Simulation Study to Verify the Appropriate k Value for Ridge Regression in Two-variable Regression Model

Hanan Duzan* and Nurul Sima Binti Mohamad Shariff
Faculty of Science and Technology, Universiti Sains Islam Malaysia (USIM), Malaysia;
Hananduzan@yahoo.com

Abstract
This study investigates the problem of using Ordinary Least Squares (OLS) estimators in the presence of multicollinearity in regression analysis. As an alternative of OLS is ridge regression, which it is believed to be superior to least-squares regression in the presence of multicollinearity. The robustness of this method is investigated and comparison is made with the least squares method via simulation studies. Our results have shown that the system stabilizes in a region of k, which k is a positive quantity less than one and whose values depend on the degree of correlation between the independent variables. The results illustrate that k is a non-linear function of the correlation between the independent variables (r_{12}).

Keywords: Least Squares Method, Linear Models, Multicollinearity, Ridge Regression

1. Introduction

The Ordinary Least Squares (OLS) estimator is the Best Linear Unbiased Estimator (BLUE) that where it can be used to investigate the linear relationships between the variables of interest. The model \( Y = X \beta + \varepsilon, \varepsilon \sim N(0, \sigma^2 I) \) and it is assumed, that \( \beta \) is linear (each of its elements is a linear function of \( y \), the dependent variable). \( E(\hat{\beta}) = \beta \). The expected value of the slope estimates \( \hat{\beta} \) is the true \( \beta \) and in the class of linear unbiased estimators of \( \beta \) the OLS estimator, \( \hat{\beta} \) with minimum variance. The OLS method has some attractive statistical properties under the following assumptions:

- \( E(\varepsilon) = 0 \) where \( \varepsilon \) and \( I \) are \((n \times 1)\) column vectors.
- \( E(\varepsilon \varepsilon') = \sigma^2 I \), where I is \((n \times n)\), i.e. the identity matrix.
- The \((n \times p)\) matrix \( X \) is non-stochastic.
- The rank of \( X \) equals to the number of columns, \( C \) in \( X \) and \( C \) is less than the number of observations \( n \).

Multicollinearity is very high-strength correlations, corresponding to singularity, among the independent variables. This phenomenon commonly occurs when a large number of independent variables are incorporated in a regression model. High-strength correlations are encountered when measuring similar dimensions and/or concepts of a phenomenon. Multicollinearity is not the only violation of the OLS assumptions. However, an accurate multicollinearity violates the assumption that \( X \) matrix is given the highest rank, which makes the OLS impossible. When a model does not reach the peak, which is the inverse of \( X \) that cannot be defined; an infinite number of least squares solutions is obtained.

Multicollinearity has several manifestations, including: 1. Small changes in the data can produce wide swings in the parameter estimates; 2. Coefficients can have high standard errors and low significance, even though they may be jointly significant and the coefficient of determination, \( R^2 \), for the regression can be quite
high; and 3. Coefficients may have the wrong sign or implausible magnitude. Multicollinearity does increase the standard error of the coefficients and the increased error means that the coefficient for the particular independent variable may not be close to 0. On the other hand, multicollinearity with a low standard error can give a significant coefficient and the researcher may not come to a conclusion with null findings.

In summary, the multicollinearity misleadingly inflates the standard error in an excessive amount. In such case, the coefficient may provide high estimates of changes in the multiple regressions when only low changes can be seen in the model or the data. Multicollinearity does not reduce the predictive power or reliability of the model as a whole but it only affects the calculations related to an individual predictor. A multiple regressions model with correlated predictors indicates good combination of the entire bundle of predictors which estimate the outcome variable. However, reliable results cannot be based on an individual predictor or on a set of predictors that are redundant. A high degree of multicollinearity can prevent the computer system from performing a matrix inversion while computing the regression coefficient or it can result in an inaccurate inversion. It is noted that in discussions of the assumptions underlying regression analyses such as OLS the phrase no multicollinearity is used sometimes to refer to absence of perfect multicollinearity, which is an expression of accurate (non-stochastic) linear relations among the regression model predictors.

Ridge regression is a technique for analyzing multiple regression models, which may be exposed to the multicollinearity problem. The OLS regression technique provides unbiased estimates, but their variances are so large that they can be far from the actual value. By adding a degree of bias to the regression estimates, ridge regression reduces the standard errors; the net effect can be highly reliable estimates of the target errors. There are a number of common biased regression techniques but the most popular one of which is ridge regression. The actual definition of ridge regression is the existence of accurate linear relationships between the variables of a regression model which we can notice. In order to identify the main predictors, it is extremely vital to deal with multicollinearity where the impact is great and the interpretation, the amendments, and the analysis occur in all the linear models. The main purpose of this study is to discuss the shortcomings OLS regression when estimating the regression coefficients in the presence of multicollinearity and to present the ridge estimator family as an alternative to the OLS procedure.

Several authors have suggested various estimation methods to reduce the biasness problem.

When developed the ridge regression technique they suggested that this method, which is also referred to as the ridge trace, can be used to solve the biasness problem. This ridge trace is a plot which illustrates the ridge regression coefficients as the main function of k. By using this ridge trace, the analyst may give a value of k at which the regression coefficients can be stabilized. Often, the regression coefficients are varied widely to get a small value of k and then they are stabilized. Choosing the smallest possible value of k (which introduces the smallest bias) ensures that the regression coefficients can remain stable. It is noted that the increasing value of k will finally drive the regression coefficients to zero. This study investigates the shortcomings of using the OLS estimators in the presence of multicollinearity with ridge regression presented as an alternative approach. The properties of ridge regression are discussed in detail and are based on the results obtained by the authors of the recent study have argued that this method is superior to the least-squares estimator in the presence of multicollinearity.

### 2. Methodology

#### 2.1 Least-Squares Estimation

Consider the following P-variable regression model

\[ Y = X\beta + \epsilon, \epsilon \sim N(0, \sigma^2I) \]  \( (1) \)

where:
- \( Y: (n \times 1) \) column vector of observations on the dependent variable \( y \);
- \( X: (n \times p) \) matrix giving \( n \) observations on \( p-1 \) variables \( X_2 \) to \( X_p \), the first column of 1’s representing the intercept term;
- \( \beta: (p \times 1) \) column vector of the unknown parameters;
- \( \epsilon: (n \times 1) \) column vector of \( n \) disturbance terms.

The least–square estimator of \( \beta \) is given by

\[ \hat{\beta} = (X'X)^{-1}X'Y \]  \( (2) \)

In model (1), the residual, \( \epsilon \), is assumed to be identically, independently, and normally distributed with a mean of zero and constant variance.
The variance covariance matrix of $\hat{\beta}$ is
\[ \text{var}(\hat{\beta}) = \sigma^2 (X'X)^{-1} \] (3)

2.1.1 Alternative Variant of the Model

The $X$ scaled variables are assumed such that $X'X$ it has the form of a correlation matrix. To recognize this, consider the following multiple linear regression model.
\[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i, \quad i = 1, 2, \ldots, n \] (4)
\[ Y_j = \tilde{\beta}_0 + \beta_1 (X_{j1} - \bar{X}_1) + \beta_2 (X_{j2} - \bar{X}_2) + \varepsilon_j \] (5)

By subtracting (4) with (5), we have:
\[ \Rightarrow Y_i - \bar{Y} = \beta_1 (X_{i1} - \bar{X}_1) + \beta_2 (X_{i2} - \bar{X}_2) + \varepsilon_i \] (6)

where,
\[ \tilde{\beta}_0 = \beta_0 + \beta_1 \bar{X}_1 + \beta_2 \bar{X}_2 \]

The variables are then standardized to:
\[ \frac{Y_i - \bar{Y}}{S_y}, \quad \frac{X_{i1} - \bar{X}_1}{S_1} \quad \text{and} \quad \frac{X_{i2} - \bar{X}_2}{S_2} \] (7)
\[ S_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \quad S_1^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_{i1} - \bar{X}_1)^2 \]
\[ \text{and} \quad S_2^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_{i2} - \bar{X}_2)^2 \] (8)

Define the following simple function of the standardized variables:
\[ Y^*_i = \frac{1}{\sqrt{n-1}} \frac{(Y_i - \bar{Y})}{s_y} \] (9)
\[ X^*_{i1} = \frac{1}{\sqrt{n-1}} \frac{(X_{i1} - \bar{X}_1)}{s_1} \] (10)
\[ \text{and} \quad X^*_{i2} = \frac{1}{\sqrt{n-1}} \frac{(X_{i2} - \bar{X}_2)}{s_2} \]

Therefore, the parameterized model with the transformed variables corresponding to model (2) is given by:
\[ Y_i^* = \beta_1^* X_{i1}^* + \beta_2^* X_{i2}^* + \varepsilon_i^* \] (11)

Note that,
\[ \beta_1^* = \left( \frac{S_{X_{i1}}}{S_{X_{i1}}} \right) \beta_1, \quad \beta_2^* = \left( \frac{S_{X_{i2}}}{S_{X_{i2}}} \right) \beta_2 \]

Then the least squares estimator of $\beta$ is given by:
\[ \hat{\beta} = (X^{*'}X^*)^{-1} X^{*'}Y^* \]
\[ X^* = \begin{bmatrix} x_{i1}^* & x_{i2}^* \\ \vdots & \vdots \\ x_{n1}^* & x_{n2}^* \end{bmatrix} \]

(13)

The $X^*$ matrix in the model can be written as follows:
\[ \Rightarrow (X^{*'}X^*) = \begin{bmatrix} \sum_{i=1}^{n} (x_{i1}^*)^2 & \sum_{i=1}^{n} (x_{i1}^* x_{i2}^*) \\ \sum_{i=1}^{n} (x_{i1}^* x_{i2}^*) & \sum_{i=1}^{n} (x_{i2}^*)^2 \end{bmatrix} \]

(14)

Since,
\[ \sum_{i=1}^{n} (x_{i1}^*)^2 = \sum_{i=1}^{n} (x_{i1} - \bar{x}_{i1})^2 = \frac{n}{n-1} \sum_{i=1}^{n} (x_{i1} - \bar{x}_{i1})^2 = S_1^2 = 1 \] (15)

Similarly for $\sum_{i=1}^{n} (x_{i2}^*)^2 = 1$
\[ \sum_{i=1}^{n} x_{i1}^* x_{i2}^* = \sum_{i=1}^{n} \frac{x_{i1} - \bar{x}_{i1}}{s_1} \frac{x_{i2} - \bar{x}_{i2}}{s_2} \]
\[ = \sum_{i=1}^{n} \frac{(x_{i1} - \bar{x}_{i1})(x_{i2} - \bar{x}_{i2})}{(n-1)s_1s_2} = \eta_{12} \]
\[ = \sqrt{\frac{\sum_{i=1}^{n} (x_{i1} - \bar{x}_{i1})^2 \sum_{i=1}^{n} (x_{i2} - \bar{x}_{i2})^2}{n-1}} \] (16)

where, $\eta_{12}$ is the simple correlation coefficient between $x_i$ and $x_j$. The $X^* X^*$ matrix for the transformed variables can be written as:
\[ (X^{*'}X^*) = \begin{bmatrix} 1 & \eta_{12} \\ \eta_{12} & 1 \end{bmatrix}, \quad -1 \leq \eta_{12} \leq 1 \] (17)

When the number of independent variable is (two) we
Simulation Study to Verify the Appropriate k Value for Ridge Regression in Two-variable Regression Model

\[ C = (X^* X^*)^{-1} = \frac{1}{1 - \frac{2}{12} r_{12}^2} \begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix} \] (18)

Equation (18) illustrates that as \( r_{12}^2 = 1 \) and hence \( \text{Var}(\hat{\beta}_j) = \sigma^2 C_{jj} \rightarrow \infty \) depending on whether \( r_{ij} \rightarrow \pm 1 \). Hence, the standard errors of the least square estimators tend to be large as the degree of collinearity between the explanatory variables increases, i.e. as \( r_{12} \) increases.

2.2 Properties of the Ridge Solution

With \( p \)-variables case, the diagonal elements of \( C = (X^* X^*)^{-1} \) can be written as follows:

\[ c_{jj} = \left(1 - R_j^2\right)^{-1}, \quad j = 1, 2, \ldots, p \]

where, \( R_j^2 \) is the coefficient of determination of the least squares regression of \( X_j \) on the remaining \((p-1)\) regressor variables.

Since \( \text{Var}(\hat{\beta}_j) = \sigma^2 C_{jj} \), then

\[ R_j^2 \rightarrow 1 \quad \text{and} \quad \text{Var}(\hat{\beta}_j) \rightarrow \infty \] (19)

The main properties of the ridge solution are:

- The length of \( \hat{\beta} \) is a decreasing function of \( k \).
- The residuals sum of squares is a monotone which increases as a function of \( k \).
- The ridge estimator, \( \hat{\beta} \), is a linear transformation of the least squares estimator, \( \hat{\beta} \).

\[ \hat{\beta}^* = (X^* X^* + kI)^{-1} (X^* X^*) \hat{\beta} \] (20)

\[ \hat{\beta}^* \] , a biased estimator of \( \hat{\beta} \).

\[ E(\hat{\beta}^*) = (X^* X^* + kI)^{-1} (X^* X^*) \beta = \beta \] (21)

- The covariance of \( \hat{\beta} \) is given by:

\[ \text{cov}(\hat{\beta}^*) = \sigma^2 (X^* X^* + kI)^{-1} (X^* X^*) (X^* X^*) (X^* X^*) + kI)^{-1} \] (22)

The mean square error of \( \hat{\beta} \) is given by

\[ \text{MSE}(\hat{\beta}) = E(\hat{\beta} - \beta)^2 \] (23)

\[ = \sigma^2 \sum_{i=1}^{2} \frac{\lambda_i}{\lambda_i + k} + k^2 \hat{\beta}^* (X^* X^* + kI)^{-2} \hat{\beta}^* \] (24)

\[ = \frac{2}{\sum_{i=1}^{2} \text{var}(\hat{\beta}^*_{i}) + 2 \sum_{i=1}^{2} \text{Bias}(\hat{\beta}^*_{i})} \] (25)

where the first term on the right hand side of equation (25) is the sum of the variance of the estimators and the second term is the sum of squared biases, which is introduced by using \( \hat{\beta}^* \) rather than \( \hat{\beta} \). It can be seen that the sum of variances is a decreasing function of \( k \), while the squared bias is an increasing function of \( k \).

\[ \text{Lim}_{\beta \in \infty} \text{MSE}(\hat{\beta}^*) \rightarrow \infty \], and hence for fixed \( k \), the ridge estimator is not minimax.

If \( \beta^T \beta \) is bounded, then there exists a \( k > 0 \) such that \( \text{MSE}(\hat{\beta}^*) < \text{MSE}(\hat{\beta}) \).

2.3 The Variance Inflation Factor (VIF)

The variance inflation factor can be computed using the equation:

\[ VIF = (1 - R_j^2)^{-1} \] (26)

where, \( R_j^2 \) is the coefficient of determination in the regression of an explanatory variable \( X_j \) on the remaining explanatory variables of the model. If \( X_j \) has a strong linear relation with other explanatory variables, then \( R_j^2 \) will be close to one and VIF values will tend to be very high. However, in the absence of any linear relations among the explanatory variables, \( R_j^2 \) will be zero and the VIF will equal one. It is known that a VIF value greater than one indicates deviation from orthogonally and has tendencies generally, when the VIF > 10, we assume that there exists highly multicollinearity, and that the Sum of Squared Errors (SSE) approaches 1. There always exists a \( k > 0 \), such that \( \hat{\beta}(k) \) has smaller MSE than \( \hat{\beta} \), which means that \( \text{MSE}(\hat{\beta}(k)) < \text{MSE}(\hat{\beta}) \). Further details on this issue have been provided by \( \beta^T \beta \). Finally, if \( R_j^2 < R_j^2 \) for all \( j \) and \( R_j^2 < 0.90 \) which implies that \( VIF = (1 - R_j^2)^{-1} \), then we should not worry about the existence of multicollinearity.

3. The Monte Carlo Design

A simulation study using 1000 samples with \( n = 10 \) was conducted to determine the appropriate value of \( k \) for ridge regression in a \( p \) variable regression model. The performance of the OLS and the different ridge regression estimators are evaluated and compared. Furthermore, a brief description of the factors that vary in our simulation study is discussed in this section.
In most simulation studies\(^1\), the Mean Squared Error (MSE), Variance Inflation Factor (VIF), and \( \mathbf{\hat{\beta}} \) of the proposed ridge estimators are calculated using a fairly low number of explanatory variables two and four are the most common value of \( (p) \). We will choose \( k \) which gives stable values of the estimated parameters with VIF and MSE for different \( k \) values. A linear regression model with correlated independent variables is considered and the different potential \( R_j \) values are computed.

For a number of values of \( k \), including the least-squares solution, when \( k = 0 \). The values of \( x_i \), \( I = 1,2 \) were generated from normal distribution with \( (0, 4) \). For given \( x_1, x_2 \) correlated variables \( 0 \leq r_{12} \leq 1 \), the \( y \) values were generated using a set of predetermined values of parameters, however, only values of \( \varepsilon_i \) to change randomly. The errors \( \varepsilon \) were generated to be \( \varepsilon \sim i.i.d. N(0, \sigma^2) \) independent and identically normally distributed with a mean of zero and variance \( \sigma^2 \).

The true values of the parameters were taken to be: 
\[
\mathbf{\beta} = (0.2, 1.2, 0.8) \quad \text{and} \quad \sigma^2 = 4
\]

One thousand data sets were used in each simulation study. Each data set was fitted by OLS and ridge regression estimation \( \mathbf{\hat{\beta}} \), VIF and SSE for different \( k \) values and different \( r_{12} \) (the correlation between \( X_1 \) and \( X_2 \)) were computed.

The mean of \( \mathbf{\hat{\beta}} \), the ridge estimates, VIF and SSE are given by the following equations:
\[
\begin{align*}
A(\mathbf{\hat{\beta}}_1) &= \frac{\text{total}(\mathbf{\hat{\beta}}_1)}{1000} \\
A(\mathbf{\hat{\beta}}_2) &= \frac{\text{total}(\mathbf{\hat{\beta}}_2)}{1000} \\
AVIF &= \frac{\text{total}(\text{VIF})}{1000} \\
ASSE &= \frac{\text{total}(\text{SSE})}{1000}
\end{align*}
\]

4. The Simulation Study

The simulation results are presented in Tables 1 to 5 and Figures 1 to 9; show that the system stabilized for the various ranges of \( k \) values based on the observed ridge trace, VIF and SSE for selected values of \( r_{12} \). For example, in Table 4, \( k = 0.009 \) is chosen as stable point solution since the smallest value of AVIF is obtained in this experiment.

For this value of \( k \), \( \mathbf{\hat{\beta}}_1 = 0.879848 \) and \( \mathbf{\hat{\beta}}_2 = 0.375681 \) are the estimated values for the parameter. We also note that the least squares estimates (when \( k = 0 \)) are \( \mathbf{\hat{\beta}}_1 = 0.888208 \) and \( \mathbf{\hat{\beta}}_2 = 0.376948 \). These values are very close to the ridge regression estimates because \( x_1 \) and \( x_2 \) have weak dependency (collinear \( r_{12} = 0.100 \)).

Table 1 gives a summary of the appropriate values of \( k \) for ridge regression estimates \( \mathbf{\hat{\beta}}_1 \) and \( \mathbf{\hat{\beta}}_2 \) for different \( r_{12} \) values. From Table 2 and Table 3 we figure out that there is a relationship between \( k \) and \( r_{12} \), and finally yield an appropriate model for this relationship that is \( K = \theta r_{12}^2 + \varepsilon \).

Based on these values (Table 1.), the developed model is: \( \hat{k} = 0.696 r_{12}^2 \).

| \( r_{12}^2 \) | \( k \) | Correlation | \( |k| \) | \( k \) | Correlation |
|---|---|---|---|---|---|
| 0.009 | 0.009 | 0.245 | 0.5530 | 0.014 | 0.1218 |
| 0.026 | 0.1674 | 0.266 | 0.5818 | 0.063 | 0.2623 |
| 0.083 | 0.3023 | 0.270 | 0.5877 | 0.084 | 0.3020 |
| 0.102 | 0.3378 | 0.301 | 0.6270 | 0.113 | 0.3569 |
| 0.119 | 0.3666 | 0.323 | 0.6540 | 0.142 | 0.4040 |
| 0.154 | 0.4239 | 0.405 | 0.7550 | 0.155 | 0.4251 |
| 0.157 | 0.4282 | 0.432 | 0.7876 | 0.207 | 0.5021 |
| 0.217 | 0.1513 | 0.465 | 0.8266 | 0.221 | 0.5190 |
| 0.222 | 0.5225 | 0.533 | 0.9030 | 0.226 | 0.5282 |
| 0.243 | 0.5508 | 0.605 | 0.9860 | 0.253 |

Table 2. No constant

| Predictor | Coefficient | Standard Deviation | T | P |
|---|---|---|---|---|
| \( r_{12}^2 \) | 0.696 | 0.01097 | 63.47 | 0.000 |

Table 3. Analysis of Variance

| Source | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| Regression | 1 | 3.4759 | 3.4759 | 4028.56 | 0.00 |
| Error | 38 | 0.03228 | 0.0009 | 0.0009 | 0.0009 |
| Total | 39 | 3.5087 | 0.0000 | 0.0000 | 0.0000 |

\( R^2=99.0\% \) (where \( R^2 \) is the coefficient of determination).
Table 4. Means of $\hat{\beta}_1, \hat{\beta}_2$ VIF and SSE when $r_{12} = 0.100$

| $K$ | $A(\hat{\beta}_1)$ | $A(\hat{\beta}_2)$ | $\text{ASSE}$ | $\text{AVIF}$ |
|-----|------------------|------------------|---------------|---------------|
| 0.000 | 0.888208 | 0.376948 | 0.001176 | 1.01018 |
| 0.002 | 0.886341 | 0.376762 | 0.002991 | 1.00812 |
| 0.004 | 0.884605 | 0.376176 | 0.004880 | 1.00605 |
| 0.006 | 0.882241 | 0.376900 | 0.006763 | 1.00402 |
| 0.007 | 0.882008 | 0.375437 | 0.007651 | 1.00299 |
| 0.009 | 0.879848 | 0.375681 | 0.009611 | 1.00098 |
| 0.010 | 0.879023 | 0.375505 | 0.010384 | 0.99995 |
| 0.011 | 0.878329 | 0.374856 | 0.011334 | 0.99896 |
| 0.015 | 0.875952 | 0.371740 | 0.015010 | 0.99493 |
| 0.120 | 0.795372 | 0.346017 | 0.101622 | 0.90008 |
| 0.130 | 0.789103 | 0.342437 | 0.109112 | 0.89198 |
| 0.140 | 0.782221 | 0.340305 | 0.116519 | 0.88403 |
| 0.150 | 0.776126 | 0.336968 | 0.123747 | 0.87622 |

Table 5. Means of $\hat{\beta}_1, \hat{\beta}_2$ VIF and SSE when $r_{12} = 0.855$

| $K$ | $A(\hat{\beta}_1)$ | $A(\hat{\beta}_2)$ | $\text{ASSE}$ | $\text{AVIF}$ |
|-----|------------------|------------------|---------------|---------------|
| 0.000 | 0.500387 | 0.537794 | 0.000140 | 3.71807 |
| 0.100 | 0.481173 | 0.503903 | 0.051351 | 2.9665 |
| 0.250 | 0.450497 | 0.464390 | 0.118970 | 1.50341 |
| 0.350 | 0.431209 | 0.442218 | 0.158937 | 1.23689 |
| 0.487 | 0.406795 | 0.415509 | 0.208145 | 1.00464 |
| 0.490 | 0.406258 | 0.414986 | 0.209172 | 1.00063 |
| 0.491 | 0.406064 | 0.414833 | 0.209504 | 0.99932 |
| 0.495 | 0.405502 | 0.414003 | 0.210843 | 0.99402 |
| 0.500 | 0.404834 | 0.412931 | 0.212518 | 0.98753 |
| 0.510 | 0.402953 | 0.411350 | 0.215854 | 0.97479 |
| 0.520 | 0.401436 | 0.409442 | 0.219154 | 0.96241 |
| 0.529 | 0.399970 | 0.407849 | 0.222093 | 0.95159 |
| 0.560 | 0.395000 | 0.402447 | 0.232088 | 0.91626 |

Table 6. Means of $\hat{\beta}_1, \hat{\beta}_2$ of VIF and SSE when $r_{12} = 0.855$

| $K$ | $A(\hat{\beta}_1)$ | $A(\hat{\beta}_2)$ | $\text{ASSE}$ | $\text{AVIF}$ |
|-----|------------------|------------------|---------------|---------------|
| 0.000 | 0.469896 | 0.533434 | 0.000135 | 36.7096 |
| 0.119 | 0.469831 | 0.476785 | 0.056737 | 4.0049 |
| 0.489 | 0.401649 | 0.403472 | 0.197733 | 1.1966 |
| 0.602 | 0.384331 | 0.385640 | 0.232758 | 1.0052 |
| 0.603 | 0.384091 | 0.385583 | 0.230353 | 1.0038 |
| 0.605 | 0.383811 | 0.385268 | 0.233649 | 1.0011 |
| 0.606 | 0.383638 | 0.385144 | 0.233944 | 0.9997 |
| 0.607 | 0.383518 | 0.384966 | 0.234243 | 0.9983 |
| 0.608 | 0.383354 | 0.384838 | 0.234531 | 0.9970 |
| 0.627 | 0.380561 | 0.382045 | 0.240096 | 0.9717 |
| 0.631 | 0.380008 | 0.381429 | 0.241264 | 0.9666 |
| 0.733 | 0.365829 | 0.367049 | 0.269719 | 0.8535 |
| 0.734 | 0.365643 | 0.366964 | 0.269991 | 0.8525 |

Figure 1. A plot of $K$ vs. $A(\hat{\beta}_1), A(\hat{\beta}_2)$ VIF and SSE from using the simulation study when $r_{12} = 0.100$. 
Figure 2. A plot of $K$ vs. $A(\beta_1^r), A(\beta_2^r)$ VIF and SSE from using the simulation study when $r_{12} = 0.855$.

Figure 3. A plot of $K$ vs. $A(\beta_1^r), A(\beta_2^r)$ VIF and SSE from using the simulation study when $r_{12} = 0.986$. 
5. Discussion and Conclusion

The main goal of this study is to investigate the appropriate value of $k$ for ridge regression in $p$-variable regression model. Since it is not possible to do that mathematically, a simulation study is conducted to study the behavior of ridge regression and one thousand random data sets are used in each simulation study. The $p$-variable linear regression model are fitted by OLS. In each simulation study, 1000 ridge regression estimate of $\hat{\beta}$, Variance Inflation Factor (VIF) and Sum Squared Error (SSE), for different $k$ and $r_{12}$ are computed. Based on the results, the relationship between $k$ and $r_{12}$ are obtained and yielded.

$$\hat{k} = 0.696r_{12}^2$$

From practical point of view, the ridge regression can be used to solve this problem in regression analysis when the multicollinearity occurs. The appropriate value of $k$ can be chosen according to Ridge trace.

6. References

1. Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics; 2000; 42(1):80–6.
2. Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics; 1970; 12(1):55–67.
3. Hoerl AE, Kennard RW. Baldwin communications in statistics-theory and methods. 1975; 4(2):105–23.
4. Alkhamisi M, Khalaf G, Shukur G. Some modifications for choosing ridge parameters. Comm Stat Theor Meth. 2006; 35(11).
5. Alkhamisi MA, Shukur G. Developing ridge parameters for SUR model. Comm Stat Theor Meth. 2008; 37(4):544–56.
6. El-Dereny M, Rashwan NI. Solving multicollinearity problem using ridge regression models. International Journal of Contemporary Mathematical Sciences. 2011; 6:585–600.
7. Judge GG. Introduction to Theory and Practice of Econometrics. New York: John Wiley and Sons; 1988.
8. Gujarati DN. Gujarati. Basic econometrics. Tata McGraw-Hill Education; 1995.
9. Gurber HJ. Improving Efficiency by Shrinkage. New York: Marcel Dekker; 1998.
10. Pasha GR, Shah MA. Application of ridge regression to multicollinear data. Journal of Research Science. 2004; 15:97–106.
11. Khalaf G. J Mod Appl Stat Meth. 2012; 11(2):15.
12. Jensen DR, Ramirez DE. Variations on ridge traces in regression. Comm Stat Simulat Comput. 2012; 41(2):265–78.
13. Singh R. A survey of ridge regression for improvement over ordinary least squares. The IUP Journal of Computational Mathematics. 2010; 3(4):54–74.