Predicting the Demographic Future of Bangladesh: Application and Comparison of ARIMA and Combined Population Forecasts

Kazi Zubair Hossain (khossain@isrt.ac.bd)
Institute of Statistical Research and Training, University of Dhaka Dhaka, Bangladesh

ABSTRACT

The future trajectory of population size and composition is crucial for informed decision-making in public and private domains. Countries with continuous population growth, like Bangladesh, need accurate and prompt population forecasts at national and local levels to efficiently use limited resources and properly utilize the rising workforce. This study used ARIMA models to forecast age-sex-specific populations of Bangladesh until 2031 and then aggregated them to produce national forecasts. Aggregated ARIMA showed one of the lowest errors among the existing methods compared to the 2022 population census count. Additionally, combined forecasts yielded the highest accuracy compared to individual methods. The results present detailed insights into the future population of Bangladesh disaggregated by age and sex.

Keywords: *Population forecasts, ARIMA, Box-Jenkins, combined forecasts, age-sex dynamics, Bangladesh*

JEL Classification: *C510, C520, C530, J110*

1. INTRODUCTION

Bangladesh, with a population exceeding 170 million, stands out as one of the world's most populous countries, ranking eighth globally, fifth in Asia, and third in South Asia (World Bank, 2023). The last fifty years' demographic transition of Bangladesh is characterized by a doubled total population, density, and life expectancy, halved annual growth rate and dependency ratio, a two-third fall in total fertility rate and family size, a four-fifth fall in infant mortality rate, and tripled literacy and urbanization rate (BBS, 2023; UN DESA, 2022). These illustrate Bangladesh's demographic successes and, at the same time, hint at the challenges. The successes can be attributed to well-planned public policies (e.g., family planning program). In the same way, to address the existing and upcoming multidimensional challenges in the face of overpopulation (e.g., efficient allocation of resources, sustainable urbanization, decent work, etc.), proper planning should be the foundation of all policies. A significant input for effective short, medium, and long-term planning is the knowledge about the future population (Wilson & Rees, 2005).

Demographic forecasts estimate the most likely future trends of different demographics based on past and current information (George et al., 2004). Population, household, and associated forecasts are essential for economic and social development planning and serve as a foundation for all major planning decisions (Booth, 2006). Education planning, healthcare planning, urban planning, employment planning, environmental planning, energy planning, transport planning, land use planning, social security planning, business planning, etc., to varying degrees, depend on population forecasts (Baxter & Williams, 1978). The utilization of demographic forecasts enables policy planners to anticipate a range of probable outcomes in the future and craft well-informed plans accordingly.

Objective population forecasting can be broadly categorized into four categories: trend extrapolation, cohort component, structural modeling, and microsimulation (Smith et al., 2013). Trend extrapolation methods rely on the extension of observed past trends. They may range from simple methods like linear extrapolation to more complex models like Autoregressive Integrated Moving Average (ARIMA) time series models. These methods are often used to estimate the overall population and are typically applied to univariate series. Although trend extrapolation methods do not present any underlying cause of population change, they are characterized by minimal data requirements, low expenses, and timeliness.

Considered a complex extrapolation method, ARIMA models aim to identify the underlying stochastic process that produces a time series by combining autoregression, moving average, and differencing processes and use it to forecast future values. ARIMA models' dynamic and stochastic framework provides a statistical basis for constructing prediction intervals around a particular forecast (Box & Jenkins, 1976). The Box-Jenkins methodology was introduced in 1970, which popularized the ARIMA models for forecasting different types of time series (Box & Jenkins, 1970).

ARIMA models have been widely used in analyzing and forecasting different demographics observed over time (Land, 1986). The first such study was conducted by Saboia (1974), where ARIMA (0, 2, 1) emerged as a suitable model to forecast the total population of Sweden for 1965 and 1970 using five-year interval population data from 1780-1960. Subsequently, ARIMA models have been applied to forecast national/sub-national populations for Sweden (Cohen, 1986), the United States (Alho & Spencer, 1997; Lee & Tuljapurkar, 1994; Pflaumer, 1992; Tayman et al., 2007), Pakistan (Zakria & Muhammad, 2009), Iraq (Abdulgader, 2016), China (Dai & Chen, 2019), and others in the last five decades. It has also been used to forecast individual demographic components such as fertility (Bell, 1997; Caleiro, 2010; Du & Chan, 2023; Keilman & Pham, 2000; Miller & McKenzie, 1984), mortality (Carter, 1996; Debón et al., 2008; Lee & Carter, 1992; McNown & Rogers, 1989; Sarpong, 2013), and migration (Beer, 1993; Bijak et al., 2019; Raymer & Wiilekens, 2008).

Planning and population forecasts are intertwined for an overpopulated country like Bangladesh to utilize its limited resources efficiently. Several sources of population forecasts exist in Bangladesh to meet that demand. In 2015, the United Nations Population Fund (UNFPA) published detailed population projections (five-year interval) for the period 2011-2061 using the cohort component method based on 2011's Population and Household Census data (Hayes, 2015). Following that, BBS also produced similar projections for 2011-2061 using the same data and method (Hossain et al., 2015). Mahsin & Hossain (2012) combined the above method with the Bayesian approach to produce five-yearly population projections for 2006-2051 based on 2001 census data. Haque et al. (2012) applied a simple trend extrapolation method named the logistic (Verhulst) model on the 1981-2006 BBS sample vital registration system's data to forecast the yearly population for 2007-2035. Later, Islam & Hoque (2014) used another simple trend extrapolation method- a two-parameter negative exponential model with an exponential growth rate method on 1991 and 2001 census data to forecast the five-year population for 2002-2031. Another universal source of population forecasts is the UN probabilistic population projections for all countries until 2100 (UN DESA, 2017a, 2022).

The question- "Where does a complex extrapolation method like ARIMA fit in this scenario in terms of performance and accuracy?" is probed in this study. To provide detailed insight into the future population structure, disaggregated by age-sex, 34 different time series for 1950-2015 are extracted from the World Population Prospects (UN DESA, 2017a). Box-Jenkins methodology is then applied to find the most suitable ARIMA model for each series, and forecasts are produced for 2016-2031. Upon aggregating the estimates for 2022, accuracy is checked against the observed population count from the latest census of 2022 (BBS, 2023). This study also probed the performance of combined forecasts by amalgamating existing forecasts. The remainder of the paper is organized as follows: section 2 contains methodology, section 3 presents the results, and section 4 discusses the findings and concludes the study.

2. METHODOLOGY

2.1 Data

For this study, 34 age-sex-specific time series containing yearly population estimates for Bangladesh have been extracted from the United Nations Population Division's World Population Prospects 2017 online database (UN DESA, 2017a). The yearly estimates are consistent with 1961, 1974, 1981, 1991, 2001, and 2011 censuses and adjusted for under-enumeration with estimates from subsequent trends in fertility, mortality, and international migration (UN DESA, 2017b). Each time series covered a period of 66 years (1950-2015). These 34 series accounted for 17 different age groups: 0-4, 5-9, 10-14, 15-19, 20-24, …, 65-69, 70-74, 75-79, and 80+, over two sexes: male and female (Figure 1). It is worth noting that at least 50 observations are preferred for univariate time series modeling (Box & Jenkins, 1976; Saboia, 1974).

2.2 ARIMA Modeling

The ARIMA model is a statistical method used for time series analysis and forecasting. In demographic forecasting, it is considered as a complex trend extrapolation method (Smith et al., 2013). It was first formalized by Box-Jenkins methodology in 1970 (Box & Jenkins, 1970). It estimates the underlying stochastic process from which a time series arises and forecasts future values based on the estimated stochastic process (Box et al., 2016). ARIMA models are specified by three values, namely (*p, d, q*)- where *p* is the order of autoregressive (AR) part (past trends), *q* is the order of moving average (MA) part (past errors), and *d* is the order of difference (ensures stationarity, i.e., constant mean and variance). ARIMA models forecast future values of a univariate time series through a linear combination of past trends (AR) and errors (MA), usually applied on stationarized series. A stationary time series can be represented by an ARMA (*p, q*) model, which can be defined as follows:

$$Z_t = (\varphi_1 Z_{t-1} + \varphi_2 Z_{t-2} + \cdots + \varphi_p Z_{t-p}) + (a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}) \quad \ldots \text{(i)}$$

Where, Z_t is the original time series (with time index t), Z_{t-i} is the i^{th} lag of the time series (i=1, 2, 3, …, p), φ_i is the AR parameter, a_t represents random errors of the time series, a_{t-k} is the random error of k^{th} lag (k=1, 2, 3, …, q), and θ_k is MA parameter. We assume Z_t to be independent of a_t and a_t follows normal distribution. Equation (i) can be rewritten as,

$$\varphi_p(B)Z_t = \theta_q(B)a_t$$

And, from this, for a non-stationary time series that has been differenced for *d* times, the ARIMA (*p, d, q*) model can be formulated as,

$$\varphi_p(B)(1-B)^d Z_t = \theta_0 + \theta_q(B)a_t$$

Where,
B is a backshift operator, i.e., $B^i Z_t = Z_{t-i}$ and $B^k a_t = a_{t-k}$
AR part: $\varphi_p(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \varphi_3 B^3 - \cdots - \varphi_p B^p$
MA part: $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - \cdots - \theta_q B^q$

θ_0 is the deterministic trend term and is often omitted from the model as it is merely needed (Wei, 2006).

This study used the Box-Jenkins iterative model-building approach to pick suitable ARIMA models for each of the 34 time series (Box et al., 2016). In this approach, a tentative model is selected (model identification), parameters are estimated (parameter estimation), and residuals are analyzed (diagnostic checking) to decide whether to consider the model for forecasting or not. The detailed implementation layout is described in Figure 2 (Box et al., 2016; Cryer & Chan, 2008; Dickey & Fuller, 1979, 1981; Hyndman & Athanasopoulos, 2018; Kwiatkowski et al., 1992; Wei, 2006).

For each series, the base year is 1950, the launch year is 2015, the base period is 1950-2015, the target year is 2031, and the forecast horizon is 2016-2031. As ARIMA forecasting is more suitable for short and medium-term (Smith et al., 2013), this study considered a forecast horizon that covers two censuses: 2022 and 2031.

2.3 Forecast Accuracy

This study aggregated 2022's ARIMA forecasts across the 34 age-sex-specific series to produce a total population estimate, similar to Tayman et al. (Tayman et al., 2007), where they aggregated state-wise estimates. Then it was compared to the observed total population of Bangladesh's Population and Housing Census 2022 (BBS, 2023) by calculating absolute percentage error (APE) as follows.

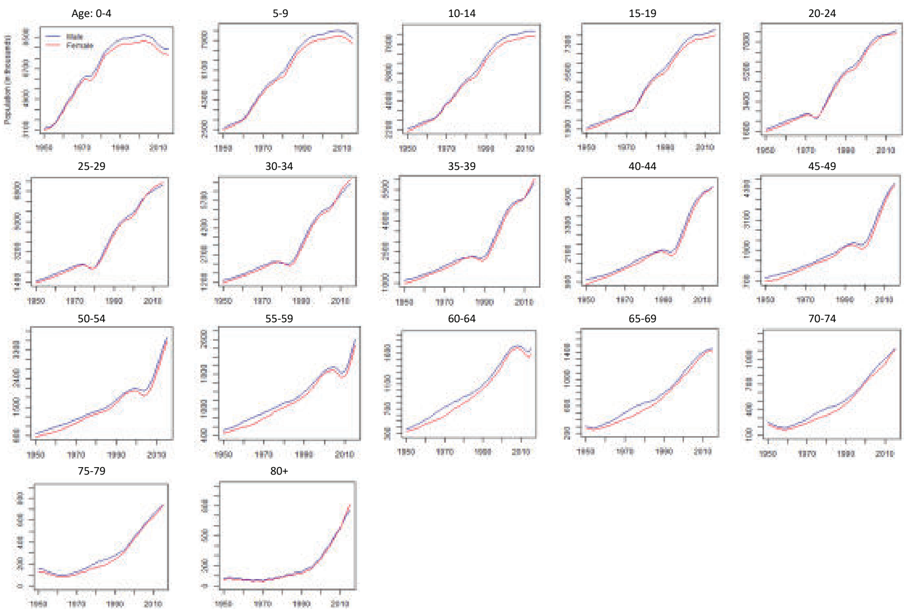
$$APE = \left|\frac{O_t - F_t}{O_t}\right| \times 100\%$$

Where, O_t is the observed value and F_t is the forecasted value.

We also calculated APE for 2022's forecasts from five other studies (Haque et al., 2012; Hayes, 2015; Islam & Hoque, 2014; Mahsin & Hossain, 2012; UN DESA, 2017b). Later, this study combined (simple mean) 2022's forecasts from these five methods with ARIMA and produced combined forecasts. As combination of different forecasts usually outperforms individual forecasts and reduces variability (Armstrong, 2001; Makridakis et al., 2022; Smith et al., 2013), we also calculated APE for it. Finally, we reviewed all the APEs to assess which method performed better in accuracy.

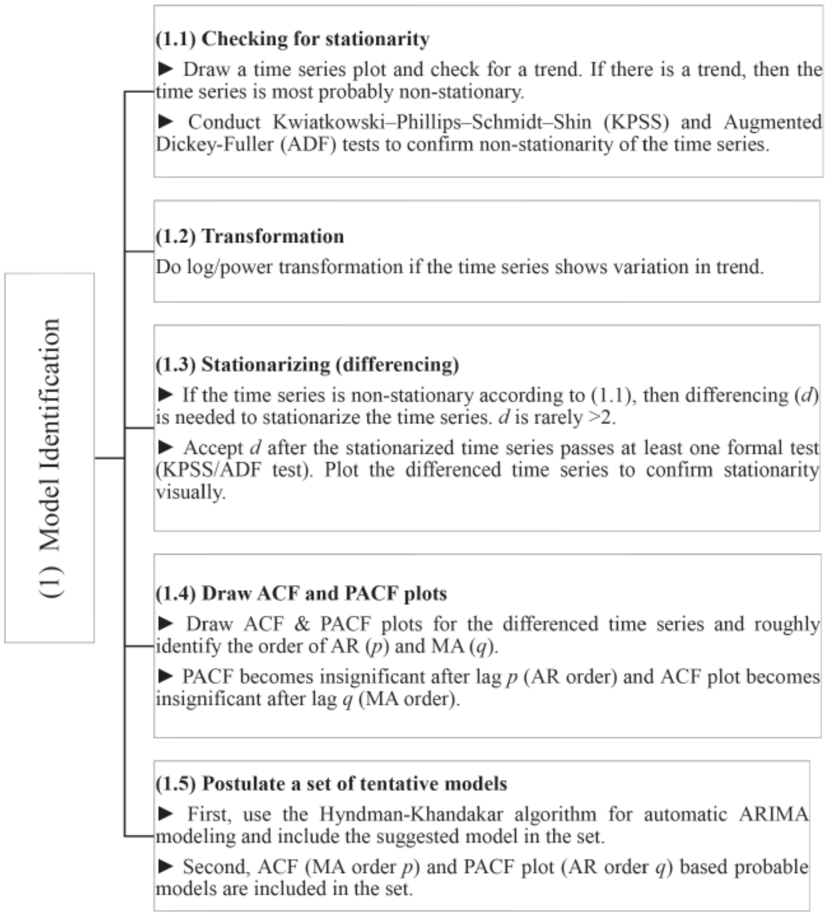
Time series plots of 34 different age-sex-specific annual population estimates (1950-2015)

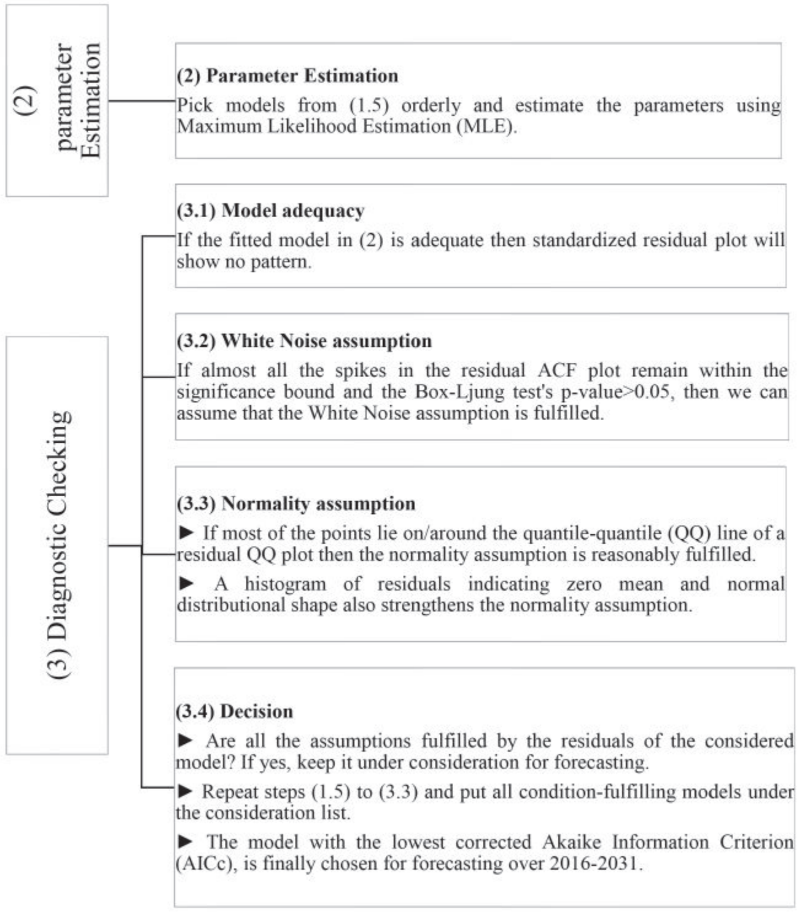
Figure 1



Implementing Box-Jenkins approach

Figure 2




(2) parameter Estimation
(2) Parameter Estimation
Pick models from (1.5) orderly and estimate the parameters using Maximum Likelihood Estimation (MLE).
(3) Diagnostic Checking
(3.1) Model adequacy
If the fitted model in (2) is adequate then standardized residual plot will show no pattern.
(3.2) White Noise assumption
If almost all the spikes in the residual ACF plot remain within the significance bound and the Box-Ljung test's p-value>0.05, then we can assume that the White Noise assumption is fulfilled.
(3.3) Normality assumption
► If most of the points lie on/around the quantile-quantile (QQ) line of a residual QQ plot then the normality assumption is reasonably fulfilled.
► A histogram of residuals indicating zero mean and normal distributional shape also strengthens the normality assumption.
(3.4) Decision
► Are all the assumptions fulfilled by the residuals of the considered model? If yes, keep it under consideration for forecasting.
► Repeat steps (1.5) to (3.3) and put all condition-fulfilling models under the consideration list.
► The model with the lowest corrected Akaike Information Criterion (AICc), is finally chosen for forecasting over 2016-2031.

3. RESULTS

3.1 Model Building

The 34 time series exhibited non-stationary characteristics indicated by an upward trend in each series (Figure 2). To further confirm this, we used two different unit root tests, namely, the KPSS and ADF tests. Rejection of the stationary null hypothesis of the KPSS test and failure to reject the non-stationary null hypothesis of the ADF test reconfirmed the non-stationarity of each series.

Checking stationarity of time series by KPSS and ADF tests

Table 1

Series	P-value of the KPSS test	P-value of the ADF test	Characteristic
Male 0-4	0.01	0.99	non-stationary
Male 5-9	0.01	0.99	non-stationary
Male 10-14	0.01	0.95	non-stationary
Male 15-19	0.01	0.53	non-stationary
Male 20-24	0.01	0.45	non-stationary
Male 25-29	0.01	0.53	non-stationary
Male 30-34	0.01	0.81	non-stationary
Male 35-39	0.01	0.80	non-stationary
Male 40-44	0.01	0.99	non-stationary
Male 45-49	0.01	0.99	non-stationary
Male 50-54	0.01	0.99	non-stationary
Male 55-59	0.01	0.95	non-stationary
Male 60-64	0.01	0.74	non-stationary
Male 65-69	0.01	0.49	non-stationary
Male 70-74	0.01	0.84	non-stationary
Male 75-79	0.01	0.70	non-stationary
Male 80+	0.01	0.99	non-stationary
Female 0-4	0.01	0.99	non-stationary
Female 5-9	0.01	0.99	non-stationary
Female 10-14	0.01	0.98	non-stationary
Female 15-19	0.01	0.67	non-stationary
Female 20-24	0.01	0.48	non-stationary
Female 25-29	0.01	0.58	non-stationary
Female 30-34	0.01	0.91	non-stationary
Female 35-39	0.01	0.93	non-stationary
Female 40-44	0.01	0.99	non-stationary
Female 45-49	0.01	0.98	non-stationary
Female 50-54	0.01	0.99	non-stationary
Female 55-59	0.01	0.66	non-stationary
Female 60-64	0.01	0.70	non-stationary
Female 65-69	0.01	0.69	non-stationary
Female 70-74	0.01	0.87	non-stationary
Female 75-79	0.01	0.96	non-stationary
Female 80+	0.01	0.99	non-stationary

Then, five out of 34 time series had to be log-transformed, and ten had to be power-transformed before differencing (Table 2) using Box-Cox transformations (Box & Cox, 1964).

Box-Cox transformation of the time series

Table 2

Series	Box-Cox λ
Male 5-9	1.18
Male 40-44	-0.79
Male 45-49	-0.65
Male 50-54	-0.46
Male 55-59	0 (log)
Male 60-64	0.26
Female 5-9	1.65
Female 40-44	-0.42
Female 45-49	0 (log)
Female 50-54	0 (log)
Female 55-59	0 (log)
Female 60-64	0 (log)
Female 65-69	0.11
Female 70-74	0.15
Female 80+	0.07

Later, all 34 non-stationary time series were differenced ‘d’ times (Table 3) and tested for stationarity by KPSS and ADF tests. Failure to reject the stationary null hypothesis of the KPSS test and/or rejection of the non-stationary null hypothesis of the ADF test confirmed the stationarity of each series. In case of different results from KPSS and ADF tests, the KPSS was preferred (Hyndman, 2014).

Differenced time series

Table 3

Series	d	P-value of the KPSS test	P-value of the ADF test	Remarks
Male 0-4	2	0.10	0.05	
Male 5-9	2	0.06	0.01	
Male 10-14	2	0.10	0.01	
Male 15-19	2	0.10	0.07	KPSS preferred
Male 20-24	2	0.10	0.01	
Male 25-29	2	0.10	0.01	
Male 30-34	2	0.10	0.39	KPSS preferred
Male 35-39	2	0.10	0.05	
Male 40-44	2	0.10	0.02	
Male 45-49	1	0.08	0.04	
Male 50-54	1	0.07	0.09	KPSS preferred

Male 55-59	1	0.10	0.10	KPSS preferred
Male 60-64	2	0.10	0.05	
Male 65-69	3	0.10	0.01	
Male 70-74	3	0.10	0.01	
Male 75-79	3	0.10	0.01	
Male 80+	2	0.10	0.01	
Female 0-4	2	0.10	0.04	
Female 5-9	2	0.08	0.01	
Female 10-14	2	0.10	0.01	
Female 15-19	2	0.10	0.05	KPSS preferred
Female 20-24	2	0.10	0.02	
Female 25-29	2	0.10	0.04	
Female 30-34	2	0.10	0.02	
Female 35-39	2	0.10	0.03	
Female 40-44	1	0.10	0.17	KPSS preferred
Female 45-49	1	0.10	0.07	KPSS preferred
Female 50-54	1	0.10	0.08	KPSS preferred
Female 55-59	1	0.10	0.09	KPSS preferred
Female 60-64	2	0.10	0.03	
Female 65-69	3	0.10	0.01	
Female 70-74	3	0.10	0.01	
Female 75-79	3	0.10	0.01	
Female 80+	3	0.10	0.01	

Seven out of 34 time series were differenced once and thrice, respectively, whereas the rest 20 were differenced twice.

For these 34 time series, we reviewed hundreds of tentative models from the Hyndman-Khandakar algorithm for automatic ARIMA modeling (Hyndman & Khandakar, 2008) and ACF-PACF plots. We selected the models with minimum AICc as such models are often the best for forecasting (Hyndman & Athanasopoulos, 2021). In Table 4, the orders of the selected models are only reported for simplicity.

Orders of the selected ARIMA models

Table 4

Series	p	d	q
Male 0-4	1	2	0
Male 5-9	2	2	2
Male 10-14	1	2	0
Male 15-19	1	2	0
Male 20-24	2	2	1
Male 25-29	2	2	3
Male 30-34	0	2	4
Male 35-39	1	2	2
Male 40-44	1	2	0
Male 45-49	2	1	1
Male 50-54	2	1	1
Male 55-59	2	1	3
Male 60-64	2	2	2
Male 65-69	0	3	1
Male 70-74	0	3	1
Male 75-79	4	3	0

Male 80+	0	2	3
Female 0-4	1	2	0
Female 5-9	2	2	4
Female 10-14	1	2	0
Female 15-19	1	2	0
Female 20-24	2	2	1
Female 25-29	2	2	0
Female 30-34	1	2	4
Female 35-39	2	2	0
Female 40-44	2	1	3
Female 45-49	2	1	0
Female 50-54	2	1	0
Female 55-59	2	1	3
Female 60-64	1	2	3
Female 65-69	0	3	1
Female 70-74	0	3	1
Female 75-79	4	3	0
Female 80+	4	3	1

The parameters of the selected models were estimated by Maximum Likelihood Estimation (Table 5). For all these models, (i) standardized residual plots showed a random pattern, which confirmed model adequacy; (ii) Box-Ljung tests showed p-value>0.05, which upholds the white noise assumption; and (iii) the majority of the residuals on the QQ plots approximately fell along the QQ line which fulfilled the Normality assumption. Moreover, Histograms of the residuals proved zero mean assumption with normal distributional shape.

The time plots and ACF-PACF plots of the differenced series, p-values of the Box-Ljung tests, and model diagnostic plots are not reported here to conserve space and are available upon request.

3.2 ARIMA Forecasting

Based on these models, each of the 34 time series was forecasted for 2016-2031. The detailed age-sex-specific forecasts are reported in Table 6. For each year of the forecast horizon, the forecasts of the 34 series were summed to produce annual national population forecasts. By 2031, Bangladesh is expected to have a population of about 18.5 crore.

Estimated parameters of the selected ARIMA models

Table 5

Time Series	(p,d,q)	drift	ϕ_1	ϕ_2	ϕ_3	ϕ_4	θ_1	θ_2	θ_3	θ_4
Male 0-4	(1,2,0)		0.272*							
			(0.121)							
Male 5-9	(2,2,2)		-0.045	-0.494**			0.370**	0.850***		
			(0.171)	(0.177)			(0.123)	(0.108)		
Male 10-14	(1,2,0)		0.450***							
			(0.112)							
Male 15-19	(1,2,0)		0.294*							
			(0.118)							
Male 20-24	(2,2,1)		1.619***	-0.738***			-0.910***			
			(0.101)	(0.079)			(0.141)			
Male 25-29	(2,2,3)		1.230***	-0.628***			-0.297	0.476***	0.296*	
			(0.170)	(0.147)			(0.184)	(0.114)	(0.126)	
Male 30-34	(0,2,4)						0.816***	0.766***	0.685***	0.714***
							(0.089)	(0.130)	(0.130)	(0.102)
Male 35-39	(1,2,2)		0.558***				0.453**	0.444**		
			(0.145)				(0.152)	(0.145)		
Male 40-44	(1,2,0)		0.494***							
			(0.107)							
Male 45-49	(2,1,1)	-0.0001	1.671***	-0.788***			-0.394			
		(0.0001)	(0.119)	(0.110)			(0.186)			
Male 50-54	(2,1,1)	-0.0004**	1.716***	-0.814***			-0.406*			
		(0.0001)	(0.119)	(0.113)			(0.188)			
Male 55-59	(2,1,3)	-0.0005**	-0.061	0.362*			1.488***	1.358***	0.797***	
		(0.0001)	(0.155)	(0.149)			(0.126)	(0.172)	(0.117)	
Male 60-64	(2,2,2)		-0.240	-0.522**			0.560***	0.895***		
			(0.159)	(0.185)			(0.102)	(0.096)		
Male 65-69	(0,3,1)						-0.816***			
							(0.074)			
Male 70-74	(0,3,1)						-0.788***			
							(0.100)			
Male 75-79	(4,3,0)		0.909***	-0.848***	-0.848***	-0.815***				
			(0.073)	(0.087)	(0.082)	(0.070)				
Male 80+	(0,2,3)						-0.603***	-0.562***	0.760***	
							(0.082)	(0.099)	(0.102)	
Female 0-4	(1,2,0)		0.270*							
			(0.120)							
Female 5-9	(2,2,4)		-1.618***	-0.975***			2.614***	3.152***	2.000***	0.652***
			(0.021)	(0.020)			(0.132)	(0.298)	(0.285)	(0.120)
Female 10-14	(1,2,0)		0.508***							
			(0.107)							
Female 15-19	(1,2,0)		0.382***							
			(114)							
Female 20-24	(2,2,1)		1.610***	-0.731***			-0.894***			
			(0.108)	(0.083)			(0.138)			
Female 25-29	(2,2,0)		1.301***	-0.543***						
			(0.102)	(0.101)						
Female 30-34	(1,2,4)		0.349**				0.764***	0.749***	0.699**	0.495*
			(0.176)				(0.153)	(0.164)	(0.214)	(0.193)
Female 35-39	(2,2,0)		1.264***	-0.511***						
			(0.107)	(0.106)						
Female 40-44	(2,1,3)	-0.0005**	0.204	0.449**			1.535***	1.394***	0.764***	
		(0.0002)	(0.155)	(0.145)			(0.130)	(0.156)	(0.088)	
Female 45-49	(2,1,0)	0.028***	1.647***	-0.756***						
		(0.005)	(0.076)	(0.076)						

Female 50-54	(2,1,0)	0.028***	1.654***	-0.759***					
		(0.005)	(0.077)	(0.078)					
Female 55-59	(2,1,3)	0.028***	0.082	0.442**			1.521***	1.374***	0.795***
		(0.007)	(0.145)	(0.143)			(0.118)	(0.187)	(0.118)
Female 60-64	(1,2,3)		0.605***				-0.373***	0.342**	-0.969***
			(0.115)				(0.088)	(0.155)	(0.212)
Female 65-69	(0,3,1)						-0.770***		
							(0.085)		
Female 70-74	(0,3,1)						-0.646***		
							(0.172)		
Female 75-79	(4,3,0)		-0.948***	-0.877***	-0.882***	-0.858***			
			(0.069)	(0.089)	(0.082)	(0.064)			
Female 80+	(4,3,1)		-1.001***	-1.002***	-0.999***	-0.984***	-0.353**		
			(0.022)	(0.026)	(0.021)	(0.012)	(0.112)		

*** p<0.001, ** p<0.01, * p<0.05
standard errors in parentheses

Age-sex-specific population forecasts of Bangladesh for 2016-2031 (in thousands)

Table 6

Age Group	2016		2017		2018		2019		2020		2021	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
0-4	7752	7423	7712	7383	7672	7343	7632	7303	7592	7263	7552	7223
5-9	7918	7615	7815	7542	7712	7474	7608	7402	7504	7330	7400	7261
10-14	8373	8021	8334	7985	8293	7947	8250	7908	8208	7868	8165	7827
15-19	8299	7899	8352	7927	8406	7955	8459	7983	8512	8010	8566	8038
20-24	7755	7531	7820	7565	7892	7612	7970	7670	8052	7738	8138	7812
25-29	7211	7355	7269	7391	7339	7424	7421	7457	7507	7491	7594	7525
30-34	6712	6970	6782	7058	6840	7136	6891	7209	6941	7280	6992	7350
35-39	5978	6192	6101	6334	6219	6464	6334	6589	6447	6713	6558	6838
40-44	5243	5106	5371	5262	5510	5413	5658	5558	5814	5697	5980	5829
45-49	4606	4522	4672	4578	4730	4616	4783	4640	4831	4653	4875	4657
50-54	3873	3737	4024	3887	4165	4018	4299	4128	4426	4220	4547	4293
55-59	2773	2621	2935	2788	3093	2949	3240	3100	3380	3244	3511	3377
60-64	1744	1646	1813	1722	1874	1786	1938	1844	2009	1897	2080	1948
65-69	1447	1394	1437	1371	1423	1341	1406	1305	1384	1264	1358	1217
70-74	1135	1121	1152	1135	1168	1146	1183	1151	1197	1153	1210	1150
75-79	758	769	775	800	790	831	804	863	818	899	834	940
80+	789	885	829	959	861	1031	892	1099	924	1165	955	1278
Subtotal	82366	80807	83193	81687	83987	82486	84768	83209	85546	83885	86315	84563
Total	163173		164880		166473		167977		169431		170878	

Age Group	2022		2023		2024		2025		2026		2027	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
0-4	7512	7182	7472	7142	7432	7102	7512	7182	7472	7142	7432	7102
5-9	7295	7185	7190	7115	7085	7041	7295	7185	7190	7115	7085	7041
10-14	8123	7787	8080	7746	8037	7706	8123	7787	8080	7746	8037	7706
15-19	8619	8065	8672	8093	8726	8120	8619	8065	8672	8093	8726	8120
20-24	8226	7892	8314	7973	8403	8054	8226	7892	8314	7973	8403	8054
25-29	7677	7561	7756	7597	7832	7634	7677	7561	7756	7597	7832	7634
30-34	7042	7419	7093	7489	7143	7559	7042	7419	7093	7489	7143	7559
35-39	6670	6966	6780	7097	6891	7229	6670	6966	6780	7097	6891	7229
40-44	6154	5955	6338	6074	6533	6188	6154	5955	6338	6074	6533	6188
45-49	4915	4654	4953	4647	4988	4637	4915	4654	4953	4647	4988	4637
50-54	4662	4349	4770	4392	4874	4422	4662	4349	4770	4392	4874	4422
55-59	3633	3503	3747	3619	3852	3727	3633	3503	3747	3619	3852	3727
60-64	2150	1998	2224	2048	2300	2098	2150	1998	2224	2048	2300	2098
65-69	1329	1167	1295	1112	1258	1054	1329	1167	1295	1112	1258	1054
70-74	1222	1143	1232	1132	1242	1117	1222	1143	1232	1132	1242	1117
75-79	848	981	861	1022	873	1064	848	981	861	1022	873	1064
80+	987	1389	1018	1497	1050	1601	987	1389	1018	1497	1050	1601
Subtotal	87064	85196	87795	85795	88519	86353	87064	85196	87795	85795	88519	86353
Total	172260		173590		174872		176106		177380		178618	

Age Group	2028		2029		2030		2031	
	Male	Female	Male	Female	Male	Female	Male	Female
0-4	7272	6941	7232	6901	7192	6861	7272	6941
5-9	6662	6744	6556	6667	6449	6590	6662	6744
10-14	7866	7544	7823	7503	7780	7462	7866	7544
15-19	8939	8230	8992	8258	9046	8285	8939	8230
20-24	8742	8354	8824	8422	8904	8489	8742	8354
25-29	8137	7780	8216	7817	8295	7853	8137	7780
30-34	7345	7837	7396	7906	7446	7976	7345	7837
35-39	7333	7762	7444	7895	7554	8027	7333	7762
40-44	7434	6582	7696	6667	7975	6746	7434	6582
45-49	5104	4587	5129	4575	5152	4565	5104	4587
50-54	5237	4457	5316	4453	5391	4447	5237	4457
55-59	4200	4081	4269	4151	4334	4216	4200	4081
60-64	2621	2309	2706	2364	2793	2421	2621	2309
65-69	1069	806	1012	743	951	681	1069	806
70-74	1272	1019	1276	986	1280	951	1272	1019
75-79	922	1262	933	1315	943	1370	922	1262
80+	1176	2207	1207	2368	1239	2528	1176	2207
Subtotal	91331	88502	92027	88991	92724	89468	91331	88502
Total	179833		181018		182192		183462	

Forecast accuracy of different methods

Table 7

Methods	Forecasted population for 2022	APE compared to PHC 2022
Aggregated ARIMA	172.26	1.43
Cohort component method* (Hayes, 2015)	172.15	1.37
UN probabilistic projection* (UN DESA, 2017a)	173.23	2.00
Bayesian cohort component method* (Mahsin & Hossain, 2012)	166.77	1.80
Logistic (Verhulst) model (Haque et al., 2012)	174.30	2.63
two-parameter negative exponential model † (Islam & Hoque, 2014)	166.47	1.98
Combined forecasts	170.39	0.33

* Medium variant scenario of fertility
† With exponential growth rate method

The total population of Bangladesh was found to be 169.83 million in the population and housing census of 2022 (BBS, 2023). Compared to this, APE for aggregated ARIMA and cohort component method were found to be 1.43 and 1.37 percent, respectively. Whereas, the same was 0.33 percent for the combined forecasts.

4. DISCUSSIONS AND CONCLUSION

This study implemented a disaggregated approach compared to usual ARIMA population forecasts by considering 34 age-sex-specific population series. Following this approach, it was possible to construct a detailed picture of the age-sex dynamics of the Bangladeshi population till 2031. Moreover, for each year in the forecast horizon, forecasts of these 34 series were aggregated to produce national population forecasts (Table 6). The underlying notion was that this should be more accurate than usual overall forecasts due to the large volume of information going into it.

To probe this, the accuracy of aggregated ARIMA was compared to five other projection methods and combined forecasts of all these methods (Table 7). In the individual methods, we found that the traditional cohort component method and aggregated ARIMA perform best with only about 1.4 APE. The cohort component method usually outperforms univariate forecasting as it considers components of population dynamics individually. However, aggregated ARIMA improved accuracy in this aspect, possibly because of the pulling of a large amount of age-sex information. Table 7 also shows that simple (deterministic) trend extrapolation methods like logistic or exponential models were well outperformed by the complex (statistical) trend extrapolation method of ARIMA.

One key drawback of the ARIMA modeling is that it can only be used for short or medium-term forecasting as it does not accommodate the components of the population dynamics. However, its key benefit is its timely implementation with limited financial and data resources. Hence, the disaggregated ARIMA approach can be used to forecast local population composition and be integrated into short and medium-term local planning. Such well-informed local planning can ease the impact of the continued growth of Bangladesh's large population on its development course in a decentralized manner.

Another key finding of this study is that combined forecasts exhibited the lowest APE compared to all individual methods. This seems plausible, as each method and dataset were different and provided useful information to produce combined forecasts. Hence, it is recommended that forecasts from several sources be combined for higher accuracy when available.

Declarations

The author has no relevant financial or non-financial interests to disclose.

Data availability

The dataset analyzed in this study is freely available on the UN DESA World Population Prospects' download center.

References

1. Abdulgader, Q. M. (2016). Time series forecasting using ARIMA methodology with application on census data in Iraq. *Science Journal of University of Zakho*, *4*(2), 258–268.
2. Alho, J. M., & Spencer, B. D. (1997). The practical specification of the expected error of population forecasts. *JOURNAL OF OFFICIAL STATISTICS-STOCKHOLM-*, *13*, 203–226.
3. Armstrong, J. S. (2001). *Combining forecasts*. Springer.
4. Baxter, R., & Williams, I. (1978). Population forecasting and uncertainty at the national and local scale. *Progress in Planning*, *9*, 1–72.
5. BBS. (2023). *Population and Housing Census 2022* [National Volume 1]. Bangladesh Bureau of Statistics (BBS), Government of Bangladesh.
6. Beer, J. D. (1993). Forecast intervals of net migration: The case of the Netherlands. *Journal of Forecasting*, *12*(7), 585–599.
7. Bell, W. R. (1997). Comparing and assessing time series methods for forecasting age-specific fertility and mortality rates. *Journal of Official Statistics*, *13*(3), 279.

8. Bijak, J., Disney, G., Findlay, A. M., Forster, J. J., Smith, P. W., & Wiśniowski, A. (2019). Assessing time series models for forecasting international migration: Lessons from the United Kingdom. *Journal of Forecasting*, *38*(5), 470–487.
9. Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, *22*(3), 547–581.
10. Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *26*(2), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x
11. Box, G. E. P., & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day.
12. Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day.
13. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis: Forecasting and control* (Fifth edition). John Wiley & Sons, Inc.
14. Caleiro, A. (2010). Forecasting the number of births in Portugal. *Forecasting Demographic Components: Fertility*, 1–9.
15. Carter, L. R. (1996). Forecasting US mortality: A comparison of Box□Jenkins ARIMA and structural time series models. *Sociological Quarterly*, *37*(1), 127–144.
16. Cohen, J. E. (1986). Population forecasts and confidence intervals for Sweden: A comparison of model-based and empirical approaches. *Demography*, *23*, 105–126.
17. Cryer, J. D., & Chan, K. (2008). *Time series analysis: With applications in R* (2nd ed). Springer.
18. Dai, J., & Chen, S. (2019). The application of ARIMA model in forecasting population data. *Journal of Physics: Conference Series*, *1324*(1), 012100.
19. Debón, A., Montes, F., & Puig, F. (2008). Modelling and forecasting mortality in Spain. *European Journal of Operational Research*, *189*(3), 624–637.
20. Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 427–431.
21. Dickey, D. A., & Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*, 1057–1072.
22. Du, S., & Chan, C.-H. (2023). Baby Boom or Baby Bust After the COVID-19 Onset in the United States? Evidence from an ARIMA Time-Series Analysis. *Population Research and Policy Review*, *42*(6), 97. https://doi.org/10.1007/s11113-023-09843-6
23. George, M. V., Smith, S. K., Swanson, D. A., & Tayman, J. (Eds.). (2004). Population Projections. In *The methods and materials of demography* (2nd ed). Elsevier/Academic Press.
24. 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.
25. Hayes, G. (2015). Future Population Prospects and their Policy Implications. In *The Impact of the Demographic Transition on Socioeconomic Development in Bangladesh: Future Prospects and Implications for Public Policy* (pp. 97–122). United Nations Population Fund (UNFPA).
26. Hossain, M. A., Latif, A. H. M. M., Rayhan, M. I., Masud, M. S., Khan, M. H. R., Alam, M. I., & Awan, N. (2015). *Population Projection of Bangladesh: Dynamics and Trends 2011-2061* [Population Projection of Bangladesh]. Bangladesh Bureau of Statistics (BBS), Government of Bangladesh.
27. Hyndman, R. J. (2014, March 12). *Unit root tests and ARIMA models*. Rob J Hyndman. https://robjhyndman.com/hyndsight/unit-root-tests/

28. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.
29. Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and practice* (Third print edition). Otexts, Online Open-Access Textbooks.
30. Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The **forecast** Package for *R*. *Journal of Statistical Software*, *27*(3). https://doi.org/10.18637/jss.v027.i03
31. Islam, M. R., & Hoque, M. N. (2014). Mathematical modeling and projecting population of Bangladesh by age and sex from 2002 to 2031. In *Emerging techniques in applied demography* (pp. 53–60). Springer.
32. Keilman, N., & Pham, D. Q. (2000). Predictive intervals for age-specific fertility. *European Journal of Population/Revue Européenne de Démographie*, *16*, 41–65.
33. Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, *54*(1–3), 159–178.
34. Land, K. C. (1986). Methods for national population forecasts: A review. *Journal of the American Statistical Association*, *81*(396), 888–901.
35. Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, *87*(419), 659–671.
36. Lee, R. D., & Tuljapurkar, S. (1994). Stochastic population forecasts for the United States: Beyond high, medium, and low. *Journal of the American Statistical Association*, *89*(428), 1175–1189.
37. Mahsin, M., & Hossain, S. S. (2012). Population forecasts for Bangladesh, using a bayesian methodology. *Journal of Health, Population, and Nutrition*, *30*(4), 456.
38. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, *38*(4), 1346–1364.
39. McNown, R., & Rogers, A. (1989). Forecasting mortality: A parameterized time series approach. *Demography*, *26*(4), 645–660.
40. Miller, R. B., & McKenzie, S. K. (1984). *Time Series Modeling of Monthly General Fertility Rates*. US Bureau of the Census.
41. Pflaumer, P. (1992). Forecasting US population totals with the Box-Jenkins approach. *International Journal of Forecasting*, *8*(3), 329–338.
42. Raymer, J., & Wiilekens, F. (2008). *International migration in Europe: Data, models and estimates*. John Wiley & Sons.
43. Saboia, J. L. (1974). Modeling and forecasting populations by time series: The Swedish case. *Demography*, *11*(3), 483–492.
44. Sarpong, S. A. (2013). Modeling and forecasting maternal mortality; an application of ARIMA models. *International Journal of Applied*, *3*(1), 19–28.
45. Smith, S. K., Tayman, J., & Swanson, D. A. (2013). *A practitioner's guide to state and local population projections*. Springer.
46. Tayman, J., Smith, S. K., & Lin, J. (2007). Precision, bias, and uncertainty for state population forecasts: An exploratory analysis of time series models. *Population Research and Policy Review*, *26*, 347–369.
47. UN DESA. (2017a). *World Population Prospects 2017, Online Edition* [dataset].
48. UN DESA. (2017b). *World Population Prospects: Methodology of the United Nations Population Estimates and Projections*. United Nations, Department of Economic and Social Affairs, Population Division.
49. UN DESA. (2022). *World Population Prospects 2022, Online Edition* [dataset].
50. Wei, W. S. (2006). *Time Series Analysis- Univariate and Multivariate Methods* (2nd ed.). Pearson.

51. Wilson, T., & Rees, P. (2005). Recent developments in population projection methodology: A review. *Population, Space and Place*, *11*(5), 337–360.
52. World Bank. (2023). *World Development Indicators* [dataset]. The World Bank. https://doi.org/10.57966/6RWY-0B07
53. Zakria, M., & Muhammad, F. (2009). Forecasting the population of Pakistan using ARIMA models. *Pakistan Journal of Agricultural Sciences*, *46*(3), 214–223.