STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **6**, September 2018, pp 427–451.
Published online in International Academic Press (www.IAPress.org)

On Shrinkage Estimation: Non-orthogonal Case

A.K.Md.Ehsanes Saleh [1]*, M. Norouzirad[2]

[1] *School of Mathematics and Statistics, Carleton University, Ottawa, Canada*
[2] *Department of Statistics, Faculty of Mathematical Sciences, Shahrood University of Technology, Shahrood, Iran*

Abstract In this paper, we consider the estimation of the parameters of the non-orthogonal regression model, when we suspect a sparsity condition. We provide with a comparative performance characteristics of the primary penalty estimators, namely, the ridge and the LASSO, with the least square estimator, restricted LSE, preliminary test and Stein-type of estimators, when the dimension of the parameter space is less than the dimension of the sample space. Using the principle of marginal distribution theory, the analysis of risks leads to the following conclusions: (i) ridge estimator outperforms least squares, preliminary test and Stein-type estimators uniformly, (ii) The restricted least squares estimator and LASSO are competitive, although LASSO lags behind the restricted least squares estimator uniformly. Both estimators outperform the least squares, preliminary test, and Stein-type estimators in a subspace, respectively. (iii) The lower bound risk expression of LASSO does not depend on the threshold parameter. (iv) Performance of the estimators depends upon the size of numbers of active coefficients, non-active coefficients, and the divergence parameter. In support of our conclusion, we prepare some tables and graphs relevant to the properties of the estimators.

Keywords Dominance; Efficiency; LASSO; PTE and Stein-type estimators; Penalty estimator; Risk function; Sparsity.

AMS 2010 subject classifications 62J07, 62H12

DOI: 10.19139/soic.v6i3.582

1. Introduction

Traditionally, we use *least squares estimators* (LSEs) for a linear model which provide *minimum variance unbiased estimators*. However, data analysts point out two deficiencies of LSEs, namely, the *prediction accuracy* and the *interpretation*. To overcome these concerns, Tibshirani [12] proposed a popular and exciting estimator called the *least absolute shrinkage and selection operator* (LASSO). It defines a continuous shrinking operation that can produce coefficients that are exactly zero and is competitive with *subset selection* and *ridge regression* estimators retaining good properties of both the estimators. The LASSO simultaneously estimates and selects the coefficients of a given linear model.

There are many shrinkage estimators, namely, the preliminary test and Stein-type estimators in the literature. These estimators only shrink toward the target value and do not select coefficients for appropriate prediction and interpretation.

Hoerl and Kennard [8] introduced the *ridge regression*, which opened the door for *penalty estimators* based on Tikhonov [13] regularization. The methodology is a minimization of least squares criterion subject to L_2 penalty. This procedure does not produce a sparse solution. However, LASSO is related to the estimators, such as non-negative garrote by Breiman [2], smoothly clipped absolute derivation (SCAD) by Fan and Li [5], elastic net by Zou and Hastie [15], adaptive LASSO by Zou [14], hard threshold LASSO by Belloni and Chernophukov [1] and

*Correspondence to: A.K.Md.Ehsanes Saleh, Email: esaleh@math.carleton.ca

ISSN 2310-5070 (online) ISSN 2311-004X (print)

many other versions. A general form of an extension of LASSO-type estimation called *the bridge estimation*, by Frank and Friedman [6], is worth pursuing.

This paper is devoted to the comparative study of the finite sample performance of the primary penalty estimators, namely, LASSO and the ridge regression estimators. They are compared to the LSE, restricted LSE (RLSE), preliminary test estimator (PTE) and Stein-type Shrinkage estimators, James-Stein estimator (JSE) and positive rule Stein estimator (PRSE). The question of comparison between the ridge regression (first discovery of penalty estimator) and Stein-type estimator is well known and is established by Draper and Craig [3] among others. So far, literature is full of scattered simulated results without any theoretical backups, and definite conclusions are not available whether the design matrix is orthogonal or non-orthogonal.

This paper points to the useful aspects of LASSO and ridge regression estimators, as well as limitations, as found in other articles. Conclusions are obtained based on the lower bound of L_2-risk expression for the LASSO estimator provided by Donoho and Johnstone [4]. The comparison of these estimators discussed here are based on mathematical analysis supported by tables of relative weighted L_2-risk efficiency (RWRE) and graphs.

In his pioneering paper, Tibshirani [12] examined the relative efficiency of the subset selection, ridge regression and the LASSO in three different scenarios, under orthogonality of the design matrix:

(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.

(c) Large number of small coefficients - ridge regression does best by a good margin, followed by LASSO and then subset selection.

The above results refer to *prediction accuracy*. Recently, Hansen [7], under the same assumption, considered the comparison of LASSO, Stein-type estimators and subset selection based on L_2-risk. His findings may be summarized as follows:

(i) Neither LASSO nor Stein-type estimator uniformly dominates one other.

(ii) 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.

Hansen [7] 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.

In our study, we discovered the following conclusions:

(i) Ridge estimator outperforms the LSE, PTE and Stein-type estimator,

(ii) The restricted estimator and LASSO are competitive, although LASSO, lags behind RLSE uniformly. Both estimators outperform the LSE, PTE, JSE and PRSE in a sub-interval of $(0, p_2 + \mathrm{tr}(\boldsymbol{M}_0))$, $\boldsymbol{M}_0 = \boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21}$ and $(0, p_2)$, respectively. As a consequence, a necessary condition for LASSO to satisfy "oracle properties" is $\mathrm{tr}(\boldsymbol{M}_0) = 0$,

(iii) Lower-bound L_2-risk of LASSO is independent of the threshold parameter. In support of our conclusion, we provided mathematical analysis, tables and graphs related our problem.

The organization of the paper is as follows: Section 2 discusses various estimators and their L_2-risk expressions using the general linear regression model. Section 3 discusses the bias and weighted L_2-risk of the estimators and multivariate normal decision theory and oracle for diagonal linear projection and related bounds of penalty estimators. Section 4 deals with the comparison of estimators and finally, section 5 presents the summary and concluding remarks along with a discussion of supporting table and graphs in our study.

2. Linear Model and the Estimators

Consider the multiple linear model,

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{X}$ is the design matrix such that $\boldsymbol{C}_n = \boldsymbol{X}^\top \boldsymbol{X}$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$, $\boldsymbol{Y} = (Y_1, \dots, Y_n)^\top$ is the response vector, $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\top$ is the n-vector of errors such that $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\top) = \sigma^2 \boldsymbol{I}_n$, σ^2 is known variance of any ϵ_i ($i = 1, \dots, n$).

It is well-known that the LSE of $\boldsymbol{\beta}$, say, $\tilde{\boldsymbol{\beta}}_n = \boldsymbol{C}_n^{-1}\boldsymbol{X}^\top \boldsymbol{Y}$, has the distribution

$$\sqrt{n}\left(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right) \sim \mathcal{N}_p(\boldsymbol{0}, \sigma^2 \boldsymbol{C}^{-1}), \quad \boldsymbol{C}^{-1} = (C^{ij}),\ i, j = 1, \dots, p. \tag{2}$$

We designate $\tilde{\boldsymbol{\beta}}_n$ as the unrestricted estimator (LSE) of $\boldsymbol{\beta}$.

In many situations, a sparse model is desired such as high-dimensional settings. Under the sparsity assumption, we partition the coefficient vector and the design matrix as

$$\boldsymbol{\beta} = \underset{p_1 \times 1}{\left(\boldsymbol{\beta}_1^\top,\right.} \ \underset{p_2 \times 1}{\left.\boldsymbol{\beta}_2^\top\right)^\top}, \qquad \boldsymbol{X} = \underset{n \times p_1}{(\boldsymbol{X}_1,} \ \underset{n \times p_2}{\boldsymbol{X}_2)}. \tag{3}$$

where $p = p_1 + p_2$.

So that (1) may also be written as

$$\boldsymbol{Y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}, \tag{4}$$

where $\boldsymbol{\beta}_1$ may stand for the main effects and $\boldsymbol{\beta}_2$ for the interaction which may be insignificant, though one is interested in the estimation and selection of main effects. Thus, the problem of estimating $\boldsymbol{\beta}$ is reduced to the estimation of $\boldsymbol{\beta}_1$ when $\boldsymbol{\beta}_2$ is suspected to be equal to $\boldsymbol{0}$. Under this setup, the LSE of $\boldsymbol{\beta}$ is

$$\tilde{\boldsymbol{\beta}}_n = \begin{pmatrix} \tilde{\boldsymbol{\beta}}_{1n} \\ \tilde{\boldsymbol{\beta}}_{2n} \end{pmatrix}, \tag{5}$$

and if $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{0}^\top)^\top$, it is

$$\hat{\boldsymbol{\beta}}_n = \begin{pmatrix} \hat{\boldsymbol{\beta}}_{1n} \\ \boldsymbol{0} \end{pmatrix}, \tag{6}$$

where $\hat{\boldsymbol{\beta}}_{1n} = (\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top \boldsymbol{Y}$.

In this research article, we are interested in the study of some shrinkage estimators steaming from the LSE and RLSE of $\boldsymbol{\beta}$ in a sparse regression model. To study the characteristic properties of various estimators, we use the two component weighted L_2-risk function

$$\mathrm{R}(\boldsymbol{\beta}_n^*; \boldsymbol{W}_1, \boldsymbol{W}_2) = E(\hat{\boldsymbol{\beta}}_{1n}^* - \boldsymbol{\beta}_1)^\top \boldsymbol{W}_1(\hat{\boldsymbol{\beta}}_{1n}^* - \boldsymbol{\beta}_1) + E(\hat{\boldsymbol{\beta}}_{2n}^* - \boldsymbol{\beta}_2)^\top \boldsymbol{W}_2(\hat{\boldsymbol{\beta}}_{2n}^* - \boldsymbol{\beta}_2), \tag{7}$$

where $\hat{\boldsymbol{\beta}}_n^*$ is any estimator of $\boldsymbol{\beta}$.

Note that the marginal distribution of $\tilde{\boldsymbol{\beta}}_{1n}$ is $\mathcal{N}_{p_1}(\boldsymbol{\beta}_1, \sigma^2 \boldsymbol{C}_{11.2}^{-1})$ and that of $\tilde{\boldsymbol{\beta}}_{2n}$ is $\mathcal{N}_{p_2}(\boldsymbol{\beta}_2, \sigma^2 \boldsymbol{C}_{22.1}^{-1})$ where $\boldsymbol{C}_{ii.j} = \boldsymbol{C}_{ii} - \boldsymbol{C}_{ij}\boldsymbol{C}_{jj}^{-1}\boldsymbol{C}_{ji}$. Hence, the weighted L_2-risk of $\tilde{\boldsymbol{\beta}}_n = (\tilde{\boldsymbol{\beta}}_{1n}^\top, \tilde{\boldsymbol{\beta}}_{2n}^\top)^\top$ is given by

$$\mathrm{R}(\tilde{\boldsymbol{\beta}}_n; \boldsymbol{C}_{11.2}, \boldsymbol{C}_{22.1}) = \sigma^2(p_1 + p_2). \tag{8}$$

Similarly, the weighted L_2-risk of $\hat{\boldsymbol{\beta}}_n = (\hat{\boldsymbol{\beta}}_{1n}^\top, \boldsymbol{0}^\top)^\top$ is given by

$$\begin{aligned} \mathrm{R}(\hat{\boldsymbol{\beta}}_n; \boldsymbol{C}_{11.2}, \boldsymbol{C}_{22.1}) &= \sigma^2 \left(\mathrm{tr}\left(\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{11.2}\right) + \frac{1}{\sigma^2}\boldsymbol{\beta}_2^\top \boldsymbol{C}_{22.1}\boldsymbol{\beta}_2 \right) \\ &= \sigma^2 \left(\mathrm{tr}\left(\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{11.2}\right) + \Delta^2 \right), \quad \Delta^2 = \frac{1}{\sigma^2}\boldsymbol{\beta}_2^\top \boldsymbol{C}_{22.1}\boldsymbol{\beta}_2, \end{aligned} \tag{9}$$

since the covariance matrix of $(\hat{\boldsymbol{\beta}}_{1n}^{\top}, \mathbf{0}^{\top})^{\top}$ is $(\sigma^2 \boldsymbol{C}_{11}^{-1}, -\boldsymbol{\beta}_2\boldsymbol{\beta}_2^{\top})^{\top}$ and computation of (7) with $\boldsymbol{W}_1 = \boldsymbol{C}_{11.2}$ and $\boldsymbol{W}_2 = \boldsymbol{C}_{22.1}$ yields the result (9).

Our focus on this paper is the comparative study of the performance properties of three penalty estimators compared to the preliminary test and Stein-type estimators. We refer to Saleh [10] for the comparative study of preliminary test and Stein-type estimators, when the design matrix is non-orthogonal. We extend the study to include the penalty estimators, which has not been theoretically done yet, except for simulation studies.

2.1. Penalty Estimators

In this paper, we consider three basic penalty estimators, namely, the (i) hard threshold estimator (HTE) (Donoho and Johnstone, [4]), (ii) least absolute shrinkage and selection operator (LASSO) by Tibshirani [12] and the (iii) ridge regression estimator (RRE) by Hoerl and Kennard [8].

Motivated by the idea that only few regression coefficients contribute signal, we consider threshold rules that retain only observe data that exceed a multiple of the noise level. Accordingly, we consider the *subset selection* rule given by Donoho and Johnstone [4] known as *hard threshold* rule as given by

$$\begin{aligned} \hat{\boldsymbol{\beta}}_n^{\mathrm{HT}}(\kappa) &= \left(\tilde{\beta}_{jn} I(|\tilde{\beta}_{jn}| > \kappa\sigma\sqrt{C^{jj}})\Big| j = 1, \dots, p\right)^{\top}, \\ &= \left(\sigma\sqrt{C^{jj}} Z_j I(|Z_j| > \kappa) | j = 1, \dots, p\right)^{\top}, \end{aligned} \tag{10}$$

where $\tilde{\beta}_{jn}$ is the jth element of $\tilde{\boldsymbol{\beta}}_n$, $I(A)$ is an indicator function of the set A, and marginally

$$Z_j = \frac{\tilde{\beta}_{jn}}{\sigma\sqrt{C^{jj}}} \sim \mathcal{N}(\Delta_j, 1), \quad j = 1, \dots, p, \tag{11}$$

where $\Delta_j = \beta_j / \sigma\sqrt{C^{jj}}$.

Here, Z_j is the test statistic for testing the null-hypothesis $\mathcal{H}_o : \beta_j = 0$ versus $\mathcal{H}_A : \beta_j \neq 0$. The quantity κ is called the threshold parameter. The components of $\hat{\boldsymbol{\beta}}_n^{\mathrm{HT}}(\kappa)$ are kept as $\tilde{\beta}_{jn}$ if they are significant and zero, otherwise. It is apparent that each component of $\hat{\boldsymbol{\beta}}_n^{\mathrm{HT}}(\kappa)$ is a preliminary-test estimator (PTE) of the predictor concerned. The components of $\hat{\boldsymbol{\beta}}_n^{\mathrm{HT}}(\kappa)$ are PTEs and discrete variables and lose some optimality properties. Hence, one may define a continuous version of (10) based on marginal distribution of $\tilde{\beta}_{jn}$ $(j = 1, \dots, p)$.

In accordance with the principle of PTE approach (see Saleh [10]), we define the Stein-type estimator as the continuous version of PTE based on the marginal distribution of $\tilde{\beta}_{jn} \sim \mathcal{N}(\beta_j, \sigma^2 C^{jj})$, $j = 1, \dots, p$ given by

$$\begin{aligned} \hat{\boldsymbol{\beta}}_n^{\mathrm{S}}(\kappa) &= \left(\tilde{\beta}_{jn} - \kappa\sigma\sqrt{C^{jj}}\frac{\tilde{\beta}_{jn}}{|\tilde{\beta}_{jn}|}\Bigg| j = 1, \dots, p\right)^{\top} \\ &= \left(\sigma\sqrt{C^{jj}}\mathrm{sgn}(Z_j)\left(|Z_j| - \kappa\right)\Big| j = 1, 2, \dots, p\right)^{\top} \\ &= \left(\hat{\beta}_{1n}^{\mathrm{S}}(\kappa), \dots, \hat{\beta}_{pn}^{\mathrm{S}}(\kappa)\right)^{\top}. \end{aligned} \tag{12}$$

See Saleh [10, Pg. 83] for some details.

Another continuous version proposed by Tibshirani [12] and Donoho and Johnstone [4] is called the LASSO. In order to develop LASSO for our case, we propose the following modified LASSO (ML) given by

$$\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}(\kappa) = \left(\hat{\beta}_{1n}^{\mathrm{ML}}(\kappa), \dots, \hat{\beta}_{pn}^{\mathrm{ML}}(\kappa)\right) \tag{13}$$

where for $j = 1, 2, \dots, p$,

$$\hat{\beta}_{jn}^{\mathrm{ML}}(\kappa) = \sigma\sqrt{C^{jj}}\,\mathrm{sgn}(Z_j)\left(|Z_j| - \kappa\right)^{+}. \tag{14}$$

The estimator $\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}(\kappa)$ defines a continuous shrinkage operation that produces sparse solution.

The formula (14) is obtained as follows:
Differentiating $(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})+2\kappa\sigma\sum_{j=1}^{p}\left(\boldsymbol{C}^{jj}\right)^{-\frac{1}{2}}|\beta_j|$ where $\kappa>0$, we obtain the following equation

$$-2\tilde{\boldsymbol{\beta}}_n+2\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}(\kappa)+2\kappa\sigma\mathrm{diag}\left([\boldsymbol{C}^{11}]^{\frac{1}{2}},\dots,[\boldsymbol{C}^{pp}]^{\frac{1}{2}}\right)\mathrm{sgn}(\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}(\kappa))=\mathbf{0} \tag{15}$$

where $\mathrm{sgn}(\boldsymbol{\beta})=(\mathrm{sgn}(\beta_1),\dots,\mathrm{sgn}(\beta_p))^\top$ and $\boldsymbol{C}^{jj}$ is the jth diagonal element of $\boldsymbol{C}^{-1}$. Now, the jth marginal component of (15) is given by $\hat{\boldsymbol{\beta}}_{jn}^{\mathrm{ML}}(\kappa)$,

$$-\tilde{\boldsymbol{\beta}}_{jn}+\hat{\boldsymbol{\beta}}_{jn}^{\mathrm{ML}}(\kappa)+\kappa\sigma\sqrt{\boldsymbol{C}^{jj}}\mathrm{sgn}\left(\hat{\boldsymbol{\beta}}_{jn}^{\mathrm{ML}}(\kappa)\right)=\mathbf{0};\ j=1,\dots,p. \tag{16}$$

Now, we have two cases:

(i) $\mathrm{sgn}\left(\hat{\boldsymbol{\beta}}_{jn}^{\mathrm{ML}}(\kappa)\right)=+1$, then (16) reduces to

$$-Z_j+\left[\frac{\hat{\boldsymbol{\beta}}_{jn}^{\mathrm{ML}}(\kappa)}{\sigma\sqrt{\boldsymbol{C}^{jj}}}\right]+\kappa=0,\quad j=1,\dots,p. \tag{17}$$

where $Z_j=\frac{\tilde{\beta_{jn}}}{\sigma\sqrt{\boldsymbol{C}^{jj}}}$. Hence,

$$0<\hat{\boldsymbol{\beta}}_{jn}^{\mathrm{ML}}(\kappa)=\sigma\sqrt{\boldsymbol{C}^{jj}}\left(Z_j-\kappa\right)=\sigma\sqrt{\boldsymbol{C}^{jj}}\left(|Z_j|-\kappa\right), \tag{18}$$

with clearly $Z_j>0$ and $|Z_j|>\kappa$.

(ii) $\mathrm{sgn}\left(\hat{\boldsymbol{\beta}}_{jn}^{\mathrm{ML}}(\kappa)\right)=-1$, then we have

$$-Z_j+\left[\frac{\hat{\boldsymbol{\beta}}_{jn}^{\mathrm{ML}}(\kappa)}{\sigma\sqrt{\boldsymbol{C}^{jj}}}\right]-\kappa=0;\quad j=1,\dots,p. \tag{19}$$

Hence,

$$\begin{aligned}0>\frac{\hat{\boldsymbol{\beta}}_{jn}^{\mathrm{ML}}(\kappa)}{\sigma\sqrt{\boldsymbol{C}^{jj}}}&=Z_j+\kappa=-|Z_j|+\kappa\\ \hat{\boldsymbol{\beta}}_{jn}^{\mathrm{ML}}(\kappa)&=-(|Z_j|-\kappa)=-\sigma\sqrt{\boldsymbol{C}^{jj}}\left(|Z_j|-\kappa\right)\end{aligned} \tag{20}$$

with clearly $Z_j<0$ and $|Z_j|>\kappa$.

(iii) For $\hat{\boldsymbol{\beta}}_{jn}^{ML}(\kappa)=0$, we have $-Z_j+\kappa\gamma=0$ for some $\gamma\in[-1,1]$. Hence, we obtain $Z_j=\kappa\gamma$ which implies $|Z_j|\le\kappa$.

Combining (18), (20), and (iii), we obtain (14).

Finally, we consider the unrestricted *ridge* regression estimators of $(\boldsymbol{\beta}_1^\top,\boldsymbol{\beta}_2^\top)^\top$. They are obtained using marginal distributions of $\tilde{\beta}_{jn}\sim\mathcal{N}(\beta_j,\sigma^2C^{jj})$, $j=1,\dots,p$, as

$$\tilde{\boldsymbol{\beta}}_n^{\mathrm{RR}}(\kappa)=\begin{pmatrix}\tilde{\boldsymbol{\beta}}_{1n}\\ \frac{1}{1+\kappa}\tilde{\boldsymbol{\beta}}_{2n}\end{pmatrix}, \tag{21}$$

to accommodate sparsity condition, see Tibshirani [12] on summary of properties discussed earlier.

In the forthcoming section, we define the traditional shrinkage estimators.

2.2. Shrinkage Estimators

We recall that the unrestricted estimator of $(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ is given by $(\tilde{\boldsymbol{\beta}}_{1n}^\top, \tilde{\boldsymbol{\beta}}_{2n}^\top)^\top$ using marginal distributions $\tilde{\boldsymbol{\beta}}_{1n} \sim \mathcal{N}_{p_1}(\boldsymbol{\beta}_1, \sigma^2 \boldsymbol{C}_{11.2}^{-1})$ and of $\tilde{\boldsymbol{\beta}}_{2n} \sim \mathcal{N}_{p_2}(\boldsymbol{\beta}_2, \sigma^2 \boldsymbol{C}_{22.1}^{-1})$.

The restricted parameter may be denoted by $\boldsymbol{\beta}_{\mathrm{R}}^\top = (\boldsymbol{\beta}_1^\top, \mathbf{0}^\top)$. Thus, the restricted estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}}_n = (\hat{\boldsymbol{\beta}}_{1n}^\top, \mathbf{0}^\top)^\top$, see (5). Next, we consider the preliminary-test estimator (PTE) of $\boldsymbol{\beta}$. For this, we first define the test statistic for testing the sparsity hypothesis $\mathcal{H}_o : \boldsymbol{\beta}_2 = \mathbf{0}$ versus $\mathcal{H}_A : \boldsymbol{\beta}_2 \neq \mathbf{0}$ as

$$\mathcal{L}_n = \frac{\tilde{\boldsymbol{\beta}}_{2n}^\top \boldsymbol{C}_{22.1} \tilde{\boldsymbol{\beta}}_{2n}}{\sigma^2}. \tag{22}$$

Indeed, $\mathcal{L}_n = \chi^2_{p_2}$ (chi-square with p_2 degrees of freedom).

Thus, define the PTE of $(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ with an upper α-level of significance as

$$\hat{\boldsymbol{\beta}}_n^{\mathrm{PT}}(\alpha) = \begin{pmatrix} \tilde{\boldsymbol{\beta}}_{1n} \\ \hat{\boldsymbol{\beta}}_{2n}^{\mathrm{PT}}(\alpha) \end{pmatrix}, \tag{23}$$

where α stands for the level of significance of the test using $\mathcal{L}_n$,

$$\hat{\boldsymbol{\beta}}_{2n}^{\mathrm{PT}}(\alpha) \quad = \quad \tilde{\boldsymbol{\beta}}_{2n} I(\mathcal{L}_n > c_\alpha). \tag{24}$$

In a similar fashion, we define the James-Stein estimator given by

$$\hat{\boldsymbol{\beta}}_n^{\mathrm{JS}} = \begin{pmatrix} \tilde{\boldsymbol{\beta}}_{1n} \\ \hat{\boldsymbol{\beta}}_{2n}^{\mathrm{JS}} \end{pmatrix}, \tag{25}$$

where

$$\hat{\boldsymbol{\beta}}_{2n}^{\mathrm{JS}} \quad = \quad \tilde{\boldsymbol{\beta}}_{2n} \left(1 - d\mathcal{L}_n^{-1}\right), \qquad d = p_2 - 2. \tag{26}$$

The estimator $\hat{\boldsymbol{\beta}}_{1n}^{\mathrm{JS}}$ is not a convex combination of $\tilde{\boldsymbol{\beta}}_{1n}$ and $\hat{\boldsymbol{\beta}}_{1n}$ may change sign opposite to the unrestricted estimator, due to the presence of the term $\left(1 - d\mathcal{L}_n^{-1}\right)$. This is the situation for $\hat{\boldsymbol{\beta}}_{2n}^{\mathrm{JS}}$ as well. To avoid this anomaly, we define the positive-rule Stein-type estimator (PRSE), $\hat{\boldsymbol{\beta}}_n^{\mathrm{S+}}$ as

$$\hat{\boldsymbol{\beta}}_n^{\mathrm{S+}} = \begin{pmatrix} \tilde{\boldsymbol{\beta}}_{1n} \\ \hat{\boldsymbol{\beta}}_{2n}^{\mathrm{S+}} \end{pmatrix}, \tag{27}$$

where

$$\hat{\boldsymbol{\beta}}_{2n}^{\mathrm{S+}} \quad = \quad \tilde{\boldsymbol{\beta}}_{2n} \left(1 - d\mathcal{L}_n^{-1}\right) I(\mathcal{L}_n > d), \quad d = p_2 - 2. \tag{28}$$

3. Bias and weighted L_2-risks of Estimators

First, we consider the bias and L_2-risk expressions of the penalty estimators.

3.1. Hard threshold estimator (Subset selection rule)

Using the results by Donoho and Johnstone [4], we write the bias and L_2-risk of the hard threshold and soft threshold estimators, under non-orthogonal design matrices.

The bias and L_2-risk expressions of $\hat{\boldsymbol{\beta}}_n^{\mathrm{HT}}(\kappa)$ are given by

$$(i) \;\; \mathbf{B}(\hat{\boldsymbol{\beta}}_n^{\mathrm{HT}}(\kappa)) = \left(-\sigma\sqrt{C^{jj}}\Delta_j H_3(\kappa^2; \Delta_j^2) \middle| \, j = 1, \ldots, p\right)^\top, \tag{29}$$

where $H_\nu(\kappa^2;\Delta_j^2)$ is the cumulative distribution function (c.d.f.) of a non-central chi-square distribution with 3 degrees of freedom and non-centrality parameter $\Delta_j^2/2$ $(j=1,\ldots,p)$ and the mean square error of $\hat{\boldsymbol{\beta}}_n^{\mathrm{HT}}(\kappa)$ is given by

$$\begin{aligned} \mathrm{R}_{HT}(\hat{\boldsymbol{\beta}}_n^{\mathrm{HT}}(\kappa)) &= \sum_{j=1}^{p} E\left[\tilde{\beta}_{jn} I\left(\left|\tilde{\beta}_{jn}\right| > \kappa\sigma\sqrt{\boldsymbol{C}^{jj}}\right) - \beta_j\right]^2 \\ &= \sigma^2 \sum_{j=1}^{p} \boldsymbol{C}^{jj}\left\{\left(1-H_3(\kappa^2;\Delta_j^2)\right) + \Delta_j^2\left(2H_3(\kappa^2;\Delta_j^2) - H_5(\kappa^2;\Delta_j^2)\right)\right\} \end{aligned} \tag{30}$$

Since

$$\left[\tilde{\beta}_{jn} I\left(\left|\tilde{\beta}_{jn}\right| > \kappa\sigma\sqrt{\boldsymbol{C}^{jj}}\right) - \beta_j\right]^2 \leq (\tilde{\beta}_{jn}-\beta_j)^2 + \beta_j^2, \quad j=1,\ldots,p.$$

Hence,

$$\mathrm{R}_{\mathrm{HT}}(\hat{\boldsymbol{\beta}}_n^{\mathrm{HT}}(\kappa)) \leq \sigma^2\,\mathrm{tr}(\boldsymbol{C}^{-1}) + \boldsymbol{\beta}^\top\boldsymbol{\beta} \quad (\text{free of } \kappa) \quad \forall \boldsymbol{\beta} \in \mathbb{R}^p \tag{31}$$

Thus, we have the revised form of Lemma 1 of Donoho and Johnstone [4].

Lemma 1

$$\mathrm{R}_{\mathrm{HT}}(\hat{\boldsymbol{\beta}}_n^{\mathrm{HT}}(\kappa)) \leq \begin{cases} \text{(i) } \sigma^2(1+\kappa^2)\,\mathrm{tr}\,\boldsymbol{C}^{-1} & \forall\, \boldsymbol{\beta}\in\mathbb{R}^p,\ \kappa>1 \\ \text{(ii) } \sigma^2\,\mathrm{tr}\,\boldsymbol{C}^{-1} + \boldsymbol{\beta}^\top\boldsymbol{\beta} & \forall\, \boldsymbol{\beta}\in\mathbb{R}^p \\ \text{(iii) } \rho_{\mathrm{HT}}(\kappa;0)\sigma^2\,\mathrm{tr}\,\boldsymbol{C}^{-1} + 1.2\,\boldsymbol{\beta}^\top\boldsymbol{\beta} & \boldsymbol{0}<\boldsymbol{\beta}<\kappa\boldsymbol{1}_p^\top \end{cases} \tag{32}$$

where $\rho_{\mathrm{HT}}(\kappa,0) = 2\left[1-\Phi(\kappa)+\kappa\Phi(-\kappa)\right]$.

The upper bound of (ii) in Lemma 1 is independent of κ. We may obtain the upper bound of the weighted L_2-risk of $\hat{\boldsymbol{\beta}}_n^{\mathrm{HT}}(\kappa)$ as given below by

$$\mathrm{R}_{\mathrm{HT}}(\hat{\boldsymbol{\beta}}_n^{\mathrm{HT}}(\kappa);\boldsymbol{C}_{11.2},\boldsymbol{C}_{22.1}) \leq \begin{cases} (i)\sigma^2(1+\kappa^2)(p_1+p_2) & \kappa>1 \\ (ii)(\sigma^2 p_1 + \boldsymbol{\beta}_1^\top\boldsymbol{C}_{11.2}\boldsymbol{\beta}_1) & \\ \quad +(\sigma^2 p_2 + \boldsymbol{\beta}_2^\top\boldsymbol{C}_{22.1}\boldsymbol{\beta}_2) & \forall\, \boldsymbol{\beta}_1\in\mathbb{R}^{p_1}, \boldsymbol{\beta}_2\in\mathbb{R}^{p_2} \\ (iii)\rho_{\mathrm{HT}}(\kappa;0)\sigma^2 p_1 + 1.2\boldsymbol{\beta}_1^\top\boldsymbol{C}_{11.2}\boldsymbol{\beta}_1 & \\ \quad +\rho_{\mathrm{HT}}(\kappa;0)\sigma^2 p_2 + 1.2\boldsymbol{\beta}_2^\top\boldsymbol{C}_{22.1}\boldsymbol{\beta}_2 & \boldsymbol{0}<\boldsymbol{\beta}<\kappa\boldsymbol{1}_p^\top \end{cases} \tag{33}$$

If we have the sparse solution with p_1 non-zero coefficients, $|\beta_j| > \sigma\sqrt{C^{jj}}$ $(j=1,\ldots,p_1)$ and p_2 zero coefficients

$$\hat{\boldsymbol{\beta}}_n^{\mathrm{HT}}(\kappa) = \begin{pmatrix} \tilde{\boldsymbol{\beta}}_{1n} \\ \boldsymbol{0} \end{pmatrix}. \tag{34}$$

Thus, the upper bound of weighted L_2-risk using Lemma 1 and (7), is given by

$$\mathrm{R}_{\mathrm{HT}}(\hat{\boldsymbol{\beta}}_n^{\mathrm{HT}}(\kappa);\boldsymbol{C}_{11.2},\boldsymbol{C}_{22.1}) \leq \sigma^2(p_1+\Delta^2), \quad \Delta^2 = \frac{1}{\sigma^2}\boldsymbol{\beta}_2^\top\boldsymbol{C}_{22.1}\boldsymbol{\beta}_2 \tag{35}$$

independent of κ.

3.2. Modified LASSO

In this section, we give expressions of bias and mean square errors and weighted L_2-risk. The bias expression for the modified LASSO is given by

$$\mathbf{B}(\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}(\kappa)) = \left(\sigma\sqrt{C^{jj}}\left[\kappa\,\left(2\Phi(\Delta_j)-1\right)\left(1-H_3(\kappa;\Delta_j)\right) + \Delta_j H_3(\kappa^2;\Delta_j)\right]\Big|\, j=1,\ldots,p\right)^\top. \tag{36}$$

The mean square error of the modified LASSO is given by

$$\mathrm{R}_{\mathrm{ST}}(\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}(\kappa), \boldsymbol{\beta}) = \sigma^2 \sum_{j=1}^{p} C^{jj} \rho_{\mathrm{ST}}(k, \Delta_j), \tag{37}$$

where

$$\begin{aligned}\rho_{\mathrm{ST}}(\kappa, \Delta_j) \quad = \quad & 1 + \kappa^2 + (\Delta_j^2 - \kappa^2 - 1)\left\{\Phi(\kappa - \Delta_j) - \Phi(-\kappa - \Delta_j)\right\} \\ & - \left\{(\kappa - \Delta_j)\varphi(\kappa + \Delta_j) + (\kappa + \Delta_j)\varphi(\kappa - \Delta_j)\right\}.\end{aligned} \tag{38}$$

and $\rho_{\mathrm{ST}}(\kappa, 0) = (1 + \kappa^2)\left(2\Phi(\kappa) - 1\right) - 2\kappa\phi(\kappa)$.

Further, Lemma 1 of Donoho and Johnstone [4] gives us the revised Lemma 2 below.

Lemma 2
Under the assumption of this section

$$\mathrm{R}_{\mathrm{ST}}(\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}(\kappa)) \leq \begin{cases} (i)\sigma^2(1+\kappa^2)\,\mathrm{tr}\,\boldsymbol{C}^{-1} & \forall\, \boldsymbol{\beta} \in \mathbb{R}^p,\ \kappa > 1 \\ (ii)\sigma^2\,\mathrm{tr}\,\boldsymbol{C}^{-1} + \boldsymbol{\beta}^\top\boldsymbol{\beta} & \forall\, \boldsymbol{\beta} \in \mathbb{R}^p \\ (iii)\sigma^2\rho_{\mathrm{ST}}(\kappa, 0)\,\mathrm{tr}\,\boldsymbol{C}^{-1} + \boldsymbol{\beta}^\top\boldsymbol{\beta} & \beta_j^2 \in \mathbb{R}^+, j = 1, \ldots, p \end{cases} \tag{39}$$

The second upper bound in Lemma 2 is free of κ. If we have sparse solution with p_1 non-zero and p_2 zero coefficients such as

$$\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}(\kappa) = \begin{pmatrix} \tilde{\boldsymbol{\beta}}_{1n} \\ \mathbf{0} \end{pmatrix}, \tag{40}$$

then the weighted L_2-risk bound using (39) is given by

$$\mathrm{R}_{\mathrm{ST}}(\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}(\kappa); \boldsymbol{C}_{11.2}, \boldsymbol{C}_{22.1}) \leq \sigma^2(p_1 + \Delta^2), \quad \Delta^2 = \frac{1}{\sigma^2}\boldsymbol{\beta}_2^\top \boldsymbol{C}_{22.1}\boldsymbol{\beta}_2, \tag{41}$$

which is independent of κ.

3.3. *Multivariate Normal Decision Theory and Oracles for Diagonal Linear Projection*

Consider the following problem in multivariate normal decision theory. We are given the LSE of $\boldsymbol{\beta}$, namely, $\tilde{\boldsymbol{\beta}}_n = (\tilde{\beta}_{1n}, \ldots, \tilde{\beta}_{pn})^\top$ according to

$$\tilde{\beta}_{jn} = \beta_j + \sigma\sqrt{C^{jj}} Z_j, \quad Z_j \sim \mathcal{N}(0, 1), \tag{42}$$

where $\sigma^2 C^{jj}$ is the marginal variance of $\tilde{\beta}_{jn}$, $j = 1, \ldots, p$, and noise level and $\{\beta_j\}_{j=1}^p$ are the object of interest. We measure the quality of the estimator based on L_2-loss and define the risk as

$$\mathrm{R}(\boldsymbol{\beta}_n^*) = E\|\boldsymbol{\beta}_n^* - \boldsymbol{\beta}\|^2. \tag{43}$$

If there is sparse solution, then use (7) formulation. We consider a family of diagonal linear projections,

$$\boldsymbol{T}_{\mathrm{DP}}(\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}, \boldsymbol{\delta}) = (\delta_1\hat{\boldsymbol{\beta}}_{1n}^{\mathrm{ML}}, \ldots, \delta_2\hat{\boldsymbol{\beta}}_{pn}^{\mathrm{ML}})^\top, \quad \delta_j \in \{0, 1\}. \tag{44}$$

Such estimators *keep* or *kill* co-ordinate. The ideal diagonal coefficients are in this case are $I(|\beta_j| > \sigma\sqrt{C^{jj}})$. These coefficients estimates those β_j's which are larger than the noise level $\sigma\sqrt{C^{jj}}$ yielding the lower bound on the risk as

$$\mathrm{R}_{\sigma^2\boldsymbol{C}^{-1}}(\boldsymbol{T}_{\mathrm{DP}}) = \sum_{j=1}^{p} \min\left(\beta_j^2, \sigma^2\boldsymbol{C}^{(jj)}\right). \tag{45}$$

As a special case of (45), we obtain

$$\mathrm{R}_{\sigma^2 \boldsymbol{C}^{-1}}(\boldsymbol{T}_{\mathrm{DP}}) = \begin{cases} \sigma^2 \operatorname{tr}(\boldsymbol{C}^{-1}) & \text{if all } |\beta_j| \geq \sigma\sqrt{C^{jj}}, j = 1, \ldots, p \\ \boldsymbol{\beta}^\top \boldsymbol{\beta} & \text{if all } |\beta_j| < \sigma\sqrt{C^{jj}}, j = 1, \ldots, p. \end{cases} \tag{46}$$

In general, the risk $\mathrm{R}_{\sigma^2 \boldsymbol{C}^{-1}}(\boldsymbol{T}_{\mathrm{DP}})$ cannot be attained for all $\boldsymbol{\beta}$ by any estimator, linear or non-linear. However, for the sparse case, if p_1 is the number of non-zero coefficients, $|\beta_j| > \sigma\sqrt{C^{jj}}$; $(j = 1, \ldots, p_1)$ and p_2 is the number of zero coefficients, then (46) reduces to the lower bound given by

$$\mathrm{R}_{\sigma^2 \boldsymbol{C}^{-1}}(\boldsymbol{T}_{\mathrm{DP}}) = \sigma^2 \operatorname{tr} \boldsymbol{C}_{11.2}^{-1} + \boldsymbol{\beta}_2^\top \boldsymbol{\beta}_2. \tag{47}$$

Consequently, the weighted L_2-risk lower bound is given by (7) as

$$\mathrm{R}_{\sigma^2 \boldsymbol{C}^{-1}}\left(\boldsymbol{T}_{\mathrm{DP}}; \boldsymbol{C}_{11.2}, \boldsymbol{C}_{22.1}\right) = \sigma^2\left(p_1 + \Delta^2\right), \quad \Delta^2 = \frac{1}{\sigma^2}\boldsymbol{\beta}_2^\top \boldsymbol{C}_{22.1}\boldsymbol{\beta}_2. \tag{48}$$

As we mentioned above that ideal risk cannot be attained in general by any estimator linear or non-linear. However, in the case of modified LASSO and hard thresholding, we revise the theorems 1-4 of Donoho and Johnstone [4] as follows.

Theorem 1
Assume (42) and (43). The modified LASSO defined by (13) with $\kappa^* = \sqrt{2\ln(p)}$ satisfies

$$\mathrm{R}_{\mathrm{ST}}(\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}(\kappa^*)) \leq (2\ln(p) + 1)\left\{\sigma^2 + \sum_{j=1}^{p} \min\left(\beta_j^2, \sigma^2 \boldsymbol{C}^{jj}\right)\right\}; \quad \forall \boldsymbol{\beta} \in \mathbb{R}^p. \tag{49}$$

The inequality says that we can mimic the performance of an oracle plus one extra parameter, σ^2 to within a factor of essentially $2\ln(p)$.

However, it is natural and more revealing to look for *optimal thresholds*, κ_p^* which yields the smallest possible constant Λ_p^* in place of $(2\ln(p) + 1)$ among soft threshold estimators. We state this in the following Theorem.

Theorem 2
Assume (42) - (43). The minimax threshold κ_p^* defined by the minimax quantities

$$\Lambda_p^* = \inf_{\kappa} \sup_{\boldsymbol{\beta}} \frac{\sigma^2 \boldsymbol{C}^{jj} \rho_{\mathrm{ST}}(\kappa, \Delta_j)}{\frac{1}{\operatorname{tr}(\boldsymbol{C}^{-1})} + \min(\Delta_j, 1)} \tag{50}$$

$$\kappa_p^* = \text{The largest } \kappa \text{ attaing } \Lambda_p^* \tag{51}$$

and satisfies the equation

$$(p+1)\rho_{\mathrm{ST}}(\kappa, 0) = \rho_{\mathrm{ST}}(\kappa, \infty) \tag{52}$$

yields the estimator

$$\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}(\kappa_p^*) = \left(\hat{\boldsymbol{\beta}}_{1n}^{\mathrm{ML}}(\kappa_p^*), \ldots, \hat{\boldsymbol{\beta}}_{pn}^{\mathrm{ML}}(\kappa_p^*)\right)^\top \tag{53}$$

which is given by

$$\mathrm{R}_{\mathrm{ST}}(\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}(\kappa_p^*)) \leq \Lambda_p^*\left\{\sigma^2 + \sum_{j=1}^{p} \min\left(\beta_j^2, \sigma^2 \boldsymbol{C}^{jj}\right)\right\}, \quad \forall \boldsymbol{\beta} \in \mathbb{R}^p. \tag{54}$$

The coefficients defined in Λ_p^* satisfy $\Lambda_p^* \leq (2\ln(p) + 1)$ and the threshold $\kappa_p^* \leq \sqrt{2\ln(p)}$. Asymptotically, as $p \to \infty$,

$$\Lambda_n^* \sim 2\ln(p), \qquad \kappa_p^* \sim (2\ln(p))^{\frac{1}{2}}.$$

Theorem 3
The following results hold under the same assumption as in Theorems 1 and 2,

$$\inf_{\tilde{\beta}} \sup_{\beta\in\mathbb{R}^p} \frac{\mathrm{R}_{\mathrm{ST}}\left(\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}(\kappa_p^*)\right)}{\sigma^2+\sum_{j=1}^p \min\left(\beta_j^2, \sigma^2 \boldsymbol{C}^{jj}\right)} \sim 2\ln(p) \quad \text{as} \quad p\to\infty. \tag{55}$$

Finally, we deal with the Theorem 4 related to the hard threshold estimator (subset selection rule).

Theorem 4
With an arbitrary sequence $\{L_n\}$, a thresholding sequence sufficiently close to $\sqrt{2\ln(p)}$, the hard threshold estimator satisfies form $L_p \approx 2\ln(p_n)$ in inequality

$$\mathrm{R}_{\mathrm{HT}}\left(\hat{\boldsymbol{\beta}}_n^{\mathrm{HT}}(\kappa_p^*)\right) \leq L_p \left\{\sigma^2+\sum_{j=1}^p \min\left(\beta_j^2, \sigma^2 \boldsymbol{C}^{jj}\right)\right\}, \quad \forall \boldsymbol{\beta}\in\mathbb{R}^p.$$

where L_p is the pth component of $\{L_n\}$.

Here, sufficiently close to $\sqrt{2\ln(p)}$ means $(1-\gamma)\ln(\ln(p)) \leq L_p - 2\ln(p) \leq o(\ln(p))$ for some $\gamma > 0$.

3.4. Ridge Regression Estimator

We have defined the ridge regression estimator as $\hat{\boldsymbol{\beta}}_n^{\mathrm{RR}}(\kappa) = (\tilde{\boldsymbol{\beta}}_{1n}^\top, \tilde{\boldsymbol{\beta}}_{2n}^\top/(1+\kappa))^\top$ in Eq. (21). The bias and L$_2$-risk are then given by

$$\begin{aligned}
\mathbf{B}(\hat{\boldsymbol{\beta}}_n^{\mathrm{RR}}(\kappa)) &= \begin{pmatrix} \mathbf{0} \\ -\frac{\kappa}{1+\kappa}\boldsymbol{\beta}_2 \end{pmatrix}, \\
\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{RR}}(\kappa)) &= \sigma^2 \operatorname{tr} \boldsymbol{C}_{11.2}^{-1} + \frac{1}{(1+\kappa)^2}\left[\sigma^2 \operatorname{tr} \boldsymbol{C}_{22.1}^{-1} + \kappa^2 \boldsymbol{\beta}_2^\top \boldsymbol{\beta}_2\right].
\end{aligned} \tag{56}$$

The weighed L$_2$-risk is then given by

$$\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{RR}}(\kappa); \boldsymbol{C}_{11.2}, \boldsymbol{C}_{22.1}) = \sigma^2 p_1 + \frac{\sigma^2}{(1+\kappa)^2}\left[p_2+\kappa^2\Delta^2\right], \quad \Delta^2 = \frac{1}{\sigma^2}\boldsymbol{\beta}_2^\top \boldsymbol{C}_{22.1}\boldsymbol{\beta}_2. \tag{57}$$

The optimum value of κ is obtained as $\kappa^* = p_2\Delta^{-2}$; so that

$$\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{RR}}(p_2\Delta^{-2}); \boldsymbol{C}_{11.2}, \boldsymbol{C}_{22.1}) = \sigma^2 p_1 + \frac{\sigma^2 p_2 \Delta^2}{(p_2+\Delta^2)}. \tag{58}$$

3.5. Shrinkage Estimators

From section 2.2, we consider the shrinkage estimators.The unrestricted estimator of $\boldsymbol{\beta}$ is $(\tilde{\boldsymbol{\beta}}_{1n}^\top, \tilde{\boldsymbol{\beta}}_{2n}^\top)^\top$ with bias $(\mathbf{0}^\top, \mathbf{0}^\top)^\top$ and weighted L$_2$-risk given by (47), while the restricted estimator of $(\boldsymbol{\beta}_1^\top, \mathbf{0}^\top)^\top$ is $(\hat{\boldsymbol{\beta}}_{1n}^\top, \mathbf{0}^\top)^\top$. Then, the bias is equal to $(\mathbf{0}^\top, -\boldsymbol{\beta}_2^\top)^\top$ and the weighted L$_2$-risk is given by (9).

Next, we consider the PTE of $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)$ given by (23). Then, the bias and weighted L$_2$-risk are given by

$$\begin{aligned}
\mathbf{B}(\hat{\boldsymbol{\beta}}_n^{\mathrm{PT}}(\alpha)) &= \begin{pmatrix} \mathbf{0} \\ -\boldsymbol{\beta}_2 H_{p_2+2}(c_\alpha; \Delta^2) \end{pmatrix} \\
\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{PT}}(\alpha); \boldsymbol{C}_{11.2}, \boldsymbol{C}_{22.1}) &= \sigma^2 p_1 + \sigma^2 p_2\left[1 - H_{p_2+2}(c_\alpha; \Delta^2)\right] \\
&\quad + \Delta^2\left[2H_{p_2+2}(c_\alpha; \Delta^2) - H_{p_2+4}(c_\alpha; \Delta^2)\right].
\end{aligned} \tag{59}$$

For the James-Stein estimator, we have

$$\mathbf{B}(\hat{\boldsymbol{\beta}}_n^{\mathrm{JS}}) = \begin{pmatrix} \mathbf{0} \\ -(p_2-2)\boldsymbol{\beta}_2 E\left[\chi_{p_2+2}^{-2}(\Delta^2)\right] \end{pmatrix},$$

$$R(\hat{\boldsymbol{\beta}}_n^{\rm JS}(\alpha); \boldsymbol{C}_{11.2}, \boldsymbol{C}_{22.1}) = \sigma^2 p_1 + \sigma^2 p_2 - \sigma^2 (p_2-2)^2 E\left[\chi_{p_2}^{-2}(\Delta^2)\right]. \tag{60}$$

Similarly, the bias and weighted L_2-risk of the PRSE are given by

$$\begin{aligned}
\mathbf{B}(\hat{\boldsymbol{\beta}}_n^{\rm S+}) &= \begin{pmatrix} \mathbf{0} \\ -(p_2-2)\boldsymbol{\beta}_2 \Big\{ H_{p_2+2}(c_\alpha; \Delta^2) \\ +E\left[\chi_{p_2+2}^{-2}(\Delta^2) I\left(\chi_{p_2+2}^{2}(\Delta^2) < p_2-2\right)\right] \Big\} \end{pmatrix}, \\
{\rm R}(\hat{\boldsymbol{\beta}}_n^{\rm S+}(\alpha); \boldsymbol{C}_{11.2}, \boldsymbol{C}_{22.1}) &= {\rm R}(\hat{\boldsymbol{\beta}}_n^{\rm JS}(\alpha); \boldsymbol{C}_{11.2}, \boldsymbol{C}_{22.1}) \\
&\quad -\sigma^2 p_2 E\left[\left(1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta^2)\right)^2 I\left(\chi_{p_2+2}^{-2}(\Delta^2) < p_2-2\right)\right] \\
&\quad +\Delta^2 \Big\{ 2E\left[\left(1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta^2)\right) I\left(\chi_{p_2+2}^{-2}(\Delta^2) < p_2-2\right)\right] \\
&\quad +E\left[\left(1-(p_2-2)\chi_{p_2+4}^{-2}(\Delta^2)\right)^2 I\left(\chi_{p_2+4}^{-2}(\Delta^2) < p_2-2\right)\right] \Big\}.
\end{aligned} \tag{61}$$

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).

4.1. Comparison of LSE with RLSE

In this case, the relative weighted L_2-risk efficiency (RWRE) of RLSE vs LSE is given by

$$\begin{aligned}
{\rm RWRE}(\hat{\boldsymbol{\beta}}_n : \tilde{\boldsymbol{\beta}}_n) &= \frac{p_1+p_2}{{\rm tr}(\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{11.2}) + \Delta^2} \\
&= \left(1+\frac{p_2}{p_1}\right)\left(1-\frac{{\rm tr}(\boldsymbol{M}_0)}{p_1}+\frac{\Delta^2}{p_1}\right)^{-1}; \quad \boldsymbol{M}_0 = \boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1}\boldsymbol{C}_{21},
\end{aligned} \tag{62}$$

which is a decreasing function of Δ^2. So, $0 \le {\rm RWRE}(\hat{\boldsymbol{\beta}}_n : \tilde{\boldsymbol{\beta}}_n) \le \left(1+\frac{p_2}{p_1}\right)\left(1-\frac{{\rm tr}\,\boldsymbol{M}_0}{p_1}\right)^{-1}$.

In order to compute ${\rm tr}(\boldsymbol{M}_0)$, we need to find $\boldsymbol{C}_{11}$, $\boldsymbol{C}_{22}$ and $\boldsymbol{C}_{12}$. These are obtained by generating explanatory variables by the following equation following McDonald and Galarneau [9],

$$x_{ij} = \sqrt{1-\rho^2}\, z_{ij} + \rho z_{ip}, \quad i=1,\ldots,n; j=1,\ldots,p. \tag{63}$$

where z_{ij} are independent $\mathcal{N}(0,1)$ pseudo-random numbers and ρ^2 is the correlation between any two explanatory variables. In this study, we take $\rho^2 = 0.1, 0.2, 0.8$ and 0.9 which shows variables are lightly collinear and severely collinear. In our case, we chose $n = 100$ and various (p_1, p_2). The resulting output is then used to compute ${\rm tr}(\boldsymbol{M}_0)$.

4.2. Comparison of LSE with PTE

Here the RWRE expression for PTE vs LSE is given by

$${\rm RWRE}(\hat{\boldsymbol{\beta}}_n^{\rm PT}(\alpha) : \tilde{\boldsymbol{\beta}}_n) = \frac{p_1+p_2}{p_1 + p_2\left(1-H_{p_2+2}(c_\alpha;\Delta^2)\right) + \Delta^2\left[2H_{p_2+2}(c_\alpha;\Delta^2) - H_{p_2+4}(c_\alpha;\Delta^2)\right]}. \tag{64}$$

Then, the PTE outperforms the LSE for

$$0 \le \Delta^2 \le \frac{p_2 H_{p_2+2}(c_\alpha;\Delta^2)}{2H_{p_2+2}(c_\alpha;\Delta^2) - H_{p_2+4}(c_\alpha;\Delta^2)} = \Delta^2_{\rm PT}. \tag{65}$$

Otherwise, LSE outperforms the PTE in the interval $(\Delta^2_{\rm PT}, \infty)$. We may mention that RWRE$(\hat{\boldsymbol{\beta}}_n^{\rm PT}(\alpha) : \tilde{\boldsymbol{\beta}}_n)$ is a decreasing function of Δ^2 with a maximum at $\Delta^2 = 0$, then decreases crossing the 1-line to a minimum at $\Delta^2 = \Delta^2_{\rm PT}(\min)$ with a value $M_{\rm PT}(\alpha)$, then increases toward 1-line.

The RWRE $\left(\hat{\boldsymbol{\beta}}_n^{\rm PT}; \tilde{\boldsymbol{\beta}}_n\right)$ belongs to the interval

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

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

$$\begin{aligned} M_{\rm PT}(\alpha) &= \left(1 + \frac{p_2}{p_1}\right)\left\{1 + \frac{p_2}{p_1}\left[1 - H_{p_2+2}(c_\alpha; \Delta^2_{\rm PT}(\min))\right]\right. \\ &\left. + \frac{\Delta^2_{\rm PT}(\min))}{p_1}\left[2H_{p_2+2}\left(c_\alpha; \Delta^2_{\rm PT}(\min)\right) - H_{p_2+4}\left(c_\alpha; \Delta^2_{\rm PT}(\min)\right)\right]\right\}^{-1}. \end{aligned}$$

The quantity $\Delta^2_{\rm PT}(\min)$ 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 $p_2 \ge 3$ to express their weighted L_2-risk (WL_2R) expressions, we assume always $p_2 \ge 3$. We have

$$\text{RWRE}\left(\hat{\boldsymbol{\beta}}_n^{\rm JS}; \tilde{\boldsymbol{\beta}}_n\right) = \left(1 + \frac{p_2}{p_1}\right)\left(1 + \frac{p_2}{p_1} - \frac{(p_2-2)^2}{p_1}E\left[\chi^{-2}_{p_2}(\Delta^2)\right]\right)^{-1}. \tag{66}$$

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 $\Delta^2 \to \infty$, its value goes to 1. Hence, for $\Delta^2 \in \mathbb{R}^+$,

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

Also,

$$\begin{aligned} \text{RWRE}\left(\hat{\boldsymbol{\beta}}_n^{\rm S+}; \tilde{\boldsymbol{\beta}}_n\right) &= \left(1 + \frac{p_2}{p_1}\right)\left(1 + \frac{p_2}{p_1} - \frac{(p_2-2)^2}{p_1}E\left[\chi^{-2}_{p_2}(\Delta^2)\right]\right. \\ &- \frac{p_2}{p_1}E\left[\left(1 - (p_2-2)\chi^{-2}_{p_2+2}(\Delta^2)\right)^2 I\left(\chi^2_{p_2+2}(\Delta^2) < (p_2-2)\right)\right] \\ &+ \frac{\Delta^2}{p_1}\left\{2E\left[\left(1 - (p_2-2)\chi^{-2}_{p_2+2}(\Delta^2)\right) I\left(\chi^2_{p_2+2}(\Delta^2) < (p_2-2)\right)\right]\right. \\ &\left.\left. - E\left[\left(1 - (p_2-2)\chi^{-2}_{p_2+4}(\Delta^2)\right)^2 I\left(\chi^2_{p_2+4}(\Delta^2) < (p_2-2)\right)\right]\right\}\right)^{-1}. \end{aligned} \tag{67}$$

So that,

$$\text{RWRE}(\hat{\boldsymbol{\beta}}_n^{\rm S+}; \tilde{\boldsymbol{\beta}}_n) \ge \text{RWRE}(\hat{\boldsymbol{\beta}}_n^{\rm JS}; \tilde{\boldsymbol{\beta}}_n) \ge 1 \qquad \forall \Delta^2 \in \mathbb{R}^+.$$

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

4.4. Comparison of LSE and RLSE with RRE

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

$$\sigma^2(p_1 + p_2) - \sigma^2 p_1 - \sigma^2\frac{p_2\Delta^2}{p_2 + \Delta^2} = \sigma^2 p_2\left(1 - \frac{\Delta^2}{p_2 + \Delta^2}\right) = \frac{\sigma^2 p_2^2}{p_2 + \Delta^2} > 0, \ \forall\ \Delta^2 \in \mathbb{R}^+ \tag{68}$$

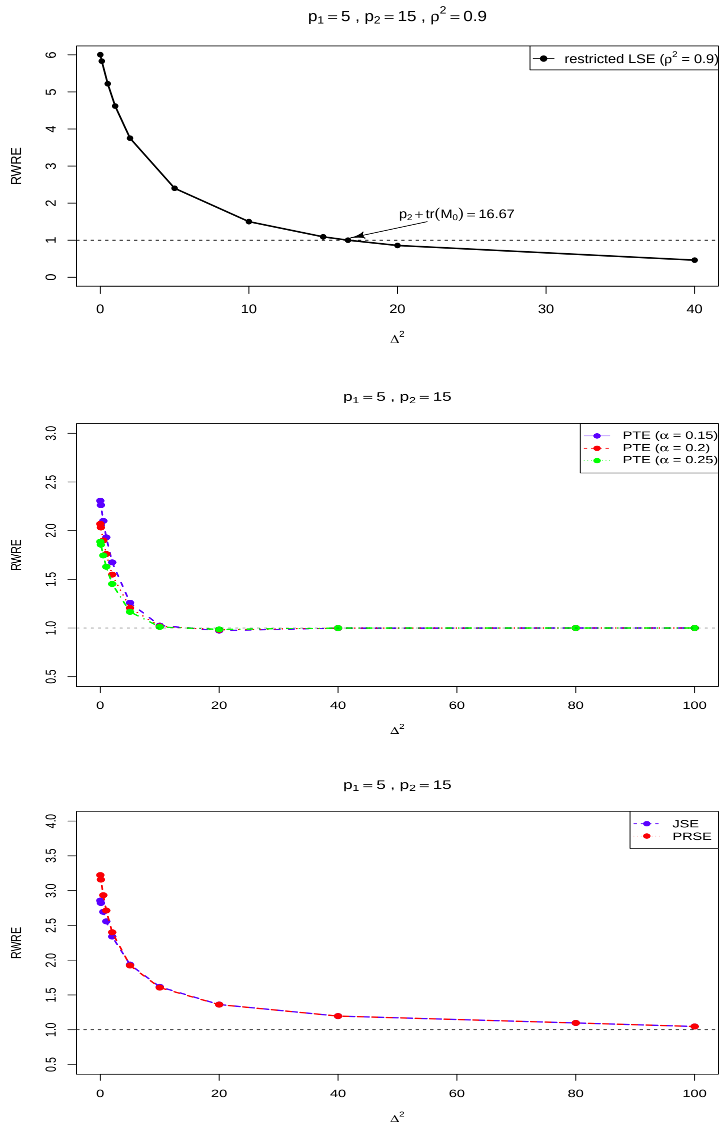

Figure 1. RWRE for the restricted, preliminary test, Stein-type and its positive rule estimators.

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

$$\begin{aligned} \sigma^2\left(\operatorname{tr} \boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{11.2}+\Delta^2\right)-\left(\sigma^2 p_1+\frac{\sigma^2 p_2\Delta^2}{p_2+\Delta^2}\right) &= \sigma^2\left\{\left[\operatorname{tr} \boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{11.2}-p_1\right]+\frac{\Delta^4}{p_2+\Delta^2}\right\} \\ &= \sigma^2\left(\frac{\Delta^4}{p_2+\Delta^2}-\operatorname{tr}(\boldsymbol{M}_0)\right) \end{aligned} \tag{69}$$

If $\Delta^2=0$, then (69) is negative. Hence, RLSE outperforms RRE at this point. Solving the equation

$$\frac{\Delta^4}{p_2+\Delta^2}=\operatorname{tr}(\boldsymbol{M}_0) \tag{70}$$

for Δ^2, we get

$$\Delta_0^2 = \frac{1}{2}\,\mathrm{tr}(\boldsymbol{M}_0)\left\{1+\sqrt{1+\frac{4p_2}{\mathrm{tr}(\boldsymbol{M}_0)}}\right\}. \tag{71}$$

If $0 \le \Delta^2 \le \Delta_0^2$, then RLSE outperform better than the RRE, and if $\Delta^2 \in \left(\Delta_0^2, \infty\right)$, RRE performs better than RLSE; Thus, neither RLSE nor RRE outperforms the other uniformly.

In addition, the RWRE of RRE versus LSE equals

$$\mathrm{RWRE}(\hat{\boldsymbol{\beta}}_n^{\mathrm{RR}}(\kappa^*) : \tilde{\boldsymbol{\beta}}_n) = \frac{p_1+p_2}{p_1+\frac{p_2\Delta^2}{p_2+\Delta^2}} = \left(1+\frac{p_2}{p_1}\right)\left(1+\frac{p_2\Delta^2}{p_1\left(p_2+\Delta^2\right)}\right)^{-1}, \tag{72}$$

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 $\Delta^2 \to \infty$. So,

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

4.5. Comparison of RRE with PTE, JSE and PRSE

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

$$\begin{aligned}&\sigma^2\left[p_2 - p_2 1 - H_{p_2+2}(c_\alpha;\Delta^2) + \Delta^2\left\{2H_{p_2+2}(c_\alpha;\Delta^2) - H_{p_2+4}(c_\alpha;\Delta^2)\right\}\right] - \frac{\sigma^2 p_2\Delta^2}{p_2+\Delta^2}\\ &= \sigma^2\left[\frac{p_2}{p_2+\Delta^2} - \left\{p_2 H_{p_2+2}(c_\alpha;\Delta^2) - \Delta^2\left(2H_{p_2+2}(c_\alpha;\Delta^2) - H_{p_2+4}(c_\alpha;\Delta^2)\right)\right\}\right] \ge 0\end{aligned} \tag{73}$$

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 $\Delta^2 \to \infty$. 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 $\Delta^2 \to \infty$. Hence, (73) is non-negative for $\Delta^2 \in \mathbb{R}^+$. Hence, the RRE uniformly performs better than PTE.

Similarly, we show RRE uniformly performs better than the JSE, i.e., the weighted L_2-risk ($\mathrm{WL}_2\mathrm{R}$) of $\hat{\boldsymbol{\beta}}_n^{\mathrm{RR}}(\kappa^*)$ and $\hat{\boldsymbol{\beta}}_n^{\mathrm{JS}}$ is given by

$$\mathrm{WL}_2\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{RR}}(\kappa^*)) \le \mathrm{WL}_2\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{JS}}), \quad \forall\, \Delta^2 \in \mathbb{R}^+. \tag{74}$$

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

$$\sigma^2\left[\frac{p_2^2}{p_2+\Delta^2} - p_2\left\{E\left[\chi_{p_2+2}^{-2}(\Delta^2)\right] + \Delta^2 E\left[\chi_{p_2+2}^{-4}(\Delta^2)\right]\right\} - (p_2^2-4)E\left[\chi_{p_2+4}^{-4}(\Delta^2)\right]\right] \ge 0 \quad \forall \Delta^2 \in \mathbb{R}^+, \tag{75}$$

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

Next, we show that the weighted L_2-risk ($\mathrm{WL}_2\mathrm{R}$) of the two estimators may be ordered as

$$\mathrm{WL}_2\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{RR}}(\kappa^*)) \le \mathrm{WL}_2\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{S+}}), \quad \forall \Delta^2 \in \mathbb{R}^+.$$

Note that

$$\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{S+}}; \boldsymbol{C}_{11.2}, \boldsymbol{C}_{22.1}) = \mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{JS}}; \boldsymbol{C}_{11.2}, \boldsymbol{C}_{22.1}) - R^*, \tag{76}$$

where

$$\mathrm{R}^* \;=\; \sigma^2 p_2 E\left[\left(1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta^2)\right)^2 I\left(\chi_{p_2+2}^{-2}(\Delta^2) < p_2-2\right)\right]$$

$$+\Delta^2\Bigg\{2E\left[\left(1-(p_2-2)\chi^{-2}_{p_2+2}(\Delta^2)\right)I\left(\chi^{-2}_{p_2+2}(\Delta^2)<p_2-2\right)\right]$$
$$-E\left[\left(1-(p_2-2)\chi^{-2}_{p_2+4}(\Delta^2)\right)^2 I\left(\chi^{-2}_{p_2+4}(\Delta^2)<p_2-2\right)\right]\Bigg\}. \quad (77)$$

Thus, we find that the $\mathrm{WL_2R}$-difference is given by

$$\begin{aligned}\mathrm{WL_2R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{S+}})-\mathrm{WL_2R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{RR}}(\kappa^*)) &= \left\{\mathrm{WL_2R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{JS}})-\mathrm{WL_2R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{RR}}(\kappa^*))\right\}-\mathrm{R}^*\\ &= \{\cdot\}\text{ is negative }-R^*(\text{non-negative})\leq 0. \end{aligned} \quad (78)$$

Hence, the RRE uniformly performs better than the PRSE.

4.6. Comparison of modified LASSO with LSE and RLSE

First note that if p_1 coefficients $|\beta_j|>\sigma\sqrt{C^{jj}}$ and p_2 coefficients are zero in a sparse solution the lower bound of the weighted $\mathrm{L_2}$-risk is given by $\sigma^2(p_1+\Delta^2)$. Thereby, we compare all estimators relative to this quantity. Hence, the weighted $\mathrm{L_2}$-risk difference between LSE and modified LASSO is given by

$$\sigma^2(p_1+p_2)-\sigma^2(p_1+\Delta^2-\mathrm{tr}(\boldsymbol{M}_0))=\sigma^2\left[(p_2+\mathrm{tr}(\boldsymbol{M}_0))-\Delta^2\right]. \quad (79)$$

Hence, if $\Delta^2\in(0,p_2+\mathrm{tr}(\boldsymbol{M}_0))$, the modified LASSO performs better than the LSE, while if $\Delta^2\in(p_2+\mathrm{tr}(\boldsymbol{M}_0),\infty)$ the LSE performs better than the modified LASSO. Consequently, neither LSE nor the modified LASSO performs better than the other uniformly.

Next we compare the RLSE and modified LASSO. In this case the weighted $\mathrm{L_2}$-risk difference is given by

$$\sigma^2(p_1+\Delta^2-\mathrm{tr}(\boldsymbol{M}_0))-\sigma^2(p_1+\Delta^2)=-\sigma^2(\mathrm{tr}(\boldsymbol{M}_0))<0. \quad (80)$$

Hence, the RLSE uniformly performs better than the modified LASSO. If $\mathrm{tr}(\boldsymbol{M}_0)=0$, MLASSO and RLSE $\mathrm{L_2}$-risk equivalent. If the LSE estimators are independent, then $\mathrm{tr}(\boldsymbol{M}_0)=0$. Hence, MLASSO satisfies the oracle properties.

4.7. Comparison of modified LASSO with PTE, JSE and PRSE

We first consider the PTE versus modified LASSO. In this case, the weighted $\mathrm{L_2}$-risk difference is given by

$$\begin{aligned}&\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{PT}}(\alpha);\boldsymbol{C}_{11.2},\boldsymbol{C}_{22.1})-\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}(\kappa^*);\boldsymbol{C}_{11.2},\boldsymbol{C}_{22.1})\\ &=\sigma^2\Bigg[p_2(1-H_{p_2+2}(c_\alpha;\Delta^2))-\Delta^2\left\{1-2H_{p_2+2}(c_\alpha;\Delta^2)+H_{p_2+4}(c_\alpha;\Delta^2)\right\}\Bigg].\\ &\geq\sigma^2 p_2(1-H_{p_2+2}(c_\alpha;0))\geq 0,\quad \text{if }\Delta^2=0.\end{aligned} \quad (81)$$

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

$$0\leq\Delta^2\leq\frac{p_2\left[1-H_{p_2+2}(c_\alpha;\Delta^2)\right]}{1-2H_{p_2+2}(c_\alpha;\Delta^2)+H_{p_2+4}(c_\alpha;\Delta^2)}. \quad (82)$$

Otherwise PTE outperforms the modified LASSO. Hence, neither outperforms the other uniformly.

Next, we consider JSE and PRSE versus the modified LASSO. In these two cases, we have weighted $\mathrm{L_2}$-risk differences given by

$$\begin{aligned}&\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{JS}};\boldsymbol{C}_{11.2},\boldsymbol{C}_{22.1})-\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}(\kappa^*);\boldsymbol{C}_{11.2},\boldsymbol{C}_{22.1})\\ &=\sigma^2\Bigg[p_1+p_2-(p_2-2)^2E\left[\chi^{-2}_{p_2+2}(\Delta^2)\right]-(p_1+\Delta^2)\Bigg]\\ &=\sigma^2\Bigg[p_2-(p_2-2)^2E\left[\chi^{-2}_{p_2+2}(\Delta^2)\right]-\Delta^2\Bigg]\end{aligned} \quad (83)$$

and from (76)

$$\begin{aligned}&\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{S+}};\boldsymbol{C}_{11.2},\boldsymbol{C}_{22.1})-\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{ML}(\kappa^*);\boldsymbol{C}_{11.2},\boldsymbol{C}_{22.1})\\&=\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{JS}};\boldsymbol{C}_{11.2},\boldsymbol{C}_{22.1})-R(\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}(\kappa^*);\boldsymbol{C}_{11.2},\boldsymbol{C}_{22.1})-\mathrm{R}^*,\end{aligned} \tag{84}$$

where R^* is given by (77). Hence, the modified LASSO outperforms the JSE as well as the PRSE in the interval

$$0\leq\Delta^2\leq p_2-(p_2-2)^2E\left[\chi_{p_2}^{-2}(\Delta^2)\right]. \tag{85}$$

Thus, neither the JSE nor PRSE outperforms the modified LASSO uniformly.

4.8. Comparison of modified LASSO with RRE

Here, the weighted L_2-risk difference is given by

$$\begin{aligned}&\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{ML}}(\kappa^*);\boldsymbol{C}_{11.2},\boldsymbol{C}_{22.1})-\mathrm{R}(\hat{\boldsymbol{\beta}}_n^{\mathrm{RR}}(\kappa^*);\boldsymbol{C}_{11.2},\boldsymbol{C}_{22.1})\\&=\sigma^2\left[(p_1+\Delta^2)-\left(p_1+\frac{p_2\Delta^2}{p_2+\Delta^2}\right)\right]=\frac{\sigma^2\Delta^2}{p_2+\Delta^2}\geq 0.\end{aligned} \tag{86}$$

Hence the RRE outperforms the modified LASSO uniformly.

5. Application

Prostate data came from the study of [11] about correlation between the level of prostate specific antigen (PSA), and a number of clinical measures in men who were about to receive radical prostatectomy. The data consist of 97 measurements on the following variables: log cancer volume (lcavol), log prostate weight (lweight), age (age), log of benign prostatic hyperplasia amount (lbph), log of capsular penetration (lcp), seminal vesicle invasion (svi), Gleason score (gleason), and percent of Gleason scores 4 or 5 (pgg45). The idea is to predict log of PSA (lpsa) from these measured variables.

A descriptions of the variables in this dataset is given in Table 1.

Table 1. Description of the variables of prostate data

Symbol	Variables	Description	Remarks
y	lpsa	Log of prostate specific antigen (PSA)	Response
β_1	lcavol	Log cancer volume	
β_2	lweight	Log prostate weight	
β_3	age	Age	Age in years
β_4	lbph	Log of bengin prostate hyperplasia amount	
β_4	svi	Seminal vesicle invasion	
β_5	lcp	Log od capsular penetration	
β_6	gleasson	Gleason score	A numeric vector
β_7	pgg45	Percent of Gleason scores 4 or 5	

The dataset is standardized, then intercept can be ignored. LASSO estimator selects “lcavol” (β_1)”, “lweight (β_2)”, and “svi (β_4)”. So, the parameter can be partitioned as

$$\boldsymbol{\beta}_1=(\beta_1,\beta_2,\beta_4)^\top,\qquad \boldsymbol{\beta}_2=(\beta_3,\beta_5,\beta_6,\beta_7)^\top.$$

In this section, we further investigated the performance of the proposed estimators, LSE, Restricted LSE, preliminary test LSE, James-Stein-type LSE, Positive-rule Stein-type LSE. Our results are based on 1000 case re-sampled samples. The performance of an estimator is evaluated by its prediction error (PE) via 10-fold cross

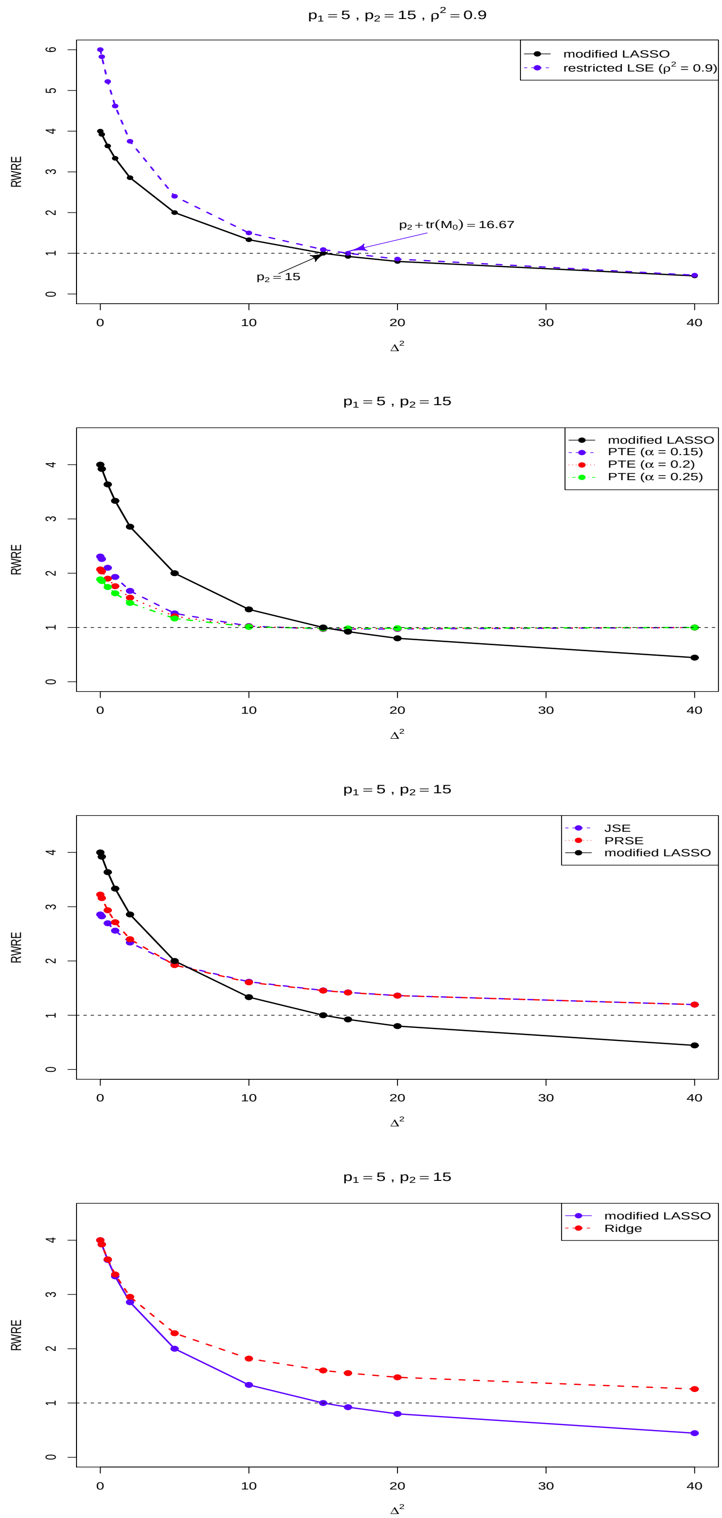

Figure 2. RWRE for the MLASSO, ridge, restricted, preliminary test, Stein-type and its positive rule estimators.

Table 2. The relative prediction errors for the estimators

	LSE	RLSE	PTLSE	JSLSE	PRSLSE
RPE	1.00	0.95	1.02	1.08	1.10

validation (CV) for each bootstrap replicate. In order to easily compare, we also calculated the relative prediction error (RPE) of an estimator with respect to the prediction error of LSE. If the RPE of an estimator is larger one, then its performance is superior to the LSE. Table 2 shows the RPE for the estimators. The Table 2 expressed the positive rule Stein-type estimator improves LSE.

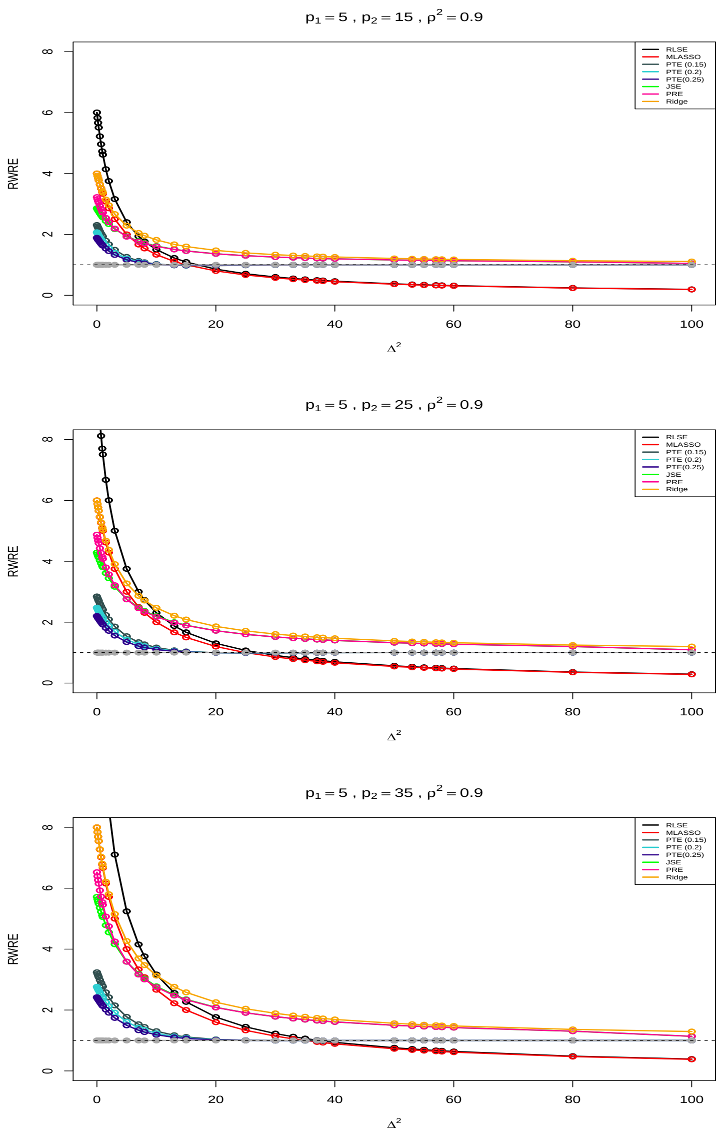

Figure 3. RWRE of estimates of a function of Δ^2 for $p_1 = 5$, $\rho^2 = 0.9$ and different p_2.

Table 3. RWRE for the estimators.

	Δ^2	LSE	RLSE $\rho^2=0.1$	RLSE $\rho^2=0.2$	RLSE $\rho^2=0.8$	RLSE $\rho^2=0.9$	MLASSO	PTE $\alpha=0.15$	PTE $\alpha=0.2$	PTE $\alpha=0.25$	JSE	PRSE	Ridge
							$p_1=5, p_2=15$						
	0	1.0000	4.9112	5.1336	5.7376	5.7864	4.0000	2.3073	2.0693	1.8862	2.8571	3.2234	4.0000
	0.1	1.0000	4.7934	5.0050	5.5775	5.6235	3.9216	2.2622	2.0324	1.8557	2.8224	3.1586	3.9221
	0.5	1.0000	4.3737	4.5493	5.0174	5.0547	3.6364	2.1000	1.8993	1.7453	2.6951	2.9343	3.6471
	1	1.0000	3.9424	4.0845	4.4579	4.4873	3.3333	1.9310	1.7602	1.6295	2.5581	2.7140	3.3684
	2	1.0000	3.2930	3.3915	3.6451	3.6648	2.8571	1.6745	1.5485	1.4526	2.3395	2.4013	2.9565
	3	1.0000	2.8273	2.8996	3.0831	3.0971	2.5000	1.4922	1.3978	1.3264	2.1729	2.1903	2.6667
$\Delta_0^2(\rho^2=0.1)$	4.20	1.0000	2.4155	2.4682	2.5999	2.6098	2.1726	1.3361	1.2689	1.2186	2.0184	2.0123	2.4141
$\Delta_0^2(\rho^2=0.2)$	4.64	1.0000	2.2937	2.3411	2.4593	2.4682	2.0736	1.2916	1.2322	1.1880	1.9713	1.9610	2.3401
	5	1.0000	2.2040	2.2477	2.3564	2.3646	2.0000	1.2595	1.2058	1.1660	1.9363	1.9237	2.2857
$\Delta_0^2(\rho^2=0.8)$	5.57	1.0000	2.0727	2.1113	2.2070	2.2142	1.8913	1.2139	1.1684	1.1349	1.8845	1.8695	2.2065
$\Delta_0^2(\rho^2=0.9)$	5.64	1.0000	2.0594	2.0975	2.1918	2.1989	1.8802	1.2093	1.1647	1.1318	1.8792	1.8640	2.1984
	7	1.0000	1.8059	1.8352	1.9070	1.9124	1.6667	1.1273	1.0979	1.0767	1.7770	1.7609	2.0465
	10	1.0000	1.4210	1.4390	1.4828	1.4861	1.3333	1.0254	1.0169	1.0114	1.6170	1.6059	1.8182
p_2	15	1.0000	1.0485	1.0583	1.0818	1.0835	1.0000	0.9728	0.9785	0.9829	1.4574	1.4535	1.6000
$p_2+\mathrm{tr}(\boldsymbol{M}_0)(\rho^2=0.1)$	15.92	1.0000	1.0000	1.0089	1.0303	1.0319	0.9558	0.9708	0.9774	0.9824	1.4363	1.4332	1.5718
$p_2+\mathrm{tr}(\boldsymbol{M}_0)(\rho^2=0.2)$	16.10	1.0000	0.9913	1.0000	1.0210	1.0225	0.9478	0.9705	0.9774	0.9824	1.4325	1.4295	1.5667
$p_2+\mathrm{tr}(\boldsymbol{M}_0)(\rho^2=0.8)$	16.51	1.0000	0.9715	0.9799	1.0000	1.0015	0.9297	0.9702	0.9773	0.9824	1.4239	1.4212	1.5552
$p_2+\mathrm{tr}(\boldsymbol{M}_0)(\rho^2=0.9)$	16.54	1.0000	0.9701	0.9785	0.9986	1.0000	0.9285	0.9701	0.9773	0.9824	1.4233	1.4206	1.5544
	20	1.0000	0.8307	0.8369	0.8515	0.8526	0.8000	0.9731	0.9808	0.9859	1.3623	1.3612	1.4737
	30	1.0000	0.5869	0.5900	0.5972	0.5978	0.5714	0.9924	0.9953	0.9969	1.2551	1.2550	1.3333
	50	1.0000	0.3699	0.3711	0.3739	0.3741	0.3636	0.9999	0.9999	1.0000	1.1596	1.1596	1.2093
	100	1.0000	0.1922	0.1925	0.1933	0.1933	0.1905	1.0000	1.0000	1.0000	1.0467	1.0467	1.1084
							$p_1=7, p_2=33$						
	0	1.0000	8.9370	9.2311	9.8003	9.8195	5.7143	2.8587	2.4977	2.2279	4.4444	4.9176	5.7143
	0.1	1.0000	8.7414	9.0225	9.5656	9.5838	5.6338	2.8171	2.4643	2.2006	4.3987	4.8424	5.6340
	0.5	1.0000	8.0379	8.2750	8.7296	8.7446	5.3333	2.6619	2.3395	2.0987	4.2270	4.5705	5.3386
	1	1.0000	7.3034	7.4985	7.8699	7.8821	5.0000	2.4906	2.2014	1.9856	4.0350	4.2840	5.0185
	2	1.0000	6.1750	6.3139	6.5753	6.5837	4.4444	2.2085	1.9732	1.7980	3.7116	3.8390	4.5016
	3	1.0000	5.3488	5.4527	5.6466	5.6528	4.0000	1.9877	1.7940	1.6500	3.4498	3.5089	4.1026
	5	1.0000	4.2198	4.2842	4.4030	4.4067	3.3333	1.6701	1.5354	1.4360	3.0520	3.0505	3.5267
	7	1.0000	3.4844	3.5282	3.6085	3.6109	2.8571	1.4594	1.3638	1.2938	2.7641	2.7453	3.1311
	10	1.0000	2.7624	2.7899	2.8398	2.8413	2.3529	1.2602	1.2023	1.1606	2.4561	2.4362	2.7258
$\Delta_0^2(\rho^2=0.1)$	10.46	1.0000	2.6772	2.7030	2.7498	2.7512	2.2908	1.2380	1.1844	1.1460	2.4179	2.3988	2.6768
$\Delta_0^2(\rho^2=0.2)$	10.79	1.0000	2.6187	2.6434	2.6882	2.6896	2.2479	1.2231	1.1724	1.1361	2.3915	2.3730	2.6431
$\Delta_0^2(\rho^2=0.8)$	11.36	1.0000	2.5239	2.5468	2.5883	2.5896	2.1777	1.1993	1.1533	1.1205	2.3483	2.3308	2.5881
$\Delta_0^2(\rho^2=0.9)$	11.38	1.0000	2.5212	2.5441	2.5856	2.5868	2.1757	1.1986	1.1528	1.1201	2.3471	2.3296	2.5866
	15	1.0000	2.0533	2.0685	2.0958	2.0966	1.8182	1.0934	1.0694	1.0529	2.1262	2.1157	2.3105
	20	1.0000	1.6339	1.6435	1.6607	1.6612	1.4815	1.0255	1.0176	1.0125	1.9171	1.9128	2.0563
	30	1.0000	1.1600	1.1649	1.1735	1.1737	1.0811	0.9951	0.9965	0.9974	1.6675	1.6670	1.7610
p_2	33	1.0000	1.0672	1.0713	1.0786	1.0788	1.0000	0.9950	0.9966	0.9976	1.6169	1.6167	1.7021
$p_2+\mathrm{tr}(\boldsymbol{M}_0)(\rho^2=0.1)$	35.52	1.0000	1.0000	1.0036	1.0100	1.0102	0.9408	0.9956	0.9971	0.9980	1.5800	1.5798	1.6593
$p_2+\mathrm{tr}(\boldsymbol{M}_0)(\rho^2=0.2)$	35.66	1.0000	0.9965	1.0000	1.0064	1.0066	0.9376	0.9956	0.9971	0.9981	1.5780	1.5779	1.6570
$p_2+\mathrm{tr}(\boldsymbol{M}_0)(\rho^2=0.8)$	35.91	1.0000	0.9902	0.9937	1.0000	1.0002	0.9321	0.9957	0.9972	0.9981	1.5746	1.5744	1.6531
$p_2+\mathrm{tr}(\boldsymbol{M}_0)(\rho^2=0.9)$	35.92	1.0000	0.9901	0.9936	0.9998	1.0000	0.9320	0.9957	0.9972	0.9981	1.5745	1.5743	1.6530
	50	1.0000	0.7342	0.7361	0.7396	0.7397	0.7018	0.9993	0.9996	0.9998	1.4311	1.4311	1.4881
	100	1.0000	0.3828	0.3834	0.3843	0.3843	0.3738	1.0000	1.0000	1.0000	1.1207	1.1207	1.2574

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 present the RWRE formula from which we prepared our tables and Figures, for quick summary.

$$
\begin{aligned}
&(i) && \text{RWRE}(\hat{\boldsymbol{\beta}}_n;\tilde{\boldsymbol{\beta}}) = \left(1+\frac{p_2}{p_1}\right)\left(1-\frac{\text{tr}(\boldsymbol{M}_0)}{p_1}+\frac{\Delta^2}{p_1}\right)^{-1}\\
&(ii) && \text{RWRE}(\hat{\boldsymbol{\beta}}_n^{\text{ML}}(\kappa);\tilde{\boldsymbol{\beta}}) = \left(1+\frac{p_2}{p_1}\right)\left(1+\frac{\Delta^2}{p_1}\right)^{-1}\\
&(iii) && \text{RWRE}(\hat{\boldsymbol{\beta}}_n^{\text{RR}}(\kappa);\tilde{\boldsymbol{\beta}}) = \left(1+\frac{p_2}{p_1}\right)\left(1+\frac{p_2\Delta^2}{p_1(p_2+\Delta^2)}\right)^{-1}\\
&(iv) && \text{RWRE}(\hat{\boldsymbol{\beta}}_n^{\text{PT}}(\alpha);\tilde{\boldsymbol{\beta}}) = \left(1+\frac{p_2}{p_1}\right)\Bigg\{1+\frac{p_2}{p_1}\left(1-H_{p_2+2}(c_\alpha;\Delta^2)\right)\\
& && \qquad +\frac{\Delta^2}{p_1}\left(2H_{p_2+2}(c_\alpha;\Delta^2)-H_{p_2+4}(c_\alpha;\Delta^2)\right)\Bigg\}^{-1}\\
&(v) && \text{RWRE}(\hat{\boldsymbol{\beta}}_n^{\text{JS}}(\alpha);\tilde{\boldsymbol{\beta}}) = \left(1+\frac{p_2}{p_1}\right)\left\{1+\frac{p_2}{p_1}-\frac{1}{p_1}(p_2-2)E\left[\chi_{p_2}^{-2}(\Delta^2)\right]\right\}^{-1}\\
&(vi) && \text{RWRE}(\hat{\boldsymbol{\beta}}_n^{\text{S+}}(\alpha);\tilde{\boldsymbol{\beta}}) = \left(1+\frac{p_2}{p_1}\right)\Bigg\{1+\frac{p_2}{p_1}-\frac{1}{p_1}(p_2-2)E\left[\chi_{p_2}^{-2}(\Delta^2)\right]\\
& && \qquad -\frac{p_2}{p_1}E\left[\left(1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta^2)\right)^2 I\left(\chi_{p_2+2}^2(\Delta^2)<(p_2-2)\right)\right]\\
& && \qquad +\frac{\Delta^2}{p_1}\Bigg[2E\left[\left(1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta^2)\right) I\left(\chi_{p_2+2}^2(\Delta^2)<(p_2-2)\right)\right]\\
& && \qquad -E\left[\left(1-(p_2-2)\chi_{p_2+4}^{-2}(\Delta^2)\right)^2 I\left(\chi_{p_2+4}^2(\Delta^2)<(p_2-2)\right)\right]\Bigg]\Bigg\}^{-1}
\end{aligned}
\tag{87}
$$

Now, we describe Table 3. This table presents relative weighted L_2-risk efficiency (RWRE) of the seven estimators for $p_1=5$, $p_2=15$ and $p_1=7$, $p_2=33$ against Δ^2-values. using a sample of size $n=100$, the $\boldsymbol{X}$ matrix is produced. Using the model given by Eq. (63) for chosen values $\rho^2=0.1,0.2$ and $0.8,0.9$. Therefore, RWRE-values of RLSE has four entries - two for low correlation and two for high correlation. Some Δ^2-values are given as p_2 and $p_2+\text{tr}(\boldsymbol{M}_0)$ for chosen ρ^2-values. Now, one may use the Table for the performance characteristics of each estimator compared to any other.

Tables 4-5 give the RWRE-values of estimators for $p_1=2,3,5$ and 7 for $p=10,20,40$ and 60.

Table 6 gives the RWRE-values of estimators for $p_1=5$ and $p_2=5,15,25,35$ and 55, and also, for $p_1=7$ and $p_2=3,13,23,33$ and 53 to see the effect of p_2 variation on RWRE.

We have the following conclusion from our study.

(i) Since the inception of the ridge regression estimator by Hoerl and Kennard [8], there have been articles comparing ridge with preliminary test and Stein-type estimators. We have now definitive conclusion that the ridge regression estimator dominates the LSE, PTE and Stein-type estimators uniformly. See Table 3 and graphs there of in Figure 2. The ridge estimator dominates the MLASSO estimator uniformly for $\Delta^2>0$, while they are L_2-risk equivalent at $\Delta^2=0$. The ridge estimator does not select variables but the MLASSO estimator does.

Table 4. RWRE of the estimators for $p = 10, 20$ and different Δ^2-value for varying p_1

	$p = 10$							
	$\Delta^2 = 0$				$\Delta^2 = 1$			
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	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
RLSE ($\rho^2 = 0.1$)	5.6865	3.7180	2.1588	1.4945	3.6238	2.7098	1.7754	1.3002
RLSE ($\rho^2 = 0.2$)	6.1123	3.9360	2.2388	1.5291	3.7913	2.8237	1.8292	1.3263
RLSE ($\rho^2 = 0.8$)	9.4588	5.0515	2.5487	1.6716	4.8587	3.3556	2.0310	1.4322
RLSE ($\rho^2 = 0.9$)	10.0940	5.1866	2.5806	1.6874	5.0216	3.4148	2.0512	1.4437
MLASSO	5.0000	3.3333	2.0000	1.4286	3.3333	2.5000	1.6667	1.2500
PTE ($\alpha = 0.15$)	2.3441	1.9787	1.5122	1.2292	1.7548	1.5541	1.2714	1.0873
PTE ($\alpha = 0.2$)	2.0655	1.7965	1.4292	1.1928	1.6044	1.4499	1.2228	1.0698
PTE ($\alpha = 0.25$)	1.8615	1.6565	1.3616	1.1626	1.4925	1.3703	1.1846	1.0564
JSE	2.5000	2.0000	1.4286	1.1111	2.1364	1.7725	1.3293	1.0781
PRSE	3.0354	2.3149	1.5625	1.1625	2.3107	1.8843	1.3825	1.1026
Ridge	5.0000	3.3333	2.0000	1.4286	3.4615	2.5806	1.7143	1.2903
	$\Delta^2 = 5$				$\Delta^2 = 10$			
LSE	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
RLSE ($\rho^2 = 0.1$)	1.4792	1.3002	1.0381	0.8553	0.8503	0.7879	0.6834	0.5991
RLSE ($\rho^2 = 0.2$)	1.5062	1.3258	1.0562	0.8665	0.8591	0.7973	0.6912	0.6046
RLSE ($\rho^2 = 0.8$)	1.6505	1.4325	1.1206	0.9105	0.9042	0.8347	0.7182	0.6257
RLSE ($\rho^2 = 0.9$)	1.6689	1.4432	1.1267	0.9152	0.9098	0.8383	0.7207	0.6279
MLASSO	1.4286	1.2500	1.0000	0.8333	0.8333	0.7692	0.6667	0.5882
PTE ($\alpha = 0.15$)	1.0515	1.0088	0.9465	0.9169	0.9208	0.9160	0.9176	0.9369
PTE ($\alpha = 0.2$)	1.0357	1.0018	0.9530	0.9323	0.9366	0.9338	0.9374	0.9545
PTE ($\alpha = 0.25$)	1.0250	0.9978	0.9591	0.9447	0.9488	0.9473	0.9517	0.9665
JSE	1.5516	1.3829	1.1546	1.0263	1.3238	1.2250	1.0865	1.0117
PRSE	1.5374	1.3729	1.1505	1.0268	1.3165	1.2199	1.0843	1.0114
Ridge	1.9697	1.6901	1.3333	1.1268	1.5517	1.4050	1.2000	1.0744
	$\Delta^2 = 20$				$\Delta^2 = 60$			
LSE	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
RLSE ($\rho^2 = 0.1$)	0.4595	0.4407	0.4060	0.3746	0.1619	0.1595	0.1547	0.1499
RLSE ($\rho^2 = 0.2$)	0.4621	0.4436	0.4087	0.3768	0.1622	0.1599	0.1551	0.1503
RLSE ($\rho^2 = 0.8$)	0.4749	0.4549	0.4180	0.3849	0.1638	0.1613	0.1564	0.1516
RLSE ($\rho^2 = 0.9$)	0.4764	0.4560	0.4188	0.3857	0.1640	0.1615	0.1566	0.1517
MLASSO	0.4545	0.4348	0.4000	0.3704	0.1613	0.1587	0.1538	0.1493
PTE ($\alpha = 0.15$)	0.9673	0.9722	0.9826	0.9922	1.0000	1.0000	1.0000	1.0000
PTE ($\alpha = 0.2$)	0.9783	0.9818	0.9890	0.9954	1.0000	1.0000	1.0000	1.0000
PTE ($\alpha = 0.25$)	0.9850	0.9877	0.9928	0.9971	1.0000	1.0000	1.0000	1.0000
JSE	1.1732	1.1201	1.0445	1.0053	1.0595	1.0412	1.0150	1.0017
PRSE	1.1728	1.1199	1.0444	1.0053	1.0595	1.0412	1.0150	1.0017
Ridge	1.2963	1.2217	1.1111	1.0407	1.1039	1.0789	1.0400	1.0145
	$p = 20$							
	$\Delta^2 = 0$				$\Delta^2 = 1$			
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	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
RLSE ($\rho^2 = 0.1$)	12.9868	8.5078	4.9239	3.3981	7.8678	5.9662	3.9506	2.9044
RLSE ($\rho^2 = 0.2$)	14.1294	9.0537	5.1464	3.5140	8.2721	6.2298	4.0925	2.9887
RLSE ($\rho^2 = 0.8$)	21.5664	11.4441	5.7417	3.7622	10.4763	7.3696	4.5917	3.2559
RLSE ($\rho^2 = 0.9$)	22.8529	11.7357	5.7920	3.7825	10.6549	7.3927	4.4907	3.1807
MLASSO	10.0000	6.6667	4.0000	2.8571	6.6667	5.0000	3.3333	2.5000
PTE ($\alpha = 0.15$)	3.2041	2.8361	2.3073	1.9458	2.4964	2.2738	1.9310	1.6797
PTE ($\alpha = 0.2$)	2.6977	2.4493	2.0693	1.7926	2.1721	2.0143	1.7602	1.5648
PTE ($\alpha = 0.25$)	2.3469	2.1698	1.8862	1.6694	1.9413	1.8244	1.6295	1.4739
JSE	5.0000	4.0000	2.8571	2.2222	4.1268	3.4258	2.5581	2.0423
PRSE	6.2792	4.7722	3.2234	2.4326	4.5790	3.7253	2.7140	2.1352
Ridge	10.0000	6.6667	4.0000	2.8571	6.7857	5.0704	3.3684	2.5225
	$\Delta^2 = 5$				$\Delta^2 = 10$			
LSE	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
RLSE ($\rho^2 = 0.1$)	3.0563	2.7196	2.2066	1.8370	1.7325	1.6188	1.4220	1.2588
RLSE ($\rho^2 = 0.2$)	3.1154	2.7732	2.2502	1.8704	1.7513	1.6377	1.4400	1.2744
RLSE ($\rho^2 = 0.8$)	3.3725	2.9630	2.3571	1.9385	1.8297	1.7021	1.4831	1.3057
RLSE ($\rho^2 = 0.9$)	3.4024	2.9822	2.3656	1.9439	1.8385	1.7084	1.4864	1.3081
MLASSO	2.8571	2.5000	2.0000	1.6667	1.6667	1.5385	1.3333	1.1765
PTE ($\alpha = 0.15$)	1.4223	1.3625	1.2595	1.1750	1.0788	1.0592	1.0254	0.9982
PTE ($\alpha = 0.2$)	1.3324	1.2864	1.2058	1.1385	1.0580	1.0429	1.0169	0.9961
PTE ($\alpha = 0.25$)	1.2671	1.2306	1.1660	1.1114	1.0437	1.0319	1.0114	0.9952
JSE	2.6519	2.3593	1.9363	1.6464	2.0283	1.8677	1.6170	1.4319
PRSE	2.6304	2.3416	1.9237	1.6370	2.0082	1.8513	1.6059	1.4243
Ridge	3.3824	2.9139	2.2857	1.8848	2.3729	2.1514	1.8182	1.5808
	$\Delta^2 = 20$				$\Delta^2 = 60$			
LSE	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
RLSE ($\rho^2 = 0.1$)	0.9283	0.8947	0.8311	0.7726	0.3250	0.3207	0.3122	0.3035
RLSE ($\rho^2 = 0.2$)	0.9337	0.9004	0.8372	0.7784	0.3256	0.3215	0.3130	0.3044
RLSE ($\rho^2 = 0.8$)	0.9555	0.9195	0.8516	0.7900	0.3282	0.3239	0.3150	0.3062
RLSE ($\rho^2 = 0.9$)	0.9579	0.9214	0.8527	0.7908	0.3285	0.3241	0.3152	0.3063
MLASSO	0.9091	0.8696	0.8000	0.7407	0.3226	0.3175	0.3077	0.2985
PTE ($\alpha = 0.15$)	0.9747	0.9737	0.9731	0.9743	1.0000	1.0000	1.0000	1.0000
PTE ($\alpha = 0.2$)	0.9813	0.9808	0.9808	0.9820	1.0000	1.0000	1.0000	1.0000
PTE ($\alpha = 0.25$)	0.9860	0.9857	0.9859	0.9870	1.0000	1.0000	1.0000	1.0000
JSE	1.5796	1.4978	1.3623	1.2560	1.2078	1.1811	1.1344	1.0957
PRSE	1.5775	1.4961	1.3612	1.2553	1.2078	1.1811	1.1344	1.0957
Ridge	1.7431	1.6408	1.4737	1.3442	1.2621	1.2310	1.1765	1.1309

Table 5. RWRE of the estimators for $p = 40, 60$ and different Δ^2-value for varying p_1

	$p = 40$							
	$\Delta^2 = 0$				$\Delta^2 = 1$			
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	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
RLSE ($\rho^2 = 0.1$)	34.9765	22.7333	13.0298	8.9503	18.6256	14.4799	9.8239	7.3123
RLSE ($\rho^2 = 0.2$)	38.1713	24.2975	13.6428	9.2480	19.4906	15.0978	10.1684	7.5097
RLSE ($\rho^2 = 0.8$)	57.9798	30.3910	15.0546	9.8031	23.6099	17.2473	10.9327	7.8715
RLSE ($\rho^2 = 0.9$)	61.3963	30.9463	15.2059	9.8320	24.1568	17.4240	11.0126	7.8901
MLASSO	20.0000	13.3333	8.0000	5.7143	13.3333	10.0000	6.6667	5.0000
PTE ($\alpha = 0.15$)	4.0480	3.7368	3.2391	2.8587	3.3255	3.1162	2.7683	2.4906
PTE ($\alpha = 0.2$)	3.2836	3.0890	2.7619	2.4977	2.7715	2.6348	2.3984	2.2014
PTE ($\alpha = 0.25$)	2.7789	2.6477	2.4196	2.2279	2.3962	2.3008	2.1314	1.9856
JSE	10.0000	8.0000	5.7143	4.4444	8.1231	6.7542	5.0518	4.0350
PRSE	12.8021	9.6941	6.5253	4.9176	9.2460	7.5071	5.4551	4.2840
Ridge	20.0000	13.3333	8.0000	5.7143	13.4483	10.0662	6.6977	5.0185
	$\Delta^2 = 5$				$\Delta^2 = 10$			
LSE	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
RLSE ($\rho^2 = 0.1$)	6.5036	5.9124	4.9539	4.2228	3.5871	3.3995	3.0591	2.7637
RLSE ($\rho^2 = 0.2$)	6.5833	6.0083	5.0479	4.3007	3.6180	3.4325	3.0918	2.7914
RLSE ($\rho^2 = 0.8$)	7.0222	6.3274	5.2212	4.4034	3.7395	3.5328	3.1591	2.8399
RLSE ($\rho^2 = 0.9$)	7.0699	6.3510	5.2394	4.4092	3.7530	3.5402	3.1657	2.8423
MLASSO	5.7143	5.0000	4.0000	3.3333	3.3333	3.0769	2.6667	2.3529
PTE ($\alpha = 0.15$)	1.9641	1.8968	1.7758	1.6701	1.3792	1.3530	1.3044	1.2602
PTE ($\alpha = 0.2$)	1.7519	1.7034	1.6146	1.5354	1.2928	1.2731	1.2362	1.2023
PTE ($\alpha = 0.25$)	1.6018	1.5652	1.4974	1.4360	1.2316	1.2163	1.1873	1.1606
JSE	4.8733	4.3528	3.5875	3.0520	3.4590	3.1966	2.7769	2.4561
PRSE	4.8815	4.3574	3.5880	3.0505	3.4186	3.1623	2.7511	2.4362
Ridge	6.2319	5.4019	4.2667	3.5267	4.0336	3.6791	3.1304	2.7258
	$\Delta^2 = 20$				$\Delta^2 = 60$			
LSE	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
RLSE ($\rho^2 = 0.1$)	1.8911	1.8377	1.7334	1.6344	0.6541	0.6476	0.6341	0.6204
RLSE ($\rho^2 = 0.2$)	1.9327	1.8759	1.7650	1.6607	0.6551	0.6488	0.6355	0.6218
RLSE ($\rho^2 = 0.8$)	1.9269	1.8694	1.7660	1.6619	0.6590	0.6523	0.6383	0.6242
RLSE ($\rho^2 = 0.9$)	1.9363	1.8780	1.7671	1.6616	0.6594	0.6525	0.6386	0.6243
MLASSO	1.8182	1.7391	1.6000	1.4815	0.6452	0.6349	0.6154	0.5970
PTE ($\alpha = 0.15$)	1.0520	1.0461	1.0352	1.0255	0.9998	0.9998	0.9998	0.9999
PTE ($\alpha = 0.2$)	1.0370	1.0327	1.0247	1.0176	0.9999	0.9999	0.9999	0.9999
PTE ($\alpha = 0.25$)	1.0271	1.0238	1.0178	1.0125	0.9999	0.9999	1.0000	1.0000
JSE	2.4152	2.2946	2.0878	1.9171	1.5195	1.4852	1.4221	1.3659
PRSE	2.4070	2.2874	2.0823	1.9128	1.5195	1.4852	1.4221	1.3659
Ridge	2.6484	2.5027	2.2564	2.0563	1.5832	1.5452	1.4757	1.4139
	$p = 60$							
	$\Delta^2 = 0$				$\Delta^2 = 1$			
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	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
RLSE ($\rho^2 = 0.1$)	78.6762	50.4531	28.3644	19.2298	33.9183	27.3362	19.2404	14.55381
RLSE ($\rho^2 = 0.2$)	85.7484	53.8968	29.6649	19.8363	35.1589	28.3182	19.8283	14.8978
RLSE ($\rho^2 = 0.8$)	130.1230	67.0712	32.6637	20.9788	40.8590	31.5747	21.1221	15.5342
RLSE ($\rho^2 = 0.9$)	138.2670	68.8123	33.1316	21.0086	41.6214	31.9529	21.3157	15.5498
MLASSO	30.0000	20.0000	12.0000	8.5714	20.0000	15.0000	10.0000	7.5000
PTE ($\alpha = 0.15$)	4.4902	4.2281	3.7863	3.4283	3.8042	3.6170	3.2930	3.0225
PTE ($\alpha = 0.2$)	3.5800	3.4220	3.1446	2.9089	3.1043	2.9869	2.7770	2.5949
PTE ($\alpha = 0.25$)	2.9925	2.8887	2.7015	2.5371	2.6418	2.5624	2.4172	2.2877
JSE	15.0000	12.0000	8.5714	6.6667	12.1220	10.0860	7.5500	6.0331
PRSE	19.3544	14.6307	9.8315	7.4030	13.9866	11.3392	8.2253	6.4531
Ridge	30.0000	20.0000	12.0000	8.5714	20.1136	15.0649	10.0299	7.5174
	$\Delta^2 = 5$				$\Delta^2 = 10$			
LSE	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
RLSE ($\rho^2 = 0.1$)	10.3945	9.6787	8.4245	7.3833	5.5695	5.3570	4.9492	4.5704
RLSE ($\rho^2 = 0.2$)	10.5082	9.7994	8.5350	7.4706	5.6020	5.3939	4.9871	4.6037
RLSE ($\rho^2 = 0.8$)	10.9671	10.1632	8.7660	7.6278	5.7299	5.5024	5.0651	4.6630
RLSE ($\rho^2 = 0.9$)	11.0214	10.2021	8.7991	7.6313	5.7447	5.5138	5.0761	4.6643
MLASSO	8.5714	7.5000	6.0000	5.0000	5.0000	4.6154	4.0000	3.5294
PTE ($\alpha = 0.15$)	2.3521	2.2844	2.1602	2.0491	1.6311	1.6029	1.5498	1.5006
PTE ($\alpha = 0.2$)	2.0425	1.9957	1.9085	1.8290	1.4873	1.4667	1.4275	1.3908
PTE ($\alpha = 0.25$)	1.8284	1.7942	1.7298	1.6701	1.3864	1.3707	1.3406	1.3122
JSE	7.0961	6.3514	5.2500	4.4745	4.8898	4.5300	3.9495	3.5015
PRSE	7.1699	6.4081	5.2855	4.4980	4.8382	4.4854	3.9152	3.4743
Ridge	9.0865	7.8981	6.2609	5.1863	5.6983	5.2140	4.4571	3.8929
	$\Delta^2 = 20$				$\Delta^2 = 60$			
LSE	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
RLSE ($\rho^2 = 0.1$)	2.8883	2.8301	2.7120	2.5941	0.9873	0.9804	0.9658	0.9504
RLSE ($\rho^2 = 0.2$)	2.8970	2.8403	2.7233	2.6048	0.9883	0.9816	0.9672	0.9519
RLSE ($\rho^2 = 0.8$)	2.9309	2.8702	2.7464	2.6237	0.9922	0.9851	0.9701	0.9544
RLSE ($\rho^2 = 0.9$)	2.9348	2.8732	2.7497	2.6241	0.9926	0.9855	0.9705	0.9544
MLASSO	2.7273	2.6087	2.4000	2.2222	0.9677	0.9524	0.9231	0.8955
PTE ($\alpha = 0.15$)	1.1500	1.1422	1.1274	1.1134	0.9996	0.9996	0.9996	0.9997
PTE ($\alpha = 0.2$)	1.1110	1.1052	1.0941	1.0837	0.9998	0.9998	0.9998	0.9998
PTE ($\alpha = 0.25$)	1.0844	1.0799	1.0714	1.0634	0.9999	0.9999	0.9999	0.9999
JSE	3.2515	3.0957	2.8258	2.6000	1.8333	1.7932	1.7187	1.6510
PRSE	3.2353	3.0812	2.8141	2.5905	1.8333	1.7932	1.7187	1.6510
Ridge	3.5562	3.3698	3.0508	2.7880	1.9053	1.8616	1.7806	1.7074

Table 6. RWRE values of estimators for $p_1 = 5, 7$, different values of p_2 and Δ^2

	LSE	RLSE $\rho^2 = 0.8$	RLSE $\rho^2 = 0.9$	MLASSO	PTE $\alpha = 0.15$	PTE $\alpha = 0.2$	PTE $\alpha = 0.25$	JSE	PRSE	Ridge
p_2	$p_1 = 5$ and $\Delta^2 = 0$									
5	1.0000	2.0072	2.0251	1.7143	1.5629	1.3933	1.2842	1.3333	1.4286	1.7143
15	1.0000	4.1405	4.1561	3.1429	2.6122	2.0622	1.7456	2.4444	2.6814	3.1429
25	1.0000	6.8120	6.8472	4.5714	3.5211	2.5509	2.0491	3.5556	3.9239	4.5714
35	1.0000	10.2879	10.3294	6.0000	4.3200	2.9270	2.2664	4.6667	5.1661	6.0000
55	1.0000	21.7292	21.7532	8.8571	5.6663	3.4741	2.5609	6.8889	7.6516	8.8571
p_2	$p_1 = 5$ and $\Delta^2 = 0.5$									
5	1.0000	1.8523	1.8675	1.6000	1.4393	1.2944	1.2082	1.2931	1.3548	1.6098
15	1.0000	3.7842	3.7973	2.9333	2.4022	1.9091	1.6345	2.3352	2.4955	2.9397
25	1.0000	6.1562	6.1850	4.2667	3.2427	2.3685	1.9247	3.3857	3.6465	4.2723
35	1.0000	9.1641	9.1969	5.6000	3.9859	2.7274	2.1365	4.4373	4.8017	5.6053
55	1.0000	18.4808	18.4984	8.2667	5.2488	3.2581	2.4294	6.5417	7.1189	8.2716
p_2	$p_1 = 5$ and $\Delta^2 = 1$									
5	1.0000	1.7195	1.7327	1.5000	1.3373	1.2163	1.1492	1.2601	1.2996	1.5319
15	1.0000	3.4844	3.4955	2.7500	2.2224	1.7803	1.5413	2.2407	2.3481	2.7717
25	1.0000	5.6157	5.6396	4.0000	3.0018	2.2116	1.8172	3.2366	3.4206	4.0193
35	1.0000	8.2618	8.2884	5.2500	3.6950	2.5533	2.0224	4.2347	4.5003	5.2683
55	1.0000	16.0790	16.0925	7.7500	4.8819	3.0662	2.3115	6.2330	6.6706	7.7673
p_2	$p_1 = 5$ and $\Delta^2 = 5$									
5	1.0000	1.0930	1.0983	1.0000	0.9453	0.9550	0.9657	1.1256	1.1223	1.2632
15	1.0000	2.1329	2.1370	1.8333	1.4112	1.2305	1.1487	1.7844	1.7746	2.0465
25	1.0000	3.2991	3.3073	2.6667	1.8651	1.4883	1.3192	2.4855	2.4782	2.8657
35	1.0000	4.6226	4.6309	3.5000	2.2877	1.7125	1.4628	3.1939	3.1944	3.6923
55	1.0000	7.8882	7.8916	5.1667	3.0478	2.0822	1.6900	4.6170	4.6436	5.3525
p_2	$p_1 = 7$ and $\Delta^2 = 0$									
3	1.0000	1.6729	1.6869	1.4286	1.3333	1.2292	1.1626	1.1111	1.1625	1.4286
13	1.0000	3.7634	3.7819	2.8571	2.4147	1.9458	1.6694	2.2222	2.4326	2.8571
23	1.0000	6.3868	6.4205	4.2857	3.3490	2.4637	1.9968	3.3333	3.6755	4.2857
33	1.0000	9.7979	9.8493	5.7143	4.1679	2.8587	2.2279	4.4444	4.9176	5.7143
53	1.0000	21.0137	21.0565	8.5714	5.5442	3.4283	2.5371	6.6667	7.4030	8.5714
p_2	$p_1 = 7$ and $\Delta^2 = 0.5$									
3	1.0000	1.5438	1.5557	1.3333	1.2307	1.1485	1.1015	1.0928	1.1282	1.3462
13	1.0000	3.4396	3.4551	2.6667	2.2203	1.8009	1.5627	2.1255	2.2661	2.6733
23	1.0000	5.7719	5.7994	4.0000	3.0831	2.2860	1.8742	3.1754	3.4159	4.0057
33	1.0000	8.7277	8.7683	5.3333	3.8441	2.6619	2.0987	4.2270	4.5705	5.3386
53	1.0000	17.8743	17.9055	8.0000	5.1338	3.2133	2.4055	6.3313	6.8869	8.0050
p_2	$p_1 = 7$ and $\Delta^2 = 1$									
3	1.0000	1.4331	1.4434	1.2500	1.1485	1.0873	1.0564	1.0781	1.1026	1.2903
13	1.0000	3.1672	3.1803	2.5000	2.0544	1.6797	1.4739	2.0423	2.1352	2.5225
23	1.0000	5.2651	5.2879	3.7500	2.8534	2.1337	1.7688	3.0371	3.2054	3.7696
33	1.0000	7.8684	7.9014	5.0000	3.5625	2.4906	1.9856	4.0350	4.2840	5.0185
53	1.0000	15.5525	15.5763	7.5000	4.7733	3.0225	2.2877	6.0331	6.4531	7.5174
p_2	$p_1 = 7$ and $\Delta^2 = 5$									
3	1.0000	0.9109	0.9151	0.8333	0.8688	0.9169	0.9447	1.0263	1.0268	1.1268
13	1.0000	1.9388	1.9437	1.6667	1.3171	1.1750	1.1114	1.6464	1.6370	1.8848
23	1.0000	3.0930	3.1008	2.5000	1.7768	1.4395	1.2873	2.3444	2.3362	2.7010
33	1.0000	4.4025	4.4128	3.3333	2.2057	1.6701	1.4360	3.0520	3.0505	3.5267
53	1.0000	7.6320	7.6378	5.0000	2.9764	2.0491	1.6701	4.4745	4.4980	5.1863

(ii) The Restricted LSE (RLSE) and modified LASSO (MLASSO) are competitive, although MLASSO lags behind RLSE uniformly. Both estimators outperform the LSE, PTE, JSE and PRSE in a sub interval of $[0, p_2]$. See Table 3.

(iii) The lower bound of L_2-risk of HTE and MLASSO is the same and independent of the threshold parameter (κ). But the upper bound of L_2-risk is dependent of κ.

Table 7. RWRE values of estimators for $p_2 = 5, 7$, different values of p_1 and Δ^2

	LSE	RLSE $\rho^2 = 0.8$	RLSE $\rho^2 = 0.9$	MLASSO	PTE $\alpha = 0.15$	PTE $\alpha = 0.2$	PTE $\alpha = 0.25$	JSE	PRSE	Ridge
p_1	$p_2 = 5$ and $\Delta^2 = 0$									
5	1.0000	2.5504	2.5767	2.0000	1.7612	1.5122	1.3616	1.4286	1.5625	2.0000
15	1.0000	1.4828	1.4862	1.3333	1.2757	1.2039	1.1531	1.1765	1.2195	1.3333
25	1.0000	1.3005	1.3018	1.2000	1.1683	1.1273	1.0971	1.1111	1.1364	1.2000
35	1.0000	1.2248	1.2255	1.1429	1.1211	1.0925	1.0711	1.0811	1.0989	1.1429
55	1.0000	1.1575	1.1577	1.0909	1.0776	1.0598	1.0463	1.0526	1.0638	1.0909
p_1	$p_2 = 5$ and $\Delta^2 = 0.5$									
5	1.0000	2.2619	2.2825	1.8182	1.5779	1.3754	1.2607	1.3737	1.4583	1.8333
15	1.0000	1.4298	1.4329	1.2903	1.2242	1.1580	1.1153	1.1574	1.1864	1.2941
25	1.0000	1.2729	1.2742	1.1765	1.1391	1.1001	1.0740	1.0997	1.1170	1.1786
35	1.0000	1.2063	1.2071	1.1268	1.1008	1.0732	1.0545	1.0730	1.0853	1.1282
55	1.0000	1.1465	1.1466	1.0811	1.0650	1.0477	1.0357	1.0475	1.0553	1.0820
p_1	$p_2 = 5$ and $\Delta^2 = 1$									
5	1.0000	2.0321	2.0487	1.6667	1.4340	1.2714	1.1846	1.3293	1.3825	1.7143
15	1.0000	1.3804	1.3834	1.2500	1.1783	1.1195	1.0845	1.1414	1.1605	1.2632
25	1.0000	1.2465	1.2477	1.1538	1.1122	1.0766	1.0548	1.0900	1.1016	1.1613
35	1.0000	1.1884	1.1891	1.1111	1.0819	1.0564	1.0405	1.0660	1.0743	1.1163
55	1.0000	1.1356	1.1358	1.0714	1.0531	1.0369	1.0267	1.0431	1.0483	1.0746
p_1	$p_2 = 5$ and $\Delta^2 = 5$									
5	1.0000	1.1209	1.1259	1.0000	0.9351	0.9465	0.9591	1.1546	1.1505	1.3333
15	1.0000	1.0818	1.0836	1.0000	0.9665	0.9725	0.9791	1.0717	1.0700	1.1429
25	1.0000	1.0688	1.0697	1.0000	0.9774	0.9815	0.9860	1.0467	1.0456	1.0909
35	1.0000	1.0622	1.0627	1.0000	0.9830	0.9861	0.9895	1.0346	1.0338	1.0667
55	1.0000	1.0557	1.0558	1.0000	0.9886	0.9907	0.9929	1.0228	1.0223	1.0435
p_1	$p_2 = 7$ and $\Delta^2 = 0$									
3	1.0000	5.0561	5.1897	3.3333	2.6044	1.9787	1.6565	2.0000	2.3149	3.3333
13	1.0000	1.7667	1.7712	1.5385	1.4451	1.3286	1.2471	1.3333	1.3967	1.5385
23	1.0000	1.4493	1.4508	1.3043	1.2584	1.1974	1.1522	1.2000	1.2336	1.3043
33	1.0000	1.3297	1.3298	1.2121	1.1820	1.1411	1.1100	1.1429	1.1655	1.2121
53	1.0000	1.2278	1.2280	1.1321	1.1144	1.0898	1.0707	1.0909	1.1046	1.1321
p_1	$p_2 = 7$ and $\Delta^2 = 0.5$									
3	1.0000	4.0353	4.1200	2.8571	2.2117	1.7335	1.4923	1.8733	2.0602	2.8846
13	1.0000	1.6920	1.6961	1.4815	1.3773	1.2683	1.1975	1.3039	1.3464	1.4851
23	1.0000	1.4151	1.4166	1.2766	1.2234	1.1642	1.1235	1.1840	1.2071	1.2784
33	1.0000	1.3079	1.3081	1.1940	1.1587	1.1183	1.0899	1.1319	1.1476	1.1952
53	1.0000	1.2154	1.2156	1.1215	1.1005	1.0759	1.0582	1.0842	1.0938	1.1222
p_1	$p_2 = 7$ and $\Delta^2 = 1$									
3	1.0000	3.3576	3.4161	2.5000	1.9266	1.5541	1.3703	1.7725	1.8843	2.5806
13	1.0000	1.6233	1.6271	1.4286	1.3166	1.2169	1.1562	1.2786	1.3066	1.4414
23	1.0000	1.3825	1.3839	1.2500	1.1909	1.1349	1.0990	1.1700	1.1854	1.2565
33	1.0000	1.2869	1.2870	1.1765	1.1367	1.0979	1.0725	1.1223	1.1329	1.1808
53	1.0000	1.2032	1.2034	1.1111	1.0871	1.0632	1.0472	1.0783	1.0849	1.1137
p_1	$p_2 = 7$ and $\Delta^2 = 5$									
3	1.0000	1.4329	1.4435	1.2500	1.0489	1.0088	0.9978	1.3829	1.3729	1.6901
13	1.0000	1.2254	1.2276	1.1111	1.0239	1.0044	0.9989	1.1607	1.1572	1.2565
23	1.0000	1.1673	1.1683	1.0714	1.0158	1.0029	0.9993	1.1017	1.0996	1.1576
33	1.0000	1.1402	1.1403	1.0526	1.0118	1.0022	0.9994	1.0744	1.0729	1.1137
53	1.0000	1.1139	1.1140	1.0345	1.0078	1.0015	0.9996	1.0484	1.0474	1.0730

(iv) Maximum of RWRE occurs at $\Delta^2 = 0$, which indicates that the LSE underperforms all estimators for any value of (p_1, p_2). Clearly, RLSE outperforms all estimators for any (p_1, p_2) at $\Delta^2 = 0$. However, as Δ^2 deviates from 0, the PTE and Stein-type estimators outperform LSE, RLSE and MLASSO. See Table 3.

(v) If p_1 is fixed and p_2 increases, the RWRE of all estimators increases. See Table 6.

(vi) If p_2 is fixed and p_1 increases, the RWRE of all estimators decreases. Then, for p_2 small and p_2 large, the MLASSO, PTE, JSE, and PRSE are competitive. See Table 7.

(vii) The PRSE is always outperform JSE. See Tables 3-7.

REFERENCES

1. A. Belloni, and V. Chernozhukov, *Least squares after model selection in high-dimensional sparse models*, Bernoulli, vol. 2, pp. 521–547, 2013.
2. L. Breiman, *Heuristics of instability and stabilization in model selection*, The Annals of Statistics, vol. 24, no. 6, pp. 2350–2383, 1996.
3. N. R. V. Draper, and R. Craig, *Ridge Regression and James- Stein Estimation: Review and Comments*, Technometrics, vol. 21, pp. 451–466., 1979.
4. D. L. Donoho, and I. M. Johnstone, *Ideal Spatial Adaptation by Wavelet Shrinkage*, Biometrika, vol. 81, no. 3, pp. 425–455, 1994.
5. J. Fan, and R. Li, *Variable selection via nonconcave penalized likelihood and its oracle properties*, Journal of the American Statistical Association, vol. 96, no. 456, pp. 1348–1360, 2001.
6. I. E. Frank, and J. H. Friedman, *A Statistical View of Some Chemometrics Regression Tools*, Technometrics, vol. 35, no. 2, pp. 109–135, 1993.
7. B. E. Hansen, *The risk of James-Stein and LASSO Shrinkage*, Econometric Reviews, vol. 35, no. 8-10, pp. 1456–1470, 2013.
8. A. E. Hoerl and R. W. Kennard, *Ridge regression biased estimation for non-orthogonal problems*, Thechnometrics, vol. 12, pp. 69–89, 1970.
9. G. C. McDonald, and D. I. Galarneau, *A Monte Carol evaluation of some ridge-type estimators*, J. Amer. Statist. Assoc, vol. 70, pp. 407–416, 1975.
10. A. K. Md. Ehsanes. Saleh, *Theory of Preliminary Test and Stein-Type Estimation with Applications*, Wiley; United Stated of America., 2006.
11. Stamey, T.A., Kabalin, J.N., McNeal, J.E., Johnstone, I.M., Freiha, F., Redwine, E.A. and Yang, N. *Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate: II. radical prostatectomy treated patients*, Journal of Urology, 141(5), 1076-1083.
12. R. Tibshirani, *Regression shrinkage and selection via the LASSO*, J. Royal. Statist. Soc. B., vol. 58, no. 1, pp. 267-288, 1996.
13. A. N. Tikhonov, *Solution of incorrectly formulated problems and the regularization method*, Doklady Academia Nauk SSSR, Translated in Soviet Mathematics 4, vol. 151, pp. 501-C504, 1963.
14. H. Zou, *The adaptive Lasso and its oracle properties*, Journal of the American Statistical Association, vol.101, no. 476, pp. 1418–1429, 2006.
15. H. Zou, and T. Hastie, *Regularization and variable selection via the elastic net*, J. R. Stat. Soc. Ser. B Stat. Methodol, vol. 67, no. 2, pp. 301-C320, 2005.