The Traditional Ordinary Least Squares Estimator under Collinearity

Ghadban AK* and Iguernane M
King Khalid University, Saudi Arabia

Abstract

In a multiple regression analysis, it is usually difficult to interpret the estimator of the individual coefficients if the explanatory variables are highly inter-correlated. Such a problem is often referred to as the multicollinearity problem. There exist several ways to solve this problem. One such way is ridge regression. Two approaches of estimating the shrinkage ridge parameter $k$ are proposed. Comparison is made with other ridge-type estimators. To investigate the performance of our proposed methods with the traditional ordinary least squares (OLS) and the other approaches for estimating the parameters of the ridge regression model, we calculate the mean squares error (MSE) using the simulation techniques. Results of the simulation study shows that the suggested ridge regression outperforms both the OLS estimator and the other ridge-type estimators in all of the different situations evaluated in this paper.

Keywords: Linear regression model; Multicollinearity; Ridge regression estimators; Simulation study

Mathematics Subject Classification: Primary 62J07; Secondary 62J05

Introduction

Consider the standard multiple linear regression model:

$$Y = X + e,$$  (1)

where $Y$ is an $(n\times 1)$ vector of responses, $X$ is an $(n \times p)$ matrix of the explanatory variables of full rank $p$, $\beta$ is a $(p \times 1)$ vector of unknown regression coefficients, and finally, $e \sim N(0, \sigma^2I)$ is an $(n \times 1)$ vector of error terms.

The OLS estimator is often used to estimate the regression coefficients $\beta$ as:

$$\hat{\beta} = (XX')^{-1}XY.$$  (2)

The standard assumption in the linear regression analysis is that all the explanatory variables are linearly independent. When this assumption is violated, the problem of multicollinearity enters into the data and it inflates the variance of an ordinary least squares estimator of the regression coefficient. Obtaining the estimators for multilinear data is an important problem in the literature. In fact, when the problem of multicollinearity is present in the measurement error ridden data, then an important issue is how to obtain the consistent estimators of regression coefficients. One of the most popular estimator for combating multicollinearity is the ridge estimator, originally proposed by Hoerl et al. [1]. They suggested a small positive number $k>0$ to be added to the diagonal elements of the $X'X$ matrix from the multiple regression and the resulting estimators are obtained as:

$$\hat{\beta}(k) = (XX + kI)^{-1}X'Y,$$  (3)

which is known as a ridge regression estimator. For a positive value of $k$, this estimator provides a smaller MSE compared to the OLS estimator, i.e.,

$$\text{MSE}(\hat{\beta}(k)) < \text{MSE}(\hat{\beta}).$$

Most of the later efforts in this area have concentrated on estimating the value of the ridge parameter $k$. Many different techniques for estimating $k$ have been proposed by different researchers, for example, Hoerl et al. [1], Hoerl et al. [2] Dempster et al. [3], Gibbons [4], Kibria [5], Khalaf et al. [6], Alkhamis et al. [7], Khalaf [8] and Khalaf [9].

The plan of the paper is as follows: in Section 2, we present different methods for estimating the parameter of ridge regression together with our proposed estimators. A simulation study has been conducted in Section 3. The simulation results are discussed in Section 4. In Section 5 we give a brief summary and conclusions.

The Proposed Ridge Regression Parameter

In case of ordinary ridge regression, many researchers have suggested different ways of estimating the ridge parameter. Hoerl et al. [1] showed, by letting $\beta_{\text{max}}$ denote the maximum of the $\beta$, that choosing:

$$\hat{k}_{\text{HE}} = \frac{\sigma^2}{\beta_{\text{max}}^2},$$  (4)

implies that $\text{MSE}(\hat{\beta}(k)) < \text{MSE}(\hat{\beta})$. The ridge estimator using $\hat{k}_{\text{HE}}$ will be denoted by HK.

Hoerl et al. [2] suggested that, the value of $k$ is chosen small enough, for which the MSE of ridge estimator is less than the MSE of OLS estimator. They showed, through simulation, that the use of the ridge with biasing parameter given by:

$$\hat{k}_{\text{HK}} = \frac{p\sigma^2}{\beta^2},$$  (5)

has a probability greater than 0.50 of producing estimator with a smaller MSE than the OLS estimator, where $\hat{\sigma}^2$ is the usual estimator of $\sigma^2$, defined by $\hat{\sigma}^2 = \frac{1}{n-p-1}(Y'X\hat{\beta} - X'X\hat{\beta})$. The ridge estimator using Eq. (5) will be denoted by HKB.

The purpose of this study is to modify the approaches of estimating $k$ mentioned in Hoerl and Kennard [1] and Hoerl et al. [2] given in

*Corresponding author: Ghadban AK, Department of Mathematics, Faculty of Science, King Khalid University, Saudi Arabia, Tel: 0172418000; E-mail: albadran50@yahoo.com

Received November 23, 2015; Accepted December 02, 2015; Published December 09, 2015

Citation: Ghadban AK, Iguernane M (2015) The Traditional Ordinary Least Squares Estimator under Collinearity. J Biom Biostat 6: 264. doi:10.4172/2155-6180.1000264

Copyright: © 2015 Ghadban AK, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
The performance of these proposed estimators will be then compared with the traditional OLS estimation and those of HK and HKB estimators in terms of MSE. This will mainly be done by means of simulations under conditions where the sample size \( n \), the number of explanatory variables \( p \), the size of each simulation \( R \) equals 5000 which corresponds to the number of replications used in the situation.

### The Simulation Results

Ridge estimators are constructed with the aim of having smaller MSE than the MSE of the OLS estimator. Improvement, if any, can therefore be studied by looking at the amount of the MSE. These MSEs are reported in Tables 1 and 2. The results of our simulation study indicate that ridge estimators outperform OLS estimator in all cases and the suggested estimators \( KI_1 \) and \( KI_2 \) performed very well in this study. They appear to offer an opportunity for large reduction in MSE, especially when the sample size and the correlation between the explanatory variables are high (Table 1).

![Citation: Ghadban AK, Iguernane M (2015) The Traditional Ordinary Least Squares Estimator under Collinearity. J Biom Biostat 6: 264.](doi:10.4172/2155-6180.1000264)
Table 1: Estimated MSE when $p = 7$.

| $\rho$ | $\sigma^2$ | $n$ | OLS | HK | HKB | $K_l$ | $K_s$ |
|-------|-----------|-----|-----|----|-----|-------|-------|
| 0.94  | $0.8^2$   | 15  | 47.988 | 16.688 | 10.105 | 2.680 | 2.805 |
|       |           | 25  | 5.773  | 3.915  | 2.268  | 1.645 | 1.755 |
|       |           | 80  | 1.3607 | 1.2231 | 0.8749 | 0.8433| 0.8977 |
|       |           | 200 | 0.5184 | 0.4978 | 0.4182 | 0.4223| 0.4427 |
| 0.98  | $0.8^2$   | 15  | 158.08 | 49.069 | 31.913 | 4.064 | 4.102 |
|       |           | 25  | 18.470 | 8.957  | 5.164  | 2.215 | 2.250 |
|       |           | 80  | 4.364  | 3.193  | 1.864  | 1.453 | 1.507 |
|       |           | 200 | 1.6500 | 1.4402 | 0.9722 | 0.9059| 0.9551 |
| 0.9   | $0.8^2$   | 15  | 6.440  | 3.459  | 2.228  | 1.539 | 1.690 |
|       |           | 25  | 0.7611 | 0.7140 | 0.6001 | 0.5955| 0.6229 |
|       |           | 80  | 0.1761 | 0.1745 | 0.1676 | 0.1671| 0.1698 |
|       |           | 200 | 0.0659 | 0.0657 | 0.0646 | 0.0647| 0.0651 |

| $\rho$ | $\sigma^2$ | $n$ | OLS | HK | HKB | $K_l$ | $K_s$ |
|-------|-----------|-----|-----|----|-----|-------|-------|
| 0.94  | $0.8^2$   | 15  | 26.715 | 10.069 | 6.155  | 2.219 | 2.383 |
|       |           | 25  | 3.049  | 2.386  | 1.489  | 1.293 | 1.407 |
|       |           | 80  | 0.7101 | 0.6728 | 0.5464 | 0.5474| 0.5761 |
|       |           | 200 | 0.2645 | 0.2594 | 0.2367 | 0.2395| 0.2466 |
| 0.98  | $0.8^2$   | 15  | 44.747 | 14.785 | 9.448  | 2.656 | 2.783 |
|       |           | 25  | 5.184  | 3.587  | 2.097  | 1.567 | 1.673 |
|       |           | 80  | 1.2181 | 1.1079 | 0.8093 | 0.7888| 0.8403 |
|       |           | 200 | 0.4629 | 0.4463 | 0.3803 | 0.3841| 0.4012 |
| 0.9   | $0.8^2$   | 15  | 150.20 | 44.75  | 30.23  | 3.92  | 3.97  |
|       |           | 25  | 16.983 | 8.301  | 4.901  | 2.175 | 2.213 |
|       |           | 80  | 3.896  | 2.923  | 1.726  | 1.384 | 1.438 |
|       |           | 200 | 1.4606 | 1.2923 | 0.8920 | 0.8437| 0.8918 |
| 0.94  | $0.8^2$   | 15  | 5.848  | 3.082  | 1.985  | 1.436 | 1.590 |
|       |           | 25  | 0.7038 | 0.6558 | 0.5655 | 0.5622| 0.5860 |
|       |           | 80  | 0.1639 | 0.1622 | 0.1559 | 0.1562| 0.1581 |
|       |           | 200 | 0.0620 | 0.0618 | 0.0609 | 0.0610| 0.0613 |
| 0.98  | $0.8^2$   | 15  | 23.002 | 9.022  | 5.371  | 2.153 | 2.339 |
|       |           | 25  | 2.750  | 2.196  | 1.383  | 1.221 | 1.326 |
|       |           | 80  | 0.6726 | 0.6394 | 0.5243 | 0.5256| 0.5521 |
|       |           | 200 | 0.2439 | 0.2395 | 0.2198 | 0.2225| 0.2288 |
If we focus on the values of $\sigma^2$, $\rho$ and the sample size $n$, we find that among the ridge estimators considered, $K_1^I$ is the best followed by HKB and $K_{1}^I$. In comparing Tables 1 and 2 which involve $p = 7$ and $p = 10$, respectively, we find that the MSEs are lowest for Table 1. This is to say that the ridge estimators are more helpful when high multicollinearity exists and the number of explanatory variables is not large.

Conclusions

In this article, we introduce two alternatives ridge estimators and study their performance using simulation techniques. Comparisons are made with other ridge type estimators evaluated elsewhere. The results from the simulation study show that the sample size, the correlation between the independent variables and the number of explanatory variables are important factors for the performance of the different estimation methods. In most of the cases, the MSE decreases when the first two factors increase.

The result also shows that, with respect to MSE criteria, the proposed ridge regression methods outperform both the OLS estimator and the estimators of Hoerl et al. [1] and Hoerl et al. [2] in all cases investigated. The use of the proposed estimators is recommended since it reduces the MSE substantially in all of the different situations investigated in this paper.

References

1. Hoerl AE, Kennard RW (1970) Ridge Regression: Biased Estimation for non-orthogonal Problems. Technometrics 12: 55-67.
2. Hoerl AE, Kennard RW, Baldwin KF (1975) Ridge Regression: Some Simulation. Communications in Statistics - Theory and Methods 4: 105-124.
3. Dempster AP, Schatzoff M, Wermuth N (1977) A simulation study of alternatives to ordinary least squares. J Amer Statist Assoc 72: 77-91.
4. Gibbons DG (1981) A simulation study of some ridge estimators. Journal of the American Statistical Association 76: 131-139.
5. Kibria BMG (2003) Performance of some New Ridge Regression Estimators.
Communication in Statistics-Theory and Methods 32: 419-435.

6. Khalaf G, Shukur G (2005) Choosing Ridge Parameters for Regression Problems. Communication in Statistics - Theory and Methods 34: 1177-1182.

7. Alkhamisi M, Shukur G (2008) Developing Ridge Parameters for SUR Model. Communication in Statistics-Theory and Methods 37: 544-564.

8. Khalaf G (2011) Ridge Regression: An Evaluation to some New Modifications. International Journal of Statistics and Analysis 1: 325-342.

9. Khalaf G (2013) A Comparison between Biased and Unbiased Estimators. JMASM 12: 293-303.

10. McDonald GC, Galarneau DI (1975) A Monte Carlo evaluation of some ridge-type estimators. J Amer Statist Assoc 70: 407-416.