

Real-time forecasting of COVID-19 prevalence in India using ARIMA model



Sitaram pandey, Amitava Samanta

Abstract: Corona virus disease (COVID -19) has changed the world completely due to unavailability of its exact treatment. It has affected 215 countries in the world in which India is no exception where COVID patients are increasing exponentially since 15th of Feb. The objective of paper is to develop a model which can predict daily new cases in India. The autoregressive integrated moving average (ARIMA) models have been used for time series prediction. The daily data of new COVID-19 cases act as an exogenous variable in this framework. The daily data cover the sample period of 15th February, 2020 to 24th May, 2020. The time variable under study is a non-stationary series as y_t is regressed with y_{t-1} and the coefficient is 1. The time series have clearly increasing trend. Results obtained revealed that the ARIMA model has a strong potential for short-term prediction. In PACF graph. Lag 1 and Lag 13 is significant. Regressed values implies Lag 1 and Lag 13 is significant in predicting the current values. The model predicted maximum COVID-19 cases in India at around 8000 during 5th June to 20th June period. As per the model, the number of new cases shall start decreasing after 20th June in India only. The results will help governments to make necessary arrangements as per the estimated cases. The limitation of this model is that it is unable to predict jerks on either lower or upper side of daily new cases. So, in case of jerks re-estimation will be required.

Keywords: COVID -19, ARIMA, Daily new cases, Short-term Prediction

I. INTRODUCTION

Corona virus disease (COVID -19) is an infectious disease caused by a newly discovered corona virus. Corona virus is a pandemic that has infected almost 56, 00,000 people and caused deaths of 3, 48,000 people across the world till 25th May, 2020. Around 24, 00,000 people recovered also without any special treatment. The first fatal case of COVID-19 pandemic was reported by Chinese authorities along with WHO on 11th January 2020 whose features are similar to pneumonia of unknown origin that caused maximum damage to developed countries like America, Spain, Russia, UK, Italy, France, Germany etc. affecting 215 countries and territories around the world and 02 international conveyances. The first case was reported in India on 15th February, 2020 and and territories around the world and 02 international conveyances

The first case was reported in India on 15th February, 2020 and thereafter the counting has reached to 1,47,144 cases with 94 percent recovery rate.. The cases in India are far below the estimated figures of some International bodies but the worst yet to come as per some renowned researchers

The lockdown measure has prevented corona virus to spread in India with a lower rates than many of the other countries. The doubling of cases is taking place in twelve days. The virus that triggered COVID-19 is transmitted mainly via droplets produced when a person who is infected from coughs, sneezes, or exhales. These droplets are too heavy to hang in the air, and quickly fall on floors or surfaces. The virus transmits from one individual to other through droplets, personal contact, and contaminated objects (fomites) similar to previous outbreaks of Severe Acute respiratory syndrome corona virus (SARS-CoV) and Middle East respiratory syndrome corona virus (MERS-CoV). On 30th January 2020, the Director-General of WHO declared the 2019-nCoV outbreak a public health emergency of International concern under the International Health Regulations (IHR) (2005). The WHO has also passed temporary recommendations in early phases to China and other countries to be prepared for containment, active surveillance, early detection, isolation and case management, contact tracing and prevention of onward spread of 2019-nCoV infection, and to share full data with WHO. So, it is a time testing situation for each individuals, healthcare workers, various departments, governments, businesses, communities & nation. It also evaluating the public health infrastructure, public distribution system, public administration, leader's will power, preparedness of education system, research & innovation of each state. It is an efficiency check of the central government strategic preparedness for crisis management, contingency and response plans related to supplies of food and essential commodities. It is also checking individual's preparedness for economic, social & hygienic factors as the suggested solution is only physical distancing. It has created turmoil in economy and stock market around the world. The measures to prevent COVID -19 like lockdown have disrupted businesses with a negative impact on economies. But before the situation reach to uncontrolled conditions, the states has to perform risk and cost-benefit analyses before continuing such restrictions. The current restrictions has aggravated the lives of lower and middle section of the society whose income was hand to mouth due to spiraling inflation and poor wage system of the country. One of the most renowned businessmen has also quoted that "The year 2020 is for survival and not to make profits". Though the awareness of risk communication and community engagement is commendable in this internet age.

Revised Manuscript Received on June 13, 2020.

* Correspondence Author

Sitaram Pandey*, Department of management, Vinoba Bhave University, Hazaribagh, India. Email: spandey1203@gmail.com

Amitava Samanta, Department of commerce and management, Vinoba Bhave University, Hazaribagh, India. Email: dramitava1@yahoo.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Though the paper is not about the discussion on the existing pandemic, still informal investigation was done to get ahead of the outbreak through short term predictions based on time series data. This paper is analyzed by using the daily data of new cases between 15th February, 2020 to 14th May, 2020.

The formal investigation is achieved by using an autoregressive (AR) model, where Number of daily new cases is used to predict its future. Since COVID-19 is an exogenous shock, so number of daily new cases in India are included as an exogenous variable in this framework. The rest of the paper is organized as follows. The next section introduces the review of literature, Section 3 depicts methodology and the data set used, Section 4 discussed empirical results, Section 5 described results and discussions while Section 6 ends with concluding remarks.

II. REVIEW OF LITERATURE

COVID, C., & Team, R. (2020) in their MMWR Morb Mortal Wkly Rep published in March reported severe outcomes among patients with corona virus disease 2019 (COVID-19). Early reports from China show that a number of deaths from corona virus disease in 2019 (COVID-19) occurred in adults aged more than 60 years and people with serious underlying conditions of health. The first detailed summary of mortality for COVID-19 patients in the United States indicates that fatalities were the largest for people aged 85 years, varying from 10 percent to 27 percent, accompanied by 3 percent to 11 percent among people aged 65 to 84 percent years, 1 percent to 3 percent among people aged 55 to 64 years and less than 1 percent among people aged 20 to 54 years and no fatalities among women. They concluded that COVID-19 resulted to severe illness, including hospitalization, admission to an intensive care unit and death, especially in older adults. Thus, all should take steps to help reduce the spread of COVID-19 and shield older people from serious infection such as social distancing [1]. Gao, Tian, & Yang (2020) discussed the breakthrough achieved through Chloroquine phosphate that has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. The 2019 corona virus disease (COVID-19) is rapidly spreading, and scientists are endeavoring to find drugs for its effective treatment in china. In multicentre clinical trials conducted in china, it is shown that chloroquine phosphate, an old drug for malaria treatment, has apparent efficacy and acceptable safety against COVID-19 associated pneumonia. It is proposed that this medication be used in the next edition of the guidelines on Pneumonia Control, Diagnosis, and care caused by COVID-19 published by the people's republic of china's national health commission for potential care of COVID-19 infection in broader communities [2]. Fanelli & Piazza (2020) analyzed and forecasted COVID-19 spreading in China, Italy and France in the time window of 22nd January to 15th March, 2020. The researcher has taken a first analysis of simple day-lag maps point and suggested that simple mean-field models can be used to collect a quantitative picture of the epidemic spreading, the height and time of the peak of confirmed infected individuals. The model placed the peak in Italy around March 21st 2020, with a peak number of infected individuals of about 26,000 and number of deaths at the end of the epidemics of about 18,000. The containment measures can only be reducing the

epidemic peak. A concerted and disciplined effort of the population as a whole would only allow decrease in epidemic peak and mortality rate [3]. Chakraborty & Ghosh (2020) in their process of short-term forecasting of the future COVID-19 cases for multiple countries and risk assessment in terms of fatality rate presented a hybrid model focused on autoregressive integrated moving average model and wavelet-based forecasting model of the number of daily confirmed cases for Canada, France, India, South Korea, and the UK. Potential disease forecasts for various countries would be valuable for efficient deployment of health care services, and can serve as an early-warning system for government policymakers and to address the second issue, they implemented an optimal regression tree algorithm to classify important causal variables that influence the case fatality rates for different countries significantly. The data-driven analysis will necessary provide in-depths insights into the study of early assessments for 50 countries which have been immensely affected [4]. Yang, Zeng, Wang, Wong, Liang, Zanin, ... & Liang (2020) sought to show how control measures like large-scale quarantine, strict controls on travel and extensive monitoring of suspected cases impacted the containment of the epidemic. The researchers used Susceptible-Exposed-Infectious-Removed (SEIR) model to derive the epidemic curve and an artificial intelligence (AI) approach, trained on the 2003 SARS data, to predict the epidemic. They claimed that the complex SEIR model has been useful in forecasting peaks and sizes of the COVID-19 epidemic and they supported the implementation of control measures in January was proved indispensable in reducing the eventual COVID-19 epidemic size [5]. Ceylan (2020) also tried to estimate the COVID-19 prevalence in Italy, Spain and France. The research indicates that the ARIMA models are appropriate for possible estimation of COVID-19 prevalence. The empirical findings will shed light on interpreting the diseases patterns and provide an indication of the epidemiological stage of certain areas [6].

III. DATA AND METHODOLOGY

The formal investigation is achieved by using autoregressive integrated moving average (ARIMA) models based on daily data of new cases in India. The daily data of new COVID-19 cases act as an exogenous variable in this framework. The daily data cover the sample period of 15th February, 2020 to 24th May, 2020. Daily data on COVID-19 are obtained from the web page of worldometer. In this research we used to model random errors that may be positively correlated with each other over time. Here autocorrelation were used to identify the independence conditions of data. An autoregressive model is when a value from a time series is regressed on previous values from that same time series. For example y_t on a y_{t-1} :

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t \quad (1)$$

In this regression model, the response variable in the previous time period has become the predictor and the errors have our usual assumptions about errors in a simple linear regression model.

The order of an auto regression is the number of immediately preceding values in the series that are used to predict the value at the present time. So, the preceding model is a first-order auto regression, written as AR(1).

A. Autocorrelation and Partial Autocorrelation

The coefficient of correlation between two values in a time series is called the **autocorrelation function (ACF)**. For example the ACF for a time series y_t is given by:

$$\text{Corr}(y_t, y_{t-1}), k= 1,2,\dots\dots\dots (2)$$

This value of k is the time gap being considered and is called the **lag**. A **lag 1** autocorrelation (i.e., $k = 1$ in the above) is the correlation between values that are one time period apart. More generally, a **lag k** autocorrelation is the correlation between values that are k time periods apart. The ACF is a way to measure the linear relationship between an observation at time t and the observations at previous times. If we assume an AR(k) model, then we may wish to only measure the association between y_t and y_{t-k} and filter out the linear influence of the random variables that lie in between (i.e., $y_{t-1}, y_{t-2}, \dots \dots \dots y_{t-(k-1)}$,) which requires a transformation on the time series. Then by calculating the correlation of the transformed time series we obtain the **partial autocorrelation function (PACF)**. The PACF is most useful for identifying the order of an autoregressive model. Specifically, sample partial autocorrelations that are significantly different from 0 indicate lagged terms of y that are useful predictors of y_t . One can model data using **partial autocorrelation function (PACF)**. The time variable under study is a non-stationary series as y_t is regressed with y_{t-1} and the coefficient is 1. The time series have clearly increasing trend.

B. ARIMA Models

ARIMA models are considered as robust and effective techniques in time series forecasting for short periods. These models are considered as the mixture of statistical properties and artificial intelligence techniques. The predictive model based on ARIMA techniques were used in this study to forecast COVID cases in India. Box-Jenkins in 1970 introduced ARIMA model. This methodology is used to identify, estimate and diagnose ARIMA models with time series data and it has outperformed many complex methods also. In ARIMA model, the future value of a variable is a linear combination of past values and past errors, expressed as follows:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \quad (3)$$

Where y_t is the actual value and ϵ_t is the random error at t , ϕ_1 and θ_q are the coefficients, p and q are integers that are often referred as autoregressive and moving average respectively. So, it is a combination of three terms auto regression, differencing & moving average [7].

IV. EMPIRICAL RESULTS

The COVID-19 dataset of daily new cases have been used to identify the order of an autoregressive model. A plot of the

daily new COVID-19 cases versus time is presented in the fig.1 below.

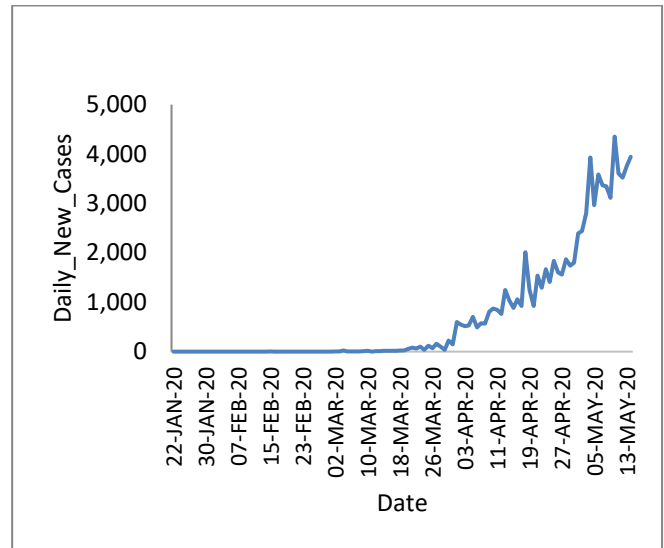


Fig. 1 : Time Series plot of daily new cases

Source : <https://www.worldometers.info/coronavirus/>

Consecutive values appear to follow one another fairly closely, suggesting an autoregressive integrated moving average (ARIMA) models model be appropriate. The time series data has increasing trend so it is a non-stationary data. To make it stationary first differencing was done. The result confirms that the series become stationary after the first-difference of the series.

We next look at a plot of autocorrelations function (ACF) & partial autocorrelations function (PACF) for the data (Fig.2).

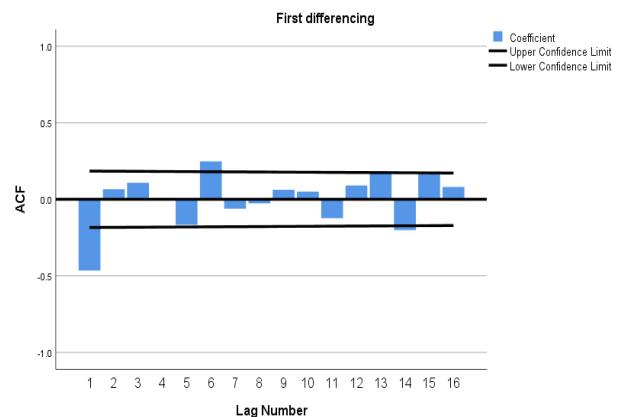


Fig. 2: Autocorrelation Function Graph of COVID-19 daily new cases

Fig. 2 is the plot of autocorrelation function Graph of COVID-19 Daily new cases at 1st Differencing. The Lag 1, Lag 5 & Lag 13 is significant in ACF function but lag 1 seems most significant.

$$y_t = \phi_1 u_{t-1} + \phi_5 u_{t-5} + \phi_{13} u_{t-13} + u_t \quad (4)$$



Real-time forecasting of COVID-19 prevalence in India using ARIMA model

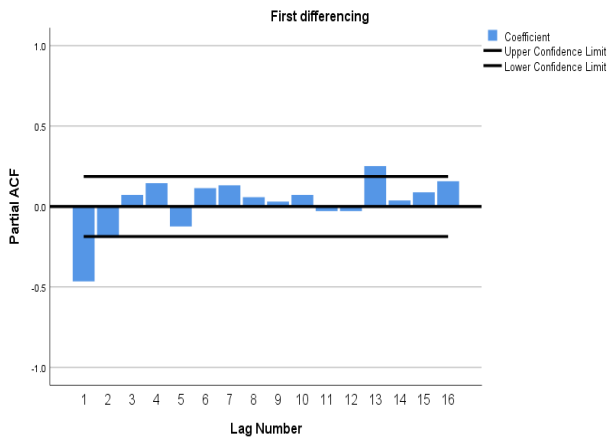


Fig. 3: Partial autocorrelation Function Graph of COVID-19 daily new cases

Fig. 3 is the plot of Partial autocorrelation function Graph of COVID-19 daily new cases at 1st Differencing In PACF graph. Lag 1 and Lag 13 is significant. Regressed values implies Lag 1 and Lag 13 is significant in predicting the current values.

$$y_t = 44.26 - 0.23y_{t-1} + 0.62y_{t-13} \quad (5)$$

Table I shows the different parameters of autoregressive (p) and moving average (q) among the several ARIMA model experimented upon. ARIMA (13,0,0) is considered best for COVID-19 prediction. The model returned the highest stationary R² and relatively lower BIC values.

Table I: Statistical Results of Different ARIMA parameters for COVID-19 daily new cases

ARIMA	BIC	Stationary R ²
(1,0,0)	11.95	0.922
(0,0,1)	13.55	0.613
(1,0,1)	11.95	0.926
(1,1,0)	11.39	0.207
(1,1,1)	11.42	0.230
(2,0,0)	11.95	0.926
(13,0,0)*	12.54	0.945

*The bold row represent the best ARIMA model among the several experiments.

Below Graph 4 is the forecasted graph of COVID-19 daily new cases in India. This analysis is based on time series data of daily new cases in India from 7th Feb 20 to 24th May 20. Maximum COVID-19 Cases around 8000 is expected around 5th June to 20th June as per the predictive model. As per the model, the number of new cases shall start decreasing after 20th June in India. The auto regression model is used to forecast the values of COVID-19 cases. Date is represented through numbers. Current position can be observed through red line in the graph (Fig.4).

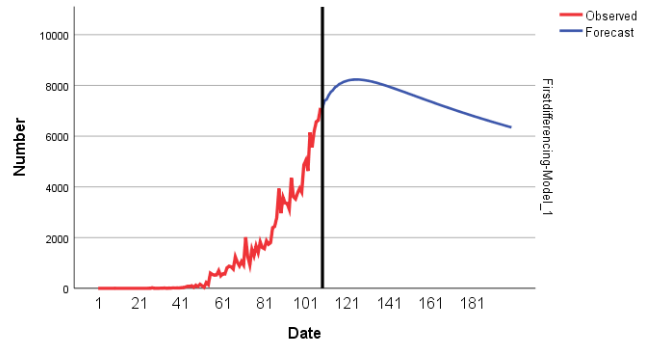


Fig.4: Forecasted Graph of COVID-19 daily new cases

V. RESULTS & DISCUSSIONS

The experimental results are discussed below. Table 2 is the result of predicted values of ARIMA (13, 0, 0) that has been found the best model for prediction of COVID-19 daily new cases in India. Figure 5 gives graphical illustration of the level of accuracy of the predicted daily cases against actual daily cases to see the performance of the ARIMA model selected.

TABLE II: Sample of empirical results of ARIMA (13, 0, 0) for Daily COVID-19 case.

Date	Actual Cases	Predicted Cases
21-Apr-20	1541	1142
22-Apr-20	1290	1343
23-Apr-20	1669	1463
24-Apr-20	1408	1566
25-Apr-20	1836	1504
26-Apr-20	1607	1795
27-Apr-20	1561	1678
28-Apr-20	1873	1695
29-Apr-20	1738	1721
30-Apr-20	1801	1867
1-May-20	2394	1819
2-May-20	2442	2298
3-May-20	2806	2455
4-May-20	3932	2822
5-May-20	2963	3724
6-May-20	3587	3368



7-May-20	3364	3662
8-May-20	3344	3500
9-May-20	3113	3615
10-May-20	4353	3176
11-May-20	3607	4231
12-May-20	3524	3726
13-May-20	3763	3794
14-May-20	3942	3576
15-May-20	3787	4031
16-May-20	4864	3763
17-May-20	5050	4720
18-May-20	4658	4958
19-May-20	6147	4977
20-May-20	5553	5836
21-May-20	6198	5841
22-May-20	6568	6291
23-May-20	6629	6587
24-May-20	7113	6940

Table II is represented through Fig. 5 to show the differences between actual and predicted cases clearly.

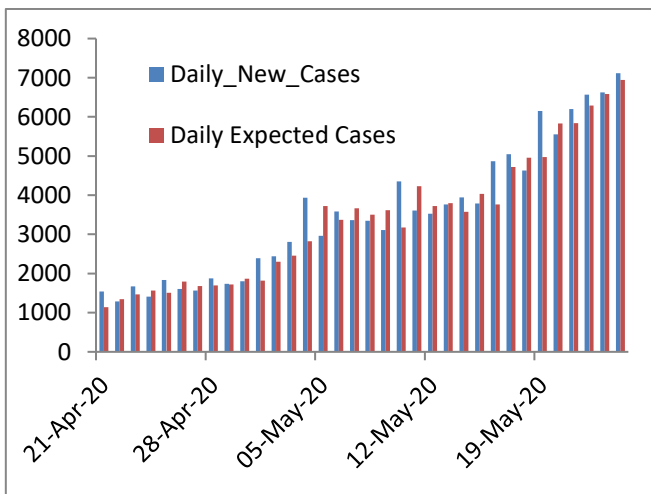


Fig.5: Graph of Actual cases Vs Predicted Cases of COVID-19

VI. CONCLUDING REMARKS

The paper presented with extensive process of building ARIMA model for daily COVID-19 cases prediction. The experimental results obtained with best ARIMA model that has been demonstrated with the potential of ARIMA models. The model used to predict COVID-19 cases on daily basis satisfactorily on short-term basis. The model

predicted maximum COVID-19 cases in India at around 8000 during 5th June to 20th June period. As per the model, the number of new cases shall start decreasing after 20th June in India. This research is done with an objective to generate optimism among government, corona warriors, society & individuals. The government can do necessary arrangements keeping in view that cases could be maximum in the month of June, so the result will guide governments in making their decisions in favor of society.

VII. LIMITATION OF THE MODEL

The model is unable to predict jerks on either lower or upper side of daily new cases. So, in case of jerks re-estimation will be required.

REFERENCES

1. C. COVID & R.Team., “Severe outcomes among patients with coronavirus disease 2019 (COVID-19)—United States”, February 12–March 16, 2020. MMWR Morb Mortal Wkly Rep, 69(12), pp. 343-346.
2. J. Gao, Z. Tian., & X. Yang , “Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies”, 2020, Bioscience trends.
3. D. Fanelli., & F. Piazza, “Analysis and forecast of COVID-19 spreading in China, Italy and France”. Chaos, Solitons & Fractals, 2020, 134, 109761.
4. T. Chakraborty & I. Ghosh, (2020). “Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis”. Chaos, Solitons & Fractals, 2020, 109850.
5. Z. Yang., Z. Zeng., K. Wang., S.S. Wong, W. Liang, M. Zanin, ... & J. Liang , “Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. Journal of Thoracic Disease”, 2020, 12(3), pp 165.
6. Z. Ceylan , “Estimation of COVID-19 prevalence in Italy, Spain, and France. Science of The Total Environment”, 2020, 138817.
7. A.A. Ariyo, A.O. Adewumi, & C.K. Ayo, “Stock price prediction using the ARIMA model”. In 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation (pp. 106-112). IEEE.
8. <https://www.worldometers.info/coronavirus/country>
9. WHO Report on 2019 Coronavirus as of 3rd Feb 20

AUTHORS PROFILE



Sitaram Pandey, is an Assistant professor & Research Scholar of University Department of Management, Vinoba Bhawe University, Hazaribag. He qualified UGC-NET with over 77% marks He is actively involved in any researches related to econometrics, time-series analysis, capital market and mathematical finances.

He has a long experiences of 12 years in teaching and consultancy of papers related to finance, commerce , statistics, operation research & quantitative finance. Won Best Business Academic Year Award, 2016 conferred by Indian Commerce Association. He also secured appreciation letter conferred by Cambridge Institute of Technology for Quality education . He is a life member of several associations. He fond of conducting workshops regularly on topics related to statistics. In his credit, there are edited chapters in Tata Mcgraw Hill, several journals of national and International repute, Confernces, workshops. He has also conducted two International confernces for Association of Indian Management Schools (AIMS) & Indian Commerce Association (ICA).

He started his career in finance field with the position of research analyst in stock market and took over fundamental and technical analysis. He has keen interest in capital & commodity markets researches and advances.



Real-time forecasting of COVID-19 prevalence in India using ARIMA model



Amitava Samanta, is an Assistant professor in University Department of Commerce and Management. He is a stalwart for the field of commerce and management. He is the real motivator and source of inspiration for various college graduates and researchers. He has more than 30 years of vast and enriched experience in the field of commerce and applied statistics. Govt. of Jharkhand has bestowed him with best teacher award in the year 2008. He is an active life member of several associations across India. He has so far presented more than 75 papers in national and international conferences in inter disciplinary subjects within and outside India. He also visited Unan University, China for paper presentation. Under his guidance, four research scholars has completed their Ph.D and seven are in the process of completion. He has also successfully completed two minor projects.. He also chaired and co-chaired several conferences & panel discussions across the country. He has a long experience of research and consultancy in the field of commerce. Currently, he is working as a Finance Officer in Vinoba Bhave University, Hazaribagh.