

REGULARIZED LINEAR REGRESSION: A STUDY IN DIFFERENT DIMENSIONAL FRAMEWORKS

*** Mohamed Bachir AOUN:**

National Higher School of Statistics and Applied Economics

Email : aoun.mohamedbachir@enssea.edu.dz

Fatiha MEZOUED :

National Higher School of Statistics and Applied Economics

Email : mezoued.fatiha@enssea.edu.dz.

Reçu le : 06/04/2025 **Accepté le :** 26/05/2025 **Publication en ligne le :** 03/06/2025

ABSTRACT:

This study concentrates on the evaluation of various regression model selection methods with sample sizes exceeding the number of features in one hand, and the number of features Surpassing sample size case in other, using parameters from [1] for low dimensional settings and [2] for high dimensional setup , focusing on regularization techniques and true subset selection accuracy. The experiments used an orthogonal, low correlation and high correlation design matrix under various number of features, with noise vectors generated from a Gaussian distribution, and varied noise levels, and observation counts. The assessment covered Least Absolute Shrinkage and Selection Operator (LASSO), Ridge, Elastic-Net, Minimax Concave Penalty (MCP) and Smoothly Clipped Absolute Deviation (SCAD) as sparse regularization methods, adaptive methods and Step-Wise selection methods. Performance was evaluated based on the frequency of correct subset selections, with higher frequencies reflecting superior methods as mentioned in [3]. In low dimensional data, all methods generally performed well across all scenarios, while some regularization methods were sensitive to sample size and noise. For high dimensional generation, all methods reach the perfect rate of recovery except the Step-Wise selection that shows perturbations and limitation in this case. This study not only provides valuable insights for practitioners in statistical model selection but also calls for future research to explore the application of these methods in real- world phenomena in all fields of study, which matches the parameters settings mentioned in this paper.

Key words: Statistical Model, High Dimension, Linear Regression, Sparse Regularization Methods, Model Selection.

* Mohamed Bachir AOUN

I INTRODUCTION

Building robust and precise statistical models is essential in many research fields, from economics to bioinformatics. In the work of [4], The new criteria consider both the number of unknown parameters and the model space's complexity as an amelioration for the classic Bayesian criteria, ensuring consistency even as covariates increase with sample size to infinity, under lasso selection. The author in [5] proposed an evaluation criterion using Fisher information for the selection under maximum level of sparsity for a true variable selection by lasso estimator. Some paper concentrate on the weakness of the two previous papers to give a robust criteria for under difficult settings between variation of sample size and SNR [6]. For the previous studies, a comparison between regularization methods based on the frequencies of recovery as a first step, under various correlation and variance settings when the number of observation is greater than the number of features[3]. Then the author evaluated the variable selection methods by including the evaluation methods in order to choose the best model selection under each scenario. Inspiring form literature, we focus on comparing the performance of different regularization and variable selection methods for high dimensional data set and regular dimensional dataset. Specifically, we analyze sparse regularization methods such as Lasso, Ridge, and Elastic Net, along with adaptive methods like Adaptive Lasso and Adaptive Elastic Net. Additionally, we study Step-Wise selection methods, including Forward Selection and Backward Elimination, and include penalty methods such as SCAD and MCP in our analysis. This study evaluates how well different modeling methods work in various features setups (orthogonal, low correlation, high correlation) by considering error variance (σ), sample size and number of features variations. By looking at these factors, the study helps understand where each method performs best or has limitations for each case, guiding practitioners in choosing the most suitable modeling method based on our finding for their specific feature design and application field.

II MATERIAL METHOD:

Linear regression is one of the most fundamental techniques used in statistical modeling and machine learning to make predictions based on one or more predictor variables. But as datasets become more complex, the risk of overfitting the model to noise rather than the actual data pattern increases exponentially. This problem has established regularized linear regression as an important enhancement of linear regression methods.

Regularization methods add penalty term on loss function which model seeks to minimize. By limiting the coefficient estimates, this penalty prevents very complex models resulting in better models on unseen data. It serves to prevent overfitting, also creates a model based on subset selection features of datasets with large X such as tokens of text or the number of attributes in a market research dataset. We review several methods that build

a series of models with increasing complexity. These methods are categorized into sparse regularization approaches, which involve optimization problems that yield sparse solutions, and Step-Wise methods, which are among the earliest techniques documented in the literature.

A Linear Regression Model

Our study focus on the linear relation between the target variable and explanatory variables, we define in a simple way the linear regression model:

$$y = x \beta + \varepsilon, \quad (1)$$

And in a complex way:

$$y = \sum_{j=1}^p x_{ij} \beta_j + \varepsilon, \quad (2)$$

Consider the scenario Y are real-valued of the target variable, and $X = (x_1, \dots, x_p)$ constitute the design matrix that contents the explanatory features. Here, ϵ represents the noise vector, and σ specifies the noise level. The primary objective is to accurately estimate the unknown regression coefficient vector β .

B Sparse Regularization Methods

We will present a detailed overview of various sparse regularization methods, exploring their principles:

Ridge Regression

In linear regression, collinearity in the design matrix X can lead to unstable solutions when using squared-error loss. Ridge regression mitigates this issue by introducing a penalty on the size of the regression coefficients, thereby minimizing the penalized residual sum of squares:

$$\hat{\beta}_{ridge} = \arg \min_{\beta} (\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2) \quad (3)$$

Where $\lambda \geq 0$ is a regularization parameter that controls the degree of shrinkage applied to the coefficients. Larger λ values result in greater shrinkage [7].

LASSO

The LASSO (Least Absolute Shrinkage and Selection Operator), introduced by [8], is a shrinkage technique using an L1 penalty to find the best fit for a linear regression model (2)

,The estimator $\hat{\beta}^{LASSO}$ minimizes the objective function similar to ridge but the absolute value of penalty term:

$$\hat{\beta}_{LASSO} = \arg \min_{\beta} (\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|) \quad (4)$$

This method encourages sparsity by setting some coefficients to zero, impacting the variable selection and a simpler, more interpretable model as λ increases.

Elastic net

Elastic Net is a regression technique that merge LASSO and ridge in a single framework by combining the squared penalty term (3) and the absolute penalty term (4) , as proposed by [9]. The objective function presents in (5):

$$\hat{\beta}_{Elastic-Net} = \arg \min_{\beta} (\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2) \quad (5)$$

Where Y is the response variable, X is the matrix of predictors, and β represents the regression coefficients. The terms $\lambda_1 \sum_{j=1}^p |\beta_j|$ (Lasso) and $\lambda_2 \sum_{j=1}^p \beta_j^2$ (Ridge) control feature selection and coefficient shrinkage, respectively.

Minimax Concave Penalty (MCP) and Smoothly Clipped Absolute Deviation (SCAD)

Minimax Concave Penalty (MCP) and Smoothly Clipped Absolute Deviation (SCAD) are quadratic spline penalty functions introduced by [10]and [11], respectively, to address the bias of (3). MCP combines Lasso and Hard Threshold (HT) penalties, defined as:

$$\rho_{MCP}(|\beta_j|) = |\beta_j| - \frac{\beta_j^2}{2\gamma\lambda} 1_{\{|\beta_j| \leq \gamma\lambda\}} + \frac{\gamma\lambda}{2} 1_{\{|\beta_j| > \gamma\lambda\}} \quad (6)$$

While SCAD is defined as:

$$\rho_{SCAD}(|\beta_j|) = |\beta_j| 1_{\{|\beta_j| \leq \lambda\}} - \frac{(|\beta_j| - \lambda)^2}{2(a-1)\lambda} 1_{\{\lambda < |\beta_j| \leq a\lambda\}} + \frac{(a+1)\lambda}{2} 1_{\{|\beta_j| > a\lambda\}} \quad (7)$$

Here, λ tunes sparsity like Lasso, while $\gamma > 1$ and $a > 2$ adjust the estimator's bias. MCP tends to the restricted least-squares when γ is close to 1, unlike SCAD. For an orthogonal design matrix X, MCP corresponds to the Firm Shrinkage estimator.

MCP tends towards restricted least-squares when γ is near 1, making it easier to solve than SCAD. For an orthogonal design matrix X , MCP corresponds to the Firm Shrinkage estimator by Bruce & Gao[12]. Firm Shrinkage and SCAD penalties are defined as:

$$\hat{\beta}_j^{FS} \begin{cases} 0 & \text{if } |Y^t X_j| \leq \lambda \\ \frac{\gamma}{\gamma-1} (Y^t X_j - \lambda \text{sgn}(Y^t X_j)) & \text{if } \lambda < |Y^t X_j| \leq \gamma \lambda \\ Y^t X_j & \text{if } |Y^t X_j| \geq \gamma \lambda \end{cases} \quad (8)$$

$$\hat{\beta}_j^{SCAD} \begin{cases} (Y^t X_j - \lambda \text{sgn}(Y^t X_j)) & \text{if } |Y^t X_j| \leq 2\lambda \\ \frac{a}{a-2} (Y^t X_j - \lambda \text{sgn}(Y^t X_j)) & \text{if } 2\lambda < |Y^t X_j| \leq a\lambda \\ Y^t X_j & \text{if } |Y^t X_j| \geq a\lambda \end{cases} \quad (9)$$

Both methods combine soft and hard thresholds. SCAD is more biased than MCP, especially when the parameter (a) is near its lower limit 2, and includes subsets I_0 (null components), I_p (penalized components), and I_N (non-penalized components).

Adaptive LASSO and Adaptive Elastic net

Adaptive Lasso and Adaptive Elastic Net improve upon Lasso and Elastic Net by using adaptive versions as introduced by [13] and [14]. The Adaptive Lasso penalty is defined as

$$\rho_{adaptive\ lasso}(|\beta_j|) = w_j |\beta_j| \quad (10)$$

And the Adaptive Elastic Net as

$$\rho_{adaptive\ net}(|\beta_j|) = w_j |\beta_j| + \frac{\lambda_2}{\lambda_1} \beta_j^2 \quad (11)$$

Where w_j is typically $|\beta_j^{init}|^{-q}$, with β^{init} being an initial solution from methods like Least-squares or Ridge regression, and q is usually 1 or 2. As we can see, the Adaptive Lasso applies the weight w_j on penalty terms in (4). For Adaptive Lasso, $w_j = 1/|\beta_j^{LS}|$ and $q=1$ aligns with the Garrote method [15].

Adaptive Lasso and Adaptive Elastic Net offer convex optimization problems, ensuring easier solutions with unique results. For weights chosen as in β^{init} with $q=2$, the estimators are:

- Adaptive Lasso:

$$\hat{\beta}_j^{adallasso} = (Y^t X_j - \tilde{\lambda} \frac{\text{sgn}(Y^t X_j)}{|Y^t X_j|^q}) 1_{|Y^t X_j|^{1+q} > \tilde{\lambda}} \quad (12)$$

- Adaptive Elastic Net:

$$\hat{\beta}_j^{adanet} = \frac{1}{1+\lambda_2} (Y^t X_j - \tilde{\lambda}_1 \frac{\text{sgn}(Y^t X_j)}{|Y^t X_j|^q}) 1_{|Y^t X_j|^{1+q} > \tilde{\lambda}_1} \quad (13)$$

[3] Illustrate these estimators with $\tilde{\lambda}=2$, $\tilde{\lambda}_1=2$, and $\lambda_2=0.1$. The hyper parameters $\tilde{\lambda}$ and $\tilde{\lambda}_1$ must differ from those in soft thresholding or elastic net. The condition for all β components to be zero is

$$\tilde{\lambda} \geq \tilde{\lambda}^{(1)} = \max_j |Y^t X_j|^{1+q} = (\tilde{\lambda}^{(1)})^{1+q} \quad (14)$$

Where $\lambda^{(1)}$ is Lasso's hyperparameter for which the first variable is included in the selection, with $q=0$ for Lasso and $q=\infty$ for hard thresholding.

CStep-Wise Methods

Step-wise methods are greedy algorithms for feature selection in regression, adding or removing variables based on their impact on model fit. At each step, they use restricted least-squares estimation to update coefficients, continuing until a stopping rule like a maximum number of variables or desired model fit is met. While they simplify variable selection and maintain accuracy, they may not guarantee optimal models and can be sensitive to the order of variable inclusion.

Forward Selection

Forward Selection begins with an empty model ($\hat{\beta}_0 = 0$) and adds variables one by one, selecting the one that most improves the model fit. The coefficients for the selected variables are estimated using restricted least-squares from (2):

$$\hat{\beta}_{LS_I} = (X_I^T X_I)^{-1} X_I^T Y \quad (15)$$

The next variable is chosen as in [16] to maximize the change in the Sum of Squared Errors (SSE):

$$j_{next} = \arg \max_{j \in \{1, \dots, p\} \setminus I} |\Delta SSE| = \arg \max_{j \in \{1, \dots, p\} \setminus I} |kY - XI\hat{\beta}_{LS_I} - kY - XI \cup \{j\}\hat{\beta}_{LS_{I \cup \{j\}}}|^2 \quad (16)$$

It uses a stopping criterion, a Fisher test with statistics as indicated in (17):

$$F_{forward} = \frac{\|X_I \hat{\beta}_I^{LS}\|^2 - \|X_{I \cup \{j_{next}\}} \hat{\beta}_{I \cup \{j_{next}\}}^{LS}\|^2}{\|Y - X_{I \cup \{j_{next}\}} \hat{\beta}_{I \cup \{j_{next}\}}^{LS}\|^2 / (n-k-1)} \quad (17)$$

Backward Elimination

Backward Elimination starts with a full model estimated by least-squares with coefficients $\hat{\beta}^{LS}$ and iteratively removes variables based on the equation (18):

$$j_{del} = \arg \min_{j \in I} \|Y - X_I \hat{\beta}_{LS_I}\|^2 - \|Y - X_{I \setminus \{j\}} \hat{\beta}_{I \setminus \{j\}}^{LS}\|^2 \quad (18)$$

It uses a stopping criterion, often a Fisher test, with statistics as shown in (19):

$$F_{backward} = \frac{\|Y - X_I \hat{\beta}_I^{LS}\|^2 - \|Y - X_{I \setminus \{j_{del}\}} \hat{\beta}_{I \setminus \{j_{del}\}}^{LS}\|^2}{\|Y - X_I \hat{\beta}_I^{LS}\|^2 / (n-k)}, \quad (19)$$

Step-Wise Regression combines Forward Selection and Backward Elimination to assess joint relevance of variables. Stopping criteria also include R^2 , adjusted R^2 , Mallows C_p , AIC, and BIC.

III THE RESEARCH FINDINGS AND DISCUSSION:

In this simulation scenarios, for n greater than p , p is fixed as the regression coefficient vector β is set to: (2, 0, 0, 4, and 0) with two non-zero coefficients. For the high dimensional settings, n and p variate proportional according to $p = n^d$ ($d=1.3$) with five nonzero coefficients (10, 20, 30, 40, 50). The design matrix X is either orthogonal or general with a correlation matrix Σ , where ρ is taken from the set $\{0, 0.2, 0.4, 0.5, 0.6, \text{ and } 0.8\}$. Given that the theoretical study from previous papers assumed a fixed design, X is generated once and fixed for the entire experiment. We generate 5000 replicates of the noise vector ϵ from a Gaussian distribution with covariance matrix $\sigma^2 I_n$, where $\sigma^2 I_n$ is taken from $\{0.2, 0.5, 1, \text{ and } 2\}$. The number of observations n varies over $\{20, 40, 60, 80 \text{ and } 100\}$ for both cases. The experiment is repeated 10 times, resulting in 10 different examples for X : one with an orthogonal design (correlation $\rho=0$), one with a general design and maximum low correlation ($\rho=0.4$), and one with a general design and maximum high correlation ($\rho=0.8$). We then compute the frequency of selection of the true subset, where a good selection is indicated by a frequency close to 1, indicating the proportion of times the correct subset is selected.

This table resume the choice of parameters for the model selection methods:

Tab.1: Hyper-Parameters for constructing models for each method

Methods	Hyper parameters
LASSO	$\lambda=1$
Ridge	$\lambda=0.1$
MCP	$\gamma=2$
Adaptive lasso	$w = (\beta_j^{LS})^{-2}$
Elastic-Net	$\lambda_2=0.3$
Adaptive Elastic-Net	$w = (\beta_j^{LS})^{-2}, \lambda_1 = 2, \lambda_2=0.3$
SCAD	$a=3.7$

Our comparison criteria is the frequency of recovery:

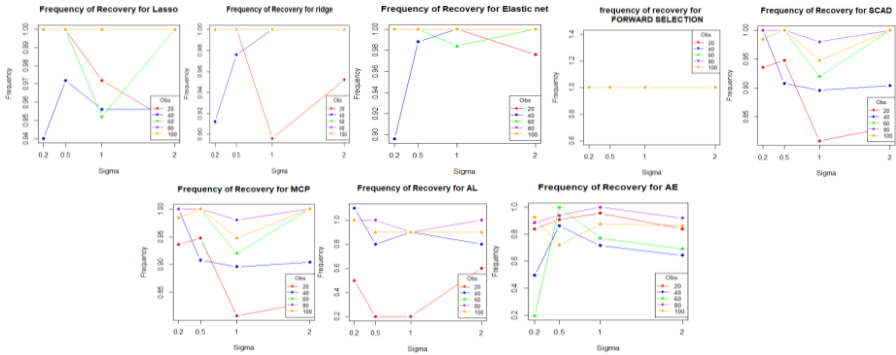
$$freq(\beta) = \frac{1}{R} \sum_{r=1}^R 1\{sgn(\hat{\beta}^r) = sgn(\beta)\}$$

With which estimated coefficients $\hat{\beta}$ matches the sign of the true parameter β across R replications.

ADetailed Analysis for Orthogonal Features Design Regular Dimensional Settings:

This part compares the of the model selection methods under various sample sizes and variances values for the orthogonal case when the sample size is greater than number of variables:

Figure 1: Average frequency of selection with respect to sample size (n) and variance σ when $n>p$ for the general design case with correlation $\rho = 0$ between the variables

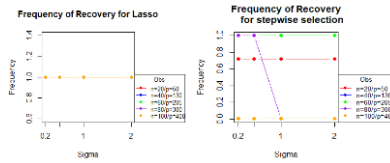


The examination of methods across different sample sizes and sigma values shows remarkable patterns. Forward and Backward Selection methods show perfect performance across all sample sizes, Lasso and Ridge methods achieve high, stable frequencies across all sample sizes, while SCAD and MCP present some variability, especially at lower sample sizes. Elastic Net's performance is stable at higher sigma values but more variable at lower ones. Adaptive Lasso and Adaptive Elastic Net generally perform better at higher sigma values but show more variability.

BDetailed Analysis for Orthogonal Features Design High Dimensional Settings:

This part analyze and discuss the performance of the model selection methods under various sample sizes and variances values for the orthogonal case when the number of features is greater than the number of observations:

Figure 2: Average frequency of selection with respect to sample size (n), number of variables(p) and variance (σ) when $p > n$ for the general design case with correlation $\rho = 0$ between the variables.



The comparative study of variable selection methods under the orthogonal features design reveals distinct performance patterns with significant implications for practical applications. On one hand, the Lasso, Ridge, Elastic Net, MCP, SCAD, Adaptive Lasso, and Adaptive Elastic Net techniques demonstrate remarkable robustness and reliability, consistently achieving perfect recovery rates across a wide range of scenarios, these methods showcased their effectiveness in handling varying sample sizes ($n = 20$ to 100) related to feature counts ($p = 50$ to 400), and noise levels ($\sigma = 0.2$ to 2). Their steady performance, regardless of the complexity of the data conditions, underscores their suitability for diverse real-world applications where data characteristics can fluctuate.

Conversely, the Step-Wise selection method exhibited a more variable and inconsistent performance. While it achieved perfect recovery in moderate scenarios ($n = 40/p = 130$ and $n = 60/p = 205$), its efficacy deteriorated significantly in smaller sample size settings ($n = 20/p = 50$) and larger, more complex datasets ($n = 80/p = 300$ and $n = 100/p = 400$), especially under higher noise conditions ($\sigma = 1, 2$).

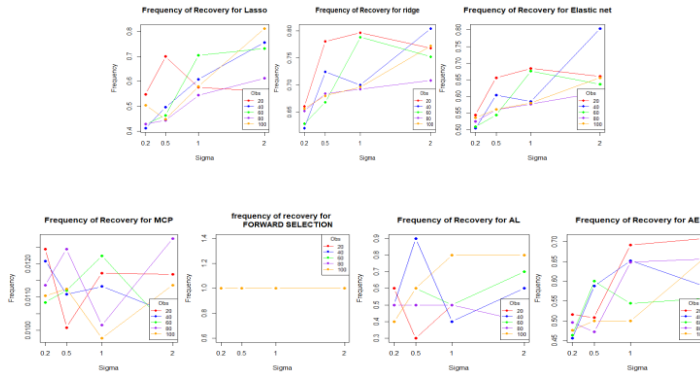
This variability in Step-Wise selection's performance suggests a sensitivity to both noise and overfitting, potentially limiting its reliability in high-dimensional or noisy

environments. The method's inconsistency underscores the need for careful consideration of its suitability for specific applications, as its effectiveness may be constrained to narrower contexts.

CDetailed Analysis for Low correlation case Regular settings:

This part analyze and discuss the performance of the model selection methods under various sample sizes and variances values when the correlation between features is low (0.4) when the sample size is greater than number of variables:

Figure 3: Average frequency of selection with respect to sample size (n) and variance (σ) when $n > p$ for the general design case with maximum correlation $\rho = 0.4$ between the variables.

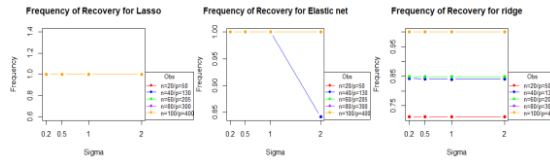


The results reveals that Lasso regression performs well at smaller sample sizes, with its effectiveness increasing as the sample size grows, especially with high sigma values -1 and 2-, achieving high frequencies of 0.812 and 0.756 respectively. Ridge regression outperforms other methods, especially with sigma 1 and 2, showing high frequencies up to 0.796. Elastic Net demonstrates stable performance across various sample sizes, with its best results aligning closely with Ridge. MCP continue to underperforms, with frequencies remaining below 0.02 across all sample sizes. Adaptive Lasso shows notable improvement with increasing sample sizes and sigma values, achieving its best performance at sigma 1. Under lower noise level, Ridge and Elastic Net outperform Lasso, with frequencies of 0.620 and 0.544 respectively, while Adaptive Lasso and Adaptive Elastic Net offer moderate improvements. At higher sigma values, Ridge remains the top performer, and Adaptive methods show strong flixibility, particularly with larger sample sizes. MCP remains ineffective under all conditions.

D Detailed Analysis for low correlated Features Design High Dimensional settings:

This part analyze and discuss the performance of the model selection methods under various sample sizes and variances values for the low correlation case when the number of features is greater than the number of observations:

Figure 4: Average frequency of selection with respect to sample size (n), number of variables (p) and variance (σ) when $p > n$ for the general design case with maximum correlation $\rho = 0.4$ between the variables.

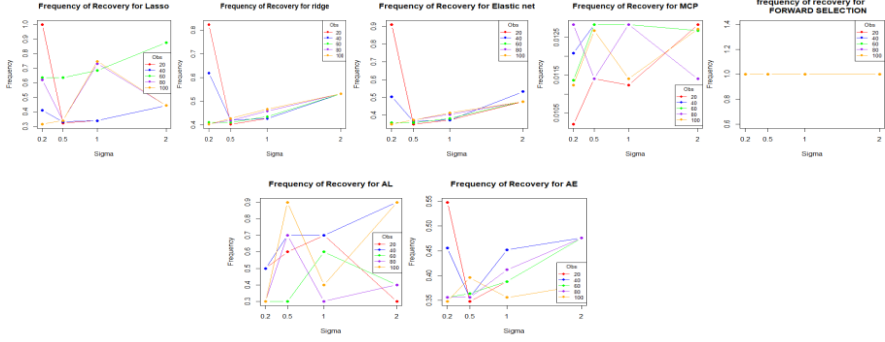


Lasso, MCP, and Adaptive Methods achieve perfect recovery across all scenarios, demonstrating their robustness and reliability in all sample sizes (n), number of features (p), or noise level (σ). This consistent effectiveness ensure them reliable for variable selection under various conditions. In contrast, Ridge shows variable performance, with recovery frequencies improving as both sample size and feature count increases to reach perfect recovery at higher complexities ($n=80/p=300$ and $n=100/p=400$), its performance diminishes with smaller datasets, especially at elevated noise levels (σ), where frequencies fluctuate between 0.712 and 0.849. This variability suggests Ridge's sensitivity to noise, necessitating cautious application in more challenging scenarios. Elastic Net also maintains strong performance, with perfect recovery in most cases. Even so, it experiences a minor decline in recovery at $n=40/p=130$ and $\sigma=2$, with a frequency of 0.841. Despite this slight decrease, its overall robustness positions it as a fllexible tool for variable selection across a range of data conditions.

E Detailed Analysis for High correlated features Regular settings:

This part analyze and discuss the performance of the model selection methods under various sample sizes and variances values for high correlation values (0.8) when the sample size is greater than number of variables:

Figure 5: Average frequency of selection with respect to sample size (n) and variance (σ) when $n > p$ for the general design case with correlation $\rho = 0.8$ between the variables.

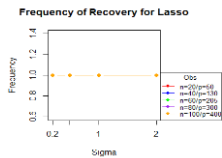


In all conditions, Step-Wise Selection achieve perfect scores, demonstrating their robustness. Lasso performs well at smaller sample sizes and lower sigma values but shows decreased effectiveness as sigma increases and sample size grows, with scores ranging from 1.000 at sigma 0.2 and sample size 20 to 0.748 at sigma 1 and sample size 100. Ridge maintains consistent performance across sample sizes and sigma values, scoring between 0.404 and 0.824, showing reliability but reduced effectiveness with larger datasets. Elastic Net also performs well at lower sigma values but struggles in larger samples, with scores from 0.912 at sigma 0.2 and sample size 20 to 0.348 at sigma 1 and sample size 100. MCP underperforms consistently, with low scores between 0.01020 and 0.01284. Adaptive Lasso and Adaptive Elastic Net shows moderate to good performance, with Adaptive Lasso showing improvements at larger sample sizes and sigma values of 0.5 and 2, while Adaptive Elastic Net remains stable across conditions. Overall, Forward and Backward Selection are the most reliable methods for feature selection across varying conditions.

F Detailed Analysis for high correlated Features Design:

This part analyze and discuss the performance of the model selection methods under various sample sizes and variances values for the high correlation case when the number of features is greater than the number of observations:

Figure 6: Average frequency of selection with respect to sample size (n), number of variables(p) and variance (σ) when $p > n$ for the general design case with correlation $\rho = 0.8$ between the variables.



The analysis comparing different variable selection methods under high correlation features framework. Techniques such as Ridge, Elastic Net, MCP, Adaptive Lasso, and Adaptive Elastic Net indicates exceptional robustness and reliability similar to Lasso, continue to achieve perfect recovery rates across various scenarios. These methods effectively manage different sample sizes ($n = 20$ to 100), feature counts ($p = 50$ to 400), and noise levels ($\sigma = 0.2$ to 2). Their consistent performance, regardless of data complexity, underscores their suitability for a wide range of real-world applications where data characteristics may vary.

IV CONCLUSION :

As a result, this analysis showcase the importance of choosing the appropriate variable selection methods based on the specific characteristics of the dataset, including sample size, dimensionality, noise levels, and correlation structure. In low-dimensional settings, Step-Wise Selection either Forward or Backward are Very efficient. Lasso performs well but struggles with higher noise and correlation. Ridge and Elastic Net demonstrate consistent performance across all scenarios, making them robust solution for variable selection. Adaptive methods offer additional flexibility, especially in high-noise and high-correlation settings.

For high-dimensional datasets, with complex correlation structures, Ridge, Elastic Net, and adaptive methods are suggested for their robustness and reliability. In contrast, Forward and Backward Selection should not be used, as their performance is highly sensitive to noise and overfitting. MCP achieved perfect recovery rate across all noise levels and correlation structures.

The best choice in model building and variable selection, the key is the behavior of the data and there characteristics, The following table resume the performance of the selection methods based on the principal aspects guided by data caracterestics:

Table 2: Recommended model construction methods across key aspects

ASPECT	DESCRIPTION
Robustness to Noise	- Ridge, Elastic Net, Adaptive Lasso, Adaptive Elastic Net
Dimensionality	- Low-dimensional settings: Forward and Backward Selection are highly effective. - High-dimensional settings: Regularization-based methods (Lasso, Ridge, Elastic Net, SCAD, MCP) and adaptive methods perform well.
Correlation Structure	- Ridge and Elastic Net: Remain stable with increasing correlation. - Adaptive methods: Show improved performance in high-correlation settings, especially at higher noise levels.
Consistency	- Ridge, Elastic Net, Adaptive Lasso, Adaptive Elastic Net: Consistently achieve high recovery rates across all scenarios, making them reliable choices for variable selection in diverse conditions.

In conclusion, the choice of variable selection method should be guided by the specific characteristics of the dataset and the research objectives. Sparse regularization and adaptive methods are generally the most reliable in high dimensional data, providing strong performance in various environments. Otherwise, Step-Wise selection select the best model in low dimensional data.

Future research should aim to apply the results of our simulation as a selection method, using evaluation criteria that best fit the variation of parameters in real data.

V REFERENCES :

- [1] Fourdrinier D, Wells M. Comparisons of selection rules of a regression model: a decision approach. *Comptes Rendus L'Académie Sci Sér I*. 1994 Jan 1;
- [2] Gohain PB, Jansson M. Robust Information Criterion for Model Selection in Sparse High-Dimensional Linear Regression Models. *IEEE Trans Signal Process*. 2023;71:2251–66.
- [3] Boisbunon A. Model selection: a decision-theoretic approach. PhD-Thesis. Rouen University. France.2013
- [4] Chen J, Chen Z. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*. 2008 Sep 1;95(3):759–71.
- [5] Owrang A, Jansson M. A Model Selection Criterion for High-Dimensional Linear Regression. *IEEE Trans Signal Process*. 2018 Jul;66(13):3436–46.
- [6] Gohain PB, Jansson M. New Improved Criterion for Model Selection in Sparse High-Dimensional Linear Regression Models. In: *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* [Internet]. Singapore, Singapore: IEEE; 2022 [cited 2025 Jan 18]. p. 5692–6.
- [7] Hoerl AE and K, Robert W. Ridge regression: applications to nonorthogonal problems. Taylor Francis. 1970;12:69--82.
- [8] Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B Methodol*. 1996 Jan;58(1):267–88.
- [9] Zou H, Hastie T. Regularization and Variable Selection Via the Elastic Net. *J R Stat Soc Ser B Stat Methodol*. 2005 Apr 1;67(2):301–20.
- [10] Zhang CH. Nearly unbiased variable selection under minimax concave penalty. *Ann Stat* [Internet]. 2010 Apr 1 [cited 2024 Jul 20];38(2).
- [11] Fan J, Li R. Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *J Am Stat Assoc*. 2001;96(456):1348–60.
- [12] BRUCE AG, GAO HY. Understanding WaveShrink: Variance and bias estimation. *Biometrika*. 1996 Dec 1;83(4):727–45.
- [13] Chung PD. The Adaptive Lasso and Its Oracle Properties Hui Zou (2006), *JASA*.
- [14] Zou H, Zhang HH. On the Adaptive Elastic-net with a Diverging Number of Parameters [Internet]. University of Minnesota; 2009 [cited 2024 Jul 20].

[15] Breiman L. Better Subset Regression Using the Nonnegative Garrote. *Technometrics*. 1995;37(4):373–84.

[16] Rawlings JO, Pantula SG, Dickey DA, editors. *Applied Regression Analysis* [Internet]. New York: Springer-Verlag; 1998 [cited 2024 Jul 20]. (Springer Texts in Statistics).