Addressing the population question in Mexico: A Box-Jenkins ARIMA approach

NYONI, THABANI

University of Zimbabwe

19 February 2019

Addressing The Population Question In Mexico: A 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 Mexico from 1960 to 2017, we 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 Mexico annual total population is I (2). Based on the AIC, the study presents the ARIMA (3, 2, 1) model as the optimal model. The diagnostic tests further reveal that the presented model is stable and that its residuals are stationary. The results of the study show that total population in Mexico will continue to rise in the next three decades and in 2050 Mexico's total population will be approximately 180 million people. Three policy prescriptions have been proposed for consideration by the government of Mexico.

Key Words: Forecasting, Mexico, 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).

Mexico experienced several decades of high population growth toward the end of the 20th century. This growth, coupled with increased female labor force participation, coincided with substantial emigration to the United States between 1970 and 2000. Overall population growth, however, is now slowing; by about 2030, it is expected that the size of the working-age population will begin to decrease. The slowing population growth, coupled with economic developments and changes in US immigration policy (including stricter border control), has resulted in a slight slowdown in Mexican immigration to the United States relative to the 1995 to 2000 period (Zuniga & Molina, 2008). In Mexico, just like in any other part of the world, population modeling and forecasting is really important for policy dialogue. This study seeks to model and forecast population of Mexico using the Box-Jenkins ARIMA approach.

LITERATURE REVIEW

Related Previous Studies

Using Box-Jenkins ARIMA models, Zakria & Muhammad (2009) modeled and forecasted population and relied on a data set ranging from 1951 - 2007; and finalized 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 based on an Autoregressive Time Trend (ATT) model making use of a data set ranging over 1965 – 2003 and finalized that there will be a downward population growth for Bangladesh for the extended period up to 2043. Ayele & Zewdie (2017) analyzed human population size and its trend in Ethiopia using Box-Jenkins ARIMA models and made use of annual data from 1961 - 2009 and proved that the most suitable model for modeling and forecasting population in Ethiopia was the ARIMA (2, 1, 2) model. In the case of Mexico, the study 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 by a backward shift operator as:

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

Where the autoregressive (AR) and moving average (MA) characteristic operators are:

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

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

and

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

Where $\emptyset$ 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 and μ_t is the disturbance term.

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).

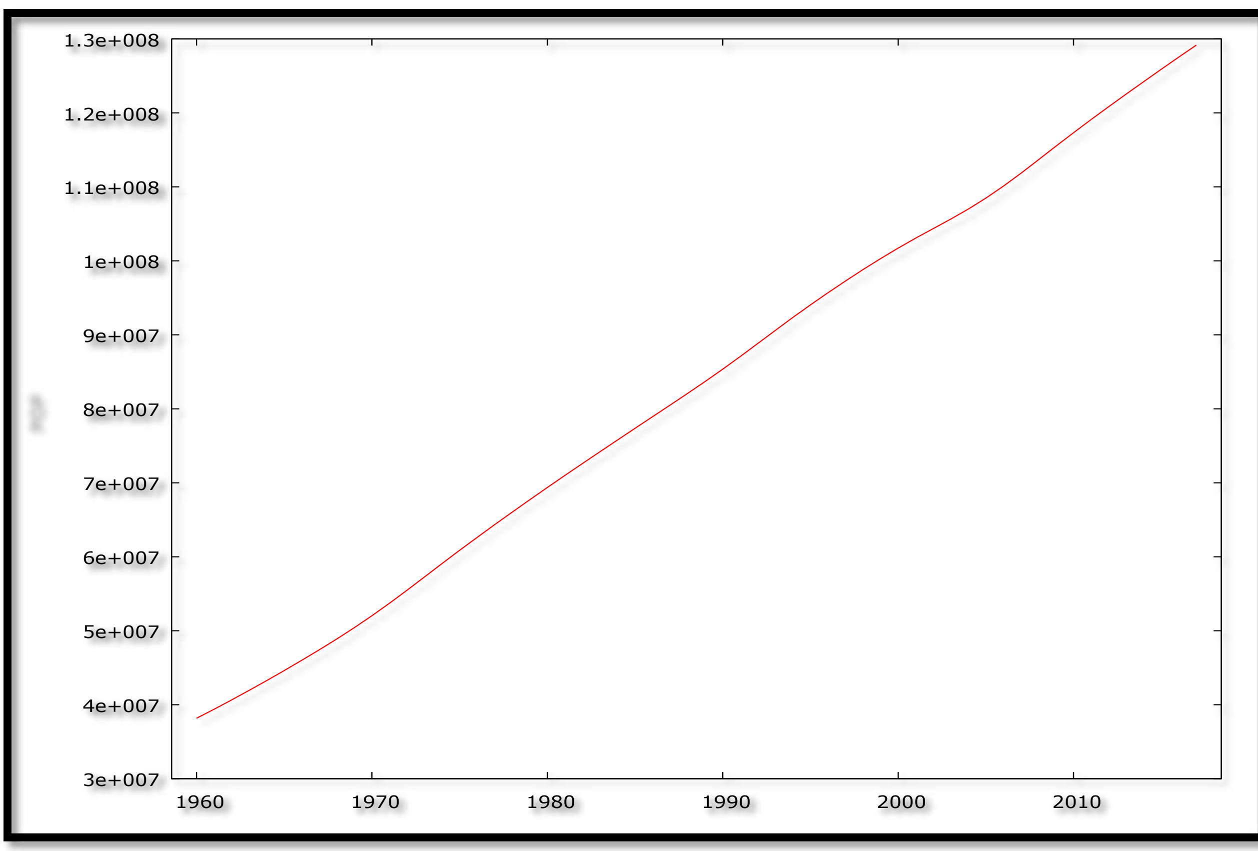
Data Collection

This piece of work is based on 58 observations of annual total population (POP, referred to as PMEX in the mathematical formulation above) in Mexico, i.e from 1960 – 2017. Our data was taken from the World Bank online database. The Word Bank online database is a well known reliable and credible source of various macroeconomic data.

Diagnostic Tests & Model Evaluation

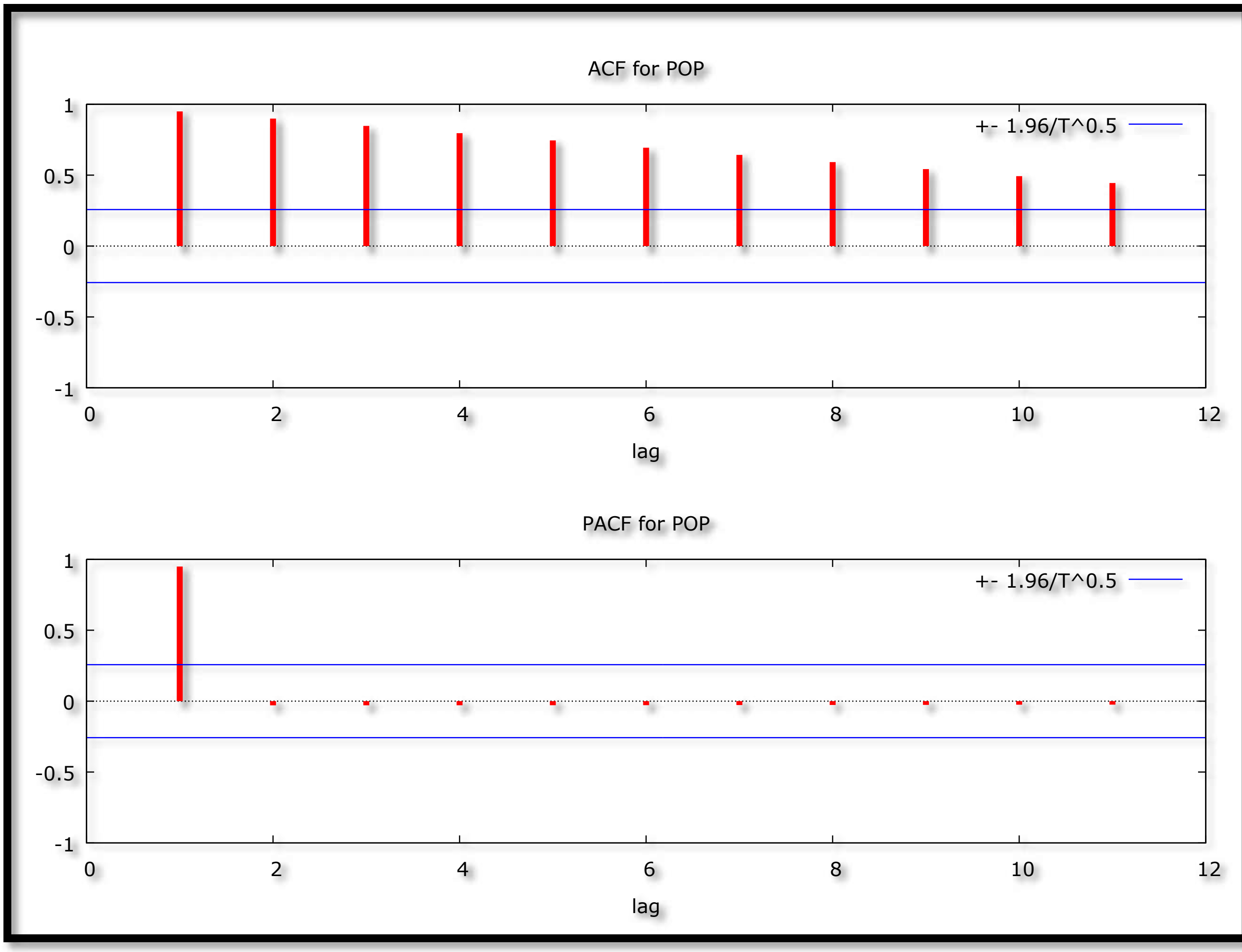
Stationarity Tests: Graphical Analysis

Figure 1



The Correlogram in Levels

Figure 2



The ADF Test

Table 1: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	0.173341	0.9683	-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	-2.820826	0.1966	-4.140858	@1%	Stationary
			-3.496960	@5%	Stationary
			-3.177579	@10%	Stationary

Table 3: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-0.276853	0.5812	-2.613030	@1%	Not stationary
			-1.947665	@5%	Not stationary
			-1.612573	@10%	Not stationary

The Correlogram (at 1st Differences)

Figure 3

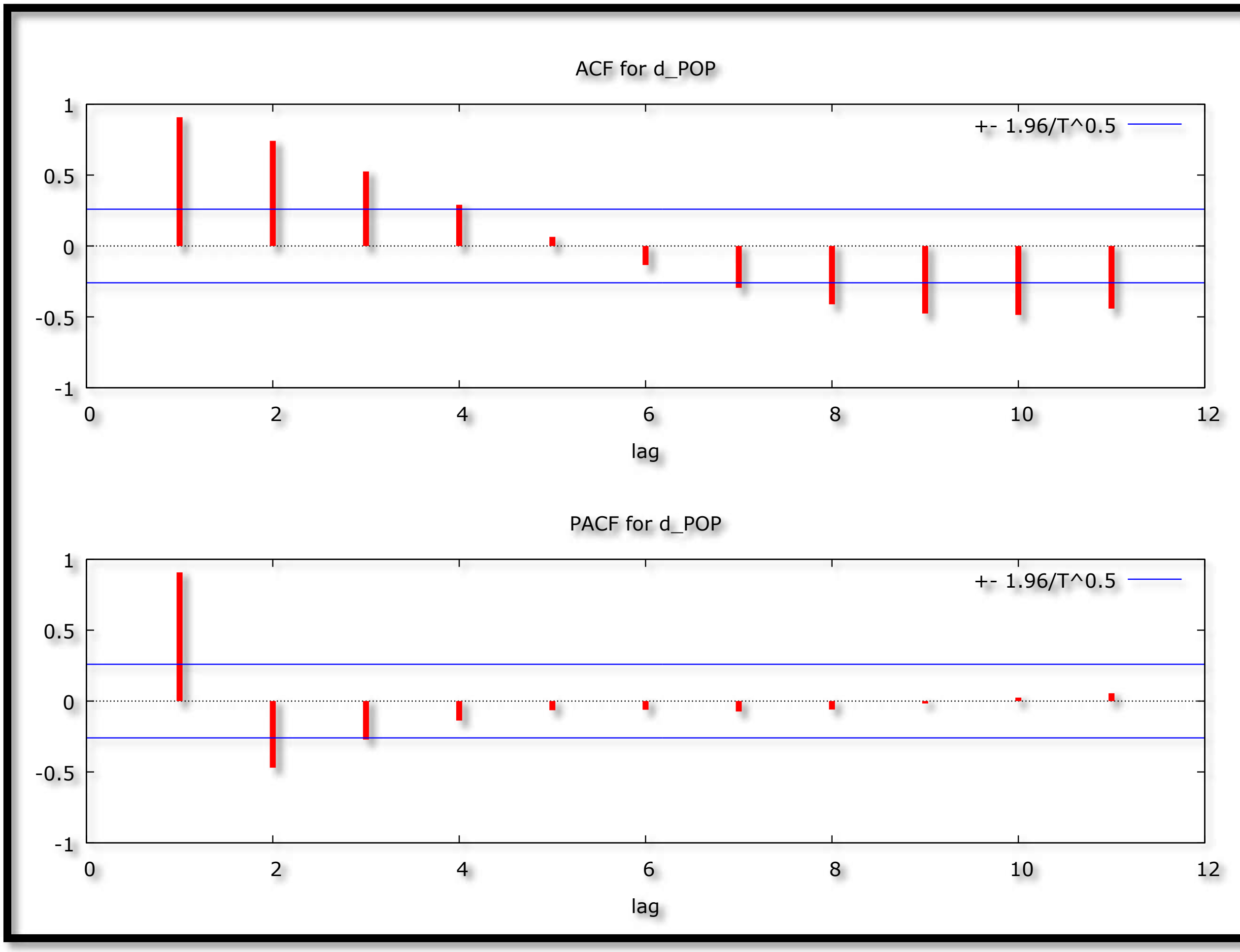


Table 4: 1st Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-3.812579	0.0050	-3.560019	@1%	Stationary
			-2.917650	@5%	Stationary
			-2.596689	@10%	Stationary

Table 5: 1st Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-2.007702	0.5817	-4.170583	@1%	Not stationary
			-3.510740	@5%	Not stationary
			-3.185512	@10%	Not stationary

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

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	0.122453	0.7165	-2.616203	@1%	Not stationary

		-1.948140	@5%	Not stationary
		-1.612320	@10%	Not stationary

Figures 1 – 3 and tables 1 – 6 demonstrate that the Mexico POP series is neither I (0) nor I (1).

The Correlogram in (2nd Differences)

Figure 4

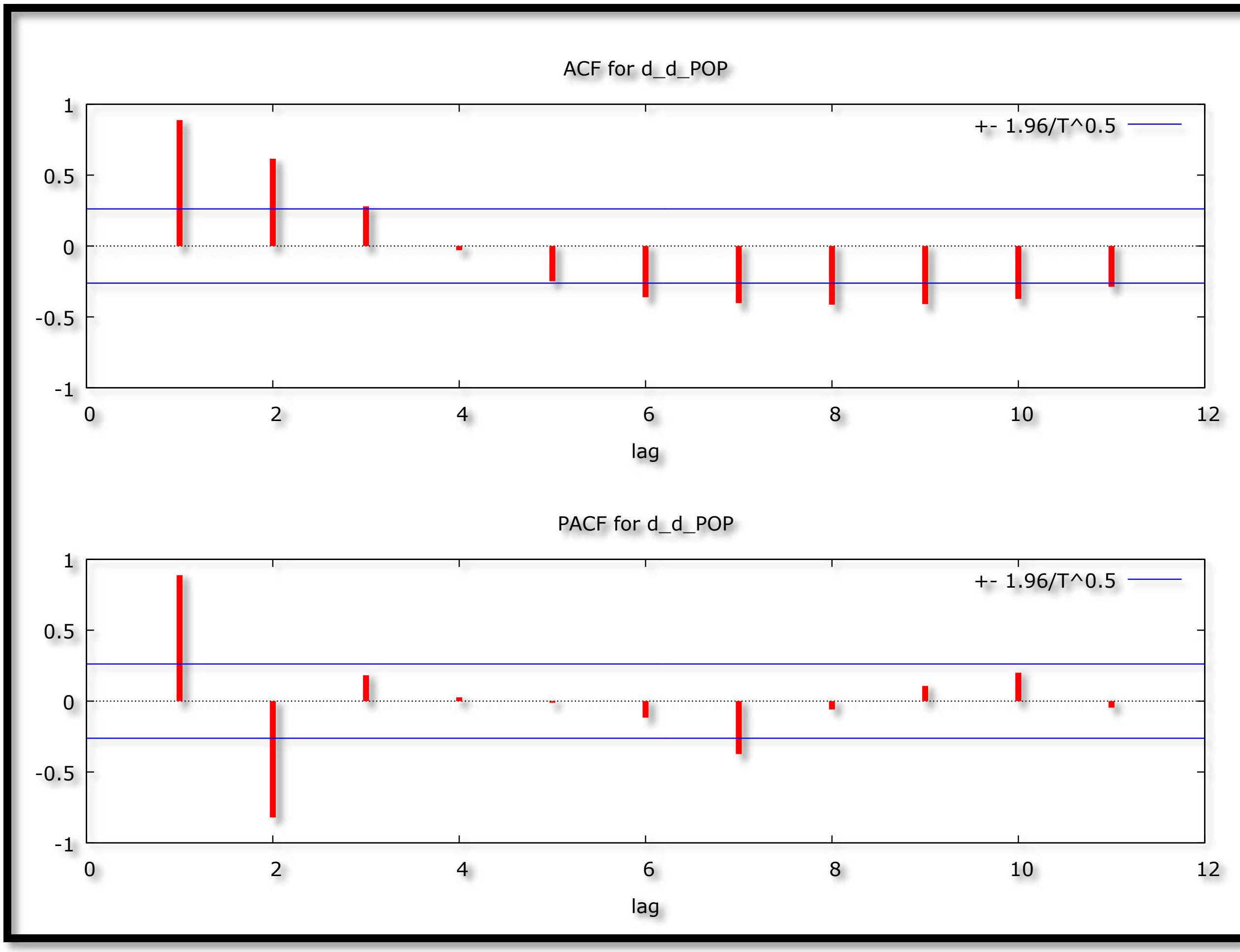


Table 7: 2nd Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-2.659419	0.0891	-3.584743	@1%	Not stationary
			-2.928142	@5%	Not stationary
			-2.602225	@10%	Stationary

Table 8: 2nd Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-2.440167	0.3550	-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	-2.675767	0.0086	-2.617364	@1%	Stationary
			-1.948313	@5%	Stationary
			-1.612229	@10%	Stationary

Figure 4 indicates the autocorrelation coefficients are only high for the first and second lags, the rest are low – relatively closer to zero, which means that the POP series is stationary in second differences. Table 8 indicates that the POP series is non-stationary and yet tables 7 and 9 concur with figure 4 that the POP series is now stationary, i.e it is an I (2) variable.

Evaluation of ARIMA models (without a constant)

Table 10

Model	AIC	U	ME	MAE	RMSE	MAPE
ARIMA (1, 2, 1)	1238.583	0.0072622	518.97	11103	14839	0.013904
ARIMA (1, 2, 0)	1286.855	0.011445	384.49	15610	22777	0.019171
ARIMA (2, 2, 0)	1211.307	0.0062145	1786.7	9028	11927	0.011924
ARIMA (3, 2, 0)	1196.213	0.0050181	1017.3	7813.5	10458	0.01024
ARIMA (4, 2, 0)	1195.885	0.0049753	1250.2	7839.2	10277	0.010286
ARIMA (3, 2, 1)	**1195.863**	**0.0049431**	1146.9	7836.8	10275	0.01026

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 on the AIC and the Theil's U in order to choose the optimal model in predicting total population in Mexico. Therefore, the ARIMA (3, 2, 1) model is chosen.

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
ε_t	-5.277735	0.0001	-3.592462	@1%	Stationary
			-2.931404	@5%	Stationary
			-2.603944	@10%	Stationary

Table 12: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
ε_t	-5.201603	0.0006	-4.186481	@1%	Stationary
			-3.518090	@5%	Stationary
			-3.189732	@10%	Stationary

Table 13: without intercept and trend & intercept

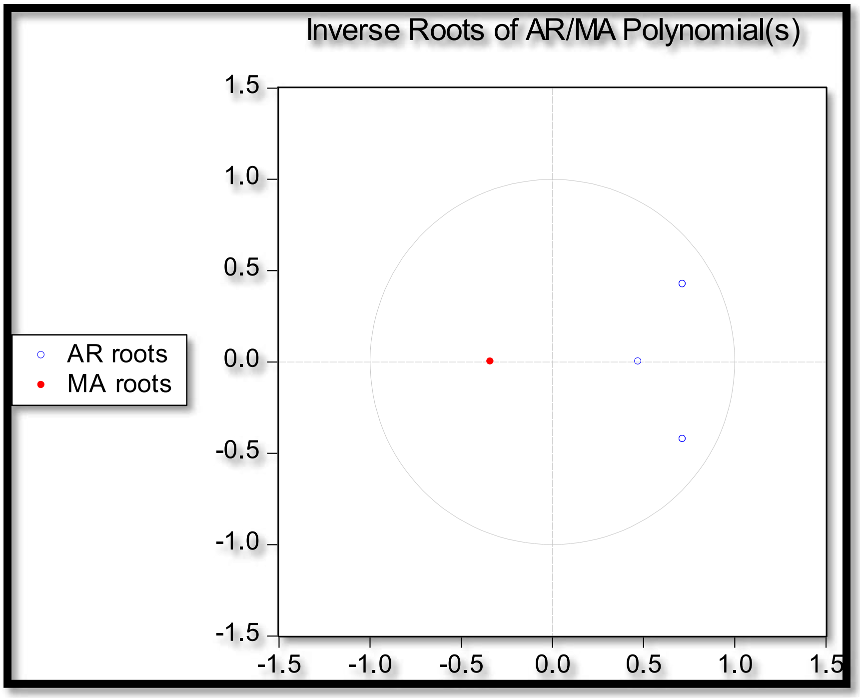
Variable	ADF Statistic	Probability	Critical Values		Conclusion
ε_t	-5.318470	0.0000	-2.619851	@1%	Stationary

		-1.948686	@5%	Stationary
		-1.612036	@10%	Stationary

Tables 11 – 13 indicate that the residuals of the ARIMA (3, 2, 1) model are stationary.

Stability Test of the ARIMA (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 quite stable.

RESULTS

Descriptive Statistics

Table 14

Description	Statistic
Mean	82824000
Median	82891000
Minimum	38174000
Maximum	129160000
Standard deviation	27267000
Skewness	-0.0024453
Excess kurtosis	-1.2119

As shown above, the mean is positive, i.e. 82824000. The wide gap between the minimum (i.e 38174000) and the maximum (i.e. 129160000) is consistent with the reality that the Mexico POP series is trending upwards. The skewness is -0.0024453 and the most striking characteristic is that it is negative, indicating that the POP series is negatively skewed and non-symmetric. Excess kurtosis is -1.2119; revealing that the POP series is not normally distributed.

Results Presentation[1]

Table 15

ARIMA (3, 2, 1) Model:

$$\Delta^2 POP_{t-1} = 1.8808\Delta^2 POP_{t-1} - 1.3305\Delta^2 POP_{t-2} + 0.3126\Delta^2 POP_{t-3} + 0.336\mu_{t-1} \ldots \ldots [5]$$

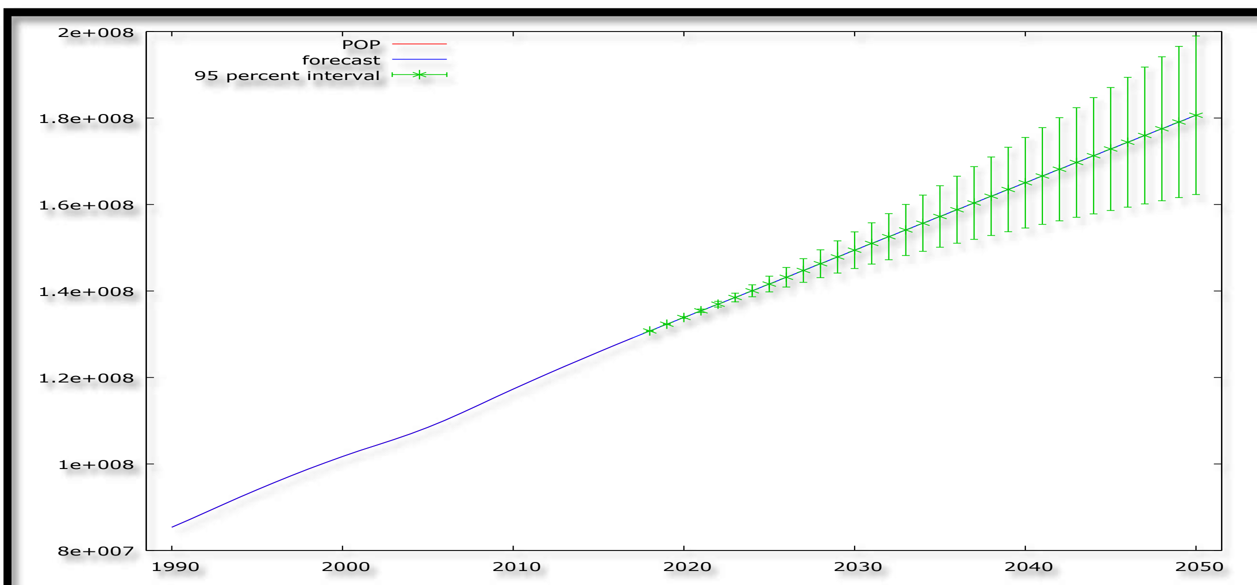
P: (0.0000) (0.0005) (0.1340) (0.1391)

S. E: (0.2246) (0.3833) (0.2086) (0.2271)

Variable	Coefficient	Standard Error	z	p-value
AR (1)	1.88084	0.224628	8.373	0.0000***
AR (2)	-1.33046	0.383275	-3.471	0.0005***
AR (3)	0.312604	0.208603	1.499	0.1340
MA (1)	0.335971	0.227118	1.479	0.1391

Forecast Graph

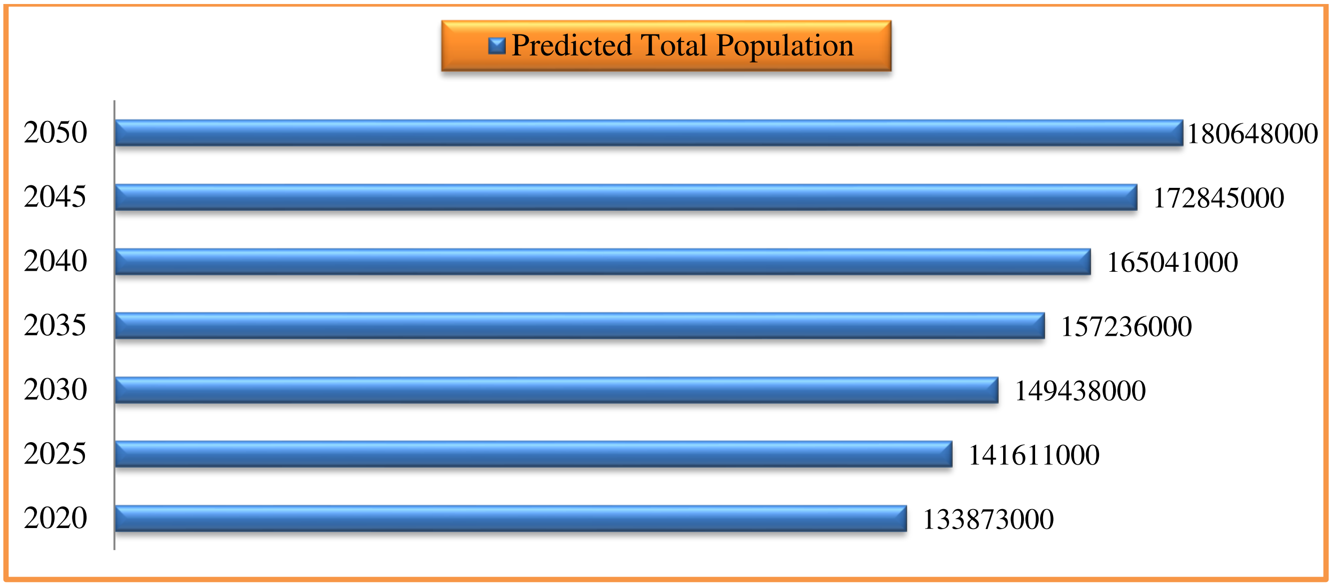
Figure 6



Predicted Total Population

Figure 7

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



Figures 6 (with a forecast range from 2018 – 2050) and 7, clearly indicate that Mexico population is indeed set to continue rising, at least for the next 3 decades. With a 95% confidence interval of 162312000 to 198984000 and a projected total population of 180648000 by 2050, the chosen ARIMA (3, 2, 1) model is consistent with the population projections by the UN (2015) which forecasted that Mexico's population will be approximately 163754000 by 2050 and is also in line with the recent population projections by the UN (2017) which forecasted that Mexico's population will be approximately 164279000 by 2050.

Policy Implications

1. The government of Mexico ought to continue investing more in infrastructural development, e.g housing, education institutions, road and telecommunication networks etc. in order to cater for the projected increase in total population.
2. The predicted gradual increase in total population in Mexico justifies the need for more and bigger companies to provide not only for the expected increase in demand for goods and services but also for the provision of employment opportunities in Mexico, especially given the high rates of unemployment persistently obtaining in Mexico.
3. The government of Mexico should continue encouraging the smaller family size norm in order to properly address the problems of spiraling population.

CONCLUSION

In the case of Mexico, the study shows that the ARIMA (3, 2, 1) model is not only stable but also the most suitable model to forecast total population for the next 3 decades. The model predicts that by 2050, Mexico's total population would be approximately, 180 million people. This is a warning signal to policy makers in Mexico, especially in light of persistant high unemployment levels and re-current economic hardships which continue to characterize Mexico. These results are indeed necessary for the government of Mexico, especially when it comes to medium-term and 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 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.

[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] 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.

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

[13] 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.

[14] United Nations (2017). World Population Prospects: The 2017 Revision, Key Findings and Advance Tables, Department of Economics and Social Affairs, Population Division, Working Paper No. ESAP/P/WP/248.

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

[16] Zuniga, E & Molina, M (2008). Demographic trends in Mexico: The implications for skilled migration, *Transatlantic Council on Migration*, New York.