STM JOURNALS

Research & Reviews: A Journal of
Immunology

ISSN: 2277-6206 (Online)
ISSN: 2349-1280 (Print)
Volume 13, Issue 1, 2023
DOI (Journal): 10.37591/RRJoI

http://stmjournals.com/index.php?journal=RRJoI&page=index

Research RRJoI

Second Wave COVID-19 Predictions and Forecasting of Confirmed Cases in West Bengal Using ARIMA Model

Sumanta Dey[1,*], Abhishek Saha[1], Pijush Dutta[2], Gour Gopal Jana[3]

Abstract

Infection and death rates surged drastically during the second wave of the COVID-19 (called delta variant) in India, owing to the destructive virus. As our country's economic load makes it more difficult to control the measures and it is critical for states such as West Bengal to forecast future cases. The present study introduced a time series forecasting model aimed at predicting and forecasting the number of confirmed and active COVID-19 cases up until July 2021. For the application of model dataset was taken from the Public Health Department in West Bengal, India. An Autoregressive integrated moving average-ARIMA (2,1,2) time series model was used to estimate the expected daily number of COVID-19 cases from April 18 to July 13, 2021. From the result analysis it was shown that confirmed cases in Bengal will be 827±200 by the end of July (July 14, 2021 to July 27, 2021) based on our forecasts which was quite satisfactory with real phenomenon. The findings of the present paper was used to forecast an increase in daily cases in West Bengal over the next month, which can assist the government in developing actions to stop the spread of virus.

Keywords: COVID-19, forecasting model, autocorrelation function (ACF), partial autocorrelation function (PACF), autoregressive integrated moving average (ARIMA)

INTRODUCTION

COVID-19 brought on by the SARS-COV-2 virus [1] became a cause of global disaster. In December 2019, the virus initially appeared in the province of Hubei, Wuhan, China. In less than a month, what started out as a few cases of pneumonia in Wuhan with no known cause, became a global disaster. More than 300,000 individuals have lost their lives and over 6 million individuals have been affected globally. In order to stop the sickness from spreading, nations have effectively been shut down, public spaces have been restricted, and several additional activity-restricting regulations have been put in place [2, 3]. The primary way of transmission of COVID-19 virus is by droplets that are expelled from a person's mouth or nose while coughing or sneezing [4]. If individuals are cautious and avoid coughing or sneezing carelessly, it might not seem lethal, but the fact that COVID-19 has spread around the world refutes the idea that it cannot be viewed as fatal [5–7].

***Author for Correspondence**
Sumanta Dey
E-mail: sumanta.dey2002@gmail.com

[1]Student, Department of Electronics and Communication Engineering, Greater Kolkata College of Engineering and Management, West Bengal, India
[2]Professor, Department of Electronics and Communication Engineering, Greater Kolkata College of Engineering and Management, West Bengal, India
[3]Assistant Professor, Department of Electronics and Communication Engineering, Greater Kolkata College of Engineering and Management, West Bengal, India

Received Date: June 28, 2023
Accepted Date: June 29, 2023
Published Date: July 15, 2023

Citation: Sumanta Dey, Abhishek Saha, Pijush Dutta, Gour Gopal Jana. Second Wave COVID-19 Predictions and Forecasting of Confirmed Cases in West Bengal Using ARIMA Model. Research & Reviews: A Journal of Immunology. 2023; 13(1): 1–8p.

The World Health Organization (WHO) has provided strategic guidance for countries planning to implement control measures, consisting of six criteria to be followed in a sequential manner. These criteria include controlling the transmission, ensuring preparedness of healthcare systems for effective contact tracing and optimal care provision, managing healthcare facilities carefully to prevent outbreaks, adhering to preventive measures as

essential services resume, addressing risks related to imported cases, and actively engaging communities to embrace the "new norm" [8, 9]. According to the World Population Prospects study (2019), it is worth noting that India has a population exceeding 1.36 billion individuals, and most of them live in metropolitan regions and cities, where they face a high risk of getting affected by COVID-19 [10, 11]. Predicting the number of confirmed and ongoing cases is of utmost importance. In this situation, every state must predict the number of confirmed and active cases so that the government may take proper steps to contain the illness. A time series model was put out in the present work to account for the likely evolution of COVID-19 in West Bengal in the near future [12, 13].

The present study was aimed to forecast the time series of COVID-19 in India based on existing data recorded from April 18, 2021 to July 17, 2021 and exhibit the patterns of the time series of the disease for 12 weeks ahead. This will help us determine roughly when the disease will be eradicated and when our life will return to normality. In later sections of the present paper, analysis dataset was used to apply on ARIMA model followed by implementation of ARIMA. The methodology was elaborated later followed by the description of model implementation.

DATA PREPARATION

A total of 87 dataset containing number of confirmed cases for each day was collected from the Health Department of West Bengal, India. The latest time stamp for the dataset was April 18, 2021 and the oldest being at July 13, 2021 (Table 1).

Table 1. Overview of COVID-19 dataset for confirmed cases.

Date	Confirmed case	Date	Confirmed case	Date	Confirmed case
18.04.21	8419	17.05.21	19003	15.06.21	3268
19.04.21	8426	18.05.21	19428	16.06.21	3187
20.04.21	9819	19.05.21	19006	17.06.21	3018
21.04.21	10784	20.05.21	19091	18.06.21	2788
22.04.21	11948	21.05.21	19847	19.06.21	2486
23.04.21	12876	22.05.21	18863	20.06.21	2184
24.04.21	14281	23.05.21	18422	21.06.21	1879
25.04.21	15884	24.05.21	17883	22.06.21	1852
26.04.21	15992	25.05.21	17005	23.06.21	1925
27.04.21	16403	26.05.21	16225	24.06.21	1923
28.04.21	17207	27.05.21	13046	25.06.21	1933
29.04.21	17403	28.05.21	12193	26.06.21	1894
30.04.21	17411	29.05.21	11514	27.06.21	1836
01.05.21	17512	30.05.21	11284	28.06.21	1761
02.05.21	17515	31.05.21	10137	29.06.21	1595
03.05.21	17501	01.06.21	9424	30.06.21	1478
04.05.21	17639	02.06.21	8927	01.07.21	1501
05.05.21	18102	03.06.21	8811	02.07.21	1422
06.05.21	184331	04.06.21	7913	03.03.21	1391
07.05.21	19216	05.06.21	7682	04.07.21	1297
08.05.21	19436	06.06.21	7002	05.07.21	885
09.05.21	19441	07.06.21	5887	06.07.21	962
10.05.21	19445	08.06.21	5427	07.07.21	982
11.05.21	20136	09.06.21	5384	08.07.21	995
12.05.21	20377	10.06.21	5274	09.07.21	990
13.05.21	208837	11.06.21	4883	10.07.21	997
14.05.21	20486	12.06.21	4286	11.07.21	924
15.05.21	19511	13.06.21	3984	12.07.21	885
16.05.21	19117	14.06.21	3519	13.07.21	863

ARIMA MODEL

George Box and Gwilym Jenkins created the ARIMA model in the year 1970 in an effort to use mathematics to characterize variations in time series [14, 15]. The objective of this model, which is based on adjusting observed values, is to minimize the discrepancy between the values produced by the model and the observed ones [16, 17]. This approach may very well be able to explain both stationary and nonstationary series' behaviors [18]. When a series' mean and variance remain constant throughout time, it is said to be stationary. Conversely, a series is said to be nonstationary when the value of the covariance relies exclusively on the interval between two time periods. In nonstationary series, random shocks increase the mean displacement and variation.

The common structure of the nonstationary model is known as ARIMA (p, d, q), where AR represents the autoregressive part with a degree of p, I represent the involvement of first difference with a degree of d, and MA represents the mean part with a degree of q [19]. The application of the ARIMA model involves three phases—identification, estimation and testing, and implementation. According to the present study, the anticipated cases are mostly utilized to help the government of West Bengal plan for the state's health infrastructure, including the number of isolation beds, intensive care unit (ICU) beds, ventilators, quarantine centers, etc. [20]. In a subsequent study, we anticipated the increase in COVID-19 patient cases during the next days based on forecasted active instances.

METHODOLOGY

The steps involved in the proposed model is discussed below (Figure 1).

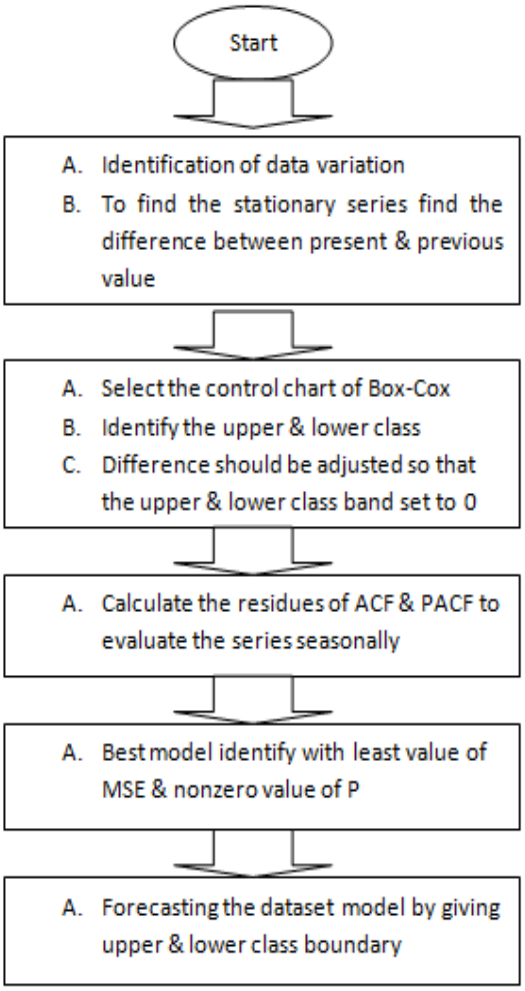

Figure 1. Flowchart of the proposed model.

IMPLEMENTATION

Checking for Stationary

The first step in applying ARIMA model is to check whether the time series is stationary or nonstationary [21, 22]. ARIMA works best when our data has stable or consistent pattern overtime, meaning that the variance and mean of the data have to remain constant overtime (Figures 2–5).

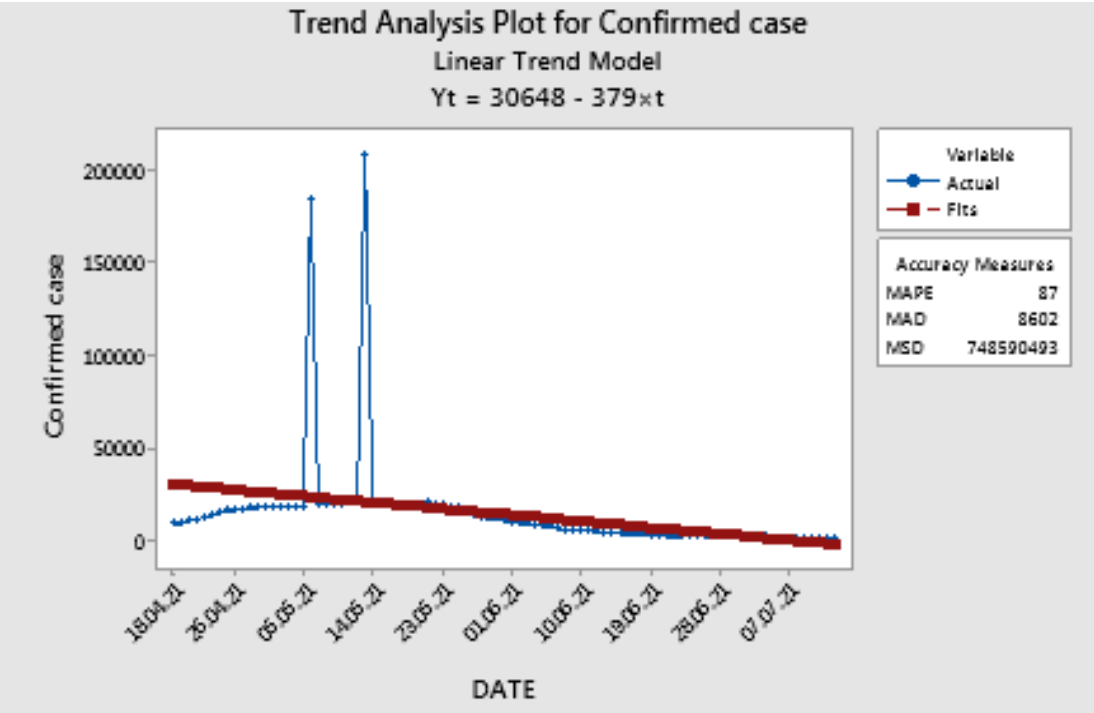

Figure 2. Trend analysis for COVID-19.

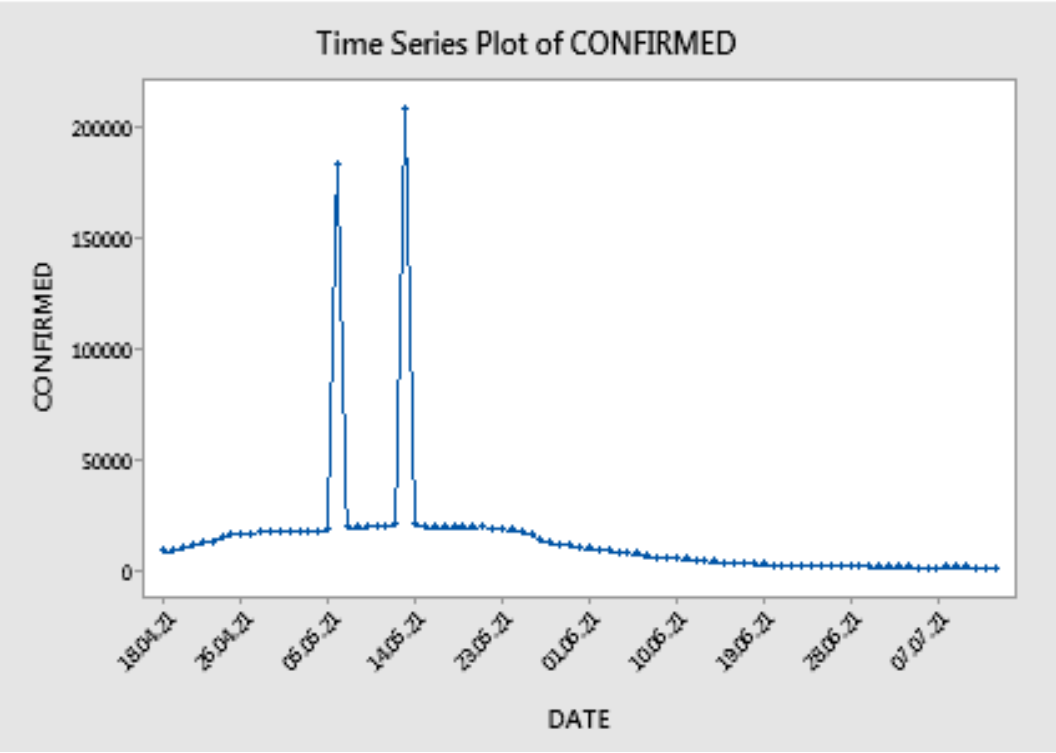

Figure 3. Plot for confirmed case and date.

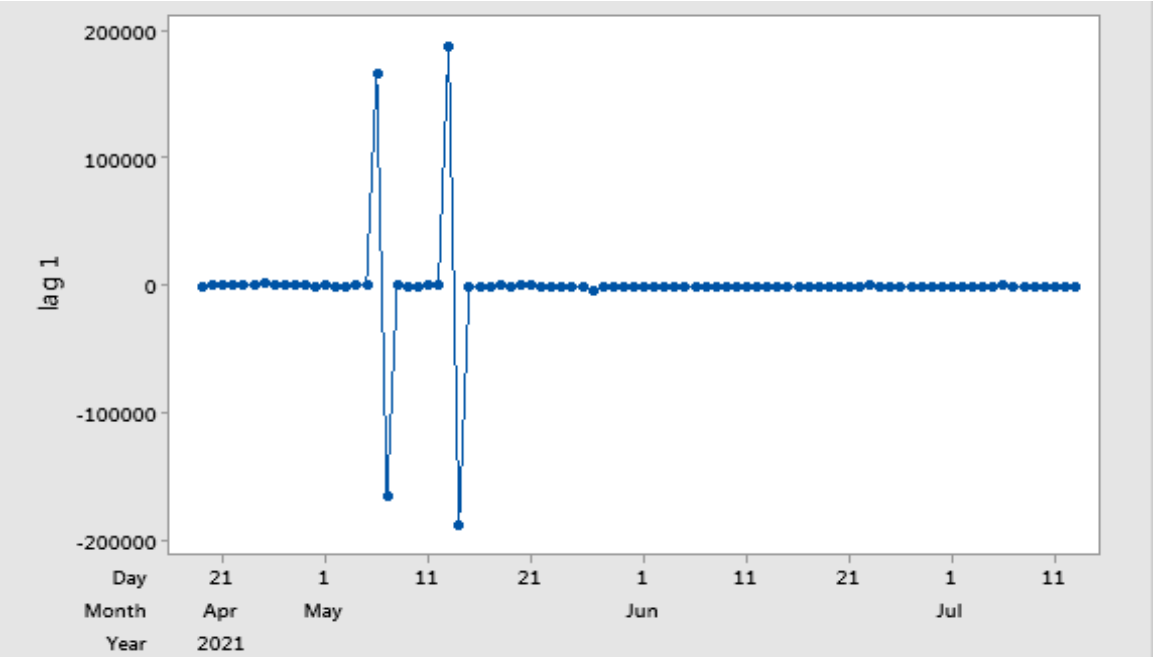

Figure 4. Plot between confirmed cases versus 1st difference with date.

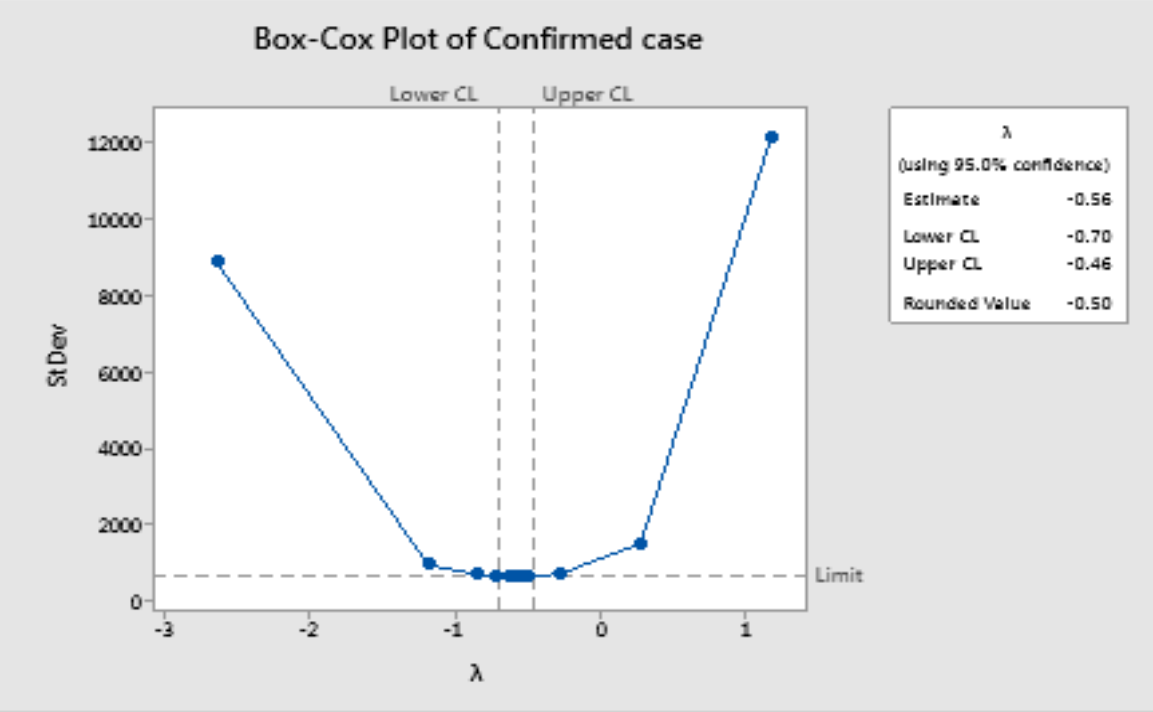

Figure 5. Box-Cox plot for confirmed case.

Forecasting Model

AR in ARIMA stands for auto regressive which is represented by p in the model. It refers to the number of lags of *Y* to be used as the predictor while the MA in ARIMA stands for moving average, represented by q in the model.

Identifying the AR term (*p*) was done by inspecting the Partial Autocorrelation (PACF) graph and MA term (*q*) was done by inspecting the Autocorrelation (ACF) graph. Both PACF and ACF graphs are shown in Figures 6–8.

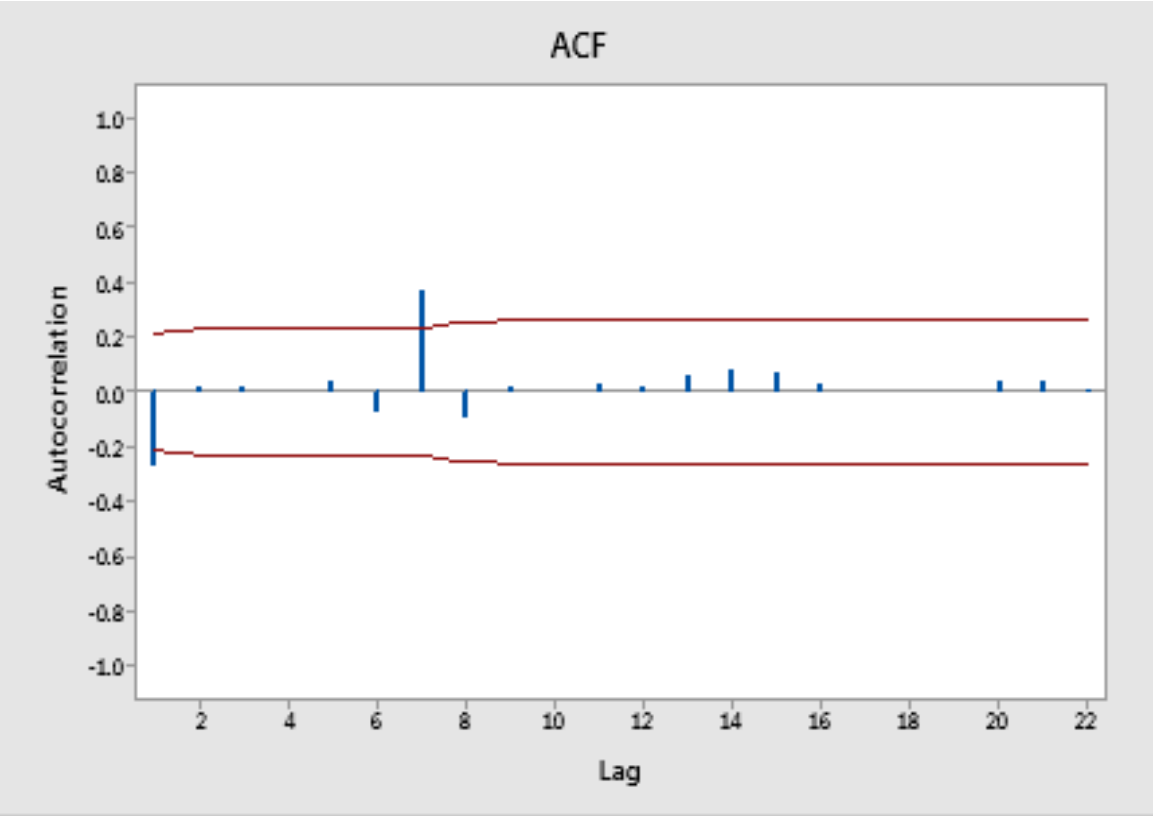

Figure 6. Autocorrelation function for confirmed case of COVID-19.

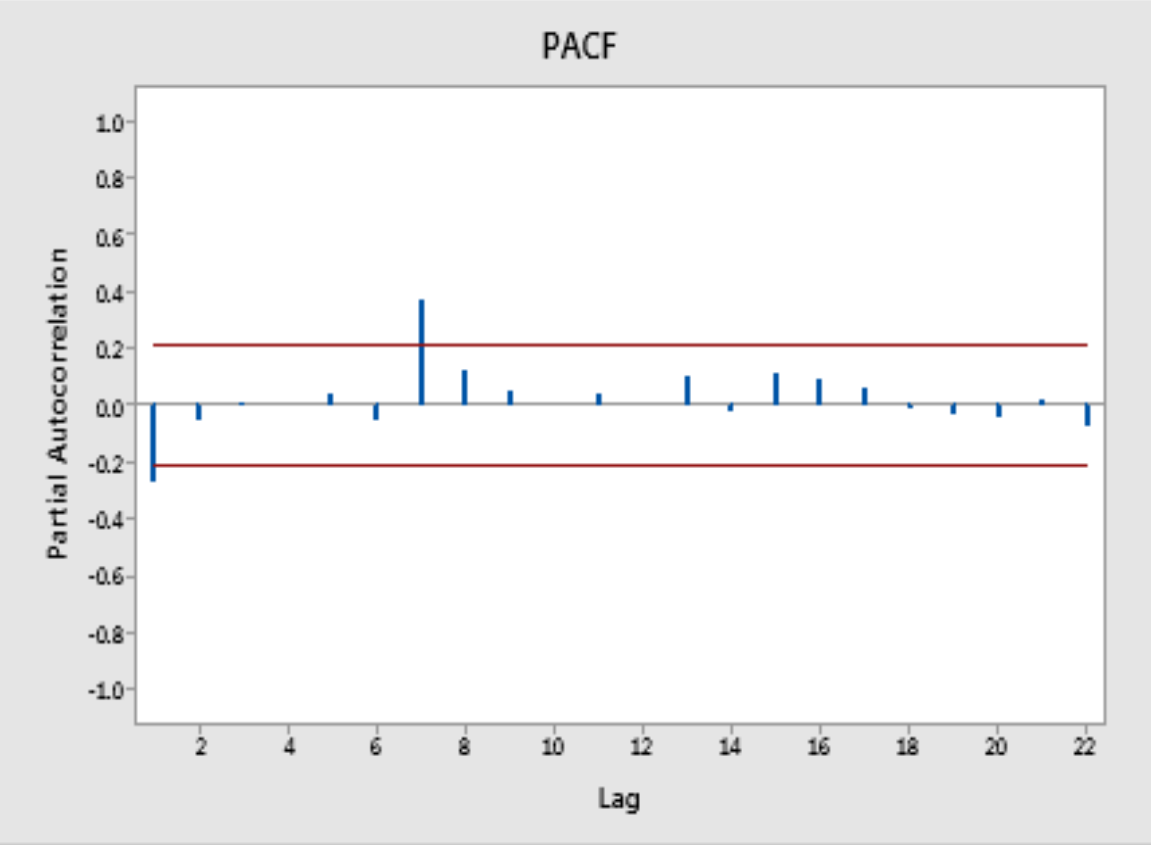

Figure 7. Partial Autocorrelation function for COVID-19.

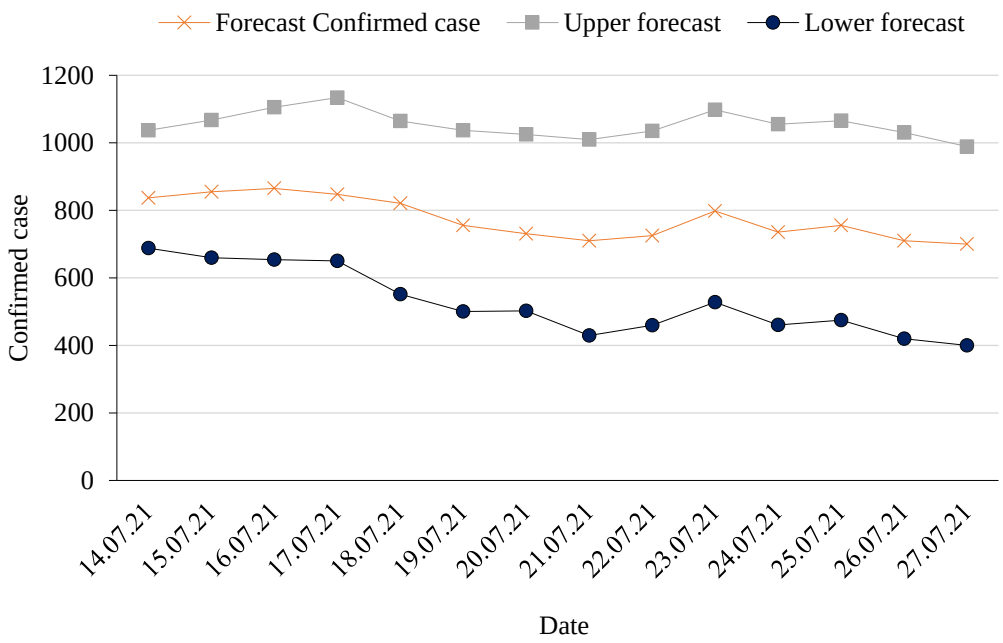

Figure 8. Plotting for forecasting of confirmed case.

CONCLUSION

Almost all nations are impacted by the COVID-19 epidemic. Country, as well as states, need to be aware of the future burden of confirmed and ongoing cases. It will assist the government in taking proactive steps to set up enough health infrastructure based on future requirements. In the present study, ARIMA and Auto ARIMA models were used to anticipate the total number of confirmed and active cases until the end of August based on time series data of COVID-19 cases in West Bengal, India. Forecasts indicated that from July 14 to July 27, 2021, there were about 827 confirmed cases in West Bengal, which was close to the actual number of verified cases.

REFERENCES

1. Ortiz-Prado E, Simbaña-Rivera K, Gomez-Barreno L, Rubio-Neira M, Guaman LP, Kyriakidis NC, Muslin C, Jaramil AMG, Barba-Ostra C, Robalino DC, San-Miguel HS, Unigarro L, Zalakeviciute R, Gadian N, Lopez-Cortes, A. Clinical, molecular, and epidemiological characterization of the SARS-CoV-2 virus and the Coronavirus Disease 2019 (COVID-19), a comprehensive literature review. Diagnostic Microbiology and Infectious Disease. 2020; 98(1): 115094.
2. Dutta P, Paul S, Kumar A. Comparative analysis of various supervised machine learning techniques for diagnosis of COVID-19. Electronic Devices, Circuits, and Systems for Biomedical Applications. 2021; 2021: 521–540p.
3. Dutta P, Paul S, Obaid AJ, Pal S, Mukhopadhyay K. Feature Selection based Artificial Intelligence Techniques for the Prediction of COVID like Diseases. Journal of Physics: Conference Series. 2021; 1963(1): 012167.
4. Bakar NA, Rosbi S. Effect of Coronavirus disease (COVID-19) to tourism industry. International Journal of Advanced Engineering Research and Science. 2020; 7(4): 189–193p.
5. Dhand R, Li J. Coughs and sneezes: their role in transmission of respiratory viral infections, including SARS-CoV-2. American Journal of Respiratory and Critical Care Medicine. 2020; 202(5): 651–659p.
6. Hafeez A, Ahmad S, Siddqui SA, Ahmad M, Mishra S. A review of COVID-19 (Coronavirus Disease-2019) diagnosis, treatments and prevention. EJMO. 2020; 4(2): 116–125p.
7. Parvin F, Islam S, Urmy Z, Ahmed S. The symptoms, contagious process, prevention and post treatment of COVID-19. European Journal of Physiotherapy and Rehabilitation Studies. 2020; 1(1): 81–105p.

8. Sarkar D, Biswas M. (Jun 2020). COVID 19 Pandemic: A real-time forecasts & prediction of confirmed cases, active cases using the ARIMA model & Public Health in West Bengal, India. Available from: https://www.medrxiv.org/content/10.1101/2020.06.06.20124180v1.full.
9. Viswa Chaitanya Chandu. (May 2020). Timeseries forecasting of COVID-19 confirmed cases with ARIMA model in the South East Asian countries of India and Thailand: A comparative case study. Available from: https://www.medrxiv.org/content/10.1101/2020.05.16.20103895v1.
10. Johnson EB, Trautman LJ. The Demographics of Death: An Early Look at COVID-19, Cultural and Racial Bias in America. Hastings Const LQ. 2020; 48: 357p.
11. Alam MS, Sultana R, Haque MA. Vulnerabilities of older adults and mitigation measures to address COVID-19 outbreak in Bangladesh: A review. Social Sciences & Humanities Open. 2022; 6(1): 100336.
12. Gupta R, Pal SK. (Mar 2020). Trend analysis and forecasting of COVID-19 outbreak in India. Available from: https://www.medrxiv.org/content/10.1101/2020.03.26.20044511v1.
13. Biswas B, Subhra Debdas, Samanta S, Mohanty S. Covid-19 data analysis of India and future forecasting using FbProphet. 2021 International Conference in Advances in Power, Signal, and Information Technology (APSIT). Bhubaneswar, India. 2021, Oct 8–10. DOI: 10.1109/APSIT52773.2021.9641221.
14. Eren ME, Solovyev N, Raff E, Nicholas C, Johnson B. COVID-19 kaggle literature organization. In: Proceedings of the ACM Symposium on Document Engineering; 2020. 1–4p.
15. Glas T. (2022). Asset pricing and investment styles in digital assets. Springer Nature. [Online] Available from: https://www.wiley.com/en-in/Time+Series+Analysis:+Forecasting +and+Control,+5th+Edition-p-9781118675021.
16. Sato RC. Disease management with ARIMA model in time series. Sao Paulo: Einstein. 2013; 11: 128–131p.
17. Box GE, Hillmer SC, Tiao GC. Analysis and modeling of seasonal time series. In: Seasonal analysis of economic time series. Cambridge: National Bureau of Economic Research; 1978. 309–344p.
18. Box G. Box and Jenkins: time series analysis, forecasting and control. In: A Very British Affair. London: Palgrave Macmillan; 2013. 161–215p.
19. Wabomba MS, Mutwiri M, Fredrick M. Modeling and forecasting Kenyan GDP using autoregressive integrated moving average (ARIMA) models. Science Journal of Applied Mathematics and Statistics. 2016; 4(2): 64–73p.
20. Goel I, Sharma S, Kashiramka S. Effects of the COVID-19 pandemic in India: an analysis of policy and technological interventions. Health Policy and Technology. 2021; 10(1): 151–164p.
21. Kumar M, Anand M. An application of time series ARIMA forecasting model for predicting sugarcane production in India. Studies in Business and Economics. 2014; 9(1): 81–94p.
22. Khashei M, Bijari M. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. Applied Soft Computing. 2011; 11(2): 2664–2675p.

