Comparative Study of LASSO, Ridge Regression, Preliminary Test and Stein-type Estimators for the Sparse Gaussian Regression Model

A. K. Md. Ehsanes Saleh [1,*], B. M. Golam Kibria [2], Florence George [2,†]

[1]*Department of Mathematics and Statistics, Carleton University, Canada*
[2]*Department of Mathematics and Statistics, Florida International University, U.S.A*

Abstract This paper compares the performance characteristics of penalty estimators, namely, LASSO and ridge regression (RR), with the least squares estimator (LSE), restricted estimator (RE), preliminary test estimator (PTE) and the Stein-type estimators. Under the assumption of orthonormal design matrix of a given regression model, we find that the RR estimator dominates the LSE, RE, PTE, Stein-type estimators and LASSO estimator uniformly, while, similar to [17], neither LASSO nor LSE, PTE and Stein-Type estimators dominates the other. Our conclusions are based on the analysis of L_2-risks and relative risk efficiencies (RRE) together with the RRE related tables and graphs.

Keywords Dominance, Efficiency, LASSO, Penalty estimator, Pre-test and Stein-type estimators, Risk function.

1. Introduction

It is well-known that the "least squares estimators (LSE)" in linear models, are unbiased with minimum variance characteristics. But data analysts point out some deficiency of LSE with regards to "prediction accuracy" and "interpretation". To overcome these two important concerns, [32] proposed a new popular and exciting penalty estimator, called, *least absolute shrinkage and selection operator (LASSO)*. It defines a continuous shrinking operation that can produce coefficients that are exactly 0 and competitive with "subset selection" and "ridge regression" retaining the good features of both of them. LASSO simultaneously estimates and selects the coefficients of a given model.

However, there are many shrinkage estimators such as "preliminary test (PT)" and Stein-type estimators (SE) in the literature. They do not select coefficients but only shrinks towards a target value.

There is an extensive literature of the preliminary test and Stein-type estimators. Most recent one is documented in [25]. Due to the immense impact of Stein's approach on "point estimation", scores of technical papers appears in the literature in various areas of applications.

In 1970, Hoerl and Kennard introduced the "ridge regression" estimator which opened the door for "penalty estimators" based on the [33]. Ridge regression combats the problem of multicollinearity in the linear models and is the precursor of the problem of estimation and selection of variables. Ridge regression (RR) methodology is

---

[*]A. K. Md. Ehsanes Saleh is a Distinguished Research Professor in School of Mathematics and Statistics, Carleton University, Ottawa, Canada K1S 5B6, Email: esaleh@math.carleton.ca
[†]B. M. Golam Kibria is a Professor and Florence George is an Associate Professor in the Department of Mathematics and Statistics,Florida International University, Modesto A. Maidique Campus Miami, FL 33199, USA,Email: kibriag@fiu.edu

a minimization of the least squares criterion subjects to L_2-penalty. Thus, the estimation subject to the penalty function was born.

LASSO is related to the estimators, such as, non-negative garotte by [8], smoothly clipped absolute derivation (SCAD) estimator by [11] and [13], elastic net by [36], adaptive LASSO by [35], hard threshold LASSO by [6] and many other versions. A general form of extension of LASSO type estimation called *the bridge estimation* by [14] 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 relative to the LSE, RE, preliminary test (PTE), James-Stein estimator (JSE) and positive rule estimator (PRE). The question of comparison between the ridge regression (first discovery of penalty estimator) and Stein-Type estimator is well known and was initiated by [10] among others yet incomplete. So far in the literature, we find scattered and partial simulated comparisons giving unfinished informations. Thus, we decided to peruse this comparison a fresh as new in terms of the L_2-risk to find their special characteristics relative to the preliminary test and Stein-Type estimators analytically to verify the conclusions by [32] and [17], who compares LASSO and Stein-Type estimators among others.

For details on preliminary test and Stein-type estimators we refer our readers to [25], [29],[30], [15], [1],[2][3], [26], [21] [21], [24], [12], [27] and very recently [28] among others. An important characteristic of LASSO is that it provides simultaneous estimation and selection of coefficients in a linear models and can be applied when the dimension of the parameters space exceeds the dimension of the sample space, while the Stein-type estimation restricts the dimension of the parameters space below the dimension of the sample space. This paper points the useful aspects of LASSO and ridge regression estimators as well as limitations as found in other papers. Conclusions are obtained based on the lower bound of the L_2-risk of the LASSO estimator provided by [9]. The comparison of seven estimators discussed here are not based on simulation experiments but based on the mathematical analysis.

Since, we are comparing several estimators where some are linear and some are non-linear, we create a level playing field by assuming orthonormal design matrix which is generally relevant for empirical practices. The main objective of this paper is to compare analytically the performance of LASSO and ridge regression estimators with LSE, RE, PTE, SE and PRSE in the sense of smaller L_2-risk.

The organization of the paper is as follows: Section 2 discusses various estimators and their risk expressions in the canonical regression models. Section 3 contains details of analysis of the relative efficiencies of the estimators with tables and graphs of the findings. Conclusions are provided in section 4.

2. Linear Model and the Estimators

Consider the multiple linear regression model in two forms

$$Y = X\beta + e, \quad \text{and} \quad Y = X_1\beta_1 + X_2\beta_2 + e, \tag{1}$$

where $X'X = I_p$ is an identity matrix, $\bar{\mathbf{x}} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_p) = \mathbf{0}$, $Y = (y_1, y_2, \ldots, y_n)'$ is an n- vector of responses, X is an $n \times p$ design matrix, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$ is an p-vector of regression coefficients and $e = (e_1, e_2, \ldots, e_n)'$ is an n- vector of errors following the $N(0, \sigma^2 I_n)$ distribution with known σ^2. The second form of the model (1), arises by partitioning $\beta = (\beta_1^T, \beta_2^T)^T$ and $X = (X_1, X_2)$, $(p = p_1 + p_2)$ where β_1 may stand for the main effects and β_2 for the interactions which may be insignificant and one is interested in the estimation main effect when β_2 is suspected to be zero (sparsity condition). It is well known that the least squares estimator (LSE) of β under the assumed conditions is given by

$$\tilde{\beta}_n = X'Y = (\tilde{\beta}_{1n}^T, \tilde{\beta}_{2n}^T)^T. \tag{2}$$

The LSE is the best linear unbiased estimator (BLUE) of β. Under normal theory $\tilde{\beta}_n \sim N_p(\beta, \sigma^2 I_p)$. Equivalently $(\tilde{\beta}_{1n}^T, \tilde{\beta}_{2n}^T)^T \sim N_p\left(\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}, \sigma^2 \begin{pmatrix} I_{p_1} & 0 \\ 0 & I_{p_2} \end{pmatrix}\right)$. We made the above simplified assumptions in order to create level playing field to compare a diverse set of shrinkage estimators of $\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$. These assumptions also

produces insight about the nature of the shrinkage that can be gleaned from the orthonormality of the design matrix. From now on we shall designate the LSE, $\tilde{\beta}_n = (\tilde{\beta}_1^T, \tilde{\beta}_2^T)^T$ as the *unrestricted estimator* (UE) of $\beta = (\beta_1^T, \beta_2^T)^T$ and $\hat{\beta}_{2n} = 0$ as the *restricted estimator* (RE) of β_2 so that the restricted estimator of $\beta_R = (\beta_1^T, 0^T)^T$ is $\hat{\beta}_n = (\tilde{\beta}_{1n}^T, 0^T)^T$. We shall study the quantitative characteristics of the estimators based on the L_2-risk defined by

$$R(\beta_n^* : I_p) = E||\beta_n^* - \beta||^2, \tag{3}$$

where β_n^* is any estimator of β. Based on (2) and (3), we see at once that the bias and the risk of LSE and RE are given respectively by,

$$
\begin{array}{rclcl}
b_1(\tilde{\beta}_{jn}) & = & 0 & \text{and} & R_1(\tilde{\beta}_{jn} : I_{p_1}) = \sigma^2 p_j, \quad j = 1, 2, \\
b_2(\hat{\beta}_{2n}) & = & -\beta_2 & \text{and} & R_2(\hat{\beta}_{2n}, I_{p_2}) = \sigma^2 \Delta^2,
\end{array} \tag{4}
$$

From equation (4), we easily find that

$$
\begin{array}{rcl}
R_1(\tilde{\beta}_n : I_p) & = & \sigma^2(p_1 + p_2) \\
R_2(\hat{\beta}_n : I_p) & = & \sigma^2(p_1 + \Delta^2)
\end{array} \tag{5}
$$

where $\Delta^2 = \frac{\beta_2' \beta_2}{\sigma^2}$ is the divergence parameter which measures the distance of $\Delta = (\Delta_{21}, \Delta_{22}, .., \ldots, \Delta_{2p_2})'$ form the origin in the R^{p_2}-space given by

$$\Delta^2 = \Delta_{21}^2 + \Delta_{22}^2 + \ldots + \Delta_{2p_2}^2, \quad \Delta_{2j} = \frac{\beta_{2j}}{\sigma}, \quad j = 1, 2, \ldots, p_2 \tag{6}$$

All our analysis will be based on $\Delta^2 \in R^+$ specifying the parameter space instead of $\Delta = (\Delta_{21}, \Delta_{22}, .., \ldots, \Delta_{2p_2})' \in R^{p_2}$. Our basic interest is to consider several shrinkage estimators which shrink towards the origin, 0. Accordingly we first, consider the *preliminary test estimator* (PTE) as

$$\hat{\beta}_n^{PT} = \left(\tilde{\beta}_{1n}^T, \tilde{\beta}_{2n}^T I(\mathcal{L}_n > c_\alpha)\right), \tag{7}$$

where $I(A)$ is the indicator function of the set A, $\mathcal{L}_n = \frac{\tilde{\beta}_{2n}' \tilde{\beta}_{2n}}{\sigma^2}$ is the test statistic for testing $H_0 : \beta_2 = 0$ and $c_\alpha = \chi_p^2(\alpha)$ is the upper α^{th} percentile of the null distribution of $\mathcal{L}_n$. It is known that $\mathcal{L}_n$ follows a non-central chi-square distribution with p_2 degrees of freedom (D.F.) and non-centrality parameter, Δ^2, which we defined as the divergence parameter. PTE is a discrete process and "keeps" or "kills" the estimators $\tilde{\beta}_n$ or $\hat{\beta}_n$. The bias-vector and L_2-risk of the PTE are respectively given by

$$
\begin{array}{rcl}
b_3(\hat{\beta}_n^{PT}) & = & \left(0^T, -\beta_2^T H_{p_2+2}\left(c_\alpha; \Delta^2\right)\right)
\end{array}
$$

$$
\begin{array}{rcl}
\text{and} \quad R_3(\hat{\beta}_n^{PT}; I_p) & = & \sigma^2 p_1 + \sigma^2 \left[p_2 - p_2 H_{p_2+2}\left(c_\alpha; \Delta^2\right)\right. \\
& + & \left. \Delta^2 \left\{2H_{p_2+2}\left(c_\alpha; \Delta^2\right) - H_{p_2+4}\left(c_\alpha^*; \Delta^2\right)\right\}\right]
\end{array} \tag{8}
$$

where H_ν-functions represent the CDF of the non-central chi-square distribution with DF ν and noncentrality parameter Δ^2 and c_α is the upper α^{th} percentile of the central chi-square distribution with ν degrees of freedom. A continuous version of PTE is the James-Stein type estimator (JSE) is defined by

$$\hat{\beta}_n^{JS} = \left(\tilde{\beta}_{1n}^T, \tilde{\beta}_{2n}^T(1 - (p_2 - 2)\mathcal{L}_n^{-1}), \quad (p_2 \geq 3)\right). \tag{9}$$

The bias and L_2-risk function of the JSE are given respectively by

$$
\begin{array}{rcl}
b_4(\hat{\beta}_n^{JS}) & = & \left(0, -\beta_2^T(p_2 - 2)E\left[\chi_{p_2+2}^{-2}(\Delta^2)\right]\right)
\end{array}
$$

$$
\begin{array}{rcl}
\text{and} \quad R_4(\hat{\beta}_n^{JS}; I_p) & = & \sigma^2 p_1 + \sigma^2 \left[p_2 - (p_2 - 2)^2 E\left[\chi_{p_2}^{-2}(\Delta^2)\right]\right],
\end{array} \tag{10}
$$

where $\chi_\nu^2(\Delta^2)$ is the standard non-central chi-square variable with ν DF and noncentrality parameter Δ^2. The main characteristic of the JSE is the reduction of L_2-risk. We obtain the JSE as a simple modification of PTE and depends

on the test-statistic, $\mathcal{L}_n$. See [25] for details. The JSE has the property of "over-shrinkage" beyond the target vector resulting in changes of sign. This is due to factor $(1 - (p_2 - 2)\mathcal{L}_n^{-1})$ whose absolute value may exceed unity. This change of value effects the interpretation. Hence, we consider the Positive-Rule Stein type shrinkage estimator (PRSE), $\hat{\beta}_n^{S+}$ defined by

$$\hat{\beta}_n^{S+} = \left(\tilde{\beta}_{1n}^T, \tilde{\beta}_{2n}^T (1 - (p_2 - 2)\mathcal{L}_n^{-1}) I(\mathcal{L}_n > (p_2 - 2)) \right). \tag{11}$$

The bias and L_2-risk function of PRSE are respectively given by

$$
\begin{aligned}
b_5(\hat{\beta}_n^{S+}) &= \left(0^T, -\beta_2^T \left\{ E\left[1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta^2) \right) I(\chi_{p_2+2}^2(\Delta^2) < p_2 - 2) \right] \right. \\
&\quad - \left. H_{p_2+2}(p_2 - 2; \Delta^2) \right\} \right)
\end{aligned}
$$

and
$$
\begin{aligned}
R_5(\hat{\beta}_n^{S+}; I_p) &= R_4(\hat{\beta}_n^{JS}; I_p) - p_2\sigma^2 E\left[\left(1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta^2) \right)^2 I\left(\chi_{p_2+2}^2(\Delta^2) < p_2 - 2 \right) \right] \\
&\quad + \sigma^2\Delta^2 \left\{ 2E\left[\left(1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta^2) I(\chi_{p_2+2}^2(\Delta^2)) < p_2 - 2) \right) \right] \right. \\
&\quad - \left. E\left[\left(1 - (p_2 - 2)\chi_{p_2+4}^{-2}(\Delta^2)^2 I(\chi_{p_2+4}^2(\Delta^2)) < p_2 - 2) \right) \right] \right\}
\end{aligned}
\tag{12}
$$

where
$$E\left[\chi_{p_2+2}^{-2}(\Delta^2)) I(\chi_{p_2+2}^2(\Delta^2) < p_2 - 2) \right]$$

is the truncated expectation of the reciprocal of a noncentral χ^2-distribution with $p_2 + 2$ degrees of freedoms and noncentrality parameter $\frac{1}{2}\Delta^2$.

Next, we consider the basic penalty estimator called the ridge regression (RR) estimator ([18]) and given by

$$\hat{\beta}_n^{RR} = \frac{1}{1+k}\tilde{\beta}_n. \tag{13}$$

This estimator is not scale invariant. If scales used to express the individual predictor variables are changed, then the ridge coefficients do not change inversely proportional to the changes in the variable scales. This ridge regression gives constant shrinkage, $\frac{1}{1+k}$. The bias and L_2-risk expression are given respectively by

$$
\begin{aligned}
b_6(\hat{\beta}_n^{RR}) &= -\frac{k}{1+k}\beta \\
\text{and} \quad R_6(\hat{\beta}_n^{RR}; I_p) &= \frac{\sigma^2}{(1+k)^2}[p + k^2\Delta^2].
\end{aligned}
\tag{14}
$$

Notice that the estimator depends on the unknown tuning parameter, k. In general ridge regression estimator combats the multicollinearity problem if the X-matrix is non-orthogonal. In this case, it is a scaled version of the LSE. It tends to null vector (0), as $k \to \infty$ and for $k = 0$, it reduces to the LSE.

It can be shown that $R_6(\hat{\beta}_n^{RR}; I_p) \leq R_1(\tilde{\beta}_n; I_p)$, whenever, $k \in (0, k_0]$ with $k_0 = \frac{p\sigma^2}{\beta'\beta} = p\Delta^{-2}$. Thus $R_6(\hat{\beta}_n^{RR}; I_p)$ in (13) at $k = k_0$ becomes,

$$R_6(\hat{\beta}_n^{RR}; I_p) = \sigma^2 \frac{p\Delta^2}{(p + \Delta^2)}.$$

However, we shall consider the ridge regression estimator as

$$\tilde{\beta}_n^{RR}(k) = \begin{pmatrix} \tilde{\beta}_{1n} \\ \tilde{\beta}_{2n}(k) \end{pmatrix} \quad \text{where} \quad \tilde{\beta}_{2n}(k) = \frac{1}{1+k}\tilde{\beta}_{2n} \tag{15}$$

for our purpose to be consistent wuth the assumption of sparsity, ie. $\beta_2 = 0$ (see Tibshirani, 1996).

Finally, we consider the LASSO (Least absolute shrinkage and selection operator) estimator due to [32] which has gone viral in the statistical literature due to its applicability in data-analysis for linear models unlike other estimators. It shrinks some coefficients and sets others to 0 and hence tries to retain good properties of subset

selection and ridge regression. The literature on LASSO related penalty estimator shows that mostly they have been studies under "orthonormal" set-up of the design matrix X. See for example [32], [9], [11] and [35] among many others. We use the orthonormality of the design matrix in our study too.

The LASSO estimator have been defined as given below,

$$\hat{\beta}_n^L(\lambda) = (\hat{\beta}_{1n}^L(\lambda), \hat{\beta}_{2n}^L(\lambda), \dots \hat{\beta}_{pn}^L(\lambda))' \tag{16}$$

where

$$\hat{\beta}_{jn}^L(\lambda) = sgn(\tilde{\beta}_{jn}) \left(|\tilde{\beta}_{jn}| - \lambda\sigma\right) I\left(|\tilde{\beta}_{jn}| > \lambda\sigma\right), \quad j = 1, 2, \dots, p. \tag{17}$$

Here, λ is the tuning parameter (threshold parameter) and according to [32], LASSO is obtained by minimizing

$$(Y - X\beta)'(Y - X\beta) + \lambda\sigma 1_p'|\beta|; \quad |\beta| = (|\beta_1|, |\beta_2|, \dots, |\beta_p|)' \tag{18}$$

which provides simultaneously estimation and selection of the components of β-vector.

Our aim is the estimation of β under L_2-risk given by (3). For this, we consider the family of diagonal linear projections (DP),

$$T_{DP}(\tilde{\beta}_n, \tau)) = (\tau_j \hat{\beta}_{jn}^L(\lambda)|j = 1, 2, \dots, p)^T, \quad \tau_j \in (0, 1) \tag{19}$$

This estimator 'keeps' or 'kills' a parameter, $\tilde{\beta}_{jn}$, ie, it does subset selection. Now, we incur a risk σ^2 if we use $\tilde{\beta}_{jn}$, and a risk β_j^2 if we sue the estimator 0 instead. Hence, our ideal choice is $I(|\beta_j| > \sigma)$ for τ_j, that is keep all those predictors whose true value is more than the noise level, σ^2. These yield the ideal risk, $R_\sigma(DP)$ given by

$$R_\sigma(DP) = \sum_{i=1}^p min(\beta_j^2, \sigma^2) \tag{20}$$

This expression is a lower bound of the risk that we can hope for. If we assume that p_1 of β_j^2's are greater than σ^2 and rest p_2 are zero, then we obtain $R_\sigma(DP) = \sigma^2(p_1 + \Delta^2)$. In this case, the lower bound of the risk of LASSO is given by

$$R_7(\lambda_p, I_p) \geq \sigma^2(p_1 + \Delta^2), \quad \lambda_p = \sqrt{2ln(p)}. \tag{21}$$

We shall use this lower bound to compare LASSO with other estimators. Finally, based on the Theorem 1 of [9]and equation (20), we have the following inequality:

$$\sum_{i=1}^p min(\beta_j^2, \sigma^2) \leq R_7(\hat{\beta}_n^L(\lambda_p) : I_p) \leq (1 + 2ln(p)) \left\{\sigma^2 + \sum_{i=1}^p min(\beta_j^2, \sigma^2)\right\}. \tag{22}$$

Next, we consider the hard threshold estimator (HTE) given by

$$\hat{\beta}_n^{HT}(\lambda) = \{\tilde{\beta}_{jn} I(|\tilde{\beta}_{jn}| > \lambda\sigma)|j = 1, \dots, p\}. \tag{23}$$

[9]shows that asymptotically LASSO comes as close as HTE to the performance of an ideal subset selector-one that uses information about the actual parameter. As such the result (22) holds for $\hat{\beta}_n^{HT}(\lambda_p)$ as well.

Finally, by Theorem 4 of Donoho and Johnstone (1994), the upper bound of (22) for a sequence of $\{\lambda_n^*\}$ of thresholds close to $\lambda_p = \sqrt{2ln(p)}$ is given by

$$R_8(\hat{\beta}_n^{HT}(\lambda) : I_p) \leq (1 + 2ln(p))\{\sigma^2 + \sum_{i=1}^p min(\beta_j^2, \sigma^2)\}, \quad \beta \in R^p \tag{24}$$

(See Tibshirani (1996), Eqn 16).

Thus, we see that

$$\sigma^2 \sum_{j=1}^{p} min(\Delta_j^2, 1) \leq R_7(\hat{\beta}_n^L(\lambda_p) : I_p) \leq \sigma^2(1 + 2ln(p))\{1 + \sum_{j=1}^{p} min(\Delta_j^2, 1)\}. \tag{25}$$

Same result holds for $\hat{\beta}_n^{HT}(\lambda_p)$. One may notice that the choice $\lambda_p = \sqrt{2ln(p)}$ gives smallest asymptotic L_2-risk for both subset selection estimator as well as LASSO estimator when we consider competing estimators.

3. Analysis of Dominance Properties of the Estimators

In this section, we compare and contrast the L_2-risks of the seven estimators discussed in section 2.

First, we note that LASSO have been proposed by [32]for shrinkage and selection of coefficients for linear and generalized regression models. The LASSO does not focus on subsets but rather defines a continuous shrinkage operation that can produce coefficients that are exactly 0 and competitive with subset selection and ridge regression in terms of prediction accuracy.

Secondly, LASSO performs best at a pont (see [7]. In our case for $\Delta^2 = 0$, it performs best among all estimators except ridge regression.

Thirdly, LASSO enjoys the "Oracle Properties: (see [35] since the design matrix is orthogonal.

In his pioneering paper,[32]examined the relative merits of the subset selection, ridge regression and the LASSO in three different scenarios: (a) Small number of large coefficients -subset selection does the best here, the LASSO not quite as well as ridge regression does quite poorly.

(b) Small to moderate numbers of moderate-size coefficients -LASSO does 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, [17] 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 LSE or Stein-Type estimators uniformly dominate one other.

(ii) Via simulation studies, he concludes that LASSO estimation is particularly sensitive to coefficient parameterization and for a significant portion of the parameter space, LASSO has higher L_2-risk than the LSE.

[17]did not specify the regions where one estimator or the other has lower L_2-risk. In his analysis, [17] used the normalized L_2-risk bounds (NRB) to arrive at his conclusion.

3.1. Comparison of LASSO and LSE

Consider the solution for LASSO. In particular, suppose that the $p_1(< p)$ coefficients satisfy the condition $\beta_j^2 > \sigma^2$ and remaining p_2 coefficients are 0 (zero). In this case, the L_2-risk difference of LSE and LASSO is given by

$$R_1(\tilde{\beta}_n : I_p) - R_7(\hat{\beta}_n^L : I_p) = \sigma^2[(p_1 + p_2) - (p_1 + \Delta^2)] = \sigma^2(p_2 - \Delta^2) \tag{26}$$

Hence, LASSO outperforms the LSE whenever

$$0 \leq \Delta^2 \leq p_2 \tag{27}$$

Otherwise, LSE outperforms LASSO in the interval (p_2, ∞). Hence, neither LSE nor LASSO outperform the other. Thus, by (20), we have the relative risk efficiency (RRE)

$$\frac{p}{(1 + 2ln(p))\{1 + \sum_{j=1}^{p} min(\Delta_j^2, 1)\}} \leq RRE[LASSO : LSE] \leq \frac{p}{\sum_{j=1}^{p} min(\Delta_j^2, 1)}. \tag{28}$$

Table 3.1: Upper bounds of the efficiencies for some selected values of (p_1, p_2) **and** $\Delta^2 = 0$

p	$\sqrt{2ln(p)}$	p_1				
		1	2	3	4	5
10	2.145	10	5	3.3	2.5	2.0
20	2.448	20	10	6.6	5.0	4.0
30	2.608	30	15	10.0	7.5	6.0
40	2.716	40	20	13.3	10.0,	8.0
60	2.862	60	30	20.0,	15.0	12.0
128	3.115	128	64	42.6,	32.0	25.6
256	3.330	256	128	85.3	64.0	51.2
512	3.532	512	256	170.6	128.0	102.4
1024	3.723	1024	512	341.3	256.0	203.8

which is a function of Δ^2. If $\Delta^2 = 0$, we get the lower bound $\frac{p}{(1+2ln(p))\{1+p_1)\}}$ and upper bound $\frac{p}{p_1} = 1 + \frac{p_2}{p_1}$. In this case, (28) becomes

$$\frac{p}{(1 + 2ln(p))(1 + p_1)} \leq RRE[LASSO : LSE] \leq 1 + \frac{p_2}{p_1} \tag{29}$$

We have displayed only the upper bounds of the $RRE[LASSO : LSE]$ for $\Delta^2 = 0$ and for $p_1 = 1, 2, 3, 4, 5$ and $p = 10, 20, 30, 40, 60, 128, 256, 512, 1024$ in Table 3.1. Some tabular values of (3.4) have been computed for $p_1 = 2, 3, 5, 7$ and $p = 10, 20, 40, 60$ and for selected value of $\Delta^2 \in [0, 100]$ and presented them in Table 3.2.

From the Table 3.1, one can see that for fixed p_1 as p increases the RRE[LASSO: LSE] increases significantly and for fixed p, as p_1 increases, LASSO loses its efficiency quickly.

Since, $p = p_1 + p_2$, then for fixed p_1, p increases if p_2 increases. Thus, we can say that for fixed p_1, if p_2 increases the RRE[LASSO: LSE] increases and fixed p_2 if p_1 increases the RRE[LASSO: LSE] decreases quickly. In general, one can say that $RRE[LASSO : LSE] \geq 1$ depending on the size of p_1 and p_2 when $\Delta^2 = 0$.

On the other hand, if $\Delta^2 = 0$, the two bounds depend on Δ^2 and bounds are decreasing function of Δ^2. Hence, $RRE[LASSO : LSE] \leq \geq 1$, depending on the size of (p_1, p_2, Δ^2). Hence, neither LASSO nor LSE outperform the other uniformly. See Figure 1.

3.2. Comparison of LASSO with RE

For the comparison of LASSO with RE, we consider the risk relative efficiency (RRE) of $\hat{\beta}_n$ relative to LSE given by

$$RRE[\hat{\beta}_n : \tilde{\beta}_n] = \frac{p}{p_1 + \Delta^2} \tag{30}$$

which is a decreasing function of Δ^2 and attains it maximum value $\left(1 + \frac{p_2}{p_1}\right)$ at $\Delta^2 = 0$ and equals unity at $\Delta^2 = p_2$. Thus, $\hat{\beta}_n$ dominates $\tilde{\beta}_n$ in the range $0 \leq \Delta^2 \leq p_2$ and $\tilde{\beta}_n$ dominates $\hat{\beta}_n$ when $\Delta^2 > p_2$. Thus none of the estimators dominate one another uniformly. The p_2 sphere with radius less that p_2 is the parameter space when $\hat{\beta}_n$ dominates $\tilde{\beta}_n$ and outside this sphere $\tilde{\beta}_n$ dominates $\hat{\beta}_n$.

Now, we know that both LASSO and RE perform better than the LSE in the interval $[0, p_2)$ and LSE performs better than both LASSO and RE in the interval (p_2, ∞). It is evident that for a significant proportion of the parameter space both LASSO and RE has higher L_2-risk than the LSE. Thus, we observe the performance characteristics of both LASSO and RE are the same indicating "oracle property" of LASSO under orthonormal design matrix.
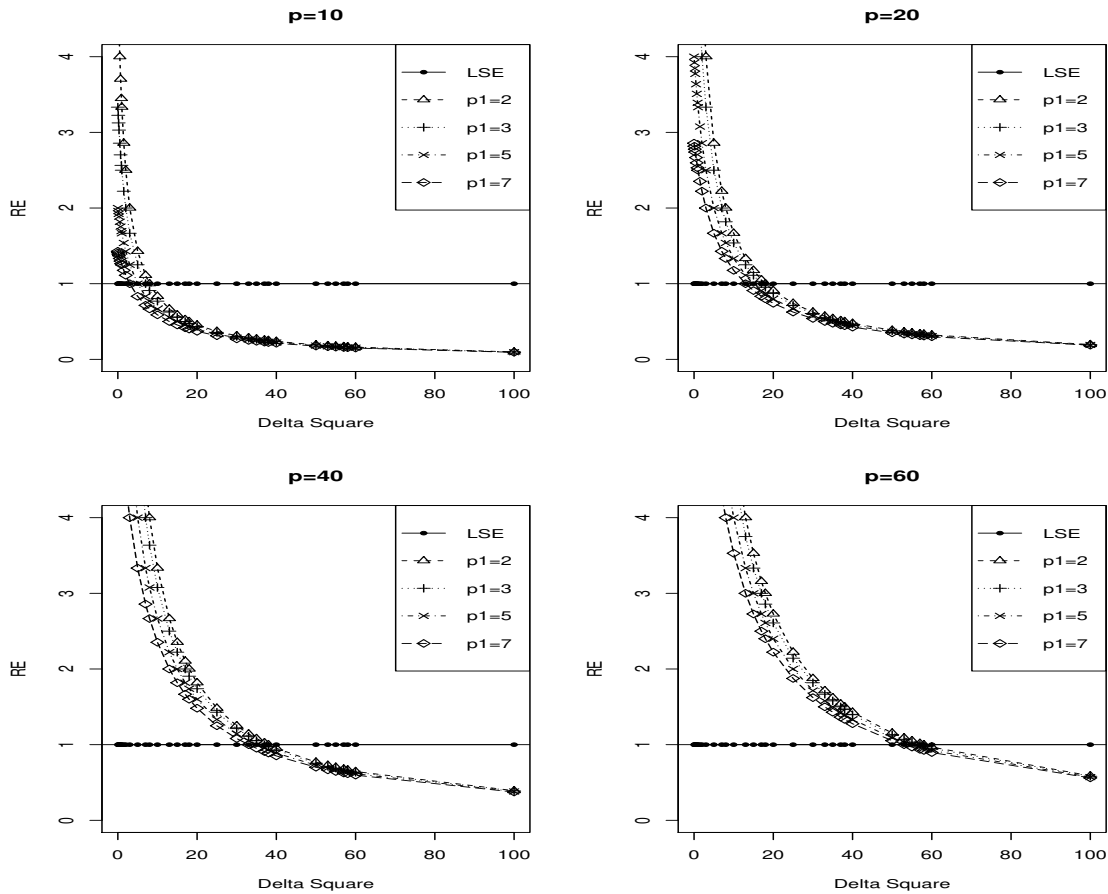
3.3. Comparison of LASSO with the PTE

In this section, we present the comparison of LASSO with the PTE. In this case, the L_2-risk difference is given by

$$
\begin{aligned}
R_3(\hat{\beta}_n^{PT} : I_p) - R_7(\hat{\beta}_n^L(\lambda_p) : I_k) &= \sigma^2 p_1 + \sigma^2 \left[p_2(1 - H_{p_2+2}\left(c_\alpha; \Delta_{(k)}^2\right)) \right. \\
&+ \left. \Delta^2 \left\{ 2H_{p_2+2}\left(c_\alpha; \Delta^2\right) - H_{p_2+4}\left(c_\alpha; \Delta^2\right) \right\} - (p_1 + \Delta^2) \right] \\
&= \sigma^2 \left[p_2(1 - H_{p_2+2}\left(c_\alpha; \Delta^2\right)) - \Delta^2(1 - 2H_{p_2+2}\left(c_\alpha; \Delta^2\right) - H_{p_2+4}\left(c_\alpha; \Delta^2\right)) \right]
\end{aligned}
\tag{31}
$$

Then, LASSO performs better than the PTE whenever,

$$
0 \leq \Delta^2 \leq \frac{p_2(1 - H_{p_2+2}\left(c_\alpha; \Delta^2\right))}{p_2(1 - 2H_{p_2+2}\left(c_\alpha; \Delta^2\right) - H_{p_2+2}\left(c_\alpha; \Delta^2\right))} = \Delta_{PTE}^2
\tag{32}
$$

Otherwise, PTE is better than the LASSO for $\Delta^2 > \Delta_{PTE}^2$. Hence neither PTE nor LASSO dominate the other uniformly for all $\alpha \in (0, 1)$. c

3.4. Comparison of LSE with JSE and PRSE

Now, we consider the comparison of JSE and LSE. It is easy to see that by simple L_2-risk difference

$$
R_1(\tilde{\beta}_n; I_p) - R_4(\hat{\beta}_n^{JS}; I_p) = \sigma^2(p_2 - 2)^2 E\left[\chi_{p_2}^{-2}(\Delta^2) \right] > 0 \quad \forall \Delta^2 \in (0, \infty).
\tag{33}
$$

Hence $\hat{\beta}_n^{JS}$ dominates uniformly over $\tilde{\beta}_n$.

Similarly, the L_2-risk difference of $\hat{\beta}_n^{S+}$ and $\tilde{\beta}_n$ is given by,

$$
\begin{aligned}
& R_4(\hat{\beta}_n^{JS}; I_p) - R_5(\hat{\beta}_n^{S+}; I_p) \\
= \quad & p_2\sigma^2 E\left[\left(1 - (p_2-2)\chi_{p_2+2}^{-2}(\Delta^2)\right)^2 I\left(\chi_{p_2+2}^2(\Delta^2) < p_2-2\right)\right] \\
- \quad & \beta'\beta\left\{E\left[\left(1 - (p_2-2)\chi_{p_2+2}^{-2}(\Delta^2)\right) I\left(\chi_{p_2+2}^2(\Delta^2) < p_2-2\right)\right]\right. \\
- \quad & \left. E\left[\left(1 - (p_2-2)\chi_{p_2+4}^{-2}(\Delta^2)\right)^2 I\left(\chi_{p_2+4}^2(\Delta^2) < p_2-2\right)\right]\right\} \leq 0
\end{aligned}
\tag{34}
$$

Thus we have the following identity

$$
R_5(\hat{\beta}_n^{S+}; I_p) \leq R_4(\hat{\beta}_n^{JS}; I_p) \leq R_1(\tilde{\beta}_n; I_p) \quad \forall \Delta^2 \in R^+.
$$

It is clear that JSE and PRSE gains are highest when the coefficients are small.

3.5. Comparison of LASSO with JSE and PRSE

How is the performance of LASSO compared to JSE? For this first we consider the L_2RE differences of JSE and the upper bound of L_2-risk of LASSO ($p_2 \geq 3$) given by

$$
\begin{aligned}
R_7(\hat{\beta}_n^L(\lambda_p) : I_p) - R_4(\hat{\beta}_n^{JS} : I_p) &= \sigma^2\left[(p_1+\Delta^2) - p + (p_2-2)^2 E\left(\chi_{p_2}^{-2}(\Delta^2)\right)\right] \\
&= \sigma^2\left[\Delta^2 - p_2 + (p_2-2)^2 E\left(\chi_{p_2}^{-2}(\Delta^2)\right)\right]
\end{aligned}
\tag{35}
$$

Hence, LASSO outperform JSE, whenever,

$$
\Omega_{\Delta^2} = \left\{\Delta^2 : 0 \leq \Delta^2 \leq p_2 - (p_2-2)^2 E\left(\chi_{p_2}^{-2}(\Delta^2)\right)\right\}
\tag{36}
$$

Otherwise, JSE outperforms LASSO in $\bar{\Omega}_{\Delta^2}$. Hence neither LASSO nor JSE dominate the others uniformly in Δ^2. Now, we will demonstrate some properties of LASSO and JSE. First, we compare the RRE[JSE:LSE] and RRE[LASSO: LSE]. They are given by

$$
\begin{aligned}
RRE[JSE : LSE] &= p\left\{p - (p_2-2)^2 E\left(\chi_{p_2}^{-2}(\Delta^2)\right)\right\}^{-1} \\
RRE[LASSO : LSE] &\geq \frac{p}{[(p_1+\Delta^2]}
\end{aligned}
\tag{37}
$$

respectively. Under $\Delta^2 = 0$, the $RRE[JSE : LSE] = 1 + \frac{p_2-2}{p_1+2}$. Thus for,

$$
\begin{aligned}
For\ p_1 = p_2 = 5, \ & RRE[JSE : LSE] = 1.429 \\
For\ p_1 = p_2 = 10, \ & RRE[JSE : LSE] = 2.142 \\
For\ p_1 = p_2 = 15, \ & RRE[JSE : LSE] = 2.857
\end{aligned}
$$

We can see that under $\Delta^2 = 0$ as p_2 increase for fixed p_1, the RRE[JSE:LSE] increases too. As a function of Δ^2, RRE[JSE:LSE] decrees from $\left(1 + \frac{p_2-2}{p_1+2}\right)$ towards unity as $\Delta^2 \to \infty$. For any fixed value of Δ^2 (say, Δ_0^2), RRE[JSE:LSE] increases with p_2 for fixed p_1. Now we observe that the RRE difference, $RRE[LASSO : LSE] - RRE[JSE : LSE] \geq 1$, whenever Ω_{Δ^2} holds.

This means that the region in which LASSO performs better than the JSE is given by (36), otherwise JSE performs better than the LASSO.

Now, we consider the comparison of LASSO and positive-rule Stein estimator and consider the L_2-risk-difference given by

$$
\begin{aligned}
R_7(\hat{\beta}_n^L(\lambda_p) : I_p) - R_5(\hat{\beta}_n^{S+} : I_p) &= \sigma^2\left[(p_1+\Delta^2) - (p_1+p_2) + (p_2-2)^2 E\left(\chi_{p_2}^{-2}(\Delta^2)\right)\right] \\
&+ p_2\sigma^2 E\left[\left(1 - (p_2-2)\chi_{p_2+2}^{-2}(\Delta^2)\right)^2 I\left(\chi_{p_2+2}^2(\Delta^2) < p_2-2\right)\right]
\end{aligned}
$$

Table 3.2: Efficiency of LASSO & RE as a function of Δ^2 for different (p_1, p_2)

Δ^2	LSE	$p_1 = 2$				$p_1 = 3$			
		p=10	p=20	p=40	p=60	p=10	p=20	p=40	p=60
0.00	1.00	5.00	10.00	20.00	30.00	3.33	6.67	13.33	20.00
0.10	1.00	4.76	9.52	19.05	28.57	3.23	6.45	12.90	19.35
0.20	1.00	4.55	9.09	18.18	27.27	3.12	6.25	12.50	18.75
0.30	1.00	4.35	8.70	17.39	26.09	3.03	6.06	12.12	18.18
0.50	1.00	4.00	8.00	16.00	24.00	2.86	5.71	11.43	17.14
0.70	1.00	3.70	7.41	14.81	22.22	2.70	5.41	10.81	16.22
0.90	1.00	3.45	6.90	13.79	20.69	2.56	5.13	10.26	15.38
1.00	1.00	3.33	6.67	13.33	20.00	2.50	5.00	10.00	15.00
1.50	1.00	2.86	5.71	11.43	17.14	2.22	4.44	8.89	13.33
2.00	1.00	2.50	5.00	10.00	15.00	2.00	4.00	8.00	12.00
3.00	1.00	2.00	4.00	8.00	12.00	1.67	3.33	6.67	10.00
5.00	1.00	1.43	2.86	5.71	8.57	1.25	2.50	5.00	7.50
7.00	1.00	1.11	2.22	4.44	6.67	1.00	2.00	4.00	6.00
8.00	1.00	1.00	2.00	4.00	6.00	0.91	1.82	3.64	5.45
10.00	1.00	0.83	1.67	3.33	5.00	0.77	1.54	3.08	4.62
13.00	1.00	0.67	1.33	2.67	4.00	0.62	1.25	2.50	3.75
15.00	1.00	0.59	1.18	2.35	3.53	0.56	1.11	2.22	3.33
17.00	1.00	0.53	1.05	2.11	3.16	0.50	1.00	2.00	3.00
18.00	1.00	0.50	1.00	2.00	3.00	0.48	0.95	1.90	2.86
20.00	1.00	0.45	0.91	1.82	2.73	0.43	0.87	1.74	2.61
25.00	1.00	0.37	0.74	1.48	2.22	0.36	0.71	1.43	2.14
30.00	1.00	0.31	0.62	1.25	1.88	0.30	0.61	1.21	1.82
33.00	1.00	0.29	0.57	1.14	1.71	0.28	0.56	1.11	1.67
35.00	1.00	0.27	0.54	1.08	1.62	0.26	0.53	1.05	1.58
37.00	1.00	0.26	0.51	1.03	1.54	0.25	0.50	1.00	1.50
38.00	1.00	0.25	0.50	1.00	1.50	0.24	0.49	0.98	1.46
40.00	1.00	0.24	0.48	0.95	1.43	0.23	0.47	0.93	1.40
50.00	1.00	0.19	0.38	0.77	1.15	0.19	0.38	0.75	1.13
53.00	1.00	0.18	0.36	0.73	1.09	0.18	0.36	0.71	1.07
55.00	1.00	0.18	0.35	0.70	1.05	0.17	0.34	0.69	1.03
57.00	1.00	0.17	0.34	0.68	1.02	0.17	0.33	0.67	1.00
58.00	1.00	0.17	0.33	0.67	1.00	0.16	0.33	0.66	0.98
60.00	1.00	0.16	0.32	0.65	0.97	0.16	0.32	0.63	0.95
100.00	1.00	0.10	0.20	0.39	0.59	0.10	0.19	0.39	0.58
Δ^2		$p_1 = 5$				$p_1 = 7$			
0.00	1.00	2.00	4.00	8.00	12.00	1.43	2.86	5.71	8.57
0.10	1.00	1.96	3.92	7.84	11.76	1.41	2.82	5.63	8.45
0.20	1.00	1.92	3.85	7.69	11.54	1.39	2.78	5.56	8.33
0.30	1.00	1.89	3.77	7.55	11.32	1.37	2.74	5.48	8.22
0.50	1.00	1.82	3.64	7.27	10.91	1.33	2.67	5.33	8.00
0.70	1.00	1.75	3.51	7.02	10.53	1.30	2.60	5.19	7.79
0.90	1.00	1.69	3.39	6.78	10.17	1.27	2.53	5.06	7.59
1.00	1.00	1.67	3.33	6.67	10.00	1.25	2.50	5.00	7.50
1.50	1.00	1.54	3.08	6.15	9.23	1.18	2.35	4.71	7.06
2.00	1.00	1.43	2.86	5.71	8.57	1.11	2.22	4.44	6.67
3.00	1.00	1.25	2.50	5.00	7.50	1.00	2.00	4.00	6.00
5.00	1.00	1.00	2.00	4.00	6.00	0.83	1.67	3.33	5.00
7.00	1.00	0.83	1.67	3.33	5.00	0.71	1.43	2.86	4.29
8.00	1.00	0.77	1.54	3.08	4.62	0.67	1.33	2.67	4.00
10.00	1.00	0.67	1.33	2.67	4.00	0.59	1.18	2.35	3.53
13.00	1.00	0.56	1.11	2.22	3.33	0.50	1.00	2.00	3.00
15.00	1.00	0.50	1.00	2.00	3.00	0.45	0.91	1.82	2.73
17.00	1.00	0.45	0.91	1.82	2.73	0.42	0.83	1.67	2.50
18.00	1.00	0.43	0.87	1.74	2.61	0.40	0.80	1.60	2.40
20.00	1.00	0.40	0.80	1.60	2.40	0.37	0.74	1.48	2.22
25.00	1.00	0.33	0.67	1.33	2.00	0.31	0.62	1.25	1.88
30.00	1.00	0.29	0.57	1.14	1.71	0.27	0.54	1.08	1.62
33.00	1.00	0.26	0.53	1.05	1.58	0.25	0.50	1.00	1.50
35.00	1.00	0.25	0.50	1.00	1.50	0.24	0.48	0.95	1.43
37.00	1.00	0.24	0.48	0.95	1.43	0.23	0.45	0.91	1.36
38.00	1.00	0.23	0.47	0.93	1.40	0.22	0.44	0.89	1.33
40.00	1.00	0.22	0.44	0.89	1.33	0.21	0.43	0.85	1.28
50.00	1.00	0.18	0.36	0.73	1.09	0.18	0.35	0.70	1.05
53.00	1.00	0.17	0.34	0.69	1.03	0.17	0.33	0.67	1.00
55.00	1.00	0.17	0.33	0.67	1.00	0.16	0.32	0.65	0.97
57.00	1.00	0.16	0.32	0.65	0.97	0.16	0.31	0.62	0.94
58.00	1.00	0.16	0.32	0.63	0.95	0.15	0.31	0.62	0.92
60.00	1.00	0.15	0.31	0.62	0.92	0.15	0.30	0.60	0.90
100.00	1.00	0.10	0.19	0.38	0.57	0.09	0.19	0.37	0.56

$$
\begin{aligned}
- \quad & \sigma^2 \Delta^2 \left\{ 2E \left[\left(1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta^2) I(\chi_{p_2+2}^2(\Delta^2)) < p_2 - 2) \right) \right] \right. \\
- \quad & \left. E \left[\left(1 - (p_2 - 2)\chi_{p_2+4}^{-2}(\Delta^2)^2 I(\chi_{p_2+4}^2(\Delta^2)) < p_2 - 2) \right) \right] \right\} \\
= \quad & \sigma^2 (\Delta^2 A - B)
\end{aligned}
\tag{38}
$$

where

$$
\begin{aligned}
A &= 1 - \left\{ 2E \left[\left(1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta^2) I(\chi_{p_2+2}^2(\Delta^2)) < p_2 - 2) \right) \right] \right. \\
&\quad - \left. E \left[\left(1 - (p_2 - 2)\chi_{p_2+4}^{-2}(\Delta^2)^2 I(\chi_{p_2+4}^2(\Delta^2)) < p_2 - 2) \right) \right] \right\} \\
B &= \sigma^2 \left\{ p_2 - (p_2 - 2)^2 E\left(\chi_{p_2}^{-2}(\Delta^2) \right) - p_2 E \left[\left(1 - (p_2 - 2)\chi_{p_2+2}^{-2}(\Delta^2) \right)^2 I \left(\chi_{p_2+2}^2(\Delta^2) < p_2 - 2 \right) \right] \right\}
\end{aligned}
\tag{39}
$$

Now, LASSO outperforms PRSE whenever $\Omega_{\Delta^2}^*$ holds, where

$$
\Omega_{\Delta^2}^* = \left\{ \Delta^2 : 0 \le \Delta^2 \le \frac{B}{A} \right\}
\tag{40}
$$

otherwise, PRSE outperform LASSO on the complementary set $\bar{\Omega}_{\Delta^2}^*$. Hence neither LASSO nor PRSE uniformly dominates the other in Δ^2. See Tables 3.3-3.4 for numerical efficiency of LASSO, JSE and PRSE.

3.6. Comparison of Ridge Regression vs LSE and RE estimators

First, recall that the bias and L_2-risk expression of ridge regression estimator are given respectively by

$$
b_6(\hat{\beta}_n^{RR}) = \begin{pmatrix} 0 \\ -\frac{k}{1+k}\beta_2 \end{pmatrix} \quad \text{and} \quad R_6(\hat{\beta}_n^{RR}; I_p) = \sigma^2 p_1 + \frac{\sigma^2}{(1+k)^2}[p_2 + k^2\Delta^2],
\tag{41}
$$

where k is the tuning parameter and $\Delta^2 = \frac{\beta_2'\beta_2}{\sigma^2}$. We first prove the following theorem on the dominance of the ridge estimator over the LSE given by.

Theorem 3.2.1 There exist always a $k > 0$ in the range $0 < k < k^* = \frac{\sigma^2}{\beta'\beta}$ such that the ridge regression estimator $\hat{\beta}_n^{RR}$ has smaller risk than the LSE.

Proof: It is obvious that for $k = 0$, $R_6(\hat{\beta}_n^{RR}; I_p) = p\sigma^2$, which is the risk of the LSE. Now, consider the two terms of $R_6(\hat{\beta}_{2n}^{RR}; I_{p_2}) = \frac{p_2\sigma^2}{(1+k)^2} + \frac{\sigma^2 k^2 \Delta^2}{(1+k)^2}$. The first term is continuous and monotonically decreasing function of k and its derivative w.r.t k approaches ∞ as $k \to 0^+$. The second term is also continuous and monotonically increasing function of k and its derivative tends to zero as $k \to 0^+$. Note that the $\frac{\sigma^2 k^2 \Delta^2}{(1+k)^2}$ converges to 0 as $k \to 0^+$. Note that $\frac{\sigma^2 k^2 \Delta^2}{(1+k)^2}$ converges to Δ^2 as $k \to \infty$. Differentiating w.r.t. k

$$
\frac{\partial R_6(\hat{\beta}_{2n}^{RR}; I_{p_2})}{\partial k} = \frac{2}{(1+k)^3}(k\beta_2'\beta_2 + p_2\sigma^2)
\tag{42}
$$

Thus, a sufficient condition for (42) to be negative is that there exist a value of k in the interval $0 < k < k^*$ such that

$$
k^* = p_2 \frac{\sigma^2}{\beta_2'\beta_2} = p_2 \Delta^{-2}
\tag{43}
$$

Substituting k^* in $R_6(\hat{\beta}_{2n}^{RR}; I_{p_2})$ we obtain

$$
R_6(\hat{\beta}_{2n}^{RR}; I_{p_2}) = \frac{\sigma^2 p_2 \Delta^2}{p_2 + \Delta^2}.
\tag{44}
$$

Table 3.3 Relative Efficiency of the estimators for $p = 10, 20$ and different Δ^2-value
for varying p_2 with fixed p_1

				p=10				
		$\Delta^2 = 0$				$\Delta^2 = 1.00$		
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
LASSO / RE	5.00	3.33	2.00	1.43	3.33	2.50	1.67	1.25
PTE.15	3.12	2.47	1.74	1.34	2.23	1.89	1.45	1.18
PTE.20	2.78	2.27	1.67	1.32	2.05	1.77	1.40	1.16
PTE.25	2.50	2.11	1.60	1.29	1.89	1.67	1.36	1.15
JSE	2.50	2.00	1.43	1.11	2.14	1.77	1.33	1.08
PR	3.04	2.31	1.56	1.16	2.46	1.98	1.42	1.11
Ridge	Inf	Inf	Inf	Inf	11.25	11.43	12.00	13.33
		$\Delta^2 = 5$				$\Delta^2 = 10$		
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LASSO / RE	1.43	1.25	1.00	0.83	0.83	0.77	0.67	0.59
PTE.15	1.20	1.13	1.02	0.94	0.98	0.96	0.94	0.94
PTE.20	1.17	1.11	1.02	0.95	0.99	0.97	0.95	0.95
PTE.25	1.15	1.10	1.02	0.96	0.99	0.98	0.96	0.96
JSE	1.55	1.38	1.15	1.03	1.32	1.23	1.09	1.01
PR	1.63	1.43	1.18	1.03	1.34	1.24	1.09	1.01
Ridge	3.25	3.43	4.00	5.33	2.25	2.43	3.00	4.33
		$\Delta^2 = 20$				$\Delta^2 = 60$		
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LASSO / RE	0.45	0.43	0.40	0.37	0.16	0.16	0.15	0.15
PTE.15	0.98	0.98	0.99	0.99	1.00	1.00	1.00	1.00
PTE.20	0.98	0.99	0.99	1.00	1.00	1.00	1.00	1.00
PTE.25	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00
JSE	1.17	1.12	1.04	1.01	1.06	1.04	1.01	1.00
PR	1.17	1.12	1.05	1.01	1.06	1.04	1.01	1.00
Ridge	1.75	1.93	2.50	3.83	1.42	1.60	2.17	3.50
				p=20				
		$\Delta^2 = 0$				$\Delta^2 = 1.00$		
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
LASSO / RE	10.00	6.67	4.00	2.86	6.67	5.00	3.33	2.50
PTE.15	4.26	3.60	2.76	2.23	3.19	2.82	2.29	1.92
PTE.20	3.57	3.13	2.50	2.08	2.77	2.50	2.10	1.80
PTE.25	3.08	2.76	2.29	1.95	2.45	2.25	1.94	1.70
JSE	5.00	4.00	2.86	2.22	4.13	3.43	2.56	2.04
PR	6.28	4.77	3.22	2.43	4.90	3.93	2.82	2.20
Ridge	Inf	Inf	Inf	Inf	21.11	21.18	21.33	21.54
		$\Delta^2 = 5$				$\Delta^2 = 10$		
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LASSO / RE	2.86	2.50	2.00	1.67	1.67	1.54	1.33	1.18
PTE.15	1.67	1.58	1.43	1.31	1.19	1.16	1.11	1.06
PTE.20	1.55	1.48	1.36	1.26	1.15	1.13	1.09	1.05
PTE.25	1.46	1.40	1.30	1.22	1.13	1.11	1.07	1.04
JSE	2.65	2.36	1.94	1.65	2.03	1.87	1.62	1.43
PR	2.84	2.50	2.02	1.70	2.08	1.91	1.64	1.45
Ridge	5.11	5.18	5.33	5.54	3.11	3.18	3.33	3.54
		$\Delta^2 = 20$				$\Delta^2 = 60$		
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LASSO / RE	0.91	0.87	0.80	0.74	0.32	0.32	0.31	0.30
PTE.15	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00
PTE.20	1.00	1.00	0.99	0.99	1.00	1.00	1.00	1.00
PTE.25	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00
JSE	1.58	1.50	1.36	1.26	1.21	1.18	1.13	1.10
PR	1.59	1.50	1.37	1.26	1.21	1.18	1.13	1.10
Ridge	2.11	2.18	2.33	2.54	1.44	1.51	1.67	1.87

Table 3.4 Relative Efficiency of the estimators for $p = 40, 60$ and different Δ^2-value for varying p_2 with fixed p_1

				p=40				
		$\Delta^2 = 0$				$\Delta^2 = 1.00$		
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
LASSO / RE	20.00	13.33	8.00	5.71	13.33	10.00	6.67	5.00
PTE.15	5.19	4.68	3.90	3.35	4.16	3.82	3.30	2.90
PTE.20	4.17	3.85	3.33	2.94	3.43	3.22	2.86	2.57
PTE.25	3.48	3.27	2.91	2.62	2.93	2.78	2.53	2.31
JSE	10.00	8.00	5.71	4.44	8.12	6.75	5.05	4.03
PR	12.80	9.69	6.53	4.92	9.81	7.88	5.65	4.41
Ridge	Inf	Inf	Inf	Inf	41.05	41.08	41.14	41.21
		$\Delta^2 = 5$				$\Delta^2 = 10$		
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LASSO / RE	5.71	5.00	4.00	3.33	3.33	3.08	2.67	2.35
PTE.15	2.31	2.21	2.04	1.89	1.54	1.51	1.44	1.38
PTE.20	2.04	1.97	1.84	1.73	1.43	1.40	1.35	1.31
PTE.25	1.84	1.79	1.70	1.61	1.35	1.33	1.29	1.25
JSE	4.87	4.35	3.59	3.05	3.46	3.20	2.78	2.46
PR	5.30	4.69	3.81	3.21	3.58	3.30	2.85	2.51
Ridge	9.05	9.08	9.14	9.21	5.05	5.08	5.14	5.21
		$\Delta^2 = 20$				$\Delta^2 = 60$		
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LASSO / RE	1.82	1.74	1.60	1.48	0.65	0.63	0.62	0.60
PTE.15	1.11	1.10	1.08	1.07	1.00	1.00	1.00	1.00
PTE.20	1.08	1.07	1.06	1.05	1.00	1.00	1.00	1.00
PTE.25	1.06	1.06	1.05	1.04	1.00	1.00	1.00	1.00
JSE	2.42	2.29	2.09	1.92	1.52	1.49	1.42	1.37
PR	2.43	2.31	2.10	1.93	1.52	1.49	1.42	1.37
Ridge	3.05	3.08	3.14	3.21	1.72	1.75	1.81	1.88
				p=60				
		$\Delta^2 = 0$				$\Delta^2 = 1.00$		
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
LASSO/ RE	30.00	20.00	12.00	8.57	20.00	15.00	10.00	7.50
PTE.15	5.61	5.19	4.53	4.01	4.66	4.37	3.90	3.51
PTE.20	4.41	4.17	3.75	3.41	3.76	3.58	3.27	3.02
PTE.25	3.64	3.48	3.20	2.96	3.16	3.04	2.83	2.65
JSE	15.00	12.00	8.57	6.67	12.12	10.09	7.55	6.03
PR	19.35	14.63	9.83	7.40	14.74	11.83	8.49	6.61
Ridge	Inf	Inf	Inf	Inf	61.03	61.05	61.09	61.13
		$\Delta^2 = 5$				$\Delta^2 = 10$		
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LASSO / RE	8.57	7.50	6.00	5.00	5.00	4.62	4.00	3.53
PTE.15	2.74	2.65	2.48	2.33	1.83	1.79	1.72	1.66
PTE.20	2.36	2.29	2.17	2.07	1.65	1.62	1.57	1.52
PTE.25	2.09	2.04	1.95	1.87	1.52	1.50	1.46	1.42
JSE	7.10	6.35	5.25	4.47	4.89	4.53	3.95	3.50
PR	7.77	6.88	5.61	4.73	5.10	4.71	4.08	3.60
Ridge	13.03	13.05	13.09	13.13	7.03	7.05	7.09	7.13
		$\Delta^2 = 20$				$\Delta^2 = 60$		
LSE	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
LASSO/ RE	2.73	2.61	2.40	2.22	0.97	0.95	0.92	0.90
PTE.15	1.22	1.21	1.19	1.17	1.00	1.00	1.00	1.00
PTE.20	1.17	1.16	1.15	1.13	1.00	1.00	1.00	1.00
PTE.25	1.13	1.12	1.11	1.10	1.00	1.00	1.00	1.00
JSE	3.25	3.10	2.83	2.60	1.83	1.79	1.72	1.65
PR	3.28	3.12	2.85	2.62	1.83	1.79	1.72	1.65
Ridge	4.03	4.05	4.09	4.13	2.03	2.05	2.09	2.13

For the comparison of RR and LSE, the L_2-risk difference is

$$R_1(\tilde{\beta}_n : I_p) - R_6(\hat{\beta}_{2n}^{RR} : I_{p_2}) = p\sigma^2 - \frac{\sigma^2 p_2 \Delta^2}{p_2 + \Delta^2} = \sigma^2 \left(p_1 + \frac{p_2^2}{p_2 + \Delta^2} \right) > 0. \tag{45}$$

Hence, $R_6(\hat{\beta}_{2n}^{RR} : I_{p_2}) < R_1(\tilde{\beta}_n : I_p) \quad \forall \Delta^2$. Thus RR estimator uniformly dominate LSE for all $\Delta^2 \in (0, \infty)$.

As for the RR estimator versus RE, we consider L_2-risk difference as

$$\sigma^2 \left[p_1 + \Delta^2 - \frac{p_2 \Delta^2}{p_2 + \Delta^2} \right] = \sigma^2 \left(p_1 + \frac{\Delta^4}{p_2 + \Delta^2} \right) > 0. \tag{46}$$

The R.H.S. of the above equation is $\sigma^2 p_1$ for $\Delta^2 = 0$ and positive when $\Delta^2 > 0$. Hence, RR uniformly dominates both LSE and RE for all Δ^2.

3.7. Comparison of Ridge Regression vs. PTE and Stein-Type Estimators

First we compare the RR estimator and PTE. The L_2-risk difference of PTE and RR is

$$\begin{aligned}
&\sigma^2 \left[p_2 - p_2 H_{p_2+2} \left(\chi_{p_2}^2(\alpha); \Delta^2 \right) + \Delta^2 \left\{ 2H_{p_2+2} \left(\chi_{p_2}^2)\alpha); \Delta^2 \right) - H_{p+4} \left(\chi_p^2(\alpha); \Delta^2 \right) \right\} \right] \\
&- \frac{\sigma^2 p_2 \Delta^2}{p_2 + \Delta^2} + \sigma^2 p_1 \\
&= \sigma^2 \left[\frac{p_2^2}{p_2 + \Delta^2} - \left\{ p_2 H_{p_2+2} \left(\chi_{p_2}^2(\alpha); \Delta^2 \right) - \Delta^2 \left(2H_{p_2+2} \left(\chi_{p_2}^2(\alpha); \Delta^2 \right) - H_{p_2+4} \left(\chi_{p_2}^2(\alpha); \Delta^2 \right) \right) \right\} \right]
\end{aligned} \tag{47}$$

For $\Delta^2 = 0$, we get for all α in $(0, 1)$

$$\sigma^2 p_1 + \sigma^2 p_2 \left[1 - H_{p_2+2} \left(\chi_{p_2}^2(\alpha); \Delta^2 \right) \right] \geq 0$$

Then the RR performs better than that of PTE for $\Delta^2 = 0$. Now, consider the case, $\Delta^2 \neq 0$ and $\alpha \in (0, 1)$. In this case the risk of PTE has minimum value $p - p_2 H_{p_2+2} \left(\chi_{p_2}^2(\alpha); \Delta^2 \right)$ at $\Delta^2 = 0$, then increases and crossing the p-line and reaches a maximum then tapers off towards the p-line. On the other hand, the optimum risk of RR has the minimum value 0, and increases monotonically towards the p-line. Hence (47) is non-negative for all $(\alpha, \Delta^2) \in (0, 1) \times R^+$. Thus, RR uniformly performs better than the PTE i.e.

$$R_6(\hat{\beta}_{2n}^{RR} : I_{p_2}) \leq R_3(\hat{\beta}_n^{PT} : I_p) \forall \Delta^2 \in R^+.$$

Now, consider the comparison of RR with JSE. In this case the L_2-risk difference is given by

$$\begin{aligned}
&\sigma^2 p_1 + \sigma^2 \left[p_2 - (p_2 - 2)^2 E \left(\chi_{p_2}^{-2}(\Delta^2) \right) \right] - \frac{p_2 \sigma^2 \Delta^2}{(p_2 + \Delta^2)} \\
&= \sigma^2 \left[p_1 + \frac{p_2^2}{p_2 + \Delta^2} - (p_2 - 2)^2 E \left(\chi_{p_2}^{-2}(\Delta^2) \right) \right].
\end{aligned} \tag{48}$$

Now, $p_1 + \frac{p^2}{p + \Delta^2}$ is a decreasing function of Δ^2 with maximum value of $p = (p_1 + p_2)$ and minimum value "p_1" while, $(p_2 - 2)^2 E \left(\chi_{p_2}^{-2}(\Delta^2) \right)$ is also a decreasing function of Δ^2 with the maximum value $(p_2 - 2)$ at $\Delta^2 = 0$ and minimum value "zero" when $\Delta \to \infty$. Hence (48) is a non-negative and RR dominate JSE uniformly in Δ^2. i.e.,

$$R_6(\hat{\beta}_{2n}^{RR} : I_{p_2}) \leq R_4(\hat{\beta}_n^S : I_p) \quad \forall \Delta^2 \in R^+.$$

Similarly, consider the L_2-risk difference of RR and PRSE as

$$R_5(\hat{\beta}_n^{S+} : I_p) - R_6(\hat{\beta}_n^{RR} : I_p) = \sigma^2 \left(p_1 + \frac{p_2^2}{(p_2 + \Delta^2)} \right) - (p_2 - 2)^2 E \left[\chi_{p_2}^{-2}(\Delta^2) \right]$$

$$
\begin{aligned}
- \quad & p_2 \sigma^2 E \left[\left(1 - (p-2)\chi_{p+2}^{-2}(\Delta^2) \right)^2 I \left(\chi_p^2(\Delta^2) < (p-2) \right) \right] \\
- \quad & \sigma^2 \Delta^2 \left\{ 2 E \left[\left(1 - (p_2-2)\chi_{p_2}^{-2}(\Delta^2) \right) I \left(\chi_{p_2+2}^2(\Delta^2) < p_2 - 2 \right) \right] \right. \\
- \quad & \left. E \left[\left(1 - (p_2-2)\chi_{p_2+4}^{-2}(\Delta^2) \right)^2 I \left(\chi_{p_2+4}^{-2}(\Delta^2) < p_2 - 2 \right) \right] \right\}
\end{aligned}
\tag{49}
$$

As $\Delta^2 \to \infty$, the above expression equals p_1 and at $\Delta^2 = 0$, it is

$$
2 + p_1 - p_2 E \left[\left(1 - (p-2)\chi_{p+2}^{-2} \right)^2 I \left(\chi_p^2 < (p-2) \right) \right] \leq (2 + p_1)
$$

Further, the risk of PRSE is monotonically increases towards the p-line while risk of RR is monotonically increasing below the graph of the risk of PRSE. Hence,

$$
R_6(\hat{\beta}_n^{RR} : I_p) \leq R_5(\hat{\beta}_n^{S+} : I_p) \leq R_4(\hat{\beta}_n^S : I_p) \leq R_1(\tilde{\beta}_n : I_p) \quad \forall \quad \Delta^2 \in R^+.
$$

3.8. Comparison of Ridge Regression and LASSO

The L_2-risk difference of sparse LASSO and ridge regression is given by

$$
\sigma^2 \left[p_1 + \Delta^2 - \frac{p_2 \Delta^2}{p_2 + \Delta^2} \right] = \sigma^2 \left(p_1 + \frac{\Delta^4}{p_2 + \Delta^2} \right) > 0.
\tag{50}
$$

Hence,

$$
R_6(\hat{\beta}_{2n}^{RR} : I_{p_2}) \leq R_7(\hat{\beta}_n^L : I_k) \quad \forall \Delta_{(k)}^2 \in R^+
$$

See Tables 3.3-3.4 for numerical efficiency of of all proposed estimators.

4. Summary and Concluding Remarks

In this paper we studied the performance of LASSO and ridge regression (RR) estimators compared to LSE, restricted, preliminary test and Stein-type estimators when the dimension of the parameters space is less than that of sample space.   Based on the analyzes of the relative efficiencies (RRE) and risks, we found that:

(i) The RR estimator uniformly dominates LSE, RE, PTE, Stein-Type estimators and LASSO;

(ii) Both RE and LASSO outperform LSE when $\Delta^2 \in [0, p_2]$ and outside this interval LSE outperforms both RE and LASSO;

(iii) LASSO outperforms the PTE when $\Delta^2 \in [0, \Delta_{PTE}^2]$, otherwise, PTE outperforms when $\Delta^2 \in (\Delta_{PTE}^2, \infty)$, where Δ_{PTE}^2 is defined by (3.7);

(iv) LASSO outperforms JSE when $\Delta^2 \in \left[0, p_2 - (p_2-2)^2 E[\chi_{p_2}^{-2}(\Delta^2)] \right)$ otherwise JSE outperforms LASSO for $\Delta^2 \in \left[p_2 - (p_2-2)^2 E[\chi_{p_2}^{-2}(\Delta^2), \infty \right)$;

(v) Similarly, LASSO outperforms PRSE when $\Delta^2 \in [0, \frac{B}{A})$ and PRSE outperforms for $\Delta^2 \in (\frac{B}{A}, \infty)$, where A and B are defined by (3.14).

  All of the above table analysis is consistent with the theoretical comparisons of the proposed estimators that are presented in section 3.

Acknowledgement

REFERENCES

1. Arashi, M. and Tabatabaey, S.M.M., *Improved variance estimation under sub-space restriction*, Journal of Multivariate Analysis, vol. 100, no. 8, pp. 1752–1760, 2009.
2. Arashi, M. and Tabatabaey, S.M.M., *A note on Stein-type estimators in elliptically contoured models*, Journal of Statistical Planning and Inference, vol. 140, no. 5, pp. 1206-1213, 2010.
3. Arashi, M. and Tabatabaey, S.M.M., *Estimation of the location parameter under LINEX loss function: multivariate case*, Metrika, vol. 72, no. 1, pp. 51-57, 2010.
4. Bancroft, T. A, *On biases in estimation due touse of preliminary tests of significance*, Annals of Mathematics and Statistics, vol. 15, no. 2, pp. 190-204, 1944.
5. Bancroft, T. A, *Analysis and inference for incompletely specified models involving the use of preliminary test(s) of significance*, Biometrics, vol. 20, no. 3, pp. 427-442, 1964.
6. Belloni, Alexandre and Chernozhukov, Victor and others, *On biases in estimation due touse of preliminary tests of significance*, Bernoulli, vol. 19, no. 2, pp.521-547, 2013.
7. Bickel, P. J, *Minimax estimation of the Mean of a normal distribution subject to doing well at a point.*, Wald Lectures. Recent Advances in Statistics, Academic Press Inc, pp. 511–528, 1983.
8. Brieman, L, *Heuristics of instability and stabilization in model selection*, The Annals of Statistics, vol. 19, no. 6, pp.2350-2383, 1996.
9. Donoho, David L and Johnstone, Jain M, *Ideal spatial adaption by wavelet shrinkage*, Biometrika, vol. 81, no. 3, pp.425–455, 1994.
10. Draper, Norman R and Van Nostrand, R Craig, *Ridge regression and James-Stein estimation: Review and Comments*, Technometrics, vol. 21, no. 4, pp.451-466, 1979.
11. Fan, Jianqing and Li, Runze, *Variable selction via nonconcave penalized likelihood and its oracle properties*, Journal of the American Statistical Association, vol. 96, no. 456, pp.1348-1360, 2005.
12. Emami, Hadi and Kiani, Sara, *Shrinkage Difference-Based Liu Estimators In Semiparametric Linear Models*, Statistics, Optimization and Information Computing, vol. 6, no. 3, pp.354–372, 2018.
13. Fan, Jianqing and Feng, Yang and Wu, Yichao, *Network exploration via the adaptive lasso and SCAD penalties*, The Annals of Applied Statistics, vol. 3, no. 2, pp.521-541, 2009.
14. Frank, I. E. and Friedman, J. H, *A statistical view of some chemometrics regression tools (with discussion)*, Technometrics, vol. 35, no. 2, pp.109-148, 1993.
15. Gruber, Marvin HJ, *Improving efficiency by shrinkage, Statistics: Textbooks and Monographs*, Springer Verlag, vol. 156, 1998.
16. Han, Chien-Pai and Bancroft, TA, *On pooling means when variance is unknown*, Journal of the American Statistical Association, vol. 63, no. 324, pp.1333–1342, 1968.
17. Hansen, Bruce E, *The risk of James-Stein and Lasso Shrinkage*, Econometric Reviews, vol. 35, no. 8-10, pp.1456–1470, 2016.
18. Hoerl, Arthur E and Kennard, Robert W, *Ridge Regression: Biased Estimation for Non-orthogonal Problems*, Technometrics, vol. 12, no. 1, pp.55–67, 1970.
19. James, William and Stein, Charles, *Estimation with quadratic loss*, roceeding of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, pp.361-379, 1992.
20. Judge, George G and Bock, Mary Ellen, *The Statistical Implications of Pre-test and Stein-rule Estimators in econometrics*, vol. 25, pp.55–67, 1970.
21. Kibria, B. M. G. and Saleh, A. K. Md. E, *Preliminary test ridge regression estimators with Student's t errors and conflicting test-statistics*, Metrika, vol. 59, no. 2, pp.105-124, 2003.
22. Kibria, B. M. G. and Saleh, A. K. Md. E, *Improving the Estimators of the Parameters of a Probit Regression Model: A Ridge Regression Approach.*, Journal of Statistical Planning and Inference, vol. 142, no. 6, pp.1421–1435, 2012.
23. Knight, K and Fu, W, *Asymptotic for a Lasso type estimators*, Annals of Statistics, vol. 28, no. 5, pp.1356-1378, 2000.
24. Roozbeh, M, *Generalized Ridge Regression Estimator in High Dimensional Sparse Regression Models*, Statistics Optimiziation and Information Computing, vol. 6, no. 5, pp.415-426, 2018.
25. Saleh, A. K. Md. E, *Theory of Preliminary Test and Stein-type Estimation with Applications*, John Wiley, vol. 6, no. 5, pp.415-426, 2018.
26. Saleh, A. K. Md. Eand Kibria, B. M. G, *On Some Ridge Regression Estimators: A Nonparametric Approach*, Journal of Nonparametric Statistics, vol. 23, no. 3, pp.819-851, 2011.
27. Saleh, K. Md. E. and M. Norouzirad, *On Shrinkage Estimation: Non-orthogonal Case*, Statistics Optimiziation and Information Computing, vol. 6, no. 3, pp.427-451, 2018.
28. Saleh, A. K. Md. E, Arashi, M. and Kibria, B. M. G. (2019), *Theory of Ridge Regression Estimation with Applications*, Wiley 2019.
29. Shalabh, *Improved Estimation in Measurement Error Models Through Stein-rule Procedure*, Journal of Multivariate Analysis, vol. 67, no. 3, pp.35-48, 1998.
30. Shalabh , *Pitman Closeness Comparison of Least Squares and Stein-rule Estimators in Linear Regression Models with Non-normal Disturbances*, The American Journal of Mathematical and Management Sciences, vol. 21, no. 1, pp.89-100, 2001.
31. Stein, C, *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, vol. I, no. 1, pp.197-206, 1956.
32. Tibshirani, R. J, *Regression Shrinkage and selection via the Lasso*, Journal of the Royal Statistical Society, vol. 58, no. 1, pp.267-288, 1996.
33. Tikhonov, A. N, *Solution of ill-posed problems*, Tranlsated in Soviet Mathematics, vol. 4, no. 1, pp.1035-1038, 1963.
34. Wasserman, L, *All of Nonparametric Statistics*, Springer,2010.
35. Zou, H, *The adaptive lasso and its oracle properties*, Journal of the American Statistical Association, vol. 101, pp.1418-1429, 2006.
36. Zou, H and Hastie, T, *Regularization and variable selection vis the elastic net*, J. of Royal Statis. Soc, pp.301-320, 2005.