OPEN DEFECATION IN CAMBODIA: A BOX-JENKINS ARIMA APPROACH

Dr. Smartson. P. NYONI[1], Thabani NYONI[2]

[1] *Medical Doctor, ZICHIRe Project, University of Zimbabwe, Harare, Zimbabwe*
[2] *MSc Economics Scholar, Department of Economics, University of Zimbabwe, Harare, Zimbabwe*

ABSTRACT

In this research paper, the Box-Jenkins ARIMA model was applied in analyzing open defecation in Cambodia. The data was collected from the online World Bank data base and covers the period 2000 – 2017. The out-of-sample forecasts cover the period 2018 – 2022. The diagnostics tests employed in this study show that the open defection series is an I (1) variable. The study finally presents the ARIMA (1, 1, 0) as the optimal model in forecasting the number of people practicing open defecation in Cambodia. The model, through the residual ADF tests, and the inverse roots of the AR/MA polynomials; has been shown to be stable and suitable for forecasting and control of open defecation in Cambodia. The results of the study indicate that by 2022 the number of open defecators will be as low as 16% of the total population. The study offers a three-fold policy recommendation for consideration by the government of Cambodia.

Keyword: - *Box-Jenkins ARIMA, Forecasting, Open Defecation*

1.0 INTRODUCTION

Open defecation in Cambodia is among the highest in the world (World Bank, 2013). Sanitation promoters in Cambodia have typically had a difficult and unsuccessful time encouraging people to drop the old habit of using the open air. As a result faecal matter; lies about in the environment, contaminating streams, soils and dust carried in the air, and augmenting the toll of disease (UNICEF, 2019). Open defecation within a community harms the physical and cognitive development of children, even children living in households that use toilets themselves. Frequently digesting feces due to poor sanitation can cause diarrhea, malnutrition and stunted growth – and thus impact negatively on a child's cognitive development (World Bank, 2013). Open defecation also has other terrible consequences for health (Coffey *et al*., 2016; Mara, 2017; UNICEF, 2018), such as the spread of bacterial, viral and parasitic infections including diarrhoea, polio, cholera, soil-transmitted helminth, trachoma infection, schistosomiasis and hookworm and is also an important cause of child stunting (Megersa *et al*., 2019) and deaths (Thiga & Cholo, 2017). Furthermore, experiencing these health hazards at young ages can ultimately limit one's earning potential in later life (World Bank, 2013). Open defecators mostly reside in rural areas: in Cambodia, 77% of the total population lives in the rural areas. The main objective of this study is to model and forecast the number of people practicing open defecation in Cambodia. This study is important for future planning regarding the reduction of the number of open defecators in Cambodia.

2.0 LITERATURE REVIEW

Coffey *et al* (2014) examined open defecation in rural North India. The authors collected data, particularly from Bihar, Haryana, Madhya, Pradesh, Rajasthan and Uttar Pradesh. Many survey respondents' behaviour revealed a preference for open defecation: over 40% of households with a working latrine have at least one member who defecates in the open. Thiga & Cholo (2017) assessed open defecation among residents of Thika East Sub-County in Kenya and established 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. In Ghana, Alhassan & Anyarayor (2018) assessed the adoption of sanitation innovations introduced in Nadowli-Kaleo district in Upper West region of Ghana as part of the efforts to attain Open Defecation Free (ODF) status and established that while effective communication of innovation resulted in widespread awareness, low income levels significantly accounted for households' inability to sustain and utilize latrines. Adhikari & Ghimire (2020) investigated various determinants of open defecation in Nepal and the results of the study indicate that Nepal still has a large number of residences without a toilet. No study has been done so far, in Cambodia, to model and forecast the number of people practicing open defecation. This study is the first of its kind in the case of Cambodia but should not be the end of the road: it is just a beginning of what ought to be done as well in the fight against open defecation in Cambodia.

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 employed to analyze the ODC series under consideration.

3.2 The Moving Average (MA) model

Given:

$$ODC_t = \alpha_0\mu_t + \alpha_1\mu_{t-1} + \cdots + \alpha_q\mu_{t-q} \ldots\ldots\ldots\ldots\ldots\ldots [1]$$

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, usually denoted as MA (q). ODC is the annual number of people (as a percentage of the total population) who practice open defecation in Cambodia 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:

$$ODC_t = \beta_1 ODC_{t-1} + \cdots + \beta_p ODC_{t-p} + \mu_t \ldots\ldots\ldots\ldots\ldots\ldots [2]$$

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

$$ODC_t = \beta_1 ODC_{t-1} + \cdots + \beta_p ODC_{t-p} + \mu_t + \alpha_1\mu_{t-1} + \cdots + \alpha_q\mu_{t-q} \ldots\ldots\ldots\ldots [3]$$

While ARMA models just like AR and MA models are meant for stationary series, reality indicates that most time series data is either I (1) or I (2). In fact, in this study, the ODC series has been found to be an I (1) variables (that is, it only became stationary after first differencing). Therefore, in this paper, the model presented below is the one that will be applied.

3.5 The Autoregressive Integrated Moving Average (ARIMA) model

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

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

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

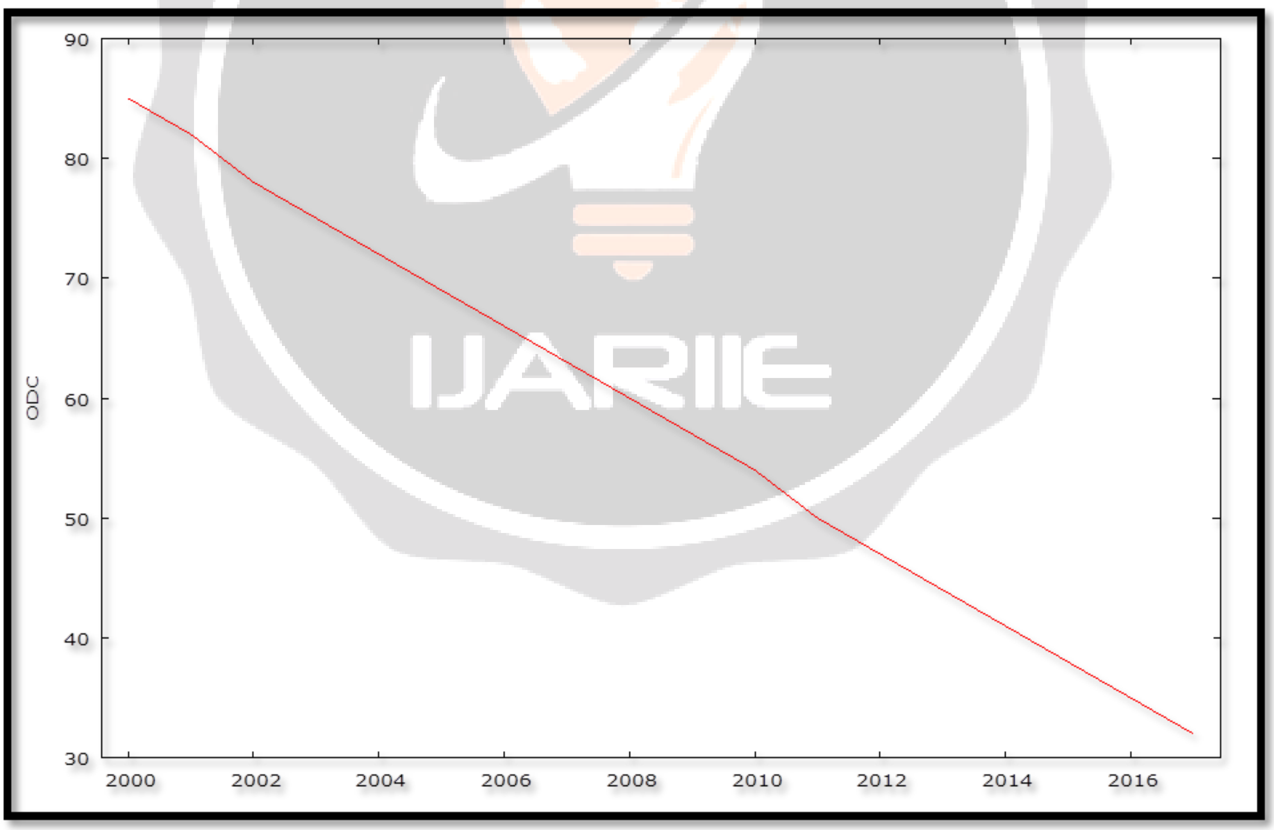
3.6 Data Collection

This study is based on annual observations (that is, from 2000 – 2017) on the number of people practicing Open Defecation [OD, denoted as ODC] (as a percentage of total population) in Cambodia. Out-of-sample forecasts will cover the period 2018 – 2022. All the data was gathered from the World Bank online database.

3.7 Diagnostic Tests & Model Evaluation

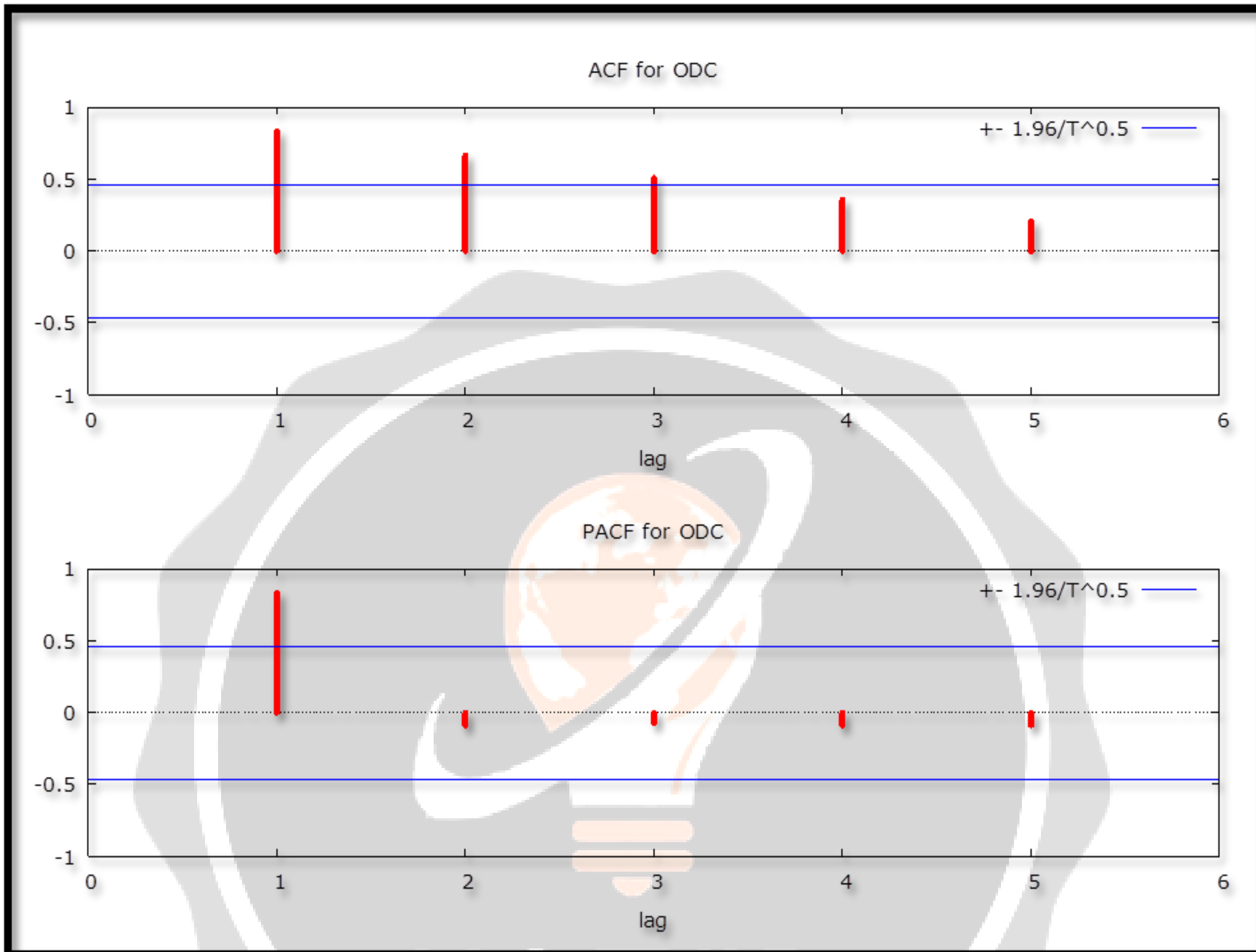
3.7.1 Stationarity Tests: Graphical Analysis

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
ODC	-0.777618	0.7998	-3.886751	@1%	Non-stationary
			-3.052169	@5%	Non-stationary
			-2.666593	@10%	Non-stationary

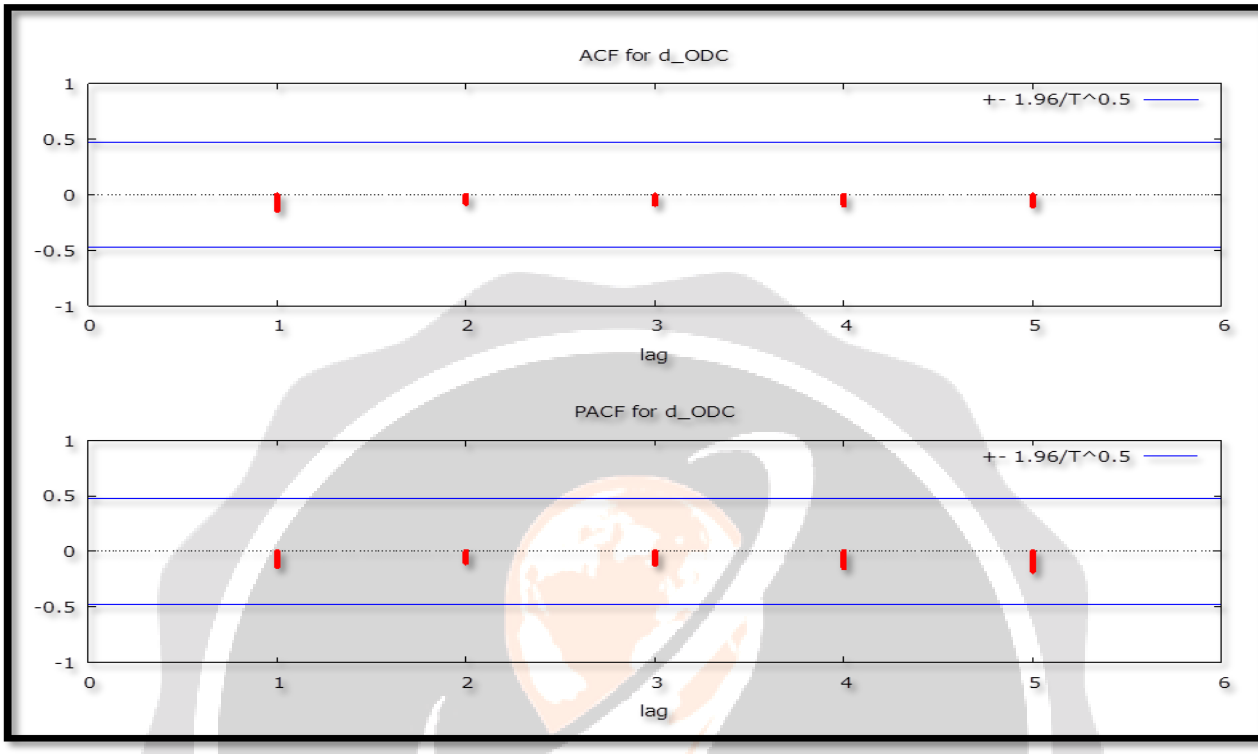
Table 2: with intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
ODC	-2.588847	0.2884	-4.616209	@1%	Non-stationary
			-3.710482	@5%	Non-stationary
			-3.297799	@10%	Non-stationary

Tables 1 and 2 show that ODC 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
ΔODC	-4.320494	0.0046	-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
ΔODC	-4.442587	0.0148	-4.667883	@1%	Non-stationary
			-3.733200	@5%	Stationary
			-3.310349	@10%	Stationary

Figure 3 as well as tables 3 and 4, indicate that ODC 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	MAE	RMSE	MAPE
ARIMA (1, 1, 0)	**15.40626**	0.090511	0.00095178	0.20438	0.31894	0.37108
ARIMA (2, 1, 0)	17.15686	0.090335	-0.0058884	0.19636	0.3178	0.35844
ARIMA (3, 1, 0)	18.80749	0.089895	-0.012394	0.18767	0.31578	0.34326

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 (1, 1, 0) model is selected.

3.8 Residual & Stability Tests

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

Table 6: with intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
R	-3.788739	0.0129	-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	-3.930838	0.0358	-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 (1, 1, 0) model; are stationary. Hence, the model is stable.

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

Figure 4: Correlogram of the Residuals

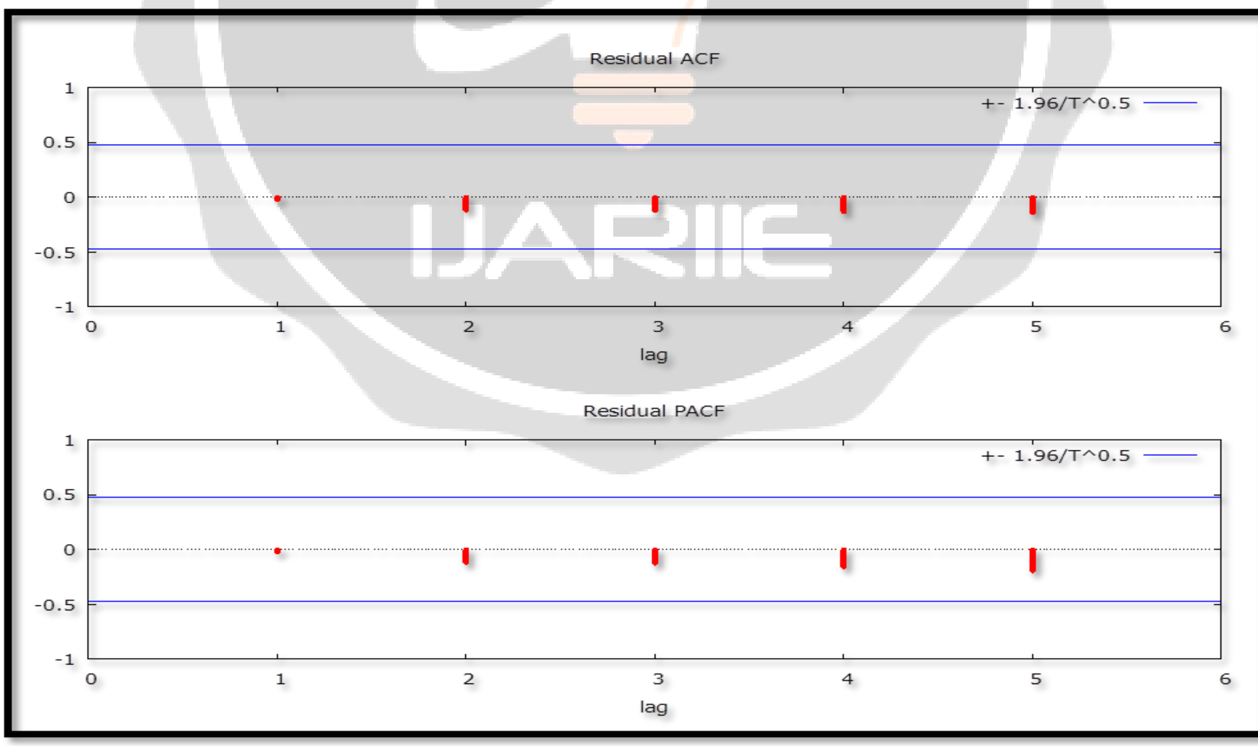
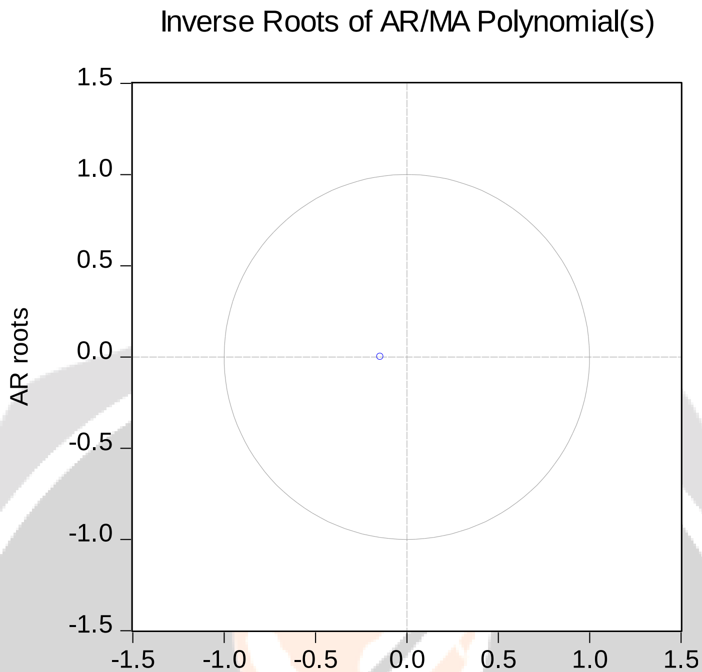


Figure 4 indicates that the estimated ARIMA (1, 1, 0) model is adequate since ACF and PACF lags are quite short and within the bands. This tells us that the "no autocorrelation" assumption is not violated in this study.

3.8.3 Stability Test of the ARIMA (1, 1, 0) 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 (1, 1, 0) model is very stable and suitable for forecasting annual number of people practicing open defecation in Cambodia.

4.0 FINDINGS

4.1 Descriptive Statistics

Table 8: Descriptive Statistics

Description	Statistic
Mean	58.222
Median	58.5
Minimum	32
Maximum	85

As shown in table 8 above, the mean is positive, that is, 58.222. This means that, over the study period, the annual average number of people practicing open defecation in Cambodia is approximately 58% of the total population. This is a warning alarm for policy makers in Cambodia with regards to the need to promote an open defecation free society. The minimum number of people practicing open defecation in Cambodia over the study period is approximately 32% of the total population, while the maximum is 85% of the total population. In fact, the number of people practicing open defecation in Cambodia has, however, continued to decline over the years from 85% in 2000 to 32% of the total population in 2017.

4.2 Results Presentation[1]

Table 9: Main Results

ARIMA (1, 1, 0) Model:				
Guided by equation [4], the chosen optimal model, the ARIMA (1, 1, 0) model can be expressed as follows: $\Delta ODC_t = -3.11932 - 0.135598\Delta ODC_{t-1} \ldots \ldots \ldots [5]$				
Variable	Coefficient	Standard Error	z	p-value
constant	-3.11932	0.0685764	-45.49	0.0000***
$\emptyset_1$	-0.135598	0.241076	-0.5625	0.5738

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

Forecast Graph

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

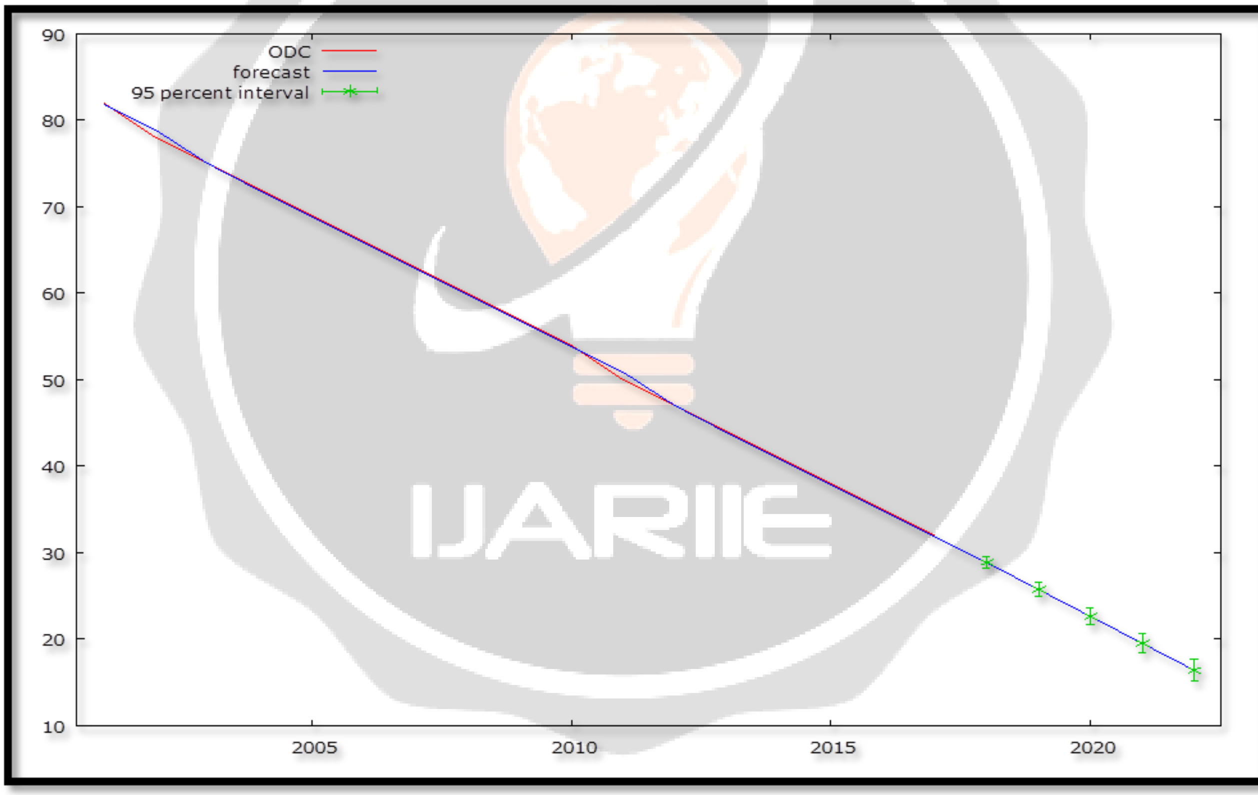


Figure 6 shows the in-and-out-of-sample forecasts of the ODC series. The out-of-sample forecasts cover the period 2018 – 2022.

[1] *The *, ** and *** imply statistical significance at 10%, 5% and 1% levels of significance; respectively;*
$\emptyset_i = \beta_i$

Predicted ODC – Out-of-Sample Forecasts Only

Table 10: Predicted ODC

Year	Predicted ODC	Standard Error	Lower Limit	Upper Limit
2018	28.86	0.319	28.24	29.49
2019	25.75	0.422	24.92	26.57
2020	22.63	0.507	21.63	23.62
2021	19.51	0.579	18.37	20.64
2022	16.39	0.644	15.13	17.65

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

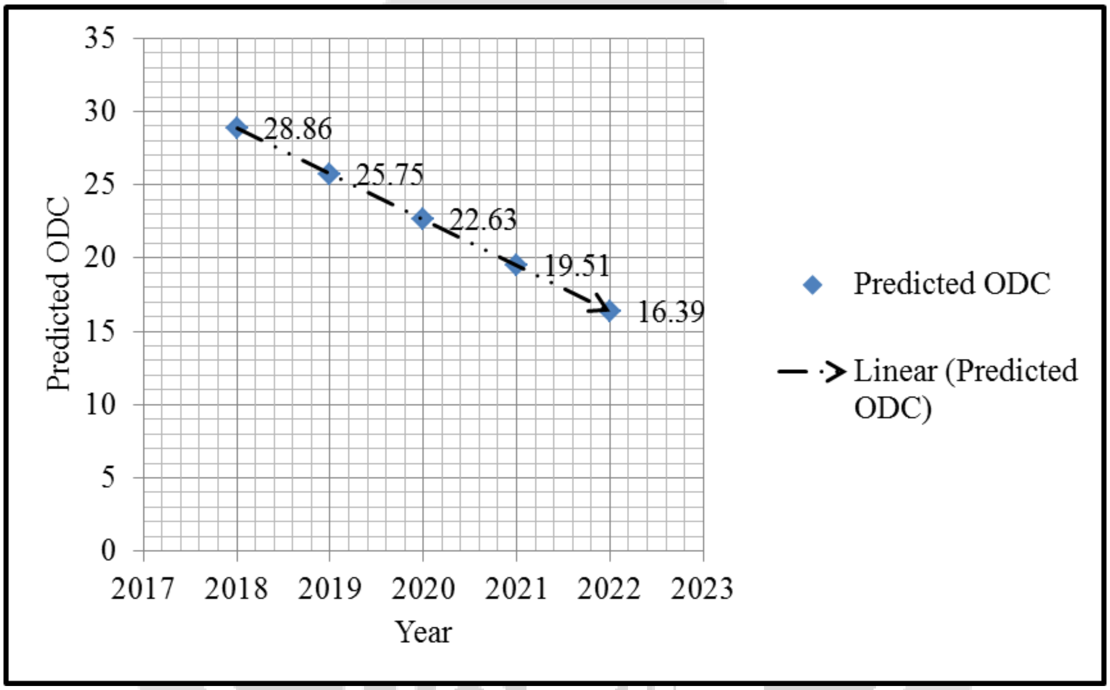


Table 10 and figure 7 show the out-of-sample forecasts only. The number of people practicing open defecation in India is projected to fall from approximately 29% in 2018 to as low as 16% of the total population by the year 2022. Even if UNICEF (2019) argues that it is not easy to end open defecation in Cambodia, the results of this study indicate that it is possible to gradually create an open defecation free Cambodia in the long run. By 2022, the number of open defecators in Cambodia will be approximately 16% of the total population and this is would be huge improvement for Cambodia. This can only materialize if the existing policy frameworks on open defecation are complemented with the recommendations suggested below:

4.3 Policy Implications

i. The government of Cambodia should continue to make toilets a status symbol, especially by intensifying the Community-Led Total Sanitation (CLTS) programs in the country.
ii. The government of Cambodia should create more demand for sanitation through teaching the public on the importance of investing in toilets.
iii. There is need for the government of Cambodia to encourage a habit of systematic hand-washing, not defecating in the open, as well as keeping toilets fly-proof. In this regard, there is need for encouragement of collective behaviour change in Cambodia.

5.0 CONCLUSION

The study reveals that the ARIMA (1, 1, 0) model is not only stable but also the most suitable model to forecast the annual number of people practicing open defecation in Cambodia over the period 2018 – 2022. The model predicts a sharp decrease in the annual number of people practicing open defecation in Cambodia. It is good for such a trend to be maintained and in this regard, a three-fold policy implication has been suggested. These findings are essential for the government of Cambodia, especially when it comes to long-term planning with regards to materializing the much needed open defecation free society.

REFERENCES

[1] Adhikari, R., & Ghimire, S. (2020). Open Defecation Free: Where Do We Need To Focus? *Health Prospect: Journal of Public Health*, 19 (1): 1 – 6.

[2] Alhassan, A., & Anyarayor, B. K. (2018). Determinants of Adoption of Open Defecation Free (ODF) Innovations: A Case Study of Nadowli-Kaleo District, Ghana, *Journal of Development and Communication Studies*, 5 (2): 54 – 69.

[3] Box, G. E. P., & Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control, *Holden Day*, San Francisco.

[4] Coffey, D., *et al*. (2014). Revealed Preference for Open Defecation: Evidence From a New Survey in Rural North India, *Economic & Political Weekly*, XLIX (38): 43 – 55.

[5] Coffey, D., *et al*. (2016). Understanding Open Defecation in Rural India, *International Growth Centre*, New Delhi.

[6] Mara, D. (2017). The Elimination of Open Defecation and Its Adverse Health Effects: A Moral Imperative for Governments and Development Professionals, *Journal of Water, Sanitation and Hygiene for Development*, 7 (1): 1 – 13.

[7] 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.

[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] 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.

[12] UNICEF (2018). Game Plan to End Open Defecation, *UNICEF*, New York.

[13] UNICEF (2019). Sanitation and Hygiene in Cambodia, *UNICEF*, New York.

[14] World Bank (2013). Growing Tall and Smart With Toilets, *World Bank*, Washington DC.