



Research Article

Modeling Time Series Poison Count Data: An Application to COVID 19 Confirmed Cases in Nigeria

¹Bashir Alhaji Mustapha, ²Harun R. Bakari and ²Yusuf A. Mohammed

¹Department of Statistics, Federal Polytechnic Damaturu, Nigeria

²Department of Mathematical Sciences, University of Maiduguri, Nigeria

*Corresponding author: hafisbash2010@gmail.com, doi.org/10.55639/607.060504

ARTICLE INFO:

Keyword:

ARIMA,
Simulation,
COVID-19 data,
Poison data,
Forecasting

ABSTRACT

Time series events are counted within time intervals usually for convenience and practical reasons. Time series data involving counts frequently follows Poison distribution. Integrated Autoregressive Moving Average (ARIMA) models were studied on simulated Poison data of different parameter and sample sizes. These models were used to capture Poison data with different phenomena. Data set were simulated from Poison process with $\lambda = 5, 10, \text{ and } 20$. ARIMA (p, d, q) were then fitted to the simulated data so as to examine the effect of the changes in parameter value of the Poison on the models' performance across the sample size. Thereafter, the same models were used to fit and forecast the daily Covid-19 data, confirmed cases from inception to February 2023. It was concluded that ARIMA (2, 1, 2) and ARIMA (1, 1, 1) are obviously the best Models at different sample sizes respectively. Based on the Covid-19 analysis, the estimated coefficient values of all ARIMA (p,d,q) strictly conforms to the bounds of the parameter between -1 and 1 and this had made the model to be stationary comparing the ARIMA models in terms of the AIC, BIC and MSE. It is clearly preferred ARIMA (1, 0, 2) to other models since their estimated AIC, BIC and MSE are smaller as compared to other models. The forecast is quite accurate, and hereby confirms what was expected. It seems that ARIMA (1, 0, 2) does a very good job in capturing the dynamic nature of the data and forecasting. It was observed that the confirmed cases increase, and then keep stationary throughout the future times.

Corresponding author: Bashir Alhaji Mustapha, Email: hafisbash2010@gmail.com
Department of Statistics, Federal Polytechnic Damaturu, Nigeria

INTRODUCTION

Time series data involving counts frequently follows Poisson distribution. This can be encountered in many biomedical and public health applications. For example, in disease surveillance, the occurrence of rare infections over time is often monitored by public health officials, and the time series collected can be used for the purpose of monitoring changes in disease activity. The primary objective of time series modeling is to study techniques and measures for drawing inferences from the past data. The models can be employed to describe and analyze the sample data and make forecasts for the future. The main advantage of time series models is that they can handle any persistence patterns in data (Abdullahi & Taylor, 2004). For time series, certain models are commonly known, like an Autoregressive Moving Average (ARMA) model. This method takes into account that past information influences the variables of today. For example, the rate of inflation of a few days ago influences the rate of inflation today. Or import and export data can be predicted by using information on how many is imported and exported in the past few months. Many examples can be given and these ARMA-models work pretty well in modeling the time series (Imam, 2020).

One special class of time series model is ARIMA models which are often associated with Box and Jenkins (1976) for their effort to systematize the whole methodology of estimating, checking and forecasting using ARIMA models (Mahesh, 2005). The Box – Jenkins method consist of three steps: identification, parameter estimation and forecasting. Among these three steps, the identification step, which involves order determination of the AR and MA part of ARIMA model, is important. This step requires statistical information such as the autocorrelation and partial autocorrelation (Box and Jenkins, 1976). The problem of estimating

the order and the parameters of an ARIMA model is still an active area of research. The Box and Jenkins Variant of the ARIMA model is predestinated for applications to non-stationary after differencing.

Amon *et al.* (2015) studied the impact of the size of the historical data on ARIMA models in formulating accuracy. The study used 286 weekly records of amount of solid waste generated in Arusha city to formulate four (4) ARIMA models using different data lengths or size. The first model, M_1 used 30 observations the second model, M_2 used 60 observations, the third model, M_3 used 120 observations and the fourth mode M_4 used 260 observations of which are the most recent. A total of 26 observations were held out for validation. The precision in forecasting was tested using MAPE, RMSE and MAD. The results indicated variation in precision, M_3 performed best in one-week ahead and 9-12 weeks ahead while M_4 performed better in 2-8 weeks and also for 13 weeks and above, M_1 was the weak model in forecasting.

Imam (2020) investigated the best ARIMA model for forecasting average daily share price indices of the series of square pharmaceuticals limited (SPL). After stationary test, the data was found to be not stationary; a differencing was carried out to obtain stationarity. The results indicated that the best ARIMA model was found to be ARIMA (2,1,2). The accurate prediction of foreign exchange is important as substantial amount of trading takes place through the currency exchange market. The prediction is affected by economic and political factors and also involves uncertainty and non-linearity. Thus, accurate prediction of exchange rate is a complex task (Nwankwo, 2014). Volatility modeling of exchange rate has many practical applications in statistics, economics and finance.

Hillmer and Tiao (1982) used ARIMA technique for the seasonal adjustment as well as introduce the decomposition of the time series

data into its mechanism like trend, seasonal and noise whereas such series will follow the assumption of the Gaussian ARIMA model. Hillmer and Tiao (1982) study the effectiveness of time series modelling (ARIMA) in forecasting stock prices. In their study, Fifty-Six (56) stocks from Seven (7) sectors were used. The accuracy of ARIMA model in predicting stock prices was evaluated. From the standard deviation of accuracy of forecasting of seven sectors, they found that Automobile sector, steel sector and the banking sector has a high standard deviation which means the values are spread over a large range, and there might be stocks for which ARIMA model does not produce good results.

Studies by some authors have modeled time series count data but failed to consider a particular distribution and the effect of sample size on the different parameters of the Poisson data. This study is aimed at modeling time series count data with emphasis to ARIMA Poisson data on confirmed cases of COVID19 in Nigeria from inception to February 2023.

METHODOLOGY

Data set were simulated in R statistical software with sample sizes of 50, 100,..., 500 from Poisson process of $\lambda_i, i = 5, 10 \text{ and } 20$. The four models under study, namely: ARIMA (1,d,1), ARIMA (1,d,2), ARIMA (2,d,1) and ARIMA (2,d,2) were then fitted to the simulated data so as to examine the effect of the proportion of changes in mean and sample size on them. Their performances were compared at different sample sizes and means of Poisson distribution. Thereafter, the forecast performances of the four fitted models were also examined at different steps ahead. All cases of simulation were randomized and replicated 1000 times each for the respective selected sample sizes. In simulation, we set our

parameters to be $\phi_1 = \phi_2 = 1$ to ensure discrete nature of poison data generated. The response Y_{ti} were generated from Poisson distribution. The four models under study were considered to analyse how well each of the model fits the selected data sets. Data were generated from linear second orders of autoregressive functions given as follows:

$$\text{Model 1. AR}(2): Y_{ti} = Y_{ti-1} + Y_{ti-2} + e_t$$

$$t = 50, 100, \dots, 500. i = 1, 2, \dots, 1000$$

Where Y_{ti} were simulated from Poisson as follows:

The Poisson regression model which is based on the Poisson distribution with probability density function

$$\frac{\lambda^y e^{-\lambda}}{y!}, \text{ for } y_i = 0, 1, 2, \dots$$

Thus, for the Poisson models $E(y_i) = V(y_i) = \mu_i$. The restrictive condition that the mean must equal the variance is often violated by overdispersed data (where variance exceeds the mean). As a result of that Poisson model is generally considered inappropriate for count data, which are usually highly skewed and over dispersed (Cameron & Trivedi, 2008).

ARIMA Model

Box-Jenkins ARIMA model has been used widely in many areas of time series analysis. Since ARIMA is among the earliest models, the capability of this model always being tested and widely used as a benchmark with other time series models. Box-Jenkins ARIMA is known as ARIMA (p,d,q) model where p is the number of autoregressive (AR) terms, d is the number of difference taken and q is the number of moving average (MA) terms. ARIMA models always assumes the variance of the data to be constant. The ARIMA (p,d,q) model can be represented by the following equation:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \tag{1}$$

Where $\varepsilon_t \sim N(0, \sigma_t^2)$, p and q are the number of autoregressive and moving average terms, respectively. The identification of modelling the conditional mean value is based on the analysis of estimated autocorrelation and partial autocorrelation (ACF, PACF). These estimators may be strongly inter-correlated. It is therefore

Transformation by Differencing

Differencing is an important technique in transforming data, which attempts to de-trend to

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t,$$

Where B is the backshift operator. We may extend the notion further and define the differences of order d as:

$$\nabla^d X_t = (1 - B)^d X_t,$$

Usually, single differencing is used to remove linear trends and double differencing is used to

$$\nabla_d X_t = X_t - X_{t-d} = (1 - B^d)X_t$$

Normally, the correct amount of differencing is the lowest order of differencing that yields a time series which fluctuates around a well-defined mean value and whose autocorrelation function (ACF) plot decays rapidly to zero, either from above or below. Thus, at every stage of differencing, we check the plots of sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF) to see where the ACF/PACF “cuts off” the bounds $\pm 1.96/n$. It is desirable to find a sample ACF that decays fairly rapidly. We say that a series is stationary if the sample ACF has very few significant spikes at very small lags and then cuts off drastically or dies down very quickly. If

recommended not to insist on unambiguous determination of the model order, but to try more models. We must not forget to carry out the verification, which is based on retrospective review of the assumptions imposed on the random errors.

control autocorrelation and achieve stationary time series. The first difference is denoted as:

remove quadratic trend. We can eliminate seasonality and trend of period d by introducing the lag d difference operator ∇_d :

the sample ACF dies slowly, the series still has some trend. If ACF has periodicity, the series has seasonality. We should do some more differencing of the data before continuing.

DATA ANALYSIS AND RESULTS

The sample of data generated at different categories and sample sizes are presented in graphs as discussed in the subsequent sections. The performances of the ARIMA models [ARIMA (1,1,1), ARIMA (1,1,2), ARIMA (2,1,1) and ARIMA (2,1,2)] and test of stationary are presented in Table 1- 12. Furthermore, the data simulated were differenced. The experiment was repeated 1000 times for the four models at each sample size.

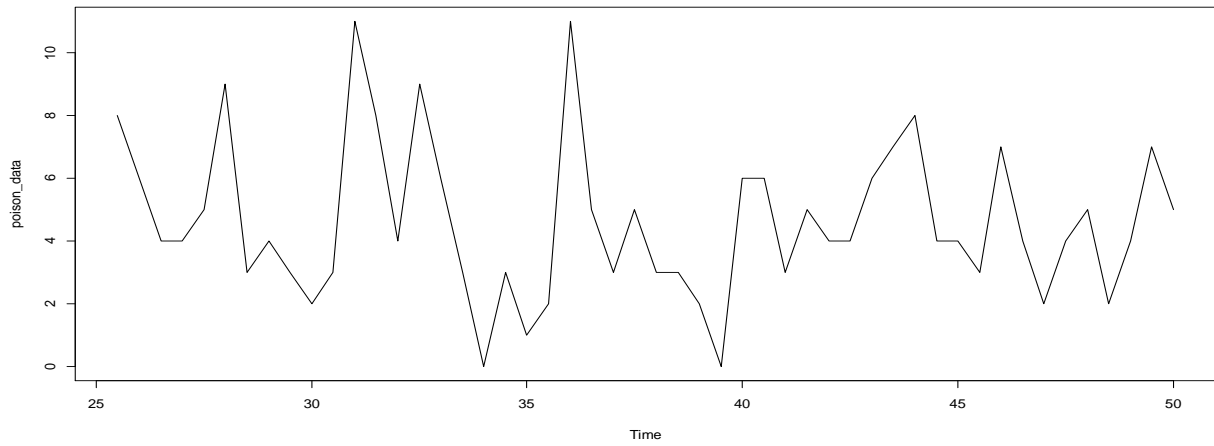


Figure 1a: Poison Data with $\lambda = 5$, and sample size $n = 50$

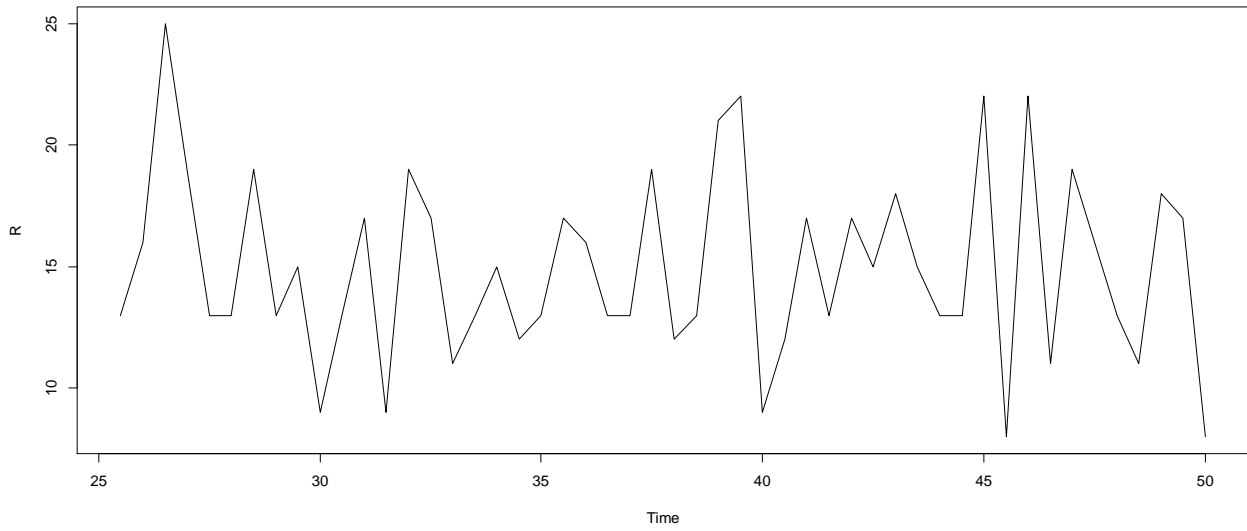


Figure 1b: Poison Data with $\lambda = 5$, and sample size $n = 100$

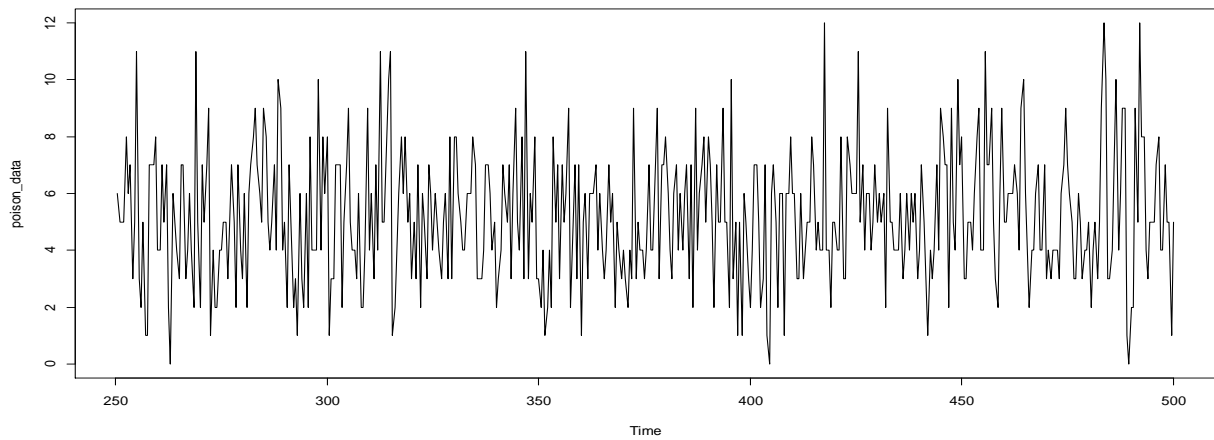


Figure 1c: Poison Data with $\lambda = 5$, and sample size $n = 500$

Table 1: ADF Test of Poison Data with $\lambda = 5$ at Different Sample Sizes

Sample Size	Values	Lag order	P-value	Hypothesis (H_0)	Decision	Remark
50	-3.0077	2	0.1885	Unit root	Accept H_0	Not stationary
100	-2.2075	3	0.492	Unit root	Accept H_0	Not stationary
150	-2.8784	3	0.2197	Unit root	Accept H_0	Not stationary
200	-3.715	4	0.02884	Unit root	Reject H_0	Stationary
250	-5.2466	4	<0.01	Unit root	Reject H_0	Stationary
300	-4.398	4	<0.01	Unit root	Reject H_0	Stationary
350	-4.7182	5	<0.01	Unit root	Reject H_0	Stationary
400	-4.6407	5	<0.01	Unit root	Reject H_0	Stationary
450	-5.8542	5	<0.01	Unit root	Reject H_0	Stationary
500	-5.2129,	5	<0.01	Unit root	Reject H_0	Stationary

Table 1 above shows ADF has the p-values which are greater than the critical value of 0.05 for lower sample sizes and we accept null hypothesis of having unit root series for the series. However, as sample size increases, the

data series is stationary due to the p-values less than 5%. It is clear that the time series plot of the poison data series and stationarity tests suggest that the data need to be transformed or differenced since it is confirmed to have a unit root at lower sample sizes.

Table 2: ADF Test of Differenced Poison Data with $\lambda = 5$ at Different Sample Sizes

Sample Size	Values	Lag order	P-value	Hypothesis (H_0)	Decision	Remark
50	-2.8068	2	0.025	Unit root	Reject H_0	Stationary
100	-3.0384,	3	0.016	Unit root	Reject H_0	Stationary
150	-4.3813	3	<0.01	Unit root	Reject H_0	Stationary
200	-3.8288	3	0.023	Unit root	Reject H_0	Stationary
250	-4.9996	4	<0.01	Unit root	Reject H_0	Stationary
300	-4.5907,	4	<0.01	Unit root	Reject H_0	Stationary
350	-3.9346	4	0.014	Unit root	Reject H_0	Stationary
400	-4.157	5	<0.01	Unit root	Reject H_0	Stationary
450	-3.8714	5	0.0173	Unit root	Reject H_0	Stationary
500	-5.565	5	<0.01	Unit root	Reject H_0	Stationary

Table 2 presents the stationarity tests for the differenced poison data with parameter $\lambda=5$ over the period of investigation with a null hypothesis of a unit root against alternative

hypothesis of a level of stationarity. Indeed, the data is stationary after the first difference hence we can proceed to fitting and forecasting of the series.

Table 3: AIC and BIC Values of ARIMA (p,1,q) with Parameter ($\lambda = 5$)

Sample Sizes	AIC					BIC		
	ARIMA (1,1,1)	ARIMA (1,1,2)	ARIMA (2,1,1)	ARIMA (2,1,2)	ARIMA (1,1,1)	ARIMA (1,1,2)	ARIMA (2,1,1)	ARIMA (2,1,2)
50	98.048	99.4584	94.4932	91.259	100.720	103.019	98.0547	95.712
100	208.108	196.034	190.555	189.614	213.021	202.584	197.106	197.801
150	311.713	298.668	303.529	283.811	317.894	306.91	311.771	294.114
200	370.996	366.734	396.660	385.635	378.066	376.161	406.087	397.419
250	497.755	467.567	446.083	452.469	505.510	477.907	456.422	465.393
300	599.586	546.161	572.275	571.993	607.898	557.244	583.358	585.846
350	706.667	714.975	686.157	631.34	715.449	727.225	697.866	645.976
400	755.901	762.409	779.56	727.905	765.088	774.66	791.811	743.218
450	892.128	776.746	845.144	800.759	901.673	894.73	857.871	816.668
500	979.827	863.746	933.499	908.944	989.692	876.899	946.652	925.385

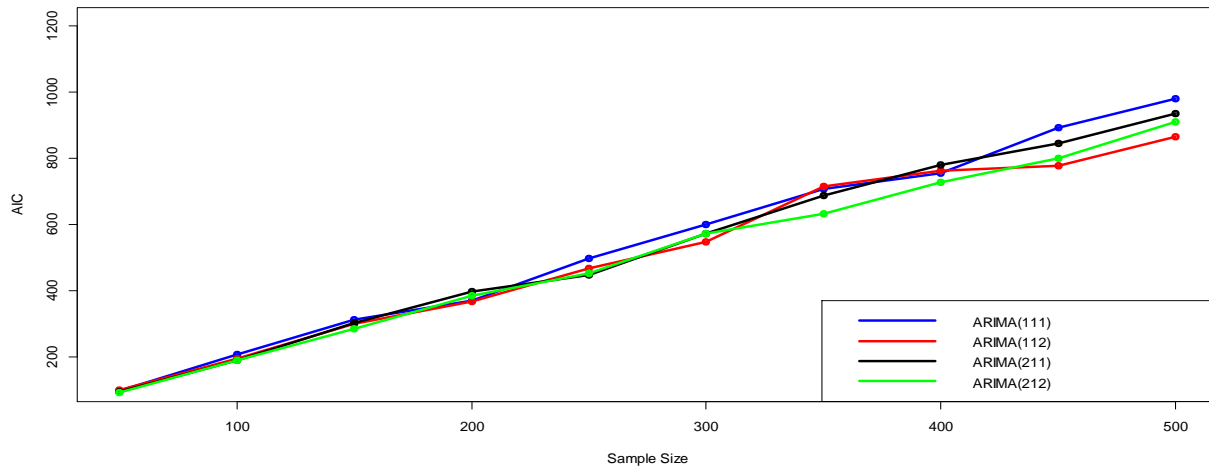


Figure 2a: AIC values of the Fitted models on Poison Data with Parameter $\lambda = 5$

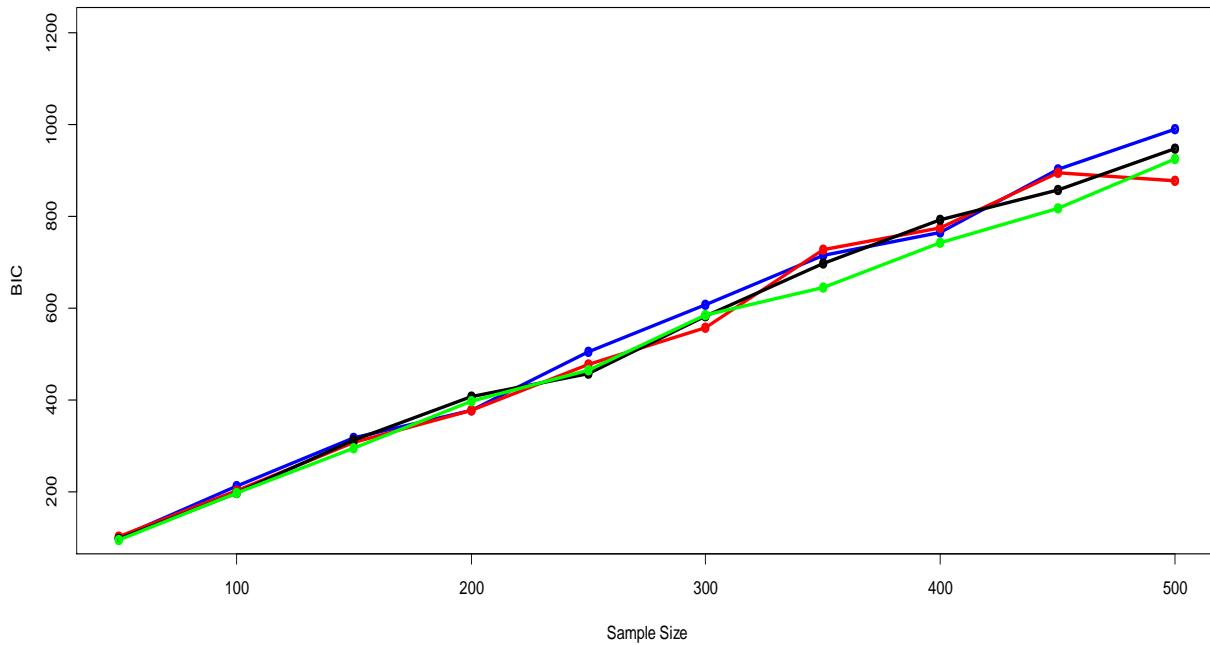


Figure 2b: BIC values of the Fitted models on Poison Data with Parameter $\lambda = 5$

Table 3 shows the fitted performances of the four ARIMA model to data simulated with underline poison distribution of parameter with the average values of AICs and BICs of each model at various sample sizes. The results obtained were plotted on the graphs as shown in

Figure 2a and 2b respectively The best fitted model based on AIC and BIC criteria is ARIMA (2,1,2) followed by ARIMA (1,1,2), . ARIMA (1,1,1) is the weak among fitted models across the periods and sample sizes.

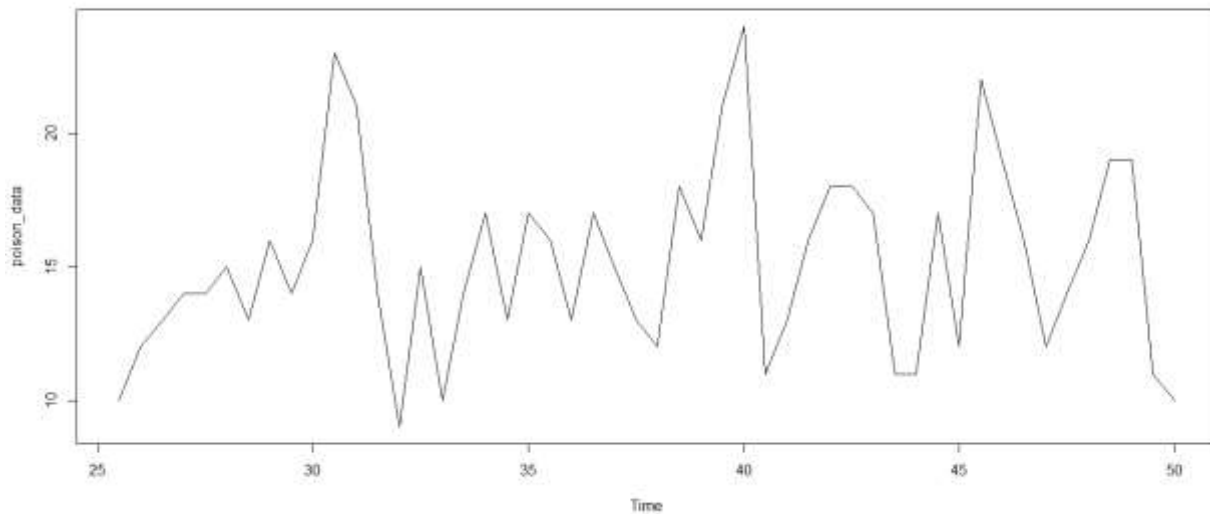


Figure 3a: Poison Data with $\lambda = 10$, and sample size $n = 50$

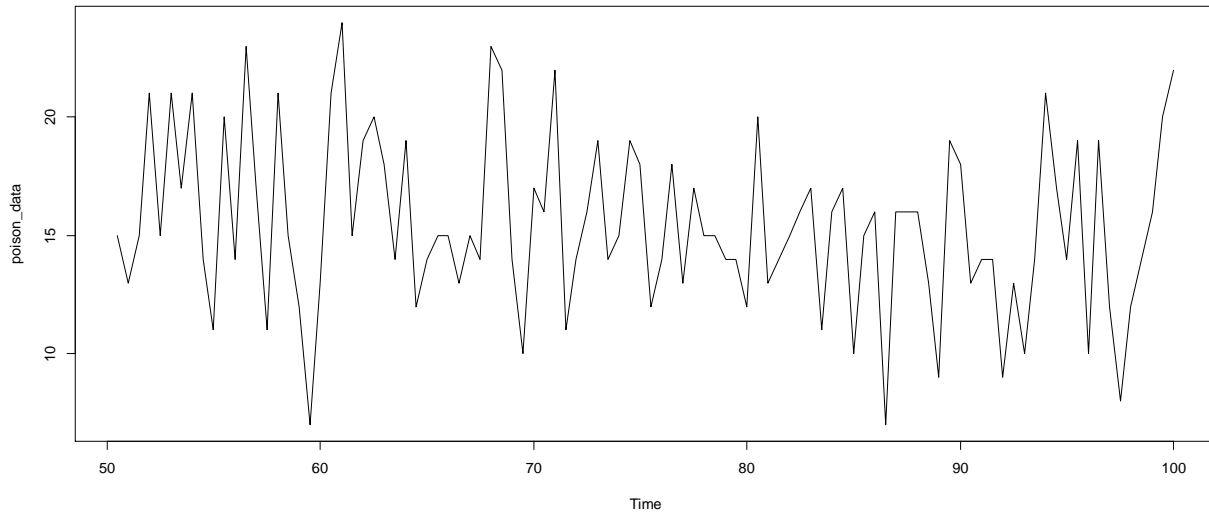


Figure 3b: Poison Data with $\lambda = 10$, and sample size $n = 100$

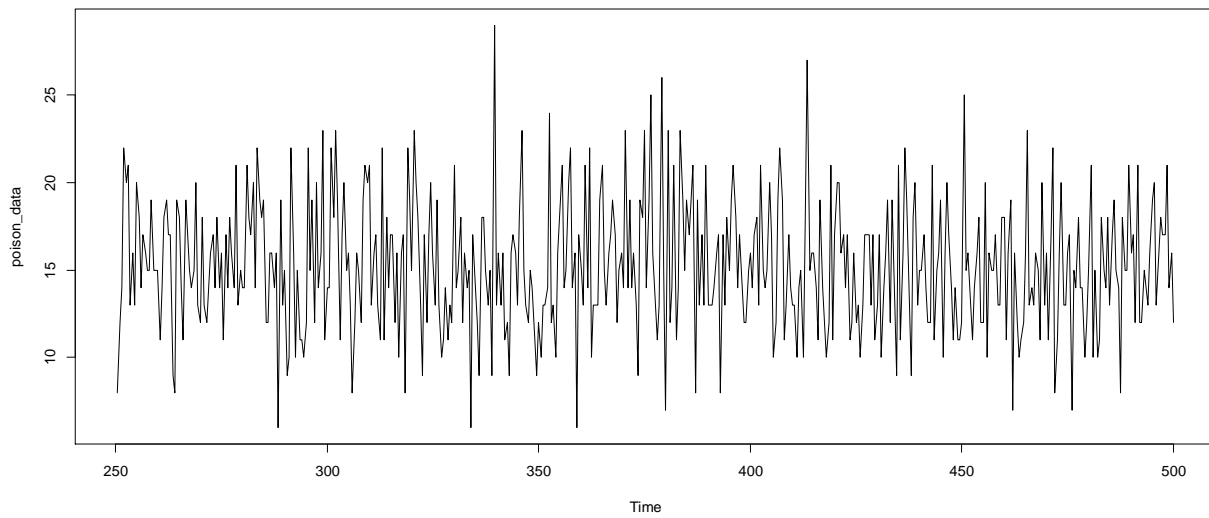


Figure 3c: Poison Data with $\lambda = 10$, and sample size $n = 500$

Table 4: ADF Test of Poison Data with $\lambda = 10$ at Different Sample Sizes

Sample Size	Values	Lag order	P-value	Hypothesis (H_0)	Decision	Remark
50	-3.6268	2	0.04208	Unit root	Reject H_0	Stationary
100	-4.2136	3	0.01146	Unit root	Reject H_0	Stationary
150	-6.1385	3	<0.01	Unit root	Reject H_0	Stationary
200	-5.4582	4	<0.01	Unit root	Reject H_0	Stationary
250	-3.3743	4	0.006277	Unit root	Reject H_0	Stationary
300	-5.1814	4	<0.01	Unit root	Reject H_0	Stationary
350	-5.0753	5	<0.01	Unit root	Reject H_0	Stationary
400	-4.7945	5	<0.01	Unit root	Reject H_0	Stationary
450	-5.593	5	<0.01	Unit root	Reject H_0	Stationary
500	-5.7945	5	<0.01	Unit root	Reject H_0	Stationary

Base on the above table, the test is stationary since the p values are less than 0.05

Table 5: AIC and BIC Values of ARIMA (p,1,q) with Parameter ($\lambda = 10$)

Sample Sizes	ARIMA (1,1,1)	ARIMA (1,1,2)	ARIMA (2,1,1)	ARIMA (2,1,2)	ARIMA (1,1,1)	ARIMA (1,1,2)	ARIMA (2,1,1)	ARIMA (2,1,2)
50	109.565	104.758	112.444	101.450	112.398	108.535	116.222	106.172
100	208.222	212.151	216.951	207.460	213.212	218.806	223.606	215.778
150	314.531	312.488	308.334	330.441	320.763	320.798	316.644	340.828
200	408.571	413.369	413.822	411.615	415.679	422.847	423.300	423.463
250	507.671	524.302	515.869	543.901	515.456	534.682	526.249	556.876
300	597.063	596.940	607.364	630.043	605.400	608.057	618.480	643.939
350	704.063	707.108	733.050	701.437	712.866	718.846	744.788	716.109
400	807.426	816.501	853.909	834.713	816.633	828.776	866.185	850.057
450	927.558	921.701	897.163	927.493	937.121	934.451	909.912	943.430
500	1034.70	1071.11	1052.56	1049.32	1044.58	1084.28	1065.73	1065.79

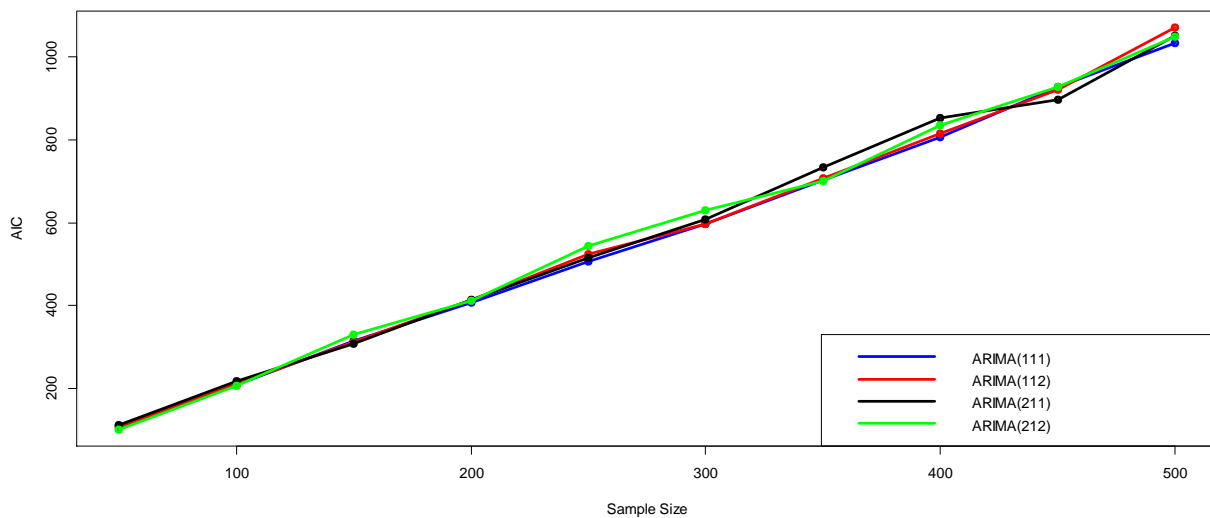


Figure 4a: AIC values of the Fitted models on Poison Data with Parameter $\lambda = 10$

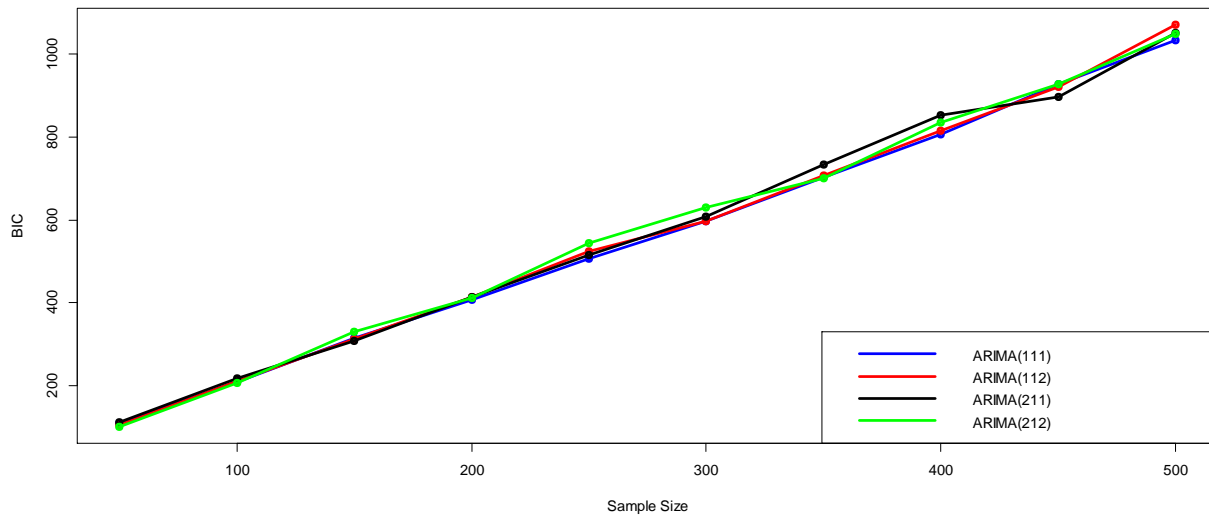


Figure 4b: BIC values of the Fitted models on Poison Data with Parameter $\lambda = 10$

Table 5 shows relative performance of the four fitted ARIMA models with the average values of AIC and BIC of each model at various sample sizes. The results obtained were plotted on the graphs as shown in Figure 4a and 4b

respectively. The best fitted model on poison data with parameter $\lambda = 10$ is ARIMA (111), at various sample sizes followed by ARIMA (2,1,2). ARIMA (1,1,2) is the weak performing Model.

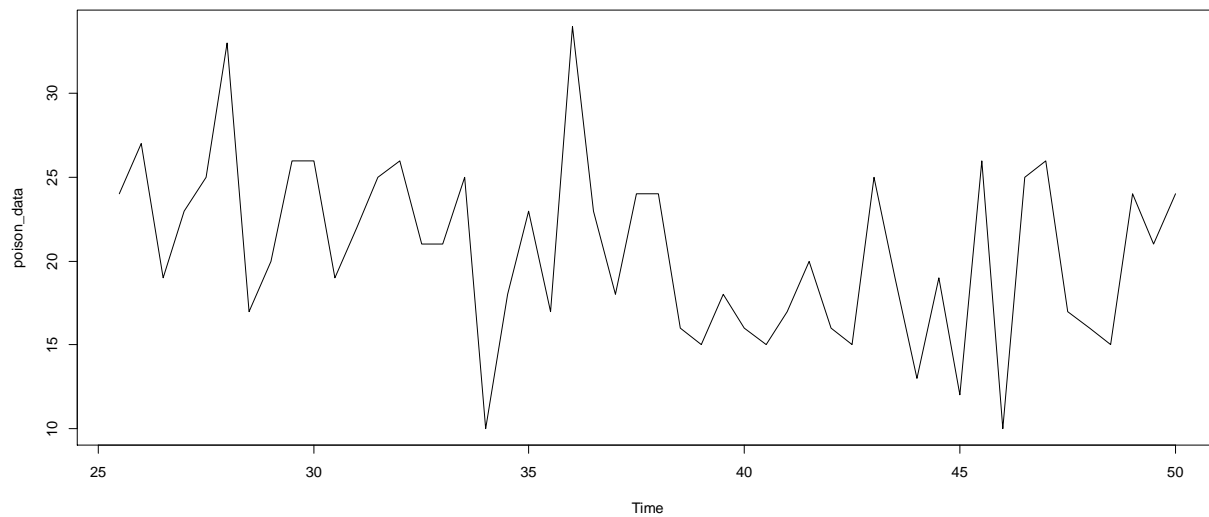


Figure 5a: Poison Data with $\lambda = 20$, and sample size $n = 50$

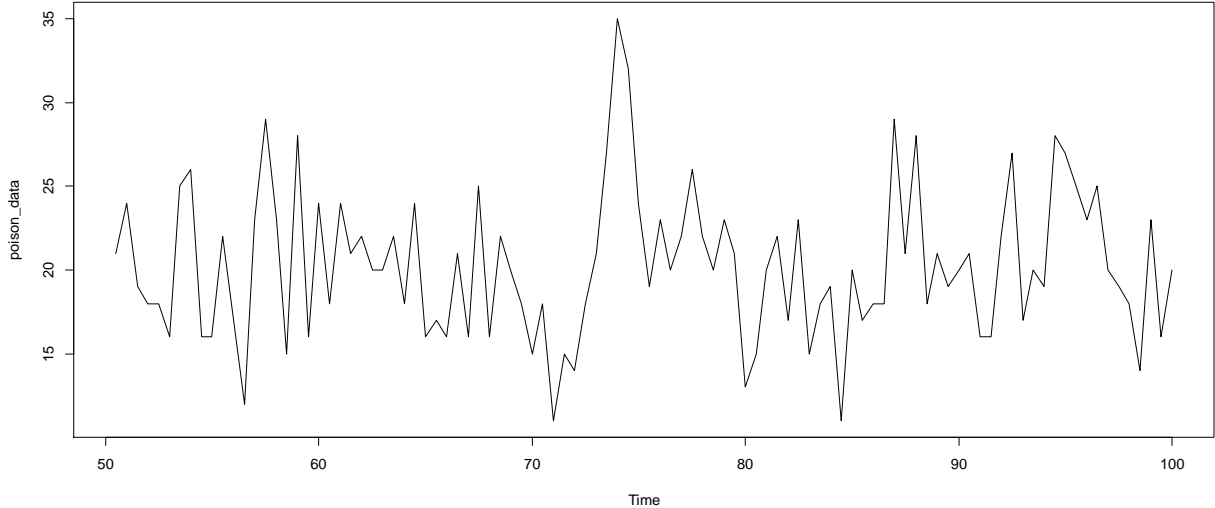


Figure 5b: Poison Data with $\lambda = 20$, and sample size $n = 100$

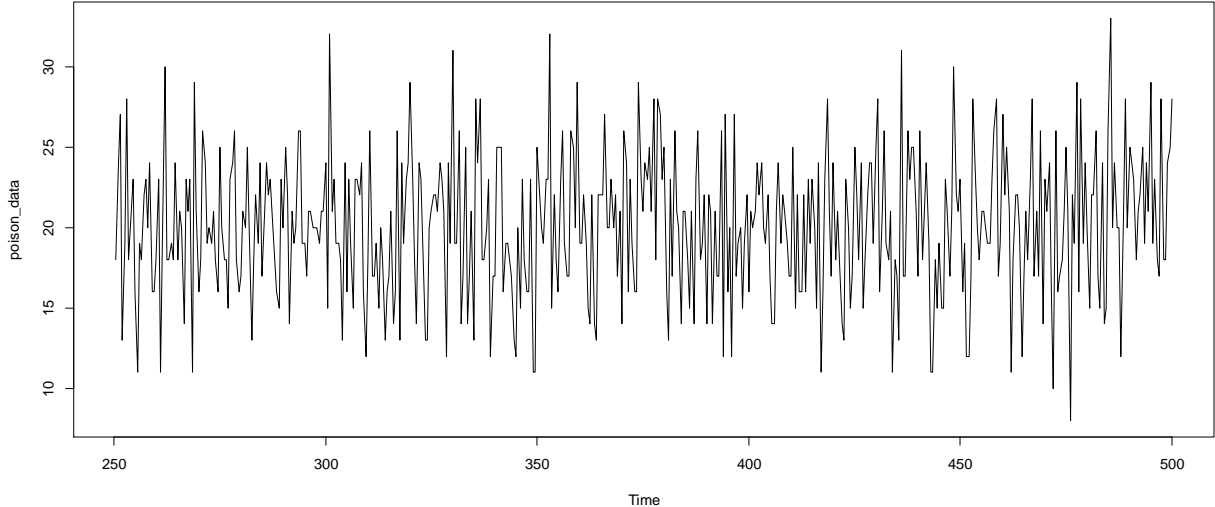


Figure 5c: Poison Data with $\lambda = 20$, and sample size $n = 500$

Table 6: ADF Test of Poison Data with $\lambda = 20$ at Different Sample Sizes

Sample Size	Values	Lag order	P-value	Hypothesis (H_0)	Decision	Remark
50	-2.8515	2	0.0248	Unit root	Reject H_0	Stationary
100	-3.946	3	0.02188	Unit root	Reject H_0	Stationary
150	-3.0749	3	0.01403	Unit root	Reject H_0	Stationary
200	-3.8491	4	0.02088	Unit root	Reject H_0	Stationary
250	-4.333	4	<0.01	Unit root	Reject H_0	Stationary
300	-4.4322	4	<0.01	Unit root	Reject H_0	Stationary
350	-4.9275	5	<0.01	Unit root	Reject H_0	Stationary
400	-5.6941	5	<0.01	Unit root	Reject H_0	Stationary
450	-5.8183	5	<0.01	Unit root	Reject H_0	Stationary
500	-5.8718	5	<0.01	Unit root	Reject H_0	Stationary

Base on the above table, the test is stationary since the p values are less than 0.05

Table 7: AIC and BIC Values of ARIMA (p,1,q) with Parameter ($\lambda = 20$)

Sample Sizes	AIC				BIC			
	ARIMA (1,1,1)	ARIMA (1,1,2)	ARIMA (2,1,1)	ARIMA (2,1,2)	ARIMA (1,1,1)	ARIMA (1,1,2)	ARIMA (2,1,1)	ARIMA (2,1,2)
50	121.293	131.669	120.510	120.556	124.127	135.447	124.288	126.278
100	248.819	239.329	244.383	235.889	253.809	245.983	251.037	254.207
150	341.452	345.023	342.613	339.917	347.684	353.333	350.924	350.305
200	495.409	491.951	483.401	443.677	502.517	501.429	492.878	455.524
250	572.069	599.048	591.894	576.839	579.854	609.428	602.275	589.814
300	687.362	695.285	692.641	698.687	695.699	706.401	703.758	712.582
350	797.261	828.101	835.319	806.991	806.065	839.838	847.057	821.663
400	945.708	933.620	912.399	953.358	954.915	945.896	924.675	968.702
450	1059.77	1050.19	1078.32	1091.31	1069.34	1062.94	1091.07	1107.25
500	1159.28	1142.56	1194.43	1188.49	1169.16	1155.74	1207.60	1204.96

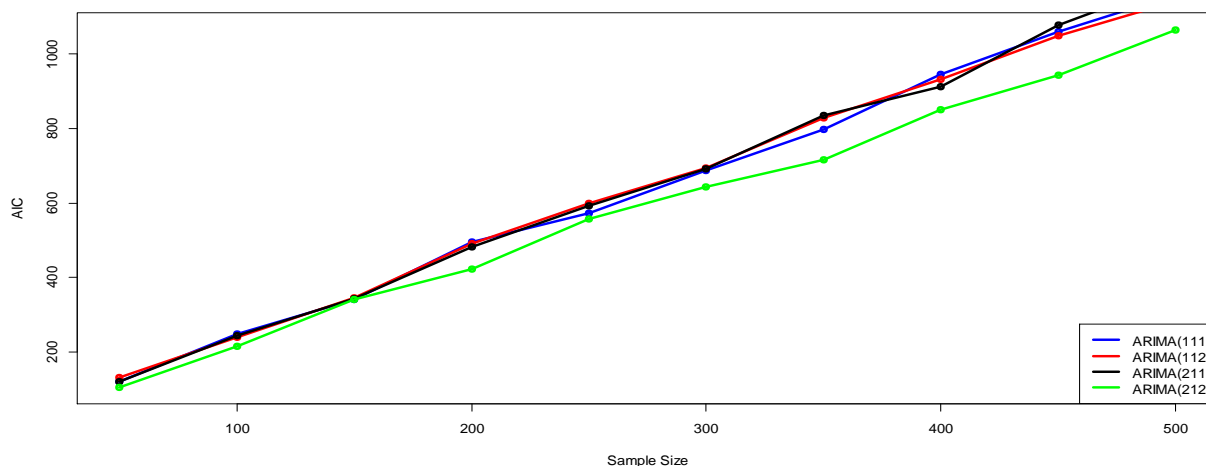


Figure 6a: AIC values of the Fitted models on Poison Data with Parameter $\lambda = 20$

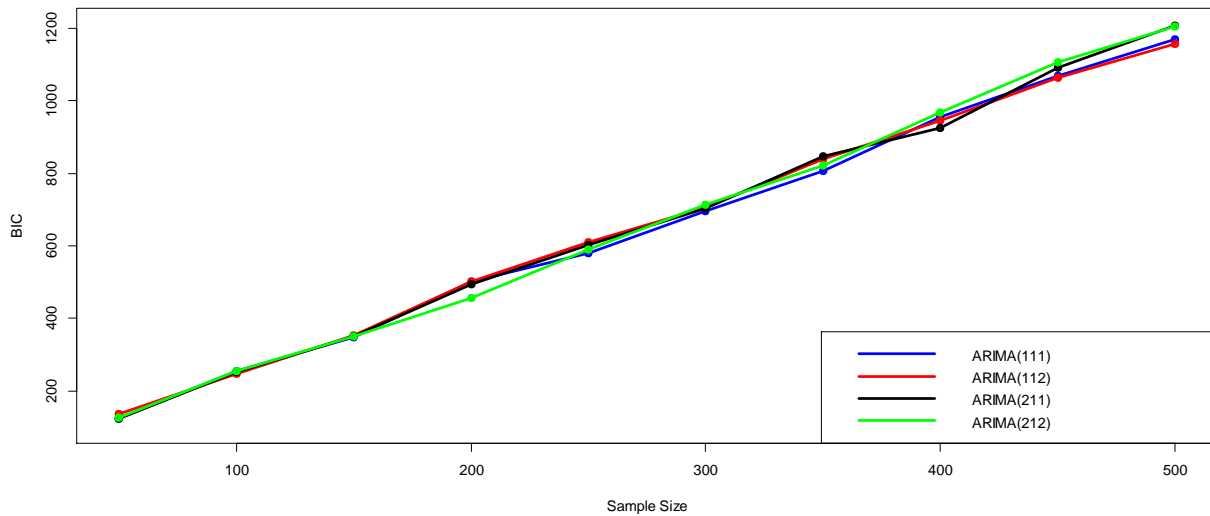


Figure 6b: BIC values of the Fitted models on Poison Data with Parameter $\lambda = 20$

Table 7 shows the fitted performances of the four ARIMA models to data simulated under poison data with $\lambda = 20$. The average values of AIC and BIC of each model at various sample sizes were recorded. The results obtained were plotted on the graphs as shown in Figure 6a and

6b respectively. From Figure 6a and 6b; the four models have close performances on the basis of AIC and BIC. However, it can be seen that ARIMA(1,1,1) Performs better than the other Models while ARIMA(1,1,2) has weak performance.

Table 8: Forecast Performance of the Models using Theil U Statistic

Steps Ahead	ARIMA (1,1,1)	ARIMA (1,1,2)	ARIMA (2,1,1)	ARIMA (2,1,2)
5	1.33013	2.84114	2.57219	3.1075
10	1.30324	2.81424	2.54529	3.09641
15	0.27634	2.78735	2.5184	3.0435
20	0.24944	2.76045	1.4915	3.0286
25	0.22255	2.73356	1.46461	3.0026
30	0.19565	2.70666	1.43771	2.9785
35	0.16876	2.67977	1.41082	2.94872
40	0.14186	2.65287	1.38392	2.92182
45	0.11497	2.62598	1.35703	2.89493
50	0.08807	2.59908	1.33013	2.86803

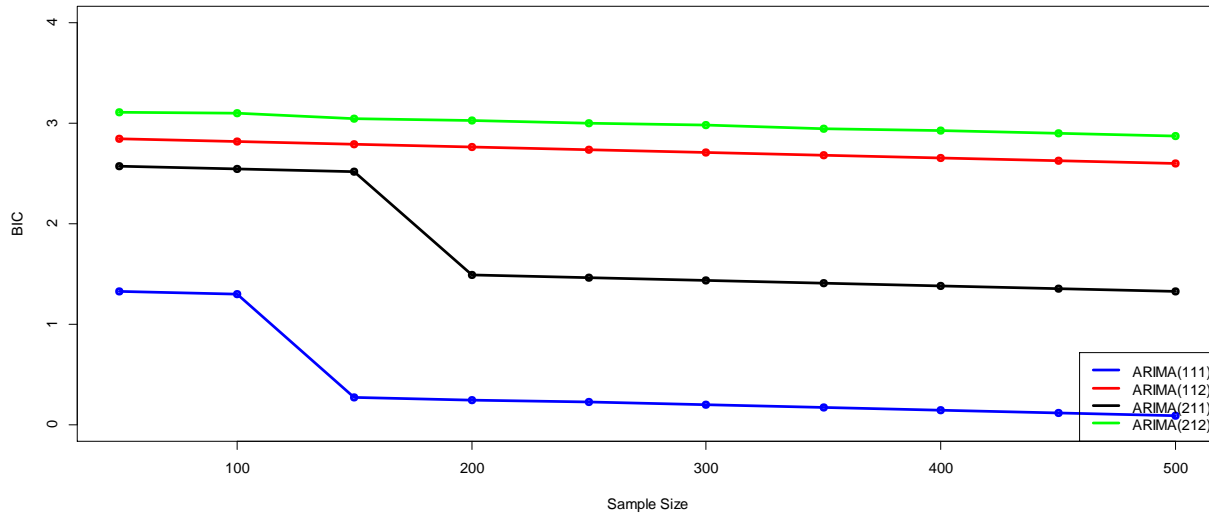


Figure 7: Forecast Performance in Table 8

Based on the Theils Analysis in table 8 above, the ARIMA (2,1,2) has the highest forecasting power due to their values greater than 1 and also greater than other values of the models across the steps ahead; this is followed by ARIMA (1,1,2) and ARIMA (2,1,1). However, the Theil values of ARIMA (1,1,1), at higher steps ahead are close to zero, hence it is not as good as other models in forecasting. Indeed, the forecasting ability of all the models decreases as steps ahead increase.

Analysis of COVID-19 data

This section consists of detail description of the analysis of Covid-19 data using ARIMA modeling method for forecasting and estimation. Most of the computational works for ARIMA models are carried out by using R software. The data obtained on daily confirmed

cases of COVID-19 were analysed to check if the data is stationary or has a unit root using Augmented Dickey Fuller (ADF). The autocorrelation and partial autocorrelation (PACF) of data were also plotted to confirm the statuary of its stationarity. The ARIMA models of different orders were used to model the COVID-19 with aim of selecting the best model for forecasting. For the purposes of the flow of the analysis, the time series data on daily cases of COVID-19 for the period of 36 months in Nigeria were used. However, the time series plots which display observations on the y-axis against equally spaced time intervals on the x-axis used to evaluate patterns and behaviour in data over time is displayed in the Figure 8 below:

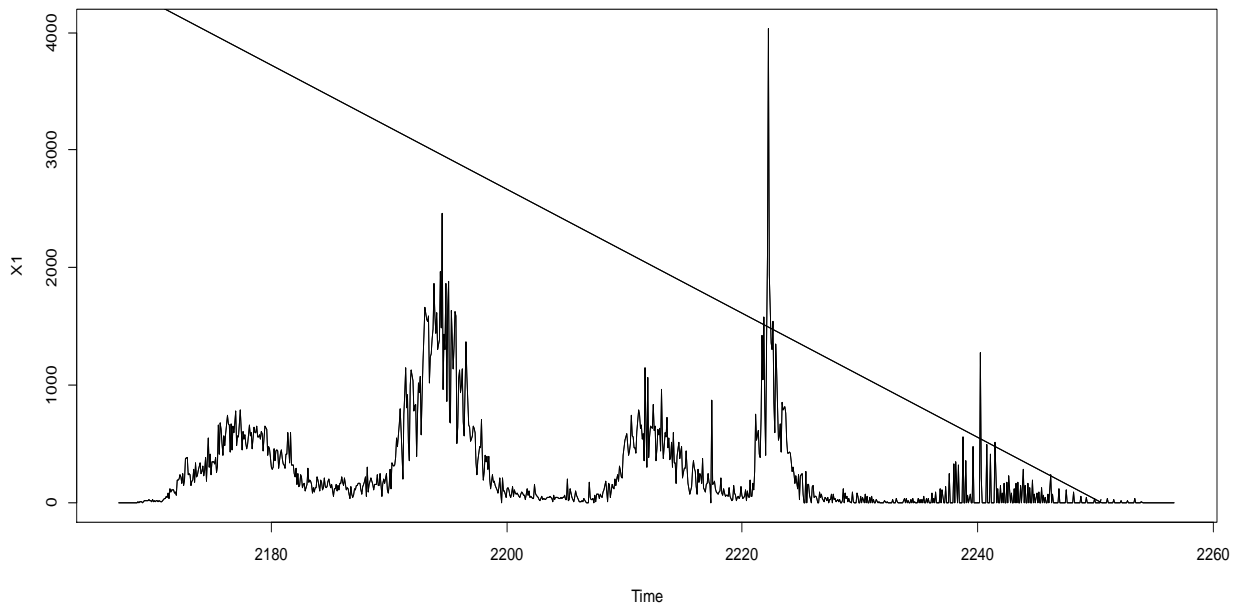


Figure 8: Plot of Confirmed Cases of Daily COVID-1

The Figure 8 above indicates clearly that the confirm cases of COVID-19 in Nigeria was not constant but rather varied from one year to the other as well as from one month to the other

with no systematically visible pattern, structural breaks, outliers, and no identifiable trend components in the time series data or no consistently increasing. It is therefore not stationary.

Table 9: Test of Stationarity using Augmented Dickey-Fuller (ADF) Test

Cases of CONVID-19	ADF Value	P-Value	Lag Order	H ₀	Decision	Remark
Total Confirmed	-3.8283	0.01758	10	No Stationarity	Reject H ₀	Stationary

Table 9 above presents ADF analyses of confirmed cases of COVID-19 with values -3.8283, with a *p*-value of 0.01758, which is less than the critical value of (α) 0.05, we reject the

null hypothesis of having a unit root series and therefore conclude that the data series is stationary and no need for differenced.

Table 10: Model Identification process

Model	Parameter				AIC	BIC	MSE
	AR(1)	AR(2)	MA(1)	MA(2)			
ARIMA(1,0,1)	0.9821 (0.0060)	-	-0.6401 (0.0269)	-	14230.9	14250.9	3.1776
ARIMA(1,0,2)	0.9848 (0.0054)	-	-0.5510 (0.0304)	0.1182 (0.0276)	14214.9	14239.9	3.1249
ARIMA(2,0,1)	0.8341 (0.0421)	-0.1677 (0.0408)	-0.7262 (0.0270)	-	14216.1	14241.1	3.1284
ARIMA(2,0,2)	0.7341 (0.1548)	-0.0188 (0.1522)	-0.5696 (0.1522)	-0.1063 (0.0994)	14216.9	14246.8	3.1259

Based on the parameters as reported in Table 10 above, The estimated coefficient values of all ARIMA (p,d,q) strictly conforms to the bounds of parameter, between -1 and 1. This has made the models to be stationary. Additionally, comparing the ARIMA models above interms of

the AIC, BIC and MSE, ARIMA(1,0,2) clearly slightly prefer to other models since their estimated AIC, BIC and MSE are smaller as compared to other models. Indeed, based on the parameter estimates and the criteria, ARIMA (1, 0, 2) is chosen as the best model to fit daily COVID-19 data.

Model Adequacy (Diagnostic) Checking of Estimated Models

Table 11: Ljung-Box Test for ARIMA (1,0,2) Model

Test Type	Chi-square Value	P-value	Decision
Ljung-Box	105.853	0.4371	Do not reject null hypothesis

The hypothesis that the Ljung-Box test is:

Null hypothesis (Ho): The residuals are uncorrelated

Alternative hypothesis (H₁): The residuals are correlated

The test is significant and its corresponding null hypothesis is rejected, if the p-value is less that chosen critical value of 0.05. From Table 11, the Ljung-Box test for the COVID-19 data, the chi-square statistic of 105.853 gives a corresponding p- value of 0.4371. Because the p-value is quite large (greater than the usually chosen a-level of 0.05), the test is not significant and therefore we do not reject the null

hypothesis, thus the residuals appear to be uncorrelated. This indicates that the residuals of the fitted ARIMA(1,0,2) model are white noise, and for that matter the model fits the series quite well (the parameters of the model are significantly not different from zero and the residuals are uncorrelated), so we can use this model to make forecasts.

Table 12: Shapiro-Wilk Normality Test for ARIMA (1,0, 2) Model

Test Type	W	P-value	Decision
Shapiro-Wilk Normality	0.9579,	0.4128	accept Null hypothesis

From the table 12, p-value of 0.4128 is quite large relatively to chosen a-level of 0.05, the test is not significant and therefore we do not reject the null hypothesis that the residual is normal. Thus, the Shapiro-Wilk test suggests that the standardized residuals are normal. This also supports the fact that the residuals of the fitted ARIMA (1,0,2) model are white noise, and for that matter the model fits the series quite well (since one of the assumptions of the residual being white noise is normality). Hence the model is stationary due to the presence of white noise.

Forecasting for Future COVID-19 Using the Best Fitted Model [ARIMA(1,0,2)]

A 12-step ahead out-of-sample forecast was conducted based on the data of the internal COVID-19 collection from the inception to February (2023) . The forecast is visually displayed in Figure 9. The 12 days of 2020-2023 were then forecasted based on data from the preceding time intervals. The forecast was obtained by using data from the previous periods to estimate COVID-19 changes for the actual period by applying the ARIMA (1,0,2) models. It can be seen in Figure 9 that the forecast is quite accurate, and hereby confirms what was expected. It seems that ARIMA (1,0,2) does a very good job in capturing the dynamic nature of the data and forecasting.

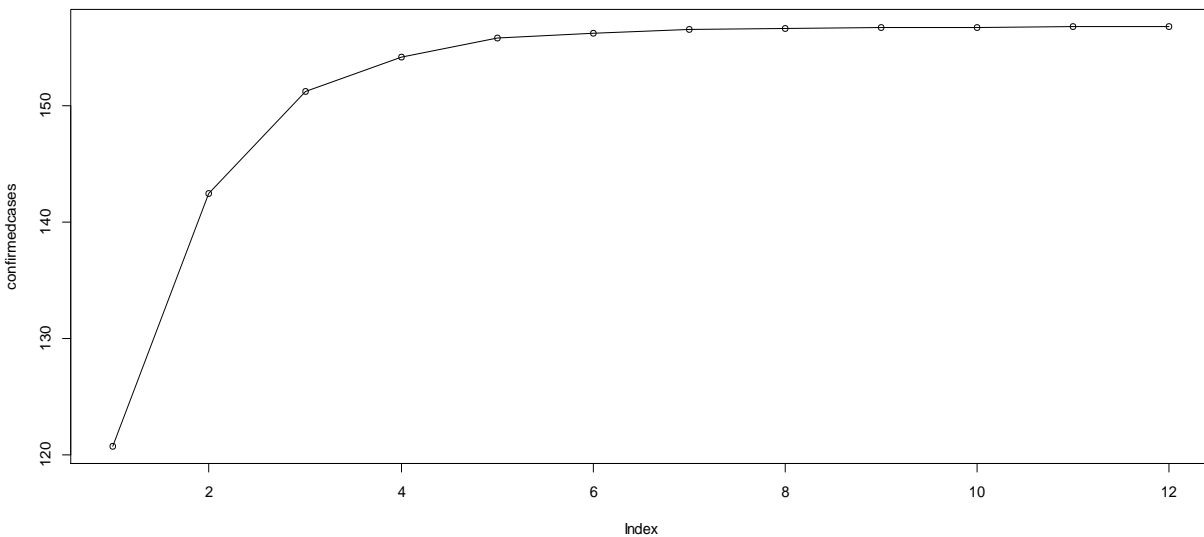


Figure 9: Forecasts from an ARIMA (1, 0, 2) Model for COVID-19 Confirmed Cases

CONCLUSION

In this study, Comparative performance of the ARIMA models of different orders was carried out on Poison data. It can be seen that, in the comparative performances of the models ARIMA (2,1,2) models are obviously preferred to be the best model that captured the Poison data when $\lambda = 5$ using the selected criteria while

ARIMA (1,1,1) is chosen as the best for the data when $\lambda = 10, 20$. Based on the forecasting ability of the models, the ARIMA (2,1,2) has the highest forecasting power due to their values greater than 1 and also greater than other values of the models across the steps ahead; this is followed by ARIMA (1,1,2) and ARIMA

(2,1,1). However, the Theil values of ARIMA (1,1,1), at higher steps ahead are close to zero, hence it is not as good as other models in forecasting. Indeed, the forecasting ability of all the models decrease as steps ahead increase.

Lastly, a class of ARIMA models that can appropriately describe the daily level of COVID-19 cases in Nigeria from the period of Inception to February(2023) are identified. The model fit to the data as defined by errors in the forecast both on aggregate and selective terms are adequate. Also the best model has an enhanced COVID-19 forecast. Therefore, based on the parameter reported in table 10, the estimated coefficient values of all ARIMA (p,d,q) strictly conforms to the bounds of the

parameter between -1 and 1 and this has made the model to be stationary comparing the ARIMA models in terms of the AIC, BIC and MSE. It is clearly preferred ARIMA (1,0,2) to other models since their estimated AIC, BIC and MSE are smaller as compared to other models. The forecast is quite accurate, and hereby confirms what was expected. It seems that ARIMA (1,0,2) does a very good job in capturing the dynamic nature of the data and forecasting. It was observed that the confirmed cases increase, and then keep stationary throughout the future times. The ARIMA (2,1,2) and ARIMA (1,1,1) models are recommended in capturing poison data with parameter $\lambda = 5$, $\lambda = 10$ and $\lambda = 20$.

REFERENCES

- Abdullahi, S. Karaman and Tayfur, Altok (2004). An experimental study on Forecasting using test processes proceeding of water simulation conference, R.G Ingalls, M.D. Rossetti, J.S Smith, and B.A Peters, eds., 437-442.
- Amon, M., Dimitry K. & Silas, M. (2015). Analyzing the impact of historical data length in non-seasonal ARIMA models Forecasting. *Mathematical theory and modelling*, 5(10), 77 – 85.
- Box, G. E. P., & Jenkins, G. M. (1976). Time series Analysis Forecasting and Control. Holden-Day Inc., San Francisco.
- Hillmer, S. C. & Tiao, G. C. (1982). An ARIMA model based Approach to seasonal Adjustment, *Journal of the American statistical Association*, 77, 63-70.
- Imam, A. (2020). Investigation of Parameter Behaviors in Stationarity of Autoregressive and Moving Average Models through Simulations. *Asian Journal of Mathematical Sciences*, 4(4), 21-32.
- Imam, A. (2022). Selecting Best Arima Model for Poison Data at Different Parameter Values. *Nigerian Journal of Mathematics and Application*, 32(2), 1 – 14.
- Nwankwo, S. C. (2014). Autoregressive integrated moving average (ARIMA) model for exchange rate (Naira to Dollar). *Academic Journal of Interdisciplinary Studies*, 3(4), 429-433.
- Tambi, M. K. (2005). Forecasting exchange rate: A Univariate out of sample approach. *The IUP Journal of Bank Management*, 4(2), 60-74.