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

Presenter: Ai-Chun Chen（陳愛群）
Advisor: Takeshi Emura

June 16, 2015

Graduate Institute of Statistics, NCU

Outline

- Introduction
- Methodology
- Thoery
- Numerical analysis
- Conclusion

Introduction – Model

- Linear model with intercept

$$\mathbf{y}_{n\times 1} = \mathbf{X}_{n\times(p+1)}\boldsymbol{\beta}_{(p+1)\times 1} + \boldsymbol{\varepsilon}_{n\times 1}, \quad \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$$

$$\text{with } \mathbf{X} = [\,\mathbf{1}\ \ \mathbf{X}_p\,] = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}$$

- Standardization of the design matrix

$$\sum_{i=1}^{n} x_{ij} = \bar{x}_j = 0 \quad \text{and} \quad \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2 = 1 \ for\ j = 1, \ldots, p$$

Introduction – Background

- Ordinary least square (OLS) estimator

Minimize the residual sum of squares (RSS)

$$\mathrm{RSS} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

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

- Pros: Unbiased and with minimum variance

Introduction – Multicollinearity problem

- Multicollinearity

$$\mathbf{X} = [\,\mathbf{1}\ \ \mathbf{X}_p\,] = [\,\mathbf{1}\ \ \underbrace{\mathbf{x}_1\ \mathbf{x}_2 \cdots \mathbf{x}_p}_{\text{Nearly linear dependent}}\,]$$

(Montgomery et al., 2012)

- Problems

There exist a least one small eigenvalue λ_j of $\mathbf{X}_p^{\mathrm{T}}\mathbf{X}_p$

$$\Rightarrow \quad \operatorname{var}(\hat{\boldsymbol{\beta}}^{\mathrm{OLS}}) = \sigma^2\left(\frac{1}{n} + \underbrace{\sum_{j=1}^{p}\frac{1}{\lambda_j}}_{\text{Too large}}\right)$$

(Jimichi, 2005)

Introduction – Mixture experiments

- Response depends only on the proportions of the ingredients in the mixture (Cornel, 2011)
- Example: Make a cake

x_1 : Flour

x_2 : Water

x_3 : Egg

Satisfy

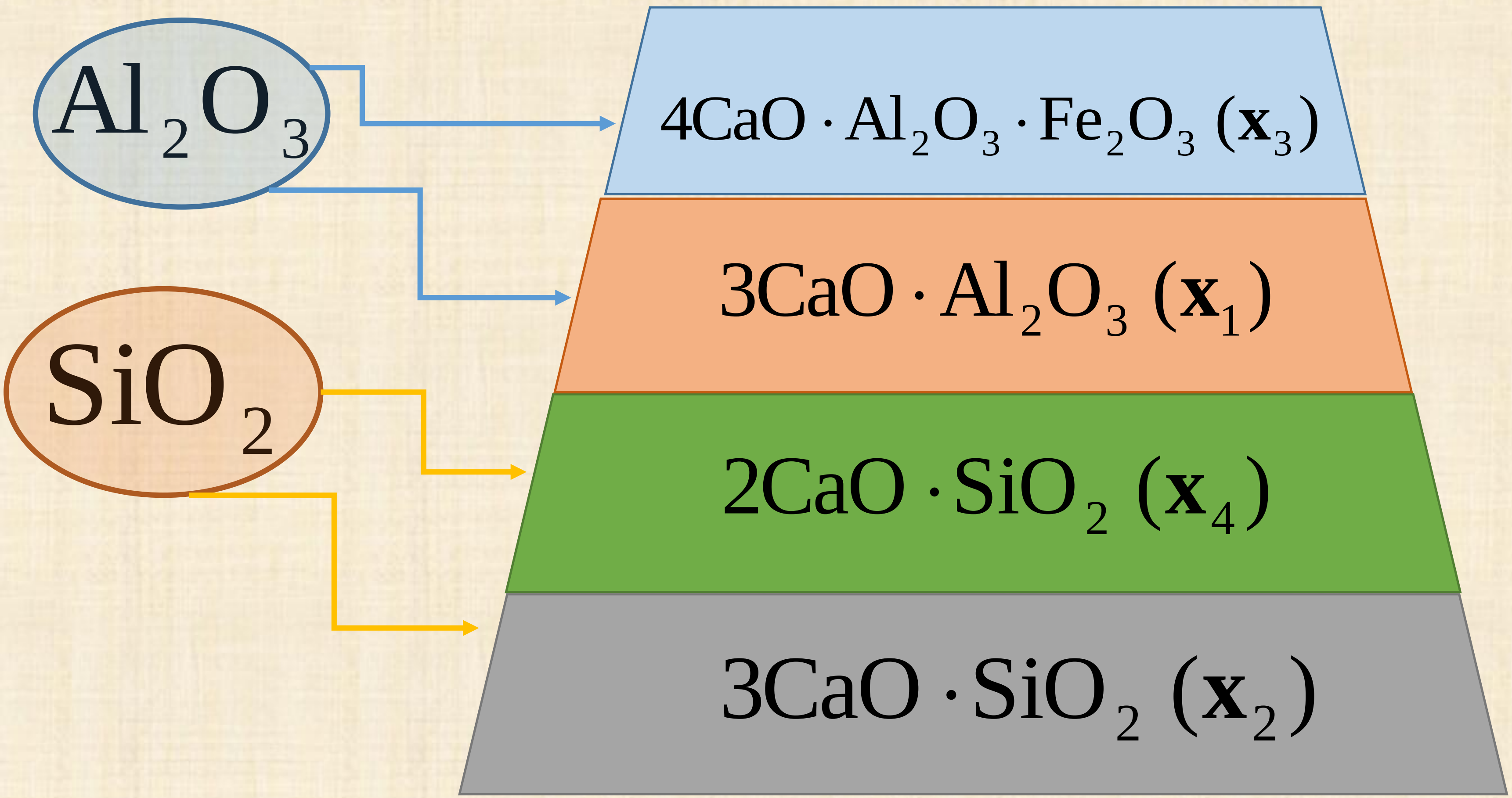
$$\sum_{j=1}^{3} x_j = 1$$

Constraints

$$0 \le L_j \le x_j \le U_j \le 1, \quad j = 1, 2, 3$$

Introduction – Motivating example

- Portland cement data (Woods et al., 1932)



Introduction – Ridge regression

- Hoerl and Kennard (1970)

$$\hat{\boldsymbol{\beta}}^{\text{Ridge}}(k) = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + k\mathbf{I}_{(p+1)})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}, \quad k \geq 0$$

Pros:

Solve the multicollinearity problem on OLS

What can improve:

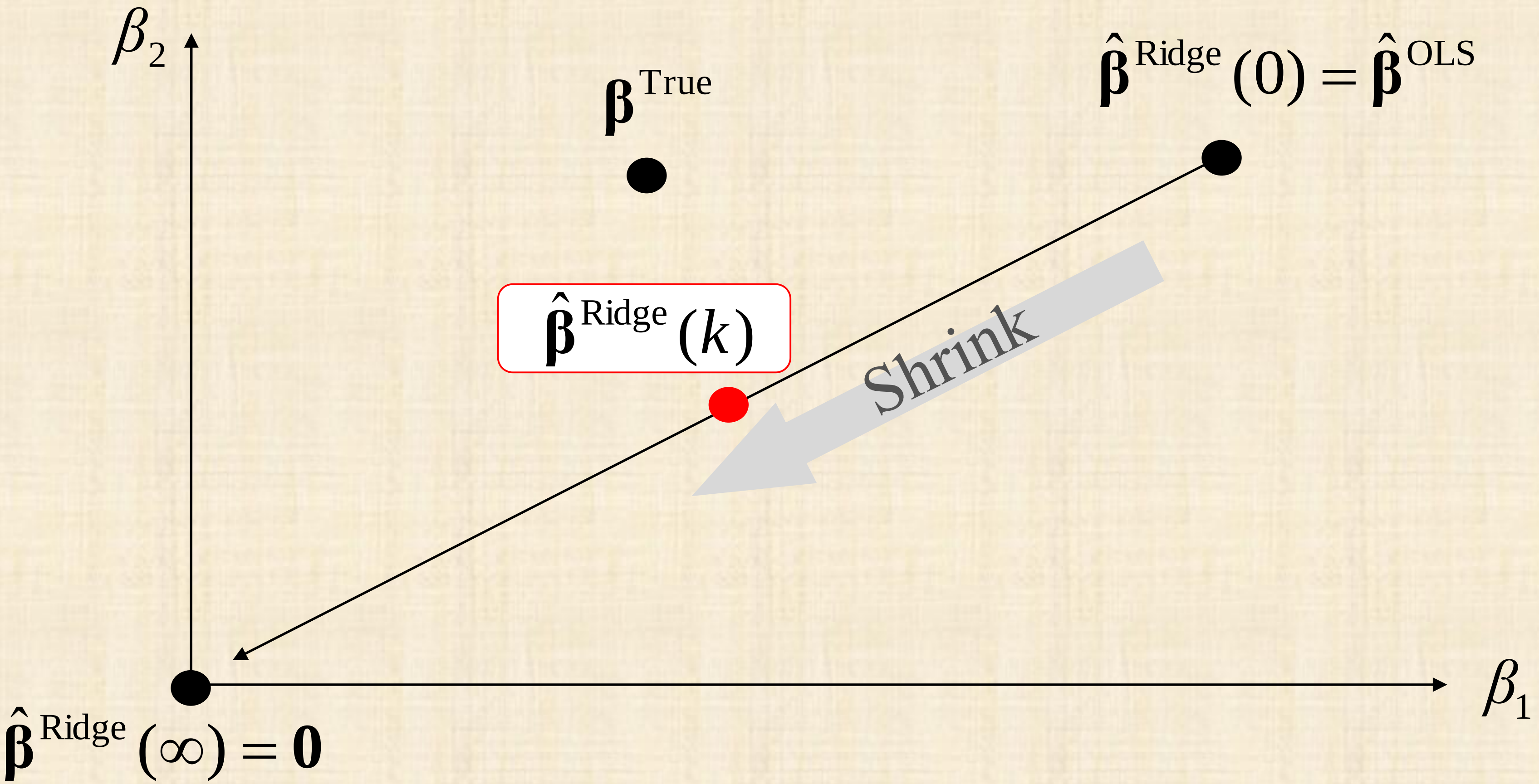
Intercept term consideration ?

Mean squared error performance ?

Introduction – Ridge regression

- In the view of RSS

$$\mathrm{RSS}^{\mathrm{Ridge}} = \mathrm{RSS} + \mathrm{penalty} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + k\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{\beta}, \quad k \geq 0$$



Introduction – Other ridge-type estimators

- Liu (1993) Liu estimator

$$\hat{\boldsymbol{\beta}}^{\text{Liu}}(d) = (\mathbf{X}^{\text{T}}\mathbf{X} + \mathbf{I})^{-1}(\mathbf{X}^{\text{T}}\mathbf{y} + d\,\hat{\boldsymbol{\beta}}^{\text{OLS}}), \quad 0 \le d \le 1$$

- Liu (2003) Liu-type estimator

$$\hat{\boldsymbol{\beta}}^{\text{Liu}}_{k,d} = (\mathbf{X}^{\text{T}}\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}^{\text{T}}\mathbf{y} - d\boldsymbol{\beta}^{*}), \quad k > 0, \quad -\infty < d < \infty$$

$\boldsymbol{\beta}^{*}$ can be any estimator of $\boldsymbol{\beta}$.

- Sakallıoğlu and Kaçıranlar (2006)

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

$$k > 0, \quad -\infty < d < \infty$$

Methodology – Proposed method

- New penalty

$$\mathrm{RSS}^{\mathrm{New}} = \mathrm{RSS} + \mathrm{penalty}^{*} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + k(\boldsymbol{\beta} - \boldsymbol{\beta}^{*})^{\mathrm{T}}(\boldsymbol{\beta} - \boldsymbol{\beta}^{*}),\ k \geq 0$$

where $\boldsymbol{\beta}^{*}$ can be any estimator of $\boldsymbol{\beta}$.

- New estimator

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

Note: Proposed is a special class of Liu-type estimator (Liu, 2003)

- How to estimate $\boldsymbol{\beta}^{*}$?

Methodology – Proposed method

Definition 1 (Compound Univariate Estimator)

i) *Use univariate model* $y_i = \beta_0 + \varepsilon_i, i = 1, \ldots, n$ *to estimate* β_0^*

$$\Rightarrow \quad \hat{\beta}_0^* = \bar{y}$$

ii) *Use univariate model* $y_i = \hat{\beta}_0^* + \beta_j x_{ij} + \varepsilon_i, i = 1, \ldots, n$ *to estimate* $\beta_j^*, \quad j = 1, \ldots, p$

$$\Rightarrow \quad \hat{\beta}_j^* = \frac{\sum_{i=1}^{n} x_{ij} y_i}{\sum_{i=1}^{n} x_{ij}{}^2}$$

iii) *Compound univariate estimator is the compound of all the univariate estimators.*

Emura et al. (2012)

Methodology – Proposed method

Emura et al. (2012)

Compound univariate estimator

$$\hat{\boldsymbol{\beta}}^{*} = \begin{bmatrix} \bar{y} \\ \dfrac{\sum_{i=1}^{n} x_{i1} y_i}{\sum_{i=1}^{n} x_{i1}^{2}} \\ \vdots \\ \dfrac{\sum_{i=1}^{n} x_{ip} y_i}{\sum_{i=1}^{n} x_{ip}^{2}} \end{bmatrix} = \{\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\mathbf{X})\}^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{\mathrm{New}}(k) &= (\mathbf{X}^{\mathrm{T}}\mathbf{X} + k\mathbf{I})^{-1}(\mathbf{X}^{\mathrm{T}}\mathbf{y} + k\hat{\boldsymbol{\beta}}^{*}) \\ &= (\mathbf{X}^{\mathrm{T}}\mathbf{X} + k\mathbf{I})^{-1}[\mathbf{X}^{\mathrm{T}}\mathbf{y} + k\{\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\mathbf{X})\}^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}] \\ &= (\mathbf{X}^{\mathrm{T}}\mathbf{X} + k\mathbf{I})^{-1}[\mathbf{I} + k\{\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\mathbf{X})\}^{-1}]\mathbf{X}^{\mathrm{T}}\mathbf{y} \end{aligned}$$

Methodology – Proposed method

- If forget or do not want to standardize

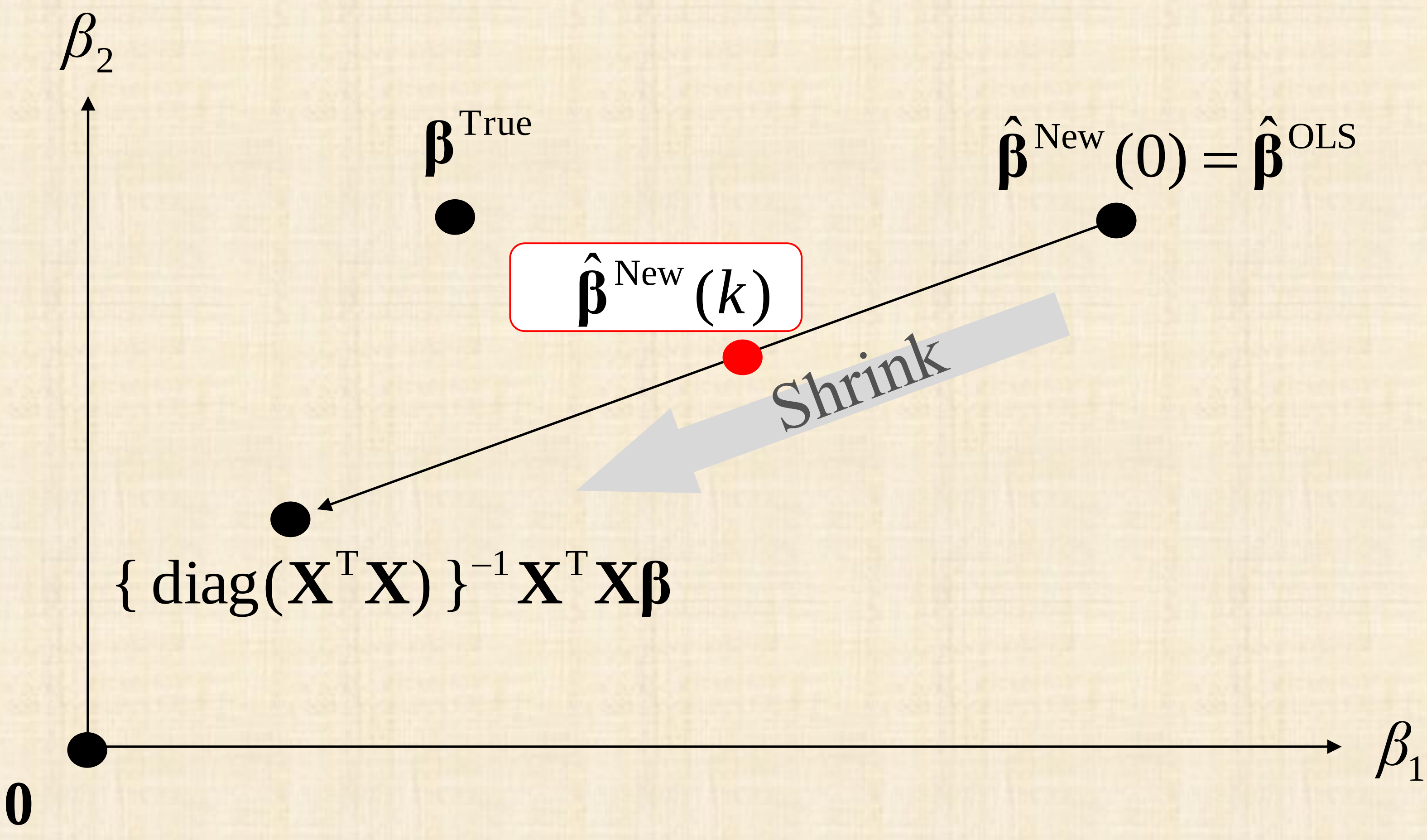
$$\hat{\boldsymbol{\beta}}^* = \begin{bmatrix} \bar{y} \\ \dfrac{\sum_{i=1}^{n} x_{i1}(y_i - \bar{y})}{\sum_{i=1}^{n} x_{i1}^2} \\ \vdots \\ \dfrac{\sum_{i=1}^{n} x_{ip}(y_i - \bar{y})}{\sum_{i=1}^{n} x_{ip}^2} \end{bmatrix}$$

Compound univariate estimator doesn't have simple form

Still use $\{\operatorname{diag}(\mathbf{X}^{\mathrm{T}}\mathbf{X})\}^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$ to estimate $\boldsymbol{\beta}^*$

Methodology – Proposed method

- Shrinkage scheme for the proposed method



Theory – Mean squared error calculation

- Total mean squared error (TMSE) calculation

Consider a linear estimator $\tilde{\boldsymbol{\beta}} = C_{(p+1)\times n}\mathbf{y}$

$$\begin{aligned}\text{TMSE}(\tilde{\boldsymbol{\beta}}) &= E(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\mathrm{T}}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\ &= \text{bias}(\tilde{\boldsymbol{\beta}})^{\mathrm{T}}\text{bias}(\tilde{\boldsymbol{\beta}}) + \text{v}(\tilde{\boldsymbol{\beta}})\end{aligned}$$

$$\text{bias}(\tilde{\boldsymbol{\beta}}) = (C\mathbf{X} - \mathbf{I})\boldsymbol{\beta} \quad \text{and} \quad \text{v}(\tilde{\boldsymbol{\beta}}) = \sigma^2 \text{trace}\, CC^{\mathrm{T}}$$

Theory – Mean squared error calculation

- We consider estimators

$$\hat{\boldsymbol{\beta}} = C^{\text{OLS}}\mathbf{y} \qquad \hat{\boldsymbol{\beta}}^{\text{Liu}}_{k,d} = C^{\text{Liu}}_{k,d}\mathbf{y}$$

$$\hat{\boldsymbol{\beta}}^{\text{Ridge}}(k) = C^{\text{Ridge}}(k)\mathbf{y} \qquad \hat{\boldsymbol{\beta}}^{\text{SK}}(k,d) = C^{\text{SK}}(k,d)\mathbf{y}$$

$$\hat{\boldsymbol{\beta}}^{\text{Liu}}(d) = C^{\text{Liu}}(d)\mathbf{y} \qquad \hat{\boldsymbol{\beta}}^{\text{New}} = C^{\text{New}}(k)\mathbf{y}$$

$$C^{\text{OLS}} = (\mathbf{X}^{\text{T}}\mathbf{X})^{-1}\mathbf{X}^{\text{T}}, \quad C^{\text{Ridge}}(k) = (\mathbf{X}^{\text{T}}\mathbf{X} + k\mathbf{I}_{(p+1)})^{-1}\mathbf{X}^{\text{T}}$$

$$C^{\text{Liu}}(d) = (\mathbf{X}^{\text{T}}\mathbf{X} + \mathbf{I}_{(p+1)})^{-1}\{\mathbf{I}_{(p+1)} + d(\mathbf{X}^{\text{T}}\mathbf{X})^{-1}\}\mathbf{X}^{\text{T}}$$

$$C^{\text{Liu}}_{k,d} = (\mathbf{X}^{\text{T}}\mathbf{X} + k\mathbf{I}_{(p+1)})^{-1}\{\mathbf{X}^{\text{T}} - dC^{\text{Ridge}}(k)\}$$

$$C^{\text{SK}}(k,d) = (\mathbf{X}^{\text{T}}\mathbf{X} + \mathbf{I}_{(p+1)})^{-1}\{\mathbf{X}^{\text{T}} + dC^{\text{Ridge}}(k)\}$$

$$C^{\text{New}}(k) = (\mathbf{X}^{\text{T}}\mathbf{X} + k\mathbf{I}_{(p+1)})^{-1}[\mathbf{I}_{(p+1)} + k\{\text{diag}(\mathbf{X}^{\text{T}}\mathbf{X})\}^{-1}]\mathbf{X}^{\text{T}}$$

Theory – Model in canonical form

Let $\lambda_1 \geq \ldots \geq \lambda_p > 0$ be the eigenvalues of $\mathbf{X}_p^{\mathrm{T}}\mathbf{X}_p$ and

$\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p$ be the correspond eigenvectors

$$\Rightarrow \quad \boldsymbol{\Gamma}_p^{\mathrm{T}}\mathbf{X}_p^{\mathrm{T}}\mathbf{X}_p\boldsymbol{\Gamma}_p = \boldsymbol{\Lambda}_p$$

where $\boldsymbol{\Gamma}_p = [\,\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p\,]$ and $\boldsymbol{\Lambda}_p = diag(\lambda_1, \ldots, \lambda_p)$

- Model in canonical form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \Rightarrow \quad \mathbf{y} = \mathbf{A}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$$

$\mathbf{A} = \mathbf{X}\boldsymbol{\Gamma},\ \boldsymbol{\alpha} = \boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\beta}$

with $\boldsymbol{\Gamma} = \begin{bmatrix} 1 & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0} & \boldsymbol{\Gamma}_p \end{bmatrix}$ and $\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathrm{diag}(n, \lambda_1, \ldots, \lambda_p)$

Theory – Mean squared error calculation

Lemma 1 (bias and total variance of new estimator)

i) Bias square

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

ii) Total variance

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

Theory – Mean squared error calculation

Theorem 1 (TMSE of the new estimator)

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

Lemma 2 (Derivatives for the bias square of new estimator)

$$\frac{d}{dk}\mathrm{bias}\{\hat{\boldsymbol{\beta}}^{\mathrm{New}}(k)\}^{\mathrm{T}}\mathrm{bias}\{\hat{\boldsymbol{\beta}}^{\mathrm{New}}(k)\}=\sum_{i=1}^{p}\frac{2k\alpha_i^2\lambda_i(\lambda_i-n+1)^2}{(\lambda_i+k)^3(n-1)^2}$$

$$\frac{d^2}{dk^2}\mathrm{bias}\{\hat{\boldsymbol{\beta}}^{\mathrm{New}}(k)\}^{\mathrm{T}}\mathrm{bias}\{\hat{\boldsymbol{\beta}}^{\mathrm{New}}(k)\}=\sum_{i=1}^{p}\frac{2\alpha_i^2\lambda_i(\lambda_i-2k)(\lambda_i-n+1)^2}{(\lambda_i+k)^4(n-1)^2}$$

Theory – Mean squared error calculation

Lemma 3 (Derivatives for total variance of new estimator)

$$\frac{d}{dk}\mathrm{v}\{\hat{\boldsymbol{\beta}}^{New}(k)\}=\sigma^2\sum_{i=1}^{p}\frac{2\lambda_i(\lambda_i-n+1)(k+n-1)}{(\lambda_i+k)^3(n-1)^2}$$

$$\frac{d^2}{dk^2}\mathrm{tr}[\mathrm{v}\{\hat{\boldsymbol{\beta}}^{New}(k)\}]=\sigma^2\sum_{i=1}^{p}\frac{2\lambda_i(\lambda_i-n+1)(\lambda_i-2k-3n+3)}{(\lambda_i+k)^4(n-1)^2}$$

Lemma 4 (Derivatives for TMSE of new estimator)

$$\frac{d}{dk}\mathrm{TMSE}\{\hat{\boldsymbol{\beta}}^{\mathrm{New}}(k)\}=\sum_{i=1}^{p}\frac{2k\alpha_i^2\lambda_i(\lambda_i-n+1)^2}{(\lambda_i+k)^3(n-1)^2}+\sigma^2\sum_{i=1}^{p}\frac{2\lambda_i(\lambda_i-n+1)(k+n-1)}{(\lambda_i+k)^3(n-1)^2}$$

$$\frac{d^2}{dk^2}\mathrm{TMSE}\{\hat{\boldsymbol{\beta}}^{\mathrm{New}}(k)\}=\sum_{i=1}^{p}\frac{2\alpha_i^2\lambda_i(\lambda_i-2k)(\lambda_i-n+1)^2}{(\lambda_i+k)^4(n-1)^2}+\sigma^2\sum_{i=1}^{p}\frac{2\lambda_i(\lambda_i-n+1)(\lambda_i-2k-3n+3)}{(\lambda_i+k)^4(n-1)^2}$$

Theory – Existence theorem

- We have that

$$\begin{cases} \lim_{k \to 0^+} \dfrac{d}{dk} \mathrm{bias}\{\hat{\boldsymbol{\beta}}^{\mathrm{New}}(k)\}^{\mathrm{T}}\{\hat{\boldsymbol{\beta}}^{\mathrm{New}}(k)\} = 0 \\ \lim_{k \to 0^+} \dfrac{d}{dk} \mathrm{v}\{\hat{\boldsymbol{\beta}}^{\mathrm{New}}(k)\} < 0 \end{cases}$$

Note: $\hat{\boldsymbol{\beta}}^{\mathrm{Ridge}}(0) = \hat{\boldsymbol{\beta}}^{\mathrm{OLS}}$

⇒ Bias square is flat at $k = 0^+$

⇒ Total variance is decreasing at $k = 0^+$

Theory – Optimal value of shrinkage parameter

- Several algorithm estimate k
- With-intercept-type model seldom consider
- Often separate to: intercept term k_0 and other term k
- Numerical minimization

$$k^{\text{Ridge}} = \arg\min_{x \geq 0} \left[x^2 \left\{ \frac{\alpha_0^2}{(n+x)^2} + \sum_{i=1}^{p} \frac{\alpha_i^2}{(\lambda_i + x)^2} \right\} + \sigma^2 \left\{ \frac{n}{(n+x)^2} + \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i + x)^2} \right\} \right]$$

$$k^{\text{New}} = \arg\min_{x \geq 0} \left[\sum_{i=1}^{p} \frac{x^2 \alpha_i^2 (\lambda_i - n + 1)^2}{(\lambda_i + x)^2 (n-1)^2} + \sigma^2 \left\{ \frac{1}{n} + \sum_{i=1}^{p} \frac{\lambda_i (x + n - 1)^2}{(\lambda_i + x)^2 (n-1)^2} \right\} \right]$$

- In real data, we use OLS estimator to replace true value

Numerical analysis – Simulation design

- Four cases:

$n = 13,\ p = 4$

Case 1 $\boldsymbol{\beta} = (50, 1, 1, 1, 1)^{\mathrm{T}}$ and $\sigma^2 = 1$

Case 2 $\boldsymbol{\beta} = (50, 1, 1, 1, 1)^{\mathrm{T}}$ and $\sigma^2 = 2$

Case 3 $\boldsymbol{\beta} = (1, 1, 1, 1, 1)^{\mathrm{T}}$ and $\sigma^2 = 1$

Case 4 $\boldsymbol{\beta} = (1, 1, 1, 1, 1)^{\mathrm{T}}$ and $\sigma^2 = 2$

Numerical analysis – Simulation design

Generate $\mathbf{X}_p = (\mathbf{x}_1, \ldots, \mathbf{x}_p)$ by

$$\begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix} \sim N_2\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, 5\mathbf{I}_2 \right), \quad \begin{bmatrix} x_{i3} \\ x_{i4} \end{bmatrix} \sim \begin{bmatrix} -x_{i1} \\ -x_{i2} \end{bmatrix} + N_2\left(\begin{bmatrix} 50 \\ 50 \end{bmatrix}, 5\mathbf{I}_2 \right)$$

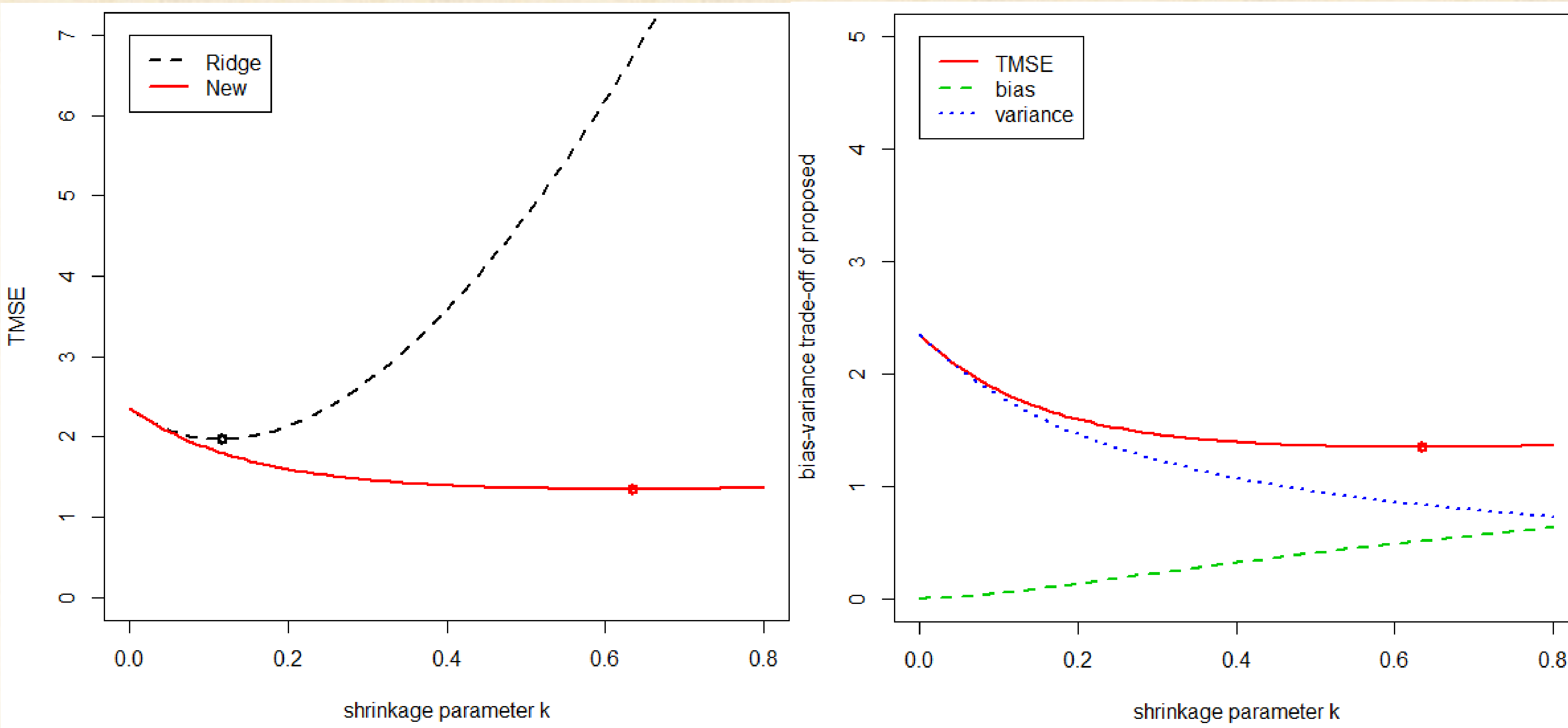
$$\text{for } i = 1, \cdots, n$$

- Sample correlation matrix of $\mathbf{X}_p$

$$\begin{array}{c c} & \begin{array}{cccc} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 \end{array} \\ \text{Sample Corr}(\mathbf{X}_p) = & \begin{bmatrix} 1.000 & -0.133 & -0.848 & 0.082 \\ -1.133 & 1.000 & 0.245 & -0.952 \\ \boxed{-0.848} & 0.245 & 1.000 & -0.139 \\ 0.082 & \boxed{-0.952} & -0.139 & 1.000 \end{bmatrix} \end{array}$$

Numerical analysis – Simulation result

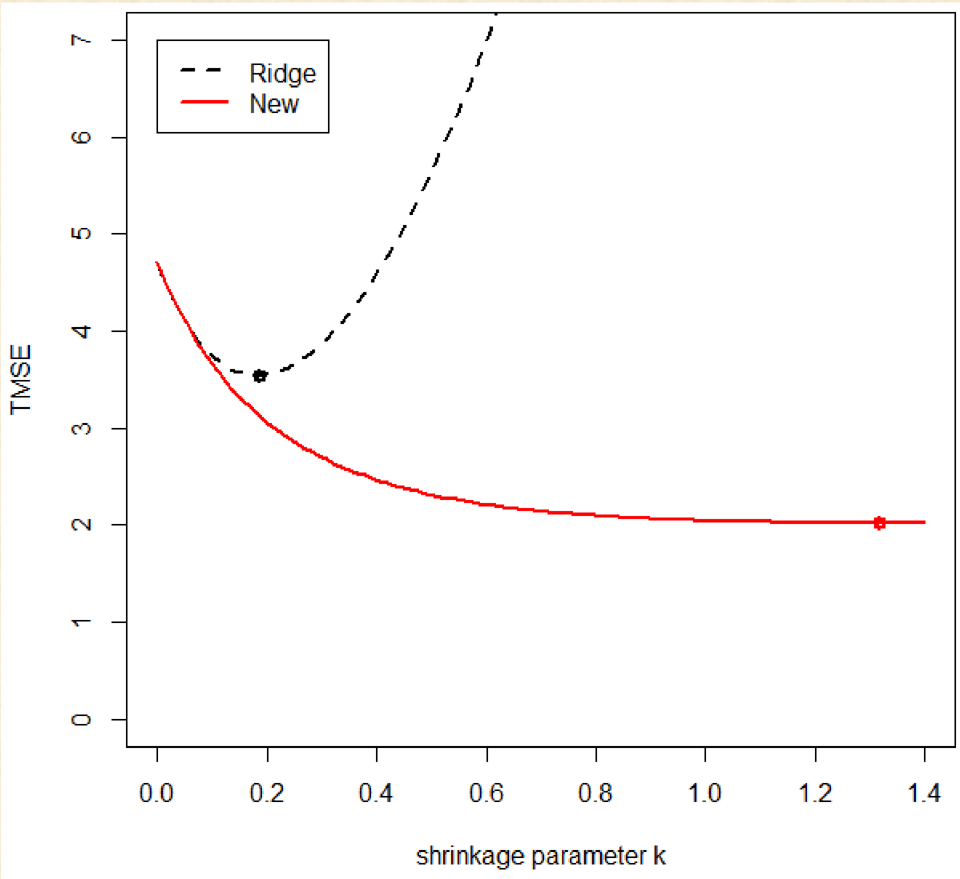
- TMSE and bias-variance trade off

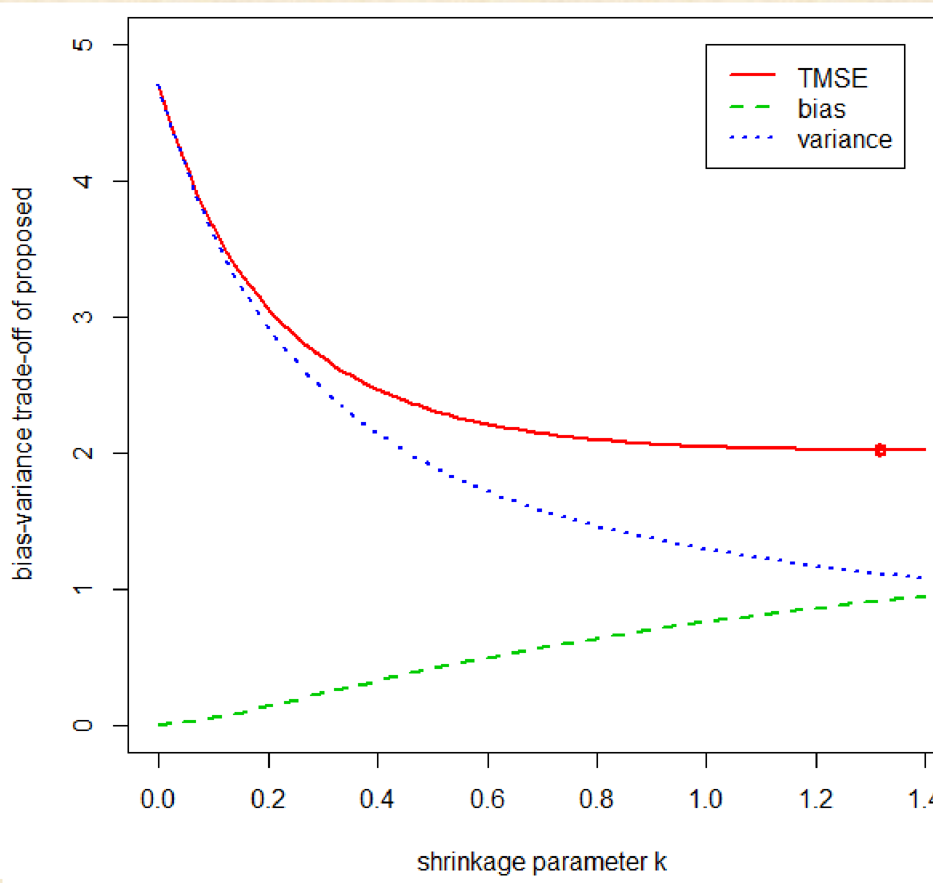


Case 1 $\boldsymbol{\beta} = (50, 1, 1, 1, 1)^T$ and $\sigma^2 = 1$

Numerical analysis – Simulation result

- TMSE and bias-variance trade off

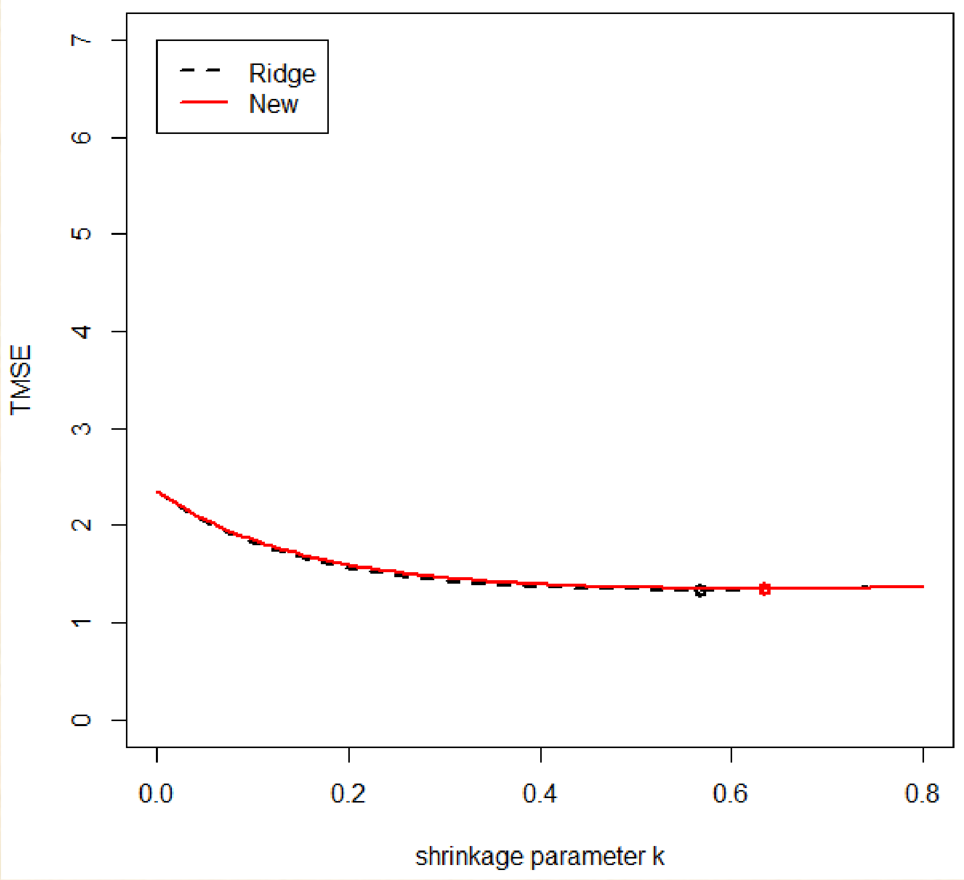


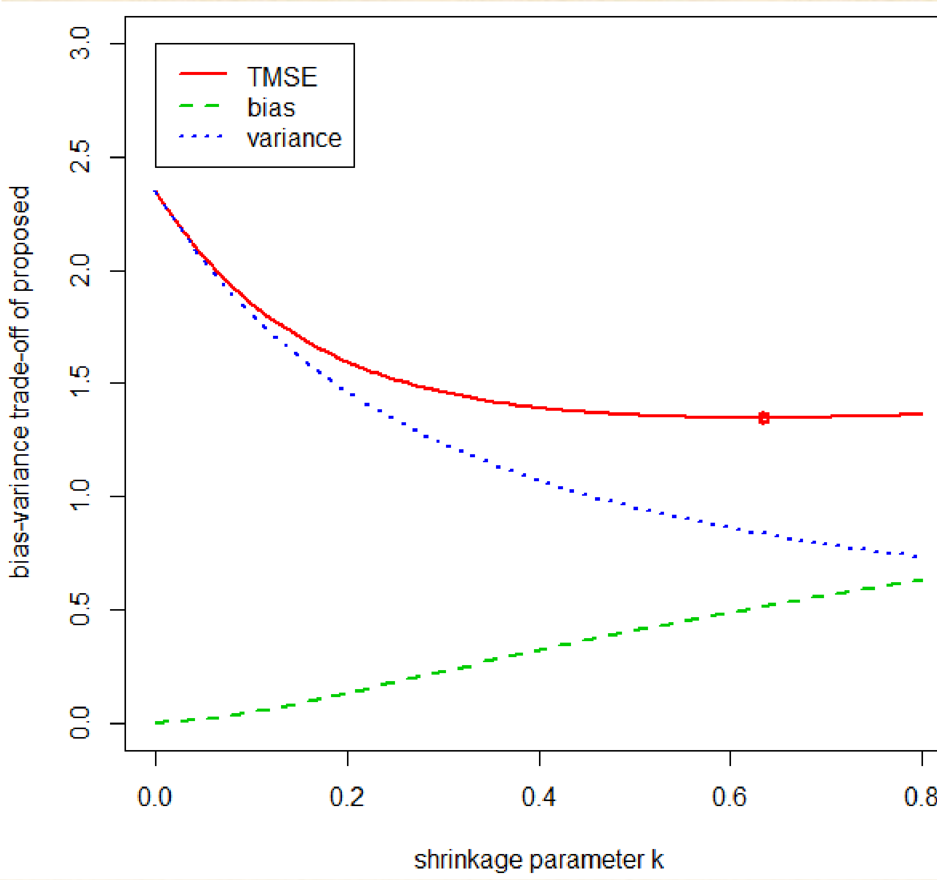


Case 2 $\quad \boldsymbol{\beta} = (50, 1, 1, 1, 1)^T$ and $\sigma^2 = 2$

Numerical analysis – Simulation result

- TMSE and bias-variance trade off

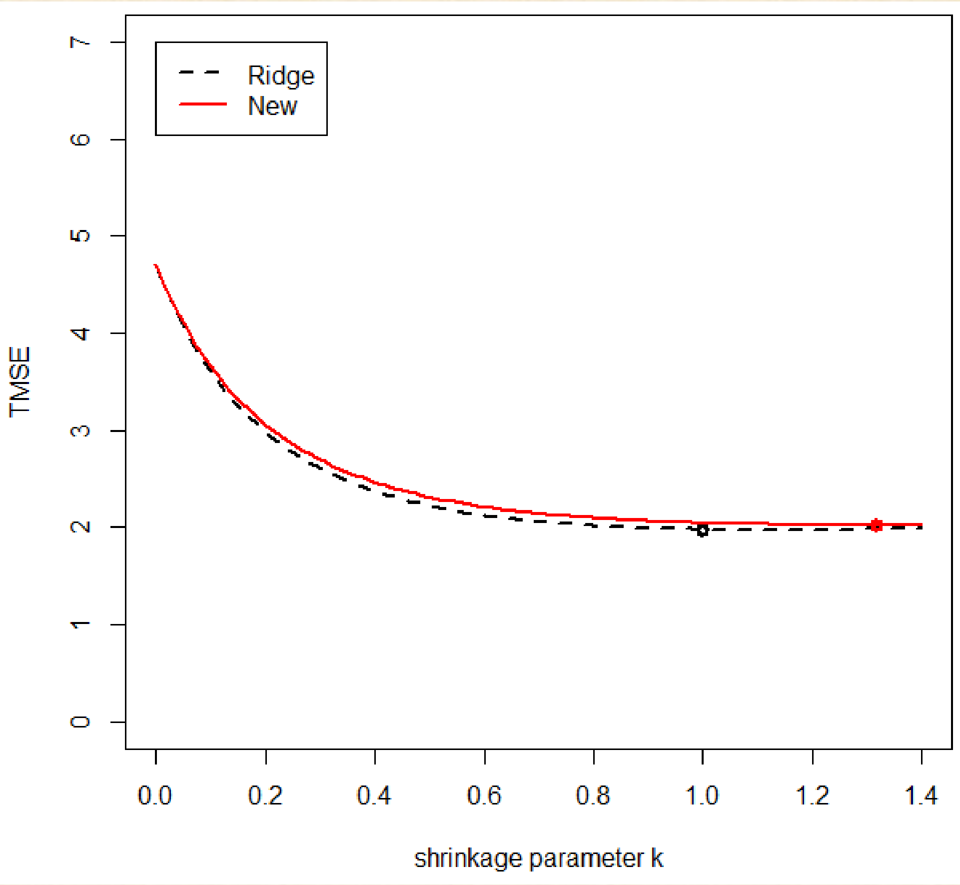


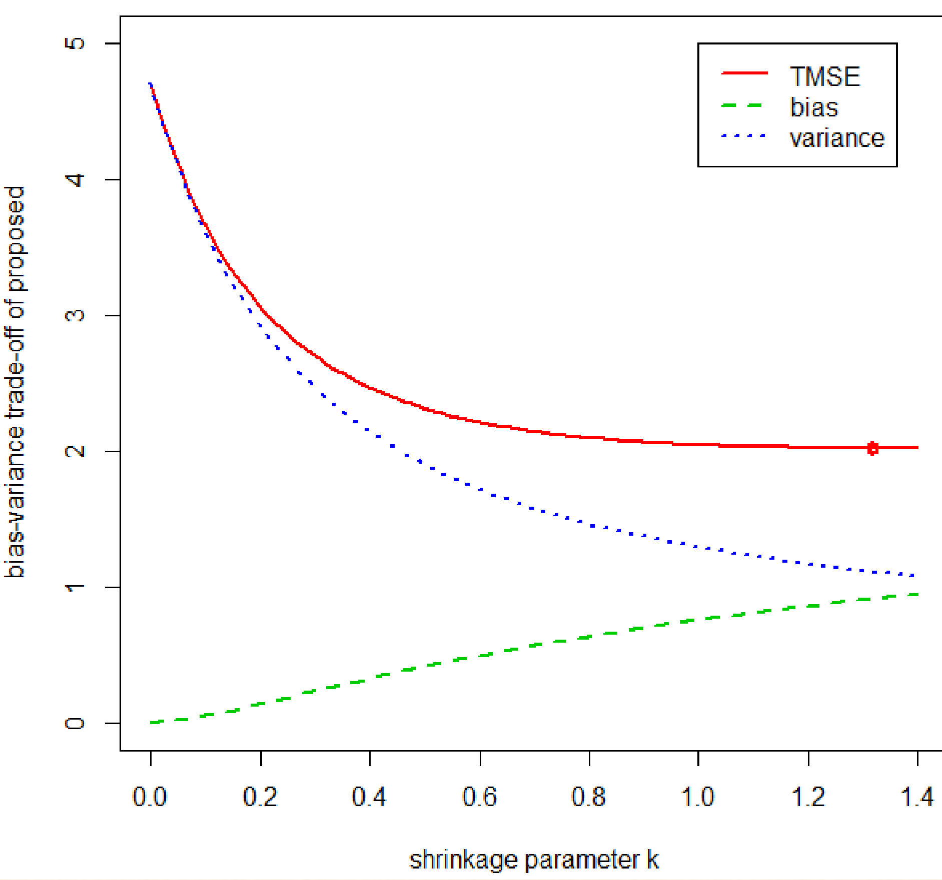


Case 3 $\boldsymbol{\beta} = (1,1,1,1,1)^T$ and $\sigma^2 = 1$

Numerical analysis – Simulation result

- TMSE and bias-variance trade off





Case 4 $\boldsymbol{\beta} = (1, 1, 1, 1, 1)^T$ and $\sigma^2 = 2$

Numerical analysis – Simulation result

- Effects of intercept term and σ^2 on ridge & new method

	Ridge	New method
Intercept term	affected	not affected
σ^2	affected	affected

The new method is more robust against the changing of intercept

Numerical analysis – Simulation result

- TMSE & Shrinkage parameter k estimation

	$\boldsymbol{\beta}=\begin{bmatrix}50\\1\\1\\1\\1\end{bmatrix}, \sigma^2=1$	$\boldsymbol{\beta}=\begin{bmatrix}50\\1\\1\\1\\1\end{bmatrix}, \sigma^2=2$	$\boldsymbol{\beta}=\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}, \sigma^2=1$	$\boldsymbol{\beta}=\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}, \sigma^2=2$
True k^{New}	0.6343	1.3171	0.6343	1.3171
True k^{Ridge}	0.1163	0.1850	0.5676	1.1234
$E(\hat{k}^{\text{New}})$	0.9144	1.4847	0.9144	1.4847
$E(\hat{k}^{\text{Ridge}})$	0.1031	0.1558	0.7143	1.3164
$\text{TMSE}\{\hat{\boldsymbol{\beta}}(\hat{k}^{\text{New}})\}$	1.3480	2.0239	1.3480	2.0239
$\text{TMSE}\{\hat{\boldsymbol{\beta}}(\hat{k}^{\text{Ridge}})\}$	1.9661	3.5442	1.3358	1.9745

Numerical analysis – Data analysis

- Portland cement data (Woods et al., 1932)

	$x1$	$x2$	$x3$	$x4$	y
1	7	26	6	60	78.5
2	1	29	15	52	74.3
3	11	56	8	20	104.3
4	11	31	8	47	87.6
5	7	52	6	33	95.9
6	11	55	9	22	109.2
7	3	71	17	6	102.7
8	1	31	22	44	72.5
9	2	54	18	22	93.1
10	21	47	4	26	115.9
11	1	40	23	34	83.8
12	11	66	9	12	113.3
13	10	68	8	12	109.4

y Heat evolved during cement hardening

$\mathbf{x}_1$ $3CaO \cdot Al_2O_3$

$\mathbf{x}_2$ $3CaO \cdot SiO_2$

$\mathbf{x}_3$ $4CaO \cdot Al_2O_3 \cdot Fe_2O_3$

$\mathbf{x}_4$ $2CaO \cdot SiO_2$

Numerical analysis – Data analysis

- Portland cement data (Woods et al., 1932)

	$x1$	$x2$	$x3$	$x4$	y
1	7	26	6	60	78.5
2	1	29	15	52	74.3
3	11	56	8	20	104.3
4	11	31	8	47	87.6
5	7	52	6	33	95.9
6	11	55	9	22	109.2
7	3	71	17	6	102.7
8	1	31	22	44	72.5
9	2	54	18	22	93.1
10	21	47	4	26	115.9
11	1	40	23	34	83.8
12	11	66	9	12	113.3
13	10	68	8	12	109.4

$\mathbf{y}$ Heat evolved during cement hardening

$\boxed{\mathbf{x}_1 \quad 3CaO \cdot Al_2O_3}$

$\mathbf{x}_2 \quad 3CaO \cdot SiO_2$

$\boxed{\mathbf{x}_3 \quad 4CaO \cdot Al_2O_3 \cdot Fe_2O_3}$

$\mathbf{x}_4 \quad 2CaO \cdot SiO_2$

Numerical analysis – Data analysis

- Sample correlation matrix

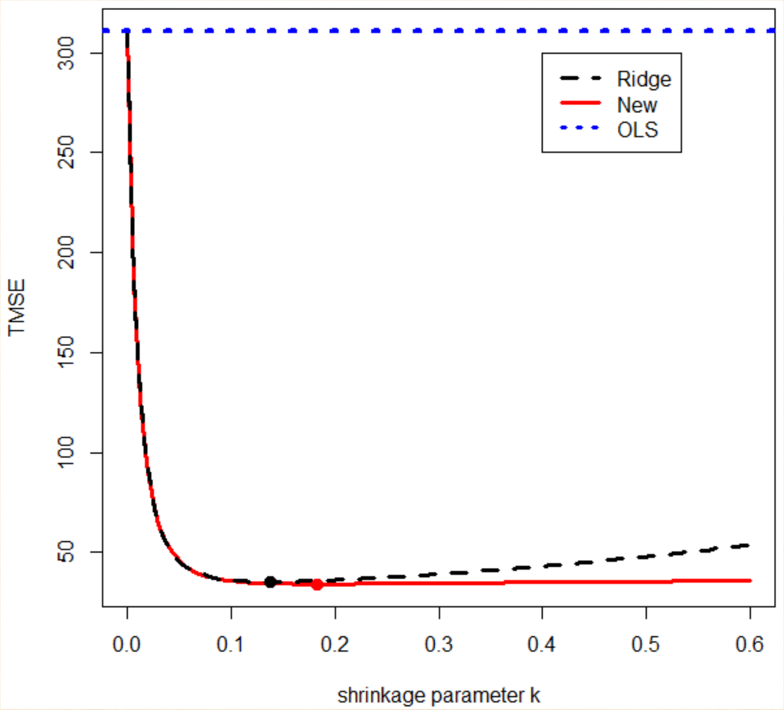
$$\text{Sample Corr}(\mathbf{X}_p) = \begin{array}{c} \begin{array}{cccc} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 \end{array} \\ \begin{bmatrix} 1.000 & 0.228 & -0.824 & -0.245 \\ 0.228 & 1.000 & -0.139 & -0.972 \\ -0.824 & -0.139 & 1.000 & 0.029 \\ -0.245 & -0.972 & 0.029 & 1.000 \end{bmatrix} \end{array}$$

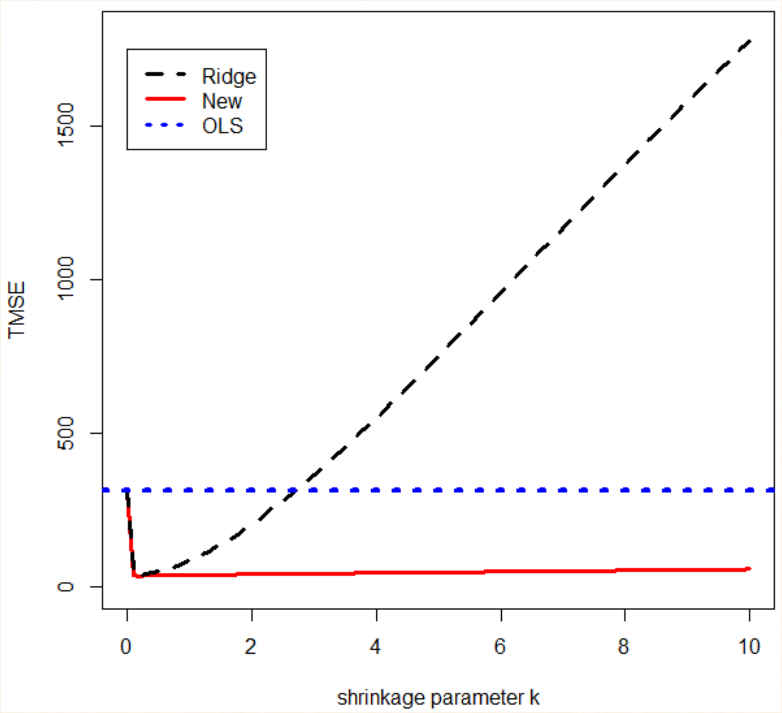
- Check eigenvalues in matrix $\mathbf{\Lambda}$

$$\mathbf{\Lambda} = \begin{bmatrix} 13 & 0 & 0 & 0 & 0 \\ 0 & 26.828 & 0 & 0 & 0 \\ 0 & 0 & 18.912 & 0 & 0 \\ 0 & 0 & 0 & 2.239 & 0 \\ 0 & 0 & 0 & 0 & 0.019 \end{bmatrix}$$

Numerical analysis – Data analysis

- TMSE on Portland cement data





	Minimun TMSE	$\hat{k}$
Ridge	35.2969	0.1372
New	34.1197	0.1826
OLS	310.7266	

Numerical analysis – Data analysis

- Case that do not standardize (Portland cement data)

$\hat{\boldsymbol\beta}$	β_0	β_1	β_2	β_3	β_4	Bias	Var	TMSE
$\hat{\boldsymbol\beta}^{\text{OLS}}$	62.41	1.55	0.51	0.10	–0.14	0	4912.09	4912.09
$\hat{\boldsymbol\beta}^{\text{Ridge}}(k), k=\hat{k}_{\text{HK}}$	27.63	1.91	0.87	0.47	0.21	1209.55	961.42	2170.55
$\hat{\boldsymbol\beta}^{\text{Ridge}}(k), k=\hat{k}^{\text{Ridge}}$	27.78	1.91	0.87	0.47	0.21	1199.62	971.40	2170.62
$\hat{\boldsymbol\beta}^{\text{Liu}}(d), d=\hat{d}_{\text{opt}}$	62.25	1.55	0.51	0.10	–0.14	0.02	4887.28	4887.30
$\hat{\boldsymbol\beta}^{\text{SK}}(k,d), k=\hat{k}_{\text{HK}}, d=\hat{d}_{\text{opt}}$	27.61	1.91	0.87	0.47	0.21	1211.46	959.50	2170.96
$\hat{\boldsymbol\beta}^{\text{New}}(k), k=\hat{k}_{\text{HK}}$	80.71	1.36	0.32	–0.09	–0.33	335.20	961.78	1296.20
$\hat{\boldsymbol\beta}^{\text{New}}(k), k=\hat{k}^{\text{New}}$	88.99	1.28	0.24	–0.18	–0.41	707.30	177.86	884.30

$$\hat{k}_{\text{HK}} = \hat{\sigma}^2 / \{ (\hat{\boldsymbol\beta}^{\text{OLS}})^{\text{T}} (\hat{\boldsymbol\beta}^{\text{OLS}}) \} = 0.00153522\,2$$ (Hoerl & Kennard, 1970)

$$\hat{k}^{\text{Ridge}} = 0.00152103\,3$$

$$\hat{k}^{\text{NEW}} = 0.00519210\,8$$

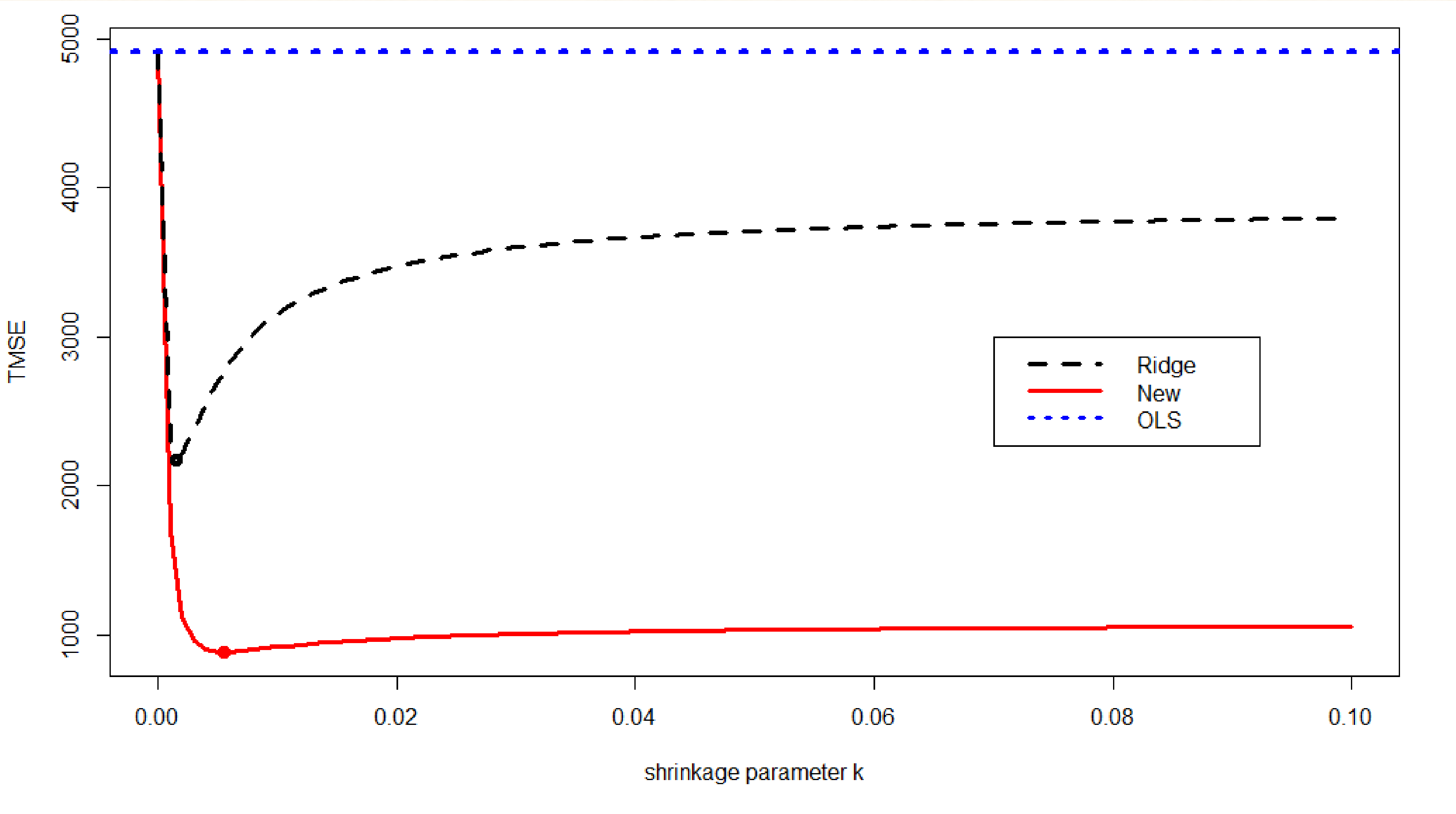
$$\hat{\sigma}^2 = 5.982955$$

$$\hat{d}_{\text{opt}} = \frac{\sum_{i=1}^{p+1} \frac{\lambda_i(\hat{\alpha}_i^2 - \hat{\sigma}^2)}{(\lambda_i+1)^2(\lambda_i+\hat{k}_{\text{HK}})}}{\sum_{i=1}^{p+1} \frac{\lambda_i(\lambda_i\hat{\alpha}_i^2 + \hat{\sigma}^2)}{(\lambda_i+1)^2(\lambda_i+\hat{k}_{\text{HK}})^2}} = 0.9974682$$

(Sakallıoğlu and Kaçıranlar, 2006)

Numerical analysis – Data analysis

- Case that do not standardize (Portland cement data)



Numerical analysis – Data analysis

- Flare data (McLean and Anderson, 1966)

	x_1	x_2	x_3	x_4	y
1	0.4	0.1	0.47	0.03	75
2	0.4	0.1	0.42	0.08	180
3	0.6	0.1	0.27	0.03	195
4	0.6	0.1	0.22	0.08	300
5	0.4	0.47	0.1	0.03	145
6	0.4	0.42	0.1	0.08	230
7	0.6	0.27	0.1	0.03	220
8	0.6	0.22	0.1	0.08	350
9	0.5	0.1	0.345	0.055	220
10	0.5	0.345	0.1	0.055	260
11	0.4	0.2725	0.2725	0.055	190
12	0.6	0.1725	0.1725	0.055	310
13	0.5	0.235	0.235	0.03	260
14	0.5	0.21	0.21	0.08	410
15	0.5	0.2225	0.2225	0.055	425

y : Amount of illumination (1,000 candles)

$\mathbf{x}_1$: Magnesium

$\mathbf{x}_2$: Sodium nitrate

$\mathbf{x}_3$: Strontium nitrate

$\mathbf{x}_4$: Binder

- Constraints

$$0.4 \le x_1 \le 0.6$$

$$0.1 \le x_2 \le 0.5$$

$$0.1 \le x_3 \le 0.5$$

$$0.03 \le x_4 \le 0.08.$$

Numerical analysis – Data analysis

- Under standardization (flare data)

$\hat{\boldsymbol{\beta}}$	β_0	β_1	β_2	β_3	β_4
$\hat{\boldsymbol{\beta}}^{\text{OLS}}$	Does not exist				
$\hat{\boldsymbol{\beta}}^{*}$	251.33	46.91	–2.38	–39.60	48.60
$\hat{\boldsymbol{\beta}}^{\text{Ridge}}(k), k=\hat{k}^{*}$	249.78	34.97	–5.75	–27.17	45.37
$\hat{\boldsymbol{\beta}}^{\text{Liu}}(d), d=\hat{d}_{\text{opt}}$	Does not exist				
$\hat{\boldsymbol{\beta}}^{\text{SK}}(k,d), k=\hat{k}^{*}, d=\hat{d}_{\text{opt}}$	240.59	33.94	–5.49	–26.40	43.67
$\hat{\boldsymbol{\beta}}^{\text{New}}(k), k=\hat{k}^{*}$	251.33	35.20	–5.79	–27.35	45.67

$$\hat{k}^{*}=\hat{\sigma}^{*2}/\{(\hat{\boldsymbol{\beta}}^{*})^{\text{T}}(\hat{\boldsymbol{\beta}}^{*})\}$$

$$=0.09338123$$

$$\hat{d}_{\text{opt}}=0.317959$$

- Un standardization (flare data)

$\hat{\boldsymbol{\beta}}$	β_0	β_1	β_2	β_3	β_4
$\hat{\boldsymbol{\beta}}^{\text{OLS}}$	Does not exist				
$\hat{\boldsymbol{\beta}}^{*}$	251.33	504.03	887.71	821.84	4294.92
$\hat{\boldsymbol{\beta}}^{\text{Ridge}}(k), k=\hat{k}^{*}$	171.66	280.23	–86.98	–242.12	220.53
$\hat{\boldsymbol{\beta}}^{\text{Liu}}(d), d=\hat{d}_{\text{opt}}$	Does not exist				
$\hat{\boldsymbol{\beta}}^{\text{SK}}(k,d), k=\hat{k}^{*}, d=\hat{d}_{\text{opt}}$	170.29	123.74	35.95	–6.62	17.21
$\hat{\boldsymbol{\beta}}^{\text{New}}(k), k=\hat{k}^{*}$	–525.89	693.64	548.72	384.24	4104.65

$$\hat{k}^{*}=0.05669671$$

$$\hat{d}_{\text{opt}}=-0.03083932$$

Conclusion

- We works on model with intercept.
- Achieves the smallest TMSE among OLS and ridge regression in some case, especially large intercept cases.
- Proposed method works on unstandardized model.

Future work

- Scheffé type model in mixture experiment
- From mixture experiments to other experiment design

THE END
THANK YOU