Covid-19 and Income Relief

Alice Ye, Amber Rashid, Dhyani Parekh, Ziling Huang
W203 Fall 2020 Tuesday 4pm PT
Lab 2

1. Introduction

Background

The onset of Covid-19 has had a myriad of expected and unexpected effects on our world and daily lives. From hospital capacities pushed to their ultimate limits, disappearing toilet paper, dolphins in the Venice Canal, and the closing of industries and businesses, there has been a balancing act in trying to return to normalcy but also minimize the spread of Covid-19.

The downturn of the American economy has been one of the biggest negative effects of the pandemic outside of the spread of the disease and death toll. One key area that both limits Covid-19 spread but also stimulates the economy is income relief. Income relief plays an essential role in reducing the necessity for people to work in jobs requiring high levels of in-person interaction which serves to limit infection rates. Though the benefits of income relief seem obvious, there have been conflicting approaches with liberal approaches encouraging income relief to prevent people to go find work and conservatives recommending “getting back to work” and fully opening the economy up again and leveraging “herd immunity.” Given its controversial but important outcomes, for our lab we have decided to explore the effects of Weekly Unemployment Income and stimulus with other variables on case rates per 100,000 people, as shown below in Figure 1.01.

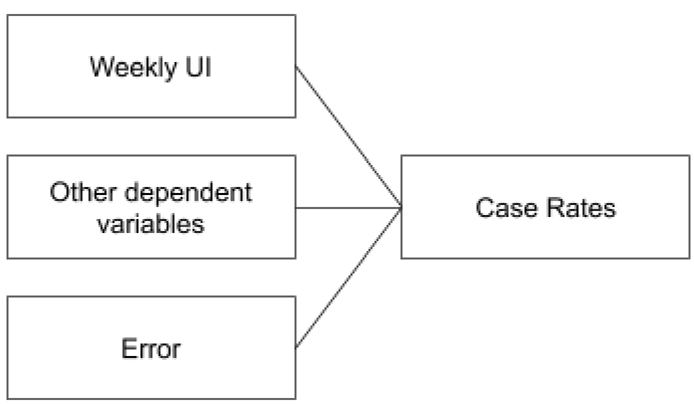


Figure 1.01 Relational Diagram for Case Rate and Weekly UI

Purpose and Research Question

For our research question, we leverage the Covid-19 dataset to explore if fiscally conservative or liberal approaches fare better for Covid-19 infection rate containment using a descriptive model.

Infection rate is operationalized and measured using “Case Rate per 100k” per State as the outcome variable. Fiscal conservativeness is operationalized using “Weekly Unemployment Insurance (UI) and Stimulus Relief” per State as the explanatory variable where we are trying to understand the direction and strength of its associative relationship with the outcome variable. The weekly UI variable is defined as the maximum amount of relief available to citizens of the state including unemployment insurance, federal pandemic relief, state level pandemic relief, and any other income relief available. We believe it is appropriate to use point-in-time, life-to-date data till end October 2020 (when the dataset was created) to measure the statistical and practical significance of the associative relationship between these variables.

The measurement goal and the research question is:

How effective was UI and stimulus relief at reducing case rate per 100k across states?

This will help us understand if the perceived benefits of income relief do indeed work for limiting the spread of disease in a pandemic and thus should be more widely adopted. For brevity, henceforth, we will be referring to the “Weekly UI and Stimulus Relief” variable as “Weekly UI” and “Case Rate per 100K” as “Case Rate”. For some figures “Weekly UI” is labeled as “UI” in order to make them easier to read and interpret.

2. Model Building Process

Preliminary Data Cleansing

For our exploratory data analysis, we needed to clean the dataset to consider the other variables for our model. We treated the data as follows:

- Transformed cases by race/ethnicity to account for strings. Values of “<.01” were converted to be 0.005 as the midpoint between 0 and 0.01.
- We created new policy variables by using the dates provided to calculating the number of days since the first diagnosed Covid-19 case in the U.S. (January 20, 2020) to when the policy was implemented.
- Removed D.C. as a state/territory in our dataset due to being an outlier for population density. D.C. had high population density due to being a governing district and not a state. It is also not considered part of the 50 states of the U.S.

- Created new variables for state characteristics by extracting State Governor's party affiliation from Governor's name and turning it into binary form where Republican = 1 and Non-Republican = 0.
- Added missing dates on policy implementation in the data through research.

Exploratory Data Analysis

We started by exploring our outcome variable - case rate per 100k:

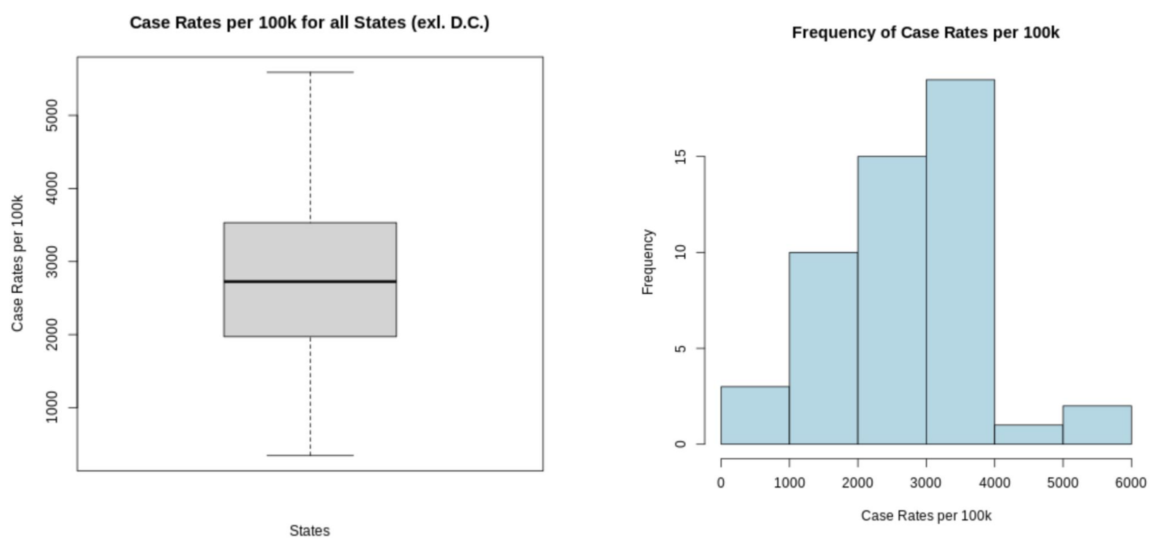


Figure 2.01 Case Rate Box Plot and Histogram

Delving into case rate, we see that it has a median value of 2,633 cases per 100,000 people, with the first and third quartile range being 2,040 to 3,516 cases. The lowest case rate is 344 people for the state of Vermont. Only 3 out of the 50 states had case rates less than 1000 - Vermont, New Hampshire, and Maine. On the other end, North Dakota, South Dakota, and New York had the highest case rates and were the only 3 states to exceed 4,000 cases per 100,000 people. This can be seen in the heavy left skewing in the histogram - there is a sharp increase post 1000 cases and sharper drop post 4000 cases. The lowest case rates are seen in the smaller in size states in the North east, whereas on the highest end we see the highest case rate in New York, the most densely populated state in the North east. No outliers were found for case rate.

Next, we explored our main dependent variable, weekly UI:

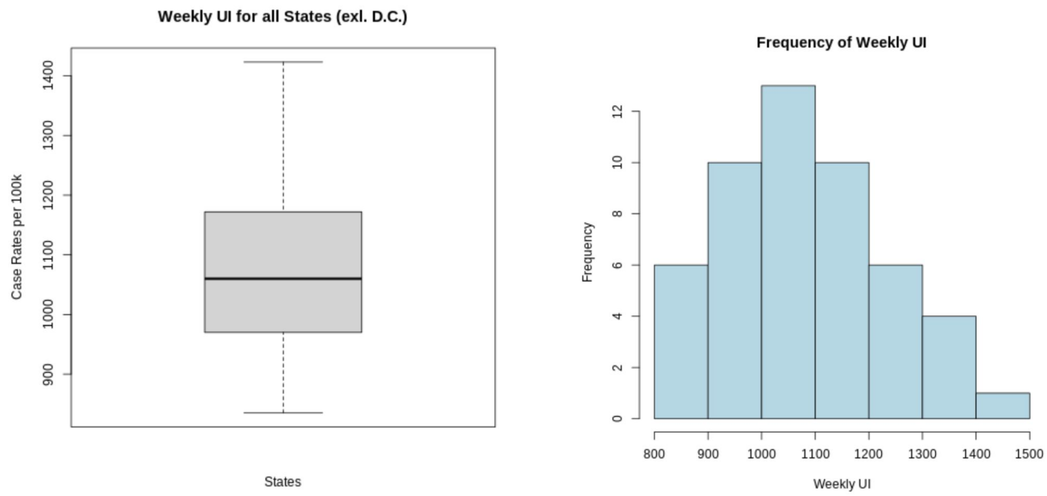


Figure 2.02 Weekly UI Box Plot and Histogram

From above, we see the maximum dollar amount of income relief available in the US was $1,432 in the state of Massachusetts. Majority of states offered maximum income relief between $974 to $1,162 (1st and 3rd quartile). Only 6 states offered income relief less than $900. No outliers were found for weekly UI.

To start out exploring our research question on income relief and Covid-19 spread, we plotted case rate against weekly UI per state to see if any preliminary relationship can be observed.

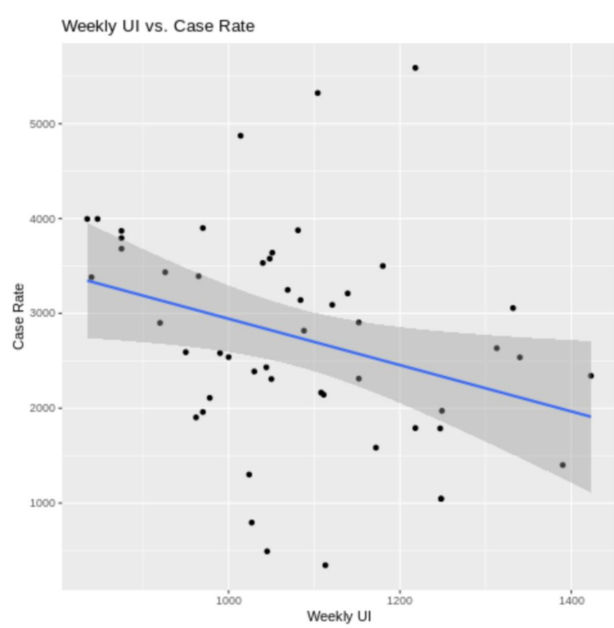


Figure 2.03 Case Rate versus Weekly UI Scatterplot

In Figure 2.03, we found an approximately linear relationship with case rate decreasing as weekly UI increased. We found a weak correlation of 0.31 between case rate and weekly UI and did not observe any improvement in the correlation from transforming the variables. Since this is a weak correlation, we hope to improve the association through our model building.

In order to further our exploration, we created correlation panels for our key variables, case rate and weekly UI, for racial, age, policy, state-level and mobility factors. We looked at the strength of the associations of the correlation coefficients, as follows:

Range	Strength of association
0	No association
0 to ±0.25	Negligible association
±0.25 to ±0.50	Weak association
±0.50 to ±0.75	Moderate association
±0.75 to ±1	Very strong association
±1	Perfect association

Figure 2.04 Strength of Correlation Values

Racial Factors

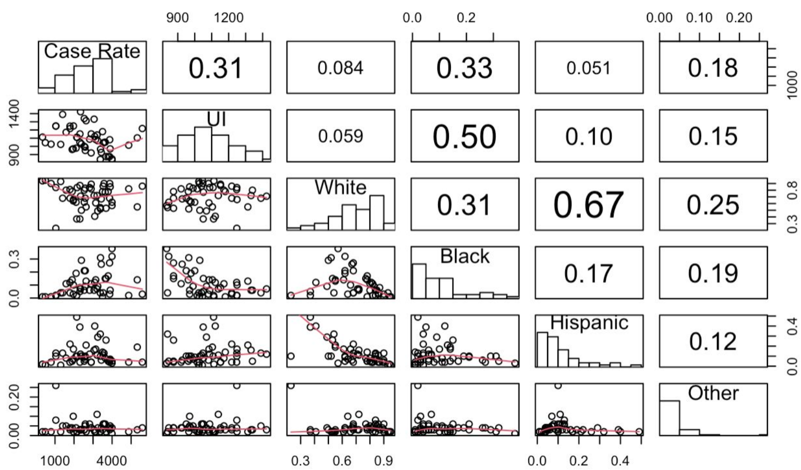


Figure 2.05 Race, Case Rate, and UI Panel

When exploring race correlations with case rate and weekly UI, in the above figure we see a weak association between case rate and the percentage of black population, but a weak to moderate association between the percentage of black population and weekly UI. This suggests that unemployment may disproportionately affect the black population, which we know to be true due to the historical and systemic issues around black access to education and opportunity.

Age Factors

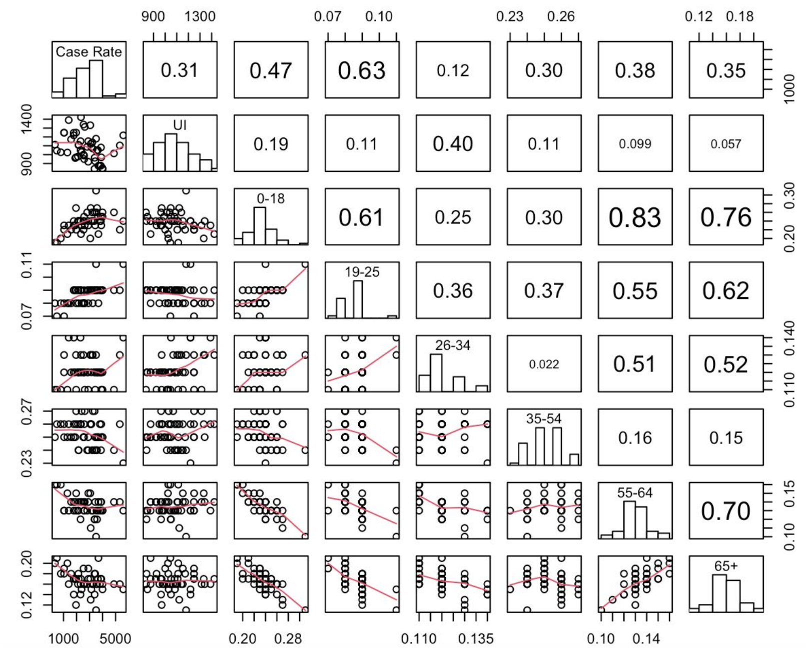


Figure 2.06 Age, Case Rate, and UI Panel

For age, we expected middle aged folks to have a higher correlation to case rates being that they are the age group which consists of most working Americans. However, we found the age groups 0-18 and 19-25 to have weak and moderate correlations as compared to the older folks, who have negligible to weak correlations. Upon investigation and further research, it does make sense that the percentage of young folks has an association with case rate for a few reasons. Young folks are more likely to congregate to crowded settings like schools, colleges, and universities and tend to have better outcomes for Covid-19. Thus, they perceive less risk of the disease and may not social distance as much, and can in turn become asymptomatic spreaders.

Age 26-34 had a weak association with weekly UI which validates that the highest age group of working Americans would also require weekly UI at some point in their lives.

Policy Factors

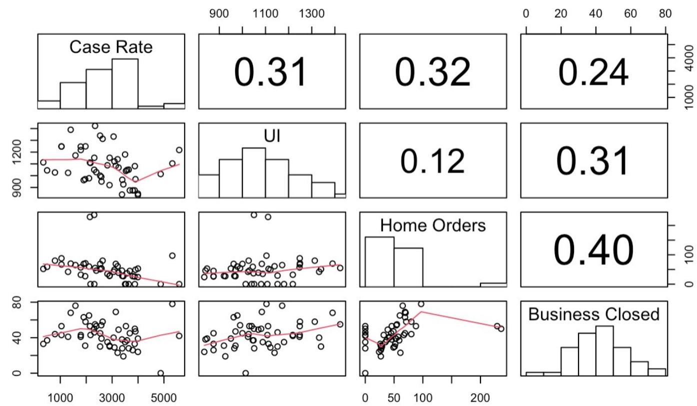


Figure 2.07 Days of Stay-at-home and closed business orders, Case Rate, and UI Panel

For policies, we first explored if there was a correlation between the number of days of stay-at-home and business closure orders and case rate where we found a weak association with stay-at-home orders. There is also a weak association between business closure and weekly UI, which suggests that business closures may have led to increasing needs for income relief due to layoffs and furloughs.

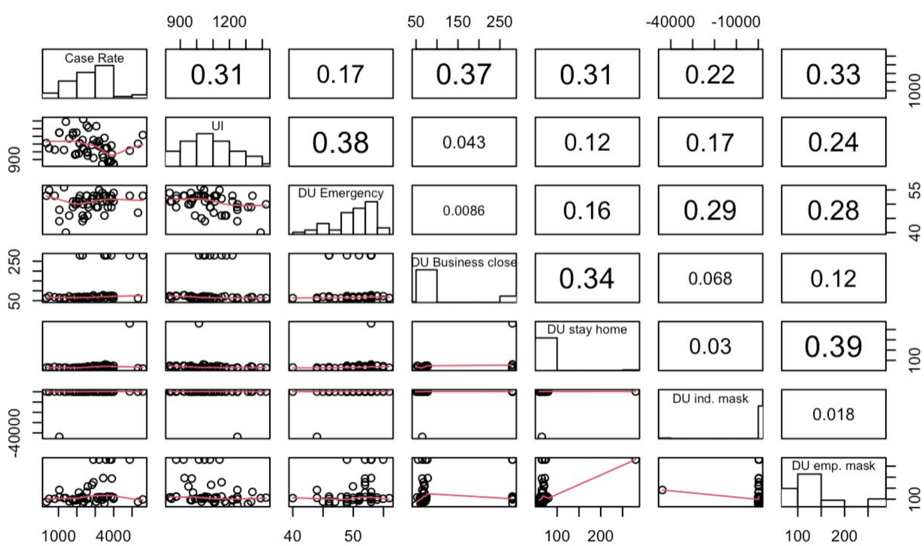


Figure 2.08 Days until emergency,stay-at-home, closed business, mask orders, Case Rate, and UI Panel

In our next exploration of policies, we took a look at how long it took until emergency, business closure, stay-at-home, individual mask, and employee mask orders were put into place on the state level. In the above panel, we see weak associations with the days until business closure, stay-at-home, and employee masks mandates, but negligible associations otherwise with case rate. For weekly UI, we see a weak association with the days until an emergency was declared which suggests that additional income relief could have been made available based on how soon a state reached an emergency declaration.

State Level Factors

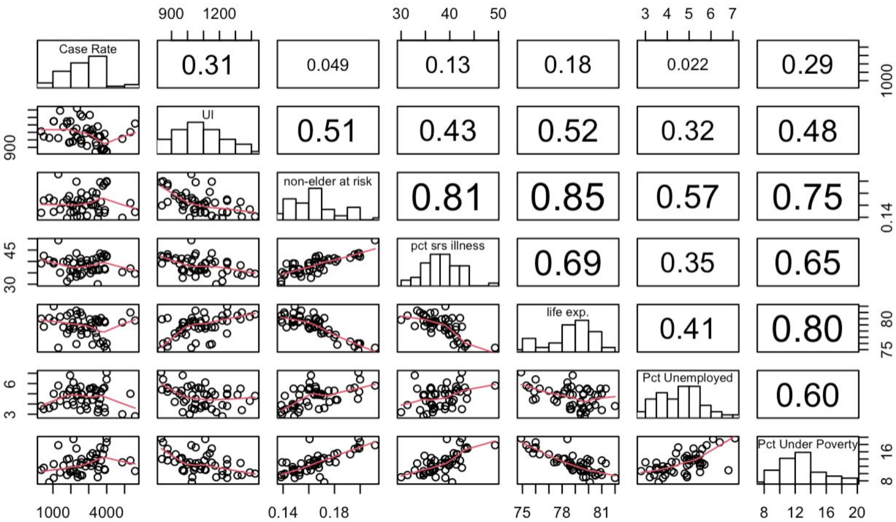


Figure 2.09 State factors, Case Rate, and UI Panel

For state level factors, we looked at the percentage of populations that were non-elderly at risk of Covid-19, at risk of a serious illness from Covid-19, life expectancy, unemployed, and under the poverty line. For case rate, there is just one weak association with those living under the poverty line. But for weekly UI, we observed moderate association with all state level factors, save a weak association with percent unemployed.

Mobility Factors

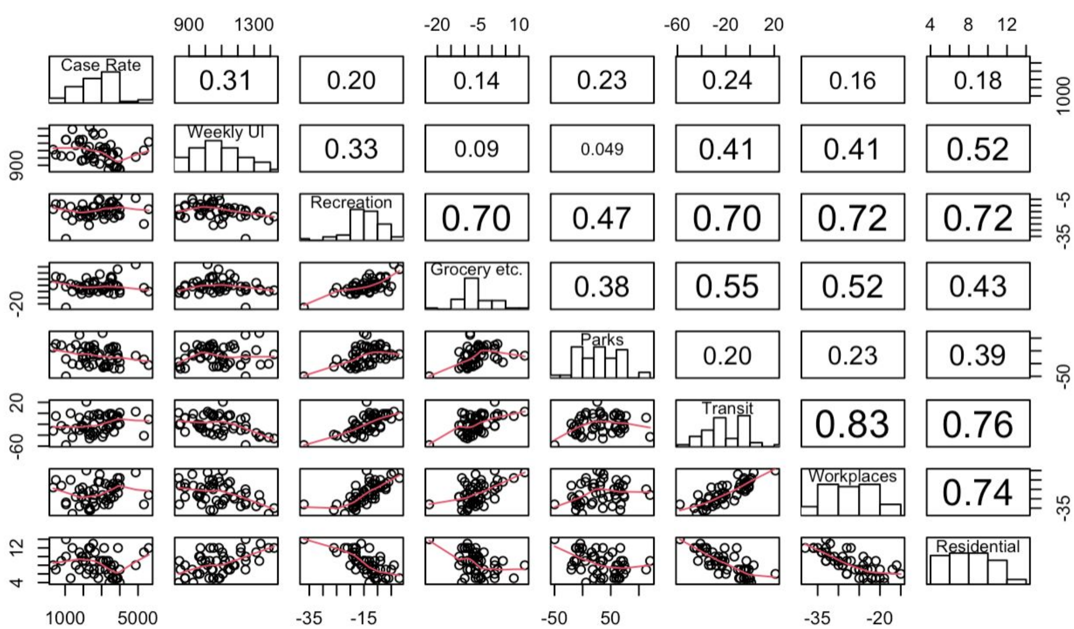


Figure 2.10 Mobility indices, Case Rate, and UI Panel

Mobility factors is an index of the level of traffic observed on a state level during the initial Covid-19 quarantine period in March. Like for state level factors, for mobility we also observe negligible associations with case rate, but moderate associations with weekly UI, namely with transit, workplace, and residential mobility. This makes sense, as weekly UI is used to deter people from seeking work, and thus it would reduce travel to workplaces and traffic at workplaces, and increase traffic in residential areas.

Additional Factors

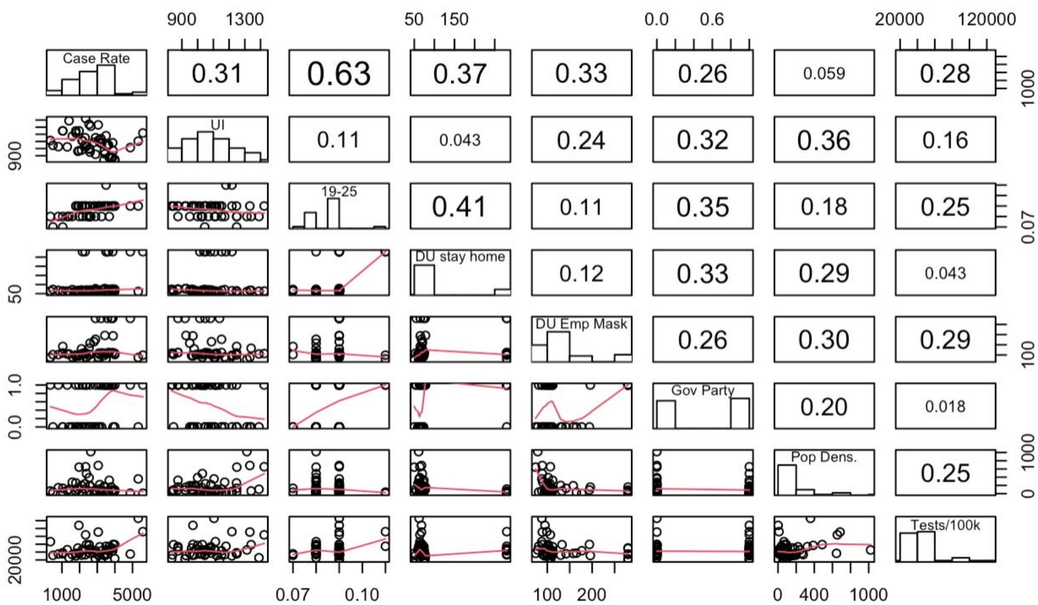


Figure 2.11 Key variables, Case Rate, and UI Panel

We lastly explore the most highly correlated variables with case rate and negligibly correlated to weekly UI, like age 19-25, days until stay-at-home orders, and days until employee mask mandate, with additional variables like the governor party, population density, and test kits available per 100,000 people.

We hypothesized that population density would be highly correlated with case rate but found it was not with a negligible correlation of ~0.06. We believe this is true because the data is on the state level and not on the city or county level. So it cannot truly capture the population density of states as they have a mix of metropolitan and rural areas. Furthermore, we hypothesize that with stay-at-home orders, population density may not have been contributing to Covid spread during that time. We also found a weak association between population density and Weekly UI, suggesting that states with higher population density needed to accommodate for more diverse socioeconomic populations.

Next, we looked at how policies and governor party affiliation related to case rate and explanatory variables of age 19-25, weekly UI, and the policies. For policies, we found a weak correlation between mask mandates for employees and case rates and a weak correlation between governor party and case rates. We found better but still weak correlations between the governor party and all other factors - age 19-25, weekly UI, and days until stay at home orders.

Finally, we explored if the level of testing on the state level was correlated with case rate, along with the other factors we found to have a weak or moderate correlation (weekly UI and age) and found that testing level (tests per 100k) had a weak correlation of 0.28 with case rate, but also weak correlation with adults age 19-25, suggesting that this age group may have a higher case rate due to being more tested than the other age groups.

Model 1: choosing to measure age

Case Rate = -5,664 + 96,997 Age 19-25

	Dependent variable:
	case_rate_per_100000
adults_19_25	96,997.020***
	(17,428.040)
Constant	−5,663.542***
	(1,518.141)
Observations	50
R^2	0.392
Adjusted R^2	0.380
Residual Std. Error	903.573 (df = 48)
F Statistic	30.976*** (df = 1; 48)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 2.12 Model 1: Baseline Model without Weekly UI

In our baseline model (as shown above in Figure 2.12), we decided to have our key variable (case rate per 100k) and 1 covariate for age. In conjunction with focusing on case rate, we chose to have one age covariate because published knowledge of Covid-19, age is a factor that is important for its infection rate. By adding 1 age covariate into our baseline model, we were able to control for age and test weekly UI and other possible covariates in models 2 and 3.

We decided to have the 1 age covariate be “percent of the state's population who are age 19-25” (abbreviated as age 19-25 henceforth) because we believe that this age group is most likely to behave differently from adults age 25+. Age 19-25 includes the population of college students and recent professionals living on their own (without their family). These adults are aware that the virus’ mortality is lower for them making them likely to congregate, be asymptomatic or follow guidelines more leniently which exposes them to Covid-19 more often. People of other ages (children 0-18 and adults 25+) are living with family and are more likely to behave according to how the whole family has decided to cope with Covid-19. Thus, we decided to operationalize age in our baseline model by using percent of adults age 19-25 in a state.

Our view is that age 19-25 is a key variable that should be in this base case model and that this is corroborated both quantitatively through our EDA and qualitatively. Firstly, the variable for age

19-25 has the highest Pearson's correlation value of 0.63 (moderate strength) amongst all of the age group variables. Secondly, this is corroborated from a qualitative perspective by Baruch Fischhoff, a professor at Carnegie Mellon University who studies human judgment and decision-making and stated that a 21 year-old is less likely than a 61-year old to have life experiences that impart the importance of preventing disease transfer.[1]

When we looked at all the correlations of age with case rate, we found 19-25 had a linear enough relationship with case rate, thus we did not need to look at transformations of the other age variables to increase the correlation.

Detailed explanation of model 1 (as shown in Figure 2.12) is in Section 4.

Model 2: choosing to measure age and UI stipend

Case Rate = -3,196 + 92,872 Age 19-25 – 1.953 Weekly UI

	Dependent variable:
	case_rate_per_100000
adults_19_25	92,872.060***
	(16,798.810)
weekly_ui	−1.953**
	(0.851)
Constant	−3,196.069*
	(1,809.375)
Observations	50
R^2	0.453
Adjusted R^2	0.430
Residual Std. Error	865.945 (df = 47)
F Statistic	19.494*** (df = 2; 47)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 2.13 Model 2: Model of Focus with Weekly UI

In model 2 (as shown above in Figure 2.13), we decided to test out the one core policy that our research question is focused on (the maximum amount an individual can get from unemployment insurance and additional stimulus per week). We feel that adding the weekly UI variable accurately represents the research question because it contains the maximum amount of money an eligible person can receive from their state's UI and additional stimulus benefits. Because of this, we decided to add only this one covariate into our model. By only choosing to add one covariate, we are exercising Occam's razor by focusing our resources and efforts on one covariate. Our exploratory data analysis also showed negligible correlation of 0.11 between the two variables age 19-25 and weekly UI and collinearity should be less of an issue but will be

1 https://www.theatlantic.com/family/archive/2020/03/coronavirus-social-distancing-socializing-bars-restaurants/608164/

tested in Section 3 of this report. We also decided not to transform weekly UI because in our EDA we found that weekly UI had a linear enough relationship with case rate. A transformation is not needed to improve its linear relationship with case rate. Not transforming weekly UI also allows for easier interpretability of the model which we felt was important.

Thus in model 2, we are able to evaluate if the weekly UI policy does explain the varied rates of infection across states while also controlling for proportion of population with Age 19-25. You will find a detailed discussion of regression summary results for Model 2 (as shown in Figure 2.13) in Section 4 of the report.

Separately, we also tested the robustness of our baseline model (model 1) with results summarized in Figure 2.14. This was to verify that our age 19-25 covariate still remains significant and is an effective covariate to hold control for as we test for weekly UI and also future covariates that we add in Model 3.

Test of Model 1 Robustness (Model 1 to Model 2)

X-Variable Name	Sign Stability	Significance of Coefficient	Robustness Ratio *(Estimated Beta/Standard Error of Estimated Beta)*
Adults Age 19-25	Remains positive	Relationship remains significant. P-value remains <0.05.(1.15e-6 to 1.38e-6)	Strongly robust since this is above 2 (5.57 to 5.53)
Overall Model	N/A	Model's ability to explain to describe associative relationship of variables with outcome remains significant with p-value <0.05.	N/A

Figure 2.14 Model 1 Robustness

Robustness of relationships of age 19-25 with case rate were evaluated based on the following 3 indicators:

1) **Sign Stability :** Signs for age 19-25 remained consistent between model 1 and model 2.

2) **Significance of Coefficient :** Significance of the associative relationship described in the model remains significant for age 19-25 with p value below 0.05 when moving from model 1 to model 2.

3) **Robustness Ratio:** Robustness persisted with a coarse guideline of strong robustness indicated by a ratio above 2 for both age 19-25 in both model 1 and model 2.

In addition, the overall model significance moving from Model 1 to Model 2 remains significant with a p-value below 0.05.

In conclusion, we find strong robustness in our baseline model with a significant positive associative relationship of age 19-25 in the population to case rates.

Model 3: choosing to measure policy and governor political party as possible other associate variables to case rate

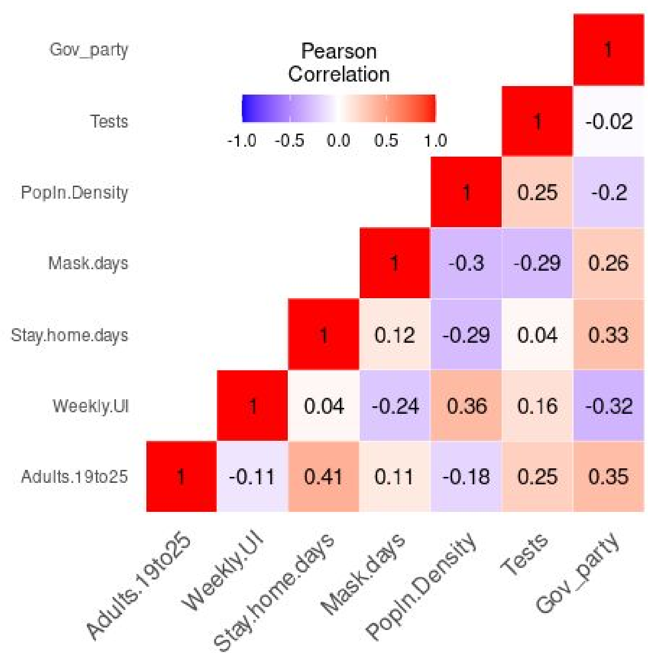
Case Rates = -2,538 + 74,330 Age 19-25 – 2.83 Weekly UI + 3.62 Days till Stay-At-Home Order +6.43 Days till Employ Mask Mandate – 308.60 If Governor is Republican + 1.37 Population Density + 0.01 Testing per 100K

	Dependent variable:
	case_rate_per_100000
adults_19_25	74,334.310***
	(17,168.180)
weekly_ui	−2.829***
	(0.848)
days_till_stay_at_home	3.618**
	(1.608)
days_till_emp_mask	6.425***
	(1.941)
gov_party	−308.645
	(248.033)
pop_density	1.366**
	(0.611)
tests_per_100k	0.013**
	(0.006)
Constant	−2,538.011
	(1,697.460)
Observations	50
R^2	0.636
Adjusted R^2	0.576
Residual Std. Error	747.225 (df = 42)
F Statistic	10.498*** (df = 7; 42)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 2.15 Model 3: Inclusive Model to Test Robustness of Weekly UI

In designing Model 3 (as shown above in Figure 2.15), we want to maximize the number of variables without introducing too much multicollinearity into the models while also bringing in unique information from the dataset that is not already covered by Model 2. For example, unemployment rate was not included because weekly UI combats unemployment and so unemployment rate does not introduce more unique information to the model. We do not include any transformations because in examining Figure 2.11 in our EDA we observe that line plots of the following variables versus case rates have linear relationships with the exception of Governor Party which is a binary variable and therefore does not make sense to transform. We demonstrate Model 2's robustness by evaluating the sensitivity of coefficients to changes in model specifications through Model 3. In Model 3, we add other control variables in the form of policy covariates such as Days till Mask Mandate is Employed, Governor Party, Days till Stay-At-Home Order was Imposed, Population Density, Testing per 100K which are weakly correlated with other independent variables in model 2 i.e. correlation is less than 0.5. Introducing variables with weak correlations is less likely to cause multicollinearity in the model. See Pearson's Correlation Heatmap for Model 3. The results of the assessment are summarized below in Figure 2.17**.**

Pearson's Correlation Heatmap for Model 3



Range	Strength of association
0	No association
0 to ±0.25	Negligible association
±0.25 to ±0.50	Weak association
±0.50 to ±0.75	Moderate association
±0.75 to ±1	Very strong association
±1	Perfect association

Figure 2.16 Model 3 Heatmap

Test of Model 2 Robustness (Model 2 to Model 3)

X-Variable Name	Sign Stability	Significance of Coefficient	Robustness Ratio *(Estimated Beta/Standard Error of Estimated Beta)*
Age 19-25	Remains positive	Relationship remains significant. P-value remains <0.05 (0 to 0).	Strongly robust since this is above 2 (5.53 to 4.33)
Weekly UI	Remains Negative	Relationship remains significant. P-value above 0.05 to below 0.05. (0.0263 to 0.00179)	Remains robust with absolute value above 2 (-2.29 to -3.34).
Overall Model	N/A	Model's ability to explain to describe associative relationship of variables with outcome remains significant with p-value <0.05.	N/A

Figure 2.17 Model 2 Robustness

Robustness of relationships of descriptive variable coefficients with the outcome variable Case Rates were evaluated based on the following 3 indicators:

1) **Sign Stability :** Signs for the 2 variables Age 19-25, Weekly UI remained consistent between model 2 and model 3.

2) **Significance of Coefficient :** Significance of the associative relationship described in the model remains significant for Age 19-25 and Weekly UI with p value below 0.05 when moving from model 2 to model 3.

3) **Robustness Ratio:** Robustness persisted with a coarse guideline of strong robustness indicated by a ratio above 2 for both Age 19-25 and Weekly UI in both model 2 and model 3.

Separately, the overall model significance moving from Model 2 to Model 3 remains significant with a p-value below 0.05.

Aside from model robustness, we also performed a model influence analysis to check for influence from outliers. Using the Cook's Distance test we find that outliers do not have a strong influence on model 2 or 3 results since none of the outliers have a Cook's distance higher than 1. See Figure 2.18 and Figure 2.19 respectively.

Cook's Distance for Model 2

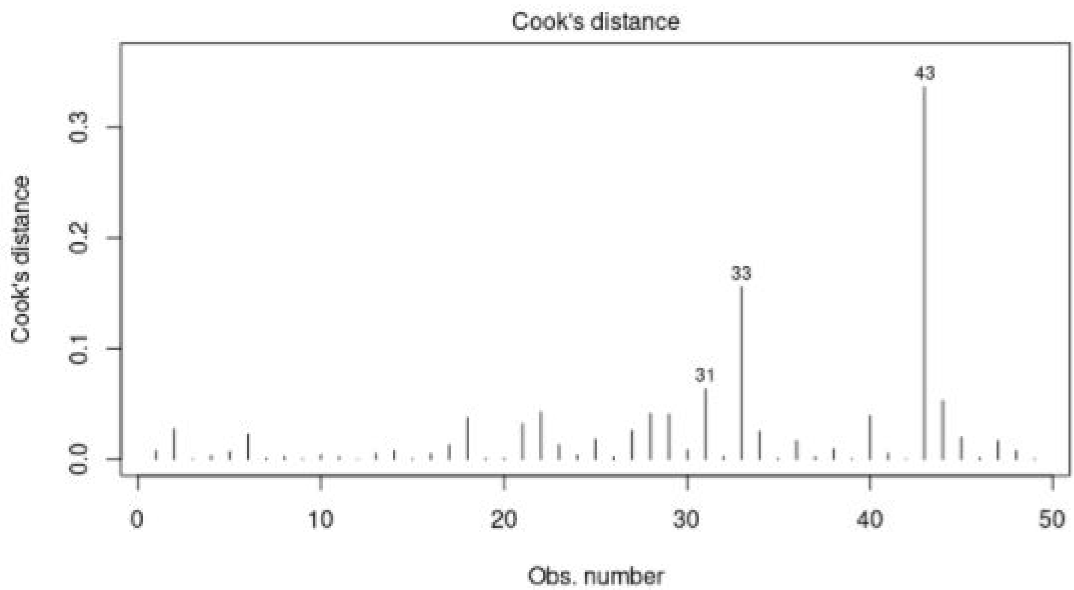


Figure 2.18 Cook's Distance Model

Cook's Distance for Model 3

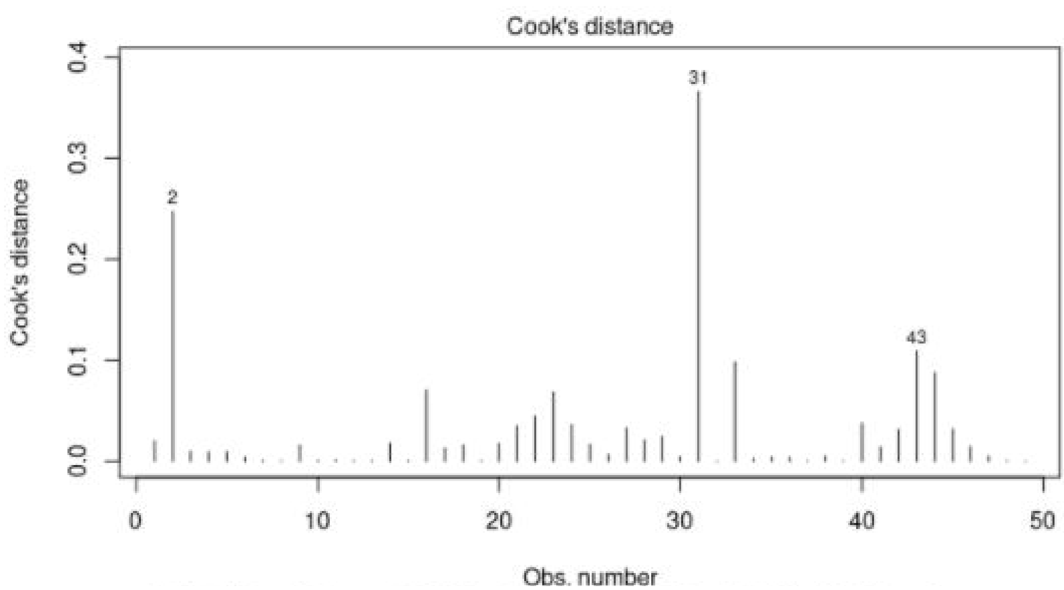


Figure 2.19 Cook's Distance Model 3

In conclusion, we find strong robustness in the significant positive associative relationship of proportion of adults Age 19-25 in the population to case rates and strong robustness in the negative associative relationship of Weekly UI to case rates.

The focus of the study is on Weekly UI and its associative effect on Case Rates. However, our group notes that there are other variables that have significant associative relationship with case rates which are laid out below in Figure 2.20.

Examining Other Variables not part of the focus of the Study

X-Variables Names	P- Value of Coefficient
Days till Stay-at-Home Order Imposed	Significant <0.05 at 0.0298.
Days till Mask Mandate Employed	Significant <0.05 at 0.00192.
If Governor Party is Republican	Not Significant >0.05 at 0.220.
Population Density per Square Mile	Significant <0.05 at 0.0306.
Tests per 100K	Significant <0.05 at 0.0275.

Figure 2.20 Other Variables

Lastly, we looked for the existence of multicollinearity in our model that could reduce the reliability of the regression coefficients. Multicollinearity is not an issue for Model 3 since the VIF is below 4 (Hair et al., 2010). The Model 3 VIF results in Screenshot 4 of the Appendix show that all X variables have a VIF between 1-2.

3. Limitations of Your Model

Models 1 to 3 will be assessed based on Classical Linear Model Assumptions because they do not meet the size requirement for Large Sample Linear Regression. We observed that all 5 assumptions were met across all three models specified. No remedial measures are required.

Please find the results of the tests and discussions of why the CLM assumptions were met in the appendix (pg 21-25).

4. Regression Table

	Dependent variable:		
	case_rate_per_100000		
	(1)	(2)	(3)
adults_19_25	96,997.020***	92,872.060***	74,334.310***
	(17,428.040)	(16,798.810)	(17,168.180)
weekly_ui		−1.953**	−2.829***
		(0.851)	(0.848)
days_till_stay_at_home			3.618**
			(1.608)
days_till_emp_mask			6.425***
			(1.941)
gov_party			−308.645
			(248.033)
pop_density			1.366**
			(0.611)
tests_per_100k			0.013**
			(0.006)
Constant	−5,663.542***	−3,196.069*	−2,538.011
	(1,518.141)	(1,809.375)	(1,697.460)
Observations	50	50	50
R^2	0.392	0.453	0.636
Adjusted R^2	0.380	0.430	0.576
Residual Std. Error	903.573 (df = 48)	865.945 (df = 47)	747.225 (df = 42)
F Statistic	30.976*** (df = 1; 48)	19.494*** (df = 2; 47)	10.498*** (df = 7; 42)
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure 4.01 Regression Table of all 3 Models

Coefficient Interpretation:

Using the regression table (as shown in Figure 4.01), we interpreted the coefficients for model 2. We chose to focus on model 2 because it contains the variables we are focused on understanding and have tested for robustness:

- **Intercept/Constant** -
  - **Interpretation:** Our intercept is -3196 which means that if a state had 0% of people aged 19-25 and there was no weekly UI or stimulus available (i.e. maximum amount is $0), then the expected value of case rate per 100,000

people would be of -3,196. While a negative case rate is not physically possible, we also know that no US state in our dataset had zero people with age 19-25 and no unemployment insurance benefit and that it is unlikely to ever happen. In addition, a minimum of 3.45% a state's population being age 19-25 would be enough to overcome the negative intercept and result in a positive case rate. A 3.45% is far below the minimum value of age 19-25 in our dataset (7% for both Maine and Hawaii) which supports our belief that the covariates would make the negative case rate unlikely to occur. Thus, we conclude the negative intercept is allowed in our model.

    - **Statistical Significance:** The coefficient of the intercept has a T-statistic of -1.766 which has a p-value of 0.0838. This means that our intercept does not meet our 95% significance level. A non-significant intercept means that when our covariates are zero then the model's outcome should be zero, not negative. This aligns with our above interpretation of the intercept and we keep the intercept in our model knowing that it is unlikely that age 19-25 and weekly UI would be 0 in reality.
    - **Practical Significance:** As indicated above in the results of the T-test, there is no practical significance of the intercept being -3,196 since we cannot have a negative case rate.

- **Coefficient of Percent of Population Age 19 to 25 (abbrv. Age 19-25):**
    - **Interpretation:** Because the variable percent of adults age 19-25 is expressed in decimal units (e.g. 5% is 0.05), then the interpretation of the coefficient is as follows: As a state had .01 percentage point more adults age 19-25 and all other characteristics were held the same, the expected value of its case rate per 100,000 people increased by 928.72.
    - **Standard Error of Coefficient:** From model 1 to model 2, the coefficient's standard error decreases from 17,428.04 to 16,798.81. This means that adding weekly UI into the model allowed it to capture some variance in the data that age 19-25 was trying to explain by itself in model 1. Resulting in the coefficient estimate for age 18 to 25 to be more precise.
    - **Statistical Significance:** The coefficient of age 19-25 has a T-statistic of 5.528 which has a p-value < 0.01. The p-value meets our 95% significance level and indicates that the relationship between age 19-25 and case rate we're seeing in our sample is highly probable to exist in the population. This means that we reject the null hypothesis that the coefficient is equal to zero and fail to reject that it is 92,872.
    - **Practical Significance:** Given that the median Covid-19 case rate among states is 2,726 per 100,000 people, the coefficient has practical significance with an

effect size that is 34.1% of the median case rate. This means that a 1 percentage point increase in percent of population who are age 19-25 can increase the number of Covid-19 cases in a state by about 34.1%. To confirm, we also used Cohen's D, which was 1.612, to calculate effect size. The effect size of age 19-25's coefficient is 0.628. To determine the strength of the effect size, we used the below table of Cohen's D threshold values (Cohen 1988).

Effect size measure	Small effect size	Medium effect size	Large effect size	Very large effect size
Odds ratio	1.5	2.5	4	10
Cohen's *d* (or one of its variants)	0.20	0.50	0.80	1.30
r	0.10	0.30	0.50	0.70
Cohen's *f*	0.10	0.25	0.40	—
Eta-squared	0.01	0.06	0.14	—

[a]Cohen, 1992, 1988; Rosenthal, 1996.

Table 4.2: Effect Size Thresholds for Cohen's D and R-Squared (Cohen 1988)

The effect size is larger than 0.5 which is our threshold for a medium effect. So we confirmed that the coefficient of age 19-25 has practical significance.

- **Coefficient of Maximum Weekly Amount of UI and Stimulus Benefits:**
  - **Interpretation:** As a state increased the maximum weekly amount of UI and stimulus a person could get by 1 US dollar, the expected value of its case rate per 100,000 people decreased by -1.953 when all other state characteristics were held the same.
  - **Standard Error of Coefficient Estimate:** The standard error of weekly UI is 0.851. This means that we are 95% confident that the value of the coefficient in the population is between -2.804 and -1.102. This standard error value is smaller than the coefficient estimate and causes the confidence interval to not contain 0. From this, we know that the coefficient estimate is statistically significant.
  - **Statistical Significance:** The coefficient of weekly UI has a T-statistic of -2.294 which has a p-value of 0.0263. The p-value meets our 95% significance level and indicates that the associative relationship between weekly UI and case rate we're seeing in our sample is highly probable to exist in the population. This means that we reject the null hypothesis that the coefficient is equal to zero and fail to reject that it is -1.953.

- **Practical Significance:** Given that the median Covid-19 case rate among states is 2,726 per 100,000 people, the coefficient is not practically significant due to its effect size of 0.07% of the median case rate. This means that a $1 increase in weekly UI can decrease the number of Covid-19 cases in a state by 0.07%. However, if the state government were to make changes to weekly UI, the changes would most likely be in larger amounts (e.g. $100) instead of smaller amounts (e.g. $1). For example, the recent CARES Act could increase weekly unemployment benefits by $600. Taking that perspective into account, the coefficient then becomes practically significant because if a state increases it's weekly UI by $100 (about 10% of median weekly UI) then case rate could decrease by 7%. Within the context Covid-19 spread, being able to reduce case rate by 7% is significant; thus, the coefficient is practically significant. To confirm, we used Cohen's D, which is -0.669, to find that the effect size of weekly UI's coefficient is 0.317. This effect size is larger than 0.2 and smaller than 0.5 which meets our threshold for a small effect size (see Table 4.2 for effect size thresholds). Similar to what was stated above, we believe that weekly UI amounts are unlikely to have low dollar amount changes, like $1, and instead will have larger changes, like $100. Thus, we accept a small effect size for single dollar amount changes in weekly UI to be practically significant.

Model 2's Statistical Significance:

From model 1 to model 2, our regression table shows that the degrees of freedom drops by 1 (from 48 to 47). This is because in both models we consistently used 50 observations and added weekly UI as a variable into model 2. The F-statistic from model 1 to model 2 decreased partially because the addition of weekly UI into our model decreased the degrees of freedom and shifted the F-statistic distribution to the right. However, the value of model 2's F-statistic in the shifted distribution remains high enough to be statistically significant at a 95% significance level. From this we can reject the null hypothesis that a model with no covariates is better and conclude that both models are useful as explaining some of the variance in case rate.

To understand if adding weekly UI improves the model's performance (e.g. is model 2 better than model 1), we ran a F-test between the two models (not shown on the regression table). Adding weekly UI into our model improved the regression sum of squares (RSS) by 3,945,815 (from 39,189,310 to 35,243,495). With the degrees of freedom and RSS of both models, we got a F-statistic of 4.918. The probability of seeing an F-statistic of this size given 49 observations and a change in degrees of freedom of 1 is about 1.16%. This probability (i.e. p-value) is below a 95% significance level which is why we concluded that adding weekly UI does improve the model's performance.

A couple measures we looked at to confirm what we found from the F-test was adjusted R-squared and residual standard error. The regression table shows that the adjusted R-squared increases from model 1 to model 2. Model 1's adjusted R-squared says that the model is able to explain 38.0% of the variance in case rate. Model 2's adjusted R-squared says that the model is able to explain 43.0% of the variance in case rate. The increase in explained variance is 5 percentage points which we considered high for adding 1 new covariate into the model. In addition, the adjusted r-squared value maintains a medium magnitude (see Table 4.2 for effect size thresholds). For residual standard error, the value decreases from 903.57 in model 1 to 865.95 in model 2 meaning that model 2's estimates of case rate fit the data more (with less residual values) than model 1. Thus, adjusted R-squared and residual standard error also indicate that adding weekly UI does improve the model's performance.

Model 2's Practical Significance:

We used Cohen's f^2 to determine model 2's effect size. To determine the strength of the effect size, we used the below table of Cohen's f^2 threshold values (Cohen 1988).

Cohen's f^2 Value	Effect Size
>= 0.02 and < 0.15	Small
>= 0.15 and < 0.35	Medium
>= 0.35	Large

Figure 4.02 Effect Size Thresholds for Cohen's f^2 (Cohen 1988)

The effect size is 0.828 which is greater than 0.35. From this we conclude that model 2 has a large effect size and is practically significant.

Model 3's Statistical Significance:

Next, we investigate whether model 3 improves upon model 2. In the regression model, we see that the degrees of freedom decreased from model 2 due to the 5 covariates that were added (from 47 to 42). This is part of what causes model 3's F-statistic to be lower than model 2's. However, model 3's F-statistic has a lower than 1% change of probability meaning that it meets a 95% significance level. This means that we can reject the null hypothesis that a model with no covariates is better than model 3 and conclude that model 3 is useful for explaining some of case rate's variance.

We ran a F-test between model 2 and model 3 to see if adding in the 5 new covariates improves the model's performance. The RSS improved by 11,792,984 and the F-statistic is 4.224 which meets a 95% significance level (p-value was 0.003). This means that the 5 new covariates (i.e. days till stay at home order, says till employee mask mandate, Governor's political party, state population density, and test rate per 100,000 people) does improve the model's performance.

We also looked at adjusted R-squared and residual standard error to confirm what the F-test found for model 3. The regression table shows that the adjusted R-squared increased from model 2 to model 3. Model 3's adjusted R-squared says that the model is able to explain 57.6% of the variance in case rate. This increased by 14.6 percentage points from model 2's adjusted R-squared which we considered a high increase for adding 5 variables into the model. According to Cohen's effect size thresholds (see Table 4.2), R-squared goes from a medium to large effect size in model 2 to model 3. For residual standard error, the value decreases from 865.95 in model 2 to 747.23 in model 3 meaning that model 3's estimates of case rate fit the data more (with less residual values) than model 2. Thus, we conclude that adjusted R-squared and residual standard error also show us that adding these 5 new covariates into model 3 improves its performance.

While our testing shows that model 3 is better than model 2 at describing the variance in case rate, we haven't tested model 3's robustness like we have for model 2. Without performing a robustness test of model 3's new covariates, we are not able to fully accept the model.

Model 3's Practical Significance:

We used Cohen's f^2 to determine model 3's effect size. The effect size is 1.747. From this we conclude that model 2 has a large effect size (see Table 4.3) and is practically significant.

5. Omitted Variables

The estimated coefficient for Weekly UI, the maximum amount an individual can get from UI and additional stimulus per week, is -1.953.

We have found 5 omitted variables of state level data that bias case rate and weekly UI, the coefficients that are most important to our study.

1. **Wealth bracket:** Given information on what the average wealth level is for each state, we would be able to better understand how wealth is associated with weekly UI and case rate.

   Relationship Triangle:

   Higher wealth levels would decrease the need for a weekly UI and those higher wealth levels would also mean individuals can afford to stay home and decrease opportunity to leave the home and contract or spread Covid-19. Alternatively, lower wealth levels increase the need for UI which would enable individuals to stay at home without employment and could lower opportunity for contracting or spreading Covid-19. See Figure 5.01.

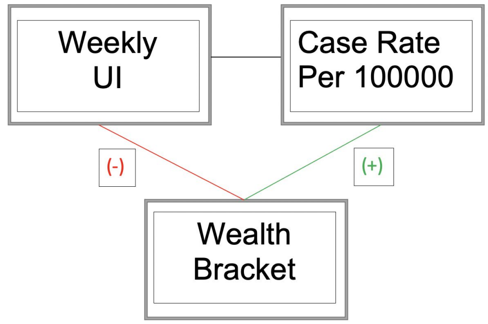


Figure 5.01 Omitted Variable - Wealth Bracket

Direction of bias:

We think the wealth bracket will have a negative correlation with weekly UI and a positive correlation with case rates. As such, we believe the direction of the omitted variable bias is away from zero.

Size of bias: Given that we lack this data, we are unable to estimate the size of the bias.

Available variable to proxy omitted variable: There are two variables provided to us that may proxy for the omitted variable wealth bracket. The two variables are Percent living under the federal poverty line (2018) and Median Annual Household Income. Both are imperfect proxies. By having the federal poverty line data for each state, we would not be able to account for wealthier states and their ability to afford staying home. Despite a federally identified poverty line, poverty thresholds could differ between states and therefore state poverty line information would be a better proxy for the wealth bracket variable. We could only account for a relationship between poverty line data and the need for UI. Although the Median Annual Household Income data is available, that is also an imperfect proxy. Due to Covid-19, annual household incomes were drastically changed for some families. As such, this data also could not perfectly proxy for wealth bracket data.

2. **Education level:** Given information on the average education level for each state, we would be able to better understand how education is associated with weekly UI and case rate.

Relationship Triangle:

Higher education levels would suggest higher pay jobs and therefore a decreased need for UI. Higher education levels would also result in a strong belief in Covid-19 risk factors and therefore the more educated would practice social distancing guidelines and contribute to lower infection case rates. See Figure 5.02.

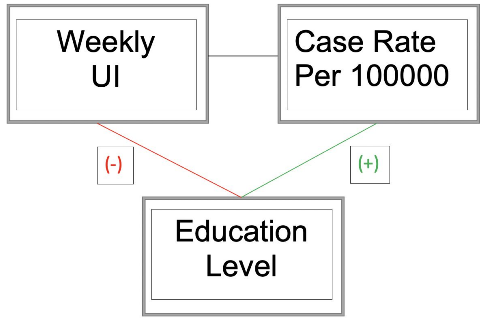


Figure 5.02 Omitted Variable - Education Level

Direction of bias:

We think education level will have a negative correlation with weekly UI and a positive correlation with case rates. As such, we believe the direction of the omitted variable bias is away from zero.

Size of bias: Given that we lack this data, we are unable to estimate the size of the bias.

Available variable to proxy omitted variable:

We do not have any variables available that may proxy for the omitted variable, education level.

3. **Number of dependents**: Given information on the average number of dependents in a household per state, we would be able to better understand how the number of dependents are associated with weekly UI and case rate.

Relationship Triangle:

A higher number of dependents would increase the need for and require a higher weekly UI. A higher number of dependents also increases the opportunity for Covid-19 infection and spread due to more individuals likely to be in one home, resulting in a higher Covid-19 case rate. See Figure 5.03.

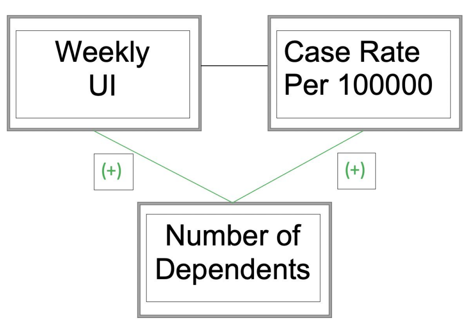


Figure 5.03 Omitted Variable - Dependents

Direction of bias:

We think the number of dependents will have a positive relationship with case rates and a positive relationship with weekly UI. As such, we believe the direction of the omitted variable bias is toward zero.

Size of bias:

Given the weekly UI coefficient is very close to zero, we estimate the size of the bias toward zero to be small.

Available variable to proxy omitted variable:

We do not have any variables available that may proxy for the omitted variable, number of dependents.

4. **Favorable Sentiment toward UI by Politicians in Media**:

The omitted variable is derived from understanding whether a state's politicians express favorable sentiment toward UI in the media or not.

Relationship Triangle:

If there is politician favorability toward weekly UI based stimulus schemes, they would feel positive toward state mask mandates and stay-at-home orders, thus lowering case rates. See Figure 5.04.

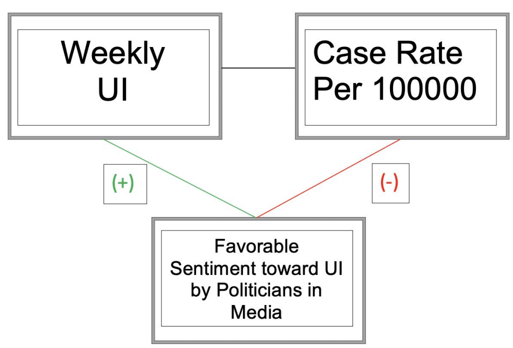


Figure 5.04 Omitted Variable - Politician Sentiments

Direction of bias:

We think favorable media sentiments will have a positive correlation with UI and a negative correlation with case rates. As such, we believe the direction of the omitted variable bias is away from zero.

Size of bias: Given that we lack this data, we are unable to estimate the size of the bias.

Available variable to proxy omitted variable:

We do not have any variables available that may proxy for the omitted variable, favorable sentiment toward UI by politicians in media.

5. **Proportion of agriculture industry in the state:**

   The omitted variable is derived from understanding what proportion of the state's industry is agricultural.

   Relationship Triangle:

   A higher proportion of the agricultural industry would result in an increased need for and a higher dependability on weekly UI because the agricultural industry was heavily impacted by the food supply chain being severely changed due to Covid-19. Due to regulations, restaurants, small business vendors, and large food chains being restricted from serving customers, they had to change their demand for raw foods from farmers. However, in having to deliver product and raw foods to grocery providers, the agricultural industry is essential, forcing workers in the agricultural industry to continue going to work which could result in higher infection case rates. See Figure 5.05.

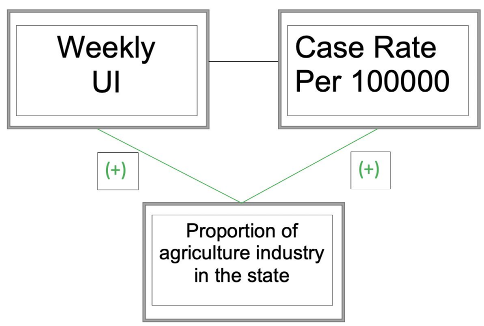


Figure 5.05 Omitted Variable - Agriculture Industry

   Direction of bias:

   We think the proportion of a state's agriculture industry will have a positive correlation to weekly UI and a positive correlation to case rate. As such, we believe the direction of the omitted variable bias is toward zero.

   Size of bias:

   Given the weekly UI coefficient is very close to zero, we estimate the size of the bias toward zero to be small.

   Available variable to proxy omitted variable:

   We do not have any variables available that may proxy for the omitted variable, proportion of agriculture industry in the state.

Are the key effects real or solely an artifact of omitted variable bias: Three of the five omitted variables will bias the coefficient away from zero. Although, we do not have information on the size of each bias, we estimate that the omitted variables would likely result in an overall bias that shifts the coefficient away from zero. This may result in an overestimation of the overall effect of the variable and potentially leading to a conclusion of false significance.

6. Conclusion

The purpose of this study was to explore if fiscally conservative or liberal approaches fare better for Covid-19 infection rate containment using a descriptive model. We conclude by analyzing our results as such:

Starting with Model 1, we chose to examine age in relation to Covid-19 case rates, based on our knowledge of how age is important to Covid-19 as a factor for infection rate. Next, in model 2, we decided to examine the core policy of Weekly UI- the maximum amount an individual can get from UI and additional stimulus per week that Republicans and Democrats have been debating. After performing our EDA, we found stronger correlation (-0.313) between case rate and UI than other policy variables which affirms our initial intuition behind this study. By examining maximum weekly UI and stimulus amount, we were able to evaluate if UI policies explain the variety in infection rate across states while controlling for proportion of population that is aged 19-25. By finding strong robustness in our baseline model with a significant positive associative relationship of proportion of adults age 19-25 in the population to case rates, we verified that our age 19-25 covariate still remains significant and is an effective covariate to hold control for maximum weekly UI and stimulus amount and also the covariates in Model 3. In Model 3, we found that the other policy related variables have significant associative relationship with case rate as well. Then, using Cohen's f^2, we were able to conclude that model 2 has a large effect size and is practically significant.

Based on our findings, we conclude that it is likely that a weekly UI offered to Americans at a time of individual economic turbulence is associated with a decrease in Covid-19 case rates. A weekly UI enables Americans to sustain their livelihood and relieves their pressure to seek employment outside the home which increases exposure to Covid-19 and increases case rates across the country. Additionally, according to Gary Burtless Brooking's report (see appendix reference link 1), in the present crisis, it is not only necessary to insure workers against the risk of earnings loss, it is also important to encourage unemployed workers to maintain a safe social distance from others to reduce the risk of disease transmission and slow the rate of infection.

We further recommend data to be collected on a more granular level (county or zip code) for weekly UI that would further improve the efficacy of the model. When that information is available, this kind of study can be concluded with greater statistical conviction on how weekly UI can effectively control the transmission of Covid-19.

Appendix

CLM Assumptions

Assumption 1 : I.I.D. Residual Errors in Data

The independence assumption is assumed to be met since this is point-in-time data on states, to prove it we ran the Durbin Watson test and showed that the p value is above 0.05 for all models and we fail to reject the null hypothesis that there is no auto-correlation and therefore infer that the errors are not correlated and the independence assumption is met. See Screenshot 1. The errors are identically distributed since they are all normally distributed, see assumption 5.

Screenshot 1

```
> #Testing the Independence (Autocorrelation) Assumption
> #Durbin Watson Test for Autocorrelation
> #The Durbin Watson examines whether the
> #errors are autocorrelated with themselves.
> #The null states that they are not autocorrelated (what we want).
> #this test is useful to verify that we havent violated the independence assumption.
> #p > 0.05, so the errors are not autocorrelated and the independence assumption is met for models 1-3
> durbinWatsonTest(model1)
 lag Autocorrelation D-W Statistic p-value
   1      0.07587106       1.83165   0.556
 Alternative hypothesis: rho != 0
> durbinWatsonTest(model2)
 lag Autocorrelation D-W Statistic p-value
   1      -0.1019172      2.198658    0.46
 Alternative hypothesis: rho != 0
> durbinWatsonTest(model3)
 lag Autocorrelation D-W Statistic p-value
   1      -0.2018058      2.374258    0.19
 Alternative hypothesis: rho != 0
```

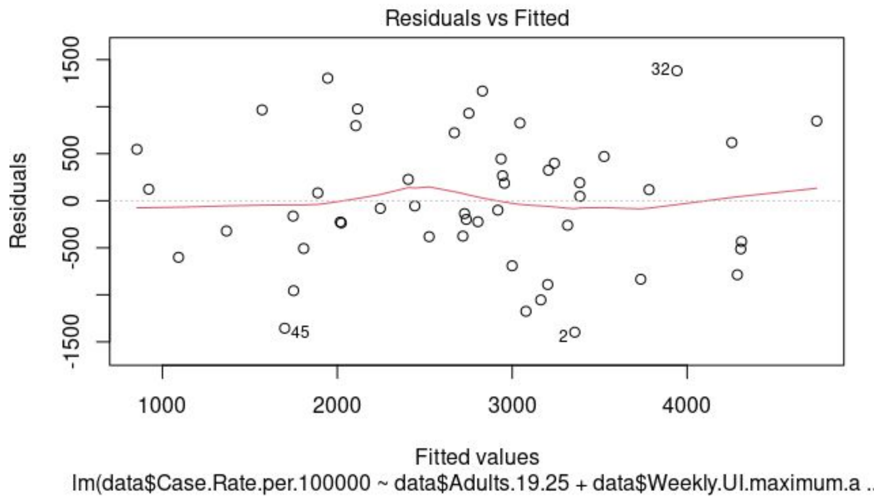
Assumption 2 : Linear Conditional Expectation

The linearity assumption is met via a plot of residuals and fitted y values, see Plot P1, P2 and P3. We observe linearity in model 1,2 and 3. Mean of the residuals is zero for all models, see Screenshot 2.

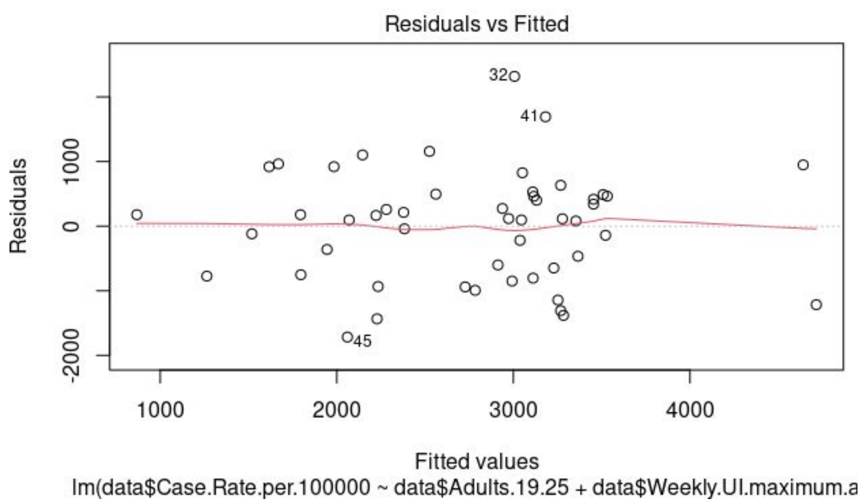
Screenshot 2

```
> #mean of residuals is zero, yes for all 3 models
> mean(model1$residuals)
[1] -9.499623e-15
> mean(model2$residuals)
[1] -1.024916e-14
> mean(model3$residuals)
[1] -3.426592e-14
```

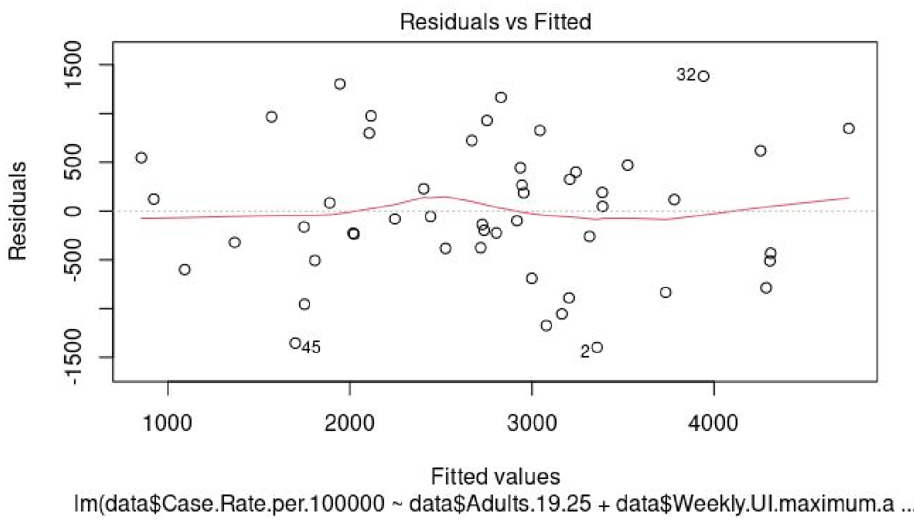
Plot P1: Residuals vs Fitted for Model 1



Plot P2: Residuals vs Fitted for Model 2



Plot P3: Residuals vs Fitted for Model 3



Assumption 3 : No Perfect Collinearity

The VIF test shows that all variables have a VIF factor below 2. A VIF factor below 4 means there is no problem with multi-collinearity (Hair et al., 2010). Therefore, we conclude there is no perfect collinearity. See Screenshots 3 to 4.

Screenshot 3

```
> vif(model2)
                                                                          data$Adults.19.25
                                                                                   1.011591
data$Weekly.UI.maximum.amount.with.extra.stimulus..through.July.31..2020...dollars.
                                                                                   1.011591
```

Screenshot 4

```
> vif(model2)
                                                          data$Adults.19.25
                                                                   1.011591
data$Weekly.UI.maximum.amount.with.extra.stimulus..through.July.31..2020...dollars.
                                                                   1.011591
> vif(model3)
                                                          data$Adults.19.25
                                                                   1.418975
data$Weekly.UI.maximum.amount.with.extra.stimulus..through.July.31..2020...dollars.
                                                                   1.349224
                                                data$days_till_stay_at_home
                                                                   1.425019
                                                    data$days_till_emp_mask
                                                                   1.236862
                                                              data$gov_party
                                                                   1.375086
                                      data$Population.density.per.square.miles
                                                                   1.419663
                                                          data$Tests.per.100K
                                                                   1.273218
>
```

Assumption 4 : Homoscedasticity of Errors

The Breusch-Pagan test shows p values above 0.05 for all 3 models, we fail to reject the null hypothesis that variance of residuals is constant and therefore infer that the residuals are homoscedastic across all models. Assumption 4 is met. See Screenshot 5.

Screenshot 5

```
> ncvTest(model1)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 0.5250953, Df = 1, p = 0.46868
> ncvTest(model2)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 0.2509857, Df = 1, p = 0.61638
> ncvTest(model3)
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 0.1605442, Df = 1, p = 0.68866
>
```

Assumption 5 : Normally Distributed Errors

The Shapiro Wilk’s test shows normality of residuals in all 3 models since all p values are above 0.05.See Screenshot 6.

Screenshot 6

```
> #residuals from model 1 are normally distributed since the p value is greater than 0.05 at 0.5017
> sresid <- studres(model1)
> shapiro.test(sresid)

        Shapiro-Wilk normality test

data:  sresid
W = 0.97878, p-value = 0.5017

> #residuals from model 2 are normally distributed since the p value is greater than 0.05 at 0.446
> sresid2 <- studres(model2)
> shapiro.test(sresid2)

        Shapiro-Wilk normality test

data:  sresid2
W = 0.97734, p-value = 0.446

> #residuals from model 3 are  normally distributed since the p value is greater than 0.05 at 0.9958
> sresid3 <- studres(model3)
> shapiro.test(sresid3)

        Shapiro-Wilk normality test

data:  sresid3
W = 0.99382, p-value = 0.9958
```

Reference Link 1:
https://www.brookings.edu/research/unemployment-insurance-as-social-protection-and-stimulus-during-the-coronavirus-crisis/

Reference Link 2:

https://www.theatlantic.com/family/archive/2020/03/coronavirus-social-distancing-socializing-bars-restaurants/608164/