Turkish Journal of Science & Technology
18(2), 319-330, 2023.

Research Paper
https://doi.org/10.55525/tjst.1244925

The Effect of the Second Stage Estimator on Model Performance in Post-LASSO Method

Murat GENÇ[1*], Ömer ÖZBİLEN[2]

[1] Department of Management Information Systems, Faculty of Economics and Administrative Sciences, Tarsus University Mersin 33400, Türkiye

[2] Department of Geomatic Engineering, Mersin University Çiftlikköy Campus Mersin, Türkiye

[*1] muratgenc@tarsus.edu.tr, [2] ozbilen@mersin.edu.tr

(Geliş/Received: 30/01/2023; Kabul/Accepted: 24/07/2023)

Abstract: Penalized linear regression methods are used for the accurate prediction of new observations and to obtain interpretable models. The performance of these methods depends on the properties of the true coefficient vector. The LASSO method is a penalized regression method that can simultaneously perform coefficient shrinkage and variable selection in a continuous process. Depending on the structure of the dataset, different estimators have been proposed to overcome the problems faced by LASSO. The estimation method used in the second stage of the post-LASSO two-stage regression method proposed as an alternative to LASSO has a considerable effect on model performance.

In this study, the performance of the post-LASSO is compared with classical penalized regression methods ridge, LASSO, elastic net, adaptive LASSO and Post-LASSO by using different estimation methods in the second stage of the post-LASSO. In addition, the effect of the magnitude and position of the signal values in the real coefficient vector on the performance of the models obtained by these methods is analyzed. The mean squared error and standard deviation of the predictions calculated on the test set are used to compare the prediction performance of the models, while the active set sizes are used to compare their performance in variable selection. According to the findings obtained from the simulation studies, the choice of the second-stage estimator and the structure of the true coefficient vector significantly affect the success of the post-LASSO method compared to other methods.

Key words: Linear Regression, LASSO, Post-LASSO, Multicollinearity.

Post-LASSO Yönteminde İkinci Aşama Tahmin Edicisinin Model Performansına Etkisi

Öz: Cezalı doğrusal regresyon yöntemleri yeni gözlemlerin doğru ön tahmini ve yorumlanabilir modeller elde edilmesi için kullanılır. Bu yöntemlerin performansı gerçek katsayı vektörünün özelliklerine bağlı olarak değişmektedir. LASSO yöntemi sürekli bir süreçte eşanlı olarak katsayı büzme ve değişken seçimi yapabilen bir cezalı regresyon yöntemidir. Veri kümesinin yapısına bağlı olarak LASSO'nun karşılaştığı problemlerin aşılabilmesi için farklı tahmin ediciler önerilmiştir. LASSO'ya alternatif olarak önerilen Post-LASSO iki aşamalı regresyon yönteminin ikinci aşamasında kullanılan tahmin yöntemi model performansı üzerinde kayda değer bir etkiye sahiptir.

Bu çalışmada Post-LASSO'nun ikinci aşamasında farklı tahminleme yöntemleri kullanılarak klasik cezalı regresyon yöntemleri olan ridge, LASSO, elastik net, uyarlanabilir LASSO ile Post-LASSO'nun performansı karşılaştırılmıştır. Ayrıca gerçek katsayı vektöründeki sinyal değerlerinin büyüklük ve konumunun söz konusu yöntemlerle elde edilen modellerin performansı üzerindeki etkisi incelenmiştir. Modellerin ön tahmin performansının karşılaştırılması için test kümesi üzerinde hesaplanan hata kareler ortalaması ve tahminlerin standart sapması; değişken seçimindeki performanslarının karşılaştırılması için aktif küme büyüklükleri kullanılmıştır. Simülasyon çalışmalarından elde edilen bulgulara göre ikinci aşama tahmin edicinin seçimi ile gerçek katsayı vektörünün yapısı Post-LASSO yönteminin diğer yöntemlere göre başarısını önemli ölçüde etkilemektedir.

Anahtar kelimeler: Doğrusal regresyon, LASSO, Post-LASSO, Çoklu İç İlişki.

1. Introduction

In statistical modeling, linear regression analysis is a technique used to estimate the relationship between a continuous response variable and explanatory variables. The regression analysis has many applications in different disciplines 1, 2, 3. The first step in forming a linear regression model is the estimation of regression coefficients.

A linear regression model is defined as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

* Corresponding author: muratgenc@tarsus.edu.tr. ORCID Number of authors: [1] 0000-0002-6335-3044, [2] 0000-0001-6110-1911

where $\mathbf{y}$ is an $n \times 1$ vector of response variables, $\mathbf{X}$ is an $n \times p$ matrix of explanatory variables, $\boldsymbol{\beta}$ is an $p \times 1$ vector of unknown coefficients, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of error terms with mean $\mathbf{0}$ and covariance matrix $\sigma^2\mathbf{I}$. The most used method for coefficient estimation in linear regression is the ordinary least squares (OLS) method. The OLS estimator of the coefficients for the regression model in Equation (1) is

$$\widehat{\boldsymbol{\beta}}_{ekk} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{2}$$

A dataset becomes ill-conditioned if there is multicollinearity, which is defined as a high degree of the linear relationship between explanatory variables. Although the OLS is often used to estimate regression coefficients, the OLS estimator yields unsatisfactory estimates due to inflation of the variance of the coefficients when the matrix of explanatory variables is ill-conditioned. Moreover, by its nature, the OLS estimator cannot estimate any coefficient as zero. Therefore, it cannot perform automatic variable selection. New estimators have been proposed to overcome these drawbacks of the OLS estimator. The ridge 4, Liu 5 restricted least squares 6 and restricted ridge estimator 7 are the methods that can produce more accurate prediction accuracy than the OLS by using some constraints on the coefficients. In contrast, the non-negative garrote 8, bridge regression 9, LASSO (least absolute shrinkage and selection operator) 10, elastic net 11 and adaptive LASSO 12 are methods that provide automatic variable selection as well as shrinkage of regression coefficients. The Bridge regression is a penalized regression method with $\|\beta\|_1^\gamma$, $\gamma > 0$ penalty function. The $\gamma = 1$ case of the penalty function of the Bridge regression corresponds to the LASSO method. The elastic net is obtained by adding the ℓ_2 norm term to the penalty function of the LASSO and is more flexible than the LASSO. The adaptive LASSO is a method based on the calculation of LASSO estimates using adaptive weights.

The Post-LASSO estimator is a two-stage regression analysis method. In the first stage of the post-LASSO, the coefficient estimates are calculated. The variables with the non-zero coefficients (signals) among these estimates are selected. In the second stage, the regression coefficients of the explanatory variables selected at the end of the first stage on the original response variable are calculated by using a certain estimator. There are various studies in the literature on such two-stage methods. 13 proposed a post-LASSO estimator based on the implementation of the LASSO in the first stage and the OLS in the second stage and showed that this estimator is at least as good as LASSO in terms of convergence speed while being less biased than LASSO. 14 takes into consideration a two-step estimation technique for estimating the interaction effects in a spatial autoregressive panel model with a potentially large spatial dimension. 15 used the post-LASSO estimator with the ridge in the second stage for selecting the nested groups of the relevant genes from microarray data. 16 proposed the double LASSO and compared its performance with some estimators performing variable selection via simulation studies and applied it to Parents' Life Satisfaction data.

In the literature, there are various studies conducted to compare different linear regression methods through simulation studies. 17 conducted a simulation study for variable selection using the bootstrap method in principal component regression for high-dimensional data analysis. 18 conducted a study on the performance of penalized regression methods on high-dimensional datasets. 19 conducted a study on the comparison of convex penalized regression methods depending on the structure of the true coefficient vector in classical datasets.

This study investigates the performance of the post-LASSO two-stage method depending on the second-stage estimator, the characteristics of the true coefficient vector and the amount and location of the signals. The classical penalized regression methods the ridge, LASSO, elastic net, adaptive LASSO, and the post-LASSO methods are used for comparison. In the second stage of the post-LASSO, the OLS, ridge and LASSO methods are used respectively and the effect of the selected method in the second stage on the performance of post-LASSO is analyzed.

In Section 2 of the study, the penalized regression methods compared are summarized and the comparison criteria are given. The characteristics of the simulation studies used in the comparison are also mentioned. In Section 3 of the study, the findings obtained from the simulation studies are presented and the results are discussed in detail. In section 4 of the study, conclusions are given, and the study is completed.

2. Material and Method

2.1. Penalized Regression Methods

Since the OLS fails to predict new observations accurately depending on the structure of the dataset and fails to select variables, various methods based on the calculation of model coefficients under certain constraints have

been proposed as an alternative to the OLS. These methods are known as penalized regression methods and are widely used 20. These methods are used to obtain stable regression coefficient estimates by dealing with the problem of inflation of the variance of the coefficients. Some of the penalized regression methods can also perform automatic variable selection.

Let $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)$ be an independent and identically distributed dataset, where $\mathbf{x}_i$ is the i-th observation vector of size $p \times 1$ and y_i is the response of i-th observation. The objective function of the linear regression model given in Equation (1) is

$$Q(\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2/2n + p_\lambda(\boldsymbol{\beta}) \tag{3}$$

where $\boldsymbol{\beta}$ is the vector of unknown coefficients, $p_\lambda(\cdot)$ is the penalty function and λ is the tuning parameter. In these methods, coefficient estimates are obtained by minimizing the objective function given in Equation (3).

The OLS estimator in Equation (2) has the smallest variance in the class of unbiased estimators. However, in the case of multicollinearity in the dataset, The OLS estimates are far from being satisfactory. To overcome this problem, penalty functions that affect the coefficient estimates in different ways have been proposed in the literature.

Among penalized regression methods, there are one-stage methods where coefficient estimates are obtained directly or two-stage methods where coefficient estimates are obtained after two stages. One-stage and two-stage penalized regression methods have been proposed whose performance varies depending on the characteristics of the datasets.

In the literature, there are various penalized regression methods that shrink the regression coefficients to achieve higher prediction accuracy than the OLS. 4 proposed ridge regression as a method based on the trade-off between the bias and variance of regression coefficients. Ridge regression aims to overcome the problem of over-inflation of variance by compromising the unbiasedness of the model coefficients. In ridge regression, coefficient estimates are obtained by solving the problem

$$\widehat{\boldsymbol{\beta}}_R = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\{\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2/2n + \lambda\|\boldsymbol{\beta}\|_2^2\} \tag{4}$$

where $\lambda > 0$ is the tuning parameter. The larger λ, the greater the shrinkage of the coefficients. The ridge regression coefficient estimates are found as

$$\widehat{\boldsymbol{\beta}}_R = \left(\mathbf{X}^T\mathbf{X} + \lambda n\mathbf{I}_p\right)^{-1}\mathbf{X}^T\mathbf{y} \tag{5}$$

by solving the problem given in Equation (4) where $\mathbf{I}_p$ is the $p \times p$ identity matrix. If the dataset is ill-conditioned, ridge regression gives more accurate preliminary prediction values than EKK. Despite its prediction success, ridge regression cannot perform automatic variable selection. Therefore, estimators that can perform variable selection have been proposed as an alternative to ridge regression.

Considering the deficiency of ridge regression in variable selection, 10 proposed the LASSO method. In the LASSO method, coefficient estimates are obtained by solving the problem.

$$\widehat{\boldsymbol{\beta}}_L = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\{\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2/2n + \lambda\|\boldsymbol{\beta}\|_1\}. \tag{6}$$

Due to the ℓ_1 norm term in Equation (6), some of the LASSO coefficient estimates become zero for sufficiently large tuning parameter values. Therefore, the LASSO is not only a coefficient shrinkage method but also an automatic variable selection method. The problem in Equation (6) does not have a closed-form solution. Therefore, various algorithms have been proposed to obtain the LASSO estimates. The least angle regression 21, alternating direction method of multipliers 22, and coordinate descent 23 are some of the algorithms that can be used to obtain the LASSO estimates.

As the correlation between explanatory variables increases, the prediction success of LASSO regression reduces 10. In this case, an alternative to the LASSO is the elastic net (ENET) 11 method. ENET is obtained by adding an ℓ_2 norm term to the penalty function of the LASSO. The ENET estimator is obtained by solving the problem

$$\widehat{\boldsymbol{\beta}}_E = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\{\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2/2n + \lambda(\alpha\|\boldsymbol{\beta}\|_1 + (1-\alpha)\|\boldsymbol{\beta}\|_2^2)\} \tag{7}$$

where $0 \leq \alpha \leq 1$ is the second tuning parameter. The ridge and LASSO correspond to $\alpha = 1$ and $\alpha = 0$ respectively in the ENET. In addition, due to the ridge-type penalization term, the ENET has the ability to group highly correlated variables. Therefore, the ENET is expected to perform better than the LASSO when there are highly correlated variables in the dataset.

In cases where the LASSO is not consistent in variable selection, the adaptive LASSO (A-LASSO) proposed by 12 can be used to estimate the model coefficients. The A-LASSO is consistent in variable selection and is approximately minimax optimal 12. The A-LASSO is based on the principle of using adaptive weights in the penalty function of the LASSO. The coefficient estimates of the A-LASSO are obtained by solving the problem

$$\widehat{\boldsymbol{\beta}}_{A-LASSO} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \left\{ \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2/2n + \sum_{j=1}^{p} w_j|\beta_j| \right\} \tag{8}$$

where w_j, $i = 1,2,\dots,p$ are adaptive weights. 12, selected the vector of adaptive weights as $\widehat{\boldsymbol{w}} = 1/|\widehat{\boldsymbol{\beta}}_{ekk}|$. Accordingly, there are two stages for the calculation of A-LASSO:

1. Calculation of the EKK coefficient estimates and the vector of adaptive weights using these estimates.
2. Solving the problem in Equation (8) by reweighting the LASSO penalty function with the weights from stage 1.

The Post-LASSO estimator is a two-stage method like the A-LASSO. However, its stages are quite different from the stages of the A-LASSO. The stages of the Post-LASSO are

1. Finding LASSO coefficient estimates and detecting the signals by solving the minimization problem given in Equation (6),
2. Regression modeling of the response variable on the subset of the original dataset corresponding to the signals.

In this study, the OLS estimator given in Equation (2) (post-LO, 13), the ridge estimator given in Equation (5) (post-LR, 15) and the LASSO estimator given in Equation (6) (post-LL, 16) are used in the second stage of post-LASSO.

2.2. Simulation Studies and Comparison Criteria

In this study, classical penalized regression methods ridge, LASSO, ENET and A-LASSO, and two-stage post-LASSO type methods post-LO, post-LR and post-LL are compared through simulation studies. In the simulation studies, datasets are generated according to the model given in Equation (1) based on the method described in 10. In each simulation study, 100 datasets consisting of 50 explanatory variables are generated. In the simulation studies, the correlation between i. and j. explanatory variables is $\rho^{|i-j|}$ and the values of ρ are chosen as 0.5 and 0.7, which are commonly used in the literature 10, 24, 25. coefficients vector on the performance of the methods, simulation studies are classified into two groups.

S1. In the first group of simulation studies, the signals of the true coefficients vector precede the noise (represented by a zero-valued coefficient). In this group, the true coefficient vector has the form $\boldsymbol{\beta} = \left[\underbrace{1.5 \dots 1.5}_{s/2}, \underbrace{0.5 \dots 0.5}_{s/2}, \underbrace{0,0, \dots ,0}_{50-s}\right]$. The number of signals, s, is taken as $s = 5,10,25,50$ to represent different sparsity levels. As the value of s increases, the sparsity level reduces.

S2. In the second group of simulation studies, the magnitude of the signals is taken equal, and the number and position of the noises are changed to examine the effect of sparsity and position. In the simulation studies in this group, the real coefficient vector has the form $\boldsymbol{\beta} = [a, \mathbf{0}_k, a, \mathbf{0}_k, \dots, a, \mathbf{0}_k]$. Here a is the signal value and k is the dimension of the zero vector. The value of k represents the sparsity level of the true coefficient vector. In this study, $a = 0.5,1,5$ and $k = 1,4,9$.

For tuning parameter selection and calculation of performance criteria, it is common to decompose the dataset into training, validation and test sets 11, 26. First, the penalized regression methods to be compared are trained on the selected tuning parameter values. The mean squared error for these models is calculated using the validation set. The model with the smallest mean squared error in the validation set is determined as the best model. The performance of the models is compared with the mean squared error on the test set.

Table 1. Quality measures of methods for S1 group simulation studies.

ρ	s	Comparison Criterion	Ridge	LASSO	ENET	A-LASSO	Post-LO	Post-LR	Post-LL
0.5	5	Median of TMSE	2.2937	0.8678	1.1947	0.9517	2.1560	1.4641	0.7039
		Standard Deviation	0.08	0.10	0.07	0.06	0.08	0.10	0.05
		Active Set Size	50	9	17	6	9	9	9
0.5	10	Median of TMSE	3.1850	1.8068	2.0322	2.2592	3.1778	2.0364	1.4044
		Standard Deviation	0.12	0.07	0.08	0.08	0.13	0.08	0.10
		Active Set Size	50	18	26	12.5	18	18	18
0.5	25	Median of TMSE	3.9187	4.0861	3.4786	5.7695	5.4639	3.2915	3.3642
		Standard Deviation	0.15	0.13	0.14	0.26	0.22	0.13	0.11
		Active Set Size	50	30	36	26	30	30	30
0.5	50	Median of TMSE	4.8628	6.9273	5.4853	9.5477	7.5741	4.3638	6.4777
		Standard Deviation	0.14	0.24	0.11	0.37	0.23	0.15	0.15
		Active Set Size	50	46	47	43	46	46	46
0.7	5	Median of TMSE	1.9134	0.7022	0.9238	1.0306	1.7544	0.9893	0.4967
		Standard Deviation	0.08	0.05	0.06	0.06	0.12	0.09	0.04
		Active Set Size	50	8	15.5	6	8	8	8
0.7	10	Median of TMSE	2.4281	1.5319	1.5066	2.5802	2.8661	1.6802	1.1579
		Standard Deviation	0.07	0.06	0.05	0.13	0.14	0.08	0.08
		Active Set Size	50	16	24	13	16	16	15
0.7	25	Median of TMSE	2.7236	3.4224	2.4956	5.4789	5.3446	2.6762	3.1084
		Standard Deviation	0.09	0.17	0.09	0.32	0.25	0.14	0.16
		Active Set Size	50	27	34.5	25	27	27	27
0.7	50	Median of TMSE	3.4368	6.3575	4.3326	9.3805	8.7705	4.4902	7.1900
		Standard Deviation	0.15	0.29	0.18	0.30	0.35	0.16	0.25
		Active Set Size	50	44	47	40	44	44	44

In this study, independent training sets of 100 observations, validation sets of 100 observations and test sets of 400 observations are formed for each simulation run. The standard deviation of the errors in Equation (1) is taken $\sigma = 3$ as in 10. $\alpha = 0.5$ is chosen in order to observe the level of difference between the ridge, LASSO and ENET methods.

Various criteria are used to compare the performance of the models produced by the methods. Let $\mathbf{\Sigma}$, the covariance matrix of the explanatory variables and $\widetilde{\boldsymbol{\beta}}$, the coefficients vector of the related penalized regression method. The mean squared error on the test set

$$TMSE = (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})^T \boldsymbol{\Sigma} (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})$$

is used to compare the performance of the models in the prediction. The standard deviations of the TMSE values are also given. Finally, the active set sizes obtained based on each method are reported as an estimate of the number of signals in the true model.

3. Findings and Discussion

The results of the S1 group simulation studies are summarized in Table 1. According to Table 1, when the correlation level between variables is low and the sparsity level is high, the methods that can select variables give a better TMSE value than ridge regression. As the sparsity level reduces, the performance of these methods decreases compared to ridge regression. The ENET is the least affected method by the ridge-type penalty term and is always better than ridge except when the sparsity level is zero. At fixed correlation values, the TMSE value of all methods increases as the sparsity level reduces. When the correlation level is low, at least one post-LASSO type method gave a superior TMSE value compared to the other methods. The post-LL is better than the other methods when the sparsity level is high while the post-LR is better when it is low. The post-LO improved over the ridge at high sparsity, however, is dominated by the ridge as the sparsity level is reduced. As the correlation level increases, the sparsity level determines the performance of the post-LASSO methods. More precisely, in terms of TMSE, the post-LASSO type methods are superior in sparse models, while in other cases the ENET is superior to other methods.

In terms of active set size, A-LASSO always yielded the sparsest models. In most cases, the post-LASSO type methods have the same sparsity level as the LASSO while the ENET produced the densest models.

The line plots of the TMSE values obtained with the ridge, LASSO, ENET, A-LASSO, Post-LO, Post-LO, Post-LR and Post-LL penalized regression methods in the simulation studies in the S1 group are shown in Figure 1. The line plots support the inferences given in Table 1.

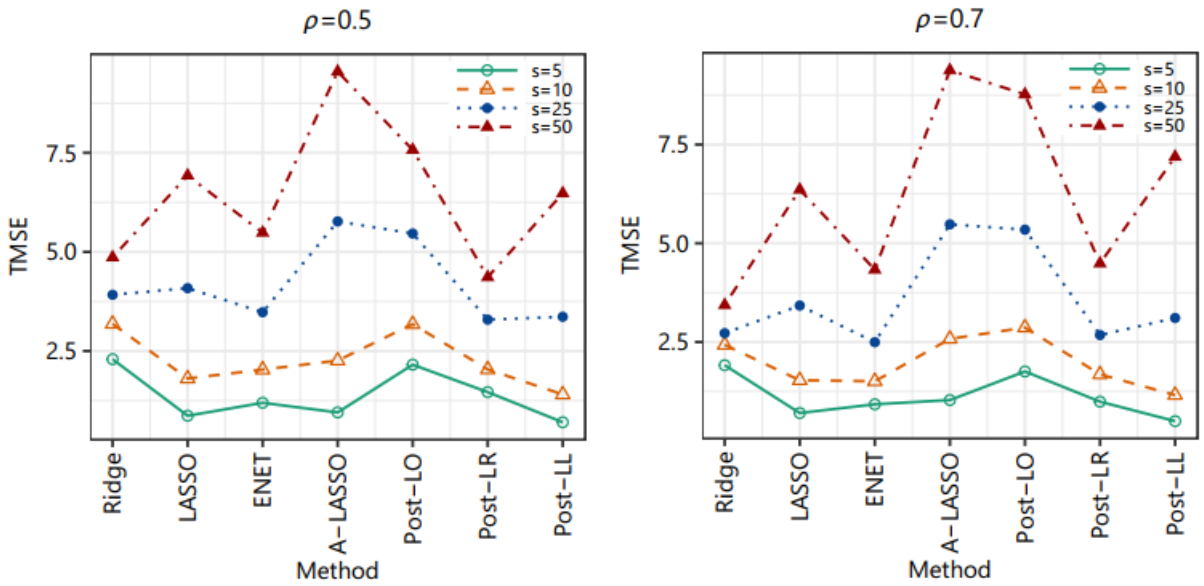

Figure 1. Line plots of TMSE values of the methods in S1 group simulation studies.

The results obtained from the S2 group simulation studies are given in Table 2,3,4. According to Table 2,3,4, when the correlation level is low and the signal value is small, at least one post-LASSO type method dominates the other methods in terms of TMSE. When the signal value is large, the A-LASSO method is more successful than the other methods at both correlation levels. As the correlation level increases, the post-LASSO type methods are superior in the case of a sparse model, while the ridge gives a better result when the model is dense. In sparse

model cases where the signal value is not small, the post-LL is superior to the post-LR, while the post-LR dominates the post-LL as the sparsity level decreases.

In terms of active set sizes, the models produced by the A-LASSO are sparser than those of the other methods. The LASSO and post-LASSO type methods mostly produced models with the same sparsity level.

Table 2. Quality measures of methods for S2 group simulation studies ($k = 1$).

ρ	a	Comparison Criterion	Ridge	LASSO	ENET	A-LASSO	Post-LO	Post-LR	Post-LL
0.5	0.5	Median of TMSE	2.5747	3.5612	3.0367	4.8732	4.7599	1.4948	2.2906
		Standard Deviation	0.09	0.09	0.08	0.09	0.18	0.08	0.11
		Active Set Size	50	29	35	22	29	29	29
0.5	1	Median of TMSE	4.7383	5.5140	4.9097	7.3843	6.7050	4.1501	4.9141
		Standard Deviation	0.12	0.18	0.13	0.31	0.26	0.14	0.22
		Active Set Size	50	39	43	34	39	39	39
0.5	5	Median of TMSE	11.2373	5.8086	8.7083	3.4424	7.3974	9.6831	5.8456
		Standard Deviation	0.47	0.20	0.35	0.11	0.26	0.44	0.21
		Active Set Size	50	41	49	27	41	41	41
0.7	0.5	Median of TMSE	1.8008	2.9484	2.3976	4.2599	6.0278	2.2575	3.3732
		Standard Deviation	0.05	0.10	0.05	0.17	0.20	0.14	0.14
		Active Set Size	50	29	35	20	29	29	29
0.7	1	Median of TMSE	3.6069	4.8343	3.9805	7.0343	7.0497	3.8334	4.9598
		Standard Deviation	0.09	0.15	0.15	0.29	0.55	0.19	0.21
		Active Set Size	50	37	43	32	37	37	37
0.7	5	Median of TMSE	11.3546	5.7803	8.7258	3.5676	7.2909	9.6696	5.7679
		Standard Deviation	0.35	0.16	0.41	0.16	0.28	0.37	0.18
		Active Set Size	50	40	48	27	40	40	40

Table 3. Quality measures of methods for S2 group simulation studies ($k = 4$).

ρ	a	Comparison Criterion	Ridge	LASSO	ENET	A-LASSO	Post-LO	Post-LR	Post-LL
0.5	0.5	Median of TMSE	1.5915	1.8618	1.7361	2.2162	2.6404	0.8594	0.9044
		Standard Deviation	0.07	0.05	0.04	0.08	0.18	0.07	0.09
		Active Set Size	50	15	19	7.5	16	16	15
0.5	1	Median of TMSE	3.3773	3.0304	2.8857	3.3551	4.2139	2.7540	2.3923
		Standard Deviation	0.07	0.17	0.14	0.16	0.26	0.12	0.10
		Active Set Size	50	24	32	17	24	24	24
0.5	5	Median of TMSE	8.8303	3.0302	6.7399	1.3167	4.7968	5.3831	2.8519
		Standard Deviation	0.35	0.15	0.26	0.07	0.28	0.29	0.17
		Active Set Size	50	25	46	12	25	25	25
0.7	0.5	Median of TMSE	1.2716	1.7106	1.5406	2.0603	3.0647	0.9003	1.1739
		Standard Deviation	0.03	0.03	0.03	0.06	0.25	0.07	0.07
		Active Set Size	50	16	22.5	11	16	16	16
0.7	1	Median of TMSE	2.7524	2.7390	2.6709	3.5810	4.4794	2.4284	2.4504
		Standard Deviation	0.08	0.07	0.08	0.11	0.26	0.08	0.13
		Active Set Size	50	24	31.5	19	24	24	24
0.7	5	Median of TMSE	8.4963	3.0106	6.7301	1.3705	4.5300	5.5549	2.9878
		Standard Deviation	0.20	0.10	0.33	0.07	0.16	0.21	0.13
		Active Set Size	50	26.5	45	12	26.5	26.5	25.5

Table 4. Quality measures of methods for S2 group simulation studies ($k = 9$).

ρ	a	Comparison Criterion	Ridge	LASSO	ENET	A-LASSO	Post-LO	Post-LR	Post-LL
0.5	0.5	Median of TMSE	0.9863	1.0334	1.0083	1.1286	2.1345	0.4742	0.5186
		Standard Deviation	0.03	0.06	0.05	0.04	0.20	0.06	0.08
		Active Set Size	50	7	10.5	4	10	10	7
0.5	1	Median of TMSE	2.3335	1.6554	1.7667	1.8022	3.086	1.8000	1.3230
		Standard Deviation	0.07	0.12	0.10	0.11	0.19	0.11	0.08
		Active Set Size	50	16	22	11	16	16	16
0.5	5	Median of TMSE	7.3924	1.6618	4.8912	0.6125	3.3622	3.5476	1.5437
		Standard Deviation	0.26	0.13	0.29	0.05	0.16	0.18	0.1
		Active Set Size	50	17	42	6	17	17	17
0.7	0.5	Median of TMSE	0.8695	1.0740	1.0122	1.1994	2.2889	0.5482	0.6073
		Standard Deviation	0.02	0.05	0.04	0.02	0.2	0.06	0.06
		Active Set Size	50	8	13	4	11	11	8
0.7	1	Median of TMSE	1.9657	1.6302	1.6825	2.1701	2.7410	1.6528	1.4061
		Standard Deviation	0.04	0.10	0.08	0.12	0.24	0.06	0.08
		Active Set Size	50	15.5	22.5	11	15.5	15.5	15
0.7	5	Median of TMSE	6.8640	1.6598	4.6428	0.6262	2.7658	3.2971	1.5978
		Standard Deviation	0.27	0.10	0.14	0.05	0.28	0.19	0.09
		Active Set Size	50	16	40	6	16	16	16

The line plots of the TMSE values obtained with the ridge, LASSO, ENET, A-LASSO, Post-LO, Post-LR and Post-LL in the S2 group simulation studies are given in Figures 2,3,4. The line plots are consistent with the analysis results given in Table 2,3,4.

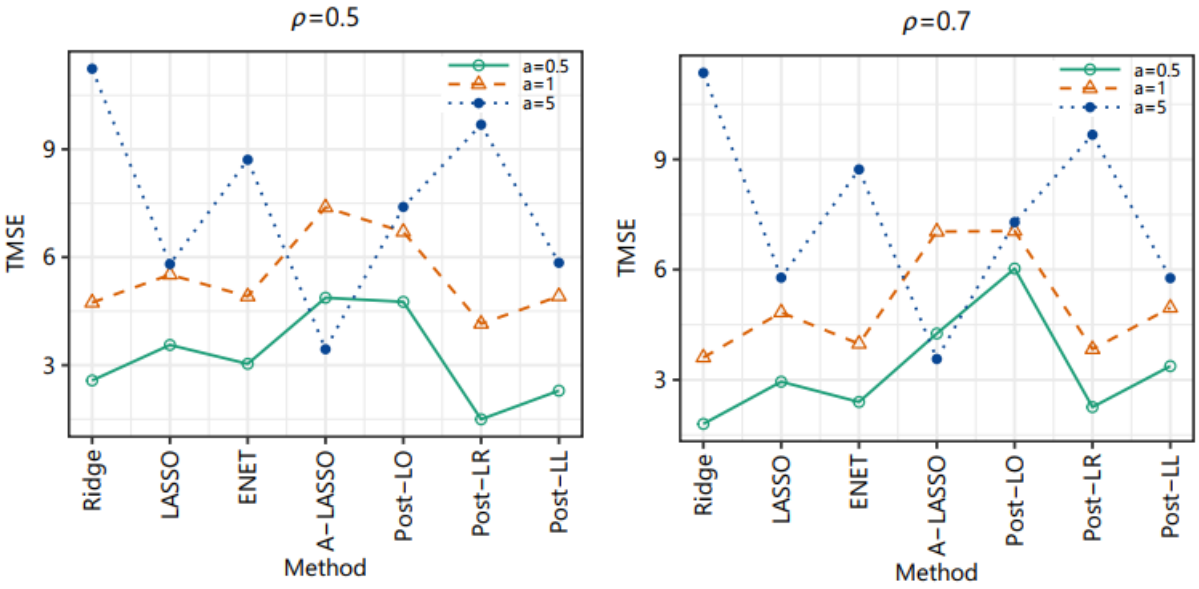

Figure 2. Line plots of TMSE values of the methods in S1 group simulation studies ($k = 1$).

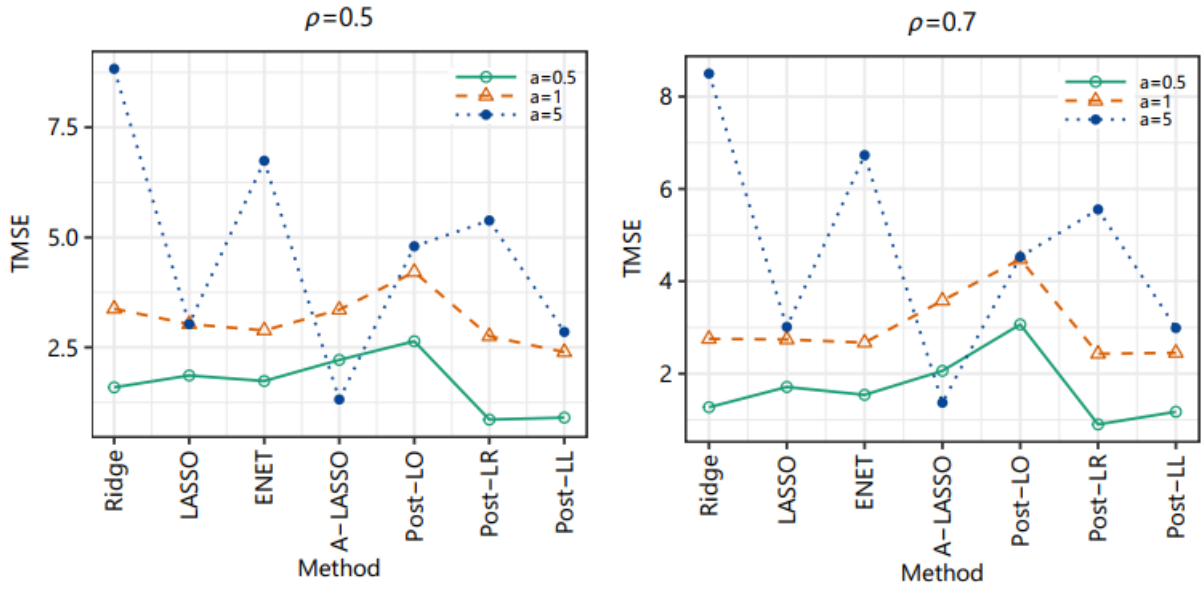

Figure 3. Line plots of TMSE values of the methods in S1 group simulation studies ($k = 4$).

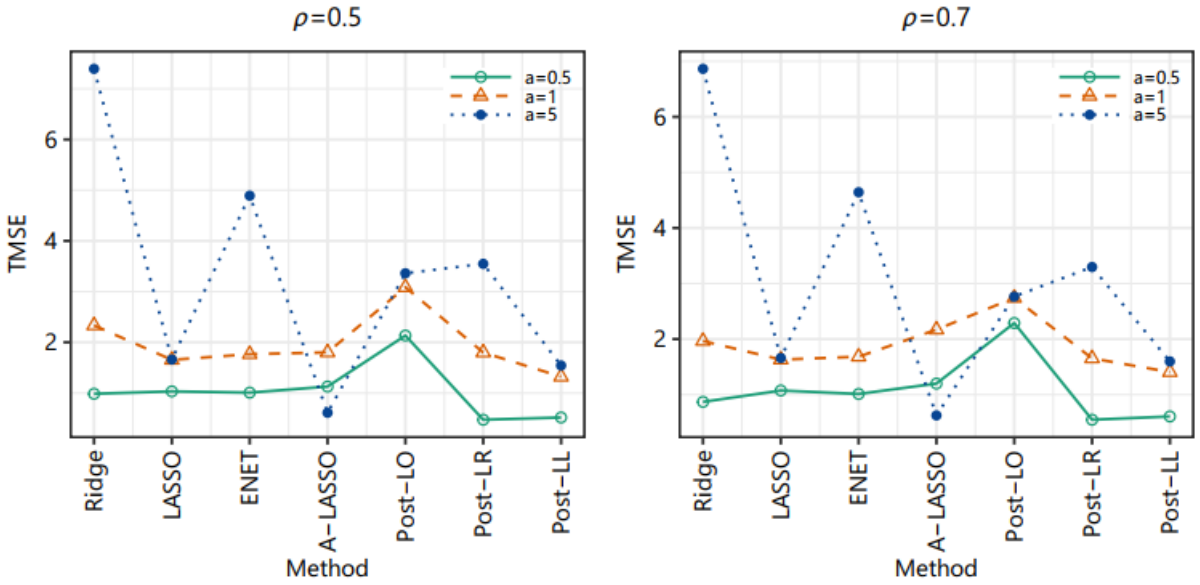

Figure 4. Line plots of TMSE values of the methods in S1 group simulation studies ($k = 9$).

4. Conclusions

In this study, the effect of the properties of the true coefficient vector on the performance of classical penalized regression methods and two-stage post-LASSO type methods is investigated. A detailed comparison of the ridge, LASSO, ENET and A-LASSO classical penalized regression methods with the post-LASSO type penalized regression methods, post-LO, post-LR, post-LR and post-LL are performed by considering the size and position of the signals.

According to the results obtained from the comparison criterion, the estimator in the second stage of the post-LASSO type methods is quite effective in the performance of these methods. In addition, the structure of the true coefficient vector of the model is very effective in the performance of classical and post-LASSO type methods. According to the active set sizes obtained by the post-LASSO type methods, the true coefficient vector and the properties of the dataset have an impact on the success of post-LASSO type methods in variable selection. With the simulation studies, the strengths and weaknesses of post-LASSO methods in terms of estimation and variable selection in models with different structures in terms of sparsity and signal magnitude are revealed.

References

[1] Montgomery DC, Runger GC, Hubele NF. Engineering Statistics. New York: John Wiley & Sons; 2009.

[2] Bzovsky S, Phillips MR, Guymer RH, Wykoff CC, Thabane L, Bhandari M, Chaudhary V. The clinician's guide to interpreting a regression analysis. Eye 2022; 36(9):1715-1717.

[3] Venkateshan SP. Mechanical Measurements. New York: John Wiley & Sons; 2015.

[4] Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 1970; 12(1):55-67.

[5] Liu K. Using Liu-type estimator to combat collinearity. Commun Stat - Theory Methods 2003; 32(5):1009-1020.

[6] Rao CR, Toutenburg H. Linear Models: Springer; 1995.

[7] Sarkar N. A new estimator combining the ridge regression and the restricted least squares methods of estimation. Commun Stat - Theory Methods 1992; 21(7):1987-2000.

[8] Breiman L. Better subset regression using the nonnegative garrote. Technometrics 1995; 37(4):373-384.

[9] Frank LE, Friedman JH. A statistical view of some chemometrics regression tools. Technometrics 1993; 35(2):109-135.

[10] Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B Methodol 1996; 58(1):267-288.

[11] Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B Methodol 2005; 67(2):301-320.

[12] Zou H. The adaptive lasso and its oracle properties. J Am Stat Assoc 2006; 101(476):1418-1429.

[13] Belloni A, Chernozhukov V. Least squares after model selection in high-dimensional sparse models. Bernoulli 2013; 19(2):521-547.

[14] Ahrens A, Bhattacharjee A. Two-step lasso estimation of the spatial weights matrix. Econometrics 2015; 3(1):128-155.

[15] De Mol C, Mosci S, Traskine M, Verri A. A regularized method for selecting nested groups of relevant genes from microarray data. J Comput Biol 2009; 16(5):677-690.

[16] Urminsky O, Hansen C, Chernozhukov V. Using double-lasso regression for principled variable selection. SSRN Working Paper No. 273374. 2016.

[17] Shahriari S, Faria S, Gonçalves AM. Variable selection methods in high-dimensional regression-A simulation study. Commun Stat - Simul Comput 2015; 44(10):2548-2561.

[18] Ahmed SE, Kim H, Yıldırım G, Yüzbaşı B. High-Dimensional Regression Under Correlated Design: An Extensive Simulation Study. International Workshop on Matrices and Statistics, Springer. 2016:145-175.

[19] Genç M. Bir Simülasyon Çalışması ile Cezalı Regresyon Yöntemlerinin Karşılaştırılması. Bilecik Şeyh Edebali Üniv Fen Bilim Derg 2022; 9(1):80-91.

[20] Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. New York: Springer series in statistics; 2001.

[21] Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. Ann Stat 2004; 32(2):407-499.

[22] Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. Found Trends Mach Learn 2011; 3(1):1-122.

[23] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010; 33(1):1-22.

[24] Chang L, Roberts S, Welsh A. Robust lasso regression using Tukey's biweight criterion. Technometrics 2018; 30(1):36-47.

[25] Chong IG, Jun CH. Performance of some variable selection methods when multicollinearity is present. Chemom Intell Lab Syst 2005; 78(1-2):103-112.

[26] Hussami N, Tibshirani RJ. A component lasso. Can J Stat 2015; 43(4):624-646.