RUHUNA JOURNAL OF SCIENCE
Vol 13 (1): 14-28, June 2022
eISSN: 2536-8400
http://doi.org/10.4038/rjs.v13i1.112

Faculty of Science
University of Ruhuna

# Generalized Stochastic Restricted LARS Algorithm

Manickavasagar Kayanan[1,2] and Pushpakanthie Wijekoon[3]

[1]*Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka*
[2] *Department of Physical Science, University of Vavuniya, Vavuniya, Sri Lanka*
[3] *Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka*

*Correspondence: *mgayanan@vau.jfn.ac.lk*; ORCID: https://orcid.org/0000-0003-2662-4383

Received: 29th July 2021; Revised: 21st April 2022; Accepted: 17th May 2022

**Abstract** The Least Absolute Shrinkage and Selection Operator (LASSO) is used to tackle both the multicollinearity issue and the variable selection concurrently in the linear regression model. The Least Angle Regression (LARS) algorithm has been used widely to produce LASSO solutions. However, this algorithm is unreliable when high multicollinearity exists among regressor variables. One solution to improve the estimation of regression parameters when multicollinearity exists is adding preliminary information about the regression coefficient to the model as either exact linear restrictions or stochastic linear restrictions. Based on this solution, this article proposed a generalized version of the stochastic restricted LARS algorithm, which combines LASSO with existing stochastic restricted estimators. Further, we examined the performance of the proposed algorithm by employing a Monte Carlo simulation study and a numerical example.

**Keywords:** LASSO, LARS, Stochastic Linear Restrictions.

## 1 Introduction

The biased estimators such as Ridge Estimator (RE) (Hoerl and Kennard 1970), Almost Unbiased Ridge Estimator (AURE) (Singh *et al*. 1986), Liu Estimator (LE) (Liu 1993), Almost Unbiased Liu Estimator (AULE) (Akdeniz and Kaçiranlar 1995), Principle Component Regression Estimator (PCRE) (Massy 1965), r-k class estimator (Baye and Parker 1984), r-d class estimator (Kaçiranlar and Sakallıoğlu 2001) and Sample Information Optimal Estimator (SIOE) (Kayanan and Wijekoon 2019) have been widely used in literature to resolve multicollinearity issue in the linear regression model. However, these estimators yield high bias when the number of explanatory variables is high, and they do not consider about irrelevant variables while fitting models. For high dimensional data, having many variables in the model and multicollinearity are major issues. To tackle these matters, Tibshirani (1996) introduced Least Absolute Shrinkage and Selection Operator (LASSO). The LASSO

is a shrinkage method that was originally used for regularization and variable selection in the linear regression model. The Least Angle Regression (LARS) (Efron *et al*. 2004) algorithm has been used to obtain the estimates of LASSO. Zou and Hastie (2005) have shown that the LASSO is unsteady when severe multicollinearity exists between the explanatory variables. Therefore, they suggested Elastic Net (ENet) estimator by combining LASSO and RE as a solution for this issue. Furthermore, they proposed LARS-EN algorithm to attain ENet solutions, which is a modified version of the LARS algorithm. Besides, Kayanan and Wijekoon (2020b) proposed a generalized version of LARS (GLARS) algorithm to combine LASSO with RE and the other biased estimators based on sample information such as AURE, LE, AULE, PCRE, r-k class estimator, and r-d class estimator. Finally, they have shown that the GLARS algorithm performs well when it combines LASSO with r-k class and r-d class estimators.

According to literature, the parameter estimation can be strengthened if prior knowledge about the regression coefficient is applied. The prior information on regression coefficients can be defined in the form of exact linear restrictions or stochastic linear restrictions. Many researchers proposed stochastic restricted estimators such as Mixed Regression Estimator (MRE) (Theil and Goldberger 1961) Stochastic Restricted Ridge Estimator (SRRE) (Li and Yang 2010), Stochastic Restricted Almost Unbiased Ridge Estimator (SRAURE) (Jibo and Hu 2014), Stochastic Restricted Liu Estimator (SRLE) (Hubert and Wijekoon 2006), Stochastic Restricted Almost Unbiased Liu Estimator (SRAULE) (Jibo and Hu 2014), Stochastic Restricted Principle Component Regression Estimator (SRPCRE) (He and Wu 2014), Stochastic Restricted r-k class estimator (SRrk) (Jibo 2014), Stochastic Restricted r-d class estimator (SRrd) (Jibo 2014), and Stochastic Restricted Optimal Estimator (SROE) (Kayanan and Wijekoon 2019) to incorporate prior information to the regression coefficient. Stochastic restricted estimators also have the same issue as biased estimators when the linear regression model contains numerous predictors. To handle this problem, Kayanan and Wijekoon (2020a) proposed a stochastic restricted LARS (SRLARS) algorithm to combine LASSO and MRE, and showed the superiority of the SRLARS over LARS algorithm.

This article proposes a generalized version of the stochastic restricted LARS algorithm, namely SRGLARS, to combine LASSO with other stochastic restricted estimators. The prediction performance of the SRGLARS algorithm was examined by employing a Monte-Carlo simulation and using a real-world example in the Root Mean Square Error (RMSE) criterion.

## 2 Model Specification and the Estimators

Consider the linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.1}$$

where $\boldsymbol{X}$ is $n \times p$ matrix of explanatory variables, $\boldsymbol{\beta}$ be the $p \times 1$ vector of unknown coefficients, and $\boldsymbol{\varepsilon}$ be the $n \times 1$ vector of disturbances such that $\boldsymbol{\varepsilon} \sim N\ (0, \sigma^2\ \boldsymbol{I})$.

Assume that there exists prior information on β, which may be expressed as a stochastic linear restriction, as (Theil & Goldberger, 1961)

$$\boldsymbol{\varphi} = \boldsymbol{R\beta} + \boldsymbol{v}, \tag{2.2}$$

where $\boldsymbol{\varphi}$ be the $q \times 1$ vector, $\boldsymbol{R}$ be the $q \times p$ matrix with rank $q$, $\boldsymbol{v}$ be the $q \times 1$ vector of disturbances, such that $\boldsymbol{v} \sim N(0, \sigma^2 \boldsymbol{W})$, $\boldsymbol{W}$ is positive definite, and $E(\boldsymbol{v\varepsilon}') = 0$.

To make the variable selection and handle multicollinearity issue by incorporating prior information defined in model (2.2), Kayanan and Wijekoon (2020a) suggested Stochastic Restricted LASSO Type Estimator (SRLASSO) for model (2.1) as

$$\widehat{\boldsymbol{\beta}}_{SRLASSO} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\{(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})\}$$
$$\text{subject to } \textstyle\sum_{j=1}^{p} |\beta_j| \leq t \text{ and } \boldsymbol{R\beta} = \boldsymbol{\varphi} - \boldsymbol{v}, \tag{2.3}$$

where $t > 0$ is a turning parameter.

Further, Kayanan and Wijekoon (2020a) proposed Stochastic Restricted LARS (SRLARS) algorithm to find the SRLASSO estimates. Note that SRLARS combines LASSO and MRE to find the estimates. To improve the SRLARS solutions, this article proposes a generalized version of SRLARS (SRGLARS) to combine LASSO with other stochastic restricted estimators such as SRRE, SRAURE, SRLE, SRAULE, SRPCRE, SRrk, SRrd and SROE.

Kayanan and Wijekoon (2018, 2019) proposed a generalized form to express the estimators MRE, SRRE, SRAURE, SRLE, SRAULE, SRPCRE, SRrk, SRrd, and SROE as

$$\widehat{\boldsymbol{\beta}}_G = \boldsymbol{G}(\boldsymbol{X'X} + \boldsymbol{R'W^{-1}R})^{-1}(\boldsymbol{X'y} + \boldsymbol{R'W^{-1}\varphi}), \tag{2.4}$$

where

$$\widehat{\boldsymbol{\beta}}_G = \begin{cases} \widehat{\boldsymbol{\beta}}_{MRE} & \text{if } \boldsymbol{G} = \boldsymbol{I} \\ \widehat{\boldsymbol{\beta}}_{SRRE} & \text{if } \boldsymbol{G} = (\boldsymbol{X'X} + k\boldsymbol{I})^{-1}\boldsymbol{X'X} \\ \widehat{\boldsymbol{\beta}}_{SRAURE} & \text{if } \boldsymbol{G} = (\boldsymbol{I} - k^2(\boldsymbol{X'X} + k\boldsymbol{I})^{-2}) \\ \widehat{\boldsymbol{\beta}}_{SRLE} & \text{if } \boldsymbol{G} = (\boldsymbol{X'X} + \boldsymbol{I})^{-1}(\boldsymbol{X'X} + d\boldsymbol{I}) \\ \widehat{\boldsymbol{\beta}}_{SRAULE} & \text{if } \boldsymbol{G} = (\boldsymbol{I} - (1-d)^2(\boldsymbol{X'X} + \boldsymbol{I})^{-2}) \\ \widehat{\boldsymbol{\beta}}_{SRPCRE} & \text{if } \boldsymbol{G} = \boldsymbol{T}_h\boldsymbol{T}_h' \\ \widehat{\boldsymbol{\beta}}_{SRrk} & \text{if } \boldsymbol{G} = \boldsymbol{T}_h\boldsymbol{T}_h'(\boldsymbol{X'X} + k\boldsymbol{I})^{-1}\boldsymbol{X'X} \\ \widehat{\boldsymbol{\beta}}_{SRrd} & \text{if } \boldsymbol{G} = \boldsymbol{T}_h\boldsymbol{T}_h'(\boldsymbol{X'X} + \boldsymbol{I})^{-1}(\boldsymbol{X'X} + d\boldsymbol{I}) \\ \widehat{\boldsymbol{\beta}}_{SROE} & \text{if } \boldsymbol{G} = \boldsymbol{\beta}_*\boldsymbol{\beta}_*'\left(\sigma^2(\boldsymbol{X'X})^{-1} + \boldsymbol{\beta}_*\boldsymbol{\beta}_*'\right)^{-1} \end{cases}$$

Note that $k > 0$ and $0 < d < 1$ are the shrinkage/regularization parameters, $\boldsymbol{I}$ is the $p \times p$ identity matrix, $\boldsymbol{T}_h = (t_1, t_2 \ldots t_h)'$ is the first $h$ columns of the eigenvectors of $\boldsymbol{X'X}$, and $\boldsymbol{\beta}_*$ is the normalized eigenvector corresponding to the largest eigenvalue of $\boldsymbol{X'X}$.

## 3 SRGLARS Algorithm

Based on Kayanan and Wijekoon (2021a) and equation (2.4), the SRGLARS algorithm for model (2.1) is outlined below:

Algorithm 1: SRGLARS

1: Standardize $\boldsymbol{X}$ to have a mean zero with a standard deviation of one, and center the $\boldsymbol{y}$ to have a mean zero.
2: Start with all estimates of the coefficients $\widehat{\boldsymbol{\beta}} = 0$ with the residuals $\boldsymbol{r} = \hat{\boldsymbol{\varepsilon}}$ and $\boldsymbol{\tau} = \widehat{\boldsymbol{v}}$.
3: Find the predictor $\boldsymbol{X_j}$ most correlated with $\boldsymbol{r}$; $j = 1, 2, \ldots, p$.
4: Move the estimate of $\hat{\beta}_j$ from 0 towards the $\widehat{\boldsymbol{\beta}}_G$ direction until some other predictor $\boldsymbol{X_k}$ has as large a correlation with the current residual as $\boldsymbol{X_j}$ does.
5: Move $\hat{\beta}_j$ and $\hat{\beta}_k$ in the direction defined by their joint $\widehat{\boldsymbol{\beta}}_G$ direction of the current residual on $(\boldsymbol{X_j}, \boldsymbol{X_k})$, until some other predictor $\boldsymbol{X_l}$ eventually earns its way into the active set.
6: If a non-zero coefficient hits zero, drop its variable from the active set of variables and recomputed the current joint $\widehat{\boldsymbol{\beta}}_G$ direction.
7: Repeat the steps 5 and 6 until SRGLARS conditions attained.

The mathematical details of the SRLARS algorithm are as follows:

Let us assume that the estimates of the coefficients $\widehat{\boldsymbol{\beta}}$ and residuals $\boldsymbol{r}$ and $\boldsymbol{\tau}$ are $\left(\widehat{\boldsymbol{\beta}}\right)_0 = \left[(\beta_1)_0, (\beta_2)_0, \ldots, \left(\beta_p\right)_0\right]' = \boldsymbol{0}, \boldsymbol{r}_0 = \boldsymbol{y}$ and $\boldsymbol{\tau}_0 = \boldsymbol{\varphi}$.

Find the predictor $(X_j)_1$ most correlated with $\boldsymbol{r}_0$.

$$(X_j)_1 = \max_j \left|Cor(X_j, \boldsymbol{r}_0)\right|; j \in (1,2, \ldots, p) \tag{3.1}$$

Then, increase the estimate of respective regression coefficient $\left(\beta_j\right)_1$ from 0 unto any other predictor $(X_j)_2$ has a high correlation with $\boldsymbol{r}_1$ as $(X_j)_1$ does. At this stage, SRGLARS moves in the equiangular direction between $(X_j)_1$ and $(X_j)_2$ rather than proceeding the path based on $(X_j)_1$.

Similarly, in $i^{th}$ run, the variable $(X_j)_1$ eventually acquires its path in the active set, and then SRGLARS moves in the equiangular direction among $(\boldsymbol{X})_i = \left[\left(X_j\right)_1, \left(X_j\right)_2, \ldots, \left(X_j\right)_i\right]'$. Proceed to add variables to the active set in this way, running in the path established by the least angle direction. During this process, $\left(\widehat{\boldsymbol{\beta}}\right)_i = \left[\left(\beta_j\right)_1, \left(\beta_j\right)_2, \ldots, \left(\beta_j\right)_i\right]'$ is updating using the following formula:

$$\left(\widehat{\boldsymbol{\beta}}\right)_i = \left(\widehat{\boldsymbol{\beta}}\right)_{i-1} + \alpha_i \boldsymbol{u}_i, \tag{3.2}$$

where $\alpha_i \in [0,1]$ which signifies how long the estimate runs in the path before another predictor enters the model and the path turns anew, and $\boldsymbol{u}_i$ is the equiangular vector.

The path vector $\boldsymbol{u}_i$ is computed using the formula given below based on the generalized form defined in equation (2.4):

$$\boldsymbol{u}_i = \boldsymbol{G}_E((\boldsymbol{E})_i'(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{R})(\boldsymbol{E})_i)^{-1}(\boldsymbol{E})_i'(\boldsymbol{X}'\boldsymbol{r}_{i-1} + \boldsymbol{R}'\boldsymbol{W}^{-1}\boldsymbol{\tau}_{i-1}), \quad (3.3)$$

where $(\boldsymbol{E})_i = \left[(\boldsymbol{e}_j)_1, (\boldsymbol{e}_j)_2, \ldots, (\boldsymbol{e}_j)_i\right]$, and $(e_j)_i$ be the $j^{\text{th}}$ standard unit vector in $R^p$, which has the record of selected variables in every succeeding steps, and $\boldsymbol{G}_E$ is a generalized matrix that can be changed by respective expressions for any of stochastic restricted estimators of our interest as outlined in Table 1.

Then, $\alpha_i$ be calculated as follows:

$$\alpha_i = \min\{\alpha_i^+, \alpha_i^-, \alpha_i^*\}, \quad (3.4)$$

where

$$\alpha_i^{\pm} = \frac{Cor\left((X_j)_i, \boldsymbol{r}_{i-1}\right) \pm Cor(X_j, \boldsymbol{r}_{i-1})}{Cor\left((X_j)_i, \boldsymbol{r}_{i-1}\right) \pm Cor(X_j, (\boldsymbol{X})_i\boldsymbol{u}_i)} \quad (3.5)$$

for any $j$ such that $\left(\beta_j\right)_{i-1} = 0$, and

$$\alpha_i^* = -\frac{(\hat{\boldsymbol{\beta}})_{i-1}}{\boldsymbol{u}_i} \quad (3.6)$$

for any $j$ such that $\left(\beta_j\right)_{i-1} \neq 0$.

If $\alpha_i = \alpha_i^*$, then $(\boldsymbol{E})_i$ is reformed by deleting the column $e_j$ from $(\boldsymbol{E})_{i-1}$. Then $\boldsymbol{r}_i$ and $\boldsymbol{\tau}_i$ can be calculated as

$$\boldsymbol{r}_i = \boldsymbol{r}_{i-1} - \alpha_i(\boldsymbol{X})_i\boldsymbol{u}_i \quad \text{and} \quad (3.7)$$

$$\boldsymbol{\tau}_i = \boldsymbol{\tau}_{i-1} - \alpha_i(\boldsymbol{R})_i\boldsymbol{u}_i, \quad (3.8)$$

where $(\boldsymbol{R})_i = \left[(\boldsymbol{R}_j)_1, (\boldsymbol{R}_j)_2, \ldots, (\boldsymbol{R}_j)_i\right]$. Then proceed to the next step where $(j)_{i+1}$ is the value of $j$ such that $\alpha_i = \alpha_i^+$ or $\alpha_i = \alpha_i^-$ or $\alpha_i = \alpha_i^*$ . Proceed with the algorithm until $\alpha_i = 1$.

Table 1: $G_E$ of the estimators for SRGLARS.

| **Estimators** | $\boldsymbol{G}_E$ |
|---|---|
| MRE | $(\boldsymbol{E})_i$ |
| SRRE | $(\boldsymbol{E})_i((\boldsymbol{E})_i'(\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})(\boldsymbol{E})_i)^{-1}(\boldsymbol{E})_i'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{E})_i$ |
| SRAURE | $(\boldsymbol{E})_i(\boldsymbol{I}_{PE} - k^2((\boldsymbol{E})_i'(\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})(\boldsymbol{E})_i)^{-2})$ |
| SRLE | $(\boldsymbol{E})_i((\boldsymbol{E})_i'(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})(\boldsymbol{E})_i)^{-1}(\boldsymbol{E})_i'(\boldsymbol{X}'\boldsymbol{X} + d\boldsymbol{I})(\boldsymbol{E})_i$ |
| SRAULE | $(\boldsymbol{E})_i(\boldsymbol{I}_{PE} - (1-d)^2((\boldsymbol{E})_i'(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})(\boldsymbol{E})_i)^{-2})$ |
| SRPCRE | $\boldsymbol{T}_{hE}\boldsymbol{T}_{hE}'(\boldsymbol{E})_i$ |
| SRrk | $\boldsymbol{T}_{hE}\boldsymbol{T}_{hE}'(\boldsymbol{E})_i((\boldsymbol{E})_i'(\boldsymbol{X}'\boldsymbol{X} + k\boldsymbol{I})(\boldsymbol{E})_i)^{-1}(\boldsymbol{E})_i'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{E})_i$ |
| SRrd | $\boldsymbol{T}_{hE}\boldsymbol{T}_{hE}'(\boldsymbol{E})_i((\boldsymbol{E})_i'(\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{I})(\boldsymbol{E})_i)^{-1}(\boldsymbol{E})_i'(\boldsymbol{X}'\boldsymbol{X} + d\boldsymbol{I})(\boldsymbol{E})_i$ |
| SROE | $(\boldsymbol{E})_i((\boldsymbol{E})_i'\boldsymbol{\beta}_*\boldsymbol{\beta}_*'(\boldsymbol{E})_i\left(\sigma^2((\boldsymbol{E})_i'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{E})_i)^{-1} + (\boldsymbol{E})_i'\boldsymbol{\beta}_*\boldsymbol{\beta}_*'(\boldsymbol{E})_i\right)^{-1}$ |

In Table 1, $\boldsymbol{I}_{PE}$ is the $p_E \times p_E$ identity matrix, $p_E$ is the amount of selected variables in each succeeding step, and $\boldsymbol{T}_{hE} = (t_1, t_2 \dots t_{hE})$ is the first $h_E$ column of the standardized eigenvectors of $(\boldsymbol{E})_i' \boldsymbol{X}' \boldsymbol{X} (\boldsymbol{E})_i$.

We can apply SRGLARS to consolidate LASSO and any of the stochastic restricted estimators listed in Table 1. The suitable value of regularization parameter $k$ or $d$ of the proposed algorithms can be chosen by 10-fold cross-validation for every $t$ as outlined in Appendix C.

We can get a separate algorithm for each stochastic restricted estimator by referring SRGLARS as LARS-MRE, LARS-SRRE, LARS-SRAURE, LARS-SRLE, LARS-SRAULE, LARS-SRPCRE, LARS-SRrk, LARS-SRrd, and LARS-SROE when GE equals to the corresponding expression of MRE, SRRE, SRAURE, SRLE, SRAULE, SRPCRE, SRrk, SRrd, and SROE, respectively.

## 4 Discussion

The performance of SRGLARS algorithm by considering different combinations of LASSO and stochastic restricted estimators were examined using RMSE criterion, which is the expected prediction error of the algorithm for each estimator, and is defined as

$$RMSE(\widehat{\boldsymbol{\beta}}) = \sqrt{\frac{1}{n_*} \left(\boldsymbol{y}_* - \boldsymbol{X}_* \widehat{\boldsymbol{\beta}}\right)' \left(\boldsymbol{y}_* - \boldsymbol{X}_* \widehat{\boldsymbol{\beta}}\right)}, \quad (4.1)$$

where $(\boldsymbol{y}_*, \boldsymbol{X}_*)$ are the new observations, $n_*$ is the size of new observations, and $\widehat{\boldsymbol{\beta}}$ is the estimated value of $\boldsymbol{\beta}$ by the corresponding algorithm. Note that we cannot provide theoretical conditions for the superiority of particular algorithms in variable selection methods. Therefore, we use a Monte Carlo simulation and a real-world example to compare the SRGLARS algorithms.

### 4.1 Simulation study

We used the following formula to generate the explanatory variables based on McDonald and Galarneau (1975):

$$x_{i,j} = \sqrt{(1 - \rho^2)} z_{i,j} + \rho z_{i,m+1}; \quad i = 1, 2, \dots, n. \quad j = 1, 2, \dots, m. \quad (4.2)$$

where $z_{i,j}$ is an independent standard normal pseudo-random number, and $\rho$ is the correlation among any two explanatory variables.

In this study, we have used 100 observations with 20 explanatory variables, in which 70 observations were used to fit the model and 30 observations were used to compute the RMSE.

A dependent variable is generated by using the following equation,

$$y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_{20} x_{i,20} + \varepsilon_i, \quad (4.3)$$

where $\varepsilon_i$ is a normal pseudo-random number with $E(\varepsilon_i) = 0$ and $V(\varepsilon_i) = \sigma^2 = 1$.

We choose $\beta = (\beta_1, \beta_2, \ldots, \beta_{20})$ as the normalized eigenvector corresponding to the largest eigenvalue of $\boldsymbol{X'X}$ for which $\boldsymbol{\beta'\beta} = 1$ (McDonald and Galarneau 1975). Prior information was defined based on Nagar and Kakwani's (1964) approach, which is described in Appendix B. Further, we have assumed the first four elements of OLSE estimates as $\boldsymbol{b}$ (see Appendix B). To study the effects of various degrees of multicollinearity on the data, we pick $\rho = (0.5,\ 0.7,\ 0.9)$, which signifies weak, moderate, and high multicollinearity, respectively. Since the execution time for the algorithm simulation is long due to cross-validation, we simulated only 50 datasets. Future simulation studies will be implemented with a greater number of simulations using cluster programming. Figures 1-3 and Tables 2-4 show the cross-validated RMSE and the median cross-validated RMSE of the SRGLARS algorithms, respectively, for 50 simulated data.

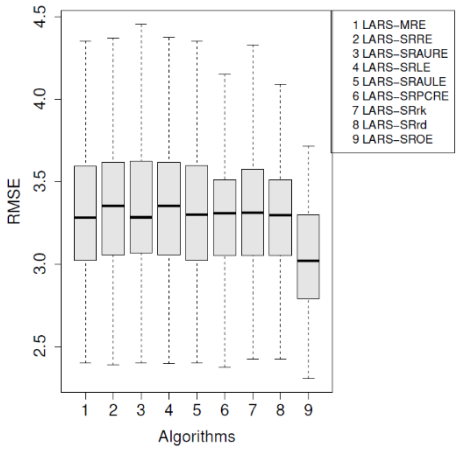



**Fig 1.** Cross-validated RMSE values of the SRGLARS algorithms when ρ = 0.5

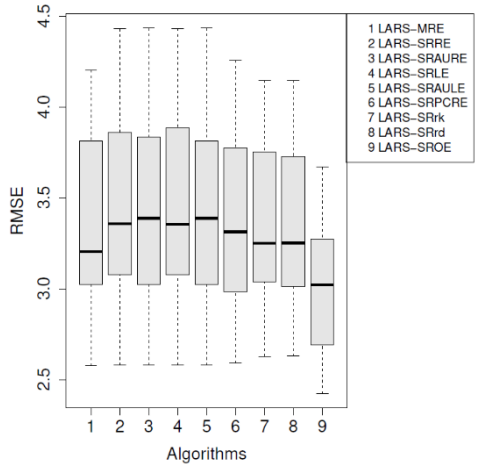



**Fig 2.** Cross-validated RMSE values of the SRGLARS algorithms when ρ = 0.7

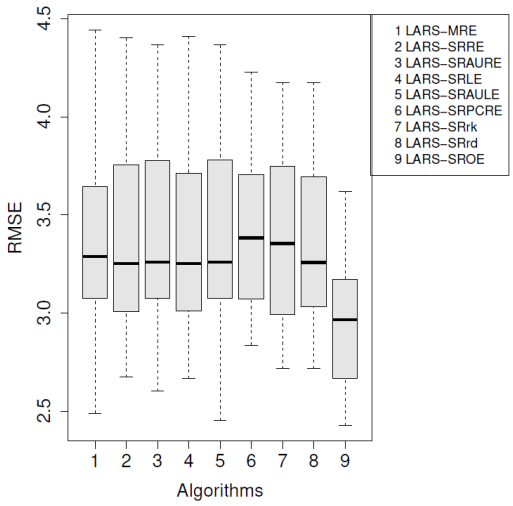



**Fig 3.** Cross-validated RMSE values of the SRGLARS algorithms when ρ = 0.9

Table 2: Median Cross-validated RMSE values of the SRGLARS algorithms when ρ = 0.5.

| **Algorithms** | **RMSE** | **(*k*, *d*)** | ***t*** | **Selected variables** |
|---|---|---|---|---|
| LARS-MRE | 3.2805 | – | 6.1630 | 15 |
| LARS-SRRE | 3.3531 | 0.3 | 6.1777 | 16 |
| LARS-SRAURE | 3.2832 | 0.1 | 6.1631 | 16 |
| LARS-SRLE | 3.3531 | 0.7 | 6.1773 | 16 |
| LARS-SRAULE | 3.3012 | 0.99 | 6.1630 | 16 |
| LARS-SRPCRE | 3.3070 | – | 6.4343 | 16 |
| LARS-SRrk | 3.3122 | 0.1 | 6.9493 | 17 |
| LARS-SRrd | 3.2976 | 0.9 | 6.9497 | 16 |
| LARS-SROE | 3.0182 | – | 4.1102 | 18 |

Table 3**:** Median Cross-validated RMSE values of the SRGLARS algorithms when ρ = 0.7.

| **Algorithms** | **RMSE** | **(*k*, *d*)** | ***t*** | **Selected variables** |
|---|---|---|---|---|
| LARS-MRE | 3.2041 | – | 6.8458 | 15 |
| LARS-SRRE | 3.3575 | 0.1 | 7.5043 | 16 |
| LARS-SRAURE | 3.3884 | 0.8 | 6.7785 | 16 |
| LARS-SRLE | 3.3537 | 0.9 | 7.5049 | 17 |
| LARS-SRAULE | 3.3886 | 0.2 | 6.7786 | 16 |
| LARS-SRPCRE | 3.3135 | – | 7.7712 | 16 |
| LARS-SRrk | 3.2525 | 1.0 | 8.5411 | 17 |
| LARS-SRrd | 3.2534 | 0.5 | 8.1097 | 17 |
| LARS-SROE | **3.0324** | – | 4.3864 | 18 |

Table 4: Median Cross-validated RMSE values of the SRGLARS algorithms when $\rho = 0.9$.

| Algorithms | RMSE | ($k$, $d$) | $t$ | Selected variables |
|---|---|---|---|---|
| LARS-MRE | 3.3391 | – | 11.6628 | 16 |
| LARS-SRRE | 3.2722 | 0.1 | 11.5897 | 17 |
| LARS-SRAURE | 3.3354 | 0.1 | 11.6606 | 16 |
| LARS-SRLE | 3.3110 | 0.99 | 11.6628 | 17 |
| LARS-SRAULE | 3.3107 | 0.99 | 11.6628 | 16 |
| LARS-SRPCRE | 3.3782 | – | 12.2288 | 16 |
| LARS-SRrk | 3.3531 | 0.7 | 12.9540 | 17 |
| LARS-SRrd | 3.3276 | 0.3 | 13.0038 | 17 |
| LARS-SROE | **2.9878** | – | 4.5654 | 18 |

From Figure 1-3, we can observe that the LARS-SROE algorithm outperformed other SRGLARS algorithms in RMSE criterion under all degrees of multicollinearity. From Tables 2-4, we observe that the SROE included more variables than the other SRGLARS algorithms, although it has minimum RMSE. Therefore, in practical situations, if a researcher intends to reduce the number of variables from the model, they may consider other SRGLARS algorithms based on the interested variables and prediction performance using the plot of coefficient paths discussed in the real-world example.

### 4.2 Real-world example

As a real-world data set, we considered the Prostate Cancer Data (Stamey *et al*. 1989), which is used by Tibshirani (1996), Efron *et al*. (2004) and Zou and Hastie (2005) to study the performance of LASSO, LARS algorithm and Enet. The Prostate Cancer Data contains 97 observations and 8 predictors such as log cancer volume (lcavol), log prostate weight (lweight), age, log of the amount of benign prostatic hyperplasia (lbph), seminal vesicle invasion (svi), log capsular penetration (lcp), Gleason score (gleason) and percentage Gleason score 4 or 5 (pgg45). The dependent variable is the log of prostate-specific antigen (lpsa). The Variance Inflation Factor (VIF) values of the predictor variables are 3.09, 2.97, 2.47, 2.05, 1.95, 1.37, 1.36 and 1.32, and the condition number is 243, which exposes high multicollinearity between the predictor variables. This data set is associated in "lasso2" R package. We have used 67 observations to fitting the model, and 30 observations to compute the RMSE. We have assumed (Nagar and Kakwani 1964) that the first three OLSE estimates of Prostate Cancer Data are unbiased, and we defined the prior information for this data based on Nagar and Kakwani's (1964) approach, as described in Appendix B.

The cross-validated RMSE of the SRGLARS algorithms are shown in Table 5, and coefficient paths of respective SRGLARS algorithm are shown in Figure 4 in Appendix A.

Table 5**:** Cross-validated RMSE values of Prostate Cancer Data using SRGLARS.

| Algorithms | RMSE | ($k$, $d$) | $t$ | Selected variables |
|---|---|---|---|---|
| LARS-MRE | 0.77784 | – | 1.5632 | 6 |
| LARS-SRRE | 0.73567 | 0.38 | 1.7329 | 8 |
| LARS-SRAURE | 0.77784 | 0.01 | 1.5632 | 6 |
| LARS-SRLE | 0.73566 | 0.62 | 1.7331 | 8 |
| LARS-SRAULE | 0.77784 | 0.99 | 1.5632 | 6 |
| LARS-SRPCRE | 0.74084 | – | 1.5656 | 7 |
| LARS-SRrk | 0.74083 | 0.04 | 1.5643 | 7 |
| LARS-SRrd | 0.74084 | 0.96 | 1.5643 | 7 |
| LARS-SROE | **0.69897** | – | 1.1226 | 8 |

Table 5 shows that the LARS-SROE algorithm outperforms other algorithms on Prostate Cancer Data, which is agreed the results obtained in the simulation study. Further, we can note that LARS-SRRE, LARS-SRLE, and LARS-SROE did not make any variable selections on Prostate Cancer Data.

Figure 4 (see Appendix A) shows that the choice of variables is different for each SRGLARS algorithms. Note that in all graphs, the order of selection of variables is the same except Figure 4(i). Further, the coefficients relevant to the variables lcp, age and gleason are very small for Fig 4((a)-(h)) and those coefficients are almost zero in Fig 4(i). This indicates that the contribution from those three variables is very low for the fitted model.

## 5 Conclusions

This study presented a generalized version of stochastic restricted LARS (SRGLARS) algorithm to combine LASSO with existing stochastic restricted estimators. We have shown the superiority of SRGLARS when it combines LASSO with SROE (LARS-SROE) using a Monte Carlo simulation and a real-world example in RMSE criterion. However, LARS-SROE is less likely to make variable selections when looking at the results directly. However, by examining the size of the coefficients as shown in the example, we can drop some irrelevant variables form the model. If a researcher wanted to reduce some variables from the model, the other combinations of the SRGLARS algorithm could be considered to fit the regression model by studying the RMSE and coefficients estimates based on the coefficients path.

**Acknowledgements**

Anonymous reviewers are acknowledged for their comments on the initial manuscript.

## References


Akdeniz F, Kaçiranlar S. 1995. On the almost unbiased generalized liu estimator and unbiased estimation of the bias and mse. *Communication in Statistics - Theory and Methods* 24(7):1789–1797.

Baye MR, Parker DF. 1984. Combining ridge and principal component regression: A money demand illustration. *Communications in Statistics-Theory and Methods* 13(2):197–205.

Efron B, Hastie T, Johnstone I, Tibshirani R. 2004. Least angle regression. *The Annals of statistics* 32(2):407–499.

He D, Wu Y. 2014. A stochastic restricted principal components regression estimator in the linear model. *The Scientific World Journal*, vol. 2014, Article ID 231506, 6p. https://doi.org/10.1155/2014/231506

Hoerl E, Kennard RW. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12:55–67.

Hubert MH, Wijekoon P. 2006. Improvement of the Liu estimator in linear regression model. *Statistical Papers* 47:471–479.

Jibo W. 2014. On the Stochastic Restricted r-k Class Estimator and Stochastic Restricted r-d Class Estimator in Linear Regression Model. *Journal of Applied Mathematics*, vol. 2014, Article ID 173836, 6 p. https://doi.org/10.1155/2014/173836

Jibo W, Hu Y. 2014. On the stochastic restricted almost unbiased estimators in linear regression model. *Communications in Statistics-Simulation and Computation* 43(2):428–440.

Kaçiranlar S, Sakallıoğlu S. 2001. Combining the Liu estimator and the principal component Regression Estimator. *Communications in Statistics-Theory and Methods* 30(12):2699–2705.

Kayanan M, Wijekoon P. 2018. Stochastic Restricted Biased Estimators in Misspecified Regression Model with Incomplete Prior Information. *Journal of Probability and Statistics*, vol. 2018, Article ID 1452181, 8 p. https://doi.org/10.1155/2018/1452181.

Kayanan M, Wijekoon P. 2019. Optimal estimators in misspecified linear regression model with an application to real-world data. *Communications in Statistics: Case Studies, Data Analysis and Applications* 4(3-4):151–163.

Kayanan M, Wijekoon P. 2020a. Stochastic Restricted LASSO-Type Estimator in the Linear Regression Model. *Journal of Probability and Statistics*, vol. 2020, Article ID 7352097, 7 p. https://doi.org/10.1155/2020/7352097.

Kayanan M, Wijekoon P. 2020b. Variable Selection via Biased Estimators in the Linear Regression Model. *Open Journal of Statistics* 10(1):113–126.

Li Y, Yang H. 2010. A new stochastic mixed ridge estimator in linear regression model. *Statistical Papers* 51:315–323.

Liu K. 1993. A new class of biased estimate in linear regression. *Communication in Statistics - Theory and Methods* 22(2):393–402.

Massy WF. 1965. Principal components regression in exploratory statistical research. *Journal of the American Statistical Association* 60:234–266.

McDonald GC, Galarneau DI. 1975. A Monte Carlo Evaluation of Some Ridge-Type Estimators. *Journal of the American Statistical Association* 70:407–416.

Nagar AL, Kakwani NC. 1964. The Bias and Moment Matrix of a Mixed Regression Estimator. *Econometrica* 32:174–182.

Singh B, Chaubey YP, Dwivedi TD. 1986. An Almost Unbiased Ridge Estimator. *The Indian Journal of Statistics* 48:342–346.

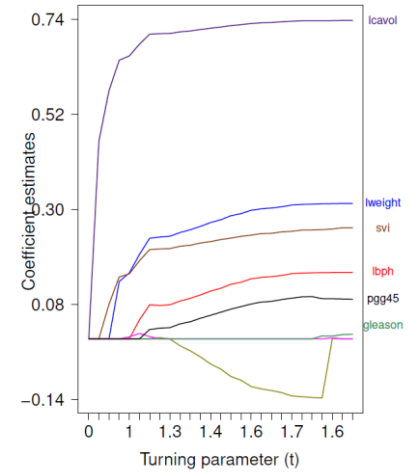
Stamey TA, Kabalin JN, Mcneal JE, Johnstone IM, Freiha F, Redwine EA, Yang N. 1989. Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate: II. radical prostatectomy treated patients. *Journal of Urology* 141:1076–1083.

Theil H, Goldberger AS. 1961. On Pure and Mixed Statistical Estimation in Economics. *International Economic Review* 2:65–78.

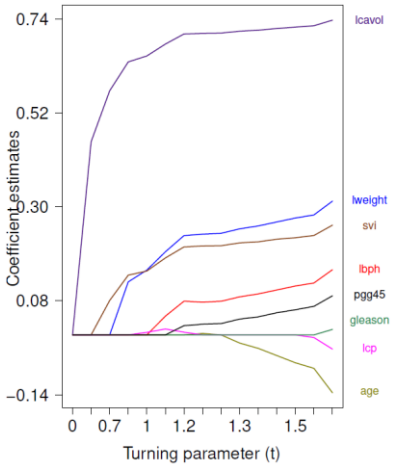
Tibshirani R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58:267–288.

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

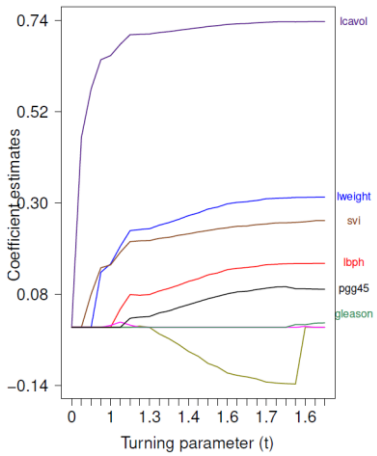
## Appendix A: Figure 4



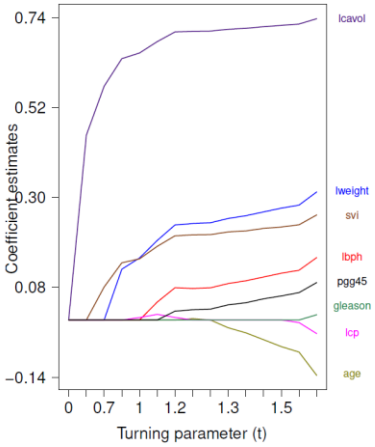

(*a*)



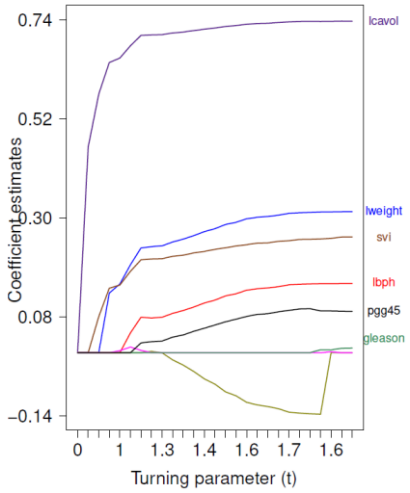

(*b*)



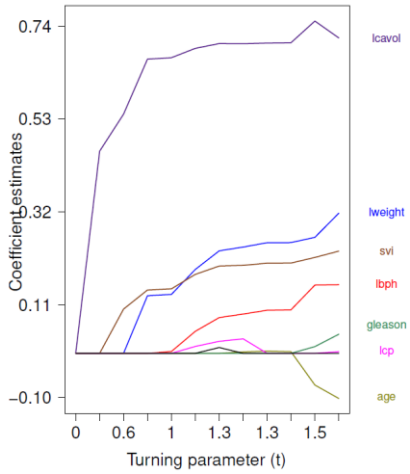

(*c*)



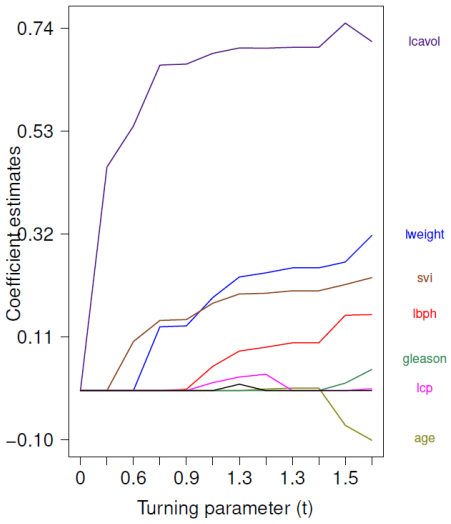

(*d*)


0.74
0.52
0.30
0.08
−0.14
Coefficient estimates
0 1 1.3 1.4 1.6 1.7 1.6
Turning parameter (t)
lcavol
lweight
svi
lbph
pgg45
gleason

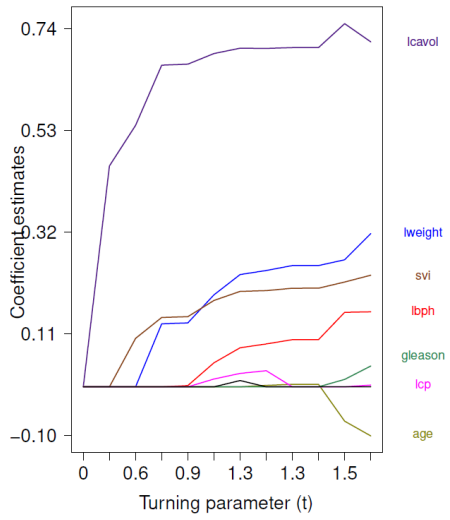
*(e)*


0.74
0.53
0.32
0.11
−0.10
Coefficient estimates
0 0.6 1 1.3 1.3 1.5
Turning parameter (t)
lcavol
lweight
svi
lbph
gleason
lcp
age

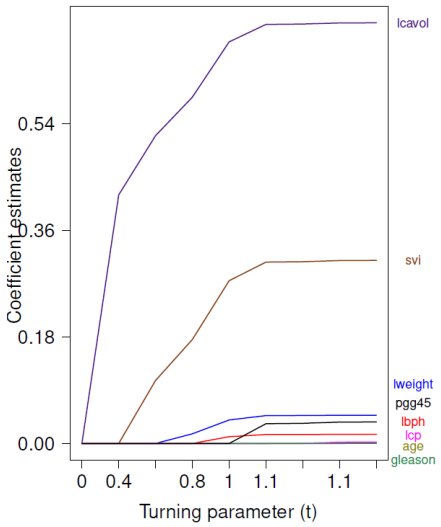
*(f)*


0.74
0.53
0.32
0.11
−0.10
Coefficient estimates
0 0.6 0.9 1.3 1.3 1.5
Turning parameter (t)
lcavol
lweight
svi
lbph
gleason
lcp
age

(*g*)


0.74
0.53
0.32
0.11
−0.10
Coefficient estimates
0 0.6 0.9 1.3 1.3 1.5
Turning parameter (t)
lcavol
lweight
svi
lbph
gleason
lcp
age

(*h*)




(*i*)

**Fig 4.** Coefficient paths of the (a) LARS-MRE, (b) LARS-SRRE, (c) LARS- SRAURE, (d) LARS-SRLE, (e) LARS-SRAULE, (f) LARS-SRPCRE, (g) LARS-SRrk (h) LARS-SRrd and (i) LARS-SROE versus $\boldsymbol{t} = \sum_{\boldsymbol{j=1}}^{\boldsymbol{p}} \left|\boldsymbol{\beta_j}\right|$ for the Prostate Cancer Data.

**Appendix B: Selection of prior information**

According to Nagar and Kakwani (1964), we can define the prior information as follows:

Let $\boldsymbol{\beta}_1$ be a vector of some selected $q$ elements of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_2$ is the rest of elements. Assume that $\boldsymbol{b}$ is the known unbiased estimates of $\boldsymbol{\beta}_1$. By using the "two sigma rule", now we can write the range of $\boldsymbol{\beta}_1$ as $\boldsymbol{b} \pm 2SE(\boldsymbol{b})$. Based on that we can set the expressions of equation (2.2) as

$$\hat{\boldsymbol{r}} = \boldsymbol{b}, \quad \widehat{\boldsymbol{R}} = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \cdots & 0 \\ 0 & 0 & \cdots & 1 & \cdots & 0 \end{pmatrix}_{q \times p}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \text{ and}$$

$$\sigma^2 \widehat{\boldsymbol{W}} = \begin{pmatrix} SE(b_1) & 0 & \cdots & 0 \\ 0 & SE(b_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & SE(b_q) \end{pmatrix}_{q \times q}.$$

**Appendix C: K-fold cross-validation to estimate shrinkage parameters, (k, d)**

**Step 1**: Split the data set into $K$ groups.

**Step 2**: For each unique group $i = 1, 2, \ldots, K$:

- Take one group as a test data set
- Take the remaining $K - 1$ groups as a training data set
- Estimate the respective estimator $\widehat{\boldsymbol{\beta}}(k, d)$ with shrinkage parameters $(k, d)$ using the training data set. Use intital values of $k, d$ as 0.01, and then compute the RMSE cross-validation errors separately as follows:

$$CV\_RMSE_i(\widehat{\boldsymbol{\beta}}(k,d)) = \sqrt{\frac{K}{n} \sum_{i=1}^{n/K} \left( y_i^* - \sum_{j=1}^{p} x_{ij}^* \hat{\beta}_j(k,d) \right)^2},$$

where $y_i^*$ and $x_{ij}^*$ are the variables belongs to the test data set.

**Step 3**: Then, find the overall cross-validation errors as follows:

$$CV\_RMSE\,(\widehat{\boldsymbol{\beta}}(k,d)) = \frac{1}{K} \sum_{i=1}^{K} CV\_RMSE_i\,(\widehat{\boldsymbol{\beta}}(k,d))$$

**Step 4**: Continue this procedure by increasing the values of $k$ or $d$ by a small increment, and choose the value of $k$ or $d$ that makes $CV\_RMSE\,(\widehat{\boldsymbol{\beta}}(k,d))$ smallest.