Covid19 Containment and Government Policies

W203: Statistics for Data Science - Section 7

Leyla Greengard, Amey Mahajan, Rakesh Walisheter

1. Introduction

In early 2020 the first cases of Covid-19 appeared in New York and soon thereafter spread through the North East region of the United States. As the disease ravaged the region, the states instituted sweeping restrictions on their populations, including at first stay-at-home orders, and closing of non-essential businesses in early March and in April instituting mask mandates. At this time other states were relatively safe. But, as the year progressed, the disease propagated through the country. As the number of cases rose some states were quick to enforce measures while others were reticent to impose restraints, be it for political reasons or in respect of individual liberties.

We intend to examine whether government actions and policies helped mitigate the spread of Covid19 in terms of the *current* number of cases reported from tests as of the date of the data, 10/31/2020.

Government policies which help during a pandemic could be either short-term in response to the rapid developing pandemic - such as imposing Stay-At-Home mandates or Mask-wearing mandates; or indeed long-term policies - such as Medicaid allocations which provide medical coverage for economically disadvantaged population who are also at high-risk of being exposed to and contracting the virus.

This study intends to highlight the influence of Medicaid-allocations and Pandemic-measures (Stayhome and Mask Mandate orders) on controlling the exponential-growth phase of the pandemic.

1.1. Intended Audience

This study is intended for government and non-government agencies which legislate and implement policies for public welfare and health & safety. The hope of this study is to act as a barometer to gauge the efficacy of pandemic-response policies and long-term healthcare policies and help stimulating a pandemic-response rulebook for the future.

Learnings in this report are also intended to help highlight the efficacy of these measure to the general public, a significant portion of which unfortunately still question the efficacy of pandemic-response measures and are oblivious to how much of an impact Medicaid has in healthcare crises. This is important because not only does the population play an important role in actively controlling the pandemic by following social-distancing and mask-guidelines, but also have a strong say in the state's medicaid allocations through their vote-powers.

1.2. Research Question

Based on the data available, this study focuses on the following policy question,

How has the number of new cases in late-stage of pandemic been controlled as a result of State-Medicaid allowance and Early adoption of pandemic-control measures?

State Medicaid Allocation

Essential frontline workers are in many ways critical to maintain a functioning society during this pandemic. They regularly put their lives at risk because they unfortunately need to do this, to have a steady source of income.

An article from California Health Care Foundation Ref-1 highlights the importance of Medicaid for essential frontline workers. But medicaid is not limited to essential workers alone and impacts many other groups of people, who are also under a high risk of contracting COVID-19; a quote from the article:

"Research has shown time and time again the varied benefits of Medicaid expansion: *lower mortality rates among older adults* with low incomes, declines in infant mortality, reductions in racial disparities in the care of cancer patients, and fewer personal bankruptcies, just to name a few."

Considering the major impact medicaid has on so many high-risk demographies, this study looks at medicaid as a long-term policy implemented in states and its impact on the number of cases.

Covid specific short term policies -

As the pandemic entered the country, the north-eastern states were the first hotspots. In response to the situation seen in these hotsots, the CDC recommended guidelines to control the pandemic and other states began to implement these policies. Primarily required policies were:

- Mask Mandates

- Stay-Home / Shelter in place
- Business closures

The effectiveness of these measures will be modeled in this study.

1.3. Operationalization

1.3.1. Dependent variable

To operationalize the exponential-growth phase of the pandemic, this report uses the column - *"Case Rate per 100000 in Last 7 Days"* in the Covid-19 dataset as the dependent variable.

The choice of analyzing cases in the last-7 days is in keeping with suggestions from research Ref-10 Ref-11, that *Super-spreader events* were majorly responsible for expediting the exponential-growth phase of the pandemic. *Occurrance of Super-spreader events* is therefore a confounding-variable which might not have a strong correlation with cases in the early stage of the pandemic in a state but is expected to be highly correlated to the exponential-growth phase.

But in the absence of data for *Occurrance of Super-spreader events* in states and the fact that the pandemic-control measures being studied in this report influence how widespread and severe a super-spreader event can be, this report focuses on cases during the last-7-days in an attempt to explain some of the variance held in *Occurrance of Super-spreader events* as well in the same model.

1.3.2 Independent variables

1.3.2.1 State's Medicaid expediture normalized to state population living under poverty.

To argue about the effectiveness of medicaid to control spread of a pandemic, we work with actual dollar-amount medicaid-expenditure which is available from the Kaiser Family Foundation data set.

Since Medicaid-expenditure is primarily allocated for the state's reported population under the poverty line, the medicaid data collected is then normalized to a per-capita-medicaid-allocation for the population living under poverty.

VARIABLE : `medicaid_per_capita_under_poverty = Total_Medicaid / (Population * Percent under poverty line)`

1.3.2.2 Timeline of states' Covid control policy implementation

The timeline variables that we have considered are all dates. To operationalize how these dates manifest as influencing the number of cases, this report considers how early a State implemented the corresponding policy. For instance, Alabama implemented a stay-home order on 04 Apr' 2020 which was 210 days from the date the data was gathered. Wyoming which never implemented a stay-home order is assigned a 0-value to represent that it never went to shut-down.

This report operationalizes these dates as the duration a policy was active by transforming each date into the number of days between that date and the date this dataset was gathered (10/31/2020). The smaller the number, the more recent the event of implementing the policy and a higher expected rate of exponential-growth.

1. Mask Mandate Implementation:

The variables we used were "Mandate face mask use by all individuals in public spaces." and "State ended statewide mask use by individuals in public spaces"

Issues:
+ Some states never issued a mask mandate. For those we set the number of days since the mandate equal to zero.
+ For the few states that ended the statewide mask use we also set the number of days since the mandate equal to zero.

2. Stay-At-Home order implementation

For the stay-at-home mandate we used the two variables "Stay at home/ shelter in place" and "End stay at home/shelter in place."
Issues:
+ In some cases the dates were set to zero. We assumed that the ones for which both "Stay at home/shelter in place" and "End stay at home/shelter in place" were 0 never had a stay-at-home mandate. For those we set the number of days since the beginning and the end of mandate equal to zero.
+ In cases where one of the variables was 0 and not the other, we used outside sources to verify the dates. As far as possible we tried to go directly to government sources. However that was not always possible, so we relied on semi-official sources. For 6-states, end of stay at home dates were not clear because the opening happened in stages. For example, in New Mexico not all counties opened at the same time. This report uses dates when orders-were issued by governments for first-time opening as the end-of-stay-at-home date. The 6-states for which data was fixed are Connecticut Ref[2], California Ref[6], New Mexico Ref[3], Oklahoma Ref[4]and Texas Ref[5].

The dataset contains other variables that point to state policies, such as the increase in unemployment payments, or the total number of tests perfomed, but we do not believe that they have a direct effect on the short term number of cases.

3.Business-Closure implementation

For the business closure implementation we used the variables "Closed other non-essential businesses" and "Began to reopen businesses statewide"
Issues:
+ The name of the first variable indicates that there were other business closures at prior dates, which were not available in the database.
+ No specific data issues.

1.3.3 Caveats About Validity of Data

The data we are using represents the time at which the regulations were issued by the state governments. As we know, mandates were not universally followed across each state's population. The mask mandates were not enforced in some states, and even when they were, many people did not follow them.

We are using the start and end dates of stay-at-home mandates as though they were uniformly followed from one day to the next. In reality, this was not the case as for example in most states the openings happened gradually for various types of business; as mentioned before, in some states like New Mexico, not all counties were opened at the same time.

The data is at the state level. This lack of granularity could hurt the analysis, especially in states that are very heterogeneous. For example, New York City has half the population of the state of New York and less than 1% of its area. Since the city has such a higher density, a stay-at-home is likely to have a very different effect on the spread of the virus than it would have on sparsely populated area. These differences do not show up in our analysis.

2. Model Building Process

2.1. Exploratory Data Analysis

2.1.1. Initial EDA

Medicaid Expenditure

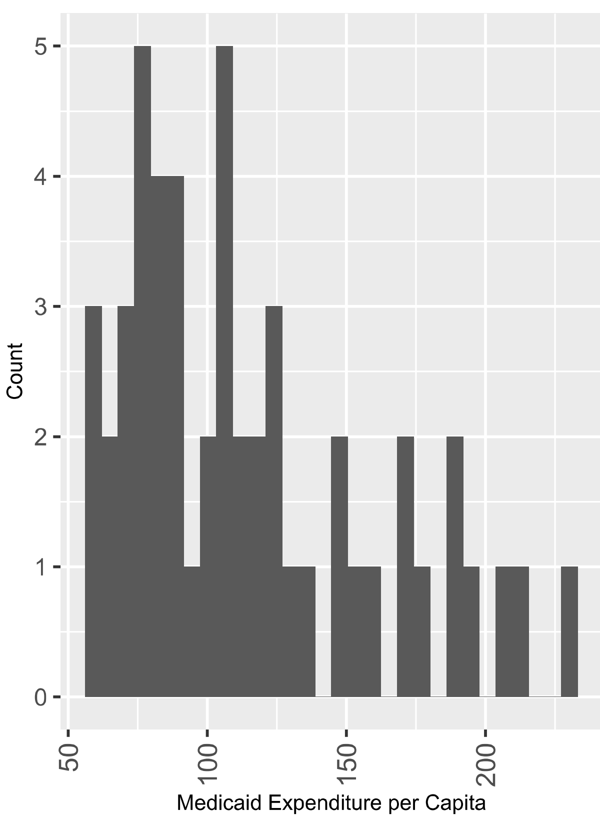
Medicaid expenditure at state level is specifically used to help people whose income is insufficient to opt for a health care plan. With limited resources, this population is more likely to have pre-existing conditions and to be essential workers and thus, more susceptible to Covid. Analyzing government efforts in tackling this challenge for the economically challenged population is hence necessary.

The given data is a percentage of total expenditure for a given state. But we want to focus the data on obtaining the medicaid expenditure per capita. We therefore use a new dataset that comprises the total medicaid expenditures for a State given as a dollar amount Ref-12. This dataset gives us the actual amounts spent by a state on medicaid. Furthermore, we are calculating the total number of people living under the poverty line from the given dataset. To finally get the per capita value for medicaid expenditure for people under the poverty line by dividing the total medicaid expenditure by the number of people below the poverty line. This transformed variable gives us insight into an overall policy-making for a state regarding medicaid for underprivileged communities. We will use this "medicaid__per__capita__under__poverty" to analyze the total number of cases.

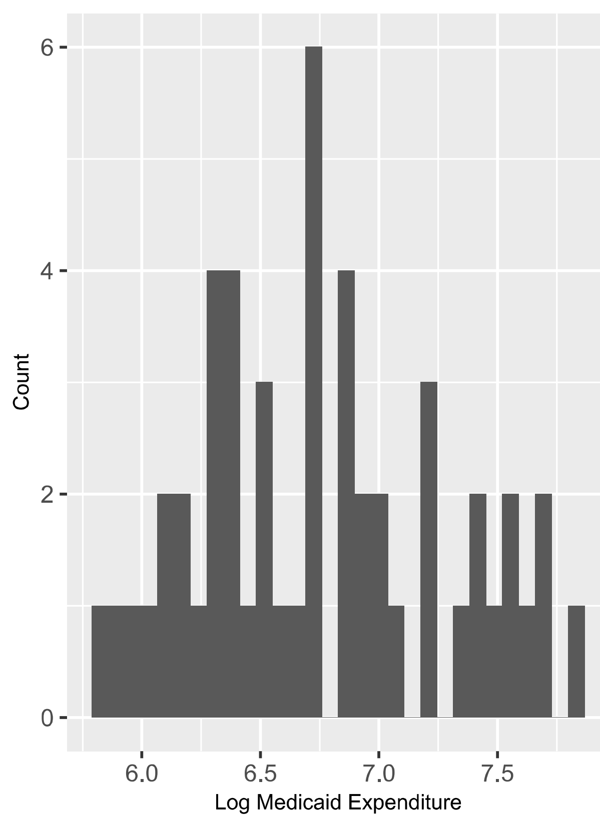
Fig-2.1.1.1 below shows that most of the data is on the left hand side of the graph. To reduce the skewness of the data we use the logarithm of the variable (see right hand side graph). We will be using log2(medicaid__per__capita__under__poverty) as our explainer variable in the model which adds resolution for clustered data s shown in Fig-2.1.1.2.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Fig−2.1.1.1 Medicaid Expenditure Per Capita (Poor Populati



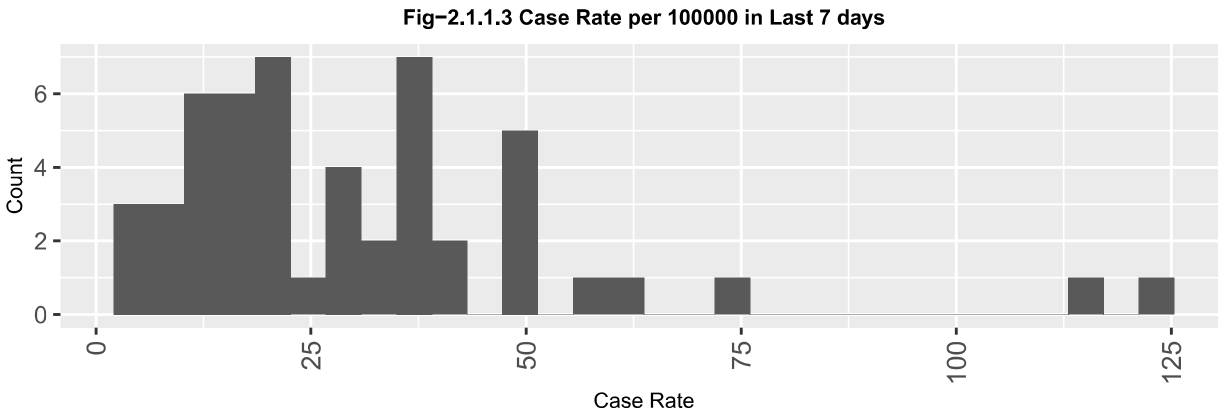
Fig−2.1.1.2 Log Medicaid Expenditure Per Capita (Poor Popula

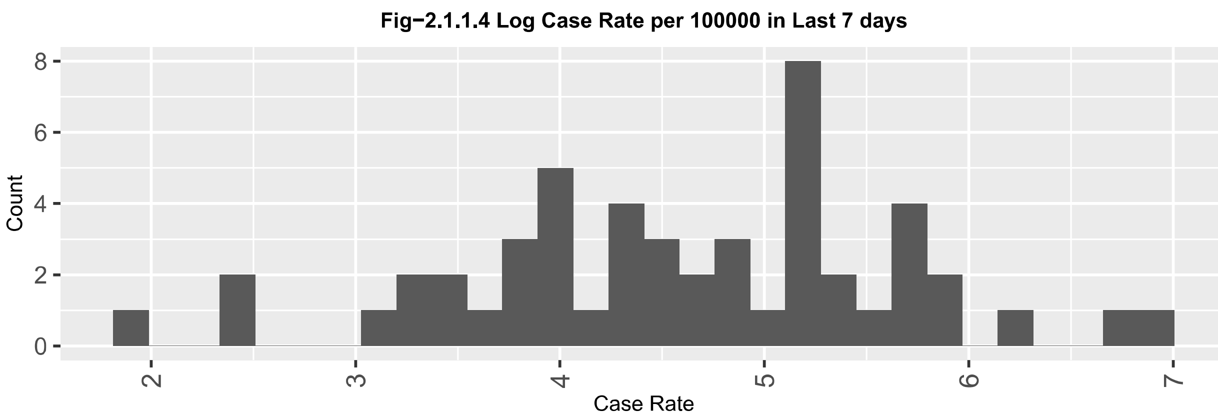


Case Rate Last-7 Days

Fig-2.1.1.3 show the histogram for distribution of 7-day-cases variable. The distribution of this variable is skewed, but a log transformation makes the overall distribution less so as seen in Fig-2.1.1.4. Hence, models in this report will use log2(normalized_cases_last7) as the output variable.

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```





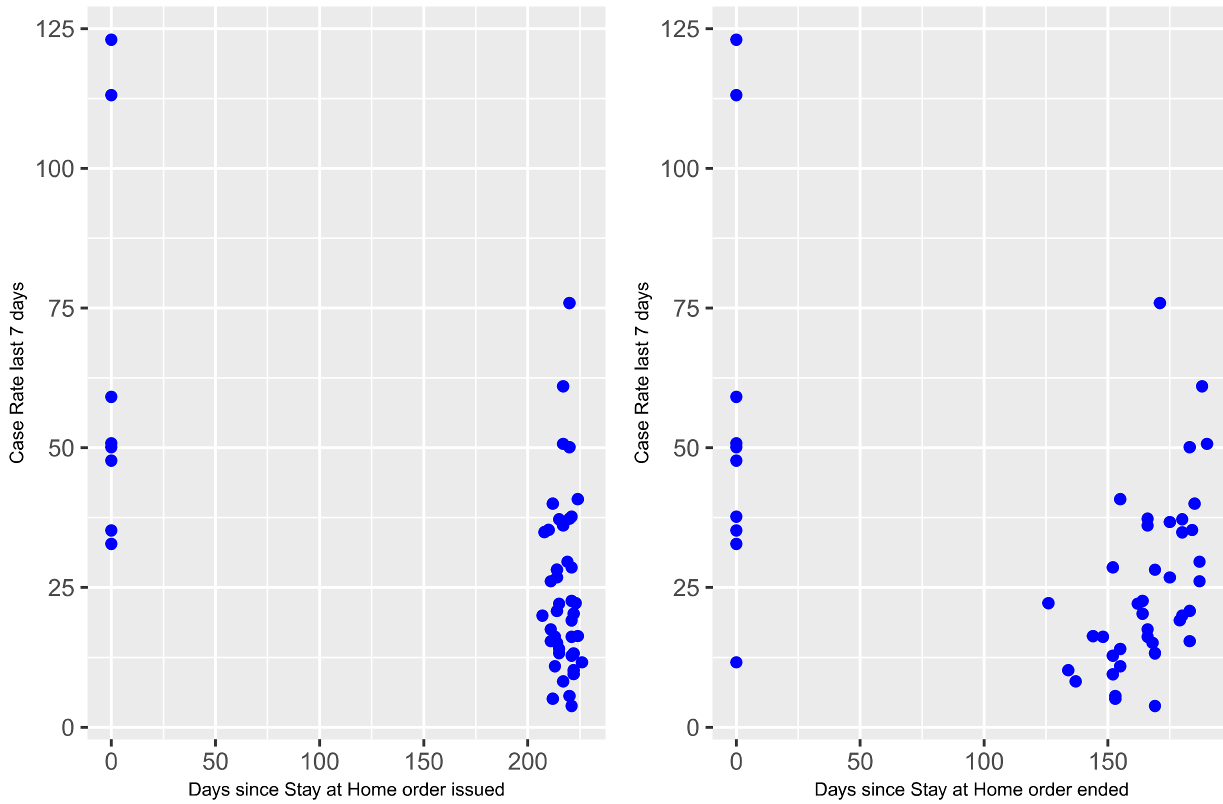
Policy Implementation Dates:

The COVID-19 pandemic forced the authorities to take major steps to tackle the spread of the infection. Given the nature of how this disease spreads, and recommendations from the CDC and the WHO, governments had three major goals: social distancing, mask wearing, and hand washing. The Covid-19 dataset addresses the first two goals using data for - Stay at Home order, Business closures, and Mask mandate. The following plots show how the case rate for the last 7 days behaves based on the number of days since these policies were implemented.

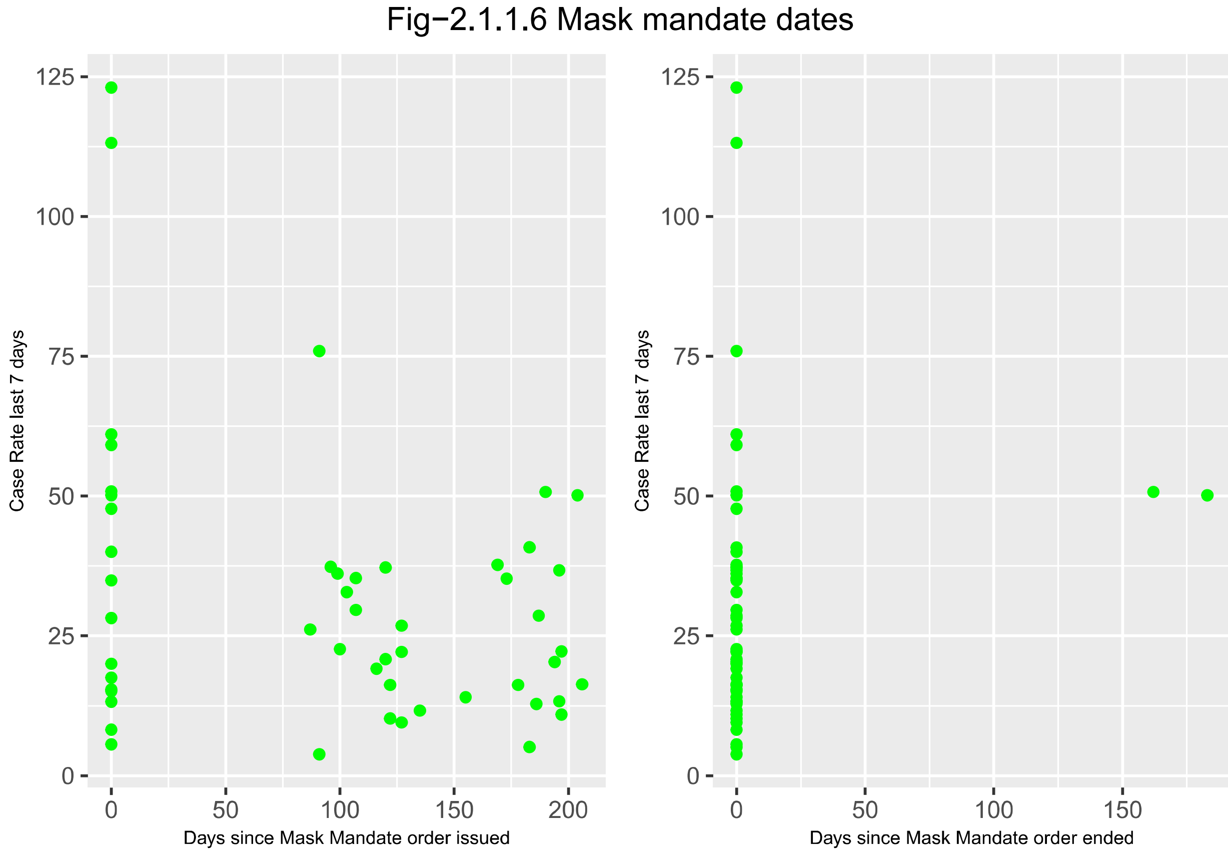
In these plots, the data-points showing 0-days since implementation correspond to states that did not institute the corresponding measure.

Stay-Home Policy Dates

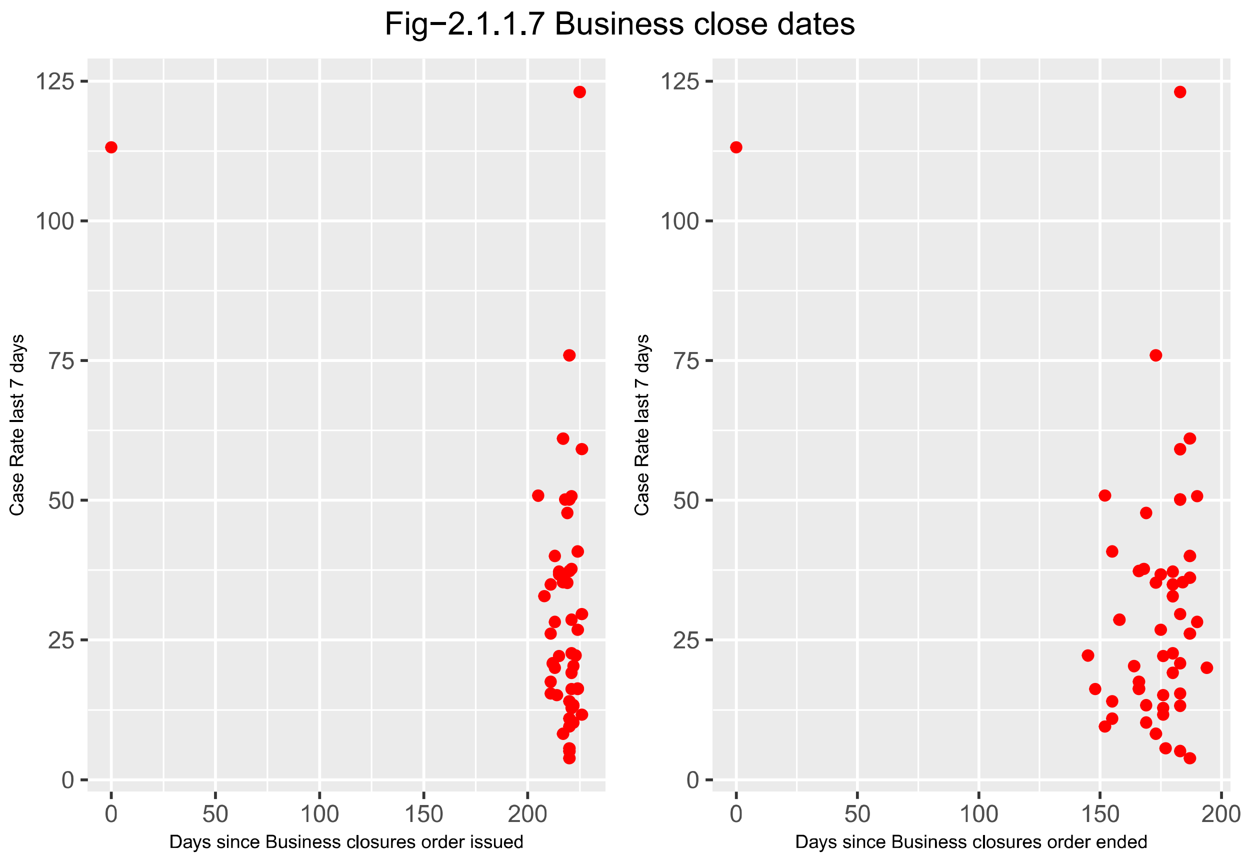
Fig−2.1.1.5 Stay at home dates



Mask-Mandate Policy Dates



The right hand side graph indicates that mask mandates are still in effect for all states that instituted them, except for 2 of them. This makes the end-mask_mandate variable unsuitable for our modelling.
We notice again that our outliers (the Dakotas) did not institute a mask mandate.



These graphs show again our two ouliers. One of them did not institute a business closure (South Dakota). The other closed businesses for a short period but still has a large number of cases (North Dakota).

Other variables

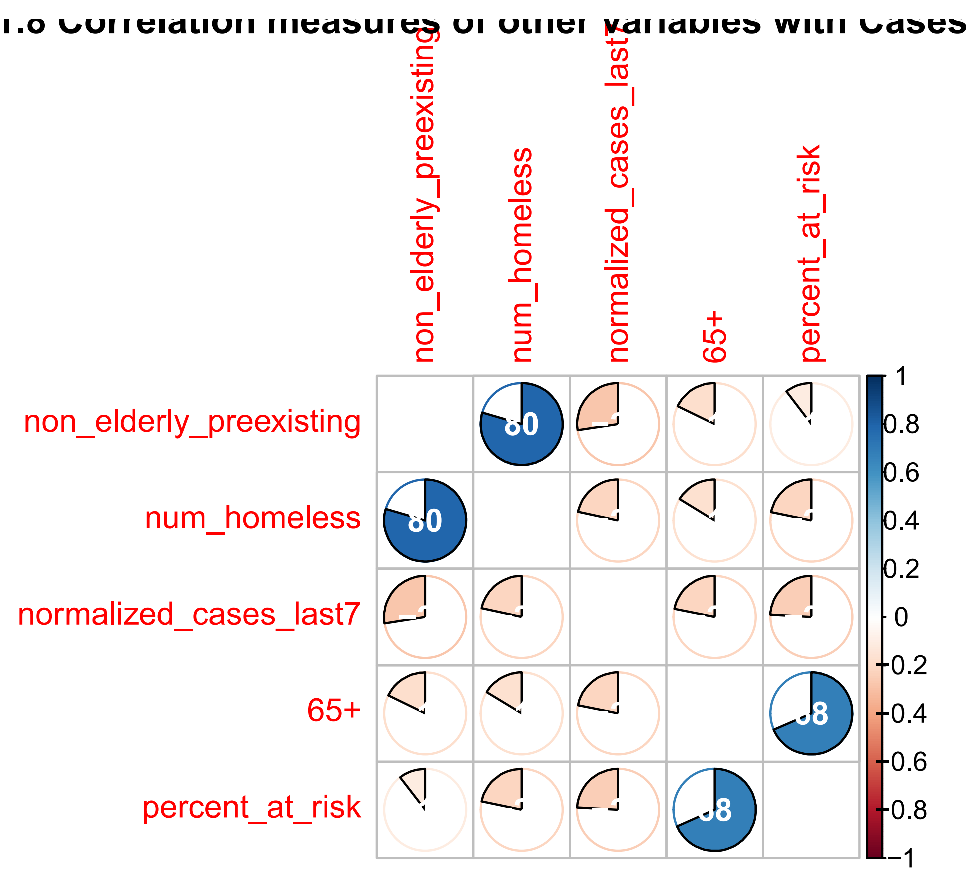
Apart from the above mentioned variables, this report explores other variables that represent state-policies which could influence spread of the the pandemic. Most of these variables are around demographics of states and cannot be directly connected with state policy but could show interesting correlations.

Number Homeless (2019): The homeless population are a section of the population at high-risk of contracting the virus. They have minimal or no access to healthcare facilities and have no way of avoiding exposure, assessing if they have been exposed to the virus. We believe that this number may influence the number of cases.

State of Emergency: We did not consider the declaration of the State of emergency as a variable because it is not a policy per se. Rather it allows governments to directly and quickly impose rules.

Although these variables are interesting to be considered, this report intends to focus on measuring the effect of specific policies: Medicaid, Stay-At-Home order and Mask-Mandate order.

The variables - "*Percent at risk for serious illness due to COVID*", "*Nonelderly Adults Who Have A Pre-Existing Condition*", "*65+*" are essential to study the severity of mortality rates but not necessarily in studying the surge in new cases and are not considered as features in this report's model. The following correlation plots indeed show no indication of a significant correlation between these variables and the 7-day case-rate.



2.2. Measurement goals

2.2.1. Explanatory Model.

We believe that a causal relationship exists between the different policy decisions taken by the state government and the spread of COVID-19. Our explainer variables in the causal plot for short term policies are the number of days since these orders were issued. Stay at home order causes minimum mobility and human interaction in public space. Similarly, closure of non-essential businesses reduces the congregation of people that can act as a super-spreader for this virus. According to the CDC, masks have been proven effective in curbing the spread of COVID-19. So, we think that a causal relationship exists between the number of days since the orders have been issued and the case rate that the states are seeing in the last 7 days. We believe that early implementation of these policies will reduce the overall infection rate for the reasons mentioned above.

Similar to the pandemic-related policies, we are of the view that medicaid allocation will also have a causal effect on the COVID-19 infection rate. We believe that states' efforts by allocating additional medicaid will help better medical facilities for the vulnerable population, thus reducing the overall infection rate. Detailed discussion can be found in section 2.2.2.

The following plot uses the transformed variables that we plan to use in our modeling exercise.

Based on the initial EDA, we are considering the transformed variables as mentioned above for our modeling exercise. Section 3 gives more details into actual model runs, and limitations to some of these variables and how we are tweaking the overall approach.

2.2.2 Detailed discussion of variables being considered.

Government policies have a significant impact on overall human mobility. Enforcing or not enforcing mask mandates can play an important role in the overall spread of infection. Similarly, enforcement of stay at home order can avoid public congregation to help subside the infection rate. Large public gatherings without proper

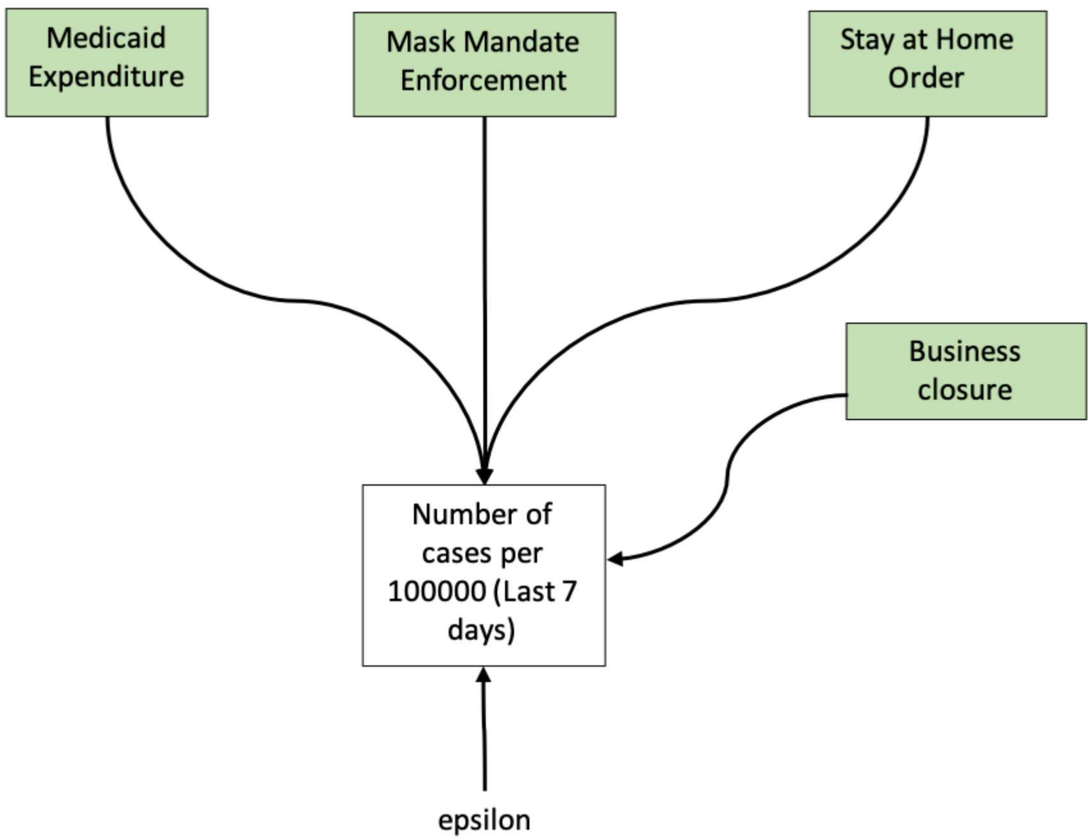


Figure 1: Causal Plot

care can cause widespread infections which can increase exponentially in a matter of a few days. According to the CDC, *"The incubation period for COVID-19 is thought to extend to 14 days, with a median time of 4-5 days from exposure to symptoms onset.(1-3) One study reported that 97.5% of people with COVID-19 who have symptoms will do so within 11.5 days of SARS-CoV-2 infection."* The fact that a person can unknowingly spread the virus in a 14 day period itself demands that stringent government regulations need to be in place to avoid human interaction. Analysis about how policies have affected COVID-19 spread is necessary to understand its impact. *We believe that there is a direct causal relationship between these policies and the case rate.* Moreover, we are of the view that it is important to understand what kind of policies need to be analyzed. State and federal governments have long-term policies like medicaid help for underprivileged in-place since many years. But due to the nature of this pandemic, the authorities have had to take some drastic measures and decisions like issuing stay at home orders, closures of non-essential businesses, mask mandates etc. to try and curb the infection.

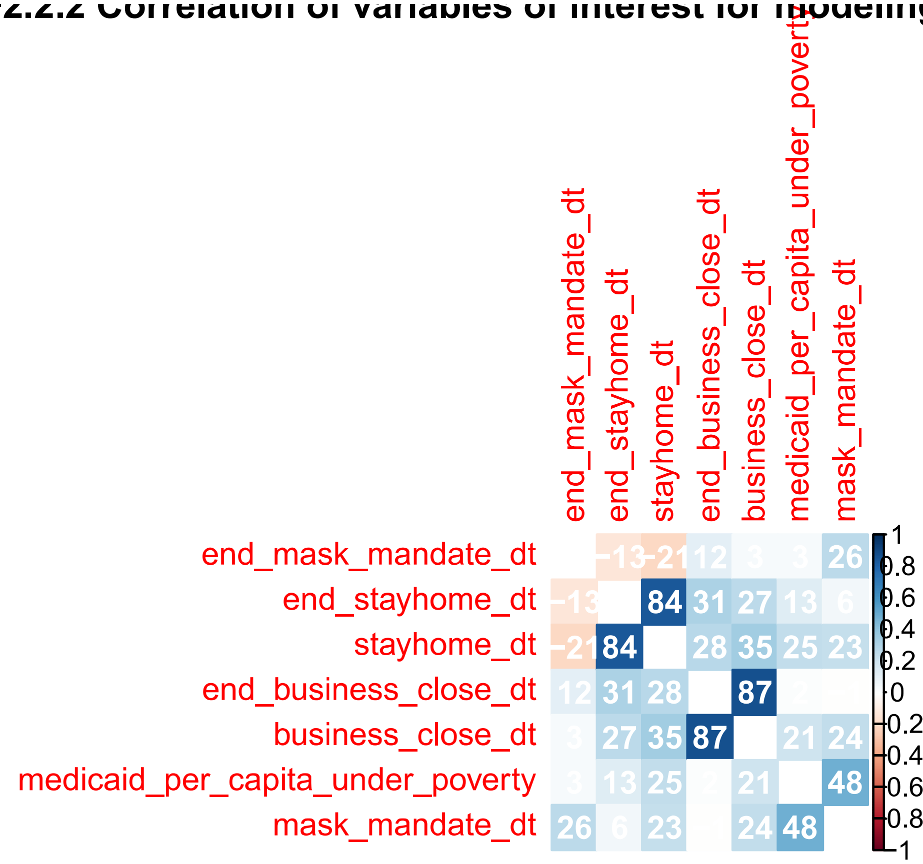
Variables that we are using for our model closely align with the overall research question in mind:

Medicaid per capita : This transformed variable that we are using provides insights into how the underdeveloped sections of the society are getting benefits from long term policies like Medicaid from the state government. In a way, such efforts can assist people with limited resources to use medical facilities. It helps the general health of this population and, as we suspect, their susceptibility to the virus.

Short-term policies : These policies include stay at home, business closure, mask mandate orders etc. that can directly affect the overall virus spread. These variables are related to how the government has tried to bring in behavioral changes by curbing freedom of movement in order to reduce the virus spread. Moreover, these are complex variables because a lot of these have delays in becoming effective, which is difficult to capture as a data point. For example, issuing orders like wearing masks and people actually following it can have some delays in becoming a ground-reality.

The box below displays the correlation between the variables used in our models. The top correlations are between the respective beginnings and ends of the business closing and the stay-at-home periods. As a result we do not use these pairs in the same models. All other pairwise correlations are below 50%

Fig–2.2.2 Correlation of variables of interest for modeling.



3. Models

3.1 Model cases over the past 7 days as influenced by Medicaid-Allocation

3.1.1 Structural Equation

$$log_2(normalized\_cases\_last7) = \beta_0 + \beta_1 \cdot log_2(medicaid\_per\_capita\_under\_poverty) + u$$

3.1.2 Transformations applied to Covariates.

Based on the EDA in section 2, we use these covariates in the model:

- Dependent Variable: Log transformed Case-count over the past 7 days.
- Independent Feature : Log Transformed medicaid-allocation per-capita under poverty in a state.

3.1.3 Overview of the Model results

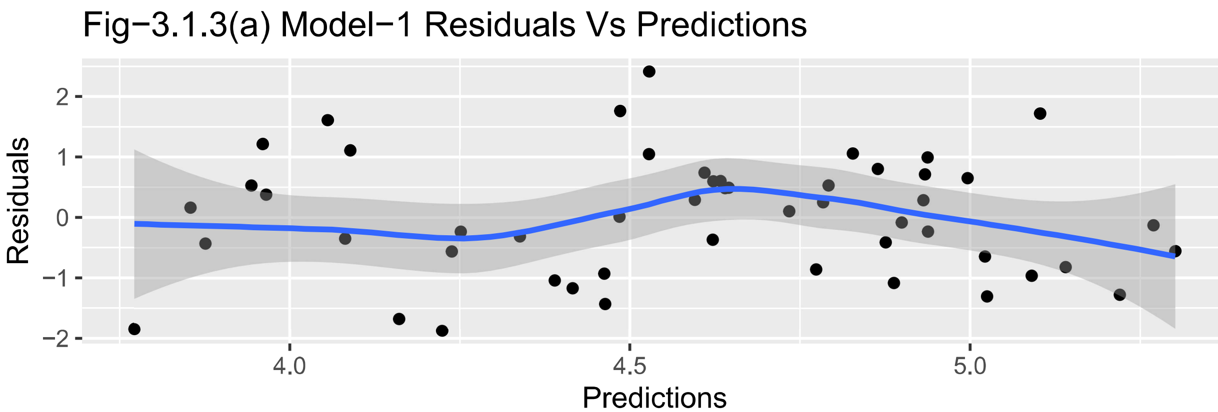
% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sat, Dec 12, 2020 - 05:03:32 PM

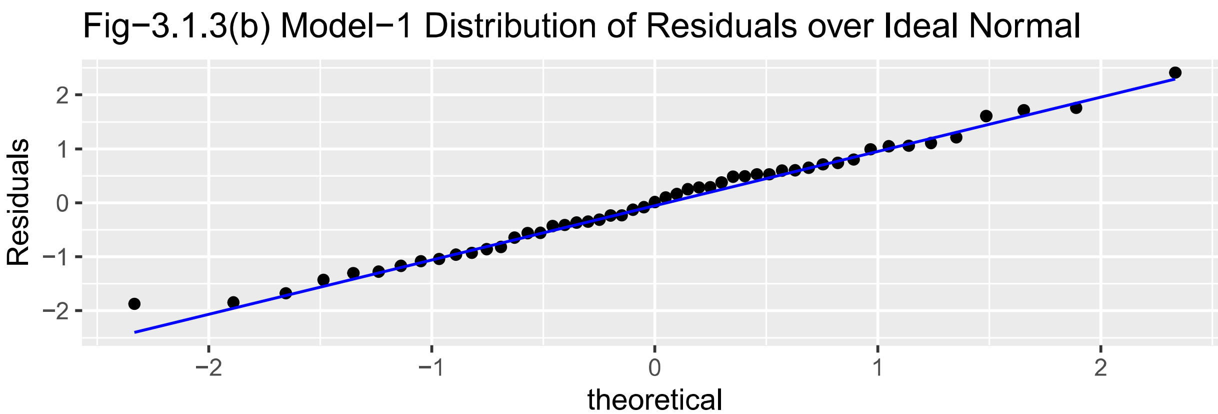
Table 1: Cases in Last-7 Days Medicaid-Allocation

	Dependent variable:
	log2(normalized_cases_last7)
log_medicaid_per_capita_under_poverty	-0.761^{***}
	(0.264)
Constant	9.731^{***}
	(1.788)
Observations	51
R^2	0.145
Adjusted R^2	0.128
Residual Std. Error	0.994 (df = 49)
F Statistic	8.308^{***} (df = 1; 49)
Note:	*p<0.1; **p<0.05; ***p<0.01

The Model-coefficient for the medicaid-allocation feature is significant as seen in the summary. Fig-3.1.3 below shows the Residuals vs Predictions plot which shows an almost linear trend and no major heteroskedasticity. The standard errors also seem to be close to normally-distributed over the full-range of predictions.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```





3.1.4 Brief explanation of practical significance of model.

To review the practical significance of the coefficient for Medicaid-allocation we look at the mean values our medicaid_per-capita and number of cases in last-7 days.

```
mean_medicaid_per_capita = mean(covid_curated$medicaid_per_capita_under_poverty)
mean_cases_last_7_days = mean(covid_curated$normalized_cases_last7)
log_mean_cases_last_7_days = log2(mean_cases_last_7_days)

mean_medicaid_per_capita
```

```
## [1] 115.2868
```

```
mean_cases_last_7_days
```

```
## [1] 31.11373
```

```
delta_log_mean_cases = (0.761 * (log2(mean_medicaid_per_capita+10) - log2(mean_medicaid_per_capita)))

delta_log_mean_cases
```

```
## [1] 0.09132538
```

```
change_cases_for10_dollar_increase = 2^delta_log_mean_cases

change_cases_for10_dollar_increase
```

```
## [1] 1.065348
```

Mean Medicaid-allocation per-capita across states as calculated from this report's data-set is ***$115.2868208***.

When all else is kept equal, if the **medicaid-allocation is increased by $10 per-capita**, working this change through this model's log-log transformations, Model-1 predicts that the normalized **cases in last-7 days will be reduced by around 1-case per 100000**.

3.2 Model cases over the past 7 days as influenced by Medicaid-Allocation and Stayhome-duration

This model introduces the **Stay-Home-date feature** into the model to model its influence on the Exponential-Growth.

3.2.1 Structural Equation

$log_2(normalized\_cases\_last7) = \beta_0 + \beta_1 \cdot stayhome\_dt + \beta_2 \cdot log_2(medicaid\_per\_capita\_under\_poverty) + u$

3.2.2 Overview of the Model results

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sat, Dec 12, 2020 - 05:03:34 PM

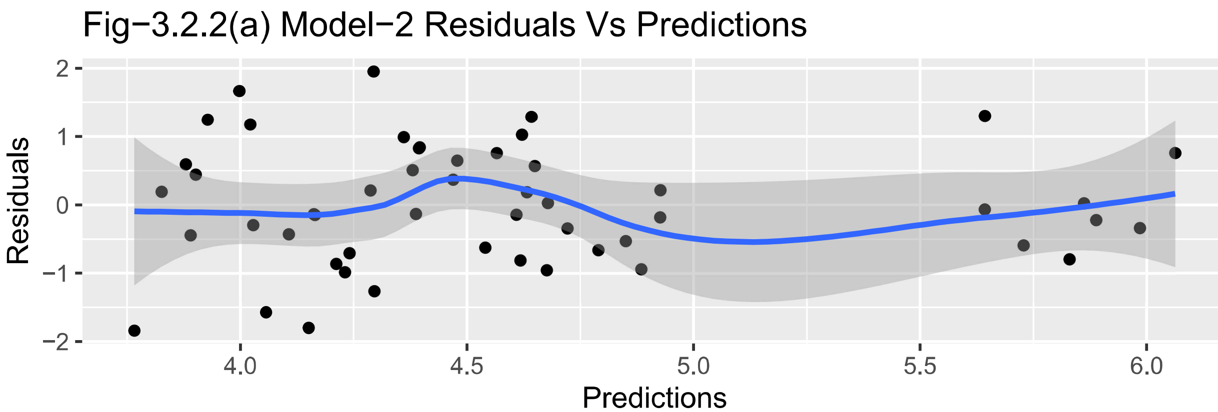
Table 2: Model2

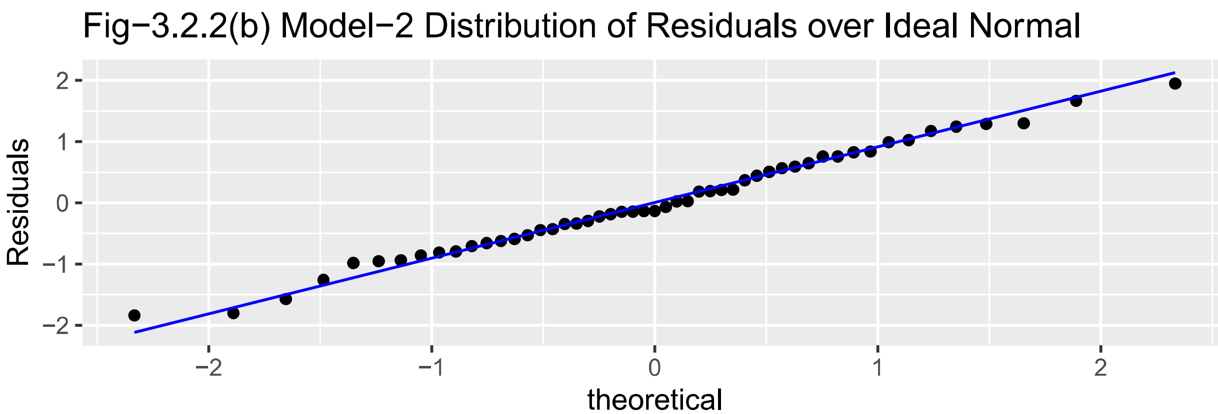
	Dependent variable:
	log_normalized_cases_last7
stayhome_dt	-0.006^{***}
	(0.002)
log_medicaid_per_capita_under_poverty	-0.556^{**}
	(0.241)
Constant	9.446^{***}
	(1.593)
Observations	51
R^2	0.337
Adjusted R^2	0.309
Residual Std. Error	0.885 (df = 48)
F Statistic	12.179^{***} (df = 2; 48)
Note:	*p<0.1; **p<0.05; ***p<0.01

In Model-2, The Model-coefficient for the medicaid-allocation feature is still significant while the feature for Stay-Home date seems to also be significant as seen in the summary above. Fig-3.2.2 below shows the Residuals vs Predictions plot which shows that the linear trend of residuals has changed but not to an extent that indicates confounding variables. The Residuals also show a more noisy heteroskedastic nature. This can be accounted for by using robust-linear-modeling tools which correct for heteroskedasticity and provide robust-coefficients. The standard errors also seem to be close to normally-distributed over the full-range of predictions.

This Model-2 will be the main model that this report uses to argue for the causal-theory set forth in Section 2.2.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```





3.2.3 Brief explanation of practical significance.

```
mean_medicaid_per_capita = mean(covid_curated$medicaid_per_capita_under_poverty)
mean_cases_last_7_days = mean(covid_curated$normalized_cases_last7)
log_mean_cases_last_7_days = log2(mean_cases_last_7_days)

mean_medicaid_per_capita
```

```
## [1] 115.2868
```

```
mean_cases_last_7_days
```

```
## [1] 31.11373
```

```
delta_log_mean_cases_model2_due_medicaid = (0.556 * (log2(mean_medicaid_per_capita+20) - log2(mean_medi

delta_log_mean_cases_model2_due_medicaid
```

```
## [1] 0.1283213
```

```
change_cases_for10_dollar_increase_model2 = 2^delta_log_mean_cases_model2_due_medicaid

change_cases_for10_dollar_increase_model2
```

```
## [1] 1.093021
```

Mean Medicaid-allocation per-capita across states as calculated from this report's data-set is ***$115.2868208***.

When all else is kept equal, if the **medicaid-allocation is increased by $20 per-capita**, working this change through this model's log-log transformations, model-2 predicts that the normalized **cases in last-7**

days will be reduced by just over 1-case per 100000.

```
delta_log_cases_model_2_due_delay = -0.006 * -5
delta_log_cases_model_2_due_delay
```

```
## [1] 0.03
```

```
change_cases_for5day_delay_model2 = 2^delta_log_cases_model_2_due_delay
change_cases_for5day_delay_model2
```

```
## [1] 1.021012
```

The other feature in this model is the number of days since the Stay-Home feature is implemented in states. To interpret the model-coefficient for this feature, consider a state which delayed implementing the Stay-home order by 5-days compared to the first states which responded to covid-hotspots. This 5-day delay is operationalized as a 5-unit decrease in variable.

When all else is kept equal, if a **state delayed its shut down by 5-days**, working this change through this model's log-normal transformations, model-2 predicts that the normalized **cases in last-7 days will be INCREASED by just over 1-case per 100000**.

3.3 Models with other important variables explored in EDA

In this model, other features identified in the Initial EDA as interesting to the causal-question about cases are included in the model to investigate their effect on coefficients and whether the features considered in Model-2 are robust to inclusion of other correlated input-variables.

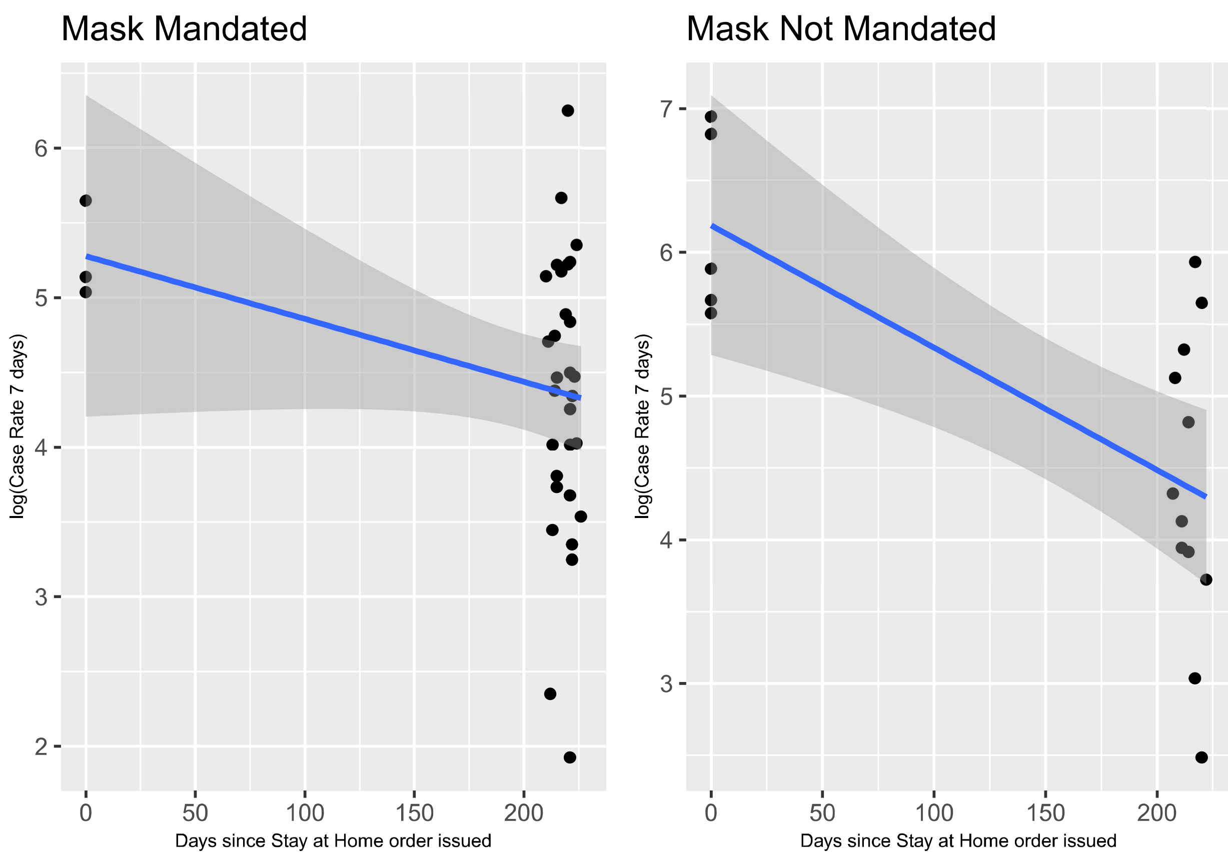
3.3.1 Structural Equation

$log_2(normalized\_cases\_last7) = \beta_0 + \beta_1 \cdot stayhome\_dt + \beta_2 \cdot log_2(medicaid\_per\_capita\_under\_poverty)$
$+ \beta_3 \cdot mask\_madate\_dt$
$+ \beta_4 \cdot business\_close\_dt\_dt$
$+ \beta_5 \cdot duration\_stay\_home$
$+ \beta_6 \cdot stayhome\_dt * mask\_mandated$
$+ u$

3.3.2 Transformations we applied and why; if any.

The stay at home directive was not uniformly issued by all states. The plot below indicates that the efficiency of the stay-at-home order was quite different (whether it was issued or not) if masks were mandated than if they were not (note the scale of the vertical axis). As such we introduce the variable mask_mandated*stayhome_dt

```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```



We also introduce the following variables: Mask mandate date: this is another lever used by most states to contain the spread of the disease.

Close of business date: around the same time as the stay-at-home mandate was introduced most states closed ono-essential businesses. This measure was implemented by more states than the stay at home measure.

Duration of the stay home period, which may have an effect since it was much shorter for some states than others.

3.3.3 Overview of the Model results

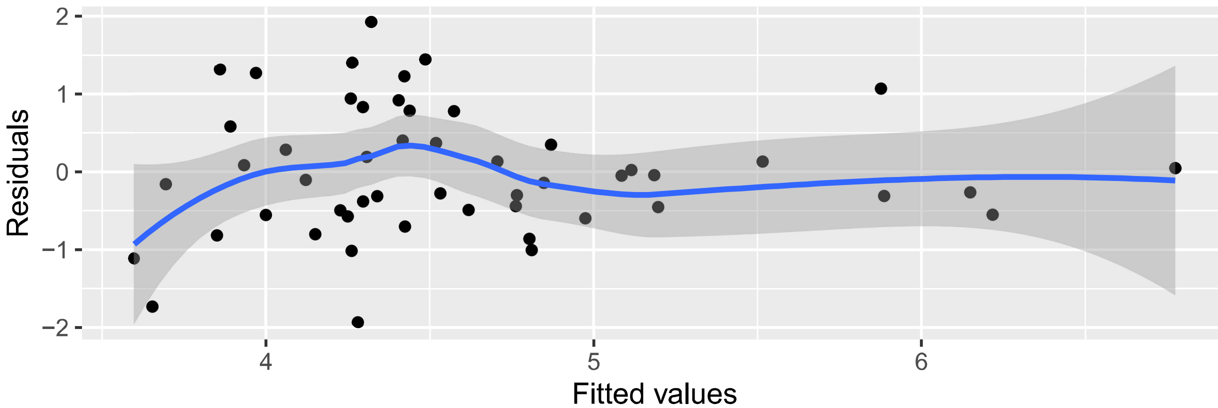
% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sat, Dec 12, 2020 - 05:03:35 PM

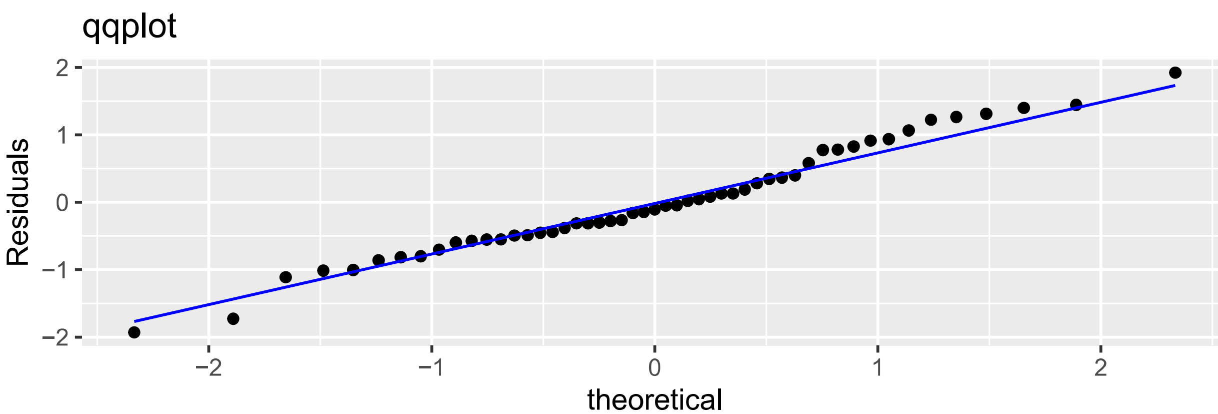
Table 3: Model-3 - Including other interesting variables

	Dependent variable:
	log_normalized_cases_last7
mask_mandate_dt	0.002
	(0.004)
stayhome_dt	-0.008^{***}
	(0.002)
mask_mandated	-1.311
	(0.940)
business_close_dt	-0.002
	(0.005)
duration_stayhome	-0.004
	(0.004)
log_medicaid_per_capita_under_poverty	-0.697^{**}
	(0.287)
stayhome_dt:mask_mandated	0.007^{*}
	(0.004)
Constant	11.011^{***}
	(1.939)
Observations	51
R^2	0.407
Adjusted R^2	0.310
Residual Std. Error	0.884 (df = 43)
F Statistic	4.215^{***} (df = 7; 43)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

A look at the graphs below indicates from top to bottom, a more pronounced heteroskedasticity of the error term than in model 2, that this model still underestimates the case rate for higher rates and overestimates it for lower cases by the same amounts as in model 2, and that the residuals distribution is deviating further from normality than in model 2.

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```





Findings:

- The date of the mask mandate is not significant. This is surprising, especially in light of the CDC's recommendations. However, as mentioned earlier, the mandate was not followed in a consistent manner.
- Duration of stay home period is not significant. We believe that the end date of the stay at home mandate.
- the date of the close of businesses is not significant. This may be due to the fact that it is a less drastic version of the stay at home measure.
- the fact that the mask mandate was enforced has only a small statistical significance (at 92%) on the effect of the stay at home mandate.
- Our previous explainers remain statistically significant, a sign of robustness of model 2.

3.3.4 Brief explanation of practical significance of model.

In terms of practical significance, we believe that there is very little or no difference between this model and the previous one.

4. Limitations

4.1 I.I.D Data

The dynamics of the Covid19 Pandemic and its spread in the US and across the world is a complex yet fascinating study. The number of confounding variables related to population-mobility within regions is endless and any variable can result in large hotspots in the cases.

For instance,

- The hotspots in Italy especially was correlated to the high number of immigrants working in the Fashion industry from China who travelled to China and back to cramped work-environments. Ref-7
- The spread in Mid-Western US states was closely tied to the Sturgis Motorcycle event in August of 2020. Ref-8
- This viral-spread behavior is obvious even at small scales such as the super-spreader events at the White-House Ref-9

In light of these geographical and demographic confounding-variables which ties cases in states, independence of data collected from the US states for this study cannot be upheld with strong arguments.

Identical-distribution of covariates being considered in these models is also not supported by strong arguments owing to the fact that the density in population differs between states to a large extent. Unfortunately, the political-initiative of governments in these states is also a confounding variable which might affect the number of cases seen in those states.

4.2 No perfect collinearity in independent features

To assess if there is high collinearity among the independent variables we can use the Variance-Inflation-Factor test. Below are two section, for Model-2 and Model-3 respectively.

VIF-values for Model-2 are close to 1 showing their presence in the model does not unjustifiably inflate the explained variance. This does not provide evidence to invalidate the Model-2.

However, the same for Model-3 which includes many other related variables shows that many of these variables inflate the explained variance due to collinearity. Considering how variables like `mask_mandated`, `stayhome_dt:mask_mandated` cause high inflation-factors, there is evidence to not accept results from Model-3.

VIF for Model-2

```r
car::vif(m2)
```

```
##                       stayhome_dt log_medicaid_per_capita_under_poverty
##                           1.05469                               1.05469
```

VIF for Model-3

```r
car::vif(m3)
```

```
##                   mask_mandate_dt                           stayhome_dt
##                          6.325783                              2.482016
##                     mask_mandated                     business_close_dt
##                         12.807035                              1.269580
##                 duration_stayhome log_medicaid_per_capita_under_poverty
##                          1.508844                              1.494417
##         stayhome_dt:mask_mandated
##                          9.274756
```

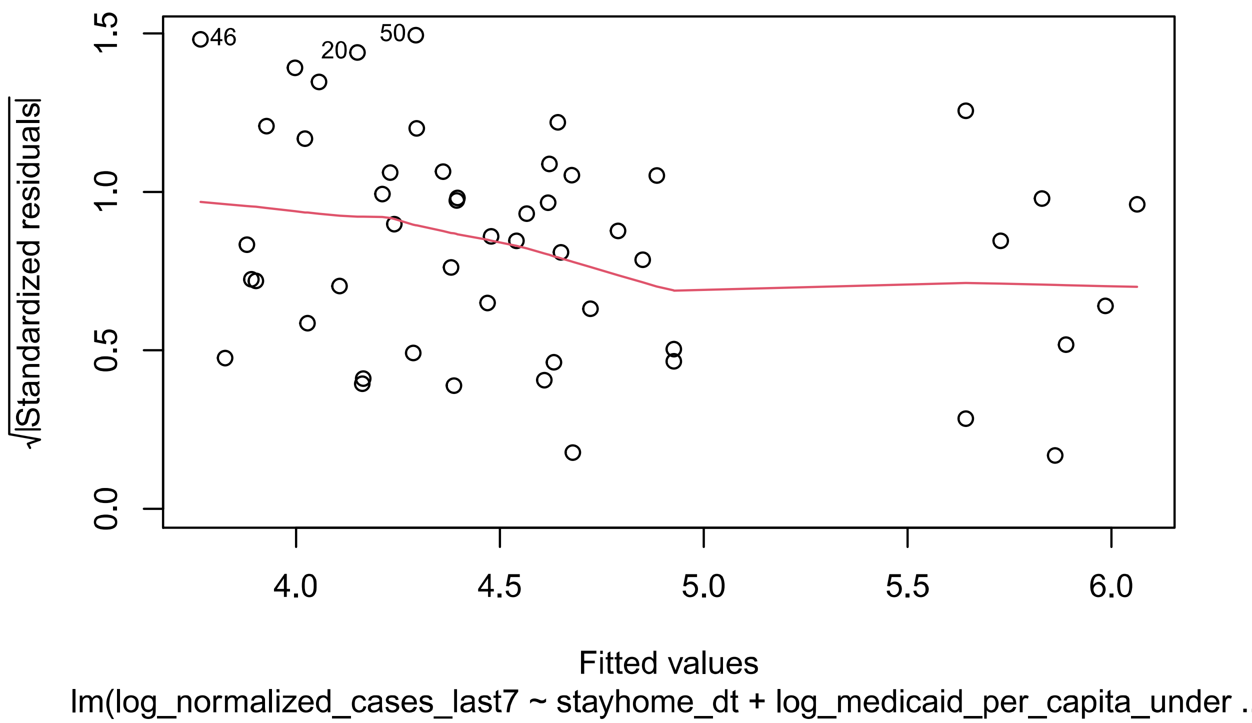
4.3 Homoskedastic distribution of residuals

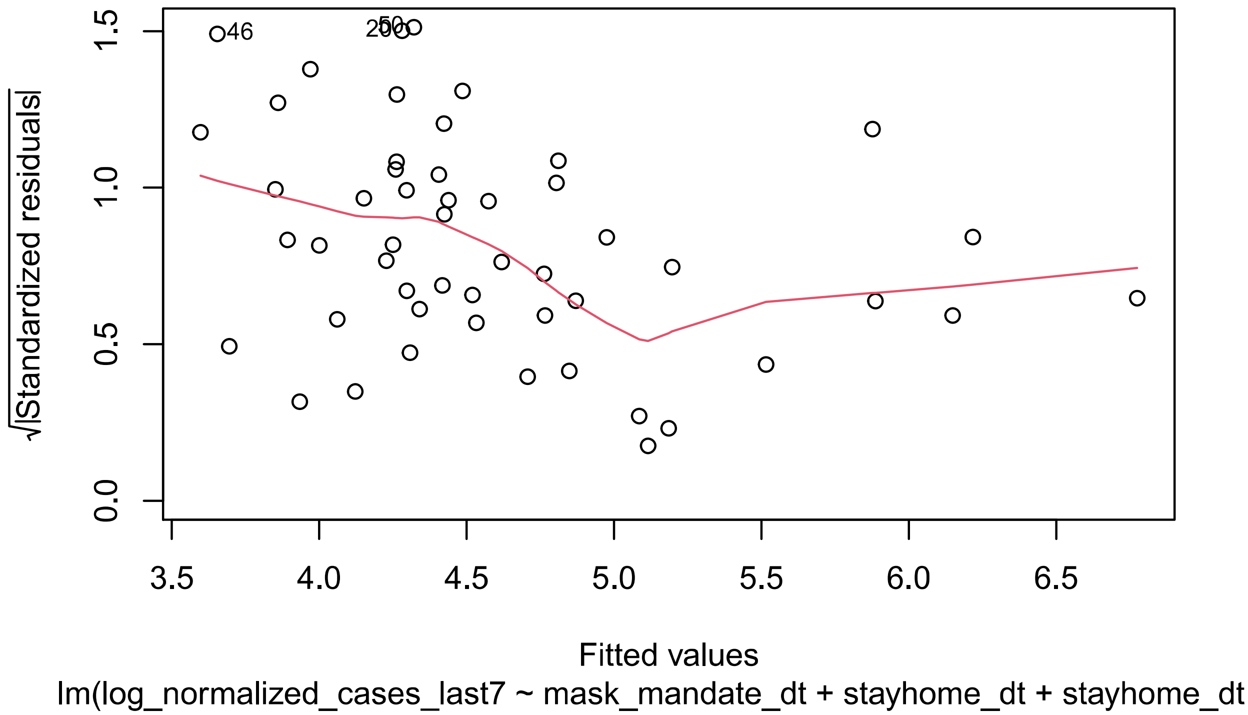
To understand if models in questions display homoskedastic residuals across the range of predictions, we look at the diagnostic plot from the models.

Fig-3.2.4.4(a) shows this plot for Model-2. The trends of Std-error in residuals shows some trends in the lower-range of predictions. This might indicate that the features Medicaid and Stayhome-dates are noisy at lower prediction ranges. Given that the trends in the Std.errors are not too extreme, and to account for these heteroskedastic tendencies in the model, we look at estimating the coefficients after adjusting for heteroskedasticity using vcovHC. From the table below, we see that the coefficients after adjusting for heteroskedastic residuals which can be utilized to explain practical significance.

```
##
## t test of coefficients:
##
##                                       Estimate Std. Error t value  Pr(>|t|)
## (Intercept)                          9.4464612  1.7317937  5.4547 1.687e-06
## stayhome_dt                         -0.0059892  0.0014011 -4.2746 9.041e-05
## log_medicaid_per_capita_under_poverty -0.5563731  0.2670344 -2.0835   0.04255
##
## (Intercept)                         ***
## stayhome_dt                         ***
## log_medicaid_per_capita_under_poverty *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Fig-3.2.4.4(b) shows this plot for Model-3. The trends of Std-error are majorly disrupted across the range of predictions which merits this model from being invalidated.





4.4 Investigation of Outliers in Data

This section investigates outliers in data-points which might have unjustified influence on the fit of the model.

Fig-4.4(a) shows outliers marked by the Cook's Distance calculated for the corresponding data point. For Model-2, Medicaid and Stayhome data from these 2 states - [***35-North Dakota , 46 - Vermont***] are identified as outliers. To mitigate the effect of outliers, robust-linear-models (rlm) could be evaluated. These models are built to account for outliers.

Fig-4.4(a) Outliers in Data in Model-2

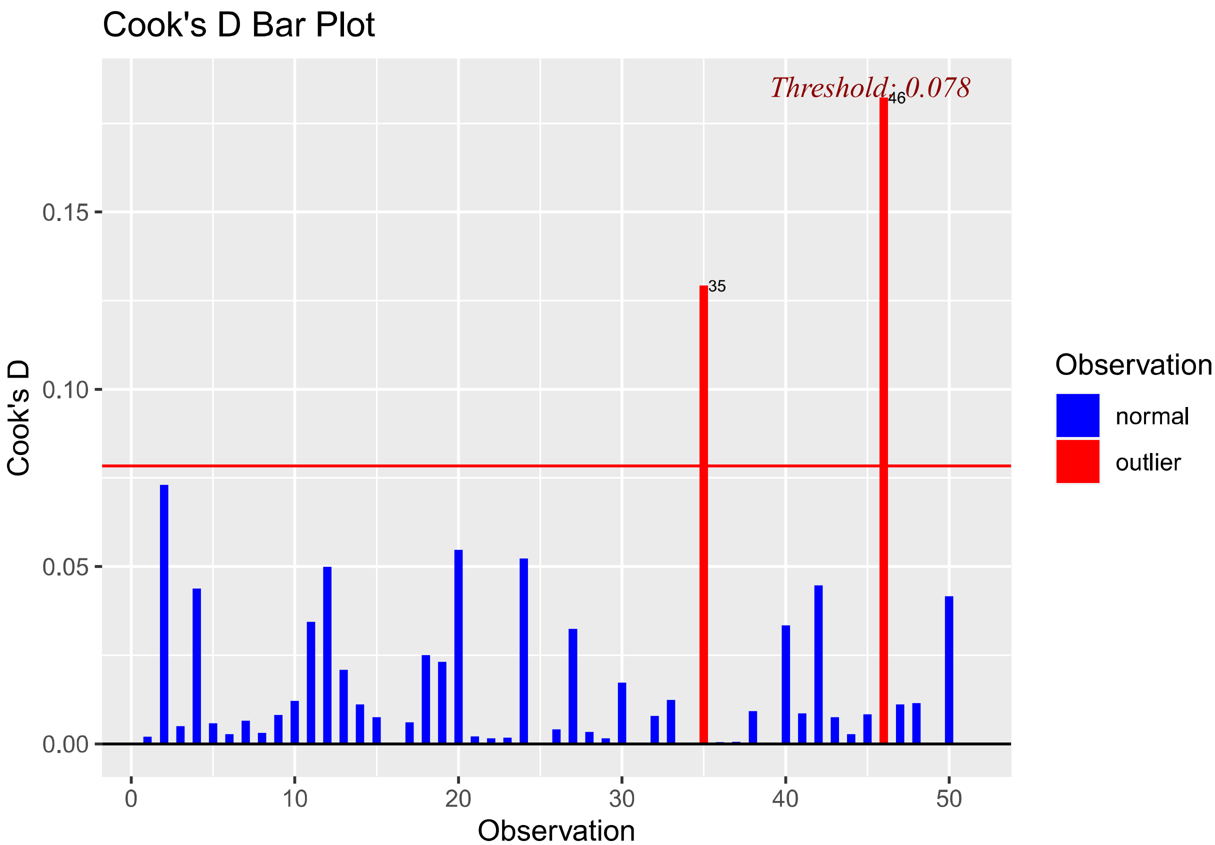
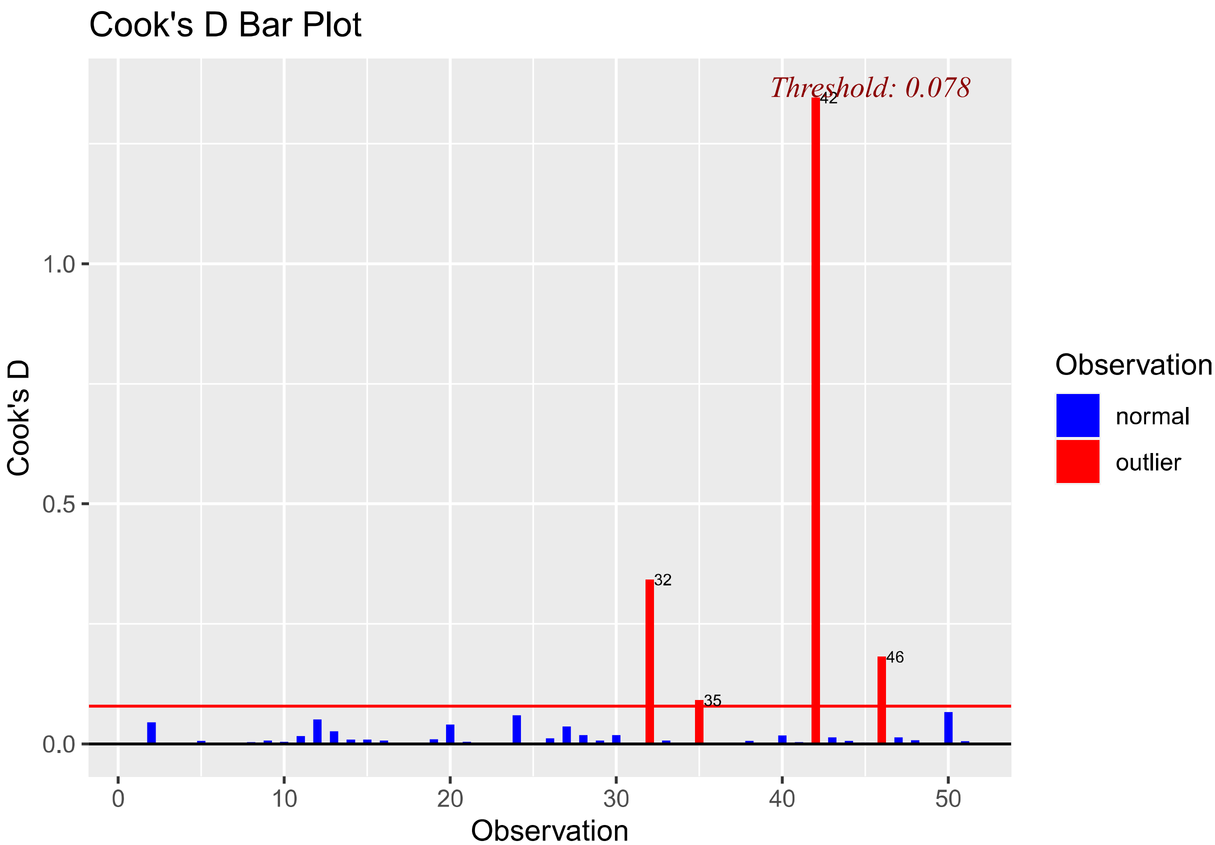


Fig-4.4(b) for Model-3 shows much sharper outliers marked by the Cook's Distance calculated for the corresponding data point. For Model-3, Medicaid and Stayhome data from 2 states : [***32-New Mexico, 35-North Dakota, 42-South Dakota, 46-Vermont***] are identified as outliers. A robust linear-model can be employed to account for these outliers.

As a note on using robust-linear-models, this study intends to further investigate why certain states are marked as outliers and if these outliers say something about the condition or whether these outliers can be omitted.

Fig-4.4(b) Outliers in Data in Model-3



Cook's D Bar Plot
Threshold: 0.078
42
32
46
35
1.0
0.5
0.0
Cook's D
0
10
20
30
40
50
Observation
Observation
normal
outlier

5. Summary Of Models

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sat, Dec 12, 2020 - 05:03:37 PM

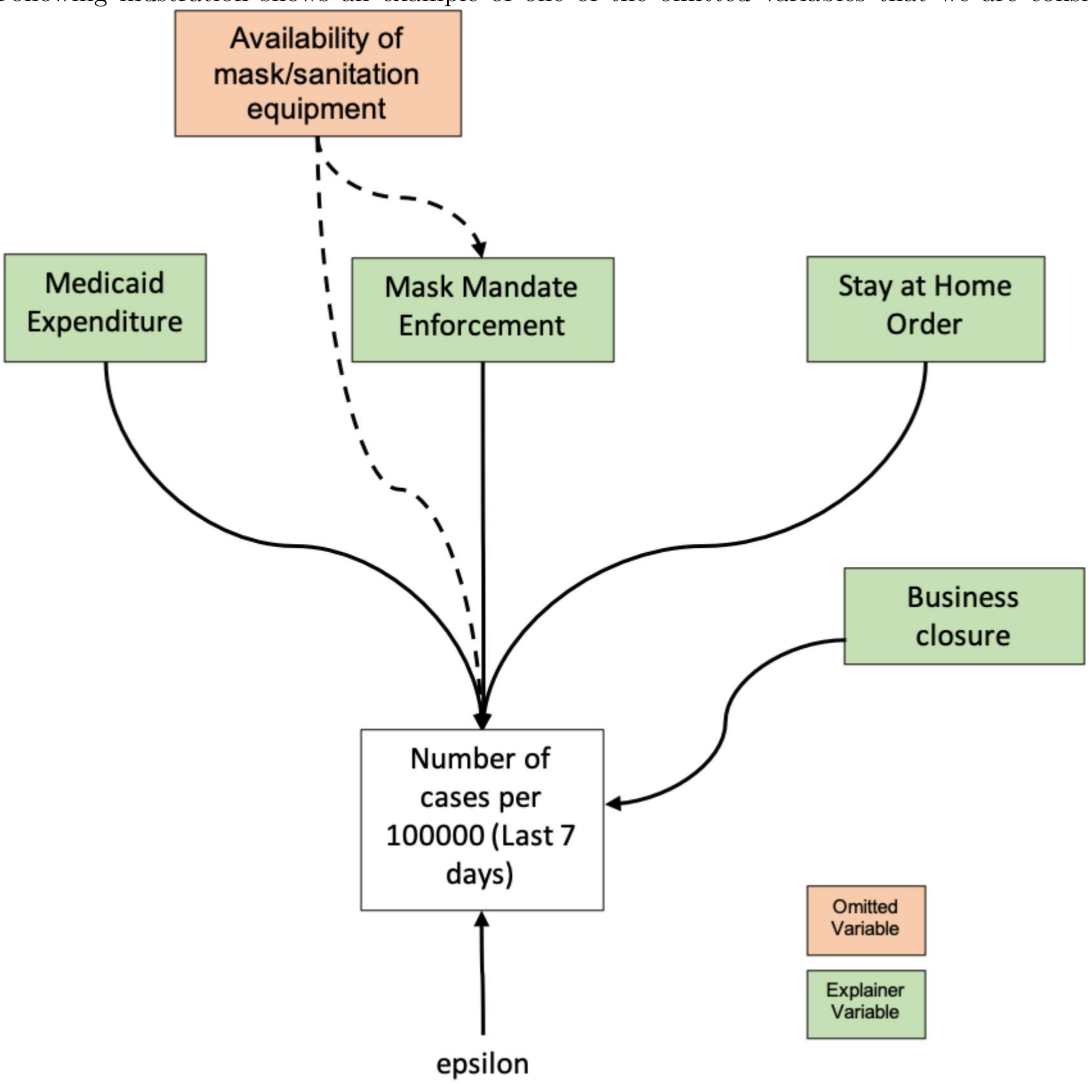
Table 4: Models for 7-day case rates

	Dependent variable:		
	log2(normalized_cases_last7)	log_normalized_cases_last7	
	(1)	(2)	(3)
mask_mandate_dt			0.002
			(0.004)
stayhome_dt		−0.006***	−0.008***
		(0.002)	(0.002)
mask_mandated			−1.311
			(0.940)
business_close_dt			−0.002
			(0.005)
duration_stayhome			−0.004
			(0.004)
log_medicaid_per_capita_under_poverty	−0.761***	−0.556**	−0.697**
	(0.264)	(0.241)	(0.287)
stayhome_dt:mask_mandated			0.007*
			(0.004)
Constant	9.731***	9.446***	11.011***
	(1.788)	(1.593)	(1.939)
Observations	51	51	51
R^2	0.145	0.337	0.407
Adjusted R^2	0.128	0.309	0.310
Residual Std. Error	0.994 (df = 49)	0.885 (df = 48)	0.884 (df = 43
F Statistic	8.308*** (df = 1; 49)	12.179*** (df = 2; 48)	4.215*** (df = 7;
Note:			*p<0.1; **p<0.05; ***p<0

As per our discussion of the three models, we believe that model 2 is the most appropriate for our research question

6. Discussion Of Omitted Variables

The explainer variables that we are using in our model around government policies are the dates/ days since the original policy was announced by the respective states. Given the nature of this pandemic, and how it demanded extreme decisions like business closures, stay at home orders etc. it is necessary to look at how and when these policies actually came into effect and became a ground reality. It is crucial to understand factors and all aspects that lead to delayed implementation of the announced policies. We are considering these delaying factors as a part of discussion around omitted variables. Our variables for this discussion revolve around how those metrics impacted the actual implementation of policies and thus the overall number of days since policies were announced. For example, even if masks were made mandatory on a particular day, given that practically fulfilling the demand for those masks took time, it affected our output variable i.e. COVID-19 case rate and the explainer variable i.e. mask mandate policy date. Following illustration shows an example of one of the omitted variables that we are considering.



Following section goes through each of these omitted variables and its significance.

Large public gatherings : This variable gives details about public meetings like protests, election rallies or any such event that can increase in-person interaction. Our data is until October and there were many such occasions of gatherings leading to the 2020 election. Such events can spike the overall case rate and cause violations of Stay at Home/ Shelter at home orders. Such violations impact the overall implementation of policies and thus contradicting our explainer variable of 'days since stay at home was issued'. This metric negatively impacts our stay at home variable and positively affects the overall case rate for the infection. Hence, the omitted variable bias is negative and towards zero.

Availability of mask/sanitation equipment : During the initial wave of the COVID-19 pandemic, there was a huge surge in demand for N-95 face masks along with sanitizers and cleaning wipes. The demand was so high that most states reported a shortage of these supplies, which demanded immediate action from the government. This variable directly impacts the surge in the infection case rate and affected when the mask mandate policy realistically came into effect. It negatively affects the mask policy and positively affects the case rate. Hence, the omitted variable bias is negative and towards zero.

Percentage of Low-income groups : Many people falling in the low income group are essential workers and have to get out to work. Moreover, their reliance on public transportation makes them even more vulnerable to getting exposed to the virus. Also, the medicaid expenditure for the state depends on this section of the population. As the number of low income groups increases because of economic meltdown, the per-capita estimates would still go down to some extent as the budget increase won't be as rapid as the increase in number of low income groups. At the same time, case rate will go up. Thus the variable will have a negative effect on the medicaid policy and a positive impact on the case rate. Thus, the omitted variable bias is negative and towards zero.

Illegal re-opening of businesses : After the initial wave of the pandemic, there were reports of non-essential businesses opening illegally, before the state law permitted them. As a result of these violations, implementation of closure policies takes a hit and causes delays in actually turning it into a reality. Also, it leads to increased mobility, further increasing the overall infections. This variable has a negative effect on the business closure policy and positive impact on the COVID-19 case rate. Thus the omitted variable bias is negative and towards zero.

Lack of public seriousness : To ensure that people are aware of the seriousness of this pandemic, every state has tried to promote awareness about social distancing and various government policies. To some extent, we believe that these policies have not been respected and acknowledged by people and there have been instances of blatant violation. There have been instances of people denouncing the policies, thus bringing down the overall effectiveness since announcement. This has a negative impact on our policy variables and a positive impact on the case rate. Hence, the overall omitted variable bias is negative and towards zero.

7. Conclusion

In early 2020 as the disease hit the north east region of the United States, most other states were relatively safe. These states instituted severe restrictions on their populations. As the disease propagated through the country. Some states were quick to enforce measures while others only did so when the rate of spread of the disease accelerated. Our analysis shows that not only did government restrictions have an effect on the spread of Covid-19, but that the earlier these restrictions went into effect the slower the spread of the disease. This is consistent with the notion that as time passes the disease propagates at an exponential rate, so any delay in instituting these policies will have a multiplicative effect on the increase in the number of cases.

In addition we see a strong indication that long term health policies, such as medicaid expenditure per capita also have an effect on the current rate of increase in the total number of covid cases (as of 10/31/2020). This is an interesting finding that should be further studied as it may have serious implications in the defense against future pandemics.

In order to further study this question it would be helpful to have access to time series of the number of cases so that government actions can be put in the context of the evolution of the disease at the state level.

8. Appendices

8.1 References

1. Medicaid Is Essential for Workers as COVID-19 Spreads : Ref-1
2. Connecticut Stay-at-Home Policy Implementation dates : Ref-2
3. New Mexico Stay-at-Home Policy Implementation dates : Ref-3
4. Oklahoma Stay-at-Home Policy Implementation dates : Ref-4
5. Texas Stay-at-Home Policy Implementation dates : Ref-5
6. California Stay-at-Home Policy Implementation dates: Ref-6
7. A cascade of causes that led to the COVID-19 tragedy in Italy and in other European Union countries: Ref-7
8. COVID-19 Outbreak Associated with a 10-Day Motorcycle Rally in a Neighboring State — Minnesota, August–September 2020: Ref-8
9. Viral genome sequencing places White House COVID-19 outbreak into phylogenetic context: Ref-9
10. Super-spreading events initiated the exponential growth phase of COVID-19 with R_0 higher than initially estimated Ref-10
11. Massachusetts superspreader: Biogen conference tied to 300,000 coronavirus cases Ref-11
12. KFF State Total Medicaid Spending Ref-12