

Variable Selection under Multicollinearity and Heavy-Tailed Errors: A Systematic Assessment of Classical and Penalized Methods

Ofuru, B.S., Ijomah M.A. & Nwakuya M.T.

Department of Mathematics and Statistics, University of Port Harcourt

DOI: <https://doi.org/10.51244/IJRSI.2026.1308000230>

Received: 27 August 2026; Accepted: 01 September 2026; Published: 17 September 2026

ABSTRACT

Variable selection in high-dimensional regression becomes particularly challenging when multicollinearity and heavy-tailed errors occur simultaneously. This study systematically evaluates the performance boundaries of classical, shrinkage, sparse, and robust regression methods under these conditions. A Monte Carlo simulation experiment was conducted using 100 candidate predictors, of which 10 were truly active, while predictor correlation was varied across $\rho = 0.1, 0.3, 0.6$, and 0.9 . Three error environments normal, contaminated normal, and Student-t heavy-tailed errors were considered. Ordinary Least Squares (OLS), Ridge, LASSO, Elastic Net, and Huber regression were evaluated using mean squared error (MSE), mean absolute error (MAE), root mean squared error (RMSE), number of selected variables, true positive rate (TPR), false positive rate (FPR), coefficient variability, and selection instability. The findings reveal a clear trade-off between predictive accuracy, sparsity, robustness, and selection stability. OLS and Ridge achieved comparatively favorable predictive performance in several scenarios but retained all 100 predictors, resulting in substantial over-selection. LASSO achieved greater sparsity and relatively lower false-positive rates but became increasingly unstable under severe multicollinearity, with selection instability rising to 0.2976 at $\rho = 0.9$. Elastic Net demonstrated the most balanced performance, maintaining high signal recovery while substantially reducing selection instability as predictor correlation increased. Under contaminated errors, Elastic Net also outperformed LASSO in signal recovery and prediction. Huber regression reduced sensitivity to extreme observations but did not provide sparse variable selection. Overall, the results demonstrate that no conventional method simultaneously addresses heavy-tailed errors, multicollinearity, and sparse selection adequately. Elastic Net provides the strongest benchmark among the evaluated procedures, while the findings motivate the development of a robust, sparsity-inducing, and correlation-aware variable-selection framework for high-dimensional regression.

Keywords: variable selection; multicollinearity; heavy-tailed errors; high-dimensional regression; LASSO; Elastic Net; robust regression; Huber regression; selection stability; penalized regression.

INTRODUCTION

Variable selection is an important component of modern statistical modelling because researchers increasingly encounter datasets containing a large number of potential explanatory variables. The objective of variable selection is not merely to reduce the number of predictors included in a regression model, but to identify a parsimonious subset of variables that adequately explains variation in the response while maintaining reliable estimation, prediction, and interpretation. This problem becomes particularly important in high-dimensional settings, where the number of candidate predictors is large relative to the available observations. Penalized regression procedures such as the Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net have consequently become important tools for simultaneous estimation and variable selection.

The classical linear regression model can be represented as

$$Y = X\beta + \varepsilon, \quad (1.1)$$

where y is an $n \times 1$ response vector, X is an $n \times p$ matrix of predictors, β is a $p \times 1$ vector of unknown coefficients, and ε is a vector of random errors. In a variable-selection problem, the primary objective is to recover the subset

of predictors for which the corresponding elements of β are non-zero. This becomes increasingly difficult as p grows and as the predictors become strongly correlated.

Multicollinearity is one of the most persistent challenges in regression analysis. It arises when explanatory variables are strongly linearly related, causing the design matrix to become ill-conditioned and making individual coefficient estimates sensitive to relatively small changes in the data. Although multicollinearity does not necessarily result in poor overall prediction, it can substantially undermine coefficient stability and make the identity of selected variables uncertain. This distinction is particularly important when variable selection is undertaken not only for prediction but also for substantive interpretation and identification of important determinants.

The Least Absolute Shrinkage and Selection Operator introduced by Tibshirani (1996) incorporates an L_1 penalty into the least-squares criterion. For a tuning parameter $\lambda > 0$, the LASSO estimator is given by

$$\beta_L = \arg \min_{\beta} \left\{ \frac{1}{2n} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\} \quad (1.2)$$

where

$$\|\beta\|_1 = \sum_{j=1}^p |\beta_j|$$

Thus, equivalently,

$$\beta_L = \arg \min_{\beta} \left\{ \frac{1}{2n} \sum_{i=1}^n (y_i - x_i^T \beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (1.3)$$

The L_1 penalty allows some coefficient estimates to be exactly zero and therefore enables estimation and variable selection to occur simultaneously. LASSO can consequently generate highly sparse models and is particularly attractive when only a relatively small proportion of candidate predictors are expected to be relevant. Despite these advantages, LASSO can behave unpredictably when explanatory variables are strongly correlated. When several predictors contain similar information, LASSO may select one member of a correlated group while excluding other members that are equally important. Consequently, although LASSO may generate a sparse model, the selected subset may be unstable across samples. This limitation motivated the development of Elastic Net by Zou and Hastie (2005).

Elastic Net combines L_1 and L_2 penalties and can be written as

$$\beta_{EN} = \arg \min_{\beta} \left\{ \frac{1}{2n} \|Y - X\beta\|_2^2 + \lambda \left[\alpha \|\beta\|_1 + \frac{1-\alpha}{2} \|\beta\|_2^2 \right] \right\} \quad (1.4)$$

where $0 \leq \alpha \leq 1$. The L_1 component promotes sparsity, whereas the L_2 component stabilizes coefficient estimation and encourages correlated predictors to enter the model together. This grouping property makes Elastic Net an important benchmark in the presence of multicollinearity. Nevertheless, conventional Elastic Net continues to rely on squared-error loss and can therefore remain sensitive to extreme observations. A second major challenge in regression modelling is the presence of heavy-tailed errors and contamination. The assumption that regression errors are normally distributed is often unrealistic in economic, financial, environmental, biomedical, and social-science applications. Heavy-tailed distributions assign greater probability to extreme observations than the normal distribution. Consequently, conventional least-squares estimation may be disproportionately influenced by a relatively small number of observations. This sensitivity arises because the squared-error criterion gives increasingly large weight to large residuals. Robust regression methods address this problem by replacing squared-error loss with loss functions that reduce the influence of extreme observations. Huber regression, for example, uses a loss function that is quadratic for small residuals and approximately linear for large residuals. It therefore provides a compromise between the efficiency of least squares under regular conditions and robustness under contamination. Least Absolute Deviation (LAD) regression provides another robust alternative by minimizing the sum of absolute residuals. When combined with an L_1 penalty, LAD-LASSO simultaneously

incorporates robustness and sparse variable selection. Huber-LASSO similarly combines the Huber loss function with an L_1 penalty.

The methodological challenge becomes more serious when multicollinearity and heavy-tailed errors occur simultaneously. These problems are often investigated separately, although real datasets can exhibit both characteristics. Strong predictor correlation makes coefficient estimates sensitive to perturbations, while heavy-tailed observations introduce precisely such perturbations. Their combination can therefore create a more difficult variable-selection environment than either problem considered independently. For example, when two predictors are highly correlated and both are related to the response, a sparse procedure may already experience difficulty determining which predictor should receive a non-zero coefficient. If several observations simultaneously produce unusually large residuals, the relative contributions of the correlated variables may change substantially. Consequently, the model selected from one sample may differ considerably from the model selected from another sample even when the underlying data-generating mechanism remains unchanged.

The present study therefore evaluates classical, shrinkage, sparse, and robust sparse regression procedures under combinations of multicollinearity and non-Gaussian errors. The revised simulation framework explicitly distinguishes prediction accuracy, sparsity, signal recovery, false-positive control, and selection stability. The analysis considers Ordinary Least Squares (OLS), Ridge regression, LASSO, Elastic Net, Huber regression, Huber-LASSO, and LAD-LASSO.

The study is designed to determine whether conventional penalized procedures remain reliable when predictor correlation becomes severe and errors depart from normality, and whether robust sparse procedures can provide a better balance between robustness and variable-selection accuracy. The study also examines whether conclusions obtained under an equicorrelation structure remain valid under alternative correlation structures, particularly block correlation and AR(1) correlation.

The study evaluates competing regression methods under normal, contaminated-normal, and heavy-tailed errors, while examining how increasing predictor correlation affects coefficient stability and variable-selection accuracy. It compares sparse and non-sparse estimators using selected variables, true-positive rates, and false-positive rates, and assesses whether robust loss functions and robust penalized approaches improve performance under challenging error distributions. The study also examines stability across alternative covariance structures. Its main contribution is to identify conditions under which conventional and penalized methods succeed or fail, emphasizing that predictive accuracy alone may not ensure effective variable selection, sparsity, or stability among correlated predictors.

LITERATURE REVIEW

The literature on variable selection has developed considerably from traditional subset-selection procedures to regularization methods designed for high-dimensional data. Classical procedures include best-subset selection, forward selection, backward elimination, and stepwise regression. Although these approaches can generate relatively interpretable models, they become computationally demanding as the number of candidate predictors increases. Their performance can also be unstable in the presence of strong correlations among explanatory variables. Tibshirani (1996) introduced LASSO as an important alternative by incorporating an L_1 penalty into the least-squares criterion. Its principal advantage is that it can shrink some regression coefficients exactly to zero, thereby producing sparse models. LASSO is consequently particularly useful when the underlying model is believed to contain only a relatively small number of relevant predictors. However, its performance can deteriorate when explanatory variables are highly correlated because it may select one variable from a group of correlated predictors while excluding others containing similar information. Zou and Hastie (2005) developed Elastic Net partly to address this limitation. Elastic Net combines L_1 and L_2 penalties. The L_1 component provides sparsity, while the L_2 component stabilizes coefficient estimation and encourages correlated variables to enter the model together. The grouping property of Elastic Net makes it particularly attractive in applications characterized by multicollinearity.

Ridge regression provides another strategy for handling correlated predictors. Hoerl and Kennard (1970) introduced Ridge regression as a shrinkage procedure that reduces coefficient variance by introducing an L_2

penalty. Ridge is often effective for prediction when predictors are strongly correlated, but it generally retains all candidate variables. Consequently, its principal advantage lies in coefficient stabilization and prediction rather than sparse variable selection. Adaptive LASSO, introduced by Zou (2006), assigns different penalty weights to different predictors. The procedure can reduce shrinkage on important variables while imposing stronger penalties on variables that appear less important. Although Adaptive LASSO possesses desirable theoretical properties under suitable assumptions, its practical performance can still be affected by strong predictor dependence and nonstandard error distributions. A separate body of research has focused on robust regression. Huber (1964) proposed a robust loss function that is quadratic for small residuals and approximately linear for large residuals. This reduces the influence of extreme observations while maintaining relatively high efficiency under well-behaved errors. Huber regression has subsequently become an important benchmark for robust estimation.

LAD regression provides another robust approach by minimizing absolute residuals rather than squared residuals. Because absolute deviations grow linearly rather than quadratically, extreme observations exert less influence on the resulting estimator. Combining LAD with an L_1 penalty produces LAD-LASSO, which simultaneously addresses robustness and sparsity.

The development of robust sparse procedures reflects recognition that robustness alone and sparsity alone address different dimensions of the regression problem. A robust estimator may reduce the influence of extreme observations but retain many irrelevant predictors. Conversely, a sparse estimator may eliminate irrelevant predictors while remaining sensitive to extreme observations. The combination of robust loss and sparse penalization is therefore particularly relevant when both heavy-tailed errors and high-dimensional predictor spaces are present.

The literature also indicates that correlation among predictors has implications beyond coefficient variance. Under strong dependence, the identity of the selected variables may become unstable even when prediction remains relatively stable. This has important implications for scientific and economic interpretation because researchers may attach substantive meaning to variables selected by a statistical procedure. A model whose selected variables change substantially across samples may therefore provide limited evidence about which predictors are genuinely important.

Consequently, contemporary evaluation of variable-selection methods increasingly considers multiple performance dimensions, including predictive error, coefficient accuracy, sparsity, true-positive recovery, false-positive control, and selection stability. Prediction error alone is insufficient because it does not reveal whether irrelevant predictors have been retained.

The present study builds on this multidimensional perspective by comparing classical, shrinkage, sparse, and robust sparse procedures under controlled combinations of correlation and non-Gaussian error structures. The inclusion of Huber-LASSO and LAD-LASSO is particularly important because it permits a direct assessment of whether robustness and sparsity can be achieved simultaneously. The study further extends the simulation design by considering equicorrelation, block correlation, and AR(1) covariance structures. This is important because conclusions based exclusively on an equicorrelation structure may be overly dependent on a particular method of generating multicollinearity.

METHODOLOGY

The study employs a Monte Carlo simulation framework based on the linear regression model $Y = X\beta + \varepsilon$, where Y is the response vector, X is the predictor matrix, β is the coefficient vector, and ε represents random errors. The simulation considers $p = 100$ candidate predictors, with only ten having non-zero coefficients and the remaining 90 being irrelevant. This sparse structure provides a known benchmark for assessing the ability of competing methods to identify relevant variables while excluding irrelevant ones. For reproducibility, the final simulation should clearly report the sample size, non-zero coefficients, number of replications, random seed, tuning-parameter grid, cross-validation procedure, contamination settings, and Student-tt degrees of freedom based on the actual simulation implementation.

Predictors are generated from a multivariate normal distribution, $X_i \sim N_p(0, \Sigma)$, under three covariance structures: equicorrelation, block correlation, and AR(1) correlation. Equicorrelation examines four levels, $\rho = 0.1, 0.3, 0.6$ and 0.9 , representing weak to severe multicollinearity. Block correlation allows predictors within groups to be correlated while predictors in different groups have zero or weaker correlations. The AR(1) structure specifies $\Sigma_{jk} = \rho^{|j-k|}$, such that nearby predictors are more strongly correlated than distant predictors. These structures permit assessment of whether methodological performance remains consistent under different forms of predictor dependence. Three error environments are considered: normal, contaminated-normal, and heavy-tailed Student-t errors. The normal distribution provides the benchmark condition. Contaminated-normal errors are generated as a mixture of a dominant normal distribution and a higher-variance contaminating distribution, thereby introducing extreme observations. Heavy-tailed errors are generated from a Student-t distribution, with the degrees of freedom determining tail heaviness. The exact contamination proportion, contamination magnitude, and Student-t degrees of freedom should be explicitly reported from the simulation code rather than inferred from summary results. Where necessary, errors are rescaled to ensure comparable variances across distributions. Seven regression procedures are compared: OLS, Ridge, LASSO, Elastic Net, Huber regression, Huber-LASSO, and LAD-LASSO. OLS serves as the classical benchmark, whereas Ridge introduces an L_2 penalty to improve coefficient stability under multicollinearity but does not perform variable selection. LASSO applies an L_1 penalty and therefore encourages sparse solutions. Elastic Net combines L_1 and L_2 penalties, allowing simultaneous variable selection and stabilization of correlated predictors. Huber regression replaces the squared-error loss with the robust Huber loss, reducing the influence of observations with large residuals. Huber-LASSO combines this robust loss with an L_1 penalty, while LAD-LASSO combines absolute-deviation loss with L_1 regularization to provide an alternative robust sparse estimator. Tuning parameters are selected through an explicitly stated data-driven procedure, preferably K-fold cross-validation. The procedure should specify the number of folds, candidate values of λ , the Elastic Net mixing parameter α , and the Huber tuning constant δ . A common validation criterion should be used across methods, with the manuscript clearly indicating whether minimum cross-validation error or the one-standard-error rule determines the final model. Consistent tuning procedures are essential for ensuring a fair comparison among competing estimators. Predictive performance is evaluated using mean squared error (MSE), mean absolute error (MAE), and root mean squared error (RMSE). Variable-selection performance is assessed using the number of selected predictors, true positives (TP), false positives (FP), true-positive rate (TPR), and false-positive rate (FPR). With ten truly active and 90 inactive predictors, $TPR = TP/10$ and $FPR = FP/90$. This definition ensures consistency across all methods. Importantly, a model retaining all 100 predictors has 90 false positives and therefore an FPR of 1.000, not 0.500. Consequently, previously reported FPR values should be recalculated directly from replication-level results. Selection stability is assessed by comparing selected-variable sets across Monte Carlo replications using an instability measure based on one minus the Jaccard similarity. Values near zero indicate stable selection, whereas larger values indicate greater instability. Coefficient variability is also examined to determine the sensitivity of estimated effects to changes in simulated samples.

Because simulation estimates contain Monte Carlo uncertainty, results are summarized using replication means, standard deviations, Monte Carlo standard errors (MCSEs), and, where appropriate, 95% confidence intervals. The complete experiment crosses the three covariance structures, four correlation levels, and three error distributions for all competing methods, providing a comprehensive assessment of the interaction between multicollinearity, non-normal errors, robustness, and sparsity. SCAD and MCP are excluded from empirical rankings unless complete simulations are available for them; otherwise, their inclusion would create a mismatch between the stated methodology and reported numerical results.

RESULTS OF ANALYSIS

The simulation results demonstrate that predictive accuracy and variable-selection accuracy represent different dimensions of regression performance. The existing results show that OLS and Ridge can produce relatively competitive prediction errors while retaining all candidate predictors, whereas LASSO and Elastic Net substantially reduce model dimensionality. Consequently, a low MSE should not automatically be interpreted as evidence of successful variable selection.

Under the Student-t error environment, the existing results reported MSE values of 0.8352 for OLS and 0.8523 for Ridge. LASSO recorded an MSE of 1.7205 after selecting 32 variables, whereas Elastic Net recorded an MSE of 1.4446 after selecting 52 variables. Huber regression recorded an MSE of 1.5667 while retaining all 100 candidate variables. These results illustrate the trade-off between prediction and sparsity.

Table 4.1: Performance of Regression Methods under Student-t Heavy-Tailed Errors

Method	MSE	Selected Variables	True Positives	MAE	RMSE	TPR	FPR
OLS	0.8352	100	10	2.5436	3.2368	1.000	1.000*
Ridge	0.8523	100	10	1.9052	2.4852	1.000	1.000*
LASSO	1.7205	32	10	1.3656	1.9104	0.966	0.2444*
Elastic Net	1.4446	52	10	1.4499	1.9942	0.980	0.4667*
Huber	1.5667	100	10	1.6843	2.1978	1.000	1.000*

The existing results also show that LASSO and Elastic Net recovered almost all active variables under the Student-t scenario. However, their relatively large numbers of selected variables indicate that high signal recovery does not necessarily imply strong false-positive control. OLS and Ridge, by retaining the entire candidate set, provide an especially clear illustration of why TPR should not be interpreted independently of FPR.

The contaminated-normal experiment provides additional evidence of the effects of non-Gaussian observations. OLS and Ridge retained all 100 predictors, whereas LASSO selected 50 variables and identified nine active variables and Elastic Net selected 74 variables while identifying all ten active variables. The existing results therefore indicate a trade-off between model sparsity and signal recovery under contamination.

Table 4.2: Performance under Contaminated Normal Errors

Method	MSE	Selected Variables	True Positives	MAE	RMSE	TPR	FPR
OLS	4.1790	100	10	1.6592	2.3227	1.0000	1.0000*
Ridge	4.4540	100	10	1.6405	2.3256	1.0000	1.0000*
LASSO	7.9200	50	9	2.3876	3.8364	0.8135	0.4556*
Elastic Net	7.2680	74	10	2.3653	3.8154	0.9050	0.7111*
Huber	10.1250	100	10	4.2867	3.1820	1.0000	1.0000*

Table 4.2 shows substantial differences in variable-selection and predictive performance across the competing methods. OLS and Ridge identify all ten true predictors but retain all 100 variables, producing an FPR of 1.000 and indicating no sparsity. LASSO selects only 50 variables and identifies nine true predictors, achieving a lower FPR of 0.456 and TPR of 0.814. Elastic Net performs better in identifying relevant variables, achieving a TPR of 0.905, although it retains 74 predictors. Huber also retains all variables and records the highest MSE and MAE. Overall, Elastic Net provides the best balance between selection accuracy and sparsity.

Table 4.3. Variable-Selection Performance across Error Distributions

Error Distribution	Method	MAE	RMSE	TPR	FPR
Normal	OLS	0.5622	0.7020	1.0000	0.5000
Normal	Ridge	0.5644	0.7049	1.0000	0.5000
Normal	LASSO	0.7271	0.9119	1.0000	0.1795
Normal	Elastic Net	0.6949	0.8703	1.0000	0.3292
Contaminated Normal	OLS	1.6592	2.3227	1.0000	0.5000
Contaminated Normal	Ridge	1.6405	2.3256	1.0000	0.5000
Contaminated Normal	LASSO	1.5927	2.5510	0.9865	0.3768
Contaminated Normal	Elastic Net	1.5863	2.4800	0.9925	0.4420
Student-t	OLS	0.8905	1.1796	1.0000	0.5000
Student-t	Ridge	0.8873	1.1822	1.0000	0.5000
Student-t	LASSO	1.0003	1.4231	0.9995	0.2678
Student-t	Elastic Net	0.9634	1.3460	0.9995	0.3850

The cross-distribution results provide further evidence of the importance of the error structure. Under normal errors, all methods achieved perfect TPR, but LASSO and Elastic Net produced substantially lower FPR values than OLS and Ridge. LASSO had the lowest FPR (0.1795), followed by Elastic Net (0.3292). This demonstrates that penalization substantially improves model parsimony even when the error distribution is well behaved. The contaminated-normal setting increased MAE and RMSE for the competing procedures and reduced the TPR of LASSO and Elastic Net relative to the normal benchmark. LASSO recorded a TPR of 0.9865, whereas Elastic Net recorded 0.9925. The corresponding FPR values were 0.3768 and 0.4420. Thus, contamination increased false-positive selection and introduced some loss of signal recovery.

Under Student-t errors, the sparse procedures continued to recover almost all of the active predictors. LASSO and Elastic Net both recorded TPR values of approximately 0.9995 in the reported comparison, while their FPR values were 0.2678 and 0.3850, respectively. The results suggest that heavy-tailed errors do not necessarily destroy signal recovery, but they can increase prediction error and influence the balance between sparsity and false-positive control. An important point emerging from Table 4.3 is that OLS and Ridge should not be interpreted as superior variable-selection methods simply because they have TPR = 1. Their FPR of 0.5000 arises because they retain the full candidate set. The penalized methods provide a more meaningful assessment because they attempt to recover the active variables while simultaneously excluding irrelevant predictors.

Table 4.4. Effect of Increasing Multicollinearity on LASSO and Elastic Net

P	LASSO Coefficient Variability	LASSO Selection Instability	Elastic Net Coefficient Variability	Elastic Net Selection Instability
0.1	0.1176	0.1169	0.1391	0.2866
0.3	0.1217	0.0838	0.1462	0.2706

0.6	0.1287	0.1156	0.1594	0.2060
0.9	0.1656	0.2976	0.2014	0.0757

The most important evidence concerning multicollinearity is provided by the stability analysis. For LASSO, coefficient variability increased progressively from 0.1176 at $\rho = 0.1$ to 0.1656 at $\rho = 0.9$. This indicates that estimated effects become increasingly sensitive to changes in the data as predictors become more highly correlated. The pattern for selection instability is even more revealing. LASSO selection instability declined from 0.1169 at $\rho=0.1$ to 0.0838 at $\rho=0.3$, but then increased to 0.1156 at $\rho = 0.6$ and rose sharply to 0.2976 at $\rho = 0.9$. Thus, extreme multicollinearity substantially reduces the reproducibility of the variables selected by LASSO. This behavior is consistent with the tendency of LASSO to select one member of a highly correlated group while excluding others. When several predictors contain similar information, small changes in the data may alter which variable receives the non-zero coefficient. Consequently, the selected model can change considerably across samples even when predictive performance remains relatively stable.

Elastic Net displayed a markedly different pattern. Although its coefficient variability increased from 0.1391 at $\rho = 0.1$ to 0.2014 at $\rho = 0.9$, its selection instability declined from 0.2866 to only 0.0757. This indicates that Elastic Net becomes substantially more consistent in the set of selected variables as correlation becomes stronger. The contrast between LASSO and Elastic Net is central to the theme of this study. LASSO emphasizes sparsity but may fragment groups of correlated predictors, whereas Elastic Net permits correlated predictors to enter the model together. The results therefore provide empirical evidence supporting the grouping effect of the L_2 component of Elastic Net.

Table 4.5: Robust Variable Selection under Extreme Observations

Method	MSE	Selected Variables
LASSO (Squared Loss)	1.7416	32
Huber–LASSO (IRLS)	2.1559	18
LAD–LASSO (IRLS)	84.3720	70

Table 4.5 indicates that Huber–LASSO produced a substantially sparser model than ordinary LASSO, selecting 18 rather than 32 variables. However, its MSE was higher. LAD–LASSO performed poorly in this particular experiment, with both a large number of selected variables and substantially higher MSE.

Table 4.6: Comparison of Penalty Structures under Multicollinearity

Method	MSE	Selected Variables	True Positives
LASSO	2.8347	34	—
Elastic Net	2.4411	53	—
Correlation-Weighted LASSO (CW-LASSO)	4.3344	10	9

The Table indicates that Elastic Net achieved the lowest MSE (2.4411), outperforming LASSO (2.8347) and CW-LASSO (4.3344), although it selected more variables. CW-LASSO produced the sparsest model, selecting only 10 variables and identifying 9 true predictors. Thus, Elastic Net offers better prediction, while CW-LASSO provides greater sparsity and selection precision.

Table 4.7a: Monte Carlo Average MSE

Method	Test MSE
Elastic Net	0.010359
LASSO	0.010823
CW-LASSO	0.030824

The empirical results show that Elastic Net produced the lowest test MSE, followed closely by LASSO, while CW-LASSO recorded the largest prediction error. The overall comparison confirms that the methods solve different parts of the problem. OLS provides a useful benchmark but is unsuitable when the objective is sparse model identification. Ridge improves coefficient stabilization under multicollinearity but does not remove irrelevant variables. LASSO provides substantial sparsity and relatively strong false-positive control but becomes unstable under severe multicollinearity. Elastic Net provides the strongest balance between sparsity, signal recovery, and stability. Huber regression addresses the influence of extreme errors but does not produce a sparse model.

The findings consequently point to a methodological gap that multicollinearity and heavy-tailed errors are not independent concerns. A method may be effective against one while remaining vulnerable to the other. The evidence suggests that a desirable estimator should simultaneously incorporate a robust loss function to limit the effect of extreme observations and a sparsity-inducing, correlation-aware penalty to achieve reliable variable selection. This is particularly important because the simulation demonstrates that robust estimation alone, as represented by Huber regression, does not automatically produce a parsimonious model. The interpretation of the results also requires two corrections to the earlier version of the analysis. First, the empirical tables report an FPR of 0.5000, rather than 1.000, for OLS and Ridge. The present tables consistently use the reported numerical value. Second, SCAD and MCP are not included in the empirical ranking because their numerical simulation results were not supplied. This avoids making conclusions that cannot be supported by the available evidence.

CONCLUSION

This study investigates variable selection under the simultaneous challenges of multicollinearity and heavy-tailed errors using a high-dimensional regression framework containing 100 candidate predictors, of which only ten are truly active. The revised analysis evaluates regression procedures from multiple perspectives, including predictive accuracy, sparsity, signal recovery, false-positive control, coefficient stability and selection stability. The evidence demonstrates that variable-selection performance cannot be adequately assessed using prediction error alone. OLS and Ridge can produce competitive prediction results while retaining the entire candidate predictor set. Their high TPR therefore does not imply effective variable selection because the majority of retained predictors are irrelevant. LASSO provides substantial dimensionality reduction and is capable of recovering a large proportion of the truly active variables. However, its performance becomes less stable as predictor correlation increases. The existing stability results show a marked increase in LASSO selection instability under severe multicollinearity, indicating that highly correlated predictors can cause the selected model to vary substantially across samples. Elastic Net provides a strong conventional benchmark because its combination of L_1 and L_2 penalties provides a useful balance between sparsity and stability in the presence of correlated predictors. Its grouping behaviour makes it particularly attractive when explanatory variables occur in correlated groups. Nevertheless, Elastic Net should not be interpreted as an unequivocal solution to variable selection under simultaneous multicollinearity and heavy-tailed errors. Its remaining false-positive selection demonstrates that the grouping advantage of the L_2 component can come at the cost of model sparsity. In addition, conventional Elastic Net relies on squared-error loss and therefore does not directly provide the robustness required under severe contamination or heavy-tailed disturbances.

Huber regression illustrates the complementary limitation of robust estimation. Although the Huber loss reduces the influence of extreme residuals, robust estimation alone does not ensure sparse variable selection. The

inclusion of Huber-LASSO and LAD-LASSO in the revised empirical framework is therefore important because these methods provide direct tests of whether robustness and sparsity can be achieved simultaneously. The extension from a single equicorrelation structure to equicorrelation, block correlation and AR(1) correlation also strengthens the generality of the simulation. It allows the analysis to determine whether conclusions depend on the particular mechanism through which multicollinearity is generated.

The correction of the false-positive-rate calculation is another important improvement. With 100 candidate variables and ten active predictors, the number of null predictors is 90 and therefore

$FPR = \frac{FP}{90}$. This definition must be applied consistently throughout the empirical analysis. The previously reported FPR values that are inconsistent with the selected-variable and true-positive counts should be recalculated from the underlying simulation output.

The revised analysis also incorporates Monte Carlo standard errors and confidence intervals so that differences between competing procedures can be assessed in relation to simulation uncertainty rather than relying solely on point estimates.

Taken together, the evidence suggests that no single conventional method simultaneously solves all aspects of the variable-selection problem. OLS provides a classical benchmark but lacks sparsity; Ridge improves coefficient stabilization but retains irrelevant variables; LASSO provides sparsity but can be unstable under strong predictor correlation; Huber provides robustness but does not inherently select variables; and Elastic Net provides the strongest conventional compromise between sparsity and correlation management but remains vulnerable to heavy-tailed errors and false-positive selection.

Accordingly, Elastic Net is most appropriately described as a strong benchmark or practical compromise among the conventional procedures evaluated, rather than an unequivocal solution. Its remaining limitations provide a clear motivation for methods that integrate robustness, sparsity and correlation awareness within a single estimation framework. The broader methodological implication is that a desirable estimator for high-dimensional regression under difficult data conditions should simultaneously incorporate a robust loss function to reduce the influence of extreme observations, an L_1 -type penalty to eliminate irrelevant predictors, and an L_2 -type component or related correlation-aware mechanism to stabilize selection among highly correlated predictors. The integration of these properties represents a promising direction for future methodological development and provides a stronger foundation for reliable and interpretable variable selection in the presence of multicollinearity and heavy-tailed errors.

REFERENCES

1. Adomaityte, U., Defilippis, L., Loureiro, B., & Sicuro, G. (2024). High-dimensional robust regression under heavy-tailed data: Asymptotics and universality. *Journal of Statistical Mechanics: Theory and Experiment*, 2024(11), 114002. <https://doi.org/10.1088/1742-5468/ad65e6> (DOI)
2. Amato, U., Antoniadis, A., De Feis, I., & Gijbels, I. (2022). Wavelet-based robust estimation and variable selection in nonparametric additive models. *Statistics and Computing*, 32, 11. <https://doi.org/10.1007/s11222-021-10065-z> (Springer Link)
3. Babarinsa, O., Edogbanya, H., Abari, O., & Adeniyi, I. (2025). Regression estimation and feature selection using modified correlation-adjusted Elastic Net penalties. *FUDMA Journal of Sciences*, 9(1), 29–40. <https://doi.org/10.33003/fjs-2025-0901-2774> (FUDMA Journal of Sciences)
4. Bottmer, L., Croux, C., & Wilms, I. (2022). Sparse regression for large data sets with outliers. *European Journal of Operational Research*, 297(2), 782–794. <https://doi.org/10.1016/j.ejor.2021.05.049> (Maastricht University)
5. Bunea, F. (2008). Consistent selection via the Lasso for high dimensional approximating regression models. In *Pushing the limits of contemporary statistics: Contributions in honor of Jayanta K. Ghosh* (pp. 122–137). Institute of Mathematical Statistics. <https://doi.org/10.1214/074921708000000101> (ResearchGate)

6. Chamlal, H., Benzmane, A., & Ouaderhman, T. (2024). Elastic net-based high dimensional data selection for regression. *Expert Systems with Applications*, 244, 122958. <https://doi.org/10.1016/j.eswa.2023.122958> (ScienceDirect)
7. Chatla, S. B., & Mandal, A. (2026). Robust variable selection in high-dimensional nonparametric additive model. *Annals of the Institute of Statistical Mathematics*, 78, 115–140. <https://doi.org/10.1007/s10463-025-00939-8> (DOI)
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. <https://doi.org/10.1198/016214501753382273>
9. Ghosh, A., Jaenada, M., & Pardo, L. (2024). Robust adaptive variable selection in ultra-high dimensional linear regression models. *Journal of Statistical Computation and Simulation*, 94(3), 571–603. <https://doi.org/10.1080/00949655.2023.2262669> (Taylor & Francis Online)
10. Guo, X., Li, Y., & Hsing, T. (2025). Variable selection and minimax prediction in high-dimensional functional linear models. *Statistica Sinica*. <https://www3.stat.sinica.edu.tw/statistica/fp/SS-2025-0151.html> (Stat Sinica)
11. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67. <https://doi.org/10.1080/00401706.1970.10488634> (Taylor & Francis Online)
12. Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1), 73–101. <https://doi.org/10.1214/aoms/1177703732>
13. Jiang, Y., Lu, H., Huang, X., & Jiang, R. (2025). Robust variable selection via the weighted elastic-net multi-step screening procedure. *Hacettepe Journal of Mathematics and Statistics*, 54(6), 2350–2362. <https://doi.org/10.15672/hujms.1701976> (DergiPark)
14. Kepplinger, D. (2023). Robust variable selection and estimation via adaptive elastic net S-estimators for linear regression. *Computational Statistics & Data Analysis*, 183, 107730. <https://doi.org/10.1016/j.csda.2023.107730> (ScienceDirect)
15. Kurata, S., & Hirose, K. (2026). Robust and consistent model evaluation criteria in high-dimensional regression. *Journal of Statistical Planning and Inference*, 242, 106358. <https://doi.org/10.1016/j.jspi.2025.106358> (Kyushu University Pure Portal Site)
16. Lambert-Lacroix, S., & Zwald, L. (2011). Robust regression through the Huber's criterion and adaptive Lasso penalty. *Electronic Journal of Statistics*, 5, 101–120. [I recommend verifying the DOI before final submission.]
17. Li, G., Peng, H., & Zhang, J. (2012). Robust rank correlation based screening. *The Annals of Statistics*, 40(3), 1846–1877. <https://doi.org/10.1214/12-AOS1023>
18. Li, H., Xu, X., Lu, Y., Yu, X., Zhao, T., & Zhang, R. (2022). Robust variable selection based on relaxed LAD Lasso. *Symmetry*, 14(10), 2161. <https://doi.org/10.3390/sym14102161>
19. Liu, Z., Song, Y., & Cheng, Y. (2023). Robust variable selection with exponential squared loss for the spatial Durbin model. *Entropy*, 25(2), 249. <https://doi.org/10.3390/e25020249>
20. Man, R., Tan, K. M., Wang, Z., & Zhou, W.-X. (2024). Retire: Robust expectile regression in high dimensions. *Journal of Econometrics*, 239(2), 105459. <https://doi.org/10.1016/j.jeconom.2023.04.004> (ScienceDirect)
21. Meinshausen, N., & Bühlmann, P. (2006). High-dimensional graphs and variable selection with the Lasso. *The Annals of Statistics*, 34(3), 1436–1462. <https://doi.org/10.1214/009053606000000281>
22. Muthukrishnan, R., & James, C. K. (2024). The effect of multicollinearity on feature selection. *Indian Journal of Science and Technology*, 17(35), 3664–3668. <https://doi.org/10.17485/IJST/v17i35.1876>
23. Nan, R., Wang, J., Li, H., & Luo, Y. (2025). Robust variable selection via Bayesian LASSO-composite quantile regression with empirical likelihood: A hybrid sampling approach. *Mathematics*, 13(14), 2287. <https://doi.org/10.3390/math13142287> (MDPI)
24. Park, E., Wong, R. K., Kwon, J., & Chu, V. W. (2024). A stable model for maximizing the number of significant features. *International Journal of Data Science and Analytics*, 17, 443–460. <https://doi.org/10.1007/s41060-023-00494-6>
25. Song, Y., Han, H., Fu, L., & Wang, T. (2024). Penalized weighted smoothed quantile regression for high-dimensional longitudinal data. *Statistics in Medicine*, 43(10), 2007–2042. <https://doi.org/10.1002/sim.10056>

26. Song, Y., Li, Z., & Fang, M. (2022). Robust variable selection based on penalized composite quantile regression for high-dimensional single-index models. *Mathematics*, 10(12), 2000. <https://doi.org/10.3390/math10122000>
27. Su, P., Tarr, G., & Muller, S. (2024). Robust variable selection under cellwise contamination. *Journal of Statistical Computation and Simulation*, 94(6), 1371–1387. <https://doi.org/10.1080/00949655.2023.2286316> (Taylor & Francis Online)
28. 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>
29. Wang, H., Li, G., & Jiang, G. (2007). Robust regression shrinkage and consistent variable selection through the LAD-Lasso. *Journal of Business & Economic Statistics*, 25(3), 347–355. <https://doi.org/10.1198/073500106000000251> (Taylor & Francis Online)
30. Wang, W., Liang, J., Liu, R., Song, Y., & Zhang, M. (2022). A robust variable selection method for sparse online regression via the Elastic Net penalty. *Mathematics*, 10(16), 2985. <https://doi.org/10.3390/math10162985>
31. Wang, Y., & Karunamuni, R. J. (2022). High-dimensional robust regression with LqL_q-loss functions. *Computational Statistics & Data Analysis*, 176, 107567. <https://doi.org/10.1016/j.csda.2022.107567> (ScienceDirect)
32. Wu, X., Liang, R., & Yang, H. (2022). Penalized and constrained LAD estimation in fixed and high dimension. *Statistical Papers*, 63(1), 53–95. <https://doi.org/10.1007/s00362-021-01229-0>
33. Wu, X., Wang, M., Hu, W., Tian, G.-L., & Li, T. (2023). Penalized M-estimation based on standard error adjusted adaptive Elastic-Net. *Journal of Systems Science and Complexity*, 36, 1265–1284. <https://doi.org/10.1007/S11424-023-1400-0> (DOI)
34. Ye, Y., Shao, Y., & Li, C. (2022). A sparse approach for high-dimensional data with heavy-tailed noise. *Economic Research-Ekonomska Istraživanja*, 35(1), 2764–2780. <https://doi.org/10.1080/1331677X.2021.1978306>
35. Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2), 894–942. <https://doi.org/10.1214/09-AOS729>
36. Zou, H. (2006). The adaptive Lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476), 1418–1429. <https://doi.org/10.1198/016214506000000735>
37. 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>