Journal of Modern Applied Statistical Methods

Volume 15 | Issue 1 Article 37

5-1-2016

Solution to the Multicollinearity Problem by Adding some Constant to the Diagonal

Hanan Duzan
Universiti Sains Islam Malaysia, hananduzan@yahoo.com

Nurul Sima Binti Mohamaed Shariff
Universiti Sains Islam Malaysia

DOI: 10.22237/jmasm/1462077360

Solution to the Multicollinearity Problem by Adding some Constant to the Diagonal

Cover Page Footnote

The work in this article has not been published before,that it is not under consideration for publication anywhere else ;that it's publication has been approved by all co-authors

Journal of Modern Applied Statistical Methods
May 2016, Vol. 15, No. 1, 752-773.

ISSN 1538 − 9472

Solution to the Multicollinearity Problem by Adding some Constant to the Diagonal

Hanan Duzan
Universiti Sains Islam Malaysia
Nilai, Malaysia

Nurul Sima Binti Mohamaed Shariff
Universiti Sains Islam Malaysia
Nilai, Malaysia

Ridge regression is an alternative to ordinary least-squares (OLS) regression. 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 through simulation studies. Our results show that the system stabilizes in a region of *k*, where *k* is a positive quantity less than one and whose values depend on the degree of correlation between the independent variables. The results also illustrate that *k* is a linear function of the correlation between the independent variables.

Keywords: Linear models, multicollinearity, least squares method, ridge regression

Introduction

The ordinary least squares (OLS) estimator is the best linear unbiased estimator (BLUE). It can be used to investigate the linear relationships between the variables of interest. The model is $\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim \mathrm{N}\left(\mathbf{o}, \sigma^2\mathbf{I}\right)$ and it is assumed that $\boldsymbol{\beta}$ is linear (each of its elements is a linear function of *y*, the dependent variable). The parameter β in OLS has the properties of being (i) unbiased where $\mathrm{E}\left(\hat{\beta}\right) = \beta$ is the expected value of the slope estimates of β, which is the true β; and (ii) consistent, where the estimator produces the minimum variance. The OLS method has some attractive statistical properties under the following assumptions:

i. $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, where $\boldsymbol{\varepsilon}$ and $\mathbf{0}$ are $(n \times 1)$ column vectors
ii. $\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is the $(n \times n)$ identity matrix
iii. The $(n \times p)$ matrix $\mathbf{X}$ is non-stochastic
iv. The rank of $\mathbf{X}$ is equal to the number of columns, *C*, in $\mathbf{X}$, and *C* is less than the number of observations, *n*

Hanan Duzan is Faculty of Science and Technology. Email at: hananduzan@yahoo.com. Dr. Shariff is Faculty of Science and Technology..

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 the matrix **X** 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: (a) small changes in the data can produce wide swings in the parameter estimates; (b) 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 (c) coefficients may have the wrong sign or implausible magnitude (Greene, 2000, p. 256). Multicollinearity increases 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, a 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 regression 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 that 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 parameters. There is a number of common biased regression techniques, the most popular 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 of OLS regression when estimating the regression coefficients in the presence of muticollinearity, 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 Hoerl and Kennard (1970a) 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 to 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. Most of the later efforts in this area have concentrated on estimating the value of the ridge parameter k. Many different techniques for estimating k have been proposed by different researchers (e.g., Hoerl & Kennard, 1970a; b; Hoerl, Kennard, & Baldwin, 1975; McDonald & Galarneau, 1975; Lawless & Wang, 1976; Dempster, Laird, & Rubin, 1977; Khalaf & Shukur, 2005; Alkhamisi, Khalaf, & Shukur, 2006; Alkhamisi & Shukur, 2008; Muniz & Kibria, 2009; Dorugade & Kashid, 2010; Jensen & Ramirez, 2012). 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 El-Dereny and Rashwan (2011), who have argued that this method is superior to the least-squares estimator in the presence of multicollinearity.

Methodology

Least-Squares Estimation

Consider the following P-variable regression model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim \mathrm{N}\left(\mathbf{o}, \sigma^2\mathbf{I}\right) \tag{1}$$

where $\mathbf{Y}$ is an $(n \times 1)$ column vector of observations on the dependent variable y; $\mathbf{X}$ is an $(n \times p)$ matrix giving n observations on $p - 1$ variables, X_2 to X_p, the first column of 1s representing the intercept term; $\boldsymbol{\beta}$ is a $(p \times 1)$ column vector of the unknown parameters; and $\boldsymbol{\varepsilon}$ is an $(n \times 1)$ column vector of n disturbance terms.

The least-square estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \tag{2}$$

In model (1), the residual, $\boldsymbol{\varepsilon}$, is assumed to be identically, independently, and normally distributed with a mean of zero and constant variance.

The variance covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$\mathrm{var}\left(\hat{\boldsymbol{\beta}}\right) = \sigma^2\left(\mathbf{X}\mathbf{X}\right)^{-1} \tag{3}$$

Alternative Variant of the Model

The **X**-scaled variables are assumed such that **X'X** has the form of a correlation matrix. To recognize this, consider the following multiple linear regression model:

$$Y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \varepsilon_i, i = 1, 2, \ldots, p \ , \tag{4}$$

$$\mathbf{Y} = \beta_0 + \beta_1 \mathbf{X}_1 + \beta_2 \mathbf{X}_2 + \ldots + \beta_p \mathbf{X}_p \ . \tag{5}$$

By subtracting (5) from (4), we get

$$Y_i - \mathbf{Y} = \beta_1\left(X_{i1} - \mathbf{X}_1\right) + \beta_2\left(X_{i2} - \mathbf{X}_2\right) + \ldots + \beta_p\left(X_{ip} - \mathbf{X}_p\right) + \varepsilon_i \ . \tag{6}$$

The variables are then standardized to

$$\frac{Y_i - \mathbf{Y}}{S_Y}, \quad \frac{X_{ij} - \mathbf{X}_j}{S_j}, \ j = 1, 2, \ldots, p \ , \tag{7}$$

$$S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \mathbf{Y}\right)^2, \quad S_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_{ij} - \mathbf{X}_j\right)^2,\ j = 1,2,\ldots,p\ . \tag{8}$$

Define the following simple function of the standardized variables:

$$Y_i^* = \frac{1}{\sqrt{n-1}}\frac{\left(Y_i - \mathbf{Y}\right)}{S_Y} \tag{9}$$

$$X_{ij}^* = \frac{1}{\sqrt{n-1}}\frac{\left(X_{ij} - \mathbf{X}_j\right)}{S_j},\ j = 1,2,\ldots,p \tag{10}$$

Therefore, the parameterized model with the transformed variables corresponding to model (1) is given by

$$Y_i^* = \beta_1^* X_{i1}^* + \beta_2^* X_{i2}^* + \ldots + \beta_p^* X_{ip}^* + \varepsilon_i^*\ . \tag{11}$$

Note that

$$\beta_j^* = \frac{\beta_j S_j}{S_Y},\ j = 1,2,\ldots,p\ .$$

Then the least squares estimator of **β** is given by

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\mathbf{Y}^*\ . \tag{12}$$

The *x** matrix in the model can be written as follows:

$$\mathbf{X}^* = \begin{bmatrix} x_{11}^* & x_{12}^* & \cdots & x_{1p}^* \\ x_{21}^* & x_{11}^* & \cdots & x_{2p}^* \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1}^* & x_{n2}^* & \cdots & x_{np}^* \end{bmatrix} \tag{13}$$

so that

$$\left(\mathbf{X}^{*\prime}\mathbf{X}^{*}\right)=\begin{bmatrix} \sum_{i=1}^{n} x_{i1}^{*2} & \sum_{i=1}^{n} x_{i1}^{*} x_{i2}^{*} & \cdots & \sum_{i=1}^{n} x_{i1}^{*} x_{ip}^{*} \\ \sum_{i=1}^{n} x_{i1}^{*} x_{i2}^{*} & \sum_{i=1}^{n} x_{i2}^{*2} & \cdots & \sum_{i=1}^{n} x_{i2}^{*} x_{ip}^{*} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_{i1}^{*} x_{ip}^{*} & \sum_{i=1}^{n} x_{i2}^{*} x_{ip}^{*} & \cdots & \sum_{i=1}^{n} x_{ip}^{*2} \end{bmatrix} \tag{14}$$

Since

$$\sum_{i=1}^{n} x_{ij}^{*2}=\sum_{i=1}^{n}\left(\frac{x_{ij}-\mathbf{x}_{j}}{\mathbf{S}_{j}\sqrt{n-1}}\right)^{2}=\frac{\sum_{i=1}^{n}\left(x_{ij}-\mathbf{x}_{j}\right)^{2}}{n-1}\div\mathbf{S}_{j}^{2}=1,\ j=1,2,\ldots,p \tag{15}$$

and

$$\begin{aligned}\sum_{i=1}^{n} x_{ij}^{*} x_{ik}^{*} &= \sum_{i=1}^{n}\left(\frac{x_{ij}-\mathbf{x}_{j}}{S_{j}\sqrt{n-1}}\right)\left(\frac{x_{ik}-\mathbf{x}_{k}}{S_{k}\sqrt{n-1}}\right) \\ &= \sum_{i=1}^{n}\frac{\left(x_{ij}-\mathbf{x}_{j}\right)\left(x_{ik}-\mathbf{x}_{k}\right)}{\left(n-1\right)S_{j}S_{k}} \\ &= \frac{\sum_{i=1}^{n}\left(x_{ij}-\mathbf{x}_{j}\right)\left(x_{ik}-\mathbf{x}_{k}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_{ij}-\mathbf{x}_{j}\right)^{2}\sum_{i=1}^{n}\left(x_{ik}-\mathbf{x}_{k}\right)^{2}}}=r_{jk},\ j,k=1,2,\ldots,p,\ j\neq k\end{aligned} \tag{16}$$

where r_{jk} is the simple correlation coefficient between $\mathbf{X}_j$ and $\mathbf{X}_K$, then the matrix for the transformed variables can be written as

$$\left(\mathbf{X}^{*\prime}\mathbf{X}^{*}\right)=\begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{12} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{1p} & r_{2p} & \cdots & 1 \end{bmatrix},\ -1\le r_{ij}\le 1,\ i\neq j\ . \tag{17}$$

When the number of independent variables is two, we have

$$C = \left(x^{*\prime} x^{*} \right)^{-1} = \frac{1}{1 - r_{12}^2} \begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix} . \tag{18}$$

For three independent variables, Equation (17) is then replaced by

$$\mathbf{C} = \left(\mathbf{X}^{*\prime} \mathbf{X}^{*} \right)^{-1} = \begin{bmatrix} \frac{1}{1 - R_{1(23)}^2} & \frac{-r_{12} + r_{13} r_{23}}{\left(1 - r_{23}^2\right)\left(1 - R_{1(23)}^2\right)} & \frac{-r_{13} + r_{12} r_{23}}{\left(1 - r_{23}^2\right)\left(1 - R_{1(23)}^2\right)} \\ \frac{-r_{12} + r_{13} r_{23}}{\left(1 - r_{23}^2\right)\left(1 - R_{1(23)}^2\right)} & \frac{1}{1 - R_{2(13)}^2} & \frac{-r_{23} + r_{12} r_{13}}{\left(1 - r_{13}^2\right)\left(1 - R_{2(13)}^2\right)} \\ \frac{-r_{13} + r_{12} r_{23}}{\left(1 - r_{23}^2\right)\left(1 - R_{1(23)}^2\right)} & \frac{-r_{23} + r_{12} r_{13}}{\left(1 - r_{13}^2\right)\left(1 - R_{2(13)}^2\right)} & \frac{1}{1 - R_{3(12)}^2} \end{bmatrix} \tag{19}$$

where

$$R_{1(23)}^2 = \frac{r_{12}^2 + r_{13}^2 - 2 r_{12} r_{13} r_{23}}{1 - r_{23}^2} \ .$$

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

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

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 $\mathrm{Var}\left(\hat{\beta}_j\right) = \sigma^2 C_{jj}$,

$$R_j^2 \to 1, \quad \mathrm{Var}\left(\hat{\beta}_j\right) \to \infty$$

Properties of the Ridge Solution

The main properties of the ridge solution are:

i. The length of $\hat{\boldsymbol{\beta}}^*$ is a decreasing function of *k*

ii. The residual sum of squares is a monotone which increases as a function of k

iii. The ridge estimator, $\hat{\boldsymbol{\beta}}^*$, is a linear transformation of the least squares estimator $\hat{\boldsymbol{\beta}}$

$$\hat{\boldsymbol{\beta}}^* = \left(\mathbf{X}^{*\prime}\mathbf{X}^* + k\mathbf{I}\right)^{-1}\left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)\hat{\boldsymbol{\beta}} \tag{20}$$

iv. $\hat{\boldsymbol{\beta}}^*$ is a biased estimator of $\boldsymbol{\beta}$.

$$\mathrm{E}\left(\hat{\boldsymbol{\beta}}^*\right) = \left(\mathbf{X}^{*\prime}\mathbf{X}^* + k\mathbf{I}\right)^{-1}\left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)\boldsymbol{\beta} \neq \boldsymbol{\beta} \tag{21}$$

v. The covariance of $\hat{\boldsymbol{\beta}}^*$, $k > 0$, is given by

$$\mathrm{cov}\left(\hat{\boldsymbol{\beta}}^*\right) = \sigma^2\left(\mathbf{X}^{*\prime}\mathbf{X}^* + k\mathbf{I}\right)^{-1}\left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)\left(\mathbf{X}^{*\prime}\mathbf{X}^* + K\mathbf{I}\right)^{-1} \tag{22}$$

vi. The mean square error (MSE) of $\hat{\boldsymbol{\beta}}^*$ is given by

$$\mathrm{MSE}\left(\hat{\boldsymbol{\beta}}^*\right) = \mathrm{E}\left(\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)'\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)\right) \tag{23}$$

$$= \sigma^2\sum_{i=1}^{2}\frac{\lambda_i}{\left(\lambda_i + k\right)} + k^2\boldsymbol{\beta}'\left(\mathbf{X}^{*\prime}\mathbf{X}^* + k\mathbf{I}\right)^{-2}\boldsymbol{\beta} \tag{24}$$

$$= \sum_{i=1}^{2}\mathrm{var}\left(\hat{\beta}_i^*\right) + \sum_{i=1}^{2}\left[\mathrm{Bias}\left(\hat{\beta}_i^*\right)\right]^2 \tag{25}$$

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 $\hat{\boldsymbol{\beta}}^*$ rather than $\hat{\boldsymbol{\beta}}$. 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

vii. $\lim_{\hat{\beta}\beta\to\infty}\mathrm{MSE}\left(\hat{\boldsymbol{\beta}}^*\right) \to \infty$ and hence, for fixed k, the ridge estimator is not minimax

viii. If $\boldsymbol{\beta}'\boldsymbol{\beta}$ is bounded, then there exists a $k > 0$ such that

$$\mathrm{MSE}\left(\hat{\boldsymbol{\beta}}^{*}\right) < \mathrm{MSE}\left(\hat{\boldsymbol{\beta}}\right)$$

The Variance Inflation Factor

The variance inflation factor (VIF) can be computed using the equation

$$\mathrm{VIF} = \left(1 - R_j^2\right)^{-1}, \tag{26}$$

where R_j 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 orthogonality and has tendencies Generally, when the VIF > 10, we assume that there is high multicollinearity in the data (Mardikyan & Çetin, 2008) 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 $\mathrm{MSE}\left(\hat{\beta}(k)\right) < \mathrm{MSE}\left(\hat{\beta}\right)$. Further details on this issue have been provided by Judge (1988), Gujarati (1995), Gruber (1998), and Pasha and Shah (2004). Finally, if $R_j^2 < R^2$ for all *j*, and $R_j^2 < 0.90$ (which implies that $\mathrm{VIF} = \left(1 - R_j^2\right)^{-1} < 10$), then there is no need to worry about existence of multicollinearity; the R^2 will be close to 1 and the VIF will be large. However, when $\mathrm{VIF}_i > 10$, the data have collineaity problems.

Monte Carlo Design

A simulation study using 1000 samples with $n = 10$ was conducted to determine the appropriate k value for ridge regression in a p-variable regression model. The performances 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 (e.g., Ghazi & Barimal, 2010), the MSE, VIF, and β 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 and small VIF and

MSE values for different k values. A linear regression model with correlated independent variables is considered, and the different potential R_j values are computed.

The values of x_i, i = 1, 2, 3, and 4 were generated from normal distribution with (0, 2). For given x_1, x_2, x_3, x_4 correlated variables $\left(0 \le R_j^2 \le 1\right)$, the y values were generated using a set of predetermined values of parameters. However, only values of ε_i (i = 1, 2,…, n) were allowed to change randomly. The errors ε_i were generated to be $\varepsilon_i \sim$ i.i.d N(0, σ^2), i.e. independent and identically normally distributed with a mean of zero and variance of σ^2. The true values of the parameters were taken to be:

$$\boldsymbol{\beta}' = \left(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4\right) = \left(0.2, 1.2, 0.8, 2.5, 1.2\right), \sigma^2 = 2 \ .$$

One thousand data sets were used in each simulation study. Each data set was fitted by least squares and ridge regression estimation. The VIF and SSE for different k values and different R_j^2 were computed.

The mean of $\hat{\beta}_j^*$, the ridge estimates, VIF, and SSE are given by the following equations:

$$\overline{\hat{\beta}_j^*} = \frac{\sum_{i=1}^{1000}\left(\hat{\beta}_j^*\right)_i}{1000}, \quad j = 1, 2, 3, 4$$

$$\overline{\text{VIF}} = \frac{\sum_{i=1}^{1000}\left(\text{VIF}\right)_i}{1000}, \quad j = 1, 2, 3, 4$$

$$\overline{\text{SSE}} = \frac{\sum_{i=1}^{1000}\left(\text{SSE}\right)_i}{1000}, \quad j = 1, 2, 3, 4$$

Results

The simulation results are presented in Tables 4 to 7 and Figures 1 to 12. The results 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_i^2, i = 1, 2, 3, 4. For example, using Table 4 with $R_1^2 = 0.863$, $R_2^2 = 0.793$, $R_3^2 = 0.793$, $R_4^2 = 0.831$, $k = 0.63$ was chosen as a stable point solution. This gives $\hat{\beta}_1^* = 0.240571$, $\hat{\beta}_2^* = 0.237122$, $\hat{\beta}_3^* = 0.174854$, and $\hat{\beta}_4^* = 0.237988$, which are quite different

from the least squares estimates when $k = 0.63$. As a matter of fact, the R_j^2 values are large.

Table 1 lists the appropriate k values for ridge regression estimates $\hat{\beta}_1^*, \hat{\beta}_2^*, \hat{\beta}_3^*, \hat{\beta}_4^*$ for different R_j^2 values. Table 2 and Table 3 indicate that there is a relationship between k and R_j^2, and that the appropriate model for the data in Table 1 is a multiple regression model given by

$$K = \theta_1 R_1^2 + \theta_2 R_2^2 + \theta_3 R_3^2 + \theta_4 R_4^2 + \varepsilon \ .$$

The developed model is

$$K = 0.174R_1^2 + 0.170R_2^2 + 0.194R_3^2 + 0.199R_4^2 + \varepsilon$$

Table 1. The appropriate k values for the different R_j^2 values or ridge regression models

k	R_1^2	R_2^2	R_3^2	R_4^2
0.18	0.348	0.073	0.320	0.124
0.20	0.356	0.126	0.161	0.345
0.23	0.079	0.219	0.436	0.427
0.29	0.295	0.376	0.301	0.300
0.32	0.423	0.483	0.420	0.378
0.38	0.726	0.426	0.480	0.478
0.39	0.608	0.495	0.228	0.732
0.42	0.798	0.871	0.083	0.631
0.46	0.927	0.367	0.441	0.907
0.47	0.565	0.724	0.592	0.711
0.50	0.656	0.918	0.518	0.884
0.51	0.830	0.822	0.503	0.666
0.52	0.503	0.666	0.830	0.822
0.55	0.356	0.954	0.917	0.846
0.56	0.956	0.967	0.539	0.766
0.58	0.939	0.954	0.703	0.782
0.59	0.874	0.884	0.853	0.861
0.60	0.960	0.976	0.986	0.976
0.63	0.863	0.793	0.793	0.831
0.66	0.940	0.783	0.847	0.932
0.72	0.951	0.957	0.703	0.972
0.71	0.912	0.968	0.830	0.967
0.76	0.970	0.955	0.973	0.974
0.78	0.903	0.999	0.994	0.999
0.81	0.984	0.999	0.995	0.999

Table 2. The estimated regression coefficients, the standard errors, and the associated *t*-tests (R^2 = 99%)

Predictor No Constant	Coef	SE Coef	T	P
R_1^2	0.17361	0.02473	7.02	0.000
R_2^2	0.17018	0.02489	6.84	0.000
R_3^2	0.19415	0.02377	8.17	0.000
R_4^2	0.19859	0.02843	6.99	0.000

Table 3. Analysis of variance

Source	DF	SS	MS	F	P
Regression	4	13.5421	3.3855	2053.91	0.000
Residual Error	56	0.0923	0.0016		
Total	60	13.6344			

Table 4. The simulation means of $\hat{\beta}_1^*, \hat{\beta}_2^*, \hat{\beta}_3^*, \hat{\beta}_4^*$ VIF and SSE when $R_1^2 = 0.863$, $R_2^2 = 0.793$, $R_3^2 = 0.793$, and $R_4^2 = 0.831$

K	$A\hat{\beta}_1^*$	$A\hat{\beta}_2^*$	$A\hat{\beta}_3^*$	$A\hat{\beta}_4^*$	$AVIF_1$	$AVIF_2$	$AVIF_3$	$AVIF_4$	*A*SSE
0.00	0.298655	0.341844	0.025654	0.381536	7.31459	4.83671	4.83955	5.90772	0.007532
0.59	0.243716	0.240542	0.173427	0.241269	1.12303	1.04726	1.06043	1.08299	0.159369
0.60	0.242624	0.239611	0.173476	0.241021	1.10790	1.03431	1.04713	1.06902	0.161539
0.61	0.241733	0.238099	0.175252	0.239907	1.09315	1.02168	1.03419	1.05542	0.162970
0.62	0.241761	0.237168	0.174536	0.239319	1.07885	1.00940	1.02159	1.04217	0.165100
0.63	0.240571	0.237122	0.174854	0.237988	1.06493	0.99739	1.00928	1.02925	0.167347
0.64	0.240347	0.236061	0.174719	0.237375	1.05134	0.98570	0.99730	1.01668	0.169175
0.65	0.238930	0.235864	0.174632	0.236797	1.03815	0.97429	0.98559	1.00444	0.171522
0.66	0.238230	0.235382	0.174621	0.235977	1.02527	0.96313	0.97417	0.99249	0.173358
0.67	0.237334	0.234947	0.174993	0.234938	1.01275	0.95225	0.96301	0.98083	0.175236
0.68	0.236631	0.233285	0.174871	0.235186	1.00053	0.94161	0.95214	0.96944	0.177417
0.69	0.236608	0.233036	0.173879	0.234140	0.98859	0.93123	0.94148	0.95836	0.179716

Table 5. The simulation means of $\hat{\beta}_1^*, \hat{\beta}_2^*, \hat{\beta}_3^*, \hat{\beta}_4^*$ VIF and SSE when $R_1^2 = 0.563$, $R_2^2 = 0.540$, $R_3^2 = 0.550$, and $R_4^2 = 0.530$

K	$A\hat{\beta}_1^*$	$A\hat{\beta}_2^*$	$A\hat{\beta}_3^*$	$A\hat{\beta}_4^*$	$AVIF_1$	$AVIF_2$	$AVIF_3$	$AVIF_4$	$ASSE$
0.00	0.397900	0.278189	0.031668	0.427235	2.28958	2.17575	2.22168	2.12754	0.012955
0.42	0.301376	0.243077	0.131435	0.316583	1.05823	1.03735	1.04590	1.02797	0.154597
0.43	0.301949	0.243416	0.130461	0.313638	1.04523	1.02492	1.03325	1.01581	0.157037
0.44	0.299678	0.241674	0.132085	0.313240	1.03254	1.01282	1.02089	1.00395	0.159689
0.45	0.296759	0.242362	0.133185	0.311539	1.02020	1.00099	1.00885	0.99237	0.162455
0.46	0.295325	0.241872	0.134068	0.309751	1.00812	0.98945	0.99712	0.98108	0.165044
0.47	0.293972	0.240496	0.135793	0.308160	0.99635	0.97821	0.98565	0.97003	0.167455
0.48	0.290813	0.240918	0.136362	0.307768	0.98488	0.96721	0.97444	0.95924	0.169584
0.49	0.292787	0.239215	0.134971	0.304933	0.97369	0.95647	0.96352	0.94872	0.173381
0.50	0.290226	0.238191	0.136188	0.304710	0.96275	0.94598	0.95284	0.93842	0.175646
0.51	0.288852	0.237685	0.138675	0.302297	0.95204	0.93573	0.94243	0.92835	0.177022
0.52	0.289634	0.238207	0.134940	0.300539	0.94160	0.92570	0.93221	0.91852	0.180999

Table 6. The simulation means of $\hat{\beta}_1^*, \hat{\beta}_2^*, \hat{\beta}_3^*, \hat{\beta}_4^*$ VIF and SSE when $R_1^2 = 0.295$, $R_2^2 = 0.376$, $R_3^2 = 0.301$, and $R_4^2 = 0.300$

K	$A\hat{\beta}_1^*$	$A\hat{\beta}_2^*$	$A\hat{\beta}_3^*$	$A\hat{\beta}_4^*$	$AVIF_1$	$AVIF_2$	$AVIF_3$	$AVIF_4$	$ASSE$
0.00	0.354471	0.337643	0.025927	0.522447	1.41774	1.60351	1.43014	1.42781	0.014170
0.25	0.309537	0.306022	0.079025	0.427219	1.01056	1.08747	1.01416	1.01374	0.125967
0.26	0.311451	0.300830	0.083100	0.423240	0.99922	1.07392	1.00270	1.00232	0.129788
0.27	0.306251	0.302779	0.080909	0.422513	0.98817	1.06073	0.99152	0.99115	0.134086
0.28	0.306893	0.300791	0.084654	0.417579	0.97737	1.04786	0.98059	0.98023	0.137238
0.29	0.303219	0.299035	0.085051	0.417393	0.96684	1.03533	0.96991	0.96959	0.141547
0.30	0.303990	0.295907	0.087936	0.413816	0.95650	1.02312	0.95947	0.95916	0.144839
0.31	0.299973	0.296581	0.088638	0.411886	0.94641	1.01117	0.94927	0.94897	0.148536
0.32	0.301425	0.297556	0.087440	0.406633	0.93654	0.99956	0.93927	0.93901	0.151623
0.33	0.299633	0.295511	0.088587	0.404863	0.92688	0.98817	0.92951	0.92926	0.155481
0.34	0.298779	0.293971	0.089124	0.402402	0.91742	0.97711	0.91997	0.91974	0.159396
0.35	0.295223	0.290853	0.089966	0.403828	0.90816	0.96626	0.91063	0.91040	0.162995

Table 7. The simulation means of $\hat{\beta}_1^*, \hat{\beta}_2^*, \hat{\beta}_3^*, \hat{\beta}_4^*$ VIF and SSE when $R_1^2 = 0.927$, $R_2^2 = 0.367$, $R_3^2 = 0.441$, and $R_4^2 = 0.907$

K	$A\hat{\beta}_1^*$	$A\hat{\beta}_2^*$	$A\hat{\beta}_3^*$	$A\hat{\beta}_4^*$	AVIF$_1$	AVIF$_2$	AVIF$_3$	AVIF$_4$	ASSE
0.00	0.570594	0.431284	0.044372	0.080629	13.71980	1.57886	1.78922	10.70640	0.023146
0.42	0.299998	0.335019	0.124706	0.235173	1.42950	0.86316	0.91980	1.28940	0.174125
0.43	0.298707	0.335742	0.122633	0.234379	1.40230	0.85500	0.91033	1.26700	0.176739
0.44	0.297638	0.331644	0.123887	0.235597	1.37620	0.84700	0.90106	1.24540	0.179077
0.45	0.295425	0.328881	0.127117	0.235160	1.35110	0.83917	0.89198	1.22460	0.181398
0.46	0.295129	0.326968	0.126581	0.234254	1.32700	0.83148	0.88312	1.20450	0.184709
0.47	0.292531	0.326450	0.128946	0.232747	1.30380	0.82396	0.87443	1.18520	0.186945
0.48	0.294073	0.322601	0.126818	0.233503	1.28150	0.81654	0.86591	1.16660	0.189423
0.49	0.290583	0.324226	0.125192	0.232659	1.26000	0.80928	0.85756	1.14860	0.193995
0.50	0.290984	0.320314	0.125871	0.232666	1.23930	0.80217	0.84941	1.13120	0.195867
0.51	0.288384	0.319893	0.127771	0.232432	1.21920	0.79517	0.84138	1.11440	0.197233
0.52	0.288367	0.318794	0.125241	0.231326	1.19990	0.78831	0.83354	1.09820	0.201373

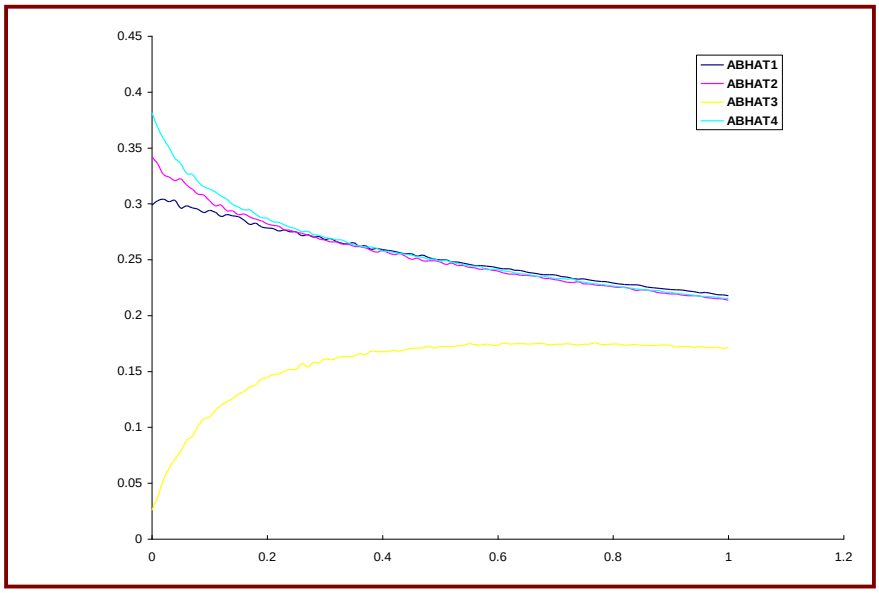

Figure 1. A plot of *k* vs $A\hat{\beta}_1^*, A\hat{\beta}_2^*, A\hat{\beta}_3^*, A\hat{\beta}_4^*$ by using the results of a simulation study when $R_1^2 = 0.863$, $R_2^2 = 0.793$, $R_3^2 = 0.793$, and $R_4^2 = 0.831$

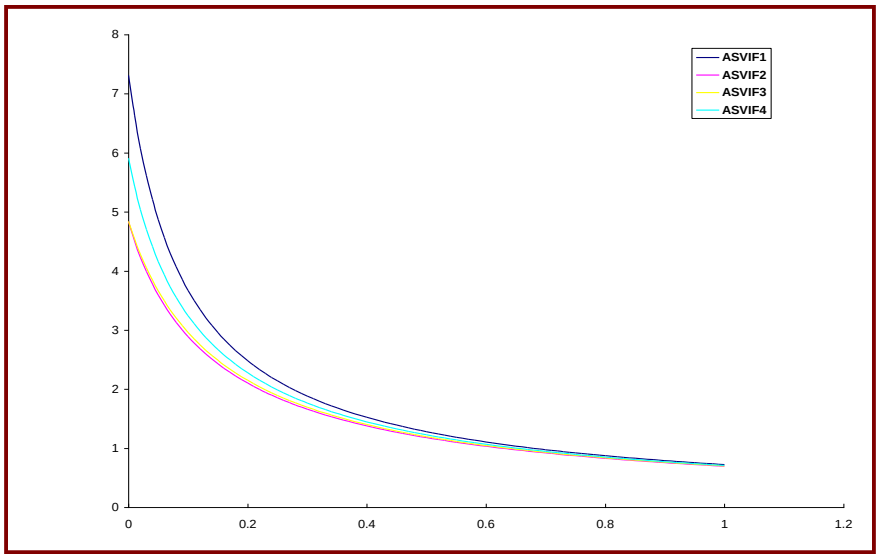

Figure 2. A plot of *k* vs $AVIF_1$, $AVIF_2$, $AVIF_3$, and $AVIF_4$ by using the results of a simulation study when $R_1^2 = 0.863$, $R_2^2 = 0.793$, $R_3^2 = 0.793$, and $R_4^2 = 0.831$

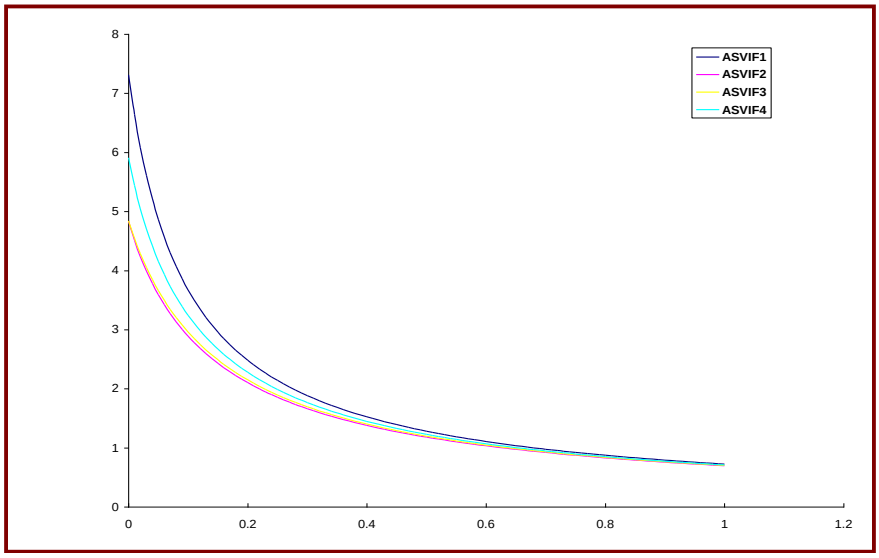

Figure 3. A plot of *k* vs SSE by using the results of a simulation study when $R_1^2 = 0.863$, $R_2^2 = 0.793$, $R_3^2 = 0.793$, and $R_4^2 = 0.831$

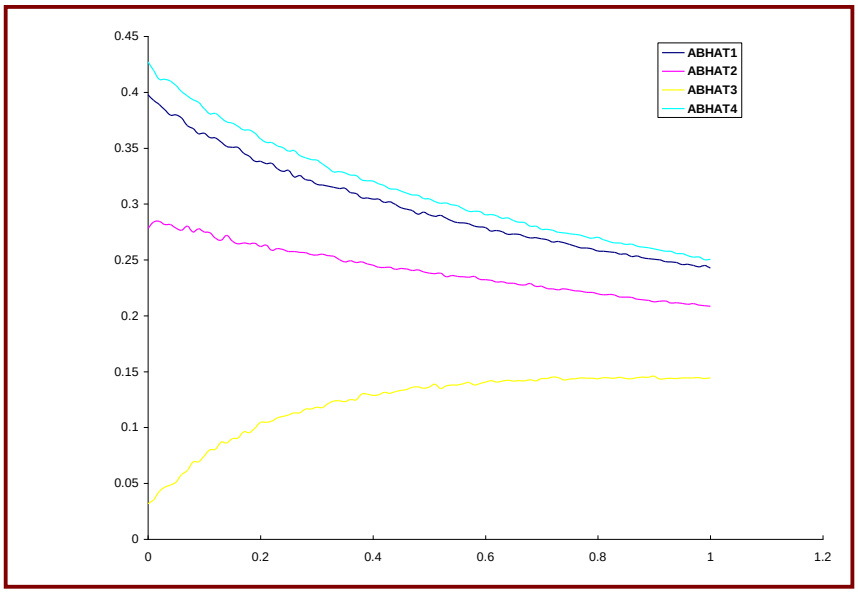

Figure 4. A plot of *k* vs $A\hat{\beta}_1^*, A\hat{\beta}_2^*, A\hat{\beta}_3^*, A\hat{\beta}_4^*$ by using the results of a simulation study when $R_1^2 = 0.563$, $R_2^2 = 0.540$, $R_3^2 = 0.550$, and $R_4^2 = 0.530$

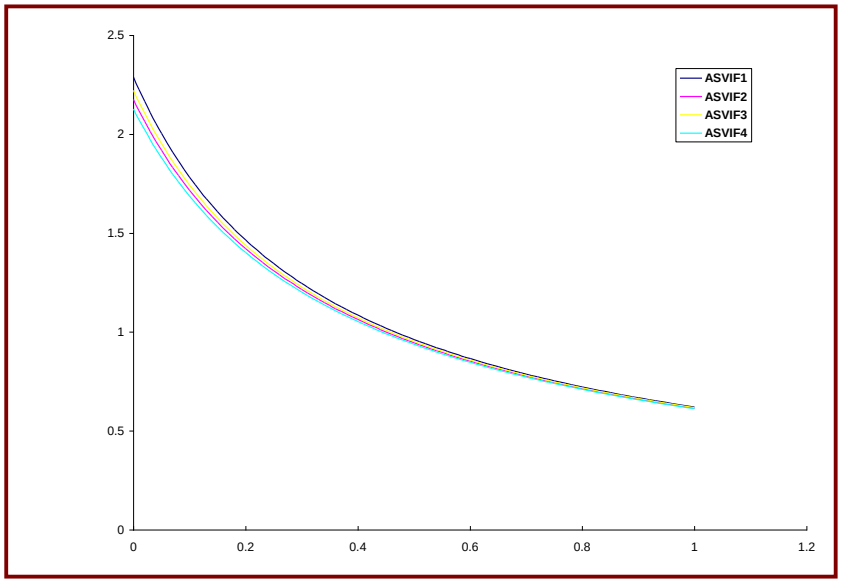

Figure 5. A plot of *k* vs $AVIF_1$, $AVIF_2$, $AVIF_3$, and $AVIF_4$ by using the results of a simulation study when $R_1^2 = 0.563$, $R_2^2 = 0.540$, $R_3^2 = 0.550$, and $R_4^2 = 0.530$

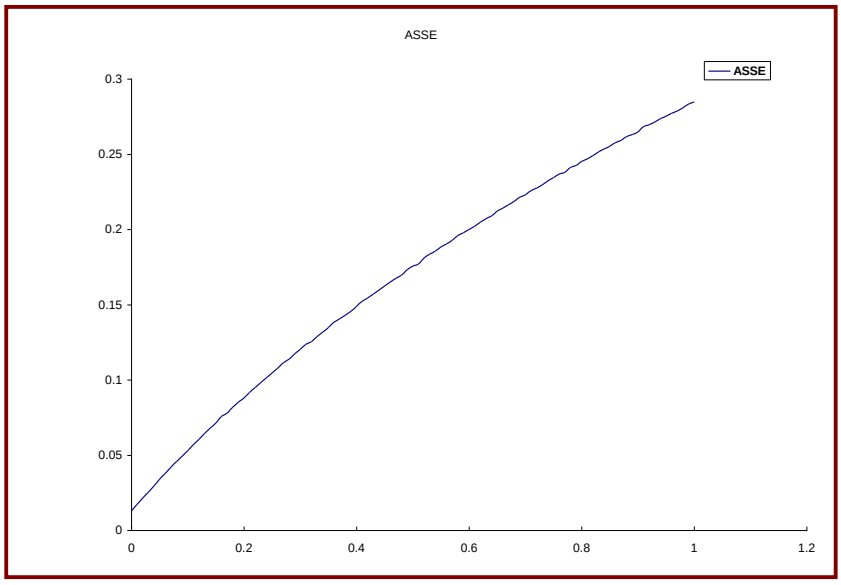

Figure 6. A plot of *k* vs SSE by using the results of a simulation study when $R_1^2 = 0.563$, $R_2^2 = 0.540$, $R_3^2 = 0.550$, and $R_4^2 = 0.530$

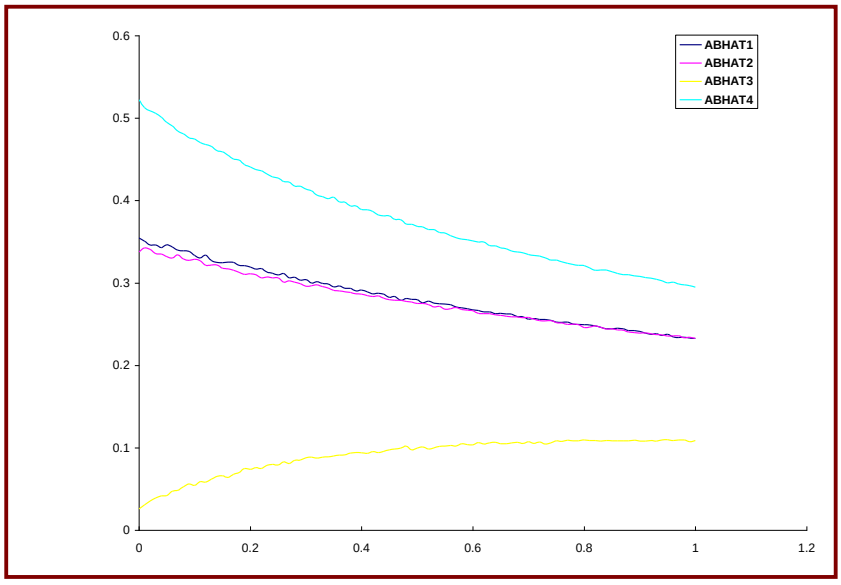

Figure 7. A plot of *k* vs $A\hat{\beta}_1^*, A\hat{\beta}_2^*, A\hat{\beta}_3^*, A\hat{\beta}_4^*$ by using the results of a simulation study when $R_1^2 = 0.295$, $R_2^2 = 0.376$, $R_3^2 = 0.301$, and $R_4^2 = 0.300$

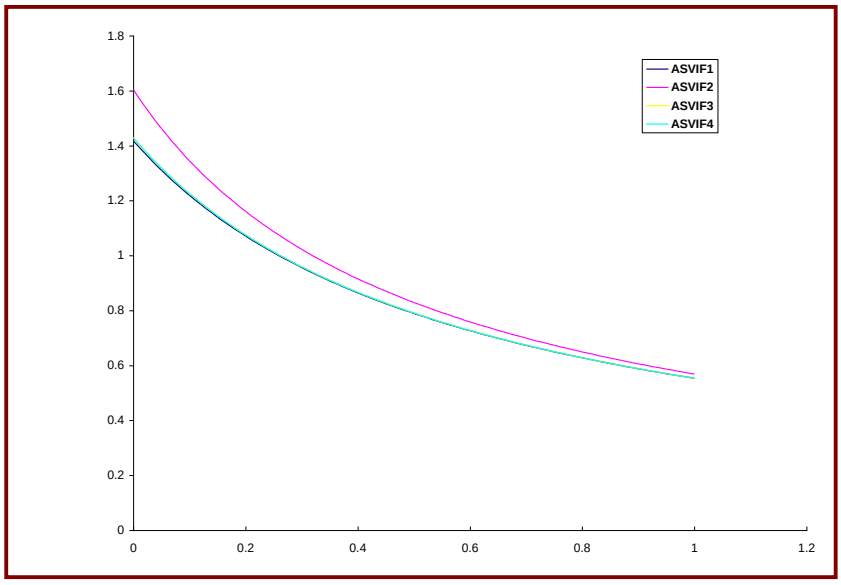

Figure 8. A plot of *k* vs $AVIF_1$, $AVIF_2$, $AVIF_3$, and $AVIF_4$ by using the results of a simulation study when $R_1^2 = 0.295$, $R_2^2 = 0.376$, $R_3^2 = 0.301$, and $R_4^2 = 0.300$

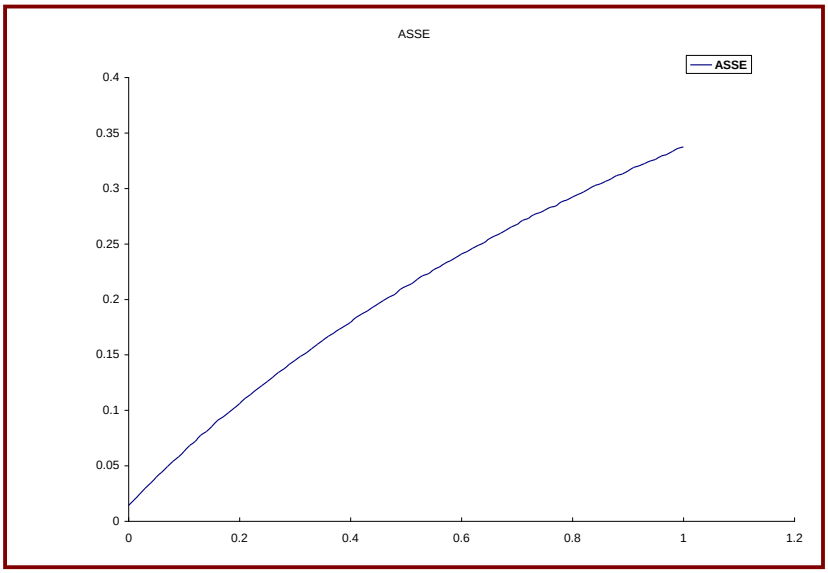

Figure 9. A plot of *k* vs SSE by using the results of a simulation study when $R_1^2 = 0.295$, $R_2^2 = 0.376$, $R_3^2 = 0.301$, and $R_4^2 = 0.300$

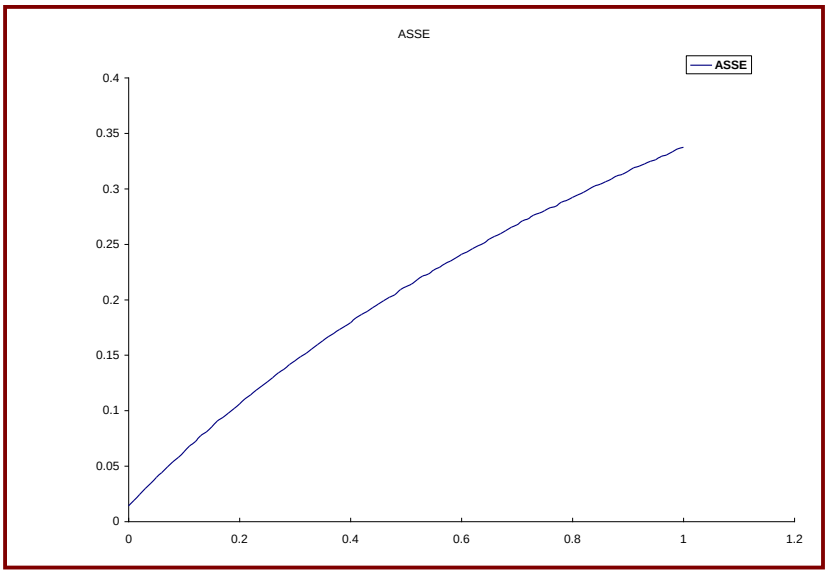

Figure 10. A plot of *k* vs $A\hat{\beta}_1^*, A\hat{\beta}_2^*, A\hat{\beta}_3^*, A\hat{\beta}_4^*$ by using the results of a simulation study when $R_1^2 = 0.927$, $R_2^2 = 0.376$, $R_3^2 = 0.441$, and $R_4^2 = 0.907$

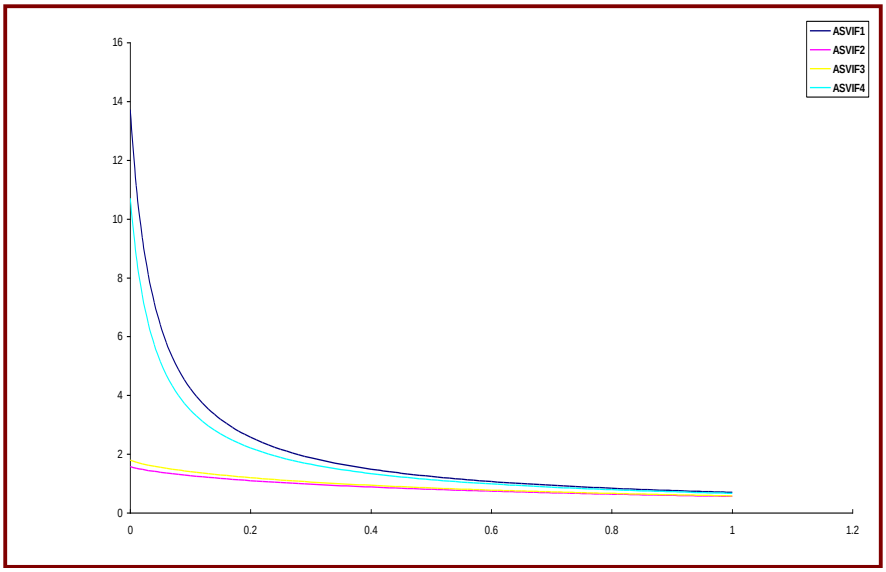

Figure 11. A plot of *k* vs $AVIF_1$, $AVIF_2$, $AVIF_3$, and $AVIF_4$ by using the results of a simulation study when $R_1^2 = 0.927$, $R_2^2 = 0.376$, $R_3^2 = 0.441$, and $R_4^2 = 0.907$

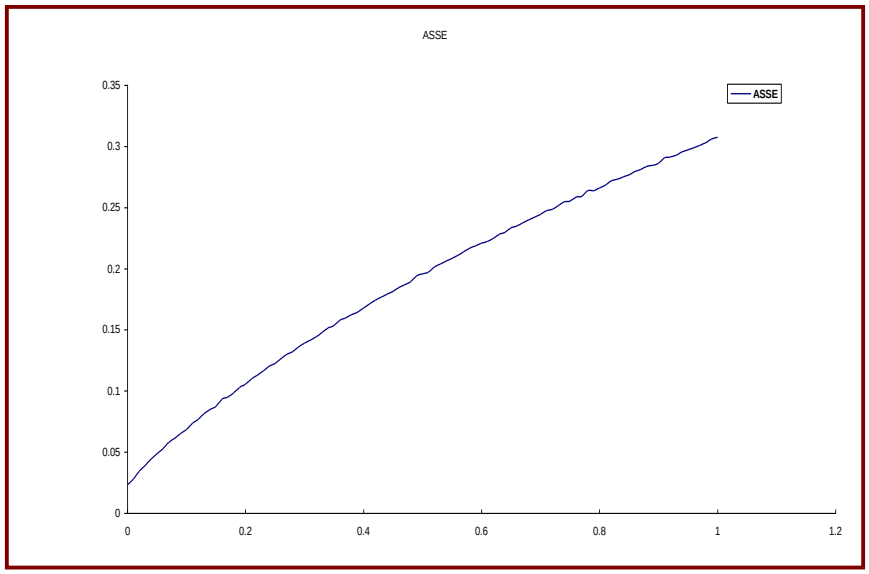

Figure 12. A plot of *k* vs SSE by using the results of a simulation study when $R_1^2 = 0.927$, $R_2^2 = 0.376$, $R_3^2 = 0.441$, and $R_4^2 = 0.907$

Discussion and Conclusion

The main goal of this study was to identify the most relevant *k* value for ridge regression in a four-variable regression model. Since it is not possible to achieve this mathematically, a simulation study was conducted to study the behavior of ridge regression in such case. It was assumed that both the form of the model and the nature of the errors, ε_i (i = 1, 2,…, n), are known. For given X_1, X_2, X_3, and X_4 correlated values $\left(0 \le R_j^2 \le 1\right)$, the y values were generated using a set of predetermined values of parameters, allowing only the values of ε_i to change randomly. The errors, ε_i, were generated such that $\varepsilon_i \sim$ i.i.d N(0, σ^2). One thousand random data sets were used in each simulation study. The p-variable linear regression model was fit by the least-squares method. Thereupon, in each simulation study, one thousand ridge regression estimates of $\hat{\boldsymbol{\beta}}^*$, VIF, and SSE for different *k* and R_j^2 values were computed. The simulation outcomes illustrate that there is a statistically-significant relationship between *k* and R_j^2. The most appropriate model to describe the relation between *k* and R_j^2 is a multiple regression model and the fitted regression equation is

$$k = 0.174R_1^2 + 0.170R_2^2 + 0.194R_3^2 + 0.199R_4^2 \quad (27)$$

From a practical point of view, when multicollinearity occurs, we can use ridge regression to solve this problem. The appropriate value of *k* can be chosen according to the ridge trace and Equation (27). In this study we have only considered four independent variables.

References

Alkhamisi, M., Khalaf, G., & Shukur, G. (2006). Some modifications for choosing ridge parameters. *Communications in Statistics – Theory and Methods, 35*(11), 2005-2020. doi: 10.1080/03610920600762905

Alkhamisi, M. A., & Shukur, G. (2008). Developing ridge parameters for SUR model. *Communications in Statistics – Theory and Methods, 37*(4), 544-564. doi: 10.1080/03610920701469152

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological), 39*(1), 1-38. Available from http://www.jstor.org/stable/2984875

Dorugade, A. V., & Kashid, D. N. (2010). Alternative method for choosing ridge parameter for regression. *Applied Mathematical Sciences, 4*(9), 447-456. Retrieved from http://www.m-hikari.com/ams/ams-2010/ams-9-12-2010/dorugadeAMS9-12-2010.pdf

El-Dereny, M., & Rashwan, N. I. (2011). Solving multicollinearity problem using ridge regression models. *International Journal of Contemporary Mathematical Sciences, 6*(12), 585-600. Retrieved from http://www.m-hikari.com/ijcms-2011/9-12-2011/rashwanIJCMS9-12-2011.pdf

Gruber, H. J. (1998). *Improving efficiency by shrinkage*. New York, NY: Marcel Dekker.

Gujarati, D. N. (1995). *Basic econometrics*. New York, NY: McGraw-Hill.

Hoerl, A. E., & Kennard, R. W. (1970a). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55-67. doi: 10.1080/00401706.1970.10488634

Hoerl, A. E., & Kennard, R. W. (1970b). Ridge regression: Applications to nonorthogonal problems. *Technometrics, 12*(1), 69-82. doi: 10.1080/00401706.1970.10488635

Hoerl, A. E., Kannard, R. W., & Baldwin, K. F. (1975). Ridge regression: Some simulations. *Communications in Statistics – Theory and Methods, 4*(2), 105-123. doi: 10.1080/03610927508827232

Jensen, D. R., & Ramirez, D. E. (2012). Variations on ridge traces in regression. *Communications in Statistics – Simulation and Computation, 41*(2), 265-278. doi: 10.1080/03610918.2011.586482

Judge, G. G. (1988). *Introduction to theory and practice of econometrics*. New York, NY: John Willy and Sons.

Khalaf, G., & Shukur, G. (2005). Choosing ridge parameter for regression problems. *Communications in Statistics – Theory and Methods, 34*(5), 1177-1182. doi: 10.1081/STA-200056836

Lawless, J. F., & Wang, P. (1976). A simulation study of ridge and other regression estimators. *Communications in Statistics – Theory and Methods, 5*(4), 307-323. doi: 10.1080/03610927608827353

Mardikyan, S., & Çetin, E. (2008). Efficient choice of biasing constant for ridge regression. *International Journal of Contemporary Mathematical Sciences, 3*(11), 527-536. Retrieved from http://www.m-hikari.com/ijcms-password2008/9-12-2008/cetinIJCMS9-12-2008.pdf

McDonald, G. C., & Galarneau, D. I. (1975). A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association, 70*(350), 407-416. doi: 10.1080/01621459.1975.10479882

Muniz, G., & Kibria, B. M. G. (2009). On some ridge regression estimators: An empirical comparisons. *Communications in Statistics – Simulation and Computation, 38*(3), 621-630. doi: 10.1080/03610910802592838

Pasha, G. R., & Shah, M. A. (2004). Application of ridge regression to multicollinear data. *Journal of Research (Science), 15*(1), 97-106. Retrieved from http://www.bzu.edu.pk/jrscience/vol15no1/15.php