

Comparison of Non-Convex Variable Selection Criteria in High-dimensional Data with Count Response

Saheed Olalekan Jabaru¹, Waheed Babatunde Yahya², Kamoru Jimoh³

²Department of Statistics, University of Ilorin, Nigeria

^{1,3}Department of Physical Sciences, Al-Hikmah University, Ilorin, Nigeria
saheedjabar@alhikmah.edu.ng¹

Abstract

Background: High-dimensional datasets with count responses present unique challenges in variable selection due to the large number of predictors and potential collinearity. While significant work has been done on selecting predictors from high-dimensional data with metrical covariates and Gaussian responses, there is a gap in comparing the performances of SCAD (Smoothly Clipped Absolute Deviation) and MCP (Minimax Concave Penalty) across different tuning parameters for count responses. Such comparisons, especially with k -fold cross-validation, are rare, highlighting the need for this study. **Aim:** This study aims to compare the effectiveness of SCAD and MCP in high-dimensional datasets where the response variable is a count, specifically for various tuning parameters and cross-validation techniques. **Method:** A simulation study was conducted using datasets with 250 predictors and a sample size of 100 modelled by a Poisson distribution (without overdispersion). The performance of SCAD and MCP was evaluated based on selection accuracy, cross-validation error, and computational efficiency. **Results:** The results showed that SCAD's performance at tested tuning parameters was similar to the standard tuning parameter ($\gamma = 3.7$) at 10-fold cross-validation only. For MCP, the variable selection performance at tuning parameters $\gamma = 4$ and $\gamma = 5$ was comparable to the standard ($\gamma = 3$), while other values differed significantly. Additionally, 10-fold cross-validation yielded the lowest cross-validation errors, and MCP selected the fewest predictors. **Conclusion:** Non-convex variable selection methods, particularly SCAD and MCP, are recommended for their robustness and accuracy in identifying relevant predictors and enhancing predictive capabilities in high-dimensional datasets with count responses.

Keywords: High-dimensional, Non-convex, k -fold, Count

1. Introduction

Count data are non-negative integers representing the occurrence frequency of discrete events, such as the number of children born to a woman or the daily accidents on a busy road (Cameron & Trivedi, 2013). These data are crucial in various fields like medicine, agriculture, and social sciences, but their analysis poses challenges like non-normality, heteroscedasticity, and an excess of zero values (Zhou et al., 2014). High-dimensional data, where the number of predictors exceeds the number of observations ($p > n$), further complicates analysis, necessitating techniques like penalized estimation to improve model accuracy and manage predictor correlations (Xiong, Tan & Kumar, 2006; Shao & Deng, 2012).

In high-dimensional settings, selecting significant predictors is essential to avoid overfitting and enhance model reliability (Mallick & Yi, 2013). Traditional methods like stepwise and best subset selection are often computationally intensive and prone to ignoring stochastic errors (Li & Liang, 2008). Penalized methods, with appropriate regularization parameters, address these issues by simultaneously estimating coefficients and selecting predictors (Fan & Li, 2001). Cross-validation, especially k-fold, is used to validate predictive models, ensuring robustness and replicability of results (Brimacombe, 2014). This study focuses on comparing significant predictors for a count response variable using non-convex penalties across diverse k-fold cross-validations and tuning parameter values (Li, 2015; Emmert-Streib & Dehmer, 2019; Li & Yang, 2021).

2. Related Works

This research focuses on applying various variable selection criteria to high-dimensional data with count data as the response variable. Count data, characterized by positive integers with a variance that increases with the mean, pose unique analytical challenges (Emmert-Streib & Dehmer, 2019). Technological advancements have facilitated the collection of extensive data with numerous features, prompting a need for effective dimension reduction techniques to extract significant features and eliminate redundancy (Jia

et al., 2022). The selection of relevant variables is crucial in fields like machine learning, pattern recognition, and data mining (Bonney, 2020; Idrees & Kamal, 2021).

LASSO and its extensions have been evaluated for their strengths and weaknesses as variable selection criteria, with findings suggesting that dataset attributes significantly influence predictor selection (Emmert-Streib & Dehmer, 2019). Studies have tested different regularization parameters to identify the best penalties for sparse models (Idrees & Kamal, 2021). Comparative analyses have shown that SCAD and MCP penalties perform well under various conditions, often yielding similar results (Bonney, 2020). These insights underscore the importance of using cross-validation and appropriate penalties to enhance the reliability and accuracy of high-dimensional data models (Jia et al., 2022).

Recent advancements in variable selection for high-dimensional data continue to focus on improving the accuracy and efficiency of model selection techniques. Key areas of exploration include penalization methods, machine learning algorithms, and robust cross-validation approaches. Penalization methods remain a cornerstone in variable selection for high-dimensional data. The LASSO (Least Absolute Shrinkage and Selection Operator) and its variants, such as Elastic Net, SCAD (Smoothly Clipped Absolute Deviation), and MCP (Minimax Concave Penalty), have been extensively studied for their ability to handle sparsity and multicollinearity. Emmert-Streib and Dehmer (2019) highlighted the strengths and limitations of LASSO-based models, emphasizing the importance of dataset characteristics in influencing predictor selection. Recent work by Bonney (2020) and Idrees and Kamal (2021) explored the comparative performance of these methods, finding that SCAD and MCP often provide better model stability and prediction accuracy under certain conditions.

Cross-validation remains a critical component in validating the robustness of predictive models. Recent studies emphasize the importance of selecting appropriate cross-validation methods to mitigate overfitting and ensure replicability. Idrees and Kamal (2021) demonstrated the effectiveness of k-fold cross-validation in evaluating the predictive performance of different penalization techniques. Their findings suggest that

10-fold cross-validation offers a consistent and reliable measure for variable selection, especially in the presence of high-dimensional data.

In summary, majority of the previous research works on dimensionality reduction of high-dimensional data focused on Gaussian/Non-Gaussian responses with continuous covariates. However, in most real-life situations, a greater number of categorical predictors have influence on the response variable. Therefore, this research work compares the variable selection ability of non-convex criteria with count responses at different values of tuning parameters.

3. Methodology

3.1 Data Simulation Process

Data for this study (response, metrical and categorical) variable were simulated. The metrical predictors were simulated using normal, uniform, and Poisson distributions. Examples of such metrical predictors in real-life data include age, birth intervals, number of births in the last five years, income, body mass index, weight, etc.

Similarly, categorical predictors were simulated as factor variables with two to six categories. Examples of such categorical variables in real-life data might include level of education, occupation, marital status, gender, residence, wealth index, etc. The categorical predictors were converted into dummy variables using one-hot encoding. One-hot encoding, also known as dummy encoding, is a technique for representing categorical variables as binary vectors in machine learning. It is applied when categorical variables need to be included in a model that only accepts numerical inputs.

Lastly, the response variable was simulated as a function of the predictors assuming Poisson distribution with parameter $\lambda_i = e^{(\beta_0 + \sum_{i=1}^n \beta_i x_i)}$. Examples of such response variable in the health and environmental sectors are the number of people that contacted COVID-19, number of children ever born, number of people diagnosed of typhoid fever in a country, etc.

3.2 Variable selection

According to Desboulets (2018), the determination of parsimonious set of variables can be achieved through test-based, screening-based or penalty-based procedure. However, penalty-based procedures that have oracle properties (i. e. has ability to identify the true non-zero coefficients) were used in this study. They are SCAD and MCP.

(a) The SCAD Method

This method proposed by Fan and Li (2001) sets a boundary for the penalty function to reduce bias (Li, 2015). The estimate of model parameters for a count response variable, assuming Poisson distribution is given by:

$$J(\beta) = - \sum_{i=1}^n \left(y_i \log(\lambda_i) - \lambda_i + \sum_{i=1}^n \rho(\beta_i, \gamma, \lambda) \right) \quad (1)$$

where:

$$\rho(\beta_i, \gamma, \lambda) = \begin{cases} \lambda |\beta_i|, & \text{if } |\beta_i| \leq \lambda \\ \frac{\gamma \lambda |\beta_i| - (\beta_i^2 + \lambda^2)}{2(\gamma - 1)}, & \text{if } \lambda < |\beta_i| \leq \gamma \lambda \\ \frac{\lambda^2(\gamma + 1)}{2}, & \text{if } |\beta_i| > \gamma \lambda \end{cases} ; \gamma > 2 \quad (2)$$

$J(\beta) =$ SCAD estimate of $\hat{\beta}$;

γ is a tuning parameter that controls the shape of the penalty function ;

$y_i =$ observed count for the i^{th} data point

$\lambda_i =$ predicted count for the i^{th} data point based on the fitted model

$\beta_i =$ value of the coefficient for a specific predictor or feature.

$\lambda =$ regularization parameter controlling the strength of the regularization terms

$\rho(\beta_i, \gamma, \lambda)$ has the following conditions:

- When $|\beta_i| \leq \lambda$, the penalty term is proportional to $|\beta_i|$.
- When $\lambda < |\beta_i| \leq \gamma \lambda$, the penalty term increases in a convex, smoothly clipped manner, encouraging sparsity by softly shrinking coefficients.

- When $|\beta_i| > \gamma\lambda$, the penalty term remains constant, effectively setting the coefficient to zero.

(b) MCP Method

Introduced by Zhang (2010), MCP criterion aimed at obtaining a sparse model with less biased regression parameters in which the penalty function varies with the amount of the covariate coefficient. The MCP estimate of $\hat{\beta}$ for a count response variable with Poisson distribution is given by:

$$J(\beta) = - \sum_{i=1}^n \left(y_i \log(\lambda_i) - \lambda_i + \sum_{i=1}^n \rho(\beta_i, \gamma, \lambda) \right) \quad (3)$$

where:

$$\rho(\beta_i, \gamma, \lambda) = \begin{cases} 0, & \text{if } |\beta_i| \leq \gamma\lambda \\ \frac{|\beta_i| - (\gamma\lambda^2 + |\beta_i|^2)}{2(\gamma\lambda - 1)}, & \text{if } \gamma\lambda < |\beta_i| \leq \lambda \\ \frac{\lambda^2(\gamma + 1)}{2}, & \text{if } |\beta_i| > \lambda \end{cases} ; \quad \gamma > 1 \quad (4)$$

$J(\beta) = \text{MCP estimate of } \hat{\beta}$;

$\gamma, y_i, \lambda_i, \beta_i, \lambda$ are as defined above

$\rho(\beta_i, \gamma, \lambda)$ is explained as follows:

- When $|\beta_i| \leq \gamma\lambda$, the penalty term is zero, meaning that no penalty is applied.
- When $\gamma\lambda < |\beta_i| \leq \lambda$, the penalty term increases in a concave, minimax manner, encouraging sparsity by smoothly shrinking coefficients.
- When $|\beta_i| > \lambda$ is greater than λ , the penalty term remains constant, effectively setting the coefficient to zero.

In this study, we compared the variable selection performance of SCAD for tuning parameters $\gamma = 3, 4, 5, 10, 15, 20, 30, 50,$ and 100 with its performance for the standard value of tuning parameter $\gamma = 3.7$ (Zhang, 2010). Similarly, we compared the variable selection performance of MCP at for tuning parameters $\gamma = 3, 4, 5, 10, 15, 20, 30, 50,$ and 100 with its performance for the standard value of tuning parameter $\gamma = 3$ (Zhang, 2010). The values were combined evaluated at $k = 4, 5$ and 10 folds cross-validation.

3.3 Performance metric of variable selection criteria

The number of variables selected, the cross-validation error (CVE) and the value of shrinkage parameter (λ) were used to assess the variable selection performances of SCAD and MCP criteria. The CVE is the error averaged across the cross-validation folds. For $k = n$ folds, cross-validation performs the fitting process n times, with each fit being performed on $\left(\frac{n-1}{n} \times 100\right)\%$ of the data set selected randomly, while the remaining $\left(\frac{1}{n} \times 100\right)\%$ was used as the test set for validation. This is repeated n times to obtain the estimate of the prediction errors $E_i; i = 1, \dots, n$ at each iteration. The average cross-validation error (CVE), also regarded as deviance is:

$$CVE_{n \text{ folds}} = \frac{1}{n} \sum_{i=1}^n E_i \quad (5)$$

A lower cross-validation error indicates a better fit, implying that the model explained the data well. λ is the value of the shrinkage parameter at the minimum CVE. Analyses are done with *ncvreg* version 3.13.0 in the R statistical software.

4. Results and discussion

Table 1. Comparison of variable selection performances of different tuning parameters γ with SCAD standard tuning parameter (Source: Author's computation, 2024)

γ		k = 4	k = 5	k = 10
3	ps	6	10	11
	CVE	0.96	0.94	0.93
	λ_{min}	0.2104	0.1864	0.1808
3.7 (Standard)	ps	14	0	11
	CVE	0.94	1.01	0.94
	λ_{min}	0.1602	0.3025	0.1754
4	ps	6	10	11
	CVE	0.96	0.94	0.93
	λ_{min}	0.2104	0.1864	0.1808
5	ps	6	10	11
	CVE	0.96	0.94	0.93
	λ_{min}	0.2104	0.1864	0.1808
10	ps	6	10	11
	CVE	0.96	0.94	0.93
	λ_{min}	0.2104	0.1864	0.1808
15	ps	6	10	11
	CVE	0.96	0.94	0.93
	λ_{min}	0.2104	0.1864	0.1808
20	ps	6	10	11
	CVE	0.96	0.94	0.93
	λ_{min}	0.2104	0.1864	0.1808
30	ps	6	10	11
	CVE	0.96	0.94	0.93
	λ_{min}	0.2104	0.1864	0.1808
50	ps	6	10	11
	CVE	0.96	0.94	0.93
	λ_{min}	0.2104	0.1864	0.1808
100	ps	6	10	11
	CVE	0.96	0.94	0.93
	λ_{min}	0.2104	0.1864	0.1808

ps=predictors selected, *CVE* = cross-validation error, *λ_{min}* = lambda at min. *CVE, std*=standard, *γ* =tuning parameter

Table 2. Comparison of variable selection performances of different tuning parameters γ with MCP standard tuning parameter (Source: Author’s computation, 2024)

γ		k = 4	k = 5	k = 10
3 Standard)	ps	5	5	7
	CVE	0.98	0.95	0.93
	λ_{min}	0.2304	0.2104	0.1921
3.7	ps	9	0	9
	CVE	0.96	1.01	0.95
	λ_{min}	0.1808	0.3025	0.1864
4	ps	5	5	9
	CVE	0.97	0.94	0.92
	λ_{min}	0.2304	0.2104	0.1864
5	ps	5	5	10
	CVE	0.97	0.94	0.92
	λ_{min}	0.2168	0.2104	0.1808
10	ps	6	10	11
	CVE	0.97	0.94	0.92
	λ_{min}	0.2104	0.1864	0.1808
15	ps	6	10	11
	CVE	0.96	0.94	0.93
	λ_{min}	0.2104	0.1864	0.1808
20	ps	6	10	11
	CVE	0.96	0.94	0.93
	λ_{min}	0.2104	0.1864	0.1808
30	ps	6	10	11
	CVE	0.96	0.94	0.93
	λ_{min}	0.2104	0.1864	0.1808
50	ps	6	10	11
	CVE	0.96	0.94	0.93
	λ_{min}	0.2104	0.1864	0.1808
100	ps	6	10	11
	CVE	0.96	0.94	0.93
	λ_{min}	0.2104	0.1864	0.1808

ps=predictors selected, *CVE* = cross-validation error, *λ_{min}* = lambda at min. *CVE*, *std*=standard, *γ* =tuning parameter

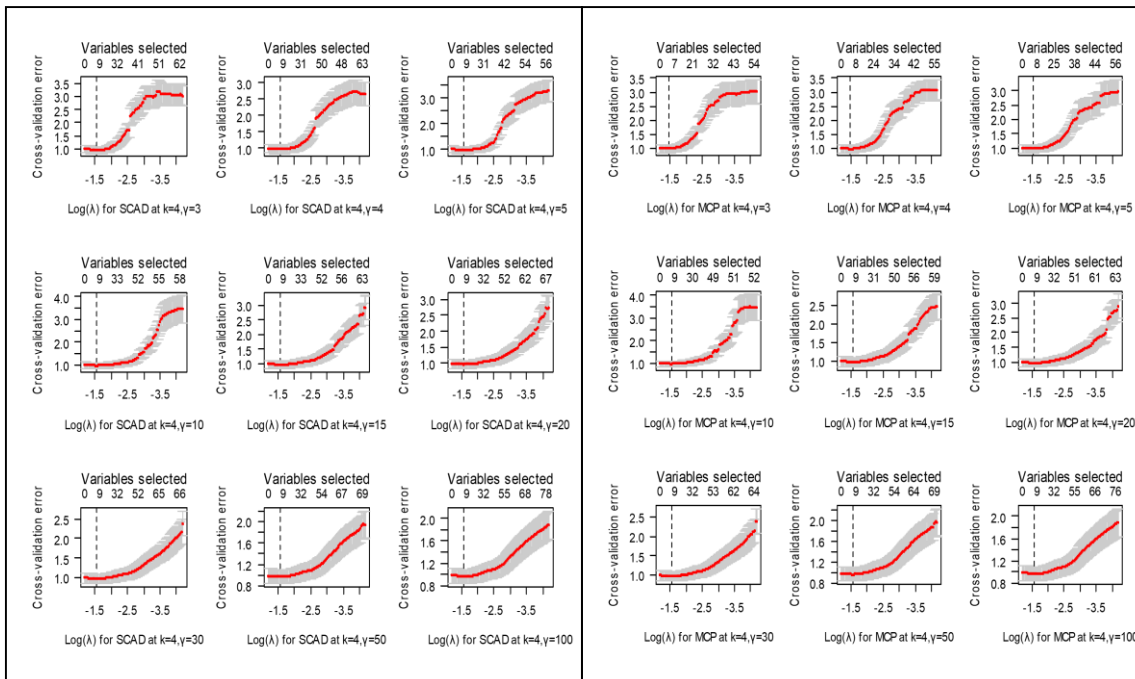


Figure 1: 4-fold cross-validation errors at different values of shrinkage parameters for SCAD (left) and MCP (right) (Source: Author’s computation, 2024)

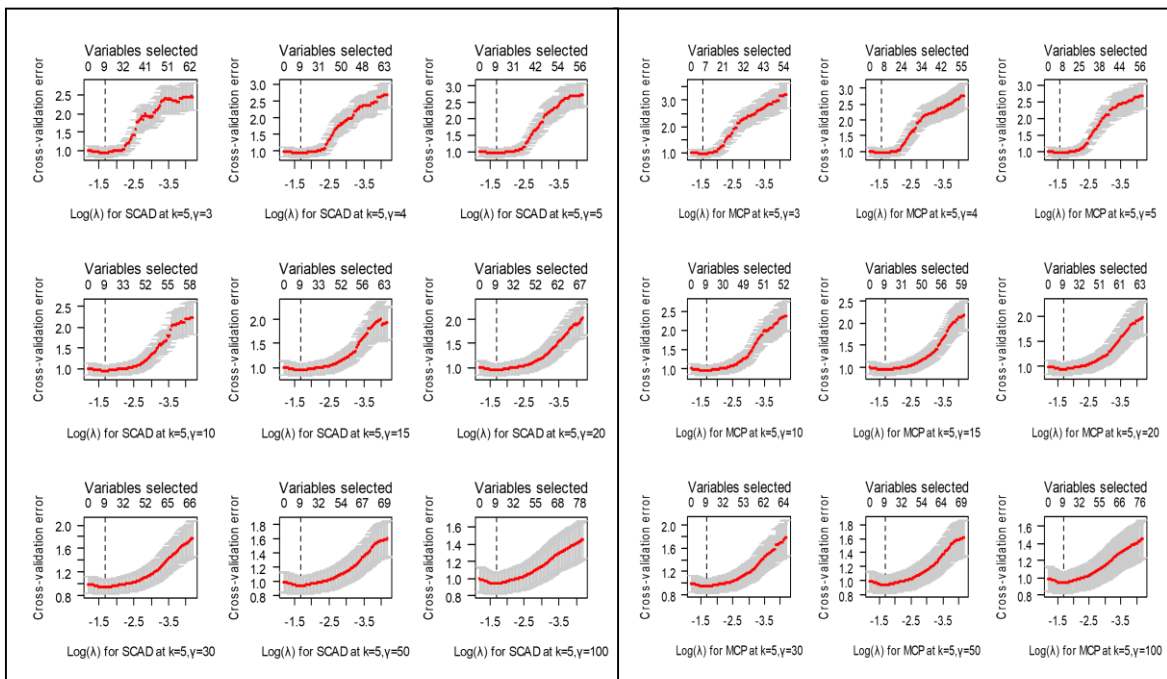


Figure 2: 5-fold cross-validation errors at different values of shrinkage parameters for SCAD (left) and MCP (right) (Source: Author’s computation, 2024)

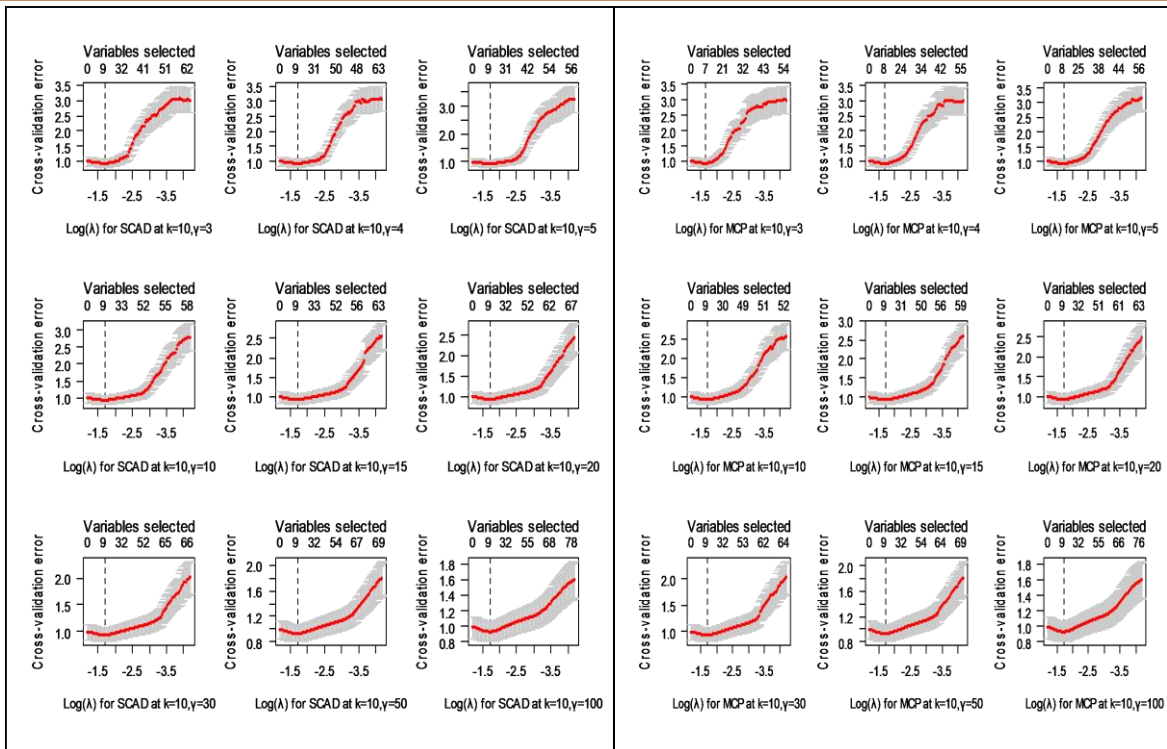


Figure 3: 10-fold cross-validation errors at different values of shrinkage parameters for SCAD (left) and MCP (right) (Source: Author’s computation, 2024)

Table 3: Summary of 95% confidence intervals of the cross-validation errors (Source: Author’s computation, 2024)

λ	SCAD			MCP		
	4-fold	5-fold	10-fold	4-fold	5-fold	10-fold
3	$1.0 \leq e \leq 3.5$	$1.0 \leq e \leq 2.5$	$1.0 \leq e \leq 3.5$	$1.0 \leq e \leq 3.5$	$1.0 \leq e \leq 3.5$ *	$1.0 \leq e \leq 3.5$ *
3.7	$1.0 \leq e \leq 3.5$	$1.0 \leq e \leq 3.0$ *	$1.0 \leq e \leq 5.0$ *	$1.0 \leq e \leq 3.5$	$1.0 \leq e \leq 4.0$ *	$1.0 \leq e \leq 4.0$ *
4	$1.0 \leq e \leq 3.0$	$1.0 \leq e \leq 3.0$ *	$1.0 \leq e \leq 3.5$	$1.0 \leq e \leq 3.5$	$1.0 \leq e \leq 3.0$	$1.0 \leq e \leq 3.5$ *
5	$1.0 \leq e \leq 3.5$	$1.0 \leq e \leq 3.0$ *	$1.0 \leq e \leq 3.5$	$1.0 \leq e \leq 3.0$	$1.0 \leq e \leq 3.0$	$1.0 \leq e \leq 3.5$ *
10	$1.0 \leq e \leq 4.0$ *	$1.0 \leq e \leq 2.5$	$1.0 \leq e \leq 3.0$	$1.0 \leq e \leq 4.0$ *	$1.0 \leq e \leq 2.5$	$1.0 \leq e \leq 2.5$
15	$1.0 \leq e \leq 3.0$	$1.0 \leq e \leq 2.0$ *	$1.0 \leq e \leq 2.5$	$1.0 \leq e \leq 2.5$	$1.0 \leq e \leq 2.5$	$1.0 \leq e \leq 3.0$
20	$1.0 \leq e \leq 3.0$	$1.0 \leq e \leq 2.0$ *	$1.0 \leq e \leq 2.5$	$1.0 \leq e \leq 3.0$	$1.0 \leq e \leq 2.0$	$1.0 \leq e \leq 2.5$
30	$1.0 \leq e \leq 2.5$	$0.8 \leq e \leq 2.0$	$1.0 \leq e \leq 2.0$ **	$1.0 \leq e \leq 2.5$	$0.8 \leq e \leq 2.0$	$1.0 \leq e \leq 2.0$ **
50	$0.8 \leq e \leq 2.5$	$0.8 \leq e \leq 1.8$	$0.8 \leq e \leq 2.0$	$0.8 \leq e \leq 2.0$ **	$0.8 \leq e \leq 1.8$	$0.8 \leq e \leq 2.0$
100	$0.8 \leq e \leq 2.0$ **	$0.8 \leq e \leq 1.6$ **	$0.8 \leq e \leq 1.8$ **	$0.8 \leq e \leq 2.0$ **	$0.8 \leq e \leq 1.6$ **	$0.8 \leq e \leq 1.8$ **

*highest confidence intervals **lowest confidence intervals, Standard tuning parameter for SCAD = 3.7, Standard tuning parameter for MCP = 3

4.1 Discussion

The comparative analysis of variable selection performance between SCAD (Smoothly Clipped Absolute Deviation) and MCP (Minimax Concave Penalty) is detailed in Table 1, showcasing the number of predictors selected, cross-validation error (CVE), and shrinkage penalty at minimum CVE (λ_{\min}) for various tuning parameters γ . This data reveals critical insights into how each method performs under different conditions. Specifically, SCAD and MCP exhibit distinct behaviors when it comes to the number of predictors selected, the error rates, and the shrinkage penalties applied during model training, which are pivotal for selecting the appropriate method in predictive modeling tasks.

For SCAD, when compared to the standard tuning parameter $\gamma = 3.7$, the number of predictors selected was notably lower for 4-fold cross-validation, higher for 5-fold, and equal for 10-fold cross-validation. This indicates that the fold number in cross-validation significantly impacts the model's complexity, which is reflected in the number of predictors SCAD selects. The lower selection of predictors at 4-fold suggests a more conservative model, whereas the higher selection at 5-fold indicates a more complex model (Fan & Li, 2001).

In terms of predictive accuracy, SCAD demonstrated lower CVE across all γ values and k-folds. This consistent reduction in CVE highlights SCAD's robustness in minimizing prediction errors, making it a potentially more reliable choice for models where prediction accuracy is critical (Zou & Li, 2008). Moreover, SCAD's shrinkage penalty λ_{\min} was lower for 4-fold and 10-fold cross-validation but higher for 5-fold, suggesting variability in the extent of penalization applied depending on the fold structure. This variability in λ_{\min} highlights the importance of tuning cross-validation folds to optimize model performance.

Conversely, MCP showed different characteristics. The number of predictors selected by MCP was equal to the standard $\gamma = 3$ at $\gamma = 4$ for 4-fold and $\gamma = 5$ for 5-fold, but higher for other γ values. This indicates that MCP generally selects more predictors as the tuning parameter γ increases, potentially leading to more complex models (Zhang, 2010). However, the CVE for MCP was higher at the standard γ for 4 and 5 folds,

suggesting that MCP may not be as effective as SCAD in minimizing prediction error at these specific settings. At 10-fold, the CVE remained stable across most γ values, indicating a level of consistency in predictive performance with increased folds.

A direct comparison between SCAD and MCP reveals that SCAD generally selected more predictors at lower γ values (3, 4, and 5) across different folds, indicating SCAD's tendency to produce more complex models under these conditions. For γ values greater than 5, both methods converged, selecting the same number of predictors. This suggests that at higher γ values, the distinction between SCAD and MCP diminishes in terms of model complexity (Breheeny & Huang, 2011). SCAD's lower CVE at $\gamma = 3, 4,$ and 5 for $k = 4$ and at $\gamma = 3$ for $k = 5$ indicates superior predictive performance under these settings. However, at $k = 10$, MCP outperformed SCAD in CVE for higher γ values, highlighting specific scenarios where MCP might be preferable.

The analysis of the shrinkage penalty λ_{\min} further differentiates the methods. SCAD had a lower λ_{\min} than MCP at $\gamma = 3, 4,$ and 5 for both 4 and 5 folds, and at $\gamma = 3$ and 4 for 10 folds, indicating a stronger penalization effect, which can lead to more regularized models (Fan & Lv, 2010). MCP's higher λ_{\min} at higher folds and γ values suggests a less aggressive shrinkage, correlating with its tendency to select more predictors and maintain model complexity. These findings underscore the necessity of carefully tuning both the γ parameter and the cross-validation fold to leverage the strengths of SCAD or MCP in different modeling contexts.

In summary, while SCAD generally provides lower CVE and tends to select more predictors at lower γ values, MCP shows stability in higher γ values and folds. The choice between SCAD and MCP should be guided by specific modeling needs, particularly the desired balance between model complexity and predictive accuracy. The variability in performance based on tuning parameters and cross-validation folds suggests that no single method is universally superior; rather, the optimal choice depends on the specific characteristics and requirements of the modeling task at hand (Hastie, Tibshirani, & Friedman, 2009).

Figures 1-3 illustrate the cross-validation errors (CVE) for different λ values, demonstrating that higher λ values lead to fewer predictors being selected and lower CVE. This trend aligns with the findings of Xiong (2014), which indicate that increasing the regularization parameter can enhance model performance by reducing overfitting. However, this observation contradicts the results presented by Lima et al. (2020), likely due to differences in the response variables used in their study, highlighting the importance of context in model performance outcomes.

In this study, 10-fold cross-validation provides the best results for minimizing error while MCP produced the most parsimonious model, a finding that corroborates the conclusions of Idrees and Kamal (2021), who emphasized MCP's efficiency in creating simpler models with fewer predictors without compromising accuracy. These results underscore the necessity of choosing appropriate regularization techniques and cross-validation methods to balance model simplicity and predictive performance effectively.

5. Conclusion

The comparative analysis of SCAD and MCP in high-dimensional datasets with count responses revealed distinct performance patterns. SCAD generally selected more predictors than MCP at lower γ values, and its cross-validation error (CVE) was consistently lower for $\gamma = 3, 4, \text{ and } 5$ at 4-fold and 5-fold cross-validation. However, at 10-fold cross-validation, MCP demonstrated superior performance in terms of lower CVE for $\gamma = 4, 5, \text{ and } 10$, while both methods showed similar performance for higher γ values.

The λ_{\min} values indicated that SCAD required less shrinkage penalty than MCP at lower γ values for 4-fold and 5-fold cross-validation, but the penalties were comparable at higher γ values. Figures 1-3 highlighted that higher λ values led to fewer predictors selected and lower CVE, emphasizing the importance of tuning parameters in model performance.

Overall, both SCAD and MCP demonstrated robust performance in variable selection for high-dimensional count response data. SCAD's advantage lies in its lower CVE at specific γ values, while MCP excels in providing a more parsimonious model, making both methods valuable depending on the specific requirements of the analysis.



References

- Bonney, D. K. (2020). General Penalized Logistic Regression for Gene Selection in High-Dimensional Microarray Data Classification (Doctoral dissertation, The University of Texas at El Paso). Available at: https://scholarworks.utep.edu/open_etd/3084
- Breheny, P., & Huang, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *Annals of Applied Statistics*, 5(1), 232-253. <https://doi.org/10.1214/10-AOAS388>
- Brimacombe, M. (2014). High-dimensional data and linear models: a review. *Open Access Medical Statistics*, 4, 17-27. Available at: <https://www.dovepress.com/high-dimensional-data-and-linear-models-a-review-peer-reviewed-article-OAMS>
- Cameron, A. C., & Trivedi, P. K. (2013). Regression analysis of count data (Vol. 53). Cambridge University Press. Available at: <https://www.cambridge.org/core/books/regression-analysis-of-count-data/8AAE78A63DE68A0F4165F1DA6F7F3EDB>
- Emmert-Streib, F., & Dehmer, M. (2019). High-dimensional LASSO-based computational regression models: regularization, shrinkage, and selection. *Machine Learning and Knowledge Extraction*, 1(1), 359-383. Available at: <https://www.mdpi.com/2504-4990/1/1/21>
- 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. Available at: <https://www.tandfonline.com/doi/abs/10.1198/016214501753382273>
- Fan, J., & Lv, J. (2010). A Selective Overview of Variable Selection in High Dimensional Feature Space. *Statistica Sinica*, 20(1), 101-148. <http://www3.stat.sinica.edu.tw/statistica/oldpdf/A20n11.pdf>
- Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., Tay, K., & Simon, N. (2021). Lasso and Elastic-Net Regularized Generalized Linear Models, Version 4.1-2. Available at: <https://glmnet.stanford.edu>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- Idrees, N., & Kamal, S. (2021). Modelling of wheat production in Punjab through the regularized regression approach while addressing collinearity. *Pakistan Journal of Agricultural Sciences*, 58(1). Available at: <https://pakjas.com.pk/papers/3358.pdf>
- James, G., Witten, D., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer. Available at: <https://link.springer.com/book/10.1007/978-1-4614-7138-7>
- Jia, W., Sun, M., Lian, J., & Hou, S. (2022). Feature dimensionality reduction: a review. *Complex & Intelligent Systems*, 8(3), 2663-2693. Available at: <https://link.springer.com/article/10.1007/s40747-022-00661-1>
- Kwon, S., Han, S., & Lee, S. (2013). A small review and further studies on the LASSO. *Journal of the Korean Data and Information Science Society*, 24(5), 1077-1088. Available at: <http://www.koreascience.or.kr/article/JAKO201302677714394.page>
- Li, R., & Liang, H. (2008). Variable selection in semiparametric regression modelling. *Annals of Statistics*, 36(1), 261. Available at: <https://projecteuclid.org/euclid.aos/1201011754>
- Li, N., & Yang, H. (2021). Nonnegative estimation and variable selection under minimax concave penalty for sparse high-dimensional linear regression models. *Statistical Papers*, 62, 661-680. Available at: <https://link.springer.com/article/10.1007/s00362-019-01119-7>
- Li, Z. (2015). High Dimensional Model Selection and Validation: A Comparison Study. *Culminating Projects in Applied Statistics*. Available at: https://repository.stcloudstate.edu/stat_etds/1
- Lima, E., Davies, P., Kaler, J., Lovatt, F., & Green, M. (2020). Variable selection for inferential models with relatively high-dimensional data: Between method heterogeneity and covariate stability as adjuncts to robust selection. *Scientific Reports*, 10(1), 8002. Available at: <https://www.nature.com/articles/s41598-020-64723-1>

- Mallick, H., & Yi, N. (2013). Bayesian methods for high dimensional linear models. *Journal of Biometrics & Biostatistics*, 5(1). Available at: <https://www.omicsonline.org/open-access/bayesian-methods-for-high-dimensional-linear-models-2155-6180.1000185.php?aid=23238>
- Shao, J., & Deng, X. (2012). Estimation in High-dimensional Linear Models with Deterministic Design Matrices. *Annals of Statistics*, 40, 812–831. Available at: <https://projecteuclid.org/euclid.aos/1336219405>
- Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288. Available at: <https://www.jstor.org/stable/2346178>
- Xiong, H., Tan, P. N., & Kumar, V. (2006). Hyperclique pattern discovery. *Data Mining and Knowledge Discovery*, 13, 219-242. Available at: <https://link.springer.com/article/10.1007/s10618-005-0026-5>
- Yang, X. S. et al. (2015). Information Analysis of High-dimensional Data and Applications. *Mathematical Problems in Engineering, Hindawi*, vol. 2015, pages 1-2, November. Available at: <https://www.hindawi.com/journals/mpe/2015/510326/>
- Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. *Annals of Statistics*, 38(2), 894-942. <https://doi.org/10.1214/09-AOS729>
- Zhang, K., Yin, F., & Xiong, S. (2014). Comparisons of penalized least squares methods by simulations. *arXiv preprint: arXiv:1405.1796*. Available at: <https://arxiv.org/abs/1405.1796>
- Zhou, H., Siegel, P. Z., Barile, J., Njai, R. S., Thompson, W. W., Kent, C., & Liao, Y. (2014). Peer Reviewed: Models for Count Data With an Application to Healthy Days Measures: Are You Driving in Screws With a Hammer? *Preventing Chronic Disease*, 11. Available at: https://www.cdc.gov/pcd/issues/2014/13_0389.htm
- Zou, H., & Li, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Annals of Statistics*, 36(4), 1509-1533. <https://doi.org/10.1214/07-AOS520>