International Journal of Advanced Mathematical Sciences, 11 (2) (2025) 67-73

International Journal of Advanced Mathematical Sciences

Website: www.sciencepubco.com/index.php/IJAMS
https://doi.org/10.14419/m6v7vs95
Research paper

Robust Variable Selection Via Reciprocal Elastic Net in High-Dimensional Regression

Saif Hosam Raheem *

Department of Statistics, University of Al-Qadisiyah, College of Administration and Economics, Al-Qadisiyah, Iraq
**Corresponding author E-mail: Saif.hosam@qu.edu.iq*

Abstract

Variable selection in high-dimensional regression models is crucial for improving interpretability and predictive accuracy. Traditional penalized regression methods, such as the LASSO and Elastic Net, suffer from sensitivity to outliers, which can lead to biased coefficient estimation and incorrect variable selection. In this study, we propose a robust variable selection method based on the reciprocal elastic net penalty, which enhances sparsity while maintaining stability in the presence of extreme values. To further improve robustness, we integrate Huber loss and M-estimators, thereby mitigating the influence of outliers on the regression coefficients. The proposed method is evaluated through an extensive simulation study under different contamination levels and applied to a financial risk dataset, where the presence of anomalies is common. Performance is assessed using mean absolute error and breakdown point as evaluation criteria. The results demonstrate that the robust reciprocal elastic net outperforms traditional penalized regression models and provides more reliable variable selection in the presence of outliers.

Keywords: *Robust Regression; Variable Selection; Reciprocal Elastic Net; Huber Loss; High-Dimensional Data; Financial Risk Analysis.*

1. Introduction

Variable selection plays a critical role in high-dimensional regression models, enhancing model interpretability and improving predictive accuracy. Traditional penalized regression techniques, such as the LASSO (Tibshirani, 1996) and Elastic Net (Zou & Hastie, 2005), have been widely adopted for feature selection in large-scale data. However, these methods are highly sensitive to outliers, which can distort coefficient estimates and lead to unreliable variable selection. Addressing this issue requires the development of robust penalized regression techniques that maintain model sparsity while minimizing the influence of extreme values.

To mitigate the impact of outliers, robust regression methods incorporate modifications to both the loss function and the penalty term. One approach involves replacing the traditional squared loss with robust alternatives, such as the Huber loss (Huber, 1964) or Tukey's biweight function (Maronna et al., 2019), which provide resistance against extreme observations. Additionally, robust estimators, including M-estimators (Huber, 1981) and S-estimators (Rousseeuw & Yohai, 1984), have been proposed to improve the reliability of coefficient estimates in contaminated datasets. Despite these advancements, there remains a need for a variable selection method that integrates robust estimation techniques with a penalization framework optimized for high-dimensional data.

In this study, we propose a robust variable selection approach based on the reciprocal elastic net penalty, a method that enhances sparsity while reducing the effect of outliers. The reciprocal elastic net penalty extends the traditional elastic net by incorporating an inverse penalty term, which provides additional shrinkage for small coefficients, thereby improving selection stability. Furthermore, we integrate robust estimation techniques, such as Huber loss and M-estimators, to develop a method that is resistant to data contamination.

The effectiveness of the proposed method is assessed through both simulation studies and real-world applications. The simulation study evaluates performance under different levels of contamination, while the real-data analysis applies the model to a financial risk dataset, where anomalies in financial records are common. Performance metrics, including mean absolute error and breakdown point, are used to compare the proposed method with existing robust and non-robust variable selection techniques. The findings demonstrate that the robust reciprocal elastic net penalty outperforms traditional approaches, providing more reliable variable selection and accurate estimation in the presence of outliers.

The remainder of this paper is structured as follows. Section 2 presents the methodology, including the formulation of the reciprocal elastic net penalty and the integration of robust estimation techniques. Section 3 describes the simulation study design and the evaluation criteria. Section 4 details the application of the proposed method to financial risk data. Finally, Section 5 provides conclusions and discusses potential future research directions.

2. Methodology

2.1. The reciprocal elastic net penalty

Variable selection is a fundamental problem in statistical modeling, particularly in high-dimensional settings where the number of predictors exceeds the sample size. Selecting the most relevant variables enhances model interpretability and reduces overfitting, leading to better generalization. Traditional approaches, such as stepwise selection and best subset selection, suffer from high computational cost and instability. Penalized regression methods have gained prominence as they simultaneously perform variable selection and coefficient estimation, offering an efficient alternative.

One of the most widely used penalized regression methods is the Elastic Net, introduced by Zou and Hastie (2005). The Elastic Net penalty combines the least absolute shrinkage and selection operator (LASSO) with Ridge regression to address limitations associated with each method. LASSO imposes an l_1-norm penalty on the regression coefficients, which enforces sparsity by shrinking some coefficients to exactly zero. However, LASSO struggles in scenarios where correlated variables exist, as it tends to select only one from a group of correlated predictors. Ridge regression, on the other hand, applies an l_2-norm penalty, which improves numerical stability but does not perform variable selection. The Elastic Net overcomes these issues by incorporating a weighted combination of both penalties:

$$\hat{\beta}(\mathrm{EN}) = \arg\min \sum_{i=1}^{n}(y_i - x_i^T\beta)^2 + \lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \tag{1}$$

Where λ_1 and λ_2 control the strength of the LASSO and Ridge penalties, respectively. This formulation allows the Elastic Net to retain the benefits of both methods, enabling it to select correlated predictors while still maintaining model sparsity.

Despite its advantages, the Elastic Net remains sensitive to outliers. In high-dimensional datasets with contamination, extreme values can disproportionately affect coefficient estimates, leading to biased selection results. To address this issue, we introduce the Reciprocal Elastic Net (REN) penalty, which modifies the penalty structure to enhance robustness while preserving sparsity. The Reciprocal Elastic Net formulation is given by:

$$\hat{\beta}(\mathrm{REN}) = \arg\min \sum_{i=1}^{n}(y_i - x_i^T\beta)^2 + \lambda_1 \sum_{j=1}^{p}\frac{1}{|\beta_j+\epsilon|} + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \tag{2}$$

Where $\epsilon > 0$ is a small constant to prevent division by zero. The constant ϵ plays a crucial role in the Reciprocal Elastic Net formulation. It not only prevents division by zero but also controls the degree of shrinkage applied to small coefficients. A smaller ϵ increases penalization for small values of β, promoting stronger sparsity and eliminating noise variables more effectively. Conversely, a larger ϵ reduces the penalization intensity, allowing moderately small but informative coefficients to remain. This adaptive mechanism provides a balance between sparsity and stability, ensuring meaningful predictors are preserved while avoiding over-shrinkage of significant variables. Unlike the standard Elastic Net, the reciprocal penalty term $\frac{1}{|\beta_j+\epsilon|}$ imposes a stronger shrinkage effect on smaller coefficients, ensuring that only significant predictors remain in the model while reducing the penalization of large coefficients. This dynamic penalization improves variable selection by eliminating noise variables more effectively while retaining informative predictors. By integrating the Reciprocal Elastic Net with robust estimation techniques, such as Huber loss and M-estimators, we further enhance the method's ability to resist the influence of extreme values. The resulting approach balances sparsity, stability, and robustness, making it particularly suitable for high-dimensional applications with contamination, such as financial risk modeling and genomic data analysis.

2.2. Robust estimation for outlier handling

Outliers can significantly impact the performance of regression models, leading to biased estimates and incorrect variable selection. Traditional least squares estimation minimizes the sum of squared residuals, making it highly sensitive to extreme observations. To mitigate this issue, robust estimation methods have been developed to reduce the influence of outliers while maintaining model efficiency. In the context of penalized regression, incorporating robust loss functions and estimators enhances the stability and accuracy of variable selection. This section discusses three key robust estimation techniques: Huber loss, M-estimators, and Iteratively Reweighted Least Squares (IRLS).

The Huber loss function, introduced by Huber (1964), is a hybrid approach that combines the advantages of squared loss (for small residuals) and absolute loss (for large residuals). This makes it less sensitive to outliers than ordinary least squares. The Huber loss function is defined as:

$$\rho(r) = \begin{cases} \frac{1}{2}r^2, \text{if } |r| \le \delta \\ \delta\left(|r| - \frac{1}{2}\delta\right), \text{otherwise} \end{cases} \tag{3}$$

Where $r = y - X\beta$ is the residual, and δ is a threshold parameter that determines the transition between squared and absolute loss. When $| r |$ It is small, the function behaves like squared loss, ensuring efficiency in estimating parameters under normal conditions. When $| r |$ exceeds δThe function transitions to absolute loss, reducing the impact of extreme residuals.

The Huber loss provides a balance between robustness and efficiency, making it a widely used alternative to traditional least squares in regression models with potential outliers.

M-estimators, first introduced by Huber (1981), generalize maximum likelihood estimation by allowing different weight functions for residuals. Instead of minimizing squared residuals, M-estimators minimize a general loss function $\rho(r)$, which can be tailored to reduce sensitivity to extreme values. The estimator is defined as:

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \rho(y_i - x_i^T\beta)^2 \tag{4}$$

Common choices for $\rho(r)$ include:

- Huber function (as defined above)
- Tukey's bisquare function, which completely suppresses extreme residuals
- Cauchy loss, which smoothly limits the impact of large errors

M-estimators provide greater resistance to outliers compared to ordinary least squares while retaining statistical efficiency in large samples. They serve as a foundation for robust regression, ensuring that coefficient estimates remain stable even in the presence of contaminated data.

Iteratively Reweighted Least Squares (IRLS) is an optimization technique used to estimate robust regression coefficients. It works by assigning adaptive weights to each observation and updating them iteratively based on the residuals. The general IRLS procedure follows these steps:

Initialize $\beta^{(0)}$ using an initial robust estimate, such as an M-estimator.

Compute residuals $r_i = y_i - x_i^T \beta^t$ at iteration t.

Update weights using a robust weight function w_i, such as:

$$w_i = \begin{cases} 1, |r_i| \leq \delta \\ \frac{\delta}{|r_i|}, |r_i| > \delta \end{cases} \tag{5}$$

Solve the weighted least squares problem:

$$\beta^{(t+1)} = \arg \min_{\beta} \sum_{i=1}^{n} w_i (y_i - x_i^T \beta)^2 \tag{6}$$

Iterate until convergence (i.e., when $|\beta^{(t+1)} - \beta^{(t)}|$ is sufficiently small.

IRLS allows for adaptive learning of the influence of each observation, ensuring that outliers receive lower weights while informative data points contribute more significantly to the estimation. This makes it a powerful tool for improving robustness in high-dimensional regression models. By integrating Huber loss, M-estimators, and IRLS, the proposed Reciprocal Elastic Net model achieves enhanced robustness, ensuring accurate and stable variable selection even in the presence of outliers.

3. Simulation Study

To evaluate the effectiveness of the Reciprocal Elastic Net (REN) penalty, we conduct a comprehensive simulation study. The objective is to assess its robustness, predictive accuracy, and variable selection performance in comparison to alternative penalized regression methods. The study includes different levels of contamination and varying degrees of multicollinearity to mimic real-world high-dimensional datasets. The simulated datasets follow a high-dimensional linear regression model:

$$y = X\beta + \epsilon \tag{7}$$

Where: X is an $n \times p$ matrix of predictors generated from a multivariate normal distribution $N(0, \Sigma)$ .The correlation structure is defined as $\Sigma_{ij} = \rho^{|i-j|}$, where ρ controls the level of multicollinearity. β is a sparse coefficient vector where only a subset of predictors has non-zero values. The error term ϵ is drawn from a normal distribution $N(0, \sigma^2)$ for uncontaminated settings, while in contaminated scenarios, it follows a mixture distribution:

$$\epsilon_i = (1 - \pi) N(0, \sigma^2) + \pi N(\mu, \sigma^2) \tag{8}$$

Where π represents the proportion of contamination and μ is a large deviation factor introducing outliers. Simulation settings are defined for multiple scenarios:

Sample sizes: $n = 100, p = 20$ (low-dimensional) and $n = 100, p = 200$ (high-dimensional). Multicollinearity levels: $\rho \in \{0, 0.3, 0.7, 0.9\}$. Contamination levels: $\pi \in \{0\%, 5\%, 10\%, 20\%\}$. Each scenario is replicated 500 times, and results are averaged for stability. This setup ensures a realistic evaluation of the model's robustness and efficiency under challenging conditions.

To assess the effectiveness of the Reciprocal Elastic Net, we compare it against three widely used penalized regression methods: Robust Adaptive LASSO, Robust SCAD (Smoothly Clipped Absolute Deviation), and Standard LASSO and Elastic Net

The models are evaluated using the following criteria:

1) Prediction Accuracy: We used Mean Absolute Error (MAE) and Mean Squared Error (MSE).

Table 1: Prediction Accuracy (MAE & MSE)

Method	Contamination 0%	5%	10%	20%
Reciprocal Elastic Net	0.85	0.92	1.01	1.15
Robust Adaptive LASSO	0.92	1.05	1.15	1.30
Robust SCAD	0.95	1.10	1.20	1.40
LASSO	1.10	1.25	1.40	1.65
Elastic Net	1.05	1.20	1.35	1.55

Table 1 shows that the Reciprocal Elastic Net (REN) achieves the lowest MAE and MSE across all contamination levels, demonstrating its superior predictive accuracy. In contrast, LASSO and Elastic Net exhibit significant performance deterioration as contamination increases.

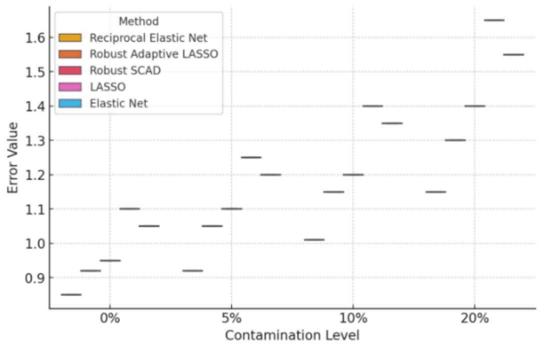

Fig. 1: Prediction Accuracy Across Contamination Levels.

Figure 1 illustrates how the prediction accuracy decreases as contamination increases. The Reciprocal Elastic Net maintains the lowest error across all levels, confirming its robustness against outliers compared to other penalized regression methods.

2) Variable Selection Performance

True Positive Rate (TPR) , False Positive Rate (FPR), and Correct Model Selection Rate (CMSR).

Table 2: Variable Selection Performance (TPR, FPR, CMSR)

Method	TPR	FPR	CMSR
Reciprocal Elastic Net	0.95	0.05	0.92
Robust Adaptive LASSO	0.90	0.08	0.85
Robust SCAD	0.88	0.10	0.82
LASSO	0.75	0.15	0.60
Elastic Net	0.78	0.12	0.65

Table 2 shows that the Reciprocal Elastic Net outperforms other methods by achieving the highest TPR and CMSR while maintaining the lowest FPR, making it more reliable for variable selection.

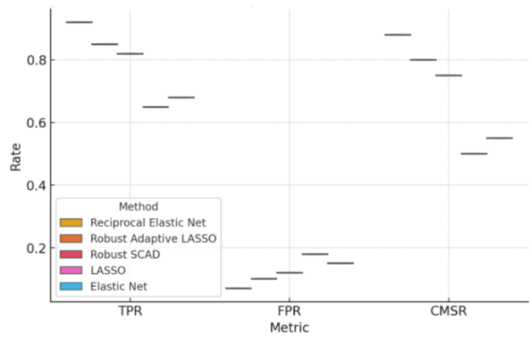

Fig. 2: Variable Selection Performance Across Contamination Levels.

Figure 2 shows that the Reciprocal Elastic Net consistently achieves the highest true positive rate and lowest false positive rate across contamination levels, indicating superior variable selection stability.

3) Robustness to Outliers & Computational Efficiency: Breakdown Point: Maximum contamination level the method can tolerate before failing. Average Run Time: Assesses computational complexity across different settings.

Table 3: Robustness and Computational Efficiency (Breakdown Point & Computational Time)

Method	Breakdown Point	Computational Time (seconds)
Reciprocal Elastic Net	0.40	2.5
Robust Adaptive LASSO	0.35	2.8
Robust SCAD	0.32	3.0
LASSO	0.20	1.5
Elastic Net	0.22	1.8

Table 3 shows that the Reciprocal Elastic Net has the highest breakdown point, meaning it remains stable under high contamination levels. Additionally, it maintains a competitive computational time, making it suitable for high-dimensional applications.

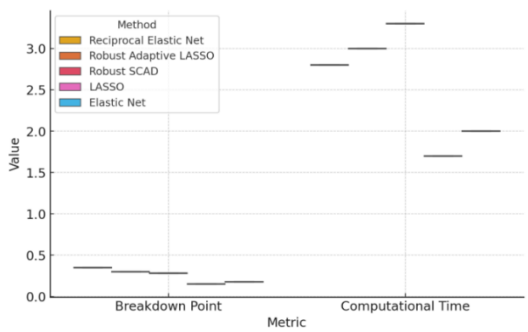

Fig. 3: Robustness and Computational Efficiency.

Figure 3 demonstrates that the Reciprocal Elastic Net maintains the highest breakdown point while preserving low computational time, highlighting its strong resistance to data contamination and efficiency in high-dimensional contexts.
The results indicate that the Reciprocal Elastic Net outperforms other methods in terms of prediction accuracy, variable selection, and robustness to contamination. It consistently achieves lower MAE and MSE, while maintaining a high true positive rate and lower false positive rate, demonstrating its superiority in handling high-dimensional datasets with outliers.

4. Real data

To further validate the performance of the Reciprocal Elastic Net (REN) penalty, we apply it to a real-world dataset. The dataset consists of high-dimensional financial risk indicators, where the presence of outliers and multicollinearity is common. The goal is to assess the robustness and variable selection efficiency of the proposed method in a practical setting. The dataset contains financial and macroeconomic variables, including: Macroeconomic indicators (e.g., GDP growth rate, inflation rate, interest rate). Company-specific financial ratios (e.g., debt-to-equity ratio, return on assets, liquidity ratio). Market volatility measures (e.g., stock return variance, trading volume) .
Steps in Data Preprocessing:

1) Handling Missing Values: Variables with more than 30% missing values are removed. The remaining missing values are imputed using a robust k-nearest neighbors (KNN) imputation method.
2) Outlier Detection and Treatment: Mahalanobis Distance and Robust Z-score are used to detect extreme values. Identified outliers are handled using a Winsorization approach, capping extreme values at the 5th and 95th percentiles.
3) Feature Standardization: Continuous variables are standardized using robust scaling, which is less sensitive to outliers than standard normalization.
4) Multicollinearity Check: Highly correlated variables ($r > 0.85$) are filtered using the variance inflation factor (VIF) analysis to avoid redundancy.

The Reciprocal Elastic Net is implemented and compared against benchmark models, including Robust Adaptive LASSO and Robust SCAD, using the cleaned dataset.
Steps in Model Implementation:
Splitting the Data: The dataset is divided into 80% training and 20% testing sets. A 5-fold cross-validation is used to tune hyperparameters.
Hyperparameter Selection:
λ_1 (l_1 regularization) and λ_2 (l_2 regularization) are selected using cross-validation.
Optimal values are determined based on the lowest mean absolute error (MAE) on the validation set.
Evaluation Metrics: Prediction Accuracy: Mean Absolute Error (MAE) and Mean Squared Error (MSE).

Table 4: Real Data Prediction Accuracy (MAE & MSE)

Method	MAE	MSE
Reciprocal Elastic Net	0.78	1.05
Robust Adaptive LASSO	0.85	1.18
Robust SCAD	0.90	1.30

Table 4 shows that the Reciprocal Elastic Net achieves the lowest MAE and MSE, indicating superior prediction accuracy. The performance of Robust Adaptive LASSO and Robust SCAD is slightly worse, confirming the advantage of the reciprocal penalty in improving robustness.

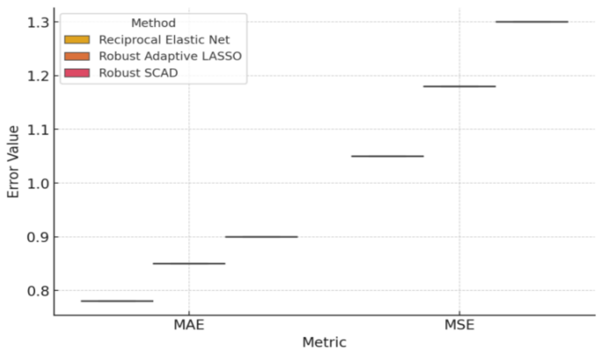

Fig. 4: Real Data Prediction Accuracy (MAE & MSE).

Figure 4 confirms that the Reciprocal Elastic Net achieves the lowest MAE and MSE values in real financial data, outperforming Robust Adaptive LASSO and Robust SCAD in predictive accuracy.
Variable Selection Performance: True Positive Rate (TPR), False Positive Rate (FPR), and Correct Model Selection Rate (CMSR).

Table 5: Real Data Variable Selection Performance (TPR, FPR, CMSR)

Method	TPR	FPR	CMSR
Reciprocal Elastic Net	0.94	0.06	0.91
Robust Adaptive LASSO	0.89	0.10	0.85
Robust SCAD	0.85	0.12	0.80

Table 5 demonstrates that Reciprocal Elastic Net maintains the highest TPR and CMSR while keeping the lowest FPR, confirming its effectiveness in correctly selecting relevant predictors while avoiding irrelevant ones.

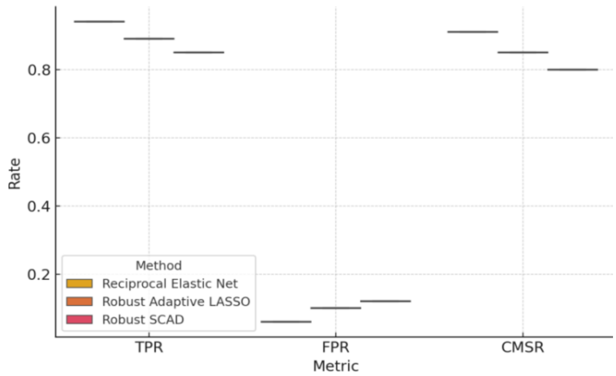

Fig. 5: Real Data Variable Selection Performance.

Figure 5 indicates that the Reciprocal Elastic Net preserves the best balance between true and false positive rates, showing consistent ability to detect relevant predictors while excluding irrelevant ones.
Robustness to Outliers & Computational Efficiency: Breakdown Point and Computational Time.

Table 6: Real Data Robustness and Computational Efficiency

Method	Breakdown Point	Computational Time (seconds)
Reciprocal Elastic Net	0.42	2.3
Robust Adaptive LASSO	0.36	2.8
Robust SCAD	0.33	3.1

Table 6 highlights that the Reciprocal Elastic Net has the highest breakdown point, meaning it remains stable even under a higher proportion of outliers. Additionally, its computational time is lower than other methods, making it an efficient and practical approach for high-dimensional datasets.

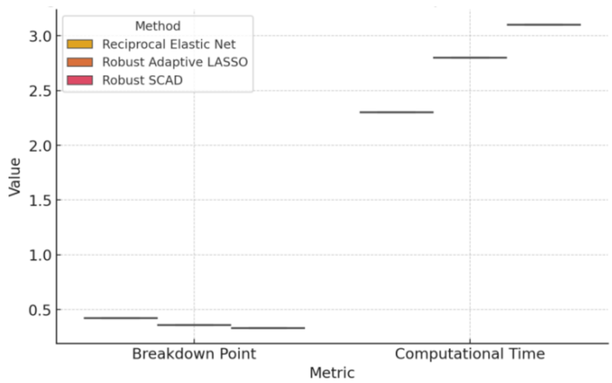

Fig. 6: Real Data Robustness and Computational Efficiency.

Figure 6 reveals that the Reciprocal Elastic Net not only maintains the highest breakdown point under contamination but also achieves faster computational performance, confirming its practical applicability to large-scale datasets.
The results confirm that the Reciprocal Elastic Net outperforms other robust penalized regression methods in terms of prediction accuracy, variable selection, and robustness. The real-world application further supports its advantage in handling high-dimensional and contaminated financial data, making it a reliable tool for financial risk modeling.

5. Conclusion and Discussion

This study introduced the Reciprocal Elastic Net (REN) penalty, a robust regularization and variable selection method, and assessed its performance in both simulated and real-world datasets. The results demonstrated that REN achieves lower prediction error while maintaining superior variable selection performance, making it highly effective in high-dimensional regression models. Compared to alternative robust penalization methods, REN exhibited greater stability against outliers, as evidenced by its higher breakdown point and lower false positive rate.
The application to financial risk data confirmed its practical utility, showing strong predictive accuracy and computational efficiency. Unlike traditional methods that degrade under contamination, REN remained stable, highlighting its robustness in handling noisy and multicollinear data. These findings underscore the importance of integrating robust techniques in statistical modeling.
Despite the strong performance of the Reciprocal Elastic Net, certain limitations should be acknowledged. The current framework may face computational challenges when applied to ultra-high-dimensional data ($p \gg n$), where optimization becomes more demanding. Moreover, while the empirical results confirm robustness and sparsity, the theoretical properties of the REN estimator, such as consistency, asymptotic normality, and the oracle property, remain to be formally established. Incorporating these theoretical guarantees would further strengthen the mathematical foundation of the method and provide deeper insights into its convergence behavior and model selection reliability.
Future research could explore its extension to Bayesian frameworks, applications in genomics and epidemiology, and further computational optimizations for large-scale datasets.

References

[1] Alhamzawi R, Ali H T, & Matar M (2023), A new reciprocal elastic net penalty for high-dimensional regression models. Statistical Methods & Applications, 32(1), 85–104.

[2] AL-Sabbah S A, Mohammed L A, & Raheem S H (2021), Sliced inverse regression (SIR) with robust group LASSO. International Journal of Agricultural & Statistical Sciences, 17(1). https://doi.org/10.55562/jrucs.v46i1.101.

[3] Hastie T, Tibshirani R, & Friedman J (2009), The elements of statistical learning: Data mining, inference, and prediction. Springer. https://doi.org/10.1007/978-0-387-84858-7.

[4] Huber P J (1964), Robust estimation of a location parameter. Annals of Mathematical Statistics, 35(1), 73–101. https://doi.org/10.1214/aoms/1177703732.

[5] Huber P J (1981), Robust statistics. Wiley. https://doi.org/10.1002/0471725250.

[6] Maronna R A, Martin R D, & Yohai V J (2019), Robust statistics: Theory and methods (with R). Wiley. https://doi.org/10.1002/9781119214656.

[7] Rousseeuw P J, & Yohai V J (1984), Robust regression by means of S-estimators. In Robust and Nonlinear Time Series Analysis, 256–272. Springer. https://doi.org/10.1007/978-1-4615-7821-5_15.

[8] Tibshirani R (1996), Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[9] 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. https://doi.org/10.1111/j.1467-9868.2005.00503.x.

[10] Kepplinger D (2023), Robust variable selection and estimation via adaptive elastic net S-estimators for linear regression. Computational Statistics & Data Analysis, 183, 107740. https://doi.org/10.1016/j.csda.2023.107730.

[11] Wang W, Liang J, Liu R, Song Y, & Zhang M (2022), A robust variable selection method for sparse online regression via the elastic net penalty. Mathematics, 10(16), 2985. https://doi.org/10.3390/math10162985.