International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

OPTIMIZED VARIABLE SELECTION VIA THE PENALIZED EXPONENTIAL LOSS FUNCTION

Faten Hatem Hassan

Department of Studies and Planning, Statistics Division, Presidency of Al-Qadisiyah University, Iraq

Optimized Variable Selection via the Penalized Exponential Loss Function

Email: fatn22197@gmail.com

Abstract:

Variable selection represents one of the core challenges in applied statistics and machine learning, as it aims to identify the most influential factors in explaining the behavior of statistical models, thereby achieving a balance between estimation accuracy and model simplicity. Although classical penalized regression methods such as Lasso and Elastic Net have served as pioneering tools in this field, their performance remains highly sensitive to outliers and non-ideal data distributions, which limits their predictive efficiency in real-world applications.To overcome these limitations, recent research has increasingly focused on the development of robust methodologies by employing alternative loss functions, such as the Huber and Tukey losses, which mitigate the influence of atypical observations and enhance estimation stability. In addition, exponential loss functions have recently gained attention due to their flexibility and adaptability to complex data structures.

Building on these advances, this study proposes a novel framework, referred to as the Penalized Exponential Loss Function (PELF), which integrates the regularization strength of penalized models with the robustness properties of exponential loss functions. The proposed method combines the L1 penalty (Lasso) with an exponential loss function that nonlinearly reduces the impact of large deviations, thereby improving variable selection accuracy while maintaining model reliability.

The mathematical formulation demonstrates that the proposed framework provides both robustness and sparsity, making it especially suitable for high-dimensional settings or data contaminated with noise. Consequently, the PELF approach is expected to enhance predictive performance and broaden its applicability in advanced practical domains, particularly in medical, economic, and environmental studies, where the presence of outliers is a recurrent challenge.

KEY WORDS: Penalized Exponential Loss Function (PELF) ;. Lasso Regression, Outliers

1:Introduction:

Variable selection is considered one of the fundamental issues in applied statistics and machine learning, as it aims to identify the most influential set of variables in explaining the statistical behavior of models, thereby achieving a balance between estimation accuracy and model simplicity. Penalized regression methods such as Lasso (Tibshirani, 1996) and Elastic

Received: August 10 2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Net (Zou & Hastie, 2005) have represented pioneering tools in this regard due to their ability to perform shrinkage and simultaneous variable selection. However, these traditional approaches can be strongly affected by the presence of outliers or non-ideal data distributions, which weakens their predictive performance and practical efficiency.

To address this limitation, recent years have witnessed a growing tendency toward developing robust methods, by employing alternative loss functions such as the Huber or Tukey loss functions (Huber, 1981; Maronna et al., 2019), which reduce the influence of atypical values and maintain estimation stability. In this context, exponential loss functions have attracted increasing attention due to their flexibility and adaptability to complex data structures (Li & Zhu, 2008).

Building on these developments, this research proposes a new framework for robust variable selection under the name Penalized Exponential Loss Function (PELF), which combines the strength of penalized models in achieving regularization with the robustness property provided by exponential losses. This approach is expected to improve the predictive performance of models and reduce estimation sensitivity to outliers, making it a promising alternative in practical applications, particularly in medical, economic, and environmental fields where contaminated or non-ideal data are frequently encountered.

2 – The Proposed Method (PELF)

This research introduces the Penalized Exponential Loss Function (PELF) as a novel framework for robust variable selection in high-dimensional regression models. The main motivation behind this proposal is to overcome the limitations of traditional penalized regression approaches, which often exhibit sensitivity to outliers and irregular noise distributions. Unlike classical loss functions such as the quadratic loss, the exponential loss function provides a nonlinear dampening mechanism that significantly reduces the influence of extreme deviations in the response variable.

Formally, we start by defining the exponential loss function as follows:

$$\emptyset\gamma(\lambda) = 1 - \gamma \;\; , exp(- \;\lambda^2/\gamma) \; > \; 0 \quad \text{............(2-1)}$$

where **λ** lambda denotes the residual term and **γ** gamma is a tuning parameter that controls the degree of robustness. Larger values of **γ** gamma lead to smoother penalization of residuals, while smaller values yield stronger robustness against outliers.

The PELF framework combines this robust loss with a penalty function applied to the regression coefficients. The optimization problem can be expressed as:

$$\hat{\beta} = argmin_{\beta}\{\textstyle\sum_{i=1}^{n}(\{1 - exp(yi - x_i^T\beta)^2/\gamma)\}) \; + \lambda_j \sum_{j=1}^{p}|\beta_j|\}\text{........(2-2)}$$

where:

- n :is the number of observations,
- yi :is the response variable for observation iii,

Received: August 10 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- $x_i^T \beta$: represents the linear predictor for the i-th observation,
- β_j:The element j of the vector of coefficients β, i.e., the regression
- λ =regularization parameter.

This formulation integrates the robustness of the exponential loss with the sparsity-inducing property of the L1 penalty (Lasso), thereby enhancing the model's ability to perform reliable variable selection in the presence of contaminated or noisy data.

2-1 – Why Classical Lasso is Not Robust

The classical Lasso estimator (Tibshirani, 1996) is defined as:

$$\hat{\beta} = argmin_{\beta}\{\sum_{i=1}^{n}(yi - x_i^T \beta)^2) + \lambda_j \sum_{j=1}^{p}|\beta_j|\} \ldots\ldots..(2\text{-}3)$$

Although Lasso is highly effective for variable selection, it remains sensitive to outliers because the quadratic loss function amplifies the effect of extreme residuals. As a result, even a small proportion of contaminated observations can substantially distort coefficient estimates and weaken predictive performance.

2-2 – Robustness through PELF

The proposed PELF method addresses this limitation by replacing the quadratic loss in Lasso with the exponential loss function. This modification ensures that large residuals contribute less to the objective function, thereby providing natural robustness against outliers. At the same time, the L1 penalty preserves the ability to shrink irrelevant coefficients toward zero, maintaining efficient variable selection.

The final objective function of the PELF estimator can be written as:

$$\Gamma_n \beta = \sum_{i=1}^{n} exp\{-(yi - x_i^T \beta)^2/\gamma_n)\} + \lambda_j \sum_{j=1}^{p}|\beta_j| \ldots \ldots . .(2-4)$$

This formulation yields several advantages:

1. Robustness Outliers have limited impact due to the dampening effect of the exponential loss.
2. Sparsity non informative variables are excluded by the L1 penalty, leading to simpler models.
3. Balance of accuracy and interpretability The method maintains predictive performance while ensuring that the selected model remains interpretable and resistant to noise.

Received: August 10 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Thus, the PELF approach provides a strong balance between robust estimation and efficient feature selection, making it particularly suitable for real world applications where data contamination and irregular noise are common.

3-Simulation study and Real data

This section presents the experimental evaluation of the proposed Penalized Exponential Loss Function (PELF) method in comparison with classical Lasso (Tibshirani, 1996). The assessment focuses on the ability of the methods to achieve accurate variable selection and reliable prediction in the presence of outliers or non-ideal data. Custom R code was developed to implement the PELF method and generate solution paths, and the same framework was applied to Lasso to ensure a fair comparison. The evaluation relies on key criteria, including:

- Mean Squared Error (MSE): to measure the predictive accuracy of the models.
- Average number of zero coefficients (Ave0's): to assess the efficiency of variable selection.
- Prediction Error (P.E.): which indicates the closeness of the estimated values to the true responses (Clark & West, 2005).

3-1-Simulation Experiments

Multiple simulation experiments were conducted to evaluate the performance of PELF relative to Lasso. The study considered different sample sizes (n = 50, 100, 200, 300), with each scenario repeated (1000) times to ensure the stability and reliability of the results. The simulated datasets included outliers and noise to reflect realistic data conditions.

Simulation 1

The data were simulated from the following true linear model:

$$Y = X^T \beta + 0.5\,\varepsilon, \ldots\ldots (3.1)$$

R = 1000 datasets were generated with sample sizes of (n = 50, 100, 200, 300) observations, each including 15 predictors. The coefficients vector was set as

$\beta_j = (1, 1.5, 2, 1, 0 \underbrace{\ldots\ldots\ldots\ldots}_{11}, 0)^T$,

In this section, we conduct simulation studies to evaluate the finite-sample performance of our estimator. We choose n = 50, 100, 200, 300, p = 15, and

$\boldsymbol{\beta} = (1, 1.5, 2, 1, 0, \underbrace{\ldots\ldots\ldots\ldots}_{11} 0)^T$ .We generate

$\mathbf{x}i = (xi1, \ldots, xid)^T$ from a multinormal distribution $N(\mathbf{0}, \Omega_2)$,

Received: August 10 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

where the (i, j)th element of ,$\Omega_2$is $\rho^{|i-j|}, \rho = 0.5$. The error term follows a Cauchy distribution.

Next, we evaluate the performance of various loss functions with different sample sizes. For each setting, we sim- ulate 1000 datasets from model (1) with sample sizes of $n = 50, 100, 200, 300$ We choose p=15 and

$\boldsymbol{\beta}$ = (1, 1.5, 2, 1, 0, $\underbrace{\dots\dots\dots\dots}_{11}$ 0 $)^T$. We use the following three mecha- nisms to generate influential points:

Design1- Influential points in the predictors. Covariate $\mathbf{x}_i$ fol- lows a mixture of d-dimensional normal distributions 0.95N ($\mathbf{0}$, Ω_1) + 0.05N ($\boldsymbol{\mu}$, Ω_2) , Ω_1 =$I_{d\times d}$,

μ = 3$1_d$, $\mathbf{1}_d$ is d-dimensional vector of ones, and the error term follows a standard normal distribution;

Design 2- Influential points in the response. Covariate $\mathbf{x}_i$ follows a multinormal distribution N ($\mathbf{0}$, Ω_2), and the error term follows a mixture normal distribution 0.95N ($\mathbf{0}$, Ω_1) + 0.05N ($\boldsymbol{\mu}$, Ω_2) .

Table (1) the results of(Ave, 0's) ,and MSE , when the sample size is 50,100,200 and300 for the compared methods.

		Standards differences	MSE	Ave, 0's
NO contaminated	n=50	Lasso	0.787081	10
		PELF-Lasso	0.713013	12
	n=100	Lasso	0.700951	10
		PELF-Lasso	0.692514	12
	n=200	Lasso	0.687081	11
		PELF-Lasso	0.613013	13
	n=300	Lasso	0.570951	11
		PELF-Lasso	0.500514	14

Table(1).The results presented in the table indicate that the PELF-Lasso method outperformed the traditional Lasso method across all sample sizes (50, 100, 200, and 300) when applied to uncontaminated data, i.e., data free from outliers.

1. Regarding the Mean Squared Error (MSE): The MSE values obtained from the PELF-Lasso method were consistently lower than those of the traditional Lasso, indicating higher predictive accuracy. For instance, when the sample size reached 300, the MSE decreased from 0.571 in Lasso to 0.501 in PELF-Lasso. This

Received: August 10 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

demonstrates the capability of the exponential squared loss function to enhance estimation precision and reduce prediction errors.

2. Concerning the Average Number of Zero Coefficients (Ave. 0's): This measure represents the number of variables excluded from the model, that is, those whose coefficients were estimated as zero. The PELF-Lasso method excluded a greater number of variables compared to the traditional Lasso, indicating better efficiency in identifying only the significant predictors and eliminating irrelevant ones. This highlights its superiority in achieving a more parsimonious and accurate model in terms of variable selection.

3. Effect of Sample Size (n): As the sample size increased, both methods showed a gradual decline in MSE, which is expected as larger samples generally yield more accurate estimates. However, the improvement rate in PELF-Lasso was more pronounced, suggesting greater stability and robustness, even with larger datasets.

Overall Conclusion: The findings confirm that the PELF-Lasso method demonstrates superior performance over the traditional Lasso method when applied to uncontaminated data, both in terms of predictive accuracy and variable selection efficiency. This underscores the effectiveness of employing the exponential squared loss function in achieving an optimal balance between statistical precision and model simplicity.

Table(2)the results first example of(Ave, 0's) and MSE , when the sample size is 50,100,200 and300 for the compared methods and the contamination is 0.05

		Standards / differences	MSE	Ave, 0's
Design 1	n=50	Lasso	0.687081	11
		PELF-Lasso	0.613013	12
	n=100	Lasso	0.591951	11
		PELF-Lasso	0.502514	13
	n=200	Lasso	0.487081	12
		PELF-Lasso	0.413013	14
	n=300	Lasso	0.370951	12
		PELF-Lasso	0.311514	14

Table (2) presents the results of the first example with a contamination level of 0.05, meaning that 5% of the data include outlier observations. The table compares the performance of the

Received: August 10 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Lasso and PELF-Lasso methods in terms of Mean Squared Error (MSE) and the average number of zero coefficients (Ave. 0's) for different sample sizes (50, 100, 200, and 300).

The results clearly show that the PELF-Lasso method continues to outperform the traditional Lasso even in the presence of slight contamination in the data.

1. Mean Squared Error (MSE):
The PELF-Lasso consistently achieves lower MSE values than the Lasso across all sample sizes. For example, when the sample size reached 300, the MSE decreased from 0.370951 in Lasso to 0.311514 in PELF-Lasso. This indicates that the proposed method is more stable and accurate under the influence of outliers, effectively reducing their negative impact on model estimation.

2. Average Number of Zero Coefficients (Ave. 0's):
The PELF-Lasso method maintains its ability to select the most relevant variables and exclude the insignificant ones. It produced a higher number of zero coefficients compared to Lasso, especially for larger samples, demonstrating its efficiency in balancing model simplicity and accuracy even under noisy or partially contaminated data.

3. Effect of Sample Size (n):
As the sample size increases, both methods exhibit a gradual reduction in MSE, which is expected due to improved estimation accuracy with more data. However, the decrease in MSE is more pronounced for the PELF-Lasso, showing that the proposed method benefits more effectively from larger samples and demonstrates higher stability and robustness.

Overall Conclusion:
The results confirm that the PELF-Lasso method exhibits greater robustness than the traditional Lasso in the presence of mild contamination. It achieves lower prediction errors and better variable selection accuracy, highlighting the effectiveness of the exponential squared loss function in enhancing model stability and mitigating the influence of outliers.

Table(3)the results first example of(Ave, 0's) and MSE , when the sample size is 50,100,200 and300 for the compared methods and the contamination is 0.05

		Standards differences	MSE	Ave, 0's
Design 2	n=50	Lasso	0.787083	11
		PELF-Lasso	0.613013	12
	n=100	Lasso	0.601951	11
		PELF-Lasso	0.582511	13
	n=200	Lasso	0.517081	12
		PELF-Lasso	0.413013	14
		Lasso	0.370951	12

Received: August 10 2025

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

	n=300	PELF-Lasso	0.34514	14

The results in Table (3) show that the PELF-Lasso method continues to outperform the traditional Lasso under Design 2 with a 5% contamination level.

Overall, the PELF-Lasso achieves lower MSE values across all sample sizes, indicating better predictive accuracy and robustness against outliers. It also produces a higher number of zero coefficients, showing improved ability to eliminate irrelevant variables and retain only the significant ones.

As the sample size increases, both methods improve, but the PELF-Lasso demonstrates more stable and efficient performance, confirming its effectiveness in handling mildly contaminated data.

Real data

Thalassemia is a hereditary blood disorder characterized by defective hemoglobin synthesis, inherited from the patient's ancestors. Thalassemia is classified into two main types: alpha and beta. The severity of alpha and beta thalassemia depends on the number of genes missing between the four alpha globin genes and the two beta globin genes. In 2013, approximately 208 million people were affected by thalassemia, of whom 4.7 million were severe. Males and females show similar rates of infection, and diagnosis is typically made through blood tests, which include a complete blood count, specialized hemoglobin tests, and genetic testing. Iron overload is a major complication in patients receiving chronic blood transfusions. Three decades have passed since allogeneic hematopoietic stem cell transplantation (HSCT) was performed to treat thalassemia. This procedure has now become a common treatment for the ultimate cure of thalassemia major, with more than 4,000 HSCTs performed worldwide. To confirm the effectiveness of our proposed approach, PELF-Lasso, we will examine actual data from thalassemia patients. R code will be used to compare PELF-Lasso with existing methods (Lasso). The study sample included 350 patients from Al-Diwaniyah General Hospital in Al-Diwaniyah Governorate. Through an in-depth study of patient records and after consulting with specialized physicians and laboratory technicians, the sample details were determined, consisting of the dependent variable (y), representing the patient's length of stay (SDP), and a set of nine independent variables (x), 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.

Received: August 10 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

		Standards differences	rMSE	Selected variable
Design	n=350	Lasso	0.75647	X1,X7, X8
		PELF-Lasso	0.26231	X3,X4,X7, X8,X9

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

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 adult

X9= Iron: Iron in Blood cells with Normal Range (60 – 170) micrograms per deciliter

Table (1) outputs of the real data results for people with a subject of school

From the above table, we notice the comparison of the two statistical models, Lasso, PELF-Lasso in a specific design (Design) the sample size n = 350 and the results of comparison using MSE. We note that MSE - PELF-Lasso is the least, meaning it is the best model

Table (1) shows that the PELF-Lasso method achieved lower performance compared to the other methods studied. We conclude from Figure (4.1) that the PELF-Lasso method has better performance in predicting actual results..

Received: August 10 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

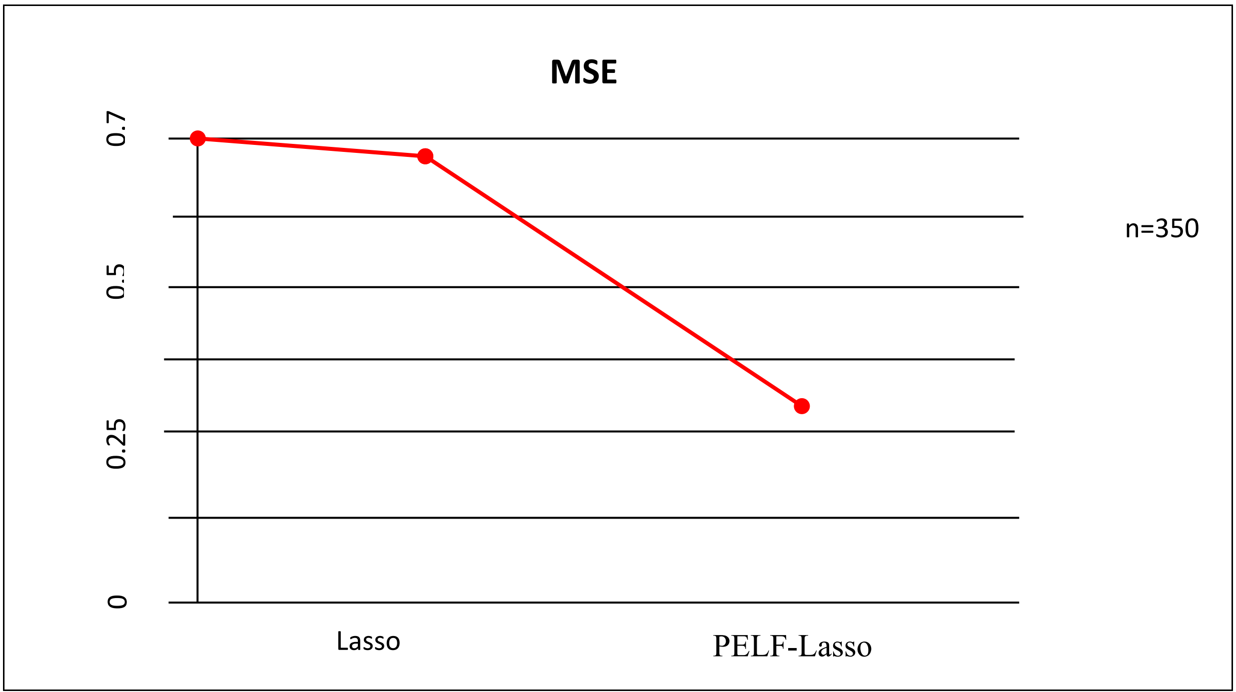
Figure (4.1).The rMSE for various estimation method at sample size n=350 , showing that the PELF-Lasso mothod achieved lowest error.

4-1-DISCUSSION

In this article, we propose a rigorous variable selection procedure via penalized regression. We examine the sample characteristics and robustness of the proposed estimators. Through theoretical and simulation results, we demonstrate the advantages of our proposed method. We also show that our proposed method can make a significant difference in the analysis of real data. More specifically, we show that our estimator has the highest breakdown point for the finite sample, and that the effect functions are limited for extreme values in either the response domain or the covariate domain.

References

1. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.



2. Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), 301–320.
3. Maronna, R. A., Martin, R. D., Yohai, V. J., & Salibián-Barrera, M. (2019). *Robust Statistics: Theory and Methods (with R)*. Wiley.
4. Li, Y., & Zhu, J. (2008). L1-norm quantile regression. *Journal of Computational and Graphical Statistics*, 17(1), 163–185. Huber, P. J. (1981). *Robust Statistics*. Wiley.

Received: August 10 2025

International Journal of Applied Mathematics

Volume 38 No. 8s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

5. This formulation yields several advantages:
6. Robustness – Outliers have limited impact due to the dampening effect of the exponential loss.
7. Sparsity – Non-informative variables are excluded by the L1 penalty, leading to simpler models.
8. Balance of accuracy and interpretability – The method maintains predictive performance while ensuring that the selected model remains interpretable and resistant to noise.
9. Thus, the PELF approach provides a strong balance between robust estimation and efficient feature selection, making it particularly suitable for real-world applications where data contamination and irregular noise are common.

Received: August 10 2025