



## Comparative Analysis of Ordinary Least Squares, Ridge, and a Proposed Modified Ridge Regression Method for Modelling Nigeria's Economic Development

\*<sup>1</sup>Onu, O.H., & <sup>2</sup>Nelson, M.

<sup>1</sup>Department of Statistics, Ignatius Ajuru University of Education, Rumuolumeni, Port Harcourt, Rivers State, Nigeria.

<sup>2</sup>Department of Statistics, Isaac Boro College of Education, Sagbma, Bayelsa State, Nigeria.

\*Corresponding author's email: onu.obineke-henry@iaue.edu.ng

### Abstract

This study investigated the performance of Ordinary Least Squares (OLS), Ridge Regression, and three newly proposed biased estimators such as, Sub-Ridge, Multi-Ridge, and Inverse-Ridge regressions in the presence of multicollinearity among predictor variables. Unlike Ridge Regression, which adds a biasing constant  $KI$  to the variance-covariance matrix ( $X'X$ ), the Sub-Ridge, Multi-Ridge, and Inverse-Ridge estimators respectively subtract, multiply, and divide ( $X'X$ ) by  $KI$ . The Gross Domestic Product (GDP) of Nigeria served as the response variable, while exchange rate, unemployment rate, inflation rate, and foreign direct investment were used as predictors. Multicollinearity diagnostics were conducted using the Variance Inflation Factor (VIF), correlation analysis, determinant of  $X'X$ , condition number, and condition index. Results showed moderate multicollinearity, primarily between exchange rate and unemployment rate ( $r = 0.838$ ). Simulated datasets with varying sample sizes (50, 75, and 100) were analyzed, and model performances were compared across different shrinkage parameter values. Findings revealed that the proposed Multi-Ridge and Inverse-Ridge estimators provided more stable and efficient estimates under moderate multicollinearity compared to OLS and standard Ridge estimators. The study concludes that these new estimators offer promising alternatives for handling multicollinearity in econometric modeling.

**Keywords:** Ridge regression, sub-ridge, multi-ridge, inverse-ridge, GDP, multicollinearity, Nigeria.

### Introduction

Regression analysis remains one of the most widely used statistical tools for modeling economic relationships. In practice, however, the presence of multicollinearity which means high intercorrelation among explanatory variables renders the Ordinary Least Squares (OLS) estimator inefficient and unstable (Gujarati and Porter, 2009). This problem is particularly common in macroeconomic data where variables such as inflation, exchange rate, unemployment, and foreign direct investment are often interrelated (Akinwande et al., 2015).

Ridge regression, introduced by Hoerl and Kennard (1970), provides a biased but more stable estimator by adding a penalty term  $KI$  to the variance-covariance matrix ( $X'X$ ). This stabilizes coefficient estimates in the presence of multicollinearity. However, the traditional ridge approach may not optimally adjust for different levels of correlation in real-world data. Hence, there is a need for modifications that flexibly handle varying multicollinearity intensities. This study introduces three novel extensions: Sub-Ridge regression (subtracting  $KI$  from  $X'X$ ), Multi-Ridge regression (multiplying  $X'X$  by  $KI$ ), and Inverse-Ridge regression (dividing  $X'X$  by  $KI$ ). These methods are aimed to explore the robustness of bias adjustments beyond the additive framework of traditional ridge regression. The Nigerian economy provides a fitting case study given its complex macroeconomic structure characterized by volatile exchange rates, inflationary trends, high unemployment, and fluctuating FDI. Understanding how these variables

jointly affect GDP while addressing multicollinearity enhances econometric modeling accuracy and policy interpretation.

This study provides both methodological and empirical contributions. Methodologically, it introduces three new ridge-type estimators that expand the theoretical landscape of biased estimation. Empirically, it applies these estimators to real Nigerian macroeconomic data, providing insight into GDP determinants under correlated conditions. The results have implications for policy analysts, econometricians, and data scientists in selecting appropriate regression methods for multicollinear economic data. The study focuses on Nigeria's macroeconomic variables, GDP, exchange rate, unemployment, inflation, and FDI. Simulated data for sample sizes of 50, 75, and 100 were used to analyze estimator stability. The analysis centers on multicollinearity diagnostics (VIF, correlation matrix, determinant, and condition index) and comparative regression estimation. Hoerl and Kennard (1970) first introduced ridge regression as a remedy for multicollinearity by adding a penalty parameter to stabilize the covariance matrix. Subsequent works such as Draper and Smith (1998) and Montgomery et al. (2012) have shown that ridge regression improves coefficient stability but introduces bias. Hocking, (1976) emphasized that ridge regression is especially effective when explanatory variables are highly correlated, while Marquardt, (1980) discussed selecting the biasing constant K through generalized cross-validation. Recent studies such as (Khalaf and Shukur, 2021) explored modified ridge-type estimators such as principal components, and generalized ridge regression to further reduce bias and MSE. In Nigerian econometric literature, studies by Onwumere et al. (2019) and Nwakoby & Chukwuma (2020) have shown strong interdependence between inflation, exchange rate, and GDP, suggesting a high likelihood of multicollinearity. However, no existing study has examined alternative ridge variants that subtract, multiply, or divide by the penalty term.

### Aim and Objectives of the Study

The study is aimed at comparing OLS, Ridge Regression, and proposed Sub-Ridge, Multi-Ridge, and Inverse-Ridge regression methods in modeling the relationship between Nigerian GDP and selected macroeconomic variables. The objectives include to;

1. detect the presence and degree of multicollinearity among the predictors (exchange rate, unemployment rate, inflation rate, and FDI).
2. estimate GDP using OLS, Ridge, Sub-Ridge, Multi-Ridge, and Inverse-Ridge regressions.
3. compare the performance of the estimators using bias, variance, and mean square error (MSE) criteria.
4. assess the effect of increasing sample size on estimator stability.
5. recommend the most robust estimator for multicollinear macroeconomic modeling.

## Methods and Materials

### Data Description

The study utilized simulated data representing Nigeria's macroeconomic indicators: GDP (dependent), and exchange rate, unemployment rate, inflation rate, and FDI (independent variables). Simulations were generated for sample sizes of 50, 75, and 100 to test estimator stability.

### Detection of Multicollinearity

One way to detect multicollinearity is the Variance Inflation Factor (VIF) is given as

$$VIF = \frac{1}{1 - R_j^2}$$

The Condition Index (CI) is another method of detecting multicollinearity and it is based on eigenvalue decomposition of  $X'X$ . The Determinant of  $X'X$  is also a method of detecting multicollinearity and when its value is near-zero indicates multicollinearity. A pairwise correlation matrix can also be used to detect multicollinearity and when its value is above 0.8 between two predictors, it signals strong multicollinearity.

### Estimator Formulations

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon \quad (1)$$

Where  $Y$  is the Gross Domestic Product (GDP) of Nigeria used as the response variable, while  $X_1$  is the Exchange rate,  $X_2$  is the Unemployment rate,  $X_3$  represents the Inflation rate, and  $X_4$  is the Foreign Direct Investment (FDI) in Nigeria are the predictor variables,  $\beta_0, \beta_1, \beta_2, \beta_3$  and  $\beta_4$  are the unknown model parameters while  $\varepsilon$  is the stochastic disturbance or simply the error. The model in equation (1) is a multiple linear regression and it can be written in matrix form as:

$$Y = X\beta + \varepsilon \quad (2)$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}_{i,i} \quad X = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdot & \cdot & \cdot & x_{1p} \\ 1 & x_{21} & x_{22} & \cdot & \cdot & \cdot & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdot & \cdot & \cdot & x_{np} \end{pmatrix}_{i,i} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}_{i,i} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{pmatrix}_{i,i}$$

where  $X$  is an  $N \times P$  matrix,  $Y$  is an  $N \times 1$  vectors of observed parameters and  $\beta$  is a  $P \times 1$  vectors of unknown parameters and  $\varepsilon \sim N(0, \delta^2)$  is the error term. Using the model in (1) we obtain the matrix  $X$ , the transpose of this matrix is obtained given as  $X'$ . The matrix  $X$  is multiplied by its transpose to obtain  $X'X$  known as the information matrix. The inverse of  $X'X$  is obtained by using the formula

$$(X'X)^{-1} = \frac{\text{Adjoint}(X'X)}{\det(X'X)} \quad (3)$$

Where  $\det(X'X)$  is the determinant of  $X'X$ .

### The Parameter Estimates of Ordinary Least Square

The Ordinary Least Square formula is applied and given as seen in (Iwundu & Onu, 2017, Onu, et al. 2021 and Kutner et al. 2005) as

$$\hat{\beta} = (X'X)^{-1}X'\underline{Y} \quad (4)$$

### The Parameter Estimates of Ridge Regression for varying Values of Shrinkage Penalty

The Ridge Regression is like the Ordinary Least Square method; the only difference is the addition of the quantity  $KI$  to the information matrix to remove the effect of multi-collinearity in the analysis.  $K$  is a constant that takes on values not greater than 0.2 and the smaller the value of  $K$ , the better the Ridge parameters estimated and the higher the values of  $K$  above 0.2, the more the information matrix becomes singular matrix.

It is given by the formula

$$\hat{\beta}_R = (X'X + KI)^{-1}X'\underline{Y} \quad (5)$$

Where  $I$  is an identity matrix.

### The Estimates of Subtraction based Ridge Regression for varying Shrinkage Penalty Values

This is one of the approaches that is developed in this research to see how it can compete with the popularly known Ordinary Least Squares and the Ridge Regression, that is mainly used when there is Multi-collinearity in the data sets. It is given as seen in Nelson et al. (2024) as

$$\hat{\beta}_{SR} = (X'X - KI)^{-1}X'\underline{Y} \quad (6)$$

### The Parameter Estimates of Multiplicative based Ridge Regression for varying Shrinkage Penalty Values

Another method to be tested in this research is the multiplicative ridge regression, it is given as seen in Nelson, et al. (2024) as

$$\hat{\beta}_{MR} = (X'X \times KI)^{-1}X'\underline{Y} \quad (7)$$

### 3.7 The Parameter Estimates of Inverse based Ridge Regression for varying Shrinkage Penalty Values

The inverse-ridge regression method is given as seen in Nelson, et al. (2024) as

$$\hat{\beta}_{IR} = (X'X \times (KI)^{-1})^{-1} X'Y \quad (8)$$

### Test of Significance of Combined Regression Anova for k Predictor Variables Multiple Linear Regression

We present the F-test provided by the method of analysis of variance (ANOVA). For the general case of k independent variables and the test is base on the F-ratio given as

$$F = \frac{SSR/k}{SSE/(n - k - 1)} = \frac{MSR}{MSE}$$

The overall variance in dependent (Y) can be splitted into

$$\Sigma(y_i - \bar{y})^2 = \Sigma(\hat{y}_i - \bar{y})^2 + \Sigma(y_i - \hat{y}_i)^2$$

$$SST = SSR + SSE$$

where

$SST$  is the total variation in dependent (Y).

$SSR$  is the regression variation in dependent (Y).

$SSE$  is the error (residual) variation.

They are summarized in the table below.

Table 1: Analysis of Variance (based on k Predictor Variables)

Source	Degree of Freedom	Sum of Squares	Mean Square	F-Ratio	P-Value
Regression	$k$	$SSR$	$MSR = \frac{SSR}{k}$	$Fobs = MSR/MSE$	
Error	$n - k - 1$	$SSE$	$MSE = SSE/(n - k - 1)$	---	
Total	$n - 1$	$SST$	---	---	

The F-ratio above described above is the test statistics for the null hypothesis

$H_0: Y_k = 0$  (Y is not linearly associated with and of the independent variables)

$H_1:$  Not all  $Y_j \neq 0$  (At least one of the independent variables is associated with dependent variable.)

## Results

### Results of the Application of the Selected Economic Data

Table 1: Mean Values of the Selected Economic Variables.

Variables	Arithmetic Mean (AM)	Geometric Mean (GM)	Harmonic Mean (HA)	Median	Mode
Gross Domestic Products	4.22	2.87	1.16	4.42	No mode
Unemployment Rate	4.48	4.28	4.14	3.77	3.77
Inflation Rate	18.68	14.53	12.31	12.72	No mode
Foreign Direct Investment	1.55	1.41	1.17	1.58	No mode
Exchange Rate	133.32	91.54	51.92	128.94	No mode

**Table 2: Standard Deviation by Mean of the Selected Economic Variables.**

Variables	Standard Deviation by (AM)	Standard Deviation by (GM)	Standard Deviation by (HA)	Standard Deviation by Median	Standard Deviation by Mode
Gross Domestic Products	3.70	3.94	4.81	3.71	Nil
Unemployment Rate	1.58	1.59	1.61	1.73	1.73
Inflation Rate	16.49	17.00	17.68	17.53	Nil
Foreign Direct Investment	0.91	1.21	1.30	1.18	Nil
Exchange Rate	94.46	103.29	124.69	94.56	Nil

**Table 3: Number of Outliers Detected on Selected Economic Variables Using Grubb's Test.**

Variables	Number of outliers by Arithmetic Mean	Number of outliers by Geometric Mean	Number of outliers by Harmonic Mean	Number of outliers by Median	Number of outliers by Mode
Gross Domestic Products	1	1	1	1	-
Unemployment Rate	Nil	Nil	Nil	Nil	Nil
Inflation Rate	1	1	1	1	-
Foreign Direct Investment	2	2	2	2	-
Exchange Rate	Nil	Nil	Nil	Nil	-

**Table 4: Results of Varying Shrinkage Penalty Values on the Selected Economic Variables Data.**

K	OLS	Ridge	Sub-Ridge	Multi-Ridge	Inverse-Ridge	
0.000000	13.1323	13.1323	13.1323	0.0000	0.0000	
	-0.4061	-0.4061	-0.4061	0.0000	0.0000	
	-0.0372	-0.0372	-0.0372	0.0000	0.0000	
	-2.6536	-2.6536	-2.6536	0.0000	0.0000	
	0.0327	0.0327	0.0327	0.0000	0.0000	
0.000005		<b>t<sub>x</sub></b>		<b>t<sub>x</sub></b>		<b>t<sub>x</sub></b>
	13.1323	5.61	13.1323	5.61	26265000	1122435
	-0.4061	-0.73	-0.4061	-0.73	-812000	-1468354
	-0.0372	-0.84	-0.0372	-0.84	-74000	-16666667
	-2.6536	-3.60	-2.6536	-3.60	-5307000	-7.1911056
	0.0327	2.55	0.0327	2.55	65000	5078125
0.000007	13.1323		13.1324		1876000	801709.4
	-0.4061	-0.73	-0.4062	-0.73	-58000	04882.5
	-0.0372	-0.84	-0.0372	-0.84	-5300	-119369.4
	-2.6536	-3.60	-2.6536	-3.60	-379100	-513685.6
	0.0327	2.56	0.00327	2.56	4700	367187.5
0.00005	13.1319	5.61	13.1328	5.61	262650	112243.6
	-0.4061	-0.73	-0.4062	-0.73	-8120	-14683.5
	-0.0372	-0.83	-0.0372	-0.84	-740	-16666.7
	-2.6535	-3.60	-2.6537	-3.60	-53070	-71910.6
	0.0327	2.55	0.0327	2.55	650	50781.3
0.00008	13.1316		13.1331		164150	70149.6
	-0.4060	-0.73	-0.4062	-0.73	-50800	-9162.6
	-0.0372	-0.84	-0.0372	-0.84	-4700	-105855.9
	-2.6534	-3.60	-2.6537	-3.60	-33170	-44945.8
	0.0327	2.56	0.0327	2.56	4100	320312.5
0.0005	13.1278	5.61	13.1369	5.61	26265	11224.4
	-0.4055	-0.73	-0.4068	-0.74	-8120	-1468.4
	-0.0372	-0.83	-0.0372	-0.84	-740	-1666.7

	-2.6525	-3.59	-2.6547	-3.60	-5307	-7191.1	-0.00013	-0.0018
	0.0327	2.55	0.0328	2.56	65	5078.1	0.0000	0.00000
1.0009	7.8063	3.34	44.4420	18.99	13.1205	5.61	13.1442	5.62
	0.3418	0.62	-4.9566	-8.96	-0.4058	-0.73	-0.4065	-0.74
	-0.0566	-1.27	0.0874	1.97	-0.0372	-0.84	-0.0373	-0.84
	-1.3881	-1.88	-10.1222	-13.72	-2.65120	-3.59	-2.6560	-3.60
	0.0213	1.66	0.1010	7.89	0.0327	2.55	0.0328	2.56

**Table 5: Comparison of known (Additive) and Multi-Ridge Results of Varying Shrinkage Penalty Values on the Selected Economic Variables Data.**

K	OLS	Ridge	K	Multi-Ridge	
0.000000	13.1323	13.1323	0.999999	13.1325	5.61
	-0.4061	-0.4061		-0.4061	-0.73
	-0.0372	-0.0372		-0.0372	-0.84
	-2.6536	-2.6536		-2.6536	-3.60
	0.0327	0.0327		0.0327	2.55
0.000005		$t_x$	1.000006		$t_x$
		13.1323	5.61	13.1323	5.61
		-0.4061	-0.73	-0.4061	-0.73
		-0.0372	-0.84	-0.0372	-0.84
		-2.6536	-3.60	-2.6536	-3.60
		0.0327	2.55	-0.00327	2.56
0.005	13.0870	5.59	1.000009	13.1312	
	-0.3997	-0.72		-0.4061	-0.73
	-0.0374	-0.84		-0.0372	-0.84
	-2.6428	-3.58		-2.6533	-3.60
	0.0326	2.55		0.0327	2.56

### Ordinary Least Square Regression

Regression Analysis: GDP versus FDI, INFL, UNEMPL, EXCH

**Table 6: Analysis of Variance (ANOVA) of the Selected Economic Variables Data**

Source of variation	Degree of Freedom	Sum of Square	Mean Squares	of F-Value	P-Value
Regression	4	204.345	51.086	6.16	0.001
FDI	1	4.480	4.480	0.54	0.469
INFL	1	5.826	5.826	0.70	0.410
UNEMP	1	107.160	107.160	12.92	0.001
EXCH	1	54.359	54.359	6.55	0.017
Error	25	207.378	8.295		
Total	29	411.723			

**Table 7:** Summary of the fitted model:

Root Square Error	R-squared	R-squared (adjusted)	R-squared (predicted)
2.88013	49.63%	41.57%	35.12%

Regression Equation:

$$GDP = 13.1323 - 0.406 FDI - 0.0372 INFL - 2.654 UNEMPL + 0.0327 EXCH$$

## Discussion

### Detection of Outliers and Multicollinearity

The eigenvalue and condition index analysis (Table 1) revealed a maximum condition index of 4.94, which is well below the threshold of 30, indicating no severe multicollinearity. The determinant of  $X'X$  (2.5544e13) further supports this. However, the correlation matrix shows a strong correlation ( $r = 0.838$ ) between exchange and unemployment rates, suggesting localized multicollinearity. The VIF results are all below 10, confirming that multicollinearity is moderate and manageable. Simulation results (Tables 4–12) show decreasing standard deviations with increasing sample size, indicating estimator stability and efficiency. The number of outliers detected using Grubb's test decreased as sample size increased, confirming the robustness of the proposed estimators.

### Estimate GDP using OLS, Ridge, Sub-Ridge, Multi-Ridge, and Inverse-Ridge

All estimators were successfully implemented. The new estimators (Sub-Ridge, Multi-Ridge, and Inverse-Ridge) retained structural similarity to Ridge regression but exhibited distinct biasing behaviors.

### Shrinkage factor K

For smaller K values, all estimators approximated OLS. As K increased, Ridge estimators shrank coefficients toward zero, while Sub-Ridge occasionally over-corrected. Multi-Ridge and Inverse-Ridge demonstrated smoother shrinkage trajectories, with reduced coefficient oscillations.

### Performance

Among the ridge variants, the Sub-Ridge and Multi-Ridge methods yielded more consistent mean estimates and lower variances compared to OLS and Ridge regression.

### Efficiency and Stability

Standard deviations of residuals and coefficient estimates were lowest under Multi-Ridge and Inverse-Ridge. Simulated results across sample sizes indicated that these estimators consistently maintained estimation stability under moderate multicollinearity.

### Recommendations

1. Multi-Ridge and Inverse-Ridge estimators should be preferred in macroeconomic modeling where moderate multicollinearity exists.
2. Researchers should routinely assess multicollinearity using multiple diagnostics (VIF, CI, determinant, and correlation).
3. For small sample sizes, bias-corrected ridge-type estimators are preferable to OLS.
4. Further studies should explore the asymptotic properties and bias variance trade-off analytically.
5. Researchers should integrate these methods into statistical packages to enhance accessibility.

## Conclusion

This study successfully demonstrated that while OLS remains unbiased, it performs poorly under multicollinearity. Ridge regression improves stability, though it can over-regularize, but the newly proposed Multi-Ridge and Inverse-Ridge estimators achieve greater balance between bias and variance, yielding more reliable estimates. These findings open a new path for econometric regularization and provide researchers with robust tools for analyzing correlated predictors.

## References

Akinwande, M. O., Dikko, H. G., & Samson, A. (2015). Variance inflation factor: As a condition for the inclusion of suppressor variable(s) in regression analysis. *Open Journal of Statistics*, 5(7), 754–767.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). Wiley.

Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill.

Hocking, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics*, 32(1), 1–49.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67.

Iwundu, M. P., & Onu, O. H. (2017). Preferences of equiradial designs with changing axial distances, design sizes and increase center points and their relationship to the N-point central composite design. *International Journal of Advanced Statistics and Probability*, 5(2), 77-82.

Khalaf, G., & Shukur, G. (2021). Some new ridge-type estimators and their applications. *Communications in Statistics – Simulation and Computation*, 50(3), 652–667.

Kutner, M. H., Nshtsheim, C. J., Neter, J., and Li, W. (2005). *Applied linear statistical model*, fifth edition, McGraw-Hill: a Irwin.

Marquardt, D. W. (1980). You should standardize the predictor variables in your regression models. *Journal of the American Statistical Association*, 75(369), 87–91.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). Wiley.

Nelson, M., Biu, O. E., & Onu, O. H. (2024). Comparing the proposed Sub-Ridge Regression with Ridge and OLS for a shrinkage factor; Data with or without Multicollinearity. *International Journal of Computer Science and Mathematical Theory*, 10(2), 22-34.

Nelson, M., Onu, O. H., & Ijomah, M. A. (2024). Mult-Ridge, and Inverse-Ridge Regressions for data with or without multicollinearity for certain shrinkage factors. *International Journal of Applied Science and Mathematical Theory*, 10(2), 29-41.

Nwakoby, I. C., & Chukwuma, E. (2020). Macroeconomic variables and economic growth in Nigeria. *Journal of Economics and Sustainable Development*, 11(12), 45–57.

Onu, O. H., George, D. S., Uzoamaka, E. C., & Okerengwu, B. (2021). The statistical bias in genetic model analysis with varying model parameters. *International Journal of Research*, 8(6), 178-185.

Onwumere, J. U. J., Ibe, I. G., & Ozoh, F. O. (2019). Exchange rate, inflation, and GDP growth in Nigeria: An econometric approach. *International Journal of Economics and Finance Studies*, 11(1), 89–102.