Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 27, Issue 3 (2025) ISSN$_{Print}$: 1816-9171
DOI: https:// doi.org/ 10.33916/ qjae.2025 .03232240

The alasso and rlasso: A Comparative Study

Mohammed H. Al-Sharoot Mayyadah Aljasimee
Alsharoot@qu.edu.iq Mayyadahjawad3@gmail.com
University of Al-Qadisiyah

Article history:
Received: 20/12/2024
Accepted: 30/12/2024
Available online: 15 /9 /2025

Corresponding Author : Mayyadah Aljasimee

Abstract : Selecting variables and estimating the coefficients of the regression model have become one of the most important factors in prediction accuracy. The problem arises when we have to choose the true subset of covariates that have a significant effect on the dependent variable, including irrelevant covariates in the model, which can lead to overfitting. In this paper, we will use two approaches of regularization, the adaptive least absolute shrinkage and selection operator (alasso) and the reciprocal lasso (rlasso), to prevent overfitting in the statistical models, and compare the performance of the two approaches using a simulation by the MCMC method. The findings of the simulation and analyses of real data demonstrate that rlasso performs comparably in various simulation studies.

Keywords: High Dimensional data, Variable Selection(VS), Regularization, adaptive lasso (alasso), Reciprocal lasso (rlasso).

INTRODUCTION: In modern statistical modeling, especially in high-dimensional settings, selecting the most relevant covariates and estimating their effects accurately is a key challenge. Including irrelevant variables can lead to overfitting and poor prediction performance.

Regularization methods have emerged as effective tools to address this problem by simultaneously performing variable selection and coefficient estimation. Among them, the adaptive lasso (alasso) and reciprocal lasso (rlasso) offer promising approaches to construct parsimonious models.

This study focuses on comparing the performance of Alasso and Rlasso using simulation experiments and real data analysis. The goal is to assess their effectiveness in producing sparse, interpretable models with high predictive accuracy.

The outline is as follows: in Section 2, we introduce several regularization approaches used in estimating coefficients and choosing variables in a linear regression model in this study, for instance, the lasso, the adaptive lasso (alasso), and the reciprocal lasso (rlasso). In Section 3, we run simulation examples to investigate the performance of two regularization approaches, namely, Alasso and Rlasso. In Section 4, we explain our approach employing the air pollution dataset. Finally, in Section 5, we provide a brief discussion of results from simulation examples and the overall performance of alasso and rlasso.

1. Methodology

1.1. Regression shrinkage using Lasso

Assume the multiple linear regression model has the following definition:

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

where $\mathbf{y} = (y_1,\dots, y_n)'$ is the response vector, $\mathrm{X} = (\mathbf{x}_1,\dots, \mathbf{x}_p)$ is the matrix of covariates, $\boldsymbol{\beta} = (\beta_1,\dots, \beta_p)'$ a regression coefficient vector, and $\boldsymbol{\varepsilon} = (\varepsilon_1,\dots, \varepsilon_n)'$ a random errors vector, where the error distribution $\varepsilon_i \sim N(0,\ \sigma^2)$. According to model (1), only a small part of the possible covariates is thought to have an impact on the dependent variable, whereas other covariates are thought to have little effect or be unimportant. Hence, covariates that are unimportant must be removed from the model.

The Ordinary Least Squares (OLS) method has issues with poor prediction performance, overfitting, and difficulty interpreting models with large variances. To improve prediction accuracy and ease model interpretation, dealing with regression models that contain a large number of covariates requires special techniques, such as regularization, which can help reduce the number of covariates and prevent overfitting. Variable selection is necessary for a parsimonious regression model using only important covariates.

It is possible to estimate the regression coefficients by reducing

$$\min_{\boldsymbol{\beta}} \left\| y - \sum_{k=1}^{p} x_k \beta_k \right\|_2^2 = \min_{\boldsymbol{\beta}} (\mathbf{y} - \mathrm{X}\boldsymbol{\beta})' (\mathbf{y} - \mathrm{X}\boldsymbol{\beta}) , \qquad (2)$$

Regularization is a penalized regression method used for VS and coefficient estimation in regression issues. It is the least absolute shrinkage and selection operator (lasso) and was suggested by Tibshirani (1996) to solve the overfitting problem when there are several studies where the number of covariates exceeds the sample size (p>n). It adds $\lambda \sum_{k=1}^{p} |\beta_k|$ is a penalty function, it is symbolized by a symbol (ℓ_1 norm) to the least squares loss function, which reduces the coefficients of unimportant covariates to zero. Thus, achieving automatic variable selection. The lasso estimator $\hat{\beta}_{lasso} = (\hat{\beta}_0 , \hat{\beta}_1 , \hat{\beta}_3 , \ldots , \hat{\beta}_p)$ may be found as minimizing the following objective function:

$$\hat{\beta}_{lasso} = \min_{\boldsymbol{\beta}} (\mathbf{y} - \mathrm{X}\boldsymbol{\beta})' (\mathbf{y} - \mathrm{X}\boldsymbol{\beta}) + \lambda \sum_{k=1}^{p} |\beta_k| , \qquad (3)$$

The lasso model's $\hat{\beta}_{lasso}$ is determined by the regularization parameter λ, where $\lambda \geq 0$, with the highest value causing the highest shrinkage (Alkenani and Yu, 2013). Small penalties lead to models with high variance but less bias, while big penalties result in models with fewer covariates and lower variance. If it is $\lambda = 0$, this means that no coefficient has been reduced towards zero in the model and that the estimators of Lasso are equal to the estimators of ordinary least squares. λ selection is important in balancing the trade-off between model complexity and predictive accuracy.

The lasso regression technique works on the premise of reducing the sum of squares of residuals according to a constraint that indicates the absolute total of coefficients that are smaller than a specific constant.

Among the prominent characteristics of the lasso regression:

1. Multicollinearity can be solved through it.
2. **Sparsity**: It lowers the model's prediction error by setting the coefficients of unimportant covariates equal to zero; effectively removing some of the covariates from the model can successfully reduce variance without considerably raising bias (Ranstam and Cook, 2018). It deals with the regression model's high number of covariates, where p is greater than n , with p denoting the number of covariates and n the sample size.
3. **Consistency**: Although lasso does a good job of setting the coefficients of unimportant covariates to zero, the lasso approach has a bias towards estimating large coefficients, making it inconsistent. As all coefficients are penalized equally by this method.
4. **Computational Efficiency**: The Lasso estimator remains a popular method for selecting and regulating variables in linear regression models due to its simplicity and arithmetic efficiency, and its performance can be improved using appropriate tuning parameters and regularization techniques.

The lasso may be efficiently calculated using a fast algorithm (LARS). Proposed by (Efron et al., 2004).

Additionally, it lacks oracle properties, which allow a method to select the true model with a probability of one quantity, which means that they may not always accurately identify the proper set of covariates that have an effect on the dependent variable. This can lead to the inclusion of irrelevant covariates in the model, which may reduce the model's predictive accuracy and interpretability. The lack of Oracle properties in Lasso is a known limitation of the approach, and researchers have proposed several modifications and additions to address this problem.

1.2. Regression shrinkage using adaptive lasso

The asymmetry of the lasso estimator in VS has been a topic of discussion in the statistical community. Inconsistency refers to the situation where the estimator fails to converge to the true value of the coefficient as the sample size increases; this can lead to biased estimates and poor prediction accuracy in some cases. Several modifications and extensions of the Lasso method have been proposed to address this problem.

In 2006, Zou demonstrated the inconsistency of the lasso estimator in VS. However, the lasso does a good job of reducing the coefficients of insignificant variables to zero, which effectively reduces the influence of unimportant covariates on the dependent variable.

In addition to the selection of variables, the Lasso approach can also be used to simultaneously estimate the coefficients of selected covariates, which makes it a popular choice for researchers who work with high-dimensional data. However, it is important to note that the lasso approach may not always be the best choice for every problem, and other regularization methods may perform better in certain situations.

Zou (2006) developed a novel regularization technique to overcome this issue. The adaptive lasso method is newer than the lasso method. In the proposed method, different regularization weights are assumed for different coefficients. This means that some coefficients are penalized more severely than others, depending on their importance in the model. The approach is known as the adaptive least absolute shrinkage and selection operator (alasso) method. However, lasso estimates are known to be biased towards large coefficients, which means that they tend to overestimate the importance of some covariates. The reason for this bias is that the lasso penalizes all coefficients equally, regardless of their size or importance.

The alasso method addresses this bias by adding adaptive weights to the lasso estimate. These adaptive weights are used to penalize different coefficients in the lasso method, which helps control bias. As a result, the coefficients of the uninteresting variables are more effectively reduced to 0, and the model becomes simpler and more interpretable. It can also improve the accuracy of predictions by reducing overfitting.

Zou's study was important because it highlighted the limitations of the lasso estimator in VS and led to the development of alternative regularization approaches that may perform better in certain situations. In addition to producing estimates that are consistent and unbiased, this technique performs better at estimating significant coefficients than lasso (Zou, 2006; Wang et al., 2007). It also minimizes bias and enhances variable selection accuracy. This means that the method is better at identifying covariates that have a significant effect on the dependent variable, leading to more accurate predictions.

The alasso may be solved using a similarly effective equation as the lasso solved by adding weight. The alasso estimator is defined by:

$$\hat{\beta}_{alasso} = \min_{\boldsymbol{\beta}} (\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta}) + \lambda \sum_{k=1}^{p} \hat{w}_k |\beta_k| , \tag{4}$$

where $\lambda \sum_{k=1}^{p} \hat{w}_k |\beta_k|$ indicates the alasso penalty.

$\hat{\boldsymbol{w}}_k = (\hat{w}_1, \hat{w}_2, \dots, \hat{w}_p)$ represents the adaptive weight vector, the weight assigned to each coefficient, which represents the significance or contribution of each covariate in predicting the dependent variable in the multiple linear regression model, is defined as follows:

$$\hat{\boldsymbol{w}}_k = \frac{1}{|\hat{\boldsymbol{\beta}}_k|^{\gamma}} , \tag{5}$$

for $k = 1, \dots, p$ and γ greater than zero. The parameter γ is a tuned parameter that controls the amount of shrinkage (Zou, 2006), which might be calculated using the cross-validation technique and $|\beta_k|$ represents the absolute value of the k-th coefficient estimate of the covariate. This adaptive penalty function helps reduce the regression coefficients towards zero, thus reducing the bias towards large coefficients, reducing model complexity, and preventing overfitting.

By penalizing large and small coefficients differently, the adaptive lasso produces less biased estimates.

The alasso penalty has been shown to perform well in simulation studies and real-world applications, particularly when covariates are of varying degrees of significance or relevance to the dependent variable.

The alasso regression has advantages:

1. According to Zou (2006), it is more attractive computationally. This is an important consideration in statistical modeling, where more complex and computationally intensive methods may not be practical or feasible for certain applications.
2. It may be solved by using the identical efficient algorithm which was used to solve the lasso, i.e., the LARS algorithm, as proposed by Efron et al., (2004).
3. It possesses oracle properties. The alasso approach also includes asymptotic guarantees of unbiasedness and normality, which means that as sample size increases, coefficient estimates become more accurate and follow a normal distribution by using the ℓ_1 penalty, which has been adaptively weighted as proposed by Zou (2006).

This method has been used in various fields such as genetics, finance, and image processing, where high-dimensional data is common. (See, Zhang and Lu, 2007; Zeng et al., 2014; Yang and Wu, 2016).

The alasso method has been shown to perform better than the lasso method in some simulation studies, especially when the number of covariates is large and the real model is sparse. However, it requires consistent initial estimates of the regression coefficients to function effectively, which are frequently lacking in high-dimensional, small sample size settings. Thus, the alasso method can be more computationally demanding than the lasso method because it requires the estimation of additional parameters.

In such cases, it is difficult to obtain reliable estimates of the regression coefficients, which are essential for the model to be accurate. Lack of accurate first estimates can lead to biased or inconsistent results, which may affect the validity of the model. However, in practice, obtaining such estimates can be difficult, and researchers often have to rely on alternative methods, such as cross-validation or smoothing, to obtain reliable estimates.

Furthermore, none of the algorithms used to calculate the alasso estimators are able to provide a correct estimate of the standard error (standard error is a measure of the variance of a statistic, such as the mean or regression coefficient, and is used to evaluate the accuracy of an estimate), is a limitation of the method.

1.3. Regression shrinkage using reciprocal lasso

In the year 2015, Song and Liang proposed a method known as rlasso, which is aimed at preventing the occurrence of overfitted models. The Rlasso method was developed for variable selection (VS) and the estimation of coefficients simultaneously. This was achieved through the utilization of a novel class of penalty functions that show discontinuity at 0. In addition, these penalty functions are observed to decrease in $(0,\infty)$ and result in coefficients that are close to 0 under infinite penalties. Because of this characteristic property, rlasso is very desirable for model selection, as noted by Song and Liang (2015) and Song (2018).

The Lasso, a regularization approach that imposes a constraint, represents the reciprocal of the sum of the absolute values of the regression model parameters. The rlasso estimator is obtained by minimizing :

$$\hat{\beta}_{rlasso} = \min_{\boldsymbol{\beta}} (\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta}) + \lambda \sum_{k=1}^{p} \frac{1}{|\beta_k|} I\{\beta_k \neq 0\}, \quad (6)$$

is used to estimate the coefficient values (represented in beta) for a multiple linear regression model. $\lambda > 0$ represents the tuning parameter. This parameter plays a crucial role in controlling the degree of penalization. Additionally, $I\{\beta_k \neq 0\}$ represents an indicator function, which means that the coefficient of the kth covariate in the linear regression model is not equal to zero. This notation is commonly used in regularization methods such as Rlasso, where the goal is to select the subset of covariates with non-zero coefficients that have the greatest influence on the dependent variable.

$\lambda \sum_{k=1}^{p} \frac{1}{|\beta_k|} I\{\beta_k \neq 0\}$, a penalty term that shrinks the coefficients towards zero and selects the subset of covariates with the greatest predictive power. It is worth mentioning that the value of λ plays a pivotal role in determining the degree of penalization, as well as the magnitude of the resulting shrinkage. In particular, it is well known that the lowest value of λ yields the highest level of shrinkage, and therefore yields a set of coefficients that are nearer to zero.

The rlasso approach differs from the lasso approach in that it uses a reciprocal penalty term instead of an absolute value penalty term. The term reciprocal penalty places more emphasis on shrinking small coefficients to zero and less emphasis on reducing large coefficients to zero.

Compared to the lasso penalty, which appears a nondecreasing nature in the interval $(0, \infty)$ and continuous, the penalty of rlasso a decreasing trend within the same interval $(0, \infty)$ while also being discontinuous at zero. Furthermore, it is worth noting that while the lasso penalty results in coefficients that are almost zero with corresponding penalties of zero, the rlasso, on the other hand, produces coefficients that are almost zero with penalties of infinity.

In theory, rlasso possesses the oracle property (Mallick et al., 2021), which means that it can identify the true subset of covariates with a high probability. However, rlasso is a lot more computationally intensive than lasso, which can make it less practical for large data sets (Song and Liang, 2015). In terms of variable selection performance, rlasso can perform better than lasso, but this comes at the cost of increased computational complexity.

The following figure (1) represents the behavior of the lasso and rlasso functions:

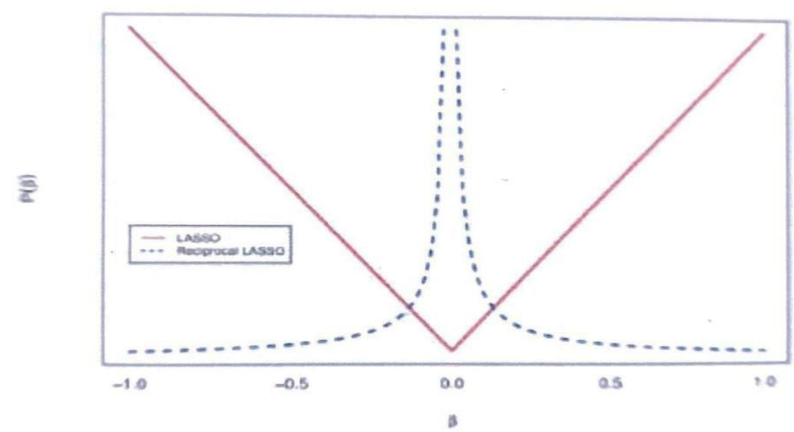

Figure (1) shows the lasso and rlasso functions (Mallick et al., 2021).

2. Simulation Study

A simulation study is conducted to evaluate and compare the performance of the Adaptive Lasso (alasso) and Reciprocal Lasso (rlasso) methods. The goal is to assess their estimation accuracy and variable selection capability in sparse settings.

We consider three different simulation models, each with 15 independent variables generated from a standard normal distribution, i.e., $X_j \sim \mathrm{Normal}(0,1), j = 1, \dots, 15$. Among the regression coefficients, only three are non-zero:

$$\beta_1 = 1, \qquad \beta_4 = 1.5, \qquad \text{and} \quad \beta_5 = 3$$

while the remaining coefficients are set to zero. This sparse configuration allows for evaluation of both estimation precision and variable selection.
To assess robustness, the error term ε is drawn from five different distributions:

- Standard normal: $N(0,1)$,
- Moderate-variance normal: $N(0,4)$,
- High-variance normal: $N(0,9)$,
- Heavy-tailed: Student's t-distribution with 3 degrees of freedom $t(3)$,
- Skewed distribution: Chi-squared distribution with 3 degrees of freedom $\chi^2(3)$.

For each scenario (model and error distribution), we perform 100 replications to ensure statistical reliability.
Model performance is evaluated using:

- Median of Mean Absolute Errors (MMAE), and
- Standard Deviation of Mean Absolute Errors (MAD).

These metrics provide insight into both the accuracy and stability of the estimation procedures across different conditions.
We note from Tables 1, 2 and 3 that the alasso method is better than the alasso method through the MMAD criterion, as it showed that this method is a robust method, as it gave the lowest values for the MMAD criterion in skewed distributions.

Table 1: the median of mean absolute deviations (MMAD), and the standard deviations of mean absolute deviation SD(MAD)

Simulation 1

Error distribution	Methods	Model 1	
		MMAD	SD(MAD)
N(0,1)	alasso	1.53904	0.87281
	rlasso	1.83865	0.75675
N(0,4)	alasso	1.78862	0.89810
	rlasso	1.71689	0.84173
N(0,9)	alasso	2.87465	1.03504
	rlasso	1.70674	0.83024
$t_{(3)}$	alasso	1.17773	0.59312
	rlasso	1.14482	0.42466
$\chi^2_{(3)}$	alasso	1.61569	0.62789
	rlasso	1.42605	0.55802

Table 2: the median of mean absolute deviations (MMAD), and the standard deviations of mean absolute deviation SD(MAD)

Simulation 2

Error distribution	Methods	Model 2	
		MMAD	SD(MAD)
N(0,1)	alasso	2.32569	0.89809
	rlasso	1.91967	0.79771
N(0,4)	alasso	1.98358	0.78282
	rlasso	1.79064	0.68584

N(0,9)	alasso	1.32405	0.68501
	rlasso	1.99527	0.44940
$t_{(3)}$	alasso	1.53569	0.77543
	rlasso	1.97139	0.72396
$\chi^2_{(3)}$	alasso	2.88228	0.88892
	rlasso	1.73949	0.77577

Table 3: the median of mean absolute deviations (MMAD), and the standard deviations of mean absolute deviation SD(MAD)

Simulation 3

Error distribution	Methods	Model 3	
		MMAD	SD(MAD)
N(0,1)	alasso	1.92215	0.72107
	rlasso	1.66564	0.57919
N(0,4)	alasso	1.85750	0.98880
	rlasso	1.03515	0.78573
N(0,9)	alasso	1.59801	0.93654
	rlasso	1.88505	1.14793
$t_{(3)}$	alasso	1.09295	0.88833
	rlasso	1.08005	0.63499
$\chi^2_{(3)}$	alasso	1.22960	0.79344
	rlasso	1.71930	0.64986

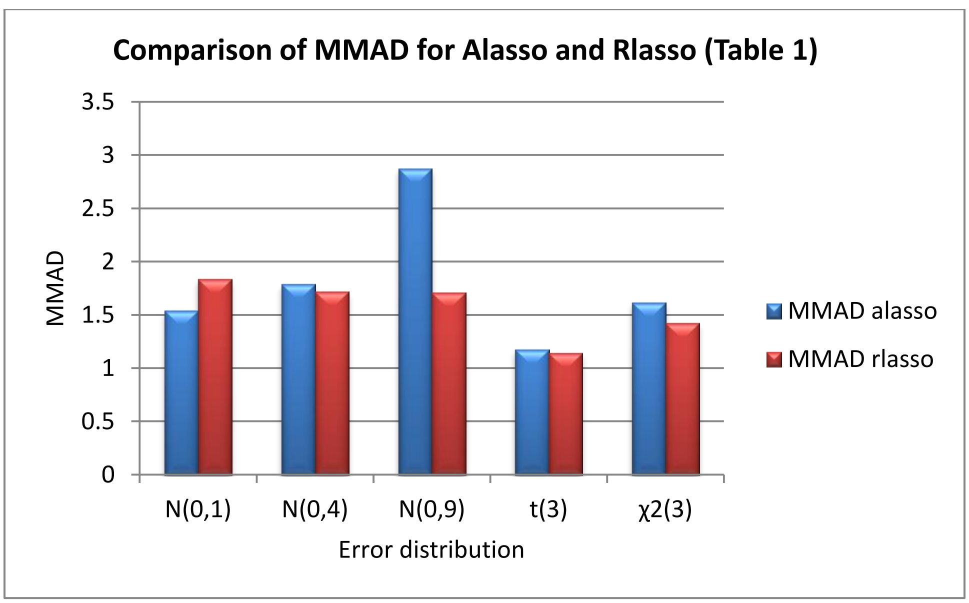

Figure 2. Comparison of MMAD for alasso and rlasso methods across different error distributions (Table 1).

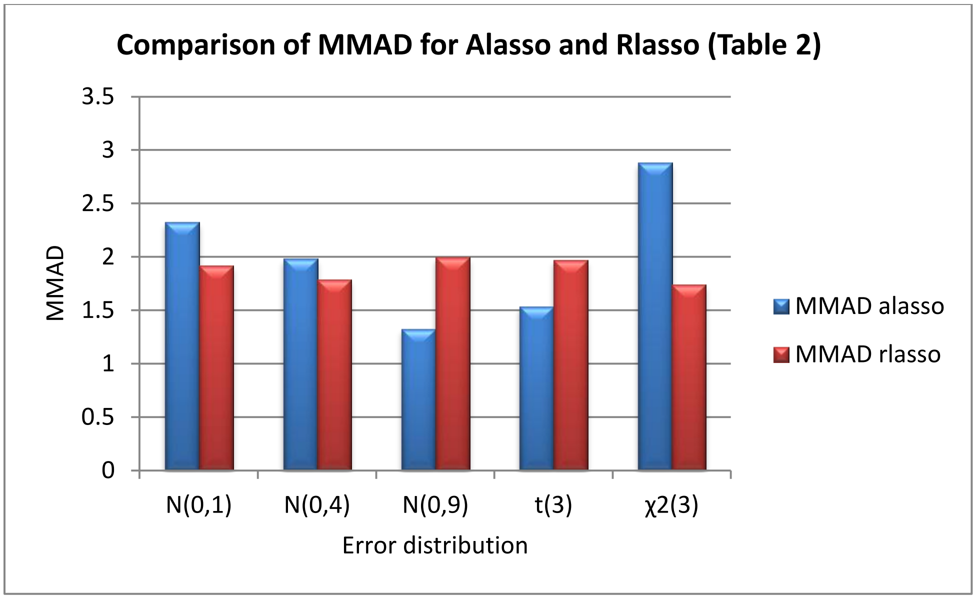

Figure 3. Comparison of MMAD for alasso and rlasso methods across different error distributions (Table 2).

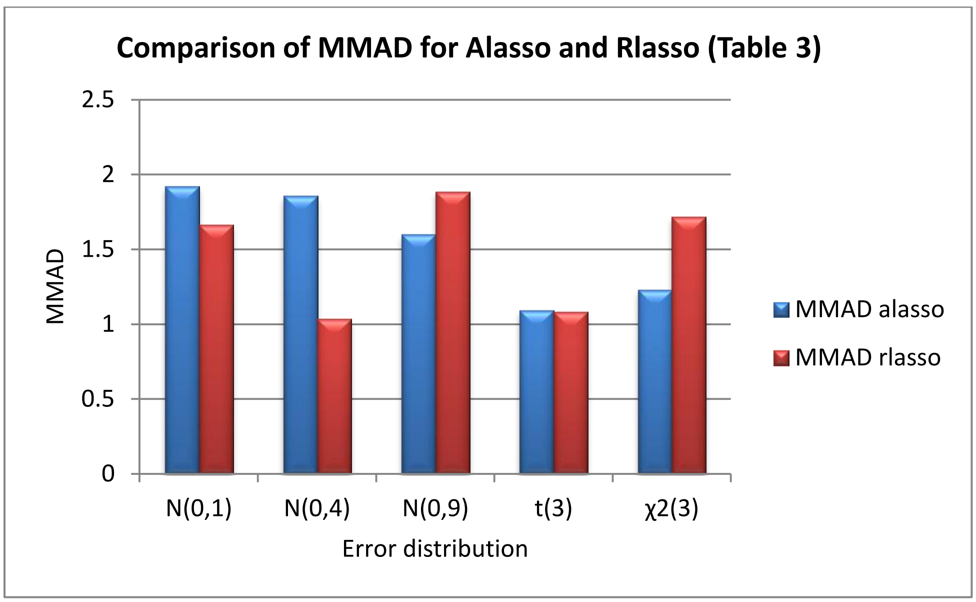

Figure 4. Comparison of MMAD for alasso and rlasso methods across different error distributions (Table 3).

The plot shows that rlasso consistently produces lower MMAD values than alasso, indicating superior estimation accuracy under all considered conditions.

3. **Real Data Application**

In this section, the air pollution dataset (Lindmark, A. and Karlsson, M. (2011)) is considered to compare the performance of our methods under study, alasso and rlasso. This dataset consists of 500 observations of air pollution recorded by the Administration of Norwegian public Roads. The response variable (y) shows the logarithm of the concentration of PM10. These data consist of seven independent variables, can be shown as follows:

X_1: show the log of the number of cars.
X_2: show the Temperature degree.
X_3: show the speed of wind.
X_4: The difference between the temperature of 25 and 2 meters above ground.
X_5: show the direction of wind.
X_6: show the hour of day.
X7: show the number of day from the first of October 2001.

Two criterions MSE and MAD are used to illustrate these methods under study as following table

Table 4: the mean absolute deviations (MAD), and the standard deviations of mean absolute deviation SD (MAD)

Methods	MSE	MAD
alasso	2.53904	1.8221
rlasso	2.33865	1.7285

We note from Tables 4 that the rlasso method is better than the alasso method through Two criterions MSE and MAD.

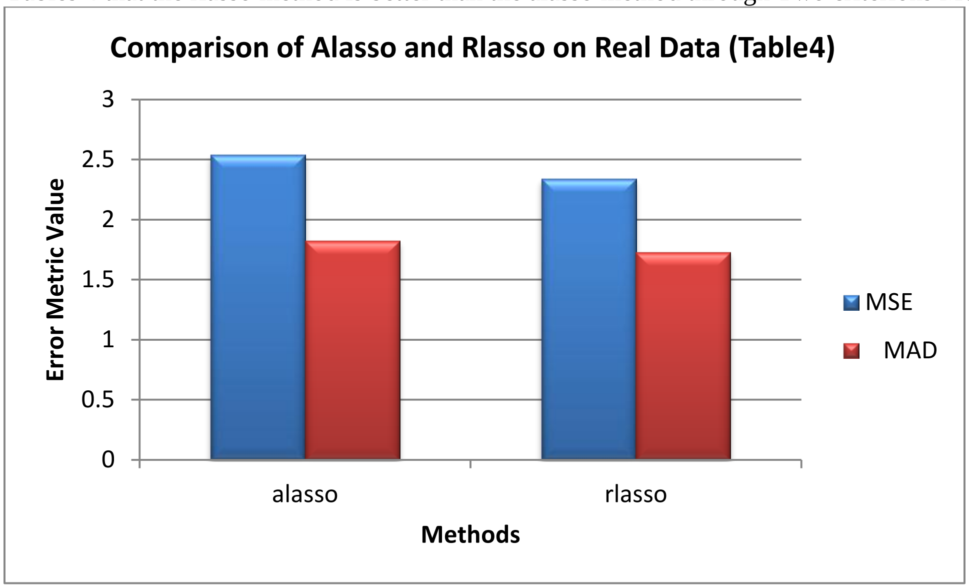

Figure 5. Comparison of MSE and mean absolute deviation (MAD) for alasso and rlasso. The results show that rlasso outperforms alasso by achieving lower error metrics, indicating improved estimation accuracy and model sparsity in a real-data context.

4. Conclusion

Regularization approaches have evolved throughout time to address the difficulties associated with studying high-dimensional data. In this paper, we describe some regularization approaches for variable selection and coefficient estimation in linear regression: lasso, rlasso, alasso, and. Also, we compare the performance of two regularization methods in this study: Alasso and Rlasso. The results are listed in Tables 1, 2,3 and 4. Our results show that rlasso performs better than alasso in selecting a high-dimensional model in various simulation studies and real data applications.

References

1. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267–288.
2. Alkenani, A. and Yu, K. (2013). Sparse MAVE with oracle penalties. Advances and Applications in Statistics 34,pp. 85–105.
3. Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. The Annals of statistics, 32(2), 407-499.
4. Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American statistical association, 101(476), 1418-1429.
5. Wang, H., Li, R., & Tsai, C. L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. Biometrika, 94(3), 553–568.
6. Zhang, H. H., & Lu, W. (2007). Adaptive Lasso for Cox's proportional hazards model. Biometrika, 94(3), 691-703.
7. Zeng, P., Wei, Y., Zhao, Y., Liu, J., Liu, L., Zhang, R., ... & Chen, F. (2014). Variable selection approach for zero-inflated count data via adaptive lasso.
8. Yang, Y., & Wu, L. (2016). Nonnegative adaptive lasso for ultra-high dimensional regression models and a two-stage method applied in financial modeling. Journal of Statistical Planning and Inference, 174, 52-67.

9. Song, Q., & Liang, F. (2015). High-dimensional variable selection with reciprocal l 1-regularization. Journal of the American Statistical Association, 110(512), 1607-1620.
10. Song, Q. (2018). An overview of reciprocal L 1-regularization for high dimensional regression data. Wiley Interdisciplinary Reviews: Computational Statistics, 10(1), e1416.
11. Ranstam, J., & Cook, J. A. (2018). LASSO regression. Journal of British Surgery, 105(10), 1348-1348.
12. Journal of Applied Statistics, 41(4), 879-894.
13. Mallick, H., Alhamzawi, R., Paul, E., & Svetnik, V. (2021). The reciprocal Bayesian lasso. Statistics in medicine, 40(22), 4830-4849.
14. Lindmark, A. and Karlsson, M. (2011). truncSP: Semi-parametric estimators of truncated regression models. R package version 1.1.