***Original Research Article***

**EFFICIENT COMBINED ESTIMATOR FOR PARAMETER ESTIMATION OF LINEAR REGRESSION MODEL WITH MULTICOLLINEARITY**

**ABSTRACT**

*The classical linear regression model relies on several key assumptions, including homoscedasticity, normality of errors, independence of observations and the absence of multicollinearity among explanatory variables (Gujarati, 2021), [1]. These assumptions are rarely fulfilled in real life situations. Multicollinearity occurs when the assumption of independent explanatory variables is violated (Alreshidi et al., 2025), [2]. There are many sources of multicollinearity, some of which are the data collection methods, the constraints placed on the model or having an overdetermined model (Paul, 2006), [3]. When multicollinearity exists in a model using conventional parameter estimation models like the Ordinary Least Squares (OLS) often leads to unstable and unreliable parameter estimates. (Alharthi & Akhtar, 2025), [4]. To address these challenges several biased estimators have been developed. Some of these are the ridge (Hoerl and Kennard, 1970), [5], liu (Liu, 1993), [6] and principal components (pc) (Hotelling, 1933), [7] estimators. Each of these existing estimators have their strengths and limitations. However, no single estimator consistently outperforms the others under all conditions. Over the years other researchers have developed combined estimators with the expectation that the combination of different estimators might inherit the individual advantages of the estimators. Following their work, in this paper effort is made to provide a combined estimator based on ridge (Hoerl and Kennard, 1970),[5] liu (Liu, 1993),[6] and principal components (pc) (Hotelling, 1933),[7] estimators. This estimator- the principal component ridge liu, leverages on the strengths of these three existing estimators. The performance of the developed estimator is compared with the existing individual estimators using Mean Square Error (MSE). Results show that the developed estimator performs better than the existing ones providing more stable and accurate parameter estimates in the presence of multicollinearity. Monte Carlo experiments a robust method for assessing statistical properties under controlled conditions with varying degrees of multicollinearity, error variance, and sample size (Oduntan, 2024), [8] were performed one thousand (1000) times on two (2) linear regression models with four (4) and seven (7) explanatory variables exhibiting five (5) degrees of multicollinearity (0.75, 0.85, 0.95, 0.99, 0.999), three(3) levels of error variance (1, 25, 100), at eight (8) sample sizes (n=10, 15, 20,30,40,50,100 and 250). The MSE criterion was used to examine the estimators and the number of times each estimator had the minimum MSE was counted at each combination of classifications. The ranking of the estimators was also done based on their MSE. Tables and figures were used to present the results of the findings. The results of the investigation revealed that when multicollinearity problems exist in linear regression models, the proposed RHKMALUMIWPC estimator is best.*

**KEYWORDS**

Ridge Estimator, Liu Estimator, Principal Component Estimator, Mean Square Error (MSE), Multicollinearity

1. **INTRODUCTION**

When the assumption of independence among explanatory variables in multiple linear regression models is violated, multicollinearity occurs (Gujarati, 2021 [1]). The Ordinary Least Squares (OLS) estimator is commonly used to estimate parameters in linear regression models. While the OLS estimator provides unbiased estimates, its performance deteriorates under multicollinearity. Specifically, the parameter estimates become inefficient, may exhibit incorrect signs, and have high standard deviations. Moreover, they become extremely sensitive to slight changes in the data (Chatterjee and Hadi, 2006 [2]; Lukman et al., 2018 [3]).

To address this issue, several biased estimators have been developed, including the Ridge Estimator proposed by Hoerl and Kennard (1970) [5], the Liu Estimator introduced by Liu (1993) [4], and the Principal Component Estimator by Hotelling (1933) [6]. This study aims to propose a new estimator that combines the Ridge, Liu, and Principal Component approaches to address multicollinearity. We also compare the performance of this combined estimator with the individual ones.

**2.0 MATERIALS AND METHODS**

**2.1 Ordinary Least Square Estimator**

Consider the standard regression model:

$Y=Xβ+U$ (1)

Where X is an n x p matrix with full rank, Y is an n x 1 vector of dependent variable, $β$ is a p x 1 vector of unknown parameters, and $U $is the error term, such that $E\left(U\right)=0$ and $E\left(UU^{1}\right)=σ^{2}I\_{n}$

provided $X^{1}X$ is invertible (Lukman et al. 2018), the OLS estimator is given as

$\hat{β}=(X^{'}X)^{-'}X^{'}Y$ (2)

Consider the regression model in equation (1) and letting Ʌ= diag $(λ\_{1,}λ\_{2},…,λ\_{p})$ be a pxp diagonal matrix of the eigenvalues of $X^{'}X$, and T be a p × p orthogonal matrix whose columns are the eigenvectors associated with $λ\_{1},λ\_{2}….λ\_{p}$ such that $TT^{'}=T^{'}T=I\_{P}$ .

Then,

 $X^{'}X= TɅT^{'}$ (3)

and

$Ʌ=T^{'}X^{'}XT$ (4)

Defining Z = XT, then

X=$ZT^{'}$ (5)

Putting eqn. (6) into eqn. (5)

$Ʌ=T^{'}X^{'}XT=Z^{'}Z$ (6)

Substituting eqn. (6) for X in eqn. (1), we obtain the equivalent model given as:

Y=$ZT^{'}β+U$ (7)

Let $T^{'}β=α$ , then:

β=T$α$ (8)

and

$Y=Zα+U$ (9)

The ordinary Least Squares estimator of $α, α\_{OLS}$ is given as:

$\hat{α}\_{OLS}=(Z^{'}Z)^{-'}Z^{'}y=Ʌ^{-'}Z^{'}Y$ (10)

**2.2 Ridge Regression Estimator**

To solve for the problem of multicollinearity, Hoerl and Kennard (1970), [5] developed the Ridge Regression estimator which proceeds by adding a small value, K, to the diagonal elements of the $X^{1}X$ matrix before computing the $\hat{β}$. This value balances the trade-off between fitting the data well and keeping the coefficients small. Using the model in equation (9), this corresponds to adding a small value, K, to the diagonal elements of the $Z^{'}Z$ matrix before computing the $\hat{α}$. Thus, the ridge regression estimator (RE) becomes:

$\hat{α}\_{RE}=(Z^{'}Z+KI)^{-'}Z^{'}Y$ (11)

Where $Z^{'}Z$ is a p x p product matrix of explanatory variables,

$Z^{'}Y$ is a p x 1 vector of the product of dependent and explanatory variables,

K = diagonal ($K\_{1},K\_{2},…,K\_{P})$,  $K\_{i}$= $\hat{\frac{σ^{2}}{\hat{α\_{i}}}}$ ≥ 0. i= 1, 2, ..., p.

When $K\_{1}=K\_{2}=….=K\_{p}=K$, K>0, then $\hat{α}\_{GRE}=\hat{α}\_{RE}$ and if K=0, then $\hat{α}\_{RE}$=$\hat{α}\_{OLS}$

In equation (11) above, K > 0 and *I* is an identity matrix. Note that if K=0 the ridge estimator (RE) equals the OLS estimator. The value of k used in this study is the one proposed by Hoerl and Kennard (1970), [5] and Fayose and Ayinde (2019), [7].

**Hoerl and Kennard Generalized Form**

K=$\hat{σ}^{2}/\hat{α}\_{i}^{2}$, i=1,2, 3…, p (12)

$\hat{σ}^{2}=\sum\_{i=1}^{n}\frac{u\_{i}^{2}}{n-p}$ (13)

 is the MSE from the OLS regression

$α\_{i }$ is the ith element of the vector $α$ from the OLS regression

p is the number of regressors and n is the sample size.

**Fayose and Ayinde Generalized Form**

$\frac{\hat{σ}}{α\_{i}^{2}}\left\{ \left[\left.\left(\left.\frac{\hat{α}\_{i}^{4}λ\_{i}^{2}}{4\hat{σ}^{2}}\right)\right.+\left(\left.\frac{6\hat{α}\_{i}^{4}λ\_{i}}{\hat{σ}^{2}}\right)\right.\right]\right.^{\frac{1}{2}}-\left(\left.\frac{\hat{α}\_{i}^{2}λ\_{i}}{2\hat{σ}^{2}}\right)\right.\right\}$(14)

$\hat{σ}^{2}=\sum\_{i=1}^{n}\frac{u\_{i}^{2}}{n-p}$ is the MSE from the OLS regression

$α\_{i }$ is the ith element of the vector $α$ from the OLS regression

p is the number of regressors and n is the sample size.

$λ\_{i }$ is the ith eigenvalue of the $Z^{1}$Z matrix

Following Fayose and Ayinde (2019), [7] seven different forms of the biasing parameter k that perform more efficiently than the generalized form were also used as k values. These different forms are maximum, minimum, Arithmetic mean, Geometric mean, Harmonic mean, Mid-range and Median.

**2.3 LIU ESTIMATOR**

Liu (1993), [4] motivated by the interpretation of the ridge estimate developed an alternative biased estimator to overcome multicollinearity for the linear regression model in equation (9) as:

$\hat{α}\_{LIU}=[\left(Z^{'}Z+I\right)^{-'}$ ($Z^{'}Z+dI) \hat{α}\_{OLS}$ (15)

where d is the Biasing parameter and is defined as:

$d\_{L}=1-\hat{σ}^{2}\left[\frac{\sum\_{i=1}^{p}\frac{1}{λ\_{i}(λ\_{i}+1)}}{\frac{\sum\_{i=1}^{p}\hat{α}\_{i}^{2}}{(λ\_{i}+1)^{2}}}\right]$ (16)

Where $\hat{α}=Q^{'}β\_{OLS}$ and $\hat{σ}^{2}$ are Ordinary Least Square estimators of α and $σ^{2}$ respectively, and *Q* is the matrix of eigenvectors corresponding to the eigenvalues o the matrix $X^{'}X$

**\_2.4 PRINCIPAL COMPONENT ESTIMATOR**

The main purpose of Principal Component Regression is to estimate the values of a response variable on the basis of selected Principal Components of the explanatory variables. It is a multivariate technique developed by Hotelling (1933), [5] for explaining a set of correlated variables by a reduced number of uncorrelated ones with maximum variances called Principal components (PCs) (Pongpiachan et al., 2024), [12]. The estimates produced using the Principal Component estimators are biased (Weeraratne et al., 2024), [13]. Two stages are involved when using principal component regression. The first stage reduces the predictor variables of the model using principal component analysis. The second stage involves using the reduced variables obtained from the principal component analysis in an OLS regression fit (Ayinde et al., 2012), [14].From our regression model in equation (9), the columns of Z, which define a new set of orthogonal regressors, such as Z=($Z\_{1}, Z\_{2}, …Z\_{p})$=[$Z\_{r}, Z\_{p-r}$] are referred to as principal components. The pxp matrix of eigen vectors T =($t\_{1}, t\_{2}, …t\_{p})$ can also be written as [$T\_{r},T\_{p-r}]$ with descending eigen values $λ\_{1}\geq λ\_{2}\geq …\geq λ\_{p}$ and that the last of these eigen values are approximately equal to zero. Thus (1) can be written as:

$Y=Xβ+U$

 =X$TT^{'}β$ +*U*

 = X$ T\_{r}T\_{r}^{'}β$ + X$ T\_{p-r}T\_{p-r}^{'}β$ +*U*

 =$Z\_{r}α\_{r}$+ $Z\_{p-r}α\_{p-r}$+ *U* (17)

Where $Z\_{r}$ contains PCs that are used be used in the regression model and $Z\_{p-r} $contain PCs that are discarded from the model. Thus, the regression equation becomes:

$Y=Z\_{r}α\_{r}+U$ (18)

**2.5 THE PROPOSED PRINCIPAL COMPONENT RIDGE LIU ESTIMATOR**

Having obtained the ridge estimator (Hoerl and Kennard, 1970),[5] as earlier stated in equation (11), the Liu estimator(Liu, 1993*),* as stated in equation (15) and the principal component estimator (Hotelling, 1933),[7] as stated in equation (16). The combination of equation (11), equation (15) and equation (16), gives the new proposed estimator derived as :

$\hat{α}\_{PCRLIU}=\left(Z\_{r}^{1}Z\_{r}+KI\right)^{-1}$ ($Z\_{r}^{1}Z\_{r}+dI) \hat{α}\_{PC}$ (19)

**2.6 MODEL FORMULATION FOR MONTE CARLO STUDY**

To investigate the performance of our proposed and existing estimators when multicollinearity is present, consider a multiple linear regression model of the form:

$Y\_{t}=α\_{0}+α\_{1}X\_{t1}+α\_{2}X\_{t2}+…+α\_{p}X\_{tp}+U\_{t}$ (20)

Where t = 1, 2,...,n; p=4, 7

Where$U\_{t}\~N(0,σ^{2}), X\_{ti}\ni t=1,2,….,n and i=1,2,….,p $ are fixed regressors

In this study, the number of explanatory variables (p) used were four (4) and seven (7). The values of the sample sizes considered are: 10, 15, 20, 30, 40, 50, 100, 250.

**2.7 Procedure for generating the Explanatory variables**

The simulation procedure used by McDonald and Galarneau (1975), [8] Wichern and Churchill (1978), [9] Gibbons (1981), [10] Kibria (2003), [11] Dorugade and Kashid (2010), [12] Dorugade (2016), [13] and Lukman et al (2018), [14] is also be used to generate the explanatory variables in this study. This is given as:

$X\_{ti}=\left(1-ρ^{2}\right)^{\frac{1}{2}}Z\_{ti}+ρZ\_{tp}$ (21)

t=1, 2, 3…, n. i=1, 2,…,p.

Where $Z\_{ti}$ is of independent standard normal distribution with mean zero and unit variance, $ρ$ is the correlation between any two explanatory variables and p is the number of explanatory variables. The $ρ$ is specified so that the correlation between any two regressors is given as $ρ^{2}$. These explanatory variables are then standardized so that $X^{1}X$ is in correlation form.

**2.8 Criterion for Investigation and Performance of Proposed Estimator.**

Evaluation, examination and comparison of the estimators were done based on the mean squares error (MSE) property of the estimators

$MSE\left(\hat{α}\right)=\frac{1}{1000}\sum\_{I=1}^{1000}\left(\hat{α}-α\right)^{1}(\hat{α}-α)$ (22)

Where $\hat{α}$ is the estimates of any of the estimators being studied. The estimator with the smallest estimated MSE is considered best. The statistical package R (4.5.0) was used to write the program that accommodated the estimators under consideration. At a particular level of error variance, multicollinearity and sample size, R gave MSE values. Statistical Package for the Social Sciences (SPSS 29.0) was further used to rank the estimators based on their MSE values.

**3.0 RESULTS & DISCUSSION**

Tables 1 and 2 summarize the simulation results using the Mean Square Error (MSE) criterion for different sample sizes, multicollinearity levels, and number of predictors, specifically when p=4 and p=7, respectively.

In Table 1, the frequency of the ridge liu Estimators with and without Principal Component with Minimum MSE Over the different Levels of Multicollinearity and Error Variance and at each Sample Size, when p=4 is displayed. From Table 1 the following can be observed:

when p=4, RHKMALUMIWPC produces the highest number of times MSE is minimum and thus is the most frequent efficient estimator. This is followed by RFAMALUMIWOPC**.** Also, In Table 2, the Frequency of the ridge liu Estimators with and without Principal Component with Minimum MSE Over the different Levels of Multicollinearity and Error Variance and at each Sample Size, when p=7 is shown. From Table 2 the following can be observed:

when p=7, RFAMALUMIWOPC produces the highest number of times MSE is minimum and thus is the most frequent efficient estimator. This is followed by RHKMALUMIWPC**.**

When the sample size is small (10-30) and p=4, RHKMALUMIWPC is the most frequent efficient estimator. However, the frequency of its efficiency decreases as the sample size increases. When the sample size is moderate (40-50), RFAMALUMIWOPC is the most frequent efficient estimator. As the sample size becomes large (100), RFAMILUHMWOPC, RHKGNWOPC, and RHKMDLUHMWOPC become the most frequent efficient estimators. For very large sample size (250), RFAHMLUHMWOPC, RFAMALUMIWOPC, RFAMILUMIWOPC are the most frequent efficient estimators.

When p=7, and the sample size is small (10-20) RHKMALUMIWPC is the most frequent efficient estimator. However, the frequency of its efficiency decreases as the sample size increases. For sample sizes greater than 50, RFAMALUMIWOPC is the most frequent efficient estimator.

In conclusion, the most efficient estimator when multicollinearity is present and p=4 is RHKMALUMIWPC. It is followed by RFAMALUMIWOPC and RFAMALUMIWPC respectively.

When p=7 the most frequent estimator with minimum MSE over the levels of multicollinearity and error variance is RFAMALUMIWOPC. It is followed by RHKMALUMIWPC, RFAMALUMIWPC and RFAGMLUHMWOPC respectively.

The most frequent estimators are pictorially presented in figures 1 and 2. In Figure 1, the Frequency of the best estimators under MSE criterion at different sample sizes when there is Multicollinearity and p=4 is shown. Also, in Figure 2 the Frequency of the best estimators under MSE criterion at different sample sizes when there is multicollinearity and p=4 is shown.

Tables 1 and 2 summarize the simulation results using the Mean Square Error (MSE) criterion for different sample sizes, multicollinearity levels, and number of predictors, specifically when p=4 and p=7, respectively.

In Table 1, the frequency of the ridge liu Estimators with and without Principal Component with Minimum MSE Over the different Levels of Multicollinearity and Error Variance and at each Sample Size, when p=4 is displayed. From Table 1 the following can be observed:

when p=4, RHKMALUMIWPC produces the highest number of times MSE is minimum and thus is the most frequent efficient estimator. This is followed by RFAMALUMIWOPC**.** Also, In Table 2, the Frequency of the ridge liu Estimators with and without Principal Component with Minimum MSE Over the different Levels of Multicollinearity and Error Variance and at each Sample Size, when p=7 is shown. From Table 2 the following can be observed:

when p=7, RFAMALUMIWOPC produces the highest number of times MSE is minimum and thus is the most frequent efficient estimator. This is followed by RHKMALUMIWPC**.**

When the sample size is small (10-30) and p=4, RHKMALUMIWPC is the most frequent efficient estimator. However, the frequency of its efficiency decreases as the sample size increases. When the sample size is moderate (40-50), RFAMALUMIWOPC is the most frequent efficient estimator. As the sample size becomes large (100), RFAMILUHMWOPC, RHKGNWOPC, and RHKMDLUHMWOPC become the most frequent efficient estimators. For very large sample size (250), RFAHMLUHMWOPC, RFAMALUMIWOPC, RFAMILUMIWOPC are the most frequent efficient estimators.

When p=7, and the sample size is small (10-20) RHKMALUMIWPC is the most frequent efficient estimator. However, the frequency of its efficiency decreases as the sample size increases. For sample sizes greater than 50, RFAMALUMIWOPC is the most frequent efficient estimator.

In conclusion, the most efficient estimator when multicollinearity is present and p=4 is RHKMALUMIWPC. It is followed by RFAMALUMIWOPC and RFAMALUMIWPC respectively.

When p=7 the most frequent estimator with minimum MSE over the levels of multicollinearity and error variance is RFAMALUMIWOPC. It is followed by RHKMALUMIWPC, RFAMALUMIWPC and RFAGMLUHMWOPC respectively.

The most frequent estimators are pictorially presented in figures 1 and 2. In Figure 1, the Frequency of the best estimators under MSE criterion at different sample sizes when there is Multicollinearity and p=4 is shown. Also, in Figure 2 the Frequency of the best estimators under MSE criterion at different sample sizes when there is multicollinearity and p=4 is shown.

**TABLES & FIGURES**

**TABLE 1:** Frequency of the Ridge LIU Estimators with and Without Principal Component with Minimum MSE Over the Levels of Multicollinearity and Error Variance at Each Sample Size When There is Multicollinearity for p=4

|  |  |
| --- | --- |
|   | SAMPLE SIZES |
| ESTIMATOR | 10 | 15 | 20 | 30 | 40 | 50 | 100 | 250 | Total | RANK |
| RHKMILUMRWPC | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 27.5 |
| **RHKMALUMIWPC** | **7** | **4** | **4** | **2** | **0** | **0** | **0** | **0** | **17** | **1** |
| RHKMALUMIWOPC | 0 | 0 | 1 | 0 | 2 | 2 | 1 | 0 | 6 | 8 |
| RHKHMWOPC | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 2 | 16 |
| **RHKHMLUMDWOPC** | **0** | **1** | **1** | **1** | **1** | **1** | **2** | **1** | **8** | **5** |
| RHKHMLUHMWOPC | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 27.5 |
| RHKGNWOPC | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 3 | 11 |
| RHKGMWOPC | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 27.5 |
| RHKGMLUHMWPC | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 27.5 |
| RHKGMLUHMWOPC | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 27.5 |
| RHKFAWPC | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 27.5 |
| RHKAMLUGNWOPC | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 27.5 |
| RHKAMLUAMWPC | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 27.5 |
| RHKAMLUAMWOPC | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 2 | 16 |
| RFAMIWOPC | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 2 | 16 |
| RFAMILUMRWPC | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 16 |
| RFAMILUMRWOPC | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 3 | 11 |
| RFAMILUMIWPC | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 27.5 |
| RFAMILUMIWOPC | 0 | 1 | 1 | 0 | 1 | 2 | 0 | 2 | 7 | 7 |
| RFAMDLUMIWOPC | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 27.5 |
| **RFAMALUMIWPC** | **2** | **3** | **2** | **2** | **0** | **0** | **0** | **0** | **9** | **3** |
| **RFAMALUMIWOPC** | **2** | **1** | **0** | **1** | **3** | **2** | **1** | **2** | **12** | **2** |
| RFAHMWOPC | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 27.5 |
| RFAHMLUMIWPC | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 27.5 |
| RFAHMLUMIWOPC | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 3 | 11 |
| **RFAHMLUMDWOPC** | **1** | **2** | **1** | **0** | **1** | **1** | **1** | **1** | **8** | **5** |
| **RFAHMLUHMWOPC** | **1** | **0** | **1** | **1** | **1** | **1** | **1** | **2** | **8** | **5** |
| RFAGNWOPC | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 27.5 |
| RFAGMWOPC | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 2 | 16 |
| RFAGMLUMIWOPC | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 2 | 16 |
| RFAGMLUHMWPC | 0 | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 4 | 9 |
| RFAAMWOPC | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 27.5 |
| RFAAMLUMIWOPC | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 2 | 16 |
| RFAAMLUHMWOPC | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 27.5 |
| LUMIWPC | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 27.5 |
|  TOTAL | **15** | **15** | **14** | **15** | **15** | **15** | **14** | **15** | **118** |  |

**NOTE : Estimator with highest frequency is bolded.**

**Figure 1:** Graphical Representation of the Frequency of the Best Estimators Under Mean Square Error Criterion at Different Sample Sizes When There is Multicollinearity and p=4

**Table 2:**Frequency of the Ridge LIU Estimators With and Without Principal Component With Minimum MSE Over the Levels of Multicollinearity and Error Variance at Each Sample Size When There is Multicollinearity for p=7

|  |  |  |
| --- | --- | --- |
|  | Sample Sizes |  |
| Estimator | 10 | 15 | 20 | 30 | 40 | 50 | 100 | 250 | Total | Rank |
| RHKMIWPC | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 16.5 | **14** |
| RHKMILUMRWOPC | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 16.5 | 20 |
| RHKMILUMIWPC | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 26.5 | 32 |
| **RHKMALUMIWPC** | **4** | **5** | **4** | **1** | **0** | **0** | **0** | **0** | **14** | **2** |
| RHKMALUMIWOPC | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 16.5 | 19 |
| RHKGNWOPC | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 26.5 | **30** |
| RHKGMWOPC | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 26.5 | 31 |
| RHKGMLUHMWPC | 0 | 0 | 0 | 2 | 0 | 2 | 1 | 0 | 5 | 6 |
| RHKGMLUHMWOPC | 0 | 1 | 2 | 1 | 1 | 0 | 0 | 0 | 5 | **6** |
| RHKAMLUAMWPC | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 16.5 | **18** |
| RHKAMLUAMWOPC | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 16.5 | **17** |
| RFAMIWOPC | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 26.5 | **29** |
| RFAMILUMRWPC | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 26.5 | 27 |
| RFAMILUMRWOPC | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 26.5 | **26** |
| RFAMILUMIWOPC | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 16.5 | 16 |
| RFAMALUMIWPC | 4 | 2 | 1 | 3 | 1 | 2 | 0 | 0 | 13 | 3 |
| **RFAMALUMIWOPC** | **4** | **1** | **1** | **1** | **2** | **2** | **2** | **3** | **16** | **1** |
| RFAHMLUMIWPC | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 16.5 | 15 |
| RFAHMLUMIWOPC | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 26.5 | **25** |
| RFAHMLUHMWPC | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 10.5 | 11 |
| RFAHMLUHMWOPC | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 4 | 8 |
| RFAGNWPC | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 16.5 | **13** |
| RFAGNWOPC | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 26.5 | 28 |
| RFAGMWPC | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 26.5 | **21** |
| RFAGMLUMIWPC | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 10.5 | 12 |
| RFAGMLUMIWOPC | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 1 | 5 | **6** |
| RFAGMLUHMWPC | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 10.5 | **10** |
| RFAGMLUHMWOPC | 0 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 10 | 4 |
| RFAAMWPC | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 10.5 | **9** |
| RFAAMLUAMWPC | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 26.5 | 24 |
| RFAAMLUAMWOPC | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 26.5 | 23 |
| LUMIWPC | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 26.5 | **22** |
|  TOTAL | **14** | **15** | **13** | **13** | **13** | **15** | **14** | **15** | **112** |  |

**NOTE : Estimator with highest frequency is bolded.**

**Figure 2:** Graphical Representation of the Frequency of the Best Estimators Under Mean Square Error Criterion at Different Sample Sizes When There is Multicollinearity and p=7

**5.0 CONCLUSION**

The study has proposed an estimator for the estimation of the parameters of a linear regression model with multicollinearity problem. The proposed estimator RHKMALUMIWPC performs better and more efficient than the existing estimators to handle the problem of multicollinearity when p=4. This study is based on simulated data, in future the developed estimator can be applied to real-world datasets from various fields to help validate its practical utility and effectiveness. It is recommended that in linear regression models with moderate multicollinearity (p = 4) the RHKMALUMIWPC estimator should be used in parameter estimation, as it outperforms traditional methods in terms of efficiency and accuracy

**Disclaimer (Artificial intelligence)**

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc) and text-to-image generators have been used during the writing or editing of manuscript.

**COMPETING INTERESTS DISCLAIMER:**

Authors have declared that they have no known competing financial interests OR non-financial interests OR personal relationships that could have appeared to influence the work reported in this paper.

**REFERENCES**

[1] Gujarati, D. N. (2021). Essentials of econometrics. Sage Publications.

[2] Alreshidi, N. A., Alrasheedi, M. A., Lukman, A. F., Alrweili, H., & Farghali, R. A. (2025). Modified Kibria–Lukman estimator for the Conway–Maxwell–Poisson regression model: Simulation and application. Mathematics, 13(5), 794.

[3] Paul, R. K. (2006). Multicollinearity: Causes, effects and remedies. IASRI, New Delhi, 1(1), 58–65.

[4] Alharthi, M. F., & Akhtar, N. (2025). Newly improved two-parameter ridge estimators: A better approach for mitigating multicollinearity in regression analysis. Axioms, 14(3), 186.

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

[6] Liu, K. (1993). A new class of biased estimate in linear regression. Communications in Statistics - Theory and Methods, 22(2), 393–402.

[7] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24(6), 417–441. https://doi.org/10.1037/h0071325

[8] Oduntan, E. (2024). A Monte Carlo study of the residuals of system estimators in the presence of multicollinearity. EKOIST Journal of Econometrics and Statistics, (41), 13–21.

[9[ Chatterjee, S., & Hadi, A. S. (2015). Regression analysis by example. John Wiley & Sons.

[10] Lukman, A. F., Haadi, A., Ayinde, K., Onate, C. A., Gbadamosi, B., & Oladejo, N. K. (2018). Some improved generalized ridge estimators and their comparison. WSEAS Transactions on Mathematics, 17, 369–376.

[11] Fayose, T. S., & Ayinde, K. (2019). Different forms biasing parameter for generalized ridge regression estimator. International Journal of Computer Applications, 181(37), 2–29.

[12] Pongpiachan, S., Wang, Q., Apiratikul, R., Tipmanee, D., Li, L., Xing, L., ... & Poshyachinda, S. (2024). Combined use of principal component analysis/multiple linear regression analysis and artificial neural network to assess the impact of meteorological parameters on fluctuation of selected PM2.5-bound elements. PLOS ONE, 19(3), e0287187.

[13] Weeraratne, N., Hunt, L., & Kurz, J. (2024). Challenges of principal component analysis in high-dimensional settings when n < p. Preprint. https://doi.org/10.21203/rs.3.rs-4033858/v1

[14] Ayinde, K., Apata, E. O., & Alaba, O. O. (2012). Estimators of linear regression model and prediction under some assumptions violation. https://doi.org/10.4236/ojs.2012.25069

[15] McDonald, G. C., & Galarneau, D. I. (1975). A Monte Carlo evaluation of some ridge-type estimators. Journal of the American Statistical Association, 70(350), 407–416. https://doi.org/10.1080/01621459.1975.10479882

[16] Wichern, D. W., & Churchill, G. A. (1978). A comparison of ridge estimators. Technometrics, 20(3), 301–311. https://doi.org/10.1080/00401706.1978.10489675

[17] Gibbons, D. G. (1981). A simulation study of some ridge estimators. Journal of the American Statistical Association, 76(373), 131–139.

[18] 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

[19] Dorugade, A. V., & Kashid, D. N. (2010). Alternative method for choosing ridge parameter for regression. Applied Mathematical Sciences, 4(9), 447–456.

[20] Dorugade, A. V. (2016). Adjusted ridge estimator and comparison with Kibria’s method in linear regression. Journal of the Association of Arab Universities for Basic and Applied Sciences, 21, 96–102.