Journal of Al-Qadisiyah for Computer Science and Mathematics Vol.18.(1) 2026,pp. Static 18–28

Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)

Constructing Regression Model Using Penalty Methods to Process High-Dimensional Data for a Factorial Experiment

Mahmood M Taher

Department of Statistics and Informatics,College of Computer Science and Mathematics, University of Mosul, Iraq. Email: Mahmood81_tahr@uomosul.edu.iq

ARTICLE INFO

Article history:
Received: 06 /11/2025
Rrevised form: 24 /12/2025
Accepted : 30/12/2025
Available online: 30 /03/2026

Keywords:

Bridge, LASSO, ALASSO, SCAD,a factorial experiment.

ABSTRACT

Research addressed high-dimensional data in a 2^3 factorial experiment model using a completely randomized design. The columns of matrix (X) represented the effects of factor levels and their interactions, along with the overall arithmetic mean, while the rows represented the number of experimental observations. The high-dimensional data problem has been addressed by using several methods, comparing them according to various criteria. Penalty methods (Bridge, LASSO, ALASSO, SCAD) were employed to select significant factors in the model. Through simulation and application of statistical criteria, the performance of these methods was compared, with results Lasso and Adaptive Lasso show poor performance in most cases, with high MSE and MAD values. CAD generally falls in between, offering better performance compared to Lasso and Adaptive Lasso, but not as good as Bridge.

MSC..

https://doi.org/10.29304/jqcsm.2026.18.12480

1. Introduction

Recently, penalized regression methods have been shown to be versatile tools for data modeling, notably when high-dimensional data is being analyzed.[6].

Over the past few decades, with the advancement of data collection techniques, a large amount of high-dimensional data has continued to appear in various biological sciences, medical, social, and economic studies. See [3]. The term "high-dimensional data" refers to the situation where the number of explanatory variables exceeds the number of observations in the data matrix.

Statistical modeling is essential in many fields, as it clarifies the relationship between the response variable and a number of explanatory variables.

There have been new results related to ridge regression. A new approach developed by authors in [2] incorporates probability distributions before the beta parameter in the ridge regression model, developing beta ridge estimators that assist in diminishing multicollinearity. The second contribution, by authors [1], presented the two-parameter modified Leo estimator to calculate the likelihood function for models describing Poisson regression, obtaining promising results. Furthermore, authors in [19] utilized the adaptive procedure of the CoCoLasso algorithm to address the high-dimensional regression model with the consideration of error observations.

∗Corresponding author: *Mahmood M Taher*

Email addresses: Mahmood81_tahr@uomosul.edu.iq

Communicated by 'sub etitor'

When designing experiments, as demonstrated in [3,13] the problem of high dimensionality of data in the matrix (X) arises after converting the mathematical model of the design into a multiple regression model through the use of the general linear model by transforming the influence of factor levels and the influence of interactions between the levels of these factors into independent variables, so that the number of columns (The effect of factor levels and the effect of interactions between the levels of these factors) is greater than the number of rows (sample size), which means that (p>n) for the experiment Therefore, classical methods cannot be used to estimate parameters [7,20]. Consequently, the high dimensions of the mathematical model for design factorial experiments affect the estimation of parameters and the selection of important parameters. A new method was followed to transform the design model into a regression equation and use penalty methods to estimate and select the important parameters.

The research aims to address the problem of high-dimensional data for a factorial experiment design matrix (2^3) applied using a completely randomized design after converting this model into a regression equation and demonstrating the impact of the high-dimensional problem and its effect depending on the model and sample size.
Study Significance and Contributions, Explaining the importance of each level of factors and the importance of the interaction between levels of factors by processing the high-dimensional data of the factorial experiment (2^3) after converting the design model to a regression model.

1.1 study Problem

The problem of high-dimensional data arises in the independent variables matrix (X) after transforming the mathematical model of the experimental design into a multiple regression model using the general linear model. In this transformation, the effect of factor levels and the effect of interactions between them are transformed into independent variables. This leads to an increase in the number of columns (the number of influences and interactions) to a point where it exceeds the number of rows (the sample size), meaning that the condition (p > n) is met in the experiment.

Consequently, classical methods for estimating parameters cannot be applied because the information matrix cannot be inverted. This negatively impacts the efficiency and accuracy of parameter estimation, as well as the process of selecting important variables in the model. Therefore, the high-dimensional nature of the mathematical model in experimental design is a major challenge that requires alternative and more suitable estimation methods to address this problem.

1.2 Study Objectives

The main objective of this research is to address the problem of high-dimensional data in factorial experiments by using several methods, comparing them according to various criteria, and confirming the results by applying simulations with different data sizes.
Study significance and contributions:
1- Developing the statistical methodology for analyzing high-dimensional data in factorial experiments.
2- This study contributes to bridging the knowledge gap regarding how to deal with high-dimensional data in factorial experiments, which is relevant to many medical and industrial fields.

2. Factorial Experiments

Factorial experiments include the effect of each factor at the same time, and each factor has two or more levels. The researcher is interested in studying the main effects of each factor in addition to the effect of the interaction between the levels of the different factors. [13].

2.1 Three-Factor Interaction Model:

In factorial experiments in which each factor has only two levels, the number of treatments will be equal to 2n. If Factor A has two levels, Factor B has two levels, and Factor C has two levels. So, the number of processors used in the experiment (23 = 8). Which ($a_0b_0c_0 = 1 . a_0b_0c_1 = c . a_0b_1c_0 = b . a_0b_1c_1 = bc . a_1b_0c_0 = a . \ a_1b_0c_1 = ac . a_1b_1c_0 = ab . a_1b_1c_1 = abc$) mathematical model of the factorial experiment is: [13].

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl} \tag{1}$$

$i = 1.2 \ , \quad j = 1.2 \ , \quad k = 1.2$

3. Bridge

The Bridge method was proposed by researchers [8]. It follows the rule $L_q - norm$, and the penalty for linear regression using Bridge can be defined as follows:

$$PLR(\beta;\lambda)^{Bridge} = \{(y - X\beta)^T(y - X\beta) + \lambda\sum_{j=1}^{p}|\beta_j|^q\} \tag{2}$$

Bridge estimators can be obtained by reducing equation (2) as follows:

$$\hat{\beta}^{Bridge} = arg\ min_\beta\{(y - X\beta)^T(y - X\beta) + \lambda\sum_{j=1}^{p}|\beta_j|^q\} \tag{3}$$

Since λ is the positive control parameter, even if q>0. When q=1, the LASSO estimators are obtained, and when q=2, the Ridge estimators are obtained, as special cases of the Bridge penalty function.

When $q \leq 1$ the bridge function is a value, it selects variables and estimates parameters simultaneously, as well as when q<1 the bridge function is non-convex, and when q=1 the bridge function is convex but cannot be differentiated at zero, while when it is a value q>1, the bridge penalty function reduces the parameter values but does not make them. It is equal to zero, and the function is convex, so the Bridge method is suitable for cases in which the variable is chosen or which contain the problem of multicollinearity [14]. In addition, Bridge estimators have the Oracle property when the number of variables is less than the sample size, and also when the number of variables exceeds the sample size. Bridge estimators, under certain conditions, can correctly distinguish between zero variables and non-zero variables with a probability approaching one [9].

Two algorithms were presented to solve equation (3) and find the bridge estimators. The first was to use the Local Quadratic Approximation (LQA), which was proposed by [7]. The second method is Local Linear Approximation (LLA), which was proposed by [20]

3.1 Local Quadratic Approximation Algorithm (Lqa)

The LQA method is used to find a solution to equation (3) for any values q>0. The bridge function can be rewritten using the local quadratic approximation as follows:

$$|\beta_j|^q \approx |\beta_{0j}|^q + \frac{q|\beta_{0j}|^{q-1}}{2|\beta_{0j}|}(\beta_j^2 - \beta_{0j}^2) \tag{4}$$

Where $\beta_0 = (\beta_{01}, \beta_{02}, \beta_{03}, \ldots\ldots \beta_{0p})$ represents a vector of initial values.

Then the minimization equation (3) can be written as a quadratic minimization function:

$$\hat{\beta} = arg\ min_\beta \left\{(y - X\beta)^T(y - X\beta) + \frac{\lambda q}{2}\sum_{j=1}^{p}|\beta_{0j}|^{q-2}\beta_j^2\right\} \tag{5}$$

Then the Newton-Raphson algorithm is used, where the values of q and λ are determined, and the vector of initial values is found $\beta_0 = (\beta_{01}, \beta_{02}, \ldots., \beta_{0p})$. The Ridge coefficients were used as initial values by [14], and then calculated

$$\hat{\beta}^b = \{X^TX + \textstyle\sum_\lambda(\hat{\beta}^{(b-1)})\}^{-1}X^Ty \tag{6}$$

$$\text{Since,}\textstyle\sum_\lambda(\hat{\beta}^{(b-1)}) = diag\left(\frac{\lambda q}{2}|\hat{\beta}_1^{(b-1)}|^{q-2}, \frac{\lambda q}{2}|\hat{\beta}_2^{(b-1)}|^{q-2}, \ldots\ldots, \frac{\lambda q}{2}|\hat{\beta}_p^{(b-1)}|^{q-2}\right) \tag{7}$$

However, b=1, 2,… the algorithm stops when the parameter results are very close.

The LQA algorithm suffers from a flaw in the variable selection process. When a variable is deleted from the algorithm, it will also be excluded from the final model. When using the LQA algorithm, it is recommended that the initial estimate contain all variables, so [14] used Ridge estimates as initial estimates because they do not select variables.

When $q \geq 1$ Equation (3) is a convex function, it can be solved without resorting to approximation.

3.2 Local Linear Approximation Algorithm (LLA)

One-step estimators, proposed by [21], are used when $0 < q < 1$. The One-Step estimator for the bridge function is obtained, assuming that $y_i^* = \sqrt{2}x_i^T, x_{ij}^* = \frac{\sqrt{2}|\beta_{0i}|^{1-q}x_{ij}}{q}$

By applying the Least Angle Regression (LARS) algorithm to the observations (y_i^*, x_{ij}^*) to solve the following reduction function:

$$\hat{\beta}^* = argmin_\beta\{(y^* - X^*\beta)^T(y^* - X^*\beta) + \lambda\sum_{j=1}^{p}|\beta_j|\} \tag{8}$$

Then

$$\beta_{1j} = \frac{\hat{\beta}_j^*|\beta_{0j}|^{1-q}}{q} \tag{9}$$

LLA selects variables without an iterative process and also reduces the cost of calculations. However, Automatic variable selection becomes ineffective with highly correlated variables or globally significant variables, thus reducing the usefulness of this method.

There are two adjustment parameters λ, q, for both algorithms. The adjustment parameter λ works to balance the sum of squares term and the penalty term, while the adjustment parameter p works to determine the rank of the penalty function. The best match is found λ, q using grid search to obtain an accurate performance evaluation.

4. Least Absolute Shrinkage and Selection Operator (LASSO)

The most frequently employed penalty term is ($L1$-$norm$) which is named as least absolute shrinkage and selection operator, abbreviated as LASSO [17] proposed LASSO by imposing ($L1$-nor) to the residual sum of squares. LASSO gets its popularity and becomes a basis for other penalized methods. This is because of its ability to perform both continuous shrinkage and variable selection simultaneously. It can exactly set the regression coefficients to zero. [17].

The regression coefficients of the penalized linear regression using LASSO are obtained as:

$$PLR(\beta.\lambda)^{LASSO} = (y - X\beta)^T(y - X\beta) + \lambda\sum_{j=1}^{p}|\beta_j| \tag{10}$$

where the penalty term $\sum_{j=1}^{p}|\beta_j|$ represents the ($L1$-$norm$) and λ is a positive tuning parameter. As $\lambda \to 0$, the OLS estimates are obtained, and as λ increases sufficiently, more coefficient estimates are shrunk to be equal to zero.

The regression coefficients are obtained from the partial linear regression using Lasso as follows:

$$\hat{\beta}_{PLR}^{\text{LASSO}} = \text{argmin}_\beta \left\{\sum_{i=1}^{n}\left(y - X\hat{\beta}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right\} \tag{11}$$

Using different norms will lead to different consequences. To illustrate the geometrical difference between ($L1$-$norm$) and ($L2$-$norm$), Figure (1) shows the geometrical representation of these penalties by considering a simple case when there are only two explanatory variables in the model. The constrained region for ($L1$–$norm$).

It is a diamond, while the region becomes a circle in ($L2$-$norm$). The OLS estimator, $\hat{\beta}$, is shown in both panels, surrounded by the elliptical contours of the RSS. For the ($L1$-$norm$), there is a positive probability that the contour may touch the rotated square at its corners so that some coefficients can be estimated as exact zeros. The constrained area due to ($L2$-$norm$), on the other hand, is smooth, and hence cannot produce a sparse solution with a probability equal to zero. [12,17].

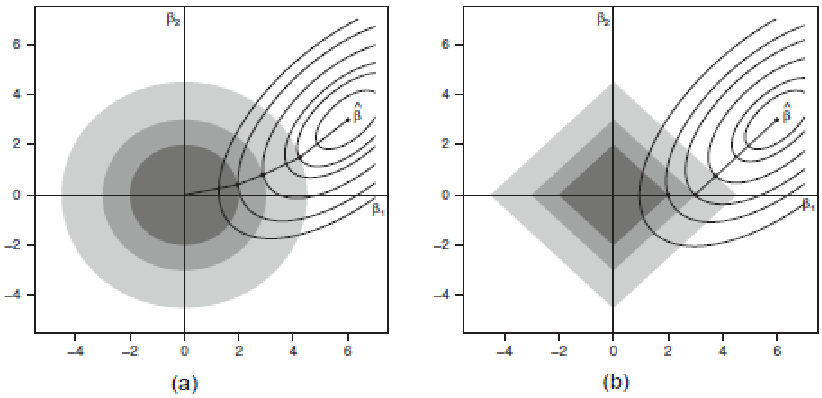

Fig 1: The Geometrical Representation for (a) ($L1-norm$) and (b) ($L2-norm$) in Non-Orthonormal Design.

5. Adaptive LASSO (ALASSO)

The adaptive LASSO (ALASSO) was proposed by Zou [20]. to alleviate the consistency problems of the LASSO, in which a data-driven weight is used with ($L1-norm$). Simply, ALASSO replaces the ($L1-norm$) by a weighted ($L1-norm$). The fundamental idea behind ALASSO is that, by allowing a relatively large amount of shrinkage for the zero coefficients and a small amount for the nonzero coefficients, it may be possible to reduce the bias of the estimation and improve the variable selection consistency.
The penalized linear regression using ALASSO is defined by

$$PLR(\beta.\lambda)^{ALASSO} = (y - X\beta)^T(y - X\beta) + \lambda \sum_{j=1}^{p} w_j |\beta_j| \quad (12)$$

Where $w = (w_j \;\;, \;\; j = 1,2, \ldots .p)^T \in \; R^p$ is a given data-driven weight vector, and it is obtained by:

$$\widehat{w}_0 = \left|\hat{\beta}_j^{initial}\right|^{-\gamma} \quad (13)$$

where (γ) is a positive constant and($\hat{\beta}_j^{\text{initial}}$) is an initial value, which is represented as an estimator for (β) obtained, for example, from OLS estimates or ridge regression estimates, are considered as initial values. For fixed value of (n) and larger values of (λ) in Eq. (5) lead to higher bias and less variability in predicted values, [11].

Further theoretical studies of ALASSO are performed when p > n, and several initial weights are proposed. [10]. studied the asymptotic properties of the ALASSO estimators, where the initial weight is chosen by the marginal regression estimator. The distribution of the ALASSO estimator was studied by [15]. In addition, the LASSO estimator itself can be used as an initial estimator [4]. Moreover, [18]. proposed to use the maximum marginal likelihood estimator as an initial weight in high-dimensional GLM.

The ALASSO estimator is defined as follows:

$$\hat{\beta}_{PLR}^{\text{ALASSO}} = \text{argmin}_{\beta} \left\{(y - X\beta)^T(y - X\beta) + \lambda \sum_{j=1}^{p} \widehat{w}_j |\beta_j|\right\} \quad (14)$$

Figure 2 displays the soft-threshold for both LASSO and ALASSO with respect to the OLS estimator. It can be seen from figure 2 that the ALASSO has no bias for the larger coefficients.

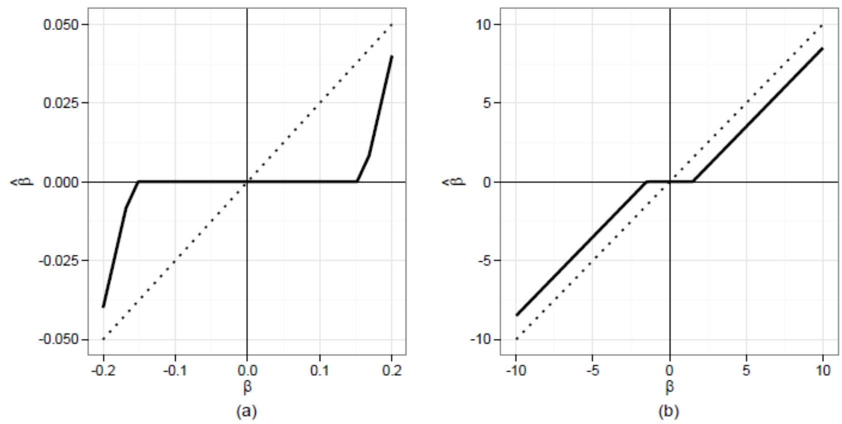

Fig 2: The Soft-Threshold Solution Under the Orthonormal Design for (a) OLS (dotted line) with ALASSO (b) with LASSO Penalized

6. SCAD (smoothly clipped absolute deviation)

The SCAD penalty function is one of the penalty methods used in estimating the parameters of linear regression models. It stands for Smoothly Clipped Absolute Deviation and was proposed by [7]. This method selects important variables and estimates regression parameters simultaneously. SCAD was proposed to address the problem of severe bias. For the large regression coefficients found in the LASSO method, this method is characterized by giving unbiased estimates of the large coefficients as well as the continuity of the selected models to ensure the stability of the model selection, in addition to the fact that SCAD works to solve the problem of consistency.
The penalty for linear regression using SCAD can be defined as follows:

$$PLR(\beta;\lambda) = \{(y - X\beta)^T(y - X\beta) + \sum_{j=1}^{p} P_\lambda|\beta_j|\} \tag{15}$$

The SCAD estimators for the penal linear regression coefficients are obtained as follows:

$$\hat{\beta}^{SCAD} = \text{argmin}_\beta \{(y - X\beta)^T(y - X\beta) + \sum_{j=1}^{p} P_\lambda|\beta_j|\} \tag{16}$$

It is $P_\lambda(.)$ the penalty limit (SCAD) and is defined as follows:

7. Regression Experiment Design

The multiple regression approach is used to show the main effects and interaction effects for two-level designs. Thus, the mathematical model for designing experiments is formulated in the form of general linear regression.

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \tag{17}$$

Where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ and $\beta = (\beta_{1 \ldots\ldots} \beta_p)$. Throughout this work, we center each input variable so that the observed mean is zero and the standard deviation is one. Mohammed et al.(2015).

7.1 Regression and the Factorial Design

Factorial experiments are characterized by the ability to measure the effect of the interaction between all factors in addition to the effect of the main factor. It is possible to use these factors and their effects with the interaction effect as a regression model after converting them to explanatory variables, assuming that we have a factorial experiment (23) with three factors (A, B, C) and two levels each factor. The mathematical model of the factorial experiment will contain the (26) effect in addition to the effect of the general arithmetic mean. [16].

1. The main effects of the levels of each factor and their number (6) effect for each level and here $(a_0, a_1, b_0, b_1, c_0, c_1)$
2. The effect of binary interaction between the levels and number of factors (12) according to each level of interaction between the two factors. Here
   AB $(a_0b_0, a_0b_1, a_1b_0, a_1b_1)$

AC $(a_0c_0 , a_0c_1 , a_1c_0 , a_1c_1)$
BC $(b_0c_0 , b_0c_1 , b_1c_0 , b_1c_1)$

3. The effect of the tripartite interaction of the levels and number of factors (8) according to the levels of the three factors, which is
$(a_0b_0c_0 , a_0b_0c_1 , a_0b_1c_0 , a_0b_1c_1 , a_1b_0c_0 , a_1b_0c_1, a_1b_1c_0, a_1b_1c_1)$

The total number of the main factors and the two and three-way interactions using the levels of each factor is the product of summing the number of factors and interactions for the previous cases is (6 + 12 + 8 = 26) which representing the explanatory variables that correspond to the factors.

Based on the above steps, a regression model for a factorial experiment (2^3) will be formed, and thus the regression equation will contain (26) an explanatory variable and (26) a parameter in addition to the parameter (β_0) as in the following equation: [16]

$$Y = \beta_0 + \sum_{i=1}^{3}\sum_{j=1}^{2}\beta_{ij}x_{ij} + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\beta_{ijkl}x_{ij}x_{ij} + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\sum_{m=1}^{3}\sum_{n=1}^{2}\beta_{ijklmn}x_{ij}x_{ij}x_{ij} + u_i \quad (18)$$

y_i : response variable ($i = 1.2. \dots . n$).

$(x_{11}, x_{11}x_{21}, \dots , x_{12}x_{22}x_{32})$: Independent variables.

β_0 : intersection parameter.

$(\beta_{11}, \dots , \beta_{122232})$ Regression parameters ($j = 1.2. \dots . 2^N$).

(N) : Number of factors.

u_i : Random error follows a normal distribution with mean (0) and variance (σ^2).

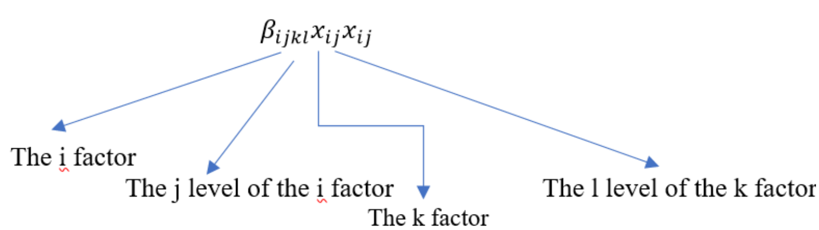

Fig 3: shows the method of encoding explanatory variables in addition to encoding parameters

Based on the equation (8), the matrix (X) will consist of (27) columns, where each column represents the effect of the levels of the main factors in addition to the effect of the interaction between all levels of the factors, and the number of rows is (16), which represents the sample size. Therefore, we note that (p>n) and thus the matrix suffers from high high-dimensional problem.

Using penalty (Bridge, LASSO, ALASSO, SCAD) methods, the high-dimensional problem will be addressed by estimating and selecting the important parameters in the experiment, where the (8) model will be employed in penalty methods as follows:

7.2 Bridge Factorial Design Regression Model

By adding the penalty (Bridge) from equation (2) to equation (18), we get:

$$Y = \beta_0 + \sum_{i=1}^{3}\sum_{j=1}^{2}\beta_{ij}x_{ij} + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\beta_{ijkl}x_{ij}x_{ij} + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\sum_{m=1}^{3}\sum_{n=1}^{2}\beta_{ijklmn}x_{ij}x_{ij}x_{ij} + u_i + \lambda\Big[|\beta_0|^q + \sum_{i=1}^{3}\sum_{j=1}^{2}\left|\beta_{ij}x_{ij}\right|^q + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\left|\beta_{ijkl}x_{ij}x_{ij}\right|^q + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\sum_{m=1}^{3}\sum_{n=1}^{2}\left|\beta_{ijklmn}x_{ij}x_{ij}x_{ij}\right|^q\Big] \quad (19)$$

- Formula using the (LQA) algorithm to find the estimators of the (Bridge) method by adding equation (5) to the equation of the proposed model (18), and thus we get equation (20) as follows:

$$Y = \beta_0 + \sum_{i=1}^{3}\sum_{j=1}^{2}\beta_{ij}x_{ij} + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\beta_{ijkl}x_{ij}x_{ij} + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\sum_{m=1}^{3}\sum_{n=1}^{2}\beta_{ijklmn}x_{ij}x_{ij}x_{ij} + u_i + \frac{\lambda q}{2}\Big[\left|\beta_{0j}\right|^{q-2}\beta_0^2 + \sum_{i=1}^{3}\sum_{j=1}^{2}\left|\beta_{0j}\right|^{q-2}\beta_{ij}^2 x_{ij} + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\left|\beta_{0j}\right|^{q-2}\beta_{ijkl}^2 x_{ij}x_{ij} + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\sum_{m=1}^{3}\sum_{n=1}^{2}\left|\beta_{0j}\right|^{q-2}\beta_{ijkl}^2 x_{ij}x_{ij}x_{ij}\Big] \quad (20)$$

- Formula using the (LLA) algorithm to find the estimators of the (Bridge) method by adding equation (8) to the equation of the proposed model (18), and thus we get equation (21) as follows:

$$y^* = \beta_0 + \sum_{i=1}^{3}\sum_{j=1}^{2}\beta_{ij}x_{ij}^* + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\beta_{ijkl}x_{ij}^*x_{ij}^* + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\sum_{m=1}^{3}\sum_{n=1}^{2}\beta_{ijklmn}x_{ij}^*x_{ij}^*x_{ij}^* + u_i + \lambda\Big[|\beta_0| + \sum_{i=1}^{3}\sum_{j=1}^{2}\left|\beta_{ij}x_{ij}^*\right| + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\left|\beta_{ijkl}x_{ij}^*x_{ij}^*\right| + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\sum_{m=1}^{3}\sum_{n=1}^{2}\left|\beta_{ijklmn}x_{ij}^*x_{ij}^*x_{ij}^*\right|\Big] \quad (21)$$

$$y_i^* = \sqrt{2}x_i^T\beta_{0i} \quad , x_{ij}^* = \frac{\sqrt{2}|\beta_{0i}|^{1-q}x_{ij}}{q}$$

7.3 Lasso Factorial Design Regression Model

When the LASSO penalty (Equation 10) is introduced into Equation (18), the resulting expression becomes:

$$Y = \beta_0 + \sum_{i=1}^{3}\sum_{j=1}^{2}\beta_{ij}x_{ij} + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\beta_{ijkl}x_{ij}x_{ij} + + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\sum_{m=1}^{3}\sum_{n=1}^{2}\beta_{ijklmn}x_{ij}x_{ij}x_{ij} + u_i + \lambda\Big[|\beta_0| + \sum_{i=1}^{3}\sum_{j=1}^{2}\left|\beta_{ij}x_{ij}\right| + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\left|\beta_{ijkl}x_{ij}x_{ij}\right| + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\sum_{m=1}^{3}\sum_{n=1}^{2}\left|\beta_{ijklmn}x_{ij}x_{ij}x_{ij}\right|\Big] \quad (22)$$

7.4 Adaptive Lasso Factorial Design Regression Model

When the ALASSO penalty (Equation 12) is introduced into Equation (18), the resulting expression becomes:

$$Y = \beta_0 + \sum_{i=1}^{3}\sum_{j=1}^{2}\beta_{ij}x_{ij} + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\beta_{ijkl}x_{ij}x_{ij} + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\sum_{m=1}^{3}\sum_{n=1}^{2}\beta_{ijklmn}x_{ij}x_{ij}x_{ij} + u_i + \lambda\begin{bmatrix} w_0|\beta_0| + \sum_{i=1}^{4}\sum_{j=1}^{2}w_{ij}\left|\beta_{ij}x_{ij}\right| + \sum_{i=1}^{4}\sum_{j=1}^{2}\sum_{k=1}^{4}\sum_{l=1}^{2}w_{ijkl}\left|\beta_{ijkl}x_{ij}x_{ij}\right| \\ + \sum_{i=1}^{4}\sum_{j=1}^{2}\sum_{k=1}^{4}\sum_{l=1}^{2}\sum_{m=1}^{4}\sum_{n=1}^{2}w_{ijklmn}\left|\beta_{ijklmn}x_{ij}x_{ij}x_{ij}\right| + \end{bmatrix} \quad (23)$$

Then the estimation method will be applied to each method to obtain the parameter values estimated by the (R) program, to obtain the results, and compare the methods.

7.5 SCAD Factorial Design Regression Model

When the SCAD penalty (Equation 15) is introduced into Equation (18), the resulting expression becomes:

$$Y = \beta_0 + \sum_{i=1}^{3}\sum_{j=1}^{2}\beta_{ij}x_{ij} + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\beta_{ijkl}x_{ij}x_{ij} + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\sum_{m=1}^{3}\sum_{n=1}^{2}\beta_{ijklmn}x_{ij}x_{ij}x_{ij} + u_i + \Big[p_\lambda|\beta_0| + \sum_{i=1}^{3}\sum_{j=1}^{2}p_\lambda\left|\beta_{ij}x_{ij}\right| + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}p_\lambda\left|\beta_{ijkl}x_{ij}x_{ij}\right| + \sum_{i=1}^{3}\sum_{j=1}^{2}\sum_{k=1}^{3}\sum_{l=1}^{2}\sum_{m=1}^{3}\sum_{n=1}^{2}p_\lambda\left|\beta_{ijklmn}x_{ij}x_{ij}x_{ij}\right|\Big] \quad (24)$$

8. Application

This study evaluates the performance of penalized regression methods (Bridge, Lasso, Adaptive Lasso, and SCAD) in high-dimensional settings with (23) factors and 26 explanatory variables. The simulation assesses parameter estimation and variable selection using criteria like (AdRsq, MSE, and MAD) across varying sample sizes (n = 5, 10, 15, 20, 25) with 1000 replications.

Simulation Steps:

1. Generate errors (ϵ) from N (0, σ^2).
2. Create a design matrix with 3 factors (coded ±1).
3. Assign fixed coefficients (β).

4. Compute response values (Y).
5. Estimate β using penalized methods.
6. Iterate over sample sizes (n).
7. Apply penalties (Bridge, Lasso, Adaptive Lasso, SCAD).
8. Evaluate performance via MSE, AdRsq, and MAD.

8.1 The Results

Table 1: shows the comparison criteria results from the estimations using the Bridge, Lasso, and adaptive Lasso, SCAD methods for all sample sizes

N	Methods	AdRsq	MSE	MAD
5	Bridge	0.6249	0.6095	0.7183
	Lasso	-0.5216	2.4729	1.3484
	ALasso	-0.5158	2.4634	1.3461
	SCAD	-0.5245	2.4776	1.3484
10	Bridge	0.9458	0.3208	0.4816
	Lasso	0.5249	2.8094	1.4907
	ALasso	0.5226	2.823	1.4952
	SCAD	0.6268	2.2072	1.1688
15	Bridge	0.8776	0.32083	0.5466
	Lasso	0.4285	2.8094	1.3801
	ALasso	0.4279	2.5719	1.3809
	SCAD	0.5175	1.8996	1.1544
20	Bridge	0.8660	0.9659	0.837
	Lasso	0.5064	3.5584	1.4335
	ALasso	0.5083	3.5443	1.4319
	SCAD	0.5217	3.4477	1.4090
25	Bridge	0.8772	1.2680	0.4578
	Lasso	0.6706	3.4019	1.4737
	ALasso	0.6700	3.4081	1.4727
	SCAD	0.7089	2.3871	1.0819

The table (1) shows the different metrics (Adjusted R-squared (AdjRsq), Mean Squared Error (MSE), and Mean Absolute Deviation (MAD) for the four methods across different sample sizes (N: 5, 10, 15, 20, 25), Here is the interpretation of the results for each sample size.

1. Sample Size 5:
   Bridge: Shows good results with Adj R-squared (0.6249), MSE (0.6095), and MAD (0.7183), providing balanced performance.
   Lasso: Shows a negative Adj R-squared (-0.5216), indicating poor model fit. The MSE (2.4729) and MAD (1.3484) are high, suggesting the predictions are far from the actual values.
   Adaptive Lasso: Performance is similar to Lasso, but with an unusually high MSE (204634), indicating a serious problem in predictions.
   SCAD: Similar to Lasso, with MSE (2.4776) and MAD (1.3484) almost identical to Lasso.
2. Sample Size 10:
   Bridge: Still the best with Adj R-squared (0.9458), showing that the model fits the data well, along with low MSE (0.3208) and MAD (0.4816).
   Lasso: Performs poorly with Adj R-squared (0.5249), indicating poor data explanation. MSE (2.8094) and MAD (1.4907) are high.
   Adaptive Lasso: Similar to Lasso, with MSE (2.823) and MAD (1.4952) being high as well.
   SCAD: Outperforms Lasso and Adaptive Lasso with Adj R-squared (0.6268), and lower MSE (2.2072) and MAD (1.1688).
3. Sample Size 15:

Bridge: Still the best with Adj R-squared (0.8776), MSE (0.32083), and MAD (0.5466).
Lasso: Shows poor performance with Adj R-squared (0.4285), MSE (2.8094), and MAD (1.3801).
Adaptive Lasso: Similar to Lasso, with MSE (2.5719) and MAD (1.3809) being slightly lower.
SCAD: Performs better than Lasso and Adaptive Lasso with Adj R-squared (0.5175), MSE (1.8996), and MAD (1.1544).

4. Sample Size 20:
   Bridge: Still excellent with Adj R-squared (0.8660), MSE (0.9659), and MAD (0.837).
   Lasso: MSE (3.5584) and MAD (1.4335) are quite high compared to Bridge.
   Adaptive Lasso: Similar to Lasso, with MSE (3.5443) and MAD (1.4319) being high.
   SCAD: MSE (3.4477) and MAD (1.4090) are high compared to Bridge, indicating that SCAD is less accurate.
5. Sample Size 25:
   Bridge: Still the best with Adj R-squared (0.8772), MSE (1.2680), and MAD (0.4578).
   Lasso: MSE (3.4019) and MAD (1.4737) remain high, indicating poor performance.
   Adaptive Lasso: Similar to Lasso, with MSE (3.4081) and MAD (1.4727).
   SCAD: Better than Lasso and Adaptive Lasso with MSE (2.3871) and MAD (1.0819).

The following graphs demonstrate the performance of the various methods.

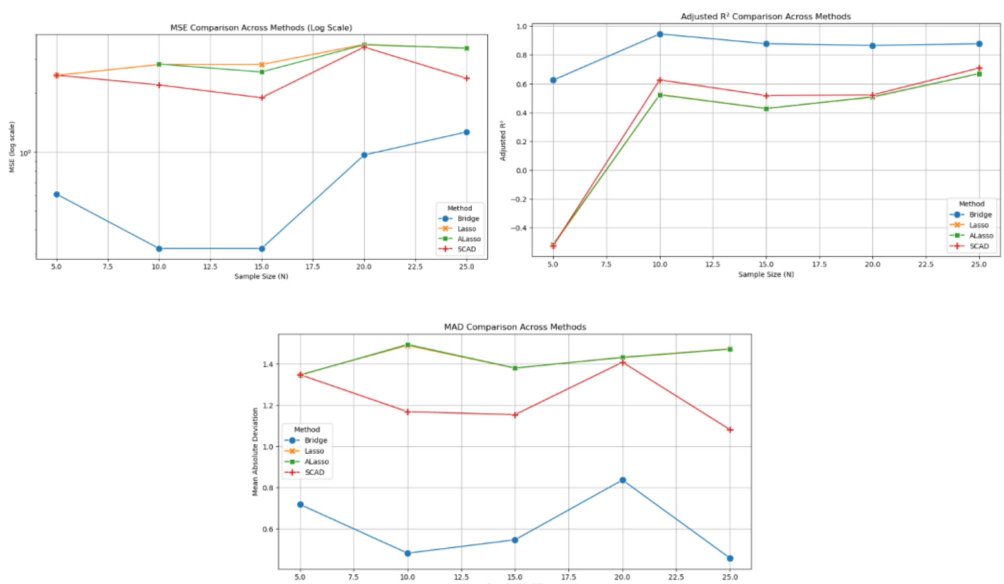

Fig 4: Comparison of estimation methods using

(MSE, Adjusted R^2, and MAD)

From Figure 4, we observe that the Bridge method shows a clear superiority in terms of accuracy (MSE), stability (MAD), and model quality (Adjusted R^2). As it has the lowest values for (MSE) and (MAD), and the highest value for (Adjusted R^2) compared to the other methods (Bridge, Lasso, adaptive Lasso, SCAD).

9. Conclusions

For small to moderate-sized datasets (n = 5, 10, and 15), the Bridge penalty generally performed better than the other methods of penalization in terms of having the highest adjusted R-square values and the lowest mean squared error and median absolute error. Conversely, LASSO, Adaptive LASSO, and SCAD generally had poor model-fit statistics with lower or negative adjusted R-square values. SCAD provided a small advantage over the other methods.

Increasing the sample size to n = 20 and 25, the result remained the same, and Bridge was the best method, giving better predictability and a lower prediction error. Though SCAD improved upon the other two methods, Lasso and Adaptive Lasso, it was still less accurate than Bridge. Lasso and Adaptive Lasso demonstrated a higher MSE and MAD, emphasizing the ineffectiveness of the methods for factorial regression analysis.

10. Recommendations

1-Use the Bridge method in high-dimensional factorial experimental designs.
2-Thesuggests expanding the study in the future to include more complex designs or non-linear models.

REFERENCES

[1] Abdelwahab, M. M., Abonazel, M. R., Hammad, A. T., and El-Masry, A. M. (2024). Modified two-parameter Liu estimator for addressing multicollinearity in the Poisson regression model. Axioms, 13(1), 46.
[2] Abonazel, M. R., and Taha, I. M. (2023). Beta ridge regression estimators: simulation and application. Communications in Statistics-Simulation and Computation, 52(9), 4280-4292.
[3] Box, G. E., Hunter, J. S., & Hunter, W. G. (2005). Statistics for experimenters. In Wiley series in probability and statistics. Hoboken, NJ: Wiley.
[4] Buhlmann, P. and Van De Geer, S. (2011). Statistics for High-Dimensional Data: Methods, Theory and Applications. NY: Springer.
[5] Efron, B., & Hastie, T. (2021). Computer age statistical inference, student edition: algorithms, evidence, and data science (Vol. 6). Cambridge University Press.
[6] ElSheikh, A., Abonazel, M. R., & Ali, M. C. (2025). A Review of Penalized Regression and Machine Learning Methods in High-Dimensional Data. The Egyptian Statistical Journal, 69(1), 250-261.
[7] Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association, 96(456), 1348-1360.
[8] Frank, L. E., & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools. Technometrics, 35(2), 109-135.
[9] Huang, J., Horowitz, J. L., & Ma, S. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models.
[10] Huang, J., Ma, S. and Zhang, C.-H. (2008). Adaptive Lasso for sparse high-dimensional regression models. Statistica Sinica. 18, 1603–1618.
[11] Hui, F. K. C., Warton, D. I. and Foster, S. D. (2015). Tuning parameter selection for the adaptive lasso using ERIC. Journal of the American Statistical Association. 110(509), 262–269.
[12] James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). Statistical learning. In An introduction to statistical learning: With applications in Python (pp. 15-67). Cham: Springer International Publishing.
[13] Montgomery, D.C. (2017), Design and Analysis of Experiments. 9th Edition. John Wily & Sons. Inc New York.
[14] Park, C., & Yoon, Y. J. (2011). Bridge regression: adaptivity and group selection. Journal of Statistical Planning and Inference, 141(11), 3506-3519.
[15] Potscher, B. M. and Schneider, U. (2009). On the distribution of the adaptive LASSO estimator. Journal of Statistical Planning and Inference. 139(8), 2775–2790.
[16] Saleem W.W. and Jaber A. G, (2022). A proposed Penalty Method for Processing High-Dimensional data in a Factorial Experiment (24). Al-Qadisiyah Journal for Administrative and Economic Sciences. Vol(24). No.3., PP .774 -790.
[17] Tibshirani, R. (1996). "Regression shrinkage and selection via the lasso". Journal of the Royal Statistical Society. Series B (Methodological). 58(1), 267–288.
[18] Wang, X. and Wang, M. (2015). Variable selection for high-dimensional generalized linear models with the weighted elastic-net procedure. Journal of Applied Statistics, In press.
[19] Yu, Q. (2025). Adaptive CoCoLasso for High-Dimensional Measurement Error Models. Entropy, 27(2), 97.
[20] Zou, H. (2006). “The Adaptive lasso and Its Oracle properties” Journal of the American statistical Association (JASA) 101 No.476., PP .1418 - 1429.
[21] Zou, H., & Li, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. Annals of statistics, 36(4), 1509.
[22] Abdelwahab, M. M., Abonazel, M. R., Hammad, A. T., and El-Masry, A. M. (2024). Modified two-parameter Liu estimator for addressing multicollinearity in the Poisson regression model. Axioms, 13(1), 46.