Analysis of the official R_0 data in Lombardy after 4th May

Francesco Naccarato

August 2020

1 Introduction and results

After Italy started *Fase 2* of the SARS-CoV-2 epidemic containment on May 4th 2020, the Italian government has released weekly reports of the situation in both the whole country and the single regions. This reports contain relevant data such as the number of new cases and the estimations of the contagion factor $R_0 = R_0(t)$.
Our goal is to use the R_0 data in constructing a SIR model for the epidemic and check whether the model prediction matches the actual data for new and total cases. We find that the estimates were pretty accurate and that the actual values surely lied in the given confidence intervals, although R_0 was probably slightly underestimated in the period immediately after 4th May. Moreover, by testing the model with several choice of a certain parameter, we make deductions on the average time an infected individual is contagious before self-isolating.
Due to the high variance in R_0 between the various regions at a fixed time t, we chose to study the evolution of the situation in just a single region, and we chose Lombardy being the first-hit region and having by far with the higher number of cases and tests.

2 Model

We recall the basics of the SIR model: the N individuals are divided in **S**usceptible, **I**nfectious and **R**emoved and the evolution of the epidemic is governed by the set of differential euqations:

$$\begin{cases} \frac{dS}{dt} = -\frac{\beta IS}{N} \\ \frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I \\ \frac{dR}{dt} = \gamma I \end{cases} \tag{1}$$

where β and γ are parameters representing respectively the average number of contacts per person per time (day) multiplied by the probability of disease transmission in a contact between a susceptible and an infectious subject, and

the reciprocal of the average time oan individual remains infectious after being infected.
It is clear that in a real scenario the value of β is not constant in time: measures like social distancing, masks, lockdowns and simple awareness can lower it substantially.
It would appear that γ should instead be constant; notice though that, in an advanced phase of the epidemic, a responsible person would self-isolate when showing first symptoms, which is not necessarily true in an early phase of the epidemic. Recall that we are analyzing *Fase 2*, in which the lockdown measures where progressively weakened after being in place for nearly 2 months. We can hereby assume that γ is constant in our period of interest, but it does not correspond to its "biological" definition: in fact, we have to take γ as the reciprocal of the average time one is infectious but does not show particularly strong symptoms. We will test our model with various choices of γ, which most certainly lies in the $[2, 8]$ days interval, based on WHO estimates for the incubation period.
The parameter R_0 is defined as:

$$R_0 = \frac{\beta}{\gamma} \tag{2}$$

and hence can be interpreted as the average number of people infected by a single infectious individual. It is clear that in the $R_0 > 1$ regime every susceptible individual will eventually become ill, hence the emphasis on lowering it under 1.
The following tables reports the weekly values of R_0 according to the Ministry of Health:

Period	4-10 May	11-17 May	18-24 May	25-31 May	1-7 June	8-14 June	15-21 June
R_0	0.62	0.51	0.75	0.91	0.9	0.82	1.01
Conf. int.	[0.59,0.64]	[0.47,0.55]	[0.72,0.84]	[0.78,1.09]	[0.75,1.02]	[0.64,0.95]	[0.71,1.37]

Table 1: R_0 values

Period	22-28 June	29-5 July	6-12 July	13-19 July	20-26 July	27-2 Aug	3-9 Aug
R_0	0.89	0.92	1.14	1	0.96	1.04	1.13
Conf. int	[0.54,1.28]	[0.66,1.26]	[0.8,1.47]	[0.71,1.36]	]0.73,1.19]	[0.84,1.27]	[0.87,1.43]

Table 2: R_0 values

We can now implement the model. We will start by fixing $\gamma = \frac{1}{5}$. Notice that we can recover β as γR_0.

```
function R=atk_rate(t)
    %Official R_0 value week-by-week
```

```
    v=[0.62,0.51,0.75,0.91,0.90,0.82,1.01,0.89,0.92,1.14,1,0.96,1.04,1.13];
    l=1+floor(t/7);
    R=v(l);
end

function y=sir(t,x)
%Implements the SIR system of ODEs
%N: Lombardy total population
%g: 1/gamma
    N=10^7;
    g=5;
    y=1/g*[-atk_rate(t)*x(1)*x(2)/N; atk_rate(t)*x(1)*x(2)/N-x(2); x(2)];
end

function model
%Implements the model
    io=37307; %initially infected
    ieff=4299; %(estimated) number of infectious
               %actually infecting others in the first week
    do=14294; %initial deaths
    reco=26504; %already recovered
    ro=reco+do+io-ieff; %initial value for R
    pop=10^7; %total population

    [t,x]=ode45(@sir, [1 97], [pop-ieff-ro, ieff, ro]);
    plot(t,x(:,2)+x(:,3)) %plotting the total number of cases
end
```

Let us explain the numbers appearing in the functions that implement the model. First, we have to understand that the issue of initial conditions is not trivial: we cannot just use the data for infected and recovered people from May 4th; officially[1], the number of *infected* people in Lombardy as of May 4th was 37307 and those who recovered were 26504. However, as can be seen by implementing the model directly with these initial conditions, the results do not match the actual evolution at all. This is basically a consequence of the same fact we mentioned while discussing the value of γ: not all the people in **I** on May 4th were actually infecting people in the following week: most of them were hospitalized or self-isolating and should therefore be considered already in **R**.

We should instead set our initial condition for **I** as the new cases between half of the $4-10$ week and half of the following week: these are the 4299 people[2] that were actually moving freely. The **R** category is thus made up by the other 33008 infected people, along with the recovered and the 14294 deaths[1]. Implementing the model this way we get:

[1]https://statistichecoronavirus.it/regioni-coronavirus-italia/lombardia/

[2]https://lab.gedidigital.it/gedi-visual/2020/coronavirus-i-contagi-in-italia/lombardia

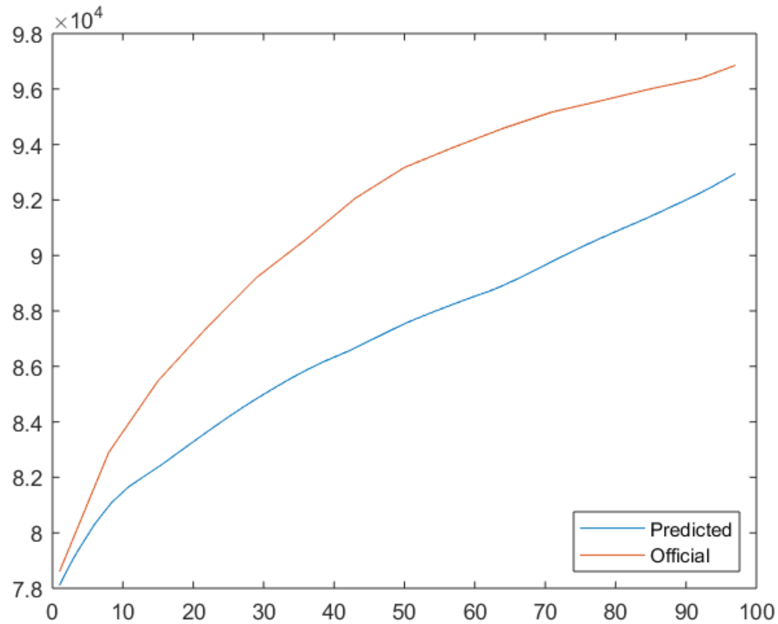

Figure 1: Actual cases vs prediction with $g = 5$

This seems a good but not perfect match for the actual curve, plotted with the simple script[1]:

```
t=[1,8,15,22,29,36,43,50,57,64,71,78,85,92,97];
c=[78605,82904,85481,87417,89205,90581,92060,93173,
93901,94580,95173,95582,96008,96381,96853];
plot(t,c)
```

The following figures show the results obtained by running the model with different values of γ:

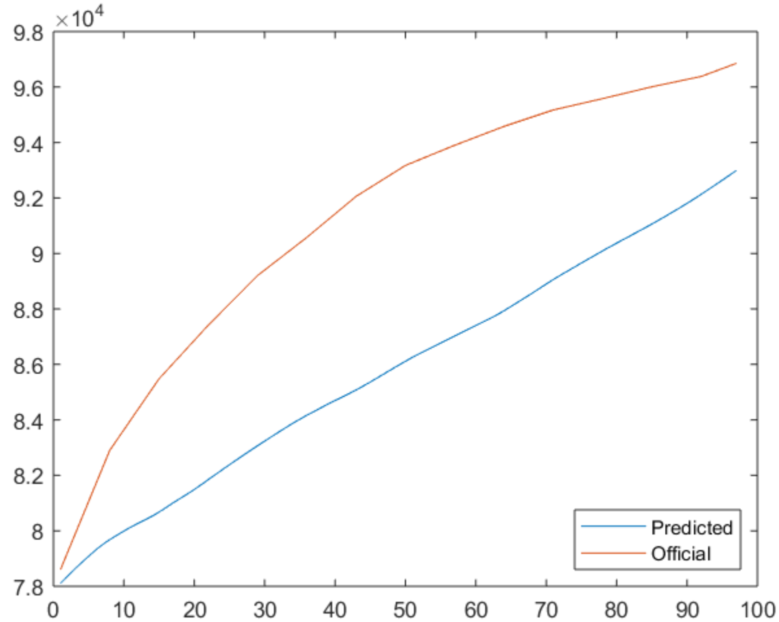

Figure 2: Actual cases vs prediction with $g = 10$

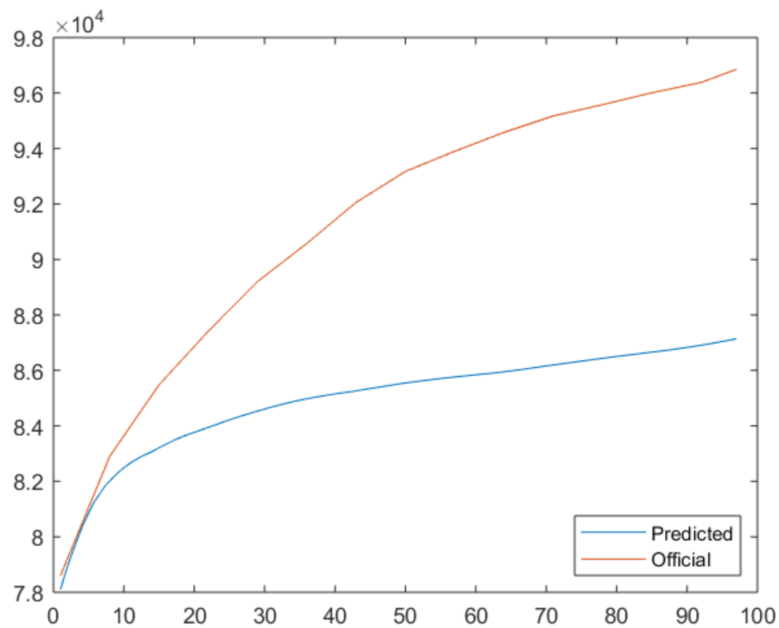

Figure 3: Actual cases vs prediction with $g = 3$

Lower values for γ do not really give a substantially different graph, while higher values, instead, give a substantially lower number of total cases, as can be expected.

3 Conclusions

The fact that the predicted curve is a bit lower than the actual one is probably caused by a slight underestimation of some of the actual values of R_0 by the authorities; nonetheless, we have to keep in mind that with *Fase 2*, interregional travel was re-opened in some exceptional cases, and that could have contributed to the difference between the two curves.
As we saw, the difference is surely not caused by the choice $\gamma = \frac{1}{5}$; moreover, the fact that with $\gamma = \frac{1}{3}$ the model does not give a good prediction gives us a good estimation for a lower bound, that is, 4, on the number of days an infected individual is potentially infectious.
We moreover notice that, although probably a bit underestimated, the official values of R_0 are not *very wrong*: in fact, if we implement the model using the *upper bounds* on the confidence intervals for the R_0s (and $g = 5$), we get this:

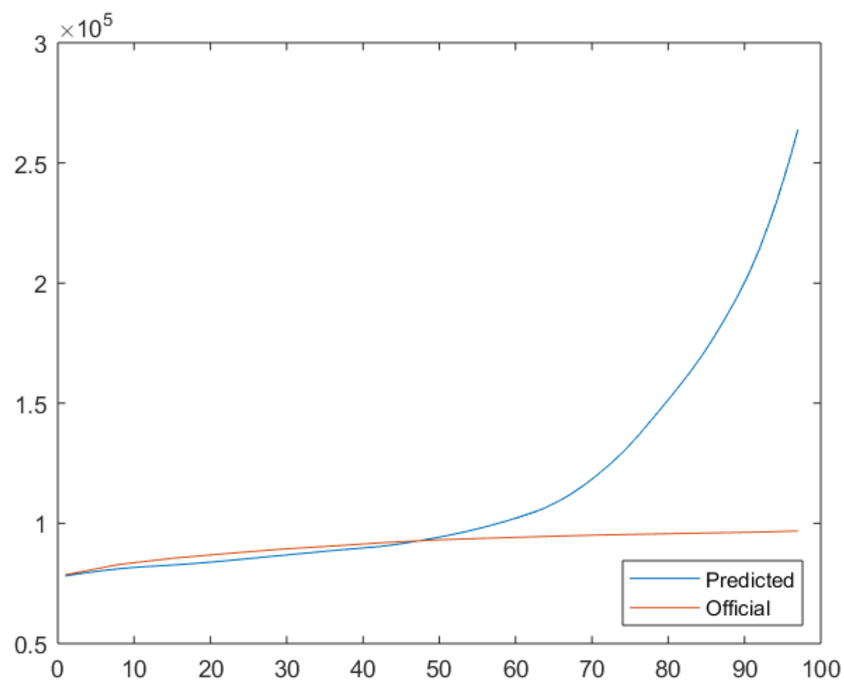

Figure 4: Actual cases vs prediction with values of R_0 chosen as the upper bounds on the official confidence intervals

which is far from the actual curve. But still, we can notice that the prediction surpasses the actual data only after mid-June: this suggests that the values for R_0 were probably underestimated in the first month after the re-opening.