Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/isl/100317

New Aspects on the Modified Group LASSO using the Least Angle Regression and Shrinkage Algorithm

Ahmed A. El Sheikh, Shaimaa L. Barakat and Salah M. Mohamed*

Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research (FSSR), Cairo University, Egypt

Received:21 Jun. 2021, Revised: 22 Jul. 2021, Accepted: 3 Aug. 2021.
Published online:1 Sep. 2021

Abstract: In this paper, we propose a new method which is a modified group lasso with least angle regression selection to improve the high dimensional linear model in explanatory data. In this approach, the data matrix becomes sparse; the column dimension increases and columns are highly correlated. We solve the problem of multicollinearity using LARS algorithm which reduces the bias and mean square error and improves the quality of the model. A high degree of multicollinearity prevents computer software packages from performing the matrix inversion required for computing the regression coefficients. Modified group lasso estimators are solved by the Least Angle Regression and Shrinkage algorithm which calculate the correlation vector, decrease the largest absolute correlation value and select best variable selection in linear regression. It is shown that the proposed method is better than Lasso, elastic net, ordinary least square, ridge regression and adaptive group lasso in various settings, particularly for large column dimension and big group sizes. Also modified group lasso with least angle regression selection is robust to parameter selection and has less variance inflation factor, less mean square error and largest determination coefficient.
Keywords: categorical variable - least angle regression selection - multicollinearity-modified group lasso - variables selection.

1 Introduction

Multicollinearity can cause serious problems in estimation and prediction when present in a set of predictors which has high dimension [1].

General regression models have been reviewed [5] with a focus on the LASSO and extensions, including the adaptive LASSO, elastic net, and group LASSO. The regularization terms which are responsible for inducing coefficient shrinkage and variable selection leading to improving performance metrics of these regression models are discussed. This makes these modern, computational regression models valuable tools for analyzing high-dimensional problems. They investigated that elastic net method is the best method.

Traditional statistical estimation procedures, such as Ordinary Least Squares (OLS) that tend to perform poorly, have high variance prediction, and may be difficult to interpret because of its large variance and covariance which means that the estimates of the parameters tend to be less precise and lead to wrong inferences, so we can use modified group lasso to solve serious problems which are caused by multicollinearity [8].

Stepwise regression procedure has been used [8] to build a regression model for describing and identifying the factors that influence the propensity to leave the service provided by cellular phone companies. The regression theory has been introduced [9] based on specific assumptions concerning the set of error random variables. They investigated that when errors are uncorrelated and have a constant variance, the ordinary least squares estimator produces the best estimates among all linear estimators.

Four variable selection methods in the context of multiple linear regression analysis have been compared [10] to select the best explanatory variables for long-term residential water demand forecasting model development. These methods were (i) stepwise selection, (ii) backward elimination, (iii) forward selection and principal component analysis (PCA). The results showed that different variable selection methods produced different multiple linear regression models with different sets of predictor variables. The selection methods (i)–(vi) showed some irrational relationships between the water demand and the predictor variables due to the presence of a high degree of correlations among the predictor variables, whereas PCA showed promising results in avoiding these irrational behaviors and minimizing multicollinearity problems.

An alternative algorithm for lasso estimator has been proposed **[**11**]** to overcome the issues in LASSO that can be combined with other exiting biased estimators called

*Corresponding author e-mail: Shaimaa.esraa@gmail.com

Almost Unbiased Ridge Estimator (AURE). They examined the performance of the proposed algorithm using a Monte-Carlo simulation study and real-world examples.

In social science studies, modified group lasso is used to solve the problem in categorical variables, such as race, gender, and nationality which are difficult to fit a linear model on such data, especially when some or all of the explanatory variables are categorical [12].

When the response variable is the only categorical variable, it is common to use modified group lasso to overcome the defects of ordinary least squares. However, when the covariates are also categorical, corresponding variables are coded using dummy variables into our design matrix [12].

In such situations, modified group lasso is often beneficial to use an alternative method and shrink the estimator towards zero vector, which has an effect on introducing some bias so as to decrease the prediction variance, with the net result of reducing the mean squared error of prediction. There is nothing better than modified group lasso (penalized estimators) which has objective functions with the addition of a penalty which is based on the parameter [12].

Group lasso in categorical data has been used [12]. They showed that group lasso has beneficial properties when dealing with categorical data. They proposed modified group lasso for improvement in categorical data. It performs better than lasso or group lasso in various settings particularly for large column dimension. They introduced a simulation study to compare the performance of lasso, group lasso and modified group lasso. They investigated that the modified group lasso is the best method.

The LARS algorithm selects the input variable that is more correlated with the response variable; it calculates the correlation vector and the largest absolute correlation value. It is computationally just as fast as forward selection [13]. It produces a full piecewise linear solution path, which is useful in cross-validation or similar attempts to tune the model. If two variables are almost equally correlated with the response, then their coefficients should increase at approximately the same rate. The algorithm thus selects the variable which is more stable. In many regression problems we are interested in finding important explanatory factors in predicting the response variable. The goal of ANOVA is often to select important main effects and interactions for accurate prediction, which amounts to the selection of groups of derived input variables. LARS algorithm is used to lessen the biased and mean square error. LARS estimators are consistent and sufficient [13].

We aim to combine modified group lasso with LARS algorithm to improve the quality of the model; reduce the biased and the mean square error; and handle multicollinearity in high dimension case. Thus, we obtain sufficient and consistent estimators. The present paper is organized as follows: the method which handles variables selection and multicollinearity in the next section; penalize regression methods is in section 3; the modified group lasso with LARS (least angle regression selection) method is proposed in section 4; theoretical properties are established in Section 5; Simulation results are reported in Section 6.

2 Material and Methods

2.1 Stepwise Method

This method can add variables (the forward selection) and can also drop variables in (backward) elimination [4]. The step wise is started with no input variables in the model (we may restart with a subset of variables and in this case, if there are more than or equal to three variables in the model, select one more significant variables x_i or delete (remove) one insignificant variable x_{ij}). Selection method compares criterion value of all models that include the first x_i and one additional x_{ij}. If the model with the additional x_{ij} gives the best criterion value, x_{ij} isn't removed in the model and no other variables are added.

2.2 Least Absolute Shrinkage and Selection Operator (LASSO)

Least absolute shrinkage and selection operator (LASSO) regression method is widely used in domains with massive datasets, such as genomics, where efficient and fast algorithms are essential. The LASSO is not robust to high correlations among predictors and will arbitrarily choose one and ignore the others and break down when all predictors are identical. The LASSO penalty expects many coefficients to be close to zero, and only a small subset to be large (and nonzero). The LASSO estimator uses L_1 norm to obtain a sparse solution to the following optimization problem see [12]. The lasso estimator is given by:

$$\frac{1}{2}\|y-x\beta\|_2^2+\lambda\sum_{j=1}^{p}\|\beta_j\| \qquad (1)$$

Where y is $(n\times 1)$ vector of response variable; x is $(n\times p)$ matrix of explanatory variables; p is the number of coefficient; (n) is the number of observation; and λ. is tuning parameter determination from the analysis data dependent on cross validation and Bayesian information computation.

2.3 Adaptive Group Lasso

Adaptive group lasso increases the flexibility of the model (consistency and efficiency). Its estimators are asymptotic. It is used when the degree of multicollinearity is medium and minimum. It reduces the number of covariates included in regression. It is used in variable selection and multicollinearity. It reduces the variance and the bias. Adaptive group lasso is used in large samples and when there are many explanatory variables. It can't be used in high dimensional [7]. Adaptive group lasso can be defined as:

$$\frac{1}{2}\|y-x\beta\|_2^2+n\sum_{j=1}^{p}\lambda_j\|\beta_j\| \qquad (2)$$

Where y , x , p , λ_j and n are defined in equation (1).

2.4 Modified Grouped Lasso

Modified grouped lasso is used in high dimension. It reduces the biased and the variance and is used in a large sample size. This method improves the quality of the model. It is used in multicollinearity and variable selection. It can be used in categorical data. Its estimators are not symbiotic see[12]. This estimator can be defined as:

$$\frac{1}{2}\|y-x\beta\|_2^2+\sqrt{n}\sum_{j=1}^{p}\lambda_j\|\beta_j\|_2 \qquad (3)$$

Where y , x , p , λ_j and n are defined in equation (1).

2.5 Elastic Net Estimator

The elastic net method overcomes the limitations of the LASSO method which uses a penalty function based on:-

$$\|\beta\|_1=\sum_{i=1}^{p}|\beta_i|$$

Using this penalty function has several limitations. For instance, in the "large *p*, small *n*" case, the LASSO selects at most effective variables before it saturates. Also if there is a group of highly correlated variables, the LASSO tends to select one variable from a group and ignore the others [15]. To overcome these limitations, the elastic net adds a quadratic part to the penalty $\|B\|^2$. When this step uses a ridge regression, its estimators are stable. This method is also used in high dimensional and categorical data. It increases the flexibility of the model and can be defined as:

$$\frac{1}{2}\|y-x\beta\|_2^2+\lambda\sum_{j=1}^{p}\|\beta_j\|+\left|\beta\sqrt{k}\right|^2 \qquad (4)$$

Where y , x are defined in equation (1); p , n and λ are defined in equation (1); $k=\frac{\sigma^2}{B'_{OLS}B_{OLS}}$, σ^2 is the variance covariance matrix of estimator ordinary least square.

2.6 Ridge Estimator

It is the best performing alternatives to the least square methods. Least square has no bias, but it has a bigger variance than the ridge regression estimator in the presence of multicollinearity. The Ridge regression estimator can improve the estimation of β_j by adding a small constant to the diagonal of the matrix, which will reduce significantly the variance influential factor in the matrix. Ridge regression is proven as an effective and efficient remedial method to deal with the general problems caused by multicollinearity see [2]. The ridge regression is defined as follows:

$$\hat{\beta}_{Ridge}=(x'x+kI)^{-1}x'y \qquad (5)$$

Where k is defined in equation (4); I is the identity matrix; y and x are defined in equation (1).

3 Penalize Regression Methods

The standard linear model (or the ordinary least squares method) performs poorly in some situations. A large multivariate data set contains a number of variables superior to the number of the samples size. A better alternative is the penalized regression allowing creating a linear regression model that is penalized for having too many variables in the model by adding a constraint in the equation. This is also known as shrinkage or regularization methods. The aim of imposing this penalty is to reduce (i.e. shrink) the coefficient values towards zero. This allows the less contributive variables to have a coefficient close to zero or equal zero. Note that the shrinkage requires selecting a tuning parameter (lambda) that determines the amount of shrinkage [7]. A tuning parameter, sometimes called a penalty parameter, controls the strength of the penalty term in methods in linear regression. It is basically the amount of shrinkage, where data values are shrunk towards a central point, like the mean. Shrinkage results in simple, sparse models which are easier to analyze than high-dimensional data models with large numbers of parameters. Tuning parameter takes the three forms:

3.1 Tuning parameter (λ_j):

Tuning parameter (λ_j) has previously been used as a model selection tool. As in model building, there are several candidate models to which λ_j tuning parameter is added, which will increase the quality of the model. By including more parameters in the model, the model becomes more complex and the estimates also tend to have greater variance. Due to this problem, tuning parameter λ_j dependent on determination on a kaiake information criterion (AIC) and cross validation for the selection of a better model which achieves a suitable trade-off between simplicity (fewer parameters) and goodness of fit (greater quality) see[7]. In the Gaussian case, tuning parameter takes the following forms:

$$\lambda_{CV} = \frac{\|y - x\beta\|}{\sigma^2} - n + 2df$$

$$\lambda_{AIC} = \log\left[\frac{1}{n}\|y - x\beta\|\right] + \frac{2df}{n}$$

$$df = \sum_{j=1}^{p} I\{\|\beta_j\| > 0\} + \sum_{j=1}^{p} \frac{\|\beta_j\|}{\beta_{OLS}},$$

Where

$\sigma^2 = \|y - x\beta\|^2 / (n - df)$ is the usual variance estimator associated with β_{ols}

3.2 Tuning Parameter d :

Tuning Parameter d is used in linear model to remedy multicollinearity so as to reduce the biased and mean square error. This tuning is used to overcome the problems of linear model in liu estimator, liu estimator two type, principle component two parameter and $(r-d)$ class estimator [14]. Tuning Parameter is fined as:

$$d_j = \frac{\sigma^2 - (\beta_{OLS})^2 k}{(\beta_{OLS})^2 + \frac{\sigma^2}{\theta_j}}$$

Where σ^2 is the variance of β_{ols} ; k is tuning parameter of ridge estimator ; θ_j is eigenvector.

3.3 Tuning Parameter k :

Tuning Parameter k is used in linear model to remedy multicollinearity so as to reduce the biased and mean square error. The tuning is used to overcome the problems of linear model in ridge estimator, principle component two parameter and $(r-k)$ class estimator see [2]. They suggested different values of tuning parameter. k is fined as:

$$k = \frac{\sigma^2}{\beta_{OLS}'\beta_{OLS}}$$

Where σ^2 is the variance of ordinary least square estimator and β_{OLS} is the estimator ordinary least square.

4 Combine between Modified Group Lasso and LARS:

Modified group lasso is used in high dimensional. It reduces the bias and variance and is used in large sample size. This method improves the quality of the model and it is used in multicollinearity and variable selection. It can be used in categorical data. Its estimators are not symbiotic see [12]. From equation (3), the modified group lasso is defined as:

$$\hat{\beta}_{\text{mod}} = \min\left(\frac{1}{2}\|y - x\beta\|_2^2 + \sqrt{n}\sum_{j=1}^{p}\lambda_j\|\beta_j\|_2\right) \qquad (6)$$

Where y is $(n \times 1)$ response vector, x is $(n \times p)$ matrix of full column rank, n is the number of observation , λ_j is tuning parameter determination from the analysis data dependent on generalized cross validation and Bayesian information computation and β_j is the vector of parameters.

Algorithm

The modified group lasso regression method overcomes multicollinearity and variable selection. Modified group lasso is one of the independent variable shrinkage regression techniques. LARS is efficient algorithm for estimating computational modified group lasso parameters. Calculation of modified group lasso parameters using LARS can use the following steps:

From equation (3), the modified group lasso is defined as:

$$\hat{\beta}_{\text{mod}} = \min\left(\frac{1}{2}\|y - x\beta\|_2^2 + \sqrt{n}\sum_{j=1}^{p}\lambda_j\|\beta_j\|_2\right)$$

x, y, n , λ_j. and β_j, are defined in equation (3).

Step (1)

Standardization the response variables (y) where $T = y - \bar{y}$

y is defined in equation (3), n is defined in equation (3). And the equation will be

$$\hat{\beta}_{\text{mod}} = \min(\frac{1}{2}\|T - x\beta\|_2^2 + \sqrt{n}\sum_{j=1}^{p}\lambda_j\|\beta_j\|_2) \qquad (7)$$

Where $T = y - \bar{y}$ is $(n \times 1)$ response variables in LARS algorithm, y is all values in $(n \times 1)$ vector, $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$, y_i is every value of response variable, n, λ_j. and β_j, x and y are defined as equation (3)

Step (2):

Standardization and normalization of the matrix of the explanatory variables x

Where $Z = \frac{x - \bar{x}_j}{\sigma_{x_j}\sqrt{n-1}}$, $\sigma_{x_j} = \sqrt{\frac{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}{n-1}}$, $\bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}$, n is the number of observation, x_{ij} every value of explanatory variables matrix. The equation will be:

$$\hat{\beta}_{\text{mod}.LARS} = \min(\frac{1}{2}\|T - Z\beta\|_2^2 + \sqrt{n}\sum_{j=1}^{p}\lambda_j\|\beta_j\|_2)$$

Where T is defined in equation (8), $Z = \frac{x - \bar{x}_j}{\sigma_{x_{ij}}\sqrt{n-1}}$ is $(n \times p)$ matrix of full column rank in LARS algorithm, $\sigma_{x_j} = \sqrt{\frac{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}{n-1}}$, $\bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}$, x_{ij} is every value in matrix $(n \times p)$. n, λ_j and β_j are defined in equation (3).

Step (3):

We select one or more significant variables and remove one or more of non-significant variables from the equation:

$$\hat{\beta}_{\text{mod}.LARS} = \min(\frac{1}{2}\|T - Z\beta\|_2^2 + \sqrt{n}\sum_{j=1}^{p}\lambda_j\|\beta_j\|_2) \qquad (9)$$

Where T and Z are defined in equation (9); n, λ_j and β_j are defined in equation (3).

Step (4):

According to [15], we compute the residual of equation (10) :

$$SSE = (T - Z\beta)^2 \qquad (10)$$

Where T, Z and β_j are defined in equation (8).

Step (5)

According to [6], we compute the mean square error of equation (10):

$$\frac{\|T - Z\beta\|_2^2}{n - p + 1} \qquad (11)$$

Where T, Z and β_j are defined in equation (9).

Step (6):

According to [3], from equation (10) and (11), we compute (3)

$$C_{\text{mod}.LARS} = SSE / MSE$$

Step (7):

According to [3], we compute the mathematical form of Mallow's criteria whose statistics is as follow:

$$M = (T - Z\beta_{OLS})^2 / \sigma_M^2$$

Where $\sigma_M^2 = \frac{\|T - Z\beta_{OLS}\|_2^2}{n - p + 1}$ (12)

Where T and Z are defined in equation (8); n, p, x and y are defined in equation (3) and $\beta_{OLS} = (x'x)^{-1}x'y$ is a vector of coefficient of ordinary least square.

Step (8):

If $C_{\text{mod}.LARS} > M$, we go to step (3) . But if $C_{\text{mod}.LARS} \le M$, we can obtain the best method.

$\therefore$ The modified group lasso with LARS algorithm is defined by minimizing:

$$\frac{1}{2}\|T - Z\beta\|_2^2 + \sqrt{n}\sum_{j=1}^{p}\lambda_j\|\beta_j\|_2 \qquad (13)$$

Where T is $(n \times 1)$ response vector; Z is $(n \times p)$ matrix of full column rank; n is the number of observation; λ_j and β_j are defined in equation (3).

5 Theoretical properties

Theorem (1):

Modified group lasso with least angle regression selection is consistent under some assumption:-

1)

$$(\hat{\beta}_{\text{mod}.LARS}+n^{-1/2}u)-(\hat{\beta}_{\text{mod}.LARS})=\frac{1}{2}\left\|T-Z(\beta+n^{-1/2}u\right\|_2^2+(\sqrt{n}\sum_{j=1}^{p}\lambda^2{}_j\left|\beta_j+n^{1/2}u\right|^2)^{1/2}$$
$$-\frac{1}{2}\|T-Z\beta\|_2^2-(\sqrt{n}\sum_{j=1}^{p}\lambda_j^2\beta_j^2)^{1/2}>1-\varepsilon_i$$

2) Under the condition $\sqrt{n}A_n \to 0$,
$A_n=\max\{\lambda_j, j\le p_0\}$;

3) $u=(x'x)$ is the full matrix; A_n is the largest value of eigenvalues;

4) $\beta_a=(\beta_1,\beta_2,.......,\beta_P)$ is the vector containing the relevant factors;

5) $\beta_b=(\beta_{p0},\beta_{p1},.......,\beta_{Pn})$ is the vector containing all the irrelevant factors thus β_a and β_b are associated with modified group lasso with LARS estimators. Suppose $\hat{\beta}_{\text{mod}.LARS}$ is the estimator of modified group lasso with LARS. Note that the estimator of modified group lasso with LARS is the local minimizer of the equation $\frac{1}{2}\|T-Z\beta\|_2^2+\lambda_j\sqrt{n}\|\beta_j\|_2$. If the true model is known, the oracle estimator (consistent) can be obtained, which is denoted by $\hat{\beta}_a$ the standard linear model implies that $\sqrt{n}(\hat{\beta}_a-\beta_a)\to N(0,\Sigma_a)$ where Σ_a is the variance covariance matrix.

Proof:

By using the first assumption, the result is :

$$(\hat{\beta}_{\text{mod}.LARS}+n^{-1/2}u)-(\hat{\beta}_{\text{mod}.LARS})=\frac{1}{2}\left\|T-Z(\beta+n^{-1/2}u\right\|_2^2+(n\sum_{j=1}^{p}\lambda_j^2\left|\beta_j+n^{1/2}u\right|^2)^{1/2}$$
$$-\frac{1}{2}\|T-Z\beta\|_2^2-(n\sum_{j=1}^{p}\lambda_j^2\beta_j^2)^{1/2}$$
$$=\frac{1}{2}u'(\frac{1}{n}Z'Z)u-u'(\frac{1}{\sqrt{n}}Z'(T-Z\beta)+(n\sum_{j=1}^{p}\lambda_j^2(\beta_j+n^{-1/2}u)^2)^{1/2}-(n\sum_{j=1}^{p}\lambda_j^2\beta_j{}^2)^{1/2}$$

By using the third, fourth and fifth assumption, the result is :

$$\frac{1}{2}u'(\frac{1}{n}Z'Z)u-u'(\frac{1}{\sqrt{n}}Z'(T-Z\beta)+(n\sum_{j=1}^{p}\lambda_j^2(\beta_j+n^{-1/2}u)^2)^{1/2}-(n\sum_{j=1}^{p}\lambda_j^2\beta_j^2)^{1/2}>$$
$$\frac{1}{2}u'(\frac{1}{n}Z'Z)u-u'(\frac{1}{n}Z'(T-Z\beta))+(n\sum_{j=1}^{p_0}\lambda_j^2\left|\beta_j+n^{-1/2}u\right|^2)^{1/2}-(n\sum_{j=1}^{p_0}\lambda_j^2\beta_j^2)^{1/2}\qquad(14)$$

By using the second assumption, the result is :

$$\frac{1}{2}u'(\frac{1}{n}Z'Z)u-u'(\frac{1}{n}Z'(T-Z\beta))+(n\sum_{j=1}^{p_0}\lambda_j^2\left|\beta_j+n^{-1/2}u\right|^2)^{1/2}-(n\sum_{j=1}^{p_0}\lambda_j^2\beta_j^2)^{1/2}=$$
$$\frac{1}{2}u'(\frac{1}{n}Z'Z)u-u'(\frac{1}{\sqrt{n}}Z'(T-Z\beta)-p_0\sqrt{n}A_n|u|\qquad(15)$$

From equations (14) and (15):

$$\therefore\frac{1}{2}u'(\frac{1}{n}Z'Z)u-u'(\frac{1}{\sqrt{n}}Z'(T-Z\beta)+(n\sum_{j=1}^{p}\lambda_j^2(\beta_j+n^{-1/2}u)^2)^{1/2}-(n\sum_{j=1}^{p}\lambda_j^2\beta_j{}^2)^{1/2}>$$
$$\frac{1}{2}u'(\frac{1}{n}Z'Z)u-u'(\frac{1}{\sqrt{n}}Z'(T-Z\beta)-p_0\sqrt{n}A_n|u|$$

Under the condition $\sqrt{n}A_n=0,\varepsilon_i=T-Z\beta;\ u'(\frac{1}{n}Z'Z)u=\sigma^2=1$ because it is normal standardized.

$$\therefore\frac{1}{2}u'(\frac{1}{n}Z'Z)u-u'(\frac{1}{\sqrt{n}}Z'(T-Z\beta)+(n\sum_{j=1}^{p}\lambda_j(\beta_j+n^{-1/2}u)^2)^{1/2}-(n\sum_{j=1}^{p}\lambda_j^2\beta_j^2)^{1/2}>1\text{-}\varepsilon$$

$\therefore$ The estimators of modified group lasso with LARS are consistent

Theorem (2):

Modified group lasso with least angle regression selection is selection consistent under some assumptions:

1) $\liminf_{n\to\infty} p(\hat{\beta}_{\text{mod}.LARS} = 0) = 1$

2) Under the condition $\sqrt{n}A_n \to 0$,
$A_n = \max\{\lambda_j, j \le p_0\}$

3) $\sqrt{n}b_n \to 0$, $b_n = \min\{\lambda_j, j \ge p_0\}$.

We know that if the probability is tending to one, all zero coefficients must be estimated exactly as zero. On the other hand, we know that the estimates for nonzero coefficients must be consistent. Such consistency implies that with probability tending to one, all the relevant variables, identified with nonzero coefficients, are consistent.

Proof:

By using the first assumption, the result is
$\liminf_{n\to\infty} p(\hat{\beta}_{\text{mod}.LARS} = 0) = 1$.

Without loss of generality, the assumption becomes $p(\hat{\beta}_p = 0) = 1$. Then the same argument can be used to show $p(\hat{\beta}_j = 0) = 1$ for any $p_o < j < p$ which implies immediately that $p(\hat{\beta}_b = 0) = 1$ for a better discussion. Z_{p_0} is defined to be the $n \times d_p$ design matrix where $p_o < d_p < p$. Z_{-p} is defined to be an $n \times (d - d_p)$ design matrix where $1 < j < p_o$. It is noted that if $(\hat{\beta}_p \ne 0)$, the penalty function $\|\beta_p\|$ becomes a differentiable function with respect to its component see [7] . Therefore, $\hat{\beta}_p$ must be

the solution of the following normal equation:

$$\frac{1}{\sqrt{n}} Z'_{p_0} \left\| T - \left[Z_{p_o}\beta_b + Z_{-p}\beta_a \right] \right\|_2^2 + \lambda_p \sqrt{n} \frac{\hat{\beta}_p}{\|\hat{\beta}_p\|} = 0$$

Where Z_{-p} is $n \times (d - d_p)$ design matrix; Z_{p_0} is the $n \times d_p$ design matrix; λ_p is eigenvalues with (p_ dimensional); $\beta_a = (\beta_1, \beta_2, \ldots\ldots, \beta_P)$ is the vector containing the relevant factors; $\beta_b = (\beta_{p0}, \beta_{p1}, \ldots\ldots, \beta_{Pn})$ is the vector containing all the irrelevant factors; n is sample size; and T is vector $(n \times 1)$ of response variables.

$$\therefore \frac{1}{\sqrt{n}} Z'_{p_0} \left\| T - Z_{p_0}\beta_b - Z_{-p}\beta_a \right\|_2^2 + \lambda_p \sqrt{n} \frac{\hat{\beta}_p}{\|\hat{\beta}_p\|} = 0$$

$$= \frac{1}{n} Z'_{-p} Z_{-p} \sqrt{n}(\beta_a - \hat{\beta}_a) + \left(\frac{1}{n} Z'_{p_0}(T - Z\beta)\right) + \frac{1}{n} Z_{p_0}{}' Z_{p_0} \sqrt{n}(\beta_b - \hat{\beta}_b) + \lambda_p \sqrt{n} \frac{\hat{\beta}_p}{\|\hat{\beta}_p\|} = 0$$

Where the second term $(\frac{1}{n} Z'_{p_0}(T - Z\,\beta))$ is the order zero when $n \to \infty$

By using the second and third assumptions:

The first and third terms are also of the same order because $\sqrt{n}(\beta_a - \hat{\beta}_a) = 0$ and $\sqrt{n}(\beta_b - \hat{\beta}_b) = 0$ according to theorem (1). The second note is that if $\hat{\beta}_p \ne 0$, there must exist a k as $\hat{\beta}_{pk} = \{\max|\hat{\beta}_{pk}| : 1 \le k \le d_p\}$ without the loss of generality. It is assumed that $k = 1$ then we must have $|\hat{\beta}_{p_1}| / \|\hat{\beta}_p\| \ge 1/d_p > 0$ in addition to $\lambda_p \sqrt{n} > \sqrt{n}b_n \to \infty$. Therefore, it is known that $\lambda_p \sqrt{n}\hat{\beta}_{pk} / \|\hat{\beta}_p\|$ dominates the first three terms in equation with probability tending to one. This simply means that this equation cannot be true as long as the sample size is sufficiently large. As a result it is concluded that with probability tending to one, the estimate $\hat{\beta}_p$ must be in apposition where $\|\hat{\beta}_p\|$ is not differentiable. Hence $\hat{\beta}_p$ has to be exactly zero. Thus, modified group lasso with LARS is selection consistency.

6 Simulation Study

In this section, a simulation is carried out to examine the performance of LASSO, adaptive group LASSO, Elastic net, modified group lasso with LARS (least angle regression selection), ordinary least square and ridge regression. The mean square error (MSE), variance inflation factor (VIF), Mallow's C*p* and determination coefficient (R square) are used to comparison. The data is used with upper fitting by generating 11 of variables of sample sizes n (n = 50, 100 and 150) using normal distribution respectively. For the model fitting, follow the convention. We choose the lambda which minimizes the estimation error to compare the performance of each method. It seems that modified group lasso with LARS (least angle regression selection) surpasses other methods. Upper fitting data is generated from library (SK learn). The multiple linear model is fitted by using mean square error, variance inflation vector and determination coefficient.

Table 1: comparison by using the measurements, n=50.

Methods measurements	VIF	MSE	Mallow's C*p*	R^2
Ridge	1.12	0.195	12.66	90.2
Lasso	1.21	0.272	28.57	89.6
Elastic net	1.20	0.270	19.78	89.9
OLS	3.60	0.630	32.98	90.0
Modified group lasso with LARS	1.10	0.190	09.63	92.8
Adaptive group lasso	1.12	0.270	19.23	89.8

In table (1), it is shown that ordinary least square has the largest mean square error and the largest value of variance inflation factor, but ordinary least square has determination coefficient bigger than determination coefficient of lasso, elastic net and adaptive group lasso. Ridge regression has mean square error less than the mean square error of lasso estimator, elastic net, ordinary least square and adaptive group lasso. Ridge regression has variance inflation factor less than the variance inflation factor of lasso, elastic net and ordinary least square. Ridge regression has determination coefficient bigger than determination coefficient of lasso, elastic net, adaptive group lasso and ordinary least square. That's why ridge regression is better than lasso, elastic net, adaptive group lasso and ordinary least square. On the other hand, it is noted that modified group lasso with LARS (least angle regression selection) has mean square error less than the mean square error of ridge regression. Modified group lasso with LARS (least angle regression selection) has variance inflation factor less than the variance inflation factor of ridge regression. Modified group lasso with LARS (least angle regression selection) has Mallow's C*p* less than the Mallow's C*p* of ridge regression. Modified group lasso with LARS (least angle regression selection) has determination coefficient bigger than the determination coefficient of ridge regression. Hence, modified group lasso with LARS (least angle regression selection) has the largest value of determination coefficient, the least value of mean square error, least value of Mallow's C*p* and least value of variance inflation factor. That's why the modified group lasso with LARS (least angle regression selection) is the best method.

Table 2: comparison by using measurements, n=100.

Methods measurements	VIF	MSE	Mallow's C*p*	R^2
Ridge	0.982	0.182	11.65	91.8
Lasso	1.030	0.250	26.31	90.4
Elastic net	0.997	0.232	19.31	91.2
OLS	2.800	0.410	31.82	91.5
Modified group lasso with LARS	0.740	0.170	08.93	94.8
Adaptive group lasso	0.988	0.232	19.01	90.8

In table (2), it is shown that ordinary least square has the largest mean square error and the largest value of variance inflation factor, but ordinary least square has determination coefficient bigger than determination coefficient of lasso, elastic net and adaptive group lasso. Ridge regression has mean square error less than the mean square error of lasso estimator, elastic net, ordinary least square and adaptive group lasso. Ridge regression has variance inflation factor less than the variance inflation factor of lasso, elastic net and ordinary least square. Ridge regression has determination coefficient bigger than determination and coefficient of lasso, elastic net, adaptive group lasso and ordinary least square. That's why ridge regression is better than lasso, elastic net, adaptive group lasso and ordinary least square. On the other hand, it is noted that modified group lasso with LARS (least angle regression selection) has mean square error less than the mean square error of ridge regression. Modified group lasso with LARS (least angle regression selection) has variance inflation factor less than the variance inflation factor of ridge regression. Modified group lasso with LARS (least angle regression selection) has Mallow's C*p* less than the Mallow's C*p* of ridge regression. Modified group lasso with LARS (least

angle regression selection) has determination coefficient bigger than the determination coefficient of ridge regression. Hence, modified group lasso with LARS (least angle regression selection) has the largest value of error, least value of Mallow's Cp and least value of variance inflation factor. That's why the modified group lasso with LARS (least angle regression selection) is the best method.

Table 3: comparison by using measurements, n=150.

Methods Measurements	VIF	MSE	Mallow's Cp	R^2
Ridge	0.642	0.112	11.02	93.70
Lasso	0.988	0.200	25.64	91.10
Elastic net	0.815	0.196	18.79	91.90
OLS	2.010	0.329	29.88	92.08
Modified group lasso with LARS	0.320	.0900	08.51	96.80
Adaptive group lasso	0.733	0.196	18.78	91.50

In table (3) , it is shown that ordinary least square has the largest mean square error and the largest value of variance inflation factor , but ordinary least square has determination coefficient bigger than determination coefficient of lasso, elastic net and adaptive group lasso. Ridge regression has mean square error less than the mean square error of lasso estimator, elastic net, ordinary least square and adaptive group lasso. Ridge regression has variance inflation factor less than the variance inflation factor of lasso, elastic net and ordinary least square. Ridge regression has determination coefficient bigger than determination coefficient of lasso, elastic net, adaptive group lasso and ordinary least square. That's why ridge regression is better than lasso, elastic net, adaptive group lasso and ordinary least square. On the other hand, it is noted that modified group lasso with LARS (least angle regression selection) has mean square error less than the mean square error of ridge regression. Modified group lasso with LARS (least angle regression selection) has variance inflation factor less than the variance inflation factor of ridge regression. Modified group lasso with LARS (least angle regression selection) has Mallow's Cp less than the Mallow's Cp of

ridge regression. Modified group lasso with LARS (least angle regression selection) has determination coefficient bigger than the determination coefficient of ridge regression. Hence, modified group lasso with LARS (least angle regression selection) has the largest value of determination coefficient, the least value of mean square error, least value of Mallow's Cp and least value of variance inflation factor. That's why the modified group lasso with LARS (least angle regression selection) is the best method.

7 Conclusions

In this paper, the modified group lasso method can be determined by LARS (least angle regression selection) algorithm. LARS (least angle regression selection) is a determination coefficient, the least value of mean square more sufficient algorithm which makes modified group lasso parameters have the least value of mean square error, the least variance inflation factor and the largest value R^2. LARS (least angle regression selection) for modified group lasso gives us coefficient vectors which have the best model with the smallest Mallow's Cp value. It is aimed to use LARS (least angle regression selection) to reduce mean square error and variance inflation factors, and consequently to improve the accuracy of the model.

Acknowledgements

We would like to thank the Editorial Board and the reviewers for their very constructive comments on the original version of the manuscript. Their comments have substantially improved the content of the paper.

References

[1] A. Dorugade, New Ridge Parameters for Ridge Regression, Journal of The Associated Statistics in Transition.,**15**, 23-36 , (2014).

[2] A.E.Hoerl, and R.W.Kennard, , Ridge Regression: Biased Estimation for Non- Orthogonal Problems", Journal of International Statistics., **55(6)**, 112-120 , (1970).

[3] A.Kazemi, H.Mohamed and H. Z. Shareef, Optimal Power Quality Monitor Placement Using Genetic Algorithm and Mallow ' s Cp", International Journal of Electrical Power and Energy Systems., **53**, 564–575, (2013).

[4] D.Anderson, D.Sweeney, , and T.Williams, Statistics for Business and Economic, Thomson Corporation, Canada, 739-745, (2008) .

[5] F.E.Streib and M. Dehmer, High –Dimensional Lasso-Based Computational Regression Models: Regularization Shrinkage and Selection, Machine Learning and Knowledge Extraction., **1**, 359-385, (2019).

[6] G.Monaco and A.Fedullo, Teaching Least Square in Matrix Notation, Ratio Mathematica., **32**,63-75, (2017).

[7] H.Wang and C.Leng, Anote on Adaptive Group Lasso,Computational Statistical and Data Analysis., **52**, 5277-5286, (2008).

[8] M.Al-Jarrah, Using Stepwise Regression to Investigate Customers' Propensity To Change Cellular Phone Providers, Global Journal Of Pure and Applied Mathematics., **13(9)** , 5013-5020, (2017).

[9] M.Giacalone, D.Panarello and R.Mattera, Multicollinearity in Regression :An Efficiency Comparsion Between L_P_ Norm and Least Square Estimators ,Springer Science and Business Media., **52**, 1831_1859 , (2018).

[10] M.Haque, A. Rahman, D.Hagare, and R.Chowdhury, A Comparative Assessment of Variable Selection Methods in Urban Water Demand Forecasting, Water., **10(4)**, 1-15 , (2018).

[11] M.Kayanan and P.Wijekoon, Variable Selection Via Biased Estimators in The Linear Regression Model ,Open Journal of Statistic., **10**, 113-126 , (2020).

[12] M.Lim, The Group-LASSO: Two Novel Application , Unpublished MSc Thesis, Stanford University., 11-55, (2013).

[13] M.Yuan and Y.Lin, Model Selection And Estimation in Regression With Grouped Variables, J. R. Statist. Soc. B., **68**,49–67 , (2006).

[14] R.A.Farghali, Generalized Liu Type Estimator For Linear Regression, IJRRAS., **38(1)**, 52-63, (2019).

[15] W.Liu, An Efficient Elastic Net With Regression Coefficients Method For Variable Selection of Spectrum Data, The National Natural Science Foundation Of China., **2(12)**, 1-13 , (2016).