Interim Write Up

Nicolas Min

3/21/2021

1. Problem Description

As a result of the COVID-19 crisis and the various non-pharmaceutical interventions (NPIs) to contain the spread of SARS-CoV-2, the U.S. saw an unprecedented spike in its claims for unemployment. And because various NPIs have been implemented at different levels in a decentralized manner, the U.S. saw significant geographic and temporal variations in its mobility and change in unemployment.

It now has been a year since everything suddenly froze in February of 2020. On a positive note, as of February 2021, the U.S. unemployment rate decreased to 6.2 percent from the record-high 14.8 percent in April 2020. In particular, according to "The Employment Situation - February 2021" released by the Bureau of Labor Statistics, employment in leisure and hospitality increased by 355,000 as "pandemic-related restrictions eased in some parts of the country". As things start to recover, people ask: "when will things get back to normal?" In the context of economics, the question becomes: "when will the U.S. recover the stable, low unemployment rate of 4% that we saw back in February 2020?" Our project intends to answer this question.

Unemployment is a major determinant of a strong aggregate demand and a healthy labor market. In this project, we will contrive a model that predicts U.S. unemployment in April/May 2021, a phase which we anticipate to see a significant decrease in COVID cases, thanks to the seasonality of the virus (assuming Holmstrom 2021 is correct) and thanks to the rapid vaccinations taking place throughout the nation. More precisely, our model will predict the *change in unemployment rate* in April/May 2021 for a given industrial sector for a given state. To do this, we will include the following variables in our model: mobility (where people are located), stringency (how strict NPIs are in a given state), sector composition (e.g., construction, financial services, leisure), and other control variables (such as the new COVID cases). Since we will be using seasonally adjusted data, we assume that our predictive model, composed of the specified variables, will generally hold true until the end of the pandemic.

2. Data Description

Below we specify the data we used for the variables in our model.

2-1. Mobility - Google Mobility data

Since the dawn of the pandemic, Google has been collecting global mobility data to inform public health officials' decisions to combat COVID-19. For each day at a given state, Google measures relative change (not the absolute visitors nor the duration) in mobility to the pre-COVID baseline level which is the median visitors to a given location from the 5-week period from Jan 3rd - Feb 6th, 2020. The entire data is relies on the Google users who have their Location History switched on. If these users visit categorized places (e.g., Parks - public garden, castle, camp ground; Transit stations - subway, taxi stand, car rental agency), Google aggregates the number of these users' location information to compute a mobility measure.

In our study, we will be focusing on the mobility data of the fifty U.S. states. As of 3/16/2021, the mobility dataset for the U.S. has 1,007,523 rows and contains retail, grocery, parks, transit, workplace, residential mobility data for each state at any given date between 2/15/2020 and 3/16/2021. (for further details, see: https://support.google.com/covid19-mobility/answer/9825414?hl=en&ref_topic=9822927).

One possible limitation of this data, as pointed out in a relevant study ("Unemployment Effects of Stay-at-Home Orders:Evidence from High Frequency Claims Data" - Baek et al. (2020)) is that the data is derived only from those with Google Accounts who opt into Location History services. However, we agree with Baek that the selection bias is unlikely to be a major concern"given Google's broad reach… (there are over 1.5 billion Gmail accounts)".

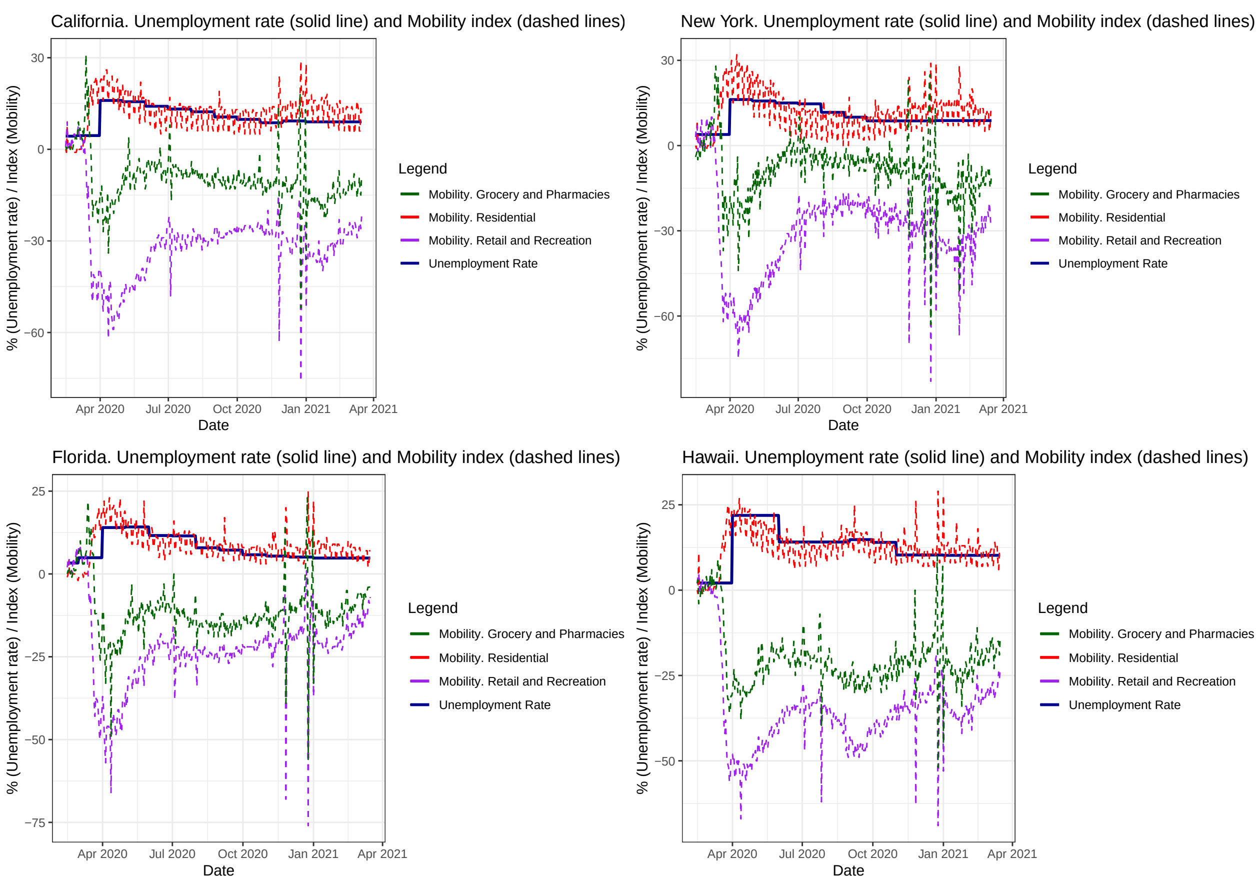
Assuming that the Google Mobility data is unbiased and accurate, we believe that the mobility information gathered from the data will serve as a good predictor of unemployment because of the direct and indirect relationship that mobility has with unemployment. Directly, less mobility would mean less demand (less transactions for over-the-counter goods, for instance), so workplaces would need less laborers. Indirectly, the shutdowns and the fear of the pandemic would lead to a drop in mobility and to an increase in unemployment.

Based on **Figure 1.**, we can observe that indeed mobility and unemployment have connections for many states. Namely, in most of the states, when the unemployment rate is high, the retail/recreational mobility is generally low (while residential mobility is high), but when the unemployment rate is low, the retail/recreational mobility is generally high (while residential mobility is low).

Finally, we noticed that the level of unemployment rose the most for Nevada, a state well known for its hospitality and leisure business, during the initial phase of the pandemic. This fact made us question whether a high dependency on person-to-person business makes a state's employment more vulnerable to the pandemic. We will delve deeper into this as we progress.

(Note: the unemployment data used here for the purpose of EDA is not the actual unemployment data we will use in our predictive model.)

Figure 1.



2-2. Stringency - Oxford Stringency Data

The Oxford COVID-19 Government Response Tracker (OxCGRT) systematically collects global information on several different common policy responses that governments have taken to respond to the pandemic. OxCGRT focuses on 18 indicators, including, but not limited to, school closures, travel restrictions, and stay-at-home orders. For the fifty U.S. states, OxCGRT focuses on 16 of these indicators. Of these 16, OxCGRT utilizes 9 ordinal indicators (school closing, workplace closing, cancel public events, restrictions on gathering size, close public transport, stay at home requirements, restrictions on internal movement, restrictions on international travel, and public information campaign) to come up with the *Stringency Index* for each state at a given date. In terms of the dimension of the data set, there are 23192 observations in the dataset as of 3/21/2021.

The data, as reported by OxCGRT, is "collected from publicly available sources such as news articles and government press releases and briefings" which are then "identified via internet searches by a team of over 50 Oxford" affiliates. In fact, for each indicator, there is an attached note that reasons why OxCGRT assigned the index it assigned for the given state at a given date. For instance, for 08/20/2020 Wyoming, OxCGRT assigned a '3' (highest measure) for the school closing index, quoting from a news article, "Wyoming schools will start under a 3 tier system under which schools can choose themselves which tier they will be in".

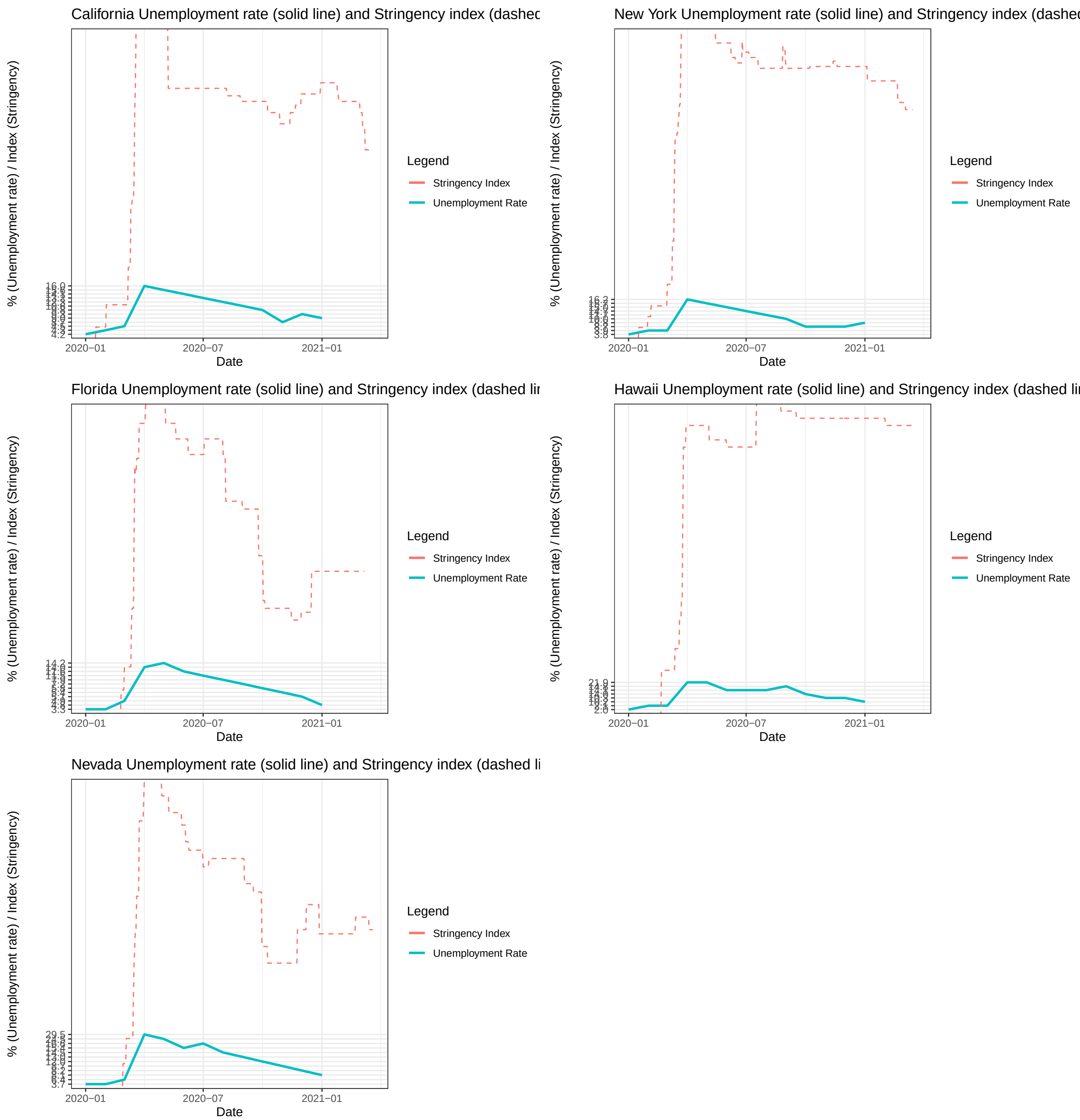
The one concern we have about this dataset is the underlying method of data collection. Investigating the attached notes, we found notes like: "While I can't find news sources to confirm… the Omaha Public Schools appear to be open in-person…". OxCGRT then assigned a '1' for Nebraska on 10/19/2020 for its school closing index. From these notes, we realized that there could be a chance that the qualitative judgment involved in the data collection might cause distortion in the stringency measure. However, given that over 50 people are involved in this judgment, we assumed there would be no systematic bias in the stringency measure. Also, we considered these judgments to be generally reasonable given that most of them (except for few) relied on the local government's official announcements.

Below, in **Figure 2.**, we have plotted two curves, unemployment and stringency, for each state. As expected, stringency and unemployment had significant relationships. Not only did we find that they follow a similar trend, we noticed that the peaks and troughs of stringency and unemployment match well for each state we observed. From this exercise, we could verify that Baek et al (2020)'s conclusion holds even for an extended time period beyond April 2020. Namely, we could verify that governmental orders to combat COVID have a significant impact on the unemployment rate throughout the era of COVID-19 pandemic.

Finally, the fact that the two curves matched particularly well for Nevada helped us confirm our thoughts from the previous section that there could be a high chance that mobility/stringency has a greater impact on the unemployment for those states with higher dependency on person-to-person oriented industries. Hence, in our next dataset, we look at the sectoral composition of each state.

(Note: the unemployment data used here for the purpose of EDA is not the actual unemployment data we will use in our predictive model.)

Figure 2.



2-3. Unemployment by state and sector - data from BLS data directory

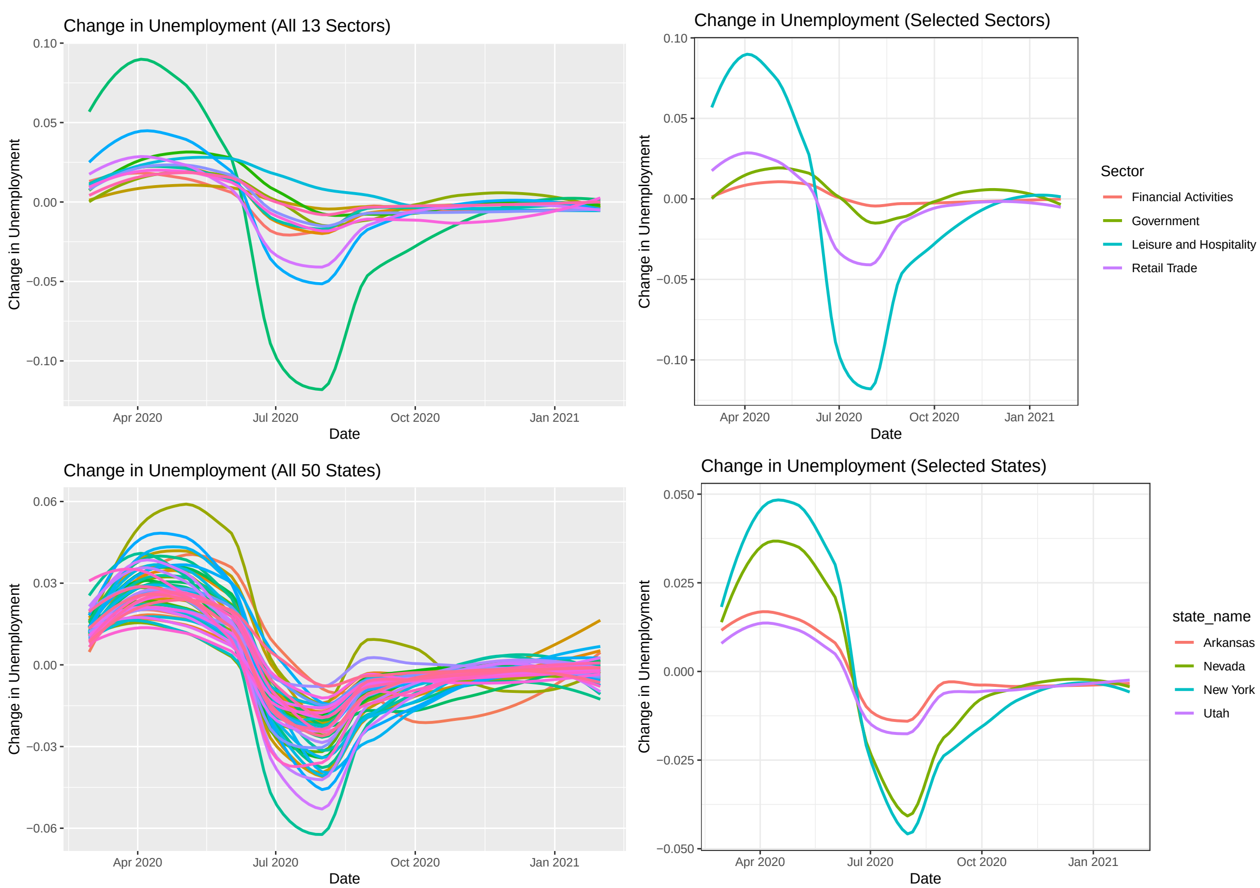
From the Bureau of Labor Statistics data directory, we imported employment data for each state for each industrial sector (13 sectors, including, but not limited to: Mining, Construction, Information, Financial Activities, and Leisure) at a given month-end date (it goes without saying that such data serve as a foundation for many statistic and economic analysis in various fields). As of 3/31/2021, the data includes 8,169,164 rows (inception year is 1939!).

The "series ID" column contains information about the sector and the state. Using this information, we re-structured our dataset such that each row corresponds to each state and each sector. In addition, we manipulated our data so that we have the *monthly change in employment* per state and per sector.

One issue with the current dataset is that it currently reports not the unemployment rate, but the number of employees in absolute numbers. Going forward, we would need to convert the number of employees into the unemployment rate, by dividing our absolute numbers by the number of people in the labor force at a given period of time. With further investigation, we plan to obtain the labor force numbers from BLS data directory. For the time being, however, the analysis in this report is made using the *number of employees* in a given state, for a given sector, and at a given month-end date. So, the *change in employment* in this report is computed assuming that the people in the labor force in a given state and for a given sector are held constant for the two months being compared.

The below plots (**Figure 3.**) observe the time trend in employment by sector, over the course of the pandemic. The plots are indeed congruent with our hypothesis that the sectors that depend heavily on person-to-person contact were more vulnerable to the mobility restrictions due to the pandemic. For instance, Government or Financial Activities sectors were less impacted by the pandemic. On the other hand, Retail Trade or Leisure and Hospitality sectors were significantly impacted by the pandemic and the relevant governmental NPIs. Also, we see that different states saw different levels of changes in unemployment due to the pandemic. In particular, the states with a greater portion of its people working in sectors that require person-to-person contact (e.g., Nevada or New York) saw a greater change in unemployment over the course of the pandemic compared to the states that do not (e.g., Arkansas or Utah).

Figure 3.



3. Methods

Based on our data analysis, we believe that the change in unemployment for a given state, time, and sector can be predicted using mobility, stringency, and sector composition in the following formulation:

$$\Delta Unemp_{s,t,sec} = \alpha + \mu_{sec} + \beta_1 Mob^2_{s,t} + \beta_2 String_{s,t} + \beta_3 Mob^2_{s,t} * Sector_{s,t} + \beta_4 String_{s,t} * Sector_{s,t} + X_{s,t} * \Gamma + \epsilon_{s,t,sec}$$

where

- $\Delta Unemp_{s,t,sec}$ is the change in unemployment for a given state (in the U.S.), at a given time (limited to the pandemic era; 02/01/2020-06/01/2021), for a given sector (one of 13)
- α is the global intercept and μ_{sec} is the local intercept that varies for each sector.
- $Mob_{s,t}$ is the seasonally adjusted residential mobility of a given state at a given time (so higher when people move less). We included a second degree polynomial term of mobility assuming that greater magnitude of mobility would have more impact on unemployment. (the R-squared without the second degree term was 4pp lower - see appendix)
- $String_{s,t}$ is the stringency level of a given state at a given time (higher index means more stringent)
- $Sector_{s,t}$ is the categorical variable.
- $X_{s,t}$ encapsulates vectors of control variables such as the number of new COVID-19 cases in a given state at a given time. Note that, in this report, control variables are not yet included.

Using data from 2/29/2020 to 1/31/2021, we will derive coefficients in our model using k-fold cross validation (k=10; the choice of k here is arbitrary). Then, using these coefficients, we plan to predict the 2021 April/May change in unemployment for a given state for a given sector. Finally, we will compute the RMSE between the predicted values and the actual values to assess the performance of our model.

In this project, we do not anticipate our model to necessarily produce the *lowest* RMSE. To minimize RMSE and let our model have the most predictive power, we should include as many features as possible that have any relevance to unemployment. However, taking this approach, the coefficient for the variables of our interest would lose most of their interpretability. Since we are interested in the interpretation of these coefficients as well as the predictive power of our model, we will instead take the approach of including only a handful of control variables in our model.

4. Results

In this report, instead of running the cross validation, we ran a linear model using the entire data set (using the formulation above) to first assess the significance of the coefficient of each variable of our interest.

Figure 4.

term	estimate	std.error	statistic	p.value
(Intercept)	-0.0052	0.0059390	-0.8780	0.3800
Mob_res_SA_sq	0.0002	0.0000240	9.7248	0.0000
SectorEducation and Health Services	-0.0061	0.0082743	-0.7411	0.4587
SectorFinancial Activities	-0.0008	0.0082743	-0.0913	0.9273
SectorGovernment	-0.0016	0.0082743	-0.1947	0.8457
SectorInformation	-0.0040	0.0083056	-0.4807	0.6307
SectorLeisure and Hospitality	-0.0459	0.0082743	-5.5452	0.0000
SectorManufacturing	-0.0052	0.0082743	-0.6250	0.5320
SectorMining and Logging	0.0055	0.0084540	0.6464	0.5181
SectorOther Services	-0.0175	0.0082743	-2.1193	0.0341
SectorProfessional and Business Services	-0.0067	0.0082743	-0.8114	0.4172
SectorRetail Trade	-0.0101	0.0082743	-1.2248	0.2207
SectorTransportation and Utilities	-0.0023	0.0082743	-0.2827	0.7774
SectorWholesale Trade	-0.0013	0.0082743	-0.1621	0.8712

term	estimate	std.error	statistic	p.value
stringency_index	-0.0004	0.0001390	-2.8735	0.0041
Mob_res_SA_sq:SectorEducation and Health Services	-0.0001	0.0000323	-2.8338	0.0046
Mob_res_SA_sq:SectorFinancial Activities	-0.0002	0.0000323	-5.9951	0.0000
Mob_res_SA_sq:SectorGovernment	-0.0002	0.0000323	-5.9308	0.0000
Mob_res_SA_sq:SectorInformation	-0.0001	0.0000324	-4.3902	0.0000
Mob_res_SA_sq:SectorLeisure and Hospitality	0.0005	0.0000323	15.6856	0.0000
Mob_res_SA_sq:SectorManufacturing	-0.0001	0.0000323	-2.4607	0.0139
Mob_res_SA_sq:SectorMining and Logging	-0.0002	0.0000340	-4.8397	0.0000
Mob_res_SA_sq:SectorOther Services	0.0001	0.0000323	3.4148	0.0006
Mob_res_SA_sq:SectorProfessional and Business Services	-0.0001	0.0000323	-3.4776	0.0005
Mob_res_SA_sq:SectorRetail Trade	0.0000	0.0000323	0.5036	0.6146
Mob_res_SA_sq:SectorTransportation and Utilities	-0.0001	0.0000323	-2.6315	0.0085
Mob_res_SA_sq:SectorWholesale Trade	-0.0001	0.0000323	-4.5234	0.0000
SectorEducation and Health Services:stringency_index	0.0004	0.0001915	1.9573	0.0503
SectorFinancial Activities:stringency_index	0.0005	0.0001915	2.4599	0.0139
SectorGovernment:stringency_index	0.0005	0.0001915	2.7544	0.0059
SectorInformation:stringency_index	0.0005	0.0001927	2.8098	0.0050
SectorLeisure and Hospitality:stringency_index	-0.0002	0.0001915	-0.8905	0.3732
SectorManufacturing:stringency_index	0.0003	0.0001915	1.7210	0.0853
SectorMining and Logging:stringency_index	0.0004	0.0001975	2.1450	0.0320
SectorOther Services:stringency_index	0.0002	0.0001915	0.9026	0.3668
SectorProfessional and Business Services:stringency_index	0.0004	0.0001915	2.1799	0.0293
SectorRetail Trade:stringency_index	0.0001	0.0001915	0.7656	0.4439
SectorTransportation and Utilities:stringency_index	0.0002	0.0001915	1.0739	0.2829
SectorWholesale Trade:stringency_index	0.0004	0.0001915	2.1630	0.0306

Based on the OLS Table, we noticed a couple of things (note that this is a preliminary analysis of our model; we expect the coefficients will be adjusted once we add more control variables in our model going forward):

- Based on the R-squared, our model could explain 23.9% of variation in the change in unemployment.
- The coefficient for the squared Mobility was statistically significant with a t value of 9.725.
- The coefficient for the stringency was statistically significant with a t value of -2.873.
- The local intercept for the Leisure and hospitality sector, which depends heavily on person-to-person contact, showed statistical significance with a t value of -5.545.
- Most of the coefficients for the interaction term, $Mob^2_{s,t} * Sector_{s,t}$ were statistically significant, while only a handful of the coefficients for the interaction term, $String_{s,t} * Sector_{s,t}$ were statistically significant. This result agrees with a separate linear regression we run in the appendix, which excludes mobility. Based on the low R-squared of 5.6% from that regression, we see that stringency itself is not enough to explain the variation of the unemployment data.

5. Conclusion

The project started from the idea that mobility, stringency, and sector composition (and some additional control variables) would sufficiently predict the change in unemployment rate of a given state, for a given sector, at a certain month-end date during the pandemic era.

In the **4.Results** section, we presented a summary of the OLS fit of *change in unemployment*($\Delta Unemp_{s,t,sec}$) on *squared mobility*, *stringency*, and *sector composition.* From the table, we identified the statistical significance of the coefficients for *squared mobility* and *stringency*, verifying that our hypothesis that mobility and stringency have a strong predictive power when it comes to predicting unemployment during the pandemic era.

Going forward, we plan to enhance our model by following the below steps.

1. Explore BLS data directory further to find the number of people in the labor force for a given state, time, and sector. Again, our current model assumes that the labor force is held constant over 2 month. Since 2 month is reasonably short, we do not consider our assumption to be preposterous. However, there is a chance that there are systematically more/less people completely giving up searching for jobs during the pandemic. Since this would induce bias in the coefficients in our model, finding labor force numbers and computing for unemployment rate could be necessary.
2. Add more control variables to our model. Some candidates include: new COVID cases, share of elderly people, 2016 Trump Vote share, and Bartik-style (method of isolating local labor demand changes) predicted job loss.

As discussed, after we enhance our model, we plan to:

1. Run a cross validation to contrive a RMSE-minimizing set of coefficients.
2. Using these coefficients, predict April/May 2021 change in unemployment data once the data for independent variables become available.
3. Assess our prediction by computing RMSE between the predicted and the real values.

Finally, it is worth noting that we will not be able to make a causal conclusion based on the coefficient we will obtain at the end of the day. This is because we face the endogeneity problem. Namely, the direction of the causal relationship could go both ways when it comes to the relationship between mobility and unemployment. To make a causal claim, we would need to limit our time frame and use addtional econometric tools so that we can treat mobility as a random assignment.

6. Appendices

6-1. OLS Table without the mobility terms

term	estimate	std.error	statistic	p.value
(Intercept)	-0.0178	0.0064458	-2.7655	0.0057
stringency_index	0.0004	0.0001258	3.0612	0.0022
SectorEducation and Health Services	-0.0013	0.0089923	-0.1433	0.8861
SectorFinancial Activities	0.0097	0.0089923	1.0772	0.2814
SectorGovernment	0.0087	0.0089923	0.9694	0.3324
SectorInformation	0.0037	0.0090261	0.4044	0.6859
SectorLeisure and Hospitality	-0.0738	0.0089923	-8.2106	0.0000
SectorManufacturing	-0.0010	0.0089923	-0.1099	0.9125
SectorMining and Logging	0.0144	0.0091764	1.5646	0.1177
SectorOther Services	-0.0238	0.0089923	-2.6419	0.0083
SectorProfessional and Business Services	-0.0007	0.0089923	-0.0812	0.9353
SectorRetail Trade	-0.0112	0.0089923	-1.2456	0.2130
SectorTransportation and Utilities	0.0021	0.0089923	0.2387	0.8114
SectorWholesale Trade	0.0065	0.0089923	0.7222	0.4702
stringency_index:SectorEducation and Health Services	0.0001	0.0001742	0.4626	0.6437
stringency_index:SectorFinancial Activities	-0.0002	0.0001742	-1.0134	0.3109
stringency_index:SectorGovernment	-0.0001	0.0001742	-0.6485	0.5167
stringency_index:SectorInformation	0.0001	0.0001754	0.4084	0.6830
stringency_index:SectorLeisure and Hospitality	0.0016	0.0001742	9.2150	0.0000
stringency_index:SectorManufacturing	0.0001	0.0001742	0.4423	0.6583
stringency_index:SectorMining and Logging	-0.0001	0.0001791	-0.7266	0.4675
stringency_index:SectorOther Services	0.0006	0.0001742	3.3126	0.0009

term	estimate	std.error	statistic	p.value
stringency_index:SectorProfessional and Business Services	0.0001	0.0001742	0.2942	0.7686
stringency_index:SectorRetail Trade	0.0002	0.0001742	1.2943	0.1956
stringency_index:SectorTransportation and Utilities	-0.0001	0.0001742	-0.3784	0.7051
stringency_index:SectorWholesale Trade	-0.0001	0.0001742	-0.3954	0.6926

6-2. OLS Table without taking mobility squared

term	estimate	std.error	statistic	p.value
(Intercept)	-0.0202	0.0059540	-3.3975	0.0007
Mob_res_SA	0.0044	0.0005981	7.3725	0.0000
SectorEducation and Health Services	-0.0008	0.0083070	-0.0931	0.9259
SectorFinancial Activities	0.0116	0.0083070	1.3929	0.1637
SectorGovernment	0.0105	0.0083070	1.2632	0.2066
SectorInformation	0.0049	0.0083379	0.5845	0.5589
SectorLeisure and Hospitality	-0.0816	0.0083070	-9.8184	0.0000
SectorManufacturing	-0.0005	0.0083070	-0.0630	0.9498
SectorMining and Logging	0.0162	0.0084757	1.9159	0.0554
SectorOther Services	-0.0259	0.0083070	-3.1182	0.0018
SectorProfessional and Business Services	0.0001	0.0083070	0.0072	0.9943
SectorRetail Trade	-0.0120	0.0083070	-1.4429	0.1491
SectorTransportation and Utilities	0.0028	0.0083070	0.3342	0.7382
SectorWholesale Trade	0.0078	0.0083070	0.9440	0.3452
stringency_index	-0.0004	0.0001572	-2.5232	0.0116
Mob_res_SA:SectorEducation and Health Services	-0.0013	0.0008195	-1.5342	0.1250
Mob_res_SA:SectorFinancial Activities	-0.0035	0.0008195	-4.3238	0.0000
Mob_res_SA:SectorGovernment	-0.0034	0.0008195	-4.1033	0.0000
Mob_res_SA:SectorInformation	-0.0024	0.0008230	-2.9208	0.0035
Mob_res_SA:SectorLeisure and Hospitality	0.0125	0.0008195	15.2675	0.0000
Mob_res_SA:SectorManufacturing	-0.0012	0.0008195	-1.4312	0.1524
Mob_res_SA:SectorMining and Logging	-0.0034	0.0008480	-4.0223	0.0001
Mob_res_SA:SectorOther Services	0.0032	0.0008195	3.8893	0.0001
Mob_res_SA:SectorProfessional and Business Services	-0.0017	0.0008195	-2.0934	0.0363
Mob_res_SA:SectorRetail Trade	0.0009	0.0008195	1.1162	0.2644
Mob_res_SA:SectorTransportation and Utilities	-0.0014	0.0008195	-1.7676	0.0772
Mob_res_SA:SectorWholesale Trade	-0.0026	0.0008195	-3.2312	0.0012
SectorEducation and Health Services:stringency_index	0.0003	0.0002171	1.3730	0.1698
SectorFinancial Activities:stringency_index	0.0005	0.0002171	2.0739	0.0381
SectorGovernment:stringency_index	0.0005	0.0002171	2.2178	0.0266
SectorInformation:stringency_index	0.0005	0.0002185	2.2570	0.0240
SectorLeisure and Hospitality:stringency_index	-0.0006	0.0002171	-2.9586	0.0031
SectorManufacturing:stringency_index	0.0003	0.0002171	1.2871	0.1981
SectorMining and Logging:stringency_index	0.0005	0.0002234	2.1243	0.0337
SectorOther Services:stringency_index	0.0000	0.0002171	-0.0053	0.9958
SectorProfessional and Business Services:stringency_index	0.0004	0.0002171	1.6158	0.1062
SectorRetail Trade:stringency_index	0.0001	0.0002171	0.2491	0.8033
SectorTransportation and Utilities:stringency_index	0.0002	0.0002171	0.8558	0.3921
SectorWholesale Trade:stringency_index	0.0004	0.0002171	1.8314	0.0671

7. Citations

- https://www.bls.gov/news.release/pdf/empsit.pdf
- https://bcf.princeton.edu/events/bengt-holmstrom-the-seasonality-of-covid-19/
- https://irle.berkeley.edu/files/2020/07/Unemployment-Effects-of-Stay-at-Home-Orders.pdf