ISSN 2303-4521
Periodicals of Engineering and Natural Sciences *Original Research*
Vol. 11, No. 3, May 2023, pp.18-28

A new shrinkage method for higher dimensions regression model to remedy of multicollinearity problem

Zainab Fadhil Ghareeb[1], Suhad Ali Shaheed AL-Temimi[2]
[1] Department of Statistics, College of Administration and Economics, Mustansiriyah University, Baghdad, Iraq.
[2] Department of Statistics, College of Administration and Economics, Mustansiriyah University, Baghdad, Iraq.

ABSTRACT

This research seeks to present new method of shrinking variables to select some basic variables from large data sets. This new shrinkage estimator is a modification of (Ridge and Adaptive Lasso) shrinkage regression method in the presence of the mixing parameter that was calculated in the Elastic-Net. The Proposed estimator is called (Improved Mixed Shrinkage Estimator (IMSHE)) to handle the problem of multicollinearity. In practice, it is difficult to achieve the required accuracy and efficiency when dealing with a big data set, especially in the case of multicollinearity problem between the explanatory variables. By using Basic shrinkage methods (Lasso, Adaptive Lasso and Elastic Net) and comparing their results with the New shrinkage method (IMSH) was applied to a set of obesity -related data containing (52) variables for a sample of (112) observations. All shrinkage methods have also been compared for efficiency through Mean Square Error (MSE) criterion and Cross Validation Parameter (CVP). The results showed that the best shrinking parameter among the four methods (Lasso, Adaptive Lasso, Elastic Net and IMSH) was for the IMSH shrinkage method, as it corresponds to the lowest (MSE) based on the cross-validation parameter test (CVP). The new proposed method IMSH achieved the optimal shrinking parameter (λ = 0.6932827) according to the (CVP) test, that leads to have minimum value of mean square error (MSE) equal (0.2576002). The results showed when the value of the regularization parameter increases, the value of the shrinkage parameter decreases to become equal to zero, so the ideal number of variables after shrinkage is (p=6).

Corresponding Author:

Zainab Fadhil Ghareeb
Department of Statistics
Mustansiriayah University
Baghdad, Iraq
Email: badeaa99@uomustansiriyah.edu.iq

1. Introduction

The development of estimation methods began in 1795 when Gauss proposed the method of least squares published by Adrine - Marie Legendre in 1805. In 1922. R. A Fisher presented the method of Maximum Likelihood, which is characterized by several advantages, including consistency, Sufficiency, efficiency, and information. The estimators of the Least Squares (LS) method have, of course, a clear and accurate interpretation, but in the case of the presence of some of the explanatory variables (p) that are not related to the response variable and are correlated with each other, then failure to exclude them leads to additional complications. Also, the estimators of the parameters resulting from the (LS) method are unlikely to be equal to Zero, which leads to the emergence of all variables in the model, and therefore the methodology of Regularization for explanatory variables has been studied for the purpose of excluding correlated variables according to accurate Statistical methodologies [1-20]. In the same context, shrinkage is one of regularization methods that are taken for fitting a regression model using all (p) predictors, but under some constraint on the size of their estimated coefficients. The importance of shrinkage lies in getting rid of the multicollinearity problem by reducing the variance of estimators in the model [1-14].

The contribution of a number of researchers in developing the methods of regularization had a significant impact on the development of many modern techniques. One of the most important of these contributions was made by the researchers Kennard and Hoerl in 1970 when they proposed the Ridge Regression method, and in 1996 the researcher Robert Tibshirani introduced the Lasso regression method (Least Absolute shrinkage and selection operator) [20].

In 2020, a number of researchers (Joe and others) studied the possibility of replacing the standard Lasso regression with the adaptive Lasso regression, as a new estimator was proposed when the sample size is smaller than the number of variables in the sample, which provides greater flexibility. As for the two researchers (Zou and Hastie), they proposed the Elastic-net method in 2005, which was considered a combination of two methods of Ridge Regression (RR) and Lasso Regression. The methods described above are considered the most common and widely used because they lead to approximation of the values of coefficients of the variables in the model to zero or very close to Zero. [3,17,18].

The hypothesis of non-existence of the multicollinearity problem between the explanatory variables is one of the most important basic hypotheses in the multiple linear regression. The process of statistical analysis of real data can face this problem, which leads to misleading results in analysis and interpretation due to the high variance of least squares estimates. Our study is concerned with looking for a shrink estimator that gives the best shrinkage for the coefficients of explanatory variables. This study aims to propose a new shrinkage estimator in regression model that estimator is a modification of the (Ridge and Adaptive Lasso) regression model in the presence of the mixing parameter that was calculated in the Elastic-Net, and it was named (Improved Mixed Shrinkage Estimator (IMSHE)) model. The proposed model contributes to overcoming the problem of multicollinearity between explanatory variables and has proven its efficiency through the accuracy of variables selection and the speed of implementation of the classification process. Also, in order to reach the best performance and select the variables to build on efficient Shrinking model. An ideal selection of variables Model was achieved by measuring the Mean Square Error (MSE) criterion.

2. Method

2.1. General linear model

The General Linear Model (GLM) is one of the most widely used models in Various fields of statistical analysis, and the Ordinary Least Squares (OLS) method is one of the most common methods for estimating the parameters of the general linear model, as Follows: [7,12,18]

$$\underline{\boldsymbol{y}}_{\boldsymbol{n}\times\boldsymbol{1}} = \boldsymbol{X}_{\boldsymbol{n}\times(\boldsymbol{p}+\boldsymbol{1})}\,\boldsymbol{\beta}_{(\boldsymbol{p}+\boldsymbol{1})\times\boldsymbol{1}} + \boldsymbol{\epsilon}_{\boldsymbol{n}\times\boldsymbol{1}} \qquad \dots (1)$$

Where y_i is dependent variable (response variable) which is a nx1 Vector. X: a matrix of explanatory variables with dimensions (p×n).β is p×1 Vector of unknown parameters. ϵ A vector with dimensions (nx1) of random errors, as $E(\epsilon) = 0$ and $Var(\epsilon) = \sigma^2_{In}$, I_n which is the identity matrix.

The matrix X is called the design matrix and contains information about the levels of the predicted variables that are obtained from the observations, and the model parameters are represented by vector $\underline{\beta}$ and are usually estimated using the (OLS) method, the sum square regression (SSR) for regression model is:

$$SSR(\underline{\beta}) = \min_{\beta} = \{(y - X\beta)'(y - XB)\} \qquad \dots (2)$$

By solving equation (2) and taking derivative with respect to $\underline{\beta}$, we get:

$$\frac{\partial SSR(\underline{\beta})}{\partial \beta} = -2X'y + X'X\beta = 0$$

$$\hat{\beta} = (X'X)^{-1}X'y$$

And when there is no problem of multicollinearity between the explanatory variables, it can easily be proven that the ordinary least squares estimator is consistent, efficient, and unbiased with respect to parameter β, which has (BLUE) properties, Also, when the explanatory variables are almost linearly related, this leads to an increase in the differences and standard errors of the model estimates, and this means a decrease in (t-statistics), meaning that the inference results are not clear (misleading).

2.2. Models of shrinkage estimators

The misleading results and the consumption of time and effort represent one of the most important challenges facing researchers when analyzing large data sets with existing multicollinearity problems. The treatment of these problems is through the use of variables selection methodology, where large data can be reduced to reasonable and more effective data sets. Lasso, Adaptive Lasso and Elastic Net regression models have been studied where shrinking techniques can be used under certain constraints to enhance model accuracy and efficiency by reducing the number of variables in the model [8,14,21].

Shrinkage methods are considered one of the most effective and preferable ways to get rid of the multicollinearity problem by reducing the variance of estimators in the model. Robert Tibshirani proposed in 1996 a new method of estimation in linear models called the Lasso for short (Least absolute shrinkage and selection Operator). [15]. In 2020, Joe and others replaced the standard Lasso model with adaptive Lasso regression and proposed a new estimator when the sample size is smaller than the number of variables in the sample, providing greater flexibility in variable selection. As for the two researchers (Zou and Hastie), they proposed the Elastic-Net method in 2005, which was considered a combination of two methods of Ridge Regression (RR) and Lasso Regression. The methods described above are considered the most common and widely used because they lead to approximation of the values of coefficients of the variables in the model to zero or very close to Zero. These methods will be reviewed in addition to the new shrinkage method as follows:

2.3. Least absolute shrinkage and selection operator estimator (lasso)

The Lasso shrinkage model (Least Absolute shrinkage and selection operator) was proposed by Tibshirani in 1996. The Lasso regression technique is based on the principle of shrinking the sum of the sum of the squares errors (SSE) depending on the existence of a constraint that represents the sum of the absolute parameters that are smaller than a specific fixed value. Therefore, based on this constraint, Lasso regression works to shrink the regression parameters and set them equal to Zero, as well as variables greater than Zero are determined after reduction and adopted as part of the model, which contributes to minimizing the prediction error and thus preserving the good features of both stepwise selection methodology and ridge regression method (RR). This method is of great importance to deal with the problem of multicollinearity between explanatory variables [7,8,11].

Lasso uses the method of (Penalty Likelihood), where it uses a penalty term of type L_1 (representing the Norm Penalty term) on the regression parameters that tend to produce sparse models, and the results have shown that Lasso estimates are consistent in the presence of appropriate conditions. In the general linear model in equation (1), the variables X_1, X_2, ……, X_p are converted to the standard formula (standardization), and therefore E(X) =0 and Var (x)=1, and L_1 penalty term is added to the model parameters and thus the solution to equation (2) can be formulated as follows: [5 ,15]

$$\hat{\beta}_{Lasso} = \hat{\beta}(\lambda) = arg\min_{\beta} (y - X\beta)'(y - X\beta);\ s.t\ \sum_{j=1}^{p-1} |\beta_j| \leq t \ \ldots\ldots (3)$$

Where (t) represents a tuning parameter and controls the amount of shrinkage and its value is calculated according to the formula $t = \Sigma|\hat{b}_{OLS}|$ and the formulation of equation (3) can be modified according to the Lagrange formula to become as follows:

$$\hat{\beta}_{Lasso} = arg\min_{\beta} \left\{(y - X\beta)'(y - X\beta) + \lambda \sum_{j=1}^{q} |\beta_j|\right\};\ q = 1,2,\ldots,p-1 \quad \ldots (4)$$

Where λ represents the regularization parameter or penalty parameter and $\lambda \sum_{j=1}^{p} |\beta_j|$ represent normal L_1, and equation (4) is called the penalty function. the data, and the results become close to the results of OLS method estimators, choosing different values for λ results in different estimators of the parameters vector $\hat{\beta}_{Lasso} = (\hat{\beta}_1 . \hat{\beta}_2 \ldots . \hat{\beta}_{k-1})^T$ which is why it is so important to choose an appropriate value for the regularization parameter. This can be achieved by using the cross - validation technique to find the (appropriate) value of the penalty parameter (shrink). The appropriate value of parameter (λ) means that value that contributes to predicting the values of the response variable with the highest possible accuracy (Less Variance). [2,8,11,20]

The cross -validation technique is based on the principle of randomly taking equal folds (Subsets) of data, estimating the model and testing it by finding the Residual Sum Squares (RSS), and then applying the chosen shrinking model to the samples (subsets) for prediction that got the lowest (RSS) as in the formula below: [5,10]

$$RSS_{\lambda s}^{q} = \sum_{i=1}^{n}\left(y_i - \sum_{j=0}^{k-1} b_j(q.\lambda s)x_{ij}\right)^2 \quad \dots (5)$$

As q represents the list of folds chosen as a test group, and its number is usually between (5-10) because it is considered appropriate and gives acceptable results. Obtaining the average values of (RSS) for all folds can be done through the following formula:

$$CV(\lambda) = MSE_{\lambda s} = \frac{1}{Q}\sum_{q=1}^{Q} RSS_{\lambda s}^{q} \quad \dots (6)$$

Where Q represents the sum of number of folds for a data set of equal length (test set) which was chosen randomly and that λs represents a grid of $\lambda \in [0,1]$ values, after which the value of λ that gives the least $MSE_{\lambda s}$ is chosen:

$$\hat{\lambda}_{min} = CV(\lambda) \quad \dots (7)$$

2.4. Adaptive least absolute shrinkage and selection operator estimator (adaptive lasso)

Adaptive Lasso estimator was proposed by the researcher [Zou, H] in (2006) with the aim of obtaining the Lasso model with Oracle properties through the use of adaptive weights to Shrinkage the coefficients in the penalty function. Oracle methodology has the following characteristics: [1,6,13,19,22]

1- Consistency in choosing the variable $lim_n P(\hat{\beta}^n = \beta^n) = 1$
2- The estimator has the asymptotic properties of the normal distribution.

The basic idea of Adaptive Lasso is to provide penalty weights for all regression coefficients, and this can be summarized in two steps. In the first step, the vector of adaptive weights that depend on the data $\widehat{\omega}$ (which are always positive values) is estimated according to the following formula:

$$\widehat{\omega} = \frac{1}{|\hat{\beta}_j|^{\gamma}} \quad \dots (8)$$

Where γ is a positive constant value $\gamma > 0$, which represents the power of the adaptive weight and is related to the high-dimensional model, and $\hat{\beta}_j$ is a constant (estimated) and initial value obtained from the method of least squares (LS) or the Ridge estimators (RR) in case the problem of multicollinearity is high. [16,18]

The second step, for the weight vector $\underline{\omega} = (\omega_1, \dots, \omega_p)^T$, Lasso estimator is reformulated according to the criterion of the Weighted Lasso methodology to minimize the objective function as follows: [22]

$$b_{ALasso} = \underset{\beta}{\operatorname{argmin}}\{(y - X\beta)'(y - X\beta)\} \ ; \ s.t \sum_{j=1}^{q} \widehat{\omega}_j |\beta_j| \le t \quad \dots (9)$$

Equation (9) can be reformulated according to Lagrange's formula, to become:

$$\hat{\beta}_{ALasso} = arg \min_{\beta}\left\{\sum_{i=1}^{n}\left(y_i - \sum_j \beta_j x_{ij}\right)^2 + \lambda \sum_{j=1}^{p} \widehat{\omega}_j |\beta_j|\right\} \quad \dots (10)$$

Both parameters (γ) and (λ) are used within the two-dimensional Cross -Validation (CV) methodology to adjust the adaptive Lasso estimator. The penalty function in equation (10) is $\lambda \sum_{j=1}^{p} \widehat{\omega}_j |\beta_j|$. To choose the value of $\hat{\beta}_j$ when p is small, i.e. ($p \ll n$), then $\hat{\beta}_j = \hat{\beta}_{OLS}$, and to choose $\hat{\beta}_j = \hat{\beta}_{RR}$ in the case of large p, i.e. ($p \gg n$). and that $\widehat{\omega}_j; j = 1,2,\dots,p$.

2.5. Elastic net estimator

In 2005, the two researchers (Zou and Hastie) proposed the Elastic Net model for the purpose of improving the Lasso method by addressing some of its constraints, especially with regard to the technique of selecting variables. The main feature of the elastic net method represented in stimulating a grouping effect, where strongly correlated predictors tend to be in or out of the model together. on the contrary, the lasso method is so tending to separate such groups and keep only the variable with the strongest correlation. Also, elastic net method is especially important and useful when the number of predictors (p) is much bigger than the number of observations (n). Conversely, the lasso is not very effective variable selection method in the case of p>n[1][3]. The mechanism of action of the Elastic Net model is based on taking the standard formula for the explanatory variables as well as for the response variable. A penalty term has been developed that combine the constraints of the (Lasso-L_1) model and the constraints of the (Ridge- L_2) model and it is represented by the two parameters $(\lambda_1.\lambda_2)$, which are fixed and non-negative values. Therefore, the solution to equation (2) based on the Elastic Net method can be formulated as follows [21]:

$$\hat{\beta}_{E.Net} = arg\min_{\beta}|y - X\beta|^2;\ s.t\ (1-\alpha)|\beta|_1 + \alpha|\beta|^2 \leq t\ for\ somet \quad \dots (11)$$

Where:

$$|\beta|^2 = \sum_{j=1}^{P} \beta_j^2$$

$$|\beta|_1 = \sum_{j=1}^{P} |\beta_j|$$

$$\alpha = \lambda_2/(\lambda_1 + \lambda_2)$$

Where $\alpha \in [0,1]$ and called (Mixing Parameter). Equation (11) can be modified according to Lagrange's formula to become as follows:

$$\hat{\beta}_{E.Net} = \text{argmin}_{\beta}\left\{(y - X\beta)'(y - X\beta) + (\lambda_1 + \lambda_2)\sum_{j=1}^{p}\left[(1-\alpha)\beta_j^2 + \alpha|\beta_j|\right]\right\} \quad ; j = 1,2,\dots,p \quad \dots (12)$$

Where $0 < \lambda_1 + \lambda_2 < 1$ and $(\lambda_1 + \lambda_2)\sum_{j=1}^{p}\left[(1-\alpha)\beta_j^2 + \alpha|\beta_j|\right]$ is penalty function

Lasso and Ridge models are both particular cases of Elastic Net model when $\alpha = 0\ and\ \alpha = 1$. The Elastic Net model allows us to select more variables and improve prediction and co-selection of groups of highly correlated variables. The relation between the three models (Elastic Net, Lasso, and Ridge) can be formulated as follows: [1,3,9,11]

$$\hat{\beta}(\alpha) = \begin{cases} \hat{\beta}_{RR} & \alpha = 0 \\ \hat{\beta}_{Enet} & 0 < \alpha < 1 \\ \hat{\beta}_{Lasso} & \alpha = 1 \end{cases} \quad \dots (13)$$

As for the selection of the two regularization parameters $(\lambda_1 + \lambda_2)$ in the elastic net estimator in the penalty term function equation, and as indicated in the Lasso regression estimator using the cross-validation technique in equation (6), but since the elastic net model has two regularization parameters then we need to implement two-dimensional cross-validation. Also, a relatively small range of values is usually chosen for λ_2, for example, (0,0.01, 0.1,1,10,100) and as for the second regularization Parameter is, its value is ten times the value of the cross-validation. [4,6,15]

2.6. The proposed improved mixed shrinkage estimator (IMSHE)

A Double Penalty Term is a combination between the Lasso and adaptive Lasso methods, with the presence of the Mixing Parameter of the elastic net method. The developed penalty term was formulated as a function of weight as follows:

$$\underline{\varrho}_j = \frac{|\hat{\beta}_j|^{1-\alpha}}{\left(|\hat{\beta}_j|^{\gamma}\right)^{\alpha}} \quad \dots (14)$$

Where $\hat{\beta}_j$ is the vector of the parameters measured by (OLS) method or by (Ridge) method when there is a high multicollinearity problem between the explanatory Variables. γ is a tunning parameter with a positive (Non-Negative) value and $\gamma > 0$ in experimental studies and α is A mixing parameter that is determined from Elastic Net model and Whose value is:

$$\alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2}$$

Where $0 < \alpha < 1$ and $0 < \lambda_1 + \lambda_2 < 1$

The value of α is controls the amount of regularization parameters (λ_1, λ_2)that give the lowest MSE and thus all the good properties of the estimator (Adaptive Lasso) and (Elastic Net) are transferred to the proposed enhanced estimator. The general form of the Enhanced shrinkage model is as follows:

$$\hat{\beta}_{IMSH} = \underset{\beta}{\operatorname{argmin}}(y - X\beta)^{`}(y - X\beta) \quad ; \qquad s.t \sum_{j=1}^{p-1} \frac{|\beta_j|^{1-\alpha}}{\left(|\beta_j|^{\gamma}\right)^{\alpha}} \leq t \qquad \ldots (15)$$

Where t represents a constant parameter that is calculated through $t = \sum\left|\hat{b}_{OLS}\right|$. Equation (14) can be reformulated according to Lagrange's formula as follows:

$$\hat{\beta}_{IMSH} = \underset{\beta}{\operatorname{argmin}}\left\{(y - X\beta)'(y - X\beta) + (\lambda_1 + \lambda_2)\sum_{j=1}^{p} \underline{\varrho}_j\left|\beta_j\right|\right\} \qquad \ldots (16)$$

The Ridge and Adaptive Lasso shrinkage estimators are a special case of Improved Mixed Shrinkage Estimator (IMSHE) when α =0 and α =1, respectively. Also, the penalty term of the adaptive Lasso model has more weight than the penalty term of the Ridge model when α approaches one and vice versa when α approaches zero. Improved Mixed Shrinkage Estimator will be transferred to the oracle estimator properties.

3. Results and discussion

The study data were collected from an obesity treatment center in Baghdad. This data on obese individuals were collected based on the indicators of the MSLCA07-Body Building Weight Test System / Human Body Fat Health Analyzer. The sample size was (n=112), that is, the number of obese people, and it was randomly drawn. The number of indicators was 52 (explanatory variables), and the person's weight was taken as a response variable, as show in Table 1.

Table 1. Represent a description of the study variables

X's	description	X's	description	X's	description
x1	Gender	**x19**	Part fat leg left	**x37**	20KHz RA
x2	Age	**x20**	Part fat leg right	**x38**	02KHz LA
x3	Height	**x21**	part muscle body	**x39**	20KHz TR
x4	total body water	**x22**	Part fat body	**x40**	20KHz RL
x5	protein	**x23**	Visceral Fat	**x41**	20KHz LL
x6	Abio-salt	**x24**	Protein	**x42**	50KHz RA
x7	Fat	**x25**	Abio- Salt	**x43**	KHz LA50
x8	Muscle	**x26**	Fat	**x44**	50KHz TR
x9	Body Mass Indicator (BMI)	**x27**	weight	**x45**	50KHz RL
x10	Percentage of Body Fat Index (PBF)	**x28**	Muscle	**x46**	50KHz LL
x11	Waist-to-hip ration (WHR	**x29**	Fat 1	**x47**	100KHz RA
x12	Moisture Ratio	**x30**	BMI 1	**x48**	100KHz LA
x13	Part muscle Arm left	**x31**	PBF 1	**x49**	100KHz TR
x14	Part muscle Arm right	**x32**	Ideal Weight	**x50**	100KHz RL
x15	Part fat Arm left	**x33**	Weight Control	**x51**	KHz LL100
x16	Part fat Arm right	**x34**	Basic metabolism	**x52**	Obesity diagnosis
x17	part muscle leg left	**x35**	Health assessment		
x18	part muscle leg right	**x36**	Physical age		

For the purpose of obtaining a matrix of correlations between the explanatory variables of the study sample, the R language program was used to analyze the statistical data, as follows:

The above correlation matrix gives a clear indication of the existence of a correlation between the explanatory variables, as indicated by blue and dark blue (positive correlation) and red and dark red (negative correlation).

For the purpose of detecting the problem of the multicollinearity between the variables of the study, the variance inflation factor test was used, according to the following equation [4,5]:

Variance Inflation factor (VIF) = $\frac{1}{1-R_j^2}$ … (17)

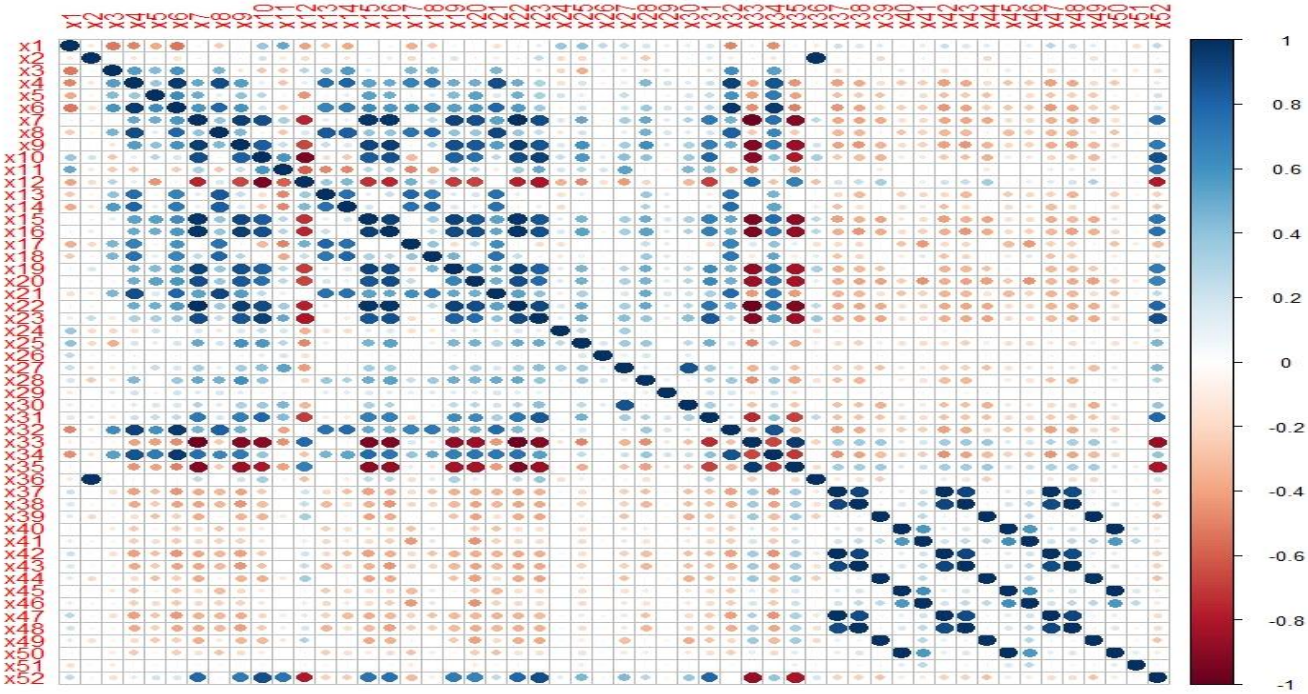

Figure 1. Matrix of correlations between explanatory variables

Where ($j = 1,2, \dots, p$) in equation (17) represents the number of explanatory variables, and VIF is calculated for each explanatory variable, and therefore the model containing p of explanatory variables means the presence of p of variance inflation factors. As for R_j^2, they represent the values of the coefficient of determination in the general linear model of the explanatory variable X_j (which will be the dependent variable and the rest of the variables $X_1, X_2, \dots, X_{j-1}, X_{j+1}, \dots, X_p$ will represent the explanatory variables). The value of VIF when it is (VIF >10), then this indicates the existence of multicollinearity problem between the explanatory variable X_j and the rest of the variables, which necessitates the exclusion of this variable from the model because it is the cause of this problem. On this basis, and by applying equation (16), if $R_j^2 > 0.90$, that is, $R_j > 0.95$ or $R_j < -0.95$, this is evidence of the existence of a problem of the multicollinearity between the explanatory variable X_j and the rest of the explanatory variables, as shown in Table 2. below (the value 1 means the presence of multicollinearity, and the value 0 means the absence of multicollinearity).

Table 2. (VIF) test to detect of multicollinearity problem

X's	VIF	detection	X's	VIF	detection	X's	VIF	detection
x1	81.0502	1	**x18**	70.1141	1	**x35**	63.5843	1
x2	653.4632	1	**x19**	2583.6199	1	**x36**	624.5083	1
x3	167.5692	1	**x20**	1516.2912	1	**x37**	3786.1842	1
x4	39674.9922	1	**x21**	35.4017	1	**x38**	4267.0815	1
x5	5223.4208	1	**x22**	54949.6262	1	**x39**	441.4946	1
x6	30019.549	1	**x23**	222.7619	1	**x40**	38890.5885	1
x7	167660.763	1	**x24**	4.6745	0	**x41**	892777.489	1
x8	28.2239	1	**x25**	4.6879	0	**x42**	2692.0258	1

x9	858.5813	1	**x26**	2.3465	0	**x43**	2776.5774	1
x10	3272.0907	1	**x27**	21.6546	1	**x44**	172.0208	1
x11	23.642	1	**x28**	6.1182	0	**x45**	1449.8673	1
x12	2231.0591	1	**x29**	1.4922	0	**x46**	893278.876	1
x13	41.9109	1	**x30**	17.4733	1	**x47**	1055.8175	1
x14	40.7056	1	**x31**	9.2427	0	**x48**	1083.4683	1
x15	456.7025	1	**x32**	87589.7625	1	**x49**	296.1177	1
x16	458.8644	1	**x33**	141873.368	1	**x50**	33597.0785	1
x17	107.9701	1	**x34**	258.0112	1	**x51**	2.3554	0
						x52	41.9769	1

A comparison criterion is extracted between the estimators of shrinkage models (Lasso, Adaptive Lasso, Elastic Net and Improved Mixed Shrinkage Estimator) by using the Mean Square Error criterion (MSE) to obtain the best estimation method, and the results are as follows:

Table 3. Shows the Best Shrinking Parameter for the four Methods

Shrinkage Method	MSE	The best shrink parameter
Lasso	0.2682562	0.4207053
Adaptive Lasso	0.2669222	0.6269165
Elastic Net	0.2628952	0.4803912
IMSH-Proposed	0.2576002	0.6932827

Table 3 shows that the proposed method (Improved Mixed Shrinkage) achieves the lowest MSE (0.2576002) and the best regularization parameter value (0.6932827) according to (CVP). The estimated values of the coefficients Regression linear model by apply the best shrinkage method (IMSH) show bellow:

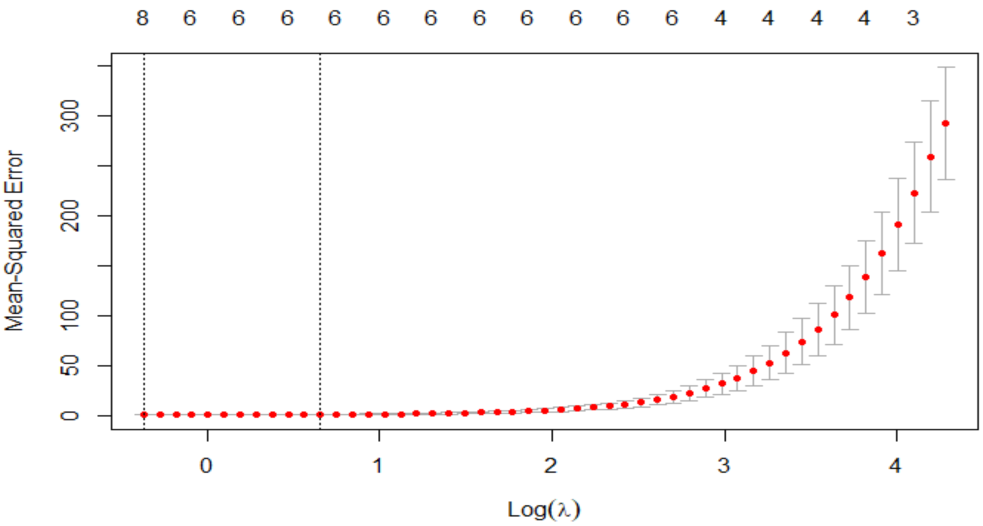

Figure 2. IMSH shrink method parameters based on cross validation Parameter test (CVP**)**

According to Figure 2 the upper axis shows the number of non- zero regression coefficients, which are (p=6) variables out of a total of (52) variables. The best shrinking parameter ($Log(\lambda) = 0.6932827$) corresponding to the lowest (MSE) was obtained.

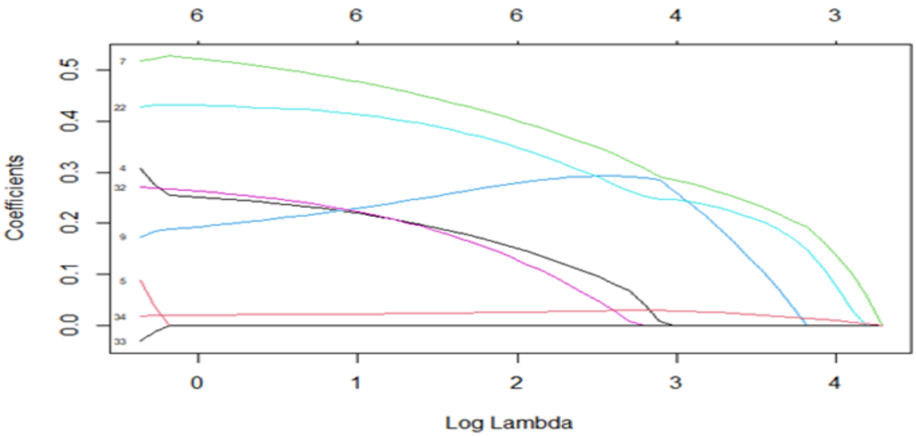

Figure 3. Paths of parameters for the IMSH method based on (Log λ)

As for Figure 3, it shows the parameter paths of the IMSH model based on (Log λ), and we can see when the value of the regularization parameter increases, it is observed that the value of the shrinkage method parameter decreases to become equal to zero, so the ideal number of variables after shrinkage is (p=6).

Table 4. Estimated Coefficients of Regression model Based on (IMSH) Method

Coff.	Estimate	Coff.	Estimate	Coff.	Estimate	Coff.	Estimate
B0	-2.7413375	**B16**	3.454421418	**B32**	0	**B48**	0
B1	0	**B17**	0	**B33**	0	**B49**	0
B2	0	**B18**	0	**B34**	0	**B50**	0
B3	0	**B19**	0.069676628	**B35**	0	**B51**	0
B4	0	**B20**	0	**B36**	0	**B52**	0
B5	0	**B21**	0.10373088	**B37**	0		
B6	10.244516	**B22**	0.327684365	**B38**	0		
B7	0.3307191	**B23**	0	**B39**	0		
B8	0	**B24**	0	**B40**	0		
B9	0.1239313	**B25**	0	**B41**	0		
B10	0	**B26**	0	**B42**	0		
B11	0	**B27**	0	**B43**	0		
B12	0	**B28**	0	**B44**	0		
B13	0	**B29**	0	**B45**	0		
B14	0	**B30**	0	**B46**	0		
B15	3.1092270	**B31**	0	**B47**	0		

Table 4 shows the estimated model after shrinkage includes the estimators of the coefficients of the explanatory variables in addition to the coefficient of the constant term as follows:

$$\hat{y} = -2.74133 + 10.244512X_6 + 0.330719X_7 + 3.109227X_{15} + 3.454421X_{16} + 0.1037308X_{21} + 0.327684X_{22}$$

Where:

- X_6 represent Abio-salts in the body.
- X_7represents the mass of fat in the body.
- X_{15}represents the left arm fat area.
- X_{16}represents the right arm fat area.
- X_{21}represents the muscles in the body.
- X_{22}represents total body fat.

The variables with mainly the greatest effect in diagnosing obesity, increasing one unit of the above variables leads to an increase in weight by (X_6 =10.244512,$X_7 = 0.330719$,X_{15}: 3.10922, X_{16}:3.454421, $X_{21} = 0.1037308$ and $X_{22} = 0.327684$).

4. Conclusions

1- The results of the tests to detect the problem of Multicollinearity were positive, as they showed the presence of Multicollinearity among most of the explanatory variables based on the variance inflation factor (VIF) test.

2- The results showed that the best shrinking parameter among the four methods (Lasso, Adaptive Lasso, Elastic Net and IMSH) was for the IMSH shrinkage method, as it corresponds to the lowest (MSE) based on the cross-validation parameter test (CVP).

3- The new proposed method IMSH achieved the optimal shrinking parameter (λ = 0.6932827) according to the (CVP) test, that leads to have minimum value of mean square error (MSE) equal (0.2576002).

4- The value of the regularization parameter increases, it is observed that the value of the shrinkage method parameter decreases to become equal to zero, so the ideal number of variables after shrinkage is (p=6).

5- The variables with mainly the greatest effect in diagnosing obesity, increasing one unit of the above variables leads to an increase in weight by (X_6 =10.244512,$X_7 = 0.330719$,X_{15}: 3.10922, X_{16}:3.454421, $X_{21} = 0.1037308$ and $X_{22} = 0.327684$).

6- The methods of estimators of shrinkage models can be applied to quantitative and qualitative data at the same time, as they have a methodology for standardizing data before statistical analysis and inference. Also, the reduction models are effective even at the level of low correlation degrees, and therefore they can be relied upon to reduce the variables with low correlation.

Declaration of competing interest

The authors declare that they have no known financial or non-financial competing interests in any material discussed in this paper.

Funding information

No funding was received from any financial organization to conduct this research.

References

[1] A. Araveeporn, "The higher-order of adaptive lasso and elastic net methods for classification on high dimensional data," *Mathematics*, vol. 9, no. 10, p. 1091, 2021.

[2] A. Hazem and A. Yousif, "Comparison between some of the robust penalized estimators using simulation," *Journal of Economics and Administrative Sciences*, vol. 100, no. 32, pp. 490–504, 2017.

[3] A. S. Al-Jawarneh, M. T. Ismail, and A. M. Awajan, "Elastic net regression and empirical mode decomposition for enhancing the accuracy of the model selection," *Int J Math, Eng, Manag Sci*, vol. 6, no. 2, pp. 564–583, 2021.

[4] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Am. Stat. Assoc.*, vol. 96, no. 456, pp. 1348–1360, 2001.

[5] F. Emmert-Streib and M. Dehmer, "High-dimensional LASSO-based computational regression models: Regularization, shrinkage, and selection," *Mach. Learn. Knowl. Extr.*, vol. 1, no. 1, pp. 359–383, 2019.

[6] F. Audrino and L. Camponovo, "Oracle properties and finite sample inference of the adaptive lasso for time series regression models," *SSRN Electron. J.*, 2013.

[7] G. J. M. Mahdi, N. J. Mohammed, and Z. I. Al-Sharea, "Regression shrinkage and selection variables via an adaptive elastic net model," *J. Phys. Conf. Ser.*, vol. 1879, no. 3, p. 032014, 2021.

[8] H. Choi, E. Song, S.-S. Hwang, and W. Lee, "A modified generalized lasso algorithm to detect local spatial clusters for count data," *Adv. Stat. Anal.*, vol. 102, no. 4, pp. 537–563, 2018.

[9]J. R. Wally Zaher and A. Hameed Yousif, "Proposing shrinkage estimator of MCP and Elastic-Net penalties in quantile regression model," *wjps*, vol. 1, no. 3, pp. 126–134, 2022.

[10] H. H. Kim and N. R. Swanson, "Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence," *J. Econom.*, vol. 178, pp. 352–367, 2014.

[11] M. Kayanan and P. Wijekoon, "Performance of LASSO and Elastic net estimators in Misspecified Linear Regression Model," *Ceylon J. Sci.*, vol. 48, no. 3, p. 293, 2019.

[12] A. F. Lukman, K. Ayinde, S. Binuomote, and O. A. Clement, "Modified ridge-type estimator to combat multicollinearity: Application to chemical data," *J. Chemom.*, vol. 33, no. 5, p. e3125, 2019.

[13] L. Wang, J. Shen, and P. F. Thall, "A modified adaptive Lasso for identifying interactions in the Cox model with the heredity constraint," *Stat. Probab. Lett.*, vol. 93, pp. 126–133, 2014.

[14] L. T. P. Thao and R. Geskus, "A comparison of model selection methods for prediction in the presence of multiply imputed data," *Biom. J.*, vol. 61, no. 2, pp. 343–356, 2019.

[15] R. Tibshirani, "Regression shrinkage and selection via the lasso: A retrospective: Regression Shrinkage and Selection via the Lasso," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 73, no. 3, pp. 273–282, 2011.

[16] X. Cui *et al.*, "Adaptive LASSO logistic regression based on particle swarm optimization for Alzheimer's disease early diagnosis," *Chemometr. Intell. Lab. Syst.*, vol. 215, no. 104316, p. 104316, 2021.

[17] S. Muhammadullah, A. Urooj, F. Khan, M. N. Alshahrani, M. Alqawba, and S. Al-Marzouki, "Comparison of weighted lag adaptive LASSO with Autometrics for covariate selection and forecasting using time-series data," *Complexity*, vol. 2022, pp. 1–10, 2022.

[18] S. A. -reda Al-Sabaah and S. M. Al-Kafishi, "Parameters Estimation of the Multiple Linear Regression Model Under Multicollinearity problem," *Journal of Economics and Administrative Sciences*, vol. 12, no. 1, pp. 1–28, 2020.

[19] S. Kwon and S. Lee, "Sufficient conditions for the oracle property in penalized linear regression," *The Korean Journal of Applied Statistics*, vol. 34, no. 2, pp. 279–293, 2021.

[20] J. C. van Houwelingen, "Shrinkage and penalized likelihood as methods to improve predictive accuracy," *Stat. Neerl.*, vol. 55, no. 1, pp. 17–34, 2001.

[21] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, 2005.

[22]H. Zou, "The adaptive lasso and its oracle properties," *J. Am. Stat. Assoc.*, vol. 101, no. 476, pp. 1418–1429, 2006.