

A Time Series Model for Predicting Human Immunodeficiency Virus in the Presence of Opportunistic Infections among Farmers in Benue State, Nigeria

David Adugh Kuhe^{1*} and Terwase Agbe²

¹Department of Statistics, Joseph Sarwuan Tarka University, Makurdi, Benue State, Nigeria

²Department of Mathematics and Computer Science, Benue State University Makurdi, Benue State, Nigeria

***Corresponding Author**

DOI: <https://doi.org/10.51244/IJRSI.2025.120500164>

Received: 29 April 2025; Accepted: 04 May 2025; Published: 19 May 2025

ABSTRACT

The aim of this study is to provide a short-term prediction of Human Immunodeficiency Virus (HIV) in the presence of opportunistic infections among farmers in Benue state, Nigeria using Autoregressive Integrated Moving Average with exogenous variables (ARIMAX) time series model. Monthly secondary data on HIV, Tuberculosis (TB), and Hepatitis B Virus (HBV) infections from January 2010 to December 2024 were sourced from the Benue State Epidemiological Unit, Makurdi. The study employed summary statistics, normality measures, time series plots; Ng-Perron modified unit root test, and ARIMAX model as methods of analysis. Employing the Box-Jenkins procedure, autocorrelation function (ACF), and partial autocorrelation function (PACF), a mixed ARIMAX (p,d,q) process was identified, with model selection based on log likelihoods (LogL), Akaike information criterion (AIC), Schwartz information criterion (SIC), and Hannan Quinn information criterion (HQC). The analysis revealed the series to be stationary in the first difference hence integrated of order one, I(1). The chosen ARIMAX (4,1,3) model, explaining 65.93% of data variability, forecasted HIV infections for 24 months from January 2025 to December 2026. The forecast depicted fluctuating trends in HIV infection rates, reflecting original dynamics, emphasizing the dynamic nature of HIV infection rates alongside opportunistic infections among farmers in Benue state, Nigeria. The forecast suggested 27,225 HIV total cases in the study area over 2025-2026, with an average monthly incidence of 1134 persons. A reliability test using actual and forecast values indicated no significant difference, affirming the reliability and accuracy of the forecasts for policy implementation. The study advocates collaborative efforts among the government of Benue state, international donor agencies, health policymakers, and stakeholders to implement robust preventive and control measures to mitigate future HIV incidences in the study area.

Keywords: ARIMAX Model, Human Immunodeficiency Virus, Opportunistic infections, Farmers, Benue state, Nigeria.

INTRODUCTION

The human immunodeficiency virus (HIV) remains a significant public health concern globally, particularly in sub-Saharan Africa, which bears the highest burden of the epidemic (UNAIDS, 2023). Nigeria, with an estimated 1.8 million people living with HIV as of 2022, continues to face substantial challenges in its prevention and control efforts (National Agency for the Control of AIDS [NACA], 2022). Benue State, often referred to as the “food basket” of the nation, is disproportionately affected, with one of the highest HIV prevalence rates in Nigeria, especially among rural farming communities (Federal Ministry of Health [FMoH], 2020). The socio-economic consequences of HIV in these communities are far-reaching, affecting not only health outcomes but also agricultural productivity and food security.

HIV seropositivity among farmers has been associated with decreased physical capacity, lower productivity, and labour shortages, ultimately contributing to increased postharvest losses (FAO, 2014). Postharvest losses significantly undermine food security and rural livelihoods, particularly in agrarian regions like Benue State. The interplay between HIV infection and food production is further compounded by the syndemic interaction with other communicable diseases such as tuberculosis (TB) and hepatitis B virus (HBV), both of which are prevalent among HIV-infected populations due to shared modes of transmission and immunosuppressive effects (WHO, 2023; Musa *et al.*, 2021).

Tuberculosis remains the leading opportunistic infection and a major cause of death among people living with HIV (WHO, 2023). Similarly, co-infection with hepatitis B is a growing concern, especially in sub-Saharan Africa, where both viruses are endemic and often interact to worsen patient outcomes (Thio, 2009). Among farming populations who typically lack access to comprehensive healthcare, these comorbidities further exacerbate health deterioration and reduce the capacity for effective postharvest handling, storage, and market engagement.

To effectively address this complex nexus of health and agriculture, there is a need for robust predictive models that can inform public health interventions and agricultural policy planning. Time series models, particularly the Autoregressive Integrated Moving Average with Exogenous variables (ARIMAX), offer a powerful framework for forecasting disease prevalence while accounting for influential external factors (Box *et al.*, 2015). The ARIMAX model's incorporation of exogenous predictors such as TB and HBV enables more precise forecasting of HIV seropositivity trends and supports evidence-based decision-making.

This study aims to apply and evaluate an ARIMAX model for predicting HIV seropositivity among farmers in Benue State, using TB and hepatitis B prevalence as exogenous variables. By modeling the dynamic interactions among these infectious diseases, the research seeks to uncover the broader implications for agricultural productivity, specifically in terms of food postharvest losses. The findings are expected to guide integrated health and agricultural interventions that address the dual burden of disease and food insecurity in rural Nigerian communities.

Many scholars have utilized different statistical models to study the prevalence of infectious diseases globally and locally. For example, Chen *et al.* (2022) conducted an ecological study that utilized an ARIMAX model to predict pulmonary tuberculosis incidence in Ningbo, China, considering air pollution and meteorological factors as exogenous variables. The model demonstrated high predictive accuracy, illustrating the utility of ARIMAX models in forecasting disease incidence when incorporating relevant external factors. Onovo *et al.* (2023) employed Bayesian statistical modeling to estimate HIV prevalence at the state level in Nigeria. Using national HIV testing services data from 2020 to 2021, the researchers adjusted for demographic, economic, biological, and societal covariates. Benue State was found to have the highest estimated HIV prevalence at 5.7% among adults aged 15-49 years. The study underscores the importance of reliable state-level estimates for effective HIV surveillance and intervention planning.

Musa *et al.* (2021) carried out a systematic review and meta-analysis to examine the prevalence of HBV among individuals living with HIV in Nigeria. The study found a significant co-infection rate, highlighting the need for integrated screening and management strategies for HIV and HBV. Kane *et al.* (2014) compared ARIMA and Random Forest models in predicting H5N1 avian influenza outbreaks in Egypt, finding that Random Forest models provided superior predictive accuracy. Imai *et al.* (2015) discussed the challenges and solutions in applying time series regression models to assess the relationship between infectious diseases and weather variables, using influenza and cholera as case studies. Zhou *et al.* (2023) Utilized an interrupted time series ARIMA model to analyze the impact of COVID-19 on the incidence rates of notifiable communicable diseases in China, revealing significant short-term declines in various disease categories.

Abu and Kotur (2022) carried out a research focusing on the impact of HIV/AIDS on farm productivity in Benue State. Comparing HIV-infected and non-infected farming households, the study found that infected farmers had significantly lower labour productivity and smaller labour forces. The findings highlight the detrimental effects of HIV/AIDS on agricultural productivity and the need for targeted interventions to support affected farmers.

The impact of infectious diseases on agricultural productivity and postharvest losses of crops are also documented in empirical literature, for example, FAO (2014) report discusses how health challenges, including HIV/AIDS, can affect agricultural productivity and postharvest handling, leading to increased food losses. Aliyu and Akor (2023) identified measures to reduce postharvest losses of vegetable crops in Benue State, emphasizing the need for adequate storage and handling practices, which can be compromised by health-related labour shortages. Agber and Aondofa (2023) focused on cassava farmers in Benue State, the study highlights how socio-economic challenges, potentially exacerbated by health issues, contribute to significant postharvest losses in the study area. Ikya and Igbokwe (2019) examined the determinants of postharvest losses among tomato farmers in Gboko Local Government Area of Benue State. The study identified many determinants of postharvest losses, noting that health-related labour constraints can impact timely harvesting and storage.

MATERIALS AND METHODS

Data Source

The data used in this work comprised monthly time series secondary data on Human Immunodeficiency Virus (HIV), Tuberculosis (TB), Hepatitis B virus infection (HBV) co-infection of farmers in Benue state. The data spanned from January, 2010 to December, 2024 and was obtained from Benue State Epidemiological Unit. The data was sieved through occupation to retain only farmers in the study. To reduce and stabilize the mean and variance the data was transformed to natural logarithms using the following formula:

$$Y'_t = \ln Y_t \quad (1)$$

Where Y'_t represent the HIV, TB or HBV series at time t and $\ln Y_t$ is the natural log of Y'_t .

Methods of Data Analysis

This study employs the following statistical tools for data analysis.

Descriptive statistics and normality measures

The mean of any given set of data is computed as:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (2)$$

The sample standard deviation is computed as:

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

where $\bar{y}$ is the sample mean, n is the sample size.

The Jarque-Bera test normality test statistic (JB) proposed by Jarque and Bera (1980, 1987) computed from the following formula:

$$JB = \frac{T}{6} \left(g_1^2 + \frac{1}{4} (g_2 - 3)^2 \right) \quad (4)$$

where g_1 is the sample skewness computed as:

$$g_1 = \frac{\mu_3}{\mu_2^{3/2}} = \left[T^{1/2} \sum_{t=1}^T (y_t - \bar{y})^3 \right] \left[\sum_{t=1}^T (y_t - \bar{y})^2 \right]^{-3/2} \quad (5)$$

and g_2 is the sample kurtosis calculated from:

$$g_2 = \frac{\mu_4}{\mu_2^2} = \left[T \sum_{t=1}^T (y_t - \bar{y})^4 \right] \left[\sum_{t=1}^T (y_t - \bar{y})^2 \right]^{-2} \quad (6)$$

where T is the total number of observations. The JB normality test checks the following pair of hypothesis:

$H_0: \hat{\mu}_3 = 0$ and $\hat{\mu}_4 = 0$ (i.e., y_t follows a normal distribution)

$H_1: \hat{\mu}_3 \neq 0$ and $\hat{\mu}_4 \neq 0$ (i.e., y_t does not follows a normal distribution).

The test rejects the null hypothesis if the p-value of the JB test statistic is less than $\alpha = 0.05$ level of significance.

Autocovariance and autocorrelation functions

For a stationary time series process $\{y_t\}$, the covariance between y_t and y_{t+k} is given as

$$\gamma_k = Cov(y_t, y_{t+k}) = E(y_t - \mu)(y_{t+k} - \mu) \quad (7)$$

and the correlation between y_t and y_{t+k} is given as

$$\rho_k = \frac{Cov(y_t, y_{t+k})}{\sqrt{var(y_t)}\sqrt{var(y_{t+k})}} = \frac{\gamma_k}{\gamma_0} \quad (8)$$

where γ_k is the autocovariance function and ρ_k is the autocorrelation function (ACF) representing the covariance and correlation between y_t and y_{t+k} from the same process separated only by k time lags. For a given observed time series $y_1, y_2, \dots, y_n$, the sample autocorrelation function (ACF) is computed as

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} = \frac{\sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}, \quad k = 0, 1, 2, \dots \quad (9)$$

where $\bar{y}$ is defined in Equation (2).

The partial autocorrelation between y_t and y_{t+k} is equal to the ordinary autocorrelation between $(y_t - \hat{y}_t)$ and $(y_{t+k} - \hat{y}_{t+k})$. Let ϕ_{kk} denotes the partial autocorrelation between y_t and y_{t+k} , then we have

$$\phi_{kk} = \frac{Cov[(y_t - \hat{y}_t)(y_{t+k} - \hat{y}_{t+k})]}{\sqrt{var(y_t - \hat{y}_t)}\sqrt{var(y_{t+k} - \hat{y}_{t+k})}} \quad (10)$$

A recursive procedure for computing the sample partial autocorrelation function, ϕ_{kk} (PACF) starting with $\phi_{11} = \hat{\rho}_1$ was given by Durbin (1960) as

$$\hat{\phi}_{k+1,k+1} = \hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_{k+1-j} / \left(1 - \sum_{j=1}^k \hat{\phi}_{kj} \hat{\rho}_j \right) \quad (11)$$

and $\hat{\phi}_{k+1,j} = \hat{\phi}_{kj} - \hat{\phi}_{k+1,k+1}\hat{\phi}_{k,k+1-j}$, $j = 1, 2, \dots, k$. This procedure also holds for computing the theoretical PACF ϕ_{kk} .

Ljung-Box Q-statistic test

A Ljung-Box Q-statistic test is a test used to investigate the presence of serial correlation or autocorrelation in the residuals of a series. The test checks the following pairs of hypotheses:

$H_0: \rho_{k,1} = \rho_{k,2} = \dots = \rho_{k,T} = 0$ (all lags correlations are zero)

$H_1: \rho_{k,1} \neq \rho_{k,2} \neq \dots \neq \rho_{k,T} \neq 0$ (there is at least one lag with non-zero correlation). The test statistic is given by:

$$Q^{(LB)} = T(T+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{T-k} \quad (12)$$

where

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

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 (Ljung and Box, 1978).

Ng and Perron (NP) modified unit root test

To check the unit root and stationarity properties of the series, Ng and Perron modified unit root test is employed because of its good power property. Ng and Perron (2001) constructed four test statistics which are based on the Generalized Least Squares detrended series Y_t^d . The four test statistics are the modified forms of Phillips & Perron Z_α and Z_t statistics, the Bhargava (1986) R_1 statistic, and the Elliot, Rothenberg & Stock Point Optimal statistic (Elliot *et al.*, 1996). First, define the term:

$$k = \sum_{t=2}^T (Y_{t-1}^d)^2 / T^2 \quad (13)$$

The four modified statistics are then written as,

$$MP_T^d = \left\{ \begin{array}{l} MZ_\alpha^d = (T^{-1}(Y_T^d)^2 - f_0)/(2k) \\ MZ_t^d = MZ_\alpha \times MSB \\ MSB^d = (k/f_0)^{0.5} \\ (-7^2k + 7T^{-1}(Y_T^d)^2)/f_0, \quad \text{if } x_t = \{1\} \\ (-13.5^2k + (1 + 13.5)T^{-1}(Y_T^d)^2)/f_0, \text{ if } x_t = \{1, t\} \end{array} \right\} \quad (14)$$

where MZ_α^d is the modified detrended Z_α transformation of the standardized estimator given by:

$$T(\hat{\alpha} - 1) = \left\{ T^{-1} \sum_{t=1}^T y_{t-1}(y_t - y_{t-1}) \right\} / \left\{ T^{-1} \sum_{t=1}^T y_{t-1}^2 \right\} \quad (14)$$

MZ_t^d is the modified detrended Z_t transformation of the conventional regression t statistic defined by:

$$t_{\alpha} = \left(\sum_{t=1}^T y_{t-1}^2 \right)^{1/2} (\hat{\alpha} - 1) / s \quad (15)$$

$$\text{where} \quad s^2 = T^{-1} \sum_{t=1}^T (y_t - \hat{\alpha} y_{t-1})^2 \quad (16)$$

MSB is the modified Bhargava R_1 statistic (Stock, 1990). The R_1 statistic is given by:

$$R_1 = \sum_{t=2}^T (y_t - y_{t-1})^2 / \sum_{t=1}^T (y_t - \bar{y})^2; \quad \bar{y} = \frac{1}{T} \sum_{t=1}^T y_t \quad (17)$$

MP_T^d is the ERS modified detrended point optimal statistic (Elliot *et al.*, 1996). The point optimal statistic is given as:

$$P_T = \frac{[S(\hat{\alpha}) - \hat{\alpha}S(1)]}{s^2} \quad (18)$$

Y_T^d is the trended series, x_t is a series of observations at time t , f_0 is the frequency zero spectrum define as:

$$f_0 = \sum_{j=-(T-1)}^{T-1} \hat{\gamma}(j) \cdot K\left(\frac{j}{l}\right) \quad (19)$$

Where l is a bandwidth parameter, T is the sample size, K is a kernel function and $\hat{\gamma}(j)$ is the j -th sample autocovariance of the residuals $\hat{u}_t$ and is given by:

$$\hat{\gamma}(j) = \sum_{t=j+1}^T (\hat{u}_t \hat{u}_{t-j}) / T \quad (20)$$

The MZ_{α} , MZ_t , MSB and MP_T statistics are collectively referred to as M tests and are used in detecting the presence of unit root in a series (Ng & Perron, 2001). In addition to the MZ_{α} and MZ_t statistics, Ng and Perron also investigated the size and power properties of the MSB statistic. Critical values for the demeaned and detrended case of this statistic were taken from (Stock, 1990).

Model Specification

Before we specify an autoregressive integrated moving average with exogenous variable (ARIMAX) model, we first specify an autoregressive (AR) process, a moving average (MA) process, autoregressive moving average (ARMA) process and autoregressive integrated moving average (ARIMA) process which are specified in the following subsections.

Autoregressive (AR) model

An autoregressive process (AR) is a process that has a significant relationship with its history observations (previous time lags). The general characteristic of autoregressive model of order p is given as

$$y_t = \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (21)$$

where y_t is the response variable at time t , α is a constant, $\phi_1, \phi_2, \dots, \phi_p$ are autoregressive parameters to be determined and ε_t is a white noise process with mean zero and variance σ^2 .

Moving Average (MA) model

A moving average process (MA) is a process that has a significant relationship with its previous random errors. The general characteristic of a moving average model of order q is given as

$$y_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (22)$$

where θ_i are the moving average parameters. The subscript on θ_i are called the orders of moving average parameters.

Autoregressive Moving average (ARMA) 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 p, q , ARMA (p, q) is specified as

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (23)$$

Equation (23) can also be written as

$$\left(1 - \sum_{i=1}^p \phi_i B^i\right) y_t = \left(1 + \sum_{j=1}^q \theta_j B^j\right) \varepsilon_t \quad (24)$$

where B is the back shift operator, the ϕ_i are the parameters of the autoregressive part of the model, the θ_j are the parameters of the moving average part and ε_t are error terms.

Autoregressive integrated moving average (ARIMA) model

Assuming that the polynomial $(1 - \sum_{i=1}^p \phi_i B^i)$ has a unitary root of multiplicity d , then it can be written as:

$$\left(1 - \sum_{i=1}^p \phi_i B^i\right) = \left(1 + \sum_{j=1}^{p-d} \theta_j B^j\right) (1 - B)^d \quad (25)$$

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

$$\left(1 - \sum_{i=1}^p \phi_i B^i\right) (1 - B)^d y_t = \left(1 + \sum_{j=1}^q \theta_j B^j\right) \varepsilon_t \quad (26)$$

An ARIMA (p, d, q) model can also be written as

$$\Phi(B) \Delta^d y_t = \Theta(B) \varepsilon_t \quad (27)$$

Where B is the back shift operator, y_t is the dependent variable in period t ,

$\Delta^d = (1 - B)^d$ denotes the difference from d degree, ε_t is the white noise process.

$\Phi(B)$ is the autoregressive polynomial of order p expressed as

$$\Phi(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - \dots - \phi_p B^p \quad (28)$$

$\Theta(B)$ is a moving average polynomial of order q given as

$$\Theta(B) = 1 - \theta_1 B^1 - \theta_2 B^2 - \dots - \theta_q B^q \quad (29)$$

ARIMAX Model

The Autoregressive Integrated Moving Average process with exogenous variable(s) (ARIMAX) model can be viewed as a time-series forecasting model using the multiple regression with ARIMA model that takes care of the residual's serial correlations. The ARIMAX model has a form as follows:

$$y_t = \beta x_t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (30)$$

An ARIMAX model can also be written using a back shift operator as follows:

$$\Phi(B)\Delta^d y_t = \mu + \varphi(B)x_t + \theta(B)\varepsilon_t \quad (31)$$

where

$$\varphi(B) = 1 - \vartheta_1 B^1 - \vartheta_2 B^2 - \dots - \vartheta_r B^r \quad (32)$$

x_t is the exogenous variable at time t .

According to Andrews *et al.* (2013) the six assumptions for building an adequate ARIMAX time series are summarized:

- (i) Residuals of the estimated ARIMAX model must be stationary;
- (ii) The residual series must not exhibit significant serial correlation/autocorrelation;
- (iii) The estimated coefficient for an exogenous variable must be statistically significant;
- (iv) An exogenous variable must not display evidence of receiving feedback from the dependent variable;
- (v) Both the dependent variable and exogenous variable must have the same level of transformation and stationarity;
- (vi) The surviving exogenous variables comprising the final model must not exhibit a significant degree of multicollinearity.

Model Identification and Selection

To identify a number of parameters needed for AR and MA processes. The idea is to match the empirical autocorrelation patterns with the theoretical ones. An autocorrelation function (ACF) plot and a partial autocorrelation function (PACF) plot are primary tools used for identifying a number of parameters for AR and MA processes respectively.

The standard method for parameter estimation is maximum likelihood estimation. The best model is selected base on the following information criteria: Akaike information criterion (AIC) due to Akaike (1974), Schwarz information Criterion (SIC) due to Schwarz (1978) and Hannan-Quinn Information Criterion (HQC) due to Hannan (1980) in conjunction with the log likelihood (LogL). Intuitively, AIC, SIC and HQC tell which model fits better to the data in term of the information loss. The information criteria are presented as:

$$AIC(P) = -2 \ln(L) + 2P \quad (33)$$

$$SIC(P) = -2 \ln(L) + P \ln(T) \quad (34)$$

$$HQC(K) = 2 \ln[\ln(T)] K - 2 \ln(L) \quad (35)$$

where P is the number of free parameters to be estimated in the model, T is the number of observations and L is the likelihood function given by:

$$L = \prod_{i=0}^n \left(\frac{1}{2\pi\sigma_i^2} \right)^{1/2} \exp \left[- \sum_{i=1}^n \frac{(y_i - \mu)^2}{2\sigma_i^2} \right] \quad (36)$$

$$\ln(L) = \ln \left[\prod_{i=1}^n \left(\frac{1}{2\pi\sigma_i^2} \right)^{1/2} \right] - \frac{1}{2} \sum_{i=1}^n \frac{(y_i - \mu)^2}{\sigma_i^2} \quad (37)$$

Given a set of time series data, the ARIMAX model with the least information criteria and large log likelihood value is the best fitting model.

Model Forecast Evaluation

We employed Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) accuracy measures to select an optimal model mode that is both parsimonious and accurately forecast the data based on minimum values of the accuracy measures.

Root mean square error (RMSE)

The Root Mean Square Error is a statistical tool for measuring the accuracy of a forecast method. It is computed as:

$$RMSE = \frac{1}{n} \sum_{t=1}^n (\hat{Y}_t - Y_t)^2 \quad (38)$$

Where $\hat{Y}_t$ is the forecast value of the series and Y_t is the actual series and n is the number of forecast observations.

Mean absolute error (MAE)

The mean absolute error (MAE) is a statistical tool for measuring the average size of the errors in a collection of predictions, without taking their directions into account. It is measured as the average absolute difference between the predicted values and the actual values and is used to assess the effectiveness of a model. It is given as:

$$MAE = \frac{1}{n} \left(\sum_{t=1}^n |\hat{Y}_t - Y_t| \right) \quad (39)$$

where Y_t is the actual value of the series at time t , $\hat{Y}_t$ is the forecasted value of the series and n is the number of observations. The lower the value of RMSE and MAE, the better the model ability to forecast future values (Pindyck & Rubinfeld, 1998).

RESULTS AND DISCUSSION

Descriptive Statistics and Normality Measures

The descriptive statistics and normality measures are computed and reported in Table 1.

The results depicted in Table 1 concerning descriptive statistics reveal that the monthly averages of HIV, TB, and HBV are all positive, indicating an upward trend in the number of cases of these diseases during the analyzed period. Moreover, the corresponding standard deviations are notably high, suggesting significant dispersion from the monthly average infection cases throughout the investigation period. The considerable disparities between the maximum and minimum values of infection cases provide further evidence of the diseases' substantial variability in the study area over the examined duration.

Regarding skewness coefficients, HIV and TB exhibit positive values, indicating that there are more extreme values on the right side of their distributions among the farming population in Benue state. Conversely, the skewness coefficient of HBV is negative, indicating more extreme values on the left side of its distribution within the study area.

Kurtosis, a measure of distribution tail thickness typically around 3 for a normal distribution, varies from this standard for all variables in the study. The skewness and kurtosis coefficients for HIV suggest a distribution

pattern that aligns with normal distribution characteristics. The Jarque-Bera test results for the HIV series do not refute the null hypothesis of normality at a 5% significance level, as the p-value ($p=0.295494$) exceeds 0.05. However, for TB and HBV, the skewness and kurtosis coefficients indicate non-normal distributions. The Jarque-Bera tests for these variables reject the null hypotheses of normality at a 5% significance level, as the p-values ($p<0.05$) fall below the threshold.

Table 1: Descriptive Statistics and Normality Measures

Variable	HIV	TB	HBV
Mean	922.9423	217.6026	208.6218
Maximum	2049.000	794.0000	398.0000
Minimum	76.00000	11.00000	17.00000
Standard Deviation	382.5985	183.0163	104.3128
Skewness	0.099296	0.841162	-0.296118
Kurtosis	3.579371	3.387442	2.076886
Jarque-Bera Statistic	2.438213	19.37211	7.818740
P-value	0.295494	0.000062	0.020053
No. of Observations	156	156	156

Graphical Examination of the Series

The initial stage in examining time series and econometric data involves plotting the original series against time to observe its graphical characteristics. This aids in comprehending both the trend and the pattern of movement within the original series. Both the original series and the differenced series are graphed over time. Figures 1 and 2 display the time plots of the original and first differenced series, respectively.

The time plots depicted in Figure 1, representing the natural log-transformed of HIV, TB, and HBV series, reveal trends characterized by irregularities, indicating that the series lack constant means (i.e., they are not mean-reverting). Moreover, the variability within the series seems non-uniform, suggesting the potential presence of changing variances over time (heteroskedasticity). These observations imply that the series are non-stationary in level. To attain stationarity, the log-transformed series have undergone first-order differencing, and the resulting time plots are showcased in Figure 2.

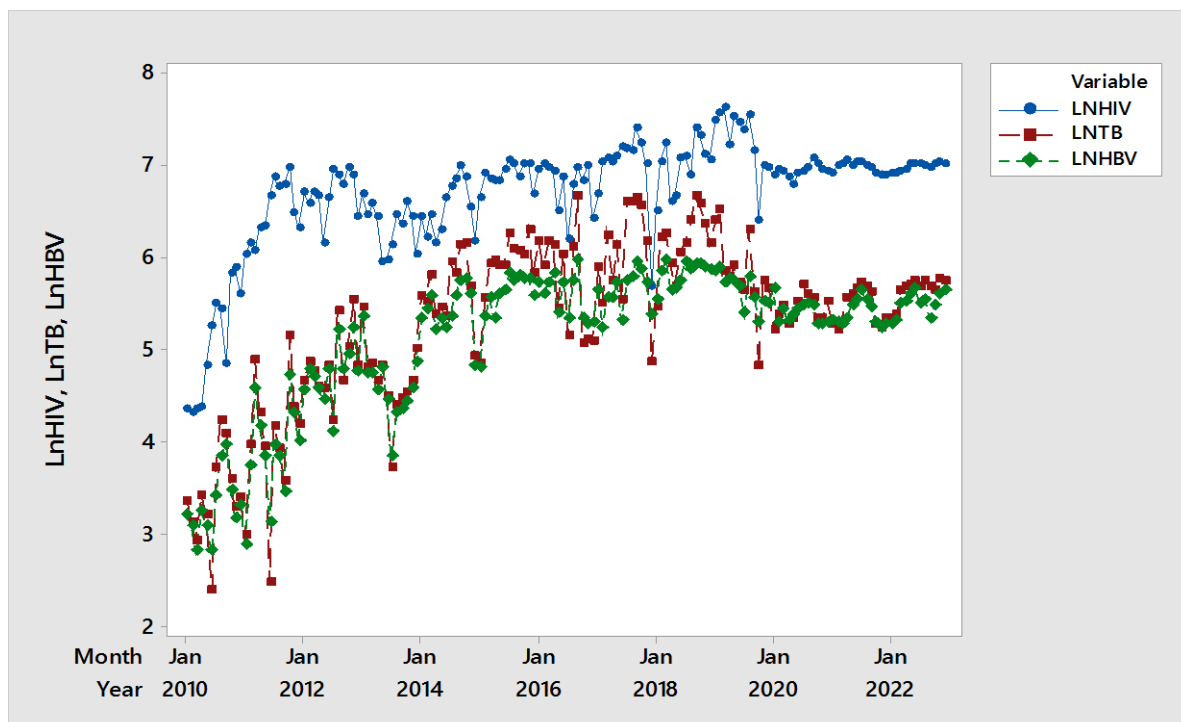


Figure 1: Time Plot of Monthly HIV, TB and HBV Infection Cases in Benue State from 2010 to 2024 (Log Transformed Series).

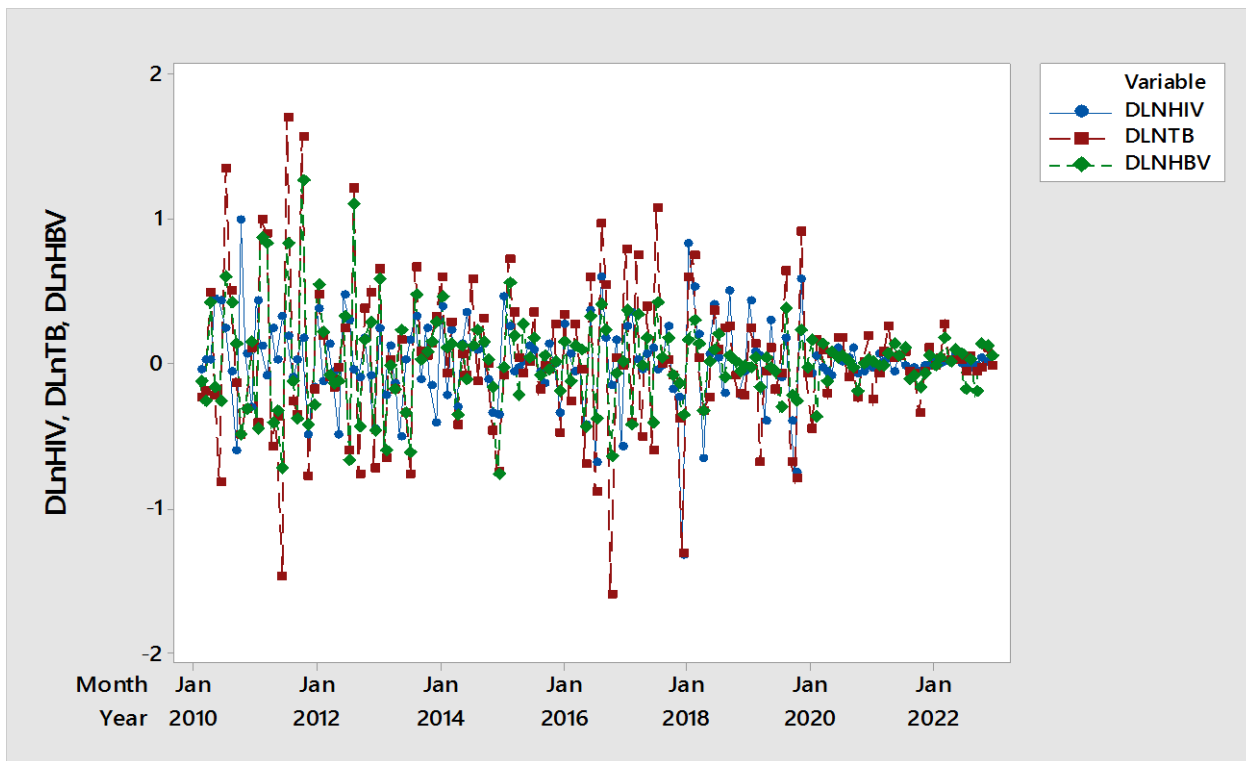


Figure 2: Time Plot of Monthly HIV, TB and HBV Infection Cases in Benue State from 2010 to 2024 (First Differenced Series).

Based on the findings depicted in Figure 2 for the first differenced series, it appears that the trends within the series exhibit a relatively smoother pattern, indicating the presence of constant means (i.e., mean reversion). Furthermore, the variability within the series appears to be consistent, suggesting that the variances may not be changing over time (homoskedasticity). These observations suggest that the first differenced series is covariance stationary. Consequently, it is inferred that the HIV, TB, and HBV infection data in Benue state are non-stationary in levels but become stationary after undergoing the first difference. Thus, the series are all integrated of order one, denoted as $I(1)$.

Ng-Perron Modified Unit Root Test Result

In order to delve deeper into the properties of unit root and stationarity, as well as confirm the order of integration of the variables under investigation, the Ng-Perron modified unit root test has been employed. The outcome of this analysis is presented in Table 2.

The results of the Ng-Perron modified unit root test, as shown in Table 2, indicate that HIV, TB, and HBV are all non-stationary in their original levels (i.e., they exhibit unit root characteristics). This non-stationarity is corroborated by the Ng-Perron M-statistics, which surpass their respective asymptotic critical values at the 5% significance level. However, when examining the first differenced series, the Ng-Perron modified unit root test suggests evidence of weak or covariance stationarity. This is evident as all four Ng-Perron M-statistics are lower than their corresponding asymptotic critical values at the 5% significance level, both for models with intercept only and for those with intercept and linear trend. Consequently, the HIV, TB, and HBV series are deemed to be integrated of order one, denoted as $I(1)$.

Table 2: Ng-Perron Modified Unit Root Test Results

Variable	Option	MZ_{α}	MZ_t	MSB	MPT
ln HIV	Intercept only	0.0018	0.0159	0.9025	47.0626
	Intercept & trend	-2.7505	-1.0767	0.3915	30.1895
$\nabla \ln HIV$	Intercept only	-119.38	-7.7259	0.0647	0.2053
	Intercept & trend	-289.65	-12.0343	0.0416	0.3147

ln TB	Intercept only	-1.6857	-0.7839	0.4650	12.4801
	Intercept & trend	-11.0467	-2.3064	0.2088	8.4765
∇ln TB	Intercept only	-78.8135	-6.2759	0.0796	0.3143
	Intercept & trend	-99.0403	-7.0366	0.0711	0.9219
ln HBV	Intercept only	0.0681	0.0590	0.8666	44.4673
	Intercept & trend	-3.5792	-1.2568	0.3511	24.1625
∇ln HBV	Intercept only	-78.0305	-6.2425	0.0800	0.3219
	Intercept & trend	-91.9921	-6.7820	0.0737	0.9906
5% Critical Values					
Intercept only		-8.1000	-1.9800	0.2330	3.1700
Intercept & trend		-17.3000	-2.9100	0.1680	5.4800

Model Identification Result

Having established the correct order of integration for the series, the subsequent step involves identifying an appropriate process to model the stationary series. Following the Box-Jenkins procedure for model identification, we examine the autocorrelogram (ACF) and partial autocorrelogram (PACF) of the stationary series, as depicted in Figure 3. The ACF and PACF plot results in Figure 3 suggest a mixed ARIMAX process, as both the ACF and PACF exhibit rapid decay towards zero. Consequently, we proceed to search for an ARIMAX (p, d, q) model that can effectively model and forecast HIV infection while accounting for TB and HBV co-infections among farmers in Benue state.

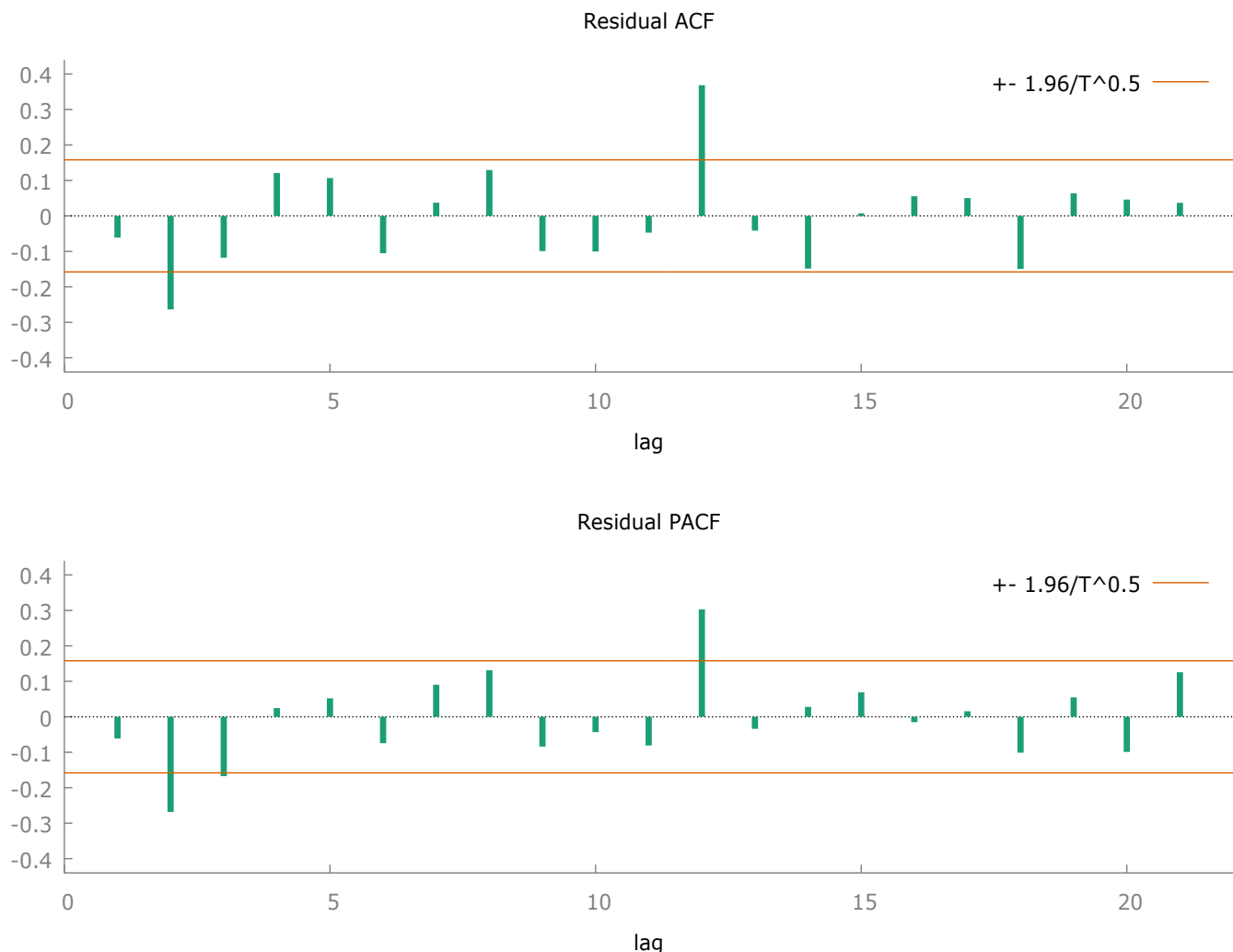


Figure 3: Autocorrelogram of Stationary Series

Model Order Selection Results

In order to find the most suitable time series model for accurately modeling and forecasting HIV infection amidst the presence of TB and HBV among the farming population in Benue state, Nigeria, we utilize the Akaike Information Criterion (AIC), Schwarz Information Criterion (SIC), and Hannan-Quinn Criterion (HQC), along with the log likelihood (LogL), to identify the optimal model. The model that exhibits the lowest information criteria values and the highest log likelihood is considered the best fitting model, which is expected to provide the most accurate fit and forecast. The outcome of the model search is presented in Table 3.

Table 3: Model Order Selection

S/n	Model	LogL	AIC	SIC	HQC
1	ARIMAX (0,1,1)	-1.9786	0.0771	0.1557	0.1090
2	ARIMAX (1,1,0)	-5.9041	0.1286	0.2075	0.1607
3	ARIMAX (1,1,1)	-0.5837	0.0725	0.1711	0.1126
4	ARIMAX (0,1,2)	0.0119	0.0644	0.1625	0.1042
5	ARIMAX (2,1,0)	-2.2545	0.0948	0.1939	0.1351
6	ARIMAX (1,1,2)	-0.3399	0.0823	0.2007	0.1304
7	ARIMAX (2,1,1)	7.3374	-0.0175	0.1014	0.0308
8	ARIMAX (2,1,2)	12.8764	-0.0768	0.0618	-0.0205
9	ARIMAX (2,1,3)	11.9024	-0.0510	0.1074	0.0134
10	ARIMAX (3,1,2)	8.0462	-0.0006	0.1585	0.0640
11	ARIMAX (1,1,3)	13.2753	-0.0815	0.0565	-0.0254
12	ARIMAX (3,1,1)	0.2124	0.0893	0.2286	0.1459
13	ARIMAX (2,1,3)	11.9026	-0.0510	0.1074	0.0134
14	ARIMAX (3,1,3)	11.3159	-0.0305	0.1486	0.0423
15	ARIMAX (1,1,4)	13.2010	-0.0675	0.0902	-0.0035
16	ARIMAX (4,1,1)	16.3177	-0.1102	0.0497	-0.0452
17	ARIMAX (2,1,4)	13.4422	-0.0581	0.1202	0.0143
18	ARIMAX (3,1,4)	11.4126	-0.0407	0.1582	0.1146
19	ARIMAX (4,1,2)	16.0834	-0.0938	0.0860	-0.0208
20	ARIMAX (4,1,3)**	28.7748	-0.2162	0.0184	-0.2350
21	ARIMAX (4,1,4)	26.4467	-0.2046	0.0152	-0.1153
22	ARIMAX (1,1,5)	19.1156	-0.1961	0.0261	-0.1264
23	ARIMAX (2,1,5)	17.7123	-0.1876	0.0257	-0.1183
24	ARIMAX (3,1,5)	15.2241	-0.1671	0.0365	-0.1175
25	ARIMAX (4,1,5)	15.3428	-0.2031	0.0266	-0.1194

Note: **denotes the model selected by the criteria

Based on the findings from Table 3 regarding model order selection, the ARIMAX (4, 1, 3) model appears to offer a statistically satisfactory representation of the provided data. This conclusion is supported by its highest log likelihood value, along with the smallest AIC, SIC, and HQC values among the options. Consequently, we designate the ARIMAX (4, 1, 3) model as the optimal and most suitable candidate for modeling and forecasting HIV infection in the presence of TB and HBV co-infection within the study area.

Model Estimation Result

Once the optimal model has been selected, the subsequent step involves estimating the parameters of the model. The outcomes of the parameter estimates for the optimal ARIMAX (4, 1, 3) model are displayed in Table 4.

From the result of the parameter estimates of Table 4, the data fits an ARIMAX (4, 1, 3) model which is presented below:

$$Y_t = 0.570007X_1 + 0.560817X_2 + 0.004937Y_{t-1} - 0.454902Y_{t-2} + 0.270107Y_{t-3} + 0.216943Y_{t-4} + \varepsilon_t - 1.075998\varepsilon_{t-1} + 0.566001\varepsilon_{t-2} - 0.489173\varepsilon_{t-3} \quad (39)$$

where Y_t = HIV infection response (dependent) variable at time t , X_1 = represents the first difference of the natural log of TB series (DLNTB) used as the first exogenous variable in the model, X_2 = represents the first difference of the natural log of HBV series (DLNHBV) used as the second exogenous variable in the model, $Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}$ = HIV infection response variables at time $t-1, t-2, t-3, t-4$ respectively, ε_t = Error term at time t and $\varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}$ = Error terms in the previous time periods which are incorporated in the response variable Y_t .

Table 4: Parameter Estimate of ARIMAX (4,1,3) Model

Variable	Coefficient	Std. Error	t-Statistic	P-value
DLNTB	0.570007	0.090725	6.282792	0.0000
DLNHBV	0.560817	0.139913	4.008315	0.0001
AR(1)	0.004937	0.204881	2.806095	0.0057
AR(2)	-0.454902	0.196422	-2.315945	0.0220
AR(3)	0.270107	0.119599	2.258440	0.0255
AR(4)	0.216943	0.084827	2.557481	0.0116
MA(1)	-1.075998	0.207759	-5.179075	0.0000
MA(2)	0.566001	0.278471	2.032535	0.0440
MA(3)	-0.489173	0.160807	-3.041998	0.0028
R-squared	0.659262	AIC	-0.216223	
Adjusted R ²	0.524747	SIC	0.018359	
Log likelihood	28.77481	HQC	-0.235045	
F-statistic	13.30609	Durbin-Watson stat	2.042812	
Prob(F-statistic)	0.000000			

The outcomes of the estimated ARIMAX (4, 1, 3) model, as presented in Table 4 and equation (39), reveal several key findings. Firstly, the AR and MA slope coefficients, along with the exogenous variables (DLNTB and DLNHV), are all statistically significant at a 5% significance level.

The coefficient of determination (R^2) for the regression model stands at 0.659262, indicating that approximately 65.93% of the total variations in HIV infection among the farming population in Benue state can be explained by the independent variables. The remaining 34.07% of unexplained variations are attributed to the error term or factors not accounted for in the model. Furthermore, the F-statistic, serving as a measure of the overall fitness of the regression parameters, yields a value of 13.30609 with a p-value of 0.00000, indicating a good fit for the regression model. Lastly, the Durbin-Watson statistic is calculated as 2.042812, suggesting the absence of positive serial correlation in the residuals of the estimated model and indicating that the model is not spurious.

ARIMAX (4, 1, 3) model validation and diagnostic checks

After fitting the model, we conduct various tests to assess its adequacy. These include the Ljung-Box Q-statistic test for serial correlation (autocorrelation), the Breusch-Godfrey serial correlation LM test, and a heteroskedasticity test for ARCH effect of the residuals from the fitted model. The outcomes of these tests are presented in Tables 5 and 6.

Furthermore, we evaluate the goodness of fit by examining the autocorrelation and partial autocorrelation plots of the residuals from the fitted model. If the majority of sample autocorrelation coefficients of the residuals fall within the limits of $\pm 1.96/\sqrt{T}$, where T represents the number of observations used to build the model, it indicates that the residuals resemble white noise, suggesting a good fit. Additionally, we scrutinize a plot displaying the residuals alongside the actual and fitted values. The ACF and PACF plot is depicted in Figure 4.

Based on the results from Tables 5 and 6, we fail to reject the null hypotheses of no serial correlation and no ARCH effect in the residuals of the fitted ARIMAX (4, 1, 3) model across all lags, as the p-values of the Ljung-Box Q-statistic test, Breusch-Godfrey serial correlation LM test, and heteroskedasticity test for ARCH effect are not statistically significant (i.e., all greater than 0.05). This indicates that the estimated model is both stationary and dynamically stable. Consequently, we conclude that the model is adequate, valid, and effective for forecasting purposes.

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

Lag	ACF	PACF	Q-Stat	P-value
1	-0.061	-0.061	0.5703	0.450
2	0.001	-0.003	0.5703	0.752
3	0.022	0.022	0.6475	0.885
4	0.041	0.043	0.9057	0.924
5	0.046	0.052	1.2425	0.941
6	0.041	0.047	1.5075	0.959
7	0.061	0.066	2.1122	0.953
8	-0.002	0.003	2.1129	0.977
9	0.224	0.221	10.241	0.331
10	-0.072	-0.053	11.084	0.351
15	0.075	0.067	15.274	0.432
20	0.053	0.017	16.666	0.675
25	-0.019	-0.012	18.178	0.835
30	0.022	0.035	19.610	0.926
35	-0.053	-0.020	22.592	0.948
36	0.058	0.069	23.267	0.950

Table 6: Test for serial Correlation and ARCH Effect

Variable	F-statistic	P-value	nR ²	P-value
Breusch-Godfrey Serial Correlation LM Test	1.678296	0.1905	3.464030	0.1769
Heteroskedasticity Test: ARCH Effect	0.552867	0.4587	0.557247	0.4554

The findings from Figure 4 indicate that nearly all sample autocorrelation coefficients of the residuals fall within the confidence bounds, suggesting that the residuals exhibit characteristics akin to white noise. This implies that the fitted model is both dynamically stable and stationary. A model deemed adequate, valid, and effective should possess the capability to forecast future values of the relevant series. In the subsequent subsection, we will assess the model's ability to forecast future values.

Forecast evaluation results

Now that our model has been validated, our focus shifts to selecting the most suitable forecast mode for predicting future relevant series. In this regard, we evaluate both in-sample and out-of-sample forecasts using two accuracy measures. The forecast mode with the lowest accuracy measures is considered the most effective for predicting HIV infection in the presence of opportunistic infections among the farming population in Benue state, Nigeria. The outcomes of the forecast comparison are outlined in Table 7.

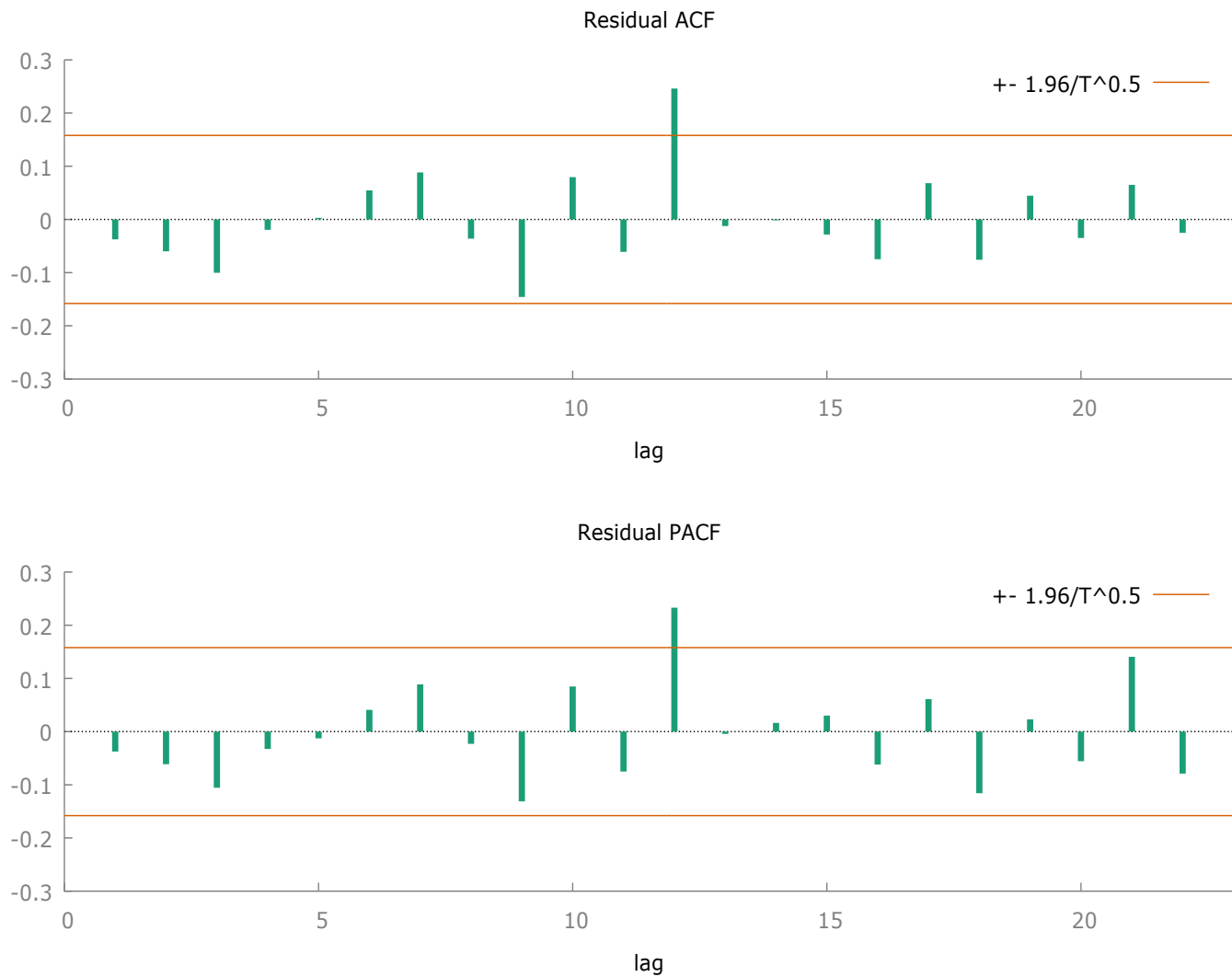


Figure 4: ACF and PACF of Residuals of the Estimated ARIMAX (4, 1, 3) Model

Table 7: Forecast Comparison using Accuracy Measures

	RMSE	MAE
In-Sample	0.326152	0.273158
Out-of-Sample**	0.213680	0.155000

Note: ** denotes forecast mode selected by accuracy measures.

Analyzing Table 7, we utilize two benchmarks, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), to compare the performance of in-sample and out-of-sample forecasts generated by the estimated ARIMAX (4, 1, 3) model. This assessment aims to evaluate the model's forecasting capability and determine which mode of forecast is superior. Notably, we observe that the RMSE and MAE values for the out-of-sample forecast are lower than those for the in-sample forecast. Based on this criterion, a smaller forecast error indicates better forecasting ability for the model. Consequently, our analysis suggests that the model is well-suited for future forecasts.

Short-Term Forecast of HIV in the presence of opportunistic infections in Benue State

Opting for the out-of-sample forecast method for the series, we employ the estimated ARIMAX (4, 1, 3) model to predict future values of HIV infection in the study area over a span of 2 years (24 months), spanning from January 2025 to December 2026. The forecast outcomes are detailed in Table 8.

Table 8: Forecast of HIV in the Presence of Opportunistic Infections in Benue State from 2025-2026

Year: Month	Forecast (natural log form)		Actual Forecast (No. of Persons)		
	Forecast	Std. error	LCL	Forecast	UCL
2024:12		---	---	1121	---
2025:01	7.03389	0.25774	685	1135	1880
2025:02	7.04711	0.30683	630	1150	2098
2025:03	7.04477	0.33112	599	1147	2195
2025:04	7.03199	0.36568	553	1132	2319
2025:05	7.0214	0.40642	505	1120	2485
2025:06	7.02242	0.44543	469	1122	2685
2025:07	7.03378	0.47764	445	1134	2893
2025:08	7.04458	0.50191	429	1145	3067
2025:09	7.04465	0.52227	412	1147	3192
2025:10	7.03423	0.54414	391	1135	3297
2025:11	7.02342	0.57031	367	1123	3433
2025:12	7.02241	0.5984	347	1122	3624
2026:01	7.03187	0.62329	334	1132	3841
2026:02	7.04259	0.64283	325	1144	4034
2026:03	7.04445	0.65927	315	1147	4174
2026:04	7.03596	0.67646	302	1137	4280
2026:05	7.0254	0.69692	289	1125	4409
2026:06	7.02277	0.71957	274	1122	4597
2026:07	7.03031	0.74063	265	1130	4827
2026:08	7.04064	0.75777	259	1142	5043
2026:09	7.04395	0.77217	253	1146	5205
2026:10	7.03734	0.78677	244	1138	5321
2026:11	7.02731	0.80385	233	1127	5447
2026:12	7.02339	0.82309	224	1123	5634
Total	168.81063			27225	
Average	7.03377625			1134.375	

Note: For 95% confidence intervals, $Z_{0.025} = 1.96$. LCL and UCL denote lower and upper confidence limits respectively.

The forecast data provided in Table 8 offer insights into the anticipated trajectory of HIV infection rates in the study area, factoring in other opportunistic infections. For instance, the forecasted value of HIV infection for January 2025 is projected to be 1135 persons, with a 95% confidence interval spanning from 685 to 1880 persons. This interval suggests that there is a 95% probability that the actual number of HIV infections in the presence of other opportunistic infections will fall within this range for the upcoming period.

Comparing this forecasted value with the number of infections reported in December 2024 (which stood at 1121 persons), we anticipate a marginal increase of 14 persons in HIV infections for January 2025. However, the confidence interval [685, 1880] implies a potential range of outcomes, indicating that HIV infections could decrease by as much as 450 persons or increase by as many as 745 persons compared to the previous month.

Further examination of the forecast reveals specific projections for HIV infections throughout 2025 and 2026. For instance, the forecast predicts at least 1147, 1122, 1147, and 1122 persons to be infected with HIV in Benue state during the months of March, June, September, and December 2025, respectively. Similarly, in 2026, approximately 1132, 1147, 1122, 1146, and 1123 persons are forecasted to contract the disease in January, March, June, September, and December, respectively. Cumulatively, the forecast suggests a total of 27,225 persons will be afflicted with HIV in the study area over the years 2025 and 2026 with an average monthly morbidity incidence of 1134 persons per month.

Moreover, the forecast unveils a fluctuating trend in HIV infection rates over time throughout the forecasted period, exhibiting both increasing and decreasing patterns. This fluctuation mirrors the trends observed in the original series, highlighting the dynamic nature of HIV infection rates in the presence of opportunistic infections among farmers in Benue state, Nigeria.

Model implications on the postharvest losses and agricultural productivity

The implications of HIV/AIDS and other opportunistic infections among farmers in relation to postharvest losses of crops and agricultural productivity in Benue state can be multifaceted:

(1) *Reduced Labour Force*: HIV/AIDS can lead to a reduced labour force among farmers due to illness and death, impacting their ability to effectively manage postharvest activities such as harvesting, storage, and transportation of crops. This reduction in manpower can result in delays in handling crops, leading to increased spoilage and losses.

(2) *Decreased Productivity*: Farmers living with HIV/AIDS may experience decreased productivity due to illness, fatigue, and weakened immune systems, affecting their capacity to properly handle crops during postharvest stages. This decreased productivity can contribute to inefficient postharvest management practices and higher rates of crop spoilage.

(3) *Limited Access to Resources*: HIV/AIDS can lead to economic challenges for affected farmers, including decreased income and limited access to resources such as agricultural inputs, storage facilities, and transportation. This can result in inadequate infrastructure and technologies for postharvest handling, increasing the susceptibility of crops to spoilage and losses.

(4) *Increased Vulnerability to Infections*: Individuals living with HIV/AIDS are more susceptible to opportunistic infections, which can further compromise their ability to effectively manage postharvest activities. These infections may exacerbate health issues, leading to absenteeism, reduced efficiency, and increased risks of contamination of crops during handling and storage.

(5) *Stigma and Discrimination*: Farmers living with HIV/AIDS may face stigma and discrimination within their communities, impacting their ability to access support networks, agricultural markets, and extension services. This social isolation can hinder their capacity to adopt improved postharvest technologies and practices, thereby increasing the likelihood of crop losses.

Addressing the implications of HIV/AIDS and opportunistic infections among farmers requires comprehensive strategies that integrate healthcare, social support, and agricultural interventions. Efforts to provide access to healthcare services, promote awareness and education on HIV/AIDS prevention and treatment, and enhance agricultural productivity and resilience can contribute to mitigating the impact of these diseases on postharvest losses of crops and improving the livelihoods of affected farmers.

Paired samples t-test result

The paired samples t-test was conducted on the actual and forecast values for an in-sample period of 24 months starting from January, 2023 to December, 2024 with 24 sample points. The paired samples statistics and correlations are presented in Table 9 while the paired samples t-test result is reported in Table 10.

The paired samples statistics reported in Table 9 showed a mean of 6.9833 for the actual series and a mean value of 6.9792 for the forecast series. These means indicate no significant difference between the actual and the forecast series. Also, the paired samples correlation for the actual and forecast series reported in the lower panel of Table 9 showed that both the actual and forecast series move in the same direction. That is, increase in the values of the actual series will lead to a corresponding increase in the values of the forecast series and vice versa.

The paired samples t-test statistic result reported in Table 10 showed a t-statistic value of 0.450 with an insignificant p-value of 0.657 ($p > 0.05$). This result indicates that there is no significant difference between

the actual and the forecast series. Thus, it is concluded that the forecast values of HIV infection in Benue state are reliable, valid and accurate and can be relied upon for policy implementation.

Table 9: Paired Samples Statistics and correlations

Paired Samples Statistics				
Variable	Mean	N	Std. Deviation	Std. Error Mean
Actual	6.9833	24	0.05130	0.01047
Forecast	6.9792	24	0.04995	0.01020
Paired Samples Correlations				
Variable	N	Correlation	p-value	
Actual & Forecast	24	0.598	0.002	

Table 10: Paired Samples Test Result

	Mean	Std. Dev.	Std. Error Mean	95% CI of the Difference		t-stat.	df	p-value
				Lower	Upper			
Actual - Forecast	0.00417	0.0454	0.0093	-0.0150	0.0233	0.450	23	0.657

CONCLUSION

This study provides a valuable contribution to understanding and predicting the burden of HIV infection in the presence of co-infections among farmers in Benue State, Nigeria. Using monthly epidemiological data spanning fifteen years (2010-2024), an Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) model was employed to forecast HIV seropositivity with Tuberculosis (TB) and Hepatitis B Virus (HBV) as significant opportunistic predictors. The ARIMAX (4,1,3) model, selected through rigorous diagnostic and model selection criteria, effectively captured the underlying patterns and variability in the data, explaining approximately 65.93% of the total variability.

The predictive analysis forecasted a total of 27,225 HIV cases for the 2025-2026 period, with an average monthly incidence of 1,134 individuals, revealing a fluctuating yet dynamic trend in HIV infection rates over time. The reliability of the model was affirmed through comparative tests between actual and forecast values, showing no statistically significant differences. This indicates that the model is robust and suitable for use in practical health planning and policy formulation.

Given the persistent burden of HIV in the agricultural population, especially in the context of immunosuppressive co-infections like TB and HBV, the findings underscore the need for integrated and data-driven interventions. The study strongly advocates for collaborative efforts involving the Benue State government, international donor organizations, public health stakeholders, and agricultural extension services to implement context-specific HIV prevention and control strategies. Emphasis should be placed on community-based awareness programmes, increased access to testing and treatment, and the strengthening of surveillance systems to prevent the worsening of HIV epidemics among vulnerable farming communities.

Ultimately, this research not only provides a methodological framework for short-term infectious disease prediction using ARIMAX models but also serves as a strategic guide for evidence-based policy-making in the fight against HIV and its consequences on rural livelihoods and food security in Benue State and similar settings.

REFERENCES

1. Abu, G. A., & Kotur, L. N. (2022). Impact of HIV/AIDS on farmers' productivity in Nigeria: Evidence from Benue State. *Direct Research Journal of Agriculture and Food Science*, 10(4), 95-102.
2. Agber, T., & Aondofa, A. S. (2023). Level of cassava post-harvest losses and the socio-economic wellbeing of Tiv farmers in Benue State. *Journal of Agriculture and Food Sciences*, 21(2), 45-52. <https://www.researchgate.net/publication/373237039>

3. Akaike, H. (1974). A new look at statistical model identification, *IEEE Transactions on Automatic Control*, AC-19: 716-723.
4. Aliyu, A., & Akor, D. (2023). Thematic measures for reducing post-harvest losses of vegetable crops among farmers in Benue State, Nigeria. *International Journal of Agricultural Sciences*, 9(2), 1-10. <https://journals.innovareacademics.in/index.php/ijags/article/view/46590>
5. Andrews, B., Dean, M., Swain, R. and Cole, C. (2013). Building ARIMA and ARIMAX models for predicting long-term disability benefit application rates in the public/private sectors. *society of actuaries*.
6. Bhargava, A. (1986). On the theory of testing for unit roots in observed time series. *Review of Economic Studies*, 53: 369-384.
7. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.
8. Chen, Y.-P., Liu, L.-F., Che, Y., Huang, J., Li, G.-X., Sang, G.-X., Xuan, Z.-Q., & He, T.-F. (2022). Modeling and Predicting Pulmonary Tuberculosis Incidence and Its Association with Air Pollution and Meteorological Factors Using an ARIMAX Model: An Ecological Study in Ningbo of China. *International Journal of Environmental Research and Public Health*, 19(9), 5385. <https://doi.org/10.3390/ijerph19095385>
9. Durbin, J. (1960). The fitting of time series models. *Review of the Institute of International Statistics*, 28, 233-244.
10. Elliot, G., Rothenberg, T. J. and Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, 64: 813-836.
11. FAO. (2014). The state of food and agriculture: Innovation in family farming. Food and Agriculture Organization of the United Nations. <https://www.fao.org/3/i4040e/i4040e.pdf>
12. Federal Ministry of Health (FMoH). (2020). National HIV & AIDS Indicator and Impact Survey (NAIIS) 2018: Technical Report. Abuja, Nigeria.
13. Hannan, E. (1980). The estimation of the order of ARMA process. *Annals of Statistics*, 8, 1071-1081.
14. Ikya, B. J., & Igbokwe, M. C. (2019). Determinants of postharvest losses among tomato farmers in Gboko Local Government Area of Benue State. *International Journal of Advanced Studies in Economics and Public Sector Management*, 7(1), 45-56.
15. Imai, C., Armstrong, B., Chalabi, Z., Mangtani, P., & Hashizume, M. (2015). Time series regression model for infectious disease and weather. *Environmental Research*, 142, 319-327.
16. Jarque, C. M. and Bera, A. K. (1980). Efficient test for normality, heteroskedasticity and serial independence of regression residuals. *Econometric Letters*, 6: 255-259.
17. Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55(2), 163-172.
18. Kane, M. J., Price, N., Scotch, M., & Rabinowitz, P. (2014). Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks. *BMC Bioinformatics* 15, 276-288. <https://doi.org/10.1186/1471-2105-15-276>.
19. Ljung, G. M., & Box, G. E .P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65, 297-303.
20. Musa, B. M., Fawibe, A. E., Sani, M. U., & Ajayi, I. O. (2021). Prevalence of hepatitis B virus infection among people living with HIV in Nigeria: A systematic review and meta-analysis. *BMC Public Health*, 21(1), 1–13. <https://doi.org/10.1186/s12889-021-10549-y>.
21. National Agency for the Control of AIDS (NACA). (2022). Nigeria HIV/AIDS Indicator and Impact Survey (NAIIS). <https://naca.gov.ng>
22. Ng, S. & Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69(6): 1519-1554.
23. Onovo, A. A., Adeyemi, A., Onime, D., Kalnoky, M., Kagniniwa, B., Dessie, M., Lee, L., Parrish, D., Adebobola, B., Ashefor, G., Ogorry, O., Goldstein, R., & Meri, H. (2023). Estimation of HIV prevalence and burden in Nigeria: a Bayesian predictive modelling study. *eClinicalMedicine*, 62, 102098.
24. Pindyck, R. S., & Rubinfeld, D. L. (1998). *Econometric Models and Economic Forecasts*, Fourth Edition, McGraw-Hill.
25. Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2): 461-464.

26. Stock, J. H. (1990). A class of tests for integration and cointegration. Mimeo, Harvard University.
27. UNAIDS. (2023). Global HIV & AIDS statistics-Fact sheet. <https://www.unaids.org/en/resources/fact-sheet>
28. Zhou, Q., Hu, J., Hu, W., Li, H., & Lin, G. –Z. (2023). Interrupted time series analysis using the ARIMA model of the impact of COVID-19 on the incidence rate of notifiable communicable diseases in China. BMC Infectious Diseases, 23, 375 (2023). <https://doi.org/10.1186/s12879-023-08229-5>