Long Run Predictions Using Gompertz Curves- A State Wise Analysis of COVID-19 Infections in India

Abhigayan Adhikary and Manoranjan Pal®

Submission Time: 12 January 2024 **Acceptance Time:**6 March 2024

ABSTRACT

The aim of this paper is to perform a State wise Analysis of the First and the Second COVID-19 Waves experienced by India using the Gompertz Curves and to estimate the maximum number of affected individuals for each wave with the best possible accuracy. A total of 21 large States are chosen for the analysis encompassing 97% of the Indian population. Data on cumulative number of cases is available till 31st October 2021. The entire dataset is segregated into two parts, i.e., the First and the Second Waves and then modelled individually by the Gompertz Curves with some generalizations.

The predicted maximum cumulative numbers of COVID-19 affected individuals are found to be quite accurate. Besides, it is found to be possible to give a methodology how one can predict these numbers with a much smaller dataset. This is important as it can help the authorities in taking an informed decision on the efficient allocation of the limited health care resources.

Keywords: COVID-19, disease modeling, Gompertz Curve, Non-linear least squares, time series, Forecasting, Prediction

JEL Codes: E0, C32, C50, C53

1. INTRODUCTION AND OVERVIEW

The current paper is a contribution to the literature on the analysis of the COVID-19 situation in India. We will explore how the different properties of the Gompertz Curves can be used as a convenient tool to study both the First and the Second Waves and offer an ingenious way to raise a timely alarm for a country to prepare itself.

Most of the Indian States are heterogeneous and have large geographic areas and populations. It is not possible to plan the course of action by taking estimates from the aggregate all India data of COVID-19 infections, because of the inherent variations among the States of India. It may be possible that the aggregate data estimates hardly match that of any of the States. Hence, there arises a need to analyze the data for each State separately. Thus, we carry out the analysis by selecting 21 Indian States, accounting for about 97% of the total Indian population as per the 2020 estimates.

Analytical research on COVID-19 can be primarily subdivided into two parts. A reasonable amount of literature has focused on proposing models for robust estimation of the daily infections while the other has attempted to model the cumulative number of infections. Given

® Abhigayan Adhikary, (email: abhigayan1999@gmail.com), ORCID:0000-0002-7420-6859

Manoranjan Pal, Economic Research Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Kolkata 700108, India (email: manoranjan.pal@gmail.com) ORCID: 0000-0002-5098-7830

these two broad categories of existing work, the present paper shall attempt to cater to the second by implementing the Gompertz Curves. One computational advantage of using the Gompertz Curves is that it can conveniently cater to both these objectives, thereby proving to be a handy tool to cater to both medical and economic aspects of the crisis.

The paper's objective is not only to pointwise map out the trajectory of cumulative infections but also to provide a robust estimate of the “maximum” number of infections that can occur in a particular wave using a smaller dataset. This is crucial, especially from an economic viewpoint, as an early and accurate overall estimate of the maximum possible infections will result in a further efficient allocation of these scarce healthcare resources when limited data is available during the early phases of a new wave of rising infections. What we go on to discuss in the consecutive sections is how the Gompertz Curve as a tool stands out to distinctively analyze the objective of this paper.

The literature on the use of Gompertz Curves in the context of COVID-19 is rather scarce. Mazuruk and Nenickova (2020) use the Gompertz Curve to model the COVID-19 cases of USA and also apply it to the data on COVID-19 deaths. They infer that in both cases, the Gompertz Curve has been able to provide a reasonably good approximation to the data. Mendietaet.al. (2020) also provide a similar conclusion from their analysis on Italy, Spain and Cuba. The authors had considered the Logistic curve and the Gompertz Curve and showed that for both countries, the Gompertz model had better estimates for the peak in confirmed cases and deaths. Rypdal and Rypdal (2020) make a detailed comparison of epidemic curves for Sweden and Norway using the Gompertz curves and they too observe that the epidemic curves for COVID-19-related deaths for most countries with a reliable reporting system are surprisingly well described by the Gompertz growth model. They also suggest that countries with rapid initial growth and slow later decay can be modelled satisfactorily using the Gompertz function.

In the Indian context, Gupta and Kumar (2020) used exploratory data analysis to report the situation in the time period of January to March and used the ARIMA model to predict future trends. They inferred that a huge surge in the number of likely COVID-19-positive cases was predicted in April and May 2020. The average that was forecasted was the detection of approximately 7000 patients in a total span of 30 days in April 2020. However, in reality, the figures were much higher.

A further enhanced study on India was done by Pandey et al. (2020). They used regression models for forecasting. According to them, expected cases may rise to about 5000 in a two-week time period. This was far more accurate than the model predicted by Gupta and Kumar (2020) however actual scenario showed a bigger upsurge.

Ghosh et al. (2020) take a unique approach to forecast COVID-19 infections in India. They consider the State wise data of infections and model them using the logistic and exponential curves. They infer that the predictions from one model might be misleading and hence suggest a linear combination of the exponential and logistic curves for the purpose of realistic predictions.

Note that since the objective of our paper differs from those of the existing literature, our results are not directly comparable with the results of the corresponding papers. Although there is no gainsaying that a pointwise mapping of the COVID-spread is necessary, however, given the huge population that needs to be catered to through relief measures in course of time, taking a myopic view and planning accordingly may lead to undesirable situations in the long run.

The paper is divided into the following sections. Section 2 highlights some of the key aspects of the Gompertz Curves and our observations from all India data followed by the State-wise analysis in Section 3. Section 4 which is the key highlight of the study focused on the Prediction of the maximum number of cumulative infections using smaller datasets and develops a criterion for robust predictions and Section 5 concludes the paper by giving a summary of the findings.

2. THE GOMPERTZ CURVES AND SOME OBSERVATIONS FOR INDIA

2.1. Gompertz Curves and it's properties

For practical purposes, we consider the Gompertz Curve (GC) as:

$$GC: y = e^{a-be^{-ct}} \; where\; a, b, c > 0, \qquad \text{... (2.1)}$$

Throughout the paper we will define:

y = cumulative frequency of daily infections,

N = Maximum cumulative frequency that can be attained (e^{a}) i.e., the asymptote,

b is the displacement on the x-axis and c is the growth rate. Hence, essentially, b and c are the shape parameters.

The following are some of the features of the Gompertz Curve:

Feature 1:

Equation (2.1) can also be extended to construct a generalized version of the Gompertz Curve as:

$$y = e^{a-be^{f(t)}} \; where\; a, b > 0,$$

where, $f(t) = c_1 t + c_2 t^2 + \cdots + c_p t^p \,, p \geq 1,$ is a polynomial in time.

Winsor (1932) mentions that if we wish to use only a finite number of terms in the power series, we must keep an odd power of t for our highest term, so as to get a finite asymptote of y. Again, taking high degree polynomial involving many parameters may lead to estimation problems along with the efficiencies of the estimates. Hence, we consider our Generalized Gompertz Curve (GGC) as:

$$GGC: y = e^{a-be^{c_1 t+c_2 t^2+c_3 t^3}}, a, b > 0, \qquad \text{... (2.2)}$$

For the GGC also, $N = e^{a}\; provided\; c_3 < 0.$

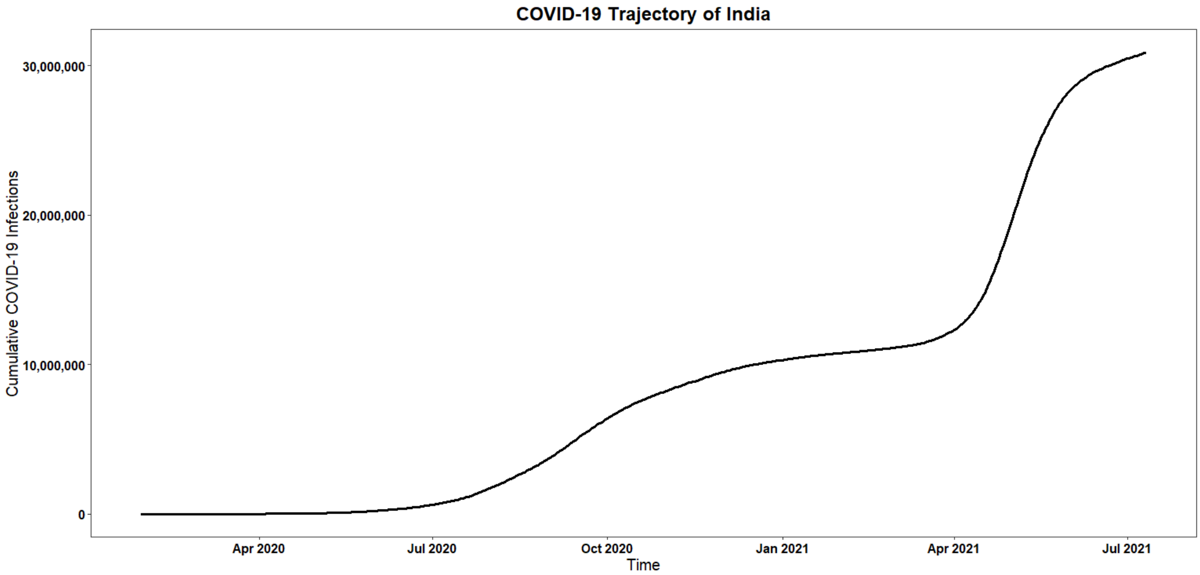
Feature 2:

The Gompertz Curves provides us with an add-on benefit of not only enabling us to make short run predictions (say for a period of 10 days, 20 days etc.) but long-term predictions as well. This is perfectly aligned with the objective of our paper as we try to make robust estimations of the maximum possible cumulative infections which is given by $\lim_{t\to\infty} y = e^{a}$.

2.2. The Indian Scenario

We first plot all India data over time as shown in Figure 2.1. As expected, the cumulative number of daily infections follows a sigmoid shape. However, since two major variants of COVID-19 drove the number of affected individuals, we observe two COVID-19 waves each having the expected sigmoid shape. Similar plots have been obtained for each of the States chosen for the analysis.

Figure 2.1: Daily Cumulative Infections over Time at the All-India Level



However, analysis of the two waves cannot be done together and hence there arises a need to split our datasets into 2 parts for the first and the second waves respectively. However, scientifically it has not been possible to demarcate a well-defined cutoff point for marking the end of the first wave and the beginning of the second wave. Hence, for the purpose of proceeding with the analysis, we need to subjectively decide on the cutoff points based on the above plot. It is worth mentioning that subjective determination of the cutoff points is not a point of grave concern because the predictions obtained by varying the cutoff points only differ marginally. The subjectively chosen cutoff points as per the author's discretion for each of the chosen States has been given in the Appendix Table A1.

The GC and the GGC can now be fitted to the data using the method of non-linear least squares (Gujarati et al.2021). However, we first need to obtain initial starting values of our parameters for non-linear least squares.

We get the initial estimates by using the following linearized model. To illustrate the steps, we use the GC:

$$y = e^{a-be^{-ct}}$$

$$\ln(y) = a - be^{-ct}$$

$$\ln(a - \ln(y)) = \ln b - ct$$

$$\ln(\bar{a} - \ln(y)) = \mathrm{b}^* + c^* t, \qquad \text{... (2.3)}$$

Hence with known $\bar{a}$ we can obtain $b = e^{b^*}$ and $c = -c^*$

Similarly, the expression for the GGC becomes:

$$\ln(\bar{a} - \ln(y)) = \mathrm{b}^* + c_1 t + c_2 t^2 + c_3 t^3, \qquad \text{... (2.4)}$$

Similarly in this case, with known $\bar{a}$ we can obtain $b = e^{b^*}, c_1, c_2, c_3$.

For obtaining the starting values of $\bar{a}$, we proceed in the following way:

For a, we know $y_{max} = e^a$. From the dataset, this can be taken as the cumulative number of infections on the last day of the wave. We increment this by 20% and initialize a as $\bar{a} = \ln(1.2\, y_{max})$. However, for partial data one may not be able to guess the initial estimate of $\bar{a}$, and must take several trial values of $\bar{a}$ and choose the one which gives the least value of the sum of the squared errors in the estimation of equation (2.4).

3. THE STATE WISE ANALYSIS

Having listed the 21 States under consideration and our observations from the all-India data, we can now evaluate whether the Gompertz Curve gives a reasonably good fit on the State level data. Data used in this paper pertains to the cumulative number of infections till 31st October 2021.

We will now use the method of nonlinear least squares for fitting the curves for both the Waves by plugging in the initial estimates. Hence, the two models under our consideration are:

- The Gompertz Curve, $GC: y = e^{a-be^{-ct}}$,
- The Generalized Gompertz Curve, $GGC: y = e^{a-be^{c_1 t + c_2 t^2 + c_3 t^3}}$,

Which model will fit better is decided on the basis of $\hat{\sigma}$, the standard deviation of the residuals, which is obtained as $\hat{\sigma} = \sqrt{\frac{RSS}{T-k}}$ (where, RSS is the Residual Sum of Squares from the fitted regression, T is the total number of data points and k indicates the number of regressors), as presented in the Tables 1 and 2.

Hence, we can observe unanimously that the GGC is a better fit in all the States for both the waves. GGC fit is a significant improvement over the GC fit. It also follows from the fact that at least one of c_2 or c_3 is significant for each state (not shown here). Note that only the $\hat{\sigma}$ values are reported as our conclusions remains unchanged with AIC criterion.

A summary of our results from the GGC Fit on the State level data has been given in the Appendix Tables A2, A3.

4. PREDICTION OF THE MAXIMUM CUMULATIVE INFECTIONS THROUGH A SMALLER DATA SET

Given the devasting intensity to which we have witnessed this pandemic, the fundamental target of any analyst would be to make an attempt to provide robust estimates of the maximum number of people that are expected to get affected by a COVID-19 wave as soon as possible.

The objective of the present section is not merely to predict the cumulative number of infections on the basis of the fitted curve but also to give reasonably good estimates through a set of data with fewer number of observations. In the initial days of the onset of COVID-19; the government and other administrative agencies were interested in having a robust estimate of the maximum possible infections in each of the States with the limited data that was available. Hence, the current discussion provides some insights on how we can obtain the best possible prediction of the maximum possible infections with a smaller subset of dataset of only a few initial days. Since the GGC is more appropriate for modelling COVID-19 infections in the Indian States, for prediction purposes we shall continue with the GGC in this section.

Suppose, our objective is to find the time point that gives the best possible prediction for a given subset of the available data: $\{y_t\}$ such that $t \in \{t_1, t_1 + 1, \cdots, t_2\}$ where $t_2 < T.$ Let us define:

$N_T =$ predicted maximum cumulative infections from the entire dataset

Thus, $N_T = e^{\widehat{a_T}}$

Also let us denote:

N_t = Predicted maximum cumulative infections obtained from the data till time t, where $t \in \{t_1, t_1 + 1, \cdots, t_2\}$. So, $N_t = e^{\widehat{a_t}}$

The optimum t value for which $| N_t - N_T|$ is minimum cannot be obtained theoretically simply because N_T is not known in this case. All we can do is to give some empirical techniques and choose the best given the data. The technique is to be empirically tested for the data of each State to arrive at some conclusion. Before we go into our proposed technique let us try to visualize the situation. Suppose t_0 is the optimum value, i.e., $Min_t| N_t - N_T|$ is attained at t = t_0. However, it is important to note that although $| N_t - N_T|$ should be a monotonically decreasing function of t, this need not necessarily be true empirically. To illustrate, let us consider the following plots from the analysis of the Second Wave for All-India data for the first 80 days, as observed from Figures 4.1 and 4.2.

Figure 4.1: A Time Series Plot of $| N_t - N_T|$ for the Second Wave for All-India

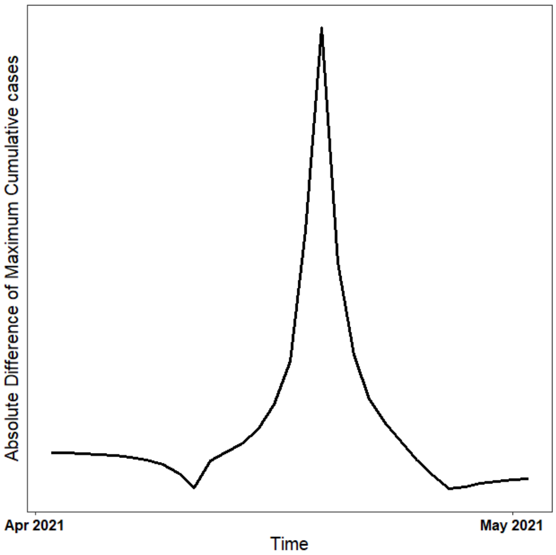
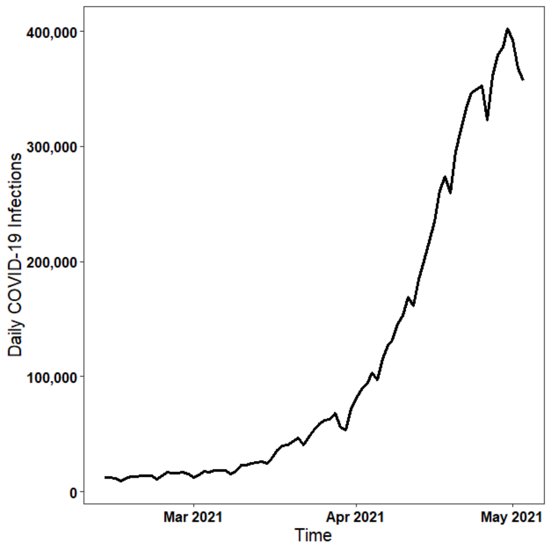


Figure 4.2: Daily Cases at the All-India Level for 80 days in the Second Wave



A possible reason for this counterintuitive observation can be attributed to the fact that COVID-19 infections had not begun in all parts of the country at around the similar time. Similar plots have been obtained with the State-level data as well. Further, at the onset, the urban areas had witnessed a higher infection rate as compared to their rural counterparts. These heterogeneities can only be eliminated by performing the analysis on the data for a further granular level for each of the states, which, however is not possible due to the lack of reliable data. Hence, from the above Figure 4.1 and 4.2, it is clear that taking more data does not necessarily lead to better predictions. However, we can make some critical observations on the above plots. It is clear that $| N_t - N_T|$ tends to get overestimated when the data shows a "sudden jump/spike" in the number of daily cases and underestimated when there is a sudden fall in the data of daily number of infected cases. Sudden rise/fall is reflected by the high positive/negative values of the consecutive differences. This leads to a criterion which is neither close to the minimum or maximum values of the first differences. Hence, to get robust predictions, criterion should be taken in such a manner that will be immune to drastic fluctuations and that is nothing but the median or close the median values of the first differences.

We proceed by fixing an interval $\{t_1, t_1 + 1, \cdots, t_2\}$. Define c_t as the cumulative number of cases at a time point t, $t = 1,2, \dots, T$. Instead of finding the daily cases by taking the difference $c_t - c_{t-1}$, we take the daily proportions $\mathrm{p_t} = \frac{\mathrm{c_t - c_{t-1}}}{\mathrm{c_{t-1}}}$ to make it comparable over different time points as the cumulative values are ever non-decreasing. We now take the first difference of $\mathrm{p_t}$ as

$$\Delta \mathrm{p_t} = (\mathrm{p_t} - \mathrm{p_{t-1}}),$$

where Δ is the first difference operator. Since $\mathrm{p_t}$ is the daily proportion of infected persons, high positive values of $\Delta \mathrm{p_t}$ reflects sudden jump and high negative values of $\Delta \mathrm{p_t}$ reflects sudden fall, we take the median of the $\Delta \mathrm{p_t}$ values and take the corresponding t value as the desired time point.

Another criterion which follows from the same logic described above is to get the optimum t value by minimizing the absolute values of the first differences. We may call it Minimum Absolute First Difference (MIFD) criterion.

The above-mentioned reasonings are justified when the daily infection rates have chances to rise as well as to fall. But in the beginning of the wave, i.e., after one month from the start of the wave, say, the daily infection rates will only rise. Hence, there is no negative differences. Given that daily infection data are increasing, MIFD criterion should be almost same as the Minimum criterion. So, instead of MIFD we may take the t value for which Δp_t is the minimum. However, it may not be a sound criterion to take, because in this case, for optimum t, p_t will be almost same as p_{t-1}. The usual growth of the daily infections is not reflected here.

In any case, let us see the efficacy of the Median and the Minimum criteria for the data of each of the chosen States for both the waves. t_1 and t_2 are taken as 30 and 40 as illustration. The findings are summarized in Tables 3 and 4. Note that N_T^{act} refers to the actual maximum number of cumulative cases for each of the waves as obtained from the data. This is essentially the cumulative infections as observed on the last day of a particular wave.

One can see from Tables 3 and 4, that except in one state each in the first and the second waves, percentage errors in the predictions by Median criterion are far less than that Minimum criterion. Thus, for both the Waves 1 and 2, the Median criterion gives better predictions than the Minimum criterion.

5. DISCUSSION AND CONCLUSION

Very recently the world has faced a sequence of waves of COVID-19 pandemic. India is too not an exception. Since complete State wise data on cumulative number of reported COVID-19 cases for the first two waves are available for India, we have made an attempt to find out the maximum number of COVID-19-affected persons using an appropriate model. Since Gompertz curve is known to give good fit to the cumulative number of affected persons in similar situations, we used the same, but generalized it so that the model accommodates the inherent variations across the States in India, and gives better fit to the data for each of the States. However, non-medical interventions like lockdowns and quarantines could not be incorporated in the model. Further, daily state level data on the spatial distribution of the COVID-19 variants was also not available. This can be seen as a limitation of the study.

The GGC was fitted to the data of major 21 States in India and the fit was very good for each State for both the waves. Before going to the State level data, we carried out the fitting for the all-India data, and the Generalized Gompertz curve gave very good fit to the data in both the waves (Pal and Adhikary, 2022).

Along with the goodness of fit, we were also concerned with the robustness of the fitting. The key finding of our paper is that the GGC is seen to be capable of being robust for long run predictions (i.e., estimating of the "maximum" number of infections that can occur in a particular wave) even with smaller datasets. In order to get an idea how robust it is, we carried out predictions using smaller datasets, but the choice of cut-off points of time has been a major problem, because the daily affected number of cases have up and down movements affecting the cumulative values also especially for the time points in the beginning of the wave. To tackle this problem, instead of taking a single point, we have taken an interval of some consecutive time points and have devised a criterion, which gives reasonably good fit to the entire data set, especially the maximum number of persons who are likely to be affected in each wave. The suggested cut-off point is the time point which corresponds to the median of the first differences of the proportional changes of cumulative number of COVID-19 affected persons. As a similar methodology is yet not available in the literature, the efficacy of the Median criterion is tested against an alternative criterion (MIFD) as discussed in Section 4 of this paper. The Median criterion has turned out to be the best and is seen to be performing reasonably well for both the First and the Second Waves with a prediction accuracy of around 85% for both the waves even with as less data as the initial 30-40 days of onset of the wave as shown in Table 3 and 4.

A closer inspection of data also reveals that more data need not necessarily mean better predictions. Once the time interval is chosen, the median criterion always gives a unique time point such that if we take data till this point, we will get reasonably good prediction for the given limited dataset. The criteria can be easily extended to any other time interval, hence is quite flexible. As a part of future research, the methodology can be readily used to study the COVID-19 waves of other countries and the impact of vaccinations can also be incorporated as an extension to enhance the efficacy of the median criterion.

Availability of data and materials: The COVID-19 Dataset on India is available at: COVID-19 India API accessed on 29th January 2022. The updated dataset can be downloaded from: https://data.covid19india.org/

Table 1: Fitted Models for the First Wave

States	$\hat{\sigma}_{GC}$	$\hat{\sigma}_{GGC}$
Telangana	2255	1654
Assam	4292	1544
Jharkhand	1846	1659
Bihar	4817	3983
Madhya Pradesh	5064	4329
Himachal Pradesh	2250	876.3
Gujarat	2192	2040
Chhattisgarh	3665	3174
West Bengal	10520	4749
Odisha	5534	2213
Uttarakhand	2371	2248
Andhra Pradesh	5702	4370
Karnataka	14290	10270
Maharashtra	48990	21070
Punjab	3867	3813
Tamil Nadu	7544	4360
Haryana	5393	4524
Uttar Pradesh	7661	7245
Rajasthan	6978	3328
Delhi	27190	8228
Kerala	14690	12820

Table 2: Fitted Models for the Second Wave

State	$\hat{\sigma}_{GC}$	$\hat{\sigma}_{GGC}$
Telangana	8433	2150
Assam	3515	3085
Jharkhand	3155	1030
Bihar	4614	1532
Madhya Pradesh	12440	2277
Himachal Pradesh	3899	2025
Gujarat	14850	2667
Chhattisgarh	9552	3981
West Bengal	12670	5505
Odisha	5436	4779
Uttarakhand	4386	1383
Andhra Pradesh	10410	9589
Karnataka	25580	15840
Maharashtra	77840	50410
Punjab	16350	11240
Tamil Nadu	35330	12340
Haryana	12590	1465
Uttar Pradesh	77840	50410
Rajasthan	12370	1956
Delhi	12680	3670
Kerala	83030	13830

Table 3: Comparison of the median and the MIFD criteria for the First Wave

States	N_T^{act}	t_{med}	Percentage error	t_{min}	Percentage error
Telangana	298057	40	13.75	33	41.64
Assam	216992	40	14.11	32	37.88
Jharkhand	119283	40	15.64	33	48.11
Bihar	261068	36	17.31	32	47.04
Madhya Pradesh	253405	39	13.69	38	45.72
Himachal Pradesh	57296	38	17.69	34	6.51
Gujarat	261838	31	18.06	30	42.15
Chhattisgarh	295949	30	15.52	40	41.78
West Bengal	573387	38	15.97	30	49.20
Odisha	336460	37	13.60	40	38.85
Uttarakhand	96964	38	14.46	30	38.91
Andhra Pradesh	887066	34	16.39	30	43.02
Karnataka	934576	37	15.31	30	47.26
Maharashtra	1906371	40	17.57	35	39.94
Punjab	171522	39	14.98	30	45.25
Tamil Nadu	837832	39	15.79	35	42.97
Haryana	266309	33	18.00	40	46.77
Uttar Pradesh	596528	35	14.71	31	39.17
Rajasthan	318021	35	16.47	33	45.60
Delhi	635639	32	17.34	30	44.50
Kerala	1167191	36	20.88	34	48.57

Table 4: Comparison of the median and the MIFD criteria for the Second Wave

States	N_T^{act}	t_{med}	Percentage error	t_{min}	Percentage error
Telangana	669932	35	2.91	34	18.93
Assam	576149	31	22.81	35	63.86
Jharkhand	347440	32	6.03	40	33.95
Bihar	725235	30	0.75	35	63.09
Madhya Pradesh	791970	31	25.46	38	3.630
Himachal Pradesh	206589	39	24.97	36	68.97
Gujarat	825085	34	20.22	30	57.67
Chhattisgarh	1003439	30	1.52	38	68.06
West Bengal	1534999	31	12.67	40	43.97
Odisha	990075	38	10.44	33	57.46
Uttarakhand	342502	33	28.66	30	47.64
Andhra Pradesh	1987051	34	12.88	30	54.80
Karnataka	2921049	31	14.65	40	52.46
Maharashtra	6363442	34	16.79	31	67.03
Punjab	599678	36	20.03	33	63.55
Tamil Nadu	2581094	37	15.68	30	39.21
Haryana	770130	33	24.76	31	63.99
Uttar Pradesh	1708836	34	21.05	38	63.83
Rajasthan	953870	36	0.26	33	64.44
Delhi	1436889	30	0.21	35	53.02
Kerala	4283494	38	14.54	34	41.20

APPENDIX:

Table A1: Subjectively determined cutoff points for the First Wave and the Second Wave

States	**First Case Reported on:**	**Start of First Wave**	**End of First Wave**	**Start of Second Wave**	**End of Second Wave**
Telangana	2 March 2020	21 March 2020	24 February 2021	26 March 2021	22 October 2021
Assam	31 March 2020	19 May 2020	24 January 2021	13 February 2021	10 August 2021
Jharkhand	31 March 2020	19 April 2020	13 February 2021	15 March 2021	10 August 2021
Bihar	22 March 2020	10 April 2020	4 February 2021	24 February 2021	9 August 2021
Madhya Pradesh	20 March 2020	8 April 2020	23 January 2021	12 February 2021	9 August 2021
Himachal Pradesh	14 March 2020	22 May 2020	27 January 2021	16 February 2021	3 August 2021
Gujarat	19 March 2020	7 April 2020	1 February 2021	11 February 2021	10 August 2021
Chhattisgarh	19 March 2020	27 May 2020	22 January 2021	2 February 2021	11 August 2021
West Bengal	17 March 2020	5 April 2020	19 February 2021	11 March 2021	10 August 2021
Odisha	16 March 2020	14 May 2020	18 February 2021	10 March 2021	11 August 2021
Uttarakhand	15 March 2020	23 May 2020	17 February 2021	10 March 2021	11 August 2021
Andhra Pradesh	12 March 2020	31 March 2020	25 January 2021	14 February 2021	11 August 2021
Karnataka	9 March 2020	28 March 2020	22 January 2021	11 February 2021	10 August 2021
Maharashtra	9 March 2020	7 April 2020	23 December 2020	12 January 2021	10 August 2021
Punjab	9 March 2020	7 April 2020	22 January 2021	12 February 2021	11 August 2021
Tamil Nadu	7 March 2020	26 March 2020	30 January 2021	20 February 2021	11 August 2021
Haryana	4 March 2020	23 March 2020	17 January 2021	6 February 2021	11 August 2021
Uttar Pradesh	4 March 2020	22 April 2020	17 January 2021	6 February 2021	11 August 2021
Rajasthan	3 March 2020	22 March 2020	5 February 2021	26 February 2021	11 August 2021
Delhi	2 March 2020	21 March 2020	4 February 2021	24 February 2021	11 August 2021
Kerala	30 January 2020	16 April 2020	11 April 2021	21 April 2021	8 September 2021

Table A2: Parameter Estimates from the GGC Model for the First Wave

States	a	b	c_1	c_2	c_3
Telangana	12.71	8.68	-0.003	-0.000096	-0.00000020
Assam	12.28	6.03	-0.0128	-0.000028	-0.00000049
Jharkhand	11.68	4.88	0.0177	-0.000295	-0.00000057
Bihar	12.53	42.76	-0.034	0.000041	-0.0000000008
Madhya Pradesh	12.60	4.49	0.003	-0.000098	-0.00000017
Himachal Pradesh	10.96	1678.26	-0.1456	0.001066	-0.00000279
Gujarat	12.61	5.38	-0.0115	0.000016	-0.00000007
Chhattisgarh	12.78	24.61	-0.0273	0.000023	-0.00000003
West Bengal	16.05	12.52	-0.0099	0.000016	-0.0000000005
Odisha	12.72	4.33	0.0079	-0.000214	-0.00000029
Uttarakhand	11.54	36.85	-0.0474	0.000197	-0.00000042
Andhra Pradesh	13.74	7.21	0.0179	-0.000314	-0.00000066
Karnataka	13.88	4.62	0.0188	-0.000262	-0.00000054
Maharashtra	14.43	6.91	-0.0174	0.000108	-0.00000061
Punjab	12.07	11.05	0.0008	-0.000159	-0.00000031
Tamil Nadu	13.75	6.41	-0.0019	-0.00011	-0.00000024
Haryana	12.55	19.06	-0.0271	0.000122	-0.00000034
Uttar Pradesh	13.34	6.70	-0.0004	-0.00015	-0.00000032
Rajasthan	12.70	15.05	-0.0263	0.000134	-0.00000038
Delhi	13.37	132.59	-0.0815	0.000521	-0.00000120
Kerala	14.01	405.63	-0.0573	0.000197	-0.00000030

Table A3: Parameter estimates from the GGC Model for the Second Wave

States	a	b	c_1	c_2	c_3
Telangana	17.30	4.85	-0.0043	0.000028	-0.00000006
Assam	13.24	0.92	0.0017	0.000055	-0.00000143
Jharkhand	12.75	1.02	0.0045	-0.000039	-0.00001028
Bihar	13.56	0.8	0.0412	-0.001027	0.00000418
Madhya Pradesh	13.68	0.94	0.0326	-0.000776	0.00000298
Himachal Pradesh	12.51	1.24	0.0256	-0.000579	0.00000225
Gujarat	13.74	0.98	0.0303	-0.000710	0.00000269
Chhattisgarh	13.93	0.98	0.0316	-0.000667	0.00000235
West Bengal	14.30	0.86	0.0289	-0.000884	0.00000383
Odisha	13.79	0.96	0.0142	-0.000370	0.00000010
Uttarakhand	12.74	1.27	-0.0068	0.000521	-0.00001247
Andhra Pradesh	14.91	1.07	0.0139	-0.000286	0.00000101
Karnataka	15.16	1.15	0.0223	-0.000467	0.00000168
Maharashtra	15.83	1.12	0.0164	-0.000322	0.00000096
Punjab	13.50	1.24	0.0161	-0.000461	0.00000177
Tamil Nadu	15.24	1.39	0.0155	-0.000347	0.00000130
Haryana	13.75	0.98	0.0278	-0.000575	0.00000207
Uttar Pradesh	14.47	0.85	0.0362	-0.000748	0.00000271
Rajasthan	13.83	0.86	0.0413	-0.001075	0.00000445
Delhi	14.18	0.87	-0.0184	0.001008	-0.00001573
Kerala	15.50	1.54	-0.0258	0.000273	-0.00000131

REFERENCES

COVID-19 India API (2022). The COVID-19 Dataset on India. https://data.covid19india.org (Accessed on January 29, 2022)

Ghosh P., R. Ghosh, B. Chakraborty (2020). COVID-19 in India: State wise Analysis and Prediction. *JMIR Public Health Surveill*, 6(3).

Gujarati Damodar N, D. C. Porter and M. Pal (2021), Basic Econometrics, *McGraw Hill, Sixth Edition, Special Indian Edition, pp. 519-532.*

Gupta R. and S. Kumar (2020). Trend Analysis and forecasting of COVID-19 outbreak in India. *MedRxiv*.

Jukić, D., G. Kralik and R. Scitovski (2004). Least-squares fitting Gompertz curve. *Journal of Computational and Applied Mathematics*, 169(2), 359-375.

Mazuruk J. and Z. Nenickova (2020). Predicting the Number of total COVID-19 cases and deaths in USA by Gompertz Curves. *Accessed: Jun, 23*

Mendieta J.F, M. Cortes, M. Cortes (2020). COVID-19 Forecasts for Cuba using Logistic Regression and Gompertz Curves. *MEDICC Review*, 22, pp.32-39.

Mittal S. (2020). An Exploratory Data Analysis of COVID-19 in India", *International Journal of Engineering Research & Technology (IJERT)*, 9(4) April-2020

Ohnishi A., Y. Namekawa and T. Fukui (2020). Universality in COVID-19 spread in view of the Gompertz Function. *Prog. Theor. Exp. Phys. 2020, 123J01, Published by Oxford University Press on behalf of the Physical Society of Japan.*

Pal, M and A. Adhikary (2022). Prediction of Number of Covid-19 Affected Persons by Using Gompertz Curves, Manuscript, *Preprint*

Pandey G., P. Chowdhury and R. Gupta (2020). SEIR and Regression Model based COVID-19 outbreak predictions in India. *arXiv preprint arXiv:2004.00958*

Rypdal K. and M. Rypdal (2020). A Parsimonious Description and cross-country Analysis of COVID-19 Epidemic Curves. *International Journal of Environmental Research and Public Health*, 17(18), p.6487.

Winsor C.P (1932). The Gompertz Curve as a Growth Curve. *Proceedings of the National Academy of Sciences*, 18(1).