The Scientific Journal of Business and Finance
https://caf.journals.ekb.eg

Combining Adaptive Group LASSO and LARS Under Certain Criteria

Shaimaa L. Barakat
Department of Applied Statistics and Econometrics, Higher institute of management , Mahala Kubra , Egypt.

Published online: **June 2024.**

To cite this article: Barakat, Shaimaa L. **Combining Adaptive Group LASSO and LARS Under Certain Criteria** .The Scientific Journal of Business and Finance, 44 (2), 86:104
DOI: 10.21608/caf. 2024.371228

***Corresponding author: Shaimaa.esraa@gmail.com**

Combining Adaptive Group LASSO and LARS Under Certain Criteria

Shaimaa L. Barakat
Department of Applied Statistics and Econometrics, Higher institute of management , Mahala Kubra , Egypt.

Article History
Received 14 May 2024, Accepted 20 June 2024, Available online June 2024

Abstract

In this paper, we propose a new method which is Adaptive Group Least Absolute Shrinkage and Selection Operation(Ada G LASSO) with the least angle regression selection to improve the high dimensional linear model in explanatory data. The addition of excessive variables to a model can lead to severe consequences. When a model contains numerous variables, it is likely that some of them will exhibit strong correlations. However, explanatory variables should ideally not possess strong relationships among themselves. This issue, known as multicollinearity, can significantly impact the interpretation of results by causing notable variations between models. It is well known that proper handling of outliers is essential in data analysis. This is even more likely in variable selection. Many variable selection methods are based on assessing minor differences in model quality or even in assessing statistics such as significance calculated from model parameters. Ada G LASSO is an attractive method that has the oracle property and is a convex penalty method. In addition the Ada G LASSO can be extended to some high-dimensional semi parametric models. Ada G LASSO performs better than Lasso, elastic net, ordinary least square and ridge regression in various settings, particularly for large column dimension and big group sizes. Also Ada G LASSO with least angle regression selection algorithm is robust to parameter selection and has less variance inflation factor, less mean square error and largest determination coefficient. .

Key wards: Ada G LASSO-least angle regression selection algorithm- mean square criteria-variance inflation factor criteria.

1. Introduction

Multicollinearity is widely recognized as one of the most critical issues in linear regression models, introducing various risks to the underlying assumptions

Multicollinearity, or near-linear dependence among the explanatory variables in a regression model, is a significant challenge in practical applications. Multicollinearity has severe implications for the ordinary least squares (OLS) estimates of the regression coefficients. If multicollinearity exists, in the method of least squares, the estimates of parameters will be generally poor, the variance of the regression coefficients may be inflated, the absolute values of the estimates will be too large and they will be unstabl(Shrestha,2020).

Redundancy:

Two or more variables can exhibit a certain degree of similarity. A sound model's loading plot will display these variables clustered closely together. When aiming for a streamlined model, it becomes feasible to eliminate redundant variables. The predictions will probably not improve, but the model will be based on fewer variables, each with a unique appearance, and it may therefore be easier to understand and interpret (Andersen and Bro, 2010).

Outliers:

It is well known that proper handling of outliers is essential in data analysis. This is even more so in variable selection. Many variable selection methods are based on assessing minor differences in model quality or even in assessing statistics such as significance calculated from model parameters. Therefore, the result of variable selection is even more sensitive to outliers than the actual model fit. For this reason, every result during the process of variable selection should therefore be complemented by careful outlier detection. This may be difficult in practice, but at the very least, the resulting model obtained after variable selection should be carefully assessed, and the variable selection may be re-run upon handling new outliers to verify the result (Andersen and Bro, 2010).

Over fitting:

If there are many more variables than samples, it is possible, by chance and over fitting, to find a certain number of variables that correlate to the property to be predicted. If such variables are chosen and the model is applied to new samples, Then, the predictions may be very poor, or there may be no relationship at all. Therefore, validation is fundamental (Andersen and Bro, 2010).

Zang and lu, (2007) proposed adaptive lasso and claimed that adaptive lasso is able to identify the true model consistency and the resulting estimator is as efficient. They investigated the use of Ada LASSO to remedy variable selection and multicollinearity, and showed that Ada LASSO is sufficient when multicollinearity is minimum and medium. . By using two theorem, they proved that adaptive lasso is consistent and obtains its tuning parameter by using generalized cross validation.

Wang and Leng, (2008) remedied G LASSO and introduced adaptive group lasso method. It is similar to Ada LASSO but has capability to select variables in a grouped manner. They investigated that adaptive group lasso estimator is able to identify the true model consistency and the resulting estimator is efficient.

Radchenko and James, (2011) proposed an approach, "Forward-Lasso Adaptive Shrinkage" (FLASH), that deals with situations where the number of explanatory variables is relatively larger than the number of observations. FLASH includes the LASSO and Forward Selection as special cases, and can be used in both the linear regression and the generalized linear model domains. As with LASSO and forward selection, FLASH iteratively adds one variable to the model in a hierarchical fashion but, unlike these methods, at each step adjusts the shrinkage so as to optimize the selection of the next variable.

Epprecht, et al (2017) compared two approaches of model selection methods for linear regression models and Adaptive LASSO. In a simulation experiment, considering a simple

setup with orthogonal candidate variables and independent data, they compared the performance of the methods concerning predictive power (out-of-sample forecast), selection of the correct model (variable selection) and parameter estimation. The case where the number of candidate variables exceeds the number of observation is taken in to consideration as well. They introduced simulation study to compare the performance of LASSO estimator and Adaptive estimator, they noted that the Adaptive LASSO estimator is the best method .

Streib and Dehmer, (2019) reviewed general regression models with a focus on the LASSO and extensions thereof, including the Adaptive LASSO, elastic net, and G LASSO. They discussed the regularization terms responsible for inducing coefficient shrinkage and variable selection leading to improved performance metrics of these regression models. This makes these modern, computational regression models valuable tools for analyzing high-dimensional problems .

El Sheikh, et al (2021) proposed a new method which is a modified GLASSO 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. The researches solved the problem of multicollinearity using LARS algorithm which reduces the bias and mean square error and improves the quality of the model. Modified GLASSO 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. They suggested that the proposed method is better than Lasso, elastic net, ordinary least square, ridge regression and Ada GLASSO in various settings, particularly for large column dimension and big group sizes. Also modified GLASSO with least angle regression selection is robust to parameter selection and has less variance inflation factor, less mean square error and largest determination coefficient.

Barakat,(2023) compared between LASSO, GLASSO, Ada LASSO, Ada GLASSO, modified GLASSO, ridge regression, elastic net, step wise, principle component and partial least square. She used real data contained 10 explanatory variables and 183 sample size. She showed that modified GLASSO, ridge and Ada GLASSO had the best methods.

2. Material and Methods

2.1 Least Absolute Shrinkage and Selection Operator (LASSO)

Least absolute shrinkage and selection operator (LASSO) regression methods are widely used in domains with massive datasets, such as genomics, where efficient and fast algorithms are essential. However, the LASSO is not robust to high correlations among predictors and will arbitrarily choose one and ignore the others, potentially leading to breakdowns when all predictors are identical. The LASSO penalty expects many coefficients to be close to zero, with only a small subset being larger (and nonzero). The LASSO estimator uses the penalized least squares criterion to obtain a sparse solution to the following optimization problem (El Sheikh, et al,2021),(Barakat,2023). The LASSO technique is inspired by ridge regression, a standard technique for shrinking coefficients. However, contrarily to the

latter, LASSO can set some coefficients to zero, resulting in an easily interpretable model (Epprecht, et al, 2017). 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_i 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,, λ. is tuning parameter determination from the analysis data dependent on cross validation and Bayesian information computation..

2.2 Adaptive Group LASSO

Adaptive GLASSO is an attractive method that has the oracle property, and it is a convex penalty method, the Ada G LASSO can be extended to some high-dimensional semi parametric models (Wang and Tian, 2017),(Barakat,2023). Ada GLASSO 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 defined as equation (1), p ,n defined as equation (1).

2.3 Elastic Net Estimator

Elastic Net simultaneously does automatic variable selection and continuous shrinkage, and it can select groups of correlated variables. Elastic Net shrinks the regression coefficients by combining L1-norm penalty (LASSO) and L2-norm penalty (ridge) together, The L1-norm part of the penalty generates a sparse model by shrinking some regression coefficients exactly to zero. The L2-norm part of the penalty removes the limitation on the number of selected variables (Liu and Li,2016), the elastic net method overcomes the limitations of the LASSO method which uses a penalty function based on:-

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

The use of this penalty function has several limitations. For instance, in the "large p, small n" case the LASSO selects at most P variables before it saturates. Also, if there is a group of highly correlated variables, then the LASSO tends to select one variable from a group and ignore the others. To overcome these limitations, the elastic net adds a quadratic part to the penalty $\|B\|^2$, which is used alone in ridge regression. whose estimators are consistency and efficient. It's used in large sample. Its estimators are stable. It is used in high dimensional, polynomial regression, multiple regression and categorical data . It increases the flexibility of the model (El Sheikh, et al, 2021),(Barakat,2023). It increases the flexibility of the model and can be defined as :

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

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

2.4 Ridge estimator

Ridge regression is motivation for improving OLS, I omitted is the fact that the estimates from such models have often a low bias and a large variance, this is related to the prediction accuracy of a model because it is known that either by shrinking the values of regression coefficients or by setting coefficients to zero (Barakat, 2023),(El Sheikh, et al ,2021). The ridge regression is defined as follows:

$$\hat{\beta}_{Ridge} = (x_{ij}{}' x_{ij} + kI)^{-1} x_{ij}{}' y_i \qquad (4)$$

Where k , the unknown constant used in estimating for different methods; I is the identity matrix y and x are defined as equation (1).

3. Combined between Adaptive Group LASSO and LARS:

LARS (least angle regression selection) algorithm in this manner acts as instinct would be expect, and furthermore is more stable. It is handily altered to create an effective algorithm for different techniques delivering comparable outcomes, similar to the tether and forward stage wise. It is powerful in settings where p > n (Epprechet, et al, 2017).

Ada GLASSO is an attractive method that has the oracle property, and it is a convex penalty method, the Ada GLASSO can be extended to some high-dimensional semi parametric models (Wang and Tian, 2017).

Algorithm :

$$Y = X\beta + \varepsilon$$

Step 1:

Compute vector of correlation coefficient between response vector (Y) and matrix of explanatory variables (X)

$$r = \frac{\sum_{i=1}^{n}(X_{ij} - \overline{X}_j)(Y_i - \overline{Y})}{\sqrt{(\sum_{i=1}^{n}(X_{ij} - \overline{X}_j)^2)(\sum_{i=1}^{n}(Y_i - \overline{Y})^2}}$$

Step 2:

Increase β_j in the direction of sign correlation (X, Y) until some other competitor X has as much correlation with current residual as does X .

Step 3:

Move β_j in the joint least squares direction for X until some other competitor X has as much correlation with the current residual.

Step 4:

Continue in this way until all predictors have been entered. Stop when correlation $(Z,T) = 0$ ∀ j, i.e. OLS solution.

Step 5:

Standardization the matrix of the explanatory variables x

Where $Z_r = \frac{X_r - \overline{X}_{rj}}{S_{xr}}$ an $S_x = \sqrt{\frac{\sum_{i=1}^{n}(X_{ij} - \overline{X}_j)^2}{n-1}}$ (5)

The equation will be :

$$Y = Z_r\beta + \varepsilon$$

Step 6:-

$$\varepsilon = (Y - Z_r\beta)$$

$$\therefore \varepsilon^2 = (Y - Z_r\beta)^2$$

Step 7:-

$$\sum_{i=1}^{n}\varepsilon^2 = \sum_{i=1}^{n}(Y - Z_r\beta)^2$$

$$\therefore \sum_{i=1}^{n}\varepsilon^2 = \|Y - Z_r\beta\|_2^2 \quad \Rightarrow \frac{\sum_{i=1}^{n}\varepsilon^2}{2} = (\frac{1}{2}\|Y - Z_r\beta\|_2^2) \quad (6)$$

Step 8:

Add the penalty $n\sum_{j=1}^{p}\lambda_j\|\beta_j\|$

$$\frac{\sum_{i=1}^{n}\varepsilon^2}{2} + n\sum_{=1}^{p}\lambda_j\|\beta_j\| = (\frac{1}{2}\|Y - Z_r\beta\|_2^2 + n\sum_{j=1}^{p}\lambda_j\|\beta_j\|)$$

Step 9:

$$\frac{\partial}{\partial\beta}\left(\frac{\sum_{i=1}^{n}\varepsilon^2}{2} + n\sum_{j=1}^{p}\lambda_j\|\beta_j\|\right)$$

The equation will be:

$$\therefore \hat{\beta}_{A.LARS} = \min(\frac{1}{2}\|Y - Z_r\beta\|_2^2 + n\sum_{j=1}^{p}\lambda_j\|\beta_j\|)$$

$$= \min Q(\beta)_{A.LARS}$$

Where $Q(\beta)_{A.LARS} = (\frac{1}{2}\|Y - Z_r\beta\|_2^2 + n\sum_{j=1}^{p}\lambda_j\|\beta_j\|)$

Theorem (1):

$\hat{\beta}_{A.LARS}$ is consistency if $\sqrt{n}\ a_n \to_p 0,$ then $\hat{\beta} - \beta = O_p(n^{-1/2})$

Where $a_n = \max\{\lambda_j, j \le p_0\}$, $b_n = \min\{\lambda_j, j > p_0\}$ and p_0 is the number of relevant variables.

Proof

let $\hat{\beta}_a$ and $\hat{\beta}_b$ be their related $\hat{\beta}_{A.LARS}$ estimators. On the off chance that one realizes the true model, the steady estimator can be acquired, which is signified by $\widetilde{\beta}_a$. Standard linear model hypothesis suggests that $\sqrt{n}(\widetilde{\beta}_a - \beta_a) \to_d N(0, \Sigma_a)$ where Σ_a is the dimensional

covariance matrix of the important elements ($d_0 = \sum_{j=1}^{p_0} d_j$). Note that the $\hat{\beta}_{A.LARS}$ objective density function $Q(\beta)_{ALARS}$ is a strictly convex function. Henceforth, as long as it can be showed that there is a neighborhood minimizer of $(\frac{1}{2}\|Y - Z_r\beta\|_2^2 + n\sum_{j=1}^{p}\lambda_j\|\beta_j\|)$, which is consistent, then, at that point, by the worldwide convexity of $(\frac{1}{2}\|Y - Z_r\beta\|_2^2 + n\sum_{j=1}^{p}\lambda_j\|\beta_j\|)$, one knows quickly that such a neighborhood minimizer should be $\hat{\beta}_{A.LARS}$. Consequently, the consistency of $\hat{\beta}_{A.LARS}$ is set up. Following a comparative thought in Fan and Li (2001), the presence of $\sqrt{n}$ consistent nearby minimizer is inferred by the way that for any $\varepsilon > 0$, there is a sufficiently large constant C, such that

$$\liminf_{n} p\left\{\inf_{u\in R^d:\|u\|=C} Q(\beta+n^{-1/2}u)_{A.LARS} > Q(\beta_{A.LARS})\right\} > 1-\varepsilon \tag{7}$$

It follows then

$$Q(\beta+n^{-1/2}u)_{A.LARS} - Q(\beta)_{A.LARS} = \frac{1}{2}\left\|Y - Z_r(\beta+n^{-1/2}u\right\|_2^2 + (n\sum_{j=1}^{p}\lambda_j\left|\beta_j + n^{1/2}u\right|)$$

$$-\frac{1}{2}\|Y - Z_r\beta\|_2^2 - n\sum_{j=1}^{p}\lambda_j\left|\beta_j\right|$$

Where Y is defined as equation (1) , Z_r is explanatory matrix with LARS (least angle regression selection) algorithm, λ_j, β_j and n are defined as equation (1) .

$$\because \frac{1}{2}\left\|Y - Z_r(\beta+n^{-1/2}u)\right\|_2^2 = \frac{1}{2}(Y - Z_r(\beta+n^{-1/2}u))^2$$

$$= \frac{1}{2}\left[Y'Y - 2Y'Z_r\beta - 2\frac{u'}{\sqrt{n}}Z_r'Y + \beta Z_r'Z_r\beta + 2\frac{u'}{\sqrt{n}}Z_r'Z_r\beta + \frac{u'}{\sqrt{n}}Z_r'Z_r\frac{u}{\sqrt{n}}\right] \tag{8}$$

$$\because -\frac{1}{2}\|Y - Z_r\beta\|_2^2 = -\frac{1}{2}(Y - Z_r\beta)^2 = \frac{-1}{2}\left[(-Z_r\beta)'(Y - Z_r\beta\right] = \frac{-1}{2}(Y'Y - 2Y'Z_r\beta + \beta Z_r'Z_r\beta) \tag{9}$$

From equations (8),(9)

$$= \frac{1}{2}u'(\frac{1}{n}Z_r'Z_r)u - u'(\frac{1}{\sqrt{n}}Z_r'(Y - Z_r\beta) \tag{10}$$

$$\therefore Q(\beta + n^{-1/2}u)_{A.LARS} - Q(\beta)_{A.LARS}$$

$$= \frac{1}{2}u'(\frac{1}{n}Z_r'Z_r)u - u'(\frac{1}{\sqrt{n}}Z_r'(Y - Z_r\beta) + (n\sum_{j=1}^{p}\lambda_j\left|\beta_j + n^{-1/2}u\right|) - (n\sum_{j=1}^{p}\lambda_j\left|\beta_j\right|) \tag{11}$$

$$\geq \frac{1}{2}u'(\frac{1}{n}Z_r'Z_r)u - u'(\frac{1}{\sqrt{n}}Z_r'(Y - Z_r\beta) + n\sum_{j=1}^{p_0}\lambda_j(\left|\beta_j + n^{-1/2}u\right| - \left|\beta_j\right|) \tag{12}$$

From equations (11),(12)

$$\therefore \frac{1}{2}u'(\frac{1}{n}Z_r'Z_r)u - u'(\frac{1}{\sqrt{n}}Z_r'(Y - Z_r\beta) + n\sum_{j=1}^{p_0}\lambda_j(\left|\beta_j + n^{-1/2}u\right| - \left|\beta_j\right|)$$

$$\geq \frac{1}{2}u'(\frac{1}{n}Z_r'Z_r)u - u'(\frac{1}{\sqrt{n}}Z_r'(Y - Z_r\beta) - p_0\sqrt{n}\,(a_n)\|u\| \quad (13)$$

where equality (13) holds because $\beta_j = 0$ for any $j > p_0$ according to the model assumption. According to the theorems condition, it is known that $\sqrt{n}\,a_n = 0_p(1)$, hence, the third term in equation (13) is $0_p(1)$. It is noted that the first term in equation (13) converges in probability to $u'\mathrm{cov}(Z_r)u$ which is a quadratic function in u, the second term $-u'(\frac{1}{\sqrt{n}}Z_r'(Y - Z_r\beta)$ is linear in u with a $O_p(1)$ coefficient. Therefore, when C is sufficiently large, the first term dominates the other two terms with an arbitrarily large probability. This implies the equation (11) and completes the proof.

Theorem (2):

Consider model $Y = Z_r\beta + \varepsilon$ we have $df = E(\widetilde{df})$ under the assumptions:

Y is $(n\times 1)$ response vector, Z_r is design matrix $(n \times p)$, β_j is the coefficient of ordinary least square, $\varepsilon \approx N(0,1)$ is error term , since $Z_r'Z_r = I$

Proof

Let $\hat{\beta}_j = (\hat{\beta}_{j1}, \hat{\beta}_{j2}, ..., \hat{\beta}_{jp})$ and $\hat{\beta}_{jLS} = (\hat{\beta}_{j1LS}, \hat{\beta}_{j2LS}., ..., \hat{\beta}_{jpLS})$ for any $\hat{\beta}_j$ that depends on Y only through $\hat{\beta}_{jLS}$, by the chain rule we have

$$\because Q(\beta)_{A.LARS} = (\frac{1}{2}\|Y - Z_r\beta\|_2^2 + n\sum_{j=1}^{p}\lambda_j\|\beta_j\|)$$

$$= (\frac{1}{2}(Y - Z_r\beta)^2 + n\sum_{j=1}^{p}\lambda_j\beta_j)$$

$$\therefore \frac{\partial Q(\beta)_{A.LARS}}{\partial\beta} = (\frac{1}{2}\times 2(Y - Z_r\beta)\ (-Z_r) + n\sum_{j=1}^{p}\lambda_j)$$

$$\hat{\beta} = (1 - \frac{(Z_r'Z_r)^{-1}\lambda_j n}{\|\hat{\beta}_{jLS}\|})\hat{\beta}_{jLS}$$

By using assumption $(Z_r'Z_r = 1)$

$$\therefore \hat{\beta} = (1 - \frac{\lambda_j n}{\|\hat{\beta}_{jLS}\|})\hat{\beta}_{jLS}$$

Where $\hat{\beta}_{jLS} = (Z_r'Z_r)^{-1}Z_r'Y$

$$\therefore \frac{\partial \hat{\beta}}{\partial \hat{\beta}_{jLS}} = 1 - \frac{n\lambda_j}{\left\|\hat{\beta}_{jLS}\right\|} + \left[\frac{n\sum_{j=1}^{p}\lambda_j \hat{\beta}_{jLS}}{\hat{\beta}_{jLS}^2}\right]$$

$$= 1 - \left[\frac{n\sum_{j=1}^{p}\lambda_j\left[\left\|\hat{\beta}_{jLS}\right\|^2 - (\beta_{jLS})^2\right]}{\hat{\beta}_{jLS}^3}\right] \qquad (14)$$

Where n and λ_j is defined as equation (1) and β_{jLS} is the coefficient of ordinary least square.

$$\therefore \sum_{j=1}^{p} \frac{\partial \hat{\beta}}{\partial \hat{\beta}_{jLS}} = \sum_{j=1}^{p} \left[1 - \left[\frac{n\sum_{j=1}^{p}\lambda_j\left[\left\|\hat{\beta}_{jLS}\right\|^2 - (\beta_{jLS})^2\right]}{\hat{\beta}_{jLS}^3}\right]\right]$$

$$\sum_{j=1}^{p}\left[\frac{\partial \hat{\beta}_j}{\partial \hat{\beta}_{jLS}}\right] = 1 - \left[\frac{n\sum_{j=1}^{p}\lambda_j(p-1)}{I(\hat{\beta}_{jLS}^3)}\right]$$

$$\therefore df\left[\frac{\partial \hat{\beta}_j}{\partial \hat{\beta}_{jLS}}\right] = 1 - \left[\frac{n\sum_{j=1}^{p}\lambda_j}{I(\hat{\beta}_{jLS}^3)}\right](p-1)$$

under the condition $I(\hat{\beta}_{jLS}) > \sum_{j=1}^{p}\lambda_j$

$$tr\left[\frac{\partial \hat{Y}}{\partial Y}\right] = tr\left[\frac{\partial Z_r \hat{\beta}}{\partial Y}\right]$$

$$\therefore tr\left[\frac{\partial \hat{Y}}{\partial Y}\right] = tr\left[\frac{\partial Z_r \hat{\beta}}{\partial \beta_{jLS}} \times \frac{\partial \beta_{jLS}}{\partial Y}\right]$$

Since $\therefore \hat{\beta}_{jLS} = (Z_r' Z_r)^{-1} Z_r' T$

$$\therefore tr\left[\frac{\partial \hat{Y}}{\partial Y}\right] = tr\left[\frac{Z_r \hat{\beta}}{\partial \beta_{jLS}} \times Z_r\right]$$

$$\therefore tr\left[\frac{\partial \hat{Y}}{\partial Y}\right] = tr\left[\frac{\hat{\beta}}{\partial \beta_{jLS}} \times Z_r{}' Z_r\right]$$

$$\therefore tr\left[\frac{\partial \hat{Y}}{\partial Y}\right] = tr\left[\frac{\hat{\beta}}{\partial \beta_{jLS}} \times I\right]$$

$$\therefore tr\left[\frac{\partial \hat{Y}}{\partial Y}\right] = \sum_{j=1}^{p} \left[\frac{\hat{\beta}}{\partial \beta_{jLS}}\right] \tag{15}$$

$$\therefore \widetilde{df}\left[\frac{\partial \hat{Y}}{\partial Y}\right] = 1 - \left[\frac{n\sum_{j=1}^{p}\lambda_j(p-1)}{I(\hat{\beta}^3_{jLS})}\right] \quad \text{under the condition } I(\hat{\beta}_{jLS}) > n\sum_{j=1}^{p}\lambda_j$$

Following a similar idea in (Yuan and Lin, 2006) an application of Stein's identity yield :

$$df = \sum_{i=1}^{n} Cov(\hat{Y}, Y)/\sigma^2$$

$$= E(\sum_{i=1}^{n} \left[\frac{\partial \hat{Y}}{\partial Y}\right])$$

$$= E[1 - \left[\frac{n\sum_{j=1}^{p}\lambda_j(p-1)}{I(\hat{\beta}^3_{jLS})}\right] \quad]$$

$$\therefore df = E\left[\widetilde{df}\right]$$

5. Simulation Study

In this section, a simulation is carried out to examine the performance of LASSO, Ada G LASSO, Elastic net, Ada G LASSO with LARS algorithm, ordinary least square and ridge regression. The mean square error (MSE), variance inflation factor (VIF) 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 200) 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 Ada GLASSO with LARS algorithm 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

Value of λ	Methods / Measurements	Ridge	Elastic Net	AGL LARS	Ada GLASSO	LASSO	OLS
VIF		No valid					14
MSE							7.97
Ad R2							92.7
Optimal λ	VIF	2.55	3.72	3.87	4.13	6.07	No valid
	MSE	.79	1.97	1.12	2.45	5.33	
	Ad R2	97.4	96	96.5	95.5	94	
$\lambda = 0.5$	VIF	2.61	3.79	3.94	4.24	6.15	
	MSE	1.07	3.12	1.99	4.02	6..9	
	Adj R2	96.5	95.8	95.77	95	93.5	
$\lambda = 0.8$	VIF	3.66	5.15	4.83	6.78	8.33	
	MSE	1.28	3.78	2.70	4.29	7.72	
	Adj R2	95.8	95.2	95.8	95	93.3	
$\lambda = 1.5$	VIF	3.96	5.17	4.94	7.25	8.74	
	MSE	1.46	3.83	2.81	4.59	7.77	
	Ad R2	95	94.5	95	94.3	92.5	

In table (1), it is shown that Elastic Net has less mean square error than LASSO estimator ; Elastic Net has less variance inflation factor than LASSO ; Elastic Net has bigger determination coefficient than LASSO , so Elastic Net is better than LASSO. Ada GLASSO has less mean square error than LASSO estimator ; Ada G LASSO has less variance inflation factor than LASSO ; Ada G LASSO has bigger determination coefficient than LASSO. That's why Ada G LASSO is better than LASSO. Ordinary least square has the largest mean square error and the largest value of variance inflation factor, but ordinary least square has bigger determination coefficient than determination coefficient of LASSO, Elastic Net and Ada G LASSO. Ada GLASSO has mean square error equal to the mean square error of Elastic Net. Ada GLASSO has less variance inflation factor than elastic net. Ridge regression has less mean square error than LASSO estimator, Elastic Net, ordinary least square and Ada G LASSO.

Ridge regression has less variance inflation factor than LASSO, Elastic Net and ordinary least square. Ridge regression has bigger determination coefficient than LASSO, Elastic Net, Ada and ordinary least square. That's why ridge regression is better than LASSO, Elastic Net, Ada GLASSO and ordinary least square. On the other hand, it is noted that Ada G LASSO with LARS estimation has less mean square error than ridge regression. Ada G LASSO with LARS estimation has less variance inflation factor than the ridge regression. Ada GLASSO with LARS estimation has bigger determination coefficient than the determination coefficient of ridge regression. Hence, Ada GLASSO with LARS estimation has the largest value of determination coefficient and the least value of mean square error and variance inflation factor. That's why the Ada GLASSO with LARS estimation is the best method.

Table 2.,Selection of variables and significant model , n=50

Value of λ	Methods / Measurements	Ridge	LASSO	AGL LARS	Elastic Net	Ada GLASSO	OLS
	Number of explanatory variables	NO Valid					All variables
	F test						11.45
	P value						0.05
$\lambda = 0.5$	Number of explanatory variables	43	11	47	35	33	NO Valid
	F test	18.230	16.23	17.18	15.39	12.33	
	P value	.005	.005	.005	.01	.025	
$\lambda = 0.8$	Number of explanatory variables	38	10	45	32	32	
	F test	17.40	15.44	16.29	15.33	12.62	
	P value	.005	.005	.005	.01	.025	
$\lambda = 1.5$	Number of explanatory variables	33	6	42	28	24	
	F test	17.244	15.41	18.644	12.64	12.41	
	P value	.005	.005	.005	.01	.025	

In table (2) it shown that ordinary least square select all variable and contains p value (0.05) so the model of ordinary least square is significant, on the other hand Ada G LASSO with LARS select large number of explanatory variable and contains least value for significant model so Ada GLASSO with LARS is the best model.

Table 3: comparison by using measurements, n=100

Value of λ	Method / Measurements	Ridge	Elastic Net	AG LARS	Ada GLASSO	LASSO	OLS
VIF		No Valid					11.16
MSE							6.29
Ad R2							93.85
λ Optimal	VIF	3.15	4.88	2.22	3.00	6.92	No Valid
	MSE	1.15	2.12	1.07	2.48	4.88	
	Ad R2	97.4	96.9	97.8	97,4	95.3	
$\lambda = 0.5$	VIF	3.93	5.39	2.82	3.57	07.38	
	MSE	1.76	2.85	1.60	2.77	5.37	
	Ad R2	96.5	96	96.7	96.7	94.31	
$\lambda = 0.8$	VIF	4.17	5.59	3.07	3.99	7.59	
	MSE	1.89	2.87	1.73	2.85	5.41	
	Ad R2	96.98	95.93	96.61	96	94	
$\lambda = 1.5$	VIF	4.23	5.69	3.13	4.15	7.74	
	MSE	1.90	3.58	1.86	3.24	5.74	
	R2	96	95.5	96	95.9	93.86	

In table (3), it is shown that Elastic Net has less mean square error than LASSO estimator ; Elastic Net has less variance inflation factor than LASSO ; Elastic Net has bigger determination coefficient than LASSO , so Elastic Net is better than LASSO. Ada GLASSO has less mean square error than LASSO estimator ; Ada G LASSO has less variance inflation factor than LASSO ; Ada G LASSO has bigger determination coefficient than LASSO. That's why Ada G LASSO is better than LASSO. Ordinary least square has the largest mean square error and the largest value of variance inflation factor, but ordinary least square has bigger determination coefficient than determination coefficient of LASSO, Elastic Net and Ada G LASSO. Ada GLASSO has mean square error equal to the mean square error of Elastic Net. Ada GLASSO has less variance inflation factor than elastic net. Ridge regression has less mean

square error than LASSO estimator, Elastic Net, ordinary least square and Ada G LASSO. Ridge regression has less variance inflation factor than LASSO, Elastic Net and ordinary least square. Ridge regression has bigger determination coefficient than LASSO, Elastic Net, Ada and ordinary least square. That's why ridge regression is better than LASSO, Elastic Net, Ada GLASSO and ordinary least square. On the other hand, it is noted that Ada G LASSO with LARS estimation has less mean square error than ridge regression. Ada G LASSO with LARS estimation has less variance inflation factor than the ridge regression. Ada GLASSO with LARS estimation has bigger determination coefficient than the determination coefficient of ridge regression. Hence, Ada GLASSO with LARS estimation has the largest value of determination coefficient and the least value of mean square error and variance inflation factor. That's why the Ada GLASSO with LARS estimation is the best method.

Table 4.Selection of variables and significant model, n=100

Value of λ	Methods / Measurements	Ridge	LASSO	AGL LARS	Elastic Net	Ada G LASSO	OLS
	Number of explanatory variables	NO Valid					All variables
	F test						11.45
	P value						.05
$\lambda = 0.5$	Number of explanatory variables	44	11	47	35	34	NO Valid
	F test	18.230	16.23	17.18	15.39	12.33	
	P value	.005	.005	.005	.01	.025	
$\lambda = 0.8$	Number of explanatory variables	38	11	46	33	32	
	F test	17.40	15.44	16.29	15.33	12.62	
	P value	.005	.005	.005	.01	.025	
$\lambda = 1.5$	Number of explanatory variables	34	6	42	28	28	
	F test	17.244	15.41	18.644	12.64	12.41	
	P value	.005	.005	.005	.01	.025	

In table (4) it shown that ordinary least square select all variable and contains p value (0.05) so the model of ordinary least square is significant, on the other hand Ada G LASSO with LARS select large number of explanatory variable and contains least value for significant model so Ada G LASSO with LARS is the best model.

Table 5.: comparison by using measurements, n=200.

Value of λ	Method / Measurements	ridge	Elastic Net	AG LARS	Ada G LASSO	LASSO	OLS
VIF		NO Valid					11.10
MSE							5.69
Ad R2							93.95
λ Optimal	VIF	3.07	4.83	2.18	2.96	6.86	NO Valid
	MSE	1.13	2.09	1.05	2.45	4.77	
	Ad R2	97.8	97	98	97,4	95.7	
$\lambda = 0.5$	VIF	3.93	5.39	2.82	3.57	07.38	
	MSE	1.69	2.82	1.58	2.75	5.24	
	Ad R2	97	96.3	97.5	97.2	94.55	
$\lambda = 0.8$	VIF	4.12	5.49	2.99	3.90	7.53	
	MSE	1.82	2.79	1.7	2.75	5.33	
	Ad R2	97	96	97	96.5	94.3	
$\lambda = 1.5$	VIF	4.23	5.63	3.09	4.03	7.68	
	MSE	1.92	3.17	1.79	2.93	5.57	
	R2	96	95.8	96.5	96.3	94	

In table (5), it is shown that Elastic Net has less mean square error than LASSO estimator ; Elastic Net has less variance inflation factor than LASSO ; Elastic Net has bigger determination coefficient than LASSO , so Elastic Net is better than LASSO. Ada GLASSO has less mean square error than LASSO estimator ; Ada G LASSO has less variance inflation factor than LASSO ; Ada G LASSO has bigger determination coefficient than LASSO. That's why Ada G LASSO is better than LASSO. Ordinary least square has the largest mean square error and the largest value of variance inflation factor, but ordinary least square has bigger determination coefficient than determination coefficient of LASSO, Elastic Net and Ada G LASSO. Ada GLASSO has mean square error equal to the mean square error of Elastic Net. Ada GLASSO has less variance inflation factor than elastic net. Ridge regression has less mean square error than LASSO estimator, Elastic Net, ordinary least square and Ada G LASSO. Ridge regression has less variance inflation factor than LASSO, Elastic Net and ordinary least square. Ridge regression has bigger determination coefficient than LASSO, Elastic Net, Ada and ordinary least square. That's why ridge regression is better than LASSO, Elastic Net, Ada

GLASSO and ordinary least square. On the other hand, it is noted that Ada G LASSO with LARS estimation has less mean square error than ridge regression. Ada G LASSO with LARS estimation has less variance inflation factor than the ridge regression. Ada GLASSO with LARS estimation has bigger determination coefficient than the determination coefficient of ridge regression. Hence, Ada GLASSO with LARS estimation has the largest value of determination coefficient and the least value of mean square error and variance inflation factor. That's why the Ada GLASSO with LARS estimation is the best method.

Table.6.Selection of variables and significant model, n=200

In table (6) it shown that ordinary least square select all variable and contains p value (0.05)

Value of λ	Methods / Measurements	Ridge	LASSO	AGL LARS	Elastic Net	Ada G LASSO	OLS
	Number of explanatory variables	NO Valid					All variable
	F test						11.45
	P value						0.5
$\lambda = 0.5$	Number of explanatory variables	44	14	47	33	28	NO Valid
	F test	18.230	16.23	17.18	15.39	12.33	
	P value	.005	.005	.005	.01	.025	
$\lambda = 0.8$	Number of explanatory variables	40	11	46	38	32	
	F test	17.40	15.44	16.29	15.33	12.62	
	P value	.005	.005	.005	.01	.025	
$\lambda = 1.5$	Number of explanatory variables	34	6	42	28	28	
	F test	17.244	15.41	18.644	12.64	12.41	
	P value	.005	.005	.005	.01	.025	

so the model of ordinary least square is significant, on the other hand Ada GLASSO with LARS select large number of explanatory variable and contains least value for significant model so Ada G LASSO with LARS is the best model.

Conclusion

In this paper, the Ada GLASSO method can be determined by LARS algorithm . LARS algorithm is a more sufficient algorithm to make Ada G LASSO parameters have the least value of mean square error, the least variance inflation factor and the largest values R Square. LARS algorithm for Ada G LASSO that calculates coefficient vectors which have the best model with the smallest variation. It is aimed to use LARS algorithm 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

Barakat,(2023)," Review of Classical Methods and Variables Selection in

Case of Multicollinearity: A Case Study with Real-Data",Journal of Commerce and Finance,vol(4),pp(226-246).

ElSheikh.A.A,Barakat.SH.LandMohamed.S.M,(2021),"New Aspect on the Modified Group LASSO Using the Least Angle Regression and Shrinkage Algorithm"InformationScienceLetter,vol(10),no(3),pp(527-536)

Epprecht, C., Guegan, D., Veiga, Á., and Joel, C., (2017), Variable Selection and Forecasting via Automated Methods for Linear Models: LASSO/adaLASSO and Autometrics, https://halshs.archives-ouvertes. fr/halshs-00917797v2.

Fan, J.and Li, R., (2001). " Variable Selection Via Non Concave Penalized Likelihood and Its Oracle Properties", Journal of the American Statistical Association vol(96), pp(1348–1360).

Liu, W and Li, Q.(2016).An efficient elastic net with regression coefficients method for variable selection of spectrum data. Journal.PONE., 12(2), 1-13.

Radchenko, P, and James, G, (2011), "Improved Variable Selection with Forward –LASSO Adaptive Shrinkage", The Annals of Applied Statistics, vol(5), no(1), pp(427–448).

Shrestha,N,(2020), "Detecting Multicollinearity in Regression Analysis",vol(8),no(2),pp(39-42).

Streib,F.E and Dehmer, M,(2019), "High –Dimensional LASSO-Based Computational Regression Models: Regularization Shrinkage and Selection", Machine Learning and Knowledge Extraction, vol(1), pp (359-385).

Wang, H and Leng, C, (2008),"Anode on Adaptive Group LASSO", Computational Statistical and Data Analysis, vol(52), pp(5277-5286..

.Wang.M and Tian.G.L,(2017),"Adaptive Group LASSO for High – Dimensional Generalized Linear Models",Springer,Stat Papers, vol(60),pp(1469-1486).

.Zhang,H.H.,Lu,W,(2007),"Adaptive LASSO For Cox's Proportional Hazard Model" ,Biometrika, vol(94),pp(691–703).