ISSN: 2394-3696 Website: ijiert.org VOLUME 7, ISSUE 7, July-2020

# UNPUZZLING INDIA'S OPEN DEFECATION PUZZLE: 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:**

A toilet is a life-saver and a dignity-protector not just a toilet! Unfortunately, WHO & UNICEF (2015) imply that the opposite is true for approximately 61% of rural and 10% of urban Indians. In fact, Coffey et al. (2014) note that most people who live in India defecate in the open and that open defecation is far more common in rural India than in urban India. Given that 70% of the total population in India lives in rural areas, it leaves a lot to be desired, especially from a public health perspective. Using annual time series data on the number of people who practice open defecation in India from 2000 – 2017, the study predicts the annual number of people who will still be practicing open defecation over the period 2018 – 2022. The study applies the Box-Jenkins ARIMA approach. The diagnostic ADF tests show that the series under consideration is an I (1) variable. Based on the AIC, the study presents the ARIMA (2, 1, 0) model as the optimal model. The diagnostic tests further reveal that the presented model is stable and its residuals are stationary in levels. The results of the study indicate that the number of people practicing open defecation in India is likely to decline over the period 2018 – 2022, from as high as 23% to as low as 11.8% of the total population. In order to sustain this desirable downwards trend, the study suggested a five-fold policy recommendation to be put into consideration, especially by the Indian government.

### **INTRODUCTION**

Open defecation is the passage of stool in an open environment (Thiga & Cholo, 2017). Of the 1.1 billion people who defecate openly, approximately 60% live in India, which means they make up more than half of the population of India (Spears, 2013). Open defecation in India, especially, in rural areas, presents a puzzle; in the sense that the open defecation rates are far higher than other developing regions where people are poorer, literacy rates are lower, and water is scarcer (Coffey et al., 2016). As the rest of the world steadily eliminates open defecation, this behaviour stubbornly persists in India. Moreso, open defecation in India is particularly threatening for health because the population density is very high (Coffey et al., 2014).

## 1.2 OBJECTIVES OF THE STUDY

- i. To investigate the years during which Open Defection was practiced by people more than 50% of the total population in India.
- ii. To forecast the number of people practicing open Defecation for the period 2018 2022.
- iii. To examine the trend of Open Defecation for the out-of-sample period.

#### 1.3 RELEVANCE OF THE STUDY

Open defecation is exceptionally widespread in India (Spears, 2013). Because open defecation has 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 (Mara et al., 2010; Spears, 2013; Chambers & Von Medeazza, 2013; Coffey et al., 2013; Ghosh et al., 2014; Strunz et al., 2014; Boisson et al., 2016; Odagiri et al., 2017; Megersa et al., 2019) and deaths (Thiga & Cholo, 2017), it is important for researchers and policy makers to model and forecast the number of people practicing open defecation. This study, hinged on the above objectives, seeks to model and forecast the number of people practicing open defecation in India. This research will go a long way in examining the possibility of unpuzzling the Indian open defecation puzzle.

### LITERATURE REVIEW

Coffey et al (2014) investigated open defecation in rural North India. The study 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. In Kenya, Thiga & Cholo (2017) assessed open defecation among residents of Thika East Sub-County. The study employed a descriptive cross-sectional design in which 20554 households were included. The study 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. The study 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 Indian study, Gauri et al. (2018) examined social norms shifting to reduce open defecation. The study survey was actually done in Uttar Pradesh and basically found out that there is a positive correlation between latrine use and social norms at baseline. In Ghana, Alhassan & Anyarayor (2018) investigated 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. Interviews were used to gather data. The study established that while effective communication of innovation resulted in widespread awareness, low income levels significantly accounted for households' inability to sustain and utilize latrines.

Nyoni (2019) forecasted total population in India using the Box-Jenkins ARIMA technique based on annual time series data on total population in India from 1960 to 2017. The study presented the ARIMA (1, 2, 3) model and concluded that total population in India will continue to sharply rise in the next three decades, thereby posing a threat to both natural and non-renewable resources. This will be a worse threat if the open defecation puzzle is left unpuzzled. Adhikari & Ghimire (2020) examined various determinants of open defecation in Nepal. Bivariate analysis was done to assess the association between dependent variables (toilet status – having and not having toilets in the household) and independent variables (demographic, socio-economic and geographical characteristics) using the Chi-square test. The multivariate logistic regression model was used to assess significant predictors for a household not having a toilet after controlling other variables. 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 India or elsewhere, to model and forecast the number of people practicing open defecation. This study is the first of its kind, and is envisioned to consolidate existing policy frameworks in the fight against open defecation in India.

### **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 OD series under consideration. Below is the mathematical intuition behind the Box-Jenkins approach to modeling and forecasting:

## 3.2 The Moving Average (MA) model

Given:

percentage of the total population) who practice open defecation at time t,  $\alpha_0$  ...  $\alpha_q$  are estimation parameters,  $\mu_t$  is the current error term while  $\mu_{t-1}$  ...  $\mu_{t-q}$  are previous error terms. Therefore:

is an MA process of order one, often denoted as MA (1). Due to the fact that previous error terms are unobserved variables, we then scale them such that  $\alpha_0=1$ . Given:

$$E(\mu_t) = 0$$

$$\forall t$$

$$(3)$$

It therefore implies that:

where  $\mu_t$  is independent with a common varience  $\sigma^2$ . Hence, we can now re – specify equation [1] as follows:

Equation [6] may be re – written as:

$$OD_{t} = \sum_{i=1}^{q} \alpha_{i} \mu_{t-i} + \mu_{t} \dots [7]$$

We can also write equation [7] as follows:

where L is the lag operator.

or as:

where:

$$a(L)=\theta(L)$$
 ......[10]

### 3.3 The Autoregressive (AR) model

Given:

Where  $\beta_1 \dots \beta_p$  are estimation parameters,  $OD_{t-1} \dots OD_{t-p}$  are previous period values of the OD series and  $\mu_t$  is as previously defined. Equation [11] is an Autoregressive (AR) process of order p, and is usually denoted as AR (p); and may also be written as:

$$OD_{t} = \sum_{i=1}^{p} \beta_{i}OD_{t-1} + \mu_{t} \dots [12]$$

Equation [12] may be re – written as:

or as:

where:

or as:

Hence:

is an AR process of order one, usually denoted as AR (1).

## 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 [11]; an ARMA (p, q) process may be specified as shown below:

or as:

by simply combining equations [7] and [12]. Equation [18] can also be written as:

where  $\phi(L)$  and  $\theta(L)$  are polynomials of orders p and q respectively, algebraically defined as:

It is not unimportant to remember that the ARMA (p, q) model, just like the AR (p) and the MA (q) models; can only be applied for stationary time series data. However, in real life, many time series are non – stationary. In fact, in this study, the OD series has been found to be an I (1) variable (that is, it only became stationary after first differencing). Based on that simple logic, ARMA models are not suitable for modeling and forecasting non – stationary time series data. In such a case, the model described below is the one that should ideally be used.

### 3.5 The Autoregressive Integrated Moving Average (ARIMA) model

A stochastic process OD<sub>t</sub> 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)

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

representation. If the sequence  $\Delta^d OD_t$  satisfies an ARMA (p, q) process; then the sequence of  $OD_t$  also satisfies the ARIMA (p, d, q) process such that:

which can also be represented as shown below:

where  $\Delta$  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] (as a percentage of total population) in India. 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

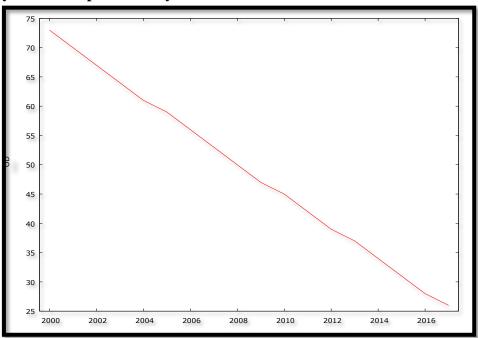


Figure 1

Figure 1 shows that the OD series is declining sharply over the period under study. There is a clear downwards trend and for this reason, we suspect that the series is non-stationary. Hence, the Augmented-Dickey-Fuller (ADF) test will be applied to verify the existence of a unit root, that is, to test for stationarity. The fact that the OD series is trending downwards point to a possible win in the fight against open defecation in India and this is encouraging for Indians who, over the years, have been popular for defecating in the open, especially the rural folks.

## 3.7.2 The Correlogram in Levels

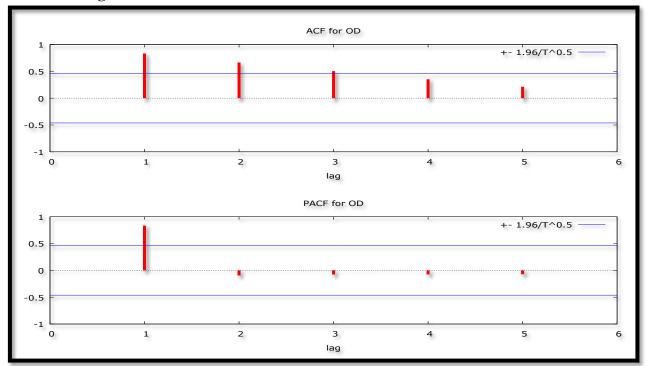


Figure 2: Correlogram in Levels

Figure 2 is the correlogram of the OD series in levels. The correlogram points to the possible existence of a unit root since the 3 lags of the ACF and first lag of the PACF lie out side the 5% confidence bands. For a stationary series, these lags could have been lying right inside the confidence bands. Hence, the correlogram also points to the fact that it is indeed necessary to carry out some confirmatory ADF tests for stationarity.

## 3.7.3 The ADF Test in Levels

Table 1: with intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
OD	-1.053428	0.7047	-3.959148	@1%	Non-stationary
			-3.081002	@5%	Non-stationary
			-2.681330	@10%	Non-stationary

Table 2: with intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
OD	-3.343622	0.0929	-4.616209	@1%	Non-stationary
			-3.710482	@5%	Non-stationary
			-3.297799	@10%	Stationary

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

## 3.7.4 The Correlogram (at First Differences)

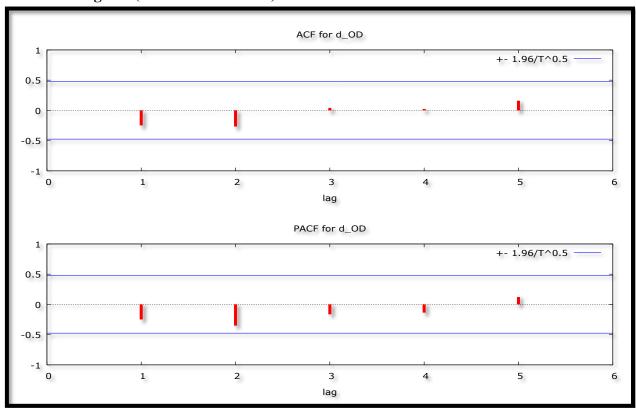


Figure 3: Correlogram (at First Differences)

### 3.7.5 The ADF Test (at First Differences)

Table 3: with intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
$\Delta OD$	-4.164133	0.0068	-3.959148	@1%	Stationary
			-3.081002	@5%	Stationary
			-2.681330	@10%	Stationary

Table 4: with intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Value	S	Conclusion
$\Delta \text{OD}$	-3.728659	0.0583	-4.886426	@1%	Non-stationary
			-3.828975	@5%	Non-stationary
			-3.362984	@10%	Stationary

Figure 3 as well as tables 3 and 4, indicate that OD 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)	23.78702	0.16466	-0.0039526	0.33562	0.40752	0.8018
ARIMA (2, 1, 0)	23.00073	0.14551	-0.014928	0.29054	0.37321	0.67177
ARIMA (3, 1, 0)	23.95547	0.13458	-0.026062	0.27595	0.36136	0.62069
ARIMA (4, 1, 0)	24.37195	0.137	-0.051587	0.28056	0.34646	0.65401

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, 1, 0) model is finally chosen.

## 3.8 Residual & Stability Tests

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

Table 6: with intercept

Variable	ADF Statistic	Probability	Critical Value	S	Conclusion
R	-3.828308	0.0119	-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	S	Conclusion
R	-4.194801	0.0227	-4.667883	@1%	Non-stationary
			-3.733200	@5%	Stationary
			-3.310349	@10%	Stationary

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

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

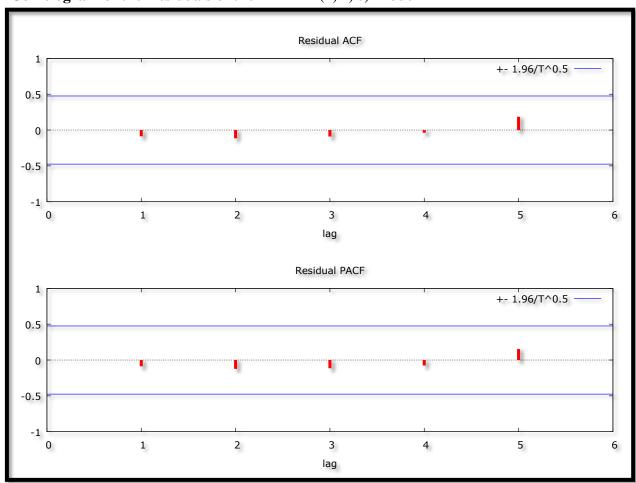
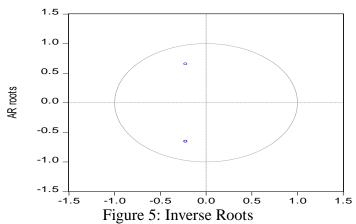


Figure 4: Correlogram of the Residuals

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

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

Inverse Roots of AR/MA Polynomial(s)



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 (2, 1, 0) model is really stable and suitable for forecasting annual number of people practicing open defecation in India.

### **FINDINGS**

### **4.1 Descriptive Statistics**

Table 8: Descriptive Statistics

Description	Statistic
Mean	49
Median	48.5
Minimum	26
Maximum	73
Standard deviation	14.769
Skewness	0.031003
Excess kurtosis	-1.1952

As shown in table 8 above, the mean is positive, that is, 49. This means that, over the study period, the annual average number of people practicing open defecation in India is approximately 49% of the total population. This is a warning alarm for Indian policy makers with regards to the need to promote an open defecation free society. The minimum number of people practicing open defecation over the study period is approximately 26% of the total population, while the maximum is 73% of the total population. In fact, the number of people practicing open defecation in India has continued to decline over the years from 73% in 2000 to 26% of the total population. This is a desirable change and there is need to intensify policies and strategies that discourage the practice of open defecation in India. The skewness is 0.031003 and the most important characteristic is that it is positive, meaning that the OD series is positively skewed and non-symmetric. Excess kurtosis is -1.1952; showing that the OD series is not normally distributed.

### 4.2 Results Presentation

Table 9: Main Results

Table 9. Walli Results							
ARIMA (2, 1, 0) Model:							
expressed as fo	Guided by equations [23] and [24], the chosen optimal model, the ARIMA (2, 1, 0) model can be expressed as follows: $\Delta OD_t = -2.77322 - 0.413975 \Delta OD_{t-1} - 0.41304 \Delta OD_{t-2}$						
Variable	Coefficient	Standard Error	Z	p-value			
constant	2 77222	0.0514494	52.0	0.0000***			

Variable	Coefficient	Standard Error	Z	p-value
constant	-2.77322	0.0514484	-53.9	0.0000***
$\emptyset_1$	-0.413975	0.246525	-1.679	0.0931*
$\emptyset_2$	-0.41304	0.247309	-1.67	0.0949*

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

## **Forecast Graph**

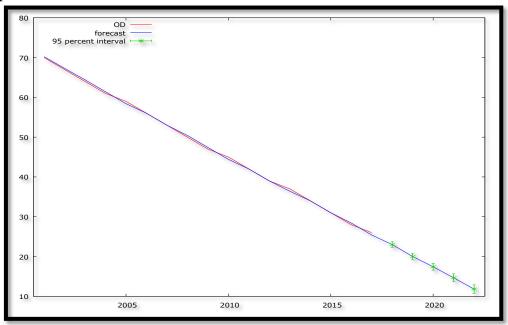


Figure 6: Forecast Graph – In & Out-of-Sample Forecasts

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

## **Predicted OD – Out-of-Sample Forecasts Only**

Table 10: Predicted OD

10010 10111001000 02							
Year	Predicted OD	Standard Error	Lower Limit	Upper Limit			
2018	23	0.371	22.27	23.73			
2019	20	0.43	19.16	20.84			
2020	17.42	0.449	16.54	18.29			
2021	14.66	0.503	13.67	15.64			
2022	11.8	0.551	10.72	12.88			

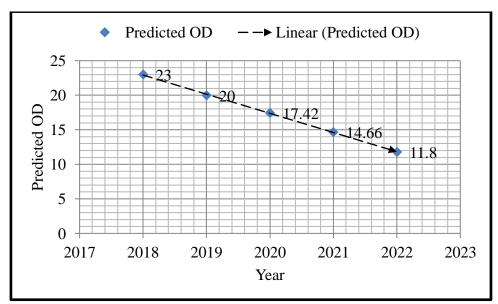


Figure 7: Graphical Analysis of Out-of-Sample Forecasts

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 23% in 2018 to as low as 11.8% of the total population by the year 2022. Indeed, it is possible to unpuzzle India's open defecation puzzle.

ISSN: 2394-3696 Website: ijiert.org VOLUME 7, ISSUE 7, July-2020

## **4.3 Policy Implications**

Open defecation is a clear typical example of a "negative externality" in which one person behavior hurts another. From a Public Economics perspective, in such situations with negative externalities, government intervention is necessary to either stop the externality or reduce its harm. Therefore, in order to continue unpuzzling the open defecation puzzle in India, the government of India ought to consider the following policy directions:

- i. The government of India should continue to make toilets a status symbol so that people stop thinking about toilets as "dark, dirty and smelly places" but rather consider toilets to be "rooms of happiness".
- ii. The government of India should create more demand for sanitation through teaching the public on the importance of investing in toilets, especially in light of disease transmission and other risks associated with open defecation.
- iii. There is need to encourage a habit of systematic hand-washing, not defecating in the open, as well as keeping toilets fly-proof. This will easily translate into an open-defecation-free India.
- iv. In Indian schools, especially in Early Childhood Development and primary schools, there is need for teachers to teach pupils to develop the habit of using toilets. This is an easy way of breaking the intergenerational cycle of open defecation.
- v. Sanitation is indeed profitable and beneficial, especially from a public health and dignity perspective, and it inevitably deserves public investment. Hence, the government of India should channel adequate financial resources towards funding open-defectaion-related projects and initiatives in all Indian states.

### **CONCLUSION**

India's open defecation rates are indeed surprising: despite rapid economic growth, improving literacy rates and widespread access to improved water sources, a significant number of rural households still resort to open defecation (Coffey et al., 2016). The study shows that the ARIMA (2, 1, 0) model is not only stable but also the most suitable model to forecast the annual number of people practicing open defecation in India over the period 2018 – 2022. The model predicts a sharp decrease in the annual number of people practicing open defecation in India. Such a trend should be maintained and in this regard, a four-fold policy implication has been suggested. These findings are essential for the government of India, 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 Defection 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) Boisson et al. (2016). Water, Sanitation and Hygiene for Accelerating and Sustaining Progress on Neglected Tropical Diseases: A New Global Strategy, International Health, 8: 19 21.
- 4) Box, G. E. P., & Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control, Holden Day, San Francisco.
- 5) Chambers, R., & Von Medeazza, G. (2013). Sanitation and Stunting in India, Economic & Political Weekly, 48 (25): 15 20.
- 6) Coffey, D., et al. (2013). Stunting Among Children: Facts and Implications, Economic & Political Weekly, 48 (34): 69 75.
- 7) 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.
- 8) Coffey, D., et al. (2016). Understanding Open Defecation in Rural India, International Growth Centre, New Delhi.
- 9) Gauri, V., Rahman, T., & Sen, I. (2018). Shifting Social Norms to Reduce Open Defection in Rural India, Working Paper No. 8684, World Bank, Washington DC.
- 10) Ghosh, A., Gupta, A., & Spears, D. (2014). Are Children in West Bengal Shorter Than Children in Bangladesh? Economic & Political Weekly, 48 (35): 1 5.

- 11) 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.
- 12) Mara, D., et al. (2010). Sanitation and Health, PLoS Medicine, 7: 1-6.
- 13) 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.
- 14) Nyoni, T (2018b). Modeling and Forecasting Inflation in Kenya: Recent Insights from ARIMA and GARCH analysis, Dimorian Review, 5 (6): 16 40.
- 15) 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.
- 16) Nyoni, T. (2018c). Box Jenkins ARIMA Approach to Predicting net FDI inflows in Zimbabwe, MPRA Paper No. 87737, University Library of Munich, Munich, Germany.
- 17) Nyoni, T. (2019). Predicting Total Population in India: A Box-Jenkins ARIMA Approach, MPRA Paper No. 92436, University Library of Munich, Munich, Germany.
- 18) Odagiri, M., et al. (2017). Enabling Factors for Sustaining Open Defection Free Communities in Rural Indonesia: A Cross-Sectional Study, International Journal of Environmental Research and Public Health, pp: 1-20.
- 19) Spears, D. (2013). How Much International Variation in Child Height Can Sanitation Explain? Working Paper No. 6351, World Bank, Washington DC.
- 20) Strunz, E. C., et al. (2014). Water, Sanitation, Hygiene and Soil-transmitted Helminth Infection: A Systematic Review and Meta Analysis, PLoS Medicine, 11: 1 9.
- 21) Thiga, L. M., & Cholo, W. (2017). An Assessment of Open Defection Among Residents of Thika East Sub-County, Kiambu County, Kenya, International Journal of Medicine Research, 2 (3): 9 20.
- 22) UNICEF (2018). Game Plan to End Open Defecation, UNICEF, New York.
- 23) WHO & UNICEF (2015). Progress on Sanitation and Drinking Water 2015 Update, WHO, Geneva.