On Some Ridge Regression Estimators: An Empirical Comparisons

GISELA MUNIZ AND B. M. GOLAM KIBRIA

Department of Statistics, Florida International University, Miami, Florida, USA

In ridge regression analysis, the estimation of the ridge parameter $k$ is an important problem. Many methods are available for estimating such a parameter. This article reviewed and proposed some estimators based on Kibria (2003) and Khalaf and Shukur (2005). A simulation study has been made and mean squared error (MSE) criteria are used to compare the performances of the estimators. We observed that under certain conditions some of the proposed estimators performed well compared to the ordinary least squared (OLS) estimator and some existing popular estimators. Finally, a numerical example has been considered to illustrate the performance of the estimators.

Keywords Bias; Estimation; MSE; Multicollinearity; Ridge regression; Simulation.

Mathematics Subject Classification 62J07; 62F10.

1. Introduction

Consider the standard multiple linear regression

$$y = X\beta + e,$$ 

(1.1)

where $y$ is an $n \times 1$ vector of responses, $X$ is an $n \times p$ observed matrix of the regressors, $\beta$ is an $p \times 1$ vector of unknown parameters, and $e$ is an $n \times 1$ vector of random errors which are distributed as normal with mean vector 0 and covariance matrix $\sigma^2I_n$ and $I_n$ is an identity matrix of order $n$. The ordinary least square estimator (OLS) of the regression coefficients $\beta$ is obtained as $\hat{\beta} = (X'X)^{-1}X'y$, and the covariance matrix of $\hat{\beta}$ is obtained as $\text{Cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}$. Both OLS and its covariance matrix heavily depend on the characteristics of the matrix $X'X$. If the $X'X$ is ill-conditioned, the OLS estimators are sensitive to a number of errors.
For example, some of the regression coefficients may be statistically insignificant or have the wrong sign, and they may result in wide confidence intervals for individual parameters (which are called unstable estimators). With these errors, it is difficult to make valid statistical inferences.

The problem of multicollinearity can be solved by collecting additional data, re-parametrizing the model, and reselecting variables. There are also two mathematical methods: “the principal components regression method” and the “ridge regression method” exist in literature. In this article, we will discuss the ridge regression method. A brief review of the literature reveals an abundance of works related to the ridge regression method. Hoerl and Kennard (1970) first proposed this method to solve the multicollinearity problem. They suggested a small positive number to be added to the diagonal elements of the $X'X$ matrix, and the resulting estimators are obtained as

$$\hat{\beta} = (X'X + kI_p)^{-1}X'y, \quad k \geq 0,$$

this is known as a ridge regression estimator. For a positive value of $k$, this estimator provides a smaller mean squared error (MSE) compared to least squared estimator (LSE). The constant $k$ ($k \geq 0$) is called a “ridge” or “biased” parameter, and it must be estimated using the real data. Although the ridge regression estimator is the most popular method for dealing with multicollinearity, it has drawbacks. The dependence on $k$ tends to result in either instability or bias. As $k \to \infty$, $\hat{\beta}(k) \to 0$ we obtain a stable, but biased estimator of $\beta$. As $k \to 0$, $\hat{\beta}(k) \to$ OLS we obtain an unbiased, but unstable, estimator of $\beta$. The expected distance between $\hat{\beta}(k)$ and $\beta$ must decrease as $k$ increases from 0. The value of $k$ that produces the best estimator, however, is not clear. The estimator $\hat{\beta}(k)$ is a complicated function of $k$. More on ridge regression method, we refer Hoerl and Kennard (1970), Lawless (1978), Saleh and Kibria (1993), Zhang and Ibrahim (2005), Saleh (2006), and recently Alkhamisi and Shukur (2008), among others.

Most of the recent efforts in the area of multicollinearity and ridge regression estimators have concentrated on estimating the value of the ridge parameter $k$. Many different techniques for estimating $k$ have been proposed or suggested by different researchers; to mention a few, Hoerl and Kennard (1970), McDonald and Galarneau (1975), Lawless and Wang (1976), Dempster et al. (1977), Gibbons (1981), Kibria (2003), Khalaf and Shukur (2005), and very recently Alkhamisi and Shukur (2008).

The purpose of this article is to apply new algorithms such as arithmetic mean, geometric mean, maximum, and square root to the approach obtained by Khalaf and Shukur (2005) and Kibria (2003) in order to define some new estimators, which are transformations of Khalaf and Shukur and Kibria works. A Monte Carlo simulation study will be made to compare the performances of the proposed estimators. The organization of the article is as follows. In Sec. 2, we review and propose some new estimators for estimating the ridge parameter $k$. A Monte Carlo simulation study has been made in Sec. 3. An application has been added in Sec. 4. Finally, some concluding remarks are presented in Sec. 5.

2. Statistical Methodology

In this section, we will discuss the statistical methodology used to analyze the estimation of the ridge parameter.
2.1. **Notations and Some Preliminaries**

Suppose there exists an orthogonal matrix \( D \) such that \( D'CD = \Lambda \), where \( \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p) \) contains the eigenvalues of the matrix \( C = X'X \). The orthogonal (canonical form) version of the multiple regression model (1.1) is

\[
Y = X'\alpha + e, \quad (2.1)
\]

where \( X^* = XD \) and \( \alpha = D'\beta \).

But, when the matrix \( X'X \) is ill conditioned (in the sense of there is a near-linear dependency among the columns of the matrix), the OLS of \( \beta \) has a large variance, and multicollinearity is said to be present. If multicollinearity is present, at least one eigen value \( \lambda_i \cong 0 \). The nearness to 0 of the smallest eigen value measures the strength of the linear multicollinearity. To make the behavior of \( X'X \) matrix more like the orthogonal case, we need to increase the eigen values. Ridge regression replace \( X'X \) with \( X'X + kI \), \((k > 0)\), which is the same as replacing the \( \lambda_i \) by \( \lambda_i + k \). This replacement counters the damaging effect of the smallest eigen value.

Then the generalized ridge regression estimators of \( \alpha \) are given as

\[
\hat{\alpha}(k) = (X'^*X^* + kI_p)^{-1}X'^*Y = (I_p + K (X'^*X^*))^{-1}\hat{\alpha} \quad (2.2)
\]

where \( K = \text{diag}(k_1, k_2, \ldots, k_p) \), \( k_i \geq 0 \), and \( \hat{\alpha} = \Lambda^{-1}X'^*Y \) is the ordinary least squares (OLS) estimates of \( \alpha \). It follows from Hoerl and Kennard (1970) that the value of \( k_i \) which minimizes the MSE(\( \hat{\alpha}(k) \)) is

\[
k_i = \frac{\sigma^2}{\hat{\alpha}_i} \quad (2.3)
\]

where \( \sigma^2 \) represents the error variance of the multiple regression model, and \( \alpha_i \) is the \( i \)th element of \( \alpha \).

2.2. **Some Ridge Estimators**

In this section, we reviewed and proposed some ridge parameters.

2.2.1. **Estimator Based on Hoerl and Kennard (1970).** Hoerl and Kennard (1970) suggested to replace \( \sigma^2 \) and \( \hat{\alpha}_i^2 \) by their corresponding unbiased estimators \( \hat{\sigma}^2 \) and \( \hat{\alpha}_i \), respectively, in (2.3). That is,

\[
\hat{k}_i = \hat{\sigma}^2 \hat{\alpha}_i, \quad (2.4)
\]

where \( \hat{\sigma}^2 = \frac{\sum e_i^2}{n-p} = \frac{(n-p)(n-1)}{n-p} \) is the residual mean square error, which is unbiased estimator of \( \sigma^2 \). Hoerl and Kennard (1970) suggested \( k \) to be

\[
k_{HK1} = \hat{k}_{HK} = \frac{\hat{\sigma}^2}{\hat{\alpha}_{\max}^2} \quad (2.5)
\]

where \( \hat{\alpha}_{\max} \) is the maximum element of \( \hat{\alpha} \). If \( \sigma^2 \) and \( \alpha \) are known, then \( \hat{k}_{HK} \) will give smaller MSE than the OLS.
2.2.2. Estimators Based on Kibria (2003). Kibria (2003) proposed some new estimators based on generalized ridge regression approach, by using the geometric mean of $\hat{k}_i$, which produces the following estimator:

$$k_{K1} = \hat{k}_{GM} = \frac{\hat{\sigma}^2}{\left(\prod_{i=1}^{p} \hat{\alpha}_i\right)^{\frac{1}{2}}}.$$  (2.6)

By using the median of $\hat{k}_i$, which produces the following estimator for $p \geq 3$,

$$k_{K2} = \hat{k}_{MED} = \text{Median}\left\{\frac{\hat{\sigma}^2}{\hat{\alpha}_i}\right\}, \quad i = 1, 2, \ldots, p.$$  (2.7)

Kibria (2003) also proposed another estimator based on the arithmetic mean. Since the estimator was not stable and not suggested by the author, we therefore omitted it.

2.2.3. Estimators Based on Khalaf and Shukur (2005). Khalaf and Shukur (2005) suggested a new method to estimate the ridge parameter $k$ as a modification of $k_{HK}$ and is given as

$$k_{S1} = \hat{k}_{KS} = \frac{t_{\max} \hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + t_{\max} \hat{\alpha}_{\max}^2}$$  (2.8)

where $t_{\max}$ is the maximum eigenvalue of $X'X$ matrix.

Following Kibria (2003) and Khalaf and Shukur (2005), Alkhamisi et al. (2006) proposed the following estimators:

$$k_{S2} = \hat{k}_{KS}^{\text{arith}} = \frac{1}{p} \sum_{i=1}^{p} \left(\frac{t_i \hat{\sigma}_i^2}{(n-p)\hat{\sigma}_i^2 + t_i \hat{\alpha}_i^2}\right), \quad i = 1, 2, \ldots, p$$  (2.9)

$$k_{S3} = \hat{k}_{KS}^{\text{max}} = \max\left(\frac{t_i \hat{\sigma}_i^2}{(n-p)\hat{\sigma}_i^2 + t_i \hat{\alpha}_i^2}\right), \quad i = 1, 2, \ldots, p$$  (2.10)

$$k_{S4} = \hat{k}_{KS}^{\text{md}} = \text{median}\left(\frac{t_i \hat{\sigma}_i^2}{(n-p)\hat{\sigma}_i^2 + t_i \hat{\alpha}_i^2}\right)$$  (2.11)

2.2.4. Some Proposed New Estimators. We applied the algorithms of geometric mean and square root to the approach obtained by Khalaf and Shukur (2005) and Kibria (2003) in order to find new estimators. The idea of the square root transformation was taken from Alkhamisi and Shukur (2008). First, following Kibria (2003), by using the geometric mean of $\left(\frac{t_i \hat{\sigma}_i^2}{(n-p)\hat{\sigma}_i^2 + t_i \hat{\alpha}_i^2}\right)$ which produces the following estimator,

$$k_{KM1} = \hat{k}_{GM}^{KS} = \left(\prod_{i=1}^{p} \frac{t_i \hat{\sigma}_i^2}{(n-p)\hat{\sigma}_i^2 + t_i \hat{\alpha}_i^2}\right)^{\frac{1}{p}}.$$  (2.12)
Now, suppose $m_i = \sqrt{\frac{\hat{\sigma}_i}{\lambda_i}}$; then following Alkhamisi and Shukur (2008), we propose the following estimators based on the square root transformations:

$$k_{KM2} = \max \left( 1, \frac{1}{m_i} \right), \quad k_{KM3} = \max (m_i), \quad k_{KM4} = \left( \prod_{i=1}^{p} \frac{1}{m_i} \right)^{\frac{1}{p}},$$

$$k_{KM5} = \left( \prod_{i=1}^{p} m_i \right)^{\frac{1}{p}}, \quad k_{KM6} = \text{median} \left( \frac{1}{m_i} \right), \quad k_{KM7} = \text{median} (m_i).$$

### 2.3. Criteria for Measuring Goodness of an Estimator

We use the mean squared error (MSE) criteria to measure the goodness of an estimator. Ridge estimators are proposed to have smaller MSE compared to OLS.

The ridge estimator of $\beta$ is $\hat{\beta}(k) = (X'X + kI_p)^{-1}X'y$, which is a biased estimator and the covariance matrix of $\hat{\beta}(k)$ is $\text{Var}(\hat{\beta}(k)) = \sigma^2(X'X + kI_p)^{-1}X'X(X'X + kI)^{-1}$. The mean squared error (MSE) of the ridge estimator is

$$\text{MSE} = \text{Var}(\hat{\beta}(k)) + \left[ \text{bias in } \hat{\beta}(k) \right]^2 = \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + k_i)^2} + \sum_{i=1}^{p} k_i^2 \frac{\lambda_i}{(\lambda_i + k)^2},$$

where $\lambda_1, \lambda_2, \ldots, \lambda_p$ are eigen values of $X'X$. The first term on the right 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 $k$. As $k$ increases, however, the variance decreases. We would like to select $k$ such that the reduction in the variance term is greater than the increase in the squared bias.

### 3. Monte Carlo Simulation

In this section, a simulation study has been conducted to compare the performance of the estimators. Even though we reviewed 14 different estimators for estimating the ridge parameter $k$, based on the preliminary simulation study, the following 10 estimators are considered in the final simulation study: $HK, K1, K2, S3, S4$, and proposed $KM1, KM2, KM4, KM5$, and $KM6$. We used the S-plus 8.0 software to complete the simulation study.

#### 3.1. Generation of Explanatory and Dependent Variables

To achieve different degrees of collinearity, following Gibbons (1981), the explanatory variable were generated using the following equation:

$$x_{ij} = \left( 1 - \gamma^2 \right)^{(1/2)} z_{ij} + \gamma z_{ip} \quad i = 1, 2, \ldots, n \quad j = 1, 2, \ldots, p, \quad (3.1)$$

where $\gamma$ represents the correlation between any two explanatory variables and $z_{ij}$ are independent standard normal pseudo-random numbers. The $n$ observations for the dependent variable are determined by the following equation:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i \quad i = 1, 2, \ldots, n, \quad (3.2)$$
where $e_i$ are independent normal pseudo-random numbers with mean 0 and variance $\sigma^2$. For our convenience, $\beta_0$ is taken to be zero.

### 3.2. Replications, Sample Size, and Correlation Coefficient

A number of factors that can affect the properties of the parameter estimators: the correlation between the explanatory variables ($\gamma$), sample size ($n$), error standard deviations ($\sigma$), and number of replications ($r$). We consider $p = 2$ independent variables for the simpler computation. Since our primary interest lies in the performance of our proposed estimators according to the strength of the multicollinearity, we used different degrees of correlation between the variables. Three different set of correlations are: $\gamma = 0.70, 0.80, 0.90$, sample size $n = 10, 20, 30, 40, 50, \text{ and } 100$, $p = 2$ and $\sigma = 0.01, 0.5, 1, 5$. For each set of explanatory variables we considered the coefficient vector that corresponded to the largest eigenvalue of $X'X$ matrix subject to the constraint that $\beta'\beta = 1$. Newhouse and Oman (1971) stated that if the mean squared error (MSE) is a function of $\beta$, $\sigma^2$, and $k$, and if the explanatory variables are fixed, then the MSE is minimized when we choose this coefficient vector. For given values of $n$, $p$, $\beta$, $\lambda$, $\gamma$, and $\sigma$, the set of explanatory variables are generated. Then the experiment was repeated 2,000 times by generating new error terms. Then the values of the ridge parameters $k$ of the different proposed estimators and the corresponding ridge estimators as well as the average MSE’s are estimated. For details about the simulation study, we refer to Gibbons (1981) and Kibria (2003), among others.

### 3.3. Condition Number and VIF

Eigenvalues and eigenvectors of the correlation matrix indicate the degree of multicollinearity. An eigenvalue that approaches to zero indicates a very strong linear dependency between regressors, while the elements of the associated eigenvector display the weights of the corresponding regressor variables in the multicollinearity. Multicollinearity can be measured in terms of the ratio of the largest and the smallest eigenvalue. This quantity is called the condition number of the correlation matrix:

$$\kappa = (\lambda_{\text{max}}/\lambda_{\text{min}})^{1/2}. \tag{3.3}$$

Large values of $\kappa$ are an indication of serious multicollinearity. Condition number of the correlation matrix $X$ is between 30–100 indicates a moderate to strong correlation and a condition numbers greater than 100 suggest severe multicollinearity (Liu, 2003). The number of eigenvalues near zero indicates the number of collinearities detected among the regressor variables.

### 3.4. Results Discussions

The simulated MSEs are presented for different sample size $n$, $\gamma$, and $\sigma$ in Tables 1 and 2. From these tables, we observed that for smaller sigma ($\sigma = 0.01$) the change in the correlation between the explanatory variables had almost no effect on the MSE of the estimators. In all situations they remained almost the same for any sample size or number of parameters, and their MSEs are very small. When $\sigma$
Table 1

For \( n = 10, 20, \ p = 2 \) and different values of \( \gamma \) and \( \sigma \)

<table>
<thead>
<tr>
<th>( \sigma )</th>
<th>OLS</th>
<th>HK</th>
<th>K1</th>
<th>K2</th>
<th>S3</th>
<th>S4</th>
<th>KM1</th>
<th>KM2</th>
<th>KM3</th>
<th>KM4</th>
<th>KM5</th>
<th>KM6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n = 10, \ p = 2, \ \gamma = 0.70 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.427</td>
<td>0.426</td>
<td>0.426</td>
<td>0.426</td>
<td>0.426</td>
<td>0.389</td>
<td>0.125</td>
<td>0.129</td>
<td>0.272</td>
<td>0.125</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.765</td>
<td>0.516</td>
<td>0.336</td>
<td>0.269</td>
<td>0.517</td>
<td>0.517</td>
<td>0.304</td>
<td>0.172</td>
<td>0.171</td>
<td>0.189</td>
<td>0.172</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.799</td>
<td>1.051</td>
<td>0.438</td>
<td>0.297</td>
<td>1.042</td>
<td>1.042</td>
<td>0.479</td>
<td>0.270</td>
<td>0.223</td>
<td>0.215</td>
<td>0.270</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>36.39</td>
<td>18.75</td>
<td>3.898</td>
<td>1.701</td>
<td>18.36</td>
<td>18.36</td>
<td>6.025</td>
<td>2.733</td>
<td>1.446</td>
<td>1.204</td>
<td>2.733</td>
<td></td>
</tr>
<tr>
<td>( n = 10, \ p = 2, \ \gamma = 0.80 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.073</td>
<td>0.073</td>
<td>0.073</td>
<td>0.073</td>
<td>0.073</td>
<td>0.073</td>
<td>0.057</td>
<td>0.121</td>
<td>0.094</td>
<td>0.025</td>
<td>0.121</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.438</td>
<td>0.244</td>
<td>0.117</td>
<td>0.084</td>
<td>0.243</td>
<td>0.243</td>
<td>0.087</td>
<td>0.065</td>
<td>0.053</td>
<td>0.047</td>
<td>0.065</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>1.608</td>
<td>0.866</td>
<td>0.133</td>
<td>0.080</td>
<td>0.840</td>
<td>0.840</td>
<td>0.172</td>
<td>0.084</td>
<td>0.082</td>
<td>0.121</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( n = 10, \ p = 2, \ \gamma = 0.90 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.073</td>
<td>0.073</td>
<td>0.073</td>
<td>0.073</td>
<td>0.073</td>
<td>0.073</td>
<td>0.052</td>
<td>0.120</td>
<td>0.094</td>
<td>0.020</td>
<td>0.120</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.618</td>
<td>0.329</td>
<td>0.125</td>
<td>0.076</td>
<td>0.313</td>
<td>0.313</td>
<td>0.087</td>
<td>0.063</td>
<td>0.051</td>
<td>0.043</td>
<td>0.063</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>2.445</td>
<td>1.267</td>
<td>0.296</td>
<td>0.137</td>
<td>1.191</td>
<td>1.191</td>
<td>0.270</td>
<td>0.133</td>
<td>0.195</td>
<td>0.133</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>61.65</td>
<td>31.81</td>
<td>4.630</td>
<td>1.542</td>
<td>30.49</td>
<td>30.49</td>
<td>7.311</td>
<td>2.890</td>
<td>1.224</td>
<td>1.222</td>
<td>2.890</td>
<td></td>
</tr>
<tr>
<td>( n = 20, \ p = 2, \ \gamma = 0.70 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.035</td>
<td>0.032</td>
<td>0.032</td>
<td>0.032</td>
<td>0.032</td>
<td>0.032</td>
<td>0.030</td>
<td>0.032</td>
<td>0.032</td>
<td>0.020</td>
<td>0.032</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.090</td>
<td>0.061</td>
<td>0.045</td>
<td>0.041</td>
<td>0.072</td>
<td>0.072</td>
<td>0.043</td>
<td>0.029</td>
<td>0.025</td>
<td>0.025</td>
<td>0.029</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.264</td>
<td>0.152</td>
<td>0.080</td>
<td>0.064</td>
<td>0.209</td>
<td>0.209</td>
<td>0.119</td>
<td>0.059</td>
<td>0.047</td>
<td>0.043</td>
<td>0.059</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>5.529</td>
<td>2.823</td>
<td>1.010</td>
<td>0.645</td>
<td>4.299</td>
<td>4.299</td>
<td>2.393</td>
<td>1.198</td>
<td>0.853</td>
<td>0.524</td>
<td>0.198</td>
<td></td>
</tr>
<tr>
<td>( n = 20, \ p = 2, \ \gamma = 0.80 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.189</td>
<td>0.189</td>
<td>0.189</td>
<td>0.189</td>
<td>0.189</td>
<td>0.189</td>
<td>0.174</td>
<td>0.055</td>
<td>0.057</td>
<td>0.123</td>
<td>0.055</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.334</td>
<td>0.225</td>
<td>0.149</td>
<td>0.127</td>
<td>0.255</td>
<td>0.255</td>
<td>0.156</td>
<td>0.077</td>
<td>0.077</td>
<td>0.086</td>
<td>0.077</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.776</td>
<td>0.444</td>
<td>0.182</td>
<td>0.127</td>
<td>0.535</td>
<td>0.535</td>
<td>0.265</td>
<td>0.117</td>
<td>0.099</td>
<td>0.098</td>
<td>0.117</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>13.77</td>
<td>6.66</td>
<td>1.312</td>
<td>0.619</td>
<td>8.96</td>
<td>8.96</td>
<td>3.500</td>
<td>1.265</td>
<td>0.703</td>
<td>0.521</td>
<td>1.265</td>
<td></td>
</tr>
<tr>
<td>( n = 20, \ p = 2, \ \gamma = 0.90 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.032</td>
<td>0.032</td>
<td>0.032</td>
<td>0.032</td>
<td>0.032</td>
<td>0.032</td>
<td>0.025</td>
<td>0.032</td>
<td>0.032</td>
<td>0.013</td>
<td>0.032</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.190</td>
<td>0.104</td>
<td>0.049</td>
<td>0.035</td>
<td>0.121</td>
<td>0.121</td>
<td>0.047</td>
<td>0.028</td>
<td>0.023</td>
<td>0.020</td>
<td>0.028</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.666</td>
<td>0.340</td>
<td>0.094</td>
<td>0.056</td>
<td>0.419</td>
<td>0.419</td>
<td>0.155</td>
<td>0.050</td>
<td>0.037</td>
<td>0.037</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>15.76</td>
<td>7.692</td>
<td>1.441</td>
<td>0.599</td>
<td>9.601</td>
<td>9.601</td>
<td>3.440</td>
<td>1.303</td>
<td>0.635</td>
<td>0.457</td>
<td>1.303</td>
<td></td>
</tr>
</tbody>
</table>

increases the higher correlation between the X’s resulted in an increase of the MSEs. In general, \( k_{KM4}, k_{KM5}, k_{KM6} \), and \( k_{K1}, k_{K2} \) performed better than others. When the sample size increases the MSE also decreases, or remained almost the same. Even \( \gamma \) and \( \sigma \) have large values, the increase in the sample size resulted in a considerable decrease of the MSE of the estimators. Again in this situation, the performance of \( k_{KM4}, k_{KM5}, k_{KM6}, \) and \( k_{K1}, k_{K2} \) are better than the rest of the estimators. When the value of \( \sigma \) increases, the MSEs also increase. For all \( \sigma \), the performance of the proposed estimators are better than the OLS in the sense of smaller MSE. However, among the proposed estimators, \( k_{KM4}, k_{KM5}, k_{KM6}, \) and \( k_{K1}, k_{K2} \) performed better than the rest. This behavior was almost constant for any sample size, and number of variables considered.
To illustrate the performance of the estimators, we consider the famous Portland cement data originally due to Woods et al. (1932). This data have been analyzed by several researchers: Hald (1952), Daniel and Wood (1980), Liu (2003), and very recently, Sakallıoğlu and Kaçıranlar (2008), among others. The data come from an experimental investigation of the heat evolved during the setting and hardening of Portland cement of varied composition and the dependence of this heat on the percentages of four compounds in the clinkers from which the cement was made. There are four explanatory variables: $X_1$: amount of tricalcium aluminate, $X_2$: amount of ticalcium silicate, $X_3$: amount of tetracalcium alumino ferrite, and $X_4$: amount of dicalcium silicate. The response variable is $Y$: heat evolved in calories per

### Table 2

For $n = 30, 100, p = 2$ and different values of $\gamma$ and $\sigma$

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>OLS</th>
<th>HK</th>
<th>K1</th>
<th>K2</th>
<th>S3</th>
<th>S4</th>
<th>KM1</th>
<th>KM2</th>
<th>KM3</th>
<th>KM4</th>
<th>KM5</th>
<th>KM6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 30, p = 2, \gamma = 0.70$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.122</td>
<td>0.121</td>
<td>0.121</td>
<td>0.121</td>
<td>0.121</td>
<td>0.117</td>
<td>0.035</td>
<td>0.037</td>
<td>0.093</td>
<td>0.035</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.161</td>
<td>0.118</td>
<td>0.089</td>
<td>0.078</td>
<td>0.145</td>
<td>0.145</td>
<td>0.107</td>
<td>0.056</td>
<td>0.055</td>
<td>0.060</td>
<td>0.056</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.279</td>
<td>0.171</td>
<td>0.098</td>
<td>0.080</td>
<td>0.240</td>
<td>0.240</td>
<td>0.158</td>
<td>0.083</td>
<td>0.073</td>
<td>0.066</td>
<td>0.083</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>4.417</td>
<td>2.277</td>
<td>0.757</td>
<td>0.465</td>
<td>3.654</td>
<td>3.654</td>
<td>2.041</td>
<td>0.795</td>
<td>0.553</td>
<td>0.402</td>
<td>0.795</td>
<td></td>
</tr>
<tr>
<td>$n = 30, p = 2, \gamma = 0.80$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.021</td>
<td>0.020</td>
<td>0.020</td>
<td>0.020</td>
<td>0.020</td>
<td>0.020</td>
<td>0.018</td>
<td>0.034</td>
<td>0.027</td>
<td>0.009</td>
<td>0.034</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.084</td>
<td>0.051</td>
<td>0.031</td>
<td>0.025</td>
<td>0.065</td>
<td>0.065</td>
<td>0.032</td>
<td>0.017</td>
<td>0.014</td>
<td>0.013</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.273</td>
<td>0.143</td>
<td>0.053</td>
<td>0.037</td>
<td>0.207</td>
<td>0.207</td>
<td>0.099</td>
<td>0.038</td>
<td>0.027</td>
<td>0.025</td>
<td>0.038</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>6.236</td>
<td>3.051</td>
<td>0.788</td>
<td>0.430</td>
<td>4.653</td>
<td>4.653</td>
<td>2.168</td>
<td>0.762</td>
<td>0.469</td>
<td>0.329</td>
<td>0.762</td>
<td></td>
</tr>
<tr>
<td>$n = 30, p = 2, \gamma = 0.90$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.021</td>
<td>0.020</td>
<td>0.020</td>
<td>0.020</td>
<td>0.020</td>
<td>0.020</td>
<td>0.016</td>
<td>0.034</td>
<td>0.027</td>
<td>0.007</td>
<td>0.034</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.123</td>
<td>0.066</td>
<td>0.030</td>
<td>0.022</td>
<td>0.087</td>
<td>0.087</td>
<td>0.035</td>
<td>0.017</td>
<td>0.014</td>
<td>0.013</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.449</td>
<td>0.230</td>
<td>0.061</td>
<td>0.034</td>
<td>0.304</td>
<td>0.304</td>
<td>0.116</td>
<td>0.034</td>
<td>0.023</td>
<td>0.023</td>
<td>0.034</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>10.65</td>
<td>5.177</td>
<td>0.686</td>
<td>0.382</td>
<td>7.235</td>
<td>7.235</td>
<td>2.178</td>
<td>0.754</td>
<td>0.385</td>
<td>0.301</td>
<td>0.754</td>
<td></td>
</tr>
<tr>
<td>$n = 100, p = 2, \gamma = 0.70$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.010</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.005</td>
<td>0.009</td>
<td>0.008</td>
<td>0.004</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.014</td>
<td>0.010</td>
<td>0.008</td>
<td>0.007</td>
<td>0.013</td>
<td>0.013</td>
<td>0.010</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.039</td>
<td>0.023</td>
<td>0.013</td>
<td>0.011</td>
<td>0.038</td>
<td>0.038</td>
<td>0.028</td>
<td>0.011</td>
<td>0.009</td>
<td>0.008</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>0.812</td>
<td>0.412</td>
<td>0.163</td>
<td>0.113</td>
<td>0.779</td>
<td>0.779</td>
<td>0.572</td>
<td>0.221</td>
<td>0.165</td>
<td>0.102</td>
<td>0.221</td>
<td></td>
</tr>
<tr>
<td>$n = 100, p = 2, \gamma = 0.80$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.010</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
<td>0.005</td>
<td>0.009</td>
<td>0.007</td>
<td>0.003</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.016</td>
<td>0.010</td>
<td>0.008</td>
<td>0.007</td>
<td>0.015</td>
<td>0.015</td>
<td>0.010</td>
<td>0.005</td>
<td>0.004</td>
<td>0.004</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.048</td>
<td>0.027</td>
<td>0.013</td>
<td>0.011</td>
<td>0.045</td>
<td>0.045</td>
<td>0.031</td>
<td>0.010</td>
<td>0.008</td>
<td>0.008</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>1.026</td>
<td>0.518</td>
<td>0.178</td>
<td>0.114</td>
<td>0.968</td>
<td>0.968</td>
<td>0.653</td>
<td>0.212</td>
<td>0.150</td>
<td>0.096</td>
<td>0.212</td>
<td></td>
</tr>
<tr>
<td>$n = 100, p = 2, \gamma = 0.90$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>0.036</td>
<td>0.034</td>
<td>0.034</td>
<td>0.034</td>
<td>0.034</td>
<td>0.034</td>
<td>0.032</td>
<td>0.010</td>
<td>0.010</td>
<td>0.023</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0.058</td>
<td>0.039</td>
<td>0.026</td>
<td>0.021</td>
<td>0.053</td>
<td>0.053</td>
<td>0.036</td>
<td>0.014</td>
<td>0.014</td>
<td>0.016</td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>0.118</td>
<td>0.067</td>
<td>0.029</td>
<td>0.021</td>
<td>0.106</td>
<td>0.106</td>
<td>0.064</td>
<td>0.021</td>
<td>0.018</td>
<td>0.017</td>
<td>0.021</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>2.115</td>
<td>1.005</td>
<td>0.222</td>
<td>0.118</td>
<td>1.861</td>
<td>1.861</td>
<td>0.960</td>
<td>0.226</td>
<td>0.132</td>
<td>0.099</td>
<td>0.226</td>
<td></td>
</tr>
</tbody>
</table>
Table 3
The MSE and the estimated regression coefficients of the estimators

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>HK</th>
<th>K1</th>
<th>K2</th>
<th>S3</th>
<th>S4</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>4912.09</td>
<td>2170.96</td>
<td>3883.604</td>
<td>3886.002</td>
<td>3891.049</td>
<td>3886.001</td>
</tr>
<tr>
<td>\hat{\beta}_0</td>
<td>62.4054</td>
<td>27.6068</td>
<td>0.1003</td>
<td>0.0809</td>
<td>0.0373</td>
<td>0.0809</td>
</tr>
<tr>
<td>\hat{\beta}_1</td>
<td>1.5511</td>
<td>1.9090</td>
<td>2.1725</td>
<td>2.1619</td>
<td>1.3973</td>
<td>2.1619</td>
</tr>
<tr>
<td>\hat{\beta}_2</td>
<td>0.5102</td>
<td>0.8688</td>
<td>1.1568</td>
<td>1.1595</td>
<td>1.3253</td>
<td>1.15949</td>
</tr>
<tr>
<td>\hat{\beta}_3</td>
<td>0.1019</td>
<td>0.4680</td>
<td>0.7435</td>
<td>0.7359</td>
<td>0.2486</td>
<td>0.7359</td>
</tr>
<tr>
<td>\hat{\beta}_4</td>
<td>-0.1441</td>
<td>0.2075</td>
<td>0.4882</td>
<td>0.4901</td>
<td>0.5962</td>
<td>0.4901</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>KM1</th>
<th>KM2</th>
<th>KM4</th>
<th>KM5</th>
<th>KM6</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>4912.09</td>
<td>3878.71</td>
<td>3889.981</td>
<td>3878.96</td>
<td>3878.96</td>
<td>3876.138</td>
</tr>
<tr>
<td>\hat{\beta}_0</td>
<td>62.4054</td>
<td>0.1398</td>
<td>0.0463</td>
<td>0.1377</td>
<td>0.1107</td>
<td>0.1605</td>
</tr>
<tr>
<td>\hat{\beta}_1</td>
<td>1.5511</td>
<td>2.1804</td>
<td>1.9050</td>
<td>2.1801</td>
<td>2.1756</td>
<td>2.1822</td>
</tr>
<tr>
<td>\hat{\beta}_2</td>
<td>0.5102</td>
<td>1.1545</td>
<td>1.2183</td>
<td>1.1546</td>
<td>1.1560</td>
<td>1.1538</td>
</tr>
<tr>
<td>\hat{\beta}_3</td>
<td>0.1019</td>
<td>0.7489</td>
<td>0.5568</td>
<td>0.7488</td>
<td>0.7456</td>
<td>0.7502</td>
</tr>
<tr>
<td>\hat{\beta}_4</td>
<td>-0.1441</td>
<td>0.4866</td>
<td>0.5284</td>
<td>0.4867</td>
<td>0.4877</td>
<td>0.4861</td>
</tr>
</tbody>
</table>

gram of cement. We consider the following linear regression model:

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon. \]  

First, we fit a homogeneous linear model without intercept to the data following Woods et al. (1932). Under this model the condition number \( (\kappa = (\lambda_{\text{max}}/\lambda_{\text{min}})^{1/2}) \) is 20.58, which indicates a weak collinearity. Now following Hald (1952) and Daniel and Wood (1980), we fit an inhomogeneous linear model with intercept to the data. The eigenvalues for the \( X \) matrix are 44676.2059, 5965.4221, 809.9521, 105.4187, and 0.0012, and the condition number is 6056.4, which implies the existence of multicollinearity in the data set. The estimated MSEs along with the ridge regression coefficients are presented in Table 3. From the table we observed that the sign for regressor \( X_4 \) has been changed and all proposed estimators performed better than the OLS in the sense that they have smaller MSE. None of them performed better than Hoerl and Kennard. Nevertheless, \( k_{KM1}, k_{KM3}, \) and \( k_{KM6} \) performed equivalently well and better than the other proposed estimators.

5. Concluding Remarks

This article considered and proposed some new estimators for estimating the ridge parameter \( k \). The performance of the ridge-type estimators depends on the variance of the random error, the correlations among the explanatory variables and the unknown coefficients vectors \( \beta \). Based on the simulation study, some conclusions might be drawn. The increase of \( \sigma \) and the increase of the correlation between the \( X \)'s variables have a negative effect in the MSE, in the sense that it also increases. When the sample size increases the MSE decreases, even when the correlation between the \( X \)'s variables and \( \sigma \) are large. In all situations, the new proposed
estimators have smaller MSE compared to ordinary least squared estimator. Three proposed estimators, $k_{KM4}$, $k_{KM5}$, $k_{KM6}$, and $k_{K1}$ and $k_{K2}$ performed better than the rest of estimators. From the example we found that all of the proposed ridge estimators are better than OLS. The analysis from the real data supported the simulation study to some extent. Finally, it appears that the proposed estimators $k_{KM4}$, $k_{KM5}$, and $k_{KM6}$ are useful and may be recommended to the practitioners.

Acknowledgment

The authors are thankful to the referee for making constructive and valuable suggestions and comments that have greatly improved the presentation of the article.

References


