SORT 42 (1) January-June 2018, 45-58 DOI: 10.2436/20.8080.02.68

# Preliminary test and Stein-type shrinkage LASSO-based estimators

M. Norouzirad and M. Arashi*

**Abstract**

Suppose the regression vector-parameter is subjected to lie in a subspace hypothesis in a linear regression model. In situations where the use of least absolute and shrinkage selection operator (LASSO) is desired, we propose a restricted LASSO estimator. To improve its performance, LASSO-type shrinkage estimators are also developed and their asymptotic performance is studied. For numerical analysis, we used relative efficiency and mean prediction error to compare the estimators which resulted in the shrinkage estimators to have better performance compared to the LASSO.

*MSC:* 62F15, 62H05.

*Keywords:* Double shrinking, LASSO, preliminary test LASSO, restricted LASSO, Stein-type shrinkage LASSO.

## 1. Introduction

Consider the linear regression model with form

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

where $\boldsymbol{Y} = (y_1, \dots, y_n)^{\mathsf{T}}$ is a vector of responses, $\boldsymbol{X}$ is an $n \times p$ non-stochastic design matrix, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^{\mathsf{T}}$ is an unknown vector of parameters, $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^{\mathsf{T}}$ is the vector of random errors, with $E(\boldsymbol{\epsilon}_n) = \boldsymbol{0}$ and $E(\boldsymbol{\epsilon}_n \boldsymbol{\epsilon}_n^{\mathsf{T}}) = \sigma^2 \boldsymbol{I}_n (\sigma^2 < \infty)$, $\boldsymbol{I}_n$ the identity matrix of order $n$.

In general, the main goal of the linear regression model (1) is the estimation of parameters and prediction of response for a given design matrix. The estimation problem is usually solved through the ordinary least squares (OLS) method. Provided $\boldsymbol{C}_n = \boldsymbol{X}^{\mathsf{T}}\boldsymbol{X}$ is well-conditioned, we use the OLS estimator given by $\tilde{\boldsymbol{\beta}}_n = \boldsymbol{C}_n^{-1}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{Y}$. The corresponding estimator of $\sigma^2$ is $s_e^2 = (\boldsymbol{Y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}_n)^{\mathsf{T}}(\boldsymbol{Y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}_n)/m$, $m = n - p$.

*Corresponding author:* m_arashi_stat@yahoo.com

* Department of Statistics, Faculty of Mathematical Sciences. Shahrood University of Technology, Shahrood, Iran

Received: February 2017

Accepted: February 2018

Assume the following regularity conditions:

**A1:** $\max_{1\leq i\leq n}\boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{C}_n^{-1}\boldsymbol{x}_i \to 0$ as $n \to \infty$ where $\boldsymbol{x}_i^{\mathrm{T}}$ is the $i$th row of design matrix $\boldsymbol{X}$.

**A2:** $\lim_{n\to\infty} n^{-1}\boldsymbol{C}_n = \boldsymbol{C}$, where $\boldsymbol{C}$ is finite and positive-definite matrix.
Then, asymptotically $\tilde{\boldsymbol{\beta}}_n \sim \mathcal{N}_p(\boldsymbol{\beta}, \sigma^2\boldsymbol{C}^{-1})$, which is independent of $(ms_e^2)/\sigma^2 \sim \chi_m^2$ (asymptotically).

Now, suppose that we are provided with some prior information about the whole or subset of covariates. This prior information can be utilized to improve the overall estimation of the regression coefficients using shrinkage estimation (Ahmed and Raheem, 2012).

There are many notable studies incorporating prior information, in the form of restrictions, to improve estimation in the sense that the restricted and shrinkage estimators have lesser risk and prediction error values.

Saleh (2006) gives extensive overviews on preliminary test and shrinkage estimators using the OLS, ridge and maximum likelihood (ML) estimators as starting points. Fallahpour et al. (2012) developed shrinkage estimators by using the weighted semi-parametric OLS estimator. Hossain and Ahmed (2014) start by maximum partial likelihood estimator and propose shrinkage and positive shrinkage estimators, while Roozbeh (2015, 2016) develops shrinkage estimators in a ridge regression. Other related studies include Hossain et al. (2015), Hossain and Howlader (2016), Hossain et al. (2016), Yuzbasi and Ahmed (2016) and Yuzbasi et al. (2017), to mention a few.

However, in this study, we have different concerns. As a prelude, Tibshirani (1996) proposed a new method for variable selection that produces an accurate, stable, and parsimonious model, called least absolute shrinkage and selection operator (LASSO) that is obtained by

$$\hat{\boldsymbol{\beta}}_n^{\mathrm{L}} = \arg\min_{\boldsymbol{\beta}}\left\{\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda_n\sum_{j=1}^{p}|\beta_j|\right\}, \qquad \lambda_n \geq 0, \qquad (2)$$

where $\lambda_n$ is the tuning parameter, controlling the level of sparsity in $\hat{\boldsymbol{\beta}}^{\mathrm{L}}$.

Now, the questions are as follows:

1. How can we build the theory if we start with the LASSO instead of using the OLS/ML estimator?
2. What will the form of shrinkage estimators be under restriction, when LASSO is used as the starting point?
3. Is it possible to derive asymptotic properties of the preliminary test and shrinkage estimators based on the LASSO?

In this paper, we cover the above issues. Hence, we organize the paper as follows: In Section 2, the restricted LASSO estimator is defined for inference under restriction and the concept of double shrinking is introduced (covering questions 1 and 2 above). Section 3 contains the asymptotic distributions of the proposed estimators (covering question 3 above). An extensive numerical study is carried out in Section 4 and we conclude our study in Section 5.

## 2. Restricted LASSO and double shrinking

The LASSO estimator has been denoted as $\hat{\boldsymbol{\beta}}_n^{\mathrm{L}}$ and termed as unrestricted LASSO estimator (ULE). Now, suppose that some non-sample information (a priori restriction on the parameters) about the covariates is available. A set of $q$ linear restrictions on the vector $\boldsymbol{\beta}$ can be written as $\boldsymbol{H\beta} = \boldsymbol{h}$. Or, we can suppose that our model is subjected to lie in the linear subspace restriction,

$$\boldsymbol{H\beta} = \boldsymbol{h}, \tag{3}$$

where $\boldsymbol{H}$ is a $q \times p$ $(q \leq p)$ matrix of known elements, and $\boldsymbol{h}$ is a $q$ vector of known components. The rank of $\boldsymbol{H}$ is $q$, which implies that the restrictions are linearly independent. This restriction may be (i) a fact known from theoretical or experimental considerations, (ii) a hypothesis that may have to be tested or (iii) an artificially imposed condition to reduce or eliminate redundancy in the description of model (Sengupta and Jammalamadaka, 2003).

Our proposal is to consider the following estimator as the restricted LASSO estimator (RLE),

$$\hat{\boldsymbol{\beta}}_n^{\mathrm{RL}} = \hat{\boldsymbol{\beta}}_n^{\mathrm{L}} - \boldsymbol{C}_n^{-1}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{C}_n^{-1}\boldsymbol{H}^{\mathrm{T}})^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}}_n^{\mathrm{L}} - \boldsymbol{h}). \tag{4}$$

The above closed form RLE cannot be achieved via routine optimization techniques. Indeed, we proposed it by the analogy of OLS estimator of $\boldsymbol{\beta}$ subject to the restriction $\boldsymbol{H\beta} = \boldsymbol{h}$.

When (3) is satisfied, $\hat{\boldsymbol{\beta}}_n^{\mathrm{RL}}$ has smaller asymptotic risk than $\hat{\boldsymbol{\beta}}_n^{\mathrm{L}}$. However, for $\boldsymbol{H\beta} \neq \boldsymbol{h}$, $\hat{\boldsymbol{\beta}}_n^{\mathrm{RL}}$ may be biased and inconsistent in many cases. Now, how can we decide on ULE or RLE, since we do not know whether the restriction holds? To solve this, it is plausible to follow Fisher's recipe and define the preliminary test LASSO estimator (PTLE) by taking $\hat{\boldsymbol{\beta}}_n^{\mathrm{L}}$ or $\hat{\boldsymbol{\beta}}_n^{\mathrm{RL}}$ according to the acceptance or rejection of the null hypothesis, $\mathscr{H}_o : \boldsymbol{H\beta} = \boldsymbol{h}$.

This estimator will have the form

$$\hat{\boldsymbol{\beta}}_n^{\mathrm{PTL}} = \hat{\boldsymbol{\beta}}_n^{\mathrm{L}} - (\hat{\boldsymbol{\beta}}_n^{\mathrm{L}} - \hat{\boldsymbol{\beta}}_n^{\mathrm{RL}})I(\mathscr{L}_n \leq \mathscr{L}_{n,\alpha}), \tag{5}$$

where $\mathscr{L}_{n,\alpha}$ is the upper $\alpha$-level critical value of the exact distribution of the test statistic $\mathscr{L}_n$ under $\mathscr{H}_o$. We will propose a relevant test statistic later in Section 3.

The PTLE is highly dependent on the level of significance $\alpha$ and has discrete nature which is simplified to one of the extremes $\hat{\boldsymbol{\beta}}_n^{\rm L}$ or $\hat{\boldsymbol{\beta}}_n^{\rm RL}$ according to the output of the test. In this respect, making use of a continuous and $\alpha$-free estimator may make more sense. Now, we propose a double shrinking idea which reflects a relevant estimator. It is well-known that the LASSO estimator shrinks coefficients toward the origin. However, when the restriction $\boldsymbol{H\beta}=\boldsymbol{h}$ is subjected to the model, it is of major importance that the estimator be shrunken toward the restricted one as well. Hence, there must be shrinking toward two directions or double shrinking concept, say. Consequently, we combine the idea of James and Stein (1961) shrinkage and LASSO to propose the following Stein-type shrinkage LASSO estimator (SSLE)

$$\hat{\boldsymbol{\beta}}_n^{\rm SSL}=\hat{\boldsymbol{\beta}}_n^{\rm L}-k_n(\hat{\boldsymbol{\beta}}_n^{\rm L}-\hat{\boldsymbol{\beta}}_n^{\rm RL})\mathscr{L}_n^{-1}, \qquad k_n=\frac{m(q-2)}{m+2}, \tag{6}$$

where $k_n$ is the shrinkage constant.

The estimator $\hat{\boldsymbol{\beta}}_n^{\rm SSL}$ may go past the estimator $\hat{\boldsymbol{\beta}}_n^{\rm RL}$. So, we define the positive-rule Stein-type shrinkage LASSO estimator (PRSSLE) given by

$$\begin{aligned}\hat{\boldsymbol{\beta}}_n^{\rm PRSSL}&=\hat{\boldsymbol{\beta}}_n^{\rm RL}+(1-k_n\mathscr{L}_n^{-1})I(\mathscr{L}_n>k_n)(\hat{\boldsymbol{\beta}}_n^{L}-\hat{\boldsymbol{\beta}}_n^{\rm RL}),\\ &=\hat{\boldsymbol{\beta}}_n^{\rm SSL}-(1-k_n\mathscr{L}_n^{-1})I(\mathscr{L}_n\le k_n)(\hat{\boldsymbol{\beta}}_n^{\rm L}-\hat{\boldsymbol{\beta}}_n^{\rm RL}).\end{aligned} \tag{7}$$

We note that, as the test based on $\mathscr{L}_n$ is consistent against fixed $\boldsymbol{\beta}$ such that $\boldsymbol{H\beta}\neq\boldsymbol{h}$, the PTLE, SSLE and PRSSLE are asymptotically equivalent to the ULE for fixed alternative. Hence, we will investigate the asymptotic risks under local alternatives and compare the performance of the estimators.

## 3. Some asymptotic results

For the purpose of this section, we consider the class of local alternatives, $\mathscr{K}_{(n)}$ defined by

$$\mathscr{K}_{(n)}:\boldsymbol{H\beta}=\boldsymbol{h}+n^{-\frac{1}{2}}\boldsymbol{\xi}, \quad \boldsymbol{\xi}=(\xi_1,\ldots,\xi_q)^{\top}\in\mathbb{R}^q.$$

Let $\hat{\boldsymbol{\beta}}_n^*$ be any estimator of $\boldsymbol{\beta}$. We define the asymptotic cumulative distribution function (c.d.f.) of $\hat{\boldsymbol{\beta}}_n^*$, under $\mathscr{K}_{(n)}$, as

$$G_p(x)=\lim_{n\to\infty}P_{\mathscr{K}_{(n)}}\left\{\sqrt{n}s_e^{-1}(\hat{\boldsymbol{\beta}}_n^*-\boldsymbol{\beta})\le\boldsymbol{x}\right\}.$$

If the asymptotic c.d.f. exists, then the asymptotic distributional bias (ADB) and quadratic bias (ADQB) are given by

$$\boldsymbol{b}(\hat{\boldsymbol{\beta}}_n^*) = \lim_{n\to\infty} E\left[\sqrt{n}(\hat{\boldsymbol{\beta}}_n^* - \boldsymbol{\beta})\right] = \int \boldsymbol{x} dG_p(\boldsymbol{x}),$$
$$B(\hat{\boldsymbol{\beta}}_n^*) = \sigma^{-2}[\boldsymbol{b}(\hat{\boldsymbol{\beta}}_n^*)]^{\mathrm{T}}\boldsymbol{C}[\boldsymbol{b}(\hat{\boldsymbol{\beta}}_n^*)],$$

respectively, where $\sigma^2\boldsymbol{C}^{-1}$ is the mean squared error (MSE)-matrix of $\tilde{\boldsymbol{\beta}}_n$ as $n \to \infty$. Defining

$$\boldsymbol{M}(\hat{\boldsymbol{\beta}}_n^*) = \int \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} dG_p(\boldsymbol{x}) = \lim_{n\to\infty} E\left[n(\hat{\boldsymbol{\beta}}_n^* - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_n^* - \boldsymbol{\beta})^{\mathrm{T}}\right],$$

as the asymptotic distributional MSE (ADMSE), we have the weighted risk of $\hat{\boldsymbol{\beta}}_n^*$ given by

$$\mathrm{R}(\hat{\boldsymbol{\beta}}_n^*) = \mathrm{tr}[\boldsymbol{M}(\hat{\boldsymbol{\beta}}_n^*)] = \lim_{n\to\infty} E[n(\hat{\boldsymbol{\beta}}_n^* - \boldsymbol{\beta})^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_n^* - \boldsymbol{\beta})]$$

as the asymptotic distributional quadratic risk (ADQR).

Suppose the LASSO is weakly consistent, i.e., $\lambda_n = o(n^{1/2})$. Up to this point, we implemented a test statistic based on the OLS estimator, however, constructing a test based on the LASSO estimator will give the same asymptotic behaviour under weak consistency. A test statistic based on the ULE will have form

$$\mathscr{L}_n = \frac{(\boldsymbol{H}\hat{\boldsymbol{\beta}}_n^{\mathrm{L}} - \boldsymbol{h})^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{C}_n^{-1}\boldsymbol{H}^{\mathrm{T}})^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}}_n^{\mathrm{L}} - \boldsymbol{h})}{s_L^2}, \tag{8}$$

where

$$s_L^2 = \frac{1}{m}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_n^{\mathrm{L}})^{\mathrm{T}}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_n^{\mathrm{L}}) \tag{9}$$

Using Theorem 2 of Knight and Fu (2000), Theorem 7.8.2.3 of Saleh (2006), and $\sqrt{n}$-consistency, we have the following important result.

**Theorem 1** *Under the assumptions of Theorem 2 and $\lambda_n = o(n^{1/2})$, we have*

**(i)** $W_n^{(1)} = \sqrt{n}(\hat{\boldsymbol{\beta}}_n^{\mathrm{L}} - \boldsymbol{\beta}) \overset{\mathscr{D}}{=} W = \sqrt{n}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$.

**(ii)** $W_n^{(2)} = \sqrt{n}(\hat{\boldsymbol{\beta}}_n^{\mathrm{RL}} - \boldsymbol{\beta}) \xrightarrow{\mathscr{D}} \mathscr{N}_p(-\boldsymbol{\delta}, \sigma^2\boldsymbol{A})$ *where* $\boldsymbol{\delta} = \boldsymbol{C}^{-1}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{C}^{-1}\boldsymbol{H}^{\mathrm{T}})^{-1}\boldsymbol{\xi}$ *and* $\boldsymbol{A} = \boldsymbol{C}^{-1} - \boldsymbol{C}^{-1}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{C}^{-1}\boldsymbol{H}^{\mathrm{T}})^{-1}\boldsymbol{H}\boldsymbol{C}^{-1}$.

**(iii)** $W_n^{(3)} = \sqrt{n}(\hat{\boldsymbol{\beta}}_n^{\mathrm{L}} - \hat{\boldsymbol{\beta}}_n^{\mathrm{RL}}) \xrightarrow{\mathscr{D}} \mathscr{N}_p(\delta, \sigma^2(C^{-1} - A))$.

**(iv)** $W_n^{(4)} = \boldsymbol{H}\hat{\boldsymbol{\beta}}_n^{\mathrm{L}} - \boldsymbol{h} \xrightarrow{\mathscr{D}} \mathscr{N}_q(\boldsymbol{H}\boldsymbol{\beta} - \boldsymbol{h}, \sigma^2(\boldsymbol{H}\boldsymbol{C}^{-1}\boldsymbol{H}^{\mathrm{T}}))$.

**(v)** $\begin{bmatrix} W_n^{(1)} \\ W_n^{(3)} \end{bmatrix} \xrightarrow{\mathscr{D}} \mathscr{N}_{2p}\left(\begin{bmatrix} 0 \\ \boldsymbol{\delta} \end{bmatrix}, \sigma^2 \begin{bmatrix} \boldsymbol{C}^{-1} & \boldsymbol{C}^{-1} - \boldsymbol{A} \\ \boldsymbol{C}^{-1} - \boldsymbol{A} & \boldsymbol{C}^{-1} - \boldsymbol{A} \end{bmatrix}\right)$.

**(vi)** $\begin{bmatrix} W_n^{(2)} \\ W_n^{(3)} \end{bmatrix} \xrightarrow{\mathscr{D}} \mathscr{N}_{2p}\left( \begin{bmatrix} \boldsymbol{\delta} \\ -\boldsymbol{\delta} \end{bmatrix}, \sigma^2 \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C}^{-1} - \boldsymbol{A} \end{bmatrix} \right).$

**(vii)** $\begin{bmatrix} W_n^{(1)} \\ W_n^{(4)} \end{bmatrix} \xrightarrow{\mathscr{D}} \mathscr{N}_{p+q}\left( \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{H\beta} - \boldsymbol{h} \end{bmatrix}, \sigma^2 \begin{bmatrix} \boldsymbol{C}^{-1} & \boldsymbol{C}^{-1}\boldsymbol{H}^{\top} \\ \boldsymbol{HC}^{-1} & \boldsymbol{HC}^{-1}\boldsymbol{H}^{\top} \end{bmatrix} \right).$

**(viii)** $\sqrt{n}(\hat{\boldsymbol{\beta}}_n^{\mathrm{SSL}} - \boldsymbol{\beta}) \overset{\mathscr{D}}{=} \boldsymbol{W} - k\left\{ \dfrac{\boldsymbol{C}^{-1}\boldsymbol{H}^{\top}(\boldsymbol{HC}^{-1}\boldsymbol{H}^{\top})^{-1}(\boldsymbol{HW}+\boldsymbol{\xi})}{\sigma^{-2}(\boldsymbol{HW}+\boldsymbol{\xi})^{T}(\boldsymbol{HC}^{-1}\boldsymbol{H}^{\top})^{-1}(\boldsymbol{HW}+\boldsymbol{\xi})} \right\}.$

**(ix)**
$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\beta}}_n^{\mathrm{PRSSL}} - \boldsymbol{\beta}) \overset{\mathscr{D}}{=} \ & \boldsymbol{W} - k\left\{ \frac{\boldsymbol{C}^{-1}\boldsymbol{H}^{\top}(\boldsymbol{HC}^{-1}\boldsymbol{H}^{\top})^{-1}(\boldsymbol{HW}+\boldsymbol{\xi})}{\sigma^{-2}(\boldsymbol{HW}+\boldsymbol{\xi})^{\top}(\boldsymbol{HC}^{-1}\boldsymbol{H}^{\top})^{-1}(\boldsymbol{HW}+\boldsymbol{\xi})} \right\} \\ & + \boldsymbol{C}^{-1}\boldsymbol{H}^{\top}(\boldsymbol{HC}^{-1}\boldsymbol{H}^{\top})^{-1}(\boldsymbol{HW}+\boldsymbol{\xi}) \\ & \times \left\{ 1 - \frac{k}{\sigma^{-2}(\boldsymbol{HW}+\boldsymbol{\xi})^{\top}(\boldsymbol{HC}^{-1}\boldsymbol{H}^{\top})^{-1}(\boldsymbol{HW}+\boldsymbol{\xi})} \right\} \\ & \times I(\mathscr{L} < k). \end{aligned}$$

*where* $\boldsymbol{W} \xrightarrow{\mathscr{D}} \mathscr{N}_p(\boldsymbol{0}, \sigma^2\boldsymbol{C}^{-1})$.

Based on the part (a) of Theorem 1, the distribution of the test statistics is obtained by Theorem 2.

**Theorem 2** *Under the foregoing regularity conditions and local alternatives $\mathscr{K}_{(n)}$, if the LASSO satisfies the weakly consistent condition, i.e., $\lambda_n = o(n^{1/2})$, the test statistics $\mathscr{L}_n$ defined in Eq. 8 converges in distribution to $\mathscr{L}$, which has the non central chi-square distribution with q degrees of freedom, non centrally parameter $\Delta^2 = \sigma^{-2}\boldsymbol{\xi}^{\top}(\boldsymbol{HC}^{-1}\boldsymbol{H}^{\top})^{-1}\boldsymbol{\xi} = \sigma^{-2}\boldsymbol{\delta}^{\top}\boldsymbol{C\delta}$ where $\boldsymbol{\delta} = \boldsymbol{C}^{-1}\boldsymbol{H}^{\top}(\boldsymbol{HC}^{-1}\boldsymbol{H}^{\top})^{-1}\boldsymbol{\xi}$, and*

$$\mathscr{L} = \frac{(\boldsymbol{HW}+\boldsymbol{\xi})^{\top}(\boldsymbol{HC}^{-1}\boldsymbol{H}^{\top})^{-1}(\boldsymbol{HW}+\boldsymbol{\xi})}{\sigma^2}.$$

*Proof.* Rewrite the numerator of test statistics in Eq. (8) as

$$\begin{aligned} & \left(\boldsymbol{H}\left(\sqrt{n}(\hat{\boldsymbol{\beta}}_n^{\mathrm{L}} - \boldsymbol{\beta})\right) + \sqrt{n}(\boldsymbol{H\beta} - \boldsymbol{h})\right)^{\top} \left(\boldsymbol{H}(n\boldsymbol{C}_n^{-1})\boldsymbol{H}^{\top}\right)^{-1} \\ & \times \left(\boldsymbol{H}\left(\sqrt{n}(\hat{\boldsymbol{\beta}}_n^{\mathrm{L}} - \boldsymbol{\beta})\right) + \sqrt{n}(\boldsymbol{H\beta} - \boldsymbol{h})\right) \end{aligned} \tag{10}$$

Using part (i) of Theorem 1, $\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n^{\mathrm{L}} - \boldsymbol{\beta}\right)$ has the same asymptotic distribution as $\boldsymbol{W}$. Hence, under $\mathscr{K}_{(n)}$ and the regularity condition **A2**, Eq. (10) has the same distribution as

$$(\boldsymbol{HW}+\boldsymbol{\xi})^{\top}\left(\boldsymbol{HC}^{-1}\boldsymbol{H}^{\top}\right)^{-1}(\boldsymbol{HW}+\boldsymbol{\xi}) \tag{11}$$

On the other hand, by (i) of Theorem 1, it is obvious that $s_L^2 \to \sigma^2$. Using this fact together with Eq. (11), the result follows by Slutsky's theorem. ■

The results of Theorems 1 and 2 can be used to derive ADB, ADQB, and ADQR.

To verify the consistency of the estimators, we have the following theorem and subsequent remarks.

**Theorem 3** *Under the foregoing regularity conditions and local alternatives $\mathscr{K}_{(n)}$, we have the following as $n \to \infty$,*

**(i)** $\hat{\boldsymbol{\beta}}_n^{\mathrm{RL}} \xrightarrow{\mathscr{P}} \operatorname{argmin}(Z) - \boldsymbol{C}^{-1}\boldsymbol{H}^{\top}(\boldsymbol{H}\boldsymbol{C}^{-1}\boldsymbol{H}^{\top})^{-1}(\boldsymbol{H}\operatorname{argmin}(Z) - \boldsymbol{h})$.

**(ii)** $\hat{\boldsymbol{\beta}}_n^{\mathrm{L}} - \hat{\boldsymbol{\beta}}_n^{\mathrm{RL}} \xrightarrow{\mathscr{P}} \boldsymbol{C}^{-1}\boldsymbol{H}^{\top}(\boldsymbol{H}\boldsymbol{C}^{-1}\boldsymbol{H}^{\top})^{-1}(\boldsymbol{H}\operatorname{argmin}(Z) - \boldsymbol{h})$.

**(iii)** $\hat{\boldsymbol{\beta}}_n^{\mathrm{PTL}} \xrightarrow{\mathscr{P}} \operatorname{argmin}(Z) - \boldsymbol{C}^{-1}\boldsymbol{H}^{\top}(\boldsymbol{H}\boldsymbol{C}^{-1}\boldsymbol{H}^{\top})^{-1}(\boldsymbol{H}\operatorname{argmin}(Z) - \boldsymbol{h})I(\mathscr{L} < \mathscr{L}_{\alpha})$.

**(iv)** $\hat{\boldsymbol{\beta}}_n^{\mathrm{SSL}} \xrightarrow{\mathscr{D}} \operatorname{argmin}(Z) - k\boldsymbol{C}^{-1}\boldsymbol{H}^{\top}(\boldsymbol{H}\boldsymbol{C}^{-1}\boldsymbol{H}^{\top})^{-1}(\boldsymbol{H}\operatorname{argmin}(Z) - \boldsymbol{h})\mathscr{L}^{-1}$.

**(v)** $\hat{\boldsymbol{\beta}}_n^{\mathrm{PRSSL}} \xrightarrow{\mathscr{D}} \operatorname{argmin}(Z) - (k\mathscr{L}^{-1} + (1 - k\mathscr{L}^{-1})I(\mathscr{L} \le k))\boldsymbol{C}^{-1}\boldsymbol{H}^{\top}(\boldsymbol{H}\boldsymbol{C}^{-1}\boldsymbol{H}^{\top})^{-1}$
$\times(\boldsymbol{H}\operatorname{argmin}(Z) - \boldsymbol{h})$.

*where $\mathscr{L}_{\alpha}$ is the upper critical value of chi-squared distribution with $q$ d.f., $k = q - 2$, and $Z(\boldsymbol{\phi}) = (\phi - \beta)^{\top}\boldsymbol{C}(\phi - \beta) + \lambda_0 \sum_{j=1}^{p} |\phi_j|$.*

*Proof.* According to Theorem 2 of Knight and Fu (2000), if $\boldsymbol{C}$ is a nonsingular matrix and $\lambda_n/n \to \lambda_0 \ge 0$, then $\hat{\boldsymbol{\beta}}_n^{\mathrm{L}} \xrightarrow{\mathscr{D}} \operatorname{argmin}(Z)$. To prove (i), by Slutsky's theorem, Eq. (4), and regularity condition (A2), we have

$$\begin{aligned}\hat{\boldsymbol{\beta}}_n^{\mathrm{L}} - \boldsymbol{C}_n^{-1}\boldsymbol{H}^{\top}(\boldsymbol{H}\boldsymbol{C}_n^{-1}\boldsymbol{H}^{\top})^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}}_n^{\mathrm{L}} - \boldsymbol{h}) \quad \xrightarrow{\mathscr{P}} \quad & \operatorname{argmin}(Z) - \boldsymbol{C}^{-1}\boldsymbol{H}^{\top}(\boldsymbol{H}\boldsymbol{C}^{-1}\boldsymbol{H}^{\top})^{-1} \\ & \times(\boldsymbol{H}\operatorname{argmin}(Z) - \boldsymbol{h}).\end{aligned}$$

(ii) By Eq. (4), we have $\hat{\boldsymbol{\beta}}_n^{\mathrm{L}} - \hat{\boldsymbol{\beta}}_n^{\mathrm{RL}} = \boldsymbol{C}_n\boldsymbol{H}^{\top}(\boldsymbol{H}\boldsymbol{C}_n^{-1}\boldsymbol{H}^{\top})^{-1}(\boldsymbol{H}\hat{\boldsymbol{\beta}}_n^{\mathrm{L}} - \boldsymbol{h})$, which converges to $\boldsymbol{C}^{-1}\boldsymbol{H}^{\top}(\boldsymbol{H}\boldsymbol{C}^{-1}\boldsymbol{H}^{\top})^{-1}(\boldsymbol{H}\operatorname{argmin}(Z) - \boldsymbol{h})$. the result follows by Slutsky's theorem and regularity condition (A2). (iv) From Theorem 2, $I(\mathscr{L}_n \le \mathscr{L}_{n,\alpha}) \xrightarrow{\mathscr{D}} I(\mathscr{L} \le \mathscr{L}_{\alpha})$. Making use of Eq. (5), (iii), and Slutsky's theorem, we have

$$\hat{\boldsymbol{\beta}}_n^{\mathrm{PTL}} \xrightarrow{\mathscr{P}} \operatorname{argmin}(Z) - \boldsymbol{C}^{-1}\boldsymbol{H}^{\top}(\boldsymbol{H}\boldsymbol{C}^{-1}\boldsymbol{H}^{\top})^{-1}(\boldsymbol{H}\operatorname{argmin}(Z) - \boldsymbol{h})I(\mathscr{L} < \mathscr{L}_{\alpha})$$

To prove (iv) and (v), since $k_n \to k = q - 2$, the result is obvious using Eq. (6), (iii), and Slutsky's theorem. ■

Similar results as in Theorem 3 can be obtained using Theorem 2 of Knight and Fu (2000).

**Remark 1** *Under the assumptions of Theorem 3 and $\lambda_n = o(n)$, we have the following results,*

**(i)** $\hat{\beta}_n^{\mathrm{RL}} \xrightarrow{\mathscr{P}} \beta - \delta; \quad \delta = C^{-1}H^{\top}(HC^{-1}H^{\top})^{-1}(H\beta - h)$.

**(ii)** $\hat{\beta}_n^{\mathrm{PTL}} \xrightarrow{\mathscr{P}} \beta - \delta I(\mathscr{L} < \mathscr{L}_\alpha)$.

**(iii)** $\hat{\beta}_n^{\mathrm{SSL}} \xrightarrow{\mathscr{P}} \beta - \delta \mathscr{L}^{-1}$.

**(iv)** $\hat{\beta}_n^{\mathrm{PRSSL}} \xrightarrow{\mathscr{P}} \beta - \left\{k\mathscr{L}^{-1} + (1 - k\mathscr{L}^{-1})I(\mathscr{L} < k)\right\}\delta$.

**Remark 2** *Under $\mathscr{H}_0$, all estimators are consistent for $\beta$.*

## 4. Numerical analysis

In this section, we evaluate performance of the proposed estimators using a simulation study along with a real example.

### 4.1. Simulation

In this section, we conduct a Monte Carlo simulation to analyse relative efficiencies with respect to different levels of sparsity. In particular, we use $\mathrm{RE}(\hat{\beta}^*; \hat{\beta}^{\mathrm{L}}) = \mathrm{R}(\hat{\beta}^{\mathrm{L}})/\mathrm{R}(\hat{\beta}^*)$, where $\hat{\beta}^*$ is one of the proposed estimators in this paper.

We generate a matrix $X$ from a multivariate normal distribution with mean vector $\mu = \mathbf{0}$ and covariance matrix $\Sigma$. The off-diagonal elements of the covariance matrix are considered to be equal to $r$ with $r = 0, 0.2, 0.9$. We consider $n = 100$ and various $p$ ranging 10, 15, and 20.

One of the most applicable $H$ and $h$ is to select variables. Sometimes, an expert claims that some variables do not affect regression model. If we suppose $\beta = (\beta_1^{\top}, \beta_2^{\top})^{\top}$, then $\beta_2 = \mathbf{0}$ is equivalent to the variables that may be ignored for predicting model.

Let us consider $\beta = \left(\beta_1^{\top}, \beta_2^{\top}\right)^{\top} = \left(\mathbf{1}_{p-q}^{\top}, \mathbf{0}_q^{\top}\right)^{\top}$, where $\mathbf{1}_{p-q}$ and $\mathbf{0}_q$ stand for the vectors of 1 and 0 with dimensions $p - q$ and $q$, respectively. In order to investigate the behaviour of the proposed estimators, we define $\Delta^* = \|\beta - \beta_0\|$, where $\beta_0 = \left(\mathbf{1}_{p-q}^{\top}, \mathbf{0}_q^{\top}\right)^{\top}$ and $\|\cdot\|$ is the Euclidean norm. If $\Delta^* = 0$, then $\beta = \beta_0$ while $\beta = (\mathbf{1}_{p-q}^{\top}, \boldsymbol{\Delta}^{\top})^{\top}$ when $\Delta^* > 0$, where $\boldsymbol{\Delta} = (\Delta, \ldots, \Delta)^{\top}$ is the $q$-dimensional vector of $\Delta$ values. When we increase the number of $\Delta^*$, it indicates the degree of violation of the null hypothesis.

In our simulation study, without loss of generality, we assume $\beta$ is a $p$-vector in which the first $s$ components of $\beta$ are $\mathbf{1}$ and other $(p - s)$ components are zero. The responses were simulated from the following model:

$$y_i = \sum_{i=1}^{p} x_i\beta_i + e_i, \qquad e_i \sim \mathscr{N}(0, 1)$$

Each realization was repeated 1000 times to obtain risk of the estimated regression parameters. Thus, risks are calculated for the ULE, RLE, PTLE, SSLE and PRSSLE. The results are tabulated in Tables 1-3.

The findings of Tables 1-3 may be summarized as:

a) When the null hypothesis is true ($\Delta^2 = 0$), RLE behaves better than other estimator. As we depart from the null hypothesis, the performance of this estimator decreases.

b) For large $\Delta^2$, the performance of estimators decreases; even, when the correlation is low, the unrestricted LASSO performs better.

c) Neither PTLE nor Stein-type shrinkage LASSO estimator dominates each other.

d) The positive rule Stein-type shrinkage LASSO uniformly dominates Stein-type LASSO estimator.

e) It is well - known that shrinkage and positive-rule shrinkage estimators are always better than unrestricted estimator. Here, the results confirm that also.

***Table 1:*** *Relative efficiencies (standard errors) of the estimators for fixed $\Delta^2$, $r = 0$, $s = 6$ different values of p.*

| | ULE | RLE | PTLE | SSLE | PRSSLE |
|---|---|---|---|---|---|
| p | | | $\Delta^2 = 0$ | | |
| 10 | 1 (0.003) | 1.63 (0.011) | 1.52 (0.011) | 1.23 (0.008) | 1.35 (0.009) |
| 15 | 1 (0.002) | 2.32 (0.013) | 2.19 (0.013) | 1.75 (0.007) | 2.05 (0.010) |
| 20 | 1 (0.001) | 3.51 (0.007) | 2.98 (0.007) | 2.44 (0.004) | 2.99 (0.006) |
| p | | | $\Delta^2 = 0.1$ | | |
| 10 | 1 (0.001) | 1.57 (0.003) | 1.46 (0.003) | 1.21 (0.003) | 1.32 (0.003) |
| 15 | 1 (0.001) | 2.18 (0.007) | 2.02 (0.007) | 1.66 (0.005) | 1.95 (0.006) |
| 20 | 1 (0.001) | 3.48 (0.015) | 2.92 (0.015) | 2.40 (0.008) | 2.98 (0.011) |
| p | | | $\Delta^2 = 0.5$ | | |
| 10 | 1 (0.001) | 0.85 (0.000) | 0.86 (0.001) | 1.05 (0.001) | 1.07 (0.001) |
| 15 | 1 (0.002) | 1.57 (0.002) | 1.42 (0.003) | 1.37 (0.003) | 1.54 (0.003) |
| 20 | 1 (0.001) | 2.86 (0.004) | 2.34 (0.004) | 2.19 (0.004) | 2.58 (0.004) |
| p | | | $\Delta^2 = 1$ | | |
| 10 | 1 (0.001) | 0.36 (0.000) | 0.90 (0.001) | 1.00 (0.001) | 1.00 (0.001) |
| 15 | 1 (0.000) | 0.82 (0.000) | 0.86 (0.000) | 1.12 (0.000) | 1.14 (0.000) |
| 20 | 1 (0.003) | 1.81 (0.002) | 1.43 (0.004) | 1.81 (0.005) | 1.90 (0.005) |
| p | | | $\Delta^2 = 5$ | | |
| 10 | 1 (0.001) | 0.02 (0.007) | 1.00 (0.007) | 0.94 (0.004) | 0.94 (0.006) |
| 15 | 1 (0.001) | 0.06 (0.000) | 1.00 (0.000) | 0.98 (0.000) | 0.98 (0.000) |
| 20 | 1 (0.001) | 0.13 (0.000) | 1.00 (0.001) | 1.00 (0.001) | 1.00 (0.001) |

***Table 2:** Relative efficiencies (standard errors) of the estimators for fixed $\Delta^2$, $r = 0.2$, $s = 6$ different values of p.*

| | ULE | RLE | PTLE | SSLE | PRSSLE |
|---|---|---|---|---|---|
| p | $\Delta^2 = 0$ | | | | |
| 10 | 1 (0.003) | 1.95 (0.014) | 1.71 (0.014) | 1.34 (0.008) | 1.44 (0.011) |
| 15 | 1 (0.002) | 2.70 (0.015) | 2.37 (0.016) | 1.90 (0.012) | 2.28 (0.014) |
| 20 | 1 (0.000) | 4.77 (0.010) | 3.55 (0.009) | 2.98 (0.006) | 3.63 (0.008) |
| p | $\Delta^2 = 0.1$ | | | | |
| 10 | 1 (0.001) | 1.93 (0.004) | 1.61 (0.004) | 1.33 (0.002) | 1.40 (0.003) |
| 15 | 1 (0.001) | 2.69 (0.009) | 2.30 (0.010) | 1.92 (0.006) | 2.21 (0.008) |
| 20 | 1 (0.001) | 4.72 (0.010) | 3.39 (0.009) | 2.96 (0.006) | 3.62 (0.008) |
| p | $\Delta^2 = 0.5$ | | | | |
| 10 | 1 (0.001) | 0.97 (0.000) | 0.91 (0.001) | 1.17 (0.002) | 1.17 (0.002) |
| 15 | 1 (0.002) | 1.87 (0.003) | 1.30 (0.004) | 1.70 (0.004) | 1.74 (0.004) |
| 20 | 1 (0.001) | 3.74 (0.005) | 1.93 (0.006) | 2.75 (0.005) | 2.91 (0.005) |
| p | $\Delta^2 = 1$ | | | | |
| 10 | 1 (0.001) | 0.37 (0.000) | 0.99 (0.001) | 1.08 (0.001) | 1.08 (0.001) |
| 15 | 1 (0.000) | 0.85 (0.000) | 0.97 (0.000) | 1.37 (0.000) | 1.37 (0.000) |
| 20 | 1 (0.003) | 1.85 (0.000) | 1.07 (0.003) | 2.10 (0.007) | 2.10 (0.007) |
| p | $\Delta^2 = 5$ | | | | |
| 10 | 1 (0.001) | 0.01 (0.001) | 1.00 (0.001) | 0.99 (0.001) | 0.99 (0.001) |
| 15 | 1 (0.001) | 0.04 (0.000) | 1.00 (0.000) | 1.02 (0.001) | 1.02 (0.001) |
| 20 | 1 (0.001) | 0.85 (0.000) | 1.00 (0.001) | 1.13 (0.001) | 1.13 (0.001) |

The linear regression model is fitted to this dataset in order to predict the response variable. The LASSO of Tibshirani (1996) (the UL in our study), restricted LASSO (RL), preliminary test LASSO (PTL), Stein-type shrinkage LASSO (SSL), and positive rule Stein-type shrinkage (PRSSL) estimators are used to estimate the unknown regression coefficients.

Since one of the biggest problems in estimation is to determine $\boldsymbol{H}$ and $\boldsymbol{h}$, we suppose that $\boldsymbol{H} = \boldsymbol{I}_7$. This choice is just for simplicity and also to avoid errors obtained by incorrect selection of parameters.

In order to show the impact of correctness or incorrectness of hypothesis, we consider the following two cases:

**Case I.** Let $\boldsymbol{h} = (0, 0, 10, 0.2, 0.7, 0.06, 0)^{\intercal}$. The null hypothesis changes into $\mathscr{H}_o : \boldsymbol{\beta} = \boldsymbol{h}$ and thus, the variables POPULATION, INCOME, and AREA are insignificant.

**Case II.** Let $\boldsymbol{h} = (0, 0, 0, 0, 0, 0, 0)^{\intercal}$. The null hypothesis changes into $\mathscr{H}_o : \boldsymbol{\beta} = \boldsymbol{0}$ and thus, all variables are insignificant.

***Table 3:*** *Relative efficiencies (standard errors) of the estimators for fixed $\Delta^2$, $r = 0.9$, $s = 6$ different values of p.*

| | ULE | RLE | PTLE | SSLE | PRSSLE |
|---|---|---|---|---|---|
| p | | | $\Delta^2 = 0$ | | |
| 10 | 1 (0.002) | 6.76 (0.051) | 3.34 (0.052) | 1.86 (0.019) | 1.86 (0.019) |
| 15 | 1 (0.001) | 8.46 (0.006) | 5.76 (0.061) | 4.28 (0.036) | 4.68 (0.042) |
| 20 | 1 (0.000) | 14.74 (0.064) | 14.48 (0.064) | 6.13 (0.065) | 11.31 (0.057) |
| p | | | $\Delta^2 = 0.1$ | | |
| 10 | 1 (0.001) | 6.35 (0.062) | 2.95 (0.017) | 1.81 (0.004) | 1.82 (0.004) |
| 15 | 1 (0.001) | 8.31 (0.004) | 5.77 (0.041) | 4.28 (0.024) | 4.68 (0.031) |
| 20 | 1 (0.001) | 14.11 (0.085) | 12.56 (0.085) | 5.96 (0.052) | 10.68 (0.082) |
| p | | | $\Delta^2 = 0.5$ | | |
| 10 | 1 (0.001) | 3.28 (0.003) | 1.39 (0.005) | 1.69 (0.005) | 1.69 (0.005) |
| 15 | 1 (0.002) | 5.41 (0.017) | 2.71 (0.020) | 3.85 (0.028) | 3.92 (0.028) |
| 20 | 1 (0.001) | 10.40 (0.020) | 7.10 (0.021) | 6.77 (0.032) | 8.68 (0.031) |
| p | | | $\Delta^2 = 1$ | | |
| 10 | 1 (0.002) | 1.18 (0.000) | 0.96 (0.001) | 1.55 (0.005) | 1.55 (0.005) |
| 15 | 1 (0.000) | 2.50 (0.000) | 1.26 (0.001) | 3.14 (0.003) | 3.14 (0.003) |
| 20 | 1 (0.002) | 5.28 (0.012) | 2.57 (0.019) | 5.33 (0.053) | 6.58 (0.053) |
| p | | | $\Delta^2 = 5$ | | |
| 10 | 1 (0.005) | 0.02 (0.000) | 1.00 (0.005) | 0.87 (0.004) | 0.87 (0.004) |
| 15 | 1 (0.002) | 0.06 (0.000) | 1.00 (0.002) | 1.58 (0.004) | 1.58 (0.004) |
| 20 | 1 (0.001) | 0.19 (0.000) | 1.00 (0.001) | 2.68 (0.006) | 2.68 (0.006) |

***Table 4:*** *Description of the variables of state.x77.*

| Variables | Description | Role |
|---|---|---|
| LifeExp | Average years of life expectancy at birth | Response |
| Population | in thousands | Predictor |
| Income | dollars per capita | Independent |
| Illiteracy | Percentage of those unable to read and write | Independent |
| Murder | number of murders and non-negligent manslaughters per 100000 people | Independent |
| HS Grad | percentage of adults who were high-school graduates | Independent |
| Frost | mean number of days per year with low temperatures below freezing | Independent |
| Area | in square miles | Independent |

### 4.2. Real data

In this section, we study the performance of proposed LASSO-based shrinkage estimators using state.x77 dataset (available by default in R software). Descriptions of the variables in this dataset are given in Table 4.

***Table 5:*** *5-fold cross validation relative average prediction errors for state data.*

| | RLE | | PTLE | | SSLE | PRSSLE |
|---|---|---|---|---|---|---|
| | | 0.01 | 0.05 | 0.10 | | |
| Case I | 22.2615 | 1.0009 | 1.0004 | 1.0004 | 1.0200 | 1.0208 |
| Case II | 1.0017 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0008 |

The performance of the estimators are evaluated using average five-fold cross validation error. By choosing 1000 as a large enough number for repeating process in a bootstrap simulation scheme, Table 5 shows the relative average prediction errors in the two cases.

Based on Table 5, RLE is the best estimator because the hypothesis $\boldsymbol{H\beta} = \boldsymbol{h}$ is nearly true, but PRSSLE has lower prediction error than other estimators in case I. This estimator is followed by SSLE. Indeed, by departing from the null hypothesis, these estimators will behave similar to the LASSO in case II. If the level of significance $\alpha$ for constructing PTLE increases, then the prediction error decreases.

## 5. Conclusion

In this paper, we proposed improved LASSO-based estimators by imposing a subspace restriction to the linear regression model. Particularly, we introduced preliminary-test LASSO, Stein-type shrinkage LASSO, and positive-rule shrinkage LASSO estimators. Asymptotic performance of the proposed estimators studied in case $n > p$. The proposed methodology for improving the LASSO can also be applied to the high-dimensional case $p > n$. Indeed the test statistic for $\mathscr{H}_o : \boldsymbol{H\beta} = \boldsymbol{h}$ plays a determining role.

In addition to the given theorems for the asymptotic behaviour of the proposed estimators, using a simulation study, we compared the performance of estimators numerically for various configurations of $p$, correlation coefficient between the predictors ($r$), and the error in variance ($\sigma^2$). For different non-centrality parameter $\Delta$, degree of model misspecification, the number of non-zero $\beta$s varied, and then the performance of estimators evaluated. We found that the positive-rule shrinkage LASSO estimator has the best performance among all. When we deviated from the null model, neither PTLE nor SSLE dominated one another and the PTLE performed better as $\alpha$ became large. Relative efficiency of the proposed estimators increased when there were more near-zero parameters in the model. As an application, a real dataset was analysed, where a five-fold cross-validation averages and standard deviations of the prediction errors were evaluated for the LASSO and its other four variants. The new estimators dominated the LASSO in average prediction error sense.

## Acknowledgments

We would like to thank the Editor, Prof. Maria Durban, and anonymous referees for their valuable suggestions and comments which substantially improved the presentation and led to adding many details in the paper. Second author M. Arashi's work is based on the research supported in part by the National Research Foundation of South Africa (Grant NO. 109214).

## References

Ahmed, S.E. and Raheem, S.M.E. (2012). Shrinkage and absolute penalty estimation in linear regression models, *Wires: Computational Statistics*, 4, 541–553.

Fallahpour, S., Ahmed, S.E. and Doksum, K.A. (2012). L1 penalty and shrinkage estimation in partially linear models with random coefficient autoregressive errors, *Applied Stochastic Models in Business and Industry*, 28, 236–250.

Hossain, S. and Ahmed, S.E. (2014). Penalized and Shrinkage Estimation in the Cox Proportional Hazards Model. *Communications in Statistics-Theory and Methods*, 43, 1026–1040.

Hossain, S., Ahmed, S.E. and Doksum, K.A. (2015). Shrinkage, pretest, and penalty estimators in generalized linear models. *Statistical Methodology*, 24, 52–68.

Hossain, S. and Ahmed, S.E. and Yi, Y. (2016). Shrinkage and pretest estimators for longitudinal data analysis under partially linear models. *Journal of Nonparametric Statistics*, DOI:10.1080/10485252.2016. 1190358.

Hossain, S. and Howlader, H. (2016). Shrinkage estimation in lognormal regression model for censored data. *Journal of Applied Statistics*, DOI: 10.1080/02664763.2016.1168365.

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

Knight, K. and Fu, W. (2000). Asymptotics for lasso-type estimators. *Annals of Statistics*, 28, 1356–1378.

Roozbeh, M. (2015). Shrinkage ridge estimators in semiparametric regression models. *Journal of Multivariate Analysis*, 136, 56–74.

Roozbeh, M. (2016). Robust ridge estimator in restricted semiparametric regression models. *Journal of Multivariate Analysis*, 147, 127–144.

Saleh, A.K.M.E. (2006). *Theory of preliminary test and stein-type estimation with applications*, John Wiley & Sons, New York.

Sengupta, D. and Jammalamadaka, S.R. (2003). *Linear models: An integrated approach*, World Scientific Publishing Company.

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

Yuzbasi, B. and Ahmed, S.E. (2016). Shrinkage and penalized estimation in semi-parametric models with multicollinear data. *Journal of Statistical Computation and Simulation*, 86, 3543–3561.

Yuzbasi, B., Ahmed, S.E. and Gungor, M. (2017). Improved Penalty Strategies in Linear Regression Models. *REVSTAT-Statistical Journal*, Accepted.