Forecasting Demographic Data of Pakistan: A Comparative Study of Time Series Models & Population Projection Methods

Amjad Ali[*],Raza Ahmed Khan[†], Dost Muhammad Khan[‡]

Abstract

This study is conducted to forecast the population of Pakistan based on different demographic variables that significantly influences the population of a country. For this purpose, annual data have been taken from Pakistan Bureau of Statistics and official website of World Bank for a period of 65 years, i.e. from 1960-2015. The data was analyzed using ARIMA (Auto Regressive Integrated Moving Average) models and deterministic methods for different demographic indicators of Pakistan. A comparison of the two methods has been made to assess the accuracy of their population projections. It was found using RMSE (Root Mean Square Error), MAE (Mean Absolute Error) and AIC (Akaike Information Criterion) that ARIMA perform better than Simple Exponential Smoothing (SES) method in forecasting the population of Pakistan. Therefore, it was concluded that ARIMA model is useful and is recommended to be used for forecasting the population of Pakistan.

***Keywords*:** demographic variables, *ARIMA* model, simple exponential smoothing method and projection methods

Introduction

Population is the total number of people living in a country, city or any particular locality (United Nations, 1995). It can be considered as an assembly of a specific type of organism living in a given area. Most of countries across the world conduct census after certain time period. The population census provides information about the complete infrastructure and size of the population during the past census years. Census data provides a detailed insight into industry needs, health care, food requirements, educational needs and the list goes on. In this regard developed countries are ahead from the developing nations because of having accurate and updated data at their disposal at the time of need.

the demographic variables such as age, gender and size of populations, occupation, income level etc. Past data show that Pakistan's population is increasing with disturbing population size of 34 million in 1951 to 207 million in 2017 with the average growth rates of 3.66%, 3.05%, 2.69% and 2.40% for the periods 1961-72, 1972-81, 1981-98, and 1998-2017 respectively. Demographic factors play a sensible role in the

[*] Department of Statistics, Islamia College, Peshawar

[†] M.Phil Scholar ,Department of Statistics Islamia College Peshawar

[‡] Assistant Professor , Department of Statistics, Abdul Wali Khan University Mardan, KP, Pakistan, Email: dostmuhammad@awkum.edu.pk

development of a country and its future planning. Therefore, it is important to know about their current trends and future values. In past studies, different methodologies and mechanisms have been used to project the demographic factors. The well-known Box-Jenkins method of time series forecasting has been used by many researchers for forecasting of demographic factors (Saboia, 1977; Land and Cantor, 1983; Carter and Lee,1996; McNown and Rogers, 1989; Knudsen et al., 1993). McDonald (1979) discussed the relationships between classical demographic deterministic forecasting models, stochastic structural econometric models and time series models to forecast the Australian total live births. The purpose of the study is to focus on transforming time series data to stationary and analyzed the properties of the models. Land and Cantor (1983) applied seasonal time series model (SARIMA) to model the U.S. monthly birth and death rates.

The well-known Box-Jenkins (1970) methodology has been used for model development in conjunction with Nerlove et al. (2014) suggestions. A second order autoregressive model with seasonal moving-average component has finally been selected for both the series. Weekly periodicities have also been found in the birth rate series. McNown and Rogers (1989) combined parameterized model mortality schedules with time series models for forecasting the U.S. mortality rate to the year 2000 by utilizing the data from 1900 to 1985. In this study the results showed that the modified time series model has good performance and absorb the shocks up to long term and enhance the posterior information of mortality rates. Lawarence (1996) made a comparative study of Box-Jenkins *ARIMA* and structural time series analysis to forecast demographic processes. They used Lee-Carter method to develop nonlinear demographic models for different demographic factors of U.S. such as total population, mortality rates, gender, race and gender combined. Their findings revealed that there were marginal differences in the performance of these two methods with a slight advantage to structural models. Zakria and Muhammad (2009) used Box-Jenkins *ARIMA* model and forecast the growth rate of Pakistan utilized census for the period of 1951 to 2007.They made forecasts for the next twenty years. In this research their findings suggest that *ARIMA (1, 2,0)* model found to be better and suitable model for projection. Zakria and Muhammad (2015) utilized Markov Chain for projection of the population age distribution for the next 40 years on available census data from 1972 to 1981. Similarly, Shitan and Lerd (2015) used A*RIMA* and *ARAR* models to forecast the total fertility rate in Malaysia. The post

sample forecast accuracy of these models have been compared. Their results showed that *ARAR* models performed better than *ARIMA* models.

In above mentioned studies, various methodologies and models have been used to model and forecast demographic factors of different countries. In the Pakistani context, Zakria and Muhammad (2015) used time series models to forecast a number of demographic factors. The main focus of the present study is to develop time series models for some of demographic factors of Pakistan and compare the forecasting ability of these models with different population projection methods. These factors include crude death rate, adult male mortality rate, adult female mortality rate, infant mortality rate, crude birth rate, adolescent fertility rate.

Problem statement: Demographic changes affect all areas of human life, e.g. economy, society, culture and even politics. Therefore, the demographic factors have an important role in policy decision making and future growth of any country. The demographic factors including adult male mortality rate, adult female mortality rate and infant mortality rate of Pakistan have been focused in this study. Thus, the main objective of this study is to develop time series models for the above mentioned demographic factors and to compare their forecasting performance with different population projection methods.

Research objectives

- To develop time series models for the understudy demographic factors of Pakistan
- To forecast Pakistan's, the demographic factors understudy through time series models and various population projection methods.
- To compare the forecast accuracy of time series models and population projection methods for understudy demographic factors.

Research methodology

Data for demographic variables including infant mortality rate, adult male mortality rate, adult female mortality rate, crude birth rate, crude death rate and adolescent fertility rate was collected from Pakistan Bureau of Statistics and official website of World Bank from 1960 to 2015. For the development of a time series models (*ARIMA*), data from 1960 to 2010 were used. Outcomes for 2011-2015 were forecasted through time series (ARIMA & SES) models and also projected through population projection methods (athematic, geometric & incremental projection methods). The forecast accuracy of time series models and

population projection methods have been compared through different forecast accuracy measures i.e. (RMSE, MSE & MAPE). For data analysis statistical software's: gretl, R - package and MS-excel 2017 have been used.

Results and Discussion

To achieve the objectives of the study different statistical methods and models have been used. These results are discussed and summarized in this section. The descriptive statistics for all variables under study have been summarized in Table-I. The data have 56 observations from 1960 to 2015. The first variable of interest is adult female mortality rate which has a downward decreasing trend. This trend shows that the data is non-stationary. As per the results of ADF test the series becomes stationary at first difference. According to AIC, HQC and SBIC results shown in Table-II, *ARIMA (1,1,0)* model has been selected for adult female mortality rate. As mentioned in the methodology section, data from 1960 to 2010 has been used for model development purpose and the remaining five values have been forecasted to check the forecast ability of each method. Therefore *ARIMA (1,1,0)* model along with other population projection methods has been used to forecast the last five values of the series. These results have been summarized in Table-3. The results of Table-3 show that *ARIMA (1,1,0)* provides precise results as compared to other methods for adult female mortality rate.

Table-I: Descriptive Statistics of Demographic Variables under Study

Demographic Variable	Mean	Minimum	Maximum	St.Dev
Adult Female Mortality Rate	231.063	152.708	371.183	62.9872
Adult Male Mortality Rate	243.292	185.625	353.771	46.6438
Infant Mortality Rate	117.218	73.5	192	30.6972

Table-II: Time Series Models Demographic Variables under Study

Demographic Variable	MODELS	MAE	RMSE	MAPE
Adult Female Mortality Rate	ARIMA(1,1,0)	0.435921	0.242626	0.109168
	ARITHMETIC	50.53637	39.4816727	0.275934303
	GEOMETRIC	32.05839	27.2100142	0.189693281
	INCREMENTAL	19.62137	16.8712070	0.117543803
	SES(ETS)	0.561934	0.213073	0.001323
Adult Male Mortality Rate	ARIMA(1,1,0)	0.470113	0.210632	0.0852559
	ARITHMETIC	24.01375	18.21384	0.101606
	GEOMETRIC	14.1015	10.38969	0.057977
	INCREMENTAL	6.078318	4.426349	0.024674
	SES(ETS)	0.390000	0.150000	0.01223
Infant Mortality Rate	ARIMA (2,2,0)	0.094944	0.0744126	0.0648326
	ARITHMETIC	15.49503	12.09	68.17992
	GEOMETRIC	33.25104	26.84451	55.32793
	INCREMENTAL	2.055383	1.879425	82.73235
	SES(ETS)	0.561934	0.213073	0.001323

Table-III: Forecast Performance of Time Series Models and Population

Demographic Variable	ARIMA -MODELS	AIC	HQC	**SBIC**
Adult Female Mortality Rate	ARIMA(1,1,0)	-1.62137	-1.6069	-1.5835
	ARIMA(2,0,0)	-1.57814	-1.54919	-1.50238
	ARIMA(0,2,0)	-1.56378	-1.56378	-1.56378
	ARIMA(1,1,1)	-1.55783	-1.52888	-1.48207
	ARIMA(2,1,0)	-1.55538	-1.52643	-1.47962
Adult Male Mortality Rate	ARIMA(1,1,0)	-1.47035	-1.45587	-1.43247
	ARIMA(0,2,0)	-1.42417	-1.42417	-1.42417
	ARIMA(2,0,0)	-1.41852	-1.38957	-1.34276
	ARIMA(1,1,1)	-1.40676	-1.37781	-1.331
	ARIMA(2,1,0)	-1.40421	-1.37526	-1.32845
Infant Mortality Rate	ARIMA(2,2,0)	-4.63202	-4.60325	-4.55697
	ARIMA(1,2,2)	-4.62845	-4.5853	-4.51588
	ARIMA(2,2,1)	-4.58472	-4.54156	-4.47215
	ARIMA(2,2,2)	-4.5697	-4.51216	-4.41961
	ARIMA(1,2,1)	-4.37185	-4.34307	-4.2968

Projection Methods

The second variable of interest is the adult male mortality rate. The descriptive statistics have shown in Table-I for adult male mortality rate. It has a minimum value of 185.625 for the year 1960 and a maximum value of 353.771 for the year 215. The series also shows a downward decreasing trend. ADF has been applied to the check the stationarity of the series and we found that the series becomes stationary after taking the first difference. Therefore, *ARIMA (1,1,0)* has been selected as the final model for adult male mortality rate. The last five values of this series have also been forecasted by using *ARIMA (1,1,0)* and different population projection methods. The results are displayed in the Table-3. It can be seen that *ARIMA (1,1,0*) has better performance as compared to other methods for adult male mortality rate.
On the same lines we select *ARIMA (2,2,0)* model for infant mortality rate. The time series *ARIMA (2,2,0)* model has good forecasting performance as compared to population projection methods for this series.

Conclusion

Findings of the current study indicated that data for the selected demographic variables has stochastic trend which has been detected by ADF- unit root test and have been made stationary with their respective order of differences. Results showed that *ARIMA* model has better forecast performance as compared to other population projection methods for all variables based on the three criteria mentioned. Additionally, the projection methods including arithmetic projection method, geometric projection method, and incremental increase projection methods were not found to be very efficient for forecasting. Moreover, this study concluded that for Pakistani context, *ARIMA* models provide a more efficient modelling for above mentioned demographic factors that can be better utilized for future planning and forecasting.

References:

Box, G. E., & Jenkins, G. M. (1970). Time series analysis, control, and forecasting. *San Francisco, CA: Holden Day*.

Carter, L. R. (1996). Forecasting US mortality: A comparison of Box-Jenkins ARIMA and structural time series models. *The Sociological Quarterly*, *37*(1), 127-144.

Knudsen, C., McNown, R., & Rogers, A. (1993). Forecasting fertility: An application of time series methods to parameterized model schedules. *Social Science Research*, *22*(1), 1-23.

Land, K. C., & Cantor, D. (1983). ARIMA models of seasonal variation in US birth and death rates.*Demography*, *20*(4), 541-568.

McDonald, J. (1979). A time series approach to forecasting Australian total live-births. *Demography*, *16*(4), 575-601.

McNown, R., & Rogers, A. (1989). Forecasting mortality: A parameterized time series approach. *Demography*, *26*(4), 645-660.

Nerlove, M., Grether, D. M., & Carvalho, J. L. (2014). *Analysis of economic time series: a synthesis*. Academic Press.

Saboia, J. L. M. (1977). Autoregressive integrated moving average (ARIMA) models for birth forecasting. *Journal of the American Statistical Association*, *72*(358), 264-270.

Shitan, M., & Ng, Y. L. (2015). Pak. J. Statist. 2015 Vol. 31 (5), 547-556 Forecasting The Total Fertility Rate In Malaysia. *Pak. J. Statist*, *31*(5), 547-556.

United Nations. Population Division. (1995). *World population prospects: the 1994 revision* (Vol. 145). United Nations.

Zakria, M., & Muhammad, F. (2009). Forecasting the population of Pakistan using ARIMA models. *Pak. J. Agri. Sci*, *46*(3), 214-223.

Zakria, M., & Muhammad, F. (2015). Projecting The Population Age Distribution Of Pakistan Using Markov Chain. *Pakistan Journal of Statistics*, *31*(2).