

MA677 Final Project

Shangchen Han

5/1/2020

Abstract

The COVID-19 Coronavirus is a new virus for us, and it still spreading very quickly in the world. In each country, more and more people get infected, and from now there is not any signal that it will stop in anytime. Thus, I choose France data from January 22 to April 25 as my whole data set which I found it from Harvard dataverse, and use SIR model and ARIMA model to predict the peak numbers and also future patients number.

SIR Model

The SIR model can be complexified to incorporate more specificities of the virus outbreak. The basic idea behind the SIR model (Susceptible - Infectious - Recovered) of communicable disease outbreaks is that there are three groups of individuals:

S: the number of susceptible to the disease. At the start of the pandemic, S is the entire population since no one is immune to the virus.

I: the number of infectious people

R: The number of recovered or deceased (or immune) individuals. They are not infectious anymore.

These groups evolve over time as the virus progresses in the population. Susceptible decreases when individuals are contaminated and move to the infectious group. As people recover or die, they go from the infected group to the recovered group. To model the dynamics of the outbreak I need three differential equations to describe the rates of change in each group, β , the infection rate, which controls the transition between S and I. γ , the removal or recovery rate, which controls the transition between I and R.

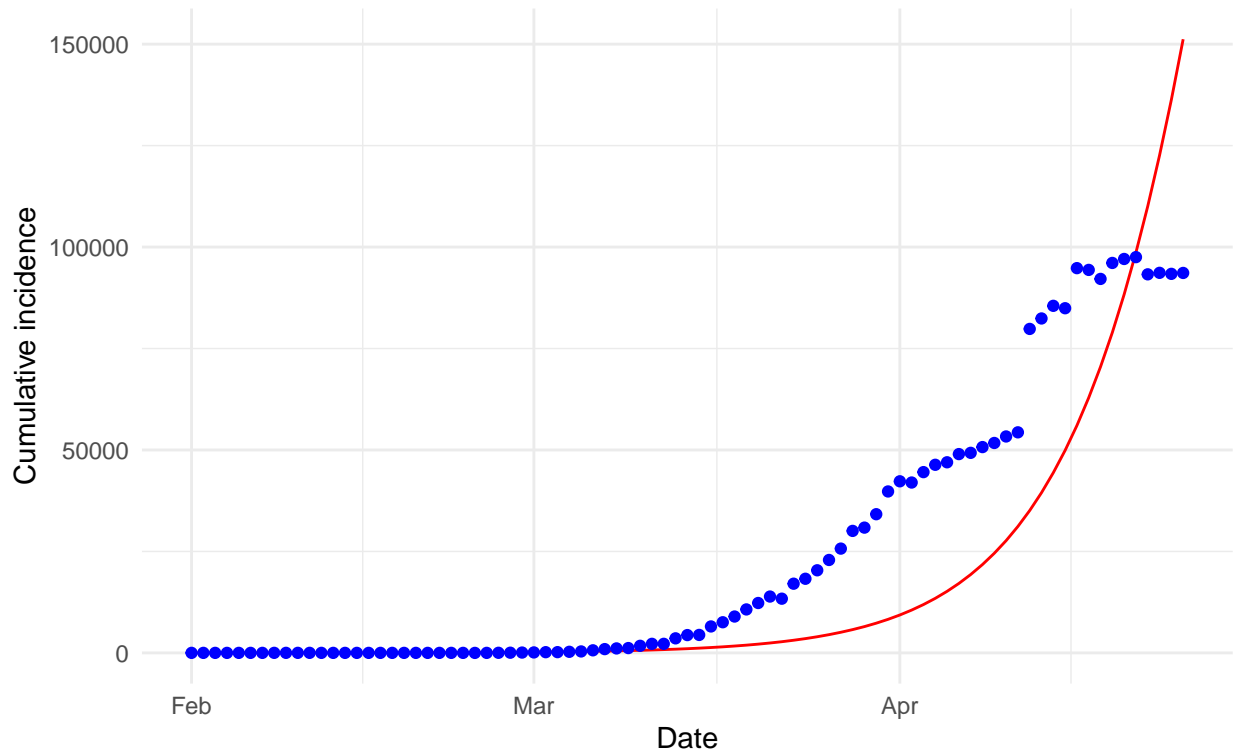
$$\frac{dS}{dt} = -\frac{\beta IS}{N}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

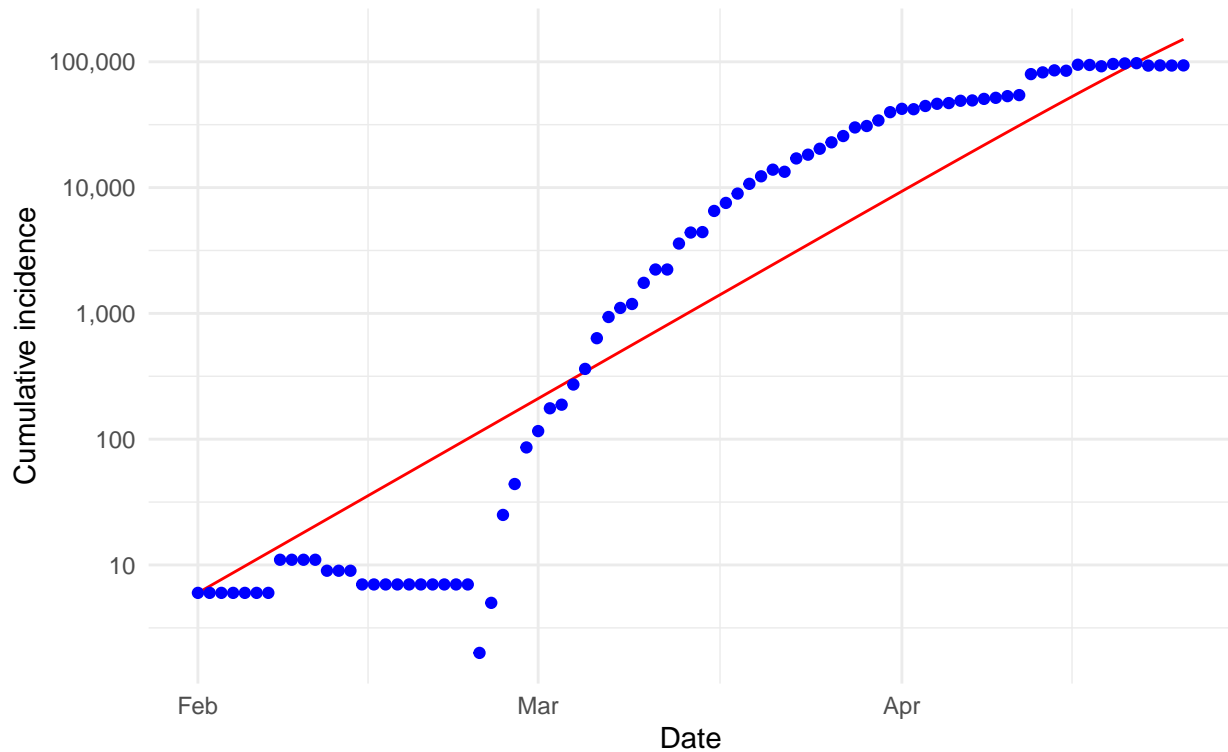
First of all, we need to minimise the sum of the squared differences between $I(t)$, which is the number of people in the infectious compartment I at time t, and the corresponding number of cases as predicted by our model $\hat{I}(t)$. In order to fit a model to the incidence data for France, we need a value N for the initial uninfected population. The population of France was 67,075,000 people, according to Wikipedia.

COVID-19 fitted vs observed cumulative active patients in France
(Red = fitted from SIR model, blue = observed)



From the above graph, roughly the observed data is following our fitted model. There is a big jump on April 11 from around 50 thousand to 75 thousand. But on the end of April the trend of increasing becomes more flat.

COVID-19 fitted vs observed cumulative active parients in France
(Red = fitted from SIR model, blue = observed)



The above graph shows that, roughly the whole February, the number of confirmed patients is not under the fitted model, and it is still low. Since the end of February, the number start to increase significantly, and from the beginning of March it surpasses our model. The data in y-axis is in log scale.

Calculating R_0

We can use our fitted model to calculate the basic reproduction number R_0 , also referred as basic reproduction ratio, and which is closely linked to β and γ . The reproduction number gives the average number of susceptible people who are infected by each infectious person. In other words, the reproduction number refers to the number of healthy people that get infected per number of sick people. When $R_0 > 1$ the disease starts spreading in a population, but not if $R_0 < 1$. Usually, the larger the value of R_0 , the harder it is to control the epidemic and the higher the probability of a pandemic.

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

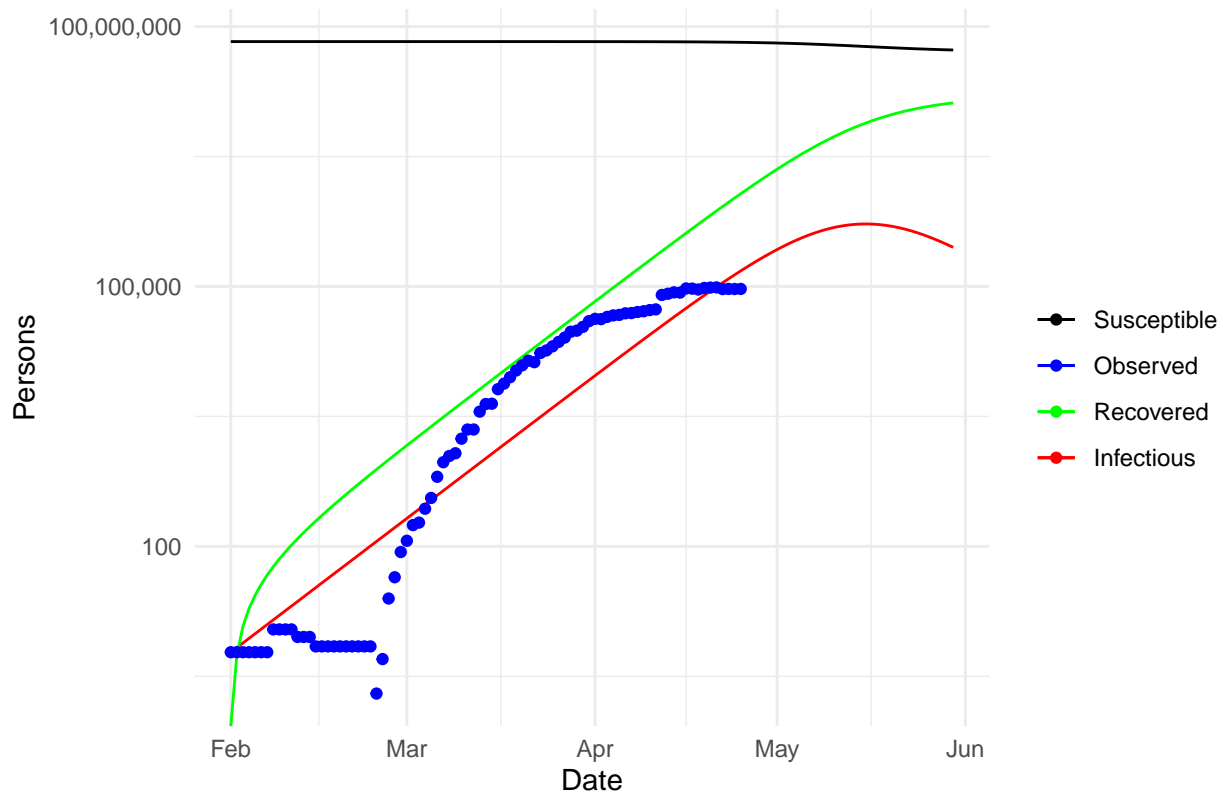
```
## [1] 1.139815
```

```
##      beta      gamma
## 1.0000000 0.8773356
```

A R_0 of 1.14 means that, on average in France, 1.14 persons are infected for each infected person. Furthermore, in the literature, it has been estimated that the reproduction number for COVID-19 is approximately 2.7. My reproduction number being lower is mainly due to the number of confirmed cases stayed constant and equal to 1 at the beginning of the pandemic. The reproduction number of 1.14 suggests that 12.3% of the population should be immunized to stop the spread of the infection. With a population in France of approximately 67 million, this translates into roughly 8.24 million people.

Analyzing the outbreak

COVID-19 fitted vs observed cumulative cases in France

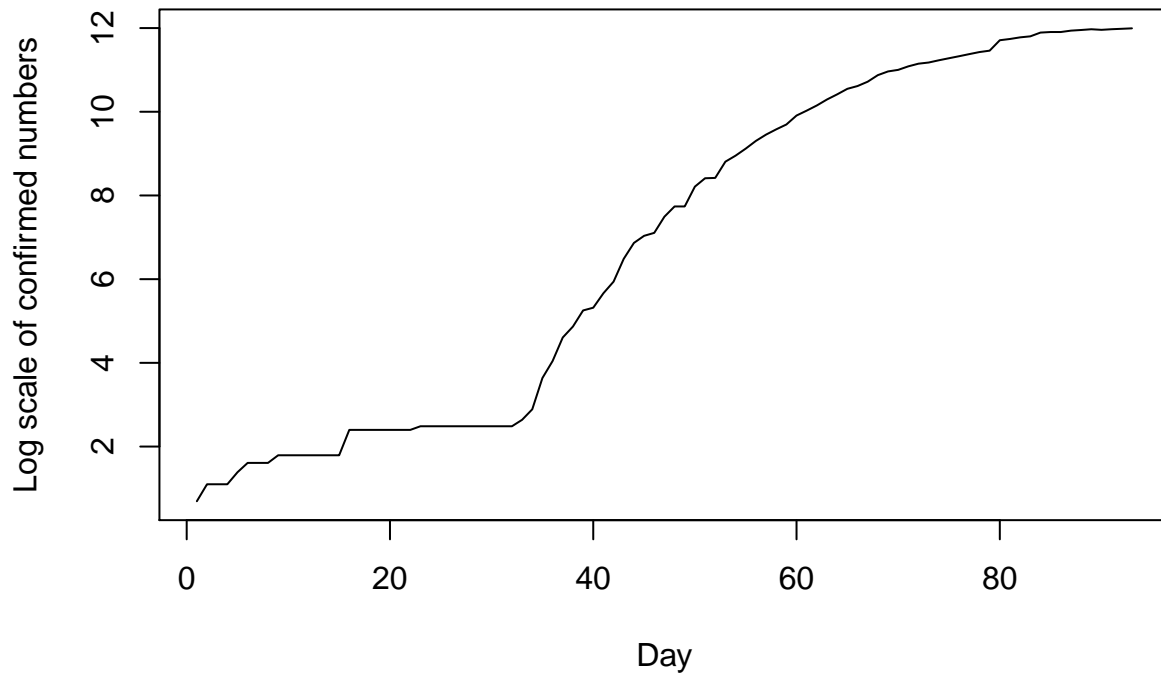


From the graph above, we could see the peak of infectious that will occur in the middle of May. The number of recovered is still waiting for the peak. In terms of number of observed seems like flat since the middle of April.

```
##          Date      I
## 106 2020-05-16 526334.9
```

This shows that the peak of incidence will reach peak on May 16th.

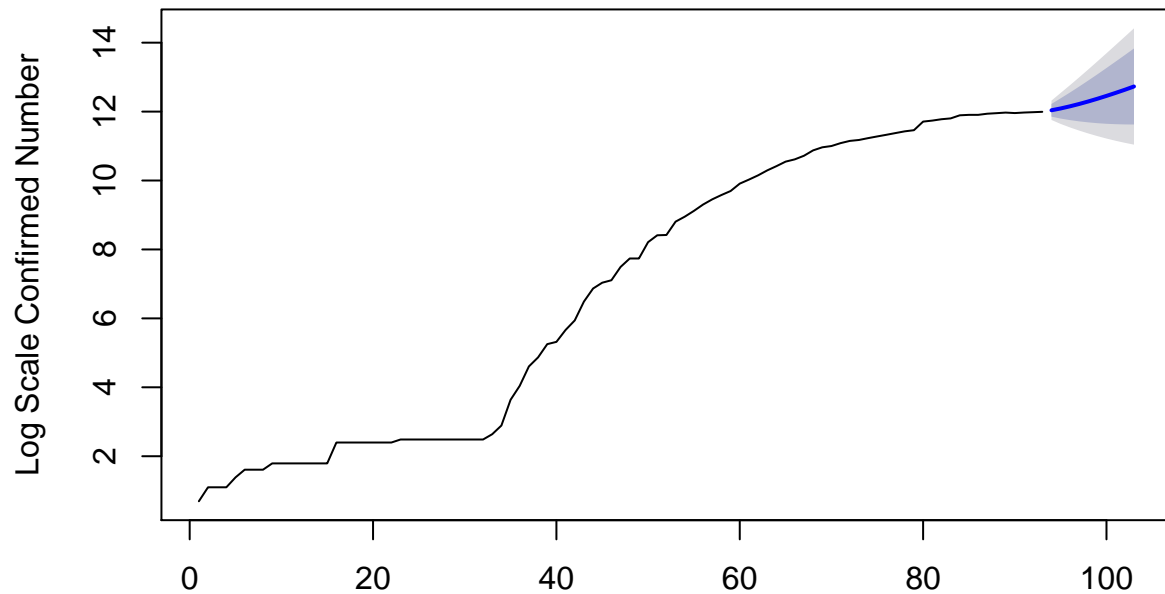
Log Number of Confirmed in France per day



I used log transformation to make the plot more readable. As we could see, the confirmed number from day 35(roughly) start to increase dramatically, and it still does not reach the peak until day 80.

##	Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
##	94	12.03798	11.85447	12.22149	11.75733	12.31864
##	95	12.09129	11.79763	12.38495	11.64218	12.54040
##	96	12.15213	11.75331	12.55094	11.54219	12.76206
##	97	12.21964	11.71703	12.72224	11.45096	12.98831
##	98	12.29307	11.68747	12.89867	11.36688	13.21926
##	99	12.37176	11.66404	13.07948	11.28940	13.45413
##	100	12.45512	11.64639	13.26386	11.21827	13.69198
##	101	12.54262	11.63420	13.45105	11.15331	13.93194
##	102	12.63380	11.62722	13.64038	11.09436	14.17323
##	103	12.72823	11.62517	13.83130	11.04124	14.41522

Forecasts Confirmed number in France by using ARIMA(1,1,1)



I used ARIMA(1,1,1) model, and as we could see the whole trend still increase in the future, and it is still waiting for the peak of confirmed number.

Reference

<https://dataverse.harvard.edu/dataverse/2019ncov>

https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology