

# Variable Selection via Biased Estimators in the Linear Regression Model

Manickavasagar Kayanan<sup>1,2</sup>, Pushpakanthie Wijekoon<sup>3</sup>

<sup>1</sup>Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka

<sup>2</sup>Department of Physical Science, Vavuniya Campus of the University of Jaffna, Vavuniya, Sri Lanka

<sup>3</sup>Department of Statistics and Computer Science, University of Peradeniya, Peradeniya, Sri Lanka

Email: mgayanan@vau.jfn.ac.lk, pushpaw@pdn.ac.lk

**How to cite this paper:** Kayanan, M. and Wijekoon, P. (2020) Variable Selection via Biased Estimators in the Linear Regression Model. *Open Journal of Statistics*, 10, 113-126.

<https://doi.org/10.4236/ojs.2020.101009>

**Received:** January 26, 2020

**Accepted:** February 25, 2020

**Published:** February 28, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Least Absolute Shrinkage and Selection Operator (LASSO) is used for variable selection as well as for handling the multicollinearity problem simultaneously in the linear regression model. LASSO produces estimates having high variance if the number of predictors is higher than the number of observations and if high multicollinearity exists among the predictor variables. To handle this problem, Elastic Net (ENet) estimator was introduced by combining LASSO and Ridge estimator (RE). The solutions of LASSO and ENet have been obtained using Least Angle Regression (LARS) and LARS-EN algorithms, respectively. In this article, we proposed an alternative algorithm to overcome the issues in LASSO that can be combined LASSO with other existing biased estimators namely Almost Unbiased Ridge Estimator (AURE), Liu Estimator (LE), Almost Unbiased Liu Estimator (AULE), Principal Component Regression Estimator (PCRE), r-k class estimator and r-d class estimator. Further, we examine the performance of the proposed algorithm using a Monte-Carlo simulation study and real-world examples. The results showed that the LARS-rk and LARS-rd algorithms, which are combined LASSO with r-k class estimator and r-d class estimator, outperformed other algorithms under the moderated and severe multicollinearity.

## Keywords

Variable Selection, Least Absolute Shrinkage and Selection Operator (LASSO), Least Angle Regression (LARS), Elastic Net (ENet), Biased Estimators

## 1. Introduction

Due to the simplicity and interpretability, linear regression models play a signif-

icant role in modern statistical methods. The linear regression model aspires to find the linear relationship between the dependent variable and the non-stochastic explanatory variables for the prediction purpose.

Let us consider the linear regression model

$$y = X\beta + \epsilon \tag{1}$$

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

It is well known that the Ordinary Least Square Estimator (OLSE) is the Best Linear Unbiased Estimator (BLUE) for finding the unknown parameter vector in model (1), which can be obtained by minimizing Error Sum of Squares (ESS),

$$ESS(\beta) = \epsilon'\epsilon = (y - X\beta)'(y - X\beta) \tag{2}$$

with respect to  $\beta$ , and defined as

$$\hat{\beta}_{OLSE} = \arg \min_{\beta} \left\{ (y - X\beta)'(y - X\beta) \right\} = (X'X)^{-1} X'y. \tag{3}$$

However, the OLSE is unstable and produces parameter estimates with high variance when multicollinearity exists on  $X$ . As a curative action to the multicollinearity problem, the biased estimators have been used by many researchers. The following biased estimators are popular in statistical literature:

- Principal Component Regression Estimator (PCRE) [1]

$$\hat{\beta}_{PCRE} = T_h (T_h' X' X T_h)^{-1} T_h' X'y = T_h T_h' \hat{\beta}_{OLSE}, \tag{4}$$

where  $T_h = (t_1, t_2, \dots, t_h)$  is the first  $h$  columns of the standardized eigenvectors of  $X'X$ .

- Ridge Estimator (RE) [2]

$$\begin{aligned} \hat{\beta}_{RE} &= \arg \min_{\beta} \left\{ (y - X\beta)'(y - X\beta) + k \sum_{j=1}^p \beta_j^2 \right\} \\ &= (X'X + kI)^{-1} X'y, \end{aligned} \tag{5}$$

where  $k > 0$  is the regularization parameter, and  $I$  is the  $p \times p$  identity matrix.

- r-k class estimator [3]

$$\hat{\beta}_{rk} = T_h (T_h' X' X T_h + kI)^{-1} T_h' X'y = T_h T_h' (X'X + kI)^{-1} X'y. \tag{6}$$

Note that r-k class estimator is a combination of PCRE and RE.

- Almost Unbiased Ridge Estimator (AURE) [4]

$$\hat{\beta}_{AURE} = (I - k^2 (X'X + kI)^{-2}) \hat{\beta}_{OLSE}. \tag{7}$$

- Liu Estimator (LE) [5]

$$\hat{\beta}_{LE} = (X'X + I)^{-1} (X'y + d \hat{\beta}_{OLSE}) = (X'X + I)^{-1} (X'X + dI) \hat{\beta}_{OLSE}, \tag{8}$$

where  $0 < d < 1$  is the regularization parameter.

- Almost Unbiased Liu Estimator (AULE) [6]

$$\hat{\beta}_{AULE} = \left( \mathbf{I} - (1-d)^2 (\mathbf{X}\mathbf{X} + \mathbf{I})^{-2} \right) \hat{\beta}_{OLSE}. \tag{9}$$

- r-d class estimator [7]

$$\begin{aligned} \hat{\beta}_{rd} &= \mathbf{T}_h (\mathbf{T}'_h \mathbf{X} \mathbf{X} \mathbf{T}_h + \mathbf{I})^{-1} \left( \mathbf{I} + d (\mathbf{T}'_h \mathbf{X} \mathbf{X} \mathbf{T}_h)^{-1} \right) \mathbf{T}'_h \mathbf{X}' \mathbf{y} \\ &= \mathbf{T}_h \mathbf{T}'_h (\mathbf{X}\mathbf{X} + \mathbf{I})^{-1} (\mathbf{X}\mathbf{X} + d\mathbf{I}) \hat{\beta}_{OLSE}. \end{aligned} \tag{10}$$

Note that r-d class estimator is a combination of PCRE and LE.

According to Kayanan and Wijekoon [8], the generalized form to represent the estimators RE, AURE, LE, AULE, PCR, r-k class estimator and r-d class estimator is given by

$$\hat{\beta}_G = \mathbf{G} \hat{\beta}_{OLSE} \tag{11}$$

where

$$\hat{\beta}_G = \begin{cases} \hat{\beta}_{RE} & \text{if } \mathbf{G} = (\mathbf{X}\mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}\mathbf{X} \\ \hat{\beta}_{AURE} & \text{if } \mathbf{G} = \left( \mathbf{I} - k^2 (\mathbf{X}\mathbf{X} + k\mathbf{I})^{-2} \right) \\ \hat{\beta}_{LE} & \text{if } \mathbf{G} = (\mathbf{X}\mathbf{X} + \mathbf{I})^{-1} (\mathbf{X}\mathbf{X} + d\mathbf{I}) \\ \hat{\beta}_{AULE} & \text{if } \mathbf{G} = \left( \mathbf{I} - (1-d)^2 (\mathbf{X}\mathbf{X} + \mathbf{I})^{-2} \right) \\ \hat{\beta}_{PCRE} & \text{if } \mathbf{G} = \mathbf{T}_h \mathbf{T}'_h \\ \hat{\beta}_{rk} & \text{if } \mathbf{G} = \mathbf{T}_h \mathbf{T}'_h (\mathbf{X}\mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}\mathbf{X} \\ \hat{\beta}_{rd} & \text{if } \mathbf{G} = \mathbf{T}_h \mathbf{T}'_h (\mathbf{X}\mathbf{X} + \mathbf{I})^{-1} (\mathbf{X}\mathbf{X} + d\mathbf{I}) \end{cases}$$

In recent studies, Kayanan and Wijekoon [8] have shown that r-k class estimator and r-d class estimator outperformed other estimators for the selected range of regularization parameter values when multicollinearity exists among the predictor variables. However, biased estimators introduce heavy bias when the number of predictor variables is high, and the final model may contain some irrelevant predictor variables as well. To handle this issue, Tibshirani [9] proposed the Least Absolute Shrinkage and Selection Operator (LASSO) as

$$\hat{\beta}_{LASSO} = \arg \min_{\beta} \left\{ (\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta) \right\} \text{ subject to } \sum_{j=1}^p |\beta_j| \leq t, \tag{12}$$

where  $t \geq 0$  is a turning parameter. Note that we cannot find an analytic solution for LASSO since  $\sum_{j=1}^p |\beta_j|$  is a non-differentiable function. Tibshirani [9] and Fu [10] have used the standard quadratic programming technique and shooting algorithm, respectively, to find solutions for LASSO. Apart from these two methods, the Least Angle Regression (LARS) algorithm proposed by Efron *et al.* [11] is a popular one in the recent literature to find LASSO solutions. The LASSO yields both the multicollinearity problem and variable selection simultaneously in the linear regression model. However, LASSO failed to outperform RE if high multicollinearity exists among predictors, and it is unsteady when the number of predictors is higher than the number of observations [12]. To over-

come this problem, Zou and Hastie [12] proposed Elastic net (ENet) estimator by combining LASSO and RE as

$$\hat{\beta}_{Enet} = \arg \min_{\beta} \left\{ (y - X\beta)'(y - X\beta) + k \sum_{j=1}^p \beta_j^2 \right\} \text{ subject to } \sum_{j=1}^p |\beta_j| \leq t. \quad (13)$$

The LARS-EN algorithm, which is a modified version of the LARS algorithm, has been used to obtain solutions for ENet.

In this work, we propose generalized version of LARS algorithm that can be combined LASSO with other biased estimators such as AURE, LE, AULE, PCRE, r-k class and r-d class estimators. Further, we compared the prediction performance of proposed algorithm with existing algorithms of LASSO and Enet using a Monte-Carlo simulation study and real-world examples. The structure of the rest of the article is as follows: Section 2 contains proposed algorithms, Section 3 shows the comparison of proposed algorithm, Section 4 concludes the article, and references are provided at the end of the paper.

## 2. Generalized Least Angle Regression (GLARS) Algorithm

Based on Efron *et al.* [11] and Hettigoda [13], now we propose GLARS algorithm as follows:

**Step 1:** Standardize the predictor variables  $X$  to have a mean of zero and a standard deviation of one, and response variable  $y$  to have a mean zero.

**Step 2:** Start with the initial estimated value of  $\beta$  as  $\hat{\beta} = 0$ , and the residual  $r_0 = y$ .

**Step 3:** Find the predictor  $X_{j_1}$  most correlated with  $r_0$  as follows

$$X_{j_1} = \max_j |Cor(X_j, r_0)|; \quad j = 1, 2, \dots, p. \quad (14)$$

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

In a similar way, the  $i^{th}$  variable  $X_{j_i}$  eventually earns its way into the active set, and then GLARS proceeds in the equiangular direction between  $X_{j_1}, X_{j_2}, \dots, X_{j_i}$ . 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 updating using the following formula:

$$\hat{\beta}_{j_i} = \hat{\beta}_{j_{(i-1)}} + \alpha_i u_i, \quad (15)$$

where  $\alpha_i$  is a value between [0,1] which represents how far the estimate moves in the direction before another variable enters the model and the direction changes again, and  $u_i$  is the equiangular vector.

The direction  $u_i$  is calculated using the following formula:

$$u_i = G_E (E_i' X' X E_i)^{-1} E_i' X' r_{i-1}, \quad (16)$$

where  $E_i$  is the matrix with column  $(e_{j_1}, e_{j_2}, \dots, e_{j_i})$ ,  $e_j$  be the  $j^{th}$  standard

unit vector in  $\mathbb{R}^p$  which has the index of variables selected in each subsequent steps, and  $\mathbf{G}_E$  is a generalized variable which can be substituted by respective expressions for any of estimators of our interest as listed in **Table 1**. Then,  $\alpha_i$  is calculated as follows:

$$\alpha_i = \min\{\alpha_{ji}^+, \alpha_{ji}^-, \alpha_{ji}^*\} \in [0,1] \tag{17}$$

where

$$\alpha_{ji}^\pm = \frac{\text{Cor}(\mathbf{r}_{i-1}, X_{ji}) \pm \text{Cor}(\mathbf{r}_{i-1}, X_j)}{\text{Cor}(\mathbf{r}_{i-1}, X_{ji}) \pm \text{Cor}(\mathbf{X}\mathbf{u}_i, X_j)} \tag{18}$$

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

$$\alpha_{ji}^* = -\frac{\hat{\beta}_{j(i-1)}}{\mathbf{u}_i} \tag{19}$$

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

**Step 4:** 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 the residual  $\mathbf{r}_i$  related to the current step is calculated as

$$\mathbf{r}_i = \mathbf{r}_{i-1} - \alpha_i \mathbf{X}\mathbf{u}_i, \tag{20}$$

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

**Step 5:** Proceed Step 3 until  $\alpha_i = 1$ .

In **Table 1**,  $\mathbf{I}_{p_E}$  is the  $p_E \times p_E$  identity matrix,  $p_E$  is the number of selected variables in each subsequent step, and  $\mathbf{T}_{h_E} = (t_1, t_2, \dots, t_{h_E})$  is the first  $h_E$  columns of the standardized eigenvectors of  $\mathbf{E}_i' \mathbf{X}' \mathbf{X} \mathbf{E}_i$ .

### 2.1. Properties of GLARS

GLARS algorithm sequentially updates the combined estimates of LASSO and other estimators. It requires  $O(m^3 + pm^2)$  operations, where  $m$  is the number of steps. The prediction performance of the GLARS is evaluated using the Root

**Table 1.**  $\mathbf{G}_E$  of the estimators for GLARS.

| Estimators | $\mathbf{G}_E$  |
|------------|---|
| OLSE       | $\mathbf{E}_i$  |
| RE         | $\mathbf{E}_i (\mathbf{E}_i' (\mathbf{X}'\mathbf{X} + k\mathbf{I}) \mathbf{E}_i)^{-1} (\mathbf{E}_i' \mathbf{X}' \mathbf{X} \mathbf{E}_i)$  |
| AURE       | $\mathbf{E}_i (\mathbf{I}_{p_E} - k^2 (\mathbf{E}_i' (\mathbf{X}'\mathbf{X} + k\mathbf{I}) \mathbf{E}_i)^{-2})$   |
| LE         | $\mathbf{E}_i (\mathbf{E}_i' (\mathbf{X}'\mathbf{X} + \mathbf{I}) \mathbf{E}_i)^{-1} (\mathbf{E}_i' (\mathbf{X}'\mathbf{X} + d\mathbf{I}) \mathbf{E}_i)$                                    |
| AULE       | $\mathbf{E}_i (\mathbf{I}_{p_E} - (1-d)^2 (\mathbf{E}_i' (\mathbf{X}'\mathbf{X} + \mathbf{I}) \mathbf{E}_i)^{-2})$  |
| PCRE       | $\mathbf{T}_{h_E} \mathbf{T}_{h_E}' \mathbf{E}_i$   |
| r-k class  | $\mathbf{T}_{h_E} \mathbf{T}_{h_E}' \mathbf{E}_i (\mathbf{E}_i' (\mathbf{X}'\mathbf{X} + k\mathbf{I}) \mathbf{E}_i)^{-1} (\mathbf{E}_i' \mathbf{X}' \mathbf{X} \mathbf{E}_i)$               |
| r-d class  | $\mathbf{T}_{h_E} \mathbf{T}_{h_E}' \mathbf{E}_i (\mathbf{E}_i' (\mathbf{X}'\mathbf{X} + \mathbf{I}) \mathbf{E}_i)^{-1} (\mathbf{E}_i' (\mathbf{X}'\mathbf{X} + d\mathbf{I}) \mathbf{E}_i)$ |

Mean Square Error (RMSE) criterion, which is described in Section 3. We can use GLARS to combine LASSO and any of estimators as listed in **Table 1**.

Note that GLARS provides LASSO and ENet solutions when  $\mathbf{G}_E$  equals to the corresponding expressions of OLSE and RE, respectively. For simplicity, we refer GLARS as LARS-LASSO, LARS-EN, LARS-AURE, LARS-LE, LARS-AULE, LARS-PCRE, LARS-rk and LARS-rd when  $\mathbf{G}_E$  equals to the corresponding expressions of OLSE, RE, AURE, LE, AULE, PCRE, r-k class and r-d class estimators, respectively.

## 2.2. Selection of Regularization Parameter Values

According to Efron *et al.* [11] and Zou and Hastie [12], the conventional tuning parameter of LARS-LASSO is  $t = \sum_{j=1}^p |\hat{\beta}_j|$ , and LARS-LASSO automatically controls it. The regularization parameter  $k$  of LARS-EN is selected using 10-fold cross-validation for each  $t$ . Similarly, we choose the regularization parameter  $k$  or  $d$  of proposed algorithms using 10-fold cross-validation for each  $t$ .

## 3. Comparison of Proposed Algorithms

Proposed algorithms are compared with the LARS-LASSO and LARS-EN algorithms using the RMSE criterion, which is the expected prediction error of the algorithms, and is defined as

$$RMSE(\hat{\boldsymbol{\beta}}) = \sqrt{\frac{1}{n} (\mathbf{y}_{new} - \mathbf{X}_{new} \hat{\boldsymbol{\beta}})' (\mathbf{y}_{new} - \mathbf{X}_{new} \hat{\boldsymbol{\beta}})} \quad (21)$$

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

### 3.1. Simulation Study

According to McDonald and Galarneau [14], first we generate the predictor variables by using the following formula:

$$x_{i,j} = \sqrt{(1-\rho^2)} z_{i,j} + \rho z_{i,m+1}; \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m. \quad (22)$$

where  $z_{i,j}$  is an independent standard normal pseudo random number, and  $\rho$  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

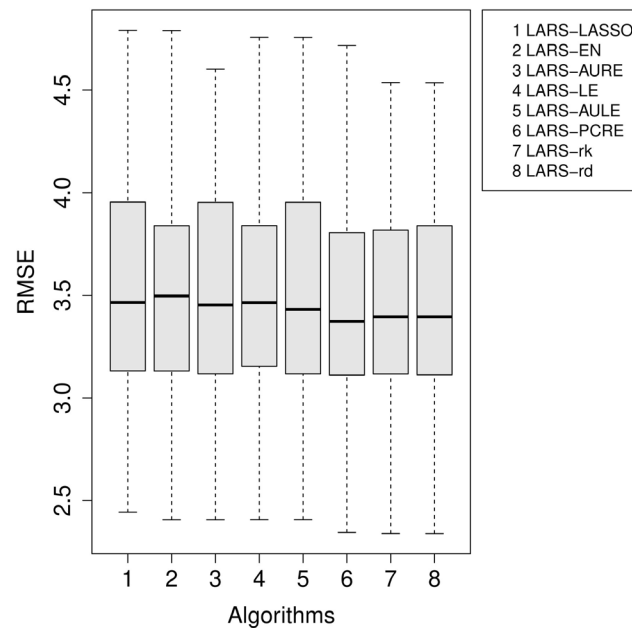
$$y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_5 x_{i,20} + \epsilon_i; \quad i = 1, 2, \dots, 100. \quad (23)$$

where  $\epsilon_i$  is a normal pseudo random number with mean zero and common variance  $\sigma^2$ .

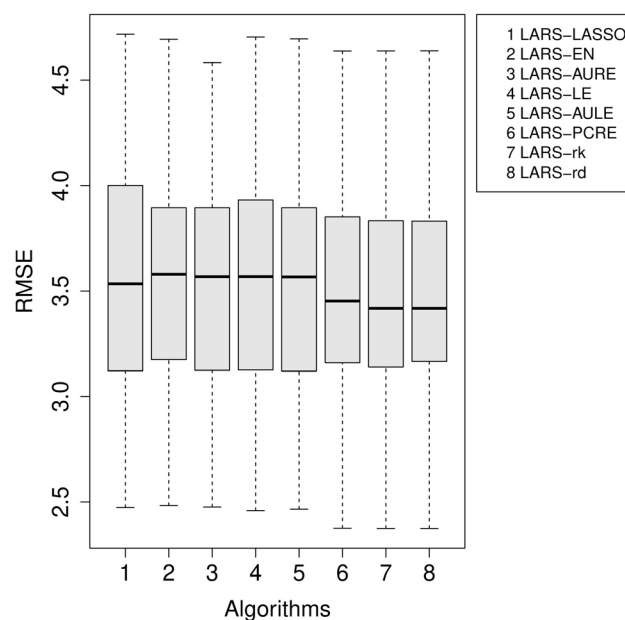
We choose  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_{20})$  as the normalized eigenvector corresponding to the largest eigenvalue of  $\mathbf{X}'\mathbf{X}$  for which  $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$ . To investigate the effects of different degrees of multicollinearity on 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. The Cross-validated RMSE of the algorithms are displayed in **Figures 1-3**, and the median cross-validated RMSE of the algorithms are displayed in **Table 2-4**.

From **Figures 1-3** and **Tables 2-4**, we can observe that LARS-PCRE, LARS-rk and LARS-rd algorithms show better performance compared to other algorithms in weak, moderated and high multicollinearity, respectively.



**Figure 1.** Cross-validated RMSE values of the algorithms when  $\rho = 0.5$ .



**Figure 2.** Cross-validated RMSE values of the algorithms when  $\rho = 0.7$ .

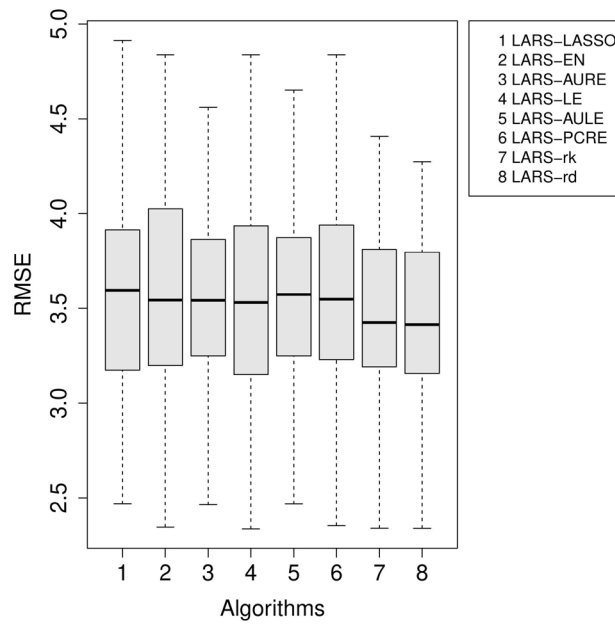


Figure 3. Cross-validated RMSE values of the algorithms when  $\rho = 0.9$ .

Table 2. Median cross-validated RMSE values when  $\rho = 0.5$ .

| Algorithms | RMSE          | $(k, d)$ | $t$    | Selected variables |
|------------|---------------|----------|--------|--------------------|
| LARS-LASSO | 3.4656        | -        | 7.8042 | 16                 |
| LARS-EN    | 3.4974        | 0.440    | 8.0932 | 17                 |
| LARS-AURE  | 3.4535        | 0.320    | 8.2268 | 17                 |
| LARS-LE    | 3.4649        | 0.800    | 7.9357 | 17                 |
| LARS-AULE  | 3.4315        | 0.990    | 8.0093 | 17                 |
| LARS-PCRE  | <b>3.3734</b> | -        | 7.6216 | 17                 |
| LARS-rk    | 3.3960        | 0.465    | 7.5012 | 17                 |
| LARS-rd    | 3.3961        | 0.625    | 7.5108 | 18                 |

Table 3. Median cross-validated RMSE values when  $\rho = 0.7$ .

| Algorithms | RMSE          | $(k, d)$ | $t$    | Selected variables |
|------------|---------------|----------|--------|--------------------|
| LARS-LASSO | 3.5342        | -        | 9.1450 | 15                 |
| LARS-EN    | 3.5796        | 0.360    | 9.4898 | 17                 |
| LARS-AURE  | 3.5682        | 0.675    | 9.5187 | 17                 |
| LARS-LE    | 3.5685        | 0.675    | 9.2909 | 17                 |
| LARS-AULE  | 3.5672        | 0.795    | 9.3737 | 17                 |
| LARS-PCRE  | 3.4526        | -        | 8.7773 | 17                 |
| LARS-rk    | <b>3.4177</b> | 0.460    | 8.5706 | 17                 |
| LARS-rd    | 3.4178        | 0.540    | 8.4988 | 17                 |



**Table 4.** Median Cross-validated RMSE values when  $\rho = 0.9$ .

| Algorithms | RMSE          | $(k, d)$ | $t$    | Selected variables |
|------------|---------------|----------|--------|--------------------|
| LARS-LASSO | 3.5954        | -        | 13.117 | 15                 |
| LARS-EN    | 3.5430        | 0.500    | 13.747 | 16                 |
| LARS-AURE  | 3.5418        | 0.695    | 13.896 | 17                 |
| LARS-LE    | 3.5306        | 0.485    | 13.712 | 16                 |
| LARS-AULE  | 3.5727        | 0.615    | 13.580 | 16                 |
| LARS-PCRE  | 3.5479        | -        | 12.391 | 16                 |
| LARS-rk    | 3.4254        | 0.550    | 12.074 | 16                 |
| LARS-rd    | <b>3.4146</b> | 0.425    | 12.384 | 16                 |

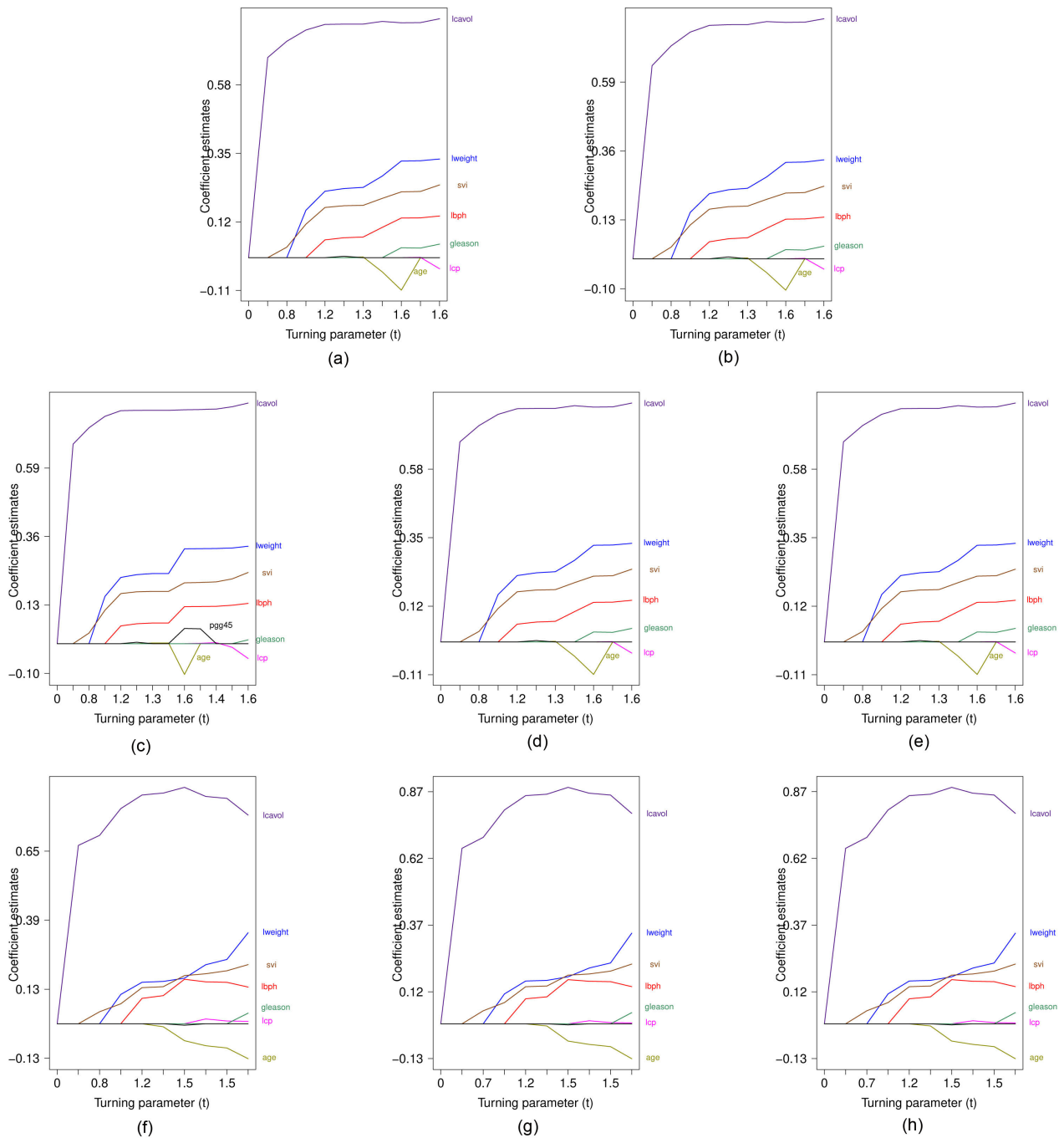
### 3.2. Real-World Examples

Two real-world examples, namely the Prostate Cancer Data [15], and the US-crime dataset [16], are considered to compare the performance of the proposed algorithms.

**Prostate Cancer Data:** In the Prostate Cancer Data, the predictors are 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 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 high multicollinearity among the predictor variables. Stamey *et al.* [15] have examined the correlation between the level of prostate specific antigen with those eight clinical measures. Further, Tibshirani [9], Efron *et al.* [11] and Zou and Hastie [12] have used this data to examine the performance of LASSO, LARS algorithm and Enet estimators. This data set is attached with “lasso2” R package. We have used 67 observations to fit the model, and 30 observations to calculate the RMSE. The cross-validated RMSE of the algorithms are displayed in **Table 5**, and Coefficient paths of each algorithm are displayed in **Figure 4**.

From **Table 5**, we can observe that LARS-rd algorithm outperforms other algorithms on Prostate Cancer Data. From **Figure 4**, we can observe that LARS-LASSO, LARS-EN, LARS-AURE, LARS- and LARS-AULE ignore the variables age and pgg45 in the final model, but LARS-PCRE, LARS-rk and LARS-rd ignore the variable pgg45 only.

**UScrime Data:** The UScrime dataset has 16 variables with 47 observations, and it is attached with “MASS” R package. This data set contains the following columns: M (percentage of males aged 14 - 24), So (indicator variable for a Southern state), Ed (mean years of schooling), Po1 (police expenditure in 1960), Po2 (police expenditure in 1959), LF (labor force participation rate), M.F (number of males per 1000 females), Pop (state population), NW (number of non-whites



**Figure 4.** Coefficient paths of the (a) LARS-LASSO; (b) LARS-EN; (c) LARS-AURE; (d) LARS-LE; (e) LARS-AULE; (f) LARS-PCRE; (g) LARS-rk and (h) LARS-rd versus  $t = \sum_{j=1}^p |\hat{\beta}_j|$  for the prostate cancer data.

per 1000 people), U1 (unemployment rate of urban males 14 - 24), U2 (unemployment rate of urban males 35 - 39), GDP (gross domestic product per head), Ineq (income inequality), Prob (probability of imprisonment), Time (average time served in state prisons),  $\gamma$  (rate of crimes in a particular category per head of population). The variable  $\gamma$  is considered as a dependent variable, and the variable So is ignored since it is categorical. Venables and Ripley [16] have

examined the effect of punishment regimes on crime rates using this dataset. The Variance Inflation Factor (VIF) values of the predictor variables of the dataset are 113.028, 104.58, 9.97, 7.43, 5.19, 5.05, 4.83, 3.84, 3.69, 2.88, 2.86, 2.75, 2.66 and 2.53, and the condition number is 923, which shows high multicollinearity among the predictor variables. For the analysis, we have used 37 observations to fit the model, and 10 observations to calculate the RMSE. The cross-validated RMSE of the algorithms are displayed in **Table 6**, and Coefficient paths of each algorithm are displayed in **Figure 5**.

From **Table 6**, we can observe that LARS-rd algorithm outperforms other algorithms on UScrime Data. From **Figure 5**, we can observe that:

- LARS-LASSO ignores the variables Ed, Ineq and Prob,
- LARS-EN, LARS-AURE, LARS-and LARS-AULE, LARS-PCRE ignore the variables Ineq and Prob, and
- LARS-rk and LARS-rd ignore the variables M, M.F and Ineq.

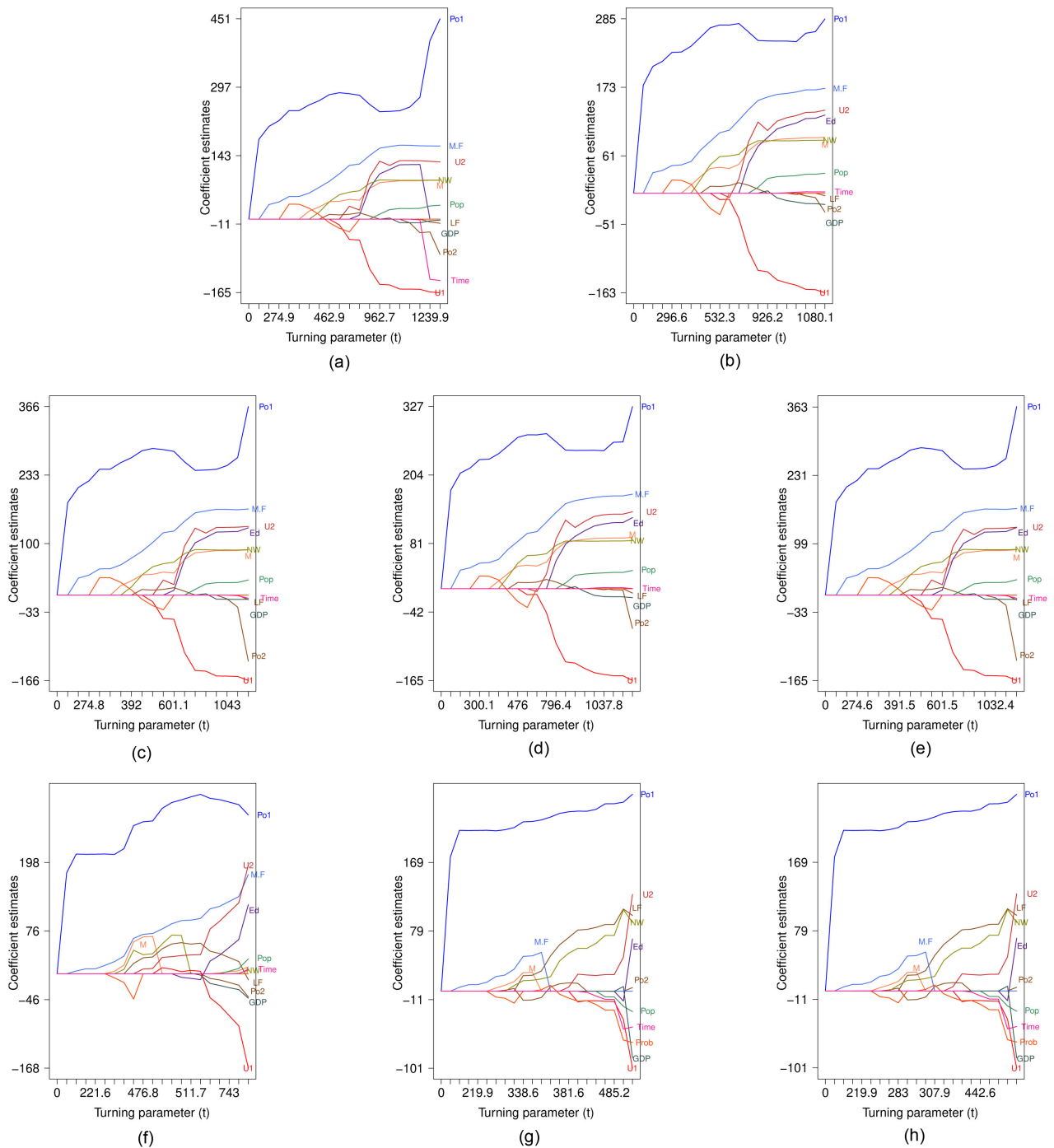
Since different algorithms choose the different combination of predictor variables in the final model, as shown in the plot of coefficient paths, the researcher can decide the most suitable model for the relevant practical situation.

**Table 5.** Cross-validated RMSE values of prostate cancer data.

| Algorithms | RMSE           | $(k, d)$ | $t$    | Selected variables |
|------------|----------------|----------|--------|--------------------|
| LARS-LASSO | 0.80057        | -        | 1.5996 | 6                  |
| LARS-EN    | 0.80039        | 1.15     | 1.5899 | 6                  |
| LARS-AURE  | 0.79548        | 2.00     | 1.5728 | 6                  |
| LARS-LE    | 0.80057        | 0.99     | 1.5995 | 6                  |
| LARS-AULE  | 0.80057        | 0.99     | 1.5996 | 6                  |
| LARS-PCRE  | 0.78191        | -        | 1.6705 | 7                  |
| LARS-rk    | 0.78027        | 0.58     | 1.6678 | 7                  |
| LARS-rd    | <b>0.78025</b> | 0.41     | 1.6678 | 7                  |

**Table 6.** Cross-validated RMSE values of UScrime data.

| Algorithms | RMSE          | $(k, d)$ | $t$  | Selected variables |
|------------|---------------|----------|------|--------------------|
| LARS-LASSO | 333.45        | -        | 1345 | 11                 |
| LARS-EN    | 255.27        | 0.71     | 1148 | 12                 |
| LARS-AURE  | 252.00        | 0.25     | 1318 | 12                 |
| LARS-LE    | 251.58        | 0.36     | 1232 | 12                 |
| LARS-AULE  | 252.51        | 0.24     | 1311 | 12                 |
| LARS-PCRE  | 265.57        | -        | 1090 | 12                 |
| LARS-rk    | 234.55        | 0.8      | 977  | 11                 |
| LARS-rd    | <b>234.38</b> | 0.2      | 980  | 11                 |



**Figure 5.** Coefficient paths of the (a) LARS-LASSO; (b) LARS-EN; (c) LARS-AURE; (d) LARS-LE; (e) LARS-AULE; (f) LARS-PCRE; (g) LARS-rk and (h) LARS-rd versus  $t = \sum_{j=1}^p |\hat{\beta}_j|$  for the UScrime Data.

### 4. Conclusions

This study clearly showed that the proposed LARS-rk and LARS-rd algorithms performed well in the high dimensional linear regression model when moderated and high multicollinearity existed among the predictor variables, respectively.

The appropriate algorithm for a particular practical problem can be chosen based on the variables of interest and prediction performance by referring to the plot of coefficient paths.

## Acknowledgements

We thank the Postgraduate Institute of Science, University of Peradeniya, Sri Lanka for providing all facilities to do this research, and also we thank the reviewers for their valuable comments.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Massy, W.F. (1965) Principal Components Regression in Exploratory Statistical Research. *Journal of the American Statistical Association*, **60**, 234-266. <https://doi.org/10.1080/01621459.1965.10480787>
- [2] Hoerl, E. and Kennard, R.W. (1970) Ridge Regression: Biased Estimation for Non-orthogonal Problems. *Technometrics*, **12**, 55-67. <https://doi.org/10.1080/00401706.1970.10488634>
- [3] Baye, M.R. and Parker, D.F. (1984) Combining Ridge and Principal Component Regression: A Money Demand Illustration. *Communications in Statistics-Theory and Methods*, **13**, 197-205. <https://doi.org/10.1080/03610928408828675>
- [4] Singh, B., Chaubey, Y.P. and Dwivedi, T.D. (1986) An Almost Unbiased Ridge Estimator. *The Indian Journal of Statistics*, **48**, 342-346.
- [5] Liu, K. (1993) A New Class of Biased Estimate in Linear Regression. *Communication in Statistics-Theory and Methods*, **22**, 393-402. <https://doi.org/10.1080/03610929308831027>
- [6] Akdeniz, F. and Kaciranlar, S. (1995) On the Almost Unbiased Generalized Liu Estimator and Unbiased Estimation of the Bias and MSE. *Communication in Statistics-Theory and Methods*, **24**, 1789-1797. <https://doi.org/10.1080/03610929508831585>
- [7] Kaciranlar, S. and Sakallioğlu, S. (2001) Combining the Liu Estimator and the Principal Component. *Communications in Statistics-Theory and Methods*, **30**, 2699-2705. <https://doi.org/10.1081/STA-100108454>
- [8] Kayanan, M. and Wijekoon, P. (2017) Performance of Existing Biased Estimators and the Respective Predictors in a Misspecified Linear Regression Model. *Open Journal of Statistics*, **7**, 876-900. <https://doi.org/10.4236/ojs.2017.75062>
- [9] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**, 267-288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- [10] Fu, W.J. (1998) Penalized Regressions: The Bridge versus the Lasso. *Journal of Computational and Graphical Statistics*, **7**, 397-416. <https://doi.org/10.1080/10618600.1998.10474784>
- [11] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004) Least Angle Regression. *The Annals of Statistics*, **32**, 407-499. <https://doi.org/10.1214/009053604000000067>

- [12] Zou, H. and Hastie, T. (2005) Regularization and Variable Selection via the Elastic net. *Journal of the Royal Statistical Society: Series B*, **67**, 301-320.  
<https://doi.org/10.1111/j.1467-9868.2005.00503.x>
- [13] Hettigoda, S. (2016) Computation of Least Angle Regression Coefficient Profiles and Lasso Estimates. Master's Thesis, University of Louisville, Louisville, KY.
- [14] McDonald, G.C. and Galarneau, D.I. (1975) A Monte Carlo Evaluation of Some Ridge-Type Estimators. *Journal of the American Statistical Association*, **70**, 407-416.  
<https://doi.org/10.1080/01621459.1975.10479882>
- [15] Stamey, T.A., Kabalin, J.N., Mcneal, J.E., Johnstone, I.M., Freiha, F., Redwine, E.A. and 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. [https://doi.org/10.1016/S0022-5347\(17\)41175-X](https://doi.org/10.1016/S0022-5347(17)41175-X)
- [16] Venables, W.N. and Ripley, B.D. (1999) *Modern Applied Statistics with S-PLUS*. 3rd Edition, Springer-Verlag, New York.  
<https://doi.org/10.1007/978-1-4757-3121-7>