Journal of Emerging Data Intelligence (JEDI)
Homepage https://journal.vu.edu.pk/JEDI
ISSN No | (Online): 3106-924X| (Print): 3106-9231

Demographic Projection: Navigating The Future of Pakistan with Box-Jenkins ARIMA Forecasting

Muhammad Rafique Daudpoto[1], Waseem Khoso[2], Ghulam Qadir Abro[3]

1 Department of Statistics, University of Sindh, Jamshoro, Sindh, Pakistan; Email: doudpoto.rafique@usindh.edu.pk

2 Department of Statistics, University of Sindh, Jamshoro, Sindh, Pakistan; Email: waseemkhoso108@gmail.com

3 Department of Statistics, Quaid i Azam university, Islamabad, Pakistan; Email: gqadir@stat.qau.edu.pk

**Correspondence: doudpoto.rafique@usindh.edu.pk*

Abstract

This research work diligently explores the global issue of overpopulation with particular emphasis on Pakistan, where rapid demographic growth presents daunting development challenges. Using the Box-Jenkins ARIMA methodology, an ARIMA (1, 1, 2) model is used to model and project the population trajectory of Pakistan for the next three decades. Challenges in making the data stationarity notwithstanding, a thoughtful analysis using correlogram tests and logarithmic transformations managed to stabilize the data and residuals. The ARIMA (1, 1, 2) model forecasts a population of around 325.9 million in the year 2050. This really underpins why the government of Pakistan needs to be quite ahead in addressing the challenges posed by such rapid population growth. High rates of unemployment, non-abating poverty, and the increasing menace of criminality are certain pressing issues that will get further escalated because of the demographic leap foreseen. This professional inquiry underlines actionable insights and calls for strategic policymaking to negotiate through the complex landscape of population dynamics and further emphasizes that proactive governance is requisite for further national development.

Keywords: Population Forecasting, ARIMA Models, Box-Jenkins Methodology, Time Series Analysis, Model Evaluation.

1. Introduction and Literature Review

According to Tartiyus et al. (2015), when the new millennium began, worldwide population estimates were approximately 6.1 billion people. The United Nations' predictions foresee an increase in numbers, first surpassing 9.2 billion by 2050 and then climbing to 11 billion by 2200, with more than 90% of the newcomers residing in the less developed areas (Todaro & Smith, 2006). As Zakaria and Muhammad (2009) argue, overpopulation has become a significant international problem particularly in third world countries. It is not only a matter of quantity of people anymore, but it is a problem of human welfare which is reflected in the accessibility of basic needs and development ambitions. Environmental degradation is among the primary consequences of rapid population growth, and it covers the extinction of the species, deforestation, desertification, and the environmental phenomenon that affects the entire globe, and this is global warming. There are also social problems like unemployment, inadequate housing, traffic congestion, and pollution, all of which are also the consequences of rapid population growth. Overall, this is a complicated situation that has brought the necessity to overcome the problems of population as one of the aspects of sustainable development highly urgent (Dominic et al, 2016). In addition, it is also possible to explain the rising crime rates in the community with the social pressure that comes with rapidly rising populations (Zakaria and Muhammad, 2009). Pakistan is marked as an area to be critically looked at in this global frame. Like most other regions in the world, growing population in Pakistan over the years has posed a stern menace to the growth of numerous projects and programs. The Box-Jenkins ARIMA technique of modeling and forecasting the population of Pakistan therefore serves the purpose of this paper in recognizing the need for informed policy dialogue. This research, with this advanced methodology, attempts to demystify the compounded demographic state of Pakistan to further equip giving insights that would then enable negotiating the challenges and opportunities that would still lie on the way of the countries developmental journey.

There is a global concern that is being screamed out loud and everyone on earth agrees upon regarding overpopulation, and the developing nations are being specifically singled out. The big size of the population is seen as the main cause of many problems and one of the areas suffering the most is food, living space, education, healthcare, and transportation. Along with these problems, it has been noticed that population pressure is also a factor behind the rise of crime in society. Therefore, various methods are being employed all over the world to control population, considering the different situations of the multicultural. A look back at the population history of Pakistan from 1947 to 1998 shows an immense increase of four times, from 33.74 million to 132.35 million. Growth rate's trends are marked; the main one is the rise during 1951-1972, later, a fall from 1972-1998. The current demographic numbers are backed by Iqbal (2007) and the 1998 population census, showing a population of 158.28 million and a growth rate of 2.69%, respectively. A total population of 156.26 million was reported by national Institute of Population Studies Islamabad (2006), with a growth rate of 1.86% thus bringing out a sharp decrease compared to the 1998 values. Stressing the significant part that accurate population figures play, the census data is recognized as one of the principal resource's indispensables for the effective national planning and economic progress. It is highlighted that the age-sex structure of the population is one of the key determinants that significantly influence the industrial product consumption pattern, employment features and consequently, the gross national production (GDP) of a nation.

The populations of the developed countries are aware of their trends, sizes, and needs, which gives them an advantage in the economic as well as the social aspect. On the other hand, the developing nations, sometimes less aware of their population changes, find it hard to attain the same economic and social level. Researchers consider the population data as a very critical one in the modern world and focus on exposing the most critical trends in population distribution. Various approaches to population forecasting are already listed and they include linear and nonlinear models such as regression, exponential and logistic regression, decay and growth models, etc. The autoregressive integrated moving average (ARIMA) method is one of the most popular ones. Furthermore, the component method of population projection, which considers fertility, mortality and migration data, is also one that remains popular among the demographers (Srinivasan, 1998).

In a local investigation that gained much attention, Zakaria & Muhammad's (2009) research proposal for population forecasting in Pakistan by means of the Box-Jenkins ARIMA models. The research, which relied on the data gathered between 1951 and 2007 concluded that ARIMA (1, 2, 0) model was the most appropriate in terms of explaining the population trends in Pakistan. Also, in a similar Asian situation, Haque et al. (2012) made a study on the forecasts for the population of Bangladesh. Going to 1991 to 2006 the results of the researchers showed that the logistic population model was the most effective in handling the population growth of Bangladesh. And, if we go to Africa, Ayele & Zewdie (2017) studied the Ethiopian human population size and patterns. The Box-Jenkins ARIMA models and yearly data 1961 to 2009 were used and therefore the study reported that ARIMA (2, 1, 2) model was the most accurate in modeling and forecasting the population trend in Ethiopia. These methods are also employed in the present study and centered on the case of Pakistan as done by Zakaria and Muhammad in 2009 and utilizes Box-Jenkins ARIMA method. The period during which the data is collected is 1960 to 2017, which will enable the effective and in-depth population modeling and prediction in Pakistan.

2. Methodology

ARIMA models are often considered to deliver more accurate forecasts than econometric techniques (song et al., 2003b). In fact, studies have indicated that they often outperform various multivariate approaches in terms of predictive accuracy (Du Preez & Witt, 2003). Earlier work also indicated that ARIMA models generally fared better than nave forecasting strategies and common smoothing techniques (Goh & Law, 2002). According to an ARIMA framework, as developed by Box and Jenkins back in the 1970s, the underlining principle of a systematic process of model identification, estimation, and diagnostic checking, informed by the principle of parsimony model should be kept as simple as possible (Asteriou and Hall, 2007). The general structure of an ARIMA (p, d, q) model can be expressed using the backward shift operator as:

$$\emptyset(B)(1 - B)\ dPPAKt = \theta(B)\mu t \quad (1)$$

where the (AR) and (MA) characteristic operators are.

$$\emptyset(B) = (1 - \emptyset 1B - \emptyset 2B\ 2 - \cdots - \emptyset pB\ p\ ) \quad (2)$$

$$(B) = (1 - \theta 1B - \theta 2B\ 2 - \cdots - \theta q\ \ ) \quad (3)$$

$$\text{And } 3\ (1 - B)AKt = \Delta dPPAKt \quad (4)$$

In this presentation, φ denotes the approximate coefficient of the autoregressive component and θ denote the approximate coefficient of the moving-average component. Where Δ differentiating operator, and d represents order of differentiating of the series. B refers to the backshift operator, and μt is the error or disturbance term.

2.1 Box-Jenkins methodology: The complexity of time series modeling

The Box-Jenkins methodology, developed by George Box and Gwilym Jenkins in the 70s, is a fantastic tool for time series analysis-particularly to forecast. This methodology consists of a coordinated series of steps. It gives the most significant role to model iteration to reach an accurate prediction. Application of the Box-Jenkins methodology is in the form of a sequence of critical steps one after the other and each stage is tasked with the development of an accurate ARIMA model. Box-Jenkins methodology, a structured time series analysis method, consists of several key steps which must be followed in the sequence given. It first processes the original time series data to achieve stationery by means of differencing which is the most significant requirement to ARIMA modeling. Next, the correlogram is referred to is the graphical view of autocorrelation and partial autocorrelation functions that provides the researcher with graphical representation of the autoregressive and moving average components that are appropriate in that they demand subjective judgments. The tentative model estimation processing the next step in the process is the tentative process in which the ARIMA model is fitted to the previously differenced data. This is followed by diagnostics testing that considers the residuals to determine the properties of white noise; in case of any deviation then model specification will be altered. The methodology is then cycled back to the next model and after that, it will go again which means that the researcher has now gotten the right, or satisfactory model according to his standards which means that the dynamic and iterative nature of the approach has been demonstrated.

2.2 Data collection

The data for carrying out ARIMA modeling is obtained from the World Bank, which is a world-renowned authority on comprehensive, accurate, and reliable statistics through its creation of a vast array of statistical databases. The researcher used 20 years (2000-2020) of data as a focus on all significant demographic indicators that are of interest in relation to the goal of the study. This dataset is made available by optimally utilizing the vast resources available at the World Bank and contains all the necessary information related to a population's size, growth rates, and various possible socio-economic factors.

3. Results and discussion

3.1 Descriptive statistics

The Table 1 presents descriptive statistics for the "year," "population," and "log transformed" variables.

Table 1: Descriptive statistics

Test statistic	Year	Population	Log transformed
Count	40.0000	40.0000	40.0000
Mean	2000.5000	156054900.0000	18.8234
Median	2000.5000	156793800.0000	18.8703
Std	11.6905	44299470.0000	0.3008
Min	1981.0000	84270200.0000	18.2495
25%	1990.7500	118256200.0000	18.5883
50%	1990.7500	156793800.0000	18.8703
75%	2010.2500	195491600.0000	19.0910
Max	2020.0000	227196700.0000	19.2413
Skewness	0.0000	-0.0136	-0.3456

Excess kurtosis	1.2000	-1.3150	-1.1321

It involves count, mean, median, standard deviation, minimum, maximum, and percentiles among other measures and provides knowledge on the distribution and the central tendency of the data. The skewness and excess kurtosis values provide information about the shape of the distributions. However, the "test statistic" column is missing, which is crucial for certain statistical tests.

3.2 Model evaluation

3.2.1 Graphical analysis (graph over time)

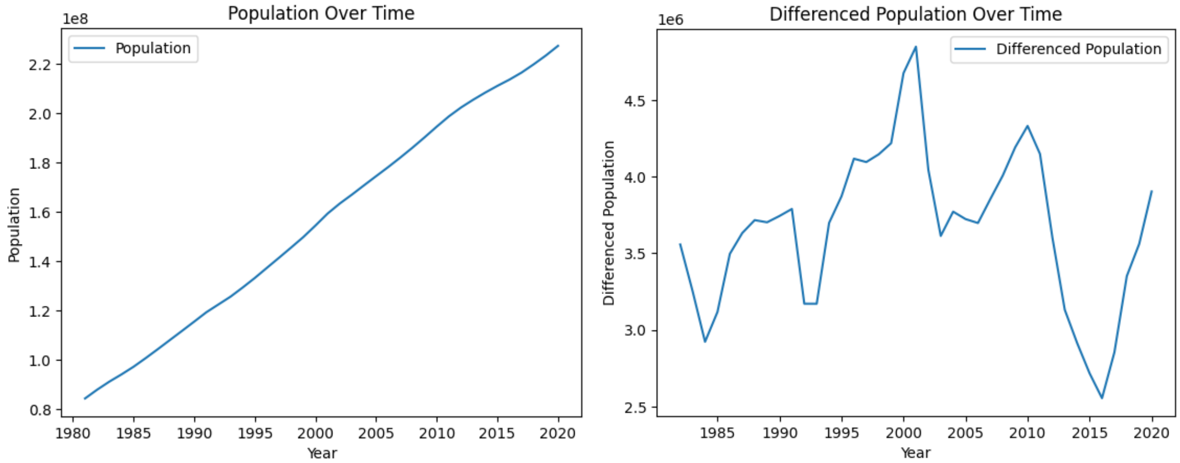

Figure 1: Population over time

3.2.2 Correlogram graph

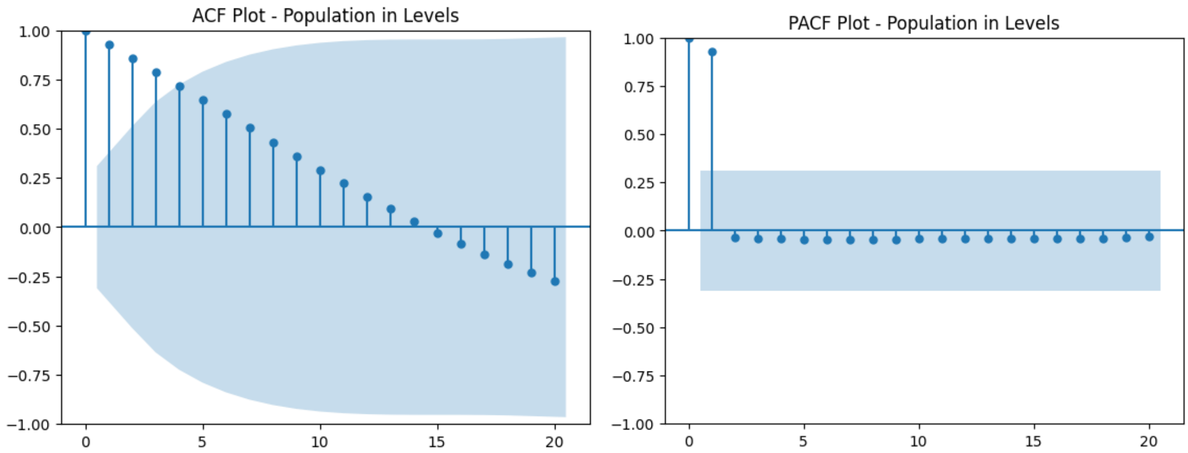

Figure 2: ACF and PACF

3.3 ADF test

Table 2: Levels-intercept

Variable	ADF statistic	Probability	Critical values	A	Conclusion

Population	-2.1613	0.2205	-3.6699	1%	Not stationary
			-2.9640	5%	Not stationary
			-2.6211	10%	Not stationary

Fail to reject the null hypothesis. The data is non-stationary.

Table 3: Levels-trend & intercept

Variable	ADF statistic	Probability	Critical values	A	Conclusion
Population	-3.2027	0.0838	-4.2970	1%	Not stationary
			-3.5685	5%	Not stationary
			-3.2184	10%	Not stationary

Fail to reject the null hypothesis. The data is non-stationary.

Table 4: Without intercept and trend & intercept

Variable	ADF statistic	Probability	Critical values	A	Conclusion
Population	-2.0295	0.0405	-2.6442	1%	Not stationary
			-1.9525	5%	Not stationary
			-1.6100	10%	Not stationary

Reject the null hypothesis. The time series is likely stationary without trend and intercept.

3.4 Correlogram (at first differences)

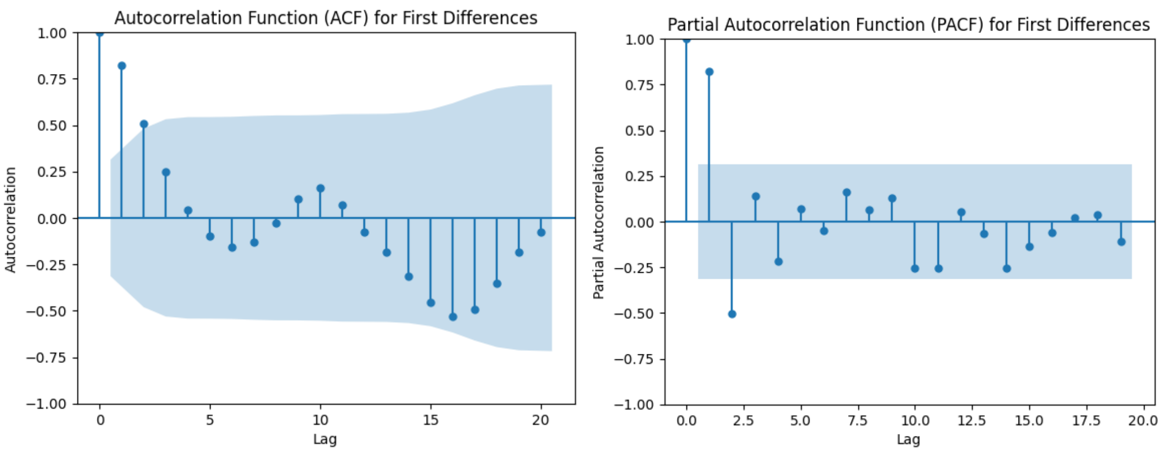

Figure 3: ACF and PACF

Table 5: 1st difference-intercept

Variable	ADF statistic	Probability	Critical values	A	Conclusion
Population	-2.3588	0.1536	-3.6889	1%	Not stationary
			-2.9719	5%	Not stationary
			-2.6252	10%	Not stationary

Fail to reject the null hypothesis. The first differences are likely non-stationary with an intercept.

Table 6: 1st difference-trend & intercept

Variable	ADF statistic	Probability	Critical values	A	Conclusion
Population	-0.2239	0.9910	-4.2970	1%	Not stationary
			-3.5685	5%	Not stationary
			-3.2184	10%	Not stationary

Fail to reject the null hypothesis. The first differences are likely non-stationary with a trend and intercept.

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

Variable	ADF statistic	Probability	Critical values	A	Conclusion
Population	-0.2493	0.5954	-2.6471	1%	Not stationary
			-1.9529	5%	Not stationary
			-1.6098	10%	Not stationary

Fail to reject the null hypothesis. The first differences are likely non-stationary without trend and intercept.

3.4.1 Correlogram in (second difference)

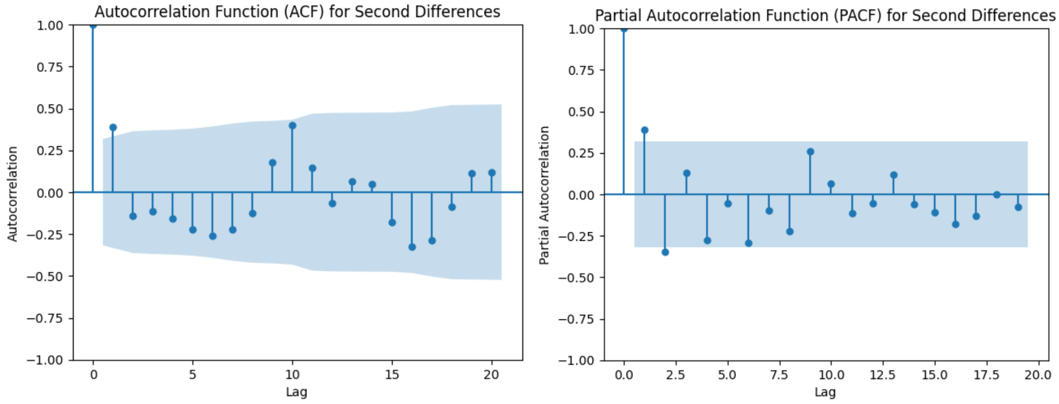

Figure 4: ACF and PACF – 2nd Difference

Table 8: 2nd Difference-intercept

Variable	ADF statistic	Probability	Critical values	A	Conclusion
Population	-1.8439	0.3588	-3.5600	1%	Not stationary
			-2.9176	5%	Not stationary
			-2.5966	10%	Not stationary

Fail to reject the null hypothesis. The second difference is likely non-stationary with intercept.

Table 9: 2nd Difference-trend & intercept

Variable	ADF statistic	Probability	Critical values	A	Conclusion
Population	-4.4490	0.1826	-4.2970	1%	Stationary
			-3.5685	5%	Stationary
			-3.2184	10%	Stationary

Reject the null hypothesis. The second difference is likely stationary with trend and intercept.

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

Variable	ADF statistic	Probability	Critical values	A	Conclusion
Population	-1.8822	0.0570	-2.6471	1%	Not stationary
			-1.9529	5%	Not stationary
			-1.6098	10%	Stationary

Fail to reject the null hypothesis. The second difference is likely non-stationary without trend and intercept.

Figure 5: 1st difference

Table 11: ADF test results for first differences in population series

Variable	ADF statistic	Probability	Critical values	A	Conclusion
Population	-2.3588	0.1536	-3.6889	1%	Not stationary
			-2.9719	5%	Not stationary
			-2.6252	10%	Not stationary

Is the series stationary? No

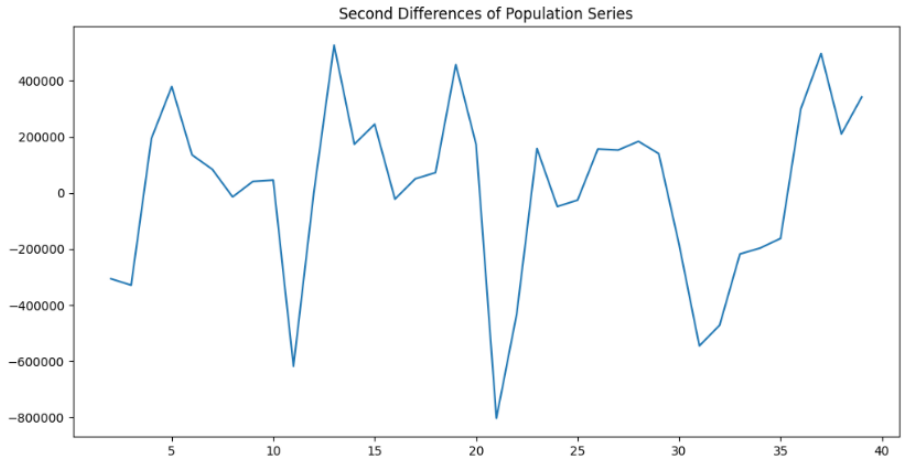

Figure 6: 2nd difference

Table 12: ADF test results for second differences of population series

Variable	ADF statistic	Probability	Critical values	A	Conclusion
Population	-1.8439	0.35886	-3.6790	1%	Not stationary
			-2.9678	5%	Not stationary

			-2.6231	10%	Not stationary

The series is no stationary.

Upon scrutinizing Figures 1 to 4 and Tables 1 to 9, it becomes apparent that the Pakistan population of Pakistan series displays non-stationary in levels, 1st differences, and 2nd differences, indicative of a discernible upward trend.

The combination of the correlogram and ADF test results gives a more complete view of how population time series data (pop) behaves as a stationary series. The "level" correlogram shows that the ACF is decaying very slowly, which means that there is a strong amount of persistence within the series and is also indicative of a non-stationary series. This pattern is validated by the ADF results presented in Tables 1-3; all three model specifications (intercept, trend plus intercept, and no-trend/no-intercept) do not reject the null hypothesis of unit root meaning that the population time series is non-stationary. Figure 3, which is the autocorrelation figure at first differences shows continued evidence of an autocorrelation pattern, as does ADF statistics presented in Tables 4-6; again, all three models do not reject the null hypothesis of unit root for all models used to determine the stationarity status, showing that even at first differences, the series is still non-stationary. A similar pattern is experienced with the second-difference correlograms (Figure 4) which demonstrated irregular decaying ACF with mixed ADF results presented in Tables 7-9. The intercept model is treated as non-stationary, while the trend and intercept model is treated as stationary with the no-trend/no-intercept model again failing to reject the null hypothesis of unit root. However, only one of the model specifications is treated as stationary thus the consistency of models is required to determine whether population series can be treated as stationary based on second differences. The summary results presented in Tables 10 and 11 confirm the conclusion that the first and second differences of population series are not stationary. Overall, the combined evidence from Figures 1 to 4 and Tables 1 to 11 clearly indicates that the Pakistan population (pop) series is dominated by a persistent upward trend and remains non-stationary under all conventional ADF test specifications.

3.5 ADF test results

Table 14: ADF test results for population series

Series type	ADF statistic	P-value	Critical value (1%)	Critical value (5%)	Critical value (10%)	Stationary
Original population	-2.1613	0.2205	-3.6699	-2.9640	-2.6211	No
First differences	-2.3588	0.1536	-3.6889	-2.9719	-2.6253	No
Seasonal differences	-2.4469	0.1289	-3.8590	-3.0420	-2.6609	No
Box-Cox transformed	-1.9542	0.3069	-3.6327	-2.9485	-2.6130	No

Log transformed	-3.1599	0.0224	-3.6327	-2.9485	-2.6130	Yes

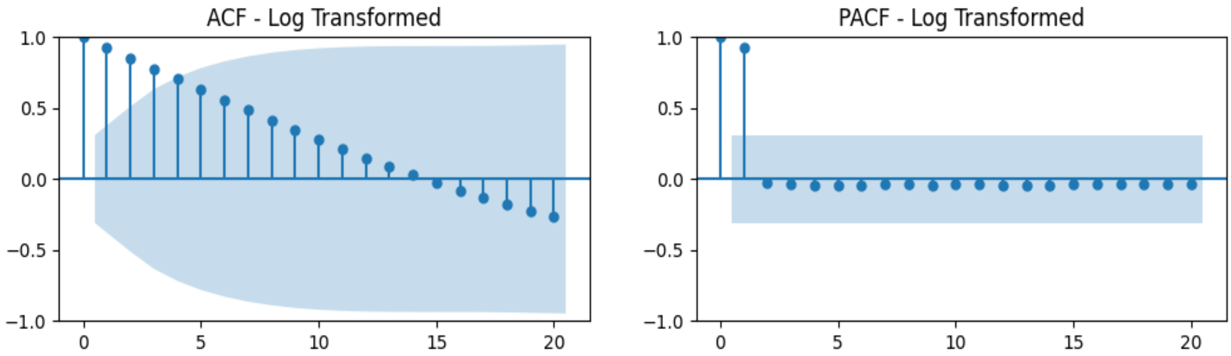

Figure 7: ACF and PACF – Log transformed

Addressing this, the augmented dickey-fuller (ADF) test was deployed, revealing that the original, first differences, seasonal differences, and Box-Cox transformed series failed to achieve stationary. However, a notable exception emerged with the log transformation, where the log-transformed series demonstrated stationary with an ADF statistic of -3.15994 and a p-value of 0.0224237. This successful transformation to stationary positions the log-transformed series as a viable foundation for subsequent analytical pursuits.

3.6 Evaluation of ARIMA models

Table 15: Evaluation of ARIMA models

Model	AIC	Theil's u	ME	MAE	MSE	MAPE
ARIMA(0, 0, 0)	20.409433	10.798661	-0.014719	0.250454	0.286468	1.331691
ARIMA (0, 0, 1)	-26.324893	5.684370	-0.011025	0.130458	0.150796	0.693581
ARIMA(0, 0, 2)	-70.488522	3.274157	-0.004160	0.072054	0.086857	0.383471
ARIMA(0, 1, 0)	-170.427606	1.000000	-0.025430	0.025430	0.026528	0.135593
ARIMA(0, 1, 1)	-217.126691	0.563890	-0.013655	0.013655	0.014959	0.072897
ARIMA(0, 1, 2)	-258.952512	0.382623	-0.007959	0.007959	0.010150	0.042592
ARIMA(0, 2, 0)	-353.636362	54.829010	0.233524	0.234466	1.454512	1.281652
ARIMA(0, 2, 1)	-363.551093	54.828995	0.233312	0.234255	1.454512	1.280537
ARIMA(0, 2, 2)	-365.629229	54.828991	0.233375	0.234219	1.454512	1.280340
ARIMA (1, 0, 0)	-165.038729	0.999279	-0.025527	0.025527	0.026509	0.136069
ARIMA (1, 0, 1)	-211.686168	0.563110	-0.013708	0.013708	0.014938	0.073160
ARIMA (1, 0, 2)	-253.725442	0.381129	-0.007904	0.007904	0.010111	0.042298

ARIMA (1, 1, 0)	-346.174610	0.261606	-0.001181	0.002612	0.006940	0.014055
ARIMA (1, 1, 1)	-360.017148	0.259735	-0.000964	0.002398	0.006890	0.012924
ARIMA (1, 1, 2)	-358.212290	0.258066	-0.001072	0.002404	0.006846	0.012950
ARIMA(1, 2, 0)	-357.688346	54.829002	0.233287	0.234341	1.454512	1.280996
ARIMA(1, 2, 1)	-363.751345	54.828992	0.233343	0.234229	1.454512	1.280396
ARIMA(1, 2, 2)	-362.916952	54.828991	0.233486	0.234209	1.454512	1.280289
ARIMA (2, 0, 0)	-340.742321	0.264464	-0.001213	0.002626	0.007016	0.014131
ARIMA (2, 0, 1)	-349.735620	0.259780	-0.001072	0.002419	0.006891	0.013034
ARIMA (2, 0, 2)	-356.022832	0.258786	-0.001031	0.002373	0.006865	0.012797
ARIMA (2, 1, 0)	-356.368303	0.260756	-0.001033	0.002432	0.006917	0.013106
ARIMA (2, 1, 1)	-351.564614	0.262133	-0.000661	0.002560	0.006954	0.013781
ARIMA (2, 1, 2)	-342.987227	0.257776	-0.001410	0.002359	0.006838	0.012711
ARIMA(2, 2, 0)	-360.929401	54.828996	0.233495	0.234293	1.454512	1.280730
ARIMA(2, 2, 1)	-361.087810	54.828993	0.233403	0.234253	1.454512	1.280521
ARIMA(2, 2, 2)	-345.637480	54.829010	0.233524	0.234466	1.454512	1.281651

In general, a model with a smaller value than the Akaike information criterion is preferred because the model fits much better than a model with a higher AIC. Similarly, Theil's u-statistic ranges between 0 and 1, and its values closer to zero indicate stronger forecasting performance. A combination of these overall evaluation measures results in the analysis below choosing the ARIMA specification, (1, 1, 2) as the most suitable model to use in this study.

3.7 Residual tests

Table 16: ADF tests for the residuals of the ARIMA (1, 1, 2) model

Variable	ADF statistic	Probability	Critical values	A	Conclusion
Population	-10127.5965	0.0000	3.6103	1%	Stationary
			-2.9391	5%	Stationary
			-2.6080	10%	Stationary

The residuals are stationary.

Table 17: ADF statistics (levels with trend and intercept)

Variable	ADF statistic	Probability	Critical values	A	Conclusion

Population	-10127.5965	0.0000	3.6103	1%	Stationary
			-2.9391	5%	Stationary
			-2.6080	10%	Stationary

The residuals are stationary.

Table 18: ADF statistics (without intercept and trend)

Variable	ADF statistic	Probability	Critical values	A	Conclusion
Population	-10127.5965	0.0000	3.6103	1%	Stationary
			-2.9391	5%	Stationary
			-2.6080	10%	Stationary

The residuals are stationary.

The ADF tests on residuals of the ARIMA (1, 1, 2) model prove that the residuals of the population variable are stationary, which shows that it is stable. This conclusion is unanimous through significance levels (1%, 5% and 10 percent) as depicted in tables 11, 12 and 13. The results suggest that the model adequately captures the underlying patterns in the data.

3.8 Stability of the model

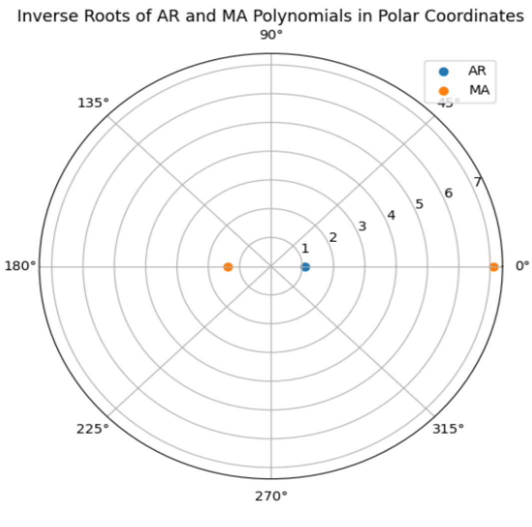

Figure 8: Inverse Roots

Inverse roots of AR polynomial	1.0845	
Inverse roots of MA polynomial	-1.3839	7.0972

Table 19: Model coefficients and statistical significance

Variable	Coefficient	Standard error	Z-score	P-value	Significance
X1	0.0271	0.0052	5.216	1.83×10^{-7}	***
Ar.l1	0.9220	0.0820	11.23	2.64×10^{-29}	***
Ma.l1	0.5816	0.2100	2.769	0.0056	**
Ma.l2	-0.1018	0.2123	-0.479	0.043164	**
Sigma2	3.453×10^{-6}	8.336×10^{-7}	4.142	3.44×10^{-5}	***

The *, **, and *** means significant at 10, 5 and 1 per cent levels of significance respectively.

The results presented in Table 12 indicate that the original form, first differences, seasonal differences, and Box–Cox transformations of the population series are still classified as non-stationary. This conclusion is based upon all the ADF statistic values falling below the critical value and their corresponding p-values remaining greater than .05 for all series. However, when transforming the series to logarithmic form, it becomes stationary when an ADF statistic is calculated at -3.15994 with a p-value of .0224, meaning it is also the only transformable series suitable for continued analysis of the population series. From this finding, an additional analysis of several ARIMA fitted models can be performed shown in Table 13. ARIMA (1,1,2) has the best overall AIC value of all fitted models (one of the lowest values) and has the smallest Theil's u-statistic of all models indicating the highest quality of forecasting accuracy and efficiency using the ARIMA (1,1,2) fitted model. The stationarity of residuals from the ARIMA (1,1,2) fitted model has also been established through ADF tests in Tables 14-16 which consistently have large negative ADF statistics with corresponding p-values of .0000 for all specifications meaning the model satisfactorily explained the underlying structure of the data. Stability diagnostics confirm the model is stable in Figure 8 because all inverse roots of the AR and MA polynomials used in the ARIMA (1,1,2) model are located outside of the unit circle. Descriptive statistics as shown in Table 17 characterize the population with an uptrend, a highly increasing trend with a mean population of 156 million and negative, but mildly negative skewness, the log transformed series has lower variability and a more symmetric distribution. At last, Table 18 provides the ARIMA (1,1,2) parameter estimates and demonstrates that the AR (1) and MA (1) terms are highly significant, whereas MA (2) is insignificant, and the error variability (sigma 2) is statistically significant, all together proving a good sufficiency of the model and its relevance to the prediction of the Pakistan population dynamics.

3.9 Predicted total population

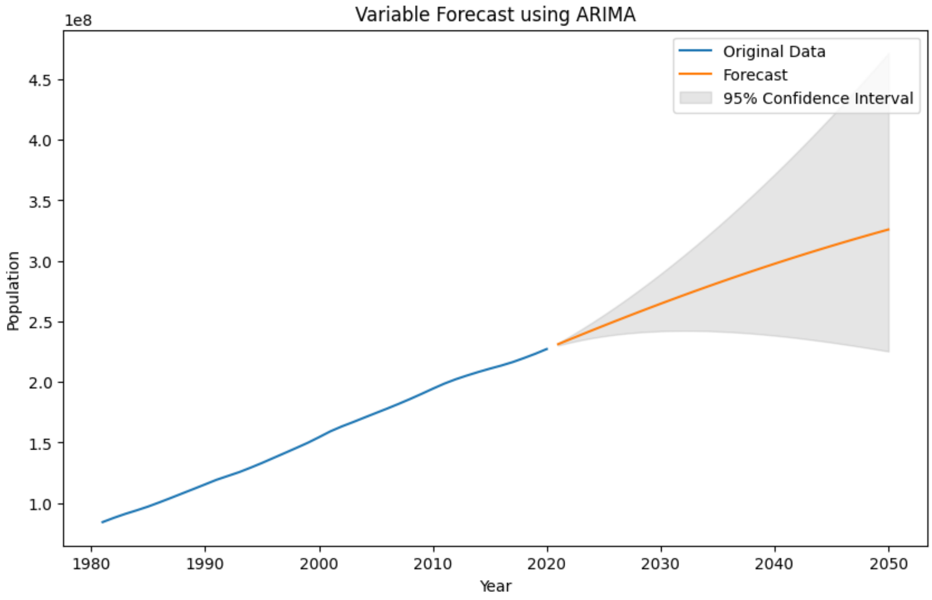

Figure 9: Prediction

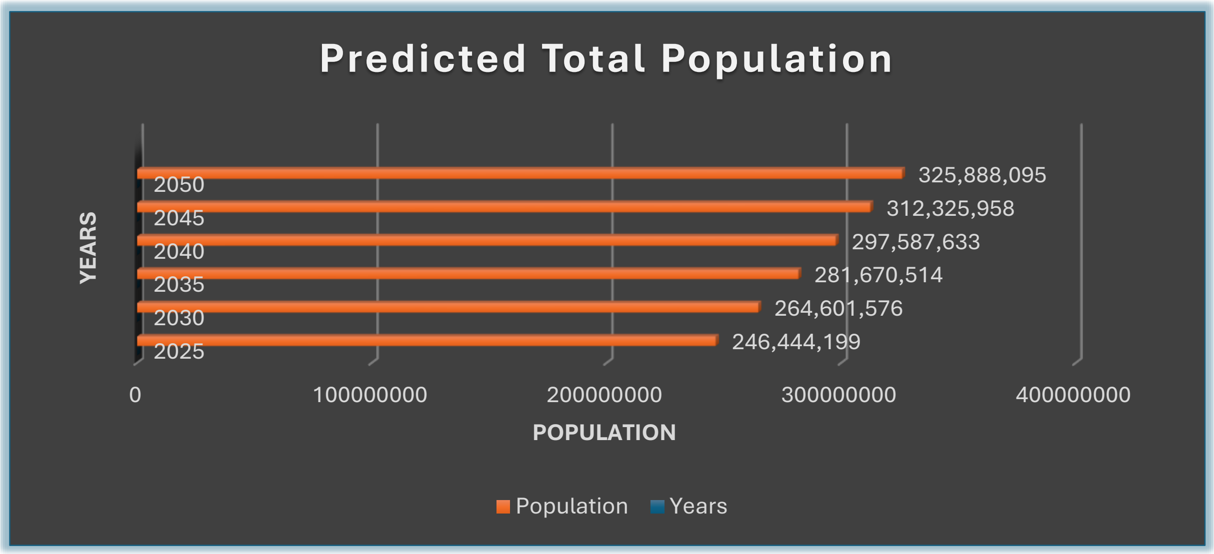

Figure 10: Predicted Population

A closer look at Figure 10, which shows the forecast picture that the years 2021-2050 will follow, would not be more persuasive in proving that the populated in Pakistan will rise dramatically and consistently in the next three decades. Using ARIMA (1, 1, 2) model, the 95 per cent confidence interval calculated is 232,158,189.43 to 471,391,431.96 with a predicted cumulative population of 324,356,000 as of the year 2050.

3.10 Policy implications

The ARIMA-based population projections suggest a sharp and sustained increase in Pakistan's population between 2021 and 2050. The population expected to reach approximately 324 million by 2050. This projected growth highlights the urgent need for effective population management policies. In this regard, comprehensive family planning campaigns should be executed and promoted to enhance awareness, education, and nationwide access to contraceptive methods. Policymakers should be informed of strategies that support the adoption of smaller family sizes, emphasizing the socioeconomic benefits of a reduced population growth rate.

Moreover, incorporating sex education into the school curriculum is essential to equip young people with the knowledge required for responsible family planning. Strengthening partnerships between public, social, and business sectors can improve resource mobilization and the effective implementation of family planning initiatives. To ensure policy effectiveness, an efficient monitoring and evaluation system should be established to assess progress and guide necessary adjustments. Overall, consistent government involvement in family planning practices, the promotion of smaller family norms, and fertility control through education are critical to mitigating the long-term impacts of rapid population growth in Pakistan.

4. Conclusion

The ARIMA (1,1,2) model has been identified as the most suitable and parsimonious choice for forecasting Pakistan's population over the next three decades. Based on the model's projections, the population is expected to reach approximately 325.9 million by the year 2050. This anticipated rise is alarming, given the multifaceted challenges that rapid population growth already poses for the country.

These results carry significant implications for policymakers. A population approaching this magnitude will place considerable pressure on Pakistan's socioeconomic infrastructure, including employment generation, poverty alleviation, healthcare, education, housing, and public safety. Without timely interventions, existing issues—such as high unemployment, persistent poverty, and rising crime rates—may intensify, further straining national resources.

The findings of this study can therefore play a vital role in informing the government's long-term planning and strategic foresight initiatives. Reliable population forecasts enable more effective resource allocation, better formulation of development policies, and proactive measures to manage demographic shifts. By highlighting the urgency of sustainable population management and evidence-based decision-making, this research contributes to a deeper understanding of the demographic trajectory of Pakistan and underscores the importance of data-driven planning for a stable and prosperous future.

References

1. Ahlburg, D. A (1998). Julian Simon and the population growth debate, Population and Development Review, 24: 317 – 327.
2. Asteriou, D. & Hall, S. G. (2007). Applied Econometrics: a modern approach, Revised Edition, Palgrave MacMillan, New York.
3. Ayele, A. W & Zewdie, M. A (2017). Modeling and forecasting Ethiopian human population size and its pattern, International Journal of Social Sciences, Arts and Humanities, 4 (3): 71 – 82.
4. Becker, G., Glaeser, E., & Murphy, K (1999). Population and economic growth, American Economic Review, 89 (2): 145 – 149.
5. Dominic, A., Oluwatoyin, M. A., & Fagbeminiyi, F. F (2016). The determinants of population growth in Nigeria: a co-integration approach, The International Journal of Humanities and Social Studies, 4 (11): 38 – 44.
6. Du Preez, J. & Witt, S. F. (2003). Univariate and multivariate time series forecasting: An application to tourism demand, International Journal of Forecasting, 19: 435 – 451.

7. Goh, C. & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic non-stationary seasonality and intervention, Tourism Management, 23: 499 –510.
8. 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.
9. Malthus, T. (1798). An essay on the principle of population, Pickering, London.
10. Nyoni, T. (2018). Modeling Forecasting Naira / USD Exchange Rate in Nigeria: a Box–Jenkins ARIMA approach, University of Munich Library – Munich Personal RePEc Archive (MPRA), Paper No. 88622.
11. Nyoni, T (2018). Modeling and Forecasting Inflation in Kenya: Recent Insights from ARIMA and GARCH analysis, Dimorian Review, 5 (6): 16 – 40.
12. Nyoni, T. (2018). Box–Jenkins ARIMA Approach to Predicting net FDI inflows in Zimbabwe, Munich University Library – Munich Personal RePEc Archive (MPRA), Paper No. 87737.
13. Solow, R. (1956). Technical change and the aggregate population function, Review of Economics and Statistics, 39: 312 – 320.
14. Solow, R. (1956). Technical change and the aggregate population function, Review of Economics and Statistics, 39: 312 – 320.
15. Song, H., Witt, S. F. & Jensen, T. C. (2003b). Tourism forecasting: accuracy of alternative econometric models, International Journal of Forecasting, 19: 123 – 141.