The lasso and rlasso: A Comparative study

Mayyadah Aljasimee [1], Rahim Alhamzawi [2]

[1, 2] Department of Statistics, College of Administration & Economics, University of Al-Qadisiyah, Iraq

[1]Mayyadahjawad3@gmail.com, [2]Rahim.alhamzawi@qu.edu.iq

ABSTRACT

Linear multiple regression models are the most widely used statistical models to illustrate the effect of a set of covariates on an outcome of interest. However, only a small number of covariates actually has an influence on the outcome of interest. The problem of choosing the true subset of covariates within a multiple linear regression model has received considerable attention over the years. In this paper, we compare the performance of two regularization approaches in this study: the least absolute shrinkage and selection operator (lasso) and the reciprocal lasso (rlasso). Simulation results show that both approaches outperform in terms of prediction accuracy. The results of both approaches (lasso and rlasso) are very similar. Our outcomes demonstrate that lasso and rlasso perform comparably in various simulation studies..

Keywords: High Dimensional data, Variable Selection, Regularization, lasso, Reciprocal lasso

1. Introduction

Suppose that model of the multiple linear regression is defined as

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

where $\mathbf{y} = (y_1,..., y_n)'$ is the response vector, $\mathrm{X} = (\mathbf{x}_1,..., \mathbf{x}_p)$ is the matrix of covariates, $\boldsymbol{\beta} = (\beta_1,... ,\beta_p)'$ a regression coefficient vector, and $\boldsymbol{\varepsilon} = (\varepsilon_1,..., \varepsilon_n)'$ a random errors vector where the error distribution $\varepsilon_i \sim N(0, \sigma^2)$. Under the

model (1), it's supposed that just a small subset of possible covariates has an effect on the dependent variable, while some covariates are not important (not effective or with very little effect), so not important covariates should be removed from the model (without simultaneously losing a lot of information). High dimensional problem has appeared in many regression applications, i.e., the number of covariates (p) is bigger than the sample size $(p > n)$, where n denotes that the sample size. The covariates selection is a critical issue in the building of a regression model. Selecting a suitable subset of covariates may typically increase prediction accuracy. The traditional statistical methods with this problem become not possible to use for statistical analysis. Thus, the difficulty of estimating coefficients and selecting the important covariates. So, high dimensional problem may result in extremely complex models.

One of the methods to reduce high dimensional data is the variable selection (VS) method. Various methods for handling VS in high dimensional linear models have been developed through the years to obtain a model with the fewest important covariates, high prediction accuracy, and ease of interpretation of the model, as well as providing model with a low cost (Guyon and Elisseeff, 2003).

There are two methods for the variable selection process: classical model selection methods for linear models, contrasting a set of candidate models, like Mallows's C_p (Mallows, 1973), Akaike information criterion (AIC; Akaike, 1974), Bayesian information criterion (BIC; Schwarz, 1978), Deviance Information Criteria (DIC; Spiegelhalter et al., 2002), and the Stochastic Search Variable Selection (SSVS; George and McCulloch, 1993).

Other new approaches to the variable selection process are the regularization methods that have been proposed for addressing the problem of model complexity through penalizing models of higher complexity. Shrinkage regression methods (also known as regularization methods or penalized likelihood) have been proposed

recently to address the overfitting issue in high dimensional linear models. Thus, variable selection and coefficient estimate may be done at the same time.

The model with high complexity has low bias and high variance, but the low complexity model has high bias and low variance. As a result, regularization approaches are frequently used to control the model complexity. Among the new methods involve the Ridge regression (Hoerl and Kennard, 1970), the bridge regression (Frank and Friedman, 1993), lasso (Tibshirani, 1996), smoothly clipped absolute deviation (SCAD; Fan and Li, 2001), the elastic net regression (Zou and Hastie, 2005), fused lasso (Tibshirani et al., 2005), adaptive lasso (alasso; Zou, 2006), group lasso (Yuan and Lin, 2006), dantzig selector (Candes and Tao, 2007), the adaptive group lasso (Wang and Leng, 2008), a self-adaptive lasso (alasso; Kang and Guo, 2009), adaptive elastic net (Zou and Zhang, 2009), minimax concave penalty (MCP; Zhang, 2010), matrix completion (Cand`es and Tao, 2010; Mazumder et al., 2011), and standardized group lasso (Simon and Tibshirani, 2012), among others.

Since lasso doesn't have oracle properties (Fan and Li, 2001), Song (2014) was the first to study the rlasso estimators with the oracle property. The rlasso approach suggested by Song and Liang (2015), as well as Song (2018) for VS and the coefficients estimate together, which is based on a novel class of penalty functions that are discontinuous at 0, decreasing in $(0,\infty)$ and also give near 0 coefficients of infinity penalties, while the traditional penalty functions are close to 0 penalties when the coefficients are close to 0 (for example, lasso and SCAD) or constant penalties (for example, ℓ_0 - penalty). Because of this characteristic property, rlasso is highly desirable for selecting models. It is able to successfully avoiding model selection that are too dense.

Similarly, from a Bayesian viewpoint, several researchers have used the Bayesian process in their approaches (see, Alhamzawi et al., 2011 ; Alshaybawee et al., 2017; Alhamzawi and Ali, 2018a, 2020) because a Bayesian method provides

an exact inference even with a small sample size, as well as an exact estimate when p is bigger than n ($p > n$, Li et al., 2010; Alhamzawi and Ali, 2018b). When identical and independent Laplace prior is assigned to every coefficient of regression, the estimations of lasso may be interpreted as a Bayesian posterior mode estimation (Tibshirani, 1996). In consequence, Park and Casella (2008) proposed the Bayesian lasso (Blasso) for models of linear regression, which uses the Laplace prior like a mixture of exponential and normal priors.

After Park and Casella (2008), more Blasso techniques have been suggested by researchers through the years (see, Huang et al., 2008; Yi and Xu, 2008; Brown and Griffin, 2010; Li et al., 2011; Legarra et al., 2011; Cai et al., 2011). Also, Blasso regression was introduced by Hans (2009). The Bayesian alasso (Balasso) and the iterative alasso were developed by Sun et al. (2010) using different adaptive weights and updating these adaptive weights iteratively. The iterative alasso approach is also much more computationally efficient than the widely utilized stepwise regression and marginal regression approaches.

Leng et al. (2014) presented the Balasso method for VS as well as an estimate of the coefficient in linear regression. Furthermore, prompted by the hierarchical Bayesian interpretation of the lasso, they gave selecting models mechanism for the Balasso through evaluating the posterior conditional mode estimations. Also in 2018, Alhamzawi and Ali introduced several Bayesian methods for obtaining the alasso as well as relevant estimators to address the drawbacks of the traditional adaptive lasso method. They considered a fully Bayesian approach to treating the alasso, which leads to a novel Gibbs sampler which has tractable full conditional posteriors. They used a scale mixture of truncated normal (SMTN) representation of the Laplace distribution to propose a novel hierarchy representation of Balasso.

Mallick et al. (2021) considered a fully Bayesian approach to the reciprocal lasso issue, based the observation that whenever the parameters of regression are allocated independent inverse Laplace priors, the rlasso estimation for linear

regression parameters may be interpreted as a Bayesian posterior mode estimation. Among the new Bayesian methods involve are a Bayesian elastic net (Bornn et al., 2010; Li and Lin, 2010; Alhamzawi, 2014; Huang et al., 2015), reversible - jump the MCMC (Chen et al., 2011), a new Bayesian lasso (Malik and Yi, 2014; Flaih et al., 2020).

In this paper, we introduce some regularization methods used in estimating the coefficients and selecting the variables in the linear regression model, such as lasso, reciprocal lasso (rlasso), Bayesian rlasso (Brlasso), adaptive lasso (alasso), and Bayesian adaptive lasso (Balasso) in Section 2. Also, we run simulation examples to investigate the performance of lasso as well as rlasso approaches in Section 3. Finally, in Section 4, we provide a summarized discussion.

2. Methods

2.1. Regularization regression using lasso

The coefficients of regression β may be estimated through minimizing

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

problem (2) leads to poor prediction performance, overfitting, and difficulty interpreting an appropriate model of least squares estimation when the number of covariates is greater than the sample size ($p > n$). Therefore, utilizing the Ordinary Least Squares (OLS) method with the aforementioned issues resulted in unstable and large variance estimations. These issues are the main reasons for using shrinkage and subset selection approaches. To ease model interpretation and improve prediction accuracy, it is usually necessary to conduct a variable selection, so that a parsimonious regression model is built by just using a few important covariates.

Regularization is a penalized regression method which is used for the regularization process for VS and the estimate of coefficients together in issues of regression. It is the least absolute shrinkage and selection operator (lasso). The lasso method was suggested by Tibshirani (1996) to solve the overfitting problem when there are many studies that have many more covariates than the sample size ($p > n$) through the addition of a penalty function (ℓ_1 norm) to the least squares loss function, which puts the coefficients of unimportant covariates equal to zero. Thus, variable selection can be achieved automatically. The lasso estimator is obtained as follows:

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

where $\lambda \geq 0$, λ is the regularization (tuning or shrinkage) parameter controlling the quantity of penalty, the highest value for λ gives a highest level of shrinkage (Alkenani and Yu, 2013). Small penalties result in the selection of big models having possibly high variance but less bias; big penalties result in the selecting models having less covariates with lower variance. The lasso regression has some notable properties:

1. It can solve the problem of multicollinearity.
2. It reduces the prediction error of the model by putting the coefficients of unimportant covariates equal to zero (Ranstam and Cook, 2018).
3. It deals with regression models that have a large number of covariates,i.e, ($p > n$).

The lasso penalty function has also been widely used in many statistical applications (Zheng, 2008; Nardi and Rinaldo, 2011; Bien et al., 2013; Wu et al., 2014; Kaul, 2014).

Although these are the attractive advantages of lasso and it has proven successful in a variety of situations, lasso has some problems that occur in the lasso method's work and it doesn't work well in certain of the following cases:

1. In practice, there are many studies that have $p > n$, but more variables than the sample size cannot be chosen by the lasso method (Zou and Hastie, 2005).
2. When there is a set of covariates that are highly correlated, the lasso will select just one of the group while ignoring the others (Zou and Hastie, 2005).
3. In the case of $(n > p)$ and when the covariates are highly correlated, as shown by experiments, the lasso regression does not choose covariates correctly (Zou and Hastie, 2005).

Although lasso performs well in putting the coefficients of unimportant covariates equal to zero, it has some drawbacks. Empirically, lasso chooses more covariates than required. The lasso approach has a bias for the estimate of large coefficients, indicating that the lasso approach is inconsistent since this approach penalizes all coefficients equally.

Consequently, lasso doesn't have oracle properties (the definition of oracle properties: according to Fan and Li (2001), the method with this property can select the true model with a probability of 1 quantity).

The lasso can be calculated using the fast algorithm which is available in the statistical program (R) called the least angle regression (LARS, Efron et al., 2004). It is the most efficient algorithm in terms of computation as well as being extremely fast. As a result, making penalized regression models highly popular in high dimensional data analysis.

2.2. Bayesian interpretation of lasso

Bayesian regularization methods were developed for difficulty obtaining statistical inference on the regression coefficients. On the other hand, a Bayesian method provides an exact inference even with a small sample size, as well as an exact estimate when p is bigger than n ($p > n$, Li et al., 2010; Alhamzawi and Ali, 2018b).

There are two steps in Bayes regularization models: First, determining the prior distribution of regression coefficients, which is the most important step in the Bayes method for variable selection and estimation of coefficients together, is the main idea in Bayesian analysis to minimize estimator variance while increasing bias. Therefore, the choice of the prior distribution must be exact because choosing an inaccurate or incorrect prior distribution without caution will lead to many problems, including Gibbs sample convergence issues and posterior estimation instabilities (Alhamzawi and Yu, 2012). Second, computing the posterior distribution (Agresti, 2010).

Park and Casella (2008) noted the penalty term's form (3). In 1996, Tibshirani proposed that when the coefficients of regression possess identical and independent Laplace (i.e., double-exponential) priors, estimations of lasso may be interpreted as posterior mode estimations, In consequence, many Blasso techniques have been suggested through the years by some other researchers subsequently employing Laplace-like priors (see for example, Figueiredo 2003; Bae and Mallick 2004; Yuan and Lin 2005). In 2008, Park and Casella considered a fully Bayesian analysis based on a conditional Laplace prior description of the form

$$\pi(\boldsymbol{\beta}/\sigma^2) = \prod_{k=1}^{p} \frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta_k|/\sqrt{\sigma^2}}, \quad (4)$$

for the regression coefficients vector $\boldsymbol{\beta}$ as well as the scale-invariant marginal prior for σ^2

$\pi(\sigma^2) = 1/\sigma^2$ on σ^2. The prior for $\boldsymbol{\beta}$ can be written as a scale mixture of normals (Andrews and Mallows, 1974; Park and Casella, 2008).

$$\frac{\lambda}{2\sqrt{\sigma^2}} e^{-\lambda|\beta|/\sqrt{\sigma^2}} = \int_0^\infty \frac{1}{\sqrt{2\pi s}} e^{-\beta^2/(2s)} \frac{\lambda^2}{2\sigma^2} e^{-\lambda^2 s/2\sigma^2} ds\,. \tag{5}$$

Under these assumptions, the Bayesian hierarchical modeling is given by (Andrews and Mallows, 1974; Park and Casella, 2008):

$$\begin{aligned} \mathbf{y}|\mu, \mathbf{X}, \boldsymbol{\beta}, \sigma^2 &\sim \mathrm{N}_n(\mu \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n), \\ \boldsymbol{\beta}|\sigma^2, s_1^2, \ldots, s_p^2 &\sim \mathrm{N}_p(\mathbf{0}_p, \sigma^2 \boldsymbol{w}_s), \\ \boldsymbol{w}_s &= diag(s_1^2, \ldots, s_p^2), \\ \sigma^2, s_1^2, \ldots, s_p^2 &\sim \pi(\sigma^2)\, w\sigma^2 \prod_{k=1}^{p} \frac{\lambda^2}{2} e^{-\lambda^2 s_k^2/2} w s_k^2\,, \end{aligned} \tag{6}$$

a scale mixture of normal (Andrews and Mallows, 1974; Park and Casella, 2008). Following this representation, Park and Casella (2008) proposed an efficient algorithm for Bayesian lasso.

2.3. Regularization regression using reciprocal lasso

To avoid overfitted models, in 2015, Song and Liang suggested the rlasso method for VS and the estimate of coefficients together, which is based on a novel class of penalty functions that are discontinuous at 0, decreasing in (0,∞) and also give near 0 coefficients of infinity penalties. Because of this, the characteristic property, rlasso is highly desirable for selecting models (Song and Liang, 2015; Song, 2018). The rlasso estimator is obtained as follows:

$$\hat{\beta}_{\mathrm{rlasso}} = (\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta}) + \lambda \sum_{k=1}^{p} \frac{1}{|\beta_k|} I\{\beta_k \neq 0\}\,, \tag{7}$$

where $\lambda > 0$ is the tuning parameter controlling the penalization degree as well as $I(.)$ is an indicator function, the lowest value for λ gives a highest level of shrinkage and gives coefficients that are near to zero. Compared to the lasso penalty, which is nondecreasing in $(0,\infty)$ and continuous, the penalty of rlasso decreases in $(0,\infty)$ and discontinuous at zero. Additionally, the lasso gives near 0 coefficients of 0 penalties, but the rlasso gives near 0 coefficients of infinity penalties. Also, the variable selection method of rlasso is very different from that of lasso. The lasso selects smaller coefficients, whereas the rlasso selects bigger ones. Theoretically, rlasso has the same oracle property (Mallick et al., 2021). The rlasso can perform much better at variable selection compared to the lasso. However, rlasso requires a lot of computational intensive (Song and Liang, 2015).

2.4. Bayesian interpretation of rlasso

Mallick et al. (2021) considered a fully Bayesian approach to the rlasso issuse, based the observation that whenever the parameters of regression are determined by independent inverse Laplace priors, the rlasso estimation for linear regression parameters may be interpreted as a Bayesian posterior mode estimation by using the following:

$$\pi(\boldsymbol{\beta}) = \prod_{k=1}^{p} \frac{\lambda}{2\beta_k^2} \exp\{-\frac{\lambda}{|\beta_k|}\} I\{\beta_k \neq 0\}, \qquad (8)$$

wherein $\lambda > 0$ denotes a scale parameter determining the prior's dispersion around 0. As a result, λ should be tiny in order to ease sparse recovery. This is rather counterintuitive considering that the majority of lasso-type shrinkage approaches penalize coefficients by a big value of λ. In specifically, the Bayesian rlasso requires a limited value of λ while the Bayesian lasso prefers a large value of λ to the best performance, which can be written as (Mallick et al., 2021)

$$\mathbf{y}^{n\times 1} | \, X, \boldsymbol{\beta}, \sigma^2 \sim N_n(X\boldsymbol{\beta}, \sigma^2 I_n),$$

$$\boldsymbol{\beta}^{p\times 1}|\mathbf{u} \sim \prod_{k=1}^{p} \frac{1}{\text{Uniform}(-u_k,u_k)}, \qquad (9)$$

$$\mathbf{u}^{p\times 1}|\lambda \sim \prod_{k=1}^{p} \text{Gamma}(2,\lambda),$$

$$\sigma^2 \sim \pi(\sigma^2),$$

following this representation, Mallick et al. (2021) proposed an efficient algorithm for Bayesian rlasso. Thus, they showed that the Bayesian approach performs better than its conventional counterpart in VS, prediction, and estimate. On the other hand, there are two reasons for using a Bayesian method to the rlasso issue. First, one of the greatest reasons for using a fully Bayesian method is the optimization issue's multimodal nature (7). Second, the Bayesian rlasso method is computationally efficient, resulting in scalable MCMC algorithms having perfect convergence as well as mixing characteristics.

2.5. The Adaptive lasso Regression

In 2006, Zou proved that the lasso estimator is inconsistent in VS, although the lasso performs well in putting the coefficients of unimportant covariates equal to zero, and the lasso approach is a common approach for VS as well as the estimate of coefficients at the same time. To address this problem, Zou (2006) proposed a new regularization method by assuming different regularization weights for different coefficients. It is called the adaptive least absolute shrinkage and selection operator (alasso) method. Lasso estimations are known to be biased to large coefficients since lasso penalizes all coefficients equally. The alasso method controls the lasso estimate's bias by adding adaptive weights, which are employed to penalize various coefficients in the lasso method. Thus, the unimportant covariates coefficients are reduced to 0 more efficiently. This method minimizes bias and improves variable selection accuracy, which creates estimates that are

consistent and unbiased, as well as it performs a better job of estimating important coefficients than lasso (Zou, 2006; Wang et al. 2007). The Alasso estimator is obtained as follows:

$$\hat{\beta}_{alasso} = (\mathbf{y} - \mathrm{X}\boldsymbol{\beta})'(\mathbf{y} - \mathrm{X}\boldsymbol{\beta}) + \lambda \sum_{k=1}^{p} \widehat{w}_k |\beta_k| , \quad (10)$$

where $\lambda \sum_{k=1}^{p} \widehat{w}_k |\beta_k|$ is known as the alasso penalty, $\widehat{\boldsymbol{w}}_k = (\widehat{w}_1, \widehat{w}_2, \ldots, \widehat{w}_p)$ denotes the adaptive weight vector, which is defined as follows:

$$\widehat{\boldsymbol{w}}_k = \frac{1}{|\widehat{\boldsymbol{\beta}}_k|^{\gamma}} , \quad (11)$$

for $k = 1, \ldots, p$ and $\gamma > 0$. The parameter γ is a tuned parameter (Zou, 2006) that may be calculated by utilizing the cross-validation method.

The alasso regression has the following advantages:

1. It is computationally more appealing (Zou, 2006).
2. It can be solved using the same efficient algorithm that was utilized to solve the lasso, i.e., the LARS algorithm (Efron et al., 2004).
3. It does have oracle properties (i.e., it selects the right subset of variables from a larger set on a consistent basis and includes asymptotic guarantees of unbiasedness and normality) by employing the ℓ_1 penalty, which is adaptively weighted (Zou, 2006).

The alasso estimator has been widely used in many statistical applications due to its good theoretical properties. (See, Zhang and Lu, 2007; Zeng et al., 2014; Yang and Wu, 2016). However, it needs consistent initial estimations of the regression coefficients, that are commonly unavailable in the high dimensional, small sample size setting. Additionally, none of the algorithms

employed to compute the alasso estimators gives a correct estimate of standard error.

2.6. The Bayesian adaptive lasso regression

A result of the alasso's suffering from collinearity induced by highly corrected covariates Because of the illness of the OLS initial estimation in $\widehat{\boldsymbol{w}}_k$'s. When the correlation between the covariates is high, the illness (ill-condition) happens, implying that $(X'X)^{-1}$ is not of full rank, resulting in coefficients estimations to be unstable.

Several authors used a Bayesian formulation in their studies, such as Sun et al. (2010) as well as Leng et al. (2014), who recently suggested Balasso for VS as well as the estimate of coefficients together in linear regression to avert the OLS initial estimations for the coefficients of regression because they estimation the adaptive weight $\widehat{\boldsymbol{w}}_k$'s automatically (Alhamzawi and Ali, 2018). Additionally, on the basis of a geometrically ergodic Markov Chain, the Bayesian approach to the alasso gives a valid standard error measure (Casella et al., 2010).

The Balasso is similar to the alasso (Zou, 2006), that develops the lasso (Tibshirani, 1996) through adding covariate-specific penalties (Sun et al., 2010). The Balasso may be obtained by using the following conditional Laplace prior with coefficient-specific tuning parameters (Feng et al., 2017):

$$\pi(\boldsymbol{\beta}/\sigma^2) = \prod_{k=1}^{p} \frac{\lambda_k}{2\sqrt{\sigma^2}} e^{-\lambda_k|\beta_k|/\sqrt{\sigma^2}} . \quad (12)$$

The Balasso, like the alasso, applies different penalties to different coefficients to improve its ability to produce better estimate and model selection outcomes.

The Balasso is a Bayesian hierarchical model. The following are the priors: (Sun et al., 2010):

$$p(B_0) \propto 1\,, \tag{13}$$

$$p(\sigma^2) \propto \frac{1}{\sigma^2}\,, \tag{14}$$

$$p(B_k/\lambda_k) = \frac{1}{2\lambda_k}\exp\left(-\frac{|B_k|}{\lambda_k}\right), \tag{15}$$

$$p(\lambda_k/\delta,\tau) = inv - Gamma(\lambda_k;\delta,\tau) = \frac{\tau^{\delta}}{\Gamma(\delta)}{\lambda_k}^{-1-\delta}\exp\left(-\frac{\tau}{\lambda_k}\right), \tag{16}$$

where $\delta > 0$ as well as $\tau > 0$ represent two hyperparameters. The Balasso was found to be very efficient, conceptually simple, simple to apply, and there is no need for any initial estimations of the regression coefficients that are useful. The Balasso, on the other hand, does not specify a point mass at zero. The regression coefficient samples would not be precisely zero. Thus, the Balasso method does not select variables, but if we observe the posterior distribution mode, it may be precisely 0.

3. Simulation Studies

We carry out simulation studies in this section to demonstrate the performance of lasso and rlasso approaches.

Where:

- Lasso: the least absolute shrinkage and selection operator.
- Rlasso: reciprocal lasso.

The data in the simulation examples were generated by

$$\boldsymbol{y}_i = x_i^{'}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, i = 1, 2, \ldots, n. \quad (1)$$

We setup the error distribution $\varepsilon_i \sim N(0, \sigma^2)$. The design matrix rows $\boldsymbol{X}$ were generated from $N(\mathbf{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ has an autoregressive correlation matrix, where $\boldsymbol{\Sigma_{ij}} = 0.50^{|\boldsymbol{i}-\boldsymbol{j}|}$ for all $1 \leq \boldsymbol{i} \leq \boldsymbol{j} \leq p$. We consider three cases for $\boldsymbol{\beta}$:

Simulation 1: $\boldsymbol{\beta}$ = (1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)
Simulation 2: $\boldsymbol{\beta}$ = (1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1)
Simulation 3: $\boldsymbol{\beta}$ = (1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)

Table 1: Mean squared errors (MSE) and standard deviations (SD) for Simulation 1

	Lasso	rLasso
MSE σ^2 =1	1.57	1.66
SD σ^2 = 1	0.30	0.32
MSE σ^2 = 4	5.64	5.95
SD σ^2 = 4	0.81	0.92
MSE σ^2 = 9	12.50	12.89
SD σ^2 = 9	2.18	2.06

Table 2: Mean squared errors (MSE) as well as standard deviations (SD) for Simulation 2

	Lasso	rLasso
MSE $\sigma^2 = 1$	1.26	1.45
SD $\sigma^2 = 1$	0.22	0.23
MSE $\sigma^2 = 4$	5.50	6.27
SD $\sigma^2 = 4$	1.22	1.21
MSE $\sigma^2 = 9$	11.78	13.43
SD $\sigma^2 = 9$	1.95	2.25

Table 3: Mean squared errors (MSE) and standard deviations (SD) for Simulation 3

	Lasso	rLasso
MSE $\sigma^2 = 1$	1.27	1.51
SD $\sigma^2 = 1$	0.26	0.30
MSE $\sigma^2 = 4$	5.10	6.16
SD $\sigma^2 = 4$	1.06	1.81
MSE $\sigma^2 = 9$	11.15	13.68
SD $\sigma^2 = 9$	1.83	2.30

4. Conclusion

Regularization methods have evolved over the years to deal with the difficulties in analysing high dimensional data. In this paper, we have described some regularization methods for variable selection and estimation of coefficients together in linear regression: lasso, Blasso, rlasso, Brlasso, alasso, and Balasso. Also, we compare the performance of two regularization approaches in this study: lasso and the rlasso. The results are listed in Tables 1, 2, and 3. The results of both approaches (lasso and rlasso) are very similar. Our outcomes demonstrate that lasso and rlasso perform comparably in choosing a high dimensional model in various simulation studies.

References

1. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of machine learning research, 3(Mar), 1157-1182.
2. Mallows CL (1973) Some comments on CP. Technometrics 15 : 661–675.
3. Akaike, H. (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), 716-723.
4. Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, 461-464.
5. Spiegelhalter DJ, Best NG, Carlin BP, Der Linde AV (2002) Bayesian measures of model complexity and fit. J R Stat Soc Series B Stat ethodol 64: 583-639.
6. George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. Journal of the American Statistical Association, 88(423), 881-889. Bayesian Analysis (2009)4, Number 1, pp. 85–118.
7. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.

8. Frank, L. E., & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. Technometrics, 35(2), 109-135.
9. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267–288.
10. Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association, 96(456), 1348-1360.
11. Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology), 67(2), 301-320.
12. Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., & Knight, K. (2005). Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(1), 91-108.
13. Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American statistical association, 101(476), 1418-1429.
14. Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1), 49-67.
15. Candes, E. and T. Tao (2007). The dantzig selector: Statistical estimation when p is much larger than n. The Annals of Statistics, 2313–2351.
16. Wang, H., & Leng, C. (2008). A note on adaptive group lasso. Computational statistics & data analysis, 52(12), 5277-5286.
17. Zou, H., & Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. Annals of statistics, 37(4), 1733.
18. Kang, J., & Guo, J. (2009). Self-adaptive Lasso and its Bayesian estimation. Working Paper.
19. Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. The Annals of statistics, 38(2), 894-942.

20. Cand`es, E. J. and T. Tao (2010). The power of convex relaxation: Near-optimal matrix completion. IEEE Transactions on Information Theory 56 (5), 2053–2080.
21. Mazumder, R., J. H. Friedman, and T. Hastie (2011). Sparsenet: Coordinate descent with nonconvex penalties. Journal of the American Statistical Association 106 (495), 1125–1138.
22. Simon, N., & Tibshirani, R. (2012). Standardization and the group lasso penalty. Statistica Sinica, 22(3), 983.
23. Song, Q. (2014). Variable Selection for Ultra High Dimensional Data (Doctoral dissertation).
24. Song, Q., & Liang, F. (2015). High-dimensional variable selection with reciprocal l 1-regularization. Journal of the American Statistical Association, 110(512), 1607-1620.
25. Song, Q. (2018). An overview of reciprocal L 1- regularization for high dimensional regression data. Wiley Interdisciplinary Reviews: Computational Statistics, 10(1), e1416.
26. Alhamzawi, R., Yu, K., Vinciotti, V., & Tucker, A. (2011). Prior elicitation for mixed quantile regression with an allometric model. *Environmetrics*, *22*(7), 911-920.
27. Alshaybawee, T., Midi, H., & Alhamzawi, R. (2017). Bayesian elastic net single index quantile regression. *Journal of Applied Statistics*, *44*(5), 853-871.
28. Alhamzawi, R., & Ali, H. T. M. (2018a). The Bayesian elastic net regression. *Communications in Statistics-Simulation and Computation*, *47*(4), 1168-1178.
29. Alhamzawi, R., & Ali, H. T. M. (2020). Brq: An r package for bayesian quantile regression. *METRON*, *78*(3), 313-328.
30. Alhamzawi, R., and Ali, H. T. M. (2018b). Bayesian tobit quantile regression with penalty. Communications in Statistics-Simulation and Computation, 47(6), 1739– 1750.

31. Li, Q., Xi, R., and Lin, N. (2010). Bayesian regularized quantile regression. Bayesian Analysis, 5(3), 533–556.
32. Park, T., & Casella, G. (2008). The bayesian lasso. Journal of the American Statistical Association, 103(482), 681-686.
33. Yi, N., & Xu, S. (2008). Bayesian LASSO for quantitative trait loci mapping. Genetics, 179(2), 1045-1055.
34. Huang, J., Ma, S., & Zhang, C. H. (2008). The iterated lasso for high-dimensional logistic regression. The University of Iowa, Department of Statistics and Actuarial Sciences, 1-20.
35. Brown, P. J., & Griffin, J. E. (2010). Inference with normal-gamma prior distributions in regression problems. Bayesian analysis, 5(1), 171-188.
36. Li, J., Das, K., Fu, G., Li, R., & Wu, R. (2011). The Bayesian lasso for genome-wide association studies. Bioinformatics, 27(4), 516-523.
37. Legarra, A., Robert-Granié, C., Croiseau, P., Guillaume, F., & Fritz, S. (2011). Improved Lasso for genomic selection. Genetics research, 93(1), 77-87.
38. Cai, X., Huang, A., & Xu, S. (2011). Fast empirical Bayesian LASSO for multiple quantitative trait locus mapping. BMC bioinformatics, 12(1), 1-13.
39. Hans, C. (2009). Bayesian lasso regression. Biometrika, 96(4), 835-845.
40. Sun, W., Ibrahim, J. G., & Zou, F. (2010). Genomewide multiple-loci mapping in experimental crosses by iterative adaptive penalized regression. Genetics, 185(1), 349-359.
41. Leng, C., Tran, M. N., & Nott, D. (2014). Bayesian adaptive lasso. *Annals of the Institute of Statistical Mathematics*, *66*(2), 221-244.
42. Alhamzawi, R., & Ali, H. T. M. (2018). The Bayesian adaptive lasso regression. Mathematical biosciences, 303, 75-82.
43. Mallick, H., Alhamzawi, R., Paul, E., & Svetnik, V. (2021). The reciprocal Bayesian lasso. Statistics in medicine, 40(22), 4830-4849.
44. Bornn, L., Gottardo, R., & Doucet, A. (2010). Grouping priors and the Bayesian elastic net. arXiv preprint arXiv:1001.4083.

45. Alhamzawi, R. (2014). Bayesian elastic net tobit quantile regression. Communications in Statistics-Simulation and Computation (just-accepted), 00–00.
46. Huang, A., Xu, S., & Cai, X. (2015). Empirical Bayesian elastic net for multiple quantitative trait locus mapping. Heredity, 114(1), 107-115.
47. Chen, X., Wang, Z. J., & McKeown, M. J. (2011). A Bayesian Lasso via reversible-jump MCMC. Signal Processing, 91(8), 1920-1932.
48. Mallick, H., & Yi, N. (2014). A new Bayesian lasso. Statistics and its interface, 7(4), 571-582.
49. Flaih, A. N., Alshaybawee, T., & Alhusseini, F. H. H. (2020). Sparsity via new Bayesian Lasso. Periodicals of Engineering and Natural Sciences, 8(1), 345-359.
50. Alkenani, A. and Yu, K. (2013). Sparse MAVE with oracle penalties. Advances and Applications in Statistics 34,pp. 85–105.
51. Ranstam, J., & Cook, J. A. (2018). LASSO regression. *Journal of British Surgery*, *105*(10), 1348-1348.
52. Zheng, S. (2008). Selection of components and degrees of smoothing via lasso in high dimensional nonparametric additive models. Computational statistics & data analysis, 53(1), 164-175.
53. Nardi, Y., & Rinaldo, A. (2011). Autoregressive process modeling via the lasso procedure. Journal of Multivariate Analysis, 102(3), 528-549.
54. Bien, J., Taylor, J., & Tibshirani, R. (2013). A lasso for hierarchical interactions. Annals of statistics, 41(3), 1111.
55. Wu, L., Yang, Y., & Liu, H. (2014). Nonnegative-lasso and application in index tracking. Computational Statistics & Data Analysis, 70, 116-126.
56. Kaul, A. (2014). Lasso with long memory regression errors. Journal of Statistical Planning and Inference, 153, 11-26.
57. Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. The Annals of statistics, 32(2), 407-499.

58. Alhamzawi, R., and Yu, K. (2012). Variable selection in quantile regression via Gibbs sampling. Journal of Applied Statistics, 39(4), 799–813.
59. Agresti, A. (2010). Analysis of ordinal categorical data (Vol. 656). John Wiley and Sons.
60. Figueiredo, M. A. T. (2003), "Adaptive Sparseness for Supervised Learning," IEEE Transactions on Pattern Analysis and Machine Intelligence, 25, 1150–1159.
61. Bae, K., and Mallick, B. K. (2004), "Gene Selection Using a Two-Level Hierarchical Bayesian Model," Bioinformatics, 20, 3423–3430.
62. Yuan, M., and Lin, Y. (2005), "Efficient Empirical Bayes Variable Selection and Esimation in Linear Models," Journal of the American Statistical Association, 100, 1215–1225.
63. Andrews, D. F., and Mallows, C. L. (1974), "Scale Mixtures of Normal Distributions," Journal of the Royal Statistical Society, Ser. B, 36, 99–102.
64. Zhang, H. H., & Lu, W. (2007). Adaptive Lasso for Cox's proportional hazards model. Biometrika, 94(3), 691-703.
65. 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. Journal of Applied Statistics, 41(4), 879-894.
66. 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.
67. Casella, G., Ghosh, M., Gill, J., & Kyung, M. (2010). Penalized regression, standard errors, and Bayesian lassos. Bayesian analysis, 5(2), 369-411.
68. Feng, X. N., Wu, H. T., & Song, X. Y. (2017). Bayesian adaptive lasso for ordinal regression with latent variables. Sociological Methods & Research, 46(4), 926-953.