On the early spreading rate of COVID-19 in India

Preprint · April 2020

DOI: 10.13140/RG.2.2.23746.12489

4 authors, including:

Rahul Chanduka
Indian Institute of Technology Bombay
1 PUBLICATION **0** CITATIONS

SEE PROFILE

Sahasrajit Sarmasarkar
Indian Institute of Technology Bombay
2 PUBLICATIONS **0** CITATIONS

SEE PROFILE

Dharmaiah Manjunath
Indian Institute of Technology Bombay
193 PUBLICATIONS **1,573** CITATIONS

SEE PROFILE

On the early spreading rate of COVID-19 in India

Himanshu Aswani, Rahul Chanduka,
Sahasrajit Sarmasarkar, D. Manjunath
Deptt of Electrical Engg., IIT Bombay
himanshuaswani123,rahulchanduka98,sahasrajit1998@gmail.com,
dmanju@ee.iitb.ac.in

April 23, 2020

Abstract

We analyse the spreading rate of the COVID-19 epidemic in India during the early days of the epidemic and compare it with that of other countries. We use three different variants of the SIR model and fit the data to the model parameters. These models account for a large number of infections not resulting in symptoms and, even from among those developing the symptoms, a non-trivial fraction of them not being tested. Based on the analysis, we have reason to believe that during these early days, the spreading rate is possibly lower in India than in other countries.

Keywords: Covid-19; spread rate estimation, Covid-10 spread in India.

1 Introduction

In this paper, we use publicly available data about the confirmed cases in India[1][2] and in other countries[3] and compare the spreading rate in India and these other countries during the early days of the epidemic. A key novelty is that we assume that only a fraction of those that develop symptoms are tested for the infection. We assume that the tests are accurate in that there are no false positives or false negatives. Our analysis

[1] https://www.covid19india.org
[2] https://covidindia.org
[3] https://coronavirus.jhu.edu/map.html

is based on a variation of the standard SIR model[4] that is widely used in modelling the spread of epidemics in a community.

We consider three variants of the SIR model, which we believe to be appropriate to the situation in a majority of the countries, including India. Further, a necessary assumption that we make is that the data is incomplete and underestimated. In the models that we describe below, we account for the fact that only those who have developed symptoms, possibly severe symptoms, are tested and those that are asymptomatic and/or mildly symptomatic are not tested. Further, we explicitly account for the fact that even among those that develop symptoms of influenza-like illnesses, possibly only a fraction have been tested.

Our key findings are as follows.

- There is reason to believe that the spreading rate in India during the early days is lower than in many other countries.
- The preceding conclusion appears to be robust to the low testing rates.
- Comparing the all-India spreading rate with that in three hot spot states indicates that the actual rate may be lower than that obtained using all India.
- A possible explanation for the latter is that in the early days, many of the infections probably came from outside India and may have caused the jump in the all-India numbers giving it a higher rate.

A more detailed analysis and discussion is available in Section 4. The following caveats should accompany these findings.

- The demographics of the infected population and those that were susceptible to be infected by them is possibly significantly different from the national demographics. The latter will become more important in the event of community spreading.
- We have assumed that the dynamics of the epidemic in India and in other communities are similar save for a scaling factor. The validity of this assumption is not clear.

[4]D. J. Daley and J. Gani, "Epidemic Modeling: An Introduction, *Cambridge University Press,* Cambridge (2005)

- The infection rates by the various severity levels are not well understood and we have used a gross simplification in our model.

We next describe the model details and the accompanying assumptions.

2 Model descriptions

Recall the standard SIR model: N is the size of the population, μ is the average number of contacts by each member of the population, p is the probability of a contact causing an infection in a susceptible member of the population, $1/\gamma$ is the average rate of recovering from the infection of each infected person. Further, at time t, $U(t)$ is the number of uninfected in the population, $I(t)$ is the number of infected, and $R(t)$ is the number that have recovered from the infection. The differential equations that model the evolution of the epidemic are as follows.

$$\begin{aligned}
\frac{dU(t)}{dt} &= -p\frac{\mu}{N}U(t) \\
\frac{dI(t)}{dt} &= p\frac{\mu}{N}U(t) - \gamma I(t) \\
\frac{dR(t)}{dt} &= \gamma I(t) \\
N &= U(t) + I(t) + R(t)
\end{aligned}$$

During the early days of the COVID-19 epidemic in all communities, the fraction of population affected by the epidemic is small compared to the total population of all the communities that have been affected. We therefore focus our attention to modelling only the growth of $I(t)$. We further assume that at time t, the infected population, $I(t)$, is made up of the following types[5].

- Asymptomatic are those who have been exposed to the infection but will never develop the symptoms. The number of such people at time t, is denoted by $A(t)$. It has been reported that the asymptomatic infect at a significantly lower rate because their viral shedding is far fewer than those that have the symptoms or

[5]Luca Ferretti, *et al,* "Quantifying dynamics of SARS-CoV-2 transmission suggests that epidemic control and avoidance is feasible through instantaneous digital contact tracing, `https://science.sciencemag.org/content/early/2020/04/09/science.abb6936.full`

will eventually develop the symptoms. This means that the p is lower for this part of the infected population.

- Presymptomatic are those who have been exposed to the infection and will eventually develop the symptoms. The number of such people at time t, is denoted by $P(t)$. They will eventually develop symptoms but before that, they have a significantly higher viral shedding than the asymptomatic. A testing protocol that only tests those with symptoms, as was practised in India [6] will not be able to identify them. Hence, they will be mixing among the population and spreading the infection. The p for this population is significantly higher than that of the asymptomatic population.

- Symptomatic are those who have been exposed to the infection and are showing symptoms of infection. The number of such people at time t, is denoted by $S(t)$. They will be tested, and since the tests are assumed accurate, they will be found positive. Thus their infectivity is reduced due to isolation. Although they are highly infectious, the μ is reduced for them through isolation.

We thus have

$$I(t) = A(t) + P(t) + S(t)$$

Based on published data, we assume that the asymptomatic will recover, on average, in twenty days from the date of infection and will stop infecting after that time. We further assume that the presymptomatic will develop symptoms, on average, after five days, i.e., they will be among the population and spreading the infection for an average of five days before they start showing symptoms and are probably tested. Due to limited testing resources, the focus on testing has been on the infected and symptomatic population; hence we assume that only a fraction β of the symptomatic population has been tested and isolated either through a home quarantine method or through hospitalisation. It also simply follows that $P(t)$ also contributes to the rate at which $S(t)$ is increasing. We define $r = p\mu$ and call it the infectiousness parameter; this is the rate at which an infected member of the population that is not isolated infects the uninfected population.

We further let α denote the fraction of the affected population that eventually show symptoms, i.e., go into the $P(t)$ class and $(1 - \alpha)$ denote those that go into $A(t)$ class. Following equations are then quite obvious regarding the temporal evolution of the

[6] https://indianexpress.com/article/explained/coronavirus-testing-india-explained-6325375/

classes. In the following, we will assume that the infectiousness of the asymptomatic is half that of the symptomatic and the presymptomatic who have the same infectiousness. We note that some estimates indicate that it could actually be about 10% [7].

$$\begin{aligned}
\frac{dA(t)}{dt} &= (1-\alpha)\ r\ \left(\frac{A(t)}{2} + P(t) + (1-\beta)\ S(t)\right) - \frac{A(t)}{20} \\
\frac{dP(t)}{dt} &= \alpha\ r\ \left(\frac{A(t)}{2} + P(t) + (1-\beta)\ S(t)\right) - \frac{P(t)}{5} \\
\frac{dS(t)}{dt} &= \frac{P(t)}{5} - \frac{S(t)}{20}
\end{aligned}$$

This gives us the following set of discrete time approximations that we will use in the fitting of the data. Our unit of time is a day.

$$\begin{aligned}
A(t) &= (1-\alpha)\ r\ \left(\frac{A(t-1)}{2} + P(t-1) + (1-\beta)\ S(t-1)\right) - \frac{19}{20}\ A(t-1) \\
P(t) &= \alpha\ r\ \left(\frac{A(t-1)}{2} + P(t-1) + (1-\beta)\ S(t-1)\right) - \frac{4}{5}P(t-1) \\
S(t) &= P(t-1) + \frac{19}{20}\ S(t-1)
\end{aligned}$$

The preceding is the basic model that is directly derived from the standard SIR model. In the next two models, we introduce more granularity in the system state by accounting for delayed spreading and transitions among classes as well as different rates of infecting susceptible population by the asymptomatic and the presymptomatic.

In the following refinements, we will only use the discrete time model.

Model 2

In this model, we have incorporated the following changes. Instead of assuming that there is a steady rate of removal of each of the three types at the average rate, we assume that the time to removal is exactly equal to the average time spent in the corresponding state. Thus we assume that the asymptomatic recovers in exactly twenty days and the presymptomatic will show symptoms in exactly five days.

[7] See Ferretti, *et al.*

Define the current infectiousness of the population as

$$I'(t) = r\left(\frac{A(t)}{2} + P(t) + (1-\beta)S(t)\right).$$

Now recall that β is the fraction of symptomatic people who were tested positive and hence quarantined. Thus we write the equation for the presymptomatic population as follows.

$$P(t+1) - P(t) = (\alpha I'(t) - \alpha I'(t-5))$$

The first term in the above expression denotes the new infections on day t that become presymptomatic whereas the subtracted term removes those people from presymptomatic population who were infected five days ago and have now become symptomatic.

Similarly we write the equation for the symptomatic population as follows

$$S(t+1) - S(t) = (\alpha I'(t-5) - \alpha I'(t-20))$$

Since the new symptomatic people added on day t are those people who were infected 5 days ago and became presymptomatic, the first term follows. The second term follows from the fact that people who were infected twenty days ago recover and hence are removed from the symptomatic population. Also we write the equation for asymptomatic population as follows.

$$A(t+1) - A(t) = ((1-\alpha)I'(t) - (1-\alpha)I'(t-20))$$

This also follows from a very similar argument where the first term denotes the new asymptomatic cases on day t whereas the second term denotes the people who were infected exactly 20 days ago.

All other variables and parameters convey the same meaning as presented in the basic model. Also, as before, we assume the infectiousness of asymptotic people to ba half of that of pre-symptomatics and asymptomatics. Thus, after some algebraic simplification, the evolution equations in this model are as follows.

$$\begin{aligned} S(t+1) - S(t) \;\; = \;\; \alpha\, r\Bigg(& \frac{\big(A(t-5) - A(t-20)\big)}{2} \; + \big(P(t-5) - P(t-20)\big) + \\ & (1-\beta)\,\big(S(t-5) - S(t-20)\big)\Bigg) \end{aligned}$$

$$
\begin{aligned}
P(t+1) - P(t) &= \alpha\, r\Big(\frac{\big(A(t) - A(t-5)\big)}{2} + \big(P(t) - P(t-5)\big) + \\
&\qquad (1-\beta)\, \big(S(t) - S(t-5)\big)\Big) \\
A(t+1) - A(t) &= (1-\alpha)\, r\, \Big(\frac{\big(A(t) - A(t-20)\big)}{2} + \big(P(t) - P(t-20)\big) + \\
&\qquad (1-\beta)\big(S(t) - S(t-20)\big)\Big)
\end{aligned}
$$

Model 3

In this variant, we assume that the asymptomatic and presymptomatic infect at significantly different rates. In other words, the people who will eventually show symptoms are more "infectious" than those who never show symptoms. Specifically, previous results for other virus outbreaks have arrived at an infection ratio of 0.1 which we too shall follow as our baseline ratio. All other variables and parameters convey the same meaning as presented in Model 1.

The equations in this case have correspondingly been modified as follows

$$
\begin{aligned}
P(t+1) - P(t) &= \alpha\, r\left(\frac{A(t)}{10} + P(t) + (1-\beta)\, S(t)\right) - \frac{P(t)}{5} \\
A(t+1) - A(t) &= (1-\alpha)\, r\left(\frac{A(t)}{10} + P(t) + (1-\beta)\, S(t)\right) - \frac{S(t)}{20} \\
S(t+1) - S(t) &= \frac{P(t)}{5} - \frac{S(t)}{20}
\end{aligned}
$$

3 Analysis Method

Recall that we have three parameters that we need to determine—infectiousness r, fraction that are asymptomatic α, and the rate of testing of the symptomatic β. For fixed β, we consider several combinations of (α, r) and fit determine the best fit log-linear plot for the number of infections as a function of time. For a fixed (α, β), the r that has the least relative mean square error between the data and the best fit plot is determined. We plot the $(\alpha - r)$-curve for different values of β for the different countries, for India, and for some hotspots in India.

The following lists a few more details on the data analysis.

- For each community we consider the data for a period of about 30 days from the time the number of infections is about 10. We call this the analysis window. The exact dates used are :
  - Germany: 26 Feb – 30 Mar
  - France: 26 Feb – 22 Mar
  - USA: 3 Mar – 1 Apr
  - India: 3 Mar – 1 Apr
  - Maharashtra: 11 Mar – 1 Apr
  - Delhi: 11 Mar – 1 Apr
  - Kerala: 10 Mar – 1 Apr
- At the start of the window, it is assumed that the initial number of pre-symptomatic people is roughly equal to the symptomatic people.
- We assume that the initial ratio of asymptomatic people to all people carrying the virus to be α. Formally, $\frac{A(0)}{I(0)} = \alpha$
- We assume that those that have been tested positive are quarantined and removed from the system; they are not infectious.
- We assume that the parameters α, β, r are constant during the initial stages of the spread of disease.
- The β parameter for a country has been chosen based on our estimation of the testing strategies there and the reported death rate. At the same time, we have done similar analysis by choosing every $\beta \in \{0.3, 0.5, 0.7, 0.9\}$.

We have run our simulations on four different countries, United States of America, Germany, France and India. The time series data utilised for these countries corresponds to the first month of steady growth in the number of cases of CoViD19. In the following, β corresponds to a measure of the degree of medical efficiency in the country under consideration. We have, therefore, assumed suitable values of β to reflect varying degrees of medical proficiency in these countries. These values are enlisted in the assumptions section.

4 Numerical results and analysis

We have plotted the actual number of reported cases and the estimated number of reported cases for a particular choice of parameters α, and β and the best fitting r for that. These plots illustrate the goodness of the model for different countries. We also show these plots for some regions in India. Finally, corresponding to each of the three models, we also show the α-r plots. From these figures, and additional extensive numerical analysis, we have the following remarks

- It is important to note that this is not a statistical model and that we cannot provide margin of errors. Further there are several strong assumptions and incomplete and still emerging data.
- We have assumed that the testing protocols in all the four countries were similar albeit with a different β. Since we have considered only the early period of the pandemic in each of the countries, this is not a bad approximation.
- Some characteristics of the $\alpha - r$ curves:
  - For a fixed value of α; as β increases r increases as well. This is reasonable because a higher β implies that a larger fraction of symptomatic people are identified and hence isolated. Thus to fit the given data, we need to have a higher spreading rate r than with a lower β.
  - For a fixed value of r, a higher β corresponds to a higher α. This is also reasonable and is argued along the same lines as before—an increase in β implies that a larger fraction of people are identified and removed from the system, and hence requires a higher number of symptomatics to spreading the infection to the values that data indicates.
  - The α-r for a fixed value of β as plotted for Model 3 is steeper than that for Model 1. Recall that the only difference between the models is that Model 3 assumes a lower (10%) ratio of infectiousness of the asymptotic to that of the symptomatic population. Thus, both the models are equivalent when $\alpha = 1$, when there are no asymptomatic people. As α reduces, the fraction of asymptomatic people increases and since their infectiousness is lower in Model 3, we need a higher increase in r to compensate.

- Recall that for India, our analysis window ends before 01 April; The effect of the lockdown is unlikely to have begun before this time. Similarly, since we have considered only the early period of the pandemic in each of the countries, effects of social distancing and lockdowns etc., had probably not yet taken effect.

- For each of the three models, the α-r plot for India appears to be lower than that of other countries. And that of the states is lower than that for the entire country. Thus we have reason to believe that in the early stages, the spreading rate in India is lower. Further, it is reasonable to assume that the μ in India is larger than in other countries. This would imply that the p is smaller than in the other countries. We reiterate that this is driven by the data and not by our medical, demographic epidemiological knowledge; we have none. Further, it is also possible that α is small. Thus there are probably a large number of asymptomatic people that effectively reduce the spreading rate.

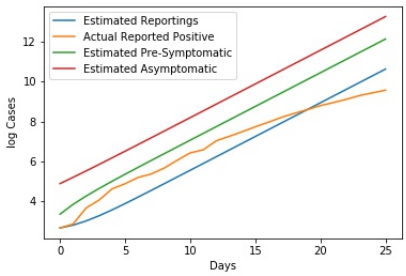
- The last point may also be seen with the following observation. In the early days, the testing protocol was to test those with symptoms similar to a COVID-19 infection and possibly had a travel history or contacts with those that were symptomatically infected. Even with this strong selection, there never has been the case that more than 5% of those tested have been found positive. Thus the β is probably not very small.

- A further observation is that for the three hotspot states chosen, the α-r curve is significantly lower than the all-India plot; i.e., the r for for any value of α is lower. A possible explanation for this is that the model assumes people are infected only because of the spreading of from human to human, but a significant portion of people might have brought the virus in from outside. In the three states, it appears that a large fraction of the cases reported in the analysis window are because of the local spreading and not infections contracted outside the community. Further, the all-India numbers may also have been affected by a few super spreading events infected a large number of people who then took it to different parts. Thus, it may be reasonable to conclude that the actual national spreading rate to be even lower than that indicated by the α-r curves.

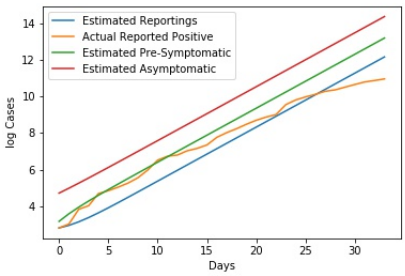
- Kerala appears to have a lower spreading rate compared to Maharashtra and Delhi for all the three models. It would be interesting to study the reason for this.

We conclude by reiterating the caveats that were outlined in Section 1. Further, contin-

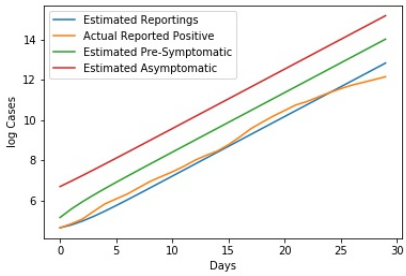
ued analysis of the data to see if the conclusions continue to hold even as the infection spreads in the larger community is essential before using these for any major policy decisions.



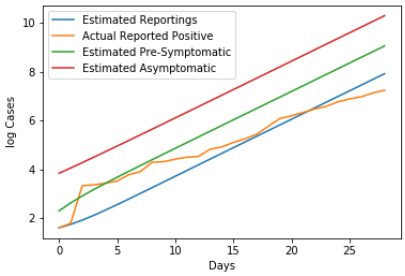
France, $\beta = 0.5$



Germany $\beta = 0.7$

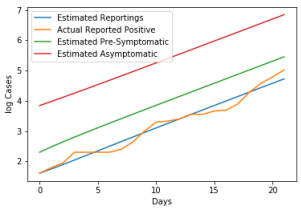


USA, $\beta = 0.6$

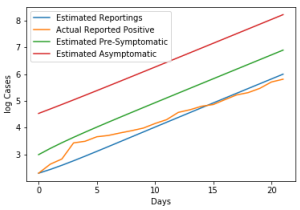


India, $\beta = 0.5$

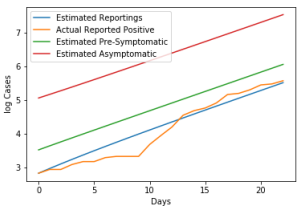
Figure 1: Model 1: Examples of estimates of reported infections, actual reported infections, estimated asymptomatic, and presymptomatic for different countries assuming $\alpha = 0.3$ and β as indicated.



Delhi



Maharashtra



Kerala

Figure 2: Model 1: Examples of estimates of reported infections, actual reported infections, estimated asymptomatic, and presymptomatic for some regions in India assuming $\beta = 0.5$, $\alpha = 0.3$.

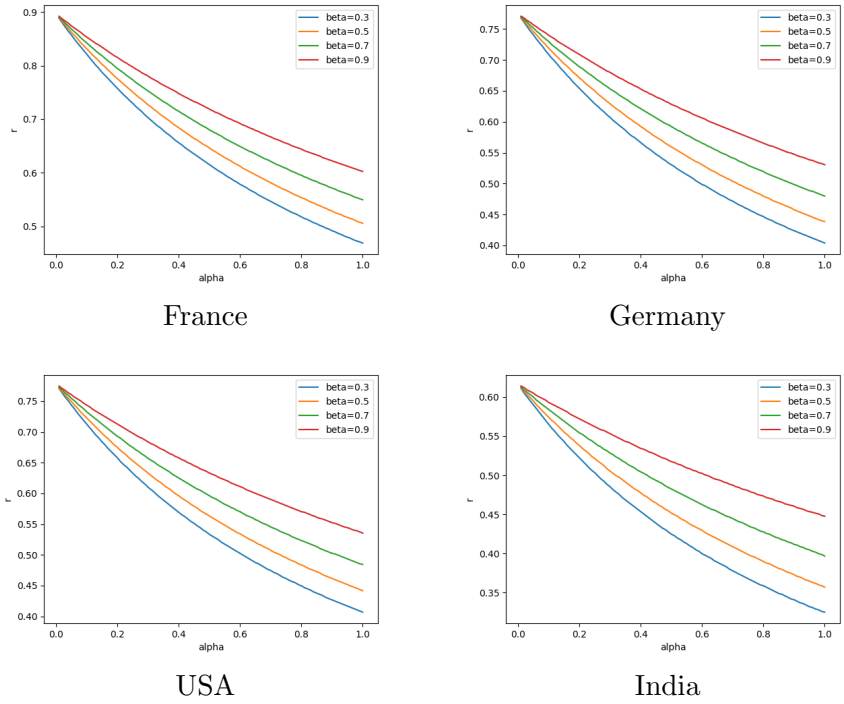


Figure 3: Model 1: α-r curve some countries.

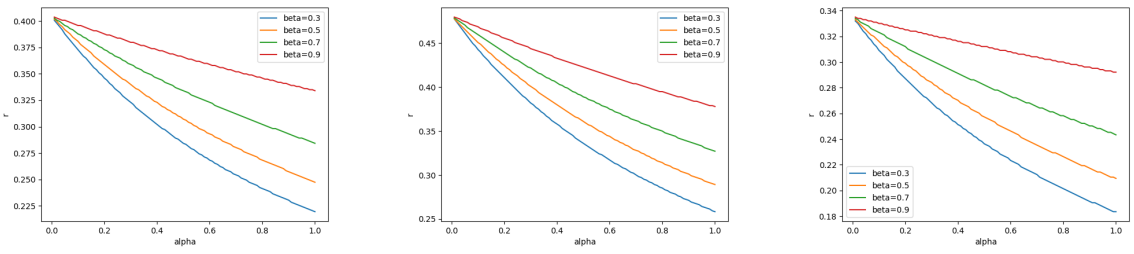
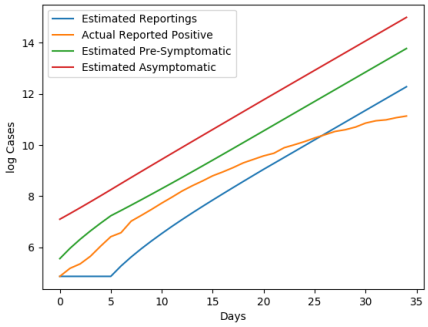
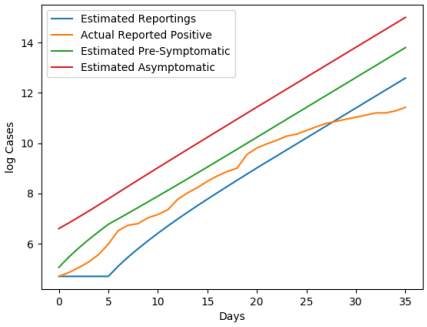


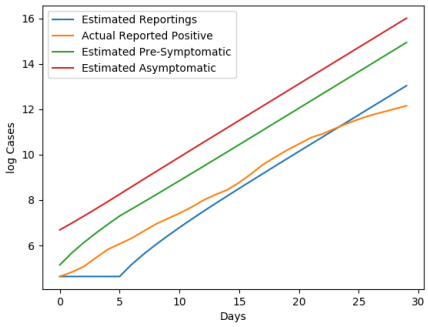
Figure 4: Model 1: α-r curve some regions in India



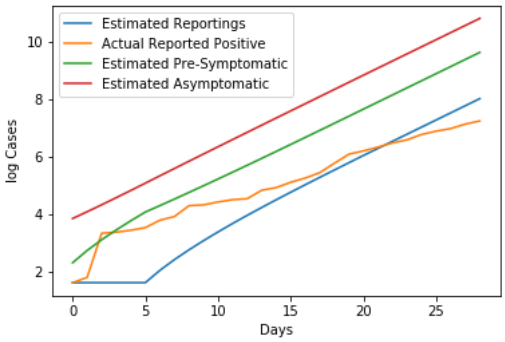
France, $\beta = 0.5$



Germany $\beta = 0.7$

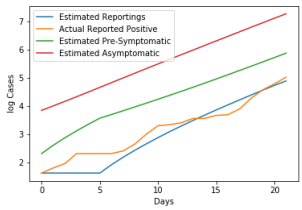


USA, $\beta = 0.6$

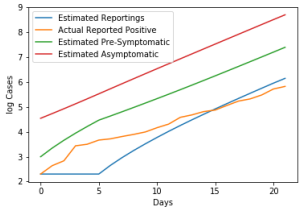


India, $\beta = 0.5$

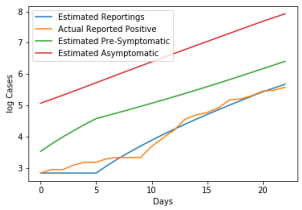
Figure 5: Model 2: Examples of estimates of reported infections, actual reported infections, estimated asymptomatic, and presymptomatic for different countries assuming $\alpha = 0.3$ and β as indicated.



Delhi

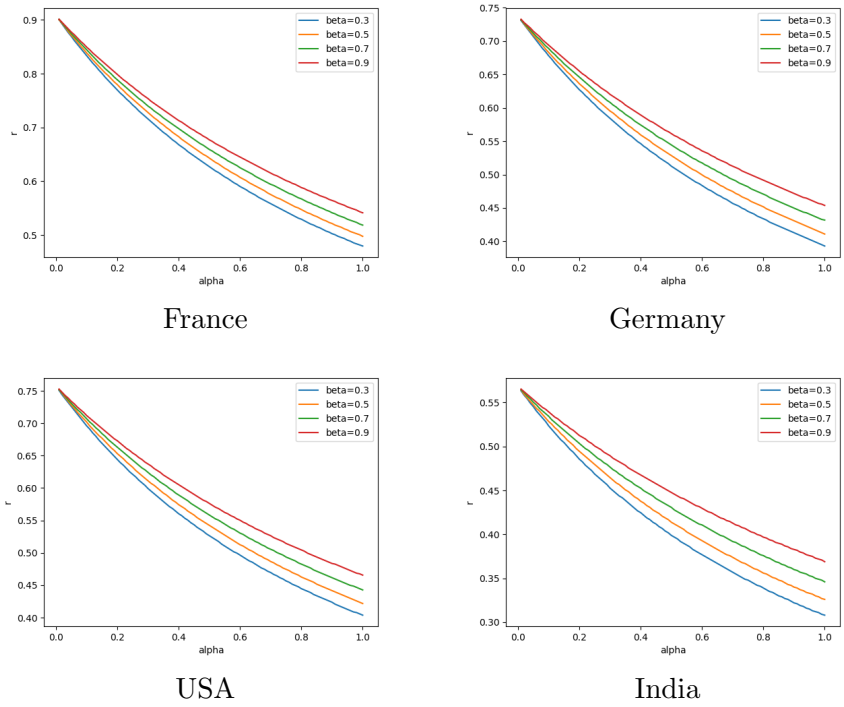


Maharashtra



Kerala

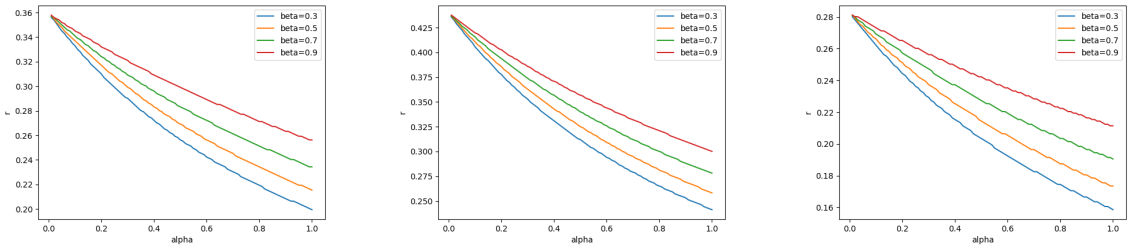
Figure 6: Model 2: Examples of estimates of reported infections, actual reported infections, estimated asymptomatic, and presymptomatic for some regions in India assuming $\beta = 0.5$, $\alpha = 0.3$.



France Germany

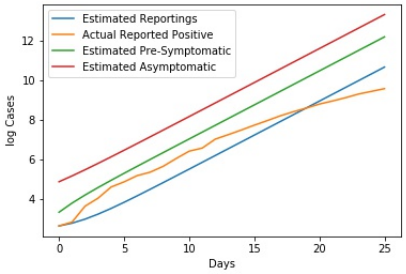
USA India

Figure 7: Model 2: α-r plots for some countries.

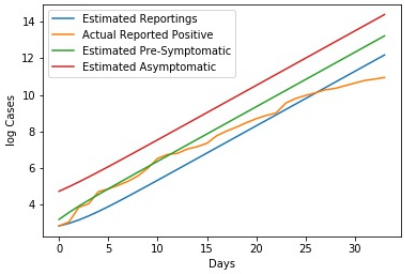


Delhi Maharashtra Kerala

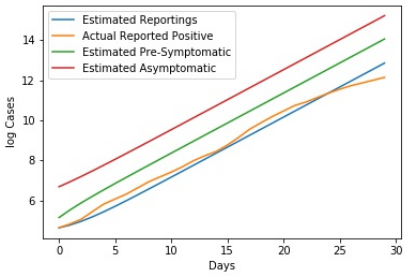
Figure 8: Model 2: α-r curve for some regions in India.



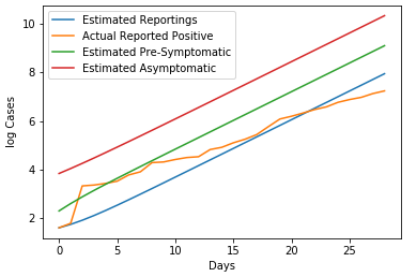
France, $\beta = 0.5$



Germany $\beta = 0.7$

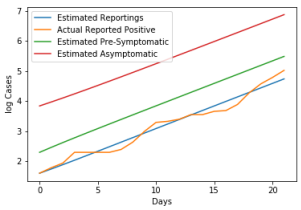


USA, $\beta = 0.6$

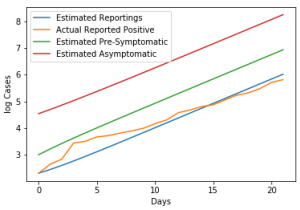


India, $\beta = 0.5$

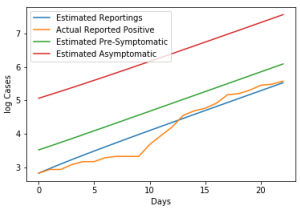
Figure 9: Model 3: Examples of estimates of reported infections, actual reported infections, estimated asymptomatic, and presymptomatic for different countries assuming $\alpha = 0.3$ and β as indicated.



Delhi

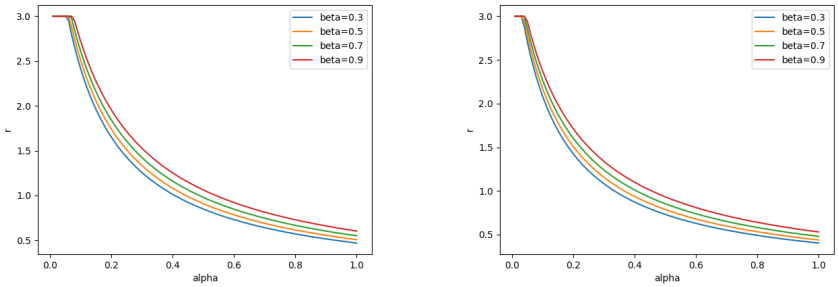


Maharashtra

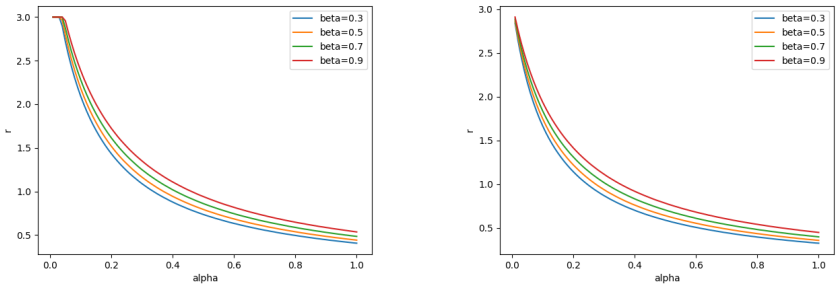


Kerala

Figure 10: Model 3: Examples of estimates of reported infections, actual reported infections, estimated asymptomatic, and presymptomatic for some regions in India assuming $\beta = 0.5$, $\alpha = 0.3$.

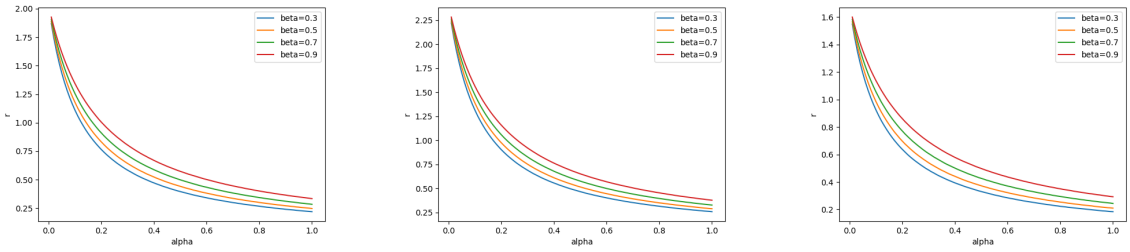


France Germany



USA India

Figure 11: Model 3: α-r plots for some countries.



Delhi Maharashtra Kerala

Figure 12: Model 3: α-r plot for some states in India.