Kybernetika

A. K. Md. Ehsanes Saleh; Radim Navrátil

Rank theory approach to ridge, LASSO, preliminary test and Stein-type estimators: Comparative study

Kybernetika, Vol. 54 (2018), No. 5, 958–977

Persistent URL: http://dml.cz/dmlcz/147537

KYBERNETIKA — VOLUME *54* (2018), NUMBER 5, PAGES 958–977

RANK THEORY APPROACH TO RIDGE, LASSO, PRELIMINARY TEST AND STEIN-TYPE ESTIMATORS: COMPARATIVE STUDY

A. K. Md. Ehsanes Saleh and Radim Navrátil

In the development of efficient predictive models, the key is to identify suitable predictors for a given linear model. For the first time, this paper provides a comparative study of ridge regression, LASSO, preliminary test and Stein-type estimators based on the theory of rank statistics. Under the orthonormal design matrix of a given linear model, we find that the rank based ridge estimator outperforms the usual rank estimator, restricted R-estimator, rank-based LASSO, preliminary test and Stein-type R-estimators uniformly. On the other hand, neither LASSO nor the usual R-estimator, preliminary test and Stein-type R-estimators outperform the other. The region of domination of LASSO over all the R-estimators (except the ridge R-estimator) is the interval around the origin of the parameter space. Finally, we observe that the L_2-risk of the restricted R-estimator equals the lower bound on the L_2-risk of LASSO. Our conclusions are based on L_2-risk analysis and relative L_2-risk efficiencies with related tables and graphs.

Keywords: efficiency of LASSO, penalty estimators, preliminary test, Stein-type estimator, ridge estimator, L_2-risk function

Classification: 62G05, 62J05, 62J07

1. INTRODUCTION

The history of estimation theory changed its course radically since James and Stein ([13] and [8]) proved that sample mean based on a sample from a p-dimensional multivariate normal distribution is inadmissible under a quadratic loss for $p \geq 3$. This result gave birth to a class of shrinkage estimators in various forms and setups. Due to the immense impact of Stein's theory, scores of technical papers appeared in the literature covering all areas of applications. Saleh and Sen (1978 – 1986) reformulated and expanded Stein's theory based on rank theory, Mtheory and quantile theory beginning in the 70's, see [11].

The next generation of "shrinkage estimators" known as penalty estimators began in the 70's with the pioneering work on "ridge regression" estimator for linear models by Hoerl and Kennard ([7]) based on the idea of "Tikhonov regularization" ([15]). "Ridge regression" estimator is the result of minimizing least squares criteria subject to some

DOI: 10.14736/kyb-2018-5-0958

quadratic restriction (L_2-function)

$$\widehat{\boldsymbol{\beta}}_n^{RR}(k) = \arg\min\{(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + k\boldsymbol{\beta}'\boldsymbol{\beta} :\ \boldsymbol{\beta} \in \mathbb{R}^p\}, \quad k > 0. \tag{1}$$

A generalized ridge regression estimator may be defined as

$$\widehat{\boldsymbol{\beta}}_n^{RS}(\mathbf{K}) = \arg\min\{(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\beta}'\mathbf{K}\boldsymbol{\beta} :\ \boldsymbol{\beta} \in \mathbb{R}^p\}, \tag{2}$$

where $\mathbf{K} = \mathrm{diag}(k_1, \ldots, k_p)$. Note that the penalty function (1) puts equal weights on the $\boldsymbol{\beta}$'s while (2) puts unequal weights.

Frank and Friedman in [6] defined a class of "bridge estimators" defined by

$$\widehat{\boldsymbol{\beta}}_n^{BE}(\lambda_n) = \arg\min\{(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + \lambda_n \mathbf{1}_p'|\boldsymbol{\beta}|^\gamma :\ \boldsymbol{\beta} \in \mathbb{R}^p\}, \tag{3}$$

where $\mathbf{1}_p'$ is a p-vector of 1's and $|\boldsymbol{\beta}|^\gamma = (|\beta_1|^\gamma, \ldots, |\beta_p|^\gamma)'$ with $\gamma > 0$.

The choice of $\gamma = 2$ gives ridge estimates, while $\gamma = 1$ relates to LASSO introduced by Tibshirani in [14]. It has become a very popular and intriguing penalty estimator. This estimator is related to the estimators such as "non-negative" garotte by Breiman ([2]), smoothly clipped absolute deviation (SCAD) by Fan and Li ([5]), elastic net by Zou and Hastie ([17]), adaptive LASSO (aLASSO) by Zou ([16]), hard threshold LASSO by Belloni and Chernozhukov ([1]) and many other versions.

This paper introduces the R-estimates and the application of marginal distribution theory to study the performance characteristics of two primary penalty estimators, namely the "ridge regression" and "LASSO" (least absolute shrinkage and selection operator) along with the preliminary test (PTE) and Stein-type estimators, for more details see Draper and Van Nostrand ([4]) and Hansen ([10]). An important characteristic of LASSO is that it provides simultaneous estimation and selection of coefficients in linear models and can be applied when the dimension of the parameter space exceeds the dimension of the sample space.

2. LINEAR MODEL AND R-ESTIMATORS

Consider the multiple regression model

$$\mathbf{Y} = \theta\mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + \mathbf{e} = \theta\mathbf{1}_n + \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}, \tag{4}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_n)'$, $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ is the $n \times p$ matrix of real numbers, θ is the intercept parameter and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ is the p-vector ($p = p_1 + p_2$) of regression parameters, where p_2-dimensional vector $\boldsymbol{\beta}_2$ may be $\mathbf{0}$. We assume that:

(1) Errors $\mathbf{e} = (e_1, \ldots, e_n)'$ are independent and identically distributed (i.i.d.) random variables with (unknown) c.d.f. F having absolutely continuous p.d.f. f with finite and nonzero Fisher information with respect to the location

$$0 < I(f) = \int_{-\infty}^{\infty} \left[-\frac{f'(x)}{f(x)}\right]^2 f(x)\,\mathrm{d}x < \infty. \tag{5}$$

(2) For the definition of linear rank statistics, we consider the score generating function $\varphi: (0,1) \mapsto \mathbb{R}$ which is assumed to be nonconstant, non-decreasing and square integrable on $(0,1)$ so that

$$A_\varphi^2 = \int_0^1 \varphi^2(u)du - \left(\int_0^1 \varphi(u)\,\mathrm{d}u\right)^2. \tag{6}$$

The scores are defined in either of the following ways:

$$a_n(i) = \mathbb{E}\varphi(U_{i:n}), \quad \text{or} \quad a_n(i) = \varphi\left(\frac{i}{n+1}\right), \quad i = 1, \ldots, n,$$

where $U_{1:n} \leq \ldots \leq U_{n:n}$ are order statistics from a sample of size n from uniform distribution $\mathcal{U}(0,1)$.

(3) Let us define

$$\mathbf{C}_n = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}}_n)(\mathbf{x}_i - \bar{\mathbf{x}}_n)', \tag{7}$$

where $\mathbf{x}_i$ is the ith row of $\mathbf{X}$ and $\bar{\mathbf{x}}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$. We assume that

$$\text{(a)} \lim_{n\to\infty} \mathbf{C}_n = \mathbf{I}_p \quad \text{and (b)} \lim_{n\to\infty} \max_{1\leq i\leq n} (\mathbf{x}_i - \bar{\mathbf{x}}_n)'\mathbf{C}_n^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}_n) = 0.$$

For the R-estimation of $\boldsymbol{\beta}$, define for $\mathbf{b} \in \mathbb{R}^p$ the rank of $Y_i - \mathbf{x}_i'\mathbf{b}$ among $Y_1 - \mathbf{x}_1'\mathbf{b}, \ldots, Y_n - \mathbf{x}_n'\mathbf{b}$ to be $R_{ni}(\mathbf{b})$. Then for each n, consider the set of scores $a_n(1) \leq \ldots \leq a_n(n)$ and define the vector of linear rank statistics

$$\mathbf{L}_n(\mathbf{b}) = (L_{n1}(\mathbf{b}), \ldots, L_{nn}(\mathbf{b}))' = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}}_n)a_n(R_{ni}(\mathbf{b})). \tag{8}$$

Since $R_{ni}(\mathbf{b})$ are translation invariant, there is no need to adjust the intercept parameter θ.

If we set $\|\mathbf{a}\| = \sum_{i=1}^{p}|a_i|$ for $\mathbf{a} = (a_1, \ldots, a_p)'$, then the unrestricted rank estimator (URE) is defined as any central point of the set

$$\mathfrak{D}_n = \{\arg\min \|\mathbf{L}_n(\mathbf{b})\| : \mathbf{b} \in \mathbb{R}^p\}. \tag{9}$$

Let us denote the URE as $\widetilde{\boldsymbol{\beta}}_n$. Then, using the uniform asymptotic linearity of Jurečková ([9])

$$\lim_{n\to\infty} P\left(\sup_{\|\boldsymbol{\omega}\|<k}\left\|\mathbf{L}_n\left(\boldsymbol{\beta} + \frac{\boldsymbol{\omega}}{\sqrt{n}}\right) - \mathbf{L}_n(\boldsymbol{\beta}) + \gamma\boldsymbol{\omega}\right\| > \epsilon\right) = 0 \tag{10}$$

for any $k > 0$ and $\epsilon > 0$. Then, it is well-known that

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \eta^2\mathbf{I}_p), \tag{11}$$

where

$$\eta^2 = \frac{A_\varphi^2}{\gamma^2}, \quad \gamma = \int_0^1 \varphi(u)\left\{-\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}\right\}\mathrm{d}u. \tag{12}$$

2.1. Penalty R-estimators

First, we note that the marginal distribution of $\sqrt{n}(\widetilde{\beta}_{jn} - \beta_j)$ is $\mathcal{N}(0, \eta^2)$ for all $j = 1, \ldots, p$. Then, we are able to define the rank based ridge regression estimator under the sparsity condition $\boldsymbol{\beta}_2 = \mathbf{0}$ as

$$(i) \quad \widehat{\boldsymbol{\beta}}_n^{RR}(k) = \begin{pmatrix} \widetilde{\boldsymbol{\beta}}_{1n} \\ \frac{1}{1+k}\widetilde{\boldsymbol{\beta}}_{2n} \end{pmatrix}, \quad k > 0, \tag{13}$$

where k is the tuning parameter.

Next, we consider the LASSO estimator (L_1-penalty)

$$(ii) \quad \widehat{\boldsymbol{\beta}}_n^{L}(\lambda) = \left(\operatorname{sign}(\widetilde{\beta}_{jn}) \left((|\widetilde{\beta}_{jn}| - \frac{\lambda}{\sqrt{n}}\eta \right)^{+} ; \ j = 1, \ldots, p \right)',$$

where λ is the tuning parameter.

Our problem is to compare the performance characteristics of "ridge" and LASSO estimators with that of Stein-type and preliminary test R-estimators with respect to asymptotic mean squared error criterion. We present the preliminary test and Stein-type R-estimators in the next section.

2.2. PTE and Stein-type R-estimators

For the model (4), if we suspect a sparsity condition that $\boldsymbol{\beta}_2 = \mathbf{0}$, then the restricted R-estimator (RE) of $(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ is $\widehat{\boldsymbol{\beta}}_n = (\widetilde{\boldsymbol{\beta}}_{1n}', \mathbf{0}')'$. For the test of the null hypothesis $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ vs. $K : \boldsymbol{\beta}_2 \neq \mathbf{0}$, the rank test statistic is given by

$$\mathfrak{L}_n = nA_n^2 \mathbf{L}_{2n}'(\mathbf{0}) \mathbf{L}_{2n}(\mathbf{0}), \tag{14}$$

where $\mathbf{L}_n(\mathbf{0}) = (\mathbf{L}_{1n}'(\mathbf{0}), \mathbf{L}_{2n}'(\mathbf{0}))'$ from (8) and

$$A_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (a_n(i) - \bar{a}_n)^2, \quad \bar{a}_n = \frac{1}{n} \sum_{i=1}^{n} a_n(i).$$

It is well-known that under model (4) and the assumptions (5)–(8) as $n \to \infty$, $\mathfrak{L}_n$ follows the χ^2 distribution with p_2 degrees of freedom (d.f.) under H_0. Then, we define the preliminary test estimator (PTE) of $(\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ as

$$\begin{aligned} \begin{pmatrix} \widehat{\boldsymbol{\beta}}_{1n}^{PT} \\ \widehat{\boldsymbol{\beta}}_{2n}^{PT} \end{pmatrix} &= \begin{pmatrix} \widetilde{\boldsymbol{\beta}}_{1n} \\ \widetilde{\boldsymbol{\beta}}_{2n} \end{pmatrix} - \left[\begin{pmatrix} \widetilde{\boldsymbol{\beta}}_{1n} \\ \widetilde{\boldsymbol{\beta}}_{2n} \end{pmatrix} - \begin{pmatrix} \widetilde{\boldsymbol{\beta}}_{1n} \\ \mathbf{0} \end{pmatrix} \right] I(\mathfrak{L}_n < \chi_{p_2}^2(\alpha)) \\ &= \begin{pmatrix} \widetilde{\boldsymbol{\beta}}_{1n} \\ \widetilde{\boldsymbol{\beta}}_{2n} - \widetilde{\boldsymbol{\beta}}_{2n} I(\mathfrak{L}_n < \chi_{p_2}^2(\alpha)) \end{pmatrix}, \end{aligned} \tag{15}$$

where $I(A)$ is the indicator function of the set A and $\chi_{p_2}^2(\alpha)$ is the $(1-\alpha)$-quantile of χ^2 distribution with p_2 degrees of freedom.

Similarly, we define the James-Stein-type R-estimator as

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}_{1n}^{JS} \\ \widehat{\boldsymbol{\beta}}_{2n}^{JS} \end{pmatrix} = \begin{pmatrix} \widetilde{\boldsymbol{\beta}}_{1n} \\ \widetilde{\boldsymbol{\beta}}_{2n}(1-(p_2-2)\mathfrak{L}_n^{-1}) \end{pmatrix}. \tag{16}$$

Finally, the positive rule Stein-type estimator is given by

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}_{1n}^{S+} \\ \widehat{\boldsymbol{\beta}}_{2n}^{S+} \end{pmatrix} = \begin{pmatrix} \widetilde{\boldsymbol{\beta}}_{1n} \\ \widetilde{\boldsymbol{\beta}}_{2n}(1-(p_2-2)\mathfrak{L}_n^{-1})I(\mathfrak{L}_n > p_2-2) \end{pmatrix}. \tag{17}$$

2.3. Asymptotic bias and L$_2$-risks of the estimators

Since $\boldsymbol{\beta}_2 = \mathbf{0}$ is uncertain, we test the null hypothesis $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ vs. $K : \boldsymbol{\beta}_2 \neq \mathbf{0}$ based on the rank statistics $\mathfrak{L}_n$ of (14). This test is consistent and its power tends to unity as $n \to \infty$ for fixed alternatives. Thus, we consider a sequence of Pitman's alternatives $K_{(n)}$ defined by

$$K_{(n)} : \boldsymbol{\beta}_n = n^{-1/2}\boldsymbol{\delta} = n^{-1/2}(\boldsymbol{\delta}_1', \boldsymbol{\delta}_2')'. \tag{18}$$

If $\boldsymbol{\delta}_2 = \mathbf{0}$, $K_{(n)} = H_0$. Then under $\{K_{(n)}\}$, the marginal asymptotic distribution of $\sqrt{n}(\widetilde{\boldsymbol{\beta}}_{jn} - \boldsymbol{\beta}_j)$ is $\mathcal{N}_{p_j}(\mathbf{0}, \eta^2\mathbf{I}_{p_j})$ for $j = 1, 2$.

Hence, the asymptotic bias (AB) and asymptotic L$_2$-risk (AL$_2$-risk) of the above estimators are given below.

Remark 2.1. The asymptotic bias is the difference of the asymptotic mean of the estimate and the real parameter value $\boldsymbol{\beta}$. Asymptotic L$_2$-risk is the sum of asymptotic mean squared errors of the components $\widehat{\beta}_{jn}$. Some of the detailed computations for the formulas below might me found in [12].

(i) Unrestricted rank estimator (URE):

$$\begin{pmatrix} \mathrm{AB}(\widetilde{\boldsymbol{\beta}}_{1n}) \\ \mathrm{AB}(\widetilde{\boldsymbol{\beta}}_{2n}) \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

$$\begin{pmatrix} \mathrm{AL}_2\text{-risk}(\widetilde{\boldsymbol{\beta}}_{1n}) \\ \mathrm{AL}_2\text{-risk}(\widetilde{\boldsymbol{\beta}}_{2n}) \end{pmatrix} = \eta^2 \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}.$$

Therefore, AL$_2$-risk$(\widetilde{\boldsymbol{\beta}}_n) = \eta^2(p_1 + p_2) = \eta^2 p$.

(ii) Restricted rank estimator (RE):

$$\begin{pmatrix} \mathrm{AB}(\widetilde{\boldsymbol{\beta}}_{1n}) \\ \mathrm{AB}(\mathbf{0}) \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ -\boldsymbol{\delta}_2 \end{pmatrix},$$

$$\begin{pmatrix} \mathrm{AL}_2\text{-risk}(\widetilde{\boldsymbol{\beta}}_{1n}) \\ \mathrm{AL}_2\text{-risk}(\mathbf{0}) \end{pmatrix} = \begin{pmatrix} \eta^2 p_1 \\ \boldsymbol{\delta}_2\boldsymbol{\delta}_2' \end{pmatrix}.$$

As a result, AL$_2$-risk$(\widehat{\boldsymbol{\beta}}_n) = \eta^2(p_1 + \Delta^2)$, where $\Delta^2 = \frac{\boldsymbol{\delta}_2'\boldsymbol{\delta}_2}{\eta^2}$.

(iii) Preliminary test estimator (PTE):

$$\begin{pmatrix} \mathrm{AB}\left(\widetilde{\boldsymbol{\beta}}_{1n}\right) \\ \mathrm{AB}\left(\widehat{\boldsymbol{\beta}}_{2n}^{PT}\right) \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ -\boldsymbol{\delta}_2 \mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha);\Delta^2) \end{pmatrix},$$

$$\begin{pmatrix} \mathrm{AL}_2\text{-risk}\left(\widetilde{\boldsymbol{\beta}}_{1n}\right) \\ \mathrm{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_{2n}^{PT}\right) \end{pmatrix}$$
$$= \begin{pmatrix} \eta^2 p_1 \\ \eta^2[p_2(1-\mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha);\Delta^2)) + \Delta^2(2\mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha);\Delta^2) - \mathcal{H}_{p_2+4}(\chi^2_{p_2}(\alpha);\Delta^2))] \end{pmatrix}.$$

Hence, the AL_2-risk of $\widehat{\boldsymbol{\beta}}_n^{PT}$ is given by

$$\eta^2[p_1 + p_2(1-\mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha);\Delta^2)) + \Delta^2(2\mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha);\Delta^2) - \mathcal{H}_{p_2+4}(\chi^2_{p_2}(\alpha);\Delta^2))],$$

where $\mathcal{H}_\nu(c;\Delta^2)$ is the c.d.f. of χ^2- distribution with ν d.f. and noncentrality parameter Δ^2 evaluated at c.

(iv) James-Stein estimator (JSE):

$$\begin{pmatrix} \mathrm{AB}\left(\widetilde{\boldsymbol{\beta}}_{1n}\right) \\ \mathrm{AB}\left(\widehat{\boldsymbol{\beta}}_{2n}^{JS}\right) \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ -\boldsymbol{\delta}_2 (p_2-2)\mathbb{E}[\chi^{-2}_{p_2+2}(\Delta^2)] \end{pmatrix},$$

$$\begin{pmatrix} \mathrm{AL}_2\text{-risk}\left(\widetilde{\boldsymbol{\beta}}_{1n}\right) \\ \mathrm{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_{2n}^{JS}\right) \end{pmatrix} = \begin{pmatrix} \eta^2 p_1 \\ \eta^2[p_2 - (p_2-2)^2\mathbb{E}[\chi^{-2}_{p_2+2}(\Delta^2)]] \end{pmatrix}.$$

Hence, the AL_2-risk of $\widehat{\boldsymbol{\beta}}_n^{JS}$ is given by the simplified form

$$\eta^2\left[p_1 + p_2 - (p_2-2)^2\mathbb{E}[\chi^{-2}_{p_2}(\Delta^2)]\right],$$

where

$$\mathbb{E}[\chi^{-2\nu}_{p_2}(\Delta^2)] = \int_0^\infty x^{-2\nu}\,\mathrm{d}\mathcal{H}_{p_2}(x;\Delta^2).$$

(v) Positive rule Stein-type estimator (PRSE):

$$\begin{pmatrix} \mathrm{AB}\left(\widetilde{\boldsymbol{\beta}}_{1n}\right) \\ \mathrm{AB}\left(\widehat{\boldsymbol{\beta}}_{2n}^{S+}\right) \end{pmatrix}$$
$$= \begin{pmatrix} \mathbf{0} \\ \mathrm{AB}\left(\widehat{\boldsymbol{\beta}}_{2n}^{JS}\right) - \boldsymbol{\delta}_2\mathbb{E}[(1-(p_2-2)\chi^{-2}_{p_2+2}(\Delta^2))I(\chi^2_{p_2+2}(\Delta^2) < p_2-2)] \end{pmatrix},$$

$$\begin{pmatrix} \mathrm{AL}_2\text{-risk}\left(\widetilde{\boldsymbol{\beta}}_{1n}\right) \\ \mathrm{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_{2n}^{S+}\right) \end{pmatrix} = \begin{pmatrix} \eta^2 p_1 \\ \mathrm{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_{2n}^{S+}\right) \end{pmatrix},$$

where

$$\begin{aligned}
&\text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_{2n}^{S+}\right) = \text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_{2n}^{JS}\right)\\
&-\eta^2 p_2 \mathbb{E}[(1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta^2))^2 I(\chi_{p_2+2}^2(\Delta^2) < p_2-2)]\\
&+\Delta^2\{2\mathbb{E}[(1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta^2))I(\chi_{p_2+2}^2(\Delta^2) < p_2-2)]\\
&+\mathbb{E}[(1-(p_2-2)\chi_{p_2+4}^{-2}(\Delta^2))I(\chi_{p_2+4}^2(\Delta^2) < p_2-2)]\}.
\end{aligned}$$

Hence, the AL$_2$-risk of $\widehat{\boldsymbol{\beta}}_n^{S+}$ is given by

$$\begin{aligned}
&\eta^2[p-(p_2-2)^2\mathbb{E}[\chi_{p_2}^{-2}(\Delta^2)]]\\
&-\eta^2[p_2\mathbb{E}[(1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta^2))^2 I(\chi_{p_2+2}^2(\Delta^2) < p_2-2)]\\
&-\Delta^2\{2\mathbb{E}[(1-(p_2-2)\chi_{p_2+2}^{-2}(\Delta^2))I(\chi_{p_2+2}^2(\Delta^2) < p_2-2)]\\
&-\mathbb{E}[(1-(p_2-2)\chi_{p_2+4}^{-2}(\Delta^2))I(\chi_{p_2+4}^2(\Delta^2) < p_2-2)]\}]\\
&= \eta^2[(p_1+p_2)-(p_2-2)^2\mathbb{E}[\chi_{p_2}^{-2}(\Delta^2)] - R^*],
\end{aligned}$$

where

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

(vi) Ridge regression (RR):

$$\begin{pmatrix} \text{AB}\left(\widetilde{\boldsymbol{\beta}}_{1n}\right) \\ \text{AB}\left(\widehat{\boldsymbol{\beta}}_{2n}^{RR}(k)\right) \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ -\frac{k}{k+1}\boldsymbol{\delta}_2, \end{pmatrix},$$

$$\begin{pmatrix} \text{AL}_2\text{-risk}\left(\widetilde{\boldsymbol{\beta}}_{1n}\right) \\ \text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_{2n}^{RR}(k)\right) \end{pmatrix} = \begin{pmatrix} \eta^2 p_1 \\ \eta^2\frac{p_2+k^2\Delta^2}{(1+k)^2} \end{pmatrix}.$$

Hence

$$\text{ADL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{RR}(k)\right) = \eta^2 p_1 + \frac{\eta^2}{(k+1)^2}(p_2+k^2\Delta^2).$$

Therefore, the optimum AL$_2$-risk $\left(\widehat{\boldsymbol{\beta}}_n^{RR}(k_{opt})\right) = \eta^2\left(p_1+\frac{p_2\Delta^2}{p_2+\Delta^2}\right)$, since $k_{opt} = p_2\Delta^{-2}$.

Consequently, the asymptotic relative L$_2$-risk efficiencies (ARRE) of the estimators are given by

$$(i)\ \text{ARRE (RE:URE)} = \left(1+\frac{p_2}{p_1}\right)\left(1+\frac{\Delta^2}{p_1}\right)^{-1},$$

$$(ii)\ \text{ARRE (LASSO:URE)} = \left(1+\frac{p_2}{p_1}\right)\left(1+\frac{\Delta^2}{p_1}\right)^{-1},$$

$$(iii)\ \text{ARRE (RR:URE)} = \left(1+\frac{p_2}{p_1}\right)\left(1+\frac{p_2\Delta^2}{p_1(p_2+\Delta^2)}\right)^{-1},$$

$$(iv)\ \text{ARRE (PTE:URE)} = \left(1+\frac{p_2}{p_1}\right)\bigg\{1+\frac{p_2}{p_1}\left(1-H_{p_2+2}(\chi^2_{p_2}(\alpha);\Delta^2)\right) + \frac{\Delta^2}{p_1}\left(2H_{p_2+2}(\chi^2_{p_2}(\alpha);\Delta^2)-H_{p_2+4}(\chi^2_{p_2}(\alpha);\Delta^2)\right)\bigg\}^{-1},$$

$$(v)\ \text{ARRE (JSE:URE)} = \left(1+\frac{p_2}{p_1}\right)\left\{1+\frac{p_2}{p_1}-\frac{1}{p_1}(p_2-2)\mathbb{E}\left[\chi^{-2}_{p_2}(\Delta^2)\right]\right\}^{-1},$$

$$(vi)\ \text{ARRE (PRSE:URE)} = \left(1+\frac{p_2}{p_1}\right)\bigg\{1+\frac{p_2}{p_1}-\frac{1}{p_1}(p_2-2)\mathbb{E}\left[\chi^{-2}_{p_2}(\Delta^2)\right] - \frac{p_2}{p_1}\mathbb{E}\left[\left(1-(p_2-2)\chi^{-2}_{p_2+2}(\Delta^2)\right)^2 I\left(\chi^2_{p_2+2}(\Delta^2)<(p_2-2)\right)\right] + \frac{\Delta^2}{p_1}\Big[2\mathbb{E}\left[\left(1-(p_2-2)\chi^{-2}_{p_2+2}(\Delta^2)\right)I\left(\chi^2_{p_2+2}(\Delta^2)<(p_2-2)\right)\right] - \mathbb{E}\left[\left(1-(p_2-2)\chi^{-2}_{p_2+4}(\Delta^2)\right)^2 I\left(\chi^2_{p_2+4}(\Delta^2)<(p_2-2)\right)\right]\Big]\bigg\}^{-1}.$$

2.4. Analysis of L_2-risk of the estimators under the sparsity condition

First, note that the sparsity condition $\boldsymbol{\beta}_2 = \mathbf{0}$ is equivalent to $\Delta^2 = 0$. Further, it is evident that the ARRE as a function of Δ^2 is decreasing and tends towards unity as $\Delta^2 \to \infty$. Clearly, under $\Delta^2 = 0$, (i), (ii) and (iii) from the previous section are equal to $\left(1+\frac{p_2}{p_1}\right)$ indicating that URE, RE and RR are all L_2-risk equivariant when sparsity conditions hold. As for PTE, JSE and PRSE, we have the following expressions:

(iv) $\left(1+\frac{p_2}{p_1}\right)\left\{1+\frac{p_2}{p_1}(1-\mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha);0))\right\}^{-1}$,

(v) $\left(1+\frac{p_2}{p_1}\right)\left\{1+\frac{2}{p_1}\right\}^{-1}$,

(vi) $\left(1+\frac{p_2}{p_1}\right)\left\{1+\frac{2}{p_1}-\frac{p_2}{p_1}\mathbb{E}\left[\left(1-(p_2-2)\chi^{-2}_{p_2+2}(0)\right)^2 I\left(\chi^2_{p_2+2}(0)<(p_2-2)\right)\right]\right\}^{-1}$.

2.5. Analysis of L_2-risk of the estimators for general Δ^2

In this section we consider the asymptotic relative risk efficiency ($ARRE$) of the above six estimators and LASSO for general Δ^2.

2.5.1. Comparison of URE and RE

In this case

$$\text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n\right) - \text{AL}_2\text{-risk}\left(\widetilde{\boldsymbol{\beta}}_n\right) = \eta^2(\Delta^2 - p_2).$$

Hence, URE outperforms RE for $\Delta^2 > p_2$ and RE outperforms URE for $\Delta^2 < p_2$. Neither RE nor URE dominates the other uniformly.

2.5.2. Comparison of URE and PTE

$$\begin{aligned}
&\text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{PT}\right) - \text{AL}_2\text{-risk}\left(\widetilde{\boldsymbol{\beta}}_n\right) \\
&= \eta^2[p_1 + p_2(1 - \mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha); \Delta^2)) \\
&+\Delta^2\{2\mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha); \Delta^2) - \mathcal{H}_{p_2+4}(\chi^2_{p_2}(\alpha); \Delta^2)\} - (p_1 + p_2)] \\
&= \eta^2\left[-p_2\mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha); \Delta^2) + \Delta^2\{2\mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha); \Delta^2) - \mathcal{H}_{p_2+4}(\chi^2_{p_2}(\alpha); \Delta^2)\}\right].
\end{aligned}$$

Hence, if

$$0 < \Delta^2 \leq \frac{p_2\mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha); \Delta^2)}{2\mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha); \Delta^2) - \mathcal{H}_{p_2+4}(\chi^2_{p_2}(\alpha); \Delta^2)},$$

then PTE dominates URE. If

$$\Delta^2 > \frac{p_2\mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha); \Delta^2)}{2\mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha); \Delta^2) - \mathcal{H}_{p_2+4}(\chi^2_{p_2}(\alpha); \Delta^2)},$$

then URE dominates PTE. Thus, neither PTE nor URE dominates the other uniformly.

2.5.3. Comparison of URE, JSE and PRSE

$$\text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{JS}\right) - \text{AL}_2\text{-risk}\left(\widetilde{\boldsymbol{\beta}}_n\right) = \eta^2\left[(p_2-2)^2\mathbb{E}[\chi^{-2}_{p_2}(\Delta^2)\right] \geq 0$$

for all $\Delta^2 \geq 0$. Hence, JSE dominates URE uniformly. Further,

$$\begin{aligned}
&\text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{JS}\right) - \text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{S+}\right) \\
&= \eta^2 p_2\mathbb{E}[(1 - (p_2-2)\chi^{-2}_{p_2+2}(\Delta^2))^2 I(\chi^2_{p_2+2}(\Delta^2) < p_2 - 2)] \\
&+\Delta^2\{2\mathbb{E}[(1 - (p_2-2)\chi^{-2}_{p_2+2}(\Delta^2))^2 I(\chi^2_{p_2+2}(\Delta^2) < p_2 - 2)] \\
&+\mathbb{E}[((p_2-2)\chi^{-2}_{p_2+4}(\Delta^2) - 1) I(\chi^2_{p_2+4}(\Delta^2) < p_2 - 2)]\}
\end{aligned}$$

for all $\Delta^2 \geq 0$, since

$$0 < \chi^2_{p_2+2}(\Delta^2) < p_2 - 2 \Rightarrow \left((p_2-2)\chi^2_{p_2+2}(\Delta^2) - 1\right) \geq 0.$$

Hence, combining with the previous result, we find that

$$\text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{S+}\right) \leq \text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{JS}\right) \leq \text{AL}_2\text{-risk}\left(\widetilde{\boldsymbol{\beta}}_n\right).$$

2.5.4. Comparison of URE and RR

$$\text{AL}_2\text{-risk}\left(\widetilde{\boldsymbol{\beta}}_n\right) - \text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{RR}(k_{opt})\right) = \eta^2 \frac{p_2^2}{\Delta^2 + p_2} > 0.$$

Hence, RR dominates uniformly over URE.

2.5.5. Comparison of RR and PTE

$$\begin{aligned}
&\text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{PT}\right) - \text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{RR}(k_{opt})\right) \\
&= \eta^2 \Bigg[p_2(1 - \mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha); \Delta^2)) + \\
&\qquad \Delta^2(2\mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha); \Delta^2) - \mathcal{H}_{p_2+4}(\chi^2_{p_2}(\alpha); \Delta^2)) - \frac{p_2\Delta^2}{p_2 + \Delta^2} \Bigg],
\end{aligned}$$

Note that the risk of PTE is an increasing function of Δ^2, crossing the p_2-line to a maximum and then it drops monotonically towards p_2-line as $\Delta^2 \to \infty$. At $\Delta^2 = 0$, the value of this risk is $p_2[1 - \mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha); 0)] < p_2$. On the other hand, $\frac{p_2\Delta^2}{p_2+\Delta^2}$ is an increasing function of Δ^2 below the p_2-line with a minimum value 0 at $\Delta^2 = 0$ and it converges to p_2 as $\Delta^2 \to \infty$. Hence, the risk difference is non-negative for all $\Delta^2 \geq 0$ and RR outperforms PTE uniformly.

2.5.6. Comparison of RR and JSE

$$\begin{aligned}
&\text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{JS}\right) - \text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{RR}(k_{opt})\right) \\
&= \eta^2 \left\{ p_2 - (p_2 - 2)^2 \mathbb{E}[\chi^{-2}_{p_2}(\Delta^2)] - \frac{p_2\Delta^2}{p_2 + \Delta^2} \right\}.
\end{aligned}$$

Note that the first function is increasing in Δ^2 with its minimum at $\Delta^2 = 0$ and as $\Delta^2 \to \infty$ it tends to p_2. The second function is also increasing in Δ^2 with a value 0 at Δ^2 and approaches the value p_2 as $\Delta^2 \to \infty$. Hence, the risk difference is non-negative for all $\Delta^2 \geq 0$. Hence, RR uniformly outperforms JSE.

2.5.7. Comparison of RR and PRSE

$$\begin{aligned}
&\text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{S+}\right) - \text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{RR}(k_{opt})\right) \\
&= \eta^2 \left\{ p_2 - (p_2 - 2)^2 \mathbb{E}[\chi^{-2}_{p_2}(\Delta^2)] - \frac{p_2\Delta^2}{p_2 + \Delta^2} - R^* \right\} \geq 0, \quad \forall\, \Delta^2 \geq 0,
\end{aligned}$$

where

$$\begin{aligned}
R^* = {} & p_2 \mathbb{E}[(1 - (p_2 - 2)\chi^{-2}_{p_2+2}(\Delta^2))^2 I(\chi^2_{p_2+2}(\Delta^2) < p_2 - 2)] \\
& - \Delta^2 \{ 2\mathbb{E}[(1 - (p_2 - 2)\chi^{-2}_{p_2+2}(\Delta^2)) I(\chi^2_{p_2+2}(\Delta^2) < p_2 - 2)] \\
& - \mathbb{E}[(1 - (p_2 - 2)\chi^{-2}_{p_2+4}(\Delta^2)) I(\chi^2_{p_2+4}(\Delta^2) < p_2 - 2)] \}.
\end{aligned}$$

Consider the risk of PRSE. It is an increasing function of Δ^2. At $\Delta^2 = 0$, its value is

$$(p_1 + 2) - p_2\mathbb{E}[(1 - (p_2 - 2)\chi^{-2}_{p_2+2}(0))I(\chi^2_{p_2+2}(0) < p_2 - 2)] \geq 0$$

and as $\Delta^2 \to \infty$, it tends to $p_1 + p_2$. For RR, at $\Delta^2 = 0$, the value is p_1 and as $\Delta^2 \to \infty$, it tends to $p_1 + p_2$. Hence, the risk difference is non-negative and therefore RR outperforms PRSE uniformly.

2.5.8. Comparison of LASSO and PTE

$$\begin{aligned}
&\text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{PT}\right) - \text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{L}\right)\\
&= \eta^2\Big[p_2(1 - \mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha); \Delta^2)\\
&\qquad -\Delta^2\{1 - 2\mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha); \Delta^2) + \mathcal{H}_{p_2+4}(\chi^2_{p_2}(\alpha); \Delta^2)\})\Big].
\end{aligned}$$

Then, for

$$0 \leq \Delta^2 \leq \frac{p_2[1 - \mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha); \Delta^2)]}{1 - 2\mathcal{H}_{p_2+2}(\chi^2_{p_2}(\alpha); \Delta^2) + \mathcal{H}_{p_2+4}(\chi^2_{p_2}(\alpha); \Delta^2)}$$

LASSO outperforms PTE. On the other hand, outside this interval PTE outperforms LASSO. Hence, neither LASSO nor PTE dominates the other uniformly.

2.5.9. Comparison of LASSO and JSE

$$\begin{aligned}
&\text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{JS}\right) - \text{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{L}\right)\\
&= \eta^2\left[p_2 - (p_2 - 2)^2\mathbb{E}[\chi^{-2}_{p_2}(\Delta^2)] - \Delta^2\right].
\end{aligned}$$

Hence, for $0 \leq \Delta^2 \leq p_2 - (p_2 - 2)^2\mathbb{E}[\chi^{-2}_{p_2}(\Delta^2)]$, LASSO outperforms JSE, otherwise JSE outperforms LASSO. Hence, neither LASSO nor JSE outperforms the other uniformly.

2.5.10. Comparison of LASSO and PRSE

$$\begin{aligned}
&\text{ADL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{S+}\right) - \text{ADL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{L}\right)\\
&= \eta^2\left[\text{ADL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{JS}\right) - R^* - (p_1 + \Delta^2)\right]\\
&= \eta^2\left[p_2 - (p_2 - 2)^2\mathbb{E}[\chi^{-2}_{p_2}(\Delta^2)] - R^* - \Delta^2\right] = \eta^2(A - \Delta^2 B),
\end{aligned}$$

where

$$\begin{aligned}
A &= p_2 - (p_2 - 2)^2\mathbb{E}[\chi^{-2}_{p_2}(\Delta^2)]\\
&\quad - p_2\mathbb{E}[(1 - (p_2 - 2)\chi^{-2}_{p_2+2}(\Delta^2))^2 I(\chi^2_{p_2+2}(\Delta^2) < p_2 - 2)]
\end{aligned}$$

and

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

Again, for $0 \le \Delta^2 \le \frac{A}{B}$, LASSO outperforms PRSE, otherwise PRSE outperforms LASSO. Neither LASSO nor PRSE dominates the other.

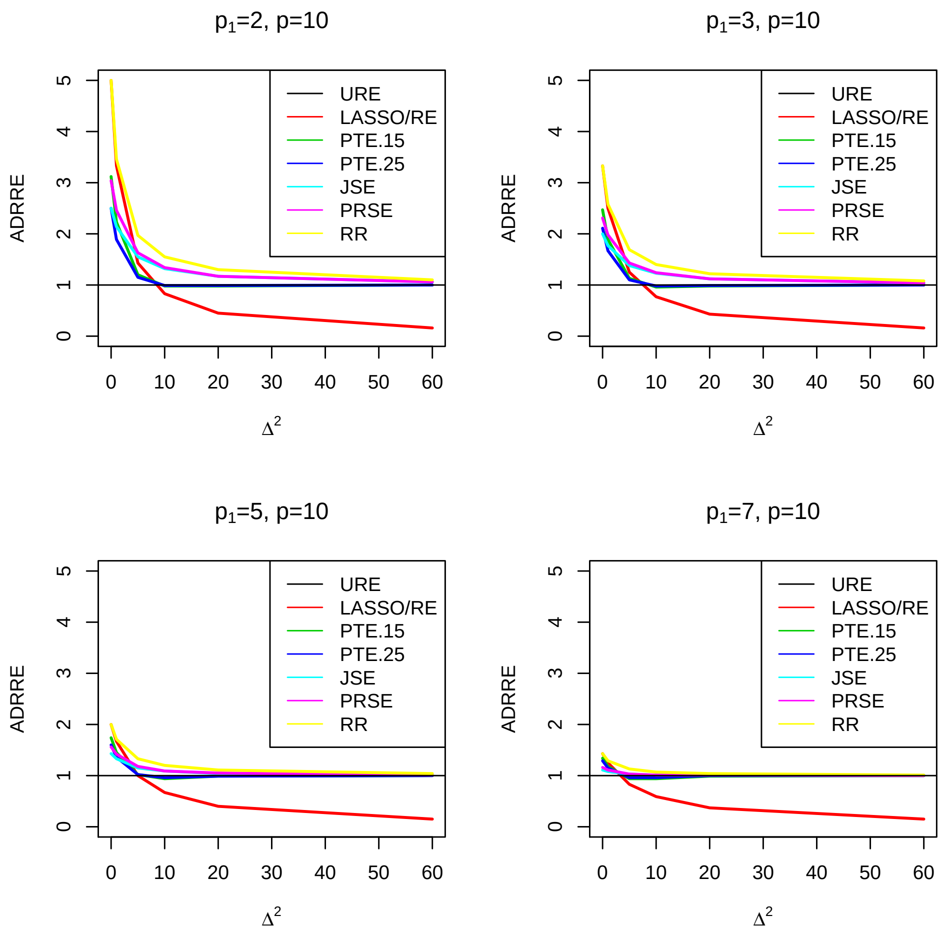

Fig. 1. ARRE of estimates as a function of Δ^2 for $p = 10$.

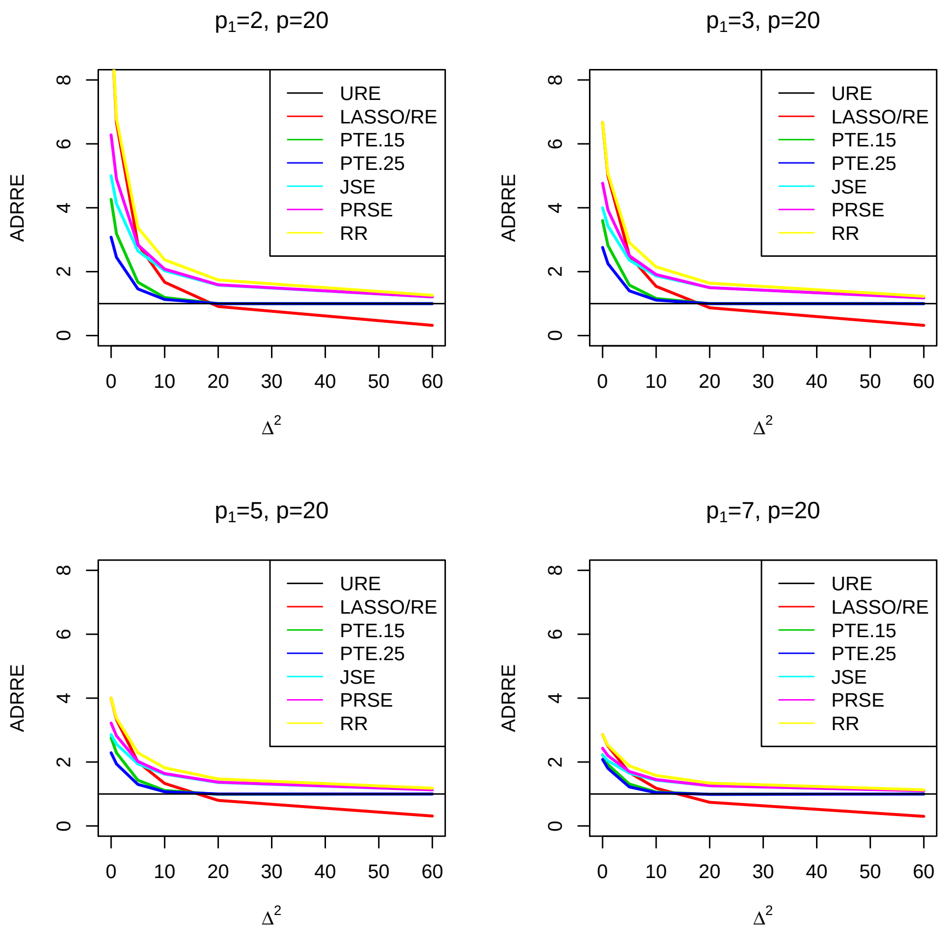

Fig. 2. ARRE of estimates as a function of Δ^2 for $p = 20$.

2.5.11. Comparison of RR with LASSO and RE

$$\mathrm{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n\right) - \mathrm{AL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{RR}(k_{opt})\right) = \eta^2 \frac{\Delta^4}{\Delta^2 + p_2} \geq 0.$$

Hence, RR uniformly dominates RE and LASSO.

3. SUMMARY AND CONCLUDING REMARKS

In this section, we discuss the contents of tables 3 − 4 and figures 1 − 4 prepared out of the analysis of the characteristics of the R-estimators.

In tables 3 − 4 ARRE's for $p = 10, 20, 40$ with $(p_1, p_2) = (5, 15)$, $(7, 33)$, $(2, 8)$, $(3, 7)$, $(4, 6)$, $(5, 5)$ are reported. Table 3 presents the ARRE values of the R-estimators for

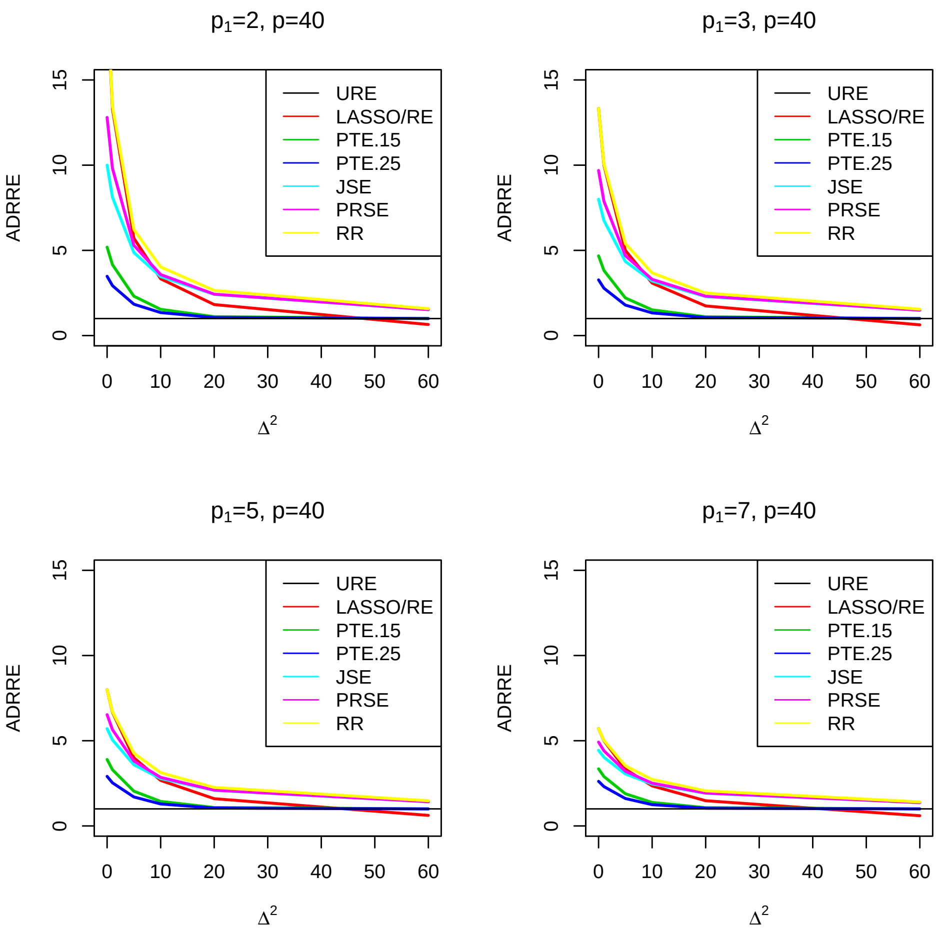

Fig. 3. ARRE of estimates as a function of Δ^2 for $p = 40$.

some selected values of Δ^2. From this table, we observe that ridge estimator uniformly dominates URE, PTE and Stein-type R-estimators. On the other hand, RE and LASSO outperform URE, PTE, JSE and PRSE in the subinterval $(0, p_2)$. If p_1 is fixed and p_2 varies, then ARRE increases for Δ^2 (see tables 3−3). However, if p_2 is fixed and p_1 varies, then the ARRE of all R-estimators decreases for each value of Δ^2. Then, for p_2 small and p_1 large the ADRRE of LASSO, PTE, JSE and PRSE are competitive – see table 4. Further, we found the order of ARRE of URE, JSE, PRSE and RR as

$$\text{ARRE(RR:URE)} \geq \text{ARRE(PRSE:URE)} \geq \text{ARRE(JSE:URE)} \geq 1$$

uniformly, and that of RR and LASSO as

$$\text{ARRE(LASSO:URE)} = \text{ARRE(RR:URE)} \quad \text{in the subinterval} \quad (0, 1).$$

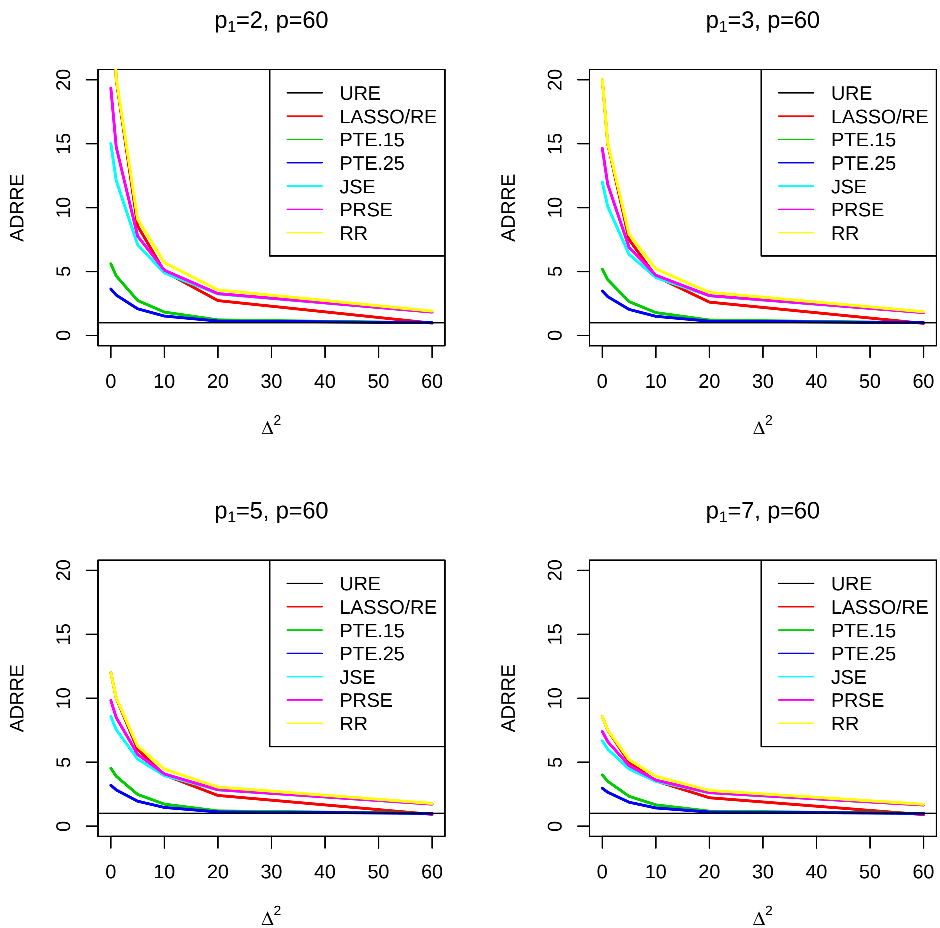

Fig. 4. ARRE of estimates as a function of Δ^2 for $p = 60$.

$$\text{and} \quad \text{ARRE(LASSO:URE)} < \text{ARRE(RR:URE)} \quad \text{in the interval} \quad [1, \infty).$$

Finally, PRSE always outperforms JSE.

We mention a few features of the ARRE expressions, using LSE method, are the same as the ones we have here, see [12]. Simulation results confirm our findings. In the simulation study, we also investigated the finite sample behavior of the R-estimates. However, the corresponding results are in accordance with those presented here for the asymptotics. It is caused by the fast convergence of the estimates to their limit distribution.

Further, we considered an ARRE expression for changing c.d.f F, but the ARRE's do not change, We studied the high-dimensional problem ($p = p_n > n$) and found the same ARRE expressions. The results of these findings will be reported in a separate paper.

		$p_1 = 2$				$p_1 = 3$			
Δ^2	URE	$p = 10$	$p = 20$	$p = 40$	$p = 60$	$p = 10$	$p = 20$	$p = 40$	$p = 60$
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

Tab. 1. ARRE of LASSO & RE as a function of Δ^2 for different (p_1, p_2).

	p=10							
	$\Delta^2 = 0$				$\Delta^2 = 1.00$			
Estimators	$p_1 = 2$	$p_1 = 3$	$p_1 = 5$	$p_1 = 7$	$p_1 = 2$	$p_1 = 3$	$p_1 = 5$	$p_1 = 7$
URE	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
PRSE	3.04	2.31	1.56	1.16	2.46	1.98	1.42	1.11
RR	5.00	3.33	2.00	1.43	3.46	2.58	1.71	1.29
	$\Delta^2 = 5$				$\Delta^2 = 10$			
URE	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
PRSE	1.63	1.43	1.18	1.03	1.34	1.24	1.09	1.01
RR	1.97	1.69	1.33	1.13	1.55	1.40	1.20	1.07
	$\Delta^2 = 20$				$\Delta^2 = 60$			
URE	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
PRSE	1.17	1.12	1.05	1.01	1.06	1.04	1.01	1.00
RR	1.30	1.22	1.11	1.04	1.10	1.08	1.04	1.01
	p=20							
	$\Delta^2 = 0$				$\Delta^2 = 1.00$			
Estimators	$p_1 = 2$	$p_1 = 3$	$p_1 = 5$	$p_1 = 7$	$p_1 = 2$	$p_1 = 3$	$p_1 = 5$	$p_1 = 7$
URE	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
PRSE	6.28	4.77	3.22	2.43	4.90	3.93	2.82	2.20
RR	10.00	6.67	4.00	2.86	6.79	5.07	3.37	2.52
	$\Delta^2 = 5$				$\Delta^2 = 10$			
URE	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
PRSE	2.84	2.50	2.02	1.70	2.08	1.91	1.64	1.45
RR	3.38	2.91	2.29	1.88	2.37	2.15	1.82	1.58
	$\Delta^2 = 20$				$\Delta^2 = 60$			
URE	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
PRSE	1.59	1.50	1.37	1.26	1.21	1.18	1.13	1.10
RR	1.74	1.64	1.47	1.34	1.26	1.23	1.18	1.13

Tab. 2. ARRE of the estimators for $p = 10, 20$ and different Δ^2-value for varying p_2 with fixed p_1.

	p=40							
	$\Delta^2 = 0$				$\Delta^2 = 1.00$			
Estimators	$p_1 = 2$	$p_1 = 3$	$p_1 = 5$	$p_1 = 7$	$p_1 = 2$	$p_1 = 3$	$p_1 = 5$	$p_1 = 7$
URE	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
PRSE	12.80	9.69	6.53	4.92	9.81	7.88	5.65	4.41
RR	20.00	13.33	8.00	5.71	13.45	10.07	6.70	5.02
	$\Delta^2 = 5$				$\Delta^2 = 10$			
URE	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
PRSE	5.30	4.69	3.81	3.21	3.58	3.30	2.85	2.51
RR	6.23	5.40	4.27	3.53	4.03	3.68	3.13	2.73
	$\Delta^2 = 20$				$\Delta^2 = 60$			
URE	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
PRSE	2.43	2.31	2.10	1.93	1.52	1.49	1.42	1.37
RR	2.65	2.50	2.26	2.06	1.58	1.55	1.48	1.41
	p=60							
	$\Delta^2 = 0$				$\Delta^2 = 1.00$			
Estimators	$p_1 = 2$	$p_1 = 3$	$p_1 = 5$	$p_1 = 7$	$p_1 = 2$	$p_1 = 3$	$p_1 = 5$	$p_1 = 7$
URE	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
PRSE	19.35	14.63	9.83	7.40	14.74	11.83	8.49	6.61
RR	30.00	20.00	12.00	8.57	20.11	15.06	10.03	7.52
	$\Delta^2 = 5$				$\Delta^2 = 10$			
URE	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
PRSE	7.77	6.88	5.61	4.73	5.10	4.71	4.08	3.60
RR	9.09	7.90	6.26	5.19	5.70	5.21	4.46	3.89
	$\Delta^2 = 20$				$\Delta^2 = 60$			
URE	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
PRSE	3.28	3.12	2.85	2.62	1.83	1.79	1.72	1.65
RR	3.56	3.37	3.05	2.79	1.91	1.86	1.78	1.71

Tab. 3. ARRE of the estimators for $p = 40, 60$ and different Δ^2-value for varying p_2 with fixed p_1.

	URE	LASSO/RE	PTE.15	PTE.20	PTE.25	JSE	PRSE	RR
p_2	$p_1 = 5$ and $\Delta^2 = 0$							
5	1.00	2.00	1.74	1.67	1.60	1.43	1.56	2.00
15	1.00	4.00	2.76	2.50	2.29	2.86	3.22	4.00
35	1.00	8.00	3.90	3.33	2.91	5.71	6.53	8.00
55	1.00	12.00	4.53	3.75	3.20	8.57	9.83	12.00
p_2	$p_1 = 5$ and $\Delta^2 = 1$							
5	1.00	1.67	1.45	1.40	1.36	1.33	1.42	1.71
15	1.00	3.33	2.29	2.10	1.94	2.56	2.82	3.37
35	1.00	6.67	3.30	2.86	2.53	5.05	5.65	6.70
55	1.00	10.00	3.90	3.27	2.83	7.55	8.49	10.03
p_2	$p_1 = 5$ and $\Delta^2 = 5$							
5	1.00	1.00	1.02	1.02	1.02	1.15	1.18	1.33
15	1.00	2.50	1.58	1.48	1.40	2.36	2.50	2.91
35	1.00	4.00	2.04	1.84	1.70	3.59	3.81	4.27
55	1.00	6.00	2.48	2.17	1.95	5.25	5.61	6.26
p_2	$p_1 = 5$ and $\Delta^2 = 10$							
5	1.00	0.67	0.94	0.95	0.96	1.09	1.09	1.20
15	1.00	1.33	1.11	1.09	1.07	1.62	1.64	1.82
35	1.00	2.67	1.44	1.35	1.29	2.78	2.85	3.13
55	1.00	4.00	1.72	1.57	1.46	3.95	4.08	4.46
p_2	$p_1 = 7$ and $\Delta^2 = 0$							
3	1.00	1.43	1.34	1.32	1.29	1.11	1.16	1.43
13	1.00	2.86	2.23	2.08	1.95	2.22	2.43	2.86
33	1.00	5.71	3.35	2.94	2.62	4.44	4.92	5.71
53	1.00	8.57	4.01	3.41	2.96	6.67	7.40	8.57
p_2	$p_1 = 7$ and $\Delta^2 = 1$							
3	1.00	1.25	1.18	1.16	1.15	1.08	1.11	1.29
13	1.00	2.50	1.92	1.80	1.70	2.04	2.20	2.52
33	1.00	5.00	2.90	2.57	2.31	4.03	4.41	5.02
53	1.00	7.50	3.51	3.02	2.65	6.03	6.61	7.52
p_2	$p_1 = 7$ and $\Delta^2 = 5$							
3	1.00	0.83	0.94	0.95	0.96	1.03	1.03	1.13
13	1.00	1.67	1.31	1.26	1.22	1.65	1.70	1.88
33	1.00	3.33	1.89	1.73	1.61	3.05	3.21	3.53
53	1.00	5.00	2.33	2.07	1.87	4.47	4.73	5.19
p_2	$p_1 = 7$ and $\Delta^2 = 10$							
3	1.00	0.59	0.94	0.95	0.96	1.01	1.01	1.07
13	1.00	1.18	1.06	1.05	1.04	1.43	1.45	1.58
33	1.00	2.35	1.38	1.31	1.25	2.46	2.51	2.73
53	1.00	3.53	1.66	1.52	1.42	3.50	3.60	3.89

Tab. 4.
ARRE of the estimators for $p_1 = 5, 7$ for different values of p_2 and Δ^2.

ACKNOWLEDGEMENTS

Authors would like to thank the two referees for their valuable comments that helped significantly to improve the paper.

The research of Radim Navrátil was supported by Student Project Grant at MU (specific research, rector's programme) MUNI/A/1204/2017.

(Received March 13, 2017)

REFERENCES

[1] A. Belloni A and V. Chernozhukov: Least squares after model selection in high-dimensional sparse models. Bernoulli *19* (2013), 521–547. DOI:10.3150/11-bej410

[2] L. Breiman: Heuristics of instability and stabilization in model selection. Ann. Statist. *24* (1996), 2350–2383. DOI:10.1214/aos/1032181158

[3] D. L. Donoho and I. M. Johnstone: Minimax estimation via wavelet shrinkage. Ann. Statist. *26* (1994), 879–921. DOI:10.1214/aos/1024691081

[4] N. R. Draper and R. C. Van Nostrand: Ridge regression and James-Stein estimation: review and comments. Technometrics *21* (1979), 451–466. DOI:10.2307/1268284

[5] J. Fan and R. Li: Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist. Assoc. *96* (2001), 1348–1360. DOI:10.1198/016214501753382273

[6] L. E. Frank and J. H. Friedman: A statistical view of some chemometrics regression tools. Technometrics *35* (1993), 109–135. DOI:10.1080/00401706.1993.10485033

[7] E. Hoerl and R. W. Kennard: Ridge regression: Biased estimation for nonorthogonal problems. Technometrics *12* (1970), 55–67. DOI:10.1080/00401706.1970.10488634

[8] W. James and C. Stein: Estimation with quadratic loss. In: Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics, University of California Press 1961, pp. 361–379.

[9] J. Jurečková: Nonparametric estimate of regression coefficients. Ann. Math. Statist. *42* (1971), 1328–1338. DOI:10.1214/aoms/1177693245

[10] B. E. Hansen: The risk of James–Stein and Lasso shrinkage. Econometric Rev. *35* (2015), 456–470.

[11] A. K. Md. E. Saleh: Theory of Preliminary test and Stein-Type Estimators with Applications. John Wiley and Sons, New York 2006. DOI:10.1002/0471773751

[12] A. K. Md. E. Saleh, M. Arashi, M. Norouzirad, and B. M. G. Kibria: On shrinkage and selection: ANOVA MODEL. J. Statist. Res. *51* (2017), 165–191.

[13] C. Stein: Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: Proc. Third Berkeley Symposium on Mathematical Statistics and Probability, University of California Press 1956, pp. 197–206.

[14] R. Tibshirani: Regression shrinkage and selection via the lasso. J. Royal Statist. Soc., Series B (Methodological) *58* (1996), 267–288.

[15] A. N. Tikhonov: Solution of incorrectly formulated problems and the regularization method. Doklady Akademii Nauk SSSR *151* (1963), 501–504.

[16] H. Zou: The adaptive lasso and its oracle properties. J. Amer. Statist. Assoc. *101* (2006), 1418–1429. DOI:10.1198/016214506000000735

[17] H. Zou and T. Hastie: Regularization and variable selection via the elastic net. J. Royal Stat. Soc. Ser. B Stat. Methodol. *67* (2005), 301–320. DOI:10.1111/j.1467-9868.2005.00503.x

A. K. Md. Ehsanes Saleh, Carleton University, Ottawa, ON. Canada.
e-mail: esaleh@math.carleton.ca

Radim Navrátil, Masaryk University, Department of Mathematics and Statistics, Brno. Czech Republic.
e-mail: navratil@math.muni.cz