-fitting, the penalty term is inevitably larger in BIC than in AIC. Categorically, the model with lowest BIC is preferable. BIC is formally defined as

$$\text{BIC} = -2 \ln \hat{L} + K \ln(n) \quad (10)$$

Where,

\hat{L} =the maximized value of the likelihood function of the model,

X=the observed data,

n=the number of data points in X and

K= the number of parameters to be estimated.

Another model selection method is the adjusted R-squared denoted R^2_a and defined as

$$R^2_a = 1 - \frac{\text{RSS}/(n-p)}{\text{TSS}/(n-p)} \quad (11)$$

Simply recall that $R^2 = 1 - \text{RSS}/\text{TSS}$. Therefore, the addition of a variable to a model can only decrease the RSS and thus increases the R^2 . The implication is that R^2 itself is not a good criterion because it would always choose the largest possible model. However, the addition of a predictor will only increase R^2 if it has some value. Meanwhile, minimizing the standard error for prediction means minimizing $\hat{\sigma}^2$ which invariably means maximizing R^2 . It is necessary to comment on the predicted residual error sum of squares (PRESS) which is a statistic used in regression analysis to provide a summary measure of the fit of a model to a sample of observations that were not themselves used to estimate the model. Briefly, it is calculated as the sum of squares of the prediction residuals for concerned observations (Allen, 1974; Tarpey, 2000). Here, the model with the lowest PRESS is preferred. Progressively, the Mallows' Cp is another technique for model selection in regression (Mallows, 1973). It is employed to assess the fit of a regression model that has been estimated using ordinary least squares estimator. It is applied during model selection where a number of predictor variables are available for predicting some outcomes and the aim is to find the best mode involving a subset of these predictors. A small Cp signals that the model is relatively precise and therefore average MSE of prediction might be a good criterion, Mallows (1973) concludes. Other model selection methods have been developed and applied by many workers and these include the Deviance Information Criterion (Vander, 2005; Celeux *et al.*, 2006), False Discovery Rate (Benjamini and Hochberg, 1995), Hannan-Quinn Information Criterion (Burnham and Anderson, 2002) and Bayes Factor (Carlin and Chib, 1995; Chen *et al.*, 2000). Although cross validation is the commonest method, it has been shown that the corrected Akaike information criterion AIC_c , outperforms AIC in model selection of q (Hurvich and Tsai, 1989; McQuarrie and Tsai, 1998). According to the papers reviewed, if the model in question is a linear regression, where k is the number of regressors, including the intercept (Findley, 1991), to minimize AIC or BIC, larger models will fit better, have smaller RSS but nevertheless use more parameters. The best choice of model will therefore balance fit with size. Since BIC has the capacity to penalize larger models more heavily, it will tend to prefer smaller models in comparison to AIC. It is on the record that this approach is beneficial in linear regression model in the presence of multicollinearity (Gujarati and Porter, 2009).

3. Combined Estimator Approach to Multicollinearity Problem

As part of the efforts to combat multicollinearity problems, the information in literature further revealed that some researchers have combined two estimation techniques with the hope that the combination will confer an advantage and become superior to the single-estimator approach. In chronological order, one of the oldest attempts was made in 1984 when Baye and Parker gave $r - k$ class estimator by combining the principal component regression estimator and the ordinary ridge regression estimator, which includes the ordinary least squares estimator, ordinary ridge regression estimator and principal component regression estimator as special cases. Interestingly, Nomura and Ohkubo (1985) obtained conditions for dominance of the $r - k$ class estimator over its special cases using the mean square error as criterion. Liu (1993) combined the merits of the stein and ordinary

ridge regression estimator to obtain an estimator called the Liu estimator. Kaciranlar and Sakallioğlu (2001) combined the Liu estimator and principal component regression estimator to get the $r - d$ class estimator and demonstrated the superiority of this estimator over the ordinary least squares estimator, Liu estimator and principal component regression estimator. Ozkale and Kaciranlar (2007) combined the ordinary ridge regression estimator and Liu estimator to have a two-parameter estimator. They obtained necessary and sufficient condition for dominance of the two-parameter estimator over the ordinary least square estimator in mean square matrix sense. Yang and Chang (2010) worked on the combination of ordinary ridge regression estimator and Liu estimator in a different manner and proposed an another two-parameter estimator and here derived necessary and sufficient conditions for superiority of the another two-parameter estimator over the ordinary least square estimator, ordinary ridge regression estimator, Liu estimator and two parameter estimator under the mean square matrix criterion. Ozkale (2012) proposed a general class of estimators, $r - (k,d)$ class estimator and assessed its performance under the mean square error criterion. The proposed estimator was a combination of the two-parameter estimator and principal component regression estimator and superiority was shown. This is interesting. Chang and Yang (2012) proposed the principal component two-parameter estimator by the combination of the principal component regression estimator and another two-parameter estimator and evaluated its performance under the mean square matrix sense. Tyagi and Chandra (2017) examined the performance of two biased estimators in the presence of multicollinearity with autocorrelated errors which include the same number of unknown parameters with the same range. In Tyagi and Chandra (2017), result suggests that for all the parametric conditions considered in the investigation, $r - (k,d)$ class estimator performs better than the principal components two-parameter estimator under scalar mean square error sense. The shift from single estimator approach to a combined estimator approach was a progressive effort targeted at solving multicollinearity problem.

4. Idea of Partitioning and Extraction of the Explanatory Variables: Need for Research and Application

Previous studies on the subject have concentrated on developing estimators regardless of effect of differences in levels of multicollinearity among regressors. However, the increased concern for a sustainable solution to multicollinearity has compelled reorientation of research focus to the development of new ideas. In responding to the call, researchers in the field should examine the influence of partitioning the regressors according to multicollinearity levels and extracting from each group to develop estimators that will estimate the parameters of a linear regression model when multicollinearity exist in a linear regression model. Recently, the idea to partition and extract in order to solve multicollinearity problem has been inspired by William (2015). From the review, it is incontrovertible that the application of partitioning and extraction of explanatory variables in solving multicollinearity problems have been grossly understudied or largely untested. Therefore, there is need to encourage research on partitioning and extraction of the explanatory variables and apply the same to solve multicollinearity problem sustainably.

5. Concluding Remarks: Suggested Research Lines that Will Tackle Multicollinearity Problems

This study reviewed old and new works on multicollinearity problems with the purpose of suggesting sustainable solution. The present study contributed to existing knowledge by bringing to limelight the influence of partitioning the regressors according to multicollinearity levels and extracting from each group to develop estimators that will estimate the parameters of a linear regression model when multicollinearity occurs. Researchers in the field of econometrics should capitalize on the findings highlighted in this review to develop innovative estimator(s) that will sustainably solve the problem of multicollinearity by addressing the following problems which form the important suggestions for future work:

- i. develop estimators for solving multicollinearity problems through principal component and partial least squares techniques employing partitioning and extraction of explanatory variables approaches.
- ii. examine the performance of newly-developed method in many linear regression models with different degrees of multicollinearity using simulated data.
- iii. compare the performance of the new method(s) with existing methods under different degrees of multicollinearity.
- iv. identify situations where the new method will produce most efficient result of the model parameters.
- v. apply the new method using real data and popularize for use.

Funding: This research received no external funding.

Acknowledgements: Federal University of Technology Akure for providing the opportunity of research and higher education. This paper emanated from the Ph.D. work of the Corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- [1] Abdi, H (2003). Partial Least Squares (PLS) Regression. In: *Encyclopedia of Social Sciences Research Methods*. Lewis-Beck M., Bryman, A., Futing, T. (eds), pp. 1–7, Sage, Thousand Oaks.
- [2] Akaike, H (1974). "A new look at the statistical model identification", *IEEE Transaction on Automatic Control* 19 (6), 716 – 723.
- [3] Alkhamisi, M; Khalaf, G; Shukur, G. (2006). Some modifications for choosing ridge parameters. *Communications in Statistics- Theory and Methods* 35 (11), 2005 – 2020.
- [4] Allen, DM (1974). The relationship between variable selection and data argumentation and a method for prediction. *Technometrics* 16, 125 – 127.
- [5] Ayinde, K; Alao, RF; Ayoola, FJ (2012). Effect of multicollinearity and autocorrelation on predictive ability of some estimators of linear regression model. *Mathematical Theory and Modeling* 2 (11), 41 – 52.
- [6] Ayinde, K; Lukman, AF (2014). Combined estimators as alternative to multicollinearity estimation methods. *International Journal of Current Research* 6 (1), 4505 – 4510.
- [7] Ayinde, K; Lukman, AF; Arowolo, OT (2015). Combined parameters estimation methods of linear regression model with multicollinearity and autocorrelation. *Journal of Asian Scientific Research* 5 (5), 243 – 250.
- [8] Baye, R; Parker, F (1984). Combining ridge and principal component regression: a moneydemand illustration. *Communications in Statistics- Theory and Methods* 13, 197 – 205.
- [9] Bair, E; Trevor, H; Paul D; Robert, T. (2006). Prediction by supervised principal components. *Journal of the American Statistical Association* 101 (473), 119 – 137.
- [10] Benjamini, Y; Hochberg, Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 57 (1), 289 – 300.
- [11] Bock, ME; Yancey, TA; Judge, GG (1973). Statistical consequences of preliminary test estimation in regression. *Journal of the American Statistical Association* 60, 234 – 246.
- [12] Burnham, AJ; MacGregor, JF; Viveros, R (2001). Interpretation of regression coefficients under a latent variable regression model. *Journal of Chemometrics* 15 (4), 265 – 284.
- [13] Burnham, KP; Anderson, DR (2002). *Model Selection and Multimodel Inference: A Practical Information Theoretic Approach* (2nd edition), Springer-Verlag.
- [14] Burnham KP; Anderson DR (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods and Research* 33 (2), 261 – 304.
- [15] Carlin, B; Chib, S (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society* 57, 473 – 484.
- [16] Celeux, G; Forbes, F; Robert, CP; Titterton, DM (2006). Deviance information criteria for missing data models (with Discussion). *Bayesian Analysis* 1, 651 – 706.
- [17] Chang, X; Yang, H (2012). Combining two-parameter and principal component regression estimators. *Statistical Papers* 53 (3), 549 – 562.
- [18] Chen, MH; Shao, QM; Ibrahim, J (2000). *Monte Carlo Methods in Bayesian Computation*. New York, Springer-Verlag.
- [19] Findley, DF (1991). Counter examples to parsimony and BIC. *Annals of the Institute of Statistical Mathematics* 43, 505 – 514.
- [20] Garthwaite, PH (1994). An interpretation of partial least squares. *Journal of American Statistical Association* 89, 122 – 127.
- [21] Geladi, P; Kowalski, B (1986). Partial least squares regression: a tutorial. *Analytica Chimica Acta* 185, 1 – 17.
- [22] Gujarati, DN; Porter, DC (2009). *Basic Econometrics*, 5th edition., New York, Mc Graw-Hill.
- [23] Helland, IS; Almoy, T (1994). Comparison of prediction methods when only a few components are relevant. *Journal of American Statistical Association* 89, 583 – 591.
- [24] Hoerl, AE; Kennard, RW (1970). Ridge regression: applications to non-orthogonal problems. *Technometrics* 12 (1), 69 – 82.
- [25] Hoerl, AE; Kennard, RW; Baldwin, KF (1975). Ridge regression: some simulation. *Communications in Statistics* 4 (2), 105 – 123.
- [26] Hurvich, CM; Tsai, CL (1989). Regression and time series model selection in small samples. *Biometrika* 76, 297 – 307.
- [27] Ismail, B; Manjula, S (2016). Estimation of linear regression model with correlated regressors in the presence of autocorrelation. *International Journal of Statistics and Applications* 6 (2), 35 – 39.
- [28] Kaciranlar, S; Sakallioğlu, S (2001). Combining the Liu estimator and the principal component regression estimator. *Communications in Statistics- Theory and Methods* 30 (12), 2699 – 2705.
- [29] Khalaf, G; Shukur, G (2005). Choosing ridge parameters for regression problems. *Communications in Statistics- Theory and Methods* 34, 1177 – 1182.
- [30] Kibria, BMG (2003). Performance of some new ridge regression estimators. *Communications in Statistics-Simulation and Computation* 32, 419 – 435.
- [31] Liu, KJ (1993). A new class of biased estimate in linear regression. *Communications in Statistics- Theory and Methods* 22 (2), 393 – 402.
- [32] Lukman, A; Arowolo, O; Ayinde, K (2014). Some robust ridge regression for handling multicollinearity and outlier. *International Journal of Sciences: Basic and Applied Research* 16 (2), 192 – 202.
- [33] Lukman, AF; Osowole, OI; Ayinde, K (2015). Two stage robust method in a linear regression model. *Journal of Modern Applied Statistical Methods* 14 (2), 53 – 67.

- [34] Lukman, AF; Ayinde, K (2017). Review and classification of the ridge parameter estimation techniques. *Hacettepe Journal of Mathematics and Statistics* 45 (5), 953 – 967.
- [35] Mallows, CL (1973). Some comments on C_p . *Technometrics* 12, 591 – 612.
- [36] Mardikyan, S; Cetin, E (2008). Efficient choice of biasing constant for ridge regression. *International Journal of Contemporary Mathematical Sciences* 3, 527 – 547.
- [37] Marquardt, DW (1970). Generalized inverse, ridge regression, biased linear estimation and non-linear estimation. *Technometrics* 12, 591 – 612.
- [38] Massy, WF (1965). Principal components regression in exploratory statistical research. *Journal of the American Statistical Association* 60 (309), 234 – 256.
- [39] Martens, H; Naes, T (1989). *Multivariate Calibration*. New York, Wiley.
- [40] McIntosh, AR; Bookstein, FL; Haxby, JV; Grady, CL (1996). Spatial pattern analysis of functional brain images using partial least squares. *Neuroimage* 3, 143 – 157.
- [41] McQuarrie, ADR; Tsai, CL (1998). *Regression and time series model selection*. Singapore, World Scientific.
- [42] Muniz, G; Kibria, BMG; Mansson, K; Shukur, G (2012). On developing ridge regression parameters: a graphical investigation. *SORT* 36 (2), 115 – 138.
- [43] Naes, T; Martens, H (1988). Principal component regression in NIR analysis: view points, background details selection of components. *Journal of Chemometrics* 2, 155 – 167.
- [44] Naes, T; Helland, IS (1993). Relevant components in regression. *Scotland Journal of Statistics* 20, 239 – 250.
- [45] Nomura, M; Ohkubo, T (1985). A note on combining ridge and principal component regression. *Communication in Statistics Theory and Methods* 14 (10), 2473 – 2487.
- [46] Olanrewaju, SO; Yahaya, HU; Nasiru, MO (2017). Effects of multicollinearity and on some estimators in a system of regression equation. *European Journal of Statistics and Probability* 5 (3), 1 – 15.
- [47] Ozkale, MR (2012). Combining the unrestricted estimators into a single estimator and a simulation study on the unrestricted estimators. *Journal of Statistical Computation and Simulation* 82 (5), 653 – 688.
- [48] Ozkale, MR; Kaciranlar, S (2007). The restricted and unrestricted two-parameter estimators. *Communication in Statistics Theory and Methods* 36 (15), 2707 – 2725.
- [49] Rosipal, R; Nicole, K (2006). Overview and Recent Advances in Partial Least Squares. In: *Subspace, Latent Structure and Feature Selection*, Saunders, C. et al. (eds) (Heidelberg: Springer-Verlag) 3940, 34 – 51.
- [50] Schwarz, GE (1978). Estimating the dimension of a model. *Annals of Statistics* 6 (2), 461 – 464.
- [51] Stein, C (1956). Inadmissibility of usual estimator for the mean of a multivariate normal distribution. In *proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, pp. 197 – 206, University of California Press.
- [52] Stone, M; Brooks, R (1990). Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression. *Journal of the Royal Statistical Society, Series B* 52 (2), 237 – 269.
- [53] Tarpey, T (2000). A note on the prediction sum of squares statistic for restricted least squares. *The American Statistician* 54, 116 – 118.
- [54] Tenenhaus, M (1998). *La Régression PLS*. Technip, Paris.
- [55] Tyagi, G; Chandra, S (2017). Note on the performance of biased estimators with autocorrelated errors. *International Journal of Mathematics and Mathematical Sciences*. Volume 2017, 12 Pages, Article ID 2045653.
- [56] Valera, GR (1978). A Comparison between Unweighted and Weighted Regression for Forecasting. *Retrospective Theses and Dissertations*, 17019.
- [57] Vander Linde, A (2005). DIC in variable selection. *Statistical Neerlandica* 59, 45 – 56.
- [58] Vinod, D; Ullah, A (1981). *Recent Advances in Regression Methods*. Marcel Dekker Inc. Publication.
- [59] Williams, R (2015). Multicollinearity. <https://www3.nd.edu/~rwilliam/> (accessed July 31, 2018).
- [60] Wold, HOA (1966). Estimation of Principal Components and Related Models by Interactive Least Squares. In: Krishnaiah, P.R. (ed). *Multivariate Analysis*, pp. 391 – 420. New York, Academic Press.
- [61] Wold, HOA (1975). Path models with latent variables. The NIPLAS approach, In: Blalock, H.M., Ajanbegian, A., Borodkin, F.M., Boudon, R., Capocchi, V. *Quantitative Sociology*, Academic Press, New York, 307 – 359.
- [62] Wold, S; Martens, H; Wold, HOA (1983). The multivariate calibration method in chemistry solved by PLS method In: *Proceedings of Conference Matrix Pencils Lecture Notes in Mathematics*, Ruhe A. and Kagstrom B., eds., Springer-Verlag. Heidelberg, 286 – 293.
- [63] Yang, H; Chang, X (2010). A new two-parameter estimator in linear regression. *Communication in Statistics Theory and Methods* 39 (6), 923 – 934.