The population question in Zimbabwe: reliable projections from the Box – Jenkins ARIMA approach

Nyoni, Thabani

University of Zimbabwe

10 September 2019

The Population Question In Zimbabwe: Reliable Projections From The Box – Jenkins ARIMA Approach

Surname: Nyoni, Name: Thabani

Department of Economics

Faculty of Social Studies

University of Zimbabwe

Harare, Zimbabwe

Email: nyonithabani35@gmail.com

Abstract

Employing annual time series data on total population in Zimbabwe from 1960 to 2017, we model and forecast total population over the next 3 decades using the Box – Jenkins ARIMA technique. Diagnostic tests indicate that Zimbabwe annual total population is neither I (1) nor I (2) but for the sake of simplicity, we assume it is I (2). Based on the AIC, the study presents the ARIMA (2, 2, 2) model as the best model. The diagnostic tests further imply that the presented model is stable and acceptable. The results of the study indicate that total population in Zimbabwe will continue to increase in the next three decades. In order to enjoy the benefits of the Ahlburg (1998) and Becker *et al* (1999) prophecy, 2 policy prescriptions have been put forward.

Key Words: ARIMA, Forecasting, Population growth, Population policy, Total population, Zimbabwe

JEL Codes: C53, Q56, R23

INTRODUCTION

At the beginning of the 21st century, the world's population was estimated to be around 6.1 billion people (Tartiyus *et al*, 2015). Forecasts produced by the United Nations place the figure at more than 9.2 billion by the year 2050 before reaching a maximum of 11 billion by 2200. More than 90% of that population will be living in the developing world (Todaro & Smith, 2006). The consequences of population growth are indeed not an issue of numbers but that of human welfare simply because population growth directly affects the provision of welfare and development. The consequences of a spiraling population manifests heavily on species extinction, deforestation, desertification, climate change and the destruction of natural ecosystems on one hand; and unemployment, mounting pressure on housing, transport traffic congestion, pollution and infrastructure security and strain on amenities (Dominic *et al*, 2016).

Population forecasts are a critical step for incorporating population concerns into comprehensive national planning strategies. Forecasts readily provide policy makers with reliable data on the evolution and nature of the Zimbabwean population, and hence are indispensable in forecasting strategies that involve public and private sector activities. More specifically, population projections are significant in three main ways. Firstly, they are crucial in determining labour supply, which is indispensable in the production of goods and services. Secondly, they define number and nature of consumers who are the ultimate beneficiaries of the production process. Thirdly, government departments need projections of future demographic parameters for planning purposes and resource allocation. Some of the variables for which projections can be made include the number of households, school environment, size and structure of the labor force, domestic incomes and consumption (ZimStats, 2015). In Zimbabwe, just like in any other country, population modeling and forecasting is essential for a well-informed and evidence-driven policy dialogue. This study endeavors to model and forecast population of Zimbabwe using the Box-Jenkins ARIMA technique.

Research Objectives

i. To analyze total population trends in Zimbabwe over the period 1960 – 2017.
ii. To determine an optimal ARIMA model for forecasting total population in Zimbabwe.
iii. To project Zimbabwe's total population over the period 2018 – 2050.

Statement of the Problem

Population growth dynamics are associated with uncertainty and hence governments and other relevant stakeholders are always seeking reliable ways to forecast total population in order to prepare themselves for handling population growth-driven uncertainties such as mounting pressure on housing & food security, excessive urbanization, rising & persistent unemployment and natural resources depletion amongst others. Policy makers can hedge against such consequences if they are able to predict the future with regards to total population growth dynamics. The importance of population forecasts in Zimbabwe has been recognized by ZimStats (2015), which has clearly highlighted the need for accurate population projections for a number of purposes such as resource allocation and strategic planning. Despite such paramount importance of total population forecasts, in Zimbabwe there is no comprehensive study which has so far attempted to model and forecast total population using the Box-Jenkins ARIMA framework. This study is envisioned to go a long way in improving population policy formulation in Zimbabwe.

LITERATURE REVIEW

Theoretical Literature Review

The Malthusian population prophecy avers that population growth is harmful to economic growth in the sense that human population grows geometrically while the means of subsistance grows arithmetically being subject to the law of diminishing returns and subsequently concludes that population growth is surely a problem to any economy. Solow (1956), in support of Malthus (1798); opined that population growth is surely a nuisance but he did not agree on the "how" part of it. How population growth would be a real problem for the economy – that's where Solow and Malthus failed to be on same pages. Solow (1956) strongly opined that an increase in the "population growth rate" would reduce the capital per worker as well as the steady-state output per worker and subsequently concluded that higher population growth could retard productivity and economic growth. Ahlburg (1998) and Becker *et al* (1999)'s lines of thoughts were not consistent with both Solow (1956) and Malthus (1798). Ahlburg (1998) reiterated that an increase in population growth would lead to an increase the need for goods and services through the "technology-pushed" and the "demand-pulled" channels and on the other hand, Becker *et al*

(1999) argued that high population growth rate induces high labour force which is the source of real wealth. In Zimbabwe, it is arguably reasonable to argue that population growth dynamics do not follow the predictions of Malthus (1798) and Solow (1956) but rather trace the projections made by Ahlburg (1998) and Becker *et al* (1999). This is especially true given that Zimbabwe is very much endowed with both skilled human resources and abundant natural resources most of which are still lying idle. Therefore an increase in population in Zimbabwe, accompanied by technological advancement, will result in improved labour productivity and hence economic growth will improve. Furthermore, an increase in population in Zimbabwe is arguably welcome, given the need for more labour force for industrial production.

Empirical Literature Review

In Pakistan, Zakria & Muhammad (2009); forecasted population using Box-Jenkins ARIMA models, and employed on a data set ranging from 1951 to 2007; and found out that the ARIMA (1, 2, 0) model was the best model. Beg & Islam (2016) looked at modeled population growth of Bangladesh from an autoregressive time trend approach based on a data set ranging over 1965 – 2003 and revealed a downward population growth for Bangladesh for the extended period up to 2043. In Ethiopia, Ayele & Zewdie (2017) analyzed human population size and its pattern in using Box-Jenkins ARIMA models and employing annual data from 1961 to 2009 and found out that the optimal model for modeling and forecasting population in Ethiopia was the ARIMA (2, 1, 2) model. In the case of Zimbabwe, we will employ the Box-Jenkins ARIMA methodology for the data set ranging from 1960 to 2017.

METHODS & MATERIALS

ARIMA Models

ARIMA models are usually considered as delivering more accurate forecasts as compared to econometric techniques (Song *et al*, 2003b). ARIMA models perform better than multivariate models in forecasting (Du Preez & Witt, 2003). Overall performance of ARIMA models is also superior to that of the naïve models and smoothing techniques (Goh & Law, 2002). ARIMA models were developed by Box and Jenkins in the 1970s and their approach of identification, estimation and diagnostics is hinged on the law of parsimony (Asteriou & Hall, 2007). The general form of the non – seasonal ARIMA (p, d, q) may be represented by a backward shift operator as shown in equation [1]:

$$\emptyset(B)(1-B)^d Z_t = \theta(B)\mu_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [1]$$

Equation [1] is simply explained in terms of the autoregressive (AR) and moving average (MA) characteristic equations as follows: AR and MA characteristic operators are given by equations [2] and [3] respectively. Equation [2] explains the repetitive habit of a series, hence the term autoregressive, meaning to say, the series depend on its own past values. Equation [3] denotes that the series is explained by the current and past values of the disturbance term or the error term:

$$\emptyset(B) = \left(1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [2]$$

$$\theta(B) = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [3]$$

and

$$(1-B)^d Z_t = \Delta^d Z_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [4]$$

Equation [4] is explained as follows: $\emptyset$ is the parameter estimate of the autoregressive component, θ is the parameter estimate of the moving average component, Δ is the difference operator, d is the difference, B is the backshift operator, μ_t is the disturbance term and Z_t is the actual total population series at time t. Equation [4] is the general presentation of the differenced total population series.

The Box – Jenkins 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, 2018i).

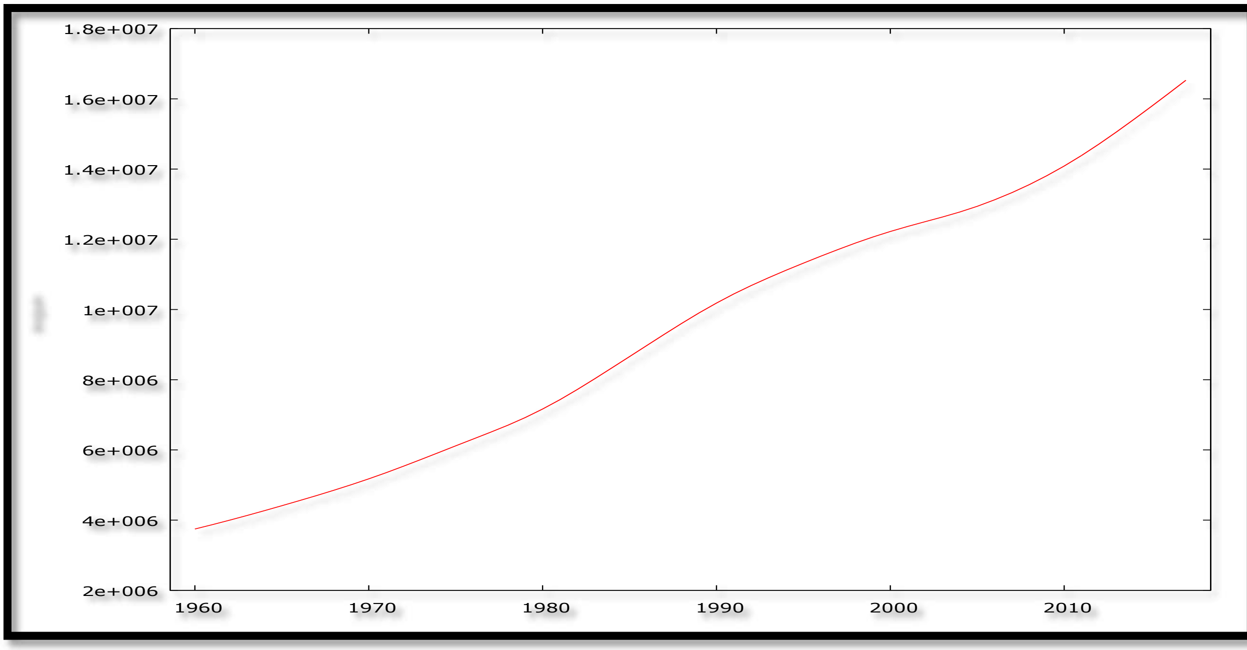
Data Collection

This research article is based on 58 observations (1960 – 2017) of annual total population (POP, referred to as Z in the mathematical formulations above) in Zimbabwe. All the data was collected from the World Bank online database, which is one of the most reliable and credible sources of various macroeconomic data.

Diagnostic Tests & Model Evaluation

Stationarity Tests: Graphical Analysis

Figure 1. Graphical Analysis

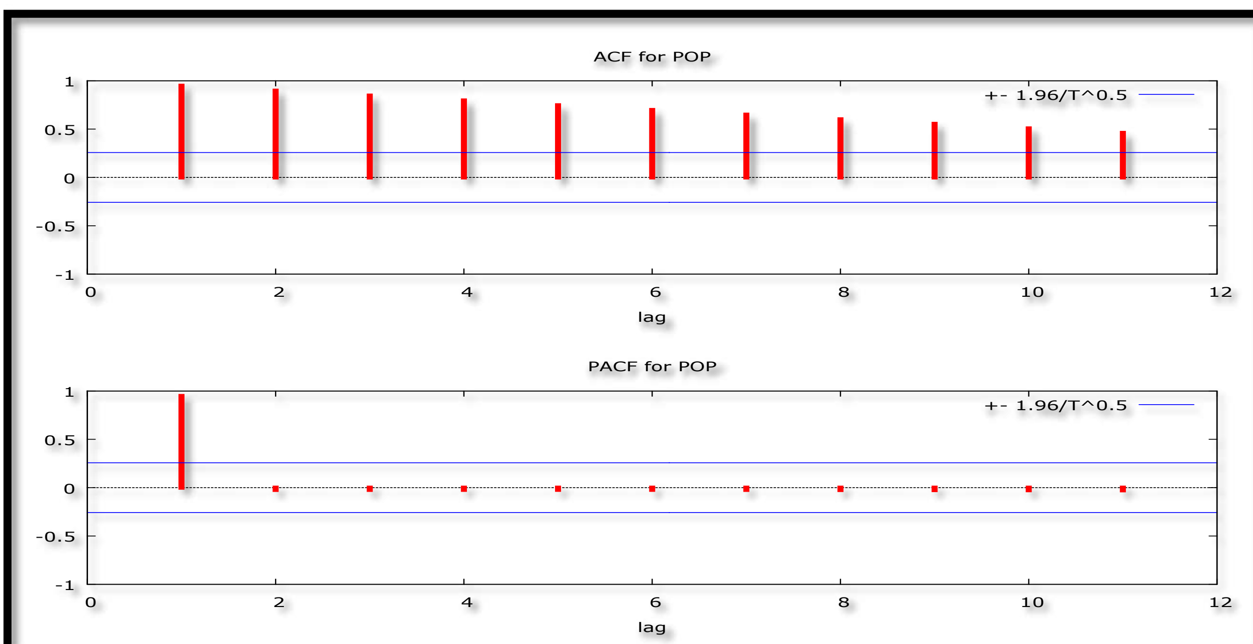


Source: Author's Own Computation

The POP variable is not stationary since it is trending upwards over the period 1960 – 2017. This means that the mean and varience of POP is changing over time.

The Correlogram in Levels

Figure 2. Correlogram in Levels



Source: Author's Own Computation

The ADF Test

Table 1: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	1.964602	0.9998	-3.560019	@1%	Not stationary
			-2.917650	@5%	Not stationary
			-2.596689	@10%	Not stationary

Source: Author's Own Computation

Table 2: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-2.204855	0.4771	-4.140858	@1%	Not stationary
			-3.496960	@5%	Not stationary
			-3.177579	@10%	Not stationary

Source: Author's Own Computation

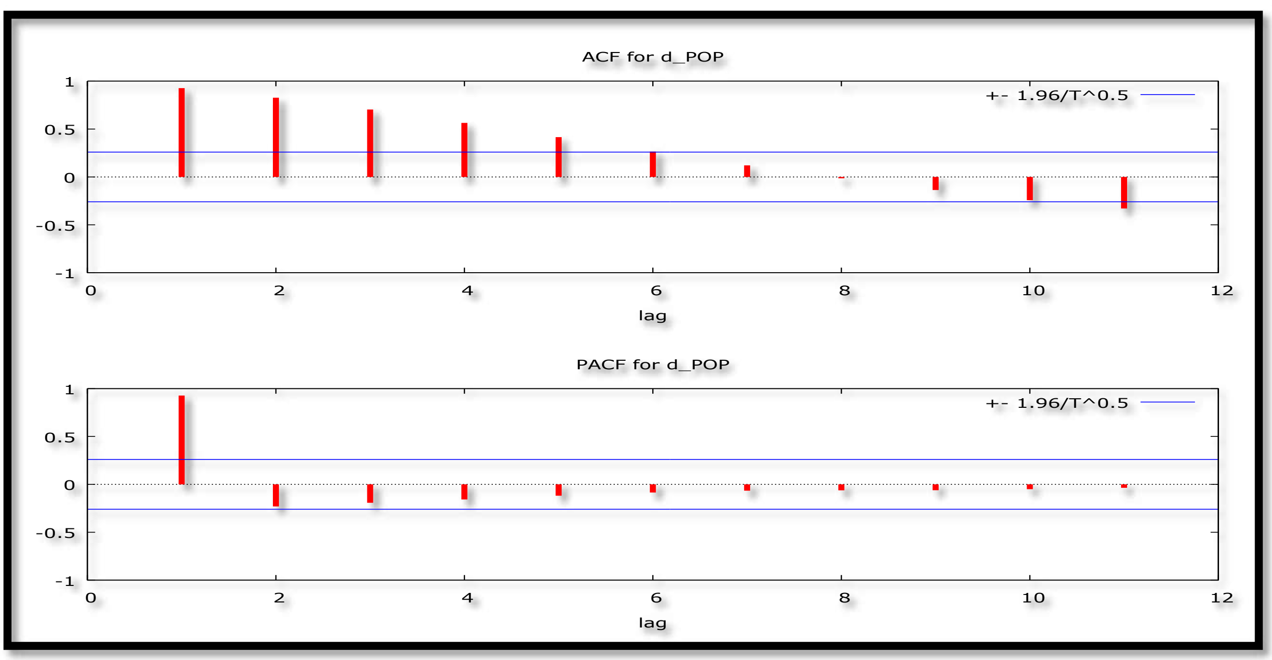
Table 3: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	2.290082	0.9941	-2.609324	@1%	Not stationary
			-1.947119	@5%	Stationary
			-1.612867	@10%	Stationary

Author's Own Computation

The Correlogram (at 1st Differences)

Figure 3. Correlogram at 1st differences



Source: Author's Own Computation

Table 4: 1st Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-2.657339	0.0883	-3.560019	@1%	Not stationary
			-2.917650	@5%	Not stationary
			-2.596689	@10%	Stationary

Source: Author's Own Computation

Table 5: 1st Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-3.433154	0.0578	-4.140858	@1%	Not stationary
			-3.496960	@5%	Not stationary
			-3.177579	@10%	Stationary

Author's Own Computation

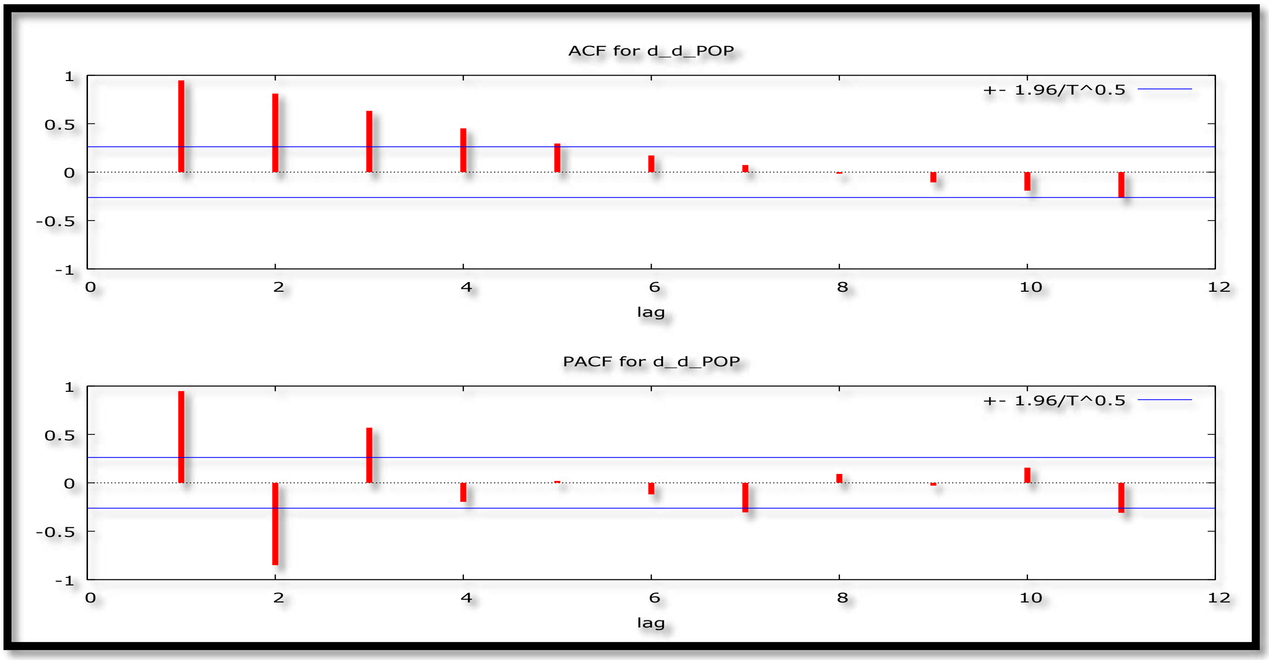
Table 6: 1st Difference-without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	0.102680	0.7110	-2.609324	@1%	Not stationary
			-1.947119	@5%	Not stationary
			-1.612867	@10%	Not stationary

Source: Author's Own Computation

The Correlogram in (2nd Differences)

Figure 4. Correlogram in 2nd differences



Source: Author's Own Computation

Table 7: 2nd Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.709768	0.4206	-3.560019	@1%	Not stationary
			-2.917650	@5%	Not stationary
			-2.596689	@10%	Not stationary

Author's Own Computation

Table 8: 2nd Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.730009	0.7239	-4.140858	@1%	Not stationary
			-3.496960	@5%	Not stationary
			-3.177579	@10%	Not stationary

Source: Author's Own Computation

Table 9: 2nd Difference-without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.546426	0.1136	-2.609324	@1%	Not stationary
			-1.947119	@5%	Not stationary
			-1.612867	@10%	Not stationary

Source: Author's Own Computation

Figures 2 – 4 and tables 1 – 9 indicate that the POP series is not I (0), not I (1), neither is it I (2). However, in this study, we will assume that the POP series is I (2).

Evaluation of ARIMA models (with a constant)

Table 10. Evaluation of ARIMA models

Model	AIC	U	ME	MAE	RMSE	MAPE
ARIMA (1, 2, 1)	1049.487	0.01073	-0.94987	1930	2551.7	0.021479
ARIMA (2, 2, 1)	1011.435	0.0073993	2.7477	1416	1760.9	0.015758
ARIMA (0, 2, 1)	1153.175	0.024506	-0.43852	5324.1	6657.6	0.052977
ARIMA (1, 2, 0)	1098.303	0.016901	-3.8478	3153.8	4086	0.034378
ARIMA (2, 2, 2)	**1005.332**	**0.0068006**	3.2247	1316.2	1626.4	0.014463

Source: Author's Own Computation

Table 10 displays the model evaluation statistics. A model with a lower AIC value is better than the one with a higher AIC value (Nyoni, 2018n). Theil's U must lie between 0 and 1, of which the closer it is to 0, the better the forecast method (Nyoni, 2018l). In this research, the researchers will consider both the AIC and Theil's U in order to choose the optimal model in terms of parsimony (AIC) and forecast accuracy (Theil's U). Hence, the ARIMA (2, 2, 2) model is chosen as the optimal model.

Residual & Stability Tests

ADF Tests of the Residuals of the ARIMA (2, 2, 2) Model

Table 11: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
H_t	-7.784168	0.0000	-3.560019	@1%	Stationary
			-2.917650	@5%	Stationary
			-2.596689	@10%	Stationary

Source: Author's Own Computation

Table 12: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
H_t	-7.707525	0.0000	-4.140858	@1%	Stationary
			-3.496960	@5%	Stationary
			-3.177579	@10%	Stationary

Source: Author's Own Computation

Table 13: without intercept and trend & intercept

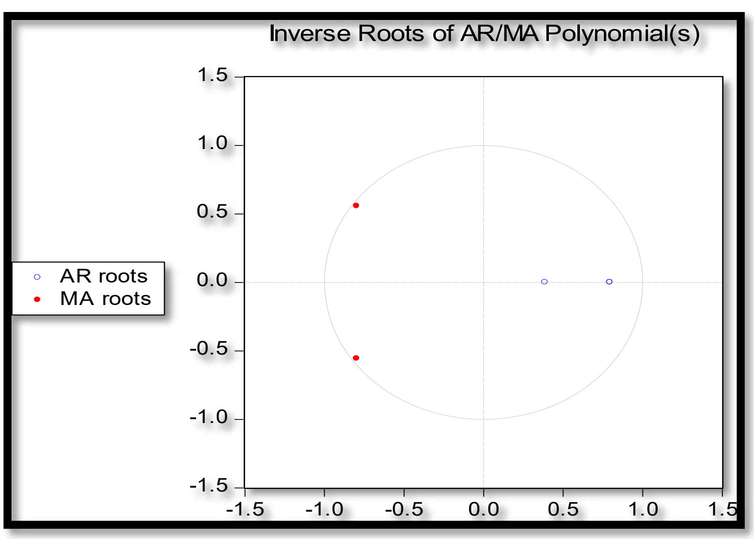
Variable	ADF Statistic	Probability	Critical Values		Conclusion
H_t	-7.853834	0.0000	-2.609324	@1%	Stationary
			-1.947119	@5%	Stationary
			-1.612867	@10%	Stationary

Source: Author's Own Computation

Tables 11, 12 and 13 above; indicate that the residuals of the ARIMA (2, 2, 2) model are stationary.

Stability Test of the ARIMA (2, 2, 2) Model

Figure 5. Inverse roots of the ARIMA (2, 2, 2) model



Source: Author's Own Computation

Figure 5 above shows that the ARIMA (2, 2, 2) model is stable since the corresponding inverse roots of the characteristic polynomial lies in the unit circle.

RESULTS & DISCUSSION

Descriptive Statistics

Table 14. Descriptive Statistics

Description	Statistic
Mean	9519600
Median	9753400
Minimum	3747400
Maximum	16530000
Standard deviation	3817000
Skewness	0.062016
Excess kurtosis	-1.2700

Source: Author's Own Computation

As shown in table 14 above, the mean is positive, i.e. 9519600. The wide gap between the minimum (i.e. 3747400) and the maximum (i.e. 16530000) is consistent with the reality that the POP series is sharply trending upwards as already shown in figure 1 above. The skewness is 0.062016 and the most important thing about it is that it is positive, indicating that the POP series is positively skewed and non-symmetric. Kurtosis is -1.27; showing that the POP series is not normally distributed.

Results Presentation[1]

Table 15. Results

ARIMA (2, 2, 2) Model:

$$\Delta^2 POP_{t-1} = 4516.07 + 1.427\Delta^2 POP_{t-1} - 0.55\Delta^2 POP_{t-2} + 1.053\mu_{t-1} + 0.483\mu_{t-2} \ldots \ldots [5]$$

P: (0.0030) (0.0000) (0.0000) (0.0000) (0.0000)

S. E: (1519.53) (0.1405) (0.1382) (0.1543) (0.1500)

Variable	Coefficient	Standard Error	z	p-value
Constant	4516.07	1519.53	2.972	0.0030***
AR (1)	1.42653	0.140474	10.16	0.0000***
AR (2)	-0.550124	0.138168	-3.982	0.0000***

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

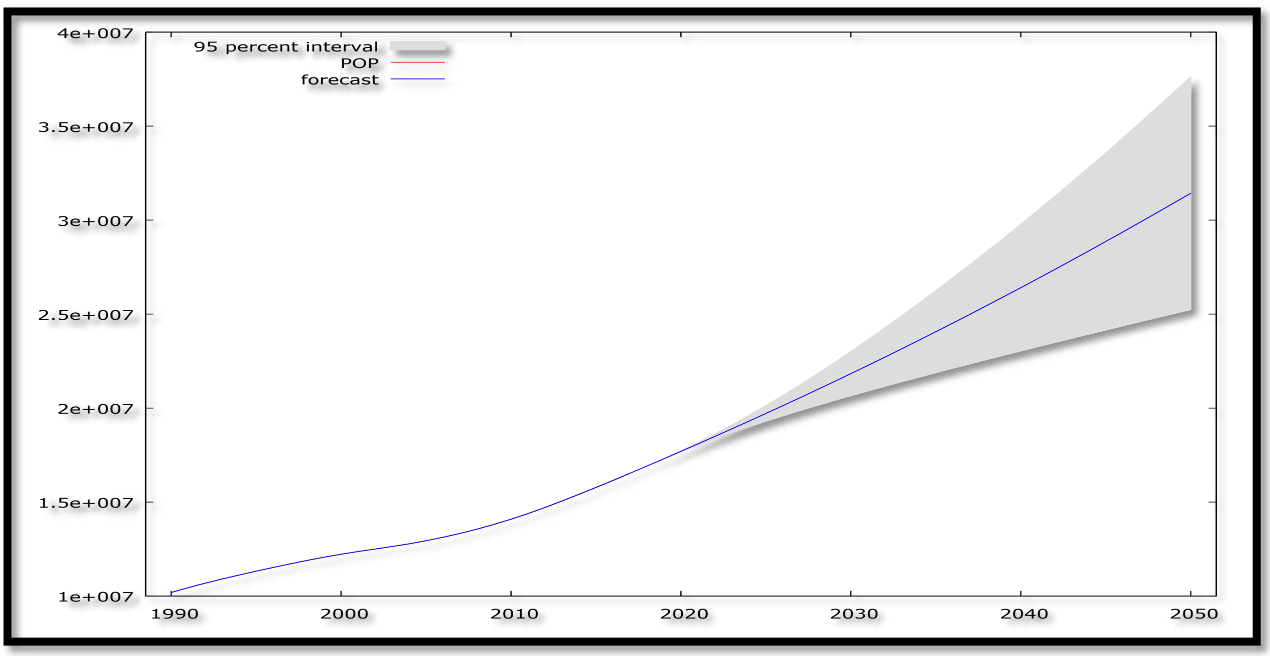
MA (1)	1.05282	0.154306	6.823	0.0000***
MA (2)	0.482762	0.149984	3.219	0.0013***

Source: Author's Own Computation

Interpretation of Results

Equation [5] is the ARIMA (2, 2, 2) model and is explained as follows: all AR and MA coefficients are statistically significant at 1% level of significance. Only the AR (2) coefficient is negative, the rest of the coefficients are positive. Previous period population and previous period unobserved shocks to population are equally important in explaining total population trends in Zimbabwe.

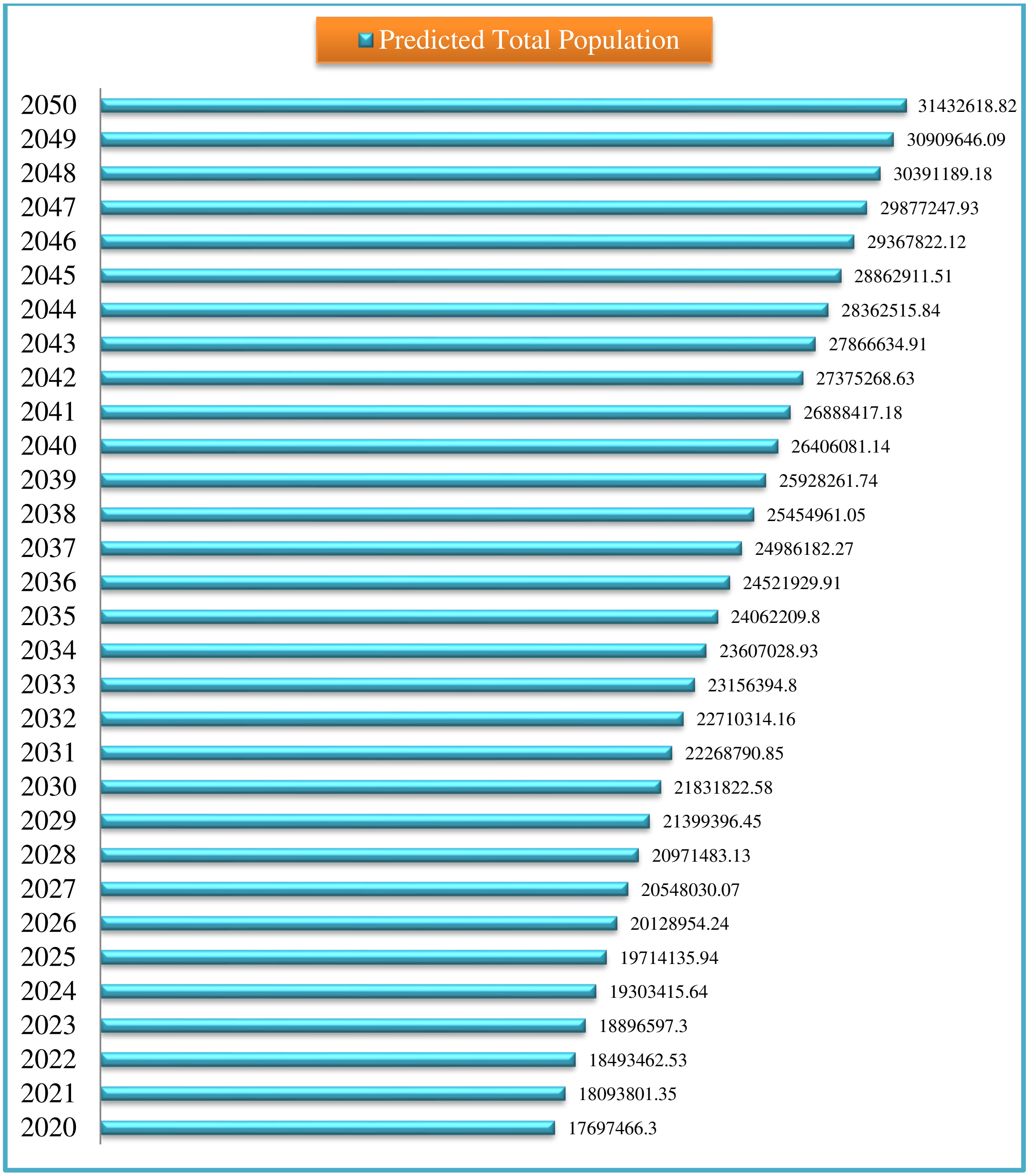
Figure 6. Forecast Graph



Source: Author's Own Computation

With a forecast range of 32 years, i.e., 2018 – 2050; figure 6, clearly indicates that Zimbabwe's total population is indeed set to increase, at least for the next 3 decades.

Figure 7. Predicted out-of-sample total population



Source: Author's Own Computation

With a forecast range of 32 years, i.e., 2018 – 2050; figure 7, clearly shows that total population in Zimbabwe, is indeed set to increase; at least for the next 3 decades. Figure 7 is consistent with figure 6 above.

Discussion of the Results

With a 95% confidence interval of 25 232 416 to 37 632 822 and a projected total population of 31 432 619 people by 2050, the ARIMA (2, 2, 2) model is consistent with the population projections by the UN (2015) which forecasted that Zimbabwe's population will be approximately 29 615 000 by 2050. The presented results of the ARIMA (2, 2, 2) model are also consistent with ZimStat (2016) whose argument is that total population in Zimbabwe will increase due to rising life expectancy levels due to reduced HIV/AIDS related deaths, improved public health systems, expanding education levels, rising incomes and urbanization. According to Nyoni & Bonga (2017), population growth is one of the fundamental factors that directly determine the supply of human resources which are indisputably critical for production. This is in line with theoretical underpinnings by Ahlburg (1998) and Becker *et al* (1999). Thus for a developing country like Zimbabwe, population growth is welcome in the sense that it is pivotal in the country's economic development trajectory. Most economically developed countries, as noted by Nyoni & Bonga (2017); have significantly high populations, for instance; the United States and China. A higher population in the case of Zimbabwe is envisaged to trigger an increase in the need for goods and services and subsequently boost technological advancement. In the long run, labour productivity will improve and Gross Domestic Product (GDP) per capita is expected to increase. These arguments make a lot of sense especially given the findings by Nyoni & Bonga (2017); that an increase in population growth in Zimbabwe will lead to positive economic development.

Policy Implications

i. The government of Zimbabwe must properly deal with health service delivery problems such as health workers' salaries and essential medicines in order to address both maternal deaths as well as crude death rates.
ii. In order to benefit more from increased population, a stable economic landscape is recommended.

CONCLUSION

Zimbabwe is blessed with vast natural resource endowments, most of which lie idle (Nyoni & Bonga, 2017). The predicted increase in population is expected to be an opportunity for industrialization because of adequate supply of labour for production. The study shows that the ARIMA (2, 2, 2) model is not only stable but also the most suitable model to forecast total population in Zimbabwe for the next 3 decades. The model predicts that by 2050, Zimbabwe's total population would be approximately, 31 million people. This is, indeed, a warning signal to policy makers in Zimbabwe, particularly with regards to infrastructural development, e.g schools and hospitals, clinics, road networks, communication networks and so forth. These findings are essential for the government of Zimbabwe, especially when it comes to medium-term and long-term planning.

REFERENCES

[1] Ahlburg, D. A (1998). Julian Simon and the population growth debate, *Population and Development Review*, 24 (7): 317 – 327.

[2] Asteriou, D. & Hall, S. G. (2007). Applied Econometrics: a modern approach, Revised Edition, *Palgrave MacMillan*, New York, United States of America.

[3] Ayele, A. W & Zewdie, M. A (2017). Modeling and forecasting Ethiopian human population size and its pattern, *International Journal of Social Sciences, Arts and Humanities*, 4 (3): 71 – 82.

[4] Becker, G., Glaeser, E., & Murphy, K (1999). Population and economic growth, *American Economic Review*, 89 (2): 145 – 149.

[5] Beg, A. B. M. R. A & Islam, M. R (2016). Forecasting and modeling population growth of Bangladesh, *American Journal of Mathematics and Statistics*, 6 (4): 190 – 195.

[6] Du Preez, J. & Witt, S. F. (2003). Univariate and multivariate time series forecasting: An application to tourism demand, *International Journal of Forecasting*, 19 (3): 435 – 451.

[7] Goh, C. & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic non-stationary seasonality and intervention, *Tourism Management*, 23 (5): 499 – 510.

[8] Malthus, T (1798). An essay of the principle of population, *Pickering*, London, United Kingdom.

[9] Nyoni, T & Bonga, W. G (2017). Population growth in Zimbabwe: A Threat to Economic Development? *DRJ – Journal of Economics and Finance*, 2 (6): 29 – 39.

[10] Nyoni, T (2018l). Modeling Forecasting Naira / USD Exchange Rate in Nigeria: a Box – Jenkins ARIMA approach, *University of Munich Library – Munich Personal RePEc Archive (MPRA)*, Paper No. 88622, Munich, Germany.

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

[12] Nyoni, T. (2018i). Box – Jenkins ARIMA Approach to Predicting net FDI inflows in Zimbabwe, *Munich University Library – Munich Personal RePEc Archive (MPRA)*, Paper No. 87737, Munich, Germany.

[13] Solow, R (1956). Technical change and the aggregate population function, *Review of Economics and Statistics*, 39 (1): 312 – 320.

[14] Song, H., Witt, S. F. & Jensen, T. C. (2003b). Tourism forecasting: accuracy of alternative econometric models, *International Journal of Forecasting*, 19 (2): 123 – 141.

[15] United Nations (2015). World Population Prospects: The 2015 Revision, Key Findings and Advance Tables, *Department of Economic and Social Affairs*, Population Division, Working Paper No. ESA/P/WP/241, New York, United States of America.

[16] Zakria, M & Muhammad, F (2009). Forecasting the population of Pakistan using ARIMA models, *Pakistan Journal of Agricultural Sciences*, 46 (3): 214 – 223.

[17] ZimStats (2015). Population Projections Thematic Report, *Zimbabwe National Statistics Agency (ZimStats)*, Harare, Zimbabwe.