The Canadian Journal of Statistics
Vol. 46, No. 4, 2018, Pages 690–704
La revue canadienne de statistique

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

A. K. Md. Ehsanes SALEH[1], Radim NAVRÁTIL[2], and Mina NOROUZIRAD[3]

[1] *School of Mathematics and Statistics, Carleton University, 1125 Colonel By Drive, Ottawa, ON K1S 5B6, Canada*
[2] *Department of Mathematics and Statistics, Masaryk University, Brno 611 37, Czech Republic*
[3] *Department of Statistics, Faculty of Mathematical Sciences, Shahrood University of Technology, P.O. Box 316-3619995181, Shahrood, Iran*

Key words and phrases: Efficiency of LASSO; L_2-risk function; Penalty estimators; Preliminary test and Stein-type estimator; Rank estimators.

MSC 2010: Primary 62G05; secondary 62J05

Abstract: In the development of efficient predictive models, the key is to identify suitable predictors to establish a prediction model for a given linear or nonlinear model. This paper provides a comparative study of ridge regression, least absolute shrinkage and selector operator (LASSO), preliminary test (PTE) 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, PTE 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 dominance of LASSO over all the R-estimators (except the ridge R-estimator) is the sparsity-dimensional interval around the origin of the parameter space. 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. *The Canadian Journal of Statistics* 46: 690–704; 2018 © 2018 Statistical Society of Canada

Résumé: Le développement de modèles prédictifs efficaces passe par le choix de prédicteurs appropriés pour un modèle linéaire ou non. Les auteurs présentent une étude comparative de la régression ridge, du lasso, du test préliminaire et des estimateurs de type Stein, basée sur la théorie des statistiques de rangs. Avec une matrice de design orthonormale pour un modèle linéaire, les auteurs constatent que l'estimateur de ridge basé sur les rangs présente une performance uniformément supérieure à l'estimateur des rangs habituel, au R-estimateur restreint, au lasso basé sur les rangs, au test préliminaire et aux R-estimateurs de type Stein. Ils constatent également qu'aucune des options parmi le lasso, le R-estimateur habituel, le test préliminaire et les R-estimateurs de type Stein ne surpasse les autres. Dans un intervalle de dimension éparse autour de l'origine de l'espace paramétrique, le lasso domine tous les R-estimateurs (sauf le R-estimateur de ridge). Les auteurs observent que le risque L_2 du R-estimateur restreint est égal à la borne inférieure du risque L_2 du lasso. Leurs conclusions reposent sur une analyse du risque L_2 et de l'efficacité relative du risque L_2 présentée par des tables et figures. *La revue canadienne de statistique* 46: 690–704; 2018 © 2018 Société statistique du Canada

Additional Supporting Information may be found in the online version of this article at the publisher's website.
* *Author to whom correspondence may be addressed.*
E-mail: esaleh@math.carleton.ca

1. INTRODUCTION

Consider the multiple regression model,

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

where $\boldsymbol{Y} = (Y_1, \ldots ,Y_n)^\top$ is the response $n \times 1$ vector, $\boldsymbol{X}$ is the $n \times p$ matrix of real numbers, $\boldsymbol{\beta}$ is the $p \times 1$ vector ($p \leq n$) of unknown regression parameters, $\mathbf{1}_n = (1, \ldots , 1)^\top$ is the $n \times 1$ vector of 1's, and θ is an intercept parameter. The error vector $\boldsymbol{e} = (e_1, \ldots , e_n)^\top$ has mutually independent components e_i which are independent and identically distributed (i.i.d.) random variables having a cumulative distribution function (c.d.f.) F defined on the real line $\mathbb{R}$.

We assume that the design matrix $\boldsymbol{X}$ has a full rank p and consider the partitioning

$$\boldsymbol{\beta} = \left(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top\right)^\top \qquad \text{and} \qquad \boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2),$$

where $\boldsymbol{\beta}_1$ is a p_1-dimensional and $\boldsymbol{\beta}_2$ is a p_2-dimensional vector with $p = p_1 + p_2$, so that (1) is rewritten as

$$\boldsymbol{Y} = \theta \mathbf{1}_n + \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e}. \tag{2}$$

Expression (2) allows us to effectively examine the settings with sparse parameters. If we suspect that $\boldsymbol{\beta}$ is sparse, we can express this by setting the sparsity condition as $\boldsymbol{\beta}_2 = \mathbf{0}$ or by making all the $\boldsymbol{\beta}_2$ components small.

Estimation and variable selection are important aspects for the development of model fitting and data analysis. The history of estimation theory changed its course radically since Stein (1956) and James & Stein (1961) proved that the 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 many areas of applications. Saleh & Sen (1978), Sen & Saleh (1987) and Saleh (2006) reformulated and expanded Stein's theory using the least squares theory, rank theory, M-theory and quantile theory beginning in the 70s.

The next generation of ''shrinkage estimators'', known as penalty estimators, began in the 1970s with the pioneering work on ''ridge regression'' estimation for linear models by Hoerl & Kennard (1970) based on the idea of ''Tikhonov regularization'' (Tikhonov, 1963). The ridge regression estimator is the result of minimizing least squares criteria subject to some quadratic restrictions (L$_2$-function),

$$\widehat{\boldsymbol{\beta}}_n^{\text{RR}}(k) = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \left\{(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + k\boldsymbol{\beta}^\top\boldsymbol{\beta}\right\} \qquad \text{with} \qquad k > 0. \tag{3}$$

A generalized ridge regression estimator may be defined as

$$\widehat{\boldsymbol{\beta}}_n^{\text{RS}}(\boldsymbol{K}) = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \left\{(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + \boldsymbol{\beta}^\top\boldsymbol{K}\boldsymbol{\beta}\right\}, \tag{4}$$

where $\boldsymbol{K} = \text{diag}\,(k_1, \ldots , k_p)$ and $k_j > 0$ for $j = 1, \ldots , p$. Note that the penalty function (3) places equal weights on the $\boldsymbol{\beta}$'s, while (4) places unequal weights.

Frank & Friedman (1993) defined a class of ''bridge estimators'' defined by

$$\widehat{\boldsymbol{\beta}}_n^{\text{BE}}(\lambda_n) = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\operatorname{argmin}} \left\{(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^\top(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) + \lambda_n\mathbf{1}_p^\top|\boldsymbol{\beta}|^\gamma\right\},$$

where $\lambda_n > 0$ and $|\boldsymbol{\beta}|^\gamma = (|\beta_1|^\gamma, \ldots , |\beta_p|^\gamma)^\top$ with $\gamma > 0$.

The choice of $\gamma = 2$ produces ridge estimates while $\gamma = 1$ relates to least absolute shrinkage and selector operator (LASSO) introduced by Tibshirani (1996). It has become a very popular and intriguing penalty estimator. This estimator is related to the estimators such as the ''non-negative'' garotte by Breiman (1996), smoothly clipped absolute deviation (SCAD) by Fan & Li (2001), elastic net by Zou & Hastie (2005), adaptive LASSO by Zou (2006), hard threshold LASSO by Belloni & Chernozhukov (2013), and by many others.

This paper introduces the R-estimators and the application of marginal distribution theory to study the performance characteristics of two primary penalty estimators, namely ''ridge regression'' and ''LASSO'' along with the preliminary test (PT) and Stein-type estimators. In this respect, the work of Draper & Van Nostrand (1979) and Hansen (2016) are informative. An important characteristic of LASSO is that it provides simultaneous estimation and selection coefficients for linear models and can be applied when the dimension of the parameter space exceeds the dimensions of the sample space.

The layout of this article is as follows. In Section 2, we define an ordinary R-estimator and introduce its improved estimator as well as a penalty R-estimator. Section 3 is devoted to asymptotic distributional bias (ADB) and L$_2$-risk (ADL$_2$-risk) of the R-estimators and Section 4 deals with the graphical and numerical assessment of the R-estimators. A real data example is discussed in Section 5.

2. LINEAR MODEL AND R-ESTIMATORS

Consider the multiple regression model (2). We assume that:

(i) Errors are i.i.d. random variables with (unknown) c.d.f. F having an absolutely continuous probability density function (p.d.f.) f with finite and nonzero Fisher information,

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

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

$$A_\varphi^2 = \int_0^1 \varphi^2(u)du - \left(\int_0^1 \varphi(u)\, du \right)^2.$$

The scores are defined in either of the following ways:

$$a_n(i) = \mathbb{E}[\varphi(U_{i:n})], \qquad \text{or} \qquad a_n(i) = \varphi\left(\frac{i}{n+1}\right), \qquad i = 1, \dots, n,$$

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

(iii) Define

$$\boldsymbol{C}_n = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{x}_i - \bar{\boldsymbol{x}}_n)(\boldsymbol{x}_i - \bar{\boldsymbol{x}}_n)^\top, \tag{5}$$

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

$$\lim_{n\to\infty} \boldsymbol{C}_n = \boldsymbol{I}_p \qquad \text{and} \quad \lim_{n\to\infty} \max_{1\le i\le n} (\boldsymbol{x}_i - \bar{\boldsymbol{x}}_n)^\top \boldsymbol{C}_n^{-1} (\boldsymbol{x}_i - \bar{\boldsymbol{x}}_n) = 0. \tag{6}$$

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

$$\boldsymbol{L}_n(\boldsymbol{b}) = (L_{n1}(\boldsymbol{b}), \dots, L_{nn}(\boldsymbol{b}))^\top = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\boldsymbol{x}_i - \bar{\boldsymbol{x}}_n) a_n(R_{ni}(\boldsymbol{b})). \tag{7}$$

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

For the R-estimation of $\boldsymbol{\beta}$, one may use the rank-based objective function due to the following result in Jaeckel (1972),

$$D_n(\boldsymbol{b}) = \sum_{i=1}^{n} (Y_i - \boldsymbol{x}_i^\top \boldsymbol{b}) a_n(R_{ni}(\boldsymbol{b})), \tag{8}$$

where $D_n(\boldsymbol{b})$ is a nonnegative, continuous, piecewise linear and convex function of $\boldsymbol{b} \in \mathbb{R}^p$. Thus, we define the unrestricted R-estimator (URE) as

$$\widetilde{\boldsymbol{\beta}}_n = \underset{\boldsymbol{b} \in \mathbb{R}^p}{\operatorname{argmin}} \, D_n(\boldsymbol{b}).$$

We note the connection between (7) and (8) as

$$\nabla D_n(\mathbf{b}) = -\boldsymbol{L}_n(\boldsymbol{b}),$$

where ∇ is the sub-gradient.

Next, we review the asymptotic uniform linearity (AUL) result due to Jurečková (1971), Jurečková & Sen (1996), and the asymptotic uniform quadraticity (AUQ) result due to Jaeckel (1972), respectively, given by

$$\lim_{n\to\infty} P\left(\sup_{\|\boldsymbol{\omega}\|<k} \left\| \boldsymbol{L}_n\left(\boldsymbol{\beta} + \frac{\boldsymbol{\omega}}{\sqrt{n}}\right) - \boldsymbol{L}_n(\boldsymbol{\beta}) + \gamma(\varphi, f)\boldsymbol{\omega} \right\| > \varepsilon \right) = 0,$$

and

$$\lim_{n\to\infty} P\left(\sup_{\|\boldsymbol{\omega}\|<k} \left\| \boldsymbol{W}_n(\boldsymbol{\omega}) \right\| > \varepsilon \right) = 0,$$

where ε is non-negative,

$$\boldsymbol{W}_n(\boldsymbol{\omega}) = \boldsymbol{D}_n(\boldsymbol{\beta} + n^{-1/2}\boldsymbol{\omega}) - \boldsymbol{D}_n(\boldsymbol{\beta}) + \gamma(\varphi, f)\boldsymbol{\omega}^\top \boldsymbol{L}_n(\boldsymbol{\beta}) - \frac{1}{2}\gamma^2(\varphi, f)\boldsymbol{\omega}^\top \boldsymbol{\omega},$$

and

$$\gamma(\varphi, f) = \int_0^1 \varphi(u) \left\{ -\frac{f'\left(F^{-1}(u)\right)}{f(F^{-1}(u))} \right\} du.$$

Thus, we conclude that as $n \to \infty$,

$$\left\| \sqrt{n}(\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) - \underset{\boldsymbol{\omega} \in \mathbb{R}^p}{\operatorname{argmin}} \, \boldsymbol{W}_n(\boldsymbol{\omega}) \right\| \xrightarrow{P} 0,$$

so that

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) - \gamma^{-1}(\varphi, f)\boldsymbol{L}_n(\boldsymbol{\beta}) \xrightarrow{P} 0.$$

Hence,

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

where $\eta^2 = A_\varphi^2/\gamma^2(\varphi, f)$.

The same results may be obtained by considering

$$\boldsymbol{L}_n^*(\boldsymbol{b}) = (L_{n(j)}(b_j, \boldsymbol{b}_{(-j)}); \;\; j = 1, \ldots, p)^\top, \tag{9}$$

where

$$L_{n(j)}(b_j, \boldsymbol{b}_{(-j)}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (x_{ij} - \bar{x}_j) a_n(R_{ni}(b_j, \boldsymbol{b}_{(-j)})), \tag{10}$$

where b_j is the jth coefficient and $\boldsymbol{b}_{(-j)}$ is a $(p-1) \times 1$ vector excluding the coefficient b_j. The components of (10) may be called marginal statistics.

Thus, using AUL and AUQ, we find that as $n \to \infty$

$$\sup_{|\omega_j|<k} |L_{n(j)}(n^{-1/2}\omega_j \boldsymbol{b}_{(-j)}) - L_{n(j)}(0, \boldsymbol{0}) + \gamma(\varphi, f)\omega_j| \xrightarrow{P} 0,$$

and

$$\sup_{|\omega_j|<k} \left| D_n\left(n^{-1/2}\omega_j \boldsymbol{b}_{(-j)}\right) - D_n(0, \boldsymbol{0}) + \gamma(\varphi, f)\omega_j L_{n(j)}(0, \boldsymbol{0}) - \frac{1}{2}\gamma^2(\varphi, f)\omega_j^2 \right| \xrightarrow{P} 0.$$

Consequently, $n \to \infty$,

$$\sqrt{n}(\widetilde{\beta}_{jn} - \beta_j) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \eta^2) \qquad \text{for all} \qquad j = 1, \ldots, n. \tag{11}$$

Note that $\mathbf{L}_n^*(b_1, \ldots, b_p)$ at (9) consists of mutually independent marginal components of (10).

2.1. Penalty R-Estimators

Based on (11) we are able to define the LASSO R-estimator following Donoho & Johnstone (1994) as

$$\begin{aligned} \widehat{\boldsymbol{\beta}}_n^{\text{LASSO}}(\lambda) &= \left(\text{sign}\left(\widetilde{\beta}_{jn}\right) \left(\widetilde{\beta}_{jn} - \frac{\lambda}{\sqrt{n}}\eta \right)^+ ; \;\; j = 1, \ldots, p \right)^\top \\ &= \left(\frac{\eta}{\sqrt{n}} \text{sign}\left(Z_j\right) (|Z_j| - \lambda)^+ ; \;\; j = 1, \ldots, p \right)^\top, \end{aligned}$$

where $Z_j = \frac{\sqrt{n}}{\eta}\widetilde{\beta}_{jn}$. Note that

$$\sqrt{n}\widehat{\beta}_{jn}^{\text{LASSO}}(\lambda) = \begin{cases} \eta\left(Z_j - \lambda\text{sign}\left(Z_j\right)\right), & \text{if } |Z_j| > \lambda, \\ 0, & \text{otherwise.} \end{cases}$$

Let $| Z_1 |> \lambda, \ldots, | Z_{p_1} |> \lambda$ and the rest of the p_2 components are 0's, then

$$\sqrt{n}\widehat{\boldsymbol{\beta}}_n^{\text{LASSO}}(\lambda) = (\eta(Z_j - \lambda\text{sign}(Z_j)), j = 1, \ldots, p_1, \mathbf{0}^\top)^\top$$

is the vector of the LASSO estimator.

Next, we consider the ridge estimator of $\boldsymbol{\beta}$, where one suspects that $\boldsymbol{\beta}_2 = \mathbf{0}$ may hold. We define

$$\widehat{\boldsymbol{\beta}}_n^{\text{RR}}(k) = \underset{(\boldsymbol{b}_1^\top, \boldsymbol{b}_2^\top)^\top \in \mathbb{R}^p}{\text{argmin}} \{D_n(\boldsymbol{b}_1, \boldsymbol{b}_2) + k\|\boldsymbol{b}_2\|^2\} = \left(\widetilde{\boldsymbol{\beta}}_{1n}^\top, \frac{1}{1+k}\widetilde{\boldsymbol{\beta}}_{2n}^\top\right)^\top.$$

Our problem is to compare the performance characteristics of ''ridge'' and LASSO estimators with that of the Stein-type and preliminary test R-estimators with respect to the asymptotic distributional 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 (2), if we suspect a sparsity condition that $\boldsymbol{\beta}_2 = \mathbf{0}$, then the restricted R-estimator (RE) of $(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ is $\widehat{\boldsymbol{\beta}}_n = (\widetilde{\boldsymbol{\beta}}_{1n}^\top, \mathbf{0}^\top)^\top$. For the test of the null hypothesis $\mathcal{H}_o : \boldsymbol{\beta}_2 = \mathbf{0}$ vs. $\mathcal{H}_A : \boldsymbol{\beta}_2 \neq \mathbf{0}$, the rank statistic is given by

$$\mathfrak{L}_n = nA_n^2\boldsymbol{L}_{2n}^\top(\mathbf{0})\boldsymbol{L}_{2n}(\mathbf{0}), \tag{12}$$

where $\boldsymbol{L}_n(\mathbf{0}) = (\boldsymbol{L}_{1n}^\top(\mathbf{0}), \boldsymbol{L}_{2n}^\top(\mathbf{0}))^\top$ from (7),

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

It is well known that under model (2) and the assumptions (5) and (6) as $n \to \infty$, $\mathfrak{L}_n$ follows the χ^2 distribution with p_2 degrees of freedom (d.f.) under $\mathcal{H}_o$. Then, we define the PTE estimator of $(\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top$ as

$$\widehat{\boldsymbol{\beta}}_n^{\text{PT}} = (\widetilde{\boldsymbol{\beta}}_{1n}^\top, \widetilde{\boldsymbol{\beta}}_{2n}^\top - \widetilde{\boldsymbol{\beta}}_{2n}^\top I(\mathfrak{L}_n < \chi_{p_2}^2(\alpha)))^\top,$$

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

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

$$\widehat{\boldsymbol{\beta}}_n^{\text{JS}} = (\widetilde{\boldsymbol{\beta}}_{1n}^\top, \widetilde{\boldsymbol{\beta}}_{2n}^\top(1 - (p_2 - 2)\mathfrak{L}_n^{-1}))^\top,$$

and the positive-rule Stein-type estimator is given by

$$\widehat{\boldsymbol{\beta}}_n^{\text{S+}} = (\widetilde{\boldsymbol{\beta}}_{1n}^\top, \widetilde{\boldsymbol{\beta}}_{2n}^\top(1 - (p_2 - 2)\mathfrak{L}_n^{-1})I(\mathfrak{L}_n > p_2 - 2))^\top.$$

3. ASYMPTOTIC DISTRIBUTIONAL BIAS AND L$_2$-RISKS OF THE ESTIMATORS

We test the null hypothesis $\mathcal{H}_o : \boldsymbol{\beta}_2 = \mathbf{0}$ vs. $\mathcal{H}_a : \boldsymbol{\beta}_2 \neq \mathbf{0}$ based on the rank statistics $\mathfrak{L}_n$ of (12). This test is consistent and its power tends to unity as $n \to \infty$ for fixed alternatives. We consider a sequence of Pitman's alternatives $\mathcal{H}_{A(n)}$ defined by

$$\mathcal{H}_{A(n)} : \boldsymbol{\beta}_n = n^{-1/2}\boldsymbol{\delta} = n^{-1/2}(\boldsymbol{\delta}_1^\top, \boldsymbol{\delta}_2^\top)^\top.$$

If $\boldsymbol{\delta}_2 = \mathbf{0}$, $\mathcal{H}_{A(n)} = \mathcal{H}_o$. Then, under $\{\mathcal{H}_{A(n)}\}_{n=1}^{\infty}$, 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 ADB and asymptotic distributional L$_2$-risks (ADL$_2$-risks) of the R-estimators are given below.

(i) URE:

$$(\mathrm{ADB}^\top(\widetilde{\boldsymbol{\beta}}_{1n}), \mathrm{ADB}^\top(\widetilde{\boldsymbol{\beta}}_{2n})) = (\mathbf{0}^\top, \mathbf{0}^\top),$$
$$(\mathrm{ADL}_2\text{-risk}(\widetilde{\boldsymbol{\beta}}_{1n}), \mathrm{ADL}_2\text{-risk}(\widetilde{\boldsymbol{\beta}}_{2n})) = \eta^2(p_1, p_2).$$

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

(ii) RE:

$$\mathrm{ADB}^\top(\widetilde{\boldsymbol{\beta}}_{1n}), \mathrm{ADB}^\top(\mathbf{0})) = (\mathbf{0}^\top, -\boldsymbol{\delta}_2^\top),$$
$$(\mathrm{ADL}_2\text{-risk}(\widetilde{\boldsymbol{\beta}}_{1n}), \mathrm{ADL}_2\text{-risk}(\mathbf{0})) = \eta^2(p_1, \Delta^2).$$

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

(iii) PTE:

$$\left(\mathrm{ADB}^\top\left(\widetilde{\boldsymbol{\beta}}_{1n}\right), \mathrm{ADB}^\top\left(\widehat{\boldsymbol{\beta}}_{2n}^{\mathrm{PT}}\right)\right) = (\mathbf{0}^\top, -\boldsymbol{\delta}_2^\top \mathcal{H}_{p_2+2}\left(\chi^2_{p_2}(\alpha); \Delta^2\right),$$
$$\left(\mathrm{ADL}_2\text{-risk}\left(\widetilde{\boldsymbol{\beta}}_{1n}\right), \mathrm{ADL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_{2n}^{\mathrm{PT}}\right)\right) = \left(\eta^2 p_1, \eta^2\left[p_2\left(1 - \mathcal{H}_{p_2+2}\left(\chi^2_{p_2}(\alpha); \Delta^2\right)\right)\right.\right.$$
$$\left.\left. + \Delta^2\left(2\mathcal{H}_{p_2+2}\left(\chi^2_{p_2}(\alpha); \Delta^2\right) - \mathcal{H}_{p_2+4}\left(\chi^2_{p_2}(\alpha); \Delta^2\right)\right)\right]\right).$$

Hence, the ADL$_2$-risk of $\widehat{\boldsymbol{\beta}}_n^{\mathrm{PT}}$ is given by

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

where $H_\nu(c;\Delta^2)$ is the c.d.f. of χ^2-distribution with ν d.f. and non-centrality parameter Δ^2 evaluated at c.

(iv) JSE:

$$\left(\mathrm{ADB}^\top\left(\widetilde{\boldsymbol{\beta}}_{1n}\right), \mathrm{ADB}^\top\left(\widehat{\boldsymbol{\beta}}_{2n}^{\mathrm{JS}}\right)\right) = \left(\mathbf{0}^\top, -\boldsymbol{\delta}_2^\top(p_2 - 2)\,\mathbb{E}\left[\chi^{-2}_{p_2+2}(\Delta^2)\right]\right),$$
$$\left(\mathrm{ADL}_2\text{-risk}\left(\widetilde{\boldsymbol{\beta}}_{1n}\right), \mathrm{ADL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_{2n}^{\mathrm{JS}}\right)\right) = \left(\eta^2 p_1, \eta^2\left[p_2 - (p_2 - 2)^2\,\mathbb{E}\left[\chi^{-2}_{p_2+2}(\Delta^2)\right]\right]\right).$$

Hence, the ADL$_2$-risk of James−Stein estimator is given by the simplified form $\eta^2[p_1+p_2-(p_2-2)^2\mathbb{E}[\chi_{p_2}^{-2}(\Delta^2)]]$, where $\mathbb{E}[\chi_{p_2}^{-2v}(\Delta^2)]=\int_0^\infty x^{-2v}dH_{p_2}(x;\Delta^2)$.

(v) PRSE:

$$\begin{aligned}\left(\mathrm{ADB}^\top\left(\widetilde{\boldsymbol{\beta}}_{1n}\right),\mathrm{ADB}^\top\left(\widehat{\boldsymbol{\beta}}_{2n}^{\mathrm{S+}}\right)\right)&=\left(\mathbf{0}^\top,\mathrm{ADB}^\top\left(\widehat{\boldsymbol{\beta}}_{2n}^{\mathrm{JS}}\right)\right.\\&\left.-\boldsymbol{\delta}_2^\top\mathbb{E}\left[\left(1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta^2)\right)I\left(\chi_{p_2+2}^{2}\left(\Delta^2\right)\le p_2-2\right)\right]\right),\\\left(\mathrm{ADL_2\text{-}risk}\left(\widetilde{\boldsymbol{\beta}}_{1n}\right),\mathrm{ADL_2\text{-}risk}\left(\widehat{\boldsymbol{\beta}}_{2n}^{\mathrm{S+}}\right)\right)&=\left(\eta^2p_1,\mathrm{ADL_2\text{-}risk}\left(\widehat{\boldsymbol{\beta}}_{2n}^{\mathrm{S+}}\right)\right),\end{aligned}$$

where

$$\begin{aligned}\mathrm{ADL_2\text{-}risk}\left(\widehat{\boldsymbol{\beta}}_{2n}^{\mathrm{S+}}\right)&=\mathrm{ADL_2\text{-}risk}\left(\widehat{\boldsymbol{\beta}}_{2n}^{\mathrm{JS}}\right)\\&-\eta^2p_2\mathbb{E}\left[\left(1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta^2)\right)^2I\left(\chi_{p_2+2}^{2}\left(\Delta^2\right)<p_2-2\right)\right]\\&+\Delta^2\Big\{2\mathbb{E}\left[\left(1-(p_2-2)\,\chi_{p_2+2}^{-2}(\Delta^2)\right)I\left(\chi_{p_2+2}^{2}\left(\Delta^2\right)<p_2-2\right)\right]\\&+\mathbb{E}\left[\left(1-(p_2-2)\,\chi_{p_2+4}^{-2}(\Delta^2)\right)I\left(\chi_{p_2+4}^{2}\left(\Delta^2\right)<p_2-2\right)\right]\Big\}.\end{aligned}$$

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

$$\eta^2\left[(p_1+p_2)-(p_2-2)^2\mathbb{E}\left[\chi_{p_2}^{-2}\left(\Delta^2\right)\right]-\mathrm{R}^*\right],$$

where

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

(vi) Relative risk:

$$\begin{aligned}\left(\mathrm{ADB}^\top\left(\widetilde{\boldsymbol{\beta}}_{1n}\right),\mathrm{ADB}^\top\left(\widehat{\boldsymbol{\beta}}_{2n}^{\mathrm{RR}}(k)\right)\right)&=\left(\mathbf{0}^\top,-\frac{k}{k+1}\boldsymbol{\delta}_2^\top\right),\\\left(\mathrm{ADL_2\text{-}risk}\left(\widetilde{\boldsymbol{\beta}}_{1n}\right),\mathrm{ADL_2\text{-}risk}\left(\widehat{\boldsymbol{\beta}}_{2n}^{\mathrm{RR}}(k)\right)\right)&=\left(\eta^2p_1,\eta^2\frac{p_2+k^2\Delta^2}{(1+k)^2}\right).\end{aligned}$$

Hence ADL$_2$-risk($\widehat{\boldsymbol{\beta}}_n^{\mathrm{RR}}(k)$) $=\eta^2p_1+\frac{\eta^2}{(k+1)^2}(p_2+k^2\Delta^2)$. Therefore, the optimum ADL$_2$-risk ($\widehat{\boldsymbol{\beta}}_n^{\mathrm{RR}}(k_{\mathrm{opt}})$) $=\eta^2\left(p_1+\frac{p_2\Delta^2}{p_2+\Delta^2}\right)$, since $k_{\mathrm{opt}}=p_2\Delta^{-2}$.

(viii) LASSO:

$$\begin{aligned}\text{ADB}(\hat{\beta}_{jn}^{\text{LASSO}}(\lambda)) &= -\{\Delta_j[\Phi(\lambda-\Delta_j)-\Phi(-\lambda-\Delta_j)] \\ &\qquad -[\varphi(\lambda-\Delta_j)-\varphi(\lambda+\Delta_j)]\} \qquad \text{for } j=1,\ldots,p\end{aligned}$$

and using Donoho & Johnstone (1994),

$$\text{ADL}_2\text{-risk}\left(\hat{\boldsymbol{\beta}}_n^{\text{LASSO}}(\lambda)\right) = \eta^2\rho_{\text{ST}}(\lambda,\boldsymbol{\delta}), \qquad \boldsymbol{\delta}=(\Delta_1,\ldots,\Delta_p)^\top,$$

where

$$\begin{aligned}\rho_{\text{ST}}(\lambda,\boldsymbol{\delta}) &= p_1(1+\lambda^2)-(1+\lambda^2)-\sum_{j=1}^{p_1}[\Phi(\lambda-\Delta_j)-\Phi(-\lambda-\Delta_j)] \\ &\quad+\sum_{j=1}^{p_1}\Delta_j^2[\Phi(\lambda-\Delta_j)-\Phi(-\lambda-\Delta_j)] \\ &\quad-\sum_{j=1}^{p_1}[(\lambda-\Delta_j)\varphi(\lambda+\Delta_j)+(\lambda+\Delta_j)\varphi(\lambda-\Delta_j)]+\Delta^2\end{aligned}$$

with $\Delta^2=\sum_{j=p_1+1}^{p}\Delta_j^2$, and where Φ and φ are the c.d.f and p.d.f of $\mathcal{N}(0,1)$, respectively.

3.1. Lower Bound for ADL$_2$-Risk for LASSO

Consider asymptotic representation of the R-estimators of $\boldsymbol{\beta}$:

$$\hat{\beta}_{jn}=\beta_j+\frac{\eta}{\sqrt{n}}Z_j, \qquad Z_j\sim\mathcal{N}(0,1), \qquad j=1,\ldots,p.$$

We wish to estimate $(\beta_1,\ldots,\beta_p)^\top$ with ADL$_2$-risk $\mathbf{R}(\boldsymbol{\beta}^*,\boldsymbol{\beta})=\mathbb{E}[n\|\boldsymbol{\beta}^*-\boldsymbol{\beta}\|^2]$. We consider a family of diagonal linear projections:

$$\text{T}_{\text{DP}}\left(\hat{\boldsymbol{\beta}}_n^{\text{LASSO}}(\lambda),\boldsymbol{\tau}\right)=\left(\tau_1\hat{\beta}_{1n}^{\text{LASSO}}(\lambda),\ldots,\tau_p\hat{\beta}_{pn}^{\text{LASSO}}(\lambda)\right)^\top$$

for $\tau_1\in(0,1),\ldots,\tau_p\in(0,1)$.

Such estimators either ‘‘keep’’ or ‘‘kill’’ each coordinate $\tilde{\beta}_{jn}$ with subset selection. We incur a risk of η^2 if we use $\tilde{\beta}_{jn}$ and a risk of β_j^2 if we use estimate 0 instead. Hence, the ideal choice of τ_j is $I(|\delta_j|>\eta)$, that is, we keep only those predictors whose true coefficient is larger than the noise level. Denote this risk $\mathbf{R}(\mathbf{T}_{\text{DP}},\boldsymbol{\delta})$ which yields the lower bound of ADL$_2$-risk of LASSO given by

$$\mathbf{R}(\text{T}_{\text{DP}},\boldsymbol{\delta})=\sum_{j=1}^{p}\min\left(\eta^2,\delta_j^2\right).$$

Suppose that there are p_1 predictors whose true value is larger than the noise level η^2, and the remaining p_2 values are estimated as zero. This configuration produces the estimate $(\tilde{\boldsymbol{\beta}}_{1n}^\top,\mathbf{0}^\top)^\top$. Then the lower bound is

$$\mathbf{R}(\text{T}_{DP},\boldsymbol{\delta})=\eta^2(p_1+\Delta^2), \qquad \Delta^2=\Delta_{p_1+1}^2+\cdots+\Delta_p^2=\frac{\boldsymbol{\delta}_2^\top\boldsymbol{\delta}_2}{\eta^2},$$

$$\leq \mathrm{ADL}_2\text{-risk}\left(\widehat{\boldsymbol{\beta}}_n^{\mathrm{LASSO}}(\lambda)\right).$$

Consequently, the asymptotic distributional L_2-risk efficiencies (ADRRE) of the estimators are

$$\mathrm{ADRRE}\left(\widehat{\boldsymbol{\beta}}_n : \widetilde{\boldsymbol{\beta}}_n\right) = \left(1 + \frac{p_2}{p_1}\right)\left(1 + \frac{\Delta^2}{p_1}\right)^{-1}, \tag{13}$$

$$\mathrm{ADRRE}\left(\widehat{\boldsymbol{\beta}}_n^{\mathrm{LASSO}} : \widetilde{\boldsymbol{\beta}}_n\right) = \left(1 + \frac{p_2}{p_1}\right)\left(1 + \frac{\Delta^2}{p_1}\right)^{-1}, \tag{14}$$

$$\mathrm{ADRRE}\left(\widehat{\boldsymbol{\beta}}_n^{\mathrm{RR}} : \widetilde{\boldsymbol{\beta}}_n\right) = \left(1 + \frac{p_2}{p_1}\right)\left(1 + \frac{p_2\Delta^2}{p_1\left(p_2 + \Delta^2\right)}\right)^{-1}, \tag{15}$$

$$\begin{aligned}\mathrm{ADRRE}\left(\widehat{\boldsymbol{\beta}}_n^{\mathrm{PT}} : \widetilde{\boldsymbol{\beta}}_n\right) = \left(1 + \frac{p_2}{p_1}\right)\Bigg\{&1 + \frac{p_2}{p_1}\Big(1 - H_{p_2+2}\Big(\chi^2_{p_2}(\alpha)\,;\Delta^2\Big)\Big)\\ &+\frac{\Delta^2}{p_1}\Big(2H_{p_2+2}\Big(\chi^2_{p_2}(\alpha)\,;\Delta^2\Big) - H_{p_2+4}\Big(\chi^2_{p_2}(\alpha);\Delta^2\Big)\Big)\Bigg\}^{-1},\end{aligned} \tag{16}$$

$$\mathrm{ADRRE}\left(\widehat{\boldsymbol{\beta}}_n^{\mathrm{JS}} : \widetilde{\boldsymbol{\beta}}_n\right) = \left(1 + \frac{p_2}{p_1}\right)\left\{1 + \frac{p_2}{p_1} - \frac{1}{p_1}(p_2-2)^2\mathbb{E}\left[\chi^{-2}_{p_2}\left(\Delta^2\right)\right]\right\}^{-1}, \tag{17}$$

$$\begin{aligned}\mathrm{ADRRE}\left(\widehat{\boldsymbol{\beta}}_n^{\mathrm{S+}} : \widetilde{\boldsymbol{\beta}}_n\right) = \left(1 + \frac{p_2}{p_1}\right)\Bigg\{&1 + \frac{p_2}{p_1} - \frac{1}{p_1}(p_2-2)^2\mathbb{E}\left[\chi^{-2}_{p_2}\left(\Delta^2\right)\right]\\ &-\frac{p_2}{p_1}\mathbb{E}\left[\left(1 - (p_2-2)\,\chi^{-2}_{p_2+2}(\Delta^2)\right)^2 I\left(\chi^2_{p_2+2}\left(\Delta^2\right) < (p_2-2)\right)\right]\\ &+\frac{\Delta^2}{p_1}\Bigg[2\mathbb{E}\left[\left(1 - (p_2-2)\,\chi^{-2}_{p_2+2}(\Delta^2)\right) I\left(\chi^2_{p_2+2}\left(\Delta^2\right) < (p_2-2)\right)\right]\\ &-\mathbb{E}\left[\left(1 - (p_2-2)\,\chi^{-2}_{p_2+4}(\Delta^2)\right)^2 I\left(\chi^2_{p_2+4}\left(\Delta^2\right) < (p_2-2)\right)\right]\Bigg]\Bigg\}^{-1}.\end{aligned} \tag{18}$$

4. GRAPHICAL AND NUMERICAL ASSESSMENT OF THE R-ESTIMATORS

We first note that ADRRE as a function of Δ^2 is decreasing and tends towards unity as $\Delta^2 \to \infty$. Clearly, under $\Delta^2 = 0$, (13)–(15), 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:

$$\mathrm{ADRRE}\left(\widehat{\boldsymbol{\beta}}_n^{\mathrm{PT}} : \widetilde{\boldsymbol{\beta}}_n\right) = \left(1 + \frac{p_2}{p_1}\right)\left\{1 + \frac{p_2}{p_1}\Big(1 - \mathcal{H}_{p_2+2}\Big(\chi^2_{p_2}(\alpha)\,;0\Big)\Big)\right\}^{-1}$$

$$\mathrm{ADRRE}\left(\widehat{\boldsymbol{\beta}}_n^{\mathrm{JS}} : \widetilde{\boldsymbol{\beta}}_n\right) = \left(1 + \frac{p_2}{p_1}\right)\left\{1 + \frac{2}{p_1}\right\}^{-1},$$

TABLE 1: Maximum ADRRE of RE compared to URE (under the null hypothesis).

	p_1			
p	2	3	4	5
10	5.00	3.33	2.50	2.00
20	10.00	6.67	5.00	4.00
30	15.00	10.00	7.50	6.00
40	20.00	13.33	10.00	8.00
60	30.00	20.00	15.00	12.00
128	64.00	42.67	32.00	25.60

$$\mathrm{ADRRE}\left(\hat{\boldsymbol{\beta}}_n^{\mathrm{S+}} : \tilde{\boldsymbol{\beta}}_n\right) = \left(1 + \frac{p_2}{p_1}\right)$$

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

Table 1 presents the ADRRE values for RE, LASSO, and RR R-estimators when $\Delta^2 = 0$ for $p_1 = 2, 3, 4, 5$ and $p = 10, 20, \ldots, 60, 128$.

In Table 2, ADRRE values for $p = 20$ with $(p_1, p_2) = (5, 15)$ is reported. In the supplementary file, a table for $(p_1, p_2) = (7, 33)$ is included.

Table 2 presents the ADRRE values of the six R-estimators for 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 ADRRE increases for Δ^2 (Figure 1). However, if p_2 is fixed and p_1 varies, then the ADRRE 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 the tables in the supplementary file for more information.

Furthermore, we found the order of ADRRE of URE, JSE, PRSE and RR as

$$\mathrm{ADRRE}\left(\hat{\boldsymbol{\beta}}_n^{\mathrm{RR}} : \tilde{\boldsymbol{\beta}}_n\right) \geq \mathrm{ADRRE}\left(\hat{\boldsymbol{\beta}}_n^{\mathrm{S+}} : \tilde{\boldsymbol{\beta}}_n\right) \geq \mathrm{ADRRE}\left(\hat{\boldsymbol{\beta}}_n^{\mathrm{JS}} : \tilde{\boldsymbol{\beta}}_n\right) \geq 1$$

uniformly, and that of RR and LASSO as

$$\mathrm{ADRRE}\left(\hat{\boldsymbol{\beta}}_n^{\mathrm{LASSO}} : \tilde{\boldsymbol{\beta}}_n\right) = \mathrm{ADRRE}\left(\hat{\boldsymbol{\beta}}_n^{\mathrm{RR}} : \tilde{\boldsymbol{\beta}}_n\right) \quad \text{in} \quad (0, 1),$$

and

$$\mathrm{ADRRE}\left(\hat{\boldsymbol{\beta}}_n^{\mathrm{LASSO}} : \tilde{\boldsymbol{\beta}}_n\right) < \mathrm{ADRRE}\left(\hat{\boldsymbol{\beta}}_n^{\mathrm{RR}} : \tilde{\boldsymbol{\beta}}_n\right) \quad \text{in} \quad [1, \infty).$$

Finally, PRSE always outperforms JSE.

We mention a few features of the ADRRE expressions, using the LSE method and the related finite sample efficiency, are the same as the ones we have here, see Saleh et al. (2017). Further, we considered an ADRRE expression for changing c.d.f. F, but the ADRRE values do not

TABLE 2: ADRRE for the estimators in case of $p_1 = 5$ and $p_2 = 15$.

			PTE	PTE	PTE			
Δ^2	URE	RE/LASSO	$\alpha = 0.15$	$\alpha = 0.2$	$\alpha = 0.25$	JSE	PRSE	RRE
0	1.00	4.00	2.31	2.07	1.89	2.86	3.22	4.00
0.1	1.00	3.92	2.26	2.03	1.86	2.82	3.18	3.92
0.5	1.00	3.64	2.10	1.90	1.75	2.70	3.01	3.65
1	1.00	3.33	1.93	1.76	1.63	2.56	2.82	3.37
2	1.00	2.86	1.67	1.55	1.45	2.34	2.53	2.96
3	1.00	2.50	1.49	1.40	1.33	2.17	2.32	2.67
5	1.00	2.00	1.26	1.21	1.17	1.94	2.02	2.29
7	1.00	1.67	1.13	1.10	1.08	1.78	1.83	2.05
10	1.00	1.33	1.03	1.02	1.01	1.62	1.64	1.82
15	1.00	1.00	0.97	0.98	0.98	1.46	1.46	1.60
20	1.00	0.80	0.97	0.98	0.99	1.36	1.36	1.47
30	1.00	0.57	0.99	1.00	1.00	1.26	1.26	1.33
50	1.00	0.36	1.00	1.00	1.00	1.16	1.16	1.21
100	1.00	0.19	1.00	1.00	1.00	1.05	1.05	1.11

change. We studied the high-dimensional problem in depth and obtained similar expressions for ADRRE. The results of this article will be reported in a separate paper as it involves substantial additional work in deriving the ADRRE expressions.

5. APPLICATION ON REAL DATA

The following example comes from McDonald & Schwing (1973). They investigated the dependence of mortality on some social and economic characteristics. More precisely, the response variable is the total age-adjusted mortality rate per 100,000 (‘‘mortality’’) for Standard Metropolitan Statistical Areas in 1959–1961. There are 15 regressors in the dataset including those that describe weather conditions such as average annual precipitation, pollution level such as relative hydrocarbon pollution potential, and demographics such as percentage of population that are old.

First, we created principal components from the original regressors to ensure an orthonormal matrix design. Explained variance by the first 10 principal components is displayed in Figure 2. Note that the overall variance is equal to the number of original regressors $p = 15$.

Then, we modeled the dependency of ‘‘mortality’’ on $p = 15$ principal components. We computed all the estimates mentioned above and used a bootstrap method to estimate their mean squared errors and relative efficiencies.

We considered $p_2 = 9$ coefficients to be small or zero (corresponding to principal components 2, 4, 5, 8, 10, 11, 13, 14, 15). We got these results thanks to the LASSO R-estimate method, which performs the model selection. The selected model was later confirmed by the tests. The model does not simply consist of the last none components (as one could expect), because they are ordered according to the content of the information about regressors $\boldsymbol{x}_i$ and this ordering does not reflect their relation to the response variable.

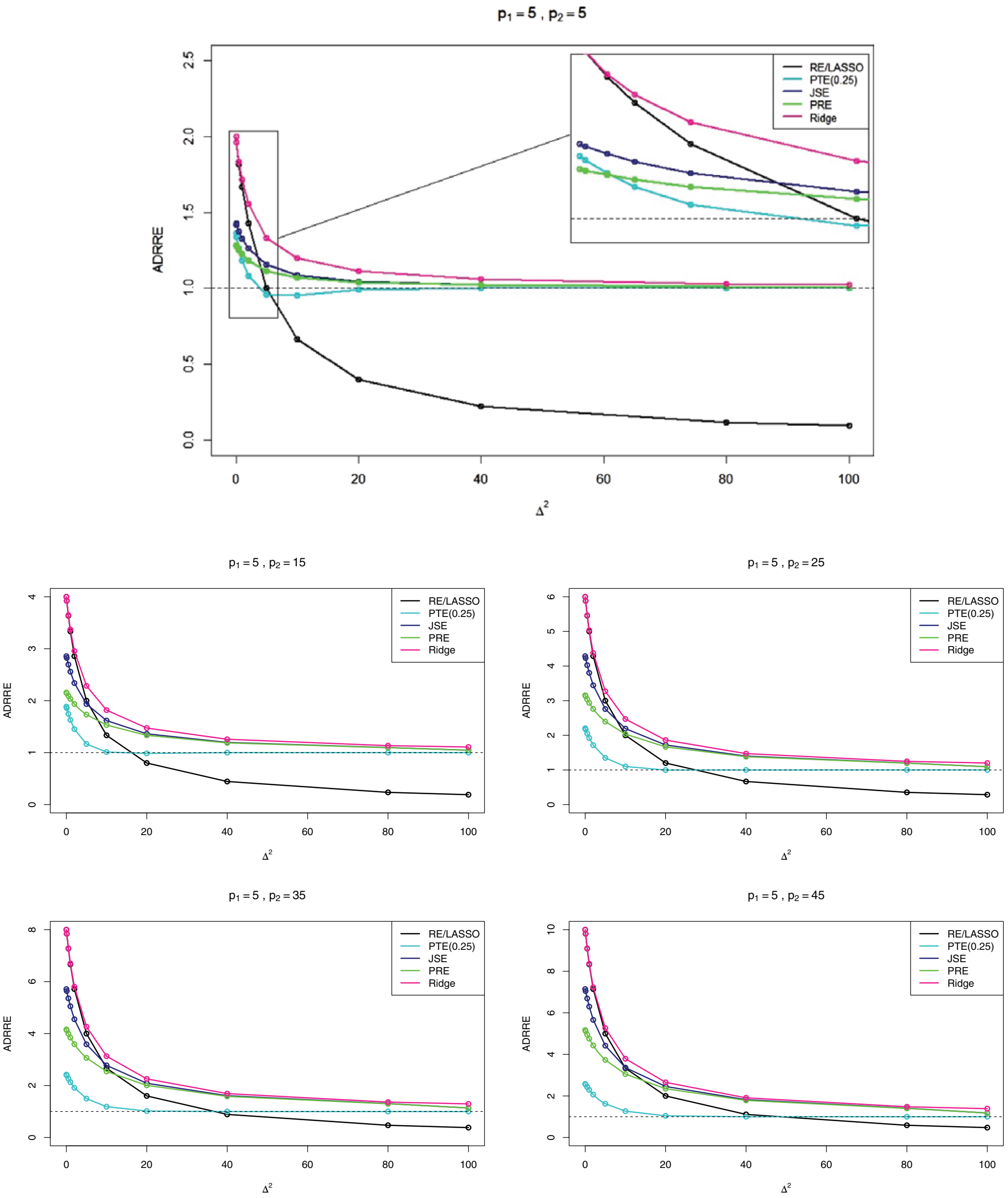

FIGURE 1: ADRRE of estimates of a function of Δ^2 for $p_1 = 5$ and different p_2.

Estimates of mean squared error are summarized in Table 3. Estimates of relative efficiencies with respect to the unrestricted R-estimate are summarized in Table 4. One may see that the results correspond with the formulas derived in previous sections.

6. DISCUSSION

Comparison of ridge regression estimator with Stein-type estimator began with the work of Draper & Van Nostrand (1979). The LASSO, with its ability to simultaneously estimate and select variables, brought in several penalty estimators. For comparison, papers appeared based on simulation studies without mathematical backup.

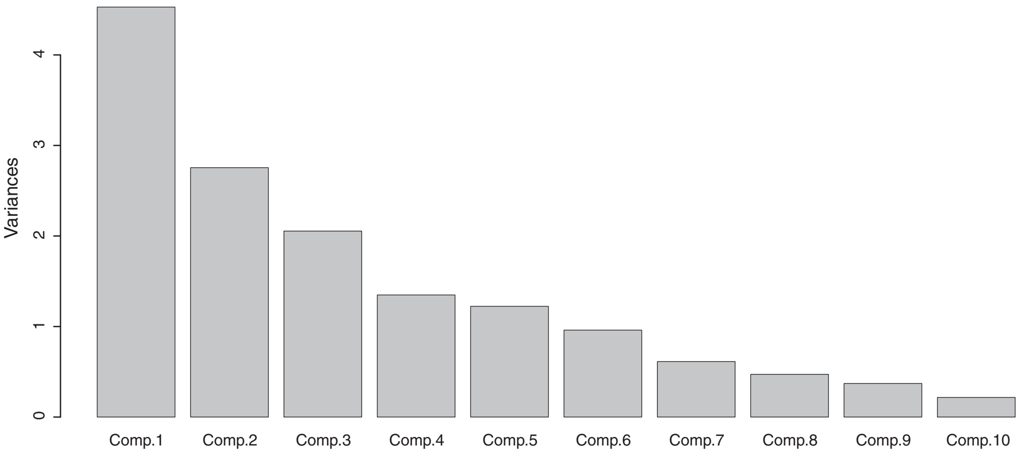

FIGURE 2: Explained variance by the first 10 principal components.

TABLE 3: Estimates of mean squared error of the R-estimates.

URE	RE/LASSO	PTE(0.15)	PTE(0.20)	PTE(0.25)	JSE	PRSE	RR
139	29.30	29.30	29.30	29.30	98.70	98.70	29.30

TABLE 4: Estimates of relative efficiencies of the R-estimates with respect to unrestricted R-estimate.

URE	RE/LASSO	PTE(0.15)	PTE(0.20)	PTE(0.25)	JSE	PRSE	RR
1.00	4.74	4.74	4.74	4.74	1.41	1.41	4.74

In this paper, we consider rank-based LASSO, ridge, preliminary test and Stein-type estimators and compare these estimators using ADL$_2$-risks.

We have demonstrated that: (i) Ridge outperforms LASSO uniformly. LASSO and Ridge are risk-equivalent in [0, 1] and Ridge outperforms LASSO uniformly when Δ^2 is in $(1, \infty)$. (ii) Neither LASSO nor PTE, JSE and positive-rule Stein-type estimator dominate the other uniformly. LASSO dominates others for $\Delta^2 \in [0, p_2]$. (iii) Ridge dominates preliminary test, James–Stein and positive-rule Stein-type estimators, uniformly. (iv) LASSO dominates the R-estimator uniformly for $\Delta^2 \in [0, p_2]$, outside this interval the R-estimator dominates.

Some of Hansen's (2016) remarks hold except that he could not find the interval described in (i)–(iv).

ACKNOWLEDGEMENTS

The authors would like to thank the editor, associate editor and the two referees for their valuable comments that helped significantly to improve the article.

REFERENCES

Belloni, A. & Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, 19, 521–547.

Breiman, L. (1996). Heuristics of instability and stabilization in model selection. *The Annals of Statistics*, 24, 2350–2383.

Donoho, D. L. & Johnstone, I. M. (1994). Minimax estimation via wavelet shrinkage. *The Annals of Statistics*, 26, 879–921.

Draper, N. R. & Van Nostrand, R. C. (1979). Ridge regression and James–Stein estimation: Review and comments. *Technometrics*, 21, 451–466.

Fan, J. & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96, 1348–1360.

Frank, L. E. & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, 35, 109–135.

Hansen, B. E. (2016). The risk of James–Stein and LASSO shrinkage. *Econometric Reviews*, 35(8–10), 1456–1470.

Hoerl, E. & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12, 55–67.

Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *The Annals of Mathematical Statistics*, 43, 1449–1458.

James, W. & Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Contributions to the Theory of Statistics*, Vol. 1. University of California Press, Berkeley, CA, pp. 361–379.

Jurečková, J. (1971). Nonparametric estimate of regression coefficients. *The Annals of Mathematical Statistics*, 42, 1328–1338.

Jurečková, J. & Sen, P. K. (1996). *Robust Statistical Procedures: Asymptotic and Interrelations*. John Wiley & Sons, New York.

McDonald, G. C. & Schwing, R. C. (1973). Instabilities of regression estimates relating air pollution to mortality. *Technometrics*, 15, 463–481.

Saleh, A. K. M. E. (2006). *Theory of Preliminary Test and Stein-Type Estimators with Applications*. John Wiley & Sons, New York.

Saleh, A. K. M. E. & Sen, P. K. (1978). Non-parametric estimation of location parameter after a preliminary test on regression. *Annals of Statistics*, 6, 154–168.

Saleh, A. K. M. E., Arashi, M., Norouzirad, M., & Kibria, B. M. G. (2017). On shrinkage and selection: ANOVA MODEL. *Journal of Statistical Research*, 51, 165–191.

Sen, P. K. & Saleh, A. K. M. E. (1987). On preliminary test and shrinkage M-estimation in linear models. *The Annals of Statistics*, 15, 1580–1592.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, 58, 267–288.

Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics Doklady*, 4, 1035–1038.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101, 1418–1429.

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B*, 67, 301–320.

Received 30 March 2017
Accepted 24 June 2018