Original Research Article

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

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.

Keywords: Hepatitis B virus, Blood donors, ARMA Model, Forecasting, Nigeria.

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 carcimona [3].

Hepatitis B is one of the prevalent diseases in the world and a major cause of morbidity and mortality [4]. According to [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 of 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. Zhang et al. [12] modeled and predicted hepatitis B incidence in Iran using time series analysis.

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 survey of HBV exposure in the average risk population and in subgroups most likely to benefit from early detection, surveillance, and treatment. Vaccination against the hepatitis B virus (HBV) is lower in Nigeria than any other West African nation of the many Sub-Saharan African countries.

Due to the severe health impact of hepatitis B infection, there is a growing need for methods that will allow forecasting and early warning with timely case detection in areas of unstable transmission, So that effective control and preventive measures can be implemented. This study contributes and extends the existing literature by modeling and providing short-term forecasts on hepatitis B virus infection among blood donors in Lafia-Nigeria using time series techniques and more recent data.

2. MATERIAL AND METHODS

2.1 Data and Source

The data use for this study comprises serologically confirmed cases of hepatitis B virus infection among blood donors in Lafia town, Nassarawa state in Nigeria from January 2007 to June 2018. The data consists of 138 monthly observations of persons believed to be residents of Lafia town. The data was obtained as secondary data from the two tertiary health institutions in Lafia town. The General Hospital Lafia and Dalhatu Araf Specialist Hospital, Lafia, Nassarawa state-Nigeria.

2.2 Some Basic Concepts

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

2.2.1 Autocorrelation Function (ACF)

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

$$\rho_k = \frac{Cov(HBV_t, HBV_{t-k})}{Var(HBV_t)} = \frac{Cov(HBV_t, HBV_{t-k})}{\sqrt{Var(HBV_t).(HBV_{t-k})}} \quad (1)$$

where $\rho_0 = 1$ and $-1 \leq \rho_k \leq 1$ otherwise. The sample autocorrelation function can be estimated by:

$$\hat{\rho}_k = \frac{\frac{1}{T-k}\sum_{t=k+1}^{T}(HBV_t - \overline{HBV_t})(HBV_{t-k} - \overline{HBV_{t-k}})}{\frac{1}{T-k}\sum_{t=k+1}^{T}(HBV_{t-k} - \overline{HBV_{t-k}})^2} \quad (2)$$

which is the OLS estimator in $HBV_t = c + \rho_k HBV_{t-k} + \mathrm{e_t}$. 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 $HBV_{t-1}, \ldots, HBV_{t-k+1}$. The PACF can be estimated as the OLS estimator $\hat{\beta}_k$ in the regression

$$HBV_t = \sigma + \beta_1 HBV_{t-1} + \beta_2 HBV_{t-2} + \cdots + \beta_k HBV_{t-k} + e_t \quad (3)$$

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

2.2.3 Stationarity of Order M

A stochastic time series process $\{HBV_t\}$ is stationary of order M if for any admissible set $\{t_1, t_2, \ldots, t_m\}$ and for any k, the joint moments of $\{HBV_{t1}, HBV_{t2}, \ldots, HBV_{tm}\}$ up to order M exists, and are equal to the joint moments of $\{HBV_{t1+k}, HBV_{t2+k}, \ldots, HBV_{tm+k}\}$ up to order M. That is $E\{(HBV_{t1})^{\alpha}(HBV_{t2})^{\beta} \ldots (HBV_{tm})^{\gamma}\} = E\{(HBV_{t1+k})^{\alpha}(HBV_{t2+k})^{\beta} \ldots (HBV_{tm+k})^{\gamma}\}$ for all $\alpha, \beta, \ldots, \gamma$ such that $\alpha + \beta + \ldots + \gamma \leq M$.

2.2.4 Weakly or Covariance Stationary

A stochastic time series process $\{HBV_t\}$ 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) $\mathrm{E}(HBV_t) = \mu$, where μ is a constant (ii) $\mathrm{E}(HBV_t - \mu)^2 = \mathrm{Var}(HBV_t) = \sigma^2$, where σ^2 is a constant and (iii) $E(HBV_t, HBV_s) = E(HBV_t, HBV_{t+k})$ 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:

$$HBV_t = \phi_0 + \phi_1 HBV_{t-1} + \varepsilon_t \quad (4)$$

where HBV_t is hepatitis B virus infection response variable at time t, ε_t is a purely random process with mean zero and variance σ^2, ϕ_0 is a constant and ϕ_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 ϕ 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 $L_1 HBV_t = HBV_{t-1}$ 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 σ^2, then the process HBV_t is said to be a moving average model of order one, MA (1) if

$$HBV_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} \quad (5)$$

Where β_1 is the moving average parameter. The subscript on β_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:

$$HBV_t = \phi_0 + \phi_1 HBV_{t-1} + \varepsilon_t - \beta_1 \varepsilon_{t-1} \quad (6)$$

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.

3.4 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 [13] and Schwarz information Criterion (SIC) due to [14].

$$AIC(P) = -2\ln(L) + 2P \quad (7)$$

$$SIC(P) = -2\ln(L) + Pln(T) \quad (8)$$

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

$$\Delta HBV_t^d = \psi_0 HBV_{t-1}^d + \psi_1 \Delta HBV_{t-1}^d + \cdots + \psi_{p-1} \Delta HBV_{t-p+1}^d + u_t \quad (9)$$

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:

$$Q^{(LB)} = T(T+2)\sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{T-k}, \quad (10)$$

$$\text{where } \hat{\rho}_k^2 = \frac{T}{T-k}\left(T \sum_{t=k+1}^{T} (\hat{\varepsilon}_t^2 - \bar{\varepsilon})(\hat{\varepsilon}_{t-k}^2 - \bar{\varepsilon}) \Big/ \sum_{t=1}^{T} (\hat{\varepsilon}_t^2 - \bar{\varepsilon})^2 \right), \text{for } \bar{\varepsilon} = T^{-1} \sum_{t=1}^{T} \varepsilon_t^2$$

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 $\widehat{HBV_t}$, respectively. The reported forecast error statistics are computed as follows:

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\sum_{t=T+1}^{T+h} \left(\widehat{HBV_t} - HBV_t\right)^2 / h}$$

$$\text{Mean Absolute Error(MAE)} = \sum_{t=T+1}^{T+h} \left|\widehat{HBV_t} - HBV_t\right| / h$$

$$\text{Mean Absolute Percentage Error(MAPE)} = 100 \times \sum_{t=T+1}^{T+h} \left|\frac{\widehat{HBV_t} - HBV_t}{HBV_t}\right| / h$$

$$\text{Theil Inequality Coefficient(TIC)} = \frac{\sqrt{\sum_{t=T+1}^{T+h}\left(\widehat{HBV_t} - HBV_t\right)^2 / h}}{\sqrt{\sum_{t=T+1}^{T+h} \widehat{HBV_t}^2 / h} + \sqrt{\sum_{t=T+1}^{T+h} HBV_t^2 / h}}$$

$$\text{Bias Proportion (BP)} = \left(\left(\sum \widehat{HBV_t}/h\right) - \overline{HBV_t}\right)^2 \Big/ \sum \left(\widehat{HBV_t} - HBV_t\right)^2 / h$$

$$\text{Variance Proportion (VP)} = \left(S_{\widehat{HBV}} - S_{HBV}\right)^2 \Big/ \sum \left(\widehat{HBV_t} - HBV_t\right)^2 / h$$

$$\text{Covariance Proportion (CP)} = 2(1-r) S_{\widehat{HBV}} S_{HBV} \Big/ \sum \left(\widehat{HBV_t} - HBV_t\right)^2 / h$$

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].

3.0 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 Figure 1.

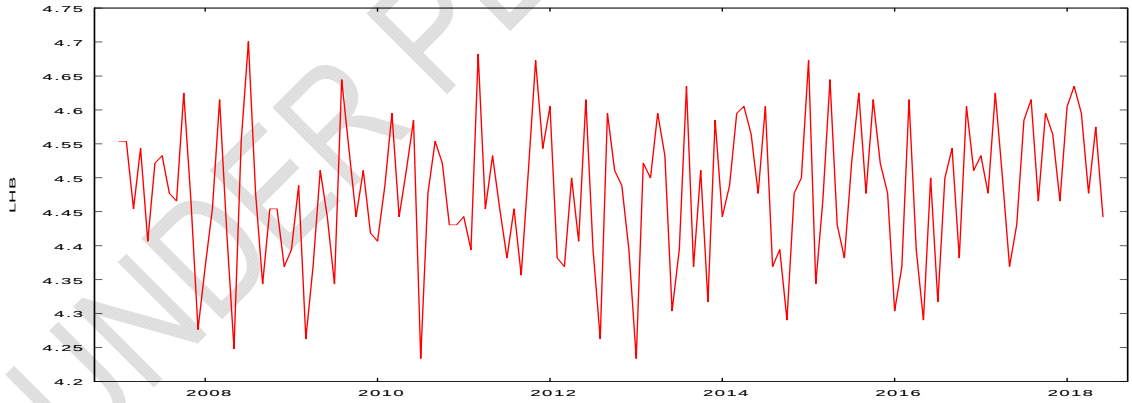

Figure 1: Time Plot of Hepatitis-B Infection in Lafia (Natural Log)

The time plot of the transformed series reported in Figure 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 Figure 2.

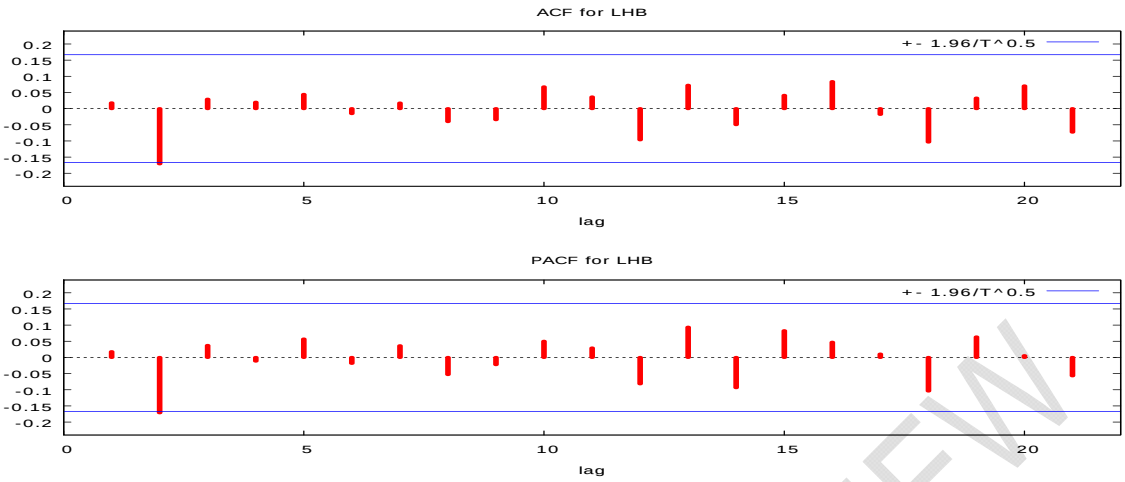

Figure 2: ACF and PACF Plot of Hepatitis-B Infection in Lafia (Natural Log)

The plots of ACF and PACF of the series reported in Figure 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.

Table 1: Autocorrelation Function and Ljung-Box Q-statistics

Lag	ACF	Std. Error	Ljung-Box Q-statistic	P-value
1	-0.0260	0.090	0.084	0.773
5	0.0374	0.089	6.270	0.281
7	-0.0433	0.088	6.711	0.460
8	-0.0343	0.087	6.865	0.551
10	0.0604	0.087	7.548	0.673
11	0.0132	0.086	7.571	0.751
12	-0.1322	0.086	9.943	0.621

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.

Table 2: Elliot-Rothenberg & Stock DF-GLS Unit Root Test Results

Option	DF-GLS Test Statistic	DF-GLS Test Critical Values 1%	5%	10%
Intercept only	-8.8306*	-2.5845	-1.9435	-1.6149
Intercept & Trend	-10.3205*	-3.5572	-3.0110	-2.7210

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 Figure 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.

Table 3: Model Order Selection Using Information Criteria and Log Likelihood

S/n	Model	AIC	BIC	LogL	R^2
1	ARMA (0,1)	-1.6222	-1.5758	99.3339	0.0012
2	ARMA (0,2)	-1.6668	-1.5971	103.0081	0.0606
3	ARMA (0,3)	-1.6692	-1.6043	104.9034	0.0871
4	ARMA (1,0)	-1.6181	-1.5714	98.2779	0.0007
5	ARMA (2,0)	-1.6479	-1.5775	100.2268	0.0497
6	ARMA (3,0)	-1.6601	-1.5794	101.9015	0.0696
7	**ARMA (1,1)****	**-1.7980**	**-1.6797**	**113.1629**	**0.6480**
8	ARMA (1,2)	-1.6509	-1.5575	102.2314	0.0649
9	ARMA (1,3)	-1.6405	-1.5718	101.8241	0.0894
10	ARMA (1,4)	-1.6307	-1.5701	103.1862	0.1290
11	ARMA (2,1)	-1.6358	-1.5419	100.5134	0.0534
12	ARMA (2,2)	-1.6815	-1.5641	104.2075	0.1169
13	ARMA (2,3)	-1.6706	-1.5297	104.5628	0.1170
14	ARMA (2,4)	-1.6519	-1.5215	103.4832	0.1392
15	ARMA (3,1)	-1.6447	-1.5276	101.2161	0.0867
16	ARMA (3,2)	-1.7852	-1.6436	110.4354	0.2199
17	ARMA (3,3)	-1.6430	-1.4778	103.1172	0.1192
18	ARMA (3,4)	-1.6516	-1.4872	104.4437	0.1152
19	ARMA (4,1)	-1.6320	-1.4896	100.6567	0.0956
20	ARMA (4,2)	-1.6491	-1.4829	102.6471	0.1262
21	ARMA (4,3)	-1.7533	-1.5634	109.6932	0.2261
22	ARMA (4,4)	-1.7348	-1.5211	109.6162	0.2251

Note: ** denotes ARMA model selected by the criteria

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 4: OLS Parameter Estimates of ARMA (1,1) Model

Variable	Coefficient	Std. Error	t-Statistic	P-value
C	4.469523	0.004667	957.5877	0.0000
AR(1)	0.724005	0.146412	4.944999	0.0000
MA(1)	0.274561	0.101923	-8.580618	0.0000
R-squared	0.647981	Akaike info criterion		-1.797973
Adjusted R^2	0.431567	Schwarz criterion		-1.679735
Log likelihood	113.1629	Hannan-Quinn criterion		-1.621347
F-statistic	7.923161	Prob(F-statistic)		0.000035
Inverted AR Roots	0 .72	Inverted MA Roots	0.87	Durbin W. 1.86016

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

$$HBV_t = 4.469523 + 0.724005HBV_{t-1} + \varepsilon_t + 0.274561\varepsilon_{t-1} \quad (11)$$

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 significant levels. The estimated model have also satisfied the stationarity condition because $\alpha_1 + \beta_1 = 0.724005 + 0.274561 = 0.998566 < 1$. This shows that the estimated ARMA (1,1) 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 Figure 3.

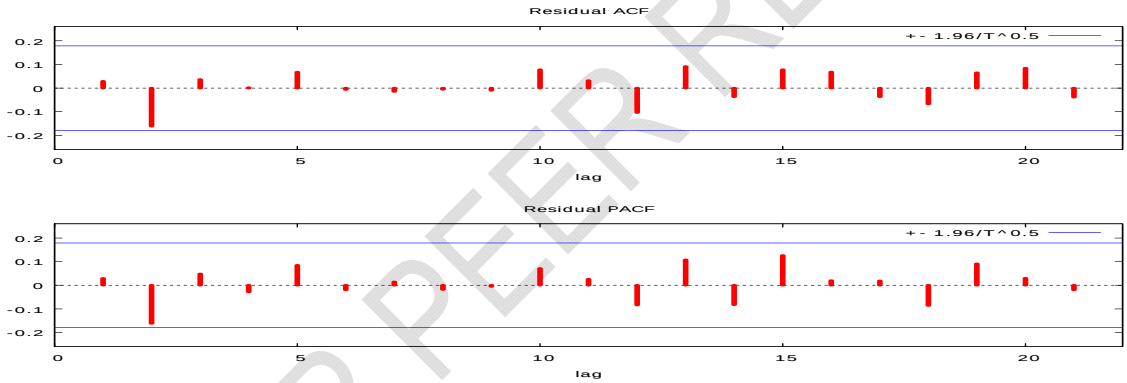

Figure 3: ACF and PACF Plot of Residual

Figure 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.

Table 5: Ljung-Box Q-statistic Test for Serial Correlation of Residuals

Lag	Q-statistics	P-value
1	0.1054	0.745
2	3.3254	0.190
3	3.4907	0.322
4	4.4911	0.479
5	4.0721	0.539
6	4.0775	0.666
7	4.1036	0.768
8	4.1080	0.847
9	4.1204	0.903
10	4.9234	0.896
11	5.0548	0.928
12	6.5083	0.888

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 AR/MA polynomials are all within a unit circle as reported in Figure 4.

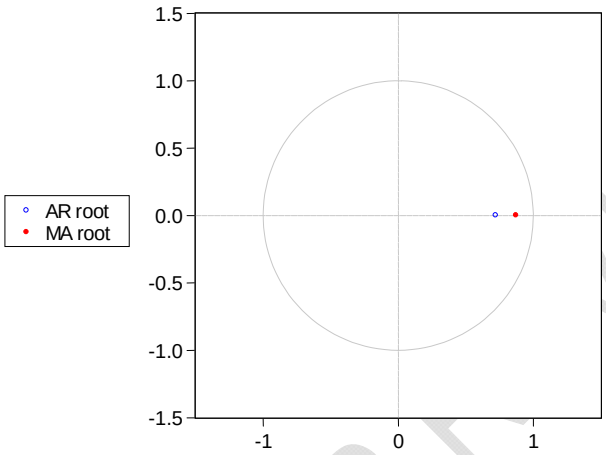

Figure 4: Inverse Roots of AR/MA Polynomials

From the root of AR and MA polynomials of the fitted model presented in Table 4, $\text{AR root} = 0.72$ and $\text{MA root} = 0.87$ and we estimate that $\tan\theta = y/x = 0.72/0.87 = 0.8276$ and $\theta = 39.61°$. Thus, the life cycle of hepatitis B virus infection among blood donor in the study area is $360°/39.61° = 9.09 \approx 9$ 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 carcimona.

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.

Table 6: Forecast Comparison Using Accuracy Measures

Forecast Mode	RMSE	MAE	MAPE	TIC	BP	VP	CP
In-Sample	0.1057	0.0855	1.9157	0.0118	0.0000	0.8899	0.1101
Out-of-Sample*	0.1034	0.0824	1.8469	0.0116	0.0000	0.6372	0.3628

Note: * denotes the forecast mode selected by accuracy measures

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).

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 use 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.

Table 7: Forecast of Hepatitis-B Infection in Lafia from July 2018-June 2019

Year: Month	Forecast (in natural log)	Actual Forecast (no of persons)	Std. Error (in log)	95% Interval (no of persons)
2018:07	4.45626	86	0.102313	[71, 106]
2018:08	4.45809	86	0.103097	[71, 106]
2018:09	4.45970	86	0.103695	[71, 106]
2018:10	4.46110	87	0.104151	[71, 106]
2018:11	4.46233	87	0.104500	[71, 106]
2018:12	4.46341	87	0.104767	[71, 106]
2019:01	4.46435	87	0.104971	[71, 107]
2019:02	4.46518	87	0.105128	[71, 107]
2019:03	4.46591	87	0.105247	[71, 107]
2019:04	4.46654	87	0.105339	[71, 107]
2019:05	4.46710	87	0.105410	[71, 107]
2019:06	4.46758	87	0.105464	[71, 107]

Note: For 95% confidence intervals, $Z_{0.025} = 1.96$

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 from the current month. The interval [71, 106] persons imply that the monthly increase may lie between 71 and 106 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 facilities provided by the state government and other NGOs and international donors in the region. This result corroborates the empirical findings of [11 & 12].

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 carcimona. 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 neighbouring 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.

REFERENCES

[1] Ciupe SM, Ribeiro RM, Nelson PW, Perelsona AS. Modeling the mechanisms of acute hepatitis B virus infection. Journal of Theoretical Biology, 2007; 247: 23–35.

[2] Kuo RJ, Cheng WC, Lien WC, Yang TJ. A medical resource forecasting model for emergency room patients with Acute Hepatitis. International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering, 2015; 9(5): 377-382.

[3] Goldstein ST, Zhou F, Hadler SC. A mathematical model to estimate global hepatitis B disease burden and vaccination impact. International Journal of Epidemiology, 2005; 34:1329-39.

[4] Carey WD. The prevalence and natural history of hepatitis B in the 21st century. Cleve Clinical Journal of Medicine, 2009; 76(3): S2-S5.

[5] Lavanchy D. Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures. Journal of Viral Hepatitis, 2004; 11:97-107.

[6] Mast EE, Alter MJ, Margolis HS. Strategies to prevent and control hepatitis B and C virus infections: A global perspective. Vaccine, 1999; 17:1730-1733.

[7] Mahoney FJ. Update on diagnosis, management, and prevention of hepatitis B virus infection. Clinical Microbiological Review, 1999; 12:351-366.

[8] Lee W. Hepatitis B virus infection. Naturtal Entomological Journal of Medicine, 1997; 337:1733-1745.

[9] Shahdoust M, Sadeghifar M, Poorolajal J, Javanrooh N, Amini P. Predicting hepatitis B monthly incidence rates using weighted markov chains and time series methods. J. Res. Health Sci., 2015; 15(1): 28-31.

[10] Gan R, Chen N, Huang D. Comparisons of forecasting for hepatitis in Guangxi Province, China by using three neural networks models. Peer J., 2016; 4: e2684.

[11] Wang YW, Shen ZZ, Jiang Y. Comparison of ARIMA and GM(1,1) models for prediction of hepatitis B in China. PLoS ONE, 2018; 13(9): e0201987.

[12] Zhang H, Huo M, Tong Y, Liu P. Time series analysis in modeling of hepatitis B incidence. Iran J. Pub. Health, 2019; 48(3): 554-555.

[13] Akaike H. A new look at statistical model identification. IEEE Transactions on Automatic Control, 1974; AC-19: 716-723.

[14] Schwarz G. Estimating the dimension of a model. Annals of Statistics, 1978; 6(2): 461-464.

[15] Elliott G, Rothenberg TJ, Stock JH. Efficient tests for an autoregressive unit root. Econometrica, 1996; 64: 813-836.

[16] Ljung GM, Box GEP. On a measure of lack of fit in time series models. Biometrica, 1978; 65: 297-303.

[17] Pindyck RS, Rubinfeld DL. Econometric models and economic forecasts, 4th edition, New York: McGraw-Hill, 1998; pp. 210-214.