Vanlalhriatsaka / Afr. J. Bio. Sc. 6(5) (2024) ISSN: 2663-2187
https://doi.org/10.33472/AFJBS.6.5.2024. 3974-3984

African Journal of Biological Sciences

A Time Series Analysis Of HIV Data In Mizoram During 2007-2022

Thanhmingliana[1], Lalpawimawha[2], Vanlalhriatsaka[2]*, L P Lalduhawma[3], Denghmingliani Zadeng[3], Anupam Kumar[2]

[1]Department of Chemistry, Mizoram University, Pachhunga University College Campus, Aizawl, Mizoram.
[2]Department of Statistics, Mizoram University, Pachhunga University College Campus, Aizawl, Mizoram.
[3]Department of Mathematics, Mizoram University, Pachhunga University College Campus, Aizawl, Mizoram.

***Corresponding Author:** Vanlalhriatsaka*
Email: Vlhriatsaka96@gmail.com

Article History
Volume 6, Issue 5, May 2024
Received: 30 Apr 2024
Accepted: 22 May 2024
Doi: 10.33472/AFJBS.6.5.2024.3974-3984

Abstract

This paper presents a comprehensive study on the uses of time series analysis techniques to HIV data, aiming to explore patterns, trends, and forecast future developments in the epidemic. With the increasing availability of longitudinal HIV data, time series analysis offers a powerful framework for understanding the dynamic nature of the disease and guiding public health interventions. This study begins by collecting and pre-processing a diverse range of HIV data, including infection rates, demographic factors, treatment utilization, and socio-economic indicators. The collected data spans multiple years, providing a temporal dimension for analysis. Time series analysis techniques, such as autoregressive integrated moving average (ARIMA) are employed to identify temporal patterns, seasonality, and lifelong trends in the data. Furthermore, this study investigates the impact of specific interventions and policies on the HIV epidemic by incorporating external factors and interventions into the time series models. These factors may include the introduction of new antiretroviral treatments, public awareness campaigns, or changes in healthcare policies. By quantifying the effects of these interventions on the HIV time series, policymakers can make informed decisions and allocate resources effectively. Moreover, this study focuses on forecasting future HIV trends using time series models. By extrapolating historical patterns and incorporating the effects of interventions, the models can provide valuable insights into future infection rates, treatment demand, and potential challenges. These forecasts aid in proactive resource planning, early detection of potential outbreaks, and the development of targeted prevention strategies. In closure, this comprehensive research highlights the significance of time series study in understanding the temporal dynamics of HIV data. By analyzing historical trends, evaluating interventions, and forecasting future developments, time series techniques serve as a valuable tool for policymakers, healthcare professionals, and researchers involved in combating the HIV epidemic. The findings of this study contribute to evidence-based decision - making and the development of effective strategies for avoidance, medication, and supervision of HIV/AIDS.

Keywords: HIV, AIDS, ARIMA, STATIONARY, MIZORAM.

1. INTRODUCTION

HIV (HUMAN IMMUNODEFICIENCY VIRUS) originally came to humans from Chimpanzees (simian immunodeficiency virus) in the years 1800s. Mizoram has the highest rate of incidence of this deadly disease among the different states of India. This virus raids the human immune system and without a strong immune system, various diseases can destroy and attack our body which may eventually lead to fatality. So, it is very important to know how HIV is transmitted and how to prevent it. HIV can be transmitted through sex without a condom, sharing of syringe, HIV-infected mother to child, getting an HIV-infected blood etc. These are the common points for the transmission of HIV. However, HIV cannot be transmitted by air, handshaking, water, mosquitoes, and any other insect sting, sharing a bathroom, dining room or bed etc.

According to the National AIDS Control Organization, Govt. of India (National AIDS Control Organization | MoHFW | GoI (naco.gov.in), Mizoram state has the highest HIV cases all over the country. HIV was first detected in Mizoram in October 1990, and within a few years, Mizoram has the highest number of HIV cases in India. Mizoram State Aids Control Society (MSACS)(MizoramSACS.org – Official Website of Mizoram State AIDS Control Society) was established in 1992 and this society plays a very important role in combating HIV in Mizoram. According to the report of MSACS, the highest transmission of HIV is from sex abuse, followed by sharing of unclean syringes. According to the MSACS report, HIV cases were found more in males than in females and the highest age group of HIV-positive cases in Mizoram is between the ages 25 – 34 and is followed by 35-49(MizoramSACS.org – Official Website of Mizoram State AIDS Control Society).

The Mizoram government launched an Anti-retroviral therapy (ART) center and this center plays a major role in fighting HIV in Mizoram. The center conducts awareness programs frequently which have been very successful, and have been quite beneficial especially for infected persons. An HIV-infected person can take this medicine free of cost at the center and are informed on the importance of taking the medicine regularly which may enable them to live a long and healthy life. The incidence of HIV and the consequences thereby could worsen unless the state government, civil society, and influential religious denominations work together to combat this threat on a war footing.

2. Objective of the Study

The objective of the research is to utilize ARIMA techniques to develop models for HIV data in Mizoram spanning the years 2007 to 2022. Furthermore, the study seeks to forecast the HIV data for Mizoram over the next 10 years.

3. Source of Data

To fulfill the objective of this present study, the data considered is secondary in nature. It has been collected from MSACS office in Ramhlun North, Aizawl, Mizoram.

4. Methodology

The Box-Jenkins Model is a mathematical approach employed to forecast data ranges by analyzing specified time series inputs. It can effectively handle different types of time series data for forecasting purposes. The model incorporates three fundamental principles – Autoregression, Differencing and Moving Average – abbreviated as p, d and q respectively. These principles are collectively referred to as ARIMA (p,d,q) in Box-Jenkins analysis. In this context, p denotes the

autoregressive component, d represent the differencing part and q represents the moving average component. The Box-Jenkins Method is a valuable tool for identifying the most suitable model fit.
Stationary test is one of the most important analysis in time series analysis. In order to conduct time series analysis using an ARIMA Model, it is crucial to determine whether the raw data is stationary or non-stationary. A stationary time series exhibits a constant mean, variance and autocorrelation structure. Hence, it is essential to carefully examine the raw data to assess its stationary nature before proceeding with ARIMA Modelling.

The next step involves checking whether the data is stationary or non-stationary. This can be done by following the procedure outlined below :-

a) The initial step is to create a diagram or plot of the raw data. By examine this plot, we can identify whether the data exhibits a stationary pattern. In the case of stationary time series, the plot will display a consistent location and scale throughout.
b) Performing ACF and PACF plots is another important step. In addition to examining the time series plot, the ACF Plot is particularly helpful in identifying non-stationary time series. In the case of a stationary time series, the ACF values decrease rapidly and approach zero. However, for non-stationary data, the ACF values decline more slowly. Therefore, analyzing the ACF plot can provide valuable insights into the stationary of the data.

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{autocovariance\ at\ lag\ k}{variance\ of\ the\ time\ series}$$

c) Dickey-Fuller (ADF) unit root test.

If the time series is stationary, it means that it exhibits a constant mean, variance and autocorrelation structure. However, if the time series displays a trend, we can use the differencing method to eliminate the linear or curvilinear trend. Once we have converted the non-stationary series into a stationary one through the differencing, the upcoming step is to determine the order of the ARIMA (p,d,q) model. In this model, p represents the autoregressive part, d represents the order of differencing and q represents the moving average part. ACF and PACF plots of the series are crucial in identifying the appropriate order of the AR and MA terms. Although these plots do not directly dictate the order of the ARIMA model, they provide valuable insights for understanding the order and selecting a suitable model that fits the time series data.
Once the parameters for the ARIMA model are determined using the ACF and PACF Plots, we can identify the significant lags. Subsequently, it is necessary to test various models up to those lags in order to determine the best model. The Akaike Information Criteria (AIC) is a criterion used for model selection among a finite set of models. The model with the lowest AIC is considered the most favorable. Ideally, the best model is the one that exhibits the minimum AIC value when compared to other models.

$$AIC = -2\ln(L) + 2k$$
$$= n\ln(\sigma_a^2) + 2k$$

Where 'σ_a^2 ' is the M.L.E and k is the number of the parameters estimated in the model.

The formula for calculating Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE) as follows; -

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

Where, y_i = prediction
x_i = true value
n = total number of data points.

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right| \mathrm{X}100\%$$

Where, n = number of times the summation iteration happens.
A_t = Actual value
F_t = Forecast value

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} |y_i - z_i|^2}{N}}$$

Where, N is th
Yi = i^{it} measurement.
Zi= corresponding prediction.

Ljung – Box Test (L.B. Test):-

The L.B test is a statistical test is used if any of the autocorrelations in a time series are significantly different from zero. It is used extensively in econometrics and other areas of time series analysis. The test is specifically applied to the residuals obtained from fitting an ARIMA Model, rather than the original series. In this analysis, the test aims to examine whether the residuals exhibits any autocorrelation. It is important to note that when conducting this test on the residuals of an estimated ARIMA model, the degrees of freedom should be adjusted to account for the parameter estimation process.

The test statistics is given by

$$\text{LB = alue} \quad \mathrm{n(n+2)} \sum_{k=1}^{h} \frac{\rho_k^2}{n-k} \sim \chi_m^2$$

The L.B test compares the calculated value of Chi-Square to the observed value by considering the sample size (n), the sample autocorrelation at lag k and the number of lag being tested. If the calculated value of Chi-Square is less than the observed value, the null hypothesis is rejected. In such cases, there is no evidence to support the presence of white noise between the tested lags.

5.1 Stationary Test

In Figure 1, the time series displays the HIV data of Mizoram from 2007 – 2022. Upon analyzing the plots, it is evident that the pattern initially increased from 2007 to 2011, followed by a decrease until 2012. Subsequently, there was another period of increase and a rapid decrease starting from 2014, followed by a significant rise. Therefore, it can be observed that the raw data exhibits non-stationary behaviors in terms of both mean and variance. To further examine the non-stationary pattern, we refer to Figure 2 and Figure 3, which present the ACF and PACF plots, respectively. The

ACF plot shows that the autocorrelation values touch the significant boundary at lag 2 and then gradually decrease. Based on this observation, we can conclude that the raw data is non-stationary data in nature. To confirm this, we conducted the ADF (Augmented Dickey-Fuller) test, resulting in a Dickey-Fuller statistics of -2.0965 and a p-value of 0.4215. Therefore, we can confirm that the HIV data of Mizoram from 2007 – 2022 is indeed non-stationary. To convert the non-stationary data into a stationary form, we should employ the differencing method.

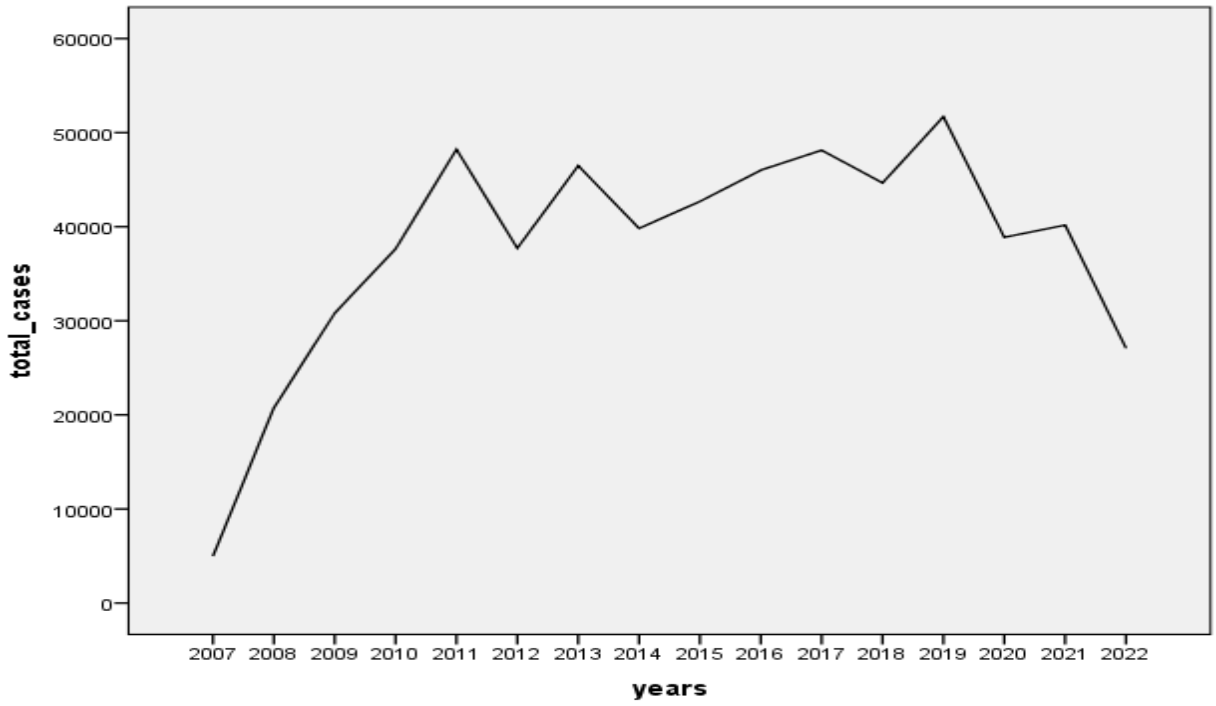


Figure 1: Time series Plot

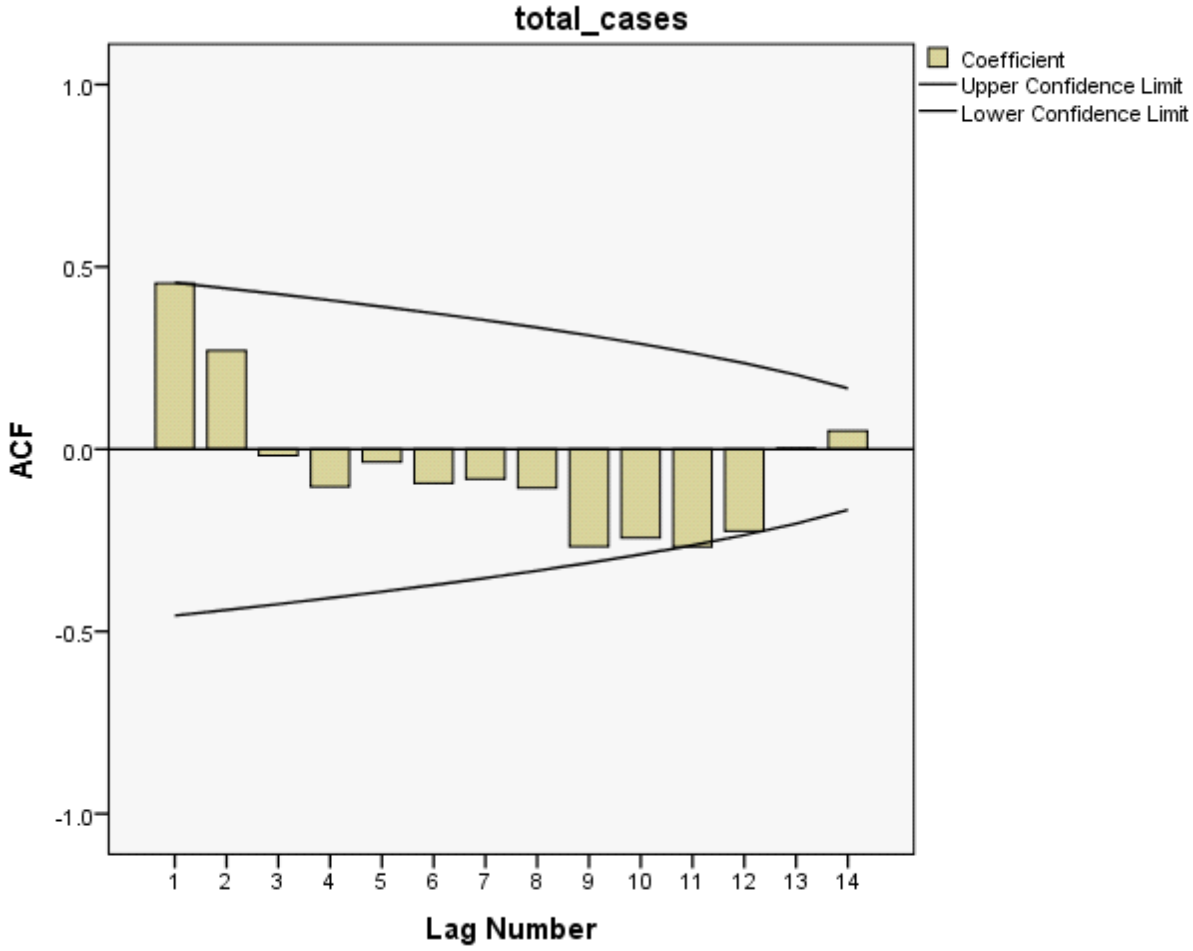


Figure 2: ACF Plot

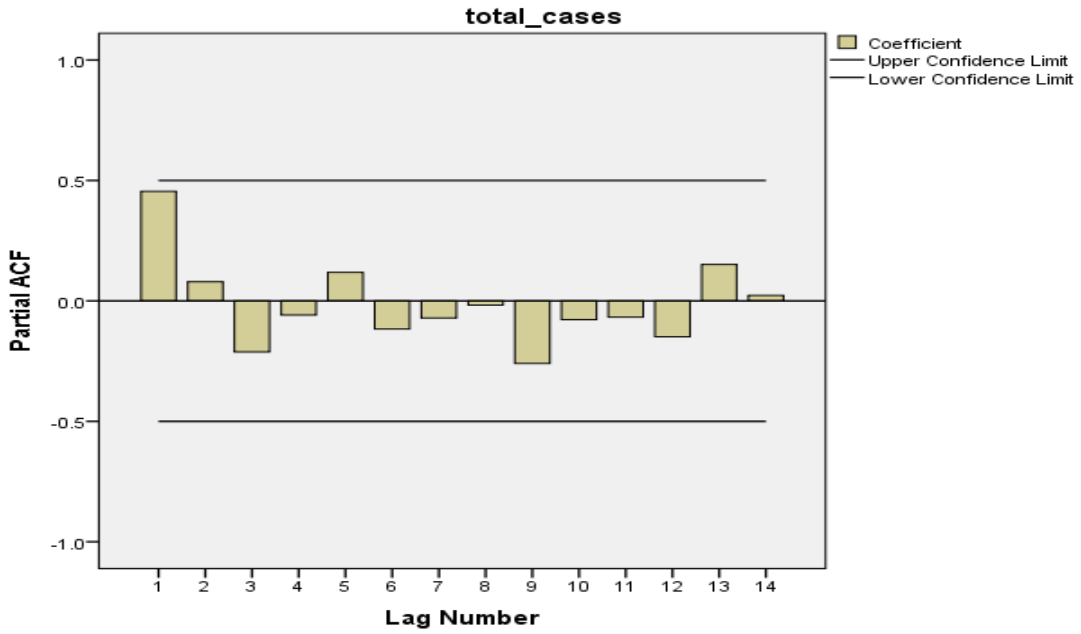


Figure 3: PACF Plot

Figure 4 illustrates that the second difference of the HIV data time series plot. Additionally, Figure 5 and Figure 6 represent the second difference of the ACF and PACF Plots, respectively. By analyzing

these plots, we can say that the non-stationary state has been eliminated, and the data has become stationary in terms of the mean. To confirm the stationary state, we performed the ADF test. The results of the test yielded a Dickey-Fuller statistic of -2.1386 and a p-value of 0.02456. Based on the ADF test, we can confidently state that the data is stationary.

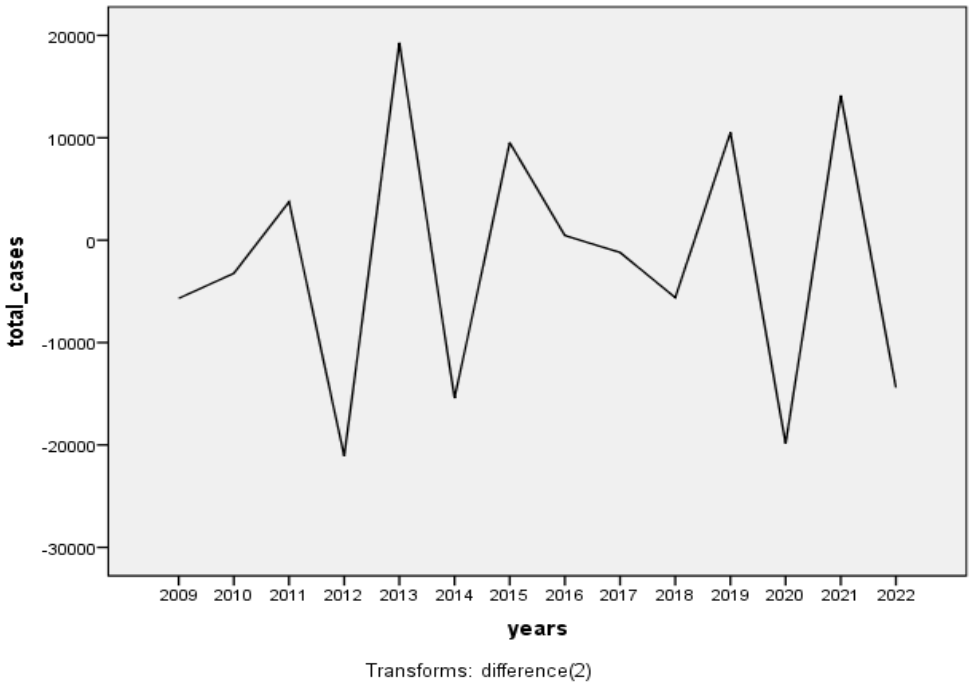


Figure 4: Time series plot after the second difference

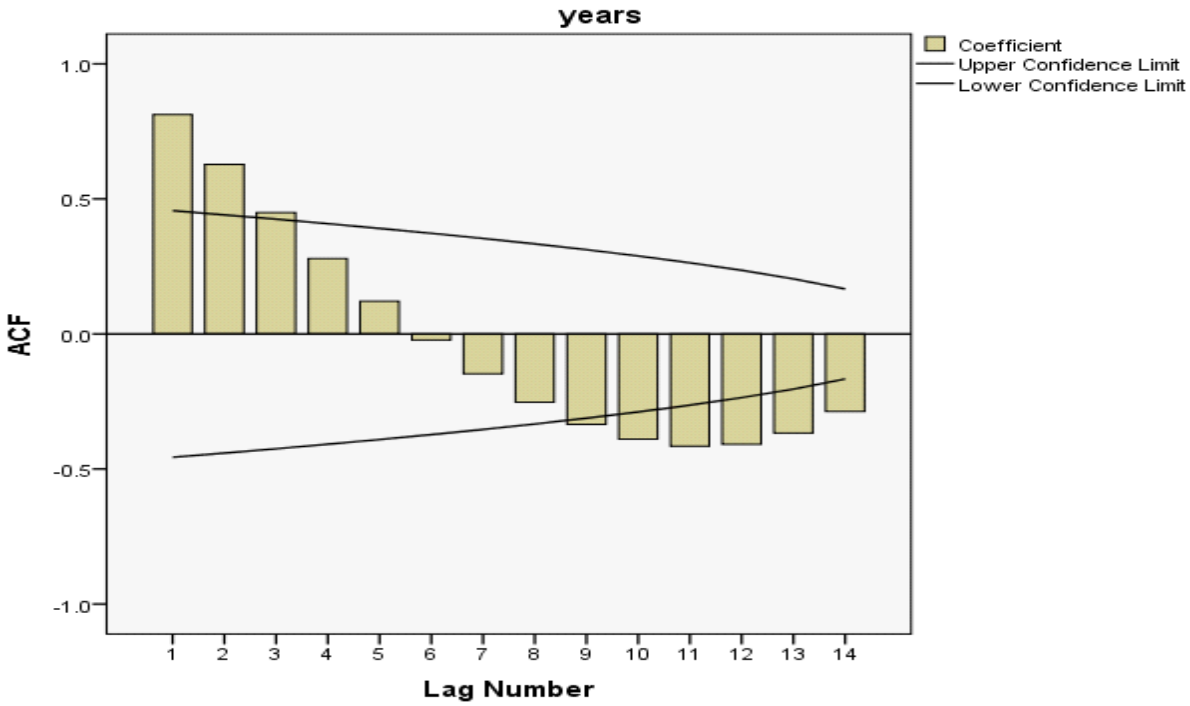


Figure 5: ACF plot after second differenced

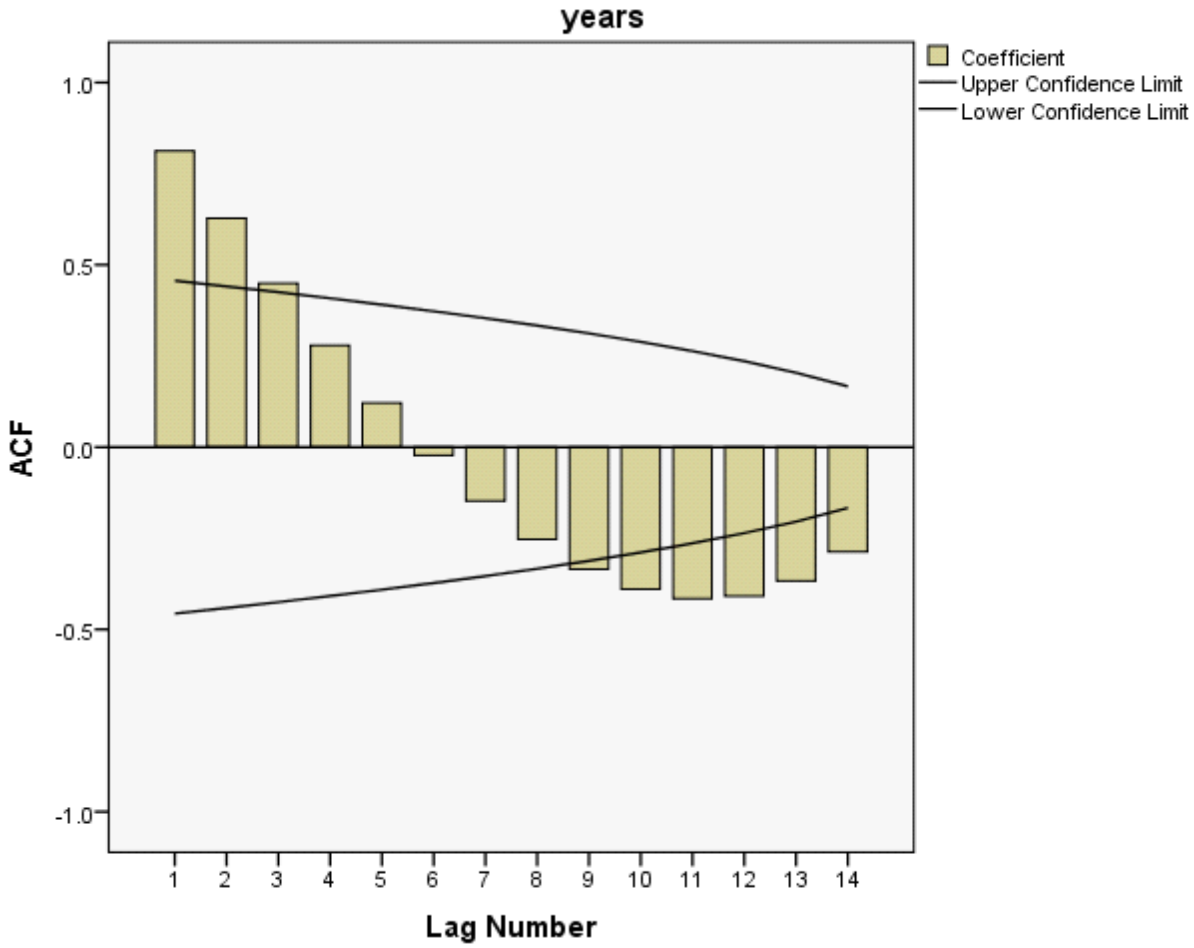


Figure 6: PACF plot after the differenced

By examining the ACF plots after performing second differencing, we observed that the plot at lag 1 touches the significant boundary and rapidly drops towards zero. Additionally, all other lag values remain below the significant boundary. This observation leads us to confirm that the time series has achieved stationary state.

5.2 Model Identification

The next step involves determining the order of the ARIMA (p,d,q) model based on the ACF and PACF plots of the second difference. In this study, the data has been differenced two times and indicating a value of 'd' as 2. By examining the ACF plots of the second difference in Figure 5, we observed that only lag 1 touches the significant boundary, while all other lags remain within the bounds. Hence, a possible value for 'q' is 1, suggesting that an MA(0) model could be the most appropriate. Moving on to the PACF plots in Figure 6, we notice that both lag 1 and lag 2 touch the significant boundary, while remaining lags stay within the bounds. Consequently, a possible value for 'p' is 2, indicating that an AR(2) model could be the most suitable choice. Therefore, possible beat fit models are as follows:-

A) ARIMA (1,2,0)
B) ARIMA (2,2,1)
C) ARIMA (2,2,0)

D) ARIMA (1,2,1)
E) ARIMA (0,2,1)
F) ARIMA (0,2,0)

The AIC, MAE and MAPE values for the different ARIMA are as follows :-

Model	AIC	RMSE	MAE	MAPE
ARIMA (1,2,0)	293	5334.092	4172.240	11.473
ARIMA (2,2,0)	295	6945.563	4976.369	12.279
ARIMA (1,2,1)	298	7170.824	4937.664	12.156
ARIMA (2,2,1)	301	7315.486	4933.369	12.132
ARIMA (0,2,1)	299	8660.587	6368.806	16.040
ARIMA (0,2,0)	302	13187.246	10128.130	25.376

Akaike Information Criteria (AIC) is a criterion used for model selection among the finite set of models. The model with the lowest AIC is the best model. A good model is the one that has minimum AIC is considered the best fit and a lower AIC indicates that a better model. So, referring to the table provided, we can determine that the best fit model is ARIMA (1,2,0) with the minimum AIC value among the other models.

5.3 Model Estimation

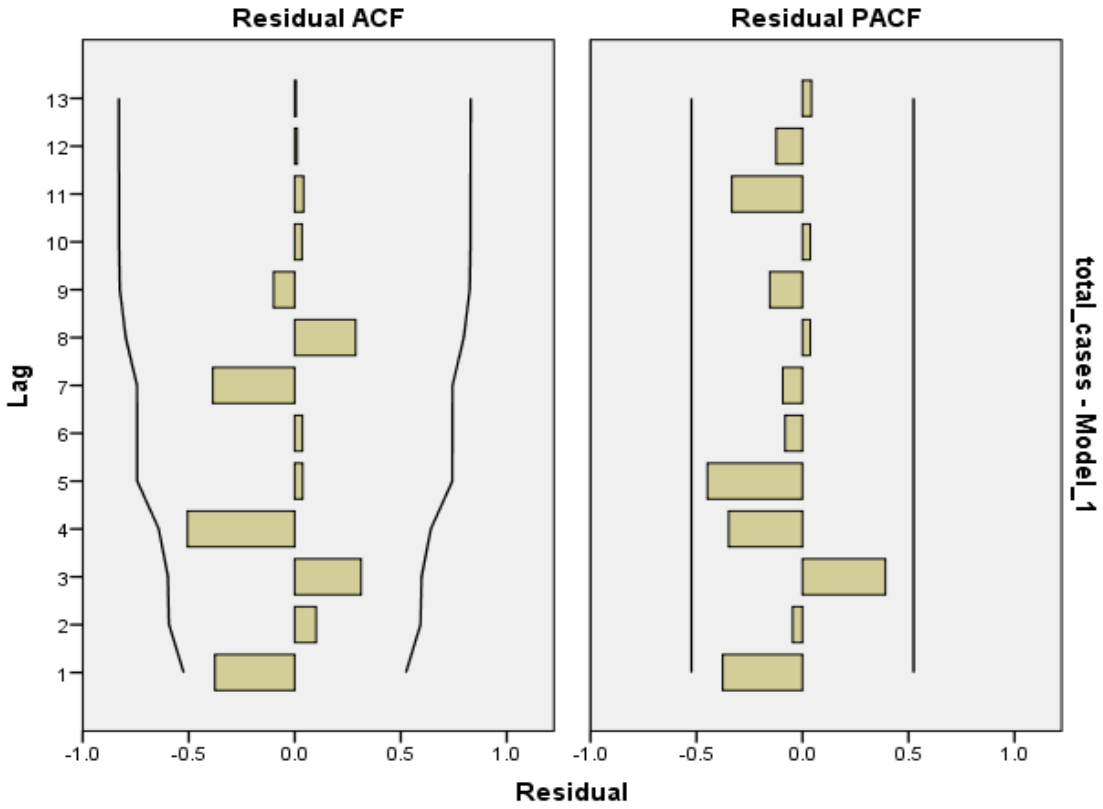
In the present study, the best fit model is ARIMA (1,2, 0) with AIC value 293 . Then, the next step is to estimate the parameters and the coefficient of estimated parameters for ARIMA (1,2,0) Model is given by

Variables	Coefficients	S.E	σ^2	Log likelihood
MA (0)	-0.8108	0.1396	53901253	-144.5

5.4 Diagnostic Check

Diagnostic checks are performed when we intend to draw inferences using a model, especially when the estimated standard errors of the parameters are utilized for such inferences. To assess the appropriateness of the fitted model, it is common practice to examine the residuals. The residual should exhibit characteristics similar to observations from a white noise process, showing no correlation with each other and being identically distributed. These diagnostic checks help us evaluate how well our model fits the data. After fitting the ARIMA(1,2,0) model, we conducted diagnostic checks using Ljung-Box test and examined the ACF and PACF plots of the residuals. The results of these checks are provided below:-

Upon analyzing the residual ACF and PACF plots, we observed that all the plots remained below the significant boundary. None of the plots touched or exceeded the significant boundary. So, we can say that the ARIMA (1,2,0) model is the most suitable model for the time series analysis of our HIV data. With Ljung Box Chi-Square Test, χ^2=12.958 , p-value =0.05 and χ^2= 27.59 (tabulated).



The above figure represents the Residual ACF and PACF Plot of the time series of HIV.

5.5 Forecasting for the next 10 years.

YEAR	FORECAST	95% Prediction Level	
		Lower	Upper
2023	25660.793	11271.323	40050.35
2024	14779.724	-7578.507	37137.95
2025	11569.363	-25282.327	48421.05
2026	2139.720	-47869.817	52149.26
2027	2247.433	-69214.313	64719.45
2028	10722.952	-94432.606	72986.70
2029	15883.692	-11877.671	87110.29
2030	23732.000	-146310.730	98846.73
2031	29401.273	-173426.797	114624.25

2032	36837.271	-202861.391	129186.85

Once we have identified the finest model ARIMA (1,2,0) , for our data, the next step is to forecast the number of HIV cases in Mizoram for the next 10 years. This prediction holds significant importance for the people of Mizoram, particularly those residing in rural areas. The table below shows the forecasted value of HIV cases in Mizoram, along with a 95% prediction interval, after fitting the ARIMA (1,2,0) model.

6. Discussion

The Box-Jenkins ARIMA model was applied to analyze HIV data from Mizoram spanning the years 2007 to 2022. In this research, we determined that the ARIMA (1,2,0) is the best model to suit our time series data. To achieve stationary state, the raw data underwent differencing two times, resulting in non-stationary data (Dickey-Fuller = -2.0965, p-value = 0.4215), which was further transformed into stationary data (Dickey-Fuller = -2.1386, p-value = 0.02456). Stationary state was confirmed by examining the second differenced data plot of ACF and PACF, leading us to denied the null hypothesis and establish stationary state. Next, we selected the ARIMA (p,d,q) parameters by analyzing the second differenced data of ACF and PACF plots. The model parameters were estimated and a diagnostic test was conducted on the residuals. With the time series now stationary, we forecasted HIV cases in Mizoram for the next 10 years, including 95% prediction intervals, using IBM SPSS. The forecasted values suggest that Mizoram will experience a lower number of HIV cases compared to previous years, but still higher than the reported cases in 2014. This indicates a downward trend in HIV cases in Mizoram.

7. CONCLUSION

From our results, we can conclude that we need more awareness about HIV and its prevention, which will contribute towards reducing new infections, combating stigma, improving approach to testing and medication and enhancing the overall well-being of individuals affected by HIV. A comprehensive and multifaceted approach to HIV education is essential to create a society that is knowledgeable, supportive and committed to ending the HIV epidemic.

References:

1. Mizoram State AIDS Control Society (MSACS), Ramhlun North Aizawl 796012.
2. S.C. Gupta and V.K.Kapoor (2013). Fundamental of Mathematical Statistics, Sultan Chand and Sons, 23, Daryaganj, New Delhi-110002, **(2013).**
3. "Meteorological Data of Mizoram 2017 and 2020", Directorate Of Economics and Statistics, Planning and Programme Implementation Department, Government Of Mizoram. Website www.des.mizoram.gov.in .
4. G.E.P. Box , G.M. Jenkin and G.C. Reinsel. Time Series Analysis and Forecasting and Control, Third Edition, Delhi, India Pearson Education, **(2003).**

5. S. Chakraborty, L. Choudhury, R. Geeta and Sarmah. Results of Appeared Candidates in Matriculation / HSLC Examination in Assam (India): A Time Series Analysis (1951-2014), **(2016).** Pp. 4-15.
6. G.S. Shilpashree. Statistical Analysis of Rainfall Pattern in Mandya District, University of Agricultural Science, Bengaluru, **(2015).**
7. S. Bisgaard and M. Kulahci. Time Series Analysis, Forecasting by Example, Hoboken,New Jersey, Wiley, A John Wiley & Sons, INC., **(2011).**
8. F. Jamal, E. Latifa, A. Zineb, E.L. Haj and L. Abdeslam. Forecasting of demand using ARIMA model, **(2018).**
9. R.K. Agrawal and R. Adhikari. An Introductory Study on Time Series Modelling and Forecasting, (2013).

Cite this article as: Vanlalhriatsaka, A Time Series Analysis Of HIV Data In Mizoram During 2007-2022, African Journal of Biological Sciences. 6(5), 1-12.
DOI: xyz