NAME: ACKSON MUTOBA

RESEARCH STUDY: ECONOMETRIC MODELING

ECONOMETRIC SOFTWARE: EViews

<u>RESEARCH TOPIC:</u>

MACROECONOMIC TIME SERIES ANALYSIS IN ZAMBIA [MULTIVARIABLE ANALYSIS]

Assignment Task.

We must download two economic variables related to Zambia from any database, e.g., Central Statistics Zambia, Bank of Zambia, International Monetary Fund, World Bank, etc., yearly from 1980 to 2019.

Required:

Introduce the chosen variables and why they are considered important for the economy. Remember to cite appropriately and avoid plagiarism.

Choose a dependent and independent variable from the downloaded variables and run a simple OLS regression of the two selected variables using E-views or STATA and generate the stationarity test at I(0) and I(1) using both the Augmented Dicky Fuller and Perron Tests. Interpret the generated results at a 5% level of significance.

Write a conclusion about what has been observed from the yearly data ranging from 1980 to 2019 and recommend what can be done to improve the economy should need to be.

1. Introduction.

   For us to measure the performance, structure, behavior, and decision-making of an economy as a whole to determine the health of the economy of any country, we measure this by certain indicators called macroeconomic indicators.

   Macroeconomic indicators are defined as specific economic data points that help investors gauge the overall health of the economy and assess its current prospects. Central banks, statistics bureaus, and other government organizations typically release this information. As an example, some popular indicators include the **Unemployment Rate**, the Consumer Price Index (**CPI**), which gives perspective on inflation, and the Gross Domestic Product (**GDP**), which shows the rate of economic growth, **Population, Interest Rates, Money Supply,** etc.

   In this paper, we focus on two variables Population and GDP, so that it helps us in understanding **economic productivity** and **growth** due to demographic changes. Talking of population variables, the National Census of Housing and Population conducted by the government of Zambia in 2022 August, an estimated aggregated population as of today stands at 19,693,423 million, Zambia Statistical Agency.

   GDP (Gross Domestic Product) is an important and common measurement of the health of an economy, it is the total monetary or market value of all the

finished goods and services produced within a country’s borders in a specific period.
There are three main types of GDP; Nominal GDP, Real GDP, and GDP per capita.

i. Nominal GDP - an assessment of economic production in an economy that includes current prices in its calculation. All goods and services counted in nominal GDP are valued at the prices that those goods and services are sold for in that year

ii. Real GDP - is an inflation-adjusted measure that reflects the number of goods and services produced by an economy in a given year, with prices held constant from year to year to separate the impact of inflation or deflation from the trend in output over time. Since GDP is based on the monetary value of goods and services, it is subject to inflation

iii. GDP per capita - is a measurement of the GDP per person in a country’s population. It indicates that the amount of output or income per person in an economy can indicate average productivity or average living standards. This is what we will use in this study from the period of 1980 to 2019, measured in US Dollars $.

2. Data Collection.

In any statistical analysis, the collection of data plays a significant part. The method of collecting information is divided into two different sections, namely primary data and secondary data.

In this process, the primary data is assembling data or information for the first time, whereas the secondary data is the data that has already been gathered or collected by others such as government publications like census data from Zambia Statistics Agency, academic journals, research institute reports, published articles from local and central government bodies, statistical synopses, online databases, etc.

In this study, we will use secondary data ensuring that we get or extract it from a reliable source such as the Zambia Statistics Agency, World Bank, IMF, and so on.

3. Data Cleaning and Reprocessing.

This is a crucial process where we will be identifying and correcting errors, inconsistencies, or missing values within followed by transforming the cleaned data further by standardizing, normalizing, etc., if necessary, in preparation for our analysis,

4. Exploratory Data Analysis [Visualization].

Exploratory Data Analysis (EDA) is an important first step in our analysis stage. It involves looking at and visualizing our two variables to understand their main features, find patterns, and discover how different parts of the data are connected. EDA helps spot any unusual data or outliers and is usually done before starting more detailed statistical analysis or building models.

To begin with, let us explore the series of variables in terms of the following;

i. Descriptive statistics such as mean, maximum, and minimum values were used to measure the variables exerted during this period from 1980 to 2019.

Date: 03/02/25 Time: 20:51
Sample: 1980 2019

	GDP_PER_C	POPULATIO
Mean	787.0655	10802523
Median	523.4088	9878818.
Maximum	1820.719	18513839
Minimum	238.2338	5802833.
Std. Dev.	492.2842	3755374.
Skewness	0.769425	0.533277
Kurtosis	2.042436	2.084754
Jarque-Bera	5.474978	3.292022
Probability	0.064733	0.192817
Sum	31482.62	4.32E+08
Sum Sq.	34230292	5.22E+15
Sum Sq. Dev.	9451407.	5.50E+14
Observations	40	40

<u>Mean</u>

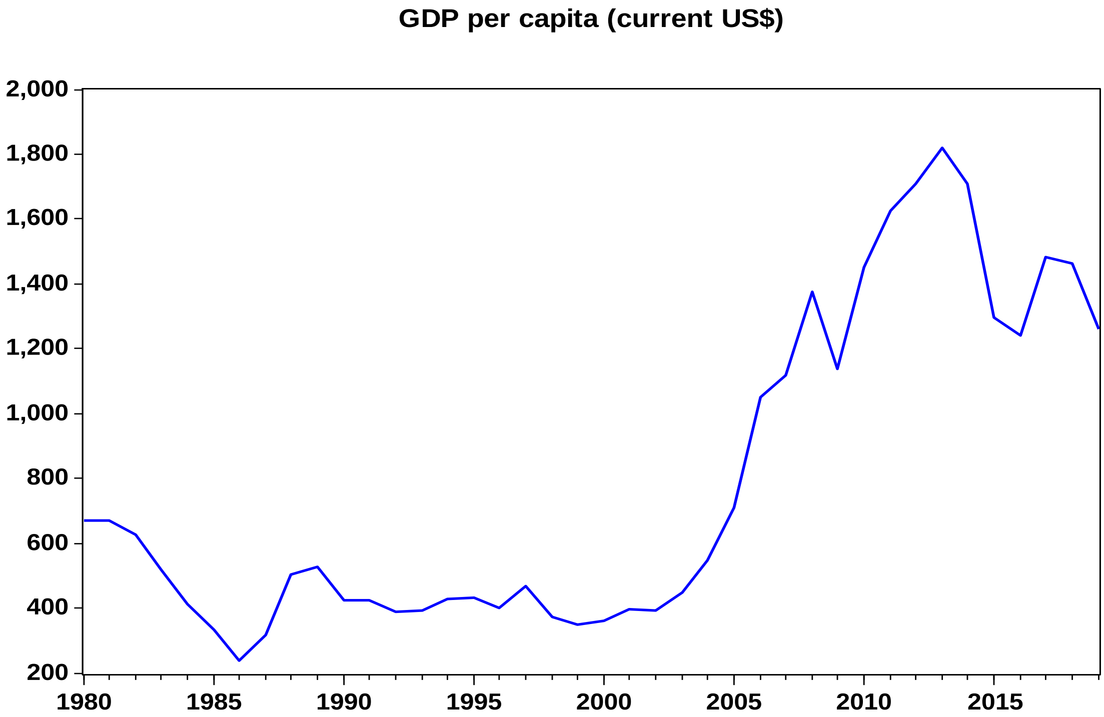
The mean GDP (787.0655) represents the average GDP over the 40 years. This gives a central tendency of the economic output and the mean population (10,802,523) represents the average population size over the same period. This helps in understanding the average demographic size.

<u>Maximum and Minimum</u>

The maximum GDP (5,802,833) and minimum GDP (18,513,839) show the range of economic output over the 40 years. This helps in understanding the variability and growth trends. The maximum population (238.2338) and minimum population (1,820.719) indicate the range of population size, which is crucial for understanding demographic changes

ii. Trend of the variables during this period.

A. GDP



The GDP data shows significant growth over the 40 years, starting at $**669.42** in 1980 and reaching $**1258.99** by 2019. This indicates a **doubling** of GDP, reflecting economic development, increased mining activity (e.g., copper), and foreign investments.

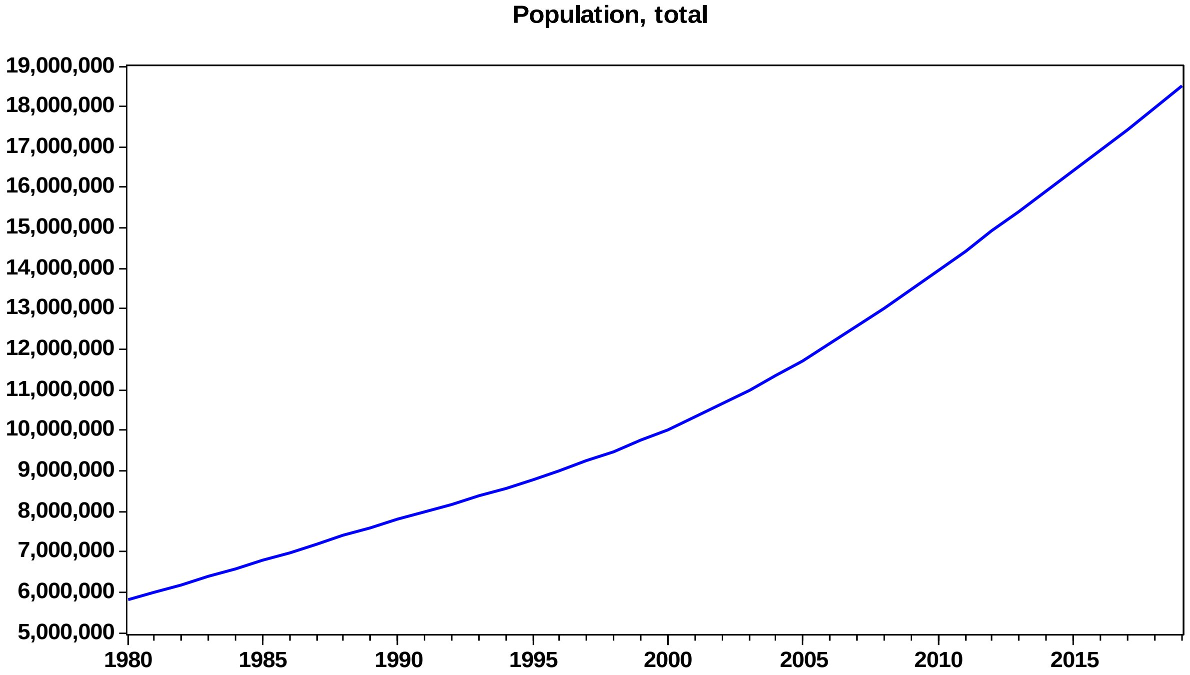
Fluctuations:

GDP growth is more volatile compared to population growth:

- Between **1980 and 1990**, GDP growth is slow and uneven, likely due to economic challenges, global recessions, or internal policy issues.
- From **1990 to 2010**, GDP growth accelerates, driven by economic reforms, mining booms, and improved global commodity prices.
- After **2010**, GDP growth slows, possibly due to global economic downturns, declining commodity prices, or domestic challenges.

Time series data such as these variables in this study aim to identify patterns, trends, seasonality, and dependencies within the data and use this information primarily as an aid to forecasting, Enders, W. (2008).

B. Population



The population data shows a steady increase over the 40 years, starting at **5,802,833** in 1980 and reaching **18,513,839** by 2019. This represents a **tripling** of the population, reflecting high birth rates, improved healthcare, and declining mortality rates.

To understand the **percentage growth** of the population over the period from **1980 to 2019**, we can use the following formula:

$$Percentage\ Growth = \left(\frac{Final\ Pop - Initial\ Pop}{Initial\ Pop}\right) \times 100$$

The population of Zambia grew by **219%** from 1980 to 2019. This means the population more than tripled over the 40 years.

Annualized Growth Rate: To understand the average annual growth rate, we can use the formula for compound annual growth rate (CAGR):

$$CAGR = \left(\frac{Final\ PopulationI}{Initial\ Population}\right)^{\frac{1}{n}} - 1$$

Where n is the number of years under consideration. The population grew at an average annual rate of **2.9%** over the 40-year period, which is relatively high and consistent with rapid population growth in many developing countries.

5. Building our Simple linear regression [E-view Software].

   a) Using E-view 14 which is an application we are using. We will open a new working file and import our Excel file.

   b) Before we proceed, we need to select which variable will act as a dependent variable and which independent variable. We will let GDP be our dependent variable and this will help in understanding economic productivity and growth due to demographic changes.

      Metrics like GDP per capita can indicate living standards, while population growth trends can influence long-term GDP growth through labor force expansion.

   c) By selecting the Quick-Estimate Equation in our E-view, then we configure our equation as follows;

      - Method: OLS or LS (Least Square and ARMA)
      - Sample: We will use the entire period 1980 – 2019
      - Equation: Where $\beta_0 = C$

$$GDP = \beta_0 + \beta_1 \cdot POPULATION$$

<u>Estimation Output.</u>

Dependent Variable: GDP
Method: Least Squares
Date: 03/04/25 Time: 05:47
Sample: 1980 2019
Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
POPULATION	0.000110	1.15E-05	9.537452	0.0000
C	-402.2245	131.8389	-3.050879	0.0041

R-squared	0.705342	Mean dependent var	787.0655
Adjusted R-squared	0.697588	S.D. dependent var	492.2842
S.E. of regression	270.7171	Akaike info criterion	14.08873
Sum squared resid	2784935.	Schwarz criterion	14.17318
Log likelihood	-279.7747	Hannan-Quinn criter.	14.11926
F-statistic	90.96299	Durbin-Watson stat	0.297613
Prob(F-statistic)	0.000000		

The population has a statistically significant positive effect on GDP at a 5% confidence interval, which implies for every 1-unit increase in Population, GDP is expected to increase by 0.000110 units.

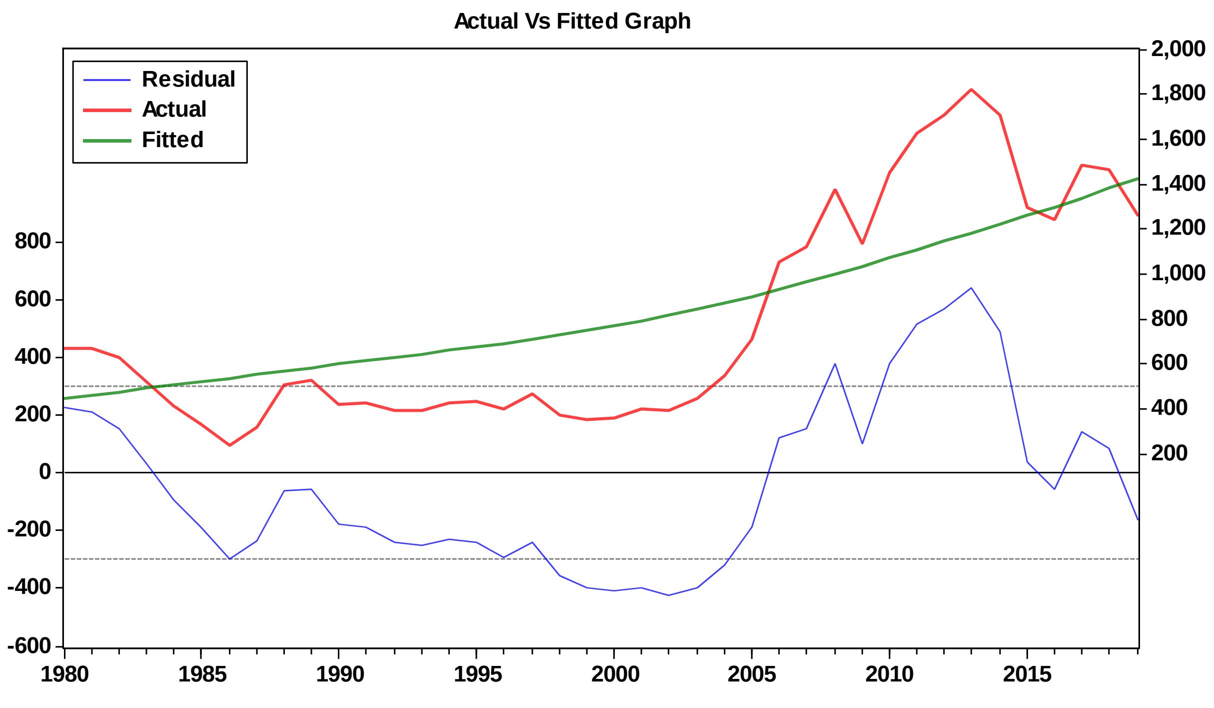
The model explains 70.53% of the variation in GDP, which is a good fit, but other factors may also play a role.

The adjusted R-squared for the number of predictors in the model is slightly lower than the R-squared, which is expected when adding variables, Nevertheless, the model's explanatory power remains robust even after adjusting for the number of predictors.

Autocorrelation is present (Durbin-Watson statistic = 0.297613), indicating that the model may be mis-specified. This could be due to omitted variables, nonlinear relationships, or time-dependent effects.

The model fits the data reasonably well based on R-squared, AIC, and SC, but improvements are needed to address autocorrelation and potentially include other explanatory variables.

Estimation Output: Actual vs Fitted, Residual Graph.



An **Actual vs Fitted graph** is a visual tool used in regression analysis to compare the **observed (actual) values** of the dependent variable (in this case, **GDP**) with the **predicted (fitted) values** generated by the regression model. This graph helps assess how well the model fits the data and identifies any patterns or discrepancies in the predictions.

6. Stationarity Test for individual Series: By checking for the existence of a Unit Root in our time series data, we can identify if our time series is stationary or non-stationary at a 5% level of significance.

   Most econometric models (e.g., regression, forecasting) require stationary data to provide meaningful results. If a series is non-stationary, using it in regression can lead to spurious results.

   i. Hypothesis: Using unit root test

      $\mathrm{H_0}$: Population has UNIT ROOT
      $\mathrm{H_1}$: Population has NO UNIT ROOT

      Interpreting Perron Test Results at 5% Significance

      - If the **test statistic** is **more** than the 5% critical value → **Fail to Reject $\mathbf{H_0}$** (series has a unit root) implying it is non-stationary.

      - If the **test statistic** is **less** than the 5% critical value → **Reject** $\mathrm{H_0}$ (series has no unit root) implying it is stationary.

   By selecting the time series data we need to test for, we carry our Time Series Diagnostic – Unit Root Tests – Standard Unit Root Test, and then we select the following configuration;

   - Test Type: ADF or Philips Perron
   - Test for unit root in: Level I(0) or 1st Difference I(1)
   - Include in Test Equation: Intercept
   - Other configurations we will set to default.

   We will do the test twice for each Test Type and at each level, meaning using ADF we will test GDP at I(0) and I(1) and Population I(0) and I(1). Similarly, we carry out the Perron test for both variables and at both required levels. Using E-view we obtained the following results particularly paying attention to test-statistic probability.

ii. Augmented Dicky Fuller (ADF).

Checks if a time series has a unit root (stationary or non-stationary) account for breaks in the trend or level.

I(0): Level (Raw Data)

Null Hypothesis: POPULATION has a unit root
Exogenous: Constant
Lag Length: 2 (Automatic - based on SIC, maxlag=9)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		2.325445	0.9999
Test critical values:	1% level	-3.621023	
	5% level	-2.943427	
	10% level	-2.610263	

Null Hypothesis: GDP has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=9)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.675889	0.8411
Test critical values:	1% level	-3.610453	
	5% level	-2.938987	
	10% level	-2.607932	

I(1): First Difference (Difference)

Null Hypothesis: D(POPULATION) has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=9)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		0.038015	0.9561
Test critical values:	1% level	-3.621023	
	5% level	-2.943427	
	10% level	-2.610263	

Null Hypothesis: D(GDP) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=9)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.936817	0.0003
Test critical values:	1% level	-3.615588	
	5% level	-2.941145	
	10% level	-2.609066	

iii. Philips Perron

The Perron Test is used to check for a unit root while accounting for structural breaks in the data.

L(0): Level (Raw Data)

Null Hypothesis: GDP has a unit root
Exogenous: Constant
Bandwidth: 1 (Newey-West automatic) using Bartlett kernel

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-0.772287	0.8158
Test critical values:	1% level	-3.610453	
	5% level	-2.938987	
	10% level	-2.607932	

Null Hypothesis: POPULATION has a unit root
Exogenous: Constant
Bandwidth: 5 (Newey-West automatic) using Bartlett kernel

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		11.23583	1.0000
Test critical values:	1% level	-3.610453	
	5% level	-2.938987	
	10% level	-2.607932	

L(1): First Difference (Difference)

Null Hypothesis: D(GDP) has a unit root
Exogenous: Constant
Bandwidth: 2 (Newey-West automatic) using Bartlett kernel

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-4.925323	0.0003
Test critical values:	1% level	-3.615588	
	5% level	-2.941145	
	10% level	-2.609066	

Null Hypothesis: D(POPULATION) has a unit root
Exogenous: Constant
Bandwidth: 4 (Newey-West automatic) using Bartlett kernel

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		0.596442	0.9878
Test critical values:	1% level	-3.615588	
	5% level	-2.941145	
	10% level	-2.609066	

7. Overall Interpretation of our Statistical Results.

Most econometric and financial time series data (e.g., forex, population, money supply, inflation) exhibit non-stationary features due to seasonality and trends in the pattern. If a series is non-stationary, using it in regression or modeling can lead to spurious results.

Augmented Dickey-Fuller Test Results.

In step 6, we carried out steps and tested for stationarity using two test types. We tested for the existence unit root in Population and GDP using ADF test type and at I(0) and we obtained test statistics of **0.999 and 0.8411** respectively. This meant that we failed to reject the null hypothesis meaning that both are non-stationery and unit root exists in both series.

Moving on to carrying out the test at I(1) or first difference, Population and GDP using ADF test type we obtained test-statistic of **0.9561 and 0.00003** respectively. This meant that we failed to reject the null hypothesis for the Population series meaning that it is still non-stationary and unit root exists even at first difference. On the other hand, with GDP at I(1) we obtained a test statistic of **0.00003** which is less than our significance level of 0.05, therefore we can reject the null hypothesis and conclude that the series is stationary at first difference**.**

Philips Perron Test Results.

This similar type of unit root test is carried out to verify our ADF results due to the difference that exists in test types. At I(0), GDP and Population test-statistic we had **0.8158** and **1.000 all greater than 0,05**, in such case we fail to reject the null hypothesis meaning all are non-stationary at I(0).

At I(1) or first difference, GDP and Population test-statistic are **0.0003** and **0.9878** respectively**.** This means that we fail to reject the null hypothesis for the Population series again, meaning it is still non-stationary and unit root exists even at first difference.

GDP at I(1) we obtained a test statistic of **0.00003** which is less than our significance level of 0.05, therefore we can reject the null hypothesis and conclude that the series is stationary at first difference**.**

Below is a Table of Test Statistics Summary Results of both ADF and Philips Perron test statistic Probabilities.

Test Type	**Variable**	**L(0)** **Raw Data**	**L(1)** **1st Difference**
ADF	Population	0.9999	0.9561
ADF	GDP	0.8411	0.00003
Philips Perron	Population	1.0000	0.9878
Philips Perron	GDP	0.8158	0.00003

SUMMARY AND CONCLUSION.

In order for us to create a meaningful and precise regression model equation, once both variables are stationary or cointegrated, you can proceed with the regression analysis.

In the earlier section 5 where we developed the LS model, we used raw data, during the further steps we concluded that raw data was non-stationary by using both Augmented Dickey-Fuller and Philips Perron Tests.

Let us understand the Problem;

- Population: Differencing twice did not make it stationary. This suggests that the Population may have a more complex structure (e.g., a higher-order integrated process, Log-Diff, or a nonlinear trend).
- GDP: Differencing once made it stationary, so it is integrated of order 1 (I(1)).

The challenge is that regression requires variables to be stationary or cointegrated. If one variable is stationary and the other is not, you cannot directly use them in a regression model, this has been highlighted to know that the regression model belt by non-stationary variables.

In this study, our focus was on Population and GDP variables, so that it helps us understand economic productivity and growth due to demographic changes. Time series models aim to identify patterns, trends, seasonality, and dependencies within the data and use this information primarily as an aid to forecasting, Enders, W. (2008), hence it is essential data for the decision-making process.

RECOMMENDATION.

1. One key note for the Ministry of National Finance and National Planning and overall government administration is to understand the annual growth rate of **2.9%** over the 40-year period for future planning which is relatively high and consistent with rapid population growth in many developing countries like Zambia, therefore, with a rapidly growing population, Zambia needs **to invest in education, healthcare, and infrastructur**e to ensure sustainable development.

   Forecasting is one of the most widely used models in time series forecasting the Autoregressive Integrated Moving Average (ARIMA) model according to Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Utilizing these terms will equip the administration in terms of planning and development goals.

2. Despite GDP growth, the high population growth rate may limit improvements in living standards. **Policies** should focus on equitable wealth distribution and poverty alleviation. Rapid population growth and economic expansion could strain natural resources. Sustainable practices are essential to balance development with environmental conservation.

3. Slow GDP growth and moderate population growth suggest economic challenges, possibly linked to global recessions or internal policy issues.

   The corrected interpretation of the data shows that Zambia’s population has tripled while GDP has doubled over the 40 years. This highlights the challenge of achieving sustainable development in the face of rapid population growth. Policymakers must focus on economic diversification, population management, and equitable development to ensure long-term prosperity.

   Much of the decline was significantly due to

   - 2008 Global Financial Crisis: Lower global demand led to another drop in copper prices, affecting GDP growth. Recessions in major economies (e.g., the U.S., China, or the EU) reduced demand for Zambia's exports, leading to lower GDP growth

   - 2015-2016 Copper Price Slump: A sharp decline in copper prices and production cuts at major mines slowed GDP growth.

   Zambia's GDP declines from the trend from 1980 to 2019 were caused by a combination of external shocks (e.g., falling copper prices, global recessions), domestic challenges (e.g., droughts, policy uncertainty, high debt), and structural issues (e.g., lack of diversification, infrastructure deficits). Addressing these issues requires a combination of economic diversification, infrastructure development, and stable policies to ensure sustainable growth.

   Zambia's economic growth and development are closely tied to its ability to manage population growth, diversify its economy, and address structural challenges. By implementing the recommended policies, Zambia can achieve sustainable growth, reduce poverty, and improve living standards for its population.

REFERENCES

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. John Wiley & Sons.

Enders, W. (2008). *Applied econometric time series*. John Wiley & Sons.

World Bank. (n.d.). *Population, total - Zambia*. Retrieved from https://databank.worldbank.org/indicator/SP.POP.TOTL/1ff4a498/Popular-Indicators#advancedDownloadOptions

Zambia Statistical Agency. (2024). *Monthly inflation: January 2024 - January 2025*. Retrieved from https://www.zamstats.gov.zm/monthly-inflation-january-2024-january-2025/