The Doubling Period can Seriously Mislead: A New Method to Track the Spread of Corona Infections

Neeraj Hatekar[1]
Pallavi Belhekar

Abstract

This paper outlines a fresh computational method to test whether the "Covid-19 Curve" is flattening out. Generally, the administration in India has been using the "doubling rate" to track the progress of the disease. We show that the doubling rate is a highly uncertain measure. The paper applies the new method to several countries and also states within India .

[1] Neeraj Hatekar teaches Economics at the University of Mumbai . Pallavi Belhekar teaches economics at Smt. MMK College of Commerce and Economics, Bandra, Mumbai. The views expressed in this paper are author's personal views and do not represent the views of the institutions to which they belong. The authors are thankful to Prof. D.M. Nachane and Prof. Avadhoot Nadkarni for extensive discussions of the ideas in this paper and to Prof. H.D. Vinod for pointing the author to his maximum entropy bootstrap method and extensive comments on a previous draft.

Executive Summary:

- **The Doubling Period, though easy to understand, is a highly uncertain quantity, with a variability that makes it of doubtful value for policy**
- **We show, using a bootstrapping technique, that even though the observed doubling rate of Covid-19 Infections in Maharashtra on 5th May was 9, it could have been anywhere between 6 to 13 with 80% chance. The feature holds generally across regions. This makes it difficult to use for policy. We need a better measure.**
- **The paper develops a new measure called r, which is helpful not only in day to day tracking of the infections, but also enables us to say whether the disease is in an exponential growth phase or a sub exponential growth phase.**
- **r could be used to classify districts / regions between green, orange and red zones more efficiently than the doubling rate, because r is relatively more stable and also amenable to statistical testing to determine classification on a daily basis.**

How Fast Does It Spread? A New Method to track the spread of Corona infections

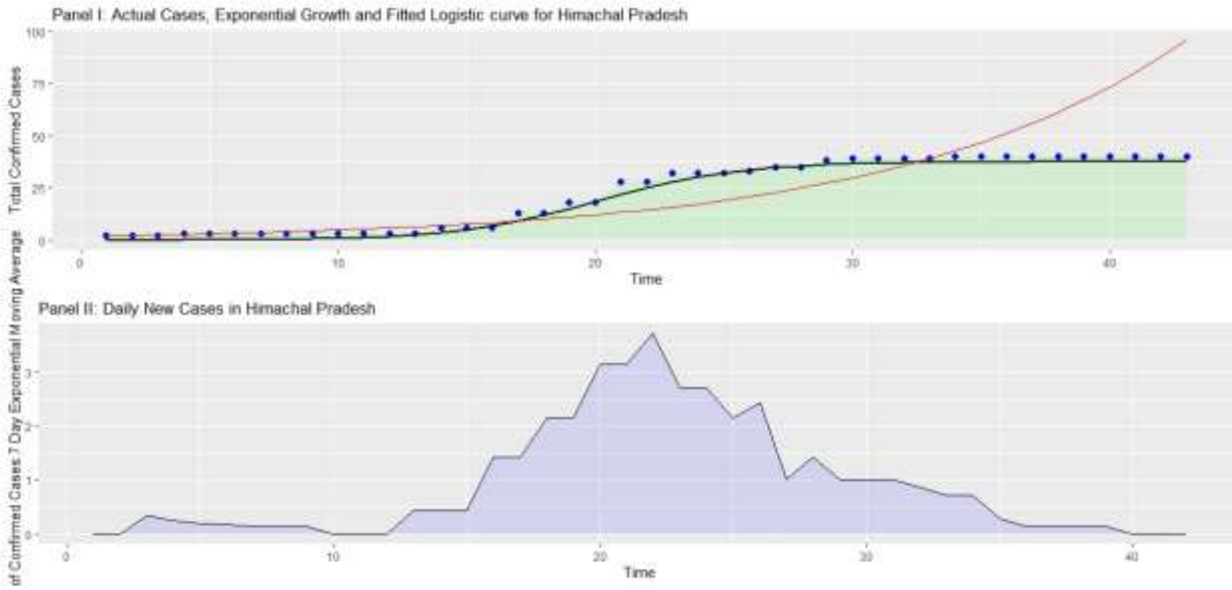
Neeraj Hatekar[2]
Pallavi Belhekar

Introduction:

The word “Exponential Growth” has entered into the public vocabulary when referring to the growth rate of virulent infections like the Covid-19. However, it would be more correct to say that Corona infections in a closed region (where people are not coming in nor are, they moving out) follow an “S” shaped trajectory, rather than the classical exponential growth. Imagine that you have a pond that is full of all nutrients needed for a thriving fish population, but in which, to start with, there are no fish. Suppose now we introduce a pair of fish in the pond. The fish will settle down, and because the conditions are exactly right, with plenty of food, start breeding. The fish population will start growing. For a while, it will explode, but it certainly cannot grow indefinitely. At some time, the food will be just sufficient to support the existing population, but not growth. At that point, the fish population in the pond will stabilize. So, there will be a period of relatively slow increase in numbers, then the numbers will increase really rapidly, and finally will stop increasing. This behavior is described by a “logistic curve”. The logistic curve is commonly used in biology to describe microbial growth. It can of course be estimated from the data. This is also how Corona infections in a closed population behave. Figure 1 below shows the time path of Corona infections in the Indian state of Himachal Pradesh.

[2] Neeraj Hatekar teaches Economics at the University of Mumbai. Pallavi Belhekar teaches economics at Smt. MMK College of Commerce and Economics, Bandra, Mumbai. The views expressed in this paper are author’s personal views and do not represent the views of the institutions to which they belong. The authors are thankful to Prof. D.M. Nachane and Prof. Avadhoot Nadkarni for extensive discussions of the ideas in this paper and to Prof. H.D. Vinod for pointing the author to his maximum entropy bootstrap method.

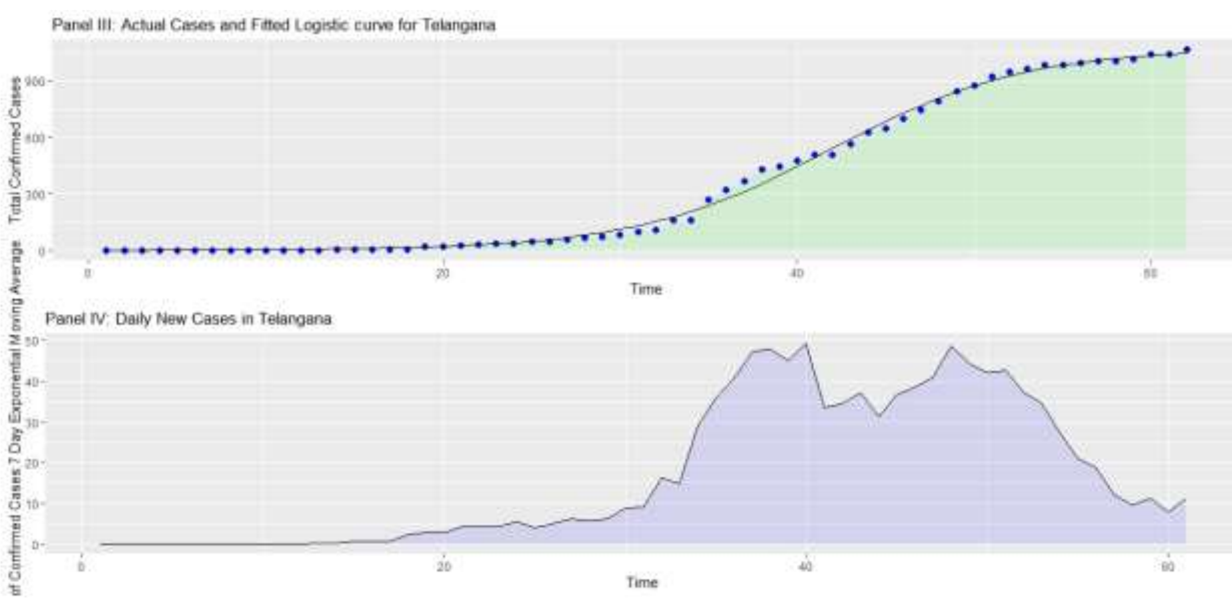
Figure 1



We have time on the X axis in both the panel. (Instead of putting real dates on the X axis, we have time periods). In the first panel, the blue dots represent the cumulative number of infections in Himachal Pradesh over time from the first infection to 3rd May 2020. The red line is a hypothetical exponential growth curve. The lower panel has new cases on the same days on the Y axis. As you can see, to start with, the actual data grows less slowly than you would expect from an exponential curve. This is when new cases are coming at a fixed rate. Then, new cases start coming faster and faster. When this happens, our blue dots start increasing faster than the exponential. The rate of increase of cases reaches a peak, and then starts to come down. At that point, you can see that the blue dots start to become flatter. This is what they mean when experts say that the "doubling period" is increasing. Eventually, the curve flattens, and cases grow at a slower pace compared to the exponential. Finally, as new cases stop coming, the curve in the upper panel flattens. That is the point at which the number of Corona infected patients becomes constant. One can always estimate the logistic curve from the data by using a technique called non-linear least squares. We have done that for the data in panel I. The fitted data are represented by the black line passing through the blue dots. As you can see, the logistic curve fits the data quite well.

Himachal Pradesh is not the only state in India that has followed tis pattern. Figure 2 shows that the southern state of Telangana has also followed the same path (it is fair to say that all states and countries also necessarily follow this path). The number of infections has been much higher, but the S shape is obvious. The blue dots in panel III are the cumulative number of infections. Initially, cases grow at a slower rate than the exponential, then the rate escalates, and finally tapers off. It all depends upon the rate at which new cases appear each day. The black line in panel III is the fitted logistic curve. In this case, the flat portion has not been reached yet, but seems eminent. Again, you can see, the logistic curve describes the data rather well.

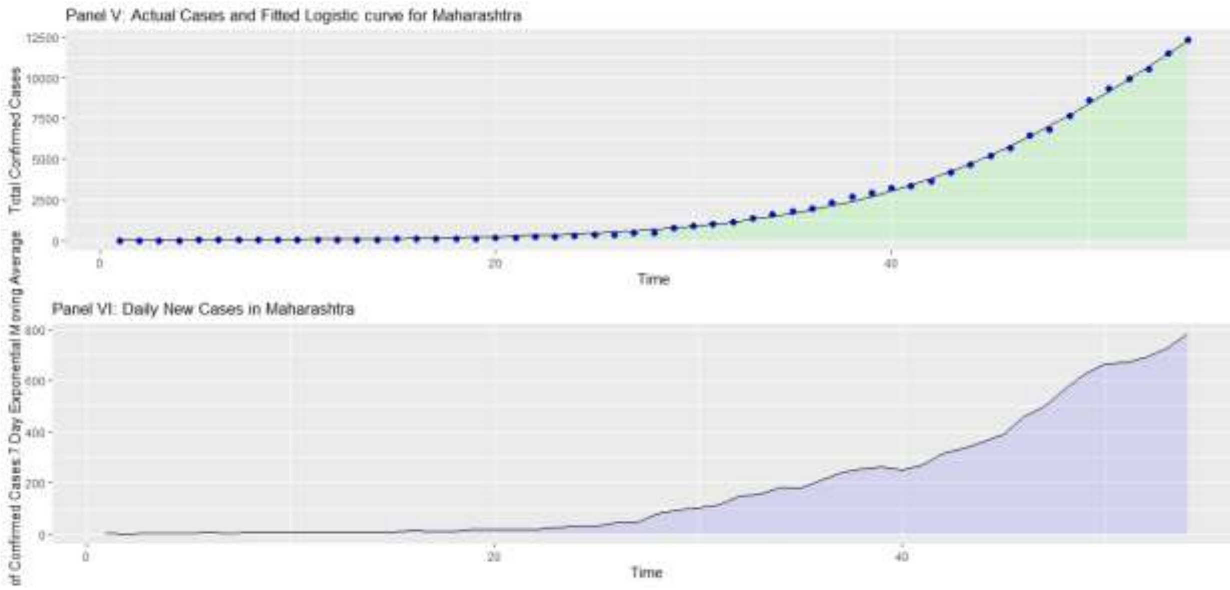
Figure 2



By its very logic, the time path of cumulative Corona cases must describe an S curve. With measures like lockdown, social distancing, the spread slows. That means the point at which the curve becomes flat appears later in time, and most importantly, at a lower height on the Y axis. Lockdowns do not ensure that the curve becomes flat; they are meant to push back that date, so that administration has more time to prepare for the load that is coming.

Not all states have reached the flat phase. Figure 3 has one of those states, Maharashtra.

Figure 3:



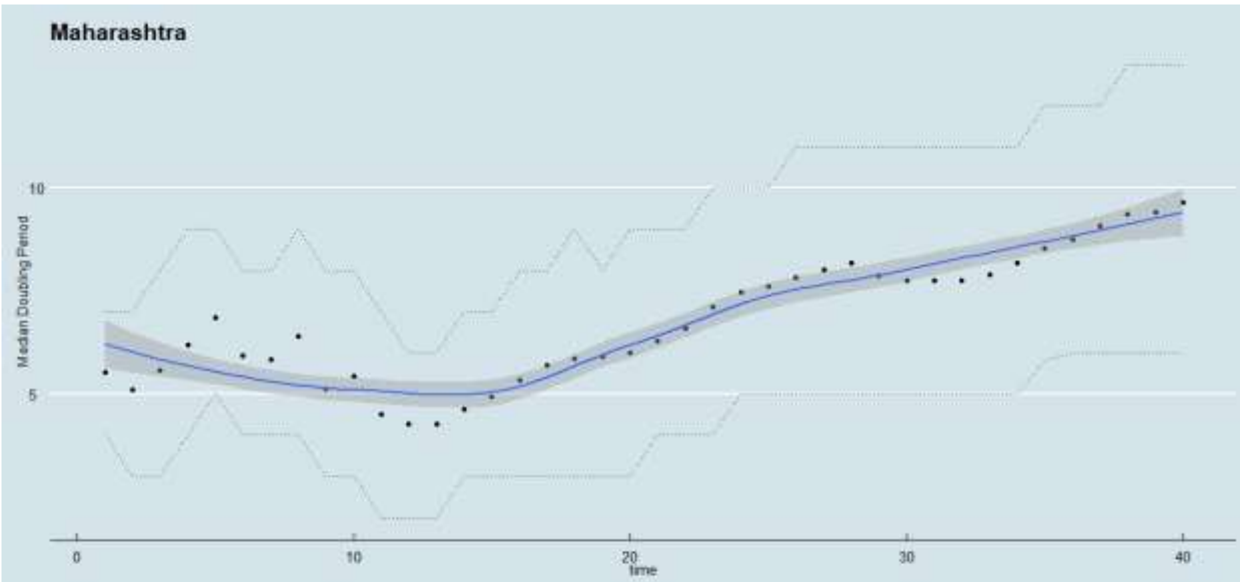
As you can see, more and more new cases are coming, and hence the blue dots are also rising. The black line in panel V is the fitted logistic curve. It describes the data well, and will continue to describe it well as the number of cases move towards the flat portion. But the all-important question that one must answer now is: “is the curve flattening? Has the black line started to grow at a slower and slower pace?” That question of course cannot be settled with a mere visual inspection of the data. Administrators use the rather imprecise term of the “doubling rate”. The “doubling rate “ is just the time required for total cases to go from X to 2X. Even though it is used very regularly, we argue that it can be rather uncertain. The important thing is to realise that the observed number of cases at any point of time, or the observed sequence of cases over any period of time , is one of several sequences that might have appeared. The number of cases on any given day is an outcome of a probabilistic process, and the observed number is just one of the many alternative values that could have appeared on that day. This property has the potential of making the observed doubling period an inaccurate measure of the actual doubling period. This is best illustrated by the case of the state of Maharashtra. The observed time series of the cases is one of many that could have appeared. Suppose we had the ability to observe several of these alternative time series (which could have appeared, but did not, by pure chance). We can now achieve this by a method referred to as maximum entropy bootstrapping.This tec hnique is used to simulate new data from existing data in a way that simulated data is useful for finding values of parameters . Basically, it works something like this: Suppose I am interested in guessing the rainfall in Mumbai on 7th July this year. Suppose I have past data on the rainfall on 7th July of the previous years, but only for 10 years. I have 10 numbers, and it would not be a great idea to estimate the rainfall on 7th July just on the basis of just those 10 observations. I need more data. I write the 10 values I already have on 10 chits of paper, put them in a shoe box, and mix them thoroughly. Then , I keep on drawing one chit at random, jotting down the number written on it, and puting it back, draw another one . Like this, we can draw a large sample. This is called bootstrapping, since we are literally lifting ourselves with our own bootstraps. When we are dealing with data involving time, we have to be especially careful because there is a certain order to the data (like 2nd February coming before 3rd February and so on). Prof. Hrishikesh Vinod of Fordham University has developed an algorithm that allows us to do this, and which can be implemented in an open source programming platform called R[3]. This gives us several (in our case 1000) replications of the same short time series.

So, we take all the data from the state of Maharashtra, right from the time it recorded the first case of Corona onn March 9 to May 5th 2020. Maharashtra is a very important state for Inndian as a whole, because as of 5th May, Maharashtra accounts for a little over 30% of all the cases in India. We first take the obseervations 1 to 15, and bootstrap 1000 replications of this series. For each of these series, we compute the doubling rate. This gives us 1000 estimates of the doubling rate for the period 9th March 24th March. We then calculate the median value of these 1000 different doublig periods, and also calculate the interval such that 80% of these values lie between this interval. This interval is very important. The wider this interval, the more uncertain our estimates. We then repeat the whole exercise for t=2 to 16, then t=3 to 17, and so on, till we

[3] Vinod H.D and J. Lopez-de-Lacalle (2009): “Maximum Entropy Bootstrap for Time Series: The meboot R Package”, Journal of Statistical Software, January 2009, vol 29, issue 5, pages 2 .19.

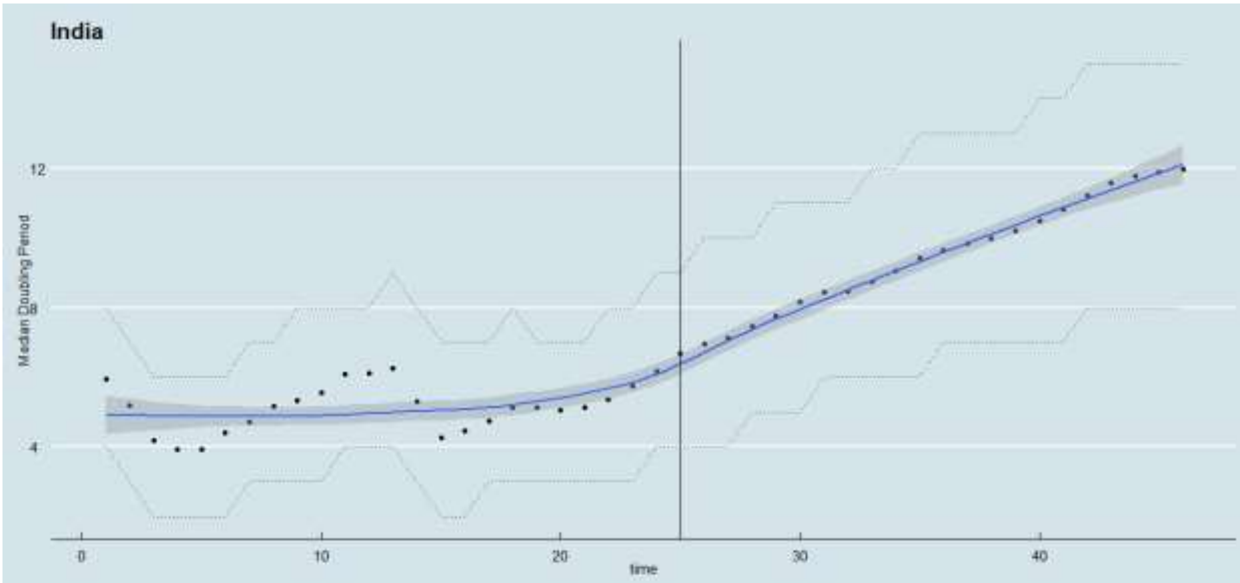
reach the end of the data. This gives us an estimate of the doubling period for each day, and also the intervals for each day. The following figure shows this.

Figure 4:



As you can see, the average doubling period has gone up, but the estimates are highly uncertain. The dotted lines are the intervals, within which the actual doubling rate will lie 80 times out of 100. So, on 5th May, though the estimated doubling period is around 9, the actual could be lying in the interval 6 to 14. This is not at all helpful. We get the same story when we look at India as a whole. The median doubling period seems to have gone up, but in reality, this is highly un certain, as the dotted lines show. Figure 5 shows the progress of the doubling rate for India.

Figure 5



You see the same feature here. Though the doubling rate has surely gone up, its estimate seems highly uncertain, which limits its value as a policy tool. The “doubling period” is regularly used as a metric by the Government of India to classify districts into designated areas such as “red zone” ,”orange zone “ and “green zone”[4]. Classifying a district as a red zone leads to serious limitations on the activities that people in these districts can do. The costs of wrongly classifying districts into various zones can be high. The doubling rate is also of limited relevance in evaluating policy impact. The black line in the graph above marks the beginning of the nationwide lockdown starting from 25th March 2020. The doubling period seems to have gone up, but that conclusion can always be challenged given the wide confidence intervals. The median doubling period has gone up from a little over 4 to about 12 days. But the observed doubling period could lie anywhere between the dotted lines. That makes inference based on doubling period problematic. of the Given something as serious as fighting Corona virus, we cannot have a measure that is as capricious as that. We need to do better. The next section outlines a methodology to do so and the subsequent section applies that to India and some major countries.

[4] Government of India, D.O.No.Z.28015/19/2020-EMR/30-4-2020 as an example

The discussion is bound to be a little technical, but most people should be able to follow at least the basic drift.

Section 1

Most people find it difficult to wrap their heads around the idea of exponential growth. In simple terms, exponential growth is when new numbers are proportional to the base. If your salary grows by 10% every year, and if you had a salary of Rs.10,00,00 in any given year, you would get an annual increment of Rs.10000. But that means that your next year's salary would be Rs. 1,10,000 . So, your increment next year would be the Rs. 10000 that you would get as a 10% increment on your original salary, plus the Rs. 1000 that you would get on the additional Rs.10,000 that was added to your salary last year. That is, you would get Rs.11,000 additionally this year, as against the Rs. 10,000 that you got this year. So, with the compound rate of growth, the additional money you got this year would be higher than the additional money you got last year. This is the basis of the exponential growth model. Even if the growth rate remains constant, exponential growth means a larger addition today compared to the addition yesterday. So, a good measure of whether the curve is in the exponential phase or the flattening phase is to check if new cases today were less than the new cases yesterday. If new cases today are actually more than the new cases yesterday, the curve is becoming steeper, being in the exponential phase. If the new cases today are less than the new cases yesterday, the curve is becoming flatter. When the new cases are just the same as the old cases, we have a point of inflection.

This is all good, but we can never know what is happening because the actual data are very noisy. That is today's cases can be more or less than yesterday's cases , not because the curve is becoming steeper or flatter, but by pure chance. That is a feature of the Corona data that most people find difficult to appreciate at first glance. If Maharashtra sees 200 additional cases by today evening, it does not mean that these 200 cases were inevitable, from the perspective of somebody thinking about it in the morning. Such a person, if she is reasonable, will think as follows " May be there is a 20% chance of observing cases just under 25, another 30% chance of observing cases between 25 and 120, and a 50 % percent chance of observing more than 120 cases". The number 200 is one of many that could have occurred that day with positive probability. Just by observing the number of new cases today and comparing them with those yesterday does not tell us anything about the underlying mechanism that is generating these cases. We need a way of separating what is happening by pure chance (the noise) and what is happening because the curve is actually changing its shape (the signal). So, let us proceed to do that.

Let new(t) be the number of new cases on date t, and let new(t-1) be the number of new cases onn the previous day. Let the ratio (new(t)/new(t-1)) be called r. Now,if r>1, we have exponential growth (new cases keep onn coming faster and faster every day), where as if r<1, we have a flattening of the curve (new cases keep on coming at a slower rate). This is our key insight.

This means that since (new(t)/new(t-1)) = r, we should have new(t)=r*new(t-1). This would mean that if we knew r, and new the number of new cases today, we should be able to predict with absolute certainty the number of new cases tomorrow. But that will never be true in the real world, which is full of uncertainty. The number of new cases tomorrow will be greater or less than what we predicted today, by pure chance. For example, some laboratories might declatre their data late, some people got might get infected accidently and so on. We will call that chance factor as e(t). Econometricians (people like me who make their living from modelling data but are not IT folks, but more likely to be found in Universities, are called econometricians. Also, as a general rule, they are never as rich as the IT folks!). So, we have , new(t)=r*new(t-1)+e(t). This is the critical idea. If r>1, this is called an explosive autoregresive process. That is, if r were 2, and we had 2 cases today, we would expect 4 tomorrow (plus or minus a randomchance term), 8 day after tomorrow and so on. On the other hand, if r <1, say it was 0.5, then if we had 10 cases today, we would expect 5 tomorrow, 2.5 day after tomorrow and son, again, admitting that there will also be a chance factor causing some plus or minus. If r <1, the process is called a stationary autoregressive process. This is a very important process in time series econometrics. So, if we want to know if the curve is flattening, we must find r. But that is difficult. There is no one single r to find from the data. So, we cannot use the whole data set from the beginning of the epidemic to date because r has in fact not been constant, but been changing. If we have historical data for say two months, our S shape implies that r has first been low, then high, then again low towards the end. It is like trying to guess how an old person looks today by putting together his photos when he was a baby, a teenager, a middle aged person and on his retiremen day. The whole historical data therefore cannot tell us whether currently the curve is flattening or not flattening. What do we then do?

My assumption from here on is that we are interested only in the most recent past. We are interested in knowing if r<1 or r>1 , right now. So, we decide to use only say the last fortnight's data for finding r. But there is a problem in that. Since we have only fifteen days data, and that data will keep changing from day to day , our estimate of r will be far from certain. If we start fifteen days from today , going backwards, we might get an estimate of r. But if we include tomorrow's data and delete the last data point in the previous series, we might get a rather different estimate r. Not a good idea. So, what do we do? In the next section, we describe what we found when we applied this the time series on new carona cases in each of the 21 Indian states for the last fortnight's daily data, and also to several other countries. It might not be too incorrect to refer to the calculated r as the virulence rate.

Section 3:

In this section, we have tried to find the value of r for all major states. We omitted the north eastern states because Corona infections are currently not significant in those states. The table below gives the findings on r for each state. For this calculation, we have used data for all days from the time the state recorded its first infection to 2nd of May 2020.

Table 1

State	State	Estimated Value of r (mean of all the r values from the 1000 replications for each state)	2.5% quantile	97.5% quantile	Standard Deviation of the Estimate
1	Andhra Pradesh	0.75031	0.575591	0.863942	0.075428

2	Assam	0.654937	0.389074	0.8438	0.120181
3	Bihar	0.195176	0.044655	0.314518	0.067058
4	Delhi	0.385887	0.23747	0.544252	0.082017
5	Gujarat	0.92188	0.745797	0.974915	0.058142
6	Haryana	0.37815	0.268594	0.461137	0.048661
7	Himachal Pradesh	0.012969	-0.09649	0.122613	0.057551
8	Jammu and Kashmir	0.621882	0.513615	0.712066	0.050065
9	Jharkhand	0.234218	0.103541	0.414851	0.080985
10	Karnataka	0.744123	0.613576	0.854414	0.063832
11	Kerala	0.702803	0.557412	0.805156	0.061093
12	Madhya Pradesh	0.716725	0.513416	0.819149	0.073373
13	Maharashtra	0.840605	0.786342	0.886354	0.026413
14	Odisha	0.128694	0.034653	0.304879	0.071454
15	Punjab	0.418767	0.343681	0.49517	0.039483
16	Rajasthan	0.76723	0.642356	0.856651	0.053581
17	Tamil Nadu	0.672031	0.492826	0.797903	0.076439
18	Telangana	0.612001	0.475375	0.747996	0.071825
19	Uttar Pradesh	0.62718	0.490829	0.712504	0.058315
20	Uttarakhand	0.570973	0.43469	0.672214	0.059618
21	West Bengal	0.865087	0.771545	0.929027	0.042105

We note several things from the table. First, the estimated r values for each of the Indian states are less than 1. This is excellent news. This is good evidence that in all the major states, the Carona curve has indeed flattened. We are approaching the flat part of the curve, though at different speeds. We also note happily that the standard deviations of our estimates are quite small, meaning our estimates are fairly reliable. That is another piece of good news. We will keep on replicating this exercise every day to keep track of the bending of the curve in all states. We note that Gujarat, Maharashtra, West Bengal and Rajasthan all have a relatively high r, meaning that the curve has bent, but not as much as inn some other states. It is interesting that Telangana is showing an r of 0.61. If you examine the data on new cases carefully, you will see that the new cases are actually turning up at the very end of the graph, in panel IV of figure 2. The graph in panel IV of figure 2 is actually constructed from a 7 day exponential moving average, so it represents a kind of a trend. So, is it reversing back in Telangana? Need to watch out. This incidence in fact illustrates the value of our method, over and above the usual “doubling period “calculation. On the other hand, if you look at Himachal Pradesh, the estimated r is almost zero. That is a good indication that the curve has plateaued out in reality. The case of Telangana shows how tricky simple visual inspection of the data can be.

The important point that is germane to the discussion is the 2.5% and 97.5% quantiles for each state , the upper and lower limits which will contain the true r 95 times out of hundred, which have been reported in the table. As we can see, these are far closer than what we found in the case of the doubling rate estimates. This is because we know from basic statistics that the variance of any quantity is proportional to its magnitudes. Doubling periods are all greater than on, and get larger and larger if public policy is successful. That means that the variance of the estimates also increases, making them more and more uncertain. This magnitudes of r are much smaller, giving us tighter confidence intervals.

Section 4

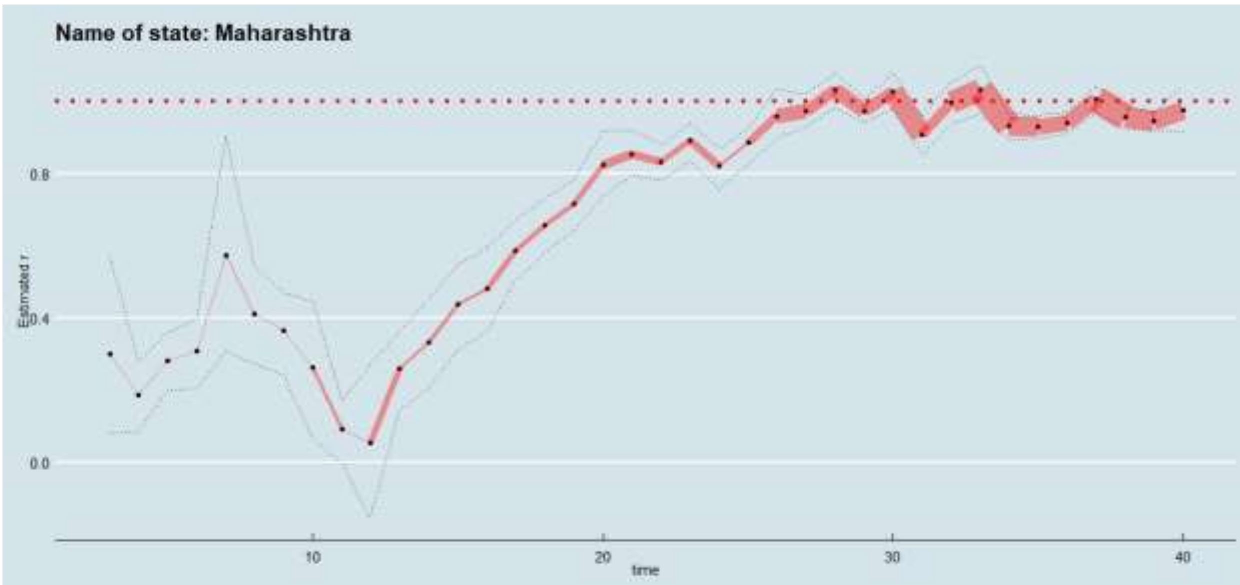
In this section, we chart out the time paths of r in every state, through estimating r by using a 15 day sliding window. So basically, for a particular state, we choose the first 1 to 15 days, bootstrap 1000 replications, generate one thousand estimates of r, and calculate the mean of these one thousand estimates. Then, we repeat the exercise for days 2 to 16, and so on till we come to the end of the series. This exercise gives us a sliding estimate of r. Periods of rising r would mean that the ratio of cases today too cases yesterday was increasing (though not necessarily by a factor greater than 1). We have done this exercise for some major states in India, which is shown inn figure 4.

Whenever r exceeds one for a state, we say that the state has been in an explosive growth phase. r less than one is the flattening out phase. As r rises, but is still less than 1, flattening out is still happening, but at a slower speed.

Of course, we would like to state a caveat here, which is particularly important for this issue. We are using data on reported cases, not the actual infections. Data generating machineries of different states are also vastly different. That surely has some bearing on then findings, but we wonder whether they would alone account for all the differences in the trajectories. The other important issue related to this is our data are not randomly collected samples, but purposive testing of mostly symptomatic people, potentially infected people and their contacts. There is a pronounced sampling bias.

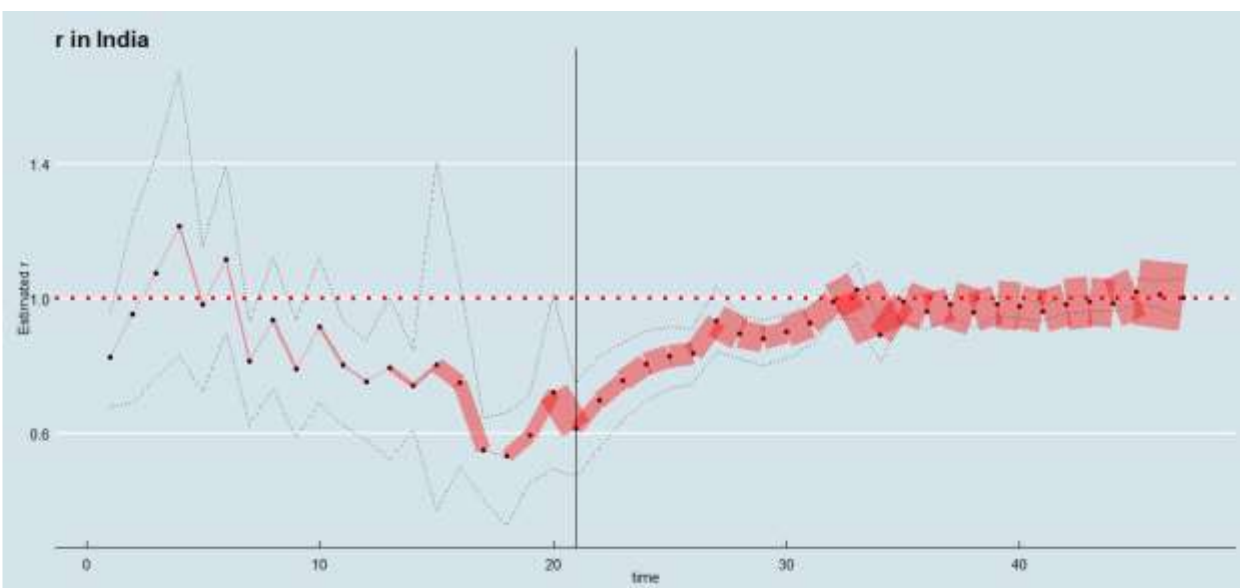
The graphs below describe the situation in each of the states in India. The red curve denotes the progress of the estimated r value over time in each of the states. The width of the curve at a point of time shows the actual size of the new cases. That is, 8 new cases today against 10 new cases yesterday and 800 new cases today against 1000 new cases yesterday would both indicate the same r, 0.8. However, the new cases in the fist case was just 8, while in the second, they were 100 times more. This needs to be distinguished. We do this by making the red curve that much thicker in the latter case, compared to the earlier case. To understand this better, let us take the example of Maharashtra given below. The curve is rising upwards, and is also becoming thicker. It is getting closer to 1, which is the dotted red line. The dotted black lines are the 80% confidence intervals. So, we can see that from around the 25th day of the graph, Maharashtra has been on the point of inflection, of moving from a non-exponential growth to exponential growth. In addition, the number of new cases have been increasing with time, as denoted by the width of the red line. This is a cause for concern, since it says that Maharashtra might get explosive on a large base of cases. What gives us satisfaction is as the number of cases increases, the dotted lines get tighter and tighter around the estimated value, which means that the precision of the estimator increases. This is the major strength of our method over and above the "doubling rate" method.

Figure 6



Let us see how r has progressed for India.

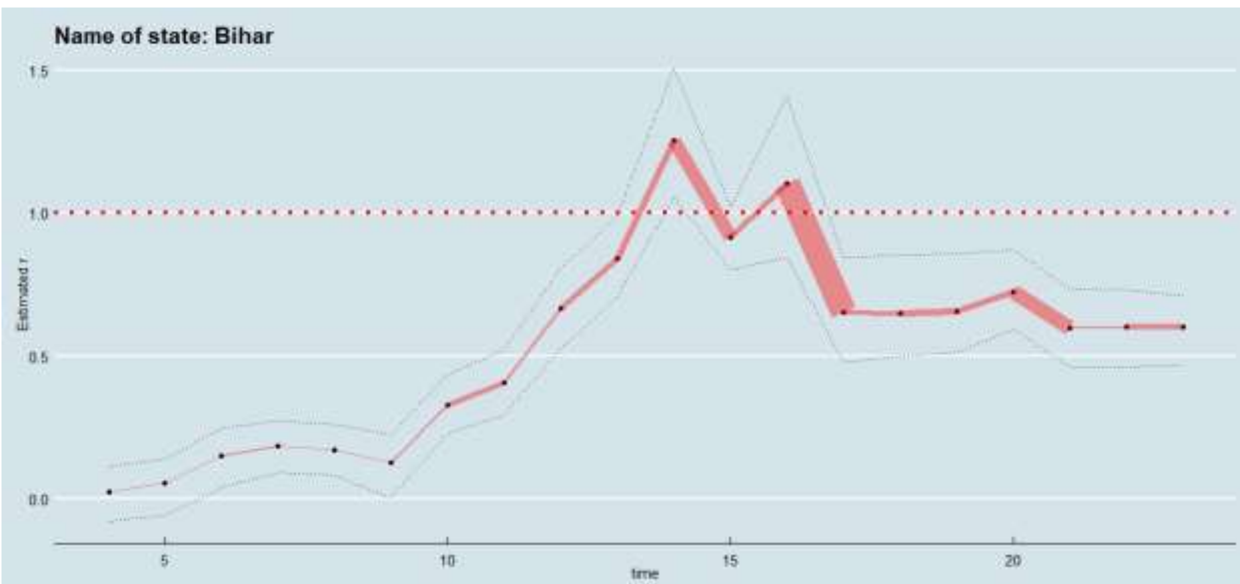
Figure 7

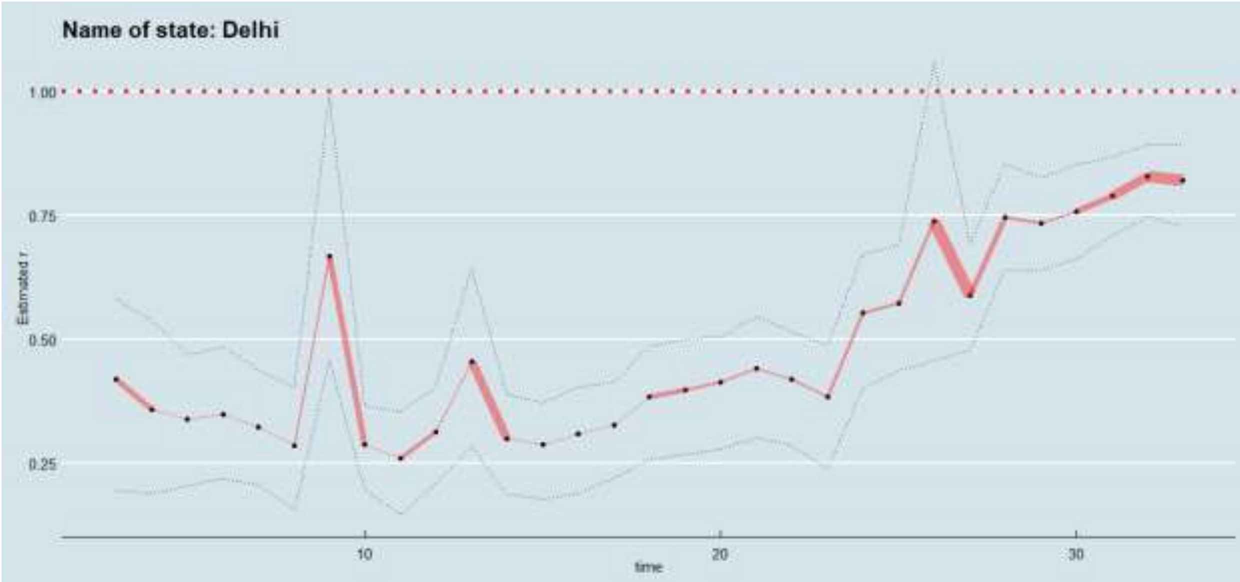


One can see that the r for India is getting closer to the exponential line, with increasing number of cases. The period to right of the black vertical line corresponds to the nationwide lockdown. As one can see, after the lockdown, India has not been in an exponential phase, but from the first week of April, has got very close to the point of switching from sub exponential to exponential growth. The dotted red line is contained between the dotted lines, which signifies India's tryst with the switching point. One can be more certain about this conclusion than that based on doubling rates because of the difference in the width of the dotted lines in the two cases.

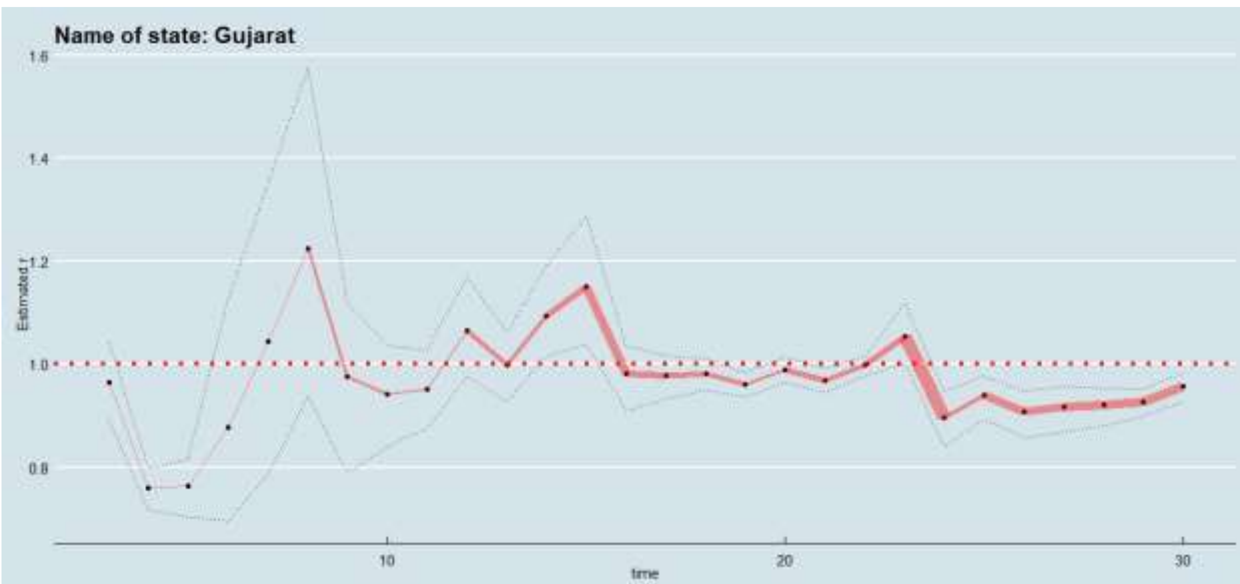
We give the graphs of other states below.

Figure 8

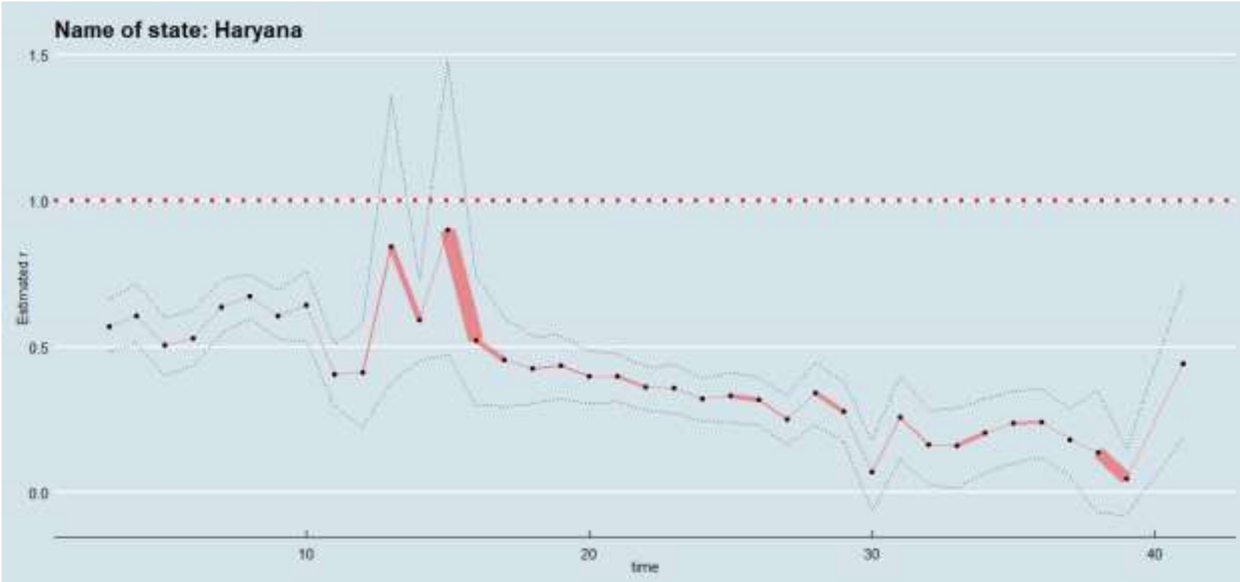


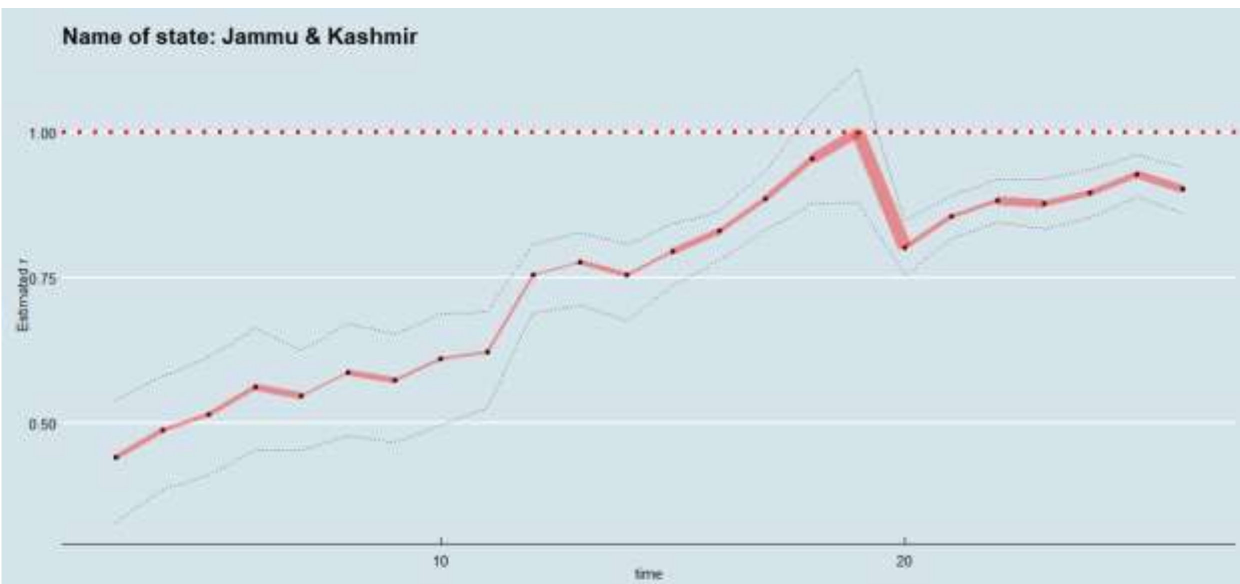


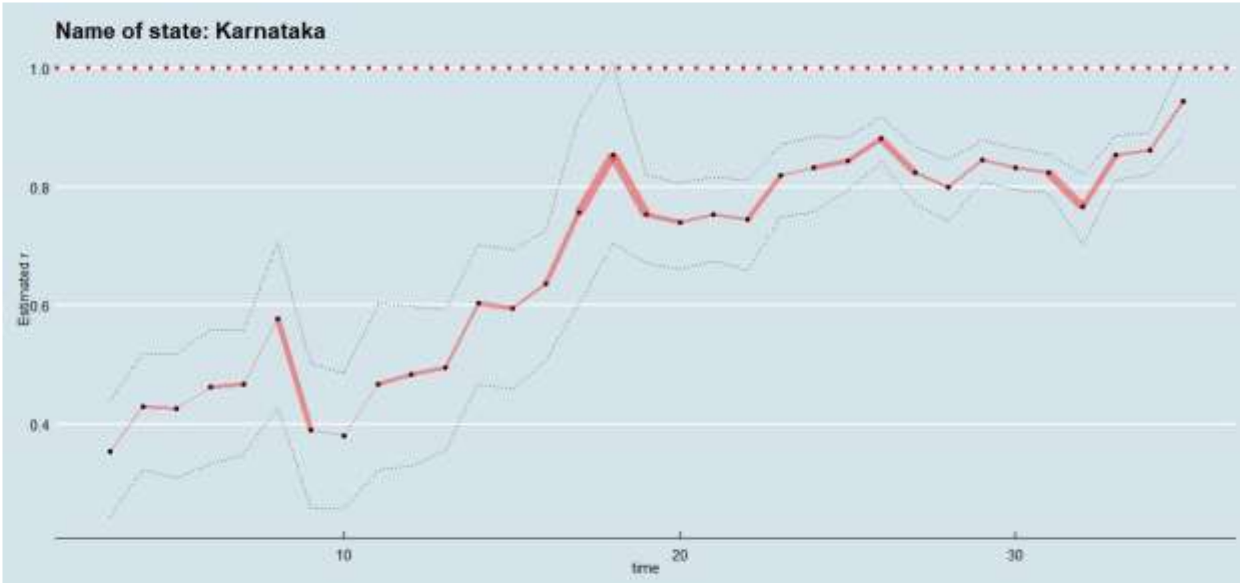
Name of state: Delhi
1.00
0.75
0.50
0.25
Estimated r
10
20
30
time

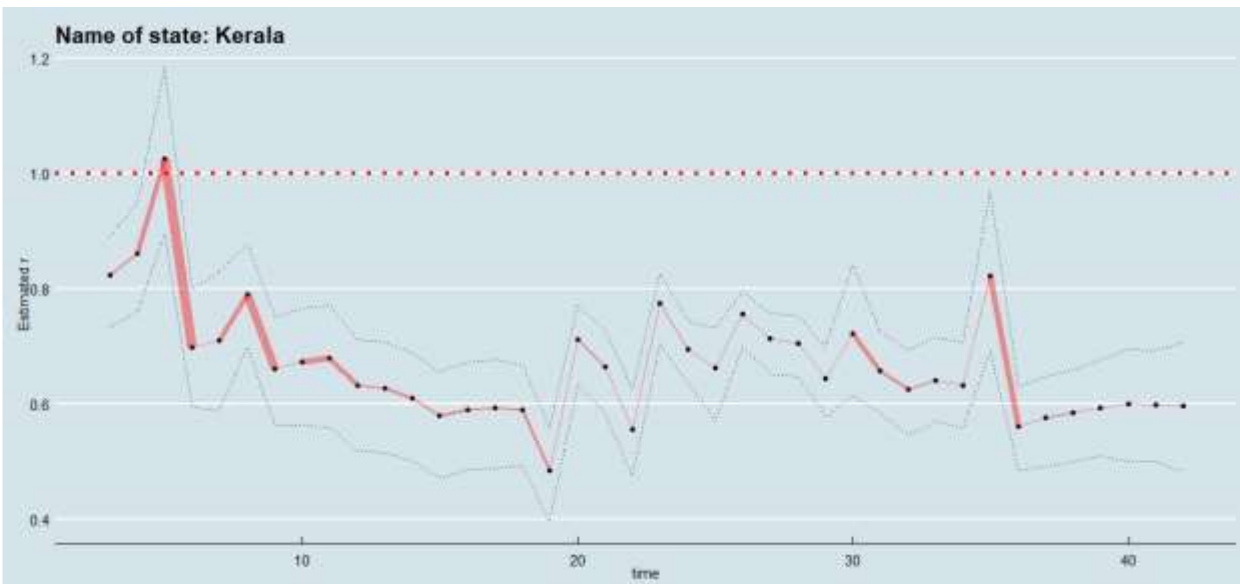


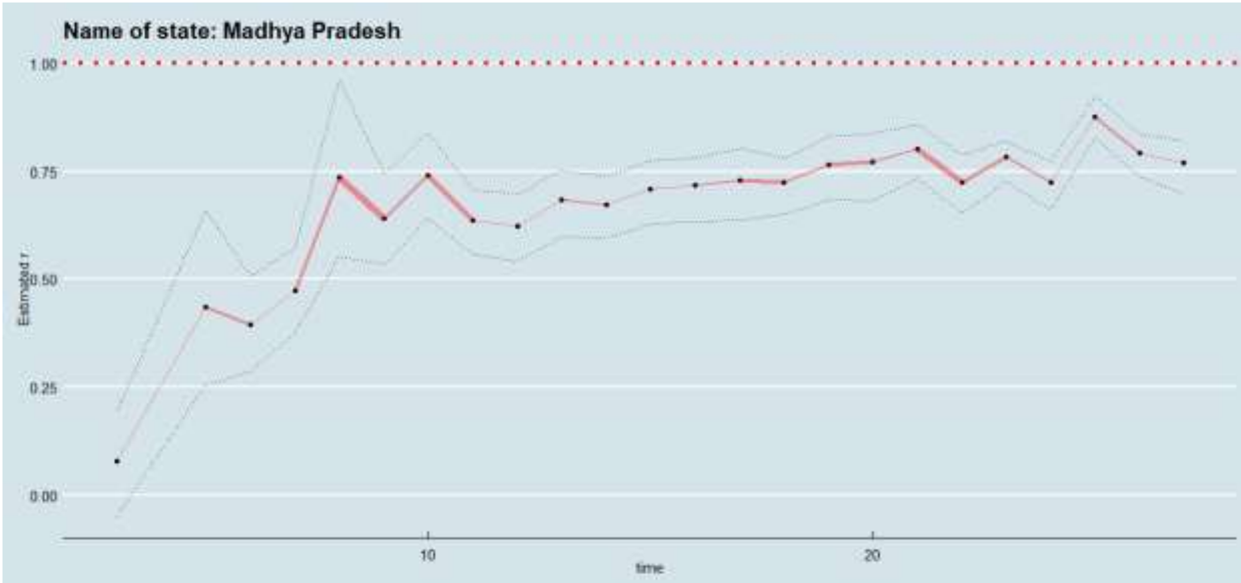
Name of state: Gujarat
1.6
1.4
1.2
1.0
0.8
Estimated r
10
20
30
time

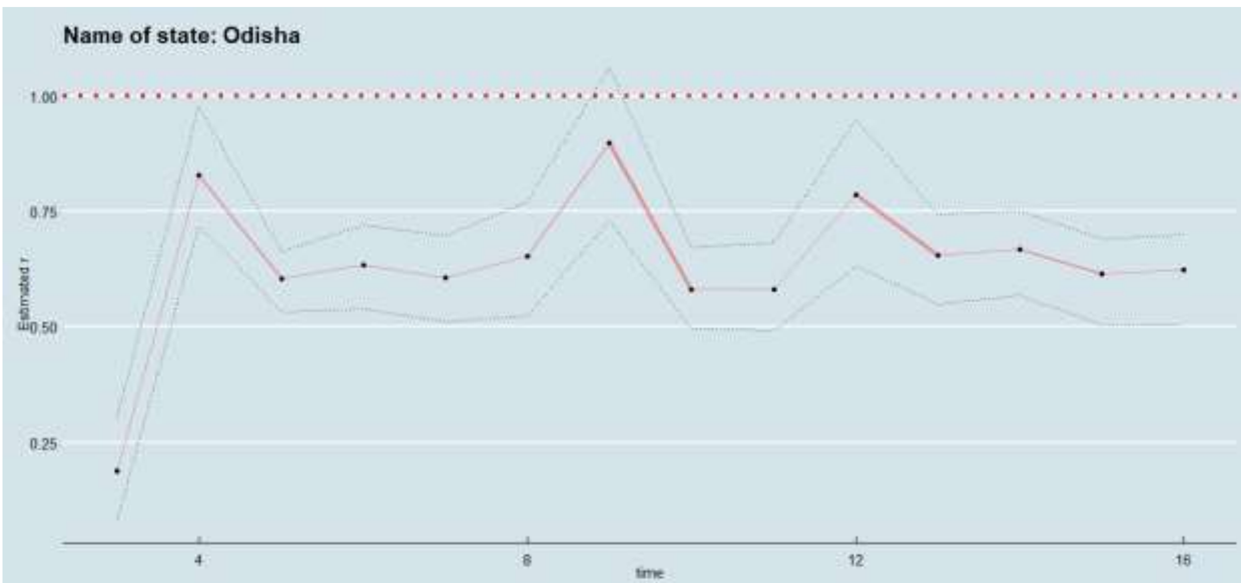

Name of state: Haryana
Estimated r
time

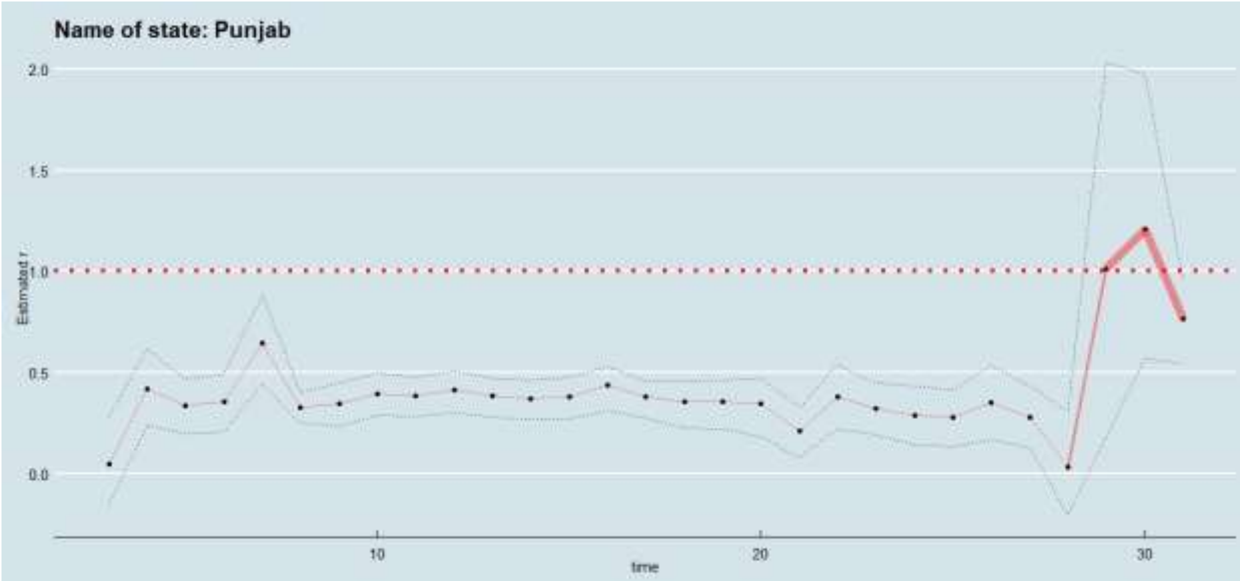

Name of state: Jammu & Kashmir
Estimated r
time

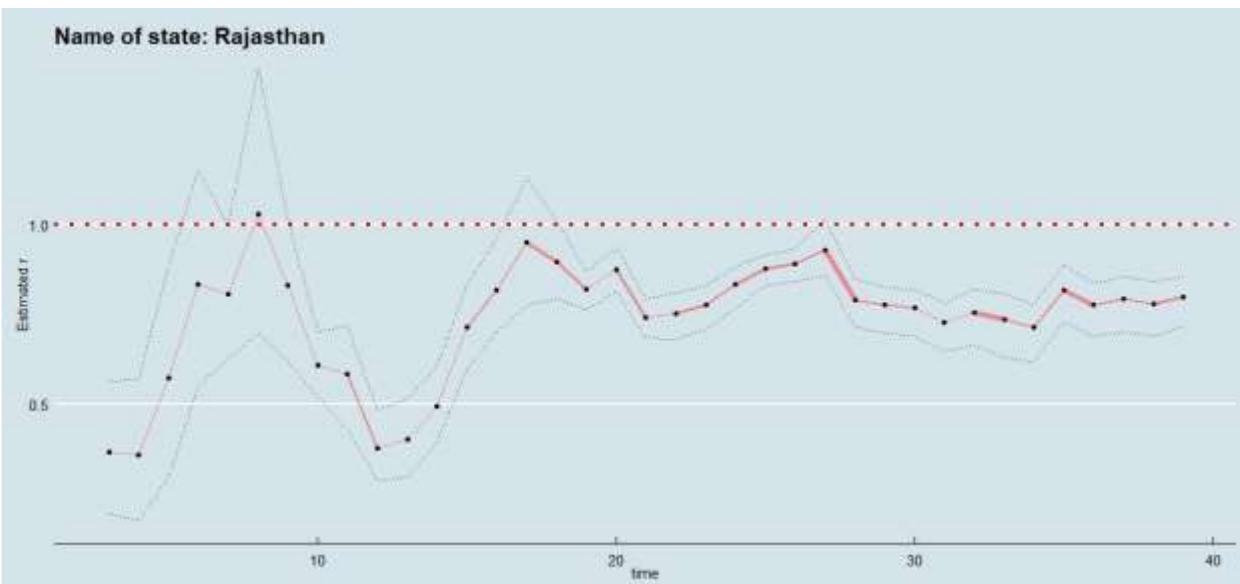

Name of state: Karnataka
1.0
0.8
0.6
0.4
Estimated r
10
20
30
time

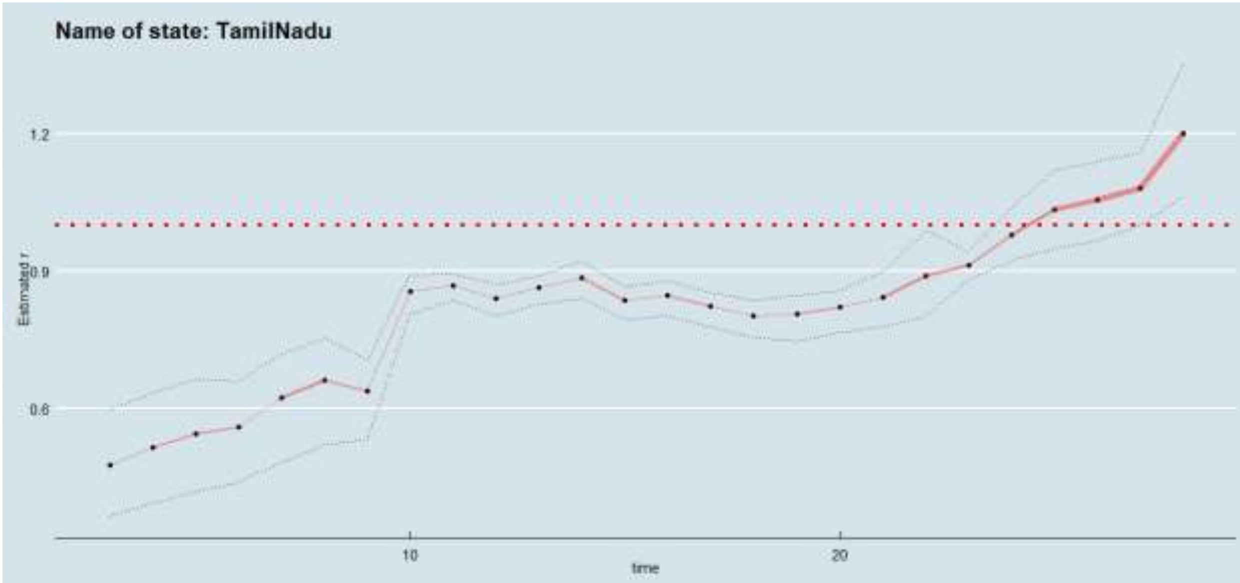

Name of state: Kerala
1.2
1.0
0.8
0.6
0.4
Estimated r
10
20
30
40
time

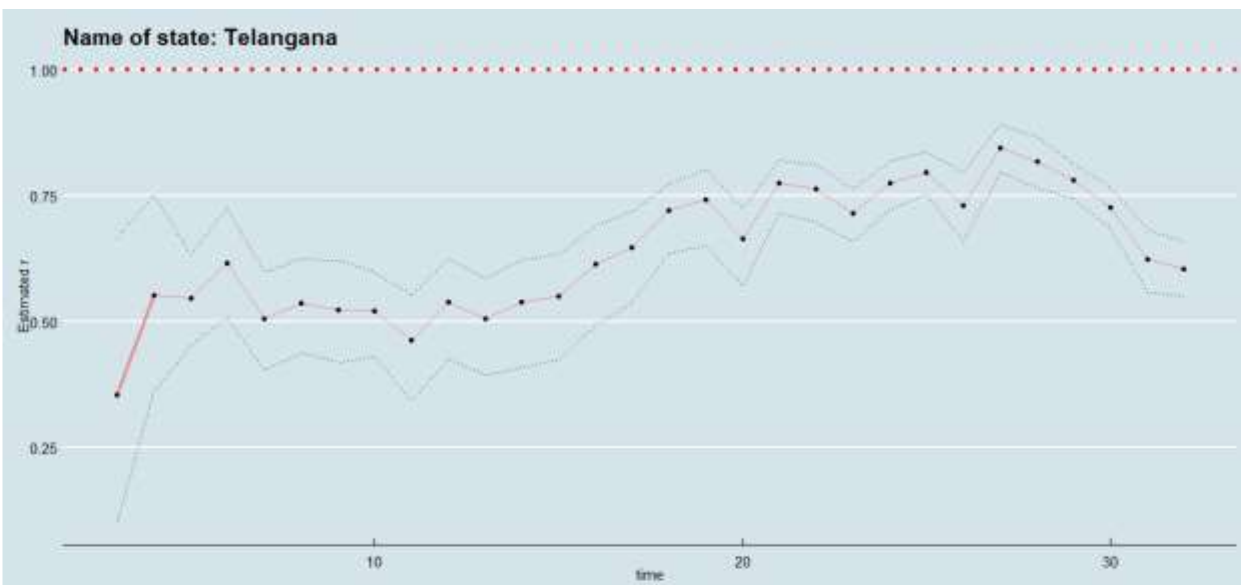

Name of state: Madhya Pradesh
1.00
0.75
0.50
0.25
0.00
Estimated r
10
20
time

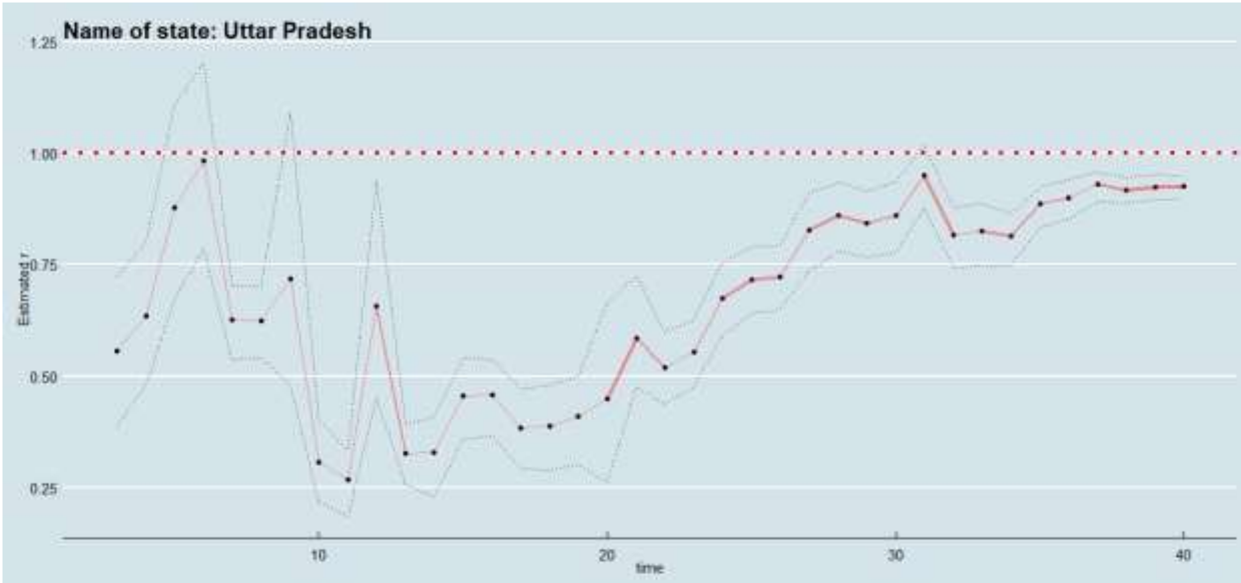

Name of state: Odisha
1.00
0.75
0.50
0.25
Estimated r
4
8
12
16
time

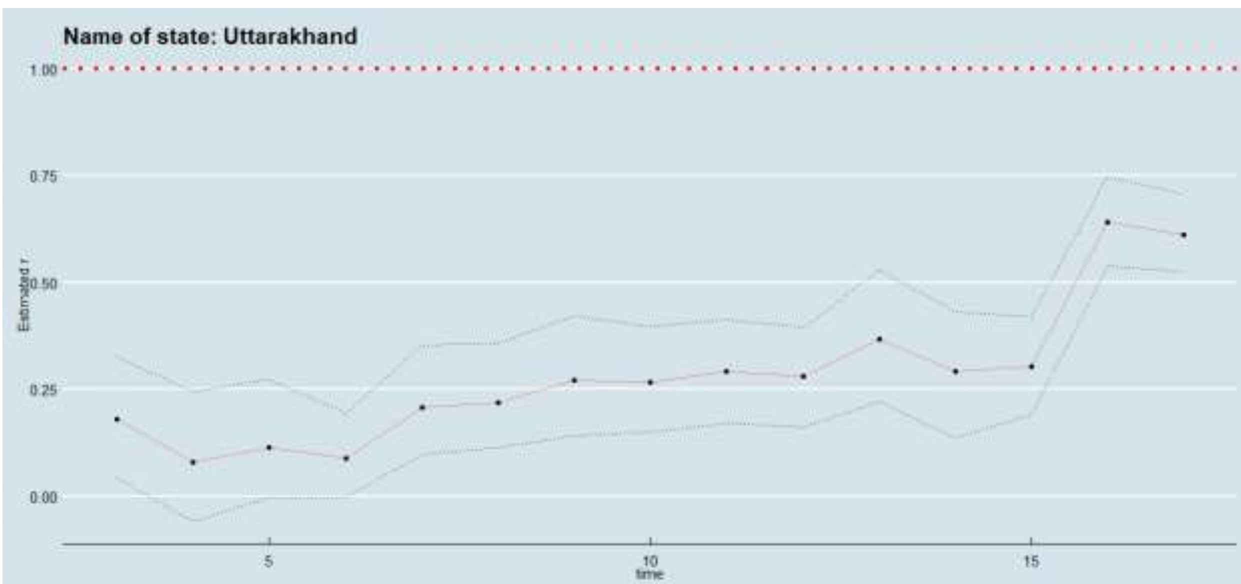

Name of state: Punjab
2.0
1.5
1.0
0.5
0.0
Estimated r
10
20
30
time

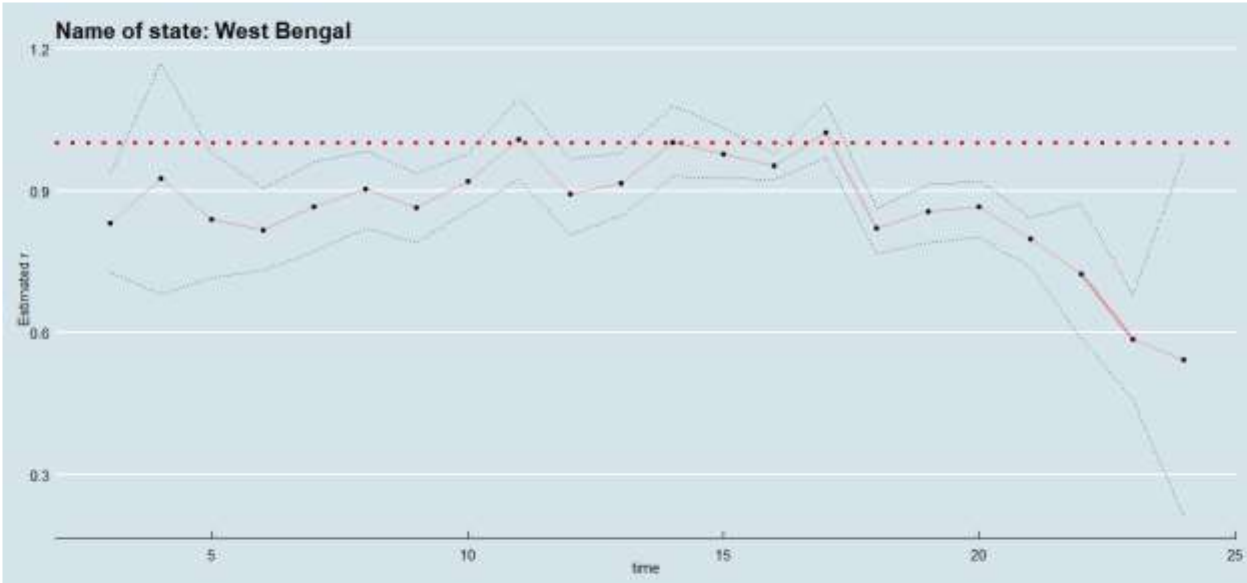

Name of state: Rajasthan
1.0
0.5
Estimated r
10
20
30
40
time


Name of state: TamilNadu
1.2
0.9
0.6
Estimated r
10
20
time


Name of state: Telangana
1.00
0.75
0.50
0.25
Estimated r
10
20
30
time


Name of state: Uttar Pradesh
1.25
1.00
0.75
0.50
0.25
Estimated r
10
20
30
40
time


Name of state: Uttarakhand
1.00
0.75
0.50
0.25
0.00
Estimated r
5
10
15
time



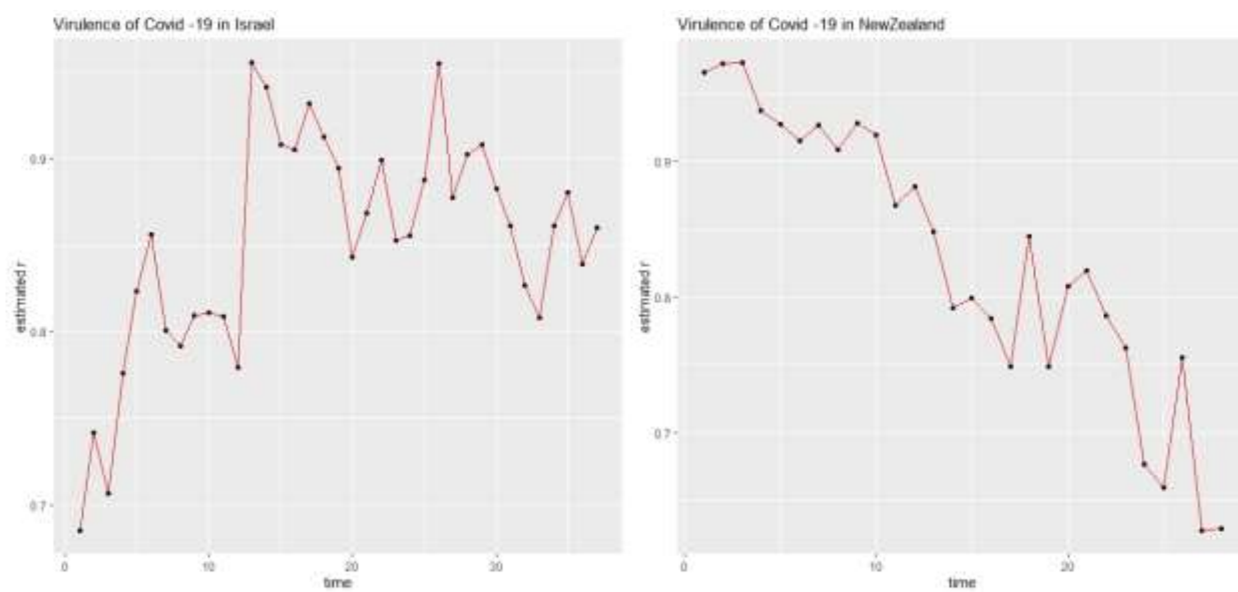
It is easy to read off what is happening in each of the states from the graphs above. Most states have not been in an exponential growth phase so far, but some major states seem on the cusp, with ever increasingly higher new numbers. Tamil Nadu has already made that transition. Haryana has performed rather well, with a declining r, and is approaching “no new cases” phase. Bihar, Delhi, Kerala, Madhya Pradesh, Orissa, Punjab, Rajasthan, Telangana, Uttarakhand and West Bengal are in the subdued growth phase ,as of now. Punjab has seen a large one time increase through 324 infected devotees returning from a Gurudwara in Maharashtra, which accounts for that one big spike.

As is easy to see from these figures, different states have had different trajectories. Maharashtra had a slow and steady move to r equaling one from about the 15th day of registering its first case to the 30th day. After that, r was steady, before starting a slow downward movement. Delhi had the reverse experience. Infection growths shot up in the initial period, and then dropped. Gujarat shot up to a little over one (the genuine exponential phase) and then has come down a notch, subsequently. West Bengal too has had the same trajectory. Madhya Pradesh is somewhat similar to Maharashtra. Bihar was steady till about the 15th day, then went very quickly to the explosive phase, before returning to more sedate flattening out.

Another interesting feature of the story is that even though all states have been under lock down from 25th March, and hence have been under more or less the same nationwide policy, their trajectories have been different. Why is that so? We hazard a guess. Corona infections are strongly localized within most states in urban areas. To illustrate, 25 municipal corporations in Maharashtra account for 95% of the total infections in Maharashtra. Therefore, what urban local bodies do is quite critical, at least as much as what the national policy is. There are bound to be differences in policy responses between municipal corporations, which probably would account for the differences.

What is the international experience? Below we report the same graphs for several countries. Here, we have used data from the date at which each country crossed the 100 cases mark till 4th May 2020. We of course cannot compare across countries because the extent of testing and the methodologies of testing are all widely different across countries. Nevertheless, comparisons across time for individual countries can be illuminating. Here, we have not indicated the size of the infection, like in the case of Indian states, since infections vary widely and we just want to illustrate the value of this tool here.

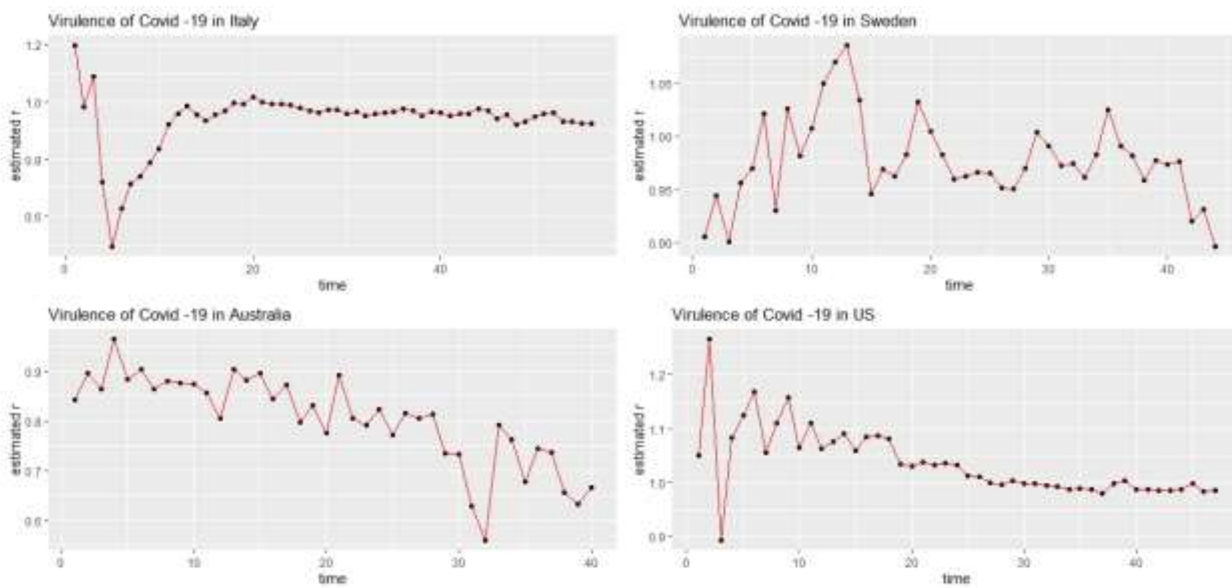
Figure 9



The left hand panel is Israel, while the right hand panel is New Zealand. Both are small countries, with very similar per capita incomes and governments rated as highly effective. Yet, the trajectory of the Corona path has been very different. New Zealand has had the fastest decline in the virulence parameter. Israel had an initial rise in r, which has subsequently been under control. The important thing is that Israel has avoided getting into the exponential phase.

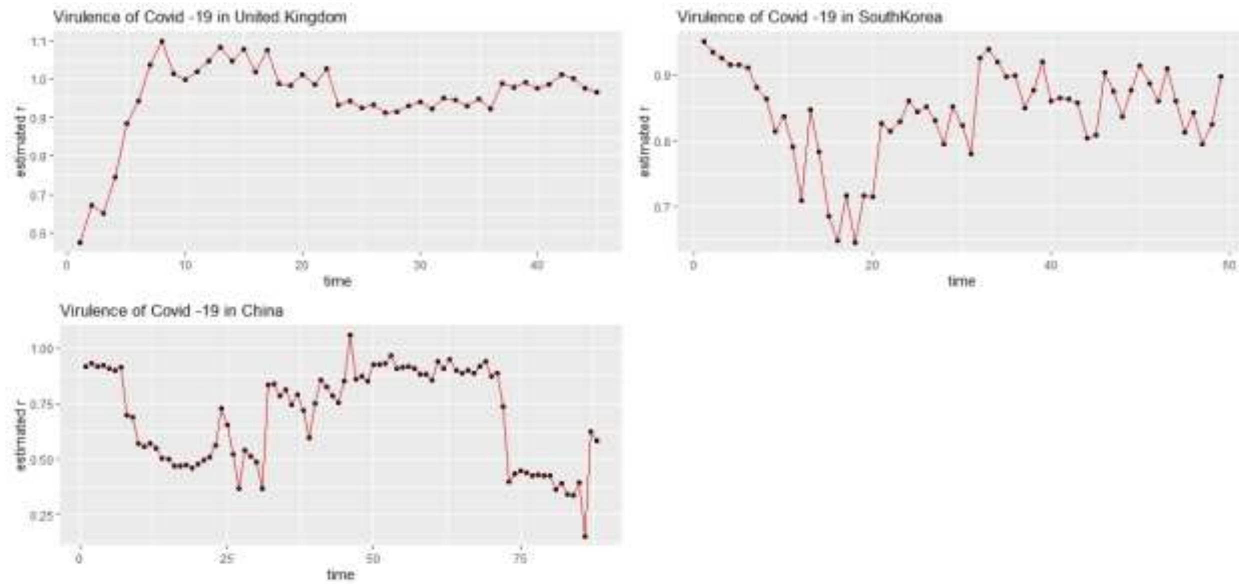
The next figure compares the trajectories of r for four countries that have taken major hits from the spread of the infection.

Figure 10



Italy as well as United States started with an exponential phase, but Italy got things under control much more quickly. In the United States, matters remained in the exponential growth phase for a month! Sweden , we know has had a unique strategy of not closing most establishments and schools. As we can see, the virulence has declined in Sweden, but no as fast a country like New Zealand that is one third poorer than Sweden and five times as populated. Figure 10 looks at China, South Korea and the United Kingdom.

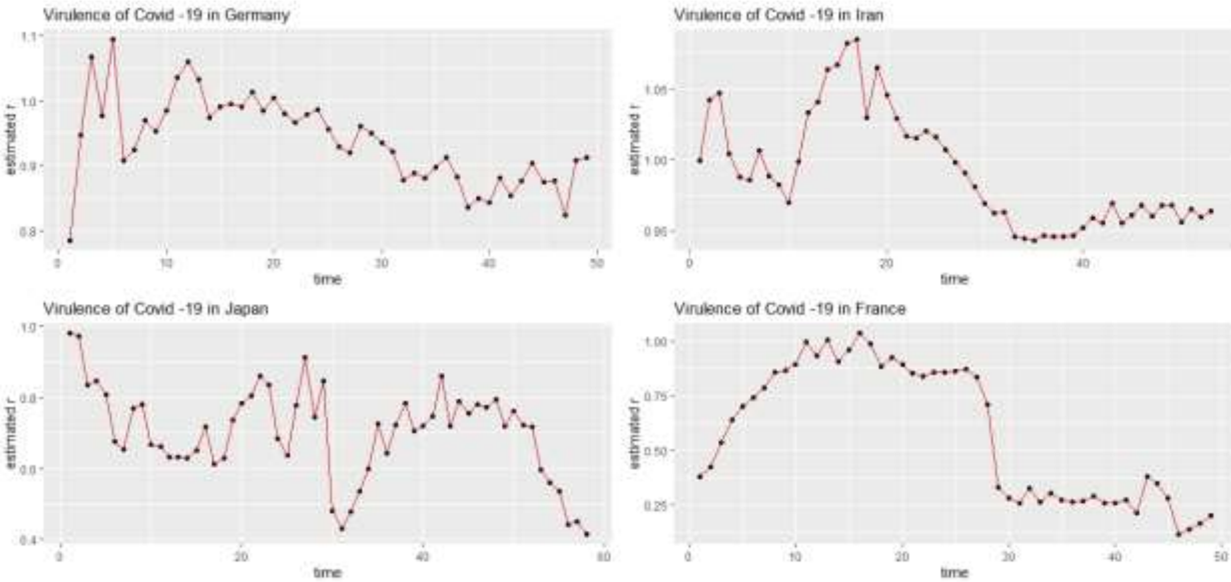
Figure 11



As one can see, China, despite the popular belief, was never in the exponential phase for any substantial amount of time. From about the 65th day, its virulence fell, rapidly, but there was a second wave in some cities at a later stage. South Korea is still continuing to experience an r fluctuating around 0.85, which is higher than some Indian states. South Korea has done testing on a massive scale, but its virulence is not going down. The battle is still on for South Korea.

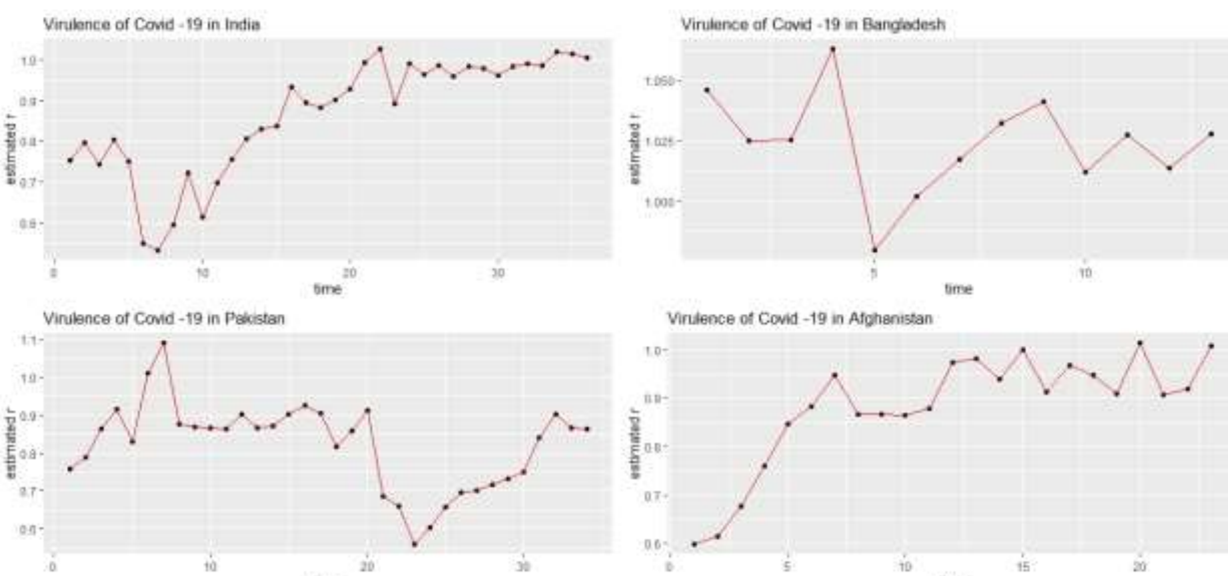
Figure 8 presents what is happening in Iran and some European countries.

Figure 12



France and Japan have won dramatic victories by driving down the virulence rate. France was a particular success , driving down the virulence in a period of little over two weeks. Germany went from an exponential growth stage to not an out of danger position of 0.9. The same can be said of Iran. Figure 12 shows what is happening in South Asia.

Figure 12



As one can see, India is slowly reaching the exponential phase, though many of its states have not. India's trajectory seems similar to Afghanistan. Bangladesh has been in the exponential phase for about two weeks now. Pakistan, after some initial gains, is now in the dangerous ninety's . So, all the South Asian economies need to watch out.

Conclusion:

In this paper, we have developed a new method to say in real time, and much less ambiguously, whether the Corona curve is indeed flattening. This is better than the rough and ready, simpler to calculate methods like the doubling period, which, as a cost of their simplicity, can be seriously misleading. Using this method on Indian states, we can say that the curve has flattened and is flattening in all major Indian states. However, whether it continues to do that will have to be watched on a daily basis. Also, what is critical to watch for the progress of the infection in India is the responses that are having at the urban municipal corporation level, as much as national level policy. The national level policies get the big press, but what is equally critical is perhaps what the municipal corporations of Indore, Agra, Mumbai are doing.

As far as the world is concerned, several major nations mostly are experiencing a value of r close to one, mostly not statistically significantly different from r. The growth may not be exponential, but is not declining either. Another set of states, New Zealand, France, Japan and China have been able to drive their virulence southwards. South Asia seems to be an emerging region for rapid corona infections in the future.

The analysis above shows the value of r in tracking the progress of Corona infections. It is more accurate and also more insightful compared to the usual "doubling period rule". r can be one of the metrics for classifying states and districts or any other regions . For instance, for districts where $r < 0.33$, despite extensive testing for at least a week, could be classified as "green zones". On the other hand, districts with an r between 0.33 and 0.67 could be classified as "orange districts", while those with $r > 0.67$ could be classified as red zones. Those with $r > 1$, where infections are likely to rise exponentially, can be classified as "critical zones". Currently, this all important distinction, between sub exponential and exponential growth, is not a part of the policy process. The progress of r could then be watched carefully, to reclassify districts if necessary. The strength here is that hypothesis tests like null of $r<0.33$ versus alternative $r>0.33$, or whichever boundary we may decide, are easily possible. At the first pass, states can be divided into regions using r as one of the criteria. At the second stage, districts within states can be classified into zones. Finally, of course, any single measure will have to be used in conjunction with other measures inn a multifaceted manner, with special attention to the quality of the data capturing system at the regional level.