

Cross Validation across Regularized Linear Regression

Received: 22-10-2025

Accepted: 19-12-2025

Mohamed Bachir AOUN¹, PhD student, Algeria.

Aoun.mohamedbachir@enssea.edu.dz.

ORCID ID : 0009-0006-6943-0107

Fatiha MEZOUED, Professor, mezoued.fatiha@enssea.edu.dz.

ORCID ID : 0009-0008-1216-1465

Abstract

This study evaluates various regression model selection methods with multi-dimensional settings, using parameters from (9) for low dimensional parameters and (12) for high dimensional setup. The experiments varies sample sizes, variables number, correlation levels, and Gaussian noise. The assessment covers LASSO, Ridge, Elastic-Net, MCP and SCAD as sparse regularization methods, adaptive methods and stepwise selection methods. Performance is evaluated based on the cross validation criterion as GCV, V-Fold Cross Validation CV-V and LOOCV. In low dimensional data, LOOCV outperform the other criterion. For high dimensional generation, CV5, CV10, and GCV are preferred depending on the model and correlation strength. The findings offer clarity for model selection and encourage future research into real-world applications across diverse fields.

Keywords : High dimension, Linear Regression, Sparse regularization methods, Cross Validation, Model selection, Evaluation Criteria.

Jel Classification Codes : C520

¹ Author correspondents

1. INTRODUCTION

Construction of a precise statistical models, selecting the optimal pertinent variables related to a certain phenomenon in any field is paramount to achieving robust and reliable predictions, particularly in the presence of varying feature correlation structures and data dimensionalities, which leads us to regularized regression as a solution for complexity. The performance of regularization-based methods such as LASSO, Ridge, Elastic Net, Minimax Concave Penalty (MCP), Adaptive LASSO, and Adaptive Elastic Net depends on the choice of cross-validation criteria—Generalized Cross-Validation (GCV), Leave-One-Out Cross-Validation (LOOCV), 5-Fold Cross-Validation (CV5), and 10-Fold Cross-Validation (CV10)—under the influence of sample size, noise levels, and the correlation status of features. This study systematically evaluates these model-criterion pairings across regular and high-dimensional settings to identify the best-performing combinations based on minimal prediction errors. By examining different scenarios combining the variation of correlation status, dimensionality, and noise, we aim to identify is the most effective Cross Validation criteria as evaluation approach for Regularization Linear Regression under diverse data scenarios. A priori, we believe that either all the cross validation, and based on its independence with noise, will perform well across different scenarios, or it will struggle in high dimension data. So, the objective of this paper is to find from an empiric application presented in simulation of data according to several scenarios varying the essential parameters related to the features structure between two different dimensionality, the most robust cross validation criteria for several Regularization methods for model construction across different data scenarios.

2. Resolution Methods:

The challenge of selecting an optimal statistical model in the presence of diverse variable correlation structures and data dimensionalities needs robust modeling methods that balance model complexity and predictive accuracy. This section focuses on the regularization path and cross-validation criteria as tools for model selection-evaluation.

2.1. Linear regression model:

In our research, we look at how the target and explanatory features connect in a linear form. We describe the linear regression model simply as follows:

$$Y = X\beta + \sigma\epsilon,$$

The vector Y are real-valued of the studied phenomena, and $X=(X_1, \dots, X_p)$ forms the design matrix containing the explanatory variables. The term ϵ represents the noise vector, while σ refers to the noise level. The main goal is to estimate the regression coefficient vector β .

2.2. Sparse regularization methods

We examine methods for constructing models in high complexity, divided into sparse regularization approaches and early stepwise methods.

We explore various methods that develop a series of models with increasing complexity. These methods are grouped into sparse regularization approaches, which solve optimization problems to produce sparse solutions, and stepwise methods, which are some of the earliest approaches documented in the literature.

2.2.1. LASSO

The Least Absolute Shrinkage and Selection Operator, introduced by(16), is a shrinkage technique using an L1 penalty to reach the best fit for a linear regression model:

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

The LASSO estimator $\hat{\beta}^{LASSO}$ minimizes the function:

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

This method encourages sparsity by setting some coefficients to zero, leading to variable selection and a simpler model as λ increases(2).

2.2.2. Ridge Regression

In linear regression, collinearity within the design matrix X can cause unstable solutions when using squared-error loss. Ridge regression solve this problem by adding a penalty on the magnitude of the regression coefficients, which helps minimize the penalized residual sum of squares.

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

Where $\lambda \geq 0$ is a regularization coefficient that controls the degree of shrinkage applied to the estimated parameters. In greater shrinkage, larger λ 's values result were established by (13).

2.2.3. Elastic net

Elastic Net is a regression technique combining Lasso and Ridge regression, as proposed by (18). The objective function is:

$$\min_{\beta \in \mathbb{R}^p} J_{enet}(\beta) = \|Y - X\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2,$$

In this context, Y is the response variable, X is the matrix of predictors, and β represents the regression coefficients. The terms $\lambda_1 \|\beta\|_1$ (Lasso) and $\lambda_2 \|\beta\|_2^2$ (Ridge) control variable selection and coefficient shrinkage, respectively.

2.2.4. Adaptive LASSO and Adaptive Elastic net

Adaptive Lasso and Adaptive Elastic Net enhance Lasso and Elastic Net by incorporating adaptive versions, as introduced by (7) and (19). The Adaptive Lasso penalty is defined as

$$\rho_{adalasso}(|\beta_j|) = w_j |\beta_j| = w_j \rho_{lasso}(|\beta_j|)$$

And for the Adaptive Elastic Net as

$$\rho_{adanet}(|\beta_j|) = w_j |\beta_j| + \frac{\lambda_2}{\lambda_1} \beta_j^2$$

Here, w_j is defined as $|\beta_j^{init}|^{-q}$, with β^{init} is an initial solution from methods like Least-squares, and q takes 1 or 2. For Adaptive Lasso, $w_j=1/|\beta_j^{LS}|$ and $q=1$ aligns with the Garrote method (4).

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

- Adaptive Lasso:

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

- Adaptive Elastic Net:

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

(2) Demonstrate these estimators with $\lambda=2$, $\lambda_1=2$, and $\lambda_2=0.1$. The condition for all β components to be zero is

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

Where $\lambda^{(1)}$ is Lasso's parameter at which step the first feature will be included in the selection.

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

SCAD and MCP represents a quadratic spline penalty functions introduced by (17)and (8), respectively, to address the bias of Lasso. MCP penalty is defined as:

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

While SCAD is defined as:

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

In this case, λ tunes sparsity similar to Lasso, while $\gamma > 1$ and $a > 2$ modify the estimator's bias. When γ approaches 1, MCP approximates restricted least-squares, in contrast to SCAD. For an orthogonal design matrix X , MCP corresponds to the Firm Shrinkage estimator.

MCP approaches restricted least-squares as γ nears 1, making it simpler to solve than SCAD. With an orthogonal design matrix X , MCP corresponds to the Firm Shrinkage estimator by (6). Firm Shrinkage and SCAD penalties are defined as:

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

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

Those methods combine soft and hard thresholds. SCAD is more biased than MCP, especially when a nears its lower limit, and

includes subsets I_0 (null components), I_P (penalized components), and I_N (non-penalized components).

2.3. Cross Validation:

Cross-validation is a technique used to evaluate the performance of a predictive model and assess its ability to generalize to new, unseen data. It involves partitioning the available dataset into multiple subsets and using them for training and testing the model in a controlled and rigorous manner. The primary goal of cross-validation is to estimate how well the model will perform on independent data, thereby helping to avoid overfitting and providing a more reliable measure of the model's effectiveness.

2.3.1. V-Fold Cross Validation:

K-fold cross-validation is a technique for evaluating predictive models. The dataset is divided into k subsets or folds as we see:

$$CV - V(\hat{\beta}) = \frac{1}{V} \sum_{i=1}^n \|Y - X\hat{\beta}^{(-v)}\|^2$$

The model is trained and evaluated k times, using a different fold as the validation set each time. Performance metrics from each fold are averaged to estimate the model's generalization performance. This method aids in model assessment, selection, and hyper-parameter tuning, providing a more reliable measure of a model's effectiveness (5). In each set (fold) training and the test would be performed precisely once during this entire process. It helps us to avoid overfitting. As we know when a model is trained using all of the data in a single shot and give the best performance

accuracy. To resist this k-fold cross-validation helps us to build the model is a generalized one (11).

To achieve this K-Fold Cross Validation, we have to split the data set into three sets, Training, Testing, and Validation, with the challenge of the volume of the data (10), here Test and Train data set will support building model and hyper-parameter assessments.

In which the model has been validated multiple times based on the value assigned as a parameter and which is called K and it should be an INTEGER (14).

Make it simple, based on the K value, the data set would be divided, and train/testing will be conducted in a sequence way equal to K time. The most common K-folds Cross validation are 5-Folds cross validation and 10-Folds Cross validation, which we are going to use in this paper.

2.3.2. Generalized Cross Validation:

Generalized Cross Validation (GCV) is a technique used in the context of regularization, particularly in linear regression and related models, to estimate the performance of a model and tune its hyper-parameters. It is an extension of the more commonly used Cross Validation (CV) technique. In traditional Cross Validation (CV), the dataset is split into training and validation sets, and the model is trained on the training set while being evaluated on the validation set. This process is repeated with various subsets of the data, and the average performance metric is utilized to estimate the model's generalization capability.

Generalized Cross Validation, on the other hand, does not require splitting the dataset into training and validation sets as we see in:

$$GCV(\hat{\beta}) = \frac{n}{n - \widehat{df}} \|Y - X\hat{\beta}\|^2$$

Where n represents the number of observations in the dataset, and $\widehat{df}$ refers to number of parameters estimated.

Instead, it uses the information present in the residuals of the model (the differences between the actual target values and the model's predictions) to estimate the model's performance and tune hyper-parameters. This makes GCV computationally more efficient and potentially less prone to overfitting, as it utilizes the entire dataset during the evaluation process (3).

2.3.3. Leave One Out Cross Validation:

Leave-One-Out Cross Validation (LOOCV), a criteria introduced by (15) and (1), can be seen as a special case of V-fold cross-validation (CV-V) where the number of folds equals the total number of observations. At each iteration, the estimator of β is computed without the i -th observation, and the empirical risk is then evaluated on this left-out observation. This approach corresponds to the following criterion:

$$LOOCV(\hat{\beta}) = \frac{1}{n} \sum_{i=1}^n \|Y - X\hat{\beta}^{(-i)}\|^2$$

3. Main Results:

3.1. Simulation:

Our scenarios for the simulation, for n exceeds p , where p is fixed as the regression coefficient vector β is defined as: (2, 0, 0, 4, and 0), containing two non-zero coefficients. In case of high dimensional setup, n and p variate proportional according to the following relation: $p = n^d$ ($d=1.3$) with five nonzero coefficients (10, 20, 30, 40, 50). The design matrix X will be considerate as orthogonal or general with a correlation matrix Σ , where ρ is taken from {0, 0.2, 0.4, 0.5, 0.6, and 0.8}. X is generated once and fixed for all the 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, and 2}. The number of observations n varies over {20, 40, 60, 80 and 100}. 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 compare the score between the criterions, where a good selection is indicated by a minimum score.

The following table precise the choice of parameters for the model selection methods:

Table 1. Parameters for model selection methods

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	$a=3.7$

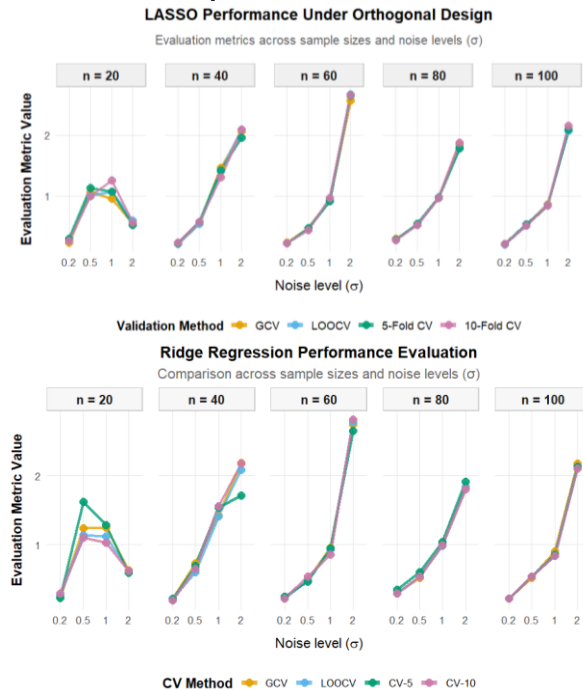
Source: collected from different literature reviews

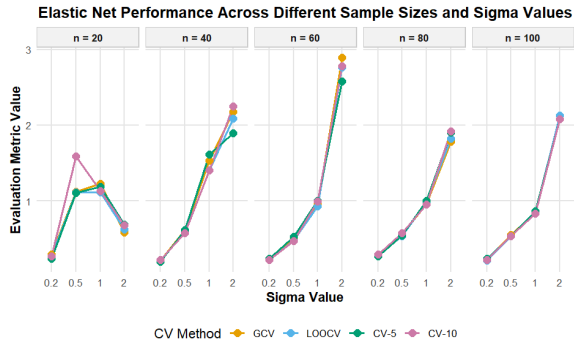
3.2. Orthogonal Design Analysis:

3.2.1. Detailed Analysis for Orthogonal Features Design Regular Settings:

The comparison of the CV criterion for the model selection methods under various sample sizes and variances values for the orthogonal case when the sample size is exceeds number of variables:

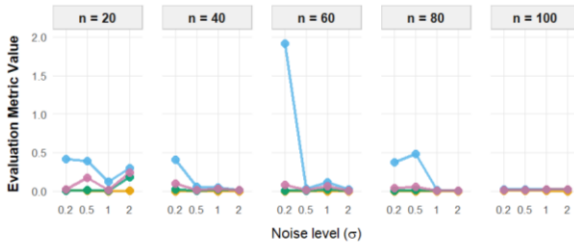
Fig 1. Cross Validation of Regularization methods with respect to sample size (n) and variance σ when $n > p$ for the general design case with correlation $\rho = 0$ between the variables.





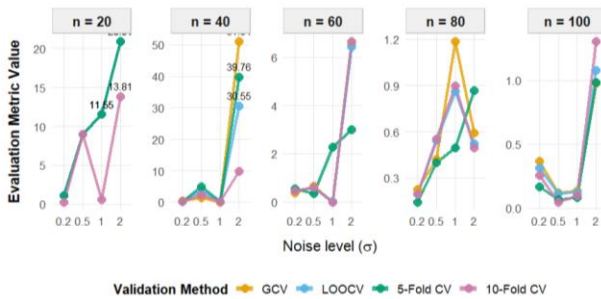
MCP Performance Under Orthogonal Design

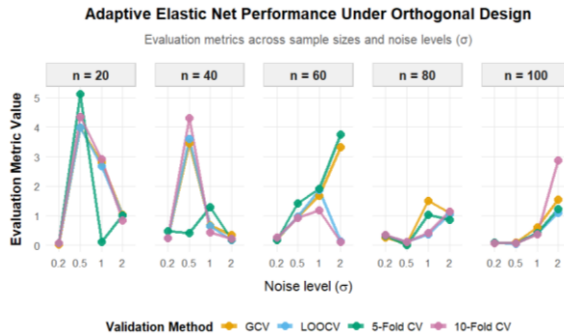
Evaluation metrics across sample sizes and noise levels (σ)



Adaptive LASSO Performance Under Orthogonal Design

Evaluation metrics across sample sizes and noise levels (σ)



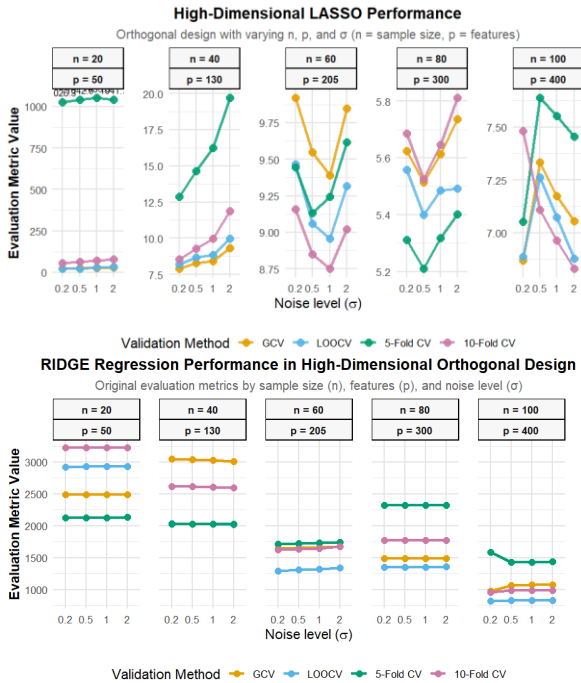


In evaluating model selection criteria (GCV, LOOCV, CV5, CV10) for LASSO, Ridge, Elastic Net, MCP, Adaptive LASSO, and Adaptive Elastic Net under orthogonal design with varying sample sizes (n) and noise levels (σ), LOOCV and CV10 stand out as the most robust choices across most models. For small sample sizes ($n \leq 40$), LOOCV and CV10 generally produce lower errors and greater stability, especially under moderate to high noise ($\sigma \geq 0.5$), with CV10 performing well in Adaptive LASSO and Adaptive Elastic Net at high σ . For moderate to large sample sizes ($n \geq 60$), differences between criteria diminish, but LOOCV and CV10 maintain a slight advantage in stability, while GCV offers computational efficiency as a practical alternative, especially for LASSO and Ridge. However, GCV is less dependable for MCP due to significant error underestimation at small n , where CV5 and CV10 are preferred. CV5 shows potential in specific cases, such as Elastic Net at small n and low σ , but can be unstable in Adaptive Elastic Net at high σ . Overall, LOOCV and CV10 provide the best combination of accuracy and stability, with model-specific recommendations favoring CV10 for Adaptive LASSO and LOOCV for Adaptive Elastic Net under high noise conditions.

3.2.2. Detailed Analysis for Orthogonal Features Design High Dimensional Settings:

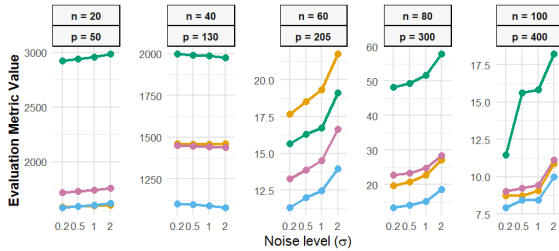
This part analyze and discuss the performance of CV for each model selection under various sample sizes and variances values for the orthogonal case when the number of features is exceeds the number of observations:

Fig 2. Cross Validation of Regularization methods with respect to sample size (n) and variance σ when $p > n$ for the general design case with correlation $\rho = 0$ between the variables.



Elastic Net Performance in High-Dimensional Orthogonal Design

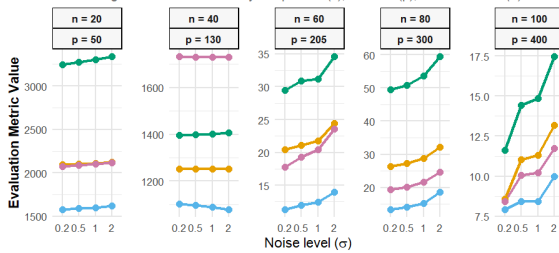
Original evaluation metrics by sample size (n), features (p), and noise level (σ)



Validation Method — GCV — LOOCV — 5-Fold CV — 10-Fold CV

Adaptive Elastic-Net Performance in High-Dimensional Orthogonal Design

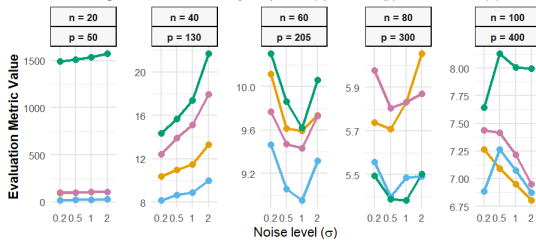
Original evaluation metrics by sample size (n), features (p), and noise level (σ)



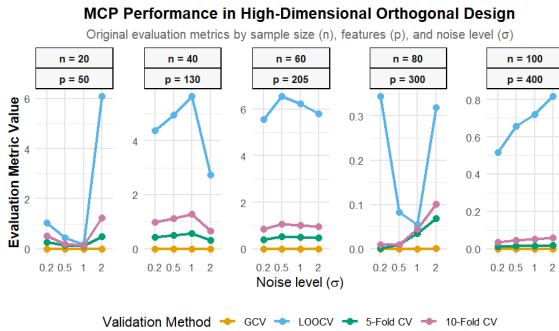
Validation Method — GCV — LOOCV — 5-Fold CV — 10-Fold CV

Adaptive LASSO Performance in High-Dimensional Orthogonal Design

Original evaluation metrics by sample size (n), features (p), and noise level (σ)



Validation Method — GCV — LOOCV — 5-Fold CV — 10-Fold CV



In high-dimensional orthogonal design settings for LASSO, Ridge, Elastic Net, MCP, Adaptive Lasso, and Adaptive Elastic Net, LOOCV emerges as the most robust and stable criterion for model selection across varying sample sizes ($n = 20\text{--}100$) and noise levels ($\sigma = 0.2\text{--}2$), particularly performing well for $n \geq 40$ and in Ridge, Elastic Net, Adaptive Lasso, and Adaptive Elastic Net, where it consistently achieves the lowest prediction errors. For small sample sizes ($n = 20$), LOOCV outperforms in most models, though GCV is competitive in LASSO and MCP under higher noise, despite its risk of over-optimism in MCP. CV5 shows significant instability at small n , especially in LASSO, Adaptive Lasso, and Adaptive Elastic Net, but becomes viable for $n \geq 60$, notably in LASSO and MCP. CV10 is a balanced alternative, often outperforming in LASSO at moderate n and high σ , but generally trails LOOCV. GCV gains efficiency and reliability for large n (≥ 80), particularly in LASSO and Adaptive Elastic Net, but underperforms in Ridge. Overall, LOOCV is recommended for its consistent accuracy and stability, with GCV as a computationally efficient option for large datasets and CV5 or CV10 as context-specific alternatives for moderate to large n .

3.2.3. Orthogonal Design Conclusion:

The following table resume the performance of evaluation criterions for each model selection method as a conclusion for the orthogonal case:

Table 2. The performance of evaluation criterions for each model selection under the orthogonal design matrix

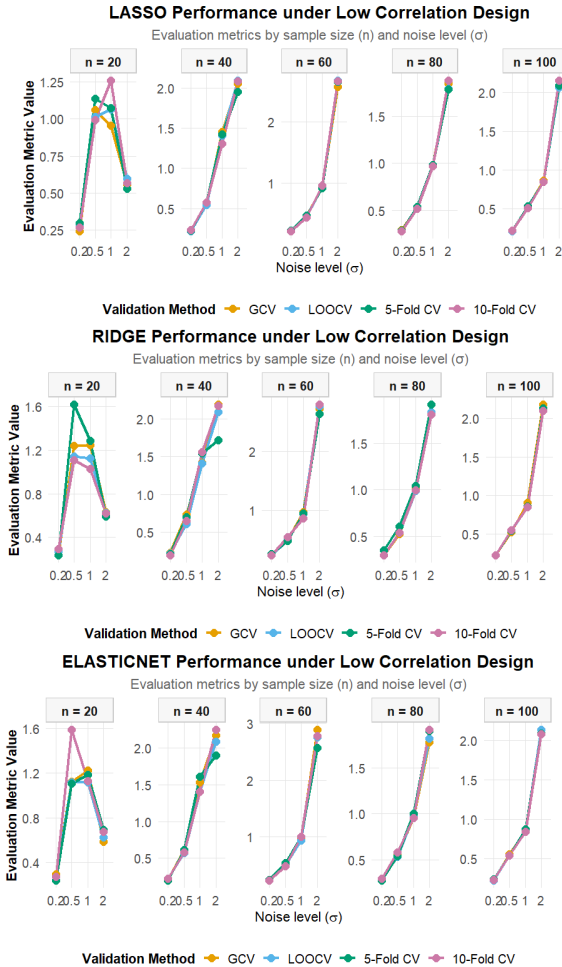
Model	Sample Size	Low σ (0.2)	Moderate-High σ (0.5-2)	Notes
LASSO	Small	LOOCV	LOOCV	GCV viable for high σ , high-dim
	Moderate	CV10	CV10	GCV strong in high-dim at $n=40$
	Large	CV5	CV10	GCV efficient in high-dim
Ridge	Small	CV5	LOOCV	CV10 as alternative for regular
	Moderate	LOOCV	LOOCV	CV10 competitive in regular
	Large	LOOCV	LOOCV	CV10 viable in high-dim
Elastic Net	Small	LOOCV	LOOCV	CV5 for regular, GCV for high-dim
	Moderate	LOOCV	LOOCV	CV10 as alternative
	Large	LOOCV	LOOCV	GCV efficient in high-dim
MCP	Small	CV5	CV5	CV10 as alternative
	Moderate	CV5	CV5	CV10 competitive
	Large	CV5	CV5	All criteria similar at $n=80+$
Adaptive Lasso	Small	LOOCV	LOOCV	CV10 viable for regular
	Moderate	LOOCV	LOOCV	CV5 strong in high-dim at $n=60$
	Large	LOOCV	GCV	CV5 competitive in high-dim
Adaptive Elastic Net	Small	LOOCV	LOOCV	GCV viable in high-dim
	Moderate	LOOCV	LOOCV	CV5 for regular at $n=40$
	Large	LOOCV	LOOCV	GCV efficient in high-dim

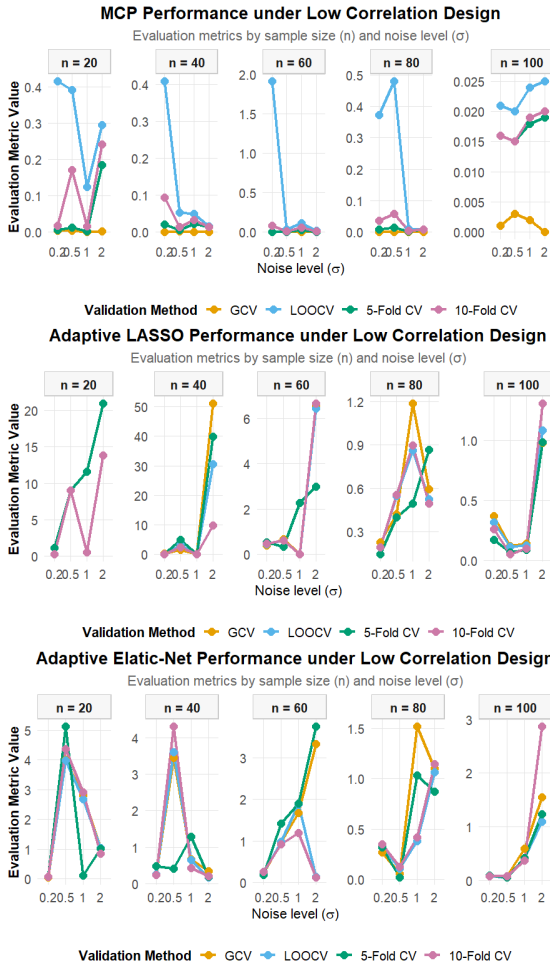
3.3. Low Correlation Design Analysis:

3.3.1. Detailed Analysis for Low correlation case Regular settings:

This part analyze and discus the performance of CV for each model selection 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:

Fig 3. Cross Validation of Regularization methods 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.





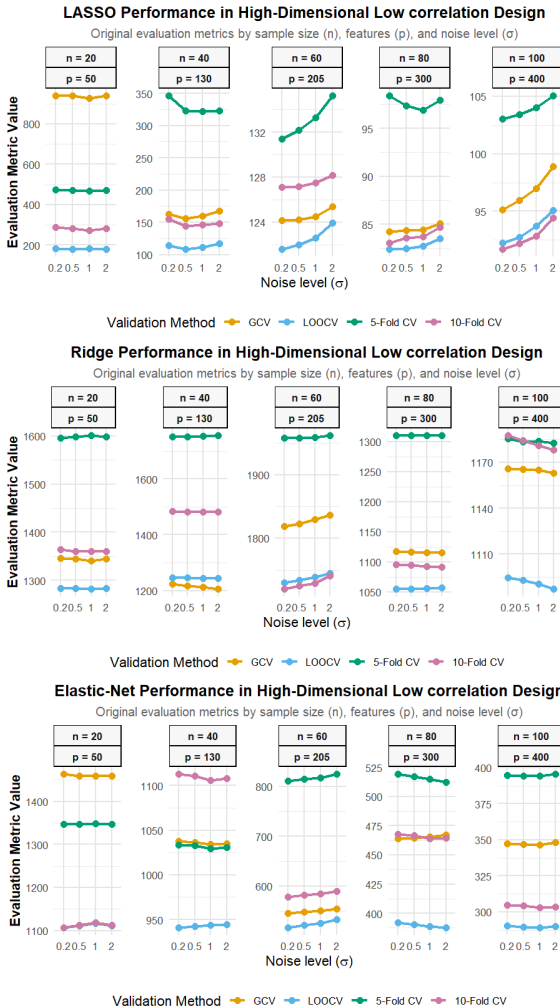
In low correlated feature settings for LASSO, Ridge, Elastic Net, MCP, Adaptive LASSO, and Adaptive Elastic Net, CV10 emerges as the most robust and consistent model selection criterion across varying sample sizes ($n = 20-100$) and noise levels ($\sigma = 0.2-2$), particularly excelling in LASSO, Ridge, Elastic Net, and Adaptive Elastic Net for moderate to high noise ($\sigma = 0.5-1$) and

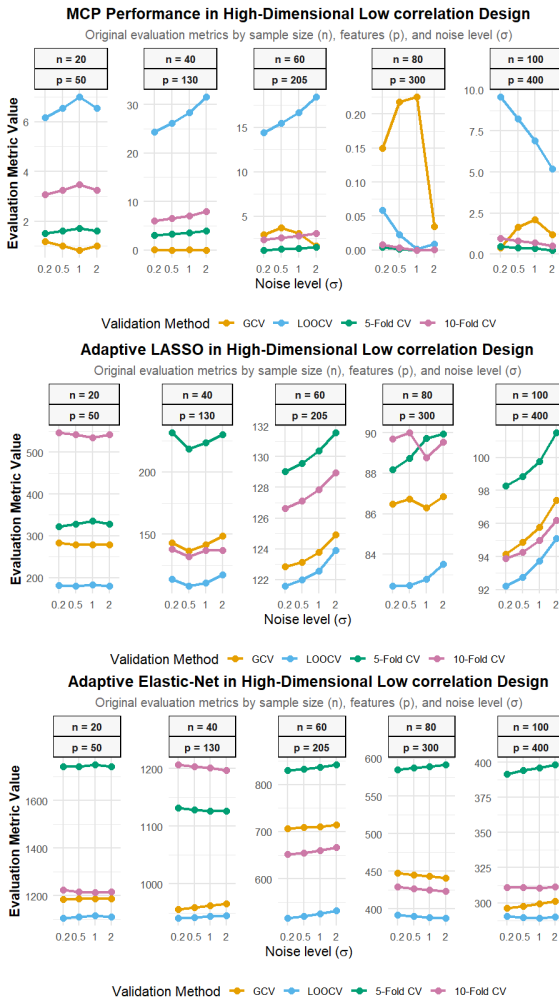
moderate to large sample sizes ($n \geq 40$). LOOCV is a strong alternative, performing well in Adaptive Elastic Net and Ridge at high noise ($\sigma = 2$) and large n , but it is less reliable for MCP due to significant instability, especially at small to moderate n (e.g., 1.915 at $n = 60$, $\sigma = 0.2$). CV5 is highly effective for MCP across all settings, yielding near-zero errors alongside GCV, and is also competitive in LASSO, Ridge, and Elastic Net at high noise ($\sigma = 2$) or low noise ($\sigma = 0.2$) for moderate n . GCV is computationally efficient and dominates in MCP with extremely low errors, but its potential for overfitting warrants caution; it is also viable for large n in LASSO, Elastic Net, and Adaptive Elastic Net at low σ . Adaptive LASSO exhibits extreme error variability at small n , making CV10 critical for stability. Low correlation increases error variability, particularly for Adaptive LASSO and MCP's LOOCV, emphasizing CV10 and CV5 as preferred choices for robust performance in these settings.

3.3.2. Detailed Analysis for low correlated Features Design High Dimensional settings:

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

Fig 4. Cross Validation of Regularization methods with respect to sample size (n) and variance σ when $p > n$ for the general design case with maximum correlation $\rho = 0.4$ between the variables.





In high-dimensional settings with low correlated features, the evaluation of model selection criteria for LASSO, Ridge, Elastic Net, MCP, Adaptive LASSO, and Adaptive Elastic Net reveals distinct performance patterns across sample sizes ($n = 20\text{--}100$) and noise levels ($\sigma = 0.2\text{--}2$). LOOCV emerges as the most robust

criterion for LASSO, Elastic Net, Adaptive LASSO, and Adaptive Elastic Net, consistently yielding the lowest prediction errors, particularly at small to moderate n (e.g., LASSO: 181.176 at $n = 20$, $\sigma = 0.2$; Elastic Net: 290.629 at $n = 100$, $\sigma = 0.2$). CV10 is a strong alternative, excelling in LASSO and Adaptive Elastic Net at larger n (e.g., LASSO: 91.720 at $n = 100$, $\sigma = 0.2$) and balancing computational efficiency. For MCP, CV5 dominates with near-zero errors at $n = 80$ (e.g., 0.004 at $\sigma = 0.2$), while LOOCV is unreliable due to high instability (e.g., 31.542 at $n = 40$, $\sigma = 2$). Ridge regression struggles significantly, with errors orders of magnitude higher than LASSO and MCP (e.g., LOOCV: 1055.399 at $n = 80$, $\sigma = 0.2$), where LOOCV and CV10 remain the most stable. GCV is computationally efficient but often underperforms, except in MCP at small n (e.g., 0.002 at $n = 40$, $\sigma = 2$). Compared to standard settings, high-dimensionality amplifies error variability, especially at small n , underscoring LOOCV and CV10's superiority for stability in LASSO-based models and CV5's effectiveness for MCP.

3.3.3. Low correlation design Conclusion:

The following table resume the performance of CV criterions for each model selection As a conclusion for the low correlation case:

Table 3. The performance of evaluation criterions for each model selection for the low correlated features

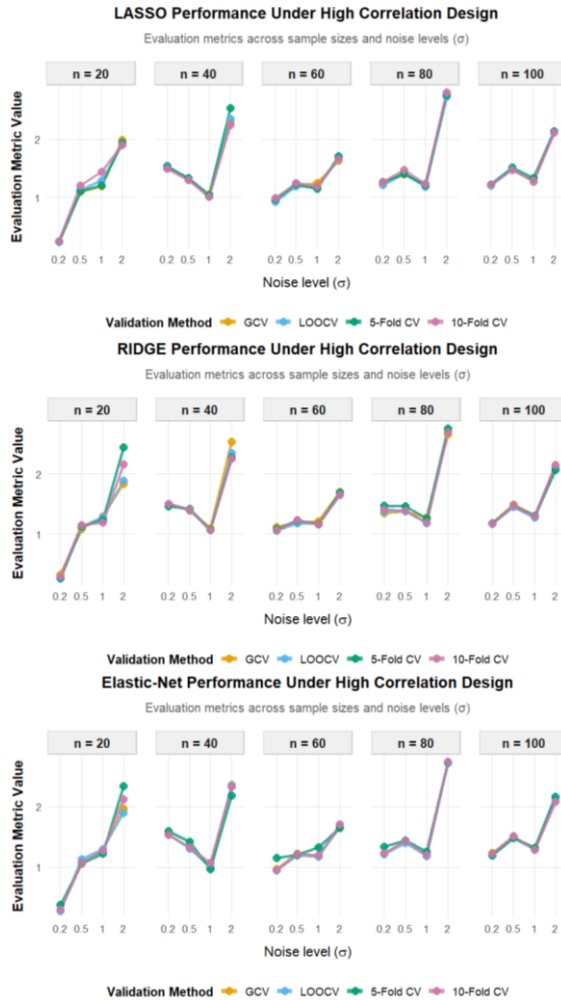
Model	Sample Size Category	Regular Setting	High-Dimensional Setting	Notes: Alternative Criterion
LASSO	Small (n = 20, 40)	CV10	LOOCV	Regular: LOOCV (robust but computationally intensive). High-D: CV10
	Moderate (n = 60)	CV10	LOOCV	Regular: LOOCV (stable for high σ). High-D: GCV
	Large (n = 80, 100)	CV10	CV10	Regular: LOOCV (viable for high σ). High-D: LOOCV
Ridge	Small (n = 20, 40)	CV10	LOOCV	Regular: LOOCV (preferred for high σ). High-D: CV10 ; GCV
	Moderate (n = 60)	CV10	CV10	Regular: LOOCV (stable alternative). High-D: LOOCV.
	Large (n = 80, 100)	CV10	LOOCV	Regular: LOOCV (robust alternative). High-D: CV10
Elastic Net	Small (n = 20, 40)	CV10	LOOCV	Regular: LOOCV (viable alternative). High-D: CV10 ; CV5
	Moderate (n = 60)	CV10	LOOCV	Regular: LOOCV (stable for high σ). High-D: GCV
	Large (n = 80, 100)	CV10	LOOCV	Regular: LOOCV (robust alternative). High-D: CV10
MCP	Small (n = 20, 40)	CV5	GCV	Regular: GCV (competitive with low errors). High-D: CV5
	Moderate (n = 60)	CV5	CV5	Regular: GCV (close in performance). High-D: CV10 (e.g., 2.405, $\sigma = 0.2$).
	Large (n = 80, 100)	CV5	CV5	Regular: GCV (viable with low errors). High-D: CV10
Adaptive LASSO	Small (n = 20, 40)	CV10	LOOCV	Regular: LOOCV (stable alternative). High-D: GCV ; CV10
	Moderate (n = 60)	CV10	LOOCV	Regular: LOOCV (preferred for high σ). High-D: GCV
	Large (n = 80, 100)	CV10	LOOCV	Regular: LOOCV (robust for high σ). High-D: CV10
Adaptive Elastic Net	Small (n = 20, 40)	CV10	LOOCV	Regular: LOOCV (preferred for high σ). High-D: GCV
	Moderate (n = 60)	CV10	LOOCV	Regular: LOOCV (stable alternative). High-D: CV10
	Large (n = 80, 100)	CV10	LOOCV	Regular: LOOCV (robust alternative). High-D: CV10

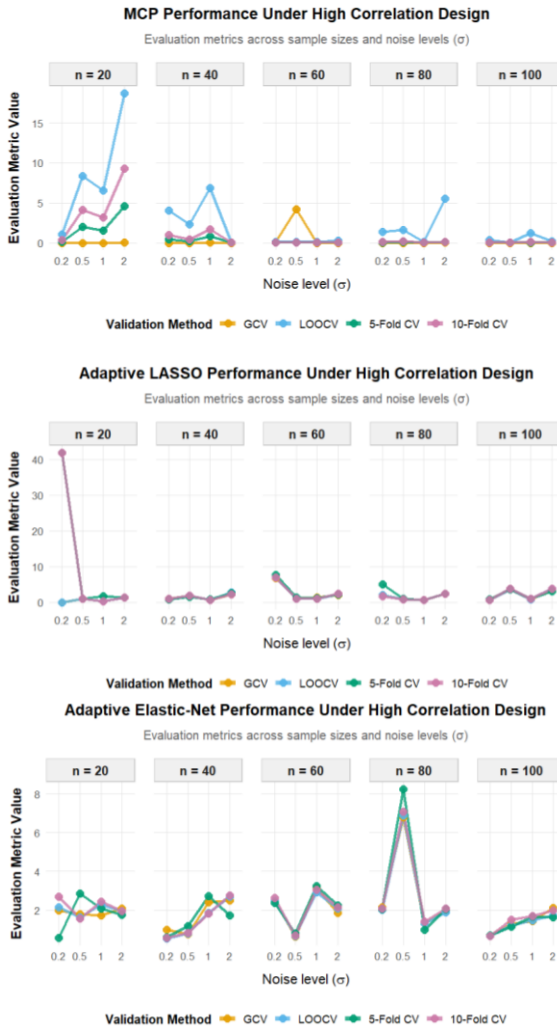
3.4. High Correlation Design Analysis:

3.4.1. Detailed Analysis for High correlated features Regular settings:

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

Fig 5. Cross Validation of Regularization methods with respect to sample size (n) and variance σ when $n > p$ for the general design case with maximum correlation $\rho = 0.8$ between the variables.





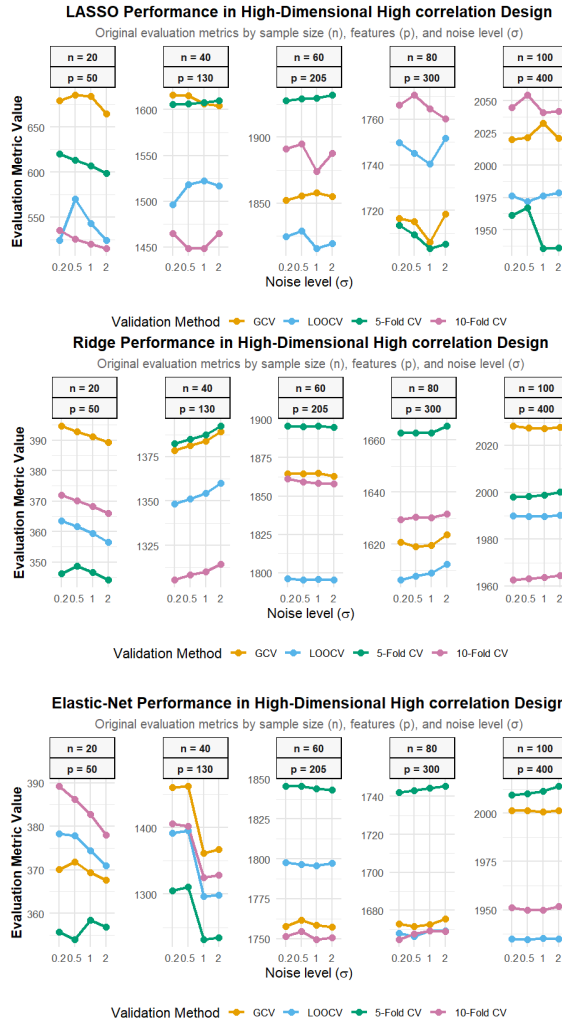
In high-dimensional settings with highly correlated features, the evaluation of model selection criteria for LASSO, Ridge, Elastic Net, MCP, Adaptive LASSO, and Adaptive Elastic Net across sample sizes (Small: $n = 20, 40$; Moderate: $n = 60$; Large: $n = 80, 100$) and noise levels ($\sigma = 0.2-2$) reveals nuanced performance

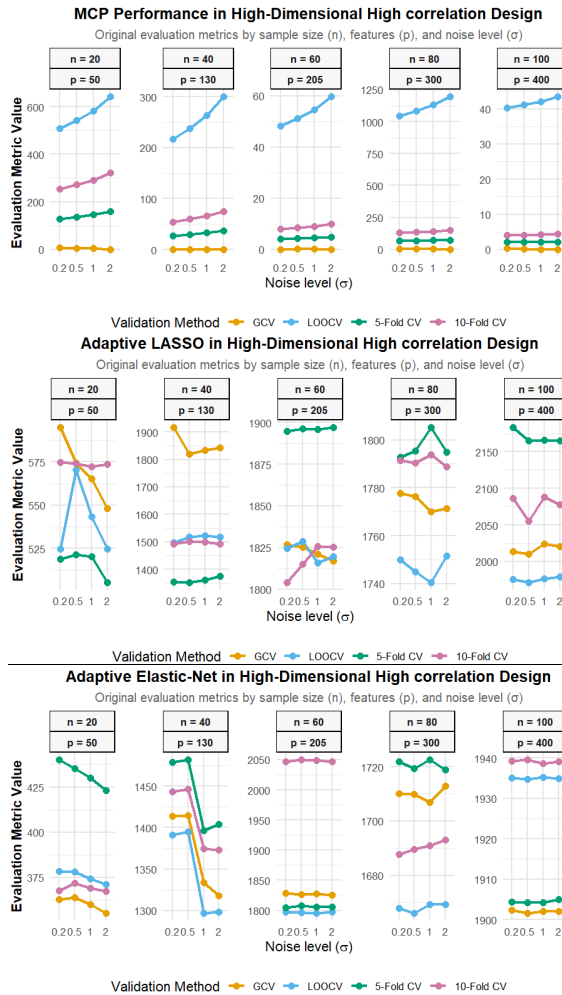
patterns compared to low correlated settings. LOOCV emerges as the most robust criterion for LASSO, Elastic Net, Adaptive LASSO, and Adaptive Elastic Net, particularly at moderate to large sample sizes, yielding the lowest prediction errors. CV10 is a strong alternative, excelling at high noise and larger sample sizes. For MCP, CV5 dominates, consistently achieving near-zero errors across moderate to large sample sizes, while LOOCV is unreliable due to extreme instability at small sample sizes. Ridge regression performs comparably to low correlated settings, with LOOCV and CV10 being stable choices. Compared to low correlated settings, high correlation increases error variability for LASSO, Elastic Net, and Adaptive variants due to challenges in sparse selection, but MCP's non-convex penalty mitigates this, making CV5 exceptionally effective. GCV is computationally efficient but less reliable, except for MCP at small sample sizes. These findings highlight LOOCV's superiority for stability in most models, CV5's strength for MCP, and the need for careful criterion selection in high correlation scenarios to balance stability and computational efficiency.

3.4.2. Detailed Analysis for high correlated Features Design:

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

Fig 6. Cross Validation of Regularization methods with respect to sample size (n) and variance σ when $p > n$ for the general design case with maximum correlation $\rho = 0.8$ between the variables.





In high-dimensional settings with highly correlated features, the evaluation of model selection criteria (GCV, LOOCV, CV5, CV10) for LASSO, Ridge, Elastic Net, MCP, Adaptive LASSO, and Adaptive Elastic Net across sample sizes ($n = 20, 40, 60, 80, 100$) and noise levels ($\sigma = 0.2-2$) reveals distinct performance patterns. For LASSO and Adaptive LASSO, LOOCV and CV10 are robust,

yielding the lowest errors at moderate to large sample sizes (e.g., LASSO: 1740.550 at $n = 80$, $\sigma = 1$), but errors increase with sample size, unlike low correlated settings where errors decrease. Ridge regression benefits from its L2 penalty, with LOOCV consistently stable across all sample sizes (e.g., 1606.440 at $n = 80$, $\sigma = 0.2$), performing better than LASSO in high correlation scenarios. Elastic Net and Adaptive Elastic Net balance sparsity and shrinkage, with LOOCV and CV10 excelling at larger sample sizes (e.g., Elastic Net: 1934.717 at $n = 100$, $\sigma = 0.5$), though they show higher errors than Ridge at small sample sizes. MCP stands out with CV5 and GCV achieving near-zero errors across most settings (e.g., CV5: 2.019 at $n = 100$, $\sigma = 0.2$), but LOOCV is unreliable due to instability (e.g., 1194.920 at $n = 80$, $\sigma = 2$). Compared to low correlated high-dimensional settings, high correlation increases error variability for LASSO, Adaptive LASSO, Elastic Net, and Adaptive Elastic Net, while MCP's non-convex penalty and Ridge's L2 regularization mitigate this effect. CV5 is recommended for MCP, LOOCV for Ridge, Elastic Net, and Adaptive Elastic Net, and CV10 for LASSO and Adaptive LASSO at moderate sample sizes, balancing stability and computational efficiency.

3.4.3. High correlation design conclusion:

The following table resume the performance of evaluation criterions for each model selection as a conclusion for the high correlation case:

Table 4. The performance of model selection methods for the high correlated features

Model	Sample Size Category	Regular Setting	High-Dimensional Setting	Notes: Alternative Criterion
LASSO	Small (n = 20, 40)	LOOCV	CV10	CV5 for Regular at high noise; LOOCV for High-Dimensional at low noise.
	Moderate (n = 60)	LOOCV	LOOCV	CV10 for Regular for efficiency; GCV for High-Dimensional for efficiency.
	Large (n = 80, 100)	LOOCV	CV5	CV5 for Regular for stability; LOOCV for High-Dimensional for robustness.
Ridge	Small (n = 20, 40)	LOOCV	CV5	CV10 for Regular for efficiency; LOOCV for High-Dimensional at high noise.
	Moderate (n = 60)	LOOCV	LOOCV	CV10 for Regular for efficiency; CV10 for High-Dimensional for stability.
	Large (n = 80, 100)	LOOCV	LOOCV	GCV for Regular for efficiency; CV10 for High-Dimensional for efficiency.
Elastic Net	Small (n = 20, 40)	LOOCV	CV5	CV10 for Regular at high noise; GCV for High-Dimensional for efficiency.
	Moderate (n = 60)	LOOCV	CV10	CV5 for Regular for stability; GCV for High-Dimensional for efficiency.
	Large (n = 80, 100)	LOOCV	LOOCV	CV5 for Regular for stability; CV10 for High-Dimensional for stability.
MCP	Small (n = 20, 40)	CV5	GCV	GCV for Regular for efficiency; CV5 for High-Dimensional for stability.
	Moderate (n = 60)	CV5	GCV	LOOCV for Regular for stability; CV5 for High-Dimensional for robustness.
	Large (n = 80, 100)	CV5	GCV	LOOCV for Regular for stability; CV5 for High-Dimensional to avoid underestimation.
Adaptive LASSO	Small (n = 20, 40)	LOOCV	CV5	CV10 for Regular for efficiency; LOOCV for High-Dimensional at moderate noise.
	Moderate (n = 60)	LOOCV	CV10	CV5 for Regular for stability; LOOCV for High-Dimensional for robustness.
	Large (n = 80, 100)	LOOCV	LOOCV	CV5 for Regular at high noise; GCV for High-Dimensional for efficiency.
Adaptive Elastic Net	Small (n = 20, 40)	LOOCV	GCV	CV10 for Regular for efficiency; CV10 for High-Dimensional for stability.
	Moderate (n = 60)	LOOCV	LOOCV	CV10 for Regular for efficiency; CV5 for High-Dimensional for stability.
	Large (n = 80, 100)	LOOCV	GCV	GCV for Regular for efficiency; CV5 for High-Dimensional for robustness.

4. Discussion

The evaluation of model selection criteria (GCV, LOOCV, CV5, CV10) for LASSO, Ridge, Elastic Net, MCP, Adaptive LASSO, and Adaptive Elastic Net across orthogonal, low, and high correlated feature settings in regular and high-dimensional contexts reveals distinct performance patterns driven by feature correlation and dimensionality.

In regular settings, LOOCV consistently delivers the best performance across all correlation statuses and sample sizes, achieving low errors for LASSO, Ridge for independent features, and Elastic Net for high correlated features.

In high-dimensional settings, orthogonal features favor LOOCV for LASSO at small samples and CV5 for Ridge at large samples, while CV5 performs best for MCP across all sample sizes. Low correlated settings show CV5 excelling for Elastic Net at large samples and for MCP, with errors decreasing with sample size.

High correlated settings see CV10 performing best for LASSO at small samples CV5 for MCP, and LOOCV for Ridge) across σ values, highlighting their robustness to multicollinearity. The most stable model-criterion pairing is Ridge with LOOCV, consistently achieving low errors across all settings and noise levels, due to its L2 regularization and LOOCV's low bias.

5. CONCLUSION

The analysis denied our starting hypothesis, the results shows that the best-performing model-criterion pairing, driven by minimal prediction errors, is dictated by sample size, dimensionality, correlation status, and noise levels, with Ridge paired with LOOCV emerging as the most stable combination across all scenarios, particularly in high-dimensional high correlated settings. LOOCV dominates in regular settings, CV10 excels for LASSO in high-dimensional orthogonal and low correlated data, CV5 for Elastic Net in low correlated large samples, and CV5 or GCV for MCP in high correlated high-dimensional contexts. Recommendations include adopting Ridge-LOOCV for robust performance, CV10 for LASSO in orthogonal settings, and CV5 for MCP in high correlated scenarios. Future research should prioritize applying these model-criterion pairings to real-world datasets from diverse domains, such as genomics, finance and medical imaging, investigating the performance of the evaluation criteria that depend on noise levels in simulation and real datasets. Prospects for advancing the field include exploring non-linear feature dependencies and noise interactions to develop more robust criteria, investigating moderate correlation strengths to bridge the gap between low and high correlated settings, and integrating Bayesian approaches to enhance model selection stability, ultimately ensuring scalable and reliable strategies across diverse and complex data landscapes.

6. Bibliography List :

[1] Allen, D. M. (1974). *The Relationship between Variable Selection and Data Agumentation and a Method for Prediction*. *Technometrics*, 16(1), 125–127. <https://doi.org/10.2307/1267500>

[2] Boisbunon, A. (2013). *Model selection: A decision-theoretic approach*, Rouen University. France.

[3] Bottegal, G. and P., & Gianluigi. (2018). *The generalized cross validation filter*. Elsevier, 90, 130--137.

[4] Breiman, L. (1995). *Better Subset Regression Using the Nonnegative Garrote*. *Technometrics*, 37(4), 373–384. <https://doi.org/10.2307/1269730>

[5] Browne, M. W. (2000). *Cross-Validation Methods*. *Journal of Mathematical Psychology*, 44(1), 108–132. <https://doi.org/10.1006/jmps.1999.1279>

[6] BRUCE, A. G., & GAO, H.-Y. (1996). *Understanding WaveShrink: Variance and bias estimation*. *Biometrika*, 83(4), 727–745. <https://doi.org/10.1093/biomet/83.4.727>

[7] Chung, P. D. (2006). *The Adaptive Lasso and Its Oracle Properties* Hui Zou , JASA.

[8] Fan, J., & Li, R. (2001). *Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties*. *Journal of the American Statistical Association*, 96(456), 1348–1360.

[9] Fourdrinier, D., & Wells, M. (1994). *Comparisons of selection rules of a regression model: A decision approach*. *Comptes Rendus de l'Académie Des Sciences. Série I*.

[10] Fourdrinier, D., & Wells, M. T. (1995). *Loss Estimation for Spherically Symmetrical Distributions*. *Journal of Multivariate Analysis*, 53(2), 311–331. <https://doi.org/10.1006/jmva.1995.1039>

[11] Fourdrinier, D., & Wells, M. T. (2012). *On Improved Loss Estimation for Shrinkage Estimators*. *Statistical Science*, 27(1). <https://doi.org/10.1214/11-STS380>

[12] Gohain, P. B., & Jansson, M. (2023). *Robust Information Criterion for Model Selection in Sparse High-Dimensional Linear Regression Models*. *IEEE Transactions on Signal Processing*, 71, 2251–2266. *IEEE Transactions on Signal Processing*. <https://doi.org/10.1109/TSP.2023.3284365>

[13] Hoerl, A. E. and K., & Robert W. (1970). *Ridge regression: Applications to nonorthogonal problems*. Taylor & Francis, 12, 69-82.

[14] Rodriguez, J. D. and P., & Aritz and Lozano, J. A. (2009). *Sensitivity analysis of k-fold cross validation in prediction error estimation*. *IEEE*, 32, 569--575.

[15] Stone, M. (1974). *Cross-Validatory Choice and Assessment of Statistical Predictions*. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 36(2), 111–133. <https://doi.org/10.1111/j.2517-6161.1974.tb00994.x>

[16] Tibshirani, R. (1996). *Regression Shrinkage and Selection Via the Lasso*. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>

[17] Zhang, C.-H. (2010). *Nearly unbiased variable selection under minimax concave penalty*. *The Annals of Statistics*, 38(2). <https://doi.org/10.1214/09-AOS729>

[18] Zou, H., & Hastie, T. (2005). *Regularization and Variable Selection Via the Elastic Net*. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2), 301–320.
<https://doi.org/10.1111/j.1467-9868.2005.00503.x>

[19] Zou, H., & Zhang, H. H. (2009). *On the Adaptive Elastic-net with a Diverging Number of Parameters*. *University of Minnesota*.