iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** XII **Month of publication:** December 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor:6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

Muticollinearity Problem in Criteria of Model Selection

T. Gangaram[1], C. Narayana[2], R.V.S.S. Nagabhushana Rao[3], K.Vasu[4], G. Mokesh Rayalu[5]
[1]*Lecturer, Department of Statistics, PVKN Govt College, Chittoor*
[2] *Assistant professor, Department of Mathematics, Sriharsha Institute of P.G. Studies, Nellore*
[3]*Assistant professor, Department of Statistics, Vikrama Simhapuri University, Nellore*
[4]*Assistant Professor, Vidya Educational Institutions, Yadamari, Chittoor*
[5]*Assistant Professor, Department of Mathematics, School of Advanced sciences ,VIT, Vellore*

Abstract: Economists and other social scientists often employ data that is multicollinear. The ordinary least squares (OLS) estimator using multicollinear data may have large sampling errors and other related features that researchers view as undesirable. The multicollinear problem may arise when some or all of the explanatory variables in a regression are highly correlated with one another. At least a brief discussion of this problem is, of course, included in every econometrics text book. A method for assessing muticollinearity and compare their method with other methods proposed in the numerical analysis literature. A serious problem that can occur in regression analysis is the presence of muticollinearity among the independent variables in a regression equation

I. INTRODUCTION

Economists seek to understand the nature and functioning of economic systems. A linear model is a set of linear relationships between a dependent variable and a set of independent variable. One objective of such an understanding is to be able to make conditional predictions of the likely future development of the system to take action to control to some degree, the evolution of the system. Considering the joint estimation procedures is of course generally more efficient than single equation procedures. The properties such as unbiasedness, efficiency, and invariance are proposed as a basis for evaluating and choosing among estimators or decision rules. These properties then form the basis for rating or ranking various estimators or rules.

II. BRIEF ABOUT GENERAL LINEAR STATISTICAL MODEL

Consider the following linear statistical model

$$Y= X\beta + \in \qquad \ldots (2.1)$$

Where

Y is a (nx1) vector of observable values of the dependent variable Y.

X is a known (nxk) no stochastic design matrix of rank k.

β is a (kx1) vector of unknown parameters

and $\in$ is (n x 1) vector of unobservable random disturbances.

With mean vector $E(\in)=0$ and finite covariance matrix

$$E(\in \in^1)= \sigma^2 I$$

Where the scalar σ^2 is unknown and I is a n^{th} order identity matrix. The specification for the real random vector $\in$ as $\in \sim(0, \sigma^2 I)$.

The (nxk) design matrix X is assumed to be composed of K fixed (explanatory) variables. The assumption that X is a matrix of fixed variables means for all the possible situations that might take place under repeated sampling, the matrix X would take on the same values. However the observable vector Y would vary from sample to sample. In estimating the unknown coefficient vector β, restrict the class of rules that are linear functions of Y, the general linear estimator or rule as.

$$\beta_0 = AY$$

Where A is some (k x n) matrix, that appropriately summarizes the information contained in y. If the use of the criterion of choosing the estimate of the coefficient vector β that makes the sum of the squared errors of the disturbance vector ε a minimum, than the quadratic form

$$l = \in^1 \in = (Y-X\beta)^1 (Y-X\beta)$$

$= Y^1Y-2\beta^1 X^1Y + \beta^1X^1 x \beta$ … (2.2)

This criterion results in the optimizing condition

$$\frac{1}{2}\frac{\partial l}{\partial \beta} = X^1X\ \beta - X^1Y = 0 \quad \text{… (2.3)}$$

Which leads to the optimizing vector

$$\hat{\beta} = \left(X^1X\right)^{-1} X^1Y \quad \text{…. (2.4)}$$

This linear rule is appropriately known as the ordinary least squares (OLS) estimator.

A. Assumptions On Classical Linear Model

The various assumptions on classical linear model are given by:

1) Assumption of Unbiasedness

$E(\in) = 0$ or $E(Y) = X\beta$ i.e., $\in_i$'s are having zero means.

2) Assumptions of Homoscedasticity

$E(\in \in^1) = \sigma^2\ I_n$ or $Var(Y) = \sigma^2 I_n$

Where σ^2 is unknown and I_n is an identity matrix of dimension n x n.

i.e., $\in_i$'s are having constant unknown variance σ^2 and are uncorrelated with each other.

3) Assumptions of Linear Independence of Regressors:

Rank of x is k, k < ni.e., the regressors (columns of X) are linearly independent.

4) Assumptions of Fixed Data Matrix

X is a nonstochastic data matrix and has the property that

$\lim_{n\to\alpha}\left(\frac{X^1X}{n}\right) = Q$;Where Q is finite and nonsingular matrix.

5) Assumption of Non-Existence of Errors in Regressors:

X is measured without error. Also X is uncorrelated with $\in$

6) Assumption of Normality

$$\in \sim N(0, I_n) \text{ or } Y \sim N(X\beta, \sigma^2\ I_n)$$

i.e., $\in$ has a multivariate normal distribution with mean vector zero and dispersion matrix $\sigma^2\ I_n$. The linear model (2.1.) with the assumptions given in (2.1) is referred to as the 'Classical General Linear Model' or 'Gauss – Markov Linear Model'.

B. Ordinary Least Squares Estimator

The hypothesized model is

$$Y_{n\times1} = X_{n\times k}\ \beta_{k\times1} + \in_{n\times1} \quad \text{…. (2.5)}$$

The first basic assumption of the model is that the vector of sample observations on y, as a linear combination of the sample observation on the explanatory X variables plus a disturbance vector, that is

$$Y = X_1\beta_1 + X_1\beta_2 + \ldots\ldots\ldots + X_k\beta_k + \in$$

Where each vector is a column vector of n elements.

$E(\in) = 0$, that is, $E(Y) = X\beta$

$E(\in \in^1) = \sigma^2 I$. Since $E(\in)=0$, $E(\in \in^1)$ is a variance matrix. This assumption gives

$$E\left(\in\in^{1}\right)=\begin{bmatrix}\sigma 2 & 0 & ..... & 0\\ 0 & \sigma^{2} & .... & 0\\ - & - & .... & -\\ 0 & 0 & .... & \sigma^{2}\end{bmatrix}$$

$\rho(X)=k$, this assumption states that the explanatory variables do not form a linearly dependent set. X is a nonstochastic matrix.

The $\in$ vector has a multivariate normal distribution that is, $\in\square\ N(0,\sigma^{2}I)$

Let b_{*} denote any arbitrary k-element vector. This serves to define a vector of errors, or residuals,

$$e_{*} = Y\text{-}Xb_{*} \qquad \ldots. (2.6)$$

The least- squares principle for choosing b_{*} is to minimize the sum of the squared residuals $e_{*}^{1}e_{*}$ from equation (2.6).

$$e_{*}^{1}e_{*}=\left(Y-Xb_{*}\right)^{1}\left(Y-Xb_{*}\right)=Y^{1}Y-2b_{*}^{1}X^{1}Y+b_{*}^{1}X^{1}Xb_{*}$$

Thus $\dfrac{\partial\left(e_{*}^{1}e_{*}\right)}{\partial b_{*}}=-2X^{1}Y+2X^{1}Xb_{*}$ … (2.7)

The resultant OLS solution for simply by b_{*} gives

$(X^{1}X)\,b = X^{1}Y$, $b=(X^{1}X)^{-1}X^{1}Y$ …. (2.8)

These are referred to as the OLS normal equations. The vector of OLS residuals is likewise denoted by e, where e = Y - Xb

….. (2.9)

using this expression to substitute for y in equation (2.9) gives $(X^{1}X)\,b=(X^{1}X)\,b+X^{1}e$.

Thus $$X^{1}e=\begin{bmatrix}X_{1}^{1}e\\ X_{2}^{1}e\\ :\\ X_{k}^{1}e\end{bmatrix}=\begin{bmatrix}0\\ 0\\ :\\ 0\end{bmatrix}=0 \qquad \ldots. (2.10)$$

This is a fundamental OLS result. The variance – covariance matrix of the OLS estimator is b-E (b) = b-$\beta=(X^{1}X)^{-1}X^{1}\in$

Thus var (b) = E $\{(X^{1}X)^{-1}X^{1}\in\in^{1}X(X^{1}X)^{-1}\}$

$=(X^{1}X)^{-1}X^{1}\sigma^{2}IX(X^{1}X)^{-1}$ (from assumptions (iii) & (iv))

var (b) = $\sigma^{2}(X^{1}X)^{-1}$ (2.11)The most important result in least-squares theory is that no other linear unbiased estimator can have smaller sampling variances than those of the OLS estimator in equation (2.11). OLS estimators are thus said to be best linear unbiased estimators (B.L.U.E), that is, to have minimum variance within the class of linear unbiased estimators.

III. SAMPLING PROPERTIES OF THE ORDINARY LEAST SQUARES ESTIMATOR

In conventional sampling theory, properties such as unbiasedness, efficiency, and invariance are proposed as a basis for evaluating and choosing estimators or decision rules. Thus properties are identified that may be desirable under certain circumstances, and these properties then form the basis for rating or ranking various estimators or rules. Since the least squares estimator b is a linear function of the random vector Y, it is also a random vector with mean vector.

$$E(b)=E((X^{1}X)^{-1}X^{1}Y)=\beta+E\left[(X^{1}X)^{-1}X^{1}\in\right]$$

$$E(b)=\beta+(X^{1}X)^{-1}X^{1}E(\in)=\beta \qquad \ldots.. (3.1)$$

Here the assumptions $E(\in)=0$ is used and that X is nonstochastic.

These results imply that the linear decision rule (3.1) is on average correct and is thus unbiased, a property that is sometimes given high priority in evaluating sampling theory estimators. The Linear unbiased rule (3.1) compares with other linear unbiased rules in terms of the precision (variance – covariance) matrix. The covariance matrix for the random vector b is

$$\begin{aligned} Cov\,(b) &= E\,[\,(b-E(b))\,(b-E(b))^1\,] = E\,[\,(b-\beta)\,(b-\beta)^1\,] \\ &= E\,[\,(X^1X)^{-1}\,X^1\,\in\in^1\,X\,(X^1X)^{-1}\,] \\ &= \sigma^2\,(X^1X)^{-1} \qquad \ldots\ldots (3.2) \end{aligned}$$

Which follows from the assumptions $E\,(\in\in^1) = \sigma^2 I$ and X is non stochastic.

Given this covariance result, the Gauss-Markov theorem provides proof that out of the class of linear unbiased rules for the statistical model $Y = X\beta + \in$ the least squares estimator is best, which is defined in terms of minimum variance. That is, any arbitrary linear unbiased estimator b_o will have a covariance matrix

$$E\,[\,(b_0-\beta)\,(b_0-\beta)^1\,] = \sigma^2\,(X^1X)^{-1} + \Delta \qquad \ldots\ldots (3.3)$$

Where Δ is a positive semi definite matrix. Therefore for the linear rule (3.3) the least square estimator is equal to or better in terms of sampling precision than all others in its class.

IV. ORDINARY LEAST SQUARES ESTIMATION FOR β AND σ^2

OLS Estimator for β:Consider the classical general linear model as:

$$Y = X\beta + \in$$

Let $\hat{\beta}$ be a column vector estimator of β. Now, the sum of squared residuals is given by,

$$e^1e = (Y - X\hat{\beta})^1(Y - X\hat{\beta}) = Y^1Y - 2\hat{\beta}^1X^1Y + \hat{\beta}^1X^1X\hat{\beta} \qquad \ldots(4.1)$$

$$= Y^1Y - 2\hat{\beta}X^1X\hat{\beta}$$

The least squares criterion results in the optimizing condition

$$\frac{1}{2}\frac{\partial(e^1e)}{\partial\hat{\beta}} = X^1X\hat{\beta} - X^1Y = 0 \qquad \ldots(4.2)$$

Which eads to optimizing vector $\hat{\beta} = (X^1X)^{-1}(X^1Y)$ ...(4.3)

known as the least squares estimator for β

Unbiased estimator for σ^2consider the OLS estimator as

$$e = Y - X\hat{\beta} = Y - X(X^1X)^{-1}X^1Y = [I_n - X(X^1X)^{-1}X^1]\,Y$$

$$= \left[I_n - X\left(X^1X\right)^{-1}X^1\right]\left(X\beta + \in\right) \qquad \ldots (4.4)$$

$$= \left[I_n - X\left(X^1X\right)^{-1}X^1\right]\in \qquad \ldots(4.5)$$

$$\Rightarrow e = M \in$$

where, $M = [I_n - X(X^1X)^{-1}X^1]$ is of dimension (nxn) and symmetric idempotent matrix. i.e.,

$M^1M = MM = M^2 = M.$ Further, $MX = 0$

The OLS residual sum of squares is given by,

$$e^1e = (M\in)^1(M\in) = \in^1 M\in \quad \ldots(4.6)$$

Consider, $\in(e^1e) = \in[e^1M\in]$

$$= E[tr(E^1ME)] = E[tr(MEE^1)] \quad = tr\{ME(\in\in^{1)}\}$$

$$= \sigma^2 tr(M) = \sigma^2 tr[I_n - X(X^1X)^{-1}X^1]$$

$$= \sigma^2[tr(I_n) - tr\{X(X^1X)^{-1}X^1\}] = \sigma^2[tr(I_n) - tr\{(X^1X)^{-1}X^1X\}]$$

$$= \sigma^2[tr(I_n) - tr(I_k)] = \sigma^2(n-k)$$

Consequently, $$E\left[\frac{e^1e}{n-k}\right] = \sigma^2 \quad \ldots. (4.7)$$

Thus, an unbiased estimator of σ^2 is

$$S^2 = \frac{e^1e}{n-k} = \frac{\left(Y - X\hat{\beta}\right)^1\left(Y - X\hat{\beta}\right)}{n-k} \quad \ldots. (4.8)$$

$$S^2 = \frac{Y^1MY}{n-k} = \frac{Y^1Y - \hat{\beta}X^1Y}{n-k} \quad \ldots(4.9)$$

A. Maximum likelihood estimators for β and σ^2

The likelihood function for the linear model as

$$L = \left(2\prod\right)^{-n/2}\left(\sigma^2\right)^{-n/2}\exp\left[\frac{\left(Y - X\beta\right)^1\left(Y - X\beta\right)}{2\sigma^2}\right] \quad \ldots (4.10)$$

To maximize L, one can set the first derivatives of log L equal to zero.

$$\frac{\partial \log L}{\partial\hat{\beta}} = 0 \Rightarrow \frac{1}{\hat{\sigma}^2}X\left(Y - X\hat{\beta}\right) = 0 \quad \ldots. (4.11)$$

$$\frac{\partial \log L}{\partial\hat{\sigma}^2} = 0 \Rightarrow \frac{-n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\left(Y - X\hat{\beta}\right)^1\left(Y - X\hat{\beta}\right) = 0 \quad \ldots (4.12)$$

Solving the equations (4.11) and (4.12) for $\hat{\beta}$ and $\hat{\sigma}^2$ one can get the maximum likelihood estimators for β and σ^2 are

$$\hat{\beta} = \left(X^1X\right)^{-1}\left(X^1Y\right) \quad \ldots. (4.13)$$

and $$\hat{\sigma}^2 = \frac{\left(Y - X\hat{\beta}\right)^1\left(Y - X\hat{\beta}\right)}{n} = \frac{e^1e}{n} \quad \ldots. (4.14)$$

V. GOODNESS OF FIT MEASURES

A. Coefficient of Multiple Determination (R^2)

In an applied work, a commonly used measure of the adequacy of an estimated linear model, is the coefficient of multiple determination (R^2). This measure has a number of interpretations. Consider the classical linear regression model as

$$Y = X\beta + \in$$

It one of the regressors in the linear model is a constant term (intercept parameter) then one can partition the total sum of squares (SST) as follows.

$$SST = SSR + SSE \quad \ldots.(5.1)$$

Where SST: Total sum of squares

$$=\sum_{i=1}^{\eta}\left(Y_i-\bar{Y}\right)^2=Y'Y-n\bar{Y}^2$$

SSR: Regression sum of squares or explained sum of squares

$$=\sum_{i=1}^{\eta}\left(\hat{Y}_i-\bar{Y}\right)^2=\hat{Y}^1\hat{Y}-n\bar{Y}^2$$

SSE : Residual sum of squares = $\sum_{i=1}^{\eta}\left(Y_i-\hat{Y}_i\right)^2=e^1e$

If the linear model is in the deviation form, then the partitioning of total sum of squares is given by

$$\hat{Y}Y=\hat{\beta}^1X^1Y+e^1e \qquad \text{....(5.2)}$$

$$(SST)=(SSR)+(SSE)$$

Now, R^2 is defined as a measure at proportion of the total variance accounted for by the linear influence at the explanatory variables.

$$R^2=\frac{SSR}{SST}=\frac{\hat{\beta}^1X^1Y}{Y^1Y} \qquad \text{....(5.3)}$$

or

$$R^2=1-\frac{SSE}{SST}=1-\frac{e^1e}{Y^1Y} \qquad \text{...(5.4)}$$

Because of the decomposition of total sum of squares in (5.2), the value of R^2 lies between 0 and l and may be interpreted as a measure of the proportion of the variation in Y explained by the estimated equation. The coefficient R (the non-negative square root of R^2) is known as the multiple correlation coefficient of Y on given explanatory variables. bviously the closer R (or) R^2 to 1, the better performance of the explanatory variables could be achieved.

B. Model Selection Using $\bar{R}^2$

To maximize $\bar{R}^2$, one may start by regressing y on all the explanatory variables in the hypothesized model, then starting a sequential deletion of the independent variable whose associated t-statistics are the smallest of all t values provides that they do not satisfy t>1. The next step will be a re-estimation of all remaining regressions; again the variable whose associated t-statistics is the smallest of all remaining values and does not satisfy t >1, will be dropped and this procedure will be continued, until all the t-values for the regression coefficients become greater than unity. There is a high chance of reaching the highest attainable $\bar{R}^2$,due such a procedure. Most step wise regression routines will allow for it.

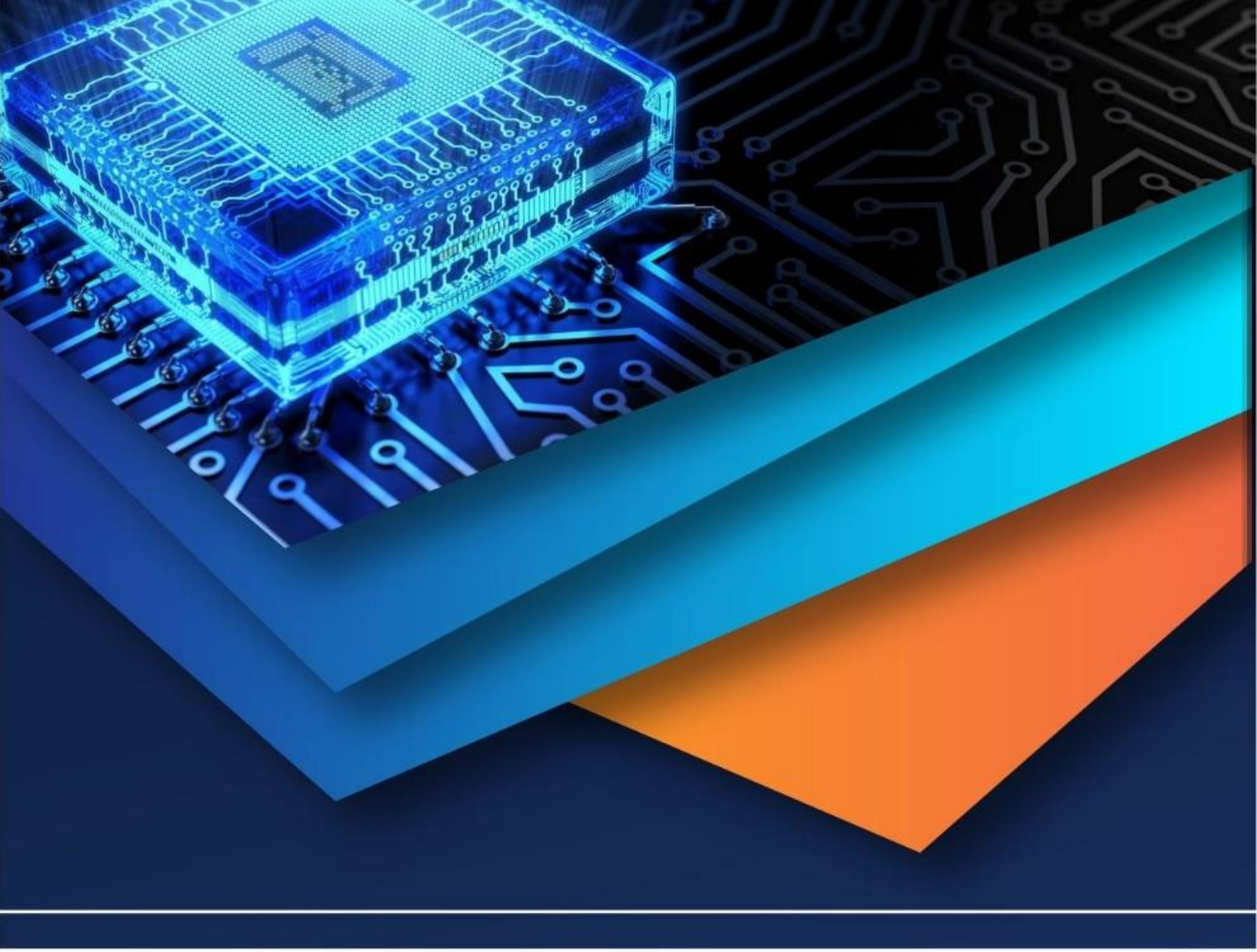
V. CONCLUSIONS

In presence of multicollinearity certain biased estimation procedures like Ridge regression, Generalized inverse estimator, principal component regression, Liu estimator are used to improve the efficiency of ordinary least squares (OLS) estimates in the linear regression model. The problem of multicollinearity in regression analysis is essentially a lack of sufficient information in the sample to permit accurate estimation of the individual parameters. The primary purpose of regression models is the prediction of the value of dependent variable with the help of many explanatory variables. In this paper we discuss about coefficient of determination and adjusted R^2 for Multicollinearity problem.

REFERENCES

[1] Bacon, R.W., and Hausman, J.A., (1974), The Relationship between Ridge regression and the minimum Mean squared Error estimator of chipman, oxford Bulletin of economics and statistics, 36, pp:115-124.
[2] Dutta, M.(1975), Econometric methods, South – Western Publishing Company, Cincinnati.

[3] Farrar, D.E., and Glauber, R.R. (1967), "Multicollinearity in Regression Analysis: The problem Revisited" Review of Economics and Statistics, 4, pp 92 – 107.
[4] Feldstein, M.S. (1973), Multicollinearity and the mean square error of alternative estimators, Econometrics 41, pp:337-342
[5] Gujarati, Damodar N. (1995). Basic Econometrics, Third Edition, McGRAW – Hill, Inc, New York
[6] Gunst, R.F. (1984). "Toward a Balanced Assessment of Collinearity Diagnostics," The American Statistician, 38, pp:79-82
[7] Hoerl, A.E. and Kennard, R.W. (1970). "Ridge Regression: Biased Estimation for Non-orthogonal problems", Technometrics, 12; pp:55-67
[8] Jodge, G.G. and Bock, M.E (1978), the statistical Implications of Pre-test and Stein-Rule Estimators in Econometrics; North-Holland Publishing Company; Amsterdam.
[9] Johnston, J. (1984) "Econometric Methods". Third Edition, McGraw Hill Book Company, Singapore.
[10] Lindley, D.V., and Smith, A.F.M. (1972), Bayes Estimates for the Linear Model, Journal of Royal Statistical Society., 34, pp:1-41
[11] Maddala, G.S. (1992), Introduction to Econometrics, Macmilla, Second Edition, New York.Marquardt, D.W.(1974). Discussion No:2 Technometrics, 16, pp:189-192
[12] Obenchain, R.L. (1977), "Classical F-Tests and confidence regions for Ridge Regression", Technometrics, 19, pp: 429 – 438
[13] Rao, C.R. and Toutenburg, H. (1995), Linear Models, Least Squares and Alternatives, Springer Verlag, New York.Webster, J.T., Gunst, R.F., and Mason, R.L. (1974). Latent root regression analysis, Technometrics, 16, pp:513-522.
[14] Zellner, A. (1971). An Introduction to Bayesian Inference in Econometrics, John Wiley & Sons, New York
[15] Ziemer, R.F., Hill, R.C. (1984). The risk of general stein – like estimator in the presence of Multicollinearity, Journal of Econometrics, North-Holland, 25, pp:205-216.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY