Shrinkage Parameters for Each Explanatory Variable Found Via Particle Swarm Optimization in Ridge Regression

Abstract

Ridge regression method is an improved method when the assumptions of independence of the explanatory variables cannot be achieved, which is also called multicollinearity problem, in regression analysis. One of the way to eliminate the multicollinearity problem is to ignore the unbiased property of $\beta$. Ridge regression estimates the regression coefficients biased in order to decrease the variance of the regression coefficients. One of the most important problems in ridge regression is to decide what the shrinkage parameter ($\lambda$) value will be. This $\lambda$ value was found to be a simple value in almost all these studies in the literature. In this study, different from those studies, we found different $\lambda$ values corresponding to each diagonal elements of variance-covariance matrix of $\beta$ instead of a single value of $\lambda$ by using a new algorithm based on particle swarm optimization. To evaluate the performance of our proposed method, the proposed method is firstly applied to real-life data sets and compared with some other studies suggested in the ridge regression literature. Finally, two different simulation studies are performed and the performance of the proposed method with different conditions is evaluated by considering other studies suggested in the ridge regression literature.

Introduction

The functional relation between a dependent variable and more than one independent variable is examined by multiple regression analysis. The purpose of the multiple regression analysis is the creation of the best model that can predict the dependent variable by using the independent variables. For this purpose, the most common method to create the best model is ordinary least square (OLS) estimates method. In this method, the sum of error squares to be minimal is calculated to predict the parameters of the model.

There are some valid assumptions for the implementation of the multiple regression analysis. These are; the absence of multicollinearity problem among independent variables, the variance of error term must be constant for all independent variables and the covariance between error term and independent variables must be equal to zero.

One of the major problems in multiple regression analysis is multicollinearity problem. If there is a full or high degree linear relationship among independent variables, this situation is called as multicollinearity. Besides, multicollinearity has some important effects on OLS estimates of the regression coefficients. In the presence of multicollinearity, the OLS regression coefficients have large variance. And also, the regression coefficients can be estimated incorrectly and the standard errors of regression coefficients can be found as exaggerated in the presence of multicollinearity. If the regression coefficients can be estimated incorrect, it can be obtained incorrect results statistically.

Therefore, ridge regression method is used to obtain stable coefficient estimates for the estimation of the regression coefficients. That means, ridge regression has been suggested to overcome the multicollinearity problem.

In the literature, it is commonly accepted that if the variance inflation factors (VIF) values are greater than 10 there is a multicollinearity problem. This is a rule of thumb and this is not exact information. Similarly, condition number can be used to determine multicollinearity problem by using rule of thumbs. As a result of, determining of multicollinearity problem can be realized by using some criteria.

The two methods most commonly used to determine the effects of multicollinearity problem are VIF and condition
The diagonal elements of $VIF_j$ are called as $VIF$ and are given by the Equation 1.

$$VIF_j = \frac{1}{1 - R_j^2} \quad j = 1, \ldots, p$$  
(1)

In this Equation, $R_j^2$ is the determination coefficient obtained from the multiple regression of $X_j$ on the remaining $(p - 1)$ regressor variables in the model.

It can be said that there is a multicollinearity problem among the relevant independent variables if these $VIF$ values increase ($VIF$ values $\geq 10$). And also, if $VIF$ values are increased, the degree of the multicollinearity increases with the increase of $VIF$ values.

Condition number method is another method to determine the multicollinearity problem which is based on the eigenvalues of $XX'$ matrix. The formula of the condition number (CN) was given in Equation 2.

$$\phi = \frac{\lambda_{\text{max}}}{\lambda_{\text{min}}}$$  
(2)

In this Equation, $\lambda$ shows the eigenvalues of $XX'$. The relationship between condition number and multicollinearity is given in Table 1.

In summary, the determining of multicollinearity problem can be done by following two rules of thumbs. The first one is that if $VIF$ values are greater than 10 multicollinearity is high. The second one is checking condition number as given in Table 1.

In addition, another problem in ridge regression is finding optimal biasing parameter $(k)$ value. This $k$ value is a very small constant determined by the researcher [1]. Several methods were proposed for finding it in the literature. These methods have been proposed in the studies of [2–22].

And also, there are many methods in the literature for ridge regression [23–29]. And also, [30] proposed some new methods that take care of the skewed eigenvalues of the matrix of explanatory variables. [31] Proposed an iterative approach to minimize the mean squared error in ridge regression. [32] Proposed new ridge parameters for ridge regression. [33] Proposed an optimal estimation for the ridge regression parameter. [34,35] Proposed some new estimators for estimating the ridge parameter.

This $k$ value was found to be a single value in almost all these studies in the literature. But in this study, we found different $k$ values corresponding to each diagonal elements of variance-covariance matrix of $\hat{\beta}$ instead of a single value of $k$ by using a new algorithm based on particle swarm optimization.

The rest part of the paper can be outlined as below:

### Table 1: Condition number and its effects.

| Condition Number | Multicollinearity             |
|------------------|-------------------------------|
| CN $\leq$100    | There is no serious multicollinearity |
| 100 $<$CN $<$1000 | Strong multicollinearity     |
| CN $\geq$1000   | Severe multicollinearity exist in the data |

The second section of the paper is about ridge regression. The methodology of the paper is given in Section 3. The implementation of our proposed method is given in Section 4. Two different simulation studies are performed under the title of simulation study and finally, discussions are presented in Section 6.

### Ridge regression

Ridge regression is a remedy used in the presence of multicollinearity problem and it was firstly proposed by [1]. Ridge regression method has two important advantages according to OLS method. One of them is to solve the multicollinearity problem and the other one is to decrease the mean square error (MSE). The solution technique of ridge regression is similar with OLS. Besides, the difference between ridge regression and OLS is the $k$ value. This $k$ value is also called as biased parameter or shrinkage parameter and it takes values between 0 and 1. This $k$ value is added to the diagonal elements of the correlation matrix and thus biased regression coefficients are obtained.

The OLS estimates of regression coefficients and ridge estimates of regression coefficients are shown in the Equations 3 and 4 respectively.

$$\hat{\beta} = (X'X)^{-1}X'Y$$  
(3)

$$\hat{\beta}_R = (X'X + kI)^{-1}X'Y$$  
(4)

As noted above, ridge regression is a biased regression method. The proof of this situation is shown in Equation 5.

$$\hat{\beta}_R = (X'X + kI)^{-1}X'Y = (X'X + kI)^{-1}(X'X)\hat{\beta} = Z\hat{\beta}$$

$$E(\hat{\beta}_R) = E(Z\hat{\beta}) = Z\beta$$  
(5)

It is clearly seen that ridge estimates of regression coefficients $(\hat{\beta}_R)$ are biased estimates. One of the most important points to be considered in the ridge regression is the $k$ value. There are many methods proposed in the literature to find the optimal $k$ value. Ridge trace is one of these methods. Ridge trace is a plot of the elements of the ridge estimator versus $k$ usually in the interval $(0, 1)$ [1].

The other methods in the literature used to find the optimal $k$ value were given in the Equations 6–14, respectively.

$$k = \frac{\sigma}{\beta} \quad \beta \quad \beta$$  
(6)

$$k = \frac{\rho \sigma}{\sum_{i=1}^{p} \lambda_i \beta_i}$$  
(7)

$$k = \frac{\rho \sigma}{\sum_{i=1}^{p} \left( \beta_i \left( \frac{\sigma}{\beta_i} \right)^{\frac{1}{2}} \right)^{\frac{1}{2}}}$$  
(8)
In the orthogonal version of the classical linear regression model: 
\[ Y = X\beta + \epsilon \]
where \( Y \) is the response vector, \( X \) is the design matrix, \( \beta \) is the vector of regression coefficients, and \( \epsilon \) is the error term. The objective function to be minimized is given by:
\[ \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 \]

And also, the generalized ridge regression estimator of Hoerl and Kennard [1, 40] is given in [41] by following Equations 17–20.

Let \( \Lambda \) and \( Q \) be the matrices of eigenvalues and eigenvectors of \( (X'X)^{-1} \). In the orthogonal version of the classical linear regression model: 
\[ Z = XQ \alpha = \alpha' \beta, \]
where \( \alpha \) is the vector of regression coefficients, \( \beta \) is the vector of regression coefficients, and \( Q \) is the matrix of eigenvalues and eigenvectors. The objective function is to find the generalized ridge estimator of \( \beta \) that minimizes the mean squared error (MSE) of the regression coefficients.

The optimization problem in the proposed method can be formulated as:
\[ \min_{\beta} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \lambda \sum_{i=1}^{p} \beta_i^2 \]

where \( \lambda \) is the penalty parameter. This is an undesirable result.

In [41], other estimation formulas for optimum shrinkage parameters are given below.

\[ \hat{k}_i = \frac{\lambda_i \hat{\sigma}^2}{(n-k)\hat{\sigma}^2 + \lambda_i \hat{\alpha}^2} \]

**Methodology**

Finding the optimal \( k \) value is an important problem in ridge regression. The \( k \) values recommended in the literature were given in the previous section. And also, there are some heuristic methods such as genetic algorithms to find the optimal \( k \) value in the literature proposed by [18, 21]. And also, [22] have found the \( k \) value by using particle swarm optimization (PSO). In all these methods suggested in the literature, this \( k \) value was found as a single value. But in this study, we found different \( k \) values corresponding to each explanatory variable instead of a single value of \( k \) by using an algorithm based on particle swarm optimization. And also, this paper is the improvement form of the study of [22].

The objective function of the paper was created by considering both mean absolute percentage error (MAPE) criterion and VIF values at the same time. The aim of the objective function is to find the optimal \( k \) values by finding the VIF values less than 10 and SSE (sum of square errors) minimum, at the same time. And also, we add a parameter \( \varepsilon(k) \) to the second part of the objective function. This parameter can be called as penalty parameter. If the VIF value corresponds to any explanatory variable is bigger than 10 the value of the objective function is increased. This is an effect of the penalty parameter. This is an undesirable result.

The optimization problem in the proposed method can be given in Equation 21.
Objective function:
\[
\min_{k_1, k_2, \ldots, k_p} MAPE\left(k_1, k_2, \ldots, k_p\right) + \xi(k_1, k_2, \ldots, k_p)
\]  
(21)
with subject to: \(0 \leq k_1, k_2, \ldots, k_p \leq 1\) \((j = 1, 2, \ldots, p)\)

where \(MAPE\left(k_1, k_2, \ldots, k_p\right)\) and \(\xi(k_1, k_2, \ldots, k_p)\) can be defined in Equations 22 and 23 respectively.

\[
MAPE\left(k_1, k_2, \ldots, k_p\right) = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{\hat{y}_i} \right|^\wedge
\]
(22)

\[
\xi(k_1, k_2, \ldots, k_p) = \begin{cases} 
0 & \forall \text{VIF}_j < 10, j = 1, 2, \ldots, p \\
\sum_{j=1}^{p} \text{VIF}_j & \text{otherwise}
\end{cases}
\]
(23)

(p shows the number of explanatory variables.)

The optimization problem defined as in (21) was solved by using PSO in the proposed method. PSO is a popular artificial intelligence technique and it was firstly proposed by [42]. The algorithm of the proposed method is given below.

Algorithm

Step 1. The parameters such as \(pn, c_1, c_2\) etc., are determined. These parameters are as follows:

\(pn\): particle number of swarm

\(c_1\): Cognitive coefficient

\(c_2\): Social coefficient interval

\(\text{max}\): Maximum iteration number

\(w\): Inertia weight

Step 2. Generate random initial positions and velocities.

The initial positions and velocities are generated by uniform distribution with \((0,1)\) parameters. Each particle has velocities up to the number of explanatory variables and each particle has positions up to the number of explanatory variables which represents \(\left(k_1, k_2, \ldots, k_p\right)\) values. \(x'_m\) represents the position of particle \(m\) at iteration \(t\) and \(v'_m\) represents the velocity of the particle \(m\) at iteration \(t\).

Step 3. The fitness function was defined as in (21) and the fitness values of the particles are calculated.

Step 4. \(P\text{best}_m\) and \(G\text{best}\) particles given in (24) and (25), respectively, are determined according to fitness values.

\[
P\text{best}_m' = (pm), m = 1, 2, \ldots, pn
\]
(24)

\[
G\text{best}' = (pg)
\]
(25)

\(P\text{best}\) is constructed by the best results obtained in the related positions at iteration \(t\). \(G\text{best}\) is the best result in the swarm at iteration \(t\).

Step 5. New velocities and positions of the particles are calculated by using the Equations given in (26) and (27).

\[
v_{m}^{t+1} = \omega v_{m}^{t} + c_1 x_{\text{rand}_1} \times (P\text{best}_m' - x_{m}^{t}) + c_2 x_{\text{rand}_2} \times (G\text{best}' - x_{m}^{t})
\]
(26)

\[
x_{m}^{t+1} = x_{m}^{t} + v_{m}^{t+1}
\]
(27)

Where \(\text{rand}_1\) and \(\text{rand}_2\) are random numbers generated from \(U(0,1)\).

Step 6. Step 3 to Step 6 is repeated until \(t<\text{max}\).

Step 7. The optimal \(\left(k_1, k_2, \ldots, k_p\right)\) values are obtained as \(G\text{best}\).

Implementation

The proposed algorithm was applied to two different and well known data sets in order to investigate of the proposed method. These two data sets named “Import Data” and “Longley Data” were used to evaluate the performance of the proposed method. Import data was analyzed by [43]. The variables of “Import Data” are; imports (IMPORT-Y), domestic production (DOPROD-X1), stock formation (STOCK-X2) and domestic consumption (CONSUM-X3), all measured in billions of French francs for the years 1949 through 1959. Both Import data and Longley data were solved by using fixed point method ([21]), iterative method ([39]), [22]’s method and the algorithm proposed in this paper. In the proposed algorithm, PSO parameters were chosen as \(pn = 30, w = 0.9, c_1 = c_2 = 2\) and \(\text{max}=100\). In the iterative ridge method the stopping criteria were chosen as \(=10^{-8}\). The results of each method were presented in Tables 2 and 3, respectively.

As we can see from Table 2, our proposed method has minimum SSE and MAPE values. And also there is no multicollinearity problem when “Import Data” solved by our proposed method. But, there is a multicollinearity problem when “Import Data” solved by FPRRM and IRRM methods because of the VIF values of these methods are bigger than 10. Although, other methods can give smaller SSE and MAPE values they do not still solve the multicollinearity problem. Because it is clearly seen that some VIF values of these methods are greater than 10.

As we can see from Table 3, our proposed method has minimum MAPE value when compared with other methods. But SSE value of our proposed method is not the smallest one. The SSE value of OLS is smaller than our proposed methods. But, it is clearly seen that the OLS method has multicollinearity problem when “Longley Data” solved by this method. But our proposed method has no multicollinearity problem.

As a result, finding \(k\) values for each explanatory variable gives better results than finding a single \(k\) value. And also, our proposed has no multicollinearity problem.

Simulation study

Two different simulation studies are performed in this section of the paper in order to show the performance of the
proposed method in different levels of multicollinearity and standard deviation of error term and the superiority of the proposed method when compared with other methods.

The First Simulation Study: In this simulation study, the proposed method was compared with ridge regression methods given in [2, 22, 39] by a simulation study. The number of observations \((n)\) was taken as 100, 500 and 1000; the standard deviation of error term \((\sigma)\) was taken as 0.01 and 1 and comparisons were made for the total 6 cases. For each case, 1000 data set including multicollinearity problem was created.

The first three independent variables were generated from standard normal distribution as given in Equation 28.

\[
X_i \sim N \left(0, 1 \right) \quad i = 1, 2, 3
\]

(28)

The last two independent variables were generated by using Equation 29. Thus, it is provided to arise multicollinearity problem for the data set by providing a high correlation between independent variables \(X_1\) and \(X_2, X_3\) and \(X_4\).

\[
X_i = U \left(10, 20 \right) + U \left(5, 20 \right) X_i + N \left(0, 7 \right) \quad i = 4, 5
\]

(29)

The observations of dependent variable were obtained using 

\[
Y = \sum_i X_i + N \left(0, \sigma \right)
\]

(30)

For each data generated in each case, \(\sum VIF^2\), SSE, MAPE and CN values are calculated by using proposed method, the studies [2, 22, 39]. The formula of SSE is given in Equation 31.

\[
SSE = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2
\]

(31)

The comparison of \(VIF\) values, SSE and MAPE obtained from OLS, FPRRM, IRRM, [22] and proposed method for Import Data.

| Variable | S.C. | VIF | [2] \((k=0.0016)\) | VIF | [2] \((k=0.0042)\) | VIF | [22] \((k=0.0090)\) | Proposed Ridge Method | k values obtained from Proposed Ridge Method |
|----------|------|-----|-------------------|-----|-------------------|-----|-------------------|-------------------------------|----------------------------------------|
| X1       | -0.34| 186.11| -0.03            | 72.09| 0.16              | 27.99| 0.29              | 9.99                          | -0.07 9.03                        |
| X2       | 0.21 | 1.02 | 0.22             | 1.00 | 0.22              | 1.00 | 0.22              | 0.98                          | 0.21 1.00                         |
| X3       | 1.30 | 186.00| 0.99             | 72.13| 0.80              | 28.01| 0.67              | 9.99                          | 1.03 9.99                         |
| MAPE     | 0.0810| 0.0086| 0.0095           | 0.1185| 0.1097           | 0.1139| 0.1185           | 0.1088                         |

S.C.: Standardized Coefficients.

The most important indicator for the comparison of methods is that VIF and CN would be small. The methods [2] and [39] do not guarantee the solution of multicollinearity problem as seen in the numerical examples. The method [22] and proposed method guarantee that all VIF values are smaller than 10. Therefore, it is suitable to compare the proposed method with [22] method in terms of SSE and MAPE criteria.

The results of median and inter quartile range (IQR) values were given between Tables 4–9.

When all tables are examined, it is clearly seen that \(\sum VIF^2\) and CN values of proposed method is lower than the other methods in all cases.

However, it is seen that the proposed method produces lower MAPE values compared to others despite producing higher SSE values. This is because the objective function of the proposed method may be depending to the MAPE.

The Second Simulation Study: A second simulation study was performed in the paper according to different levels of multicollinearity problem and standard deviation of error term. The regressors were generated by using Equations 32–36 given by (44).

\[
w_{ij} \sim N \left(0, 1 \right) \quad i = 1, 2, \ldots, n \quad j = 1, 2, \ldots, 6
\]

(32)

\[
x_{ij} = (1 - \rho^2)^{1/2} w_{ij} + \rho w_i, \quad i = 1, 2, \ldots, n \quad j = 1, 2, 3
\]

(33)

\[
x_{ij} = w_{ij}, \quad i = 1, 2, \ldots, n \quad j = 4, 5
\]

(34)

The observations of dependent variable were obtained using 

\[
Y = \sum_i X_i + N \left(0, \sigma \right)
\]

(30)

The comparison of \(VIF\) values, SSE and MAPE obtained from OLS, FPRRM, IRRM, [22] and proposed method for Longley Data.

| Variable | S.C. | VIF | [2] \((k=0.0003)\) | VIF | [2] \((k=0.0006)\) | VIF | [22] \((k=0.0172)\) | Proposed Ridge Method | k values obtained from Proposed Ridge Method |
|----------|------|-----|-------------------|-----|-------------------|-----|-------------------|-------------------------------|----------------------------------------|
| X1       | 0.05 | 135.53| -0.01            | 87.32| -0.01             | 87.32| 0.25              | 9.99                          | -0.002 0.59                      |
| X2       | -1.01 | 1788.51| -0.25            | 472.15| -0.25             | 472.15| 0.34              | 1.55                          | -0.001 0.00                      |
| X3       | -0.54 | 33.62| -0.43             | 10.95| -0.43             | 10.95| -0.28             | 2.55                          | -0.37 3.03                      |
| X4       | -0.20 | 3.59 | -0.18             | 2.88 | -0.18             | 2.88 | -0.11             | 1.94                          | -0.15 2.17                      |
| X5       | -0.10 | 399.15| -0.28             | 180.21| -0.28             | 180.21| 0.16              | 7.32                          | -0.10 5.94                      |
| X6       | 2.48 | 758.98| 1.88             | 309.82| 1.88             | 309.82| 0.46              | 4.07                          | 1.39 9.99                       |
| MAPE     | 0.0045| 0.0050| 0.0050           | 0.0123| 0.0050           | 0.0123| 0.0138           | 0.0826                         |

S.C.: Standardized Coefficients.

Citation: Bas E, Egioglu E, Uslu VR (2016) Shrinkage Parameters for Each Explanatory Variable Found Via Particle Swarm Optimization in Ridge Regression. Trends Comput Sci Inf Technol 1(1): 031-039. DOI: http://dx.doi.org/10.17352/tcsit.000005
\[ e_i \sim N(0, \sigma) ; i = 1, 2, \ldots, n \]  
\[ y_i = \sum_{j=1}^{5} \beta_j x_{i,j} + e_i ; i = 1, 2, \ldots, n \]

Where \( w_{i,j} \) independent standard normal are pseudorandom numbers and \( \rho^2 \) is theoretical correlation between any two explanatory variables.

Simulation study was conducted for a total of 8 cases for \( n=500, \sigma = 0.01 \).

**Table 4:** Simulation results for \( n=100, \sigma = 0.01 \)

| Method  | Median | [39] | [2] | [22] | Proposed Method |
|---------|--------|------|-----|------|-----------------|
| SSE     | 0.0001259 | 0.0001259 | 0.0001363 | 0.0000127 | 0.00001576 |
| IQR     | 0.0000756 | 0.0000756 | 0.0000931 | 0.0001270 | 0.00012317 |

**Table 5:** Simulation results for \( n=500, \sigma = 0.01 \)

| Method  | Median | [39] | [2] | [22] | Proposed Method |
|---------|--------|------|-----|------|-----------------|
| SSE     | 0.0001259 | 0.0001259 | 0.0001363 | 0.0000127 | 0.00001576 |
| IQR     | 0.0000756 | 0.0000756 | 0.0000931 | 0.0001270 | 0.00012317 |

Sample size is 100, \( (n=100) \), standard deviation of the standard deviation of error term \( (\sigma = 0.01, 0.1, 1, 5) \) and different degrees of multiple connections \( (\rho = 0.99, 0.999) \) (Tables 10 - 17).

It is clearly seen that in the tables of the simulation Study 2, \( \Sigma \gamma_{RF} \) and CN values of the proposed method do not change significantly when standard deviation of error term values are changed. \( \Sigma \gamma_{RF} \) And CN values of the proposed method are increased dramatically when multicollinearity is increased. And also there is no a hardly ever change to be seen in the MAPE values of the proposed method with the reasonable standard deviation of error term values \( (\sigma = 0.01,0.1) \) or there is a decrease to be seen in the MAPE values of the proposed method when multicollinearity is increased.

In this simulation study, different levels of standard deviation of error term are also employed. As a result of this simulation study it is clearly seen that when standard deviation of error term value is greater than 1 and >1 the model has very big deviation from linear regression model because MAPE values are obtained about 60 and this value is not suitable. And also, it is clearly seen that in the tables of the simulation study 2, the prediction performance of the proposed is affected quite negatively when standard deviation of error term is increased.

**Discussion**

There are some valid assumptions to create a model in multiple regression analysis. One of them is that it should not be multicollinearity problem among independent variables. Ridge regression method is often used in the literature when there is a multicollinearity problem among independent variables.

But, ridge regression has also some problems. One of the most important problems in ridge regression is to decide what
the shrinkage parameter \((k)\) value will be. There are many studies in the literature to find the optimal \(k\) value. In these studies, this \(k\) value was found to be a single value. But in this study, we found different \(k\) values corresponding to each explanatory variable instead of a single value of \(k\) by using a new algorithm based on particle swarm optimization. And also, the proposed method was supported by two simulation studies. Besides, it is an important novelty for ridge regression literature.

### Table 10: Simulation results for \(n=100, \rho = 0.99, \sigma = 0.1\)

| Method | \([39]\) | \([2]\) | \([22]\) | Proposed Method |
|--------|--------|--------|--------|-----------------|
| SSE    | Median | 0.10852 | 0.10852 | 0.08735 | 0.10578         |
|        | IQR    | 0.07191 | 0.07191 | 0.05225 | 0.07341         |
| MAPE   | Median | 0.08589 | 0.08589 | 0.09975 | 0.08738         |
|        | IQR    | 0.02448 | 0.02448 | 0.02849 | 0.02516         |
| \(\sum VIF^2\) | Median | 0.36885 | 0.36885 | 0.20309 | 0.14416         |
|        | IQR    | 0.17361 | 0.17361 | 0.12232 | 0.31462         |
| CN     | Median | 174.00479 | 174.00480 | 150.87468 | 142.72940       |
|        | IQR    | 57.08409 | 57.08410 | 55.21013 | 59.74116         |

### Table 11: Simulation results for \(n=100, \rho = 0.99, \sigma = 0.01\)

| Method | \([39]\) | \([2]\) | \([22]\) | Proposed Method |
|--------|--------|--------|--------|-----------------|
| SSE    | Median | 0.01102 | 0.01102 | 0.00863 | 0.01069         |
|        | IQR    | 0.00670 | 0.00671 | 0.00310 | 0.00595         |
| MAPE   | Median | 0.00089 | 0.00089 | 0.00113 | 0.00093         |
|        | IQR    | 0.00288 | 0.00288 | 0.00085 | 0.00032         |
| \(\sum VIF^2\) | Median | 3.70E-01 | 3.70E-01 | 3.33E-01 | 3.03E-01         |
|        | IQR    | 2.11E-01 | 2.11E-01 | 2.06E-01 | 2.13E-01         |
| CN     | Median | 171.30708 | 171.30710 | 167.37020 | 163.00550       |
|        | IQR    | 57.08409 | 57.08410 | 55.21013 | 59.74116         |

### Table 12: Simulation results for \(n=100, \rho = 0.99, \sigma = 1\)

| Method | \([39]\) | \([2]\) | \([22]\) | Proposed Method |
|--------|--------|--------|--------|-----------------|
| SSE    | Median | 1.05139 | 1.052158 | 1.009223 | 1.041334         |
|        | IQR    | 0.89362 | 0.89368 | 0.894413 | 0.894033         |
| MAPE   | Median | 7.806875 | 7.806608 | 7.805308 | 7.803432         |
|        | IQR    | 1.835182 | 1.832176 | 1.831011 | 1.828457         |
| \(\sum VIF^2\) | Median | 0.073703 | 0.082648 | 0.14884 | 0.074125         |
|        | IQR    | 0.031235 | 0.036015 | 0.070147 | 0.265688         |
| CN     | Median | 110.7648 | 114.7465 | 141.0569 | 121.504          |
|        | IQR    | 25.48441 | 25.1948 | 30.39541 | 51.88191         |

### Table 13: Simulation results for \(n=100, \rho = 0.99, \sigma = 5\)

| Method | \([39]\) | \([2]\) | \([22]\) | Proposed Method |
|--------|--------|--------|--------|-----------------|
| SSE    | Median | 1.872207 | 2.004527 | 2.009562 | 2.013568         |
|        | IQR    | 0.889942 | 0.999704 | 1.121944 | 1.141227         |
| MAPE   | Median | 67.69349 | 66.27068 | 65.8774 | 65.92444         |
|        | IQR    | 8.280155 | 8.75298 | 8.700223 | 8.611794         |
| \(\sum VIF^2\) | Median | 8.19E-05 | 0.002858 | 0.138193 | 0.073933         |
|        | IQR    | 5.31E-05 | 0.009572 | 0.114742 | 0.163497         |
| CN     | Median | 9.919616 | 49.19599 | 138.8491 | 118.3783         |
|        | IQR    | 4.636951 | 37.61656 | 40.51028 | 37.26164         |

### References
1. Hoerl AE, Kennard RW (1970) Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12: 55–67. Link: https://goo.gl/5ZV56T
2. Hoerl AE, Kennard RW, BaldwinKF (1975) Ridge regression: some simulations. Communications in Statistics 4: 105–123. Link: https://goo.gl/QoF3L3
3. McDonald GC, Galanerou DI (1975) A Monte Carlo evaluation of some ridge-type estimators. Journal of the American Statistical Association 70: 407–412. Link: https://goo.gl/TZN2co

4. Lawless JF, Wang P (1976) A simulation study of ridge and other regression estimators. Communications in Statistics – Theory and Methods 14: 1589–1604. Link: https://goo.gl/WRUz0p

5. Hocking RR, Speed FM, Lynn MJ (1976) A class of biased estimators in linear regression. Technometrics 18: 425–437. Link: https://goo.gl/N Ej sRY

6. Gunst RF, Mason RL (1977) Biased estimation in regression: an evaluation using mean squared error. Journal of the American Statistical Association 72: 616–628. Link: https://goo.gl/hXfXln

7. Wichern D, Curchill G (1978) A comparison of ridge estimators. Technometrics 20: 301-311. Link: https://goo.gl/U60lUQ

8. Lawless JF (1978) Ridge and related estimation procedure Theory and Methods. Communications in Statistics 7: 139–164. Link: https://goo.gl/ncEyME

9. Nordberg L (1982) A procedure for determination of a good ridge parameter in linear regression, Communications in Statistics 11: 285–309. Link: https://goo.gl/pNqtc2

10. Saleh AK, Kibria BM (1993) Performances of some new preliminary test ridge regression estimators and their properties. Communications in Statistics – Theory and Methods 22: 2747–2764. Link: https://goo.gl/4XqQNd

11. Haq MS, Kibria BMG (1996) A shrinkage estimator for the restricted linear regression model: ridge regression approach. Journal of Applied Statistical Science 3: 301–316. Link: https://goo.gl/sjCZrw

12. Kibria BM (2003) Performance of some new ridge regression estimators. Communications in Statistics – Simulation and Computation 32: 419–435. Link: https://goo.gl/JoJ6a

13. Pasha GR, Shah MA (2004) Application of ridge regression to multicollinear data. Journal of Research Science 15: 97–106. Link: https://goo.gl/SeP2i5

14. Khalaf G, Shukur G (2005) Choosing ridge parameter for regression problem. Communications in Statistics – Theory and Methods 34: 1177–1182. Link: https://goo.gl/Nu1Xs4

15. Norfiza A, Maizah HA, Robin A (2006) A comparative study on some methods for handling multicollinearity problems. Mathematika 22: 109–119. Link: https://goo.gl/Tlyeqj

16. Alkhmisi MA, Shukur G (2007) A Monte Carlo study of recent ridge parameter methods. Communications in Statistics – Simulation and Computation 36: 535–547. Link: https://goo.gl/Mv2FY

17. Mardikyan S, Cetin E (2008) Efficient choice of biasing constant for ridge regression. Int. J. Contemp. Math. Sciences, 3: 527–536. Link: https://goo.gl/oG9sH

18. Prago-Alejo RJ, Torre-Trevino LM, Pina-Monarrez MR (2008) Optimal determination of k constant of ridge regression using a simple genetic algorithm. Electronics robotics and Automotive Mechanics Conference. Link: https://goo.gl/uPyg0B

19. Dorugade AV, Kashid DN (2010) Alternative method for choosing ridge parameter for regression. Applied Mathematical Sciences 4: 447–456. Link: https://goo.gl/E7MYY5

20. Al-Hassan Y (2010) Performance of new ridge regression estimators. Journal of the Association of Arab Universities for Basic and Applied Science 9: 23–26. Link: https://goo.gl/zTjoEe

21. Ahn JJ, Byun HW, Oh KJ, Kim TY (2012) Using ridge regression with genetic algorithm to enhance real estate appraisal forecasting. Expert Systems with Applications 39: 8369–8379. Link: https://goo.gl/7M0Udi

22. Uslu VR, Egiroglu E, Bas E (2014) Finding optimal value for the shrinkage parameter in ridge regression via particle swarm optimization. American Journal of Intelligent Systems 4: 142-147. Link: https://goo.gl/U06GuG

23. Chitsaz S, Ahmed SE (2012) Shrinkage estimation for the regression parameter matrix in multivariate regression model. Journal of Statistical Computation and Simulation 82: 309-323. Link: https://goo.gl/tIDz2U

24. Firinguetti L (1997) Ridge regression in the context of a system of seemingly unrelated regression equations. Journal of Statistical Computation and Simulation 56: 145-162. Link: https://goo.gl/wWRQue

25. Halawa AM, El Bassiouni MY (2000) Tests of regression coefficients under ridge regression models. Journal of Statistical Computation and Simulation 65: 341-356. Link: https://goo.gl/LUQbW

26. Dorugade AV, Kashid DN (2010) Variable selection in linear regression based on ridge estimator. Journal of Statistical Computation and Simulation 80: 1211-1224. Link: https://goo.gl/A0WJa7

27. Golam Kibria BM (2004) Performance of the shrinkage preliminary test ridge regression estimators based on the conflicting of W, LR and LM tests, Journal of Statistical Computation and Simulation 74: 793-810. Link: https://goo.gl/TMqLyE

28. Roozbeh M, Arashi M, Nourmand HA (2011) Ridge regression methodology in partial linear models with correlated errors. Journal of Statistical Computation and Simulation 81: 517-528. Link: https://goo.gl/Y24Nz

29. Simpson JR, Montgomery DC (1996) A biased-robust regression technique for the combined outlier-multicollinearity problem. Journal of Statistical Computation and Simulation 56: 1-22. Link: https://goo.gl/ogKF7z

30. Uzuke CA, Mbegbu JI, Nwosu CR (2015) Performance of kibria, khalaf and shurkur’s methods when the eigenvalues are skewed. Communications in Statistics – Simulation and Computation Link: https://goo.gl/Vdvglo

31. Wong KY, Chiu SN (2015) an iterative approach to minimize the mean squared error in ridge regression. Computational Statistics 30: 625-639. Link: https://goo.gl/rdHIKp

32. Dorugade AV (2014) new ridge parameters for ridge regression. Journal of the Association of Arab Universities for Basic and Applied Sciences 15: 94-99. Link: https://goo.gl/wpNPWF

33. Khalaf G (2013) An optimal estimation for the ridge regression parameter. Journal of Fundamental and Applied Statistics 5: 11-19. Link: https://goo.gl/boAbde

34. Muniz G, Golam Kibria BM, Månsson K, Ghazi S (2012) On developing ridge regression parameters: a graphical investigation. Sort-Statistics and Simulation 30: 621-630. Link: https://goo.gl/wqCKh

35. Muniz G, Golam Kibria BM (2009) On some ridge regression estimators: an empirical comparisons. Communications in Statistics - Simulation and Computation 38: 621-630. Link: https://goo.gl/wqCKh

36. Nomura M (1988) On the almost unbiased ridge regression estimation. Communications in Statistics – Simulation and Computation 17: 729–743. Link: https://goo.gl/SX03MM

37. Montgomery DC, Peck EA, Vining GG (2006) Introduction to Linear Regression Analysis. John Wiley and Sons. Link: https://goo.gl/MT3qXY

38. Batah FS, Ramnathan T, Gore SD (2008) The efficiency of modified jackknife and ridge type regression estimators: a comparison. Surveys in Mathematics and Its Applications 3: 111–122. Link: https://goo.gl/bwKx8c

39. Hoerl AE, Kennard RW (1976) Ridge regression: iterative estimation of W, LR and LM statistics. Technometrics 18: 425–437. Link: https://goo.gl/sjCZrw

40. Hoerl AE, Kennard RW (1970) Ridge Regression: Applications
41. Firinguetti L (1999) A generalized ridge regression estimator and its finite sample properties. Communications in Statistics-Theory and Methods 28: 1217-1229. Link: https://goo.gl/VTzhib

42. Kennedy J, Eberhart R (1995) Particle swarm optimization. In Proceedings of IEEE International Conference on Neural Networks, Piscataway, NJ, USA, IEEE Press, 1942–1948

43. Chatterjee S, Hadi (2006) A Regression Analysis by Example. John Wiley and Sons. Link: https://goo.gl/Dx6iqn

44. Gibbons DG (1981) A simulation study of some ridge estimators. Journal of the American Statistical Association 76:131–139. Link: https://goo.gl/XMzqJs