

COVID-19 Analysis

Introduction

Through the analysis of Korean, Italy and US data to detect the lagging effect and predict what will happen next in US.

Data and Analysis

Date:

The data are all from Kaggle or from GitHub

- Korea data: <https://www.kaggle.com/kimjihoo/coronavirusdataset>
- Italy data: <https://www.kaggle.com/sudalairajkumar/covid19-in-italy>
- USA data: <https://www.kaggle.com/sudalairajkumar/covid19-in-usa>

Visualization:

From those raw datasets, we extracted the daily update on the cumulative total of test cases, positive cases, released cases and decreased cases. The three plots in appendix show the time series of those four tracking metrics in the 3 countries. However, those plots are not easy to digest or to register. To remove the effect of the scale(population) and the temporal effort, we calculated the DoD change on the cumulative positive cases [see Fig.1.]. To better visualize the plots, we scale the time periods all to 60 days. The upper left is for Korea from date 2020-01-21 to 2020-03-12, the upper right is for Italy from date 2020-02-24 to 2020-03-16, while the bottom one is for US from 2020-03-03 to 2020-03-15. The intuitive explanation is that when the DoD is bigger than 1, we will see an increase in the total confirmed cases, while DoD is smaller than 1, we will see a decrease in the total confirmed cases.

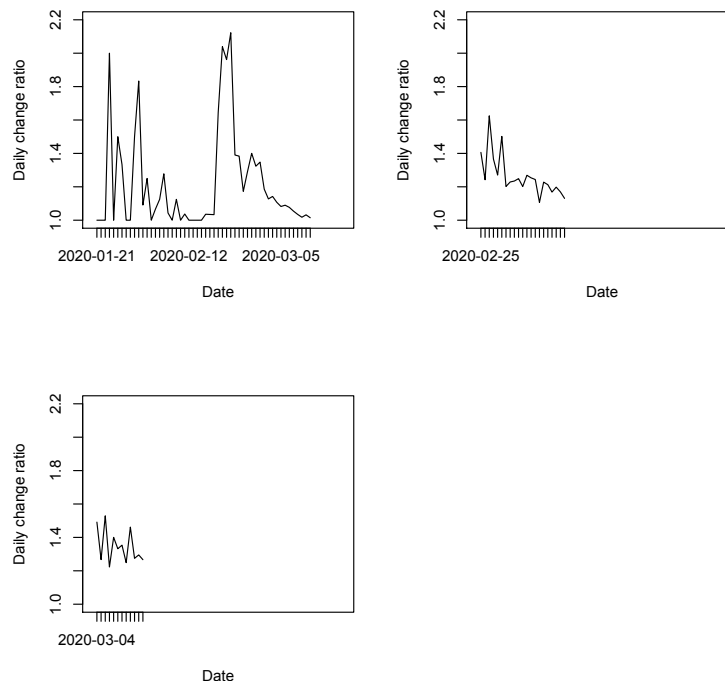
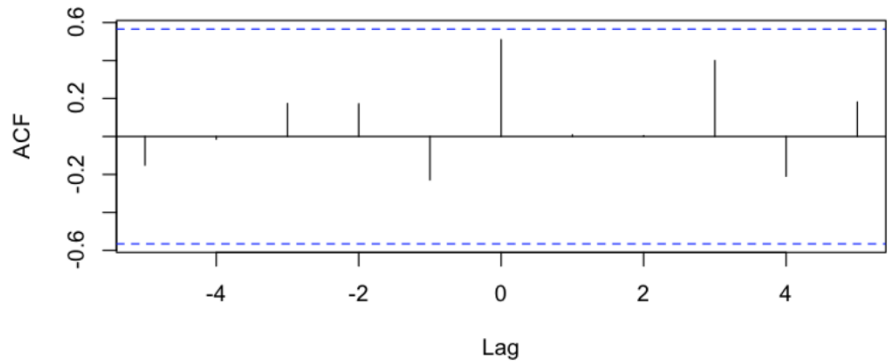


Figure 1: DoD (%) change of total confirmed cases in Korea (upper left), Italy (upper right) and US (bottom).

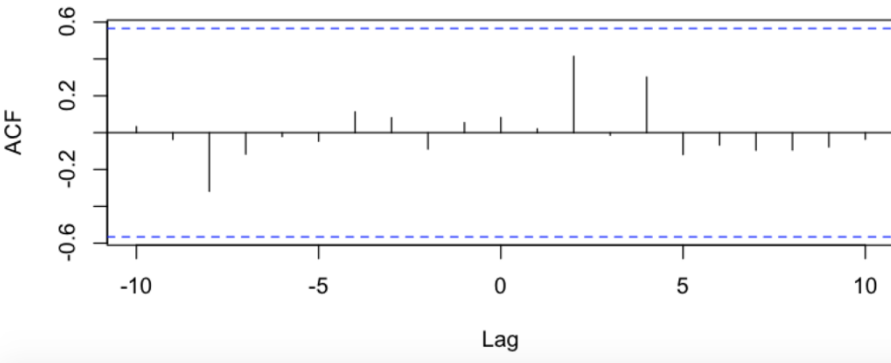
Through the visual inspection, we can see the Italy curve and US curve look similar. We used sample cross correlation function (CCF) to identify the lagging of between those two curves and confirmed with our

26 hypothesis. The correlation coefficient with lagging 0 is the largest with value at 0.51 [See Fig.2.]. Given the date
27 difference in the dataset, we can say 2020-02-25 data in Italy is similar to the 2020-03-04 data in US (**Italy is**
28 **leading US by around 10 days**).



29
30 *Figure 2: CCF between DoD change in Italy and US.*

31
32 For Korean data, we restrict the data from 2020-02-18 onward and run the same CCF. We can say 2020-02-20
33 data in Korea is similar to the 2020-03-04 data in US (**Korea is leading US by around 15 days**).



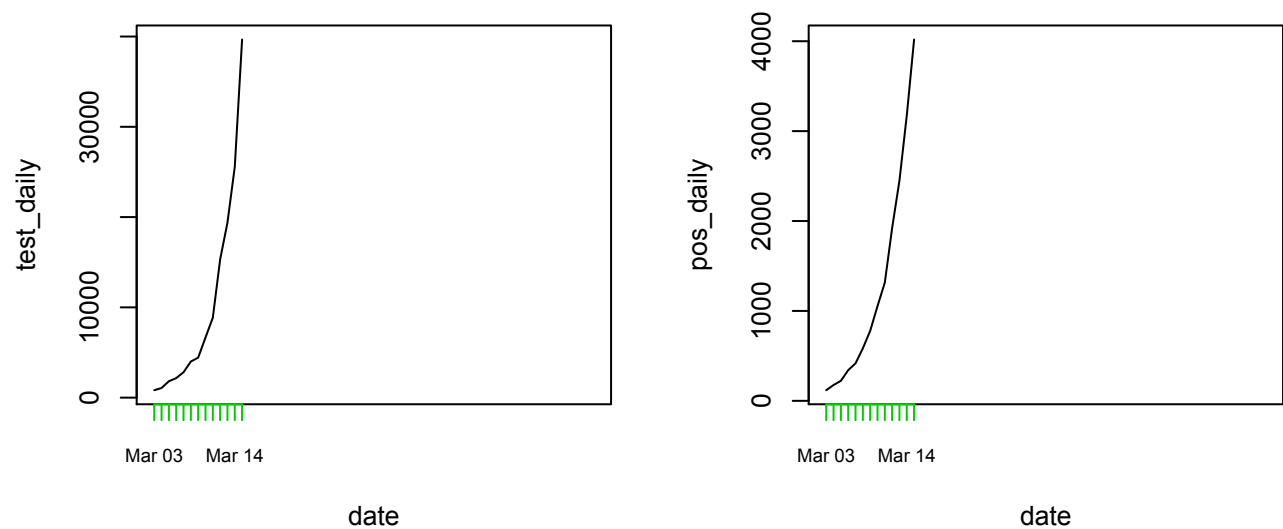
34
35 Projections:
36 Based on the learned lagging, we tried to run the projection use Korean and Italian's DoD data [See Fig.3].
37

38 **Next**

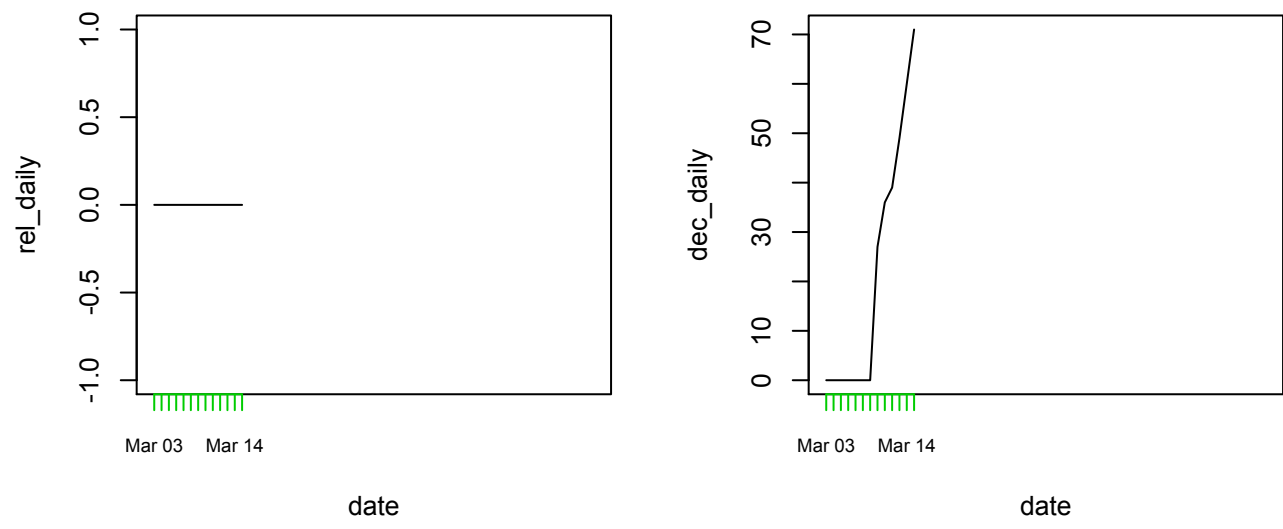
39 More analysis can be done on the test capacity, death rate and relationship among all the metrics.

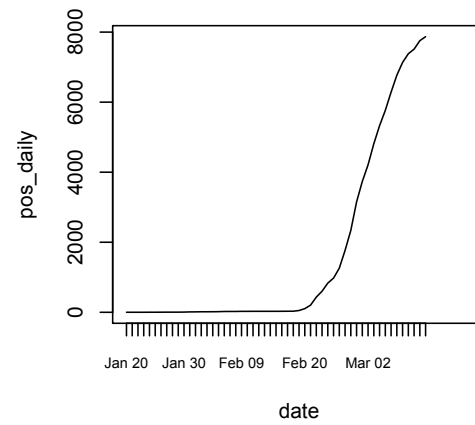
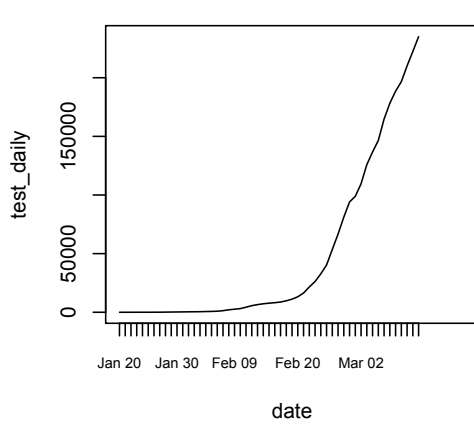
Date	Based on Korea	Based on Italy
3/3/20	118	118
3/4/20	176	176
3/5/20	223	223
3/6/20	341	341
3/7/20	417	417
3/8/20	584	584
3/9/20	778	778
3/10/20	1053	1053
3/11/20	1315	1315
3/12/20	1922	1922
3/13/20	2450	2450
3/14/20	3173	3173
3/15/20	4019	4019
3/16/20	4592	5038
3/17/20	5084	6266
3/18/20	5502	6933
3/19/20	5996	8513
3/20/20	6457	10325
3/21/20	6807	12065
3/22/20	7044	14454
3/23/20	7169	16906
3/24/20	7400	19115
3/25/20	7508	

Figure 3: Projects with Different scenarios. I am hoping we will follow Korean trajectory, but based on 03/16 and 03/17 data, US is more closely tracking Italy ;(.

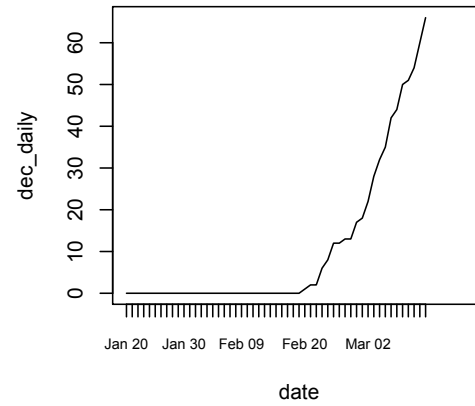
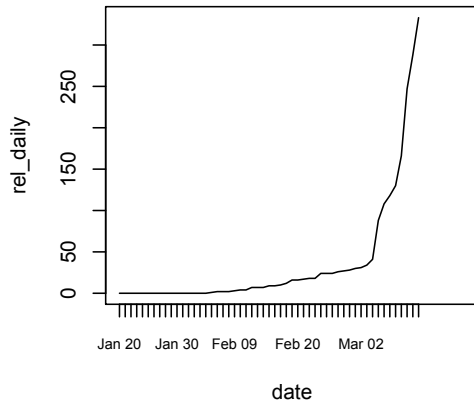


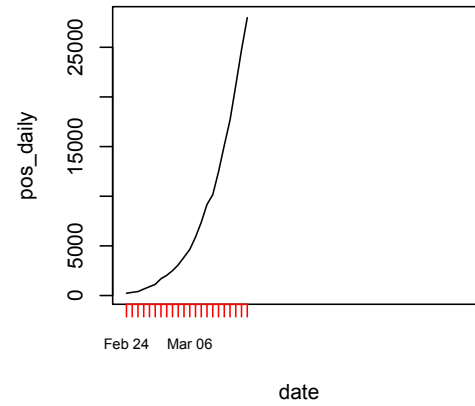
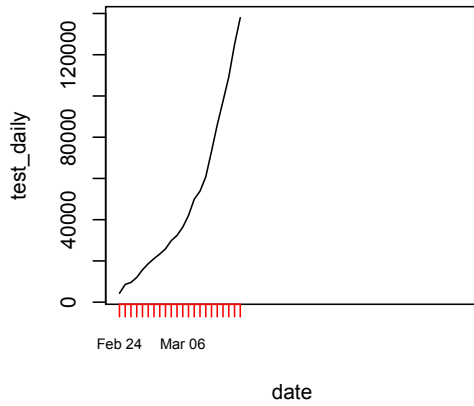
**USA: Time Series(total tested, positive case, released, deceased)
from 2020-03-04 to 2020-03-16**





**Korea: Time Series(total tested, positive case, released, deceased)
from 2020-01-20 to 2020-03-12**





**Italy: Time Series(total tested, positive case, released, deceased)
from 2020-02-24 to 2020-03-16**

