Performance of Some New Ridge Regression Estimators

Article *in* Communication in Statistics- Simulation and Computation · January 2003

DOI: 10.1081/SAC-120017499

B M Golam Kibria
Florida International University

COMMUNICATIONS IN STATISTICS
Simulation and Computation®
Vol. 32, No. 2, pp. 419–435, 2003

Performance of Some New Ridge Regression Estimators

B. M. Golam Kibria*

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

ABSTRACT

In the ridge regression analysis, the estimation of ridge parameter k is an important problem. Many methods are available for estimating such a parameter. This article has considered some of these methods and also proposed some new estimators based on *generalized ridge regression* approach. A simulation study has been made to evaluate the performance of proposed estimators based on the minimum mean squared error (MSE) criterion. The simulation study indicates that under certain conditions the proposed estimators perform well compared to least squares estimators (LSE) and other popular existing estimators. Finally, a numerical example has been analyzed and its findings support the simulation results to some extent.

*Correspondence: B. M. Golam Kibria, Department of Statistics, Florida International University, Miami, FL 33199, USA; E-mail: kibriag@fiu.edu. AQ1

DOI: 10.1081/SAC-120017499 0361-0918 (Print); 1532-4141 (Online)

Key Words: MSE; Ridge regression; Signal-to-noise; Simulation study.

1. INTRODUCTION

In multiple linear regression model, we usually assume that the explanatory variables are independent. However, in practice, there may be strong or near to strong linear relationships among the explanatory variables. In that case the independence assumptions are no longer valid, which causes the problem of multicollinearity. In the presence of multicollinearity, it is impossible to estimate the unique effects of individual variables in the regression equation. Moreover, the regression coefficient will be experienced with unduly large sampling variance which affects both inference and prediction. Therefore, multicollinearity becomes one of the serious problem in the linear regression analysis. In literature, there are various methods existing to solve this problem. Among them, "ridge regression" is the most popular one which has many usefulness in real life.

To describe the ridge regression, we consider the following multiple linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \tag{1}$$

where $\mathbf{y}$ is a $n \times 1$ vector of observations, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients, $\mathbf{X}$ is a $n \times p$ known design matrix of rank p and $\mathbf{e}$ is a $n \times 1$ vector random variable, which is distributed as multivariate normal with mean vector $\mathbf{0}$ and variance–covariance matrix $\sigma^2\mathbf{I}_n$, $\mathbf{I}_n$ is a identity matrix of order n. The usual least squares estimate (LSE) or the maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = \mathbf{C}^{-1}\mathbf{X}'\mathbf{y}, \tag{2}$$

which will heavily depend on the characteristics of the matrix $\mathbf{C} = \mathbf{X}'\mathbf{X}$. If the $\mathbf{C}$ matrix is ill conditioned (near dependency among various columns of $\mathbf{C}$ or $\det(\mathbf{X}'\mathbf{X} \approx 0)$), the LSE are sensitive to a number of errors, for example, some of the regression coefficients may be statistically insignificant with wrong sign and meaningful statistical inference become impossible for the practitioners. Hoerl and Kennard (1970) found that multicollinearity is a common problem in the field of engineering. To resolve this problem they suggested to use $\mathbf{C}(k) = \mathbf{C} + kI_p, (k \geq 0)$ instead of $\mathbf{C}$, for estimating $\boldsymbol{\beta}$. The resulting estimators are given as

$$\hat{\boldsymbol{\beta}}(k) = (\mathbf{C} + k\mathbf{I}_p)^{-1}\mathbf{X}'\mathbf{y}, \tag{3}$$

which are known as ridge regression estimators (RRE). The constant, $k > 0$ is known as or biasing or ridge parameter. As k increases from zero and continues upto infinity, the regression estimates tend toward zero. Though these estimators result in biased, for certain value of k, they yield *minimum mean square error* (MMSE) compared to the LSE (see Hoerl and Kennard (1970)). However, the $\mathrm{MSE}(\hat{\boldsymbol{\beta}}(k))$ will depend on unknown parameters k, $\boldsymbol{\beta}$ and σ^2, which cannot be calculated in practice. But k, has to be estimated from the real data instead. Much of the discussions on ridge regression concern the problem of finding good empirical value of k. Many different techniques for estimating k have been proposed or suggested by different researchers. Dempster et al. (1977), Gibbons (1981), Golub et al. (1979), Gunst and Mason (1977), Haq and Kibria (1996), Hemmerle and Brantle (1978), Hocking et al. (1976), Hoerl and Kennard (1970), Hoerl et al. (1975), Kibria (1996), Lawless (1978), Lawless and Wang (1976), McDonald and Galarneau (1975), Nordberg (1982), Saleh and Kibria (1993), Singh and Tracy (1999), Wencheko (2000), Wichern and Churchill (1978) to mention a few.

The objective of the paper is to investigate some of the existing popular techniques that are available in literature and to make a comparison among them based on mean square properties. Moreover, we suggest some methods or procedures which are according our experience yields **AQ2** minimum MSE i.e., the good k values. The organization of the paper is as follows. We review and propose some new methods for estimating the ridge parameter k and consider a criterion for comparison of the estimators in Sec. 2. Section 3 describes the Monte Carlo simulation. Simulation results are discussed in Sec. 4. An example has been considered in Sec. 5. Finally, the concluding remarks is presented in Sec. 6.

2. PROPOSED ESTIMATORS AND PERFORMANCE CRITERIA

In this section, we introduce the *generalized ridge regression* and propose some estimators for estimating ridge parameter k. Suppose, there exists an orthogonal matrix $\mathbf{D}$ such that $\mathbf{D}'\mathbf{C}\mathbf{D} = \boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$ contains the eigen values of the matrix $\mathbf{C}$. The orthogonal (canonical form) version of the model (1) is

$$\mathbf{Y} = \mathbf{X}^*\boldsymbol{\alpha} + \mathbf{e}, \tag{4}$$

where $\mathbf{X}^* = \mathbf{X}\mathbf{D}$ and $\boldsymbol{\alpha} = \mathbf{D}'\boldsymbol{\beta}$. Then the generalized ridge regression estimators is given as follows:

$$\hat{\boldsymbol{\alpha}}(k) = (\mathbf{X}'^*\mathbf{X}^* + \mathbf{K})^{-1}\mathbf{X}'^*\mathbf{Y}, \tag{5}$$

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

$$\mathrm{MSE}(\hat{\boldsymbol{\alpha}}(k)) = \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + k_i)^2} + \sum_{i=1}^{p} \frac{k_i^2 \alpha_i^2}{(\lambda_i + k_i)^2}, \tag{6}$$

when

$$k_i = \frac{\sigma^2}{\alpha_i^2}, \tag{7}$$

where σ^2 represents the error variance of model (1), α_i is the ith element of $\boldsymbol{\alpha}$. Hocking et al. (1976) shows that for known optimal k_i, the generalized ridge regression estimator is superior to all other estimators within the class of biased estimators they considered. However, the optimal value of k_i fully depends on the unknown σ^2 and α_i, and they must be estimated from the observed data. Hoerl and Kennard (1970), suggested to replace σ^2 and α_i^2 by their corresponding unbiased estimators. That is,

$$\hat{k}_i = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}, \tag{8}$$

where $\hat{\sigma}^2 = (\sum \hat{e}_i^2)/(n-p)$ is the residual mean square estimate, which is unbiased estimator of σ^2 and $\hat{\alpha}_i$ is the tth element of $\hat{\boldsymbol{\alpha}}$, which is an unbiased estimator of $\boldsymbol{\alpha}$. Now we will review and propose some new methods based on the Eq. (8) and they are as follows:

1. Hoerl and Kennard (1970) suggests k is to be (thereafter $\hat{k}_{\mathrm{HK}}$ or HK),

   $$\hat{k}_{\mathrm{HK}} = \frac{\hat{\sigma}^2}{\hat{\alpha}_{\max}^2}, \tag{9}$$

   where $\hat{\alpha}_{\max}$ is the maximum element of $\hat{\boldsymbol{\alpha}}$. If σ^2 and $\boldsymbol{\alpha}$ are known then $\hat{k}_{\mathrm{HK}}$ will give smaller MSE than the LSE.
2. Hoerl, Kennard and Baldwin (1975) (thereafter $\hat{k}_{\mathrm{HKB}}$ or HKB), proposed a different estimator of k by taking the harmonic mean of $\hat{k}_i$ in Eq. (8). That is

   $$\hat{k}_{\mathrm{HKB}} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p} \alpha_i^2} = \frac{p\hat{\sigma}^2}{\hat{\boldsymbol{\alpha}}'\hat{\boldsymbol{\alpha}}} = \frac{p\hat{\sigma}^2}{\hat{\boldsymbol{\beta}}'\hat{\boldsymbol{\beta}}}, \tag{10}$$

   where $\hat{\boldsymbol{\beta}}$ is the LSE of $\boldsymbol{\beta}$.

3. From the Bayesian point of view, Lawless and Wang (1976) (thereafter $\hat{k}_{\mathrm{LW}}$ or LW), proposed the following estimator:

$$\hat{k}_{\mathrm{LW}} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p} \lambda_i \hat{\alpha}_i^2} = \frac{p\hat{\sigma}^2}{\hat{\boldsymbol{\alpha}}' X' X \hat{\boldsymbol{\alpha}}} = \frac{p\hat{\sigma}^2}{\hat{\boldsymbol{\beta}}' \boldsymbol{\Lambda} \hat{\boldsymbol{\beta}}}. \tag{11}$$

4. Hocking, Speed and Lynn (1979) (thereafter $\hat{k}_{\mathrm{HSL}}$ or HSL), suggests the following estimator for k

$$\hat{k}_{\mathrm{HSL}} = \hat{\sigma}^2 \frac{\sum_{i=1}^{p} (\lambda_i \hat{\alpha}_i)^2}{(\sum_{i=1}^{p} \lambda_i \hat{\alpha}_i^2)^2}. \tag{12}$$

5. New estimator ($\hat{k}_{\mathrm{AM}}$ or AM): We propose an estimator of k by using the arithmetic mean of $\hat{k}_i$ in Eq. (8), which produces the following estimator,

$$\hat{k}_{\mathrm{AM}} = \frac{1}{p} \sum_{i=1}^{p} \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}. \tag{13}$$

6. New estimator ($\hat{k}_{\mathrm{GM}}$ or GM): We propose to estimate k by using the geometric mean of $\hat{k}_i$ in Eq. (8), which produces the following new estimator:

$$\hat{k}_{\mathrm{GM}} = \frac{\hat{\sigma}^2}{(\prod_{i=1}^{p} \hat{\alpha}_i^2)^{\frac{1}{p}}}. \tag{14}$$

7. New estimator ($\hat{k}_{\mathrm{MED}}$ or MED): We propose to estimate k by using the median of $\hat{k}_i$ in Eq. (8), which produces the following new estimator for $p \geq 3$:

$$\hat{k}_{\mathrm{MED}} = \mathrm{Median}\left\{\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}\right\}, \quad i = 1, 2, \ldots, p. \tag{15}$$

It is observed from Eqs. (9)–(15) that for letting $\hat{\alpha}_i = \hat{\alpha}$ $(i = 1, 2, \ldots, p)$, all the proposed estimators except HSL become $(\hat{\sigma}^2/\hat{\alpha}^2)$. Among all estimators, but the LW and HSL, no weighting by λ_i are required as they are independent of λ_i. The estimators in Eqs. (9)–(11) have already been considered by Wichern and Churchill (1978) along with one iterative method. We will compare these estimators based on the criteria given as follows.

To compare the proposed estimators, a criterion for measuring "goodness" of a estimator is needed. Following Dempster et al. (1977), Lawless and Wang (1976) and Gibbons (1981), mean square error (MSE) criteria are used throughout our study to measure the goodness of an estimator. Ridge estimators are proposed with an attempt to have smaller MSE compared to LSE.

For any particular estimator $\tilde{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, with $\tilde{\boldsymbol{\alpha}} = \mathbf{D}'\tilde{\boldsymbol{\beta}}$ as the corresponding estimate of $\boldsymbol{\alpha}$, the total mean square error for the estimator is given by

$$\mathrm{MSE}(\tilde{\boldsymbol{\beta}}) = E(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sum_{i=1}^{p} E(\tilde{\alpha}_i - \alpha_i)^2 = \mathrm{MSE}(\tilde{\boldsymbol{\alpha}}), \tag{16}$$

which heavily depends on λ_i, α_i and σ^2 (see Eq. (6)). Since, theoretically the estimators in Eqs. (9)–(15) are very hard to compare, we will compare them through a simulation study which is discussed in the following section.

3. THE MONTE CARLO SIMULATION

In this section, we will discuss the simulation study to compare the performance of different estimators. To achieve different degrees of collinearity, following McDonald and Galarneau (1975) and Gibbons (1981) the explanatory variables were generated using the following device

$$x_{ij} = (1-\gamma^2)^{(1/2)} z_{ij} + \gamma z_{ip}, \quad i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, p, \tag{17}$$

where z_{ij} are independent standard normal pseudo-random numbers, and γ is specified so that the correlation between any two explanatory variables is given by γ^2. These variables are then standardized so that $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{y}$ are in correlation forms. Four different set of correlation corresponding to $\gamma = 0.70, 0.80, 0.90$ and 0.99 are considered.

Then n observations for the dependent variable are determined by

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots \beta_p x_{ip} + e_i, \quad i = 1, 2, \ldots, n, \tag{18}$$

where e_i are independent normal $(0, \sigma^2)$ pseudo-random numbers, and β_0 is taken to be identically zero. We vary the sample sizes between 15 and 25 and explanatory variables, between 4 and 10. Eight values of σ are investigated in this study, which are 0.01, 0.1, 0.25, 0.5, 1, 4, 9, and 20. For each set of explanatory variables one choice for the coefficient vectors is considered. Newhouse and Woman (1971) stated that if the mean squared error is a function of $\boldsymbol{\beta}$, σ^2 and k and, if the explanatory variables are fixed, then the MSE is minimized when $\boldsymbol{\beta}$ is the normalized eigen vectors corresponding to the largest eigen value of $\mathbf{X}'\mathbf{X}$ matrix subject to constraint that $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$. Hence, we selected the coefficients $\beta_1, \beta_2, \ldots, \beta_p$

as normalized eigen vectors corresponding to the largest eigen values of $\mathbf{X}'\mathbf{X}$ matrix so that $\beta'\beta = 1$. We could have as well used normalized eigen vectors corresponding to the smallest eigen value, however the conclusion about the performance of estimators in both cases do not change greatly. The eigen values and the regression coefficients of $\mathbf{X}'\mathbf{X}$ for different set of n, p, γ and ρ^2 are given in Table 4.1. T1

We consider a relationship between σ^2 and the signal to noise ratio as,

$$\rho^2 = \frac{\boldsymbol{\beta}'\boldsymbol{\beta}}{\sigma^2}.$$

The values of ρ^2 corresponding to σ are 1000, 100, 16, 4, 1, 0.063, 0.0123, and 0.0025 respectively.

For given values of n, p, $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$, γ and ρ^2, the set of explanatory variables are generated. Then the experiment was repeated 2000 times by generating new error terms in Eq. (18). For each replicate $r(r = 1, 2, \ldots, 2000)$,

Table 4.1. Values of λ and β are used in simulation for different n and p.

γ	0.70	0.80	0.90	0.99	0.70	0.80	0.90	0.99
n		$n=15$,	$p=4$			$n=25$,	$p=10$	
λ_1	1.7115	3.2872	3.1683	3.9135	6.8480	7.1640	8.5162	9.8656
λ_2	0.9975	0.3291	0.4523	0.0534	0.6831	0.6762	0.3641	0.0302
λ_3	0.7239	0.2464	0.2095	0.0225	0.6023	0.6006	0.3022	0.0300
λ_4	0.5671	0.1373	0.1699	0.0106	0.4727	0.4635	0.2455	0.0215
λ_5					0.3465	0.1833	0.1914	0.0382
λ_6					0.3370	0.2492	0.1324	0.0154
λ_7					0.2451	0.2194	0.1159	0.0080
λ_8					0.1969	0.1214	0.0718	0.0063
λ_9					0.1336	0.0918	0.0551	0.0028
λ_{10}					0.1071	0.0552	0.0247	0.0020
β_1	0.5539	0.5079	0.4518	0.5029	0.3270	0.3251	0.3181	0.3169
β_2	0.3769	0.4876	0.5242	0.4977	0.3228	0.3238	0.3316	0.3161
β_3	0.5059	0.4953	0.5143	0.5026	0.3163	0.2931	0.3007	0.3165
β_4	0.5433	0.5088	0.5065	0.4968	0.2926	0.2995	0.3243	0.3174
β_5					0.3072	0.3286	0.3149	0.3160
β_6					0.3326	0.3104	0.3146	0.3146
β_7					0.3158	0.3231	0.3251	0.3317
β_8					0.3270	0.3196	0.3236	0.3159
β_9					0.3047	0.2903	0.3022	0.3155
β_{10}					0.3142	0.3448	0.3057	0.3163

the values of k of different proposed estimators and the corresponding ridge estimators were calculated using

$$\hat{\boldsymbol{\alpha}}(k) = (\boldsymbol{\Lambda} + \hat{k} \times \mathbf{I}_p)^{-1}\mathbf{X}^{*\prime}\mathbf{y}, \quad \hat{k} = \hat{k}_{\mathrm{HK}}, \hat{k}_{\mathrm{HKB}}, \ldots, \hat{k}_{\mathrm{MED}}. \tag{19}$$

Then the MSEs for the estimators are calculated as follows

$$\mathrm{MSE}(\hat{\alpha}_k) = \frac{1}{2000}\sum_{r=1}^{2000}(\hat{\alpha}_{(r)} - \alpha)'(\hat{\alpha}_{(r)} - \alpha). \tag{20}$$

The simulated mean squared errors are summarized in Tables 4.2–4.9. The proportion of replications for which the least squares estimators produced a smaller MSE than the ridge regression estimator were calculated and presented in the parenthesis. T2–T9

Several major simulation studies have been conducted to compare LSE with RRE by various researchers: McDonald and Galarneau (1975), Hoerl et al. (1975) examined LSE and RRE, Lawless and Wang (1976) studied LSE, RRE and Principle component estimators (PCE) and Hocking et al. (1976) studied the RRE, PCE and Shrunken estimators to mention a few. All of them concluded that ridge type estimators perform better than the LSE most of the times. In the following section, we will discuss the performance of ridge type estimators.

Table 4.2. Estimated MSE with $n = 15$, $p = 4$, $\gamma = 0.70$ and $\lambda_1/\lambda_4 = 3.0$.

ρ^2	LS	HK	HKB	LW	HSL	AM	GM	MED
10000	0.056	0.056	0.056	0.056	0.056	0.552	0.123	0.342
		(100)	(100)	(100)	(100)	(100)	(100)	(100)
100	0.092	0.093	0.098	0.096	0.094	0.562	0.247	0.362
		(98.2)	(98.6)	(98.3)	(98.2)	(99.9)	(99.7)	(99.7)
16	0.251	0.251	0.267	0.257	0.251	0.631	0.409	0.454
		(61.6)	(70.3)	(66.6)	(61.7)	(91.3)	(83.2)	(83.1)
4	0.536	0.519	0.538	0.524	0.515	0.747	0.627	0.637
		(41.7)	(50.8)	(48)	(43.1)	(74.7)	(63.1)	(60.4)
1	0.861	0.805	0.793	0.779	0.793	0.866	0.814	0.813
		(26.4)	(33.9)	(33.0)	(28.9)	(51.6)	(43.0)	(41.0)
.0625	1.202	1.088	1.031	1.016	1.064	0.984	0.990	0.994
		(13.7)	(17.8)	(18)	(15.6)	(26.6)	(22.4)	(21.6)
.0123	1.282	1.154	1.087	1.073	1.126	1.012	1.032	1.037
		(11.7)	(15.0)	(15.2)	(13.6)	(21.3)	(18.3)	(17.6)
.0025	1.3	1.168	1.100	1.086	1.140	1.022	1.044	1.046
		(11.7)	(14.3)	(14.6)	(13.3)	(20.4)	(17.4)	(16.8)

Table 4.3. Estimated MSE with $n = 15, p = 4, \gamma = 0.80$ and $\lambda_1/\lambda_4 = 23.9$.

ρ^2	LS	HK	HKB	LW	HSL	AM	GM	MED
10000	0.202	0.202	0.202	0.202	0.202	0.438	0.221	0.288
		(100)	(100)	(100)	(100)	(100)	(100)	(100)
100	0.253	0.251	0.248	0.250	0.251	0.445	0.277	0.309
		(26.0)	(31.8)	(26.4)	(26.0)	(78.9)	(56.6)	(62.4)
16	0.487	0.419	0.388	0.399	0.403	0.522	0.397	0.417
		(4.9)	(9.8)	(5.4)	(4.9)	(46.2)	(24.4)	(25.5)
4	0.941	0.727	0.627	0.568	0.568	0.649	0.581	0.586
		(2.8)	(6.4)	(3.3)	(2.9)	(25.4)	(12.7)	(12.5)
1.0	1.464	1.112	0.927	0.766	0.832	0.804	0.799	0.792
		(2.7)	(4.3)	(3.7)	(3.5)	(12.8)	(7.9)	(6.8)
0.0625	1.92	1.486	1.246	1.010	1.222	0.964	1.023	1.041
		(1.20)	(1.7)	(1.7)	(1.8)	(4.7)	(2.9)	(2.8)
0.0123	1.974	1.540	1.297	1.061	1.281	1.004	1.066	1.084
		(1.0)	(2.0)	(1.9)	(2.05)	(3.9)	(2.6)	(2.6)
0.0025	2.011	1.573	1.326	1.084	1.310	1.021	1.091	1.108
		(1.0)	(1.7)	(1.9)	(2.0)	(3.7)	(2.3)	(2.6)

Table 4.4. Estimated MSE with $n = 15, p = 4, \gamma = 0.90$ and $\lambda_1/\lambda_4 = 18.7$.

ρ^2	LS	HK	HKB	LW	HSL	AM	GM	MED
10000	0.193	0.193	0.193	0.193	0.193	0.451	0.214	0.29
		(100)	(100)	(100)	(100)	(100)	(100)	(100)
100	0.242	0.24	0.238	0.239	0.240	0.467	0.278	0.313
		(31.9)	(36.8)	(32.4)	(32.0)	(83.9)	(62.7)	(68.8)
16	0.448	0.393	0.368	0.378	0.383	0.537	0.392	0.411
		(6.7)	(11.9)	(7.1)	(6.7)	(52.8)	(30.7)	(31.6)
4	0.87	0.681	0.600	0.556	0.557	0.647	0.575	0.582
		(3.4)	(8.1)	(4.1)	(3.6)	(28.8)	(16.4)	(15.5)
1	1.361	1.054	0.895	0.765	0.830	0.803	0.790	0.784
		(3.7)	(5.8)	(4.7)	(4.5)	(15.1)	(9.7)	(8.7)
.0625	1.817	1.434	1.217	1.009	1.211	0.961	1.007	1.023
		(2.1)	(2.8)	(3.1)	(3.0)	(6.1)	(4.3)	(3.9)
.0123	1.885	1.489	1.270	1.064	1.263	1.000	1.058	1.070
		(1.2)	(1.9)	(2.3)	(2.3)	(4.7)	(3.4)	(3.2)
.0025	1.886	1.500	1.289	1.09	1.280	1.024	1.087	1.098
		(1.1)	(1.9)	(1.9)	(1.9)	(4.4)	(3.1)	(2.8)

Table 4.5. Estimated MSE with $n = 15$, $p = 4$, $\gamma = 0.99$ and $\lambda_1/\lambda_4 = 369.2$.

ρ^2	LS	HK	HKB	LW	HSL	AM	GM	MED
10000	0.249	0.249	0.248	0.249	0.249	0.331	0.25	0.26
		(20.9)	(21.3)	(20.9)	(20.9)	(80.8)	(45.6)	(62.0)
100	0.649	0.493	0.413	0.42	0.421	0.379	0.348	0.366
		(0.1)	(0.2)	(0.1)	(0.1)	(14.2)	(3.1)	(4.5)
16	2.359	1.458	0.968	0.482	0.472	0.529	0.597	0.567
		(0.1)	(0.2)	(0.1)	(0.1)	(4.5)	(1.1)	(1.2)
4	6.021	3.606	2.201	0.536	0.516	0.659	0.989	0.906
		(0.1)	(0.2)	(0.1)	(0.1)	(1.4)	(0.4)	(0.4)
1	9.483	5.661	3.401	0.684	0.924	0.739	1.217	1.312
		(0.1)	(0.2)	(0.1)	(0.1)	(0.7)	(0.3)	(0.3)
.0625	11.741	6.933	4.233	0.933	2.038	0.929	1.307	1.764
		(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)
.0123	12.147	7.309	4.479	0.993	2.197	0.981	1.403	1.858
		(0)	(0)	(0.05)	(0.05)	(0.1)	(0)	(0)
.0025	12.29	7.439	4.58	1.031	2.425	1.011	1.425	1.894
		(0.1)	(0.1)	(0.2)	(0.2)	(0.2)	(0.2)	(0.2)

Table 4.6. Estimated MSE with $n = 25$, $p = 10$, $\gamma = 0.70$ and $\lambda_1/\lambda_{10} = 63.9$.

ρ^2	LS	HK	HKB	LW	HSL	AM	GM	MED
10000	0.382	0.382	0.382	0.382	0.382	0.652	0.413	0.427
		(100)	(100)	(100)	(100)	(100)	(100)	(100)
100	0.443	0.438	0.425	0.437	0.438	0.654	0.443	0.443
		(0.7)	(2.55)	(0.85)	(0.7)	(87.15)	(43.4)	(41.4)
16	0.703	0.619	0.538	0.578	0.595	0.687	0.513	0.516
		(0.05)	(0.4)	(0.05)	(0.05)	(44)	(7.1)	(5.3)
4	1.256	0.998	0.759	0.717	0.726	0.756	0.649	0.67
		(0)	(0.1)	(0)	(0)	(11.3)	(1.35)	(0.7)
1	1.894	1.458	1.04	0.849	0.877	0.852	0.827	0.85
		(0)	(0.05)	(0)	(0)	(1.05)	(0.05)	(0.05)
.0625	2.445	1.898	1.348	1.038	1.326	0.968	1.023	1.073
		(0)	(0)	(0)	(0)	(0.2)	(0.05)	(0)
.0123	2.476	1.925	1.382	1.076	1.382	0.991	1.059	1.113
		(0)	(0)	(0)	(0)	(0.1)	(0)	(0)
.0025	2.486	1.93	1.395	1.098	1.395	1.008	1.084	1.135
		(0)	(0)	(0)	(0)	(0)	(0)	(0)

Table 4.7. Estimated MSE with $n = 25$, $p = 10$, $\gamma = 0.80$ and $\lambda_1/\lambda_{10} = 129.8$.

ρ^2	LS	HK	HKB	LW	HSL	AM	GM	MED
10000	0.393	0.393	0.393	0.393	0.393	0.647	0.417	0.429
		(98.9)	(99)	(98.9)	(98.9)	(100)	(100)	(100)
100	0.474	0.463	0.439	0.459	0.462	0.656	0.445	0.445
		(0.05)	(0.75)	(0.05)	(0.05)	(78.75)	(26.9)	(24.25)
16	0.83	0.695	0.567	0.597	0.619	0.682	0.515	0.522
		(0)	(0.05)	(0)	(0)	(31.8)	(3.15)	(2.9)
4	1.543	1.164	0.811	0.712	0.717	0.746	0.649	0.672
		(0)	(0)	(0)	(0)	(6)	(0.4)	(0.25)
1	2.373	1.74	1.133	0.835	0.852	0.844	0.827	0.846
		(0)	(0)	(0)	(0)	(0.65)	(0)	(0)
.0625	3.095	2.287	1.491	1.029	1.337	0.964	1.038	1.095
		(0)	(0)	(0)	(0)	(0.05)	(0)	(0)
.0123	3.193	2.367	1.545	1.071	1.416	0.991	1.078	1.137
		(0)	(0)	(0)	(0)	(0.05)	(0)	(0)
.0025	3.23	2.397	1.569	1.09	1.463	1.003	1.096	1.155
		(0)	(0)	(0)	(0)	(0)	(0)	(0)

Table 4.8. Estimated MSE with $n = 25$, $p = 10$, $\gamma = 0.90$ and $\lambda_1/\lambda_{10} = 344.8$.

ρ^2	LS	HK	HKB	LW	HSL	AM	GM	MED
10000	0.433	0.433	0.433	0.433	0.433	0.613	0.444	0.449
		(66.1)	(67.5)	(66.1)	(66.1)	(100)	(99.1)	(99.4)
100	0.557	0.528	0.488	0.521	0.524	0.619	0.469	0.47
		(0)	(0)	(0)	(0)	(51.8)	(5.7)	(5.2)
16	1.114	0.876	0.657	0.658	0.668	0.646	0.544	0.564
		(0)	(0)	(0)	(0)	(11.2)	(0.6)	(0.4)
4	2.374	1.694	1.033	0.737	0.721	0.721	0.699	0.731
		(0)	(0)	(0)	(0)	(0.9)	(0)	(0)
1	3.949	2.748	1.55	0.832	0.83	0.838	0.913	0.956
		(0)	(0)	(0)	(0)	(0.2)	(0.05)	(0.1)
.0625	5.167	3.601	2.023	1.004	1.501	0.957	1.116	1.242
		(0)	(0)	(0)	(0)	(0)	(0)	(0)
.0123	5.373	3.757	2.122	1.054	1.678	0.99	1.156	1.291
		(0)	(0)	(0)	(0)	(0)	(0)	(0)
.0025	5.216	3.61	2.068	1.073	1.636	1.005	1.188	1.328
		(0)	(0)	(0)	(0)	(0)	(0)	(0)

Table 4.9. Estimated MSE with $n = 25$, $p = 10$, $\gamma = 0.99$ and $\lambda_1/\lambda_{10} = 4932.8$.

ρ^2	LS	HK	HKB	LW	HSL	AM	GM	MED
10000	0.479	0.478	0.474	0.478	0.478	0.520	0.467	0.468
		(0)	(0)	(0)	(0)	(47.2)	(2.5)	(3.1)
100	1.858	1.336	0.831	0.687	0.687	0.548	0.568	0.605
		(0)	(0)	(0)	(0)	(1.3)	(0)	(0)
16	8.216	5.285	2.397	0.662	0.650	0.644	0.934	0.981
		(0)	(0)	(0)	(0)	(0)	(0)	(0)
4	22.393	14.096	5.889	0.642	0.621	0.773	1.631	1.915
		(0)	(0)	(0)	(0)	(0)	(0)	(0)
1	42.587	26.831	11.124	0.734	0.754	0.826	2.405	3.243
		(0)	(0)	(0)	(0)	(0)	(0)	(0)
.0625	55.958	34.528	14.018	0.934	2.769	0.942	2.573	4.318
		(0)	(0)	(0)	(0)	(0.1)	(0)	(0)
.0123	58.132	36.436	15.047	0.976	3.441	0.975	2.609	4.54
		(0)	(0)	(0)	(0)	(0)	(0)	(0)
.0025	58.65	36.838	15.193	1.004	3.524	1	2.685	4.669
		(0)	(0)	(0)	(0)	(0)	(0)	(0)

4. THE SIMULATED RESULTS

The main results of the simulation have been presented in a two-way tables (Tables 4.2–4.9) of estimators versus ρ^2 for different γ, n, p and the ratio of (λ_1/λ_p) to display the main results as concisely as possible. Therefore, to compare the performance of the proposed estimators, we will consider the following criteria.

4.1. Performance as a Function of ρ^2

For any given value of $\rho^2(<10,000)$, the proposed estimators dominate the least squared estimator (dominance in the sense of having smaller MSE). It is noticed from Tables 4.2–4.9 that for all the cases, the MSEs increase with the decrease in ρ^2, which supports the simulation results of Hoerl et al. (1975) and Lawless and Wang (1976). It is also noticed that the proportion of replication for which the LSE have smaller MSE than the proposed estimators decreases as ρ^2 decreases and approaches to zero when ρ^2 is equal to one or less. Note in Tables 4.4–4.9 that estimators GM and LW perform about equally well but better than HKB most of the times. For $\rho^2 \leq 1$, both GM and AM perform better than LW and HKB, however, for $\rho^2 \geq 1$, LW perform better than GM and HBK most of the times. Sometimes, the performance of GM estimator are somewhat between HKB and LW estimators. Unlike LW estimator, GM does not depend on the eigen values of $\mathbf{X}'\mathbf{X}$ matrix. Surprisingly the MED performs better than HKB most of the times and HSL some of the times. Lawless and Wang (1976) considered various estimators for k along with HKB and LW and observed that their estimators outperformed by a sizeable amount than the HKB estimator for small and moderate values of ρ^2. It is noted that our simulation results support their simulation results. However, HKB perform better than LW and GM for small and moderate correlation and higher values of ρ^2.

4.2. Performance as a Function of γ and (λ_1/λ_p)

For given n, p, and σ, the proportion of replication for which the LSE produce smaller MSE than proposed estimators decreases as γ increases. GM and MED perform as good as LW, HKB and better than others for the small and moderate correlations among the explanatory variables. It is evident from Tables 4.2–4.9 that the degree of multicollinearity (as a measure of (λ_1/λ_p)) affects the relative performance of the ridge estimators. The proportion of replications for which LSE produces smaller MSE compared to proposed estimators becomes smaller as the ratio, (λ_1/λ_p) becomes larger. But for a fixed value of ρ^2, the MSE of the estimators increases with the increase of (λ_1/λ_p). All estimators perform at least as well as the least squares for $\rho^2 < 10000$ over the values of (λ_1/λ_p). For given n, p, ρ^2, and γ, the MSE increases for most of the estimators as the ratio increases. For small value of (λ_1/λ_p), HKB performs little better than others. However, for wider spread of λ, i.e., large ratio of (λ_1/λ_p), all proposed estimators outperform the least squares estimators.

4.3. Performance as a Function of n and p

For a given ρ^2 and γ, as the sample size and the number of explanatory variables increase, the MSE of the proposed estimators increases and the proportion of replication for which LSE produce smaller MSE than the proposed estimators decreases. It is observed that for a small sample size and moderate correlation HKB performs better than LW and GM when $\rho^2 > 1$. However, as n and p increase the performance of LW and GM become better than HKB estimator.

5. AN EXAMPLE

In this section we consider an example, which has been taken from Brownlee (1965) to compare the performance of the proposed estimators. This example contains 21 days of operation of a plant for the oxidation of ammonia (NH_3) to nitric acid (HNO_3). We consider the following linear regression model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + e, \tag{21}$$

where Y is the percent of ingoing NH_3 that is lost by escaping in the unabsorbed nitric oxides ($\times 10$), which is an inverse measure of the yield of the overall efficiency for the plant, X_1 is the operation day, X_2 is the rate of operation of the plant, X_3 is the temperature of the cooling water in the coils of the absorbing tower for the nitric oxides and X_4 is the concentration of HNO_3 in the absorbing liquid (coded by minus 50, times 10). For details about the data, see Brownlee (1965). The correlation matrix of the variables in model (21) is presented in Table 5.1. **T5.1**

It is observed from Table 5.1 that the explanatory variables are moderate to highly correlated. Moreover, $(\lambda_1/\lambda_p) = (246264.3/65.89) = 3737.51$, which implies the existence of multicollinearity in the data set. So it is adequate to compare the proposed ridge estimators with this real data set.

The MSE of the proposed estimators are estimated by

$$\mathrm{MSE}(\hat{\beta}_n) = \hat{\sigma}^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + \hat{k})^2} + \hat{k}^2 \sum_{i=1}^{p} \frac{\alpha_i^2}{(\lambda_i + \hat{k})^2}, \tag{22}$$

where $\hat{k} = \hat{k}_{\mathrm{HK}}, \hat{k}_{\mathrm{HKB}}, \ldots, \hat{k}_{\mathrm{MED}}$.

The estimated MSE along with the ridge regression coefficients are presented in Table 5.2. **T5.2**

Table 5.1. Correlations among the variables.

	X_1	X_2	X_3	X_4	Y
X_1	1.000	−0.696	−0.767	−0.426	−0.811
X_2	−0.696	1.000	0.782	0.500	0.920
X_3	−0.767	0.782	1.000	0.391	0.876
X_4	−0.426	0.500	0.391	1.000	0.400
Y	−0.811	0.920	0.876	0.400	1.00

Table 5.2. The MSE and the estimated regression coefficients of the estimators.

Estimators	LSE	HK	HKB	LW	HSL	AM	GM	MED
MSE	9.069	0.142	0.115	0.171	0.192	0.152	0.117	0.153
$\hat{\beta}_1$	−0.534	−0.540	−0.547	−0.534	−0.511	−0.538	−0.551	−0.537
$\hat{\beta}_2$	0.668	0.673	0.678	0.668	0.600	0.640	0.671	0.639
$\hat{\beta}_3$	0.640	0.587	0.505	0.640	0.251	0.302	0.390	0.300
$\hat{\beta}_4$	−0.353	−0.343	−0.325	−0.353	−0.212	−0.250	−0.291	−0.249

Here we are not discussing the consequences of the multicollinearity, as it is available in Hoerl and Kennard (1970), among other papers on ridge regression. From Table 5.2, we observe that all the proposed estimators perform better than LSE in the sense of having smaller MSE. However, GM and HKB perform equivalently and little better than other estimators.

6. SUMMARY AND CONCLUDING REMARKS

This article has considered and proposed some new estimators for estimating the ridge parameter k. The performance of the ridge-type estimators depends on the signal to noise ratio i.e., variance of the random error, the correlations (γ) among the explanatory variables and the unknown coefficients vectors ($\boldsymbol{\beta}$). Based on the simulation study, some conclusion might be drawn. However, these conclusions are restricted to the set of experimental conditions which were investigated. When the signal-to-noise ratio is large ($\rho^2 \geq 10,000$), the performance of LSE is reasonably better than all proposed estimators for all degree of multicollinearity condition. Otherwise, all of the proposed estimators have smaller MSE than the ordinary least squared estimators. It is evident from the simulation results that estimators GM and LW perform equivalently well, but a little better than HKB. For small or moderate correlation, GM performs better than LW. However, for strong correlation, LW performs better than GM and HKB. It is noted that AM performs quite well compared to other estimators. However, being a highly biased estimator (see Hoerl and Kennard (1970)), it is not advisable to use in practice. Most of the times, the proposed estimators perform better than the LW, HK, HKB, HSL for large σ or small signal-to-noise ratio. Our result does not contradict with that of Hoerl et al. (1970), Lawless and Wang (1976) and Gibbon (1981) among others. From the example we found that GM and HKB perform equivalently well and a little better than other estimators. In conclusion, GM estimator might be considered to estimate the ridge parameter k, as one of the good estimators. However, more study is required before making any definite statement.

ACKNOWLEDGMENT

The author is thankful to the College of Arts and Sciences of Florida International University for providing him some research facilities.

REFERENCES

Brownlee, K. A. (1965). *Statistical Theory and Methodology in Science and Engineering*. 2nd ed. New York: Wiley.

Dempster, A. P., Schatzoff, M., Wermuth, N. (1977). A simulation study of alternatives to ordinary least squares. *Journal of the American Statistical Association* 72:77–91.

Gibbons, D. G. (1981). A simulation study of some ridge estimators. *Journal of the American Statistical Association* 76:131–139.

Golub, G. H., Health, M., Wahaba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* 21:215–223.

Gunst, R. F., Mason R. L. (1977). Biased estimation in regression: An evaluation using mean squared error. *Journal of the American Statistical Association* 72:616–628.

Haq, M. S., Kibria, B. M. G. (1996). A shrinkage estimator for the restricted linear regression model: ridge regression approach. *Journal of Applied Statistical Science* 3(4):301–316.

Hemmerle, W. J., Brantle, T. F. (1978). Explicit and Constraint generalized ridge estimator. *Technometrics* 2:109–120.

Hoerl, A. E., Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics* 12:55–67.

Hoerl, A. E., Kennard, R. W., Baldwin, K. F. (1975). Ridge regression: some simulation. *Communications in Statistics* 4:105–123.

Hocking, R. R., Speed, F. M., Lynn, M. J. (1976). A class of biased estimators in linear regression. *Technometrics* 18:425–438.

Kibria, B. M. G. (1996). On preliminary test ridge regression estimators for linear restriction in a regression model with non-normal disturbances. *Communications in Statistics* A25:2349–2369.

Lawless, J. F. (1978). Ridge and related estimation procedure. *Communications in Statistics* A7:139–164.

Lawless, J. F., Wang, P. (1976). A simulation study of ridge and other regression estimators. *Communications in Statistics* A5:307–323.

McDonald, G. C., Galarneau, D. I. (1975). A monte carlo evaluation of some Ridge-type estimators. *Journal of the American Statistical Association* 70:407–416.

Newhouse, J. P., Oman, S. D. (1971). An evaluation of ridge estimators. Rand Corporation, P-716-PR.

Nordberg, L. (1982). A procedure for determination of a good ridge parameter in linear regression. *Communications in Statistics* A11:285–309.

Saleh, A. K. Md., Kibria, B. M. G. (1993). Performances of some new preliminary test ridge regression estimators and their properties. *Communications in Statistics—Theory and Methods* 22:2747–2764.

Singh, S., Tracy, D. S. (1999). Ridge-regression using scrambled responses. *Metrika* 147–157.

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

Wencheko, E. (2000). Estimation of the signal-to-noise in the linear regression model. *Statistical Papers* 41:327–343.