Simulation of COVID-19 in different countries using SIR-F model

Meifang Li
13043390
University of Amsterdam
Amsterdam, the Netherlands

Yuhao Qian
13011456
University of Amsterdam
Amsterdam, the Netherlands

ABSTRACT

In this report, we built a SIR-F model to study the spread of COVID-19 in four different countries. We also fit real-world data into the SIR-F model in different phases based on the change of S-R trend and obtained the estimated R_0 through different periods. We explored the relationship between the stringency of lockdown strategies and estimated R_0, also used the estimated parameters to predict the number of infections and deaths in the coming 30 days.

1 INTRODUCTION

The COVID-19 firstly occurred in December 2019 in Wuhan (Hubei Province, China) and has rapidly rise concerns of all scientists around the world[1]. On March 11, it has become a pandemic declared by the World Health Organization(WHO) which has spread around the globe[4]. Almost every country in the world suffered from severe effects caused by COVID-19. Currently, the number of confirmed cases is still increasing worldwide and the pressure on global healthcare systems has been considerable, with more than 41.57 million cases and 1.13 million deaths by October 23, 2020.

Under such circumstances, mathematical models are required to estimate disease transmission, recovery, deaths, and other significant parameters separately for various countries. For the classic SIR model, the R class usually represents the recovered individuals. However, these assumptions are not representative enough in the presence of massive deaths caused by the COVID-19. Thus, We would introduce the SIR-F model, which is a derived-SIR customized by Covsirphy Development Team. This SIR-F model divides the susceptible class into confirmed and un-categorized and introduces a new class: fatal with a confirmation. We could fit real-world data into the SIR-F model in different phases based on the change of S-R trend and obtain the estimated reproductive ratio through different periods. The estimated parameters can also be used to make a prediction of infections and death numbers in the future.

In the absence of highly effective treatment or approved vaccines, the most wide-used strategies against the transmission of this virus are social distancing, allied with contact tracing, self-quarantine, and wearing a face mask. Governmental responses exhibit significant nuance and heterogeneity based on their projections[5]. Social distancing policies include school and workplace closure, cancellation of public events and gatherings in addition to movement restrictions with public transport closure, and prohibition of internal and international travel. The University of Oxford, Blavatnik School of Government has documented these policies according to date, categorized them into specific groups, and each graded according to policy stringency. Maximum policy stringency was only found to be associated with reduced mortality or case numbers for international travel restrictions[2].

In our study, we used COVID-19 datasets from John Hopkins University in the form of time-series, spanning January to October 2020. These datasets are used to estimate the parameters of the SIR-F model to understand the effects and estimate the trend of the disease in different countries, namely China, Japan, Italy, and Singapore in this report. We also explore the impact of lockdown strategies with regard to their strength in these four countries.

2 THEORY

2.1 SIR-F Model

SIR-F model, customized by Covsirphy Development Team, is a mathematical model derived from the most famous SIR model. The basic SIR model has Susceptible, Infectious, and Recovered these three classes.There are five classes in SIR-F model: susceptible(S), confirmed and un-categorized(S^*), confirmed and diagonalized as infectious(I),recovered(R) and fatal with confirmation(F). In this pandemic, some people died before they are confirmed as COVID-19 case, so the new term S^* is necessary. The data provided by John Hopkins University have cumulative confirmed cases, recovered, and death. The confirmed cases consists infectious(I),recovered(R) and fatal(F). We set the total population is $N = S + I + R + F$ and the elapsed time from the starting date is T. The SIR-F model can be presented by:

$$\frac{dS}{dT} = -\beta\frac{SI}{N} \tag{1}$$

$$\frac{dI}{dT} = (1-\alpha_1)\beta\frac{SI}{N} - (\gamma+\alpha_2)I \tag{2}$$

$$\frac{dR}{dT} = \gamma I \tag{3}$$

$$\frac{dF}{dT} = \alpha_1\beta\frac{SI}{N} + \alpha_2 I \tag{4}$$

If we set $(S, I, R, F) = N \times (x, y, z, w)$ and

$$(T, \alpha_1, \alpha_2, \beta, \gamma) = (\tau t, \theta, \tau^{-1}\kappa, \tau^{-1}, \tau^{-1}\rho, \tau^{-1}\sigma)$$

the ODE system above turns to:

$$\frac{dx}{dt} = -\rho xy \tag{5}$$

$$\frac{dy}{dt} = \rho(1-\theta)xy - (\sigma+\kappa)y \tag{6}$$

$$\frac{dz}{dt} = \sigma y \tag{7}$$

$$\frac{dw}{dt} = \rho\theta xy + \kappa y \tag{8}$$

where N is the total population and τ is a coefficient for simplicity. The basic reproduction rate of this model is calculated by:

$$R_0 = \rho\frac{(1-\alpha_1)}{(\gamma+\alpha_2)}$$

The parameters $(\theta, \kappa, \rho, \sigma)$ of example data would be estimated by the hyperparameter optimization method *covsirphy.simulation.estimator*[3].

2.2 Stringency index

Publicly available data on the national response to the COVID-19 pandemic was extracted from the Oxford COVID-19 Government Response Tracker, Blavatnik School of Government database[6]. This tracker tracks individual policy measures across 17 indicators and the following nine of them as shown in table1 are used to calculate the stringency index. Each sub-index score (I) for any given indicator (j) on any given day (t), is calculated by the function described in equation9 based on the following parameters:

- the maximum value of the indicator(N_j)
- whether that indicator has a flag ($F_j = 1$ if the indicator has a flag variable, or 0 if the indicator does not have a flag variable)
- the recorded policy value on the ordinal scale($v_{j,t}$)
- the recorded binary flag for that indicator($f_{i,t}$)

$$I_{j,t} = 100\frac{v_{j,t} - 0.5(F_j - f_{j,t})}{N_j} \tag{9}$$

This normalizes the different ordinal scales to produce a sub-index score between 0 and 100 where each full point on the ordinal scale is equally spaced. For indicators that do have a flag variable, if this flag is recorded as 0 then this is treated as a half-step between ordinal values. Note that the database only contains flag values if the indicator has a non-zero value. If a government has no policy for a given indicator (ie the indicator equals zero) then the corresponding flag is blank/null in the database. To calculate the index, this is equivalent to a sub-index score of zero. In other words, $I_{j,t} = 0$ if $v_{j,t} = 0$. Additionally, the data is not always fully complete and sometimes indicators are missing and they make the conservative assumption that an absence of data corresponds to a sub-index score $I_{j,t}$ of zero.

Table 1: Indicator of different policies

Indicator	Policy	Max.Value(N_j)	Flag(F_j)
C1	School Closure	3 (0, 1, 2, 3)	yes=1
C2	Workplace Closure	3 (0, 1, 2, 3)	yes=1
C3	Cancelling of Public Events	2 (0, 1, 2)	yes=1
C4	Restriction on Gathering	4 (0, 1, 2, 3, 4)	yes=1
C5	Closure of Public Transport	2 (0, 1, 2)	yes=1
C6	Stay at Home Restriction	3 (0, 1, 2, 3)	yes=1
C7	Domestic Travel Restriction	2 (0, 1, 2)	yes=1
C8	International Travel Restriction	4 (0, 1, 2, 3, 4)	no=0
H1	Public Information	2(0, 1, 2)	yes=1

The stringency index is simple averages of the individual component indicators. This is described in equation10 below where k is the number of component indicators in an index and I_j is the sub-index score for an individual indicator.

$$index = \frac{1}{k}\sum_{j=1}^{k} I_j \tag{10}$$

3 EXPERIMENT

In this section, we would fit real-world data into the SIR-F model in different phases based on the change of S-R trend and obtain the estimated R_0 through different periods. Additionally, we would estimate future trends using the most recent estimated R_0 and explore the relationship between the stringency index and estimated R_0.

3.1 Set phases for S-R trend

It this not only difficult but also inaccurate to fit the model using all the data at one time. So we would better separate the whole time interval into several phases. Before we set phases for S-R trend analysis, we would like to transform S as a function of R. Firstly we have the following equations in the SIR-F model:

$$\frac{dS}{dT} = -\beta\frac{SI}{N} \tag{11}$$

$$\frac{dR}{dT} = \gamma I \tag{12}$$

$$\tag{13}$$

After the substitution we have:

$$\frac{dS}{dR} = -\frac{\beta}{N\gamma}S \tag{14}$$

$$\tag{15}$$

Integrating 14 we get:

$$S(R) = Ne^{-\frac{R\beta}{N\gamma}} \tag{16}$$

$$\tag{17}$$

This leads to:

$$\log S_{(R)} = -aR + b, a = \frac{\beta}{N\gamma}, b = \log N \tag{18}$$

Because $R(t)$ is a nonnegative number, $\log S$ decreases constantly. The relation between $\log S_{(R)}$ and R is a straight line. The slope of this line would change if the parameters (β, γ) change. We use *ruptures* [7] to detect the change of these two parameters and thus set the phase. The period between two detected change point is defined as a phase.

3.2 Estimate Future Trend

We wanted to know how many cases will be in 30 days if the parameter values will not be changed from the last phase so we added a phase with 30 days from the date of the last record. We would estimate the future trend using the SIR-F model and the initial values would be the last number of the last phase in these four countries.

3.3 Stringency index and Rt

As mentioned above, we have successfully divided past days into different phases and obtained the estimated R_0 for each phase. We then get the stringency index of these days using the methods explained in theory and explore the relationship between the reproductive ratio of Covid-19 and strategies of lockdown in different countries.

4 RESULTS AND DISCUSSION

4.1 China

As shown in Figure1, China passed eight phases. During the initial phase and the first phase (from 8 May to 17 February 2020),

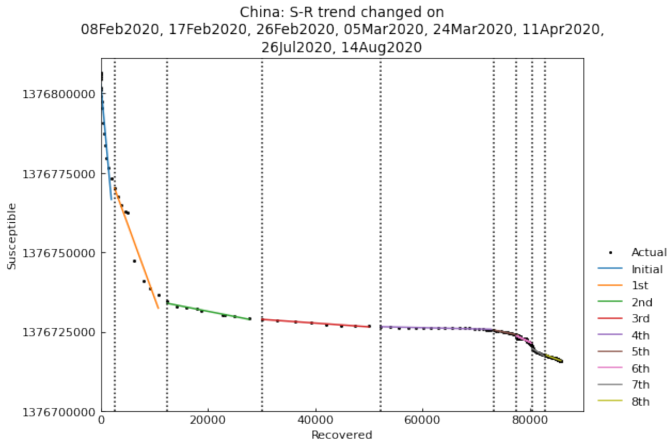

Figure 1: S-R trend for China. Using *covsirphy.phase.trend* to get 9 phases in total. The estimated R_0 for different phases is 45.66, 3.97, 0.20, 0.09, 0.04, 0.45, 0.50, 0.81, 0.21

the log($Susceptible$) dropped sharply because the number of infectious surged and peaked on 17 February in China. In the next four phases, the number of recovered increased continuously because the Chinese government put great effort into controlling the situation. Figure2 showed the changes in stringency and Rt (R_0 as a function of t) throughout ten months. The largest $R_0 = 45.66$ occurred in February because COVID-19 hit Wuhan city firstly and the government had zero preparation. The Chinese government released drastic policy quickly: Wuhan city locked down, school and workplace closed and people were required to stay at home. These quick responses prevented the first wave from getting even worse. Chinese government applied these strict policies to other cities in China so the stringency remained high until April. It was noticeable that China had a mild increase in importing cases from overseas in August; this was because a large number of students and employees returned to China owing to the second outbreak in Europe. Using the estimated parameters in the eighth phase, we predicted the numbers in the next 30 days. As the result shown in Figure3, there would not have a significant increase in all the four classes, which meant China has done a great job in preventing the second outbreak.

4.2 Japan

Figure4 showed that Japan passed through 10 phases. Actually from the start of February to the end of April, the highest $R_0 = 9.87$ last more than 80 days before dropped to a significantly low level. Figure5 also demonstrates the fact that the government slowly raised the policy stringency during the first peak of the epidemic. Though the largest stringency index was below 50, which is relatively lower than other countries, the government still controlled the situation successfully and R_0 declined significantly before May. Compared with the first response in China, we noticed the restrictions released by Japan was not effective enough, given the fact that R_0 remained peaked for three months. Interestingly, when the government slowly loosed lockdown strategies in the next months after the middle of May, R_0 rebounded in July. Then the government

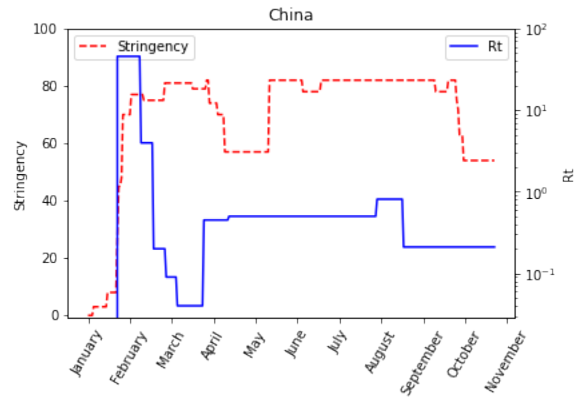

Figure 2: China: Stringency and R_t

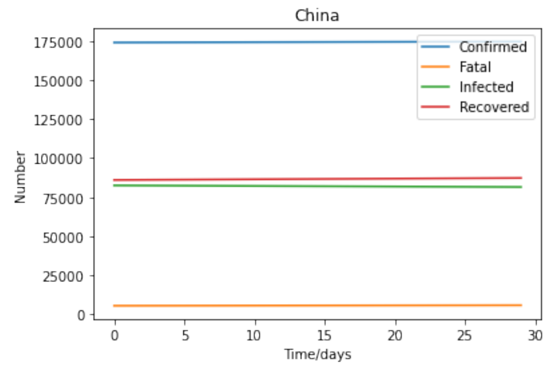

Figure 3: China: prediction between 21 Oct and 19 Nov. The predicted number of confirmed, fatal, infected and recovered at 19 Nov is 174594, 5818, 81506, 87270 respectively.

tightened policies accordingly. Because of the slight adjustment in stringency, R_0 dropped below one and remain relatively stable since then. These trends could reveal that the stringency index has a noticeable effect on the reproductive ratio but this effect is not instantly. This means there could be a lag between releasing policy and its effectiveness.

Through estimated parameters of the tenth phase, we obtained predicted numbers for the next 30 days. Figure6 told that there would be a steady increase by approximately 500 per day both in confirmed and recovered cases. This prediction meant Japan has achieved a balance between stringency and the reproductive ratio. In other words, the Japanese government, to some extent, has eased the pressure of the second outbreak but more strict policies should be published if they want to fully control the situation.

4.3 Italy

As can be seen in Figure7, Italy had two significant outbreaks through 12 phases – one in the early start of February like other countries in the world, the other one from the end of July till now. The first pandemic outbreak heightened the pressure on the Italian healthcare system and caused massive deaths so the government proposed extremely strict policies from March and R_0 slowly went down to below one at the end of April. We could also notice a second

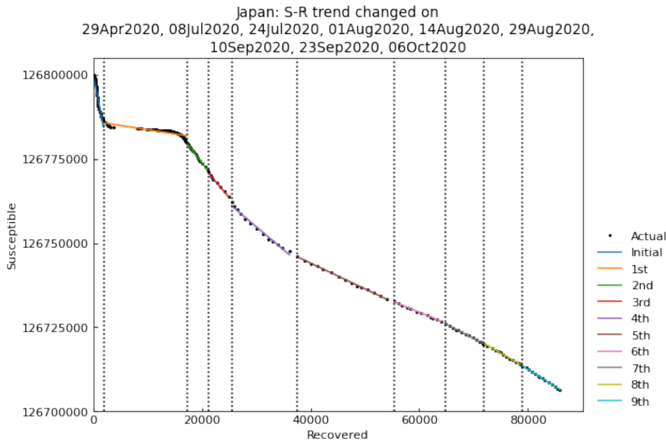

Figure 4: S-R trend for Japan. Using *covsirphy.phase.trend* to get 10 phases in total. The estimated R_0 for different phases is 9.87, 0.24, 2.12, 2.12, 1.68, 0.70, 0.62, 0.73, 0.89, 1.00

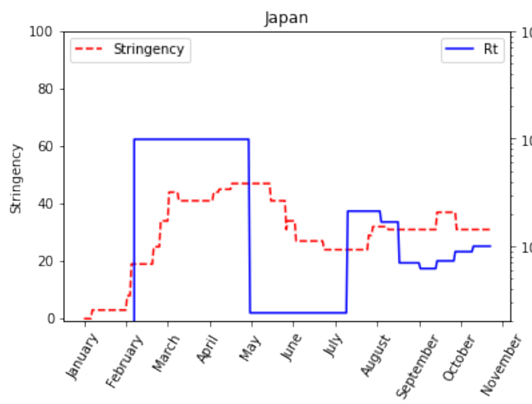

Figure 5: Japan: Stringency and R_t

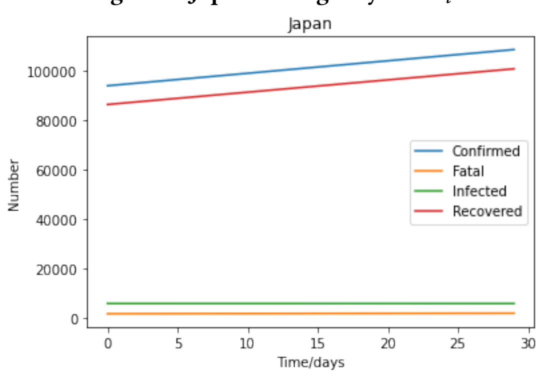

Figure 6: Japan: prediction between 21 Oct and 19 Nov. The predicted number of confirmed, fatal, infected and recovered at 19 Nov is 108778,1943,5857,100978 respectively.

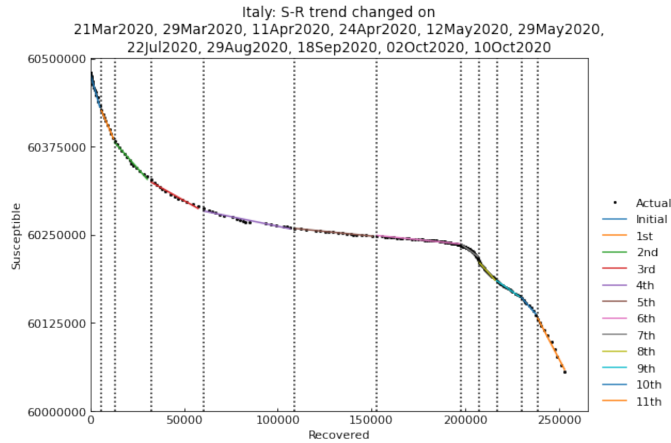

Figure 7: S-R trend for Italy. Using *covsirphy.phase.trend* to get 12 phases in total. The estimated R_0 for different phases is 21.26, 4.52, 2.36, 1.46, 0.47, 0.23, 0.24, 1.45, 2.70, 1.74, 2.65, 4.50

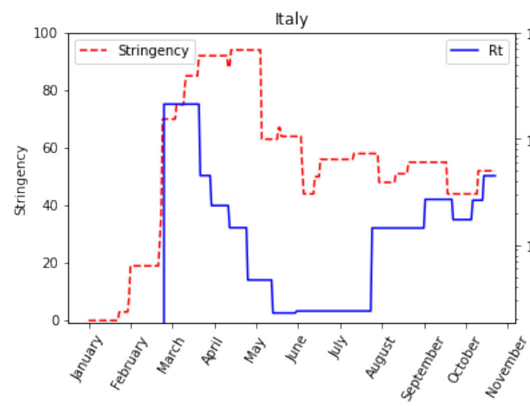

Figure 8: Italy: Stringency and R_t

wave in Figure8. As the stringency decreased from May, R_0 started to see a significant rise to above one at the end of July and even grew larger and larger till now but we cannot see the effectiveness of the government's reaction due to the relatively remaining stable stringency.

Thus, we used the estimated parameters in the twelfth phase to predict the numbers in the next 30 days. As shown in Figure9, the confirmed cases would increase by more than 12000 per day. The real-world situation is even worse than we predict when comparing the public data on 21/22/23 Oct with our estimation. The government of Italy is in urgent need to propose more strict lockdown regulations otherwise this second wave would cause more severe effects than the first one.

4.4 Singapore

Singapore is one of the best countries which controlled the outbreak of COVID-19 well. According to Figure10, Singapore only had one significant outbreak in May and a subtle fluctuation in

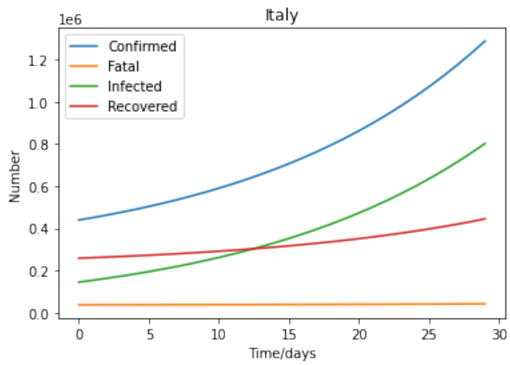

Figure 9: Italy: prediction between 21 Oct and 19 Nov. The predicted number of confirmed, fatal, infected and recovered at 19 Nov is 1287754,41588,801899,444267 respectively.

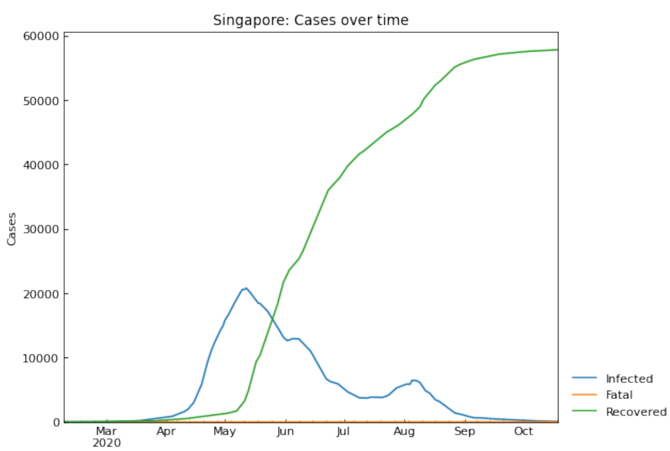

Figure 10: Singapore: cases from February to October.

early August. In Figure11, there were 10 phases detected in Singapore. The S-R trend for Singapore was similar to that in China, which had a sharp decrease in the early stage and declined slowly in the later stages. As we can see in Figure12, the government of Singapore released a series of restrictions from January to May to enhance their stringency, started by cessation flight from Wuhan and mandatory quarantine for travelers. Since the first outbreak controlled well, the Singapore government reopened school and workplace in June but the restriction for the gathering was still at a high level. Owing to the successful control of the Singapore government, the increase in confirmed and recovered number were subtle in our prediction. The recovered curve almost covered the confirmed coincidentally in the Figure13, as these two data were very close.

5 CONCLUSION

As COVID-19 seems to spread continuously for a long time, the governments should continue to deal with this severe situation. In this report, we built the SIR-F model to fit the spread of COVID-19 in China, Japan, Italy, and Singapore from 1 January to 20 October in 2020. Through *covsirphy.simulation.estimator* we detected several

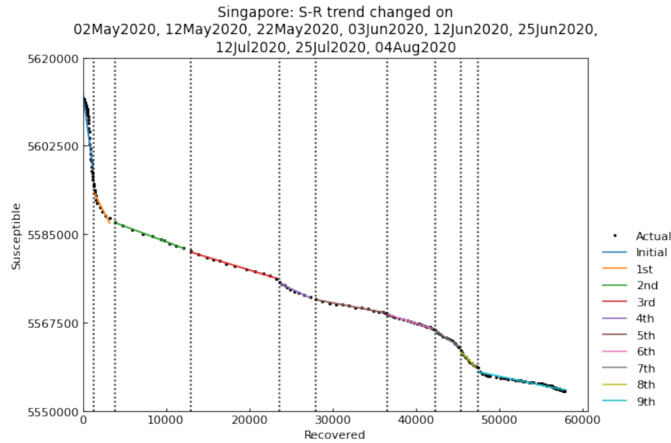

Figure 11: S-R trend for Singapore. The estimated R_0 for 10 phases is 0.02, 3.22, 0.55, 0.49, 1.07, 0.35, 0.46, 1.13, 1.57, 0.45.

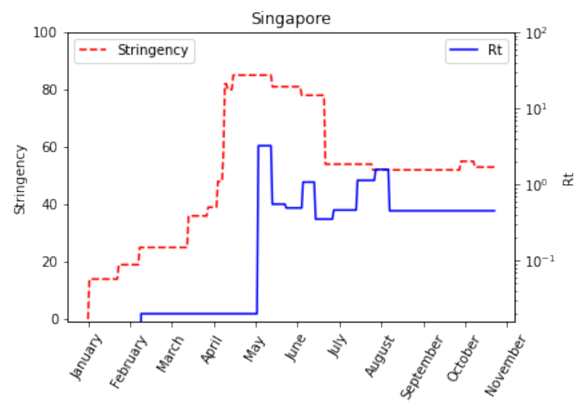

Figure 12: Singapore: Stringency and R_t.

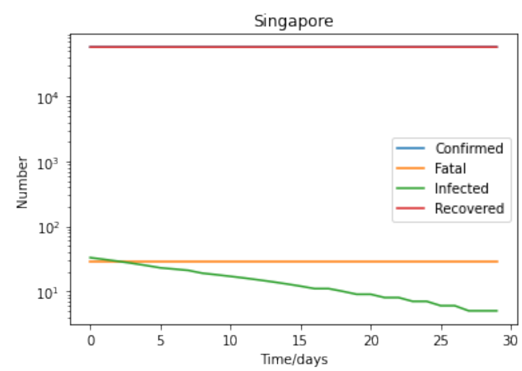

Figure 13: Singapore: prediction between 21 Oct and 19 Nov. The predicted number of confirmed, fatal, infected and recovered at 19 Nov is 58052, 29, 5, 58018 respectively.

phases and estimated the parameters $(\theta, \kappa, \rho, \sigma)$ respectively. Based on the parameters in the last phase, we predicted the number of confirmed cases, fatal, infected, and recovered in 30 days. China and Singapore have been doing well in preventing the second breakout since August, whereas Italy is suffering from the second outbreak and the government should apply more restrictions as they did in April. As for Japan, the government should publish stricter restrictions to deal with the increasing R_0 since September. We also showed that the stringency level has a great impact on reproduction rate R_t: higher stringency could lower R_t. Besides, there could be a lag between releasing policy and showing its effectiveness. This lag also varied among different countries.

We still have a lot of things to do in future research. First of all, we explained the relation between stringency and R_t intuitively; we would like to apply quantitative analysis to these two indexes. Then, we are going to figure out which is the most effective measure to lower the reproduction rate R_0. Last but not least, we could estimate the parameters in the SIR-F model more frequently because of the spread of COVID-19 and the government's responses change quickly. More frequent estimation, for example 7 days, might lead to more accurate R_0.

REFERENCES

[1] MAMT Baldé. 2020. Fitting SIR model to COVID-19 pandemic data and comparative forecasting with machine learning. *medRxiv* (May 2020). https://doi.org/10.1101/2020.04.26.20081042

[2] Dimitris I Papadopoulos; Ivo Donkov; Konstantinos Charitopoulos; Samuel Bishara. 2020. The impact of lockdown measures on COVID-19: a worldwide comparison. *medRxiv* (June 2020). https://doi.org/10.1101/2020.05.22.20106476

[3] CovsirPhy Development Team (2020). 2020. *CovsirPhy*. Python package for COVID-19 analysis with SIR-derived ODE models. https://github.com/lisphilar/covid19-sir

[4] CG Antonopoulos I Cooper, A Mondal. 2020. A SIR model assumption for the spread of COVID-19 in different communities. *Elsevier* 139 (October 2020 2020). https://doi.org/10.1016/j.chaos.2020.110057

[5] Di Gennaro F.; Pizzol D.; Marotta C.; Antunes M.; Racalbuto V.; Veronese N.; Smith L. 2020. Coronavirus diseases (COVID-19) current status and future perspectives: a narrative review. *Int. J. Environ. Res. Public Health* 17 (April 2020).

[6] Blavatnik School of Government. 2020. *Coronavirus Government Response Tracker*. University of Oxford. Retrieved October, 2020 from https://covidtracker.bsg.ox.ac.uk/

[7] Charles Truong, Laurent Oudre, and Nicolas Vayatis. 2020. Selective review of offline change point detection methods. *Signal Processing* 167 (2020), 107299.