Science Forum (Journal of Pure and Applied Sciences)

Journal Homepage: https://atbuscienceforum.com.ng/index.php/jpas

Comparative Analysis of the Classical and Robust Regression Models for Predicting Life Expectancy in Nigeria

Hezekiah David, Nwaosu, S.C and Ikughur, J.A

Department of Statistics, Joseph Sarwuan Tarka University Makurdi, Benue State Nigeria

Abstract

Macroeconomic and demographic datasets from developing economies are frequently affected by extreme observations, structural breaks, heteroskedasticity, and non-normal error distributions arising from economic instability, policy reforms, financial crises, and other exogenous shocks. These characteristics often violate the assumptions underlying the Classical Ordinary Least Squares (OLS) regression model, resulting in biased parameter estimates, inefficient inference, and unreliable policy conclusions. Robust regression techniques provide an effective alternative by reducing the influence of outliers and influential observations, thereby producing more stable and reliable parameter estimates under non-ideal data conditions. This study evaluates the suitability of robust regression for modeling life expectancy in Nigeria by comparing its predictive performance with that of the Classical OLS regression model. Gross Domestic Product (GDP) and mortality rate were employed as explanatory variables, while life expectancy served as the response variable. Both modeling approaches were fitted to the data and their estimated coefficients were examined to assess the consistency and reliability of the relationships between the selected predictors and life expectancy. The empirical results revealed that mortality rate exerted a negative effect on life expectancy in both the Classical OLS and robust regression models, indicating that higher mortality levels are associated with reduced life expectancy. However, notable differences emerged in the estimated impact of GDP. While the OLS model produced a positive but statistically non-significant relationship between GDP and life expectancy, the robust regression model identified GDP as having a positive and statistically significant influence on life expectancy in Nigeria. This disparity suggests that outliers and influential observations may have obscured the true relationship under the OLS framework. Consequently, the findings demonstrate that robust regression provides more reliable estimates in the presence of anomalous observations and should be preferred over OLS when datasets exhibit outlier contamination or violations of classical regression assumptions.

Keywords
Life Expectancy,
GDP,
Mortality Rate,
Robust Regression,
OLS

ISSN 1119-4618
JPAS
Faculty of Science, ATBU Bauchi

Corresponding Author: Ikughur, J.A., Department of Statistics, Joseph Sarwuan Tarka University Makurdi, Benue State Nigeria.
Email: davidhezekiah15@gmail.com, *ikughur.ja@uam.edu.ng, sylvesternwaosu88@gmail.com, ORCID: https://orcid.org/0000-0003-1276-668X

Introduction:

Life expectancy is the expected number of years that a newborn could live considering the prevailing mortality rates and is often employed to provide assessment of the effectiveness of health care and economic policies across countries. This index is an indicator of population health and socio-economic development that reflects the effect of income levels, health care schemes, environmental situations, and demographic factors within an economy (World Health Organization (WHO), 2022).

On the other hand, economic performance commonly proxied by Gross Domestic Product (GDP), is one of the major determinants of health outcomes and longevity. According to health economics theory, higher GDP enhances a country's capacity to invest in health infrastructure, medical technology, education, nutrition, and sanitation, which in turn improves life expectancy (Grossman, 1972; Bloom *et al.*, 2024). Studies from developing economies have shown that sustained economic growth has close association with life longevity, although the magnitude of this relationship depends on income distribution, quality of governance, and public health spending efficiency (Cutler *et al.*, 2018).

Another critical demographic factor is mortality rate which is also seen to directly influence life expectancy. High mortality rates, particularly infant and adult mortality, substantially reduce average life expectancy and often indicate poor access to quality health care, widespread poverty, malnutrition, and exposure to communicable diseases (United Nations Development Programme (UNDP), 2021).

Empirical studies examining the factors affecting life expectancy in Nigeria as well as other developing countries have predominantly relied on conventional estimation techniques such as Ordinary Least Squares (OLS) regression (Eboh and Uche, 2018; Olaniyan and Sunkanmi, 2020). However, macroeconomic and demographic numerical observations obtained from developing economies are often characterized by outliers, structural breaks, heteroskedasticity, and non-normal error distributions arising from economic shocks, epidemics, and policy changes amongst others. These data characteristics can bias OLS estimates and lead to misleading inferences (Rousseeuw and Leroy, 2003).

Robust regression techniques offer a methodological advantage by minimizing the influence of extreme observations and providing more reliable parameter estimates in the presence of violations of classical regression assumptions (Huber and Ronchetti, 2009). To validate his assertion, robust regression model if performed with the view to compare its outcomes with the classical ordinary least squares regression model in attempt to predict life expectancy considering GDP and Mortality rate of Nigeria.

2.0 Literature Review.

In this section, literatures related to the variables of study and their relationships are presented in other to support the proposition in this study.

2.1 Life expectancy, Mortality and Gross Domestic Products

Life expectancy as average number of years a person is expected to live from time of birth. It serves as a critical indicator for assessing the country's progress toward sustainable development and human capital accumulation. It constitutes one of the three core components of the Human Development Index (HDI), alongside education and income, highlighting the recognition that human well-being extends beyond economic output alone (UNDP, 2023).

Gross Domestic Product (GDP) is the most commonly used measure of economic performance and national income. It denotes the total monetary value of all final goods and services produced within a country's geographical boundaries during a specified period, usually one year (Samuelson and Nordhaus, 2010). GDP is often interpreted as an indicator of a country's capacity to generate income and finance public goods and services, including healthcare, education, and infrastructure. Some economic theory advocates that higher GDP levels enable governments and households to invest more in health-related inputs such as hospitals, medical personnel, nutrition, clean water, and sanitation, thereby improving health outcomes and life expectancy (Bloom and Canning, 2000).

Mortality rate is the frequency of deaths occurring in a defined population within a specified period, commonly expressed per 1,000 or 100,000 individuals (WHO, 2023). Mortality is a fundamental demographic process and a direct determinant of life expectancy. Various mortality indicators are used in demographic and health studies, namely; crude death rate, infant mortality rate, under-five mortality rate, adult mortality rate, and maternal mortality ratio. High mortality rates in Nigeria are largely attributable to preventable causes such as malaria, pneumonia, diarrhea, malnutrition, unsafe childbirth practices, and poor sanitation. Socioeconomic factors such as poverty, low female education, rural residence, and limited health infrastructure, further exacerbate mortality risks (World Bank, 2024). Mortality rate therefore serves as both a reflection of health system performance and a determinant of demographic and economic outcomes.

2.2 Interrelationship between GDP, mortality rate, and life expectancy

Life expectancy, GDP, and mortality rate are interconnected through complex feedback mechanisms. Economic growth, as measured by GDP, can reduce mortality rates and increase life expectancy by improving access to healthcare, education, nutrition, and living conditions. Conversely, high mortality rates have the tendency to reduce life expectancy and weaken human capital, thereby limiting labor productivity and economic growth (Bloom, Canning, and Sevilla, 2004).

In Nigeria, the persistence of high mortality rates despite GDP growth suggests that economic expansion alone is insufficient to improve health outcomes. The effectiveness of GDP in enhancing life expectancy depends on the allocation of resources, equity in income distribution, and the efficiency of health systems. Therefore, mortality rate acts as a crucial mediating variable in the relationship between GDP and life expectancy. Understanding these interrelationships provides the conceptual foundation for empirical analyses that model the impact of GDP and mortality rate on life expectancy in Nigeria.

2.3 Health production function theory

The Health Production Function Theory, pioneered by Grossman (1972), conceptualizes health outcomes as a durable form of capital that individuals and societies invest in to enhance productivity, utility, and longevity. According to this theory, health outcomes such as life expectancy are produced through a combination of inputs including income, healthcare services, education, nutrition, environmental quality, and lifestyle choices. Grossman (1972) argues that higher income levels increase both the demand for and supply of health inputs. Empirical extensions of this theory demonstrate that increases in national income are related with improvements in survival rates and reductions in mortality (Bloom and Canning, 2000; Cutler, Deaton, and Lleras-Muney, 2006).

In the Nigerian context, the Health Production Function Theory suggests that low life expectancy is partly a consequence of inadequate investment in health-related inputs. Despite periods of GDP growth, weak public health expenditure, inefficient allocation of resources, and limited access to quality healthcare have constrained the production of health outcomes (World Bank, 2024). Consequently, GDP is theoretically expected to exert a positive influence on life expectancy and a negative influence on mortality rate, although this effect may be weakened by institutional inefficiencies.

2.4 Empirical Review

Bloom and Canning (2000), employing panel data methodologies for both developed and developing countries, showed that higher income levels are related with improved health outcomes and hence, increased life expectancy. The findings in their study showed that GDP growth enhances government capacity to invest in health infrastructure and disease prevention, leading to lower mortality and longer lifespan. Meanwhile, Cutler *et al.* (2006) used international panel data to establish that infant and child mortality account for a significant proportion of gains in life expectancy in developing countries.

Deaton (2003) further argues that income growth alone does not guarantee improved life expectancy unless it translates into effective public health interventions that reduce mortality. This finding is particularly relevant for countries like Nigeria, where GDP growth has not consistently reduced mortality rates.

Adeniyi and Abiodun (2011) examined the association between economic growth and health indicators in Nigeria using statistical method called ordinary least squares (OLS) estimation. The findings in the study showed that GDP per capita has positive influence on life expectancy, but the magnitude of the effect is constrained by income inequality and inefficient public spending while Bakare and Olubokun (2011) examined the relationship between Nigeria's economic growth, mortality, and health outcomes using time-series data. Their results reveal that while GDP growth contributes positively to life expectancy, mortality indicators exert a stronger and more statistically significant influence.

Lawanson (2014) investigated how public health expenditure and economic growth on health outcomes in Nigerian are connected using time-series data. The study found that GDP growth has a weak statistically positive effect on life expectancy, suggesting that economic growth alone is insufficient without effective health sector investment. Meanwhile, Ogunjimi and Adebayo (2017) employ autoregressive distributed lag (ARDL) modeling to assess the effect of health indicators on life expectancy in Nigeria. The findings showed that infant mortality rate has strong negative effect on life expectancy in both the short run and long run.

A study by Akinwale and Adekunle (2020) applies vector error correction modeling (VECM) to assess the long-run relationship among GDP, mortality rate, and life expectancy in Nigeria. The findings confirm a long-run equilibrium relationship, with mortality rate exerting a dominant negative effect on life expectancy, while GDP shows a positive but moderate impact.

In Asia, Rana *et al.* (2018) showed that GDP growth significantly improves life expectancy only when accompanied by reductions in infant mortality and increased public health expenditure. Their panel study reinforces the significance of mortality control as a transmission mechanism between economic growth and longevity.

3.0 Methods

3.1 Source of Data and Analysis

The study made use of secondary data which was obtained from international databases. Annual data on Nigeria's life expectancy and Gross domestic product (GDP) were sourced from the World Development Indicators (WDI) published by the World Bank. Data on Mortality rate was obtained from the World Health Organization (WHO) for the period 1960-2021.

3.2 Model Specification

This section presents classical regression modeling and the Robust Regression (Huber Regression) and M-estimation Technique.

Consider the linear regression model given in Equation (1)

$$y = X'\beta + e \qquad (1)$$

Where X is a vector of independent variables, β is the vector of regression coefficient and e is the vector of the error term and y is the explained variables. The ordinary least square (OLS) method minimizes the sum of squared function as shown in Equation (2)

$$Min \sum \ (y_i - X_i\beta)^2 = Min \sum_{i-1}^{n} e^2 \qquad (2)$$

Where the parameters are given by Equation (3)

$$\beta = (X'X)^{-1}(X'Y) \qquad (3)$$

Whose parameters are expected to be significant.

3.2.1 Huber robust regression

The classical regression assumes linearity, homoscedasticity, uncorrelated errors and free of outliers. However, when the data under consideration appears to be affected by outliers and other forms of noise, classical OLS does not provide accurate estimates. In order to obtain improved accuracy, the robust regression model is most appropriate. Let the model under this condition be given by Equation (4)

$$y_i = X_i{}'\beta + e_j \qquad (4)$$

where

$$\hat{\beta} = \arg Min\ \beta \ \sum_{i=1}^{n} P(r_i) \qquad (5)$$

Where P is the robust loss function as given by Equation (6)

$$P(r) = \begin{cases} \frac{1}{2}r^2 \ if \ |r| \le \delta \\ (|r| - \frac{1}{2}r^2\delta) \ for \ |r| > \delta \end{cases} \qquad (6)$$

To obtain the model's parameters, the M-estimation technique developed by Huber (1964, 1973, 2004) is used. Thus, instead of minimizing the sum of squared residuals (SSR), a robust M-estimator minimizes the sum of the less rapidly increasing function of the residuals as shown by Equation (7)

$$min \sum_{i-1}^{n} p(y_i - \sum X_{ij}\beta_j) = \min \sum_{i=1}^{n} p(e_i) \qquad (7)$$

The solution is not scale equivalent and thus the residuals must be standardized by a robust estimate of their scale $\hat{\sigma}_e$ which is estimated simultaneously. However, for the M-estimates of location, the median absolute deviation (MAD) is often used. Taking the derivative of (13) and solving produces the score function as shown by Equation (8)

$$\sum_{i=1}^{n} \varphi(y_i - \sum \frac{X_{ij}\beta_j}{\hat{\sigma}})X_{ik} = \min \sum_{i=1}^{n} \varphi(\frac{e_i}{\hat{\sigma}_e})\, X_i = 0 \qquad (8)$$

The key assumptions for robust regression include outlier resistance, model form, error distribution and influence of observations.

In standard linear regression, we estimate parameters B by minimizing the sum of squared residuals:

$$\hat{B} = \arg min_{\theta} \sum (y_i - X_i B) \quad (9)$$

This minimization problem leads to the normal equations.

$$(X^T X)\hat{B} = X^T \quad (10)$$

Since squaring residuals amplifies large errors, OLS is very sensitive to outliers. To make regressions robust to outliers, we replace the squared loss with a general function p.

$$\hat{B} = \arg min_{\theta} \sum p(y_i - X_i B) \quad (11)$$

Taking the derivative with respect to $B = \sum \psi(y_i - X_i B) X_i$

where $\psi(u) = \frac{\partial p(u)}{\partial u}$ is the influence function. For certain p functions, the estimations procedure leads to an iteratively reweighted least square (IRLS) approach.

$$\sum w_i (y_i - X_i B) X_i = 0 \quad (12)$$

Where the weight function is:

$$w_i = \frac{\psi(y_i - X_i B)}{y_i - X_i B} \quad (13)$$

At each iteration, we solve a weighted least squares problem:

$$B^{(t+1)} = (X^T W^{(t)} X)^{-1} X^T W^{(t)} y \quad (14)$$

Where $W^{(t)}$ is the diagonal weight matrix at iteration t.

34 Diagnostic test: The following diagnostic test are discussed in this section. These include, Cook's D - test and the studentized residual test.

The Cook's distance test is a diagnostic statistic used in regression analysis to identify influential observations. For the *i-th* observations, the Cook's Distance is defined as shown in Equation (15)

$$D_i = \frac{(\hat{\beta} - \hat{\beta}_i)' (X'X)(\hat{\beta} - \hat{\beta}_i)}{k\hat{\sigma}^2} \quad (15)$$

Where, D_i = Cook's Distance for observation *I*, $\hat{\beta}$ = vector of estimated coefficients using all observations, $\hat{\beta}_i$ = vector of estimated coefficients with *i-th* observations removed, X = matrix of explanatory variables, k = number of parameters in the model (including the intercept) and $\hat{\sigma}^2$ = estimated variance of the error term. An observation is considered influential if $D_i > \frac{4}{n}$ or $D_i > 1$, where, n denotes the sample size.

On the other hand, the Studentised residual test is a standardized form of the ordinary least squares (OLS) residual that accounts for non-constant variance of residuals caused by leverage. It is used to detect outliers in regression analysis. The internally studentized residual for the *i-th* observation is given by Equation (16)

$$r_i = \frac{e_i}{\hat{\sigma}\sqrt{1 - h_{ii}}} \quad (16)$$

Where, r_i = studentized residual for observation I, e_i = OLS residual for observation i, $\hat{\sigma}$ = estimated standard deviation of the error term and h_{ii} = leverage of the *i-th* observation (diagonal element of the hat matrix). The Externally studentized residual (deleted residual) is defined as shown by Equation (17)

$$t_i = \frac{e_i}{\hat{\sigma}_{ii}\sqrt{1-h_{ii}}} \qquad (17)$$

Where, $\hat{\sigma}_i$ = estimated standard deviation of the error term computing after deleting the *i-th* observation. This form follows a t-distribution with $n-k-1$ degree of freedom and is preferred for formal outlier testing. An observation is considered a potential outlier if $|r_i| > 2 \ or \ |r_i| > 3$. More strictly for the externally studentized residual $|t_i| > t_{\frac{\alpha}{2}, n-k-1}$

Results

The trend analysis of gross domestic product, mortality rate and life expectancy are presented in Figure 1-3. Diagnostic test of the present of outliers is shown on Table 1, Classical Ordinary Least Square Regression model and the Robust Regression Model of Mortality rate (MR) and Gross Domestic Product (GDP) on Life Expectancy (LE) in Nigeria are presented in Tables 2 and 3 respectively.

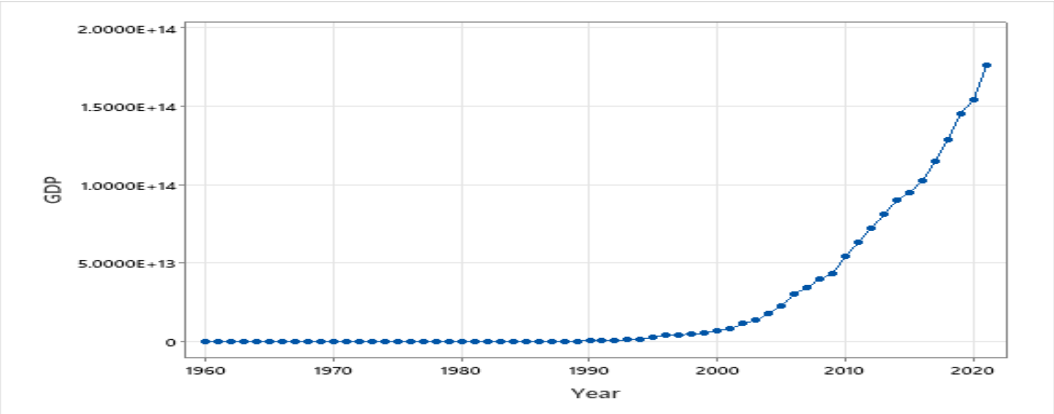

Figure 1: Trend of GDP variable for the period 1960-2021

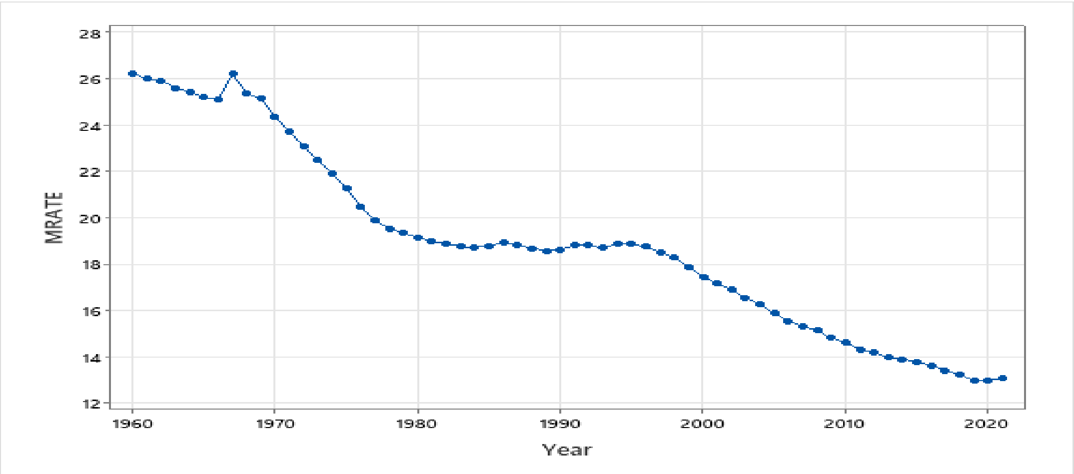

Figure 2: Trend of Mortality Rate (MRATE) variable for the period 1960-2021

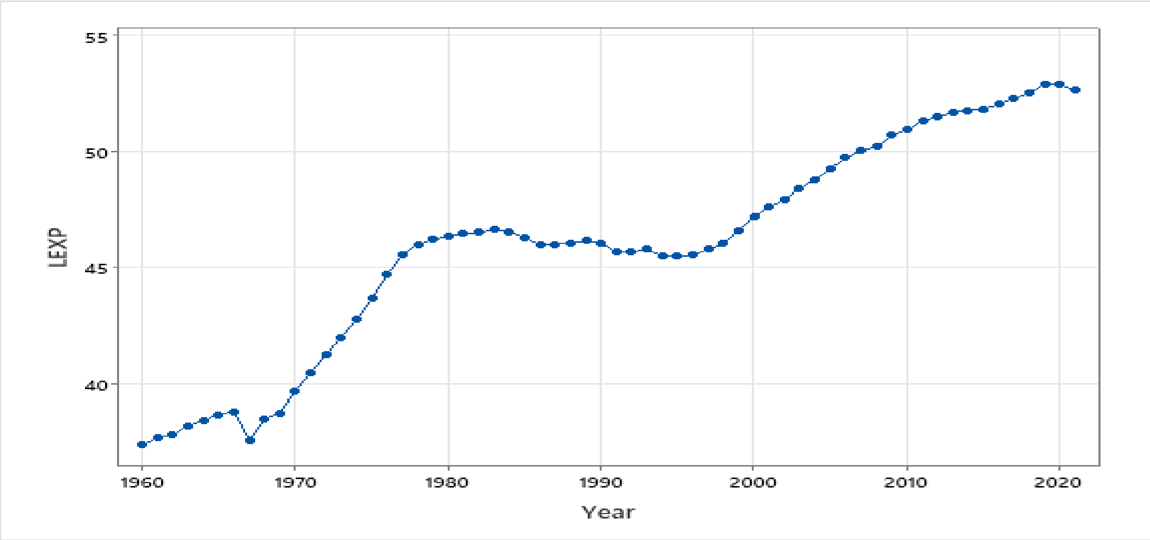

Figure 3: Trend of Life Expectancy (LEXP) variable for the period 1960-2021

Table 1: Diagnostic Test

Test	|r|	D.	Statistics	p-value
R Square			0.9689	
Studentised Residual (r)	2.702054			
Cook's d		0.552337		
F-statistics			6799.36	0.0000

Table 2: Ordinary Least Square Regression Estimates of Mortality rate (MR) and Gross Domestic Product (GDP) on Life Expectancy (LE) in Nigeria

Variables	Coefficient	Std. Error	t.	p-value
LMR	-1.139931	.0156792	-73.07	0.000
LGDP	1.10e-15	1.40e-15	0.78	0.436
C	67.48835	0.3249132	207.71	0.000

LMR = Log Mortality rate, LGDP = Log Gross domestic product Prob > F = 0.0000, R-squared = 0.9945, Adj R-squared = 0.9943 and F(2, 59) = 5341.95

Table 3: Robust Regression Model of Mortality rate (MR) and Gross Domestic Product (GDP) on Life Expectancy (LE) in Nigeria

Variables	Coefficient	Std. Error	t.	p-value
LMR	-1.132024	.0139435	--81.19	0.000
LGDP	2.57e-15	1.25e-15	2.06	0.044
C	67.22864	0.2889464	232.67	0.000

LMR = Log Mortality rate, LGDP = Log Gross domestic product, C = constant, F(2, 59) = 6799.36 and Prob > F = 0.0000

Discussion of Results

5.1 Results of Descriptive Statistics

Figures 1-3 presents the time plot of the variables of interest. It is noticeable that the GDP of Nigeria appeared stable for 61 years and began to increase at faster rate.

Diagnostic Test Results: The diagnostic test of the robust regression model reported on Table 1 shows the R-square 0.9689 indicated that 96.89% of the variability in the life expectancy (LEXP) is explained by Mortality Rate (LMR) and Gross Domestic Product (LGDP) signifying a good fit. The F-statistic of 6799.36 with a p-value of 0.0000 confirms the overall significance of the robust regression model as the null hypothesis that all regression coefficients are equal to zero is rejected. The studentized residual (2.702054) and cook's d (0.552337) revealed the existence of outliers as the absolute value of the studentized residual is greater than 2 even as Cook's d is greater than 4/N.

Results of Ordinary Least Square Regression: The classical regression model on Table 2 shows that only LMR with negative coefficient of 1.1399 is significant since the p-value is less than 0.05 while LGDP coefficient with 1.10e-15 is not significant as p-value is greater than 0.05 level of significance. The OLS regression model is also significant as p-value of the F statistic = 0.0000 < 0.0000 level of significance and with R^2 of 99.43% of the variations.

Results of Robust Regression Analysis: The robust regression model coefficients on Table 3 showed LMR and LGDP were all significant in the model since the p-value are less than 0.05 significant level thereby rejection of the null hypothesis.

The present of outliers in the dataset suppress the impact of LGDP in the OLS regression model making its coefficient to be insignificant. This challenge is overcome by the use of robust regression model whose coefficient are all significant without loss of precision.

The findings in this study when OLS procedure is used contradicts the findings by Adeniyi and Abiodun (2011) while agreeing with with the findings of Ogunjimi and Adebayo 2017) and Akinwale and Adekunle (2020) which sowed that mortality has negative effect of life expectancy. Findings from the robust regression model's outcome are in line with findings of Deaton (2003) which posited that economic growth accompanied by reduced could among others, improve life expectancy.

Conclusion and Recommendation

This study investigated the impact of mortality rate and GDP on economic growth in Nigeria spanning the period from 1960-2021 using classical OLS regression and robust regression analysis.

i. Mortality rate had negative effect on life expectancy in both the classical and robust regression models, it is necessary for the Nigerian government to concentrate on reducing mortality rate
ii. For the robust regression, GDP showed a significant positive impact on life expectancy in Nigeria while the classical OLS showed a non-significant coefficient.
iii. Following the outcomes above, it can be concluded that where outliers exist in dataset, robust regression would be more appropriate under this scenario.

REFERENCES

Adeniyi, O. M., and Abiodun, L. N. (2011). Health Expenditure and Economic Growth in Nigeria. *Journal of Economics and Sustainable Development*, **2**(4): 38-47.

Akinwale, Y. O., and Adekunle, I. A. (2020). Economic Growth, Mortality and Life Expectancy in Nigeria: A VECM approach. *Journal of Health Economics and Development*, **5**(2): 55-71.

Bakare, A. S., and Olubokun, S. (2011). Health Care Expenditure and Economic Growth in Nigeria: An Empirical Study. *Journal of Emerging Trends in Economics and Management Sciences*, **2**(2): 83-87.

Bloom, D. E., and Canning, D. (2000). The Health and Wealth of Nations. *Science*, **287**(5456): 1207-1209.

Bloom, D. E., Canning, D., and Sevilla, J. (2024). The Effect of Health on Economic Growth: A Production Function Approach. *World Development*, **32**(1): 1-13.

Cutler, D. M., Deaton, A., and Lleras-Muney, A. (2018). The Determinants of Mortality. *Journal of Economic Perspectives*, **20**(3): 97-120.

Deaton, A. (2018). Health, Inequality, and Economic Development. *Journal of Economic Literature*, **41**(1): 113-158.

Eboh, E. C., and Uche, C. U. (2018). Health Outcomes and Economic Growth in Nigeria. *African Development Review*, **30**(2): 210-223.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, **80**(2): 223-255.

Huber, P. J., and Ronchetti, E. M. (2009). Robust Statistics (2nd ed.). Hoboken, NJ: Wiley.

Lawanson, A. O. (2014). Health Care Financing in Nigeria: Implications for Life Expectancy. *African Journal of Economic Policy*, **21**(1): 1-22.

Olaniyan, O., and Sunkanmi, O. A. (2020). Determinants of Life Expectancy in Nigeria. International *Journal of Social Economics*, **47**(3): 345-360.

Rana, R. H., Alam, K., and Gow, J. (2018). Health Expenditure, Income and Life Expectancy: Evidence from Asia. *Economic Analysis and Policy*, **60**, 45-58.
Rousseeuw, P. J., and Leroy, A. M. (2003). Robust Regression and Outlier Detection. Hoboken, NJ: Wiley, 684.

Samuelson, P. A., and Nordhaus, W. D. (2010). Economics (19th ed.). McGraw-Hill.

UNICEF. (2023). Maternal, Newborn and Child Health Statistics for Nigeria. UNICEF, 5-89

United Nations Development Programme (UNDP). (2021). Human Development Report. New York: UNDP, 67-183.

United Nations Development Programme (UNDP). (2023). Human Development Report. UNDP.

World Bank. (2024). World Development Indicators. World Bank

World Health Organization (WHO). (2022). World Health Statistics. Geneva: WHO.