연세경제연구 제Ⅲ권, 제2호
Yonsei Economic Studies, Vol. Ⅲ, No. 2 (September 1996)

Limited Use of Ridge Regression

Wonwoo Lee

Ridge regression is the most popular approach to the multicollinearity problem when it presents in the multiple linear regression model. However, it is not a panacea since the performance of ridge regression is dependent upon the values of the parameters.

Ⅰ. Introduction

The multiple linear regression(MLR) model

$$y = X\beta + \varepsilon$$

has been broadly used in various application areas, where y is an $(n \times 1)$ responses vector, X is the $(n \times p)$ standardized regressors matrix, β is a $(p \times 1)$ parameter vector, and ε is an $(n \times 1)$ error vector with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2 I$. However, in practice, the researchers are confronted with the multicollinearity problem seriously so that the alternative methods to the least squares (LS) method have been proposed. Among them, there is no most powerful alternative because multicollinearity works curiously enough (Sujan [8], Draper and Nostrand [4], etc.). Furthermore, several authors (Newhouse and Oman [7], Bingham and Larntz [2], Draper and Nostrand [4]) have founded the fact that the ridge estimators did worse than the LS estimators for at least some values of the parameter vectors. Also, Lee [6] has shown, with simulation studies that for some choices of the parameter vectors the principal component regression should cautiously be used.

In this paper, consider reparameterizations of the MLR model as follows;

President, AID Consulting Institute, 625-11 Shinsa-dong, Gangnam-gu, Seoul, 135-120, Korea.

$$y = X\beta + \varepsilon \quad (1)$$
$$= Z\alpha + \varepsilon \text{ (eigenvalue decomposition)} \quad (2)$$
$$= U\gamma + \varepsilon \text{ (singular value decomposition)} \quad (3)$$

where $Z = XV$, $\alpha = V'\beta$, $X = ULV'$, and $\gamma = LV'\beta$ (Belsley, Kuh, and Welsch [1]). Note that the columns of V are the eigenvectors of $X'X$, and those of U are the eigenvectors corresponding to the positive eigen values of XX'. Furthermore, $L = \Lambda^{1/2}$, where Λ is the diagonal matrix of the eigenvalues of $X'X$ and, without loss of generality, the eigenvalues, $\lambda_1, \cdots, \lambda_p$ are ordered in magnitude, i.e., $\lambda_1 \geq \cdots \geq \lambda p$.

In Section 2, based on the framework of fractional principal components regression, the condition for which the ridge estimators work better than the LS estimators in terms of mean square error(MSE) will be developed. Problems for ridge regression are displayed with respect to the values of the parameters in Section 3 and the problems are also graphically illustrated. Finally, the concluding remarks are in Section 4.

Ⅱ. Conditions for $MSE(a_R) \leq MSE(a_{LS})$

When multicollinearity is present in the MLR model, the class of the fractional principal components(FPC) estimators can be defined as the class of the biased estimators which are the potential substitutes for the LS estimator. One typical member of the class is expressed by

$$a_{FPC} = F\, a_{LS}, \quad (4)$$

where a_{LS} is the LS estimator of α in (2) and $F = diag(f_1, \cdots, f_p)$, $0 \leq f_i \leq 1$, $i = 1, \cdots, p$. The ith fraction, f_i of ridge regression will be

$$f_i = \frac{\lambda_i}{\lambda_i + k}$$

and its fracion matrix, F is

$$F = \Lambda(\Lambda + kI)^{-1}, \quad (5)$$

where k is a positive shrinkage parameter. Thus, the ridge estimator of α can be written as

$$a_R = \Lambda(\Lambda + kI)^{-1} a_{LS} \tag{6}$$

Some properties of the ridge estimator are as follows; (*i*) it is biased since $E(a_R) = F\alpha = \alpha - (I-F)\alpha$ and (*ii*) the variance-covariance matrix of a_R is $Var(a_R) = \sigma^2 F\Lambda^{-1}F$. In order to compare the performances of a_R and a_{LS} in terms of MSE, it is of interest to consider the matrix form of MSE, denoted by $MtxMSE(a_R)$,

$$\begin{aligned} MtxMSE(a_R) &= Var(a_R) + Bias(a_R)\,Bias(a_R)' \\ &= \sigma^2 F\Lambda^{-1}F + (I-F)\alpha\,\alpha'(I-F) \end{aligned} \tag{7}$$

That is, in order for a_R to improve over a_{LS} either $MSE(a_{LS}) - MSE(a_R)$ or $MtxMSE(a_{LS}) - MtxMSE(a_R)$ can be used since the former is the nonnegative if and only if the latter is positive semi-definite (see Theobald [9]).

Therefore, the condition for which $MtxMSE(a_{LS}) - MtxMSE(a_R)$ is positive semi-definte can be obtained by showing that $\Lambda^{-1}(I-F^2) - \sigma^{-2}(I-F)\alpha\,\alpha'(I-F)$ is positive semi-difinite. Consequently, using Cauchy-Schwarz inequality, the necessary and sufficient condition for $MSE(a_R) \leq MSE(a_{LS})$ is

$$\alpha'\Lambda(I-F)^2(I-F^2)^{-1}\alpha \leq \sigma^2 \tag{8}$$

(see Lee [6]). By considering the individual diagonals, the term $(I-F)^2(I-F^2)^{-1}$ in (8) can be described as $k(2\Lambda + kI)^{-1}$ so that the condition is rewritten as

$$k\,\alpha'\Lambda(2\Lambda + kI)^{-1}\alpha \leq \sigma^2. \tag{9}$$

Or, equivalently, since $\gamma = \Lambda^{1/2}\alpha$, the condition in terms of γ is

$$\gamma'(2k^{-1}\Lambda + I)^{-1}\gamma \leq \sigma^2. \tag{10}$$

The inequality in (8) is dependent upon (*i*) the choice of α, (*ii*) the degree of multicollinearity as indicated by Λ, and (*iii*) the value of σ^2. The combinations of three factors make the ridge estimator out of use at some situations in the next section.

Ⅲ. Limitations of Ridge Regression

It has been said that the ridge estimator is not always better than the LS estimator even though multicollinearity is inspected in the MLR model. Since the necessary and sufficient condition for $MSE(a_R) \leq MSE(a_{LS})$ in (8) does not always hold, the structures of the parameters (α , σ^2) with the degree of multicollinearity play important roles in using of ridge regression.

For simplicity, assuming the number of the regressors is 2 and the squared norm of γ is given as $\gamma' \gamma = c$, where c is a constant, consider the ellipse $\gamma'(2k^{-1}\Lambda + I)^{-1}\gamma = \sigma^2$ and the circle $\gamma' \gamma = c$ in the γ_1 and γ_2 axes. Then, if

$$\text{(Case I)} \quad \sqrt{c} < \sigma\sqrt{1+2\lambda_2/k} \tag{11}$$

as shown in Figure 1, the ellipse includes the circle since $\lambda_1 > \lambda_2$. In this situation, since any parameter vector γ lies inside the ellipse, the condition in (8) holds. In other words, ridge regressing works wells in terms of MSE.

The interesting case is when the ellipse meets the circle as in Figure 2. If

$$\text{(Case II)} \quad \sigma\sqrt{1+2\lambda_2/k} \leq \sqrt{c} \leq \sigma\sqrt{1+2\lambda_1/k} \tag{12}$$

then for only the limited parameter vector γ the ridge estimator can be recommended instead of the LS estimator.

In other words, the parameter vector γ along the chords AB and CD satisfy the condition (10) But for any parameter vector along BC and DA of the circle, the LS estimation outperforms the ridge estimation in terms of MSE. Note that the coordinates of A, B, C, and D with respect to (γ_1, γ_2) are

$$\left(\pm\left[\left(\sigma^2 - \frac{kc}{2\lambda_2+k}\right)\Big/\left(\frac{k}{2\lambda_2+k} - \frac{k}{2\lambda_1+k}\right)\right]^{1/2},\right.$$

$$\left.\pm\left[\left(\sigma^2 - \frac{kc}{2\lambda_1+k}\right)\Big/\left(\frac{k}{2\lambda_2+k} - \frac{k}{2\lambda_1+k}\right)\right]^{1/2}\right), \tag{13}$$

respectively.

Finally, if the ellipse in inside the fixed circle (see Figure 3), that is,

$$\text{(Case III)} \quad \sigma\sqrt{1+2\lambda_2/k} \leq \sqrt{c}, \tag{14}$$

then the ridge estimator should not be used. It is the case for which the LS estimator has smaller MSE than the ridge estimator.

Furthermore, (Case II) provides the guideline of the value of the signal-to-noise ratio in terms of γ, $\gamma'\gamma/\sigma^2$, For the two regressors, if

Figure 1 (Case I) Ellipse of $\gamma'(2k^{-1}\Lambda+I)^{-1}\gamma=\sigma^2$

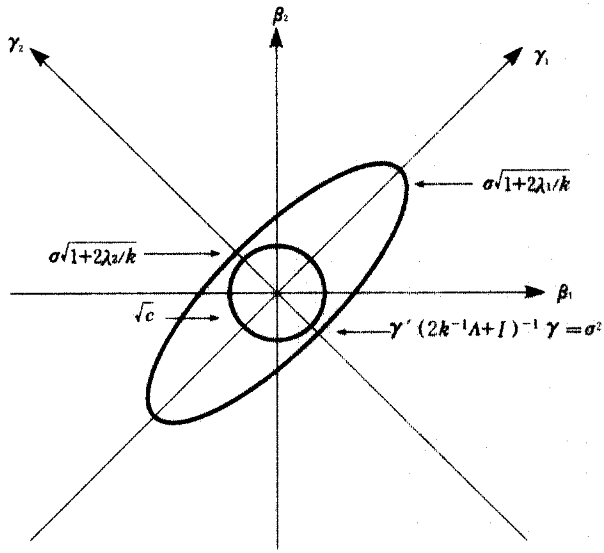


Figure 2 (Case II) Ellipse of $\gamma'(2k^{-1}\Lambda+I)^{-1}\gamma=\sigma^2$

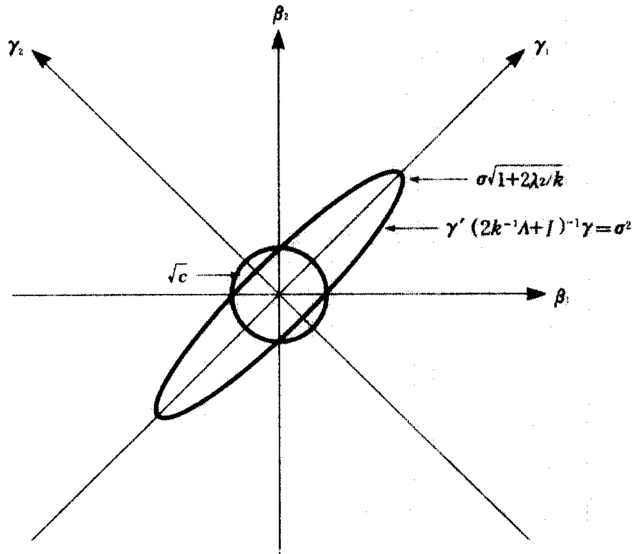
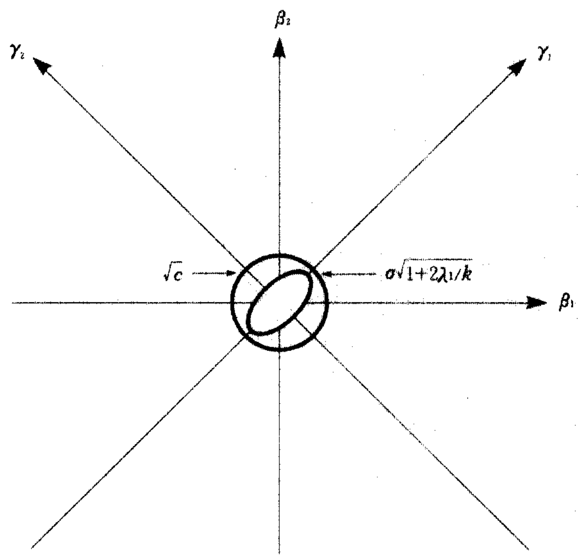


Figure 3 (Case Ⅱ) Ellipse of $\gamma' (2k^{-1}\Lambda + I)^{-1} \gamma = \sigma^2$



$$1 + 2\lambda_1/k \leq \sigma^{-2} \gamma' \gamma \qquad (15)$$

then ridge regression is of no use. Sujan [8] and Gunst and Mason [5] have felt that ridge regression tends to give worse results with very high signal-to-noise ratios. Even for the general MLR model, it is not difficult to argue that the inequality in (15) is the guideline for the use of ridge regression when λ_1 is the largest eigenvalue of $X'X$.

Ⅳ. Concluding Remarks

In the multiple linear regression model the presence of multicollinearity is a serious problem which cannot be completely resolved. In general, it is said that ridge estimator works well when multicollinearity problem presents in the MLR model. However, the use of ridge regression should be scrutinized. It is not a panacea.

In this paper, the choice of rules for selecting the ridge shrinkage, k, is open to discussion.

◈ *REFERENCES* ◈

1. Belsley, D. A., Kuh, E. and Welsch, R. E., *Regression Diagnostic : Identifying Influential Data and Sources of Collinearity*, 1980, New York : John Wiley and Sons Inc..

2. Bingham, C. and Larntz, K. Comment, *Journal of the American Statistical Association*, 72, 1977, pp. 97~102.

3. Dempster, A. P., Schatzoff, M. and Wermuth, N., "A Simulation Study of Alternatives to Ordinary Least Squares," *Journal of the American Statistical Association*, 72, 1977, pp. 77~91.

4. Draper, N. R. and Nostrand, R. C. V., "Ridge Regression and James-Stein Estimation : Review and Comments," *Technometrics*, 21, 1979, pp. 451~466.

5. Gunst, R. F. and Mason, R. L., "Biased Estimation in Regression : An Evaluation Using Mean Squared Error," *Journal of the American Statistical Association*, 72, 1977, pp. 616~628.

6. Lee, Wonwoo, "A New Deletion Criterion of Principal Components Regression with Orientations of the Parameters," *Journal of the Korean Statistical Society*, 16(2), 1987, pp. 55~70.

7. Newhouse, J. P. and Oman, S. D., "An Evaluation of Ridge Estimators," *Rand*, R−716−PR, 1971, pp. 1~28.

8. Sujan, I., "Methods for Solving the Problem of Multicollinearity in Econometric Models: A Survey," *Multicollinearity and Biased Estimation*, 1980, Göttingen: Vandenhoeck & Ruprecht.

9. Theobald, C. M., "Generalizations of Mean Square Error Applied to Ridge Regression," *Journal of the Royal Statistical Society*, B. 36, 1974, pp. 103~106.

10. Toro-Vizcarrondo, C. and Wallace, T. D., "A Test of the Mean Square Error Criterion for Restrictions in Linear Regression," *Journal of the American Statistical Association*, 63, 1968, pp. 558~572.

11. Vinod, H. D., "Simulation and Extension of a Minimum Mean Squared Error Estimator in Comparison with Stein's," *Technometrics*, 18, 1976, pp. 491~496.