Taylor & Francis
Taylor & Francis Group

Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

A simulation study on SPSS ridge regression and ordinary least squares regression procedures for multicollinearity data

John Zhang & Mahmud Ibrahim

To cite this article: John Zhang & Mahmud Ibrahim (2005) A simulation study on SPSS ridge regression and ordinary least squares regression procedures for multicollinearity data, Journal of Applied Statistics, 32:6, 571-588, DOI: 10.1080/02664760500078946

To link to this article: https://doi.org/10.1080/02664760500078946

Published online: 12 Apr 2011.

Journal of Applied Statistics
Vol. 32, No. 6, 571–588, August 2005

Routledge
Taylor & Francis Group

A Simulation Study on SPSS Ridge Regression and Ordinary Least Squares Regression Procedures for Multicollinearity Data

JOHN ZHANG* & MAHMUD IBRAHIM**

**Department of Mathematics, Indiana University of PA, Indiana, USA, **Department of Mathematics, Syracuse University, USA*

ABSTRACT *This study compares the SPSS ordinary least squares (OLS) regression and ridge regression procedures in dealing with multicollinearity data. The LS regression method is one of the most frequently applied statistical procedures in application. It is well documented that the LS method is extremely unreliable in parameter estimation while the independent variables are dependent (multicollinearity problem). The Ridge Regression procedure deals with the multicollinearity problem by introducing a small bias in the parameter estimation. The application of Ridge Regression involves the selection of a bias parameter and it is not clear if it works better in applications. This study uses a Monte Carlo method to compare the results of OLS procedure with the Ridge Regression procedure in SPSS.*

KEY WORDS: Ridge regression, least squares regression, eigenvalues, eigenvectors, simulation

Introduction

SPSS is the statistical software of choice in Indiana University of PA, USA. Many applications require a regression method. In the Applied Research Lab, a statistical consulting arm of the graduate school in our university, we often face the problem of parameter estimation when there is multicollinearity. One simple solution to this problem is to drop some of the highly correlated variables. This strategy usually works well. However, there are times when the variables are too important to be excluded from the analysis. One strategy is to apply Ridge Regression in such cases. Under some conditions, Ridge Regression has, in theory, been shown to be effective in dealing with multicollinearity. However, it was unclear in many applications that these conditions were satisfied. Indeed, we were never sure if Ridge Regression provides a better model in these applications. In many cases, we are also interested in how well the SPSS Ridge Regression procedure works. We thus conducted this simulation study to evaluate the SPSS ridge regression.

Correspondence Address: John Zhang, Department of Mathematics, Indiana University of PA, Indiana PA, 15705, USA. Email: john.zhang@iup.edu

0266-4763 Print/1360-0532 Online/05/060571–18
DOI: 10.1080/02664760500078946

It is well known that OLS estimates of parameters are unstable when the vectors of the explanatory variables are multicollinear. Multicollinearity refers to the situation where the explanatory variables are not orthogonal. In addition to the problem of estimation, multicollinearity makes misleading or erroneous inferences on: identifying the relative effects of the explanatory variables; prediction; and selection of an appropriate set of variables for the model, etc.

Hoerl & Kennard (1970) have demonstrated that some of these undesirable effects of multicollinearity can be reduced by using 'ridge' estimates in place of the least squares estimates. The ridge estimates depend on a parameter, k that is determined by the data in practice. Several mechanical rules and a graphical procedure, known as the ridge trace, have been proposed for the selection of k.

The organization of this thesis is as follows: the next section provides background information. The mechanical rule and evaluation criteria are outlined in third section. The fourth section provides a detailed simulation procedure. The results of the simulations are presented in the fifth section. The final section gives concluding remarks.

Background Information

Consider the multiple linear regression model: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, where $\mathbf{y}$ is a $(n \times 1)$ vector of observations on the dependent variable, $\mathbf{X}$ is a $(n \times p)$ fixed matrix of observations on the explanatory variables, $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of unknown regression coefficients, and $\mathbf{e}$ is a $(n \times 1)$ vector of errors assumed to be normally distributed with $\boldsymbol{E}(\mathbf{e}) = \boldsymbol{0}$ and $\boldsymbol{E}(\mathbf{ee}') = \sigma^2\mathbf{I}_n$. The usual estimator for $\boldsymbol{\beta}$ is the least squares estimator given by $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

When the vector of predictor variables is multicollinear, the least squares estimates are likely to be large in absolute value and even with a wrong sign. The problem is a result of the fact that $(\mathbf{X}'\mathbf{X})$ is near singular. The Gauss–Markov property gives the assurance that the least squares estimator has minimum variance in the class of unbiased linear estimators, but there is no guarantee that this variance will be small.

One way to alleviate this problem is to drop the requirement that the estimator of $\boldsymbol{\beta}$ be unbiased. Suppose there is a biased estimator of $\boldsymbol{\beta}$, say $\hat{\beta}^*$, that has a smaller variance than the unbiased estimator $\hat{\beta}$. Consider the mean squared error of the estimator $\hat{\beta}^*$:

$$\mathbf{MSE}(\hat{\beta}^*) = \boldsymbol{E}(\hat{\beta}^* - \boldsymbol{\beta})^2 = \mathbf{V}\text{ar}(\hat{\beta}^*) + [\boldsymbol{E}(\hat{\beta}^*) - \boldsymbol{\beta}]^2$$

or

$$\mathbf{MSE}(\hat{\beta}^*) = \mathbf{V}\text{ar}(\hat{\beta}^*) + (\text{bias in } \hat{\beta}^*)^2$$

It should be noted that the **MSE** is just the distance from $\hat{\beta}^*$ to $\boldsymbol{\beta}$. By allowing a small bias in $\hat{\beta}^*$, the variance of $\hat{\beta}^*$ can be made smaller. Consequently confidence intervals on $\boldsymbol{\beta}$ would be narrower using the biased estimator. The small variance for the biased estimator also implies that $\hat{\beta}^*$ is a more stable estimator of $\boldsymbol{\beta}$ than is the unbiased estimator $\hat{\beta}$. Hence a model using $\hat{\beta}^*$ may have better predictive power.

Following this line of thinking, Hoerl & Kennard suggested that the least squares estimators be replaced by the 'ridge' estimators $\hat{\beta}(k)$, where, for a fixed $k > 0$, $\hat{\beta}(k) = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}\mathbf{X}'\mathbf{y}$.

The ridge estimator is a linear transformation of the least squares estimator since

$$\begin{aligned}\hat{\beta}(k) &= (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}\mathbf{X}'\mathbf{y}\\ &= (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}(\mathbf{X}'\mathbf{X})\hat{\beta}\\ &= \mathbf{Z}_k\hat{\beta}\end{aligned}$$

Therefore

$$\boldsymbol{E}[\hat{\beta}(k)] = \boldsymbol{E}(\mathbf{Z}_k\hat{\beta}) = \mathbf{Z}_k\boldsymbol{\beta}$$

and $\hat{\beta}(k)$ is a biased estimator of $\boldsymbol{\beta}$.

The covariance matrix of $\hat{\beta}(k)$ is

$$\mathbf{V}\text{ar}(\hat{\beta}(k)) = \sigma^2(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}$$

The mean squared error of the ridge estimator is

$$\begin{aligned}\mathbf{MSE} &= \mathbf{V}\text{ar}(\hat{\beta}(k)) + [\text{bias in } \hat{\beta}(k)]^2\\ &= \sigma^2\mathbf{Tr}[(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}] + k^2\boldsymbol{\beta}'(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-2}\boldsymbol{\beta}\\ &= \sigma^2\sum_{j=1}^{p}\frac{\lambda_j}{(\lambda_j + k)^2} + k^2\boldsymbol{\beta}'(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-2}\boldsymbol{\beta},\end{aligned}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_p$ are eigenvalues of $\mathbf{X}'\mathbf{X}$. The first term on the right-hand side of the above equation is the sum of the variances of the parameters in $\hat{\beta}(k)$ and the second term is the square of the bias. If $k > 0$, the bias in $\hat{\beta}(k)$ increases with k. However, the variance decreases as k increases. We would like to select a k such that the reduction in the variance term is greater than the increase in the squared bias. If this can be done, the mean squared error of the ridge estimator $\hat{\beta}(k)$ will be less than the variance of the least squares estimator $\hat{\beta}$. Hoerl & Kennard (1970) proved that there exists a non-zero value of k for which the **MSE** of $\hat{\beta}(k)$ is less than the variance of the least squares estimator $\hat{\beta}$, provided that $\boldsymbol{\beta}'\boldsymbol{\beta}$ is bounded.

The residual sum of squares is

$$\begin{aligned}\mathbf{SSE} &= [(\mathbf{y} - \mathbf{X}\hat{\beta}(k)]'[(\mathbf{y} - \mathbf{X}\hat{\beta}(k)]\\ &= (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta}) + [\hat{\beta}(k) - \hat{\beta}]'\mathbf{X}'\mathbf{X}[\hat{\beta}(k) - \hat{\beta}].\end{aligned}$$

Since the first term on the right-hand side of the preceding equation is the residual sum of squares for the least squares estimator $\hat{\beta}$, we see that as k increases, the residual sum of squares increases. Consequently because the total sum of squares is fixed, $\boldsymbol{R}^2$ decreases as k increases. Therefore, the ridge estimate will not necessarily provide the best 'fit' to the Data. This should not be of concern, when the objective is to provide a stable set of parameter estimates. The ridge regression may result in an equation that does a better job of predicting future observations than the least squares regression (although there is no conclusive proof that this will happen).

The Evaluation Criteria and Selection of Parameter k

To compare estimators of the unknown coefficient vector $\boldsymbol{\beta}$, a criterion for measuring the 'goodness' of an estimator is needed. A criterion is the total mean squared error function defined by

$$\boldsymbol{E}(\boldsymbol{L}^*) = \mathbf{E}[(\hat{\beta}^* - \boldsymbol{\beta})'(\hat{\beta}^* - \boldsymbol{\beta})]$$

The total mean squared error is the sum of the mean squared errors for the individual coefficients. Since it represents the expected squared distance between $\boldsymbol{\beta}$ and $\hat{\beta}^*$, a 'good' estimator is characterized by a relatively small mean squared error.

Hoerl & Kennard (1970) have demonstrated the existence of a range of k values for which the associated ridge estimators are better with respect to mean squared error than the least squares estimator. The ridge parameter, k in their argument is assumed to be fixed, and the range of k values for which the ridge estimators are demonstrated to be better than the least squares depends on the unknown coefficient vector $\boldsymbol{\beta}$ as well as σ^2. Thus, no constant value of k can be assured to yield a ridge estimator that is better than least squares for all unknown coefficients vectors. Consequently, several 'rules' have been proposed for choosing k.

The choice of k used in this study is the one proposed by Hoerl *et al.* (1975). The appropriate choice for k suggested is

$$k = p\hat{\sigma}^2/\hat{\beta}'\hat{\beta}$$

where $\hat{\sigma}^2$, an estimate of σ^2, is defined by

$$\hat{\sigma}^2 = (\mathbf{y} - \mathbf{X}\hat{\beta})'(\mathbf{y} - \mathbf{X}\hat{\beta})/(n - p - 1)$$

The estimates $\hat{\beta}$ and $\hat{\sigma}^2$ are obtained from the least squares solution. It should be pointed out that since the choice of k depends on the particular sample under investigation, the properties associated with ridge regression for fixed k may not hold. In addition, the least squares estimation procedure corresponds to a choice of $k = 0$.

Simulation

The simulations in this study are similar to Gibbons (1981). One hundred observations are generated using the statistical software package, *SPSS* for each of three explanatory variables. In the notation of Chapter 2, $p = 3$ and $n = 100$. The explanatory variables are generated by

$$\chi_{ij} = (1 - \rho^2)z_{ij} + \rho z_{i4} \qquad i = 1, \ldots, n; \quad j = 1, \ldots, p$$

where z_{i1}, z_{i2}, z_{i3} and z_{i4} are independent standard normal pseudo-random numbers, and ρ is specified so that the correlation between any two explanatory variables is given by ρ^2. These variables are then standardized so that $\mathbf{X}'\mathbf{X}$ is in correlation form. Four different sets of correlations are considered corresponding to $\rho = 0.80$, 0.90, 0.95 and 0.99.

For each set of explanatory variables so constructed, two choices for the true coefficient vector $\boldsymbol{\beta}$ are considered. The mean square error function, $\boldsymbol{E}[\boldsymbol{L}(k)]$, associated with ridge estimation depends on the explanatory variables (through the λ_j), on σ^2 and on $\boldsymbol{\beta}$.

Newhouse & Oman (1971) have noted that if $\boldsymbol{E}[\boldsymbol{L}(k)]$ is regarded as a function of $\boldsymbol{\beta}$ with σ^2, k and the explanatory variables fixed, then subject to the constraint that $\|\boldsymbol{\beta}\| = 1$, $\boldsymbol{E}[\boldsymbol{L}(k)]$ is minimized when $\boldsymbol{\beta}$ is the normalized eigenvector corresponding to the largest eigenvalue of the $\mathbf{X}'\mathbf{X}$ matrix. Similarly, $\boldsymbol{E}[\boldsymbol{L}(k)]$ is maximized when $\boldsymbol{\beta}$ is the normalized eigenvector corresponding to the smallest eigenvalue of the $\mathbf{X}'\mathbf{X}$ matrix. It is understood that these remarks are not necessarily true when k is not fixed. We use the normalized eigenvectors corresponding to the largest and smallest eigenvalues of the correlation matrix in our simulation and denote the two vectors as $\boldsymbol{\beta}_{\mathrm{L}}$ and $\boldsymbol{\beta}_{\mathrm{S}}$ respectively.

Observations on the dependent variable are determined by

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1\chi_{j1} + \beta_2\chi_{i2} + \beta_3\chi_{i3} + e_i \\
i &= 1, \ \ldots, n,
\end{aligned}
$$

where the e_i are independent normal pseudo-random numbers with mean zero and variance σ^2, and β_0 is taken to be identically zero. Seven values of σ are investigated in this study: 0.01, 0.1, 0.2, 0.3, 0.4, 0.5 and 1.0.

For each set of dependent and explanatory variables so constructed, the class of estimators defined by

$$\hat{\beta}(k) = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}\mathbf{X}'\mathbf{y}$$

are determined. The variables are standardized so that $\mathbf{X}'\mathbf{y}$ represents the vector of correlations of the dependent variable with each explanatory variable. The standardized least squares and ridge regression coefficients are computed. The LS regression estimates are computed by using the linear regression program on *SPSS* while the ridge regression estimates are computed by first augmenting the standardized data as follows:

$$\mathbf{X}_{\mathrm{A}} = \begin{bmatrix} \mathrm{X} \\ \sqrt{k}\mathrm{I}_p \end{bmatrix}, \quad \mathbf{y}_{\mathrm{A}} = \begin{bmatrix} y \\ 0_p \end{bmatrix}$$

where $\sqrt{k}\mathrm{I}_p$ is a $p \times p$ diagonal matrix with diagonal elements equal to the square root of the biasing parameter and 0_p is a $p \times 1$ vector of zeros. The estimates are then computed using the *SPSS* linear regression program:

$$\hat{\beta}(k) = (\mathbf{X}_{\mathrm{A}}'\mathbf{X}_{\mathrm{A}})^{-1}\mathbf{X}_{\mathrm{A}}'\mathbf{y}_{\mathrm{A}} = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}\mathbf{X}'\mathbf{y}$$

The estimated length of the true standardized coefficient vector is determined from an unbiased estimator of $\boldsymbol{\beta}'\boldsymbol{\beta}$ obtained as follows:

$$Q = \hat{\beta}'\hat{\beta} - \hat{\sigma}^2 \sum_{1}^{p} \lambda_j^{-1}$$

where the λ_j, $j = 1, \ldots, p$, represent the eigenvalues of the correlation matrix, $\hat{\beta}$ represents the standardized least squares estimate, and $\hat{\sigma}^2$ is the estimate of the variance for the standardized model. The standardized coefficients are transformed back to the orig-

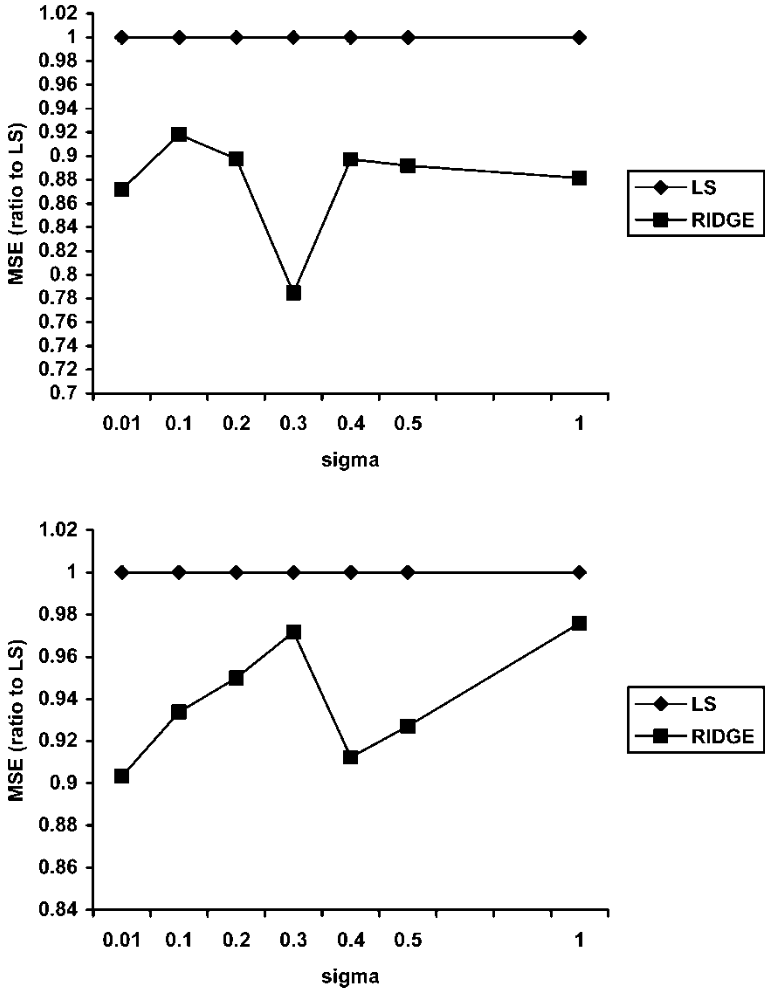


Figure 1. Ratio of the estimated **MSE** of ridge regression to the estimated **MSE** of the LS regression as a function of σ ($\rho = 0.80$). The top (bottom) graph presents results for $\boldsymbol{\beta}_L$ ($\boldsymbol{\beta}_S$)

inal model (Wichern & Churchhill, 1978). A constant term is estimated for the model by

$$\hat{\beta}_0(k) = \bar{y} - \sum_{1}^{3} \hat{\beta}_0(k)\bar{\chi}_j$$

where

$$\bar{y} = \frac{\sum_1^n y_i}{n} \quad \text{and} \quad \bar{\chi}_j = \frac{\sum_1^n x_{ij}}{n}$$

$j = 1, 2, 3$. The mean squared error and the k values are computed for the least squares and ridge regression methods. The estimation procedure that performed the best (minimum squared error) and the estimation procedure that performed the worst (maximum squared error) are noted.

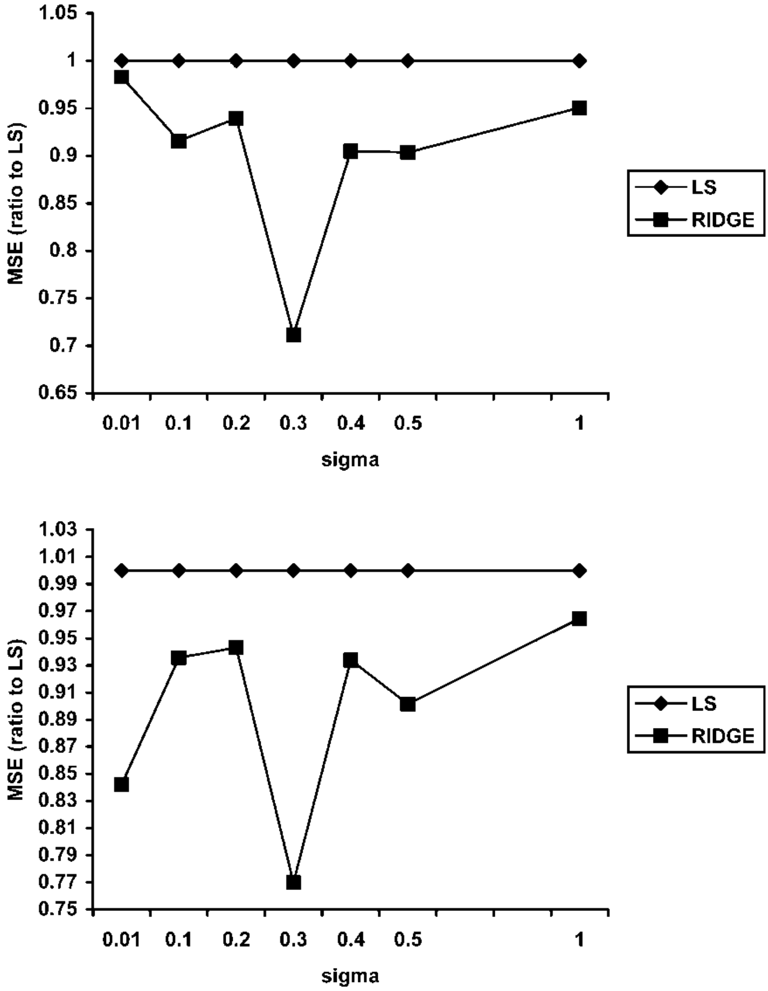


Figure 2. Ratio of the estimated **MSE** of ridge regression to the estimated **MSE** of the LS regression as a function of σ ($\rho = 0.90$). The top (bottom) graph presents results for $\boldsymbol{\beta}_L$ ($\boldsymbol{\beta}_S$)

Additional samples of size $n = 100$ are generated. The $\mathbf{X}'\mathbf{X}$ matrix and $\boldsymbol{\beta}$ vector remain fixed, while the random error and hence dependent variable change. After five samples have been generated, the following quantities are computed for each estimator:

(1) Average squared error (estimated mean squared error)
(2) Average k value
(3) Average $\boldsymbol{R}^2$
(4) The ratio of the estimated **MSE** for a particular ridge estimator to the estimated **MSE**

Some Simulation Results

Ridge estimators are constructed with the aim of having smaller **MSE** than least squares (LS) estimators. A measure of improvement is the ratio of the estimated **MSE** for a

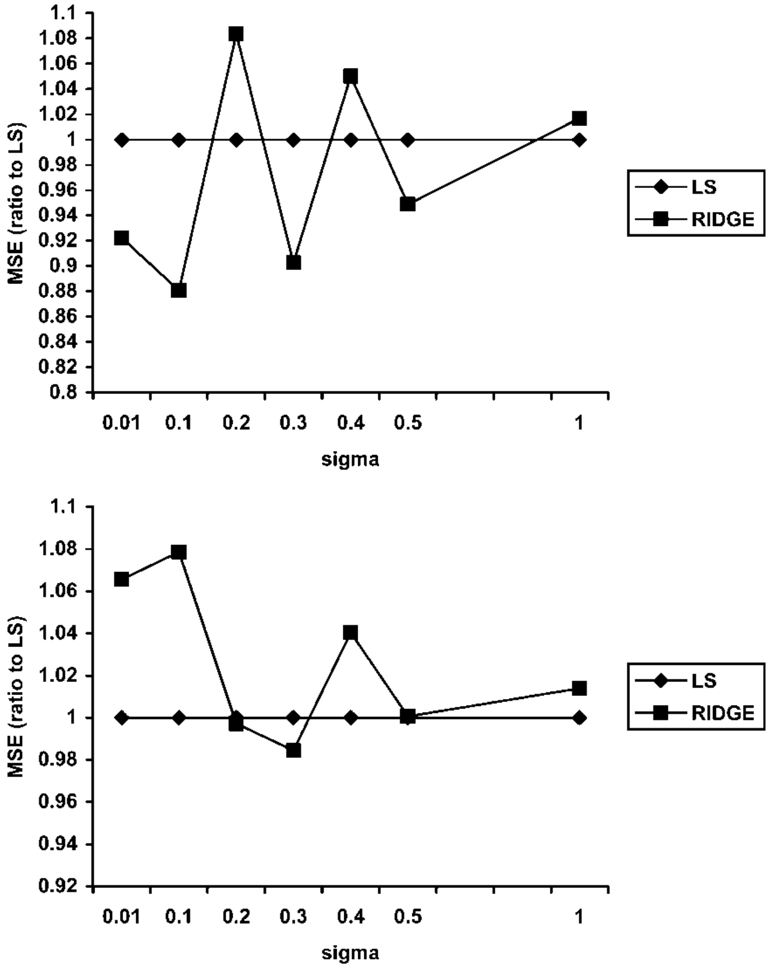


Figure 3. Ratio of the estimated **MSE** of ridge regression to the estimated **MSE** of the LS regression as a function of σ ($\rho = 0.95$). The top (bottom) graph presents results for $\boldsymbol{\beta}_L$ ($\boldsymbol{\beta}_S$)

particular ridge estimator to the estimated **MSE** for LS. This measure, denoted M is convenient in that it allows the results for all (ρ, σ, $\boldsymbol{\beta}$) to be represented on the same scale. The ratio is plotted in Figures 1 through 4. Each figure represents a fixed value of ρ and presents the ratio M for each of the estimators as a function of σ. The top graph presents the results for $\boldsymbol{\beta}_L$ and the bottom graph for $\boldsymbol{\beta}_S$. Each point plotted represents the average of five repeated simulations. In each run, the **MSE** and $\boldsymbol{R}^2$ for the LS estimator as well as the ridge estimator are computed for each σ. The average **MSE** and average $\boldsymbol{R}^2$ for each estimator are recorded.

From Figures 1–4, we have the following summary.

$\boldsymbol{\beta} = \boldsymbol{\beta}_L$

(1) The ridge estimator is at least as good as the LS estimator; that is $M \leq 1$ for the (ρ, σ) combination when $\rho = 0.80$ or $\rho = 0.90$. However, mixed results are obtained for the (ρ, σ) when $\rho = 0.95$ or $\rho = 0.99$.

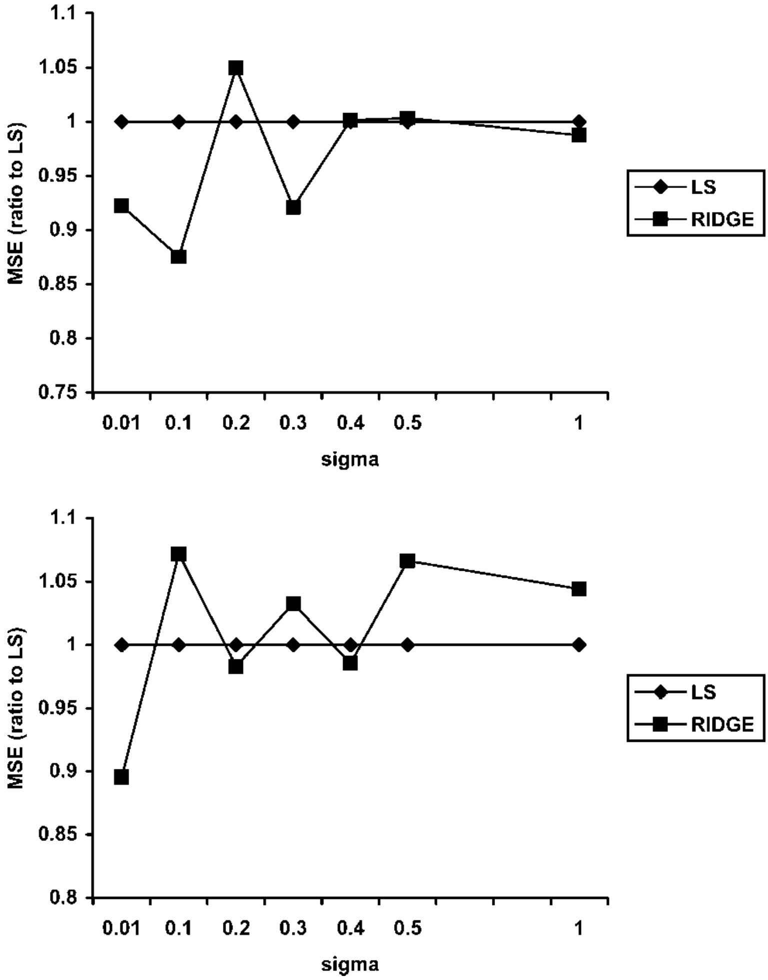


Figure 4. Ratio of the estimated **MSE** of ridge regression to the estimated **MSE** of the LS regression as a function of σ ($\rho = 0.99$). The top (bottom) graph presents results for $\boldsymbol{\beta}_L$ ($\boldsymbol{\beta}_S$)

(2) The best performance of the ridge estimator is observed when $\rho = 0.80$ and the worst performance is observed when $\rho = 0.95$.
(3) The least value of M, 0.78 is observed when $\rho = 0.80$ and the highest value of M, 1.08 is observed when $\rho = 0.95$. That is $0.78 \leq M \leq 1.08$.

$\boldsymbol{\beta} = \boldsymbol{\beta}_S$

(1) The ridge estimator is at least as good as the LS estimator; that is $M \leq 1$ for the (ρ, σ) combination when $\rho = 0.80$ or $\rho = 0.90$. However, mixed results are obtained for the (ρ, σ) when $\rho = 0.95$ or $\rho = 0.99$.
(2) The best performance of the ridge estimator is observed when $\rho = 0.90$ and the worst performance is observed when $\rho = 0.99$.
(3) The least value of M, 0.77 is observed when $\rho = 0.90$ and the highest value of M; 1.07 is observed when $\rho = 0.99$. That is $0.77 \leq M \leq 1.07$.

Table 1. Average R^2 values of LS coefficients obtained in simulation

		ρ			
σ	β	0.80	0.90	0.95	0.99
0.01	β_L	1.000	1.000	1.000	1.000
	β_S	1.000	1.000	1.000	1.000
0.1	β_L	1.000	1.000	1.000	1.000
	β_S	1.000	1.000	1.000	1.000
0.2	β_L	0.999	0.999	0.999	0.999
	β_S	0.998	0.998	0.998	0.998
0.3	β_L	0.994	0.993	0.993	0.994
	β_S	0.990	0.992	0.991	0.992
0.4	β_L	0.982	0.979	0.977	0.979
	β_S	0.971	0.973	0.972	0.973
0.5	β_L	0.948	0.945	0.953	0.951
	β_S	0.925	0.934	0.935	0.936
1	β_L	0.560	0.559	0.558	0.537
	β_S	0.482	0.445	0.464	0.452

The quantity tabulated in Table 1 gives the average R^2 in five samples based on the LS coefficients, while the quantity tabulated in Table 2 is based on the ridge coefficients.

For a fixed (ρ, σ) combination, R^2 is generally smaller when $\beta = \beta_S$ than when $\beta = \beta_L$. Low R^2 values (≤ 0.5) are observed when $\sigma = 1$ and $\beta = \beta_S$ in both Tables 1 and 2. The largest R^2 value, 1 is observed when $0.01 \leq \sigma \leq 0.2$ for all β; and the smallest R^2 value,

Table 2. Average R^2 values of ridge coefficients obtained in simulation

		ρ			
σ	β	0.80	0.90	0.95	0.99
0.01	β_L	1.000	1.000	1.000	1.000
	β_S	1.000	1.000	1.000	1.000
0.1	β_L	1.000	1.000	1.000	1.000
	β_S	1.000	1.000	1.000	1.000
0.2	β_L	0.999	0.999	0.999	0.999
	β_S	0.998	0.998	0.998	0.998
0.3	β_L	0.994	0.993	0.993	0.993
	β_S	0.991	0.992	0.992	0.991
0.4	β_L	0.980	0.978	0.978	0.978
	β_S	0.968	0.971	0.972	0.973
0.5	β_L	0.952	0.948	0.951	0.949
	β_S	0.926	0.932	0.933	0.929
1	β_L	0.541	0.556	0.520	0.526
	β_S	0.460	0.420	0.454	0.409

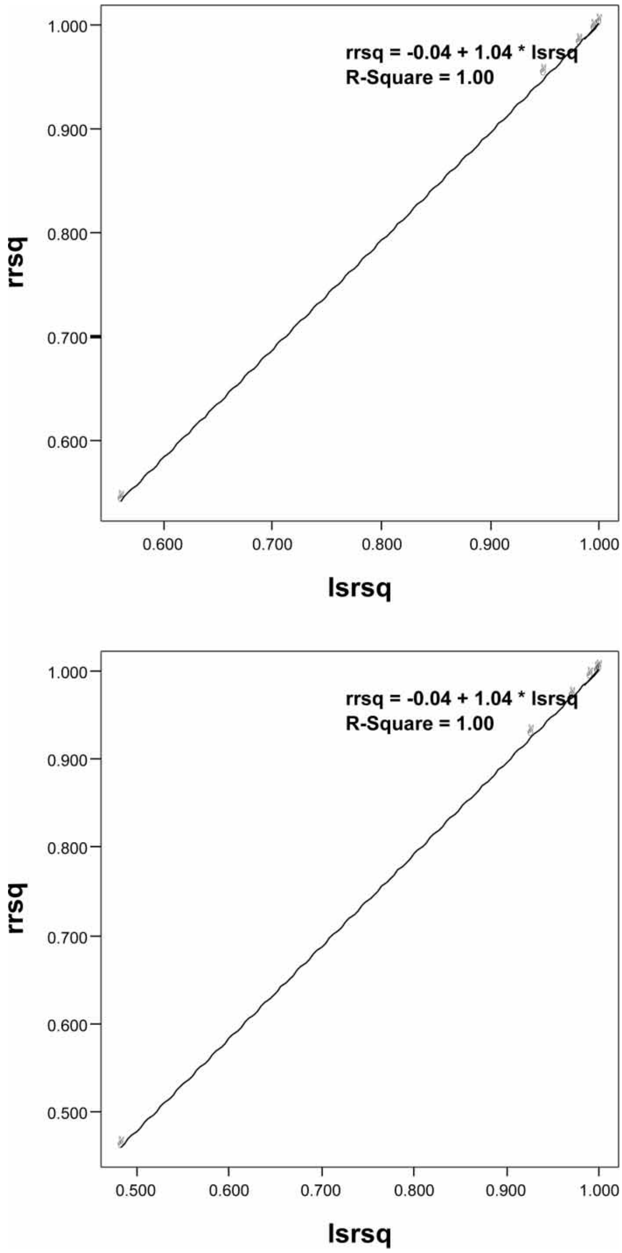
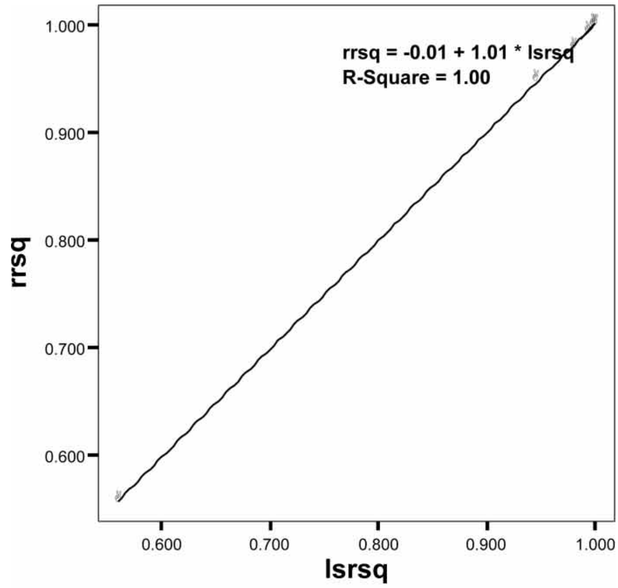


Figure 5. Scatterplot of $\boldsymbol{R}^2$ values of ridge coefficients (**rrsq**) versus $\boldsymbol{R}^2$ values of LS coefficients (**lsrsq**) for $\rho = 0.80$. The top (bottom) graph presents results for $\boldsymbol{\beta}_L$ ($\boldsymbol{\beta}_S$)

0.409 is observed in Table 2 when $\boldsymbol{\beta} = \boldsymbol{\beta}_S$, $\sigma = 1$ and $\rho = 0.99$. The range of values in Table 1 is $0.445 \leq \boldsymbol{R}^2 \leq 1$ and that of Table 2 is $0.420 \leq \boldsymbol{R}^2 \leq 1$.

The $\boldsymbol{R}^2$ values of the ridge coefficients are plotted against the $\boldsymbol{R}^2$ values of the least squares coefficients. Figures 5–8 present the straight-line relationship between the $\boldsymbol{R}^2$ values of the ridge coefficients and the $\boldsymbol{R}^2$ values of the LS coefficients for $\rho = 0.80$, $\rho = 0.90$, $\rho = 0.95$ and $\rho = 0.99$ respectively.



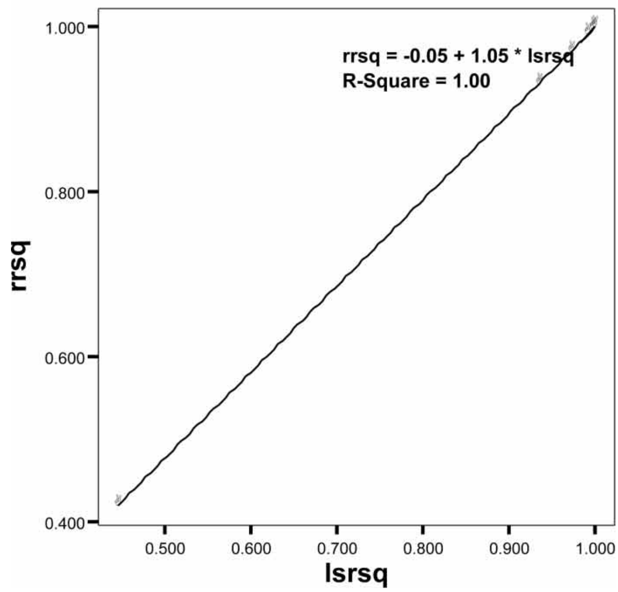
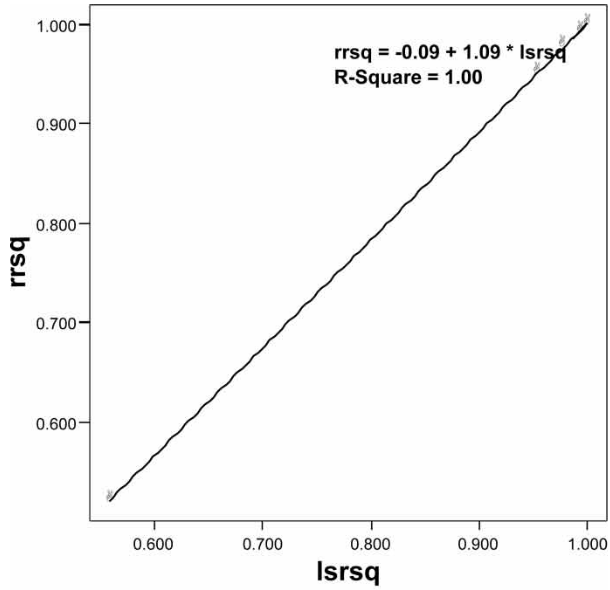


Figure 6. Scatterplot of R^2 values of ridge coefficients (**rrsq**) versus R^2 values of LS coefficients (**lsrsq**) for $\rho = 0.90$. The top (bottom) graph presents results for $\boldsymbol{\beta}_L$ ($\boldsymbol{\beta}_S$).

Figures 5–8 indicates that the difference is minimal in R^2 between Ridge and OLS regression.

The average of the k values observed in five samples are recorded for the ridge estimation procedure in each (ρ, σ, $\boldsymbol{\beta}$) combination. Figures 9–12 present the results for $\rho = 0.80$, $\rho = 0.90$, $\rho = 0.95$ and $\rho = 0.99$ respectively.



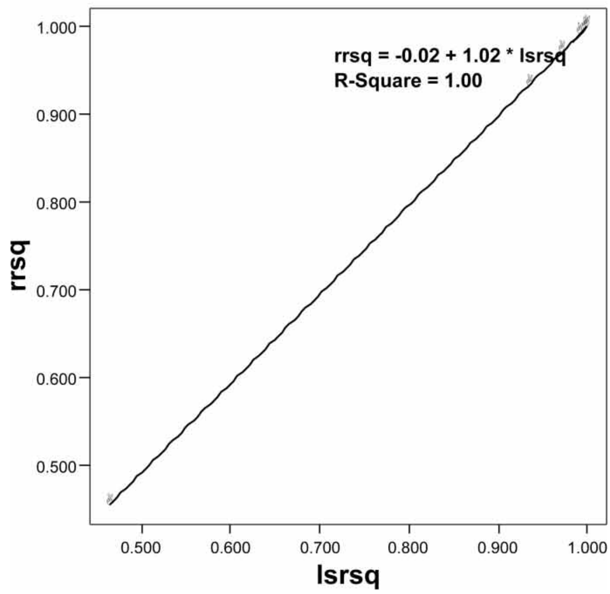


Figure 7. Scatterplot of $\boldsymbol{R}^2$ values of ridge coefficients (**rrsq**) versus $\boldsymbol{R}^2$ values of LS coefficients (**lsrsq**) for $\rho = 0.95$. The top (bottom) graph presents results for $\boldsymbol{\beta}_{\mathrm{L}}$ ($\boldsymbol{\beta}_{\mathrm{S}}$)

For $\boldsymbol{\beta} = \boldsymbol{\beta}_{\mathrm{L}}$, the average k is an increasing function of σ, with the minimum value 2.88×10^{-8} and the maximum value 3.15. Similarly for $\boldsymbol{\beta} = \boldsymbol{\beta}_{\mathrm{S}}$, the average k is an increasing function of σ, with the minimum value 2.83×10^{-8} and the maximum value 2.88.

For $\boldsymbol{\beta} = \boldsymbol{\beta}_{\mathrm{L}}$, the average k is an increasing function of σ, with the minimum value 3.263×10^{-8} and the maximum value 2.716. Similarly for $\boldsymbol{\beta} = \boldsymbol{\beta}_{\mathrm{S}}$, the average k is an

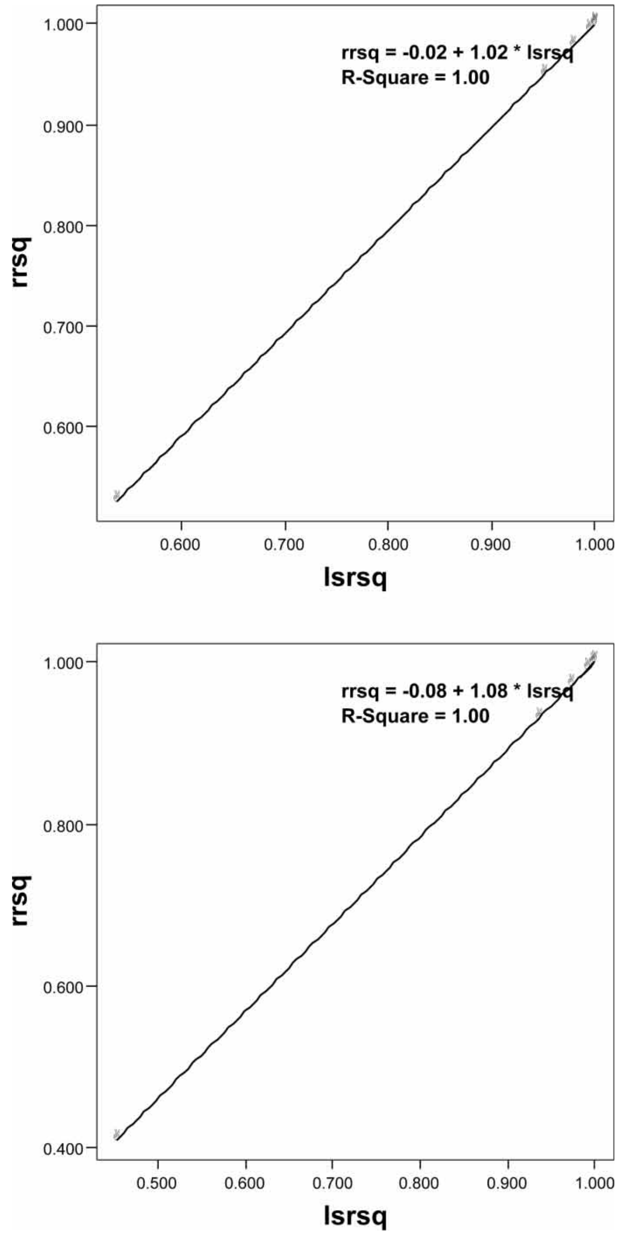
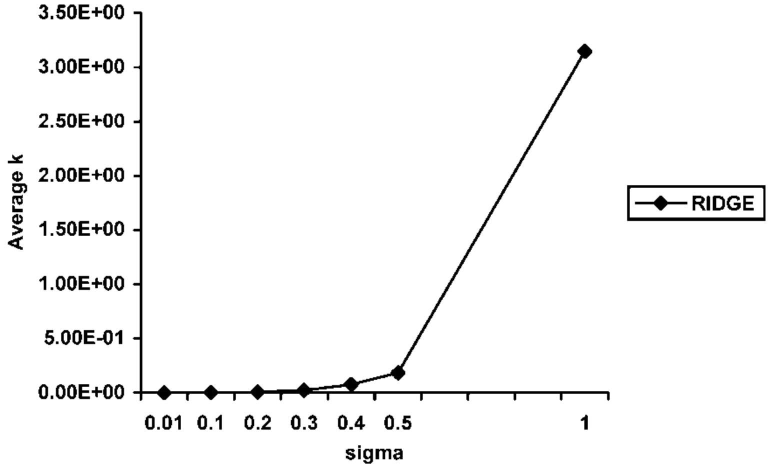


Figure 8. Scatterplot of $\boldsymbol{R}^2$ values of ridge coefficients (**rrsq**) versus $\boldsymbol{R}^2$ values of LS coefficients (**lsrsq**) for $\rho = 0.99$. The top (bottom) graph presents results for $\boldsymbol{\beta}_{\mathrm{L}}$ ($\boldsymbol{\beta}_{\mathrm{S}}$)

increasing function of σ, with the minimum value 2.873×10^{-8} and the maximum value 3.601.

For $\boldsymbol{\beta} = \boldsymbol{\beta}_{\mathrm{L}}$, the average k is an increasing function of σ, with the minimum value 3.097×10^{-8} and the maximum value 2.75. Similarly for $\boldsymbol{\beta} = \boldsymbol{\beta}_{\mathrm{S}}$, the average k is an increasing function of σ, with the minimum value 2.816×10^{-8} and the maximum value 3.33.



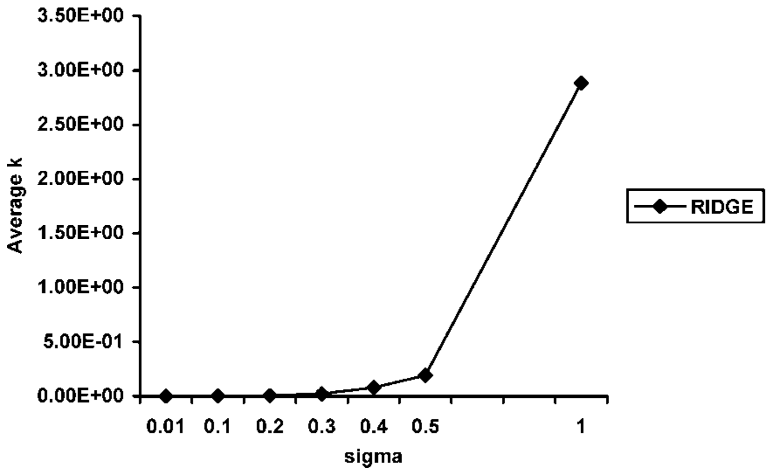


Figure 9. The average k value as a function of σ ($\rho = 0.80$). The top (bottom) graph presents results for $\boldsymbol{\beta}_L$ ($\boldsymbol{\beta}_S$)

For $\boldsymbol{\beta} = \boldsymbol{\beta}_L$, the average k is an increasing function of σ, with the minimum value 3.022×10^{-8} and the maximum value 3.215. Similarly for $\boldsymbol{\beta} = \boldsymbol{\beta}_S$, the average k is an increasing function of σ, with the minimum value 2.945×10^{-8} and the maximum value 3.819.

Concluding Remarks

The performance of the evaluated ridge estimator, as well as the potential performance of any ridge-type estimator, depends on the variance of the random error, the correlations among the explanatory variables and the unknown coefficient vector. The simulations described in this study were conducted in such a way that the performance of the estimators can be observed as one of these factors is changed while the remaining two are fixed. The ridge estimator evaluated in this study has not been shown to be

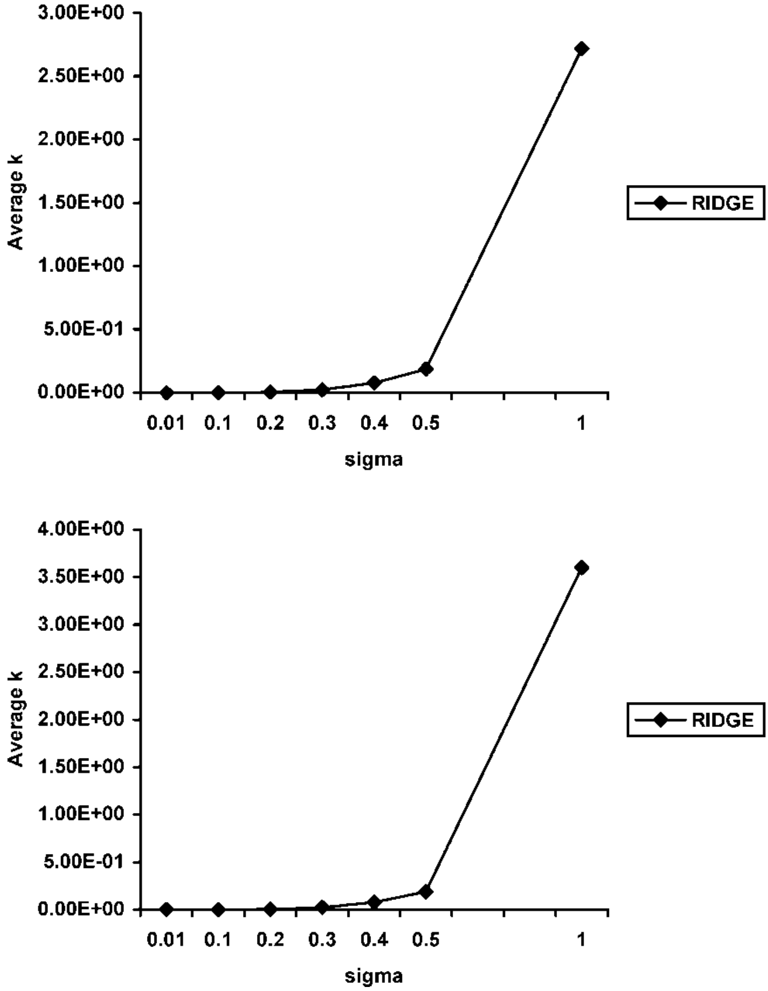


Figure 10. The average k value as a function of σ ($\rho = 0.90$). The top (bottom) graph presents results for $\boldsymbol{\beta}_L$ ($\boldsymbol{\beta}_S$)

better than the least squares estimator in all cases. It achieves a substantial reduction in **MSE** in some cases while increasing the **MSE** somewhat in others. The ratio of the **MSE** for the ridge estimator to the **MSE** for the least squares estimator, M ranges from 0.77 to 1.08 in the cases investigated. The range of the $\boldsymbol{R}^2$ values for the least squares estimator is from 0.445 to 1 while the range of values for the ridge estimator is from 0.420 to 1. The average k value for the ridge estimator is from 2.83×10^{-8} to 3.15.

From the scatterplots of the $\boldsymbol{R}^2$ values of the ridge coefficients against the $\boldsymbol{R}^2$ values of the LS coefficients, the square of the correlation coefficient ($\boldsymbol{R}^2$) for $\rho = 0.80$, $\rho = 0.90$, $\rho = 0.95$ and $\rho = 0.99$ is 1 in each case. This shows that $\boldsymbol{R}^2$ cannot be used to determine which estimator is better in this particular study.

For $\rho = 0.80$ and $\rho = 0.90$, ridge regression has been shown to be better in terms of yielding a lower **MSE** than the least squares estimator. However, when $\rho = 0.95$

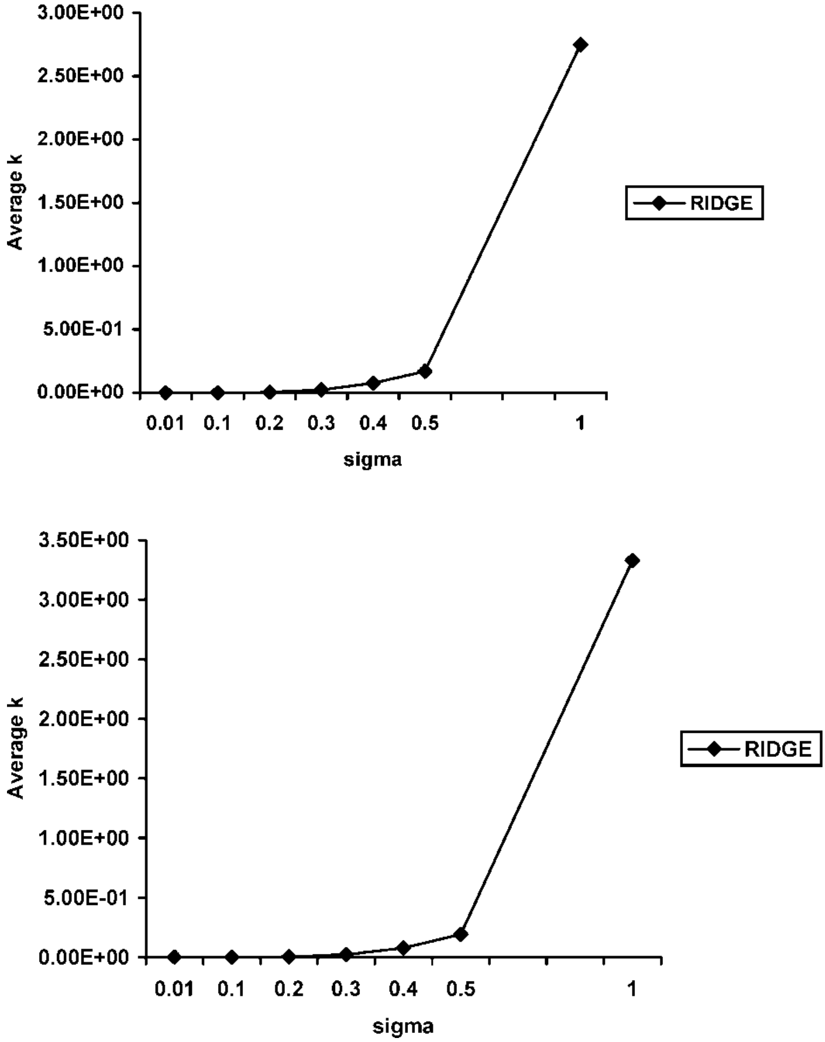
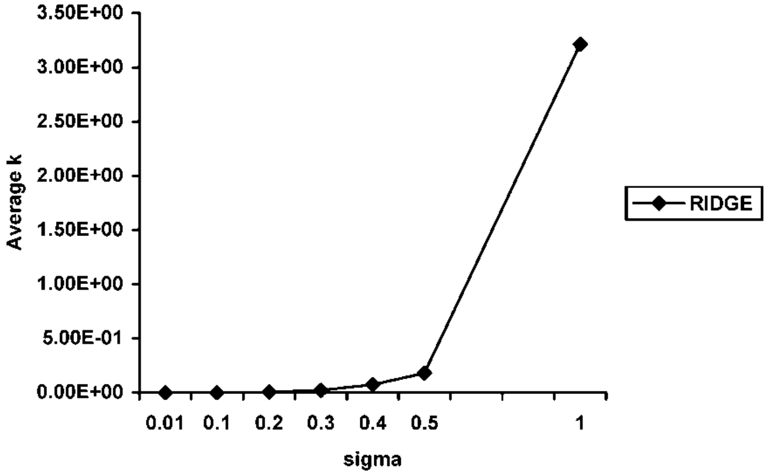


Figure 11. The average k value as a function of σ ($\rho = 0.95$). The top (bottom) graph presents results for $\boldsymbol{\beta}_L$ ($\boldsymbol{\beta}_S$)

and $\rho = 0.99$, the results from the simulations are inconclusive. This indicates that while ridge regression may be effective when multicollinearity is not serious, it is not effective when the explanatory variables are highly correlated. This contrasts sharply with the results obtained in the simulation studies conducted by Hoerl *et al.* (1975) and Gibbons (1981). Their results show that the ridge regression performed better overall; and there were substantial reduction in **MSE** in the cases of severe multicollinearity. Also, better **MSE** estimates were obtained when $\boldsymbol{\beta} = \boldsymbol{\beta}_L$ than when $\boldsymbol{\beta} = \boldsymbol{\beta}_S$.

The lack of agreement between this study and that undertaken by Hoerl *et al.* (1975) and Gibbons (1981) could be attributed to factors like the precision level of the statistical software package, *SPSS* employed in the simulations.



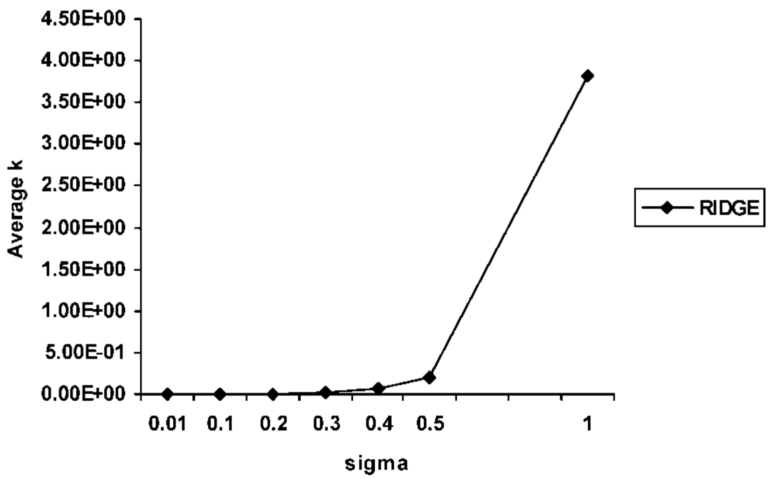


Figure 12. The average k value as a function of σ ($\rho = 0.99$). The top (bottom) graph presents results for $\boldsymbol{\beta}_L$ ($\boldsymbol{\beta}_S$)

In conclusion, this study has not shown conclusively that ridge regression yields better estimates than the least squares regression overall. On the contrary, in cases of a high degree of multicollinearity, the least squares regression sometimes produces better estimates than the ridge regression. Therefore the widespread use of ridge regression in SPSS will not be recommended without further study.

References

Gibbons, D. G. (1981) A simulation study of some ridge estimators, *Journal of the American Statistical Association*, pp. 131–139

Hoerl, A. E. & Kennard, R. W. (1970) Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, 12, pp. 55–82.

Hoerl, A. E., Kennard, R. W. & Baldwin, F. W. (1975) Ridge regression: some simulations, *Communication in Statistics*, 4, pp. 105–123.

Newhouse, J. P. & Oman, S. D. (1971) An evaluation of ridge estimators, *Rand Report*, #R–716–PR, 1–28.

Wichern, D. W. & Churchill, G. A. (1978) A comparison of ridge estimators, *Technometrics*, 20, pp. 301–311.