Final Project

By Catherine Jackson, Hussen Mohammed Ibrahim, Ivanna Kreshchenetska

---

CS130 - Statistical Modeling: Prediction and Causal Inference

Prof. Diamond

April 21, 2023

Summary

This report will focus on the paper ‘Stay Home Save Lives: A Machine Learning Approach to Causal Inference to Evaluate Impact of Social Distancing in the US’ by Syed Muhammad Ishraque Osman and Nazmus Sakib (published by ResearchSquare, 2021).

The paper by Osman & Sakib (2021) focuses on answering the causal inference question of whether social distancing methods are effective at ‘slowing down the contagion of a pandemic’. Specifically, they are looking at the impact of social distancing measures on the spread of Coronavirus-19 (COVID-19) in the United States.

The authors use a treated group consisting of states that implemented social distancing measures prior to March 26th, 2020. They employ a control pool (from which they construct a control group) which consists of states that implemented social distancing measures on or after March 26th, 2020. The treatment date is considered to be March 26th, 2020.

One of the problems that they faced when conducting this experiment was that we don’t have a single treated unit and a pool of control units; we have multiple treated units. And this units didn’t even start receiving the treatment at the same time, so these two caveats add complexity to the problem and prevent us from cinducting a simple synthetic control to estimate the treatment effect.

They used a combination of different methods to account for the problems mentioned above. To address the fact that we have multiple treated units and we want to estimate the treatmnet effect for all of them, they created a generalized treated unit from the multiple individual treated units and matched this to the synthetic unit that was also created from the pool of donor states. Since the traditional synthetic control method works for a single treated unit, they used the generalized synthetic control method to address this issue.

A generalised synthetic control method (GSCM) was used here to determine the impact of social distancing measures on the spread of Covid-19 in the United States. Like matching, GSCM aims to create a control group that is similar to the treated group, similar here meaning that the distribution of covariates/confounding variables is significantly similar between the two groups. Osman & Sakib (2021) use a GSCM to create a control group from the control pool using a weighted combination of states. The method aims to create a control group that matches the covariates for the treated group as closely as possible in the pre-treatment period (before March 26th, 2020).

The second issue as mentioned above was the fact that different states implemented the treatment at different points in time. They address this problem by using methods that are commonly used in panel data analysis. When we have data for units that are fundamentally different individually but across time as well, we can use fixed effects to control for those differences. In our case, the states are different from each other individually but also they

changed across time as well. To account for this, the used state and time fixed effect models, both separately and in combination.

The state fixed effect model is used to address for factors that vary across states but stay the same for a given state over time. The population of a state is a good example of this. It doesn't vary much across time for a given state (and thus its name – state fixed effect), but varies across states and is an important confounder in our study. By using state fixed effects, they're able to control for the impact of this confounder.

The time fixed effect model on the other hand is used to control for factors that remain the smae across states but vary over time. This any covariate that remains the same for all states at a given moment in time but varies over time would be controlled for using this model. The level of preparedness including the availability of masks for example would be something that we can control using this model. This and other federally administered responded to the pandemic in general can be controlled for using this model since they vary over time but not across individual states.

Finally, a combination of the state and time fixed effect models above was used along with the generalized synthetic control method to account for the different challenges mentioned earlier. Even though using this combination of methods is useful for accounting for the different biases that might affect our results, they didn't fully exploit the advantages that comes from using panel data and having access to treatment status over time. The limitations of their model are discussed in more detail in the critiques section below.

The data for this paper was accessed via Harvard Dataverse and contains two datasets, one with deaths per day per state, and the other with cases per day per state. You can find the raw csv file with the dataset with COVID-19 case counts here and the dataset with the deaths here. They were preprocessed and then combined to form one dataset containing new columns with death and infection rates as well as cumulative cases.

Critiques

Though Osman & Sakib (2021) found a statistically significant difference between control and treatment outcomes, there are some flaws in their methodology that are worth pointing out.

1. Social distancing measures are effective at 'slowing down the contagion of a pandemic' is the conclusion made by the researchers. However, this is not necessarily the case. They are specifically researching the impact of social distancing measures on the spread of COVID-19 in the United States, thus their conclusions are only valid within this scope.

2. The authors fail to mention the assumptions that they have made using the GSCM.
    2.1. It relies on the assumption that the treated states are similar to the control states prior to the intervention. This includes unobserved confounders. There could, for example, be some other factor that heavily influenced the impact of social

distancing measures on the spread of COVID-19. For example, the treated states could have more individuals that take other precautions to prevent the spread of COVID-19. In this case, the authors' results do not accurately represent the causal impact of social distancing measures specifically.

2.2. GSCM also assumes that states are completely independent, which is not necessarily the case. As soon as the first state implemented social distancing measures, there would likely be individuals across the country that followed this advice, even in states where it wasn't yet a legal necessity.

3. The report only focuses on the growth rate of infections plus deaths as an outcome. There are several other factors that are important to measure the 'impact' of social distancing measures, such as hospitalisations from COVID-19.

4. The dates when treatment states implemented social distancing measures are different, which scews the results on the effectiveness of social distancing. The impact of social distancing measures may take time to manifest, and the delay between implementation and seeing the effects could vary depending on factors such as population density and compliance with the measures. The researchers identify the treated group as states that implemented social distancing measures before March 26th, 2020. Some of these states implemented social distancing measures on March 24th, 2020 while others implemented them on March 19th . Even within states, measures were implemented at different times. The results will not accurately reflect any specific implementation date.

4.1. They could have addressed the discrepancy by utilizing a crucial feature in panel data analysis that enables units to receive treatment at different times and change their treatment status over time. Although they mention that individuals would be categorized as part of the control group if social distancing measures were implemented before March 26, and as part of the treated group if they were not, this assumption is problematic due to the rapid changes in people's awareness of the disease, even over a short period of time. If they had used panel data analysis, they would have been able to account for different treatment dates, albeit at the expense of limiting their ability to use GSCM. It is clear that different methods offer solutions to different problems, and the method they chose failed to address a significant issue. This would not have been an issue if it had been explicitly described and discussed, but unfortunately, they failed to do so.

5. As mentioned above, the compliance of individuals to the social distancing measures implemented varies widely across states. This means that the effect of the law itself might be very different from the effect of the actual treatment. Just like in randomized encouragement designs we have ITT and LATE, we can think of applying that framework in this case as well. The language that they used throughout the paper was not consistent. At some points, they referred to the effect as the effect of implementing social

distancing measures and in others, they referred to it as the effect of social distancing. This subtle difference makes abig difference because compliance rate is not the same in different states and is not something that we can easily control for using time or state fixed effects, so their language needs to be scrutinized. And the impact of this further exacerbated by the fact that compliance rate affects our results highly even more so than in other cases due to the fact that this is a transmittable disease and as such, the fact that other complied to the treatment could also affect individuals who didn't comply to the treatment and vice versa.

Replication

After some minor changes (i.e. changes to function and package names), we were able to run Osman & Sakib's code in our own environment. You can access the adjusted codebase here.

In Figure 1, you can see the table for the average treatment effect of Social Distancing we reproduced. To create it, the authors used R programming language with the following packages: readr, dplyr, tidyr, stringr, and doBy.

The code first extracts and formats the data from Johns Hopkins University Center for Systems Science and Engineering. Then, the code aggregates the cumulative cases and deaths per day per state and calculates each state's daily number of cases and deaths. As mentioned before in the Summary section, the authors used a treated group consisting of states that implemented social distancing measures prior to March 26th, 2020, and constructed a control group of the weighted combinations of states out of the remaining states using GSCM. The covariates used in this study were related to the demographics, population density, and climate of each state, as well as the number of confirmed cases and deaths in each state before the treatment date. They are optimized to create counterfactuals for the treated group that are as close to treatment in the pre-treatment period as possible.

The figure below visualizes the average treatment effect (ATE) or the difference in cumulative cases between the treated and control groups. The x-axis represents the time period from January 21st, 2020, to April 13th, 2020, and the y-axis represents the COVID-19 cases registered. The blue line represents the average infection rates for the control group, and the black line shows these rates for the treatment group.

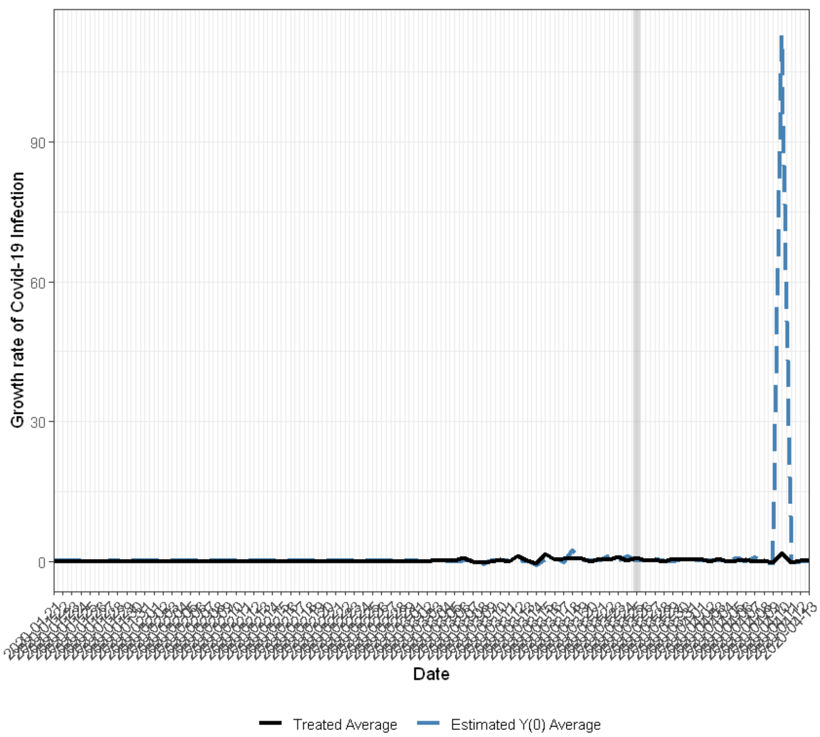

Figure 1. Average Treatment Effect of Social Distancing

Based on this plot, the infection rates for treatment and control groups come very close to each other in the pre-treatment era. This implies that the counterfactuals do a nice job of mimicking treatment units in our data. Therefore, we decently accounted for the selected confounders and can be more confident observing and interpreting the differences between treatment and control in the post-treatment period.

The peak in the blue line shows the skyrocketing number of COVID-19 cases registered compared to the black treatment group, which has just a slight increase of just around a couple of instances in the post-treatment era. Such a difference in peaks suggests that treatment effectively mitigates COVID-19 cases according to the data and its pipeline used in the paper.

Extension

We decided to extend the paper by conducting serveral placebo tests. We performed three tests. The first one is the placebo test by setting a angle state as the treatment unit and finding a synthetic control for that state. For the second one, we perfomred a placebo test in space where we treat each state as the treated unit and find the treatment effect for it. This will allow us to see if there is an actual treatment effect for treated states or if we would get the same result if we chose a unit that hasn't been treated as well. And finally, we also performed a placebo test in time. As mentioned above, the treatment date used in the analysis is March 26, 2020. So we wanted to see waht would happen if we change the treatment time to earlier dates.

SCM for Montana as an Individual

For the first extension, we use Montana as the only treated unit instead of combining different treated units into one. The rationale behind choosing Montana is that it was the earliest state to implement social distancing measures. We then create a synthetic control for it and estimate the treatment effect, The treatment date for Montana is March 16 instead of March 26, which is the treatment date we used in the earlier parts of the analysis.

Since Montana was early to adopt social distancing, we expect to see a much higher effect for it as compared to the treatment unit that is a combination of the different treated units. In our analysis of Montana as a single treated unit, we found the effect to be much higher than the original analysis.

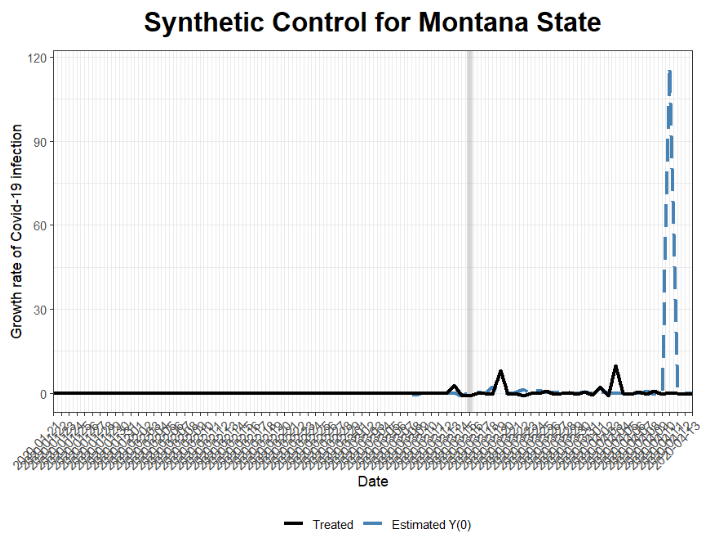

Figure 2. The trend for the outcome variable, growth rate of daily cases, for Montana and its Synthetic control unit over time

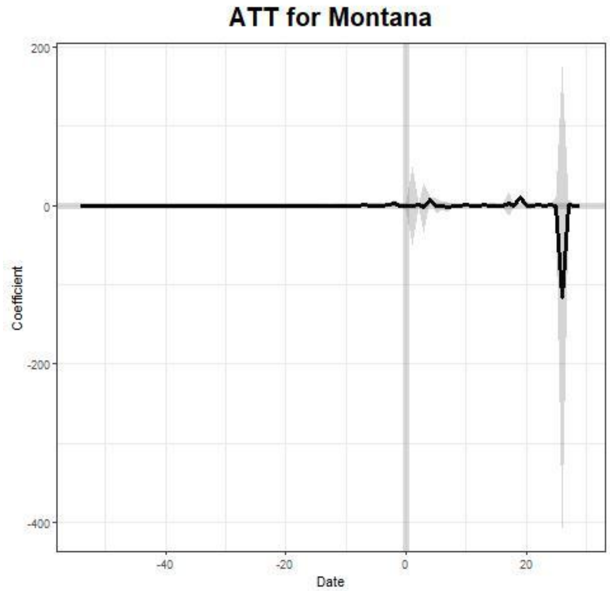

Figure 3. The average treatment effect for Montana. The treatment date is March 16, 2020.

Placebo in space

For the placebo test in space, our main treated unit is Montana and we are testing if the results we see for it are actually there. We do this by assuming that each of the states are treated in each iteration and the rest of the states will be in the control pool. For this case, when Montana is not being treated, we place it in the donor pool as well. Below, we see the results of the placebo test. The results are all over the place, and Motana doesn't stand out in any way.

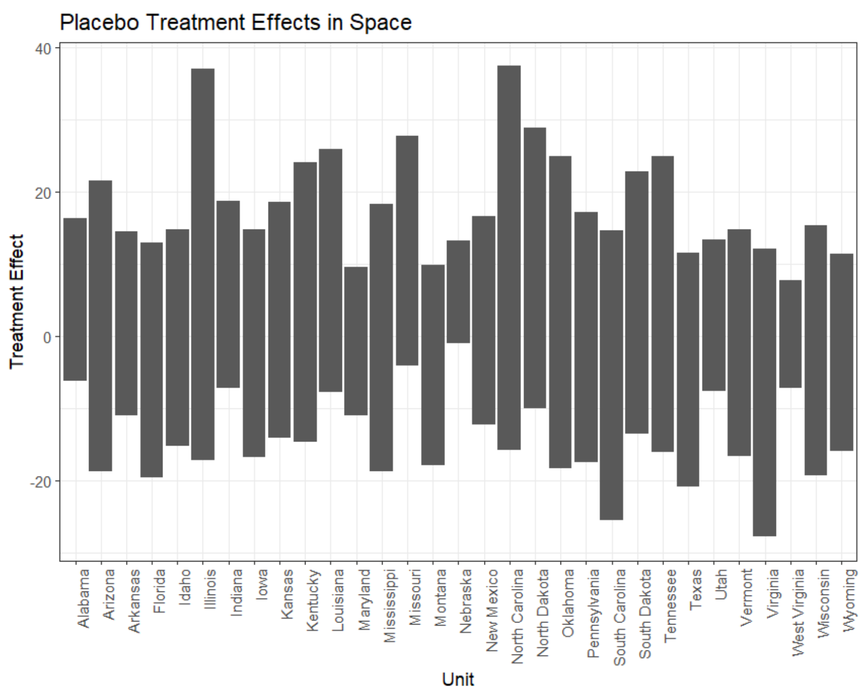

Figure 4. Placebo test in space where Montana is the actual treated unit. We see that it's treatment effect is not significant since it crosses zero, as do the results for all of the states

We can also get the distribution for the RMSPEs using the placebo test. The following figure plots the distribution of the average treatment effects using the placebo test. The red line is the effect for Montana.

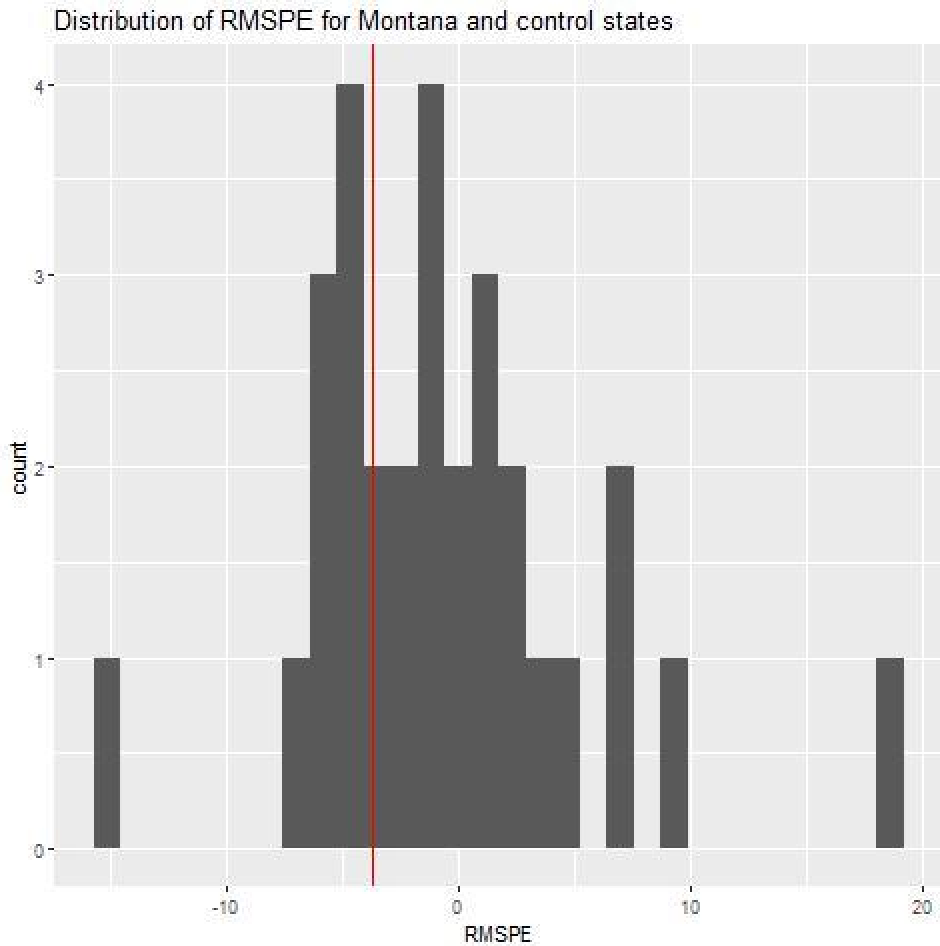

Figure 5. Distribution of RMSPEs.

The red line represents Montana and it's value is -3.693. We can see that's it's not an extreme case because of the treatment.

Placebo in time

For the placebo test in time, we are still considering Montana as our single treated unit. But to test if we can see the effect at different times when the policy was not actually implemented, we can change the treatment date from March 16, 2020 when it was actually implemented to an earlier date and calculate the effect. We can then compare the result from the actual treatment date to the placebo date(s).

Just like for the placebo test in space, we get results for the treatment effect for different dates. The figure below shows the treatment effect along with the standard error for the different dates.

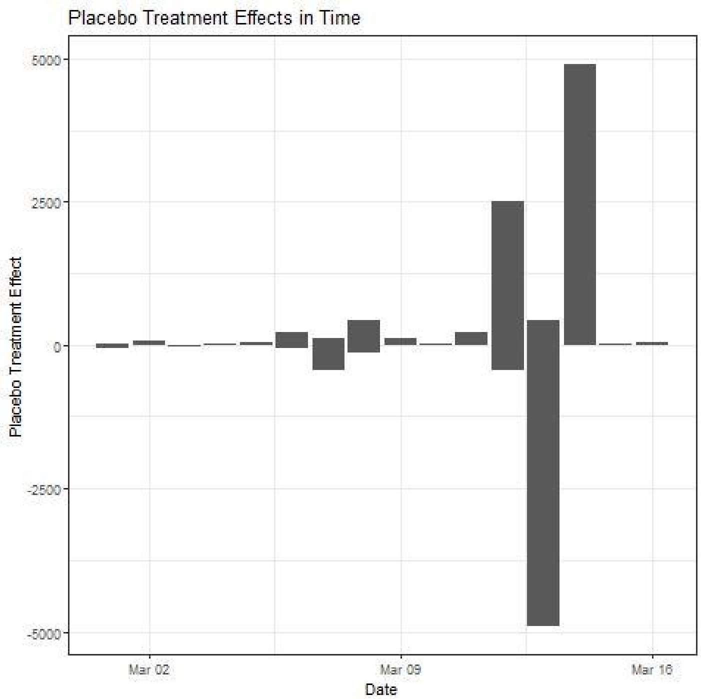

Figure 6. Placebo test in time where March 16 is the actual treated date. We see that it's treatment effect is not significant since it crosses zero, as do the results for almost all of the dates

We can also get the distribution for the RMSPEs using the placebo test. The following figure plots the distribution of the average treatment effects using the placebo test. The red line is the effect for for the actual treatment date – March 16.

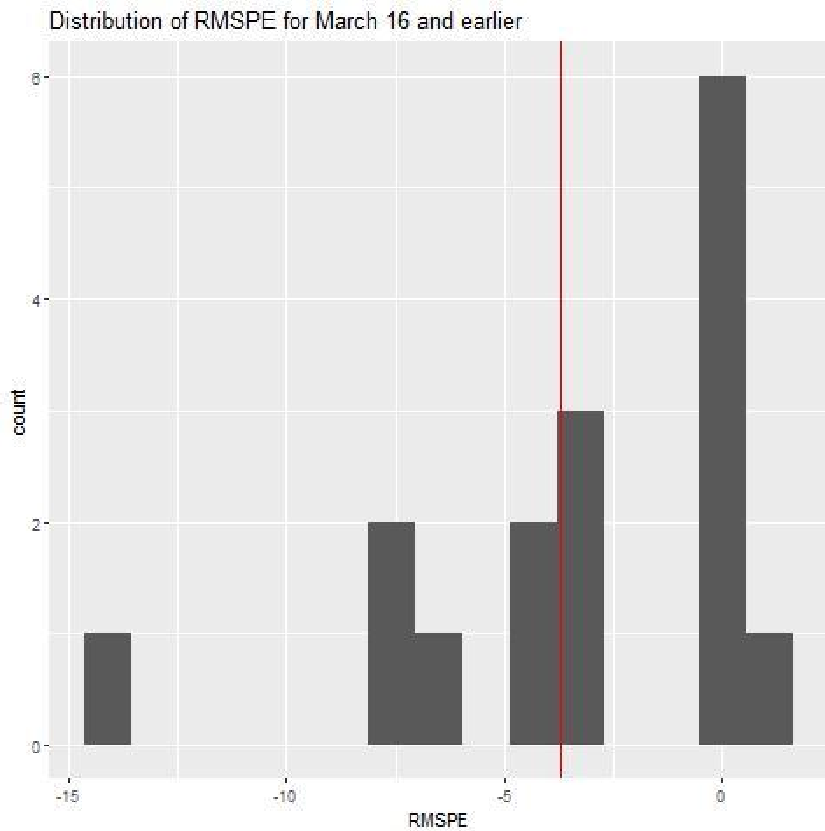

Figure 7. Distribution of RMSPEs.

The red line represents March 16 and it's value is -3.693. We can see that's it's not an extreme case because of the treatment.

Therefore, in both placebo tests, we see that the effect is not significant. Our original analysis in Montana also gave us the treatment effect of -3.69 with a confidence interval of [-13.13391 5.74737], which reduces our confidence in the first place. And conducting the placebo test in space and time gave us more confidence that the effect is not significant.

Discussion

The original analysis in the paper uses a GSCM to estimate the effect of adopting social distancing measures for treated states in the US. We summarized their methodlogy, critiqued some of their methods and assumptions (or lack there of), and replicated there findings. We extended the paper by focusing on a single treated unit (Montana) and performing placebo tests in space and in time. Both of these analysis showed that the results are not robust.

Studying the causal effect of adopting social distancing measures is a complicated topic because people react in uncertain ways and the causal effect not only changes from place to place but daily as well as people gain awareness of the disease and this would affect the growth rate of daily cases, but not necessarily as a result of the social distancing measures.

AI Tools Declaration

No AI tools were used in the creation of this assignment.

Contributions

All three of us worked on the interpretation of the paper and the replication code together. Catherine wrote up the summary and critique sections of the paper and did some of the extension code. Ivanna wrote up the replication section of the paper and made modifications to the summary and critique sections. Hussen wrote the remaining parts of the summary and critique sections and finished the extension code and wrote up the extension and discussion sections of the paper.

References

Gujarati, D. (2011). Econometrics by Example. http://zalamsyah.staff.unja.ac.id/wp-content/uploads/sites/286/2019/11/7-Econometrics-by-Example-Gujarati.pdf

Osman, S., & Sakib, N. (2020, a). socialdistance_data_cleaning.R - Harvard Dataverse. Harvard.edu. https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/OHLWEU/NNASEK&version=2.0

Osman, S., & Sakib, N. (2020, b). socialdistancing_model.R - Harvard Dataverse. Harvard.edu. https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.7910/DVN/OHLWEU/51ZCJA&version=2.0

Osman, & Sakib, N. (2020, c). Stay Home Save Lives: A Machine Learning Approach to Causal Inference to Evaluate Impact of Social Distancing in the US. Harvard Dataverse. https://doi.org/10.7910/DVN/OHLWEU

HC Appendix

#sampling: We pointed out the potential biases that can be introduced by the sampling methods used in the paper. For example, we highlighted how building a treatment group where each unit introduces treatment on different days might skew the results we get about the treatment effect. Also, we critique selecting control states similar to the treated states prior to the intervention as this way, the authors don't account for the unobserved confounding variables never discussed in the paper studied.

#biasmitigation: