OPEN DEFECATION 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 people who practice open defecation in Kenya from 2000 – 2017, the study predicts the annual number of people who will still be practicing open defecation over the period 2018 – 2021. The authors apply the Box-Jenkins ARIMA methodology. The diagnostic ADF tests show that the ODK series under consideration is an I (1) variable. Based on the AIC, the study presents the ARIMA (0, 1, 2) model as the optimal model. The diagnostic tests further indicate that the presented model is quite stable and its residuals are stationary in levels. The results of the study indicate that the number of people practicing open defecation in Kenya is likely to decline, although slightly, over the period 2018 – 2022, from approximately 9.9% to almost 8.2% of the total population. Hence, it is possible for Kenya to completely eliminate the practice of open defecation by 2030. The study basically suggested a 3-fold policy recommendation to be put into consideration, especially by the government of Kenya.

1.0 INTRODUCTION

Open defecation can be defined as defecation in the fields, bushes, and bodies of water or other open spaces (UN, 2015). Globally, over a billion people defecate in the open (*ibid*). Approximately 215 million in Sub-Saharan Africa (SSA) are open defecators (Galan *et al*., 2013). In the case of Kenya, at least 5.6 million people are open defecators (WSP, 2014) and this simply implies that Kenya is still facing major sanitation challenges, especially in rural areas (Busienei *et al*., 2019). There is need to end open defecation due to its negative impacts, especially on human beings (Desai *et al*., 2015). Such negative impacts include the spread of bacterial, viral and parasitic infections including diarrhoea, polio, cholera, soil-transmitted helminth, trachoma infection, schistosomiasis and hookworm as well as child stunting (Megersa *et al*., 2019) and deaths (Thiga & Cholo, 2017). Kenya's policy on sanitation aims to achieve and sustain Open Defecation Free (ODF) status in the entire country by 2030 (Njuguna, 2019). Eliminating open defecation is increasingly seen as a key health outcome (Okullo *et al*., 2017; UNICEF, 2018; Busienei *et al*., 2019). Therefore, it has become fundamental for public health researchers and policy makers to model and forecast the number of people practicing open defecation in order to formulate evidence-driven policies to end open defecation. The main goal of this study is to predict the annual number of open defecators in Kenya over the period 2018 – 2021. This study, besides being the first of its kind in the case of Kenya, will go a long way in uncovering the possibility of ending open defecation in the country.

1.2 OBJECTIVES OF THE STUDY

i. To investigate the years during which open defection was practiced by people more than 10% of the total population in Kenya.
ii. To forecast the number of people practicing open defecation in Kenya for the period 2018 – 2021.
iii. To examine the trend of open defecation in Kenya for the out-of-sample period.

2.0 LITERATURE REVIEW

Njuguna & Muruka (2017) looked at open defecation trends among the 47 counties in, newly created in 2013 in Kenya. The study made use of four data sets on open defecation, unimproved water supply coverage, poverty levels and population density. Their findings basically show that the average open defecation rate across the 47 counties was 23.5% and the median rate was 6.9% and also that poverty was the most significant predictor accounting for 68.4% of the variance in open defecation after controlling for unimproved water supply and population density. Consistently, Thiga & Cholo (2017) examined open defecation among residents of Thika East Sub-County in Kenya. The study made use of a descriptive cross-sectional design in which 20554 households were included. The study concluded that 23.3% of the sampled homesteads did not have latrines and that members of these households were defecating in the fields, neighbor latrines or public toilets. The study also concluded that open defecation was a predominant norm practiced in most of the communities and it had negative effects on human health, water and air pollution. In a recent Kenya study, Njuguna (2019) sought to explore progress made in attaining sustainable goal number 6 at the household level with a focus on poor households. Kenya demographic and health survey for 2003, 2008 and 2014 respectively were analyzed. Descriptive analysis and bivariate logistic regression was done with open defecation status as the dependent variable. Independent variables were poverty status, place of residence, region where household was located, absence of farm animals, gender and education level of household head. The results of the study basically indicate that, in Kenya, the burden of open defecation has increased among poor households, more so amongst the poorest. No study has been done to forecast the number of open defecators in Kenya. This study is the first of its kind in the case of Kenya and is expected to speed-up the elimination of open defecation in Kenya.

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 ODK series under consideration.

3.2 The Moving Average (MA) model

Given:

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

where μ_t is a pure random process with mean zero and varience σ^2. Equation [1] is called a Moving Average (MA) process of order q, usually denoted as MA (q). ODK is the annual number of people (as a percentage of the total population) who practice open defecation 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:

$$\mathrm{ODK}_t = \sum_{i=1}^{p} \beta_i \mathrm{ODK}_{t-i} + \mu_t \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots [2]$$

Where $\beta_1 \ldots \beta_p$ are estimation parameters, $\mathrm{ODK}_{t-1} \ldots \mathrm{ODK}_{t-p}$ are previous period values of the ODK 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 merely a 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:

$$\mathrm{ODK}_t = \sum_{i=1}^{p} \beta_i \mathrm{ODK}_{t-i} + \sum_{i=1}^{q} \alpha_i \mu_{t-i} + \mu_t \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots [3]$$

3.5 The Autoregressive Integrated Moving Average (ARIMA) model

A stochastic process ODK_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 $\Delta^d \mathrm{ODK}_t$ satisfies an ARMA (p, q) process; then the sequence of ODK_t also satisfies the ARIMA (p, d, q) process such that:

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

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

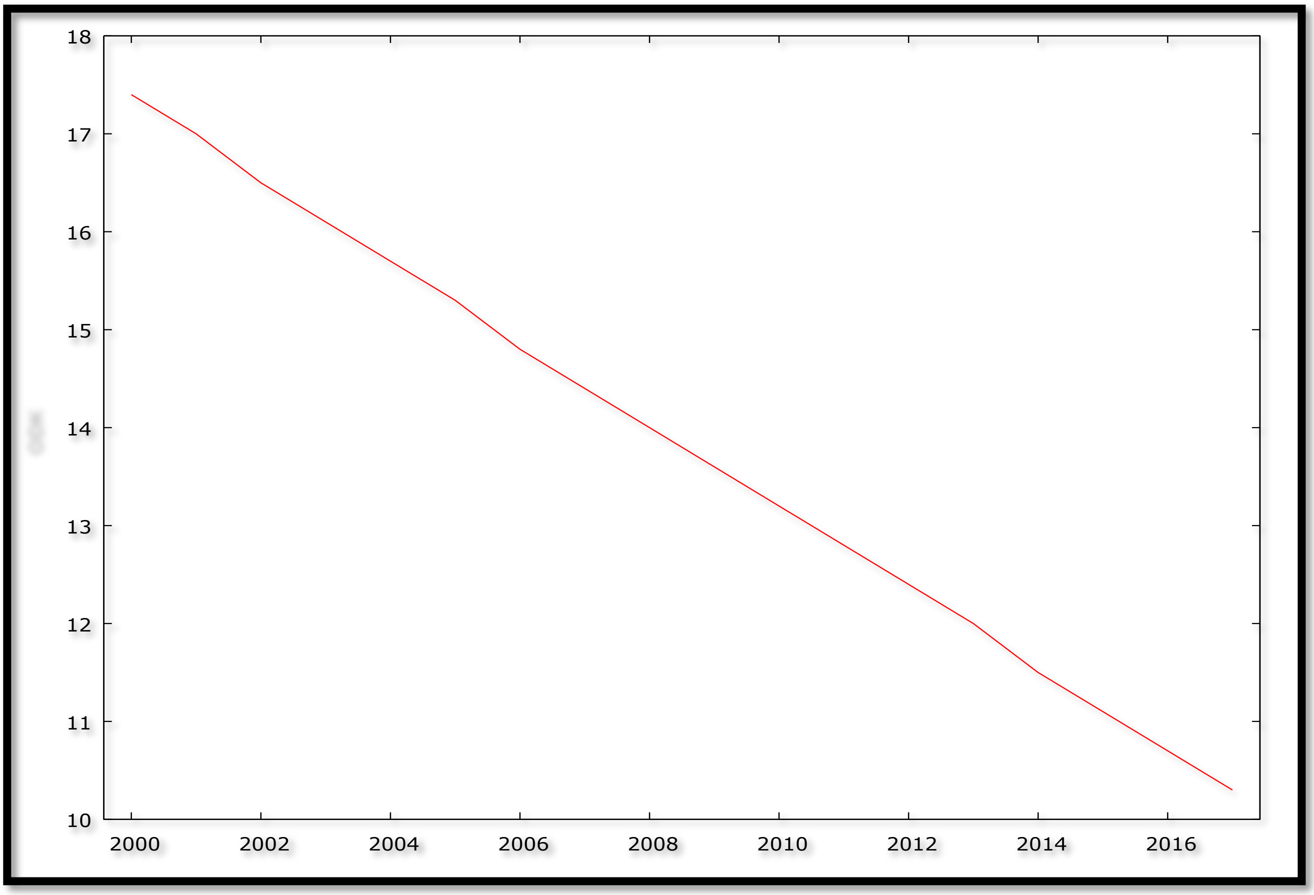
3.6 Data Collection

This study is based on annual observations (that is, from 2000 – 2017) on the number of people practicing Open Defecation [OD, denoted as ODK] (as a percentage of total population) in Kenya. Out-of-sample forecasts will cover the period 2018 – 2021. All the data was gathered from the World Bank online database.

3.7 Diagnostic Tests & Model Evaluation

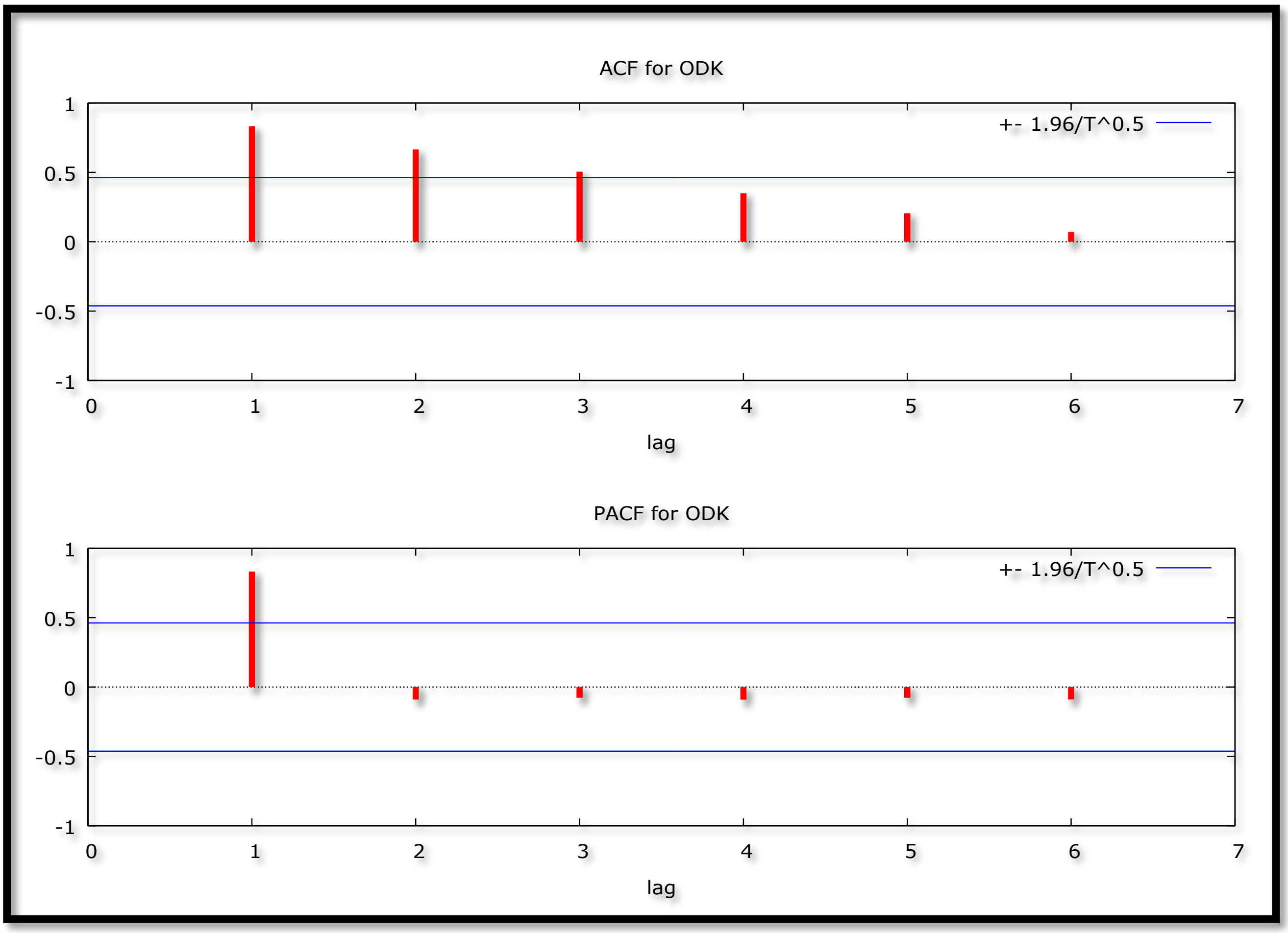
3.7.1 Stationarity Tests: Graphical Analysis

Figure 1



3.7.2 The Correlogram in Levels

Figure 2: Correlogram in Levels



3.7.3 The ADF Test in Levels

Table 1: with intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
ODK	-0.656297	0.8325	-3.886751	@1%	Non-stationary
			-3.052169	@5%	Non-stationary
			-2.666593	@10%	Non-stationary

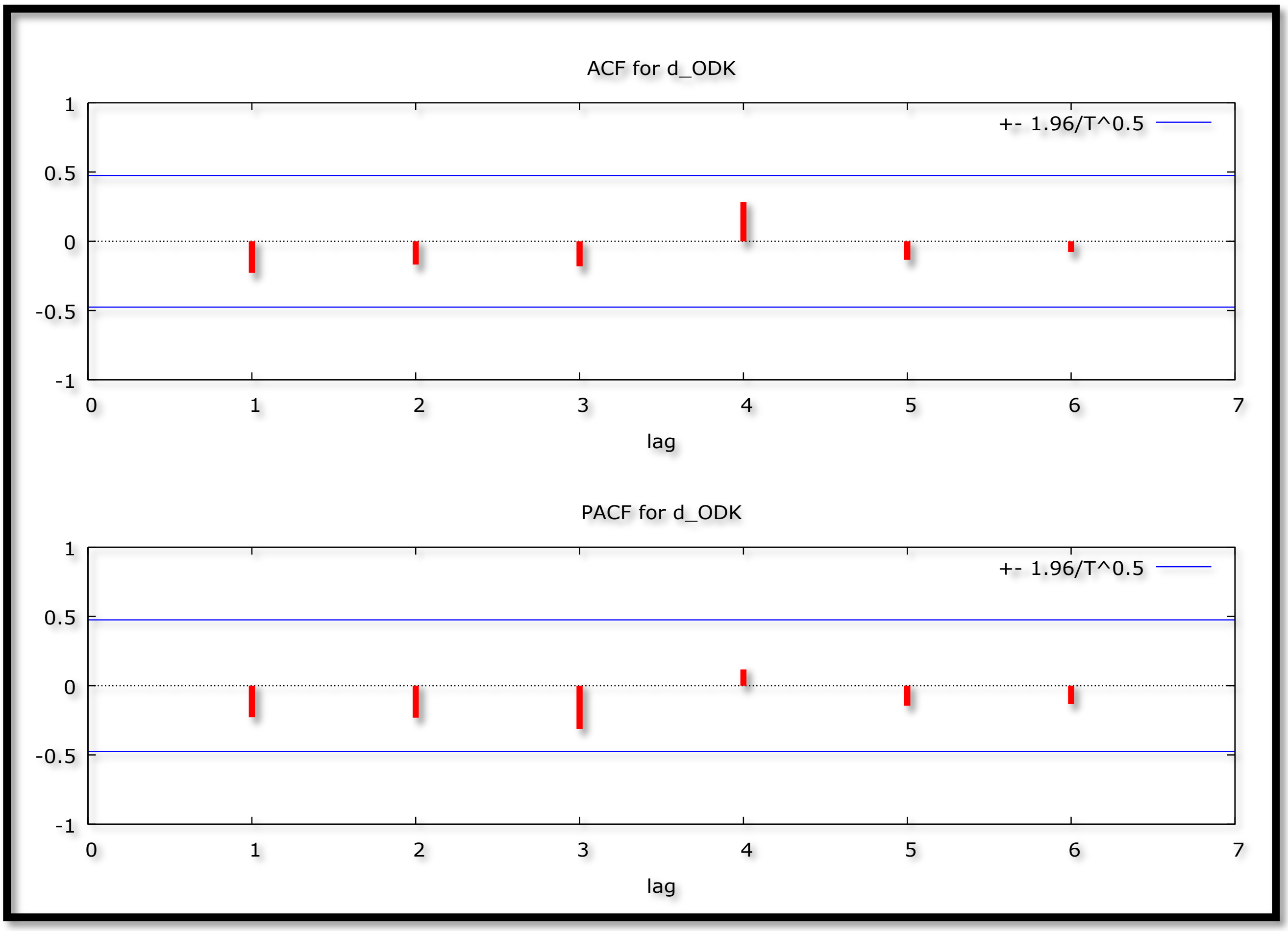
Table 2: with intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
ODK	-2.812499	0.2119	-4.616209	@1%	Non-stationary
			-3.710482	@5%	Non-stationary
			-3.297799	@10%	Non-stationary

Tables 1 and 2 show that ODK is not stationary in levels as already suggested by figures 1 and 2.

3.7.4 The Correlogram (at First Differences)

Figure 3: Correlogram (at First Differences)



3.7.5 The ADF Test (at First Differences)

Table 3: with intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
ΔODK	-4.732864	0.0021	-3.920350	@1%	Stationary
			-3.065585	@5%	Stationary
			-2.673459	@10%	Stationary

Table 4: with intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
ΔODK	-4.811247	0.0078	-4.667883	@1%	Stationary
			-3.733200	@5%	Stationary
			-3.310349	@10%	Stationary

Figure 3 as well as tables 3 and 4, indicate that ODK is an I (1) variable.

3.7.6 Evaluation of ARIMA models (with a constant)

Table 5: Evaluation of ARIMA Models (with a constant)

Model	AIC	U	ME	RMSE	MAPE
ARIMA (1, 1, 0)	-57.70539	0.086961	-0.00023356	0.037114	0.20623
ARIMA (2, 1, 0)	-56.77865	0.084765	-0.00081816	0.036195	0.18946
ARIMA (3, 1, 0)	-57.01947	0.079688	-0.0018685	0.03424	0.16174
ARIMA (4, 1, 0)	-55.45155	0.078919	-0.0014861	0.033461	0.15479
ARIMA (5, 1, 0)	-54.41204	0.077605	-0.0018598	0.033021	0.1599
ARIMA (0, 1, 1)	-58.71106	0.084355	-0.0010757	0.036048	0.22115
ARIMA (0, 1, 2)	**-59.07730**	0.07784	-0.0062055	0.034348	0.17929
ARIMA (0, 1, 3)	-57.39662	0.076915	-0.0051821	0.033892	0.16925
ARIMA (0, 1, 4)	-56.44597	0.07564	-0.0063451	0.03307	0.17613
ARIMA (0, 1, 5)	-54.64374	0.075523	-0.0047748	0.032763	0.17175

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, 1, 2) model is finally chosen.

3.8 Residual & Stability Tests

3.8.1 ADF Test (*in levels*) of the Residuals of the ARIMA (0, 1, 2) Model

Table 6: with intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
R	-3.735671	0.0142	-3.920350	@1%	Non-stationary
			-3.065585	@5%	Stationary
			-2.673459	@10%	Stationary

Table 7: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
R	-4.513944	0.0131	-4.667883	@1%	Non-stationary
			-3.733200	@5%	Stationary
			-3.310349	@10%	Stationary

Tables 6 and 7 indicate that the residuals of the chosen optimal model, the ARIMA (0, 1, 2) model; are stationary. Hence, the model is stable.

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

Figure 4: Correlogram of the Residuals

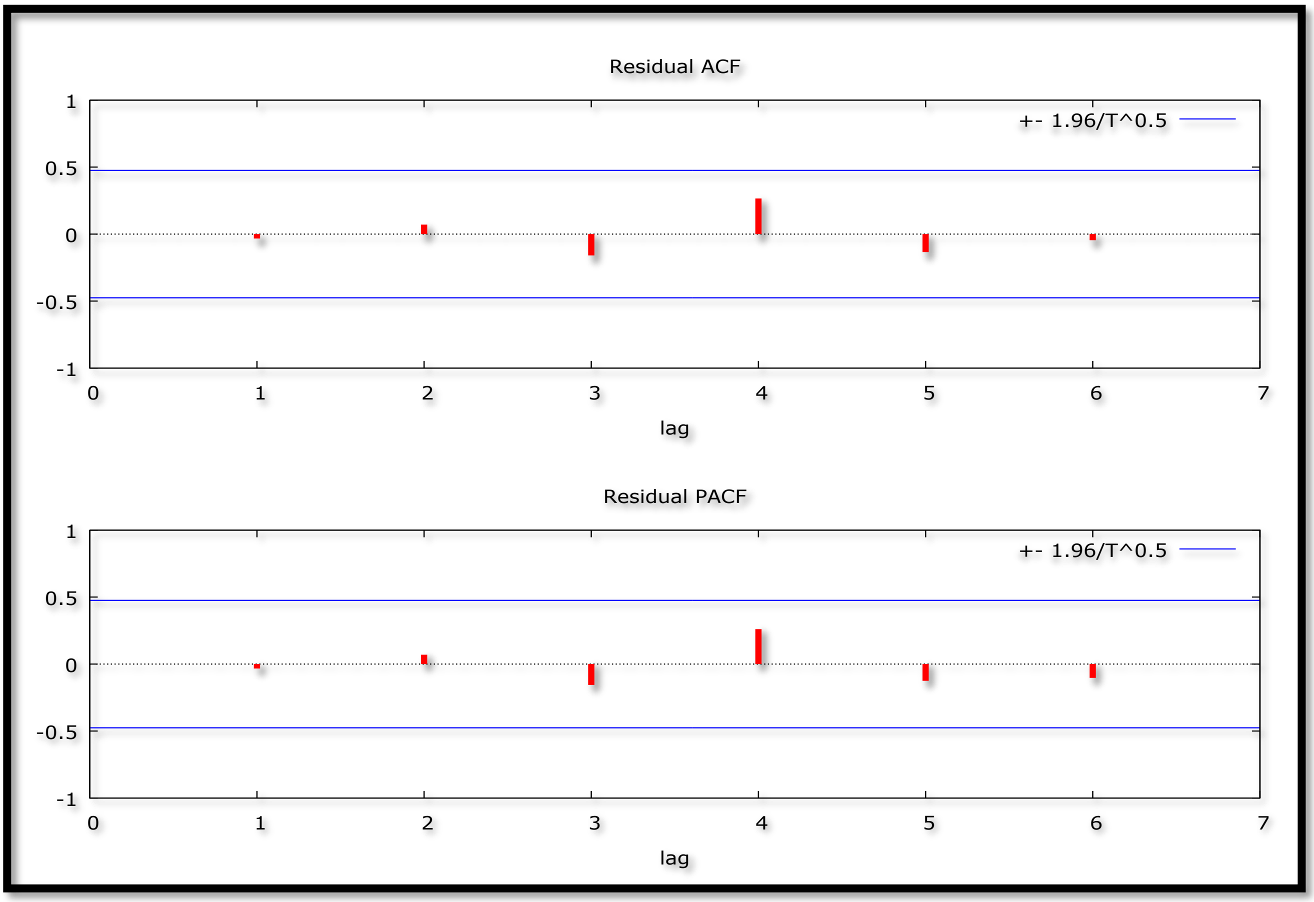
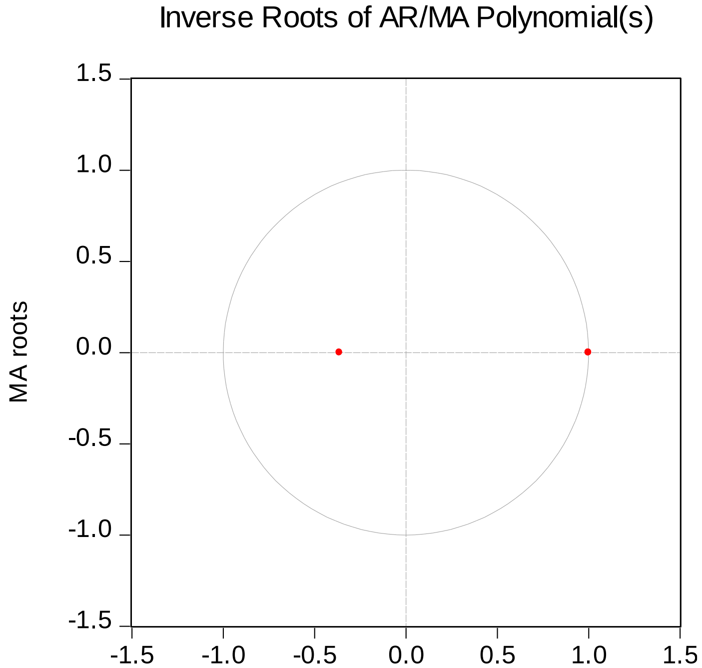


Figure 4 indicates that the estimated model is adequate since ACF and PACF lags are quite short and within the bands. This implies that the "no autocorrelation" assumption is not violated in this study.

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

Figure 5: Inverse Roots



Since all the AR roots lie inside the unit circle, it implies that the estimated ARIMA process is (covariance) stationary; thus confirming that the ARIMA (0, 1, 2) model is stable and suitable for forecasting annual number of people practicing open defecation in Kenya.

4.0 FINDINGS

4.1 Descriptive Statistics

Table 8: Descriptive Statistics

Description	Statistic
Mean	13.822
Median	13.8
Minimum	10.3
Maximum	17.4

As shown in table 8 above, the mean is positive, that is, 13.822. This means that, over the study period, the annual average number of people practicing open defecation in Kenya is approximately 14% of the total population. The minimum number of people practicing open defecation in Kenya over the study period is approximately 10.3% of the total population, while the maximum is 17.4% of the total population.

4.2 Results Presentation[1]

Table 9: Main Results

ARIMA (0, 1, 2) Model:

Guided by equation [4], the chosen optimal model, the ARIMA (0, 1, 2) model can be expressed as follows:

$$\Delta ODK_t = -0.416484 - 0.627602\mu_{t-1} - 0.372398\mu_{t-2} \ldots \ldots \ldots \ldots \ldots [5]$$

Variable	Coefficient	Standard Error	z	p-value
constant	-0.416484	0.00187130	-222.6	0.0000***
α_1	-0.627602	0.274501	-2.286	0.0222**
α_2	-0.372398	0.202572	-1.838	0.0660*

Table 9 shows the main results of the ARIMA (0, 1, 2) model.

Forecast Graph

Figure 6: Forecast Graph – In & Out-of-Sample Forecasts

[1] *The *, ** and *** imply statistical significance at 10%, 5% and 1% levels of significance; respectively.*

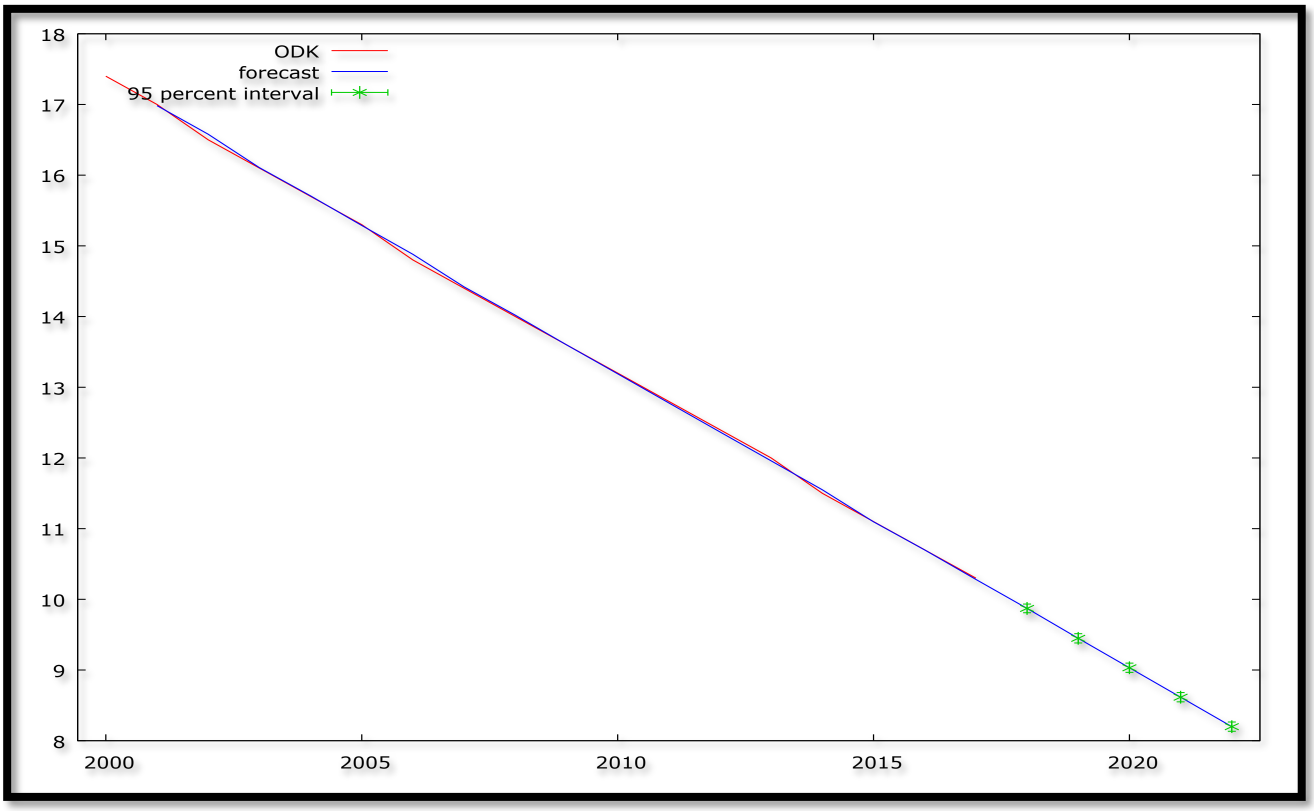


Figure 6 shows the in-and-out-of-sample forecasts of the ODK series. The out-of-sample forecasts cover the period 2018 – 2022.

Predicted ODK – Out-of-Sample Forecasts Only

Table 10: Predicted

Year	Predicted ODK	Standard Error	Lower Limit	Upper Limit
2018	9.9	0.03	9.8	9.9
2019	9.4	0.03	9.4	9.5
2020	9.0	0.03	9.0	9.1
2021	8.6	0.03	8.5	8.7
2022	8.2	0.03	8.1	8.3

Figure 7: Graphical Analysis of Out-of-Sample Forecasts

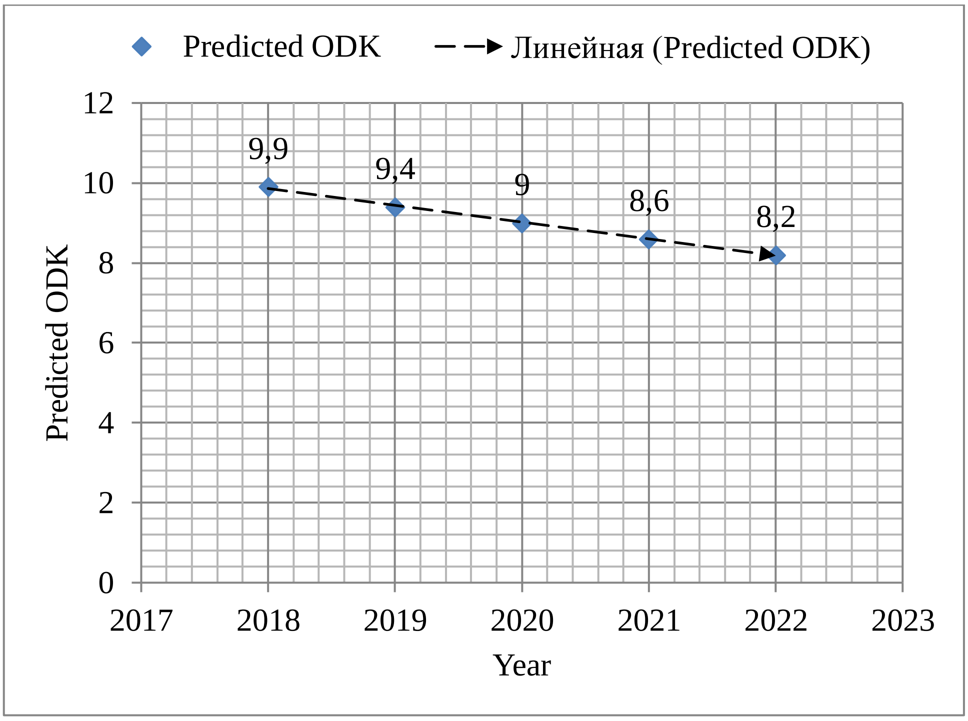


Table 10 and figure 7 show the out-of-sample forecasts only. The number of people practicing open defecation in Kenya is projected to slightly fall from approximately 9.9% in 2018 to 8.2% of the total population by the year 2022. On this downwards trajectory, Kenya has the potential to eliminate the practice of open defecation by 2030; especially if the current government considers the policy directions suggested below.

4.3 Policy Implications

i. The government of Kenya should continue to make toilets a status symbol. In this regard, the government of Kenya ought to take charge of providing good-quality sanitation facilities in poor communities where people are not able to build themselves decent sanitation facilities.
ii. The government of Kenya should continue to create demand for sanitation through teaching the public on the importance of investing in toilets.
iii. There is need for the government of Kenya to keep on encouraging a habit of systematic hand-washing, and not defecating in the open.

5.0 CONCLUSION

The study shows that the ARIMA (0, 1, 2) model is not only stable but also the most suitable model to forecast the annual number of people practicing open defecation in Kenya over the period 2018 – 2022. The model predicts a slight decrease in the annual number of people practicing open defecation in Kenya. These results are essential for the government of Kenya, especially for long-term planning with regards to materializing the much needed open defecation free society.

REFERENCES

[1] Box, G. E. P., & Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control, *Holden Day*, San Francisco.
[2] Busienei, P. J., Ogendi, G. M., & Mokua, M. A. (2019). Latrine Structure, Design, and Conditions and the Practice of Open Defecation in Lodwar Town, Turkana County, Kenya: A Qualitative Methods Research, *Environmental Health Insights*, 13: 1 – 11.
[3] Desai, R., McFarlane, C., & Graham, S. (2015). The Politics of Open Defecation: Informality, Body and Infrastructure in Mumbai, *Antipode*, 47: 98 – 120.
[4] Galan, I. D., Kim, S. S., & Graham, J. P. (2013). Exploring Changes in Open Defecation Prevalence in Sub-Saharan Africa Based on National Level Indices, *BMC Public Health*, 13 (527): 1 – 9.
[5] Megersa, S., Benti, T., & Sahiledengle, B. (2019). Prevalence of Diarrhoea and Its Associated Factors Among Under-Five Children in Open Defecation Free and Non-Open Defecation Free Households in Goba District Southeast Ethiopia: A Comparative Cross-Sectional Study, *Clinics in Mother and Child Health*, 16 (324): 1 – 9.
[6] Njuguna, J. (2019). Progress in Sanitation Among Poor Households in Kenya: Evidence From Demographic and Health Surveys, *BMC Public Health*, 19 (135): 1 – 8.
[7] Njuguna, J., & Muruka, J. (2017). Open Defecation in Newly Created Kenyan Counties: A Situational Analysis, *Journal of Health Care for the Poor and Underserved*, 28 (1): 71 – 78.

[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] Okullo, J. O., Moturi, W. N., & Ogendi, G. M. (2017). Open Defecation and its Effects on the Bacteriological Quality of Drinking Water Sources in Isiolo County, Kenya, *Environmental Health Insights*, 11: 1 – 8.
[12] Thiga, L. M., & Cholo, W. (2017). An Assessment of Open Defecation Among Residents of Thika East Sub-County, Kiambu County, Kenya, *International Journal of Medicine Research*, 2 (3): 9 – 20.
[13] UN (2015). Fact Sheet on Sanitation, *UN*, New York.
[14] UNICEF (2018). Game Plan to End Open Defecation, *UNICEF*, New York.
[15] WSP (2014). Economic Impacts of Poor Sanitation in Africa: Kenya, *WSP Country Office*, Kenya.