Forecasting the population of Brazil using the Box-Jenkins ARIMA approach

NYONI, THABANI

University of Zimbabwe

25 February 2019

FORECASTING THE POPULATION OF BRAZIL USING 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 Brazil 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 Brazil annual total population is non-stationary in all levels; for simplicity purposes, the study has assumed that the POP series is I (2). Based on the AIC, the study presents the ARIMA (6, 2, 0) model as the optimal model. The diagnostic tests further indicate that the presented model is stable and that its residuals are stationary. The results of the study reveal that total population in Brazil will continue to rise in the next three decades and in 2050 Brazil's total population will be approximately 256 million people. Four policy prescriptions have been suggested for consideration by the government of Brazil.

Key Words: Brazil, 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).

Brazil, the world's largest country in land area, also ranks fifth in total population size, with 156.7 million persons in 1993. Among the 45 countries of the Latin American and Caribbean region, Brazil ranks number one in population and land area, with its population size greatly exceeding that of its neighbors and accounting for one-third of the region's population. Brazil's population more than tripled in size since 1950, but fertility decline reduced the average annual growth rate from a peak of over 2.9% in the 1950s to 1.4% in the early 1990s. The absolute population increase peaked in the late 1980s and is now declining. Currently, Brazil is adding about 2.3 million people to its population each year (Center for International Research, 1993). In Brazil, 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 Brazil using the Box-Jenkins ARIMA approach.

REVIEW OF PREVIOUS STUDIES

Using ARIMA models, Zakria & Muhammad (2009) forecasted population 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 making use of a data set ranging over 1965 – 2003 and illustrated that there will be a downward population growth for Bangladesh for the extended period up to 2043. Ayele & Zewdie (2017) investigated human population size and its pattern in Ethiopia using ARIMA models and made use of annual data from 1961 - 2009 and demonstrated that the most suitable model for modeling and forecasting population in Ethiopia was the ARIMA (2, 1, 2) model. In the case of Brazil, 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 follows:

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

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

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

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

and

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

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).

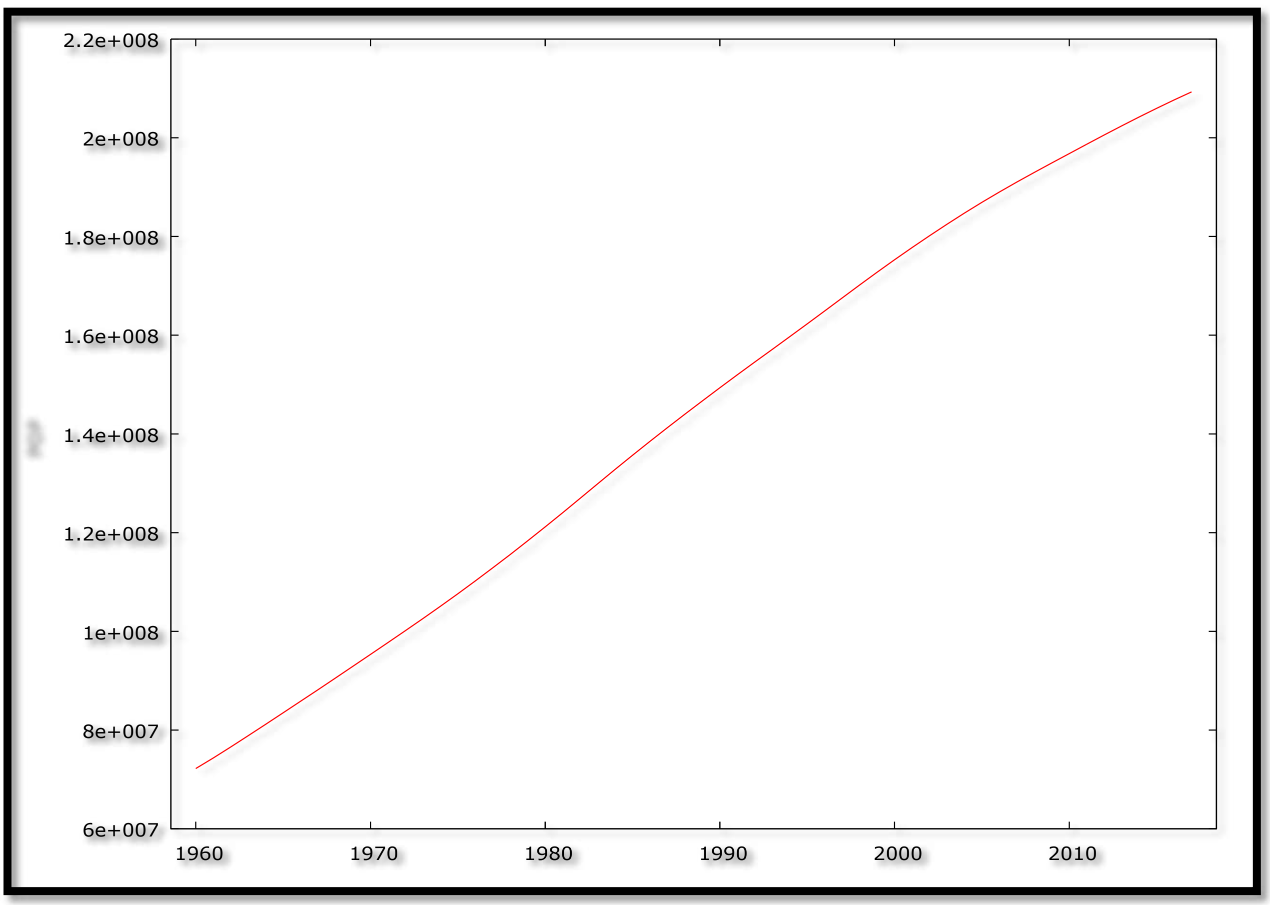
Data Collection

This paper is based on 58 observations of annual total population in Brazil (POP, referred to as P in the mathematical formulations above). Our data was taken from the World Bank online database, whose recognition, integrity and credibility is well above board.

Diagnostic Tests & Model Evaluation

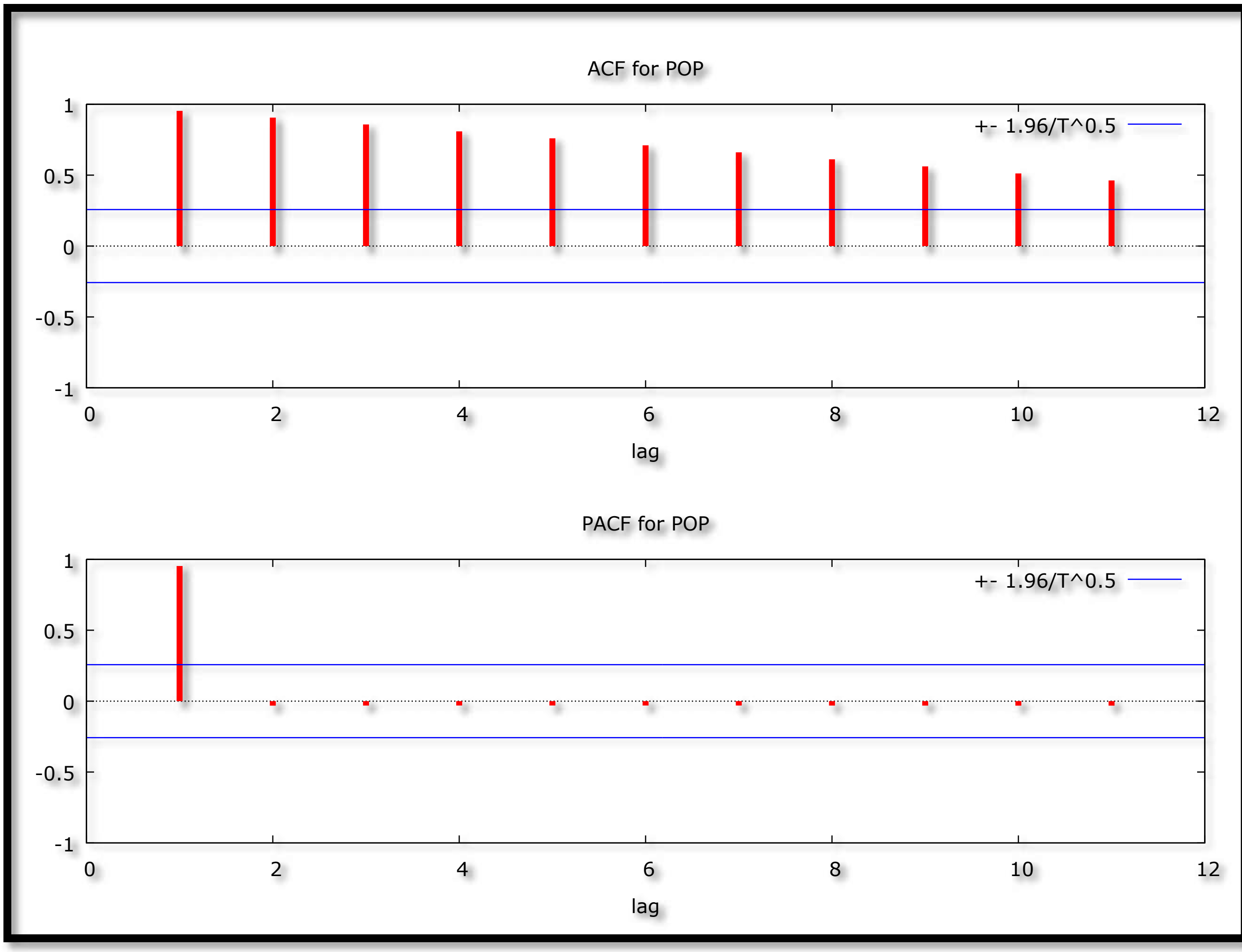
Stationarity Tests: Graphical Analysis

Figure 1



The Correlogram in Levels

Figure 2



The ADF Test

Table 1: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-5.709463	0.0000	-3.574446	@1%	Stationary
			-2.923780	@5%	Stationary
			-2.599925	@10%	Stationary

Table 2: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-2.883677	0.1767	-4.161144	@1%	Not stationary
			-3.506374	@5%	Not stationary
			-3.183002	@10%	Not stationary

Table 3: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-3.264192	0.0016	-2.613030	@1%	Stationary
			-1.947665	@5%	Stationary
			-1.612573	@10%	Stationary

Tables 1 and 3 indicate that the POP series is I (0) and yet table 2 indicates that the POP series is non-stationary in levels. We therefore, proceed to test for stationarity in first differences to further analyze the stationarity of the POP series. In most cases, it is rare for a sharply upward trending series to be stationary in levels and hence the justifications to further analyze the POP series.

The Correlogram (at 1st Differences)

Figure 3

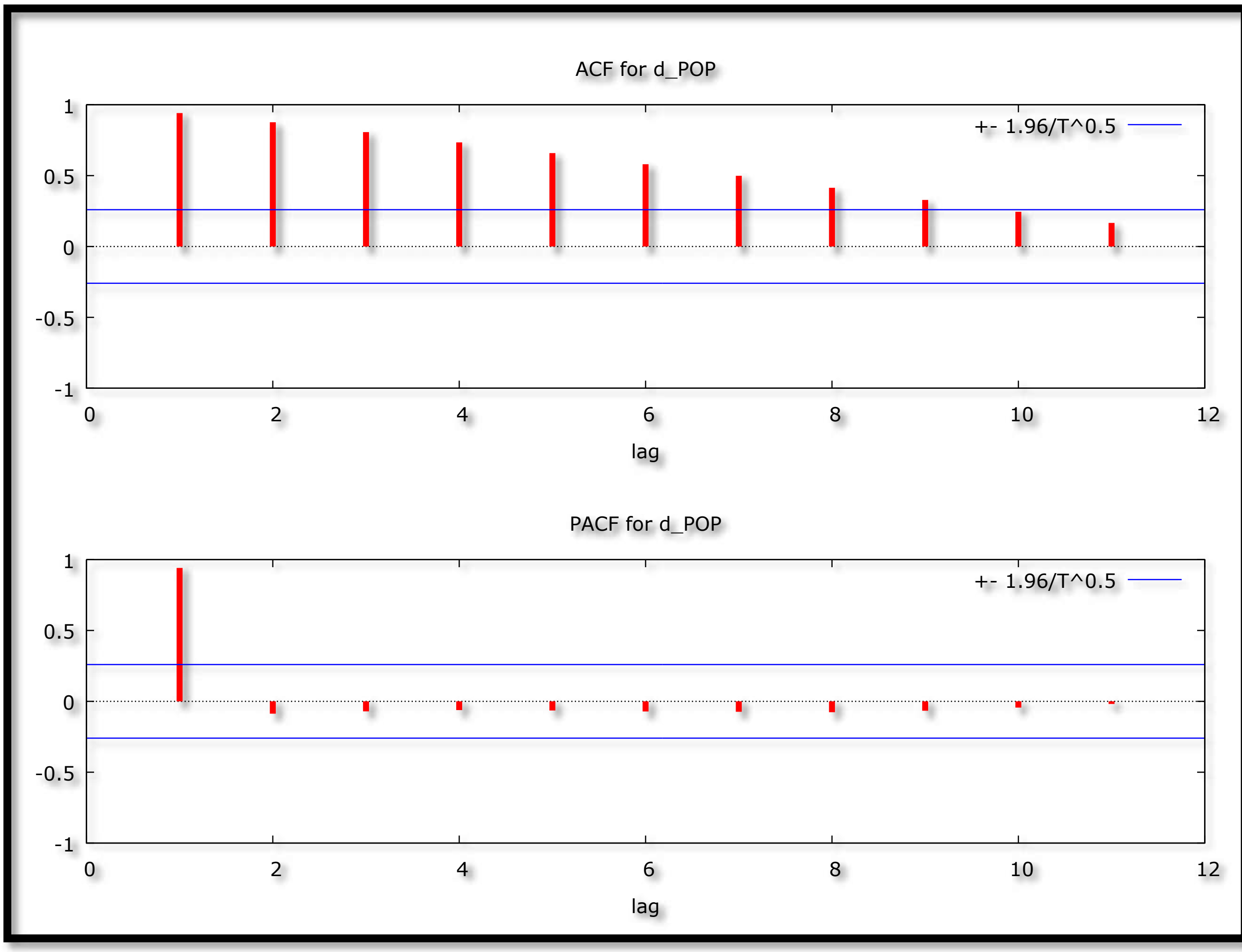


Table 4: 1st Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.907748	0.3260	-3.581152	@1%	Not stationary
			-2.926622	@5%	Not stationary
			-2.601424	@10%	Not stationary

Table 5: 1st Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-2.372631	0.3884	-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	-3.970497	0.0002	-2.616203	@1%	Stationary
			-1.948140	@5%	Stationary
			-1.612320	@10%	Stationary

While tables 4 and 5 show that the POP series non-stationary in first differences, table 6 indicates that the POP series is I (1). The researcher will go ahead and test for stationarity in second differences in second differences.

The Correlogram in (2nd Differences)

Figure 4

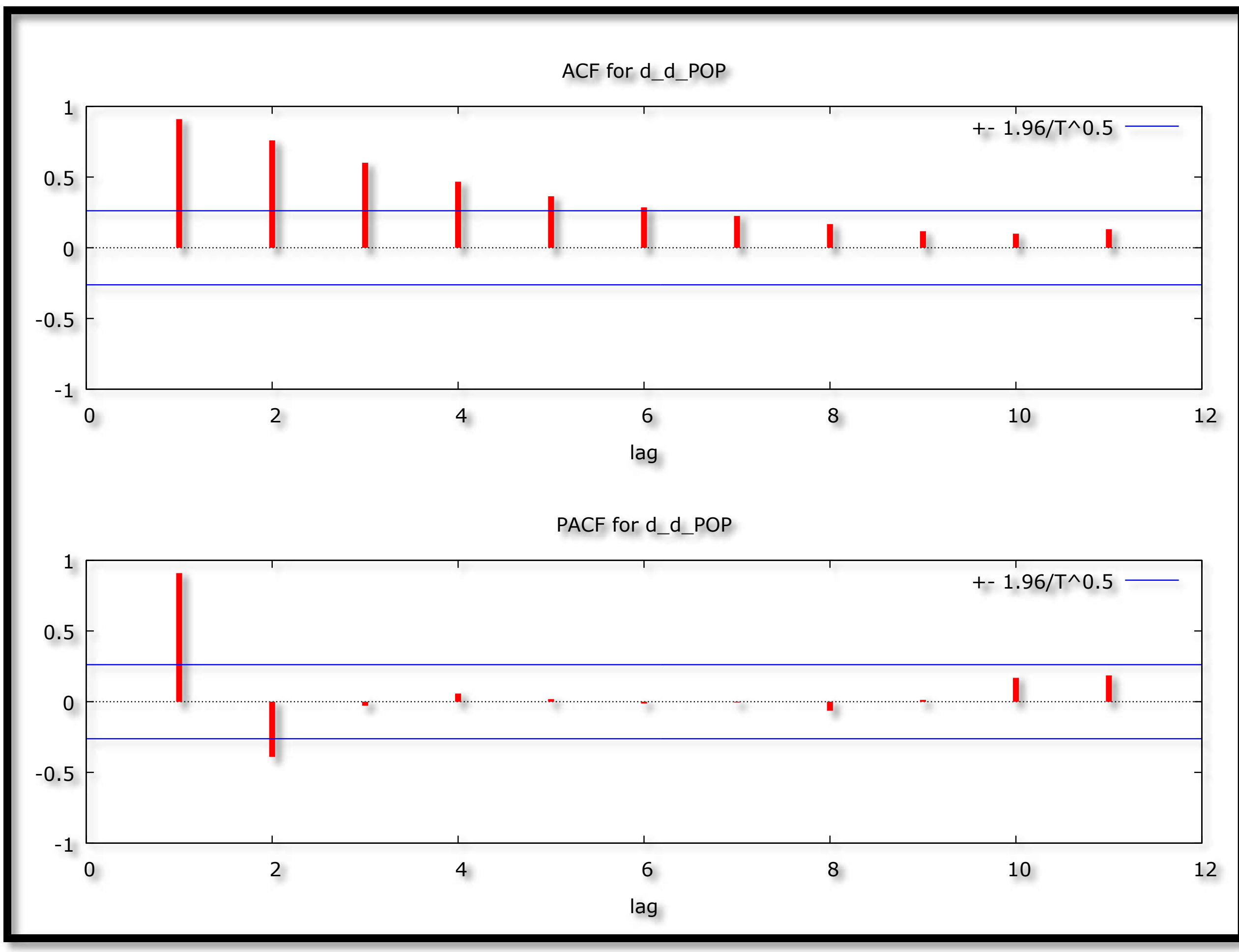


Table 7: 2nd Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-0.395784	0.9012	-3.581152	@1%	Not stationary

			-2.926622	@5%	Not stationary
			-2.601424	@10%	Not stationary

Table 8: 2nd Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.376028	0.8549	-4.170583	@1%	Not stationary
			-3.510740	@5%	Not stationary
			-3.185512	@10%	Not stationary

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

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-0.162081	0.6221	-2.617364	@1%	Not stationary
			-1.948313	@5%	Not stationary
			-1.612229	@10%	Not stationary

Tables 7 – 9 confirm that the POP series is non-stationary. This is quite acceptable for sharply upward trending series.

Evaluation of ARIMA models (without a constant)

Table 10

Model	AIC	U	ME	MAE	RMSE	MAPE
ARIMA (1, 2, 0)	1232.558	0.0053331	-1255.8	11531	16961	0.0091098
ARIMA (2, 2, 0)	1177.056	0.0032717	326.03	8017.1	13089	0.0066943
ARIMA (3, 2, 0)	1154.597	0.0027459	281.49	6529.8	12210	0.0056956
ARIMA (4, 2, 0)	1151.562	0.0026275	396.83	6416.9	12068	0.0011724
ARIMA (5, 2, 0)	1150.334	0.0025729	370.18	6345.3	11983	0.0055375
ARIMA (6, 2, 0)	**1147.931**	**0.002488**	464.43	6129.3	11869	0.0053618

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 Brazil. Therefore, the ARIMA (6, 2, 0) model is chosen.

Residual & Stability Tests

ADF Tests of the Residuals of the ARIMA (6, 2, 0) Model

Table 11: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
ε_t	-3.026617	0.0411	-3.610453	@1%	Not stationary
			-2.938987	@5%	Stationary
			-2.607932	@10%	Stationary

Table 12: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
ε_t	-9.558579	0.0000	-4.205004	@1%	Stationary

		-3.526609	@5%	Stationary
		-3.194611	@10%	Stationary

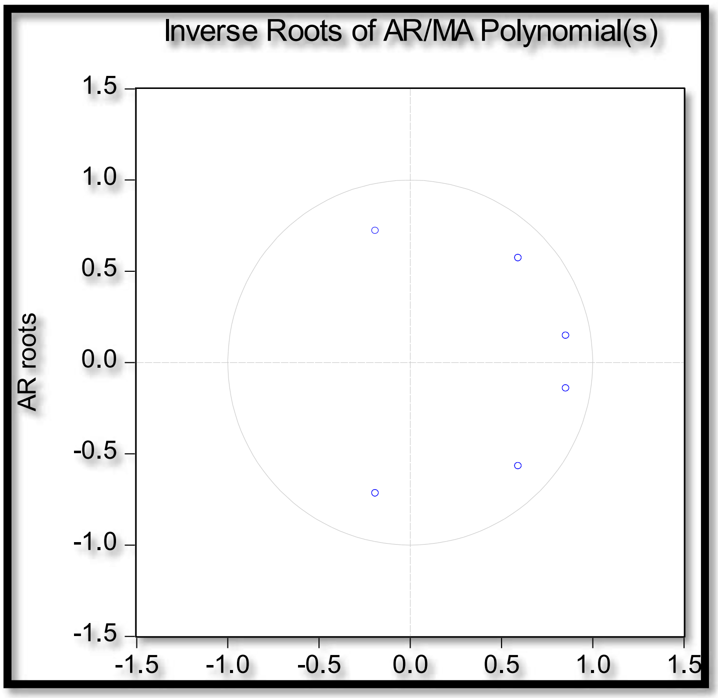
Table 13: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
ε_t	-1.857507	0.0609	-2.625606	@1%	Not stationary
			-1.949609	@5%	Not stationary
			-1.611593	@10%	Stationary

Tables 11 – 13 indicate that the residuals of the ARIMA (6, 2, 0) model are stationary.

Stability Test of the ARIMA (6, 2, 0) Model

Figure 5



Since the corresponding inverse roots of the characteristic polynomial lie in the unit circle, it illustrates that the chosen ARIMA (6, 2, 0) model is quite stable.

RESULTS

Descriptive Statistics

Table 14

Description	Statistic
Mean	143480000
Median	145350000
Minimum	72208000
Maximum	209290000

Standard deviation	42680000
Skewness	-0.078367
Excess kurtosis	-1.3081

As shown above, the mean is positive, i.e. 143480000. The wide gap between the minimum (i.e 72208000) and the maximum (i.e. 209290000) is consistent with the reality that the China POP series is trending upwards. The skewness is -0.078367 and the most striking characteristic is that it is negative, indicating that the POP series is negatively skewed and non-symmetric. Excess kurtosis is -1.3081; showing that the POP series is not normally distributed.

Results Presentation[1]

Table 15

ARIMA (6, 2, 0) Model:

$$\Delta^2 POP_{t-1} = 2.55\Delta^2 POP_{t-1} - 2.97\Delta^2 POP_{t-2} + 2.37\Delta^2 POP_{t-3} - 1.64\Delta^2 POP_{t-4} + 0.93\Delta^2 POP_{t-5} - 0.29\Delta^2 POP_{t-6} \ldots\ldots\ldots\ldots [5]$$

P: (0.0000) (0.0000) (0.0000) (0.0008) (0.0078) (0.0332)

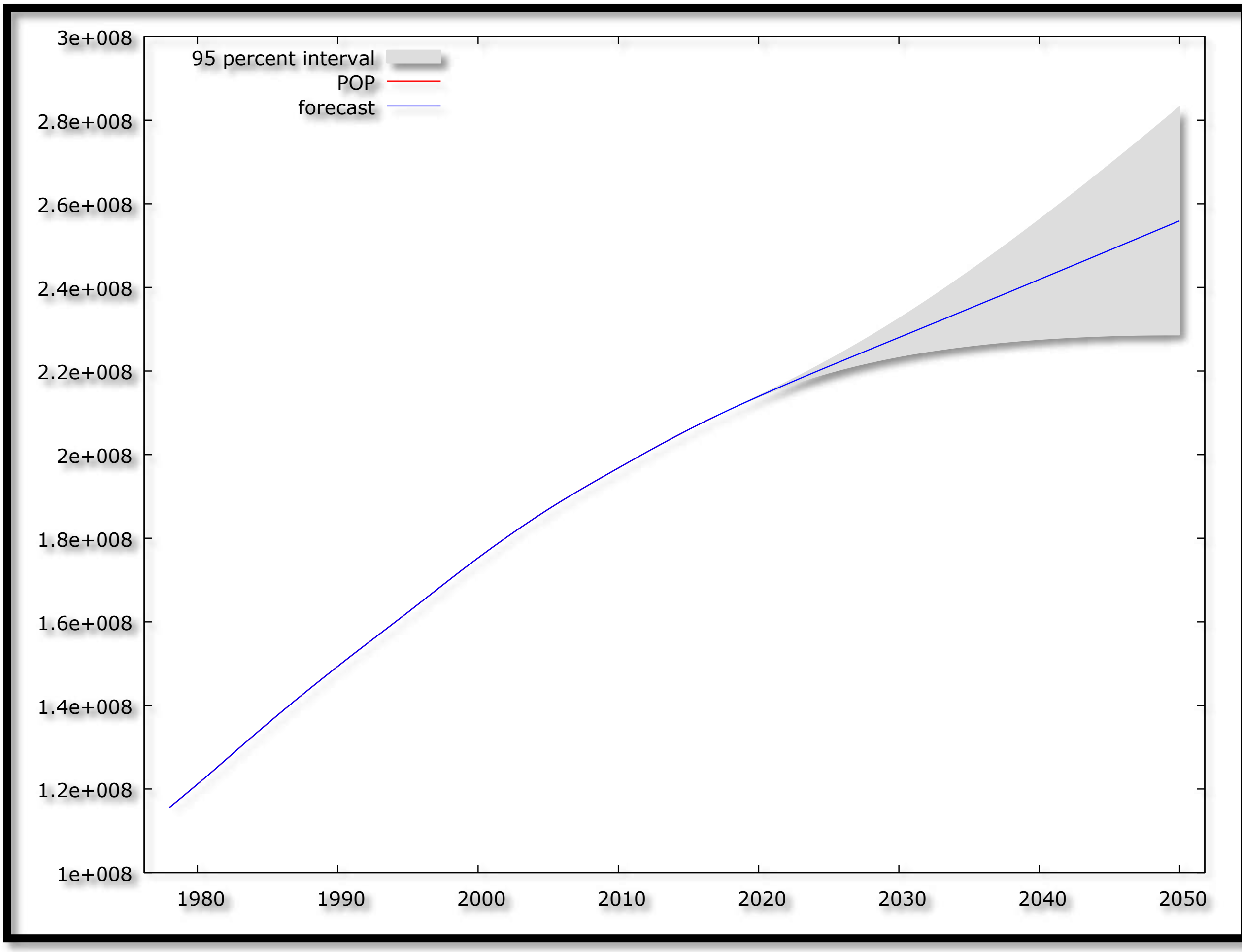
S. E: (0.1325) (0.1352) (0.4893) (0.4881) (0.3499) (0.1308)

Variable	Coefficient	Standard Error	z	p-value
AR (1)	2.54718	0.132508	19.22	0.0000***
AR (2)	-2.97218	0.1351849	-8.447	0.0000***
AR (3)	2.37238	0.489305	4.848	0.0000***
AR (4)	-1.63644	0.488103	-3.353	0.0008***
AR (5)	0.930725	0.349861	2.660	0.0078***
AR (6)	-0.278646	0.130809	-2.130	0.0332**

Forecast Graph

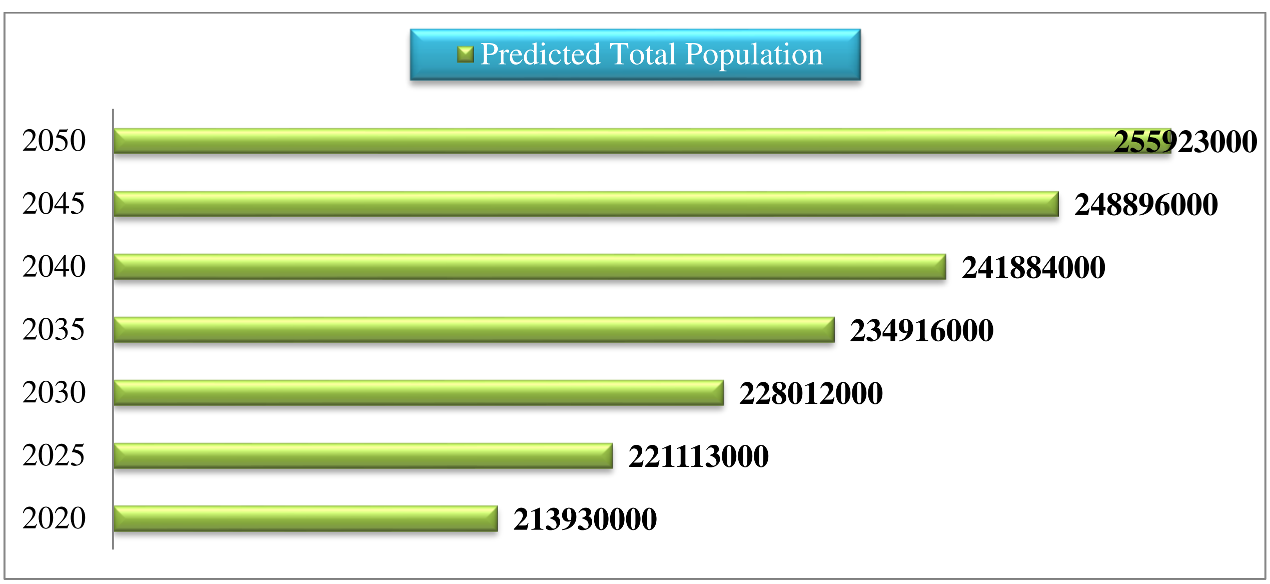
Figure 6

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



Predicted Total Population

Figure 7



Figures 6 (with a forecast range from 2018 – 2050) and 7, clearly indicate that Brazil population is indeed set to continue rising gradually, at least for the next 3 decades. With a 95% confidence interval of 228659000 to 283187000 and a projected total population of 255923000 by 2050, the chosen ARIMA (6, 2, 0) model is consistent with the population projections by the UN (2015) which forecasted that Brazil's population will be approximately 238270000 by 2050 and is also in line with the recent population projections by the UN (2017) which forecasted that Brazil's population will be approximately 232688000 by 2050.

Policy Implications

a) The government of Brazil ought to continue investing more in infrastructural development in order to cater for the projected increase in total population.
b) The predicted gradual increase in total population in Brazil justifies the need for more and bigger companies to provide for the expected increase in demand for goods and services.
c) The elderly population will increase from about 11% of the working-age population in 2005 to 49% by 2050, while the school-age population will decline from about 50% of the working-age population in 2005 to 29% by 2050. These shifts in population age structure will lead to substantial additional fiscal pressure on publicly financed healthcare and pensions, along with substantial reductions in fiscal pressures for publicly financed education (Gragnolati *et al*, 2011). This also justifies the need for the government of Brazil to plan for enough resource mobilization in order to take care of a very large number of the elderly in 3 decades' time.
d) The government of Brazil should continue encouraging the smaller family size norm.

CONCLUSION

In the case of Brazil, the study shows that the ARIMA (6, 2, 0) 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, Brazil's total population would be approximately, 256 million people. This is a warning signal to policy makers in Brazil. These results are quite necessary for the government of Brazil, 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] Center for International Research (1993). Population Trends – Brazil, US Department of Commerce, *Economics and Statistics Administration*, Bureau of the Census.

[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] 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.

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

[8] Gragnolati, M., Jorgenson, O. H., Rocha, R & Fruttero, A (2011). Growing old in an older Brazil: implications of population ageing on growth, poverty, public finance and service delivery, *World Bank*, Washington DC.

[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). 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.

[11] 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.

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

[13] 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.

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

[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.

[16] 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.

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