Duke Datathon 2020

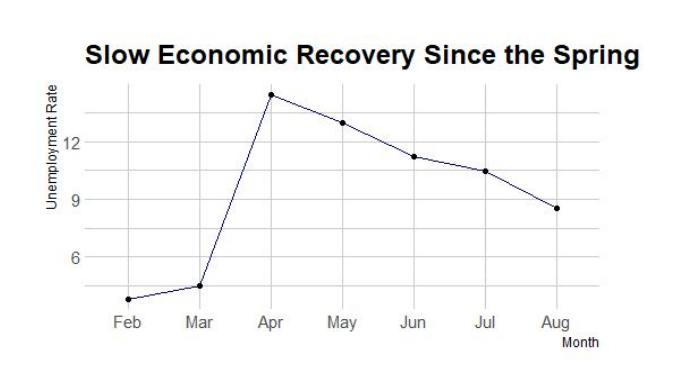
'I like to move it, move it': The Effects of post-COVID Mobility on Unemployment

Team Bae-sian Statistics

Duke

Abstract

The COVID-19 pandemic and the resulting social distancing measures have drastically changed our daily lives, especially our economic well-being. We investigated movement trends and various socioeconomic factors to predict differences in unemployment at the county level across the United States during the pandemic. Through backwards selection analysis of variables from movement trends, unemployment, education, and poverty data, we found average movement for retail, grocery, and workplaces to be the strongest predictors for unemployment rates. Furthermore, we found that a significant interaction effect existed between average retail movement and average grocery movement. By monitoring movement rates and consumer spending in these areas, we could identify and predict areas of high unemployment.



Introduction

Our lives have changed drastically since the onset of the COVID-19 pandemic, and social distancing measures have resulted in shifts in working and spending habits. This has had a profound economic impact, changing job dynamics across the world (Gharehgozli 2020). The effects of the pandemic on jobs are felt unequally across different regions of the United States. Our team sought to investigate which socioeconomic factors could serve as predictors of unemployment during the COVID-19 pandemic.

Objectives

A. **Understanding the relationship between movement and county-level unemployment during the COVID-19 pandemic:** How does movement in retail, workplace, and grocery spaces relate to unemployment?
B. **Predict future shifts in unemployment based on previous data:** Using conclusions from part A, how do we expect unemployment to change in the future, given current trends in movement/mobility? Which developments have remained stable throughout the course of the pandemic and are likely to continue? Which counties will see the most significant increases or decreases in unemployment?

Methodology

- Our team combined Google's COVID-19 Community Mobility Reports with three demographic datasets from the Economic Research Service that include data on unemployment and median household income, educational attainment, and poverty estimates. We selected the most recent year's data from each dataset (either from 2018 or 2019), and we analyzed the data at the county level.
- We **randomly sampled** 500 counties from all U.S. counties to fit our model to avoid spatial correlation (counties closer to each other will have similar/correlated unemployment and movement rates).
- **Response:** We created the variable unemployment_change by taking the difference of unemployment rates in August 2020 and Feb 2020 (Aug - Feb).
- **Predictors:** We selected 7 variables we believed to be likely predictors of unemployment_change: retail_avg, grocery_avg , workplaces_avg, residential_avg , medhhinc_2018, pct_bachelor_or_higher, and pctpovall_2018.

Variable Selection

- Through **backwards selection**, we found retail_avg, grocery_avg, and workplaces_avg to be significant predictors ($p < 0.05$). All else held constant, for each 1% increase in retail_avg, grocery_avg, or workplaces_avg, the unemployment rate is predicted to decrease by 0.073%, 0.015%, and 0.035%, respectively.
  - Interestingly, all other demographic data were not found to be significant, whereas variables created from the Google Mobility Data were kept as **significant predictors** of unemployment change
- Through a **nested F-test**, we found the interaction term between retail_avg and grocery_avg to be significant, meaning that there is a possible relationship between retail movement on change in unemployment on grocery movement, and vice versa.

$$\widehat{\text{unemployment\_change}} = 1.266 - 0.73\ \text{retail\_avg} - 0.015\ \text{grocery\_avg} - 0.035\ \text{workplaces\_avg} + 0.002\ \text{retail\_avg:grocery\_avg}$$

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	1.266	0.515	2.458	0.014	0.254	2.278
retail_avg	-0.073	0.013	-5.680	0.000	-0.099	-0.048
grocery_avg	-0.015	0.014	-1.068	0.286	-0.044	0.013
workplaces_avg	-0.035	0.023	-1.537	0.125	-0.080	0.010
retail_avg:grocery_avg	0.002	0.000	3.770	0.000	0.001	0.002

Duke

Predictions

If all else continues, our model predicts these counties to have the largest unemployment rate *increases* over the next 6 months:

1. New York County, NY (+8.7%),
2. Arlington County, VA (+6.9%),
3. Santa Clara County, CA (+6.8%),
4. Kauai County, HI (+6.5%),
5. Maui County, HI (+6.5%).

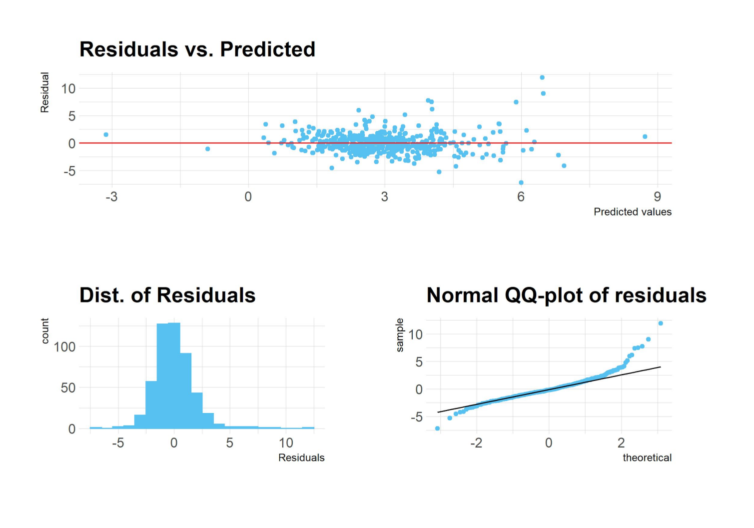
If all else continues, our model predicts these counties to have the largest unemployment rate *decreases* over the next 6 months:

1. Cape May County, NJ (-3.3%),
2. Dare County, NC (-3.1%),
3. Door County, WI (-2.5%),
4. Le Sueur County, MN (-0.9%),
5. Taney County, MO (-0.9%).

Conditions

The four conditions for linear regression are satisfied:

1. **Linearity:** Satisfied - there is no distinguishable pattern in the residuals. They are randomly scattered.
2. **Normality:** Satisfied - the points fall along a straight diagonal line on the normal quantile plot.
3. **Constant Variance:** Satisfied - the vertical spread of the residuals remains relatively constant across the plot.
4. **Independence:** Satisfied - the error for one county does not tell us anything about the error for another county; we randomly sampled 500 counties in our dataset, avoiding spatial correlation.



Duke

Future Economic Impact during COVID-19 Pandemic

(1) Model expects places with higher movement in retail, grocery, and other workplaces to have lower immediate unemployment.
(2) If trends continue, areas with lower consumer spending in grocery and retail, our model expects sustained high unemployment.

Limits/Next Steps

(1) This model is unable to assess the effect of increased COVID-19 numbers on *long-term* unemployment and economic downturn, because it focuses on unemployment and short-term movement data. In other words, it cannot forecast the effect of increased movement on increased COVID-19 numbers, and the economic effect that that has on an individual county in the long-run cannot be determined by our model.
(2) We are unable to discern short-term (1-2 years) versus longer-term economic effects (5 years).
(3) Counties that include multiple large cities with more diverse economies are less likely to have accurate predictions due to generalization when grouping them all together into the same county.
(4) We do not account for metrics of economic conditions outside of unemployment and economic movement (i.e. stocks, e-commerce, etc.).
(5) This model is correlative, not causal.

Duke

References

- GitHub Repo: https://github.com/ben-j-wallace/Datathon2020
- Google COVID-19 Community Mobility Reports: https://www.google.com/covid19/mobility/
- County-Level Socioeconomic Data Sets (Economic Research Service): https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/
- Bureau of Labor Statistics: https://www.bls.gov/lau/
- Impact of COVID-19 on the Economic Output of the US Outbreak's Epicenter: https://link.springer.com/article/10.1007/s41885-020-00069-w

Appendix

- **unemployment_change**: unemployed percentage in Aug 2020 minus unemployed percentage in Feb 2020
- **retail_avg**: average daily retail movement in county from Feb 2020 to Aug 2020
- **grocery_avg**: average daily grocery movement in county from Feb 2020 to Aug 2020
- **workplaces_avg**: average daily workplace movement in county from Feb 2020 to Aug 2020
- **residential_avg**: average daily residential movement in county from Feb 2020 to Aug 2020
- **medhhinc_2018**: median household income in 2018
- **pct_bachelor_or_higher**: percent with a bachelor's degree or higher
- **pctpovall_2018**: percent of people in poverty in 2018

Duke