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)

Robust Variables Selection via the Reciprocal Lasso with an Exponential Squared Loss Function

Ali. J. Alkenani [a] **, *Naàel. H. Taklef*** [b,*]

[a] *Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Al Diwaniyah, Iraq.* ali.alkenani@qu.edu.iq

[b] *Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Al Diwaniyah, Iraq. nnaael405@gmail.com*

ARTICLE INFO

Article history:
Received: 00 /06/2025
Rrevised form: 25 /08/2025
Accepted : 27 /08/2025
Available online: 30 /09/2025

Keywords:

Exponential ,loss function, Lasso, Reciprocal Lasso

ABSTRACT

Variable selection procedures in regression models have gained increasing attention. They can be used to perform variable selection and are expected to yield accurate estimates. However, there are a number of assumptions in regression models and one of them related with assuming the normal distribution for the error term. The variables selection methods are not robust to outliers. One of the variables selection methods is the rLasso and it is not robust to outliers. A method has been proposed to make rLasso robust to variable selection by relying on an exponential squared loss function instead of the least squares loss function, which is sensitive to outliers. The effectiveness of the proposed method, which is robust to outliers in X and Y, has been demonstrated through simulations and real data.In this paper, we propose robust penalized regression estimator based on an exponential squared loss function, the motivation behind this proposal is that it enables us to introduce robust variable selection method. performance is outperforms the exsisting methods when the outliers are founded. Specifically, under different setting. We conducted simulation studies to compare our proposed method with some existing methods. Our simulation study shows that the proposed method performs well in terms of mean square error and average number of zero coefficients in the presence of significant outliers.

https://doi.org/10.29304/jqcsm.2025.17.32288

Introduction

Regression analysis is the most widely used tool in applied statistical studies, and is most prevalent in models containing large amounts of data and variables. Therefore, researchers search for explanatory variables that they believe have a significant effect on the dependent variable under study, there are several reasons that motivate researchers to search for important explanatory variables, including that some variables are not considered essential to the model, they contain measurement errors, and their effect may be similar to that of other variables. Therefore, researchers seek to narrow down the range of explanatory variables used in the final model, to arrive at the final model, several methods for variable selection have been proposed in the literature, including stepwise regression (Efroimson, 1960), the Akaike Information Measure AIC (Akaike, 1973) the BIC (Schwartz, 1978) Least absolute shrinkage and selection operator, Lasso (Tibshirani, 1996). Group Lasso (Yuan and Lin, 2006). Adaptive Lasso(Song and Liang, 2015). Smoothly trimmed absolute deviation (SCAD) (Fan and Li, 2001). The above mentioned variable selection methods select the variables of interest and estimate the parameters, they are not robust to outliers because they rely on the least squares loss function, which is sensitive to outliers, penalized quantile regression has gained increasing attention among researchers, and many different $\emptyset(\bullet)$ loss functions and

∗Corresponding author: Naàel. H. Taklef

Email addresses: *nnaael405@gmail.com*

Communicated by 'sub etitor'

penalty functions have been proposed and studied, for example, Wang et al. (2007) introduced LAD-Lasso, Johnson and Peng (2008) introduced rank-based variable selection. Wu and Liu (2009) introduced penalized quantile regression. Wang and Li. (2009) used a weighted Wilcoxon type SCAD method for robust variable selection. Ling (2010) introduced variable selection via regularized rank regression, in regression models, the choice of loss function determines the robustness of the resulting estimators. Wang et al.) 2013) propose an efficient penalty selection method based on an exponential squared loss function. Although the rLasso method solves some of the problems inherent in Lasso. It is also inefficient in handling outliers due to its reliance on a least squares loss function, this leads us to propose an efficient version of rLasso based on an exponential squared loss function, it is worth noting that the proposed method (ELS-rLasso) will be effective in dealing with outliers in X and Y, because the exponential squared loss function is effective against outliers in X and Y. It is not entirely surprising that our method performs well, as the exponential squared loss function with rLasso achieved similar success in the classification problem.

2-Variable selection with robust loss function:

Identifying explanatory variables with a real impact is one of the most significant challenges in statistical learning. Significant progress has been made in this regard through the use of penalized regression methods as a means of variable selection, these methods are based on a unified theoretical framework and offer high flexibility in terms of the possibility of applying multiple types of penalties, such as the bridge regression method introduced by (Frank and Friedman ,1993) and the Lasso technique developed by (Tibshirani ,1996) are widely used approaches in regression analysis, SCAD (Fan and Li, 2001) and regression using adaptive Lasso (Zou, 2006) it is worth noting that these methods are based on least squares, which is known for its high sensitivity to outliers, especially in small samples, which can lead to inaccurate results when selecting variables. For this reason, when data contain extreme values, it is preferable to replace least squares with a more resilient approach, in this context. Fan and Li (2001) presented pioneering work in which they proposed a general framework for robust penalized regressions, aimed at reducing the likelihood of outliers. Wang et al (2013) proposed a new estimator based on the exponential squared loss function with a Lasso penalty, they also developed a new regression framework that combines the exponential squared loss function with the Lasso penalty, aiming to enhance the model's immunity to outliers while maintaining efficient parameter selection, theoretical and experimental results demonstrate the superiority of this approach in environments characterized by conflicting or anomalous data, here is the following equation:

$$\Gamma_n\beta = \sum_{i=1}^{n} exp\{-(yi - x_i^T\beta)^2/\gamma_n)\} + n\sum_{j=1}^{d} p_{\lambda nj}(|\beta_j|) \dots\dots\dots\dots (2-1)$$

For the parameter β, we are dealing with a special case of Equation (2-1) when γ ∈ γ ∈ (0, + ∞). Since $\beta \in (n)$ represents a regularization parameter, it must be chosen adaptively depending on the characteristics of the data, for this reason, the objective function variable selection is a crucial step in building statistical models, especially for high dimensional problems, which are common challenges in many areas of scientific research and discovery today.

2-1-2-The proposed method(ELS-rLasso)

This section focuses on the proposed method. In addition to its robustness, the proposed method demonstrates nearoptimal performance and empirically outperforms compared with the existing methods when applied to real data. It should be noted, the use of the exponential squared loss function in the context of rLasso shows its good performance when applied to the experimental and real datasets and outperforming the non robust rLasso method and other methods. The exponential squared loss function is as follows

$$\emptyset_\gamma(a)=1\text{-}\exp(-a^2/\gamma),\dots\dots(2\text{-}2)$$

The proposed method relies on two key components to improve both robustness and variable selection.First, it uses the Exponential Squared Loss function .This loss function is designed to reduce the effect of large errors by assigning them lower weight. Compared to least squared loss, it makes the model less sensitive to outliers. The parameter γ controls how strongly large errors are down-weighted, which adds flexibility to the model's robustness.Second, the method incorporates a modified penalty called reciprocal Lasso (rLasso). Instead of using the classical Lasso penalty based on the absolute values of the coefficients, rLasso applies a penalty based on the

reciprocal of those values.This modification prevents overly shrinking large coefficients, which often represent important variables in the model. As a result, it maintains the benefits of variable selection while improving the accuracy of coefficient estimation especially when the data contains variables with different levels of influence. By combining the exponential squared loss function with the reciprocal penalty, the proposed method offers a more balanced approach that improves performance on real and complex datasets, outperforming the classical Lasso and non robust methods

Although the rLasso method is an improvement over the traditional Lasso in terms of reducing bias in estimating large parameters, it is not a statistically robust method. This limitation is due to the nature of the penalty function used, which is based on the reciprocal of the absolute values of the parameters, as any parameter approaches zero, the penalty imposed on it becomes very large, resulting in excessive sensitivity to small changes in the data,this behavior makes the model numerically unstable and weakens its ability to deal with outliers or high noise. Furthermore, the penalty function in this method is not always convex, making it difficult to guarantee stable solutions when using optimization algorithms. As a result, the rLasso method does not provide the sufficient degree of robustness required by models in environments characterized by irregularity or high noise. (Hahn et al., 2020)

The reciprocal Lasso method is a method for regularizing high dimensional linear models, it aims to reduce bias in the estimation of large coefficients by applying a penalty based on the reciprocal of the absolute values of the coefficients. However, this method itself lacks statistical robustness, as it is highly sensitive to outliers and noise in the data, this is due to the unstable nature of the penalty function as the coefficients approach zero, and its often reliance on a squares loss function which is sensitive. To overcome these limitations, the r Lasso method is combined with an exponential squared loss function, a robust loss function that reduces the impact of outliers, this combination enhances the model's ability to resist noise and data distortion, improving the reliability and generalizability of the estimates, by adopting an exponential squared loss function with an rLasso. The model benefits from this combining to obtain robustness and variable selection. Wang et al (2013) proposed a new estimator based on the exponential squares loss function with a Lasso penalty, they also developed a new regression framework that combines the exponential squared loss function with a Lasso penalty, aiming to enhance the model's immunity to outliers while maintaining efficient parameter selection. Theoretical and experimental results have demonstrated the superiority of this approach in environments characterized by conflicting or anomalous data, as indicated in Equation (2-1) we will work on developing the above equation and making it more efficient indoing robust variables selection, as shown in equation (2-3).

$$\Gamma_n(\beta) = \sum_{i=1}^{n} exp\{-(yi - x_i^T\beta)^2/\gamma_n)\} + \sum_{j=1}^{p} \frac{\lambda}{|\beta_j|} \dots\dots\dots..(2-3)$$

$\Gamma_n\beta$: Objective Function

N: Number of observations (number of rows or cases in the data).

yi: The true value of the response(dependent variable)for observation number i.

γ_n:tuning parameter.

P:Number of variables.

β :Vector of unknow parameters.

λ :regularization parameter.

This proposed methodology aims to construct a linear regression model that is highly immune to outliers while maintaining an efficient ability to select the important variables, this is achieved by combining an exponential squared loss function, which ensures that the impact of extreme data deviations is minimized, with a penalty function based on the inverse of the absolute value of the regression coefficients, enhancing the model's ability to reduce non-significant coefficients to as close to zero as possible, this method combines robust estimation with simplification accuracy, contributing to a more efficient and reliable model in highly dispersed data environments.

3-1- Update the tuning parameter γ_n

In the research presented by (Wang et al. ,2013) the tuning parameter γn of the exponential squared loss function was systematically updated to ensure improved model performance and increased resilience to outliers, the update begins with an initial value for γn, and the optimal value is then selected by trying a range of different values using cross validation or model selection criteria these methods help achieve a balance between model accuracy and noise tolerance, such that γn is adjusted to reduce the excessive influence of outliers without negatively impacting prediction accuracy, these steps ensure that the loss function is more resilient and stable during the estimation process, in the research presented by (Wang et al. ,2013) the following algorithm is implemented in the next two steps:

1- Find the pseudo outlier set of the sample:

Suppose we have $D_n = \{(x_i, y_i), \ldots\ldots (x_n, y_n)\}$ calculate $r_i(\hat{\beta}_n) = Y_i - X_i^t \widehat{\beta_n}$

$S_n = 1.4826 \times median_i \left| r_i(\hat{\beta}_n) - median_j\, r_j(\hat{\beta}_n)) \right|$ and i=1,……..n

Then: take the pseudo outlier set. $D_m = \left\{(x_i, y_i): \left| r_i(\hat{\beta}_n) \right| \geq 2.5 S_n\right\}$

Set $m = \left\{1 \leq i \leq n: \left|r_i(\hat{\beta}_n)\right| \geq 2.5\, S_n\right\}$ and $D_{n-m} = D_n / D_m$

2- Update the tuning parameter γ_n.

Let γ_n be the minimization of the matrix determinant. $\det\left(\hat{V}(\gamma)\right)$ in the set $G = \{\gamma: \delta(y) \in (0, 1]\}$ when $\delta(y)$is the determinant operator of the matrix.

$\hat{V}(\gamma) = \left\{\hat{I}_i(\hat{\beta}_n)\right\}^{-1} \widehat{\sum}_2 \left\{\hat{I}_i(\hat{\beta}_n)\right\}^{-1}$,and

$\hat{I}_i(\hat{\beta}_n) = \frac{2}{\gamma}\left\{\frac{1}{n}\sum_{i=1}^n \exp\left(-r_i^2(\hat{\beta}_n)/\gamma\right)\left(\frac{2r_i^2(\hat{\beta}_n)}{\gamma} - 1\right)\right\} \times \left(\frac{1}{n}\sum_{i=1}^n X_i X_i^T\right)$.....(3-1)

$\sum_2 = COV\left\{\exp\left(-r_i^2(\hat{\beta}_n)/\gamma\right)\frac{2r_i^2(\hat{\beta}_n)}{\gamma} xi, \ldots\ldots, \exp\left(-r_i^2(\hat{\beta}_n)/\gamma\right)\frac{2r_i^2(\hat{\beta}_n)}{\gamma} xn\right\}$....(3-2)

3-Update $\check{\beta}_n$:

With fixed λ_{nj}= log(n)/(n|$\check{\beta}_{nj}$ |), and the selected γ_n in Step 2, and update $\check{\beta}_n$by maximizing (3-9).In practice, (Wang et al. ,2013) set the MM-estimator $\check{\beta}_n$ as the initial estimate, that is, $\check{\beta}_n$ = $\tilde{\beta}_n$, then repeat Steps 1–3 until $\check{\beta}_n$and γ_n converges.Note that λ_{nj} is fixed within the iterations. One may update $\check{\beta}_n$ in Step 3 with λ_{nj} selected by cross-validation; however, his approach requires huge computation.

The adjustment in Wang's proposal will be performed in the second step of Equation (3-1) and chosen in a robust manner, in this thesis, we used the MCD method to compute the covairance matrix estimater instead of the classical covairance matrix estimaes which is employed (wang et al. 2013).It is well known that MCD giveus robust estimates the covairance matrix , The Minimum Covariance Determinant (MCD) is a robust method used to estimate the mean and the covariance matrix by selecting a subset of the data that minimizes the determinant of the covariance matrix. It automatically excludes outliers during the selection process, resulting in estimates that are less sensitive to atypical observations. This method is highly resistant to outliers, with a breakdown point of up to 50%, making it more reliable than traditional techniques(Rousseeuw and Driessen, 1999). We will work with Equation (3-1) and estimate the mean and variance using MCD. Let γ_n be the minimization of the matrix determinant. $\det\left(\hat{V}(\gamma)\right)$ In appointment $G = \{\gamma: \delta(y) \in (0, 1]\}$ When $\delta(y)$is the determinant operator of the,matrix

$\hat{V}(\gamma) = \left\{\hat{I}_i(\hat{\beta}_n)\right\}^{-1} \widehat{\sum}_{2MCD} \left\{\hat{I}_i(\hat{\beta}_n)\right\}^{-1}$

$\hat{I}_i(\hat{\beta}_n) = \left(\frac{2}{\gamma}\left\{\frac{1}{n}\sum_{i=1}^n \exp\left(-r_i^2(\hat{\beta}_n)/\gamma\right)\left(\frac{2r_i^2(\hat{\beta}_n)}{\gamma} - 1\right) \times \left(\frac{1}{n}\sum_{i=1}^n X_i X_i^T\right)\right\}\right)$.(3-3)

$\widehat{\sum}_{2\,MCD} = cov_{MCD}\left\{\exp\left(-r_i^2(\hat{\beta}_n)/\gamma\right)\frac{2r_i^2(\hat{\beta}_n)}{\gamma} xi, \ldots\ldots, \exp\left(-r_i^2(\hat{\beta}_n)/\gamma\right)\frac{2r_i^2(\hat{\beta}_n)}{\gamma} xn\right\}$....(3-4)

X_i: $observation\ vector\ i$

h: $Size\ of\ subset\ H$

H: $1,2,\dots,n$ $\quad a\ subset\ of\ observation\ of\ size\ h$

$X_i X_i^T$: $outer\ product\ of\ X_i$

$\check{\beta}_n$: $Estimated\ parameter\ vector$

$r_i^2(\beta_n)$: $squared\ residual\ for\ observation\ i$

exp(.): $Exponential\ function$

$\hat{I}_i(\beta_n)$: $Estimated\ Fisher\ information\ (or\ loss\ derivative)$

3-1-1- Choosing the tuning parameter λ.

In their study of variable selection using the exponential squared loss function, Wang et al.(2013) relied on a cross-validation technique to select the regularization parameter λ, which controls the severity of the penalty imposed on the model coefficients, this approach aims to achieve a balance between the quality of prediction and the number of variables included in the model, this is achieved by partitioning the data into multiple parts and evaluating the model's performance at different values of λ. The value that achieves the best predictive performance for data not used in training is chosen, this approach is suitable when using an unconventional and robust loss function such as the exponential squared loss, which makes relying on direct analytical methods for selection impossible or inaccurate. Thus, researchers were able to develop a model that efficiently selects important variables while minimizing the influence of outliers, enhancing the model's accuracy and reliability in environments characterized by atypical data (Ni et al., 2005). A set of model selection criteria was employed to determine the optimal value of λ, including the generalized Cross-Validation (GCV), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) (Schwartz, 1978), and the Risk Inflation Criterion (RIC) (Shi and Tsai, 2002). Unlike the study by Wang et al. (2013), which utilized BIC criterion, the current study proposed a robust version of AIC criterion vai the exponential squared loss function. The AIC criterion is as follows:

$$\text{AIC}= n\ log(RSS/n) + 2\ \text{p}(\lambda),\dots\dots \quad \dots\dots\dots\dots \quad .. \ (3\text{-}7)$$

$$\text{Where, RSS} = \sum_{i=1}^{n}(yi - X_i^t\ \beta)^2 \ \dots\dots\dots\dots\dots\dots(3\text{-}8)$$

Where, RSS refers to the residual sum of squares in the rLasso model, measuring the discrepancy between predicted and actual values., representing the predictors retained in the final model and p(λ) represents the number of non zero parameters in the model, this modification ensures that the selection criterion is robust to outliers.Since the AIC criterion is not robust to outliers because it relies on a non robust sum of squares (RSS) based estimates, it is possible to propose a robust version of it as follows.

$$\text{RAIC= n. log\{RRSS\}+2p(}\lambda\text{)}\dots\dots\dots\dots\dots\dots\dots\dots(3\text{-}9)$$

$$\text{Where, RRSS} = \sum_{i=1}^{n} exp\{-(Yi - X_i^t\ \beta)^2/\gamma_n\} \dots\dots\dots(3-10)$$

where γ_n represents the regularization parameter that controls the degree of robustness.

4-1-Simulation study and Real data

This section presents the experimental (simulation) aspect, aiming to evaluate the efficiency of the ELS-rLasso method in terms of prediction accuracy and variable selection, the ELS-rLasso method will be compared with the studied methods Lasso (Tibshirani, 1996) robust Lasso (Wang et al., 2013) and rLasso (Song and Liang,2015) the researcher wrote code for the proposed ELS-rLasso method to arrive at a solution path for the studied comparison methods, the proposed method and the comparison methods will be compared based on a variety of criteria, including the MSE criterion and the average number of zero coefficients (Ave, 0's) a simulation was conducted to evaluate the feasibility of the proposed ELS-rLasso method a typical simulation is detailed below Different data sizes (100, 200, 300, and 400) were used, and the simulation was repeated (1500) times, with the simulation being naturally generated.

4-1-1-Simulation 1

In this section, we conduct simulation studies to evaluate the performance of the proposed method on a limited sample. Data were simulated from the following true linear model:

		Standards / **differences**	MSE	Ave, 0's

$$Y = \mathrm{X}^{\mathrm{T}}\beta + 0.5\,\varepsilon,$$

R = 1500 datasets were generated with sample sizes of 100 ,200,300and 400 observations, each including 27 predictors. The coefficients vector was set as $\beta_j = (1, 1.5, 2, 1, 0 \underbrace{\ldots\ldots\ldots\ldots}_{23}, 0)^T$,

We choose n = 100,200,300 and 400, p = 27, and we generate

xi = (xi1, . . . , xid)^T from a multinormal distribution N(0,Ω_2),where the (i, j)th element of ,$\Omega_2$is $\rho^{|i-j|}$, ρ = 0.5 . The error term follows a cauchy distribution. We use the following three mecha nisms to generate influential points

Design1- Influential points in the predictors, covariate x_i follows a mixture of d-dimensional normal distributions 0.95N (0, Ω_1) + 0.05N (μ, Ω_2) ,=$I_{d\times d}$,

μ = 31_d , 1_d is d-dimensional vector of ones, and the error term follows a standard normal distribution.

Design 2- Influential points in the response, covariate x_i follows a multi normal distribution N (0, Ω_2), and the error term follows a mixture normal distribution 0.95N (0, Ω_1) + 0.05N (μ, Ω_2) .

Design 3- Influential points in both the predictors and response, covariate x_i follows a mixture of d-dimensional normal dis- tributions 0.95N (0, Ω_1) + 0.05N (μ, Ω_2) and the error term follows a Cauchy distribution.For each mechanism mentioned above , compare the performance

of four methods , the Lasso method introduced by (Tibshirani ,1996) with the rlasso method introduced by (Zeilieis ,2011) and our method [ESL (exponential squared loss)-rLasso].

Table (1) the results of(Ave, 0's) ,and MSE , when the sample size is 100,200,300 and400 for the compared methods.

		Lasso	0.887081	10
		rLasso	0.813013	15
		Rob-Lasso	0.742460	21
		ELS-rLasso	0.616461	23
	n=100	Lasso	0.848951	12
		rLasso	0.892514	16
		Rob-Lasso	0.750332	21
		ELS-rLasso	0.616224	23
	n=200	Lasso	0.537055	12
		rLasso	0.518849	21
		Rob-Lasso	0.492674	22
		ELS-rLasso	0.380460	23
	n=300	Lasso	0.471436	14
		rLasso	0.433999	20
		Rob-Lasso	0.309398	22
		ELS-rLasso	0.201136	23
	n=400			

Table(1) contains the performance comparison results of four methods (Lasso, rLasso, Rob-Lasso, and ELS-rLasso) (no contaminatd), with sample sizes of n = 100, 200, 300, and 400, the results are measured using MSE and Ave, 0's, indicating that ELS-rLasso performed best in accuracy (Ave, 0's = 23): rLasso, Lasso, and Rob-LASSO had relatively high MSE. ELS-rLasso achieves the best balance between accuracy and reducing the number of variables, especially when n is large (300, 400). ELS-rLasso achieves very accurate performance.

Table(2)the results first example of(Ave, 0's) and MSE , when the sample size is 100,200,300 and400 for the compared methods and the contamination is 0.05

		Lasso	0.573387	10
		rLasso	0.552293	15
Design 1	n=200	Rob-Lasso	0.389881	19
		ELS-rLasso	0.268058	23
	n=300	Lasso	0.426149	11
		rLasso	0.418683	16
		Rob-Lasso	0.225253	20
		ELS-rLasso	0.125119	23
	n=400	Lasso	0.332771	11
		rLasso	0.285669	16
		Rob-Lasso	0.189563	20
		ELS-rLasso	0.057659	23
Design		Lasso	0.360345	12

		Lasso	0.770986	11
		rLasso	0.661347	15
	n=100	Rob-Lasso	0.527066	15
		ELS-rLasso	0.500159	22
	n=200	Lasso	0.686982	11
		rLasso	0.580569	16
		Rob-Lasso	0.404764	18
		ELS-rLasso	0.347095	22

Standards differences MSE Ave_0's

Table (2) above shows the comparison of the four statistical models Lasso, rlasso - Rob-Lasso, ELS-rLasso in a specific design (Design1) with sample size n = 100, 200, 300 and 400 from the comparison results using MSE when the sample size increases we gradually notice that ELS-rLasso is the best model variant.

2		rLasso	0.301476	16
	n=300	Rob-Lasso	0.281523	19
		ELS-rLasso	0.200595	23
		Lasso	0.224857	12
		rLasso	0.218633	17
	n=400	Rob-Lasso	0.168759	20
	n=100	ELS-rLasso	0.110610	23

		Standards differences	MSE	Ave, 0's
		Lasso	0.973432	10
		rLasso	0.941347	12
		Rob-rLasso	0.725040	18
		ELS-rLasso	0.711139	21

Table (3) the results two example of(Ave, 0's)and MSE , when the sample size is 100,200,300 and400 for the compared methods and the contamination is 0.05

Table (3) above shows the comparison of the four statistical models Lasso, rlasso - Rob-Lasso, ELS-rLasso in a specific design (Design 2) with sample size n = 100, 200, 300 and 400 from the comparison results using MSE when the sample size increases we gradually notice that ELS-rLasso is the best model variant.

Table (4) the results three example of(Ave, 0's)and MSE , when the sample size is 100,200,300 and400 for the compared methods and the contamination is 0.05

Design 3	n=200	Lasso	0.878994	10
		rLasso	0.808367	13
		Rob-Lasso	0.608672	18
		ELS-rLasso	0.5631462	22
	n=300	Lasso	0.741722	12
		rLasso	0.632717	15
		Rob-Lasso	0.566277	18
		ELS-rLasso	0.330505	22
	n=400	Lasso	0.504547	12
		rLasso	0.445188	16
		Rob-Lasso	0.208555	19
		ELS-rLasso	0.193119	23

Table (4) above shows the comparison of the four statistical models Lasso, rlasso - Rob-Lasso, ELS-rLasso in a specific design (Design 3) with sample size n = 100, 200, 300 and 400 from the comparison results using MSE when the sample size increases we gradually notice that ELS-rLasso is the best model variant. When the sample size increases, we notice a clear improvement in the transmission of the proposed model, it is always the best, when we contaminated the error X, Y of the model, we gradually notice that ELS-rLasso is the best selection model.

4-2-Real data

Thalassemia is a hereditary blood disorder characterized by the abnormal synthesis of hemoglobin, inherited from one's progenitors. thalassemia is categorized into two main types alpha and beta, the severity of alpha and beta thalassemia depends on the quantity of absent genes among the four alpha globin genes and the two beta globin genes, thalassemia is prevalent in the mediter ranean region, africa, the middle east, the Indian subcontinent, and southeast asia in 2013, approximately 208 million individuals were affected by thalassemia, with 4.7 million individuals afflicted with severe disease, both males and females exhibit comparable disease rates, diagnosis is generally conducted through blood tests, which encompass a complete blood count, specialized hemoglobin assays, and genetic testing. A significant complication in patients receiving chronic transfusions is iron overload, three decades have elapsed since the inaugural hemopoi etic stem cell transplant (HSCT) for thalassemia, and this procedure is now a prevalent treatment for the definitive cure of thalassemia major, with over 3,000 HSCTs conducted globally , to confirm the effectiveness of our suggested approach, ELS-rLasso, we will examine actual data, which is represented by patient data from thalassemia. used to compare ELS-rLasso with the examined methods (Lasso, rLasso, , and ROB-Lasso), the study sample included 501 patients from Al-Samawah General Hospital Al-Samawah Governorate, and through an indepth study of patient records and after consulting with some specialized doctors and technicians in the laboratory, the details of the sample were determined, consisting of the dependent variable (y), which represents the survival duration of the patient (SDP), and a set of independent variables (x), numbering 11 variables, the details of which are as follows:

X1= Gender: Patient sex (0 for Female and 1 for Male)

X2= Age: Patient's age in years

X3= RBC: Red Blood Cell with Normal Ranges: (for Males 4.2 - 5.7, Females 3.8 - 5.0) 10^6/microliter.

X4= MCV: Mean Corpuscular Volume with Normal Ranges 82.5 - 98 fL for adults.

X5= MCH: Mean Corpuscular Hemoglobin is the average amount of hemoglobin in the average red cell with Normal Ranges (27 – 32) picograms for adults.

		Standards / differences	P.E	Selected variable
Design	n=501	Lasso	0.57852	X7, X11
		rLasso	0.354731	X3, X7, ,X11
		Rob-Lasso	0.233441	X3,X7 ,X9 ,X11
		ELS-rLasso	0.013960	X3,X4,X5, X9,X10,X11

X6 =HGB: Hemoglobin with Normal Ranges: (for Males 13.6 - 16.9, Females 11.9 - 14.8) Grams per deciliter.

X7= S: Subunits of HLPC Test with Normal Range = 0%.
X8= . HBA2: Subunits of HLPC Test with Normal Range (1-3) % for adults
X9=. HBA: Subunits of HLPC Test with Normal Range (> 97) % for adults
X10= HBF: Subunits of HLPC Test with Normal Range (< 1) % for adults
X11= Iron: Iron in Blood cells with Normal Range (60 – 170) micrograms per deciliter.

4-2-1- Prediction error

Prediction error (P.E.) represents the accuracy of a model's predictions. Ideally, P.E. should be near zero, indicating that the estimatesaer closely aligned with the true values, it measures how accurately an estimator assesses a parameter or how well a regression line fits a set of data points

Table (5) outputs of the real data results for people with a subject of school

From the above table, we notice the comparison of the four statistical models, Lasso, rLasso, ROB-Lasso, ELS-rLasso in a specific design (Design) the sample size n = 501 and the results of comparison using MSE, we note that MSE - ELS-rLasso is the least, meaning it is the best model.

From Table (5) it is clear that the ELS-rLasso method achieved lower performance compared to the other methods studied. We conclude from Figure (4.1) that the ELS-rLasso method has better performance in predicting actual results.

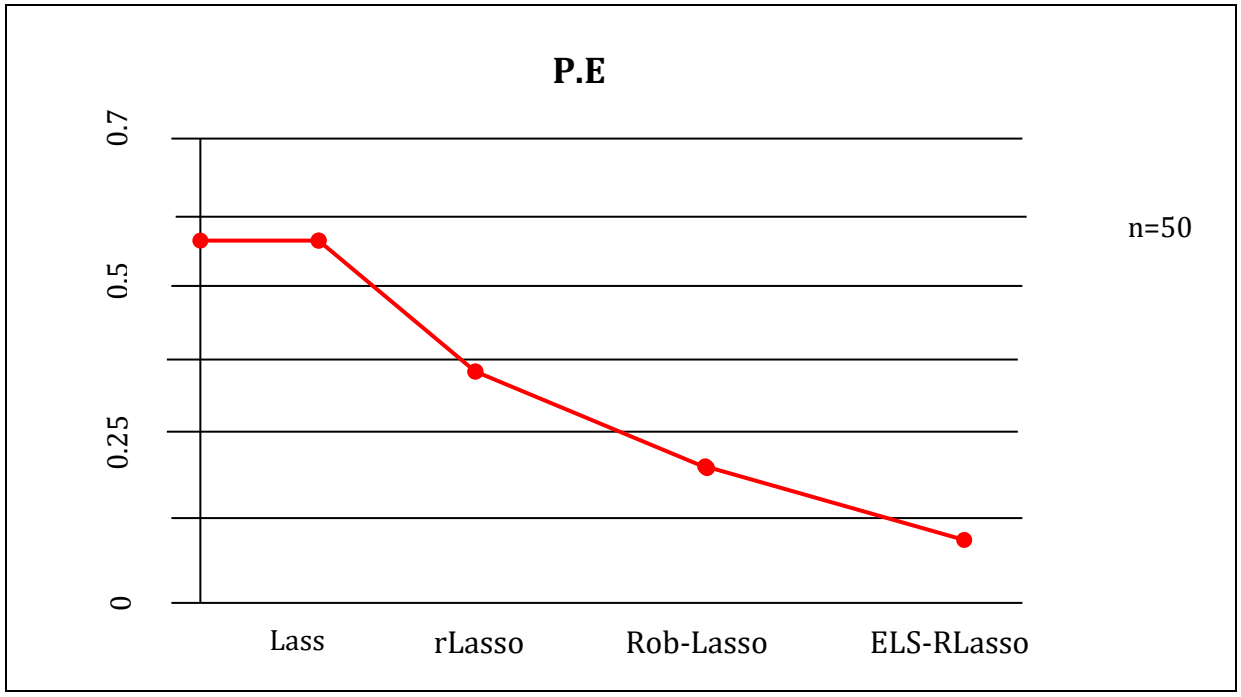


Figure (4.1).The prediction error (P.E) for various estimation method at sample size n=501 , showing that the ELS-rLasso mothod achieved lowest error.

5-Conclusion:

The proposed method (ELS-rLasso) demonstrated high efficiency in accurately selecting important variables, especially when there are outliers or noisy values in the data, compared to traditional methods such as Lasso or rLasso and Rob-Lasso and the robust method demonstrated superiority in terms of statistical model stability, maintaining prediction quality even in the presence of deviations from classical assumptions .Combining the rLasso with a exponential squared loss function enhanced the model's flexibility, allowing it to adapt to complex data characteristics without sacrificing interpretability and Simulation experiments showed that the proposed method gives the best positive outcome selection rate and increases the probability of detecting truly influential variables.

Application results on real data supported the effectiveness of the proposed approach, constructing a simpler and more accurate model without being severely affected by outliers.

References

1. Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csáki (Eds.), Proceedings of the 2nd International Symposium on Information Theory (pp. 267–281). Budapest: Akadémiai Kiadó.
2. Efroymson, M. A. (1960). Multiple regression analysis. In B. L. Magidson (Ed.), Mathematical Methods for Digital Computers (pp. 191–203). John Wiley & Sons.
3. Fan, J., and Li, R. (2001), “Variable Selection via Nonconcave Penalized Like-lihood and Its Oracle Properties,” Journal of the American Statistical Asso-ciation, 96, 1348–1360.

4. Frank, I. E., & Friedman, J. H. (1993). A statistical view of some chemometrics regression tools (with discussion). Technometrics, 35(2), 109–148.
5. Hahn, G., Lutz, S., Lange, C., & Munz, M. (2020). Inverse Lasso: A New Approach to Sparse Estimation with Nonconvex Penalty. Statistical Papers, 61(3), 995–1014. https://doi.org/10.1007/s00362-018-
6. Johnson, B. A., & Peng, L. (2008). Rank-based Variable Selection. Journal of Nonparametric Statistics, 20(3), 241–25
7. Li, J., Sun, F., & Zhang, W. (2023). Robust variable selection in spatial single-index varying-coefficient models using exponential squared loss. Entropy, 25(2), 230. https://doi.org/10.3390/e25020230
8. Mishra, P., & Liland, K. H. (2023). Iterative re weighted covariates selection for robust feature selection modelling in the presence of outliers (irCovSel). Journal of Chemometrics, 37(2), Article e3458. https://doi.org/10.1002/cem.3458
9. Robust variable selection using weighted Wilcoxon-type SCAD.Statistics and Computing, 19(3), 263–276. https://doi.org/10.1007/s11222-008-9089-7
10. Rousseeuw, P. J., & Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. Technometrics, 41(3), 212–223. https://doi.org/10.1080/00401706.1999.10485670
11. Shen, X., Pan, W., & Zhu, Y. (2014). Robust variable selection with nonconvex penalized M-estimators. Journal of the American Statistical Association, 109(505), 668–680.
12. Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2013). A sparse-group lasso. Journal of Computational and Graphical Statistics, 22(2),
13. Song, Q., & Liang, H. (2015). Variable selection in linear models with reciprocal Lasso. Journal of Statistical Planning and Inference, 166, 119–127. https://doi.org/10.1016/j.jspi.2015.04.001
14. Wang, H., Li, G., and Jiang, G. (2013). Robust Variable Selection with Exponential Squared Loss. Journal of the American Statistical Association.
15. Wang, X., Li, Y., and Leng, C. (2020). Adaptive Lasso for high-dimensional linear regression. Journal of Computational and Graphical Statistics, 29(1), 154–163. https://doi.org/10.1080/10618600.2019.1572217
16. Yohai, V., and Zamar, R. (1988), "High Breakdown-Point Estimates of Regres-sion by Means of the Minimization of an Efficient Scale," Journal of the American Statistical Association, 83, 406–413
17. Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(1), 49–67. https://doi.org/10.1111/j.1467-9868.2005.00532.x
18. Zou, H. (2006), "The Adaptive Lasso and Its Oracle Properties," Journal of the American Statistical Association, 101, 1418–1429.