Performance of a New Ridge Regression Estimator

Al-Hassan, Yazid M.

Department of Mathematics, Royal Institute of Technology (KTH)
Stockholm, Sweden
yazidah@kth.se ; yazid_12111980@yahoo.com

ABSTRACT
Ridge regression estimator has been introduced as an alternative to the ordinary least squares estimator (OLS) in the presence of multicollinearity. Several studies concerning ridge regression have dealt with the choice of the ridge parameter. Many algorithms for the ridge parameter have been proposed in the statistical literature. In this article, a new method for estimating ridge parameter is proposed. A simulation study has been made to evaluate the performance of the proposed estimator based on the mean squared error (MSE) criterion. The evaluation has been done by comparing the MSEs of the proposed estimator with other well-known estimators. In the presence of multicollinearity, the simulation study indicates that under certain conditions the proposed estimator performs better than other estimators.

KEYWORDS: Mean squared error; Monte Carlo simulations; Multicollinearity; Ridge parameter; Ridge regression.

INTRODUCTION

Consider the standard model for multiple linear regression (Draper and Smith, 1998)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \qquad (1)$$

where $\mathbf{y}$ is an $n \times 1$ column vector of observations on the dependent variable, $\mathbf{X}$ is an $n \times p$ fixed matrix of observations on the explanatory variables and is of full rank p $(p \le n)$, $\boldsymbol{\beta}$ is a $p \times 1$ unknown column vector of regression coefficients, and $\mathbf{e}$ is an $n \times 1$ vector of random errors; $\mathrm{E}(\mathbf{e}) = \mathbf{0}, \mathrm{E}(\mathbf{ee}') = \sigma^2 \mathbf{I}_n$, where $\mathbf{I}_n$ denotes the $n \times n$ identity matrix and the prime denotes the transpose of a matrix. The variables are assumed to be standardized so that $\mathbf{X}'\mathbf{X}$ is in the form of correlation matrix, and the vector $\mathbf{X}'\mathbf{y}$ is the vector of correlation coefficients of the dependent variable with each explanatory variable. The ordinary least squares (OLS) estimator, $\hat{\boldsymbol{\beta}}$, of the parameters is given by (Draper and Smith, 1998)

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \qquad (2)$$

Clearly, $\hat{\boldsymbol{\beta}}$ is an unbiased estimator of $\boldsymbol{\beta}$. Let $\lambda_1, \lambda_2, ..., \lambda_p$ denotes the eigenvalues of $\mathbf{X}'\mathbf{X}$. The mean squared error (MSE) of the components of $\hat{\boldsymbol{\beta}}$ is given by (Draper and Smith, 1998)

$$\mathrm{MSE}(\boldsymbol{\beta}) = \mathrm{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sigma^2 \sum_{i=1}^{p} \frac{1}{\lambda_i} \qquad (3)$$

In application of multiple linear regression, the matrix $\mathbf{X}'\mathbf{X}$ might be nearly singular, that is, λ_i is small for some value of i. This is due to some inter-relation between the explanatory variables. The relation is technically called multicollinearity. The OLS estimator of regression coefficients tends to become "unstable" in the presence of multicollinearity. More precisely, the variance of the estimates of some of the regression coefficients becomes large. This is clear from (3).

Many attempts have been made to improve the OLS estimation procedure. In general, there are two approaches. One approach centers on finding (biased) estimators which have smaller MSE than the OLS estimators. Ridge regression, as well as many shrinkage type of estimators (Stein, 1960; Sclove, 1968), is one example. This approach does not directly address itself to the issue of multicollinearity, even though multicollinearity is often the situation where the aforementioned procedures (or estimators) are used.

Among these estimators, the ridge estimator points indirectly to the issue of multicollinearity by constraining the length of the coefficient estimator (Hocking, 1976). In contrast, the second approach deals straightforward with the dependency nature of the explanatory variables. The principal components regression, as well as the latent root regression and the factor analysis approach, is one such example.

Hoerl and Kennard (1970a, 1970b) proposed the ridge estimator as an alternative to the OLS estimator for use in the presence of multicollinearity. The ridge estimator is given by

$$\hat{\boldsymbol{\beta}}_R = (\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}'\mathbf{y}\,, \tag{4}$$

where $\mathbf{I}$ denotes an identity matrix and k is a positive number known as ridge parameter. The corresponding MSE is given by

$$\mathbf{MSE}(\hat{\boldsymbol{\beta}}_R) = \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + k)^2} + k^2 \boldsymbol{\beta}'(\mathbf{X}'\mathbf{X} + k\mathbf{I})^{-2}\boldsymbol{\beta} \tag{5}$$

Though this estimator results in bias, for a certain value of k, it yields minimum MSE compared to the OLS estimator (Hoerl and Kennard, 1970a). However, the MSE($\hat{\boldsymbol{\beta}}_R$) depends on unknown parameters k, $\boldsymbol{\beta}$ and σ^2, which can't be calculated in practice, but k has to be estimated from the real data instead.

Several methods for estimating k have been proposed (Hoerl and Kennard ,1970a; Hoerl et al. 1975; McDonald and Galarneau ,1975; Lawless and Wang ,1976; Hocking et al. ,1976;Wichern and Churchill ,1978, Nordberg, 1982; Saleh and Kibria ,1993; Singh and Tracy ,1999;, Wencheko ,2000; Kibria , 2003; Khalaf and Shukur ,2005;, Alkhamisi et al. ,2006 and Alkhamisi and Shukur ,2007).

Alkhamisi and Shukur (2007) suggested a new approach to estimate the ridge parameter. They also proposed some new estimators by adding $1/\lambda_{\max}$ to some well-known estimators, where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{X}'\mathbf{X}$. The authors used Monte Carlo experiments and the MSE criterion to compare the proposed estimators with some well-known estimators.

The purpose of this study is to apply the modification mentioned in Alkhamisi and Shukur (2007) to the estimator proposed by Hocking et al. (1976) in order to define a new estimator. A Monte Carlo comparison will be made using the MSE criterion to compare the performances of the proposed estimator with the OLS estimator and the estimators of Hocking et al. , 1976 and Hoerl and Kennard ,1970.

METHODOLOGY

It is convenient to express the regression model (1) in the canonical form. Suppose that there exists an orthogonal matrix $\mathbf{D}$ such that $\mathbf{D}'\mathbf{C}\mathbf{D} = \mathbf{\Lambda}$, where $\mathbf{C} = \mathbf{X}'\mathbf{X}$ and $\mathbf{\Lambda} = diag(\lambda_1, \lambda_2, ..., \lambda_p)$ contains the eigenvalues of the matrix $\mathbf{C}$, then the canonical form of the model (1) is

$$\mathbf{y} = \mathbf{X}^*\boldsymbol{\alpha} + \mathbf{e}, \quad (6)$$

where $\mathbf{X}^* = \mathbf{XD}$ and $\boldsymbol{\alpha} = \mathbf{D}'\boldsymbol{\beta}$. Then the OLS estimator is given as follows

$$\hat{\boldsymbol{\alpha}} = \mathbf{\Lambda}^{-1}\mathbf{X}^{*'}\mathbf{y} \quad (7)$$

and so we can write the ridge estimator as

$$\hat{\boldsymbol{\alpha}}_R = (\mathbf{X}^{*'}\mathbf{X}^* + \mathbf{K})^{-1}\mathbf{X}^{*'}\mathbf{y}, \quad (8)$$

where $\mathbf{K} = diag(k_1, k_2, ..., k_p)$, $k_i > 0$. Equation (8) is called *the general form of ridge regression* (Hoerl and Kennard, 1970a). It follows from Hoerl and Kennard (1970a) that the value of k_i which minimizes the MSE($\hat{\boldsymbol{\alpha}}_R$), where

$$\mathbf{MSE}(\hat{\boldsymbol{\alpha}}_R) = \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + k_i)^2} + \sum_{i=1}^{p} \frac{k_i^2 \alpha_i^2}{(\lambda_i + k_i)^2}, \quad (9)$$

is

$$k_i = \frac{\sigma^2}{\alpha_i^2}, \quad (10)$$

where σ^2 represents the error variance of model (1), α_i is the ith element of $\boldsymbol{\alpha}$.

Equation (10) gives a value of k_i that fully depends on the unknowns σ^2 and α_i, and must be estimated from the observed data. Hoerl and Kennard (1970a) suggested the replacement of σ^2 and α_i by their corresponding unbiased estimators, that is,

$$\hat{k}_i = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}, \quad (11)$$

where $\hat{\sigma}^2 = \sum e_i^2 / n - p$ is the residual mean square estimate, which is an unbiased estimator of σ^2, and $\hat{\alpha}_i$ is the ith element of $\hat{\boldsymbol{\alpha}}$, which is an unbiased estimator of $\boldsymbol{\alpha}$. They found that the best method for achieving a better estimate $\hat{\boldsymbol{\alpha}}_R$ is to use $k_i = k$ for all i, and they suggested k to be $\hat{k}_{HK}$ (or HK) where

$$\hat{k}_{HK} = \frac{\hat{\sigma}^2}{\max(\hat{\alpha}_i)}. \quad (12)$$

If σ^2 and $\boldsymbol{\alpha}$ are known, then $\hat{k}_{HK}$ is sufficient to give ridge estimators having smaller MSE than the OLS estimator.

Hocking et al. (1976) suggested the following estimator, $\hat{k}_{HSL}$ (or HSL), for k:

$$\hat{k}_{HSL} = \hat{\sigma}^2 \frac{\sum_{i=1}^{p} (\lambda_i \hat{\alpha}_i)^2}{(\sum_{i=1}^{p} \lambda_i \hat{\alpha}_i^2)^2}. \quad (13)$$

We now apply the modification mentioned in Alkhamisi and Shukur (2007) to the estimator proposed by Hocking et al. (1976), $\hat{k}_{HSL}$, to obtain our new estimator $\hat{k}_{NHSL}$ (or NHSL):

$$\hat{k}_{NHSL} = \frac{\hat{\sigma}^2 \lambda_{\max} \sum_{i=1}^{p} (\lambda_i \hat{\alpha}_i)^2 + (\sum_{i=1}^{p} \lambda_i \hat{\alpha}_i^2)^2}{\lambda_{\max} (\sum_{i=1}^{p} \lambda_i \hat{\alpha}_i^2)^2}$$

$$= \hat{\sigma}^2 \frac{\sum_{i=1}^{p} (\lambda_i \hat{\alpha}_i)^2}{(\sum_{i=1}^{p} \lambda_i \hat{\alpha}_i^2)^2} + \frac{1}{\lambda_{\max}}$$

$$= \hat{k}_{HSL} + \frac{1}{\lambda_{\max}} \qquad (14)$$

Since $1/\lambda_{\max} > 0$, $\hat{k}_{NHSL}$ is grater than $\hat{k}_{HSL}$.

The Simulation Study

In this section, we use Monte Carlo simulation to investigate the properties of OLS, HK, HSL and NHSL. A comparison is then made based on the MSE criterion. Although many estimators can be considered in this simulation study (Kibria ,2003; Khalaf and Shukur ,2005; Alkhamisi et al. 2006; Alkhamisi and Shukur ,2007 and Al-Hassan ,2008), we will only consider OLS, HK and HSL estimators and compare them with NHSL. We made these choices for the following reasons:

1. Our interest here lies in studying the properties of NHSL as an alternative of OLS in the presence of multicollinearity.
2. HK estimator is the first ridge estimator that was proposed among all other estimators. Moreover, most of studies concerned with proposing new ridge estimators or comparing ridge estimators to each other take HK estimator in consideration.
3. By construction, NHSL is a modified version of HSL, so we thought that it is necessary to make a comparison between them.

Therefore, we can say that it is convenient to make the comparison among OLS, HK, HSL and NHSL estimators. But, at the same time, we have to note that more investigation of NHSL is needed in future. This may be done by making comparisons between NHSL and other ridge estimators.

Following McDonald and Galarneau (1975), Wichern and Churchill (1978), Gibbons (1981) and Kibria (2003), the explanatory variables were generated using the device

$$x_{ij} = (1-\gamma^2)^{\frac{1}{2}} z_{ij} + \gamma z_{ip} \; , i = 1,2,\ldots,n, \; j = 1,2,\ldots,p \qquad (15)$$

where z_{ij} are independent standard normal pseudo-random numbers, γ is specified so that the correlation between any two explanatory variables is given by γ^2, and p is the number of explanatory variables. The variables are then standardized so that $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{y}$ are in correlation forms. Different sets of correlation are considered corresponding to γ = 0.7, 0.8, 0.9 and 0.99. Using the condition number C_N, it can be shown that these values of γ will include a wide range of low, moderate and high correlations between variables. The n observations for the dependent variable y are determined by

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + e_i \quad , \; i = 1,\ldots,n \qquad (16)$$

where e_i are independent normal (0, σ^2) pseudo-numbers and β_0 is taken to be identically zero, and p is defined as in (15). We used three different sample sizes: 15, 25 and 30 with 5, 10 and 20 explanatory variables respectively. These choices of p are taken to study the behavior of the estimators for small, moderate and large number of explanatory variables.

For each set of explanatory variables, one choice for the coefficient vectors is considered. The MSE function depends on the explanatory variables (through λ_i), on σ^2 and on $\boldsymbol{\beta}$. It was noted (Newhouse and Oman ,1971) that if MSE is regarded as a function of $\boldsymbol{\beta}$ with σ^2, k and the explanatory variables are fixed, then, subject to the constraint that $\|\boldsymbol{\beta}\| = 1$, the MSE is minimized when $\boldsymbol{\beta}$ is the normalized eigenvector corresponding to the largest eigenvalue of the matrix $\mathbf{C}$. We didn't use normalized eigenvectors corresponding to the smallest eigenvalue because the conclusion about the performance of estimators in both cases will not change greatly (Kibria, 2003).

For given values of p , n and γ , the experiment was repeated 1000 times by generating 1000 samples. For each replicate r (r =1, 2,… , 1000), the values of k of different proposed estimators and the corresponding ridge estimators were calculated using

$$\hat{\boldsymbol{\alpha}}_R = (\boldsymbol{\Lambda} + \hat{k}\mathbf{I})^{-1}\mathbf{X}^{*\prime}\mathbf{y} \;\; , \; \hat{k} = \hat{k}_{HK} \, , \; \hat{k}_{HSL} \, , \; \hat{k}_{NHSL} \qquad (17)$$

Then the MSEs for estimators are calculated as follows

$$\text{MSE}(\hat{\alpha}_R) = \frac{1}{1000}\sum_{r=1}^{1000}(\hat{\alpha}_{(r)} - \alpha)'(\hat{\alpha}_{(r)} - \alpha) \qquad (18)$$

Results of the Simulation Study

In this section we present the results of our Monte Carlo experiments. Our primary interest here lies in comparing the MSEs of the considered estimators. The main results of simulation are summarized in Tables 1-3 below. To compare the performances of the considered estimators, we calculate the MSEs of each one. We consider the estimator that leads to the minimum MSE to be the best. It is worth mentioning here that we used the statistics package *Minitab14* to do all calculations that were made in this article.

Table 1. Estimated MSEs and the values of C_N with p = 5 and n =15.

	Estimators				
γ	OLS	HK	HSL	NHSL	C_N
0.99	0.0369229	0.0328831	0.0336641	0.0268559	879.62
0.9	0.0270955	0.0248932	0.0250849	0.0213376	80.79
0.8	0.0204500	0.0191982	0.0192228	0.0173943	35.29
0.7	0.0159889	0.0153413	0.0153420	0.0132687	19.39

Table 2. Estimated MSEs and the values of C_N with $p = 10$ and $n = 25$.

	Estimators				
γ	OLS	HK	HSL	NHSL	C_N
0.99	0.0179593	0.0173544	0.0174193	0.0156275	2743.34
0.9	0.0123661	0.0121384	0.0121454	0.0106224	240.29
0.8	0.0088826	0.0087805	0.0087810	0.0083277	99.86
0.7	0.0070337	0.0069838	0.0069838	0.0066757	57.91

Table 3. Estimated MSEs and the values of C_N with $p = 20$ and $n = 30$.

	Estimators				
γ	OLS	HK	HSL	NHSL	C_N
0.99	0.01385120	0.01348810	0.01350310	0.00661062	23185.20
0.9	0.00877632	0.00867149	0.00867239	0.00414520	2002.15
0.8	0.00633292	0.00628823	0.00628828	0.00304895	830.08
0.7	0.00481253	0.00479185	0.00479185	0.00465087	477.50

Looking at tables 1-3, we can see that the HK, HSL and NHSL are better than the OLS, and the NHSL performs better than the HK and HSL. The results also reveal that for high correlations, i.e., when γ=0.9 and 0.99, the HK and HSL perform almost equivalently. However, the HK produces somewhat lower MSEs than the HSL for all sets of correlation. Moreover, it is observed that for given n and p, the MSEs for all estimators increase as the correlation among the explanatory variables increases. In an opposite manner, for given γ, as the sample size and the number of explanatory variables increase, the MSEs of all estimators decrease.

CONCLUSION

In this article we have investigated the properties of a new proposed method for estimating the ridge parameter in the presence of multicollinearity. The investigation has been done using Monte Carlo experiments, where levels of correlation, the numbers of explanatory variables and the sample sizes have been varied. For each combination we have 1000 replications. The evaluation of our estimator has been done by comparing the MSEs of our proposed estimator with the OLS estimator and the estimators of Hocking et al. (1976) and Hoerl and Kennard (1970). We found that our estimator uniformly dominates the other estimators.

REFERENCES

1. Al-Hassan, Y. (2008). A Monte Carlo Evaluation of Some Ridge Estimators. *Jordan Journal of Applied Science* (JJAS), 10 (2): 101-110.
2. Alkhamisi, M. A., Khalaf, G. and Shukur, G. (2006). Some modifications for choosing ridge parameter. *Communications in Statistics-Theory and Method*s, 35:2005-2020.
3. Alkhamisi, M. A. and Shukur, G. (2007) A Monte Carlo study of recent ridge parameters. *Communications in Statistics, Simulation and Computation*, 36 (3): 535-547.

4. Draper, N. R. and Smith, H. (1998) *Applied Regression Analysis*, 3ed Ed., New York John Wiley and Sons,.
5. Gibbons, D. G. (1981). A simulation study of some ridge estimators. *Journal of the American Statistical Association*, 76 (373):131–139.
6. Hocking, R. R.(1976). The Analysis and Selection of Variables in Linear Regression. *Biometrics*, 32:1-49.
7. Hocking, R. R. , Speed, F. M. and Lynn, M. J. (1976). A class of biased estimators in linear regression. *Technometrics*, 18 (4):425-437.
8. Hoerl, A. E. and Kennard, R. W. (1970a). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12:55-67.
9. Hoerl, A. E. and Kennard, R. W. (1970b). Ridge regression: applications to nonorthogonal problems. *Technometrics*, 12:69-82.
10. Hoerl, A. E., Kennard, R. W. and Baldwin, K. F. (1975). Ridge regression: some simulation. *Communications in Statistics,* 4:105–123.
11. Khalaf, G. and Shukur, G. (2005). Choosing ridge parameter for regression problems. . *Communications in Statistics-Theory and Method*s, 34:1177-1182.
12. Kibria, B. M. (2003). Performance of some new ridge regression estimators. *Communications in Statistics- Simulation and Computation*, 32(2):419-435.
13. Lawless, J. F. (1978). Ridge and related estimation procedure. *Communications in Statistics,* A7:139–164.
14. Lawless, J. F. and Wang, P. (1976). A simulation study of ridge and other regression estimators. *Communications in Statistics*, A5:307–323.
15. McDonald, G. C. and Galarneau, D. I. (1975). A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*, 70 (350):407-412.
16. Newhouse, J. P. and Oman, S. D. (1971). An evaluation of ridge estimators. *Rand Corporation*, P-716-PR:1-28.
17. Nordberg, L. (1982). A procedure for determination of a good ridge parameter in linear regression. *Communications in Statistics*. A11:285–309.
18. Saleh, A. K. and Kibria, B. M. (1993). Performances of some new preliminary test ridge regression estimators and their properties. *Communications in Statistics-Theory and Methods*, 22:2747–2764.
19. Scolve, S. L. (1968). Improved Estimators for Coefficients in Linear Regression, Journal of American Statistical Association, 63:597-606.
20. Singh, S. and Tracy, D. S. (1999). Ridge-regression using scrambled responses. *Metrika*:147–157.
21. Stein, C. M. (1960). Multiple Regression, Contributions to Probability and Statistics, USA, Stanford University Press, 424-443.
22. Wencheko, E. (2000). Estimation of the signal-to-noise in the linear regression model. *Statistical Papers,* 41:327–343.
23. Wichern, D. and Churchill, G. (1978). A comparison of ridge estimators, *Technometrics*, 20:301–311.

تقييم أداء لمقدر انحدار حرف جديد

الحسن، يزيد

قسم الرياضيات، المعهد الملكي للتكنولوجيا، ستوكهولم، السويد
yazidah@kth.se ; yazid_12111980@yahoo.com

الملخص

تم اقتراح مقدر انحراف الحرف كبديل لمقدر المربعات الصغرى عند وجود ارتباط متعدد بين المتغيرات التوضيحية. اهتمت العديد من الدراسات المتعلقة بانحدار الحرف بطرق اختيار معلمة الحرف. ونتيجة لذلك، تم تقديم عدد من الطرق لاختيار معلمة الحرف. في هذا البحث تم اقتراح طريقة جديدة لاختيار معلمة الحرف، فلقد تم القيام بتقييم الطريقة المقترحة باستخدام المحاكاة و بمقارنة متوسط مربعات الخطأ للمقدر المقترح مع مقدرات أخرى. أشارت عملية التقييم إلى انه في حال و جود ارتباط متعدد بين المتغيرات التوضيحية فان المقدر المقترح يُظهر أفضلية على المقدرات الأخرى.