Analysis of a SEIR model with social distancing considerations in the paradigm of COVID-19 outbreak in India

Suryadeepto Nag [*]   Siddhartha P. Chakrabarty [†]

July 28, 2021

Abstract

A novel approach of adapting social distancing consideration into a SEIR model is presented, where susceptible, exposed and unidentified compartments are collated under the umbrella of the social-distanced compartment. Another key characteristic of the model is the inclusion of the nature of social distancing to be contingent on the rate of change of the active cases. The methodology and the results exhibiting an excellent fit to the data (upto 3rd March 2021) are presented, in case of the COVID-19 outbreak in India. The model attributed the apparently extensive social distancing, to the socio-geographical factors, unique to India. Also the data exhibited greater rate of infection from a diagnosed case as compared to undetected infection. Finally, it is demonstrated that a very conservative estimate of undiagnosed cases is at least 75% of the total number of cases.

Keywords: SEIR Model; Social Distancing; Parameter Estimation

1   Introduction

A comprehensive data-driven analysis of the first wave, and modeling of the second (and likely subsequent) waves for Germany and Spain was carried out in [1]. This was accomplished making use of a SEIR model, by dividing the population into susceptible, infected (but not infectious), infected (as well as infectious) and recovered. Out of the model parameters incorporated, the parameters of transmission rate and mortality rate were considered to be time dependent. The model also assumed the complete interaction of the entire population, the impact of natural birth and death being negligible for the time window considered, and little to no possibility of reinfection of the recovered patients. In addition, two more equations were considered separately to allow for the tracking of accumulated number of deaths, and current infections. The time dependent parameters for rates of transmission and mortality were taken to be of the form of specific sigmoid functions, to accommodate for the confinement measures and quarantine of diagnosed cases. Since most of the cases went undetected, the probability of detection had to be incorporated into the model paradigm. Finally the data fitting was done using the accumulated data for infected, deaths, recovered and currently infected, using the least squares approach. The analysis carried out in case of Germany and Spain suggested that a second wave of magnitude larger than the current wave is very likely.

A new model which adds a quarantine compartment to a SEIR model was proposed and analyzed in [2]. The primary objective of this SEIRQ model was to accommodate the impact of implementation of quarantine, in the paradigm of restricting contact in the population. The study was carried out in case of the outbreak in Poland,

---

[*]Indian Institute of Science Education and Research Pune, Pune-411008, Maharashtra, India, e-mail: suryadeepto.nag@students.iiserpune.ac.in

[†]Department of Mathematics, Indian Institute of Technology Guwahati, Guwahati-781039, Assam, India, e-mail: pratim@iitg.ac.in

with the focus being on the time window for which social distancing regulation were in place, while considering different scenarios for easing of the restrictions, in conjunction with enhanced testing and contact tracing. In another work on SEIR type model [3], infectious disease modeling was carried out incorporating quarantine, as well as testing. The motivation of this work was achieving greater clarity about the impact of testing individuals who are asymptomatic, and also planned quarantine, on the evolutionary dynamics of coronavirus. Adoption of quarantine without testing, can result in quarantine cases, which can be known, unknown or recovered. However, testing enables a bifurcation between leading to quarantine of infected individuals and non-infected individuals being released from quarantine. In [3], a parameter is determined so as to distinguish between the various transmission rates in the quarantined population, which would make it possible to execute a comparative analysis vis-a-vis, with only testing, as well as quarantine protocols put in place. In conclusion, the authors observed that more testing along with effective quarantine protocols, can lead to reduction in the economic cost of the outbreak, as well as less logistical burden for hospitals.

An improved version of the basic SEIR model was adopted for the modeling of COVID-19 outbreak in [4], to incorporate specific dynamic compartments and parameters pertinent to the epidemic. The focus of the work was on public health strategies in response to the outbreak, particularly the aspects of enforcement of lockdown and their relaxation, adopted across geographical locations. The methodology was illustrated in case of a small college town and incorporated the control strategies adopted over time, in the state of New York. The adapted model took into account a series of factors, including transmission, asymptomatic as well as presymptomatic transmission, mortality, immunity and age variability in the epidemiological profile. In addition to the SIR compartments, the model accounted for compartments of latently infected individuals, asymptomatic individuals (who may infect others, despite not presenting any symptoms), presymptomatic individuals (who may infect others and are likely to exhibit symptoms soon) and the fatalities. The key driver of modeling the social dynamics was by allowing for age-specific mobility, especially the daily travel routine. In the context of the outbreak in India, Gupta et al. [5] focused on the estimation of three epidemiological aspects, namely, estimation of parameters, effectiveness of the lockdown imposed in March 2020, and strategies of relaxation in the lockdown. The data from the lockdown period was analyzed to determine the basic, as well as the time varying reproduction number. An adaptation of the SEIR model was incorporated in the study setup to include a mandated quarantine (but not preemptive social distancing) and the key model parameters were estimated using the incidence data.

A key idea behind the model presented in this paper is to factor in the public response to the growth and progression of a pandemic. In [6] Reluga visited this crucial question in the context of how the public reluctance to put up with the costs of social distancing during an epidemic, results is limiting the effectiveness of using social distancing as a control mechanism. The author proposed a differential game model to ascertain the best possible approach of social distancing behavior during an outbreak. This was accomplished by analyzing the extent of value of social distancing by the determination of equilibrium behavior, for different cost functions. The conclusion was that the key drivers of this analysis are the basic reproduction number and the baseline indicator of efficacy resulting from social distancing. In particular, the benefits of social distancing is optimal for basic reproduction numbers in the vicinity of 2, and sole reliance on social distancing, in absence of other interventions, including vaccination, limits the recovery of the cost of infection to a ceiling of 30%.

2 MATHEMATICAL MODEL

We propose and present a fairly generalized model for COVID-19, in the paradigm of the SEIRQ model setup, motivated by the model in [2]. The model takes into consideration the classical modeling approach of epidemics by incorporating the susceptible, exposed, infectious and removed individuals. This is then extended to a version of the classical SEIR model, to include the impact of social distancing/isolation. In addition, the infectious and the recovered compartments are each subdivided into two compartments. Also the social distanced compartment

2

is divided into three compartments, in addition to the death compartment. The schematic representation of the model is given in Figure 1.
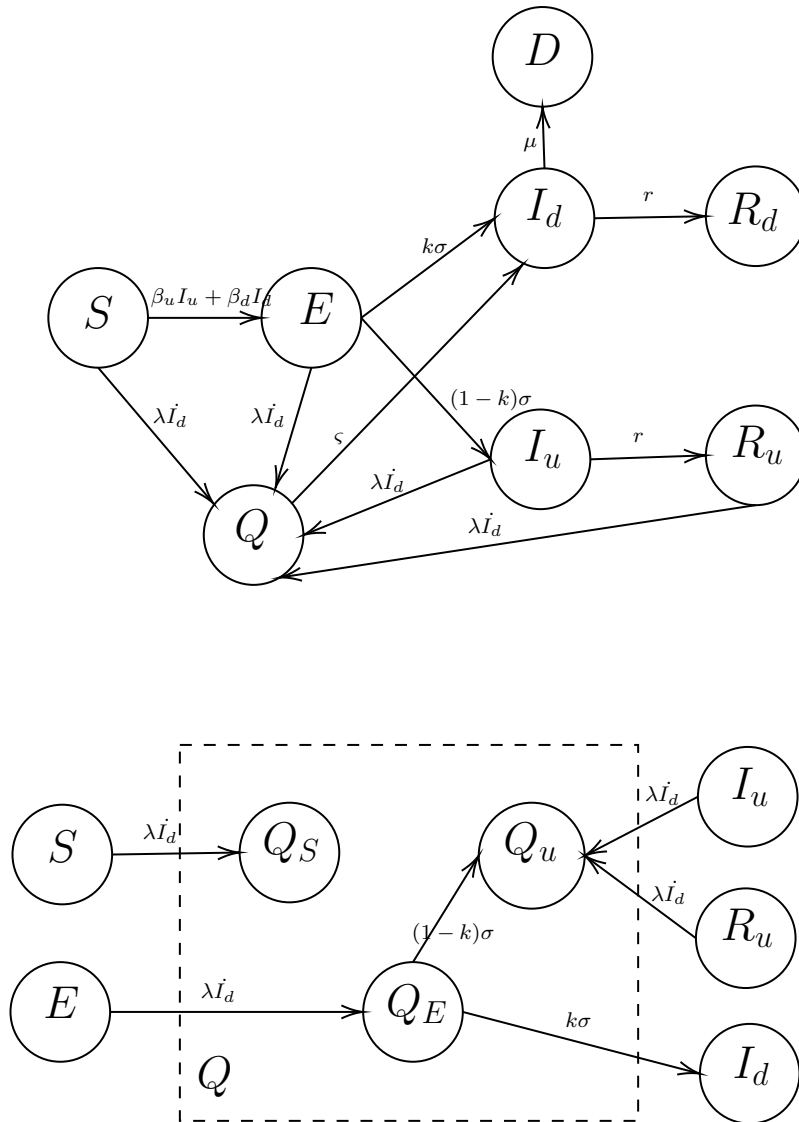

Figure 1: Schematic representation of the mathematical model.

We assume that the infectious population are of two types, namely, that which has been detected, and that which has not been detected. Further, the additional assumption is that the undetected cases are either mild or asymptomatic in nature, since it is reasonable to expect that cases which require medical attention will very likely be detected, due to visible symptoms. This was true especially in the first wave of the pandemic in India when the load on the healthcare system was not as much as that in the second wave. This assumption is crucial since it will have a ramification on the decisions pertaining to the adoption of social distancing practices. There are two other implicit assumptions inherent in the model, namely, that the chances of an individual who has adopted social distancing habits, getting infected is negligible, and that all individuals who have been infected once are not likely to experience a reinfection.

As seen from Figure 1, the modeling representation can be made in two parts. Although this two part setup is equivalent to a single system, wherein each compartment interacts independently, it is useful to classify Q_S (social distanced susceptible), Q_E (social distanced exposed) and Q_u (social distanced unidentified), under the umbrella of the "social distanced class" Q (We stick with Q to describe the variable, following literature on SEIRQ models).

(i) In the absence of social-distancing norms being in place, a susceptible individual (S) becomes exposed (E). However, once exposed, after the incubation period, the individual may either be asymptomatic or develop a mild illness, and may be undiagnosed for COVID-19 (I_u) or get tested positive for the infection (I_d). If they recover, they respectively move to the classes of undetected recovered (R_u) or detected recovered (R_d). For some cases of serious infections in I_d, the case may lead to a fatality and the individual goes to the death (D) compartment.

(ii) However, due to widespread public apprehension, resulting primarily from progressively rapid rise in case loads, as well as restrictions imposed by the health authorities, individuals may adopt the practice of social distancing. The susceptible individuals go on to the Q_S compartment, and are safe as long as they continue to be in Q_S. Since exposed individuals are not aware that they have been already exposed to the virus, they may apprehensively move to the Q_E compartment. In case they develop symptoms and test positive, they move to I_d, otherwise they will join other un-diagnosed infected persons who (unaware of their infection) preemptively practice social distancing in Q_u. While individuals in Q_u may be infected, they cannot spread the infection as they are themselves in social distancing. Un-diagnosed yet recovered individuals may also want to social distance and since it does not make a difference to the model (they cannot infect), we can group them with Q_u. Note that in Figure 1, the outflow from Q to I_d shown as ςQ, is actually the flow shown from Q_E to I_d which is $k\sigma Q_E$.

In order to model the social distancing, we assume that the rate with which people move in and out of social distancing is proportional to the rate of change of active cases. If there is a surge of new cases, then people are likely to be apprehensive and adopt greater precautions, while if there are few cases and a large number of recoveries, people are more likely to venture out rather than stay at home. This, however applies to those people who have not contracted the disease yet (to their knowledge). If X is a compartment and Q_X is the corresponding compartment of those in X who have socially distanced, then we model the flow by,

$$\dot{Q}_X = \lambda \dot{I}_d X,$$

where λ is a constant and I_d is the number of detected active infections. Such a relationship implies that as the number of active diagnosed cases increases (positive $\dot{I}_d$), more individuals will isolate themselves in apprehension, and as the number of cases decrease, individuals will stop isolating themselves. Obviously, the rate with which people isolate will depend only on detected cases. Thus the mathematical model for the schematic representation in Figure 1, is given by,

$$
\begin{aligned}
\dot{S} &= -(\beta_u I_u + \beta_d I_d)S - \lambda \dot{I}_d S \\
\dot{E} &= (\beta_u I_u + \beta_d I_d)S - \sigma E - \lambda \dot{I}_d E \\
\dot{I}_u &= (1-k)\sigma E - r I_u - \lambda \dot{I}_d I_u \\
\dot{I}_d &= k\sigma(E + Q_E) - r I_d - \mu I_d \\
\dot{R}_u &= r I_u - \lambda \dot{I}_d R_u \\
\dot{R}_d &= r I_d \\
\dot{D} &= \mu I_d \\
\dot{Q}_S &= \lambda \dot{I}_d S \\
\dot{Q}_E &= \lambda \dot{I}_d E - \sigma Q_E \\
\dot{Q}_u &= \lambda \dot{I}_d (I_u + R_u) + (1-k)\sigma Q_E,
\end{aligned}
\tag{2.1}
$$

subject to the constraint,

$$S + E + I_u + I_d + R_u + R_d + D + Q_S + Q_E + Q_u = N.$$

Here, the state variables represent the compartments as described (and labeled) above. The constraint equation ensures that all the compartments add up to the population size N.

3 METHODOLOGY

We use data downloaded from the website of COVID-19 tracker in India [7]. It is a publicly available database for the data on the spread COVID-19 pandemic in India. We use daily time series on cumulative numbers of total confirmed cases, total recoveries and total deaths in India. The time series of the diagnosed cases and the diagnosed cases for the period of 30th January 2020 to 3rd March 2021 and the period of 12th September 2020 to 3rd March 2021, are given in Figures 2 and 3, respectively. We restrict our study to the retreat of the virus from the
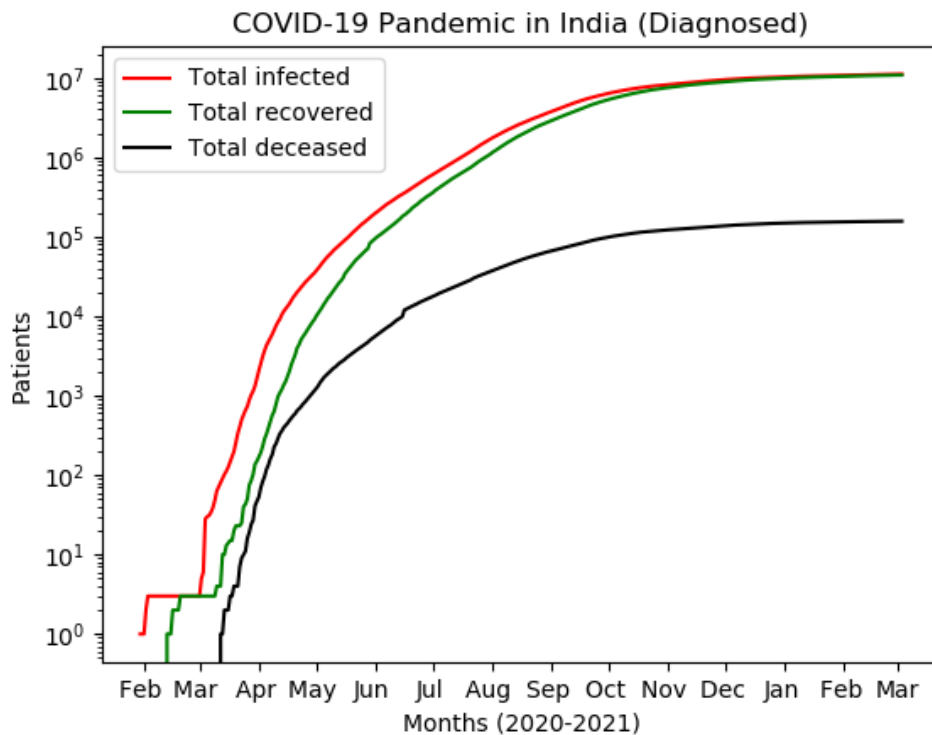

Figure 2: Time Series for the Diagnosed between February 2020 to 3rd March 2021

peak. This was done primarily for two reasons. Firstly, the Government of India made periodic policy decisions and therefore for majority of mid-2020, the "un-lockdown" happened in phases, and it would be unreasonable to expect the parameters to have constant values through these periods. Another reason is that the number of cases earlier were much lesser than at the peak and small errors in the data turn out to be large fractions of the actual number. Moreover there are discrepancies in the data, for instance, on 16th June 2020, there was a sudden spike in the number of deaths (greater than 2000). These problems do not affect the latter part of the time series. We consider the data after 12th September 2020, as this is when there was an almost consistent drop in the number of new daily infections for the remaining months of 2020 and the beginning of 2021. And we consider data only up to 3rd March 2021, before the second wave of the pandemic emerged. The reason that we do this is because of a different strain of virus being responsible for the surge of cases in the second wave, which would imply newer parameters, and since we cannot quite identify when the newer strain began causing infections, and what fraction of cases were due to the old and the new strain, we assume the presence of only one strain causing infections in the period considered. Other assumptions like the number of undetected patients dying from the illness being negligible would also become unrealistic for the second wave since a large number of those affected died in the
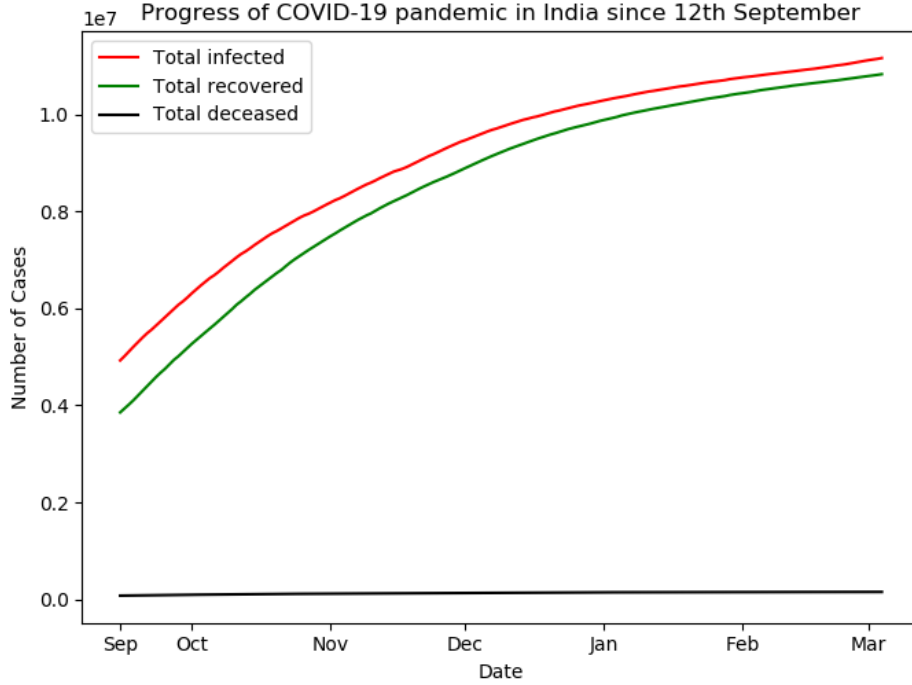
5

Figure 3: Time Series for the Diagnosed between 12th September 2020 to 3rd March 2021

second wave due to a shortage of test kits and hospital beds, as well as individual reluctance to get tested/admitted. In the following subsections we elaborate upon the approach adopted for estimation of the model parameters.

3.1 DEATH RATE AND RECOVERY RATE

The equations for the death rate and the recovery rate can be solved independent of the other parameters, from the data of recoveries and deaths. To calculate the death rate, we solve the equation,

$$\dot{D} = \mu I_d.$$

We have available daily data for both $\dot{D}$ and I_d as a function of time. A key assumption in this analysis is that deaths occur only from reported cases. We believe this assumption to be reasonable because any case with serious COVID-19 symptoms is likely to be reported, and most undetected cases are mild and asymptomatic in nature. We observe from the data that the quantity $\dfrac{\dot{D}}{I_d}$ appears to be exponential, and, accordingly, instead of considering μ to be a constant, it would be more prudent to model it like an exponential, as we see the death rate falling off with time and saturating at a lower level. This could be because of the healthcare system becoming more adept with dealing with COVID-19 patients, or because of the virus losing strength. Accordingly, we consider,

$$\mu := \mu(t) = \mu_0 e^{-\delta t}$$

To estimate μ_0 and δ we take the log of $\dfrac{\dot{D}}{I_d}$ and fit a straight line to it by using least squares estimation. We use the `curve_fit` function from the python package Scipy.optimize for this purpose. Subsequently, we plot $\dot{R}_d$ against I_d. Unlike the death rate, we observe that the daily number of recovery cases is roughly a straight line with respect to the number of active cases, and therefore we can assume r to be a constant, which we estimate from the data of daily recoveries, using `curve_fit` on the data points for $\dfrac{R_d}{I_d}$.

6

3.2 Choice of Known and Unknown Parameters

In all our system of ODEs require 10 initial conditions and has 9 constants. But the estimation of these 19 parameters, will be computationally expensive and lead to greater inaccuracy in the estimation. Accordingly, we would like to strike a balance between being parsimonious in the number of unknown parameters, and the realism of referring to literature for the value of parameters. Fortunately, we know some initial conditions from the data itself *i.e.,* $I_d(0)$, $R_d(0)$ and $D(0)$. Additionally, we can estimate some parameters like the recovery rate and death rate simply from observing the data, as has been done in the preceding Subsection 3.1. Apart from this, the population of the country is a known fact and need not be estimated from the data, and since the average time taken for an exposed person to become infectious is a universal biological fact for the virus and does not depend on a person's environment, we consider σ to be a known parameter as well, and fix its value from literature. The nature of the physical quantities that the remaining 11 parameters represent is such that they depend on the geography, demographics, healthcare system and government measures in place. For instance β_u and β_d depend on "how infectious" are the persons from classes I_u and I_d, respectively, strongly dependent on how the people in these classes behave, government policy decisions, and of course the socio-economic circumstance in the country. Similarly the constant k which determines what fraction of the infected population is diagnosed, will depend on government standards for testing criteria, the availability of test kits, the willingness of people to get tested, etc, and thus needs to be determined specifically for different countries and geo-political areas. The same is true for the constant λ and needless to say, for the unknown initial conditions. Therefore we expect that the value of these parameters will be unique to the place of study (in this case, India), and hence they serve as the unknown parameters that we hope to estimate.

3.3 Estimation of Unknown Parameters

For estimating the unknown parameters, we use Python's Scipy.integrate package. We first use the `odeint` function to solve the system of ODEs, and define the objective function for the problem. The function numerically solves the system of ODEs. Therefore for a parameter vector Θ, we have a solution of equation (2.1). To get a good estimate of Θ from the data, we would like to minimize the bias *i.e.,* reduce the training error. Therefore we aim to find Θ such that the solution of the system of ODEs is most compatible with the data we have. Since we have three sets of data in the confirmed cases, recovered cases and deaths, we would want a loss function minimized such that it the total discrepancy between the solution of the ODEs and the data is minimum. For this we rewrite our dataset to be a concatenated array of confirmed cases, recoveries and deaths. We then use the `curve_fit` function to find the parameters that fit the data the best for that objective function. The `curve_fit` function works on the objective function which returns a concatenated array of the solution of the system of the ODEs in confirmed cases, recoveries and deaths. For the estimation, we specify reasonable upper and lower bounds for the parameters. We make guesses about the bounds of the initial conditions based on the data available, and we set the bounds for λ such that none of the compartments attain negative occupancies in the duration of the study. For better results, we also start with an initial guess of what the parameters could be. Then the `curve_fit` function returns the optimum set of parameters.

3.4 Error Analysis

To get a measure of the error in our estimation of the parameters, we plot the training error and calculate the 95% confidence intervals for each of the estimated unknown parameters. For the training error, we plot the absolute difference between the data and the simulation for the estimated parameters, and scale the error by the value of the solution, as determined by the simulation at that point in time. We individually look at the relative training errors in each of the data for confirmed cases, recovered cases and deaths.

We also calculate confidence intervals by a method called parametric bootstrap [8]. In particular, we use a method called residual bootstrap [9]. A residual bootstrap is simply a method of repeated sampling (with replacement) of the residues obtained at each data point (training error), and using this to generate a synthetic distribution for the population. In this method, we simulate the model with the estimated parameters and calculate the residual vector. Here, we construct the residual vector by concatenating the residuals left in each of the time series of confirmed cases, reported recoveries and deaths. We assume the set of residuals to represent the distribution of residuals and generate 1000 new time series wherein the synthetic (bootstrapped) data is generated by adding synthetic residuals randomly sampled (with replacement) from the real set of residuals to the predicted data (from the model). For each of these synthetic data sets, we estimate the parameters in the method described above. This gives us 1000 estimated values for each parameter after bootstrapping. This set of parameters is assumed to be the distribution for the population, and the 95% confidence interval for each parameter is estimated from them by simply considering the data points at the 2.5-th and the 97.5-th percentiles.

4 RESULTS AND CONCLUSION

In this Section, we present the results from the methodology outlined in Section 3, being implemented on the dataset. We begin with the presentation of the estimates of the parameters representing the death rates and the recovery rates, and subsequently enumerate the estimates of the remaining unknown parameters, including the initial conditions for our model system (equation (2.1)), which is the main focus of this paper.

Recall that the death rate parameter was taken in the exponential form $\mu := \mu_0(t) = e^{-\delta t}$, and hence our estimate of the parameters μ_0 and δ (resulting from the fit of $\ln\dfrac{\dot{D}}{I_d}$) are tabulated in Table 1. In a similar manner, the recovery rate r estimated from fitting the data of $\ln\dfrac{R_d}{I_d}$ is also tabulated in Table 1. An important result of the preliminary estimation is that the death rate has seen a downward trend over time. This could possibly be attributed to greater awareness which led to earlier hospitalization of patients, overall better healthcare facilities, and medical institutions being better equipped with dealing with patients. It may also be due to mutations in the virus that reduce its strength over time. The fitted data using the estimated death rate and the recovery rate are presented in Figures 4 and 5, respectively.

Parameter	Value	Source
μ_0	0.09586	Estimated using Subsection 3.1
δ	0.00086	Estimated using Subsection 3.1
r_d	0.09105	Estimated using Subsection 3.1

Table 1: Estimation of Death Rate and Recovery Rate.

The values of the known parameters as whose choice was delineated in Subsection 3.2 are tabulated in Table 2. Based on the values estimated (as tabulated in Table 1) and the known parameters, along with confidence intervals (as enumerated in Table 2), the unknown parameters, along with confidence intervals, were estimated (as elaborated in Subsection 3.3) and are tabulated in Table 3. The fit for the model prediction in case of confirmed, deceased and recovered cases are presented in Figures 6, 7 and 8, respectively. Finally, the error analysis during the training phase is depicted in Figure 9.

We can observe from Figures 6, 8 and 7 that the model has fit the data well. Further, we can observe from Figure 9, that the errors for the confirmed and the recovered cases never exceeds 1%, which given the large data set, and the involvement of 8 parameters, is a highly satisfactory fit. We see that from the end of February, the error

8

Parameter	Value	Source
$I_d(0)$	990458	Estimated using Subsection 3.2
$R_d(0)$	3856265	Estimated using Subsection 3.2
$D(0)$	80224	Estimated using Subsection 3.2
σ	$\dfrac{1}{3.5}$	[2]
r	0.09105	Estimated using Subsection 3.2
μ_0	0.09586	Estimated using Subsection 3.2
δ	0.00086	Estimated using Subsection 3.2
N	1.3664×10^9	World Bank

Table 2: Estimation of Known Parameters.

Estimated Parameters	Estimated value	95% confidence intervals
$E(0)$	1356533	$[26682, 3040818]$
$I_u(0)$	3734197	$[103624, 4513229]$
$R_u(0)$	9719530	$[8699709, 9835259]$
$Q_s(0)$	999937439	$[974171025, 999999137]$
$Q_e(0)$	13074	$[105, 98743]$
$Q_u(0)$	999411	$[857484, 999886]$
β_u	0.2141	$[0.1714, 0.4476]$
β_d	0.5666	$[0.2533, 0.6692]$
k	0.2544	$[0.1239, 0.2913]$
λ	7.4308×10^{-5}	$[6.520 \times 10^{-6}, 0.0010]$

Table 3: Estimation of Unknown Parameters.

in the confirmed cases begin to grow rapidly, as can be seen in Figure 6, as well where the data starts exceeding the curve of best fit. This is most likely due to the onset of the second wave of infections, and the double-mutant strain. However, note that, the error in recoveries in this time period did not shoot up because of the time-lag in the infections and recoveries, which implies that a very small amount of the data can really be attributed to the period where the double-mutant variant was active, which almost entirely validates our initial assumption that only one strain of the virus was active in the period considered.

If we look at the error in the death rate, we see that it tends to be slightly higher, at a maximum of about 2.5% of the data. This is because of a multiple reasons. Firstly, since our fitting algorithm involved a least squares minimization of the absolute errors in the confirmed, recovered, and fatal cases, the relative error is more likely to be more in the case of deaths, because the total number of confirmed cases and recoveries is much higher and the error will also be proportional, and the fitting algorithm will prioritize restricting the errors here, compared to the smaller errors in the number of deaths. That being said, it is important to note that the absolute training error (at a 2.5% level for deaths) is still much lesser than those recorded in the case of confirmed and recovered cases. The second reason for this error results from the caveat of the assumption that the death rate behaves like a perfect exponential decay. While the initial period and tail follows an exponential, there is a dip and a rise in the death rate in the October-December period. We observe from Figure 9 that this is also the period in which the training error for the number of deaths is maximum. We can also see this in Figure 7 as well, where the data for

the number of deaths first dips and then rises back to the curve of fit in this period.

From the estimated parameters we can draw a few significant conclusions. Firstly, the results suggest that there is a very large number of susceptible persons who are practicing social distancing. The explanation of such a large number is that the model includes the population of all those people in India who are susceptible but protected from infection by a variety of temporary factors. This includes the large numbers of people voluntarily social distancing because of the virus at its peak, and it also includes large numbers of people from those districts of the country which did not have any cases, or a large fraction of people from geographical regions where cases were sparse. Now while these people have not actively social distanced, they are for all practical purposes isolated from infectious individuals. While ideally, to get a measure of the extent of social distancing, we should have only been looking at areas where the epidemic is actually spreading, but then it would have been impossible to quantify the total number of people in the system, which we have had to assume is the population of the country. The fact that the large number estimated in the social distanced-susceptible compartment is due to geographically secure individuals and not a flaw in the model, is evident from the estimated initial conditions for Q_E and Q_u, which are much lesser fractions of the initial numbers of E and $(I_u + R_u)$ respectively, despite λ having the same value for all. This is because there would be no individuals in the Q_E and Q_u compartments in geographically isolated regions, but there would be a large number of individuals in Q_S.

Another important conclusion that we can draw is that $\beta_d > \beta_u$ which means that the rate of infection from a diagnosed individual is greater than from an undetected infection. This is in line with the empirical results investigated in [10, 11, 12]. Each of these studies confirmed from empirical results that asymptomatic patients had a lower relative transmissibility or infectivity than symptomatic patients. Moreover the value estimated from our fit is also in the neighbourhood of these empirical results. An explanation for this may come from the reasonable assumption that the undiagnosed cases are usually mild or asymptomatic, a consequence of which would be a lower viral load.

The final conclusion we can draw is that the fraction of cases which go undetected are in a majority. According to the estimate, only a little over a fourth ($k = 0.2544$) of the cases are actually diagnosed. However, due to the nature of the infections, any attempt to estimate the number of asymptomatic patients will be an underestimate. There are still studies which estimate the number of known asymptomatic patients as over 30% [13]. Along with mild cases, the figure for undetected cases could be even higher, especially due to an overall stress on the healthcare system and a shortage of beds when the virus was at its peak along with expensive treatment and quarantining procedure which further discourages individuals from undergoing tests, unless they show significant symptoms. Another reason for individuals being reluctant to get tested and undergo treatment maybe the social stigma that has been observed in multiple instances, including isolating patients visiting hospitals for non-covid treatment.
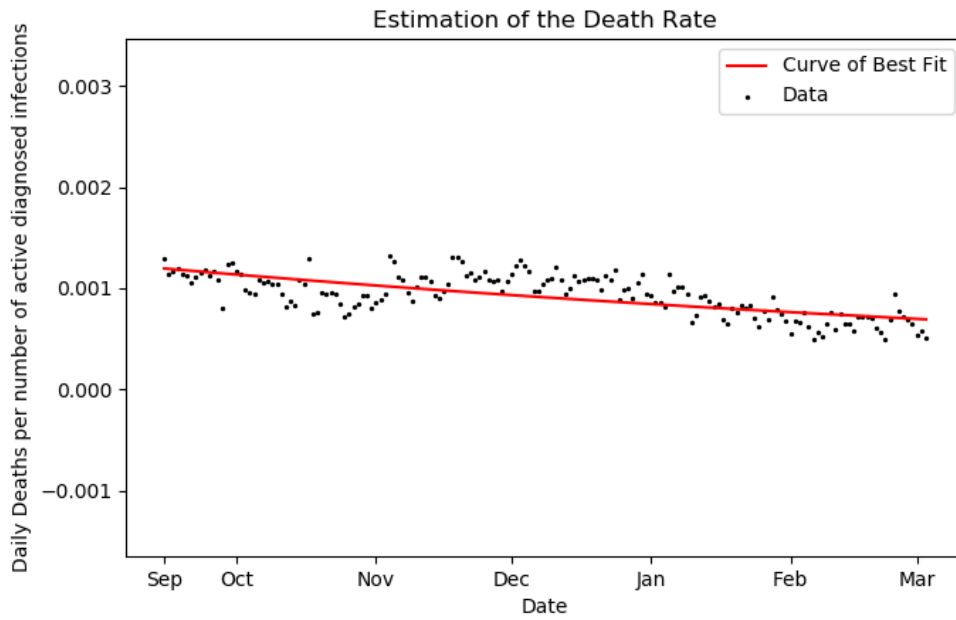
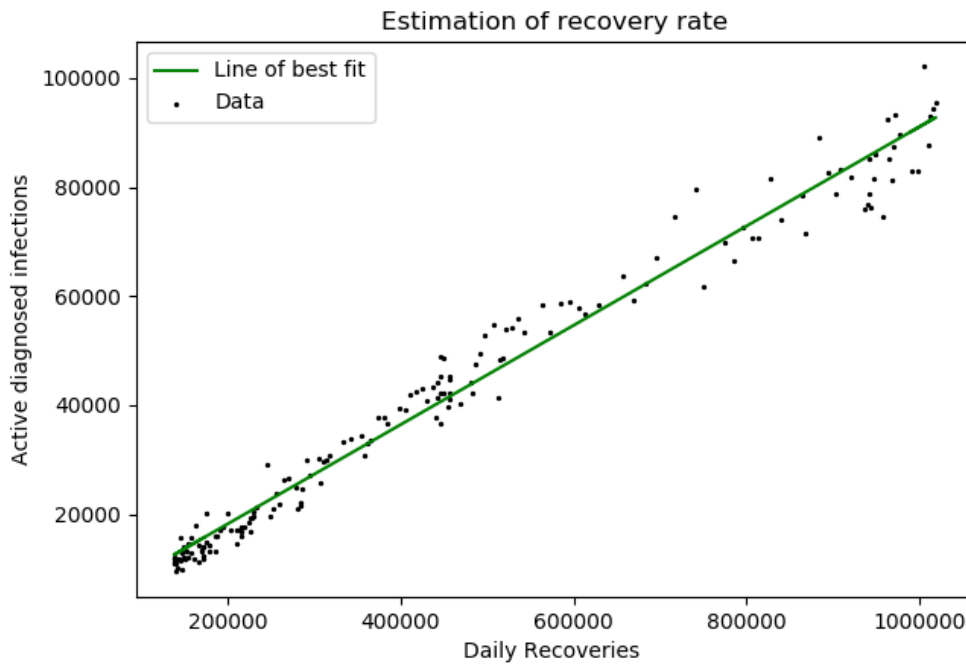Figure 4: Fitting of Death Rate to the Data
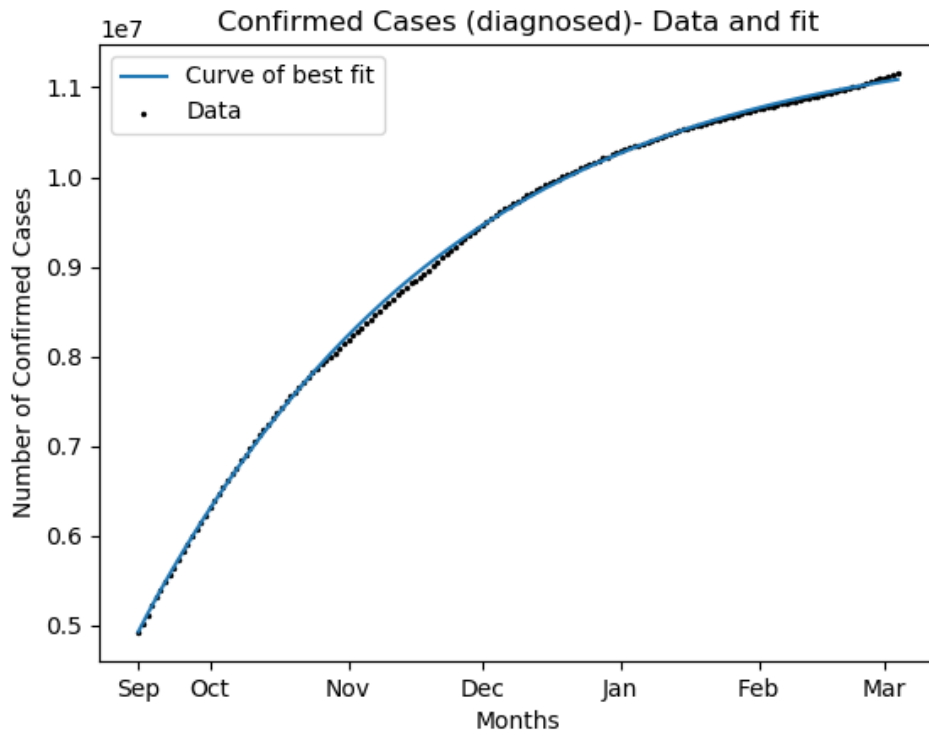

Figure 5: Fitting of Recovery Rate to the Data

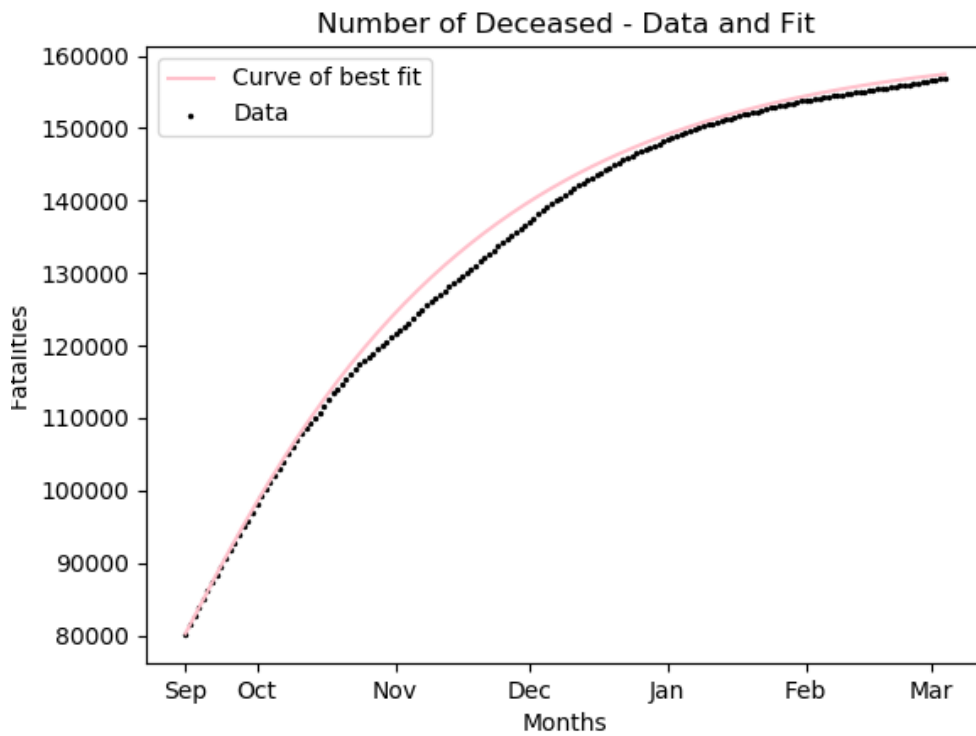Figure 6: Fitting of Confirmed Rate to the Data

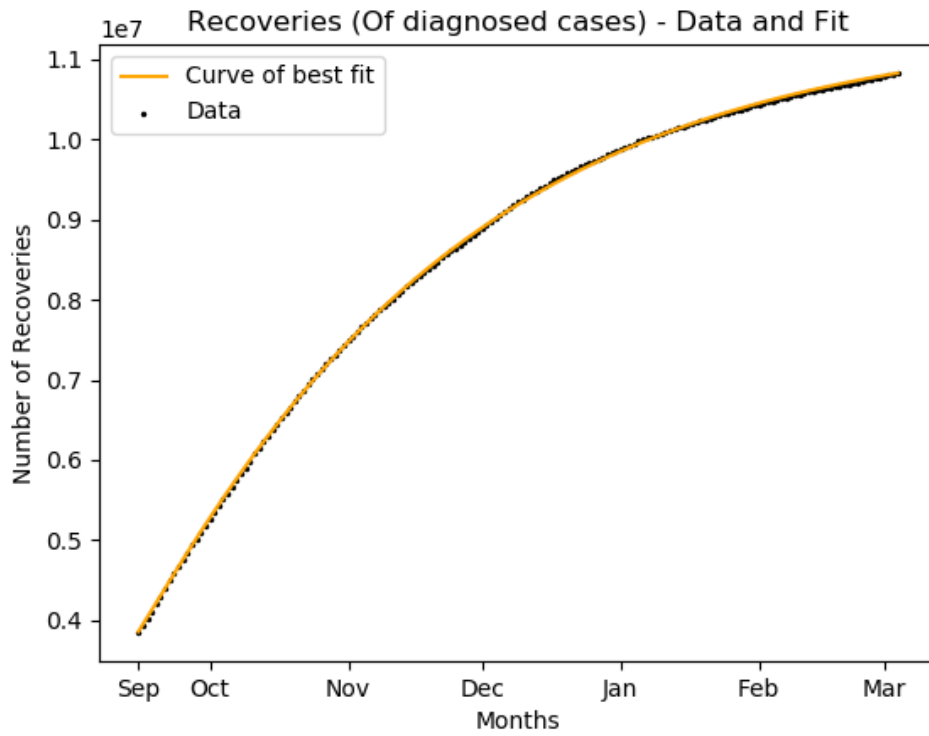Figure 7: Fitting of the Total Deceased Cases Data to the Model

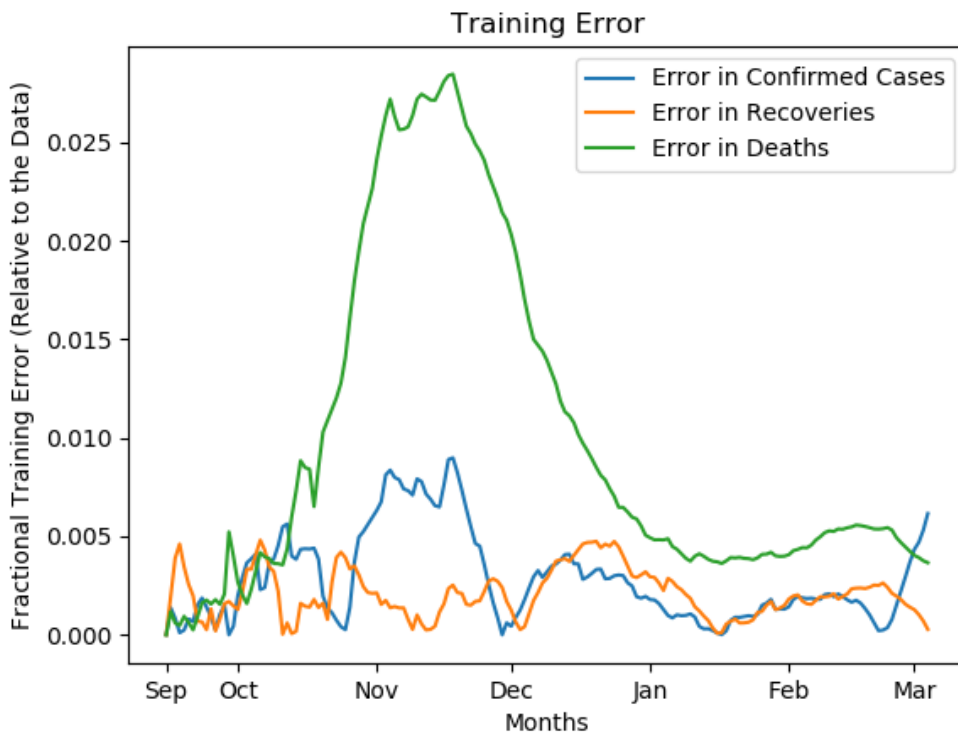Figure 8: Fitting of the Total Recovered Cases Data to the Model

Figure 9: Training Error

13

REFERENCES

[1] de Castro, F., 2020. Modelling of the second (and subsequent) waves of the coronavirus epidemic. Spain and Germany as case studies. medRxiv 2020.06.12.20129429.

[2] Piasecki, T., Mucha, P.B. and Rosińska, M., 2020. A new SEIR type model including quarantine effects and its application to analysis of Covid-19 pandemia in Poland in March-April 2020. arXiv preprint arXiv:2005.14532.

[3] Berger, D.W., Herkenhoff, K.F. and Mongey, S., 2020. An seir infectious disease model with testing and conditional quarantine (No. w26901). National Bureau of Economic Research.

[4] Radulescu, A., Williams, C. and Cavanagh, K., 2020. Management strategies in a SEIR-type model of COVID 19 community spread. Scientific reports, 10(1), pp.1-16.

[5] Gupta, M., Mohanta, S.S., Rao, A., Parameswaran, G.G., Agarwal, M., Arora, M., Mazumder, A., Lohiya, A., Behera, P., Bansal, A. and Kumar, R., 2021. Transmission dynamics of the COVID-19 epidemic in India and modeling optimal lockdown exit strategies. International Journal of Infectious Diseases, 103, pp.579-589.

[6] Reluga, T.C., 2010. Game theory of social distancing in response to an epidemic. PLoS Comput Biol, 6(5), p.e1000793.

[7] COVID-19 India Data, https://www.covid19india.org/

[8] Efron, B. and Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical science, pp.54-75.

[9] Residual Bootstrap, http://brainimaging.waisman.wisc.edu/~chung/asymmetry/wild.bootstrap.pdf

[10] He, D., Zhao, S., Lin, Q., Zhuang, Z., Cao, P., Wang, M.H. and Yang, L., 2020. The relative transmissibility of asymptomatic COVID-19 infections among close contacts. International Journal of Infectious Diseases, 94, pp.145-147.

[11] Nakajo, K. and Nishiura, H., 2021. Transmissibility of asymptomatic COVID-19: Data from Japanese clusters. International Journal of Infectious Diseases, 105, pp.236-238.

[12] Sayampanathan, A.A., Heng, C.S., Pin, P.H., Pang, J., Leong, T.Y. and Lee, V.J., 2021. Infectivity of asymptomatic versus symptomatic COVID-19. The Lancet, 397(10269), pp.93-94.

[13] Nishiura, H., Kobayashi, T., Miyama, T., Suzuki, A., Jung, S.M., Hayashi, K., Kinoshita, R., Yang, Y., Yuan, B., Akhmetzhanov, A.R. and Linton, N.M., 2020. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). International journal of infectious diseases, 94, p.154.