medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

A data first approach to modelling Covid-19

Jayanti Prasad*

Khagol-20, 38/1, Panchavti, Pashan, Pune (India) - 411008

June 2, 2020

Abstract

The primary data for Covid-19 pandemic is in the form of time series for the number of confirmed, recovered and dead cases. This data is updated every day and is available for most countries from multiple sources such as [Gar20b, iD20]. In this work we present a two step procedure for model fitting to Covid-19 data. In the first step, time dependent transmission coefficients are constructed directly from the data and, in the second step, measures of those (minimum, maximum, mean, median etc.,) are used to set priors for fitting models to data. We call this approach a "data driven approach" or "data first approach". This scheme is complementary to Bayesian approach and can be used with or without that for parameter estimation. We use the procedure to fit a set of SIR and SIRD models, with time dependent contact rate, to Covid-19 data for a set of most affected countries. We find that SIR and SIRD models with constant transmission coefficients cannot fit Covid-19 data for most countries (mainly because social distancing, lockdown etc., make those time dependent). We find that any time dependent contact rate decaying with time can help to fit SIR and SIRD models for most of the countries. We also present constraints on transmission coefficients and basic reproduction number $\mathfrak{R}_0$, as well as effective reproduction number $\mathfrak{R}(t)$. The main contributions of our work are as follows. (1) presenting a two step procedure for model fitting to Covid-19 data (2) constraining transmission coefficients as well as $\mathfrak{R}_0$ and $\mathfrak{R}(t)$, for a set of countries and (3) releasing a python package **PyCov19** [Pra20b] that can used to fit a class of compartmental models, with time varying coefficients, to Covid-19 data.

1 Introduction

At present the world is going through an unprecedented crisis of pandemic Covid-19 caused by a novel form of coronavirus, named Sars-CoV-2 which was passed to the human from bats in the Wuhan city of China, some time in December 2019 [Org20a, Org20b, ea20h, ea20t, ea20v, ea20d, ea20r, ea20l]. Till the end of May 2020 the virus has reached almost all the parts of the world resulting in more than six million people infected and more than a quarter million deaths [Wor20]. The measures to contain the virus medically by developing a vaccine are going on war footing. However, the success is still

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

*prasad.jayanti@gmail.com

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

expected to be a few years away [ea20f]. Till a fraction of the population develop (herd) immunity or the vaccine is ready, the only means to contain the pandemic are social measures (social distancing, contact tracing etc.,) and enhanced hygiene practices [ea06, ea20s, ea20p].

Some of the most important problems related to Covid-19 research are (1) estimating the controlling parameters of the pandemic, (2) making short term predictions using mathematical-statistical modeling which can help in mitigating policies (3) simulating the growth of the epidemic by taking into account as many contributing effects as possible and (4) quantifying the impact of mitigation measures, such as lockdown etc [ea20j].

Modeling Covid-19 pandemic with compartmental models of Kermack and McKendrick (for an introduction see [JR08, Li18, BC18]) has been one of the most active problems in the recent times [ea20p, ea20a, ea20c, ea20e, FP20, ea20m, Oli20]. There have been alternative approaches also such as [IM20] where statistical considerations are being taken into account for predictions. In one of the studies [FP20] it is argued that the data for the confirmed, recovered and dead, all three can easily fit a power law model with similar coefficients. The main attractive feature of these data driven approaches is that the complexity of the model being considered is determined by the data and not by theoretical expectations.

In the present work we follow a middle approach and fit two compartmental models, named SIR and SIRD with some modification, to the Covid-19 data. One of the main reasons to consider these models has been that the Covid-19 data is available only for the **S**usceptible, **I**nfected, **R**ecovered and **D**ead compartments (for the notations used here and other places in the present work see table (1)). It may be true that a large fraction of the population which may be **E**xposed (defined later) play an important role in the dynamics of the pandemic however, it is hard to get reliable numbers for that. Apart from that, a large number of undocumented cases [ea20l] may also have significant influence on the spread of the pandemic.

A brief summary of the work presented here is as follows.

In §2 we give a brief introduction to the compartmental models and introduce the notations and variables used in the work. In particular, we discuss the SIR model in §2.1 and the SEIR and the SIRD models in §2.2 and §2.3 respectively. One of the major parts of the work presented here is to study the time dependence of the contact rate β, we introduce a set of parametric models of $\beta(t)$ in §2.4. We discuss the time series data used in the study in the §3 by giving an example of Italy which is one of the most affected countries. The main results of our work are given in §4 and in §5. In §4 we discuss the reconstruction (regression) procedure for the set of transmission coefficients as well as for the effective reproduction (defined later), number $\mathfrak{R}(t)$. Parameter estimation is discussed in §5. The main conclusions of our work with a summary and some important points are discussed in §6.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

2 Mathematical Modeling

Mathematical modeling of epidemics that started a century back with the seminal work of Kermack and McKendrick (see [NC08, JR08, Li18] for an introduction) is still the main framework most commonly used. The main idea of the Kermack and McKendrick's compartmental models is that every individual in a society belongs to one of the M compartments and the total number of individuals belonging to different compartments keep changing with time. The minimum value M can have is two, for the Susceptible-Infected-Susceptible (SIS) model, in which the recovery does not guarantee that one will not get the infection again [JR08].

During an epidemic phase an individual can go through many stages from being perfectly healthy to the recovered one after an infection, with or without any immunity (short or long term) or may die. If we represent every stage with a compartment and keep the track of the number of individuals in each compartment then we can easily model the dynamics of the epidemic. This approach is very similar to the approach taken in astronomy where we count the number of stars in different stages of their life to understand the stellar evolution.

In principle we can have any number of logical compartments but in practice we should consider only those compartments for which we have the counts data, in particular for model fitting. Taking into account the fact that we have data only for the number of confirmed, recovered and dead population, the only compartmental model that meets the requirement is the SIRD model. If we consider the recovered and dead together we get the SIR model as is discussed in the next section.

One of the important compartments that also is commonly considered is the 'exposed' one and represents the population which have received the infection but cannot pass to others, before a certain period called the incubation period. If we consider exposed population also then we get the SEIR model that also is discussed below. Three compartmental models SIR, SIRD, and SEIR are shown in the (a), (b) and (c) panels of Figure (1) respectively (for more detail one can refer to [JR08, Het00, ea97, Oli20]).

If we identify the compartments with the nodes of a graph then the transmission between different compartments, as is represented by a set of coefficients, can be considered the edges of the graph. Some of the nodes may have multiple edges and some of the edges could be bi-directional also. The main challenge of the modeling a pandemic like Covid-19 is not the scarcity of mathematical models but it is of the reliable data for the compartments being considered.

2.1 SIR Model

The most basic compartmental model is the SIR model which is shown in (a) of the Figure(1) and is described by the following set of equations:

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

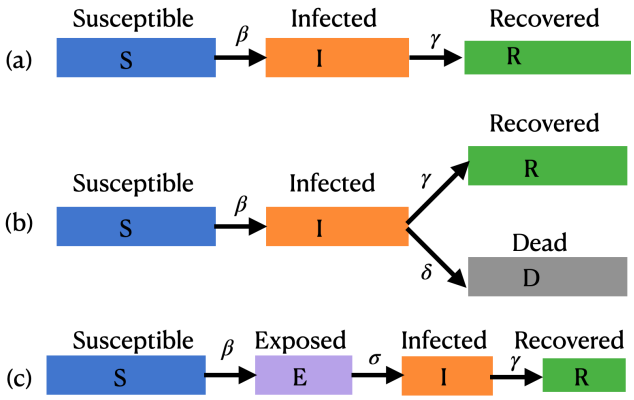

Figure 1: In the compartmental models the total population N is divided into a set of compartments as shown in the figure. The number of compartments and their connectivity depends on the detail of the model being considered. If we consider these compartments as nodes of a graph then there are transmission coefficients for every connecting edge that determine how effective that edge is in changing the population of the connected compartments. In (a) and (c) representing SIR and SEIR models, the compartments are connected in a linear way, however, for the case (b), representing the SIRD model, there is a branching also. Since the total population must remain a constant so the rates of change along all the connecting edges must add to zero.

S. No.	**Notation**	**Description**
1	**S**, **I**, **R**, **D**, **E**	Name of the compartments
2	S, I, R, D, E	Population in the compartments
3	$\mathcal{C}, \mathcal{R}, \mathcal{D}$	Cumulative numbers
4	$\mathfrak{R}_0$	Basic reproduction number
5	$\mathfrak{R}(t)$	Effective reproduction number

Table 1: Notations used in the work

$$\begin{aligned} \frac{dS}{dt} &= -\beta \frac{S}{N} I \\ \frac{dI}{dt} &= \beta \frac{S}{N} I - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{1}$$

Here β and γ are the transmission coefficients, also called the contact rate and the recovery removal rate respectively, and $1/\beta$ and $1/\gamma$ represent the mean duration of infectiousness and the average period of infectivity (see [Het00, ea03a]), respectively, (see Table (1 for notations).

In general there is some time lag between acquiring an infection and becoming infectious. However, in the SIR model it is ignored and an assumption is made that individuals become infectious immediately upon getting an infection. This is a very strong assumption and the main reasons for

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

making this is that we do not know reliably how many people are actually 'exposed', or have the virus but are still not infectious (cannot pass it to others). One of the ways to address this problem could be contact tracing and assuming that anyone who has come into contact with an infected person is an exposed one. However, this assumption is as strong as the assumption made in the SIR model.

If we do not consider the birth, death and movement of people then the following condition must be satisfied.

$$S + I + R = N \text{ (constant) } \text{ or } \frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0, \tag{2}$$

here S, I and R is the population of the **S**, **I** and **R** compartments respectively.

In equation (1) the transmission coefficient β is one of the most important parameters of the epidemic dynamics and can be written as the product of the contact rates (the average number of contacts per person per time) and the transmission probability (the probability of disease transmission on contact between a susceptible and an infectious person). As has been mentioned that the transmission coefficient γ can be identified with the recovery rate which is nothing but the inverse of the infectious period (during which an infected person can pass the virus to other healthy people).

In general, the equations (1) is solved with the following initial conditions:

$$S(t=0) = S_0 > 0, \quad I(t=0) = I_0 > 0, \quad \text{and} \quad R(t=0) = R_0 = 0. \tag{3}$$

The second equation from 1 can be written as:

$$\frac{dI}{dt} = \beta \left(\frac{S}{N} - \frac{\gamma}{\beta} \right) I, \tag{4}$$

and for $S/N > \gamma/\beta$ we get a positive infection rate.

Here we define one of the most important parameters of an epidemic in terms of the ratio β/γ, called the **basic reproduction number** $\mathfrak{R}_0$, when considered a constant, and called the **effective reproduction number** $\mathfrak{R}(t)$, when considered a function of time. The most common definition [JR08] of it is that it is *the average number of secondary cases arising from an average primary case in an entirely susceptible population.* Note that in the text we may also use just "reproduction number" and the meaning of it will depend on the context. Some of the studies such as [ea20a] call $\mathfrak{R}_0$ and $\mathfrak{R}(t)$ both as basic reproduction number, however, we follow the convention used in [ea03b, ea03a, Cob20].

The basis reproduction number $\mathfrak{R}_0$ is the main measure which quantifies the transmissibility of the virus and $\mathfrak{R}_0 > 1$, sets a chain of transmissions leading an exponential growth of the pandemic. We can keep $\mathfrak{R}_0 < 1$, by minimizing the contact rates (social distancing etc.,), lowering the infectiousness of the infected people (by treating them or putting them in a quarantine etc.) and reducing the susceptibility of the healthy people by vaccination etc., (for detail see [ea05]).

The SIR model is one of the most basic models and can be easily generalized by one or more of the following ways:

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

1. **Adding more compartments:** Depending on the type of pandemic and other details we are interested in we can add more compartments to the SIR model. These compartments can fit in between the existing ones (for example as shown in Figure (1) (c) for SEIR case) or can branch out from the existing once (as shown in Figure (1) (b) for the SIRD case). With every new compartment added we must include the transmission coefficients for the connecting edges and also need an initial population for the new compartment being added [BC18, BS20, ea20b, ea20u].

2. **Heterogeneous population:** In the basic SIR model we consider a homogeneous population which share transmission coefficients such as β and γ. This may not be true in practice, for example, people from different groups (based on age, medical conditions, gender etc, etc) may have different contact rate β and/or recovery rate γ [ea20p, RR20].

3. **Variable transmission coefficients:** Theoretically transmission coefficients such as β and γ are considered constant, however, in practice they can vary with time due to multiple factors as the pandemic spreads [Cob20, P.20, GPD20]. For example, social distancing and other precautions such as hand wash etc., may help to lower β. In a similar way a better understanding of the disease and the urgency with which medical and testing facilities are brought online may improve the recovery rate γ.

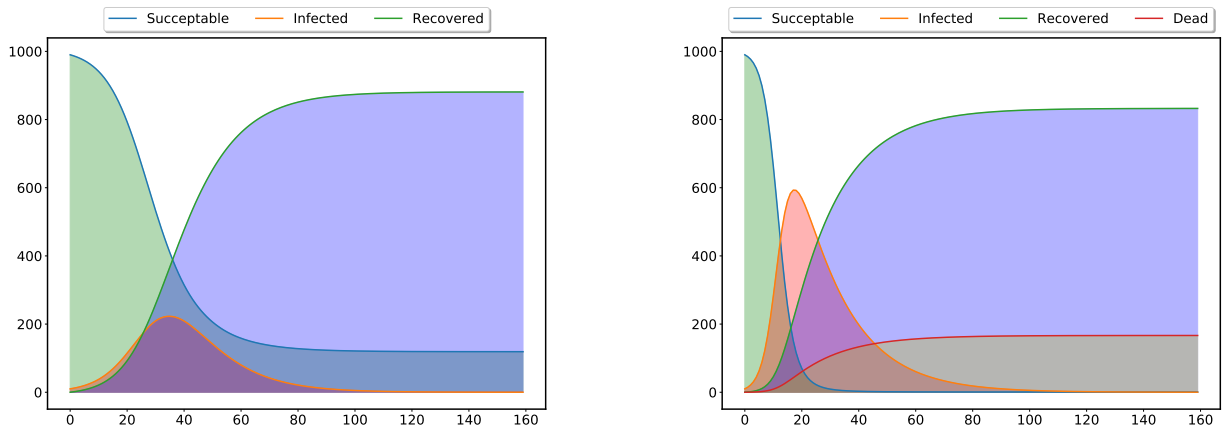

Figure 2: The left panel in the figure shows the standard SIR model with three type of population (compartments) - Susceptible, Infected and Recovered and the right panel shows the SIRD model with four type of population - on extra compartment for the dead population. The dynamics of the epidemic depends on the transmission coefficients β, γ and δ and the starting values of the different types of population. The values of the parameters and the initial population in different compartments used in the figure are only for the purpose on an illustration.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

2.2 SEIR model

If we relax the assumption that the people who get the infection become infectious instantly and consider a latent period to the the onset of infectiousness there is a fraction of population (compartment) which has been exposed to the virus but will become infectious only after some latent period $1/\sigma$, then the model is called **S**usceptible-**E**xposed-**I**nfected-**R**ecovered (SEIR) model represented by the following set of equations [JR08]

$$\begin{aligned} \frac{dS}{dt} &= -\beta \frac{S}{N} I \\ \frac{dE}{dt} &= \beta \frac{S}{N} I - \sigma E \\ \frac{dI}{dt} &= \sigma E - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{5}$$

Note that if we combine the second and third equation above we get:

$$\frac{d(E+I)}{dt} = \beta \frac{S}{N} I - \gamma I = \beta \left(\frac{S}{N} - \frac{\gamma}{\beta} \right) I \tag{6}$$

From the above equation we can see that population in the **E** and **I** compartments together can grow with time only when the fraction of the susceptible population is greater than the inverse of the reproduction number :

$$\frac{S}{N} > \frac{1}{\mathfrak{R}} \quad \text{where} \quad \mathfrak{R} = \frac{\beta}{\gamma} \tag{7}$$

There are many forms of SEIR equations which are in common use (see [BC18, ea20p, ea20u, ea20b, P.20, ea20n]) however, equation (5) is the simplest one and does not include natural deaths. One of the common practices with the SEIR model has been to consider the incubation period $1/\sigma$ a constant, and estimates it from some other observations. The SEIR modal is quite complex as compared to the SIR model and we cannot find the number of exposed people exactly at time $t = 0$ for evolving the equations and so the approach used to define $\mathfrak{R}$ no longer works. Thanks to the new generation matrix models [BC18] it is still possible to write $\mathfrak{R}$ in a close form for this case also.

2.3 SIRD model

One of the serious drawbacks of the SIR model is that people who recover and who die are treated in the same way - there are no separate compartments for the dead and recovered people. This drawback can be addressed by separating the compartments for the dead and recovered population as is done in the SIRD model described with the following set of equations (for a detail discussion see [ea20a, Vil20]).

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

$$\begin{aligned}
\frac{dS}{dt} &= -\beta\frac{S}{N}I \\
\frac{dI}{dt} &= \beta\frac{S}{N}I - (\delta+\gamma)I \\
\frac{dR}{dt} &= \gamma I \\
\frac{dD}{dt} &= \delta I
\end{aligned} \tag{8}$$

Here a new transmission coefficient δ has been introduced which we can identify with the death rate. One of the advantages of the SIRD model is that it has three transmission coefficients β, γ and δ and we have the data for three time series $I(t), R(t)$ and $D(t)$ available so it is possible to compute the time dependency of all the three coefficients as well as the reproduction number $\mathfrak{R}$.

The aim of any mitigation measures may be one or more of the followings:

1. Lower the contact or infection rate β.
2. Lower the mortality rate δ.
3. Increase the recovery rate γ.

The SIRD model provides us a framework to estimate or fit all these parameters. In one of the coming sections we will discuss how we can reconstruct the transmission coefficients β, δ and γ as well as $\mathfrak{R}$ from the data by a direct reconstruction approach.

The basic reproduction rate for the SIRD model can be written in the following way [ea20a]:

$$\mathfrak{R} = \frac{\beta}{\gamma+\delta}, \tag{9}$$

or,

$$\frac{1}{\mathfrak{R}} = \frac{1}{\mathfrak{R}_\gamma} + \frac{1}{\mathfrak{R}_\delta}, \tag{10}$$

where $\mathfrak{R}_\gamma = \beta/\gamma$ and $\mathfrak{R}_\delta = \beta/\delta$. If apart from death and recovery there is some other channel that can lower the population in the **I** compartment, for example if infected people move out from that region with transmission coefficient η then we can write:

$$\frac{1}{\mathfrak{R}} = \frac{1}{\mathfrak{R}_\gamma} + \frac{1}{\mathfrak{R}_\delta} + \frac{1}{\mathfrak{R}_\eta}, \tag{11}$$

with $\mathfrak{R}_\eta = \beta/\eta$. A more realistic model will have multiple compartments (nodes), either connected in series or some branching out from others, with data to constrain the transmission coefficients (edges). Apart from this, realistic models may also require to consider different transmission coefficients of different subgroups (based on age etc.,). Incorporating, all these considerations will lead to very complex models having very less connect with the actual data we have.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

2.4 Time dependent β models

As a pandemic triggers various containment measures [ea20s, DG20, ea20k, SNC20, RR20, ea20c] such as lockdown, social distancing, improved hygiene practices etc., are taken and that lead to transmission coefficients such as β becoming time dependent [GPD20, ea20m, FP20, ea20e, ea20i]. Apart from this, the drop in the susceptible population also decreases β (see [ea03a, Cob20]).

Lockdown has been one of the most common mitigation measures followed all over the world and, in its extreme form, we can assume that once it starts the contact rate between susceptible and infectious people drops to zero. In general, the lockdown starts on a fixed day t_l and has a duration (time scale) we call τ (we will be using both τ and corresponding decay rate $\mu = 1/\tau$ in the discussion). We can incorporate these two parameters into the modeling of $\beta(t)$ in many different ways and a set of three common choices is given below:

1. **Polynomial Suppression** [ea20m]:

$$\beta(t) = \begin{cases} \beta_0, & \text{if} \quad t \leq t_l \\ \beta_0 \left[1 - \mu \left(\frac{t-t_l}{t}\right)\right], & \text{otherwise} \end{cases} \tag{12}$$

   This model is discussed here just for an example and we do not expect the variation of $\beta(t)$ as slow as linear one. This expression shows that $\beta(t)$ starts with an initial value β_0 and after time t_l it starts decreasing linearly with a constant rate of $\mu = 1/\tau$ and finally becomes $\beta_0(1-\mu)$ at $t = \infty$.

2. tanh **Suppression**:

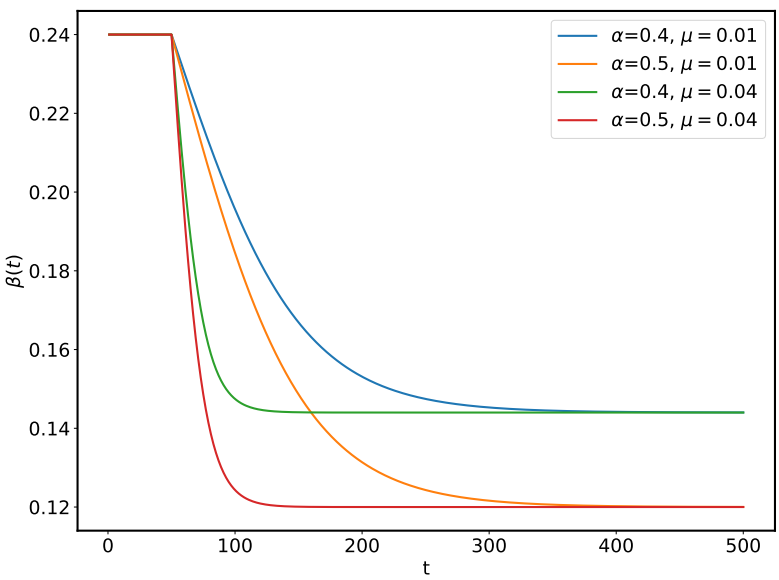

Figure 3: This figure shows the effect of α and μ on $\beta(t)$ for the **tanh** model.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

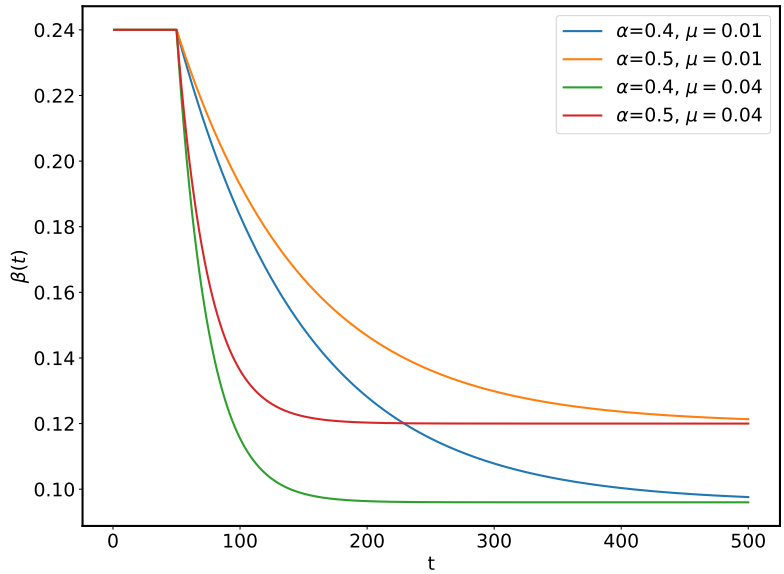

Figure 4: This figure shows the effect of α and μ on $\beta(t)$ for the **exp** model.

$$\beta(t) = \begin{cases} \beta_0, & \text{if} \quad t \leq t_l \\ \beta_0 \left[1 - \alpha * \tanh\left(\frac{t-t_l}{\tau}\right)\right], & \text{otherwise} \end{cases} \tag{13}$$

This form of suppression of $\beta(t)$ starts with a constant value β_0 at some $t = t_l$ and keeps decaying for period represented by τ and finally settles to a final value $\beta_0(1-\alpha)$ as is shown in Figure (3). This can be written in the following way also:

$$\alpha = 1 - \frac{\beta_{\text{min}}}{\beta_{\text{max}}} \tag{14}$$

From equation (13) we can also write :

$$\mu = \frac{1}{\tau} = \frac{1}{t - t_l} \operatorname{arctanh}\left[\frac{1}{\alpha}\left(1 - \frac{\beta(t)}{\beta_0}\right)\right] \tag{15}$$

Equations (14) and equation (15) are important to find the priors for α and μ once we know the priors for β and this will be discussed again in §4 and will be used in parameter estimation in §5.

3. **Exponential Suppression** [FP20, ea20o, ea20n]:

$$\beta(t) = \begin{cases} \beta_0, & \text{if} \quad t \leq t_l \\ \beta_0 \left[(1-\alpha)\exp\left(\frac{-(t-t_l)}{\tau}\right) + \alpha\right], & \text{otherwise} \end{cases} \tag{16}$$

This model is similar to the tanh model and in this case also $\beta(t)$ starts from some initial value β_0 and after decreasing for a period and finally approaches to a constant value $\beta_0\alpha$ at $t = \infty$ as

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

is shown in Figure (4). Note that the transmission coefficient β may decay with time without any intervention also as is discussed in [ea97] for plants.

In this case also we can write :

$$\alpha = \frac{\beta_{\min}}{\beta_{\max}}, \tag{17}$$

and

$$\mu = \frac{1}{\tau} = -\frac{1}{t - t_l} \log\left[\frac{1}{1-\alpha}\left(\frac{\beta(t)}{\beta_0} - \alpha\right)\right] \tag{18}$$

These equations also will be used to find the priors for parameter estimation.

In one of the studies [ea20o] it has been argued that even the time of recovery $1/\gamma$ may also vary with time due to the improvement in medical understanding of the epidemic and facilities and that also can be modeled as an exponential function. There have been other physically well motivated exponentially decaying forms also such as given in [Vil20] in which β starts from starting value β_0 and decay with rate $1/\tau$ finally becomes β_1.

$$\beta(t) = \beta_0(t)\exp(-t/\tau) + \beta_1(1 - \exp(-t/\tau)). \tag{19}$$

The author argues that β_1 depends on the policy decisions leading to behavioral changes. This model is different from the model we are considering only in the respect that it considers the "lockdown" from the beginning i.e., $t = 0$.

The time dependent β models as are discussed above and shown in Figure (5) share a common property that before a certain time t_l, that we can identify with the day on which lockdown starts, β has a constant value β_0 and after that it starts decreasing with a rate that depend on the parameter $\mu = 1/\tau$. The effect of the suppression in β is controlled by the parameter μ and for its zeros values all the models become constant β models. From Figure (5) we can conclude that different models can lead to the same amounts of "flattening" of the curve with a different choice of parameters so there is no preferred model for the suppression.

The SIR model with constant transmission coefficients is applicable only in the situation when the pandemic is let to grow without any intervention. In the real world once a pandemic starts interventions of different kinds (social, medical etc.,) are considered to reduce the rate at which the the epidemic spreads. These interventions can be easily taken into account by considering a time dependent (decaying) growth rate (β). As we can see from the above figure that a decaying (exponentially) β helps to contain the disease by lowering the height of the peak as shown in Figure (5).

3 Data

The primary data for Covid-19 is in terms of three times series for the count of confirmed $\mathcal{C}(t)$, recovered $\mathcal{R}(t)$ and dead $\mathcal{D}(t)$, persons for every country. By definition all the three times series

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

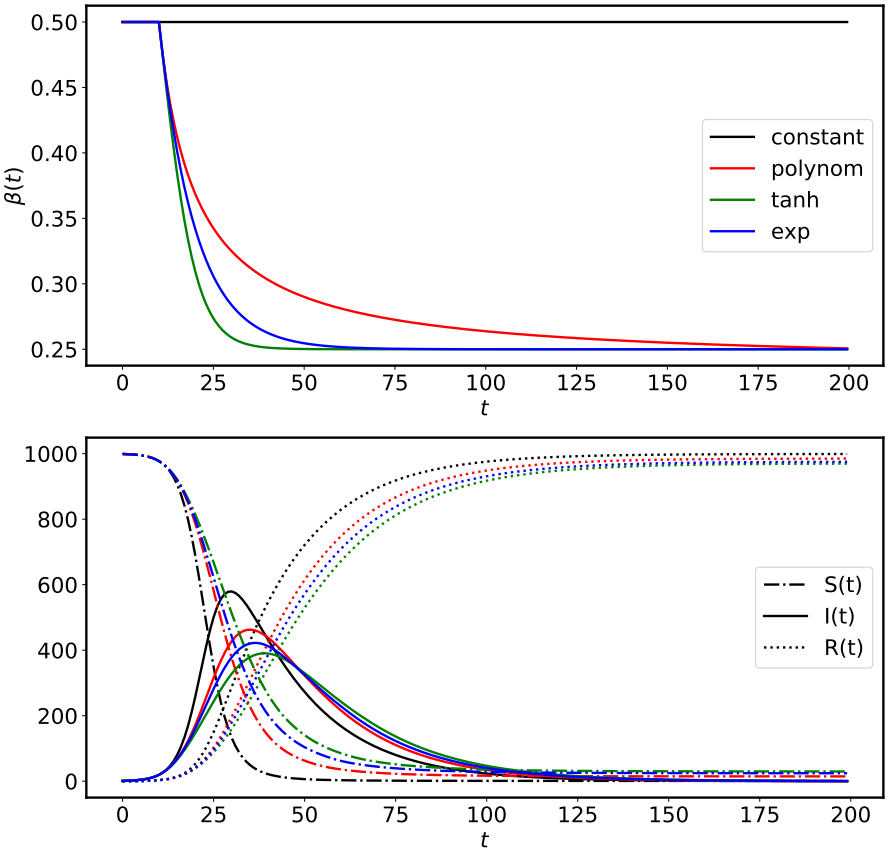

Figure 5: The top panel in the figure shows the β suppression models as are discussed in the text. The solid and dashed line in the bottom panel show the number of infected people $I(t)$ and the number of recovered people $R(t)$ for the SIR model corresponding to different β models. From the bottom panel we can see that all the models have similar effects, however, for the tanh case the suppression in the peak of $I(t)$ is maximum. This figure is for an illustration the values of parameters have been chosen carefully to highlight the effects.

are non-decreasing functions of time, as are shown in Figure (6) for Italy. The data for Covid-19 is provided by the John Hopkins University [ea20g, Gar20a, Gar20b] and is updated on a daily basis and interactive tools are also provided for data exploration. The worldometer website [Wor20] and **our world in data** [iD20] also provide an up to-date data for most countries with some extra information, such as the numbers for the active cases, critical cases and the test conducted for a million population of the country.

The time series which are shown in Figure (6) grow exponentially in the beginning and then settle to a slower power law growth. If we look at the daily new cases, as shown in Figure (7), they look like broad peaked functions with increasing fluctuations around the the current value. Time series for a set of countries used in the analysis are shown in the Figure (8).

There are many factors, known and unknown, which determine the behavior of these time series,

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

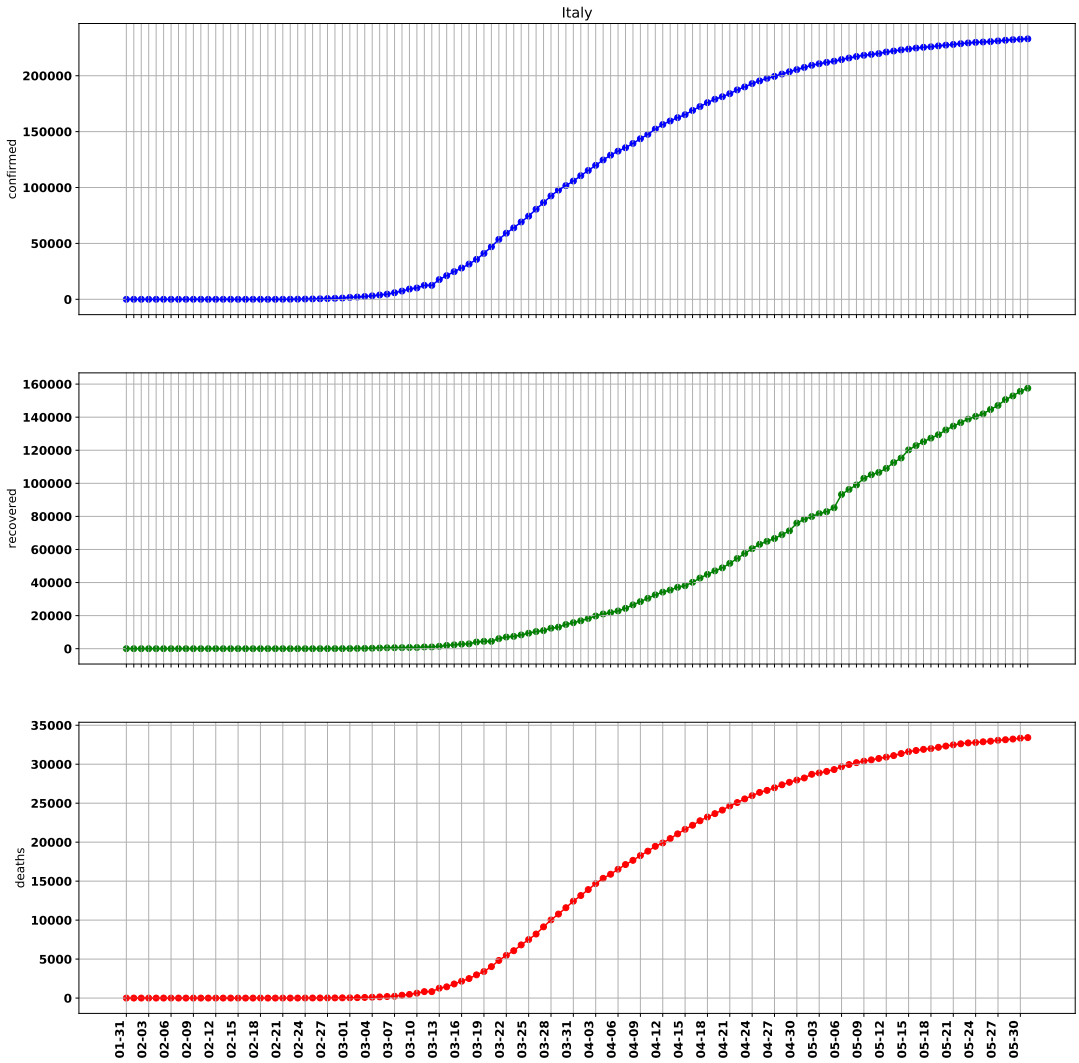

Figure 6: The total number of confirmed, recovered and dead case for Italy.

such as the nature of the diseases/virus, the health profile of the population, availability of medical infrastructure, age-distribution, social mixing, personal hygiene and may be geographical location and genetic makeup of the population etc. Compartmental models, such as SIR, SIRD, SEIR etc., help to simplify the problem by replacing a large number of controlling parameters with a very small set of well motivated parameters - the transmission coefficients.

The compartmental models predict that how the population in different compartments change with time. In the SIR model the population in **S** and **R** compartments can only decrease and increase respectively, however, in the **I** compartment it can increases as well as decreases. On the onset of the epidemic almost everyone is in the **S** compartment with a very small fraction in **I** and no one in the **R** compartment. At the end of the pandemic everyone is in the **R** compartment with no one in **S**

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

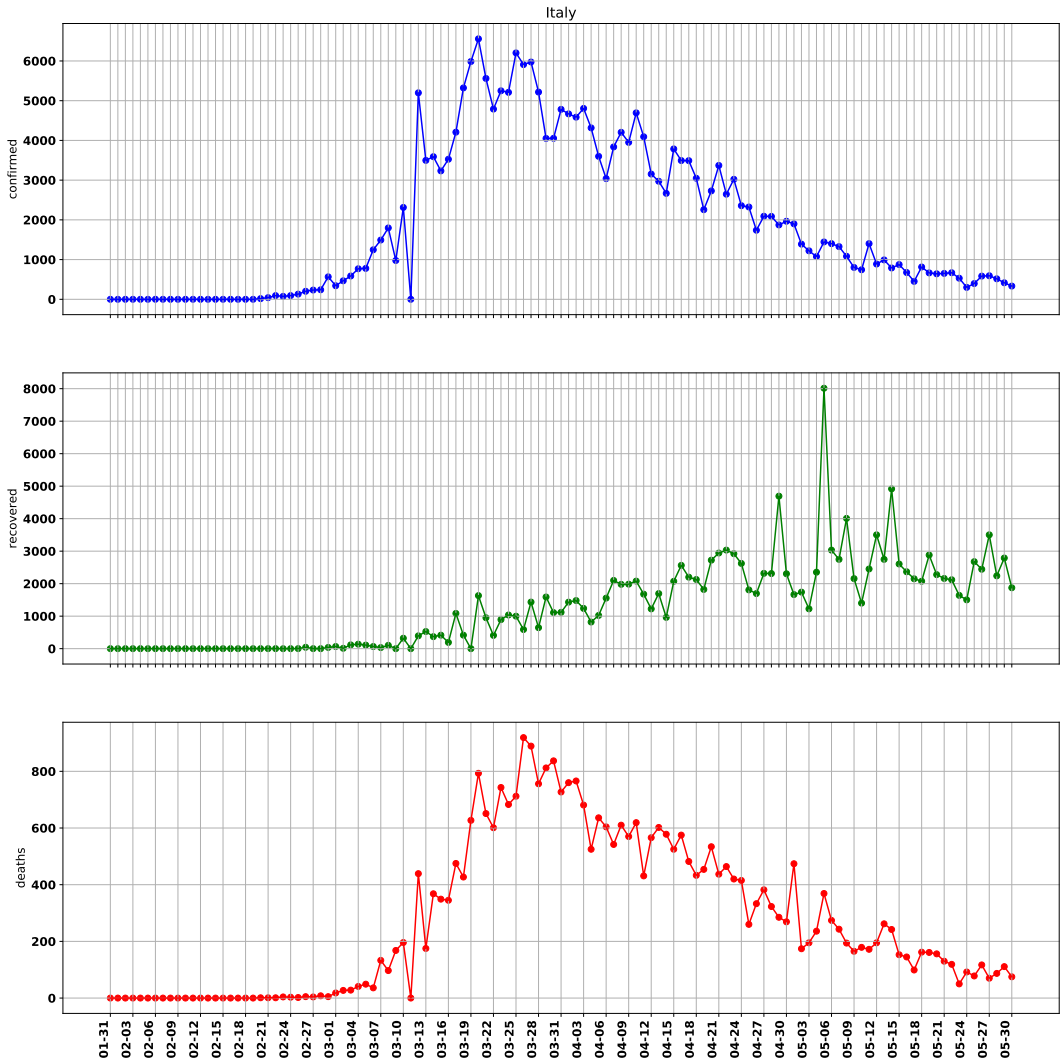

Figure 7: The number of new confirmed, recovered and dead case for Italy.

or **I** compartment (see Figure (1). The same happens for SIRD model also where the decrease in **I** compartment happens due to recovery (R) and deaths (D).

The time series $I(t)$ for the population in compartment **I** can be obtained by subtracting $\mathcal{R}$ and $\mathcal{D}$ from $\mathcal{C}$.

$$I(t) = \mathcal{C}(t) - \mathcal{R}(t) - \mathcal{D}(t) \tag{20}$$

For a set of countries the time series of $I(t)$ are shown in Figure (8). The time series $\mathcal{S}, \mathcal{I}, \mathcal{R}$ and $\mathcal{D}$ have very weak dependency on each other - the only constrains they have to satisfy is that the sum total of the population in different compartment must add to the total population. The number of people $I(t)$ in the **I** compartment at time t does depend on all the three transmission rates β, γ, and

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

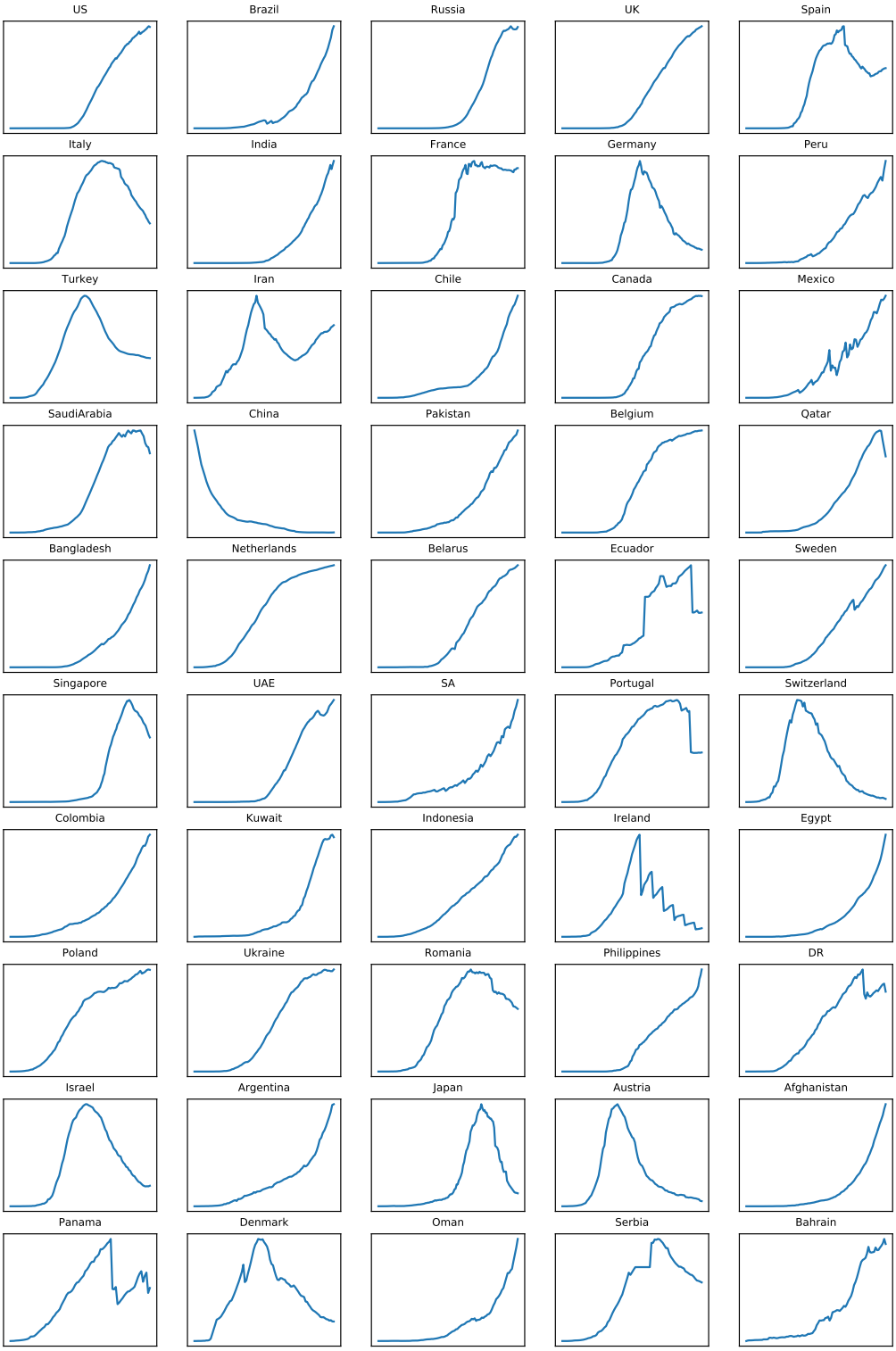

Figure 8: Time series for a set of countries we used for modeling.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

S. No	country	starting_date	num_days	population
1	**US**	2020-02-21	101	331,002,651
2	**Brazil**	2020-03-10	83	212,559,417
3	**Russia**	2020-03-13	80	145,934,462
4	**UK**	2020-03-01	92	67,886,011
5	**Spain**	2020-02-28	94	46,754,778
6	**Italy**	2020-02-22	100	60,461,826
7	**India**	2020-03-05	88	1,380,004,385
8	**France**	2020-02-28	94	65,273,511
9	**Germany**	2020-02-27	95	83,783,942
10	**Peru**	2020-03-13	80	32,971,854
11	**Turkey**	2020-03-17	76	84,339,067
12	**Iran**	2020-02-23	98	83,992,949
13	**Chile**	2020-03-13	80	19,116,201
14	**Canada**	2020-03-04	89	37,742,154
15	**Mexico**	2020-03-15	78	128,932,753
16	**SaudiArabia**	2020-03-12	81	34,813,871
17	**China**	2020-03-11	82	1,439,323,776
18	**Pakistan**	2020-03-13	80	220,892,340
19	**Belgium**	2020-03-05	88	11,589,623
20	**Qatar**	2020-03-11	82	2,881,053
21	**Bangladesh**	2020-03-24	65	164,689,383
22	**Netherlands**	2020-03-04	89	17,134,872
23	**Belarus**	2020-03-16	77	9,449,323
24	**Ecuador**	2020-03-14	79	17,643,054
25	**Sweden**	2020-03-04	89	10,099,265
26	**Singapore**	2020-02-05	117	5,850,342
27	**UAE**	2020-03-07	86	9,890,402
28	**SA**	2020-03-14	79	59,308,690
29	**Portugal**	2020-03-08	85	10,196,709
30	**Switzerland**	2020-03-01	92	8,654,622
31	**Colombia**	2020-03-15	78	50,882,891
32	**Kuwait**	2020-02-26	96	4,270,571
33	**Indonesia**	2020-03-11	82	273,523,615
34	**Ireland**	2020-03-10	83	4,937,786
35	**Egypt**	2020-03-08	85	102,334,404
36	**Poland**	2020-03-11	82	37,846,611
37	**Ukraine**	2020-03-20	73	43,733,762
38	**Romania**	2020-03-11	82	19,237,691
39	**Philippines**	2020-03-10	83	109,581,078
40	**DR**	2020-03-19	74	10,847,910
41	**Israel**	2020-03-08	85	8,655,535
42	**Argentina**	2020-03-13	80	45,195,774
43	**Japan**	2020-02-06	108	126,476,461
44	**Austria**	2020-03-04	89	9,006,398
45	**Afghanistan**	2020-03-22	71	38,928,346
46	**Panama**	2020-03-13	80	4,314,767
47	**Denmark**	2020-03-08	85	5,792,202
48	**Oman**	2020-03-18	75	5,106,626
49	**Serbia**	2020-03-13	80	8,737,371
50	**Bahrain**	2020-02-26	96	1,701,575

Table 2: A summary for the countries used for modelling

δ, therefore it is a good measure which we can fit to a compartmental model, such as SIR or SIRD and can get constraints.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

4 Reconstruction

In this and next section we present the main results of the study in the form of demonstrating a reconstruction procedure for the time dependent transmission coefficients $\beta(t)$, $\gamma(t)$, $\delta(t)$ and the effective reproduction number $\mathfrak{R}(t)$. We consider an example of Italy for this procedure. Note that this approach is common and can be used to understand the variation of the transmission coefficients with time as a result of interventions. The main advantages of this approach is that there are no parameters to adjust and so the results are easy to reproduce.

The approach we use here is similar to as used in [ea20e, GPD20]. In this approach the evolution equations are written in a discretized form as shown in equation (21).

$$\begin{aligned} S_{t+1} &= S_t - \beta(t)\frac{S_t}{N}I_t \\ I_{t+1} &= I_t + \beta(t)\frac{S_t}{N}I_t - \gamma(t)I_t \\ R_{t+1} &= R_t + \gamma(t)I_t \end{aligned} \tag{21}$$

From the third equation we can write:

$$\gamma(t) = \frac{R_{t+1} - R_t}{I_t}, \tag{22}$$

and using this and second equation from (21) we get,

$$\beta(t) = \frac{(I_{t+1} - I_t) + (R_{t+1} - R_t)}{I_t}. \tag{23}$$

Note that by definition $R_{t+1} \geq R_t$, so $\gamma(t) \geq 0$, however, we may have $I_{t+1} \leq I_t$ also, $\beta(t)$ may become negative also once the population in the compartment **I** starts decreasing.

Here an important assumption is being made and that is the fraction of susceptible population S/N is close to unity which may be true at the beginning of the epidemic. Once we have expressions for the time dependent β and γ we can also written an expression for the time dependent reproduction number in the following way:

$$\mathfrak{R}_{SIR}(t) = \frac{\beta(t)}{\gamma(t)} = \frac{(I_{t+1} - I_t) + (R_{t+1} - R_t)}{R_{t+1} - R_t} = 1 + \frac{I_{t+1} - I_t}{R_{t+1} - R_t} \tag{24}$$

Following the similar procedure we can write the SIRD equations in the following discretized form:

$$\begin{aligned} S_{t+1} &= S_t - \beta(t)\frac{S_t}{N}I_t \\ I_{t+1} &= I_t + \beta(t)\frac{S_t}{N}I_t - (\gamma(t) + \delta(t))I_t \\ R_{t+1} &= R_t + \gamma(t)I_t \\ D_{t+1} &= D_t + \delta(t)I_t \end{aligned} \tag{25}$$

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

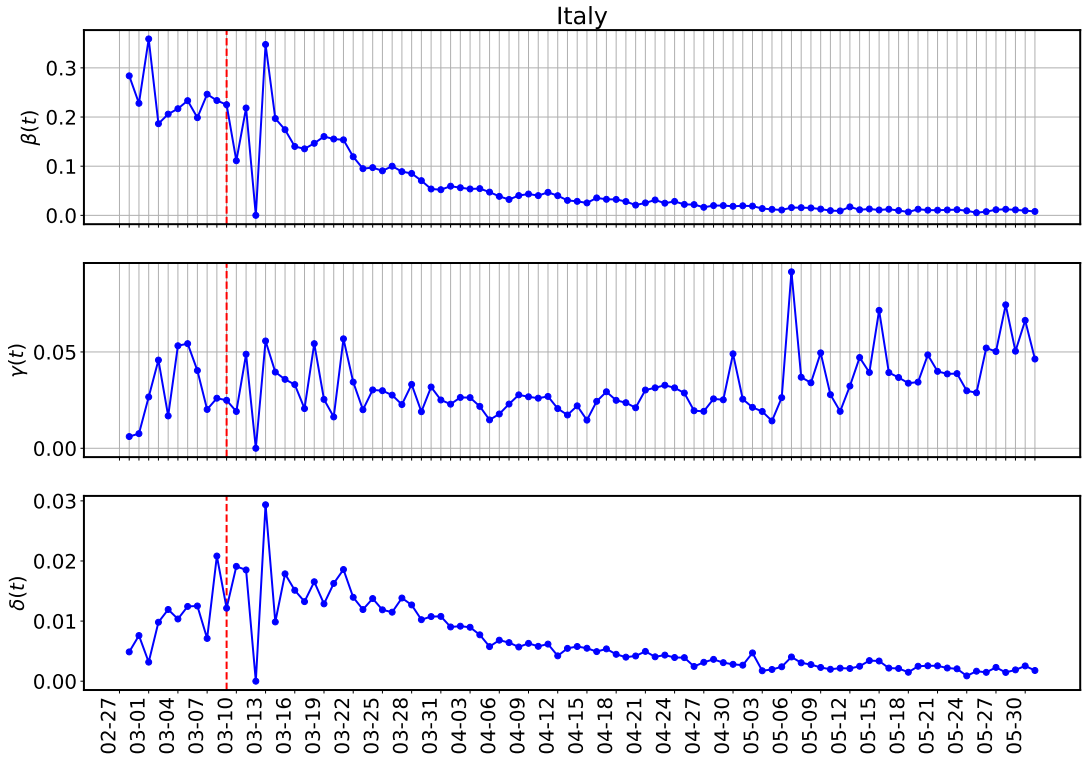

Figure 9: Reconstruction of the transmission coefficients $\beta(t)$, $\gamma(t)$ and $\delta(t)$ with the SIRD model for Italy.

From these questions we can write :

$$\begin{aligned}
\beta(t) &= \frac{(I_{t+1} - I_t) + [(R_{t+1} - R_t) + (D_{t+1} - D_t)]}{I_t} \\
\gamma(t) &= \frac{R_{t+1} - R_t}{I_t} \\
\delta(t) &= \frac{D_{t+1} - D_t}{I_t},
\end{aligned} \tag{26}$$

and can write the expression for the reproduction number:

$$\mathfrak{R}_{SIRD}(t) = \frac{\beta}{\gamma + \delta} = \frac{\Delta I_t + \Delta R_t + \Delta D_t}{\Delta R_t + \Delta D_t} = 1 + \frac{\Delta I_t}{\Delta R_t + \Delta D_t}, \tag{27}$$

where $\Delta X_t = X_{t+1} - X_t$ with $X = I, R$ and D. This equation is identical to equation (24) if we do not count dead and recovered separately i.e., replace $\Delta R_t + \Delta D_t$ with ΔR_t. One of the interpretations of R is that it is a ratio of two rates and so in case we are interested finding out two separates measures for γ and δ, we can also write :

$$\mathfrak{R}_\gamma = \frac{\beta}{\gamma} \quad \text{and} \quad \mathfrak{R}_\delta = \frac{\beta}{\delta}, \tag{28}$$

and so,

$$\frac{1}{\mathfrak{R}} = \frac{1}{\mathfrak{R}_\gamma} + \frac{1}{\mathfrak{R}_\delta} \tag{29}$$

The procedure as discussed above can be used to know the variation of the transmission coefficients β, γ, δ and effective reproduction $\mathfrak{R}(t)$with time. In order to follow this procedure we need to abandon

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

first few data points which have very high noise. As explained above occasionally we may also have negative values of $\mathfrak{R}(t)$.

In figure (9) we show the reconstruction for $\beta(t), \gamma(t)$ and $\delta(t)$ for Italy with SIRD. From this figure it is clear that all the three transmission coefficients vary with time and the variation of $\beta(t)$ is maximum. The same trend is observed for other countries also. In the next section will fit a parameterized form of $\beta(t)$ which we obtain from the model-fitting to the data shown in the figure.

Once we have reconstructed $\beta(t), \gamma(t)$ and $\delta(t)$ we can easily get $\mathfrak{R}(t)$for SIR and SIRD model.

4.1 Priors

One of the import uses of the reconstruction procedure we have discussed here is to find the priors (minimum, maximum and best fit) values for the parameters to be fitted. Once we have the estimates for $\beta(t), \gamma(t)$ and $\delta(t)$ from the above procedure we can easily find $X_{\min}, X_{\max}, X_0$, values (with $X = \beta, \gamma, \delta$. Here, X_0 is the approximate point for the parameter that is needed in many optimization procedure which iteratively find the solution. Since in the present work we use parametric form of $\beta(t)$, so we need priors for the parameters of $\beta(t)$ i.e., β_1, α, μ and τ which can find from the reconstructed $\beta(t)$ (see §2.4 for detail).

5 Model fitting and parameter estimation

We consider a set of six compartmental models, three belonging to the SIR and three to the SIRD class. The models are different from each other in terms of the choice for the epidemiological class (SIR or SIRD) or the model for the contact rate $\beta(t)$ (see §2.4 for detail). A summary of the models is given in Table (3).

Model	**Epidemiological model**	$\beta(t)$ **Model**	**Parameters**
1	SIR	exp	$(\beta, \gamma), (\beta_0, \alpha, \mu, t_l)$
2	SIR	tanh	$(\beta, \gamma), (\beta_0, \alpha, \mu, t_l)$
3	SIRD	exp	$(\beta, \gamma, \delta), (\beta_0, \alpha, \mu, t_l)$
4	SIRD	tanh	$(\beta, \gamma, \delta), (\beta_0, \alpha, \mu, t_l)$
5	SIR	exp	$(\beta, \gamma), (\beta_0, \mu)$
6	SIRD	exp	$(\beta, \gamma, \delta), (\beta_0, \mu)$

Table 3: A summary of the models being considered for the analysis

Note that in the model (5) and (6), $\beta(t)$ starts decaying from the very beginning (in place of starting from a particular day representing the date of the lockdown) with a constant rate μ.

In any fitting procedure the choice of the loss function depends on what we wish to fit. In the

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Variable	Name	Description
$\mathbb{D}$	Primary Data	$\mathcal{C}, \mathcal{R}, \mathcal{D}$
θ	Fitting parameters	As defined in the last column of Table (3)
X_l^o	Fitting data	$I, R+D$ for SIR and $I,\ R,\ D$ for SIRD
X_l^p	Predicted data	prediction from the model
n_t	Days	Number of days for which we have data
n_d	Number of time series used	2 for SIR and 3 for SIRD
c	country	label for the country
w_l	weight	Used for fitting multiple time series together

Table 4: A summary of the variables

common least square fitting we use the sum of the squares of the offsets as the loss function. However, there is a problem here with the data we have for that choice. The time series we wish to fit have small values at the beginning and very large values at the later stage, so the the fitting is biased towards the points which have large values. One of the solutions for this could be to fit the log of the time series but then the fitting becomes biased toward small values, in the beginning (or later stage when the peaks falls).

We decide to use the loss function of the ordinary least square which fits the data points close to the peak (having higher values) more accurately than other data points. We found this useful for the following two reasons:

1. The peak in the time series is an important feature, in particular its location and height, therefore any loss function biased towards it is justified.

2. For the short term predictions only the data points close to the dates of prediction is important, so using a loss function that fits later points (having higher values) more accurately than the noisy data points in the beginning is favorable.

The loss function which we used for fitting the data to SIR and SIR models is given below.

$$\mathcal{L}^c(\theta|\mathbb{D}) = \frac{1}{n_d}\sum_{l=1}^{l=n_d}\frac{w_l}{n_t}\sum_{i=1}^{i=n_t}(X_{l,i}^o - X_{l,i}^p)^2 \tag{30}$$

The variables used in the above equations are defined in table (4):

Once we have estimated the fitting parameters of our model we can find a smooth representation of $\beta(t)$ using the best-fit values of the parameters β_0, α, μ and t_l as is shown in Figure 11 for Italy with SIR model and **exp** model for $\beta(t)$. The first few data points are very noise and so are not shown in the figure. The similar exercise is also done for $\mathfrak{R}(t)$ in Figure (12). One of the shortcomings for the case of $\mathfrak{R}(t)$ is that we have assumed that γ as a constant although the data does show the variation

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

of γ and δ with time (see Figure (9). We have also shown the beta model fitting for India and US in the Figure (13) and (14).

For SIR and SIRD models we fit multiple time series together so we must weight the sum of the squares of the offsets for different time series since they have very different values - the value of $I(t)$ is generally few orders of magnitude higher than $R(t)$ and $D(t)$. We use the following weights for this purpose:

$$w_l = 1 - \frac{\bar{X}_l}{\sum_l \bar{X}_l}, \tag{31}$$

where $\bar{X}_l$ is the average of the time series $X(t)$.

We use the **solve_ivp** and **minimize** modules from the `Scipy` [Sci20] for integrating the differential equations and minimize the cost function respectively. The loss function given by equation (30) represents the Root Mean Square Deviation (RMSD) and we use its final value as a measure of the goodness of the fit.

A list of fitting parameter for the different models is given in Table (3). For the SIR class of models, model (1) and (2) we have five fitting parameters named, $\gamma, \beta_0, \alpha, \mu, t_l$ and for the SIRD models, model (3) and (4) we have six fitting parameters named $\gamma, \beta_0, \alpha, \mu, t_l$ and δ. As we can notice that for model (1) to model (4) four of the parameters are associated with $\beta(t)$ and for model (5) and (6) the variation of $\beta(t)$ is controlled by just two parameter - β_0, the initial value of β and its decay rate $\mu = 1/\tau$.

The best fit values of the fitting parameters with their 90 % CI (standard deviation) as well as the median values are given Table (5). The tables also give the estimate for the effective reproduction number $\mathfrak{R}(t)$which is a derived quantity here. Note that for computing $\mathfrak{R}_0$we have extrapolated the value of $\mathfrak{R}(t)$to the last date for which the data is being used here.

A histogram of the effective reproduction number for the different models being considered is shown in Figure (10) and detail values of that for different countries, which include the average values as well as 90% CI (StdDev), are given in Table (6).

6 Summary and conclusions

Covid-19 is a global crisis and understanding its impact on different systems of the modern human life (medical, social, economical etc.,) and the responses presented is an important exercise to carry out. We understand that despite being a global phenomenon, the impact of Covid-19 in terms of the loss to life and the resourced being exhausted depend on the local conditions as well as on the mitigation measures taken locally. However, we believe that the global picture of the crisis does help to plan and take policy decisions at the local scale also.

Full understanding of any pandemic, in particular like Covid-19 which does not have any other examples in the history (in terms of the scale and impact), may become available only when it is over

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Models	Parametes						
	S.No.	**Parameter**	**Aver**	**Std Dev**	**Median**	**Min**	**Max**
Model (1)	1	γ	0.0487	0.0659	0.0335	0.0050	0.4440
	2	β_0	0.2528	0.1145	0.2440	0.1000	0.7040
	3	α	0.1900	0.1858	0.1000	0.1000	1.0000
	4	μ	0.0808	0.0279	0.1000	0.0010	0.1000
	5	t_l	21.4087	18.4903	17.4585	0.0000	81.9410
	6	$\mathfrak{R}(t)$	1.6141	1.3228	1.1575	0.3520	7.4700
Model (2)	1	γ	0.0386	0.0335	0.0305	0.0040	0.1870
	2	β_0	0.2784	0.1176	0.2640	0.1000	0.5570
	3	α	0.8342	0.2266	0.9140	0.1000	1.0000
	4	μ	0.0582	0.0307	0.0495	0.0010	0.1000
	5	t_l	17.2522	20.4181	6.3750	0.0000	73.0000
	6	$\mathfrak{R}(t)$	1.5368	1.4330	1.1935	0.0000	7.7130
Model (3)	1	δ	0.0049	0.0093	0.0030	0.0010	0.0670
	2	γ	0.0357	0.0246	0.0290	0.0010	0.1000
	3	β_0	0.2551	0.0847	0.2555	0.1000	0.4390
	4	α	0.1610	0.1264	0.1000	0.1000	0.8790
	5	μ	0.0804	0.0275	0.1000	0.0070	0.1000
	6	t_l	15.4674	14.6679	12.8050	0.0000	61.7040
	7	$\mathfrak{R}(t)$	1.6448	1.3119	1.2010	0.0950	7.5620
Model (4)	1	δ	0.0051	0.0094	0.0030	0.0010	0.0640
	2	γ	0.0361	0.0246	0.0315	0.0010	0.1000
	3	β_0	0.2685	0.1044	0.2555	0.1000	0.4710
	4	α	0.8606	0.1831	0.9425	0.1130	1.0000
	5	μ	0.0626	0.0300	0.0565	0.0030	0.1000
	6	t_l	16.4373	20.2824	8.3325	0.0000	95.0790
	7	$\mathfrak{R}(t)$	1.3892	1.4310	1.1750	0.0000	7.8070
Model (5)	1	γ	0.0304	0.0214	0.0255	0.0010	0.0950
	2	β_0	0.3432	0.1971	0.2970	0.1000	1.0000
	3	μ	0.1165	0.2580	0.0360	0.0040	1.0000
	4	$\mathfrak{R}(t)$	0.9590	0.8869	0.8290	0.0000	3.4780
Model (6)	1	δ	0.0036	0.0039	0.0030	0.0010	0.0270
	2	γ	0.0321	0.0216	0.0275	0.0010	0.1000
	3	β_0	0.3236	0.1586	0.2890	0.1000	0.6750
	4	μ	0.0347	0.0232	0.0335	0.0030	0.1000
	5	$\mathfrak{R}(t)$	1.1612	1.1458	0.9470	0.0000	5.9240

Table 5: A summary of the fitting parameters for all the models.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

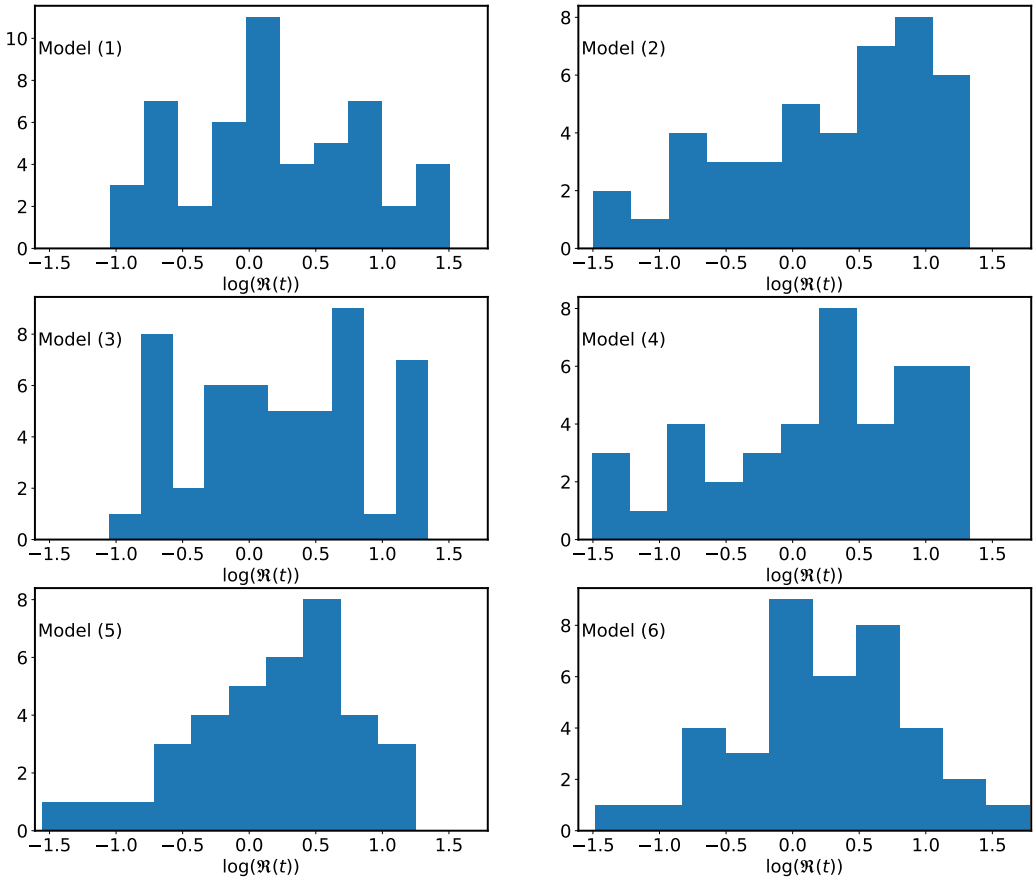

Figure 10: Histogram of the effective reproduction number $\mathfrak{R}(t)$with different models.

and the facts and figure presented here may have very short life. However, we still believe that any quick timely insight may help a lot in terms of the planning for the worse. Knowing very well that all the mathematical models are wrong but some are useful, we believe that mathematical models which are presented in this work may help to develop some insight about the crisis. A brief summary of the work presented here is as follows.

In §1 we have given a very brief introduction of the problem being addressed and reviewed some of the key works about Covid-19 which motivated the present work. A brief introduction of the mathematical framework used in the work in §2, in particular we have review a set of compartmental models SIR, SEIR and SIRD in §2. We have also discussed a set of of parametric models for one of the transmission coefficients $\beta(t)$, in §2.4. We have discussed the data being used in the work in §3.

The main results of the present work are discussed in §4 and §5. In §4 we have reviewed a reconstruction procedure for the transmission coefficients and basic reproduction number $\mathfrak{R}(t)$. This procedure does not depend on the choice of any parameter and can be easily generalized for other similar models also. We have presented the best-fit values of the parameters with their 90% CI, §5 in

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

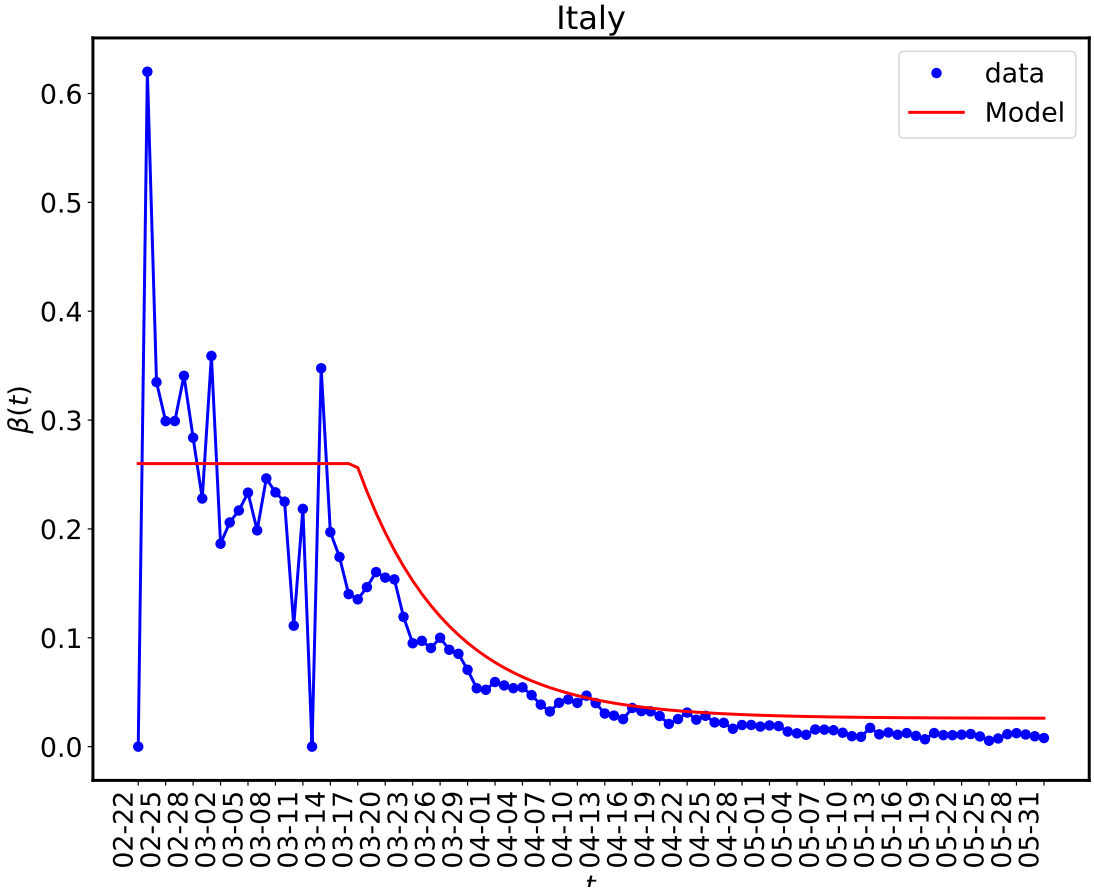

Figure 11: In this figure we have shown the reconstructed $\beta(t)$ and a smoother version which we have obtained by model fitting for the SIR case with **exp** model for $\beta(t)$. Similar exercise can be done for other countries also. Since the data points are very noise so it does not make much sense to use more complex model for $\beta(t)$ that what has been used here.

the form a set of tables. We have presented the values of the parameters in the following two forms:

1. Model based

2. Country based

All the fitting parameters for the models being considered are summarized in Table (5) and full lists of parameters for different countries with different models is given at [Pra20a]. Here we only give the values of the effective reproduction for the countries (see Table (2) on the last date for which we have the data in Table (6).

The work we presented here assumes that spreading of a pandemic like Covid-19 happens homogeneously in space and time, however, we know that it is far from true. As the experience [ea03b] shows that "super-spread" events (SSEs) or rare events where, one particular infectious person interacts with a very large number of susceptible people over a short period of time have the maximum impact. In these situation the average measures like $\mathfrak{R}_0$ are not very informative. In the present work we data

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

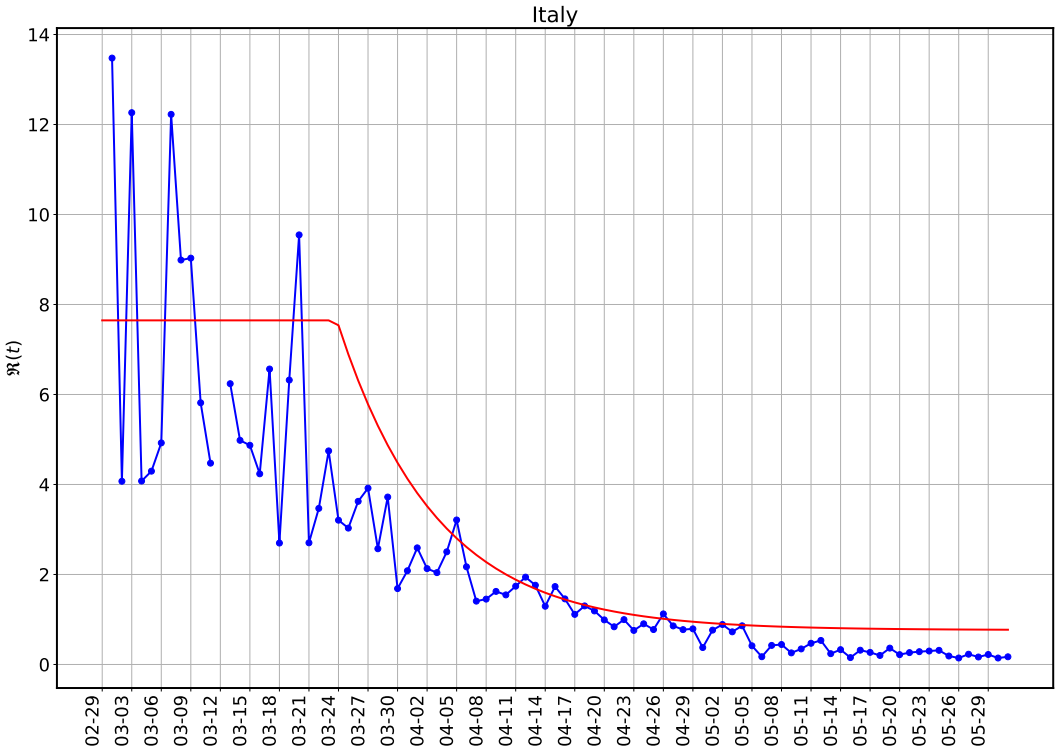

Figure 12: Reconstructucted effective reproduction number $\mathfrak{R}(t)$for the same set of parameters as are used in Figure (11)

for a set of countries to constrain the parameters of the SIR and SIRD model one similar exercise with SIRD model for India is done in [ea20q].

In the present work we have considered only time variation of β, however, from the data we can see that other transmission coefficients such as γ and δ also change with time, although not that much, mainly because they depend more on the nature of the disease and less on the mitigation and other social measures. Significant change in the values of γ and δ can take place only due to medical interventions.

Acknowledgment

The author would like to thank Dr. Gaurav Goswami for comments and feedback. At present the author works as an independent researcher and data scientist and the work presented here is not supported by any public or private agency. The author will be thankful to any agency, individual or individuals who come forward to sponsor/support this and other similar works on Covid-19.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

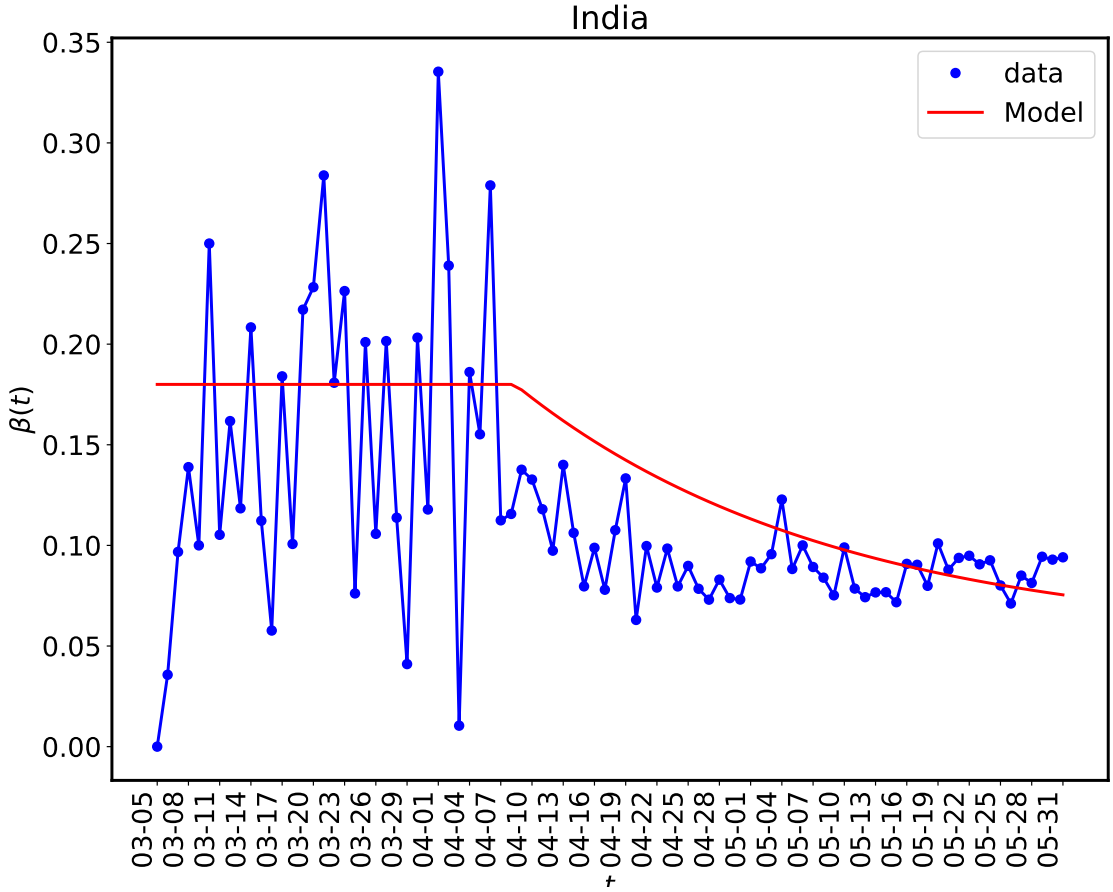

Figure 13: The same as in Figure (11) for India.

References

[BC18] Julie Blackwood and Lauren M. Childs. An introduction to compartmental modeling for the budding infectious disease modeler. *Letters in Biomathematics*, 5:1:195–221, 2018. doi:10.1080/23737867.2018.1509026.

[BS20] Kathakali Biswas and Parongama Sen. Space-time dependence of corona virus (covid-19) outbreak. *ArXiv e-prints*, 2020, 2003.03149.

[Cob20] Sarah Cobey. Modeling infectious disease dynamics. *Science*, 2020. doi:10.1126/science.abb5659 (2020).

[DG20] Emanuele Daddi and Mauro Giavalisco. Early forecasts of the evolution of the covid-19 outbreaks and quantitative assessment of the effectiveness of countering measures. *ArXiv e-prints*, 2020, 2004.08365.

[ea97] C. A. Gilligan et al. Analysis and fitting of an sir model with host response to infection load for a plant disease. *Philos Trans R Soc Lond B Biol Sci.*, 352(1351):353–364, 1997. doi:10.1098/rstb.1997.0026.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

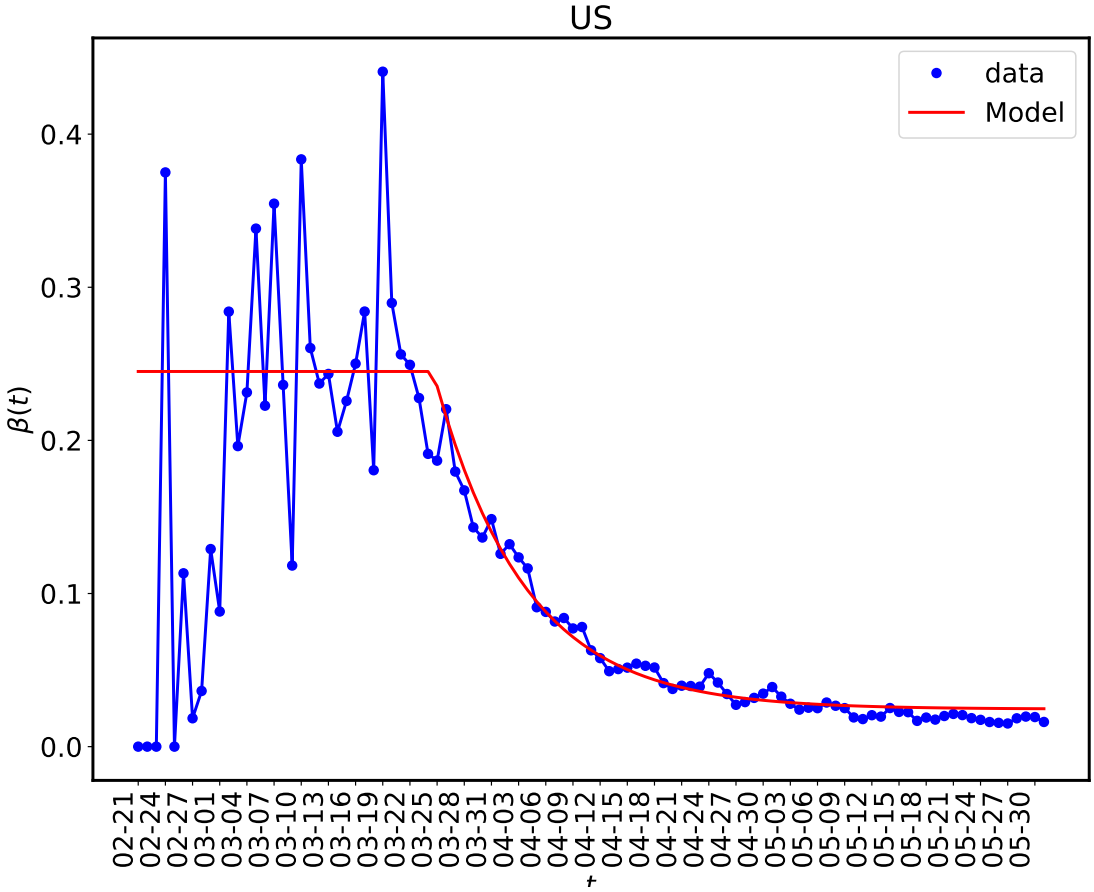

Figure 14: The same as in Figure (11) for US.

[ea03a] Lipsitch et al. Transmission dynamics and control of severe acute respiratory syndrome. *Science*, 300(5627):1966–1970, 2003. doi:10.1126/science.1086616.

[ea03b] Riley et al. Transmission dynamics of the etiological agent of sars in hong kong: Impact of public health interventions. *Science*, 300:1961–1966, 2003. doi:10.1126/science.1086478.

[ea05] Neil M. Ferguson et al. Strategies for containing an emerging influenza pandemic in southeast asia. *Nature*, 437:209–214, 2005. doi:10.1038/nature04017.

[ea06] Neil M. Ferguson et al. Strategies for mitigating an influenza pandemic. *Nature*, 442:448–452, 2006. doi:10.1038/nature04795.

[ea20a] Anastassopoulou et al. Data-based analysis, modelling and forecasting of the covid-19 outbreak. *PLOS ONE*, 2020. doi:10.1371/journal.pone.0230405.

[ea20b] Arcede et al. Accounting for symptomatic and asymptomatic in a seir-type model of covid-19. *ArXiv e-prints*, 2020, 2004.01805.

[ea20c] Castilho et al. Assessing the efficiency of different control strategies for the coronavirus (covid-19) epidemic. *ArXiv e-prints*, 2020, 2004.03539.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

S. No	country	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)	Aver	StdDev
1	US	2.233	1.54	2.166	1.493	0.904	0.957	1.549	0.519
2	Brazil	2.048	2.091	2.048	2.092	1.498	1.57	1.891	0.254
3	Russia	1.195	0.704	1.052	0.688	1.65	1.649	1.156	0.392
4	UK	4.188	3.608	3.721	3.257	0.854	0.904	2.755	1.354
5	Spain	0.739	0.516	0.731	0.511	0.0	0.074	0.429	0.292
6	Italy	0.76	0.129	0.758	0.109	0.115	0.125	0.333	0.302
7	India	1.643	1.65	1.625	1.588	1.644	1.655	1.634	0.023
8	France	1.134	0.668	1.126	0.733	0.121	0.126	0.651	0.413
9	Germany	0.482	0.243	0.486	0.249	0.0	0.054	0.252	0.187
10	Peru	1.653	1.681	1.637	1.673	1.415	1.422	1.58	0.115
11	Turkey	0.649	0.184	0.635	0.033	0.072	0.075	0.275	0.264
12	Iran	0.895	0.899	0.888	0.898	0.158	0.166	0.651	0.346
13	Chile	2.219	2.257	2.22	2.257	1.848	1.849	2.108	0.184
14	Canada	1.009	1.064	1.005	1.05	0.508	0.507	0.857	0.248
15	Mexico	1.181	1.194	1.172	1.175	0.0	0.822	0.924	0.434
16	SaudiArabia	0.568	0.412	0.566	0.459	1.127	1.13	0.71	0.301
17	China	0.535	0.534	0.095	0.0	0.0	0.0	0.194	0.243
18	Pakistan	2.226	2.27	2.19	2.742	1.698	1.712	2.14	0.358
19	Belgium	1.316	1.193	1.304	1.199	0.212	0.228	0.909	0.489
20	Qatar	3.763	3.764	3.643	3.645	3.478	3.38	3.612	0.141
21	Bangladesh	3.547	3.791	3.386	3.795	2.619	2.653	3.298	0.489
22	Netherlands	4.502	2.646	3.823	2.421	0.037	0.015	2.241	1.714
23	Belarus	1.329	1.869	1.371	1.379	1.322	1.32	1.432	0.197
24	Ecuador	1.041	0.303	1.033	0.221	1.062	1.034	0.782	0.369
25	Sweden	2.188	2.075	2.157	1.247	0.804	0.876	1.558	0.599
26	Singapore	1.62	1.429	2.191	0.138	1.657	2.165	1.533	0.685
27	UAE	0.861	0.892	0.907	1.232	1.312	0.999	1.034	0.175
28	SA	1.672	1.678	1.665	1.671	1.599	1.6	1.648	0.034
29	Portugal	1.094	0.968	1.66	0.342	0.937	1.027	1.005	0.384
30	Switzerland	0.453	0.007	0.452	0.001	0.016	0.017	0.158	0.209
31	Colombia	3.165	3.278	3.179	3.267	2.089	3.672	3.108	0.486
32	Kuwait	1.078	2.612	0.544	1.175	2.689	2.672	1.795	0.885
33	Indonesia	2.467	2.521	2.331	2.384	1.367	1.444	2.086	0.485
34	Ireland	0.378	0.084	0.523	0.095	0.111	0.118	0.218	0.17
35	Egypt	3.099	3.165	3.116	3.174	2.424	2.476	2.909	0.326
36	Poland	1.223	1.292	1.23	1.283	0.519	0.542	1.015	0.343
37	Ukraine	1.115	2.4	1.099	0.866	0.957	0.975	1.235	0.528
38	Romania	0.581	0.224	1.047	0.265	0.321	0.335	0.462	0.285
39	Philippines	2.389	2.376	2.334	2.414	0.0	0.937	1.742	0.94
40	DR	0.772	0.595	0.741	0.631	0.73	0.772	0.707	0.069
41	Israel	0.57	0.002	0.556	0.002	0.048	0.05	0.205	0.254
42	Argentina	2.604	2.688	2.62	1.349	1.904	1.945	2.185	0.492
43	Japan	0.352	0.001	0.348	0.001	0.653	0.664	0.336	0.269
44	Austria	0.769	0.0	0.45	0.0	0.007	0.007	0.206	0.3
45	Afghanistan	7.47	7.713	7.562	7.807	0.0	5.924	6.079	2.792
46	Panama	0.504	0.434	0.495	0.464	0.517	0.54	0.492	0.035
47	Denmark	0.489	0.426	0.483	0.413	0.117	0.082	0.335	0.169
48	Oman	0.976	3.591	3.418	3.518	2.684	2.647	2.806	0.902
49	Serbia	0.766	0.082	0.76	0.08	0.452	0.466	0.434	0.279
50	Bahrain	1.195	1.097	1.69	1.972	1.693	1.683	1.555	0.307

Table 6: Effective reproduction number $\mathfrak{R}(t)$ on, May 31, 2020, for countries with different models

[ea20d] Chen et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: a descriptive study. *The Lancet*, 395:507–513, 2020. doi:10.1016/S0140-6736(20)30211-7.

[ea20e] Chen et al. A time-dependent sir model for covid-19 with undetectable infected persons. *ArXiv e-prints*, 2020, 2003.00122.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

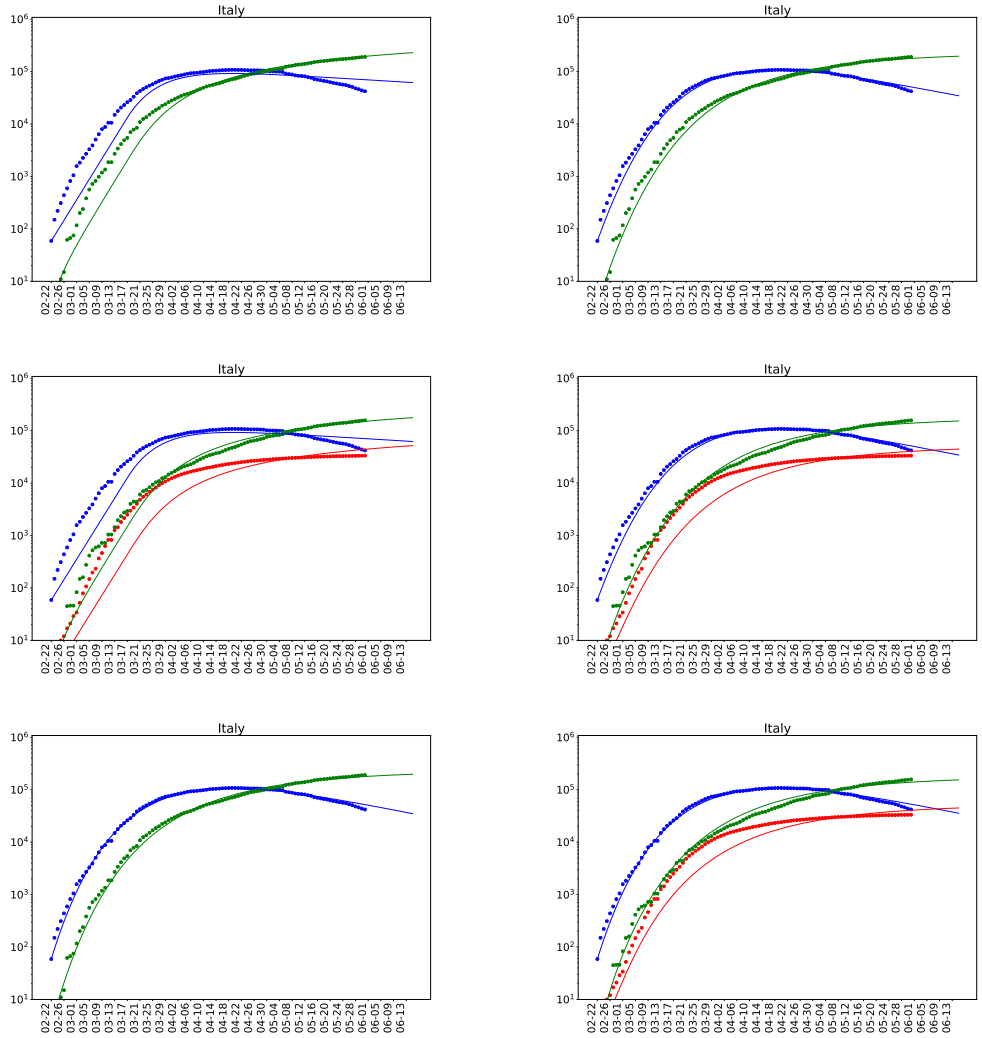

Figure 15: Fitting models with data for Italy. The panel from the top left to the bottom right are for the Models (1) to Model (6) respectively.

[ea20f] Corey et al. A strategic approach to covid-19 vaccine r&d. *Science*, 2020. doi:10.1126/science.abc5312.

[ea20g] Dong et al. An interactive web-based dashboard to track covid-19 in real time. *The Lancet*, 20:533–534, 2020. doi:10.1016/S1473-3099(20)30120-1.

[ea20h] Gorbalenya et al. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-ncov and naming it sars-cov-2. *Nat Microbiology*, 5:536–544, 2020. doi:10.1.1/jpb001.

[ea20i] Hotz et al. Monitoring the spread of covid-19 by estimating reproduction numbers over time. *ArXiv e-prints*, 2020, 2004.08557.

[ea20j] Jonas Dehning et al. Inferring change points in the spread of covid-19 reveals the effectiveness of interventions. *Science*, 2020. doi:10.1126/science.abb9789.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[ea20k] Kucharski et al. Early dynamics of transmission and control of covid-19: a mathematical modelling study. *The Lancet*, 20(5):553–558, 2020. doi:10.1016/S1473-3099(20)30144-4.

[ea20l] Li et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov-2). *Science*, 368(6490):489–493, 2020. doi:10.1126/science.abb3221.

[ea20m] Loli et al. Preliminary analysis of covid-19 spread in italy with an adaptive seird model. *ArXiv e-prints*, 2020, 2003.09909.

[ea20n] Lopez et al. A modified seir model to predict the covid-19 outbreak in spain and italy: simulating control scenarios and multi-scale epidemics. *MedRxiv e-prints*, 2020, 2003.09909. doi:10.1101/2020.03.27.20045005.

[ea20o] Pone et al. A data driven analysis and forecast of an seiard epidemic model for covid-19 in mexico. *ArXiv e-prints*, 2020, 2004.08288.

[ea20p] Prem et al. The effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in wuhan, china: a modelling study. *The Lancet*, 5:261–270, 2020. doi:10.1016/S2468-2667(20)30073-6.

[ea20q] Saptarshi Chatterjee et al. Studying the progress of covid-19 outbreak in india using sird model. *medRxiv*, 2020. doi:10.1101/2020.05.

[ea20r] Sun et al. Covid-19: Epidemiology, evolution, and cross-disciplinary perspectives. *Trends in Molecular Medicine*, 26(5):483–495, 2020. doi:10.1016/j.molmed.2020.02.008.

[ea20s] Walker et al. The global impact of covid-19 and strategies for mitigation and suppression. *Imperial College COVID-19 Response Team*, 2020. doi:10.25561/77735.

[ea20t] Wu et al. A new coronavirus associated with human respiratory disease in china. *Nature*, 579:270–273, 2020. doi:10.1038/s41586-020-2012-7.

[ea20u] Yang et al. Modified seir and ai prediction of the epidemics trend of covid-19 in china under public health interventions. *Journal of thoracic disease*, 12(3):165–174, 2020. doi:10.21037/jtd.2020.02.64.

[ea20v] Zhou et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579:270–273, 2020. doi:10.1038/s41586-020-2012-7.

[FP20] Duccio Fanelli and Francesco Piazza. Analysis and forecast of covid-19 spreading in china, italy and france. *Chaos, Solitons and Fractals*, 134:109761, 2020, 2003.06031. doi:10.1016/j.chaos.2020.109761.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[Gar20a] Lauren Gardner. Mapping 2019-ncov. `https://systems.jhu.edu/research/public-health/ncov/`, 2020.

[Gar20b] Lauren Gardner. Novel coronavirus (covid-19) cases. `https://github.com/CSSEGISandData/COVID-19`, 2020.

[GPD20] Gaurav Goswami, Jayanti Prasad, and Mansi Dhuria. Extracting the effective contact rate of covid-19 pandemic. *ArXiv e-prints*, 2020, 2004.07750.

[Het00] Herbert W. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000.

[iD20] Our World in Data. Data on covid-19 (coronavirus) confirmed cases, deaths, and tests. `https://github.com/owid/covid-19-data`, 2020.

[IM20] Robert Marsland III and Pankaj Mehta. Data-driven modeling reveals a universal dynamic underlying the covid-19 pandemic under social distancing. *ArXiv e-prints*, 2020, 2004.10666.

[JR08] Keeling Matt J. and Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2008.

[Li18] Michael Y Li. *An Introduction to Mathematical Modeling of Infectious Diseases*. Springer International Publishing, 2018.

[NC08] Grassly NC and Fraser C. Mathematical models of infectious disease transmission. *Nat Rev Microbiol.*, 6:477–87, 2008. doi:10.1038/nrmicro1845.

[Oli20] Gonçalo Oliveira. Refined compartmental models, asymptomatic carriers and covid-19. *ArXiv e-prints*, 2020, 2004.14780. doi:10.1101/2020.04.14.20065128.

[Org20a] World Health Organization. Novel coronavirus (2019-ncov) situation report - 1. `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports`, 2020.

[Org20b] World Health Organization. Report of the who-china joint mission on coronavirus disease 2019 (covid-19). `https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19)`, 2020.

[P.20] Tales P. A time-dependent seir model to analyse the evolution of the sars-covid-2 epidemic outbreak in portugal. *Bull World Health Organ*, 2020. doi:10.2471/BLT.20.258392.

[Pra20a] Jayanti Prasad. A data first approach to covid-19. `https://github.com/jayanti-prasad/Covid-19/tree/master/project2/`, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.22.20110171; this version posted June 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[Pra20b] Jayanti Prasad. A python package for fitting covid-19 data. `https://github.com/jayanti-prasad/PyCov19`, 2020.

[RR20] Singh R. and Adhikari R. Age-structured impact of social distancing on the covid-19 epidemic in india. *ArXiv e-prints*, 2020, 2003.12055.

[Sci20] SciPy. Scientific computing tools for python. `https://www.scipy.org/`, 2020.

[SNC20] Tridip Sardar, Sk Shahid Nadim, and Joydev Chattopadhyay. Assessment of 21 days lockdown effect in some states and overall india: A predictive mathematical study on covid-19 outbreak. *ArXiv e-prints*, 2020, 2004.03487.

[Vil20] Villaverde. Estimating and simulating a sird model of covid-19 for many countries, states, and cities. `https://cepr.org/active/publications/discussion_papers/dp.php?dpno=14711`, 2020.

[Wor20] Worldometers. Covid-19 coronavirus pandemic. `https://www.worldometers.info/coronavirus/`, 2020.