Will the United States of America (USA) be a beneficiary of the Alburg (1998) and Becker et al (1999) prophecies? Recent insights from the Box-Jenkins ARIMA approach

NYONI, THABANI

University of Zimbabwe, Department of Economics

19 February 2019

Will The United States Of America (USA) Be A Beneficiary Of The Ahlburg (1998) and Becker *et al* (1999) Prophecies? Recent Insights From The Box – Jenkins ARIMA 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 the USA from 1960 to 2017, we model and forecast total population over the next 3 decades using the Box – Jenkins ARIMA approach. Diagnostic tests show that USA annual total population data is I (2). Based on the AIC, the study presents the ARIMA (0, 2, 3) model. The diagnostic tests indicate that the presented model is very stable and quite suitable. The results of the study reveal that total population in USA will continue to sharply rise in the next three decades. Considering a highly educated labor force, coupled with latest technological advancements, USA is likely to be one of the first beneficiaries of the Ahlburg (1998) and Becker *et al* (1999) prophecies. In order to stay in the realm of the aforementioned prophecies, USA should take note of the 3-fold policy recommendations put forward.

Key Words: Population, Forecasting, USA

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). The need for population forecasts is hardly disputed. In politics, in public administration, and in business, far-reaching decisions are made which depend on the future development of the population (Pflaumer, 2012). Population modeling and forecasting in the US just like in ay other country; is important for policy dialogue. This study attempts to model and forecast population of the US using the Box-Jenkins ARIMA technique.

LITERATURE REVIEW

Theoretical Literature Review

The population theory propounded by Malthus (1798) posits that population growth is really bad for economic growth and development and Malthus (1798) attributes this to the argument that human population grows geometrically while the means of subsistance grows arithmetically being subject to the law of diminishing returns. The applicability of the Malthusian population prophecy is not universal, in the USA, this prophecy has arguably tumbled; primarily due to a highly educated labour force as well as technological advancements and innovation. In a slight disagreement with the basic propositions of the Malthus (1798) population theory, Solow (1956) averred that an increase in the "population growth rate" not in the "population level"; would reduce the capital per worker as well as the steady-state output per worker and concluded that higher population growth could harm productivity and economic growth. Ahlburg (1998) and Becker *et al* (1999) argued against Solow (1956) and Malthus (1798) and tried to show that population growth is not always bad for growth. Ahlburg (1998) pointed that an increase in population growth leads to an increase the need for goods and services through the "technology-pushed" and the "demand-pulled" channels while Becker *et al* (1999), basically in the synonymous line of thought; stressed that high population growth rate apparently induces high labour force which is the source of real wealth.

Empirical Literature Review

Pflaumer (1992) employed the Box-Jenkins technique for forecasting the US population and revealed that the US population can be satisfactorily described by an ARIMA (2, 2, 0) process and consequently confirmed that this model is equivalent to a Parabolic Trend model or Stevens model when making long-term population forecasts. Zakria & Muhammad (2009) analyzed population dynamics in Pakistan using Box-Jenkins ARIMA models, and relied on a data set ranging from 1951 to 2007; and concluded that the ARIMA (1, 2, 0) model was the best model. Haque *et al* (2012) studied Bangladesh population projections using the Logistic Population model with a data set ranging from 1991 to 2006 and established that the Logistic Population model has the best fit for population growth in Bangladesh. Pflaumer (2012), in another US study, forecasted population using the Gompertz Growth Curve employing data over the period 1890 – 2010 and basically established that the accuracy of some simple time series models is better than the accuracy of more complex models. Ayele & Zewdie (2017) investigated human population size and its pattern in Ethiopia using Box-Jenkins ARIMA models and employing annual data from 1961 to 2009 and revealed that the optimal model for modeling and forecasting population in Ethiopia was the ARIMA (2, 1, 2) model. In this piece of work, the Box-Jenkins ARIMA technique will be employed for the data set ranging from 1960 to 2017.

MATERIALS & METHODS

The Autoregressive Integrated Moving Average (ARIMA) model

ARIMA models are a set of models that describe the process (for example, POP_t) as a function of its own lags and white noise process (Box & Jenkins, 1974). Making predicting in time series using univariate approach is best done by employing the ARIMA models (Alnaa & Ahiakpor, 2011). A stochastic process POP_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 $\Delta^d POP_t$ satisfies and ARMA (p, q) process; then the sequence of POP_t also satisfies the ARIMA (p, d, q) process such that:

$$\Delta^d POP_t = \sum_{i=1}^{p} \beta_i \Delta^d POP_{t-i} + \sum_{i=1}^{q} \alpha_i \mu_{t-i} + \mu_t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [1]$$

which we can also re – write using the lag operator (L) notation as follows:

$$\Delta^d POP_t = \sum_{i=1}^{p} \beta_i \Delta^d L^i POP_t + \sum_{i=1}^{q} \alpha_i L^i \mu_t + \mu_t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [2]$$

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

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).

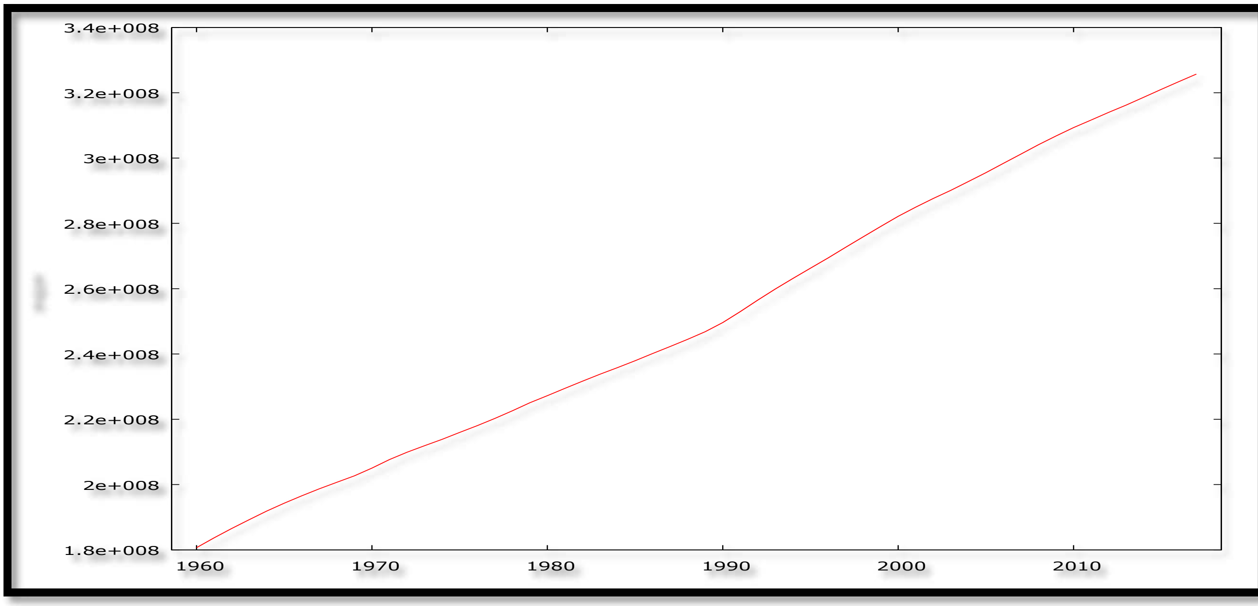
Data Collection

This paper is based on 58 observations of annual total population in the United States of America (USA), i.e. from 1960 – 2017. All the data was taken from the World Bank online database. The Word Bank online database is a reliable source of various macroeconomic data on literally all countries in the world; therefore the author chose this source on the basis of its credibility and integrity.

Diagnostic Tests & Model Evaluation

Stationarity Tests: Graphical Analysis

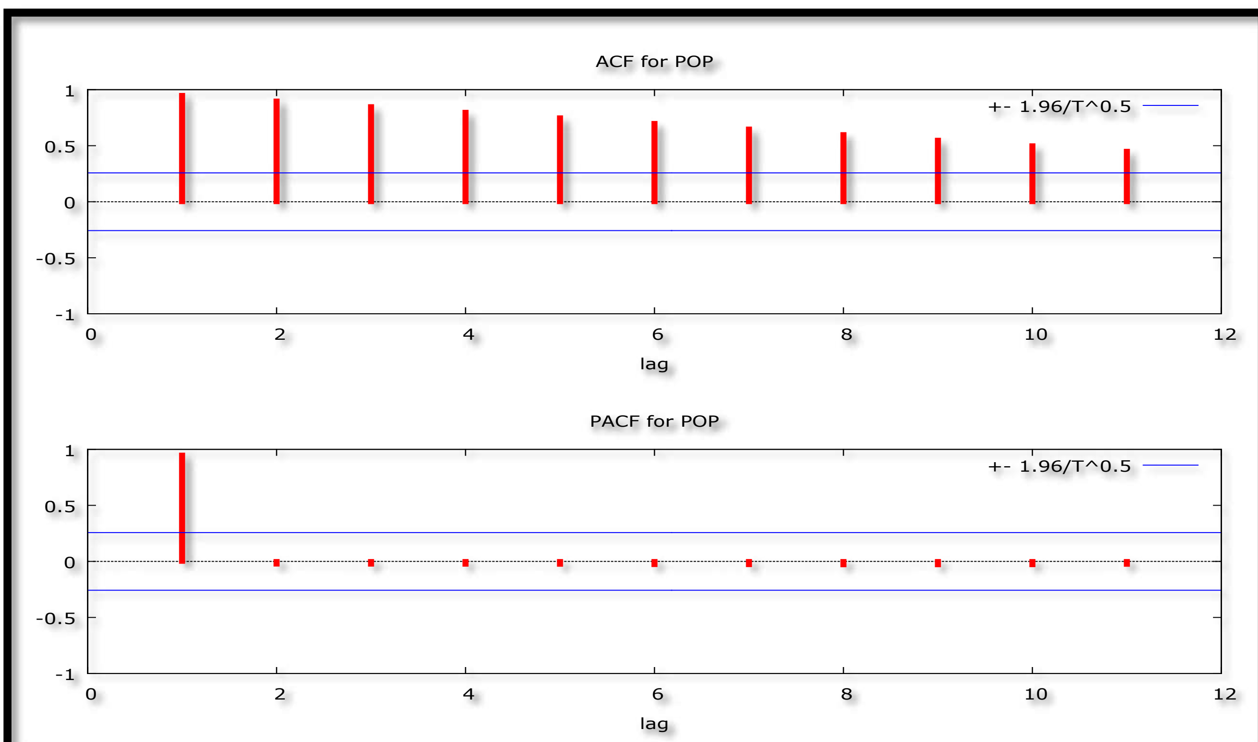
Figure 1



The POP variable, graphically shown above; is not stationary since it is trending upwards over the period 1960 – 2017 and this actually implies 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	0.749822	0.9922	-3.555023	@1%	Not stationary
			-2.915522	@5%	Not stationary
			-2.595565	@10%	Not stationary

Table 2: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-2.235023	0.4612	-4.133838	@1%	Not stationary
			-3.493692	@5%	Not stationary
			-3.175693	@10%	Not stationary

Table 3: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	1.914356	0.9857	-2.607686	@1%	Not stationary
			-1.946878	@5%	Not stationary
			-1.612999	@10%	Not stationary

The Correlogram (at 1st Differences)

Figure 3

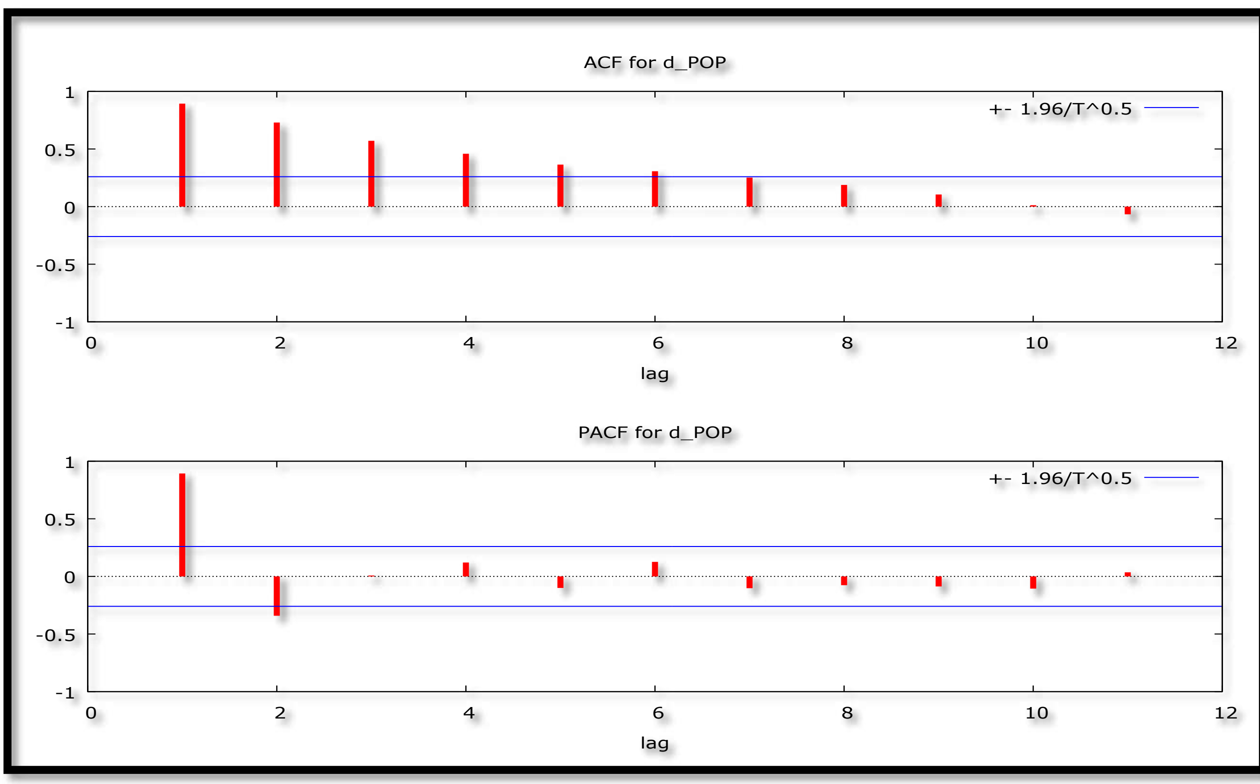


Table 4: 1st Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-2.434220	0.1373	-3.555023	@1%	Not stationary
			-2.915522	@5%	Not stationary
			-2.595565	@10%	Not stationary

Table 5: 1st Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-2.589288	0.2866	-4.133838	@1%	Not stationary
			-3.493692	@5%	Not stationary
			-3.175693	@10%	Not stationary

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

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-0.604728	0.4509	-2.607686	@1%	Not stationary
			-1.946878	@5%	Not stationary
			-1.612999	@10%	Not stationary

As illustrated above in figures 2 and 3 as well as tables 1 – 6, the POP series is not stationary at both levels and in first differences.

The Correlogram in (2nd Differences)

Figure 4

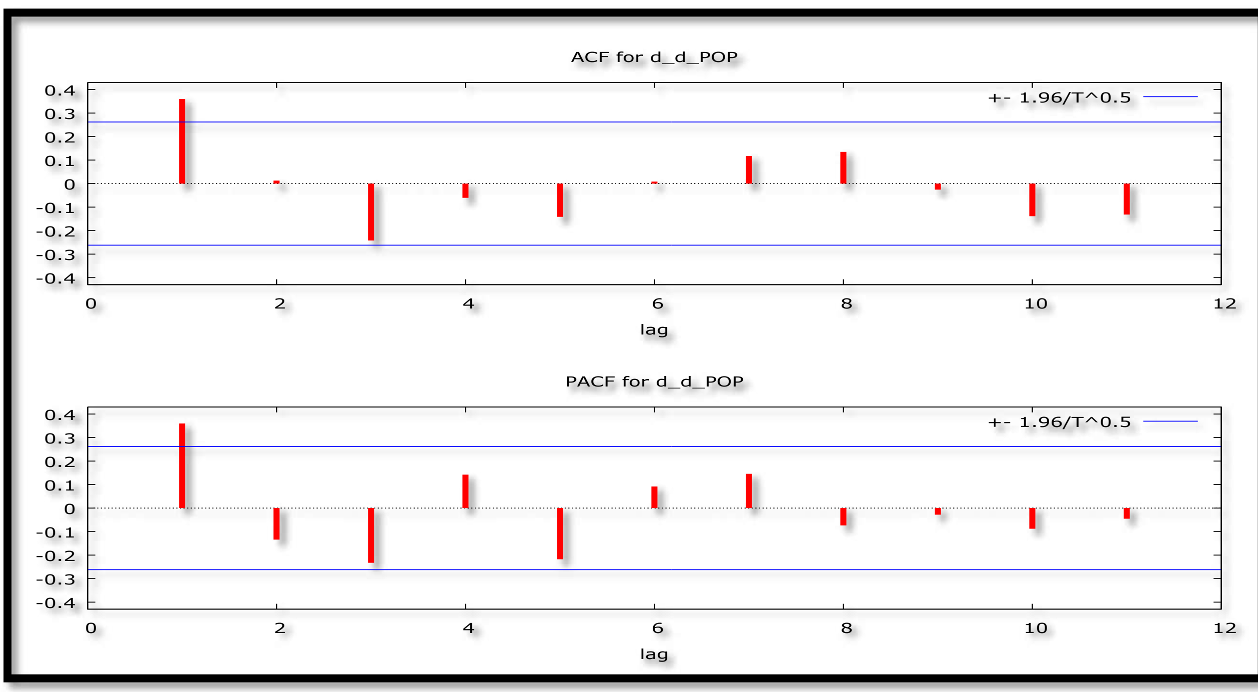


Table 7: 2nd Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-5.031552	0.0001	-3.555023	@1%	Stationary
			-2.915522	@5%	Stationary
			-2.595565	@10%	Stationary

Table 8: 2nd Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-4.821125	0.0014	-4.140858	@1%	Stationary
			-3.496960	@5%	Stationary
			-3.177579	@10%	Stationary

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

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-5.071680	0.0000	-2.607686	@1%	Stationary
			-1.946878	@5%	Stationary
			-1.612999	@10%	Stationary

Figure 4 above indicates that most of the autocorrelation coefficients are now closer to zero, with the exception of the first lag. This is generally a feature of a stationary series. Tables 7 – 9, confirm that the POP series became stationary after taking second differences and hence it is I (2).

Evaluation of ARIMA models (without a constant)

Table 10

Model	AIC	U	ME	MAE	RMSE	MAPE
ARIMA (1, 2, 0)	1511.588	0.069573	-8399.3	0.0000129	0.000017	0.052965
ARIMA (2, 2, 0)	1512.644	0.068464	-9385.5	0.000013	0.0000168	0.053349
ARIMA (3, 2, 0)	1511.639	0.067863	-11412	0.0000123	0.0000164	0.050821
ARIMA (0, 2, 1)	1512.115	0.069426	-9824.8	0.0000134	0.000017	0.054783
ARIMA (0, 2, 2)	1511.106	0.068451	-7407.9	0.0000124	0.0000166	0.05091
ARIMA (0, 2, 3)	**1507.380**	**0.064673**	-10430	0.0000117	0.0000158	0.048479
ARIMA (1, 2, 1)	1513.197	0.06908	-8682.6	0.000013	0.000017	0.053048
ARIMA (1, 2, 2)	1508.335	0.065266	-7614.4	0.0000119	0.0000159	0.048909
ARIMA (1, 2, 3)	1509.23	**0.064573**	-9834.6	0.0000117	0.0000157	0.048497
ARIMA (2, 2, 1)	1513.37	0.068016	-11702	0.0000128	0.0000166	0.052567
ARIMA (3, 2, 2)	1511.389	0.064602	-8999.6	0.0000119	0.0000158	0.048889

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 paper will consider only the AIC in selecting the optimal model. Therefore, the ARIMA (0, 2, 3) model is chosen.

Residual & Stability Tests

ADF Tests of the Residuals of the ARIMA (0, 2, 3) Model

Table 11: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
W_t	-7.507096	0.0000	-3.555023	@1%	Stationary
			-2.915522	@5%	Stationary
			-2.595565	@10%	Stationary

Table 12: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
W_t	-7.437924	0.0000	-4.133838	@1%	Stationary
			-3.493692	@5%	Stationary
			-3.175693	@10%	Stationary

Table 13: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
W_t	-7.562630	0.0000	-2.607686	@1%	Stationary
			-1.946878	@5%	Stationary
			-1.612999	@10%	Stationary

Tables 11, 12 and 13 show that the residuals of the ARIMA (0, 2, 3) model are stationary.

Stability Test of the ARIMA (0, 2, 3) Model

Figure 5

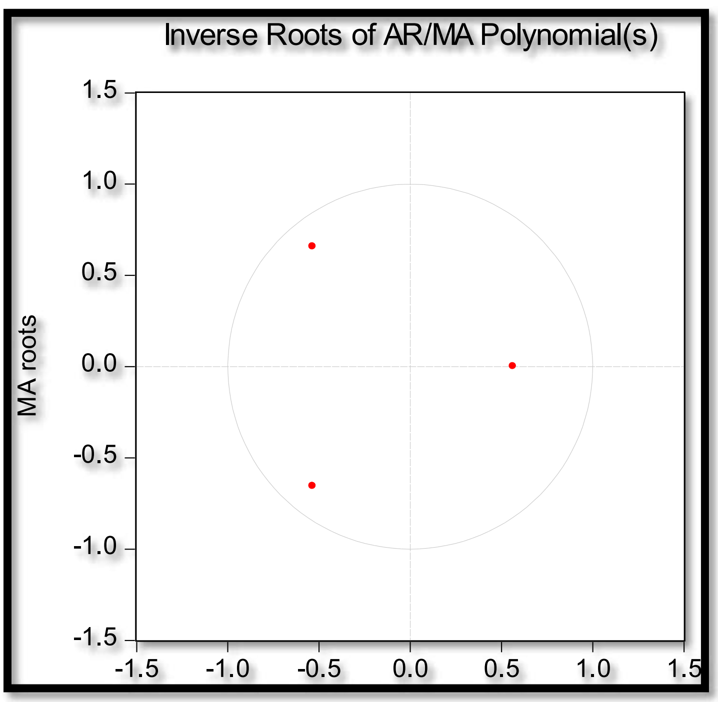


Figure 5 above indicates that the ARIMA (0, 2, 3) model, is quite stable as expected, as the corresponding inverse roots of the characteristic polynomial lie in the unit circle.

RESULTS & DISCUSSION

Descriptive Statistics

Table 14

Description	Statistic
Mean	251270000
Median	245660000
Minimum	180670000
Maximum	325720000
Standard deviation	43555000
Skewness	0.13392
Excess kurtosis	-1.2451

The mean, as shown in the table above; is positive, i.e. 251 270 000. The wide gap between the minimum, i.e., 180 670 000 and the maximum, i.e., 325 720 000 is consistent with the reality that the POP series is on an upwards trajectory. The skewness is 0.13392 and it is positive, revealing that the POP series is positively skewed and non-symmetric. Excess kurtosis is -1.2451 showing that the POP series is not normally distributed.

Results Presentation[1]

Table 15

ARIMA (0, 2, 3) Model:

$$\Delta^2 POP_{t-1} = 0.487\mu_{t-1} + 0.1\mu_{t-2} - 0.349\mu_{t-3} \ldots \quad [3]$$

P: (0.0001) (0.4828) (0.0067)

S. E: (0.128159) (0.142164) (0.128796)

Variable	Coefficient	Standard Error	z	p-value
MA (1)	0.487078	0.128159	3.801	0.0001***
MA (2)	0.099770	0.142164	0.7018	0.4828
MA (3)	-0.349127	0.128796	-2.711	0.0067***

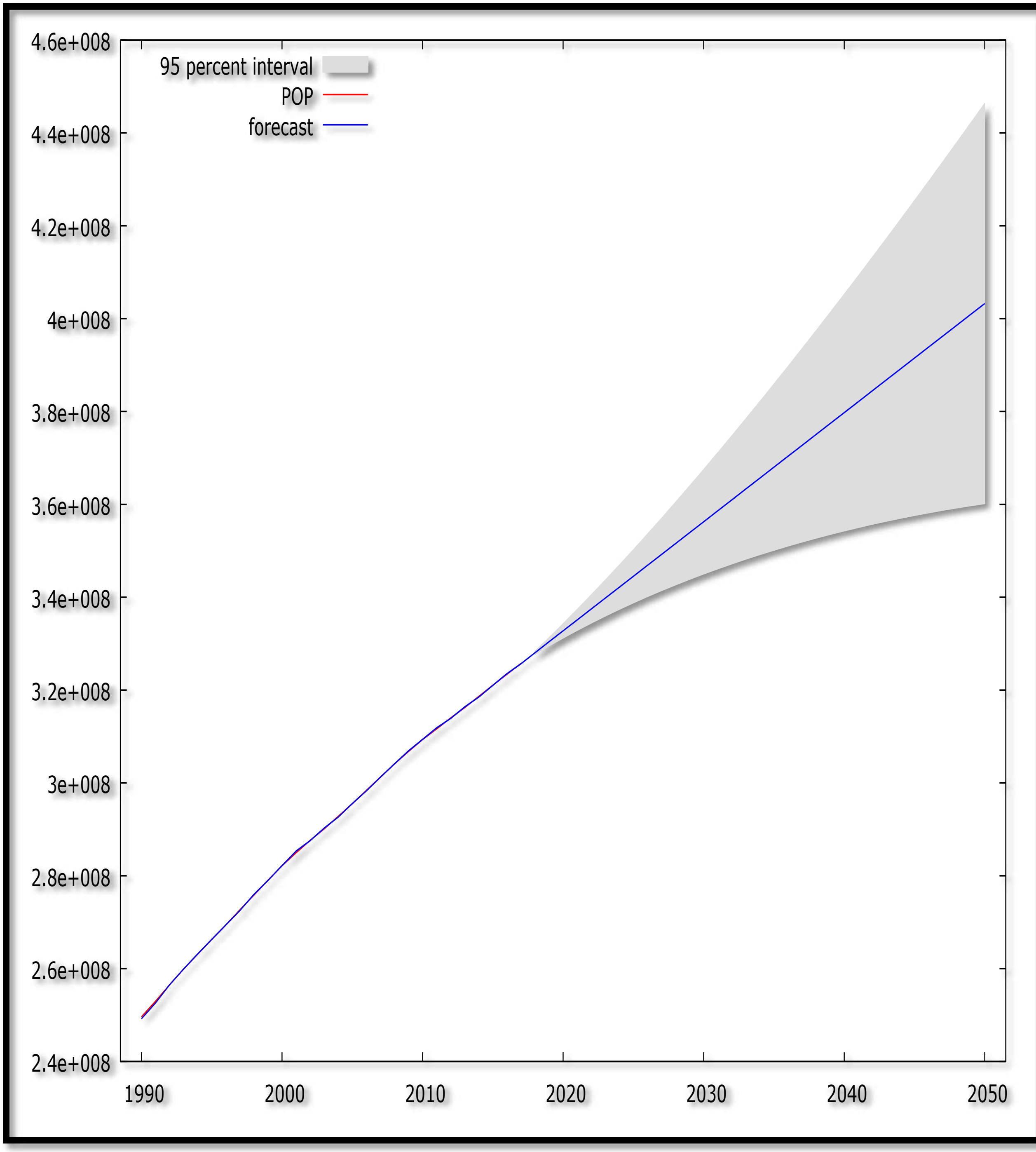
Interpretation of Results

The coefficients of the MA (1) and MA (3) terms are statistically significant at 1% level of significance, the MA (1) coefficient is positive (i.e. 0.487078) while the MA (3) coefficient is negative (i.e. -0.349127). It is quite clear that the MA (1) and MA (3) coefficients are more relevant in explaining population dynamics in the US. The MA (2) coefficient is positive (i.e. 0.099770) but statistically insignificant and thus less important in explaining US population dynamics over the period under study.

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

Forecast Graph

Figure 6



Predicted Total Population

Figure 7


Predicted Total Population
2050 403257000
2049 400907000
2048 398557000
2047 396207000
2046 393857000
2045 391507000
2044 389157000
2043 386807000
2042 384457000
2041 382107000
2040 379757000
2039 377407000
2038 375057000
2037 372707000
2036 370357000
2035 368006000
2034 365656000
2033 363306000
2032 360956000
2031 358606000
2030 356256000
2029 353906000
2028 351556000
2027 349206000
2026 346856000
2025 344506000
2024 342156000
2023 339806000
2022 337456000
2021 335106000
2020 332756000
2019 330406000
2018 328038000

Figures 6 (with a forecast range of 32 years, i.e.; 2018 – 2050) and 7, clearly indicate that USA population is indeed set to continue rising sharply, at least for the next 3 decades, ceteris paribus. With a 95% confidence interval of 360 199 000 to 446 315 000 and a projected total population of 403 257 000 by 2050, the ARIMA (0, 2, 3) model is consistent with the population projections by the UN (2015) which forecasted that US total population will be approximately 388 865 000 by 2050. Our model is also consistent with the population projections by the US Census Bureau (2018) which forecasted that the US will grow by 78 million people in the next 4 decades, from about 326 million to 404 million between 2017 and 2060. The optimal model, our ARIMA (0, 2, 3) model is also line with US population projections done by Colby & Ortman (2014) who forecasted that between 2014 and 2060, the US population will increase from 319 million to 417 million, reaching 400 million in 2051. A growing population, in the US, is arguably an opportunity for growth given US's educated labor force and the technological advancements prevalent in the US. This study argues that the US is and is likely to continue fulfilling the Ahlburg (1998) and Becker *et al* (1999) population prophecy.

Policy Implications

i. For the US to continue wondering the in the realms of the Ahlburg (1998) and Becker *et al* (1999) population prophecies, there is need to maintain a highly educated and trained workforce. Technological advancements and innovation should continue in order to continuously improve production processes and national output.
ii. The US policy makers ought to encourage a culture of entrepreneurship and creativity in order to circumvent the likely challenge of unemployment due to a large population.
iii. Since a large population basically translates into an increased demand for goods and services, the US business community should expand their business operations in order to cater for the expected increase in demand for commodities.

CONCLUSION

The ARIMA (0, 2, 3) model is a suitable and most parsimonious model to forecast the population of the USA for the next 3 decades. The model predicts that by 2050, USA's population would be nearly, 403 million. The results of this endeavor are important for the US government, especially in terms of planning for the future.

REFERENCES

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

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

[4] Colby, S. L & Ortman, J. M (2014). Projections of the size and composition of the US population: 2014 to 2060, Current Population Report, *US Census Bureau*, pp: 25 – 1143.

[5] 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.

[6] Haque, M., Ahmed, F., Anam, S., & Kabir, R (2012). Future population projection of Bangladesh by growth rate modeling using logistic population model, *Annals of Pure and Applied Mathematics*, 1 (2): 192 – 202.

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

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

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

[10] 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.

[11] Pflaumer, P (1992). Forecasting US population totals with the Box-Jenkins approach, *International Journal of Forecasting*, 8: 329 – 338.

[12] Pflaumer, P (2012). Forecasting the US population with the Gompertz Growth curve, *Social Statistics Section – JSM*, pp: 4967 – 4981.

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

[14] 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.

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

[16] 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.

[17] United States Census Bureau (2018). Demographic turning points for the United States: population projections for 2020 to 2060, *USCB*, March Issue.

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