Predicting total population in India: A Box-Jenkins ARIMA approach

NYONI, THABANI

University of Zimbabwe, Department of Economics

25 February 2019

PREDICTING TOTAL POPULATION IN INDIA: 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 India from 1960 to 2017, we model and forecast total population over the next 3 decades using the Box – Jenkins ARIMA technique. Diagnostic tests show that Indian annual total population data is I (2). Based on both the AIC and Theil's U, the study presents the ARIMA (1, 2, 3) model. The diagnostic tests further confirm that the presented model is stable and quite acceptable. The results of the study reveal that total population in India will continue to sharply rise in the next three decades, thereby posing a threat to both natural and non-renewable resources. In order to deal with the threats posed by a large population in India, the study recommends family planning practices amongst other policy prescriptions.

Key Words: Population, Forecasting, India

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). India's population the one billion mark in 2000. It's population is likely to pass China's as the world's largest within 20 years (Population Foundation of India, 2007). 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 problem of forecasting or projecting population has always been an important task for the demographers (Smith, 1987). Population modeling and forecasting is important for policy dialogue, especially with regards to the future threat to natural resources, persistant unemployment and worsening poverty levels in India. This study endeavors to model and forecast population of India using the Box-Jenkins ARIMA technique.

LITERATURE REVIEW

Theoretical Literature Review

The population theory by Malthus (1798) posits that population growth is not good to economic growth and development and this is basically attributed to the argument that human population grows geometrically while the means of subsistance grows arithmetically being subject to the law of diminishing returns. While the applicability of the Malthusian population prophecy is not universal, in India is arguably relevant given perpetually shrinking employment opportunities and increasing poverty levels for an ordinary Indian. In partial disagreement with Malthus (1798), Solow (1956) argued that an increase in the "population growth rate" instead of "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). Ahlburg (1998) stressed that an increase in population growth would lead to an increase the need for goods and services through the "technology-pushed" and the "demand-pulled" channels while Becker *et al* (1999) argued that high population growth rate induces high labour force which is the source of real wealth and in such scenarios, population growth could not be a problem but an opportunity for growth.

Empirical Literature Review

Rahul *et al* (2007) projected Indian population using the MCMC technique using the data set ranging from 1901 to 2001 and established that the MCMC tool is suitable for fitting population data in India. Zakria & Muhammad (2009), in the case of Pakistan; looked at population dynamics using Box-Jenkins ARIMA models, and relied on a data set ranging from 1951 to 2007; and established that the ARIMA (1, 2, 0) model was the best model. Haque *et al* (2012) investigated Bangladesh population projections using the Logistic Population model with a data set ranging from 1991 to 2006 and revealed that the logistic population model has the best fit for population growth in Bangladesh. Ayele & Zewdie (2017) studied human population size and its pattern in Ethiopia using Box-Jenkins ARIMA models and employing annual data from 1961 to 2009 and concluded that the best model for modeling and forecasting population in Ethiopia was the ARIMA (2, 1, 2) model. In this study, the Box-Jenkins ARIMA methodology will be employed for the data set ranging from 1960 to 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:

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

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

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

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

and

$$(1-B)^{d}PIN_{t}=\Delta^{d}PIN_{t} \ldots \ldots \ldots [4]$$

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).

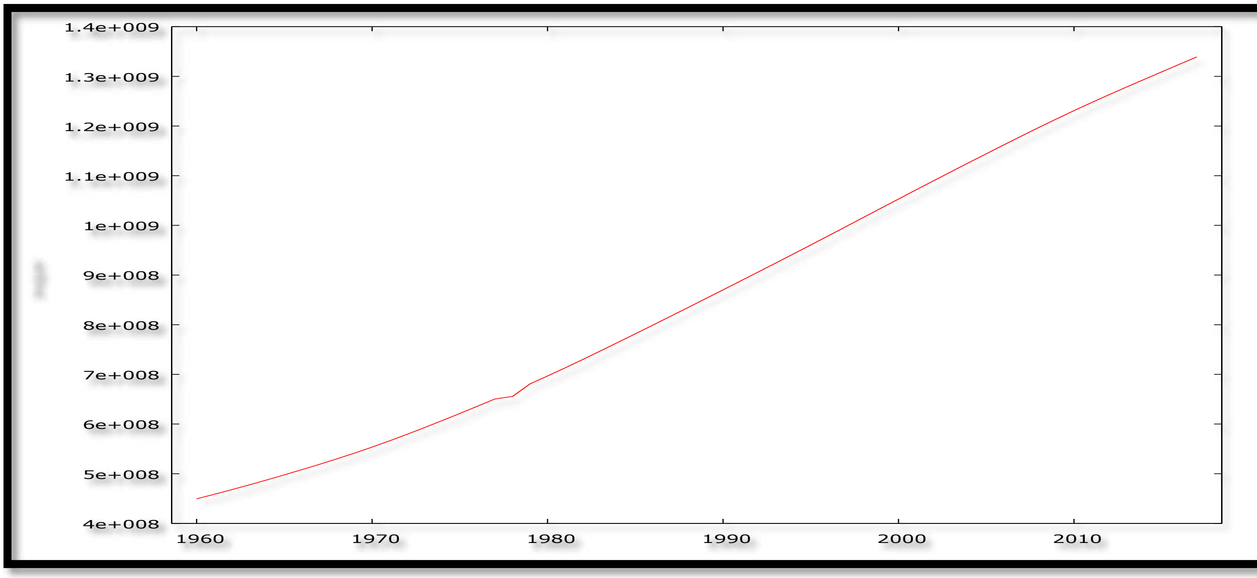
Data Collection

This research work is based on 58 observations of Indian annual total population (POP, referred to as PIN in the mathematical formulations above). All the data was adapted from the World Bank online database, a well known reliable source of various macroeconomic data.

Diagnostic Tests & Model Evaluation

Stationarity Tests: Graphical Analysis

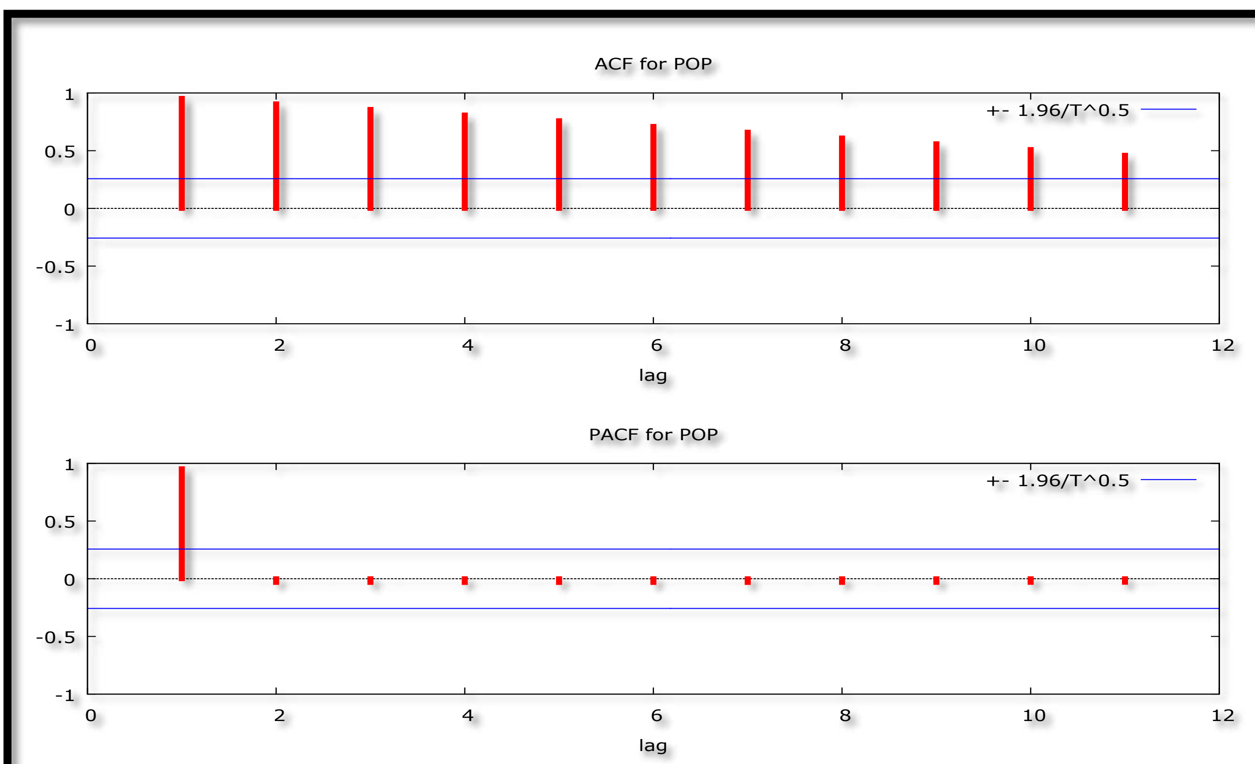
Figure 1



The POP variable, as shown in figure 1 above, is not stationary since it is trending upwards over the period 1960 – 2017. Thus 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	-1.331701	0.6083	-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	-5.674758	0.0001	-4.127338	@1%	Stationary
			-3.490662	@5%	Stationary
			-3.173943	@10%	Stationary

Table 3: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.542900	0.1143	-2.609324	@1%	Not stationary
			-1.947119	@5%	Not stationary
			-1.612867	@10%	Not stationary

The Correlogram (at 1st Differences)

Figure 3

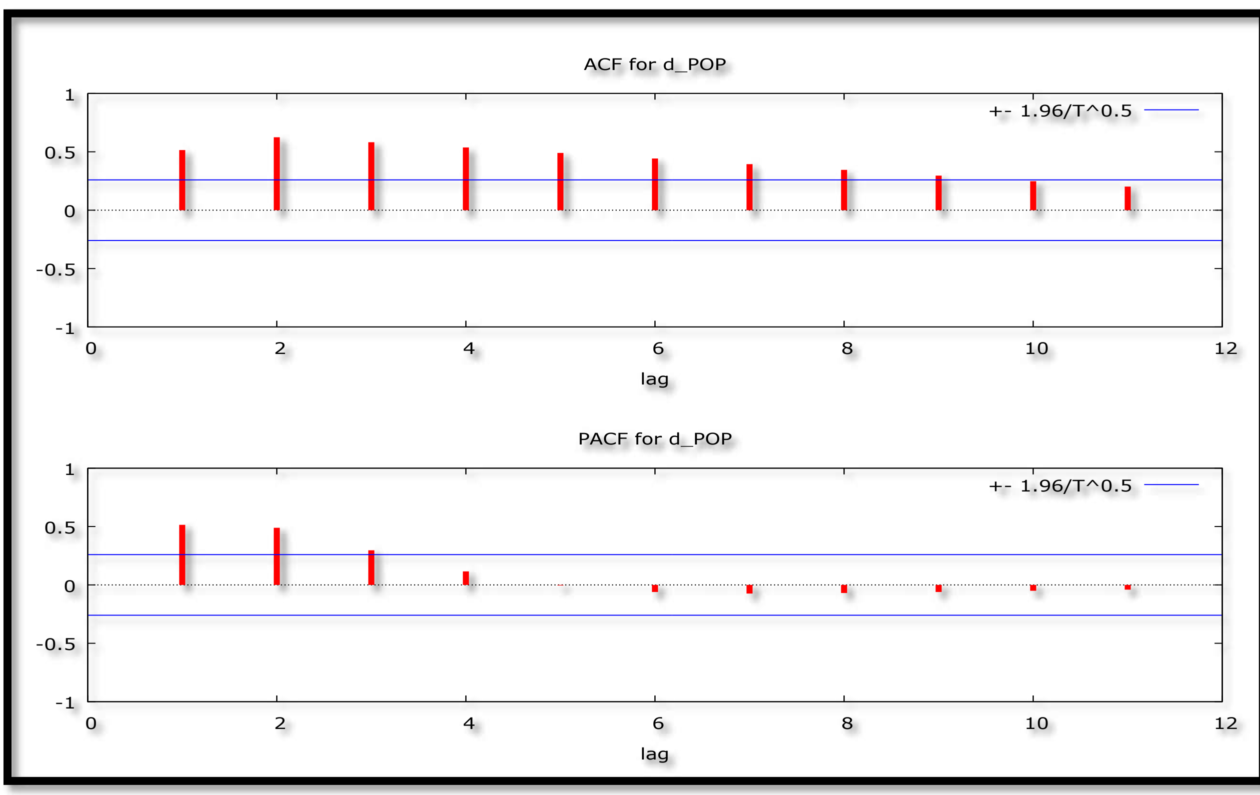


Table 4: 1st Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-2.264128	0.1871	-3.557472	@1%	Not stationary
			-2.916566	@5%	Not stationary
			-2.596116	@10%	Not stationary

Table 5: 1st Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-0.709656	0.9670	-4.140858	@1%	Not stationary
			-3.496960	@5%	Not stationary
			-3.177579	@10%	Not stationary

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

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	0.621389	0.8475	-2.609324	@1%	Not stationary
			-1.947119	@5%	Not stationary
			-1.612867	@10%	Not stationary

Figures above, i.e. 2 and 3 and tables above, i.e. 1 – 6 basically indicate that the POP series is not stationary at both levels and in first differences; hence the need to difference the POP series for the second time.

The Correlogram in (2nd Differences)

Figure 4

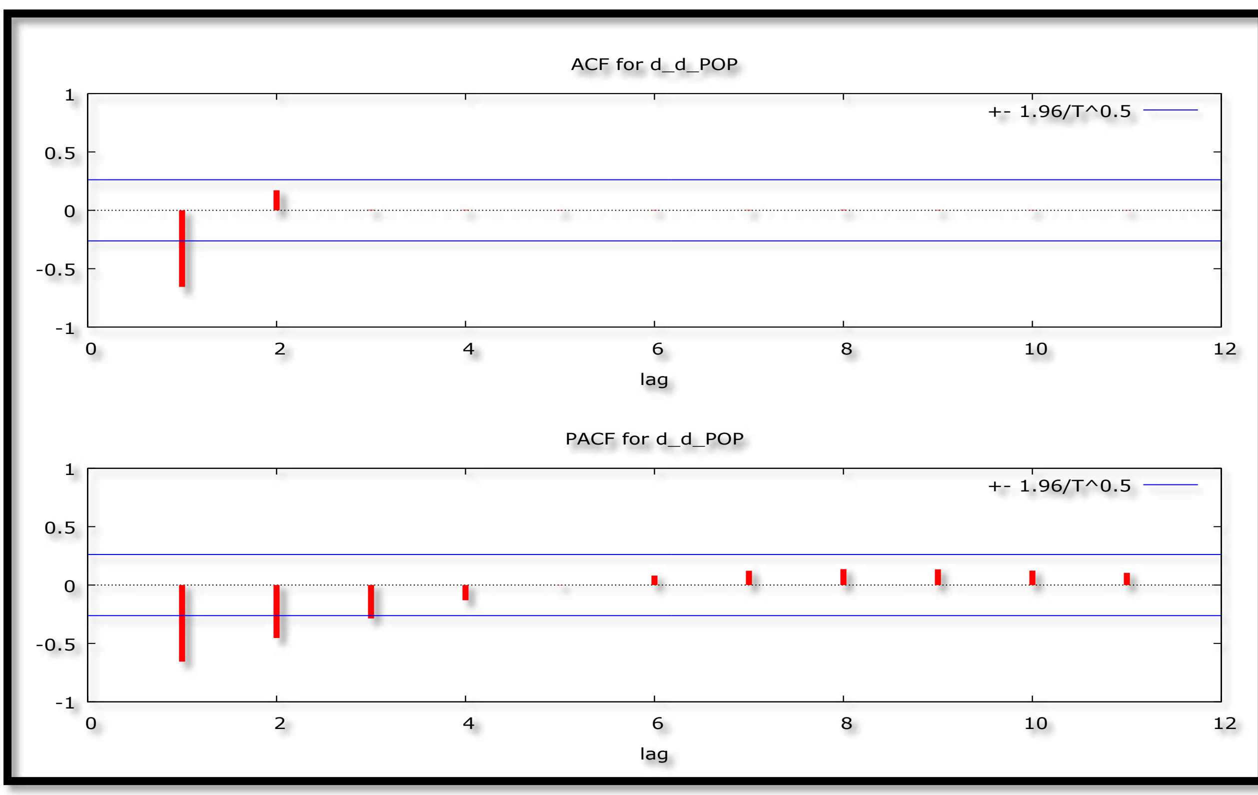


Table 7: 2nd Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-7.807916	0.0000	-3.560019	@1%	Stationary
			-2.917650	@5%	Stationary
			-2.596689	@10%	Stationary

Table 8: 2nd Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-7.114147	0.0000	-4.144584	@1%	Stationary
			-3.498692	@5%	Stationary
			-3.178578	@10%	Stationary

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

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-7.737225	0.0000	-2.609324	@1%	Stationary
			-1.947119	@5%	Stationary
			-1.612867	@10%	Stationary

Figure 4 above shows that most of the autocorrelation coefficients are now closer to zero (expect at lag 1) which is, basically, a feature of a stationary series. Tables 7 – 9, all indicate that the POP series is now stationary and thus an I (2) variable.

Evaluation of ARIMA models (without a constant)

Table 10

Model	AIC	U	MAPE
ARIMA (1, 2, 1)	1801.077	0.16897	0.15332
ARIMA (2, 2, 1)	1801.028	0.1657	0.15542
ARIMA (1, 2, 2)	1796.667	0.15878	0.1557
ARIMA (1, 2, 3)	**1794.048**	**0.1491**	0.14725
ARIMA (1, 2, 4)	1796.023	**0.1491**	0.14669
ARIMA (2, 2, 2)	1794.959	0.15207	0.14472

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 researcher will rely on the AIC and Theil's U in order to select the optimal model. Hence, the ARIMA (1, 2, 3) model is chosen.

Residual & Stability Tests

ADF Tests of the Residuals of the ARIMA (1, 2, 3) Model

Table 11: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
r_t	-7.339556	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	-7.605320	0.0000	-4.137279	@1%	Stationary
			-3.495295	@5%	Stationary
			-3.176618	@10%	Stationary

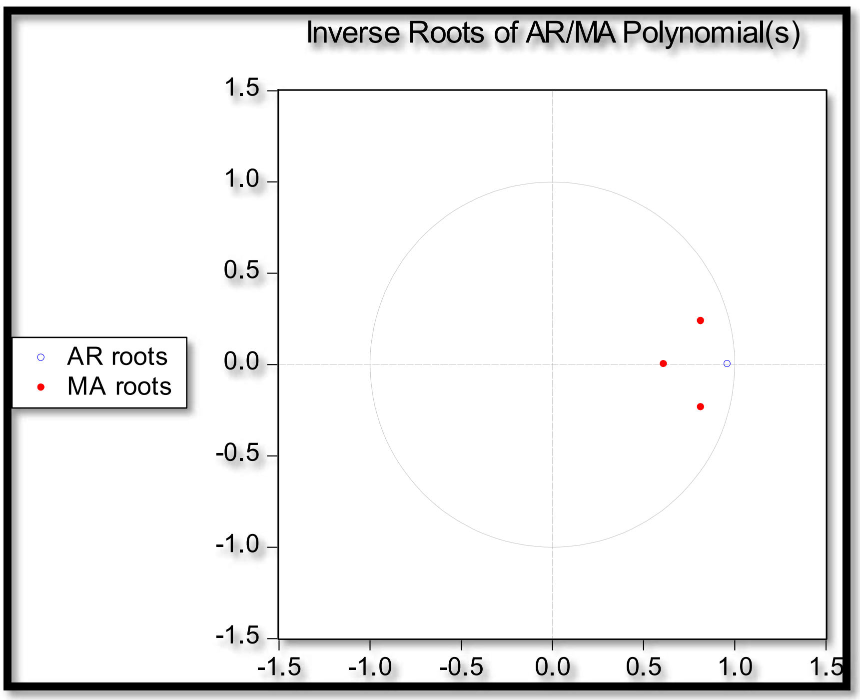
Table 13: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
r_t	-7.290935	0.0000	-2.608490	@1%	Stationary
			-1.946996	@5%	Stationary
			-1.612934	@10%	Stationary

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

Stability Test of the ARIMA (1, 2, 3) Model

Figure 5



As shown above in figure 5, the ARIMA (1, 2, 3) model, is quite stable, as the corresponding inverse roots of the characteristic polynomial lie in the unit circle.

RESULTS & DISCUSSION

Descriptive Statistics

Table 14

Description	Statistic
Mean	861470000
Median	843380000

Minimum	449480000
Maximum	133920000
Standard deviation	278510000
Skewness	0.14852
Excess kurtosis	-1.2951

As shown in the table above, the mean is positive, i.e. 861470000. The wide gap between the minimum (i.e. 449480000) and the maximum (i.e. 133920000) is consistent with the reality that the POP series is sharply trending upwards. The skewness is 0.14852 and it's positive, indicating that the POP series is positively skewed and non-symmetric. Kurtosis is -1.2951; indicating that the POP series is not normally distributed.

Results Presentation[1]

Table 15

ARIMA (1, 2, 3) Model:

$$\Delta^2 POP_{t-1} = 0.9577\Delta^2 POP_{t-1} - 2.1883\mu_{t-1} + 1.6321\mu_{t-2} - 0.399\mu_{t-3} \ldots \ldots \ldots \ldots \ldots . \ldots \ldots [5]$$

P: (0.0000) (0.0000) (0.0000) (0.0045)

S. E: (0.0594) (0.1446) (0.2596) (0.1406)

Variable	Coefficient	Standard Error	z	p-value
AR (1)	0.957716	0.0594305	16.11	0.0000***
MA (1)	-2.18834	0.144606	-15.13	0.0000***
MA (2)	1.63213	0.259582	6.288	0.0000***
MA (3)	-0.399048	0.140640	-2.837	0.0045***

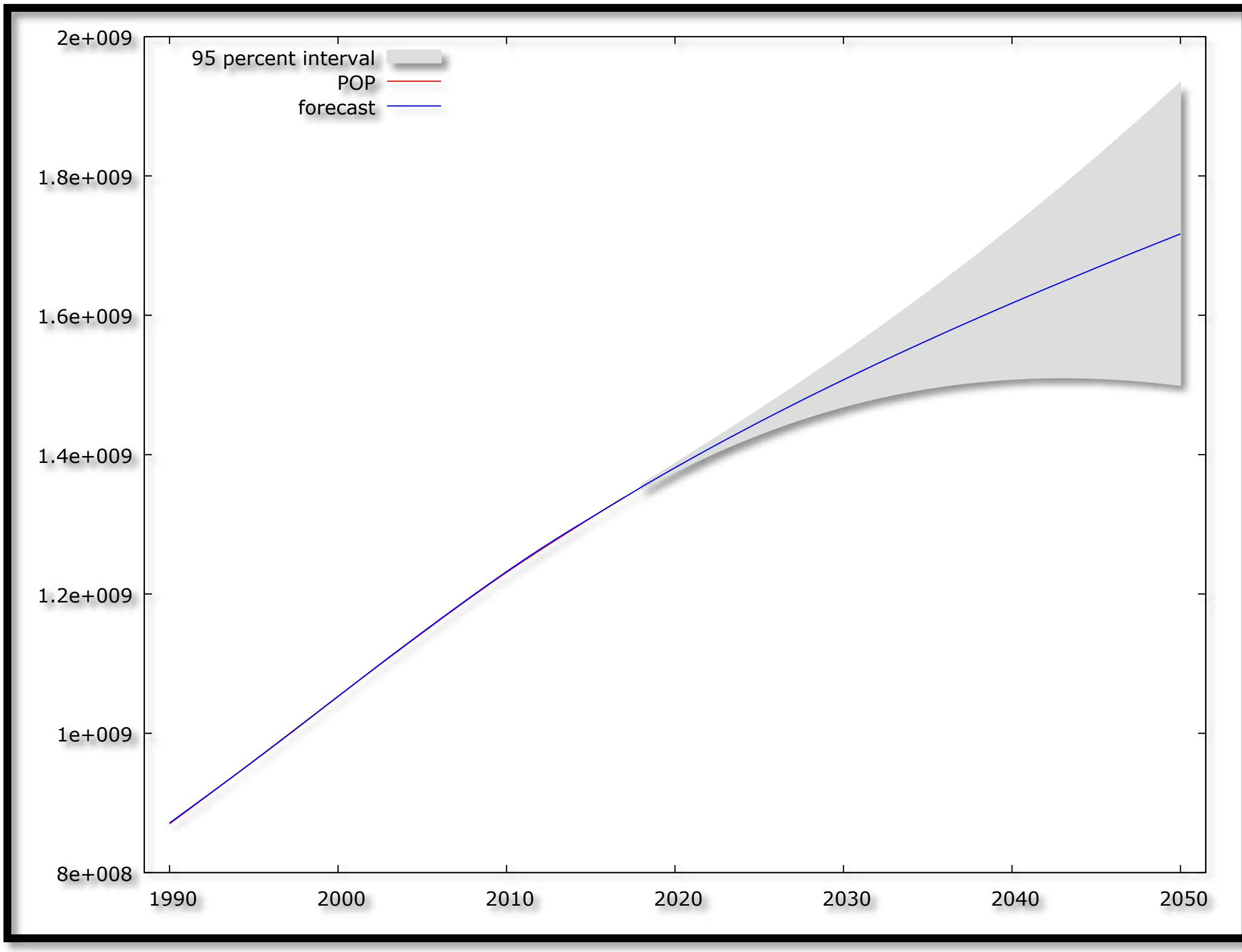
Interpretation of Results

All coefficients of the AR and MA terms are significant at 1% level of significance, indicating the equal relevance of previous period population levels as well as previous period shocks in explaining current and future population levels in India.

Forecast Graph

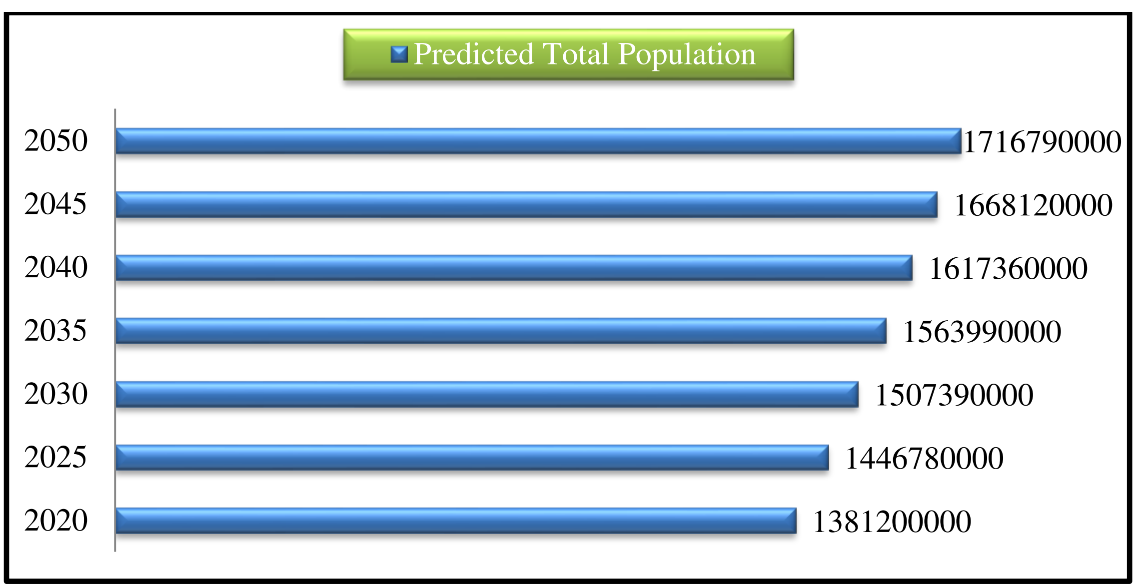
Figure 6

[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 of 32 years, that is; 2018 – 2050) and 7, clearly indicate that India population is set to continue rising sharply, at least for the next 3 decades; unless and until stern population control measures are put in place. With a 95% confidence interval of 1 499 420 000 to 1 934 170 000 and a projected total population of 1 716 790 000 by 2050, our model is consistent with the population projections by the UN (2015) which forecasted that India's population will be approximately 1 705 333 000 by 2050.

Policy Implications

i. There is need, for the Indian government; to put in place and enforce family planning policies and practices.
ii. The government of India should continue to promote the smaller family size norm.
iii. Sex education should be consistently delivered in order to control fertility in India.

CONCLUSION

The ARIMA (1, 2, 3) model is an appropriate model to forecast the population of India for the next 3 decades. Population growth is a real threat to the future of India especially considering the fact that India is currently experiencing excessively high levels of unemployment and poverty is still widespread. The findings of this research are particulary important for the government of India, especially in terms of future planning.

REFERENCES

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

[2] Asteriou, D. & Hall, S. G. (2007). Applied Econometrics: a modern approach, Revised Edition, *Palgrave MacMillan*, New York.

[3] 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.

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

[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] 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.

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

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

[12] 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.

[13] Population Foundation of India (2007). The future population of India: a long-range demographic view, *Population Foundation of India*.

[14] Rahul, G., Pandey, S & Singh, O. P (2007). Population projection of India using MCMC technique: a WinBUGS implementation, *Demography India*, 36 (1): 145 – 154.

[15] Smith, S. K (1987). Tests of forecast accuracy and bias for county population projections, *Journal of the American Statistical Association*, 82: 991 – 1003.

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

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

[18] 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.

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

[20] 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.

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