Comparative Study of Time Domain Models in the Analysis of Aggregated Crime Rate in Nigeria

N. O. Adeboye[1]; K. O Omosanya[2]; D. A. Agunbiade[3]

[1]Department of Mathematics and Statistics,
Federal Polytechnic, Ilaro, Nigeria.
e-mail: nureni.adeboye@federalpolyilaro.edu.ng

[2]Department of Statistics,
Federal School of Statistics,
Ibadan, Nigeria
e-mail: olalekanomosanya90@gmail.com

[3]Department of Mathematical Sciences
Olabisi Onabanjo University,
Ago-Iwoye, Nigeria

Abstract — **The high rate of crime in Nigeria is a serious concern for the government and individuals because of its economic implication, serious damages and likely loss of lives and properties. Nigeria is one of the countries with a high rate of crime because of unemployment, economic hardship and crime friendly environment. It has, therefore, become imperative to ascertain the appropriate time-domain modeling for the crime rate in the country. As a result of this, ARIMA and SARIMA models were examined on the aggregated crime committed in Nigeria between the years 2005 and 2015. It was discovered that the data was not stationary and stationarity was achieved through 2nd order differencing and 2nd order seasonal differencing for ARIMA and SARIMA models respectively. Augmented Dickey-Fuller test was then performed on the differenced series in order to check for its acceptability. The AIC, BIC, AICc, AME and RMSE were used to select the models that really provided the best fit for the series and it revealed that both ARIMA (5,2,4) and SARIMA $(4,2,3)(1,2,1)_{12}$ are the best. The chosen SARIMA model, however, outclassed the ARIMA model due to its overall lower values of model selection criteria used and its better forecasting accuracy. The chosen model was then used to forecast the crime rate in Nigeria for five years and the results revealed an expectation of a downward trend in crime rate in Nigeria.**

Keywords - ARIMA, SARIMA, Crime, Differencing, Stationarity.

I. Introduction

The prevalence of crime all over the world today is a cause for serious concern for all and Sundry. It undermines the social fabric by eroding the sense of safety and security and equally constitutes a problem when its incidence is as rampant in the society as to constitute a threat to the security of persons and property. According to [1], the costs of crime is tangible and intangible on economic or social, direct or indirect, physical or psychology, individual or community, and it is from these costs that the consequences of crime are decided.

According to [2], "a crime is held to be an offense" and for it to be known as such it must be reported and recorded by the police or other investigators [3]. However, at the moment there is no universal definition of crimes relating to the psychological and economic conditions of people in respect to crime sites. An act may be a crime in once society, but not in another [4], [5], [6], [7], [8] and [9].

Crime is a threat to the economic, political and social security of a nation and a major factor associated with underdevelopment because it discourages both local and foreign investments, reduces the quality of life and destroys human & social capital, damaged relationship between citizens and the states, thus undermining democracy, rule of law and the ability of the country to promote development. All these factors and the needs to overcome their consequences necessitate the needs for a time domain modeling of crime rate. The essence is to check the superiority between two time domain models of ARIMA and SARIMA in carrying out crime analysis in order to provide insight into the impact of seasonal or non-seasonal effects on the crime rate in the country. This will help to contribute to efforts in combating crime as opined by [10] that crime analysis is a major component in specific police crime reduction strategies [11] examined different crime models believed to have captured obvious heterogeneity that existed within the population where crimes are committed and established that there appear to be uneven periods of time between arrests even for people with fixed number of arrests in a short period of time.

[12] used ARIMA to make short term forecasting of property crime for one city of china. With the given data of property crime for 50 weeks, the fitted ARIMA model was used to predict the crime amount of 1week. The research showed that the ARIMA model had higher fitting and forecasting accuracy than exponential smoothing and therefore, would be helpful for the local police stations and municipal governments in decision making and crime suppression. They opined that AR (1) is suitable for crime sample distributing by week and IMA (1, 1) by day. Thus, AR (1) had higher accuracy in fitting and forecasting than ARMA (1, 1). [13] applied ARIMA with intervention model to evaluate the impact of September 11th terrorist attack on Asian economic crisis while [14] investigated the effects of ORS SIKAP on road accident in Malaysia using intervention analyses by assessing the intervention effect in comparison with the standard ARIMA model, and hence, to obtain the best model for forecasting purpose and the findings reveal that there was a drop in the number of road accidents during OPS SIKAP.

[15] assessed the effects of a demerit points system introduced in July 2003 on the prevalence of seat belt use and the number of road traffic death and injuries through the application of intervention analysis and the research revealed that an estimated number of 1545 deaths and 91772 injuries were prevented in 18months after the introduction of the legislation.

[16] conducted three empirical studies detecting the determinants of crime in English and Wales using time-series analyses to look for co-integrating relationships between property crimes and unemployment as well as law enforcement instruments, employing panel data and corresponding techniques to control for area-specific fixed effects as well as the endogeneity of law enforcement variables and allowed crime rate to have spatial spillover effect.

It is pertinent to note that the majority of the existing literature dwelled majorly on ARIMA modeling of crime around the globe, hence the above reviewed. However, in various independent research works of [17], [18], [19] and [20], it was affirmed that SARIMA model outperformed other models in estimation and forecasting power especially when it comes to financial time series modeling.

II. MATERIALS AND METHODS

This examined thoroughly the basic plots, definitions and concepts of time series analysis assumption, condition, principles and processes involved in the application of autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA). The data for this research comprises of secondary data on the crime rate in Nigeria collected from the Federal Office of Statistics website for the periods of sixteen years (2000-2015).

ARIMA Model

The model is generally referred to as an ARIMA (p,d,q) model where p, d and q are integers greater than or equal to zero and refer to the order of the autoregressive, integrated, and moving average parts of the model respectively. A time series X_t is said to follow an integrated autoregressive moving average model if the dth difference $W_t = \nabla^d X_t$ is a stationary ARMA process. If W_t follows an ARMA (p, q) model, we say that X_t is an ARMA (p, d, q) process. Consider an ARIMA (p, 1, q) process where d = 1. With $W_t = X_t - X_{t-1}$ we have:

$$W_t = \emptyset_1 W_{t-1} + \emptyset_2 W_{t-2} + \cdots + \emptyset_p W_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (1)$$

Or, in terms of the observed series,

$$X_t - X_{t-1} = \emptyset_1 (X_{t-1} - X_{t-2}) + \emptyset_2 (X_{t-2} - X_{t-3}) + \cdots + \emptyset_p (X_{t-p} - X_{t-p-1}) + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \quad (2)$$

Equation (2) may be written as:

$$X_t = (1 + \emptyset_1) X_{t-1} + (\emptyset_2 - \emptyset_1) X_{t-2} + (\emptyset_3 - \emptyset_2) X_{t-3} + \cdots + (\emptyset_p - \emptyset_{p-1}) X_{t-p} - \emptyset_p X_{t-p-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

This is called the difference equation form of the model. Notice that it appears to be an ARMA (p + 1, q) process. This can be written in lag form as:

$$\left(1 - \sum_{i=1}^{p} \emptyset_i B^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i B^i\right) \varepsilon_t \quad (3)$$

Assume that the polynomial $\left(1 - \sum_{i=1}^{p} \emptyset_i B^i\right)$ has a unitary root of multiplicity d. then it can be re-written as:

$$\left(1 - \sum_{i=1}^{p} \emptyset_i B^i\right) = \left(1 + \sum_{i=1}^{p-d} \omega_i B^i\right) (1 - B)^d \quad (4)$$

An ARIMA (p,d,q) process expresses this polynomial factorization property, and is given by:

$$\left(1-\sum_{i=1}^{p}\omega_i B^i\right)(1-B)^d X_t = \left(1+\sum_{i=1}^{q}\theta_i B^i\right)\varepsilon_t \quad (5)$$

SARIMA Model

An important tool in modeling non-stationary seasonal processes is the seasonal difference. The seasonal difference of period **s** for the series $[X_t]$ is denoted by $\nabla_s X_t$ and is defined as:

$$\nabla_s X_t = X_t - X_{t-s} \quad (6)$$

For a series of length n, the seasonal difference series will be of length n-s; that is, **s** data values are lost due to seasonal differencing.

In a non-stationary seasonal model, a process X_t is said to be a multiplicative seasonal ARIMA model with non-seasonal (regular) orders p, d and q, seasonal orders P, D and Q and seasonal period **s** if the differenced series

$$W_t \nabla_s^D X_t \quad (7)$$

satisfies an ARMA (p,q) x (P Q)$_s$ model with seasonal period **s**. We say that $[X_t]$ is an ARIMA (p, d, q) x (P, D, Q)s model with seasonal period **s**.

In practice, many time series contains a seasonal periodic component which repeats every s observations. Box-Jenkins has generalized the ARIMA model to deal with seasonality and defines a general multiplicative seasonal ARIMA model in the form:

$$\emptyset(B)\Phi(B)(1-B)(1-B^{12})X_t = \theta(B)\Theta(B^{12})\varepsilon_t \quad (8)$$

where B denotes the backward shift operator, $\emptyset, \Phi, \theta$ and Θ are polynomials for order p,P, q, and Q respectively. X_t is the observed time series and ε_t represent an unobserved white noise series. i.e a sequence of independently (uncorrelated) identically distributed random variables with zero mean and constant variance σ_ε^2.

All the identified parameters for the two models shall be estimated using the method of maximum likelihood and diagnostic checks shall be carried out using Autocorrelation function (ACF) Partial autocorrelation function (PACF) plots, Ljung-Box test, The Akaike Information Criterion (AIC), Improved AIC (AICc) and Schwarz's Bayesian Criteria (SBC). However, in selecting the best out of the two models, the forecasting error (minimum forecast errors) is to be compared with each other. Forecast accuracy measures like, Absolute Mean Error (AME) and Root Mean Square Error (RMSE) will be used in this research work.

Forecasting Using ARIMA and SARIMA Models.

Once an adequate and satisfactory model is fitted to the series of interest, forecasts can be generated using the model. Consider the general ARIMA model of equation (1).

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \quad (9)$$

The one-step ahead forecast for time t+1 is given by:

$$y_t = \phi_1 y_t + \phi_2 y_{t-1} + \ldots + \phi_p y_{t-p+1} + e_{t+1} - \theta_1 e_t - \theta_2 e_{t-1} - \cdots - \theta_q e_{t-q+1} \quad (10)$$

Except for e_{t+1} , the random shock at time t+1, all other parameters are known.

Thus, setting $e_{t+1} = 0$, its true expected values, the one-step ahead forecasts can be generated.

III. ANALYSIS AND RESULTS

The results for this research are categorized into three different sections.

Exploratory Data Analysis (EDA)

The table and graph below present the summary statistics and time plot of the crime rate respectively.

Table 1: Summary of the Original Data

Min	1stquarter	Median	Mean	3rdquarter	Max
10456	16210	19100	19021	21849	29011

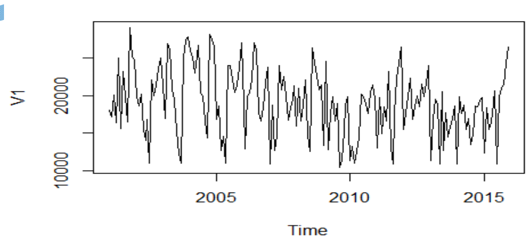

Fig 1 – Time Plot of Crime rate

Specification of ARIMA Model

Table 2: Augmented Dickey-Fuller Test for Series Stationarity of differenced data

Dickey-Fuller Test Statistic	Lag order	P-value
-18.024	36	0.000

.

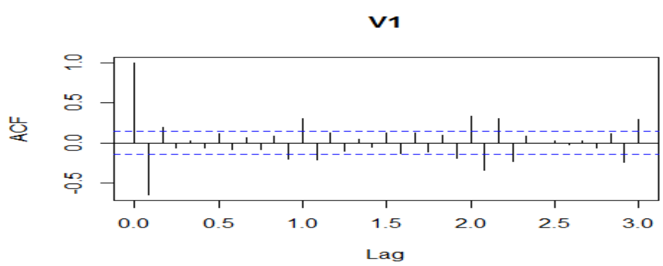

Fig 2 – ACF of 2nd order differenced crime data

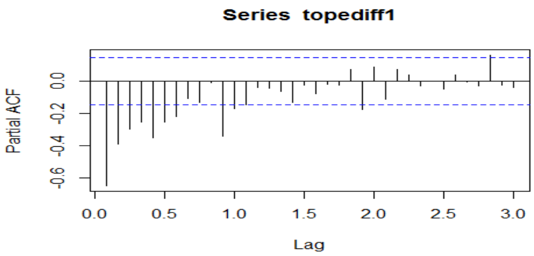

Fig 3 – PACF of 2nd order differenced crime data

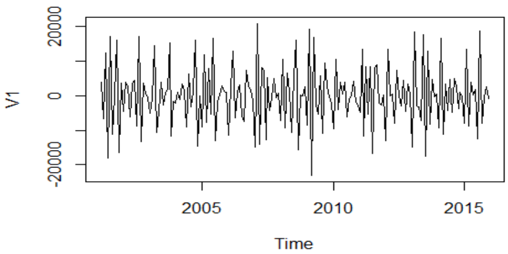

Fig 4 - Plot of 2nd Order differenced crime rate data

Table 3: ARIMA Models for Crime Rate

Model	AIC	BIC	AICc	RMSE	AME
(5,2,4)	3491	3522.813	3492.054	4019.268	3174.23
(5,2,5)	3497.91	3532.908	3499.21	4019.647	3239.911
(5,2,3)	3494.89	3523.53	3495.736	4036.845	3252.895
(5,2,2)	3510.35	3535.808	3511.005	4300.84	3454.895
(6,3,4)	3503.91	3538.844	3505.208	4187.265	3377.911

Table 4: SARIMA Models for Crime Rate

Model	AIC	BIC	AICc	RMSE	AME
$(2,2,2)(1,2,1)_{12}$	3106.658	3127.927	3107.144	4213.728	2997.777
$(3,2,2)(1,2,1)_{12}$	3123.582	3147.878	3147.878	4452.053	3073.694
$(3;2;4)(1,2,1)_{12}$	3108.534	3138.904	3138.904	4079.284	2866.235
$(4,2,3)(1,2,1)_{12}$	3104.943	3135.313	3105.313	4018.882	2885.193

Table 5: Parameter Estimates of Selected ARIMA (5, 2, 4) and SARIMA $(4,2,3)(1,2,1)_{12}$

Order	Coefficient	Std Error	AIC
ARIMA - AR (5)	-0.1339	0.0810	3491
- MA (4)	0.9426	0.1037	
SARIMA – AR (4)	-0.0539	0.0845	3104.9
- **MA (3)**	0.9971	0.0761	
	-0.3980	0.0767	
- **SAR (1)**	-0.9973	0.0733	
- **SMA (1)**			

IV. DISCUSSIONS

Table 1 presents the summary statistics for the Crime rate. Averagely, there are 19021 reported crimes monthly and the highest number of crimes on a monthly basis is 29011. The minimum so far (within the last 15 years) on a monthly basis is 10456 being the number of crimes committed.

According to Figure 1, the pattern of the graph indicates series non-stationarity due to its upward and downward trend as well as a little seasonal variation which also shows that the series is stochastic in nature. The autocorrelation plots in Figures 2 and 3 indicate significant spikes and intervals up to lag 36 but not much of seasonal variation.

As a result of this, stationarity was achieved by 2nd order differencing of the data set and performing Augmented Dickey-Fuller Test on the differenced series in order to check for its acceptability. Figure 4 shows the time series plot of the differenced data and it can be observed that the upward trend has disappeared thereby making the series to be stationary. With few iterations on model-building strategy, the best possible ARIMA and SARIMA models for the series were discovered to be ARIMA (5, 2, 4) and SARIMA $(4,2,3)(1,2,1)_{12}$ respectively. This was the result of their lower values of diagnostic checks as presented in Tables 3 and 4 above. These values were further compared for the two chosen models and It was observed that SARIMA model $(4,2,3)(1,2,1)_{12}$ is having the least values in all the diagnostic values compared to the ARIMA model, thus SARIMA $(4,2,3)(1,2,1)_{12}$ was selected as the best model for forecasting the crime rate data.

V. CONCLUSION

Having used all necessary approaches in line with the set goals of this research, there is no doubt that the main purpose has been fully realized. SARIMA $(4, 2, 3)(1, 2, 1)_{12}$ model outperformed the chosen ARIMA model and is therefore recommended for the modeling and forecasting of crime-related data. From the forecasted figures, the values exhibited a nearly constant trend for the aggregated crime data with an occasional periodic increase in the month of February, April, June, August, October and December throughout the five years forecasted periods. Thus, crime rates in Nigeria can be said to have been seasonally induced over time due to the occasional increases experienced in the listed months.

REFERENCES

[1] Onoge, O. O. *Crime and Control in Nigeria. The Problem of Robbery in Unknown Dimensions of Social Problems in Nigeria*, Kuru, NIPSS, 1999

[2] Scott, J. *Oxford Dictionary of Sociology*, Fourth Edition, Oxford University Press, New York. 2014

[3] Scott, I. and Marshall, E. *Oxford Dictionary of Sociology*, Oxford University Press, New York, 2009

[4] Quinney, R. Structural Characteristics, Population Areas, and Crime Rates in the United States. *The Journal of Crime Law, Criminology and Police Science*. 57(1), 45 – 52, 1980

[5] Levith, S.D. The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation. *The quarterly Journal of Economics*, Mit Press. 111, 319-351. 1996

[6] Fieldman, M. P. The Psychology of Crime, Cambridge University Press, 1997

[7] Dambazau, A. B. *Criminology and Criminal Justice*, 2nd Edition, Ibadan University Press, 2007

[8] Draca, M, Machin, S., Witt, R. Panic on the streets of London: Police, crime, and the July 2005 terror attacks. *The American Economic Review, vol. 101 (5), 2157–2181, 2011*

[9] Usman, U., Yakubu, M. & Bello, A. Z. An Investigation of the Rate of Crime in Sokoto State Using Principal Component Analysis. *Nigeria Journal of Basic and Applied Science*. 20(2), 157-160, 2012

[10] SANTOS, R. B (2014). THE EFFECTIVENESS OF CRIME ANALYSIS FOR CRIME REDUCTION. CURE OR DIAGNOSIS? *JOURNAL OF CONTEMPORARY CRIMINAL JUSTICE, 30(2), 2014*

[11] Bushway, S. D and Tahamont, S. Modeling Long-Term Criminal Careers. What Happened to the Variability? *Journal of Research in Crime and Delinquency, 53(3), 2016*

[12] Chang, G. L., Chen, L. & Carder, E. C. Intervention Analysis for the Impacts of the 65mph Speed Limit on Rural Interstate Highway Fatalities. *Journal of Safety Research. 24(1), 33-53, 19993*

[13] Jennifer, C. H., Hgien-Hung, M .K. & Liu, H. H. Interventions Affecting Air Transport Passenger Demand in Taiwan. *African Journal of Business Management.* 4(10), 2121-2131, 2010

[14] Yaacob, W. F., Husting, W. Z., Abd-Aziz, N. and Nordin, N. I. An Intervention Model of Road Accidents: The Case of OPS Sikap Intervention. *Journal of Applied Science. 11, 1105 – 1112, 2014*

[15] Zambon, F., Fedeli, U., Christiana, Y., Marchesan, M., Ayoosa, F., Brodo, S. & Spolane, P. Evidence-based Policy on Road Safety: The Effect of the Demerit Points System on Seat belt Use and Health Outcomes. *Journals of Epidemiology and Community Health*. 877- 881, 2007

[16] Han, L., Bandyopadhyay, S., Bhattacharya, S. Determinants of violent and property crimes in England and Wales: A Panel data analysis. *Applied Economics, vol. 45(34), 4820–4830. 2013*

[17] Olowe, R. A. Modelling Naira/Dollar Exchange Rate Volatility: Application of Garch and Asymmetric Models", *International Review of Business Research Papers, 5(3), 377 – 398, 2009*

[18] Etuk, E. H. "Forecasting Nigerian Naira-US Dollar Exchange Rates by a Seasonal Arima model", *American Journal of Scientific Research , vol. 59, 71 – 78, 2012*

[19] Etuk, E. H. "Forecasting Nigerian Naira-Euro Exchange Rates by a Seasonal Arima model", *American Journal of Scientific Research , vol. 60,* 60 – 67, 2013

[20] Adeboye, N. O and Fagoyinbo, I. S. Fitting of Seasonal Autoregressive Integrated Moving Average to the Nigerian Stock Exchange Trading Activities. *Edited Proceedings of 1st International Conference of Nigeria Statistical Society, 2017.*