Journal of Modern Applied Statistical Methods

Volume 23 | Issue 2 Article 4

The Performance of Ridge Regression, LASSO, and Elastic-Net in Controlling Multicollinearity: A Simulation and Application

Netti Herawati
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia

Ameliana Wijayanti
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia

Agus Sutrisno
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia

Nusyirwan
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia

Misgiyati
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia

Received: 31-10-2023; Revised: 02-12-2023; Accepted: 30-12-2023; Published: 09-01-2024.

Recommended Citation

Netti Herawati, Ameliana Wijayanti, Agus Sutrisno, Nusyirwan, Misgiyati (2024). The Performance of Ridge Regression, LASSO, and Elastic-Net in Controlling Multicollinearity: A Simulation and Application. Journal of Modern Applied Statistical Methods, 23(2), https://doi.org/10.56801/Jmasm.V23.i2.4

Journal of Modern Applied Statistical Methods
2024, Vol. 23, No. 2,
Doi: 10.56801/Jmasm.V23.i2.4

ISSN 1538 – 9472

The Performance of Ridge Regression, LASSO, and Elastic-Net in Controlling Multicollinearity: A Simulation and Application

Netti Herawati
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia

Ameliana Wijayanti
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia

Agus Sutrisno
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia

Nusyirwan
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia

Misgiyati
Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Lampung, Indonesia

This research aims to compare the performance of Ordinary Least Square (OLS), Least Absolute Shrinkage and Selection Operator (LASSO), Ridge Regression (RR) and Elastic-Net in controlling multicollinearity problems between independent variables in multiple regression analysis using simulation data and case data. Data simulation uses a multiple regression model with p = 6 with a high level of multicollinearity (ρ = 0.99) at several sample sizes (n = 25, 50, 75). The best method is measured based on the smallest Average Mean Square Error (AMSE) and AIC values. The research results show that Elastic-Net is the best method for simulated data compared to LASSO and Ridge because it has the smallest AMSE and AIC values for each sample size studied. Similar things were also obtained when applying these three methods to data on stunting toddler cases in Indonesia which had high multicollinearity. By using the best method, namely the Elastic Net method, real data shows that cases of stunted toddlers in Indonesia are influenced by the percentage of toddlers who are malnourished ($\boldsymbol{x_1}$), the percentage of toddlers who receive exclusive breast milk ($\boldsymbol{x_2}$), the percentage of toddlers whose growth is monitored ($\boldsymbol{x_4}$), coverage of health services for pregnant women ($\boldsymbol{x_6}$), number of nutrition workers ($\boldsymbol{x_7}$), percentage of households with adequate drinking water ($\boldsymbol{x_9}$), percentage of households with adequate sanitation ($\boldsymbol{x_{10}}$), human development index ($\boldsymbol{x_{11}}$), and population density ($\boldsymbol{x_{13}}$).

Keywords: Ridge; LASSO; Elastic-Net; Multicollinearity; Toddler stunting.

1. Introduction

Regression analysis is a method of data analysis that is often used in modeling the relationship between the dependent variable and one or more independent variables. One type of regression analysis is multiple linear regression analysis. In multiple linear regression analysis, it is not uncommon for specific problems to arise during the analysis. One of them is the problem of multicollinearity. According to [1], one of the assumptions in the regression analysis that must be fulfilled is the absence of multicollinearity. Multicollinearity is a condition that appears in multiple regression analysis when one independent variable is correlated with another independent variable. Multicollinearity can create inaccurate estimates of the regression coefficients, inflate the standard errors of the regression coefficients, deflate the partial t-tests for theregression coefficients, give false, nonsignificant, p-values, and degrade the predictability of the model [1, 2]. Multicollinearity is a serious problem, where in cases of high multicollinearity, it results in making inaccurate decisions or increasing the chance of accepting the wrong hypothesis. Therefore it is very important to find the most suitable method to deal with multicollinearity [3]. According to [4], there are several ways to detect the presence of multicollinearity including looking at the correlation between independent variables and using the Variance Inflation Factor (VIF). As for the method to overcome the problem of multicollinearity, one way is by shrinking the estimated coefficients. The shrinkage method is often referred to as the regularization method. The regularization method can shrink the parameters to near zero relative to the least squares estimate. The regularization methods that are often used are Regression Ridge, Least Absolute Shrinkage and Selection Operator (LASSO), and Elastic-Net [5]. Ridge Regression is a technique to stabilize the value of the regression coefficient due to multicollinearity problems. By adding a degree of bias to the regression estimate, RR reduces the standard error and obtains a more accurate estimate of the regression coefficient than the OLS. Meanwhile, LASSO and Elastic-Net overcome the problem of multicollinearity by reducing the regression coefficients of the independent variables that have a high correlation close to zero or exactly zero. This study will explore ridge regression, LASSO and Elastic-Net in dealing with multicollinearity problems in multiple regression analysis.

2. Regulized Regression

One method that can be used to estimate parameters is Ordinary Least Squares (OLS). This method requires the absence of multicollinearity between independent variables. If the independent variable has multicollinearity, the estimate of the regression coefficient may be imprecise. This method is used to estimate β by minimizing the sum of squared errors. If the data consists of n observations $\{y_i, x_i\}_{i=1}^{n}$ and each observation i includes a scalar response y_i and a vector of p predictors (regressors) x_{ij} for j=1,…,m, a multiple linear regression model can be written as n the matrix form the model as $Y = X\beta + \varepsilon$ where Y_{nx1} is the vector dependent variable, X_{nxm} represents the explanatory variables, β_{mx1} is the regression coefficients to be estimated, and ε_{mx1} represents the errors or residuals. $\hat{\beta}^{OLS} = (X^T X)^{-1} X^T Y$ is estimated regression coefficients using OLS by minimizing the squared distances between the observed and the predicted dependent variable [2, 4]. To have unbiased OLS estimation of the model, some assumptions should be satisfied. Those assumptions are that the errors have an expected value of zero, that the independent variables are non-random, that the independent variables are linearly independent (non-multicollinearity), that the disturbance are homoscedastic and not autocorrelated. If the independent variables have multicollinearity the estimates of coefficient regression may be imprecise.

2.1 Ridge Regression

Ridge regression introduced by [6] is one method for deal with multicollinearity problems. The ridge regression technique is based on addition the ridge parameter (λ) to the diagonal of the $X'X$ matrix forms a new matrix ($X'X+\lambda I$). is called ridge regression because diagonal one in the correlation matrix can be described as ridge [7]. The ridge regression coefficients estimator is

$$\hat{\beta}_R = (X^T X + \lambda I)^{-1} X^T Y, \quad \lambda \geq 0$$

when $\lambda = 0$, the ridge estimator become as the OLS. If > 0 the ridge estimator will be biased against the $\hat{\beta}^{OLS}$ but tends to be more accurate than the least squares estimator. Ridge regression can also be written in Lagrangian form:

$$\hat{\beta}^{RIDGE} = \arg\min_{\beta} \|y - X\beta\|_2^2 + \lambda\|\beta\|_2^2$$

Ridge regression has the ability to solve problems multicollinearity by limiting the estimated coefficients, hence, it reduces the estimator's variance but introduces some bias [8].

2.2 Least Absolute Shrinkage and Selection Operator

Least Absolute Shrinkage and Selection Operator (LASSO) introduced by [9] is a method that aims to reduce the regression coefficients of independent variables that have a high correlation with errors to exactly zero or close to zero. LASSO regression can also be written in Lagrangian form:

$$\hat{\beta}^{LASSO} = \arg\min_{\beta} \|y - X\beta\|_2^2 + \lambda\|\beta\|_1$$

with the condition $\|\beta\|_1 \leq \lambda$, where λ is a tuning parameter that controls the shrinkage of the LASSO coefficient with $\lambda \geq 0$. If $\lambda < \lambda_0$ with $\lambda_0 = ||\hat{\beta}_j||_1$ it will cause the shrinkage coefficient to approach zero or exactly zero, so LASSO helps as a variable selection [9, 13]. Like ridge, the absolute value penalty of the. LASSO coefficient introduces shrinkage towards zero. However, on ridge regression, some of the coefficients are not shrinks to exactly zero.

2.3 Elastic-Net

According to [10], the Elastic-Net method can shrink the regression coefficient exactly to zero, besides that this method can also perform variable selection simultaneously with Elastic-Net penalties which are written as follows:

$$\sum_{j=1}^{p}\left[\alpha\left|\beta_j\right| + (1-\alpha){\beta_j}^2\right]$$

with $\alpha = \frac{\lambda_1}{\lambda_1+\lambda_2}$, $0 \leq \alpha \leq 1$

The coefficient estimator on Elastic-Net can be written as follows:

$$\hat{\beta}^{net} = \sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + \lambda_2\sum_{j=1}^{p}{\beta_j}^2 + \lambda_1\sum_{j=1}^{p}\left|\beta_j\right|$$

Elastic-Net can be used to solve problems from LASSO. Where the LASSO Regression has disadvantages include; when p > n then LASSO only chooses n variables included in the model, if there is a set of variables with high correlation, then LASSO only tends to choose one variable from the group and doesn't care which one is selected, and when p < n, LASSO performance is dominated by Ridge Regression.Multicollinearity is the existence of a linear relationship between independent variables. where multicollinearity can occur in either some or all of the independent variables in the multiple linear regression model [1]. One way to detect multicollinearity is to use the Variation Inflation Factor (VIF). VIF value can be calculated by the following formula:

$$VIF_j = \frac{1}{1-R_j^2}$$

if the VIF value > 10, it can be concluded significantly that there is multicollinearity between the independent variables and one way to overcome multicollinearity is using the Ridge Regression, LASSO and Elastic-Net.

2.4 Measurement of Performance

To assess performance on the method studied in multicollinearity handling, will be evaluated using Average Mean Square Error (AMSE) of regression coefficient $\boldsymbol{\beta}$ is measured. AMSE is defined as:

$$AMSE\left(\hat{\beta}\right) = \frac{1}{n}\sum_{j=1}^{m}\left\|\hat{\beta}^{(j)} - \beta\right\|^2$$

where $\widehat{\boldsymbol{\beta}}^{(j)}$ denotes the estimated parameter in the *j*-th simulation. AMSE value close to zero indicates that the slope and intercept are correctly estimated [11]. Akaike Information Criterion (AIC) is also used as the performance criterion with formula :

$$AIC_C = 2k - 2\ln(\hat{L})$$

Where, $\hat{L} = p(x|\hat{\theta}, N), \hat{\theta}$ are the parameter values that maximize the likelihood function, x = the observed data, n = the number of data points in x, and k = the number of parameters estimated by the model [14]. The best model is indicated by the lowest values of AIC.

3. Materials and Methods

The data used in this study are simulated data and real data. Research with simulated data is done by $p = 6$, with sample size $n = 25, 50, 75$ and $\beta_0 = 0$; $\beta_1 = \beta_2 = \ldots = \beta_6 = 1$ with the true model as $Y = X\beta + \varepsilon$. Following [12], to obtain the multicollinearity in each data set, Xp is generated using Monte Carlo's simulation using formula:

$$X_{ij} = (1-\rho^2)^{1/2} u_{ij} + \rho u_{i(p+1)}, i = 1,2, \ldots, n, j = 1.2, \ldots, p$$

Where u_{ij} are independent standard normal pseudo-random numbers and ρ is specified so that the theoretical correlation between any two independent variables is given by ρ^2. Dependent variable ($\boldsymbol{Y}$) for each p independent variables is from $Y = X\beta + \varepsilon$ with β parameters vectors are chosen arbitrarily for p= 6 and $\varepsilon \sim N\ (0, 1)$. Application to real data used in this study is data on cases of stunting toddlers in Indonesia with the number of observations n = 34 and the number of independent variables p = 13. The variables used in the original data are the percentage of toddlers who are malnourished (x_1), the percentage of children under five receiving exclusive breastfeeding (x_2), the percentage of children receiving complete basic immunization (x_3), the percentage of children under five being monitored for growth (x_4), the percentage of children receiving DPT-HB-Hib3 immunization (x_5), coverage of health services for pregnant women (x_6), number of nutrition workers (x_7), number of hospitals (x_8), percentage of households with proper drinking water (x_9), percentage of households with proper sanitation (x_{10}), development index human population (x_{11}), the number of poor people (x_{12}), and population density (x_{13}). To quantify the amount of multicollinearity in the data set, the inflation factor variance (VIF) is examined. Cross validation is used to find the value of λ for Ridge, LASSO and Elastic-Net. The performance of the OLS, Ridge, LASSO and Elastic-Net methods was compared based on the AMSE and AIC values.

4. Results and Discussion

The initial VIF values of simulated data is designed to have a high correlation (ρ = 0.99) between 2, 3, and 6 independent variables. Consequently, the VIF value of the corresponding variable greater than 10 indicates there is a problem of multicollinearity in these variables. The experiment was repeated 1000 times to get accurate estimation results. The simulation results can be seen in Table 1.

Table 1. VIF between variables

Multicollinearity	**Variables**	**VIF**
Between 2 variables	x_1	40.746.478
	x_2	41.988.552
	x_3	1.335.763
	x_4	1.364.037
	x_5	1.138.785
	x_6	1.117.702
Between 3 variables	x_1	43.820.829
	x_2	56.075.303
	x_3	42.204.721
	x_4	1.455.187
	x_5	1.472.338
	x_6	1.376210
Between 6 variables	x_1	60.10378
	x_2	86.63088
	x_3	44.82822
	x_4	81.03695
	x_5	104.1332
	x_6	72.46558

A comparison of the AMSE of the four methods is presented in Table 2. In this table it can be seen that OLS has the highest AMSE value compared to the other three methods at each level of multicollinearity followed by LASSO. On the other hand, ridge gives a lower AMSE score than OLS and LASSO but still higher compared to Elastic-Net. Lowest AMSE given by Elastic-Net at every case. The same results can be seen in the comparison of AIC values in Table 3. This clearly shows that Elastic-Net is the most accurate estimator when there is a severe multicollinearity problem. The results also show that the sample size has an effect on the AMSE and AIC score. The higher it is the larger the sample size used, the lower the respective AMSE and AIC value estimator. One of them can be seen in the simulation data with 2 independent variables which contain high multicollinearity, where the AMSE value decreases when n gets bigger. From Table 2 and Table 3, it can also be seen that Elastic-Net has the lowest AMSE and AIC value compared to OLS, Ridge, and LASSO.

Table 2. Average Mean Square Error of OLS, Ridge, Lasso, and Elastic-Net

Number of Multicollinearity	n	AMSE			
		OLS	Ridge	LASSO	Elastic-Net
	25	3.233	0.600	2.347	0.004
2	50	1.304	0.300	1.130	0.003
	75	1.099	0.200	1.001	0.002
	25	6.252	1.200	4.229	0.018
3	50	2.963	0.500	2.428	0.013
	75	1.890	0.300	1.766	0.003
	25	16.659	1.624	8.966	0.013
6	50	5.831	0.795	4.684	0.011
	75	4.116	0.622	3.552	0.006

Table 3. Akaike Information Criterion (AIC) of OLS, Ridge, Lasso, and Elastic-Net

Number of Multicollinearity	n	AMSE			
		OLS	Ridge	LASSO	Elastic-Net
	25	43.34	12.349	14.512	1.041
2	50	27.27	12.430	14.037	7.436
	75	21.08	12.231	14.000	7.779
	25	59.82	15.140	15.730	2.990
3	50	68.31	11.946	14.444	3.456
	75	61.74	11.724	14.171	3.741
	25	84.32	22.078	17.562	24.93
6	50	102.16	12.662	15.228	7.124
	75	120.12	12.046	14.788	4.172

To make it easier to understand the AMSE comparison of the four methods, the AMSE values are also presented graphically and can be seen in Figures 1 – 3.

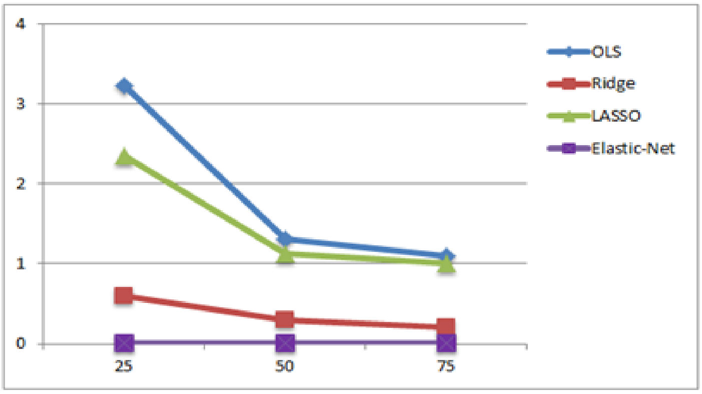

Figure 1. AMSE of OLS, Ridge, LASSO, and Elastic-Net Contain Multicollinearity in 2 Independent Variables

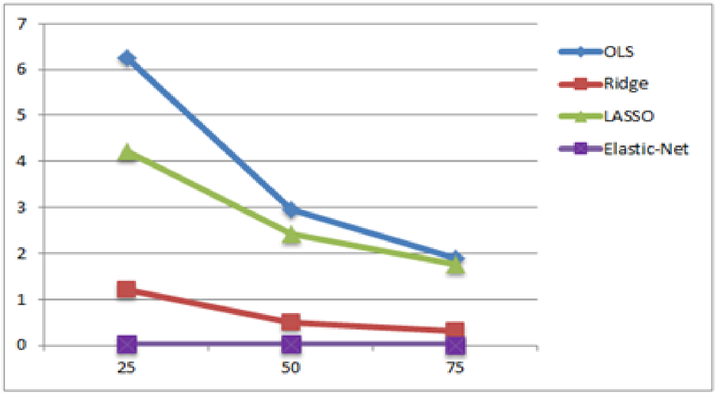

Figure 2. AMSE of OLS, Ridge, LASSO, and Elastic-Net Contain Multicollinearity in 3 Independent Variables

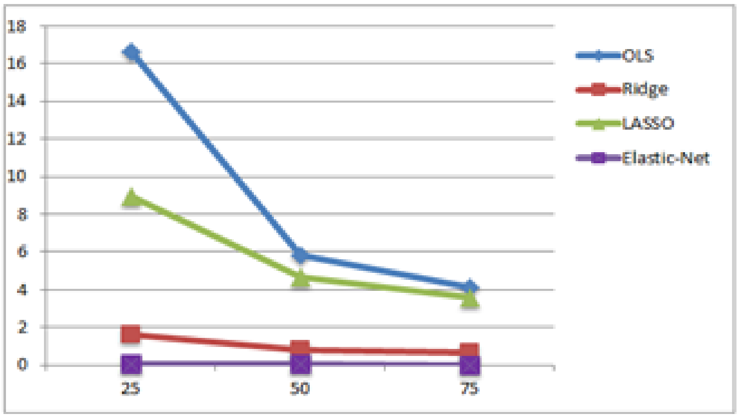

Figure 3. AMSE of OLS, Ridge, LASSO, and Elastic-Net Contain Multicollinearity in 6 Independent Variables

Analysis using data on stunted toddlers in Indonesia consisting of 13 independent variables found that there were 6 independent variables with VIF values > 10, which indicated that a multicollinearity problem was detected in them. In Table 4 it is shown that the variables X_3, X_5, X_6, X_7, X_8, dan X_{12} have a correlation with each other. The VIF value is more than 10. Therefore, Ridge regression, LASSO, and Elastic-Net methods will be applied so that the correlation between these variables can be controlled. The results of applying these three methods and when compared with the OLS method can be seen in Table 5. From Table 5, it can be seen that AMSE value of the Elastic-Net method is the smallest compared to OLS, Ridge and LASSO. This means that the simulation data results are accurate enough to be applied to real data. The results of the differences in AMSE and AIC values can also be seen in Figure 4-5.

Table 4. VIF Value

Variable	**VIF**
X_1	1,85119
X_2	2,44541
X_3	479,00167
X_4	2,22135
X_5	365,17829
X_6	38,01047
X_7	17,09611
X_8	42,36026
X_9	1,94832
X_{10}	1,59373
X_{11}	4,26647
X_{12}	35,73663
X_{13}	2,07855

Table 5. AMSE & AIC from OLS, RIDGE, LASSO, and Elastic-Net On Real Data

Method	**AMSE**	**AIC**
OLS	23.37169	135.151
Ridge	6.96006	93.966
LASSO	4.58480	79.773
Elastic-Net	4.27151	77.366

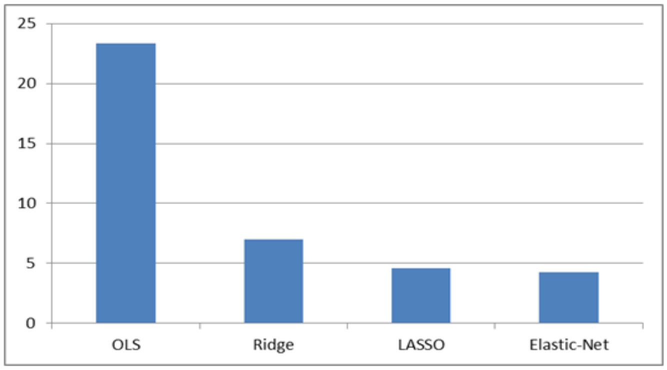

Figure 4. AMSE from OLS, Ridge, LASSO, and Elastic-Net on Real Data

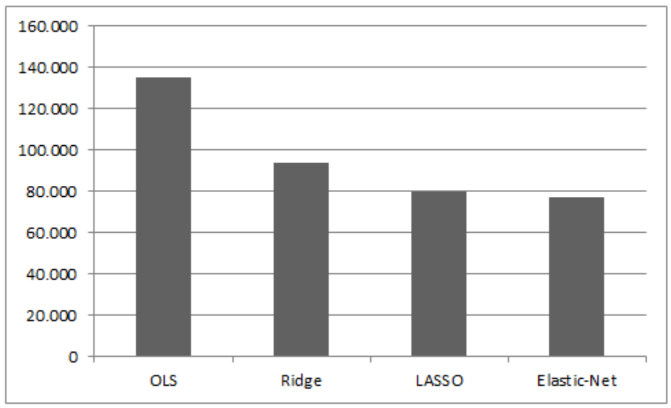

Figure 5. AIC from OLS, Ridge, LASSO, and Elastic-Net on Real Data

From Figures 4-5, it can be seen that Elastic-Net is the best method to control multicollinearity in stunting toddlers in Indonesia. Based on the Elastic-Net method, an analysis was carried out of the variables that influence stunting toddlers in Indonesia. The results show that percentage of toddlers who are malnourished ($\boldsymbol{x_1}$), percentage of toddlers who receive exclusive breast milk ($\boldsymbol{x_2}$), percentage of toddlers whose growth is monitored ($\boldsymbol{x_4}$), coverage of health services for pregnant women ($\boldsymbol{x_6}$), number of nutrition workers ($\boldsymbol{x_7}$), percentage of households with adequate drinking water ($\boldsymbol{x_9}$), percentage of households with adequate sanitation ($\boldsymbol{x_{10}}$), human development index ($\boldsymbol{x_{11}}$), and population density ($\boldsymbol{x_{13}}$) are the variables that influence the stunting toddlers in Indonesia.

5. Conclusions

Based on the simulation results at p = 6 and the number of data n = 25, 50, and 75 containing severe multicollinearity between 2, 3, and 6 independent variables, Elastic-Net can solve multicollinearity problems better than the other three methods. As for the original data, the smallest AMSE and AIC value is obtained by using Elastic-Net. Overall it can be concluded, using both simulated data and real data, Elastic-Net is able to solve multicollinearity problems better than OLS, Ridge, and LASSO.

References

Gujarati D. N. & Porter, D. C. Basic Econometrics. Boston: McGraw-Hill Irwin, 2009.

Draper, N.R. and Smith, H. Applied Regression Analysis. 3rd edition. New York: Wiley, 1998.

Judge, G.G., Introduction to Theory and Practice of Econometrics. New York: John Willy and Sons, 1988.

Montgomery, D. C., Peck, E. A., & Vining, G.G. Introduction to Linear Regression Analysis. Fourth Edition. New York: John Willey and Sons, 2012.

James, G., Witten, D., Hastie, T., & Tibshirani, R. An Introduction to Statistical Learning-with Applications in R. New York: Springer, 2013.

Hoerl, A. E. 1962. Application of Ridge Analysis to Regression Problems. Chem. Eng. Prog., 58, 54-59.

Hoerl, A.E. and Kennard, R.W. 2000. Ridge Regression: Biased Estimation for nonorthogonal Problems. Technometrics, 42, 80-86.

James, G., Witten D., Hastie T., Tibshirani R An Introduction to Statistical Learning: With Applications in R. New York: Springer Publishing Company, Inc., 2013.

Tibshirani, R. 1996. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1): 267-288.

Zou, H. & Hastie, T. 2005. Regularization and Variable Selection Via the Elastic Net. Journal of the Royal Statistical Society. Series B: Statistical Methodology, 67(2): 301–320.

Herawati, N., Nisa, K., Setiawan, E., Nusyirwan, & Tiryono. 2018. Regularized Multiple Regression Methods to Deal with Severe Multicollinearity. International Journal of Statistics and Applications, 8(4): 167-172.

McDonald, G.C. and Galarneau, D.I. 1975. AMonte Carlo Evaluation of Some Ridge-Type Estimators. J. Amer. Statist. Asoc., 70(350): 407 – 416.

Hastie, T., Tibshirani, R., Mainwright, M. 2015. Statistical Learning with Sparsity The LASSO and Generalization. USA: Chapman and Hall/CRC Press.

Akaike, H. 1973. Information Theory and An Extension of The Maximum Likelihood Principle. In B.N. Petrow and F. Csaki (eds), Second International Symposium on Information Theory (pp.267-281). Budapest: Academiai Kiado