Evaluating Ordinary Least Squares Estimators and Ridge Estimators for Predicting Nigeria's Economic Growth from Agricultural Outputs

Ogunnusi Oluwatobi Nurudeen* & Ajibode IlesanmiAkanbi
Department of Mathematics and Statistics, Federal Polytechnic, Ilaro. P.M.B. 50, Nigeria.
Corresponding Email: oluwatobi.ogunnusi@federalpolyilaro.edu.ng

Introduction

One of the key drivers of Nigeria's economy is the agricultural sector, making it crucial to understand how agricultural outputs contribute to economic growth. Traditional econometric models like Ordinary Least Squares (OLS) have been widely used for economic forecasting. However, OLS is highly sensitive to multicollinearity, a common issue when working with highly correlated predictors such as crop production, fishing, forestry, and livestock. To address this limitation, Ridge Regression (RR), a regularisation technique, was employed to improve model stability and predictive accuracy. This study aims to compare OLS and RR in modelling Nigeria's economic growth using agricultural output data, to identify the more efficient predictive model.

Methodology

This study utilised secondary data obtained from the Central Bank of Nigeria (CBN) Statistical Bulletin, covering the period from 1981 to 2023, every month. The dependent variable was Nigeria's Gross Domestic Product (GDP), while the independent variables included four major agricultural output categories: crop production, fishing, forestry, and livestock. OLS and RR were fitted into the model for evaluation. The OLS regression, under the Gauss-Markov theorem, has the smallest variance among all linear unbiased estimators when errors are homoscedastic and uncorrelated, but it proved inefficient and unstable in large samples when multicollinearity is an issue among the predictor variables, leading to large variances and unreliable predictions. Given a dataset with n observations and p predictors, the linear regression model is presented as:

$$Y = X\beta + \epsilon \qquad (1)$$

Where$Y$ is the $n \times 1$ vector of response variable; X is the $n \times p$ matrix of predictors; β is the $p \times 1$ vector of regression coefficients; and ϵ is the $n \times 1$ vector of error terms.

Using the error sum of squares, the estimator of equation (1) is expressed as:

$$\hat{\beta} = (X^T X)^{-1} X^T y \qquad (2)$$

However, RR applies an L_2 penalty (sum of squared coefficients) to equation (1), reducing the variance caused by collinearity among predictors, making it efficient, but it proved bias in its estimates due to the penalty attached for actualization of the reduced variances. The penalty is estimated in matrix form as:

$$\min_{\beta} ||Y - X\beta||^2 + \lambda ||\beta||^2 \qquad (3)$$

Where: $\lambda \geq 0$ is the regularisation parameter, controlling the strength of the penalty. Larger λ values shrink coefficients more aggressively, reducing variance but increasing bias. The ridge estimator is as deduced in equation (3) as:

$$\hat{\beta}_{ridge} = (X^T X + \lambda I)^{-1} X^T y \qquad (4)$$

$I$is the $p \times p$ identity matrix. The term λI ensures invertibility even when $X^T X$ is singular (i.e., in the presence of multicollinearity). Data analysis was conducted using *R* software version 4.2.1, employing *ggplot2* for data visualisation, *moments* for computing skewness and kurtosis, *haven* for importing and managing datasets, for model performance evaluation, *car* for OLS model fitting development and *glmnet* for implementing Ridge Regression.

Data Analysis and Model Fitting

An initial OLS model was estimated, and the significance of each regressor was assessed using *p*-values. The presence of multicollinearity was examined using the Variance Inflation Factor (VIF), which revealed severe multicollinearity, with values of 154.785 (crop production), 118.468 (fishing), 57.461 (forestry), and 177.546 **(livestock)**. This indicated the need for a regularisation approach.

To mitigate multicollinearity, Ridge Regression was applied, with the tuning parameter (λ) optimised using k-fold cross-validation. Model performance was evaluated using Root Mean Square Error of Prediction (RMSEP)for both training and test datasets.

Findings

The OLS estimates indicated that all four agricultural output variables significantly influenced economic growth ($p < 0.05$**)**. However, due to multicollinearity, OLS produced unstable estimates. Ridge Regression effectively shrank the regression coefficients, reducing prediction error.