Covid Spread Modelling

Zheng Ren

2023-05-08

Set Up Section

To reproduce the simulation results, the following steps should be followed:

1. install the package from GitHub and import the package:

```
devtools::install_github('Zheng206/Covid19/covidmc')
library(covidmc)
data(covid_df)
```

2. read in simulation results for the purpose of illustration, download addresses are as follows:

- **UK bootstrap results:**
  https://raw.githubusercontent.com/Zheng206/Covid19/main/data/uk_tune.csv
- **US bootstrap results:**
  https://raw.githubusercontent.com/Zheng206/Covid19/main/data/us_new.csv
- **Time Statistics results:**
  https://raw.githubusercontent.com/Zheng206/Covid19/main/data/time_statistics.csv

3. R CMD Check

Introduction

The coronavirus disease 2019 (COVID-19) epidemic was officially declared as a pandemic in early March 2020 by WHO and has caused heavy life losses and damages to social-economics globally. A better understanding of the spread of virus is essential for preventing potential eruptions of the disease, better allocating social and medical resources, and diminishing the overall time that a pandemic lasts.

This study is based on the paper "A novel Monte carlo simulation procedure for modelling COVID-19 spread over time"(Xie 2020), aiming to develop a COVID-19 spread dynamics model that can

be used as a decision making tool for disease control. The proposed method is characterized by a Monte Carlo Simulation model, which captured infection rate and population changes, and treated each individual in a population as a random point.

This study builds on top of Xie's model. Beyond estimated daily active cases and new cases, the proposed model includes daily estimated total number of confirmed cases in the outputs as well for parameter tuning. This study also used real world COVID data for a more accurate infection rate pattern and population limits, aiming to better adapt the proposed model to the real world situation.

Method

The key parameters and assumptions of the proposed model were: (1) the infection rate parameter "Rt"; (2) the actual number of people being infected by each active case is determined by a Poisson distribution with the mean Rt; (3) the exact number of days for someone getting infected follows a negative binomial distribution with the mean/expected time "muT" in days; (4) The limit of the number of people and the immunity proportion in a study population; (5) Other initial conditions including the initial number of infectious people, the observation period (in days), and the simulation period (in days) of a simulation study.

Real world COVID data (https://ourworldindata.org/coronavirus) is then applied to tune model parameters for the country of interest. We used the real infection rate since March 1, 2020 as our infection rate pattern in the proposed model, and used country estimated population as the population limits. We assumed the immunity proportion to be 0.001 and the initial number of infectious individuals to be the sum of newly confirmed cases over the period of two weeks prior to 1 March, 2020. Finally we tried different ranges of muT and sizeV, and selected the ones that minimize the sum of squared difference between the real daily total cases and the simulated total cases.

To count for the uncertainty in a Monte Carlo simulation study, a bootstrap approach was followed to obtain the estimated median and interquartile values of the number of the active cases and the daily new cases over the observation period. Each simulation model was run for 100 times and the median values were considered as the most likely estimation. The uncertainty level was characterized by the interquartile range.

Simulation Replication

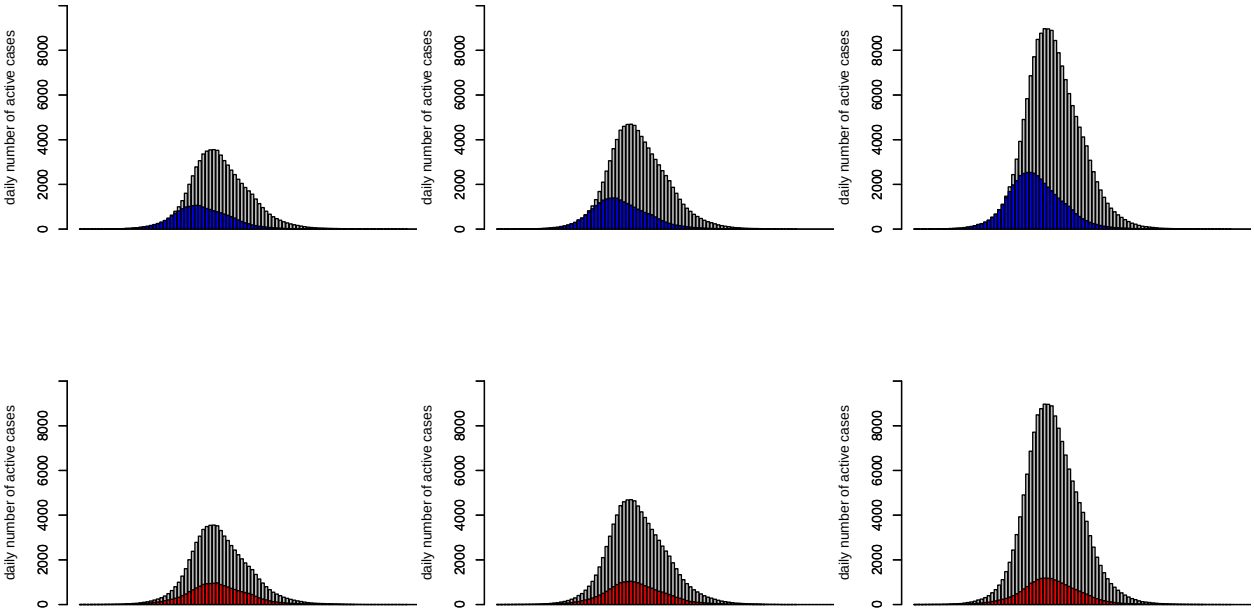
This section replicates Xie's simulation results for 12 disease transmission scenarios in the paper "A novel Monte Carlo Simulation procedure for modelling COVID-19 spread over time". This simulation is based on three hypothetical infection rate patterns and three different simulation settings. The three hypothetical infection rate patterns are as follows:

- **pattern 1:** rr = c(rep(3,30), 2,2,1.5,1.5, 1.2,1.2,rep(0.8,4),rep(0.5,10),rep(0.1,50))
- **pattern 2:** rr = c(rep(3,25), 2,2,1.5,1.5, 1.2,1.2,rep(0.8,4),rep(0.5,10),rep(0.1,55))
- **pattern 3:** rr = c(rep(2.5,30), 2,2,1.5,1.5, 1.2,1.2,rep(0.8,4),rep(0.5,10),rep(0.1,50))

The three simulation settings are:

- nd = 100, muT = 4, sizeV = 1
- nd = 100, muT = 4, sizeV = 0.9
- nd = 100, muT = 3.6, sizeV = 1

The three panels in the top row showed that by only five days earlier of enforcing intervention, the number of the active cases would peak earlier accordingly with a much lower peak value (black curves versus blue curves). The three panels in the bottom row showed that by reducing Rt from 3 to 2.5 for the first 30 days, the number of the active cases would reach the peak on the same day but with a much lower peak value (black curves versus red curves). Reducing the muT values (3.6 versus 4 days) would increase the peak level quite substantially.



Simulation Time Statistics Summary

To have a better understanding of the time the proposed model requires for a single iteration, we tested on two simulation settings. One simulation setting has a simulation period of 100 days and the second one has one of 200 days. As we can see, the time a single iteration take to run significantly increased when we doubled the simulation period. Therefore, it is rather essential to include parallel processing or even high performance computing clusters in the step of tuning parameters and bootstrapping.

Table 1: Time Statistics Table

expr	min	lq	mean	median	uq	max	neval	cld
TransSimu 100	44.834	122707.9	305584.9	216799.5	440830	1492320	100	a
TransSimu 200	52.250	12233080.9	11815451.9	12360211.1	12554110	13466216	100	b

Simulation Extension on US and UK data

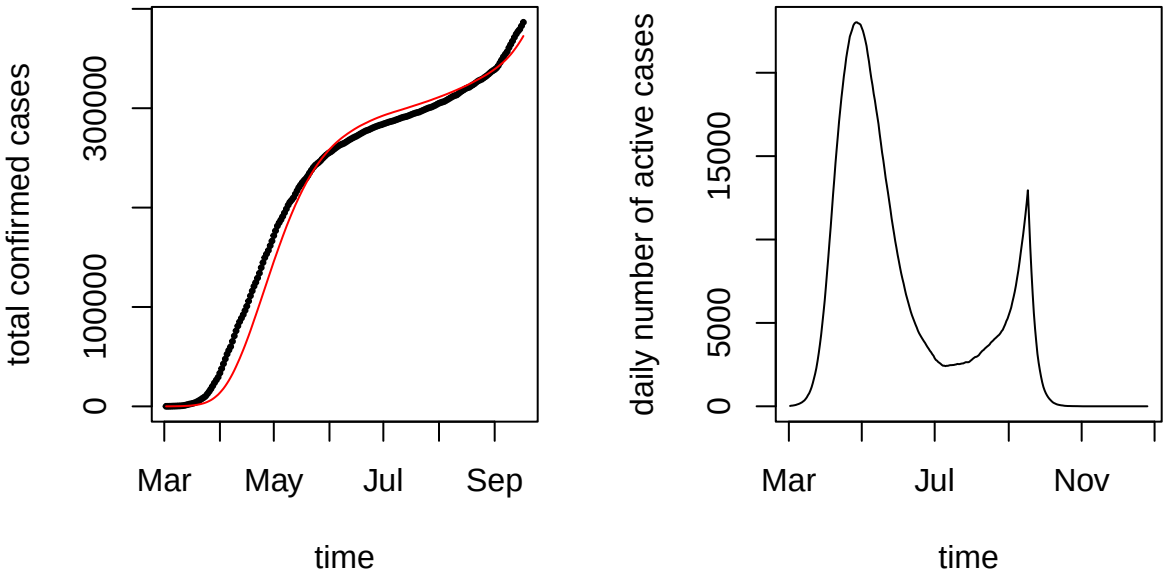
This section uses US and UK real COVID data to better tune the model, aiming to get a more accurate prediction of daily active cases and new cases for better policy decision making. Based on the spread model outputs, we can better understand how infection rate and population limits affect daily active cases and new cases in these two countries.

UK

Model Set Up

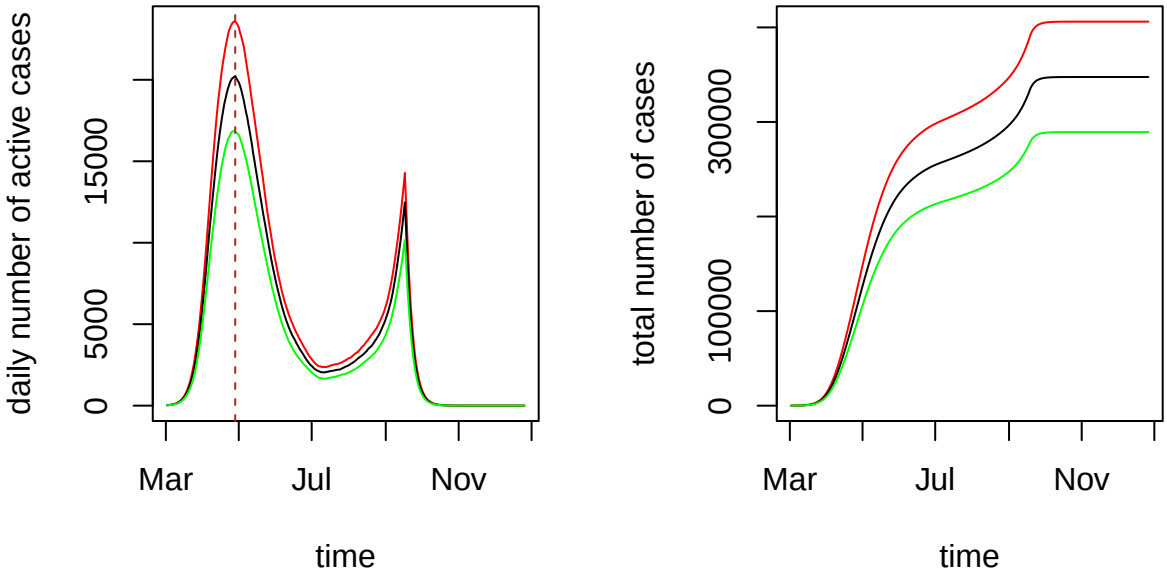
To first set up the model, we chose 300 days as our observation period and set a simulation period of 200 days. We used the real COVID data from March 1, 2020 to December 16, 2020 for both US and UK.

After applying UK data to the model, we observed that the simulated daily total number of cases are closest to the actual ones when parameter muT is 3.65 and parameter sizeV is 0.8. As we can see on the left plot, the black dots represent the true total confirmed cases in the UK, whereas the red line represents model predictions. It seems that the proposed model slightly underestimated total confirmed cases before May 2020, but overlaps with the true data pretty well after June 2020.



Based on the model outputs, we identified 2 peaks of daily active cases, one is 2020-04-27, the other is 2020-09-17. Comparing the two peaks, the first peak is worse than the second one.

Bootstrap Results

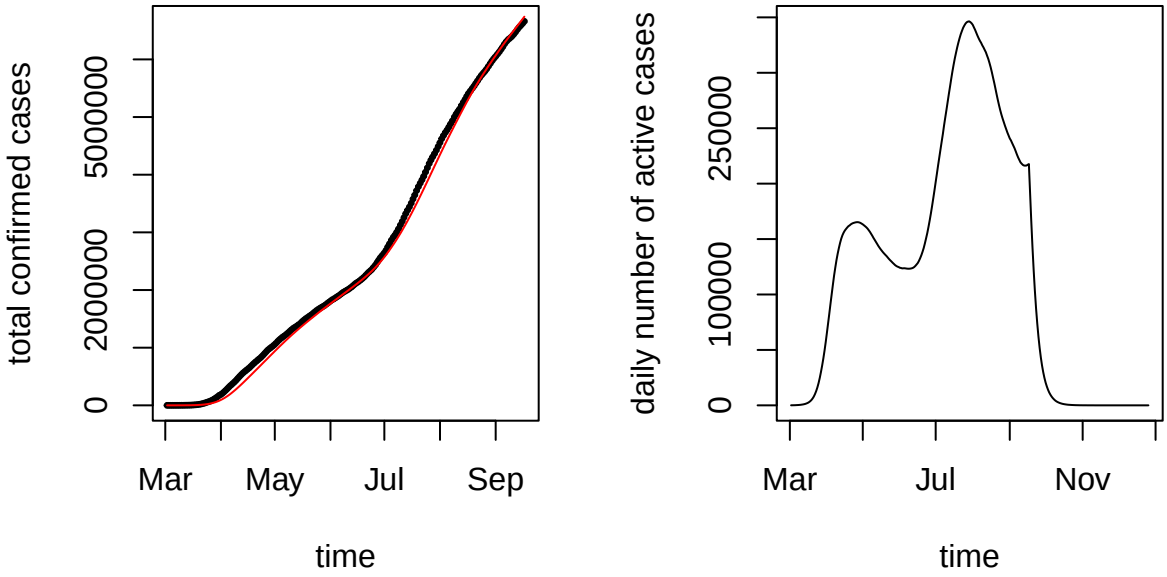


The bootstrap result shows that the first peak of daily active cases is around 2020-04-28 and the peak value is 20236; the second peak of daily active cases is around 2020-09-17 and the peak value is 12474.

US

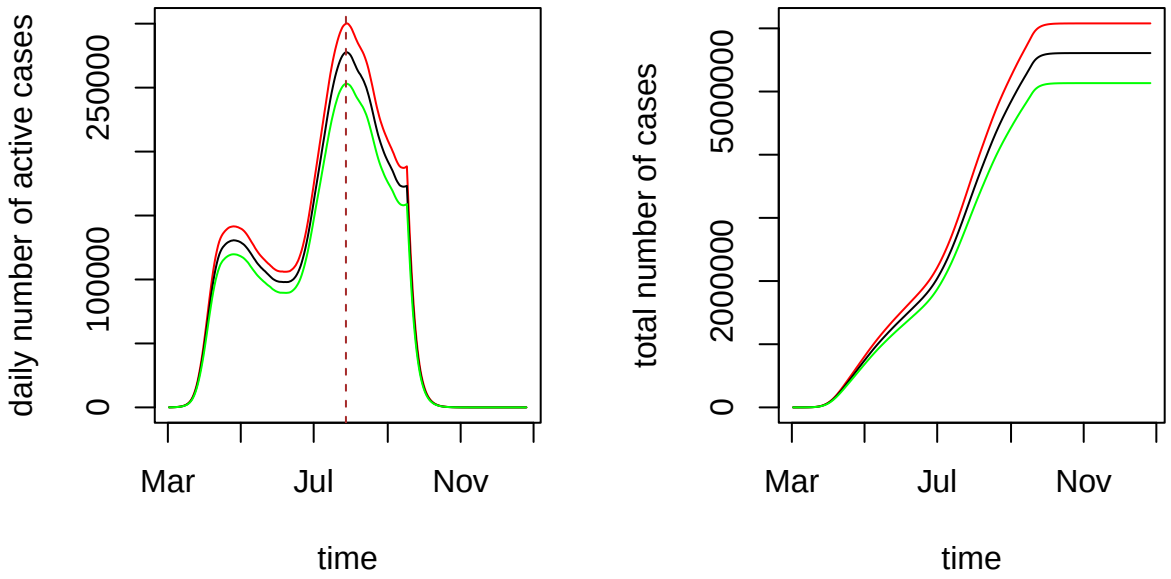
Model Set Up

After applying US data to the model, we observed that parameter muT is best to be 5 and parameter sizeV is best to be 1.1. As we can see in the left plot, the red line overlaps extremely well with the black dots.



Based on the model outputs, we also identified 2 peaks of daily active cases, one is 2020-04-26, the other is 2020-07-29. The second peak is worse than the first one.

Bootstrap Results



The bootstrap result shows that the first peak of daily active cases is around 2020-04-25 and the peak value is 130650; the second peak of daily active cases is around 2020-07-28 and the peak value is 277662.

Simulation Result Discussion

Based on UK and US simulation results, we can see a clear difference in the daily active cases pattern. Multiple factors can contribute to this difference, but we focused more on three of them: (1) the mean time infectious people can infect others, (2) population limits and (3) infection rate pattern in this study.

The mean infectious time (muT): The United Kingdom has a mean infectious time of 3.65 days whereas the United States has a mean infectious time of 5 days. In other words, an infectious individual in US poses a longer threat to the rest of people in the US.

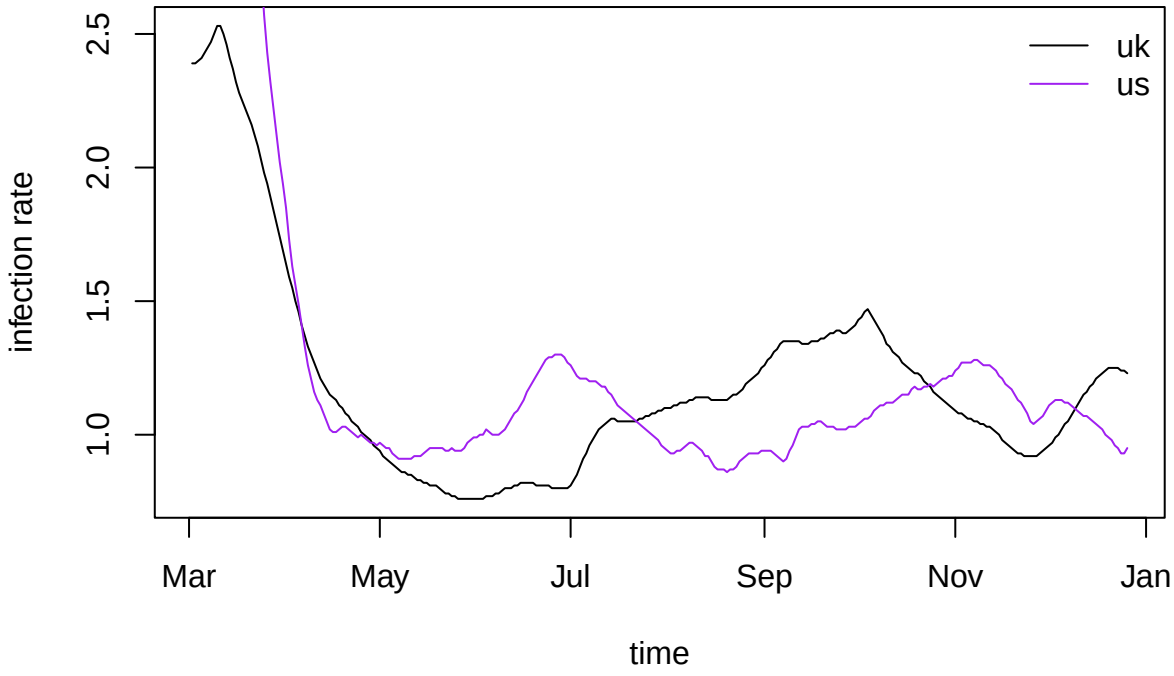
The population limits: UK has an estimated population of 67508936 whereas US has an estimated population of 338289856. The larger population in the US also explains a general greater number of daily active cases in the US, especially at peaks.

The infection rate pattern (Rt): In general, UK has a smaller infection rate over time, as we can see in the quantile summary table (Table 2). However, the different infection rate patterns over time in both countries are associated with different peak times of daily active cases.

Table 2: Infection Rate Summary Table

	0%	25%	50%	75%	100%
uk_infection	0.76	0.94	1.11	1.31	2.53
us_infection	0.86	0.97	1.06	1.21	3.61

In the UK, the infection rate continuously increased from June to October in 2020. This increases seems to explain the second peak in September; In the US, the inflation rate stopped dropping at an infection rate around 1 in May 2020 and continuously increased afterwards until July 2020. This infection rate pattern seems to be associated with the second peak in July 2020. Therefore, a continuous increase of infection rate seems to contribute to a peak of daily active cases. It is recommended to control the time that a continuous increase of infection rate last in order to prevent the eruption of active cases.



Discussion

This study provides us a tool to better predict the spread of covid among people based on historical data. The proposed simulation model also shows a good potential in performing what-if analysis for decision making for combating COVID-19 in specific and any other infectious diseases in general.

However, this study also has some limitations. First of all, the current approach heavily relies on assumptions and historical data. Further researches should be conducted to test if the current approach is still robust when some assumptions are violated. Besides, the proposed model currently takes a lot of computing power especially when we have a long simulation period. Further improvements are expected to increase the efficiency of this model. Lastly, a better cost function should be designed to better tune model parameters.

Reference

Xie, Gang. 2020. "A Novel Monte Carlo Simulation Procedure for Modelling COVID-19 Spread over Time." *Scientific Reports* 10 (1). https://doi.org/10.1038/s41598-020-70091-1.