IOSR Journal of Humanities And Social Science (IOSR-JHSS)
Volume 25, Issue 7, Series 16 (July. 2020) 32-37
e-ISSN: 2279-0837, p-ISSN: 2279-0845.
www.iosrjournals.org

Projections for COVID 19 Pandemic in Assam using Autoregressive Integrated Moving Average (ARIMA) Models

Jyotsnali Chetia
Department of Economics,
Sibsagar Girls' College, Sivasagar, Assam India

ABSTRACT: The current outbreak of COVID 19 pandemic creates a serious threat to all over the world. Most of the nations including India have gone into a lockdown situation. The disease is spread by respiratory droplets and communication pathways. Fever, cough, shortness of breath to pneumonia, kidney failure and even death are some of the symptoms of this disease which can take 2- 14 days to appear in human body(WHO). Assam reported its first COVID 19 confirmed case on 31 of March and first death was reported on 10th April. The upsurge of COVID 19 pandemic in the state has been traced to persons who attended the conference of the Tablighi Jamaat religious organization at Nizamuddin Markaz, Delhi and also boost up due to some pilgrims of Ajmer Sharif Dargah, Rajasthan who arrived in the Silchar city of Cachar district on 6th May,2020. The state has also begun witnessing a spike in the number of COVID 19 cases due to returning of migrant workers from different parts of the country. The daily count of number of cases of the disease has been increasing in an alarming rate. As of 21st July 2020, the Government of Assam has confirmed 26772 confirmed cases of COVID 19 including 18033 recoveries, 3 migrants and 64 deaths in the state. This paper is an attempt to forecast the number of cases of COVID 19 in Assam for the days to come. Univariate time series modeling has been applied to forecast the number of COVID- 19 infected cases that can be expected in upcoming days in Assam. The study finds that there is an increasing trend in the actual and forecasted number of cases of COVID- 19 which is for serious concern for the people as well as the government. In order to prevent the spread of the virus an increase in the intensity of prevailing interventions is required. At the same time additional and strict intervention and active support of the Central Government, international organizations and the people of Assam is very much essential.

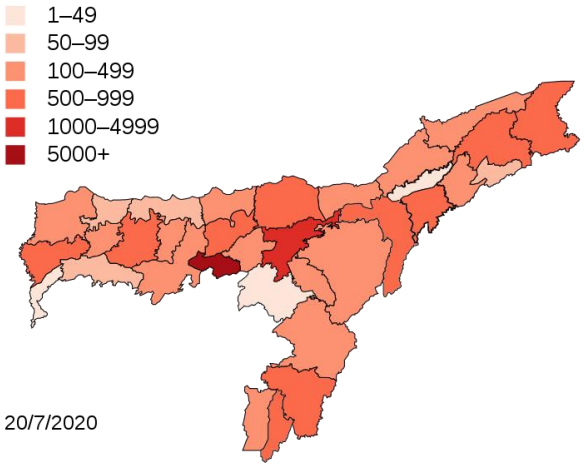
KEY WORD: Univariate Time Series, Novel coronavirous, ARIMA, BIC, Forecasting.

Date of Submission: 16-07-2020 Date of Acceptance: 31-07-2020

I. INTRODUCTION

Assam is situated in the North-East region of India- bordering seven States viz. Arunachal Pradesh, Manipur, Mizoram, Meghalaya, Nagaland, Tripura and West Bengal and two countries viz. Bhutan and Bangladesh.The geographical area of Assam is 78438 sq. kilometer of which 98.4 percent is rural area. The state shares about 2.4 percent of the country's total geographical area and provides shelter to 2.6 percent population of the country. Assam is the 15th most populous state in India as per 2011 census with 312.05 lakh population. As per National Family Health Survey Report (NFHS-IV) Assam shows a critical state of affairs in case of some vital statistics viz. birth rate (22.4), death rate(7.2), Infant mortality rate (49), under-5 mortality rate (2.3), maternal mortality rate(300) etc. The status of health infrastructure in the state is very indigent. According to Economic Survey,2016-17 there are 25 civil hospitals, 13 sub-divisional civil hospitals,1014 PHCs, 62 FRUs, 151 CHCs and 4621 Sub Centres with 18356 numbers of total beds in the state at the end of 2015. The number of available medical and paramedical staff in Assam is 5004 including Ayurvedic and Homeopathic doctors in 2015. As on 1.1.2013 average population served per Govt. hospitals was 30556 and average population per bed in hospital was 3062. Most surprisingly the number of government doctors was 12 per lakh population. In such a pathetic situation, if the number of cases of COVID 19 increases at existing rate government has to struggle to accommodate the critically ill patients. The present paper is a humble attempt to analyze the present situation and to give statistical evidence on potential evolution of COVID 19 pandemic in the state. The current situation of COVID 19 pandemic in Assam can be witnessed from Figure-1.

Figure-1: Map of districts with confirmed cases of COVID 19 in Assam as of 20th July,2020 [©wikipedia]



II. MATERIALS & METHODS

The ARIMA methodology is also called as Box- Jenkins methodology. This methodology is mainly concerned with identifying the stochastic process of the time-series and accurately predicting the future values for the series. Unlike regression models, the Box-Jenkins type models allow Y_t to be explained by past, or lagged, values of Y itself and stochastic error terms (Gujrati & Sangitha,2009). A stochastic process is either stationary or non-stationary. As the most time series are non-stationary in nature so the first step of B-J model is to reduce the non-stationary series to a stationary series by taking the first differences. This is so because the ARIMA model refers to only stationary time series. In this paper time series data regarding COVID 19 pandemic in Assam were collected. Data for the time period from 31st March to 15th July has been used for the model building, while data for 16th July,2020 to 21th July ,2020 are taken for model validation. On the basis of best fitted model data forecasting has been made up to 15thAugust, 2020.

Autoregressive (AR) model

ARIMA models, which stands for Autoregressive Integrated Moving Average models. The term integrated means the trends has been removed; if the series has no significant trend, the models are known as ARMA models .The AR (p) refers to the autoregressive model of order p. The AR (p) model can be expressed as:

$$\mathbf{Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ldots\ldots\ldots\ldots\ldots + \alpha_p Y_{t-p} + \mu_t}$$

Where, Y_t = response variable at time period t; $\mathbf{Y_{t-1}, Y_{t-2} \ldots\ldots\ldots Y_{t-p}}$ are the respective variables at different time with lags**;** $\boldsymbol{\alpha_0, \alpha_1, \ldots\ldots\ldots\alpha_p}$ are the parameters and μ_t is the error term

Moving Average (MA) model

The MA (q) refers to the moving average model of order q. The MA (q) model can be written as

$$\mathbf{Y_t = \mu_t + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \ldots\ldots\ldots\ldots\ldots\ldots + \beta_q \varepsilon_{t-q} + \nu_t}$$

Where μ_t is the constant mean of the series $\beta_1 \ldots\ldots\ldots \beta_p$ are the parameters of the estimated error term; ε_t is the error term.

ARIMA model can be formed by combining both the models

$$\mathbf{Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ldots\ldots\ldots\ldots\ldots + \alpha_p Y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \ldots\ldots\ldots\ldots + \beta_q \varepsilon_{t-q} + \nu_t}$$

If Yt is stationary at level or I(0) or at first difference I(1) determines the order of integration which is called ARIMA model.

ARMA (p,q) models can be used when the data are stationary. When a time series is studied based on the dependence relationship between the time-lagged values of the forecast variance and the past error terms, an ARIMA model is more appropriate and it can be used when the time series is non- stationary (Manikandan *et.al.*,2016)

Given a set of time series data, one can easily calculate the mean, variance, autocorrelation function (ACF) and partial autocorrelation function (PACF) of the series. Based on the nature of ACF and PACF appropriate ARIMA models are worked out, but the final decision is made once the model is estimated and diagnosed. In this step one can see that whether the chosen model fits the data reasonably well. Here one simple test of the chosen model is to see if the residuals estimated from this model white noise; if yes, one can accept

the particular fit and if they are not, one should start the process afresh, thus the Box-Jenkins methodology is an iterative process (Misra *et.al*.,2013)

ARIMA forecasting models involves four steps

Identification: The first step is to identify the appropriate order of ARIMA (p,d,q) model. Identification of ARIMA model implies selection of order of AR (p), MA (q) and I (d). The main tools used in model specification and selection of order p and q involve plotting of the autocorrelation function(ACF) and partial autocorrelation(PACF) against the lag length. The ACF specify the order of q and PACF specify the order of p. On the basis of the nature of ACF and PACF suitable ARIMA models are worked out. The ACF shows autocorrelation coefficients at different lag length with 95 percent confidence interval whether they are statistically significantly different from zero or not(Pradhan,2012).

Estimation: Next step is to estimate the parameters of the AR and MA terms included in the model. This generally involves the use of a simple ordinary least squares estimation process.

Diagnostic testing: Step three is to diagnosis whether the chosen model fits the data which is done by checking the residual term obtained from the ARIMA model by applying the ACF and PACF functions.

Forecasting: After completion of the above three steps of forecasting, the final step is to obtain the forecasted values by estimating the appropriate model devoid of all the problems. ARIMA models are useful for forecasting due to the use of lagged variables.

III. RESULTS AND DISCUSSION

Data regarding the number of COVID- 19 cases in Assam were collected from www.covid19assam.org.in.These data when plotted on a graph shows a linear trend of COVID-19 confirmed cases in Assam (Figure-2).

Figure 2: Trend of COVID 19 cases in Assam

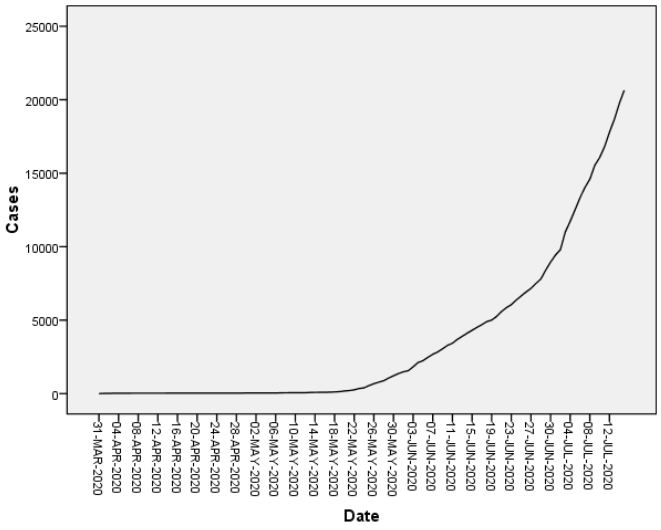


Figure 3 : Autocorrelation Function

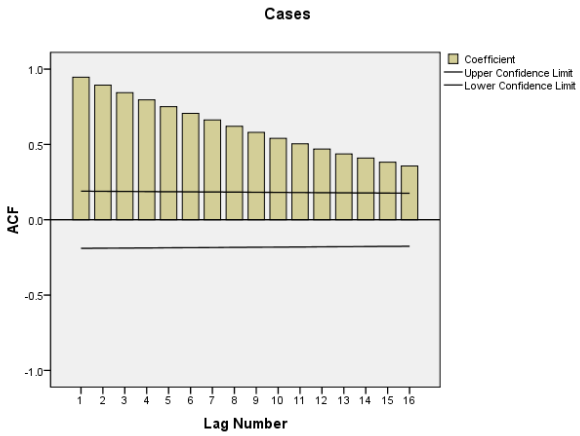
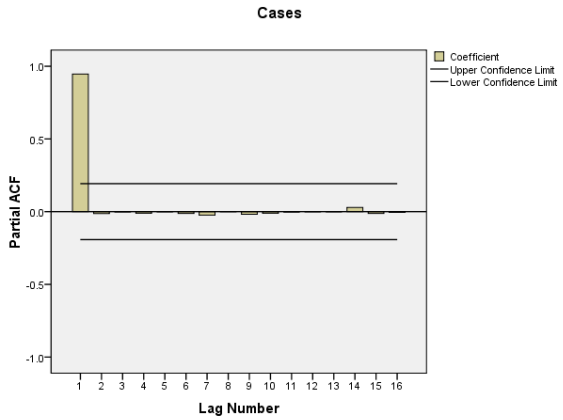


Figure 4: Partial autocorrelation function



Autocorrelation function (**Figure 3**) shows that the series has significant positive autocorrelation to a large number of lag, so a higher order differentiation is required. Partial autocorrelation function (**Figure 4**) plot has a significant spike only at lag- 1, meaning that all the higher order autocorrelations are effectively explained by the lag -1 autocorrelation**.** Thus a reasonable conclusion is that the first difference of COVID- 19 cases is best characterized as following a first-order autoregressive process.

Table 1: Selection of best fitted ARIMA model

ARIMA	R square	BIC Value
1,0,0	0.967	13.854
1,1,0	0.999	9.827
1,1,1	1.000	9.596
0,1,1	0.998	10.716
0,1,0	0.997	11.421
1,0,1	0.963	13.974

Henceforth ARIMA models (p,d,q) suitable for such scenario was adopted. The selection of the appropriate model was done on the basis of BIC value. R^2 represents how much variance of the data is explained by the model, the R^2= 1.000 means the model completely fit and explained all variance. BIC is criteria commonly used for model selection. A lower BIC value indicates a better fit as compared to others. In this paper the ARIMA (1,1,1) was found appropriate with goodness of fit (R^2= 1.000) and minimum BIC value (9.596).

The model verification is done by examining the autocorrelation and correlation of the residuals of various orders. Residuals are useful in checking whether a particular model has adequately captured the information in the data or not. If the residual are uncorrelated the model is a good forecasting model. It is evident that the ACF and PACF residuals are uncorrelated, implying that the ARIMA (1,1,1) model is a good fit(figure 5)

Figure 5: Residuals of ACF & PACF

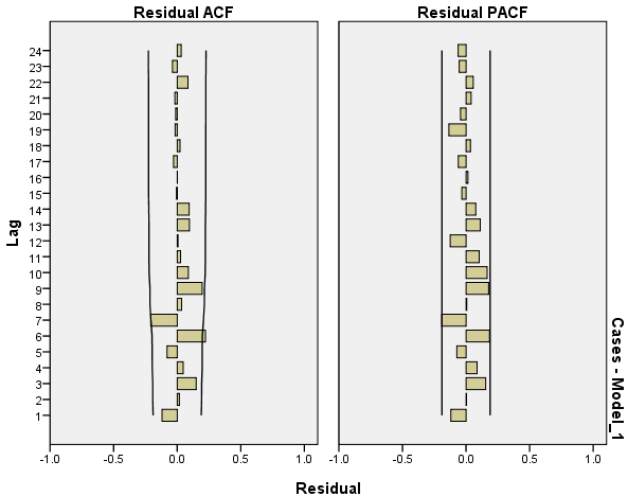


Figure 6: Actual and forecasted cases of COVID 19 in Assam

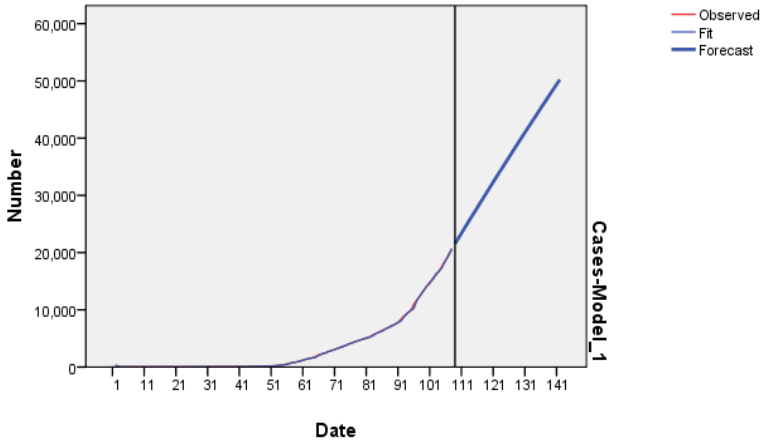


Table-2 Actual and forecasted cases ARIMA (1,1,1)

Date	Actual no of cases	Forecasted no of cases	Lower confidence limit	Upper confidence limit
15 July,2020	19754	20762	20507	21018
16 July,2020	20646	21477	21221	21732
17 July,2020	21864	22252	21702	22802
18 July,2020	22981	22976	22088	23864
19 July,2020	23999	23655	22400	24911
20 July,2020	25092	24293	22649	25936
21 July,2020	26772	24893	22847	26938
22 July,2020	-	25458	23002	27914
23 July,2020		25993	23121	28865
24 july,2020	-	26500	23210	29789
25 July,2020	-	26981	23272	30689
26 July,2020	-	27439	23313	31564
27 July,2020	-	27876	23336	32416
28 July,2020	-	28294	23344	33244
29 July,2020	-	28695	23339	34051
30 July,2020	-	29080	23323	34838

31 July,2020	-	29451	23299	35604
01Aug,2020	-	29810	23268	36352
02 Aug2020	-	30156	23231	37082
03 Aug2020	-	30492	23190	37795
04 Aug2020	-	30819	23146	38492
05 Aug,2020	-	31136	23099	39174
06 Aug,2020	-	31446	23051	39841
07 Aug,2020	-	31749	23002	40496
08 Aug,2020	-	32045	22953	41137
09 Aug,2020	-	32336	22904	41767
10 Aug,2020	-	32621	22856	42385
11 Aug,2020	-	32901	22809	42992
12 Aug,2020	-	33176	22763	43589
13 Aug,2020	-	33448	22719	44177
14 Aug2020	-	33716	22677	44755
15 Aug2020	-	33981	22637	45325

The Figure 6 shows the observed value, fit value and forecasted value of COVID 19 confirmed cases in Assam. It is clear from table-2 that there is an increasing trend in both actual and forecasted number of cases of COVID 19 with approximately 1000 per day. The proximity in the actual and forecasted values also ensured the validity of the ARIMA (1,1,1) model.

IV. CONCLUSION:

In this paper ARIMA model has been used to forecast the number of COVID-19 cases in India on data collected from Government of Assam official reports. Forecasting has done by applying ARIMA (1,1,1) model. The trend in forecasted values from 22nd July to 15th August, 2020 reveals that there is a sharp increase in the number of COVID-19 cases which is for serious concern for the people as well as the government. Government of Assam is taking all necessary steps to face the challenge and the threat posed by the rapidly growing pandemic of COVID-19 the Corona Virus. In order to prevent the spread of the virus an increase in the intensity of prevailing interventions is required. At the same time additional and strict intervention and active support of the Central Government, international organizations and the people of Assam is very much essential.

REFERENCES

[1]. World Health Organization. Coronavirus disease 2019 (COVID-19): situation report. Available at https://www.who.int/emergencies/diseases/novelcoronavirus-2019/situation-reports; 2020. 67.

[2]. Gujarati,D.N & Sangeetha.G.(2007) *Basic Econometrices*.4th Edition. New Delhi:Tata McGraw Hill Education Private Limited.

[3]. Manikandan,M. Velavan A,Singh Z,Purty AJ, Bazroy J, Kannan S.Forecasting(2016).Forecasting the trend in cases of Ebola virus disease in West African countries using autoregressive integrated moving average models. *International Journal of Community Medicine and Public Health*.Mar;3(3):615-618.Available from https://dx.doi.org/10.18203/2394-6040.ijcmpk20160620.

[4]. Mishra,P, Sarkar C, Vishwajith KP, Dhekale BS, Sahu PK.(2013).Instability and forecasting using ARIMA model in area, production and productivity of onion in India. *Journal of Crop and Weed*. 9(2):96-101.

[5]. Pradhan PC.(2012).Application of ARIMA Model for Forecasting Agricultural Productivity in India.*Journal of Agriculture and Social Sciences* 8, 50-56.

[6]. Ministry of Health & Family welfare, http://covidindia.org/portal,last access 21st July 2020.

[7]. Pamel, EC(2000).*Logistic Regression A Primer*. New Delhi: Sage Publications.

[8]. Goswami,K., Bharali, S.& Hazarika,J.(2020).Projections for COVID-19 Pandemic in India and effect of temperature and humidity. *Diabetes & Metabolic Syndrome: Clinical Research&Reviews*.14,801-805.Availablefrom https//:doi.org/10.1016/j.dsx.2020.05.045.

[9]. Government of Assam (2017),Economic Survey,2016-17, Assam, Directorate of Economics and Statistics, Assam Planning and Development Department,.

[10]. Government of India (2017), National Family Health Survey-4, Ministry of Health and Family Welfare,India.