



Modeling and forecasting population growth rate in Bangladesh: Using Arima models

Md Muhaiminul Islam Selim¹, Md Ahasan Ul Haque², Liton Chandra Voumik^{3*}

¹⁻³ Lecturer, Department of Economics, Noakhali Science and Technology University, Noakhali, Bangladesh

Abstract

Background/Objectives: The tremendous increase in population is one of the most crucial problems of Bangladesh. Currently, one of the most densely populated countries and this bulky size of the population is now one of the main concerns in this country. This paper aims to apply the ARIMA method to model and forecast population growth rates in Bangladesh.

Methods/ Statistical analysis: The data was collected from World Development Indicators (WDI) and it has been collected over 40 years by WDI, World Bank. We applied Phillips–Perron (PP) and Augmented Dickey-Fuller (ADF) tests to investigate the stationary character of the data. Stata and R statistical software was used to build a structure of the ARIMA method to model and forecast the population growth rates.

Findings: In this study, the population of Bangladesh from 1979 to 2018 is modeled using ARIMA (P, I, Q) methodology. The model was validated by the lowest values of AIC and BIC, less than 5% of P-values, graphical presentations of ACF and PACF plots, and corrgrams. Using these models, the numeric figure of future population growths are drawn and forecasted. The population growth rate for the next first and second decades is also forecasted using the ARIMA (1, 1, 1) model. If the decreasing trend persists, the population growth rate of Bangladesh would be approximately 0.56% in 2028 and 0.35% in 2038. Statistical outcomes illustrate that Bangladesh's population growth rate is a decreasing trend that will continue declining in the future.

Improvements/Applications: This finding will help policymakers and academicians to formulate population-related strategies and policies more precisely.

Keywords: population growth rate, Arima, time series, AIC and BIC, population forecasting

1. Introduction

The land area of Bangladesh is small but it is one of the most populated countries. About 166.46 million people live in this small country with a total land area of, 1, 47, 570 square kilometers^[1]. About 1253 people survive per square kilometer^[2]. The main reasons behind the excessive increase in the population are early marriage, polygamy, illiteracy, tropical climate, and the least amount of recreational conveniences. This overpopulation causes child labor, food shortage, crime, communication problem, health problem, deforestation, shelter problem, pollution, unemployment, and poverty. For solving these issues the population growth rate should be controlled. The public consciousness and literacy rate should be improved. Moreover, strong procedures will have to be taken against early marriage and polygamy. Furthermore, people will have to be motivated to adopt family planning policies, contraceptive uses, and recreational facilities. However, the government of Bangladesh, several private organizations, and NGOs has given much pressure on the population control by identifying it as one of the major problems. So, we should help the government in this respect.

2. How Bangladesh has been trying to reduce this population problem

In 1971, the population of Bangladesh was slightly over 6.5 crores^[3]. The figure has more than 2.5 times over the next three decades, invoking immediate family planning campaigns. The campaign, two kids are enough, boy or girl does not matter, was successful in dropping the TFR or total fertility rate 6.4 in 1971

to 2.05 today, which was 3.0 in 2001. The government plan is to reduce the rate to 2 percent by 2022. There have been no major changes in the main ingredients of the population policy of Bangladesh from the time it was announced in the 1st 5-year plan for population control was formulated. Realizing the threat of overpopulation, the government of Bangladesh has given the highest priority to population control in its development program and proposes to achieve it through 1) improved health and education, 2) reduced infant, child, and maternal mortality, 3) increasing the socioeconomic status of women, 4) changing the outlook of the rural population through personal contact and the use of mass media, 5) delivering contraceptive services to people at or near home, 6) enlisting community support for population planning, 7) involving all ministries concerned with development in population planning and control, 8) encouraging and supporting non-governmental organizations in family planning, and 9) adopting a system of incentives and legal and social measures, 10) availing easy and cheap contraceptives. As the government has taken many initiatives to control the population, the family planning program becomes successful. According to the economic survey of Bangladesh now population growth rate is 1.36 percent which was 1.57 percent in 2001. Contraceptive prevalence rates (CPRs) continued to rise steadily throughout the 1990s, from 44.6% (1993/1994) to a peak of 58.1% in 2004, followed by a slight fall to 55.8% in 2007 and now it is 62.4 percent point by the last survey. 75 percent of CPRs are targeted by the government within 2022. Population control is a multi-task

strategy that has to depend on cultural, economic, social, and educational changes. Bangladesh's government has taken initiatives to get rid of this problem.

3. Why Bangladesh need to minimize the size of the population

Bangladesh is seeking the tag of a developed country by 2041, on the 70th birthday of the nation's independence^[4]. But, unless the nation can effectively address overpopulation, the goal will remain a nightmare. Overpopulation is creating many problems that are hindering the growth potentiality of the country. Broadly the sectors like education, healthcare, rural economy, and employment are affected severely. In 1995 primary enrollment was 95% with teacher-student ratio of 1:70. According to the Bangladesh planning commission, if universal primary education is to be achieved and if the teacher-student ratio is to be improved even to 1:50, the number of resources dedicated to primary education has to be doubled. Around the year 2000, the doctor-population ratio was 1:5506 and it was 1:1724 in 2018. Though the scenario has improved, we need to go for a long way. Overpopulation is directly hampering the rural economy by raising the landless. The government is struggling to reduce poverty at it's targeted level but overpopulation is a great obstacle here. Food safety will not be sustainable if the country is overpopulated. According to the Social outlook 2018 and APE of ILO, youth (age 15-24) unemployment increased from 6.32 percent in 2000 to 12.8 percent in 2017. The Labor Force Survey (2016-17) found that around 30 percent of the total youth are neither in education nor training. These data are showing the severity of the problem. Now population density is 1103 per square kilometers which is highest in South Asia. So, before crossing the carrying capacity minimization of the size of our population is a must.

4. Literature Reviews

To forecast the population different types of simple, double, and triple exponential models, simple and multiple regression models, several growth models, logistic regression, and other advanced models are being used^[5]. ARIMA model is one of the most popular models to forecast and model the time series data in different disciplines. The most relevant population forecasting research to our paper is referred here. Using the data between 1998 and 2015, Osman G. showed that the total population of Kuwait is likely to arrive at 5 crores by 2020, and 7 crores in 2030^[6]. In another empirical study, Zakria and Faqir modeled the population of Pakistan from 1951 to 2007 using an ARIMA (1, 2, 0) and forecasted the population of Pakistan for the next two decades. Their finding is that the population of Pakistan would be nearly 23 crores in 2027 if the population growth trend will nonstop^[7]. Maria *et al.* (2014)^[8] made projections for Portugal applying a multi-state cohort-component model. However, the cohort-component models tend to overestimate the slow-growing populations and underestimate fast-growing ones^[8]. Hazlenah *et al.* (2013)^[9] applied the Lee-Carter model to forecast the age-specific fertility curves in the Malaysian Peninsula^[9]. They frequently used the probabilistic approaches in their literature. Tom Wilson (2013)^[10] applied a probabilistic model to measure a substantial population forecast in the Greater Sydney region,

NSW, Australia. The authors find that UN estimates underestimate the uncertainty in population projections though these forecasts are often conditional on significant amounts of uncertainty^[10]. Stefan R. (2008)^[11] provided a very clear primer on population forecast errors. He projected that mean algebraic percentage errors or MALPE and mean absolute percentage errors or MAPE are the most commonly used standards to determine the precision and the bias in population projections^[11]. In their article Guangqing C. and Paul R. Voss (2011)^[4] reinforces the argument that Spatio-temporal regression or the extrapolation techniques work better for small geographical area forecasts^[12]. Again, to estimate the mortality rates of different forms Cary C. L. T., Shuai Y. (2015)^[13], and Tzulung L., Cary C. L. T., (2015) applied linear regression approach and structural models^[13,14]. To forecast US population trajectories Jeff Tayman *et al.* (2007), also used ARIMA methodology^[15]. Vladimir Mkhitarian *et al.* (2015) used a similar model to the region of the Russian Federation. They applied autoregressive methods to analysis demographic procedures and fatigue function for diverse population groups of the metro area^[16]. In this paper, the well-liked ARIMA model is also used to forecast the population growth rate.

5. Research Objectives

- To assessment the data type, stationarity or non-stationarity, of population growth rate over the period.
- To study autocorrelation in the observed time series of population growth rate.
- To forecast population growth rate using an appropriate ARIMA (P, I, Q) Model.
- To check model fitness using AIC, BIC or, and p-values.

6. Model

To forecast the population growth rate, we applied a simple ARIMA model in this paper. The ARIMA model was popularized by G. Box and G. Jenkins in the early 1970s [17]. An ARIMA model can be articulated as

$$Y_t = \alpha + \sum \phi_p Y_{t-p} + \sum \theta_q \varepsilon_{t-q} \dots \dots \quad (1)$$

Where, Y_t = Population growth rates at different periods t , Y_{t-p} = Population growth rate at different periods $(t-p)$, θ_q = Random shock at period q , ε_{t-q} = Random error term at period t , and $(\alpha, \phi_p, \theta_q)$ are parameters to be estimated.

Since the 80s decade, this model has gained enormous popularity due to its versatility in many areas in business and economics. The parameters of the ARIMA (P, I, Q) model are defined as follows: P is here how many lags included in the model. I is here the number of how many times are differenced. The order of the moving average here is Q. After establishing ARIMA, the next step is model forecasting. In the model forecasting in this paper, we focused on the out-of-sample forecasting.

To build up a reliable forecasting model, the following three factors will be considered

1. Accuracy level
2. Data and information availability
3. The time horizon for forecasting

Table 1: Descriptive Statistics

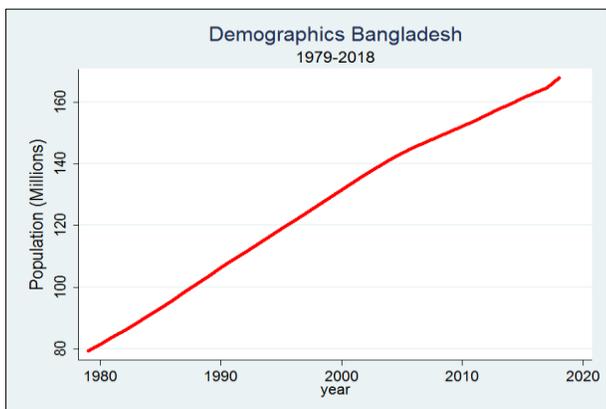
Variable	Observations	Mean	Standard Dev.	Min	Max
Year	40	1998.5	11.69	1979	2018
Population growth rate	40	1.9245	.637	1.011	2.82
Total Population	40	1.26e+08	2.70e+07	7.92e+07	1.68e+08

Source: Author's calculation

Table 1 shows that there are 40 years of observations were collected. The mean population growth rate is 1.924. The range of population growth rates is 1.01 and 2.82.

7. Data and Methods of Analysis

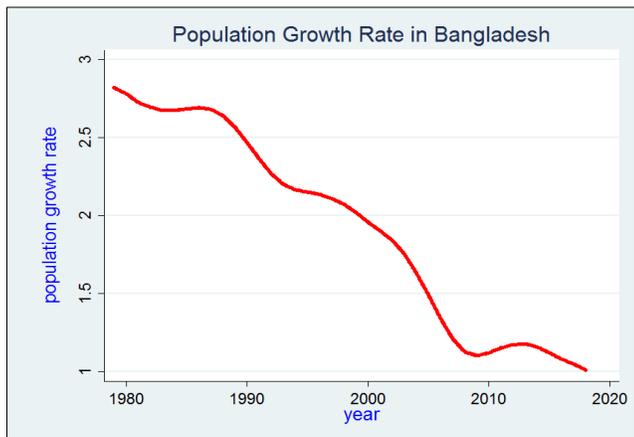
To forecast population growth, data on this macroeconomic variable was collected from 1979 to 2018 from the World Development Indicators (WDI), World Bank. Table 1 shows that the data file consists of 40 observations. A graphical representation of data reveals that population series follows an increasing pattern over this period (Fig 1).



Source: Author's calculation

Fig 1: The total population in Bangladesh

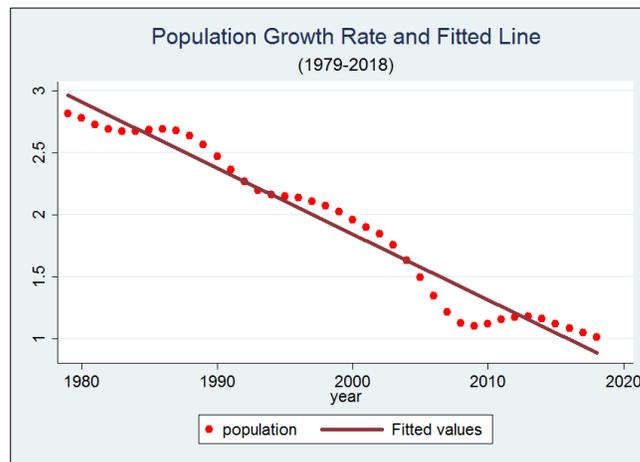
Figure 1 shows the total population size in Bangladesh over 40 years from 1979 to 2018. The figure was given as the value of the millions of the population.



Source: Author's calculation

Fig 2: Line of population growth rates in Bangladesh

Figure 2 illustrates the line of population growth rates over four decades. The line shows population growth rates have been decreasing sharply.



Source: Author's calculation

Fig 3: Population growth rate with fitted line

In figure 3, we added a fitted line. It also shows population growth rates have been decreasing sharply.

Table 2: Level data- without trend & lags (ADF test)

Variable	ADF Statistic	Probability	Critical Values	Conclusion
Population Growth Rate	-0.132	0.9260	-3.655 @1%	Not stationary
			-2.961 @5%	Not stationary
			-2.61 @10%	Not stationary

Source: Author's calculation

Table 3: Level data- with trend and lags (ADF test)

Variable	ADF Statistic	Probability	Critical Values	Conclusion
Population Growth Rate	-2.699	0.2366	-4.270 @1%	Not stationary
			-3.552 @5%	Not stationary
			-3.22 @10%	Not stationary

Source: Author's calculation

Tables 2 and 3 present the ADF test result with and without lags of level data of the population growth series from 1979-2008.

Table 4: 1st Difference-without trend & lags (ADF test)

Variable	ADF Statistic	Probability	Critical Values	Conclusion
Population Growth Rate	-2.645	0.5113	-3.662 @1%	Not stationary
			-2.964 @5%	Not stationary
			-2.614 @10%	Stationary

Source: Author's calculation

Table 5: 1st Difference-with trend & lags (ADF test)

Variable	ADF Statistic	Probability	Critical Values	Conclusion
Population Growth Rate	-4.487	0.04	-4.279 @1%	Stationary
			-3.556 @5%	Stationary
			-3.214 @10%	Stationary

Source: Author's calculation

Tables 4 and 5 present the ADF test results with and without trend and lags of the population growth after the first differences during 1979-2018.

Table 6: Level data- without trend & lags (PP test)

Variable	ADF Statistic	Probability	Critical Values	Conclusion	
Population Growth Rate	Z(rho)-0.258	0.9211	-18.152	@1%	Not stationary
	Z(t)-0.330		-3.655		
			-12.941	@5%	Not stationary
			-2.961		
			-10.480	@10%	Not stationary
			-2.613		

Source: Author's calculation

Table 7: Level data- with trend and lags (PP test)

Variable	ADF Statistic	Probability	Critical Values	Conclusion	
Population Growth Rate	Z(rho)-6.643	0.6888	-24.292	@1%	Not stationary
	Z(t)-1.833		-4.251		
			-18.964	@5%	Not stationary
			-3.544		
			-16.272	@10%	Not stationary
			-3.2.6		

Source: Author's calculation

Tables 6 and 7 present the PP test result with and without the lag of the population growth rates of level data of the population growth series during 1979-2008.

Table 8: 1st Difference-without trend & lags (PP test)

Variable	ADF Statistic	Probability	Critical Values	Conclusion	
Population Growth Rate	Z(rho)-10.779	0.1638	-18.084	@1%	Not stationary
	Z(t)-2.326		-3.662		
			-12.916	@5%	Not stationary
			-2.964		
			-10.460	@10%	Stationary
			-2.614		

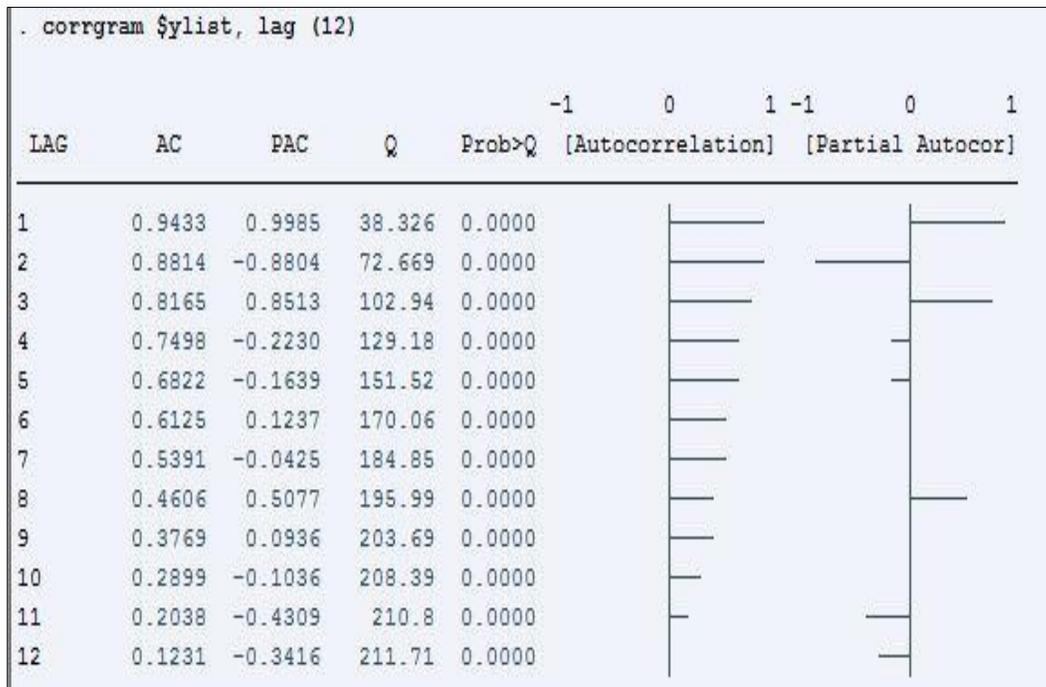
Source: Author's calculation

Table 9: 1st Difference-with trend & lags (PP test)

Variable	ADF Statistic	Probability	Critical Values	Conclusion	
Population Growth Rate	Z(rho) -18.926	0.4558	-24.164	@1%	Not stationary
	Z(t) -3.661		-4.260		
			-18.888	@5%	Stationary
			-3.548		
			-16.224	@10%	Stationary
			-3.209		

Source: Author's calculation

Tables 8 and 9 present the PP test result with and without trend and lags of the population growth after the first differences during 1979-2018.

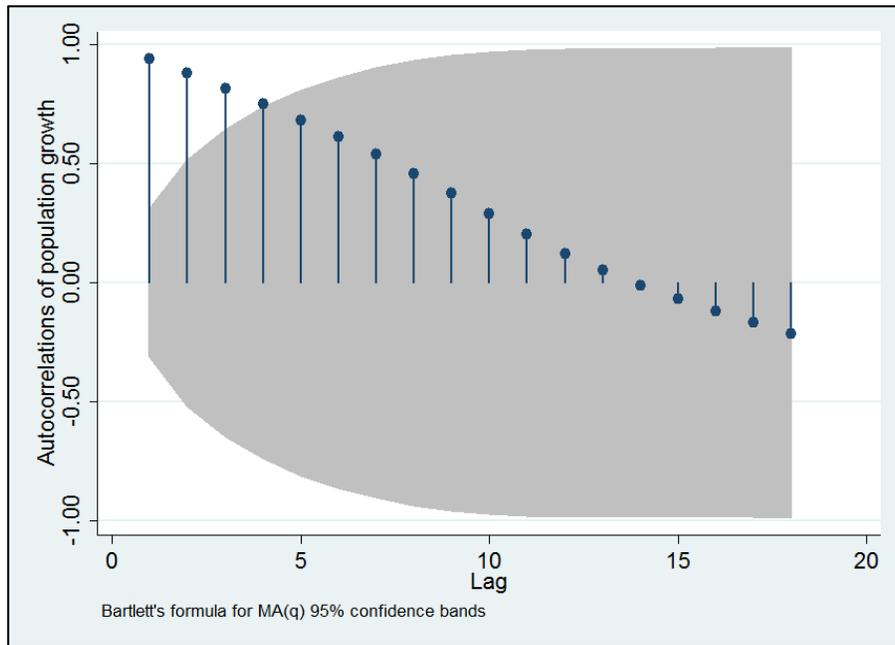


Source: Author's calculation

Fig 4: Correlogram of growth (Level data)

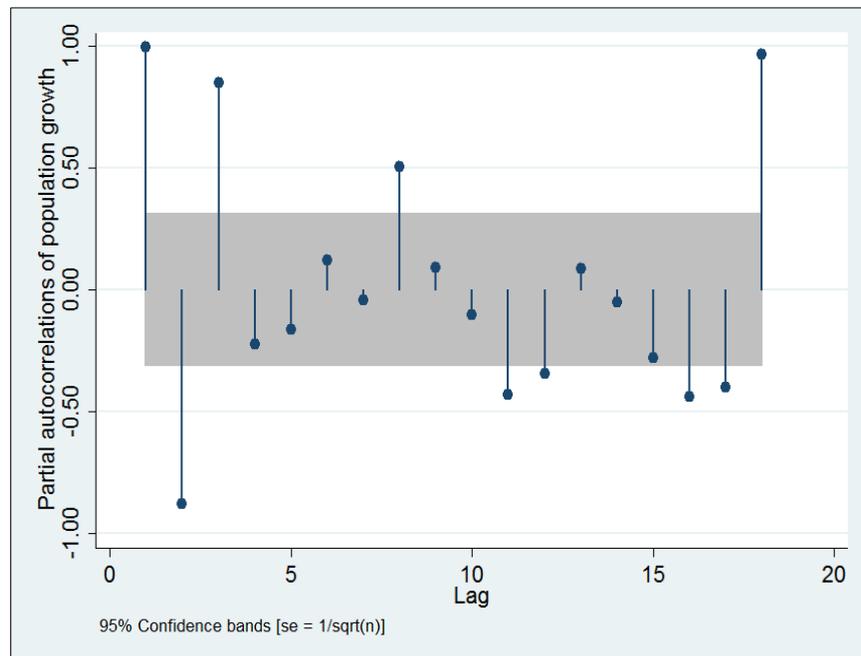
In figure 4, we showed a correlogram of the 40 observations of population growth series,

which deals with autocorrelation of population growth and its diverse lags with respective t-statistics.



Source: Author's calculation

Fig 5a: ACF of population growth rate (Level data)

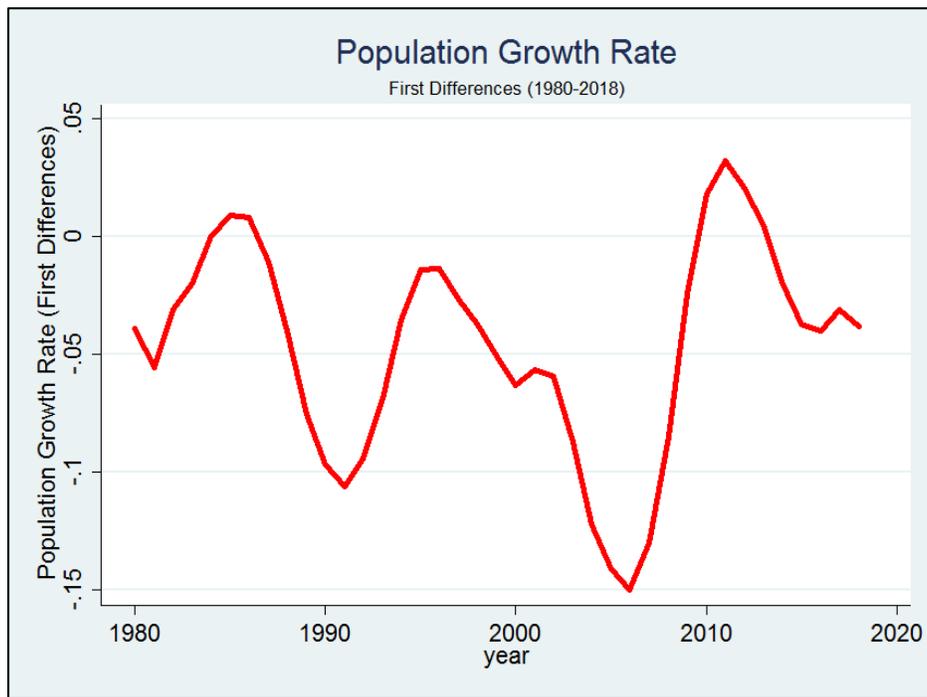


Source: Author's calculation

Fig 5b: PACF of population growth (Level data)

Figure 5a shows that there are significant spikes at ACF at few lags, and after these lags, the ACFs slowly decline. We can conclude that it is a non-stationary time series. Again, from Figure 5b, Partial Autocorrelation (PACF) of the level data series is estimated. We see that it has some significant spike at different

lags. So, we can accept the null (H_0) that the population growth rate series of 40 observations is non-stationary. Since the ACF and PACF have spikes, the first differences data can be applied for this model.

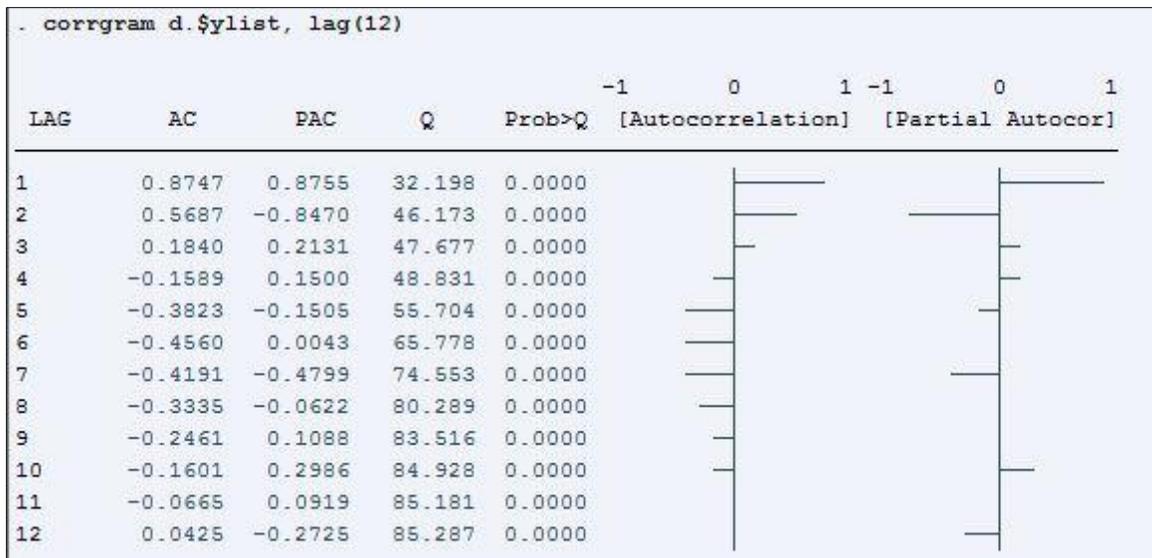


Source: Author's calculation

Fig 6: First Differences Population Growth Rates in Bangladesh

Figure 6 shows the first differenced data of population growth rates over four decades.

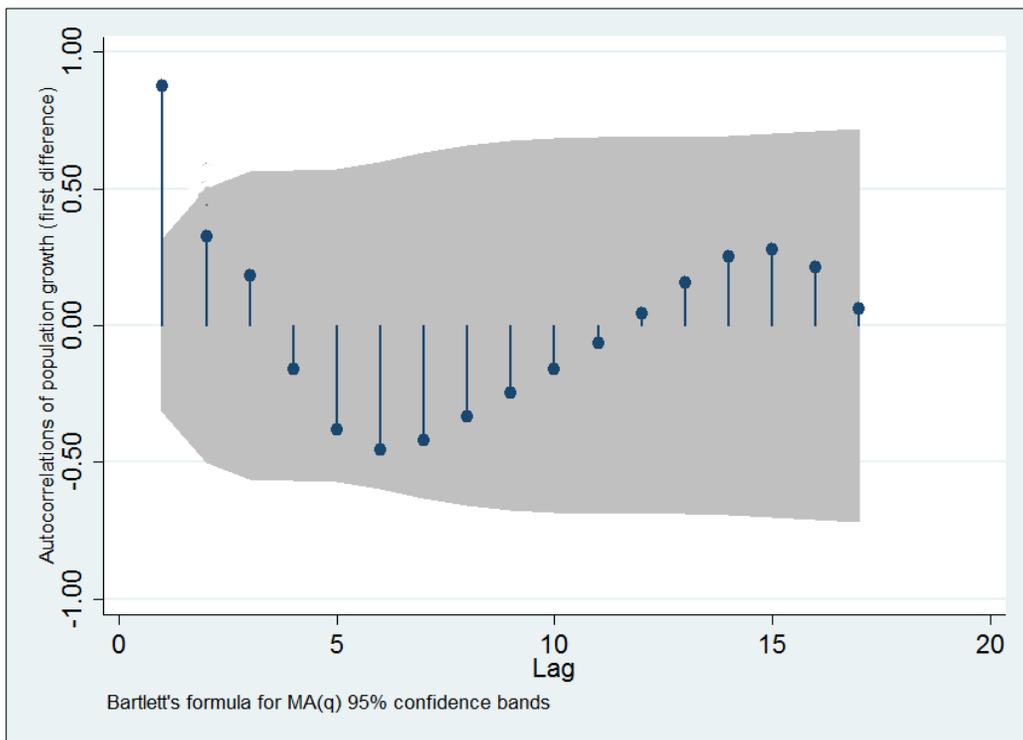
The line illustrates the first differences of population growth rates have been fluctuating.



Source: Author's calculation

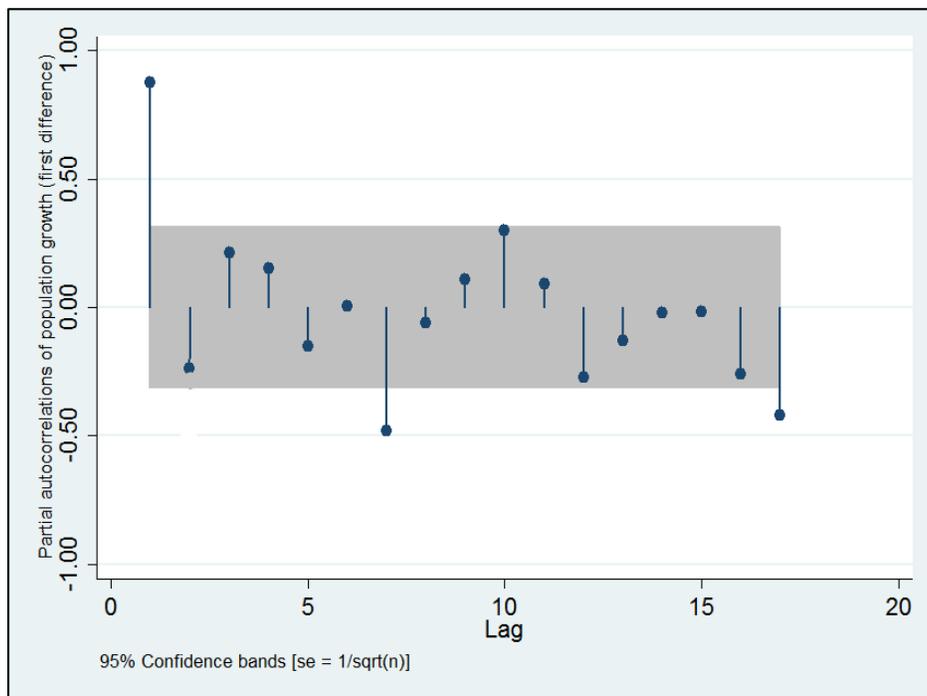
Fig 7: Corrogram of growth (first-order difference)

In figure 7, we showed a correlogram of the first difference data of population growth series.



Source: Author's calculation

Fig 8a: ACF of population growth (first difference data)



Source: Author's calculation

Fig 8b: PACF of population growth (first difference data)

From figures 8a and 8b, it clearly shows that the ACF and PACF first difference series has only one significant spike at any lags. To make simple our model we can conclude that now our time

series is stationary. ARIMA models with first differences, therefore, are recommended for the time series.

Table 10: Evaluation of ARIMA models (I)

Arima →	Arima (1,1,1)	Arima (0,1,1)	Arima(1,1,0)	Arima (2,1,0)	Arima (0,1,2)
Population - cons	-0.0434*** (-6.72)	-0.0459*** (-5.60)	-0.0445 (-1.60)	-0.0455*** (-5.04)	-0.0449*** (-5.33)
ARMA L.ar	0.810*** (7.65)		0.857*** (10.22)	1.593*** (15.45)	
L2.ar				-0.823*** (-8.13)	
L.ma	0.717*** (4.27)	1.000 (.)			1.680*** (39.58)
L2.ma					1.000 (.)
Sigma_cons	0.0153*** (7.36)	0.0242*** (7.21)	0.0217*** (7.94)	0.0114*** (11.21)	0.0147*** (8.03)
N	39	39	39	39	39
AIC	-224.6	-171.9	-180.9	-226.7	-205.1
BIC	-217.9	-168.6	-175.9	-220.1	-200.2

Source: Author’s calculation
 t statistics in parentheses * p< 0.05, ** p < 0.01, *** p < 0.001

Table 11: Evaluation of ARIMA models (II)

Arima →	Arima (2, 1, 2)	Arima (3, 1, 3)	Arima (2, 1, 1)	Arima (2, 2, 2)	Arima (1, 2, 1)
Population – cons	-0.0448*** (-3.56)	-0.0439*** (-2.94)	-0.0445*** (-4.32)	-0.00131 (-0.26)	-0.00233 (-0.23)
ARMA L.ar	1.472*** (6.65)	1.908 (0.51)	1.545*** (11.69)	1.351*** (14.22)	0.606*** (3.45)
L2.ar	-0.734* (-2.13)	-1.533** (-2.59)	-0.783* (-1.92)	-0.836* (-2.23)	
L3.ar		0.439 (1.23)			
L.ma	0.305 (1.32)	-0.385 (-0.97)	0.203 (1.14)	-0.755 (-0.00)	0.435* (2.33)
L2.ma	0.233 (0.93)	0.740 (.)		1.000 (0.00)	
L3.ma		0.353 (0.71)			
Sigma_cons	0.0109*** (10.32)	0.000895*** (5.54)	0.0112*** (9.78)	0.00977*** (0.00)	0.0139*** (6.99)
N	39	39	39	38	38
AIC	-215.7	-213.3	-221.0	-216.2	-208.1
BIC	-207.7	-210.7	-114.7	-209.4	-201.6

Source: Author’s calculation
 t statistics in parentheses * p< 0.05, ** p < 0.01, *** p < 0.001

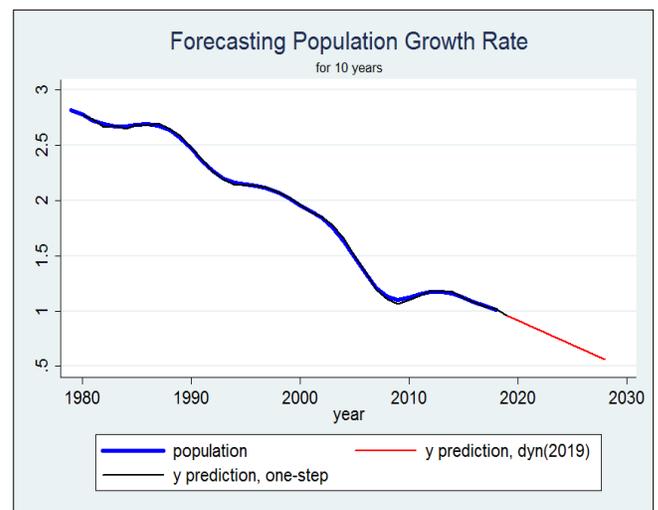
In tables 10 and 11, various ARMA and ARIMA models with several orders of difference, autoregressive, and moving average terms were compared based on their performance and significance levels. By using statistics such as P-value and minimum AIC and BIC these models were checked and verified. ARIMA (1, 1, 1) is the best choice among several models based on minimum values of AIC (-224.6) and BIC (-17.9) and P-value*** (0.012). Thus, the ARIMA (1, 1, 1) model is used in this paper for future forecasting. Therefore, other than within sample forecasts, this study also estimated one and two decades out of sample forecasts of the model to measure the forecast ability. Results indicate that population growth will continue to decline in Bangladesh. The 10-year forecast of Bangladeshi population growth rates is presented in table 12.

Table 12: Comparative Trend of Growth Rate for the Next 10 Years Using ARIMA (1, 1, 1)

Year	Projected Population Growth Rate	Lower limit	Upper Limit
2019	.963	.9142	1.013
2020	.917	.8677	.9665
2021	.871	.8217	.9205
2022	.825	.7762	.8750
2023	.780	.7311	.8299
2024	.735	.6864	.7852
2025	.691	.6419	.7407
2026	.647	.5977	.6964
2027	.602	.5535	.6523
2028	.558	.5096	.6083

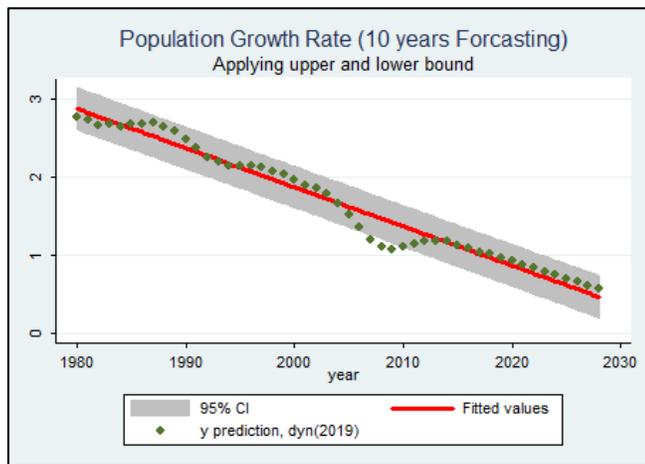
Source: Author’s calculation

Table 12 contains the forecasted population for the subsequently first decades using ARIMA (1, 1, 1). The growth rates are declined from .963 to .558 in 2nd column. The 1st column shows future years and 2nd column shows the forecasted population for the next 10 years. The 3rd and 4th columns show the lower and upper regions of the forecasted population during the years 2019-2028.



Source: Author’s calculation

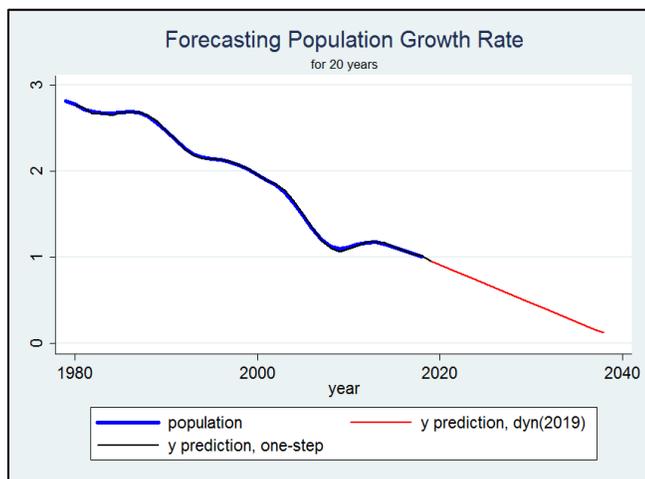
Fig 9: Population growth rate forecasting for 10 years



Source: Author's calculation

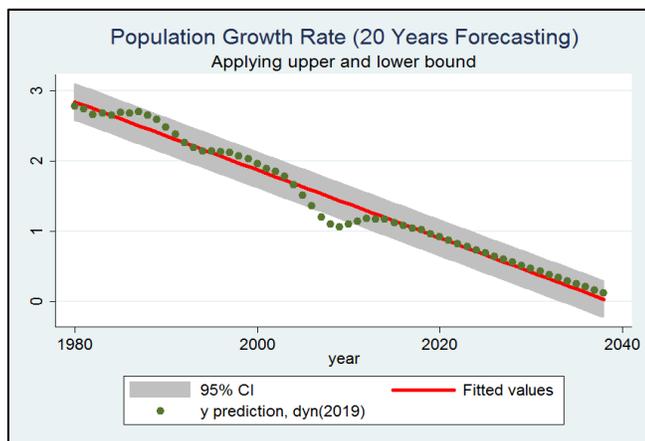
Fig 10: Population growth forecasting with upper and lower bounds

In figures 09 and 10, forecasted population growth lines are illustrated for the next decade with an ARIMA (1,1,1) model. The forecasted line illustrated in figure 09 and the upper and lower bound is included in figure 10.



Source: Author's calculation

Fig 11: Population growth rate forecasting for 20 years



Source: Author's calculation

Fig 12: Population Growth Rate forecasting for 20 years with upper and lower bounds

In figures 11 and 12, forecasted population growth lines are illustrated for the next two decades with an ARIMA (1, 1, 1) model. The forecasted line illustrated in figure 11 and in figure 12, the upper and lower bound are included.

7. Conclusion

The ARIMA (1, 1, 1) model is an appropriate and the most suitable model to forecast the population growth rates for the next decades. The AIC and BIC's values for this model are -224.6 and -217.9 which are the minimum among the nearer other ARIMA models given in tables 10 and 11. Also, P-value is 0.012 which determines the significance of your model. The model predicts that by 2028 and 2038, Bangladesh's population growth rate would be approximate, 50% and 35%; unless and until more stern population control policies and strategies are implemented in Bangladesh. This clearly shows that indeed population growth won't be a real threat to the future of Bangladesh, though Bangladesh is currently experiencing high levels of unemployment, food shortage, crime, and poverty. The declining population growth rates show that Bangladesh will be able to control the population. These findings are close to the findings reported by all other national and international bureau. These findings are particularly important for the government of Bangladesh as well as other organizations, particularly when it comes to planning for the upcoming decades. Though we forecasted population growth rates for 10 and 20 years, it is recommended that researchers should be alert when forecasting for more than 5 years.

References

1. World population review. Retrieved from <http://worldpopulationreview.com/countries/bangladesh-population/>
2. Population division, Department of Economic and Social Affairs. World Population Prospects: The 2019 Revision.
3. Population Pyramids of the World from 1950 to 2100. Retrieved from <https://www.populationpyramid.net/bangladesh/1971/>
4. Biru P Paul. Becoming a developed country by 2041, 2017. Retrieved from <https://www.thedailystar.net/opinion/open-sky-0/becoming-developed-country-2041-1398349>
5. Krishnamurthy S. *Basic Demographic Techniques and Applications*, Thousand Oaks, CA: Sage Publications, Inc, 1998.
6. Osman G. Forecasting Population and Demographic Composition of Kuwait until 2030, International Journal of Economics and Financial Issues. 2016; 6(4):1429-1435.
7. Zakria M, Faqir M. Forecasting the Population of Pakistan Using ARIMA Models, Pak. J Agriculture and Science. 2009; 46(3).
8. Maria OM, Rodrigues I, Teresa R. Multistate projections by level of education for Portugal, 2011-2031. Journal of Population Research. 2014; 31(4):317-343.
9. Hazlenah H, Abdul AA. Structural Modeling of age Specific Fertility Curves in Peninsular Malaysia: An Approach of Lee Carter Method. Paper Presented at the AIP Conference Proceedings. 2013; 1571:p1062-1068.

10. Wilson Tom. Quantifying the uncertainty of regional demographic forecasts. *Applied Geography*, 42. doi: 10.1016/j.apgeog. 2013; 05(006):108-115.
11. Stefan R. Population forecast errors: A primer for planners. *Journal of Planning Education and Research*. 2008; 27(4):417-430.
12. Guangqing C, Paul RV. Small-area population forecasting: Borrowing strength across space and time. *Population, Space and Place*. 2011; 17(5):505-520.
13. Cary CLT, Shuai Y. A linear regression approach to modeling mortality rates of different forms. *North American Actuarial Journal*. 2015; 19(1):1-23.
14. Tzulung L, Cary CLT. A simple linear regression approach to modeling and forecasting mortality rates. *Journal of Forecasting*. 2015; 34(7):543-559.
15. Jeff Tayman, Stanley KS, Lin J. Precision, bias, and uncertainty for state population forecasts: An exploratory analysis of time series models. *Popul Research and Policy Rev*. 2007; 26:347-369.
16. Vladimir M, Mikhaylova SS, Sarycheva TV, Bakumenko, LP. Long-term forecasting of the number of labor pensioners in the region. *Mediterranean Journal of Social Sciences*. 2015; 6(3):351-360.
17. Hall SG, Asteriou D. ARIMA models and the Box- Jenkins Methodology. *Applied Econometrics*. 2011; 2(2):265-286