



COMPARATIVE PERFORMANCE OF CLASSICAL AND REGULARIZED REGRESSION MODELS FOR SMALL- SAMPLE CONTINUOUS DATA

Vyankatesh P Bhaurkar¹, Arpita Deodikar²

Assistant Professor, Mechanical Engineering, Mechanical Engineering, AI, ML, Data Science, Material science, vibrations, wastewater treatment, Sanjivani College of Engineering Kopergaon, Email I'd: vyankatesh.bhaurkar@gmail.com

²Symbiosis Institute of Technology, Symbiosis International (Deemed) University, Pune 412115, India

Corresponding author: vyankatesh.bhaurkar@gmail.com

ARTICLE INFO

Received: 27/11/2025

Revised: 07/03/2026

Accepted: 15/04/2026

Published: 29/04/2026

DOI: <https://doi.org/10.69980/8mghe57>

Volume:12, Issue:02, April 2026

Abstract

This study compared the predictive performance and estimation stability of ordinary least squares (OLS), Ridge, Lasso, and Elastic Net regression under systematically varied small-sample conditions. An empirical dataset containing 21 numerical predictors was evaluated using repeated random samples of 30, 50, 75, 100, 150, and 200 observations, with an independent 300-observation test set retained for evaluation. Model performance was assessed using root mean squared error (RMSE), mean absolute error, coefficient of determination, coefficient variability, and variable-selection stability. Strong multicollinearity was identified among several predictors, with variance inflation factors approaching 20. At $n = 30$, OLS produced an RMSE of 71.02, compared with 31.81 for Ridge, 27.31 for Lasso, and 26.14 for Elastic Net. Prediction error and coefficient variability generally decreased as sample size increased, while performance differences between classical and regularized estimation progressively narrowed. Friedman and Holm-adjusted Wilcoxon tests confirmed significantly lower RMSE for the regularized estimators relative to OLS across all sample-size conditions. Lasso produced the sparsest solutions, whereas Elastic Net retained broader groups of correlated predictors. Overall, regularization substantially improved predictive and coefficient stability when observations were limited and predictors were strongly correlated, with its relative advantage diminishing as sample availability increased.

Keywords: Small-sample regression, Regularization, Ridge regression, Lasso regression, Elastic Net

1. Introduction

Regression analysis is a basic statistical technique used to explore the association between a continuous dependent variable and several independent variables. Ordinary least squares (OLS) is widely used due to its ease of computation, easily-interpreted coefficients and known statistical properties. However, when there is not a large sample compared to the predictor structure, its reliability may be reduced. Small samples offer less information to estimate regression parameters, and may make the model more susceptible to individual observations, impacting the accuracy of predictions and the stability of the model. Multivariable prediction modelling with continuous outcomes is thus dependent on adequate sample size, which should be determined to take account of model complexity and expected predictive performance, rather than following traditional rules (Riley et al., 2019). Such difficulties are more apparent if explanatory variables are highly correlated. Even when the overall model fit seems good, multicollinearity can lead to high variances of the coefficients and to unstable estimates. OLS is especially sensitive because correlated predictors include some overlap of information which is difficult to decompose with limited observations. Ridge regression tackles this issue by shrinking the regression coefficients, which helps to minimize the variance caused by the multicollinearity (Schreiber-Gregory, 2018). Multicollinearity can impact statistical estimation and predictive modelling more generally, so identification and control of multicollinearity is important in data-driven analyses (Chan et al., 2022). In Ridge regression, further development of Ridge-based estimators also clearly indicates the importance of dealing with predictor dependence in linear regression (Dar et al., 2023).

The unstable estimation is a systematic way to control it using Regularization of model fitting by adding some penalties. While ridge has an L2 penalty which continuously pushes down the coefficients toward zero, lasso has an L1 penalty which can push down some of the coefficients all the way to zero. Elastic Net is a fusion of both penalties which shrinks the coefficients and selects variables with simultaneous shrinkage and variable selection and is robust to correlated predictors. These approaches change the bias–variance balance by intentionally adding variance to the estimate to lower the variance of the estimate and possibly gain generality (James et al., 2021). Thus, regularization has gradually emerged as an essential method for regularizing model complexity and ensuring model robustness in statistical and machine learning applications (Tian & Zhang, 2022). In small sample regression, one needs to worry about variables selected for the model. While a fewer number of predictors may improve parsimony, selected predictors can vary significantly from one sample to another. Therefore, careful assessment of selection procedures is required before being able to conclude that a selected model is stable or reproducible (Heinze et al., 2018). The variable-retention properties of penalized approaches also vary, yielding different sparsity and predictive performance with alternative penalties (Greenwood et al., 2020). The relative performance of the penalized regression estimators also relies on the underlying sample and predictor setup as shown in simulation experiments (Genç, 2022).

The availability of samples is at the core of these relationships. Finite-sample analyses have shown that prediction, variable selection and predictor ranking can be quite different when the information that allows for estimation is limited (Wang et al., 2020). For reliable estimates of predictive performance, in independent validation, an appropriate sample size is also crucial (Archer et al., 2021). The results of broader experimental comparisons of regression methods have further revealed that method and dataset can differ significantly in predictive performance, favoring empirical comparisons to a uniform performance of one of the regression methods across all analytical situations (Fernández-Delgado et al., 2019). The accuracy of predictions, however, does not sufficiently define regression reliability. Models with comparable average prediction errors can have very different coefficient stability and sensitivity to variations in the training observations. When sampling variability matters, predictive-stability criteria offer an available approach to evaluate penalized linear models and choose the penalty parameters (Dustin et al., 2024). In low dimensional regression, even in cases where there are more observations than predictors and predictors are correlated, a meaningful estimation instability can occur due to shrinkage (Kipruto & Sauerbrei, 2024).

However, there is little comparative literature that combines predictions of accuracy, coefficient stability, variable selection practices, and statistical model comparisons as the sample size varies systematically. A detailed analysis of these dimensions can shed light on the relative performance of classical and regularized regression estimators with increasing sample size. Accordingly, this study pursued three objectives:

1. To compare predictive performance of classical and regularized regression models across systematically varied small-sample conditions.
2. To evaluate coefficient stability and prediction variability of regression estimators under limited samples and correlated predictors.
3. To determine how increasing sample size influences the relative behavior of OLS, Ridge, Lasso, and Elastic Net.

2. Materials and Methods

2.1 Dataset and Study Variables

The data analyzed was a numeric set with 6999 observations in the training set and an independent set of 300 observations in the test set. There were 21 continuous predictor variables in each observation representing activity and resource properties of the system, such as read/write operations, system calls, process activity, page faults, and memory and swap availability and related computation properties. The dependent variable was continuous with the range of 0 to 99. The 6,999 observations served as the empirical sampling pool from which controlled small-sample training sets were repeatedly sampled. Throughout model development and evaluation a separate test set of 300 observations (PMLB regression datasets, 2018) was maintained. This allowed to directly assess the regression behavior under systematically reduced sample availability while maintaining the empirical predictor distributions and correlation structure.

2.2 Data Preprocessing and Exploratory Analysis

Data quality was assessed prior to model estimation and included missing data, duplicates, variable type, variable range and variable distribution. There were no values missing or duplicate observations found. A continuous variable was presented as the mean, standard deviation, median, minimum and maximum. Pearson correlation coefficients were computed to measure the linear relationship between the predictors and the response variable, and between individual predictors and the response variable. The multicollinearity was also checked by variance inflation factors (VIFs). All predictors were standardized to zero mean and unit variance since the regression coefficients were penalized and it is important to have the same scale for all predictors. Only one of the training sets was sampled for the standardisation parameters and these were then used for the other independent test sets, to avoid information leakage in the evaluation of the model.

2.3 Small-Sample Experimental Design

A subsampling strategy was repeated to estimate the performance of the model in the case of limited sample size. Six training-sample conditions were examined: $n = 30, 50, 75, 100, 150$, and 200 . For each sample size, 30 independent random samples were taken from the 6,999 observation training pool. In each replication, the same observations were provided to all four regression models so that the models could be compared independently of any differences in samples. The design resulted in 180 independent training subsets and 720 primary model fits for each of the four regression approaches, with the exception of the internal hyper parameter fitting approach. For all sizes of sample size, and replications, the independent 300-observation test set was the same, giving a common basis for evaluating and quantifying changes in terms of changes in the test set due to having different training samples.

2.4 Regression Model Specification

Four regression estimators were evaluated: ordinary least squares (OLS), Ridge, Lasso, and Elastic Net. OLS represented the classical unregularized benchmark and estimated regression coefficients by minimizing the residual sum of squares. Ridge regression incorporated an L2penalty to shrink regression coefficients and reduce estimation variance in the presence of correlated predictors:

$$\hat{\beta}^{Ridge} = \arg \min_{\beta} \left[\sum_{i=1}^n (y_i - \beta_0 - X_i \beta)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right].$$

Lasso regression incorporated an L1penalty, enabling simultaneous coefficient shrinkage and variable selection by reducing selected coefficients exactly to zero:

$$\hat{\beta}^{Lasso} = \arg \min_{\beta} \left[\sum_{i=1}^n (y_i - \beta_0 - X_i \beta)^2 + \lambda \sum_{j=1}^p |\beta_j| \right].$$

Elastic Net combined L1 and L2 regularization:

$$\hat{\beta}^{EN} = \arg \min_{\beta} \left[\sum_{i=1}^n (y_i - \beta_0 - X_i \beta)^2 + \lambda \left(\alpha \sum_{j=1}^p |\beta_j| + (1 - \alpha) \sum_{j=1}^p \beta_j^2 \right) \right].$$

Here, λ represented overall regularization strength, whereas α determined the relative contributions of the L1 and L2 penalties. The four-model framework therefore captured unregularized estimation, continuous coefficient shrinkage, sparse variable selection, and combined shrinkage-selection behavior.

2.5 Hyperparameter Selection and Model Validation

Optimization of regularization parameters was carried out separately in each training set sampled. For Ridge and Lasso models, the penalty parameter λ needed to be chosen, while for Elastic Net, optimization of the regularization strength and the L₁/L₂ mixing parameter was required. Internal cross-validation was used to test which candidate combinations of parameter values had the best validation prediction error, and those combinations that had the lowest prediction error were kept. Only the samples used as the training set were used for all the steps of the preprocessing and for the estimation of the parameters and hyperparameters. The independent 300-observation test set was not used in the pre-processing estimation, model fitting, or parameter selection stages, but was only used to evaluate the performance in the end. The hyperparameters of the regularizations were not optimized by OLS, and the hyperparameters of the predictors and observations sampled for OLS were the same as for the penalized models.

2.6 Predictive Performance and Stability Assessment

Predictive accuracy was evaluated using root mean squared error (RMSE), mean absolute error (MAE), and the coefficient of determination (R²). RMSE was calculated as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2},$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|.$$

The coefficient of determination was calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}.$$

Performance metrics were calculated on the independent test set for every model, sample size, and replication and summarized using means and standard deviations. Coefficient stability was quantified from the variability of standardized regression coefficients across repeated samples, including mean and maximum coefficient standard deviations. Prediction stability was assessed from between-replication variability in RMSE. For Lasso and Elastic Net, variable-selection stability was additionally quantified as the proportion of repeated samples in which each predictor retained a non-zero coefficient.

2.7 Statistical Comparison of Regression Performance

A matched repeated measures design was used for statistical comparisons because the same observations were used to test the four regression methods in each replication. The differences between the OLS, Ridge, Lasso and Elastic Net were first tested separately at each sample size using the Friedman test. A set of significant omnibus results was followed by paired Wilcoxon signed rank comparisons. The Holm procedure was used to adjust for family-wise error due to multiple pairwise tests. To measure the size and direction of the difference in performances at each level an index of rank-biserial effect was computed. A two tailed cutoff of $p < 0.05$ was used to determine statistical significance. To assess the behavior of the model across small-sample conditions, interpretation included predictive error, coefficient variability, variable-selection stability, adjusted statistical significance, and effect magnitude.

3. Results

3.1 Dataset Characteristics and Predictor Structure

The training dataset contained 6,999 observations with 21 numerical predictors and one quantitative outcome, while an independent set of 300 observations was retained for model evaluation. No missing values or duplicate observations were identified. The outcome ranged from 0 to 99, with a mean of 83.94 ± 18.47 and a median of 89. Considerable variation was observed across predictors. The strongest associations with the outcome were observed for freeswap ($r = 0.680$), runqsz ($r = -0.622$), vflt ($r = -0.420$), pflt ($r = -0.372$), and fork ($r = -0.363$). Table 1 shows the descriptive characteristics of the principal variables.

Table 1. Descriptive Characteristics of the Principal Study Variables

Variable	Mean	SD	Median	Minimum	Maximum
scall	2306.94	1638.31	2048.00	109.00	12493.00
sread	209.88	190.55	165.00	6.00	4449.00
swrite	149.36	149.56	117.00	7.00	4590.00
fork	1.89	2.50	0.80	0.00	20.12
exec	2.80	5.25	1.20	0.00	59.56
pflt	109.71	114.56	63.67	0.00	899.80
vflt	185.08	191.63	120.84	0.20	1365.00
runqsz	19.81	128.05	2.00	1.00	2823.00
freemem	1777.58	2488.92	582.00	55.00	12027.00
freeswap	1328982.32	423497.94	1292978.00	2.00	2243187.00
Target	83.94	18.47	89.00	0.00	99.00



Figure 1. Correlation Structure Among Predictor Variables

Figure 1 shows pronounced correlation clusters among the predictor variables. The strongest relationships occurred between fork and vflt ($r = 0.941$), pflt and vflt ($r = 0.938$), fork and pflt ($r = 0.933$), ppgout and pgfree ($r = 0.921$), and pgin and ppgin ($r = 0.921$), indicating substantial redundancy within the predictor space.

3.2 Multicollinearity and Predictor–Outcome Relationships

Multicollinearity diagnostics further confirmed the strong dependence among predictors. Table 2 shows that VIF values reached 19.81 for vflt, 18.83 for fork, 18.34 for pgfree, and 15.42 for ppgout. In contrast, the two predictors most strongly associated with the outcome, freeswap and runqsz, had relatively low VIF values of 2.23 and 1.28, respectively. The simultaneous presence of strong predictor–outcome associations and substantial inter-predictor dependence established a challenging setting for coefficient estimation under restricted sample sizes.

Table 2. Predictor–Outcome Correlations and Multicollinearity Diagnostics

Predictor	Correlation with Target	VIF
freeswap	0.680	2.23
runqsz	−0.622	1.28
vflt	−0.420	19.81
pflt	−0.372	11.83
fork	−0.363	18.83
sread	−0.344	6.31
rchar	−0.334	2.01
scall	−0.322	2.49
swrite	−0.289	5.06
exec	−0.288	3.21
freemem	0.271	1.70
pgin	−0.244	7.43
ppgin	−0.230	7.82
ppgout	−0.214	15.42
pgfree	−0.213	18.34
pgscan	−0.178	7.75

3.3 Predictive Performance Across Small-Sample Conditions

Predictive performance differed markedly between OLS and the regularized estimators, particularly under the smallest training samples. Table 3 shows that at $n = 30$, OLS produced an RMSE of 71.02 ± 67.46 , whereas Ridge, Lasso, and Elastic Net achieved 31.81 ± 41.80 , 27.31 ± 17.99 , and 26.14 ± 17.47 , respectively. Corresponding MAE values were 19.07 for OLS and 7.34–8.69 for the regularized models.

Prediction error declined substantially as sample size increased. At $n = 100$, RMSE decreased to 17.34 for OLS, 13.11 for Ridge, 12.77 for Lasso, and 12.30 for Elastic Net. At $n = 150$, Elastic Net recorded an RMSE of 11.06 ± 3.98 and R^2 of 0.503, compared with an RMSE of 14.19 ± 5.69 and R^2 of 0.161 for OLS. At $n = 200$, the performance gap narrowed,

Table 3. Predictive Performance Across Small-Sample Training Conditions

Sample size	Model	RMSE, mean $\pm$ SD	MAE, mean $\pm$ SD	Mean R ²
30	OLS	71.02 $\pm$ 67.46	19.07 $\pm$ 12.13	-33.076
	Ridge	31.81 $\pm$ 41.80	8.69 $\pm$ 6.21	-8.746
	Lasso	27.31 $\pm$ 17.99	7.40 $\pm$ 3.07	-2.822
	Elastic Net	26.14 $\pm$ 17.47	7.34 $\pm$ 3.01	-2.529
50	OLS	31.71 $\pm$ 15.77	9.27 $\pm$ 3.34	-3.497
	Ridge	19.76 $\pm$ 14.10	6.41 $\pm$ 1.83	-1.102
	Lasso	22.86 $\pm$ 16.27	6.45 $\pm$ 1.82	-1.809
	Elastic Net	22.62 $\pm$ 16.13	6.62 $\pm$ 1.80	-1.754
75	OLS	23.39 $\pm$ 12.08	7.24 $\pm$ 2.83	-1.484
	Ridge	15.86 $\pm$ 10.03	5.46 $\pm$ 1.42	-0.259
	Lasso	16.64 $\pm$ 10.73	5.66 $\pm$ 1.61	-0.400
	Elastic Net	15.74 $\pm$ 10.04	5.33 $\pm$ 1.30	-0.245
100	OLS	17.34 $\pm$ 9.54	6.81 $\pm$ 1.29	-0.402
	Ridge	13.11 $\pm$ 5.68	5.46 $\pm$ 0.72	0.267
	Lasso	12.77 $\pm$ 6.03	5.49 $\pm$ 0.97	0.285
	Elastic Net	12.30 $\pm$ 5.09	5.28 $\pm$ 0.71	0.364
150	OLS	14.19 $\pm$ 5.69	5.97 $\pm$ 1.26	0.161
	Ridge	11.24 $\pm$ 4.28	5.07 $\pm$ 0.82	0.480
	Lasso	12.45 $\pm$ 4.86	5.61 $\pm$ 1.38	0.358
	Elastic Net	11.06 $\pm$ 3.98	5.02 $\pm$ 0.84	0.503
200	OLS	13.28 $\pm$ 5.50	5.37 $\pm$ 1.31	0.258
	Ridge	11.88 $\pm$ 4.76	4.82 $\pm$ 1.03	0.412
	Lasso	12.03 $\pm$ 4.87	5.17 $\pm$ 1.25	0.395
	Elastic Net	11.71 $\pm$ 4.91	4.83 $\pm$ 1.02	0.421

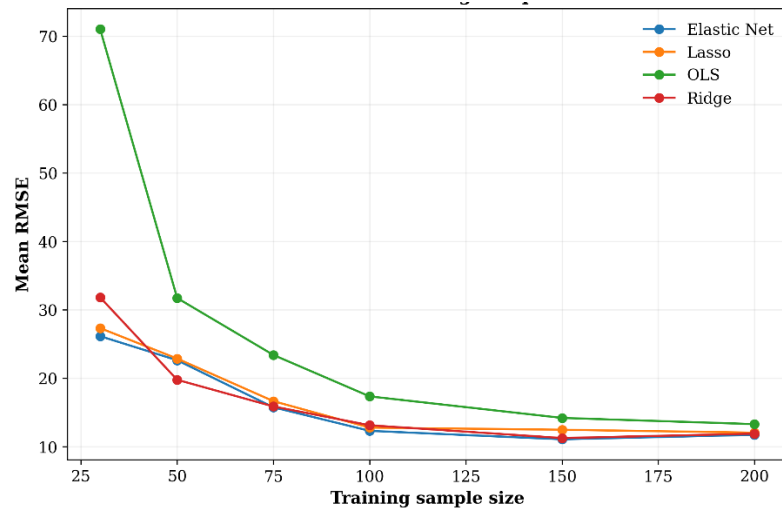


Figure 2. RMSE Profiles of Regression Models Across Training Sample Sizes

Figure 2 shows a pronounced decline in RMSE between $n = 30$ and $n = 100$, followed by comparatively smaller changes at larger sample sizes. The separation between OLS and the regularized models was greatest under the smallest sample condition and progressively narrowed with increasing sample availability.

3.4 Prediction Variability and Coefficient Stability

Regularization substantially improved coefficient stability. At $n = 30$, the mean standard deviation of standardized coefficients was 11.515 for OLS, compared with 2.793 for Ridge, 2.544 for Lasso, and 2.440 for Elastic Net. Table 4 shows that coefficient variability declined across all models as sample size increased but remained consistently greater for OLS. At $n = 200$, mean coefficient variability decreased to 2.134 for OLS, 0.927 for Ridge, 0.881 for Lasso, and 0.831 for Elastic Net.

Prediction variability exhibited a similar pattern. At $n = 30$, the RMSE standard deviation was 67.459 for OLS, compared with 41.802 for Ridge, 17.994 for Lasso, and 17.468 for Elastic Net. At $n = 150$, RMSE variability had declined to 5.693 for OLS and 3.978–4.862 across the regularized estimators.

Table 4. Coefficient and Prediction Stability Across Sample Sizes

n	Model	Mean coefficient SD	Maximum coefficient SD	RMSE SD
30	OLS	11.515	40.981	67.459
	Ridge	2.793	9.473	41.802
	Lasso	2.544	9.813	17.994
	Elastic Net	2.440	9.455	17.468
50	OLS	6.827	15.836	15.774
	Ridge	1.915	4.484	14.097
	Lasso	1.460	3.725	16.270
	Elastic Net	1.574	4.231	16.130
75	OLS	4.137	9.173	12.083
	Ridge	1.327	4.036	10.034
	Lasso	1.259	4.252	10.727
	Elastic Net	1.184	3.853	10.035
100	OLS	3.881	8.635	9.543
	Ridge	1.357	3.098	5.682
	Lasso	1.572	4.253	6.032
	Elastic Net	1.008	3.159	5.087
150	OLS	3.210	7.829	5.693
	Ridge	1.101	2.964	4.277
	Lasso	1.244	4.519	4.862
	Elastic Net	1.046	2.773	3.978
200	OLS	2.134	4.561	5.500
	Ridge	0.927	2.566	4.764

	Lasso	0.881	3.159	4.871
	Elastic Net	0.831	2.674	4.908

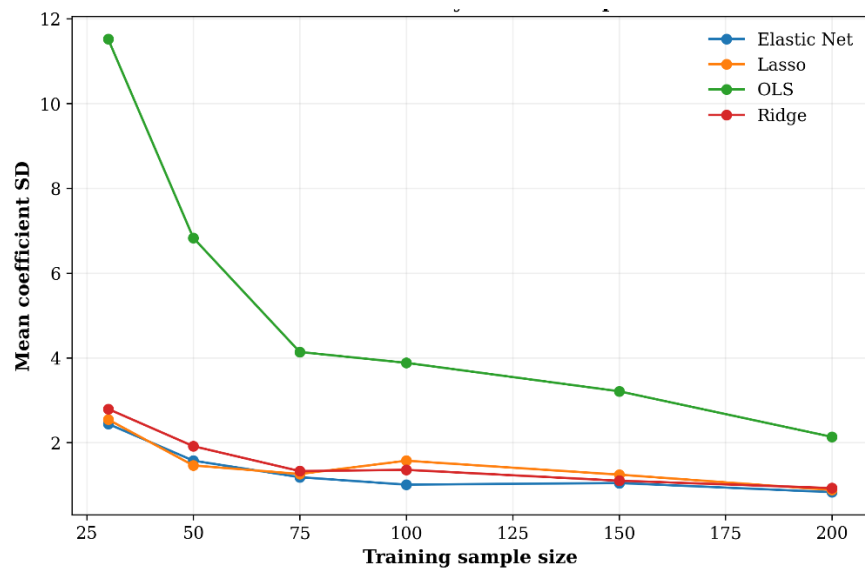


Figure 3. Coefficient Stability Across Increasing Sample Sizes

Figure 3 shows a consistent decline in coefficient variability with increasing sample size. OLS exhibited the greatest coefficient dispersion throughout the analysis, whereas regularized regression produced markedly more stable estimates, particularly under the smallest sample conditions.

3.5 Statistical Comparison of Model Performance

Significant differences in RMSE among the four models were observed at every sample size. Friedman tests yielded $\chi^2(3) = 42.92$ at $n = 30$, $\chi^2(3) = 36.88$ at $n = 50$, $\chi^2(3) = 50.12$ at $n = 75$, $\chi^2(3) = 47.48$ at $n = 100$, $\chi^2(3) = 38.92$ at $n = 150$, and $\chi^2(3) = 29.88$ at $n = 200$; all $p < 0.001$.

Holm-adjusted paired Wilcoxon comparisons demonstrated significantly lower RMSE for each regularized estimator relative to OLS across all sample-size conditions. Table 5 shows that the magnitude of the difference was greatest at $n = 30$, where OLS exceeded Ridge by 39.215 RMSE units, Lasso by 43.709, and Elastic Net by 44.888. At $n = 200$, these differences narrowed to 1.398, 1.244, and 1.563, respectively.

Table 5. Pairwise Comparison of OLS With Regularized Regression Models

n	Comparison	Mean RMSE difference*	Holm-adjusted p	Rank-biserial effect
30	OLS vs Ridge	39.215	<0.001	0.901
30	OLS vs Lasso	43.709	<0.001	0.931
30	OLS vs Elastic Net	44.888	<0.001	0.914
50	OLS vs Ridge	11.955	<0.001	0.970
50	OLS vs Lasso	8.853	<0.001	0.854
50	OLS vs Elastic Net	9.096	<0.001	0.901
75	OLS vs Ridge	7.530	<0.001	0.996
75	OLS vs Lasso	6.755	<0.001	0.970
75	OLS vs Elastic Net	7.652	<0.001	0.987
100	OLS vs Ridge	4.224	<0.001	0.944
100	OLS vs Lasso	4.565	<0.001	1.000

100	OLS vs Elastic Net	5.040	<0.001	0.978
150	OLS vs Ridge	2.946	<0.001	0.923
150	OLS vs Lasso	1.731	0.003	0.660
150	OLS vs Elastic Net	3.126	<0.001	0.914
200	OLS vs Ridge	1.398	<0.001	0.841
200	OLS vs Lasso	1.244	0.005	0.647
200	OLS vs Elastic Net	1.563	<0.001	0.905

*Positive values indicate higher RMSE for OLS.

Differences among the regularized models were smaller. Ridge and Elastic Net did not differ significantly at $n = 30, 50, 75, 150$, or 200 after multiplicity adjustment. Elastic Net showed a modest advantage over Ridge at $n = 100$ (adjusted $p = 0.046$) and lower RMSE than Lasso at $n = 75$ (adjusted $p = 0.039$) and $n = 150$ (adjusted $p < 0.001$).

3.6 Regularization and Variable-Selection Patterns

Variable-retention patterns differed substantially between Lasso and Elastic Net. Lasso generated comparatively sparse solutions, retaining a mean of 10.63 predictors at $n = 30$, 11.27 at $n = 50$, 10.57 at $n = 75$, 9.03 at $n = 100$, 8.77 at $n = 150$, and 9.77 at $n = 200$. Elastic Net retained broader predictor sets, with corresponding means of 18.00, 17.70, 16.43, 16.40, 17.77, and 17.03.

At $n = 200$, runqsz was retained in 100% of both Lasso and Elastic Net replications. freeswap was retained in 97% of Lasso and 100% of Elastic Net models. Elastic Net additionally retained pflt in 100% and fork and vflt in 97% of replications.

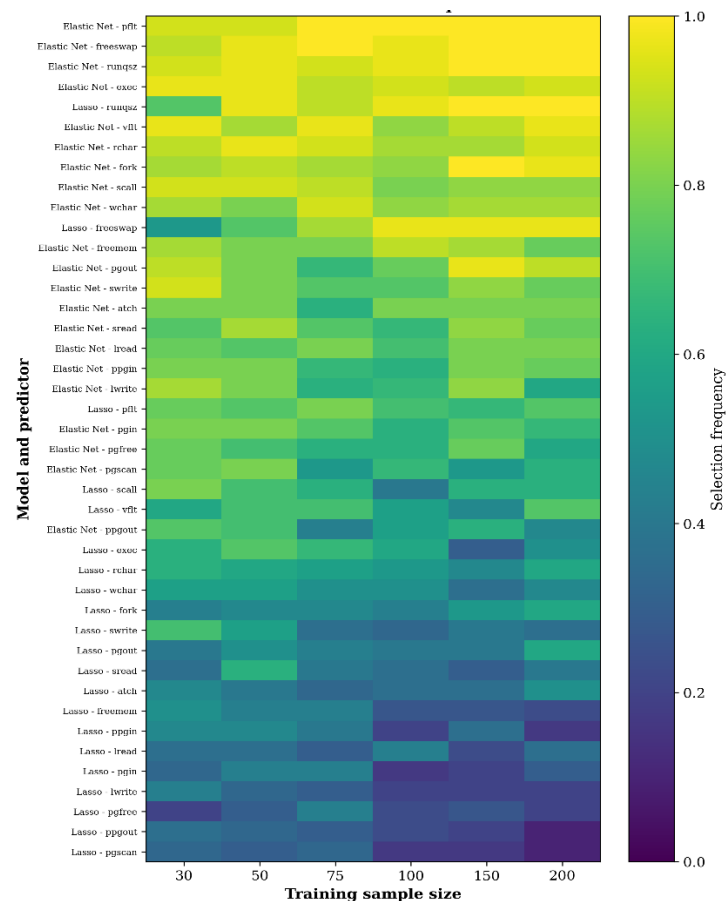


Figure 4. Predictor-Selection Frequencies Across Small-Sample Conditions

Figure 4 shows the contrasting sparsity patterns of Lasso and Elastic Net. Lasso retained a smaller subset of predictors, whereas Elastic Net preserved a broader group of correlated variables. Selection consistency became increasingly concentrated around the strongest predictors as sample size increased.

Overall, restricted sample availability was associated with pronounced increases in prediction error and coefficient Volume-12 | Issue-02 | April 2026

instability, with the greatest deterioration observed for OLS. Regularization substantially reduced estimation instability under small-sample conditions, while differences between classical and regularized regression progressively narrowed as sample size increased.

4. Discussion

The results show that sample availability had a significant impact on the predictive accuracy and coefficient stability of both classical regression and regularized regression models. The differences were greatest when the samples were severely restricted. At $n = 30$, OLS produced an RMSE of 71.02, compared with 31.81 for Ridge, 27.31 for Lasso, and 26.14 for Elastic Net. Overall, the prediction errors decreased with larger number of training samples and the gap between the OLS and penalized estimators was not very large. This trend suggests that the regularization had more influence when fewer observations were used for model estimation. Similar worries have been voiced for prediction models built on tiny datasets; penalization and shrinkage can be as sensitive to insufficient information as can be the model parameters (Riley et al., 2021), highlighting the fact that regularization can't remove the essential role of sample size.

The poor performance of OLS with the smallest samples was highly correlated with the high correlation among the predictors. Several pairs of predictors had correlation values above 0.90, with values close to 20 for corresponding VIF values. When this is the case, coefficient variance is increased by overlapping information in the predictors, and the OLS estimates are especially vulnerable to the composition of the sample. As shown by the mean coefficient SD of 11.515 with $n = 30$ for OLS, the values for the regularized estimators were between 2.440 and 2.793. Similar simulation and applied results have demonstrated that Ridge, Lasso and Elastic Net can perform better to control the negative effects of multicollinearity than unregularized estimation, depending on the correlation structure of the problem (Herawati et al., 2024).

The coefficient instability was significantly reduced for all the sample sizes analyzed upon regularization. Ridge did so by using continuous L_2 shrinkage, and Lasso and Elastic Net used shrinkage and various degrees of variable selection. At $n = 200$, mean coefficient SD remained lower for Ridge (0.927), Lasso (0.881), and Elastic Net (0.831) than for OLS (2.134). The results show that regularization did not only lead to better prediction accuracy, but it also caused an improvement in robustness to the changes in training sample composition. In limited-data settings, evidence also shows that regularized regression can be helpful when multicollinearity makes it difficult to obtain good estimates of the weights of individual predictors (Sztepanacz & Houle, 2024).

There were also different variable retention patterns. The sparseness of the solutions was relatively low for both lasso and elastic net, with about 9–11 predictors kept for lasso and 16–18 for elastic net. This separation is in line with the penalization they impose: L_1 component of Lasso imposes a penalty encouraging coefficient values to be zero, Elastic Net imposes L_1 selection plus L_2 shrinkage and can therefore allow groups of correlated variables. The relationships between runqsz and freeswap were relatively high and the two variables remained stable at higher samples. The coefficient recovery and retention differences we found between Ridge and Elastic Net are also found to be strongly dependent on the underlying predictor configuration, further emphasizing the need for accounting for the correlation structure when interpreting penalized estimates (Gana, 2022).

The most restricted and intermediate conditions seemed to be the extremes that showed the greatest effect on sample size. Elastic Net RMSE decreased from 26.14 at $n = 30$ to 12.30 at $n = 100$, while OLS decreased from 71.02 to 17.34. The improvement was not significant beyond that and there was a slight variation between $n = 150$ and 200. The results therefore confirm the general decrease in error as the amount of information increases, rather than a strictly monotonic error function for each successive size of information. Similarly, the Mean R^2 values were positive for the regularized models at $n = 100$ and for all four models at $n = 150$. More importantly, shrinkage does not always lead to better prediction under all data configurations, and it is dependent on various factors such as sample size, predictor structure or penalty estimation (Van Calster et al., 2020).

These observations were reinforced by the inferential comparisons. Friedman tests revealed significant differences in performance for all sample sizes and Holm adjusted Wilcoxon comparisons revealed that all regularized estimators had significantly lower RMSE than OLS. However, these differences were reduced significantly as the sample size increased. The OLS–Elastic Net RMSE difference declined from 44.888 at $n = 30$ to 1.563 at $n = 200$. There were significant difference between the Ridge and Lasso models, but differences between the Ridge and Elastic Net were relatively small and were not significantly different in all sample sizes. The absence of universal dominance is in line with other comparative evidence that the performance of OLS and penalized estimators depends on the statistical features of the data (Nayem et al., 2024).

The results have real-world applications for statistical analysis of small continuous datasets. Regardless of how accurate the prediction is, the regression coefficients can be very unstable, especially if the predictors are highly correlated. The repeated-sampling evaluation is therefore a valuable source of information on the robustness of fitted relationships, apart from the conventional test-set metrics. Regularization was most beneficial in the situation of fewer observations and more significant multicollinearity, and the larger the number of observations, the smaller the difference between penalized and classical estimation.

In this study, only one empirical predictor structure was considered and four linear regression estimators were evaluated. Differences observed might therefore vary with dimensionality, noise levels, signal strength and correlation structure. However, the consistent evidence obtained from the controlled sample-size framework, independent evaluation set, repeated sampling, stability assessment, and matched statistical comparisons all demonstrate that regularization significantly enhanced the stability of the estimation when the sample size was restricted. The overall results illustrate that, in addition to the difference between the estimators themselves, the comparative performance of classical and

5. Conclusion

This study revealed that classical regression models and regularized regression models performed differently when the data was continuous, and that this difference was significant with respect to sample size and predictor dependence. Overall, OLS was found to have significantly higher prediction error and more instability in the coefficients than Ridge, Lasso and Elastic Net in the most restrictive sample conditions. Regularization effects such as controlling the magnitude of the coefficients and the variance of the strong correlated predictors reduced this sensitivity. This benefit was greatest at the lowest number of sample, and diminished as more observations were added. Different regularized methods exhibited different features, not a most idealistic one. Ridge offered coefficient stability while Lasso yielded comparatively sparse models, and Elastic Net both stabilized coefficients and retained more correlated predictors. The differences in performance were significant between OLS and the regularized estimators for all sample sizes, and were smaller and less consistent between the penalized approaches. Sample size generally had a positive impact on predictive accuracy and stability of the coefficients, with the best increase between severely restricted and intermediate sample conditions. The results highlight the need to jointly consider all factors with regard to the selection of regression models when the sample size is small and the observations are continuous, namely sample size, multicollinearity, predictive error and coefficient stability. When observations are sparse and predictors are highly correlated, regularization is a particularly helpful technique; its usefulness diminishes as more information is available to estimate. The results thus indicate that it is not the best to use the same estimator for all conditions of data, but rather a sample-size-sensitive approach to regression modelling.

References

1. Archer, L., Snell, K. I., Ensor, J., Hudda, M. T., Collins, G. S., & Riley, R. D. (2021). Minimum sample size for external validation of a clinical prediction model with a continuous outcome. *Statistics in Medicine*, 40(1), 133-146.
2. Chan, J. Y. L., Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., & Chen, Y. L. (2022). Mitigating the multicollinearity problem and its machine learning approach: a review. *Mathematics*, 10(8), 1283.
3. Dar, I. S., Chand, S., Shabbir, M., & Kibria, B. G. (2023). Condition-index based new ridge regression estimator for linear regression model with multicollinearity. *Kuwait Journal of Science*, 50(2), 91-96.
4. Dustin, D., Clarke, B., & Clarke, J. (2024). Predictive stability criteria for penalty selection in linear models: D. Dustin et al. *Computational Statistics*, 39(3), 1241-1280.
5. Fernández-Delgado, M., Sirsat, M. S., Cernadas, E., Alawadi, S., Barro, S., & Febrero-Bande, M. (2019). An extensive experimental survey of regression methods. *Neural Networks*, 111, 11-34.
6. Gana, R. (2022). Ridge regression and the elastic net: How do they do as finders of true regressors and their coefficients?. *Mathematics*, 10(17), 3057.
7. Genç, M. (2022). Bir Simülasyon Çalışması ile Cezalı Regresyon Yöntemlerinin Karşılaştırılması. *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi*, 9(1), 80-91.
8. Greenwood, C. J., Youssef, G. J., Letcher, P., Macdonald, J. A., Hagg, L. J., Sanson, A., ... & Olsson, C. A. (2020). A comparison of penalised regression methods for informing the selection of predictive markers. *PloS one*, 15(11), e0242730.
9. Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection—a review and recommendations for the practicing statistician. *Biometrical journal*, 60(3), 431-449.
10. Herawati, N., Sutrisno, A., Nusyirwan, N., & Misgiyati, M. (2024). The performance of ridge regression, LASSO, and elastic-net in controlling multicollinearity: A simulation and application. *Journal of Modern Applied Statistical Methods*, 23(2), 1-13.
11. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: with applications in R* (Vol. 2). New York: springer.
12. Kipruto, E., & Sauerbrei, W. (2024). Post-Estimation Shrinkage in Full and Selected Linear Regression Models in Low-Dimensional Data Revisited. *Biometrical Journal*, 66(7), e202300368.
13. Nayem, H. M., Aziz, S., & Kibria, B. G. (2024). Comparison among ordinary least squares, ridge, lasso, and elastic net estimators in the presence of outliers: Simulation and application. *International Journal of Statistical Sciences*, 25-48.
14. PMLB regression datasets. (2018). Regression datasets. <https://doi.org/10.5281/zenodo.1475816>
15. Riley, R. D., Snell, K. I., Ensor, J., Burke, D. L., Harrell Jr, F. E., Moons, K. G., & Collins, G. S. (2019). Minimum sample size for developing a multivariable prediction model: Part I—Continuous outcomes. *Statistics in medicine*, 38(7), 1262-1275.
16. Riley, R. D., Snell, K. I., Martin, G. P., Whittle, R., Archer, L., Sperrin, M., & Collins, G. S. (2021). Penalization and shrinkage methods produced unreliable clinical prediction models especially when sample size was small. *Journal of Clinical Epidemiology*, 132, 88-96.
17. Schreiber-Gregory, D. N. (2018). Ridge Regression and multicollinearity: An in-depth review. *Model Assisted Statistics and Applications*, 13(4), 359-365.
18. Sztapanacz, J. L., & Houle, D. (2024). Regularized regression can improve estimates of multivariate selection in the face of multicollinearity and limited data. *Evolution Letters*, 8(3), 361-373.
19. Tian, Y., & Zhang, Y. (2022). A comprehensive survey on regularization strategies in machine learning. *Information Fusion*, 80, 146-166.

20. Van Calster, B., van Smeden, M., De Cock, B., & Steyerberg, E. W. (2020). Regression shrinkage methods for clinical prediction models do not guarantee improved performance: simulation study. *Statistical methods in medical research*, 29(11), 3166-3178.
21. Wang, F., Mukherjee, S., Richardson, S., & Hill, S. M. (2020). High-dimensional regression in practice: an empirical study of finite-sample prediction, variable selection and ranking: Wang et al. *Statistics and computing*, 30(3), 697-719.

EPH – International Journal of Mathematics and Statistics

<https://ejjms.com/index.php/ms/index>



This work is licensed under a Creative Commons Attribution 4.0 International License. The authors retain ownership of the copyright for their article, You are free to: Share — copy and redistribute the material in any medium or format for any purpose, even commercially. To see the complete license contents,

please visit <http://creativecommons.org/licenses/by/4.0/>.