International Journal of Healthcare and Medical Sciences
ISSN(e): 2414-2999, ISSN(p): 2415-5233
Vol. 12, Issue. 1, pp: 12-21, 2026
URL: https://arpgweb.com/journal/journal/13
DOI: https://doi.org/10.32861/ijhms.121.12.21

Academic Research Publishing Group

Original Research Open Access

Modelling and Forecasting of Measles cases in Nigeria Using the Integer Autoregressive (INAR) Model with Negative Binomial (NB) Distribution Hybrid

Nasiru M. Olakorede (Corresponding Author)
No. 15 Sokode Crescent, Wuse Zone 5, Abuja
Department of Statistics, Faculty of Science, University of Abuja, Abuja, Nigeria
Email: nasiru.olakorede@uniabuja.edu.ng

Sa'adatu Liman Ahmed
No. 15 Sokode Crescent, Wuse Zone 5, Abuja
Department of Statistics, Faculty of Science, University of Abuja, Abuja, Nigeria

Zira Sharon Dia
Department of Statistics, Faculty of Science, University of Abuja, Abuja, Nigeria

Article History
Received: 21 December, 2025
Revised: 26 January, 2026
Accepted: 12 February, 2026
Published: 22 February, 2026

Abstract

This study examines the measles case series in Nigeria from 2018 to 2022, analyzing trends, modeling, and forecasting using the INAR model with NB hybrid. The average number of measles cases per year was 1084 persons, with a minimum of 254 persons and a maximum of 2924 persons The total number of measles cases recorded during this period was 53,114 persons The data was found to be non-normally distributed (p-value = 0.005), overdispersed (p-value < 0.05), and autocorrelated (p-value < 0.05). An Integer AutoRegressive (INAR) model of order 1 was found to be the most suitable model, with an estimated ar1 parameter of -06885 (p-value < 0.05). The model was combined with a Negative Binomial distribution hybrid to fit the data, with an overdispersion parameter of k=203 and a mean of 108394 The hybrid modeling approach, with lead time forecasting method, predicted that measles cases will tend to increase in the nearest future, with a minimum record of 1407 persons and a maximum record of 1443 persons The results provide insights into the dynamics of measles cases in Nigeria and inform public health strategies for combating measles outbreaks.

Keywords: Modelling ; Forecasting ; Integer ; Autoregressive ; INAR ; NB ; Hybrid.

1. Introduction

Measles is a highly contagious disease caused by a virus known as rubella virus. Before the introduction of measles vaccine in 1963 and widespread vaccination, major epidemics occurred approximately every 2 – 3 years and measles caused an estimated 2.6 Million deaths each year according to W.H.O. the disease remains one of the leading causes of death among young children globally, despite the availability of a safe effective vaccine. Approximately 110000 people died from measles in 2017 – mostly children under the age of 5 years "World Health Organization WHO", Measles is caused by a virus in the paramyxovirus family and it is normal passed through direct contact and through the air. The virus infects the respiratory tract, then spread throughout the body. Measles is a human disease and is now known to occur in animals. Accelerated immunization activities have had a major impact on reducing measles deaths. In 2016, an estimated 90000 people died from measles an 84% drop from more than estimated 550000 in 2000 according to a new report published by leading organization. This mark the first time global measles death has fallen below 100,000 per year according to joint news release CDC/GAVI/UNICEF/WHO.

Measles cases in Nigeria is becoming alarming and worrisome as quite a number of cases has been recorded in the past, with clear evidence of what might be responsible for the unstable cases recorded. In this study, a univariate time series model will be employed, to identify the model that best describe the measles cases data in order to predict future cases with a view to aiding planning strategies for eradication of measles in Nigeria.

A time series is a collection of observations y_t, observed sequentially with time t. For continuous time series, the observations are measured continuously over some time interval, while for discrete-time series, the observations are measured at sequential integer values over fixed time intervals. In this study, we employed the discrete-time series Integer-valued Autoregressive (INAR) model with Negative Binomial distribution hybrid.

The primary objective of time series analysis is to establish a hypothetical probability model, the time series model, to represent the data. Once an appropriate family of models is chosen, it becomes possible to estimate the parameters in the models, assess the goodness-of-fit to the data, and potentially use the proposed models to enhance understanding of the underlying mechanism. Upon obtaining a satisfactory model, it can be utilized for further study, such as making predictions for future observations or applying it in a specific field.

Recently, there has been increased attention in the literature towards analysis cases of measles using different approaches. This interest stems at ensuring that the health agencies are assisted in their measles prevention and containment efforts. But building an appropriate time series forecasting model still remain an area yet to be fully developed.

Stephanie *et al*. (2024) estimated county-level measles risk using a machine learning model with 17 predictor variables, which was trained on 2014 and 2018 united states county-level measles case data and tested on data from 2019. They compared the predicted and actual location of 2019 measles cases. The study concluded that the machine learning model accurately predicted a majority of the united states counties at high risk for measles and could be used as a framework by state and natural health agencies in their measles prevention and containment efforts.

Timothy and Oluwaseyi (2023), employed co-integration technique to investigate the effect of measles outbreaks on under-five mortality in Nigeria between 1986 and 2021. The result of their findings shows a positive effect of measles outbreaks on under-five deaths in Nigeria.

Abdul-Kuddus and Mohiuddin (2023), employed mathematical modelling framework to explore the different interventions scenarios for programmatic measles control in Bangladesh from 2000 and 2019. The study concluded that scenario which combines enhanced treatment for exposed and infected population, first and second doses of vaccine is the most effective at rapidly reducing the total number of measles incidence and under-5 mortalities in Bangladesh.

Tsegaye, *et al.* (2022), employed the used of descriptive and logistic regression to model measles outbreaks in Guradamole District of Bale Zone, South Eastern Ethiopia. They identified 98 cases with overall attack rate of 12/1000 population and a case fertility rate (CFR) of 7%.

Nomhwange *et al.* (2022), employed retrospective research method and descriptive statistics to assess measles outbreaks response immunization during the COVID-19 pandemic in Borno State, Nigeria. The study found that a total of 181,634 children aged 9 months were vaccinated against measles with 27,961 receiving the measles vaccine for the first time.

In recent years, the urgency of addressing measles outbreaks has gained considerable attention amid global public health challenges. As highlighted in previous studies, there is a pressing need for robust analytical methods to inform health policy and intervention strategies, particularly within the context of Nigeria, where measles remains a significant public health concern. This contribution aims to extend the existing body of literature by providing an empirical comprehensive time series analysis of measles cases in Nigeria from 2018 to 2022, utilizing a hybrid model that incorporates a negative binomial distribution to address the over-dispersed nature of the incidence data.

The negative binomial distribution is particularly suited for modeling count data with over-dispersion, a common characteristic in public health data, especially for infectious diseases like measles. Previous studies have indicated the effectiveness of traditional machine learning and co-integration approaches, yet the application of time series methodologies with appropriate statistical forecasting model has not been adequately explored in the literature. The proposed analysis will harness both autoregressive and moving average components alongside the negative binomial framework to better capture the temporal dynamics and variability in measles cases. This framework has been comprehensively looked at in the paper by Nasiru and Olanrewaju (2023), where they carried out a simulation study to investigate the forecasting accuracy of count data using the Integer-Valued first order Autoregressive INAR(1) model with negative binomial distribution innovation.

The objectives of this study were to:

i. examine the trends and patterns of Measles cases in Nigeria;
ii. fit an integer-valued autoregressive moving average model with negative binomial hybrid to the data; and
iii. employ the best fitted model to forecast future cases of measles in Nigeria

2. Methodology

2.1. Integer Autoregressive Moving Average (INARMA) Models

The INARMA models for count time series constitute an Integer-valued counterpart to the conventional autoregressive moving-average (ARMA) models. They adapt the ARMA recursion to the count-data case using binomial thinning operation. The basic INARMA models have stochastic properties being analogous to those of the stationary ARMA models, and also several extensions have been developed. In particular, there exist modifications of the INARMA approach to deal with bounded counts or multivariate counts.

The INARMA (p,q) process is given by the following difference equation as defined by Alzaid and Al-Osh (1990)

$$y_t = \sum_{i=1}^{p} \alpha_i \circ y_{t-i} + z_t + \sum_{i=1}^{q} \beta_i \circ z_{t-i} \quad (2.1)$$

Where y_t is the integer-valued time series of data, α is the integer-valued autoregressive model's parameters, β is the integer-valued moving average model's parameters, $\alpha_i, \beta_i \in (0,1)$, z_t is the innovation distribution, and "$\circ$" thinning operator are performed independently of each other and also of the corresponding operations at previous times.

2.2. The Binomial Thinning Operator

Binomial thinning (or thinning) operation, commonly known as Galton-Watson branching process, are probabilistic operators that can be used to handle integer values. This popular binomial thinning operator was primarily introduced by Steutel and Van Horn (1979) to defined discrete self-decomposability, or precisely of infinitely divisible if $\psi(t) = [\psi(t)]^n$ for all positive integer n, where $\psi(t)$ is itself a characteristic function. The fundamental interpretation and generalization of binomial thinning operation are reviewed in this section:

Suppose Y is a non-negative integer-valued random variable. Then, for any $\alpha \in [0,1]$, the thinning operation "$\circ$" is defined by:

$$\alpha \circ Y = \sum_{i=1}^{y} x_i \quad (2.2)$$

Where $\{X_i\}$ is a sequence of i.i.d. Bernoulli random variables, independent of Y, and with a constant probability that the variable will take the value of unity:

$$P\ X_i = 1 - P\ X_i = 0 = \alpha \quad (2.3)$$

Some of the properties of the thinning operation can be obtained as follows:

(1) $0 \circ Y = 0$

(2) $1 \circ Y = Y$

(3) $\alpha \circ (\beta \circ Y) \stackrel{d}{=} (\alpha\beta) \circ Y$

(4) $(\circ Y) = \alpha\ E(Y)$

(5) $E\ (\alpha \circ Y)^2 = \alpha^2 (\ Y^2\ + \alpha\ 1 - \alpha\ (Y)$

(6) var $\alpha \circ Y = \alpha^2$var $Y + \alpha\ 1 - \alpha\ E(Y)$

2.3. Integer-Valued First Order Autoregressive (INAR(1)) Model

The Integer-valued first order Autoregressive INAR(1) model is defined by:

$$y_t = \alpha \circ y_{t-1} + z_t \quad (2.4)$$

Where $\alpha \in (0,1)$, and z_t is a sequence of i.i.d non-negative integer-valued random variables, independent of $y_t \sim (\mu_{z,}\ \sigma^2{}_{z,})$, z_t and y_{t-1} are assumed to be stochastically independent for all points in time, and the thinning operator "$\circ$" is defined via:

$$\alpha \circ y = \sum_{i=1}^{y} x_i \quad (2.5)$$

Where x_i is a sequence of independently and identically distributed (i.i.d.), Bernoulli random variables, independent of y, and with a constant probability that the variable will take value of unity.

$$P(x_t=1) = 1\text{-}p(x_t=0) = \alpha \quad (2.6)$$

The process obtained by equation (2.4) is stationary and it resembles the Gaussian AR(1) process except that it is nonlinear due to the thinning operation "$\circ$" replacing the scalar multiplication in continuous models.

Equation (2.4) shows that, based on the definition of the thinning operation, the memory of an INAR(1) model decays exponentially (Al-Osh and Alzaid, 1987).

Brännäs and Hellstrorn (2001), have studied in details; the two independence limitations we have assumed so far — independence of $\{X_i\}$ in the thinning operation, and independence of z_t and Y_{t-1}.

It is pertinent to add that the probability α is assumed to be constant here. Alzaid and Al-Osh (1993) developed a model in which this probability of retaining an element is not constant. Also, Zheng *et al.* (2007) developed a random coefficient model where α_t are i.i.d. random variables that can take values in the interval [0,1).

The mean and variance of the process $\{Y_t\}$are respectively:

$$E\ (Y_t) = a^t\ E\ (Y_0) + \ \mu_z\ \sum_{j=0}^{t-1} a^j \quad (2.7)$$

$$var(Y_t) = \ a^{2t}\ var\ Y_0 + \ (1 - a) \sum_{j=1}^{t} a^{2j-1}\ E\ \left(Y_{t-j}\right) + \ \sigma_z^2\ \sum_{j=1}^{t} a^{2\,(j-1)} \quad (2.8)$$

2.3. The Yule-Walker Estimation For Integer-Valued First Order Autoregressive (INAR(1)) Model

The Yule-Walker estimator for α in an INAR(1) model was found by Al-Osh and Alzaid (1987) to be as follows:

$$\hat{\alpha} = \frac{\sum_{t=0}^{n-1}(Y_t - \bar{\bar{Y}})(Y_{t+1} - \bar{\bar{Y}})}{\sum_{t=0}^{n}(Y_t - \bar{\bar{Y}})^2} \quad (2.9)$$

where $\bar{Y}$ is the sample mean.

2.4. Conditional Least Squares Estimation For Integer-Valued First Order Autoregressive (INAR(*1*)) Model

It can be easily seen that in the INAR(1) model, Y_t given Y_{t-1} is still a random variable due to the definition of the thinning operation. The conditional mean of Y_t given Y_{t-1}, which is the best one-step-ahead predictor (Brännäs and Hall, 2001), is:

$$E\ (Y_t/\ Y_{t-1} = \alpha Y_{t-1} + \lambda = (\boldsymbol{\theta}, Y_{t-1}) \quad (2.10)$$

where $\boldsymbol{\theta}=(\alpha,\lambda)'$ is the vector of parameters to be estimated. Al-Osh and Alzaid (1987) employ a procedure developed by Klimko and Nelson (1978) and derive the estimators for α as follows:

$$\hat{\alpha} = \frac{\sum_{t=1}^{n}\ Y_t Y_{t-1} - (\sum_{t=1}^{n}\ Y_t\ \sum_{t=1}^{n}\ Y_{t-1})/n}{\sum_{t=1}^{n} Y_{t-1}^2 - (\sum_{t=1}^{n} Y_{t-1})^2/n} \quad (2.11)$$

2.5. Forecasting over Lead Time

In this section, forecasting over a lead time is discussed. Lead time forecasting has applications in many areas, particularly in time series analyses.

2.5.1. Lead Time Forecasting for an INAR(1) Model

For the INAR(1) process of $Y_t = \alpha\ ^\circ\ Y_{t-1} + Z_t$, the cumulative Y over lead time l is given by:

$$\sum_{j=1}^{l+1} Y_{t+j} = Y_{t+1} + Y_{t+2} + \cdots + Y_{t+l+1}$$
$$= (\alpha \circ Y_{t-1} + Z_{t+1}) + (\alpha^2 \circ Y_t + \alpha \circ Z_{t+1} + Z_{t+2})$$
$$+ \cdots + (\alpha^{l+1} \circ Y_t + \alpha^l \circ Z_{t+1} + \alpha^{l-1} \circ Z_{t+2} + \cdots + Z_{t+l+1}) \quad (2.12)$$

Because $\alpha \circ X + \beta \circ X \neq (\alpha + \beta) \circ X$, the above equation can be written as:

$$\sum_{j=1}^{l+1} Y_{t+j} = \sum_{j=1}^{l+1} \sum_{i=1}^{n_j^1} \psi_{ij}^1 \circ Y_t + \sum_{j=1}^{l+1} \sum_{i=1}^{n_j^2} \psi_{ij}^2 \circ Z_{t+k_{ij}} \quad (2.13)$$

Where n_j^1 is the number of Y_t terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ in equation (2.12), ψ_{ij}^1 is the corresponding coefficient for each Y_t, n_j^2 is the number of $Z_{t+k_{ij}}$ terms in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ in equation (2.12), ψ_{ij}^2 is the corresponding coefficient for each $Z_{t+k_{ij}}$. All of these terms are explained below.

It can be seen that because the process is an integer autoregressive of order one, each of $\{Y_{t+j}\}_{j=1}^{l+1}$ yields only one Y_t, in equation (2.12); therefore, n_j^1=1. The corresponding coefficient for Y_t in each of $\{Y_{t+j}\}_{j=1}^{l+1}$ (sayY_{t+2}) is obtained from α thinned the coefficient of Y_t in the previous term (in this case Y_{t+1}). As a result, $\psi_{ij}^1 = \alpha^j$.

It can be seen from equation (2.12) that due to the repeated substitution of Y_{t+j}, the number of $Z_{t+k_{ij}}$ increases in each of $\{Y_{t+j}\}_{j=1}^{l+1}$. This number, shown by n_j^2, can be obtained from n_{j-1}^2+1. This means that each of $\{Y_{t+j}\}_{j=1}^{l+1}$ (sayY_{t+2}) has one of more Z compared to the previous one (which is Y_{t+1}in this case). The corresponding coefficient for each $Z_{t+k_{ij}}$ shown by Ψ_{ij}^2, is α thinned the corresponding coefficient in the previous term $\alpha \circ \Psi_{1(j+1)}^2$.) $t + k_{ij}$ is the subscript of innovation terms in each of $\{Y_{t+j}\}_{j=1}^{l+j}$ and from equation (2.12) it can be easily seen that k_{ij} is given by

$$k_{ij} = \begin{cases} k_{i(j-1)} & for\ 1 \le i \le n_{j-1}^2 \\ j & for\ n_{j-1}^2 < i \le n_j^2 \end{cases} for\ j = 1, \ldots, l+1 \quad (2.14)$$

Based on the equation (2.13), the conditional expected value of the aggregated process:

$$E\left(\sum_{j=1}^{l+1} Y_{t+j} \mid Y_t\right) = \left(\sum_{j=1}^{l+1}\sum_{i=1}^{n_j^1} \Psi_{ij}^1\right) Y_t + \left(\sum_{j=1}^{l+1}\sum_{i=1}^{n_j^2} \Psi_{ij}^2\right)\mu = \frac{\alpha(1-\alpha^{l+1})}{1-\alpha} Y_t$$
$$+ \left(\sum_{j=1}^{l+1}\sum_{i=1}^{j} \alpha^{i-1}\right)\mu = \frac{\alpha(1-\alpha^{l+1})}{1-\alpha} Y_t + \frac{\mu}{1-\alpha}\left[(l+1) - \sum_{j=1}^{l+1} \alpha^j\right] \quad (2.15)$$

Therefore, at time T, when Y_T is observed, the lead time forecast can be obtained from:

$$E\left(\sum_{j=1}^{l+1} Y_{T+j} \mid Y_T\right) = \frac{\alpha(1-\alpha^{l+1})}{1-\alpha} Y_T + \frac{\mu}{1-\alpha}\left[(l+1) - \sum_{j=1}^{l+1} \alpha^j\right] \quad (2.16)$$

2.6. The Negative Binomial (Nb) Distribution

The innovation distribution assumed in this study is the negative binomial distribution. The negative binomial distribution has two parameters: the mean μ and the shape parameter or the dispersion parameter k, which is commonly considered to be fixed to measure overdispersion. For a sample of counts X that fits a negative binomial distribution (X ~ NB(u, k)), the variance of the distribution is

$\mu + \mu^2 / k$. The probability that the variable X takes the value x is:

$$\text{Prb[X=x]} = \frac{\Gamma(x+k)}{x!\Gamma(k)}\left(\frac{\mu}{\mu+k}\right)^x \left(1+\frac{\mu}{k}\right)^{-k} = \frac{(x+k-1)(x+k-2)\ldots(k+1)k}{x!}\left(\frac{\mu}{\mu+k}\right)^x \left(1+\frac{\mu}{k}\right)^{-k}, \mu, k>0,$$

x=0,1,2,… (2.17)

Where $\Gamma(.)$ denotes the gamma function defined by:

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt. \quad (2.18)$$

From the probability density function of the negative binomial distribution, it can be seen that k is an essential part of the model. Estimation of k is thus important given a sample of counts.

3. Analysis and Results

This section presents the utilized data and analysis results obtained from the study. This comprises the descriptive analysis of the occurrences of Measles in Nigeria from the year 2018 to 2022 The data were secondarily sourced from statista data website (https://wwwstatistacom/statistics/1126801/suspected-measles-cases-in-nigeria-by-status/) The analysis also includes the trend analysis and adoption of a hybrid of Negative Binomial distribution and Integer ARMA in forecasting the Measles occurrences. The data was analyzed with the aid of R statistical package.

3.1. Descriptive Statistics and Trend Analysis

Table 3.1 presents the descriptive statistics of the monthly occurrences of Measles in Nigeria from the year 2018 to 2022 As observed from the table, for the periods examined, the average, minimum and maximum recorded

persons with measles disease were 1084-persons, 254-persons and 2924-persons respectively Also, the table depicts a total of 53,114-persons were recorded to have the disease between January 2018 to January 2022 The skewness, kurtosis and Jarque-bera statistics reveal the spread (ie normality nature) of the data over for the period examined However, while relying on the Jarque-bera probability value of 0005, the study deduced non-normality for the recorded measles disease Thus, subsequent chat presents the time series plot of the recorded measles disease

Table 3.1. Descriptive Statistics

	Measles (MSL)
Mean	1084
Median	801
Maximum	2924
Minimum	254
Std Dev	76944
Skewness	114
Kurtosis	291
Jarque-Bera	1067
Probability	0005
Sum	53114
Observations	49

Source: Researchers' Computations from R output

Moreover, Figure 3.1 presents the time series plot of the recorded measles disease in Nigeria According to the figure, the number of persons recorded with measles were relatively constant at roughly 500-persons monthly, between January 2018 to January 2019 The recorded measles cases rose abruptly in February 2019 to 2348-persons and increased significantly to the highest record of 2924-persons in September 2019 The recorded measles cases dropped sharply in January 2020 and was relatively constant at roughly 834-persons monthly, between January 2020 to December 2021 The recorded measles cases were observed to drop to the lowest record of 254-persons in January 2022

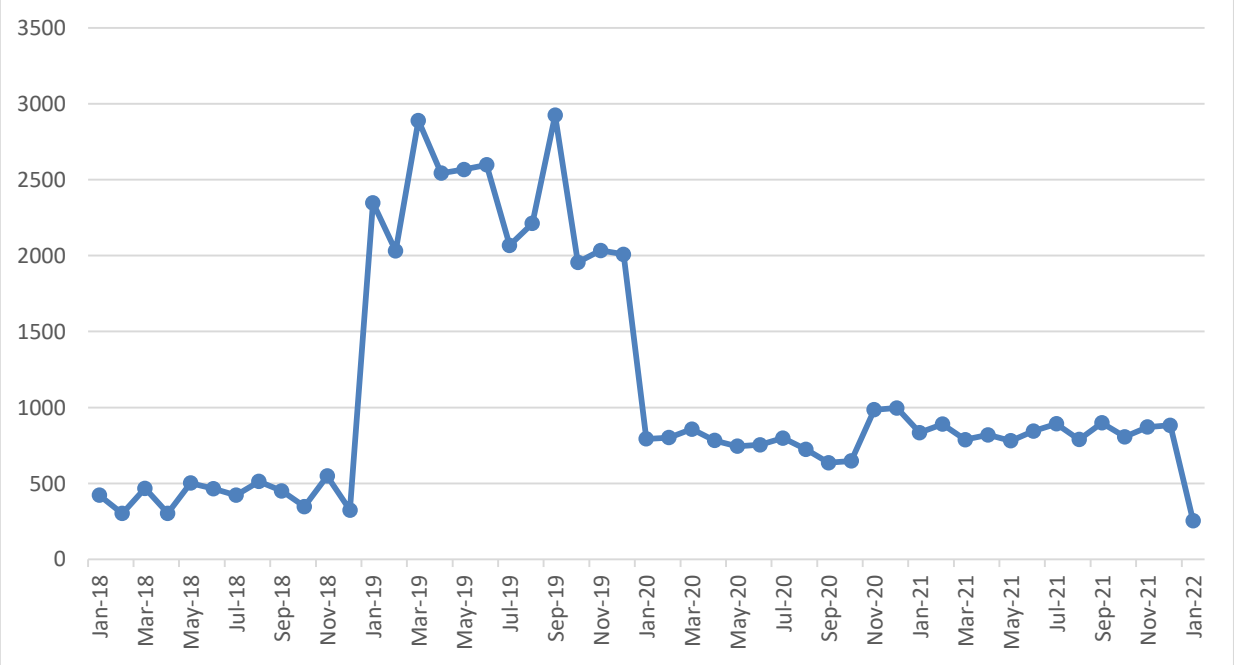

Figure 3.1. Time Series Plot of the Measles Occurrences

3.2. Modelling and Forecasting of the Measles Occurence

This section present headways as well as analysis results of the adoption of a hybrid of Negative Binomial distribution and Integer ARIMA in forecasting the Measles occurrences

3.2.1. Overdispersion Test, Autocorrelation Test and Negative Binomial Distribution Fit

Table3.2 reports the results of the Overdispersion test, Autocorrelation test as well as Negative Binomial distribution fit of the recorded measles diseases According to the Overdispersion test results, the null hypothesis of no true dispersion can be rejected for measles series since the test $p - value\ (12735e - 07) < 0.05$. Thus there is presence of overdispersion in the recorded measles cases series Likewise, according to the autocorrelation test results in Table 3.2, the null hypothesis (i.e., no autocorrelation in the series residuals) can be rejected for the series since the test $p - value\ (22e - 16) < 0.05$. Thus, there is presence of autocorrelation in the recorded measles cases

series Consequently, following the established of overdispersion and autocorrelation in the recorded measles cases series, the negative binomial distribution was fitted for the series The results returned 203 and 108394 as the overdispersion parameter (k) and mean of the recorded measles series respectively

Table-3.2. Over-Dispersion and Autocorrelation Tests Results

Test	Statistic	P-value
No True Overdispersion	Z=5009	2735e-07 < 005
No True Autocorrelation	X^2=96704	22e-16 < 005
Negative Binomial Distribution Fit	k=203 μ=108394	AIC=77041 Loglikelihood =-38320

3.2.2. Stationarity Test: Augmented Dickey Fuller (ADF)

Table3.3 reports the Augmented Dickey-Fuller (ADF) test also known as Stationarity test results for the recorded measles diseases The ADF-test returns non-stationarity for the measles series at level and first difference Evidence from the second difference stationarity test return a significant p-value for the test, which signify the rejection of null hypothesis (i.e., non-stationary) in favour of the alternative hypothesis Therefore, the recorded measles disease integration order is 2

Table-3.3. Stationarity Test (ie ADF) Results

Series	Order of Integration	Lag	Dickey-Fuller	p-value
Measles (MSL)	2	3	-43304	001 < 005

Source: Researchers' Computations

3.2.3. Model Identification: Correlogram

In the course to estimate an efficient integer autoregressive type model for the recorded measles cases, it commences with the examination of the series correlogram Figure3.2 and Figure3.3 present the series ACF and PACF plots respectively As observed the ACFs depict a wave decaying of the lags while the PACFs depict a cut-off at lag 1 of the series Hence, the correlogram (i.e., ACF and PACF plots) clearly signposts that the series follows AR order model process Therefore, following the behaviours of both ACF and PACF plots candidate models were examined, next section discusses this in details

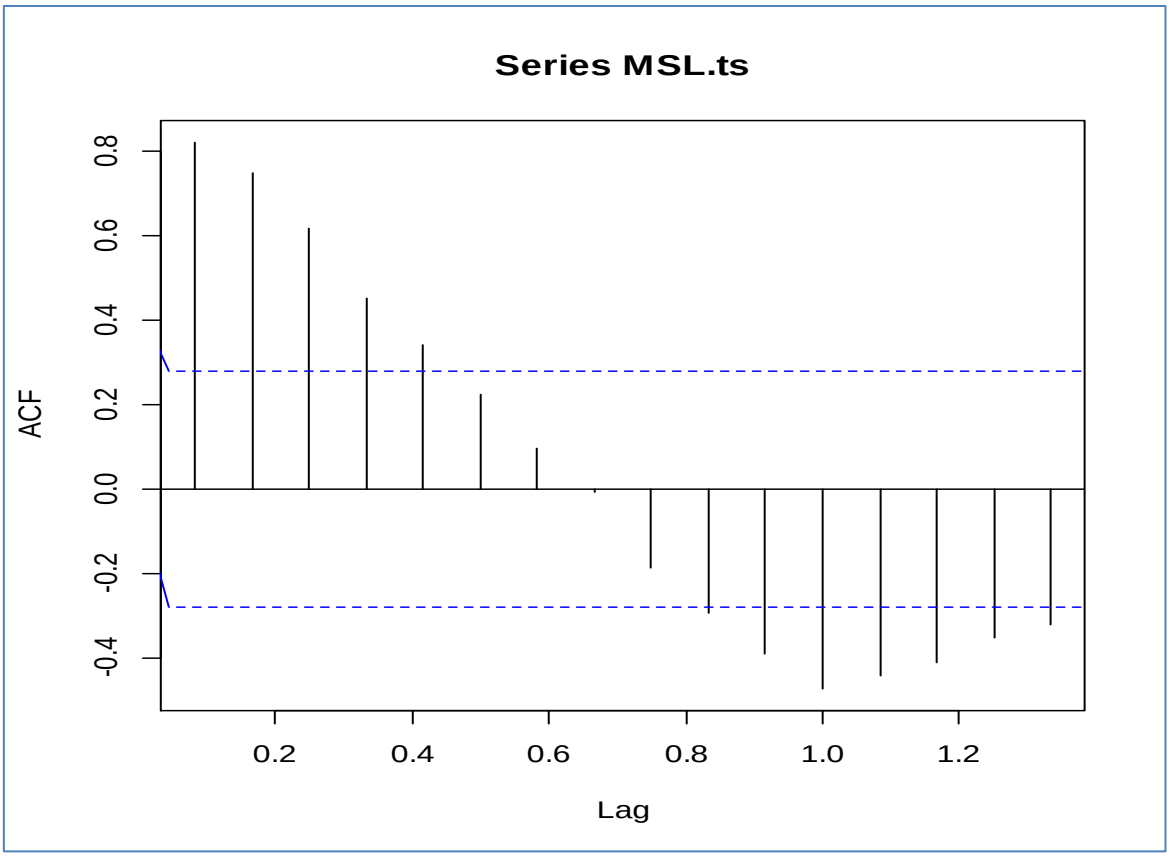

Figure-3.2. ACF Plots of the Measles Series

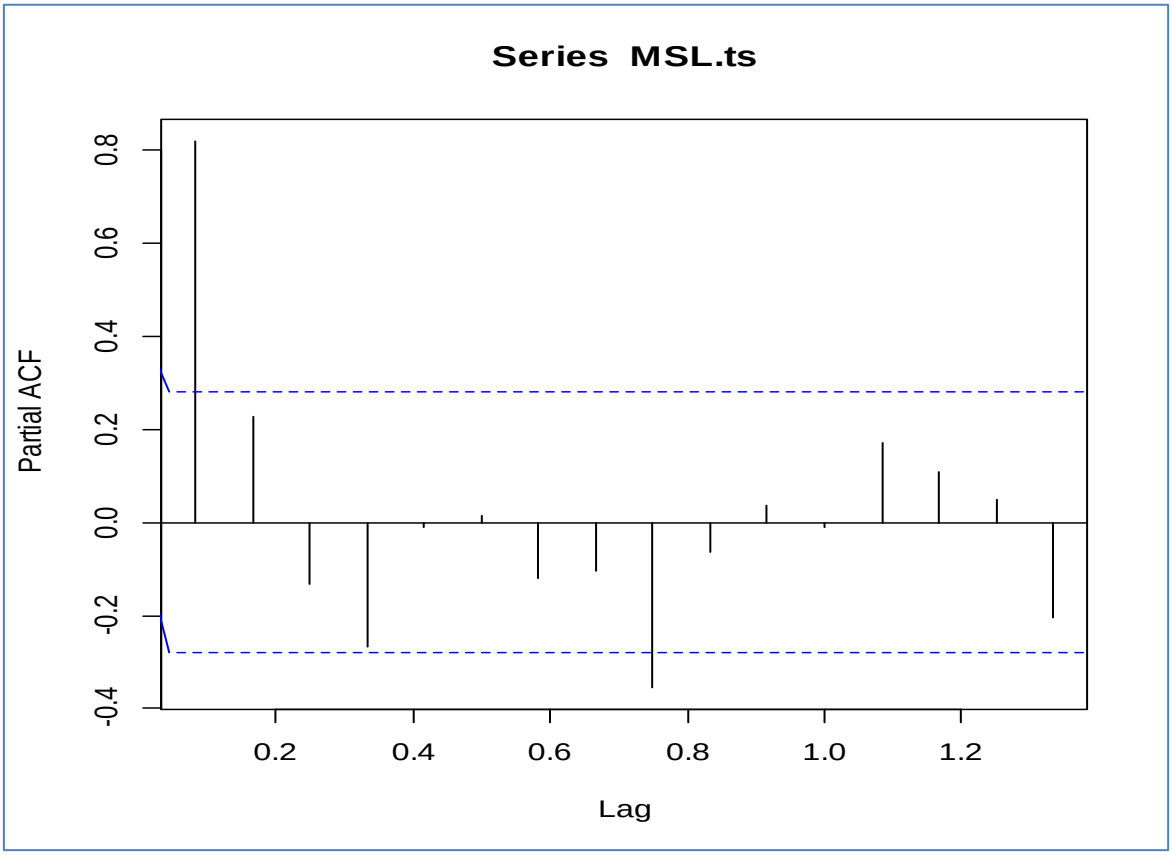

Figure-3.3. PACF Plots of the Measles Series

3.2.4. Integer Autoregressive (INAR) Model Estimation

Prior to the estimation of the INAR model for the prediction of the recorded measles cases in Nigeria, Table3.4 presents a summary of candidate INAR models with their respective log-likelihood values in the pursuit to select the most parsimonious and robust model Using the specification measures, among the 5 estimated INAR models for the recorded measles cases, INAR(1) returned with the highest Log-likelihood of -36028 Thus, INAR(1) returns as the most parsimonious INAR model for the recorded measles series Table3.5 presents the model's estimations Following the aforementioned, Table3.5 presents the model's estimation of most parsimonious INR models for the measles disease. The conditional least square estimator, as proposed by Nasiru and Olanrewaju (2023), was employed in this research to estimate the parameter of INAR(1) model. The INAR(1) estimation for the recorded measles cases reveals significant for the estimated ar1 parameter (since its p-value 1359e-10 < 0.05) with coefficient of -06885 Also, Table3.5 depicts the log-likelihood value of -36028 and a model variance level of 266509 The $INAR(1)$ model for the measles case is therefore stated as:

$$MSL_t = -06885 MSL_{t-1} \quad (3.1)$$

Consequently, the model was further diagnosed for goodness of fit. The results are subsequently discussed

Table-3.4. Integer Autoregressive Candidate Models

Model	Log-likelihood
INAR(1)	**-36028**
INAR(2)	-35487
INAR(3)	-35513
INAR(4)	-35557
INAR(5)	-35487

Source: Researchers' Compilations from R Outputs

Table-3.5. $INAR(1)$ Model Estimation for the Recorded Measles Cases

Parameters	Estimate	Std Error	z-value	p-value	Model Summary
ar1	-06885	01072	-64205	1359e-10 < 005	Log-likelihood = -36028 Sigma^2 = 266509

Source: Researchers' Compilations from R Outputs

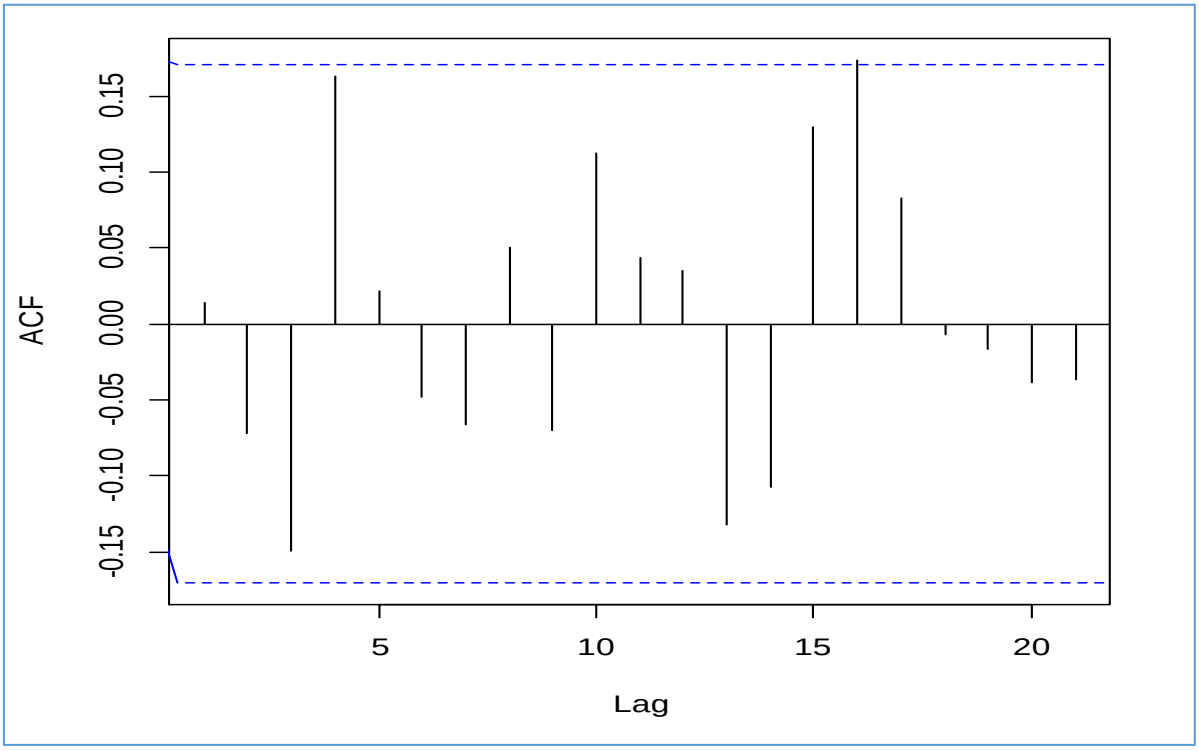

Figure-3.4. ACF Plots of the $INAR(1)$ Residuals

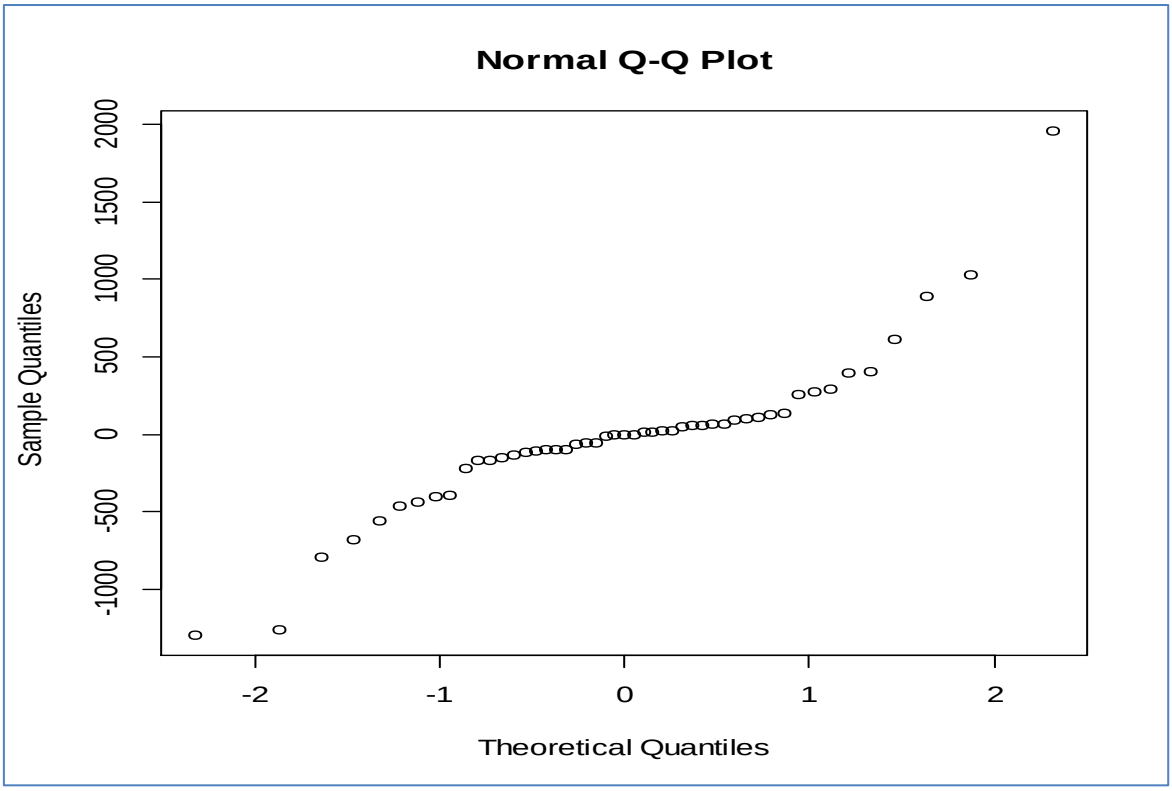

Figure-3.5. Normal Q-Q Plot of the $INAR(1)$ Residuals

Table-3.6. Ljung-Box Test of the Estimated Models

Lag	X-squared	P-value
1	00279	0567
2	07236	0566
3	37774	0387
4	74798	0213
5	75434	0153

Source: Researchers' Compilations from R Outputs

According to Figure3.4 and Figure3.5 depicting the ACF plots and Normal Q-Q plots of $INAR(1)$ residuals respectively, the ACF plots of the model return that none of the ACF lags was seen to be significant. Similarly, the Normal Q-Q plots show that the residuals are normal as virtually all residual rest on a line and are not all over the place Lastly, Ljung-Box test results in Table3.6 depicts no any p-values is less than 0.05 level, hence the results show no trace autocorrelation in the residuals of the model Based on the foregoing, it can be deduced that the model $INAR(1)$ was appropriately fitted and statistically suitable to be adopted for forecasting of the recorded measles cases in Nigeria.

3.2.5. The Hybrid of INAR(1) and Negative Binomial Distribution (NBD) for the Forecast of Measles Cases

Following that all the goodness of fit tests are in support of the assumption that there is no pattern in the residuals of the models, the study progresses to forecast the recorded measles cases by combining the estimated INAR(1) parameter and NBD parameter in the Lead Time Forecast model The lead time forecast is model as follows:

$$Y_{T+j} = \frac{\alpha(1-\alpha^{l+1})}{1-\alpha} Y_T + \frac{\mu}{1-a}\left[(l+1) - \sum_{j=1}^{l+1} \alpha^j\right] \qquad 3.2$$

Following the estimated INAR(1) model the $\alpha = ar1 = -06885$ and also based on the estimated NBD the $\mu = 108394$ And for the lead time l = 1, $Y_T = 254$ (the last point for the measles cases to be forecast), this study obtain the point forecast at this stage couple with the minimum mean square error That is: For l = 3, Y_T=the point forecast for l = 1; For l = 5, Y_T= the point forecast for l = 3; For l = 7, Y_T= the point forecast for l = 5; For l = 9, Y_T= the point forecast for l = 7; and For l = 15, Y_T= the point forecast for l = 9

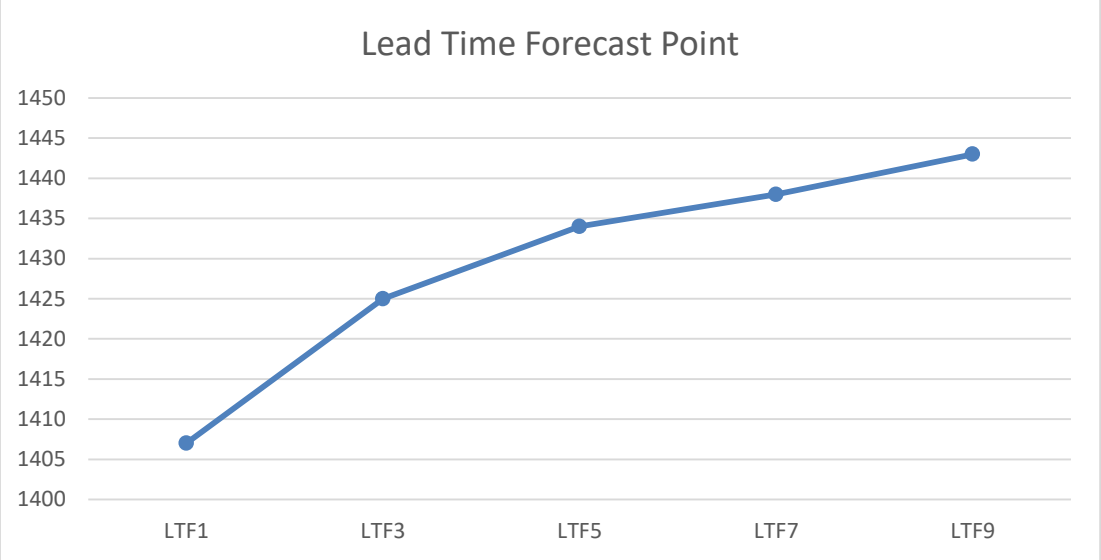

Figure-3.6. The Lead Time Forecast Points Graph of the Hybrid of INAR(1) and NBD) Forecast Model

Table-3.7. Hybrid of INAR(1) and NBD) for the Lead Time Forecast of Measles Cases

Forecast	Lead Time Forecast				
	1	3	5	7	9
Forecast Point	140655 ≡ 1407	142525 ≡ 1425	143411 ≡ 1434	143831 ≡ 1438	144333 ≡ 1443
MSE	6840548	6964678	7025963	7055562	7091392

Figure3.6 and Table3.7 present the hybrid lead time forecast model performances and accuracy measure criteria, i.e., Mean Square Error (MSE) According to Figure3.6 and Table3.7 which present the lead time forecast for lead time 1, 3, 5, 7, and 9 remarkably reveal that the recorded measles cases would tend to increase in the nearest future as the results depict a minimum record of 1407-persons and maximum record of 1443-persons

4. Summary of Findings

In summary**,** the average number of measles cases in Nigeria from 2018 to 2022 was 1084 persons, with a minimum of 254 persons and a maximum of 2924 persons The total number of measles cases recorded during this period was 53,114 persons The Jarque-Bera test revealed that the data is not normally distributed (p-value = 0005) The time series plot shows that the number of measles cases was relatively constant at around 500 persons per month between January 2018 and January 2019. The number of cases then increased sharply in February 2019 and reached a peak of 2924 persons in September 2019. The number of cases dropped sharply in January 2020 and remained relatively constant at around 834 persons per month until December 2021. The number of cases dropped to a low of 254 persons in January 2022. The overdispersion test revealed that there is overdispersion in the measles case series

(p-value < 005). The autocorrelation test revealed that there is autocorrelation in the measles case series (p-value < 005). The Augmented Dickey-Fuller (ADF) test shows that the measles series is non-stationary at level and first difference, but stationary at second difference, with the integration order of 2.

The correlogram plots showed that the series follows an AR order model process Five candidate INAR models were examined, and the INAR(1) model was found to be the most parsimonious and robust model. The estimation of the INAR(1) model shows that the estimated ar1 parameter was -06885, which is significant at 5% level (p-value < 005) The ACF plots and Normal Q-Q plots of INAR(1) residuals shows no significant patterns or autocorrelation in the residuals The Ljung-Box test also shows no evidence of autocorrelation in the residuals.

The negative binomial distribution was fitted to the data, and the results showed that the overdispersion parameter (k) was 203 and the mean of the recorded measles series was 108394 The study employed a hybrid modeling approach, combining the Negative Binomial distribution The lead time forecast model was estimated using the INAR(1) model parameters and NBD parameters The model predicted that the recorded measles cases would tend to increase in the nearest future, with a minimum record of 1407 persons and a maximum record of 1443 persons.

5. Conclusion

Based on our findings, the following conclusions were drawn:

i. Measles cases exhibited a non-constant pattern, with varying numbers of cases from one month to another, displaying overdispersion and autocorrelation in the time series
ii. The INAR(1) model provided the best fit to the data, emerging as the most parsimonious and robust model Furthermore, the estimated parameter was statistically significant (p < 005), supporting its relevance
iii. Using a hybrid modeling approach that incorporated the fitted INAR(1) model and the NBD parameter with lead time forecasting, our results suggest that measles cases are likely to increase in the near future, with a projected range of 1407 to 1443 cases

References

Abdul-Kuddus, D., Rahman, A., & Mohidin, M. (2023). Analysis of the different interventions scenario for programmatic measles control in Bangladesh: A modelling study. *PLOS ONE, 18(6): e0283082*. https://doi.org/10.1371/ journal. pone.0283082.

Al-Osh, M. A. and A. A. Alzaid (1987). First order integer-valued autoregressive (INAR(1)) processes. *Journal of Time Series Analysis* 8(3): 261-275.

Alzaid AA and Al-Osh MA (1993) Generalized Poisson ARMA processes. Annals of the Institute of Statistical Mathematics, 45, 223–32.

Alzaid, A. A. and M. Al-Osh (1990). An integer-valued pth-order autoregressive structure (INAR(p)) process. *Journal of Applied Probability* 27(2): 314-324.

Brännäs, K. and J. Hellström (2001). Generalized integer-valued autoregression. *Econometric Reviews* 20(4): 425-443.

Klimko, L. A. and P. I. Nelson (1978). On conditional least squares estimation for stochastic processes. *Annals of Statistics* 6(3): 629-642.

Nasiru Mukaila Olakorede & Olanrewaju Samuel Olayemi (2023). Integer-Valued First Order

Autoregressive (INAR(!)) Model with Negative Binomial (NB) Innovation for the Forecasting of Time Series Count Data. International Journal of Statistics and Probability; Vol.12, No.6; November 2023. URL: https://doi.org/10.5539/ijsp.v12n6p23

Nomhwange, T., Abede, M., Jean Baptiste, A.E., Musa, A., Yusuf, A., Luce, R., & Masresha

B. (2022). Measles outbreak response immunization during the COVID-19 pandemic: Lesson from Borno State, Nigeria. *Pan African Medical Journal. 41(104)*. https://www.panafrican-med-journal.com//content/article/41/104/full

Stephanie A. Kujowski, Boshu Ru, Nelson Lee Afanador, James H. Conway, Richard Boumgartner, Monjiri Pawskar (2024). Prediction of Measles Cases in US Counties: A Machine Learning approach. Volume 42, Issue 26, 2 December 2024, 126289.

Stutel, F. W. and K. van Harn (1979). Discrete analogues of self-decomposability and stability. *Annals of Probability* **7**(5): 893-899.

Tsegaye, G., Gezegn, Y., Bulcha, G.G., & Berhanu, N. (2022). Measles Outbreak Investigation in Guradamole District of Bale Zone, South Eastern Ethiopia. *Journal of Taylor and Francis Group*. https://www.tandfonline.com/loi/didr20.

Zheng, H., I. V. Basawa and S. Datta (2007). First-order random coefficient integer- valued autoregressive processes. *Journal of Statistical Planning and Inference* 137(1): 212-229