



Modeling Nigeria's Unemployment Rates: Box-Jenkins Approach

Agog Nathan Samuel^{1*}; Bako Sunday Samuel¹; David Ikwuoche John²;
Peter Magdalene¹; Yahaya Jibril Kajuru³

¹Department of Mathematical Sciences, Kaduna State University, Nigeria.

²Department of Mathematics and Statistics, Federal University Wukari, Nigeria.

³Institute of Education, Ahmadu Bello University Zaria, Nigeria.

e-mail: nathan.agog@kasu.edu.ng^{1*}

Abstract

One of the key determinants of economic growth and stability of a nation is the macroeconomic policy of a government. In Nigeria, despite the macroeconomic reforms and prudent policies pursued by the government, such reforms and policies have not translated to increase in employment opportunities and poverty reduction. The aim of this paper is to find the most suitable time series model of annual unemployment rates in Nigeria using Box-Jenkins methodology and to examine the precision of forecast on this model. ARIMA (2, 2, 1), ARIMA (2, 2, 2), ARIMA (1, 2, 1) and ARIMA (1, 2, 2) were fitted to the unemployment data. The Akaike Information Criterion and the Bayesian information Criterion suggested ARIMA(1,2,1) as the best model, however, evaluating a five year forecast accuracy using different measures of accuracy suggested that ARIMA(2,2,2) minimizes the forecast error for the unemployment data. The projection made from 2020 to 2024 shows that unemployment rate will constantly increase in Nigeria if government at all levels do not implement economic policies that will translate to reduction in the unemployment rate. This projection does not take into cognizance any impact or effect of COVID-19 pandemic.

Keywords-Box-Jenkins, ARIMA, Box-Jenkins methodology, measures of accuracy, forecast error

1. Introduction

National Bureau of Statistics (NBS) in its Q2 report in August, 2020 indicates that Nigeria's unemployment rate has increased from 23.1% in 2018 to 27.1% in the second quarter of 2020. This indicates that about 21,764,617 (twenty one million, seven hundred and sixty four thousand, six hundred and seventeen) Nigerians are unemployed (NBS, 2020). Unemployment is very high in Nigeria and mostly prevalent among the youths. The government of Nigeria despite its willingness to tackle unemployment still finds it difficult to fulfill their aspirations. Lately, the Executive and Legislative arms of government in Nigeria were faced with the battle of who superintends over the employment of 774,000 temporary workers in the N-power scheme. Unemployment is one of the significant microeconomic variables. It is one of the most difficult economic problems facing the governments. High unemployment has both direct and indirect impacts on poverty, crime and socio-political problems that are also increasing (Nikolaos *et al.* 2018). According to Dritsaki (2016), ARIMA models are suitable for short run forecasts because they give more emphasis on the recent past rather than distant past. Etuk, Uchendu and Edema

(2012) applied the Box-Jenkins methodology on monthly data for the unemployment rates of Nigeria. They came up with ARIMA (1,2,1) as the best model for the data used. Adenomon (2017) used an ARIMA (2, 1, 2) model to forecast the unemployment rates of Nigeria. The researcher used the smallest value of mean square error (MSE) as an accuracy measure for identifying the most suitable model to forecast the unemployment rate. Dobre and Alexandrou (2008) used ARIMA (2, 1, 2) on a monthly data between 1998-2007 for Romania. The same model was used to forecast Romania's unemployment rate for the months of 2008. In the research carried out by Power and Gasser (2012), ARIMA (1, 1, 0) was selected as the best model for forecasting unemployment rates in Canada. Dritsaki (2016) applied the Box-Jenkins methodology where SARIMA (0,2,1)(1,2,1)₁₂ was selected and used to predict the monthly unemployment rate of Greek in both dynamic and static process. The Mean Squared Error (MSE), Mean Absolute Error (MAE) criteria and Theils' coefficient gave a good guide for the prediction accuracy. The aim of this paper is to find the most suitable time series model for annual unemployment rates in Nigeria using Box-Jenkins methodology and to examine the precision of the forecast using the estimated model.

2. Materials and Method

This study utilized unemployment rates data from 1976 to 2020 obtained from abstracts and annual reports of the Central Bank of Nigeria and Nigeria Bureau of Statistics (NBS) Statistical Bulletin. The ARMA or ARIMA model is used for forecasting. A typical ARMA model combines both the moving average (MA) model and the autoregressive (AR) model. The autoregressive moving average (ARMA) is given as

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \dots + \varphi_q \varepsilon_{t-q} \quad (1)$$

$$Y_t = \sum_{k=1}^p \alpha_k Y_{t-k} + \sum_{k=1}^q \varphi_k \varepsilon_{t-k} + \varepsilon_t \quad (2)$$

Equation (2) can be expressed as

$$\alpha(B)Y_t = \varphi(B)\varepsilon_t \quad (3)$$

where, $\alpha(B) = 1 - \alpha_1 B - \alpha_1 B^2 - \dots - \alpha_p B^p$ and $\varphi(B) = 1 - \varphi_1 B - \varphi_1 B^2 - \dots - \varphi_q B^q$.

If Z_t is a stationary series obtained after d differencing from Y_t series then we obtain

$$Z_t = \nabla^d Y_t = (1 - B)^d Y_t \quad (4)$$

If Z_t follows an ARMA (p, d) model, we can conclude that $\{Y_t\}$ is an ARIMA (p, d, q). Hence the expression of the ARIMA(p, d, q) model is given as

$$(1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p)(1 - B)^d Y_t = (1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_q B^q) \quad (5)$$

$$(1 - B)^d \alpha_p(B) Y_t = \varphi_q(B) \varepsilon_t \quad (6)$$

The expression in (6) consist of an Autoregressive (AR) process which handles memory from previous values, an integrated procedure which makes the data stationary and a Moving-Average (MA) which represents terms of previous error. The Box-Jenkins' technique applies the ARMA or ARIMA models to obtain the best model that fits the time-series data. The stages involved in Box-Jenkins technique are as follows;

2.1 Model Identification: In developing a Box–Jenkins model, it starts with determining whether the time series data is stationary or not. A time series data is said to be Stationary when the mean and the variance are constant over a period of time. To make the data stationary, it is recommended to difference the data. The ACF and PACF is plotted immediately the data becomes stationary, this is carried out in order to identify the order of the ARMA model (i.e. p and q).

2.2 Parameter Estimation: The methods of Maximum likelihood estimation (MLE), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are applied for the parameter estimation. The maximum likelihood estimation (MLE) method obtains values of the parameters that maximize the probability of obtaining the actual value of the parameter we are estimating. The likelihood function is as follows;

$$\log L = -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^T \epsilon_t^2 \quad (7)$$

where, T = time, $t = 1, 2, 3, \dots, T$, ϵ_t = error and σ^2 = constant variance.

The log likelihood with maximal value is selected as the best model. The Akaike Information Criterion (AIC) is useful for comparing models fitted to the same series. It is written as

$$AIC = -2 \ln L + 2N \quad (8)$$

where L = likelihood of the data and N = number of parameters in the model.

The Bayesian Information Criterion (BIC) penalizes the parameters more strongly than the AIC. It is written as

$$BIC = AIC + N \log(T) \quad (9)$$

ARIMA model that has the lowest AIC, AICc, BIC with largest log likelihood will be the preferred model.

3.3 Diagnostic Model Assessment: The ACF, PACF, chi-square and the Ljung-Box Q statistics obtained from the residual of the ARIMA model are used for diagnostic checking. The model is considered fit if the residual satisfies the assumption of regression model and assumptions for a stationary process. The residuals should be white noise.

3.4 Forecasting: Accuracy measures such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Square Error (MASE) are used for measuring the performance accuracy of the ARIMA model. Any of the temporary fitted ARIMA models which has the smallest statistics is considered the best model for forecasting.

$$MASE = \frac{1}{N} \sum_{t=1}^N |\hat{Y}_t - Y_t|^2 \quad (10)$$

$$MAE = \frac{1}{N} \sum_{t=1}^N |\hat{Y}_t - Y_t| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{Y}_t - Y_t)^2} \quad (12)$$

where, Y_t = actual value of variable Y at time t , $\hat{Y}_t$ = predicted value of Y at time t , N = number of observations.

3. Results and Discussion

Figure 1 shows the time series plots of unemployment rates in Nigeria from 1972-2020. The plot of the observations displays the series on the y-axis against the time interval on the x-axis showing the patterns in data over time.

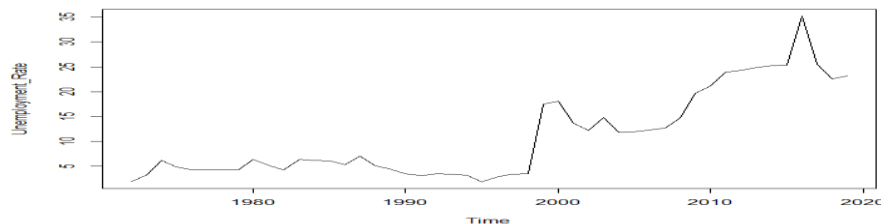


Figure 1: Time Series Plot of Unemployment Rates in Nigeria.

The plot of the series reveals several fluctuations across the years indicating significant variation over time. The data attains stationarity after the second difference, with the Augmented Dickey-Fuller Test (Dickey-Fuller = -2.0844, Lag order = 3, p-value = 0.5404). Since the p-value is greater than the significant level at 0.05 we therefore fail to reject the null hypothesis that Unemployment rate data is not stationary.

Test for Stationarity

The unemployment rate data attains stationarity after the second difference, with the Augmented Dickey-Fuller Test (Dickey-Fuller = -4.9195, Lag order = 3, p-value = 0.01). Since the p-value is less than 0.05, we therefore reject the null hypothesis and accept the alternative hypothesis that the series data for unemployment rate stationary.

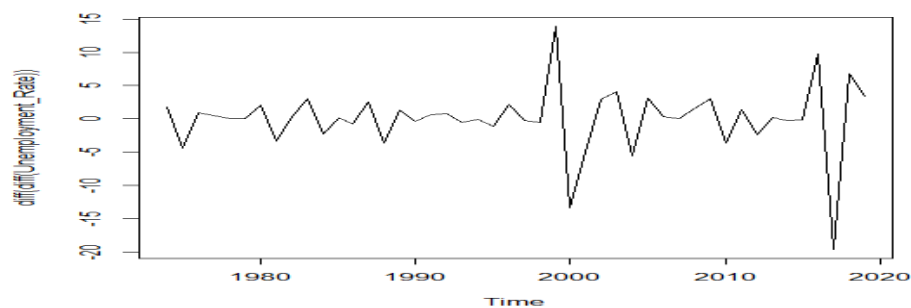


Figure 2: Time Series Plot for the differenced data

Model Identification

Since the data for the unemployment rate became stationary at the second lag difference, it means that the model to be considered is of this form ARIMA (p, 2, q).

In Figure 3, the plots of the autocorrelation and partial autocorrelation function of the stationary time series data are examined. The ACF decay exponentially after lag 1 and the PACF dies off after two significant spikes at lag 1 and 2, suggesting an ARIMA (2, 2, 1).

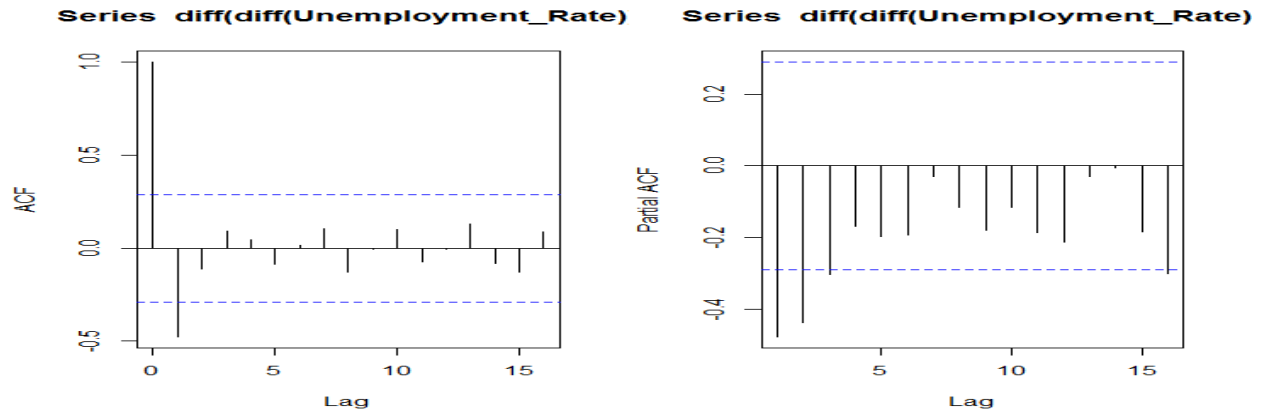


Figure 3: ACF and PACF of the Second differenced data of Unemployment rate in Nigeria.

3.2 Model Estimation

Based on the ACF and PACF, four tentative models are chosen and they include; ARIMA (2, 2, 1), ARIMA (2, 2, 2), ARIMA (1, 2, 1) and ARIMA (1, 2, 2). These tentative ARIMA models are assessed by choosing the least AIC, log likelihood and BIC criteria to serve as the best ARIMA model.

Table 1: Best ARIMA model for Unemployment Rate Based on Information Criteria values

Variable	ARIMA(2,2,1)	ARIMA (2,2,2)	ARIMA(1,2,1)	ARIMA(1,2,2)
AIC	250.39	252.38	249.75	251.46
AICc	251.37	253.88	250.32	252.44
BIC	257.71	261.53	255.23	258.78
Log likelihood	-121.2	-121.19	-121.87	-121.73

The results in Table 1 indicates that the best model is ARIMA (1, 2, 1), since it has the smallest values of AIC, AICc, BIC and largest value of log likelihood. ARIMA (1, 2, 1) is selected as the best model among the tentative models.

Table 2: ARIMA (1, 2, 1) model for Unemployment Rate in Nigeria

Coefficients	AR1	MA1
	-0.0958	-1.0000
SE	0.1463	0.0655
Sigma ² estimated as 10.73: log likelihood=-121.87		
AIC=249.75 AICc=250.32 BIC=255.23		

Table 2 shows the coefficient of the parameters for the selected model, ARIMA (1, 2, 1).

3.3 Diagnostic Checking

The diagnostic checking analyzes the residuals as well as model comparisons. The Standardized residuals are independent identically distributed sequence with zero mean and variance one. The ACF and PACF of the standardized residuals show no significant lags. Also, the Ljung Box result ($\chi^2 = 9.5085$, $df = 12$, $p\text{-value} = 0.659$) shows that the residuals are stationary and constitute white noise. Thus the selected model is appropriate to represent the series and it is a good fit.



Fig 4 Residual plot of Unemployment Rates in Nigeria from 1972-2019

3.4 Forecasting Unemployment Rates in Nigeria

This stage examines any of the tentative ARIMA models with the smallest RMSE, MAE and MASE on the estimated parameter results. The model with the smallest statistics is used for forecasting.

Table 3: Accuracy Measures

Variable	ARIMA(2,2,1)	ARIMA (2,2,2)	ARIMA(1,2,1)	ARIMA(1,2,2)
RMSE	3.1471	3.1453	3.2069	3.1764
MAE	1.7928	1.7910	1.8128	1.7918
MASE	0.9878	0.9869	0.9988	0.9873

From the four tentative ARIMA models, it shows that ARIMA (2,2,2) is the best model for forecasting unemployment rate in Nigeria. The parameters of ARIMA (2, 2, 2) from the result of the analysis are; $\alpha_1 = -0.0221$, $\alpha_2 = -0.1630$, $\varphi_1 = -1.1020$ and $\varphi_2 = 0.1020$. These parameters are used for modeling the unemployment rates. ARIMA(2,2,2) model can be obtained from equation (6) as follows;

$$Y_t = (2 + \alpha_1)Y_{t-1} - (1 + 2\alpha_1 - \alpha_2)Y_{t-2} + (\alpha_1 - 2\alpha_2)Y_{t-3} + \alpha_2Y_{t-4} + \epsilon_t + \varphi_1\epsilon_{t-1} + \varphi_2\epsilon_{t-2}$$

$$Y_t = 1.9779Y_{t-1} - 1.1188Y_{t-2} + 0.0221Y_{t-3} - 0.0221Y_{t-4} + \epsilon_t - 1.102\epsilon_{t-1} + 1.102\epsilon_{t-2}$$

(13)

Future unemployment rates of Nigeria are forecasted using equation (13).

Table 4: Five Years Forecast of Unemployment Rate in Nigeria

Years	Point Forecast	Lower 95% Limit	Upper 95% Limit
2020	24.2981	17.6368	30.9593
2021	24.7245	15.7767	33.6724
2022	25.0542	14.7918	35.3167
2023	25.5118	13.9714	37.0523
2024	25.9824	13.1783	38.7865

ARIMA (2,2,2) model was used to forecast unemployment rate of Nigeria for 5 years from 2020 to 2024. Table 4 shows the forecast of unemployment rates in Nigeria from 2020 to 2024. The result of the forecast shows a gradual increase in the number of unemployment rates in Nigeria.

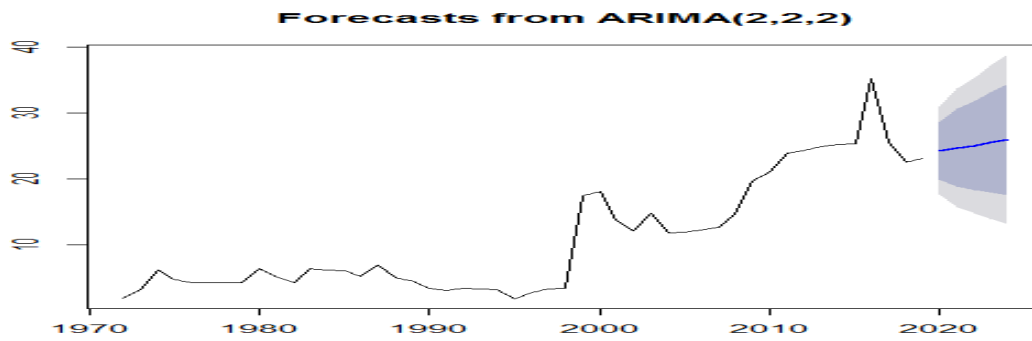


Figure 5: Time series Plot of Unemployment Rate in Nigeria and its Forecast and Confidence Interval

4. Conclusion

This study presented the time series plot of the unemployment rate in Nigeria. The time series plot indicated that the mean and variance were not constant, hence the need for differencing of the observed data to attain stationarity. The data attain stationarity after the second difference. Four tentative time series models were proposed (based on ACF and PACF) where ARIMA (2, 2, 2) was observed to be the best time fit model for forecasting unemployment rate in Nigeria. The model is given as;

$$Y_t = 1.9779Y_{t-1} - 1.1188Y_{t-2} + 0.0221Y_{t-3} - 0.0221Y_{t-4} + \epsilon_t - 1.102\epsilon_{t-1} + 1.102\epsilon_{t-2}.$$

The model was considered the best since the values of RMSE, MAE and MASE were the smallest as compared to other models. ARIMA (2,2,2) was used to make projection of unemployment rate in Nigeria for the period of five (5) years. The projection made from 2020 to 2024 shows that unemployment rate will constantly increase in Nigeria if government at all levels did not strategize ways to tackle unemployment.

References

- Adenomon, M.O. (2017). Modeling and Forecasting Unemployment rates in Nigeria using ARIMA model. *FUW Trends in Science & Technology Journal*, 2(1B): 525-531
- Dobre, I. and Alexandrou, A.A. (2008). Modelling unemployment rate using Box-Jenkins procedure. *Journal of Applied Quantitative Methods*, 3(2): 156-166.
- Dritsaki, C. (2016). Forecast of SARIMA Models: An application of unemployment rates of Greece. *American Journal of Applied Mathematics and Statistics*, 4(5): 136-148.
- Etuk, E.H., Uchendu, B. and Edema, U.V. (2012). ARIMA fit to Nigerian unemployment data. *Journal of Basic and Applied Scientific Research*, 2(6): 5964-5970
- National Bureau of Statistics (2020). Labour Force Statistics: Unemployment and Underemployment Report (August Q2, 2020).
- Nkwatoh, L. S. (2012). Forecasting Unemployment Rates in Nigeria Using Univariate Time Series Models, *International Journal of Business and Commerce*, 1(12): 33-46.
- Nikolaos, D., Stergios, A., Tasos, S. and Ioannis, S. (2018). Forecasting unemployment rates in Greece. *International Journal of Sciences; Basic and Applied Research*, 37(1):43-55.
- Power, D and Gasser, K. (2012). Forecasting Future Unemployment Rates. ECON 452 First Report.