Forecasting coronavirus in Italy with SIRD modelling

Gabrielle Littlefair

Oral: `https://imperial.cloud.panopto.eu/Panopto/Pages/Viewer.aspx?id=7d1f1304-438b-40e3-b586-abd7011a8903`

Objectives

- Find estimates for the contact, death and recovery rates of coronavirus in Italy using least squares regression.
- Identify any issues with the model and data.
- Extrapolate the model to see how coronavirus may progress in the upcoming months.

Introduction

The first case of coronavirus in Italy was reported on the 31st January 2020. Between then and the 6th June 2020, 33,846 people died and the total number of cases rose to 234,801 [8]. 23.3% of Italy's population is over the age of 65 [5], making it the second oldest population in the world. This may explain why Italy seems especially vulnerable. I have modelled Italy's outbreak using SIRD modelling. Modelling accurate rates of infection, recovery and death, along with the basic reproduction number, R_0, can be quite challenging due to the high proportion of infections that are undetected. The number of infections has been estimated to be as high as 63 times as large as the number recorded [3].

SIRD Modelling

SIRD modelling is based on four different groups within the population: those who are susceptible (S); those who are infected (I); those who have recovered (R); and those who have died (D).
The governing equations of my model are as follows:

$$S + I \xrightarrow{\beta} 2I$$
$$I \xrightarrow{\gamma} R$$
$$I \xrightarrow{\delta} D$$

Where β is the contact rate, γ is the rate of recovery, and δ is the rate of death. From these equations, the following system of ODEs can be found and solved [6]:

$$\frac{dS}{dt} = -\frac{\beta}{N}SI$$
$$\frac{dI}{dt} = \frac{\beta}{N}SI - (\gamma + \delta)I$$
$$\frac{dR}{dt} = \gamma I$$
$$\frac{dD}{dt} = \delta I$$

Fitting the Model to the Data

I used least squares regression from python's lmfit module to fit my SIRD model to my data [7]. I began by assuming that my rates β, γ and δ were all constant. However, after plotting the results of this model, I quickly realised that my value of β needed to decrease with time. Therefore, I decided to define $\beta(t)$ as a function instead. I found this function intuitively using a negative exponential model multiplied by the β value found. I also assumed that the initial number of susceptible in Italy was equal to the population: 60,461,828 [1]. The results for recovery and death rates were as follows:

$$\gamma = 0.0234$$
$$\delta = 0.0064$$

The following values of β correspond to the start of the data and the end of lockdown:

$$18/02/2020 \quad \beta = 0.1473$$
$$14/05/2020 \quad \beta = 0.0080$$

From these values of β we can see that at the end of lockdown, the infection rate was much lower, as expected, due to much fewer contacts between those in the susceptible group and those in the infected group.

The Model

Figure 1:Data vs Model plotted between 18/02/2020 and 07/06/2020

R_0

An important feature of modelling epidemics is the basic reproduction number R_0. This number is the number of secondary infections resulting from a single primary infection. The reason this value is so important is that it is an indication of whether the disease will die out ($R_0 < 1$), or if it will become an endemic ($R_0 > 1$) [4]. This value changes when measures are implemented that reduce the rate of infection, like lockdown. R_0 can be found using the following equation [7]:

$$R_0 = \frac{\beta(t)}{\gamma}$$

Using my model, I have found values of R_0 at different times:

$$18/02/2020 \quad R_0 = 6.2929$$
$$22/04/2020 \quad R_0 = 1.0052$$
$$23/04/2020 \quad R_0 = 0.9542$$
$$14/05/2020 \quad R_0 = 0.3405$$

R_0 is first below 1, meaning that the disease has begun dying out, on the 23rd April 2020, in the middle of the lockdown period. This suggests that lockdown was effective. When Italy began relaxing its lockdown measures, R_0 was very small, however as the lockdown eases and contact rates increase that value could easily rise once again.

Assumptions

SIRD modelling has many drawbacks. First of all, the assumes that once you have been infected you are then immune to the virus, which has not been proved or disproved for coronavirus yet. If this assumption is proved wrong, the model would be unreliable until another category is added in. My model assumes that you become infectious when you become infected (i.e. when you test positive), however, according to Harvard Medical School [2], you actually become infectious up to 72 hours before you show any symptoms. Another assumption is that there are no births in the population. There are also problems with the accuracy and reliability of the data [8] being used. Many cases of coronavirus go unreported and untested. There could also be a high proportion of asymptomatic cases [3], which will also remain unrecorded.
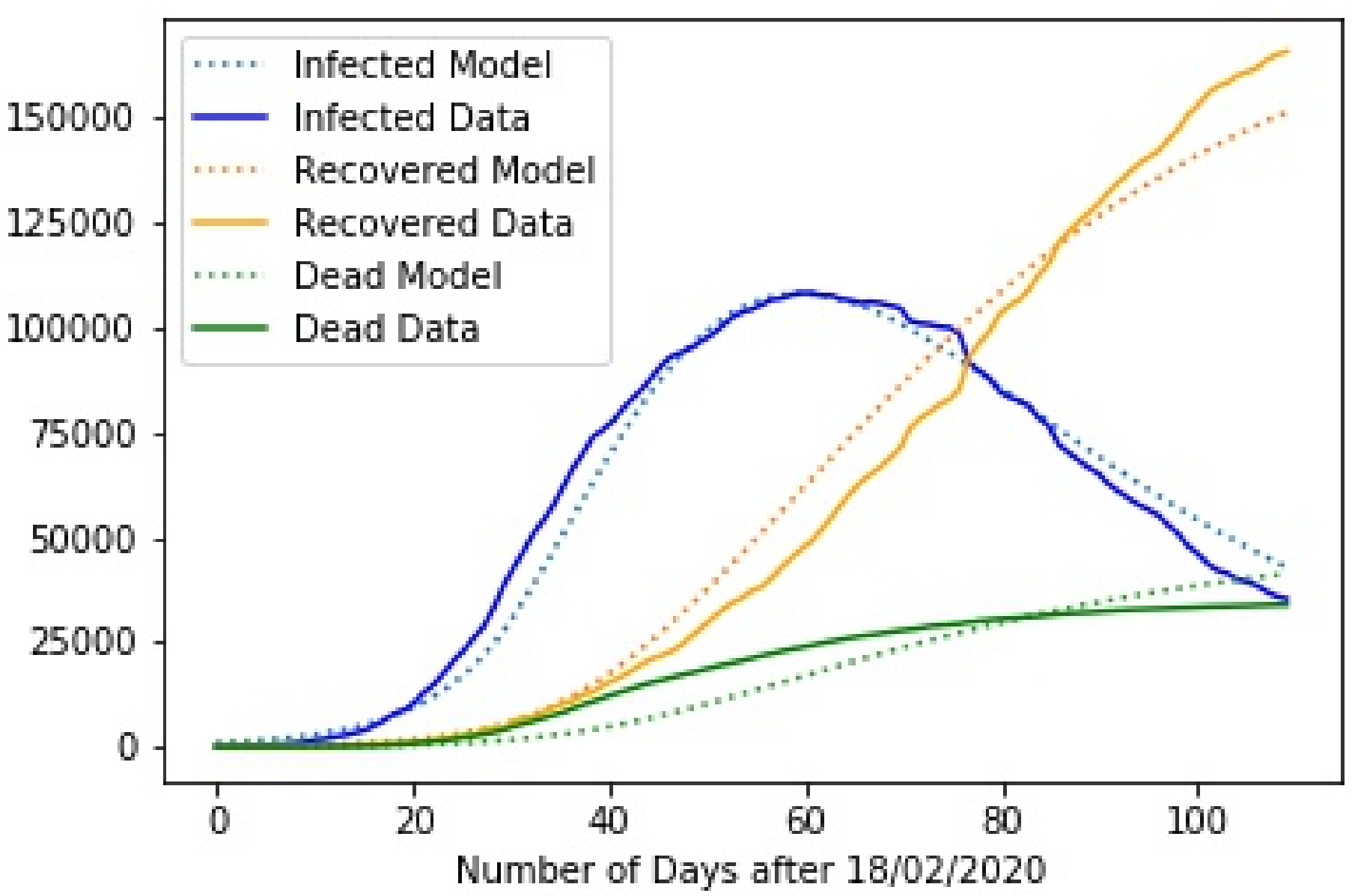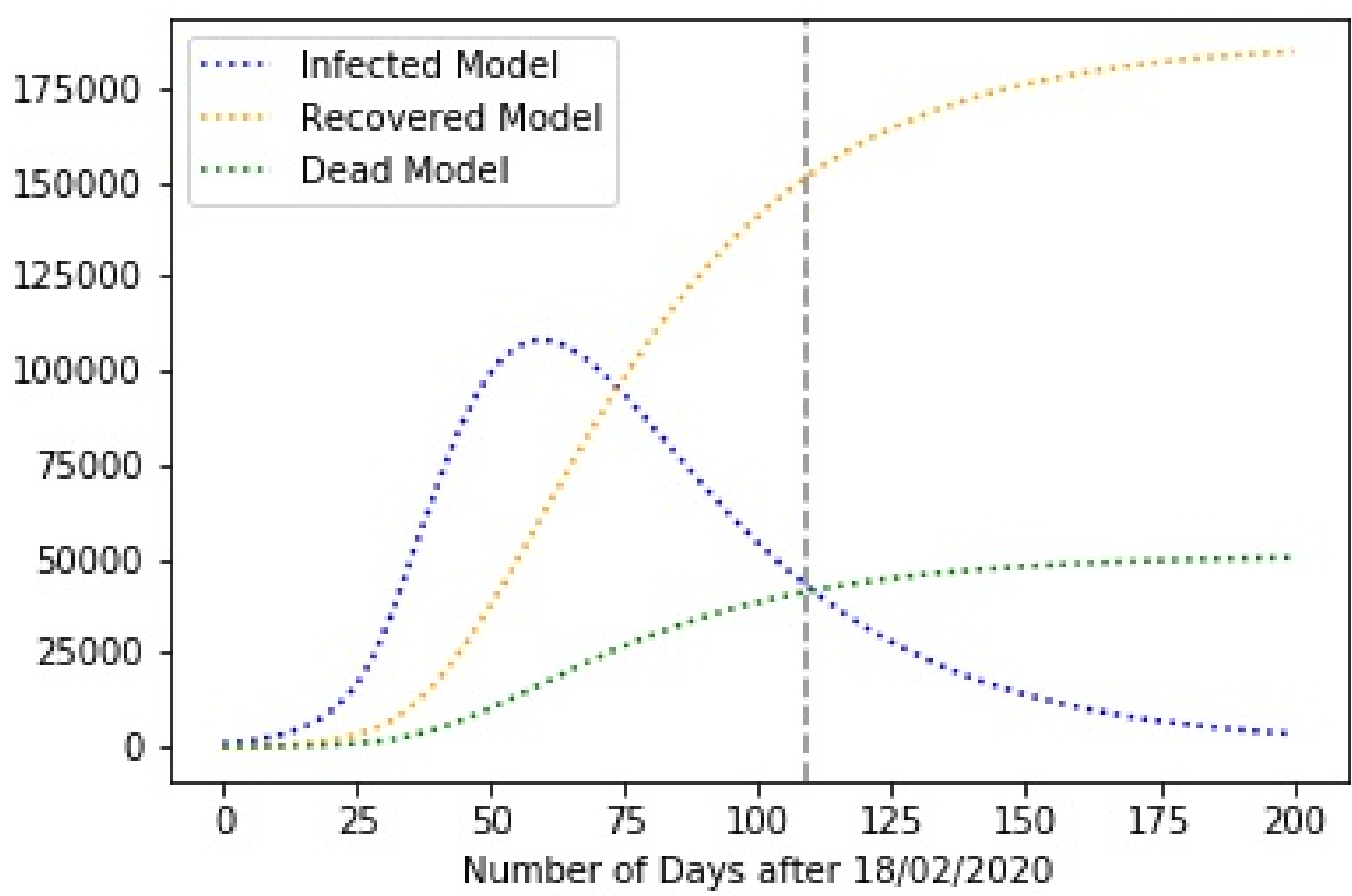
Forecast

Figure 2:Model plotted between 18/02/2020 and 06/08/2020

Conclusion

Although I do not believe that the model I have generated is very accurate, it seems that Italy is very much on the way towards eradicating coronavirus. However, the unknown proportion of the population that has coronavirus and is asymptomatic, along with the unrecorded cases, could mean that we may begin to see an upward trend in the number of cases. Potentially even a second peak. Going forward, I would like to add in more compartments into my model (such as the exposed compartment), in order to increase the accuracy of the model. This exposed category would remove the need for a dampener on the rate of infection, as a contact rate would then also be estimated, making a much more reliable model. I would also like to compare various countries which have now left lockdown in order to see how exiting the lockdown has affected their infection rates and whether a second spike seems likely.

References

[1] "Our world in data coronavirus." [Online]. Available: https://ourworldindata.org/coronavirus

[2] "Harvard health coronavirus." [Online]. Available: https://www.health.harvard.edu/diseases-and-conditions/if-youve-been-exposed-to-the-coronavirus

[3] G. C. Calafiore, C. Novara, and C. Possieri, "A modified sir model for the covid-19 contagion in italy," Mar 31, 2020. [Online]. Available: https://arxiv.org/abs/2003.14391

[4] K. Nixon and L. Servitje, *Endemic*. London: Palgrave Macmillan Limited, 2016. [Online]. Available: https://ebookcentral.proquest.com/lib/[SITE_ID]/detail.action?docID=4720003

[5] J. B. Dowd, L. Andriano, D. M. Brazel, V. Rotondi, P. Block, X. Ding, Y. Liu, and M. C. Mills, "Demographic science aids in understanding the spread and fatality rates of covid-19," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 18, pp. 9696–9698, May 5, 2020. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/32300018

[6] . K. S. E. Model, "Introduction to epidemic modeling."

[7] J. Fernández-Villaverde, "Estimating and simulating a sird model of covid-19 for many countries, states, and cities," 2020. [Online]. Available: http://www.econis.eu/PPNSET?PPN=1698547927

[8] "John hopkins university coronavirus data." [Online]. Available: https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases

Imperial College London