

## **DUAL METHODOLOGICAL APPROACH OF RANDOM FOREST REGRESSOR AND ORDINARY LEAST SQUARES IN ASSESSING THE IMPACT OF LABOUR FORCE ON GROSS DOMESTIC PRODUCT (GDP) IN NIGERIA.**

Authors: Nnah Bright Ahuraezemma, Okereke-Jude, E.O, Okeudo Justice Ogbonnaya

Department of Statistics, Faculty of physical Sciences, Abia State University, Uturu.

Corresponding Author: Nnah Bright Ahuraezemma

[nnahbright46@gmail.com](mailto:nnahbright46@gmail.com), 08108775265

### **ABSTRACT**

This study examines the impact of total labour force on Gross Domestic Product (GDP) in Nigeria using a comparative methodological framework that integrates Ordinary Least Squares (OLS) regression and Random Forest regression. While classical economic theory assumes a linear relationship between labour and output, emerging data science techniques suggest that macroeconomic relationships may exhibit nonlinear dynamics. The objective of this study is to determine whether the labour to GDP relationship in Nigeria is predominantly linear or potentially nonlinear. Annual time series data were analyzed using OLS to estimate the linear effect of labour force on GDP, followed by Random Forest regression to capture possible nonlinear structures and improve predictive performance. The OLS results reveal a positive and statistically significant relationship between labour force and GDP, explaining approximately 79.5% of the variation in economic output. However, the Random Forest model outperformed OLS, explaining 87.3% of GDP variation and significantly reducing prediction errors (RMSE, MSE, and MAE). Based on comparative model performance, the findings suggest that although labour force significantly influences GDP, the relationship is not strictly linear but potentially nonlinear and structurally complex. The study concludes that integrating machine learning techniques with traditional econometric methods enhances macroeconomic analysis and policy interpretation. Policy implications emphasize labour productivity, human capital development, and data-driven economic planning. This research contributes to the growing literature bridging applied statistics, machine learning, and economic development analysis in developing economics.

**Keywords:** Labour force, GDP, Regression, OLS, Random forest, Machine learning

### **Introduction**

Labour force dynamics constitute a fundamental pillar of economic growth, particularly in developing economies where human capital remains a primary factor of production. Classical and endogenous growth theories emphasize labour as a key input that enhances output through

productivity improvements, innovation, and capital accumulation Romer (1990); Solow (1956). The size, participation, and effectiveness of the labour force directly influence aggregate output by determining both the scale and efficiency of productive activities. For countries with rapidly expanding populations such as Nigeria, labour force development represents both a potential demographic dividend and a critical economic challenge.

Nigeria has one of the largest labour forces in Africa; however, economic growth has remained volatile, with persistently high unemployment and underemployment rates. Despite periods of GDP expansion, labour absorption has been weak, suggesting that increases in labour supply have not translated proportionately into productive output National Bureau of Statistics (2023). This phenomenon indicates that the relationship between labour force indicators and GDP in Nigeria may be nonlinear and influenced by interacting structural factors, including informality, skill mismatches, demographic pressures, and macroeconomic instability.

Empirical investigations into the labour force to GDP nexus in Nigeria have predominantly employed conventional econometric techniques such as Ordinary Least Squares (OLS), vector autoregression (VAR), and cointegration methods Akeju & Olanipekun (2014); Onifade et al. (2020). While these approaches offer valuable insights into long-run relationships and causal links, they rely on restrictive assumptions of linearity, parameter stability, and normally distributed errors. In the Nigerian context characterized by structural breaks, policy shocks, and labour market rigidities these assumptions may be violated, potentially leading to biased or incomplete inferences.

Recent advances in machine learning provide alternative analytical tools capable of addressing these methodological limitations. Random Forest (RF), an ensemble learning algorithm introduced by Breiman (2001), is particularly effective in modeling complex nonlinear relationships and high-order interactions without imposing strict parametric assumptions. In economic and labour market research, RF has demonstrated superior predictive performance relative to traditional econometric models, especially in environments marked by structural complexity and data irregularities Athey and Imbens (2019); Mullainathan and Spiess (2017). Furthermore, RF produces variable importance measures that enhance interpretability and support evidence-based policy formulation.

Despite the increasing application of machine learning techniques in economic analysis, their use in examining labour force–GDP relationships in Nigeria remains limited. Existing studies largely prioritize inferential analysis within linear frameworks, with little emphasis on predictive accuracy or the identification of nonlinear effects. More critically, there is limited empirical evidence on the relative importance of total labour force in explaining variations in Nigeria's GDP.

### **Conceptual Framework of the Study**

The conceptual framework of this study is anchored on the theoretical and empirical proposition that labour force size is a fundamental driver of aggregate economic output. In classical growth theory, labour is identified as a primary factor of production that directly contributes to output expansion by increasing productive capacity Solow (1956). Similarly, endogenous growth theory emphasizes that the scale of labour enhances output through learning-by-doing and productivity spillovers Romer (1990). Within this framework, total labour force is conceptualized as an exogenous input that influences Gross Domestic Product (GDP) through its contribution to production activities across sectors of the economy.

In the Nigerian context, the relevance of total labour force is amplified by rapid population growth and expanding working-age demographics. An increase in the total labour force potentially raises the volume of goods and services produced, thereby increasing GDP. However, the labour force to GDP relationship may not follow a simple linear pattern due to structural characteristics such as informality, labour market rigidities, capital constraints, and technological limitations. These complexities suggest the presence of nonlinear and threshold effects, where the marginal contribution of labour force to GDP varies across different economic conditions.

To capture this complexity, the present study adopts a Random Forest regression framework, which allows the relationship between total labour force and GDP to be modeled as a flexible nonlinear function rather than a fixed linear association. Random Forest aggregates multiple decision trees generated through bootstrap sampling and random feature selection, enabling it to learn complex functional forms without imposing restrictive assumptions Breiman (2001). This approach aligns with recent methodological advancements in economic analysis that advocate

the use of machine learning techniques for modeling structurally complex macroeconomic relationships Athey and Imbens (2019); Mullainathan and Spiess (2017); David Chikezie & Nnah Bright (2025).

Within the conceptual framework, total labour force serves as the independent variable, while GDP represents the dependent variable. The Random Forest model acts as the analytical mechanism through which the nonlinear mapping from labour force to GDP is estimated. The framework also allows for the evaluation of predictive strength by comparing Random Forest outputs with those obtained from Ordinary Least Squares (OLS) model. This comparative structure strengthens the empirical contribution by demonstrating whether nonlinear machine learning methods provide superior explanatory and predictive performance in assessing the labour growth in Nigeria.

Generally, the conceptual framework posits that changes in total labour force exert a direct and potentially nonlinear influence on GDP, and that Random Forest regression provides a more robust tool for uncovering this relationship than traditional linear econometric approaches.

## **Methodology**

### **Research Design**

This study adopts a quantitative ex post facto research design, as it relies on historical macroeconomic data that cannot be manipulated by the researcher. The design is appropriate for examining the relationship between total labour force and Gross Domestic Product (GDP) in Nigeria over time. In addition, a dual modeling approach is employed, combining a machine learning technique (Random Forest regression) with a traditional econometric method (Ordinary Least Squares) to evaluate performance and robustness.

### **Population of the Study**

The population of the study comprises all annual macroeconomic observations on total labour force and Gross Domestic Product (GDP) in Nigeria. These observations reflect the entire

Nigerian economy and cover the full scope of labour force dynamics and economic output within the country

### **Sample and Sampling Technique**

The sample consists of annual time-series data for Nigeria covering the period 1990–2023 (subject to data availability). A census sampling technique is adopted, whereby all available observations within the study period are included. This approach is appropriate given the limited number of macroeconomic observations and ensures comprehensive coverage of labour force to GDP dynamics.

### **Sources of Data**

The study utilizes secondary data obtained from authoritative and internationally recognized source

- **World Bank site, World Development Indicators (WDI)** – Total labour force and GDP.

This source was selected for their reliability, consistency, and widespread use in empirical macroeconomic research World Bank (2023).

### **Method of Data Analysis**

The data analysis is conducted in two main stages:

#### **Random Forest Regression**

Random Forest regression is employed to model the relationship between total labour force and GDP. The algorithm constructs an ensemble of decision trees using bootstrap sampling and aggregates their predictions to improve accuracy and reduce overfitting Breiman (2001). This method enables the capture of nonlinear relationships between total labour force and GDP without imposing restrictive parametric assumptions.

Mathematical Formulation:

$$\hat{Y}(x) = \frac{1}{B} \sum_{b=1}^B f_b(x) \quad (1)$$

for  $b = 1 - B$

Where

- $\hat{Y}(x)$  represents the predicted value of GDP.
- $B$  is the total number of decision trees.
- $f_b(x)$  is the prediction from the  $b$ -th decision tree.

Each tree is trained on a bootstrap sample of the data and at each node, a random subset of predictors is considered for splitting. The number of trees (ntree), the number of predictors randomly selected at each split (mtry), and the maximum depth of trees.

Random Forest is widely used in economic forecasting due to its robustness.

Model performance is evaluated using:

- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Error (MAE)
- Coefficient of Determination ( $R^2$ )

### Ordinary Least Squares (OLS) Regression

For comparison, a conventional OLS regression model is estimated as:

$$GDP_t = \beta_0 + \beta_1 LF_t + \varepsilon_t \quad (2)$$

Where;

$GDP_t$  represents Gross Domestic Product,

$LF_t$  denotes total labour force,

$\varepsilon_t$  is the error term.

OLS results provide a benchmark for assessing whether Random Forest offers superior predictive performance.

## Research Hypothesis

The study tests the following hypothesis:

- **Null Hypothesis ( $H_0$ ):** Total labour force has no significant impact on Gross Domestic Product (GDP) in Nigeria.
- **Alternative Hypothesis ( $H_1$ ):** Total labour force has a significant impact on Gross Domestic Product (GDP) in Nigeria.

## Decision Rule

For the OLS regression:

- Reject the null hypothesis ( $H_0$ ) if the p-value of the labour force coefficient is less than 0.05.
- Fail to reject  $H_0$  if the p-value is greater than or equal to 0.05.

For the Random Forest model:

- Model effectiveness is established if the Random Forest demonstrates lower prediction error (RMSE, MSE, MAE) and higher  $R^2$  compared to the OLS model.

## Decision Criterion

- If Random Forest outperforms OLS in predictive accuracy, the study concludes that total labour force has a meaningful and potentially nonlinear impact on GDP in Nigeria.
- If OLS performs comparably or better, the relationship is interpreted as predominantly linear.
- Policy conclusions are drawn based on the model exhibiting superior empirical performance.

## RESULTS AND DISCUSSION OF FINDINGS

### Introduction

The empirical results of the Dual Methodological analysis between Ordinary Least Squares (OLS) regression and Random Forest regression in assessing the impact of total labour force on Gross Domestic Product (GDP) in Nigeria. The objective is to determine whether the relationship between labour force and GDP is predominantly linear or exhibits nonlinear characteristics.

### Results from Ordinary Least Squares (OLS)

#### Estimated Model

The fitted regression model is:

$$\text{GDP}_t = -450.8 + 1.043 \times 10^{-5} \text{LabourForce}_t + \varepsilon_t \quad (3)$$

#### Interpretation of Coefficients

##### Intercept (-450.8)

- The intercept is  $-450.8$  and statistically significant ( $p < 0.001$ ).
- This represents the predicted GDP when labour force equals zero.
- Although statistically significant, it has limited economic interpretation because a zero labour force is unrealistic.

##### Labour Force Coefficient (1.043e-05)

- The coefficient is  $0.00001043$ .
- It is positive and highly statistically significant ( $p < 0.001$ ).
- $t$ -value =  $10.405$ , indicating strong evidence against the null hypothesis.

**Economic Meaning:**

A one-unit increase in total labour force increases GDP by approximately **0.00001043 units**, holding other factors constant.

Labour force is measured in millions, this implies that an increase of one million persons in the labour force leads to a meaningful increase in GDP.

**Model Goodness-of-Fit**

**R-squared = 0.7945 (79.45%)**

- About 79.45% of the variation in GDP is explained by total labour force.
- This indicates a strong linear relationship.

**Adjusted R-squared = 0.7872**

- Slightly lower than  $R^2$ , confirming model reliability without overfitting.

**F-statistic = 108.3 ( $p < 0.001$ )**

- The model is statistically significant overall.
- Labour force significantly explains GDP variation.

**Hypothesis Testing****Null Hypothesis ( $H_0$ ):**

Labour force has no significant impact on GDP in Nigeria.

**Alternative Hypothesis ( $H_1$ ):**

Labour force has significant impact on GDP in Nigeria.

**Decision:**

Since p-value ( $3.98e-11$ )  $< 0.05$ , **reject  $H_0$** .

**Conclusion:**

Total labour force has a positive and statistically significant linear impact on GDP in Nigeria.

The OLS regression results reveal a positive and statistically significant relationship between total labour force and GDP in Nigeria.

**Key Findings:**

- The labour force coefficient is positive and statistically significant ( $p < 0.001$ ).
- $R^2 = 0.795$ , indicating that approximately 79.5% of the variation in GDP is explained by total labour force.
- The overall model is statistically significant (F-statistic  $p < 0.001$ ).

**Interpretation**

The positive coefficient implies that increases in Nigeria's labour force are associated with increases in GDP. This finding aligns with classical growth theory, where labour is considered a primary factor of production contributing to economic output.

The high  $R^2$  value indicates that labour force plays a major role in explaining GDP fluctuations during the study period. However, while the linear model performs strongly, it assumes a constant marginal effect of labour on GDP.

**Results from Random Forest Regression**

The Random Forest model was estimated to examine whether a nonlinear specification improves predictive performance.

**Performance Metrics Comparison**

| Model         | $R^2$ | RMSE   | MSE      | MAE    |
|---------------|-------|--------|----------|--------|
| OLS           | 0.795 | 262.21 | 68,752.1 | 175.87 |
| Random Forest | 0.873 | 170.40 | 29,036.5 | 129.11 |

## Interpretation

The Random Forest model:

- Explains 87.3% of the variation in GDP, outperforming OLS.
- Produces substantially lower RMSE, MSE, and MAE values.
- Demonstrates superior predictive accuracy.

This indicates that Random Forest captures additional structural patterns in the labour to GDP relationship that are not fully explained by a simple linear model.

## Comparative Discussion

Based on the established decision criterion:

- If Random Forest outperforms OLS → relationship is potentially nonlinear.
- If OLS performs comparably or better → relationship is predominantly linear.

Since Random Forest outperforms OLS across all evaluation metrics, the study concludes that:

The impact of total labour force on GDP in Nigeria is meaningful and potentially nonlinear.

While OLS confirms a strong positive linear association, Random Forest suggests the presence of nonlinearities such as:

- Diminishing or increasing returns to labour
- Structural shifts in labour productivity
- Threshold effects in labour absorption
- Interaction effects between labour supply and economic capacity

This implies that GDP does not increase at a constant rate with labour force expansion.

## Policy Implications

Given that Random Forest exhibits superior empirical performance, policy recommendations are grounded in its findings.

1. Labour force expansion alone may not guarantee proportional GDP growth.
2. Emphasis should be placed on:
  - Labour productivity enhancement
  - Skill development and human capital investment
  - Sectoral diversification
  - Efficient labour market absorption
3. Policymaking should incorporate data-driven and flexible modeling approaches capable of capturing nonlinear economic dynamics.

## Summary of Findings

The empirical evidence shows that:

- Labour force has a positive and statistically significant impact on GDP in Nigeria.
- OLS explains a substantial proportion of GDP variation (79.5%).
- Random Forest improves explanatory power to 87.3% and reduces prediction errors significantly.
- The labour to GDP relationship is therefore not strictly linear but exhibits nonlinear characteristics.

## Conclusion

The findings confirm that total labour force is a critical determinant of economic growth in Nigeria. The OLS results establish a strong positive linear association between labour force and GDP, consistent with classical and neoclassical production theories.

However, the superior predictive performance of the Random Forest model indicates that the relationship extends beyond simple linear assumptions. The labour to GDP nexus likely involves:

- Nonlinear productivity effects
- Structural labour market dynamics
- Threshold effects in labour absorption
- Variations in returns to labour over time

Therefore, economic growth in Nigeria responds to labour force changes in a structurally complex manner rather than at a constant marginal rate.

The integration of machine learning with classical econometrics provides deeper empirical insight and strengthens policy interpretation.

## References

Akeju, K. F., & Olanipekun, D. B. (2014). Unemployment and economic growth in Nigeria. *Journal of Economics and Sustainable Development*, 5(4), 138–144.

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11, 685–725. <https://doi.org/10.1146/annurev-economics-080217-053433>.

Breiman, L. (2001). Random forests. (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.

David C. Chikezie & Nnah Bright Ahuraezemma. (2025). Bridging Ordinary Least Square(OLS) and Machine Learning(Random Forest Regression):An empirical study of fisheries' contribution to GDP in Nigeria. *International journal of business, management, behavioral research and innovations*, 2(1). <https://doi.org/10.5281/zenodo.17923703>.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87–106. <https://doi.org/10.1257/jep.31.2.87>

National Bureau of Statistics. (2023). *Labour force statistics: Unemployment and underemployment report*. National Bureau of Statistics.

Onifade, S. T., Çevik, S., Erdoğan, S., Asongu, S., & Bekun, F. V. (2020). An empirical retrospect of the impacts of government expenditures on economic growth: New evidence from the Nigerian economy. *Journal of Economic Structures*, 9(6), 1–13. <https://doi.org/10.1186/s40008-020-0186-7>

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5, Part 2), S71–S102. <https://doi.org/10.1086/261725>

Solow, R. M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 70(1), 65–94. <https://doi.org/10.2307/1884513>

World Bank. (2023). *World development indicators*. World Bank.

RSS-NLG 2026 Conference Proceedings