What will be Botswana's population in 2050? Evidence from the Box-Jenkins approach

NYONI, THABANI

University of Zimbabwe

5 May 2019

What Will Be Botswana's Population In 2050? Evidence From The Box – Jenkins Approach

Nyoni, Thabani

Department of Economics

University of Zimbabwe

Harare, Zimbabwe

Email: nyonithabani35@gmail.com

Abstract

Employing annual time series data on total population in Botswana from 1960 to 2017, I model and forecast total population over the next 3 decades using the Box – Jenkins ARIMA technique. Diagnostic tests such as the ADF tests show that Botswana annual total population is neither I (1) nor I (2) but for simplicity purposes, the researcher has assumed it is I (2). Based on the AIC, the study presents the ARIMA (3, 2, 1) model as the optimal model. The diagnostic tests further indicate that the presented model is indeed stable. The results of the study reveal that total population in Botswana will continue to rise in the next three decades and in 2050 Botswana's total population will be approximately 3 665 140 people. In order to benefit from an increase in total population in Botswana, 3 policy recommendations have been suggested.

Key Words: Botswana, Forecasting, Population

JEL Codes: C53, Q56, R23

INTRODUCTION

As the 21st century began, the world's population was estimated to be almost 6.1 billion people (Tartiyus *et al*, 2015). Projections 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. Over 90% of that population will inhabit the developing world (Todaro & Smith, 2006). The problem of population growth is basically not a problem of numbers but that of human welfare as it affects the provision of welfare and development. The consequences of rapidly growing population manifests heavily on species extinction, deforestation, desertification, climate change and the destruction of natural ecosystems on one hand; and unemployment, pressure on housing, transport traffic congestion, pollution and infrastructure security and stain on amenities (Dominic *et al*, 2016). Population figures are required for planning in every sector of a population (Statistics Botswana, 2015). In Botswana, just like in any other part of the world, population modeling and forecasting is important for policy dialogue. This study endeavors to model and forecast population of Botswana using the Box-Jenkins ARIMA technique.

REVIEW OF PREVIOUS STUDIES

Zakria & Muhammad (2009), in Pakistan; forecasted population using ARIMA models, and relied on a data set ranging from 1951 - 2007; and established that the ARIMA (1, 2, 0) model was the suitable model for forecasting total population in Pakistan. Beg & Islam (2016) looked at population growth of Bangladesh using an Autoregressive Time Trend (ATT) model based on a data set ranging over 1965 – 2003 and concluded that there will be a downward population growth for Bangladesh for the extended period up to 2043. Ayele & Zewdie (2017) studied human population size and its pattern in Ethiopia using ARIMA models and employing annual data from 1961 - 2009 and found out that the most suitable model for modeling and forecasting population in Ethiopia was the ARIMA (2, 1, 2) model. In the case of Botswana, the researcher will employ the Box-Jenkins ARIMA technique for the data set ranging from 1960 - 2017.

MATERIALS & METHODS

ARIMA Models

ARIMA models are often considered as delivering more accurate forecasts then econometric techniques (Song *et al*, 2003b). ARIMA models outperform multivariate models in forecasting performance (du Preez & Witt, 2003). Overall performance of ARIMA models is 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 based on the principle of parsimony (Asteriou & Hall, 2007). The general form of the ARIMA (p, d, q) can be represented as follows:

$$\emptyset_p(L)\Delta^d P_t = \theta_q(L)\mu_t \quad \ldots \ldots [1]$$

Where P_t is the total population in Botswana at time, t.

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, 2018).

Data Collection

This study is based on 58 observations of annual total population in Botswana. All the data was gathered from the World Bank.

Diagnostic Tests & Model Evaluation

Stationarity Tests: Graphical Analysis

Figure 1

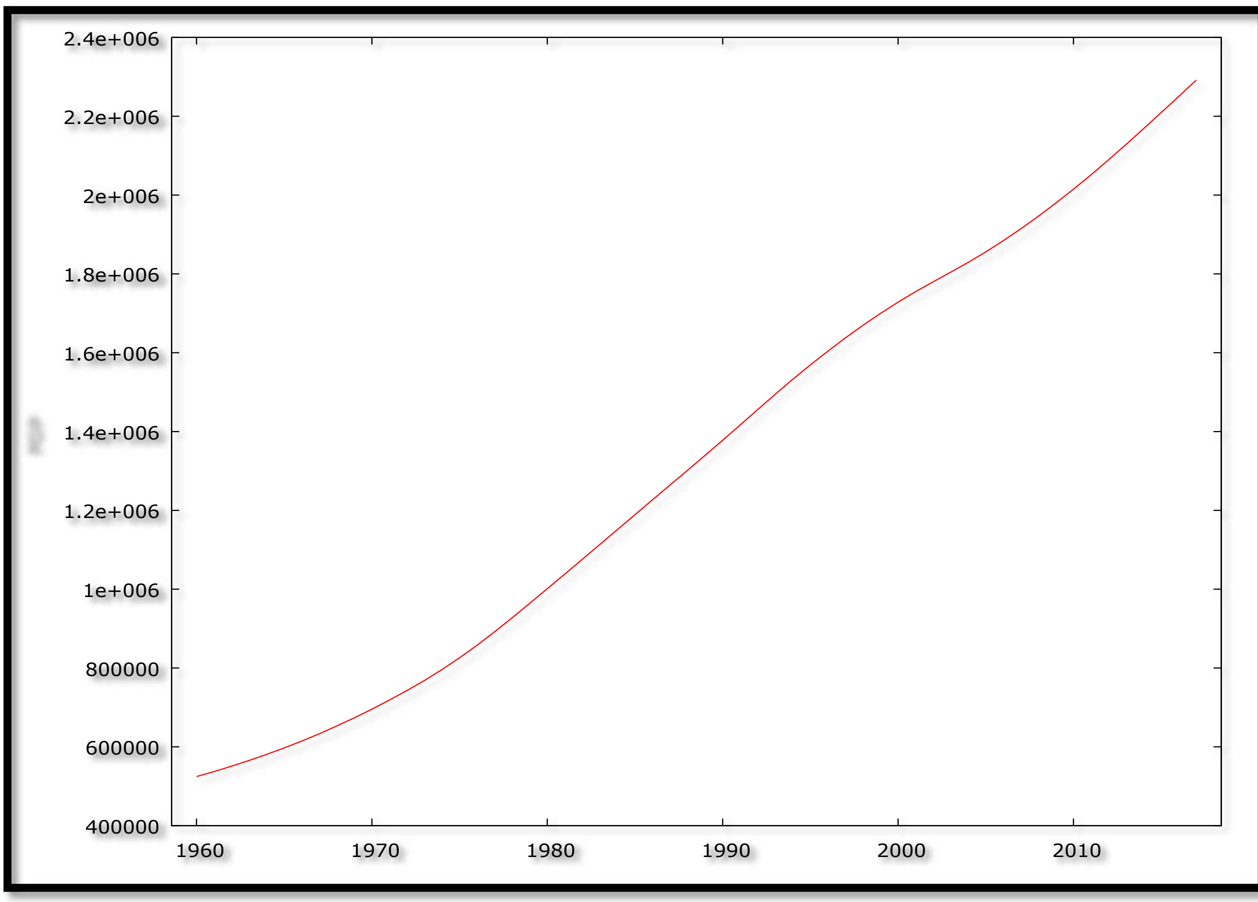
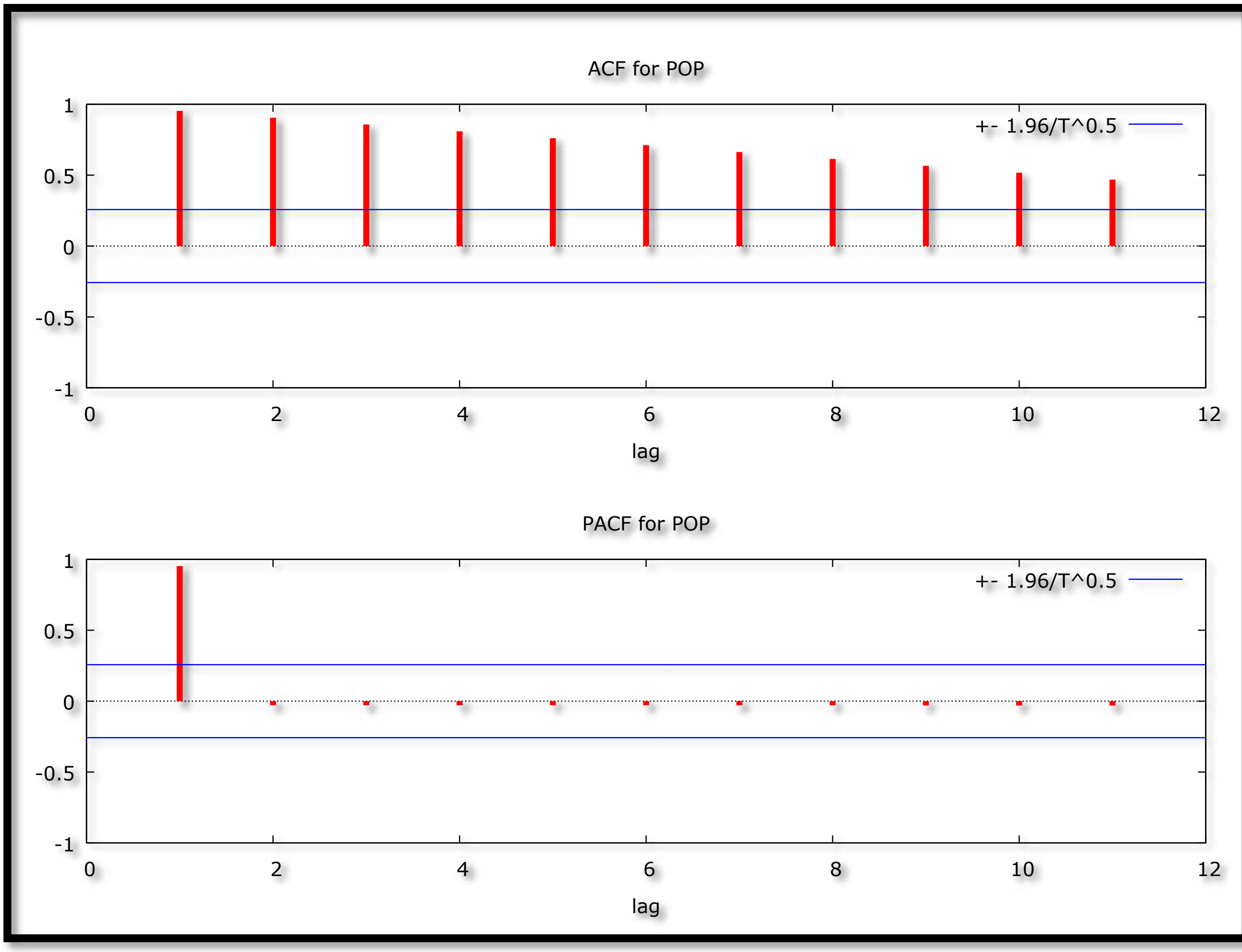


Figure 1 above indicates that the Botswana POP variable is not stationary since it is trending upwards over the period 1960 – 2017. This basically proves that the mean and varience of POP is changing over time.

The Correlogram in Levels

Figure 2



The ADF Test

Table 1: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	1.347594	0.9986	-3.560019	@1%	Not stationary
			-2.917650	@5%	Not stationary
			-2.596689	@10%	Not stationary

Table 2: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-2.982960	0.1466	-4.140858	@1%	Not stationary
			-3.496960	@5%	Not stationary
			-3.177579	@10%	Not stationary

Table 3: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	1.317657	0.9509	-2.609324	@1%	Not stationary
			-1.947119	@5%	Not stationary
			-1.612867	@10%	Not stationary

The Correlogram (at 1st Differences)

Figure 3

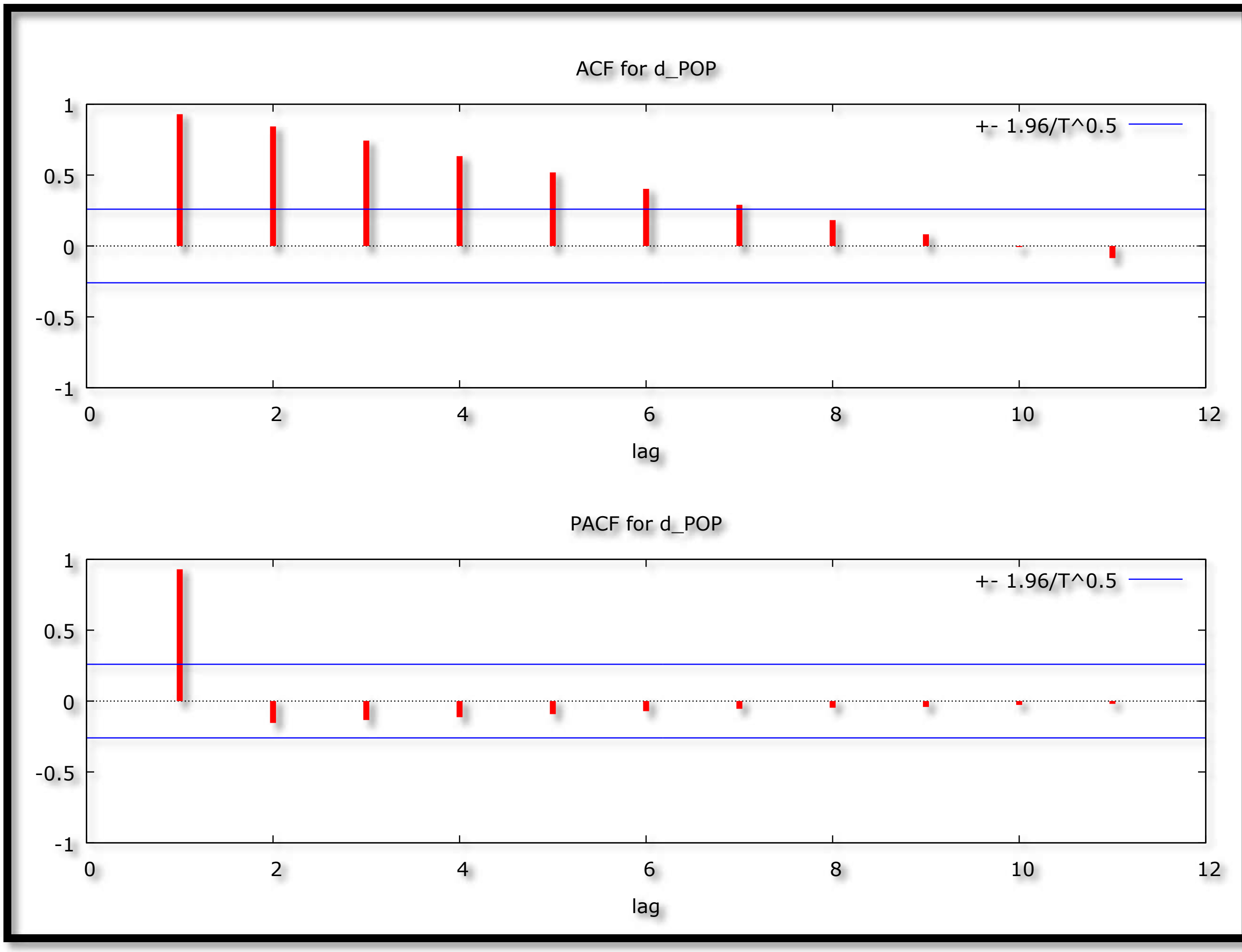


Table 4: 1st Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-2.625230	0.0944	-3.560019	@1%	Not stationary
			-2.917650	@5%	Not stationary
			-2.596689	@10%	Not stationary

Table 5: 1st Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-5.293058	0.0004	-4.170583	@1%	Stationary
			-3.510740	@5%	Stationary
			-3.185512	@10%	Stationary

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

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	0.390357	0.7930	-2.610192	@1%	Not stationary

		-1.947248	@5%	Not stationary
		-1.612797	@10%	Not stationary

Figures above, i.e. 2 and 3 and tables above, i.e. 1, 2, 3, 4 and 6 indicate that the Botswana POP series is not stationary at both levels and in first differences. However, table 5 indicates that the POP series is stationary at all levels. Due to the overwhelming inconsistency in the above stationarity tests, the researcher will go ahead and further test for stationarity in second differences in order to verify the stationarity of the Botswana POP series.

The Correlogram in (2nd Differences)

Figure 4

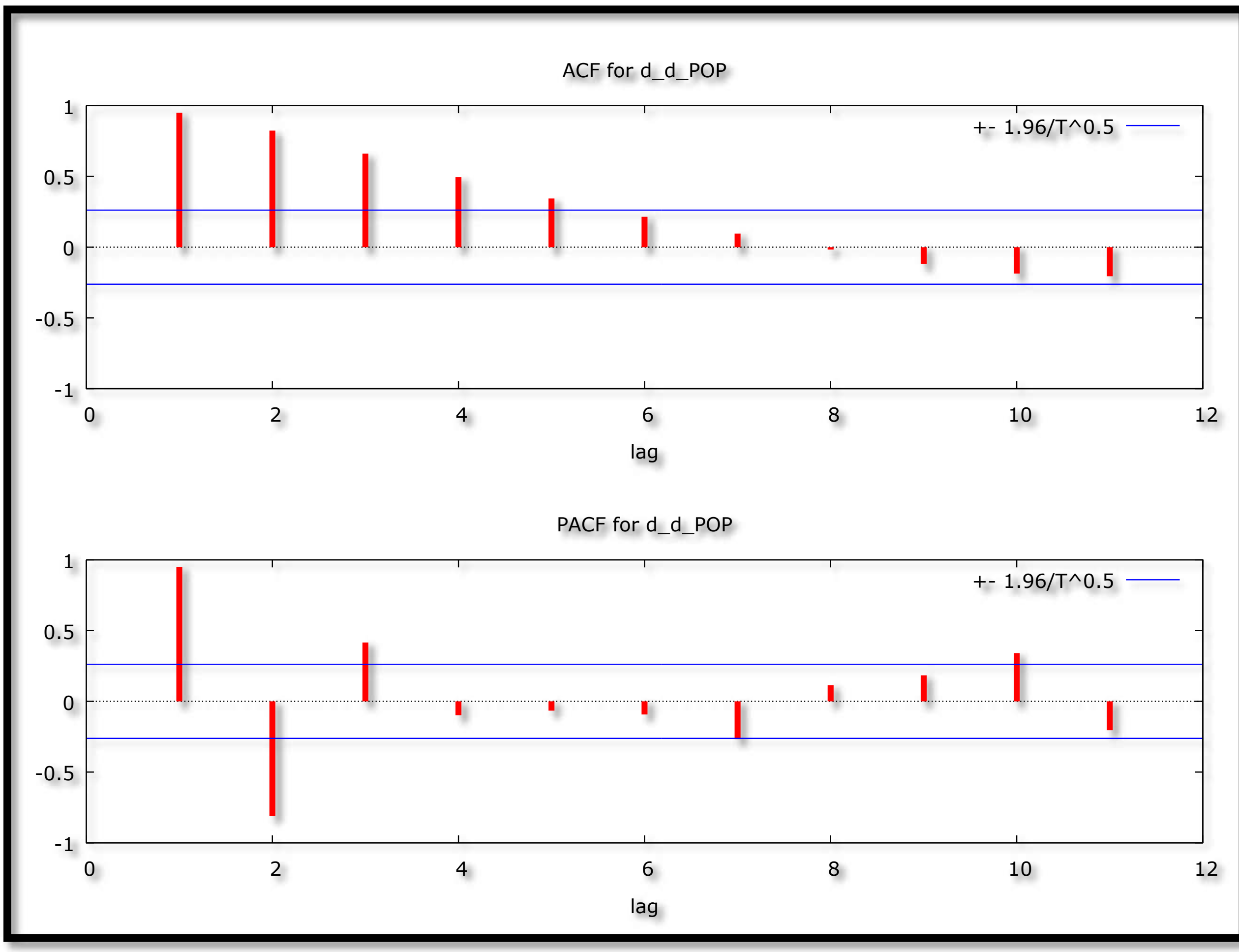


Table 7: 2nd Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.942591	0.3105	-3.584743	@1%	Not stationary
			-2.928142	@5%	Not stationary
			-2.602225	@10%	Not stationary

Table 8: 2nd Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.644556	0.7590	-4.175640	@1%	Not stationary
			-3.513075	@5%	Not stationary
			-3.186854	@10%	Not stationary

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

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.902523	0.0552	-2.617364	@1%	Not stationary
			-1.948313	@5%	Not stationary
			-1.612229	@10%	Stationary

Figure 4 and tables 7 – 9 demonstrate that the Botswana POP series is not stationary in second differences, but for simplicity purposes, the study will assume that it's stationary after taking second differences [I(2)].

Evaluation of ARIMA models (without a constant)

Table 10

Model	AIC	U	ME	MAE	RMSE	MAPE
ARIMA (1, 2, 1)	769.4967	0.0061509	26.319	173.2	242.61	0.015358
ARIMA (1, 2, 0)	818.9994	0.0095278	30.882	249.73	361.05	0.021143
ARIMA (2, 2, 0)	759.1453	0.0058232	50.243	164.72	226.49	0.014927
ARIMA (0, 2, 1)	887.5321	0.023295	272.16	551.85	645.52	0.050733
ARIMA (3, 2, 1)	**738.7884**	0.004669	33.013	139.74	194.06	0.01277
ARIMA (4, 2, 1)	740.4827	0.0046889	34.191	139.39	193.69	0.012785
ARIMA (5, 2, 1)	742.4184	0.0046794	33.674	139.19	193.62	0.012753
ARIMA (3, 2, 0)	742.0285	0.0047007	31.889	141.25	200.31	0.012805
ARIMA (4, 2, 0)	739.4796	0.0047038	35.58	140.01	194.86	0.012868
ARIMA (5, 2, 0)	740.9157	0.0046748	33.812	139.56	194.2	0.012785

A model with a lower AIC value is better than the one with a higher AIC value (Nyoni, 2018). Theil's U must lie between 0 and 1, of which the closer it is to 0, the better the forecast method (Nyoni, 2018). The study will consider the AIC and the Theil's U in order to choose the best model. Therefore, for forecasting total population in Botswana, the ARIMA (3, 2, 1) model is carefully selected.

Residual & Stability Tests

ADF Tests of the Residuals of the ARIMA (3, 2, 1) Model

Table 11: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
K_t	-3.831159	0.0053	-3.592462	@1%	Stationary
			-2.931404	@5%	Stationary
			-2.603944	@10%	Stationary

Table 12: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values	Conclusion

K_t	-3.788582	0.0268	-4.186481	@1%	Not stationary
			-3.518090	@5%	Stationary
			-3.189732	@10%	Stationary

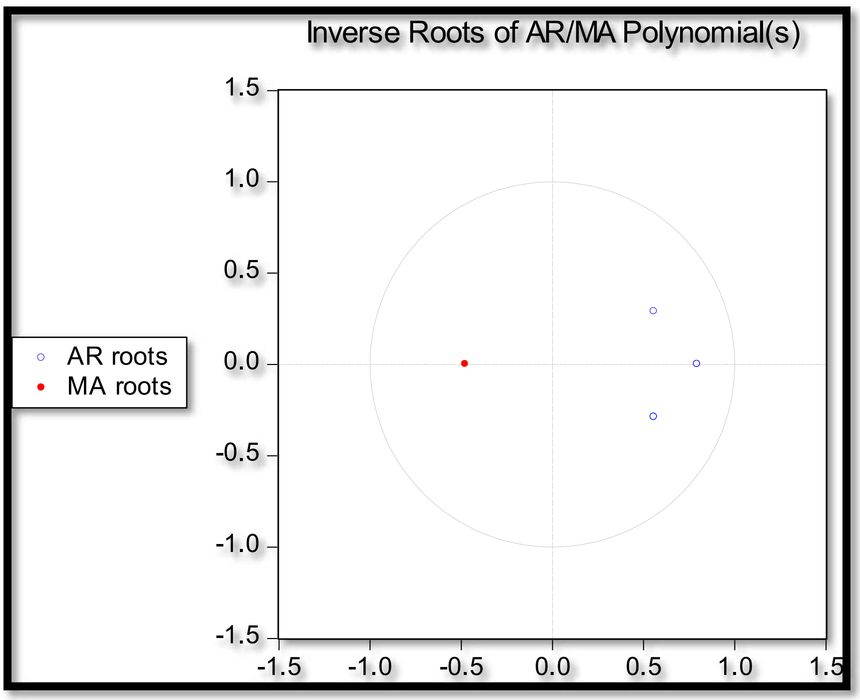
Table 13: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
K_t	-3.886665	0.0003	-2.619851	@1%	Stationary
			-1.948686	@5%	Stationary
			-1.612036	@10%	Stationary

Tables 11, 12 and 13 demonstrate that the residuals of the ARIMA (3, 2, 1) model are stationary.

Stability Test of the ARMIA (3, 2, 1) Model

Figure 5



Since the corresponding inverse roots of the characteristic polynomial lie in the unit circle, it illustrates that the chosen ARIMA (3, 2, 1) model is indeed stable.

RESULTS

Descriptive Statistics

Table 14

Description	Statistic
Mean	1326900
Median	1320700
Minimum	524550

Maximum	2291700
Standard deviation	550590
Skewness	0.084825
Excess kurtosis	-1.3248

As shown above, the mean is positive, i.e. 1326900. The wide gap between the minimum (i.e 524550) and the maximum (i.e. 2291700) is consistent with the reality that the Botswana POP series is sharply trending upwards over the period under study. The skewness is 0.084825 and the most striking characteristic is that it is positive, indicating that the Botswana POP series is positively skewed and non-symmetric. Excess kurtosis is -1.3248; showing that the Botswana POP series is not normally distributed.

Results Presentation[1]

Table 15

ARIMA (3, 2, 1) Model:

$$\Delta^2 POP_{t-1} = 1.89\Delta^2 POP_{t-1} - 1.249\Delta^2 POP_{t-2} + 0.301\Delta^2 POP_{t-3} + 0.469\mu_{t-1} \ldots\ldots\ldots [2]$$

P: (0.0000) (0.0005) (0.1057) (0.0162)

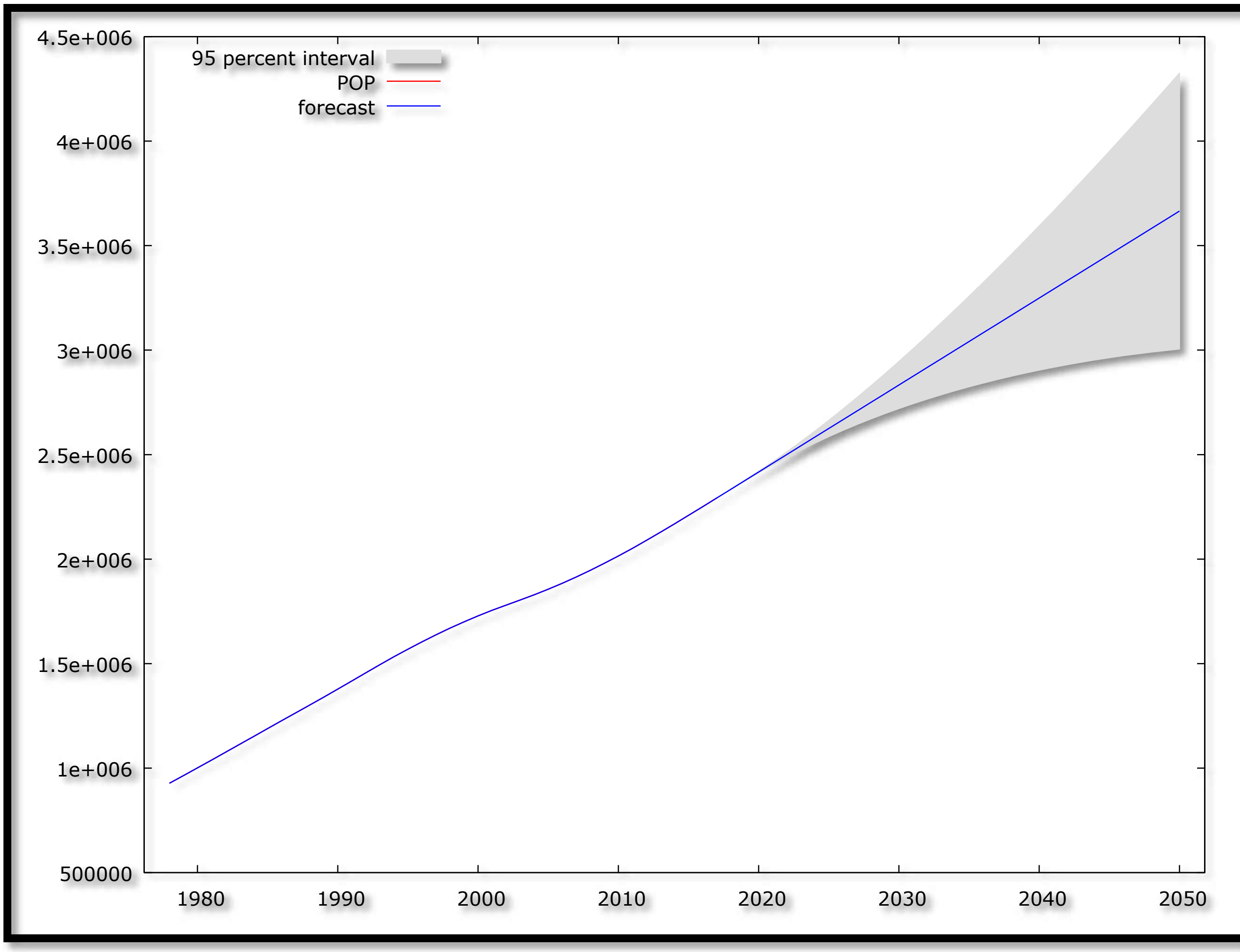
S. E: (0.1995) (0.3600) (0.1862) (0.1948)

Variable	Coefficient	Standard Error	z	p-value
AR (1)	1.89007	0.199500	9.474	0.0000***
AR (2)	-1.24908	0.359936	-3.47	0.0005***
AR (3)	0.301295	0.186238	1.618	0.1057
MA (1)	0.468669	0.194837	2.405	0.0162**

Forecast Graph

Figure 6

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



Predicted Total Population (2018 – 2050)

Table 16

Year	Prediction	Std. Error	95% Confidence Interval
2018	2333261.99	153.461	2332961.21 - 2333562.77
2019	2374951.05	686.275	2373605.98 - 2376296.13
2020	2416658.62	1811.811	2413107.54 - 2420209.70
2021	2458351.40	3685.705	2451127.56 - 2465575.25
2022	2500019.67	6397.017	2487481.75 - 2512557.60
2023	2541665.65	9978.929	2522107.30 - 2561223.99
2024	2583295.64	14425.922	2555021.35 - 2611569.93
2025	2624915.91	19709.867	2586285.28 - 2663546.53

2026	2666531.02	25792.007	2615979.61 - 2717082.42
2027	2708143.72	32630.558	2644189.01 - 2772098.44
2028	2749755.40	40184.814	2670994.61 - 2828516.19
2029	2791366.57	48416.951	2696471.09 - 2886262.05
2030	2832977.35	57292.519	2720686.08 - 2945268.63
2031	2874587.72	66780.311	2743700.72 - 3005474.72
2032	2916197.64	76852.015	2765570.46 - 3066824.82
2033	2957807.11	87481.848	2786345.83 - 3129268.38
2034	2999416.16	98646.233	2806073.09 - 3192759.22
2035	3041024.86	110323.558	2824794.66 - 3257255.06
2036	3082633.27	122493.972	2842549.49 - 3322717.04
2037	3124241.45	135139.226	2859373.43 - 3389109.46
2038	3165849.45	148242.532	2875299.42 - 3456399.47
2039	3207457.31	161788.435	2890357.80 - 3524556.82
2040	3249065.07	175762.694	2904576.52 - 3593553.62
2041	3290672.74	190152.168	2917981.34 - 3663364.14
2042	3332280.35	204944.719	2930596.08 - 3733964.62
2043	3373887.90	220129.109	2942442.78 - 3805333.03
2044	3415495.42	235694.919	2953541.86 - 3877448.97
2045	3457102.89	251632.466	2963912.32 - 3950293.46
2046	3498710.35	267932.733	2973571.84 - 4023848.85
2047	3540317.78	284587.308	2982536.90 - 4098098.65
2048	3581925.19	301588.324	2990822.94 - 4173027.44
2049	3623532.59	318928.410	2998444.39 - 4248620.79
2050	3665139.98	336600.643	3005414.84 - 4324865.11

Figure 6 (with a forecast range from 2018 – 2050) and table 16, clearly show that Botswana population is indeed set to continue rising sharply, in the next 3 decades. With a 95% confidence interval of 3005415 to 4324865 and a projected total population of 3665140 by 2050, the chosen

ARIMA (3, 2, 1) model is consistent with the population projections by the UN (2015) which forecasted that Botswana's population will be approximately 3389000 by 2050.

Policy Implications

i. The government of Botswana should invest more in infrastructural development in order to cater for the expected increase in total population.
ii. The projected increase in total population justifies the need for more companies to provide for the expected increase in demand for goods and services.
iii. There is need to improve health service delivery in Botswana in order to ensure a healthier society, especially in light of such a likely increase in total population.

CONCLUSION

The ARIMA (3, 2, 1) model is not only stable but also the most suitable model to forecast the population of Botswana for the next 3 decades. The model predicts that by 2050, Botswana's population would be approximately, 3.6 million people. This is a warning signal to policy makers in Botswana, especially with regards to infrastructural development, for example, schools and hospitals. These findings are essential for the government of Botswana, especially when it comes to long-term planning.

REFERENCES

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

[2] 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.

[3] 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.

[4] Dominic, A., Oluwatoyin, M. A., & Fagbeminiyi, F. F (2016). The determinants of population growth in Nigeria: a co-integration approach, *The International Journal of Humanities and Social Studies*, 4 (11): 38 – 44.

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

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

[7] Nyoni, T (2018). 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.

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

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

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

[11] Statistics Botswana (2015). Botswana Population Projections (2011 – 2026), *Statistics Botswana*, Gaborone.

[12] Tartiyus, E. H., Dauda, T. M., & Peter, A (2015). Impact of population growth on economic growth in Nigeria, *IOSR Journal of Humanities and Social Science (IOSR-JHSS)*, 20 (4): 115 – 123.

[13] Todaro, M & Smith, S (2006). Economic Development, 9th Edition, *Vrinda* Publications, New Delhi.

[14] 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.

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