Box-Jenkins ARIMA approach to predicting total population in Russia

NYONI, THABANI

University of Zimbabwe

19 February 2019

BOX – JENKINS ARIMA APPROACH TO PREDICTING TOTAL POPULATION IN RUSSIA

Nyoni, Thabani

Department of Economics

University of Zimbabwe

Harare, Zimbabwe

Email: nyonithabani35@gmail.com

Abstract

Employing annual time series data on total population in Russia 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 Russia annual total population is I (2). Based on the AIC, the study presents the ARIMA (1, 2, 1) model as the optimal model. The diagnostic tests further indicate that the presented model is quite stable and that its residuals are stationary as well. The results of the study reveal that total population in Russia will continue to rise, but slowly, in the next three decades and in 2050 Russia's total population will be approximately 147 million people. Three policy prescriptions have been suggested for consideration by the government of the federation of Russia.

Key Words: Forecasting, Population, Russia

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). Russia is facing demographic challenges that are common to all developed countries, but significantly aggravated by a range of historic circumstances that have become highly favourable over the course of many demographic processes. Among the main challenges are very high mortality, very low fertility and, as a result, the continued negative natural increase and overall population decline in the country. Now, these challenges are exacerbated by new ones, connected with a worsening are balance, the decrease in working-age population and the growth of dependency ratio, especially as a consequence of an ageing population (Vishnevsky, 2009). In Russia, 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 Russia using the Box-Jenkins ARIMA approach.

REVIEW OF PREVIOUS STUDIES

Employing 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) investigated population growth of Bangladesh using an Autoregressive Time Trend (ATT) model making use of a data set from 1965 to 2003 and demonstrated that there will be a downward population growth for Bangladesh for the extended period up to 2043. Ayele & Zewdie (2017) looked at human population size and its trends in Ethiopia using Box-Jenkins ARIMA models and made use of annual data from 1961 to 2009 and showed that the most suitable model for modeling and forecasting population in Ethiopia was the ARIMA (2, 1, 2) model. In the case of Russia, 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 mathematical formulation of the ARIMA (p, d, q) model using lag polynomials can be simply written as:

$$\left(1-\sum_{i=1}^{p}\emptyset_i L^i\right)(1-L)^d POP_t = \left(1+\sum_{j=1}^{q}\theta_j L^j\right)\mu_t \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..[1]$$

Where p and q are orders of the autoregressive (AR) and moving average (MA) components respectively and d is the number of times the series is differenced.

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).

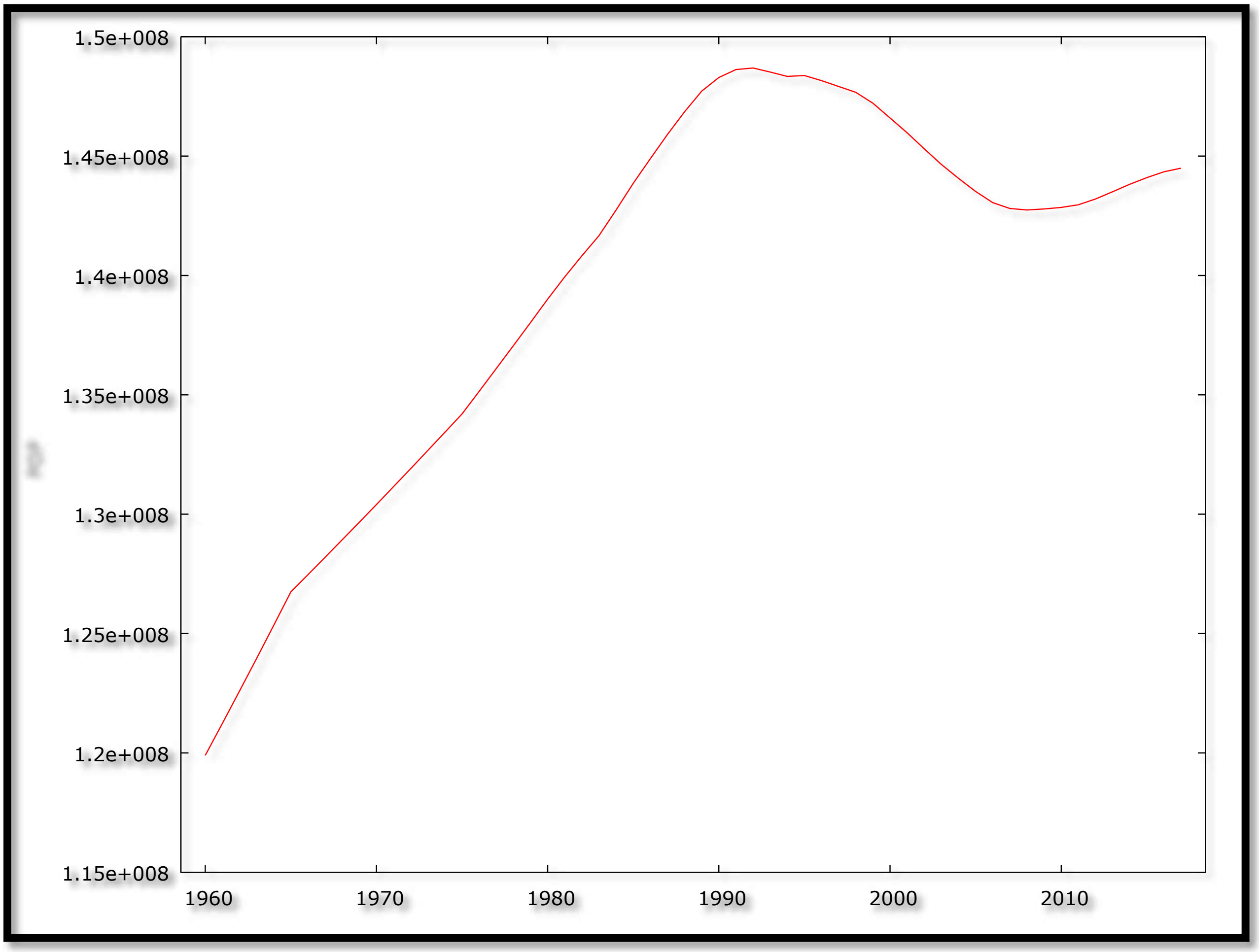
Data Collection

This study is based on 58 observations of annual total population in Russia. All the data was obtained from the World Bank online database. This database is a reliable source of various macroeconomic data; therefore the researcher chose this source on the basis of its credibility and integrity.

Diagnostic Tests & Model Evaluation

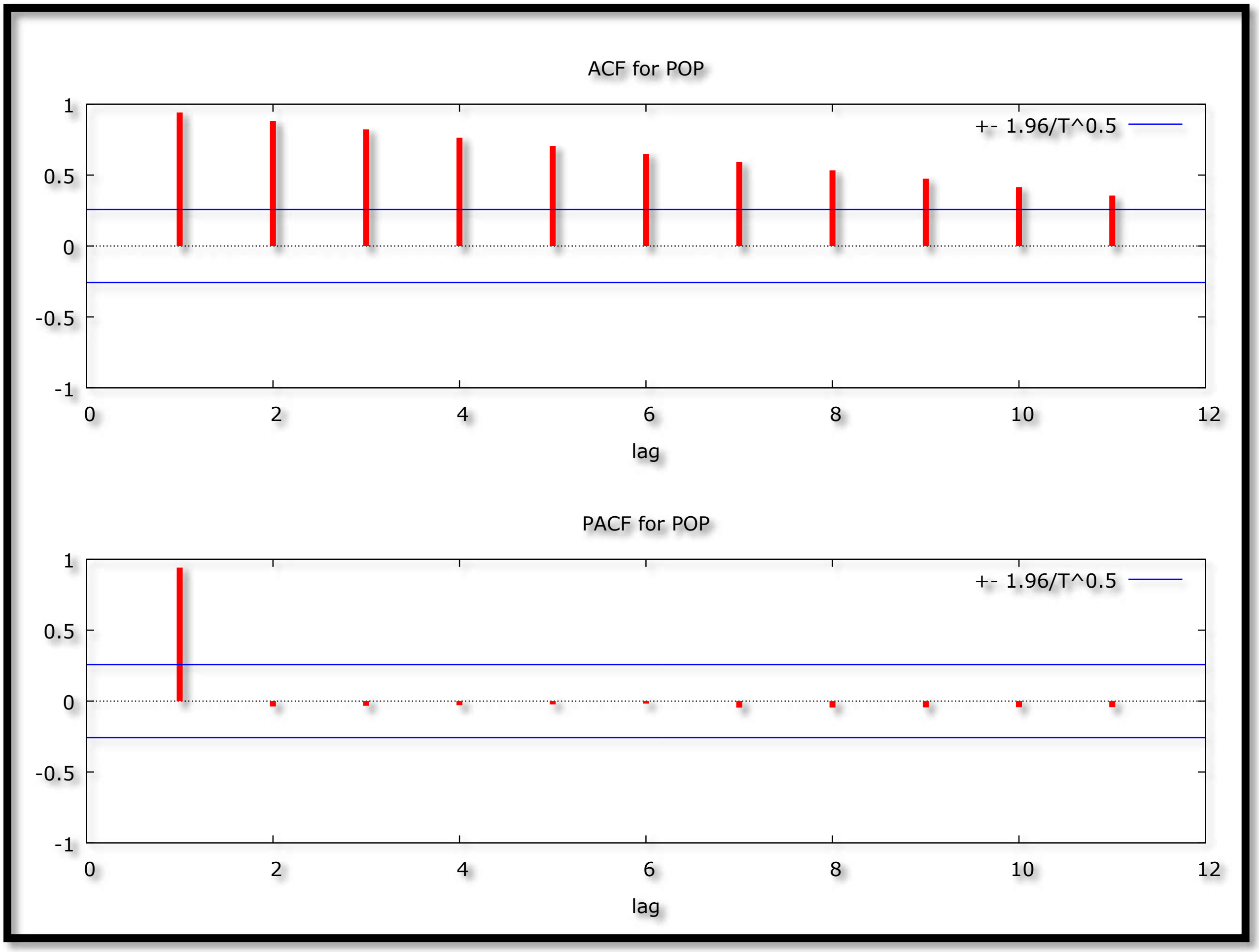
Stationarity Tests: Graphical Analysis

Figure 1



The Correlogram in Levels

Figure 2



The ADF Test

Table 1: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.829948	0.3625	-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	-2.457734	0.3472	-4.130526	@1%	Not stationary
			-3.492149	@5%	Not stationary
			-3.174802	@10%	Not stationary

Table 3: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	0.107374	0.7126	-2.607686	@1%	Not stationary

			-1.946878	@5%	Not stationary
			-1.612999	@10%	Not stationary

The Correlogram (at 1^{st} Differences)

Figure 3

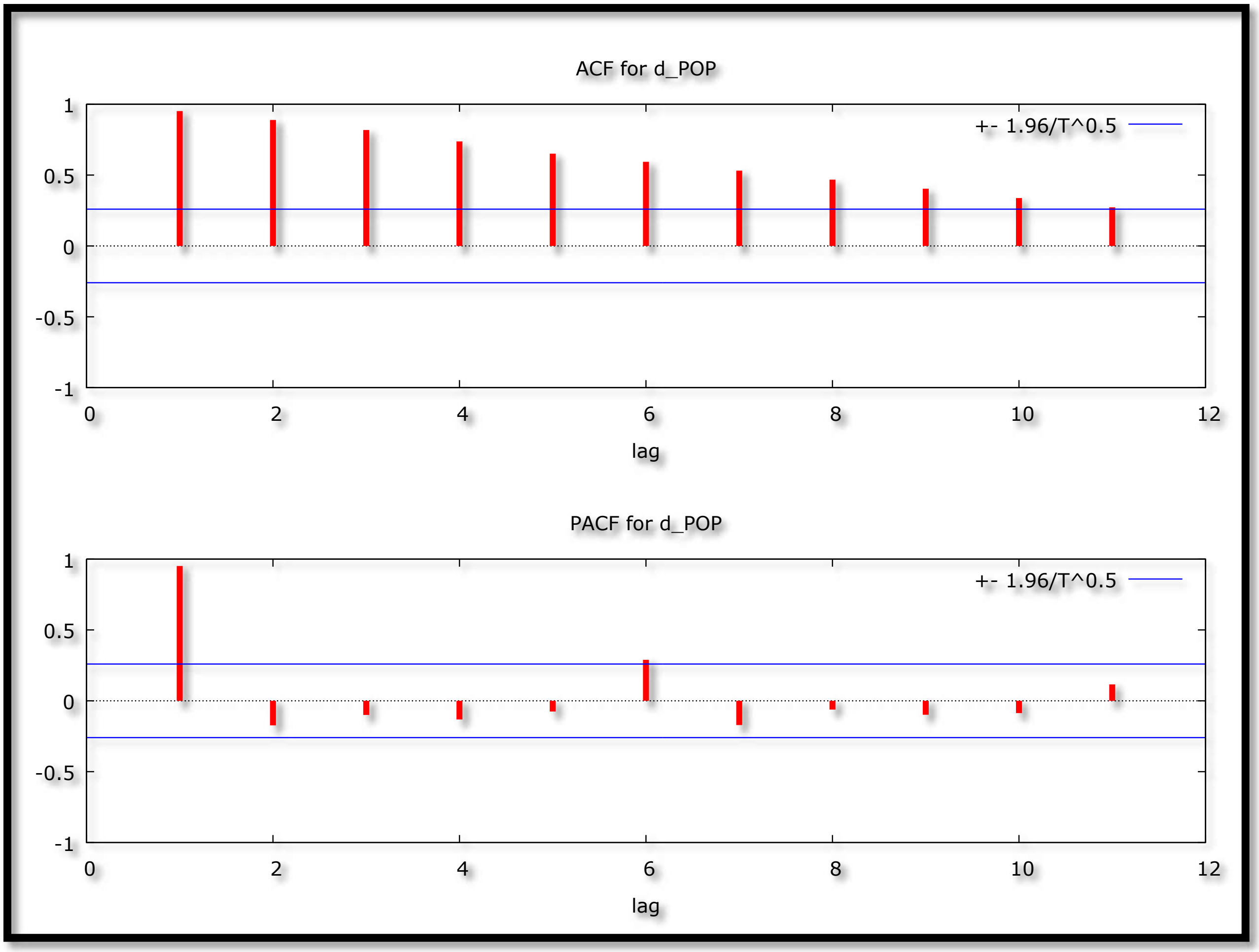


Table 4: 1^{st} Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.628881	0.4613	-3.555023	@1%	Not stationary
			-2.915522	@5%	Not stationary
			-2.595565	@10%	Not stationary

Table 5: 1^{st} Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.505618	0.8158	-4.133838	@1%	Not stationary
			-3.493692	@5%	Not stationary
			-3.175693	@10%	Not stationary

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

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.859943	0.0604	-2.607686	@1%	Not stationary
			-1.946878	@5%	Not stationary
			-1.612999	@10%	Stationary

The Correlogram in (2nd Differences)

Figure 4

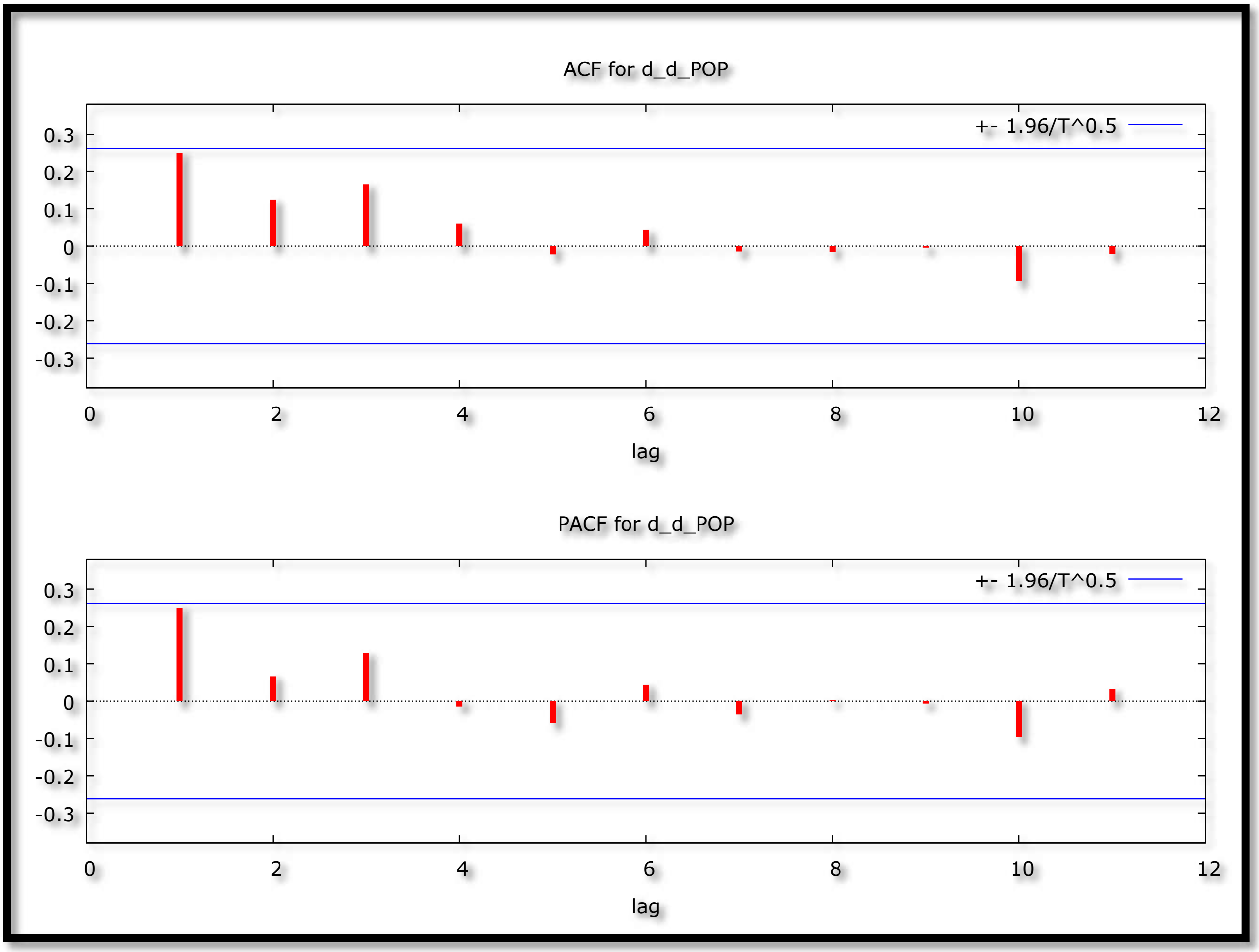


Table 7: 2nd Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-5.619845	0.0000	-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	-5.666705	0.0001	-4.133838	@1%	Stationary
			-3.493692	@5%	Stationary
			-3.175693	@10%	Stationary

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

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-5.565460	0.0000	-2.607686	@1%	Stationary
			-1.946878	@5%	Stationary
			-1.612999	@10%	Stationary

Figure 1 – 3 and tables 1 – 6 show that the POP series is neither I (0) nor I (1). Figure 4 and tables 7 – 9 indicate that the POP series is stationary after taking second differences and thus an I (2) variable.

Evaluation of ARIMA models (without a constant)

Table 10

Model	AIC	U	ME	MAPE
ARIMA (1, 2, 1)	**1490. 886**	**0.19346**	-13566	0.058666
ARIMA (2, 2, 1)	1492.869	0.19345	-13486	0.05832
ARIMA (3, 2, 1)	1494.294	0.19269	-13461	0.058279
ARIMA (0, 2, 1)	1490.2	0.19534	-17435	0.058459
ARIMA (1, 2, 2)	1492.86	**0.19345**	-13.452	0.058749

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 Russia. Therefore, the ARIMA (1, 2, 1) model is preferred.

Residual & Stability Tests

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

Table 11: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
r_t	-7.174202	0.0000	-3.557472	@1%	Stationary
			-2.916566	@5%	Stationary
			-2.596116	@10%	Stationary

Table 12: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
r_t	-7.173032	0.0000	-4.137279	@1%	Stationary
			-3.495295	@5%	Stationary
			-3.176618	@10%	Stationary

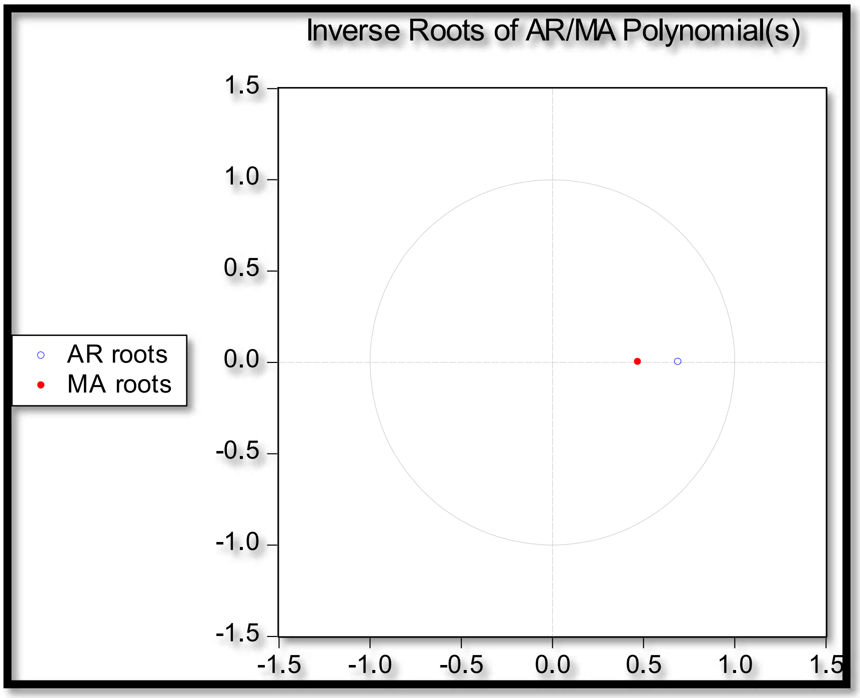
Table 13: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
r_t	-7.171142	0.0000	-2.608490	@1%	Stationary
			-1.946996	@5%	Stationary
			-1.612934	@10%	Stationary

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

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

FINDINGS

Descriptive Statistics

Table 14

Description	Statistic
Mean	139660000
Median	142910000
Minimum	119900000
Maximum	148690000
Standard deviation	8172700
Skewness	-0.87108
Excess kurtosis	-0.442265

As shown above, the mean is positive, i.e. 139660000. The wide gap between the minimum (i.e 119900000) and the maximum (i.e. 148690000) is consistent with the reality that the Russia POP series is trending upwards. The skewness is -0.87108 and the most striking characteristic is that it is negative, indicating that the POP series is negatively skewed and non-symmetric. Excess kurtosis is -0.442265; showing that the POP series is not normally distributed.

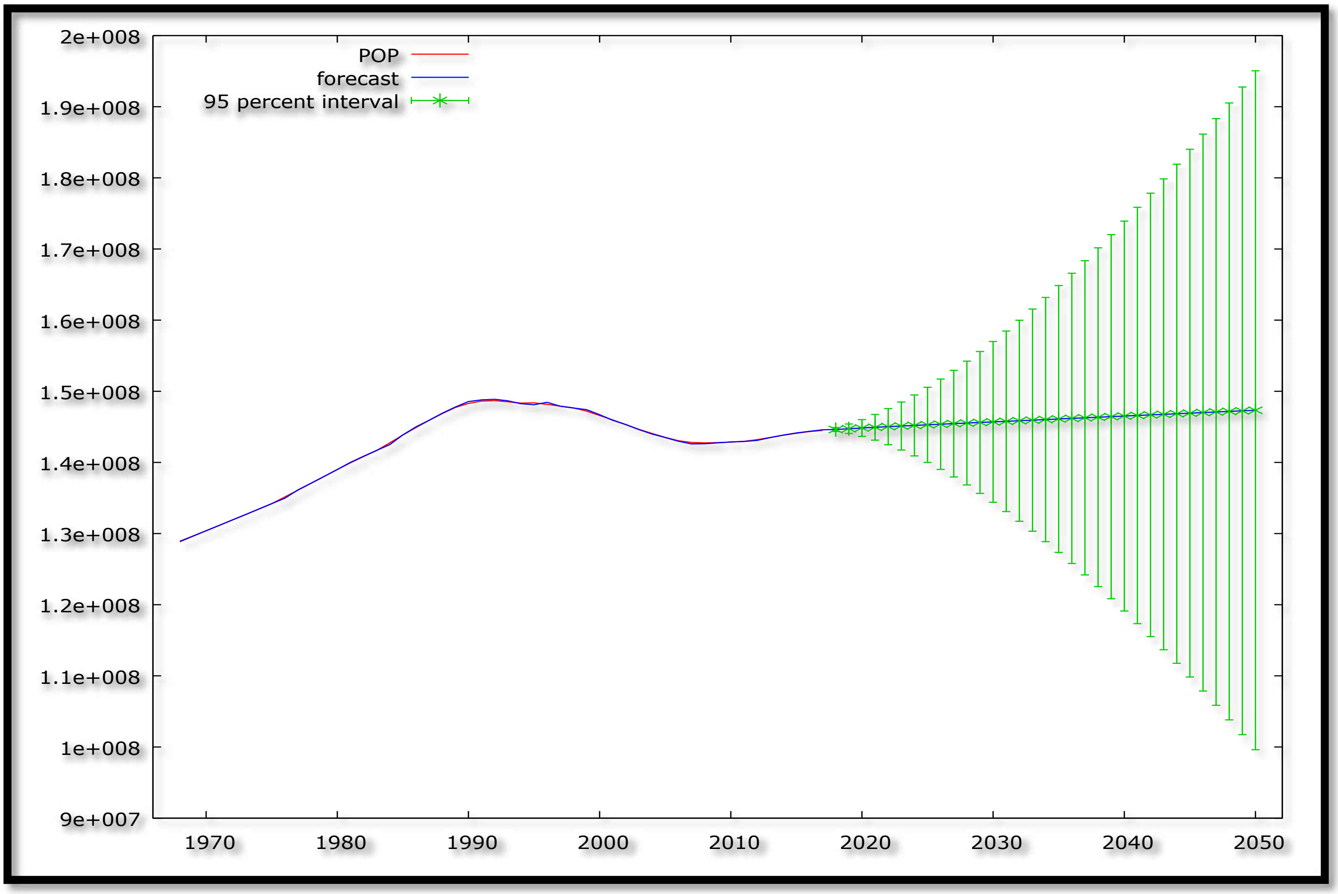
Results Presentation[1]

Table 15

ARIMA (1, 2, 1) Model:				
$\Delta^2 POP_{t-1} = 0.663471\Delta^2 POP_{t-1} - 0.442301\mu_{t-1} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.\ldots\ldots\ldots\ldots$ [2]				
P: (0.0376) (0.2493)				
S. E: (0.319036) (0.383893)				
Variable	Coefficient	Standard Error	z	p-value
AR (1)	0.663471	0.319036	2.08	0.0376**
MA (1)	-0.442301	0.383893	-1.152	0.2493

Forecast Graph

Figure 6

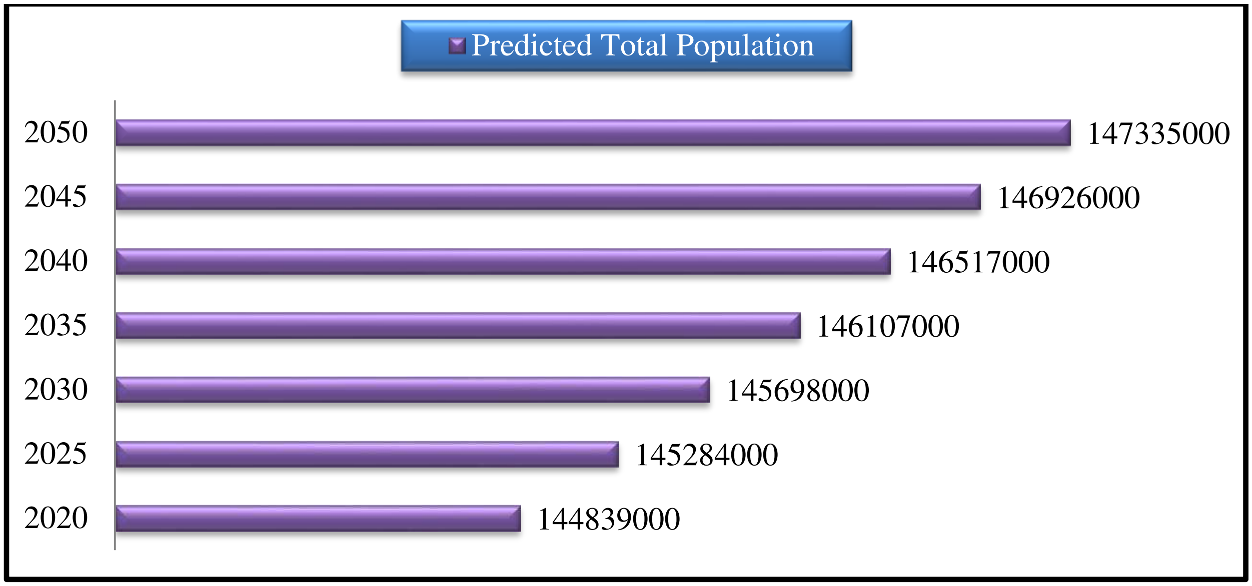


---

[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 Russia population is indeed set to continue rising but slowly, at least for the next 3 decades. With a 95% confidence interval of 99622100 to 195048000 and a projected total population of 147335000 by 2050, the chosen ARIMA (1, 2, 1) model is consistent with the population projections by the UN (2015) which forecasted that Russia's population will be approximately 128599000 by 2050 and is also in line with the recent population projections by the UN (2017) which forecasted that Russian's population will be approximately 132731000 by 2050.

Policy Implications

i. The government of the Federation of Russia ought to continue investing more in infrastructural development in order to cater for the projected increase in total population. Russia is also "greying" – therefore, there should be more health facilities in order to take care of the elderly.
ii. The predicted gradual increase in total population in Russia justifies the need for more and bigger companies to provide for the expected increase in demand for goods and services.
iii. The government of Russia should continue encouraging the smaller family size norm.

CONCLUSION

In the case of Russia, the study shows that the ARIMA (1, 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, Russia's total population would be approximately, 147 million people. This is a warning signal to policy makers in Russia, especially given the rising population of the elderly. These results are quite invaluable for the government of Russia, particulary 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] Vishnevsky, A (2009). The challenges of Russia's demographic crisis, *IFRI Russia/NIS Center*.

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