Mathematical Forum ISSN: 0972-9852
Vol. 30, 2023

A TIME SERIES ANALYSIS OF COVID-19 DEATH RATES DURING SECOND WAVE IN INDIA AND IN THE STATE OF ASSAM USING ARIMA MODEL

KUKI KALPITA MAHANTA[1] and NITYARAJ CHETIA[2]

[1] *Department of Statistics, Dibrugarh University, Dibrugarh-786004, Assam, India*

[2]*Department of Statistics, Dibrugarh University, Dibrugarh-786004, Assam, India*

Email: [1]kalpitamahanta05@gmail.com , [2] nityarajchetia454@gmail.com

Received on: 18/01/2023
Accepted on: 08/09/2023

Abstract

Covid-19, the most threatening virus that was discovered in the year 2019 has spread its influence all across the world and caused millions of death. Till now the whole world has faced two perilous waves of this pandemic and therefore in a way it has become very important for us to analyse the COVID cases to predict our future days. So this research work intends to study through time series analysis about the death rates of India along with the state Assam and then make a comparison among them for a time period under the second wave of COVID-19 pandemic. ARIMA Methodology is employed here to forecast the future death rates and in addition the forecasted death rates are also compared with the original death rates. According to the diagnostic check analysis the models are adequately fitted to the death rate data. Our selected model gives us forecasted death rates for next one month and up to a large extent they match up with the original death rates. So in future, the government and health workers can take the help of these models to predict future covid-19 death rates and it might prove substantially beneficial for them.

Keywords: COVID-19 death rate, ARIMA model, forecast, Assam, India.

2010 AMS classification: 37M10

1. Introduction

Over the last two years the scenario of the people of the whole world has changed due to an infectious virus known as COVID-19. The virus was discovered in the last part of the year 2019 and affected more than 26.1 crore people across the globe, and caused almost more than 51.9 lakh of deaths. In India, more than 3.4 crore cases were reported out of which more than 4.6 lakh deaths occurred during the last two years. Since it is a newly discovered disease, therefore medical science without a doubt has faced varied challenges to control it. Moreover, also due to the lack of awareness of people, the virus has penetrated into the lives of people thereby impacting the large number of people. Now constant effort has been made by health workers, and at the same time different awareness programmes too have been conducted by the government. With the coming vaccination programme, the worsening situation caused due to Covid-19 has seemed to come under control. But according to some scientists, as some situations have gone beyond control, therefore one must now have to adapt with the virus or must learn to survive with the virus for the coming years. The death rate of COVID patients of a country indicates the clear picture of the medical facility of the country and also at the same time shows to what extent the government of a particular country has succeeded to control the disease. During the second wave of COVID pandemic, after United States, India became the country with the second highest COVID cases in the world. So a very immediate necessity has come up to analyse COVID cases, especially in the context of India.

Tandon et al. [9] developed a time series model to forecast the future COVID-19 cases in India. Salgotra et al. [7] also proposed a time series model based on genetic programming for the analysis of confirmed and death cases across three most pretentious states of India- Maharashtra, Gujarat and Delhi as well as for whole India. Khan & Gupta [4] used an ARIMA model to predict the COVID-19 infected cases in upcoming days in India on the data collected from 31st January 2020 to 25th March 2020 and verified it using the data collected from 26th March 2020 to 04th April 2020. Singh et al. [8] produced a prediction of confirmed, deceased and recovered cases using the SVM model covering all the COVID cases from 22nd January, 2020 to 25th April, 2020 across worldwide. Maleki et al. [6] used a family of models, Autoregressive time series models based on two-piece scale mixture normal distributions to analyze the real world time series data of confirmed and recovered COVID-19 cases. Kumar & Susan [5] modelled and forecasted the COVID-19 outbreak using ARIMA methodology and Prophet time series forecasting of whole world and 10 mostly affected countries taking the data from the period 22nd January, 2020 to 20th May, 2020.

The medical facilities in the north-eastern states in India are not up to the mark as compared to other states of the country. Assam is the most populous and most developed state among the North-Eastern states of India. Majority of its population lives in rural area and the state government as well as central government take many schemes to improve the overall health facility of the state. Since time series analysis and forecasting of death rate and recovery rate of COVID cases helps in planning and monitoring of the disease, therefore, this research work was done to study the time series analysis of the death rates of India along with the state Assam and made a comparison among them for a time period under second wave of COVID-19 pandemic.

2. Methodology:

The specific objectives of the present study are to determine the trend of death rates, to detect of the structural break points of death rates, testing of stationarity of the death rates and to choose a best fitted ARIMA model to predict covid-19 death rates for India along with the state Assam and made a comparison among them for a time period under second wave of COVID-19 pandemic.

The data used in this study are secondary in nature. The COVID-19 death cases are collected from Aarogya Setu App by Government of India for the period 4th April to 31st August, 2021 and are converted to death rates. Jarque–Bera [3] test is used here to check the normality of the data. To detect the presence of structural break points F- test is performed here and to know about the stationarity of the data graphically Auto Correlation Function (Acf) and Partial Auto Correlation Function (Pacf) are plotted here and to test it statistically Augmented Dickey-Fuller [2] tests are also used here. To forecast the death rates, ARIMA methodology introduced by Box and Jenkins [1] is used.

In a ARIMA (p, d, q) model, p denotes the number of auto regression parameters, d denotes the order of differentiation and q denotes the number of parameters of moving average. A simple ARIMA model is shown below-

$$\varphi(B)(1-B)^d y_t = \mu + \theta(B)e_t \qquad (1)$$

$\phi(B)$ and $\theta(B)$ polynomials are the autoregressive and moving average components of orders p and q respectfully.

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - .... - \varphi_p B^p \qquad (2)$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots\ldots - \theta_q B^q \qquad (3)$$

The models choice criterias like AIC, Log-likelihood has also been used here to choose the best fitted model. For the diagnostic checking, the significance of the AR and MA coefficients are carried out through Z test. For model accuracy, the values of Error measures like Mean Error (ME), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) etc. are calculated here. To check the normality of the residuals of the ARIMA Models, "Normal Q-Q plot" is also constructed here.

3. Analysis of Data:

In this paper an attempt has been made to analyse the collected data related to the COVID-19 cases in India and in the state Assam from April to August 2021. In this regard, the methodology as mentioned in section 2 has been used and accordingly conclusions have been drawn. During the 154 days of study period from 4th April to 31st August, 2021 the death rate is minimum for India on 25th April (1.103105 percent) and for Assam on 11th May (0 percent). And it is maximum for India on 6th July (1.340588 percent) and for Assam on 30th July (0.501881 percent). The average death rate for India is 1.257238 percent and for Assam is 0.922391 percent. The standard deviation of the death rates data for India is 0.087552 and for Assam it is 0.029451.

According to the Jarque-Bera Test results the data were normal in nature. Fig-1 and Fig- 2 shows the time series plot of the data and they exhibit stochastic trends.

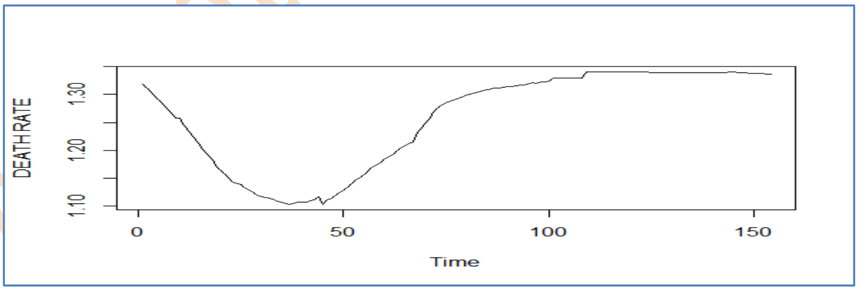

Fig. 1: Time series plot of death rates of India from April to August, 2021.

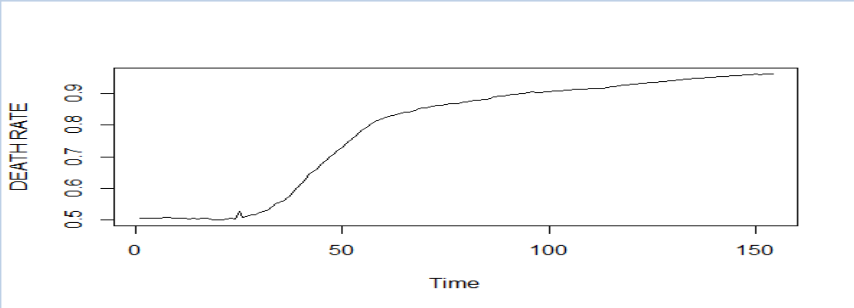

Fig. 2: Time series plot of death rates of Assam from April to August, 2021.

For the detection of the structural break points in the data, F test is performed here the results of the test are shown in Table 1. The p-value is less than 0.05, which means existence of Structural Break points and the R software identifies the structural Break points in the data. For India, the break points are as follows: on 25th April, it was 0.1493506 percent, on 19th May it got increased to 0.3051948 percent, then on 11th June the percentage was exceeded to 0.4545455 percent and then finally on 6th July, the percentage reached to 0.6168831 percent. Likewise for Assam, the breakpoints are: on 11th May (0.2532468 percent), on 3rd June (0.4025974 percent), on 26th June (0.5519481 percent) and on 30th July (0.7727273 percent).

Table 1: F-test for Structural Brake Points

Test For	Hypothesis	Test Statistic	p-value
India	H_0: No of Structural Break points exists H_1: At least one Structural Break points exists	F = 441.82	p-value < $2.2\text{x}10^{-16}$
Assam	H_0: No of Structural Break points exists H_1: At least one Structural Break points exists	F = 1178	p-value < $2.2\text{x}10^{-16}$

To know about the stationarity of the data graphically, the Auto Correlation Function (Acf) and Partial Auto Correlation Function (Pacf) are plotted here and are shown from Fig 3 to Fig 6. Since in Fig 3 and Fig 5 most of the bars have traversed the upper limit, and in Fig 4 and Fig 6 only one bar has traversed the upper limit, thus we may conclude that data is not stationary in nature. For statistical significance Augmented Dickey-Fuller Test was used here and the results are given in Table 2.

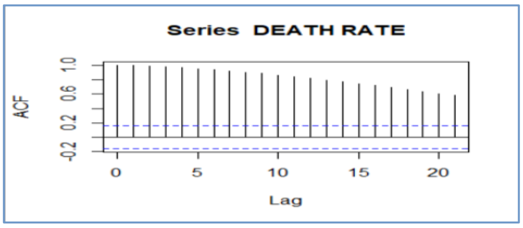

Fig. 3: Auto Correlation function (Acf) for death rates of India.

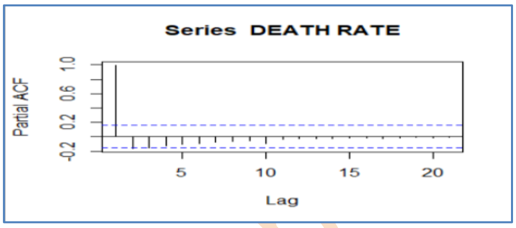

Fig. 4: Partial Auto Correlation function (Pacf) or death rates of India.

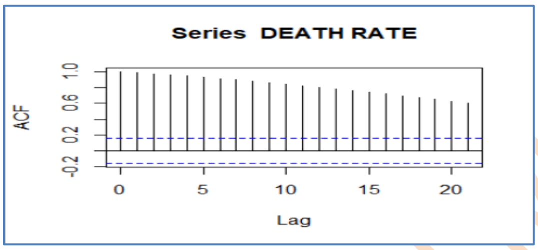

Fig. 5: Auto Correlation function (Acf) for death rates of Assam.

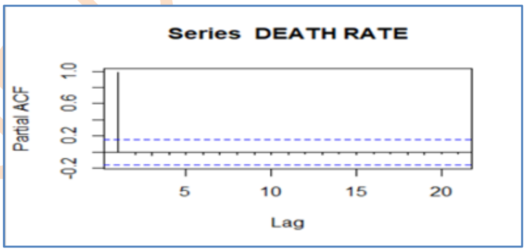

Fig. 6: Partial Auto Correlation function (Pacf) for death rates of Assam.

Table 2: Augmented Dickey-Fuller Test

Test For	**Hypothesis**	**Test Statistic**	**Lag order**	**p-value**
India	H_0: Stationary data H_1: Non-Stationary data	Dickey-Fuller= -3.3401	5	0.06714
Assam	H_0: Stationary data H_1: Non-Stationary data	Dickey-Fuller = -2.6068	5	0.3235

As the p-value greater than 0.05, indicates non- stationarity of data. We have to take the help of simple differentiation to make the data stationary.

Thus, New death rate = $Y_t - Y_{t-1}$ (4)

After the differentiation, the new death rate displays the pattern as shown in fig-7 &8. Now, we have again plotted the Auto Correlation Function (Acf) and Partial Auto Correlation Function (Pacf) and are depicted through Fig-9 to 12. In the Acf plot most of the bars are within their limits, so now we may conclude that after taking the differences, the data becomes stationary. Statistically we can show that by using Augmented Dickey-Fuller Test and Table 3 shows the test results.

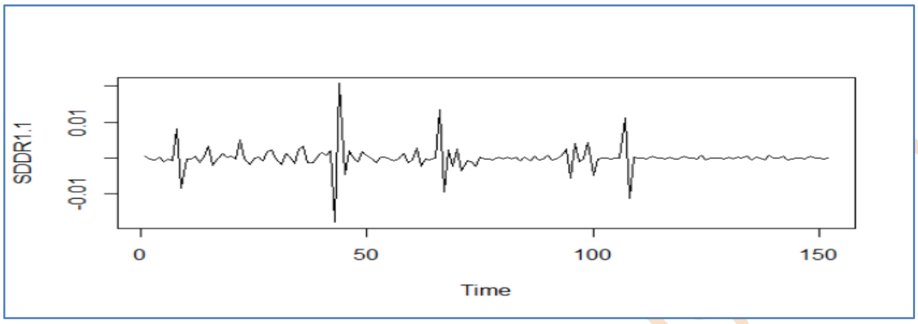

Fig. 7: Time series plot of death rate for India after taking their differences

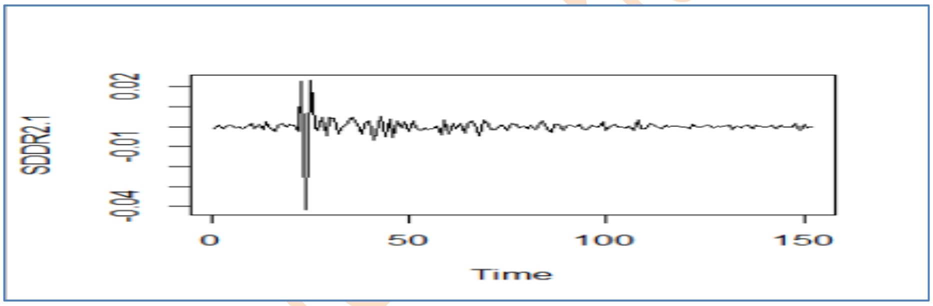

Fig. 8: Time series plot of death rate for Assam after taking their differences

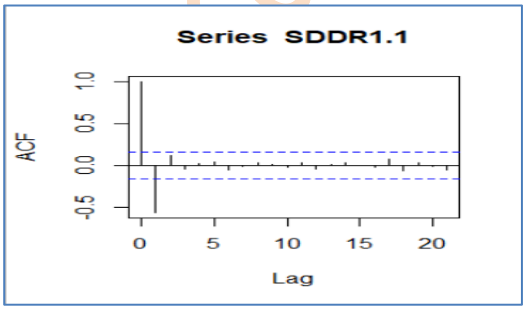

Fig. 9: Auto Correlation Function (Acf) for the death rates of India after taking their differences.

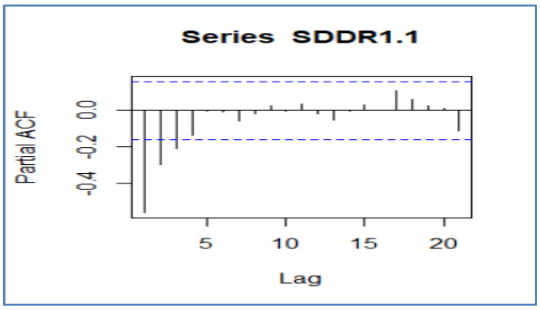

Fig. 10: Partial Auto Correlation Function (pacf) for the death rate of India after taking their differences.

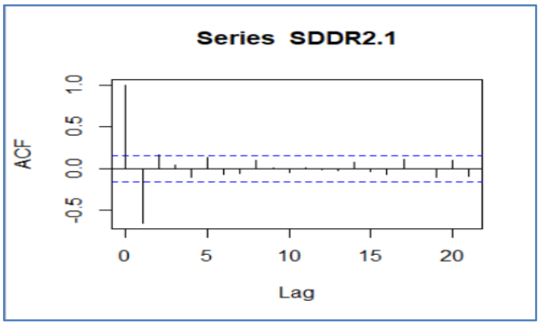

Fig. 11: Auto Correlation Function (Acf) for the death rates of Assam after taking their differences.

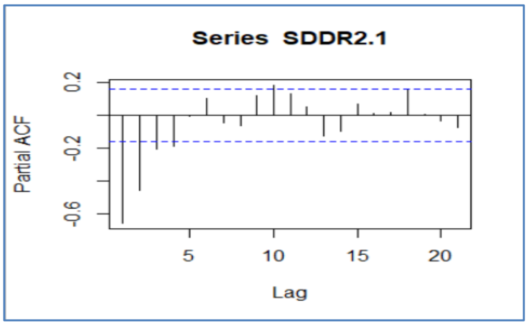

Fig. 12: Partial Auto Correlation Function (pacf) for the death rate of Assam after taking their difference

Table 3: Augmented Dickey-Fuller Test

Test For	Hypothesis	Test Statistic	Lag order	p-value
India	H_0: Stationary data H_1: Non-Stationary data	Dickey-Fuller = -6.3488	5	0.01
Assam	H_0: Stationary data H_1: Non-Stationary data	Dickey-Fuller = -6.3488	5	0.01

In Table 3 the p-value is less than 0.05, indicates stationarity of the data. Since after taking the first difference the data becomes stationary so the order of the 'd' is fixed as 1 in the ARIMA(p,d,q) model. Keeping it constant, five ARIMA models are recommended here to fit the data and among them the finest model will be selected according to some model choice criterias. The models are presented in Table 4.

Table 4: AIC and Log Likelihood of the fitted ARIMA models

For	Model	ARIMA Order	AIC	Log likelihood
India	**Model 1.1**	(1,1,0)	-1059.03	531.51
	Model 1.2	(0,1,1)	-1067.4	538.7
	Model 1.3	(1,1,1)	-1065.56	535.78
	Model 1.4	(2,1,1)	-1065.56	536.78
	Model 1.5	(2,1,2)	-1064.03	537.01
Assam	**Model 2.1**	(1,1,1)	-1065.56	535.78
	Model 2.2	(0,1,3)	-1065.01	536.51
	Model 2.3	(3,1,1,)	-1064.5	537.25
	Model 2.4	(2,1,3)	-1064.74	538.37
	Model 2.5	(0,1,1)	-1067.4	539.7

From the Table 4, it is observed that in case of Model 1.2 for India and Model 2.5 for Assam i.e ARIMA (0,1,1), ARIMA (0,1,1) the value of AIC is lesser and Log-likelihood is greater as compared to the others. So here Model 1.2 and Model 2.5 are used to forecast the death rates of India and Assam. To know about the significance of the AR and MA coefficients, Z test is employed here and the test results are shown in the Table 5.

Table 5: Z test of the AR and MA coefficients

For	Coefficients	Estimate	Std. Error	Z value	Pr (>|z|)
India	**MA1**	-0.32093	0.17465	-1.8375	0.006614
Assam	**MA1**	-1.627921	0.430589	-3.7807	0.0001564

Table 5 shows that, the p-values are less than 0.01. Thus both the MA coefficients are greatly significant. The Error measures Mean Error (ME), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Absolute Scaled Error (MASE) and Auto Correlation of Errors at Lag 1(ACF1) are presented in the Table 6.

Table 6: Error Measures for the Models

Model for	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
India	**-168.143**	**1221.599**	**510.9575**	**-0.0937734**	**0.2808962**	**1.016272**	**0.1825707**
Assam	**-0.00012**	**0.145943**	**0.094664**	**-16.31145**	**414.4458**	**0.778404**	**0.00853529**

Fig 13 and Fig 14 depicted the "Normal Q-Q plot" of the residuals. Since the points are obtained approximately on a straight line, indicates that the residuals follow Normal distribution.

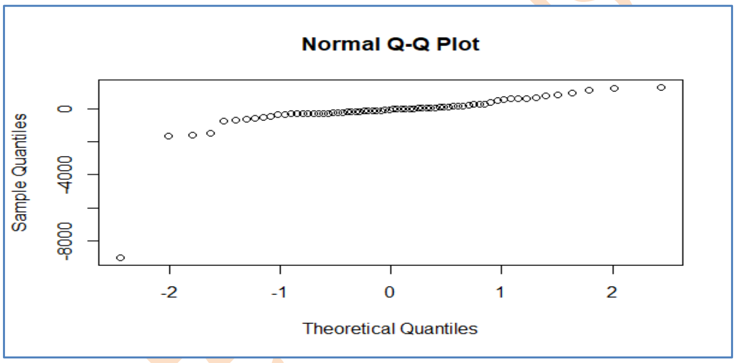

Fig. 13: Normal Q-Q plot of the residuals of ARIMA (0,1,1) Model for India.

.

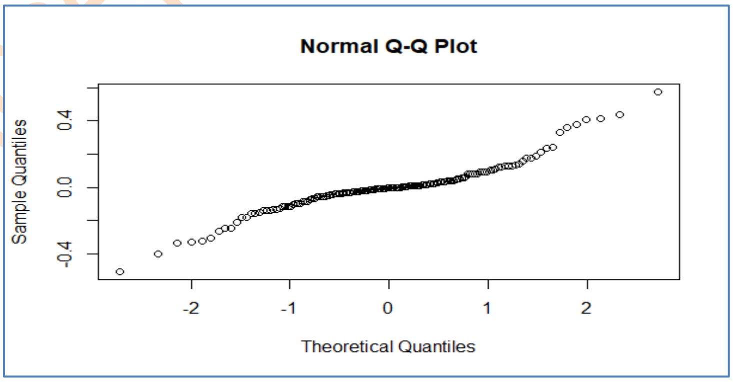

Fig. 14: Normal Q-Q plot of the residuals of ARIMA (0,1,1) Model for Assam.

For comparison purpose the forecasted death rates (Point Forecast) and the actual death rates for next 30 days depending on ARIMA Models are calculated here and are presented in Table 7 and the forecasted plots are shown in Fig 15 and Fig 16.

Table 7: Forecasted and actual death rates for next one month

Date	**India**		**Assam**	
	Original Value	**Forecasted Value**	**Original Value**	**Forecasted Value**
05-09-2021	1.335406853	1.335723	0.96220965	0.9619776
06-09-2021	1.334495149	1.335198	0.96270144	0.9622987
07-09-2021	1.334112026	1.334688	0.96352444	0.9626199
08-09-2021	1.333790861	1.334194	0.96427674	0.9629411
09-09-2021	1.332979038	1.333716	0.96425695	0.9632622
10-09-2021	1.332357537	1.333252	0.96422095	0.9635834
11-09-2021	1.33194593	1.332802	0.96441977	0.9639045
12-09-2021	1.331817108	1.332366	0.96495564	0.9642257
13-09-2021	1.33138429	1.331944	0.96638408	0.9645469
14-09-2021	1.331386618	1.331534	0.96664423	0.964868
15-09-2021	1.331153049	1.331137	0.96735451	0.9651892
16-09-2021	1.331225218	1.330753	0.96797551	0.9655103
17-09-2021	1.330726887	1.33038	0.96838226	0.9658315
18-09-2021	1.330232553	1.330018	0.96895591	0.9661527
19-09-2021	1.329932529	1.329668	0.96869881	0.9664738

20-09-2021	1.329611772	1.329329	0.96944987	0.966795
21-09-2021	1.329327547	1.328999	0.97038383	0.9671161
22-09-2021	1.329400792	1.32868	0.97067114	0.9674373
23-09-2021	1.328976566	1.328371	0.97067938	0.9677585
24-09-2021	1.328681701	1.328072	0.97234807	0.9680796
25-09-2021	1.328373882	1.327781	0.97225555	0.9684008
26-09-2021	1.328028367	1.3275	0.97231452	0.9687219
27-09-2021	1.327821021	1.327227	0.9728297	0.9690431
28-09-2021	1.327611617	1.326962	0.97349388	0.9693643
29-09-2021	1.32798971	1.326706	0.97452533	0.9696854
30-09-2021	1.327985375	1.326457	0.97509584	0.9700066
01-10-2021	1.327754584	1.326216	0.97531731	0.9703278
02-10-2021	1.327490131	1.325983	0.97532475	0.9706489
03-10-2021	1.32731498	1.325757	0.97559033	0.9709701
04-10-2021	1.327031046	1.325537	0.97544389	0.9712912
05-10-2021	1.327909971	1.325325	0.97639738	0.9716124

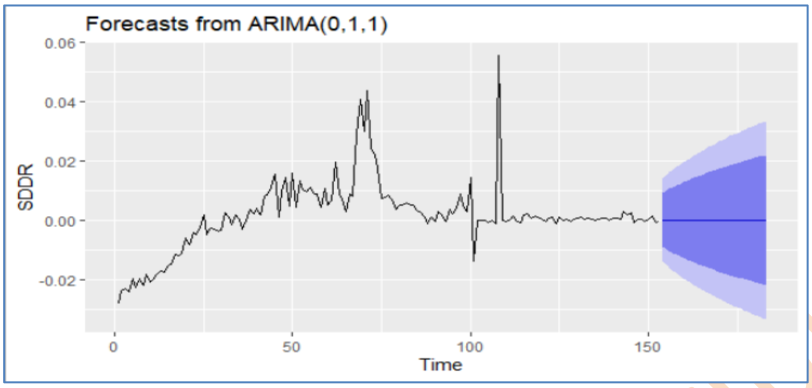

Fig. 15: Plot of forecasted death rate for India using ARIMA (0,1,1) Model.

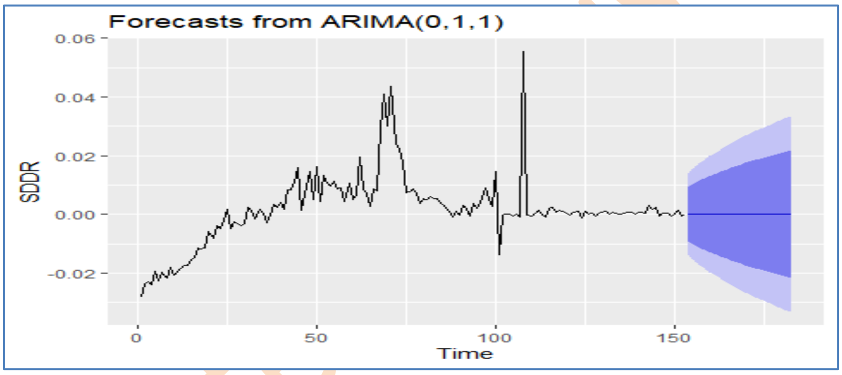

Fig. 16: Plot of forecasted death rate for Assam using ARIMA (0,1,1) Model.

4. Summary and Conclusion:

From the study, it is observed that the 2nd wave of COVID-19 has turned out to be less severe in the state of Assam as compared to India. The Maximum death rate for Assam is lower as compared to that of India. Again, the average death rate for the state Assam is slightly lower as compared to the country's average death rate which indicates that health workers as well as Assam state government were successful up to some extent to fight with COVID-19. The study also reveals the presence of high variability in the death rates for India as compared to Assam.

The death rate data are normally distributed and it exhibits stochastic trends. Originally the death rate data are not stationary but after the first difference it becomes stationary. For the death rates of India and Assam ARIMA (0,1,1) model is obtained to be the best fitted model according to some model choice criteria. According to the diagnostic analysis results, the models are adequately fitted to the death rate data. Our selected model gives us forecasted death rates for next one month and up to a large extent they match up with the original death rates. So in future, the government and health workers can take the help of these models to predict covid-19 death rates and it might prove substantially beneficial for them.

References:

[1]. Box, G. E. & Jenkins. G.M. (1970), Time Series Analysis: Forecasting and Control.

[2]. Dickey, D. A., & Fuller, W. A. (1979), Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association; 74(366a) 427-431.

[3]. Jarque, Carlos M.; Bera, Anil K. (1980), Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics Letters; 6 (3)255–259.

[4]. Khan, F. M., & Gupta, R. ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India (2020), Journal of Safety Science and Resilience;1(1)12-18.

[5]. Kumar, N., & Susan, S. (2020), Covid-19 pandemic prediction using time series forecasting models. 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT); 1-7.

[6]. Maleki, M., Mahmoudi, M. R., Wraith, D., & Pho, K. H. (2020) Time series modelling to forecast the confirmed and recovered cases of COVID-19. Travel medicine and infectious disease; 37, 101742.

[7]. Salgotra, R., Gandomi, M., & Gandomi, A. H. (2020), Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming. Chaos, Solitons & Fractals; 138, 109945.

[8]. Singh, V., Poonia, R. C., Kumar, S., Dass, P., Agarwal, P., Bhatnagar, V., et al. (2020), Prediction of COVID-19 corona virus pandemic based on time series data using Support Vector Machine. Journal of Discrete Mathematical Sciences and Cryptography; 23(8) 1583-1597.

[9]. Tandon, H., Ranjan, P., Chakraborty, T., & Suhag, V. (2020) Coronavirus (COVID-19): ARIMA based time-series analysis to forecast near future. arXiv preprint arXiv:2004.07859.