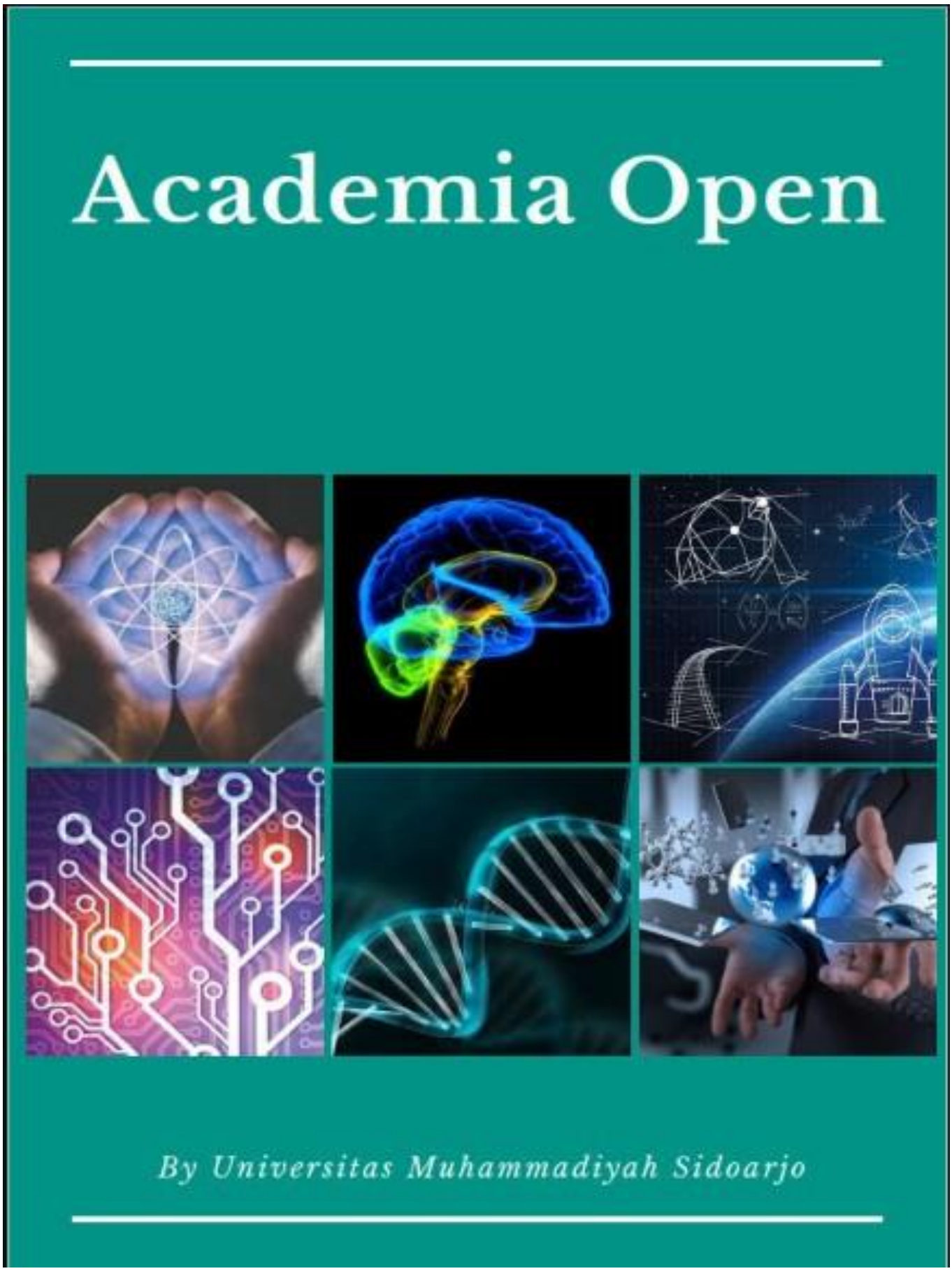
Academia Open

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.15042



ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Academia Open

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.15042

Table Of Contents

Academia Open

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.15042

Originality Statement

The author[s] declare that this article is their own work and to the best of their knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the published of any other published materials, except where due acknowledgement is made in the article. Any contribution made to the research by others, with whom author[s] have work, is explicitly acknowledged in the article.

Conflict of Interest Statement

The author[s] declare that this article was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright Statement

Academia Open

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.15042

EDITORIAL TEAM

Complete list of editorial team (link)

Complete list of indexing services for this journal (link)

How to submit to this journal (link)

Academia Open

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.15042

Article information

Check this article update (crossmark)

Check this article impact (*)

Save this article to Mendeley

(*) Time for indexing process is various, depends on indexing database platform

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Academia Open

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.15042

A Comparison Some Penalized Variable Selection Methods For Analysis General Linear Regression Model With Application

Tareq Azeez Salih, tazeez@uowasit.edu.iq(*)

College of Administration and Economic, Wasit University, Iraq

Ali Hameed Yousif, ahameed@uowasit.edu.iq

College of Administration and Economic, Wasit University, Iraq

Zainab Kadhum Mezher, zmezher@uowasit.edu.iq

College of Administration and Economic, Wasit University, Iraq

(*) Corresponding author

Abstract

General Background The general linear regression model balances nonparametric flexibility with parametric explanatory power. **Specific Background** High-dimensional data requires robust regularization to isolate critical predictors. **Knowledge Gap** Classical approaches falter with small datasets or multicollinearity, triggering the curse of dimensionality. **Aims** This study compares smoothly clipped absolute deviation methods and the mini-max concave penalty approach for simultaneous parameter estimation and covariate selection. **Results** Simulations across various sample sizes and correlation intensities show that the SCAD-L2 penalty consistently achieves the lowest average error. **Novelty** Empirical implementation using clinical records from Al-Kut Teaching Hospital confirms SCAD-L2 isolates critical variables by driving non-significant coefficients to zero. **Implications** This optimized methodology enhances predictive accuracy and simplifies model architectures without losing explanatory control.

Keywords: General Linear Regression, Variable Selection, Penalty Function, SCAD-L2 Method, MCP Method

Key Findings Highlights

SCAD-L2 consistently outperforms alternative regularization methods across diverse sample sizes and high correlation configurations.

Empirical medical database validation successfully isolated critical blood urea determinants by eliminating non-significant clinical indicators.

The integrated mathematical algorithm achieves stable parameter convergence while minimizing the overall mean absolute error.

Published date: 2026-07-15

Academia Open

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.15042

1. Introduction

Parametric methods are often inappropriate for data that is small or has no known distribution. Whereas, nonparametric methods can contribute to some extent in obtaining a practical ability to avoid the problem Multicollinarity contained in this type of data As the study of general linear regression model, will elicit the problem curse of dimensionality Which most researchers suffer from and restricts them and prevents their progress towards generalizing a univariate case to a bivariate case and then to a multivariate case.

Accordingly, the issue of developing penalized methods for analyzing big data is one of the things that are required.

In a large measure to modern methods represented by parametric statistical methods for data analysis, which provide us with valid inferences in the event that the conditions are there is a non-linear structure of the data Most economic models follow the behavior of semi-parametric models, which, as their name As the parametric part represents the regression function that is supposed to be linear in the observations of its explanatory variables, while the non-parametric part includes the distribution of the population suggests, is general linear regression model of great importance in the field of statistics and has a wide range of applications in the fields of biostatistics, medicine, finance, business and others Semi-parametric models are of great importance when non-parametric models perform poorly For example, when the problem is dimensionality, it leads to high variance estimates . From the foregoing, general linear regression model has become more common in the statistical literature than other models because it preserves the flexibility of nonparametric models and maintains the explanatory power of parametric models,

In this research, the general linear regression model was studied, and penalized estimation methods were used in its analysis, which are smoothly clipped absolute deviation (SCAD-L1) penalty function method, are smoothly clipped absolute deviation (SCAD-L2) penalty function method with mini-max concave penalty (MCP) penalty function method (MCP) method . The simulation method was used for the purpose of comparing the various estimation methods and for several experiments, and finding the best estimation method based on the comparison standard, Average Mean Absolute Error (AMAE). And then using real data to compare the various estimation methods and reach the most important conclusions of this research.

2. The General Linear Regression Model

The general regression model is used to analyze the relationship between a variable called the response variable and several variables called the explanatory variables, considering the response variable as a function of the explanatory variables. Based on the nature of the relationship between the response variable and the explanatory variables, there are two types of general (multiple) regression models. The first is the general linear regression model, based on the assumption that the relationship between the variables is linear. The second is the general nonlinear regression model, based on the assumption that the relationship between the variables is nonlinear. Our study will focus on the general linear regression model. The general linear regression model is based on the assumption of a linear relationship between the response variable and a number of explanatory variables and a random error term. This means that the relationship between the response variable and the explanatory variables is described by a linear function.

The general linear regression model is based on the assumption of a linear relationship between the response variable and a number of explanatory variables and a random error term. For a random sample of size n, assuming that the response variable is represented by the variable Y_i and the explanatory variables by the variables X_ij where (i = 1, 2, ..., n) and (j=1, 2, ..., p), and where p represents the number of explanatory variables included in the model, and the random error term is represented by the variable U_i, then the general formula for the general linear regression model for each observation of the response variable randomly selected from a sample of size n is as follows :

$$Y_i=\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_p X_p+u_i \quad ,i=1,2,...n \qquad .... (1)$$

The above template can be written in shorthand using the sum symbol as follows:

$$Y_i= \sum_{j=0}^{p} [\beta_j X_{ij}+ u_i] \qquad .... (2)$$

i = 1,2,3,.....n and j = 0,1,2,...,p , X_0j=1

Matrices can be used to write the model formula as follows:

$$Y=X\beta +U \qquad ....(3)$$

Since:

Y: is a value vector of the response variable of degree n*1

β: is the vector representing the unknown model parameters that need to be estimated to order (p+1)*1.

X: represents the matrix of explanatory variables of order n*(p+1).

U:represents the vector representing the random error values of order n*1

3. Variable selection

Sometimes a large number of independent variables, Xi, is available for a given modeling problem, and not all of these predictor variables may contribute equally well to the explanation of the predicted variable Y. Some of the independent variables may not contribute at all to the model.

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.15042

Thus we have to select from these variables to obtain a model which contains as little variables as possible while still being the best model, Therefore, some punitive methods will be used to select the significant variable and exclude the non-significant variable from the study model. Variable selection is a technique used in statistical modeling to identify the most significant covariates , ensuring that only the most relevant factors are included, such as in General linear regression model. This process enhances model accuracy by filtering out less important variables, allowing for a more streamlined and effective analysis. Through variable selection, researchers can better understand the relationships within their data and improve predictive performance. And the methods used to solve this problem are the MCP method and SCAD-L1 and SCAD-L2 .

4. Advantages and disadvantages for linear regression model

4.1 Advantages

1- Linear regression is a relatively simple algorithm, making it easy to understand and implement. The coefficients of the linear regression model can be interpreted as the change in the dependent variable for a one-unit change in the independent variable, providing insights into the relationships between variables.

2- Linear regression is computationally efficient and can handle large datasets effectively. It can be trained quickly on large datasets, making it suitable for real-time applications.

3- Linear regression is relatively robust to outliers compared to other machine learning algorithms. Outliers may have a smaller impact on the overall model performance.

4.2 Disadvantages

1- Linear regression assumes a linear relationship between the dependent and independent variables. If the relationship is not linear, the model may not perform well.

2- Linear regression is sensitive to multicollinearity, which occurs when there is a high correlation between independent variables. Multicollinearity can inflate the variance of the coefficients and lead to unstable model predictions.

3- Linear regression provides limited explanatory power for complex relationships between variables.

5. Tuning Parameter and Penalty Parameter :[1],[2].[6]

A tuning parameter (λ), sometimes called a penalty parameter, controls the strength of the penalty term in ridge regression and lasso 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.

When λ = 0, no parameters are eliminated. The estimate is equal to the one found with linear regression.

As λ increases, more and more coefficients are set to zero and eliminated.

When λ = ∞, all coefficients are eliminated.

There is a trade-off between bias and variance in resulting estimators. As λ increases, bias increases and as λ decreases, variance increases. For example, setting your tuning parameter to a low value results in a more manageable number of model parameters and lower bias, but at the expense of a much larger variance.

Penalty Parameter L1 and L2 Penalties

Tuning parameters are part of a process called regularization, which works by biasing data towards particular values. Popular regularization methods use either an L1 or L2 penalty (or sometimes, a combination of both):

L1 penalties limits the size of the coefficients and can result in sparse models (i.e. models with a small number of coefficients); Some coefficients are eliminated.

L2 penalties do not result in sparse models because all coefficients are shrunk by the same factor and none are eliminated.

6. Selection of the tuning parameter:[1],[8]

The choice of λ is critical. In practice λ is usually selected by minimizing the generalized cross validation criterion

$$\mathrm{GCV}_\lambda = \frac{\|y - X\hat{\beta}_\lambda\|^2}{n\left(1-\frac{DF_\lambda}{n}\right)^2} = \frac{\hat{\sigma}_\lambda^2}{(1-DF_\lambda/n)^2} \qquad \text{----------(4)}$$

Where

$$\hat{\sigma}_\lambda^2 = n^{-1}\|y - X\hat{\beta}_\lambda\|^2$$

DF_λ : *is the generalised degrees of freedom given by*

$$DF_\lambda = tr\{X(X'X + n\Sigma_\lambda)^{-1}X'\} \text{ and}$$

$$\Sigma = diag\left\{\frac{P_\lambda'(|\hat{\beta}_{\lambda 1}|)}{|\hat{\beta}_{\lambda 1}|}, \ldots, \frac{P_\lambda'(\hat{\beta}_{\lambda d})}{|\hat{\beta}_{\lambda d}|}\right\}$$

The diagonal elements of Σ_λ are coefficients of quadratic terms in the local quadratic approximation to the smoothly clipped absolute deviation penalty function $P_\lambda(\cdot)$ (Fan & Li, 2001). Since some coefficients of the estimator of β are exactly equal to zero, DF_λ is calculated by replacing Σ_λ

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Academia Open
Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.15042

with its submatrix corresponding to the selected covariates, and by replacing $\Sigma_\lambda$with its corresponding submatrix. The resulting optimal tuning parameter is $\hat{\lambda}_{GCV} = argmin_\lambda \text{GCV}_\lambda$.

7-penalty function . [8],[10]

Penalization method are used As a procedure for estimating and selecting the significant variable simultaneously To perform the process of selecting the significant variable, some formulas of the following penalty functions were used selecting the significant variable, some formulas of the following penalty functions were used

Penalty function	Formula
1L- SCAD Is called l1- penalty function	$= \lambda\|\beta_j\| p_\lambda(\|\beta_j\|)$ if $\|\beta_j\| < \lambda$, with $q = 1$
L2- SCAD Is called L2- penalty function	$= \lambda\|\beta_j\|^2 p_\lambda(\|\beta_j\|)$ if $\|\beta_j\| < \lambda$, with $q = 2$
SCAD -Lq Is called bridge penalty function	$= \lambda\|\beta_j\|^q p_\lambda(\|\beta_j\|)$ if $\|\beta_j\| < \lambda$, with $0 < q < 1$
MCP	$=\lambda\|\beta\| - (\frac{\beta^2}{2a}) p_\lambda(\|\beta\|)$ if $\|\beta\| \leq a\lambda$, with $a > 1$

8- Methods of penalized Estimation

There are many special statistical methods estimating and selecting the significant variable for general linear regression model Which leads to a high improvement in the accuracy of the model as well as its computational efficiency The following are some of the statistical methods used:

8-1 Smoothly Clipped Absolute Deviation(SCAD) –L2 with L1 Method :[1],[4],[14],[15]

That is a penalty function –Lq It is suggested by the researcher Tibshirani in year (1996) And it has the following :

$= \lambda|\beta_j|^q$ with $q > 0$, Whereas: $p_\lambda(|\beta_j|)$

q: Represent The shrinkage parameter

and penalty function Special cases are as follows:

1) If the value $0 < q < 1$ The penalty function is called bridge penalty And it has the following form :

$= \lambda|\beta_j|^q$ with $0 < q < 1$ $p_\lambda(|\beta_j|)$

2) If the value $q=0$ The penalty function is called L0- penalty function And it has the following form :

$= \lambda$ with $q = 0$ $p_\lambda(|\beta_j|)$

3) If the value $q=1$ The penalty function is called L1- penalty function And it has the following form :

$= \lambda|\beta_j|$ with $q = 1 p_\lambda(|\beta_j|)$

4) If the value $q=2$ The penalty function is called L2- penalty function (ridge penalty function) And it has the following form :

$= \lambda|\beta_j|^2$ with $q = 2$ $p_\lambda(|\beta_j|)$

In partial regression and general linear regression model $y = X\beta + \epsilon$

This method has three properties Transactions estimators which includes the following :(Unbiasedness, Sparsity and Continuity)

It is defined by the following new partial least squares function:

$$L_{\lambda_1,\lambda_2}.\text{SCAD} - l_2(\beta) = \frac{1}{2}\|y - X\beta\|^2 + \sum_{j=1}^{d} f_{\lambda_1}(\beta) + \lambda_2 \|\beta\|^2 \text{ ---------- (5)}$$

Whereas:

: Represent L_2 normally and $\|\beta\|^2 = \sum_{j=1}^{d} \beta_j^2$ $\|.\|^2$

: Represent SCAD penalty function It is defined by the following equation . $f_\lambda(\beta$

Academia Open
Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.15042

$$f_{\lambda(\beta)} = \begin{cases} \lambda|\beta| & , \text{if } 0 \le |\beta| < \lambda \\ \frac{-\beta^2 - 2a\lambda|\beta| + \lambda^2|}{2(a-1)} & , \text{if } \lambda \le |\beta| < a\lambda \\ \frac{(a+1)\lambda^2}{2} & , \text{Other wise} \end{cases} \quad \text{---------- (6)}$$

Whereas:

The constant (a) has a value greater than two Where the researchers used Fan & Li(a=3.7) for average sample size . This method is suggested by researchers Fan and Li) in year (2001) .

And to estimate the coefficients using the penal least squares method We first suppose that the matrix X is Orthonormal whereas $X^TX = I$ and Under this assumption We can integrate estimator SCAD – L_2 or SCAD – L_1 With other partial least squares including SCAD, assuming that:

$$P_{\lambda(\beta)} = \sum_{j=1}^{d} p_\lambda(\beta_j) \quad \text{---------- (7)}$$

When my penalty limits is used.

The formula for the partial least squares function can be written as follows:

$$L_\lambda(\beta) = \frac{1}{2}\|y - X\beta\|^2 + P_\lambda(\beta)$$

$$L_\lambda(\beta) = \frac{1}{2}\|y - \hat{y}\|^2 + \frac{1}{2}\|y - X\beta\|^2 + \sum_{j=1}^{d} P_\lambda(\beta_j)$$

$$L_\lambda(\beta) = \frac{1}{2}\|y - \hat{y}\|^2 + \frac{1}{2}\left\|\hat{\beta}_{ols} - \beta\right\|^2 + \sum_{j=1}^{d} P_\lambda(\beta_j)$$

Whereas:

: It is the estimator of the ordinary least squares(OLS) $\hat{\beta}_{ols} = X^T y$

: whereas $\|y - \hat{y}\|^2$ does not depend on β $\hat{y} = X\hat{\beta}_{ols}$

And to reduce the penalty squares function $L_\lambda(\beta)$

It is equivalent to reducing the following function :

$$L_\lambda(\beta) = \frac{1}{2}\left\|\hat{\beta}_{ols} - \beta\right\|^2 + \sum_{j=1}^{d} P_\lambda(\beta_j)$$

and the first partial derivative of β_j It is given by:

$$\frac{\partial}{\partial \beta_j} L_\lambda(\beta) = \beta_j - \hat{\beta}_{j,ols} + \acute{P}_\lambda(\beta_j)$$

Whereas:

$\hat{\beta}_{j,ols}$:Represent jth elements for $\hat{\beta}_{ols}$ By setting all the first partial derivatives equal to zero We get the estimate for jth It is given according to the following formula:

$$\hat{\beta}_j = \hat{\beta}_{j,ols} - \acute{P}_\lambda(\hat{\beta}_j) \quad \text{---------- (8)}$$

And penalty function SCAD know through $P_{\lambda,SCAD(\beta)} = \sum_{j=1}^{d} f_\lambda(\beta_j)$

Then estimated SCAD for ith It is given according to the following formula:

$$\hat{\beta}_{j,SCAD} = \left\{ \begin{array}{ll} \text{Sing}\left(\hat{\beta}_{j,ols}\right)\left(|\hat{\beta}_{j,ols}| - \lambda\right)+ & , \text{if } |\hat{\beta}_{j,ols}| \le 2\lambda \\ \frac{(a-1)\hat{\beta}_{j,ols} - a\lambda.\text{sign}(\hat{\beta}_{j,ols})}{a-2} & , \quad \text{if } 2\lambda < |\hat{\beta}_{j,ols}| \le a\lambda \\ \hat{\beta}_{j,ols} & \text{Otherwise} \end{array} \right\} \text{----------(9)}$$

Where:

$\hat{\beta}_{j,ols}$:is Estimation(unbiased) When the estimator has an absolute value greater than (aλ) . And as the penalty least squares were defined to SCAD - L2, the penalty function is :

$$P_{\lambda_1,\lambda_2}, SCAD.\, L_2(\beta) = \sum_{j=1}^{d}(f_{\lambda_1}(\beta_j) + \lambda_2\hat{\beta}_j) \quad \text{----------(10)}$$

Whereas:

Academia Open
Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.15042

$P_{\lambda_1,\lambda_2}(\beta) = f_{\lambda_1}(\beta) + \lambda_2\beta^2$

And it has the following first derivative:

$$P'_{\lambda_1,\lambda_2}(\beta) = \left\{ \begin{array}{ll} \lambda_1 \, \mathrm{Sing}(\beta) + 2\lambda_2\beta & , \text{if } 0 \le |\beta| \le \lambda_1 \\ \frac{a\lambda_1 \, \mathrm{sign}(\beta) - \beta}{a-1} + 2\lambda_2\beta & , \quad \text{if } \lambda_1 \le |\beta| < a\lambda_1 \\ 2\lambda_2\beta & , \quad \text{Otherwise} \end{array} \right\} \text{------(11)}$$

We put the first partial derivative of the partial squares SCAD-L2 equal to zero And we get solutions Naive SCAD-L2 to jth and It is given according to the following equation:

$$\hat{\beta}_{j,naive} = \left\{ \begin{array}{ll} \frac{\mathrm{Sing}\left(\hat{\beta}_{j,ols}\right)\left(|\hat{\beta}_{j,ols}| - \lambda_1\right) +}{1 + 2\lambda_2} & , \text{if } |\hat{\beta}_{j,ols}| \le 2\lambda_1(1+\lambda_2) \\ \frac{(a-1)\hat{\beta}_{j,ols} - a\lambda_1 . \, \mathrm{sign}\,(\hat{\beta}_{j,ols})}{-1 + (a-1)(1+2\lambda_2)} & , \quad \text{if } 2\lambda_1(1+\lambda_2) < |\hat{\beta}_{j,ols}| \le a\lambda_1(1+2\lambda_2) \\ \frac{\hat{\beta}_{j,ols}}{1+2\lambda_2} & \text{Otherwise} \end{array} \right\}$$

Specifically, it is:

$\hat{\beta}_{j,naive} = \hat{\beta}_{j,ols}/(1+2\lambda_2)$ ---------- (12)

for the greatest value to $|\hat{\beta}_{j,ols}|$ and is biased estimator Finally, it is estimator SCAD-L2 It is defined by the following formula :

$\hat{\beta}_{SCAD-L_2} = (1+2\lambda_2)\hat{\beta}_{naive}$ ---------- (13)

And it is an unbiased estimator.

Note that the value of the constant parameter (a=3.7) and values of tuning parameters penalty ($\lambda_1 = 1$, $\lambda_2 = 0.5$).

And in the general linear regression model then design matrix (X) may not be orthonormal then the estimator $\hat{\beta}_{naive}$ the estimate can be obtained by reducing the following equation:

--- (14) $L_{\lambda_1,\lambda_2,SCAD} - L_2(\beta) = \frac{1}{2}(y - X\beta)^T(y - X\beta) + \sum_{j=1}^{d} f_{\lambda_1}(\beta_j) + \lambda_2\|\beta\|^2$

In year (2007) the researchers found Huang and Xie Until that SCAD method It has two main advantages :

1- The process of selecting a variable with the SCAD method is continuous Hence; it is more stable than the usual variable selection methods As a subset selection method which is a discrete and unstable process.

2- The SCAD method is computationally intensive for high dimensional data On the other hand, the calculation of the usual variable selection methods is the method of choosing a subset It is complex and not feasible to calculate when the explanatory variable-P-dimensions are large .

Method estimates can be obtained SCAD-L1 for semi parametric single index model Through the following formula:

$Q\,(\hat{g}, \hat{\beta}) = arg\ min_{\beta}\ \{\sum_{i=1}^{n}(Yi - g(X_i^T\,\beta))\}^2 + n\sum_{j=1}^{p}(\lambda|\beta_j|)$

While method estimates are obtained SCAD-L2 for semi parametric single index model through the following formula:

$Q\,(\hat{g}, \hat{\beta}) = arg\ min_{\beta}\ \{\sum_{i=1}^{n}(Yi - g(X_i^T\,\beta))\}^2 + n\sum_{j=1}^{p}(\lambda|\beta_j|^2)$

where
L2 penalty:
$p_\lambda(|\beta_j|) = \lambda|\beta_j|^2 = \lambda\sum_{j=1}^{d}\beta_j^2$
L1 penalty: $p_\lambda(|\beta_j|) = \lambda|\beta| = \lambda\sum_{j=1}^{d}\beta_j$
Where tuning parameters is: $\lambda_1 = 1\ and\ \lambda_2 = 0.5$

8.2: MCP Method. [5],[7],[9],[13]

Mini max Concave Penalty Method

Working method MCP On estimating and selecting significant parameters at the same time And it is a penalty function MCP from Concave penalty functions Which is efficient in estimation and selection If the oracle properties are true .

A penalty function has been proposed MCP by researcher Zhang (2010) According to the following formula:

Academia Open
Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.15042

$$P_{MCP}(\beta) = \begin{cases} \lambda|\beta| - \frac{\beta^2}{2a} & if\ |\beta| \le a\lambda \\ \frac{a\lambda^2}{2} & if\ otherwise \end{cases} \quad ----------(15)$$

where:

λ : *represent* tuning parameter

a : Another *represent* tuning parameter And it's worth (a > 1)

And when MCP Concavity results in penalty loss points At certain thresholds for variable selection as well as unbiased Penalty function of MCP can be understood Through the derivative of the function and as follow:

$$P'_{MCP}(\beta) = \begin{cases} sing\ (\beta)\left(\lambda - \frac{|\beta|}{a}\right) & if\ |\beta| \le a\lambda \\ 0 & if\ otherwise \end{cases} \quad ---------- (16)$$ It matches the derivative of the following function:

$$P'_{MCP}(\beta) = \begin{cases} \left(\lambda - \frac{|\beta|}{a}\right) & if\ |\beta| \le a\lambda \\ 0 & if\ otherwise \end{cases}$$

As a penalty function MCP It starts with the same penalty rate as in the rest of the other penalty functions But the penalties are minimizes or reduced until they reach $\beta \ge a\lambda$.

Also, until the rate of penalization reaches zero , And minimizes the partial regression of a function (MCP) It is given as follows:

$+n\sum_{J=1}^{P} P_{MCP}(|\beta_k|)$ ---------- (17) $\sum_{i=1}^{n}(Y_i - X_i^T\beta)^2$

It has been suggested MCP method by researchers Wang and Yin in year (2008) Concerning the estimation and selection of significant variables as follows:

$$\hat{\beta}^{MCP} = \text{argmin}\ \{\sum_{i=1}^{n}\sum_{j=1}^{n}\{ Y_i - a_j - (X_i - X_j)^T\beta b_j\}^2 W_{ij} + n\sum_{j=1}^{n} P_{MCP}(|\beta_j|)$$
$$\beta: \|\beta\| = 1$$
$$a_j, b_j,\ j = 1,2, \dots, n$$

where:

$$W_{ij} = k_h\left(X_{ij}^T\hat{\beta}^{(0)}\right) = \frac{k\left(\frac{X_i^T\hat{\beta}^{(0)} - X_j^T\hat{\beta}^{(0)}}{h}\right)}{\sum_{i=1}^{n} k\left(\frac{X_i^T\hat{\beta}^{(0)} - X_j^T\hat{\beta}^{(0)}}{h}\right)}$$

Because the penalize estimator can be known for penalty function – MCP for general linear model as follows:

$$\hat{\beta}^{MCP} = \arg\min\ \{\sum_{i=1}^{n}(Y_I - \sum_{j=1}^{p}\beta_j X_{ij}\}^2 + \lambda|\beta| - \left(\frac{\beta^2}{2a}\right)\}$$
$$\beta: \|\beta\| = 1$$
$$a_j, b_j,\ j = 1,2, \dots, n$$

It represents the estimator $\lambda|\beta| - \left(\frac{\beta^2}{2a}\right)\ penalty\ function - MCP$

Under restraint $|\beta| \le a\lambda$ It can take the following form:

$$\hat{\beta}^{MCP} = \text{argmin}\{\sum_{i=1}^{n}(Y_I - \sum_{j=1}^{p}\beta_j X_{ij}\}^2 + \left(a\frac{\lambda^2}{2}\right)\}$$
$$\beta: \|\beta\| = 1$$
$$a_j, b_j,\ j = 1,2, \dots, n$$

It represents the amount $a\frac{\lambda^2}{2}$ penalty function –MCP under the following constraint$|\beta| > a\lambda$

Algorithm can be explained MCP According to the following steps:

Step (1): We assume an initial estimate of the parameter $\hat{\beta}^{(0)}$

Using the ordinary least squares method (OLS).

Step (2) : is installed$\hat{\beta}^{(0)} = \hat{\beta}$ Parameters vector is calculated $\hat{a}_j$, $\hat{b}_j$Through the following formula :

Academia Open
Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.15042

$$\begin{pmatrix}\hat{a}_j\\ \hat{b}_j\end{pmatrix}=\left\{\sum_{i,j}^{n} w_{ij}^{\hat{\beta}^{(0)}}\begin{pmatrix}1\\ X_{ij}^T\hat{\beta}\end{pmatrix}\begin{pmatrix}1\\ X_{ij}^T\hat{\beta}\end{pmatrix}^T\right\}^{-1}.\sum_{i,j}^{n} w_{ij}^{\hat{\beta}^{(0)}}\begin{pmatrix}1\\ X_{ij}^T\hat{\beta}\end{pmatrix} Y_i$$

Where

$$w_{ij}^{\hat{\beta}^{(0)}}=k_h\left(X_{ij}^T\hat{\beta}^{(0)}\right)=\frac{k\left(\frac{x_i^T\hat{\beta}^{(0)}-x_j^T\hat{\beta}^{(0)}}{h}\right)}{\sum_{i=1}^{n}k\left(\frac{x_i^T\hat{\beta}^{(0)}-x_j^T\hat{\beta}^{(0)}}{h}\right)}$$

Step (3) : is installed $\hat{a}_j$, $\hat{b}_j$ Through the following formula : And estimation vector parameters

$$\hat{\beta}^{MCP}=\operatorname{argmin}\left\{\sum_{j=1}^{n}\sum_{i=1}^{n}\{Y_i-\hat{a}_j-\hat{b}_jX_{ij}^T\beta\}^2 w_{ij}^{\hat{\beta}^{(0)}}+\lambda|\beta|-\left(\frac{\beta^2}{2a}\right)\right\}$$

$\beta: \|\beta\|=1$

$a_j, b_j,\ i,j=1,2,\dots,n$

represents the amount $\lambda|\beta|-\left(\frac{\beta^2}{2a}\right)$ penalty function – MCP under the following constraint $|\beta|\le a\lambda$

Step (4): The two steps are repeated second and third with

$$\hat{\beta}^{(0)}=\text{sign}\,(\hat{\beta}_1)\frac{\hat{\beta}}{\|\hat{\beta}\|}$$

Even convergence and the last vector is an estimator vector MCP for $\hat{\beta}^{(0)}$and known through $\hat{\beta}^{MCP}$ And the final estimation of the link function g(.) is

$\hat{a}_j=\hat{g}\,(u\,,\hat{\beta}^{MCP}\,)$.

9- Comparison standard Mean Absolute Error (MAE) [6],[9]:

Mean Absolute Error is an evaluation metric used to calculate the accuracy of a regression model. MAE measures the average absolute difference between the predicted values and actual values.

Mathematically error (AMAE) to can be calculated by the following formula which is used for simulation experiments:

$$AMAE=\frac{1}{n}E\sum_{i=1}^{n}\{|Y_i-\hat{Y}_i|\} \quad \text{---------- (18)}$$

Here

- n: is the number of observations
- Yi: represents the actual values.
- $\hat{Y}_i$: represents the predicted values

Lower MAE value indicates better model performance.

And Standard of Mean Average Error (MAE) to can be calculated by the following formula which is used for real data:

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|Y_i-\hat{Y}_i| \quad \text{---------- (19)}$$

10- Simulation studies. [3],[11]

Simulation experiments are performed using (200) data set each consisting of a random sample size of (n=25, n=100) observation With a repetition of 200 times for each experiment, from the following model : $Y=X^T\beta+\sigma\in$ as follows :

With p=8, Where X= ($x_1, x_2, x_3, x_4, x_5, x_6, x_7 x_8$) , $\beta=(3,1.5,0,0,2,0,0,0)^T$

With Xi are iid $\sim N(0,\sigma^2)$ where: $\in_i \sim N(0,\sigma^2)$ with $\sigma_\in^2=3$ and correlation between Xi and Xj is

$=\rho$ and $X\sim N(0,\Sigma)$ for i,j=1,2,…,,8 with three values of where explored 0.25, 0.5, 0.95 with $\in_i$ and X are inedpendent . $\sum_{ij}=\rho^{|i-j|}$

Table1 : Represents the estimated value and standard error of the parameter vector for all estimation methods when $p=8, \rho=0.25$ when $\sigma_\in^2=3$

Academia Open

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.15042

n	variable	methods					
		SCAD-L1		SCAD-L2		MCP	
		Average($\hat{\beta}$)	S.E	Average($\hat{\beta}$)	S.E	Average($\hat{\beta}$)	S.E
25	1	0.84505	0.27199	0.72501	0.08098	0.37684	0.19629
	2	0.21996	0.14532	0.43123	0.11939	0.00000	0.20185
	3	-0.24059	0.35695	0.00000	0.26837	0.72871	0.35001
	4	0.24237	0.14612	0.00000	0.07477	0.34603	0.34259
	5	0.34001	0.01432	0.00000	0.07477	0.00000	0.26220
	6	0.00000	0.28175	0.44680	0.17817	0.36853	0.13134
	7	-0.05173	0.15982	-0.29792	0.14854	0.00000	0.25735
	8	0.00000	0.28137	0.00000	0.15701	0.26720	0.05913
	AMAE	1.90770		1.64146		1.82769	
100	1	0.81697	0.01237	0.58138	0.14702	0.45589	0.49159
	2	0.29344	0.10247	0.64420	0.27487	0.24395	0.12531
	3	0.10966	0.09839	0.00000	0.04904	0.42042	0.22920
	4	0.23564	0.20006	0.00000	0.18023	0.29044	0.27128
	5	0.37229	0.05798	0.46218	0.06602	0.37670	0.07930
	6	0.11388	0.02778	-0.13085	0.03653	0.23358	0.19128
	7	0.10678	0.01121	0.09644	0.17097	0.42778	0.00814
	8	0.12613	0.14684	-0.08335	0.10416	0.30342	0.12878
	AMAE	1.9738		1.83737		1.84204	

Table2 : Represents the estimated value and standard error of the parameter vector for all estimation methods when $p = 8, \rho = 0.50$ when $\sigma_{\epsilon}^2 = 3$

n	variable	methods					
		SCAD-L1		SCAD-L2		MCP	
		Average($\hat{\beta}$)	S.E	Average($\hat{\beta}$)	S.E	Average($\hat{\beta}$)	S.E
25	1	0.75663	0.18803	0.83783	0.29513	0.09460	0.45326
	2	0.00000	0.35708	0.00000	0.40229	0.67793	0.18385
	3	-0.14142	0.13409	-0.28415	0.20440	-0.33916	0.27669
	4	0.39319	0.25518	0.00000	0.14028	0.00000	0.15510
	5	0.21409	0.13567	0.16174	0.12026	0.39223	0.09010
	6	-0.18378	0.13621	0.42017	0.28319	-0.15565	0.13858
	7	0.15771	0.22553	0.12074	0.29035	-0.31889	0.25184
	8	0.32543	0.10699	0.00000	0.15984	0.36965	0.18576
	AMAE	1.87793		1.73744		2.08680	
100	1	0.75022	0.06445	0.62241	0.13081	0.87565	0.10151
	2	0.51880	0.12719	0.72092	0.28673	0.25826	0.08871
	3	0.08314	0.05905	0.00000	0.09471	-0.25722	0.21946
	4	0.22118	0.13819	-0.01740	0.07830	0.16551	0.16383
	5	0.19270	0.10488	0.30406	0.04850	0.23688	0.10863
	6	0.23696	0.06522	0.00000	0.14530	0.00000	0.00000
	7	0.06767	0.06108	0.00000	0.10942	0.09531	0.05503
	8	0.11962	0.0655	0.01100	0.14572	-0.08721	0.17273
	AMAE	1.95681		1.93395		1.95870	

Table3 : Represents the estimated value and standard error of the parameter vector for all estimation methods when $p = 8, \rho = 0.95$ when $\sigma_{\epsilon}^2 = 3$

		Average($\hat{\beta}$)	S.E	Average($\hat{\beta}$)	S.E	Average($\hat{\beta}$)	S.E
25	1	0.61799	0.27916	0.32745	0.21450	0.31460	0.16434
	2	0.58986	0.6575	0.77660	0.17211	0.35089	0.42195
	3	0.20036	0.18741	0.00000	0.27750	0.07759	0.55147
	4	0.00000	0.18007	-0.30401	0.15927	-0.23875	0.34779
	5	0.30629	0.19753	0.43526	0.25130	0.00000	0.28433
	6	0.30767	0.24797	0.00000	0.35706	0.72639	0.41938
	7	-0.17455	0.46513	0.00000	0.22133	-0.41468	0.34838
	8	0.10508	0.25526	-0.08784	0.20654	-0.13197	0.07619
	AMAE	2.21470		1.92358		2.32542	
100	1	0.63083	0.09314	0.32658	0.24080	0.40496	0.06216
	2	0.06508	0.06801	0.65626	0.43701	0.27488	0.14691
	3	0.33954	0.32289	-0.42135	0.17457	0.00000	0.20046
	4	-0.28179	0.19618	0.00000	0.05741	-0.07507	0.15017
	5	0.63250	0.04551	0.46984	0.03834	0.63770	0.21024
	6	0.00000	0.21132	-0.08016	0.28626	0.00000	0.00000
	7	-0.05523	0.05805	0.00000	0.05753	-0.49104	0.58376
	8	0.00000	0.14073	0.24077	0.10497	0.32713	0.32227
	AMAE	2.27630		2.42669		2.49066	

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Academia Open
Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.15042

Table 4: The criteria represents comparison (AMAE) between methods when ($n = 25$ and $\rho = 0.25, 0.50, 0.95$)

methods	$n = 25, \rho = 0.25$	$n = 25, \rho = 0.50$	$n = 25, \rho = 0.95$
	AMAF	AMAE	AMAE
$SCADL_1$	1.90770	1.87793	2.06273
$SCADL_2$	1.64146	1.73744	1.92358
MCP	1.82769	2.08680	2.32542
The Best methods	SCAD-L_2	SCAD-L_2	SCAD-L_2

Table 5: The criteria represents comparison (AMAE) between methods when ($n = 100$ and $\rho = 0.25, 0.50, 0.95$)

methods	$n = 100, \rho = 0.25$	$n = 100, \rho = 0.50$	$n = 100, \rho = 0.95$
	AMAE	AMAE	AMAE
$SCADL_1$	1.92738	1.95681	2.27630
$SCADL_2$	1.83737	1.93393	2.42669
MCP	1.84204	1.95870	2.49066
The Best methods	SCAD-L_2	SCAD-L_2	SCAD-L_1

We note through Tables No (1) ,(2),(3)When sample size n=25 And the number of variables P=8 and Correlation coefficient value (ρ) = 0.25 ,0.50 ,0.95

the results showed that method (SCAD-L2) This is the best Method to estimate vector parameters for model Because it gives less value for Average Mean Absolute Error (AMAE) then (MCP) method

When sample size n=100 And the number of variables P=8 and Correlation coefficient value (ρ) = 0.25

In general, the results showed that (SCAD-L2) method This is the best Method to estimate vector parameters for model Because it gives less value for Average Mean Absolute Error (AMAE) then SCAD-L1 method

When sample size n=25,100 And the number of variables P=8 and Correlation coefficient value (ρ) = 0.5 *and* 0.95

In general, the results showed that (SCAD-L2) method This is the best Method to estimate vector parameters for model Because it gives less value for Average Mean Absolute Error (AMAE) then SCAD-L1 method.

Overall, this(SCAD-L2) method proved superior to other estimation methods in all simulation experiments .

11-A real data Application

A random sample was taken from Al-Kut Teaching Hospital From 30 patients in 2024

This data will be analyzed using some penalized methods to based on the general linear regression model .

The study variables can be explained as follows:
Dependent variable (Yi): It represents the percentage of urea in the blood
The independent variables are represented by the following factors:
X1: Represents the level of the enzyme (GOT) in the blood for liver function
X2: Represents the measurement of the enzyme (GPT) in the blood for liver function.
X3: Represents the level of ferritin in the blood .
X4: Represents the percentage of calcium in the blood.
X5 : Represents the level of sugar in the blood.
X6: Represents the level of thyroid hormone in the blood (T3).
X7: Represents the level of thyroid hormone in the blood (T4).
X8: Represents the level of thyroid hormone in the blood (T5h).
A real data in the https://top4top.io/downloadf-3662wywk31-docx.html
Analyze and Explanation the results:
On dependent general linear regression model (GLRM) to predict the value of the dependent variable which represents the percentage of urea in the blood It is explained below:

$$Y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + \beta_8 X_{i8} + \epsilon_i$$

Table 6: Represents the estimated value and standard error of the parameter vector for Real data of all estimation methods when $n = 30, p = 8$

Variable	methods					
	SCAD-L_1		SCAD-L_2		MCP	
	Average($\hat{\beta}$)	S.E	Average($\hat{\beta}$)	S.E	Average($\hat{\beta}$)	S.E
x_1	0.77605	0.12040	0.77660	0.24080	0.06216	0.64315

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.15042

x_2	-0.16731	0.21850	-0.16681	0.43701	0.14691	0.00000
x_3	0.23020	0.08728	0.22994	0.17457	0.20046	0.24692
x_4	0.00000	0.02737	0.00000	0.05474	0.15017	0.16973
x_5	0.30738	0.01917	0.30759	0.03834	0.21029	0.32893
x_6	0.19407	0.14313	0.19343	0.28606	0.00000	0.37962
x_7	0.42960	0.02876	0.42904	0.05753	0.58376	0.42812
x_8	0.00571	0.05248	0.00000	0.10497	0.32227	0.24693
MAE	5.78662		5.78639		5.80662	

Table7: The criteria represent comparison (MAE) between The methods

Methods	MAE
SCAD-L_1	5.78662
SCAD-L_2	5.78639
MCP	5.80662
The best	SCAD-L_2

Through the results of Table No (6) and (7) of the real data analysis Shows that method SCAD-L2 It is the best compared to other methods in estimation general linear regression model because it gave less value for Mean Absolute Error

(MAE) = 5.78639

12- Conclusion

Through the results obtained by relying on some statistical methods in estimating and selecting parameters general linear regression model (GLRM) The results were shown in the experimental side and the applied side that method SCAD-L2 A clear advantage over the rest of the estimation methods based on the mean Absolute error Depending on the results of the best method SCAD-L2 The factors affecting elevated blood urea levels, represented as the dependent variable, were identified. It was found that all factors affect elevated blood urea levels except for the X4 is the percentage of calcium in the blood equal zero and X8 is the level of thyroid hormone in the blood (T5h) equal zero , which had no significant effect on the dependent variable (Y) .

References

1. Florentine Bunea et al., "Penalized least squares regression methods and applications to neuroimaging," NeuroImage, vol. 55, pp. 1519-1527, 2011.
2. J. L. Horowitz, "Semi parametric models," Department of Economics, Northwestern University, U.S.A., Working Paper No. 17, 2004.
3. J. Liu, R. Zhang, W. Zhao, and Y. Lu, "A robust and efficient estimation method for single index models," Journal of Multivariate Analysis, vol. 122, pp. 226-238, 2013.
4. K. Jiratchayat and C. Bumrungsup, "Penalized Linear Regression Methods where the Predictors Have Grouping Effect," Thailand Statistician, vol. 17, no. 2, pp. 212-222, 2019.
5. M. Lu et al., "Application of penalized linear regression methods to the selection of environmental naturopathy," Biometrika: Biomarker Research, vol. 5, p. 9, 2017.
6. O. Akkus, "Xplore Package for The Popular Parametric and The semi-parametric single index models," Gazi University Journal of Science, vol. 24, no. 4, pp. 753-762, 2011.
7. Q. Wang and W. Yao, "An adaptive estimation of MAVE," Journal of Multivariate Analysis, vol. 104, pp. 88-100, 2012.
8. St. Johns, "Variable selection in multivariate Multiple Regression," M.S. thesis, Dept. Math. Stat., Memorial University, 2015.
9. T. Wang, P. Xu, and L. Zhu, "Penalized minimum Average variance estimation," Statistica Sinica, vol. 23, pp. 543-569, 2013.
10. X. Liu, "Penalized variable selection for semi-parametric Regression models," Ph.D. dissertation, University of Rochester, New York, 2011.
11. X. Guo, T. Wang, and L. Zhu, "Model checking for parametric single index models: a dimension reduction model adaptive approach," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 78, no. 5, pp. 1013-1035, 2016.
12. Y. Xia, H. Tong, W. K. Li, and L. Zhu, "An adaptive estimation of dimension reduction space," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 64, no. 3, pp. 363-410, 2002.

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo

Vol. 11 No. 2 (2026): December
DOI: 10.21070/acopen.11.2026.15042

13. Y. Xia, W. Härdle, and O. Linton, "Optimal smoothing for a computationally and statistically Efficient single index Estimator," SFB 649 Discussion Papers, No. 2009-028, Humboldt University, Berlin, Germany, 2009.

14. Y. Xia and W. Härdle, "Semi-parametric estimation of partially linear single index models," Journal of Multivariate Analysis, vol. 97, pp. 1162-1184, 2006.

15. L. Zeng and J. Xie, "Group variable selection via SCAD-L2," Statistics: A Journal of Theoretical and Applied Statistics, vol. 48, no. 1, pp. 49-66, 2012.

ISSN 2714-7444 (online), https://acopen.umsida.ac.id, published by Universitas Muhammadiyah Sidoarjo