IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2018

Estimation of Parameter Vector Using Ridge and Principal Components Regression Techniques for Non-Orthogonal Data

Alona Biswa[1] Rajeshwar Singh[2]

Research Scholar, Department of Statistics, North Eastern Hill University, Shillong[1]

Associate Professor, Department of Statistics, North Eastern Hill University, Shillong[2]

Abstract: Method of least squares estimate gives imprecise estimates in handling numerous multi-collinear explanatory variables. After detecting the presence of multi-co linearity, techniques of ridge and principal components regressions are employed to study the impacts of explanatory variables on dependent variable. Using ridge regression the degree of multi-co linearity is reduced only but the technique of principal components regression gives better estimate with greater degree. This paper discusses about the application and comparison of the techniques of ridge and principal components regressions for the estimation of parameter vector together with the conflicting claim on the discovery of the technique of ridge regression based on available documents.

Keywords: Conflicting Claim on Discovery of Ridge Regression, Multi-co linearity, Principal Components Regression and Ridge Regression.

1. INTRODUCTION

The term multi-co linearity is due to R. Frisch. Today multi-co linearity means the existence of an approximate linear relationship among several explanatory variables. Orthogonality illustrates about non-existence of multi-co linearity and this is one of the assumptions of classical linear regression model. Curtis and Ghosh (2011) discussed that the crux of the problem is that the linear regression model assumes each explanatory variable has an independent effect on the response that can be encapsulated in the explanatory variable's regression coefficient. In case of highly multi-collinear explanatory variables the data do not hold much information on the independent effects of each explanatory variable, offering rise of standard errors for the regression coefficients. Niemelä-Nyrhinen and Leskinen (2014) discussed Structural Equation Modeling (SEM) to alleviate the effects of multi-co linearity, using the LISREL program (a statistical software package used in SEM developed by Karl G. Jöreskog during 1970s). The Ordinary Least Squares (OLS) procedure cannot be applied in these situations or it may be said that OLS procedure is no longer valid to get the estimates of parameter vector. These situations have motivated the researchers to squeeze out maximum information from whatever data have at their disposal and invented some very ingenious techniques like: Ridge Regression (RR), Principal Components Regression (PCR), generalized inverse regression, dynamic programming, partial least squares regression, Tychonoff Regularization and others. These could fruitfully be applied in case of existence of multi-co linearity in the data. This paper looks into and applies techniques of RR and PCR to get the estimates of coefficient vector in case of multi-co linearity. Hoerl and Kennard (1970 a, b) proposed the technique of RR suggesting to add a small positive quantity in $X'X$ before inverting it. Ridge Estimates (RE) are biased but have minimum mean squared error as compared to the variance of estimates in case of OLS. PCR is another superior technique to obtain estimate of parameter vector in the presence of multi-co linearity. Actually all explanatory are constructed into new orthogonal variables, called Principal Components (PCs) through Principal Components Analysis (PCA) and then OLS technique is applied to get the estimate of those coefficients. Finally, original parameter vector is obtained through the established relationship during the construction. Both estimators provide biased coefficient estimators with the relatively smaller variation than the variance of the OLS estimator and both are the most popular regression methods that have been applied to collinear data. Instead of smoothly shrinking the coordinates on the PCs, PCR either does not shrink a coordinate at all or stable. First of all, multi-co linearity is detected from the available data using Condition Index (CI), condition number, eigen values, tolerance and Variance-Inflating Factor (VIF). These methods are new methods for detecting the multi-co linearity in the data.

2. CORRECTIVE MEASURES

When multi-co linearity is detected in given data and then it is required to proceed for its corrective measures. Only two techniques, viz., Ridge Regression (RR) and Principal Components Regression (PCR) of several corrective measures are considered in this paper for the study.

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2018

An Illustration:

Table 1: Data on Y, X_1, X_2 and X_3

Y	X_1	X_2	X_3
15.90	149.30	4.20	108.10
16.40	161.20	4.10	114.80
19.00	171.50	3.10	123.20
19.10	175.50	3.10	126.90
18.80	180.80	1.10	132.10
20.40	190.70	2.20	137.70
22.70	202.10	2.10	146.00
26.50	212.40	5.60	154.10
28.10	226.10	5.00	162.30
27.60	231.90	5.10	164.30
26.30	239.00	0.70	167.60

Source: E. Mallinvaud, Statistical methods of Econometrics, 2nd ed. (Amsterdam: North Holland, 1970), p. 19 Box plot is plotted using data given in Table 1 and no outlier in Y, X_1, X_2 and X_3 is visible in all boxplots drawn. Thus box plots in Fig 1 demonstrate that there is no outlier present in the data but dependent variable Y and explanatory variables X_1, X_2 and X_3 are skewed and asymmetrically distributed. This gives general idea about distributions of Y, X_1, X_2 and X_3. Fig 2 given below illustrates about matrix plots for variables Y, X_1, X_2 and X_3 and it reflects about the presence of multi-co linearity between X_1 and X_2, X_1 and X_3 and no collinearity between X_2 and X_3 in the given data. This recommends that some better numerical methods should be utilized for its detection.

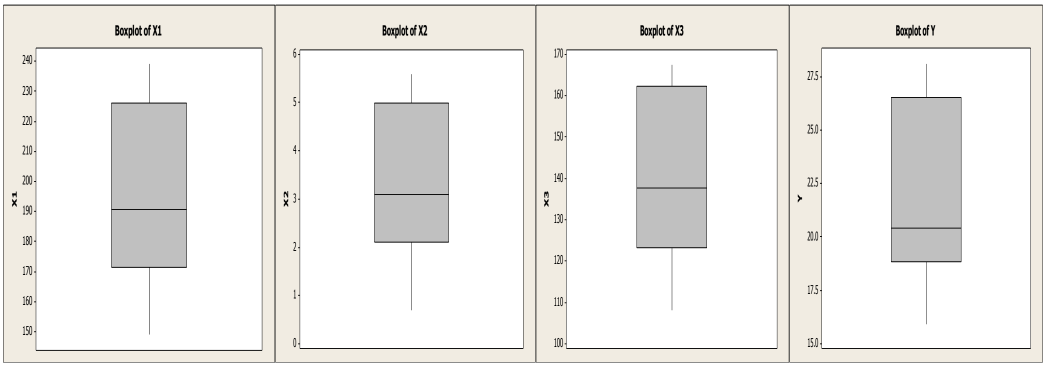

Fig 1: Box plots of Y, X_1, X_2 and X_3

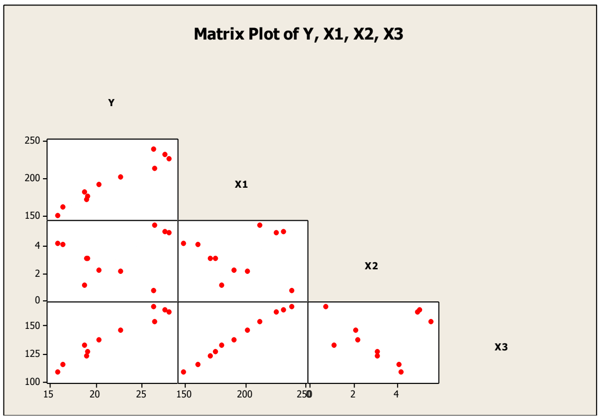

Fig 2: Matrix Plot

Tables 2 and 3 demonstrate about the presence of multi-collinearity using R^2 with its adjusted R^2 and ANOVA. This further suggests that some well known techniques like PCR and RR are required to proceed further for the estimation of coefficients in lieu of the technique of OLS and they are discussed in sub-sections 2.1 and 2.2 respectively.

 DOI 10.17148/IARJSET.2018.578

Table 2: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	0.996[a]	0.992	0.988	0.48887

a. Predictors: (Constant), X_3, X_2, X_1

Table 3: ANOVA Table

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	204.776	3	68.259	285.610	.000[a]
Residual	1.673	7	.239		
Total	206.449	10			

a. Predictors: (Constant), X_3, X_2, X_1

b. Dependent Variable: Y

2.1 PRINCIPAL COMPONENTS REGRESSION

Section 2 describes that data available in Table 1 exhibit the multi-collinearity. In this section the technique of PCR will be discussed with its application. PCR is actually completed in three steps. In first step, a new set of PCs is constructed from given multi-collinear explanatory variables using PCA technique and all PCs constructed thus are orthogonal or uncorrelated to each other. In second step, OLS technique is applied on constructed PCs to get estimate of new parameter vector and in the third step original parameter vector is computed from the estimated value of new parameter vector through established relationship during the second step.

Let $X_1, X_2, \ldots, X_p$ are p-explanatory variables under following classical linear regression model

(1) $Y = X\beta + \varepsilon$

If explanatory variables are correlated then a set of PCs is constructed. After it consider Z = XV and $\alpha = V'\beta$. Hence $Z'Z = V'X'XV = \wedge$. (1) is now written as

(2) $Y = XVV'\beta + u = Z\alpha + u$ as $VV' = I$

Since Z-values are orthogonal and OLS estimate of α is $\hat{\alpha}$. Then

(3) $\hat{\alpha} = (Z'Z)^{-1} Z'Y = \wedge^{-1} Z'Y$

The estimate of original parameter vector β is computed using the relation $\alpha = V'\beta$, which gives $\hat{\alpha} = V'\hat{\beta}$.

Table 4: Descriptive Statistics

	Mean	Std. Deviation
Y	21.8909	4.54367
X_1	194.5909	29.99952
X_2	3.3000	1.64924
X_3	139.7364	20.63440

The technique of PCR is now developed using data from Table 1 with the help of software. Table 4 gives descriptive statistics of Y, X_1, X_2 and X_3 and Table 5 describes about three components and standardized explanatory variables X_1', X_2' and X_3' to get estimates of C_1, C_2 and C_3, which are described as

(4) $C_1 = 0.999 X_1' + 0.062 X_2' + 0.999 X_3'$

(5) $C_2 = -0.036 X_1' + 0.998 X_2' - 0.026 X_3'$

(6) $C_3 = 0.037 X_1' + 0.002 X_2' - 0.037 X_3'$

Table 5: Standardized Independent Variables and Components

Component	Eigen Values	Variance %	Cummulative Variance %	Standardized Independent Variable		
				X'_1	X'_2	X'_3
C_1	1.999	66.638	66.638	0.999	0.062	0.999
C_2	0.999	33.272	99.910	-0.036	0.998	-0.026
C_3	0.003	0.090	100.00	0.037	0.000	-0.037

Now OLS technique is applied on PCs extracted and the $\hat{\alpha}$ vector is estimated. After it, the estimate of original parameter vector $\hat{\beta}$ is obtained using $\hat{\alpha} = V'\hat{\beta}$. After simplification the best standardized fitting linear regression equation is obtained as

(7) $\hat{Y} = 0.966\ X_1 + 0.240\ X_2 + 0.971\ X_3$

2.2 RIDGE REGRESSION

Hoerl and Kennard (1970 a, b) proposed the technique of RR utilizing $\hat{\beta}_R = (X'X + kI)^{-1} X'Y$ in lieu of $\hat{\beta} = (X'X)^{-1} X'Y$, where $\hat{\beta}_R$ and $\hat{\beta}$ are the ridge and OLS estimates of the parameter vector, β respectively. The genesis of RR reclines with a paper by Hoerl (1959) in which he discussed optimization from response surface point of view. Tichonoff (1943) provided a method of regularization in Russian language utilized for ill-conditioned problems and this is popularly known as 'Tikhonov Regularization (TR)'. TR is also recognized as RR in Statistics. Anders (2001) suggested that RR is an application of TR, a method that has been explored in the approximation theory literature for about as long as RR has been used in Statistics. Singh (2011) concluded that it is highly appropriate and legitimate to give credit to Tychonoff's TR having findings more general in nature than H-K's RR having contextual in nature. H-K expounded the finite dimensional case only of TR under Statistical approach, while Tychonoff's TR is a general case.

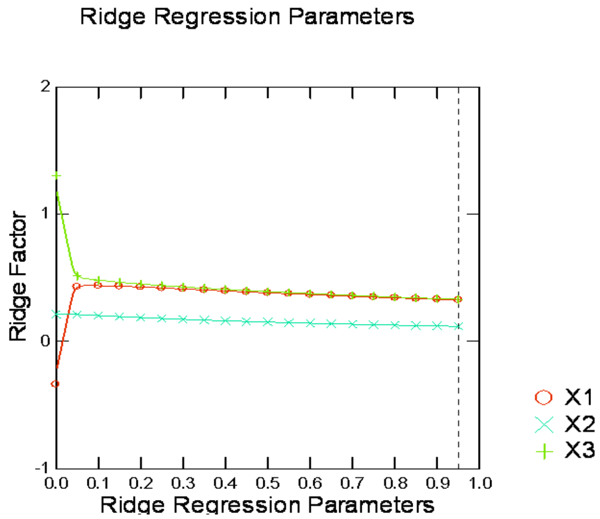

Fig 3: Ridge Regression Parameters

Fig 3 demonstrates parameters of RR showing presence of multi-co linearity. Again data given in Table 1 is used to apply RR technique for getting the estimates of coefficients as given in Table 6 below. H-K suggested the value of k from 0 to 1.

The value of k lies between 0.250 and 0.350 to get almost stability in estimates of RR coefficients by observing Table 6 and they are $X_1 = 0.411$, $X_2 = 0.173$ and $X_3 = 0.427$ at k = 0.300 and regression coefficients is stable with k = 0.300. The corresponding fitting of linear regression equation is as

(8) $\hat{Y} = 0.411\ X_1 + 0.173\ X_2 + 0.427\ X_3$

 DOI 10.17148/IARJSET.2018.578

Table 6: RR Standardized Coefficients

Standardized Ridge Coefficients			
LAMBDA	X_1	X_2	X_3
0.000	-0.339	0.213	1.303
0.050	0.427	0.211	0.513
0.100	0.436	0.202	0.481
0.150	0.433	0.194	0.464
0.200	0.426	0.186	0.450
0.250	0.419	0.179	0.438
0.300	0.411	0.173	0.427
0.350	0.403	0.167	0.417
0.400	0.396	0.161	0.408
0.450	0.388	0.156	0.399
0.500	0.381	0.151	0.391
0.550	0.374	0.147	0.383
0.600	0.367	0.142	0.376
0.650	0.360	0.138	0.368
0.700	0.354	0.135	0.361
0.750	0.348	0.131	0.355
0.800	0.342	0.127	0.348
0.850	0.336	0.124	0.342

Between PCR and RR, it is better to apply PCR technique to get the estimates of coefficients. As we know that the multi-co linearity never is removed from the data. Using RR its degree is reduced only. But in PCR, the multi-collinear explanatory variables are transformed into orthogonal or non multi-collinear PCs using PCA technique. The concept of multi-co linearity does not arise in all constructed PCs and the technique of PCR is easily applied to get the estimate of parameter vector without any doubt about the presence multi-co linearity.

3. CONCLUDING REMARKS

Multi-co linearity is a statistical event where a perfect (very rare) or exact relationship exists among explanatory variables and in this situation; it is not easy to obtain reliable estimates of their individual coefficients by OLS technique. Presence of multi-co linearity can cause serious problems with the estimation of β vector with its interpretation. OLS estimators are imprecisely estimated due to presence of multi-co linearity in the data. Multi-co linearity is a big problem, if goal is to understand how the various X variables impact on Y. It became very necessary to detect and solve the issue of multi-co linearity before estimating the parameter based on fitted regression model. VIF, tolerance, condition number, condition index and others are used to detect the multi-co linearity. Remedial measures facilitate in obtaining estimates of coefficients in the presence of multi-co linearity. RR is a statistical technique to obtain estimate of coefficients with multi-collinear explanatory variables and the technique developed by H-K in 1970 but discovery of RR by Tychonoff in 1943 was much prior to H-K. Thus controversy lies between Tychonov and H-K on propriety of its discovery. Tychonov's discovery on RR should be given credit than that of H-K's invention of RR because Tychonov's discovery was much earlier and general in nature.

Stability in estimates of RR coefficients are $X_1 = 0.411$, $X_2 = 0.173$ and $X_3 = 0.427$ at $k = 0.300$ and regression coefficients is stable with the $k = 0.300$. The corresponding fitting linear regression equation is:

$$\hat{Y} = 0.411\, X_1 + 0.173\, X_2 + 0.427\, X_3$$

PCR is an additional statistical technique to obtain estimate of coefficients in the presence of multi-co linearity. The PCR estimate gives following fitting of linear regression equation:

$$\hat{Y} = 0.966\, X_1 + 0.240\, X_2 + 0.971\, X_3$$

PCR provided excellent result than RR during the study applied on Table 1. Because in PCR, PCs are extracted first and they are uncorrelated or orthogonal. Hence OLS technique is applied on extracted PCs to obtain the estimates of coefficients. This property makes the PCR superior to RR. Both, PCR and RR are biased estimators with greater efficiency than OLS estimator.

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology
ISO 3297:2007 Certified
Vol. 5, Issue 7, July 2018

REFERENCES

[1] Anders, B. (2001), "Ridge Regression and Inverse Problems", Stockholm University, Sweden.
[2] Curtis, SM and Ghosh S K (2011), "A Bayesian Approach to Multi-co linearity and the Simultaneous Selection and Clustering of Predictors in Linear Regression", Journal of Statistical Theory and Practice, Vol. 5 (4), pp.715-735.
[3] Hoerl AE and Kennard RW (1959), "Optimum Solution of Many Variables Equations", Chemical Engineering Process, Vol. 55, pp. 11-21.
[4] Hoerl A E, Kennard R W. (1970 a), "Ridge Regression: Biased Estimation of Non orthogonal Problems", Techno metrics, Vol. 12 (1), pp. 55-67.
[5] Hoerl A E, Kennard R W. (1970 b), "Ridge Regression: Application to Non orthogonal Problems", Techno metrics, Vol. 12 (1), pp. 69-82.
[6] Niemelä-Nyrhinen J, Leskinen E. (2014), "Multi-co linearity in Marketing Models: Notes on the Application of Ridge Trace Estimation in Structural Equation Modeling", The Electronic Journal of Business Research Methods, Vol. 12 (1), pp. 3-15.
[7] Singh R. (2011), "Two Famous Conflicting Claims in Econometrics", Asian-African Journal of Econometrics, Vol. 11 (1), pp. 185-186.
[8] Tychonoff A N. (1943), "Об устойчивости обратных задач [On the Stability of Inverse Problems]. Doklady Akademii Nauk SSSR, Vol. 39 (5), pp. 195-198.