Ridge Regression Estimators with the Problem of Multicollinearity

Maie M. Kamel
Statistic Department, Faculty of Commerce
Tanta Univeristy, Tanta, Egypt
maie.m.kamel@gmail.com

Sarah F. Aboud
Egyptian Academy of Computers, Information & Management Technology
Ministry of High Education, Tanta, Egypt
Sara_aboud@yahoo.com

Copyright © 2013 Maie M. Kamel and Sarah F. Aboud. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

The study aims to illustrate the negative effect of the Multicollinearity problem upon the specimen, identify the way of Ridge Regression as a way to deal with the Multicollinearity problem, focus on some of the estimators of Ridge regression as (James and Stein, Bhattacharya, Heuristic) and identify which estimator from the previously mentioned estimators is highly preferable to be used, to estimate the parameters of a model which faces the Multicollinearity problem. Minimum mean-square error (MSE) has been used as the best measure for estimator. Application has been done on specific data for return on total assets of a bank after making sure that this data faces the Multicollinearity problem. Also, simulation method was used to generate fabricated data sets, which gave more space in the application. According to the study we can see that James and Stein’s estimator has got the minimum mean square error (MSE). Consequently the study recommends its usage to estimate model parameters which face the Multicollinearity problem.
Keywords: Ridge Regression – Multicollinearity – Heuristic – Bhattacharya – James and Stein – Simulation

1-Introduction

Several years ago some researchers in Econometric field used ordinary least square (OLS) method to estimate the parameters of linear regression model. The Ordinary least square method is based on a number of assumptions; one of these assumptions says that there is no linear relationship between the explanatory variables, in case of dropping this hypothesis the multicollinearity problem will appear. The multicollinearity problem threatens both the assumption and usage of ordinary least square (OLS). In case the model is not full ranked, it won’t be possible to find the inverse matrix of explanatory variables; consequently, we will get infinite solutions according to (OLS) method. On the other hand if the relationship degree between the explanatory variables is much enough and we use the (OLS) method we will find that:

- The estimations of model parameters have different results in addition to that the standard errors for these estimations increase when the relationship between explanatory variables increases.

- More standard errors of estimators mean a breadth of confidence interval in the parameters of population.

- Increasing risk of type two errors (accepting wrong assumptions) because of the width of confidence interval due to an increase in the standard error of the estimation.

- If the linear relationship is high, we get higher $R^2$ however; most parameters of the relationship are non-significant statistics by using t test.

The existence or nonexistence of the linear relationship between the variables isn’t the point but the strength degree is. We previously mentioned that there is a linear relationship between the explanatory variables (mathematical variables) not a random relationship, so this phenomenon concerns with the sample not the society for whom the sample was selected. So we don’t test the existence of these relationships but we measure the strength degree in any sample. Due to the negative impact caused by Multicollinearity problem, there are many ways to deal with this problem and one of these methods is Ridge regression (RR).
2-Ridge Regression

Ridge Regression represents one of the methods which deal with multicollinearity problem and ridge regression estimators as Horel and Kennard represented it in 1970. It is presented in low values from the distinctive values (Eigen Values) of the matrix (X’ X) in the model (Y = Xb + u). These values are added to the elements of the main diameter in the matrix (X’ X) in the estimators of least squares to gain the ridge regression estimators which represent biased estimators depending on adding low positive values which called low added values (k) with (ridge parameters) or (biasing factor). It could be clarified by the following linear regression equations:

\[ Y = Xb + e \]  

If the model parameters were estimated according to the (OLS) method, we will find that \( b^0 \) will be estimated as follows:

\[ b^0 = (X'X)^{-1}X'y \]  

But if the model parameters are estimated according to (RR) method, we will find that the ordinary ridge regression estimator \( b^{\text{ORR}} \) will be estimated according to the following formula:

\[ B^{\text{ORR}} = (X'X + kl)^{-1}X'y \]  

There are several ridge regression estimators due to the lack of consistent law to identify the value (k), whereas \( k > 0 \) denotes to biasing factor. Hence the idea of this research is represented to illustrate and show some of ridge regression estimators as James and Stein’s parameter, Bhattacharya’s parameter and Heuristic’s parameter. In trying to understand which one of these parameters is able to help us in determining the value of biasing factor which is less than MSE.

3-Ridge Regression Estimators

General Ridge Regression estimator (\( b^{\text{GRR}} \))

\[ b^{\text{GRR}} = (XX + GKG)^{-1} X' Y \]  

where:

(G) Represents the distinctive columns in the matrix (X’ X).

(K) Represents orthogonal design matrix.

Suppose that all \( K_i \) values are equal in formula no. (4) and has \( K \) value, it means that \( (k_i = k) \). So we find that general regression rate estimator (GRRE) denotes to ordinary regression rate estimator(ORRE) which is estimated by the following formula:

\[ b^{\text{ORR}} = (X'X + kl)^{-1}X'y \]
James and Stein estimator ($b^S$)

In the following regression model

$$Y = H b_{(orth)} + e$$  \hspace{1cm} (6)

This model denotes to an orthogonal model Where (Y) symbolizes vector of dependant variables, and (H) symbolizes the matrix of explanatory variables for an orthogonal model. also $b_{(orth)}$ symbolizes vector regression coefficients for an orthogonal model. and (e) denotes to random error vector.

In 1961 James Stein represented ($b^S_{(orth)}$) estimator [(vector regression coefficients) in orthogonal model ] which is estimated according to the following formula:

$$b^S_{(orth)} = Sb^o_{(orth)}$$ \hspace{1cm} (7)

$b^o_{(orth)}$ Denotes to (OLS) estimator for vector regression coefficients for the illustrated model in the formula no. (6).

(S) had been illustrated before by James & Stein in 1961 as the following:

$$S = \frac{\frac{2}{n-p+2} v}{\|b^o_{(orth)}\|^2}$$ \hspace{1cm} (8)

- Where: (P) denotes to the number of regression coefficients in the model.
  - (n) denotes to views sample (sampler’s size).
  - (v) denotes to the total sum squared error which is illustrated in formula no.(6) where \[ V=\sum e^\prime e \].

**The relationship between $b^o_{(orth)}$ estimator and $b^o$ estimator.**

According to (OLS) estimator which is illustrated in regression model and formula no.(6) we find that :

$$b^o_{(orth)}=(H' H)^{-1} H' Y$$ \hspace{1cm} (9)

where: $H' H=1$

$\therefore b^o_{(orth)} = H' Y$

$\therefore X = H \Lambda_G^\frac{1}{2} G'$

$\therefore b^o_{(orth)} = X' \Lambda_G^\frac{1}{2} G^{-1} Y$

by multiplying the two Sides by $(X'X)$

$\therefore b^o_{(orth)} = (X'X)^{-1} (X'X)^{-1} X Y$

$\therefore b^o_{(orth)} = G \Lambda_G^\frac{1}{2} G^{-1} b^o$

by substituting in the formula

$(X'X) = G \Lambda_G^\frac{1}{2} G'$

$\therefore b^o_{(orth)} = G \Lambda_G^\frac{1}{2} G^{-1} b^o$

$\therefore b^o_{(orth)} = \Lambda_G^\frac{1}{2} G' b^o$ \hspace{1cm} (10)
In 1961 James & Stein presented \( \hat{b}^s \) estimator in case that vector regression coefficient has more than two compound \( (p \geq 3) \) it can be illustrated in this formula as follows:

\[
\hat{b}^s = S \hat{b}^o
\]  

(11)

where: \( \hat{b}^o \) denotes to (OLS) as it was previously explained in formula (2)

**Bhattacharya Estimator** \( (\hat{b}^{BH}) \)

Bhattacharya estimator can be obtained by the following steps:

**First:** Bhattacharya defined \( (\hat{b}^{BH}(\text{orth})) \) estimator which denotes to an evaluation for regression coefficients in an orthogonal regression model as follows:

\[
\hat{b}^{BH}(\text{orth}) = \Delta \hat{b}^o(\text{orth})
\]  

(12)

Where \( \Delta \) is a diametrical matrix, the main diameter elements is represented in \( \Delta_i \).

**Second:** The \( \Delta_i \) elements can be obtained by (reverse of \( \Delta \))as follows:

\[
\Delta_i = \Delta^p_{i+1}
\]  

(13)

**Note:** the symbol (-) above the variable denotes to reverse relationship for this variable.

**Third:** \( \Delta_i \) values was illustrated with the following formula:

\[
\Delta_i = \frac{\sum_{j=1}^{p} \alpha_j f_j(\hat{b}^o(\text{orth}), v)}{\sum_{j=1}^{p} \alpha_j}
\]  

(14)

**Fourth:** Bhattacharya considered the non-negative values \( \hat{\alpha}_i \) as follows:

\[
\hat{\alpha}_i = \lambda_{p-i+1}^{-1} - \lambda_{p-i}^{-1} \quad \text{for } i=1,2,3,\ldots, p
\]  

(15)

**Fifth:** Bhattacharya defined \( (f_i(\hat{b}^o(\text{orth}), v)) \)

\[
f_i(\hat{b}^o(\text{orth}), v) = \begin{cases} 
1 & \text{if } i=1,2 \\
1 - \frac{i-2}{n-p+2} \cdot \frac{v}{\|\hat{b}^o(\text{orth})\|^2} & \text{if } i=3,4,\ldots, p
\end{cases}
\]  

(16)

Where

\[
\|\hat{b}^o(\text{orth})\|^2 = \sum_{j=1}^{j} (\hat{b}^o(\text{orth})_j)^2
\]  

(17)

**Sixth:** By reference to formula no.(10) which connects between \( \hat{b} \) and \( \hat{b}^o(\text{orth}) \).

\[
\therefore \hat{b}^{BH} = G \Lambda^{-\frac{1}{2}} b^{BH}(\text{orth})
\]  

(18)

By substituting in \( b^{BH}(\text{orth}) \) from formula no.(14)

\[
\therefore \hat{b}^{BH} = G \Lambda^{-\frac{1}{2}} \hat{b}^o(\text{orth})
\]  

(19)

By substituting in the \( b^o(\text{orth}) \) according to formula nom. (10)
\[ \therefore b^{BH} = G \Delta^{\frac{1}{2}} G' b^O \] \quad (20)
\[ \therefore b^{BH} = G \Delta G' b^O \] \quad (21)

**Heuristic Estimator** \( b^H \)

It represents an extension to Bhattacharya estimator, it’s also known that this estimator matches between the results which each one of (GRR) and (Bhattacharya) had reached to. It can also be possible evaluate Heuristic \( b^H \) according to the following formula:

\[ b^H = G \delta G' b^O \] \quad (22)

\[ \delta = \begin{bmatrix}
\delta_1 & 0 & 0 & 0 & 0 \\
0 & \delta_2 & 0 & 0 & 0 \\
0 & 0 & . & 0 & . \\
0 & 0 & 0 & . & . \\
0 & 0 & . & . & \delta_p
\end{bmatrix} \] \quad (23)

The elements of the main diameter represented in \( \delta_i \) in the diametrical matrix .

\[ \delta_i = \frac{\sum_{i=1}^n \alpha_i g_i \left( \tilde{O}^{\text{Orth}}, v \right)}{\sum_{i=1}^n \alpha_i} \] \quad (24)

In the previous formula \( g_i \left( \tilde{O}^{\text{Orth}}, v \right) \) denotes to reorder process, \( f_i \left( \tilde{O}^{\text{Orth}}, v \right) \) in order to be non-increasing .

**4-Using Mean Square of Error to Compare Among study Estimators**

Comparing between Ordinary Least Square and Ridge Regression Estimators is being made to evaluate model parameters in the presence of multicollinearity problem according to mean square of error criteria.

The function error of **ordinary least square**(OLS) can be expressed as follows:

\[ \therefore \text{MSE}(b^O) = E \left( b^O_{(\text{Orth})} - b_{(\text{Orth})} \right)' \Delta^{-1} \left( b^O_{(\text{Orth})} - b_{(\text{Orth})} \right) \] \quad (25)

The function error for **James and Stein**’s estimator for \( b \) can be expressed as follows:

\[ \text{MSE}(b^S) = E \left( b^S_{(\text{Orth})} - b_{(\text{Orth})} \right)' \Delta^{-1} \left( b^S_{(\text{Orth})} - b_{(\text{Orth})} \right) \] \quad (26)

The function error for **Bhattacharya**’s estimator can be expressed as follows:

\[ \text{MSE}(b^{BH}) = E \left( b^{BH}_{(\text{Orth})} - b_{(\text{Orth})} \right)' \Delta^{-1} \left( b^{BH}_{(\text{Orth})} - b_{(\text{Orth})} \right) \] \quad (27)

The function error for **Heuristic**’s estimator can be expressed as follows:

\[ \text{MSE}(b^H) = E \left( b^H_{(\text{Orth})} - b_{(\text{Orth})} \right)' \Delta^{-1} \left( b^H_{(\text{Orth})} - b_{(\text{Orth})} \right) \] \quad (28)
5-Simulation Technique

The researchers have used HKB (Hoerl, Kennard, and Baldwin) style for simulation to generate different collocations from regression coefficients.

This kind is distinctively known with its easy style in application in addition to that it gives a certain range to choose regression coefficients by using Signal to Noise Ratio, and it’s easy to illustrate the steps of (HKB) applications as follows:

1-Picked up randomly a certain number (say 10) in a certain range [1 -2500] to represent the SNR (Signal to Noise Ratio) which also represents the total squared bias divided on the contrast, and it can be expressed as follows:

\[ SNR = \frac{b_i^{(2)} - b_i^{(1)}}{\sigma^2} \]  

where:

* \(b_i^{(2)}\) represents the numbers.
* \(b_i^{(1)}\) represents the number of explanatory variables in the model and they are (7) and takes the symbol (w).
* \(\sigma^2\) is the high, it was also found that the high \(\sigma\) is gives a great chance to the other estimators to overcome the ordinary least square.

SNR values were represented in (1,4,9,25,48,81,200,400,800,2500)

2- The values of matrix of explanatory variables (X) are inserted as follows:

\[
X = \begin{bmatrix}
1 & 290 & 413 & 69 & 543 & 329 & 79 & 420 \\
1 & 349 & 524 & 69 & 524 & 338 & 80 & 555 \\
1 & 158 & 1031 & 113 & 520 & 348 & 81 & 572 \\
1 & 217 & 1052 & 101 & 429 & 317 & 85 & 510 \\
1 & 75 & 1342 & 110 & 393 & 338 & 84 & 578 \\
1 & 69 & 1389 & 121 & 474 & 386 & 100 & 615 \\
1 & 52 & 1484 & 106 & 485 & 398 & 95 & 749 \\
1 & 121 & 1523 & 115 & 406 & 431 & 111 & 804 \\
1 & 358 & 1093 & 119 & 667 & 487 & 145 & 745 \\
1 & 320 & 1160 & 138 & 643 & 531 & 169 & 937 \\
1 & 220 & 981 & 258 & 695 & 597 & 228 & 974 \\
1 & 469 & 385 & 312 & 894 & 669 & 241 & 1133 \\
1 & 167 & 802 & 402 & 738 & 657 & 241 & 1240
\end{bmatrix}
\]

3- From each value from Signal-to-noise ratio (SNR) we pick some of random numbers (r) from the standard normal distribution equal m.

where:

* \((r)\) represents the numbers.
* \((m)\) represents the number of explanatory variables in the model and they are (7) and takes the symbol (w).
* \(r = 500\) in this study.

Finally the number of samples which have been generated are 5000 sample (10 x 500), as (10) represents the number of Signal-to-noise ratio values, which have been identified in step No. (1).
4-To calculate the regression coefficients \( b_{(orth)} \) for each generated random sample (where \( b_{(orth)} \) represents regression coefficient in the orthogonal regression model.)

\[
b_{(orth)}_i = \frac{SNR \cdot w_i}{\sqrt{\sum w_i^2}}
\]  \hspace{1cm} (29)

5- A 500 different groups will be generated from the random variable error in the normal distribution with expected value equal to zero and one standard deviation and its symbol is (e) to represent the error element in the orthogonal regression model.

6- Values of the dependant variable will be calculated (Y) for each sample among 500 samples.

6-The study result

Table (1) clarifies the results which the study has reached to:

1) The first column represents the one – tenth value in SNR.

2) The second column represents the outcome of using (OLS) method to evaluate data regression coefficients which have been generated by using simulation technique.

3) The third column shows the results of using (James and Stein) regression coefficients evaluation for data which has been generated by using simulation method.

4) The forth column shows the results of using Bhattacharya estimator.

5) The fifth column shows the results of using Heuristic estimator.
Table (1)
(MSE for each Estimator)

<table>
<thead>
<tr>
<th>SNR</th>
<th>OLS</th>
<th>Stein</th>
<th>BH</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.6034023</td>
<td>.85147788</td>
<td>1.6033469</td>
<td>1.5334987</td>
</tr>
<tr>
<td>4</td>
<td>1.4749686</td>
<td>1.2627039</td>
<td>1.4749509</td>
<td>1.4749505</td>
</tr>
<tr>
<td>9</td>
<td>1.6280015</td>
<td>.8296062</td>
<td>1.6279435</td>
<td>1.665581</td>
</tr>
<tr>
<td>25</td>
<td>1.6655513</td>
<td>1.6743449</td>
<td>1.6655746</td>
<td>1.665581</td>
</tr>
<tr>
<td>49</td>
<td>1.4380137</td>
<td>1.4318862</td>
<td>1.4380535</td>
<td>1.4380169</td>
</tr>
<tr>
<td>81</td>
<td>1.4365162</td>
<td>1.4395838</td>
<td>1.4366661</td>
<td>1.4365238</td>
</tr>
<tr>
<td>200</td>
<td>2.1090713</td>
<td>2.1103591</td>
<td>2.114007</td>
<td>2.1092927</td>
</tr>
<tr>
<td>400</td>
<td>1.4971057</td>
<td>.6719066</td>
<td>1.4970512</td>
<td>1.4970521</td>
</tr>
<tr>
<td>900</td>
<td>1.8583551</td>
<td>1.8589462</td>
<td>1.8691939</td>
<td>1.8611626</td>
</tr>
<tr>
<td>2500</td>
<td>1.5895881</td>
<td>1.5895775</td>
<td>2.0674158</td>
<td>1.9292857</td>
</tr>
</tbody>
</table>

- The sign √ denotes to mean square estimator in comparison to all study estimators.
- The sign * denotes to lowest mean square error in ridge regression estimators.

The study results can be clarified by the following Chart.
- The horizontal line in chart (1) represents the study estimator [(OLS), Stein, Heuristic, Bhattacharya] respectively.
- The vertical line represents mean square error & the ten illustrated lines in the chart represent SNR estimators.
- In comparison to the rest of ridge regression estimators; mean square error was found to be the best estimator which got the least estimation by the applied study, in (SNR=1, 4, 9, 49, 400, 2500).
- In comparison to the rest of ridge regression estimators, Heuristic estimator has got the least mean square error by the applied study, in (SNR=25, 81, 200).
- In comparison to the rest of ridge regression estimators, (OLS) estimator has got the least mean square error by the applied study, in (SNR=25, 81, 200, 900).

Conclusion

According to the data study and the results which were found, James and Stein was the best estimator among the study estimators as it got the least mean square error. It’s recommended to be used in estimating regression coefficients for a model which face linear Multicollinearity problem. It is worth noting that Heuristic estimator was the second best...
estimator which got the least mean square error in regression coefficients, which requires greater attention while preparing studies based on Heuristic estimator.

References


Received: February 21, 2013