

Solving Multicollinearity Problem Using Ridge Regression Models

Yewon Kim

12/03/2015

Introduction

In this paper, they introduce many different Methods of ridge regression to solve multicollinearity problem. These methods include Ordinary Ridge Regression(ORR), Generalized Ridge Regression(GRR), and Directed Ridge Regression(DRR).

Some properties of ridge regression estimators and methods of selecting biased ridge regression parameter are discussed. They use data simulation to make comparison between methods of ridge regression and Ordinary Least Squares (OLS) method. According to a results of this study, they found that all methods of ridge regression are better than OLS method when the Multicollinearity is exist.

Multicollinearity

Multicollinearity refers to a situation in which or more predictor variables in a multiple regression Model are highly correlated if Multicollinearity is perfect (EXACT), the regression coefficients are indeterminate and their standard errors are infinite, if it is less than perfect. The regression coefficients although determinate but posses large standard errors, which means that the coefficients can not be estimated with great accuracy (Gujarati, 1995).

Methods used to detect multicollinearity

- ▶ Compute the correlation matrix of predictors variables
- ▶ Eigen structure of $X^T X$
- ▶ Variance inflation factor (VIF)
- ▶ Checking the relationship between the F and T tests

Effects

- ▶ High variance of coefficients may reduce the precision of estimation.
- ▶ Multicollinearity can result in coefficients appearing to have the wrong sign.
- ▶ Estimates of coefficients may be sensitive to particular sets of sample data.
- ▶ Some variables may be dropped from the model although, they are important in the population.
- ▶ The coefficients are sensitive of to the presence of small number inaccurate data values (more details in Judge 1988, Gujarati; 1995).

The ordinary ridge regression (ORR)

$$Y = X\beta + \epsilon$$

where Y is $(n \times 1)$ vector of the response variable values, X is $(n \times p)$ matrix contains the values of P predictor variables and this matrix is full Rank (matrix of rank p), β is a $(p \times 1)$ vector of unknown coefficients, and ϵ is a $(n \times 1)$ vector of normally distributed random errors with zero mean and common variance σ^2 . Note that, Both X 's and Y have been standardized.

The ordinary ridge regression (ORR)

The ordinary least square (OLS) estimate $\hat{\beta}$ of β is obtained by:
$$\hat{\beta} = (X^T X)^{-1} X^T Y, \quad \text{VAR}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}, \quad \text{MSE}(\hat{\beta}) = \sigma^2 \sum_{i=1}^P \frac{1}{\lambda_i}$$

The ridge solution is given by:

$$\hat{\beta}(K) = (X^T X + KI)^{-1} X^T Y, \quad K \geq 0$$

Note that, if $K=0$, the ridge estimator become as the OLS. If all K 's are the same, the resulting estimators are called the ordinary ridge estimators (John, 1998).

The ordinary ridge regression (ORR)

$$\text{MSE}(\hat{\beta}(K)) = \hat{\sigma}^2 \sum_{i=1}^P \frac{\lambda_i}{\lambda_i + K} + K^2 \hat{\beta}^T (X^T X + KI)^{-2} \hat{\beta}.$$

(More details see Judge, 1988, Gujarati; 1995, Gruber 1998, Pasha and Shah 2004) This means that $\text{MSE}(\hat{\beta}(K)) < \text{MSE}(\hat{\beta})$. There always exists a $K > 0$, such that $\text{MSE}(\hat{\beta}(K))$ has smaller than $\text{MSE}(\hat{\beta})$

The generalized ridge regression (GRR)

Let P is a $(p \times p)$ matrix with columns as eigenvectors of $X^T X$.
Then the linear model can be written as

$$Y = X\beta + \epsilon = (XP)(P^T\beta) + \epsilon = X^*\alpha + \epsilon$$

The ridge estimator for α is given by

$$\alpha(\hat{K}) = (X^{*T}X^* + K)^{-1}X^{*T}Y$$

The directed ridge regression (DRR)

Guilkey and Murphy (1975), proposed a technique called Directed Ridge Regression. This method of estimation based on the relationship between the eigenvalues of $X^T X$ and the variance of α_j . Since $\text{Var}(\alpha_j) = \sigma^2 \Lambda^{-1}$, relatively precise estimation is achieved for corresponding to large eigenvalues, while relatively imprecise estimation is achieved for α_j corresponding to small eigenvalues.

As a result of adjusting only those elements of Λ^{-1} corresponding to the small eigenvalues of $X^T X$, the DRR estimator results in an estimate of α_j that is less biased than the resulting from GRR estimator.

Choice of ridge parameter K

The ridge regression estimator does not provide a unique solution to the problem of multicollinearity but provides a family of solutions. These solutions depend on the value of K (the ridge biasing parameter). For example:

Hoerl, Kennard and Baldwin (1975), $\hat{K}(HKB) = P\hat{\sigma}^2/\hat{\beta}^T\hat{\beta}$ and
Lawless and Wang (1976), $\hat{K}(LW) = P\hat{\sigma}^2/\hat{\beta}^T X^T X \hat{\beta}$

Example

In this research, they simulate a set of data using SAS package, where the correlation coefficients between the predictor variables (X 's) are large (the number of predictor variables in this study are six variables).

Table (1) Correlation Matrix

| | X_1 | X_2 | X_3 | X_4 | X_5 | X_6 |
|-------|-------|-------|-------|-------|-------|-------|
| X_1 | 1 | | | | | |
| X_2 | 0.944 | 1 | | | | |
| X_3 | 0.957 | 0.948 | 1 | | | |
| X_4 | 0.959 | 0.973 | 0.946 | 1 | | |
| X_5 | 0.972 | 0.969 | 0.973 | 0.963 | 1 | |
| X_6 | 0.936 | 0.961 | 0.979 | 0.937 | 0.970 | 1 |

First : A sample of size 50 observations is simulated for 2000 iterations, and the following indicators are computed as :

Example

- The eigenvalues of the predictor correlation matrix are as follow :

$$\lambda_1 = 5.798$$

$$\lambda_2 = 0.084$$

$$\lambda_3 = 0.064$$

$$\lambda_4 = 0.027$$

$$\lambda_5 = 0.015$$

$$\lambda_6 = 0.011$$

- The variance inflation factor (VIF) for all variables X's are as follow :

$$\text{VIF}(X_1) = 24.7$$

$$\text{VIF}(X_2) = 37.01$$

$$\text{VIF}(X_3) = 42.47$$

$$\text{VIF}(X_4) = 29.03$$

$$\text{VIF}(X_5) = 46.93$$

$$\text{VIF}(X_6) = 43.22$$

Example

Using both OLS method and all methods of Ridge Regression to analyze the simulated data, they get the following results :

Table (2)

Estimates of coefficients and standard deviation using OLS and RR method

| OLS | | ORR ₁ | | ORR ₂ | | GRR | | DRR | |
|--------|-------|------------------|-------|------------------|-------|-------|-------|-------|-------|
| Coef | STDEV | Coef | STDEV | Coef | STDEV | Coef | STDEV | Coef | STDEV |
| 1.301 | 0.206 | 1.172 | 0.204 | 1.287 | 0.204 | 0.378 | 0.007 | 0.378 | 0.007 |
| -0.403 | 0.241 | -0.333 | 0.205 | -6.395 | 0.236 | 0.478 | 0.159 | 0.470 | 0.164 |
| -0.115 | 0.218 | -.034 | 0.184 | -0.106 | 0.212 | 0.789 | 0.211 | 0.783 | 0.215 |
| 0.565 | 0.203 | 0.507 | 0.171 | 0.559 | 0.159 | 0.323 | 0.161 | 0.439 | 0.150 |
| -0.714 | 0.399 | -0.553 | 0.27 | -0.697 | 0.293 | 0.909 | 0.379 | 1.024 | 0.411 |
| 0.309 | 0.357 | 0.180 | 0.203 | 0.294 | 0.221 | 0.488 | 0.311 | 0.674 | 0.341 |

Example

| Method | MSE |
|--------|-------|
| OLS | 0.432 |
| ORR1 | 0.36 |
| ORR2 | 0.403 |
| GRR | 0.322 |
| DRR | 0.42 |

Example

Table (3) R^2 for each Model

| R-Square | |
|-----------------|-------|
| OLS | 0.779 |
| ORR1 | 0.975 |
| ORR2 | 0.781 |
| GRR | 0.911 |
| DRR | 0.911 |

Example

From the previous results, it is obvious that :

- ▶ All models of RR have smaller standard deviation than OLS.
- ▶ All models of RR have smaller MSE of regression coefficient than OLS.
- ▶ While, all models of RR have larger R^2 than OLS. consequently, all models of RR are better than OLS when the multicollinearity problem is exist in data.

Conclusion

In This research, they referred to the multicollinearity problem, methods of detecting of this problem and effect on a result of multiple regression model. Also, they introduced many different models of ridge regression to solve this problem and make a comparison between RR methods and OLS by using a simulation data (2000 replications). Based on the standard deviation, MSE and R^2 for estimators of each model, they noted that all ridge regression models are better than ordinary least square when the multicollinearity problem is exist and the best model is the generalized ridge regression because it has smaller MSE of estimators, smaller standard deviation for most estimators and has larger coefficient of determination

References



M.El-Dereny and N.I.Rashwan

Solving Multicollinearity Problem Using Ridge Regression Models

Int.J.Contemp.Math.Science, Vol.6,2011 12,585-600

The End