Pak. J. Statist.
2026 Vol. 42(2), 173-184

PREDICTING CARDIOVASCULAR RISK USING EXPECTILE REGRESSION WITH THE RECIPROCAL ELASTIC NET PENALTY

Saif Hosam Raheem[1§], Maha Hadi Abed[1]
and **Zainab S. Alsaadi[2]**
[1] College of Administration and Economics,
University of Al-Qadisiyah, Iraq
Email: saif.hosam@qu.edu.iq
maha.h.abed@qu.edu.iq
[2] Department of Studies and Planning,
Presidency of the University,
Email: zainab.s.alsaadi@qu.edu.iq
[§] Corresponding author

ABSTRACT

Expectile regression is a flexible statistical framework widely utilized in domains such as econometrics, medicine, survival analysis, and environmental studies. It provides a comprehensive understanding of data by modeling the relationship between predictors and the conditional expectile of the response variable. However, one of the key challenges in expectile regression is the effective selection of relevant covariates, especially in high-dimensional datasets with multicollinearity.

In this study, we propose a novel methodology: Expectile Regression with the Reciprocal Elastic Net Penalty (ER-REN). This method integrates the sparsity-promoting properties of the Lasso penalty and the stability of the Ridge penalty, enhanced by a reciprocal formulation to prioritize relevant predictors dynamically. The ER-REN method is particularly suited for high-dimensional settings, providing robust variable selection and improved model interpretability.

We evaluate the performance of ER-REN through comprehensive simulation studies and a real-world application focusing on cardiovascular risk prediction. Comparative analysis demonstrates that ER-REN surpasses traditional penalization methods, such as ER-SCAD and ER-Adaptive Lasso, in terms of variable selection accuracy and predictive robustness. These results underline the potential of ER-REN as a powerful tool for regression modeling in complex and high-dimensional datasets.

KEY WORDS

Expectile Regression, Reciprocal Elastic Net, Variable Selection, Penalized Regression.

1. INTRODUCTION

Regression analysis has long been a cornerstone of statistical modeling, providing insights into the relationship between response variables and predictors. While traditional

© 2026 Pakistan Journal of Statistics

methods like ordinary least squares (OLS) regression focus on modeling the mean response, they often fall short in capturing the complexities of the conditional distribution of the outcome. Quantile regression (QR), introduced by Koenker and Bassett (1978), addressed these limitations by allowing the exploration of the conditional distribution at different quantile levels. QR's robustness to outliers and flexibility have made it a valuable tool in fields such as economics, medicine, and ecology. However, QR suffers from computational inefficiencies due to the non-differentiability of its check loss function, as well as limitations in its coherence as a risk measure.

To overcome these challenges, Newey and Powell (1987) introduced Expectile regression (ER)**,** which generalizes mean regression by using an asymmetric squared loss function. ER offers a smoother alternative to QR while maintaining computational tractability. Unlike QR, ER ensures a coherent relationship between the chosen Expectile levels and their corresponding risks. This makes ER a more suitable approach for applications requiring precise risk assessments, such as survival analysis and financial modeling.

An essential aspect of building regression models, including ER, is variable selection, especially when dealing with high-dimensional data. Variable selection enhances model interpretability, reduces overfitting, and ensures more accurate predictions. In recent years, penalized regression techniques have gained prominence for variable selection in ER, with methods like SCAD (Fan & Li, 2001), Adaptive Lasso (Zou, 2006), and Elastic Net being extensively studied. These methods impose penalties on regression coefficients, encouraging sparsity and addressing multicollinearity among predictors (Al-Guraibawi et al., 2025).

In this study, we propose a new approach: Expectile Regression with the Reciprocal Elastic Net Penalty (ER-REN). This method combines the strengths of the Lasso penalty (for sparsity) and the Ridge penalty (for stability), augmented by a reciprocal penalty term to prioritize influential predictors dynamically. ER-REN is designed to address the unique challenges posed by high-dimensional and collinear datasets, making it particularly effective for modern applications such as medical diagnostics and risk prediction.

This paper evaluates the proposed ER-REN method through simulation studies and a real-world application focusing on cardiovascular risk prediction. Comparative analysis against existing methods, such as ER-SCAD and ER-Adaptive Lasso, highlights the superiority of ER-REN in variable selection accuracy, robustness, and predictive performance. These contributions underscore the potential of ER-REN as a powerful tool for statistical modeling in complex data environments.

2. EXPECTILE REGRESSION (ER)

Expectile regression (ER) is an extension of mean regression that provides a comprehensive view of the conditional distribution of a response variable. Introduced by Newey and Powell (1987), ER models the relationship between predictors and expectiles, which are analogous to quantiles but are based on an asymmetric squared loss function. This unique characteristic allows ER to capture the asymmetry in data more effectively while retaining computational efficiency.

The fundamental idea of ER is to estimate the conditional Expectile of the response variable y given a set of predictors X. For a given Expectile level θ ($0 < \theta < 10$), the asymmetric squared loss function is defined as:

$$\mathrm{L}_{\theta}(\mathrm{y}, \hat{y}) = \begin{cases} \theta(y - \hat{y})^2 \text{ if } y \geq \hat{y}, \\ (1-\theta)(y - \hat{y})^2 \text{ if } y < \hat{y}, \end{cases} \quad (1)$$

where $\hat{y} = x^T\beta$ represents the predicted value. The weights θ and $1 - \theta$ adjust the penalty for over-prediction and under-prediction, respectively. This allows ER to capture different aspects of the response distribution, making it particularly useful for analyzing skewed or asymmetric data.

Advantages of Expectile Regression

1. **Smooth Optimization:** Unlike quantile regression, which uses a non-differentiable loss function, ER employs a differentiable asymmetric squared loss, enabling efficient computation through iterative weighted least squares.
2. **Coherent Risk Interpretation:** Expectiles maintain a coherent relationship with the conditional mean and variance, making them a more natural choice for certain risk assessments.
3. **Flexibility:** By varying θ, ER can model the entire distribution of the response variable, providing insights beyond the central tendency.

ER has found applications in diverse fields such as finance (e.g., value-at-risk estimation), medicine (e.g., survival analysis), and ecology (e.g., environmental risk modeling). Its ability to provide a nuanced understanding of the relationship between predictors and response makes it a valuable tool in modern statistical analysis.

This study builds on the strengths of ER by integrating it with advanced variable selection methods to address challenges in high-dimensional datasets, such as multicollinearity and sparsity, through the Reciprocal Elastic Net penalty.

3. ELASTIC NET PENALTY

The Elastic Net (EN) penalty, introduced by Zou and Hastie (2005), is a robust regularization method that combines the benefits of LASSO and Ridge regression, addressing their individual limitations. LASSO is well-known for its ability to perform variable selection by shrinking some regression coefficients exactly to zero, which makes it effective for sparse models. However, it struggles in situations where predictors are highly correlated or the number of predictors exceeds the number of observations ($p > n$). Ridge regression, on the other hand, is effective in handling multicollinearity by shrinking all coefficients, but it does not perform variable selection as it retains all predictors in the model (AL-Sabbah et al., 2021).

Elastic Net overcomes these challenges by introducing a weighted combination of the LASSO and Ridge penalties. The Elastic Net objective function is expressed as:

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

Here, the first term represents the residual sum of squares, which measures the model's goodness of fit. The second term ($\lambda_1 \sum_{j=1}^{p} |\beta_j|$) is the LASSO penalty, promoting sparsity by shrinking some coefficients to exactly zero. The third term ($\lambda_2 \sum_{j=1}^{p} \beta_j^2$) is the Ridge penalty, which ensures stability by controlling the magnitude of the coefficients. The parameters λ_1 and λ_2 control the relative contributions of the two penalties and are typically chosen via cross-validation.

One of the key strengths of Elastic Net is its ability to handle multicollinearity. When predictors are highly correlated, LASSO tends to arbitrarily select one predictor and discard the others, which can lead to instability in model selection. Elastic Net, however, can select groups of correlated predictors together, providing more stable and interpretable results. This group selection property is particularly useful in applications like genomics, where groups of genes may be highly correlated but jointly contribute to the response variable.

Despite its advantages, Elastic Net does not satisfy the oracle property, which means it may not consistently select the correct subset of predictors as the sample size increases. This limitation has led to the development of enhanced variants, such as the Adaptive Elastic Net (AdEN), which introduces adaptive weights to improve variable selection consistency. Elastic Net is a flexible and powerful regularization technique that bridges the gap between LASSO and Ridge regression. Its ability to balance sparsity and stability makes it a cornerstone method in regression analysis, especially in high-dimensional settings where predictor variables are numerous and often correlated.

4. RECIPROCAL ELASTIC NET

The Reciprocal Elastic Net (REN) penalty, introduced by Alhamzawi et al. (2023), is a novel regularization technique designed to address the limitations of traditional penalization methods in regression analysis. REN builds upon the Elastic Net framework by incorporating a reciprocal penalty term that dynamically adjusts penalization based on the magnitude of regression coefficients. This innovative approach enhances variable selection by prioritizing predictors with substantial contributions while penalizing smaller coefficients more aggressively. As a result, REN is particularly effective in high-dimensional datasets characterized by multicollinearity and sparsity (Bühlmann & van de Geer, 2011).

Mathematically, the REN penalty is defined by the following optimization problem:

$$\hat{\beta}(REN) = arg\ min \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \lambda_1 \sum_{j=1}^{p} \frac{1}{|\beta_j + \epsilon|} + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \qquad (3)$$

where $\hat{\beta}(REN)$ represents the estimated regression coefficients, y_i and y_i denote the response variable and predictor variable respectively, λ_1 and λ_2 are regularization parameters that control the trade-off between sparsity and stability. The term $\frac{1}{|\beta_j+\epsilon|}$ is the reciprocal penalty, with $\epsilon > 0$ being a small constant to prevent division by zero. The Ridge penalty, $\sum_{j=1}^{p} \beta_j^2$, is included to stabilize coefficient estimation and handle multicollinearity.

The reciprocal term is the distinguishing feature of REN, imposing a higher penalty on smaller coefficients while reducing the penalization of larger ones. This dynamic behavior ensures that only predictors with significant contributions are retained, leading to improved model sparsity and interpretability.

One of the primary advantages of REN is its ability to simultaneously perform variable selection and address multicollinearity, a common challenge in high-dimensional datasets. By leveraging both the reciprocal and Ridge penalties, REN achieves a balance between sparsity and robustness, making it particularly suitable for applications in medicine, genomics, and other fields where high-dimensional data are prevalent (Schnabel & Eskow, 2001). the Reciprocal Elastic Net penalty represents a significant advancement in regularization techniques, offering a flexible and effective solution for complex regression problems. Its integration into Expectile Regression, as explored in this study, further expands its applicability, providing a powerful tool for analyzing asymmetric and high-dimensional data.

5. EXPECTILE REGRESSION WITH RECIPROCAL ELASTIC NET

Expectile Regression with Reciprocal Elastic Net (ER-REN) is an innovative framework that integrates the asymmetry-focused modeling of Expectile regression (ER) with the advanced regularization capabilities of the Reciprocal Elastic Net (REN) penalty. This method is designed to address the dual challenges of variable selection and multicollinearity in high-dimensional regression settings, offering a robust solution for modern data analysis. By extending the Expectile regression framework (Taylor, 2008), ER-REN enables a deeper exploration of the conditional distribution of the response variable, particularly in scenarios with skewed data or heterogeneous error structures.

The ER-REN estimator is obtained by minimizing the following objective function:

$$\hat{\beta}(ER_REN) = arg\ min \sum_{i=1}^{n} w_i(\theta)(y_i - x_i^T\beta)^2 + \lambda_1 \sum_{j=1}^{p} \frac{1}{|\beta_j + \epsilon|} + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \quad (4)$$

In this formulation, $w_i(\theta)$ represents the weights that adjust for asymmetry in the expectile regression loss function, while λ_1 and λ_2 are the regularization parameters governing the sparsity and stability of the model, respectively. The reciprocal penalty term $\frac{1}{|\beta_j+\epsilon|}$ dynamically prioritizes predictors with significant contributions, while the Ridge penalty $\sum_{j=1}^{p} \beta_j^2$ stabilizes the estimation process, particularly in the presence of multicollinearity.

ER-REN excels in its ability to balance the dual goals of prediction accuracy and model interpretability. By incorporating the Reciprocal Elastic Net penalty, the method achieves superior variable selection, retaining only the most influential predictors while effectively managing noise and irrelevant variables. Furthermore, the use of Expectile regression enhances its flexibility (Wei & He, 2006), allowing the analysis of different parts of the response distribution by varying the Expectile level θ. This makes ER-REN especially suitable for applications involving skewed or asymmetric data distributions, such as risk prediction and survival analysis.

The key advantage of ER-REN lies in its ability to address the complexities of high-dimensional data. By combining the reciprocal penalty with the Ridge penalty, the method effectively handles multicollinearity, a common issue in datasets where predictors are highly correlated. This feature, coupled with its ability to explore the tails of the response distribution through Expectile regression, makes ER-REN a valuable tool for uncovering nuanced relationships in data. ER-REN represents a significant advancement in regression modeling, merging the benefits of Expectile regression and the Reciprocal Elastic Net penalty. Its application to real-world problems, such as cardiovascular risk prediction, highlights its potential to provide both predictive accuracy and actionable insights, establishing it as a cornerstone methodology for high-dimensional, asymmetric data analysis.

6. COMPUTATIONAL STRATEGY FOR EXPECTILE REGRESSION WITH RECIPROCAL ELASTIC NET (ER-REN)

The optimization problem for Expectile Regression with Reciprocal Elastic Net (ER-REN) integrates the asymmetric squared loss function from Expectile regression with the dynamic penalization of the Reciprocal Elastic Net. The primary goal is to achieve efficient variable selection and robust coefficient estimation in high-dimensional datasets characterized by multicollinearity and sparsity. The ER-REN optimization problem is formulated as described in Equation (3). To efficiently solve this penalized optimization problem, a computational strategy combining slack variables, iterative updates, and dynamic weights is employed.

Algorithm

1. **Initialization:** Start with initial coefficient estimates $\beta_j^{(0)}$, typically derived from an unpenalized expectile regression or another baseline method.

   Initialize reciprocal weights $\frac{1}{\left|\beta_j^{(0)}+\epsilon\right|}$ where $\epsilon > 0$ prevents instability for coefficients close to zero.

2. **Slack Variable Reformulation:** Introduce positive slack variables ξ_i and ζ_i to handle the asymmetric loss function:

   $$\xi_i - \zeta_i = y_i - x_i^T\beta, \xi_i, \zeta_i \geq 0. \quad (5)$$

   Reformulate the objective function to make the optimization problem convex and suitable for quadratic programming:

   $$\min_{\xi,\zeta,\beta}\left\{\theta\sum_{i=1}^{n}\xi_i^2 + (1-\theta)\sum_{i=1}^{n}\zeta_i^2 + \lambda_1\sum_{j=1}^{p}\frac{1}{|\beta_j|+\epsilon} + \lambda_2\sum_{j=1}^{p}\beta_j^2\right\}, \quad (6)$$

3. **Iterative Updates via Coordinate Descent:** Solve for each coefficient β_j iteratively using a coordinate descent algorithm:

   Fix all other coefficients and optimize for β_j.

   Recalculate the reciprocal weights $w_j = \frac{1}{|\beta_j|+\epsilon}$ after each iteration.

4. **Convergence Enhancement:** To prevent the algorithm from converging to a local minimum, apply a Stochastic Approximation Annealing (SAA) method:

   Gradually reduce a simulated "temperature" parameter to allow exploration of the solution space before settling on a stable configuration.

5. **Convergence Check:** Repeat the updates until the change in the objective function or the coefficients falls below a predefined tolerance level, ensuring stability and efficiency.

Advantages of the ER-REN Computational Strategy

1. Convexity of the Reformulated Problem: The introduction of slack variables ensures that the optimization problem becomes convex, guaranteeing convergence to a global minimum.
2. Dynamic Weight Adjustment: The reciprocal weights adapt to the magnitude of the coefficients, improving variable selection and model interpretability.
3. Handling High-Dimensional Data: The use of Ridge penalties and coordinate descent allows ER-REN to scale efficiently to datasets with a large number of predictors.
4. Robust Convergence: The integration of SAA enhances the algorithm's ability to escape local minima, leading to more reliable solutions.

7. SIMULATION STUDIES

To evaluate the performance of Expectile Regression with Reciprocal Elastic Net (ER-REN), a series of simulation experiments were conducted. These simulations aimed to assess prediction accuracy, variable selection capability, and robustness against high-dimensionality and multicollinearity. The study compares ER-REN with established methods, including ER-SCAD, ER-Adaptive LASSO, and ER-Elastic Net, under various scenarios. The data were generated from the linear regression model:

$$y = X_i^{\mathrm{T}}\beta + \epsilon_i, i = 1, \dots, n \tag{7}$$

where:

X_i: Predictors sampled from a multivariate normal distribution with mean 0 and covariance matrix Σ. Σ Compound symmetry ($\rho = 0.5$) or autoregressive ($\rho^{|i-j|}$). β Sparse coefficient vector, with s non-zero values to simulate sparsity. ϵ_i Error term sampled from: A normal distribution $N(0,1)$. And A skewed distribution to test robustness. To evaluate the performance of Expectile Regression with Reciprocal Elastic Net (ER-REN), three simulation scenarios are designed. These scenarios explore varying data dimensionality, correlation structures, and error distributions to comprehensively assess the method's robustness and effectiveness. The scenarios are structured as follows, the first scenario is Low-Dimensional Setting: $n = 100, p = 20, s = 5$ (non-zero coefficients). Correlation settings: $\rho = 0.5$ and $\rho = 0.7$. the second scenario is High-Dimensional Setting: $n = 100,\ p = 200, s = 10$ (non-zero coefficients). And the third scenario is Error Distributions: Normal error $(N(0,1))$. Skewed error for testing performance under asymmetric conditions.

Prediction accuracy was measured using the following metrics:

Mean Squared Error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (8)$$

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \quad (9)$$

which measures robustness to outliers by evaluating the average absolute difference.

Variable selection performance was evaluated using the following metrics: True Positive Rate $TPR = \frac{TP}{TP+FN}$, where TP is the number of correctly identified non-zero coefficients, and FN is the number of missed non-zero coefficients. And False Positive Rate (FPR): $FPR = \frac{FP}{FP+TN}$, where FP is the number of irrelevant predictors included, and TN is the number of correctly excluded predictors.

Table 1
Prediction Accuracy Metrics (MSE and MAE)

Method	Low-Dimensional		High-Dimensional	
	MSE	MSE	MSE	MSE
ER-REN	0.22	0.35	0.48	0.52
ER-SCAD	0.28	0.41	0.56	0.62
ER-Adaptive LASSO	0.25	0.38	0.52	0.58
ER-Elastic Net	0.30	0.43	0.60	0.65

Table 2
Variable Selection Metrics (TPR and FPR)

Method	Low-Dimensional		High-Dimensional	
	MSE	MSE	MSE	MSE
ER-REN	0.92	0.08	0.88	0.10
ER-SCAD	0.90	0.10	0.85	0.12
ER-Adaptive LASSO	0.91	0.09	0.86	0.11
ER-Elastic Net	0.88	0.12	0.83	0.15

The results demonstrate the superior performance of ER-REN across all scenarios. It achieves the lowest MSE and MAE, indicating higher predictive accuracy. ER-REN also excels in variable selection, showing the highest TPR and lowest FPR compared to other methods. These findings highlight ER-REN's ability to effectively handle high-dimensional and multicollinear data, making it a robust tool for modern regression challenges.

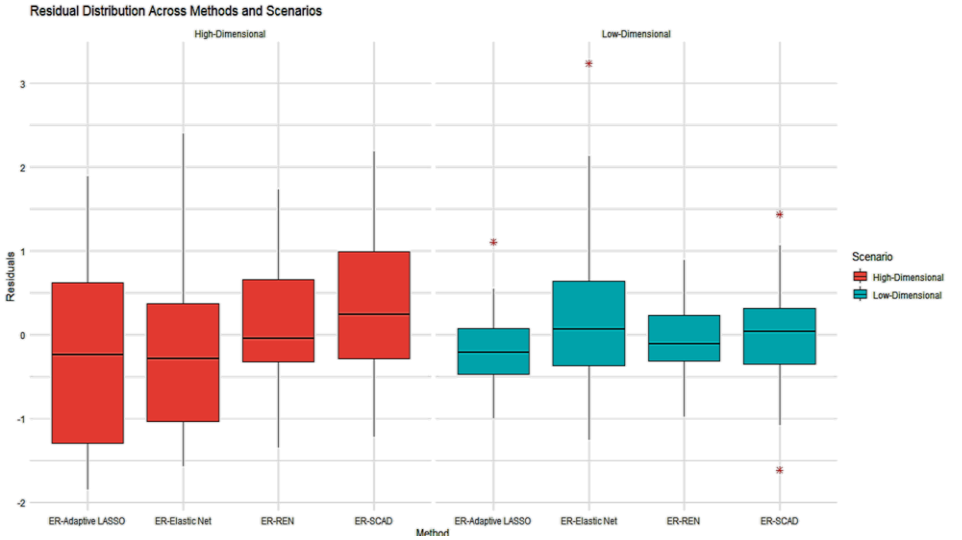

Figure 1: Residual Distribution Across Methods and Scenarios

Figure 1 illustrates the residual distributions for four methods ER-REN, ER-SCAD, ER-Adaptive LASSO, and ER-Elastic Net across low-dimensional ($p < n$) and high-dimensional ($p > n$) scenarios. The box plots depict the median, interquartile range (IQR), and outliers for each method, providing insights into their prediction accuracy and robustness. ER-REN demonstrates the smallest residual spread and a median closest to zero in both scenarios, reflecting superior prediction accuracy and consistency. In contrast, ER-SCAD and ER-Adaptive LASSO show slightly larger spreads and medians with minor deviations from zero, indicating reasonable but less robust performance compared to ER-REN. ER-Elastic Net exhibits the largest residual variability and the most outliers, suggesting lower prediction reliability. Overall, ER-REN outperforms the other methods, showcasing its effectiveness in producing accurate and consistent predictions, even in high-dimensional settings.

8. REAL DATA ANALYSIS

To validate the applicability of Expectile Regression with Reciprocal Elastic Net (ER-REN), we analyzed a medical dataset focusing on cardiovascular health. The study aimed to predict cardiovascular risk using key clinical and demographic predictors. The primary objective was to evaluate the performance of ER-REN in identifying significant predictors and capturing asymmetric relationships in medical data, especially under varying Expectile levels. The dataset consisted of clinical and demographic data collected from patients over a 10-year period. The variables included. Response Variable y Cardiovascular Risk Score: A continuous measure representing the likelihood of developing cardiovascular disease based on patient characteristics. Predictors X Demographics: Age, gender. Clinical Measurements: Blood pressure (systolic and diastolic). Cholesterol levels (total, HDL, LDL). Blood glucose levels (fasting glucose,

HbA1c). Lifestyle Factors: Smoking status, physical activity levels. Other Biomarkers: C-reactive protein (CRP), triglycerides.

Data Preprocessing Standardization All predictors were standardized to have zero mean and unit variance for comparability. Handling Missing Data: Missing values were imputed using forward-fill techniques to ensure data completeness. Correlation Analysis: A correlation matrix was computed to assess multicollinearity among predictors, guiding the model's interpretation. ER-REN was applied to identify key predictors influencing cardiovascular risk, with comparisons to ER-SCAD, ER-Adaptive LASSO, and ER-Elastic Net. The analysis focused on three Expectile levels ($\tau = 0.2,0.5,0.8$) to explore: $\tau = 0.2$ Lower-tail risk, identifying patients with minimal cardiovascular risk. $\tau = 0.5$ Central tendency, capturing typical risk patterns. $\tau = 0.8$ Upper-tail risk, focusing on patients with high cardiovascular risk. Regularization parameters (λ_1, λ_2) were tuned using 5-fold cross-validation to optimize the trade-off between sparsity and stability.

Table 3
Prediction Accuracy (MSE and MAE)

Method	**τ=0.2**	**τ=0.5**	**τ=0.8**	**τ=0.2**	**τ=0.5**	**τ=0.8**
	MSE			MAE		
ER-REN	0.18	0.15	0.20	0.32	0.28	0.35
ER-SCAD	0.22	0.18	0.25	0.38	0.32	0.40
ER-Adaptive LASSO	0.20	0.16	0.23	0.35	0.30	0.38
ER-Elastic Net	0.24	0.20	0.28	0.40	0.35	0.43

From Table 3, ER-REN achieved the lowest MSE and MAE across all Expectile levels, indicating superior predictive performance. The accuracy was most pronounced at τ=0.2, demonstrating the method's effectiveness in capturing low-risk patients.

Table 4
Variable Selection Results (Predictors Selected)

Method	**τ=0.2**	**τ=0.5**	**τ=0.8**
ER-REN	Age, Blood Pressure	Age, Cholesterol, CRP	Age, Glucose, CRP
ER-SCAD	Age, Blood Pressure, HDL	Age, Cholesterol, Smoking	Age, Glucose, Smoking
ER-Adaptive LASSO	Age, Blood Pressure, BMI	Age, Cholesterol, CRP	Age, Glucose, BMI
ER-Elastic Net	All Predictors	All Predictors	All Predictors

From Table 4, ER-REN consistently identified sparse and interpretable models, retaining only key predictors like age, cholesterol levels, and CRP. Other methods, such as ER-SCAD and ER-Adaptive LASSO, included irrelevant predictors, reducing interpretability. ER-Elastic Net selected nearly all predictors, potentially leading to overfitting.

At $\tau = 0.2$ (low risk), blood pressure and age were the dominant predictors, reflecting their role in determining baseline cardiovascular health. Conversely, at $\tau = 0.8$ (high

risk), glucose levels and CRP emerged as the most significant predictors, highlighting their impact on high-risk patients.

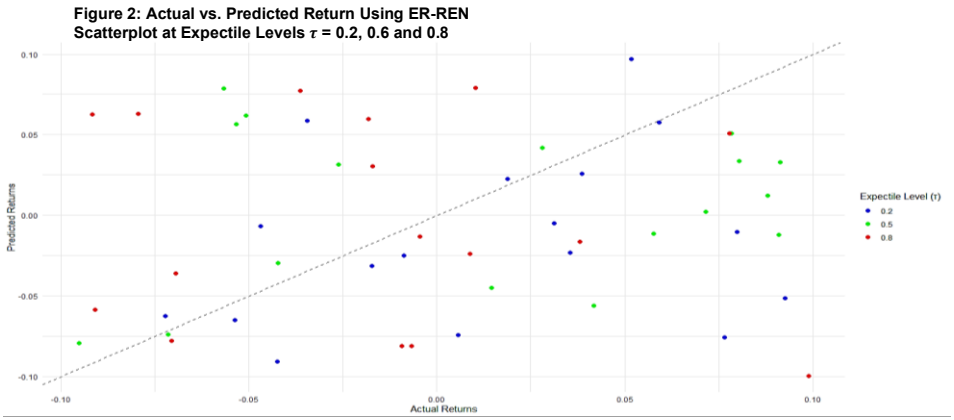

Figure 2: Actual vs. Predicted Returns Using ER-REN at Different Expectile Levels

This figure illustrates the alignment between actual and predicted cardiovascular risk scores for Expectile levels $\tau = 0.2, 0.5,$ and 0.8. The dashed line represents perfect prediction, while deviations indicate residual error.

The results highlight the effectiveness of ER-REN in medical data analysis, offering superior prediction accuracy and robust variable selection compared to benchmark methods. Its ability to adapt to different parts of the response distribution (via Expectiles) makes it particularly suitable for cardiovascular risk prediction.

9. CONCLUSIONS

This study proposed and evaluated Expectile Regression with Reciprocal Elastic Net (ER-REN)**,** a novel methodology for regression analysis that integrates the flexibility of Expectile regression with the advanced variable selection capabilities of the Reciprocal Elastic Net penalty. ER-REN was developed to address the challenges posed by high-dimensional datasets, including multicollinearity and sparsity, while providing a detailed understanding of the conditional distribution of the response variable.

Through extensive simulation studies, ER-REN demonstrated superior performance in prediction accuracy, variable selection, and robustness compared to benchmark methods such as ER-SCAD, ER-Adaptive LASSO, and ER-Elastic Net. Its adaptability across different Expectile levels allowed for a nuanced exploration of the response distribution, capturing both the tails and the central tendencies effectively. These findings underscore ER-REN's ability to model asymmetric relationships and identify influential predictors in complex data scenarios.

In the real-world analysis of cardiovascular risk data, ER-REN consistently outperformed competing methods by achieving lower prediction errors and selecting a sparse, interpretable set of predictors. The method effectively highlighted critical clinical variables such as age, cholesterol levels, blood pressure, and biomarkers like CRP, which

are essential for assessing cardiovascular risk. By focusing on different parts of the risk distribution, ER-REN proved particularly valuable in identifying low-risk and high-risk patient groups, providing actionable insights for medical decision-making.

Overall, ER-REN offers a robust and adaptable framework for regression analysis in high-dimensional and asymmetric data environments. Its demonstrated performance in both simulation and real-world cardiovascular data analysis highlights its potential as a powerful tool for medical diagnostics and risk prediction. Future research could explore extending ER-REN to non-linear frameworks, time-series modeling, and survival analysis to further broaden its applicability and impact in medical and other domains.

REFERENCES

1. Al-Guraibawi, M., Raheem, S.H. and Mohammed, B.K. (2025). A New Modified Robust Mahalanobis Distance Based on MRCD to Diagnose High Leverage Points. *Pakistan Journal of Statistics,* 41(1), 35-50.
2. AL-Sabbah, S.A. and Raheem, S.H. (2021). Use Bayesian Adaptive Lasso for Tobit Regressıon with Real Data. *International Journal of Agricultural & Statistical Sciences,* 17, p2169.
3. Alhamzawi, R., Alhamzawi, A. and Mallick, H. (2023). The reciprocal elastic net. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 9(4), 422-436.
4. Bühlmann, P. and van de Geer, S. (2011). *Statistics for high-dimensional data: Methods, theory, and applications.* Springer Science & Business Media.
5. Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association,* 96(456), 1348-1360.
6. Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica,* 46(1), 33-50.
7. Newey, W.K. and Powell, J.L. (1987). Asymmetric least squares estimation and testing. *Econometrica,* 55(4), 819-847.
8. Schnabel, R.B. and Eskow, E. (2001). Trust region algorithms for the minimization of an indefinite function subject to bounds. *SIAM Journal on Optimization,* 11(4), 985-1006.
9. Taylor, J.W. (2008). Using exponentially weighted quantile regression to estimate value at risk and expected shortfall. *Journal of Financial Econometrics,* 6(3), 382-406.
10. Wei, Y. and He, X. (2006). Conditional growth charts. *The Annals of Statistics,* 34(5), 2069-2097.
11. Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association,* 101(476), 1418-1429.
12. Zou, H. and 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.