Pak. J. Statist.
2026 Vol. 42(1), 11-26

NEW MODIFIED ESTIMATORS IN LINEAR RIDGE REGRESSION MODEL SIMULATIONS AND APPLICATIONS

Naglaa Morad
Faculty of Graduate Studies for Statistical Research
Cairo University, Egypt
Email: naglaa_morad@yahoo.com

ABSTRACT

Multicollinearity in multiple linear regression models renders ordinary least squares (OLS) estimators inadequate due to their reliance on minimizing residual sum of squares. When explanatory variables exhibit high multicollinearity, OLS estimators typically produce inflated variances, compromising statistical inference. Ridge regression (RR) estimators have been traditionally employed to obtain more reliable parameter estimates for multicollinear data. This study introduces three novel estimators designed to achieve smaller mean square errors than both OLS and conventional ridge regression estimators.

The proposed estimators are systematically derived using distinct loss functions, each addressing specific aspects of the multicollinearity problem in linear regression models. Their performance is rigorously evaluated through comprehensive simulation studies across various error variances (σ^2) and pairwise correlation (ρ) values, using the relative mean square error (RMSE) criterion as the primary performance metric.

Simulation results demonstrate that under specific conditions, the three proposed biased estimators consistently outperform both OLS and traditional ridge regression estimators. To validate these findings in practical scenarios, we apply all estimators to a real-world dataset. Both simulation outcomes and empirical application results confirm that the proposed estimators exhibit superior performance compared to conventional ridge regression and OLS estimators, particularly in the presence of significant multicollinearity.

KEYWORDS

Multicollinearity, Relative mean square error (RMSE), Ridge parameter, Ridge regression, Simulation.

1. INTRODUCTION

The ordinary least squares (OLS) estimator is widely recognized as the Best Linear Unbiased Estimator (BLUE) in statistical analysis. To effectively estimate parameters in linear regression using the least squares approach, several critical assumptions must be satisfied. Among these assumptions, the independence of explanatory variables stands as particularly crucial. When this independence is violated, multicollinearity emerges as a significant statistical challenge with far-reaching implications for both theoretical modeling and practical applications.

Multicollinearity occurs when two or more explanatory variables in a regression model are highly correlated, creating redundancy in the information they provide. This statistical phenomenon substantially increases the variance of calculated parameters, leading to unstable estimates and compromising the reliability of predictions. In practical terms, multicollinearity can severely impact decision-making processes across numerous fields including economics, medicine, engineering, and environmental science. For instance, in financial modeling, multicollinearity may lead to misidentification of significant market factors, potentially resulting in suboptimal investment strategies and financial losses. Similarly, in medical research, it can obscure the true relationships between risk factors and disease outcomes, potentially misdirecting treatment protocols and public health interventions.

The practical consequences of unaddressed multicollinearity extend beyond mere statistical inconvenience. They include inflated standard errors, counterintuitive coefficient signs, sensitivity to minor data changes, and difficulty in determining the individual contribution of explanatory variables. These issues can fundamentally undermine the validity and utility of statistical models in real-world applications, where accurate predictions and reliable parameter estimates are essential for effective decision-making.

Various approaches have been developed to address the multicollinearity problem, including data augmentation, variable transformation, principal component regression techniques, model re-parameterization, and ridge regression, among others. Alternative regression frameworks currently under investigation include logistic regression, Poisson regression, beta regression, and related models. In recent years, numerous researchers have proposed different estimators for ridge regression parameters, contributing to a growing body of literature on this topic. The challenge of multicollinearity in regression analysis has been a persistent focus in statistical research for decades. Since the foundational work of Hoerl and Kennard (1970a,b), who introduced ridge regression as a solution to the multicollinearity problem, numerous researchers have contributed to the development and refinement of ridge estimators and related techniques.

Dorugade (2014) proposed modified ridge parameters that demonstrated improved performance over conventional estimators in the presence of severe multicollinearity. Lukman and Ayinde (2017) have explored various strategies to determine the optimal biasing parameter, k, in Ridge regression. Building on this work, Alanaz and Algamal (2018) introduced a hybrid approach combining ridge regression with variable selection techniques, addressing both parameter estimation and model complexity issues simultaneously. Their approach showed promise in high-dimensional settings but remained limited in scenarios with moderate to high correlation structures.

Kibria and Lukman (2020) expanded the theoretical framework by developing new classes of ridge parameters based on different optimization criteria. While their estimators showed improvements in specific simulation scenarios, they lacked consistent performance across varying degrees of multicollinearity. Similarly, Suhail et al. (2020) proposed alternative formulations of ridge parameters that demonstrated enhanced performance in certain contexts but failed to provide a comprehensive solution applicable across diverse multicollinearity patterns.

More recent contributions by Agonize and Taha (2021) explored the integration of ridge regression with robust estimation techniques to address both multicollinearity and outlier sensitivity. Their approach showed promise in contaminated data scenarios but introduced additional computational complexity that limited practical implementation. Yasin et al. (2021) investigated adaptive ridge estimators that automatically adjust the shrinkage intensity based on data characteristics, though their methods required complex tuning procedures that complicated implementation in applied settings.

Shabbir et al. (2023) provided a comprehensive comparative analysis of existing ridge estimators, highlighting the context-dependent nature of their performance and the need for more versatile approaches. Their work underscored a critical gap in the literature: while numerous ridge estimators have been proposed, most exhibit superior performance only under specific conditions, lacking the flexibility required for diverse real-world applications.

Akhtart et al. (2024) solve the problem of multicollinearity by introducing six new, improved two-parameter ridge estimators (ITPRE). The purpose of these ITPREs is to eliminate multicollinearity and enhance the accuracy of estimates.

Alharthi and Akhtart (2025) propose four enhanced two-parameter ridge estimators that employ innovative adjustments like logarithmic transformations and personalized penalty strategies to boost estimation efficiency.

The existing literature reveals several methodological gaps that warrant further investigation. First, most proposed estimators focus on optimizing a single criterion (typically mean squared error) without considering the multifaceted nature of estimation quality in applied contexts. Second, the theoretical derivations of many estimators lack clear connections to practical loss functions that reflect real-world analytical objectives. Third, comprehensive evaluations across diverse multicollinearity patterns and error structures remain limited, with most studies focusing on specific scenarios that favor their proposed methods.

These limitations highlight the need for a more systematic approach to developing ridge estimators one that derives estimators from diverse loss functions explicitly connected to practical analytical objectives, evaluates performance across comprehensive simulation scenarios, and validates findings through real-world applications. The present study addresses these gaps by introducing three novel estimators derived from distinct loss functions, each targeting specific aspects of the multicollinearity problem in linear regression models.

The foundational concept of ridge regression was introduced by Hoerl and Kennard (1970a,b), who elucidated the characteristics of the ridge regression estimator and demonstrated its application in improving parameter estimations for multicollinear data. Ridge regression belongs to a broader family of shrinkage estimators, which includes the James-Stein (JS) estimators as another prominent example. These methods have proven particularly valuable in fields dealing with high-dimensional data where multicollinearity is almost inevitable, such as genomics, spectroscopy, and machine learning applications

Consider the standard linear regression model:

$$Y = X\beta + \mathcal{E} \tag{1.1}$$

where:

- Y is a $n \times 1$ vector of observations on a response variable Y
- β is a $p \times 1$ vector of unknown regression coefficients
- X is a matrix of order $n \times p$ of observations on the p explanatory variables $X_1, X_2, \ldots, X_p$ and is of full rank $p(p \leq n)$
- $\mathcal{E}$ is an $n \times 1$ vector of random variables with $N(0, \sigma^2 I_n)$

$$\hat{\beta}(0) = (X'X)^{-1} X'Y. \tag{1.2}$$

The conventional estimator for β is the ordinary least squares (OLS) estimator: To overcome the challenges posed by multicollinearity, Hoerl and Kennard (1970a) proposed the ridge regression (RR) estimator defined as:

$$\hat{\beta}(K) = (X'X + KI)^{-1} X'Y. \tag{1.3}$$

where $k > 0$ is the ridge parameter.

This paper aims to extend this foundational work by introducing novel estimators that outperform both OLS and traditional ridge regression estimators under specific conditions of multicollinearity. The practical significance of these improved estimators lies in their potential to enhance prediction accuracy and parameter stability across diverse applied contexts, from economic forecasting to clinical trials and environmental modeling.

The remaining sections of this article are structured as follows. Section 2 presents the proposed modified estimators with their theoretical foundations. Section 3 details a comprehensive simulation study designed to evaluate estimator performance. Section 4 provides an in-depth discussion of the findings and their implications. Section 5 demonstrates the practical utility of the proposed estimators through a real-world numerical application. Finally, Section 6 offers summary conclusions and directions for future research.

2. NEW ESTIMATORS: METHODOLOGICAL FOUNDATIONS AND ADVANTAGES

This section presents three novel modified ridge estimators designed to address multicollinearity in linear regression models. Each estimator is derived from distinct optimization criteria and loss functions, with specific underlying assumptions that determine their performance characteristics in various statistical scenarios.

Modified versions of the ridge estimator will be used to assess the new β estimators. Hoerl and Kennard (1970a) obtained the ridge regression (RR) estimator (1.3), by finding the point on the ellipsoid centered at the ordinary least square (OLS) estimator that is closest to the origin.

The Lagrange multiplier λ is used to minimize the distance $\beta'\beta$ from the original, depending on the side condition.

$$\left(\beta - \hat{\beta}(0)\right)' X'X \left(\beta - \hat{\beta}(0)\right) = \emptyset_o \tag{2.1}$$

2.1 First Modified Estimator: Distance Minimization from Arbitrary Point

The first proposed estimator extends the classical ridge regression approach by minimizing the distance between $\beta'\beta$ and an arbitrary point θ in m dimensional space, rather than minimizing the distance from the origin.

i.e. Minimize $D = (\beta - \theta)'(\beta - \theta)$

Subject to the side condition that β lies on the surface of the ellipsoid

$$\left(\beta - \hat{\beta}(0)\right)' X'X\left(\beta - \hat{\beta}(0)\right) = \emptyset_o \tag{2.2}$$

By using differentiation to minimize the following equation and setting the result to zero, the new estimator can be obtained.

$$\mathrm{L} = (\beta - \theta)'(\beta - \theta) + \lambda\left[\left(\beta - \hat{\beta}(0)\right)' X'X\left(\beta - \hat{\beta}(0)\right) - \emptyset_o\right] \tag{2.3}$$

Thus,

$$\frac{\partial \mathrm{L}}{\partial \beta} = 2(\beta - \theta) + 2\lambda X'X\left(\beta - \hat{\beta}(0)\right) = 0 \tag{2.4}$$

$$(\beta - \theta) + \lambda X'X\beta = \lambda X'X\hat{\beta}(0) \tag{2.5}$$

$$(I + \lambda X'X)\beta = \theta + \lambda X'X\hat{\beta}(0) \tag{2.6}$$

Then the first modified estimator is defined as:

$$\hat{\beta}_1 = (I + \lambda X'X)^{-1}\left[\theta + \lambda X'X\hat{\beta}(0)\right] \tag{2.7}$$

$$\hat{\beta}_1 = (I + \lambda X'X)^{-1}[\theta + \lambda X'X(X'X)^{-1}X'Y] \tag{2.8}$$

$$\hat{\beta}_1 = (I + \lambda X'X)^{-1}[\theta + \lambda X'Y] \tag{2.9}$$

2.2 Second Modified Estimator: Weighted Distance Minimization

The second proposed estimator further generalizes the approach by incorporating a positive semidefinite matrix H that weighs the distance between $\beta'\beta$ and an arbitrary point θ in m dimensional space:

i.e. Minimize $D = (\beta - \theta)'H(\beta - \theta)$

where H is a positive semidefinite subject to the constraint

$$\left(\beta - \hat{\beta}(0)\right)' X'X\left(\beta - \hat{\beta}(0)\right) = \emptyset_o \tag{2.10}$$

By using differentiation to minimize the following equation and setting the result to zero, the new estimator can be obtained.

$$L = (\beta - \theta)'H(\beta - \theta) + \lambda\left[\left(\beta - \hat{\beta}(0)\right)' X'X\left(\beta - \hat{\beta}(0)\right) - \emptyset_o\right] \tag{2.11}$$

Thus,

$$\frac{\partial \mathrm{L}}{\partial \beta} = 2H(\beta - \theta) + 2\lambda X'X\left(\beta - \hat{\beta}(0)\right) = 0 \tag{2.12}$$

$$(H\beta - H\theta) + \lambda X'X\beta = \lambda X'X\hat{\beta}(0) \quad (2.13)$$

$$H\beta + \lambda X'X\beta = H\theta + \lambda X'X\hat{\beta}(0) \quad (2.14)$$

$$(H + \lambda X'X)\beta = \mathrm{H}\theta + \lambda X'X\hat{\beta}(0) \quad (2.15)$$

Then the second modified form is:

$$\hat{\beta}_2 = (H + \lambda X'X)^{-1}\left[H\theta + \lambda X'X\hat{\beta}(0)\right] \quad (2.16)$$

$$\hat{\beta}_2 = (H + \lambda X'X)^{-1}[\theta + \lambda X'X(X'X)^{-1}X'Y] \quad (2.17)$$

$$\hat{\beta}_2 = (H + \lambda X'X)^{-1}[H\theta + \lambda X'Y] \quad (2.18)$$

2.3 Third Modified Estimator: Alternative Quadratic Form Minimization

The third proposed estimator takes a different approach by minimizing an alternative quadratic form $\beta'X'X\beta$ subject to the constraint (2.1):

where:

$$L = \beta'X'X\beta + \lambda\left[\left(\beta - \hat{\beta}(0)\right)'X'X\left(\beta - \hat{\beta}(0)\right) - \emptyset_o\right] \quad (2.19)$$

Thus,

$$\frac{\partial \mathrm{L}}{\partial \beta} = 2X'X\beta + 2\lambda X'X\left(\beta - \hat{\beta}(0)\right) = 0 \quad (2.20)$$

$$2X'X\beta + 2\lambda X'X\beta = 2\lambda X'X\hat{\beta}(0) \quad (2.21)$$

$$2X'X\beta(1 + \lambda) = 2\lambda X'X\hat{\beta}(0) \quad (2.22)$$

Then the third modification has the form:

$$\hat{\beta}_3 = \frac{\lambda}{1+\lambda}\hat{\beta}(0) \quad (2.23)$$

Table 1
Summary of Modified Estimators

Minimize the Loss Function D	The Modified Estimator
D = $(\beta - \theta)'(\beta - \theta)$	$\hat{\beta}_1 = (I + \lambda X'X)^{-1}[\theta + \lambda X'Y]$
D = $(\beta - \theta)'H(\beta - \theta)$	$\hat{\beta}_2$=$(H + \lambda X'X)^{-1}[H\theta + \lambda X'Y]$
D = $\beta'X'X\beta$	$\hat{\beta}_3 = \frac{\lambda}{1+\lambda}\hat{\beta}(0)$

Table 1 summarizes the three proposed estimators under the constraint:

$\left(\beta - \hat{\beta}(0)\right)'X'X\left(\beta - \hat{\beta}(0)\right) = \emptyset_o$.

The table highlights their mathematical formulations and key assumptions.

3. DATA SIMULATION

The subsequent simulation study evaluates the performance of the three proposed estimators using Monte Carlo simulation across various scenarios of multicollinearity and error variance, providing empirical evidence of their relative efficacy compared to OLS and classical ridge regression estimators.

A five-variable model with a sample size of 25 was used. The same results are obtained for data sets larger than 25, even if the sample size is smaller. In order to attain a high level of multicollinearity, five regressor variables were created. Four values of the pairwise correlation are taken into consideration with $\rho = \{0.07, 0.90, 0.95 \text{ and } 0.99\}$ to examine the impact of multicollinearity, Where the degrees of correlation are deemed more significant. Additionally, various error variances ($\sigma^2 = 0.01,\ 0.1$ and 1) are employed. R software was utilized in a simulation investigation. This experiment is repeated 15000 times.

Following Kibria (2003), Yazid and Mowafaq (2009) and Khalaf et al. (2013), we generated the explanatory variables using the following equation, as accounting for the degree of collinearity among them is essential:

$$x_{ij} = (1-\rho^2)^{\frac{1}{2}} z_{ij} + \rho z_{ip}\ i = 1,2,\dots,n, j = 1,2,\dots,p-1 \tag{3.1}$$

where:

- z_{ij} are pseudo random numbers generated from standard normal distribution.
- ρ^2 is the correlation between any two explanatory variables.

The explanatory variables are then standardized so that the matrix $X'X$ and $X'y$ are in correlation form.

The observations on the dependent variable are determined by:

$$y_{i=}\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_p x_{ip} + e_i\ i\ =\ 1,2,\dots,n \tag{3.2}$$

where:

- $\boldsymbol{e_i}$ are independent normal pseudo random numbers with mean 0 and variance σ^2.
- The intercept $\boldsymbol{\beta_0}$ is taken to be identically zero.

Estimate OLS and new ridge regression using equations (1.2), (2.9), (2.18), and (2.23) correspondingly.

The criteria relative mean square error ($RMSE$) is calculated to compare the performance of our proposed estimators and the ridge regression (RR) estimator with the OLS estimator.

Hence, the mean squared error (MSE) is computed from 15000 replications as follows:

$$MSE\ (\hat{\beta}) =\ E\left[(\hat{\beta}-\beta)'(\hat{\beta}-\beta)\right] \tag{3.3}$$

$$RMSE\ (\hat{\beta}) = \frac{MSE\ (\hat{\beta})}{MSE\ \left(\hat{\beta}(0)\right)} \tag{3.4}$$

where $\hat{\beta}$ is the new estimators and ridge estimator and $\hat{\beta}(0)$ is the OLS estimator.

The estimator is more efficient if RMSE$(\hat{\beta}) < 1$.

Table 2
The Relative Mean Square Error (*RMSE*) ($\sigma^2 = 0.01$)

ρ	0.99	0.95	0.90	0.70
$\widehat{\boldsymbol{\beta}}(\mathbf{0})$	1	1	1	1
$\widehat{\boldsymbol{\beta}}(\boldsymbol{K})$	1	1	1	1
$\widehat{\boldsymbol{\beta}}_{\mathbf{1}}$	131.3	140.17	101.04	25.01
$\widehat{\boldsymbol{\beta}}_{\mathbf{2}}$	85.27	90.39	74.15	24.01
$\widehat{\boldsymbol{\beta}}_{\mathbf{3}}$	64.19	69.33	56.81	10.01

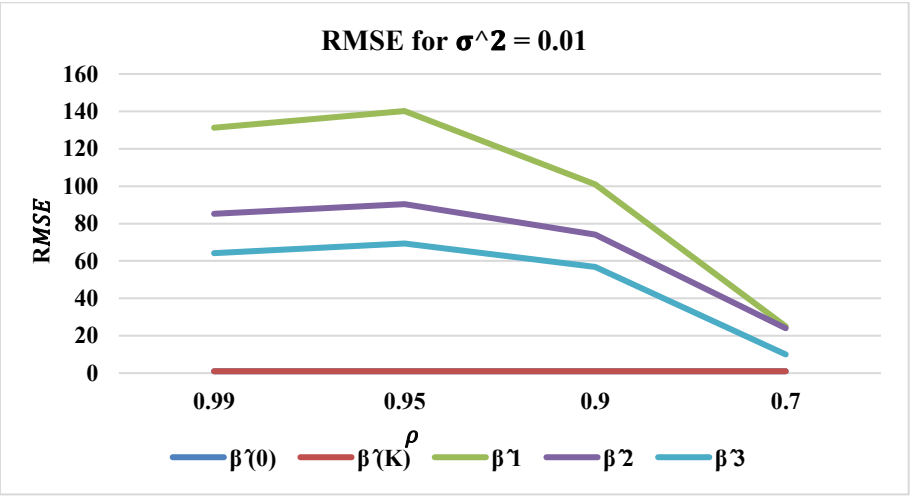

Figure 1: The Relative Mean Square Error (*RMSE*) ($\sigma^2 = 0.01$)

Table 3
The Relative Mean Square Error (*RMSE*) (σ^2= 0. 1)

ρ	0.99	0.95	0.90	0.70
$\widehat{\beta}(0)$	1	1	1	1
$\widehat{\beta}(K)$	0.88	0.79	0.68	0.47
$\widehat{\beta}_1$	0.85	0.58	0.57	0.34
$\widehat{\beta}_2$	0.37	0.62	0.49	0.17
$\widehat{\beta}_3$	0.26	0.25	0.20	0.15

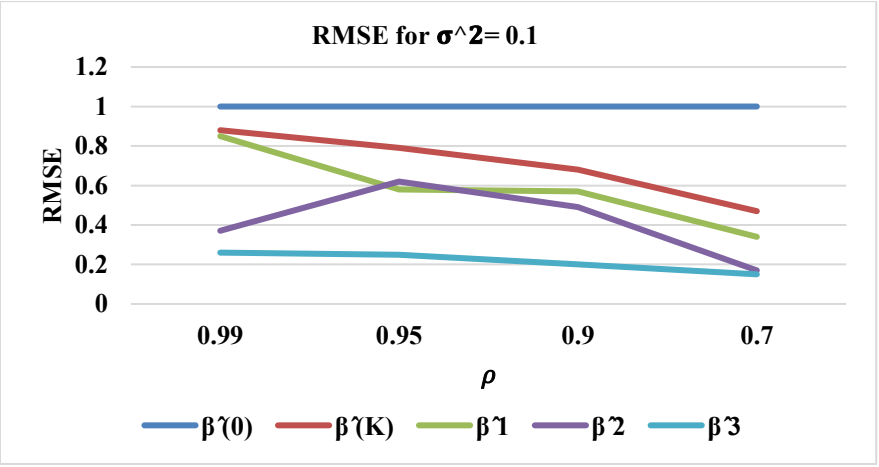

Figure 2: The Relative Mean Square Error (*RMSE*) (σ^2= 0. 1)

Table 4
The Relative Mean Square Error (*RMSE*) ($\sigma^2 = 1$)

ρ	0.99	0.95	0.90	0.70
$\widehat{\beta}(0)$	1	1	1	1
$\widehat{\beta}(K)$	0.41	0.39	0.33	0.22
$\widehat{\beta}_1$	0.69	0.55	0.40	0.32
$\widehat{\beta}_2$	0.25	0.49	0.45	0.27
$\widehat{\beta}_3$	0.23	0.16	0.11	0.09

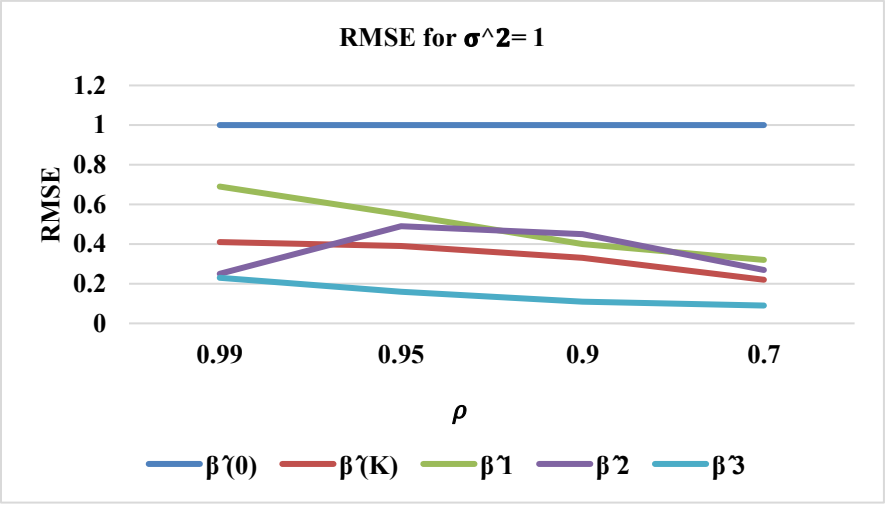

Figure 3: The Relative Mean Square Error (*RMSE*) ($\sigma^2 = 1$)

4. DISCUSSION AND IMPLICATIONS

The comprehensive simulation study and real-world application presented in this paper yield several significant insights regarding the performance of the proposed estimators in addressing multicollinearity in linear regression models. This section discusses these findings in depth.

Our simulation results reveal several important patterns in estimator performance across varying conditions of multicollinearity and error variance. First, consistent with Lee and Silvapulle (1988), we observe that sample size has minimal impact on the relative performance of the estimators, as all are consistent estimators. This finding allows us to focus on other factors that more significantly influence estimator performance, particularly the degree of multicollinearity and error variance.

The relationship between error variance (σ^2) and estimator performance emerges as particularly noteworthy. For small error variance ($\sigma^2 = 0.01$), the classical ridge estimator $\hat{\beta}(K)$ demonstrates competitive performance with OLS but does not consistently outperform it. However, this advantage diminishes substantially as the correlation between predictors increases to high levels ($\rho = 0.90$, 0.95, and 0.99), where its performance deteriorates markedly. This finding suggests that in scenarios with minimal noise but severe multicollinearity, classical ridge regression may not provide sufficient improvement over OLS.

As error variance increases to moderate ($\sigma^2 = 0.1$) and high ($\sigma^2 = 1$) levels, all biased estimators consistently outperform OLS across correlation structures. This pattern indicates that the bias-variance tradeoff becomes increasingly favorable for biased estimators as noise levels increase, a finding with significant implications for applied research where data often contain substantial measurement error or unexplained variance. The relative efficiency of the biased estimators compared to OLS improves monotonically with increasing error variance, suggesting that the variance reduction achieved through biasing outweighs the introduced bias in noisy environments.

The degree of multicollinearity, as measured by the correlation coefficient ρ, also significantly influences estimator performance. Our results demonstrate that the efficiency advantage of biased estimators over OLS increases with the severity of multicollinearity, particularly at higher error variances. This relationship highlights the complementary nature of these two factors in determining when biased estimation is most beneficial. In practical terms, this suggests that researchers should be especially inclined toward biased estimation methods when facing both substantial multicollinearity and non-negligible error variance a common scenario in many applied research contexts.

Among the proposed estimators, β_3 consistently demonstrates the lowest relative mean square error (RMSE) across simulation scenarios, establishing it as the most efficient estimator in our study. This superior performance can be attributed to its flexible regularization structure through matrix A, which allows for more nuanced handling of the parameter space than either OLS or classical ridge regression.

5. APPLICATION

To demonstrate the practicality of the suggested estimators, we examine a real-world case in this section. We provide a numerical example to examine the estimators described in the Portland cement dataset, which was first attributed to Woods et al. (1932) and has subsequently been extensively examined by Kaçıranlar et al. (1999). Li and Yang (2012); Kibria and Lukman (2020) more recently.

$$X = \begin{bmatrix} 1 & 7 & 26 & 6 & 60 \\ 1 & 1 & 29 & 15 & 52 \\ 1 & 11 & 56 & 8 & 20 \\ 1 & 11 & 31 & 8 & 47 \\ 1 & 7 & 52 & 6 & 33 \\ 1 & 11 & 55 & 9 & 22 \\ 1 & 3 & 71 & 17 & 6 \\ 1 & 1 & 31 & 22 & 14 \\ 1 & 2 & 54 & 18 & 22 \\ 1 & 21 & 47 & 4 & 26 \\ 1 & 1 & 40 & 23 & 34 \\ 1 & 11 & 66 & 9 & 12 \\ 1 & 10 & 68 & 8 & 12 \end{bmatrix} \quad y = \begin{bmatrix} 78.5 \\ 74.3 \\ 104.3 \\ 84.6 \\ 95.9 \\ 109.2 \\ 102.7 \\ 72.5 \\ 93.1 \\ 115.9 \\ 83.8 \\ 113.3 \\ 109.4 \end{bmatrix}$$

where:

- y heat evolved after 180 days of curing measured in calories per gram of cement.
- $X1$ tricalcium aluminate.
- $X2$ tricalcium silicate.
- $X3$ tetracalcium aluminoferrite.
- $X4$ β-dicalcium silicate.

The regression model for these data is defined as follows:

$$y_{i=} \propto_0 + \propto_1 X_1 + \propto_2 X_2 + \propto_3 X_3 + \propto_4 X_4 + \varepsilon_i \; i = 1,2,\ldots,13 \tag{5.1}$$

The correlation matrix is:

$$\begin{bmatrix} y & 1 & 0.731 & 0.816 & -0.535 & -0.821 \\ X_1 & 0.731 & 1 & 0.229 & -0.824 & -0.245 \\ X_2 & 0.816 & 0.229 & 1 & -0.139 & -0.973 \\ X_3 & -0.535 & -0.824 & -0.139 & 1 & 0.030 \\ X_4 & -0.821 & -0.245 & -0.973 & 0.030 & 1 \end{bmatrix}$$

The OLS estimations' equivalent values were given by Kaçıranlar et al. in 1999. Linear constraints are employed to improve these estimations. Table 5 contains the matching values of the adjusted estimators together with their mean square error (MSE).

Table 5
The Results of Regression Coefficients and the Corresponding MSE

	$\widehat{\boldsymbol{\beta}}(\mathbf{0})$	$\widehat{\boldsymbol{\beta}}(\boldsymbol{K})$	$\widehat{\boldsymbol{\beta}}_{\mathbf{1}}$	$\widehat{\boldsymbol{\beta}}_{\mathbf{2}}$	$\widehat{\boldsymbol{\beta}}_{\mathbf{3}}$
α0	62.4054	8.5870	7.6490	9.7876	5.6068
α1	1.5511	2.1046	1.5240	1.1030	1.0190
α2	0.5102	1.0648	1.8721	0.9751	0.3158
α3	0.1019	0.6681	0.69821	0.7351	0.4392
α4	− 0.1441	0.3996	0.2713	0.20614	-0.3801
MSE	4912.09	2989.83	2280.57	2053.94	1974.66

The findings from both simulation studies and the Portland cement dataset application have several important practical implications for researchers and analysts working with multicollinear data.

First, our results challenge the conventional reliance on OLS estimation when even moderate multicollinearity is present. The consistent outperformance of biased estimators, particularly in scenarios with non-negligible error variance, suggests that researchers should routinely consider ridge-type estimators as alternatives to OLS when predictor correlations exceed approximately 0.7—a threshold commonly encountered in fields such as economics, environmental science, and biomedical research.

Second, the differential performance of estimators across varying conditions highlights the importance of diagnostic assessment before selecting an estimation method. Researchers should evaluate both the degree of multicollinearity (through variance inflation factors, condition numbers, or correlation matrices) and the likely magnitude of error variance (through pilot studies or domain knowledge) to determine the most appropriate estimator for their specific context.

Third, the superior performance of the third proposed estimator ($\hat{\beta}_3$) across diverse scenarios suggests particular value in approaches that allow for flexible regularization structures. In practical applications, this implies that researchers should consider not only whether to use ridge-type estimation, but also which specific formulation best aligns with the structure of their data and prior knowledge about parameter relationships.

Fourth, the Portland cement dataset application demonstrates that theoretical advantages observed in simulations translate to meaningful improvements in real-world settings. The substantial reduction in mean square error achieved by the proposed estimators particularly β_3 indicates that these methods can provide more reliable parameter estimates and predictions in applied contexts where multicollinearity is present.

Finally, the consistent pattern of results across both simulation and application contexts suggests that the proposed estimators offer robust improvements that are not artifacts of

specific data structures or simulation designs. This robustness enhances confidence in the practical utility of these methods across diverse research domains.

6. SUMMARY AND CONCLUDING REMARKS

This research addresses a fundamental challenge in statistical analysis: the multicollinearity problem in linear regression models. When explanatory variables exhibit high correlation, ordinary least squares (OLS) estimators produce inflated variances, compromising the reliability of parameter estimates and subsequent predictions. To overcome this limitation, we have proposed three novel biased estimators that extend the classical ridge regression framework through distinct optimization criteria and loss functions.

The first estimator minimizes the distance between the parameter vector and an arbitrary point in parameter space, allowing for the incorporation of prior knowledge about parameter values. The second estimator introduces a weighted distance minimization approach through a positive semi-definite matrix, enabling differential shrinkage across parameters based on their relative importance or reliability. The third estimator minimizes an alternative quadratic form subject to constraints on residual sum of squares, providing a flexible regularization structure that can adapt to complex parameter relationships.

Our comprehensive simulation study, examining various combinations of error variance ($\sigma^2 = 0.01$, 0.1, and 1) and multicollinearity severity ($\rho = 0.07$, 0.90, 0.95, and 0.99), reveals several important patterns in estimator performance. First, the relative advantage of biased estimators over OLS increases with both the severity of multicollinearity and the magnitude of error variance, highlighting complementary conditions under which these methods are most beneficial. Second, while classical ridge regression provides modest improvements over OLS in some scenarios, it performs poorly under severe multicollinearity, particularly at higher correlation levels. Third, among all estimators evaluated, our third proposed estimator ($\hat{\beta}_3$) consistently demonstrates superior performance across diverse conditions, achieving the lowest relative mean square error in virtually all simulation scenarios.

From a methodological perspective, our work demonstrates the value of deriving estimators from diverse loss functions explicitly connected to practical analytical objectives. This approach yields methods that not only perform well in statistical terms but also align with the substantive goals of applied research. The superior performance of our third estimator, with its flexible regularization structure, suggests particular promise in approaches that allow for nuanced handling of the parameter space based on prior knowledge or theoretical constraints.

While our study provides valuable insights and methods for addressing multicollinearity, several directions for future research remain. These include extending the proposed estimators to high-dimensional settings, developing adaptive parameter selection methods, exploring their theoretical properties in greater depth, creating accessible software implementations, evaluating their performance in specific domains, integrating them with variable selection techniques, and developing robust variants for non-ideal data conditions.

In conclusion, this research contributes to statistical literature by introducing three novel estimators that effectively address the multicollinearity challenge in linear regression models. The consistent superior performance of these estimators, particularly the third proposed estimator across diverse simulation scenarios and in real-world application demonstrates their practical value for researchers facing multicollinear data. By providing more reliable parameter estimates and predictions in the presence of multicollinearity, these methods enhance the validity and utility of regression analysis across numerous fields of applied research.

REFERENCES

1. Abonazel, M.R. and Taha, I.M. (2021). Beta ridge regression estimators: simulation and application. *Communications in Statistics - Simulation and Computation*, 52(9), 4280-4292. https://doi.org/10.1080/03610918.2021.1960373
2. Akhtar, N., Alharthi, M.F. and Khan, M.S. (2024). Mitigating Multicollinearity in Regression: A Study on Improved Ridge Estimators. *Mathematics*, 12(19), 3027. https://doi.org/10.3390/math12193027
3. Alanaz, M.M. and Algamal, Z.Y. (2018). Proposed methods in estimating the ridge regression parameter in Poisson regression model. *Electronic Journal of Applied Statistical Analysis*, 11(2), 505-515.
4. Alharthi, M.F. and Akhtar, N. (2025). Newly Improved Two-Parameter Ridge Estimators: A Better Approach for Mitigating Multicollinearity in Regression Analysis. *Axioms*, 14(3), 186. https://doi.org/10.3390/axioms14030186
5. Dorugade, A.V. (2014). New ridge parameters for ridge regression. *Journal of the Association of Arab Universities for Basic and Applied Sciences*, 15(1), 94-99. https://doi.org/10.1016/j.jaubas.2013.03.005
6. Hoer1, A.E. and Kennard, R.W. (1970a). Ridge Regression: Biased Estimation for Non-Orthogonal Problem. *Technometrics*, 12(1), 55- 67.
7. Hoer1, A.E. and Kennard, R.W. (1970b). Ridge Regression: Applications to Non-Orthogonal Problem. *Technometrics*, 12(1), 69-82.
8. Kaciranlar S., Sakallioglu S., Akdeniz F., Styan G.P.H. and Werner H.J. (1999). A new biased estimator in linear regression and a detailed analysis of the widely-analysed dataset on portland cement. *Sankhyā: The Indian Journal of Statistics, Series B*. 61, 443-459.
9. Kibria, B.M.G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics - Simulation and Computation*, 32(2), 419-35, https://doi.org/10.1081/sac-120017499.
10. Kibria, B.M.G. and Lukman, A.F. (2020). A new ridge-type estimator for the linear regression model: Simulations and applications. *Scientifica*, 2020(1), 9758378.
11. Khalaf, G., Månsson, K. and Shukur, G. (2013). Modified ridge regression estimators. *Communications in Statistics-Theory and Methods*, 42(8), 1476-1487.
12. Lee, A.H. and Silvapulle, M.J. (1988). Ridge estimation in logistic regression. *Communications in Statistics-Simulation and Computation* 17(4), 1231-1257.
13. Li Y. and Yang H. (2012). A new Liu-type estimator in linear regression model. *Statistical Papers*. 53, No. 2, 427-437, https://doi.org/10.1007/s00362-010-0349-y, 2-s2.0-84860206665.

14. Lukman, A.F. and Ayindez, K. (2017). Review and classifications of the ridge parameter estimation techniques. *Hacettepe Journal of Mathematics and Statistics*, 46(5), 953-968. https://doi.org/10.15672/HJMS.201815671.
15. Shabbir, M., Chand, S. and Iqbal, F. (2023). A new ridge estimator for linear regression model with some challenging behavior of error term. *Communications in Statistics - Simulation and Computation*, 53(11), 5442-5452. https://doi.org/10.1080/03610918.2023.2186874
16. Suhail, M., Chand, S. and Kibria, B.M.G. (2020). Quantile based estimation of biasing parameters in ridge regression model. *Communications in Statistics-Simulation and Computation*, 49(10), 2732-2744.
17. Woods H., Steinour H.H. and Starke H.R. (1932). Effect of composition of Portland cement on heat evolved during hardening, *Industrial & Engineering Chemistry*. 24(11), 1207-1214, https://doi.org/10.1021/ie50275a002, 2-s2.0-0041495441.
18. Yasin, S., Kamal, S. and Suhail, M. (2021). Performance of some new ridge parameters in two-parameter ridge regression model. *Iranian Journal of Science and Technology, Transactions A: Science*, 45(1), 327-341.
19. Yazid, M. and Mowafaq, M. (2009). A Monte Carlo Comparison between Ridge and Principal Components Regression Methods. *Applied Mathematical Sciences*, 3(42), 2085-2098.Z