EFFECT OF SOCIAL DISTANCING ON COVID-19 INFECTION RATE: A PANEL VAR ANALYSIS

Jingjing Fan, Lencer Ogutu, Elizabeth Combs, Cathylin Wang

correlation.one
DATA SCIENCE FOR ALL™

INTRODUCTION

The COVID-19 pandemic has upended life as we know it, introducing new concepts and ways of relating to each other.

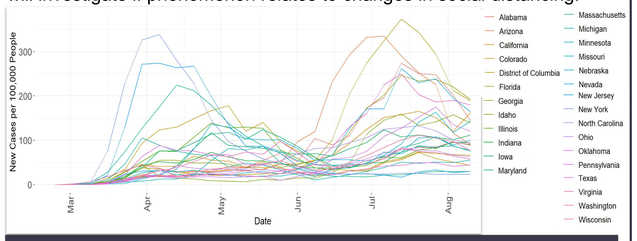
Key Questions:

- ❑ *Are social distancing metrics predictive of new US COVID-19 cases?*
- ❑ *Do building entry and workplace activity impact new case counts?*
- ❑ *How do varied responses across states impact the way social distancing relates to new cases?*

Our Data:

- **COVID-19 confirmed case data** tracks the new cases daily at the county-level, aggregated by state for this project.
- **Social distancing level** measures the percent change in traffic from the January baseline across categories.
- **Building entry percentage** consists of the percent change in building entry from early February.

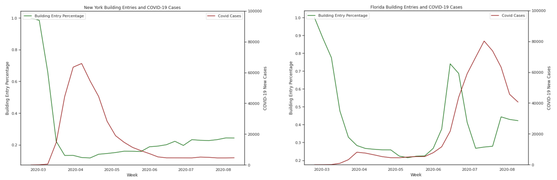
New COVID-19 cases per 100k by state shows two distinct waves. We will investigate if phenomenon relates to changes in social distancing.



EXPLORATORY ANALYSIS

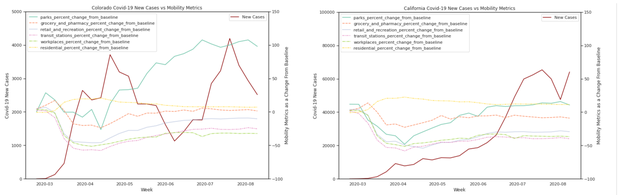
Building entry and new cases

- A spike in building entry rate is followed by a delayed spike in number of new cases (as seen evidently in Florida). Additionally, the response to rise in number of cases is a drop in building entry percentage potentially due to stricter enforcement of lockdown procedures.



Mobility metrics and new cases

- States exhibiting infection trends similar to New York and Colorado show similar trends in mobility evidenced by significantly high mobility in parks especially around June when states were reopening.

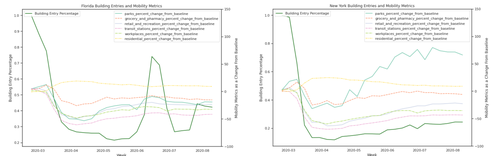


EXPLORATORY ANALYSIS CONT.

- Florida and California-like states did not exhibit much change in mobility variables except the dip in April when these states had strict lockdown orders.

Mobility metrics and building entry

- There was high correlation between building entry percentage and mobility to workplaces showing that many of the buildings entered were work buildings
- States that ended up recording a second wave had a high peak in percentage of building entry around June, like Florida below, when those states reopened. This, however, did not reflect on mobility data.



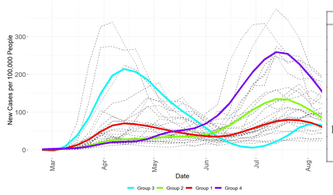
- From the cross analysis of the variables, data on buildings and mobility between March and mid- April was reactionary to the rising new COVID-19 cases. We therefore did not use this data in our analysis.
- Additionally, given the change point in the infections per state time series at around early June, marking a distinction in the two infection waves, we decided to use two models to analyze the data across the two periods.

METHODS

Panel Data: Multivariate time series data for multiple states.

Preprocessing:

- Group states by new-case trajectory over time using Bayesian Functional Clustering to account for policy differences between states.
- Lag social distancing variables to account for delayed diagnoses. Lags are selected for each group using CCF plots.



Group 1	Group 2	Group 3	Group 4
Colorado	California	District of Columbia	Alabama
Illinois	Idaho	New Jersey	Arizona
Maryland	Iowa	Massachusetts	Florida
Indiana	Missouri	New York	Georgia
Michigan	North Carolina		Nevada
Minnesota	Oklahoma		Texas
Nebraska	Wisconsin		
Ohio			
Pennsylvania			
Virginia			
Washington			

Dynamic Panel Model (DPM):

- Linear model with shared parameters and unobserved heterogeneity (also called random effects).
- Dependent variable: new cases
- Independent variables: all Mobility and social distancing measurements.
- Goal: recover the regression parameters and use them for prediction.

$$y_{i,t} = \boldsymbol{\beta}\mathbf{X}_{i,t} + \mu_i + \varepsilon_{i,t}$$

Fixed Effects set up:

$$\tilde{y}_{i,t} = y_{i,t} - \bar{y}_i = \boldsymbol{\beta}(\mathbf{X}_{i,t} - \bar{\mathbf{X}}_i) + (\varepsilon_{i,t} - \bar{\varepsilon}_i)$$

$$\bar{\mathbf{X}}_i = \frac{1}{T}\sum_t \mathbf{X}_{i,t}$$

$$\bar{\varepsilon}_i = \frac{1}{T}\sum_t \varepsilon_{i,t},$$

METHODS CONT.

Panel Vector Autoregression (PVAR):

- Adding a vector autoregressive component to a DPM. Ex:

$$\mathbf{y}_t = \boldsymbol{\nu} + \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + \cdots + \mathbf{A}_L\mathbf{y}_{t-L} + \varepsilon_t.$$

- Endogenous variables (y): New cases, building entry percentage, workplace utility
- Exogenous variables (X): all other mobility metrics
- Goal: find a granger causal relationship between the endogenous variables. Isolate the effect that going to work or going into buildings has on new cases.

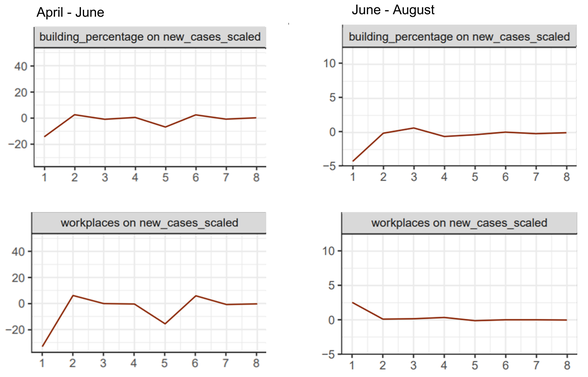
$$\mathbf{y}_{i,t} = \mu_i + \sum_{\ell}^{L} \mathbf{A}_\ell \mathbf{y}_{i,t-\ell} + \mathbf{B}\mathbf{X}_{i,t} + \varepsilon_{i,t}$$

RESULTS

Impulse Response Graphs:

Group 3: DC, NJ, MA, NY:

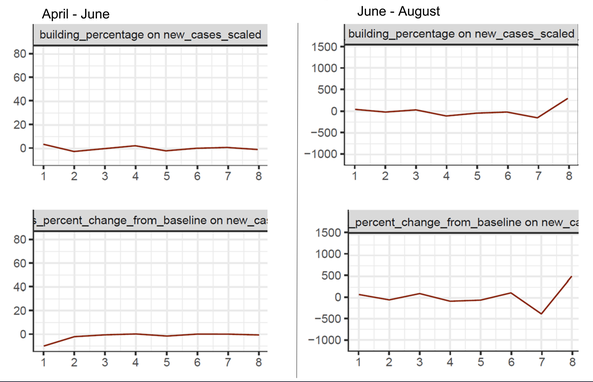
- Unit shock to building and workplace mobility → increased COVID-19 cases 2 weeks later in both time periods.
- Case count stabilizes faster in June - August than April - June.



Group 4: AL, AZ, FL, GA, NV, TX:

- April - June: Unit shock to building and workplace mobility → increased COVID-19 cases 2-4 weeks later.
- June - August: Unit shock to building and workplace mobility → initial increase in COVID-19 followed by oscillation:

↑ Mobility → ↑ Transmission → ↓ Mobility → ↓ Transmission → ↑ Mobility

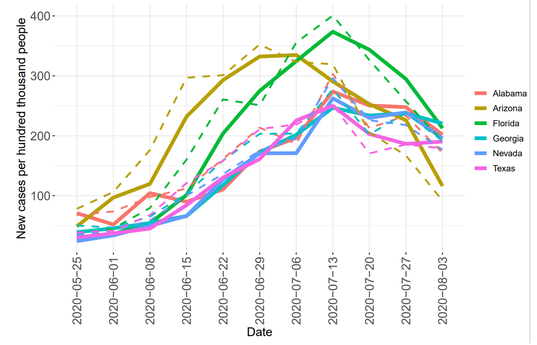


RESULTS CONT.

- Mobility data from all place categories can be used to predict COVID-19 cases for the next week.
- The model gets more accurate over time because there are more data points

Group 3: DC, NJ, MA, NY:

Group 4: AL, AZ, FL, GA, NV, TX:



CONCLUSIONS

Key Takeaways:

- ✓ Mobility and social distancing has both predictive power and are highly influential for disease transmission.
- ✓ Different groups of states are doing better/worse depending on their policies.

Next Steps:

- Analyze policy differences between states.
- Examining interactions between the different types of mobility.

ACKNOWLEDGEMENTS

Special thanks to Sergio Mastrogiovanni and Savannah Thais for mentorship throughout this project and Correlation One for hosting the Data Science for All Women's Summit.

REFERENCES

https://penntoday.upenn.edu/news/coronavirus-or-covid-glossary-help-navigate-pandemic-vocabulary
https://newsforkids.net/fastfacts/coronavirus-words/
https://www.cnn.com/2020/03/16/us/nyc-coronavirus/index.html
https://www.google.com/covid19/mobility/
https://www.openpath.com/social-distancing-index
https://raw.githubusercontent.com/CSSEGISandData/COVID-19/