

# Forecasting Zambia's Gross Domestic Product Using Time Series Autoregressive Integrated Moving Average (ARIMA) Model

Chikumbe Evans Sankwa  
Department of Economics and Finance  
Kwame Nkrumah University, Kabwe, Zambia

Sikota Sharper  
Department of Economics and Finance  
Kwame Nkrumah University, Kabwe, Zambia

**Abstract:-** Gross Domestic Product is one of the social indicators of development. This study attempts to model Zambia's Gross domestic product using the Autoregressive Integrated Moving Average (ARIMA) model. This model has proved to help many countries during economic recession or when there is any disruption in the economic system due to pandemics or natural disasters. The study utilized a time series dataset from 1960 to 2018. The best model that fit the data set, following the selection model criteria, was ARIMA (5,2,0) model with the lowest Akaike's Information Criteria(AIC) and Bayesian Information Criteria (BIC) and smallest volatility. The study results showed that, on average, Zambia's gross domestic product will continue to rise over the next eight years. However, few recession (decline) points are expected in the period 2020 to 2022. It is hoped that the forecasts would be useful for researchers in Zambia, including the fiscal and monetary policy makers.

**Keywords:-** ARIMA Model, Gross Domestic Product, Forecasting.

## I. INTRODUCTION

Zambia is one of the Sub-Saharan countries that is striving hard to attain the status of upper middle- income country since 2008. However, due to regime changes and conflicting policies, the country has experienced economic down-turn and ups. To achieve sustainable development, the government has embarked on massive infrastructure development. Foreign Direct Investment is one area that need to be enhanced in order to boost investor confidence. And to do this, government through the Ministry of Finance release the budget projections every year. The budget normally gives some forecast of GDP in order to help investors form some business expectations for the future.

For example, it has been projected that real GDP will grow at 2%, 4.5% and 4.6% for 2019, 2020 and 2021, respectively (Zambia National Budget Review, 2019). But how accurate are these predictions given by the National Budget document? The Autoregressive Integrated Moving Average (ARIMA) model becomes the tool of proving such facts. Its chosen because of its proven accuracy by scholars of econometrics. This paper attempts to forecast GDP for the next eight years.

## II. LITERATURE REVIEW

Raymond (1997) summarized the two questions that must be answered to identify the nature of data in a stochastic process:

- whether the realizations are a random process
- does the series have any trend?

After answering the above questions, the next step is to do *identification*, *parameter estimation* and *testing for model validity*. If a series is random, the correlation between successive values in a time series is close to zero. If the realizations depend on each other, then, the Box-Jenkins or ARIMA model is appropriate for the time series analysis.

Mohad and Ahmed (2019) approximated the ARIMA model for forecasting Egyptian gross domestic product for the annual data from 1965 to 2016. Their research utilized both in-sample and out-of-sample ARIMA estimations in order to enhance model accuracy and minimize forecast errors.

Bipasha and Bani (2012) modelled the forecasting of GDP growth rates in India. GDP data on growth rates were collected over a period of 60 years from various publications of Reserve Bank of India. They tested some hypotheses to prove the precondition to estimate the most parsimonious model to forecast GDP and its growth rates in some periods ahead. An ARIMA (1, 2, 2) model was fitted to estimate the parameters of autoregressive and moving average components of this model. Results suggested that only one period of autoregressive and moving average terms were statistically significant. Absolute values of forecasted GDP indicated an increasing trend and its respective growth rates revealed an opposite trend in future. The results were meant to help policy makers and managers to formulate economic and business strategies that are accurate in planning investment decisions.

Contreras et al (2003) conducted a study that used the ARIMA technique, to predict next-day electricity prices both for spot markets and long-term contracts in Spain and Californian markets.

From the range of many time series analysis, the Autoregressive Integrated Moving Average (ARIMA) is a popular technique of forecasting univariate time series variables such as exchange rate, inflation, mortality rate etc.

This study attempted to predict the future Gross Domestic Product values for a period of eight (8) years in Zambia by fitting an ARIMA model on the time series data from 1960 to 2018.

### III. METHODOLOGY

#### ❖ Modeling Procedure

This study utilized the Autoregressive Integrated Moving Average (ARIMA) model also known as Box-Jenkins Methodology to forecast Zambia's Gross Domestic Product for the next eight years. The main reason of choosing ARIMA model in forecasting emanates from the fact that the model assumes and takes into account the non-zero autocorrelation between the successive values of the time series data.

The ARIMA model combines the Autoregressive Model and Moving Average part:

$$Y_t = \mu + \rho_1 Y_{t-1} + \rho_2 Y_{t-2} + \dots + \rho_p Y_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

or

$$\rho(L)Y_t = \theta(L)\epsilon_t$$

The steps involved in ARIMA modelling includes the following:

#### ➤ Identification

Identification involves determining the values (p, d, and q) of the model. The *p* and *q* are the order of the autoregressive and moving average parts, respectively. The letter "d" determines the stationarity or non-stationarity of the data. If the data is non-stationary, what is the order of differentiation (d), that will make it stationary?

Making sure that the variables are stationary, identifying seasonality in the series, and using the plots of

the Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) of the series to identify which autoregressive or moving-average component should be used in the model.

#### ➤ Estimation

This stage requires the estimation of parameters using the Maximum Likelihood Estimation (MLE) or least-squares estimation procedures.

#### ➤ Diagnostic Checking

This step involves testing whether the estimated model conforms to the specifications of a stationary univariate process. The residuals should be independent of each other and constant in mean and variance over time; plotting the ACF and PACF of the residuals are helpful to identify misspecification. If the estimation is inadequate, we have to return to step one and attempt to build a better model. Moreover, the estimated model should be compared with other ARIMA models to choose the best model for the data. The two common criteria used in model selection: Akaike's Information Criterion (AIC) and Bayesian Information Criteria (BIC) which are defined by:

$$AIC = 2m - 2 \ln(\hat{L}), BIC = \ln(n)m - 2 \ln(\hat{L}),$$

where  $\hat{L}$  denotes the maximum value of the likelihood function for the model, *m* is the number of parameters estimated by the model, and *n* is the number of observations. Practically, AIC and BIC are used with the classical criterion: the Mean Squared Error (MSE).

#### ➤ Forecasting or Predicting

When the selected ARIMA model conforms to the specifications of a stationary univariate stochastic process, then we can use this model for forecasting the out of sample data.

### IV. PRESENTATION AND DISCUSSION OF RESULTS

#### ➤ Step 1: Identification

In the identification section, the study checked whether the time series data is stationary or not. We check this by first plotting a two-way graph and then use the correlogram, ACF and PACF to determine stationarity or non-stationarity of data.



Fig 1:- Gross Domestic Product for Zambia from 1960 to 2018

The graph in figure 1 shows that the data has an upward trend implying its non-stationary. The correlograms (see figure 2 below) also show the same results. Furthermore, figure 2 shows that lag 1 up to 4 are outside the 95% confidence interval which is evidence that we need to difference the GDP series.

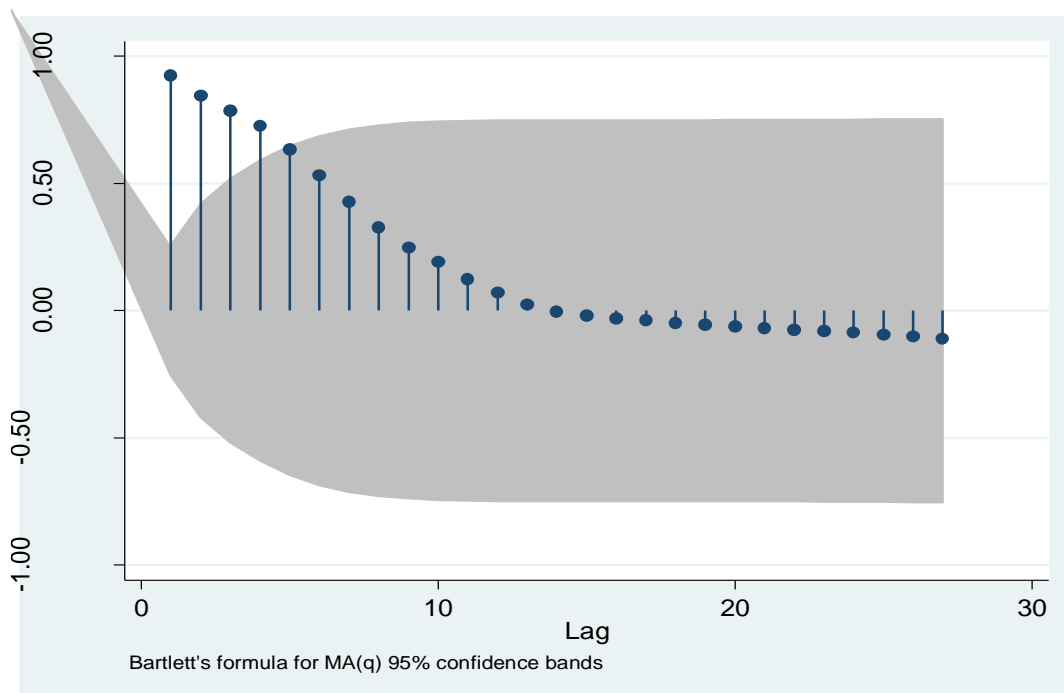


Fig 2:- Autocorrelation function

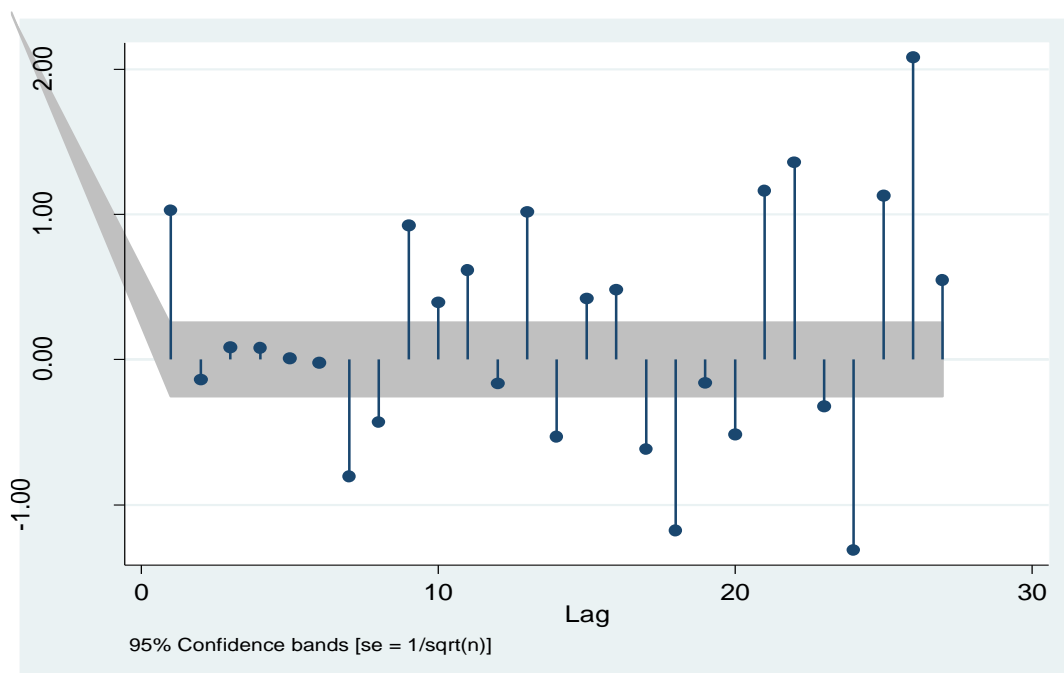


Fig 3:- Partial Autocorrelation Function

The figures 1, 2 and 3, clearly show that there is need to make the data stationary by differencing it. In order to make the time series stationary, we first smoothen the peaks by using common logarithm and difference the variable twice ( $d=2$ ), the autocorrelation function and the partial autocorrelation function below in figures 4 and 5 shows that the process is now stationary.

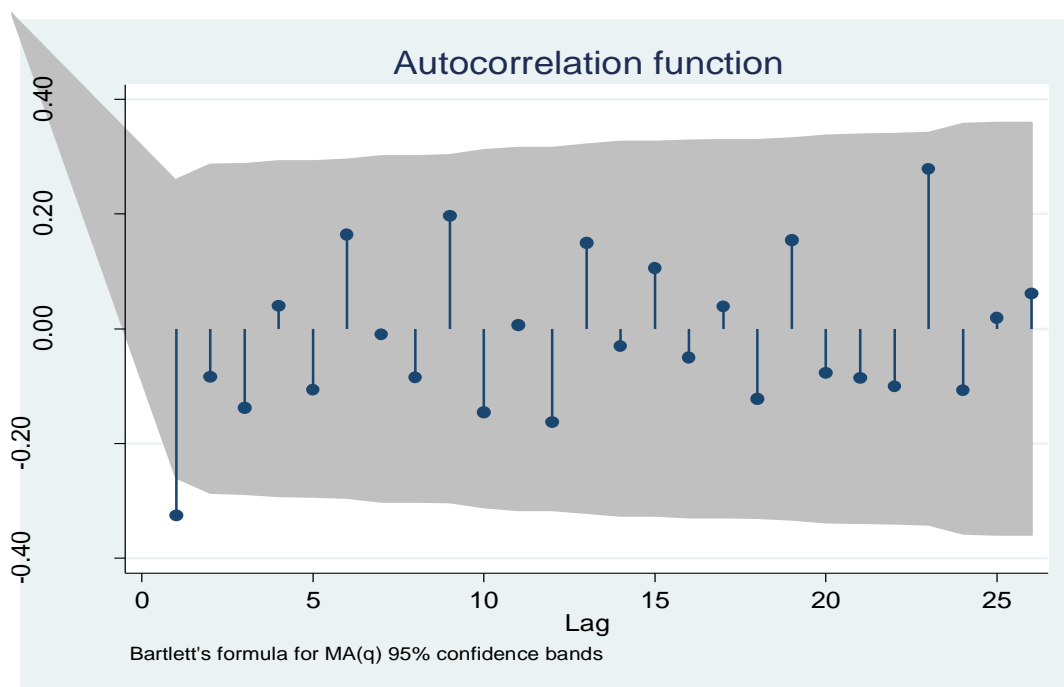


Fig 4:- ACF for the differenced log GDP

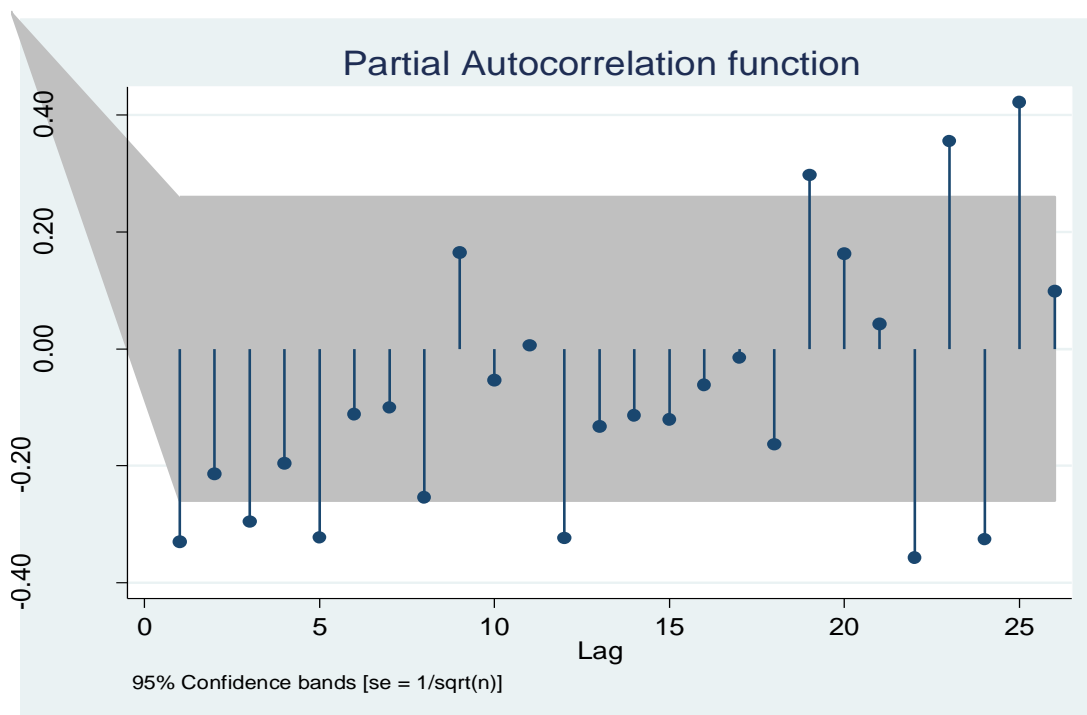


Fig 5:- Partial Autocorrelation function of log of GDP differenced twice.

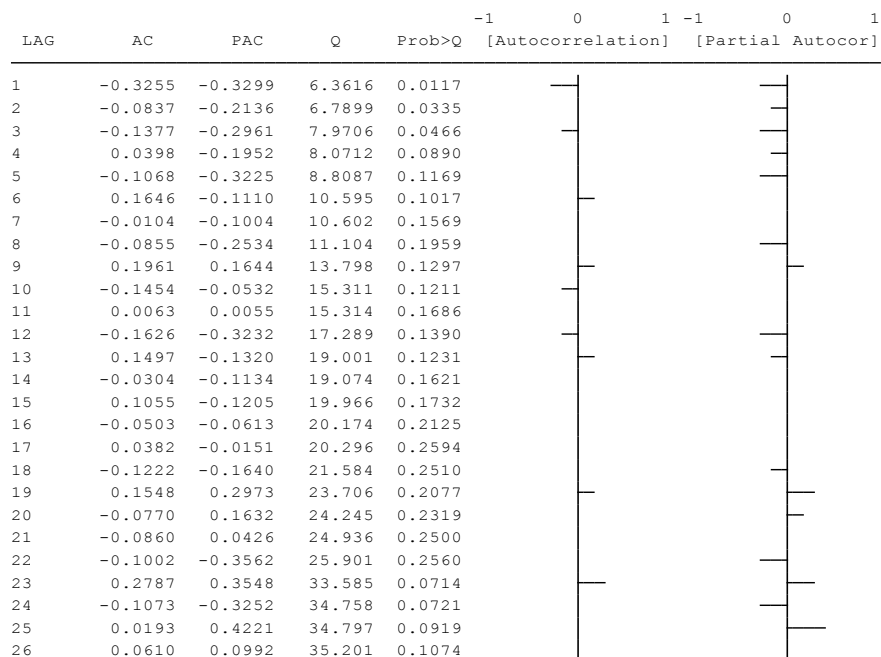


Fig 6:- Correlogram differenced log of gross domestic product.

The correlogram above shows that most of the coefficient are not significant except lags 1 to 3. This suggests that we can have an AR with three lags while the autocorrelation is suggesting 1 lag for the moving average.

The formal way of testing for stationarity is by using the Augmented Dickey-Fuller test. The null hypothesis states that the process is not stationary. Table 1 below now shows the results for the test which is statistically significant. This means the process is stationary. We reject the null hypothesis.

Dickey-Fuller test for unit root Number of obs = 57

Test Statistic	Interpolated Dickey-Fuller		
	1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-5.814	-3.570	-2.597

MacKinnon approximate p-value for Z(t) = 0.0000

Table 1:- Dickey-Fuller test for stationarity

Since the data is now stationary as evidenced above, we now turn to the second step where we need to estimate the parameters:

➤ Step 2: Parameter estimation

Looking at the shape of the autocorrelation function and partial autocorrelation function, our tentative ARIMA model will have the form *ARIMA(p, d, q)* where

$$d = 2, p \geq 0, \text{ and } q \geq 0$$

After trial and testing, the following models were adopted:

- arima(1,2,0)*
- arima(3,2,0)*
- arima(4,2,0)*
- arima(5,2,0)*

To select the suitable forecasting model, we chose the one with the lowest Akaike's Information Criterion(AIC) and Bayesian Information Criteria (BIC)

ARIMA MODEL	Parameters					
	Constant	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)
(1,2,0)	0.00168	-0.324218				
(3,2,0)	0.0002	-0.447584	-0.323541	-0.283149		
(4,2,0)	-0.0004	-0.503063	-0.382261	-0.371659	-0.1865273	
(5,2,0)	-0.0009	-0.563637	-0.495927	-0.483612	-0.3491418	-0.3033387

Table 2:- Parameter estimation and lag selection

From Table 2; we deduced that the ARIMA (5,2,0) model can best fit the data because of having the lowest AIC and BIC values (see table 3 below). It can best predict the future values of gross domestic product of Zambia.

ARIMA MODEL	Sigma	Loglikelihood	AIC	BIC	Significant figures
(1,2,0)	0.194539	12.38099	-18.76197	-12.63282	1
(3,2,0)	0.182165	15.95687	-21.91373	-11.69848	3
(4,2,0)	0.178824	16.93759	-21.87518	-9.61687	3
(5,2,0)	0.169797	19.6481	-25.2962	-10.99484	3

Table 3:- Model selection criterion using AIC and BIC

The ARIMA (5,2,0) was chosen for having the lowest AIC and volatility as shown by the value of sigma.

➤ Step3: Diagnostic checking

Having chosen a particular ARIMA model, and having estimated its parameters, we can now check whether the chosen model fits the data reasonably well, for it is possible that another ARIMA model might do the job as well. One simple test of the chosen model is to see if the residuals estimated from this model are white noise; if they are, we can accept the particular fit; if not, we must start over.

To test white noise validity, we can use Bartlett’s B statistic. Below is the graph showing the distribution of the disturbance terms:

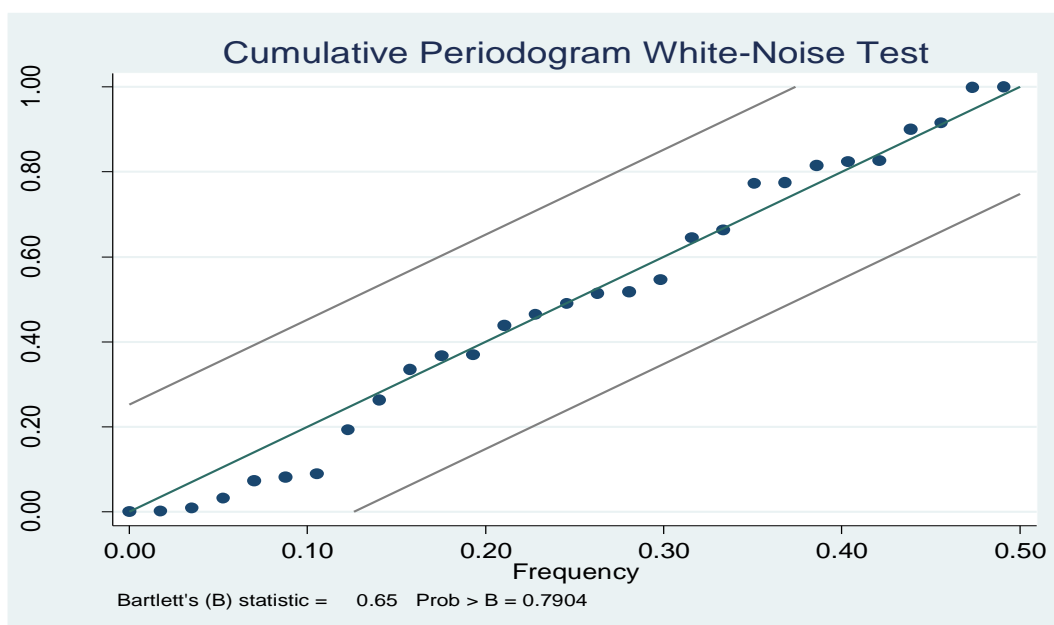


Fig 7:- Bartlett’s B statistic showing the distribution of the error terms.

The residuals are white noise and are all within the 95% confidence interval overtime as seen in Figure 7 using Bartlett’s B statistic. Another checking we can use is the Portmanteau test. The null hypothesis is that the residuals are not white noise and the alternative is; the residuals are white-noise. Table 4 below shows that, the Q statistic is statistically significant hence, we reject the null hypothesis and conclude that the residuals are white noise.

-----  
**Portmanteau (Q) statistic = 26.6610**  
**Prob > chi2(26) = 0.4273**  
 -----

Table 4:- Portmanteau Test for white noise

➤ Step 4: Forecasting the model

This study projected the gross domestic product for a period of 8 years up to the period of the year 2026. This period is particularly important because of major economic shocks like Covid-19, debt burden, droughts, volatility of the exchange rate and the pending 2021 general elections. Hence, the need to understand the business cycles of the country.

YEAR	GDP Forecasts (Billions)
2019	28.99369
2020	28.60539
2021	25.16788
2022	24.89649
2023	27.268
2024	28.44018
2025	30.3351
2026	30.3816

Table 5:- showing forecast values of GDP for the next eight years

To check how close the fitted values are, to the actual, figure 8 below shows the actual gross domestic product and their fitted counterpart.

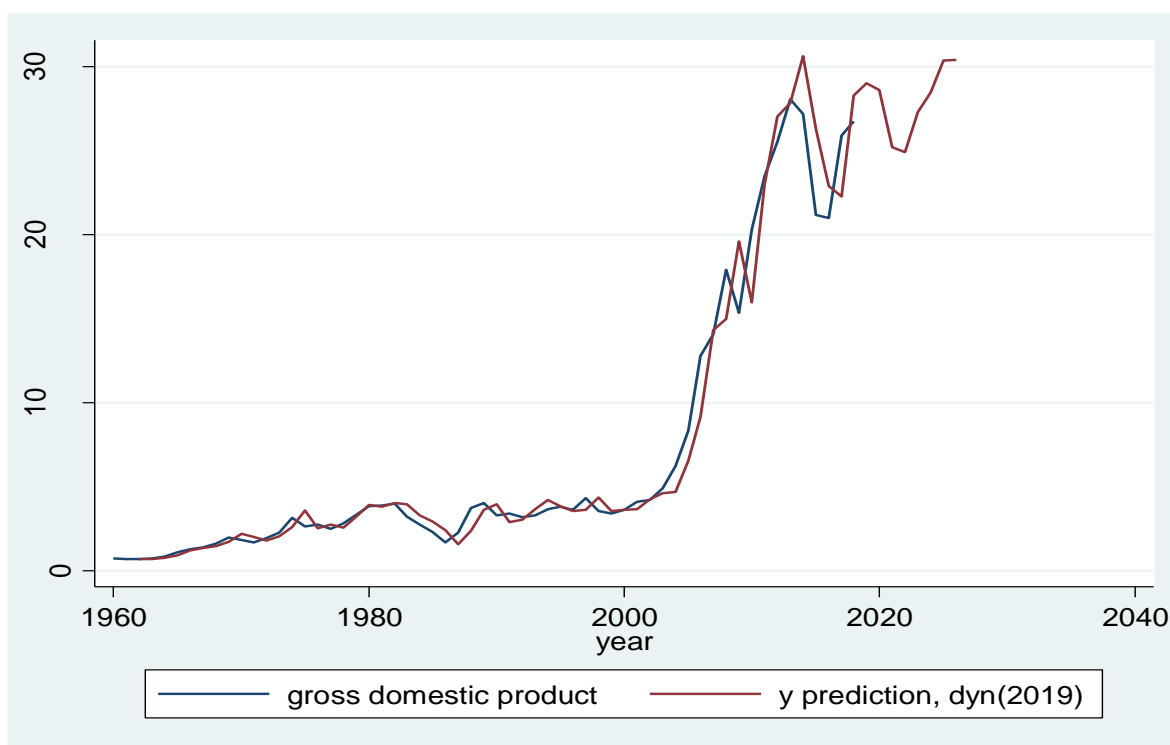


Fig 8:- Graph showing the actual GDP and the fitted values of GDP

To generalize the predicted values, it is hoped that, on average, the nominal gross domestic product will continue to rise over this period (see Figure 9 below). However, it is expected that around 2020 and 2022, there will be a sharp decline as seen in Table 6, showing the actual forecasted values of GDP.

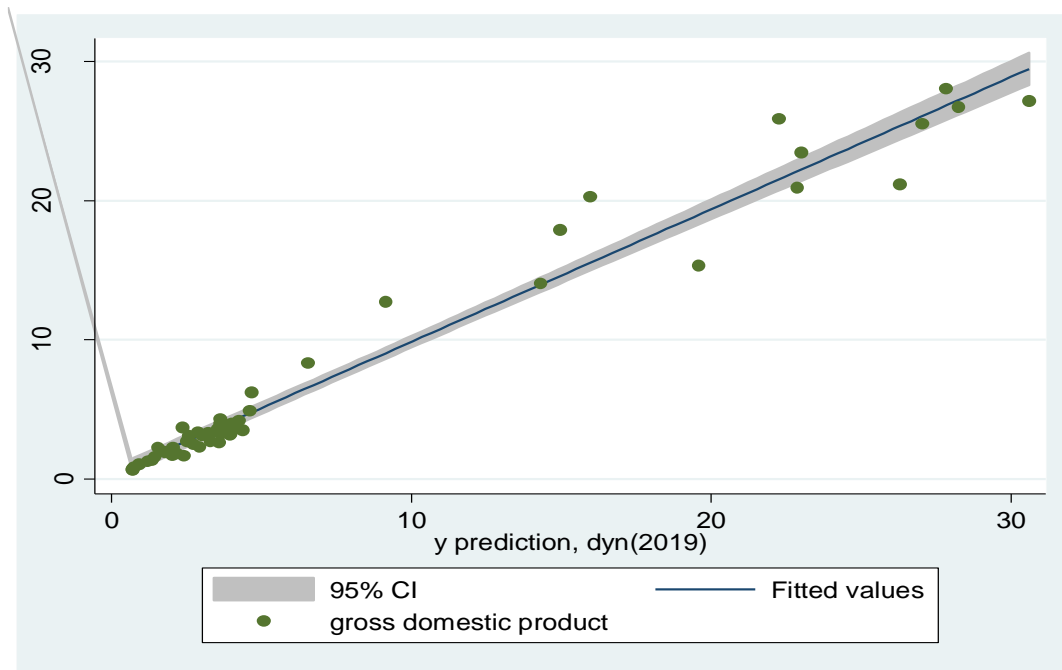


Fig 9:- predicted values of GDP

## V. CONCLUSION

From the data analyzed, it is expected that Zambia's gross domestic product will continue, on average, rising over the next few years except the between 2020 and 2023. This information is necessary as it will help stakeholders to put some safety nets that will cushion the economy in times of down-turns or recessions. The best ARIMA model that fit the data set was ARIMA (5,2,0). Most of the parameters are significant, meaning that the predicted values solely depended on past or lagged values of GDP.

One of the limitation of this study is that we only had access to yearly data and not quarterly data. And the data used may not be exact due to the rebasing the Zambian kwacha in 2012. Further, this analysis utilized only one dependent variable that was regressed on its past lags. There is need to see how different variables like inflation, gross domestic product, exchange rate can interact in the market. Hence, the need to investigate these relationship using VAR or VEC models in the next paper.

## REFERENCES

- [1]. Agrawal, V, (2018), GDP modeling and Forecasting using ARIMA, an empirical study from India, Central European University.
- [2]. Bipasha, M. and Bani, C.,(2012), Forecasting GDP Growth Rates Of India: An Empirical Study, International Journal of Economics and Management Sciences, Vol. 1, No. 9, 2012, pp. 52-58
- [3]. Contreras, J., Espinola, R., Nogales, F.J., Conejo, A.J., (2003), ARIMA Models to Predict Nextday Electricity Prices, IEEE Transactions on Power Systems, Vol. 18, No. 3, pp. 1014- 1020.
- [4]. Ljung, G.M. and Box, G.E.P. (1978), On a measure of lack of fit in time series models, *Biometrika*, Vol. 67, pp. 297-303.
- [5]. Mohad, R. A. and Ahmed, I. A., (2019), Forecasting Egyptian GDP using ARIMA models, *Reports on Economics and Finance*, Vol. 5, No. 1, pp. 35- 47.
- [6]. Raymond Y.C. Tse (1997), An application of the ARIMA model to real-estate prices in Hong Kong, *Journal of Property Finance*, Vol. 8, No. 2, pp.152 – 163.