Adults newly infected with hiv in burundi: a box-jenkins arima approach

Dr. Smartson. P. NYONI
ZICHIRe Project, University of Zimbabwe, Harare, Zimbabwe
Mr. Thabani NYONI
Department of Economics, University of Zimbabwe, Harare, Zimbabwe

ABSTRACT

Using annual time series data on the number of adults (ages 15 and above) newly infected with HIV in Burundi from 1990 – 2018, the study predicts the annual number of adults who will be newly infected with HIV over the period 2019 – 2025. The study applied the Box-Jenkins ARIMA methodology. The diagnostic ADF tests as well as correlogram analysis show that the G series under consideration is an I (2) variable. Based on the AIC, the study presents the ARIMA (0, 2, 1) model as the optimal model. The residual correlogram and the inverse roots of the applied model further reveal that the presented model is stable and suitable for forecasting new HIV infections in adults in Burundi. The results of the study indicate that the number of new HIV infections in adults in Burundi will most likely decline, over the period 2019 – 2023, from approximately 698 to almost 90 new HIV infections. By 2025, Burundi could experience her first zero new HIV infections in adults! This implies that, despite the fact that Vision Burundi 2025 is a highly ambitious blue-print; Vision Burundi 2025 will largely be achieved as far as HIV/AIDS prevention and control is concerned.

1.0 INTRODUCTION

Burundi, an East African country, is one of the world's poorest countries. The country was severely affected by a civil war from 1993 to 2003. Burundi's total population is approximately 11 million inhabitants. Burundi's first HIV cases were discovered in 1983 (Oxfam International, 2003). As a response to the HIV pandemic, the country's national AIDS program was launched in 1988 and apparently includes prevention, testing, care and treatment activities. HIV services have been progressively decentralized into primary health facilities throughout in the country (ISTEEBU, 2010). HIV/AIDS remains one of the main causes of mortality and also appears to be one of the greatest socio-economic and medical threats in Burundi. Vision Burundi 2025 gives priority to the control of HIV/AIDS, whose sero-prevalence today amounts to the rates of 5% for the urban population, 4% for the semi-urban population and 2.5% for the rural environment (UNDP, 2011). In Burundi, HIV prevalence is 1.4% (ISTEEBU, 2010; Barankanira *et al*., 2016). The main goal of this study is to forecast the number of adults newly infected with HIV in Burundi over the period 2019 – 2030. This study will go a long way in assessing the possibility of ending the HIV pandemic in the country.

2.0 LITERATURE REVIEW

In a Zimbabwean study, Mahomva *et al*. (2006) investigated HIV prevalence in Zimbabwe using data reported from 4 Antenatal Clinic (ANC) surveys conducted between 2000 and 2004, 2 small local studies in Zimbabwe conducted from 1997 through 2003, 4 general population surveys from 1999 through 2003 and service statistics covering 1999 through 2004. The authors discovered that HIV prevalence among pregnant women attending ANCs declined substantially from 32.1% in 2000 to 23.9% in 2004. The study finally concluded that there is increased evidence that a decline in HIV prevalence in Zimbabwe is actually happening in the population. In another Zimbabwean study, Gregson *et al* (2010) examined the contributions of rising mortality, falling HIV incidence and sexual behaviour change to the decline in HIV prevalence in Zimbabwe. The study's methodological technique was hinged on comprehensive review and secondary analysis of national and local sources on trends in HIV prevalence, HIV incidence, mortality and sexual behaviour covering the period 1985 – 2007. The study found out that HIV prevalence in Zimbabwe fell from 29.3% in 1997 to 15.6% in 2007. In a Burundian study, Barankanira *et al*. (2016) investigated the spatial heterogeneity of HIV prevalence in Burundi and then assessed the association of social and behavioral characteristics with HIV infection accounting for the spatial heterogeneity. The researchers used data from the 2010 Demographic and Health Survey. A geospatial approach was applied for data analysis. The study showed that overall HIV prevalence was 1.4%. None of the reviewed studies attempted to forecast new HIV infections in Burundi. Given the fact that HIV in Burundi is more prevalent in adults than children, this study will shade light on the possibility of reasonably controlling the spread of HIV/AIDS amongst adults in the country.

3.0 METHODODOLOGY

3.1 The Box – Jenkins (1970) Methodology

The first step towards model selection is to difference the series in order to achieve stationarity. Once this process is over, the researcher will then examine the correlogram in order to decide on the appropriate orders of the AR and MA components. It is important to highlight the fact that this procedure (of choosing the AR and MA components) is biased towards the use of personal judgement because there are no clear – cut rules on how to decide on the appropriate AR and MA components. Therefore, experience plays a pivotal role in this regard. The next step is the estimation of the tentative model, after which diagnostic testing shall follow. Diagnostic checking is usually done by generating the set of residuals and testing whether they satisfy the characteristics of a white noise process. If not, there would be need for model re –

specification and repetition of the same process; this time from the second stage. The process may go on and on until an appropriate model is identified (Nyoni, 2018c). This approach will be used to analyze the G series under consideration.

3.2 The Moving Average (MA) model

Given:

$$G_t = \sum_{i=1}^{q} \alpha_i \mu_{t-i} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots [1]$$

where μ_t is a purely random process with mean zero and varience σ^2. Equation [1] is reffered to as a Moving Average (MA) process of order q, commonly denoted as MA (q). G is the annual number of adults newly infected with HIV in Burundi at time t, $\alpha_0 \ldots \alpha_q$ are estimation parameters, μ_t is the current error term while $\mu_{t-1} \ldots \mu_{t-q}$ are previous error terms.

3.3 The Autoregressive (AR) model

Given:

$$G_t = \sum_{i=1}^{p} \beta_i G_{t-i} + \mu_t \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots [2]$$

Where $\beta_1 \ldots \beta_p$ are estimation parameters, $G_{t-1} \ldots G_{t-p}$ are previous period values of the G series and μ_t is as previously defined. Equation [2] is an Autoregressive (AR) process of order p, and is usually denoted as AR (p).

3.4 The Autoregressive Moving Average (ARMA) model

An ARMA (p, q) process is just a mere combination of AR (p) and MA (q) processes. Thus, by combining equations [1] and [2]; an ARMA (p, q) process may be specified as shown below:

$$G_t = \sum_{i=1}^{p} \beta_i G_{t-i} + \sum_{i=1}^{q} \alpha_i \mu_{t-i} + \mu_t \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots [3]$$

3.5 The Autoregressive Integrated Moving Average (ARIMA) model

A stochastic process G_t is referred to as an Autoregressive Integrated Moving Average (ARIMA) [p, d, q] process if it is integrated of order "d" [I (d)] and the "d" times differenced process has an ARMA (p, q) representation. If the sequence $\Delta^d G_t$ satisfies an ARMA (p, q) process; then the sequence of G_t also satisfies the ARIMA (p, d, q) process such that:

$$\Delta^d G_t = \sum_{i=1}^{p} \beta_i \Delta^d G_{t-i} + \sum_{i=1}^{q} \alpha_i \mu_{t-i} + \mu_t \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots [4]$$

where Δ is the difference operator, vector $\beta \in \mathbb{R}^p$ and $\alpha \in \mathbb{R}^q$.

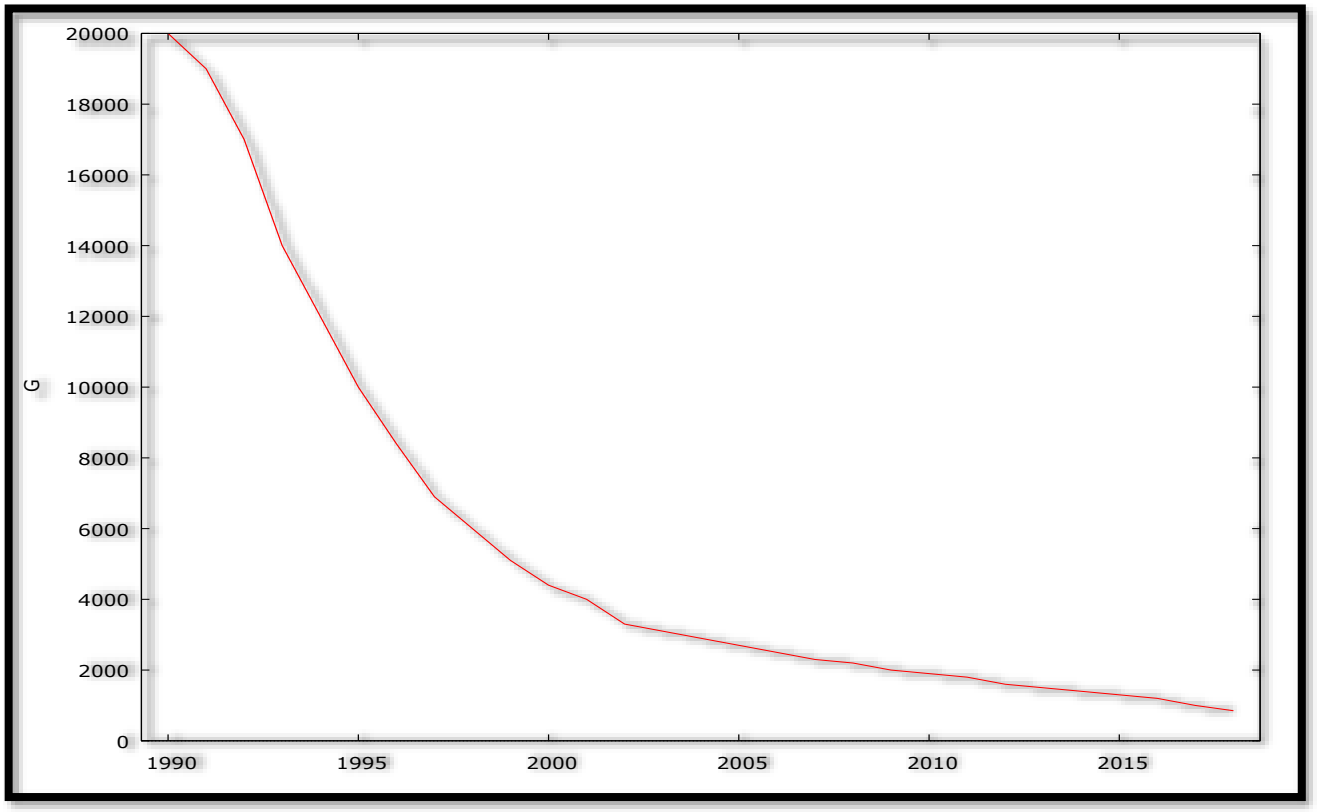
3.6 Data Collection

This study is based on annual observations (that is, from 1990 – 2018) on the number of new HIV infections adults (ages 15 years and above) [denoted as G] in Burundi. Out-of-sample forecasts will cover the period 2019 – 2025. All the data was gathered from the World Bank online database.

3.7 Diagnostic Tests & Model Evaluation

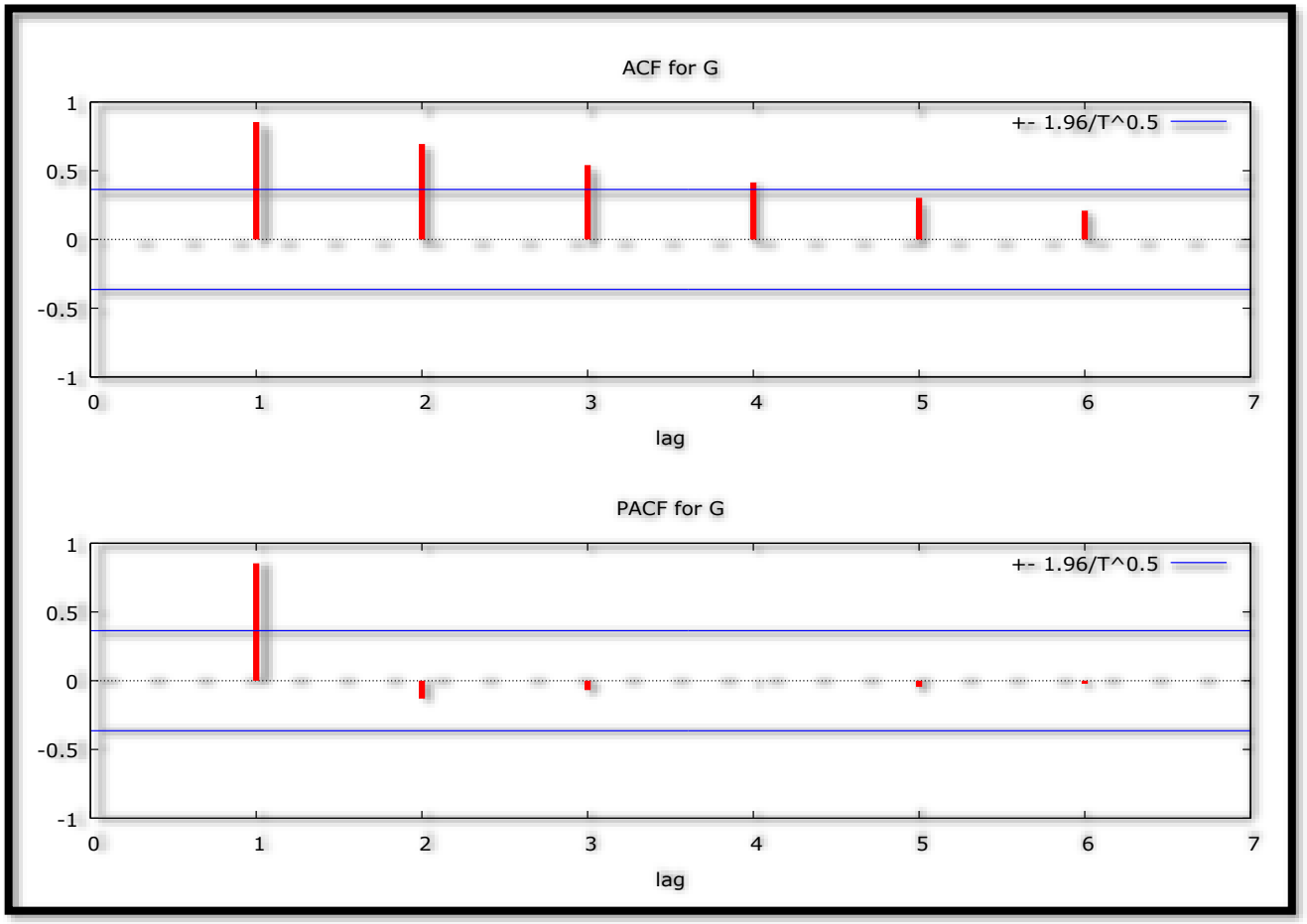
3.7.1 Stationarity Tests: Graphical Analysis

Figure 1



3.7.2 The Correlogram in Levels

Figure 2: Correlogram in Levels



3.7.3 The ADF Test in Levels

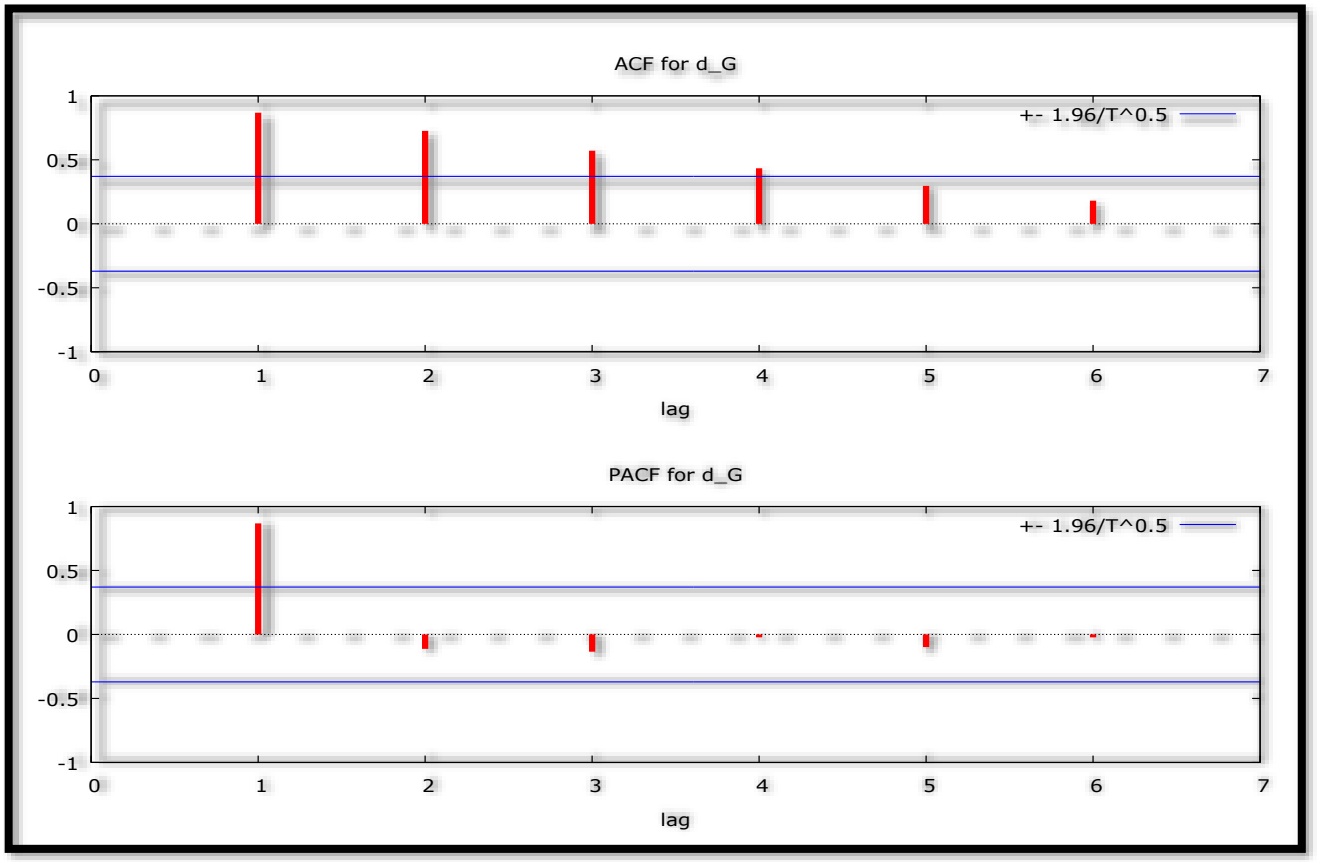
Table 1: without trend and intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
G	-2.209013	0.0291	-2.674290	@1%	Non-stationary
			-1.957204	@5%	Stationary
			-1.608175	@10%	Stationary

Table 1 shows that G is stationary in levels. However, these results are not in line with what is being suggested by figure 1: which shows that the series under consideration is on a monotonically decreasing trajectory. This apparently indicates that the variable G cannot be an I(0) process. We, therefore proceed to test for stationarity at first differences.

3.7.4 The Correlogram (at First Differences)

Figure 3: Correlogram (at First Differences)



3.7.5 The ADF Test (at First Differences)

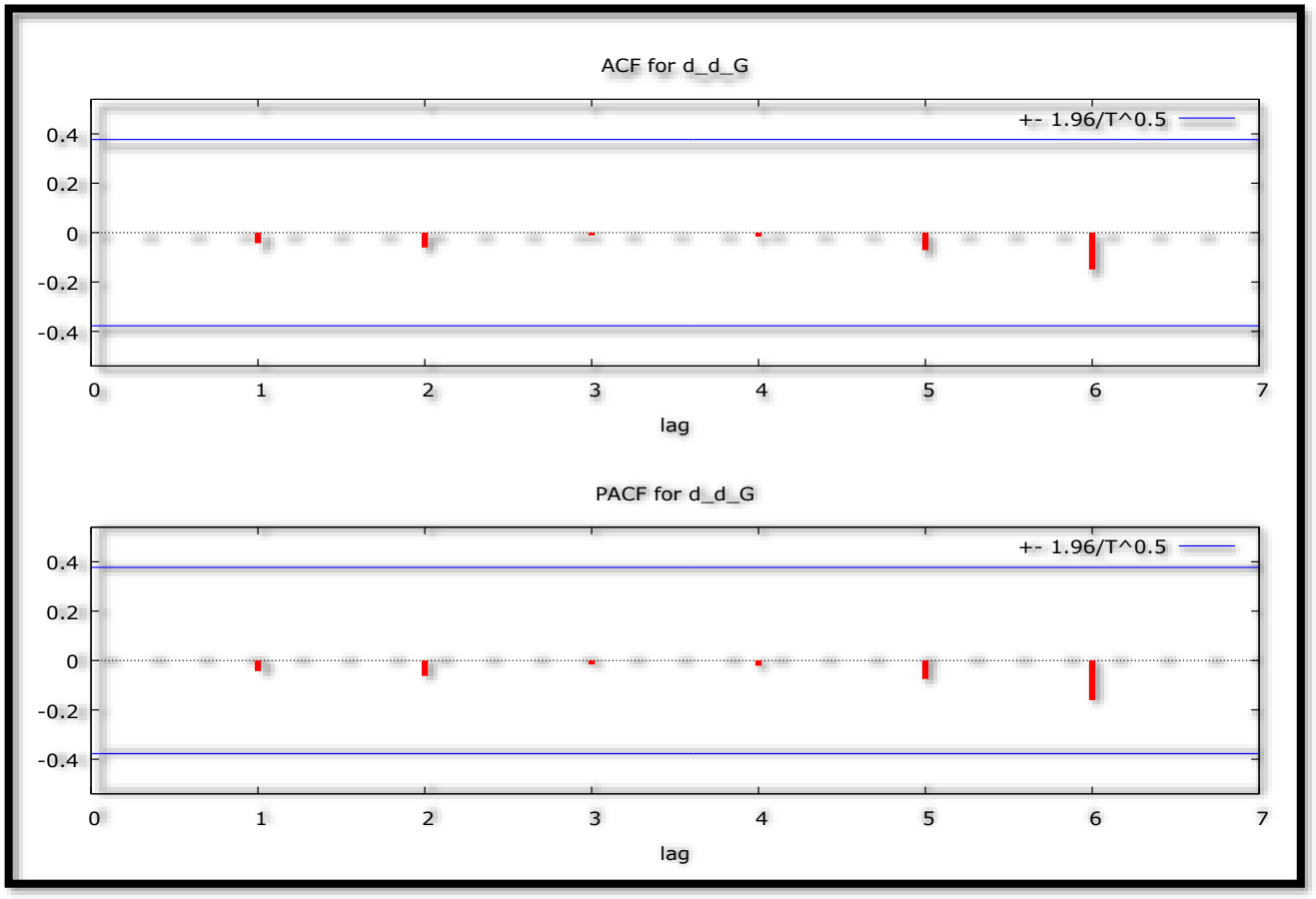
Table 2: without trend and intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
ΔG	-8.963294	0.0000	-2.660720	@1%	Stationary
			-1.955020	@5%	Stationary
			-1.609070	@10%	Stationary

Table 2 shows that G is stationary at first differences. However, figure 3 suggests a possibility of non-stationarity. Therefore, in order to confirm this, the study further tests for stationary at second differences.

3.7.6 The Correlogram (at Second Differences)

Figure 4: Correlogram (at Second Differences)



3.7.7 The ADF Test (at Second Differences)

Table 3: without trend and intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
$\Delta^2 G$	-2.056167	0.0404	-2.664853	@1%	Non-stationary
			-1.955681	@5%	Stationary
			-1.608793	@10%	Stationary

Figure 4 as well as table 3 consistently show that G is an I (2) variable.

3.7.8 Evaluation of ARIMA models (without a constant)

Table 4: Evaluation of ARIMA Models (without a constant)

Model	AIC	U	ME	RMSE	MAPE
ARIMA (1, 2, 0)	402.4750	0.4545	32.543	387.75	4.4074
ARIMA (2, 2, 0)	404.3521	0.46182	37.016	387.8	4.5486
ARIMA (3, 2, 0)	406.3296	0.45418	35.381	388.2	4.546
ARIMA (0, 2, 1)	**402.4713**	0.45382	32.765	387.74	4.412

A model with a lower AIC value is better than the one with a higher AIC value (Nyoni, 2018b) Similarly, the U statistic can be used to find a better model in the sense that it must lie between 0 and 1, of which the closer it is to 0, the better the forecast method (Nyoni, 2018a). In this research paper, only the AIC is used to select the optimal model. Therefore, the ARIMA (0, 2, 1) model is finally selected.

3.8 Residual & Stability Tests

3.8.1 Correlogram of the Residuals of the ARIMA (0, 2, 1) Model

Figure 5: Correlogram of the Residuals

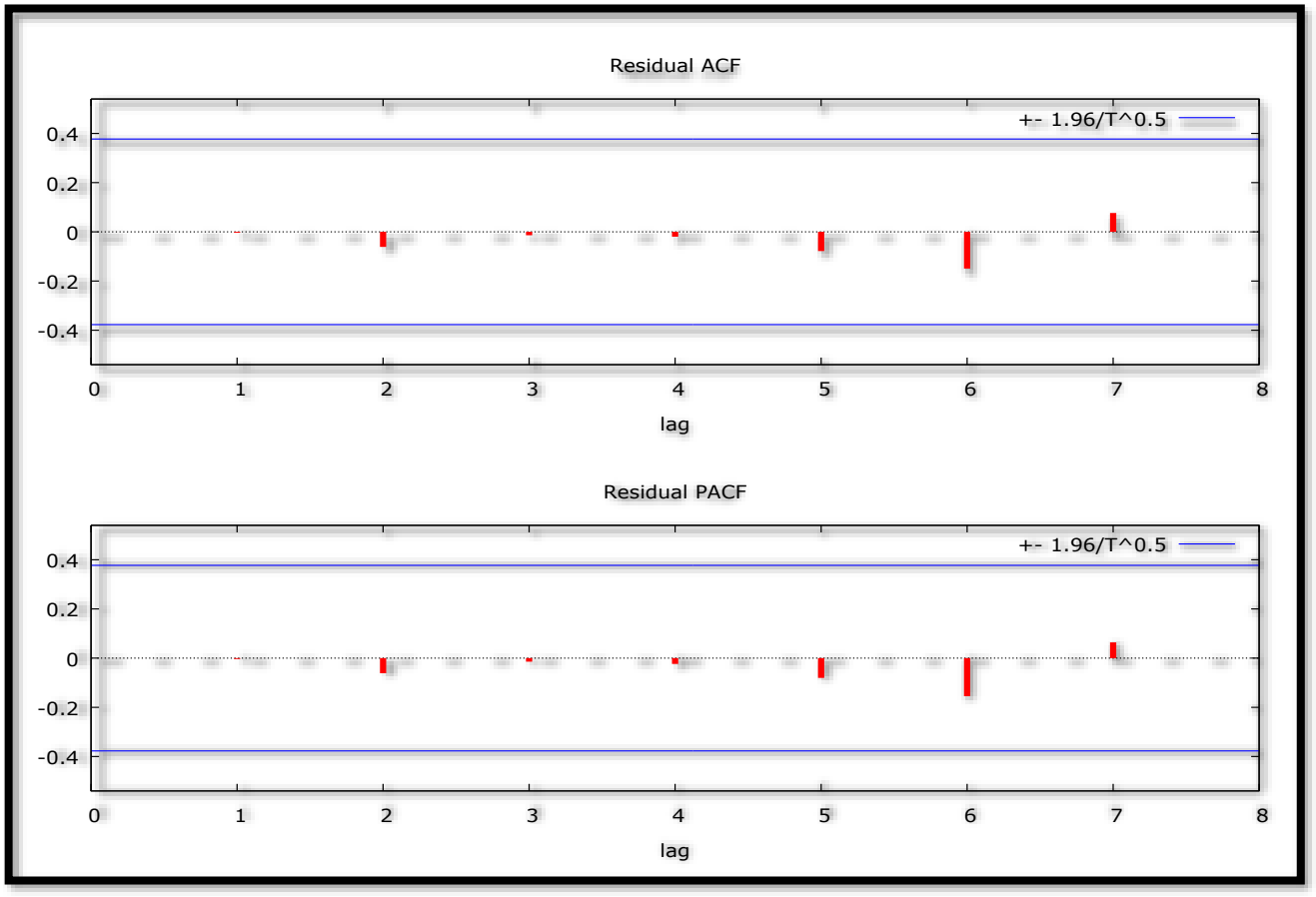
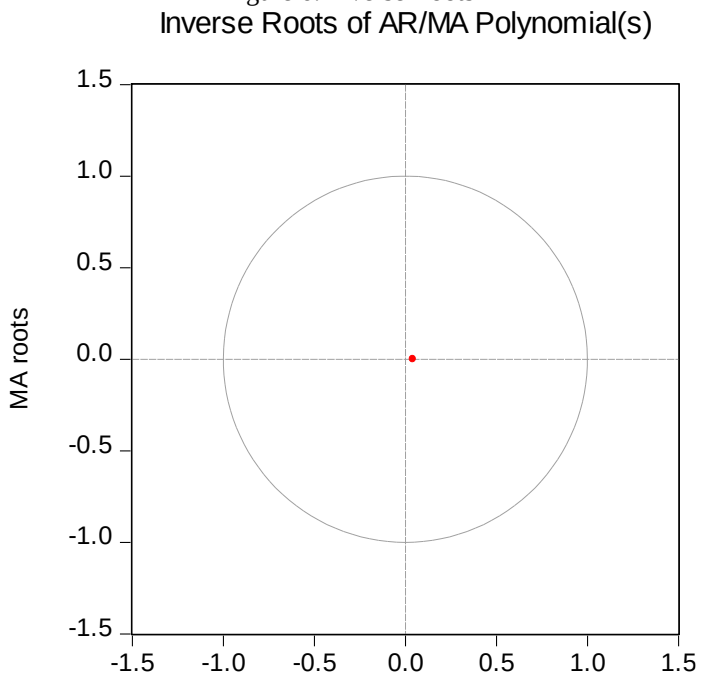


Figure 5 indicates that the estimated ARIMA (0, 2, 1) model is adequate since ACF and PACF lags are quite short and within the bands. This implies that the "no autocorrelation" assumption is not violated in this work.

3.8.2 Stability Test of the ARIMA (0, 2, 1) Model

Figure 6: Inverse Roots



Since all the roots lie inside the unit circle, it implies that the estimated ARIMA process is (covariance) stationary; thus confirming that the ARIMA (0, 2, 1) model is stable and suitable for forecasting annual number of adults newly infected with HIV in Burundi.

4.0 FINDINGS

4.1 Descriptive Statistics

Table 5: Descriptive Statistics

Description	Statistic
Mean	5529.3
Median	2900
Minimum	850
Maximum	20000

As shown in table 5 above, the annual average number of adults newly infected with HIV in Burundi is approximately 5529 people. The minimum number of new infections is 850 while the maximum is 20000. The wide gap between the minimum and the maximum is attributed to the fact that new infections have been sharply decreasing in Burundi since 1990.

4.2 Results Presentation

Table 6: Main Results

ARIMA (0, 2, 1) Model: Guided by equation [4], the chosen optimal model, the ARIMA (0, 2, 1) model can be expressed as follows: $\Delta^2 G_t = -0.0412421\mu_{t-1}$ …………………………………………………[5]				
Variable	Coefficient	Standard Error	z	p-value
μ_1	-0.0412421	0.192460	-0.2143	0.8303

Table 6 shows the main results of the ARIMA (0, 2, 1) model.

Forecast Graph

Figure 7: Forecast Graph – In & Out-of-Sample Forecasts

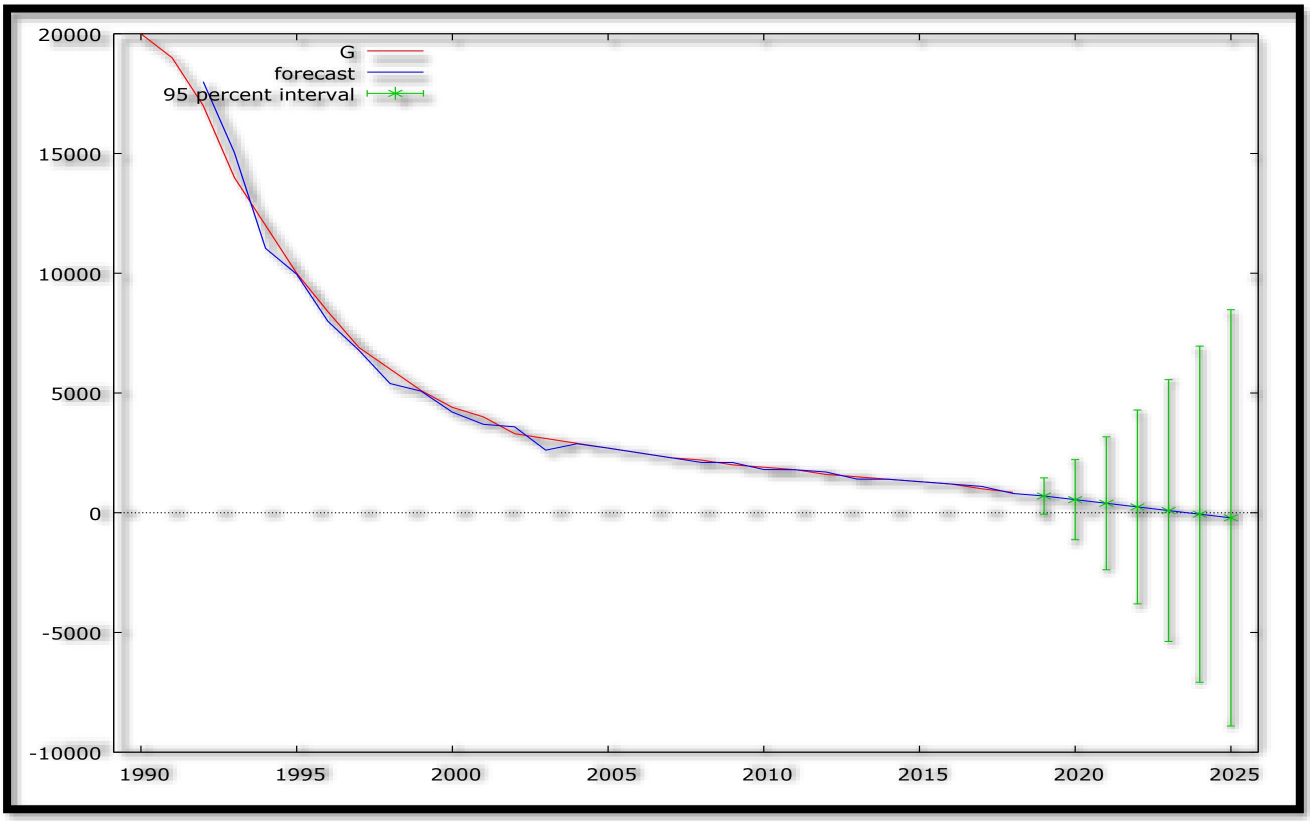


Figure 7 shows the in-and-out-of-sample forecasts of the G series. The out-of-sample forecasts cover the period 2019 – 2025.

Predicted G– Out-of-Sample Forecasts Only

Table 7: Predicted

Year	Prediction	Standard Error	95% Confidence Interval
2019	698.108	387.660	(-61.6916, 1457.91)

2020	546.216	852.564	(-1124.78, 2217.21)
2021	394.324	1416.35	(-2381.66, 3170.31)
2022	242.432	2064.96	(-3804.82, 4289.68)
2023	90.5399	2788.77	(-5375.35, 5556.43)
2024	-61.3521	3580.76	(-7079.52, 6956.81)
2025	-213.244	4435.56	(-8906.78, 8480.29)

Table 7 and figure 7 show the out-of-sample forecasts only. The number of adults newly infected with HIV is projected to continue to decrease in Burundi, from 698 new infections in 2019 to as low as 91 new infections by 2023. This is shows that, in the near future, Burundi can possibly win the war against the HIV pandemic; especially if the country continues to effectively engage on control and preventive measures. Our model shows that Vision Burundi 2025 will be achieved with regards to control and prevention of the HIV/AIDS pandemic in Burundi.

5.0 CONCLUSION

The study shows that the ARIMA (0, 2, 1) model is a stable and suitable model to forecast the annual number of adults newly infected with HIV in Burundi over the period 2019 – 2025. The model predicts a commendable decrease in the annual number of new HIV infections in adults in the country. The study is important for the government of Burundi, especially with regards to the fight against the HIV pandemic. The study recommends that the government of Burundi should continue scaling up HIV prevention and treatment access, especially to the vulnerable groups who were displaced during the civil war time. Special emphasis ought to be directed towards behavior change interventions such as increased condom use and reduction of sexual partners. Burundi, just like any other East African country, is a low-circumcision country; therefore, there is need for up scaling of medical male circumcision as an additional HIV prevention strategy.

REFERENCES

[1] Barankanira, E., *et al*. (2016). Spatial Analysis of HIV Infection and Associated Individual Characteristics in Burundi: Indicators for Effective Prevention, *BMC Public Health*, 16 (118): 1 – 11.

[2] Gregson, S., *et al*. (2010). HIV Decline in Zimbabwe due to Reductions in Risky Sex? Evidence From a Comprehensive Epidemiological Review, *International Journal of Epidemiology*, 39 (2010): 1311 – 1323.

[3] ISTEEBU (2010). ISTEEBU, Bujumbura.

[4] Mahomva, A., *et al*. (2006). HIV Prevalence and Trends From Data in Zimbabwe, 1997 – 2004, *Sexually Transmitted Infections*, 82 (1): 42 – 47.

[5] Nyoni, T (2018b). Modeling and Forecasting Inflation in Kenya: Recent Insights from ARIMA and GARCH analysis, *Dimorian Review*, 5 (6): 16 – 40.

[6] Nyoni, T. (2018a). Modeling and Forecasting Naira/USD Exchange Rate in Nigeria: A Box-Jenkins ARIMA Approach, MPRA Paper No. 88622, *University Library of Munich*, Munich, Germany.

[7] Nyoni, T. (2018c). Box – Jenkins ARIMA Approach to Predicting net FDI inflows in Zimbabwe, MPRA Paper No. 87737, *University Library of Munich*, Munich, Germany.

[8] Oxfam International (2003). A Situation Analysis on the HIV/AIDS Epidemic in Burundi, Gitega, Burundi.

[9] UNDP (2011). Vision Burundi 2025, *UNDP*, Burundi.