An ARMA Model for Short-term Prediction of Hepatitis B Virus Seropositivity among Blood Donors in Lafia-Nigeria

David Adugh Kuhe¹* and Thomas Akwana Obed²

¹Department of Mathematics/Statistics/Computer Science, University of Agriculture, Makurdi, Benue State, Nigeria.
²Department of Basic Sciences, College of Agriculture, Lafia, Nassarawa State, Nigeria.

Authors’ contributions

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JSRR/2019/v24i130142

Received 02 February 2019
Accepted 13 April 2019
Published 18 July 2019

Original Research Article

ABSTRACT

In this paper, we attempt to search for an optimal Autoregressive Moving Average (ARMA) model that best forecast hepatitis B virus infection among blood donors in Lafia-Nigeria. The study uses monthly data in Lafia-Nigeria for the period of 11 years 6 months from January 2007 to June 2018. The data was obtained as secondary data from General Hospital Lafia and Dalhatu Araf Specialist Hospital, Lafia. The time series and stationarity properties of the data are explored using time plots and Dickey-Fuller Generalized Least Squares unit root test. The results indicate that the series is integrated of order zero, I(0). An ARMA (p,q) model in line with Box-Jenkins procedure was employed to model the time series data. The result shows that ARMA (1,1) was the best candidate to model and forecast hepatitis B virus infection among blood donors in Lafia-Nigeria. Critical analysis of the model shows that the HBV infection is chronic among blood donors in the study area. The estimated ARMA (1,1) model was then used to forecast future values of hepatitis B infection among blood donors in Lafia-Nigeria from July 2018 to June 2019. The forecast shows a stable level of infection for the forecasted period. The study provided some policy recommendations.
1. INTRODUCTION

Hepatitis B is a highly contagious liver disease caused by infection with the hepatitis B virus (HBV). The hepatitis B virus is known as a blood-borne virus because it is transmitted from one person to another. The virus is spread when blood, semen, saliva, vaginal fluids (including menstrual blood) and other bodily fluids from an infected person enter the body of an uninfected person. Possible methods of transmission include: transfer from mother to baby during birth, being pricked with a contaminated needle, close contact with a person with HBV, sex (oral, vaginal, and anal), using an infected toothbrush or razor. Symptoms may not occur for a few days or longer after contracting the virus. However, one is still contagious, even without symptoms. Symptoms of hepatitis B may not be apparent for months or years. However, common symptoms include: dark urine, joint pain, loss of appetite, fever, abdominal discomfort, weakness, yellowing of the whites of the eyes (sclera) and skin (jaundice).

The complications of HBV without early treatment include: liver scarring (cirrhosis), liver failure, kidney cancer, kidney failure and liver cancer. Another possible complication is hepatitis D infection. It is only people with HBV that can contract hepatitis D. A combined infection can cause serious liver problems.

Hepatitis B virus infects liver cells (hepatocytes) and can cause both acute and chronic disease. Acute hepatitis lasts for less than 6 months while chronic hepatitis lasts for more than 6 months [1]. Acute infection does not usually require treatment. Most people overcome an acute infection on their own. Chronic infection requires antiviral medications for treatment which help in fighting the virus and may also reduce the risk of future liver complications [2]. Persons with chronic hepatitis B virus (CHB) are at risk of developing serious sequelae, such as cirrhosis and hepatocellular carcinoma [3].

Hepatitis B is one of the prevalent diseases in the world and a major cause of morbidity and mortality [4]. According to Lavanchy [5] an approximate population of 2 billion people worldwide has been infected with the hepatitis B virus (HBV). Despite the availability of highly effective vaccine against hepatitis B virus there are still over 350 million chronic carriers worldwide, of whom possibly one million die annually from cirrhosis and/or hepatocellular carcinoma [6]. HBV infection accounts for 0.6 to 1.2 million global deaths annually [7,8].

Empirical literature on modeling and forecasting of hepatitis B virus are well documented in the literature, see for example: [9] conducted a historical cohort study on HBV incidence in the Hamadan Province of west o Iran from 2004 to 2012. They employed Weighted Markov Chain (WMC) method and two time series models including Holt Exponential Smoothing (HES) and SARIMA model. The results of the different methods were compared to correct percentages of predicted incidence rates. The overall incidence rate of HBV was estimated to decrease over time. The comparison of results of the three models indicated that in respect of the existing seasonality trend and non-stationarity, the HES had the most accurate prediction of the incidence rates. Gan et al. [10] conducted a study to compare and evaluate the prediction of hepatitis in Guangxi Province of China using back propagation neural networks based genetic algorithm (BPNN-GA), generalized regression neural networks (GRNN), and wavelet neural networks (WNN). In order to compare the results of forecasting, the data obtained from 2004 to 2013 and 2014 were used as modeling and forecasting samples, respectively. The results showed that when the small data set of hepatitis had seasonal fluctuation, the prediction result by BPNN-GA was better than the other two methods. The WNN method was more suitable for predicting the large data set of hepatitis that had seasonal fluctuation; it was the same for the GRNN method when the data increased steadily. Wang et al. [11] modeled and compared ARIMA model and Grey model (GM (1,1) model) for forecasting hepatitis B incidence in China using monthly data from March, 2010 to October, 2017. ARIMA model showed better hepatitis B fitting and forecasting performance than GM (1,1) model. The forecast results indicated that hepatitis B incidence in China might have a slight fluctuation for the forecasted period of November, 2017 to March, 2018.

In Nigeria, HBV is reported to be the most common cause of liver disease. Several authors have reported on the prevalence of HBV among sub-populations in Nigeria with varying estimates depending on population studied and methods used. However, there is no reliable nationwide
2.2 Some Basic Concepts

Let \( \{HBV_i\} \) be a stochastic time series process. We define \( HBV_i \) as a sequence of hepatitis B virus infection indexed by time. We shall be using \( HBV_i \) to refer to a series throughout our study.

2.2.1 Autocorrelation Function (ACF)

We define the Autocorrelation function (ACF) of a stationary series \( \{HBV_i\} \) as:

\[
\rho_k = \frac{\text{cov}(HBV_i, HBV_{i-k})}{\sqrt{\text{var}(HBV_i)}\sqrt{\text{var}(HBV_{i-k})}} = \frac{\text{cov}(HBV_i, HBV_{i-k})}{\text{var}(HBV_i)}
\]

which is the OLS estimator in \( HBV_i = c + \rho_k HBV_{i-k} + e_i \). The 95% confidence bounds are given by \( \pm 1.96/\sqrt{T} \), where \( T \) is the number of observations.

2.2.2 Partial Autocorrelation Function (ACF)

The partial Autocorrelation Function (PACF) is the correlation between \( HBV_t \) and \( HBV_{t-k} \) after the data has been corrected for intermediate lags \( HBV_{t-1},...,HBV_{t-k+1} \). The PACF can be estimated as the OLS estimator \( \hat{\beta}_k \) in the regression

\[
HBV_t = \beta_1 HBV_{t-1} + \beta_2 HBV_{t-2} + \cdots + \beta_k HBV_{t-k} + e_t
\]

where the intermediate lags are included. Under the assumption of white noise, \( \beta_1 = \beta_2 = \cdots = 0 \), it holds that \( \text{var}(\hat{\beta}_k) = T^{-1} \).

2.2.3 Stationarity of order M

A stochastic time series process \( \{HBV_i\} \) is stationary of order \( M \) if for any admissible set \( \{t_1, t_2, ..., t_m\} \) and for any \( k \), the joint moments of \( \{HBV_{t_1}, HBV_{t_2}, ..., HBV_{t_m}\} \) up to order \( M \) exists, and are equal to the joint moments of \( \{HBV_{t_1+k}, HBV_{t_2+k}, ..., HBV_{t_m+k}\} \) up to order \( M \). That is \( E((HBV_{t_1})^\alpha(HBV_{t_2})^\beta \cdots (HBV_{t_m})^\gamma) = E((HBV_{t_1+k})^\alpha(HBV_{t_2+k})^\beta \cdots (HBV_{t_m+k})^\gamma) \) for all \( \alpha, \beta, \gamma \) such that \( \alpha + \beta + \cdots + \gamma \leq M \).

2.2.4 Weakly or covariance stationary

A stochastic time series process \( \{HBV_i\} \) is said to be weakly or covariance stationary if its mean and variance are constant over time and its covariance function depends only on the time lag. A covariance stationary series satisfies the following conditions:

(i) \( E(HBV_i) = \mu \), where \( \mu \) is a constant
(ii) \( E(HBV_i - \mu)^2 = \text{var}(HBV_i) = \sigma^2 \), where \( \sigma^2 \) is a constant and
(iii) \( E(HBV_i, HBV_{i+s}) = E(HBV_i, HBV_{i+s+k}) \) is a function of \( s - t = k \) only where \( k \) is the lag.

2.3 Model Specification

2.3.1 Autoregressive model

An autoregressive model of order one, AR (1) is specified as:

\[
HBV_t = \phi_0 + \phi_1 HBV_{t-1} + e_t
\]
where \( HBV_t \) is hepatitis B virus infection response variable at time \( t \), \( \varepsilon_t \) is a purely random process with mean zero and variance \( \sigma^2 \), \( \phi_0 \) is a constant and \( \phi_1 \) is an autoregressive parameter and the subscript \( 1 \) is the order of the autoregressive parameters which increase with increases in \( HBV_t \). The values of \( \phi \) which would make the process to be stationary are such that the roots of the polynomial equation \( \phi[L] = 0 \) lie outside the unit circle in the complex plane. \( L \) is the lag operator such that \( L_t \, HBV_t = HBV_{t-1} \) and \( \phi L = 1 - \phi L \).

### 2.3.2 Moving average model

Suppose that \( \{\varepsilon_t\} \) is a white noise process with mean zero and variance \( \sigma^2 \), then the process \( HBV_t \) is said to be a moving average model of order one, \( MA(1) \) if

\[
HBV_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} \tag{5}
\]

Where \( \beta_1 \) is the moving average parameter. The subscript on \( \beta_1 \) is called the order of moving average parameter.

### 2.3.3 Autoregressive moving average model

A stochastic process resulting from the combination of autoregressive and moving average models is called an Autoregressive Moving Average (ARMA) model. An ARMA model of order one, \( ARMA(1,1) \) is specified as:

\[
HBV_t = \phi_0 + \phi_1 HBV_{t-1} + \varepsilon_t - \beta_1 \varepsilon_{t-1} \tag{6}
\]

To obtain stationarity for this model the equation \( \phi[L] = 0 \) has its roots outside the unit circle and the root of \( \beta[L] = 0 \) must lie outside the unit circle for the process to be invertible. Equation (6) is the theoretical model which serves as a basic framework of our analysis.

### 2.4 Model Order Selection

We used the following information criteria for model order selection in conjunction with log likelihood function: Akaike Information Criterion (AIC) due to Akaike [13] and Schwarz information Criterion (SIC) due to Schwarz [14].

\[
AIC(P) = -2 \ln(L) + 2P \tag{7}
\]

\[
SIC(P) = -2 \ln(L) + P \ln(T) \tag{8}
\]

where \( P \) is the number of free parameters to be estimated in the model, \( T \) is the number of observations and \( L \) is the maximum likelihood function.

### 2.5 Some Statistical Tests

#### 2.5.1 Dickey-fuller generalized least squares (DF GLS) unit root test

If \( HBV_t \) is the series under investigation, the DF GLS test is based on testing:

- \( H_0: \psi = 0 \) (The series contains unit root) against
- \( H_1: \psi < 0 \) (The series is stationary) in the following regression

\[
\Delta HBV_t^d = \psi_0 HBV_{t-1}^d + \psi_1 \Delta HBV_{t-1}^d + \cdots + \psi_{p-1} \Delta HBV_{t-p+1}^d + u_t \tag{9}
\]

where \( HBV_t^d \) is the detrended series. Detrending depends on whether a constant or a constant and trend are included in the model. We reject \( H_0 \) if the DF-GLS test statistic is less than the critical value of the test at the designated test sizes. Elliot et al. [15] show that de-trending in this way produces a test that has good power properties.

#### 2.5.2 Portmanteau test

A Portmanteau test is a test used for investigating the presence of autocorrelation in time series. The test checks the following pairs of hypotheses:

- \( H_0: \rho_{k,1} = \rho_{k,2} = \cdots = \rho_{k,T} = 0 \) (all lags correlations are zero) against:
- \( H_1: \rho_{k,1} \neq \rho_{k,2} \neq \cdots \neq \rho_{k,T} \neq 0 \) (there is at least one lag with non-zero correlation). The test statistic is given by:

\[
Q^{(LR)} = T(T+2) \sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{T-k} \tag{10}
\]

where

\[
\hat{\rho}_k^2 = \frac{T}{T-k} \left( T \sum_{t=k+1}^{T} (\hat{\varepsilon}_t^2 - \bar{\varepsilon}) (\hat{\varepsilon}_{t-k}^2 - \bar{\varepsilon}) / \sum_{t=1}^{T} (\hat{\varepsilon}_t^2 - \bar{\varepsilon})^2 \right),
\]

for \( \bar{\varepsilon} = T^{-1} \sum_{t=1}^{T} \varepsilon_t^2 \)

denotes the autocorrelation estimate of squared standardized residuals at \( k \) lags. \( T \) is the sample size, \( Q \) is the sample autocorrelation at lag \( k \). We reject \( H_0 \) if p-value is less than \( \alpha = 0.05 \) level of significance [16].
2.6 Forecast and Forecast Evaluation

Suppose the sample we wish to forecast is \( j = T + 1, T + 2, \ldots, T + h \), and denote the actual and forecasted value in period \( t \) as \( HBV_t \) and \( \hat{HBV}_t \), respectively. The reported forecast error statistics are computed as follows:

Root Mean Square Error (RMSE) = \( \sqrt{\frac{\sum_{t=T+1}^{T+h} (HBV_t - \hat{HBV}_t)^2}{h}} \)

Mean Absolute Error (MAE) = \( \frac{\sum_{t=T+1}^{T+h} |HBV_t - \hat{HBV}_t|}{h} \)

Mean Absolute Percentage Error (MAPE) = \( 100 \times \frac{\sum_{t=T+1}^{T+h} |\frac{HBV_t - \hat{HBV}_t}{HBV_t}|}{h} \)

Theil Inequality Coefficient (TIC) = \( \frac{\sqrt{\sum_{t=T+1}^{T+h} (HBV_t - \hat{HBV}_t)^2}}{\sqrt{\sum_{t=T+1}^{T+h} HBV_t^2} / h + \sqrt{\sum_{t=T+1}^{T+h} \hat{HBV}_t^2} / h} \)

Bias Proportion (BP) = \( (\sum HBV_t / h - \hat{HBV}_t) / \sum HBV_t - \hat{HBV}_t)^2 / h \)

Variance Proportion (VP) = \( (S_{HBV} - S_{\hat{HBV}})^2 / \sum (HBV_t - \hat{HBV}_t)^2 / h \)

Covariance Proportion (CP) = \( 2(1 - r)S_{HBV}S_{\hat{HBV}} / \sum (HBV_t - \hat{HBV}_t)^2 / h \)

where \( h \) is the number of steps ahead that we want to predict, and \( T \) is the total sample size. For additional discussion of forecast evaluation see [17].

3. RESULTS AND DISCUSSION

3.1 Graphical Examination of the Series

The data generating process of the series are first examined using time plot after transforming the original series into natural logarithms. The result of time plot of the series is presented in Fig. 1.

The time plot of the transformed series reported in Fig. 1 indicates a stable and smooth trend which suggests that the mean and variance of the series are constant over time (homoskedastic). This means that the natural log of the series in level is weakly stationary. Although, we will further investigate this by considering the autoregressive function (ACF) and partial autoregressive function (PACF) of the series reported in Fig. 2.

The plots of ACF and PACF of the series reported in Fig. 2 suggest that the series in stationary in level since all the lags are inside the confidence bounds. This is an indication that the residual of the series are purely random process. This also shows that the series is independent of time (i.e., the infection in the current month does not depend on the infection of the previous month and vice versa). We also consider the Q-statistics for autocorrelation of the series. The result is presented in Table 1.

The p-values of the Q-statistics of the series reported in Table 1 are highly statistically insignificant. This is one of the properties of a dynamically stable and stationary series whose residuals are purely random process. The Q-statistics of the ACF thus help to confirm that the series is stationary in level.

3.2 Dickey Fuller (DF) GLS Unit Root Test

To further confirm the stationarity of the series in level as shown by the result of time plot, ACF and PACF plots as well as Q-statistic test and to know the order of integration of the series, we conduct unit root test in level of the series using Dickey-Fuller Generalized Least Squares unit root test procedure. The result of the DF-GLS unit root test in level is reported in Table 2.
Fig. 1. Time plot of hepatitis-B Infection in Lafia (Natural Log)

Fig. 2. ACF and PACF plot of Hepatitis-B infection in Lafia (Natural Log)

Table 1. Autocorrelation function and ljung-box Q-statistics

<table>
<thead>
<tr>
<th>Lag</th>
<th>ACF</th>
<th>Std. error</th>
<th>Ljung-Box Q-statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.0260</td>
<td>0.090</td>
<td>0.084</td>
<td>0.773</td>
</tr>
<tr>
<td>5</td>
<td>0.0374</td>
<td>0.089</td>
<td>6.270</td>
<td>0.281</td>
</tr>
<tr>
<td>7</td>
<td>-0.0433</td>
<td>0.088</td>
<td>6.711</td>
<td>0.460</td>
</tr>
<tr>
<td>8</td>
<td>-0.0343</td>
<td>0.087</td>
<td>6.865</td>
<td>0.551</td>
</tr>
<tr>
<td>10</td>
<td>0.0604</td>
<td>0.087</td>
<td>7.548</td>
<td>0.673</td>
</tr>
<tr>
<td>11</td>
<td>0.0132</td>
<td>0.086</td>
<td>7.571</td>
<td>0.751</td>
</tr>
<tr>
<td>12</td>
<td>-0.1322</td>
<td>0.086</td>
<td>9.943</td>
<td>0.621</td>
</tr>
</tbody>
</table>
Table 2. Elliot-rothenberg and stock DF-GLS unit root test results

<table>
<thead>
<tr>
<th>Option</th>
<th>DF-GLS test statistic</th>
<th>DF-GLS test critical values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1%</td>
</tr>
<tr>
<td>Intercept only</td>
<td>-8.8306*</td>
<td>-2.5845</td>
</tr>
<tr>
<td>Intercept &amp; Trend</td>
<td>-10.3205*</td>
<td>-3.5572</td>
</tr>
</tbody>
</table>

Note: * denotes the significant of the DF-GLS test statistic at 1% significance level

The unit root test result reported in Table 2, shows that the series is weakly stationary in level since the DF-GLS test statistics both with intercept only and with intercept and trend are all less (or more negative) than the critical values of the test at the conventional test sizes. This shows that the series is integrated of order zero, I(0). That is, stationary in level. Having obtained the order of integration of the series, we proceed with other analysis using the stationary series.

3.3 Selection of Model Order

The spikes of ACF and PACF in Fig. 3 both decayed quickly to zero. This suggest a mixed ARMA model for the series while the DF-GLS unit root test shows the order of integration of the series to be zero, I(0). We need to marry these two basic ideas to search for an optimal ARMA (p,d,q) model using information criteria, log likelihood and $R^2$ statistic bearing in mind that $d = 0$. The result is reported in Table 3.

The result of Table 3 indicates that ARMA (1,1) model has the least information criteria, largest log likelihood and highest $R^2$. Based on Box-Jenkins procedure, this seems to describe our time series data more adequately. We therefore select ARMA (1,1) as the best candidate to model and forecast hepatitis B virus infection among volunteer blood donors in Lafia, Nassarawa state-Nigeria. The parameter estimates of ARMA (1,1) are presented in Table 4.

Table 3. Model order selection using information criteria and Log likelihood

<table>
<thead>
<tr>
<th>S/n</th>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
<th>LogL</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ARMA (0,1)</td>
<td>-1.6222</td>
<td>-1.5758</td>
<td>99.3339</td>
<td>0.0012</td>
</tr>
<tr>
<td>2</td>
<td>ARMA (0,2)</td>
<td>-1.6668</td>
<td>-1.5971</td>
<td>103.0081</td>
<td>0.0606</td>
</tr>
<tr>
<td>3</td>
<td>ARMA (0,3)</td>
<td>-1.6692</td>
<td>-1.6043</td>
<td>104.9034</td>
<td>0.0871</td>
</tr>
<tr>
<td>4</td>
<td>ARMA (1,0)</td>
<td>-1.6181</td>
<td>-1.5714</td>
<td>98.2779</td>
<td>0.0007</td>
</tr>
<tr>
<td>5</td>
<td>ARMA (1,1)**</td>
<td>-1.7980</td>
<td>-1.6797</td>
<td>113.1629</td>
<td>0.6480</td>
</tr>
<tr>
<td>6</td>
<td>ARMA (1,2)</td>
<td>-1.6509</td>
<td>-1.5575</td>
<td>102.2314</td>
<td>0.0649</td>
</tr>
<tr>
<td>7</td>
<td>ARMA (1,3)</td>
<td>-1.6405</td>
<td>-1.5718</td>
<td>101.8241</td>
<td>0.0894</td>
</tr>
<tr>
<td>8</td>
<td>ARMA (1,4)</td>
<td>-1.6307</td>
<td>-1.5701</td>
<td>103.1862</td>
<td>0.1290</td>
</tr>
<tr>
<td>9</td>
<td>ARMA (2,1)</td>
<td>-1.6358</td>
<td>-1.5419</td>
<td>100.5134</td>
<td>0.0534</td>
</tr>
<tr>
<td>10</td>
<td>ARMA (2,2)</td>
<td>-1.6815</td>
<td>-1.5641</td>
<td>104.2075</td>
<td>0.1169</td>
</tr>
<tr>
<td>11</td>
<td>ARMA (2,3)</td>
<td>-1.6706</td>
<td>-1.5297</td>
<td>104.5628</td>
<td>0.1170</td>
</tr>
<tr>
<td>12</td>
<td>ARMA (2,4)</td>
<td>-1.6519</td>
<td>-1.5215</td>
<td>103.4832</td>
<td>0.1392</td>
</tr>
<tr>
<td>13</td>
<td>ARMA (3,1)</td>
<td>-1.6447</td>
<td>-1.5276</td>
<td>101.2161</td>
<td>0.0867</td>
</tr>
<tr>
<td>14</td>
<td>ARMA (3,2)</td>
<td>-1.7852</td>
<td>-1.6436</td>
<td>110.4354</td>
<td>0.2199</td>
</tr>
<tr>
<td>15</td>
<td>ARMA (3,3)</td>
<td>-1.6430</td>
<td>-1.4778</td>
<td>103.1172</td>
<td>0.1192</td>
</tr>
<tr>
<td>16</td>
<td>ARMA (3,4)</td>
<td>-1.6516</td>
<td>-1.4872</td>
<td>104.4437</td>
<td>0.1152</td>
</tr>
<tr>
<td>17</td>
<td>ARMA (4,1)</td>
<td>-1.6320</td>
<td>-1.4896</td>
<td>100.6567</td>
<td>0.0956</td>
</tr>
<tr>
<td>18</td>
<td>ARMA (4,2)</td>
<td>-1.6491</td>
<td>-1.4829</td>
<td>102.6471</td>
<td>0.1262</td>
</tr>
<tr>
<td>19</td>
<td>ARMA (4,3)</td>
<td>-1.7533</td>
<td>-1.5634</td>
<td>109.6932</td>
<td>0.2261</td>
</tr>
<tr>
<td>20</td>
<td>ARMA (4,4)</td>
<td>-1.7348</td>
<td>-1.5211</td>
<td>109.6162</td>
<td>0.2251</td>
</tr>
</tbody>
</table>

Note: ** denotes ARMA model selected by the criteria
3.4 ACF and PACF

We examine the adequacy and goodness of fit of the model by means of plotting the ACF and PACF of residuals. If all the sample autocorrelation coefficients of the residuals are within the 95% confidence bounds, then the residuals are white noise indicating that the model is a good fit. The ACF and PACF plots are presented in Fig. 3.

Fig. 3 shows that all the sample autocorrelation coefficients of the residuals are within the confidence bounds indicating that the residuals are white noise and the fitted model is stable and stationary.

We also conduct Ljung-Box Q-statistic test of serial correlation (autocorrelation) for residuals of the fitted model. The result of the test is presented in Table 5.

From the result of Table 5, the null Hypothesis of no serial correlation in the residuals of the fitted model at all lags is accepted since the p-values of the Q-statistics are all greater than 0.05. This shows that the estimated model is stationary and dynamically stable.

3.4.2 Stability and invertibility analysis

Another evidence to show that the estimated model is dynamically stable is that the inverse roots of the ARMA model is dynamically stable.

Table 4. OLS parameter estimates of ARMA (1,1) model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. error</th>
<th>t-statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>4.469523</td>
<td>0.004667</td>
<td>957.5877</td>
<td>0.0000</td>
</tr>
<tr>
<td>AR(1)</td>
<td>0.724005</td>
<td>0.146412</td>
<td>4.944999</td>
<td>0.0000</td>
</tr>
<tr>
<td>MA(1)</td>
<td>0.274561</td>
<td>0.101923</td>
<td>-8.580618</td>
<td>0.0000</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.647981</td>
<td>0.647981</td>
<td>0.274561</td>
<td>0.724005</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.431567</td>
<td>0.431567</td>
<td>0.146412</td>
<td>0.724005</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>113.1629</td>
<td>113.1629</td>
<td>7.923161</td>
<td>0.000035</td>
</tr>
<tr>
<td>F-statistic</td>
<td>7.923161</td>
<td>7.923161</td>
<td>1.86016</td>
<td>0.186016</td>
</tr>
<tr>
<td>Inverted AR Roots</td>
<td>0.72</td>
<td>Inverted MA Roots</td>
<td>0.87</td>
<td>Durbin W.</td>
</tr>
</tbody>
</table>

From the result of Table 4, the estimated ARMA (1,1) model is represented in equation (11):

$$H_{BV_t} = 4.469523 + 0.724005 H_{BV_{t-1}} + \varepsilon_t + 0.274561 \varepsilon_{t-1}$$

(11)

The result of equation (11) shows that the intercept (C) is positively related with hepatitis B infection and statistically significant. This implies that the predicted value of hepatitis B infection will be 4.469523 units in log form (i.e., approximately 87 persons) if all other explanatory variables are kept constant. The AR and MA slope coefficients of the model are all statistically significant at marginal significance levels. The estimated model have also satisfied the stationarity condition because $\alpha + \beta_1 = 0.724005 + 0.274561 = 0.998566 < 1$. This shows that the estimated ARMA (1,1) model is stationary.

The coefficient of determination ($R^2$) of the regression model is 0.647981 indicating that about 64.80% of the total variations in hepatitis B infection has been explained by independent variables while the remaining 35.20% unexplained variations is being accounted for by the error term or by factors not included in the model. The F-statistic is a goodness of fit test which measures the overall fitness of the regression parameters. F=7.923161 with a p-value of 0.000035 indicates that the regression model is a good fit. The value of Durbin Watson statistic is 1.86016 which is greater than $R^2$ and $R^2$ adjusted indicating that the model is not spurious.

3.4 Model Validation and Diagnostic Checks

We now validate our model by carrying out residual diagnostic check on the estimated ARMA (1,1) model.

3.4.1 ACF and PACF plots of residual

We examine the adequacy and goodness of fit of the model by means of plotting the ACF and PACF of residuals. If all the sample autocorrelation coefficients of the residuals are within the 95% confidence bounds, then the residuals are white noise indicating that the model is a good fit. The ACF and PACF plots are presented in Fig. 3.

Fig. 3 shows that all the sample autocorrelation coefficients of the residuals are within the confidence bounds indicating that the residuals are white noise and the fitted model is stable and stationary.

Table 5. Ljung-Box Q-statistic test for serial correlation of residuals

<table>
<thead>
<tr>
<th>Lag</th>
<th>Q-statistic</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1054</td>
<td>0.745</td>
</tr>
<tr>
<td>2</td>
<td>3.3254</td>
<td>0.190</td>
</tr>
<tr>
<td>3</td>
<td>3.4907</td>
<td>0.322</td>
</tr>
<tr>
<td>4</td>
<td>4.4911</td>
<td>0.479</td>
</tr>
<tr>
<td>5</td>
<td>4.0721</td>
<td>0.539</td>
</tr>
<tr>
<td>6</td>
<td>4.0775</td>
<td>0.666</td>
</tr>
<tr>
<td>7</td>
<td>4.1036</td>
<td>0.768</td>
</tr>
<tr>
<td>8</td>
<td>4.1080</td>
<td>0.847</td>
</tr>
<tr>
<td>9</td>
<td>4.1204</td>
<td>0.903</td>
</tr>
<tr>
<td>10</td>
<td>4.9234</td>
<td>0.896</td>
</tr>
<tr>
<td>11</td>
<td>5.0548</td>
<td>0.928</td>
</tr>
<tr>
<td>12</td>
<td>6.5083</td>
<td>0.888</td>
</tr>
</tbody>
</table>
roots of AR/MA polynomials are all within a unit circle as reported in Fig. 4.

From the root of AR and MA polynomials of the fitted model presented in Table 4, AR root = 0.72 and MA root = 0.87 and we estimate that \( \tan \theta = y/x = 0.72/0.87 = 0.8276 \) and \( \theta = 39.61^\circ \). Thus, the life cycle of hepatitis B virus infection among blood donor in the study area is \( 360^\circ/39.61^\circ = 9.09 \approx 9 \) months and we say that hepatitis B virus infection among blood donors in Lafia-Nigeria has a life cycle of 9 months which could be describe as chronic, a disease condition in which if not properly treated will lead to severe liver complications and high risk of developing serious sequelae, such as cirrhosis and hepatocellular carcinoma.

3.4.3 Forecast evaluation

Having validated our model, we now seek an appropriate forecast mode that best forecast future relevant series. Here we consider in-sample and out-of-sample forecasts using seven accuracy measures. The forecast mode with the least accuracy measures stands as the best to predict hepatitis B virus infection among blood donors in Lafia-Nigeria. The result of forecast comparison is presented in Table 6.

The accuracy measures automatically select out-of-sample forecast mode for our model. This is because the out-of-sample forecast has the least accuracy measures except for covariance proportion (CP).
facilities provided by the state government and other NGOs and international donors in the region. This result corroborates the empirical findings of Wang et al. [11] & Zhang et al. [12].

3.4.4 Short-term forecast of hepatitis B infection in Lafia-Nigeria

Using the out-of-sample forecast approach for the series, the estimated ARMA (1,1) model is used to forecast future values of hepatitis B virus infection among volunteer blood donors in Lafia-Nigeria for the period of 1 year (12 months) starting from July 2018 to June 2019. The result of the forecast is presented in Table 7.

The forecast value for the month of July 2018 is 86 persons with a 95% confidence interval of [71, 106] persons. By this we are 95% confident that the outcome for the next period will fall within this interval. Comparing with the monthly infection in June 2018 (85 persons), we predict that in July 2018 the hepatitis B virus infection will slightly increase at least by 1 person or at most by 20 persons (i.e. it may increase at least by 1 person or at most by 20 persons) in July 2018. The forecasts for the following months show a stable level in the virus infection in Lafia-Nigeria. The confidence intervals of the forecast suggest a stable level of infection during the forecasted period of July 2018 to June 2019. This implies that hepatitis B virus infection among volunteer blood donors in Lafia-Nigeria will remain stable within the years 2018 and 2019. This could possibly be as a result of better and improved control and preventive measures, enhanced awareness and campaign strategies, medical care and treatment.

4. CONCLUSION AND RECOMMENDATIONS

In this paper, attempt has been made to search for an optimal Autoregressive Moving Average (ARMA) model that best forecast hepatitis B virus infection among blood donors in Lafia-Nigeria. The study uses monthly data in Lafia-Nigeria for the period of 11 years 6 months from January 2007 to June 2018. The data comprises of 138 consecutive observations and was obtained as secondary data from General Hospital Lafia and Dalhatu Araf Specialist Hospital, Lafia. The time series and stationarity properties of the data are explored using time plot, ACF and PACF plots and Dickey-Fuller Generalized Least Squares unit root test. The results indicate that the series is stationary in level and hence integrated of order zero, I(0). An ARMA (p,q) model in line with Box-Jenkins procedure were employed to model the time series data. The result shows that ARMA (1,1) was the best candidate to model and forecast hepatitis B virus infection among blood donors in Lafia-Nigeria. The analysis of the model shows that hepatitis B infection is chronic among blood donors in Lafia-Nigeria. Persons with chronic HBV infection are at high risk of developing serious sequelae, such as cirrhosis and hepatocellular carcimoma. The estimated
ARMA (1,1) model was then used to forecast future values of hepatitis B infection among blood donors in Lafia-Nigeria from July 2018 to June 2019. The forecast shows a stable level of infection for the forecasted period.

Based on the findings of this study the following recommendations/suggestions are hereby presented:

i. To further reduce the spread of HBV, government in collaboration with public health authorities need to educate the community and health care providers about HBV transmission routes based on known HBV epidemiology in Lafia and its neighboring communities.

ii. Hepatitis B vaccine programme should be initiated with a target of reducing the infection rate from its current state.

iii. Future research should be carried out with focus on factors associated with hyper-endemic levels of HBV infection in the community.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES


