Journal of Statistics Applications & Probability
*An International Journal*

http://dx.doi.org/10.18576/jsap/120326

# Robustness of Predictive Performance of Arima Models Using Birth Rate of Tamilnadu

D. Karunanidhi* and S. Sasikala

Department of Statistics, Thanthai Periyar Government Arts & Science College, Trichy-23, Tamil Nadu, India

Received: 2 Nov. 2022, Revised: 22 Dec. 2022, Accepted: 9 Feb. 2023.
Published online: 1 Sep. 2023.

**Abstract:** Birth Rate is one of the important measures in Vital Statistics. In Policy decision making, Birth rate of the state plays a major role and based on that Government makes decision for people welfare. Predicting births and birth rates are fundamental factors in predicting the future population of states. Time series models are one of the best methods for forecasting future values. In time series models, ARIMA models are the class of models which are used to predict the values which can be made the stationarity. In this study, the state TamilNadu birth rates from 1950-2019 has been considered to forecasts the birth rate of the state using Box-Jenkins methods. The ARIMA (5,1,1) models were found to have lower normalised Bayesian information criterion (BIC) and Akaike information criterion (AIC) values, making them more acceptable. The ARIMA (0,1,0) model was used to predict the birth rate for the next 30 years, and the results indicated that the birth rate will decline in succeeding years. In conclusion, co-efficient of robust ARIMA model compared with the classical ARIMA model and resulted the best ARIMA models to forecasts the birth rate of the state TamilNadu.

**Keywords:** Time series, ARIMA models, Birth Rate, Augmented Dickey-Fuller Test.

## 1 Introduction

Demographic statistics are essential for every country/program state's implementation, monitoring, and assessment, as well as, for research in a variety of sectors.Usually, demographic information is gathered from a variety of sources, including sample surveys, the civil registration system, and population censuses. Birth rate, death rate, infant mortality rate, and so, forth are the primary demographic data indicators. Every state's Civil Registration System (CRS) maintains an ongoing record of important occurrences in the population, such as births, deaths, marriages, and other life events. Those registration documents are most helpful for legal documentation and secondarily as a source of vital statistics.

TamilNadu is one of the well-developed states in South India. Fertility is the rate at which a population grows by having children, and it is typically measured by comparing the number of births to the size of a certain population segment. The most significant fertility indicators include the crude birth rate, age-specific fertility rate, total fertility rate, gross and net reproduction rates, and others. One indicator of fertility is the crude birth rate (CBR), which is correlated with the number of live births in the population. It is determined by dividing the total number of births in a given year by the total number of people who have ever lived in that year at the midpoint (or average). In this study, TamilNadu crude birth rate for 70 years from 1950 to 2019 have been considered. The source of the data is Civil Registration System.

**Objective of the Study:**

- To check the stationarity of the given dataset using Augmented Dickey Filler Test
- To fit the best model ARIMA (p, d, q)
- Plot Auto Correlation Function and Partial Auto Correlation Function and to results the correlation with the lags
- Forecasts the birth rate for next 30 years using fitted ARIMA (p, d, q)
- Compare the Classical and Robust ARIMA models and using the co-efficients of the models results the best model of ARIMA

*Corresponding author e-mail: dkarunanidhi001@gmail.com

## 2 Methodology

Time Series Analysis:

One of the strongest statistical analysis techniques for modelling and illuminating time-dependent successions of data is time series analysis. The next step in applying a model, time series forecasting predicts future values based on previously known values. The best method to fit a model and forecast future time series data is ARIMA.

ARIMA Models**:**

An ARIMA model is a sort of statistical approach used to analyse and predict time series data. In Auto-Regressive Integrated Moving Average, both the integration notation and Simple Auto Regressive Moving Average (ARMA) are generalised (ARIMA). Three parameters—p (lag order), d (degree of differencing), and q—are used in ARIMA models (order of moving average). The Box-Jenkins method is frequently used in the various ARIMA iterations.

Box Jenkins Method**:**

The integrated autoregressive moving average time series models can be found, fitted, checked, and used using the Box-Jenkins technique. Given by, is the universal Box-Jenkins ARIMA (p, d, q) model for m.

$$m_t = \delta_1\, m_{t-1} + \delta_2\, m_{t-2} + \cdots + \delta_p\, m_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2\, a_{t-2} - \cdots\ \theta_q\, a_{t-q} \ldots\ldots (1)$$

Where and are unknown parameters, and are identically independent normal errors with a mean of zero, q is the number of lagged values of the error terms indicating the order of the model's moving averages (MA) dimensions, p is the number of delayed values of and it represents the order of auto-regressive (AR) dimensions, d is the number of times m is varied, and so forth.

The three procedures of identification, estimation, testing, and application that the ARIMA model goes through in order to provide a forecast are referred to as "tenure integrated."

**Identification Model:** The Box-Jenkins approach is based on the assumption that the time

series data is stationary. The mean, variance (or standard deviation), and auto-covariance are constant in stationary data. If the first order not satisfied the stationarity conditions that means weak stationary or second – order. The data is non-stationary then Box-Jenkins procedure recommended differencing the data one or more times to attain the stationarity.

Selecting the proper (p, q) values of the ARMA model for the stationarity series is done at the identification stage based on two characteristics, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) (Auto-correlation Function).

**Estimation and Testing the Model:** In Box-Jenkins methods are using many approaches for estimation such as, Ordinary Least Square estimation, Maximum Likelihood Method, Non-Linear estimation method and Moment's methods and so on. The adequate model for the given data predicted using the error measures such Root Square Error (RSE), Mean Absolute Error (MAE), Akaike Information Criterion (AIC), Hannan-Quninn Criterion (HQC) and Bayesian Information Criterion (BIC) and so on.

In this work, AIC and BIC are evaluated and matched to the best of ARIMA (p, d, q).

**Akaike's Information Criterion (AIC):**

AIC determines the variance of the model residuals, which is affected by the number of estimated parameters. The AIC model with the minimum mean squared error and least number of coefficients. Typically, ARIMA models make a choice using AIC measures.

AIC = $2\ln(RMSE) + \frac{2C}{n}$

$$Where\ RMSE = \sqrt{\frac{\sum_{t=1}^{n} a_t^2}{n-c}}$$

Where c is the number of parameters, n is the number of observations, and t is the time period. The error that is the most straightforwardly calculated is called RMSE. The fitted model's estimated co-efficient count is denoted by the letter C.

**Bayesian Information Criterion (BIC):**

A statistical metric for contrasting and comparing time series models is the Bayesian Information Criterion, often known as the Schwarz Criterion. It and the AIC measure are closely connected. When parameter values were increased, BIC produced the best goodness of fit to the model.

BIC is provided by (mathematically),

$$BIC = -2\,\frac{t}{n} + \frac{k.\ \ln n}{n}$$

The minimum value of BIC resulted the best goodness of fit among number of models.

**Test for Stationarity: Augmented Dickey-Fuller (ADF) test:**

The enhanced Dickey Fuller test is one of the best ways to detect whether or not a time series of data is stationary. The time series data in this study are not stationary, according to the null hypothesis (H0), and they are stationary, according to the alternative hypothesis (H1). By appropriately differencing the data in $d^{th}$ order and applying the ADF test to the differenced model, the hypothesis is evaluated.

**Ljung-Box Test:**

In time series analysis, the Ljung-Box Test is playing a vital role. It is the test that explicit the presence of auto correlation for the series of estimated statistical model.

Ljung-Box test (Q) = $\frac{n(n+2)\sum p_k^2}{(n-k)}$

The test statistic resulted that the residuals of the estimated model independently distributed or not using the p values.

**Robust ARIMA models:**

In time series analysis, the robust methodology is appropriate method for forecasts the smoothened data and it is suitable for forecasts the univariate time series in the presence of outliers. In robust techniques, the parameters are robustified. The robust filtering method prevents the impact of one outlier from spreading to a number of consecutive residuals.

## 3 Results and Discussion:

The data for the study taken from the official website of Government of India crsorgi.gov.in and considered the birth rate for the state TamilNadu from 1950 to 2019. Using the past known data, 25 years of future data prediction will attain with the reliable models.

**Table.1. Birth Rate of TamilNadu (1950-2019)**

| Year | Birth rate | Year | Birth rate | Year | Birth rate | Year | Birth rate | Year | Birth rate |
|---|---|---|---|---|---|---|---|---|---|
| 1950 | 39.9 | 1961 | 38.5 | 1972 | 32.4 | 1983 | 27.8 | 1994 | 19.2 |
| 1951 | 39.2 | 1962 | 38.4 | 1973 | 30 | 1984 | 28 | 1995 | 20.3 |
| 1952 | 38.7 | 1963 | 35.3 | 1974 | 29.2 | 1985 | 24.7 | 1996 | 19.5 |
| 1953 | 39.5 | 1964 | 38.4 | 1975 | 30.7 | 1986 | 23.8 | 1997 | 19 |
| 1954 | 40.2 | 1965 | 32.8 | 1976 | 30.7 | 1987 | 24 | 1998 | 19.2 |
| 1955 | 39.8 | 1966 | 35.3 | 1977 | 29.8 | 1988 | 22.7 | 1999 | 19.3 |
| 1956 | 39 | 1967 | 34.5 | 1978 | 28.8 | 1989 | 23.1 | 2000 | 19.3 |
| 1957 | 38.8 | 1968 | 31.9 | 1979 | 28.9 | 1990 | 21.6 | 2001 | 19 |
| 1958 | 38.4 | 1969 | 27.19 | 1980 | 27.9 | 1991 | 20.8 | 2002 | 18.5 |
| 1959 | 40.8 | 1970 | 30 | 1981 | 28 | 1992 | 20.7 | 2003 | 18.3 |
| 1960 | 41.7 | 1971 | 31.4 | 1982 | 27.7 | 1993 | 19.5 | 2004 | 17.1 |

**Table.1. Birth Rate of TamilNadu (1950-2019) (Cont..)**

| Year | Birth rate | Year | Birth rate |
|---|---|---|---|
| 2005 | 16.5 | 2016 | 15 |
| 2006 | 16.2 | 2017 | 14.9 |
| 2007 | 15.8 | 2018 | 14.7 |
| 2008 | 16 | 2019 | 14.2 |
| 2009 | 16.3 | | |
| 2010 | 15.9 | | |
| 2011 | 15.9 | | |
| 2012 | 15.7 | | |
| 2013 | 15.6 | | |
| 2014 | 15.4 | | |
| 2015 | 15.2 | | |

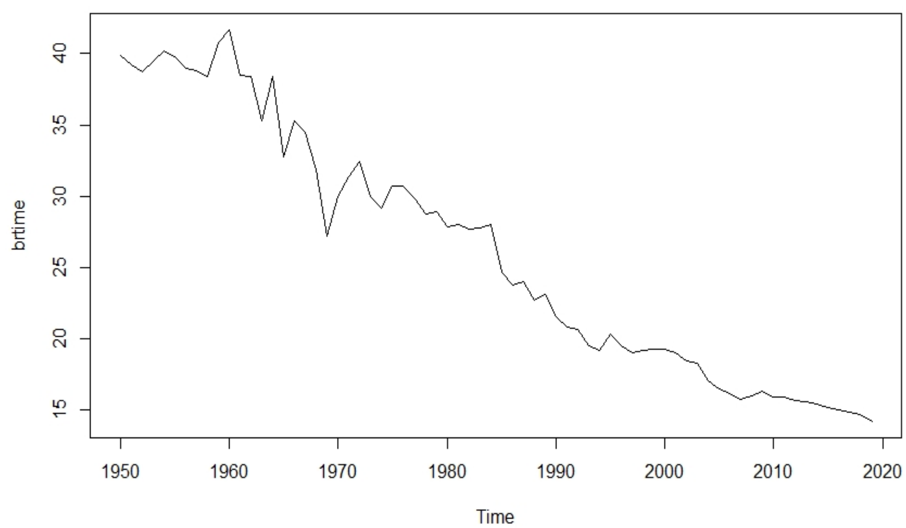


**Fig.1 The graphical presentation of Birth Rate of TamilNadu (1950-2019)**

**Table 2 Auto-Correlation Function (ACF) co-efficients with Lag values for Birth Rate of TamilNadu**

| Lag Value | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| ACF | 1.000 | 0.955 | 0.920 | 0.884 | 0.852 | 0.809 | 0.767 | 0.724 | 0.683 | 0.638 |
| ACF _df | 1.000 | -0.620 | 0.225 | -0.316 | 0.334 | -0.151 | 0.074 | -0.125 | 0.190 | -0.091 |
| Lag Value | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | |
| ACF | 0.590 | 0.543 | 0.500 | 0.451 | 0.408 | 0.361 | 0.331 | 0.293 | 0.252 | |
| ACF _ df | -0.143 | 0.129 | 0.015 | 0.065 | -0.107 | -0.043 | 0.019 | 0.134 | -0.092 | |

**Table 3 Partial Auto-Correlation Function (PACF) co-efficients with Lag=18 values for Birth Rate of TamilNadu**

| Lag Value | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| PACF | 0.955 | 0.083 | -0.009 | 0.022 | -0.132 | -0.040 | -0.027 | -0.019 | -0.040 | -0.051 |
| PACF df | | -0.620 | -0.258 | -0.522 | -0.306 | -0.200 | -0.190 | -0.206 | 0.039 | 0.240 |
| Lag Value | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | |
| PACF | -0.078 | -0.016 | -0.001 | -0.075 | 0.039 | -0.075 | 0.152 | -0.057 | -0.097 | |
| PACF df | -0.135 | -0.160 | -0.194 | -0.179 | 0.014 | -0.054 | -0.244 | -0.135 | -0.015 | |

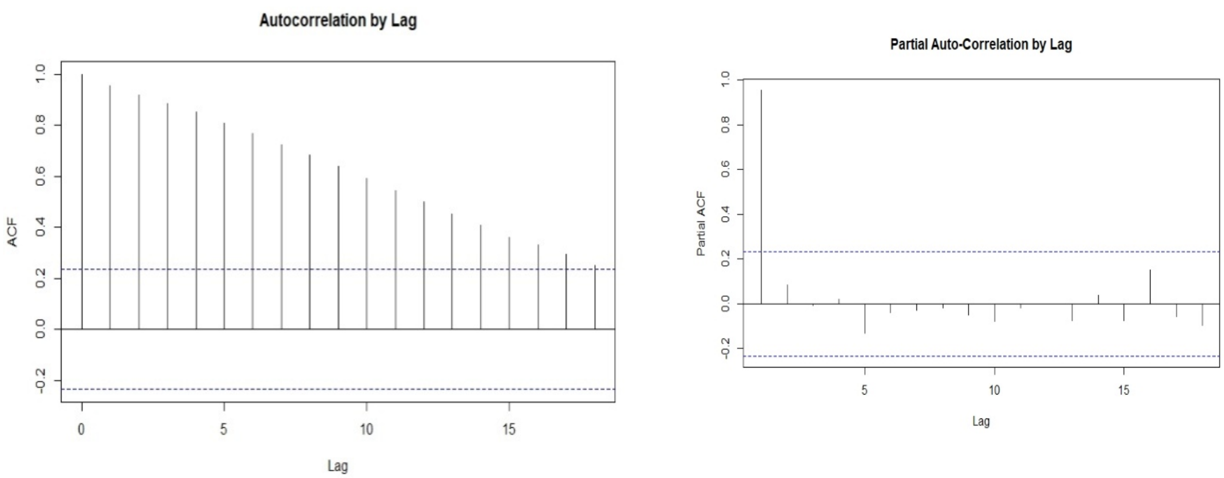


**Fig 2. ACF and PACF at the Lag=18 for Birth Rate of TamilNadu (1950-2019)**

From Fig.1. Shows the downward decreasing trend and it shows that the given time is non-stationary. Fig. 2. shows the plots of ACF and PACF respectively. The autocorrelation values are progressively declining from a first order auto-correlation co-efficient till to the termination. In Table 4, ADF test concluded that the acceptance of Null hypothesis, (i.e.) the Birth Rate of TamilNadu (1950-2019) is follows non-stationarity.

**Differencing the data series to reach the stationarity:**

The Box-Jenkins approach requires stationary time series data. To examine the stationarity of the time series data, a plot correlogram is utilised. Whether the data's time series is stationary or not was determined as a consequence. In the event that stationarity is not achieved, the differencing approach is used to get rid of changes in the series, producing a constant mean and 3.d variance over time. The transferred series after first order differencing is shown in Fig. 3, and the ACF and PACF plots after first order differencing are shown in Figs. 4 and 5, respectively.

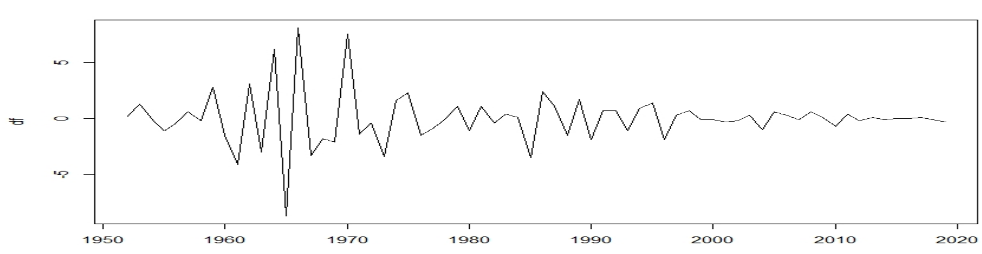


**Fig 3. First Differencing of the Time-series plot**

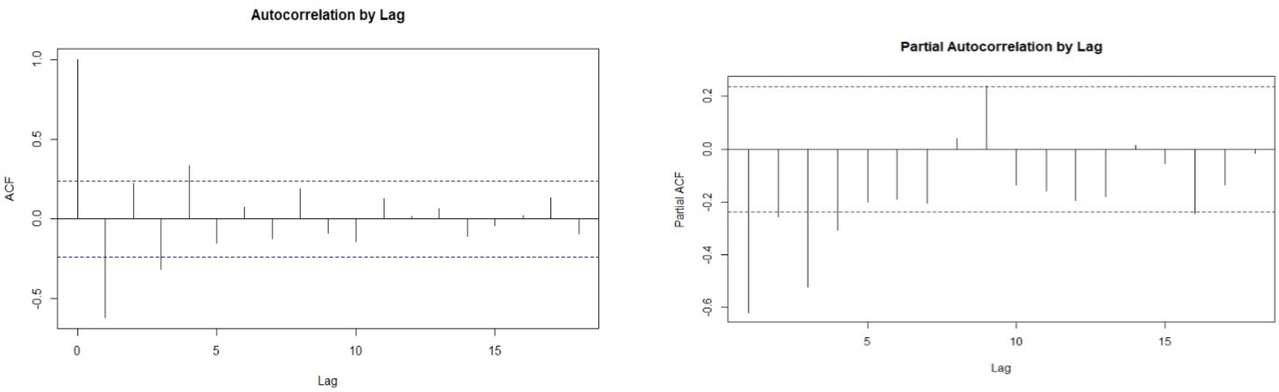


**Fig.4. ACF and PACF Plots for transformed data of the series**

The ADF unit root test is reported in this study under the hypothesis as:

Birth rate is non-stationary, rejecting the null hypothesis

Alternative Theory: The Birth Rate Is Stable

**Table. 4. Augmented Dickey-Fuller Test (ADF) to check stationarity of Birth Rate of TamilNadu**

| | Time series data | Transformed series ($1^{st}$ order differencing) |
|---|---|---|
| **ADF -Test statistic** | -1.5818 | -6.9445 |
| **Lag order** | 4 | 4 |
| **p-value** | 0.7453 | 0.01 |
| **Level of Significance** | 5% | 5% |

From Table.4 it is concluded that the stationarity attained by the $1^{st}$ order differentiated data with the 5% level of significance. The alternative hypothesis is accepted (i.e.) $p<0.05$, the alternative hypothesis accepted. Then the transformed time series data (with $1^{st}$ order differencing) is stationary.

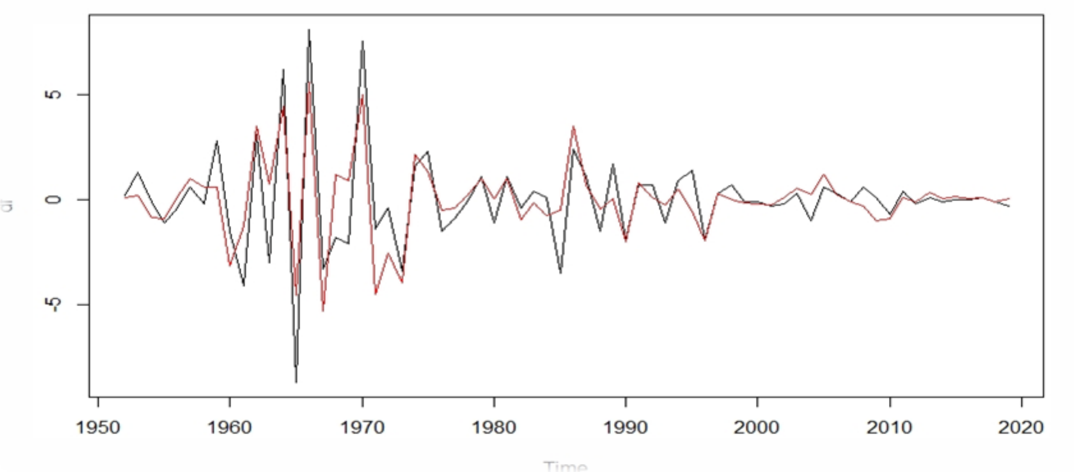


**Fig.7. The fitted model of ARIMA (0,1,1) compared with the transformed data with 1st order differencing**

In this study, the stationary has been attained after the 1st order differencing of the birth rate of TamilNadu. Here, RMSE and MAE are the two measures to find the accuracy of the ARIMA models and it is also resulted the goodness of fit to the model. AIC and BIC values are used to find the appropriate model of the transformed data.

**Table .5 .ARIMA Model selection criteria**

| Models | RMSE | MAE | AIC | BIC |
|---|---|---|---|---|
| ARIMA (1,1,1) | 1.8370 | 1.1644 | 284.23 | 290.84 |
| ARIMA (1,1,0) | 2.7308 | 1.7954 | 330.58 | 334.98 |
| ARIMA (2,1,0) | 2.5490 | 1.5938 | 284.30 | 330.24 |
| ARIMA (2,1,1) | 1.7740 | 1.0941 | 283.13 | 290.95 |
| ARIMA (2,1,2) | 1.3844 | 0.9250 | 257.92 | 268.95 |
| ARIMA (3,1,0) | 2.1451 | 1.4027 | 303.46 | 312.28 |
| ARIMA (3,1,1) | 1.5059 | 1.0297 | 263.90 | 274.92 |
| ARIMA (3,1,2) | 1.3379 | 0.9024 | 253.80 | 267.03 |
| ARIMA (4,1,0) | 1.9005 | 1.2580 | 290.18 | 301.20 |
| ARIMA (4,1,1) | 1.4273 | 0.9720 | 259.64 | 272.87 |
| ARIMA (4,1,2) | 1.3023 | 0.9219 | 254.20 | 269.95 |
| ARIMA (5,1,0) | 1.7780 | 1.1569 | 283.89 | 297.12 |

| | | | | |
|---|---|---|---|---|
| **ARIMA (5,1,1)** | **1.3001** | **0.9005** | **252.36** | **265.68** |
| ARIMA (5,1,2) | 1.3001 | 0.9179 | 256.06 | 273.70 |

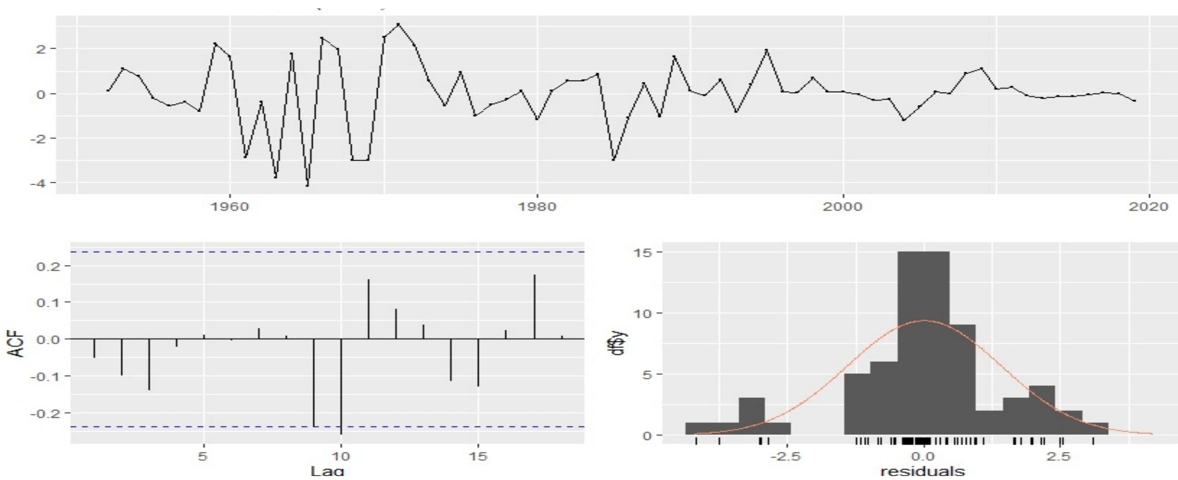


**Fig.8 Residuals from ARIMA (5, 1 ,1) with zero mean**

**Table.6. Estimated model parameters for the ARIMA (5, 1, 1) method**

| Variables | Co-efficient | SE |
|---|---|---|
| Constant | -12.08 | 1.3943 |
| AR (1) | -1.1413 | 0.1185 |
| AR (2) | -1.0396 | 0.1707 |
| AR (3) | -0.9762 | 0.1728 |
| AR (4) | -0.5239 | 0.1678 |
| **AR (5)** | **-0.1987** | **0.1152** |
| MA (1) | -0.92 | 0.0439 |

AIC and BIC values were used to calculate the estimated ARIMA model. To determine the goodness of fit for the estimated models, the AIC and BIC values are utilised. As the best fit to the converted (1st order differenced data) Birth rate of Tamil Nadu, ARIMA (5,1,1) is derived from Tab. 5. Using metrics like AIC, BIC, MAPE, and MAE values, which are the least among the number of models, the best model in this study, ARIMA (5,1,1), has been determined. The model ARIMA (5,1,1) follows normality, with mean 0 and variance 1, and the ACF and estimated models are similarly presented in Fig.8.

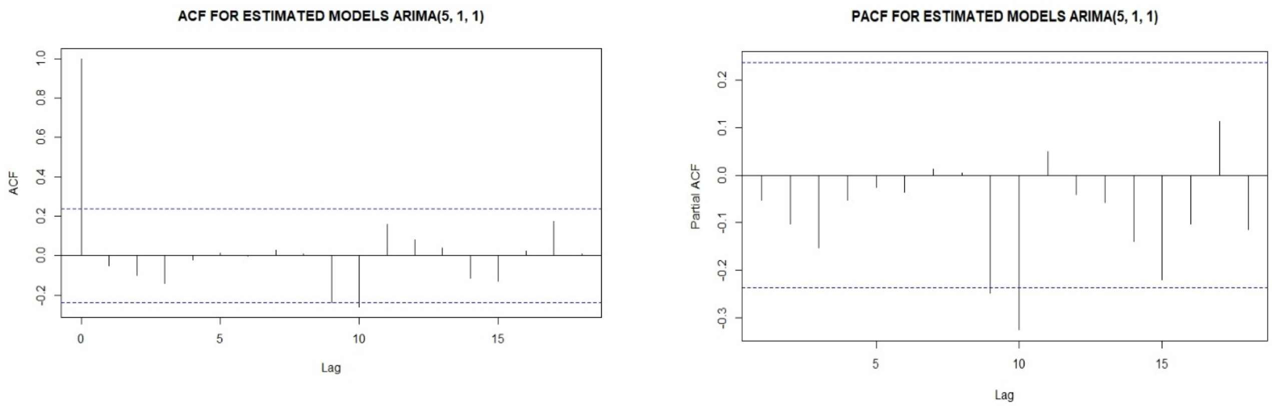


**Fig. 9. ACF and PACF for the ARIMA estimated model (5,1,1)**

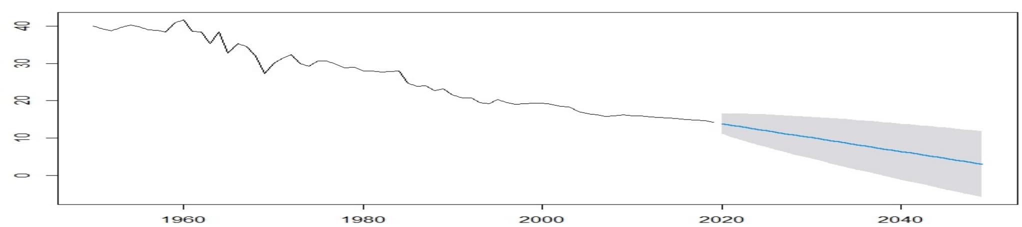

**Fig. 10 Forecasts from ARIMA (5, 1, 1)**

**Table.7. Forecasted values of Birth Rate of TamilNadu for 30 years**

| Year | Forecasted | 95% confidence Level | |
|---|---|---|---|
| | | Lower Level | Upper Level |
| 2020 | 13.8234 | 11.1529 | 16.4937 |
| 2021 | 13.4515 | 10.3567 | 16.5462 |
| 2022 | 13.0796 | 9.6122 | 16.5470 |
| 2023 | 12.7077 | 8.9039 | 16.5115 |
| 2024 | 12.3358 | 8.2230 | 16.4485 |
| 2025 | 11.9639 | 7.5638 | 16.3639 |
| 2026 | 11.5920 | 6.9223 | 16.2617 |
| 2027 | 11.2201 | 6.2955 | 16.1447 |
| 2028 | 10.8483 | 5.6813 | 16.0152 |
| 2029 | 10.4764 | 5.0778 | 15.8748 |
| 2030 | 10.1045 | 4.4840 | 15.7249 |
| 2031 | 9.7326 | 3.8986 | 15.5665 |

| | | | |
|---|---|---|---|
| 2032 | 9.3606 | 3.3208 | 15.4006 |
| 2033 | 8.9887 | 2.7497 | 15.2279 |
| 2034 | 8.6168 | 2.1847 | 15.0490 |
| 2035 | 8.2450 | 1.6255 | 14.8645 |
| 2036 | 7.8731 | 1.0714 | 14.6748 |
| 2037 | 7.5012 | 0.5219 | 14.4805 |
| 2038 | 7.1293 | -0.0229 | 14.2816 |
| 2039 | 6.7574 | -0.5638 | 14.0787 |
| 2040 | 6.3855 | -1.1009 | 13.8720 |
| 2041 | 6.0137 | -1.6344 | 13.6617 |
| 2042 | 5.6417 | -2.1645 | 13.4481 |
| 2043 | 5.2698 | -2.6915 | 13.2319 |
| 2044 | 4.8979 | -3.2155 | 13.0115 |
| 2045 | 4.5260 | -3.7368 | 12.7890 |
| 2046 | 4.1542 | -4.2554 | 12.5638 |
| 2047 | 3.7823 | -4.7712 | 12.3369 |
| 2048 | 3.4104 | -5.2852 | 12.1059 |
| 2049 | 3.0836 | -5.7955 | 11.8736 |

From Tab. 7 and Fig.10, we concluded that the birth rate of the state TamilNadu may decline from 14% (2019) to 3% (2049). If the birth rate is decreases, the population size also decrease for the future days.

**Table. 8: Ljung-Box test to assess ARIMA's autocorrelation (5,1,1)**

| **Model** | **Number of Predictors** | **Ljung-Box Q(10)** | | |
|---|---|---|---|---|
| | | **Statistic** | **DF** | **p-value** |
| Birth Rate (df) | 1 | **19.035** | **10** | **0.03982** |

According to, Tab. 7, the ARIMA (5, 1, 1) model is independently distributed, proving that there is no correlation in the data. Thus, it can be said that ARIMA (5, 1, 1) is the best model to predict Tamil Nadu's birth rate.

**Robustified the ARIMA (p, q, d) model for the Birth Rate TamilNadu:**

In this paper, we planned to study the robustness of extrapolative recital of ARIMA models. The birth rate of TamilNadu has been taken to study the pattern and related measures calculated. The Robustified ARIMA (p, d, q) models are studied and found the model based on the outlier presence in the data. In Tab. 9 the measures of various Robust ARIMA models have been presented and the model ARIMA (3, 1, 1) which resulted with standard deviation 0.0217 and the degrees of freedom 70 with 65 residuals. The regression co-efficient of the model is -0.008 (intercept).

**Table. 9 Robust ARIMA (3,1,1) with relative outliers and its statistics**

| | | | | | | |
|---|---|---|---|---|---|---|
| **AR co-efficients** | AR (1) : -0.1640 | AR (2) : -0.2292 | AR (3) : -0.3691 | | | |
| **No.of outliers Detected** | 6 | | | | | |
| **Outlier Index** | 14 | 16 | 20 | 36 | 41 | 55 |
| **Outlier Type** | AO | LS | AO | LS | LS | LS |
| **Outlier Impact** | -0.0665 | -0.1510 | -0.0944 | -0.1301 | -0.0820 | -0.0818 |
| **Outlier t-statistics** | 3.3019 | 6.2863 | 5.0861 | 5.4177 | 3.8177 | 3.900 |

Using the Tab. 9, the ARIMA (3, 1, 1) model is used to forecast the data which was having the minimal outliers' measures with 65 residuals value. The test statistics revealed that there are no extraordinary forecasting errors present in outlier samples. Though, a classical forecasting which based on the past behaviour is not present in the study and robust ARIMA models better to compensate these type of anomalies in the contrast to ARIMA.

**Table.10. Robust ARIMA models with the co-efficients**

| **ARModel** | **AR (p)** | **Co-efficients** | **Std deviation** | **Outliers detected** | **Intercept (regression)** |
|---|---|---|---|---|---|
| ARIMA (1, 1, 1) | AR(1) | 0.1576 | 0.0278 | 11 | -0.01 |
| ARIMA (2,1,0) | AR(1) | 0.2990 | 0.0271 | 8 | - |
| | AR(2) | -0.065 | | | |
| ARIMA (2,1,1) | AR(1) | 0.2076 | 0.025 | 12 | -0.086 |
| | AR(2) | -0.1868 | | | |
| ARIMA (3,1,0) | AR(1) | 0.3351 | 0.0251 | 10 | -0.092 |
| | AR(2) | -0.1735 | | | |
| | AR(3) | 0.2879 | | | |
| **ARIMA (3, 1 ,1)** | **AR(1)** | **-0.1640** | **0.0217** | **6** | **-0.008** |
| | **AR(2)** | **-0.2292** | | | |
| | **AR(3)** | **-0.3961** | | | |
| ARIMA (4,1,1) | AR(1) | 0.2550 | 0.0235 | 11 | -0.0091 |
| | AR(2) | -0.2449 | | | |

| | | | | | |
|---|---|---|---|---|---|
| | AR(3) | 0.2255 | | | |
| | AR(4) | 0.0235 | | | |
| ARIMA(4,1,0) | AR(1) | 0.3221 | 0.0255 | 8 | |
| | AR(2) | -0.1713 | | | |
| | AR(3) | 0.2719 | | | |
| | AR(4) | -0.0383 | | | |
| ARIMA(5,1,0) | AR(1) | 0.2511 | 0.022 | 11 | |
| | AR(2) | -0.1388 | | | |
| | AR(3) | 0.3376 | | | |
| | AR(4) | 0.1589 | | | |
| | AR(5) | 0.2696 | | | |
| ARIMA (5,1,1) | AR(1) | -0.1055 | 0.0212 | 8 | |
| | AR(2) | -0.0867 | | | |
| | AR(3) | 0.1240 | | | |
| | AR(4) | 0.2400 | | | |
| | AR(5) | 0.3521 | | | |
| ARIMA (6,1,0) | AR(1) | 0.2486 | 0.022 | 11 | |
| | AR(2) | -0.1228 | | | |
| | AR(3) | 0.3618 | | | |
| | AR(4) | 0.1619 | | | |
| | AR(5) | 0.2765 | | | |
| | AR(6) | 0.0000 | | | |
| ARIMA (6,1,1) | AR(1) | 0.2500 | 0.0194 | 11 | |
| | AR(2) | -0.1362 | | | |
| | AR(3) | 0.2969 | | | |
| | AR(4) | 0.1015 | | | |
| | AR(5) | 0.2186 | | | |
| | AR(6) | -0.0924 | | | |

## Conclusions of the study:

In this study, the birth rate of TamilNadu has been taken into account to study the future birth rate of the state. The data taken from 1950 to 2019, nearby 71 past records considered and studied. Based on the relative measures of accuracy, it is found that the best model ARIMA (5, 1, 1). The estimated model ARIMA (5, 1 ,1) forecasted the future birth rate of the state TamilNadu for next 30 years. Main objective of the study fulfilled with robustness measures for ARIMA models studied and resulted that best model is ARIMA (3, 1, 1) with minimum standard deviation and outlier measures. A probable explanation would be the forecasts might have had additional knowledge for their prediction not extractable from the past observations. The findings of robust ARIMA model supports the basic ARIMA models to predict the best model with corrected or adjusted by the author using the additional knowledge of using the robustness.

The forecasted values of the state TamilNadu results the decreasing values and it may reflect in future population census. If the position continues in the state, the population size of the state may decrease in future. It is suggested that to the Government of TamilNadu should take necessary steps to reduce the infertility rates using various medical advancements.

## References


[1] Qais Mustafa Abdulqader, Time Series Forecasting Using Arima Methodology with Application on Census Data in Iraq, *Journal of University of Zakho*, **4(A), No.2**, 258-268, 2016.

[2] R. Swaminathan, V. Shanta, J. Ferlay, S. Balasubramanian, F. Bray, R. Sankaranarayanan, Trends in cancer incidence in Chennai city (1982–2006) and statewide predictions of future burden in Tamil Nadu (2007–16), *The National Medical Journal of India*, **24(2),** 2011

[3] Ntebogang Dinah Moroke, The robustness and accuracy of Box-Jenkins ARIMA in modeling and forecasting household debt in South Africa, *Journal of Economics and Behavioral Studies*, 6(9), 748-759, 2014

[4] Buah Ahoba Masha, Zigli David Delali and Annan Wobir Reuben, Time Series Analysis of Monthly Birth Rate in Ghana, *International Journal of Applied Science and Mathematics*, 7(5), 2394-2894, 2020.

[5] Andres Hernandez-Matamoros, Hamido Fujita, Toshitaka Hayashi and Hector Perez-Meana, Forecasting of COVID19 per regions using ARIMA models and polynomial functions, *Applied Soft Computing*, 96(2), 2020.

[6] Azhari A. ELHAG a and Hanaa Abu-Zinadah, Forecasting under Applying Machine Learning and Statistical Models, *Forecasting under Applied Statistical and Multilayer Thermal Science*, 24(1), 131-137 (2020).

[7] Dr.S.Appavu Alias Balamurugan, Balaji Ganesh Rajagopal, and Dr. Kumar Parasuram, A Robust Framework to Detect Moving Vehicles in Different Road Conditions in India, *Journal of Theoretical and Applied Informtion Technology*, 97 (01), (2019).

[8] Arul Earnest, Sue M Evans, Fanny Sampurno and Jeremy Millar, Forecasting annual incidence and mortality rate for prostate cancer in Australia until 2022 using autoregressive integrated moving average (ARIMA) models, *BMJ Open.* 9(8): e031331, 2019.

[9] K. Senthamarai Kannan and E. Sakthivel, Fuzzy Time Series Model and ARIMA Model -A Comparative Study, *Indian journal of Applied Research*,.4(8),2014.

[10] Kumar Manoj and Anand Madhu, An Application of Time series ARIMA Forecasting Model for predicting sugarcane Production in India, *Studies in Business and Economics* 9(1):81 – 94.

[11] Ewa Chodakowska, Joanicjusz Nazarko and Laq1ukasz Nazarko, ARIMA Models in Electrical Load Forecasting and Their Robustness to Noise, *Energies*, 14, 7952,2021.