International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Time Series Approach to Forecasting Birth Rate in India

Pankaj Kumar D. Parmar[1], Dr. Sanjay Patel[2]

Research Scholar, Assistant Professor, M.G. Science Institute, Ahmedabad, India
Email: *parmar.pan0303[at]gmail.com*
Mobile No: +919510289439

Assistant Professor, HOD, Department of Statistics, Udhana Citizen Commerce College, Surat, India

Abstract: ***This research study seeks to model and forecast the Birth rate in India using Box- Jenkins method. Yearly data from1960 to 2022 was collected from World Development Indicators (World Bank). SPSS statistical software, E-views and MS Excel were used for the analysis of the data. ARIMA (1, 1, 1) model was found to be the most suitable model with the least normalised Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC) values amidst the model test in the work. The developed model was then used to forecast for the years 2023 to 2030 and the results showed that the Birth rate will decrease in the next future years. In conclusion, it was established that there was a negative relationship between the dependent variable (Birth rate) and the independent variable (time/year).***

Keywords: Time Series, Birth Rate, Box-Jenkins Method

1. Introduction

India is the second-most populous nation on earth. India ranks eighth internationally in terms of land area. India's population makes up roughly 17.70% of the global population. The three states with the most inhabitants in India are Uttar Pradesh, Maharashtra, and Bihar. Sikkim, in comparison, has the lowest population density in the nation. China now has the highest population in the world, but India's birth rate and population growth rate suggest that India may overtake China's population in the near future. Here are some fascinating stats on various demographic features of India.The gathering of birth history data to improve data availability for tracking the pace of fertility drop has been a significant component of the demographic and health surveys carried out throughout the nation.

The total number of live births per 1,000 of the population in a year is the birth rate, to put it simply. A big population rise might result from high birth rates, which wouldn't be good for India because it would pose several challenges to the nation's development and economy.

Due to inadequate documentation of the district's and the country's birth rates, this crucial aspect has been overlooked. Determining the trend and making plans for the country's economy have grown more challenging.

In order to anticipate birth rates in India, this work aims to both analyse the trend pattern of birth rates and fit an ARIMA model.

2. Literature Review

Carter and Lee (1986) used mathematical demographic and statistical time series methodologies in their work. They used univariate time series, transfer function, and multivariate time series models to find the best models of fertility and nuptality indices, and the researchers then used a demographic simultaneous-equation model to anticipate births and marriages in conjunction with total births and first marriages. This forecasting approach was used to anticipate first marriages and total births in the United States.

Toth et al. (2000) compared short-time rainfall prediction models for real-time flood forecasting. Completely different structures of auto-regressive moving average (ARMA) models, ANN and NearestNeighbours approaches were applied for forecasting storm rainfall occurring within the Sieve river basin, Italy, within the amount 1992- 1996 with lead times variable from 1 to 6 h. The ANN adaptive activity application proved to be stable for lead times longer than three hours, however inadequate for reproducing low rainfall events. Abraham et al. (2001) used an ANN with scaled conjugate gradient algorithmic rule (ANN-SCGA) and evolving fuzzy neural network (EfuNN) for predicting the rainfall time series. In the study, monthly rainfall was used as input data for training model. The authors analysed eighty-seven years of rainfall information in Kerala, a state within the southern a part of the Indian dry land. The empirical results showed that neuro-fuzzy systems were economical in terms of getting higher performance time and lower error rates five compared to the pure neural network approach. nevertheless, rainfall is one in all the twenty most complicated and tough parts of the geophysical science cycle to grasp and to model due to the tremendous range of variation over a wide range of scales both in space and time.

Buah Ahoba Masha (2020), "Time Series Analysis of Monthly Birth Rate in Ghana".In this study, he examined Ghana's birth rate using the Box- Jenkins technique on a monthly basis.He obtained monthly birth rate data from the Ghana Statistical Service for the research period of January 2014 to December 2019.Among the models tested in the paper, he discovered that the ARIMA (1, 1, 1) model had the best fit, with the lowest normalised Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) values. The generated model was applied to the 12-month birth rate for the year 2020. Among the models tested in the study, the ARIMA (1, 1, 1) model was determined to have the best fit, with the lowest normalised Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) values. In conclusion, a negative association was discovered between the dependent variable (Birth rate) and the independent variable (time/months).

Muhammad Zakria (2009), "Forecasting the population of Pakistan using ARIMA model". The study focus on forecast thepopulation trend of Pakistan for the next 20 years onthe basis of previous trends using a stochastic ARIMAmodel.In this empirical study, population ofPakistan from 1951 to 2007 is modelled using Box Jenkins ARIMA methodology. For this Purpose, he collected the data of Pakistan population for the period of 1951 to 2007.The population of Pakistan isalso forecasted for the next 20 years using the parsimonious ARIMA (1, 2, 0) model. He found that current growth ratetrend continues, the population of Pakistan would be approximately 230.7 million in 2027.The result of projected population by parsimonious ARIMA model is close to the projected population ofPakistan by different bureaus. These bureaus reported 229 million population of Pakistan for the year 2025. He concludes that the estimates provided in study usingARIMA (1,2,0) are close to other researcher's findingand are equally important for Government of Pakistanas well as Non-Government Organizations for futureplanning and projects.

Anonymous, (N.D) (2023), "Forecasting the Birth Rate in the Philippines using ARIMA Model" The ARIMA Model will be used to forecast the birth rate in the Philippines in this article. Furthermore, the dataset used in this study is from the United Nations - World Population Prospects, which is the official United Nations population estimates and projections prepared by the Population Division of the United Nations Secretariat's Department of Economic and Social Affairs. It gives population estimates for 237 nations or territories from 1950 to the present, supported by assessments of historical demographic patterns. However, this analysis will only include 53 annual observations from 1971 to 2023. The best fitted model is ARIMA (0,2,0) in this study. He concludes the birth rate in the Philippines will slowly decline as the year increases.

3. Scope and Limitation

This study simply employs software to simulate the ARIMA and anticipate the birth rate in India. Furthermore, the data for this study was acquired from the World Development Indicators (World Bank) website Prospects. Furthermore, from 1960 to 2022, this research will collect63 yearly observations.

4. Research Methodology

The major data source was a World Development Indicators (World Bank) database named annual Birth Rate 1960-2022. The relevant data may be found as a time series for the historical data of India's birth rate. Using software, this research will plot the time series and see if it exhibits constant behaviour or changes over a longer period of time. This research will examine the original ACF and PACF series to see if differencing is required. The Akaike Information Criterion will be utilised in this study to select the model that will be employed for the time series. The ACF and PACF residuals will be plotted using software in this study to determine if the generated model is stationary.The resulting ARIMA model will be used to forecast the Birth Rate Data (BRD) model. The anticipated model's output values will be plotted. The Ljung-Box test will be used in this study to determine if the time series has autocorrelation. The void The residuals are independently distributed, according to Hypothesis H_0. The alternative hypothesis is that the residuals are not distributed separately and have a serial correlation.

5. Finding

5.1 Examining the Stationarity of the Birth Rate

The time series plot of the Yearly birth ratio data is shown in Fig 1. The time horizon of the dataset 1960 to 2022.

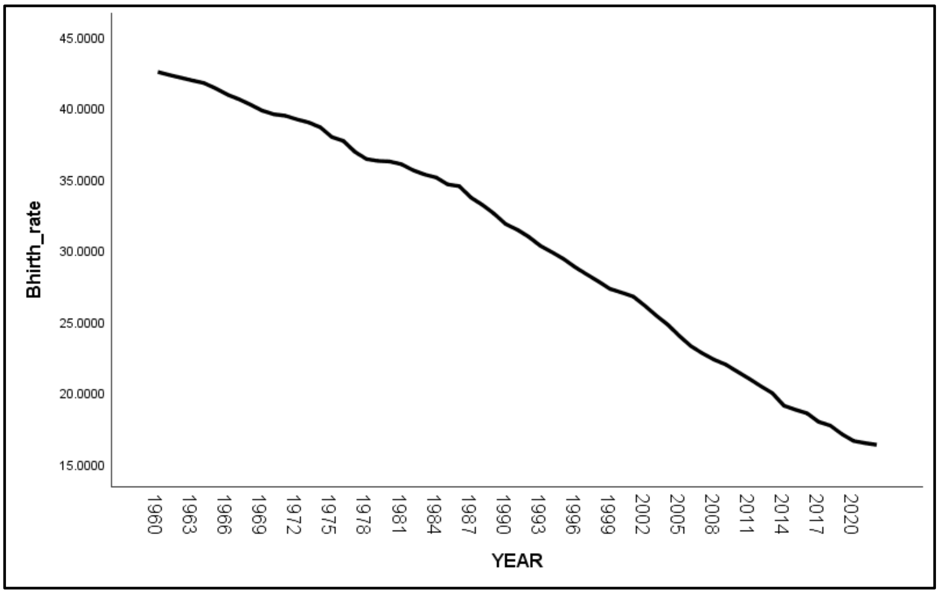


Chart 1: Time series plot of Birth Rate in India from 1960 to 2022

Paper ID: SR231229144243 DOI: https://dx.doi.org/10.21275/SR231229144243

The overall shape of the trend of India's yearly birth rate may be seen in Fig. 1. In other words, the birth rate's overall trend does not appear the same throughout time, making it a nonstationary series.

5.2 ANOVA Test of Stationarity

Regression analysis is used to determine if the independent variable in the raw data and the series' dependent variable are related.

- Null Hypothesis: Birth Rate is Stationary.
- Alternate Hypothesis: Birth Rate is not stationary.

Significance level = 5%.

Decision Rule: Do not reject the null hypothesis—that the birth is steady—if the p value above the level of significance. If not, the birth is not stationary.

Table 1: ANOVA Test of Stationarity

Model	Sum of Squares	DF	Mean Square	F	Sig.
Regression	4245.372	1	4245.272	5746.814	.000[b]
Residual	45.063	61	0.739		
Total	4290.435	62			
a. Dependent Variable: Birth-rate					
b. Predictors: (Constant), Years					

From Table 1, the p-value is 0.000(0.000 < 0.05) which is less than the 5% significance level. Therefore, the null hypothesis is rejected indicating that the of birth rates series is not stationary.

5.3. Augmented Dickey Fuller (ADF) Root Test

The hypothesis is established as follows, and the Augmented Dickey Fuller unit root test is run at 5% level of significant:

- Null Hypothesis: Birth Rate is Not Stationary (i.e. it has a unit root).
- Alternate Hypothesis: Birth Rate is Stationary (i.e. it has no unit root).

Decision Criteria: The null hypothesis is rejected, i.e., the birth rate is stationary, if the absolute value of the t-statistic is larger than the absolute of the critical value (provided at the level of significance). If not, the birth rate is not stable.

Table 2: Augmented Dickey Fuller Unit Root Test

ADF Test		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		1.2843	0.9983
Test critical values:	1% level	-3.5420	
	5% level	-2.9100	
	10% level	-2.5926	

From Table 2, the ADF t-statistic value is 2.2366 and probability value is 0.9983 (0.9983> 0.05) which is greater than 0.05, so we don't reject the null hypothesis at 5% level of significance. Therefore, the birth rate series is not stationary.

5.4 Differencing the Series to Achieve Stationarity

The Box-Jenkins technique makes the assumption that time series data are stationary. In order to do that, the time series' correlogram is examined to determine whether or not it is stationary. If not, the data are subjected to the process of differencing to eliminate the fluctuations in the series and maintain a consistent mean and variance throughout time.

Table 3: Augmented Dickey Fuller Unit Root Test

ADF Test		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.4592	0.000
Test critical values:	1% level	-3.5420	
	5% level	-2.5926	
	10% level	-2.5926	

After the first order differencing the ADF t-statistics value is -5.4592 and P-value is 0.000 which is less than 0.05 (0.000< 0.05), So, we rejected the null hypothesis at 5% level of significance. There for birth rate series is stationary.

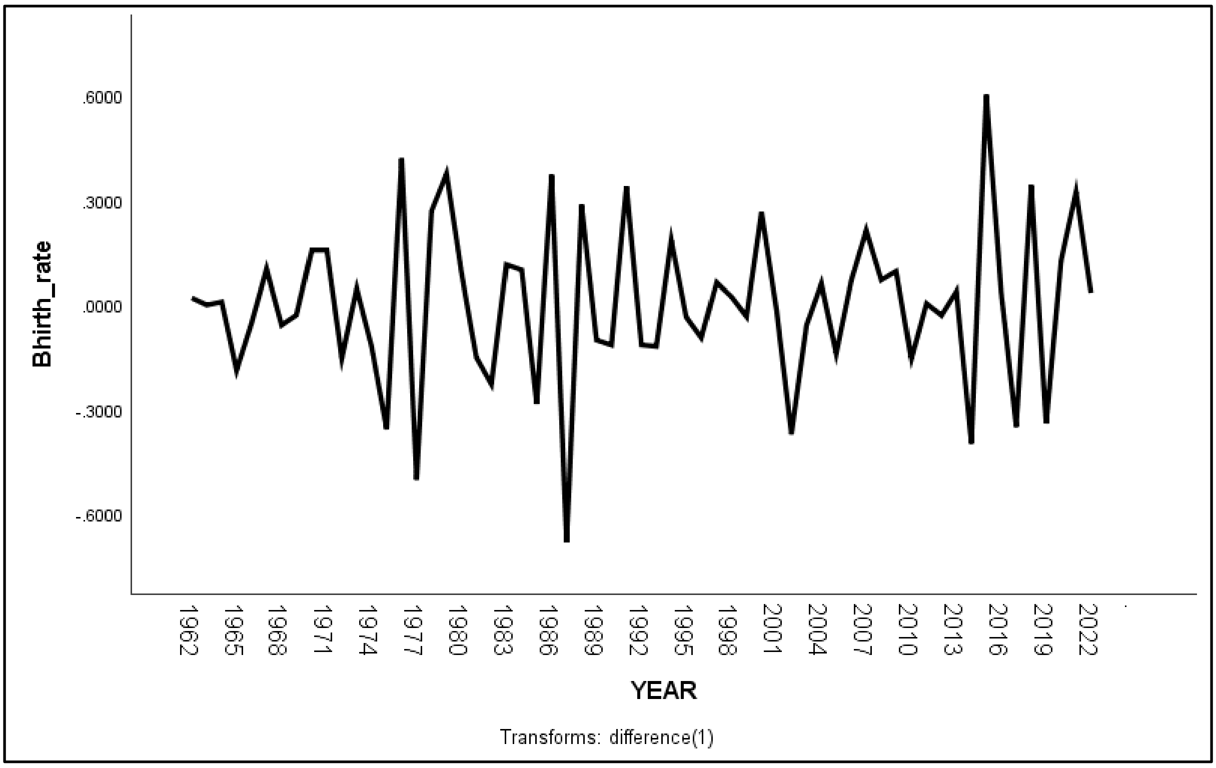
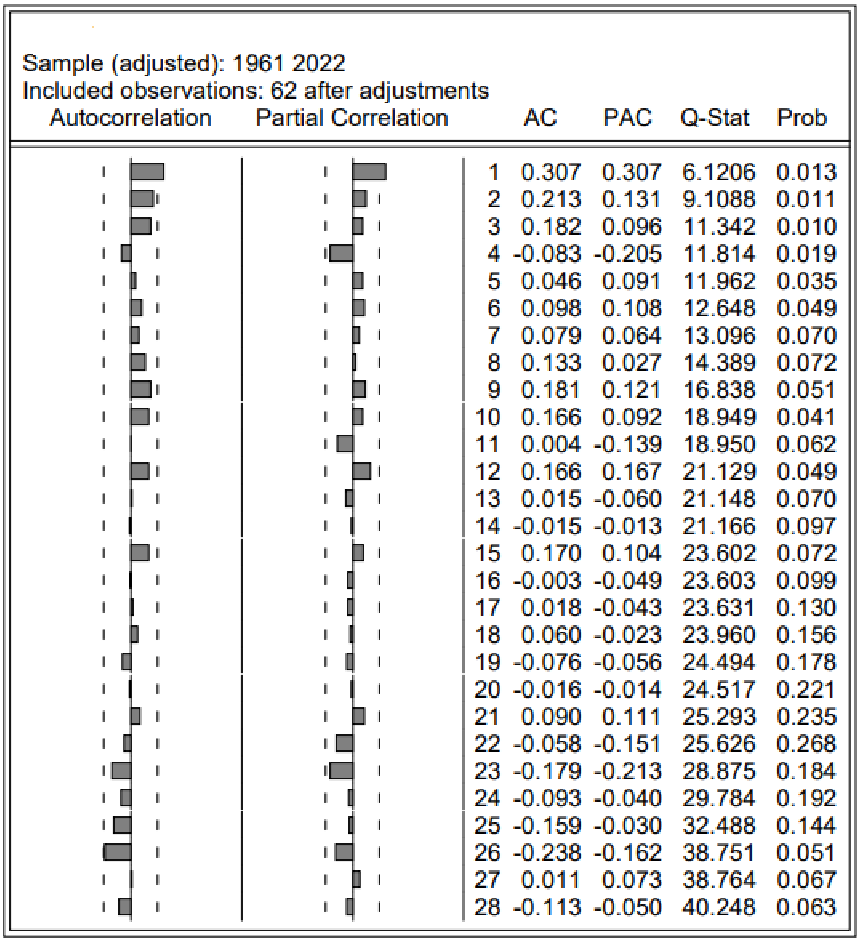


Chart. 2: A Correlogram of the first Differencing of the Time Series Plot.

5.5 Identification of Tentative Model

The Box Jenkins technique may be utilised to determine the potential Autoregressive (AR), Moving Average (MA), and their corresponding orders.For time series data, the average (ARIMA) and ARIMA (p, d, q) models would be suitable. Since the data was only varied once, it can be inferred that ARIMA (p, 1, q), where p stands for the autoregressive (AR), which can be acquired from the ACF, and q stands for the moving average (MA), which can be obtained from the PACF, will be applied to the data. When a stable time series has been obtained by differencing, the correlation is investigated to determine the proper ordering of the AR and MA components. To do this, ACF and PACF plots of the differenced data are studied.

Paper ID: SR231229144243 DOI: https://dx.doi.org/10.21275/SR231229144243



Sample (adjusted): 1961 2022
Included observations: 62 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.307	0.307	6.1206	0.013
		2	0.213	0.131	9.1088	0.011
		3	0.182	0.096	11.342	0.010
		4	-0.083	-0.205	11.814	0.019
		5	0.046	0.091	11.962	0.035
		6	0.098	0.108	12.648	0.049
		7	0.079	0.064	13.096	0.070
		8	0.133	0.027	14.389	0.072
		9	0.181	0.121	16.838	0.051
		10	0.166	0.092	18.949	0.041
		11	0.004	-0.139	18.950	0.062
		12	0.166	0.167	21.129	0.049
		13	0.015	-0.060	21.148	0.070
		14	-0.015	-0.013	21.166	0.097
		15	0.170	0.104	23.602	0.072
		16	-0.003	-0.049	23.603	0.099
		17	0.018	-0.043	23.631	0.130
		18	0.060	-0.023	23.960	0.156
		19	-0.076	-0.056	24.494	0.178
		20	-0.016	-0.014	24.517	0.221
		21	0.090	0.111	25.293	0.235
		22	-0.058	-0.151	25.626	0.268
		23	-0.179	-0.213	28.875	0.184
		24	-0.093	-0.040	29.784	0.192
		25	-0.159	-0.030	32.488	0.144
		26	-0.238	-0.162	38.751	0.051
		27	0.011	0.073	38.764	0.067
		28	-0.113	-0.050	40.248	0.063

Chart 3: A Correlogram of ACF and PCF with First Order Differencing.

From the ACF graph, the plot shows the autocorrelation function of first differencing of birth rates at various lags. It was found out that, the ACF has a positive significant spike at lag 1. From the above PACF plot, it has positive significant spike atlag1. So, maybe there is auto-correlation problem in the data and moving average problem in the series. The final possible model is **ARIMA (1, 1, 1)** from chart 3.

5.6. Ljung-Box Test of Serial Correlation for ARIMA (1, 1, 1)

The following is a summary of the test's hypothesis:

- Null Hypothesis H_0 :β_0= 0 (No Serial Correlation).
- Alternate Hypothesis H_1 :$\beta_0 \neq$ 0 (Serial Correlation).

Decision Rule: Reject the null hypothesis if the significant result is less than the significance threshold, which is 5%, and do not do so otherwise.

Table 4: Model Statistics for ARIMA(1,1,1)

Model	Number of Predictors	Model Fit statistics		Ljung-Box Q(18)		
		Stationary R-squared	R-squared	Statistics	DF	Sig.
Birth_Rate-Model_1	1	0.454	1.000	7.013	16	0.973

Interpretation:
We do not reject the null hypothesis, which suggests that the Ljung-Box Q(18) is not statistically significant at the 5% level of significance, based on the observation in Table 4 above that the model's p-value is 0.973 (0.973 > 0.05), which is bigger than 0.05 (the 5% level of significance). The ARMA (1, 1, 1) model is the recommended model since it is independently distributed and lacks serial correlations.

Table 5: ARIMA (1, 1, 1) Model Parameters

	Parameters	Estimate	SE	T-statistics	Sig.
Birth_Rate-Model_1 (Natural Logarithm)	Constant	0.653	0.110	5.926	0.000
	AR(1)	-0.632	0.303	-2.084	0.042
	MA(1)	-0.803	0.234	-3.425	0.001

Interpretation:
The ARIMA (1,1,1) model's parameter and constant P-value of time are both less than 0.05 in Table 5 above, indicating statistical significance at the 5% level. As a result, we can include both parameters for construct the model.

The prediction equation for the ARIMA (1,1,1) model with constant term would be defined by equation (1) based on Table 5.

Model Equation: $\widehat{Y}_t = \mu + Y_{t-1} + \alpha(Y_{t-1} - Y_{t-2}) + \emptyset\varepsilon_{t-1} + \varepsilon_t$**....................(1)**

μ =Constant,
Y_{t-1} = Time with 1 difference,
ε = Random error,
Y_{t-1} = Lag 1 of Y_t
Y_{t-2} = Lag 2 of Y_t,
ε_t= Random Error Term,
ε_{t-1}= Lag 1 of the Random Error Term
α= Auto Regressive (AR) Coefficient,
ϕ= Moving Average (MA) Coefficient

The parameter estimates for ARIMA (1,1,1) is shown in Table 5 above along with the associated significance level. Equation (2) is produced by plugging the constant values into Equation (1).

$\widehat{Y}_t = 1.9213 + Y_{t-1} + 0.5315(Y_{t-1} - Y_{t-2}) + 0.4479\varepsilon_{t-1} + \varepsilon_t$....................(2)

Equation (2) is the forecasting equation that shows the Indian birth rate trend.

5.7. Model Forecasting

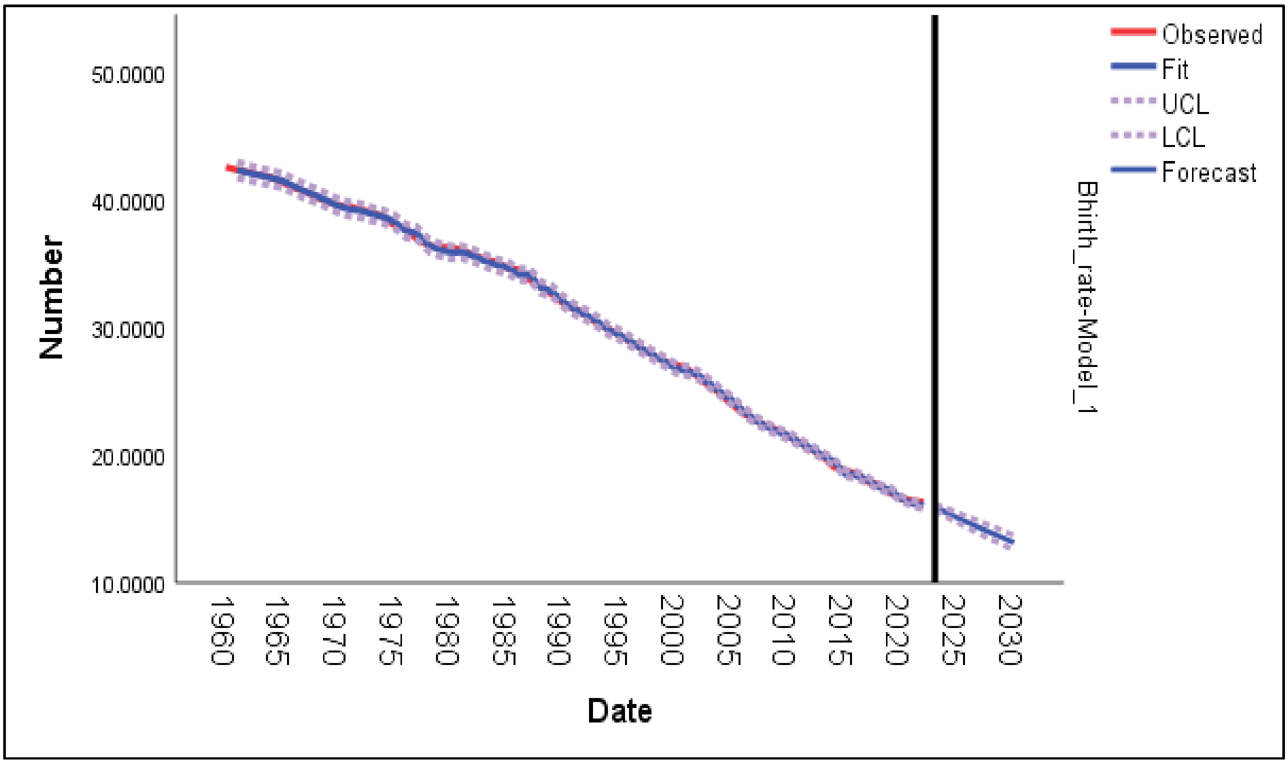


Chart 5: Forecasted Values of the Birth Rate in India from 2022 to 2030

Paper ID: SR231229144243 DOI: https://dx.doi.org/10.21275/SR231229144243

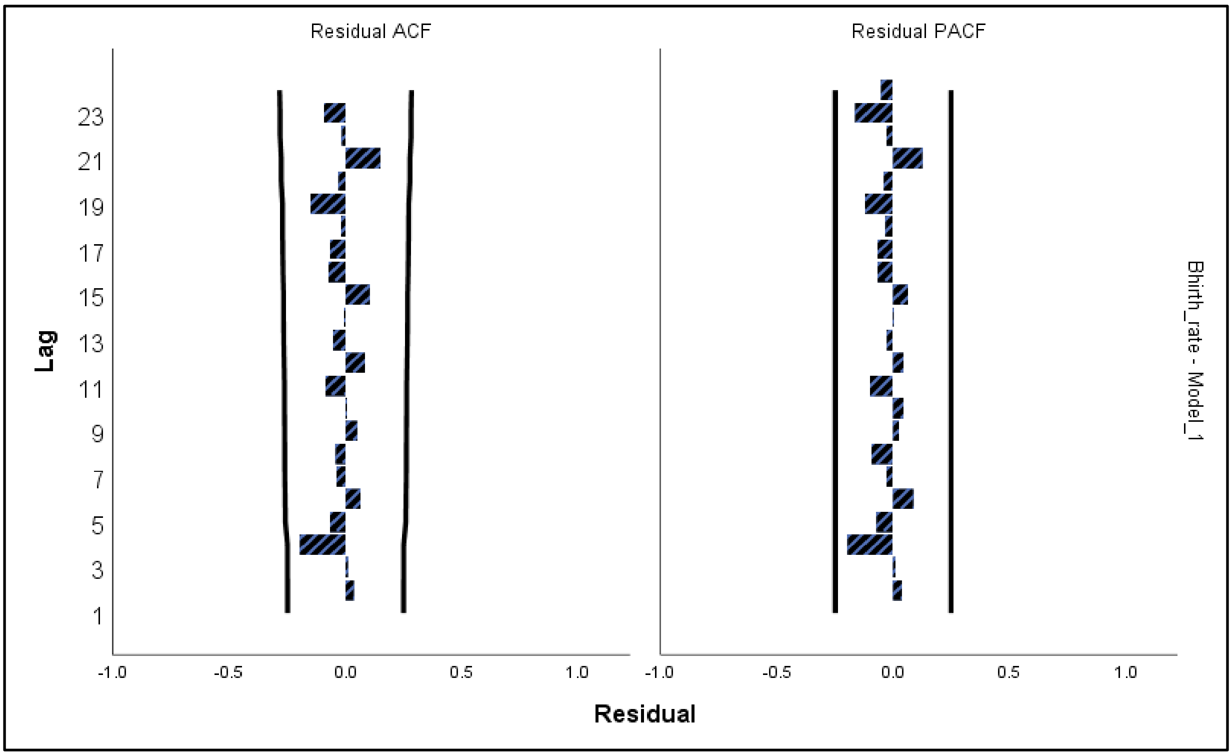


Chart 6: Forecasted residual ACF and PCF Birth Rate in India from 2022 to 2030.

Table 6: Forecasted Values of the Birth Rate in India from 2023 to 2030

Year	Forecasted Value	95% Confidence Level	
		UCL	LCL
2023	15.9153	16.1417	15.6913
2024	15.4814	15.8216	15.1468
2025	15.0869	15.4904	14.6913
2026	14.6774	15.1370	14.2284
2027	14.2866	14.7869	13.7991
2028	13.8939	14.4295	13.3734
2029	13.5122	14.0756	12.9659
2030	13.1337	13.7206	12.5658

6. Conclusion

It was discovered during this analysis that the original series data for the annual birth rate in India (1960 to 2022) was not stationary, thus first differencing was carried out to achieve stationarity. To determine if the initial differenced data were stationary, the ADF test was used. We discovered from the ACF and PACF graph that the data are free of autocorrelation issues, making ARIMA (1, 1, 1) an adequate fit for the trend in the annual birth rate in India from 2023 to 2030.We observed from the forecasting values that the birth trend of India goes to downward in the future which may be due to awareness cause by government efforts toward population control.

References

[1] Nuzula, Isnaini. (2019). The Approach of box Jenkins Time Series Analysis for Predicting Stock Price on LQ45 Stock Index. Journal of profit. pp. 18-25.
[2] Zhang Z., Moore, J. C., (2015), “Autoregressive Moving Average Models” Mathematical and Physical Fundamentals of Climate Change.
[3] Buah Ahoba Masha (2020), “Time Series Analysis of Monthly Birth Rate in Ghana”.
[4] https://stats.stackexchange.com/
[5] https://www.studocu.com
[6] Vijay Kumar G (2021), “Statistical Tests to Check Stationarity in Time Series”
[7] David A. Dickey. "Stationarity Issues in Time Series Models" (PDF). 2.sas.com. Retrieved 2016-06-26.
[8] "Unit Root Testing — arch 4.19+14.g318309ac documentation". *arch.readthedocs.io*. Retrieved 2021-10-18.
[9] Fuller, W. A. (1976). Introduction to Statistical Time Series.New York: John Wiley and Sons. ISBN 0-471-28715-6.
[10] World Development Indicators (World Bank).