Int. J. Contemp. Math. Sciences, Vol. 6, 2011, no. 12, 585 - 600

Solving Multicollinearity Problem

Using Ridge Regression Models

M. El-Dereny and N. I. Rashwan

Department of Statistics & Mathematics

Tanta University, Faculty of Commerce, Egypt
nasrrashwan@yahoo.com

Abstract

In this paper, we introduce many different Methods of ridge regression to solve multicollinearity problem. These Methods include ordinary ridge regression (ORR), Generalized ridge regression (GRR), and Directed ridge regression (DRR). Properties of ridge regression estimators and Methods of selecting biased ridge regression Parameter are discussed. We use data simulation to make comparison between Methods of ridge regression and ordinary least squares (OLS) Method. According to a results of This study, we found That all Methods of ridge regression are better than OLS Method when the Multicollinearity is exist.

Keywords: Ordinary ridge regression, Generalized ridge regression, Directed ridge regression, Multicollinearity, biased ridge parameter, and ordinary least squares

1-Introduction

Multicollinearity refers to a situation in which or more predictor variables in a multiple regression Model are highly correlated if Multicollinearity is perfect

(EXACT), the regression coefficients are indeterminate and their standard errors are infinite, if it is less than perfect. The regression coefficients although determinate but posses large standard errors, which means that the coefficients can not be estimated with great accuracy (Gujarati, 1995).

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 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. (Vinod and Ullah, 1981).

There are many methods used to detect multicollinearity, among these methods :

i) Compute the correlation matrix of predictors variables, a high value for the correlation between two variables may indicate that the variables are collinear. This method is easy, but it can not produce a clear estimate of the rate (degree) of multicollinearity.

ii) Eigen structure of $X'X$, let $\lambda_1, \lambda_2, \ldots, \lambda_p$ be The eigenvalues of $X'X$ (in correlation form). When at least one eigenvalue is close to zero, then multicollinearity is exist (Greene, (1993), Walker, (1999)).

iii) Condition number : there are several methods to compute the condition number (ϕ) which indicate degree of multicollinearity Vinod and Ullah, (1981), suggested that the condition number is given by :

$$\phi_1 = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \quad \ldots\ldots\ldots \ 1$$

Montgomery and Peck, (1992) defined the condition number as the ratio of λ_{max} and λ_{min} as :

$$\phi_2 = \frac{\lambda_{max}}{\lambda_{min}} \ldots\ldots\ldots \quad \ldots 2$$

Where λ_{max} is the largest eigenvalue.

λ_{min} is the smallest eigenvalue

if $\lambda_{min} = 0$, Then ϕ is infinite, which means that perfect multicollinearity between predictor variables. If λ_{max} and λ_{min} are equal, Then $\phi = 1$ and the predictors are said to be orthogonal. Pagel and Lunneborg, (1985) suggested that the condition number is :

$$\phi_3 = \sum_{i=1}^{P} \frac{1}{\lambda_i} \qquad \ldots\ldots\ldots \quad \ldots\ldots\ldots \; 3$$

Some conventions are that if ϕ lies between 5 and 30 it considered that multicollinearity goes from Moderate to strong. iv) variance inflation factor (VIF) can be computed as follow :

$$VIF = \frac{1}{1 - R_j^2} \qquad \ldots\ldots\ldots \quad \ldots \; 4$$

Where R_j^2 is the coefficient of determination in the regression of explanatory variable X_j on the remaining explanatory variables of the model. Generally, when VIF > 10, we assume There exists highly multicollinearity (Sana and Eyup, 2008).

v) Checking the relationship between the F and T tests might provide some indication of the presence of multicollinearity. If the overall significance of the model is good by using F-test, but individually the coefficients are not significant by using t-test, then the model might suffer from multicollinearity.

Multicollinearity has several effects, these are described as follow :

- High variance of coefficients may reduce the precision of estimation.
- Multicollinearity can result in coefficients appearing to have the wrong sign.
- Estimates of coefficients may be sensitive to particular sets of sample data.

- Some variables may be dropped from the model although, they are important in the population.
- The coefficients are sensitive of to the presence of small number inaccurate data values (more details in Judge 1988, Gujarat; 1995).

Because multicollinearity is a serious problem when we need to make inferences or looking for predictive models. So it is very important for us to find a better method to deal with multicollinearity. Therefore, The main objective in this paper, is to introduce different models of ridge regression to solve multicollinearity problem and make comparison between these models of ridge regression with the ordinary least square method.

2-The ordinary ridge regression (ORR).

Consider the standard model for multiple linear regression :

$$Y = X\beta + E \qquad \ldots 5$$

Where Y is (n × 1) vector of the dependent variable values, X is (n×p) matrix contains the values of P predictor variables and this matrix is full Rank (matrix of rank p), β is a (p × 1) vector of unknown coefficients, and E is a (n × 1) vector of normally distributed random errors with zero mean and common variance $\sigma^2 I$. Note that, Both X's and Y have been standardized.

The OLS estimate $\hat{\beta}$ of β is obtained by minimizing the residual sum of squares, and are given by:

$$(Y - X\hat{\beta})'(Y - X\hat{\beta}),$$

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

$$Var(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1} ..$$

and
$$MSE(\hat{\beta}) = \hat{\sigma}^2 \operatorname{trace}(X'X)^{-1}$$

$$= \hat{\sigma}^2 \sum_{i=1}^{P} \frac{1}{\lambda_i} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 6$$

Where $\hat{\sigma}^2$ is the mean squares error. This estimator $\hat{\beta}$ is an unbiased and has a minimum variance. However, if $X'X$ is ill-conditioned (singular), the OLS estimate tend to become too and some of coefficients have wrong sign (Wethrill, 1986). In order to prevent these difficulties of OLS, Hoerl and Kennard (1970), suggested the ridge regression as an alternative procedure to the OLS method in regression analysis, especially, multicollinearity is exist. The ridge technique is based on adding a biasing constants K's to the diagonal of $X'X$ matrix before computing $\hat{\beta}$'s by using method of Hoerl and Kennard (2000). Therefore, the ridge solution is given by :

$$\hat{\beta}(K) = (X'X + KI)^{-1}X'Y \ , K \geq 0 \ \ .... 7$$

Where K is ridge parameter and I is identity matrix. Note that if K = 0, the ridge estimator become as the OLS. If all K's are the same, the resulting estimators are called the ordinary ridge estimators (John, 1998).

3-Properties of ordinary ridge regression estimator

The ridge regression estimator has several properties, which can be summarized as follow :

- From equation (7), by taking expectation on both sides, then

  $$E\left(\hat{\beta}(K)\right) = A_k \beta, \quad \text{where} \quad A_k = [I + K(X'X)^{-1}]^{-1} \text{ and}$$

  $$\text{Var}\left(\hat{\beta}(K)\right) = \hat{\sigma}^2 A_k (X'X)^{-1} A_k'.$$

  So $\hat{\beta}(K)$ is a biased estimator but reduce the variance of the estimate.

- $\hat{\beta}$ (K) is the coefficient vector with minimum length, this means that K > 0 always exists, for which the squares length of $\hat{\beta}$ from β on average. Thus the ridge estimate shrinks the OLS estimate.

- Because $\hat{\beta}(K) = [I + K(X'X)^{-1}]^{-1}\hat{\beta}$, the ridge estimator is a linear transformation of the OLS.

- The sum of the squared residuals is an increasing function of K.
- The mean squares error of $\hat{\beta}(K)$ is given by :

$$MSE\left(\hat{\beta}(K)\right) = E[\hat{\beta}(K) - \beta)'(\hat{\beta}(K) - \beta)]$$

$$= \hat{\sigma}^2 \ \text{trace}[A_k (X'X)^{-1} A'_k] + \hat{\beta}'(I - A_k)'(I - A_k)\hat{\beta}$$

$$= \hat{\sigma}^2 \sum_{i=1}^{P} \frac{\lambda_i}{\lambda_i + k} + K^2 \hat{\beta}'(X'X + KI)^{-2} \hat{\beta} \ldots\ldots\ldots\ldots 8$$

Note that, the first term of the right hand in equation (8) is the trace of the dispersion matrix of the $\hat{\beta}$ (K) and the second term is the square length of the bias vector.

- From the previous property, we find that, for K > 0, the variance term is monotone decreasing function of K and The squares bias is monotone increasing function of K. Therefore the suitable choice if K is determined by striking a balance between two terms, so we select K which achieved reduce in variance larger than the increase is bias.
- There always exists a K > 0, such that $\hat{\beta}(K)$ has smaller MSE than $\hat{\beta}$,
- this means that MSE $\left(\hat{\beta}(K)\right)$ < MSE $(\hat{\beta})$. *(More details see Judge, 1988, Gujarat; 1995, Gruber 1998, Pasha and Shah 2004)*

4-The generalized Ridge regression (GRR) :

Let P is a (p × p) matrix with columns as eigenvectors (V_1, V_2, …. V_P) of $X'X$, $P'P = I$ and hence $P'(X'X)P = \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots \lambda_p)$. Then we can rewrite the linear model as :

$$Y = X\beta + E$$

$$= (XP)(P'\beta) + E$$

$$\therefore Y = X^* \alpha + E \qquad \ldots 9$$

Where $X^* = XP$, $\alpha = P'\beta$

This model is called canonical linear model or uncorrelated components model. The OLS estimate for α is given by :

$$\hat{\alpha} = (X^{*\prime}X^{*})^{-1}X^{*\prime}Y$$

$$= \Lambda^{-1}X^{*\prime}Y \text{..........................} 10$$

and Var $(\hat{\alpha}) = \hat{\sigma}^2(X^{*\prime}X^{*})^{-1} = \hat{\sigma}^2\Lambda^{-1}$

which is diagonal. This show the important property of this parameterization since the elements $\hat{\alpha}_1, \hat{\alpha}_2, ..., \hat{\alpha}_p$ of $\hat{\alpha}$ are uncorrelated (Pagel and Lunneborg 1985).

Note that the original regression coefficient of OLS is $\hat{\beta} = P\hat{\alpha}.$

The ridge estimator for α is given by :

$$\hat{\alpha}(K) = (X^{*\prime}X^{*} + K)^{-1}X^{*\prime}Y$$

$$= (\Lambda + K)^{-1}X^{*\prime}Y \text{.......... ....} 11$$

$$= (I + K\Lambda^{-1})^{-1}\hat{\alpha} = C_k\hat{\alpha} = \text{diag} \ (\frac{\lambda_i}{\lambda_i + K_i})$$

where K is the diagonal matrix with entries (k_1, k_2, …, k_p). This estimate is known as a generalized ridge estimate (Sang 2001) and the mean squares error of $(K)\hat{\alpha}$ is given by :

$$\text{MSE} \ (\hat{\alpha}(k) = \hat{\sigma}^2 \text{trace} \ (C_k\Lambda^{-1}C'_k) + (C_k - I)\hat{\alpha} \ \hat{\alpha}'(C_k - I)'$$

$$= \hat{\sigma}^2 \sum_{i=1}^{P} \frac{\lambda_i}{(\lambda_i + k_i)^2} + \sum_{i=1}^{P} \frac{K_i^2\hat{\alpha}_i^2}{(\lambda_i + k_i)^2} \text{.......... ...} 12$$

Vinond and Ullah, (1981) prove That $\text{MSE}(\hat{\alpha}(K)) \leq \text{MSE}(\hat{\alpha}),$

when matrix K satisfies $K_i = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2} = \text{diag}\ (\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}, \frac{\hat{\sigma}^2}{\hat{\alpha}_2^2}, \ldots\ldots, \frac{\hat{\sigma}^2}{\hat{\alpha}_p^2}).$

These values of K are minimize the MSE of generalized ridge estimator, note that the original form of ridge can be converted back from the canonical from by :

$$\hat{\beta}(K) = P\,\hat{\alpha}(K) \ldots\ldots\ldots \quad \ldots\ldots\ldots\ldots \quad \ldots\ldots \ 13$$

Horal and Kennard (2000), show that the generalized regression has some properties that can be summarized as follow :

* $\hat{\alpha}(K) = (X^{*'} X^{*})^{-1} X^{*'} y = C_k \hat{\alpha}$

*$E(\hat{\alpha}(K)) = C_k \alpha$ which refer that $\hat{\alpha}(K)$ is a biased estimator of α.

* For $K \neq 0$ where 0 is a (p*p) zero matrix, if $\| \hat{B} \| \neq 0$, we always have

$\| \hat{\alpha}\ (K) \|^2 \leq \| \hat{\alpha} \|^2$

* There exists $K > 0$ such MSE $(\hat{\alpha}(K)) \leq \text{MSE}(\hat{\alpha})$

* Generalized ridge regression is a form of Bayes estimation. (Gurber, 1998).

5-The directed ridge regression (DRR) :

Guilkey and Murphy (1975), proposed a technique called directed ridge regression. This method of estimation based on the relationship between the eigenvalues of $X'X$ and the variance of $\hat{\alpha}_i$

Since Var $(\hat{\alpha}) = \sigma^2 \Lambda^{-1}$. relatively precise estimation is achieved for corresponding to large eigenvalues, while relatively imprecise estimation is achieved for α_i corresponding to small eigenvalues.

As a result of adjusting only those elements of Λ corresponding to the small eigenvalues of $X'X$. The DRR estimator results in an estimate of $\hat{\alpha}_i$ that is less biased than the resulting from GRR estimator.

The steps of computing the directed ridge estimators are summarized as follow :

1 - Find mean squares error by OLS $\hat{\sigma}^2$ and $\hat{\alpha} = \Lambda^{-1} X^{*\prime} Y$.

2 - Find the eignvalues λ_i, matrix of eigenvectors of $X^{\prime}X$ and $X^* = PX$.

3 - Find $K_i(0) = \hat{\sigma}^2 / \alpha_i^2$, note that K is only added to diagonal element of $\lambda_i \geq 10^{-c} \lambda_{max}$ where c is arbitrary.

4 - Compute the directed ridge estimator

$$\hat{\alpha}(dk)^{(0)} = (\Lambda + KI)^{-1} X^{*\prime} Y$$

5 - Re-estimated

$$K_i(1) = \hat{\sigma}^2 / \hat{\alpha}_i (dk)^{(0)}$$

6 - Repeat steps 4 and 5 until obtaining the optimal values of K_i, which make the difference of squared length of $\hat{\alpha}(dk)^{(i)2}$ and $\hat{\alpha}(dk)^{(i+1)2}$ is very small. Let m is the number of iterations then

$$\hat{\alpha}(dk)^{(m)} = (\Lambda + KI)^{-1} X^{*\prime} Y$$

where K is the diagonal matrix with entries $[(K_1(m\text{-}1), K_2(m\text{-}1), \ldots K_p(m\text{-}1)]$. Note that the original regression of ORR is

$$\hat{\alpha}(K) = P\hat{\alpha}(dk)^{(m)}$$

6-Choice of ridge parameter (K)

The ridge regression estimator does not provide a unique solution to the problem of multicollinearity but provides a family of solution. These solutions depend on the value of K (the ridge biasing parameter). So there are many suggestions for computing K can be summarized as follow :

- Ridge trace :

  Hoerl and Kennard (1970), suggested a graphical method called ridge trace to select the value of the ridge parameter K. This is a plot of the values of individual components of $\hat{\beta}$ (K) against a range of values of K (0 < K < 1). They select the minimum value of K for which all $\hat{\beta}$ (K) become stable and the wrong signs of same coefficients become the correct and also the residual sum of square is not too large compared to its minimum at subjective way because different persons can select different values of K even if are looking at the same ridge trace (John, 1998). So, there are many other objective ways to calculate the ridge parameter are summarized below :

- Hoerl and Kennard (1970), proposed method to select K_i in the case of generalized ridge estimation where, $k_i = \frac{\hat{\sigma}^2}{\alpha_i^2}$
- Hoerl, Kennard and Baldwin (1975), suggested anther method to select a single K value of all K_i, This method is called the ordinary ridge estimator. The K value by using this method can be computed by :

$$\hat{K}(HKB) = P\hat{\sigma}^2 / \hat{\beta}'\beta \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots 14$$

where P is the number of predictor variables

$\hat{\sigma}^2$ is the Mean squared error by OLS

$\hat{\beta}$ is the vector of estimators by OLS

- Lawless and Wang (1976) gave the following method to compute the value of K (Hoerl and Kennard, 2000).

$$\hat{K}(LW) = P\hat{\sigma}^2 / \hat{\beta}'X'X\hat{\beta} \quad \ldots\ldots\ldots\ldots\ldots 15$$

- McDonaled and Galarneau (1975) have proposed that K be selected as the solution to the following equation. (John, 1998).

$$\hat{\beta}'(k)\ \hat{\beta}(k) = \hat{\beta}'\ \hat{\beta} - \hat{\sigma}^2 (X'X)^{-1} \quad \ldots\ldots\ldots\ldots 16$$

- Hoerl and Kennard (1976), suggested anther method of estimating K by using the iterative method (Angshuman, 1998) which is described as follow :

1 - Compute $\hat{\beta}$ and $\hat{\sigma}^2$

2 - Find $K(a_0) = P\hat{\sigma}^2 / \hat{\beta}'\hat{\beta}$

and compute $\left(K(a_0)\right) = [X'X + K(a_0)I]^{-1} X'Y$

3 - Find $K(a_1) = P\hat{\sigma}^2 / \hat{\beta}'\left(K(a_0)\right)\hat{\beta}\left(K(a_0)\right)$

and compute $\hat{\beta}\left(K(a_1)\right) = (X'X + K(a_1)I)^{-1} X'Y$

4 - Find $K(a_2) = P\ \hat{\sigma}^2 \hat{\beta}'\left(K(a_1)\right)\hat{\beta}\left(K(a_1)\right)$ then compute

$\hat{\beta}\left(K(a_2)\right) = (X'X + K(a_2)I)^{-1} X'Y$ *and continue*

until $[K(a_{i-1}) - K(a_i)] / K(a_i) > \delta$,

where $\delta = 20\,T^{-1.3}$ and $T = \dfrac{\text{trace}(X'X)^{-1}}{n}$

- Dempster, Schatzoff and Wermuth (1977), suggested method to select K by the solution of K in the equation

$$P = \sum_{i=1}^{P} C_i^2 / (\hat{\sigma}^2 K^{-1} + \sigma^2 \lambda_i)$$

Where C_i is the i^{th} element of unstandardized uncorrelated components of $\hat{\beta}(K)$. (more details about these methods, see Myers 1990, Hoerl and Kennard 2000, Kibria 2003 and Munize and Kibria 2009).

In this study, we depend on the following method to compute the value of K :

- $\hat{K}(HKB) = P\hat{\sigma}^2 / \hat{\beta}'\hat{\beta}$

  $K(Lw) = P\hat{\sigma}^2 / \hat{\beta}'\ X'X\ \hat{\beta}$ in the case of the ordinary ridge regression model.

- $\hat{K} = \hat{\sigma}^2 / \hat{\alpha}_i^2$ ، $i = 1,2,\ldots, p$

in the case of generalized ridge regression model.

- Directed method with assumption that C = 2 in the case of the directed ridge regression model.

7-Comparison between the OLS method and ridge regression methods

In this study, we compare between ridge regression method and OLS method depending on the standard deviation for individual estimators, mean squares error of estimators for each method and the predictive accuracy of model based on the coefficient of determination R^2 for each method. We will use simulated data in the study to make comparison between OLS and RR methods.

8-Results and discussion

In this research, we simulate a set of data using SAS package, where the correlation coefficients between the predictor variables (X's) are large (the number of predictor variables in this study are six variables). Table (1) shows the correlation matrix based on a set of simulated data

Table (1) Correlation Matrix

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1					
X_2	0.944	1				
X_3	0.957	0.948	1			
X_4	0.959	0.973	0.946	1		
X_5	0.972	0.969	0.973	0.963	1	
X_6	0.936	0.961	0.979	0.937	0.970	1

First : A sample of size 50 observations is simulated for 2000 iterations, and the following indicators are computed as :

- The eigenvalues of the predictor correlation matrix are as follow :

$\lambda_1 = 5.798$ $\lambda_2 = 0.084$ $\lambda_3 = 0.064$

$\lambda_4 = 0.027$ $\lambda_5 = 0.015$ $\lambda_6 = 0.011$

- The variance inflation factor (VIF) for all variables X's are as follow :

VIF (X_1) = 24.7 VIF (X_2) = 37.01 VIF (X_3) = 42.47

VIF (X_4) = 29.03 VIF (X_5) = 46.93 VIF (X_6) = 43.22

- The condition number ϕ_1, ϕ_2 and ϕ_3 as follow :

$\phi_1 = 24$ $\phi_2 = 527$ $\phi_3 = 222$

From the previous indicators, it is obvious that there are a serious multicollinearity problem because there is one of eigenvalues (λ_6) close to zero, all VIF's values more than 10, and the different types of condition number more than 5.

Second : Using both OLS method and all methods of RR to analyze the simulated data, we get the following results :

- The regression coefficients and standard deviations of these coefficients can be summarized in Table (2) :

Table (2)

Estimates of coefficients and standard deviation using OLS and RR method

OLS		ORR_1		ORR_2		GRR		DRR	
Coef	STDEV	Coef	STDEV	Coef	STDEV	Coef	STDEV	Coef	STDEV
1.301	0.206	1.172	0.204	1.287	0.204	0.378	0.007	0.378	0.007
-0.403	0.241	-0.333	0.205	-6.395	0.236	0.478	0.159	0.470	0.164
-0.115	0.218	-.034	0.184	-0.106	0.212	0.789	0.211	0.783	0.215
0.565	0.203	0.507	0.171	0.559	0.159	0.323	0.161	0.439	0.150
-0.714	0.399	-0.553	0.27	-0.697	0.293	0.909	0.379	1.024	0.411
0.309	0.357	0.180	0.203	0.294	0.221	0.488	0.311	0.674	0.341

Note that ,

we indicate to the ordinary ridge regression when $\hat{K}$ (HKB)= $P\hat{\sigma}^2/\hat{\beta}'\beta$ by ORR_1 and the ordinary ridge regression ,when $\hat{K}(Lw) = P\hat{\sigma}^2/\hat{\beta}'X'X\hat{\beta}$ by ORR_2 . Here, we must refer to the values of ridge parameter (K) in all methods of RR. These value can be summarized as follow :

$\hat{K}$ (HKB) = 0.157 $\hat{K}$(Lw) = 0.012

$\hat{K}_i$ (GRR) = [0.52 0.33 0.14 0.47 0.125 0.276]

$\hat{K}_i$ (DRR) = [0.53 0.41 0.166 0 0 0]

- Computing the mean squares error of regression coefficient for all method, we found that

  MSE (B's) OLS = 0.432 MSE ($\hat{B}$ (k)) ORR1 = 0.360

  MSE ($\hat{B}$(k)) ORR2 = 0.403 MSE$(\hat{\alpha}(k))$GRR = 0.322

  MSE$(\hat{\alpha}(k))$ DRR = 0.42

- Computing the coefficient of determination R^2 for each model, we obtain the following result. This result can be summarized in Table (3) as follow.

Table (3) R^2 for each Model

R-Square	
OLS	0.779
ORR1	0.975
ORR2	0.781
GRR	0.911
DRR	0.911

From the previous results, it is obvious that :

- All models of RR have smaller standard deviation than OLS.
- All models of RR have smaller MSE of regression coefficient than OLS.
- While, all models of RR have larger R^2 than OLS. consequently, all models of RR are better than OLS when the multicollinearity problem is exist in data.

9-Conclusions

in This research, we referred to the multicollinearity problem, methods of detecting of this problem and effect on a result of multiple regression model. Also, we introduced many different models of ridge regression to solve this problem and we make a comparison between RR methods and OLS by using a simulation data (2000 replications). Based on the standard deviation, mean square errors and coefficient of determination for estimators of each model. We noted that all ridge regression models are better than ordinary least square when the multicollinearity problem is exist and the best model is the generalized ridge regression because it has smaller MSE of estimators, smaller standard deviation for most estimators and has larger coefficient of determination

References

[1] S. Angshuman, *Application of Ridge Regression for Improved Estimation of Parameters in compartmental Models*, Ph.D. Dissertation, Washington University, 1998.

[2] W.H. Green, *Econometric Analysis*, Mac-Millan, New York,1993.

[3] D.K. Guilkey, and J.L. Murphy, *Directed Ridge Regression Techniques in cases of Multicollinearity*, Journal of American Statistical Association, 70(1975) , 767-775.

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

[5] H.J. Gurber, *Improving Efficiency by Shrinkage*, Marcel Dekker, New York, 1998.

[6] A.E. Hoerl, and R.W. Kennard, *Ridge Regression* : Biased Estimation for non orthogonal Problems, Technometrics, 42 (2000), 80 – 86.

[7] T.M. John, *Mean Squared Error Properties of the Ridge Regression Estimated Linear Probability Models*, Ph.D. Dissertation, University of Delaware, 1998.

[8] G.G. Judge, *Introduction to Theory and Practice of Econometrics*, John Willy and Sons, New York,1988.

[9] B.M Kibria, *Performance of Some Ridge Regression Estimators*, Commun.Statist – Simulation and computation,32 (2003), 419 – 435.

[10] D.C. Montgomery, and E.A. Peck, *Introduction to Linear Regression Analysis*, John Willy and Sons, New York,1992.

[11] G. Munize, and B.M. Kibria, *on Some Ridge Regression estimator : An Empirical Comparison*, Commun. Statist – Simulation and Computation, 38(2009), 621 – 630.

[12] R.H. Myers, *Classical and Modern Regression with Applications*, PWS – KENT Publishing Co., Boston,1990.

[13] M.D. Pagel, and C.E. Lunneborg, *Empirical Evaluation of Ridge Regression*, Psychological Bulletin, 97(1985), 342-355.

[14] G.R. Pasha, and M.A. Shah, *Application of Ridge Regression to Multicollinear Data*, J.res.Sci., 15(2004), 97-106.

[15] M. Sana, and. C. Eyup, *Efficient Choice of Biasing Constant for Ridge Regression*, Int. J. Contemp. Math. Sciences, 3(2008), 527 – 536.

[16] H. L. Song, *Comparison of Regression Techniques for certain patterned Matrices in the errors in variables Model*, Ph.D. Dissertation, American University, Washington, 2001.

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

[18] E.Walker, *Detection of Collinearity–Influential observations*, Communications in Statistics, Theory and Methods,18 (1989), 1675 – 1690.

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

Received: September, 2010