

ON SHRINKAGE AND SELECTION: ANOVA MODEL

A. K. MD. EHSANES SALEH*

School of Mathematics and Statistics, Carleton University, Ottawa, Canada
Email: esaleh@math.carleton.ca

M. ARASHI

Department of Statistics, School of Mathematical Sciences
Shahrood University of Technology, Shahrood, Iran
Email: m_arashi_stat@yahoo.com

M. NOROUZIRAD

Department of Statistics, School of Mathematical Sciences
Shahrood University of Technology, Shahrood, Iran
Email: mina.norouzirad@gmail.com

B M GOALM KIBRIA

Department of Mathematics and Statistics
Florida International University, Miami, FL, USA
Email: kibriag@fiu.edu

SUMMARY

This paper considers the estimation of the parameters of an ANOVA model when sparsity is suspected. Accordingly, we consider the least square estimator (LSE), restricted LSE, preliminary test and Stein-type estimators, together with three penalty estimators, namely, the ridge estimator, subset selection rules (hard threshold estimator) and the LASSO (soft threshold estimator). We compare and contrast the L_2 -risk of all the estimators with the lower bound of L_2 -risk of LASSO in a family of diagonal projection scheme which is also the lower bound of the exact L_2 -risk of LASSO. The result of this comparison is that neither LASSO nor the LSE, preliminary test, and Stein-type estimators outperform each other uniformly. However, when the model is sparse, LASSO outperforms all estimators except "ridge" estimator since both LASSO and ridge are L_2 -risk equivalent under sparsity. We also find that LASSO and the restricted LSE are L_2 -risk equivalent and both outperform all estimators (except ridge) depending on the dimension of sparsity. Finally, ridge estimator outperforms all estimators uniformly. Our findings are based on L_2 -risk of estimators and lower bound of the risk of LASSO together with tables of efficiency and graphical display of efficiency and not based on simulation.

Keywords and phrases: ANOVA model, Dominance, Efficiency, LASSO, PTE and Stein-type estimators, Penalty estimator, Risk function, Sparsity

AMS Classification: 62F15 (primary), 62H05 (secondary)

* Corresponding author

1 Introduction

An important model belonging to the class of general linear hypothesis is the analysis of variance (ANOVA) model. In this model, we consider the assessment of p treatment effects by considering samples experiments of sizes n_1, n_2, \dots, n_p respectively with the responses $\{(y_{i1}, \dots, y_{in_i})^T; i = 1, 2, \dots, p\}$ which satisfy the model, $y_{ij} = \theta_i + e_{ij}$ ($j = 1, 2, \dots, n_i$, $i = 1, 2, \dots, p$). The main objective of the paper is the selection of the treatments which would yield best results. Accordingly, we consider the penalty estimators, namely, ridge, subset selection rule and LASSO together with the classical shrinkage estimators, namely, the preliminary test estimator and the Stein-type estimators such as James-Stein estimator (JSE) and positive-rule Stein-type estimator (PRSE) of $\theta = (\theta_1, \dots, \theta_p)^T$. For LASSO and the related, see Breiman (1995), Fan and Li (2001), Zou and Hastie (2005), Zou (2006) among others and for preliminary test and Stein-type estimators, see Saleh (2006).

The paper points to the useful “selection” aspect of LASSO and ridge estimators as well as limitation found in other papers. Our conclusions are based on the ideal L_2 -risk of LASSO of an oracle which would supply with an optimal coefficients in a diagonal projection scheme given by Donoho and Johnstone (1994, Pg. 437). The comparison of the estimators considered here are based on mathematical analysis as well as by tables of L_2 -risk efficiencies and graphs and not by simulation.

In his pioneering paper, Tibshirani (1996) examined the relative performance of the subset selection, ridge regression and LASSO in three different scenarios, under orthogonal design matrix in a linear regression model: (a) Small number of large coefficients - subset selection does the best here, the LASSO not quite as well, ridge regression does quite poorly, (b) Small to moderate numbers moderate-size coefficients - LASSO does the best followed by ridge regression and then subset selection, and (c) Large number of small coefficients - ridge regression does best by a good margin, followed by LASSO and then subset selection. These results refer to *prediction accuracy*.

Recently, Hansen (2016) considered the comparison of LASSO, Stein-type estimators and subset selection based on the upper bounds of L_2 -risk under infeasible condition: all parameters may be zero. His findings may be summarized as follows: (i) The condition above yielded an infeasible estimate equal to $\mathbf{0}$ -vector with the L_2 -risk equal to the divergence parameter, Δ^2 . This led to doubt the “oracle properties” of LASSO. (ii) Neither LASSO nor Stein-type estimators uniformly dominate one other, (iii) Via simulation studies, he concludes that LASSO estimation is particularly sensitive to coefficient parametrization and for a significant portion of the parameter space, LASSO has higher L_2 -risk than the LSE. He did not specify the regions where one estimator or the other has lower L_2 -risk. In his analysis, he used the normalized L_2 -risk bounds (NRB) to arrive at his conclusion with a total sparse model.

In our study, we discovered the following conclusions:

- (i) The ridge estimator outperforms the LSE, preliminary test, and Stein-type estimators (JSE and PRSE) uniformly. The ridge dominates LASSO and restricted LSE

uniformly for $\Delta^2 > 0$ and at $\Delta^2 = 0$, they are L_2 -risk equivalent where Δ^2 is the divergence parameter.

- (ii) The restricted LSE (RLSE) and LASSO are L_2 -risk equivalent and satisfy “oracle properties”.
- (iii) Under the family of “Diagonal Linear Projection”, the “Ideal” L_2 - risk of LASSO and subset rule (hard threshold estimator) are same and do not depend on the threshold parameter (κ) under sparse condition.
- (iv) Performance of estimators compared to the LSE depends on the size p_1, p_2 , and the divergence parameter, Δ^2 . RLSE, HTE and LASSO perform best compared to LSE on the interval $[0, p_2]$. On the other hand, LASSO approximately outperforms Stein-type estimator inside the interval $[0, p_1]$. LASSO is worse off beyond the interval, $[0, p_1]$.

The organization of the paper is as follows: Section 2 discusses various estimators and L_2 -risk expressions using ANOVA model. Section 3 discusses the bias and L_2 -risks of the estimators, multivariate normal decision theory and oracles for diagonal variant linear projection. Section 4 deals with the comparisons of the estimators. To illustrate the methodologies of this paper, a numerical data analysis is given in section 5. Section 6 presents summary of our results along with superiority tables of efficiency and graphs related to the comparisons.

2 Model, Estimation and Tests

Consider the ANOVA model

$$\mathbf{Y} = \mathbf{B}\boldsymbol{\theta} + \boldsymbol{\epsilon} = \mathbf{B}_1\boldsymbol{\theta}_1 + \mathbf{B}_2\boldsymbol{\theta}_2 + \boldsymbol{\epsilon}, \quad (2.1)$$

where $\mathbf{Y} = (y_{11}, \dots, y_{1n_1}, \dots, y_{p_1}, \dots, y_{pn_p})^\top$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{p_1}, \theta_{p_1+1}, \dots, \theta_p)^\top$ is the unknown vector can be partitioned as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ where $\boldsymbol{\theta}_1 = (\theta_1, \dots, \theta_{p_1})^\top$ and $\boldsymbol{\theta}_2 = (\theta_{p_1+1}, \dots, \theta_p)^\top$. The error vector $\boldsymbol{\epsilon}$ is $(\epsilon_{11}, \dots, \epsilon_{1n_1}, \dots, \epsilon_{p_1}, \dots, \epsilon_{pn_p})^\top$ with $\boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. The notation \mathbf{B} stands for a block-diagonal vector of $(\mathbf{1}_{n_1}, \dots, \mathbf{1}_{n_p})$ which can subdivide into two matrices \mathbf{B}_1 and \mathbf{B}_2 as $(\mathbf{B}_1, \mathbf{B}_2)$ where $\mathbf{1}_{n_i} = (1, \dots, 1)^\top$ is an n_i tuples of 1s, \mathbf{I}_n is the n -dimensional identity matrix, where $n = n_1 + \dots + n_p$ and σ^2 is the known variance of the errors.

Our objective is to estimate and select the treatments $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$ when we suspect that the subset $\boldsymbol{\theta}_2 = (\theta_{p_1+1}, \dots, \theta_p)^\top$ may be $\mathbf{0}$, i.e. ineffective. Thus, we consider the model (2.1) and discuss the LSE of $\boldsymbol{\theta}$ in the Subsection 2.1.

2.1 Estimation of Treatments Effects

First, we consider the unrestricted LSE of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ given by

$$\begin{aligned}\tilde{\boldsymbol{\theta}}_n &= \operatorname{argmin}_{\boldsymbol{\theta}} \{(\mathbf{Y} - \mathbf{B}_1\boldsymbol{\theta}_1 - \mathbf{B}_2\boldsymbol{\theta}_2)^\top(\mathbf{Y} - \mathbf{B}_1\boldsymbol{\theta}_1 - \mathbf{B}_2\boldsymbol{\theta}_2)\} \\ &= \begin{pmatrix} \mathbf{B}_1^\top\mathbf{B}_1 & \mathbf{B}_1^\top\mathbf{B}_2 \\ \mathbf{B}_2^\top\mathbf{B}_1 & \mathbf{B}_2^\top\mathbf{B}_2 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{B}_1^\top\mathbf{Y} \\ \mathbf{B}_2^\top\mathbf{Y} \end{pmatrix} = \begin{pmatrix} \mathbf{N}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_2 \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{B}_1^\top\mathbf{Y} \\ \mathbf{B}_2^\top\mathbf{Y} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{N}_1^{-1}\mathbf{B}_1^\top\mathbf{Y} \\ \mathbf{N}_2^{-1}\mathbf{B}_2^\top\mathbf{Y} \end{pmatrix} = \begin{pmatrix} \tilde{\boldsymbol{\theta}}_{1n} \\ \tilde{\boldsymbol{\theta}}_{2n} \end{pmatrix},\end{aligned}\quad (2.2)$$

where $\mathbf{N} = \mathbf{B}^\top\mathbf{B} = \operatorname{diag}(n_1, \dots, n_p)$, $\mathbf{N}_1 = \operatorname{diag}(n_1, \dots, n_{p_1})$, and $\mathbf{N}_2 = \operatorname{diag}(n_{p_1+1}, \dots, n_p)$.

In case that σ^2 is unknown, the BLUE of σ^2 is given by

$$s_n^2 = (n - p)^{-1}(\mathbf{Y} - \mathbf{B}_1\tilde{\boldsymbol{\theta}}_{1n} - \mathbf{B}_2\tilde{\boldsymbol{\theta}}_{2n})^\top(\mathbf{Y} - \mathbf{B}_1\tilde{\boldsymbol{\theta}}_{1n} - \mathbf{B}_2\tilde{\boldsymbol{\theta}}_{2n}).$$

Clearly, $\tilde{\boldsymbol{\theta}}_n \sim \mathcal{N}_p(\boldsymbol{\theta}, \sigma^2\mathbf{N}^{-1})$ is independent of ms_n^2/σ^2 ($m = n - p$), which follows a central χ^2 -distribution with m degrees of freedom (d.f.).

When $\boldsymbol{\theta}_2 = \mathbf{0}$, then the restricted LSE (RLSE) of $\boldsymbol{\theta}_R = (\boldsymbol{\theta}_1^\top, \mathbf{0}^\top)^\top$ is given by $\hat{\boldsymbol{\theta}}_R = (\tilde{\boldsymbol{\theta}}_{1n}^\top, \mathbf{0}^\top)^\top$, where $\tilde{\boldsymbol{\theta}}_{1n} = \mathbf{N}_1^{-1}\mathbf{B}_1^\top\mathbf{Y}$.

2.2 Test of Significance

For the test of $\mathcal{H}_o : \boldsymbol{\theta}_2 = \mathbf{0}$ versus $\mathcal{H}_A : \boldsymbol{\theta}_2 \neq \mathbf{0}$, we consider the statistic \mathcal{L}_n given by

$$\mathcal{L}_n = \frac{1}{\sigma^2} \tilde{\boldsymbol{\theta}}_{2n}^\top \mathbf{N}_2 \tilde{\boldsymbol{\theta}}_{2n}, \quad \text{if } \sigma^2 \text{ is known} \quad (2.3a)$$

$$= \frac{1}{p_2 s_n^2} \tilde{\boldsymbol{\theta}}_{2n}^\top \mathbf{N}_2 \tilde{\boldsymbol{\theta}}_{2n}, \quad \text{if } \sigma^2 \text{ is unknown} \quad (2.3b)$$

Under a null-hypothesis \mathcal{H}_o , the null-distribution of \mathcal{L}_n is the central χ^2 -distribution with p_2 d.f. when σ^2 is known and the central F -distribution with (p_2, m) d.f. in the case of σ^2 is unknown, respectively. Under the alternative hypothesis, \mathcal{H}_A , the test statistics \mathcal{L}_n follows the non-central version of the mentioned densities. In both cases, the non-centrality parameter is $\Delta^2 = \boldsymbol{\theta}_2^\top \mathbf{N}_2 \boldsymbol{\theta}_2 / \sigma^2$. In this paper, we always assume that σ^2 is known, then \mathcal{L}_n follows a chi square distribution with p_2 d.f.

Further, we note that

$$\tilde{\theta}_{jn} \sim \mathcal{N}(\theta_j, \sigma^2 n_j^{-1}), \quad j = 1, \dots, p$$

so that $\mathcal{Z}_j = \sqrt{n_j} \tilde{\theta}_{jn} / \sigma \sim \mathcal{N}(\Delta_j, 1)$, where $\Delta_j = \sqrt{n_j} \theta_j / \sigma$. Thus, one may use \mathcal{Z}_j to test the null-hypothesis $\mathcal{H}_o^{(j)} : \theta_j = 0$ versus $\mathcal{H}_A^{(j)} : \theta_j \neq 0$, $j = p_1 + 1, \dots, p$.

In this paper, we are interested in studying of three penalty estimators, namely, (i) the subset rule called ‘‘hard threshold estimator’’ (HTE), (ii) LASSO or the ‘‘soft threshold

estimator” (STE) and (iii) the “ridge regression estimator” (RRE) together with the classical preliminary test estimator (PTE) and Stein-type estimators such as “James-Stein estimator” (JSE) and “positive-rule Stein-type estimator” (PRSE).

2.3 Penalty Estimators

In this section, we shall discuss the penalty estimators. Define the HTE as

$$\begin{aligned}\hat{\boldsymbol{\theta}}_n^{HT}(\kappa) &= \left(\tilde{\theta}_{jn}I(|\tilde{\theta}_{jn}| > \kappa\sigma n_j^{-\frac{1}{2}}) | j = 1, \dots, p\right)^\top \\ &= \left(\sigma n_j^{-\frac{1}{2}} \mathcal{Z}_j I(|\mathcal{Z}_j| > \kappa) | j = 1, \dots, p\right)^\top,\end{aligned}$$

where κ is a positive threshold parameter.

This estimator is discrete in nature and may be extremely variables and unstable due to the fact that small change in the data can result in a very different models and can reduce the prediction accuracy. As such we obtain the continuous version of $\hat{\boldsymbol{\theta}}_n^{HT}(\kappa)$, the LASSO is defined by

$$\hat{\boldsymbol{\theta}}_n^L(\lambda) = \arg \min_{\boldsymbol{\theta}} (\mathbf{Y} - \mathbf{B}\boldsymbol{\theta})^\top (\mathbf{Y} - \mathbf{B}\boldsymbol{\theta}) + 2\lambda\sigma \sum_{j=1}^p \sqrt{n_j}\kappa|\theta_j|,$$

where $|\boldsymbol{\theta}| = (|\theta_1|, \dots, |\theta_p|)^\top$, yielding the equation

$$\mathbf{B}^\top \mathbf{B}\boldsymbol{\theta} - \mathbf{B}^\top \mathbf{Y} + \lambda\sigma \mathbf{N}^{\frac{1}{2}} \text{sgn}(\boldsymbol{\theta}) = \mathbf{0}$$

or

$$\hat{\boldsymbol{\theta}}_n^L(\lambda) - \tilde{\boldsymbol{\theta}}_n + \frac{1}{2}\lambda\sigma \mathbf{N}^{-\frac{1}{2}} \text{sgn}(\hat{\boldsymbol{\theta}}_n^L(\lambda)) = \mathbf{0} \tag{2.4}$$

Now, the j th component of (2.4) is given by

$$\hat{\theta}_{jn}^L(\lambda) - \tilde{\theta}_{jn} + \lambda\sigma n_j^{-\frac{1}{2}} \text{sgn}(\hat{\theta}_{jn}^L(\lambda)) = 0. \tag{2.5}$$

Then, we consider three cases:

(i) $\text{sgn}(\hat{\theta}_{jn}^L(\lambda)) = +1$, then, (2.5) reduces to

$$0 < \frac{\hat{\theta}_{jn}^L(\lambda)}{\sigma n_j^{-\frac{1}{2}}} - \frac{\tilde{\theta}_{jn}}{\sigma n_j^{-\frac{1}{2}}} + \lambda = 0$$

Hence,

$$0 < \hat{\theta}_{jn}^L(\lambda) = \sigma n_j^{-\frac{1}{2}} (\mathcal{Z}_j - \lambda) = \sigma n_j^{-\frac{1}{2}} (|\mathcal{Z}_j| - \lambda) \tag{2.6}$$

with clearly $\mathcal{Z}_j > 0$ and $|\mathcal{Z}_j| > \lambda$.

(ii) $\text{sgn}(\hat{\theta}_{jn}^L(\lambda)) = -1$, then we have

$$0 > \frac{\hat{\theta}_{jn}^L(\lambda)}{\sigma n_j^{-\frac{1}{2}}} = \mathcal{Z}_j + \lambda = -(|\mathcal{Z}_j| - \lambda) \quad (2.7)$$

with clearly $\mathcal{Z}_j < 0$ and $|\mathcal{Z}_j| > \lambda$, and

(iii) For $\hat{\theta}_{jn}^L(\lambda) = 0$, we have $-\mathcal{Z}_j + \lambda\gamma = 0$ for some $\gamma \in (-1, 1)$. Hence, we obtain $\mathcal{Z}_j = \lambda\gamma$ which implies $|\mathcal{Z}_j| < \lambda$.

Combining (2.6)-(2.7) and (iii), we obtain

$$\hat{\theta}_{jn}^L(\lambda) = \sigma n_j^{-\frac{1}{2}} \text{sgn}(\mathcal{Z}_j)(|\mathcal{Z}_j| - \lambda)^+, \quad j = 1, \dots, p$$

where $a^+ = \max(0, a)$. Hence, the LASSO is given by

$$\hat{\boldsymbol{\theta}}_n^L(\lambda) = \left(\sigma n_j^{-\frac{1}{2}} \text{sgn}(\mathcal{Z}_j)(|\mathcal{Z}_j| - \lambda)^+ | j = 1, \dots, p \right)^\top.$$

Next, we consider the ridge regression estimator given by

$$\hat{\boldsymbol{\theta}}_n^{RR}(k) = \begin{pmatrix} \tilde{\boldsymbol{\theta}}_{1n} \\ \frac{1}{1+k} \tilde{\boldsymbol{\theta}}_{2n} \end{pmatrix}, \quad \kappa \in \mathbb{R}^+ \quad (2.8)$$

to accommodate sparse condition.

We may obtain $\tilde{\boldsymbol{\theta}}_n^{RR}(k)$ equal to $\tilde{\boldsymbol{\theta}}_{1n}$ when $\boldsymbol{\theta}_2 = \mathbf{0}$ and $\tilde{\boldsymbol{\theta}}_{2n}(k) = \frac{1}{1+k} \tilde{\boldsymbol{\theta}}_{2n}$ by minimizing the objective function,

$$(\mathbf{Y} - \mathbf{B}_1 \tilde{\boldsymbol{\theta}}_{1n} - \mathbf{B}_2 \boldsymbol{\theta}_2)^\top (\mathbf{Y} - \mathbf{B}_1 \tilde{\boldsymbol{\theta}}_{1n} - \mathbf{B}_2 \boldsymbol{\theta}_2) + k \boldsymbol{\theta}_2^\top \mathbf{N}_2 \boldsymbol{\theta}_2$$

with respect to $\boldsymbol{\theta}_2$. Thus, Eq. (2.8) is “feasible estimator of $\boldsymbol{\theta}$ ” when $\boldsymbol{\theta}_2$ consists of small sized parameters.

2.4 Preliminary Test and Stein-type Estimators

We recall that the unrestricted estimator of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ is given by $(\tilde{\boldsymbol{\theta}}_{1n}^\top, \tilde{\boldsymbol{\theta}}_{2n}^\top)^\top$ with marginal distribution $\tilde{\boldsymbol{\theta}}_{1n} \sim \mathcal{N}_{p_1}(\boldsymbol{\theta}_1, \sigma^2 \mathbf{N}_1^{-1})$ and $\tilde{\boldsymbol{\theta}}_{2n} \sim \mathcal{N}_{p_2}(\boldsymbol{\theta}_2, \sigma^2 \mathbf{N}_2^{-1})$, respectively. The restricted estimator of $(\boldsymbol{\theta}_1^\top, \mathbf{0}^\top)^\top$ is $(\tilde{\boldsymbol{\theta}}_{1n}^\top, \mathbf{0}^\top)^\top$. Similarly, the preliminary test estimator of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}}_n^{PT}(\alpha) = \begin{pmatrix} \tilde{\boldsymbol{\theta}}_{1n} \\ \tilde{\boldsymbol{\theta}}_{2n} I(\mathcal{L}_n > c_\alpha) \end{pmatrix},$$

where $I(A)$ is the indicator function of the set A , \mathcal{L}_n is the test statistic given in section 2.2 and c_α is the α -level critical value.

Similarly, the James-Stein estimator (JSE) is given by

$$\hat{\boldsymbol{\theta}}_n^{JS} = \begin{pmatrix} \tilde{\boldsymbol{\theta}}_{1n} \\ \tilde{\boldsymbol{\theta}}_{2n}(1 - (p_2 - 2)\mathcal{L}_n^{-1}) \end{pmatrix}, \quad p_2 \geq 3$$

and the positive-rule Stein-type estimator (PRSE) is given by

$$\hat{\boldsymbol{\theta}}_n^{S+} = \begin{pmatrix} \tilde{\boldsymbol{\theta}}_{1n} \\ \hat{\boldsymbol{\theta}}_{2n}^{JS} I(\mathcal{L}_n > p_2 - 2) \end{pmatrix}.$$

3 Bias and Weighted L_2 -risks of Estimators

This section contains the bias and the weighted L_2 -risk expressions of the estimators. We study the comparative performance of the seven estimators defined above based on the weighted L_2 -risks defined by

$$R(\boldsymbol{\theta}_n^* : \mathbf{W}_1, \mathbf{W}_2) = \mathbb{E}[(\boldsymbol{\theta}_{1n}^* - \boldsymbol{\theta}_1)^\top \mathbf{W}_1 (\boldsymbol{\theta}_{1n}^* - \boldsymbol{\theta}_1)] + \mathbb{E}[(\boldsymbol{\theta}_{2n}^* - \boldsymbol{\theta}_2)^\top \mathbf{W}_2 (\boldsymbol{\theta}_{2n}^* - \boldsymbol{\theta}_2)], \quad (3.1)$$

where $\boldsymbol{\theta}_n^* = (\boldsymbol{\theta}_{1n}^{*\top}, \boldsymbol{\theta}_{2n}^{*\top})^\top$ is any estimator of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, and \mathbf{W}_1 and \mathbf{W}_2 are weight matrices. For convenience, when $\mathbf{W}_1 = \mathbf{I}_{p_1}$ and $\mathbf{W}_2 = \mathbf{I}_{p_2}$, we get the mean square error (MSE) and write $R(\boldsymbol{\theta}_n^* : \mathbf{I}_p) = \mathbb{E}[\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}\|^2]$.

First, we note that for LSE,

$$B_1(\tilde{\boldsymbol{\theta}}_n) = \mathbf{0} \quad \text{and} \quad R_1(\tilde{\boldsymbol{\theta}}_n : \mathbf{I}_p) = \sigma^2(p_1 + p_2)$$

and for RLSE, $\hat{\boldsymbol{\theta}}_R = (\tilde{\boldsymbol{\theta}}_{1n}^\top, \mathbf{0}^\top)^\top$, we have

$$B_2(\hat{\boldsymbol{\theta}}_R) = (\mathbf{0}^\top, \boldsymbol{\theta}_2^\top) \quad \text{and} \quad R_2(\hat{\boldsymbol{\theta}}_R; \mathbf{I}_p) = \sigma^2(p_1 + \Delta^2).$$

3.1 Hard Threshold Estimator (Subset Selection Rule)

The bias of this estimator is given by

$$B(\hat{\boldsymbol{\theta}}_n^{HT}(\kappa)) = \left(-\sigma n_j^{-\frac{1}{2}} \Delta_j H_3(\kappa^2; \Delta_j^2) | j = 1, \dots, p \right)^\top,$$

where $H_\nu(\cdot; \Delta_j^2)$ is the cumulative distribution function (c.d.f.) of a non-central χ^2 -distribution with ν d.f. and non-centrality parameter Δ_j^2 ($j = 1, \dots, p$).

The MSE of $\hat{\boldsymbol{\theta}}_n^{HT}(\kappa)$ is given by

$$\begin{aligned} R(\hat{\boldsymbol{\theta}}_n^{HT}(\kappa) : \mathbf{I}_p) &= \sum_{j=1}^p \mathbb{E}[\tilde{\theta}_{jn} I(|\tilde{\theta}_{jn}| > \kappa \sigma n_j^{-\frac{1}{2}}) - \theta_j]^2 \\ &= \sigma^2 \sum_{j=1}^p n_j^{-1} \{ (1 - H_3(\kappa^2; \Delta_j^2)) + \Delta_j^2 (2H_3(\kappa^2; \Delta_j^2) - H_5(\kappa^2; \Delta_j^2)) \}. \end{aligned} \quad (3.2)$$

Since $[\tilde{\theta}_{jn}I(|\tilde{\theta}_{jn}| > \kappa\sigma n_j^{-\frac{1}{2}}) - \theta_j]^2 \leq (\tilde{\theta}_{jn} - \theta_j)^2 + \theta_j^2$, we obtain

$$R(\hat{\boldsymbol{\theta}}_n^{HT}(\kappa) : \mathbf{I}_p) \leq \sigma^2 \text{tr} \mathbf{N}^{-1} + \boldsymbol{\theta}^\top \boldsymbol{\theta} \quad (\text{free of } \kappa).$$

Following Donoho and Johnstone (1994), one can show that the following holds.

Lemma 3.1. *Under the assumed regularity conditions*

$$R(\hat{\boldsymbol{\theta}}_n^{HT}(\kappa) : \mathbf{I}_p) \leq \begin{cases} (i) & \sigma^2(1 + \kappa^2) \text{tr} \mathbf{N}^{-1} & \forall \boldsymbol{\theta} \in \mathbb{R}^p, \kappa > 1, \\ (ii) & \sigma^2 \text{tr} \mathbf{N}^{-1} + \boldsymbol{\theta}^\top \boldsymbol{\theta} & \forall \boldsymbol{\theta} \in \mathbb{R}^p, \\ (iii) & \sigma^2 \rho_{HT}(\kappa, 0) \text{tr} \mathbf{N}^{-1} + 1.2 \boldsymbol{\theta}^\top \boldsymbol{\theta} & 0 < \boldsymbol{\theta} < \kappa \mathbf{1}_p^\top \end{cases}$$

where $\rho_{HT}(\kappa, 0) = 2[(1 - \Phi(\kappa)) + \kappa\varphi(\kappa)]$, and $\varphi(\cdot)$ and $\Phi(\cdot)$ are the probability density function (p.d.f.) and c.d.f. of standard normal distribution, respectively.

Theorem 1. *Under the assumed regularity conditions, the weighted L_2 -risk bounds are given by*

$$R(\hat{\boldsymbol{\theta}}_n^{HT}(\kappa) : \mathbf{N}_1, \mathbf{N}_2) \leq \begin{cases} (i) & \sigma^2(1 + \kappa^2)(p_1 + p_2) & \kappa > 1, \\ (ii) & \sigma^2(p_1 + p_2) + \boldsymbol{\theta}_1^\top \mathbf{N}_1 \boldsymbol{\theta}_1 + \boldsymbol{\theta}_2^\top \mathbf{N}_2 \boldsymbol{\theta}_2 & \forall \boldsymbol{\theta} \in \mathbb{R}^p \\ (iii) & \sigma^2 \rho_{HT}(\kappa, 0)(p_1 + p_2) \\ & + 1.2\{\boldsymbol{\theta}_1^\top \mathbf{N}_1 \boldsymbol{\theta}_1 + \boldsymbol{\theta}_2^\top \mathbf{N}_2 \boldsymbol{\theta}_2\} & 0 < \boldsymbol{\theta} < k \mathbf{1}_p^\top \end{cases}$$

If the solution of $\hat{\boldsymbol{\theta}}_n^{HT}(\kappa)$ has the configuration $(\tilde{\boldsymbol{\theta}}_{1n}^\top, \mathbf{0}^\top)^\top$, then the L_2 -risk of $\hat{\boldsymbol{\theta}}_n^{HT}(\kappa)$ is given by

$$R_3(\hat{\boldsymbol{\theta}}_n^{HT}(\kappa) : \mathbf{N}_1, \mathbf{N}_2) = \sigma^2[p_1 + \Delta^2],$$

independent of κ .

3.2 LASSO Estimator

The bias expression of LASSO estimator is given by

$$B(\boldsymbol{\theta}_n^L(\lambda)) = \left(\sigma n_j^{-\frac{1}{2}} [\lambda(2\Phi(\Delta_j) - 1)]; j = 1, \dots, p_1; -\Delta_{p_1+1} \dots \Delta_p \right)^\top$$

The MSE of the LASSO estimator has the form

$$R(\hat{\boldsymbol{\theta}}_n^L(\lambda) : \mathbf{I}_p) = \sigma^2 \sum_{j=1}^{p_1} n_j^{-1} \rho_{ST}(\lambda, \Delta_j) + \Delta^2,$$

where

$$\begin{aligned} \rho_{ST}(\lambda, \Delta_j) &= (1 + \lambda^2)\{1 - \Phi(\lambda - \Delta_j) + \Phi(-\lambda - \Delta_j)\} + \Delta_j^2\{\Phi(\lambda - \Delta_j) - \Phi(-\lambda - \Delta_j)\} \\ &\quad - \{(\lambda - \Delta_j)\varphi(\lambda + \Delta_j) + (\lambda + \Delta_j)\varphi(\lambda - \Delta_j)\} \end{aligned}$$

Thus due to Donoho and Johnstone (1994, appendix 2), we have the following result.

Lemma 3.2. *Under the assumed regularity conditions*

$$\mathbf{R}(\hat{\boldsymbol{\theta}}_n^L(\lambda) : \mathbf{I}_p) \leq \begin{cases} (i) & \sigma^2(1 + \lambda^2) \operatorname{tr} \mathbf{N}^{-1} & \forall \boldsymbol{\theta} \in \mathbb{R}^p, \kappa > 1 \\ (ii) & \sigma^2 \operatorname{tr} \mathbf{N}^{-1} + \boldsymbol{\theta}^\top \boldsymbol{\theta} & \forall \boldsymbol{\theta} \in \mathbb{R}^p \\ (iii) & \sigma^2 \rho_{ST}(\lambda, 0) \operatorname{tr} \mathbf{N}^{-1} + 1.2\boldsymbol{\theta}^\top \boldsymbol{\theta} & \forall \boldsymbol{\theta} \in \mathbb{R}^p \end{cases}$$

where $\rho_{ST}(\lambda, 0) = 2 [(1 + \lambda^2)(1 - \Phi(\lambda)) - \kappa\phi(\lambda)]$.

If the solution of $\hat{\boldsymbol{\theta}}_n^L(\lambda)$ has the configuration $(\hat{\boldsymbol{\theta}}_{1n}^\top, \mathbf{0}^\top)$, then the L_2 -risk of $\hat{\boldsymbol{\theta}}_n^L(\lambda)$ is given by

$$\mathbf{R}_4(\hat{\boldsymbol{\theta}}_n^L(\lambda) : \mathbf{N}_1, \mathbf{N}_2) = \sigma^2(p_1 + \Delta^2). \tag{3.3}$$

Thus, we note that

$$\mathbf{R}_2(\hat{\boldsymbol{\theta}}_n : \mathbf{N}_1, \mathbf{N}_2) = \mathbf{R}_3(\hat{\boldsymbol{\theta}}_n^{HT}(\kappa) : \mathbf{N}_1, \mathbf{N}_2) = \mathbf{R}_4(\hat{\boldsymbol{\theta}}_n^L(\lambda) : \mathbf{N}_1, \mathbf{N}_2) = \sigma^2(p_1 + \Delta^2).$$

To proof L_2 risk of LASSO, we consider the multivariate decision theory. We are given the LSE of $\boldsymbol{\theta}$ as $\tilde{\boldsymbol{\theta}}_n = (\tilde{\theta}_{1n}, \dots, \tilde{\theta}_{pn})^\top$ according to

$$\tilde{\theta}_{jn} = \theta_j + \sigma n_j^{-\frac{1}{2}} \mathcal{Z}_j, \quad \mathcal{Z}_j \sim \mathcal{N}(0, 1),$$

where $\sigma n_j^{-\frac{1}{2}}$ is the marginal variance of $\tilde{\theta}_{jn}$ and noise level, and $\{\theta_j\}_{j=1, \dots, p}$ are the treatment effects of interest. We measure the quality of the estimators based on the L_2 -risk, $\mathbf{R}(\tilde{\boldsymbol{\theta}}_n : \mathbf{I}_p) = \mathbb{E}[\|\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|^2]$. Note that, for sparse solution, we use (3.1).

Consider the family of diagonal linear projections,

$$\mathbf{T}_{DP}(\hat{\boldsymbol{\theta}}_n^L(\lambda) : \boldsymbol{\delta}) = (\delta_1 \hat{\theta}_{1n}^L(\lambda), \dots, \delta_p \hat{\theta}_{pn}^L(\lambda))^\top \tag{3.4}$$

with $\boldsymbol{\delta} = (\delta_1, \dots, \delta_p)^\top$, $\delta_j \in (0, 1)$, $j = 1, \dots, p$. Such estimators “kill” or “keep” the coordinates.

Suppose we had available an oracle which would supply for us the coefficients δ_j optimal for use in the diagonal projection scheme (3.4). These “ideal” coefficients are $\delta_j = I(|\theta_j| > \sigma n_j^{-\frac{1}{2}})$, ideal diagonal projections consist of estimating only those θ_j , larger than its noise, $\sigma n_j^{-\frac{1}{2}}$ ($j = 1, 2, \dots, p$). These yield the “ideal” L_2 -risk given by (3.5) below.

Then, the ideal diagonal coordinates in our study are $I(|\theta_j| > \sigma n_j^{-\frac{1}{2}})$. These coordinates estimate those treatment effects θ_j which are larger than the noise level $\sigma n_j^{-1/2}$, yielding the “ideal” L_2 -risk as

$$R(\hat{\boldsymbol{\theta}}_n^L(\lambda) : \mathbf{I}_p) = \sum_{j=1}^p \min(\theta_j^2, \sigma^2 n_j^{-1}) = \begin{cases} \sigma^2 \text{tr } \mathbf{N}^{-1} & \forall |\theta_j| > \sigma n_j^{-\frac{1}{2}}, j = 1, \dots, p, \\ \boldsymbol{\theta}^\top \boldsymbol{\theta} & \forall |\theta_j| < \sigma n_j^{-\frac{1}{2}}, j = 1, \dots, p. \end{cases} \quad (3.5)$$

Thus, we find that ideal lower bound of the L_2 -risk of $\hat{\boldsymbol{\theta}}_n^L(\lambda)$ leading to “keep” or “kill” solution is $(\tilde{\boldsymbol{\theta}}_{1n}^\top, \mathbf{0}^\top)^\top$. Thus, $R(\text{DP}; \mathbf{N}_1, \mathbf{N}_2) = \sigma^2(p_1 + \Delta^2)$ which is the same as (3.5).

In general, the L_2 -risk given by (3.5) cannot be achieved for all $\boldsymbol{\theta}$ by any linear or nonlinear estimator of treatment effects. However, in the sparse case, if p_1 treatment effects $|\theta_j|$ exceed $\sigma n_j^{-1/2}$ and p_2 coefficients are null, then we obtain the ideal L_2 -risk given by

$$R(\hat{\boldsymbol{\theta}}_n^L(\lambda) : \mathbf{I}_p) = \sigma^2 \text{tr } \mathbf{N}_1^{-1} + \boldsymbol{\theta}_2^\top \boldsymbol{\theta}_2.$$

This ideal L_2 -risk happen to be the lower bound of L_2 -risk given by (3.2). We shall use this ideal L_2 -risk to compare with the L_2 -risk of other estimators. Consequently, the lower bound of the weighted L_2 -risk is given by

$$R(\hat{\boldsymbol{\theta}}_n^L(\lambda) : \mathbf{N}_1, \mathbf{N}_2) = \sigma^2(p_1 + \Delta^2).$$

3.3 Ridge Regression Estimator

Recall that the ridge regression estimator is given by

$$\hat{\boldsymbol{\theta}}_n^{RR}(k) = \begin{pmatrix} \tilde{\boldsymbol{\theta}}_{1n} \\ \frac{1}{1+k} \tilde{\boldsymbol{\theta}}_{2n} \end{pmatrix}, \quad k \in \mathbb{R}^+.$$

The bias and MSE of $\hat{\boldsymbol{\theta}}_n^{RR}(k)$ have forms

$$B(\hat{\boldsymbol{\theta}}_n^{RR}(k)) = \begin{pmatrix} \mathbf{0} \\ -\frac{k}{1+k} \boldsymbol{\theta}_2 \end{pmatrix}$$

and

$$R(\hat{\boldsymbol{\theta}}_n^{RR}(k) : \mathbf{I}_p) = \sigma^2 \text{tr } \mathbf{N}_1^{-1} + \frac{1}{(1+k)^2} (\sigma^2 \text{tr } \mathbf{N}_2^{-1} + k^2 \boldsymbol{\theta}_2^\top \boldsymbol{\theta}_2).$$

Hence, the weighted L_2 -risk is obtained as

$$R(\hat{\boldsymbol{\theta}}_n^{RR}(k) : \mathbf{N}_1, \mathbf{N}_2) = \sigma^2 p_1 + \frac{\sigma^2}{(1+k)^2} (p_2 + k^2 \Delta^2).$$

One may find the optimum value $k = k_o = p_2 \Delta^{-2}$, yielding

$$R_5(\hat{\boldsymbol{\theta}}_n^{RR}(k_o) : \mathbf{N}_1, \mathbf{N}_2) = \sigma^2 \left(p_1 + \frac{p_2 \Delta^2}{p_2 + \Delta^2} \right).$$

4 Comparison of Estimators

In this section, we compare various estimators with respect to the unrestricted estimator (LSE), in term of relative weighted L_2 -risk efficiency (RWRE).

We recall that for sparse solution, the L_2 -risk of LASSO is $\sigma^2(p_1 + \Delta^2)$ (Eq. 3.3) which is also the “ideal” L_2 -risk in an “ideal” diagonal projection scheme. Therefore, we shall use $\sigma^2(p_1 + \Delta^2)$ to compare the L_2 -risks function of other estimators.

4.1 Comparison of LSE with RLSE

Recall that the restricted LSE (RLSE) is given by $\hat{\boldsymbol{\theta}}_n = (\tilde{\boldsymbol{\theta}}_{1n}^\top, \mathbf{0}^\top)^\top$. In this case, the relative weighted L_2 -risk efficiency (RWRE) of RLSE versus LSE is given by

$$\text{RWRE}(\hat{\boldsymbol{\theta}}_n : \tilde{\boldsymbol{\theta}}_n) = \frac{p_1 + p_2}{p_1 + \Delta^2} = \left(1 + \frac{p_2}{p_1}\right) \left(1 + \frac{\Delta^2}{p_1}\right)^{-1},$$

which is a decreasing function of Δ^2 . So, $0 \leq \text{RWRE}(\hat{\boldsymbol{\theta}}_n : \tilde{\boldsymbol{\theta}}_n) \leq \left(1 + \frac{p_2}{p_1}\right)$.

4.2 Comparison of LSE with PTE

Here, it is easy to see that

$$R_6(\hat{\boldsymbol{\theta}}_n^{PT}(\alpha) : \mathbf{N}_1, \mathbf{N}_2) = p_1 + p_2 (1 - H_{p_2+2}(c_\alpha; \Delta^2)) + \Delta^2 [2H_{p_2+2}(c_\alpha; \Delta^2) - H_{p_2+4}(c_\alpha; \Delta^2)].$$

Then, the RWRE expression for PTE versus LSE is given by

$$\text{RWRE}(\hat{\boldsymbol{\theta}}_n^{PT}(\alpha) : \tilde{\boldsymbol{\theta}}_n) = \frac{p_1 + p_2}{p_1 + p_2 (1 - H_{p_2+2}(c_\alpha; \Delta^2)) + \Delta^2 [2H_{p_2+2}(c_\alpha; \Delta^2) - H_{p_2+4}(c_\alpha; \Delta^2)]}.$$

Then, the PTE outperforms the LSE for

$$0 \leq \Delta^2 \leq \frac{p_2 H_{p_2+2}(c_\alpha; \Delta^2)}{2H_{p_2+2}(c_\alpha; \Delta^2) - H_{p_2+4}(c_\alpha; \Delta^2)} = \Delta_{PT}^2. \quad (4.1)$$

Otherwise, LSE outperforms the PTE in the interval (Δ_{PT}^2, ∞) . We may mention that $\text{RWRE}(\hat{\boldsymbol{\theta}}_n^{PT}(\alpha) : \tilde{\boldsymbol{\theta}}_n)$ is a decreasing function of Δ^2 with a maximum at $\Delta^2 = 0$, then decreases crossing the 1-line to a minimum at $\Delta^2 = \Delta_{PT}^2(\text{min})$ with a value $M_{PT}(\alpha)$ then increases toward 1-line. This means the gains in efficiency of PTE is the highest in the interval given by Eq. (4.1) and loss in efficiency can be noticed outside it.

The RWRE $(\hat{\boldsymbol{\theta}}_n^{PT}; \tilde{\boldsymbol{\theta}}_n)$ belongs to the interval

$$M_{PT}(\alpha) \leq \text{RWRE}(\hat{\boldsymbol{\theta}}_n^{PT}(\alpha); \tilde{\boldsymbol{\theta}}_n) \leq \left(1 + \frac{p_2}{p_1}\right) \left(1 + \frac{p_2}{p_1} [1 - H_{p_2+2}(c_\alpha; 0)]\right)^{-1},$$

where $M_{PT}(\alpha)$ depends on the size α and given by

$$M_{PT}(\alpha) = \left(1 + \frac{p_2}{p_1}\right) \left\{1 + \frac{p_2}{p_1} [1 - H_{p_2+2}(c_\alpha; \Delta_{PT}^2(\min))] + \frac{\Delta_{PT}^2(\min)}{p_1} [2H_{p_2+2}(c_\alpha; \Delta_{PT}^2(\min)) - H_{p_2+4}(c_\alpha; \Delta_{PT}^2(\min))]\right\}^{-1}.$$

The quantity $\Delta_{PT}^2(\min)$ is the value Δ^2 at which the RWRE value is minimum.

4.3 Comparison of LSE with JSE and PRSE

Since JSE and PRSE need $p_2 \geq 3$ to express their weighted L_2 -risk (WL_2R) expressions, we assume always $p_2 \geq 3$. First, note that

$$R_7(\hat{\theta}_n^{JS}; \mathbf{N}_1, \mathbf{N}_2) = p_1 + p_2 - (p_2 - 2)^2 E[\chi_{p_2}^{-2}(\Delta^2)].$$

As a result, we obtain

$$\text{RWRE}(\hat{\theta}_n^{JS}; \tilde{\theta}_n) = \left(1 + \frac{p_2}{p_1}\right) \left(1 + \frac{p_2}{p_1} - \frac{(p_2 - 2)^2}{p_1} E[\chi_{p_2}^{-2}(\Delta^2)]\right)^{-1}.$$

It is a decreasing function of Δ^2 . At $\Delta^2 = 0$, its value is $\left(1 + \frac{p_2}{p_1}\right) \left(1 + \frac{2}{p_1}\right)^{-1}$ and when $\Delta^2 \rightarrow \infty$, its value goes to 1. Hence, for $\Delta^2 \in \mathbb{R}^+$,

$$1 \leq \left(1 + \frac{p_2}{p_1}\right) \left(1 + \frac{p_2}{p_1} - \frac{(p_2 - 2)^2}{p_1} E[\chi_{p_2}^{-2}(\Delta^2)]\right)^{-1} \leq \left(1 + \frac{p_2}{p_1}\right) \left(1 + \frac{2}{p_1}\right)^{-1}.$$

Hence, the gains in efficiency is the highest when Δ^2 is small and drops towards 1 when Δ^2 is the largest. Also,

$$\begin{aligned} \text{RWRE}(\hat{\theta}_n^{S+}; \tilde{\theta}_n) &= \left(1 + \frac{p_2}{p_1}\right) \left(1 + \frac{p_2}{p_1} - \frac{(p_2 - 2)^2}{p_1} E[\chi_{p_2}^{-2}(\Delta^2)] \right. \\ &\quad \left. - \frac{p_2}{p_1} E\left[(1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta^2))^2 I(\chi_{p_2+2}^2(\Delta^2) < (p_2 - 2))\right] \right. \\ &\quad \left. + \frac{\Delta^2}{p_1} \left\{2E\left[(1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta^2)) I(\chi_{p_2+2}^2(\Delta^2) < (p_2 - 2))\right] \right. \right. \\ &\quad \left. \left. - E\left[(1 - (p_2 - 2)\chi_{p_2+4}^{-2}(\Delta^2))^2 I(\chi_{p_2+4}^2(\Delta^2) < (p_2 - 2))\right]\right\} \right)^{-1}. \end{aligned}$$

So that,

$$\text{RWRE}(\hat{\theta}_n^{S+}; \tilde{\theta}_n) \geq \text{RWRE}(\hat{\theta}_n^{JS}; \tilde{\theta}_n) \geq 1 \quad \forall \Delta^2 \in \mathbb{R}^+.$$

We also provide graphical representation (Figure 1) of RWRE of the estimators.

In the next subsection, we show that RRE uniformly dominates all other estimators although it does not select variables.

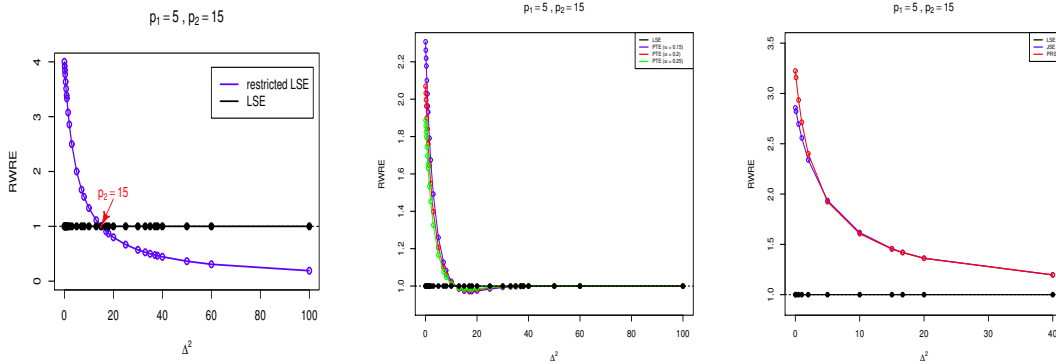


Figure 1: RWRE for the restricted, preliminary test, Stein-type and its positive rule estimators.

4.4 Comparison of LSE and RLSE with RRE

First we consider weighted L_2 -risk difference of LSE and RRE given by

$$\sigma^2(p_1 + p_2) - \sigma^2 p_1 - \sigma^2 \frac{p_2 \Delta^2}{p_2 + \Delta^2} = \sigma^2 p_2 \left(1 - \frac{\Delta^2}{p_2 + \Delta^2} \right) = \frac{\sigma^2 p_2^2}{p_2 + \Delta^2} > 0, \forall \Delta^2 \in \mathbb{R}^+$$

Hence, RRE outperforms the LSE uniformly. Similarly, for the RLSE and RRE, the weighted L_2 -risk difference is given by

$$\sigma^2 (p_1 + \Delta^2) - \left(\sigma^2 p_1 + \frac{\sigma^2 p_2 \Delta^2}{p_2 + \Delta^2} \right) = \frac{\sigma^2 \Delta^4}{p_2 + \Delta^2} > 0.$$

Therefore, RRE performs better than RLSE uniformly.

In addition, the RWRE of RRE versus LSE equals

$$\text{RWRE}(\hat{\theta}_n^{RR}(k_o) : \tilde{\theta}_n) = \frac{p_1 + p_2}{p_1 + \frac{p_2 \Delta^2}{p_2 + \Delta^2}} = \left(1 + \frac{p_2}{p_1} \right) \left(1 + \frac{p_2 \Delta^2}{p_1 (p_2 + \Delta^2)} \right)^{-1},$$

which is a decreasing function of Δ^2 with maximum $\left(1 + \frac{p_2}{p_1} \right)$ at $\Delta^2 = 0$ and minimum 1 as $\Delta^2 \rightarrow \infty$. So,

$$1 \leq \left(1 + \frac{p_2}{p_1} \right) \left(1 + \frac{p_2}{p_1 \left(1 + \frac{\Delta^2}{p_2} \right)} \right)^{-1} \leq 1 + \frac{p_2}{p_1}; \quad \forall \Delta^2 \in \mathbb{R}^+.$$

4.5 Comparison of RRE with PTE, JSE and PRSE

Here, the weighted L_2 -risk difference of PTE and RRE is given by

$$\begin{aligned} & \sigma^2 \left[p_2 - p_2 1 - H_{p_2+2}(c_\alpha; \Delta^2) + \Delta^2 \{ 2H_{p_2+2}(c_\alpha; \Delta^2) - H_{p_2+4}(c_\alpha; \Delta^2) \} \right] - \frac{\sigma^2 p_2 \Delta^2}{p_2 + \Delta^2} \\ & = \sigma^2 \left[\frac{p_2}{p_2 + \Delta^2} - \{ p_2 H_{p_2+2}(c_\alpha; \Delta^2) - \Delta^2 (2H_{p_2+2}(c_\alpha; \Delta^2) - H_{p_2+4}(c_\alpha; \Delta^2)) \} \right] \geq 0. \end{aligned} \quad (4.2)$$

Since the first term is a decreasing function of Δ^2 with a maximum value p_2 at $\Delta^2 = 0$ and tends to 0 as $\Delta^2 \rightarrow \infty$. The second function in the bracket is also decreasing in Δ^2 with maximum $p_2 H_{p_2+2}(c_\alpha; 0)$ at $\Delta^2 = 0$ which is less than p_2 and the function tends to 0 as $\Delta^2 \rightarrow \infty$. Hence, Eq. (4.2) is non-negative for $\Delta^2 \in \mathbb{R}^+$. Thus, the RRE uniformly performs better than PTE.

Similarly, we show that RRE uniformly performs better than the JSE, i.e.,

$$\text{WL}_2\text{R}(\hat{\theta}_n^{RR}(k_o)) \leq \text{WL}_2\text{R}(\hat{\theta}_n^{JS}), \quad \forall \Delta^2 \in \mathbb{R}^+.$$

The weighted L_2 -risk difference of JSE and RRE is given by

$$\sigma^2 \left[\frac{p_2^2}{p_2 + \Delta^2} - p_2 \{ E[\chi_{p_2+2}^{-2}(\Delta^2)] + \Delta^2 E[\chi_{p_2+2}^{-4}(\Delta^2)] \} - (p_2^2 - 4) E[\chi_{p_2+4}^{-4}(\Delta^2)] \right] \geq 0$$

$\forall \Delta^2 \in \mathbb{R}^+$,

since the first function decreases with a maximum value p_2 at Δ^2 , also the second function decreases with a maximum value $1 (\leq p_2)$ and tends to 0 as $\Delta^2 \rightarrow \infty$. For that reason, the two functions are one below other one and the difference is non-negative for $\Delta^2 \in \mathbb{R}^+$.

Next, we show that the weighted L_2 -risk (WL_2R) of the two estimators may be ordered as

$$\text{WL}_2\text{R}(\hat{\theta}_n^{RR}(k_o)) \leq \text{WL}_2\text{R}(\hat{\theta}_n^{S+}), \quad \forall \Delta^2 \in \mathbb{R}^+.$$

Note that

$$\text{R}_8(\hat{\theta}_n^{S+} : \mathbf{N}_1, \mathbf{N}_2) = \text{R}_7(\hat{\theta}_n^{JS} : \mathbf{N}_1, \mathbf{N}_2) - R^*, \quad (4.3)$$

where

$$\begin{aligned} R^* &= \sigma^2 p_2 E \left[(1 - (p_2 - 2) \chi_{p_2+2}^{-2}(\Delta^2))^2 I(\chi_{p_2+2}^{-2}(\Delta^2) < p_2 - 2) \right] \\ &+ \Delta^2 \left\{ 2E \left[(1 - (p_2 - 2) \chi_{p_2+2}^{-2}(\Delta^2)) I(\chi_{p_2+2}^{-2}(\Delta^2) < p_2 - 2) \right] \right. \\ &\left. - E \left[(1 - (p_2 - 2) \chi_{p_2+4}^{-2}(\Delta^2))^2 I(\chi_{p_2+4}^{-2}(\Delta^2) < p_2 - 2) \right] \right\}. \end{aligned} \quad (4.4)$$

Thus, we find that the WL_2R -difference is given by

$$\begin{aligned} \text{WL}_2\text{R}(\hat{\theta}_n^{S+}) - \text{WL}_2\text{R}(\hat{\theta}_n^{RR}(k_o)) &= \{ \text{WL}_2\text{R}(\hat{\theta}_n^{JS}) - \text{WL}_2\text{R}(\hat{\theta}_n^{RR}(k_o)) \} - R^* \\ &= \{ \cdot \} \text{is negative} - R^* (\text{non-negative}) \leq 0. \end{aligned} \quad (4.5)$$

Hence, the RRE uniformly performs better than the PRSE.

4.6 Comparison of LASSO with LSE and RLSE

First note that if we have for p_1 coefficients, $|\beta_j| > \sigma n_j^{-1/2}$ and also p_2 coefficients are zero in a sparse solution, then the “ideal” weighted L_2 -risk is given by $\sigma^2(p_1 + \Delta^2)$. Thereby, we compare all estimators relative to this quantity. Hence, weighted L_2 -risk difference between LSE and LASSO is given by

$$\sigma^2(p_1 + p_2) - \sigma^2(p_1 + \Delta^2) = \sigma^2 [p_2 - \Delta^2].$$

Hence, if $\Delta^2 \in (0, p_2)$, the LASSO performs better than the LSE, while if $\Delta^2 \in (p_2, \infty)$ the LSE performs better than the LASSO. Consequently, neither LSE nor the LASSO performs better than the other, uniformly.

Next, we compare the RLSE and LASSO. In this case, the weighted L_2 -risk difference is given by

$$\sigma^2(p_1 + \Delta^2) - \sigma^2(p_1 + \Delta^2) = 0.$$

Hence, LASSO and RLSE are L_2 -risk equivalent. And consequently, the LASSO satisfies the oracle properties.

4.7 Comparison of LASSO with PTE, JSE and PRSE

We first consider the PTE versus LASSO. In this case, the weighted L_2 -risk difference is given by

$$R_6(\hat{\theta}_n^{PT}(\alpha) : \mathbf{N}_1, \mathbf{N}_2) - R_4(\hat{\theta}_n^L(\lambda) : \mathbf{N}_1, \mathbf{N}_2) \tag{4.6}$$

$$= \sigma^2 [p_2(1 - H_{p_2+2}(c_\alpha; \Delta^2)) - \Delta^2 \{1 - 2H_{p_2+2}(c_\alpha; \Delta^2) + H_{p_2+4}(c_\alpha; \Delta^2)\}] \tag{4.7}$$

$$\geq \sigma^2 p_2(1 - H_{p_2+2}(c_\alpha; 0)) \geq 0, \quad \text{if } \Delta^2 = 0.$$

Hence, the LASSO outperform the PTE when $\Delta^2 = 0$. But, when $\Delta^2 \neq 0$, the LASSO outperforms the PTE for

$$0 \leq \Delta^2 \leq \frac{p_2 [1 - H_{p_2+2}(c_\alpha; \Delta^2)]}{1 - 2H_{p_2+2}(c_\alpha; \Delta^2) + H_{p_2+4}(c_\alpha; \Delta^2)}.$$

Otherwise, PTE outperforms the LASSO. Hence, neither LASSO nor PTE outmatches the other uniformly.

Next, we consider JSE and PRSE versus the LASSO. In these two cases, we have weighted L_2 -risk differences given by

$$\begin{aligned} R_7(\hat{\theta}_n^{JS} : \mathbf{N}_1, \mathbf{N}_2) - R_4(\hat{\theta}_n^L(\lambda) : \mathbf{N}_1, \mathbf{N}_2) \\ = \sigma^2 \left[p_1 + p_2 - (p_2 - 2)^2 E [\chi_{p_2+2}^{-2}(\Delta^2)] - (p_1 + \Delta^2) \right] \\ = \sigma^2 \left[p_2 - (p_2 - 2)^2 E [\chi_{p_2+2}^{-2}(\Delta^2)] - \Delta^2 \right] \end{aligned} \tag{4.8}$$

and from (4.3),

$$R_8(\hat{\theta}_n^{S+} : N_1, N_2) - R_4(\hat{\theta}_n^L(\lambda) : N_1, N_2) = R(\hat{\theta}_n^{JS} : N_1, N_2) - R(\hat{\theta}_n^L(\lambda) : N_1, N_2) - R^*,$$

where R^* is given by (4.4). Therefore, the LASSO outperforms the JSE as well as the PRSE in the interval $[0, p_2 - (p_2 - 2)^2 E[\chi_{p_2}^{-2}(\Delta^2)]]$. Thus, neither JSE nor the PRSE outperform the LASSO uniformly.

4.8 Comparison of LASSO with RRE

Here, the weighted L_2 -risk difference is given by

$$\begin{aligned} R_4(\hat{\theta}_n^L(\lambda); N_1, N_2) - R_5(\hat{\theta}_n^{RR}(k_o); N_1, N_2) &= \sigma^2 \left[(p_1 + \Delta^2) - \left(p_1 + \frac{p_2 \Delta^2}{p_2 + \Delta^2} \right) \right] \\ &= \frac{\sigma^2 \Delta^2}{p_2 + \Delta^2} \geq 0. \end{aligned}$$

Hence the RRE outperforms the LASSO uniformly.

In the Figure 2, the comparisons of LASSO with other estimators are showed.

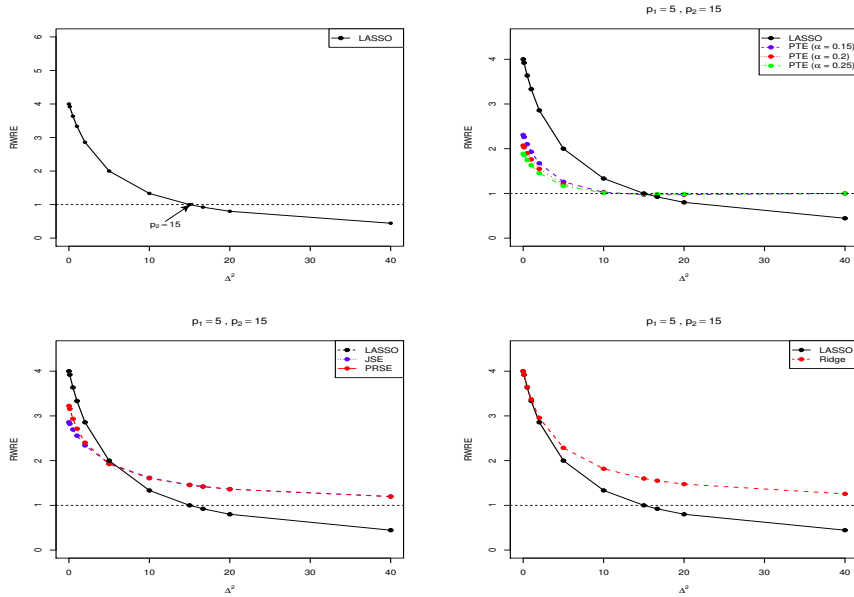


Figure 2: RWRE for the LASSO, ridge, restricted, preliminary test, Stein-type and its positive rule estimators.

5 Application

To illustrate the methodologies in Sections 4 and 5, we will consider the following numerical example in this section. The data were generated from $N(\theta_i, 1.25)$ distribution, where $\theta_i = (3, 1.5, 2.5, 0, 0, 4, 0, 0, 4.5, 0)'$ and $\sigma = 1.25$. The respective sample sizes are $n_i = (10, 15, 12, 20, 10, 15, 12, 20, 10, 16)'$. The generated data and some summary statistics are presented in the following Table 1.

Table 1: One way ANOVA table

	y1	y2	y3	y4	y5	y6	y7	y8	y9	y10
1	3.963	2.754	2.156	-0.776	0.589	5.134	0.050	2.372	1.984	-0.817
2	2.557	2.878	3.173	-0.516	-2.100	1.963	-1.055	1.475	3.358	0.953
3	0.948	1.364	1.440	-0.469	-1.447	2.120	0.928	0.435	7.123	-0.174
4	4.254	2.426	2.635	-0.192	0.078	3.385	-0.747	-2.213	3.843	-1.403
5	2.496	2.391	3.280	0.983	-1.605	3.997	1.095	1.409	5.122	-0.276
6	2.684	-0.168	6.687	-2.806	0.857	4.102	-1.857	0.985	3.814	0.819
7	3.684	1.104	0.973	0.729	-0.619	5.161	0.795	-0.376	5.232	-0.345
8	1.807	3.429	0.755	1.269	-2.191	3.661	0.578	0.161	5.310	1.000
9	2.029	1.744	2.511	1.106	-0.136	3.519	-0.445	0.662	3.897	0.362
10	0.466	3.411	0.186	-1.891	-2.289	5.997	-0.627	-0.555	4.066	0.632
11		2.264	3.382	1.192		3.502	-0.515	-0.552		0.458
12		3.445	4.780	-0.060		2.444	-2.082	-0.604		1.317
13		1.444		-1.658		5.942		0.272		-0.690
14		-0.154		-1.187		2.200		0.001		1.291
15		1.451		-0.659		3.270		-1.739		-0.860
16				1.144				-1.125		1.559
17				1.468				-2.784		
18				-0.014				0.572		
19				-0.243				-0.012		
20				0.652				0.719		
\bar{y}_i	2.489	1.986	2.663	-0.096	-0.886	3.760	-0.323	-0.45	4.375	0.239
s_i	1.243	1.168	1.815	1.185	1.190	1.320	1.048	1.264	1.390	0.906

First we compute the LASSO estimator and find the following

$$\hat{\theta}_n^{LASSO} = (2.535, 1.120, 2.075, 0.000, 0.000, 3.620, 0.000, 0.000, 4.035, 0.000)'$$

Since, LASSO kills $\hat{\theta}_4, \hat{\theta}_5, \hat{\theta}_7, \theta_8,$ and θ_{10} , we will be testing the following hypothesis

$$H_0 : \theta_2 = 0, \quad vs \quad H_a : \theta_2 \neq 0$$

where $\theta_2 = (\theta_4, \theta_5, \theta_7, \theta_8, \theta_{10})'$.

Note that the estimate of the overall mean is $\hat{\theta}_0 = 1.250$. The computed value of the test statistic for known $\sigma = 1.25$ is 6.561, which is less than the critical value of $\chi_{0.95,5}^2 = 9.488$, hence the null hypothesis will not be rejected. The estimated values of different estimators

Table 2: Estimated values of different estimators

	$\bar{\theta}$	RE	PT	Lasso	SE	PR
1	2.489	2.489	2.489	2.535	2.489	2.489
2	1.986	1.986	1.986	1.120	1.986	1.986
3	2.663	2.663	2.663	2.075	2.663	2.663
4	3.760	3.760	3.760	0.000	3.760	3.760
5	4.375	4.375	4.375	0.000	4.375	4.375
6	-0.096	0.000	0.000	3.620	-0.067	-0.067
7	-0.886	0.000	0.000	0.000	-0.616	-0.616
8	-0.323	0.000	0.000	0.000	-0.225	-0.225
9	-0.045	0.000	0.000	4.035	-0.031	-0.031
10	0.239	2.489	2.489	0.000	0.166	0.166

for θ are given in the following Table 2. For the computation of the estimators we used formulas in Section 2 and prepared the Table 1. For the comparison of weighted risks among the estimators, see Tables 3 to 8.

6 Summary and Concluding Remarks

In this section, we discuss the contents of the Tables 3 - 7 presented as confirmatory evidence of the theoretical findings of the estimators.

First, we note that we have two classes of estimators, namely, the traditional PTE and Stein-type estimators and the penalty estimators. The restricted LSE plays an important role due to the fact that LASSO belongs to the class of restricted estimators. We have the following conclusions from our study.

- (i) Since the inception of the ridge regression estimator by Hoerl and Kennard (1970), there have been articles comparing ridge with PTE and Stein-type estimators. From this study, we conclude that the ridge regression estimator dominates the LSE, PTE and Stein-type estimators uniformly. The ridge estimator dominates the LASSO estimator uniformly for Δ^2 greater than 0. They are L_2 -risk equivalent at $\Delta^2 = 0$ and at this point LASSO dominates all other estimators. The ridge estimator does not select variables but the LASSO estimator does. See Table 3 and graphs there of in Figure 3.
- (ii) The Restricted LSE (RLSE) and LASSO are L_2 -risk equivalent. Hence, LASSO satisfies “oracle properties”.
- (iii) The “ideal” L_2 -risk of HTE and LASSO does not depend on the threshold parameter, but the upper bound does. See Donoho and Johnstone (1994).
- (iv) The RWRE of estimators compared to the LSE depends upon the size of p_1 , p_2 and divergence parameter, Δ^2 . LASSO/RLSE and ridge outperform all of estimators when Δ^2 is 0.

- (v) The LASSO satisfies the “oracle properties” and it dominates LSE, PTE, JSE, and PRSE in the sub-interval of $[0, p_1)$. In this case, with a small number of active parameters, the LASSO and HTE performs best followed by ridge as pointed out by Tibshirani (1996).
- (vi) If p_1 is fixed and p_2 increases, the RWRE of all estimators increases. See Table 6.
- (vii) If p_2 is fixed and p_1 increases, the RWRE of all estimators decreases. Then, for a given p_2 small and p_1 large, the LASSO, PTE, JSE, and PRSE are competitive. See Table 7.
- (viii) The PRSE is always outperform JSE. See Tables 3-7.
- (ix) We illustrated the findings of the paper by an numeral application in Section 5.

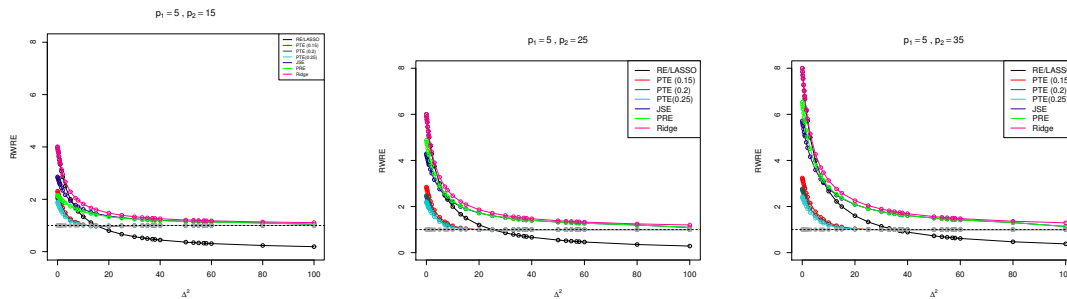


Figure 3: RWRE of estimates of a function of Δ^2 for $p_1 = 5$, and different p_2 .

Acknowledgments

Authors are thankful to referees for their constructive comments. Thanks are due to the technical help by Ms. Verna Bolivar. This article was partially completed while author, B. M. Golam Kibria was on sabbatical leave (Fall 2017). He is grateful to Florida International University for awarding the sabbatical leave which gave him excellent research facilities

References

Breiman, L. (1995), “Better subset regression using non-negative garrote,” *Technometrics*, 37, 373–384.

Donoho, D. L. and Johnstone, I. M (1994), “Ideal Spatial Adaptation by Wavelet Shrinkage,” *Biometrika*, 81(3), 425–455.

Table 3: RWRE for the estimators

(p_1, p_2, Δ^2)	LSE	RLSE/LASSO	PTE			JSE	PRSE	Ridge
			$\alpha = 0.15$	$\alpha = 0.2$	$\alpha = 0.25$			
(5, 15, 0)	1	4.00	2.30	2.07	1.89	2.86	3.22	4.00
(5, 15, 0.1)	1	3.92	2.26	2.03	1.85	2.82	3.16	3.92
(5, 15, 0.5)	1	3.64	2.10	1.89	1.74	2.69	2.93	3.64
(5, 15, 1)	1	3.33	1.93	1.76	1.63	2.56	2.71	3.36
(5, 15, 2)	1	2.86	1.67	1.55	1.45	2.33	2.40	2.96
(5, 15, 3)	1	2.50	1.49	1.40	1.33	2.17	2.19	2.67
(5, 15, 5)	1	2.00	1.26	1.21	1.17	1.94	1.92	2.26
(5, 15, 7)	1	1.67	1.13	1.10	1.08	1.78	1.77	2.04
(5, 15, 10)	1	1.33	1.02	1.02	1.01	1.62	1.60	1.81
(5, 15, 15)	1	1.00	0.97	0.97	0.98	1.46	1.45	1.60
(5, 15, 20)	1	0.80	0.97	0.98	0.98	1.36	1.36	1.47
(5, 15, 30)	1	0.57	0.99	0.99	0.99	1.25	1.25	1.33
(5, 15, 50)	1	0.36	0.99	0.99	1.00	1.16	1.16	1.21
(5, 15, 100)	1	0.19	1.00	1.00	1.00	1.05	1.05	1.11
(7, 33, 0)	1	5.71	2.86	2.50	2.23	4.44	4.92	5.71
(7, 33, 0.1)	1	5.63	2.82	2.46	2.20	4.40	4.84	5.63
(7, 33, 0.5)	1	5.33	2.66	2.34	2.10	4.23	4.57	5.34
(7, 33, 1)	1	5.00	2.49	2.20	1.98	4.03	4.28	5.02
(7, 33, 2)	1	4.44	2.21	1.97	1.80	3.71	3.84	4.50
(7, 33, 3)	1	4.00	1.99	1.79	1.65	3.45	3.51	4.10
(7, 33, 5)	1	3.33	1.67	1.53	1.43	3.05	3.05	3.53
(7, 33, 7)	1	2.86	1.46	1.36	1.29	2.76	2.74	3.13
(7, 33, 10)	1	2.35	1.26	1.20	1.16	2.46	2.44	2.72
(7, 33, 15)	1	1.82	1.09	1.07	1.05	2.13	2.11	2.31
(7, 33, 20)	1	1.48	1.02	1.02	1.01	1.92	1.91	2.06
(7, 33, 30)	1	1.08	0.99	0.99	0.99	1.67	1.67	1.76
(7, 33, 33)	1	1.00	0.99	0.99	0.99	1.62	1.62	1.70
(7, 33, 50)	1	0.70	0.99	0.99	0.99	1.43	1.43	1.49
(7, 33, 100)	1	0.37	1.00	1.00	1.00	1.12	1.12	1.25

Table 4: RWRE of the estimators for $p \in \{10, 20\}$ and different Δ^2 -value for varying p_1

Estimators	$p = 10$							
	$\Delta^2 = 0$				$\Delta^2 = 1$			
	$p_1 = 2$	$p_1 = 3$	$p_1 = 5$	$p_1 = 7$	$p_1 = 2$	$p_1 = 3$	$p_1 = 5$	$p_1 = 7$
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RLSE/LASSO	5.00	3.33	2.00	1.43	3.33	2.50	1.67	1.25
PTE ($\alpha = 0.15$)	2.34	1.98	1.51	1.23	1.75	1.55	1.27	1.09
PTE ($\alpha = 0.2$)	2.06	1.80	1.43	1.19	1.60	1.45	1.22	1.07
PTE ($\alpha = 0.25$)	1.86	1.66	1.36	1.16	1.49	1.37	1.18	1.06
JSE	2.50	2.00	1.43	1.11	2.14	1.77	1.33	1.08
PRSE	3.03	2.31	1.56	1.16	2.31	1.88	1.38	1.10
Ridge	5.00	3.33	2.00	1.43	3.46	2.58	1.71	1.29
$p = 10$								
$\Delta^2 = 5$				$\Delta^2 = 10$				
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RLSE/LASSO	1.43	1.25	1.00	0.83	0.83	0.77	0.67	0.59
PTE ($\alpha = 0.15$)	1.05	1.01	0.95	0.92	0.92	0.92	0.92	0.94
PTE ($\alpha = 0.2$)	1.03	1.00	0.95	0.93	0.94	0.93	0.94	0.95
PTE ($\alpha = 0.25$)	1.02	0.99	0.96	0.94	0.95	0.95	0.95	0.97
JSE	1.55	1.38	1.15	1.03	1.33	1.22	1.09	1.01
PRSE	1.53	1.37	1.15	1.03	1.32	1.22	1.08	1.01
Ridge	1.97	1.69	1.33	1.13	1.55	1.40	1.20	1.07
$p = 10$								
$\Delta^2 = 20$				$\Delta^2 = 60$				
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RLSE/LASSO	0.45	0.43	0.40	0.37	0.16	0.16	0.15	0.15
PTE ($\alpha = 0.15$)	0.97	0.97	0.98	0.99	1.00	1.00	1.00	1.00
PTE ($\alpha = 0.2$)	0.98	0.98	0.99	0.99	1.00	1.00	1.00	1.00
PTE ($\alpha = 0.25$)	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00
JSE	1.17	1.12	1.04	1.00	1.06	1.04	1.01	1.00
PRSE	1.17	1.12	1.04	1.00	1.05	1.04	1.01	1.00
Ridge	1.30	1.22	1.11	1.04	1.10	1.08	1.04	1.01
$p = 20$								
$\Delta^2 = 0$				$\Delta^2 = 1$				
Estimators	$p_1 = 2$	$p_1 = 3$	$p_1 = 5$	$p_1 = 7$	$p_1 = 2$	$p_1 = 3$	$p_1 = 5$	$p_1 = 7$
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RLSE/LASSO	10.00	6.67	4.00	2.85	6.67	5.00	3.33	2.50
PTE ($\alpha = 0.15$)	3.20	2.84	2.31	1.95	2.50	2.27	1.93	1.68
PTE ($\alpha = 0.2$)	2.70	2.45	2.07	1.80	2.17	2.01	1.76	1.56
PTE ($\alpha = 0.25$)	2.35	2.17	1.89	1.67	1.94	1.82	1.63	1.47
JSE	5.00	4.00	2.86	2.22	4.13	3.42	2.56	2.04
PRSE	6.28	4.77	3.22	2.43	4.58	3.72	2.71	2.13
Ridge	10.00	6.67	4.00	2.86	6.78	5.07	3.37	2.52
$p = 20$								
$\Delta^2 = 5$				$\Delta^2 = 10$				
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RLSE/LASSO	2.86	2.50	2.00	1.67	1.67	1.54	1.33	1.18
PTE ($\alpha = 0.15$)	1.42	1.36	1.25	1.17	1.08	1.06	1.02	0.99
PTE ($\alpha = 0.2$)	1.33	1.29	1.20	1.14	1.06	1.04	1.02	0.99
PTE ($\alpha = 0.25$)	1.27	1.23	1.17	1.11	1.04	1.03	1.01	0.99
JSE	2.65	2.36	1.94	1.65	2.03	1.87	1.62	1.43
PRSE	2.63	2.34	1.92	1.64	2.01	1.85	1.60	1.42
Ridge	3.38	2.91	2.28	1.88	2.37	2.15	1.82	1.58
$p = 20$								
$\Delta^2 = 20$				$\Delta^2 = 60$				
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RLSE/LASSO	0.91	0.87	0.80	0.74	0.32	0.32	0.31	0.30
PTE ($\alpha = 0.15$)	0.97	0.97	0.97	0.97	1.00	1.00	1.00	1.00
PTE ($\alpha = 0.2$)	0.98	0.98	0.98	0.98	1.00	1.00	1.00	1.00
PTE ($\alpha = 0.25$)	0.99	0.98	0.98	0.99	1.00	1.00	1.00	1.00
JSE	1.58	1.51	1.36	1.26	1.21	1.18	1.13	1.09
PRSE	1.58	1.50	1.36	1.25	1.21	1.18	1.13	1.09
Ridge	1.74	1.64	1.47	1.34	1.26	1.23	1.18	1.13

Table 5: RWRE of the estimators for $p \in \{40, 60\}$ and different Δ^2 -value for varying p_1

Estimators	$p = 40$							
	$\Delta^2 = 0$				$\Delta^2 = 1$			
	$p_1 = 2$	$p_1 = 3$	$p_1 = 5$	$p_1 = 7$	$p_1 = 2$	$p_1 = 3$	$p_1 = 5$	$p_1 = 7$
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RLSE/LASSO	20.00	13.33	8.00	5.71	13.33	10.00	6.67	5.00
PTE ($\alpha = 0.15$)	4.05	3.74	3.24	2.86	3.32	3.12	2.77	2.49
PTE ($\alpha = 0.2$)	3.29	3.09	2.76	2.50	2.77	2.64	2.40	2.20
PTE ($\alpha = 0.25$)	2.78	2.65	2.42	2.23	2.40	2.30	2.13	1.98
JSE	10.00	8.00	5.71	4.44	8.12	6.75	5.05	4.03
PRSE	12.80	9.69	6.52	4.92	9.25	7.51	5.45	4.28
Ridge	20.00	13.33	8.00	5.71	13.45	10.07	6.70	5.02
$p = 40$								
$\Delta^2 = 5$				$\Delta^2 = 10$				
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RLSE/LASSO	5.71	5.00	4.00	3.33	3.33	3.08	2.67	2.35
PTE ($\alpha = 0.15$)	1.9641	1.8968	1.7758	1.6701	1.3792	1.3530	1.3044	1.2602
PTE ($\alpha = 0.2$)	1.75	1.70	1.61	1.53	1.29	1.27	1.24	1.20
PTE ($\alpha = 0.25$)	1.60	1.56	1.50	1.44	1.23	1.22	1.19	1.16
JSE	4.87	4.35	3.59	3.05	3.46	3.20	2.78	2.46
PRSE	4.88	4.36	3.59	3.05	3.42	3.16	2.75	2.44
Ridge	6.23	5.40	4.27	3.53	4.03	3.68	3.13	2.72
$p = 40$								
$\Delta^2 = 20$				$\Delta^2 = 60$				
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RLSE/LASSO	1.82	1.74	1.60	1.48	0.64	0.63	0.61	0.60
PTE ($\alpha = 0.15$)	1.05	1.05	1.03	1.02	0.99	0.99	0.99	0.99
PTE ($\alpha = 0.2$)	1.04	1.03	1.02	1.02	0.99	0.99	0.99	0.99
PTE ($\alpha = 0.25$)	1.03	1.02	1.02	1.01	0.99	0.99	1.00	1.00
JSE	2.41	2.2946	2.09	1.92	1.52	1.48	1.42	1.36
PRSE	2.41	2.29	2.08	1.91	1.52	1.48	1.42	1.36
Ridge	2.65	2.50	2.26	2.06	1.58	1.54	1.47	1.41
$p = 60$								
$\Delta^2 = 0$				$\Delta^2 = 1$				
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RLSE/LASSO	30.00	20.00	12.00	8.57	20.00	15.00	10.00	7.50
PTE ($\alpha = 0.15$)	4.49	4.23	3.79	3.43	3.80	3.62	3.29	3.02
PTE ($\alpha = 0.2$)	3.58	3.42	3.14	2.91	3.10	2.99	2.78	2.59
PTE ($\alpha = 0.25$)	2.99	2.89	2.70	2.54	2.64	2.56	2.42	2.29
JSE	15.00	12.00	8.57	6.67	12.12	10.09	7.55	6.03
PRSE	19.35	14.63	9.83	7.40	13.99	11.34	8.22	6.45
Ridge	30.00	20.00	12.00	8.57	20.11	15.06	10.03	7.52
$p = 60$								
$\Delta^2 = 5$				$\Delta^2 = 10$				
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RLSE/LASSO	8.57	7.50	6.00	5.00	5.00	4.61	4.0000	3.53
PTE ($\alpha = 0.15$)	2.35	2.28	2.16	2.05	1.63	1.60	1.55	1.50
PTE ($\alpha = 0.2$)	2.04	1.99	1.91	1.83	1.49	1.47	1.43	1.39
PTE ($\alpha = 0.25$)	1.83	1.79	1.73	1.67	1.39	1.37	1.34	1.31
JSE	7.10	6.35	5.25	4.47	4.89	4.53	3.94	3.50
PRSE	7.17	6.41	5.28	4.50	4.84	4.48	3.91	3.47
Ridge	9.09	7.90	6.26	5.19	5.70	5.21	4.45	3.89
$p = 60$								
$\Delta^2 = 20$				$\Delta^2 = 60$				
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RLSE/LASSO	2.73	2.61	2.40	2.22	0.97	0.95	0.92	0.89
PTE ($\alpha = 0.15$)	1.15	1.14	1.13	1.11	0.99	0.99	0.99	0.99
PTE ($\alpha = 0.2$)	1.11	1.10	1.09	1.08	0.99	0.99	0.99	0.99
PTE ($\alpha = 0.25$)	1.08	1.08	1.07	1.06	0.99	0.99	0.99	0.99
JSE	3.25	3.09	2.82	2.60	1.83	1.79	1.72	1.65
PRSE	3.23	3.08	2.81	2.59	1.83	1.79	1.72	1.65
Ridge	3.55	3.37	3.05	2.79	1.90	1.86	1.78	1.71

- Fan, J. and Li, R. (2001) "Variable selection via non-concave penalized likelihood and oracle properties," *Journal of the American Statistical Association*, 96, 1348–1360.
- Hansen, B. E. (2016), "The risk of James-Stein and LASSO Shrinkage," *Econometric Reviews*, 35(8-10), 1456–1470.
- Hoerl A. E., Kennard. R. W. (1970), "Ridge regression biased estimation for non-orthogonal problems," *Thechnometrics*, 12,69–89.
- Saleh, A. K. Md. E. (2006), "Theory of Preliminary Test and Stein-Type Estimation with Applications," Wiley; USA.
- Tibshirani, R. (1996), "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society, Series B*, 58(1), 267–288.
- Zou, H. (2006), "The adaptive LASSO and its oracle properties," *Journal of the American Statistical Association*, 101, 1418–1429.
- Zou, H. and Hastie, T. (2005), "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B*, 63, 301–320.

Table 6: RWRE values of estimators for $p_1 \in \{5, 7\}$, different values of p_2 and Δ^2

(p_1, p_2, Δ^2)	LSE	RLSE/LASSO	PTE			JSE	PRSE	Ridge
			$\alpha = 0.15$	$\alpha = 0.2$	$\alpha = 0.25$			
(5, 5)	1.00	2.00	1.76	1.51	1.36	1.43	1.56	2.00
(5, 15, 0)	1.00	4.00	3.11	2.31	1.89	2.86	3.22	4.00
(5, 25, 0)	1.00	6.00	4.23	2.84	2.20	4.28	4.87	6.00
(5, 35, 0)	1.00	8.00	5.18	3.24	2.42	5.71	6.52	8.00
(5, 55, 0)	1.00	12.00	6.71	3.79	2.70	8.57	9.83	12.00
(5, 5, .5)	1.00	1.82	1.58	1.37	1.26	1.37	1.46	1.83
(5, 15, .5)	1.00	3.64	2.79	2.10	1.74	2.70	2.93	3.65
(5, 25, .5)	1.00	5.45	3.81	2.61	2.05	4.03	4.43	5.46
(5, 35, .5)	1.00	7.27	4.68	2.98	2.26	5.36	5.93	7.28
(5, 55, .5)	1.00	10.91	6.11	3.52	2.55	8.02	8.94	10.92
(5, 5, 1)	1.00	1.67	1.43	1.27	1.18	1.33	1.38	1.71
(5, 15, 1)	1.00	3.33	2.53	1.93	1.63	2.56	2.71	3.37
(5, 25, 1)	1.00	5.00	3.46	2.41	1.92	3.80	4.08	5.03
(5, 35, 1)	1.00	6.67	4.27	2.77	2.13	5.05	5.45	6.70
(5, 55, 1)	1.00	10.00	5.61	3.29	2.42	7.55	8.22	10.03
(5, 5, 5)	1.0000	1.00	0.93	0.95	0.96	1.15	1.15	1.33
(5, 15, 5)	1.00	2.00	1.47	1.26	1.17	1.94	1.92	2.28
(5, 25, 5)	1.00	3.00	1.98	1.54	1.35	2.76	2.75	3.27
(5, 35, 5)	1.00	4.00	2.44	1.77	1.50	3.59	3.59	4.27
(5, 55, 5)	1.00	6.00	3.27	2.16	1.73	5.25	5.28	6.26
(7, 5, 0)	1.00	1.43	1.33	1.23	1.16	1.11	1.16	1.43
(7, 15, 0)	1.00	2.86	2.41	1.94	1.67	2.22	2.43	2.86
(7, 25, 0)	1.00	4.28	3.35	2.46	2.00	3.33	3.67	4.28
(7, 35, 0)	1.00	5.71	4.17	2.86	2.23	4.44	4.92	5.71
(7, 55, 0)	1.00	8.57	5.54	3.43	2.53	6.67	7.40	8.57
(7, 5, .5)	1.00	1.33	1.23	1.15	1.10	1.09	1.13	1.35
(7, 15, .5)	1.00	2.67	2.22	1.80	1.56	2.12	2.27	2.67
(7, 25, .5)	1.00	4.00	3.08	2.29	1.87	3.17	3.41	4.00
(7, 35, .5)	1.00	5.33	3.84	2.66	2.10	4.23	4.57	5.34
(7, 55, .5)	1.00	8.00	5.13	3.21	2.40	6.33	6.89	8.00
(7, 5, 1)	1.00	1.25	1.15	1.09	1.06	1.08	1.10	1.29
(7, 15, 1)	1.00	2.50	2.05	1.68	1.47	2.04	2.13	2.52
(7, 25, 1)	1.00	3.75	2.85	2.13	1.77	3.03	3.20	3.77
(7, 35, 1)	1.00	5.00	3.56	2.49	1.98	4.03	4.28	5.01
(7, 55, 1)	1.00	7.50	4.77	3.02	2.29	6.03	6.45	7.52
(7, 5, 5)	1.00	0.83	0.87	0.92	0.94	1.03	1.03	1.13
(7, 15, 5)	1.00	1.67	1.32	1.17	1.11	1.65	1.64	1.88
(7, 25, 5)	1.00	2.50	1.78	1.44	1.29	2.34	2.34	2.70
(7, 35, 5)	1.00	3.33	2.20	1.67	1.44	3.05	3.05	3.53
(7, 55, 5)	1.00	5.00	2.98	2.05	1.67	4.47	4.50	5.19

Table 7: RWRE values of estimators for $p_2 \in \{5, 7\}$, different values of p_1 and Δ^2

(p_1, p_2, Δ^2)	LSE	RLSE/LASSO	PTE			JSE	PRSE	Ridge
			$\alpha = 0.15$	$\alpha = 0.2$	$\alpha = 0.25$			
(5, 5, 0)	1.00	2.00	1.76	1.51	1.36	1.43	1.56	2.00
(15, 5, 0)	1.00	1.33	1.27	1.20	1.15	1.18	1.22	1.33
(25, 5, 0)	1.00	1.20	1.17	1.127	1.10	1.11	1.14	1.20
(35, 5, 0)	1.00	1.14	1.12	1.09	1.07	1.08	1.10	1.14
(55, 5, 0)	1.00	1.09	1.08	1.06	1.04	1.05	1.06	1.09
(5, 5, 0.5)	1.00	1.82	1.58	1.37	1.26	1.34	1.46	1.83
(15, 5, 0.5)	1.00	1.29	1.22	1.16	1.11	1.16	1.19	1.29
(25, 5, 0.5)	1.00	1.18	1.14	1.10	1.07	1.10	1.12	1.18
(35, 5, 0.5)	1.00	1.13	1.10	1.07	1.05	1.07	1.08	1.13
(55, 5, 0.5)	1.00	1.08	1.06	1.05	1.03	1.05	1.05	1.08
(5, 5, 1)	1.00	1.67	1.43	1.27	1.18	1.33	1.38	1.71
(15, 5, 1)	1.00	1.25	1.18	1.12	1.08	1.14	1.16	1.26
(25, 5, 1)	1.00	1.15	1.11	1.08	1.05	1.09	1.10	1.16
(35, 5, 1)	1.00	1.11	1.08	1.06	1.04	1.07	1.07	1.12
(55, 5, 1)	1.00	1.07	1.05	1.04	1.03	1.04	1.05	1.07
(5, 5, 5)	1.00	1.00	0.93	0.95	0.96	1.15	1.15	1.33
(15, 5, 5)	1.00	1.00	0.97	0.97	0.98	1.07	1.07	1.14
(25, 5, 5)	1.00	1.00	0.98	0.98	0.98	1.05	1.04	1.09
(35, 5, 5)	1.00	1.00	0.98	0.99	0.99	1.03	1.03	1.07
(55, 5, 5)	1.00	1.00	0.99	0.99	0.99	1.02	1.02	1.04
(3, 7, 0)	1.00	3.33	2.60	1.98	1.66	2.00	2.31	3.33
(13, 7, 0)	1.00	1.54	1.44	1.33	1.24	1.33	1.40	1.54
(23, 7, 0)	1.00	1.30	1.26	1.20	1.15	1.20	1.23	1.30
(33, 7, 0)	1.00	1.21	1.18	1.14	1.11	1.14	1.16	1.21
(53, 7, 0)	1.00	1.13	1.11	1.09	1.07	1.09	1.10	1.13
(3, 7, 0.5)	1.00	2.86	2.21	1.73	1.49	1.87	2.06	2.88
(13, 7, 0.5)	1.00	1.48	1.38	1.27	1.20	1.30	1.35	1.48
(23, 7, 0.5)	1.00	1.28	1.22	1.16	1.12	1.18	1.20	1.28
(33, 7, 0.5)	1.00	1.19	1.16	1.12	1.09	1.13	1.15	1.19
(53, 7, 0.5)	1.00	1.12	1.10	1.07	1.06	1.08	1.09	1.12
(3, 7, 1)	1.00	2.50	1.93	1.55	1.37	1.77	1.88	2.58
(13, 7, 1)	1.00	1.43	1.32	1.22	1.16	1.28	1.31	1.44
(23, 7, 1)	1.00	1.25	1.19	1.13	1.10	1.17	1.18	1.26
(33, 7, 1)	1.00	1.18	1.14	1.10	1.07	1.12	1.13	1.18
(53, 7, 1)	1.00	1.11	1.09	1.06	1.05	1.08	1.08	1.11
(3, 7, 5)	1.00	1.25	1.04	1.01	0.99	1.38	1.372	1.69
(13, 7, 5)	1.00	1.11	1.02	1.00	0.99	1.16	1.15	1.26
(23, 7, 5)	1.00	1.07	1.01	1.00	0.99	1.10	1.10	1.16
(33, 7, 5)	1.00	1.05	1.01	1.00	0.99	1.07	1.07	1.11
(53, 7, 5)	1.00	1.03	1.01	1.00	0.99	1.05	1.05	1.07

Table 8: Sample efficiency table of estimators under Hansens method

p	Δ^2	LSE	RLSE/LASSO	PTE			JSE	PRSE	Ridge
				($\alpha = 0.15$)	($\alpha = 0.2$)	($\alpha = 0.25$)			
10	0	1.00	∞	2.68	3.38	4.03	5.00	7.03	∞
	0.1	1.00	100.00	2.84	3.54	4.18	4.81	6.56	101.00
	0.5	1.00	20.00	3.46	4.15	4.78	4.19	5.23	21.00
	1	1.00	10.00	4.20	4.87	5.47	3.65	4.25	11.00
	2	1.00	5.00	5.56	6.17	6.69	2.97	3.21	6.00
	5	1.00	2.00	8.59	8.90	9.13	2.09	2.12	3.00
	10	1.00	1.00	10.63	10.52	10.43	1.61	1.61	2.00
	20	1.00	0.50	10.43	10.29	10.20	1.32	1.32	1.50
	50	1.00	0.20	10.00	10.00	10.00	1.13	1.13	1.20
	100	1.00	0.10	10.00	10.00	10.00	1.04	1.04	1.10
20	0	1.00	∞	4.62	5.91	7.13	10.00	15.04	∞
	0.1	1.00	200.00	4.80	6.10	7.32	9.57	13.90	201.00
	0.5	1.00	40.00	5.53	6.84	8.07	8.20	10.73	41.00
	1	1.00	20.00	6.44	7.76	8.97	6.99	8.46	21.00
	2	1.00	10.00	8.19	9.50	10.68	5.48	6.09	11.00
	5	1.00	4.00	12.80	13.91	14.82	3.54	3.63	5.00
	10	1.00	2.00	17.80	18.31	18.68	2.46	2.47	3.00
	20	1.00	1.00	20.45	20.34	20.26	1.78	1.78	2.00
	50	1.00	0.40	20.01	20.00	20.00	1.32	1.32	1.40
	100	1.00	0.20	20.00	20.00	20.00	1.09	1.09	1.20
30	0	1.00	∞	6.45	8.31	10.08	15.00	23.41	∞
	0.1	1.00	300.00	6.66	8.52	10.30	14.33	21.55	301.00
	0.5	1.00	60.00	7.46	9.36	11.15	12.20	16.45	61.00
	1	1.00	30.00	8.47	10.40	12.20	10.33	12.80	31.00
	2	1.00	15.00	10.46	12.43	14.23	7.99	9.04	16.00
	5	1.00	6.00	16.06	17.94	19.54	4.97	5.15	7.00
	10	1.00	3.00	23.29	24.58	25.57	3.30	3.32	4.00
	20	1.00	1.50	29.49	29.66	29.77	2.24	2.24	2.50
	50	1.00	0.60	30.02	30.01	30.01	1.52	1.52	1.60
	100	1.00	0.30	30.00	30.00	30.00	1.14	1.14	1.30
40	0	1.00	∞	8.23	10.65	12.97	20.00	31.99	∞
	0.1	1.00	400.00	8.45	10.88	13.20	19.10	29.39	401.00
	0.5	1.00	80.00	9.32	11.79	14.14	16.20	22.28	81.00
	1	1.00	40.00	10.41	12.93	15.31	13.66	17.23	41.00
	2	1.00	20.00	12.59	15.18	17.58	10.49	12.03	21.00
	5	1.00	8.00	18.91	21.52	23.80	6.41	6.68	9.00
	10	1.00	4.00	27.86	29.96	31.64	4.14	4.18	5.00
	20	1.00	2.00	37.57	38.23	38.68	2.70	2.70	3.00
	50	1.00	0.80	40.04	40.03	40.02	1.71	1.71	1.80
	100	1.00	0.40	40.00	40.00	40.00	1.18	1.18	1.40

A Appendix

We now apply Hansen's method for ANOVA when it is suspected that $\boldsymbol{\theta}$ may be $\mathbf{0}$. In this case, the estimators are: LSE of $\boldsymbol{\theta}$ is $\tilde{\boldsymbol{\theta}}_n$; RLSE ($\hat{\boldsymbol{\theta}}_n$) and LASSO ($\hat{\boldsymbol{\theta}}_n^L$) as $\mathbf{0}$ vector; PTE of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_n^{PT} = \tilde{\boldsymbol{\theta}}_n I(\mathcal{L}_n \geq c_\alpha)$; JSE of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_n^{JS} = \tilde{\boldsymbol{\theta}}_n(1 - (p-2)\mathcal{L}_n^{-1})$; PRSE of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_n^{S+} = \tilde{\boldsymbol{\theta}}_n(1 - (p-2)\mathcal{L}_n^{-1})^{-1}$, ridge estimator of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_n^{ridge}(k) = \frac{1}{1+k}\tilde{\boldsymbol{\theta}}_n$.

Accordingly, the L_2 -risks are

$$\begin{aligned} R(\tilde{\boldsymbol{\theta}}_n, \mathbf{N}) &= \sigma^2 p, \\ R(\hat{\boldsymbol{\theta}}_n, \mathbf{N}) &= \sigma^2 \Delta^{*2}, \quad \Delta^{*2} = \frac{1}{\sigma^2} \boldsymbol{\theta}^T \mathbf{N} \boldsymbol{\theta} \\ R(\hat{\boldsymbol{\theta}}_n^L, \mathbf{N}) &= \sigma^2 \Delta^{*2}, \\ R(\hat{\boldsymbol{\theta}}_n^{PT}, \mathbf{N}) &= \sigma^2 p (1 - H_{p+2}(c_\alpha, \Delta^{*2})) + \sigma^2 \Delta^2 \{2H_{p+2}(c_\alpha, \Delta^{*2}) - H_{p+4}(c_\alpha, \Delta^{*2})\} \\ R(\hat{\boldsymbol{\theta}}_n^{JS}, \mathbf{N}) &= \sigma^2 (p - (p-2)^2 E[\chi_{p_2+2}^{-2}(\Delta^{*2})]) \\ R(\hat{\boldsymbol{\theta}}_n^{S+}, \mathbf{N}) &= R(\hat{\boldsymbol{\theta}}_n^{JS}, \mathbf{N}) - \sigma^2 p E \left[(1 - (p-2)\chi_{p_2+2}^{-2}(\Delta^{*2}))^2 I(\chi_{p_2+2}^2(\Delta^{*2}) \leq p-2) \right] \\ &\quad + \sigma^2 \Delta^{*2} \{2E[(1 - (p-2)\chi_{p_2+2}^{-2}(\Delta^{*2})) I(\chi_{p_2+2}^2(\Delta^{*2}) \leq p-2)] \\ &\quad - E[(1 - (p-2)\chi_{p_2+4}^{-2}(\Delta^{*2}))^2 I(\chi_{p_2+4}^2(\Delta^{*2}) \leq p-2)]\} \\ R(\hat{\boldsymbol{\theta}}_n^{RR}, \mathbf{N}) &= \frac{p\Delta^{*2}}{p + \Delta^{*2}} \end{aligned}$$

The efficiency table is given in Table 8.

Note that the LASSO solution always puts some (say p_2) parameters equal to 0 and others (p_1) as $\tilde{\theta}_{jn}$. Then, the oracle solution of the risk is $\sigma^2(p_1 + \Delta^2)$ where $\Delta^{*2} > \Delta^2$ so that efficiency is $p/(p_1 + \Delta^2)$. This happens under our assumptions and not under Hansen's assumption. We get Δ^{*2} corresponding to the infeasible estimator, $\mathbf{0}$. Our assumptions are the right for the study of LASSO and related.

Received: May 8, 2017

Accepted: January 6, 2018