How much did restricting mobility help in controlling the spread of COVID?

Satheesh Joseph, Paco Valdez, Yi Zhang

Introduction

Overview

The Covid-19 pandemic has taken more than 2.5 million lives worldwide. In the U.S. alone, more than half a million people died because of it. COVID-19 is thought to spread mainly through close contact from person to person, including among people who are physically near each other (within about 6 feet).[1]

As the virus spread, many cities and other areas have implemented a shutdown policy that significantly restricted the mobility of the population. Specifically, many companies have opted for a complete work-from-home policy.

As a team of Data Scientists, we're interested in assessing the effectiveness of restricting mobility in controlling the spread of Covid-19.

Research Question

Rather than investigating the effect of social distancing on a micro and personal interaction level, we are interested in its effect on a policy level. Specifically, the Research Question we're asking is:

`How much did restricting social mobility help in controlling the spread of COVID?`

The basic causal theory we're working under can be expressed by the diagram below:

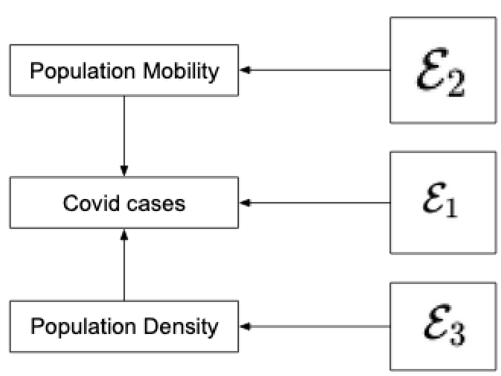

Figure 1: The Basic Causal Theory

Operationalization

To operationalize the research question, we propose the following plan for the causal models.

Firstly, we will use the New York Times Covid-19 data for the number of Covid cases per State. And we will use the COVID-19 Community Mobility Report for the population mobility scores. For the base model, we will use the "residential mobility score" as a relatively stable representative of the general population's social mobility.

Secondly, mobility have drastically different effect on the spread of the virus depending on the population density. Thus, we believe that having the variable mobility alone as the causal factor would be misleading. Therefore, we feel that the population density per sq. mile should be a mandatory variable for all the models. This data will come from the COVID-19 US State Policy Database.

And Finally, we noticed that the Covid cases metric is a time-series data. Given that Covid symptoms typically start between 2 to 14 days after exposure[1], we would build in a time delay between the time of the mobility score and the Covid cases..

We would break the time from March 2020 to March 2021 into 52 weeks, and stagger the mobility score and the number of cases by 1 week. Therefore the simplest linear model would be:

$$total\_cases\_in\_week_{i+1} = \beta_0 + \beta_1 average\_mobility\_in\_week_i + \beta_2 population\_density$$

For example, one specific data point would be:

```
x1 = average mobility score of California between 2020-03-01 and 2020-03-06
x2 = population density of the California
y = total number of cases in California between 2020-03-07 and 2020-03-13
```

For a more advanced model, we propose to use some control variables to assess the difference mobility score makes in relation to other factors, notably we'll include whether the "Stay at Home" order is in effect and whether the face mask policy is in effect.

For the most inclusive model, we will explore all the other mobility scores, including to workplaces, parks, grocery & pharmacy, as well as retail & recreation.

The models

Data Exploration

The Linear Models

The Regression Table

Model Limitations

Classic Linear Model assessment

Omitted Variables

Conclusion

References:

[1]: https://www.cdc.gov/coronavirus/2019-ncov/faq.html