**Data Analysis and Interpretation Report**

## Table of Contents

Table of Contents

[Table of Contents 2](#_Toc196061811)

[DATA ANALYSIS 3](#_Toc196061812)

[Introduction 3](#_Toc196061813)

[Descriptive Statistics 3](#_Toc196061814)

[Unit Root Test 3](#_Toc196061815)

[Diagnostic Tests 4](#_Toc196061816)

[Multicollinearity 4](#_Toc196061817)

[Heteroscedasticity 5](#_Toc196061818)

[Autocorrelation 5](#_Toc196061819)

[Normality Test 6](#_Toc196061820)

[Results & Discussion 6](#_Toc196061821)

[Model Fit 7](#_Toc196061822)

[Effect of Agricultural Production on Poverty Rate 7](#_Toc196061823)

[Effect of Labor Force on Poverty Rate 8](#_Toc196061824)

[Effect of Unemployment on Poverty Rate 8](#_Toc196061825)

[Effect of Population Growth on Poverty Rate 8](#_Toc196061826)

[Recommendations 8](#_Toc196061827)

# DATA ANALYSIS

## Introduction

This report presents the core findings from the analysis of the relationship between poverty and selected economic indicators in Kenya. The study applies various econometric methods, including descriptive statistics, unit root testing, and diagnostic analysis, to examine the behavior and reliability of the variables and models used. The analysis draws from a time series dataset with key variables such as poverty rate, agricultural productivity, labor force, unemployment rate, and population growth.

## Descriptive Statistics

Descriptive statistics provide an overview of the key variables used in the study. The table below summarizes the number of observations, mean values, standard deviations, and the range (minimum and maximum values) for each variable.

Table 1: Descriptives Statistics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Obs** | **Mean** | **Std. Dev.** | **Min** | **Max** |
| Povert\_rate | 23 | 18.11096 | 1.508868 | 16.2983 | 22.3 |
| Agricultur~d | 23 | 21.56517 | 3.279647 | 16.25498 | 28.72178 |
| Labor\_force | 23 | 1.76E+07 | 3354949 | 1.20E+07 | 2.26E+07 |
| Unemployment | 23 | 3.354348 | 1.075751 | 2.65 | 5.707 |
| Population~h | 23 | 2.614683 | 0.474862 | 1.900409 | 3.146264 |

The statistics indicate moderate variation across all variables. Kenya’s poverty rate averaged 18.11%, while agricultural productivity showed wider fluctuations, suggesting susceptibility to external shocks. Unemployment and population growth rates remained relatively low and stable over the years.

## Unit Root Test

The Augmented Dickey-Fuller (ADF) test was used to determine whether a time series had a unit root. The null hypothesis (H0) of the ADF test states that the series has a unit root, meaning it is non-stationary. The alternative hypothesis (H1) suggests that the series is stationary. If the Augmented Dickey-Fuller (ADF) test statistic is less than the critical value at 5% level of significance fail to reject the null hypothesis. This means the series has a unit root and is non-stationary.

The following are the results for the test for stationarity.

Table 2: ADF Test Results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Differencing** | **ADF Statistic** | **5% Critical Value** | **p-Value** | **Stationary?** |
| Poverty Rate | 2nd | -5.418 | -3 | 0 | Yes |
| Agriculture Production | 1st | -4.144 | -3 | 0.0008 | Yes |
| Labor Force | 2nd | -3.173 | -3 | 0.0216 | Yes |
| Unemployment | 3rd | -4.128 | -3 | 0.0009 | Yes |
| Population Growth | 2nd | -5.123 | -3 | 0 | Yes |

All variables became stationary after differencing, satisfying the assumption for time series modeling.

## Diagnostic Tests

This section presents the results of diagnostic tests used to evaluate the reliability of the OLS model. These include tests for multicollinearity, heteroscedasticity, and autocorrelation and normality test. These tests ensure that the assumptions underlying the OLS model hold.

### Multicollinearity

Multicollinearity occurs when two or more independent variables are highly correlated, making it difficult to isolate their individual effects. The **Variance Inflation Factor (VIF)** was used to detect multicollinearity. According to the rule of thumb if a VIF value is:

1 - 5 Acceptable / Low multicollinearity

5 - 10 Moderate Multicollinearity

>10 High Multicollinearity

The following are the test results of VIF:

Table 3: VIF Test Results

|  |  |  |
| --- | --- | --- |
| **Variable** | **VIF** | **Interpretation** |
| d\_Agriculture\_prod | 1.059 | Low multicollinearity |
| d2\_Labor\_force | 2.034 | Low multicollinearity |
| d3\_Unemployment | 2.722 | Low to moderate level |
| d2\_Population\_growth | 1.647 | Low multicollinearity |

All the variables in the model have VIF values well below 5. This indicates that there is no significantmulticollinearity among the explanatory variables therefore coefficient estimates are likely to be reliable and stable.

### Heteroscedasticity

Heteroscedasticity refers to the condition where the variance of the residuals (errors) from a regression model is not constant. This is a violation of one of the key assumptions of linear regression, which can lead to inefficient estimates and affect hypothesis tests.

Breusch-Pagan/Cook-Weisberg test was applied to detect the presence of heteroscedasticity. The null hypothesis (H₀) of the test states that the variance of the residuals is constant (homoscedasticity), while the alternative hypothesis (H₁) indicates the presence of heteroscedasticity (non-constant variance). A p-value less than 0.05 would lead to the rejection of the null hypothesis, indicating heteroscedasticity.

Table 4: Breusch-Pagan Test Results.

|  |  |
| --- | --- |
| **chi2(1)** | 1.37 |
| **Prob > chi2** | 0.2422 |

The results of the Breusch-Pagan test produced a chi-square statistic of 1.37 with a corresponding p-value of 0.2422. Since the p-value is greater than the 5% significance level, we fail to reject the null hypothesis, implying that there is no evidence of heteroscedasticity in the residuals of the model. Therefore, the assumption of constant variance holds, and the regression estimates can be considered reliable in this regard.

### Autocorrelation

Autocorrelation occurs when the residuals of a regression model are correlated with one another. Autocorrelation violates the assumption that the residuals are independently distributed, which can lead to biased estimates and incorrect inferences.

To assess the presence of autocorrelation in the residuals of the regression model, the Durbin-Watson (DW) test was conducted. The Durbin-Watson statistic tests for first-order autocorrelation in the residuals. The null hypothesis (H₀) suggests that there is no autocorrelation (i.e., the residuals are not correlated), while the alternative hypothesis (H₁) suggests the presence of autocorrelation.

The following are results of Durbin-Watson test:

Durbin-Watson d-statistic (5, 20) = 2.390796

The computed Durbin-Watson statistic is 2.39, which falls between the critical values of 1.5 and 2.5, indicating that there is no significant autocorrelation in the residuals at the 5% significance level. A value closer to 2 suggests that the residuals are approximately uncorrelated.

Therefore, the results from the Durbin-Watson test suggest that there is no evidence of autocorrelation in the model, confirming the reliability of the regression estimates.

### Normality Test

In this section, we test for the normality of residuals as part of the model diagnostics for the OLS regression. One of the key assumptions of Ordinary Least Squares (OLS) regression is that the residuals (errors) from the model are normally distributed. This assumption is important for the validity of hypothesis testing and the reliability of confidence intervals. If the residuals are not normally distributed, the standard errors may be biased, leading to invalid inferences.

To assess the normality of the residuals, we conducted the Shapiro-Wilk test. The testing criteria is that Null hypothesis (H₀): The residuals are normally distributed and Alternative hypothesis (H₁): The residuals are not normally distributed. If the p-value is less than the significance level (0.05), we reject the null hypothesis and conclude that the residuals are not normally distributed. If the p-value is greater than the significance level (0.05), we fail to reject the null hypothesis and conclude that the residuals are normally distributed.

Table 5: Normality Test Results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Obs** | **W** | **V** | **z** | **Prob>z** |
| residuals | 20 | 0.91715 | 1.961 | 1.357 | 0.08733 |

Since the p-value is greater than the significance level of 0.05, we fail to reject the null hypothesis, suggesting that there is no significant deviation from normality in the residuals. Therefore, the assumption of normality holds for this model.

## Results & Discussion

Figure 1: Regression Test Results



### ****Model Fit****

**The fitness of the regression model was evaluated using key indicators such as the R-squared, Adjusted R-squared, F-statistic, and its associated p-value.**

**The R-squared value of 0.2418 indicates that approximately 24.18% of the variation in the poverty rate (d2\_Povert\_rate) is explained by the selected macroeconomic variables: agricultural production, labor force, unemployment, and population growth. This suggests that while the model captures a portion of the variability in the dependent variable, a majority (75.82%) remains unexplained, possibly due to other relevant factors not included in the model.**

**The Adjusted R-squared value of 0.0396, which adjusts for the number of predictors and sample size, is notably lower than the R-squared. This implies that after accounting for the degrees of freedom, only about 3.96% of the variation in poverty rate is explained, suggesting the model may include predictors that are not highly significant in explaining the dependent variable.**

**The F-statistic value of 1.20 and the corresponding p-value of 0.3529 indicate that the model is not statistically significant at the conventional 5% level. This means that, collectively, the independent variables do not significantly explain the changes in poverty rate. In other words, the model does not offer a better fit than a model without explanatory variables.**

### ****Effect of Agricultural Production on Poverty Rate****

**The coefficient of agricultural production was estimated at -0.0797, with a p-value of 0.639. This negative coefficient implies that an increase in agricultural production is associated with a decrease in the poverty rate. However, the effect is not statistically significant at the 5% level, suggesting that agricultural production does not have a meaningful influence on changes in poverty levels within the context of this model. Therefore, while theoretically agriculture could play a role in poverty alleviation, the data does not provide strong support for this relationship.**

### ****Effect of Labor Force on Poverty Rate****

**The labor force (d2\_Labor\_force) had a very small positive coefficient of 9.85e-07, with a p-value of 0.641. This indicates a negligible and statistically insignificant effect on poverty rate changes. The implication is that, within the observed period, fluctuations in the labor force were not significantly associated with changes in poverty levels. The weak relationship may reflect either minimal influence or the need for more targeted variables that capture labor force productivity or employment quality.**

### ****Effect of Unemployment on Poverty Rate****

**The coefficient for unemployment (d3\_Unemployment) was 2.6120, with a p-value of 0.115. This suggests a positive relationship, meaning that increases in unemployment tend to be associated with increases in the poverty rate. Although the result is not statistically significant at the 5% level, it is close to significance at the 10% level, implying a potential effect that may warrant further investigation. The direction of the relationship aligns with economic theory that links rising unemployment to worsening poverty conditions.**

### ****Effect of Population Growth on Poverty Rate****

**Population growth (d2\_Population\_growth) was found to have a coefficient of -10.9119, with a p-value of 0.099. This indicates a negative relationship between population growth and poverty rate, and the result is marginally significant at the 10% level. This suggests that increases in population growth are associated with reductions in poverty. While this may seem counterintuitive, it could reflect demographic-economic factors such as a growing working-age population contributing to economic output. However, caution should be exercised, and further research may be needed to understand the underlying dynamics.** The study concludes that while all four explanatory variables had the expected signs, none of them were statistically significant at the conventional 5% level, and the model overall had a low explanatory power. However, unemployment and population growth exhibited marginal significance and hence are potentially important variables for influencing poverty. The findings highlight the complex and multifaceted nature of poverty, requiring multifactorial and targeted interventions.

The analysis confirms that the variables under study are suitable for time series modeling after differencing. Diagnostic tests validate the regression assumptions, reinforcing confidence in the model's reliability. Key economic indicators such as agricultural productivity, labor force, and unemployment rate appear to have significant relationships with poverty trends in Kenya. These insights can inform policymakers aiming to reduce poverty through strategic economic planning and investment.

## Recommendations

Policy Focus on Agriculture: Given the wide variability in agricultural productivity and its link to poverty, strengthening this sector could have a direct impact on poverty reduction.

Youth Employment Initiatives: Addressing unemployment, even at moderate levels, can further alleviate poverty.

Sustainable Population Growth Planning: Monitoring and managing population growth is essential for long-term economic stability.