Using Ridge Regression model to solving Multicollinearity problem

Dr.Samira Muhammad salh

Abstract: In this (paepar) research ,we introduce two different method to solve multicollinearity problem .these methods include ordinary least square (OLS) and ordinary ridge regression (ORR),and using data simulation to comparison between methods ,for three different sample size (25,50,100).According to a results ,we found that ridge regression (ORR) are better than OLS Method when the Multicollinearity is exist. And we conclude that the sample size affects the results of estimated value, whenever the sample size increases, the case results of the methods of Estimation stables more.

Keywords: multiple regressions, Multicollinearity problem, ordinary least square, ordinary ridge regression.

1.INTRODUCTION:

Multiple regressions is a statistical technique that allows us to predict someone's score on one variable on the basis of their scores on several other variables. An example might help. Suppose we were interested in predicting how much an individual enjoys their job. Variables such as salary, extent of academic qualifications, age, sex, number of years in full-time employment and socioeconomic status might all contribute towards job satisfaction. If we collected data on all of these variables, perhaps by surveying a few hundred members of the public, we would be able to see how many and which of these variables gave rise to the most accurate prediction of job satisfaction. We might find that job satisfaction is most accurately predicted by type of occupation, salary and years in full-time employment, with the other variables not helping us to predict job satisfaction. When using multiple regression in psychology, many researchers use the term "independent variables" to identify those variables that they think will influence some other "dependent variable". We prefer to use the term "predictor variables" for those variables that may be useful in predicting the scores on another variable that we call the "criterion variable". Thus, in our example above, type of occupation, salary and years in full-time employment would emerge as significant predictor variables, which allow us to estimate the criterion variable – how satisfied someone is likely to be with their job. As we have pointed out before, human behavior is inherently noisy and therefore it is not possible to produce totally accurate predictions, but multiple regression allows us to identify a set of predictor variables which together provide a useful estimate of a participant's likely score on a criterion variable. If two variables are correlated, then knowing the score on one variable will allow you to predict the score on the other variable. The stronger the correlation, the closer the scores will fall to the regression line and therefore the more accurate the prediction. Multiple regression is simply an extension of this principle, where we predict one variable on the basis of several other variables. Having more than one predictor variable is useful when predicting human behavior, as our actions, thoughts and emotions are all likely to be influenced by some combination of several factors. Using multiple regression we can test theories (or models) about precisely which set of variables is influencing our behavior.

Dr.Samira muhamad Salh ,Department of statistics, University of sulamani(samira196902@yahoo.com)

Human behavior is rather variable and therefore difficult to predict. What we are doing in both ANOVA and multiple regression is seeking to account for the variance in the scores we observe. Thus, in the example above, people might vary greatly in their levels of job satisfaction. Some of this variance will be accounted for by the variables we have identified. For example, we might be able to say that salary accounts for a fairly large percentage of the variance in job satisfaction, and hence it is very useful to know someone's salary when trying to predict their job satisfaction. You might now be able to see that the ideas here are rather similar to those underlying ANOVA . In ANOVA we are trying to determine how much of the variance is accounted for by our manipulation of the independent variables (relative to the percentage of the variance we cannot account for). In multiple regression we do not directly manipulate the IVs but instead just measure the naturally occurring levels of the variables and see if this helps us predict the score on the dependent variable (or criterion variable). Thus, ANOVA is actually a rather specific and restricted example of the general approach adopted in multiple regression.To put another way, in ANOVA we can directly manipulate the factors and measure the resulting change in the dependent variable. In multiple regression we simply measure the naturally occurring scores on a number of predictor variables and try to establish which set of the observed variables gives rise to the best prediction of the criterion variable. A current trend in statistics is to emphases the similarity between multiple regression and ANOVA, and between correlation and the t-test. All of these statistical techniques are basically seeking to do the same thing – explain the variance in the level of one variable on the basis of the level of one or more other variables. These other variables might be manipulated directly in the case of controlled experiments, or be observed in the case of surveys or observational studies, but

ISSN 2229-5518
the underlying principle is the same. Thus, although we have given separate chapters to each of these procedures they are fundamentally all the same procedure.

The usage of multiple egressions:-

1. You can use this statistical technique when exploring linear relationships between the predictor and criterion variables – that is, when the relationship follows a straight line. (To examine non-linear relationships, special techniques can be used.)

2.The criterion variable that you are seeking to predict should be measured on a continuous scale (such as interval or ratio scale).There is a separate regression method called logistic regression that can be used for dichotomous dependent variables (not covered here).

3. The predictor variables that you select should be measured on a ratio, interval, or ordinal scale. A nominal predictor variable is legitimate but only if it is dichotomous, i.e. there are no more that two categories. For example, sex is (masculine, feminine and androgynous) could not be coded as a single variable. Instead, you would create three different variables each with two categories (masculine/not masculine; feminine/not feminine and androgynous/not androgynous). The term dummy variable is used to describe this type of dichotomous variable.

4. Multiple regressions require a large number of observations. The number of cases (participants) must substantially exceed the number of predictor variables you are using in your regression. The absolute minimum is that you have five times as many participants as predictor variables. A more acceptable ratio is 10:1, but some people argue that this should be as high as 40:1 for some statistical selection methods.

2. TERMINOLOGY:

There are certain terms we need to clarify to allow you to understand the results of this statistical technique:

2.1 R, R Square, Adjusted R Square:

R is a measure of the correlation between the observed value and the predicted value of the criterion variable. In our example this would be the correlation between the levels of job satisfaction reported by our participants and the levels predicted for them by our predictor variables. R Square (R^2) is the square of this measure of correlation and indicates the proportion of the variance in the criterion variable which is accounted for by our model – in our example the proportion of the variance in the job satisfaction scores accounted for by our set of predictor variables (salary, etc.). In essence, this is a measure of how good a prediction of the criterion variable we can make by knowing the predictor variables. However, R square tends to somewhat over-estimate the success of the model when applied to the real world, so an Adjusted R Square value is calculated which takes into account the number of variables in the model and the number of observations (participants) our model is based on. This Adjusted R Square value gives the most useful measure of the success of our model. If, for example we have an Adjusted R Square value of 0.75 we can say that our model has accounted for 75% of the variance in the criterion variable.

2.2 MULTICOLINEARITY PROBLEM

Multicolinearity is occurred where two or more explanatory random variables are correlated linearly by a strong linear relation, that it is not easy to separate the effect of each of them from the (Response) dependent variable. These strong correlations always made the multiple correlations are almost necessary. Multicolinearity is happened if the observations of one of the explanatory (xi 's) are equals or one of the explanatory is a function of the others, then in order to apply (OLS) method for estimation, it is required to summarize the assumption related with multicolinearity. As follow: (There is no exact or Simi-linear relationship among the explanatory or independent variables, moreover it is necessary that the number of parameters have been estimated must be less than the sample size(n) under consideration).

2.3 THE (OLS) ASSUMPTIONS IN (GLM):-

The General linear model formula is given by :

$$\underline{Y} = \underline{X}\beta + \underline{U} \qquad U \sim N(0,\sigma 2u) \text{ ....(1)}$$

$$E(Ui) = 0 \quad , \qquad \text{Then } E(Y) = X\beta$$

1. Var(Ui) = σ2u means that the variance of r.v U, is constant for all terms for all independent r.v's. Then U ~ N(0, σ2u)
2. covariance (UiUj) = E (UiUj) = 0 That means the various values of r.v (U)are uncorrelated with each other
3. E(UiXi) = 0 , or the values of (Ui's) are uncorrelated withany one of the explanatory variables (Xi's) this assumption is achieved by fixing the values of (Xi) with the same value in other sample.
4. It is required that the explanatory (Xi;s) are uncorrelated, so the fitted multiple regression model is doesn't including Multicolinarity.
5. The linear relation that required to be estimated by a multiple regression model is required to be (identified), i.e. the economic model under consideration have a discriminate form and doesn't contains the same variables that exist in other relation in the same problem under consideration.

2.4 SOME PROPERTIES OF DEPENDENT RANDOM VARIABLE IN (GLM):

Concerning with the properties for dependent random variable (Yi), are:

1. (Yi) is distributed normal with mean given by:

$$\overline{Y} = E(Y_i) = \beta_o + \beta_1 X_i \qquad \text{...(2)}$$

2. The variance of (Yi) is given by:

$$Var(Y_i) = E[(Y_i) - E(Y_i)]^2 = \sigma_u^2 \qquad \text{...(3)}$$

These properties can be formulated as follow:

ISSN 2229-5518

$$Y_i \sim N(\beta_o + \beta_1 X_i, \sigma_u^2)$$

After illustrating the main assumptions for (GLM), now one can conclude the effect of multicolinearity that concerning with assumption (5). Means, if the explanatory r.v;s are exactly correlated , then the estimation of parameters of the linear model is impossible, that's because the inverse matrix of fisher information matrix (X′X) would be not valid, this is returned to the fact that the determinant of (X′X) is vanished (Zero), moreover the variance for the estimators attains to infinity($\propto$) . But in other hand if the explanatory r.v;s are highly correlated it is possible to estimate the parameters but the value of determinant of (X′X) is very small that makes the variance of estimators very large and attains to unexpected results about the model, see the following variance formula.

$$Var(\underline{b}) = S_e^2 * adj(X'X) / |X'X| \qquad ...(4)$$

The students can conclude the above illustrations either |X′X| = 0, or have a small value. If multicolinearity is happened for the above reasons, it may causes or attains to a wrong conclusions that some of Explanatory(Xi's) are not important due to the (t) test for each of these parameters, so the fitted model is not capable to arise and explain the separate effect for each explanatory with respect to the highly correlation among them. Multicolinearity refers to a situation in which or more predictor variables in a multiple regression Model are highly correlated if Multicolinearity 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 cannot be estimated with great accuracy (Gujarati, 1995).We can define multicolinearity Through the concept of orthogonality , when the predictors are orthogonal or uncorrelated, all eigen values of the design matrix are equal to one and the design matrix is full rank. if at least one eigen value 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 cannot produce a clear estimate of the rate (degree) of multicolinearity.

ii)Eigen structure of $X'X$, let λ1, λ2, ….., λp 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 multicolinearity Vinod and Ullah,(1981), suggested that the condition number is given by :

$$\phi_1 = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} \qquad ...(5)$$

Montgomery and Peck, (1992) defined the condition number as the ratio of λmax and λmin as :

$$\phi_2 = \frac{\lambda_{\max}}{\lambda_{\min}} \qquad ...(6)$$

Where λmax is the largest eigenvalue.

λmin is the smallest eigenvalue

if λmin = 0, Then φ is infinite, which means that perfect multicollinearity between predictor variables. If λmax and λmin are equal, Then φ = 1 and the predictors are said to be orthogonal. Pagel and Lunneborg, (1985) suggested that the condition number is :

$$\phi_3 = \sum_{i=1}^{p} \frac{1}{\lambda_i} \qquad ...(7)$$

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 :

$$VIF = \frac{1}{1 - R_j^2} \qquad ...(8)$$

Where is the coefficient of determination in the regression of explanatory Variable Xj on the remaining explanatory variables of the model? Generally, when VIF > 10, we assume there exists highly multicolinearity (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 multicolinearity 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 multicolinearity. 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.

2.5 THE ORDINARY RIDGE REGRESSION (ORR).

Consider the standard model for multiple linear regressions:

$$Y = X\beta + E \qquad ...(9)$$

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 σ2I.Note that, Both X's and Y have been standardized.

The OLS estimate β^ of β is obtained by minimizing the residual sum of squares, and is given by: and

$$
\begin{aligned}
&(Y - X\hat{\beta})'(Y - X\hat{\beta}),\\
&\hat{\beta} = (X'X)^{-1}X'Y\\
&\mathrm{Var}(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1}\\
&\mathrm{MSE}(\hat{\beta}) = \hat{\sigma}^2 \ \mathrm{trace}(X'X)^{-1}\\
&\qquad = \hat{\sigma}^2 \sum_{i=1}^{p} \frac{1}{\lambda_i} \qquad \ldots(10)
\end{aligned}
$$

Where $\hat{\sigma}^2$ is the mean squares error. This $\hat{\beta}$ estimator 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 by using method of Hoerl and Kennard (2000). Therefore, the ridge solution is given by :

$$\hat{\beta}(K) = (X'X+KI)^{-1}X'Y \quad ,K \geq 0 \quad \ldots(11)$$

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).

2.6 Properties of Ordinary Ridge Regression Estimator

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

- From equation (11), by taking expectation on both sides, then

$$
\begin{aligned}
&E(\hat{\beta}(K)) = A_k \beta\\
&\textit{Where} \quad A_k = \left[I + K(X'X)^{-1} \right]^{-1}\\
&\textit{and} \quad \mathrm{var}(\hat{\beta}(K)) = \hat{\sigma}^2 A_k (X'X)^{-1} A_k'
\end{aligned}
$$

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. -Because $\hat{\beta}(K) = \left[I + K(X'X)^{-1} \right]^{-1} \hat{\beta}$ 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 :

$$
\begin{aligned}
MSE(\hat{\beta}(K)) &= E\left[\left(\hat{\beta}(K) - \beta\right)' \left(\hat{\beta}(K) - \beta\right) \right]\\
&= \hat{\sigma}^2 trace\left[A_k (X'X)^{-1} A_k' \right] + \hat{\beta}'(I - A_k)'(I - A_k)\hat{\beta}\\
&= \hat{\sigma}^2 \sum_{i=1}^{p} \frac{\lambda_i}{\lambda_i + k} + K^2 \hat{\beta}'(X'X + KI)^{-2} \hat{\beta} \qquad \ldots(12)
\end{aligned}
$$

Note that, the first term of the right hand in equation (12) 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 $MSE\,\hat{\beta}(K) < MSE(\hat{\beta})$. (More details see Judge, 1988, Gujarat; 1995, Gruber 1998, Pasha and Shah 2004)

Testing of Multicolinearity:-

The most common test to detect multicolinearity problem is the test which named by (Farrar-Glauber)test, which depending on the statistic (χ2 – Chi Square) to test the following hypothesis:

1. Ho : (Xj) orthogonal V/S H1: (Xj) Not orthogonal

2. The test statistic is given by:

$$\chi^2 = -[n - 1 - (2k + 5)/6] * Ln|D| \quad \ldots(13)$$

such that:

n: No. of observations

k: No. of Explanatory (Xi's)

Ln|D| : the natural logarithm for the absolute value for determinant of correlation matrix among explanatory. But the correlation matrix is given by:

$$D = \begin{bmatrix} 1 & r_{12} & r_{13}\ldots\ldots & r_{1k} \\ r_{21} & 1 & r_{23}\ldots\ldots & r_{2k} \\ . & . & . & . \\ r_{1k} & r_{2k} & r_{3k}\ldots\ldots & 1 \end{bmatrix}$$

3. Comparing Chi-Square computed from (2), with the theoretical Chi-Square fro statistical table for a specific significant level (%5, or %1), and [k(k-1)/2] degrees of freedom (d.f).

4. It's decided to reject Ho if χ2cal.> χ2table. Which means that Multicolinearity exists among explanatory r.v's. And there is no reason to reject(accept) Ho if χ2cal.< χ2table. (Repeat the test if χ2cal.= χ2table).

Note: It is important to explain what does (Orthogonal) means in a mathematical explanation:

The orthogonal vectors:

let us we introduce the expression:

$$\cos(\theta) = \frac{x'y}{(x'x)^{1/2}(y'y)^{1/2}} \qquad \ldots(14)$$

θ : is the angle between two vectors X, and Y

These two vectors are orthogonal if the cos(θ) is nearly equal one. But if exactly one these two vectors are said to be orthonormal or there length normalized to unity.

After detecting multicolinearity due to the test above, the another step is to determined which one of explanatory caused multicolinearity.This reorganization can be done by making another test named by fisher test or commonly (F Test). This test is made as follow:

1. Let X1,X2,X3,X4 is a set of explanatory in (GLM). Taking X1 First.

2. Calculating the correlation of determination R2 or multiple correlations for X1 as a dependent variable and the remainder (X2X3X4) are independents, using the following formula:

$$R^2_{1.(234)} = 1-(1-r^2_{12})(1-r^2_{13.2})(1-r^2_{14.23}) \qquad \ldots(15)$$

such that each of (r13.2 , r14.23) are represents the partial correlation coefficients.

3. Let Xj is the explanatory one which causes multicolinearity then to test the hypothesis: Ho: X1 causes Multicolinearity or $R^2_{j.(234\ldots k)} = 0$ V/S H1: X1 does not causes Multicolinearity $R^2_{j.(234\ldots k)} \neq 0$

4. The test statistic is given by:

Let Xj is the explanatory one which causes multicolinearity then: $F_{(j)cal.} = \frac{R^2_{j.234}/(k-1)}{(1-R_{j.234})/(n-k)}$ ....(16)

5. The rejection region:

Comparing Fcal if > Ftable with (k-1),(n-k) d.f and (α: level of significant), then the decision would be reject Ho or X1 indeed causes multicolinearity.

Note :In order to compute the partial correlation coefficients the following mathematical formula can be used.2nd order partial correlation coefficient:

$$\rho_{ij.hc} = \frac{\rho_{ij.c} - \rho_{ih.c}\rho_{jh.c}}{\sqrt{(1-\rho^2_{ih.c})(1-\rho^2_{jh.c})}} \qquad \ldots(17)$$

First order partial correlation coefficient:

$$\rho_{ij.h} = \frac{\rho_{ij} - \rho_{ih}\rho_{jh}}{\sqrt{(1-\rho^2_{ih})(1-\rho^2_{jh})}}$$

such that ρ_{ij} are simple correlation coefficients from matrix (D). After determining the explanatory r.v let (Xj) which causes multicolinearity, the next step is to determine (that Xj with which one of remainders causes this multico) this step could be done for all remainders one after another. Then the following statistical hypothesis which concerned with the second order correlation coefficients must be applied;

Ho : rij.(1,2,3….k)= 0 V/S H1 : rij.(1,2,3….k) $\neq$ 0

The test statistic is given by :

$$t_{cal} = \frac{r_{ij.(12\ldots k)}\sqrt{n-k}}{\sqrt{1-r^2_{ij,(12\ldots k)}}} \qquad \ldots(18)$$

And compare this value with ttable for (n-k) d.f and (α) level , we reject Ho if tcal > ttable , then at last and after all these tests are made for all explanatory r,v's then all those causes multicolinearity are determined.

Applicative Side:

In this research ,we simulate a three set of data using sas package of (larg ,medium and small) size which are (100,50,25)respectively ,where the correlation coefficients between the predictor variable (xs) are large (the number of

ISSN 2229-5518

predictor variable in this study are six variable).table(1,2,3) shows the correlation matrix based on set of simulate data.

VIF	N=50	N=100
$VIF(X_1)$	12.32	17.92
$VIF(X_2)$	12.95	16.39
$VIF(X_3)$	14.16	14.61
$VIF(X_4)$	18.93	12.85
$VIF(X_5)$	12.88	17.98
$VIF(X_6)$	13.55	11.99

Table(1) correlation matrix of size(100)

Table (2) correlation matrix of size(50)

	X1	X2	X3	X4	X5	X6
X1	1	0.943	0.956	0.962	0.970	0.890
X2	0.943	1	0.934	0.925	0.981	0.912
X3	0.956	0.934	1	0.912	0.943	0.913
X4	0.962	0.925	0.912	1	0.885	0.927
X5	0.970	0.981	0.943	0.885	1	0.941
X6	0.890	0.912	0.913	0.927	0.941	1

Table (3) correlation matrix of size(25)

	X1	X2	X3	X4	X5	X6
X1	1	0.890	0.856	0.943	0.974	0.904
X2	0.890	1	0.912	0.873	0.981	0.894
X3	0.856	0.912	1	0.914	0.942	0.956
X4	0.943	0.873	0.914	1	0.887	0.900
X5	0.974	0.981	0.942	0.887	1	0.962
X6	0.904	0.894	0.956	0.900	0.962	1

And then we can find the eigenvalues of the predictor

Eigenvalues	N=25	N=50	N=100
λ_1	5.6132	0.0009	0.0271
λ_2	-0.0388	0.0204	0.0417
λ_3	0.0260	0.0469	0.0752
λ_4	0.1655	0.1255	0.1191
λ_5	0.1170 + 0.0300i	0.1363	0.1289
λ_6	0.1170 - 0.0300i	5.6700	5.6621

correlation matrix are as follow:

We find and show that the sample size affects the results ,while the two eigenvalues of sample size (n=25) are complex number (which may not be all distinct) has two additional non-real roots (λ_5, λ_6) then ignore this sample. whenever the sample size increases, the case results of the methods of Estimation stables more.

The variance inflation factor (VIF) for all variable xs are as follow:

	X1	X2	X3	X4	X5	X6
X1	1	0.923	0.941	0.973	0.964	0.975
X2	0.923	1	0.934	0.962	0.893	0.902
X3	0.941	0.934	1	0.945	0.938	0.889
X4	0.973	0.962	0.945	1	0.904	0.952
X5	0.964	0.893	0.938	0.904	1	0.913
X6	0.975	0.902	0.889	0.952	0.913	1

The

N=100			
OLS		0RR	
cofe	stdev	cofe	stdev
1.890	0.299	0.375	0.071
-0.621	0.215	-0.425	0.172
-0.223	0.212	-0.108	0.192
0.710	0.171	0.553	0.150
-0.694	0.379	1.276	0.089
0.294	0.311	0.675	0.198

condition number ($\phi_1, \phi_2,$ and ϕ_3) as follow:

C.numbers	N=50	N=100
ϕ_1	79.37	14
ϕ_2	630	208
ϕ_3	6	6

From the previous indicators, it is obvious that there are a serious multicollinearity problem because there is one of eigenvalues (λ_1) for n=50 close to zero, all VIFs values more than 10, and the different types of condition number more than 5.

The regression coefficients and standard deviations of these coefficients can be summarized in table (4 &5).by using both OLS methods of RR to analyze the simulated Data ,we get the following results.

Table (4)Regression coefficients and Standard deviations

N=50			
OLS		0RR	
cofe	stdev	cofe	stdev
1.201	0.250	1.892	0.230
-0.503	0.323	-0.433	0.189
-0.214	0.201	-0.031	0.185
0.643	0.304	0.180	0.202
-0.517	0.324	-0.506	0.291
0.218	0.381	0.205	0.294

Table (5) Regression coefficients and Standard deviations

And in the end(later) we can find the mean square error and coefficient of determination(R2) for (OLS) and (ORR) Methods respectively .we obtain the following result.this result can be in summarized in table (6).

Table (6) MSE and R-Square for each sample

	MSE		R-Square	
sample	OLS	ORR	OLS	ORR
n=100	0.413	0.372	0.793	0.965
n=50	0.401	0.324	0.701	0.932

From the previous results, it is obvious that:
-all values of ORR have smaller standard deviation than OLS .
-all values of ORR have smaller MSE of regression coefficient than OLS .
While, model of RR have larger (R2) than OLS).consequently, ORR method better than OLS when the multicollinearity problem is exist in dada.

Conclusions:

In this research ,we referred to the multicollinearity problem, method of detecting of this problem and effect on result of multiple regression model.also we introduced model of ridge regression to solve this problem and we make a comparison between ORR and OLS methods .based on standard deviation, mean square error and coefficient of determinations .we note that all values of ORR have smaller standard deviation than OLS when the multicollinearity problem exist and ORR it has smaller MSE of estimators ,smaller standard deviation for most estimators and has larger coefficient of determination. We conclude that the sample size affects the results of estimated value, whenever the sample size increases, the case results of the methods of Estimation stables more.

REFERENCES:

1. W.H. Green, Econometric A analysis, Mac- Millan, New York,1993.
2.H.Wethrill, Evaluation of ordinary Ridge Regression, Bulletin of Mathematical Statiatics,18(1986),1-35.
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. G.G.Jude, Introduction to theory and practice of Econometric, John Willy and son, New York, 1988.
6. D.C. Montgomery, and E.A. Peck, Introduction to linear regression analysis, John willy and Sons, New York, 1992.
7. R.H.Myers, Classical and Modern Regression with application, PWS-KENT publishing Co., Boston, 1990.
8. M.D. Pagel, and C.E. Lunneborg, Empirical Evaluation of Ridge Regression, Psychological Bulletin,97(1985),342-355.
9. G.R. Pasha, and M.A. Shah, application of Ridge Regression to Multicollinear Data, J.res.sci, 15(2004),97-106.
10. H.D .Vinod, and A. Ullah, Rrecent Advances in Regression Models ,Marcel Dekker, New York,1981.
11. E. Walker, Detection of Collinearity –Influential observations, Communications in statistics, theory and methods,18(1989),1675-1690.
12.M.Sana,and.C.Eyup,Efficient choice of Biasing Constant for Ridge Regression,Int.J. Contemp .Math.Science,3(2008),527-536.

Dr.Samira Muhammad salh

Department of statistics, University of sulamani

Samira196902@yahoo.com