ARIMA Model for Birth Rate Forecasting in Azerbaijan

Makrufa Hajirahimova
Institute of Information Technology of the Ministry of Science and Education of the Republic of Azerbaijan
Baku, Azerbaijan
hmakrufav@gmail.com

Aybeniz Aliyeva
Institute of Information Technology of the Ministry of Science and Education of the Republic of Azerbaijan
Baku, Azerbaijan
aliyeva.a.s@mail.ru

***Abstract*—The purpose of this study is to evaluate Autoregressive Integrated Moving Average (ARIMA) model ability to forecast the births rate in Azerbaijan. In the analysis, the Box-Jenkins methodology was followed when building the suggested model. Yearly data from 1990 to 2023 was collected from the website of the Statistics Committee of the Republic of Azerbaijan. Python programming languages and its Scikit Learn library were used for the analysis of the data. Besides, Akaike's information criterion (AIC) and Bayesian Information Criteria (BIC) are used to select the best ARIMA model, compared to another estimated models. The prediction results of the models are evaluated using the mean absolute percentage error (MAPE) and the root mean square error (RMSE) . Comparing the predicted data from the ARIMA models shows that the correct selection of model parameters, it possible to fairly accurately predict the births rates. Obtained predictions reflect the dynamic of the births, and can be useful for the demographers and government officials.**

Keywords—time series forecasting, Birth Rate, Box-Jenkins method, ARIMA model, MAPE, RMSE

I. INTRODUCTION

Recently, the decline in the birth rate has caused serious concern at the global level. In many developed countries, the decline in population growth leads to a demographic crisis. Thus, with the decline in the population, the number of elderly people increases, and the number of working-age young people decreases. This leads to a shortage of labor, the collapse of the pension system, and an economic slowdown. The decline in working-age people reduces economic production and tax revenues. This complicates the sustainability of social security systems. As a result of the decline in human resources, changes in social dynamics occur.

Currently, Azerbaijan is among the countries facing a decline in the birth rate. Compared to the previous year, the birth rate per thousand people has decreased from 11.2 to 10.1. If this trend continues, the population decline in the future may also have a negative impact on the demographic and economic development of our country.

Thus, the decline in the birth rate leads to serious problems both globally and locally. If this problem is not solved in a timely manner, serious difficulties may arise in the social, economic and cultural spheres.

Forecasting birth rates for the near future allows us to observe changes that will occur in the age structure of the population. These interpretations can guide policymakers to focus on socio-economic development and comprehensive healthcare system strengthening as crucial strategies for raising the birth level.

II. LITERATURE REVIEW

Birth and death forecasts can be made (produced) using univariate or multivariate statistical time series methods, structural models or machine learning (ML) methods [1]. The ARIMA model is widely used in the forecasting of birth time series, along with ML methods.

The ARIMA models and AR models are used for forecast the Total Fertility Rate (TFR) in Malaysia and the forecasting performance of these models is evaluated by using post sample forecasting accuracy criterion. It was found that the AR model could be the most appropriate model for forecasting the TFR in Malaysia [2].

D. Lancheros-Cuestaet al. [3] tested the application of an ARIMA forecasting model to obtain future estimates for birth and death data in Colombia. The analyzed data has been sorted as national and by departments as well. Close approximations in the years of study were obtained from the general data. The actual birth and death data for 2016 for eleven departments were compared with the ARIMA results. This model was found to fit properly with the data used in this study and show reliable prediction with low percentage error. Also, the best strategy for analyzing time series is to have a large amount of specific data. When the data is generalized, errors are not taken into account, and if there are big errors, the model may tend to be inaccurate.

Y.Wang [4] applied the ARIMA and the ETS models to forecast the short-term birth rate in China in the next five years. The historical fertility rate data in China from 1964 to 2021 used in study. From the Results shown that the time series plots of these two models have similar trends. Author found that the ARIMA model was more appropriate for predicting the short-term birth rate in China. As a result, was utilized the ARIMA(0,0,1) model for prediction.

B. A. Masha et al. [5] examined Ghana's birth rate using the Box- Jenkins technique on a monthly basis. Monthly data ranging from January, 2014 through to December 2019 was collected from Ghana Statistical Service. They discovered that the ARIMA (1, 1, 1) model had the best fit, with the lowest normalized BIC and AIC values among the tested models. This model was applied to the 12-month birth rate for the year 2020.

III. Materials and Methods

A. Dataset

In this study, we use demographic data for Azerbaijan comprising data on live births and deaths (non-fetal deaths) broken down by years provided by Statistics Committee of the Republic of Azerbaijan. This information is obtained based on monthly statistical elaboration of birth statements compiled by registration Office of the Ministry of Justice. The data is located on the website of the statistical committee and the website gives the option to choose which data to download by year [14]. The required data for the study, (births and deaths data ranging from January, 1990 through to December 2023) was downloaded from the website of Statistics Committee. A database is created in Excel with all the downloaded information from the website, sorting the data by year and births rate. With this information, the necessary conditions for the model to forecast accurate predictions are determined. The univariate ARIMA model was applied to predict the births rate for the next sevsn years. The data analysis has been performed using Python programming languages and its Scikit Learn library.

B. ARIMA model

ARIMA is a statistical model used for analyzing and forecasting time series data. It also popularly is known as the Box-Jenkins method [15]. The ARIMA is approach providing a flexible and structured way to model time series data based on its prior observations as well as past prediction errors. ARIMA models are, the most general class of models for forecasting a time series which can be made to be “stationary” [16]. A stationary series has no trend (increase or decrease), its fluctuations around its mean value have a constant amplitude, and it moves in a consistent manner. The ARIMA forecasting equation is a regression-type equation in which the predictors consist of lags of the dependent variable or lagged forecast errors. An ARIMA model is classified as an "ARIMA (p, d, q)" model, where:

- p is the number of autoregressive (AR) terms, also known as the lag order.
- d is the minimum number of differences needed for stationarity. Also, it known as the degree of differencing. If the time series is stationary, then d = 0.
- and q is the size of the ’moving average’ (MA) window. q is the number of lagged forecast errors in the prediction equation.

The steps (stages) of the ARIMA model building methodology are presented in Fig. 2.

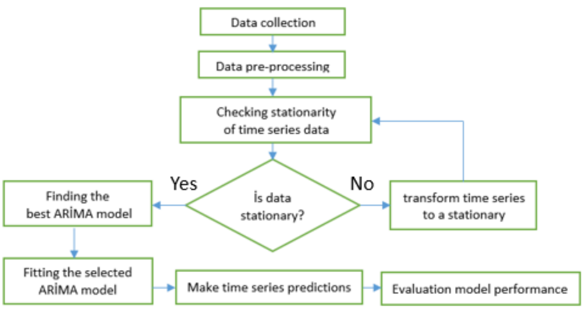

Fig. 1. The stages of the ARIMA model building methodology.

In the analysis, the Box-Jenkins methodology followed when building an ARIMA model. This methodology consists of four steps. The first step is called identification, and the aim of this step is to make sure that the series is stationary and identifying appropriate lags for the components of the AR and MA. The Auto Correlation Function (ACF) and the Partial Auto Correlation Function (PACF) are used to find out the appropriate values of p, d, q in the first step, and determine the best ARİMA model.

The second step in the analysis is to estimate of the parameters in the model. The process of least-squares estimation is usually used for this [15].

The next (third) step is diagnostic testing, which check residuals of the fitted model to determine the adequacy of the chosen ARIMA model. If the residuals are not white noise, then the first, second, and third steps are repeated again using the new values for p, d, and q [17].

Forecasting is the fourth step where the ARIMA time series models may be used to predict future values. The best ARIMA model was selected based on minimal AIC and BIC values. The Ljung–Box test was used to evaluate the stability of the model.

The performance of the models in terms of predicting the uncertainty of future births and deaths is determined by the 95% prediction intervals associated with each set of forecasts. If the deviations from the 95% prediction interval are too large, the model is considered to fail in terms of predicting the uncertainty.

The performance of the learning models is measured using the value of the MAPE and RMSE. MAPE expresses the average absolute percent difference between actual and predicted values. The RMSE measure is defined as the square root of mean square error. Lower values of these measures generally signify better model accuracy[18].

II. Results and Discussions

In In this section, experiments are conducted and their results are summarized. Our target is to build an ARIMA model that can predict the number of births of deaths for the next 7 years (2024-2030 years) in Azerbaijan.

The dataset used in our study is based on yearly births rate in Azerbaijan. This data was plotted on a graph, as shown in Figure 2. The births dataset is divided into training (85%) and testing (15%).

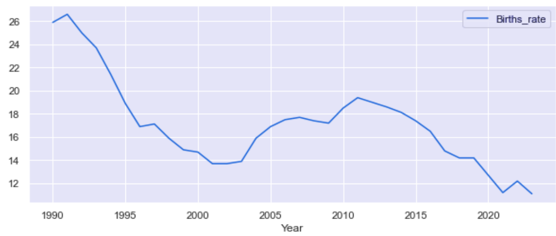

Fig. 2. Time series for births rate

ARIMA models usually applied on stationary time series. For this, we first check the stationarity of the time series. The time series presents a non-stationary characteristic in Figure 1. The ACF and PACF measurements for stationarity also indicate that the time series is not stationary as shown in Figure 3 and Figure 4.

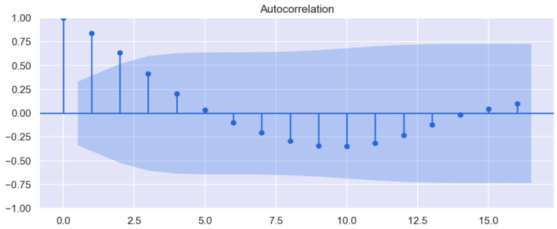

Fig. 3. ACF before differencing

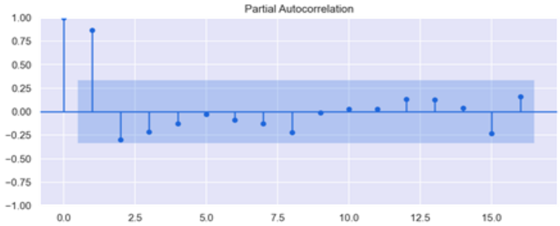

Fig 4. PACF before differencing

In order to investigate the stationarity of time series, in this study, Augmented Dickey -Fuller (ADF) test was applied. It was implemented in Python. The following ADF test results obtained in Python also show that the data are not stationary.

```
In [50]: def test_stationarity(timeseries):
             result = adfuller(timeseries)
             print('ADF Statistic:', result[0])
             print('p-value:', result[1])
             if result[1] <= 0.05:
                 print("Stationary")
             else:
                 print("Not Stationary")

         test_stationarity(df['Births_rate'])

         ADF Statistic: -2.702819525241609
         p-value: 0.07355163756626018
         Not Stationary
```

After having one time difference to time series, the new times series is more stationary as shown in Figure 5. The ACF and PACF for the new time series prove its stationarity. The following results obtained in Python also show that the data are stationary. Consequently, d for ARIMA will equal 1 since the time series become stationary after the first difference.

```
In [51]: #first order differencing
         df_diff = df.diff().dropna()
         test_stationarity(df_diff['Births_rate'])

         ADF Statistic: -3.4685301884089825
         p-value: 0.008833058475173199
         Stationary
```

As seen that the p-value of ADF test is less than 0.05, it indicates, that the birth data is stationary, so there is no need for differencing. Note that, differencing is required for stationarity of time series when the p-value is greater than 0.05.

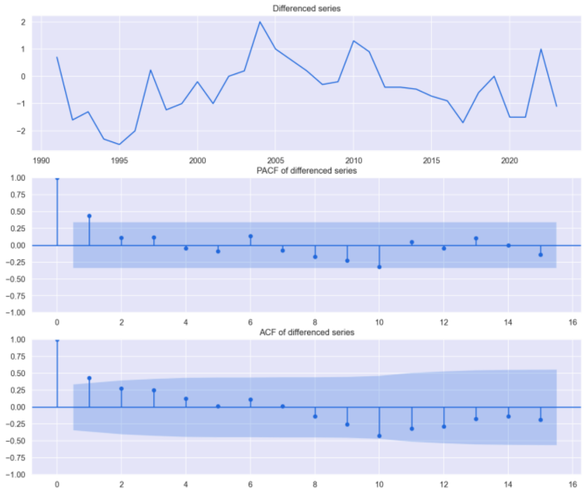

Fig. 5. ACF (a) and PACF (b) plot for births time series data.

Figure 4 Time series, ACF and PACF for births data after differencing

The best suitable ARIMA model for birth time series was selected based on minimal AIC and BİC values. In this study, the least AIC value is 96.502, and BIC value is 99.378 (as given in Table II), and the optimal model is ARIMA (1 ,1 ,0) with the lowest MAPE values of 10.2% and RMSE values of 1.157

TABLE II. CHOOSING THE BEST ARIMA MODEL FOR THE BIRTHS DATA

Model	AIC	BIC	MAPE	RMSE
(1,1,1)	98.187	102.116	0.107	1.410
(0,1,0)	101.565	106.486	0.166	2.229
(1,1,0)	**96.502**	**99.378**	**0.102**	**1.157**
(0,1,1))	97.648	101.885	0.171	2.271
(2,1,0,)	98.310	102.373	0.124	1.3791
(2,1,1)	100.172	105.604	0.127	1.411

After identification the value of p, d and q, the next step in the Box-Jenkins methodology is to estimate the parameters. Accordingly, the parameters of model have been estimated and presented in Table III.

TABLE III. PARAMETER ESTIMATE FOR THE ARIMA (1, 1 0) MODEL

Variable	Coefficient	Stand. Error	t-statistic	p-value
AR(1)	0.525	0.155	-6.502	0.001

Investigating the results of these estimates shows that the AR for ARIMA(1,1, 0) model is statistically significantly different from zero (less than 0.05), the test statistic is statistically insignificant and the residuals of selected model conforms to the characteristics of white noise process. Therefore, this model may be used to forecast the births. A visual demonstration of the fitted values of the ARIMA (1,1, 0) model, and actual values for births data are illustrated in Fig. 4.

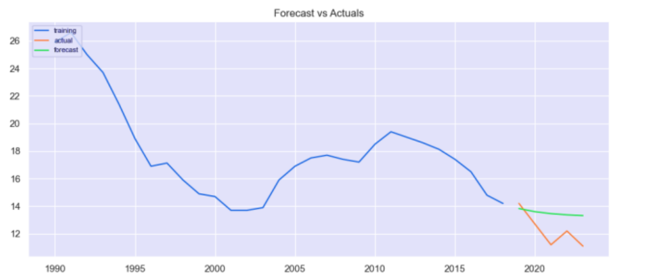

Fig 6. Actual values and the fitted values of the ARIMA (1,1, 0) model for births data.

As can be seen from Fig 4. that the ARIMA (1,1,0) model analyzed in terms of predicting the uncertainty of future births performs well. Here, there is a high probability that actual birth observations will be within the 95% prediction intervals. In Table 4, the forecasted values of ARIMA (1, 1, 0) model and the actual values were presented for each year.

TABLE IV. THE PREDICTED VALUES OF ARIMA (1, 1, 0) MODEL AND THE ACTUAL VALUES

Year	Actual	Predicted number of births
2019	14.2	13.91
2020	11.2	13.60
2021	11.2	13.46
2022	12.2	13.37
2023	11.1	13.32

So, ARIMA (1,1,0) model is determined to be the best fit model for the birth data. The prediction of births in Azerbaijan over the next 8 years using the ARIMA (1,1,0) model is calculated, and illustrated in Fig. 7.

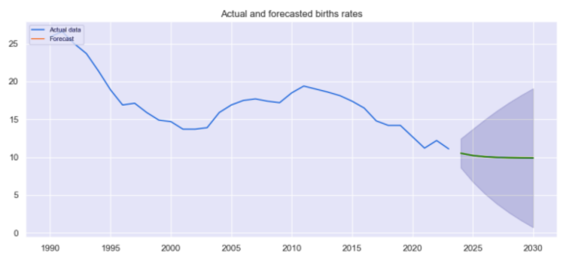

Fig. 7. Prediction the number of births for the next years.

Table V shows the forecasting of births for the next years in Azerbaijan (from 2024 to 2030). As can be seen from the Fig. 7 and the Table 5, the number of births in the country is decrease.

TABLE V. FORECASTING THE BIRTHS RATE FOR THE NEXT EIGHT YEARS IN AZERBAIJAN

Predicted date	Forecasted value of births (fc_series)	Minimum prediction (lower-series)	Maximum prediction (upper_series)
2024	10.522	8.609	12.435
2025	10.219	6.729	13.708
2026	10.059	5.154	14.964
2027	9.975	3.817	16.134
2028	9.931	2.658	17.205
2029	9.908	1.633	18.183
2030	9.896	0.711	19.081

IV. CONCLUSION

The proposed models fit properly with the data used in this research, and show a valid prediction that present a low percentage error. Therefore, it can be considered reliable for decision making aimed at timely prevention of the demographic challenges that the country may face. Thus, by effectively using the ARIMA model, reliable predictions of birth rates for the near future can be obtained, which can be useful for demographers and government officials, especially health and social sector officials, in timely and systematic management of child health indicators, providing financial support to families, determining the sustainability of pension and social security systems, and strategic planning.

ARIMA models are not explicitly designed to capture seasonal patterns in the data, which can lead to inaccuracies in forecasting for datasets with significant seasonal variations. For this purpose, it is planned to use the SARIMA model and machine learning methods in future research.

REFERENCES

[1] S. E. Vollset., E. Goren, Ch.-W. Yuan, J. Cao et al., “Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for the Global Burden of Disease Study,” Lancet, 2020, vol. 396, pp. 1285–1306. https://doi.org/10.1016/S0140-6736(20)30677-2

[2] M. Shitan, “Forecasting the Total Fertility Rate in Malaysia,” Pakistan Journal of Statistics, 2015, vol. 31, no. 5, pp. 547-556.

[3] D. Lancheros-Cuesta, C. D. Bermudez, S. Bermudez, and G. Marulanda, “Predictive Model of Births and Deaths with ARIMA,” Computer Aided Systems Theory – EUROCAST 2019: 17th International Conference, Las Palmas de Gran Canaria, Spain, Revised Selected Papers, February 2019, Part I, pp. 52–58. https://doi.org/10.1007/978-3-030-45093-9_7

[4] Y. Wang, “Analysis of China’s Birth Rate Prediction Based on Time Series,” Proceedings of the 2nd International Conference on Financial Technology and Business Analysis, 2023, pp, 205-217. DOI: 10.54254/2754-1169/55/20231009

[5] B. A. Masha, Z. D. Delali, and A.W. Reuben, “Time Series Analysis of Monthly Birth Rate in Ghana,” International Journal of Applied Science and Mathematics, 2020, vol. 7, no. 5, pp. 2394-2894.

[6] State Statistical Committee of the Republic of Azerbaijan: Demographic indicators of Azerbaijan, Statistical yearbook, 2023, Baku, stat.gov.az

[7] F. M. Khan, R. Gupta, “ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India,” Journal of Safety Science and Resilience, 2020, vol. 1, pp. 12–18. https://doi.org/10.1016/j.jnlssr.2020.06.007

[8] M. M. Hassan, T. Mirza, "Using Time Series Forecasting for Analysis of GDP Growth in India," International Journal of Education and Management Engineering (IJEME), 2021, vol. 11, no. 3, pp. 40-49. 2021. DOI: 10.5815/ijeme.2021.03.05

[9] G. C. Tiao, Time Series: “ARIMA Methods, in International Encyclopedia of the Social & Behavioral Sciences,” 2001, https://www.sciencedirect.com/economics-econometrics-and-finance/arma-model

[10] M. Sh.Hajirahimova, A.S. Aliyeva, “Development of a Prediction Model on Demographic Indicators based on Machine Learning Methods: Azerbaijan Example,” International Journal of Education and Management Engineering (IJEME), 2023, vol. 13, no. 2, pp. 1-9. https://doi.org/10.5815/ijeme.2023.02.01