NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY [IJIERT]
ISSN: 2394-3696 Website: ijiert.org
VOLUME 7, ISSUE 10, Oct.-2020

ADULTS NEWLY INFECTED WITH HIV IN KENYA: 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 Kenya from 1990 – 2018, the study predicts the annual number of adults who will be newly infected with HIV over the period 2019 – 2030. The study employed the Box-Jenkins ARIMA methodology. The diagnostic ADF tests show that the B series under consideration is an I (0) variable. Based on the AIC, the study presents the ARIMA (2, 0, 4) model as the optimal model. The residual correlogram further reveals that the presented ARIMA (2, 0, 4) model is stable. The results of the study indicate that the number of new HIV infections in adults in Kenya is expected to continue to fall from the estimated 36229 new infections to nearly 4726 new infections by 2030. This would be a significant improvement in Kenya's quest for an AIDS-free society. Amongst other policy conclusions, the study basically encourages the relevant authorities in Kenya to continue scaling up HIV prevention and treatment access; with special emphasis on behavior change interventions such as increased condom use and reduction of sexual partners.

INTRODUCTION
The HIV epidemic in Kenya has evolved, since the first case was diagnosed in 1984 (KNACC, 2014), and has remained a major cause of preventable morbidity and mortality in Kenya (Oluoch et al., 2011; Braithwaite et al., 2014). Kenya has the joint third largest HIV epidemic in the world (alongside Tanzania) with 1.6 million people living with HIV in 2018 (UNAIDS, 2019). In the same year, 25000 people died from AIDS-related illnesses (Avert, 2019). While this is still high, the death rate has declined steadily from 64000 in 2010 (UNAIDS, 2019). The first case of HIV in Kenya was reported in 1984. Almost a decade later, HIV was one of the major causes of illness in the country, putting huge demands on the healthcare system as well as the economy. By 1996, 10.5% Kenyans were already living with HIV. Thanks to the national HIV programmes, prevalence has almost halved since then, standing at 5.9% by 2015. This positive public health outcome is mainly due to the rapid scaling up of HIV treatment and care (KNACC, 2014). In 2016 alone, 64% of people living with HIV were on treatment, 51% of who were virally suppressed (UNAIDS, 2017). Kenya's HIV epidemic is driven by sexual transmission; over 90% of all HIV infections are due to sexual transmission (Montana et al., 2008) and is generalized, meaning it affects all sections of the population including children, young people, adults, women and men. As of 2015, 660000 children were recorded as being orphaned by HIV/AIDS (UNAIDS, 2015). The major goal of this study is to predict the number of adults newly infected with HIV in Kenya over the period 2019 – 2030. This study will go a long way in assessing the possibility of ending the HIV epidemic in the country.

LITERATURE REVIEW
Braithwaite et al. (2014) estimated the portion of HIV infections attributable to unhealthy alcohol use and also evaluated the impact of hypothetical interventions and directed at unhealthy alcohol use on HIV infections and deaths. The study was based on a transmission and progression simulation model. The study indicated that the effects of behaviors accompanying unhealthy alcohol consumption are responsible for 13% of new HIV infections in Kenya. Employing a 2-stage cluster sampling survey and covering the period October 2012 to February 2013, Kimanga et al. (2016) analyzed HIV prevalence and incidence among adults aged 15-64 in Kenya. The study found out that HIV prevalence has declined in Kenya. In a recent Kenyan study, Kiplagat et al. (2018) examined the effect of age on retention post ART initiation. The study applied the logistic and cox regression models as well as Pearson's Chi-square and Mann-Whitney U tests.

The study concluded that a higher proportion of older adults was initiated on ART and have better retention in care at 12, 24 and 36 months post ART initiation than younger adults. To the best of our knowledge no study has attempted to forecast new HIV infections in Kenya. We chose to focus on new HIV infections on adults because most HIV infections have been detected in this group as compared to others groups.

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 B series under consideration.

3.2 The Moving Average (MA) model

Given:

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

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). B is the annual number of adults newly infected with HIV in Kenya 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:

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

Where $\beta_1 \ldots \beta_p$ are estimation parameters, $B_{t-1} \ldots B_{t-p}$ are previous period values of the B 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:

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

3.5 The Autoregressive Integrated Moving Average (ARIMA) model

A stochastic process B_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 $\Delta^d B_t$ satisfies an ARMA (p, q) process; then the sequence of B_t also satisfies the ARIMA (p, d, q) process such that:

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

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 in adults (ages 15 years and above) [denoted as B] in Kenya. Out-of-sample forecasts will cover the period 2019 – 2030. All the data was gathered from the World Bank online database.

3.7 Diagnostic Tests & Model Evaluation

3.7.1 The ADF Test in Levels

Table 1: without trend and intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
B	-3.757453	0.0006	-2.664853	@1%	Stationary
			-1.955681	@5%	Stationary
			-1.608793	@10%	Stationary

Table 1 shows that B is stationary in levels.

3.7.2 Evaluation of ARIMA models (without a constant)

Table 2: Evaluation of ARIMA Models (without a constant)

Model	AIC	U	ME	RMSE	MAPE
ARIMA (1, 0, 4)	600.3988	0.64167	5583.2	32317	7.4034
ARIMA (1, 0, 3)	603.3370	0.67157	4927.4	32313	7.9088
ARIMA (1, 0, 2)	603.8861	0.6373	4469.7	32353	7.5575
ARIMA (1, 0, 1)	619.3844	0.70719	3414	32706	7.7339
ARIMA (1, 0, 0)	630.9828	0.96734	1675.4	33317	9.9957
ARIMA (4, 0, 1)	598.8631	0.63683	6118.7	32339	7.3454
ARIMA (4, 0, 3)	598.1360	0.67728	6270.9	32324	7.6961
ARIMA (4, 0, 2)	597.3773	0.68907	6157.4	32341	7.8578
ARIMA (2, 0, 4)	**595.7255**	0.62242	7270	32279	7.4223
ARIMA (4, 0, 4)	598.7411	0.62343	6470.6	32312	7.3345
ARIMA (2, 0, 2)	601.0840	0.62134	6069.6	32312	7.254
ARIMA (3, 0, 1)	609.0042	0.66964	7484.8	32391	7.6097
ARIMA (2, 0, 1)	609.3800	0.5988	6762.2	32407	6.7758
ARIMA (3, 0, 4)	597.1279	0.60175	6490.9	32270	6.9077
ARIMA (4, 0, 0)	597.2758	0.6354	6189.4	32.338	7.2539

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 (2, 0, 4) model is finally chosen. It is imperative to note that the selected model is equivalent to (or the same as) the ARMA (2, 4) model. However, for consistency purposes, throughout the study, we choose to describe the model as ARIMA (2, 0, 4) model.

3.8 Residual Tests

3.8.1 Correlogram of the Residuals of the ARIMA (2, 0, 4) Model

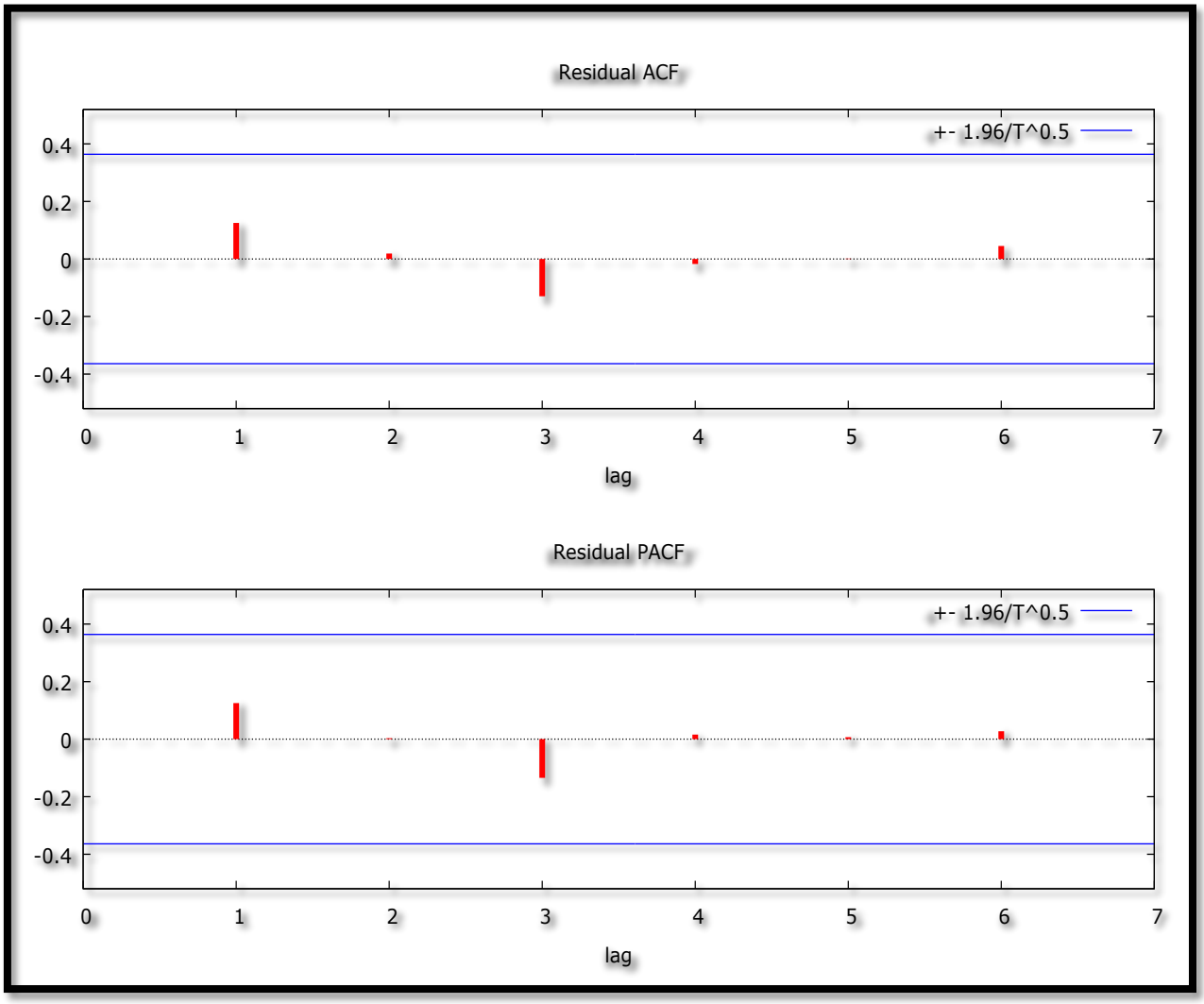

Figure 1: Correlogram of the Residuals

Figure 1 indicates that the estimated optimal 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 study.

FINDINGS

4.1 Descriptive Statistics

Table 3: Descriptive Statistics

Description	Statistic
Mean	89517
Median	65000
Minimum	38000
Maximum	210000

Over the study period, the annual average number new HIV infections in Kenya was 89517. The minimum number of new HIV infections was 38000 while the maximum was 210000. The monotonic decrease in the number of new infections from as high as 210000 to as low as 38000, points to the effectiveness of HIV/AIDS intervention programmes in Kenya.

4.2 Results Presentation

Table 4: Main Results

ARIMA (2, 0, 4) Model:

Guided by equation [4], the chosen optimal model, the ARIMA (2, 0, 4) model can be mathematically expressed as follows:

$$B_t = 1.77663B_{t-1} - 0.818512B_{t-2} - 0.0590467\mu_{t-1} + 1.16136\mu_{t-2} - 0.0590477\mu_{t-3} + 0.999995\mu_{t-4} \ldots\ldots\ldots [5]$$

Variable	Coefficient	Standard Error	z	p-value
β_1	1.77663	0.131523	13.51	0.0000***
β_2	-0.818512	0.134086	-6.104	0.0000***
α_1	-0.0590467	0.251987	-0.2343	0.8147
α_2	1.16136	0.279969	4.148	0.0003***
α_3	-0.0590477	0.247038	-0.2390	0.8111
α_4	0.999995	0.411025	2.433	0.0150**

Table 4 shows the main results of the ARIMA (2, 0, 4) model.

Forecast Graph

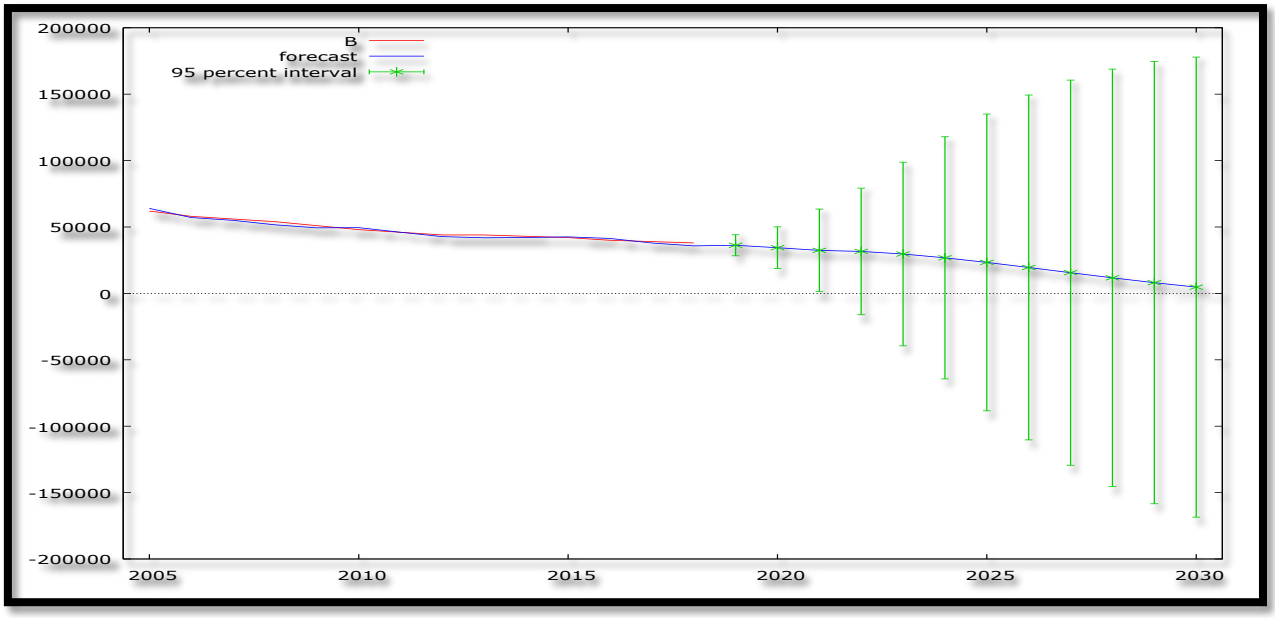

Figure 2: Forecast Graph – In & Out-of-Sample Forecasts

Figure 2 shows the in-and-out-of-sample forecasts of the B series. The out-of-sample forecasts cover the period 2019 – 2030.

Predicted B– Out-of-Sample Forecasts Only

Table 5: Predicted

Year	Predicted B	Standard Error	95% Confidence Interval
2019	36228.7	4035.35	(28319.6, 44137.8)
2020	34395.3	8020.20	(18676.0, 50114.6)
2021	32388.3	15872.7	(1278.31, 63498.3)
2022	31601.7	24318.4	(-16061.5, 79264.9)
2023	29634.4	35277.6	(-39508.4, 98777.3)
2024	26783.1	46519.4	(-64393.3, 117960)
2025	23327.7	57010.5	(-88410.7, 135066)
2026	19522.4	66241.8	(-110309, 149354)
2027	15590.2	73985.4	(-129419, 160599)
2028	11718.7	80192.4	(-145456, 168893)
2029	8059.12	84936.3	(-158413, 174531)
2030	4726.19	88372.8	(-168481, 177934)

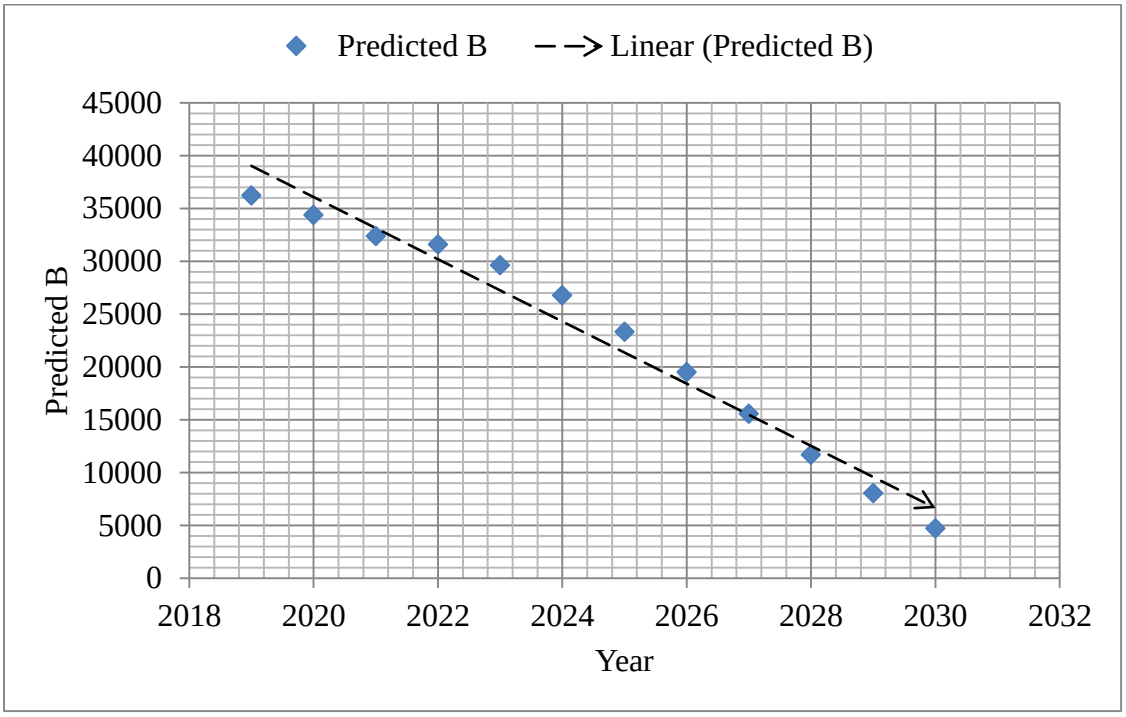

Figure 3: Graphical Analysis of Out-of-Sample Forecasts

Table 5 and figure 3 show the out-of-sample forecasts only. The number of new HIV infections in adults in Kenya is projected to fall from approximately 36229 new infections in 2019 to 4726 new infections by the year 2030. The results are consistent with previous studies such as Kimanga et al. (2016) and the KNACC (2018) which have already reported that HIV prevalence is on a downward trajectory in Kenya.

CONCLUSION

The study shows that the ARIMA (2, 0, 4) model is not only stable but also the most suitable model to forecast the annual number of new HIV infections in Kenya over the period 2019 – 2030. The model predicts a commendable monotonic decrease in the annual number new HIV infections in adults in the country. The study recommends that the government of Kenya should continue scaling up HIV prevention and treatment access; with special emphasis on behavior change interventions such as increased condom use and reduction of sexual partners as well as desisting from alcohol & drug abuse. Moreso, there is need for up scaling of medical male circumcision as an additional HIV prevention strategy in Kenya.

REFERENCES

1) Avert (2019). HIV and AIDS in Kenya, Avert, Nairobi.
2) Braithwaite, R. S., et al. (2014). Impact of Interventions Targeting Unhealthy Alcohol Use in Kenya on HIV Transmission and AIDS-related Deaths, Alcoholism: Clinical and Experimental Research, 38 (4): 1059 – 1067.
3) Kimanga, D. O., et al. (2016). Prevalence and Incidence of HIV Infection, Trends, and Risk Factors Among Persons Aged 15-64 Years in Kenya: Results From a Nationally Representative Study, Journal of Acquired Immune Deficiency Syndrome, 66 (1): 1 – 25.
4) Kiplagat, J., et al. (2018). Retention in Care Among Older Adults Living With HIV in Western Kenya: A Retrospective Observational Cohort Study, PLoS ONE, 13 (3): 1 – 12.
5) KNACC (2014). Kenya AIDS Strategic Framework 2014/2015 – 2018/2019, KNACC, Nairobi.
6) KNACC (2018). Kenya HIV Estimates, KNACC, Nairobi.

7) Montana, L., Mishra, V., & Hong, R., et al. (2008). Comparison of HIV Prevalence Estimates From Antenatal Care Surveillance and Population-based Surveys in Sub-Saharan Africa, Sexually Transmitted Infections, 84 (1): 78 – 84.
8) Nyoni, T (2018b). Modeling and Forecasting Inflation in Kenya: Recent Insights from ARIMA and GARCH analysis, Dimorian Review, 5 (6): 16 – 40.
9) 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.
10) Nyoni, T. (2018c). Box – Jenkins ARIMA Approach to Predicting net FDI inflows in Zimbabwe, MPRA Paper No. 87737, University Library of Munich, Munich, Germany.
11) Oluoch, T., et al. (2011). Correlates of HIV Infection Among Sexually Active Adults in Kenya: A National Population-based Survey, The Open AIDS Journal, 5: 125 – 134.
12) UNAIDS (2015). HIV/AIDS Estimates, UNAIDS, New York.
13) UNAIDS (2017). UNAIDS Information, UNAIDS, New York.
14) UNAIDS (2019). AIDS Information, UNAIDS, New York.