Afro-Asian Journal of Scientific Research (AAJSR)
المجلة الأفروآسيوية للبحث العلمي
E-ISSN: 2959-6505
Volume 4, Issue 1, 2026
Page No: 226-241
Website: https://aajsr.com/index.php/aajsr/index
SJIFactor 2024: 5.028 ISI 2025: 0.915 معامل التأثير العربي (AIF) 2025: 0.76

The ISLasso Estimator for Multicollinearity and High-Dimensional Problems in Linear Regression with Continuous and Categorical Outcomes

Ibrahim Al-Barghathi[1], Omar A. El-Sharif[2], Abir Hassan Elgihawi[3], Abdelbaset Abdalla[4*], Ahmed M. Mami[5]
[1,2,3,4,5]Department of Statistics, Faculty of Science, University of Benghazi, Benghazi, Libya

مقدر *ISLasso* لمعالجة مشاكل التعدد الخطي والأبعاد العالية في الانحدار الخطي مع متغير الاستجابة مستمر وفئوي

إبراهيم بالقاسم البرغثي[1]، عمر أبو بكر الشريف[2]، عبير حسن الجحاوي[3]، عبد الباسط عبد الله العمروني[4*]، أحمد محمد مامي[5]
[1،2،3،4،5]قسم الاحصاء، كلية العلوم، جامعة بنغازي، بنغازي، ليبيا

*Corresponding author: abdelbaset.abdalla@uob.edu.ly

Received: December 25, 2025 Accepted: February 02, 2026 Published: February 09, 2026

Abstract:
This paper compares the efficiency of seven penalized regression estimators-Ridge, LASSO, Elastic Net (ENET), Adaptive LASSO (ALASSO), Adaptive Elastic Net (AENET), ISLasso, and MAENet-designed to improve predictive performance under conditions of multicollinearity and high dimensionality. By incorporating penalty terms into the ordinary least squares (OLS) framework, these methods reduce variance at the cost of introducing some bias, thereby shrinking coefficient estimates toward zero. The study evaluates estimator performance for both continuous and binary dependent variables using Mean Squared Error (MSE) and Mean Absolute Error (MAE). Analysis is based on simulated datasets with varying sample sizes, numbers of predictors, and correlation levels ($\rho = 0.0$ to 0.80), with 500 replications per simulation condition.

Keywords: ridge regression, LASSO regression, elastic net regression (ENET), adaptive Lasso regression (ALASSO), adaptive elastic net regression (AENET), ISLasso, MAENet, high-dimensional data, multicollinearity, penalized regression.

الملخص:
تقارن هذه الورقة البحثية كفاءة سبعة مُقدِّرات انحدار جزائية – ريدج Ridge، ولاسو Lasso، والشبكة المرنة (ENET)، ولاسو التكيفي (ALASSO)، والشبكة المرنة التكيفية (AENET)، وISLasso، وMAENet - المصممة لتحسين الأداء التنبؤي في ظل وجود مشكلة التعدد الخطي والأبعاد العالية. من خلال دمج حدود الجزاء في إطار المربعات الصغرى العادية (OLS)، تُقلل هذه الطرق التباين على حساب إدخال بعض التحيز، مما يُقلل تقديرات المعاملات نحو الصفر. تُقيِّم الدراسة أداء المُقدِّر لكل من المتغيرات التابعة المستمرة والثنائية باستخدام متوسط مربع الخطأ (MSE) ومتوسط الخطأ المطلق (MAE). يعتمد التحليل على مجموعات بيانات مُحاكاة بأحجام عينات وأعداد مُتنبئات ومستويات ارتباط مُتفاوتة ($\rho = 0.0$ إلى 0.80)، مع 500 تكرار لكل حالة محاكاة.

الكلمات المفتاحية: انحدار ريدج Ridge، انحدار لاسو Lasso، انحدار الشبكة المرنة (ENET)، انحدار لاسو التكيفي (ALASSO)، انحدار الشبكة المرنة التكيفي (AENET)، ISLasso، MAENet، البيانات عالية الأبعاد، التعدد الخطي، مقدرات الانحدار الجزائية.

Introduction:

The introduction should be typed in Arial with font size 10. Author can select Normal style setting from Styles of this template. The simplest way is to replace (copy-paste) the content with your own material. In this section, highlight the importance of topic, making general statements about the topic and presenting an overview on current research on the subject. Your introduction should clearly identify the subject area of interest.

Reference should be cited at appropriate point in the text by number(s) in square brackets in line with the text. e.g.: '.... was reported earlier [1, 2].'

The actual authors can be referred to, but the reference number(s) must always be given. e.g.: 'Barnaby and Jones [3] obtained a different....'

The goal is to identify a model that best fits the data while remaining simple and meaningful. Model fit is assessed using goodness-of-fit measures, and parsimony is achieved through effective model selection techniques.

Let Y denote the response variable (also called the endogenous variable) and denote the explanatory variables (also called exogenous variables). The relationship between the response variable and the explanatory variables can be expressed in vector form as Regression analysis is a fundamental statistical methodology for modeling the relationship between a response (dependent) variable and one or more predictor (independent) variables. Its primary objectives are to **understand** this underlying relationship and to estimate or predict future outcomes. The nature of the response variable, whether continuous or categorical, determines the specific modeling framework, influencing choices such as the link function and the criteria for assessing model fit. Consequently, careful model specification is essential for obtaining accurate and interpretable results.

The overarching aim is to identify a model that achieves an optimal balance: it must fit the observed data well while remaining parsimonious and substantively meaningful. This involves evaluating goodness-of-fit using appropriate metrics and applying disciplined model selection techniques to ensure simplicity without sacrificing explanatory power.

Formally, let **Y** represent the response variable (endogenous variable) and $\mathbf{X_1, X_2, ..., X_p}$ denote the explanatory variables (exogenous variables). The general relationship can be expressed in vector form as:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$
$$\mathbf{E}(\mathbf{Y}) = \mathbf{X\beta}$$

Where:

- **Y** is the $n\times 1$ response vector.
- **X** is the $n\times(p+1)$ design matrix, which includes a column of ones for the intercept.
- β is the $(p+1)\times 1$ parameter vector of regression coefficients.
- ε is the $n\times 1$ error vector, typically assumed to follow $\varepsilon \sim N(0,\sigma^2 I_n)$.

Regression analysis is a foundational statistical tool for modelling relationships between a response variable and a set of predictors. The ordinary least squares (OLS) estimator is the most common approach due to its simplicity and desirable properties under ideal conditions. However, OLS relies on several critical assumptions, linearity, homoscedasticity, normality of errors, and no multicollinearity, to produce unbiased, consistent, and efficient estimates. Violations of these assumptions can lead to biased inferences, unstable predictions, and ultimately, unreliable conclusions.

Two particularly common and challenging violations in applied settings are multicollinearity and the curse of dimensionality. Multicollinearity occurs when predictors are highly correlated, leading to inflated variances of coefficient estimates and complicating the interpretation of individual effects. Mathematically, it results in a near-singular design matrix (X^TX), causing instability in its inverse and, consequently, in the OLS estimates. Indicators include high variance inflation factors (VIFs) and substantial sensitivity of coefficients to minor data changes. In parallel, the curse of dimensionality emerges in high-dimensional data (where the number of predictors p approaches or exceeds the sample size n), manifesting as overfitting, poor model generalization, and increased computational complexity.

To address these limitations, a class of penalized regression methods has been developed by introducing a regularization term into the OLS loss function to shrink coefficient estimates toward zero. This controlled introduction of bias often yields a favorable trade-off by significantly reducing variance and enhancing model stability. Seminal contributions include ridge regression (Hoerl & Kennard, 1970), which applies an L2-penalty to handle multicollinearity; the LASSO (Tibshirani, 1996), which uses an L1

-penalty to achieve sparse variable selection; and the elastic net (Zou & Hastie, 2005), which combines both penalties to leverage their respective advantages. Subsequent extensions, such as the adaptive LASSO, adaptive elastic net, and more recent variants like the induced smoothed LASSO (ISLasso) and multi-stage adaptive elastic net (MAENet), have further refined the balance between selection consistency, estimation accuracy, and predictive performance, especially in high-dimensional regimes.

Despite the proliferation of these methods, a comprehensive comparative analysis of their performance under controlled conditions of multicollinearity and dimensionality, particularly for both continuous and binary response variables, remains a pertinent research gap. Most simulation studies focus on a subset of these estimators or on a single type of outcome, leaving practitioners without clear guidance for method selection across varied data scenarios.

To bridge this gap, this paper conducts an extensive simulation study to evaluate and compare the performance of seven prominent penalized regression estimators: Ridge, LASSO, Adaptive LASSO, Induced Smoothed LASSO (ISLasso), Elastic Net, Adaptive Elastic Net, and Multi-stage Adaptive Elastic Net (MAENet). We systematically vary key data characteristics, including sample size, number of predictors, and degree of correlation among them. Performance is assessed using Mean Squared Error (MSE) and Mean Absolute Error (MAE) for both continuous and binary outcomes. The findings aim to provide a clear, empirical basis for selecting an appropriate penalized regression method tailored to specific data challenges in predictive modelling.

Material and methods:

Regularization addresses two core challenges in regression: overfitting from excessively large coefficients and instability from collinearity among predictors (Hoerl & Kennard, 1970). In the linear model $Y = X\beta + \varepsilon$, where X is the $n \times p$ design matrix, traditional least squares estimates can become highly variable when predictors are correlated. This instability is exacerbated in high-dimensional settings (p approaching or exceeding n), increasing the risk of models that fit noise rather than the underlying signal.

Penalized regression methods introduce a constraint on the magnitude of the coefficient vector **β**. By shrinking coefficients toward zero, these methods introduce a small bias to achieve a substantial reduction in variance, thereby stabilizing predictions and improving generalization performance. This study implements and compares seven such estimators, each defined by a specific penalty function added to the ordinary least squares loss.

Controlling Regression Coefficients:

There are two popular penalties usually used to regularize the regression coefficients:

1. **L_1 Penalty (LASSO):** The regression coefficients are constrained to $\sum_{j=1}^{p}|\beta_j| \leq C$ **.** Some coefficients become exactly zero, effectively performing variable selection. This leads to the LASSO (Least Absolute Shrinkage and Selection Operator) estimator. LASSO is often used to identify a subset of important predictors (Tibshirani, 1996). It can be used instead of **stepwise regression** or **best-subsets selection.**
2. **L_2 Penalty (Ridge Regression): The** regression coefficients **are constrained to** $\sum_{j=1}^{p}\beta_j^2 \leq C$. Shrinks coefficients **continuously toward zero** but **never exactly to zero**. Ridge Regression is particularly useful when predictors are highly collinear, making it effective for datasets with many correlated variables. Its main goal is to improve prediction accuracy rather than perform variable selection. Both Ridge and related shrinkage methods constrain the size of coefficients through a hyperparameter (smaller values enforce stronger shrinkage). This introduces a small bias in exchange for a significant reduction in variance, exemplifying the bias–variance tradeoff. While Ordinary Least Squares (OLS) estimates have low bias but can exhibit high variance under multicollinearity or many predictors, shrinkage methods enhance model stability, generalization, and interpretability by controlling coefficient magnitudes.

Introduction to Ridge Regression:

Multicollinearity occurs when two or more explanatory variables in a regression model are **highly correlated**. This causes the **matrix** $\boldsymbol{X^tX}$to be **close to singular** (i.e., not invertible).

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{OLS}} = (\boldsymbol{X^tX})^{-1}\boldsymbol{X^t}\,\boldsymbol{Y}$$

As a result, the **OLS estimator becomes unstable.** The **small changes** in data lead to **large changes** in $\boldsymbol{\beta\hat{}\_OLS}$, and the **variance of the estimators becomes considerably larger**.

Ridge regression **modifies** the OLS solution by adding a **penalty term** λI to the matrix $\boldsymbol{X^tX}$, thereby improving **numerical stability**.

$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{Ridge}} = (\boldsymbol{X^tX} + \lambda\mathrm{I})^{-1}\boldsymbol{X^t}\,\boldsymbol{Y}$$

Where λ > 0 is the **ridge penalty (tuning parameter),** and I is the identity matrix.

The immediate Effect is that $X^{\wedge}t\,X + \lambda I$ is guaranteed to be **nonsingular** for any λ > 0. The estimator is **more stable,** and the **variance is reduced**, though a **small bias** is introduced.
The **MSE** of an estimator $\hat{\boldsymbol{\beta}}$ is given by:

$$\text{MSE}(\hat{\beta}) = \text{E}||\hat{\beta} - \beta||^2 \qquad (2.1)$$

Remember that the MSE is a commonly used measure for assessing the quality of estimation, which consists of two parts: the squared bias and the variance, and can be written in the following form:

$$E\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|^2 = \sum_j E(b_j - \beta_j)^2 = \sum_j \{E(b_j) - \beta_j\}^2 + \sum_j Var(b_j) \qquad (2.2)$$

OLS has zero bias, but potentially **high variance** under multicollinearity.

The ridge estimator achieves this stabilization through the **bias–variance trade-off**. By introducing a controlled bias via the penalty parameter λ, it sharply reduces the variance of the coefficient estimates. This often results in a lower overall mean squared error (MSE) and superior out-of-sample prediction accuracy compared to the ordinary least squares estimator (Baltagi, 2001; Weisberg, 2005). Formally, ridge regression can be framed as a constrained least squares optimization problem, yielding a biased estimator with substantially reduced sampling variance.

Regression in High Dimensions:

High-dimensional regression, where the number of predictors p*p* meets or exceeds the sample size n*n*, presents a distinct inferential challenge. The ordinary least squares (OLS) estimator is undefined when *p*>*n* and becomes highly unstable as *p* approaches *n*, a scenario that invalidates classical statistical assumptions and necessitates specialized methods. This study focuses on **regularization** as a primary strategy, which addresses high-dimensionality by imposing structural constraints (e.g., sparsity or coefficient shrinkage) to enable stable estimation and prediction.

Theoretical Development of Ridge Shrinkage Estimator:

Ridge regression was first proposed by Hoerl & Kennard (1970) as a solution to the problems caused by multicollinearity among predictor variables. It serves as an alternative to the ordinary least squares (OLS) method in such cases. For any least square estimator *b*, the least squares objective function can be reformulated, reaching its minimum at the ordinary least squares estimate $\|\hat{\beta}^{LS}\|$. This leads to a quadratic expression in *b*.

$$\text{Q(b)} = \|y - X\hat{\beta}^{LS} + X\hat{\beta}^{LS} - Xb\|^2$$
$$= (y - X\hat{\beta}^{LS})^t(y - X\hat{\beta}^{LS}) + (\text{b} - \hat{\beta}^{LS})^t X^t X(\text{b} - \hat{\beta}^{LS}) \qquad (2.5)$$
$$= Q_{min} + \phi(b)$$

The level curves (contours) of the quadratic function Q(b) form hyper-ellipsoids, all centered around the ordinary least squares estimate $\hat{\beta}^{\wedge}LS$. Based on equation (2.5), it is reasonable to expect that any deviation from the point of minimum Q_min tends to occur in a direction that reduces the magnitude (length) of b. In Ridge Shrinkage Regression, the corresponding optimization problem is defined as:

$$\text{minimizing } \|\beta\|^2 \text{subject to } (\beta - \hat{\beta}^{LS})^t X^t X(\beta - \hat{\beta}^{LS}) = \phi_0 \qquad (2.6)$$

for a given constant $\phi_0$this constraint ensures that the residual sum of squares Q(β) remains relatively close to its minimum value Q_min. Figure 2.1 illustrates this by showing the contour lines of the residual sum of squares along with the L_2 ridge shrinkage constraint in a two-dimensional setting (Kotz and Nadarajah, 2004).

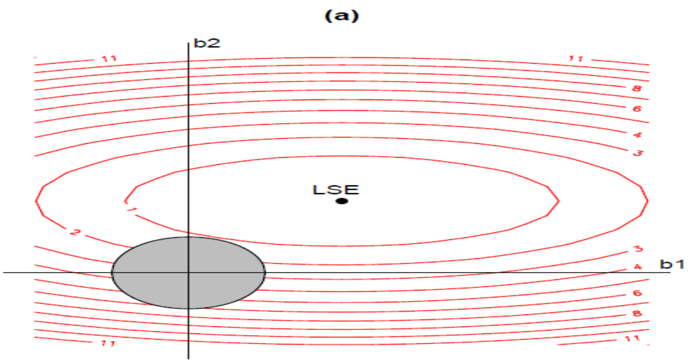

Figure (1): Contours of the Sum of Squares of the Residual and the Constraint Functions in Ridge Shrinkage Regression.

LASSO Regression:

In high-dimensional settings, a **sparsity assumption**, that only a subset of predictors truly influences the response, is often critical. This makes variable selection essential. Ridge regression, while stabilizing estimates, does not perform selection, as its L_2 penalty shrinks all coefficients without setting any to zero. The LASSO overcomes this by substituting an L_1 penalty in its objective function:

$$\mathrm{PLS}_{LASSO}\ (\beta_1,\lambda_1) = \mathrm{OLS}(\boldsymbol{\beta})+\lambda_1 P_1(\boldsymbol{\beta})$$
$$=\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|_2^2+\lambda_1\|\boldsymbol{\beta}\|_1$$
$$=\sum(yi-\beta_0-\sum \beta_j x_{ij}\ )^2\ +\lambda 1\sum|\beta j| \qquad (2.7)$$

Or:

$$\beta_{LASSO} = \text{argmin}\ \boldsymbol{\beta}\in\mathbb{R}\ p\ [\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\|^2 + \lambda 1\|\boldsymbol{\beta}\|] \qquad (2.8)$$

In the LASSO objective function, the second term $\lambda_1\|\beta\|$ = $\lambda_1\sum_{j=1}^{p}|\beta j|$ serves as a penalty for coefficient size, with the tuning parameter $\lambda_1 > 0$ controlling the degree of shrinkage. This parameter plays a critical role in determining the behavior of the LASSO estimator. When λ_1 is close to zero, the LASSO estimate $\hat{\beta}$ Lasso (λ_1) closely resembles the Ordinary Least Squares (OLS) solution. On the other hand, as λ_1 becomes very large, the estimator approaches a zero vector (0p′), effectively driving all coefficients to zero.

LASSO differs from Ridge Regression in how it shrinks coefficients: while Ridge reduces all coefficients without eliminating any, LASSO can shrink some coefficients exactly to zero as the regularization parameter increases. This allows simultaneous variable selection and shrinkage, giving LASSO its name ("Least Absolute Shrinkage and Selection Operator"). Geometrically, LASSO's L1-norm constraint forms a diamond-shaped region, and the residual sum of squares contours often intersect at vertices, producing zero coefficients. In contrast, Ridge's L2-norm constraint is circular, rarely yielding zero coefficients. In higher dimensions, LASSO's polyhedral constraint increases the likelihood of sparse models, making this a key advantage for variable selection (Tibshirani, 1996; Hastie et al., 2015; Efron and Hastie, 2021).

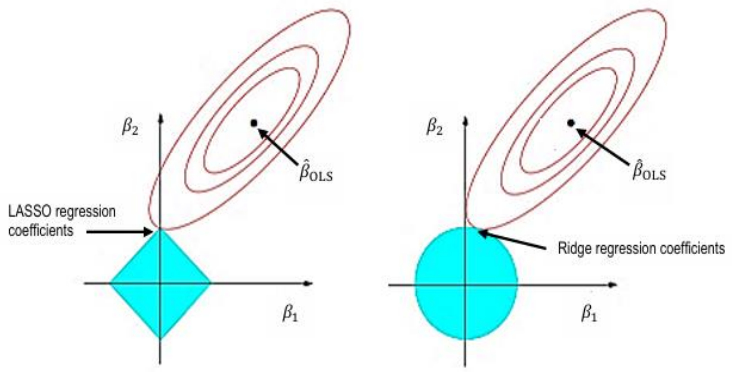

Figure (2): Estimation picture for the L_1-norm on the left and L_2-norm on the right (Hastie et al., 2015).

The Adaptive Lasso:

The Adaptive Lasso (ALasso), introduced by Zou (2006), improves upon traditional Lasso by enhancing variable selection accuracy and reducing bias in coefficient estimation. Unlike standard Lasso, which applies a uniform penalty, ALasso uses data-driven, weighted L1 penalties based on initial estimates (often from OLS), where stronger predictors receive lighter penalties and weaker ones are heavily penalized. This selective regularization allows ALasso to achieve the "oracle property," meaning it can asymptotically identify the correct subset of relevant predictors and estimate their coefficients as accurately as if the true model were known. Consequently, Adaptive Lasso offers both sparsity and precision, making it highly effective for high-dimensional modelling.

Computational Implementation Details:

The application of the Adaptive Lasso in this research follows a structured, two-stage procedure. In the First stage, an initial consistent estimate of the regression coefficients, denoted as $\hat{\beta}$ initial, is obtained through (Specify initial estimation method, e.g., Ordinary Least Squares (OLS) regression or Ridge regression).

These initial estimates are then used to construct the adaptive weights, $\hat{\omega}_j = \left(|\hat{\beta}\ initial, j|\right)^{-\gamma}$ for each predictor j. Following Zou (2006), a common choice for γ is typically 0.5, 1, or 2, with (chosen value for gamma, e.g., γ=1) being adopted for its balance between performance and computational stability.

In the Second stage, the Adaptive Lasso model is fitted by minimizing the weighted. $L1$ penalized objective function:

$$L(\beta) = \sum_{i=1}^{n}(y_i - x_i'\beta)^2 + \lambda\sum_{i=1}^{n} w_j|\beta_j|$$

This optimization problem is a convex optimization problem, meaning it does not suffer from multiple local minima and can be efficiently solved to obtain the global minimizer. The crucial regularization parameter, λ, will be rigorously tuned through (Specify the cross-validation method, e.g., 10-fold cross-validation) to identify the value that optimizes predictive performance and model generalizability.

Elastic Net Estimator:

To address the issue of double shrinkage present in Naïve Elastic Net (NENET) regression, Zou and Hastie (2005) introduced the **corrected Elastic Net (ENET)** estimator. This estimator is essentially a rescaled version of the Naïve Elastic Net solution. To construct it, starting with the original dataset (y, X) and two regularization parameters λ_1 for sparsity and λ_2 for grouping, similar to the Naïve Elastic Net, a modified or **augmented dataset** (y^*, X^*) is created using the following approach:

$$X_{(n+p)*p} = 1\frac{1}{\sqrt{1+\lambda_2}}(X\frac{X}{\sqrt{\lambda_2}}I),\ y_{(n+p)}=\binom{y}{0}$$

Let y = $\lambda_1/\sqrt{1+\lambda_2}$ and $\beta^* = \sqrt{1+\lambda_2}\beta$ Then the penalized least squares (PLS) of the Naïve Elastic Net (NENET) converts to a LASSO-type problem

$$\text{PLS}(\beta^*, Y) = \text{OLS}(\beta^*) + YP_1(\beta^*)$$
$$= \|Y^* - X^*\beta^*\|_2^2 + \text{Y}\|\beta^*\|_1$$

This means that the Naïve Elastic Net solves a LASSO-type problem.

$$\beta\hat{\beta}^*_{LASSO} = \arg \underbrace{min}_{\beta^*\in\mathbb{R}^p}\left(\|y^* - X^*\beta^*\|_2^2 + \frac{\lambda_1}{\sqrt{1+\lambda_2}}\|\beta^*\|_1 \right)$$

According to Zou and Hastie (2005), the Elastic Net (corrected) estimates **β**̂EN are defines by

$$\hat{\beta}_{ENET} = \sqrt{1+\lambda_2}\,\hat{\beta}^*_{LASSO}.$$

Moreover, the Naïve Elastic Net vector of estimators is defined by:

$$\hat{\beta}_{NENET} = \left(1/\sqrt{1+\lambda_2}\right)\hat{\beta}^*_{LASSO}$$

Thus, the vanilla Elastic Net vector of estimators is given by:

$$\hat{\beta}\_ENET = (1+\lambda\_2)\ \hat{\beta}\_NENET = (1+1/2\,\lambda(1-\alpha))\,\hat{\beta}\_NENET$$

The Elastic Net improves upon the Naïve Elastic Net by rescaling its coefficients, which corrects excessive shrinkage while preserving the ability to select variables and group correlated predictors. Compared to LASSO and Ridge, Elastic Net performs strongly, addressing LASSO's limitations and improving variable selection efficiency and overall predictive performance.

Elastic Net Estimator Advantages:

Elastic Net is an advanced regularization method tailored for analyzing high-dimensional datasets, offering a more effective alternative to traditional LASSO and Ridge regression techniques. Its key strength lies in the combination of the advantages of both methods, effectively overcoming their individual shortcomings to create a more flexible and reliable modelling approach.

At its core, Elastic Net modifies the standard regression objective by incorporating a penalty term. This term promotes a compromise between accurately fitting the training data and maintaining smaller coefficient values, which helps prevent overfitting and enhances the model's ability to generalize to unseen data.

Elastic Net combines two penalties to enhance model performance:

1. **LASSO (L_1)** promotes sparsity by shrinking some coefficients to zero, enabling variable selection and simpler models. However, it struggles with highly correlated predictors, often selecting only one and discarding the rest.
2. **Ridge (L_2)** handles multicollinearity by shrinking all coefficients toward zero without eliminating any. It retains all correlated variables, improving model stability when predictors are interrelated.

Elastic Net combines the strengths of LASSO and Ridge regression through two key parameters:

- **Lambda (λ):** Controls the overall level of regularization, higher values lead to greater shrinkage and simpler models.
- **Alpha (α):** Determines the balance between L_1 and L_2 penalties:
  - **α = 1:** Equivalent to LASSO, emphasizing variable selection.
  - **α = 0:** Equivalent to Ridge, focusing on multicollinearity.
  - **0 < α < 1:** Balances both effects, selects key variables while also retaining correlated predictors.

This interaction makes Elastic Net especially useful for high-dimensional data with correlated features.

Adaptive Elastic Net Estimator:

In high-dimensional data scenarios, where the number of parameters increases with the sample size, selecting and estimating models becomes increasingly difficult. An ideal solution should possess the **oracle property** (as discussed in *JASA, 2001* and *Annals of Statistics, 2004*), ensuring optimal performance as data scales. Additionally, high dimensionality often leads to **collinearity**, which any effective method must address. Most existing variable selection techniques struggle to satisfy both of these essential criteria. To overcome this, the adaptive elastic-net is proposed. This approach cleverly merges quadratic (Ridge-style) regularization with LASSO shrinkage that uses adaptive weights. The method is shown to satisfy the oracle property under broad conditions and, based on simulation studies, demonstrates superior handling of collinearity compared to other oracle-based methods, which makes it particularly effective even when sample sizes are limited.

The adaptive elastic-net addresses this by combining quadratic regularization with adaptively weighted LASSO. It is proven to meet the oracle property under general conditions and shows superior performance in managing collinearity, particularly in small sample sizes.

Zou (2006) proposed the following adaptive lasso estimator.

$$\hat{\beta}_{(AdaLasso)} = \underbrace{\arg min}_{\beta} \; \|y - X\beta\|_2^2 + \lambda \textstyle\sum_{j=1}^{P} \hat{\omega}_j \left|\beta_j\right|.$$

where $\{(wj\}_{i=1}^{p}$ are the adaptive data-driven weights, and can be computed by wj =(| $\widehat{\beta_j^{ini}}$ |), where γ is a positive constant, and βini is an initial root-n consistent estimate of β.

Zou (2006) demonstrated that the adaptive LASSO can perform as well as the oracle when the regularization parameter (λ) is properly chosen. Building on this, Candes et al. (2008) applied adaptive LASSO concepts to enhance sparsity in signal recovery through reweighted L_1 minimization. However, L_1-based methods like LASSO often perform poorly in the presence of highly correlated predictors**,** a common issue in high-dimensional settings. Even when variables are independently generated, high dimensionality can lead to large sample correlations (Fan and Lv, 2008).

The elastic-net estimator is defined as follows:

$$\hat{\beta}_{(enet)} = \left(1 + \frac{\lambda_2}{n}\right)\left\{\underbrace{\arg min}_{\beta} \; \|y - X\beta\|_2^2 + \lambda_2\|\beta\|_2^2 + \lambda_1\|\beta\|_1\right\}.$$

When predictors are standardized (mean zero and L_2-norm one), the regularization adjustment changes from 1 +λ_1 to 1+λ_2, as noted by Zou and Hastie (2005). In Elastic Net, the L_1 component enables variable selection, while the L_2 component stabilizes solution paths, enhancing prediction accuracy (Donoho et al., 1995).

The adaptive elastic-net combines two improvements over LASSO: the adaptive LASSO, which achieves the oracle property**,** and the elastic-net, which handles collinearity. However, the adaptive LASSO suffers from instability in high-dimensional settings, while the elastic-net does not satisfy the oracle property (Zou and Hastie, 2005), (Zou, 2006).

The Induced Smoothed Lasso (ISLasso) Estimator:

Although LASSO is effective for prediction and variable selection, it has significant limitations in statistical inference and hypothesis testing due to the non-smooth L_1 penalty. This causes several problems:

- Difficulty in computing reliable standard errors
- Inability to derive accurate p-values
- Zero standard errors for coefficients shrunk to zero, preventing uncertainty estimation

As a result, LASSO is less suitable for analyzing variable significance and making formal inferences, limiting its use beyond predictive tasks.

The Induced Smoothed Lasso (IS-Lasso) Framework:

To address the inferential limitations of traditional Lasso, the Induced Smoothed Lasso (ISLasso) framework introduces a novel method that facilitates dependable hypothesis testing in high-dimensional regression contexts. At its core, ISLasso substitutes the non-differentiable absolute value function in the Lasso penalty with a strategically designed smooth approximation. This smoothing is deliberately derived using a defined statistical mechanism, rather than being applied arbitrarily.

ISLasso accomplishes this by formulating a new estimating equation. This equation is derived by averaging the non-smooth score function, originally from the standard Lasso objective, across scaled normal perturbations of the parameters. This averaging process effectively smooths the sharp edges of the L_1 penalty, resulting in a smooth and differentiable estimating function. The ISLasso algorithm generally follows an iterative approach, in which the smooth estimating function and its associated covariance matrix are alternately updated until a specified convergence condition is satisfied.

Advantages and Key Features of ISLasso Estimator:

The Induced Smoothed Lasso (ISLasso) enhances traditional Lasso for high-dimensional regression by enabling statistically valid inference. Its main advantages include:

- Providing reliable p-values for all coefficients for rigorous significance testing.
- Delivering superior inferential performance with more accurate confidence intervals.
- Achieving computational efficiency, converging with few Newton-Raphson iterations.
- Maintaining asymptotic equivalence to the standard Lasso for consistency as sample size grows.

Overall, ISLasso is a robust, efficient tool for valid statistical inference in high-dimensional settings. (Zhu et al. , 2021).

Multi-Stage Adaptive Elastic Net (MSAEnet) Estimator:

Regularization methods like Lasso and Elastic-Net can generate models that capture true signals but also include many irrelevant variables, which is especially problematic in applications like biomarker discovery, where misclassification can lead to costly follow-up studies.

The Multi-Step Adaptive Elastic-Net (MSAEnet) Procedure:

To effectively minimize the number of false-positive variables while preserving estimation accuracy, **Xiao and Xu (2015)** introduced the **Multi-step Adaptive Elastic-Net (MSAEnet).** This procedure extends the adaptive Elastic Net by employing a **multi-stage iterative procedure**. At each stage, the regularization is re-applied using updated information from the previous step, progressively refining both variable selection and coefficient estimation.

Methodology:

The **MSAEnet method** operates through the following **iterative process**:

1. **Initialization:** Set the initial adaptive weights w_j for all p predictor variables to 1, i.e., w_j≡1 for all j=1,2,...,p.
2. **Iterative Estimation (for** k=1,2,...,M **steps):**

- At each iteration k, the method performs an **adaptive Elastic-Net regression.**
- This involves minimizing an objective function that combines both the **L_2 (Ridge)** and **weighted L_1 (Lasso)** penalties.
- The L_1 penalty term uses the adaptive weights $w_j(k-1)$ computed from the previous iteration, allowing the model to adjust the shrinkage applied to each variable dynamically.
- The regularization function at step k is therefore tailored using updated information, refining both variable selection and estimation accuracy at each stage.

$$\lambda_2^{*(k)}\|\beta\|_2^2 + \lambda_1^{*(k)} \sum_{j=1}^{p} w_j^{(k-1)} |\beta_j|.$$

The regularization parameters, $\lambda_2^{*(k)}$ and , $\lambda_1^{*(k)}$, are determined by cross-validation to achieve prediction optimality for that specific step. The resulting estimator from this step is denoted as $\beta^{(k)}$

- The adaptive weights are then updated for the next iteration using the formula:

$$w_j^{(k)} = \beta_j^{(k-1)}$$

In other words, variables that had smaller (but non-zero) coefficient estimates in the previous iteration are assigned **larger weights** in the current step. This results in **stronger penalization**, increasing the likelihood that these variables will be removed in subsequent iterations.

- Importantly, when k=1 (i.e., only one iteration), the procedure simplifies to the **standard Elastic-Net**.
- When k=2 (two iterations), it becomes equivalent to the **adaptive Elastic-Net**.

The regularization parameters $\lambda_1^{(k)}$ and $\lambda_2^{(k)}$ can also be reparametrized as $\lambda^{*(k)}$ and α(k), where α(k) determines the balance between the L_1 and L_2 penalties (Xiao and Xu, 2015).

Performance and Benefits of MSAEnet:

MSAEnet demonstrates strong performance in high-dimensional regression, consistently outperforming other regularization methods. Key benefits include:

- **Reduction of false positives:** Effectively eliminates irrelevant variables, even with high predictor correlation, crucial in costly validation settings like biomarker discovery.
- **High estimation accuracy:** Maintains or improves predictive performance compared to Adaptive Lasso and Adaptive Elastic Net through iterative refinement.
- **Handling multicollinearity:** Incorporates an L_2 (Ridge) component to retain correlated, important variables and support grouping effects.
- **Improved stability:** Shows consistent variable selection with lower variability in false-positive counts, especially under low correlation.

Overall, MSAEnet provides a robust balance of false-positive control, predictive accuracy, and interpretability for high-dimensional data. (Xiao and Xu, 2015).

The Simulation Study:

Two simulation studies were conducted for comparison, with Ridge included mainly for reference. All analyses were performed in R.

Description of The Experiment:

This simulation study compares seven penalized regression estimators under varying predictor correlations (ρ = 0.00, 0.25, 0.50, 0.75) to assess their effectiveness in handling multicollinearity. Two scenarios are examined:

- **Setting 1:** Data generated from a Gaussian distribution (Xiao & Xu, 2015).
- **Setting 2:** Data generated for logistic (binomial) regression.

The number of predictors is set at p = 60, 100, 250, with corresponding sample sizes:

- p=60 → n=30, 40, 50
- p=100 → n=25, 50, 75
- p=250 → n=100, 150, 200

These configurations provide a systematic framework for evaluating estimator performance under high-dimensional conditions.

The Numerical Summary and Graphics of Setting 1:

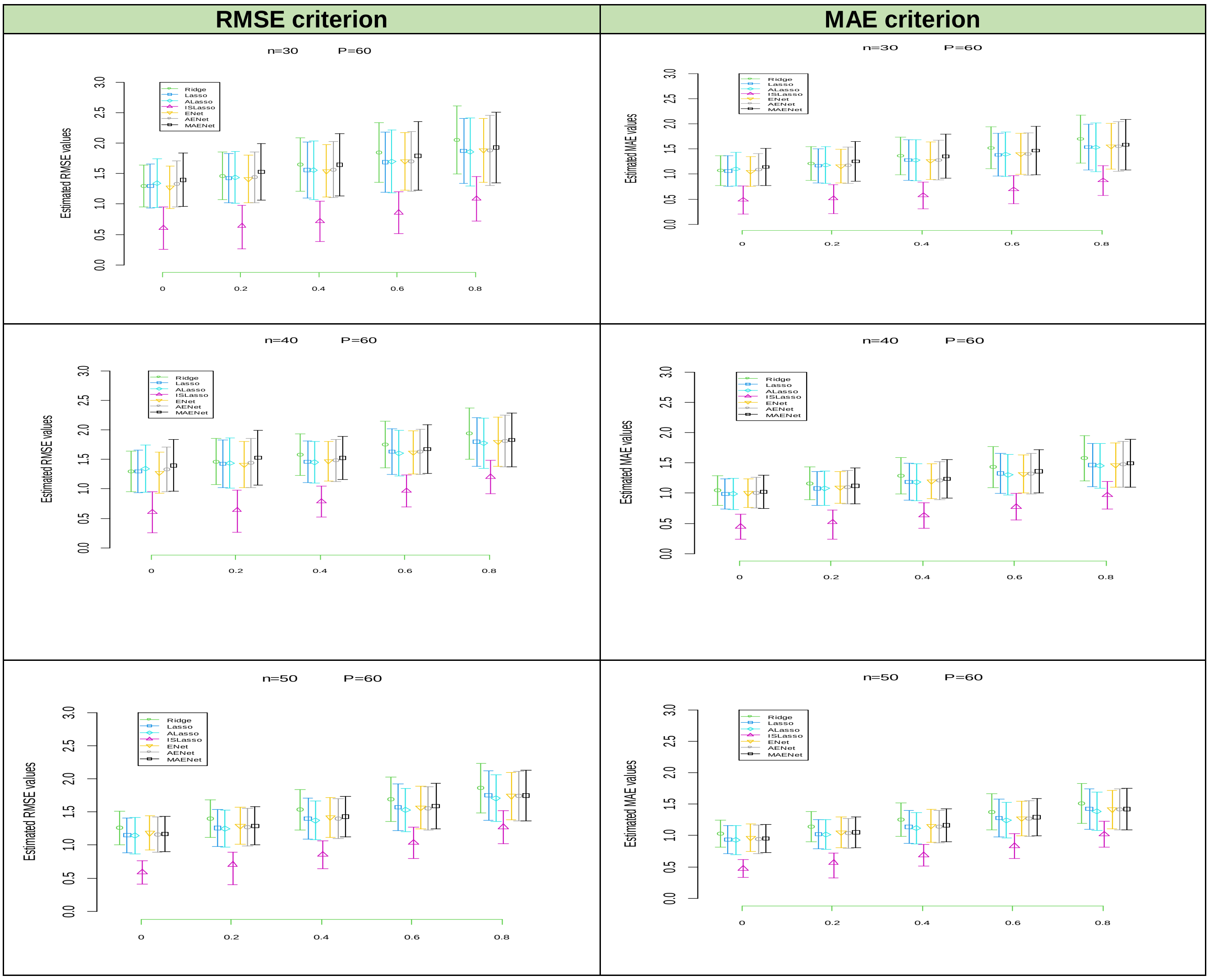

Figure (3): The Box plots demonstrate how various choices of correlation affect the mean (standard deviation) of the RMSE and MAE values for the seven proposed estimators for p= **60**in setting 1.

From Figure 3, we reach the following remarks:

- As the correlation among predictors increases, both RMSE and MAE rise in all models.
- **ISLasso performance:** ISLasso consistently gives the lowest RMSE and MAE across all scenarios, making it the most efficient method when correlations are weak or moderate.
- **When correlations increase:** ISLasso's efficiency declines, and **Lasso** or **Adaptive Lasso (ALasso)** become better alternative.
- **Model rankings:**

- n = 30: ISLasso best, followed by Elastic Net at low correlations; Lasso and ALasso better at high correlations. Ridge and MAENet perform the worst under strong correlation.
- n = 40: ISLasso best; ALasso generally second, but Lasso second when there is no correlation. ALasso is closest in relative efficiency when correlation exists.
- n = 50: ISLasso best, followed by ALasso; Ridge performs worst. ALasso is consistently closest to ISLasso across all correlation levels.

Overall, **ISLasso is the most efficient estimator**, with **Adaptive Lasso** (and sometimes Lasso) as the next best choice, while **Ridge and MAENet** are the least effective under strong correlation.

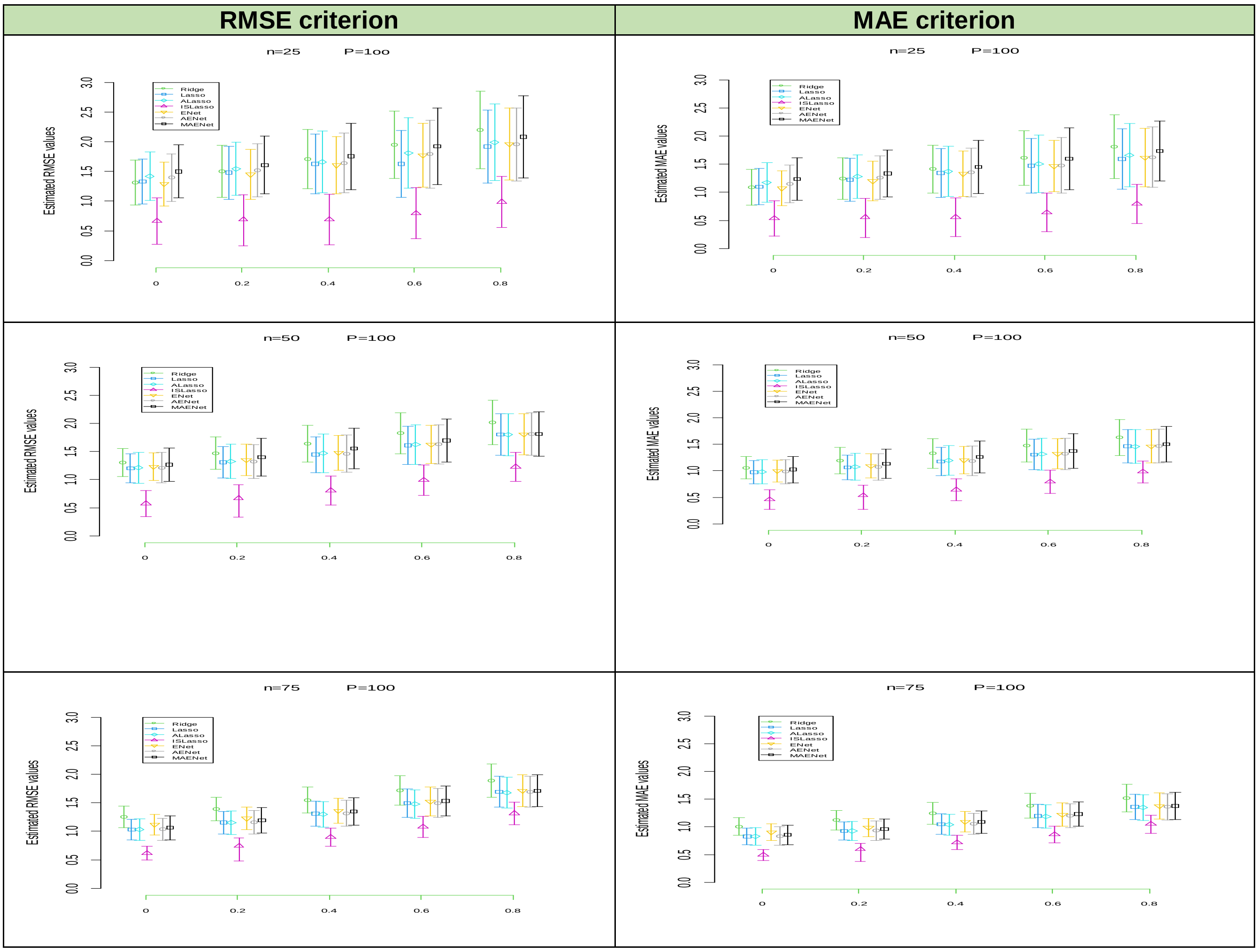

Figure 4: The Box plots demonstrate how various choices of correlation affect the mean (standard deviation) of the RMSE and MAE values for the seven proposed estimators for p=100 in setting 1.

From Figure 4, we reach the following remarks:

- **General trend:** Increasing correlation among predictors raises RMSE and MAE across all models. ISLasso consistently provides the lowest error values, making it the most efficient in all cases.
- **n = 25:**
  - ISLasso best overall.
  - Elastic Net second when correlation is absent/weak.
  - Lasso outperforms Elastic Net under strong correlation.
  - Relative efficiency: Elastic Net is closest to ISLasso unless correlation is very strong, in which case Lasso is closer. Ridge is competitive only when no correlation exists.
- **n = 50:**
  - ISLasso best overall.
  - Lasso is generally second-best, except under strong correlation, where Adaptive Lasso slightly outperforms it.
  - Relative efficiency: Lasso is closest to ISLasso in most cases; Elastic Net becomes closer under strong correlation.
- **n = 75:**

- ISLasso best overall.
- Lasso second when no correlation exists.
- Adaptive Lasso becomes the best second option whenever correlation is present (regardless of strength).
- Relative efficiency: Lasso is closest to ISLasso with no correlation; Adaptive Lasso is closest under correlated settings.

Overall, **ISLasso dominates across all scenarios**, while **Elastic Net, Lasso, or Adaptive Lasso** emerge as the second-best depending on correlation strength and sample size.

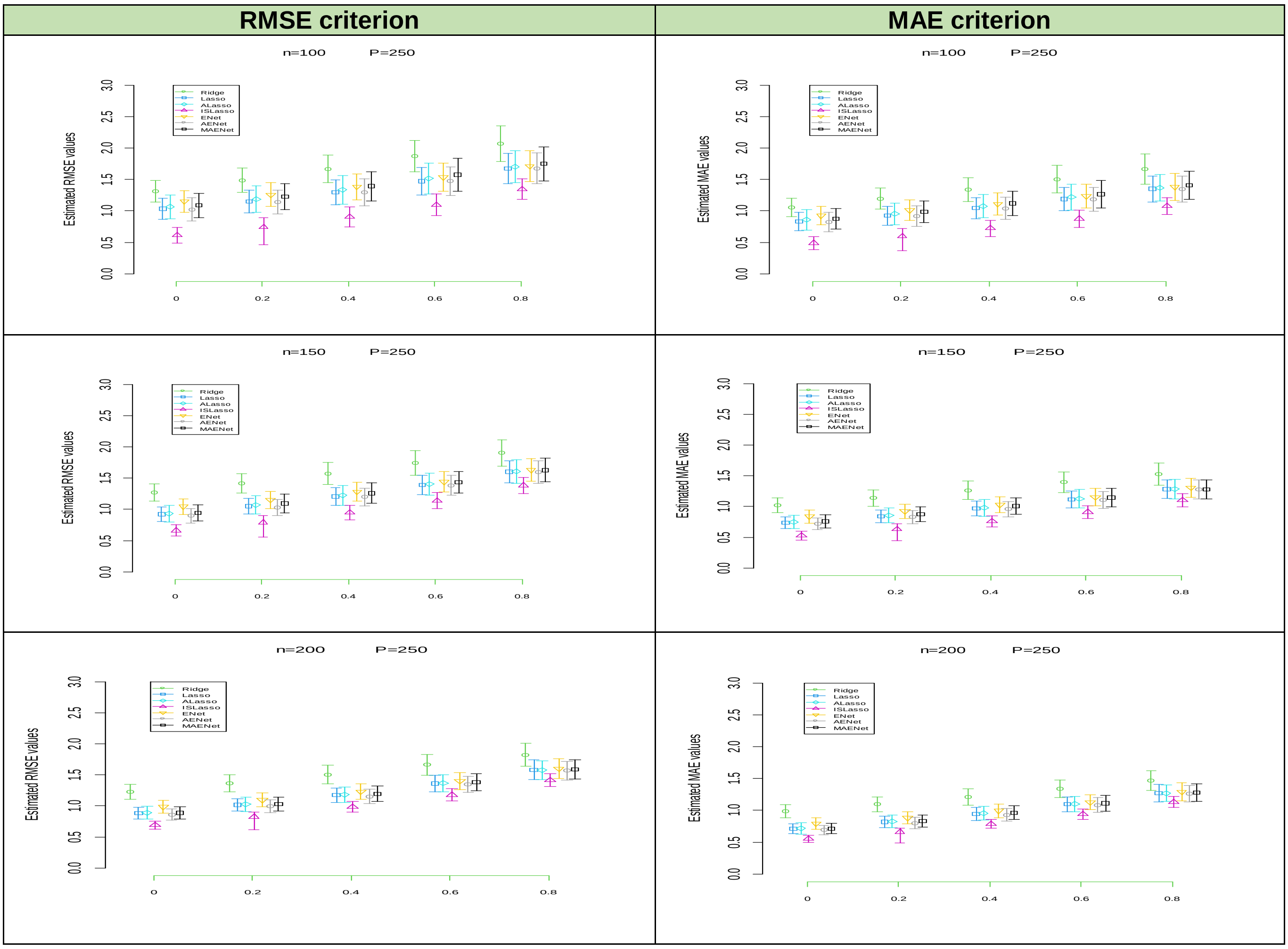

Figure 5: The Box plots demonstrate how various choices of correlation affect the mean (standard deviation) of the RMSE and MAE values for the seven proposed estimators for p=250 in setting 1.

From Figure 5, we reach the following remarks:

- **General trend:** Across all scenarios, RMSE and MAE increase as predictor correlations strengthen. ISLasso consistently achieves the lowest error values, making it the most efficient estimator.
- **n = 100:**
  - ISLasso best overall.
  - Adaptive Elastic Net (AENet) is second-best when correlation is weak or absent.
  - Lasso becomes the second-best under strong correlation.
  - Relative efficiency: AENet is closest to ISLasso, except under strong correlation, where Lasso is slightly closer.
- **n = 150:**
  - ISLasso best overall.
  - Adaptive Elastic Net consistently the second-best across all correlation levels.
  - Relative efficiency: AENet is closest to ISLasso, followed by Lasso.
- **n = 200:**

- ISLasso best overall.
- Adaptive Elastic Net is again the second-best under all correlation levels.
- Relative efficiency: AENet is consistently closest to ISLasso, followed by Lasso.

Overall, **ISLasso dominates across all settings**. The **Adaptive Elastic Net** is the most reliable second-best, with **Lasso** only competing closely under strong correlation when n is small

The Numerical Summary and Graphics of Setting 2:

In Setting 2, we generate simulation data for logistic (binomial) regression models. Again, several choices of the number of independent variables (p=60, 100, and 250). Several sample sizes were considered, when p = 60 the choices of n =30, 40, and 50 and when p = 100 the choices of n =31, 50, and 75, and when p = 250 the choices of n =100, 150, and 200, to form the logistic regression function, so the binary logistic regression model is defined as

$$\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x$$

Simulations were conducted over 300 independent runs, varying predictor correlations (ρ = 0.00, 0.20, 0.40, 0.60, 0.80), three choices of predictors (p), and three sample sizes (n). Results are presented in Figures 3.4–3.6. The **Multi-Stage Adaptive Lasso (MAENet)** was excluded due to its inapplicability in these scenarios.

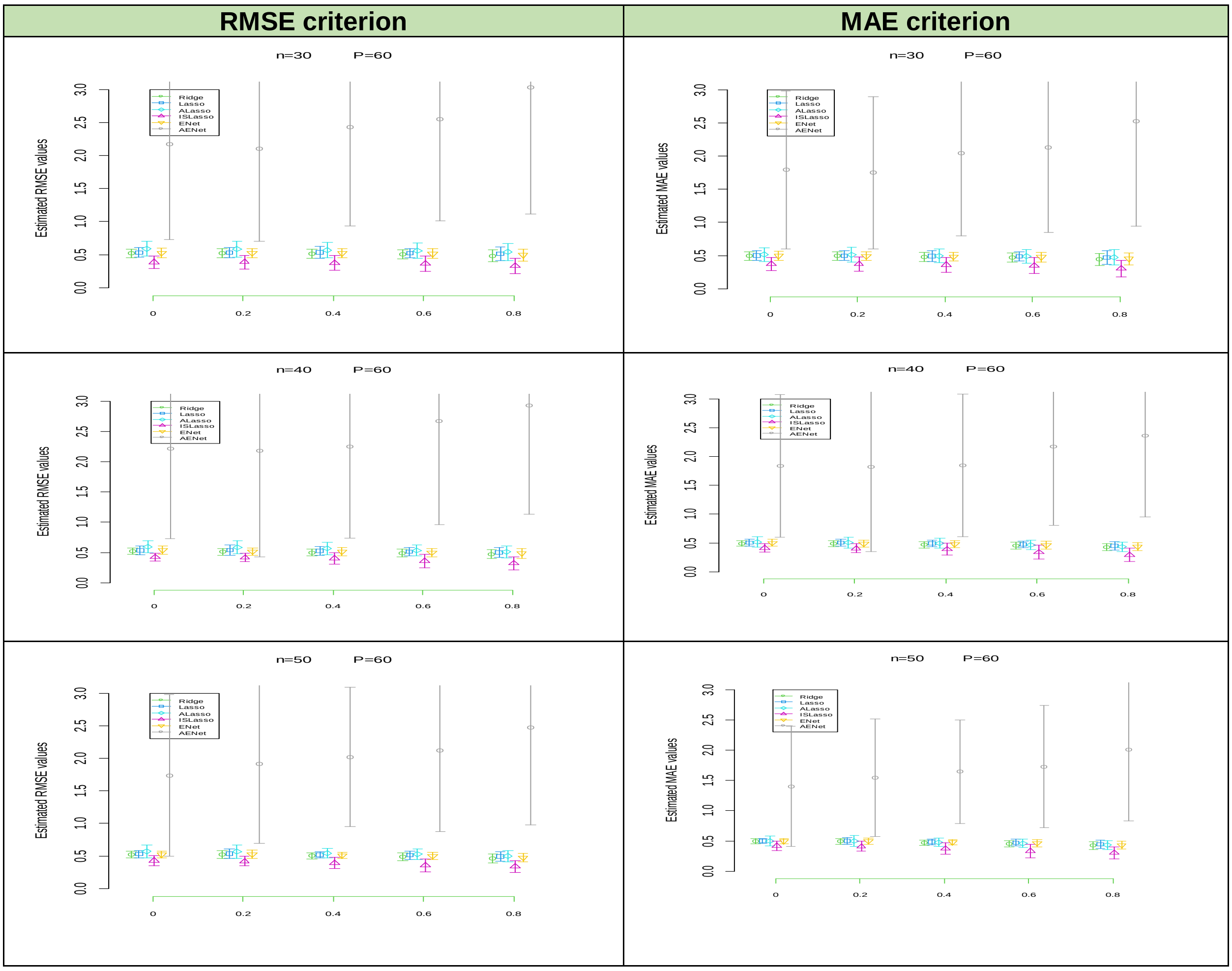

Figure 6: The Box plots demonstrate how various choices of correlation affect the mean (standard deviation) of the RMSE and MAE values for the seven proposed estimators for p= **60**in setting 2.

From Figure 6, we reach the following remarks:

- **General trend:** As the correlation among predictors increases, RMSE and MAE decrease for most models, except for Adaptive Elastic Net (AENet), which worsens with higher correlation.
- **ISLasso performance:**
  - Consistently achieves the lowest RMSE and MAE at all correlation levels and sample sizes.
  - Efficiency improves as correlation increases.

- Ridge is generally the second-best performer.
- **Other models:**
  - Elastic Net (ENet) and Adaptive Elastic Net (AENet) perform poorly relative to ISLasso across all correlation levels.
  - AENet is the least efficient estimator in every scenario.
- **MAE criterion:** Confirms ISLasso as the most efficient, followed by Ridge.

Conclusion: ISLasso dominates across all correlation levels and sample sizes for p = 60, while Elastic Net and Adaptive Elastic Net are consistently less efficient.

RMSE criterion	MAE criterion

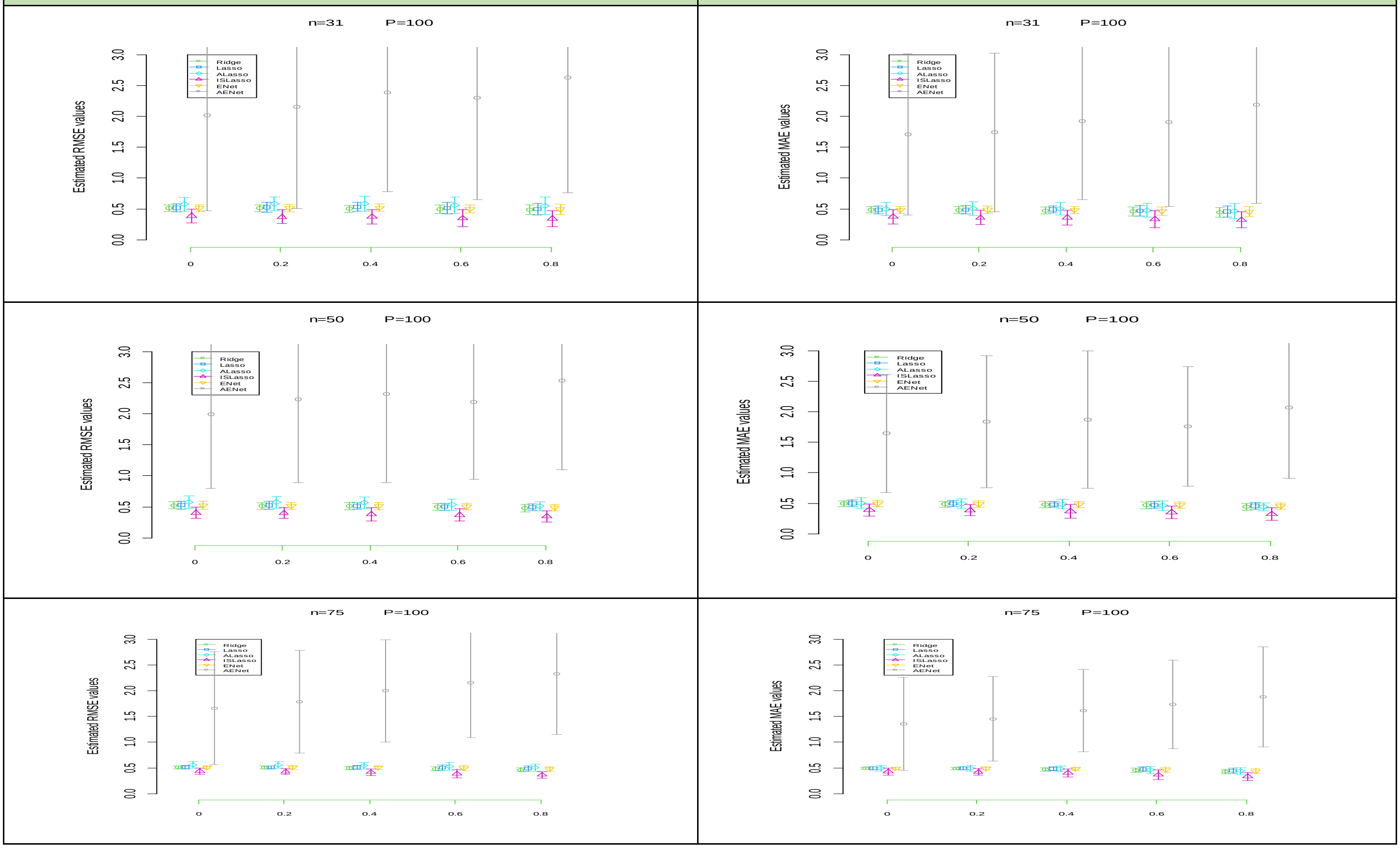

Figure 7: The Box plots demonstrate how various choices of correlation affect the mean (standard deviation) of the RMSE and MAE values for the seven proposed estimators for p=100 in setting 2.

From Figure 7, we reach the following remarks:

- **General trend:** As predictor correlation increases, RMSE and MAE decrease for most models, except **Adaptive Elastic Net (AENet),** which shows increasing errors with higher correlation.
- **ISLasso performance:**
  - Consistently achieves the lowest RMSE and MAE across all correlation levels and sample sizes.
  - Efficiency improves with increasing correlation.
  - Ridge is generally the second-best estimator.
- **Other models:**
  - Elastic Net and Adaptive Elastic Net perform poorly relative to ISLasso.
  - AENet is the least efficient across all scenarios.
- **MAE criterion:** Confirms ISLasso as the most efficient, followed by Ridge.

Conclusion: ISLasso is the dominant estimator for p = 100, while Ridge is second-best, and Elastic Net and Adaptive Elastic Net are consistently less efficient.

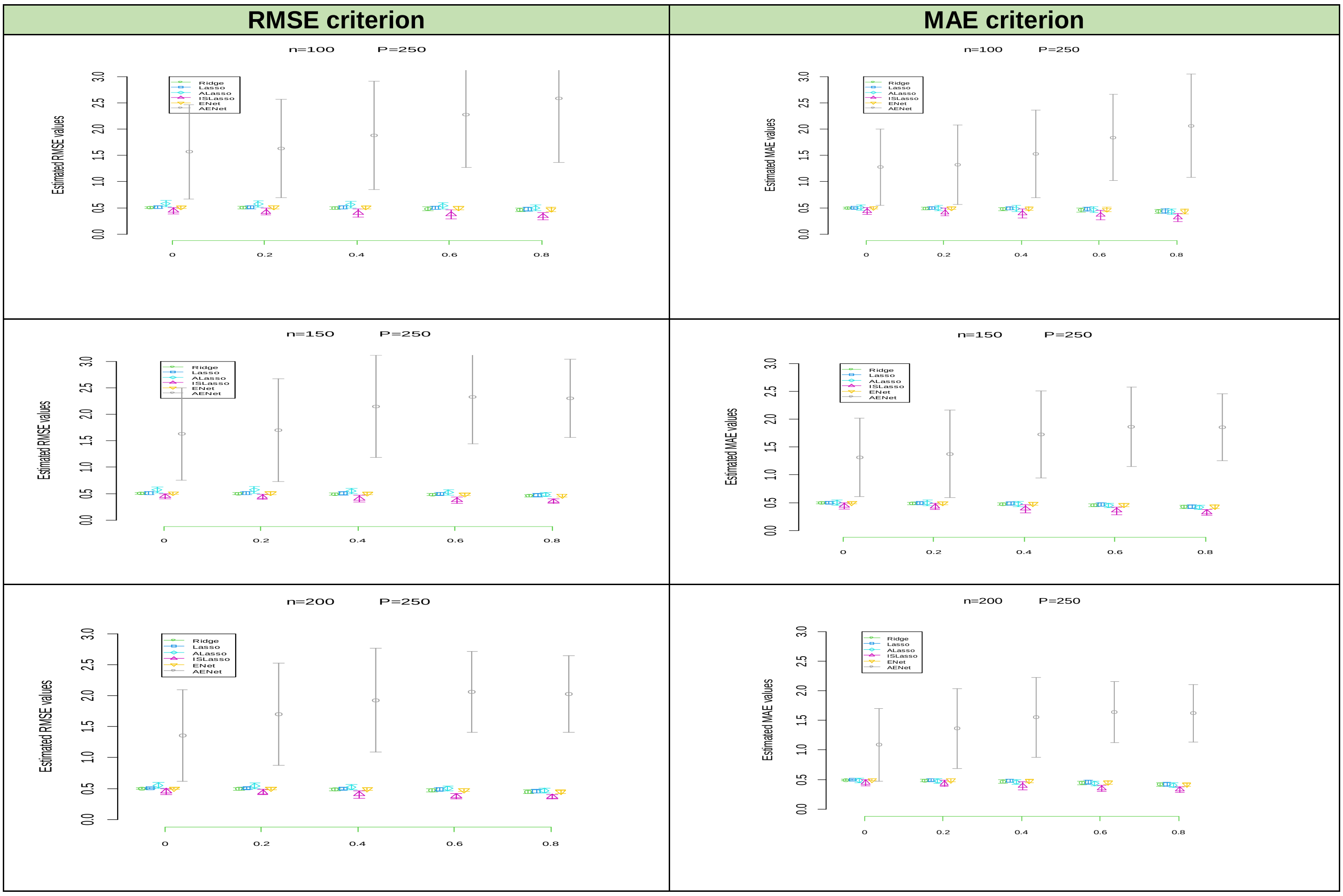

Figure 8: The Box plots demonstrate how various choices of correlation affect the mean (standard deviation) of the RMSE and MAE values for the seven proposed estimators for p=250 in setting 2.

Figure 8 reveals several consistent patterns across simulation scenarios for p=250p=250:

1. **Correlation Effects:** For most estimators, predictive error (RMSE and MAE) decreased as predictor correlation increased. The notable exception was the **Adaptive Elastic Net (AENet)**, whose errors grew with higher correlation, indicating instability in high-dimensional, correlated settings.
2. **Dominant Performance of ISLasso:** The **Induced Smoothed LASSO (ISLasso)** achieved the lowest RMSE and MAE across all correlation levels and sample sizes. Its efficiency improved with increasing correlation, suggesting a particular advantage in managing multicollinearity.
3. **Performance Hierarchy: Ridge regression** emerged as the second-best performer overall, while standard **Elastic Net** and **Adaptive Elastic Net** performed poorly relative to both ISLasso and Ridge.
4. **Validation Across Metrics:** The Mean Absolute Error (MAE) criterion confirmed the RMSE results, reinforcing ISLasso's superiority and Ridge's position as a robust alternative.
5. **Methodological Implication:** The strong performance of ISLasso with logistic loss demonstrates its effective extension to binary classification, where it maintains simultaneous feature selection and stable parameter estimation in high dimensions.

For high-dimensional scenarios with p=250, ISLasso is the most efficient and reliable estimator. Its performance is robust to increasing correlation and generalizes effectively from Gaussian to logistic regression frameworks. Ridge regression serves as a computationally simple and stable benchmark, particularly for binary outcomes. Practitioners should be cautious with Adaptive Elastic Net in correlated, high-dimensional settings due to its inconsistent performance.

There are several benefits of using the ISLasso model with binary outcomes, that includes:

1. ISLasso model Handles correlated predictors effectively.
2. The ISLasso model selects relevant variables for classification.
3. The ISLasso model can improve predictive accuracy and interpretability in binary classification tasks.

Conclusion:

This study conducted an extensive simulation-based comparison of seven penalized regression estimators-Ridge, LASSO, Elastic Net, Adaptive LASSO, Adaptive Elastic Net, Induced Smoothed LASSO (ISLasso), and Multi-stage Adaptive Elastic Net (MAENet), under systematically varied conditions of multicollinearity, dimensionality, sample size, and outcome type (continuous and binary). The central and most consistent finding is the superior performance of the ISLasso estimator. Across both Gaussian and logistic regression frameworks, ISLasso achieved the lowest prediction error (RMSE and MAE) in the vast majority of scenarios. Its efficiency was particularly pronounced under moderate to high correlation, highlighting its robustness in overcoming the twin challenges of multicollinearity and high dimensionality. This can be attributed to its smoothing mechanism, which stabilizes coefficient estimates while preserving the variable selection capability of the Lasso penalty, thereby optimizing the bias-variance trade-off.

The performance hierarchy of the other methods revealed important context-dependent insights:

- For continuous outcomes, the Adaptive LASSO consistently emerged as a strong second choice, especially when correlations were high, due to its oracle properties and adaptive weighting.
- For binary outcomes, Ridge regression proved to be a surprisingly effective and stable benchmark, reliably ranking second. This underscores that the simple L_2 penalty can provide substantial stabilization in high-dimensional classification problems.
- In contrast, the Elastic Net and Adaptive Elastic Net (AENet) exhibited less reliable performance. While sometimes competitive in low-correlation Gaussian settings, they showed significant instability, with AENet's error increasing with correlation in logistic models, marking it as the least efficient estimator in our study.

Limitations and Future Work:

This study was based on simulated data with specific correlation structures (equicorrelation) and linear signal. Future research should validate these findings on real-world datasets with more complex covariance patterns and nonlinear relationships. Furthermore, investigating the computational efficiency and scalability of these estimators, particularly ISLasso and MAENet, with ultra-high-dimensional data (p>>1000p>>1000) would be valuable.

Practical Recommendations Based on our empirical evidence, we offer the following guidance for practitioners:

1. **Primary Recommendation:** For high-dimensional regression problems with correlated predictors, the Induced Smoothed LASSO (ISLasso) is the recommended first-choice estimator due to its optimal balance of prediction accuracy, stability, and inferential capability.
2. **Practical Alternatives:** Use Adaptive LASSO for Gaussian responses where interpretable variable selection is paramount. For logistic regression, Ridge regression offers a robust, simple, and computationally efficient alternative.
3. **Use with Caution:** The Adaptive Elastic Net should be applied cautiously in settings with correlated predictors, as its performance can degrade significantly.

In summary, this research provides a clear empirical basis for selecting penalized regression methods. It establishes ISLasso as a leading-edge estimator for modern statistical learning challenges, effectively bridging the gap between prediction, variable selection, and reliable inference.

References

[1] Baltagi, B. H. (2001). Econometric analysis of panel data (2nd ed.). John Wiley & Sons.

[2] Candès, E. J., & Tao, T. (2007). The Dantzig selector: Statistical estimation when *p* is much larger than *n*. The Annals of Statistics, 35(6), 2313–2351.

[3] Candès, E. J., Wakin, M. B., & Boyd, S. P. (2008). Enhancing sparsity by reweighted ℓ1 minimization. Journal of Fourier Analysis and Applications, 14(5–6), 877–905.

[4] Donoho, D. L., Johnstone, I. M., Hoch, J. C., & Stern, A. S. (1995). Maximum entropy and the nearly black object. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 41–81.

[5] Efron, B., & Hastie, T. (2021). Computer age statistical inference: Algorithms, evidence, and data science. Cambridge University Press.

[6] Fan, J., & Lv, J. (2008). Sure independence screening for ultrahigh-dimensional feature space. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 70(5), 849–911.

[7] Hastie, T., Tibshirani, R., & Friedman, J. (2015). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.

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

[9] Kotz, S., & Nadarajah, S. (2004). Multivariate t-distributions and their applications. Cambridge University Press.

[10] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267–288.
[11] Weisberg, S. (2005). Applied linear regression (3rd ed.). John Wiley & Sons.
[12] Xiao, N., & Xu, Q. S. (2015). Multi-step adaptive elastic-net: Reducing false positives in high-dimensional variable selection. Journal of Statistical Computation and Simulation, 85(18), 3755–3765.
[13] Zhu, R., Yu, Y., Li, R., & Zhu, Z. (2021). Induced smoothed lasso for inference in high-dimensional regression. Journal of the American Statistical Association, 116(534), 821–834.
[14] Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 101(476), 1418–1429.
[15] 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.