



On Modeling Murder Crimes in Nigeria

Obubu Maxwell*

Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria

Ikediwa Udoka Chinedu

Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria

Anabike Charles Ifeanyi

Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria

Nwokike Chukwudike C.

Department of Statistics, Abia State University, Nigeria

Abstract

This paper examines the modelling and forecasting Murder crimes using Auto-Regressive Integrated Moving Average models (ARIMA). Twenty-nine years data obtained from Nigeria Information Resource Center were used to make predictions. Among the most effective approaches for analyzing time series data is the method propounded by Box and Jenkins, the Autoregressive Integrated Moving Average (ARIMA). The augmented Dickey-Fuller test for unit root was applied to the data set to investigate for Stationarity, the data set was found to be non-stationary hence transformed using first-order differencing to make them Stationary. The Stationarities were confirmed with time series plots. Statistical analysis was performed using GRETL software package from which, ARIMA (0, 1, 0) was found to be the best and adequate model for Murder crimes. Forecasted values suggest that Murder would slightly be on the increase.

Keywords: Crime; Forecasting; ARIMA Model; Murder; Box-jenkins method; Akaike information criteria; Bayesian information criteria; Hannan-quinn criteria.



CC BY: [Creative Commons Attribution License 4.0](https://creativecommons.org/licenses/by/4.0/)

1. Introduction

Crime prevention is one of the important roles of the police system in any country [1]. One of the components of crime prevention is crime forecasting. Police will require crime forecasting to make operational and tactical strategies in the future, as like to allocate police patrols in the right area, install CCTV in the right place and plan other operations [2]. In a strict legal definition, a crime is a violation of criminal law, which in most societies can be defined broadly as any 'act or omission forbidden by law on pain of punishment'. The crime forecasting is an emerging approach in criminological research [3, 4]. Crime forecasting is not widely practised by police [5]. While there are numerous econometric studies of crime, one is hard-pressed to find police departments or other police organizations making regular use of forecasting for policing. From a more proactive standpoint, problem-oriented policing efforts may be enhanced by a more accurate scanning of areas with crime problems, in that one can examine both distributions of past crimes and predictions of future concentrations. The ability to predict can serve as a valuable source of knowledge for law enforcement agencies, both from tactical as well as strategic perspectives [6]. Forecasting can help a police department's performance by strategic deployment efforts and efficient investigation direction. The purpose of this paper is to examine the modelling and forecasting of murder rate using Auto Regressive Integrated Moving Average Models (ARIMA).

2. Materials and Method

In this paper, we have used the Crime data on Murder Cases for past 29 years (1986 -2015). The crime data were sourced from Nigeria Information Resource Centre. We have used GRETL (Gnu Regression, Econometrics and Time-series Library) software for plotting the graphs and analysis of the data set.

2.1. Time Series

Time series can be defined as a collection of observations or measurements on quantitative variables made sequentially or in a uniform set of time period, usually daily, weekly, monthly, quarterly, annually, e.t.c Time series analysis comprises methods or processes that break down a series into components and explainable portions that allow trends to be identified, estimates and forecasts to be made. Basically, time series analysis attempts to understand the underlying context of the data points through the use of a model to forecast future values based on known past values. Such time series models include ARIMA, GARCH, TARCH, EGARCH, FIGARCH, and CGARCH and so on but the main focus of this study is based on ARIMA model.

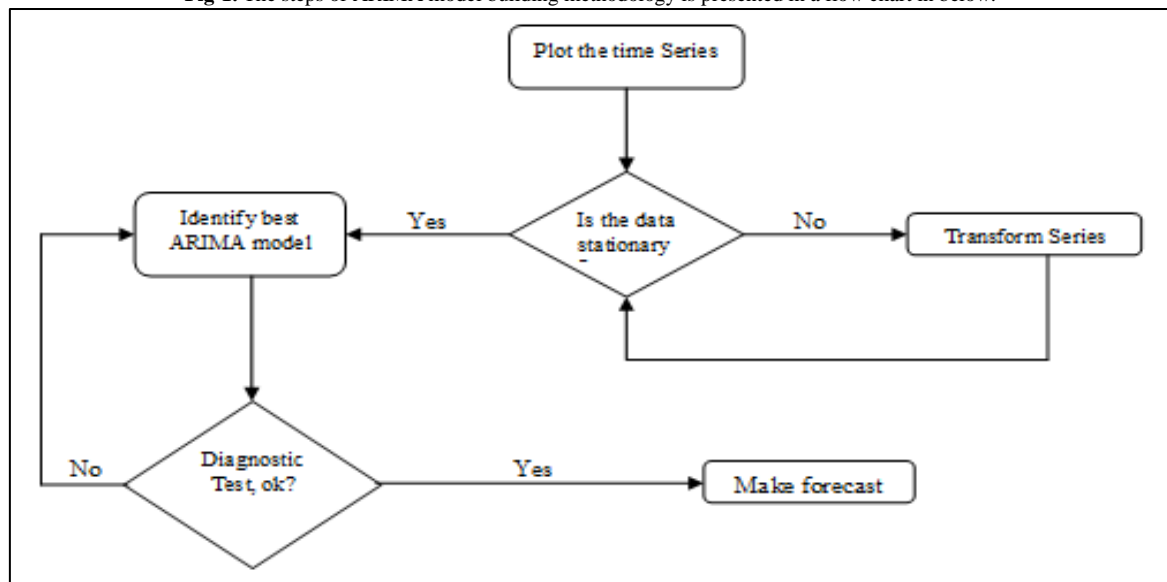
2.2. ARIMA (Autoregressive Integrated Moving Average)

ARIMA model was introduced by Box and Jenkins (hence also known as Box-Jenkins model) in 1960. It is an extrapolation method for forecasting and, like any other such method, it requires only the historical time series data on the variable under-forecasting. ARIMA models are the most general class of models for forecasting a time series [7, 8]. Normally, the ARIMA model is represented as $ARIMA(p,d,q)$ where p is the number of autoregressive terms, d is the number of non-seasonal differences, and q is the number of lagged forecast errors in the prediction equation. The identification of the appropriate ARIMA model for a time series begins with the process of finding integer, usually very small (e.g., 0, 1, or 2), values of p , d , and q that model the patterns in the data. When the value is 0, the element is not needed in the model. The middle element, d , also known as trend component is investigated before p and q . The goal is to determine if the process is stationary and, if not, to make it stationary before determining the values of p and q . The augmented Dickey–Fuller (ADF) test is the most widely used test for checking the stationarity of a series. If $d = 0$, the model becomes ARMA, which is a linear stationary model. ARIMA (i.e. $d > 0$) is a linear non-stationary model [9]. If the underlying time series is non-stationary, taking the difference of the series with itself 'd' times makes it stationary, and then ARMA is applied onto the differenced series. A stationary process has a constant mean and variance over the time period. There are various methods available to make a time series stationary. Normally differencing techniques are used to transform a time series from a non-stationary to stationary by subtracting each datum in a series from its predecessor. If a series is stationary without any differencing it is designated as $I(0)$, or integrated of order 0. On the other hand, a series that has stationary first differences is designated $I(1)$, or integrated of order 1. The term 'shock' is used to indicate an unexpected change in the value of a variable (or error). For a stationary series, a shock will gradually die away. In other words, the effect of a shock during time 't' will have a smaller effect on time 't+1', a still smaller effect in time 't+2', etc. The lags of the differenced series appearing in the forecasting equations are called "auto-regressive" terms. The auto-regressive components represent the memory of the process for preceding observations. The value of p is the number of auto-regressive components in an ARIMA (p, d, q) model. The value of p is 0 if there is no relationship between adjacent observations. When the value of p is 1, there is a relationship between observations at lag 1 and the correlation coefficient ϕ_1 is the magnitude of the relationship. When the value of p is 2, there is a relationship between observations at lag 2 and the correlation coefficient ϕ_2 is the magnitude of the relationship. Thus p is the number of correlations we need to model the relationship. The lags of the forecast errors are called "moving average" terms. The moving average components represent the memory of the process for preceding random shocks. The value q indicates the number of moving average components in an ARIMA (p, d, q). When q is zero, there are no moving average components [10-12]. When q is 1, there is a relationship between the current score and the random shock at lag 1 and the correlation coefficient θ_1 represents the magnitude of the relationship. When q is 2, there is a relationship between the current score and the random shock at lag 2, and the correlation coefficient θ_2 represents the magnitude of the relationship. When one of the terms is zero, it's usual to drop AR, I or MA component. For example, $I(1)$ model is $ARIMA(0,1,0)$, and a $MA(1)$ model is $ARIMA(0,0,1)$. The autocorrelation function (ACF) and partial correlation function (PACF) are very important for the definition of the internal structure of the analyzed series. The models can be identified through patterns in their autocorrelation functions (ACFs) and partial autocorrelation functions (PACFs).

2.3. Steps of ARIMA Model

- (i) Identification of ARIMA (p,d,q) structure
- (ii) Estimating the coefficients of the formulation
- (iii) Fitting test on the estimated residuals
- (iv) Forecasting the future outcomes based on the historical data

Fig-1. The steps of ARIMA model building methodology is presented in a flow chart in below.



3. Results and Discussion

Fig-2. Time Series plot for Murder data series

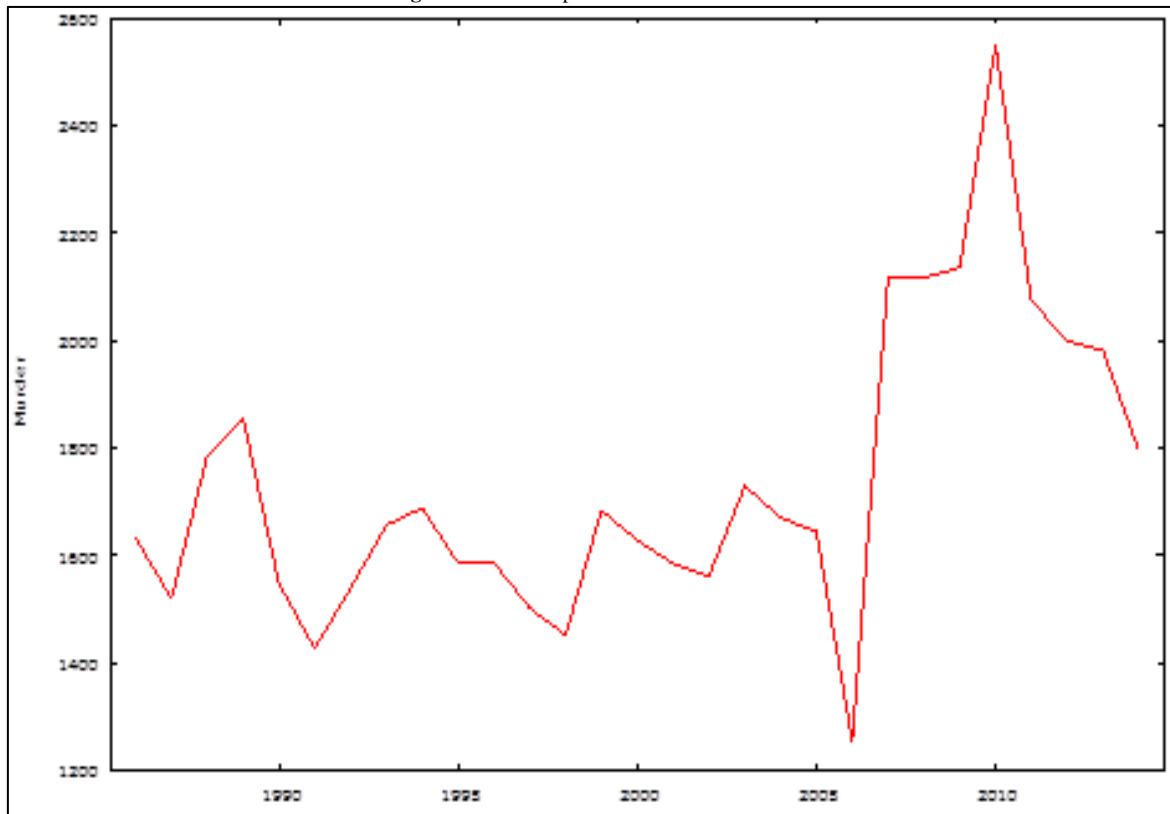


Fig-3. Correlogram for murder data series

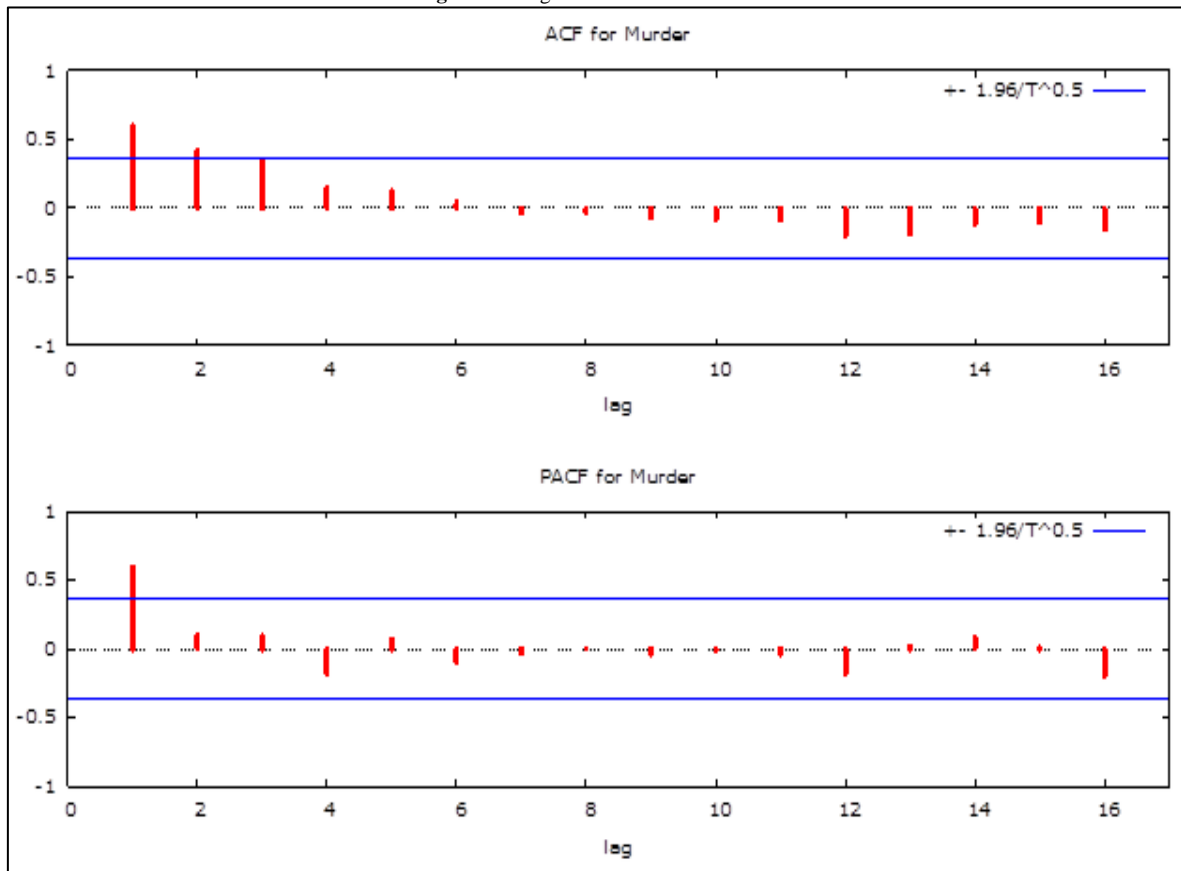


Table-1. Augmented Dickey-Fuller test for stationarity of Murder data series

D	t-statistic	p-value	α -value
0	-0.2427	0.5897	0.05
1	-6.952	4.27e-08	0.05

Fig-4. Time Series plot for differenced Murder data series

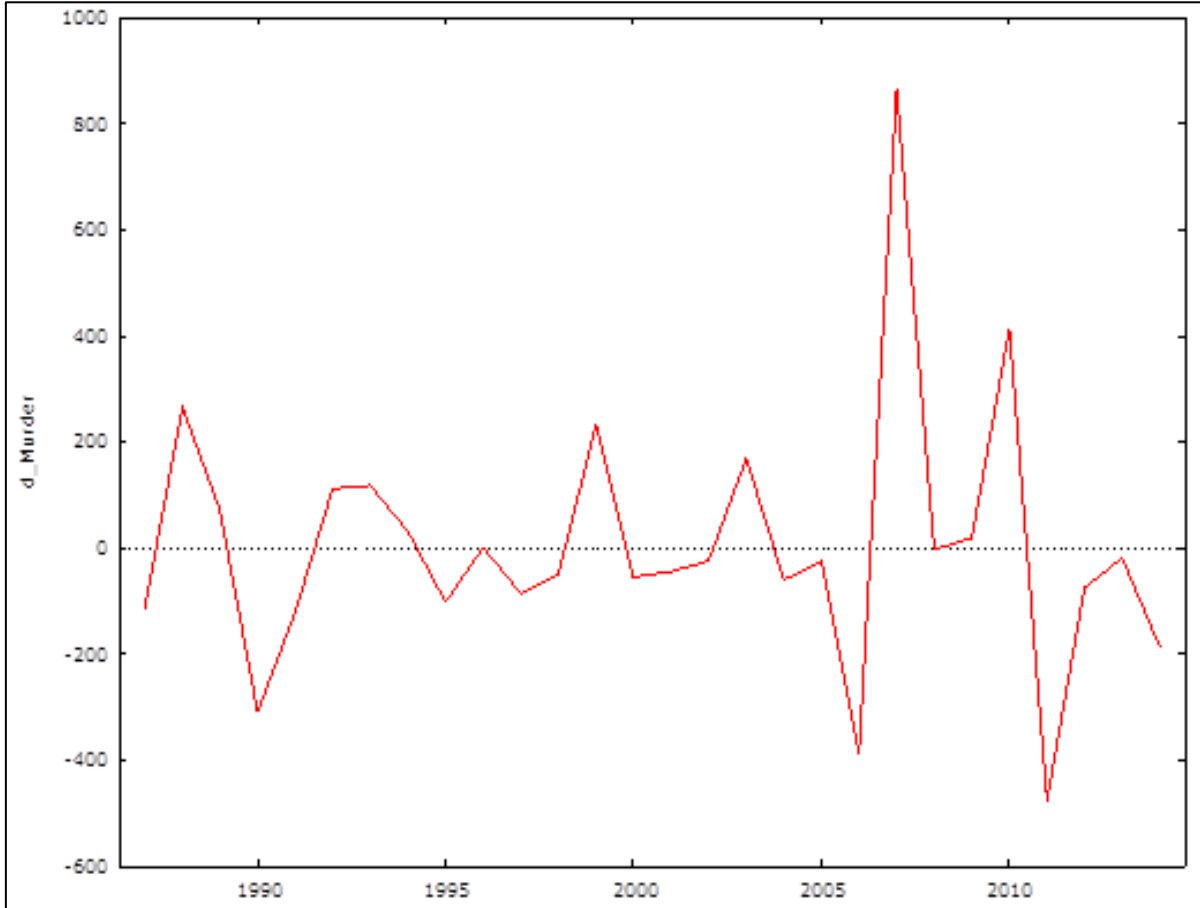
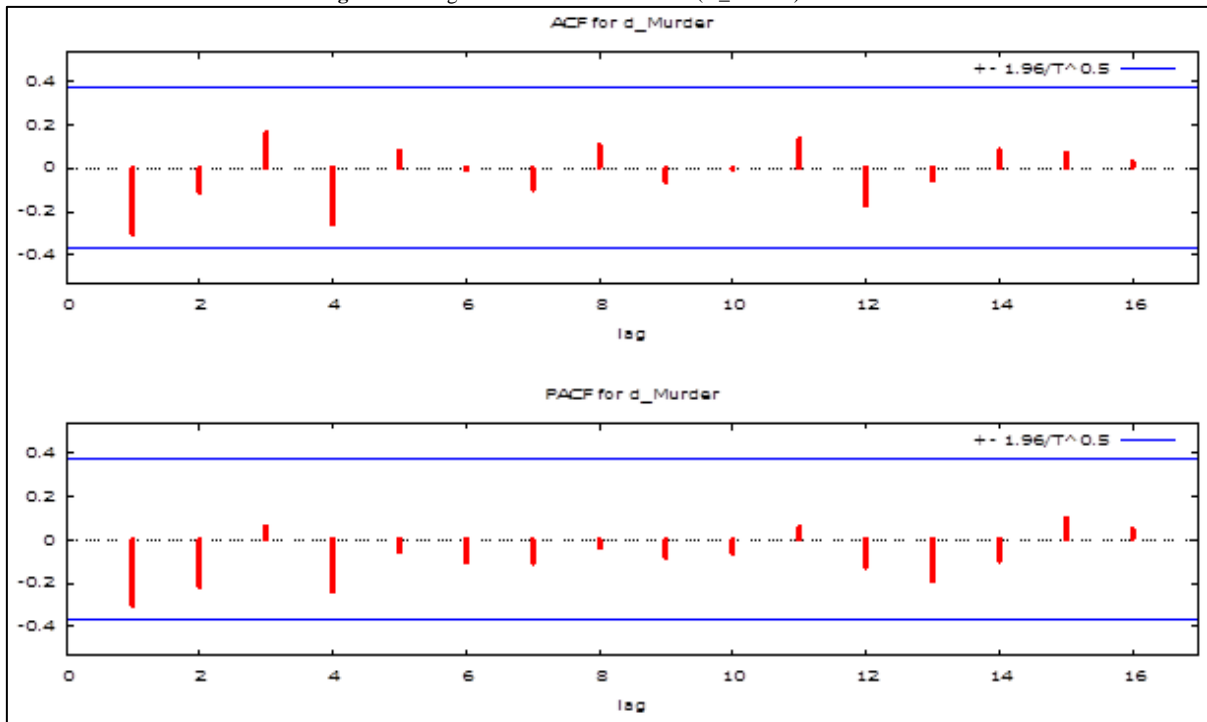


Fig-5. Correlogram for differenced murder (d_Murder) data series



The time series plot and correlogram of the murder data series on [fig 2](#) and [fig 3](#) respectively show a strong evidence of a non stationary series, the non stationarity was also confirmed with the help of the augmented dickey-fuller (ADF) test on [table 1](#), which tests the null hypothesis that murder data series follows a unit root process which was accepted at 5% alpha level. Implying that the murder data series is non-stationary. By using first order differencing transformation, we obtained a t-statistic lesser than what was obtained at $d = 0$, and a p-value lesser than 5% alpha level. Thus, we select the condition that $d = 1$ and transform the data using first-order differencing to make it stationary as seen in [fig 4](#). The Auto-Correlation Function (ACF) and Partial Auto-Correlation Function (PACF) plot of the differenced series is shown in [fig 5](#).

Table-2. Identification of Best ARIMA model for Murder

ARIMA Model	AIC	BIC	HQC
101	401.8915	407.3607	403.6044
010	389.4619	390.7941	389.8692
110	390.7801	394.7768	392.0019
011	389.4489	393.4455	390.6707
111	388.7527	392.7494	389.9745
210	389.4914	393.4881	390.7132
211	390.8039	396.1327	392.4330

Akaike information criteria (AIC), Bayesian information criteria (BIC), and Hannan-Quinn criteria (HQC).

Seven tentative models were entertained, and the model with the minimum AIC, BIC and HQC was chosen as the best model. To verify that the chosen ARIMA (0,1,0) is an appropriate model for murder, a diagnostic check is done using residual ACF/PACF plot at different lags and testing the significance of the correlations up to 16 lags by Q statistic and respective p-values [13, 14].

3.1. Diagnostic Check on the Best Model for Murder / Model Verification

Fig-6. Correlogram of residuals for Murder

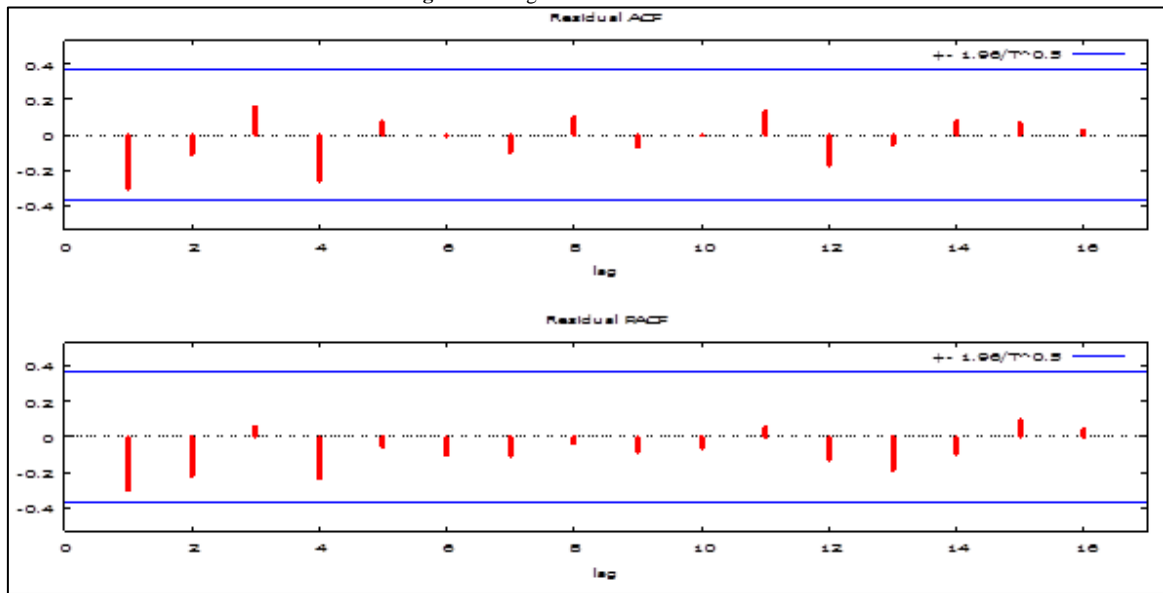


Fig 6 above clearly shows evidence of random walk as the values are within the bounds and undulate about zero. Hence we uphold the first order differencing.

Table-3. Residual autocorrelation function

LAG	ACF	PACF	Q-stat.	[p-value]
1	-0.3024	-0.3024	2.8458	[0.092]
2	-0.1096	-0.2213	3.2338	[0.199]
3	0.1607	0.0617	4.1014	[0.251]
4	-0.2583	-0.2373	6.4366	[0.169]
5	0.0769	-0.0566	6.6529	[0.248]
6	-0.0070	-0.1024	6.6547	[0.354]
7	-0.0987	-0.1080	7.0448	[0.424]
8	0.1042	-0.0397	7.5007	[0.484]
9	-0.0678	-0.0862	7.7041	[0.564]
10	-0.0055	-0.0646	7.7056	[0.658]
11	0.1335	0.0583	8.5860	[0.660]
12	-0.1738	-0.1264	10.1721	[0.601]
13	-0.0555	-0.1912	10.3445	[0.666]
14	0.0826	-0.0981	10.7536	[0.705]
15	0.0690	0.0998	11.0615	[0.748]
16	0.0318	0.0449	11.1321	[0.801]

Clearly, none of these correlations is significantly different from zero at a reasonable level. The ACF and PACF of the residuals also indicate a good fit of the model. This proves that the selected ARIMA model is an appropriate model.

Table-4. Crime Forecasting for Murder using ARIMA (0,1,0) For 95% confidence intervals, $z(0.025) = 1.96$

Obs	Murder	Prediction	Std. Error	95% interval
2016	undefined	1806.57	352.405	(1115.87, 2497.27)
2017	undefined	1812.36	431.606	(966.424, 2658.29)
2018	undefined	1818.14	498.376	(841.344, 2794.94)
2019	undefined	1821.65	501.624	(761.401, 2885.92)
2020	undefined	1824.55	505.124	(604.361, 2947.44)

From the forecast on Table 4, we see that murder would increase in 2019, and in subsequent years.

4. Conclusion

ARIMA (0,1,0) has been successfully used to forecast Murder crimes in Nigeria using a twenty-nine-year data series. The murder was found to be on the increase in the forecasted period. Hence, the government should endeavour to sanitize the Nigerian Police Force and other security agencies. The role of the police as enshrined in the Constitution of the Federal Republic of Nigeria, as it concerns crime prevention and control cannot be over-emphasized, the development of a society largely depends on the rate of crime. If the crime rate is high, it could scare away or discourage investors. Also, the safety of lives is an issue of priority in any given society. Consequently, the government needs to increase its funding for the police so that modern and adequate equipment to combat crime could be acquired. However, such funds are to be properly monitored to avoid diversion into private pockets and also rid the Nigerian police of its corrupt elements. The family institution must also play its role by monitoring all the people in the family particularly the youth, as the police cannot fight Murder crime alone. Intelligence gathering is a veritable tool for crime prevention and control. This is why the members of the public must wake up to assist the police in its functions. The public should be ready at all times to give necessary information to the police. This is because crime is a challenge that all and sundry must confront headlong for a stable and safe society.

References

- [1] Agyemang, B., 2012. *Autoregressive integrated moving average (ARIMA) Intervention analysis model for the major crimes in Ghana (the case of eastern region)*. M. Sc Thesis for Masters of Philosophy in Applied Mathematics, Nkrumah University of Science and Technology Ghana.
- [2] Anand, K. S. and Ekata, A. K. S., 2012. "Applicability of box jenkins ARIMA model in crime forecasting: A case study of counterfeiting in Gujarat State." *Int. J. Advance Research in Computer Engineering and Technology*, vol. 1, pp. 494-497.
- [3] Box, G. E. and Jenkins, G. M. *Time series analysis: Forecasting and control*. San Francisco: Holden Day.
- [4] Bowerman, B. L., Connell, R. T., and Koehler, A. B., 2005. *Forecasting time series and regression: An applied approach*. Belmont, CA: Thomson.
- [5] Chen, P., Yuan, H., and Shu, X., 2008. "Forecasting crime using ARIMA model." In *Proceedings of the 5th International Conference on Fuzzy Systems and Knowledge Discovery, October 18-20, 2008, Jinan, Shandong, China*. pp. 627-630.
- [6] Gorr, W., Olligschlaeger, A., and Thompson, Y., 2003. "Short term forecasting of crime." *Int. J. Criminol. Sociol. Theory*, vol. 4, pp. 648-656.
- [7] Khashei, M., Bijari, M., and Ardali, G. A. R., 2009. "Improvement of auto-regressive integrated moving average models using fuzzy logic and artificial neural networks (ANNs)." *Neurocomputing*, vol. 72, pp. 956-967.
- [8] Obubu and Nwokolo, P. C., 2016. "Modeling and forecasting of armed robbery rate in Nigeria using autoregressive integrated moving average (ARIMA) models." *Journal of Natural Science Research (IISTE)*, vol. 8, pp. 5-15.
- [9] Obubu, Oyafajo, O. A., Anyawu, I. F., and Olayemi, J. I., 2019. "Modeling typhoid mortality with box-jenkins autoregressive integrated moving average models." *Scholars Journal of Physics, Statistics, and Mathematics*, vol. 6, pp. 29-34.
- [10] Maxwell, O., Friday, A. I., and Chukwudike, N. C., 2019. "A theoretical analysis of the odd generalized exponentiated inverse Lomax distribution." *Biom Biostat Int. J.*, vol. 8, pp. 17-22.
- [11] Osuji, G. A., Obubu, M., and Nwosu, C. A., 2016. "Stock investment decision in Nigeria; A pc approach." *World Journal of Multidisciplinary and Contemporary Research*, vol. 2, pp. 1-11.
- [12] Osuji, G. A., Obubu, M., and Nwosu, C. A., 2016. "Preconception sex selection using proper ovulation timing." *World Journal of Probability and Statistics Research*, vol. 2, pp. 1-12.
- [13] Osuji, G. A., Okoro, C. N., Obubu, M., and Obiora-Ilouno, H. O., 2016. "Effect of akaike information criterion on model selection in analyzing auto-crash variables." *International Journal of Sciences: Basic and Applied Research*, vol. 26, pp. 98-109.
- [14] Maxwell, O., Happiness, O. I., Alice, U. C., and Chinedu, I. U., 2018. "An empirical assessment of the impact of Nigerian all share index, market capitalization, and number of equities on gross domestic product." *Open Journal of Statistics*, vol. 8, pp. 584-602. Available: <https://doi.org/10.4236/ojs.2018.83038>