Journal of Pharmacognosy and Phytochemistry

Available online at www.phytojournal.com

J P P Journal of Pharmacognosy and Phytochemistry

E-ISSN: 2278-4136
P-ISSN: 2349-8234
www.phytojournal.com
JPP 2020; Sp9(2): 354-358
Received: 10-01-2020
Accepted: 12-02-2020

Megha Goyal
Department of Business Management, CCS Haryana Agricultural University, Hisar, Haryana, India

Suman Ghalawat
Department of Business Management, CCS Haryana Agricultural University, Hisar, Haryana, India

Nitin Agarwal
Research scholar, Guru Jambheshwar University, Hisar, Haryana, India

Neelam
Department of Business Administration, National Institute of technology, Kurukshetra, Haryana, India

Corresponding Author:
Nitin Agarwal
Research scholar, Guru Jambheshwar University, Hisar, Haryana, India

Trend and forecast of Covid-19 in India

Megha Goyal, Suman Ghalawat, Nitin Agarwal and Neelam

Abstract
COVID-19, an extremely transferrable virus, was reported firstly in Wuhan, China, in December-2019. Since then, there has been an aggressive growth in the figure of such cases around the globe. As of 29th May 2020, there have been 5,704,736 confirm cases, including 357,736 deaths, reported to WHO. The total figure of confirmed infected individuals in India is 165,799 and 4971 deaths to date. The figure shows the outbreak of the virus in various countries. In the current study, an effort has been made to analyze, and forecasting confirmed cases of corona-virus in India. Autoregressive integrated moving average (ARIMA) model by Box and Jenkins (1976) is used for prediction. This technique uses historical data of univariate time series to investigate its particular trend and estimate the forthcoming cycle. The daily data of pandemic confirm cases from 15thfeb 2020 to 25th may 2020 is used for model development, and five days ahead is projected. The explanations indicated increasing inclination in COVID patients.

Keywords: ARIMA, trend, forecast, Covid-19 and coronavirus

Introduction
Today, the planet faces an unparalleled catastrophe due to the pandemic, first found in December 2019 in Wuhan, China. The World Health Organization has demarcated the corona outbreak as an ancestor of viruses, ranging from high fever and the common cold to coronavirus (MERS) and serious acute respiratory syndrome coronavirus. In absolute wild beasts, the disease circulates and has the potential to spread from animals to humans. These diseases, besides other indicators of high fever and the common cold, may cause respiratory symptoms in humans. There are no precise therapies for coronaviruses.
Nonetheless, by sustaining simple individual hygiene and community distance from infected individuals, one can prevent infection. On 11 March 2020, the WHO announced Coronavirus disease as a global pandemic at the end of 2019. The virus has broadened transversely 213 nations and regions across the globe, with further added five million confirmed cases in total. In India, the epidemic was first spotted in a student who came back from Wuhan on 30 January 2020 in Kerala. The total (collective) number of confirmed infected individuals is more than 165,799 till date (29 May 2020) in all over India. Figure 1 displays the day-to-day development of coronavirus cases in India. Subsequently to the initial circumstances on 30thJanuary-3rd February 2020, there were no confirmed cases for around a month. Around 2 March 2020, the instances emerged again. Such cases are linked to persons who have been expatriate or arrived from the nations impacted by this disease. From 20 March 2020 onwards, the regular number of coronavirus cases is increasing exponentially at the India level.

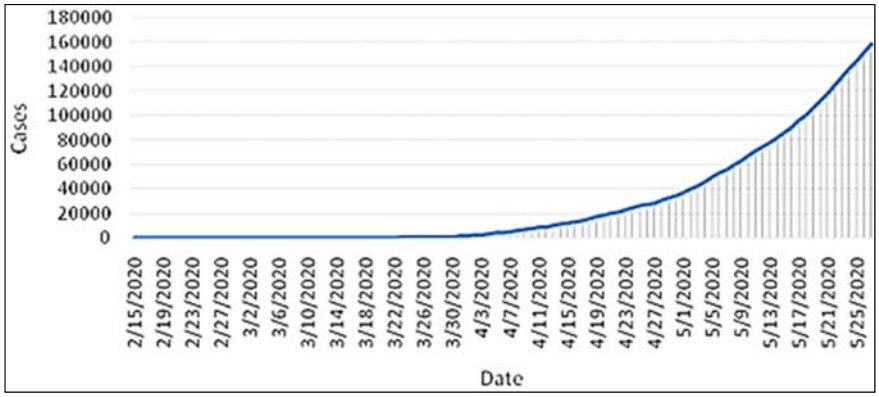

Fig 1: From 20 March 2020 onwards, the regular number of coronavirus cases is increasing exponentially at the India level.

As follows, there are 4 phases of epidemic, reliant on the virus spread forms. Stage 1: Introduced diseased cases with travel history from pandemic-hit nations are experienced by one country/region through this stage. Stage 2: A nation/area gets novel contagions from people who didn't have a history of travel but came into connection with people identified in phase 1. Stage 3: This is population spread; novel infectivity ensues in an individual who has not been in interaction through a diseased one or someone by way of the past of virus-hit nations traveling during this time. Step 4: The spread of the virus is virtually irrepressible at this point, and the country may have several large groups of infection. Many news outlets regularly say or challenge that India is actually in stage 3. Different Indian states at different points in time are at various stages of the disease. It is problematic to mark a Covid-19 phase at the India level. It will broaden distortion amongst ordinary individuals. Those, which are in stage 3, necessitate faster intervention than others. On the other side, states that are in phases 1 and 2, will concentrate on halting the spread of Covid-19.

Data Source

The study has used Indian COVID-19 data accessible widely. The primary data has been collected from the various websites of the Ministry of Health and Family Welfare, India, and form website www.worldometer.info>coronavirus.

The overall objectives of the studies are as follows:

1. To analyze whether the sequence of explanations under contemplation is stationary or not. If the observations are not stable, the data must be altered into fixed utilizing appropriate transformation.
2. To indicate the best ARIMA model subsequent the principle of parsimony and selection measures.
3. To access the trend of confirmed cases COVID-19 from 15th February 2020 to 25th May 2020.
4. To forecast the confirmed cases COVID-19 for the next five-point sufficiently.

Some of the studies linked to that are Palash *et al*. (May 2020) [6] developed three growth models to forecast infected populations for each of India's infected states over the next 30 days. The effect on the regular infected-rate of preventive measures is addressed for each state. Kumar *et al*. [2020] stated the approximate quantity of individuals in the Haryana, rural community, of the country with whom a person may come into exchange within 24 hours to be 17. They described interaction within 3 feet as having a direct discussion that may or may not has involved bodily interaction. The value of their paper's contact-rate parameter is 0.70. Only a handful of all individuals who come into connection with a person infected with this virus can be infected with the virus. Samuel P. C. Brand *et al*. [April 2020] [7] provide estimates of the possible occurrence rate and severity of pandemic in Kenya. The occurrence rate is based on the detected rate of development and age distribution of reported cases in China while taking into account the demographic and terrestrial variations between Kenya and China. Tomer and Gupta [April 2020] used data-driven approximation approaches such as long-term memory and curve suitable to predict the frequency of this outbreak cases in India 30 days ahead, and the consequence of preemptive procedures such as community isolation and lock-down on the epidemic spread. The estimation of the several parameters obtained by the proposed approach (figure of positive and recovered cases, etc.) is reliable within an impressive range. It would be a valuable resource for health and administrator's representatives.

Methodology

Autoregressive Integrated Moving Average Modeling

Box and Jenkins first promoted the univariate ARIMA approach for building the model. The method referred to as Univariate Box-Jenkins (UBJ) model. This model construction procedure performed in the resulting phases:

a. Understanding of the combined moving average configuration autoregressive where the order is (p, d, q);
b. Coefficients estimate;
c. A study carried out on the expected residuals, and the model exposed to a series of analytical tests and
d. Prediction for upcoming points from given data collection.

The overall functional forms of the model are:

$$\phi_p(B)\Delta^d y_t = c + \theta_q(B)a_t \quad \text{.....(1)}$$

where, y = Variable under forecasting

B = Lag operator

a = Error term ($Y - \hat{Y}$, where $\hat{Y}$ is the estimated value of Y)

t = time subscript

$\phi_p(B)$= non-seasonal ARi.e. the autoregressive operator, represented as a polynomial in the back-shift operator

Δ^d= non-seasonal difference

$\theta_q(B)$= non-seasonal MA i.e. the moving-average operator, represented as a polynomial in the back-shift operator

ϕ's and θ's are the parameters to be estimated

The order of the model AR and MA can be articulated here by p and q correspondingly. The amount of differences in time series stated over d. It should be distinguished that d, p, and q are all set of integers, showing the estimated residual for each time span. The model should be distributed independently as a random normal set of results, for ideal circumstances.

1. Identification of the Integrated Autoregressive Moving Average Model Where the Order is (p, d, q)

Initial stage of constructing an ARIMA model in equation (1) is to find the stationarity of observation data from time series. If the sequence of observations does not show stationarity, this must be transformed by correct differentiation into

stationary sequences. In order to obtain stationarity, the ACF or Auto Correlation Function and the PACF or Partial Auto Correlation Function of the given time series data sets must be drawn at the primary phase, if the series is not stationary then properly proper differenced order is needed to stationary the sequence being considered. The p and q of the model is to be equipped using the Correlogram analysis, this is based on iterative method. Here the BIC is tested alternatively speaking the Bayesian Knowledge Criteria to check for the goodness of fit. For this model, AR (p) the Auto Correlation function is tailing off at level p but the Partial Auto Correlation function cuts off for moving average model, MA (q). The Auto Correlation Function cuts off whereas the Partial Auto Correlation Function is tailing off in the order of q. Moreover ARMA (p, q) neither of the Auto Correlation Function or the Partial Auto Correlation Function is tailing off.

2. Coefficient Estimation

The standard model established in the equation (1) requires to be evaluated over iterative method till the sum of the squares of the residual in its least is attained.

3. Diagnostic Testing

The diagnostic testing will test the appropriateness of the built model. This involves the method of inspecting the residuals from the model thus suited for inspection, if the possibility of non-randomness exists. Here from the residuals correlogram is calculated, it is found out to what extent there is significant difference from zero among the coefficients. To test for the randomness of the model 's residuals the Ljung Box Statistic (Ljung and Box, 1978) is utilized. The formulation for the Ljung Box statistics is: $Q = N(N+2)\sum_{k=1}^{K}\frac{\hat{\rho_k}^2}{N-k}$(2)

Here N specifies the frequency of observations, the order of autocorrelation of the interval is represented by $\hat{\rho_k}$. Approximately the statistics Q follows a chi-square distribution with a df (K-m), where m is the frequency of parameters calculated in the ARIMA model. If Q is large it articulates that the residual autocorrelations as a set are different significantly from zero and random shocks of assessed model are possibly auto interrelated, then reformulate the model.

4. Future forecasting from given data collection.

Finally, the ARIMA model is used to predict performance; these limits give us the confidence interval with upper and lower limits. The root mean square estimation is used to define the dimension errors, which lets us determine the model's strength.

$$\text{RMSE}=\sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t-\hat{y_t})^2} \quad (3)$$

Equation (3) displays the appearance of the RMSE. Where $y_t$actual observation and is$\hat{y_t}$ is estimated observation.

Results and Discussion

Model Identification

The current model only anticipated after the set of remarks (sample data) taken for prediction is transformed into a stationary sequence. The stable series means the results vary only from a constant variance and a mean over time. First, it established whether or not the sample data sequence of explanations is stationary. By appropriate separation order of the data series under review, non-stationary in the mean is modified. Thus, first-order differencing introduced by analyzing the behavior of the time-series i.e., d=1. In the ARIMA model, the job now is to discover the valid values for p, q. The stationary series correlogram and partial correlogram is examined i.e. the first differential order of the sample results. Figure (2) displays the ACF and the PACF of the observation sample at first differencing.

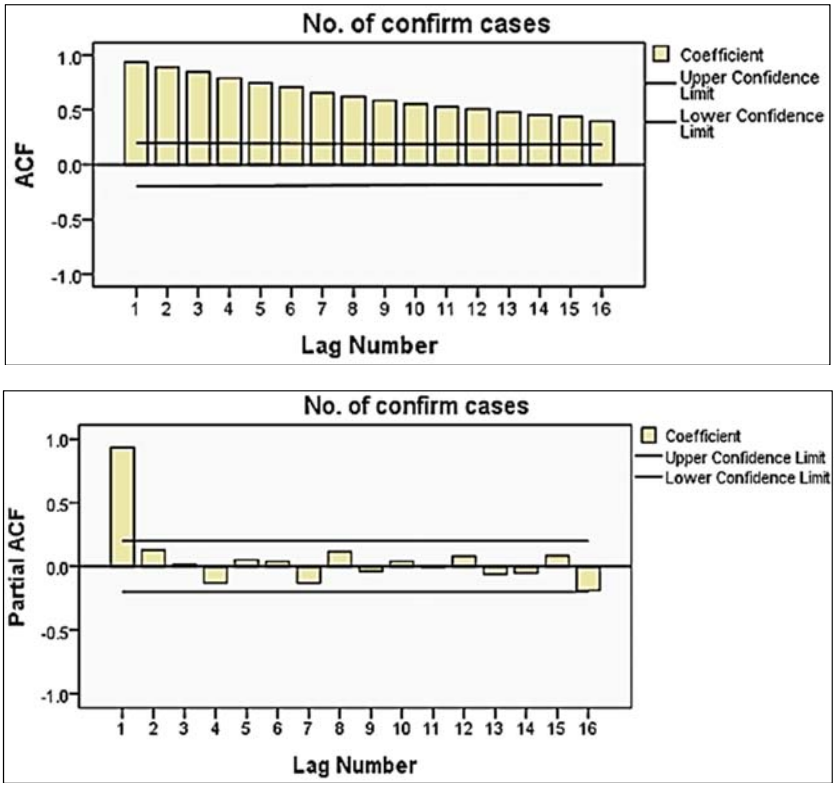

Fig 2: ACF and PACF of COVID-19 confirm cases

Figure (2) shows the Auto Correlation Function found to be tailing off, and the Partial Auto Correlation Function cutoff to order 1. So, conversing to the identification principle elucidated the model is AR (1) Different ARIMA models fitted by the iterative method, the model with the smallest standardized BIC values was chosen. The model (1, 1, 1) has the lowest BIC requirements; this model is selected, Table (III) illustrates all the structured BIC collection of remarks produced for the various this model.

Table 1: ARIMA (p, d, q) with Normalized BIC criteria

ARIMA (p, d, q)	Normalized BIC
ARIMA (1,1,0)	12.65
ARIMA (1,1,1)	12.64

Model Estimation

Autoregressive Integrated Moving Average model fitting is performed where specific coefficients that indicate significant coefficient are present. The test results described in table-2 produce all significant parameters.

Table 2: Parameter estimates of ARIMA model

Model		Estimates	Std Error	t	Approx Sig
ARIMA (1,1,0)	AR1	0.984	0.026	38.50	0.00
	Constant	2434.18	273430	0.88	0.37
ARIMA (1,1,1)	AR1	0.997	0.017	60A.17	0.00
	MA1	0.429	0.112	3.83	0.00
	Constant	3054.51	9374.91	0.32	0.74

Diagnostic Checking

The essence of the systematic form of residuals of both the fitted models is investigated in the process of diagnostic screening. The model fitted with ARIMA (1, 1, 1) indicates the lagged dependent variable values, i.e., autoregressive component. The residual ACF, along with the related't' tests and the Ljung and Box (1978), indicated a Chi-squared test were used to verify random shocks to be white noise (Table 3).

Table 3: Analytical examination of residuals autocorrelations: Covid-19 cases

Model	Number of Predictors	Model Fit statistics				
		R-squared	RMSE	MAPE	Ljung-Box Q Statistics	Sig.
ARIMA (1,1,0)	0	.925	533.12	1012.59	40344	.001
ARTMA(1,1,1)	0	.930	518.92	1107.73	19.259	.255

Forecasting

The observed and forecasted values of COVID-19 cases along with percent deviation are shown in Table 4.

Table 4: Percent deviation between observed & forecasted values from 26th May to 30th May.

Dates	Observed	Forecast	Percent Deviations
26-05-2020	150793	145359.8	-0.28
27-05-2020	158086	151530.2	-0.49
28-05-2020	165386	158100.3	-0.01
29-05-2020	173491	164660.3	-0.44
30-05-2020	181827	171210.4	-1.31

$$\text{Percentdeviation(RD\%)} = \frac{(\text{forecasted yield} - \text{actual yield})}{\text{actual yield}} \times 100$$

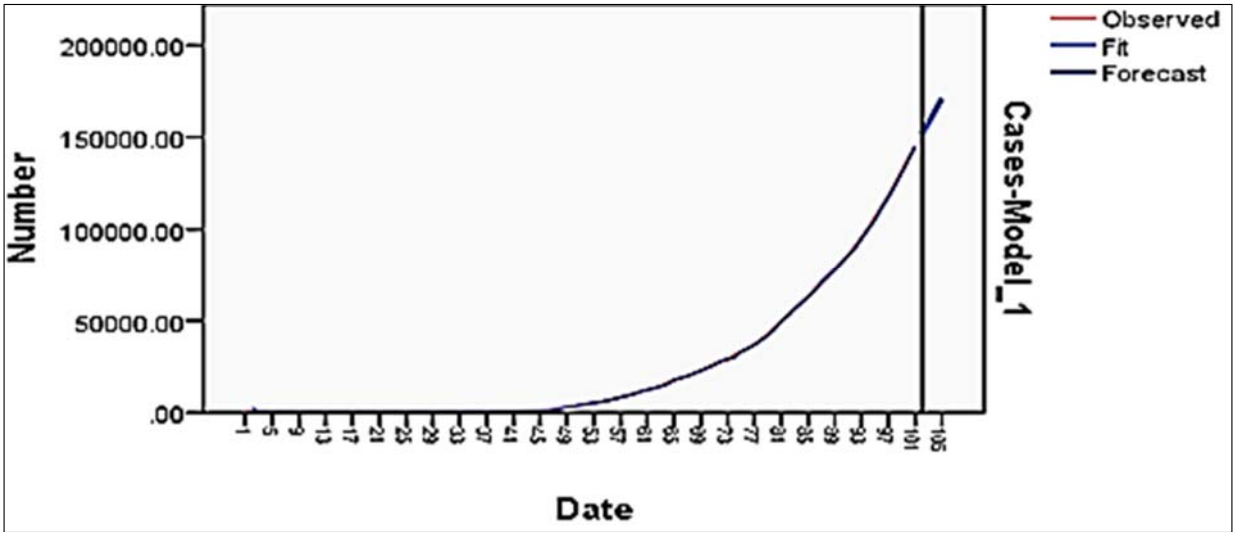

Fig 3: Forecasting Performance of the model represent in Figure-3

Conclusions

In this article, the ARIMA (1, 1, 1) model could be used successfully for modeling as well as forecasting of daily confirmed coronavirus cases in India. It has been found that there is a significant increasing trend in COVID cases in India. The forecast values from date 26th May to 30th May are close to the actual costs as percent deviation of the forecasted, and observed figures are in acceptable limits shown in table-4, and figure-3 shows that the proposed model perfectly fits the confirmed cases of the epidemic.

References

1. Akaike H. Fitting autoregressive models for prediction. Ann. Inst. Statist. Math. 1969; 21:243-47.
2. Anuradha Tomar, Neeraj Gupta. Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. Science of the Total Environment. 2020; 728:138-762.
3. Box GEP, Jenkins GM. Time series analysis: Forecasting and control. Holden Day, San, 1976.
4. Kumar S, Gosain M, Sharma H, *et al*. Who interacts with whom? Social mixing insights from a rural population in India. Lau EH, ed. PLoS One. 2018; 13(12):e0209039. Doi: 10.1371/journal.pone.0209039
5. Ljung GM, Box GEP. Ona measure of lack of fit in time series models. Biometrika. 1978; 65:297-303.
6. Palash Ghosh, Rik Ghosh, Bibhas Chakraborty. Covid-19 in India: State-wise Analysis and Prediction, 2020. https://doi.org/10.1101/2020.04.24.20077792.
7. Samuel PC, Brand Rabia Aziza *et al*. Forecasting the scale of the COVID-19 epidemic in Kenya, 2020. Doi: https://doi.org/10.1101/2020.04.09.20059865
8. MoHFW | Home. https://www.mohfw.gov.in/. Accessed May 30, 2020.
9. Covid-19 Tracker | India. https://www.covid19india.org/. Accessed, 2020. coronavirus pandemic in India - Wikipedia.
10. https://en.wikipedia.org/wiki/2020_coronavirus_pandemic_in_India#Statistics. Accessed May 30, 2020.
11. https://www.worldometer.info>coronavirus/. Accessed May 30, 2020.