A Box-Jenkins ARIMA approach to the population question in Pakistan: A reliable prognosis

NYONI, THABANI

UNIVERSITY OF ZIMBABWE

25 February 2019

A BOX – JENKINS ARIMA APPROACH TO THE POPULATION QUESTION IN PAKISTAN: A RELIABLE PROGNOSIS

Nyoni, Thabani

Department of Economics

University of Zimbabwe

Harare, Zimbabwe

Email: nyonithabani35@gmail.com

Abstract

Employing annual time series data on total population in Pakistan from 1960 to 2017, we model and forecast total population over the next 3 decades using the Box – Jenkins ARIMA technique. Based on the minimum AIC and Theil's U, the study presents the ARIMA (3, 2, 1) model. The diagnostic tests indicate that the presented model is stable. The results of the study reveal that total population in Pakistan will continue to sharply rise within the next three decades, for up to approximately 324 million people by 2050. In order to address the threats posed by such a population explosion, 3 policy recommendations have been put forward.

Key Words: Forecasting, Pakistan, 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). Nowadays, the major issue of the world is overpopulation especially of the developing countries (Zakria & Muhammad, 2009). 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).

Furthermore, the crime rate among the societies also rises due to heavy pressure of the population (Zakria & Muhammad, 2009). In Pakistan, just like in any other part of the world, population modeling and forecasting is invaluable for policy dialogue, especially given the fact that the sharp rising of population during the past decades has threatened the development efforts in Pakistan. This study endeavors to model and forecast population of Pakistan using the Box-Jenkins ARIMA technique.

LITERATURE REVIEW

Theoretical Literature Review

The Malthus' population theory generally uncovers the effect of spiraling population on economic growth, of which Malthus (1798), later on supported by Solow (1956), reiterates that population growth is a threat to economic growth and development. While Solow's propositions were basically consistent with the basic Malthusian framework, he rather focused on the term "population growth rate" unlike Malthus who preferred the term "population level". As time went on Solow and Malthus faced serious criticism, mainly from Ahlburg (1998) and Becker *et al* (1999) who strongly argued that population growth was actually good and strongly refuted the Malthusian population explanation. Ahlburg's arguments were based on the "technology-pushed" and "demand-pulled" dynamics while Becker and his team concentrated on "high labor – a source of real wealth". This paper will let us know where Pakistan is going with regards to population dynamics.

Empirical Literature Review

In a well known local study, Zakria & Muhammad (2009) forecasted population using Box-Jenkins ARIMA models, and relied on a data set ranging from 1951 to 2007; and found out that the ARIMA (1, 2, 0) model was the optimal model in Pakistan. Haque *et al* (2012), in yet another Asian study, closer to home; analyzed Bangladesh population projections using the Logistic Population model with a data set ranging from 1991 to 2006 and found out that the Logistic Population model has the best fit for population growth in Bangladesh. In Africa, 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 finalized that the best model for modeling and forecasting population in Ethiopia was the ARIMA (2, 1, 2) model. In the case of Pakistan, just like Zakria & Muhammad (2009); the paper will adopt the Box-Jenkins ARIMA methodology 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:

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

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

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

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

and

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

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).

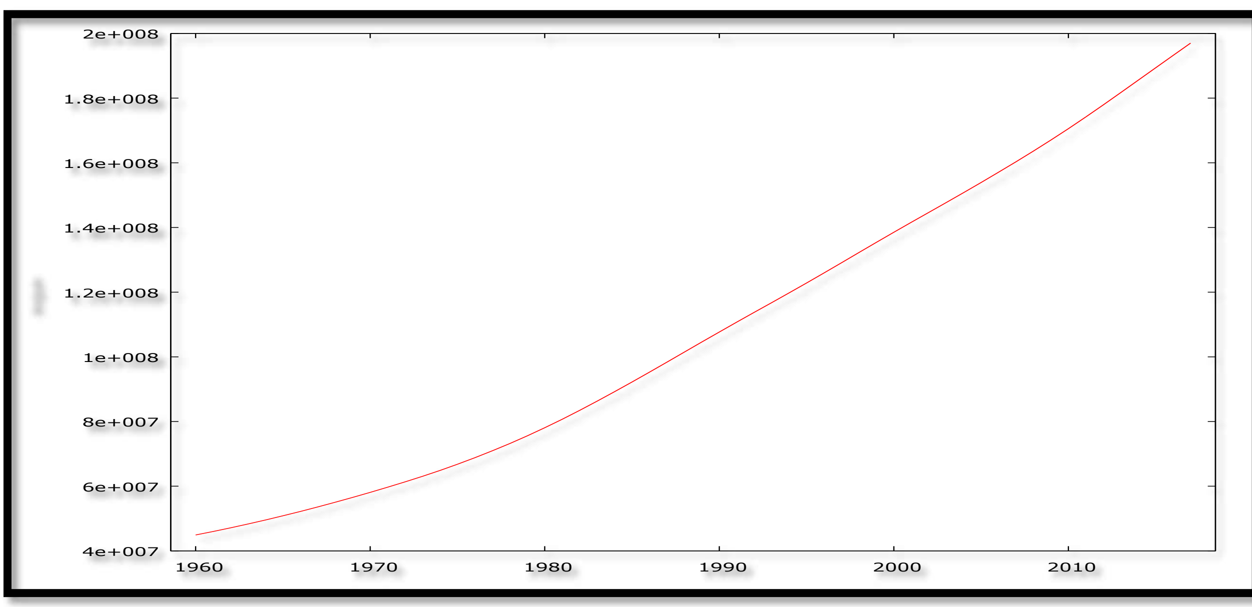
Data Collection

This research work is based on 58 observations of annual total population (POP, referred to as PPAK in the mathematical formulation above) in Pakistan. All the data was gathered from the World Bank, which is a reliable and credible source of macroeconomic data.

Diagnostic Tests & Model Evaluation

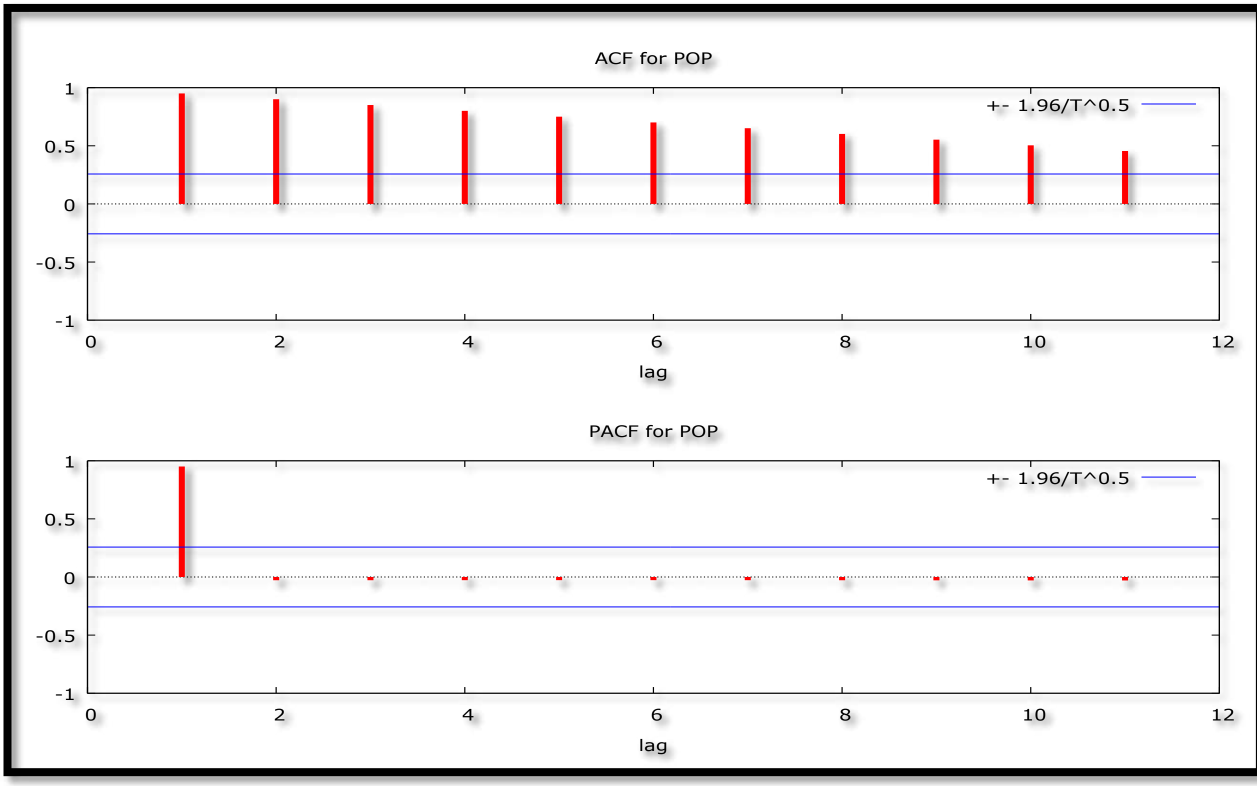
Stationarity Tests: Graphical Analysis

Figure 1



The Correlogram in Levels

Figure 2



The ADF Test

Table 1: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	1.401319	0.9988	-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	-3.206813	0.0942	-4.140858	@1%	Not stationary
			-3.496960	@5%	Not stationary
			-3.177579	@10%	Stationary

Table 3: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	0.579605	0.8384	-2.609324	@1%	Not stationary
			-1.947119	@5%	Not stationary
			-1.612867	@10%	Not stationary

The Correlogram (at 1st Differences)

Figure 3

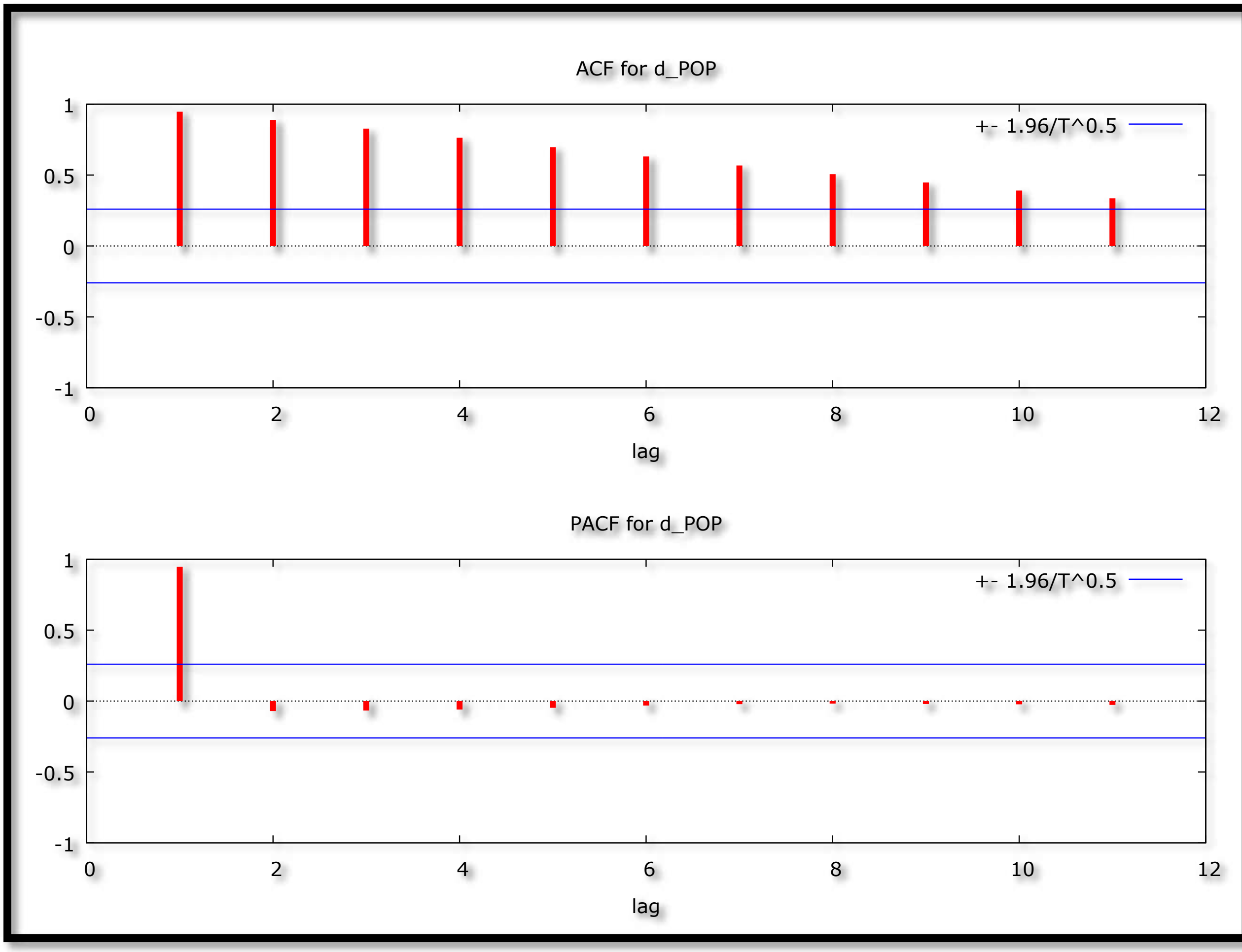


Table 4: 1st Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.681348	0.4347	-3.560019	@1%	Stationary
			-2.917256	@5%	Stationary
			-2.596689	@10%	Stationary

Table 5: 1st Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-2.358025	0.3966	-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	0.583166	0.8392	-2.609324	@1%	Not stationary

		-1.947119	@5%	Not stationary
		-1.612867	@10%	Not stationary

The Correlogram in (2nd Differences)

Figure 4

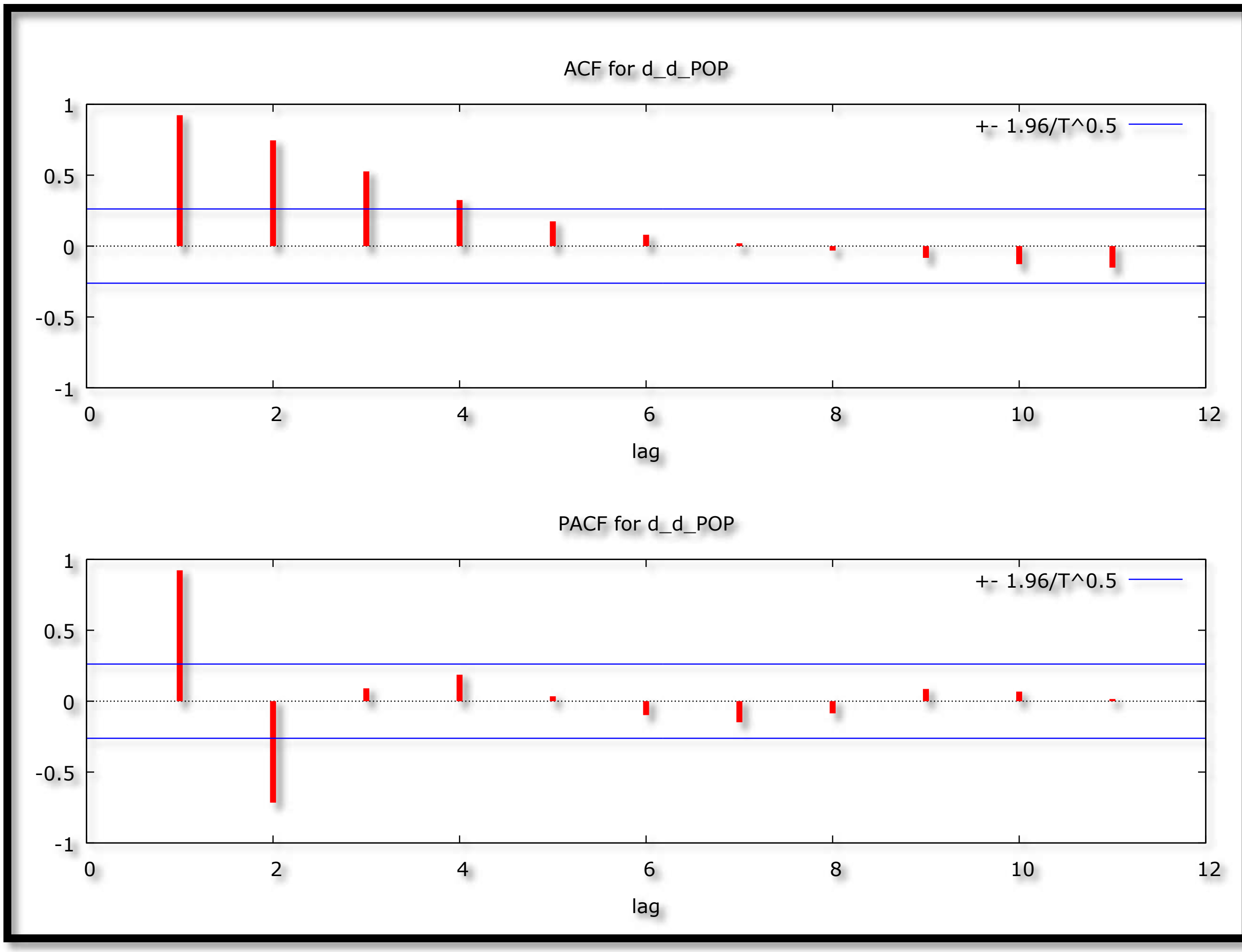


Table 7: 2nd Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.806613	0.3734	-3.560019	@1%	Not stationary
			-2.917650	@5%	Not stationary
			-2.596689	@10%	Not stationary

Table 8: 2nd Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-2.016180	0.5792	-4.140858	@1%	Not stationary
			-3.496960	@5%	Not stationary
			-3.177579	@10%	Not stationary

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

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.289121	0.1797	-2.609324	@1%	Not stationary
			-1.947119	@5%	Not stationary
			-1.612867	@10%	Not stationary

Figures 1 – 4 and tables 1 – 9 indicate that the Pakistan POP series is not stationary in levels, in first differences and in second differences. This is characteristic of sharply upwards trending time series and is consistent with the observation that total population in Pakistan is spiraling. However, for analytical purposes of this study, we assume that the Pakistan POP series is I (2).

Evaluation of ARIMA models (without a constant)

Table 10

Model	AIC	U	ME	MAE	RMSE	MAPE
ARIMA (1, 2, 1)	1204.061	0.003452	1611.8	8872.8	12785	0.0092371
ARIMA (1, 2, 0)	1254.124	0.005566	1672	14032	18229	0.013691
ARIMA (0, 2, 1)	1337.749	0.017024	25429	30737	36085	0.036519
ARIMA (2, 2, 0)	1194.464	0.0033739	3554.1	8513.2	12122	0.0092104
ARIMA (2, 2, 1)	1171.180	0.002637	2715	6892.7	10676	0.007653
ARIMA (3, 2, 1)	**1163.852**	**0.0024088**	1934	6309.7	10254	0.0071341
ARIMA (3, 2, 0)	1164.603	0.0025007	1821.6	6284.4	10374	0.0071424

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 study will consider the AIC and the Theil's U in order to choose the best model. Therefore, the ARIMA (3, 2, 1) model is selected.

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
u_t	-7.181837	0.0000	-3.562669	@1%	Stationary
			-2.918778	@5%	Stationary
			-2.597285	@10%	Stationary

Table 12: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
u_t	-7.175537	0.0000	-4.144584	@1%	Stationary
			-3.498692	@5%	Stationary
			-3.178578	@10%	Stationary

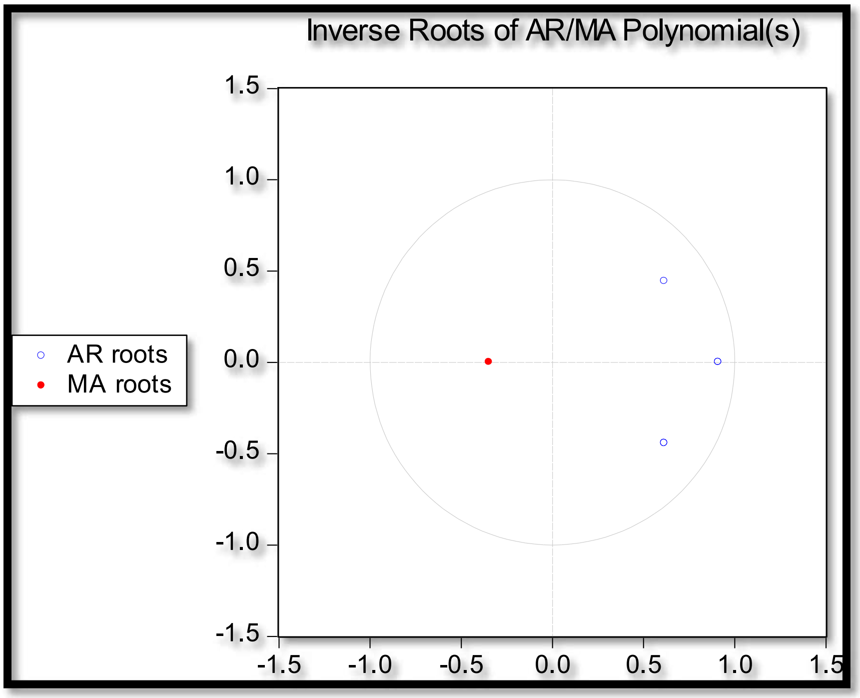
Table 13: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
u_t	-7.134249	0.0000	-2.610192	@1%	Stationary
			-1.947248	@5%	Stationary
			-1.612797	@10%	Stationary

Tables 11, 12 and 13 show 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 indicates that the selected ARIMA (3, 2, 1) model is stable.

FINDINGS

Descriptive Statistics

Table 14

Description	Statistic
Mean	108360000
Median	103030000
Minimum	44908000
Maximum	197020000
Standard deviation	46706000
Skewness	0.30091
Excess kurtosis	-1.1996

As shown above, the mean is positive, i.e. 108360000. The wide gap between the minimum (i.e 44908000) and the maximum (i.e. 197020000) is consistent with the reality that the Pakistan POP series is sharply trending upwards. This simply means that Pakistan population is spiraling and apparently posing a threat to the economy. The skewness is 0.30091 and the most striking characteristic is that it is positive, indicating that the POP series is positively skewed and non-symmetric. Excess kurtosis is -1.1996; showing that the POP series is not normally distributed.

Results Presentation[1]

Table 15

ARIMA (3, 2, 1) Model:

$$\Delta^2 POP_{t-1} = 2.1169\Delta^2 POP_{t-1} - 1.6516\Delta^2 POP_{t-2} + 0.5035\Delta^2 POP_{t-3} + 0.34\mu_{t-1} \ldots\ldots\ldots [5]$$

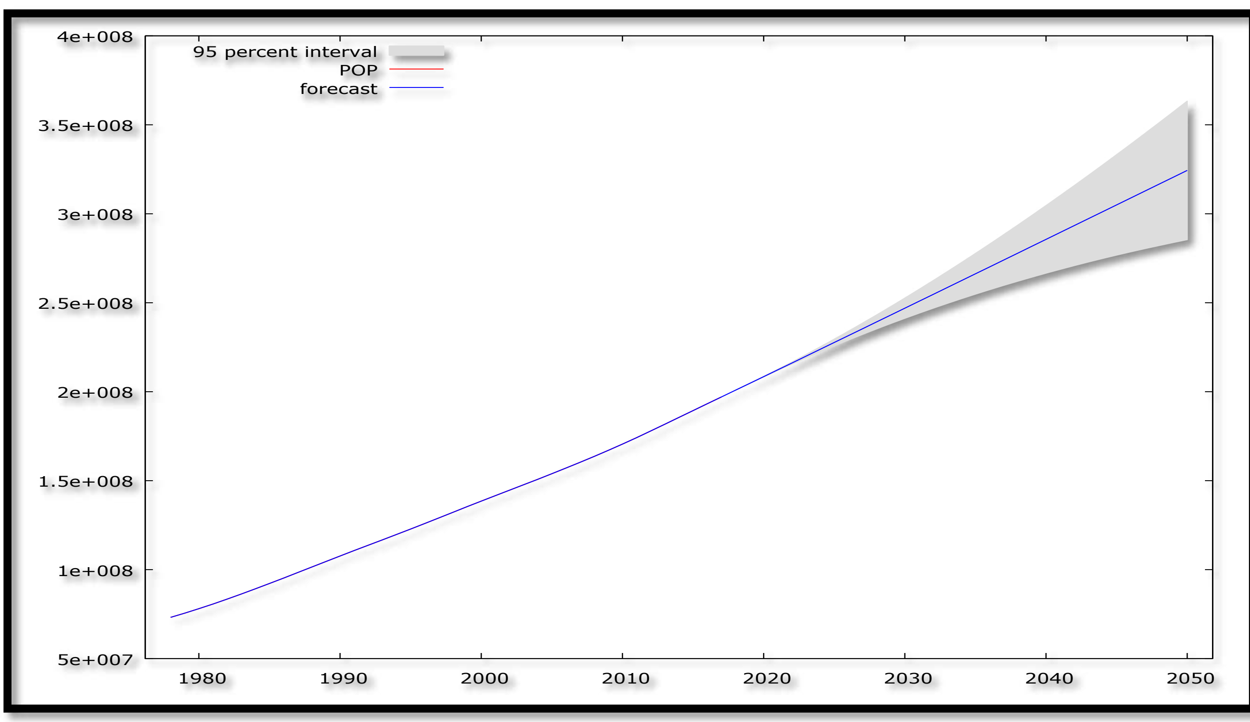
P: (0.0000) (0.0000) (0.0012) (0.0661)

S. E: (0.1658) (0.3026) (0.1551) (0.1856)

Variable	Coefficient	Standard Error	z	p-value
AR (1)	2.11694	0.165757	12.77	0.0000***
AR (2)	-1.65155	0.302643	-5.457	0.0000***
AR (3)	0.503532	0.155075	3.247	0.0012***
MA (1)	0.341026	0.185597	1.837	0.0661*

Forecast Graph

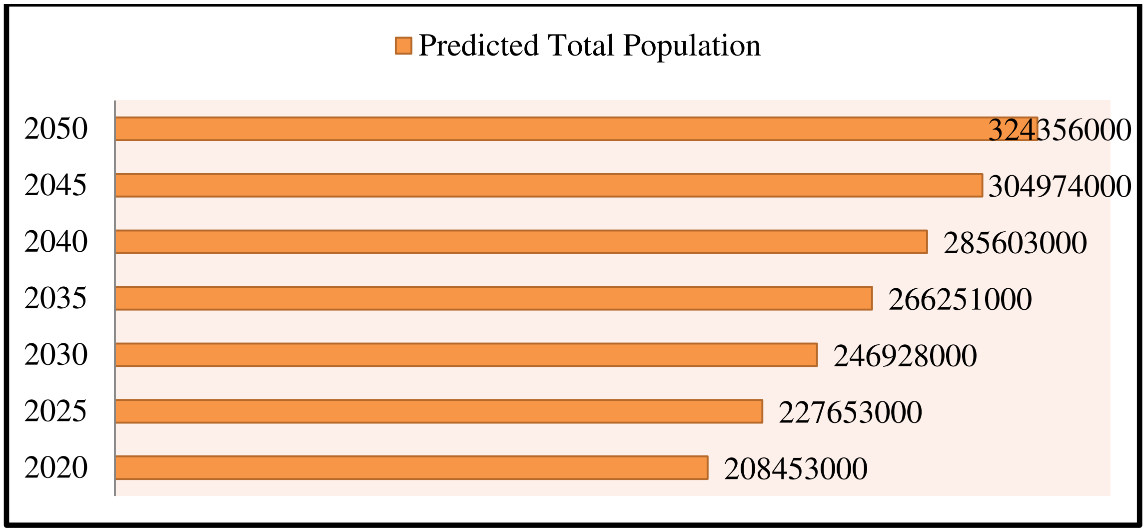
Figure 6



[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 shows that Pakistan population is indeed set to continue rising sharply, at least for the next 3 decades. With a 95% confidence interval of 285407000 to 363306000 and a projected total population of 324356000 by 2050, the chosen ARIMA (3, 2, 1) model is consistent with the population projections by the UN (2015) which forecasted that Pakistan's population will be approximately 309640000 by 2050.

Policy Implications

i. The government of Pakistan ought to enforce consistent family planning practices.
ii. The government of Pakistan should promote the smaller family size norm.
iii. The government of Pakistan should engage in sex education in order to control fertility in Pakistan.

CONCLUSION

The ARIMA (3, 2, 1) model is not only acceptable but also the most parsimonious model to forecast the population of Pakistan for the next 3 decades. The model predicts that by 2050, Pakistan's population would be approximately, 324 million. This clearly proves that population growth is a real threat to the future of Pakistan especially considering the fact that Pakistan is currently experiencing high levels of unemployment and poverty & crimes are still rampant. These findings are essential for the government of Pakistan, especially when it comes to planning for the future.

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

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

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

[16] 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.

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

[18] 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.

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