A class of Liu-type estimator based on ridge regression under multicollinearity with an application to mixture experiments

Ai-Chun Chen (陳 愛 群)
Advisor: Takeshi Emura
Graduate Institute of Statistics, National Central University

Abstract: In the linear regression, the least squared estimator does not perform well when multicollinearity exists. The problem of multicollinearity occurs in industrial mixture experiments, where regressors are constrained. Hoerl and Kennard (1970) proposed the ordinary ridge estimator to overcome the problem of the least squared estimator under multicollinearity. This paper considers a special class of Liu-type estimators (1993, 2003). We derive the theoretical formula of the mean square error for the proposed method. We perform simulations to compare the proposed estimator with the ridge estimator in terms of mean square error. We demonstrate this special class using the dataset on Portland cement with mixture experiment (Woods et al., 1932).

1. Introduction

In a regression model, when multicollinearity exists, the ordinary least square (OLS) estimator is inappropriate. Ridge regression is an alternative estimator derived by Hoerl and Kennard (1970). Ridge regression aims to reduce the large variance by shrinking the OLS estimator toward zero. Hoerl and Kennard (1970) showed that there exists some range of shrinkage parameter such that the total mean squared error (MSE) is smaller than that of the OLS. Multicollinearity arises in mixture experiments, where explanatory variables are proportions of a mixture. Ridge has been successfully applied to overcome the multicollinearity among the proportion (Jang and Anderson-Cook, 2014). We proposed a ridge-type estimator and study the properties by theory and simulation.

2. Background

Usual model with intercept is

$$\mathbf{y}_{n\times 1} = \mathbf{X}_{n\times(p+1)}\boldsymbol{\beta}_{(p+1)\times 1} + \boldsymbol{\varepsilon}_{n\times 1}, \quad \mathbf{X} = [\,\mathbf{1}\;\mathbf{X}_p\,],$$

where $\boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I})$, and the design matrix is fixed and standardized. The model is written in the canonical form:

$$\mathbf{y} = \mathbf{A}\boldsymbol{\alpha} + \boldsymbol{\varepsilon},$$

where $\mathbf{A} = \mathbf{X}\boldsymbol{\Gamma}, \boldsymbol{\alpha} = \boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\beta}$ and $\boldsymbol{\Gamma}_{(p+1)\times(p+1)}$ is an orthogonal matrix defined as:

$$\boldsymbol{\Gamma} = \begin{bmatrix} 1 & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0} & \boldsymbol{\Gamma}_p \end{bmatrix},$$

It follows that $\boldsymbol{\Gamma}_p^{\mathrm{T}}\mathbf{X}_p^{\mathrm{T}}\mathbf{X}_p\boldsymbol{\Gamma}_p = \boldsymbol{\Lambda}_p = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ and $\lambda_1 \geq \cdots \geq \lambda_p > 0$, where λ_i are eigenvalues of $\mathbf{X}_p^{\mathrm{T}}\mathbf{X}_p$.

2.1 Multicollinearity and mixture experiments

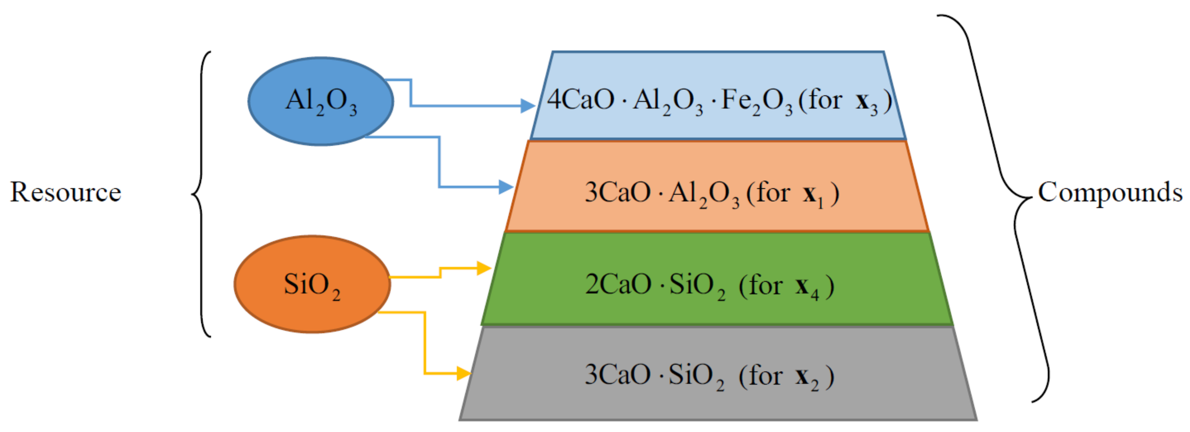
The ordinary least squared (OLS) estimator of $\boldsymbol{\beta}$ is solved by minimizing the residual sum of squares.

$$\hat{\boldsymbol{\beta}}^{\mathrm{OLS}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}.$$

And

$$\mathrm{tr}\{\,\mathrm{V}(\hat{\boldsymbol{\beta}}^{\mathrm{OLS}})\,\} = \sigma^2\left(\frac{1}{n} + \sum_{j=1}^{p}\frac{1}{\lambda_j}\right).$$

When multicollinearity occurs, there exists a least one small eigenvalue, say $\lambda_r = 0.001$.Then, The total variance of OLS estimator is quite large.
A motivating example to illustrate the multicollinearity is the dataset on Portland cement (Woods et al., 1932).



No. of Cement	SiO_2	Al_2O_3	Fe_2O_3	CaO
1	27.7	3.8	2.0	65.0
2	26.0	3.5	5.1	63.1
3	21.9	5.7	2.8	65.0
4	24.6	5.8	2.8	64.2
5	25.0	3.9	2.1	66.6

No. of Cement	$4CaO \cdot Al_2O_3 \cdot Fe_2O_3$ $\mathbf{x}_3$	$3CaO \cdot Al_2O_3$ $\mathbf{x}_1$	$2CaO \cdot SiO_2$ $\mathbf{x}_4$	$3CaO \cdot SiO$ $\mathbf{x}_2$
1	6	7	60	26
2	15	1	52	29
3	8	11	20	56
4	8	11	47	31
5	6	7	33	52

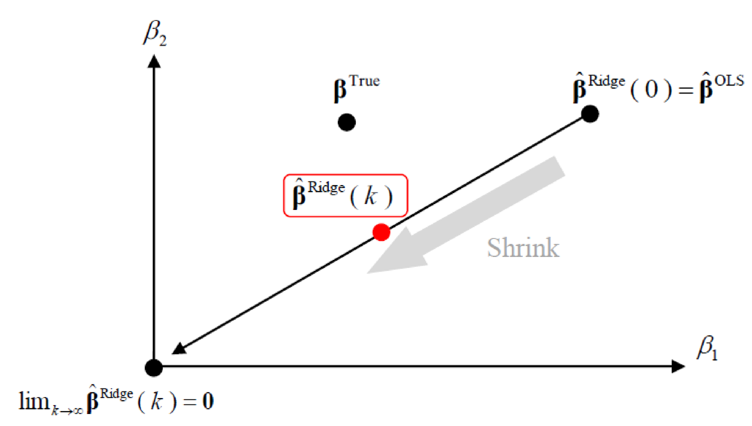
2.2 Ridge regression

The original definition of ridge is to add a diagonal matrix to the information matrix of the OLS estimator

$$\hat{\boldsymbol{\beta}}^{\mathrm{Ridge}}(k) = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

In another view, the ridge regression gives a penalty to the residual sum of squares (RSS),

$$\mathrm{RSS}^{\mathrm{Ridge}} = \|\,\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\,\|^2 + k\|\,\boldsymbol{\beta}\,\|^2.$$



3. Proposed method

Consider a penalized RSS, defined as

$$\mathrm{RSS}^{\mathrm{Ridge}} = \|\,\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\,\|^2 + k\|\,\boldsymbol{\beta} - \boldsymbol{\beta}^*\,\|^2.$$

where $\boldsymbol{\beta}^*$ can be any estimator of $\boldsymbol{\beta}$.
Thus, the new estimator is defined by

$$\hat{\boldsymbol{\beta}}^{\mathrm{New}}(k) = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}^{\mathrm{T}}\mathbf{y} + k\boldsymbol{\beta}^*)$$

Following Emura et al. (2012), we estimate $\boldsymbol{\beta}^*$ by the so-called compound univariate estimator.

Definition (Compound Univariate Estimator):
We use a univariate model

$$y_i = \beta_0 + \varepsilon_i$$

to estimate β_0^*. *And again use univariate model*

$$y_i = \hat{\beta}_0^* + \beta_j x_{ij} + \varepsilon_i$$

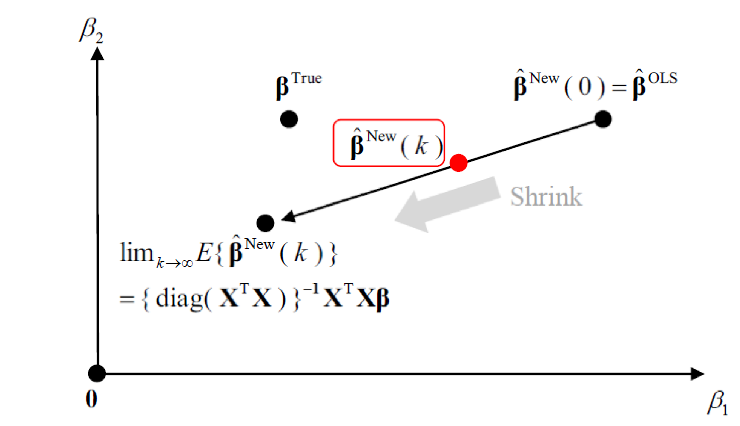
to estimate β_j^*.

The compound univariate estimator can be expressed as

$$\hat{\boldsymbol{\beta}}^* = \{\,\mathrm{diag}(\,\mathbf{X}^{\mathrm{T}}\mathbf{X}\,)\,\}^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}.$$

Therefore, the proposed estimator takes the form

$$\hat{\boldsymbol{\beta}}^{\mathrm{New}}(k) = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + k\mathbf{I})^{-1}[\,\mathbf{I} + k\{\,\mathrm{diag}(\,\mathbf{X}^{\mathrm{T}}\mathbf{X}\,)\,\}\,]\mathbf{X}^{\mathrm{T}}\mathbf{y}.$$



4. Theory

4.1 Mean squared error (MSE) calculation

Consider a linear estimator of the form $\tilde{\boldsymbol{\beta}} = C\mathbf{y}$. Then, the MSE of $\tilde{\boldsymbol{\beta}} = C\mathbf{y}$ is calculated as

$$MSE(\tilde{\boldsymbol{\beta}}) = bias(\tilde{\boldsymbol{\beta}})bias(\tilde{\boldsymbol{\beta}})^{\mathrm{T}} + \mathrm{var}(\tilde{\boldsymbol{\beta}}).$$

We provide the bias and variance of the proposed estimator

$$\mathrm{bias}\{\hat{\boldsymbol{\beta}}^{\mathrm{New}}(k)\}^{\mathrm{T}}\mathrm{bias}\{\hat{\boldsymbol{\beta}}^{\mathrm{New}}(k)\} = \sum_{i=1}^{p}\frac{k^2\alpha_i^2(\lambda_i - n + 1)^2}{(\lambda_i + k)^2(n-1)^2},$$

$$\mathrm{v}\{\hat{\boldsymbol{\beta}}^{\mathrm{New}}(k)\} = \sigma^2\left\{\frac{1}{n} + \sum_{i=1}^{p}\frac{\lambda_i(k + n - 1)^2}{(\lambda_i + k)^2(n-1)^2}\right\}.$$

where α_i is the i-th component of $\boldsymbol{\alpha}$ as in background.

The most valuable feature of the MSE function is the value of 1st order derivatives in the neighborhood of the zero. We have shown that

$$\lim_{k\to 0^+}\frac{d}{dk}\mathrm{B}(k) = 0 \qquad \lim_{k\to 0^+}\frac{d}{dk}\mathrm{V}(k) < 0.$$

5. Simulation

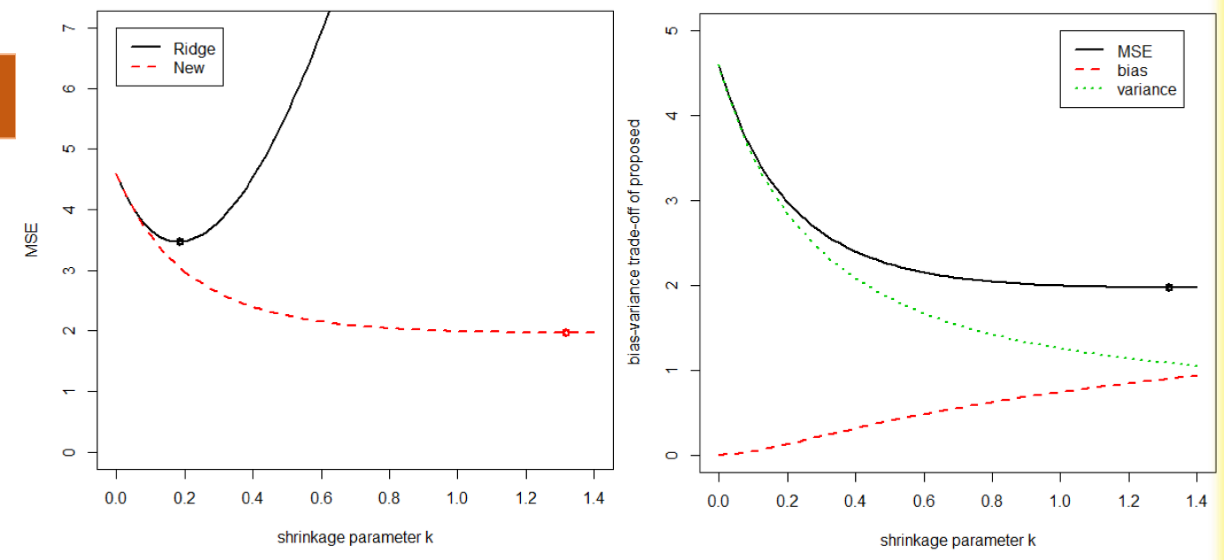
We perform our simulation under $n = 13$ and $p = 4$ for

$$\sigma^2 = 1 \text{ or } 2 \text{ and } \boldsymbol{\beta} = (\beta_0, 1, 1, 1, 1)^{\mathrm{T}}, \beta_0 = 50 \text{ or } 1.$$
$$\mu_1 = 12, \mu_2 = 25, \tau = 5.$$

Generate design matrix by following:

$$\begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix} \sim N_2\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \tau^2\mathbf{I}_2\right), \quad \begin{bmatrix} x_{i3} \\ x_{i4} \end{bmatrix} \sim \begin{bmatrix} -x_{i1} \\ -x_{i2} \end{bmatrix} + N_2\left(\begin{bmatrix} 50 \\ 50 \end{bmatrix}, \tau^2\mathbf{I}_2\right)$$

where we set "seed(1)" to generate "rnorm" function in R package.



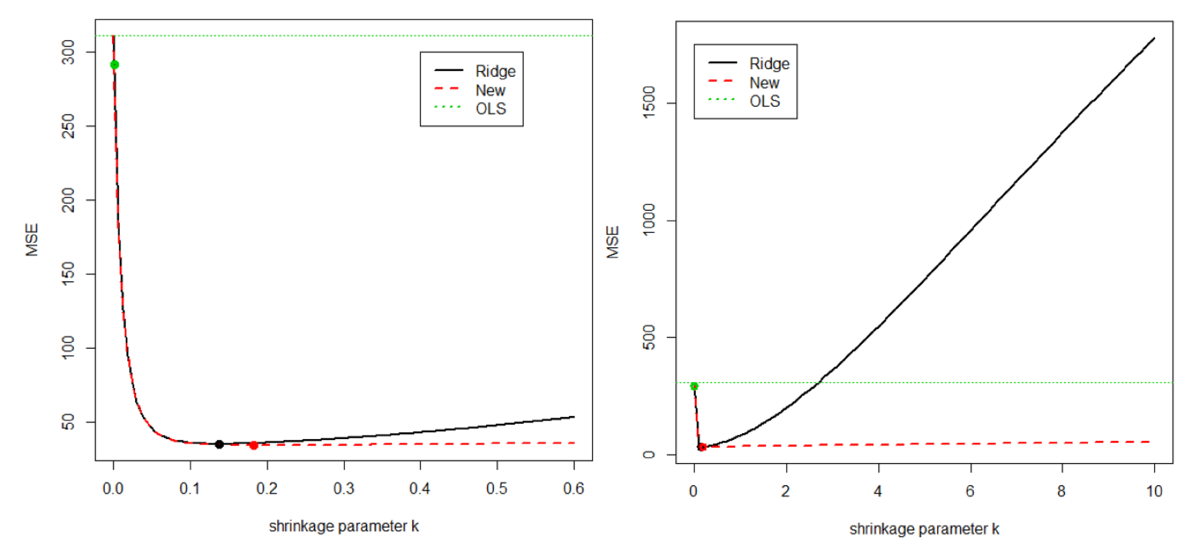
Simulation of MSE condition on $\boldsymbol{\beta} = (50, 1, 1, 1, 1)^{\mathrm{T}}$ and $\sigma^2 = 2$

6. Data analysis

We use the Portland cement data to check the proposed estimator. This data has $n = 13$ and $p = 4$. The sample correlation matrix of $\mathbf{X}_p$ is

$$\text{Sample Corr}(\,\mathbf{X}_p\,) = \begin{bmatrix} 1.000 & -0.133 & -0.848 & 0.082 \\ -1.133 & 1.000 & 0.245 & -0.952 \\ -0.848 & 0.245 & 1.000 & -0.139 \\ 0.082 & -0.952 & -0.139 & 1.000 \end{bmatrix}.$$

where we can have that multicollinearity occurs.



If we forget to standardized the model, it brings a result as follow:

$\hat{\boldsymbol{\beta}}$	β_0	β_1	β_2	β_3	β_4	Bias	Var	$\hat{m}$
$\hat{\boldsymbol{\beta}}^{\mathrm{OLS}}$	62.40	1.55	0.51	0.10	–0.14	0	4912.09	4912.09
$\hat{\boldsymbol{\beta}}^{\mathrm{Ridge}}(\hat{k}^{\mathrm{Ridge}})$	27.50	1.91	0.86	0.46	0.20	1218.73	952.25	2170.98
$\hat{\boldsymbol{\beta}}^{\mathrm{New}}(\hat{k}^{\mathrm{New}})$	64.14	1.53	0.49	0.08	–0.16	724.71	159.84	884.56

7. Conclusion

1. Proposed a new method for regressor coefficient estimation.
2. The method provides good estimate on MSE criterion in some case.

8. Reference

1. Emura, T., Chen Y. H. and Chen H. Y. (2012). *PLoS ONE* **7**, DOI: 10.1371/journal.pone.0047627.
2. Hoerl, A. E. and Kennard, R. W. (1970). *Technometrics* **12**, 55-67.