Restricted Ridge Regression estimator as a parameter estimation in multiple linear regression model for multicollinearity case

F A O Rumere, S M Soemartojo and Y Widyaningsih

Department of Mathematics, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia

Corresponding author’s email: yekti@sci.ui.ac.id

Abstract. Multiple Linear Regression analysis is one of the techniques in statistics that is used to analyse the relationship between a dependent variable and two or more independent (regressor) variables. Ordinary Least Square method is commonly used to estimate the parameters. Most frequently occurring problem in multiple linear regression analysis is the presence of multicollinearity. Multicollinearity in the least square estimation produces estimation with a large variance, so another method is needed to overcome the multicollinearity. The method is called ridge regression. In this method, a constant bias ridge \( k \) is added to \( XX' \) matrix. A study had developed the method by using the prior information of the parameter \( \beta \) and introduced the Restricted Ridge Regression method. Prior information of the parameter \( \beta \) was defined as a non-sample information arising from past experiences and the opinions of an expert with similar situations and containing the same parameters \( \beta \). This study explains the use of Restricted Ridge Regression method in overcoming the multicollinearity in regression model. Based on an application on a data set, \( \beta \) of Restricted Ridge Regression has smallest mean square error (MSE) than \( \beta \) of Ordinary Least Square.

Keywords: Multicollinearity, regressor, restricted ridge regression.

1. Introduction

Consider the multiple linear regression models denoted by

\[
y = X\beta + \epsilon
\]

which \( y \) is \( n \times 1 \) observation (dependent variable) matrix, \( X \) is \( n \times p \) regressor variables matrix, \( \beta \) is \( p \times 1 \) parameter regression vector and \( \epsilon \) is \( nx1 \) random error vector with \( E(\epsilon) = 0 \) and \( Var(\epsilon) = \sigma^2 I \) [1]. Gauss-Markov explains that the least square estimator \( \hat{\beta}_{ls} = (X'X)^{-1}X'y \) is an unbiased estimator with minimum variance if the classical assumption of error is satisfied [2]. However, when there is a linear dependency between regressor variables, it can cause the least squares estimator become less effective. Multicollinearity is a condition in which regressor variables in multiple linear regression model is almost linearly dependent. This condition causes the variance of least squares estimator tends to be large and the estimator becomes unstable. This resulted in a less precise interpretation of the regression model [1].
Multiple linear regression model that experienced multicollinearity, resulted in the least squares estimation to be unstable in assessing the coefficient of regression parameters. In 1970, Hoerl and Kennard introduced a new method to solve the problem, the method is known as ridge regression. The basic principle of the ridge regression method is to add the bias constant $k$ to the main diagonal of $X'X$ matrix which is obtained by minimizing the sum of squares error function of multiple linear regression models with subject to $\hat{\beta}'\hat{\beta} < c$, $c$ is an optimal value [3].

Ridge regression estimator is denoted by

$$\hat{\beta}_R = (X'X + kI)^{-1}X'y$$  \hspace{1cm} (1)

where:

- $y$ is $n \times 1$ observation vector
- $X$ is $n \times p$ regressor variable matrix
- $\hat{\beta}_R$ is $p \times 1$ ridge estimator
- $k$ is constant bias ridge, $k > 0$

This method produces ridge estimator that is bias and depends on a constant $k$ (ridge bias constant), so it is necessary to find a constant $k$ that can optimize ridge estimates by producing variance and a small bias [3]. Grob and Sarkar developed a method that utilized prior information from the $\beta$ parameters and the ridge regression (RR) method [4, 5]. Restricted Ridge Regression (RRR) method is expected to overcome multicollinearity and produce small MSE values.

2. Experimental

2.1. Prior information of parameter $\beta$

Berger defines the prior information for the $\beta$ parameter which is a non-sample information arising from past experiences with similar situations and containing the same $\beta$ parameters [6]. Shalabh et al. said that prior information can be expressed in 3 forms of exact, stochastic, and inequality restriction [7]. This paper discusses more deeply about the exact linear restriction symbolized as the form of $R\beta = r$, where $R$ is a nonzero matrix, $m \times p$ with $\text{rank}(R) = m < p$ and $r$ is an $m \times 1$ vector.

2.2. Restricted ridge regression estimator

The Restricted Ridge Regression method estimates the $\beta$ parameter by minimizing the sum of squares error function of the multiple linear regression model with its constraints.

Minimize

$$\langle y - X\hat{\beta} \rangle'(y - X\hat{\beta})$$  \hspace{1cm} (2)

Subject to $R\hat{\beta} = r$ and $\hat{\beta}'\hat{\beta} < u$, where $u$ is a limitation constant, $R\beta = r$ is the prior information of the parameters $\beta$ and $\hat{\beta}'\hat{\beta} < u$ are the limits required for a regression model to reduce multicollinearity problems. The Lagrange method is used to find Restricted Ridge Regression estimator as below [8, 9].

$$\hat{\beta}_{RR} = \hat{\beta}_R - (X'X + kI)^{-1}X'r((R(X'X + kI)^{-1}R')^{-1}(R\hat{\beta}_R - r))$$

Let $S_k = X'X + kI$

$$\hat{\beta}_{RR} = \hat{\beta}_R - S_k^{-1}R'(RS_k^{-1}R')^{-1}(R\hat{\beta}_R - r), \hspace{1cm} k > 0$$

or $\hat{\beta}_{RR}$ can be expressed as

$$\hat{\beta}_{RR} = M_k'x - S_k^{-1}R'(RS_k^{-1}R')^{-1}r$$  \hspace{1cm} (3)

where $M_k = (S_k^{-1} - S_k^{-1}R'(RS_k^{-1}R')^{-1}RS_k^{-1})$ and $k > 0$
Bias of $\hat{\beta}_{RR}$

$$E(\hat{\beta}_{RR}) - \beta = -kM_k\beta$$

(4)

Covariance of $\hat{\beta}_{RR}$

$$\text{Cov}(\hat{\beta}_{RR}) = E((\hat{\beta}_{RR} - E(\hat{\beta}_{RR}))(\hat{\beta}_{RR} - E(\hat{\beta}_{RR}))')$$

$$= E(M_kX'\epsilon)(M_kX'\epsilon)'$$

$$= \sigma^2 M_kX'XM_k$$

(5)

Mean Square Error (MSE) of $\hat{\beta}_{RR}$

$$\text{MSE}(\hat{\beta}_{RR}) = \text{Var}(\hat{\beta}_{RR}) + (\text{bias}(\hat{\beta}_{RR}))^2$$

$$= \sigma^2 M_kX'XM_k + (-kM_k\beta)(-kM_k\beta)'$$

$$= \sigma^2 M_kX'XM_k + k^2 M_k\beta'\beta M_k$$

(6)

2.3. Choosing of bias ridge $k$

Hoerl and Kennard suggested a graphical method called the ridge trace method to select the value of $k$. Ridge trace is a plot of elements $\hat{\beta}_{RR}$ versus $k$ where $k \in (0, 1]$ [3]. Marquardt and Snee suggested to use 25 values of $k$ at the interval $(0, 1]$ [10]. The proper selection of $k$ at the interval $(0, 1)$ when multicollinearity occurs can be seen from the stability of regression coefficient and the MSE of the estimator [1].

3. Results and discussion

The data in this paper were taken from a paper entitled *A New Biased Estimator in Linear Regression and A Detailed Analysis of the Analytical-Analyzed Dataset on Portland Cement* by Kaciranlar et al. [11]. The standardized data, a dependent variable and 4 regressors are shown at table 1.

Table 1. The standardized data

| $Y$     | $X_1$          | $X_2$         | $X_3$          | $X_4$          |
|---------|----------------|---------------|----------------|----------------|
| -1.1249262 | -0.078461     | -1.4236884    | -0.9007209     | 1.7923096      |
| -1.4041124 | -1.0984538    | -1.2308973    | 0.5044037      | 1.3143603      |
| 0.5900749 | 0.6015342     | 0.504223      | -0.588471      | -0.5974365     |
| -0.5200227 | 0.6015342     | -1.1023698    | -0.588471      | 1.0156421      |
| 0.0317025 | -0.078461     | 0.2471681     | -0.9007209     | 0.179231       |
| 0.9157922 | 0.6015342     | 0.4399593     | -0.432346      | -0.4779492     |
| 0.4837182 | -0.7584562    | 1.4681787     | 0.8166536      | -1.4338477     |
| -1.5237636 | -1.0984538    | -1.1023698    | 1.5972783      | 0.8364111      |
| -0.1544217 | -0.928455      | 0.3756956     | 0.9727785      | -0.4779492     |
| 1.3611606 | 2.3015222     | -0.0741504    | -1.2129708     | -0.2389746     |
| -0.7726197 | -1.0984538    | -0.5239964    | 1.7534033      | 0.2389746      |
| 1.1883311 | 0.6015342     | 1.1468601     | -0.432346      | -1.0753857     |
| 0.92908673 | 0.43153542    | 1.27538753    | -0.5884710     | -1.0753857     |
Multicollinearity test in the data can be done in several ways by looking at the Variance Inflation Factor (VIF) value of each regressor variable as shown at table 2.

The Restricted Ridge Regression method has a working principle that is by using the prior information of $\beta$ parameter in the form of linear restriction equation $R\beta = r$ which obtained in previous research. In this study, the prior information is derived from the parameter value of $\beta$ in the ridge regression estimation, where, ridge estimators from equation 1 are

$$(\hat{\beta}_R)' = [0.5479 \quad 0.3927 \quad -0.0054 \quad -0.2912]$$

defined

$$R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Thus the linear restriction equation $r' = R\beta$ (from section 2) is as the following, where $r'$ is a transpose of $r$.

$$r' = [0.5479 \quad 0.3927 \quad -0.0054 \quad -0.2912]$$

Table 3 shows the Restricted Ridge Regression estimator based on equation 3 for some values of $k \in (0,1]$. MSE is mean square errors, where in this computation MSE is the mean of sum squares of the differences between observation values, $y_i$ and its predictions, $\hat{y}_i$.

**Table 2. Values of variance inflation factor.**

|      | $X_1$  | $X_2$  | $X_3$  | $X_4$  |
|------|--------|--------|--------|--------|
|      | 40.44369 | 241.76083 | 48.30423 | 277.33374 |

**Table 3. Restricted Ridge Regression estimator for $k \in (0,1]$.**

|      | $k = 0.1$ | $k = 0.2$ | $k = 0.3$ | $k = 0.4$ | $k = 0.5$ |
|------|-----------|-----------|-----------|-----------|-----------|
| $\hat{\beta}_1$ | 0.5479 | 0.5479 | 0.5479 | 0.5479 | 0.5479 |
| $\hat{\beta}_2$ | 0.3927 | 0.3927 | 0.3927 | 0.3927 | 0.3927 |
| $\hat{\beta}_3$ | -0.0054 | -0.0054 | -0.0054 | -0.0054 | -0.0054 |
| $\hat{\beta}_4$ | -0.2912 | -0.2912 | -0.2912 | -0.2912 | -0.2912 |
| MSE  | $1.3877e-87$ | $1.3877e-87$ | $1.3877e-87$ | $1.3877e-87$ | $1.3877e-87$ |

|      | $k = 0.6$ | $k = 0.7$ | $k = 0.8$ | $k = 0.9$ | $k = 1$ |
|------|-----------|-----------|-----------|-----------|----------|
| $\hat{\beta}_1$ | 0.5479 | 0.5479 | 0.5479 | 0.5479 | 0.5479 |
| $\hat{\beta}_2$ | 0.3927 | 0.3927 | 0.3927 | 0.3927 | 0.3927 |
| $\hat{\beta}_3$ | -0.0054 | -0.0054 | -0.0054 | -0.0054 | -0.0054 |
| $\hat{\beta}_4$ | -0.2912 | -0.2912 | -0.2912 | -0.2912 | -0.2912 |
| MSE  | $1.3877e-09$ | $1.3877e-09$ | $1.3877e-09$ | $1.3877e-09$ | $1.3877e-09$ |
Table 4. The values of $\hat{\beta}_{LS}$, $\hat{\beta}_{R}$, $\hat{\beta}_{RR}$

| Regressor | $\hat{\beta}_{LS}$ | $\hat{\beta}_{R}$ | $\hat{\beta}_{RR}$ |
|-----------|---------------------|-------------------|-------------------|
| $X_1$     | 0.6837              | 0.5479            | 0.5479            |
| $X_2$     | 0.7833              | 0.3927            | 0.3927            |
| $X_3$     | 0.1256              | −0.0054           | −0.0054           |
| $X_4$     | 0.0873              | −0.2912           | −0.2912           |
| MSE       | 1.281               | 0.037             | 1.387692e − 09    |

Table 4 shows the $\hat{\beta}_{LS}$, $\hat{\beta}_{R}$, $\hat{\beta}_{RR}$, parameters estimation of ordinary least square, ridge regression, and Restricted Ridge Regression, respectively. It shows that parameters of Restricted Ridge Regression have the smallest MSE. It means this estimation gives stable regression estimator.

4. Conclusion

Multiple linear regression model has unknown value regression coefficient parameter so it is necessary to estimate parameters using least square method. However, regression model often has problems with the presence of multicollinearity, characterized by a linear relationship between regressor variables and it cause the least square estimator ($\hat{\beta}_{LS}$) become unstable. This paper uses the application of the Restricted Ridge Regression method in the "Portland Cement datasets" and the results show that multicollinearity causes $\hat{\beta}_{LS}$ to be unstable to estimate the model, $\hat{\beta}_{R}$ produces stable estimates with MSE 0.037 and $\hat{\beta}_{RR}$ also produces stable estimates with MSE 1.387692e-09, where MSE ($\hat{\beta}_{RR}$) < MSE ($\hat{\beta}_{R}$).

Acknowledgment

This work was financially supported by Universitas Indonesia under research grant PITTA A 2019 with grant contract number NKB-0451/ UN2.R3.1/HKP.05.00/2019.

References

[1] Montgomery D C, Peck E A and Vining G G 2001 *Introduction to Linear Regression Analysis* (United States: John Wiley & Sons)
[2] Gujarati D N 2003 *Basic Econometrics* 4th edition (New York: Mc Graw-Hill)
[3] Hoerl A E and Kennard R W 1970 *Technometrics* 12 55-67
[4] Groβ J 2003 *Stat. Probabil. Lett.* 65 57-64
[5] Sarkar N 1992 *Commun. Stat-Theor. M.* 21 1987-2000
[6] Berger J 1980 *Ann. Statist.* 8 716-61
[7] Shalabh S, Garg G and Misra N 2009 J. Multivar. Anal. 100 1498-520
[8] Bertsekas D P 2014 *Constrained Optimization and Lagrange Multiplier Methods* (United States: Academic Press)
[9] El-Dereny M and Rashwan N I 2011 *Int. J. Contemp. Math. Science* 6 585-600
[10] Marquardt D W and Snee R D 1975 *Am. Stat.* 29 3-20
[11] Kaciranlar S, Sakallioglu S, Akdeniz F, Styan G P H and Werner H J 1999 *Sankhya* 61 443-59