W203 Lab 02

Association of Changes in Mobility and Policy with COVID-19 Case Counts

Jun Qian, Lucas Schroyer, Ryan Mitchell, Oliver Chang

Research Question & Feature Selection

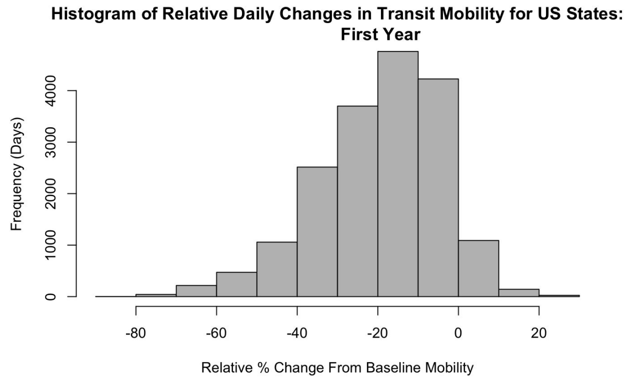
What features from our mobility and policy databases, if any, are highly associated with changes in COVID-19 state-level case counts?

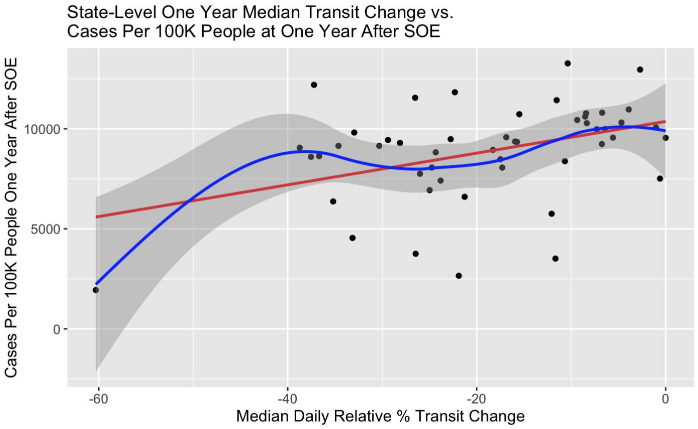
Features of Interest	Model Applicability	Source
Transit Mobility (% Change)	Models 1, 2, 3	Google's COVID-19 Community Mobility Report
Population Density (People/sq. mi.)	Models 2, 3	US Census Data
Population Under the Age of 25 (%)	Models 2, 3	US Census Data
Days of Mask Mandate	Models 3	US Policy Database
Days of Unemployment Benefits Increase	Models 3	US Policy Database
Amt of Unemployment Benefit Increase	Models 3	US Policy Database
Days of Close Business Mandate	Models 3	US Policy Database
Days of Travel Quarantine Mandate	Models 3	US Policy Database
Days of Stay At Home Mandate	Models 3	US Policy Database

Analysis Steps

1. **Formulate** research question and identify data sources
2. **Perform** data cleaning and preliminary analysis for the following datasets:
    a. **US Census Data –** Aggregate county-level population and area data to the state level
    b. **Google's COVID-19 Community Mobility Report –** Using census population weights, create a weighted state level dataframe of mobility percent changes by sector
    c. **COVID-19 US State Policy Database–** Read in seven unique policy workbooks and join into one policy dataframe by state
    d. **NYT COVID-19 Database –** Read in and join with other datasets by state and date to create the master dataframe
3. **Conduct** Exploratory Data Analysis (EDA) and Create Three Descriptive Models using OLS
4. **Compare** statistical significance and key associations from the three models
5. **Assess** model limitations and study conclusions
6. **Complete** presentation and report

Model 1: cases_per_100k= β0 + β1 median_transit_change + u

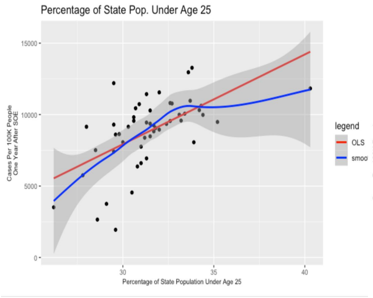


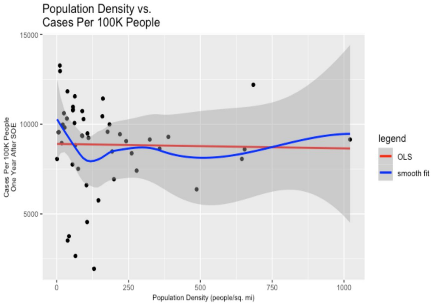


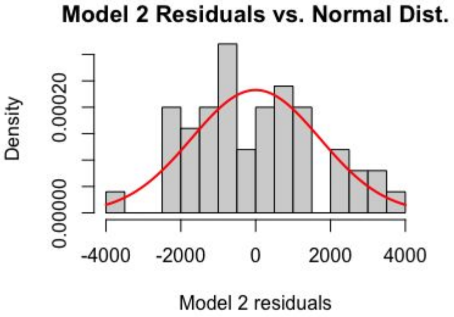
Key Points:

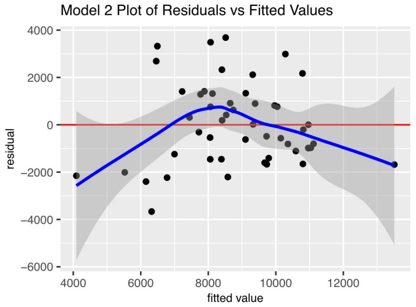
- Adjusted R-squared = 0.151 (df = 48)
- β estimation: [10370**, 79**]

Model 2: cases_per_100k= β_0 + β_1 median_transit_change + β_2 pop_age_0_24 + β_3 population_density + u









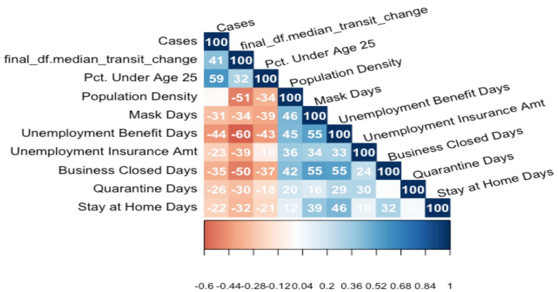
Key Points:

- F statistics=16.012***, Adjusted R-squared: 0.48 (df=46),
- β estimation: [-10160.59**, 82.22***, 628.59***,4.67**]
- Robust sd: [5070.54, 31.48 , 154.76, 2.21], total sd 1769.78

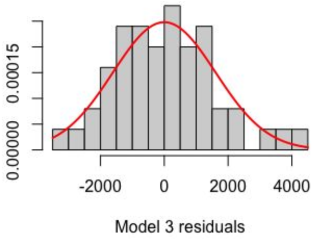
Model 3:

cases_per_100k= β0 + β1 median_transit_change + β2 pop_age_0_24 + β3 population_density + β4 mask_mandate_days + β5 unemployment_benefits_days + β6 increased_weekly_unemployment_ insurance_amt_thru_jul31 + β7 business_closed_days + β8 travel_quarantine_ mandate_days + β9 stay_at_home_days + u

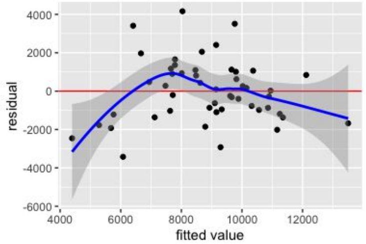
Correlation Scatter Matrix



Model 3 residual vs. Normal Dist.



Model 3 residual vs. Model fit



Key Points:

- F statistics=5.888***, Adjusted R-squared: 0.47 (df=40)
- β estimates for transit change turns insignificant, additional covariates all insignificant
- β estimates for transit change, age group, population density [54.35, 569.30***, 5.79**]
- Total robust sd 1780.51
- vif values all <= 2.3. No substantial multi-colinearity.

Conclusion

The following conclusions are based on our descriptive model analysis of the case counts, mobility, and policy data

Model 1

Model 1 shows the 1% median transit changes is associated with approx 80 case counts but model fit can be improved.

Model 2 Parameters

The parameter estimate of Model 2' s two additional covariates (pct of pop of age under 25 and population density) are significant

Model 2 Robustness

Model 2 sees better F stats and Adj r-squared. Residuals of the model has smaller (robust) sd and closer to normal distribution.

Model 3

Model 3's all six additional policy covariates are insignificant. The transit change also turned insignificant

Model limitations

IID violation, Heteroskedasiticity, Omitted variable such as temperature,mask compliance, test availability etc, family wise error

Based on these findings, we believe that Model 2 is our most parsimonious model with the transit change, percent of population under 25 years old and density as our independent variables.

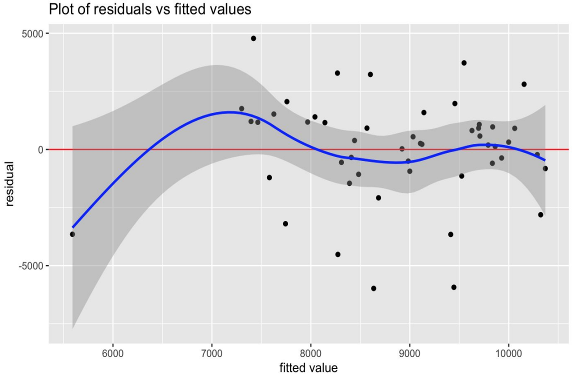
Works Cited

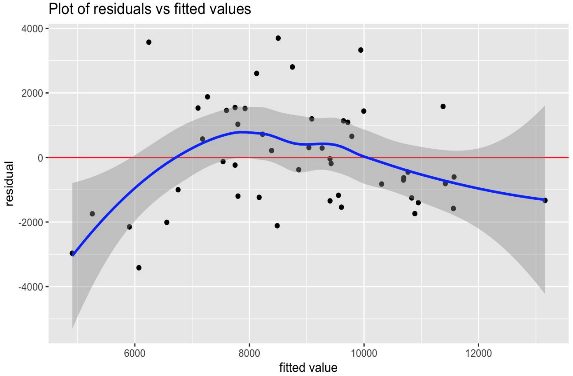
- Raifman J, Nocka K, Jones D, Bor J, Lipson S, Jay J, and Chan P. (2020). "COVID-19 US state policy database." www.tinyurl.com/statepolicies. A database of state policy responses to the pandemic, compiled by researchers at the Boston University School of Public Health.
- COVID-19 Community Mobility Report, https://www.google.com/covid19/mobility. A Google dataset that includes state-level measurements of individual mobility
- The American Community Survey, https://data.census.gov/cedsci/table?q=ACS&g=0100000US.04000.001&tid=ACSDP1Y2019.DP05&moe=false&hidePreview=true)
- New York Times, https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-states.cs. For COVID-19 case counts

Appendix

CLM for Model 2 Assumptions 1 & 2

- IID – Fail to reject the observations are not IID due to clustering of policy and case counts between states of close proximity and temporal clustering of case case counts and mobility data
- Linear Conditional Expectation - Fail to reject that they are LCE





CLM for Model 2 Assumptions 3, 4, 5

- No Perfect Colinearity: No Perfect Colinearity or Near Perfect Colinearity. VIF all <2

```
median_transit_change       pop_pct_age_0_24     population_density
             1.393597               1.168902               1.419158
```

- Homoskedasticity : can reject the null hypothesis that variance of the error term is homoskedastic

```
##
##  studentized Breusch-Pagan test
##
## data:  lm.base
## BP = 8.3386, df = 3, p-value = 0.03951
```

- Normality : can not reject the null hypothesis that residual is normal

```
##
##  Shapiro-Wilk normality test
##
## data:  model_2_final$residuals
## W = 0.97275, p-value = 0.2984
```

F-Test Between Model Selection

F-test for model 1 and 2:

```
## Analysis of Variance Table
##
## Model 1: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24
## Model 2: cases_per_100k_at_365d ~ median_transit_change
##   Res.Df       RSS Df Sum of Sq      F    Pr(>F)
## 1     47 176654365
## 2     48 245004785 -1 -68350419 18.185 9.595e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

F-test for model 2 and 3:

```
## Analysis of Variance Table
##
## Model 1: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +
##     population_density + mask_mandate_days + unemployment_benefits_days +
##     increased_weekly_unemployment_insurance_amt_thru_jul31 +
##     business_closed_days_round1 + travel_quarantine_mandate_days +
##     stay_at_home_days
## Model 2: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +
##     population_density
##   Res.Df       RSS Df Sum of Sq      F Pr(>F)
## 1     40 126808331
## 2     46 144077125 -6 -17268794 0.9079 0.4991
```

Models 1-3 OLS Regression Table

Table 1: OLS models for COVID-19 Spread

	Dependent variable:		
	cases_per_100k_at_365d		
	(1)	(2)	(3)
median_transit_change	79.250** (32.943)	82.223*** (31.481)	54.348 (45.057)
pop_pct_age_0_24		628.592*** (154.756)	569.302*** (160.445)
population_density		4.665** (2.213)	5.790** (2.789)
mask_mandate_days			-1.188 (3.116)
unemployment_benefits_days			-3.873 (5.252)
increased_weekly_unemployment_insurance_amt_thru_jul31			-2.339 (2.110)
business_closed_days_round1			-16.598 (23.841)
travel_quarantine_mandate_days			-1.779 (3.121)
stay_at_home_days			1.841 (5.388)
Constant	10,370.170*** (610.184)	-10,160.590** (5,070.544)	-4,556.676 (5,891.585)
Observations	50	50	50
R2	0.168	0.511	0.570
Adjusted R2	0.151	0.479	0.473
Residual Std. Error	2,259.262 (df = 48)	1,769.777 (df = 46)	1,780.508 (df = 40)
F Statistic	9.714*** (df = 1; 48)	16.018*** (df = 3; 46)	5.880*** (df = 9; 40)
Note:			*p<0.1; **p<0.05; ***p<0.01

OMV and Other Model Biases Discussion

Omitted Feature	Correlation to Target Feature	Correlation to Selected Feature	Bias Direction
Temperature	Positive	Positive (Mobility)	Positive
Mask Compliance	Negative	Positive (Population Density)	Negative
COVID Testing Availability	Positive	Positive (Population Density)	Positive
Percentage People with Mass Transit	Positive	Positive (Population Density)	Positive
Percentage People WFH	Negative	Negative (Mobility)	Positive