


Research Article

Stochastic Restricted LASSO-Type Estimator in the Linear Regression Model

Manickavasagar Kayanan ^{1,2} and Pushpakanthie Wijekoon³

¹Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka

²Department of Physical Science, Vavuniya Campus of the University of Jaffna, Vavuniya, Sri Lanka

³Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka

Correspondence should be addressed to Manickavasagar Kayanan; mgayanan@vau.jfn.ac.lk

Received 13 November 2019; Revised 9 February 2020; Accepted 29 February 2020; Published 30 March 2020

Academic Editor: Yaozhong Hu

Copyright © 2020 Manickavasagar Kayanan and Pushpakanthie Wijekoon. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Among several variable selection methods, LASSO is the most desirable estimation procedure for handling regularization and variable selection simultaneously in the high-dimensional linear regression models when multicollinearity exists among the predictor variables. Since LASSO is unstable under high multicollinearity, the elastic-net (Enet) estimator has been used to overcome this issue. According to the literature, the estimation of regression parameters can be improved by adding prior information about regression coefficients to the model, which is available in the form of exact or stochastic linear restrictions. In this article, we proposed a stochastic restricted LASSO-type estimator (SRLASSO) by incorporating stochastic linear restrictions. Furthermore, we compared the performance of SRLASSO with LASSO and Enet in root mean square error (RMSE) criterion and mean absolute prediction error (MAPE) criterion based on a Monte Carlo simulation study. Finally, a real-world example was used to demonstrate the performance of SRLASSO.

1. Introduction

Let us consider the linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (1)$$

where \mathbf{y} is the $n \times 1$ vector of observations on the dependent variable, \mathbf{X} is the $n \times p$ matrix of observations on the non-stochastic predictor variables, $\boldsymbol{\beta}$ is a $p \times 1$ vectors of unknown coefficients, and $\boldsymbol{\epsilon}$ is the $n \times 1$ vector of random error terms, which is independent and identically normally distributed with the mean zero and common variance σ^2 , that is, $\mathbf{E}(\boldsymbol{\epsilon}) = 0$ and $\mathbf{E}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}') = \boldsymbol{\Omega} = \sigma^2\mathbf{I}$.

It is well-known that ordinary least square estimator (OLSE) is a best linear unbiased estimator for model (1), and it is defined as

$$\hat{\boldsymbol{\beta}}_{\text{OLSE}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (2)$$

Furthermore, researchers [1, 2] have shown that parameter estimation is improved when prior information for

regression coefficients is available, which can be in the form of exact linear restrictions or stochastic linear restrictions. Let us assume that there exists prior information on $\boldsymbol{\beta}$ in the form of stochastic linear restriction as

$$\boldsymbol{\phi} = \mathbf{R}\boldsymbol{\beta} + \mathbf{v}, \quad (3)$$

where $\boldsymbol{\phi}$ is the $q \times 1$ vector, \mathbf{R} is the $q \times p$ matrix with rank q , and \mathbf{v} is the $q \times 1$ vector of disturbances, such that $\mathbf{E}(\mathbf{v}) = 0$, $D(\mathbf{v}) = \mathbf{E}(\mathbf{v}\mathbf{v}') = \boldsymbol{\Psi} = \sigma^2\mathbf{W}$ (\mathbf{W} is positive definite) and $\mathbf{E}(\mathbf{v}\boldsymbol{\epsilon}') = 0$. Note that equation (3) will be the exact linear restriction when $\mathbf{v} = 0$.

Theil and Goldberger [2] proposed the mixed regression estimator (MRE) by combining the models (1) and (3), and it is defined as

$$\hat{\boldsymbol{\beta}}_{\text{MRE}} = (\mathbf{X}'\mathbf{X} + \mathbf{R}'\mathbf{W}^{-1}\mathbf{R})^{-1}(\mathbf{X}'\mathbf{y} + \mathbf{R}'\mathbf{W}^{-1}\boldsymbol{\phi}). \quad (4)$$

According to the literature, it was noted that OLSE and MRE are unstable when the number of predictors is high. In

this case, the variable selection methods such as forward selection, backward selection, and step-wise selection have been used. However, these methods are also unstable when multicollinearity exists among predictor variables. As a remedial solution to this problem, Tibshirani [3] proposed the least absolute shrinkage and selection operator (LASSO) by considering the model (1) to handle both multicollinearity and variable selection simultaneously in the high-dimensional linear regression model. The LASSO estimator is defined as

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\text{LASSO}} &= \arg \min_{\boldsymbol{\beta}} \{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\} \\ \text{subject to } &\sum_{j=1}^p |\beta_j| \leq t, \end{aligned} \quad (5)$$

where $t \geq 0$ is a turning parameter. The solutions of LASSO can be obtained using either the standard quadratic programming technique or least angle regression (LARS) [4] algorithm. According to Zou et al. [5, 6], LASSO is unstable when high multicollinearity exists among predictor variables. Therefore, they proposed elastic-net (Enet) estimator as an alternative to LASSO to handle this issue. The Enet estimator is defined as

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\text{Enet}} &= \arg \min_{\boldsymbol{\beta}} \left\{ (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + k \sum_{j=1}^p \beta_j^2 \right\} \\ \text{subject to } &\sum_{j=1}^p |\beta_j| \leq t. \end{aligned} \quad (6)$$

The Enet solutions can be obtained using LARS-EN algorithm, which is the modified version of LARS algorithm.

Norouzirad et al. [7] and Tuuç and Arslan [8] attempted to combine LASSO with the exact linear restriction, and their work did not well define how the exact restriction is incorporated since there is no analytical solution for LASSO. In this article, we proposed stochastic restricted LASSO-type estimator (SRLASSO) by combining LASSO and stochastic restrictions. Furthermore, we compared the performance of SRLASSO with LASSO and Enet in root mean square error (RMSE) criterion and mean absolute prediction error (MAPE) criterion using a Monte Carlo simulation study and a real-world example. The structure of the rest of the article is as follows: Section 2 describes SRLASSO and algorithm to find SRLASSO solutions, Section 3 shows the performance of SRLASSO, and Section 4 concludes the article and references provided at the end of the paper.

2. Stochastic Restricted LASSO-Type Estimator (SRLASSO)

By considering equation (3) as an additional constrain, we define stochastic restricted LASSO-type estimator (SRLASSO) as

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\text{SRLASSO}} &= \arg \min_{\boldsymbol{\beta}} \{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\} \\ \text{subject to } &\sum_{j=1}^p |\beta_j| \leq t, \\ &\mathbf{R}\boldsymbol{\beta} = \boldsymbol{\phi} - \mathbf{v}. \end{aligned} \quad (7)$$

We can view this as a quadratic optimization problem. Here, we have $2^p + q$ constraints. However, this method is not suitable in a practical situation if p is large. So, we propose a stochastic restricted LARS (SRLARS) algorithm, which is the modified version of the LARS algorithm, to find SRLASSO solutions. In SRLARS, we consolidate MRE with LARS.

2.1. Stochastic Restricted LARS (SRLARS). Standardize the predictor variables \mathbf{X} to have a mean of zero and a standard deviation of one, and response variable \mathbf{y} to have a mean zero. Let residuals of the models (1) and (3) be $\mathbf{r} = \hat{\boldsymbol{\epsilon}}$ and $\boldsymbol{\tau} = \hat{\mathbf{v}}$, respectively.

Step 1. Start with $\hat{\boldsymbol{\beta}} = \mathbf{0}$, $\mathbf{r}_0 = \mathbf{y}$, and $\boldsymbol{\tau}_0 = \boldsymbol{\phi}$.

Step 2. Find the predictor \mathbf{X}_{j_1} most correlated with \mathbf{r}_0 as follows:

$$\mathbf{X}_{j_1} = \max_j |\text{Cor}(\mathbf{X}_j, \mathbf{r}_0)|; \quad j = 1, 2, \dots, p. \quad (8)$$

Let $\hat{\beta}_{j_1}$ be the regression coefficient of \mathbf{X}_{j_1} . Then, increase the estimate of $\hat{\beta}_{j_1}$ from 0 toward MRE until any other predictor \mathbf{X}_{j_2} has a high correlation with the current residual as \mathbf{X}_{j_1} does. At this point, SRLARS proceeds in the equiangular direction between the two predictors \mathbf{X}_{j_1} and \mathbf{X}_{j_2} instead of continuing in the direction based on \mathbf{X}_{j_1} .

In a similar way, i^{th} variable \mathbf{X}_{j_i} eventually earns its way into the active set, and then SRLARS proceeds in the equiangular direction between \mathbf{X}_{j_1} , \mathbf{X}_{j_2} , \dots , \mathbf{X}_{j_i} . Continue adding variables to the active set in this way moving in the direction defined by the least angle direction. In the intermediate steps, the coefficient estimates are updated using the following formula:

$$\hat{\boldsymbol{\beta}}_{j_i} = \hat{\boldsymbol{\beta}}_{j_{(i-1)}} + \alpha_i \mathbf{u}_i, \quad (9)$$

where α_i is a value between 0 and 1 which represents how far the estimate moves in the direction before another variable enters the model and the direction changes again, and \mathbf{u}_i is the equiangular vector.

The direction \mathbf{u}_i is calculated using the following formula based on MRE:

$$\mathbf{u}_i = \mathbf{E}_i (\mathbf{E}_i' (\mathbf{X}'\mathbf{X} + \mathbf{R}'\mathbf{W}^{-1}\mathbf{R}) \mathbf{E}_i)^{-1} \mathbf{E}_i' (\mathbf{X}'\mathbf{r}_{i-1} + \mathbf{R}'\mathbf{W}^{-1}\boldsymbol{\tau}_{i-1}), \quad (10)$$

where \mathbf{E}_i is the matrix with column $(e_{j_1}, e_{j_2}, \dots, e_{j_i})$ and e_j be the j^{th} standard unit vector in \mathbb{R}^p , which has the index of variables selected in each subsequent step.

Then, α_i is calculated as follows:

$$\alpha_i = \min\{\alpha_{ji}^+, \alpha_{ji}^-, \alpha_{ji}^*\} \in [0, 1], \quad (11)$$

where

$$\alpha_{ji}^\pm = \frac{\text{Cor}(\mathbf{r}_{i-1}, \mathbf{X}_{ji}) \pm \text{Cor}(\mathbf{r}_{i-1}, \mathbf{X}_j)}{\text{Cor}(\mathbf{r}_{i-1}, \mathbf{X}_{ji}) \pm \text{Cor}(\mathbf{X}\mathbf{u}_i, \mathbf{X}_j)}, \quad (12)$$

for any j such that $\hat{\beta}_{j(i-1)} = 0$ and

$$\alpha_{ji}^* = \frac{\hat{\beta}_{j(i-1)}}{\mathbf{u}_i}, \quad (13)$$

for any j such that $\hat{\beta}_{j(i-1)} \neq 0$.

Step 3. If $\alpha_i = \alpha_{ji}^*$, then \mathbf{E}_i is the matrix formed by removing the column e_j from \mathbf{E}_{i-1} . Then \mathbf{r}_i and $\boldsymbol{\tau}_i$ related to the current step is calculated as

$$\begin{aligned} \mathbf{r}_i &= \mathbf{r}_{i-1} - \alpha_i \mathbf{X}\mathbf{u}_i, \\ \boldsymbol{\tau}_i &= \boldsymbol{\tau}_{i-1} - \alpha_i \mathbf{R}\mathbf{u}_i, \end{aligned} \quad (14)$$

and then move to the next step where j_{i+1} is the value of j such that $\alpha_i = \alpha_{j_i}^+$ or $\alpha_i = \alpha_{j_i}^-$ or $\alpha_{j_i}^*$.

Step 4. Proceed Step 2 until $\alpha_i = 1$.

2.2. Properties of SRLARS. SRLARS algorithm sequentially updates the SRLASSO estimates. It requires $O(m^3 + pm^2)$ operations, where m is the number of steps. The prediction performance of SRLARS is evaluated using the RMSE criterion and MAPE criterion, which is described in Section 3. According to Efron et al. [4], the conventional tuning parameter is $t = \sum_{j=1}^p |\hat{\beta}_j|$, and the suitable value of turning parameter t for the particular problem is selected using K-fold cross-validation.

2.3. Selection of Prior Information. According to Nagar and Kakwani [9], we can define the prior information as follows: Let $\boldsymbol{\beta}_1$ be a vector of some selected q elements of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_2$ be the rest of elements. Assume that \mathbf{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 $\mathbf{b} \pm 2\text{SE}(\mathbf{b})$. Based on that, we can set the expressions of equation (3) as $\hat{\mathbf{r}} = \mathbf{b}$,

$$\hat{\mathbf{R}} = \begin{pmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & \dots & 0 \end{pmatrix}_{q \times p}, \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}, \quad \text{and}$$

$$\sigma^2 \hat{\mathbf{W}} = \begin{pmatrix} \text{SE}(\mathbf{b}_1) & 0 & \dots & 0 \\ 0 & \text{SE}(\mathbf{b}_2) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \text{SE}(\mathbf{b}_l) \end{pmatrix}_{q \times q}.$$

3. Performance of SRLASSO

SRLASSO is compared with LASSO and Enet using the RMSE criterion and MAPE criterion, which are the expected prediction errors of the algorithms, and are defined as

$$\text{RMSE}(\hat{\boldsymbol{\beta}}) = \sqrt{\frac{1}{n} (\mathbf{y}_{\text{new}} - \mathbf{X}_{\text{new}} \hat{\boldsymbol{\beta}})' (\mathbf{y}_{\text{new}} - \mathbf{X}_{\text{new}} \hat{\boldsymbol{\beta}})}, \quad (15)$$

$$\text{MAPE}(\hat{\boldsymbol{\beta}}) = \frac{1}{n} \sum |y_{\text{new}} - \mathbf{X}_{\text{new}} \hat{\boldsymbol{\beta}}|,$$

where $(\mathbf{y}_{\text{new}}, \mathbf{X}_{\text{new}})$ denotes the new data which are not used to obtain the parameter estimates, n is the number of new observations, and $\hat{\boldsymbol{\beta}}$ is the estimated value of $\boldsymbol{\beta}$ using the respective algorithm. A Monte Carlo simulation study and a real-world example are used for the comparison.

3.1. Simulation Study. According to McDonald and Galarnau [10], first we generate the predictor variables by using the following formula:

$$\mathbf{x}_{i,j} = \sqrt{(1 - \rho^2)} z_{i,j} + \rho z_{i,m+1}; \quad i = 1, 2, \dots, n, j = 1, 2, \dots, m, \quad (16)$$

where $z_{i,j}$ is an independent standard normal pseudo random number and ρ is the theoretical correlation between any two explanatory variables.

In this study, we have used a linear regression model of 100 observations and 20 predictors. A dependent variable is generated by using the following equation:

$$y_i = \boldsymbol{\beta}_1 \mathbf{x}_{i,1} + \boldsymbol{\beta}_2 \mathbf{x}_{i,2} + \dots + \boldsymbol{\beta}_5 \mathbf{x}_{i,20} + \epsilon_i; \quad i = 1, 2, \dots, 100, \quad (17)$$

where ϵ_i is a normal pseudo random number with a mean zero and common variance σ^2 .

We choose $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_{20})$ as the normalized eigenvector corresponding to the largest eigenvalue of $\mathbf{X}'\mathbf{X}$ for which $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$. To define the prior information according to Section 2.3, we assume that OLSE estimates of the first four elements of $\boldsymbol{\beta}$ are unbiased, which are the estimates of \mathbf{b} . To investigate the effects of different degrees of multicollinearity on the estimators, we choose $\rho = (0.5, 0.7, 0.9)$, which represents weak, moderated, and high multicollinearity. For the analysis, we have simulated 50 data sets consisting of 50 observations to fit the model and 50 observations to calculate the RMSE and MAPE. The cross-validated RMSE and MAPE of the estimators are displayed in Figure 1 and Figure 2, respectively. The median cross-validated RMSE and MAPE of the estimators are displayed in Table 1.

From Figures 1 and 2 and Table 1, we can observe that SRLASSO always shows better performance compared to LASSO and Enet in both RMSE criterion and MAPE criterion under all degrees of multicollinearity.

3.2. Real-World Example. As a numerical example, the well-known Prostate Cancer Data [11] was used to compare the performance of SRLASSO. This data set is attached with ‘‘lasso2’’ R package. In the Prostate Cancer Data, the predictors are the following eight clinical measures: 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

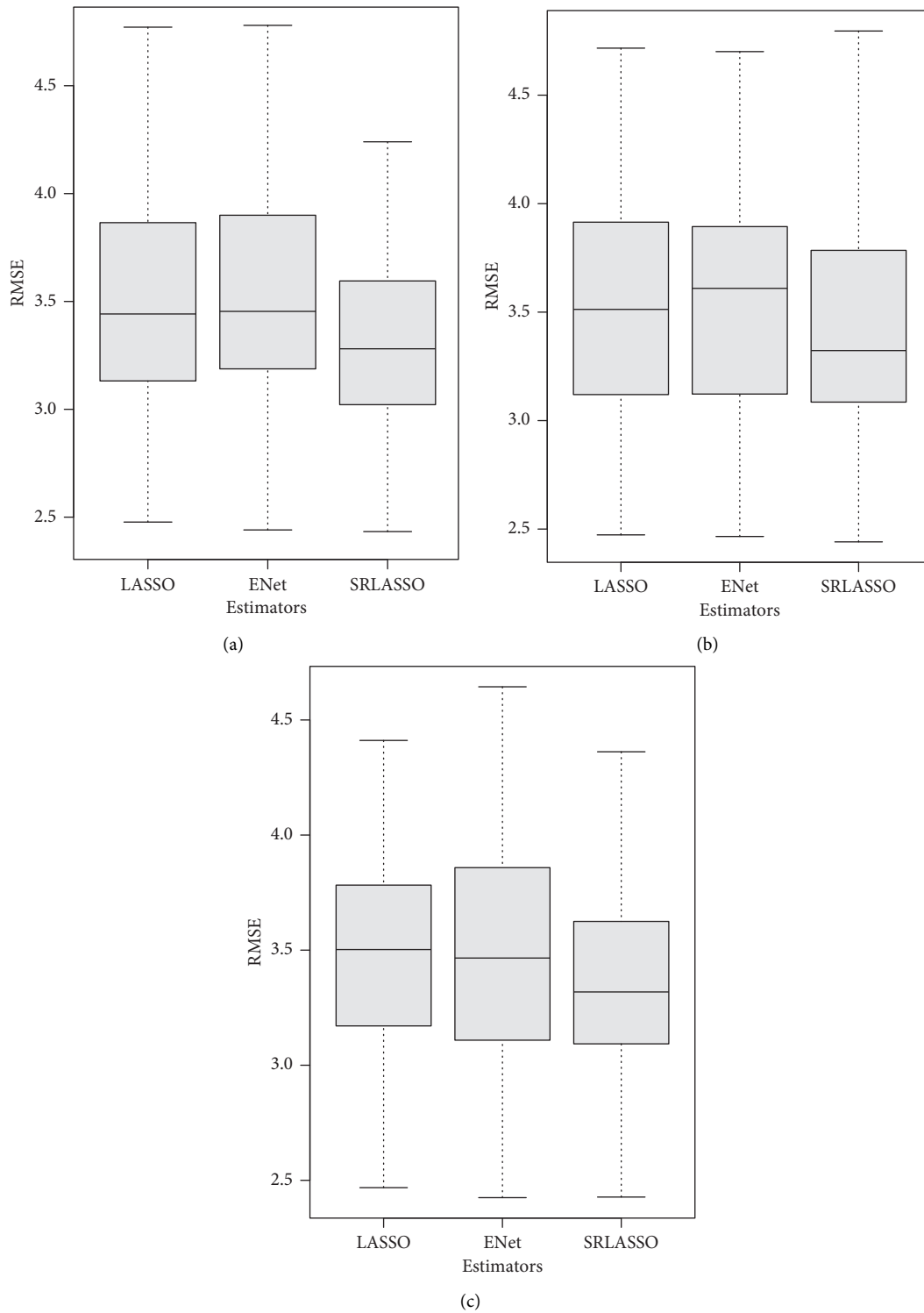


FIGURE 1: Cross-validated RMSE values of the estimators when (a) $\rho = 0.5$, (b) $\rho = 0.7$, and (c) $\rho = 0.9$.

score (gleason), and percentage Gleason score 4 or 5 (pgg45). The response is the log of prostate specific antigen (lpsa), and the dataset has 97 observations. The variance inflation factor (VIF) values of the predictor variables of the dataset 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 shows evidence of multicollinearity among the predictor variables. Stamey et al. [11] have examined the correlation between the level of prostate specific antigen with those eight clinical measures. Furthermore, Tibshirani [3] and Tibshirani et al. [4] have used this data to

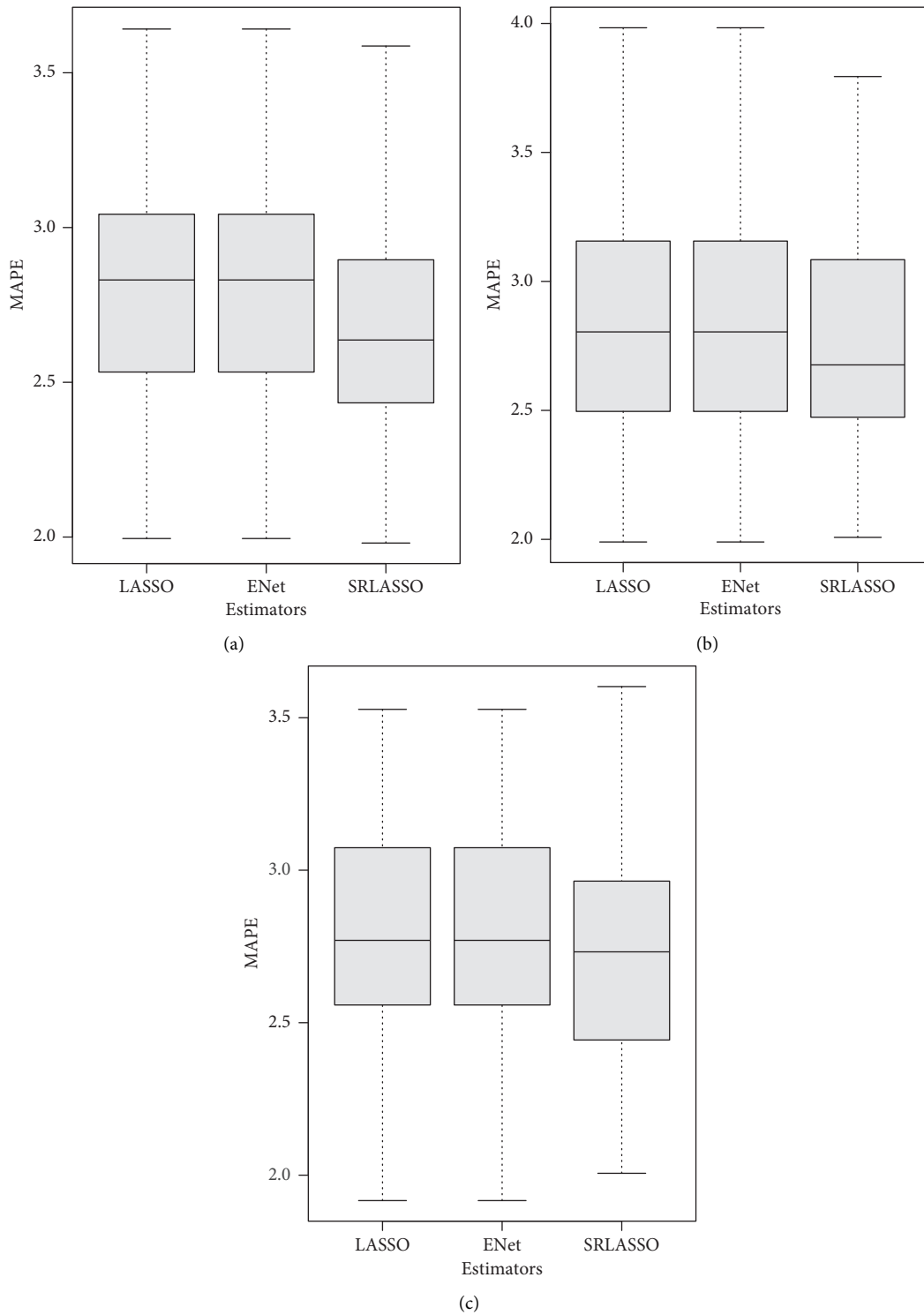


FIGURE 2: Cross-validated MAPE values of the estimators when (a) $\rho = 0.5$, (b) $\rho = 0.7$, and (c) $\rho = 0.9$.

examine the performance of LASSO and LARS algorithm, respectively. We have used 67 observations to fit the model and 30 observations to calculate RMSE and MAPE. We assume that OLSE estimates of the first three regression coefficients of Prostate Cancer Data are unbiased and we defined the prior information for this data based on Section 2.3.

The cross-validated RMSE and MAPE of the estimators are displayed in Table 2, and coefficient paths of each estimator are displayed in Figure 3.

From Table 2, we can observe that SRLASSO outperforms LASSO and Enet on Prostate Cancer Data in both RMSE criterion and MAPE criterion. Furthermore, we can

TABLE 1: Median cross-validated RMSE and MAPE values.

	Estimators	RMSE	MAPE	t	k	Selected variables
$\rho = 0.5$	LASSO	3.440	2.830	7.60	—	16
	Enet	3.454	2.854	7.14	0.01	17
	SRLASSO	3.280	2.63	6.16	—	15
$\rho = 0.7$	LASSO	3.512	2.804	8.66	—	15
	Enet	3.609	2.879	8.95	0.01	17
	SRLASSO	3.322	2.676	6.43	—	15
$\rho = 0.9$	LASSO	3.502	2.769	13.48	—	15
	Enet	3.466	2.814	14.47	0.01	16
	SRLASSO	3.318	2.732	11.66	—	15

TABLE 2: Cross-validated RMSE and MAPE values of Prostate Cancer Data.

Estimators	RMSE	MAPE	t	k
LASSO	0.8567	0.6496	1.2263	—
Enet	0.8063	0.6010	1.2897	0.01
SRLASSO	0.7833	0.5803	1.0415	—

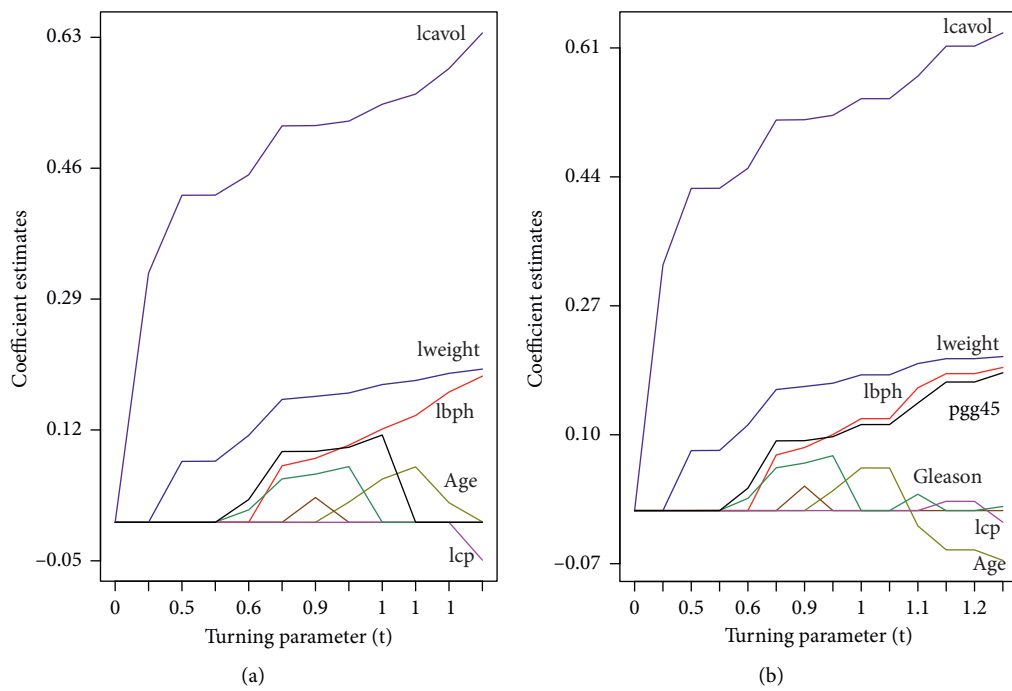


FIGURE 3: Continued.

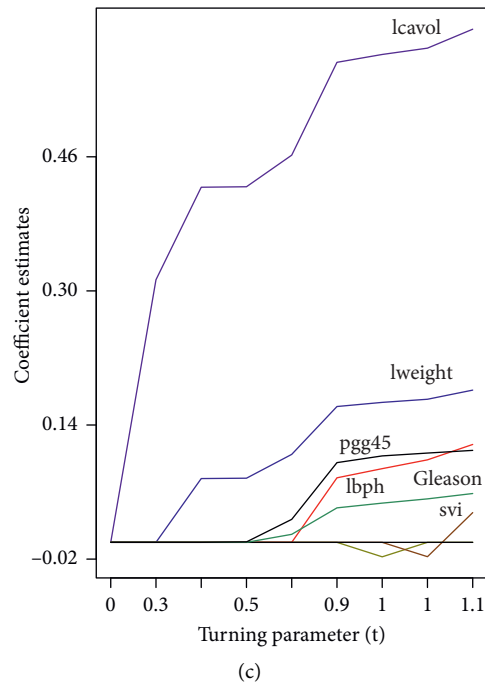


FIGURE 3: Coefficient paths of the (a) LASSO, (b) Enet, and (c) SRLASSO versus $t = \sum_{j=1}^p |\hat{\beta}_j|$ for the Prostate Cancer Data.

note that the selection of variables is different for each estimator by comparing Figures 3(a)–3(c).

4. Conclusions

This study clearly showed that SRLASSO does a better performance than LASSO and Enet in both RMSE criterion and MAPE criterion when multicollinearity exists among the predictor variables. Therefore, SRLASSO can be used as an alternative estimator of LASSO and Enet if prior information is accessible on the regression coefficients. The proposed SRLARS algorithm can be used to obtain SRLASSO solutions.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] J. Durbin, "A note on regression when there is extraneous information about one of the coefficients," *Journal of the American Statistical Association*, vol. 48, no. 264, pp. 799–808, 1953.
- [2] H. Theil and A. S. Goldberger, "On pure and mixed statistical estimation in economics," *International Economic Review*, vol. 2, no. 1, pp. 65–78, 1961.
- [3] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [4] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [5] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.
- [6] Z. Y. Algamal, M. H. Lee, A. M. Al-Fakih, and M. Aziz, "High-dimensional QSAR prediction of anticancer potency of imidazo 4, 5-b pyridine derivatives using adjusted adaptive LASSO," *Journal of Chemometrics*, vol. 29, no. 10, pp. 547–556, 2015.
- [7] M. Norouzirad, M. Arashi, and A. K. Saleh, "Restricted LASSO and double shrinking," 2015, <https://arxiv.org/abs/1505.02913>.
- [8] Y. Tuuç and O. Arslan, "Variable selection in restricted linear regression models," 2017, <https://arxiv.org/abs/1710.04105>.
- [9] A. L. Nagar and N. C. Kakwani, "The bias and moment matrix of a mixed regression estimator," *Econometrica*, vol. 32, no. 1/2, pp. 174–182, 1964.
- [10] G. C. McDonald and D. I. Galarneau, "A Monte Carlo evaluation of some ridge-type estimators," *Journal of the American Statistical Association*, vol. 70, no. 350, pp. 407–416, 1975.
- [11] T. A. Stamey, J. N. Kabalin, J. E. McNeal et al., "Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate. II. Radical prostatectomy treated patients," *Journal of Urology*, vol. 141, no. 5, pp. 1076–1083, 1989.