International Journal of Advanced Statistics and Probability, 13 (1) (2026) 103-114

International Journal of Advanced Statistics and Probability

Website: www.sciencepubco.com/index.php/IJASP
https://doi.org/10.14419/swvhrx34
Research paper

Optimal Biasing Parameter Selection for Shrinkage Estimators In High-Dimensional Multicollinear Systems

Afeez Abolaji Lawal *, Oluwayemisi Oyeronke Alaba, Ibraheem Sanusi

Department of Statistics, Faculty of Science, University of Ibadan
**Corresponding author E-mail: afeezabolaji91@gmail.com*

Received: 20-04-2026, Accepted: 06-07-2026, Published: 25-07-2026

Abstract

Severe multicollinearity remains a challenge in regression analysis, leading to reduced estimator efficiency. A biased parameter estimator is a popular approach that provides various remedies for the problem, but the performance of this estimator will depend on the selection of the biasing parameter. This paper explores and compares different biasing parameter selection strategies, including one-and two-parameter biased estimators, emphasizing fixed, data-driven, cross-validation, and machine learning-based solutions in different sample sizes and predictor degree relationships. The study used a Monte Carlo simulation that generated data at the conditions of strong multicollinearity (ρ_x ranging from 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, and 0.99), varying error variance (σ = 3, 5, 10, 15, and 20), and sample sizes (n = 100, 500, 1000, 5000, and 10,000). The performance of the estimators was analyzed with MSE, which allows the systematic comparison of different methods for selection of biasing parameters under different types of multicollinearity and sample sizes. The results consistently indicate that the machine learning–based selection provides the lowest MSE in all cases and is increasingly advantageous as increasing multicollinearity is encountered and sample size reduces. Data-driven methods provide competitive outcomes and clearly improve over fixed parameter and cross-validation. Cross-validating methods experience inefficiency under more complex multicollinearity scenarios. To validate the simulation studies, a real-world data setting was applied using the Ames housing dataset. The machine learning approach also produced the best predictive performance, recording the lowest MSE, substantially outperforming the fixed, data-driven, and cross-validation methods in the real-data application. Thus, in the presence of severe multicollinearity, practitioners and researchers should go beyond applying fixed or only cross-validation–based rules but use machine learning or data-driven guided approaches.

Keywords: *Cross-Validation; Data-Driven Methods; Machine Learning; Multicollinearity.*

1. Introduction

Multicollinearity remains one of the problems of multiple regression analyses in various fields, including econometrics, biostatistics, and social sciences. For severe multicollinearity, the Ordinary Least Squares (OLS) estimator, while unbiased, is unstable, with the variance and covariances magnified, the standard errors increased, and the parameter estimates less predictive (poor precision and interpretability). This complicates both hypothesis testing and inferential interpretations when the sample size is small and in moderate conditions often encountered in empirical research (Lukman et al., 2025; Akhtar and Alharthi, 2025). To overcome these drawbacks, investigators have devised biased estimation methods, which are performed through deliberate shrinkage in the estimation process in return for a low overall estimation error. Among these methods, Liu-type estimators and ridge regression are considered. Ridge regression is a biasing parameter (BP) introduced to punish a large coefficient value. A slight bias is balanced to achieve a lower total MSE than OLS under multicollinearity conditions (Luqman et al., 2025). Nevertheless, in spite of this long-standing literature, the ridge estimator's performance is highly sensitive to the choice of BP for k (Akhtar and Alharthi, 2025). The wrong choice of parameter can cause a decline in the estimator's performance, undercorrection of multicollinearity, and an overfitting estimate, which hides true relationships (Luqman et al., 2025).
Biasing parameter selection is not limited to one-parameter shrinkage estimators. Two-parameter extensions like modified Liu and Kibria–Lukman estimators increase degrees of freedom that introduce theoretical performance advantages and yet increase parameter optimization difficulty (El-doakly, 2024). New-stage literature simulations show that two-parameter estimators can be statistically superior to other models for the reduction of MSE, despite their dependence on parameter selection and the data architecture (El-doakly, 2024; Dawoud, Abonazel, and Awwad, 2022). Existing studies like Kibria and Lukman (2020) admit that multicollinearity severity and sample size are significant influencers on estimator performance, but few works systematically examine how different methodologies of BP selection perform in varying sample sizes and under varying levels of multicollinearity, and in particular for extreme correlations, like near-singular predictor matrices, in real-world studies. A few recent advances aim for BP to be more sensitive to data properties (e.g., adjusting ridge penalty terms according to eigen structure or condition indices of the design matrix), which seem to offer promising enhancements in prediction accuracy and the stabilization of prediction outcomes (Akhtar and Alharthi, 2025).

An increasing number of simulation-based works also confirm the necessity for a comprehensive assessment of BP. Research on large ridge parameter sets has shown that moderate parameter selection should not be too far away from resolving moderate multicollinearity (as in the case of strongly correlated predictors, large sample sizes, or condition-adjusted or data-driven parameters that lower MSE in different conditions, like changing the characteristics of dimensionality, noise structure, and size of the predictor sample) (Kibria and Lukman, 2020). Nevertheless, these data show that the selection of a BP is crucial and not trivial in the use of biased estimators. However, despite such methodological innovations, the literature rarely systematically examines how the choice mechanism works when sample size and multicollinearity are heterogeneous. The importance of the knowledge gap is that sample size and predictor correlations are common features of real-world data, and their interaction with BP selection can have a large impact on estimator performance (Akhtar and Alharthi, 2025; Luqman et al., 2025).

Specifically, we systematically measure the performance of BP selection methods for one-parameter biased regression estimators across a range of sample sizes and multicollinearities. Instead of merely assessing alternative selection methods, the current work systematically examines whether or not BP selection is the most effective in small- and large-sample settings and for increasing sample size, which provides efficient performance. More precisely, classical fixed (analytical) BP rules are compared with data-driven, cross-validation- and machine-learning-inspired selection procedures, with equal emphasis on interpretability of the model. Based on an extensive Monte Carlo model, estimator performance is primarily tested on the MSE criterion, allowing the controlled and systematic assessment of the response of each selection approach to varying sample sizes (those with small and large sample sizes) and the strength of correlation between predictor variables.

2. Literature Review

In the shrinkage estimation literature, there has been a long-standing debate on how to determine the BP, whether it should be obtained by theoretical optimization or estimated by adaptive data-driven procedures. Theoretical approaches are often based on analytical derivations aiming at the minimization of the estimator risk under some assumptions on the data-generating process. These methods are mathematically consistent and interpretable, but may perform poorly when underlying assumptions are violated, or the degree of multicollinearity is substantially different from the theoretical conditions under which they were derived (Hoerl and Kennard, 1970; Liu, 1993). Adaptive methods for selecting BP use sample information to decide the shrinkage intensity and are generally more sensitive to the correlation structure among the explanatory variables that is observed. However, recent empirical evidence shows that adaptive approaches tend to perform better in finite samples, at least in terms of Mean Squared Error (MSE), by adapting to the specific characteristic of the observed data (Akhtar and Alharthi, 2025; Luqman et al., 2025). However, these methods have been criticized for being sensitive to the particular sample conditions, which may threaten their efficiency and generalizability to different multicollinearity scenarios. As a result, neither theoretical nor adaptive approaches have completely resolved the challenge of selecting optimal BP. This unresolved issue continues to motivate methodological developments in shrinkage estimation.

In this regard, the recent developments of the Kibria–Lukman family of estimators have been mainly directed towards improving the estimation efficiency by changing the mechanism of the shrinkage and the corresponding biasing parameter expressions. Studies like Dawoud et al. (2022), Ugwuowo et al. (2023), and Akay et al. (2023) reported that the MSE has been significantly decreased in comparison to the classical ridge and Liu estimators under several simulation conditions. The results show the potential benefits of the Kibria–Lukman approach in dealing with multicollinearity. However, a common concern in the literature is that the performance improvements are often attributed to the new biasing parameter specifications, rather than to improvements in the estimators' fundamental structure. Therefore, it is difficult to determine if the efficiency improvements observed are due to the estimator architecture or the particular strategy used to select its parameters. Furthermore, existing studies have largely concentrated on proposing new estimator variants and comparing their performance using simulation experiments, often without providing a comprehensive theoretical explanation for the conditions under which specific BPs are expected to perform optimally. This has resulted in a growing body of estimator modifications but limited understanding of the mechanisms linking BP choice to estimator performance. Consequently, the literature lacks a coherent framework capable of explaining how factors such as multicollinearity severity, sample size, error variance, and data structure influence the effectiveness of different parameter selection strategies.

Another important limitation concerns the common assumption that multicollinearity is a homogeneous phenomenon. Ayinde et al. (2021) challenged this perspective by demonstrating that multicollinearity may vary considerably across subsets of explanatory variables and that traditional shrinkage methods often fail to account for such heterogeneity. This observation suggests that a single rule for selecting BP may not be equally effective across all correlation structures, thereby raising questions about the robustness and transferability of existing approaches. While many studies have attempted to identify estimators that attain the minimum MSE for particular simulation scenarios, relatively little attention has been paid to the reasons why certain BP selection methods work under some conditions and fail under others. Lack of such knowledge limits the generality of empirical findings and the development of shrinkage estimation procedures that are universally applicable. Therefore, a systematic study of the BP determination that combines theoretical aspects and empirical performance in different multicollinearity scenarios is needed. In this sense, the current study adds a contribution to the literature by departing from the traditional approach of proposing another variant of the estimator. Instead, the emphasis is on the basic problem of BP determination to arrive at a more robust, theoretically principled, and empirically validated approach to parameter selection. The study attempts to clarify the circumstances under which different shrinkage strategies are most effective by analyzing the interaction between the choice of BP, the degree of multicollinearity, and the estimator performance.

3. Methodology

Consider the standard linear regression model.

$$y = X\beta + \varepsilon \tag{3.1}$$

Where $y \in \Re^{n\times 1}$ is the response vector, $X \in \Re^{n\times p}$ is the design matrix of full column rank, $\beta \in \Re^{p\times 1}$ is the vector of unknown regression coefficients, and $\varepsilon \sim N(0, \sigma^2 I_n)$ is the stochastic error term. The OLS estimator from equation (3.1) is given by:

$$\hat{\beta}_{OLS} = (X'X)^{-1}X'y \tag{3.2}$$

When predictors are highly correlated, $X'X$ becomes ill-conditioned or nearly singular

3.1. Ridge regression estimator

The concept of ridge regression was first proposed by Hoerl and Kennard (1970) to handle the multicollinearity problem. A positive value of the ridge parameter k introduces bias into the estimator but often results in a lower mean squared error (MSE) than the ordinary least squares estimator, particularly in the presence of multicollinearity. Thus, the resulting estimator is obtained as:

$$\hat{\beta}_R(k) = (X'X + kI_p))^{-1}X'y = W\hat{\beta} \qquad k > 0 \tag{3.3}$$

Where $W = [Ip + kC^{-1}]^{-1}$, $k \geq 0$ is the biasing (ridge) parameter, $C = X'X$, and I_p is an identity matrix of order p. This is known as the ridge regression estimator. Since $[X'X + kI_p]$ is always invertible for $k > 0$, there exists a unique solution for $\hat{\beta}_R(k)$. The ridge regression estimator is biased; however, it generally has a smaller mean squared error (MSE) than the OLS estimator when multicollinearity is present. From (3.3), it can be observed that as k→0, the ridge estimator converges to the OLS estimator $(\hat{\beta}_R(k) \to \hat{\beta}_{OLS})$. Conversely, as $k \to \infty$, the ridge estimator shrinks towards the zero vector $(\hat{\beta}_R(k) \to 0)$. Several studies at different periods of time worked in this area of research and developed and proposed different estimators such as Hoerl and Kennard (1970) and the modified ridge regression estimator (MRR), the Liu estimator (1993), and Kibria-Lukman (2020).

3.1.1. Statistical properties of the ridge estimator

1) Bias of Ridge Estimator

$$E[\hat{\beta}_R(k)] = (X'X + kI)^{-1}X'X\beta = B(k)\beta \tag{3.4}$$

Where

$$B(k) = (X'X + kI)^{-1}X'X$$

Thus, the bias is:

$$\text{Bias}[\hat{\beta}_R(k)] = (B(k) - I)\beta \tag{3.5}$$

2) Variance of Ridge Estimator

$$\text{Var}[\hat{\beta}_R(k)] = \sigma^2(X'X + kI)^{-1}X'X(X'X + kI)^{-1} \tag{3.6}$$

3) Mean Squared Error
The scalar MSE criterion is:

$$\text{MSE}(\hat{\beta}) = E[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)] \tag{3.7}$$

For ridge regression:

$$\text{MSE}(\hat{\beta}_R) = \underbrace{\| \text{Bias}(\hat{\beta}_R) \|^2}_{\text{Squared Bias}} + \underbrace{\text{tr}[\text{Var}(\hat{\beta}_R)]}_{\text{Variance}} \tag{3.8}$$

3.2. Theoretical Properties of Ridge-type Biasing Parameter Strategies

3.2.1. Bias–Variance Trade-off Framework

The superiority of ridge-type estimators arises from the fundamental bias–variance trade-off. While the Ordinary Least Squares (OLS) estimator is unbiased, $E(\hat{\beta}_{OLS}) = \beta$, its variance becomes excessively large under multicollinearity because the matrix $X'X$ approaches singularity. The covariance matrix of the OLS estimator is given as;

$$\text{Var}(\hat{\beta}_{OLS}) = \sigma^2(X'X)^{-1} \tag{3.9}$$

When the minimum eigenvalue of $X'X$ tends toward zero,

$$\lambda_{min}(X'X) \to 0 \tag{3.10}$$

The variance inflates substantially, i.e. $\text{Var}(\hat{\beta}_{OLS}) \to \infty$. Ridge regression stabilizes estimation by introducing a shrinkage parameter. k

$$\hat{\beta}_R(k) = (X'X + kI)^{-1}X'y \tag{3.11}$$

Which produces a biased but lower-variance estimator. The Mean Squared Error (MSE) for ridge regression decomposition is:

$$\text{MSE}(\hat{\beta}_R) = \| (B(k) - I)\beta \|^2 + \sigma^2\text{tr}[(X'X + kI)^{-1}X'X(X'X + kI)^{-1}] \tag{3.12}$$

Thus, although ridge estimation introduces bias, the reduction in variance can dominate the increase in bias, leading to lower overall MSE.

3.2.2. Spectral Representation and Stability

Using spectral decomposition

$$X'X = Q\Lambda Q' \tag{3.13}$$

Where:

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p) \tag{3.14}$$

The ridge estimator becomes:

$$\hat{\beta}_R = Q(\Lambda + kI)^{-1}\Lambda Q'\hat{\beta}_{OLS} \tag{3.15}$$

The variance of the j-th component is:

$$\text{Var}(\hat{\beta}_{R,j}) = \sigma^2 \frac{\lambda_j}{(\lambda_j+k)^2} \tag{3.16}$$

Under severe multicollinearity $\lambda_j \approx 0$, the OLS variance becomes unstable, whereas ridge estimation remains bounded because $\lambda_j + k > 0$ for all $k > 0$. Hence, ridge-type estimators are numerically stable even when the design matrix is nearly singular.

3.2.3. Theoretical Motivation for Data-driven Biasing Parameters

The optimal ridge parameter minimizing component-wise MSE is:

$$k_j^* = \frac{\sigma^2}{\beta_j^2} \tag{3.17}$$

Which depends on unknown population quantities. Fixed ridge strategies assume $k = c$ for some constant c, irrespective of sample size, predictor dimension, correlation structure, or noise level. This creates theoretical inefficiency because the optimal shrinkage level varies across data-generating processes. Data-driven ridge strategies instead estimate $k = f(n, p, \hat{\sigma}^2, \lambda_j)$ allowing shrinkage intensity to adapt to the observed structure of the data.

3.2.4. Asymptotic behavior

Assume:

$$\frac{1}{n}X'X \to \Sigma \tag{3.18}$$

Where Σ is positive definite. For Ridge regression;

$$\hat{\beta}_R(k) = (X'X + kI)^{-1}X'y \tag{3.19}$$

If $k_n \to 0$ and $\frac{k_n}{n} \to 0$, then $\hat{\beta}_R(k_n) \xrightarrow{p} \beta$. Therefore, ridge estimators are asymptotically consistent provided the shrinkage parameter decreases appropriately with sample size.

For adaptive/data-driven estimators, $\hat{k}_n \xrightarrow{p} k^*$ implies $\hat{\beta}_R(\hat{k}_n) \xrightarrow{p} \beta$ under standard regularity conditions. Thus, adaptive ridge estimators preserve consistency while improving finite-sample efficiency.

3.2.5. Asymptotic mean squared error (AMSE)

The asymptotic MSE of ridge estimation is:

$$\text{AMSE}(\hat{\beta}_R) = \lim_{n\to\infty} E\left[(\hat{\beta}_R - \beta)'(\hat{\beta}_R - \beta)\right] \tag{3.20}$$

Using eigenvalue decomposition:

$$\text{AMSE}(\hat{\beta}_R) = \sum_{j=1}^{p}\left[\frac{\sigma^2\lambda_j}{(\lambda_j+k)^2} + \frac{k^2\beta_j^2}{(\lambda_j+k)^2}\right] \tag{3.21}$$

Where p is the number of explanatory variables, σ^2 is the error variance, λ_j is the j[th] eigenvalue of the matrix $X'X$, β_j is the j[th] transformed regression coefficient corresponding to the eigenvector associated with λ_j, k ridge parameter, and $\hat{\beta}_R$.

3.2.6. Robustness under multicollinearity

Under near-singular correlation structures, $\det(X'X) \approx 0$. OLS estimators become unstable and highly sensitive to small perturbations in the data. Ridge-type estimators improve robustness because $X'X + kI$ is always positive definite for $k > 0$.Therefore:

$$\kappa(X'X + kI) < \kappa(X'X) \tag{3.22}$$

Where $\kappa(\cdot)$denotes the condition number. This reduction in condition number improves numerical stability, coefficient robustness, and predictive reliability.

3.3. Fixed ridge parameter selection

Fixed ridge methods specify k as a constant independent of the data realization. Common choices include:

$$k = c, c > 0$$

Or heuristic rules such as:

$$k = \frac{\sigma^2}{\hat{\beta}'_{OLS}\hat{\beta}_{OLS}} \tag{3.23}$$

The Kibria Lukman Estimator
A one-parameter estimator was proposed by minimizing the following objective function:

$$(y - X\beta)'(y - X\beta) + k\left[(\beta + \hat{\beta})'(\beta + \hat{\beta}) - c\right] \tag{3.24}$$

With respect to β, will yield the normal equations

$$(X'X + KI_P)\beta = X'y - k\hat{\beta} \tag{3.25}$$

Where k is the nonnegative constant. The solution to the above gives the new estimator as:

$$\hat{\beta}_{KL} = (S + KI_P)^{-1}(S - KI_P)\hat{\beta} = W(K)M(K)\hat{\beta} \tag{3.26}$$

Where $S = X'X, W(K) = [I_P + kS^{-1}]^{-1}$ and $M(K) = [I_P - kS^{-1}]$. The proposed estimator is called the Kibria-Lukman (KL) estimator and denoted by $\hat{\beta}_{KL}$.
Where k > 0 is the biasing parameter (Kibria and Lukman 2020).

3.4. Data-driven BP selection

Data-driven methods explicitly express k as a function of observable sample characteristics:

$$k = f(n, p, \hat{\sigma}^2, \lambda_1, \ldots, \lambda_p) \tag{3.27}$$

Where λ_j are eigenvalues of $X'X$.
Spectral Decomposition
Let,

$$X'X = Q\Lambda Q' \tag{3.28}$$

Then the ridge estimator becomes;

$$\hat{\beta}_R = Q(\Lambda + kI)^{-1}\Lambda Q'\hat{\beta}_{OLS} \tag{3.29}$$

Optimal k(MSE Minimization)
For the j-th component

$$MSE_j(k) = \frac{\sigma^2\lambda_j}{(\lambda_j+k)^2} + \frac{k^2\beta_j^2}{(\lambda_j+k)^2} \tag{3.30}$$

Minimizing with respect to k:

$$\frac{\partial MSE_j(k)}{\partial k} = 0 \Rightarrow k_j^* = \frac{\sigma^2}{\beta_j^2} \tag{3.31}$$

Since β_j is unknown, data-driven estimators replace it with sample-based proxies, yielding adaptive ridge parameters such as the Covariance Adjusted Ridge Estimator (CARE) and Stein-type Parameter Shrinkage (SPS) estimators.

3.5. Cross-Validation-Based Ridge Parameter Selection

Cross-validation (CV) selects k by minimizing out-of-sample prediction error.
K-Fold CV Criterion

$$CV(k) = \frac{1}{K}\sum_{i=1}^{K} \| y_{(i)} - X_{(i)}\hat{\beta}_R^{(-i)}(k) \|^2 \tag{3.32}$$

The optimal parameter is:

$$k_{CV} = \arg\min_k CV(k) \tag{3.33}$$

Under regularity conditions:

$$CV(k) \xrightarrow{p} E[\text{Prediction Error}(k)] \tag{3.34}$$

Thus, CV provides an asymptotically unbiased estimator of predictive risk.

3.6. Machine Learning Approach

3.6.1. Conceptual framework

In this study, machine learning (ML) is employed as an automated hyperparameter optimization framework for selecting the optimal ridge biasing parameter. Specifically, the study treats the ridge parameter selection problem as a supervised learning and regularized optimization task, where the objective is to minimize out-of-sample prediction error through algorithmic learning. Unlike fixed or analytically derived ridge parameters, the machine learning approach does not prespecify the shrinkage parameter. k. Instead, the parameter is learned adaptively from the data using resampling-based optimization procedures. Thus, the machine learning framework in this study belongs to the class of supervised regularized learning with:

- predictors/features: explanatory variables X,
- target/output: response variable y,
- learning algorithm: ridge regression with automated tuning,
- optimization criterion: prediction risk minimization.

3.6.2. Learning Algorithm

The machine learning model adopted is L2-regularized linear regression (ridge regression) implemented using the glmnet algorithm developed by Friedman et al. (2010). The ridge optimization problem is defined as:

$$\hat{\beta}_\lambda^{ML} = \arg\min_\beta \left[\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\beta)^2 + \lambda \| \beta \|_2^2\right] \tag{3.35}$$

Where:
λ is the machine learning regularization hyperparameter,
$\| \beta \|_2^2 = \sum_{j=1}^{p} \beta_j^2$ is the L2 penalty term.
The closed-form ridge solution becomes:

$$\hat{\beta}_\lambda^{ML} = (X'X + n\lambda I_p)^{-1}X'y \tag{3.36}$$

With the equivalence $k = n\lambda$ linking the ML regularization parameter λ to the classical ridge parameter k.

3.6.3. Feature engineering and input structure

The machine learning model uses the explanatory variables: $X = [X_1, X_2, \dots, X_p]$ as predictive features, while the dependent variable y serves as the prediction target. To ensure numerical stability and prevent scale dominance among predictors, all features are standardized before model fitting:

$$X_{ij}^* = \frac{X_{ij} - \bar{X}_j}{s_j} \tag{3.37}$$

Where:
$\bar{X}_j$ is the mean of the j-th predictor,
s_j is the standard deviation.
Standardization is necessary because ridge penalties are scale-sensitive.

3.6.4. Hyperparameter tuning procedure

The ridge penalty parameter was selected using K-fold cross-validation (CV) within the machine learning framework.
The dataset was partitioned into $K = 10$ folds:

$$\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_{10}\} \tag{3.38}$$

For each candidate value of λ:
1) The model was trained on $K+1$ folds,
2) Validation was performed on the remaining fold,
3) Prediction error was computed,
4) The process was repeated across all folds.
The cross-validation criterion is:

$$CV(\lambda) = \frac{1}{K}\sum_{k=1}^{K} \frac{1}{n_k} \sum_{i\in\mathcal{D}_k} (y_i - \hat{y}_i^{(-k)})^2 \tag{3.39}$$

The optimal hyperparameter is obtained as:

$$\lambda^* = \arg\min_{\lambda} CV(\lambda) \tag{3.40}$$

The corresponding machine learning ridge estimator is then:

$$\hat{\beta}_{ML} = \hat{\beta}_{\lambda^*} \tag{3.41}$$

3.6.5. Prevention of overfitting

Several mechanisms were implemented to reduce overfitting and improve generalization performance:
1) L2 Regularization: The ridge penalty shrinks coefficient magnitudes toward zero:

$$\lambda \sum_{j=1}^{p} \beta_j^2 \tag{3.42}$$

Thereby reducing estimator variance and preventing unstable coefficient inflation under multicollinearity.
2) Cross-Validation: The use of 10-fold cross-validation ensures that model performance is evaluated on unseen data, reducing optimism bias.
3) Standardization: Feature scaling prevents predictors with large magnitudes from dominating the optimization process.
4) Automated Hyperparameter Optimization: Instead of arbitrarily selecting k, the learning algorithm searches across a grid of candidate penalties $\lambda \in \{10^{-4}, 10^{-3}, \dots, 10^{2}\}$to identify the penalty producing minimum prediction error.

3.6.6. Computational workflow

The machine learning implementation followed the workflow below:
Step 1: Generate multicollinear predictors

$$X_{ij} = \sqrt{1-\rho^2} z_{ij} + \rho z_{ip} \tag{3.43}$$

Step 2: Standardize predictors

$$X \rightarrow X^* \tag{3.44}$$

Step 3: Partition data into K folds ($K = 10$).
Step 4: Train ridge model across candidate λ values using the glmnet optimization algorithm.
Step 5: Compute validation error using mean squared prediction error.

3.7. Monte Carlo simulation design

Following Kibria-Lukman (2020) in generating X's are generated using the following equation:

$$X_{ij} = (1-\rho^2)^{\frac{1}{2}} z_{ij} + \rho z_{ip}\ ,\ \ i = 1,2,\dots,n,\ \ j = 1,2,\dots,p \tag{3.45}$$

Where $X \sim N(0,\Sigma), \Sigma_{ij} = \rho^{|i-j|}$, z_{ij} are independent standard normal pseudo-random numbers and ρ represents the correlation between any two Xs. These variables are standardized so that X'X and X'y are in correlation form. The response variable for the k-th equation is generated as:

$$y_{ik} = \beta_{0k} + \beta_{1k}X_{i1} + \beta_{2k}X_{i2} + \cdots + \beta_{pk}X_{ip} + \varepsilon_{ik},\ \ i = 1,2,\dots,n,\ k = 1,2,\dots,m \tag{3.46}$$

The disturbance vector $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{im})'$is assumed to follow a multivariate normal distribution, $\varepsilon_i \sim N(0,\Omega)$ with:
$\rho \in \{0.7, 0.8, 0.9, 0.95\}$
$\beta_k = (1,1,\dots,1)',\ \ \forall\, k$
$n \in \{100, 500, 1000, 5000, 10000\}$
$\sigma \in \{3,5,10,15,20\}$

3.8. Performance Evaluation Criterion

The primary evaluation metric is given as;

$$\widehat{\mathrm{MSE}} = \frac{1}{\mathrm{R}} \sum_{\mathrm{r}=1}^{\mathrm{R}} (\hat{\beta}^{(\mathrm{r})} - \beta)'(\hat{\beta}^{(\mathrm{r})} - \beta) \quad (3.47)$$

Where R is the number of replications.

3.9. Variable measurement

Measurements are taken from property assessment records in Ames, Iowa (original dataset compiled by Dean De Cock, 2011).

Variable	Measurement	Notes
Sale Price	Continuous (USD, numeric)	Final sale price of the house, recorded at transaction.
Gr Liv Area	Continuous (square feet)	Above-ground living area (excluding basement), measured from property records.
Garage Area	Continuous (square feet)	Size of the garage, measured from property records.
Total Bsmt SF	Continuous (square feet)	Total basement area, measured from property records.
Full Bath	Discrete (count)	Number of full bathrooms above ground, recorded in property assessment.

Source: Dean De Cock, 2011.

4. Results

The Mean Squared Error (MSE) values for different levels of multicollinearity (ρ= 0.70, 0.80, 0.90, 0.95, 0.99), sample sizes and error variances are shown in Tables 1–7. When n = 100, ρ = 0.70 and σ = 3, the MSE values of fixed, data-driven, cross-validation, and machine learning methods are 0.73699, 0.49797, 0.75482, and 0.45425, respectively. For n = 1000, ρ = 0.85 and σ = 10, the MSE values of fixed, data-driven, cross-validation, and machine learning methods are 1.3123, 0.6640, 1.4279, and 0.5311, respectively. The machine learning-based BP selection method produced the lowest MSE values for all simulation scenarios (see Tables 1 – 7). The advantage of this approach was especially clear in small-sample settings and under conditions of high error variance, where differences among competing methods were most pronounced. The data-driven approach was second best in terms of performance, giving lower MSE values than the fixed and cross-validation approaches for most of the experimental conditions. Under moderate sample sizes and larger error variance, it was more advantageous.

Comparison of the fixed and cross-validation approaches did not reveal consistent superiority of one approach over the other. The fixed approach was sometimes able to obtain lower values of MSE, but both methods showed larger estimation errors with respect to the machine learning and data-driven procedures at all times. All the parameter selection methods exhibited large decreases in MSE with increasing sample size. The differences between competing approaches became smaller and smaller. For large sample sizes, the MSE values of all methods converged, suggesting asymptotic equivalence in estimator performance. Nevertheless, the machine learning-based approach maintained the smallest MSE values throughout, even under severe multicollinearity.

Simulation results also indicated some threshold effects in terms of sample size and error variance. Under small sample and high variance conditions, the machine learning-based method exhibited significant reductions in MSE compared to all competing approaches. The machine learning method was still the most efficient method for moderate sample sizes ($1{,}000 \leq n \leq 5{,}000$), and the data-driven approach was always second. For large sample sizes ($n \geq 5{,}000$), performance differences between the methods were negligible. Nevertheless, the machine learning approach yielded the lowest MSE values for all levels of multicollinearity considered. The results highlight the advantages of adaptive BP selection methods, especially the machine learning-based procedure, in terms of finite-sample performance, as compared to traditional fixed and cross-validation procedures.

The results from the empirical analysis were presented in Tables 8 – 10. Table 8 gives the descriptive statistics of the research variables. The mean of the sale price of houses is $180,796.06 (standard deviation = $79,886.69). This shows that there is high variability in the value of the properties within the data set. Also, the predictor variables and their interaction terms have high variability. Table 9 gives the variance inflation factor (VIF) results, which show the presence of serious multicollinearity in the explanatory variables. The VIFs of some predictors are greater than the threshold of 10. Therefore, the predictors are highly collinear, and the use of shrinkage estimation techniques can be applied. Table 10 shows the comparison of the predictive power of the different selection methods of the biasing parameter. The machine learning (ML) approach performed better in terms of the MSE (1,728,378), RMSE (41.57), and MAE (27.05). It therefore outperformed the Data-Driven Ridge, Fixed Ridge, and CV Ridge approaches in the selection of the biasing parameters.

Table 1: MSE Comparison with Different Biasing Parameter Selection when $\rho = 0.7$

σ	n	Fixed	Data Driven	Cross Validation	Machine Learning
3	100	0.73699	0.49797	0.75482	0.45425
	500	0.14528	0.13641	0.14681	0.11758
	1000	0.07201	0.06995	0.07240	0.05983
	5000	0.01413	0.01406	0.01414	0.01364
	10000	0.00729	0.00727	0.00729	0.00787
5	100	1.87147	1.33914	1.95695	0.93267
	500	0.37986	0.30275	0.39032	0.26335
	1000	0.20532	0.18669	0.20860	0.16099
	5000	0.03981	0.03919	0.03995	0.03463
	10000	0.02027	0.02011	0.02030	0.01824
10	100	6.89508	2.72769	6.66616	2.56007
	500	1.48058	1.05595	1.60461	0.75525
	1000	0.75722	0.50635	0.79813	0.45126
	5000	0.15912	0.14847	0.16129	0.12912
	10000	0.08026	0.07775	0.08083	0.07060
15	100	14.02793	3.56822	11.12742	4.39969
	500	3.19556	1.86931	3.58518	1.39989
	1000	1.66411	1.18215	1.82647	0.83252
	5000	0.36862	0.29923	0.37942	0.25984
	10000	0.18523	0.17051	0.18812	0.14695
20	100	25.48157	4.84205	16.48837	7.25949
	500	5.50800	2.57455	6.20198	2.11584

	1000	2.76245	1.79943	3.13385	1.20767
	5000	0.64478	0.45986	0.67610	0.40490
	10000	0.31231	0.26673	0.32060	0.22574

Table 2: MSE Comparison with Different Biasing Parameter Selection when $\rho = 0.75$

σ	n	Fixed	Data Driven	Cross Validation	Machine Learning
	100	0.82723	0.52030	0.84596	0.46226
	500	0.17234	0.15931	0.17442	0.13078
3	1000	0.08329	0.08047	0.08381	0.06570
	5000	0.01660	0.01649	0.01662	0.01461
	10000	0.00838	0.00836	0.00839	0.00832
	100	2.17942	1.39628	2.27893	0.94415
	500	0.44723	0.34002	0.46140	0.28137
5	1000	0.22883	0.20355	0.23301	0.16588
	5000	0.04562	0.04480	0.04581	0.03816
	10000	0.02353	0.02332	0.02358	0.02025
	100	7.46410	2.62048	7.19442	2.45617
	500	1.67656	1.13207	1.83162	0.76244
10	1000	0.87622	0.56250	0.93000	0.46961
	5000	0.18558	0.17021	0.18854	0.14085
	10000	0.09279	0.08927	0.09357	0.07760
	100	16.45899	3.68037	12.82735	5.02977
	500	3.48668	1.96081	3.94776	1.35084
15	1000	1.86854	1.18525	2.07138	0.84702
	5000	0.39804	0.31414	0.41155	0.25537
	10000	0.20297	0.18372	0.20664	0.14987
	100	27.77489	4.48541	16.54909	6.93874
	500	5.91414	2.47782	6.72424	2.02190
20	1000	3.13847	1.83888	3.59698	1.23461
	5000	0.70652	0.47725	0.74626	0.39785
	10000	0.37471	0.30246	0.38631	0.24756

Table 3: MSE Comparison with Different Biasing Parameter Selection when $\rho = 0.8$

σ	n	Fixed	Data Driven	Cross Validation	Machine Learning
	100	0.95390	0.54717	0.97642	0.46798
	500	0.20381	0.18437	0.20666	0.14133
3	1000	0.10074	0.09644	0.10150	0.07328
	5000	0.02124	0.02107	0.02127	0.01689
	10000	0.00988	0.00984	0.00988	0.00880
	100	2.64279	1.36069	2.76475	1.02741
	500	0.55817	0.36562	0.57895	0.30996
5	1000	0.27389	0.23393	0.27981	0.17913
	5000	0.05526	0.05402	0.05554	0.04313
	10000	0.02812	0.02781	0.02819	0.02253
	100	9.74288	2.70216	9.29262	3.05198
	500	2.02411	1.15142	2.23443	0.82693
10	1000	1.04163	0.63732	1.11592	0.49291
	5000	0.22436	0.20005	0.22871	0.15270
	10000	0.11145	0.10619	0.11258	0.08627
	100	19.70122	3.75653	14.85561	5.39329
	500	4.15641	1.97096	4.75437	1.42568
15	1000	2.20076	1.21832	2.47146	0.88078
	5000	0.48214	0.34527	0.50178	0.27439
	10000	0.25181	0.22054	0.25735	0.16829
	100	35.13954	4.97992	21.35565	8.69638
	500	7.02809	2.49072	8.02278	2.14293
20	1000	3.83192	1.75842	4.44213	1.34639
	5000	0.85816	0.56150	0.91498	0.42990
	10000	0.44191	0.32850	0.45840	0.25772

Table 4: MSE Comparison with Different Biasing Parameter Selection when $\rho = 0.85$

σ	n	Fixed	Data Driven	Cross Validation	Machine Learning
	100	1.32209	0.65187	1.35346	0.54562
	500	0.26333	0.22688	0.26787	0.15713
3	1000	0.13675	0.12866	0.13805	0.08795
	5000	0.02780	0.02751	0.02785	0.01916
	10000	0.01314	0.01307	0.01315	0.01002
	100	3.35597	1.51734	3.50178	1.14944
	500	0.68821	0.43670	0.71923	0.33239
5	1000	0.35724	0.28400	0.36704	0.20252
	5000	0.07388	0.07165	0.07436	0.05186
	10000	0.03667	0.03614	0.03679	0.02628
	100	11.61720	2.49871	10.72305	3.22450
	500	2.67811	1.26837	2.99375	0.96109
10	1000	1.31232	0.66396	1.42788	0.53110
	5000	0.28619	0.24322	0.29331	0.16876
	10000	0.14721	0.13768	0.14914	0.10086
15	100	24.59691	3.76338	17.64973	5.87364
	500	5.40581	1.95703	6.23490	1.65389

	1000	2.84504	1.34609	3.25088	0.99104
	5000	0.62126	0.38861	0.65329	0.30069
	10000	0.31507	0.25882	0.32395	0.18154
	100	44.25766	5.46036	25.05371	9.71941
	500	8.93457	2.38001	10.20000	2.45327
20	1000	4.86181	1.83527	5.72468	1.43958
	5000	1.07212	0.55861	1.15997	0.45302
	10000	0.53375	0.34489	0.55908	0.26712

Table 5: MSE Comparison with Different Biasing Parameter Selection when $\rho = 0.9$

σ	n	Fixed	Data Driven	Cross Validation	Machine Learning
	100	1.74117	0.76348	1.78046	0.57454
	500	0.38069	0.29580	0.38954	0.17829
3	1000	0.18189	0.16539	0.18432	0.09406
	5000	0.03901	0.03841	0.03912	0.02143
	10000	0.01917	0.01902	0.01920	0.01135
	100	4.41716	1.45531	4.56247	1.24584
	500	1.03187	0.53522	1.09468	0.39815
5	1000	0.52488	0.34274	0.54480	0.23611
	5000	0.10462	0.09982	0.10559	0.06054
	10000	0.05313	0.05197	0.05338	0.03117
	100	15.86685	2.74670	13.91507	3.96803
	500	3.83897	1.40090	4.37421	1.19341
10	1000	1.86851	0.80561	2.08147	0.61509
	5000	0.41635	0.31661	0.43103	0.19727
	10000	0.20483	0.18392	0.20870	0.11492
	100	36.21065	4.26781	24.65973	8.21752
	500	7.68504	1.96915	8.96646	2.08748
15	1000	4.14295	1.37674	4.84378	1.22876
	5000	0.90084	0.45596	0.96451	0.35030
	10000	0.45058	0.31098	0.46854	0.20820
	100	66.06941	7.33185	37.22600	15.21499
	500	12.21496	2.44599	13.92003	2.85920
20	1000	6.70562	1.73304	8.05543	1.77838
	5000	1.56640	0.67944	1.73727	0.54695
	10000	0.77954	0.40292	0.83095	0.30767

Table 6: MSE Comparison with Different Biasing Parameter Selection when $\rho = 0.95$

σ	n	Fixed	Data Driven	Cross Validation	Machine Learning
	100	3.42489	0.96656	3.46598	0.74899
	500	0.66310	0.34577	0.69006	0.18744
3	1000	0.37095	0.28834	0.37968	0.11187
	5000	0.07443	0.07209	0.07482	0.02330
	10000	0.03667	0.03612	0.03677	0.01222
	100	8.36333	1.70522	8.38589	1.69512
	500	1.86628	0.62400	2.04989	0.47798
5	1000	0.95519	0.37763	1.01826	0.27347
	5000	0.19975	0.17959	0.20324	0.07099
	10000	0.10400	0.09934	0.10493	0.03706
	100	31.70857	3.62770	25.90979	6.75646
	500	6.27676	1.40628	7.42967	1.48348
10	1000	3.47578	1.04664	4.05640	0.86380
	5000	0.79049	0.37688	0.84023	0.25631
	10000	0.40712	0.30295	0.42151	0.15110
	100	70.19759	6.71097	45.34422	15.41402
	500	13.85950	2.11078	16.23335	3.07274
15	1000	7.27087	1.57391	8.85526	1.67808
	5000	1.64668	0.54642	1.84187	0.45255
	10000	0.85631	0.40424	0.91755	0.27092
	100	117.87660	9.80820	61.43129	24.45048
	500	23.27397	2.76324	25.73769	4.73580
20	1000	12.30430	1.99752	15.09569	2.74414
	5000	2.76619	0.85887	3.23408	0.70437
	10000	1.42261	0.52581	1.58222	0.38269

Table 7: MSE Comparison with Different Biasing Parameter Selection when $\rho = 0.99$

σ	n	Fixed	Data Driven	Cross Validation	Machine Learning
	100	14.87923	1.66512	11.57390	0.56786
	500	3.14854	0.58386	3.40647	0.13181
3	1000	1.63481	0.37751	1.75295	0.06533
	5000	0.35409	0.27786	0.36220	0.01367
	10000	0.17985	0.16432	0.18205	0.00750
	100	37.79664	3.17305	26.85065	1.66913
	500	7.91370	1.11462	9.17462	0.42201
5	1000	4.03342	0.75186	4.71690	0.21741
	5000	0.91885	0.35829	0.98312	0.04571
	10000	0.49220	0.32650	0.51143	0.02349
10	100	153.65050	13.47384	88.50551	11.58575
	500	30.49830	2.90200	34.86872	2.40166

	1000	14.25091	1.42652	17.95925	1.16131
	5000	3.19217	0.59004	3.81135	0.26778
	10000	1.77256	0.39761	2.00086	0.15194
	100	329.93040	29.78177	148.66190	30.95568
	500	65.57060	6.00899	69.04351	7.31261
15	1000	32.77834	3.06215	40.05806	3.60616
	5000	7.14523	1.05285	9.06731	0.79539
	10000	3.71269	0.63643	4.49238	0.39982
	100	567.40070	45.89659	229.99420	65.11880
	500	113.23550	9.47328	107.05100	14.82107
20	1000	54.58299	4.31016	63.44476	6.55396
	5000	12.06811	1.25728	15.85197	1.54759
	10000	6.39396	0.84392	8.08533	0.83401

Table 8: Summary Statistics

Variables	Mean	SD
Sale Price	180796.06	79886.69
Gr Liv Area	1499.69	505.51
Garage Area	472.66	215.19
Total Bsmt SF	1051.26	440.97
Full Bath	1.57	0.55
Total Bsmt SF*Total_Bsmt_SF	1299524.83	1301327.33
Gr_Liv_Area*Garage_Area	761469.11	551606.40
Full_Bath*Garage_Area	788.80	523.58

Table 9: Multicollinearity Test

Variables	VIF
Gr Liv Area	10.55756
Garage Area	10.98126
Total Bsmt SF*Total Bsmt SF	7.593681
Gr Liv Area*Garage Area	36.99556
Full Bath*Garage Area	34.84937
Total Bsmt SF	6.459277
Full Bath	10.80597

Table 10: MSE Comparison with Various Biasing Parameter Selection

Method	MSE	RMSE	MAE
Machine Learning	1728378	41.57377	27.04924
Data Driven	35146330	187.4735	182.8011
Fixed	35175128	187.5503	183.0162
Cross Validation	35921395	189.5294	182.2069

4.2. Discussion of findings

The results show that adaptive BP selection methods provide better estimation performance than fixed and cross-validation methods at all levels of multicollinearity considered. Specifically, the machine learning-based approach always gave the lowest MSE values, and the data-driven approach was generally second. These results suggest that parameter selection strategies that make use of information from the observed data are more successful in controlling estimation error than traditional approaches. The advantages of the adaptive methods were more pronounced in the presence of severe multicollinearity, small sample sizes, and high error variance. Our finding is in line with recent studies that suggest flexible determination of shrinkage parameters can boost the performance of estimators in difficult data scenarios (Akhtar and Alharthi, 2025; Luqman et al., 2025). Similarly, Luqman et al. (2025) showed that flexible, data-dependent biasing parameter selection enhances estimator performance, particularly in challenging regression settings. The performance gap between the methods decreased with increasing sample size, but the machine learning-based approach still showed a small advantage in all scenarios. From a practical point of view, the results show that the performance of an estimator depends not only on the estimator itself but also on the way of estimating the biasing parameter. Therefore, adaptive approaches, especially those based on machine learning, could offer a more robust framework for parameter selection bias under multicollinearity conditions.

5. Conclusion

The results of this study demonstrated that the selection of the biasing parameter using machine learning consistently yields the smallest MSE for different degrees of multicollinearity, sample sizes, and variances of errors. It is particularly useful in small-sample, high-variance settings and provides a competitive alternative to data-driven methods that typically outperform fixed and cross-validation methods. The empirical application to the Ames Housing dataset further confirms the effectiveness of the machine learning approach in the presence of severe multicollinearity, reaching the highest predictive accuracy among the considered methods. In conclusion, the results emphasize the importance of determining the biasing parameter in shrinkage estimation and indicate that adaptive methods, especially machine learning-based methods, offer a robust framework for improving the performance of estimators. Future work may extend this work through the investigation of alternative data-generating mechanisms, computational efficiency, and hybrid parameter selection strategies.

References

[1] Akay, K. U., Ertan, E., and Erkoç, A. (2023). A New biased estimator and variations based on the Kibria Lukman Estimator. *Istanbul Journal of Mathematics*, *1*(2), 74-85. https://doi.org/10.26650/ijmath.2023.00009.

[2] Akhtar, N., and Alharthi, M. F. (2025). *Enhancing accuracy in modelling highly multicollinear data using alternative shrinkage parameters for ridge regression methods. Scientific Reports, 15*, 10774. https://doi.org/10.1038/s41598-025-94857-7.

[3] Ayinde, K., Alabi, O. O., and Nwosu, U. I. (2021). Solving multicollinearity problem in linear regression model: The review suggests new idea of partitioning and extraction of the explanatory variables. *Journal of Mathematics and Statistics Studies*, *2*(1), 12-20. https://doi.org/10.32996/jmss.2021.2.1.2.

[4] Dawoud, I., Abonazel, M. R., and Awwad, F. A. (2022). Generalized Kibria-Lukman estimator: method, simulation, and application. *Frontiers in Applied Mathematics and Statistics*, *8*, 880086. https://doi.org/10.3389/fams.2022.880086.

[5] De Cock, D. (2011). Ames, Iowa: Alternative to the Boston housing data as an end-of-semester regression project. *Journal of Statistics Education*, *19*(3). https://doi.org/10.1080/10691898.2011.11889627.

[6] El-doakly, W. S. H. (2024). *A comparative study of some two-parameter ridge-type and Liu-type estimators to combat the multicollinearity problem in regression models: Simulation and application. Journal of Statistics and Econometrics, 54*(4), 105–136. https://doi.org/10.21608/jsec.2024.395938.

[7] Hoerl, A. E., and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634.

[8] Kibria, B. G., and Lukman, A. F. (2020). A new ridge-type estimator for the linear regression model: simulations and applications. *Scientifica*, *2020*(1). https://doi.org/10.1155/2020/9758378.

[9] Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics Simulation and Computation, 32*(2), 419–435. https://doi.org/10.1081/SAC-120017499.

[10] Liu, K. (1993). A new class of biased estimate in linear regression. *Communications in Statistics—Theory and Methods, 22*(2), 393–402. https://doi.org/10.1080/03610929308831027.

[11] Luqman, A. F., Ayinde, K., and Dawoud, I. (2025). *Data-driven biasing parameter selection strategies for shrinkage estimators under severe multicollinearity*. Statistical Papers, 66(2), 1123–1145.

[12] Luqman, M., Bhatti, S. H., Aydin, D., and Jamil, M. (2025). *Addressing multicollinearity in general linear model: A novel approach for ridge parameter with performance comparison. PLOS ONE, 20*(10), e0335072. https://doi.org/10.1371/journal.pone.0335072.

[13] Ugwuowo, F. I., Oranye, H. E., and Arum, K. C. (2023). On the jackknife Kibria-Lukman estimator for the linear regression model. *Communications in Statistics - Simulation and Computation*, *52*(12), 6116–6128. https://doi.org/10.1080/03610918.2021.2007401.