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

NYONI, THABANI

University of Zimbabwe

25 February 2019

What Will Be China'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 China 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 China annual total population is I (2). Based on the AIC, the study presents the ARIMA (0, 2, 1) model as the optimal model. The diagnostic tests further indicate that the presented model is indeed stable although its residuals are non-stationary. The results of the study reveal that total population in China will continue to rise in the next three decades and in 2050 China's total population will be approximately 1.6 billion people. Three policy recommendations have been suggested for consideration by the Chinese government.

Key Words: China, 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). As a large population country, China's population characteristics include large population base, low cultural quality, accelerated aging process, high sex ratio at birth, and low urbanization (Wang & Zhong, 2017). Therefore, the population issue has drawn great attention, and it is mainly reflected in the demographic structure and population policies. In the 1980s, in order to control the population growth strictly, China began to implement the Family Planning Policy (Meng *et al*, 2014). Following this policy measure, the birth rate of China dropped from 35.54% in 1970 to 11.52% in 2010; the natural growth rate dropped from27.93% in 1970 to 4.95% in 2010 (Jiang & Quan, 2018). This policy has effectively contained the problem of excessive population growth (Wang & Wang, 2018). Due to the first drop in the total labor force population in China in 2012 and the problem of aging in recent years, the party and government have formulated the policy of "Two Children Alone" and have developed into a policy of "Full Two Children" (Han, 2012). In China, just like in any other part of the world, population modeling and forecasting is very important for policy dialogue. This study seeks to model and forecast population of China using the Box-Jenkins ARIMA approach.

REVIEW OF PREVIOUS STUDIES

Zakria & Muhammad (2009) forecasted population using Box-Jenkins 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 employing 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 reiterated that the most suitable model for modeling and forecasting population in Ethiopia was the ARIMA (2, 1, 2) model. In the case of China, 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 by a backward shift operator as:

$$\emptyset(B)(1-B)^{d}POP_{t}=\theta(B)\mu_{t} \ldots [1]$$

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

$$\emptyset(B)=\left(1-\emptyset_{1}B-\emptyset_{2}B^{2}-\cdots-\emptyset_{p}B^{p}\right) \ldots [2]$$

$$\theta(B)=\left(1-\theta_{1}B-\theta_{2}B^{2}-\cdots-\theta_{q}B^{q}\right) \ldots [3]$$

and

$$(1-B)^d POP_t = \Delta^d POP_t \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots . \dots \dots \dots \dots \dots .. [4]$$

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).

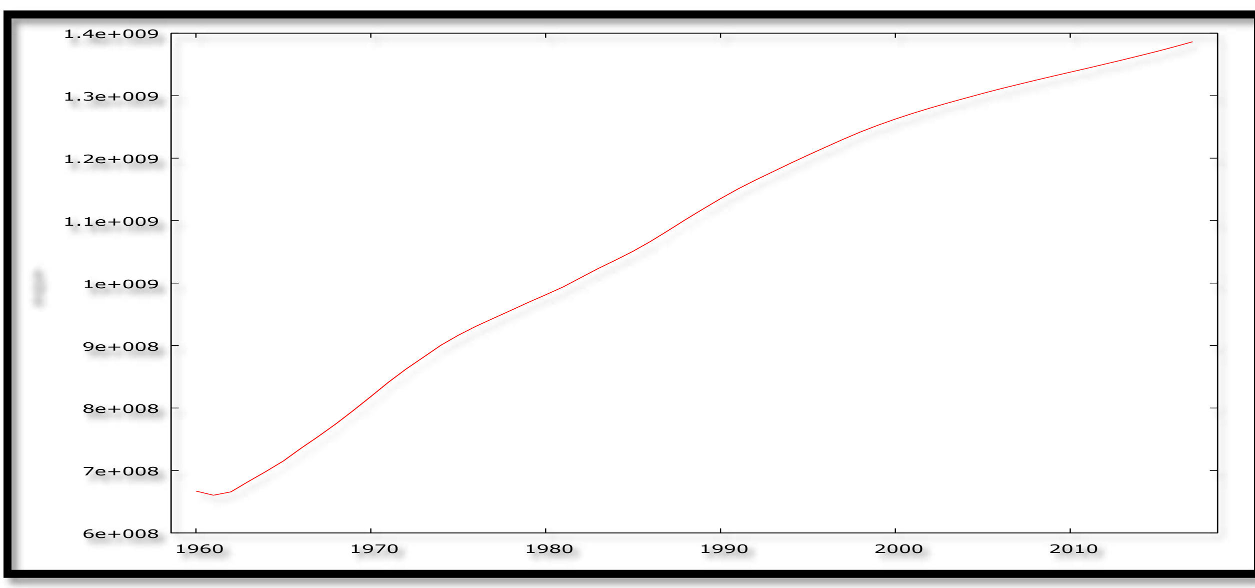
Data Collection

This study is based on 58 observations of annual total population in China; that is from 1960 to 2017. All the data was gathered from the World Bank online database, a reliable source of various macroeconomic data on literally all countries in the world. Thus the author had to choose this source on the basis of its credibility and recognition.

Diagnostic Tests & Model Evaluation

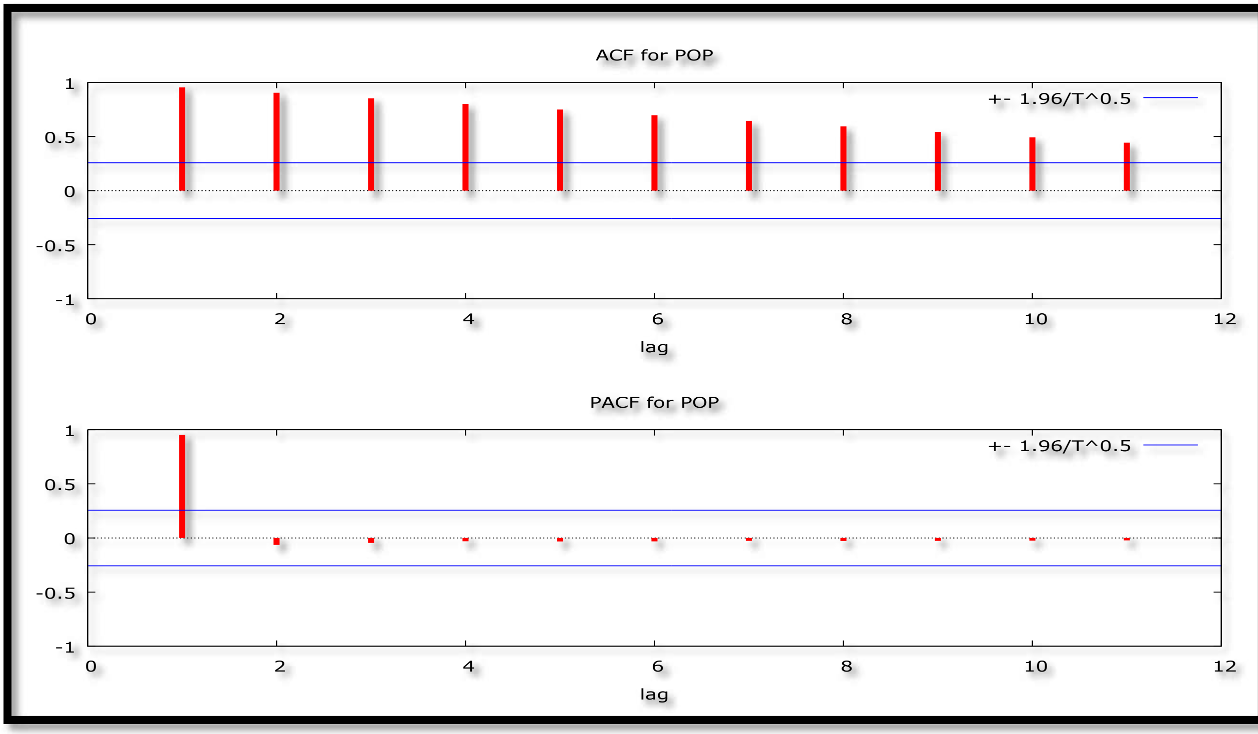
Stationarity Tests: Graphical Analysis

Figure 1



The Correlogram in Levels

Figure 2



The ADF Test

Table 1: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-2.042474	0.2684	-3.577723	@1%	Not stationary
			-2.925169	@5%	Not stationary
			-2.600658	@10%	Not stationary

Table 2: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	0.557964	0.9992	-4.165756	@1%	Not stationary
			-3.508508	@5%	Not stationary
			-3.184230	@10%	Not stationary

Table 3: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	0.258089	0.7567	-2.615093	@1%	Not stationary
			-1.947975	@5%	Not stationary
			-1.612408	@10%	Not stationary

The Correlogram (at 1st Differences)

Figure 3

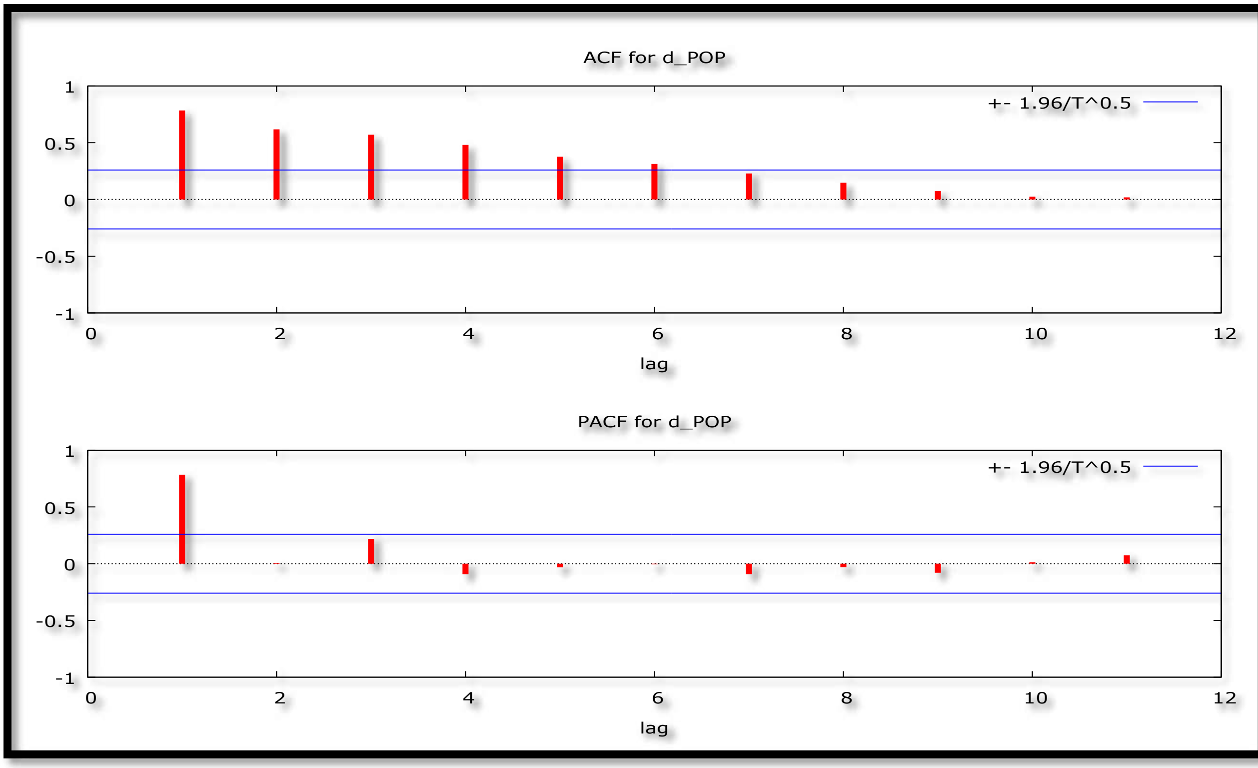


Table 4: 1st Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.950022	0.3076	-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	-2.522159	0.3166	-4.165756	@1%	Not stationary
			-3.508508	@5%	Not stationary
			-3.184230	@10%	Not stationary

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

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.561088	0.1103	-2.615093	@1%	Not stationary
			-1.947975	@5%	Not stationary
			-1.612408	@10%	Not stationary

Figures 1 – 3 and tables 1 – 6 demonstrate that the POP series is non-stationary in levels and in first differences.

The Correlogram in (2nd Differences)

Figure 4

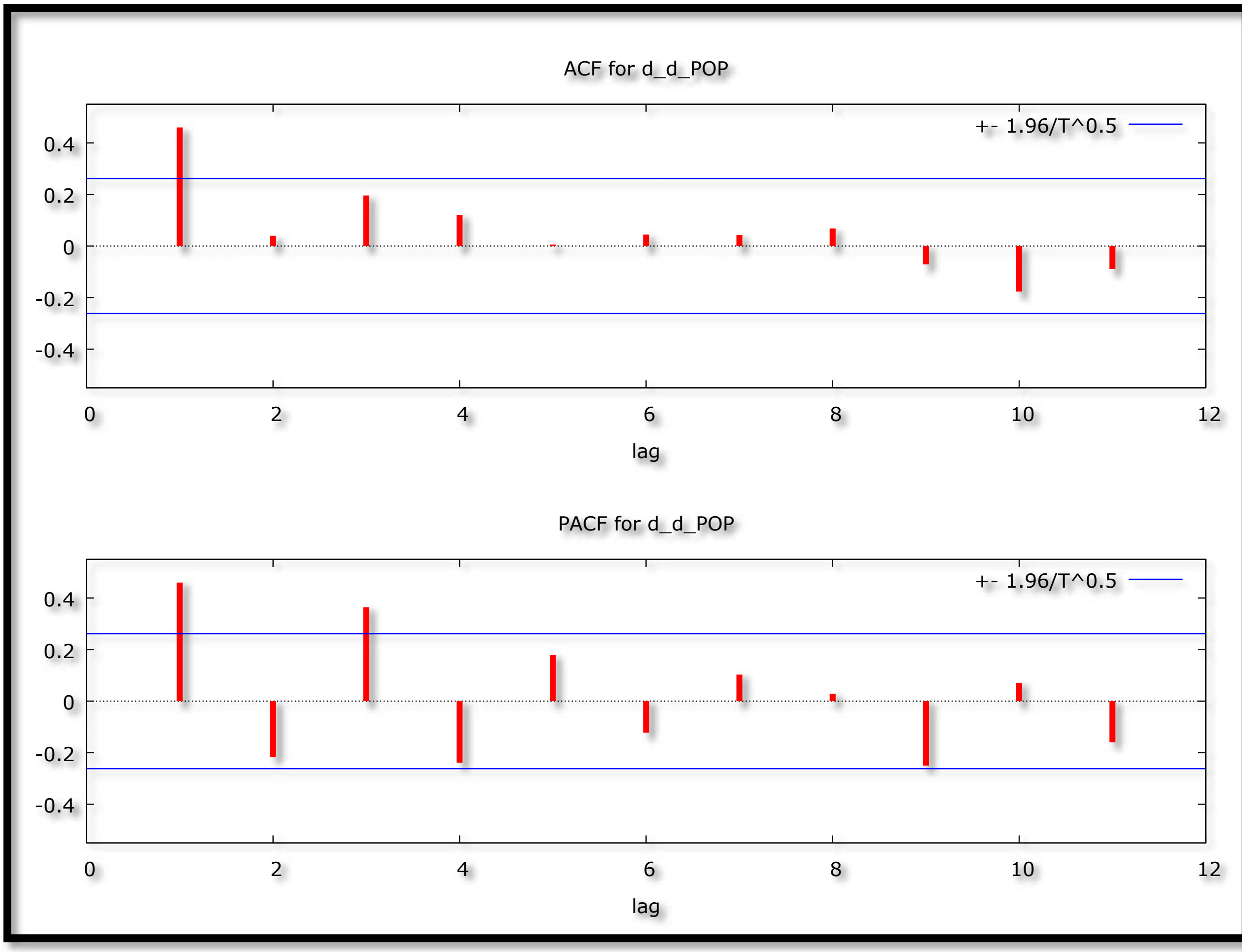


Table 7: 2nd Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-3.206226	0.0258	-3.577723	@1%	Not stationary
			-2.925169	@5%	Stationary
			-2.600658	@10%	Stationary

Table 8: 2nd Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-3.173422	0.1022	-4.165756	@1%	Not stationary
			-3.508508	@5%	Not stationary
			-3.184230	@10%	Not stationary

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

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-2.924307	0.0043	-2.615093	@1%	Stationary
			-1.947975	@5%	Stationary
			-1.612408	@10%	Stationary

While table 8 indicates that the POP series is non-stationary, tables 7 and 9 show that the POP series 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)	1778.197	0.12094	169320	1032200	2144300	0.12631
ARIMA (2, 2, 1)	1780.096	0.12209	151490	1027200	2150000	0.12585
ARIMA (1, 2, 0)	1784.353	0.14663	59482	1068700	2307600	0.13138
ARIMA (1, 2, 2)	1778.601	0.12953	40915	1046500	2198300	0.12811
ARIMA (1, 2, 3)	1779.233	0.14463	-1944.1	998110	2294400	0.12121
ARIMA (0, 2, 1)	**1777.362**	**0.11663**	187130	1019700	2114100	0.12439

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 China. Therefore, the ARIMA (0, 2, 1) model is chosen.

Residual & Stability Tests

ADF Tests of the Residuals of the ARIMA (0, 2, 1) Model

Table 11: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
R_t	-2.042474	0.2684	-3.577723	@1%	Not stationary
			-2.925169	@5%	Not stationary
			-2.600658	@10%	Not stationary

Table 12: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
R_t	0.557964	0.9992	-4.165756	@1%	Not stationary
			-3.508508	@5%	Not stationary
			-3.184230	@10%	Not stationary

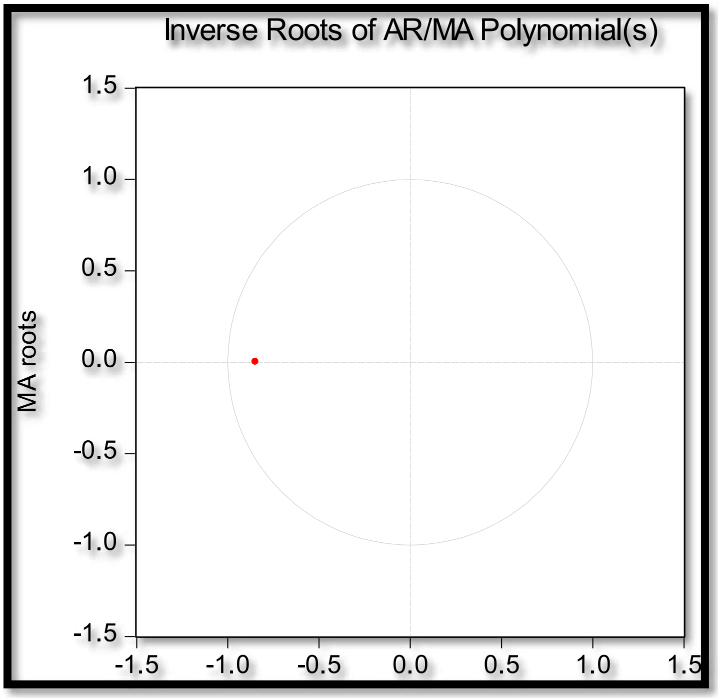
Table 13: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
R_t	0.258089	0.7567	-2.615093	@1%	Not stationary
			-1.947975	@5%	Not stationary
			-1.612408	@10%	Not stationary

Tables 11 – 13 indicate that the residuals of the ARIMA (0, 2, 1) model are non-stationary.

Stability Test of the ARIMA (0, 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 (0, 2, 1) model is quite stable.

RESULTS

Descriptive Statistics

Table 14

Description	Statistic
Mean	1077300000
Median	1110100000
Minimum	660330000
Maximum	1386400000
Standard deviation	229150000
Skewness	-0.36394
Excess kurtosis	-1.1582

As shown above, the mean is positive, i.e. 1077300000. The wide gap between the minimum (i.e 660330000) and the maximum (i.e. 1386400000) is consistent with the reality that the China POP series is trending upwards. The skewness is -0.36394 and the most striking characteristic is that it is negative, indicating that the POP series is negatively skewed and non-symmetric. Nyoni & Bonga (2017) emphasize that the rule of thumb for kurtosis is that it must be approximately 3 for normally distributed variables but in this paper, excess kurtosis is -1.1582; showing that the POP series is not normally distributed.

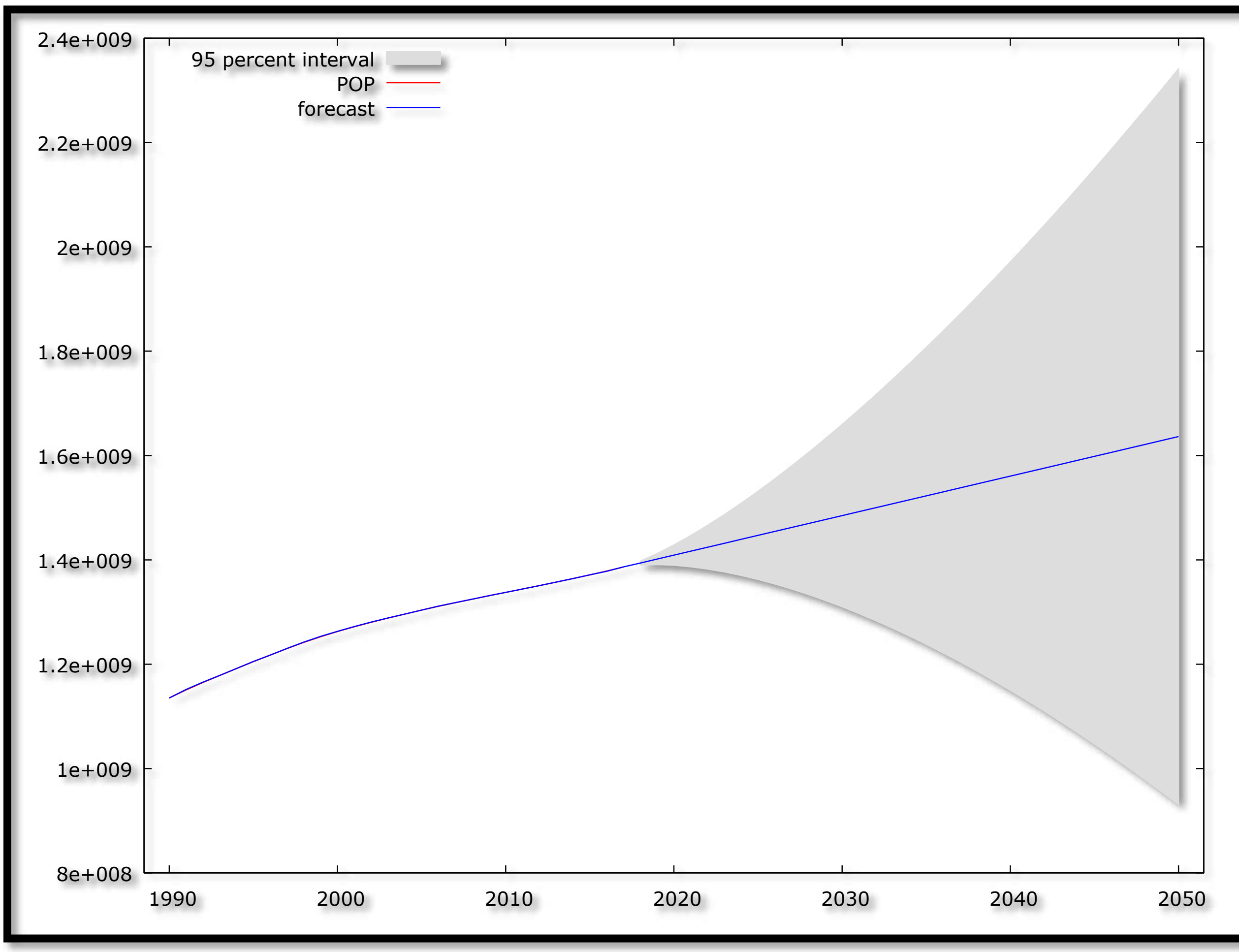
Results Presentation[1]

Table 15

ARIMA (0, 2, 1) Model:				
$\Delta^2 POP_{t-1} = 0.819298\mu_{t-1}$ … … … … … … … … … … … … … … … … … . … … … … … … … … … … … … [5]				
P:	(0.0000)			
S. E:	(0.0873057)			
Variable	Coefficient	Standard Error	z	p-value
MA (1)	0.819298	0.0873057	9.384	0.0000***

Forecast Graph

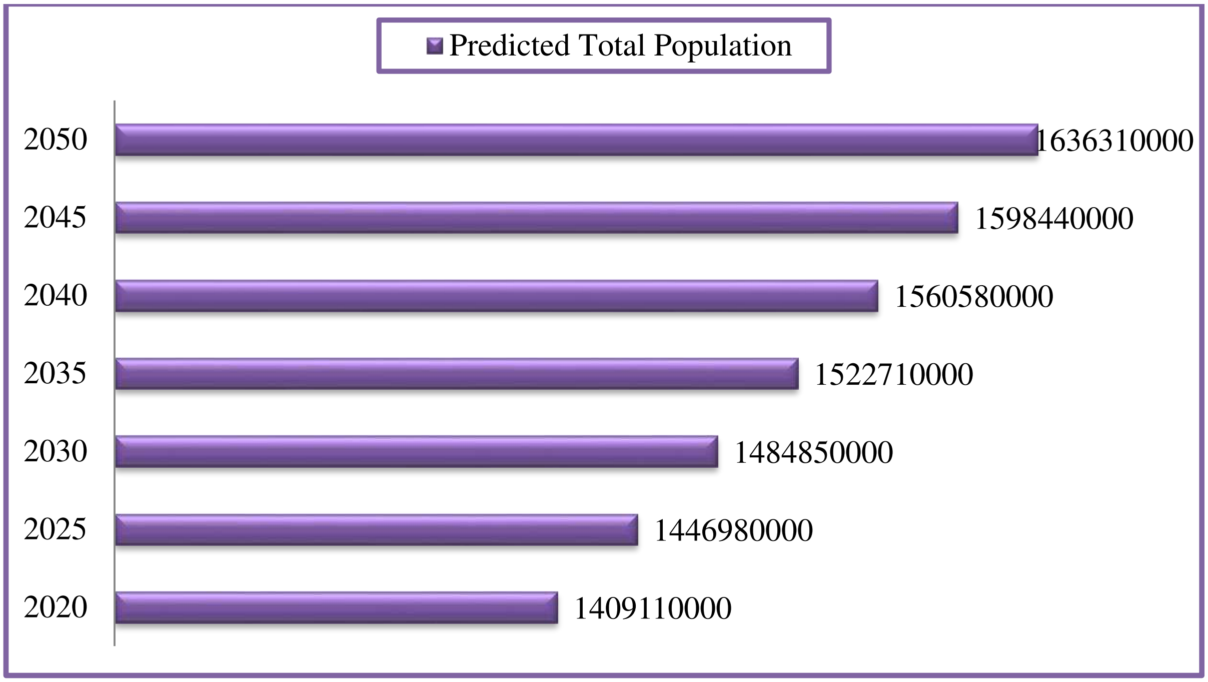
Figure 6



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

Predicted Total Population

Figure 7



China's expected continued population growth until 2050 is apparently inevitable, even if the decline of fertility accelerates. Figures 6 (with a forecast range from 2018 – 2050) and 7, clearly indicate that China population is indeed set to continue rising gradually, at least for the next 3 decades. With a 95% confidence interval of 931320000 to 2341300000 and a projected total population of 1636310000 by 2050, the chosen ARIMA (0, 2, 1) model is consistent with the population projections by the UN (2015) which forecasted that China's population will be approximately 1348056000 by 2050. Our optimal model is also consistent with the population projections by the UN (2017) which forecasted that China's population will be approximately 1364457 by 2050.

Policy Implications

1. The government of China ought to continue investing more in infrastructural development in order to cater for the expected increase in total population.
2. The predicted gradual increase in total population in China justifies the need for more and bigger companies to provide for the anticipated increase in demand for goods and services.
3. The government of China ought to continue promoting the smaller family size norm.

CONCLUSION

In the case of China, the study shows that the ARIMA (0, 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, China's total population would be approximately, 1.6 billion people. This is a warning signal to policy makers in China. These findings are essential for the government of China, 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] Han, X (2012). Leslie model of China's future population strategy simulation, *Dongbai University of Finance and Economics*.

[8] Jiang, Y & Quan, M (2018). China's population structure, pension insurance and household consumption, *Economic Survey*.

[9] Meng, L., Li, C & Hu, G (2014). Prediction of China's population structure based on PDE model, *Chinese Journal of Population, Resources and Environment*, 24 (2): 132 – 144.

[10] 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.

[11] 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.

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

[13] 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.

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

[15] 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.

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

[17] 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.

[18] 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.

[19] Wang, P & Wang, J (2018). Study on population structure change and policy development based on Leslie model, *Humanities and Social Sciences*, 6 (1): 12 – 18.

[20] Wang, Y & Zhong, F (2017). Changes in the birthrate of China population and prediction since 1990, *Population and Economy*.

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