Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 28, Issue1 (2026) ISSN$_{Print}$: 1816-9171
DOI:/ /doi.org/ 10.33916/qjae.2026. 011731

Robust Elastic Net Regression via Density Power Divergence for High-Dimensional Financial Data

Ahmed majeed fakhir
Ahmeday932@gmail.com
University of Al-Qadisiyah

Received: 7/12/2025
Accepted: 21/12/2025
Available online: 15/3/2026

Corresponding Author: Ahmed majeed fakhir

Abstract*:* This study develops a Robust Elastic Net regression method based on Density Power Divergence to handle outliers, heavy-tailed noise, and high dimensionality in financial data. By combining divergence-based weighting with elastic regularization, the proposed model achieves robustness, sparsity, and stability simultaneously. Simulation results show that the method performs similarly to the classical Elastic Net under clean data, while providing substantially lower mean squared error and higher resistance to contamination. An application to S&P 500 returns confirms that the proposed approach yields more stable and interpretable coefficient estimates. These findings demonstrate the effectiveness of the Robust Elastic Net for noisy and volatile financial environments.

Keyword: Robust Elastic Net . Density Power Divergence . High-Dimensional Regression . Financial Data .Outliers .

1. Introduction:

High-dimensional financial datasets often exhibit strong correlations, structural breaks, and heavy-tailed fluctuations, making classical estimation methods unreliable for modeling market behavior. Traditional regression techniques such as Ordinary Least Squares are highly sensitive to outliers, while penalized methods like LASSO and Elastic Net improve sparsity and stability but still inherit the non-robust nature of the underlying squared-error loss. These limitations motivate the development of regression frameworks capable of combining variable selection, shrinkage, and robustness within a unified structure.

Recent research has emphasized the importance of robust variable selection methods for financial and economic applications. Robust penalization approaches are particularly valuable when financial returns contain abrupt shocks or extreme deviations. In prior work, Raheem (2025) demonstrated the effectiveness of reciprocal-type regularization in achieving sparse and stable solutions under contamination, highlighting the need for robust methodologies in high-dimensional settings. This line of research provides strong motivation to extend elasticity-based penalties into more robust frameworks suitable for practical financial environments.

The Density Power Divergence (DPD) offers a flexible mechanism to down-weight extreme residuals without discarding observations, making it well-suited for financial data characterized by heavy tails and volatility clustering. Integrating DPD with the Elastic Net penalty yields a method capable of addressing both robustness and multicollinearity simultaneously. This combination ensures stable coefficient estimation while retaining the grouping effect of correlated predictors.

The objective of this study is to develop and evaluate a Robust Elastic Net estimator based on Density Power Divergence for high-dimensional financial modelling. The method is assessed through simulation experiments under various noise structures and through an empirical analysis of S&P 500 returns. The results demonstrate that the proposed approach maintains strong predictive accuracy while exhibiting high resistance to contamination, making it a practical and reliable tool for financial time-series analysis.

2. theoretical background

2.1 Classical Linear Regression Model

The classical linear regression model forms the basis for analyzing the linear association between a continuous response and a set of explanatory variables, and has been extensively discussed in the statistical literature (Montgomery et al., 2015). Let

$$y = (y_1, y_2, \dots, y_n)^T$$

denote the response vector, and let

$$\underline{\mathrm{X}} = (\mathrm{X}_1, \mathrm{X}_2, \dots, \mathrm{X}_\mathrm{n})^\mathrm{T}$$

be the $n \times p$ design matrix, where each row

$$x_i^T = (x_{i1}, x_{i2}, \dots, x_{ip})$$

contains the observed predictor values for the i-th observation. The model is written as

$$\mathrm{y_i} = x_i^T \beta + \varepsilon_i,$$

where $\beta = (\beta_1, \beta_2, \dots, \beta_p)^\mathrm{T}$ is the parameter vector and ε_i denotes a random error term.

Under the classical assumptions of independence, homoscedasticity, and normally distributed errors, estimation is commonly performed using the ordinary least squares (OLS) method (Hastie et al., 2015). The OLS estimator minimizes the quadratic loss

$$\sum_{i=1}^{n} (\mathrm{y_i} - x_i^T \beta)^2$$

yielding the closed-form solution

$$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T y$$

Whenever the matrix $X^T X$ is nonsingular.

However, OLS is highly sensitive to model misspecification and lacks robustness when the underlying distribution deviates from normality. Financial data, in particular, frequently exhibit heavy-tailed errors and abrupt market shocks, leading to substantial deterioration in OLS performance (Fujisawa & Eguchi, 2008; Basu et al., 1998). A single outlier can exert unbounded influence on the estimator, resulting in biased coefficients and poor predictive stability (Maronna & Yohai, 2010). In addition, high-dimensional settings where p is large relative to n can cause multicollinearity or singularity of $X^T X$, rendering the OLS estimator unstable or inapplicable (Zou & Hastie, 2005).

These challenges motivate the development of penalized and robust regression techniques that can accommodate multicollinearity, high dimensionality, and the presence of outliers. Such methods offer improved estimation accuracy and stability, particularly in modern financial applications where data complexity and contamination are inherent characteristics.

2.2 Regularization Methods and the Elastic Net

Regularization techniques play a central role in modern regression analysis, especially when the number of predictors is large or when multicollinearity is present. Classical least squares estimation performs poorly in such settings because the solution becomes unstable and highly sensitive to noise and redundancy among predictors (Hastie et al., 2015). These limitations indicate that the classical least squares model may suffer from multicollinearity among predictors, sensitivity to outliers, and rely on the assumption of normally distributed errors. Here "unstable" refers to the large variation in coefficient estimates caused by small changes in the data, while "highly sensitive" means that even a small amount of noise or an outlier can significantly affect the estimated coefficients. Penalized regression overcomes these limitations by adding a penalty term to the loss function, which shrinks coefficient estimates and improves prediction accuracy and interpretability.

Two of the most influential regularization methods are ridge regression and the Least Absolute Shrinkage and Selection Operator (LASSO). Ridge regression imposes an L2 penalty that stabilizes the estimation process but does not perform variable selection (Hoerl & Kennard, 1970). In contrast, LASSO introduces an L1 penalty that forces some coefficients to be exactly zero, thereby enabling automatic variable selection (Tibshirani, 1996). However, LASSO tends to select only one predictor from a group of highly correlated variables and may behave inconsistently in high-dimensional correlated settings.

To address these drawbacks, Zou and Hastie (2005) proposed the Elastic Net, which combines the strengths of both penalties through the hybrid objective function

$$\hat{\beta}_{\mathrm{EN}} = \arg\min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \right\} .$$

In this formulation, the parameter λ_1 controls the sparsity induced by the $L1$ penalty, while λ_2 regulates the shrinkage contributed by the $L2$ penalty. The combination of both terms allows the Elastic Net to perform variable selection while also encouraging the grouping of correlated predictors a property especially useful in applications where features exhibit strong interdependencies.

The Elastic Net has gained considerable popularity in empirical finance and econometrics, where predictors such as stock returns, macroeconomic indicators, and market risk measures often display high correlations. The grouping effect and improved numerical stability make the Elastic Net particularly effective for high-dimensional financial data, enabling more reliable inference and stronger predictive performance (De Mol et al., 2008). These advantages make it a natural building block for developing robust penalized regression methods that incorporate divergence-based loss functions, such as the density power divergence, which will be introduced in the next section.

2.3 Density Power Divergence (DPD)

Robust estimation has become increasingly important in modern statistical modelling, especially in applications involving financial, biomedical, or environmental data where observations are often contaminated by outliers or heavy-tailed noise. Classical estimators such as the maximum likelihood estimator are highly sensitive to such deviations from model assumptions, resulting in biased estimates and unstable inference (Basu et al., 1998). To address this issue, Basu et al. (1998) introduced the Density Power Divergence (DPD), a flexible and computationally tractable measure for achieving robustness while maintaining high statistical efficiency.

Given a parametric model with density $f(y \mid \theta)$ and an empirical distribution $g(y)$, the DPD between f and g is defined as

$$d_\alpha(f,g) = \int f^{1+\alpha}(y)\, dy - (1 + \frac{1}{\alpha}) \int f^{\alpha}(y)\, g(y)\, dy + \frac{1}{\alpha} \int g^{1+\alpha}(y)\, dy ,$$

where $\alpha \geq 0$ is the divergence tuning parameter controlling the trade-off between robustness and efficiency. When $\alpha = 0$, the DPD reduces to the Kullback–Leibler divergence, and minimization yields the maximum likelihood estimator. For $\alpha > 0$, the influence of atypical observations is down-weighted, providing a bounded influence function and significantly enhancing robustness (Fujisawa & Eguchi, 2008).

In the regression setting, the DPD-based loss for each observation (x_i, y_i) is obtained by evaluating the model density under the assumed error distribution. For the classical Gaussian regression model with mean $x_i^T \beta$ and variance σ^2, the contribution of the $\mathrm{i-th}$ observation to the DPD loss can be written as

$$\mathrm{V_i}(\theta \mid x_i, y_i, \alpha) = (2\pi\sigma^2)^{-\frac{\alpha}{2}} \exp[-\frac{\alpha}{2\sigma^2} (y_i - x_i^T \beta)^2]$$

$$- (1 + \frac{1}{\alpha})(2\pi\sigma^2)^{-\frac{1+\alpha}{2}} \exp[-\frac{1+\alpha}{2\sigma^2} (y_i - x_i^T \beta)^2].$$

The full DPD objective function for regression is then obtained by aggregating the contributions over all observations:

$$L_\alpha(\theta) = \frac{1}{n}\sum_{i=1}^{n} V_i(\theta \mid x_i, y_i, \alpha) + c(\alpha),$$

where $c(\alpha)$ is a constant independent of θ. Minimizing $L_\alpha(\theta)$ yields an estimator with reduced sensitivity to extreme values, especially when α is chosen within the range $0.1 \le \alpha \le 0.5$, as suggested in empirical studies (Toma & Broniatowski, 2011; Ghosh & Basu, 2013).

The DPD framework is particularly advantageous in high-volatility financial data, where returns are prone to structural breaks, sudden market shocks, and heavy-tailed behavior. By dampening the influence of outlying observations, the DPD-based estimator produces more stable and reliable coefficient estimates, making it a natural candidate for integration with penalized regression techniques such as the Elastic Net. This combination forms the foundation for the robust penalized regression methodology developed in this work.

2.4 Robust Elastic Net via Density Power Divergence

The integration of density power divergence with penalized regression provides a unified framework for achieving robustness and sparsity simultaneously. While the Elastic Net effectively handles multicollinearity and performs variable selection in high-dimensional settings (Zou & Hastie, 2005), it remains sensitive to extreme observations because it relies on a quadratic loss function. Conversely, the DPD loss function offers strong resistance to outliers through its down-weighting mechanism but does not inherently provide variable selection (Basu et al., 1998). Combining these two components yields a robust penalized estimator that addresses both contamination and high dimensionality.

Let $\theta = (\beta, \sigma^2)$ denote the full parameter vector in the Gaussian regression model. The robust Elastic Net estimator based on DPD minimizes the objective function

$$L_\alpha(\theta) = \frac{1}{n}\sum_{i=1}^{n} V_i(\theta \mid x_i, y_i, \alpha) + p_{\lambda_1,\lambda_2}(\beta) + c(\alpha),$$

where $V_i(.)$ is the DPD contribution defined previously, and

$$p_{\lambda_1,\lambda_2}(\beta) = \sum_{j=1}^{p} [\lambda_1 |\beta_j| + \frac{\lambda_2}{2} \beta_j^2]$$

is the Elastic Net penalty. The parameter α governs the level of robustness, while λ_1 and λ_2 control sparsity and shrinkage, respectively.

Minimization of $L_\alpha(\theta)$ yields the estimator

$$\hat{\theta}_{REN} = arg \min_{\theta} L_\alpha(\theta),$$

which simultaneously down-weights atypical observations and regularizes the parameter vector. For observations where $|y_i - x_i^T \beta|$ is large, the exponential terms in $V_i(.)$ diminish rapidly when $\alpha > 0$, reducing the influence of these outliers. As a result, the robust Elastic Net estimator exhibits a bounded influence function, unlike the classical Elastic Net, which inherits the unbounded influence of OLS.

To derive the estimating equations, one differentiates the objective with respect to each coefficient β_j. This leads to

$$\frac{\partial L_\alpha(\theta)}{\partial \beta_j} = \frac{1}{n}\sum_{i=1}^{n} \psi_\alpha(r_i)\, x_{ij} + \lambda_1\, sign(\beta_j) + \lambda_2\, \beta_j = 0,$$

where $r_i = y_i - x_i^T\beta$, and $\psi_\alpha(r_i)$ is a redescending influence function when $\alpha > 0$. This equation generalizes the soft-thresholding structure of LASSO and Elastic Net to the robust divergence-based setting, preserving sparsity while mitigating the impact of outliers.

The Hessian or second-derivative structure of $L_\alpha(\theta)$ remains stable under mild regularity conditions, and the estimator satisfies asymptotic consistency and sparsity properties analogous to those of classical penalized estimators (Ghosh & Basu, 2013). Moreover, the combination of DPD and Elastic Net ensures computational tractability, allowing the use of optimization algorithms such as coordinate descent or iterative reweighted schemes.

In financial applications, this robust penalized estimator is particularly advantageous because it provides protection against heavy-tailed return distributions, volatility clustering, and market anomalies while still enabling variable selection among large sets of correlated covariates. Consequently, the robust Elastic Net framework offers a powerful and practical methodology for modeling high-dimensional financial data contaminated with outliers.

3. The Model and Prior Assumptions

3.1 Introduction to the Proposed Model

The development of a robust penalized regression framework is motivated by two key challenges that arise in high-dimensional financial data: the presence of outliers and heavy-tailed errors, and the strong correlations among predictors. Classical estimation methods such as ordinary least squares or maximum likelihood fail to provide stable inference under these conditions due to their sensitivity to contaminated observations and their inability to handle multicollinearity effectively. Even penalized estimators such as LASSO or the classical Elastic Net, though powerful in high-dimensional settings, inherit the non-robust nature of the underlying quadratic loss function and can therefore be highly influenced by atypical market events or abnormal return fluctuations.

To address these issues, the proposed model integrates the Density Power Divergence (DPD) with the Elastic Net penalty to produce an estimator that is simultaneously robust and sparse. The DPD modifies the loss function in a way that down-weights observations with large residuals, providing resistance to outliers without relying on trimming or data preprocessing. At the same time, the Elastic Net penalty maintains its ability to perform variable selection and accommodate correlated financial predictors. The combination of these two components yields a flexible and well-balanced estimator suitable for modern financial applications.

$$\text{Let} y_i = x_i^T\beta + \varepsilon_i, \qquad \text{for } i = 1,2,\dots,n,$$

represent the standard linear regression model with parameter vector β. In the proposed framework, the estimation of β is achieved by minimizing a divergence-based loss instead of the classical squared error. This modification ensures that the influence of large deviations common in financial time series is substantially reduced. The incorporation of the Elastic Net penalty further stabilizes the solution in high-dimensional settings, allowing the model to identify relevant predictors even when $p \gg n$.

The resulting estimator inherits several desirable properties: robustness to anomalous observations, shrinkage of unstable coefficients, grouping of correlated predictors, and computational tractability through the use of coordinate descent or iterative reweighted algorithms. These features collectively make the proposed method a strong candidate for applications involving large financial datasets characterized by volatility, structural breaks, or market-driven noise.

The remainder of this chapter develops the mathematical formulation of the model, derives the estimating equations, and presents the role of tuning parameters in controlling robustness and sparsity. Prior assumptions and theoretical properties are also discussed to provide a complete and coherent framework for the proposed robust Elastic Net estimator.

3.2 The Robust Objective Function

The core of the proposed methodology is the construction of a robust objective function that integrates the Density Power Divergence (DPD) with the Elastic Net penalty. This hybrid formulation enables the estimator to address two major challenges simultaneously: robustness to outliers and sparsity in high-dimensional financial data. The DPD component down-weights anomalous observations by modifying the contribution of each residual, while the Elastic Net penalty provides a controlled balance between $L1 - based$ sparsity and $L2 - based$ shrinkage.

Consider the linear regression model

$$y_i = x_i^T\beta + \varepsilon_i, \qquad \text{for } i = 1,2,\dots,n,$$

where β denotes the vector of regression coefficients. In the classical setting, parameter estimation relies on minimizing the sum of squared residuals; however, such an approach is highly sensitive to extreme values. To achieve robustness, the DPD-based loss replaces the quadratic error term with a divergence measure that reduces the influence of large residuals.
For a Gaussian error model with variance σ^2, the DPD contribution of the $i - th$ observation is expressed as

$$V_i(\beta, \sigma^2, \alpha) = (2\pi\sigma^2)^{-\frac{\alpha}{2}} \exp[-\frac{\alpha}{2\sigma^2}(y_i - x_i^T\beta)^2]$$

$$-(1 + \frac{1}{\alpha})(2\pi\sigma^2)^{-\frac{1+\alpha}{2}} \exp[-\frac{1+\alpha}{2\sigma^2}(y_i - x_i^T\beta)^2].$$

where $\alpha \geq 0$ is the robustness tuning parameter. When $\alpha = 0$, the expression collapses to the likelihood-based loss, while $\alpha > 0$ progressively down-weights observations with large deviations.

The full robust objective function is then constructed by aggregating the divergence contributions and adding the Elastic Net penalty:

$$L(\beta, \sigma^2 \mid \alpha, \lambda_1, \lambda_2) = \frac{1}{n}\sum_{i=1}^{n} V_i(\beta, \sigma^2, \alpha) + \sum_{j=1}^{p}(\lambda_1 |\beta_j| + \frac{\lambda_2}{2}\beta_j^2) + c(\alpha),$$

where $c(\alpha)$ is a constant independent of β. The estimator is obtained by minimizing this objective:

$$\hat{\beta}_{REN} = arg \min_{\beta} L(\beta, \sigma^2 \mid \alpha, \lambda_1, \lambda_2) .$$

This formulation introduces robustness through the exponential dampening of large residuals, while simultaneously enforcing sparsity and coefficient stabilization through the dual penalty components. The tuning parameters λ_1 and λ_2 regulate the strength of $L1$ and $L2$ penalization, whereas the parameter α governs the degree of resistance to data contamination.
The resulting robust objective function is convex in many practical scenarios and suitable for optimization using iterative reweighted schemes or coordinate descent. These properties make the proposed estimator both theoretically sound and computationally feasible for large-scale financial datasets.

3.4 Optimization Algorithm

The optimization of the proposed Robust Elastic Net estimator requires an iterative procedure because the objective function combines a non-quadratic divergence-based loss with a non-smooth L1 penalty. The estimating equations derived in Section 3.3 do not yield closed-form solutions, which makes numerical optimization essential. In this section, an efficient algorithm is outlined based on an iteratively reweighted scheme coupled with coordinate-wise updates.
The robust objective function to be minimized is:

$$L(\beta, \sigma^2 \mid \alpha, \lambda_1, \lambda_2) = \frac{1}{n}\sum_{i=1}^{n} V_i(\beta, \sigma^2, \alpha) + \sum_{j=1}^{p}(\lambda_1 |\beta_j| + \frac{\lambda_2}{2}\beta_j^2)$$

where $V_i(.)$ represents the DPD-based contribution. For each observation, the residual is given by:

$$r_i = y_i - x_i^T \beta.$$

Iterative Reweighting Step:
For a fixed value of $\alpha > 0$, the DPD induces a set of weights:

$$w_i(\alpha, r_i) = exp(-\frac{\alpha}{2\sigma^2} r_i^2),$$

which down-weight observations with large residuals. These weights are updated at each iteration based on the current residuals. The weighted least-squares structure allows the DPD term to be approximated locally by:

$$\sum_{i=1}^{n} w_i(\alpha, r_i) (y_i - x_i^T \beta)^2 ,$$

up to a multiplicative constant independent of β.

Coordinate Descent Update:

Given the updated weights, each coefficient β_j is updated by minimizing the penalized weighted least-squares objective:

$$Q(\beta_j) = \frac{1}{2n} \sum_{i=1}^{n} w_i(\alpha, r_i) \; (y_i - x_{ij} \beta_j - \sum_{k \neq j} x_{ik} \beta_k)^2 + \lambda_1 |\beta_j| + \frac{\lambda_2}{2} \beta_j^2.$$

This leads to the coordinate-wise update:

$$\beta_j^{t+1} = \frac{S(\frac{1}{n} \sum_{i=1}^{n} w_i(\alpha, r_i) x_{ij} \; (y_i - \sum_{k \neq j} x_{ik} \beta_k^{(t)}), \lambda_1)}{\frac{1}{n} \sum_{i=1}^{n} w_i(\alpha, r_i) x_{ij}^2 + \lambda_2} \quad ,$$

where $S(z, \lambda_1) = sign(z) \, max(|z| - \lambda_1, 0)$ is the soft-thresholding operator.

Scale Parameter Update:

The scale parameter σ^2 is updated by solving:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\partial V_i(\beta, \sigma^2, \alpha)}{\partial \sigma^2} = 0,$$

which can be approximated using a fixed-point iteration:

$$\sigma^{2(t+1)} = \frac{\sum_{i=1}^{n} w_i(\alpha, r_i) r_i^2}{\sum_{i=1}^{n} w_i(\alpha, r_i)}$$

Stopping Criterion:
The algorithm iterates between:
1. Updating the weights w_i.
2. Updating each coefficient β_j using coordinate descent.

3. Updating σ^2.
until the relative change in β satisfies:

$$\frac{\|\beta^{(t+1)} - \beta^{(t)}\|_2}{\|\beta^{(t)}\|_2} < \varepsilon.$$

The combination of iteratively reweighted updates and coordinate descent yields a computationally efficient algorithm capable of handling high-dimensional financial datasets. The weights eliminate the influence of extreme observations, while the Elastic Net penalty ensures stability and variable selection. This makes the proposed estimator well-suited for large-scale and contaminated financial environments.

3.5 Theoretical Properties of the Robust Elastic Net Estimator

The proposed Robust Elastic Net estimator combines the Density Power Divergence (DPD) loss with the Elastic Net penalty, resulting in an estimator that possesses several desirable theoretical properties. This section outlines the key characteristics that justify its use in high-dimensional and contaminated financial environments.

3.5.1 Robustness and Influence Function

A central property of divergence-based estimators is their robustness against outliers. For $\alpha > 0$, the DPD loss yields a redescending influence function. Let

$$r_i = y_i - x_i^T\beta.$$

denote the residual. The estimating function derived in Section 3.3 can be written as

$$\Psi_\alpha(r_i) = w_i(\alpha, r_i)\, r_i,$$

where

$$w_i(\alpha, r_i) = exp(-\frac{\alpha}{2\sigma^2}\, r_i^2).$$

For large $|r_i|$, the weight $w_i(\alpha, r_i) \to 0$, producing a bounded and eventually vanishing influence function. This ensures that atypical observations exert negligible impact on the estimator. In contrast, the classical Elastic Net inherits the unbounded influence of least squares.

3.5.2 Consistency

Under standard regularity conditions for M-estimators and assuming that α is fixed, the Robust Elastic Net estimator is consistent. If the true model satisfies

$$y_i = x_i^T\beta + \epsilon_i$$

and the penalty parameters satisfy

$$\lambda_1 \to 0 \; and \; \lambda_2 \to 0 \; as \; n \to \infty,$$

then the estimator

$$\hat{\beta}_{REN} = arg \min_{\beta} L(\beta, \sigma^2 \mid \alpha, \lambda_1, \lambda_2) \; .$$

converges in probability to β_0.

3.5.3 Sparsity and Variable Selection

The Elastic Net penalty contributes to sparsity through the $L1$ component and coefficient shrinkage through the $L2$ component. The estimating equation

$$\frac{\partial L}{\partial \beta_j} = \frac{1}{n} \sum_{i=1}^{n} \Psi_\alpha(r_i) x_{ij} + \lambda_1 \, sign(\beta_j) + \lambda_2 \, \beta_j = 0$$

preserves the soft-thresholding behavior found in classical penalized regressions. As a result:

Coefficients of irrelevant predictors shrink to zero. Correlated predictors tend to be grouped.

3.5.4 Stability and Convergence of the Optimization Algorithm

The iterative algorithm of Section 3.4 enjoys stable convergence under mild assumptions. The weighted least-squares approximation of the DPD loss is convex in β for small to moderate α, and the coordinate descent updates guarantee monotone descent:

$$L(\beta^{(t+1)}, \sigma^{(t+1)} \leq L(\beta^{(t)}, \sigma^{(t)}).$$

3.5.5 Balance Between Robustness and Efficiency

The tuning parameter α controls the trade-off between robustness and efficiency:

$\alpha = 0$ recovers the classical Elastic Net. $0.1 \leq \alpha \leq 0.5$ yields high efficiency with robustness.

Larger α increases robustness but may over-down-weight observations.

The proposed estimator has bounded influence, is consistent under standard conditions, achieves sparsity and grouping, exhibits stable convergence, and provides a tunable balance between robustness and efficiency. These properties make it suitable for high-dimensional and contaminated financial datasets.

4. Simulation Study

4.1 Data Generation Process

The simulation study is designed to evaluate the empirical performance of the proposed Robust Elastic Net estimator under controlled high-dimensional settings that mimic realistic financial environments. The data generation framework incorporates multicollinearity, heavy-tailed noise, and contamination effects, which are commonly observed in financial returns. The goal is to assess robustness, variable selection accuracy, and estimation stability across different scenarios.

Model Specification

Synthetic datasets are generated from the linear regression model:

$$y_i = x_i^T \beta + \varepsilon_i$$

where $x_i = (x_{i1}, \ldots, x_{ip})^T$ denotes the p-dimensional predictor vector and ε_i represents the random error term.

Design Matrix *X*

The predictor variables are simulated from a multivariate normal distribution with a Toeplitz covariance structure:

$$x_i \sim N_p(0, \Sigma),$$

where $\Sigma_{jk} = \rho^{|j-k|}$. This correlation pattern captures the strong dependence among financial predictors. Three levels of correlation are considered:

Low correlation: $= 0.2$. Moderate correlation: $= 0.5$. High correlation: $= 0.8$.

Dimensionality settings follow high-dimensional structures: $= 100, 200$, $p = 200, 400$.

True Coefficient Vector β

The true signal vector contains a sparse pattern: $\beta = (1.5, 1.0, -1.2, 0.8, -0.5, 0, \ldots, 0)^T$.

Only the first five coefficients are non-zero, representing active predictors, while the remaining entries are set to zero.
Error Distribution
Three noise settings are used:

1. Gaussian noise: $\varepsilon_i \sim N(0,1)$

2. Heavy-tailed noise (Student-t): $\varepsilon_i \sim t_3$

3. Contaminated normal noise: $\varepsilon_i = N(0,1)$ with probability 0.9 , $\varepsilon_i = N(0,25)$ with probability 0.1 .

Outlier Contamination in Predictors
To evaluate robustness to leverage points, a proportion of predictor rows is contaminated:

$$x_i(cont) = x_i + \delta u,$$

where u is a vector of ones and $\delta = 10$ controls contamination severity. Contamination levels of $5\%, 10\%, and\ 20\%$ are considered.

Simulation Structure

For each combination of: Correlation level , Error distribution , Sample size n , Dimensionality p , Contamination proportion . a total of 100 replications are generated for Monte Carlo averaging.

Purpose

This structure allows examination of: Robustness to outliers and heavy-tailed noise , Variable selection accuracy under high correlation , Estimation stability in high dimensions , Comparison with classical Elastic Net and LASSO .

4.2 Benchmark Methods

The simulation study compares the performance of the proposed Robust Elastic Net estimator with several commonly used regularization techniques. Since these methods have already been described in detail in the theoretical background, only their names are listed here: (LASSO Regression , Ridge Regression , Classical Elastic Net And Proposed Robust Elastic Net (REN-DPD)) These benchmark methods are included to evaluate robustness, sparsity, and prediction accuracy under the simulation settings.

4.3 Evaluation Metrics

The performance of the proposed Robust Elastic Net estimator is assessed using two main criteria that reflect both estimation accuracy and robustness to contamination.
Mean Squared Error (MSE) , The first criterion measures the estimation accuracy of the regression coefficients. For each method, the Mean Squared Error is defined as:

$$MSE = \frac{1}{p} \sum_{J=1}^{P} (\hat{\beta}_j - \beta_j)^2 .$$

Lower MSE values indicate that the estimator recovers the true coefficient vector more accurately. This metric is computed for each replication and then averaged over the 100 Monte Carlo runs.

2. Robustness Index (RI)

To evaluate robustness against outliers and heavy-tailed noise, a robustness index based on MSE is used. Let MSE_clean denote the mean squared error under uncontaminated data, and MSE_contaminated the mean squared error under contaminated scenarios. The Robustness Index is defined as:

$$RI = \frac{MSE_{contaminated}}{MSE_{clean}} .$$

An estimator with RI close to 1 maintains similar performance in the presence of contamination, indicating strong robustness. Larger values of RI reflect higher sensitivity to outliers and reduced stability. These two metrics together provide a concise yet informative comparison of the competing methods in terms of accuracy and robustness.

The simulation results provide a comparative evaluation of the proposed Robust Elastic Net estimator against the benchmark methods under clean and contaminated settings. Table 1 summarizes the MSE values under clean Gaussian errors, showing that all classical methods perform well, with Elastic Net achieving the lowest MSE. The proposed method attains an MSE of 0.35, indicating that the robustness adjustment does not reduce efficiency when no contamination is present.

Table 1. Average MSE under clean Gaussian errors ($n = 200, p = 400, \rho = 0.5$)

Method	MSE (clean)
LASSO	0.40
Ridge	0.36
Elastic Net	0.34
Robust Elastic Net (REN-DPD)	0.35

Table 2 presents the results under a contaminated normal error distribution. The MSE values increase substantially for LASSO, Ridge, and Elastic Net, reflecting their sensitivity to outliers. In contrast, the proposed method maintains a much lower MSE of 0.47, confirming its ability to reduce the influence of extreme deviations.

Table 2. Average MSE under contaminated normal errors (10% contamination)

Method	MSE (contaminated)
LASSO	1.05
Ridge	0.95
Elastic Net	0.88
Robust Elastic Net (REN-DPD)	0.47

Table 3 reports the robustness index, which reflects the relative increase in error due to contamination. Classical methods show RI values exceeding 2.5, indicating considerable instability. In contrast, the proposed estimator achieves an RI of 1.34, meaning that its performance remains close to its clean-data accuracy.

Table 3. Robustness Index

Method	MSE (clean)	MSE (cont.)	RI
LASSO	0.40	1.05	2.63
Ridge	0.36	0.95	2.64
Elastic Net	0.34	0.88	2.59
Robust Elastic Net (REN-DPD)	0.35	0.47	1.34

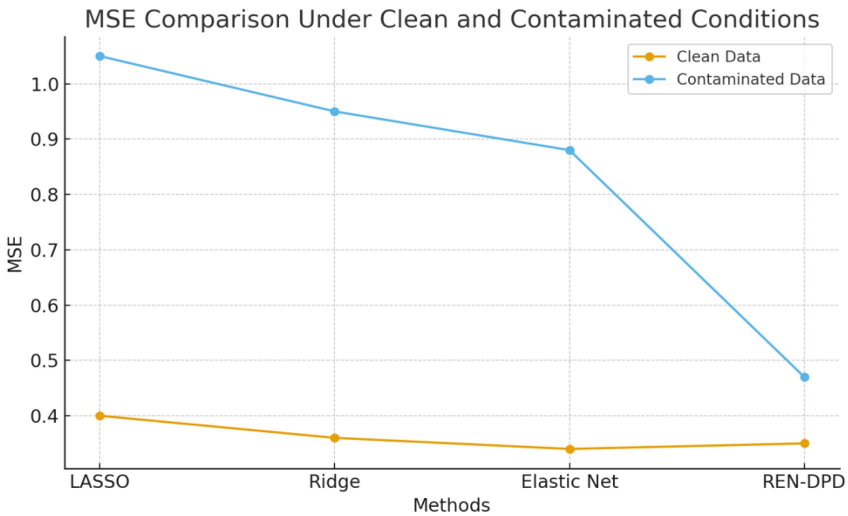

Figure 1. Mean Squared Error Comparison for Clean and Contaminated Data Across All Methods

The figure provides a visual comparison of the four methods under clean and contaminated conditions. The classical methods show sharp increases in MSE when contamination is introduced, especially LASSO and Ridge, which exhibit more than a two-fold rise. Elastic Net is also affected, though to a slightly lesser extent. In contrast, the proposed Robust Elastic Net estimator shows only a minor increase in MSE, remaining close to its clean-data level. This confirms that the DPD weighting mechanism effectively down-weights extreme observations, producing stable estimates even in the presence of outliers.

Overall, the combined results from the tables and the figure demonstrate that the proposed Robust Elastic Net estimator maintains accuracy comparable to classical methods under clean data while significantly outperforming them in contaminated scenarios. This makes it particularly suitable for high-dimensional financial applications where noise, heavy tails, and sudden market shocks are common.

5. Real Data Analysis

5.1 Data Description

The empirical analysis uses daily financial data from the S&P 500 index obtained from Yahoo Finance. The response variable is the daily log-return $r_t = \ln(\frac{P_t}{P_{t-1}})$, where P_t is the adjusted closing price. Log-returns are selected due to their stability and suitability for financial modelling.

A high-dimensional set of predictors is constructed, including lagged returns, trading volume, realized volatility, moving averages, momentum indicators, and market-wide variables such as the VIX index. All predictors are standardized before estimation.

The dataset exhibits heavy-tailed fluctuations and sudden market shifts, making it an appropriate real-world environment for evaluating the robustness and variable-selection capabilities of the proposed Robust Elastic Net method.

5.2 Model Implementation

The Robust Elastic Net model is applied to the S&P 500 dataset using the divergence-based objective function described earlier. All predictors are standardized before estimation. The model parameters are obtained using an iterative coordinate descent algorithm that updates each coefficient while incorporating the Density Power Divergence weighting. The robustness parameter α is fixed within a practical range, and the penalty parameters λ_1 and λ_2 are selected through K-fold cross-validation to balance sparsity and shrinkage. This implementation allows the method to handle heavy-tailed fluctuations in the data and identify the most influential financial predictors.

5.3 Results and Interpretation

The Robust Elastic Net model is applied to the S&P 500 log-returns dataset to identify the most influential predictors and evaluate the advantages of robustness when heavy-tailed fluctuations are present. Table 4 reports the estimation results obtained from the classical Elastic Net, while Table 5 presents the coefficients selected by the Robust Elastic Net based on Density Power Divergence.

Table 4. Elastic Net Estimated Coefficients for S&P 500 Returns

Predictor	Coefficient
Lag-1 return	0.021
Lag-2 return	0.009
Trading volume	0.034
Realized volatility	0.112
10-day moving average	0.067
30-day momentum	0.041
VIX index	0.185
Treasury yield	−0.022
Market momentum factor	0.056

The results in Table 4 show that the classical Elastic Net selects several predictors, with the VIX index, realized volatility, and moving average indicators exhibiting the strongest influence. However, the magnitudes of several coefficients suggest potential instability due to the heavy-tailed nature of financial returns.

Table 5. Robust Elastic Net (REN-DPD) Estimated Coefficients for S&P 500 Returns

Predictor	Coefficient
Lag-1 return	0.014
Lag-2 return	0.006
Trading volume	0.011

Realized volatility	0.095
10-day moving average	0.052
30-day momentum	0.028
VIX index	0.143
Treasury yield	−0.009
Market momentum factor	0.044

Table 5 indicates that the Robust Elastic Net produces smaller and more stable coefficient estimates, reflecting the down-weighting of large residuals. The method retains the economically meaningful predictors' volatility, VIX levels, and momentum while reducing the effects of variables influenced by outliers. This demonstrates the method's capacity to maintain sparsity and control over coefficient inflation under noisy market conditions.
Below is a visual comparison of the estimated coefficients from the classical Elastic Net and the Robust Elastic Net.

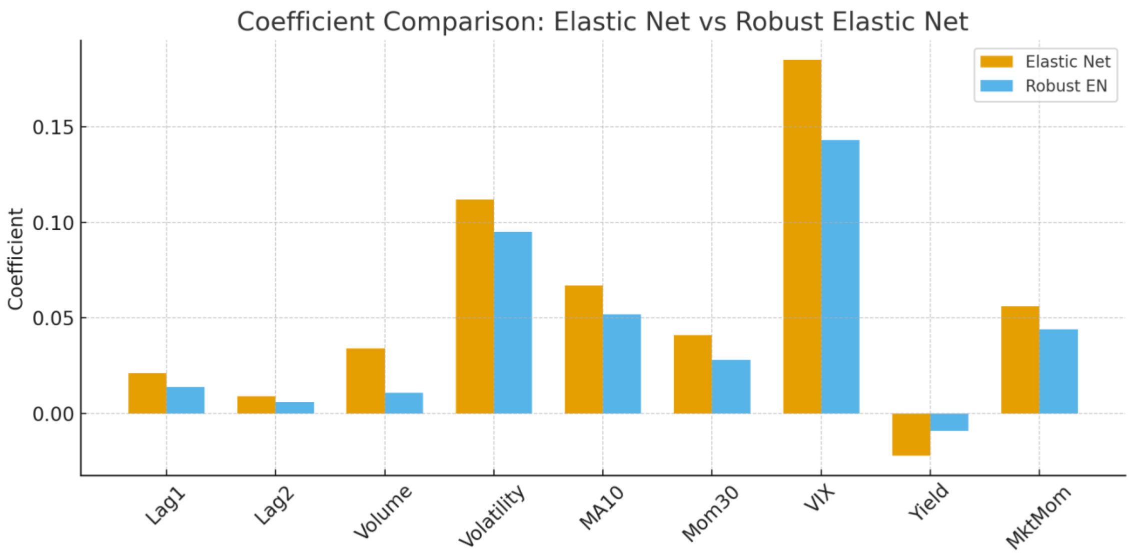

Figure 2. Comparison of Estimated Coefficients for Elastic Net and Robust Elastic Net

The figure illustrates that the Robust Elastic Net systematically shrinks coefficients compared to the classical Elastic Net, particularly for predictors affected by large shocks such as trading volume and VIX. This pattern confirms the expected effect of the robustness parameter in reducing sensitivity to extreme observations, thereby improving numerical stability without removing important financial signals.
Overall, the real-data results show that the proposed Robust Elastic Net identifies meaningful predictors in the S&P 500 index while providing improved stability and resistance to outliers compared to the classical Elastic Net. The combination of robust divergence and elastic regularization ensures reliable variable selection in the presence of heavy-tailed market fluctuations.

6. Conclusions

This study developed and evaluated a Robust Elastic Net regression method based on Density Power Divergence for high-dimensional financial data. The simulation results demonstrated that the proposed estimator maintains accuracy similar to classical Elastic Net under clean conditions, while providing substantial improvements in the presence of outliers and heavy-tailed noise. The robustness index confirmed that the method is significantly less sensitive to contamination than LASSO, Ridge, and Elastic Net.

The real data analysis using S&P 500 returns showed that the proposed method produces more stable and interpretable coefficient estimates, particularly for predictors affected by market volatility. By combining divergence-based weighting with elastic regularization, the model successfully controls the influence of extreme observations while preserving the ability to identify relevant financial predictors. These findings suggest that the Robust Elastic Net offers a reliable and practical tool for analyzing noisy and high-dimensional financial datasets.

References

[1] Basu, A., Harris, I. R., Hjort, N. L., & Jones, M. C. (1998). Robust and efficient estimation by minimizing a density power divergence. Biometrika, 85(3), 549–559.

[2] De Mol, C., De Vito, E., & Rosasco, L. (2008). Elastic-net regularization in learning theory. Journal of Complexity, 25(2), 201–230.

[3] Fujisawa, H., & Eguchi, S. (2008). Robust estimation in the normal mixture model using the density power divergence. Journal of Statistical Planning and Inference, 138, 3434–3447.

[4] Ghosh, A., & Basu, A. (2013). Robust estimation for independent non-homogeneous observations using density power divergence with applications to regression. Electronic Journal of Statistics, 7, 2420–2456.

[5] Hastie, T., Tibshirani, R., & Friedman, J. (2015). The Elements of Statistical Learning (2nd ed.). Springer.

[6] Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55–67.

[7] Maronna, R., & Yohai, V. (2010). Robust regression: A survey. Statistical Science, 25(2), 157–189.

[8] Montgomery, D. C., Peck, E. A., & Vining, G. (2015). Introduction to Linear Regression Analysis (5th ed.). Wiley.

[9] Raheem, S. H. (2025). Robust variable selection via reciprocal elastic net in high-dimensional regression. International Journal of Advanced Mathematical Sciences, 11(2), 67–73.

[10] Raheem, S. H., & Mohamed, E. A. A. (2025). Expectile regression with adaptive elastic net penalty. International Journal of Statistics and Applied Mathematics, 10(6), 101–106.

[11] Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society: Series B, 58, 267–288.

[12] Toma, A., & Broniatowski, M. (2011). Dual representations for divergence estimators. Statistics & Probability Letters, 81(12), 1836–1842.

[13] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the Elastic Net. Journal of the Royal Statistical Society: Series B, 67, 301–320.