Population dynamics in Gambia: an ARIMA approach

NYONI, THABANI and MUTONGI, CHIPO and MUNYARADZI, NYONI

University of Zimbabwe, Midlands State University, Ministry of Primary and Secondary Education

7 May 2019

POPULATION DYNAMICS IN GAMBIA: AN ARIMA APPROACH

Nyoni, Thabani
Department of Economics
University of Zimbabwe
Harare, Zimbabwe
Email: nyonithabani35@gmail.com

Dr. **Mutongi**, Chipo (Ph.D)
Dept. Bus. Mgt.
Midlands State University
Gweru, Zimbabwe
mutongic@gmail.com

Nyoni, Munyaradzi
Min. Sec. & Pri. Edu.
Gov. Zim.

Abstract

Employing annual time series data on total population in Gambia from 1960 to 2017, I 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 Gambia annual total population is I (2). Based on the AIC, the study presents the ARIMA (3, 2, 1) model and our diagnostic tests also indicate that the presented model is stable. The results of the study reveal that total population in Gambia will continue to gradually rise in the next three decades. In order to take advantage of the expected increase in total population in Gambia, 4 policy recommendations have been proposed for consideration by the Gambian policy makers.

Key Words: Forecasting, Gambia, 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).

The Gambia has a steady population growth rate of about 3 per cent and a total population of around 2 million inhabitants. The population of the country is young and more than 50 percent live in urban areas. Poverty is a major problem in the Gambia and manifests itself in its low ranking in the 2015 human development index, where it is ranked 175 out of 188 countries (UN, 2016). The Gambia's young population has the potential to provide labour to all sectors and could ultimately lead to equitable growth (Ministry of Lands and Regional Government, 2015). In Gambia, just like in any other part of the world, population modeling and forecasting is critical for policy dialogue. This study endeavors to model and forecast population of the Gambia using the Box-Jenkins ARIMA technique.

REVIEW OF PREVIOUS STUDIES

Table 1

Author(s) / Year	Country	Period	Methodology	Major Findings
Zakria & Muhammad (2009)	Pakistan	1951 – 2007	Box-Jenkins ARIMA Model	ARIMA (1, 2, 0) is the optimal model
Haque *et al* (2012)	Bangladesh	1991 – 2006	Logistic Population Model (LPM)	The LPM has the best fit for population growth in Bangladesh
Beg & Islam (2016)	Bangladesh	1965 – 2003	Autoregressive Time Trend Model	Downward population growth for Bangladesh for the extended period up to 2043
Ayele & Zewdie (2017)	Ethiopia	1961 – 2009	Box-Jenkins ARIMA Model	ARIMA (2, 1, 2) Model is the optimal model

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:

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

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

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

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 Gambia, i.e. 1960 – 2017, gathered from the World Bank online database.

Diagnostic Tests & Model Evaluation

Stationarity Tests: Graphical Analysis

Figure 1

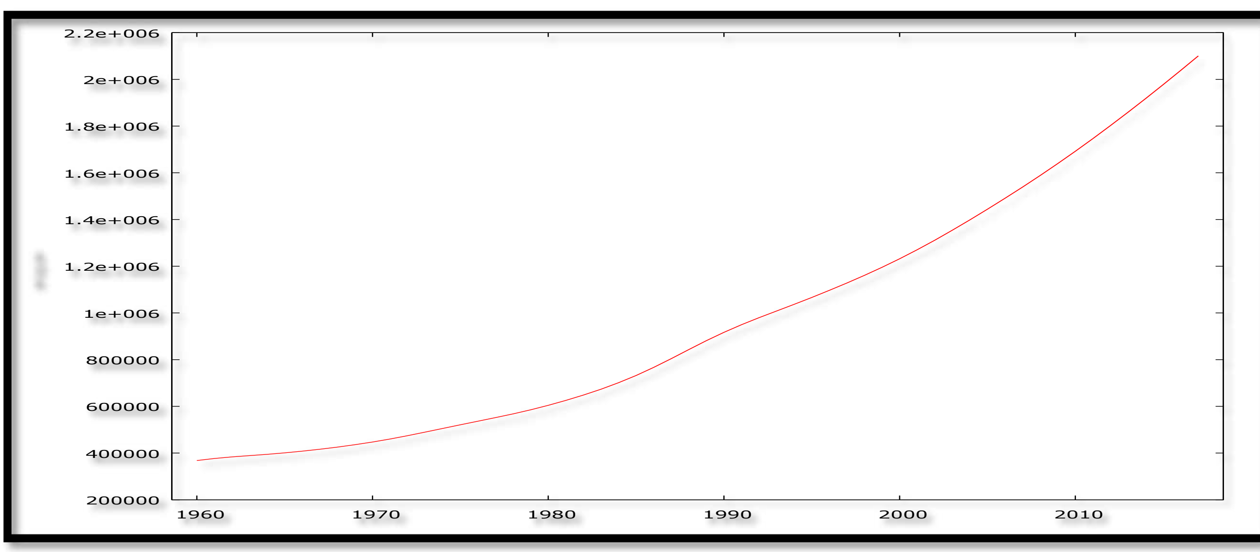
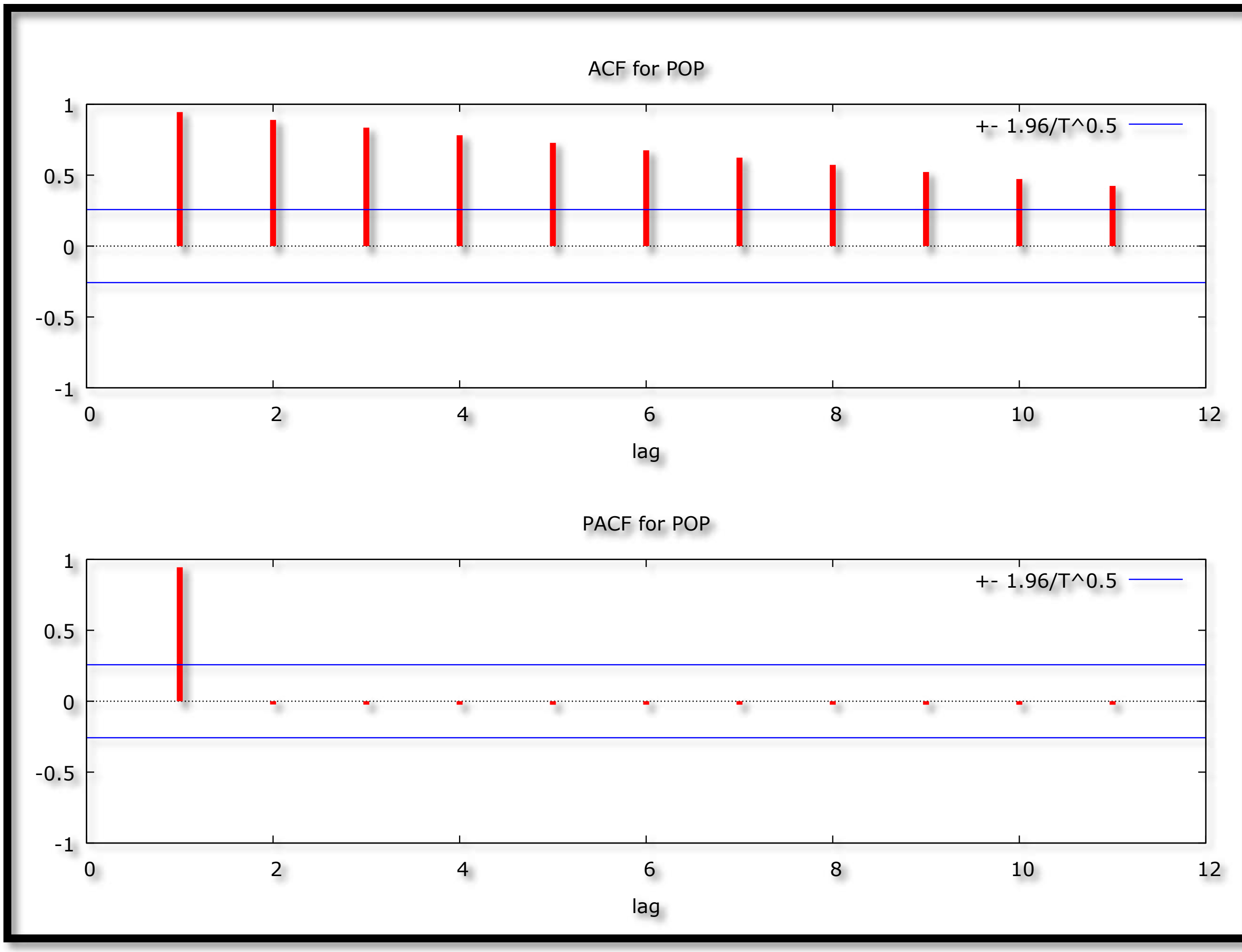


Figure 1 above indicates that the Gambia POP variable is not stationary since it is trending upwards over the period 1960 – 2017. This basically points to the notion that the mean and varience of POP is changing over time.

The Correlogram in Levels

Figure 2



The ADF Test

Table 2: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	2.575586	1.0000	-3.560019	@1%	Not stationary
			-2.917650	@5%	Not stationary
			-2.596689	@10%	Not stationary

Table 3: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	1.198963	0.9999	-4.140858	@1%	Not stationary

			-3.496960	@5%	Not stationary
			-3.177579	@10%	Not stationary

Table 4: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	2.766260	0.9983	-2.609324	@1%	Not stationary
			-1.947119	@5%	Not stationary
			-1.612867	@10%	Not stationary

The Correlogram (at 1st Differences)

Figure 3

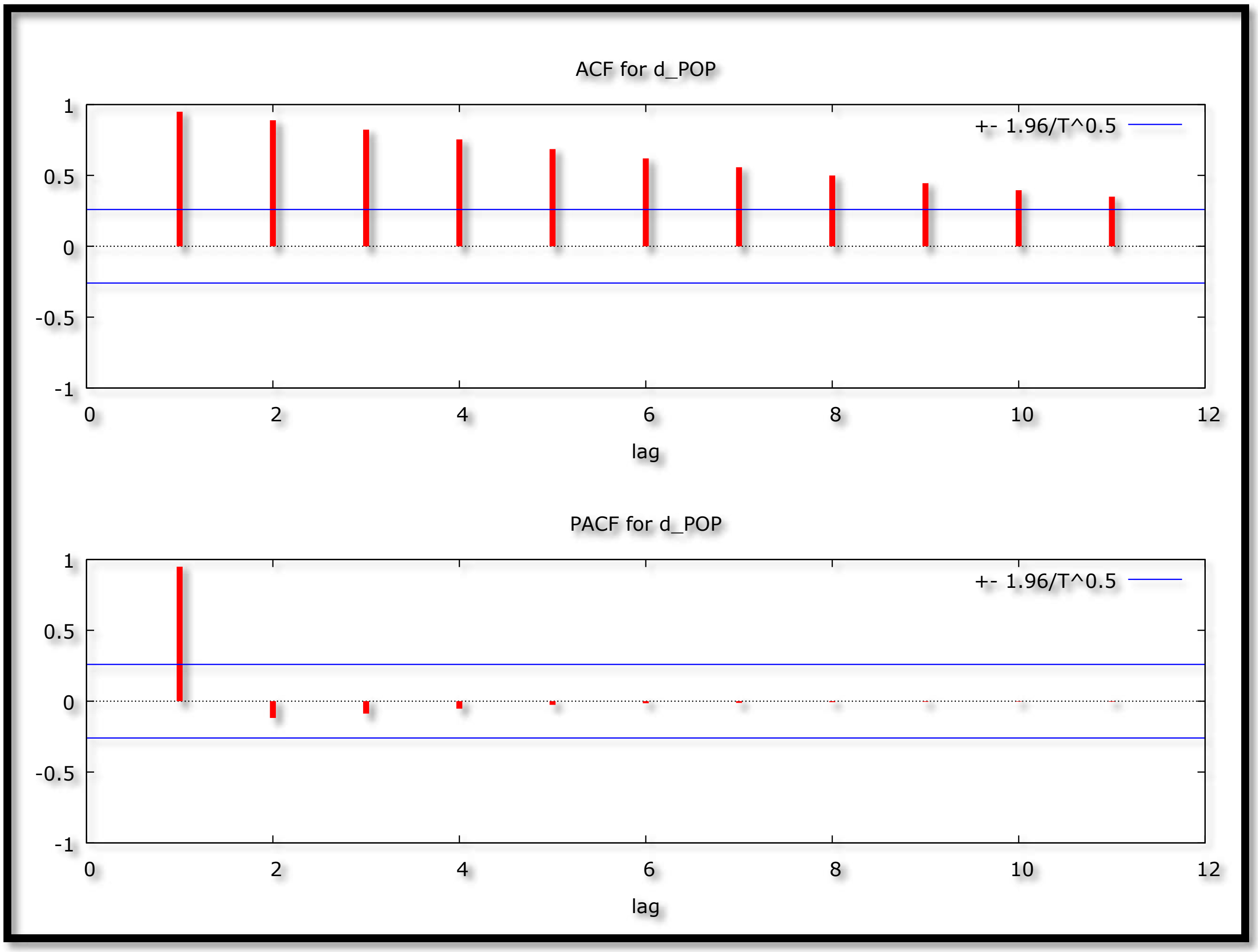


Table 5: 1st Difference-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	1.291012	0.9983	-3.571310	@1%	Not stationary
			-2.922449	@5%	Not stationary
			-2.599224	@10%	Not stationary

Table 6: 1st Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-3.299515	0.0791	-4.170583	@1%	Not stationary
			-3.510740	@5%	Not stationary
			-3.185512	@10%	Stationary

Table 7: 1st Difference-without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	1.525858	0.9672	-2.609324	@1%	Not stationary
			-1.947119	@5%	Not stationary
			-1.612687	@10%	Not stationary

Figures above, i.e. 2 and 3 and tables above, i.e. 2 to 7 indicate that the Gambia POP series is not stationary in levels and in first differences.

The Correlogram in (2nd Differences)

Figure 4

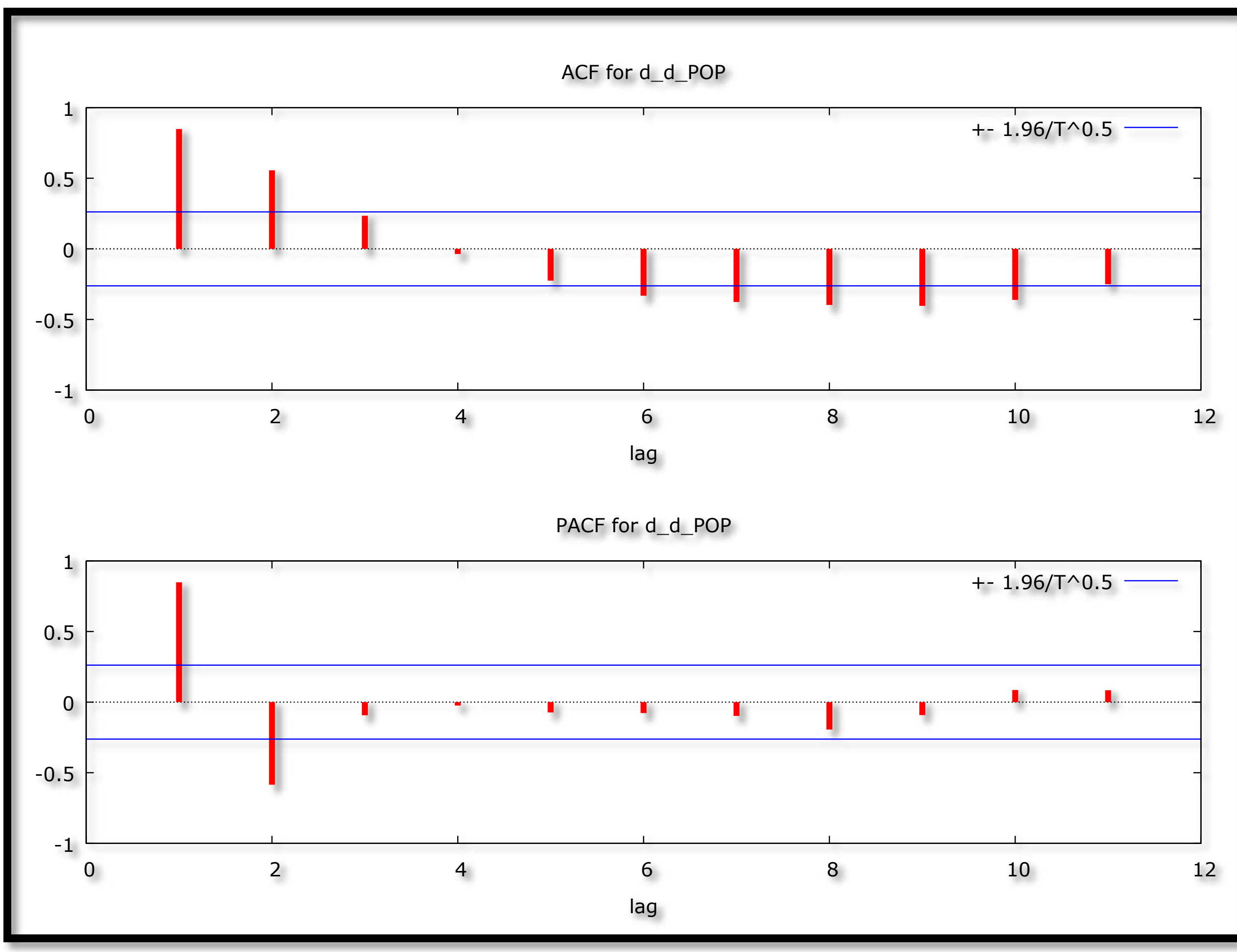


Table 8: 2nd Difference-intercept

Variable	ADF Statistic	Probability	Critical Values	Conclusion

POP	-4.043850	0.0027	-3.571310	@1%	Stationary
			-2.922449	@5%	Stationary
			-2.599224	@10%	Stationary

Table 9: 2nd Difference-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-4.394714	0.0052	-4.156734	@1%	Stationary
			-3.504330	@5%	Stationary
			-3.181826	@10%	Stationary

Table 10: 2nd Difference-without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
POP	-1.725385	0.0800	-2.609324	@1%	Not stationary
			-1.947119	@5%	Not stationary
			-1.612867	@10%	Stationary

Figure 4 shows that most of the autocorrelation coefficients are around zero pointing to the notion that the Gambia POP series could be stationary in second differences; only at the first and second lags are the autocorrelations coefficients quite high. Tables 8 and 9 illustrate that the Gambia POP series is stationary in second differences. Table 10 indicates the POP series is only stationary at 10% level of significance.

Evaluation of ARIMA models (without a constant)

Table 11

Model	AIC	U	ME	MAE	RMSE	MAPE
ARIMA (1, 2, 1)	835.3101	0.01856	46.162	306.09	471.23	0.047322
ARIMA (1, 2, 0)	884.2296	0.028127	88.321	457.04	664.74	0.068489
ARIMA (0, 2, 1)	923.0804	0.035217	483.06	775.87	892.8	0.098278
ARIMA (2, 2, 1)	803.0702	0.014138	59.479	264.17	393.92	0.040691
ARIMA (3, 2, 1)	**800.4067**	0.013906	40.986	249.33	386.17	0.039351
ARIMA (4, 2, 1)	802.4067	0.013905	41.001	249.34	386.17	0.039351
ARIMA (5, 2, 1)	803.8917	0.013873	37.257	247.59	385.31	0.039162
ARIMA (6, 2, 1)	804.2669	0.013633	43.902	241.24	382.64	0.038188
ARIMA (2, 2, 0)	822.2132	0.016506	93.114	315.98	437	0.047321
ARIMA (3, 2, 0)	801.3062	0.014415	37.656	244.97	391.02	0.03942
ARIMA (4, 2, 0)	801.3495	0.014051	42.877	250.24	387.74	0.039594
ARIMA (5, 2, 0)	802.1704	0.013927	37.069	248.3	385.78	0.039312
ARIMA (6, 2, 0)	802.8298	0.013686	41.292	243.36	383.54	0.038456

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 in order to choose the optimal model for forecasting total population in Gambia. Therefore, for forecasting total population in Gambia, the ARIMA (3, 2, 1) model is preferred.

Residual & Stability Tests

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

Table 12: Levels-intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
V_t	-4.505414	0.0008	-3.592462	@1%	Stationary
			-2.931404	@5%	Stationary
			-2.603944	@10%	Stationary

Table 13: Levels-trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
V_t	-4.420373	0.0054	-4.186481	@1%	Stationary
			-3.518090	@5%	Stationary
			-3.189732	@10%	Stationary

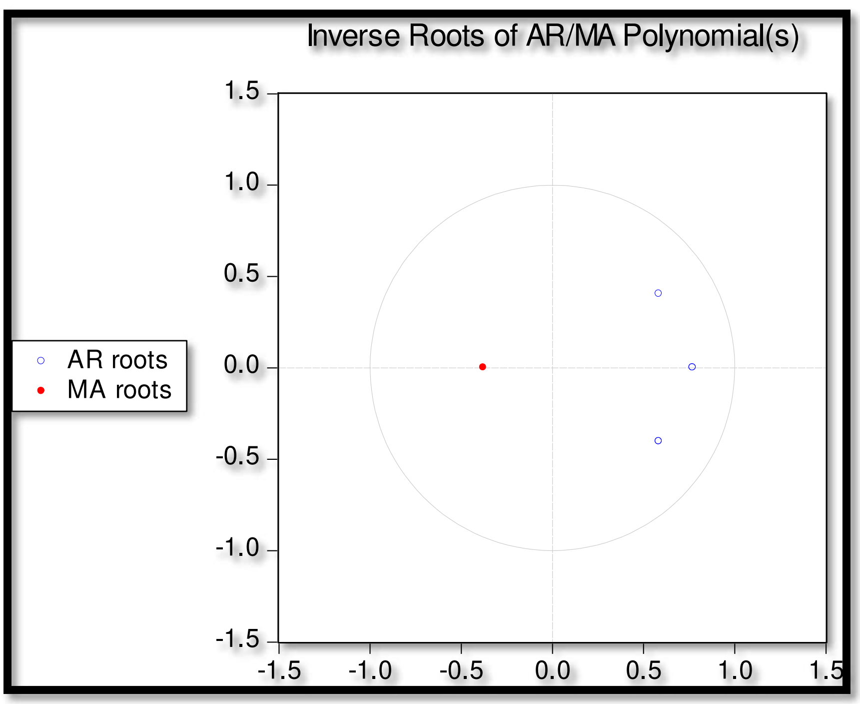
Table 14: without intercept and trend & intercept

Variable	ADF Statistic	Probability	Critical Values		Conclusion
V_t	-7.160562	0.0000	-2.610192	@1%	Stationary
			-1.947248	@5%	Stationary
			-1.612797	@10%	Stationary

Tables 11, 12 and 13 demonstrate that the residuals of the selected optimal model, 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 simply proves that the chosen ARIMA (3, 2, 1) model is stable.

FINDINGS

Descriptive Statistics

Table 15

Description	Statistic
Mean	971720
Median	862090
Minimum	367930
Maximum	2100600
Standard deviation	519200
Skewness	0.60261
Excess kurtosis	-0.85562

As shown above, the mean is positive, i.e. 971720. The wide gap between the minimum (i.e 367930) and the maximum (i.e. 2100600) is consistent with the observation that the Gambian POP series is gradually trending upwards over the period under study. The skewness is 0.60261 and the most essential characteristic is that it is positive, indicating that the Gambian POP series is positively skewed and non-symmetric. Excess kurtosis is -0.85562; showing that the Gambian POP series is not normally distributed.

Results Presentation[1]

Table 16

ARIMA (3, 2, 1) Model:

$$\Delta^2 POP_{t-1} = 1.9496\Delta^2 POP_{t-1} - 1.4256\Delta^2 POP_{t-2} + 0.3919\Delta^2 POP_{t-3} + 0.3791\mu_{t-1} \; \ldots . [5]$$

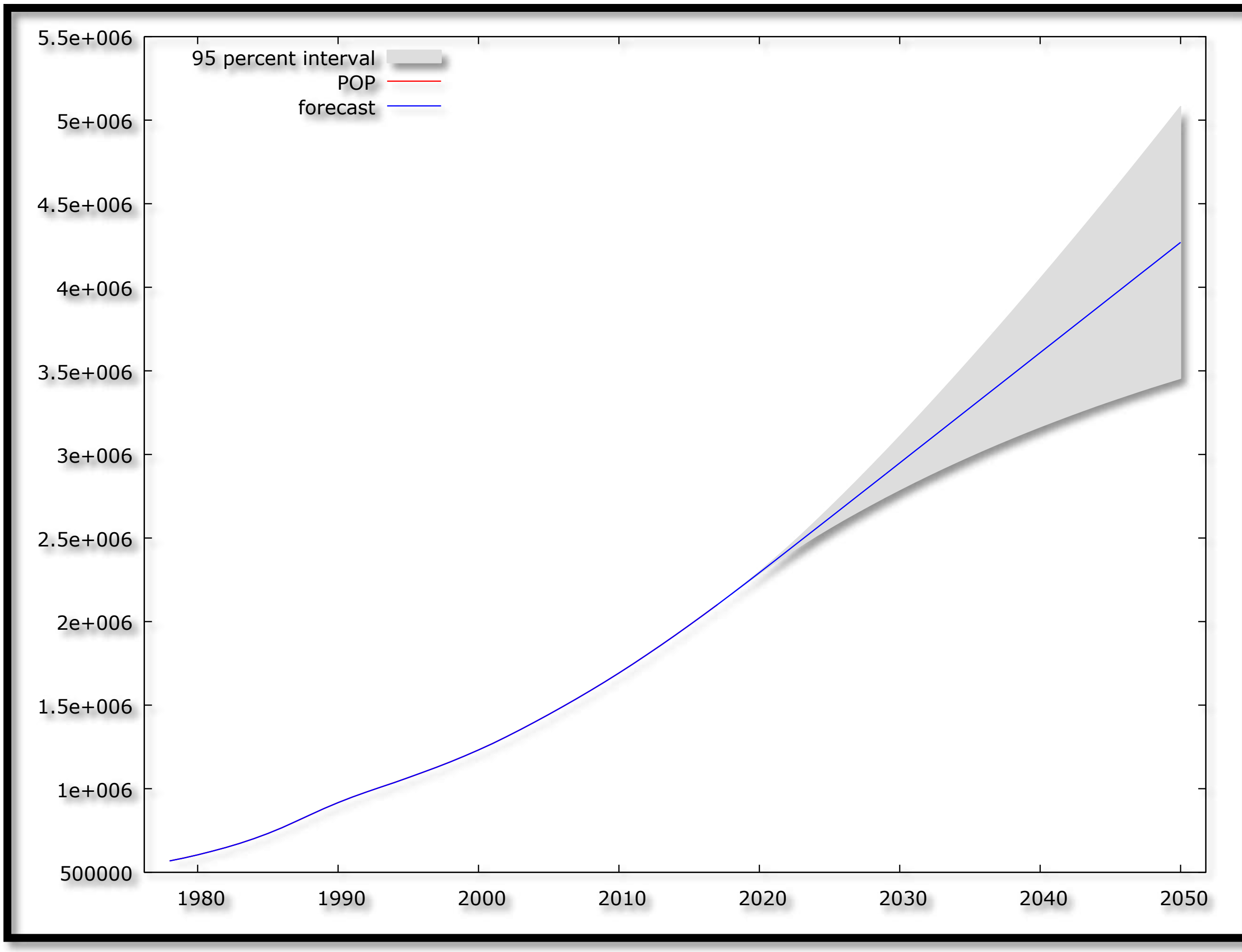
P:	(0000)	(0.0000)	(0.0268)	(0.0525)
S. E:	(0.1912)	(0.3341)	(0.177)	(0.1955)

Variable	Coefficient	Standard Error	z	p-value
AR (1)	1.94964	0.191208	10.2	0.0000
AR (2)	-1.42564	0.334107	-4.267	0.0000
AR (3)	0.391932	0.176969	2.215	0.0268
MA (1)	0.379139	0.195537	1.939	0.0525

Forecast Graph

Figure 6

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



Predicted Total Population

Table 17

Year	Actual	Prediction	Std. Error	95% Confidence Interval
2000	1231844.00	1231690.78		
2001	1270495.00	1270213.99		
2002	1311349.00	1311483.93		
2003	1354194.00	1354035.49		
2004	1398573.00	1398712.99		
2005	1444204.00	1443914.66		
2006	1491021.00	1490979.06		
2007	1539116.00	1538982.50		

2008	1588572.00	1588553.15		
2009	1639560.00	1639331.48		
2010	1692149.00	1692182.09		
2011	1746363.00	1746196.18		
2012	1802125.00	1802126.41		
2013	1859324.00	1859215.33		
2014	1917852.00	1917795.84		
2015	1977590.00	1977550.44		
2016	2038501.00	2038370.60		
2017	2100568.00	2100544.23		
2018		2163699.77	266.801	2163176.85 - 2164222.69
2019		2227719.14	1185.340	2225395.92 - 2230042.37
2020		2292404.16	3108.880	2286310.86 - 2298497.45
2021		2357538.81	6258.808	2345271.78 - 2369805.85
2022		2422949.06	10704.871	2401967.90 - 2443930.22
2023		2488516.44	16397.134	2456378.65 - 2520654.24
2024		2554173.55	23218.160	2508666.79 - 2599680.31
2025		2619889.55	31031.231	2559069.46 - 2680709.65
2026		2685654.07	39712.635	2607818.74 - 2763489.41
2027		2751464.38	49166.360	2655100.09 - 2847828.68
2028		2817317.88	59325.394	2701042.24 - 2933593.51
2029		2883209.30	70145.601	2745726.45 - 3020692.16
2030		2949131.07	81597.322	2789203.25 - 3109058.88
2031		3015074.82	93658.016	2831508.48 - 3198641.16
2032		3081033.07	106307.374	2872674.44 - 3289391.69
2033		3147000.11	119524.989	2912735.44 - 3381264.79
2034		3212972.27	133289.914	2951728.83 - 3474215.70

2035	3278947.51	147581.224	2989693.63 - 3568201.40
2036	3344924.96	162378.855	3026668.25 - 3663181.67
2037	3410904.29	177664.263	3062688.73 - 3759119.85
2038	3476885.36	193420.738	3097787.69 - 3855983.04
2039	3542868.02	209633.389	3131994.13 - 3953741.91
2040	3608852.00	226288.932	3165333.84 - 4052370.16
2041	3674837.01	243375.394	3197830.00 - 4151844.02
2042	3740822.74	260881.832	3229503.75 - 4252141.74
2043	3806808.94	278798.121	3260374.66 - 4353243.22
2044	3872795.42	297114.803	3290461.11 - 4455129.74
2045	3938782.08	315823.009	3319780.35 - 4557783.80
2046	4004768.84	334914.402	3348348.67 - 4661189.00
2047	4070755.68	354381.142	3376181.41 - 4765329.96
2048	4136742.60	374215.860	3403292.99 - 4870192.21
2049	4202729.58	394411.620	3429697.01 - 4975762.15
2050	4268716.62	414961.882	3455406.27 - 5082026.96

Figure 6 (with a forecast range from 2018 – 2050) and table 17, clearly show that Gambia population is set to continue rising gradually, in the next 3 decades. With a 95% confidence interval of 3455406 to 5082027 and a projected total population of 4268717 by 2050, the chosen ARIMA (3, 2, 1) model is consistent with the population projections by the UN (2015) which forecasted that Gambia's population will be approximately 4981000 by 2050. According to the Gambia Bureau of Statistics (2013), the steady increase in population size has policy implications for all sectors particulary the education, health, housing and agriculture sectors.

Policy Implications

1) The Gambian government should invest more in infrastructural development in order to cater for the expected increase in total population.
2) The projected increase in total population justifies the need for more and bigger companies to provide for the anticipated increase in demand for goods and services.
3) The anticipated increase in total population of the Gambia signifies the likely increase in the demand for land for both residential and agriculture purposes.
4) The Gambian government should take action so as to improve health service delivery in the country in order to ensure a healthier society, especially in light of such a likely increase in total population.

CONCLUSION

In the case of Gambia, the study shows that the ARIMA (3, 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, Gambia's total population would be approximately, 4.3 million people. This is a warning signal to policy makers in Gambia, especially with regards to infrastructural development, e.g schools and hospitals. These findings are essential for the Gambian government, especially when it comes to 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] 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.

[8] Ministry of Land and Regional Government (2015). Habitat III – National Report, *Republic of Gambia*.

[9] 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.

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

[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] The Gambia Bureau of Statistics (2013). The Gambia 2013, Population and Housing Census Preliminary Results, *The Gambia*, Serrekunda.

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

[16] UN (2016). The Gambia – Country Profile, *UN*, Addis Ababa office.

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