

The Peak and Size of COVID-19 in India: SARIMA and Forecast

Suryakant Yadav¹, Neha Yadav² and Pawan Kumar Yadav*³

Abstract: Following the USA, India ranks the second position globally for COVID-19 cases in the pandemic year 2020. The study intends to explore the epidemiological stage of COVID-19 disease by estimating the most warranted features peak and size of COVID-19 cases in India. Data for this study was retrieved from a publicly available COVID19-India application programming interface (API). Exponential model was applied to estimate the growth rates of daily COVID-19 cases. Seasonal auto regressive integrated moving average (SARIMA) model was developed for the growth rates to predict daily COVID-19 cases. The exponential model unravels a shift and a modest decline in the growth of daily COVID-19 cases. The study shows that the SARIMA model is suitable for projecting daily COVID-19 cases. The forecasted peak value of daily COVID-19 cases was approximately 104,000 COVID-19 cases on 19 September 2020, whereas the real-time peak value was 97,861 COVID-19 cases conspicuous on 16 September 2020. The projected size of COVID-19 disease was 105 lakhs versus 103 lakhs at the end of December 2020. The forecasts and projections is adjoining to the real-time peak value of daily COVID-19 cases in India and successfully explores the epidemiological stage of COVID-19 disease in India.

Keywords: COVID-19, SARS-CoV-2, Exponential model, ARIMA, Growth and decay rates, Pandemic.

Introduction

Many top ranked countries in pandemic such as the United States of America (USA), Brazil, Russia, United Kingdom (UK) has experienced the second wave of Coronavirus disease 2019 (COVID-19) disease caused by the virus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Atangana, 2020; WHO, 2020). India is one of the countries not yet shown the second wave of COVID-19 pandemic. However, following the USA, India ranks the second position in the size of COVID-19 disease (Dong, Du, & Gardner, 2020; JHU CSSE, 2020). Given such trends, India shows a different predicament. Unlike other countries, India is the only country that showed an upsurge of COVID-19 with a delayed and high peak value, while other countries, including the USA, have shown the first peak much before India, and later on, ups and downs have been the course of the COVID-19 pandemic disease.

However, in August 2020, India breached the USA's peak value and reached the highest count of daily COVID-19 cases in the world. The exponential rise in COVID-19 cases was observed that delayed the peak value indicating a subtle impact of NPIs in India. In India, some conclusive factors such as income support, debt relief (Dev, 2020; Sengupta & Jha, 2020) and internal movement restrictions (Bhagat et al., 2020) might have led to the delay in attaining a peak

*Corresponding Author

¹ Assistant Professor, Department of Bio-Statistics and Epidemiology, International Institute for Population Sciences, Mumbai. Email: suryakant11@gmail.com

² Research Scholar, Centre of Social Medicine and Community Health, Jawaharlal Nehru University, New Delhi. Email: dr.nehayaadav@gmail.com

³ Research Scholar, Department of Bio-Statistics and Epidemiology, International Institute for Population Sciences, Mumbai. Email: pawankumaryadavbhu@gmail.com

value. After some conditional relaxations during April 2020 (MOHFW, 2020a), India implemented a cluster-containment strategy by red, orange, and green zones concerning movement restrictions during May 2020 (MOHFW, 2020b), where green coloured has the least restrictions and red zone had the strict restrictions.

Acknowledging data constraints in many countries, researchers have examined the turbulent rise in COVID-19 cases imperatively for analysing reproduction rate and case fatality rate on priority based on data-driven modelling (Atangana, 2020; Yadav et al., 2020) and for analysing epidemiological parameters such as growth rates, transmission rate, incubation period, recovery rate using mathematical modelling (Biggerstaff et al., 2020; Djaoue et al., 2020; Mohamadou et al., 2020). Alternatively, forecast and projection of the peak and size of the COVID-19 pandemic are warranted features (Bhardwaj, 2020; Ceylan, 2020; Jewell et al., 2020; Moreau, 2020; Singh et al., 2020). Numerous studies have investigated the time-series data preferably using statistical models such as vector autoregressive (Khan et al., 2020), autoregressive integrated moving average (ARIMA), exponential family (Ogundokun et al., 2020), compartmental (Krishnamurthy et al., 2020; Nabi, 2020) and machine learning (Chimmula & Zhang, 2020; Mehta et al., 2020) models.

This study aims to comprehend the epidemiological stage of COVID-19 disease in India. The rise of COVID-19 cases indicates an exponential rise in the last seven months from 30 January 2020 to August 2020. Therefore, it is important to analyse the daily confirmed cases of COVID-19 disease in India to estimate the peak value and the size of the COVID-19 cases noting the end of September and December 2020 as the critical periods. We explored the COVID-19 disease data by applying exponential models and developing the Seasonal Auto Regressive Integrated Moving Average (SARIMA) model and performed the model's assessment. Based on the application of these models, we analysed the growth rates of COVID-19 cases and provided an expected count of the peak and an expected size of COVID-19 cases in India.

Materials and methods

Data source

We retrieved data from COVID19-India application programming interface (API) an open access and publicly available database of COVID-19 disease cases (COVID19-India API, 2020). We retrieved data between the dates 30 January 2020 and 31 August 2020, and this period of five-and-a-half months is used for the analytical purpose of daily COVID-19 cases. Besides, the daily COVID-19 cases for the period between 01 September 2020 and 30 September 2020 was used for validating the outcomes for the next 30 days.

Exponential model

The exponential model is expressed as

$$y_t = C * e^{kt} \quad (1)$$

where t is time or date, y_t is the number of daily COVID-19 cases at time t , C is the initial value, and $B=e^k$ with k as the growth or decay rate.

Forecast and projection of COVID-19 cases

Auto regressive integrated moving average (ARIMA) was applied to the parameter k of exponential model (Ceylan, 2020; Singh et al., 2020) from mid-March 2020 to 31 August 2020 to predict daily cases from 01 September 2020 to until 28 February 2021. ARIMA (p, d, q), where p is the order of autoregression, d is the degree of difference, q is the order of moving average, is expressed as

$$y_t = \alpha + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-n} + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \dots \phi_n \varepsilon_{n-1} + \varepsilon_t \quad (2)$$

where y_t is the differenced series of COVID-19 cases, y_{t-n} are the lagged values of y_t of order n , ε_t is white noise, ε_{n-1} is the lagged errors of order n , α is an intercept term, β is an autoregressive parameter, and ϕ is a moving average parameter.

An ARIMA model is efficient in dealing with non-stationary and stationary time series data (Duan & Zhang, 2020; Lotto et al., 2017). The seasonal ARIMA or SARIMA for time series is referred to as an ARIMA (p, d, q) (P, D, Q)[m] where (P, D, Q) represents the (p, d, q) for the seasonal part of the time series, and m refers to the number of observations per cycle. The SARIMA model's accuracy is measured by root mean square error (RMSE) and mean absolute percentage error (MAPE).

The natural logarithm of the constant 'C' of the exponential model showed a linear increase during the studied period. A simple linear regression was applied to predict the values of 'C' overtime.

$$\ln(C) = \mu + \beta * time \quad (3)$$

where, μ is constant and β is the slope of the model.

The Pearson correlation test was applied to measure the linear association between projected values and real-time data between 01 September 2020 and until the peak timing. Statistical softwares used were STATA (StataCorp, 2019), R (R Core Team, 2017) and RStudio (RStudio Team, 2020).

Results

Exponential model

The study included 3,687,944 confirmed cases of COVID-19 disease for analyses during the period of seven months between 30 Jan 2020 and 31 August 2020. Besides, the study used COVID-19 cases in September 2020 for validating the outcomes. Figure 1 shows the non-linear fits based on the Equation (1) over the daily COVID-19 cases. It also shows the adjusted R-square of the exponential model's natural logarithm from 15 March 2020 until 31 August 2020 as the sample size was small until the mid of March 2020.

The adjusted R-square values (brown coloured connected circles) of the exponential model's natural logarithm was more than 85% from May to August 2020. This high adjusted R-square confirms the large proportion of explained variance by the exponential model's natural logarithm and ascertains the model's best fit over COVID-19 cases. The model parameters 'C' and

‘B’ ($=e^k$) of the exponential model were estimated by applying the nonlinear estimation method. These model parameters ‘C’ and ‘B’ are significant ($P < 0.001$) since mid-April 2020. The parameter ‘B’ or the slope of the exponential model explains the acceleration or deceleration of COVID-19 cases. The slope parameter ‘B’ was 1.09 ($P < .001$) by mid-April, which declined to 1.047 ($P < .001$) and 1.023 ($P < .001$), respectively, by the end of May 2020 and August 2020 (Table 1). Based on these two slope parameters, the plot of non-linear fits shown until May 2020 and August 2020 (deep-sky-blue (dark and light) coloured long-dash line) displays the exponential increase in daily COVID-19 cases and a deceleration in COVID-19 cases (Figure 1).

Figure 1: Exponential fits over daily COVID-19 cases, India, March to August 2020

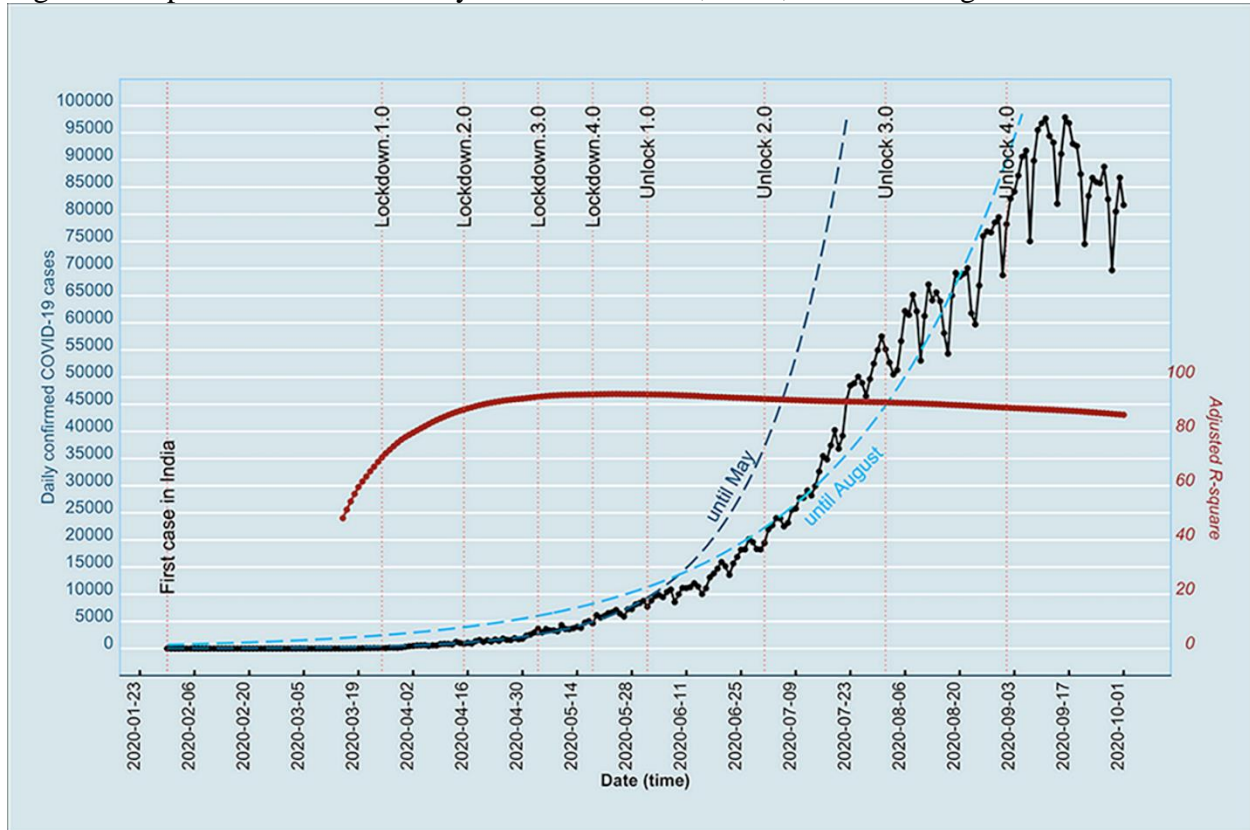


Table 1: Model fitting of an exponential model and its parameter estimates, India, March-August 2020

	Time period from 15 Mar 20	COVID-19 cases	C	B	SE (%)	Adj. R ²
Nonlinear	Until May 20	190,775	31.862	1.047	3.443	
	Until Aug 20	3,766,328	718.437	1.022	5.846	
Linear	Until May 20	190,775	0.326	0.092		93.686
	Until Aug 20	3,766,328	2.153	0.058		88.738

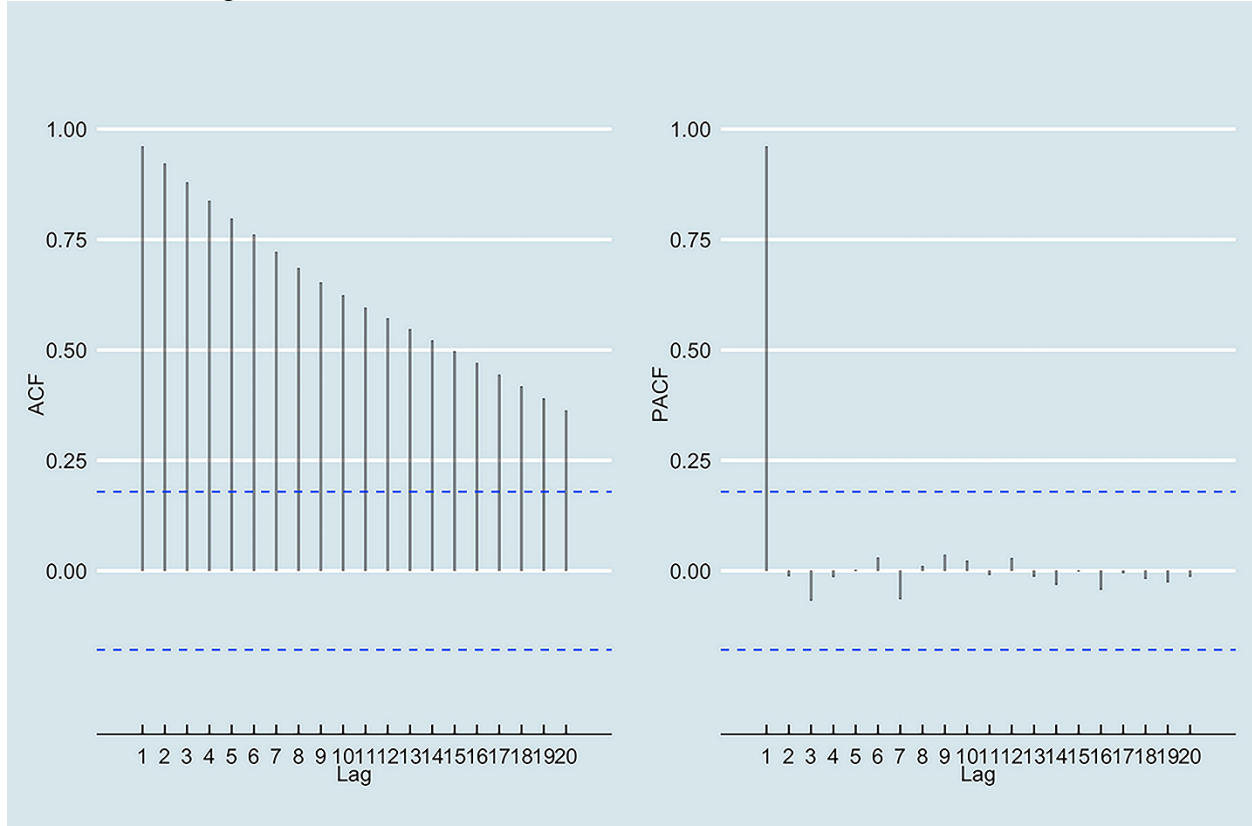
C and B ($=e^k$) are constant and acceleration or deceleration, respectively, of exponential model, SE: Standard error, Adj. R²: Adjusted R-square

Forecast, projection and the peak and size of COVID-19 cases

Model assessment: ARIMA/SARIMA

We performed ARIMA/SARIMA model using Equation (2) to obtain the forecasts and projections of daily COVID-19 cases. The time series plot of growth rates, autocorrelation function (ACF), and partial autocorrelation function (PACF) (Figure 2) reveal seasonality and non-stationary in growth rates of daily COVID-19 cases.

Figure 2: Autocorrelation and Partial autocorrelation for the daily variation in COVID-19 cases, India, March-August 2020



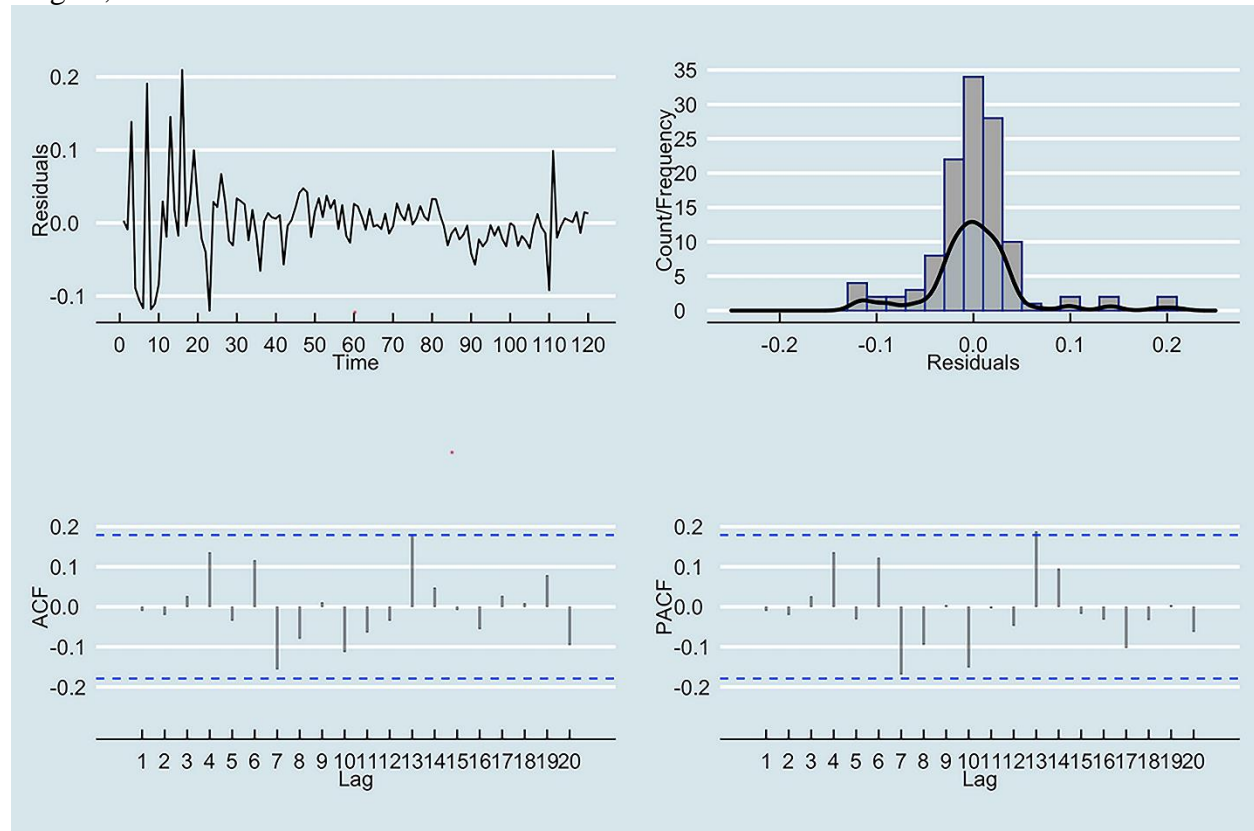
Based on the ACF and PACF of the ARIMA model, the second difference was applied for making the time series stationary. The corrected AIC (AICc) value of the model ARIMA (2, 2, 1) was -347.3 . Again, based on the seasonal part of ACF and PACF, the SARIMA (2, 2, 1) (1, 0, 1)[7] was applied that had the lowest AICc value of -357.3 (Table 2). A seven-day cycle was applied as the moving average of seven days shows a linear trend in COVID-19 cases. The developed SARIMA model’s diagnosis confirms no trends in the residuals, no outliers, and nearly constant variance (Figure 2). Specifically, the ACF plot of residuals shows no significant autocorrelations. The plot of the residuals shows a standard normal variate. The *P* values for the Ljung-Box statistic (Q) were above 0.05 (Table 2). In sum, the diagnosis of residuals confirms that the developed SARIMA (2, 2, 1) (1, 0, 1)[7] model is appropriate to the trends and seasonality in growth rates. Root mean square error (RMSE) and mean absolute percentage error (MAPE) of the SARIMA model were at 0.049 and 0.775, respectively, which is very small.

Table 2: ARIMA and SARIMA model fit statistics, India, March-August 2020

	BIC [€]	AIC ^{\$}	AICc [¥]	RMSE [‡]	MAPE ^γ	Ljung-Box Q	P^{\wedge}
ARIMA (2, 2, 1)	-336.59	-347.67	-347.32	0.052	0.91	23.87	0.013
SARIMA (2, 2, 1)(1, 0, 1)[7]	-341.38	-358.01	-357.25	0.049	0.78	15.89	0.069

[€]: Bayesian Information Criteria, ^{\$}: Akaike Information Criteria, [¥]: Corrected Akaike Information Criteria, [‡]: Root Mean Square Error; ^γ: Mean Absolute Percentage Error; Q: Ljung-Box statistic; [^]: P values

Figure 3: Diagnosis of the residuals of SARIMA model for COVID-19 cases, India, March-August,2020



SARIMA and exponential model

Table 3 shows the estimates of the coefficients of SARIMA (2,2,1) (1,0,1) [7] model. The coefficients of applied SARIMA model are significant at 1% level. The estimates of autoregressive model of order 1 (AR1) and of order 2 (AR2) of the SARIMA (2,2,1) (1,0,1) [7] are 0.182 ($P=0.089$) and 0.160 ($P=0.177$), respectively. The estimates of moving average of order 1 (MA1) is -0.955 ($P<.001$). The estimates of the seasonal autoregressive model of order 1 (SAR1) and a

The Peak and Size of COVID-19 in India: SARIMA and Forecast

moving average of order 1 (SMA1) are 0.940 ($P < .001$) and -0.778 ($P < .001$), respectively. The significant values of MA1, SAR1, and SMA1 of the SARIMA model reveal valuable effect on the estimates of COVID-19 cases. The SAR1 has a positive effect on the growth rate and hence on COVID-19 cases. Whereas, MA1 and SMA1 have negative effect on the growth rate. Altogether, there is a deceleration in the growth rate, and so, it would show a decline overtime in COVID-19 cases.

Table 3: Parameter estimation of the SARIMA model, COVID-19 disease, India, March to August 2020

	Model	Estimate	Std. Error	z^{\S}	P^{\wedge}
ARIMA	Difference	2.000			
	AR1	0.113	0.106	1.059	0.289
	AR2	0.018	0.113	0.161	0.871
	MA1	-0.900	0.0583	-15.426	0.000
SARIMA	Difference	2.000			
	AR1	0.182	0.107	1.696	0.089
	AR2	0.160	0.119	1.348	0.177
	MA1	-0.955	0.051	-18.557	0.000
	SAR1	0.940	0.052	18.020	0.000
	SMA1	-0.778	0.116	-6.699	0.000

\S : z statistic; \wedge : P values

Table 4 shows the parameter estimates of the Equation (3) for extrapolating the constant 'C'. The values of the slope parameter and adjusted R-square are 0.06 and 75.3, respectively. The beta coefficient value is 0.060 which is small and significant at 1% level.

Table 4: Parameter model of the regression model for projection of the constant 'C'

Linear Model	Coefficient	t^{\S}	P^{\wedge}	Adj. R^2
μ	-5.25	-14.10	0.000	0.753
β	0.060	22.74	0.000	

\S : t statistic; \wedge : P values, Adj. R^2 : Adjusted R-square, μ : constant, β : slope.

Forecast and projection: peak and size of COVID-19 cases

Figure 4 shows the trends in daily COVID-19 cases (grey coloured circles with pipes) until August 2020 and projected daily COVID-19 cases (green coloured circles with pipes) based on SARIMA model (Equation 3) and exponential model (Equation 1) from 1 September 2020 until 28 February 2021.

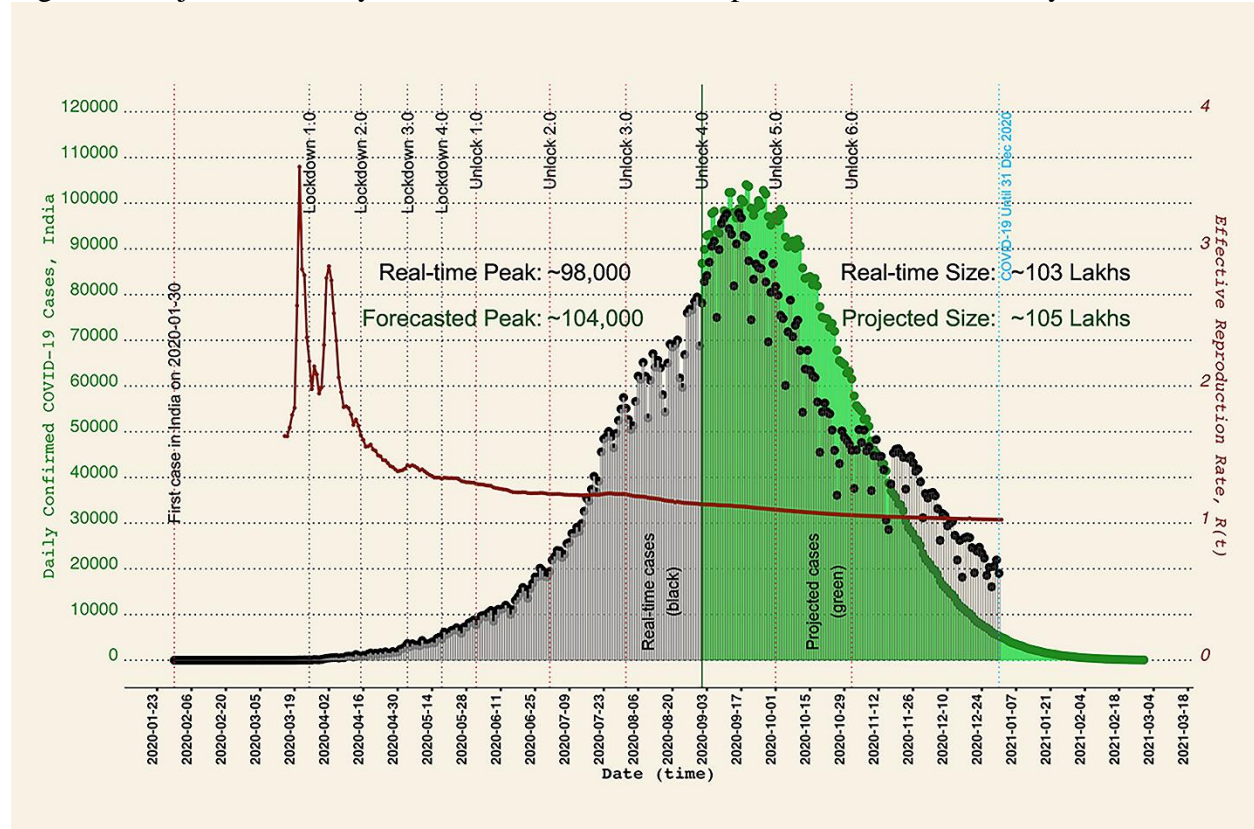
The estimates of SARIMA model shown in Table 3 for forecasting growth rates between 01 September 2020 and 28 February 2021. Along with estimates of constant 'C' shown in Table 4, the projected growth rates were used in the exponential Equation (1) to estimate projected values of COVID-19 cases until 28 February 2021.

The forecasted peak value is estimated at 104,081 counts of COVID-19 cases on 19 September 2020 compared to the real-time peak value of 97,861 COVID-19 cases on 16 September 2020. The Pearson correlation coefficient between daily COVID-19 cases and real-time daily

COVID-19 cases was 67% for the period of 19 days between 1 and 19 September 2020 and 57% between 19 and 30 September 2020.

The projections reckon the cumulative number of COVID-19 cases, i.e., the size of COVID-19 disease accounts for a total of 105 lakhs at the end of December 2020.

Figure 4: Projection of daily COVID-19 cases, India, September 2020 to February 2021.



Discussion

The study demonstrates the epidemiological stage of COVID-19 disease in India by estimating the peak value and the size of COVID-19 cases based on 3,687,944 number of COVID-19 cases between 30 Jan and 31 August 2020. A two-parameter exponential model was applied to the COVID-19 cases for examining the growth or decay rate of COVID-19 cases. A SARIMA model was developed to forecast the growth or decay rate for future time, and hence COVID-19 cases were projected.

The results show that the adjusted R-square of the exponential model was higher than 85% in the studied period (Table 1). The fit of the exponential model shows a shift that confirms a deceleration in the growth rates (Figure 1).

Based on the developed SARIMA model (Table 2), the projection shows a slow, cyclic decline in the growth rates of daily COVID-19 cases since September 2020. Projections show the forecasted peak value of 104,081 counts of daily COVID-19 cases on 19 September 2020 versus

The Peak and Size of COVID-19 in India: SARIMA and Forecast

real-time peak value of 97,861 COVID-19 cases on 16 September 2020 (Figure 4). The declining growth rates underpinned by the significant negative values of SMA1 and MA1 of the SARIMA model affirms the peak value since the third week of September 2020 (Table 3).

More prominently, we predicted the timing of the peak and its magnitude very close to real-time data, for the first time. In comparison to real-time data, the projected daily COVID-19 cases were nearly coinciding with the original data points between 1 September and 30 September 2020. It ensures an independent cross-check compared to other sensitivity analysis, such as comparing other countries and training data. Furthermore, projections demonstrate that the total cumulative number of COVID-19 cases would be approximately 105 lakhs by the end of December 2020. It is confirmed that projections are correct as the real-time COVID-19 cases accounts for 103 lakhs COVID-19 cases at the end of December 2020.

Knowledge of the peak value and its timing is crucial information for government authorities which they are looking forward to. Particularly for India, the precautionary measures vary by the states of India, and the same follows for standard measures such as social distancing, self-quarantined, asymptomatic carriers and transmission rate (Gurdasani et al., 2021; Nabi, 2020; Sarkar et al., 2020). This study used more than five months' data and hence presented the results for a long-term prediction. For the first time, the study shows a close forecast of the peak value of the daily COVID-19 cases, its timing and the size of COVID-19 disease in India.

Limitations of the study

The study is based on daily COVID-19 cases during seven months between the dates 30 January 2020 and 31 August 2020. The projection analysis was performed at the national level only because smaller states have limitations for projections given small samples and differences in the length of period for COVID-19 disease. Apart from statistical and time series assumptions, ARIMA or SARIMA can make projections within its limits of a linear relationship (Zhang et al., 2013) and depend on requiring a new cycle (Nobre et al., 2001; Wang et al., 2020). The behavioural aspects of population and socio-economic and demographics characteristics of the population were not available for consideration in trend analysis and projections because of data constraints. In view of these data constraints, we do have trend and seasonality into account for robust analysis and long-term projection.

Conclusion

The study shows that the SARIMA model is suitable for projecting daily confirmed cases of COVID-19 disease. The study based on moments of the distribution of the daily COVID-19 cases unravels the uncertainty about the peak and the size of COVID-19 disease in India. Based on the applied methodology, we show the peak value of 104,081 counts of COVID-19 cases on 19 September 2020, whereas the real-time peak value of 97,861 COVID-19 cases was conspicuous on 16 September 2020. The forecast and projection of daily COVID-19 cases are very close to the real-time peak value and the size of COVID-19 cases for India. The projected COVID-19 cases are 105 lakhs versus real-time COVID-19 cases of 103 lakhs by the end of December 2020. The study successfully explores the epidemiological stage of COVID-19 disease in India. The study

strongly suggests keeping track of the growth rates of daily COVID-19 cases obtained from the exponential model to understand the flattening and size of COVID-19 disease in India.

Acknowledgements

Authors are grateful to Prof. K. S. James, the Director of International Institute for Population Sciences (IIPS), Mumbai, for approving this research work in the COVID-19 Study Series, and the first draft of this research work was displayed on the website of IIPS which can be retrieved from https://iipsindia.ac.in/sites/default/files/iips_covid19_poci.pdf (doi: [10.13140/RG.2.2.22503.88486](https://doi.org/10.13140/RG.2.2.22503.88486)). We are also thankful to the members of the group “IIPS Team of Researchers on Estimation and Projection of COVID 19 Cases” for their help. The authors are grateful to the reviewers for their comments and suggestions.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Abbreviations

COVID-19: Coronavirus disease of 2019; SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2; ARIMA: Auto regressive integrated moving average; SARIMA: Seasonal auto regressive integrated moving average; H1N1: Hemagglutinin Type 1 and Neuraminidase Type 1; MERS-CoV: Middle east respiratory syndrome coronavirus; SARS-CoV: Severe acute respiratory syndrome coronavirus. ASHA: Accredited social health activist; ANM: Auxiliary nurse midwife; ACF: Autocorrelation function; PACF: Partial autocorrelation function; AIC: Akaike information criteria; RMSE: Root mean square error; MAPE: Mean absolute percentage error; AR1: Autoregressive model of order 1; AR2: Autoregressive model of order 2; SAR1: Seasonal autoregressive model of order 1; SMA1: Seasonal moving average of order 1; MA1: Moving average of order 1; NPIs: Nonpharmaceutical interventions.

References

- Atangana, A., 2020, Modelling the spread of COVID-19 with new fractal-fractional operators: Can the lockdown save mankind before vaccination? *Chaos Solitons Fractals*, 136 (109860), 109860. doi:10.1016/j.chaos.2020.109860
- Bhagat, R. B., Reshmi, R. S., Sahoo, H., Roy, A. K., and Govil, D., 2020, The COVID-19, Migration and Livelihood in India: Challenges and Policy Issues. *Migration Letters*, 17(5), 705-718. doi:10.33182/ml.v17i5.1048
- Bhardwaj, R., 2020, A Predictive Model for the Evolution of COVID-19. *Transactions of the Indian National Academy of Engineering*, 5(2), 133-140. doi:10.1007/s41403-020-00130-w
- Biggerstaff, M., Cowling, B. J., Cucunuba, Z. M., Dinh, L., Ferguson, N. M., Gao, H., . . . Group, W. C.-M. P., 2020, Early Insights from Statistical and Mathematical Modeling of Key Epidemiologic Parameters of COVID-19. *Emerg Infect Dis*, 26(11), e1-e14. doi:10.3201/eid2611.201074
- Ceylan, Z., 2020, Estimation of COVID-19 prevalence in Italy, Spain, and France. *Sci Total Environ*, 729, 138817. doi:10.1016/j.scitotenv.2020.138817

The Peak and Size of COVID-19 in India: SARIMA and Forecast

- Chimmula, V. K. R., and Zhang, L., 2020, Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos Solitons Fractals*, 135, 109864. doi:10.1016/j.chaos.2020.109864
- COVID19-India API., 2020, A volunteer-driven, crowd-sourced database for COVID-19 stats & patient tracing in India (Online). Retrieved from: <https://api.covid19india.org/>
- Dev, S. M., 2020, Income Support Through Cash Transfers and Employment Guarantee Schemes During the Pandemic Times in India. *Indian J Labour Econ*, 1-6. doi:10.1007/s41027-020-00268-9
- Djaoue, S., Guilsou Kolaye, G., Abboubakar, H., Abba Ari, A. A., and Damakoa, I., 2020, Mathematical modeling, analysis and numerical simulation of the COVID-19 transmission with mitigation of control strategies used in Cameroon. *Chaos Solitons Fractals*, 139, 110281. doi:10.1016/j.chaos.2020.110281
- Dong, E., Du, H., and Gardner, L., 2020, An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*, 20(5), 533-534. doi:10.1016/S1473-3099(20)30120-1
- Duan, X., and Zhang, X., 2020, ARIMA modelling and forecasting of irregularly patterned COVID-19 outbreaks using Japanese and South Korean data. *Data Brief*, 31, 105779. doi:10.1016/j.dib.2020.105779
- Gurdasani, D., Alwan, N. A., Greenhalgh, T., Hyde, Z., Johnson, L., McKee, M., . . . Ziauddeen, H., 2021, School reopening without robust COVID-19 mitigation risks accelerating the pandemic. *The Lancet*. doi:10.1016/s0140-6736(21)00622-x
- Jewell, N. P., Lewnard, J. A., & Jewell, B. L. (2020). Predictive Mathematical Models of the COVID-19 Pandemic: Underlying Principles and Value of Projections. *JAMA*, 323(19), 1893-1894. doi:10.1001/jama.2020.6585
- JHU CSSE., 2020, COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Retrieved from <https://github.com/CSSEGISandData/COVID-19>
- Khan, F., Saeed, A., and Ali, S., 2020, Modelling and forecasting of new cases, deaths and recover cases of COVID-19 by using Vector Autoregressive model in Pakistan. *Chaos Solitons Fractals*, 140, 110189. doi:10.1016/j.chaos.2020.110189
- Krishnamurthy, K., Ambikapathy, B., Kumar, A., and Britto, L., 2020, Prediction of the Transition From Subexponential to the Exponential Transmission of SARS-CoV-2 in Chennai, India: Epidemic Nowcasting. *JMIR Public Health Surveill*, 6(3), e21152. doi:10.2196/21152
- Lotto, M., Ayala Aguirre, P. E., Rios, D., Andrade Moreira Machado, M. A., Pereira Cruvinel, A. F., and Cruvinel, T., 2017, Analysis of the interests of Google users on toothache information. *PLoS One*, 12(10), e0186059. doi:10.1371/journal.pone.0186059
- Mehta, M., Julaiti, J., Griffin, P., and Kumara, S., 2020, Early Stage Machine Learning-Based Prediction of US County Vulnerability to the COVID-19 Pandemic: Machine Learning Approach. *JMIR Public Health Surveill*, 6(3), e19446. doi:10.2196/19446
- Mohamadou, Y., Halidou, A., and Kapen, P. T., 2020, A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. *Applied Intelligence*, 50(11), 3913-3925. doi:10.1007/s10489-020-01770-9
- MoHFW, 2020a, Containment plan for large COVID-19 outbreaks. Retrieved from New Delhi: https://prsindia.org/files/covid19/notifications/1689.IND_containment_April4.pdf
- MoHFW, 2020b, Information on identifying areas in need of critical intervention and classifying them as green zones, orange zones, and red zones based on prevalence of COVID-19. (D.O.No.Z.28015/19/2020-EMR). Retrieved from Nirman Bhawan, New

- Delhi:[https://prsindia.org/files/covid19/notifications/Ltr%20to%20CS%2030%20april%20\(2\).pdf](https://prsindia.org/files/covid19/notifications/Ltr%20to%20CS%2030%20april%20(2).pdf)
- Moreau, V. H., 2020, Forecast predictions for the COVID-19 pandemic in Brazil by statistical modeling using the Weibull distribution for daily new cases and deaths. *Braz J Microbiol*, 51(3), 1109-1115. doi:10.1007/s42770-020-00331-z
- Nabi, K. N., 2020, Forecasting COVID-19 pandemic: A data-driven analysis. *Chaos Solitons Fractals*, 139(110046), 110046. doi:10.1016/j.chaos.2020.110046
- Nobre, F. F., Monteiro, A. B., Telles, P. R., and Williamson, G. D., 2001, Dynamic linear model and SARIMA: a comparison of their forecasting performance in epidemiology. *Stat Med*, 20(20), 3051-3069. doi:10.1002/sim.963
- Ogundokun, R. O., Lukman, A. F., Kibria, G. B. M., Awotunde, J. B., and Aladeitan, B. B., 2020, Predictive modelling of COVID-19 confirmed cases in Nigeria. *Infect Dis Model*, 5, 543-548. doi:10.1016/j.idm.2020.08.003
- R Core Team, 2017, R: A language and environment for statistical computing. Retrieved from Vienna, Austria <https://www.R-project.org/>
- RStudio Team, 2020, RStudio: Integrated Development for R. Retrieved from MA: <http://www.rstudio.com/>
- Sarkar, K., Khajanchi, S., and Nieto, J. J., 2020, Modeling and forecasting the COVID-19 pandemic in India. *Chaos Solitons Fractals*, 139(110049), 110049. doi:10.1016/j.chaos.2020.110049
- Sengupta, S., and Jha, M. K., 2020, Social Policy, COVID-19 and Impoverished Migrants: Challenges and Prospects in Locked Down India. *The International Journal of Community and Social Development*, 2(2), 152-172. doi:10.1177/2516602620933715
- Singh, R. K., Rani, M., Bhagavathula, A. S., Sah, R., Rodriguez-Morales, A. J., Kalita, H., . . . Kumar, P., 2020, Prediction of the COVID-19 Pandemic for the Top 15 Affected Countries: Advanced Autoregressive Integrated Moving Average (ARIMA) Model. *JMIR Public Health Surveill*, 6(2), e19115. doi:10.2196/19115
- StataCorp., 2019, Stata Statistical Software: Release 16. Retrieved from College Station, TX:
- Wang, Y., Xu, C., Li, Y., Wu, W., Gui, L., Ren, J., and Yao, S., 2020, An Advanced Data-Driven Hybrid Model of SARIMA-NNAR for Tuberculosis Incidence Time Series Forecasting in Qinghai Province, China. *Infect Drug Resist*, 13, 867-880. doi:10.2147/IDR.S232854
- WHO, 2020, Naming the coronavirus disease (COVID-19) and the virus that causes it. Retrieved from [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it)
- Yadav, S., Yadav, P. K., Kishore, M., Unisa, S., Chander, S., Ram, U., . . . Dey, S., 2020, Basic Reproduction Rate and Case Fatality Rate of COVID-19: Application of Meta-analysis. *Demography India*, 49(Special Issue), 76-97.
- Zhang, X., Liu, Y., Yang, M., Zhang, T., Young, A. A., and Li, X., 2013, Comparative study of four time series methods in forecasting typhoid fever incidence in China. *PLoS One*, 8(5), e63116. doi:10.1371/journal.pone.0063116