

UNEMPLOYMENT TRENDS AMIDST COVID-19 AND ITS GOVERNANCE

LauraMargaret Burbach (burbach), Ali Pelczar (apelczar), and Jonathan Rockower (jrockower)

Project Summary

This project aims to model county-level unemployment in the United States using population demographics, political information, COVID-19 case trends, and government management in response to the COVID-19 pandemic. As of May 1, 2020, the Centers for Disease Control and Prevention (CDC) reported 1,062,466 total cases in the United States, 5.9% of which have proven fatal.¹ Given the virus's communicability via respiratory droplets, its fatality, and its impact on health care resources, states, counties, and cities have enacted various restrictions on businesses, services, and personal behavior in an effort to reduce its spread.

Both the disease and these restrictions have led to increased unemployment in the United States. The nationwide unemployment rate rose from 3.5% in February 2020 to 4.4% in March 2020, and it is expected to continue rising, with a record 30 million Americans filing for unemployment in the last six weeks (mid-March to the end of April).² The economic impact has been unsettling to many citizens; several states and cities have seen protests as the stay-at-home orders have closed stores, restaurants, and schools. As of May 1, more than half of states had plans to partially reopen their economies over the coming days and weeks, while federal social distancing guidelines expired April 30.³

The question on the nation's mind is whether reopening states will curtail the steep drop in economic activity, though many remain worried for their health and are therefore reluctant to return to their everyday routines as before. While this question will not be answered with certainty for several months, for now we can try to understand the true impact of COVID-19 cases and state at home orders on unemployment. With so much attention focused on these state- and national-level trends, we will look instead at the county level.

Our work aims to determine the extent to which differences in government reaction and case outbreaks can predict trends in unemployment at the county level.

As pandemic and restriction management continues, state- and county-level policymakers could use our results to inform their restrictions based on the results seen with certain restrictions by demographically similar areas. Looking at our results retrospectively, infectious disease specialists and the general public would be able to use our results to evaluate the effectiveness of various policies and leadership during the time of COVID-19.

¹ [COVID-19 Cases in the United States](#), Centers of Disease Control and Prevention

² [30 Million Have Sought US Unemployment Aid Since Virus Hit](#), New York Times

³ [Most US states are reopening within days but some need time 'to remain vigilant'](#), CNN

Our regression will include regressors that directly relate to the management of COVID-19, such as the type of stay-at-home order and the date when it was enacted, as well as control variables for more static factors like population density, socioeconomic makeup, and political trends.

Based on the evaluation metrics of our regression model, high coefficients for the management-related variables will yield findings that are more interesting to our intended audiences. These results will not be as relevant if our findings show that changes in unemployment are not responsive to variations in COVID-19 management practices, though this would suggest a different story for how local policy responses to COVID-19 impacted the economy.

Data

Our analysis will use county-level data compiled from several sources. We will use the dataset maintained by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University on case counts and testing rates.⁴ Unemployment information will come from the Bureau of Labor Statistics,⁵ while a variety of county-level descriptor variables (demographics by race, population density, access to internet) will come from the American Community Survey⁶ and the U.S. Census Bureau.⁷ Data on the dates that stay at home orders were enacted will be drawn from the New York Times⁸ and a dataset maintained by a team of policy researchers at Boston University.⁹ The U.S. Department of Agriculture provides datasets on educational attainment and poverty by county.¹⁰ We will also include the political affiliation of the state governor, using a dataset provided by the Kaiser Family Foundation¹¹, and the political leaning of the county as measured by the split between Republican and Democratic votes in the 2016 presidential election.¹²

We plan to perform a regression on this data, as discussed in the following section, *Machine Learning*. While the data we will use in our analysis will be detailed at the county level, the overall trends national trends that inform our research questions can be illustrated with the national-level data described here:

⁴ [Novel Coronavirus \(COVID-19\) Cases](#), Johns Hopkins University

⁵ [Local Area Unemployment Statistics County Data](#), Bureau of Labor Statistics

⁶ [American Community Survey Data](#), United States Census

⁷ [County Population Totals](#), United States Census

⁸ [See Which States and Cities Have Told Residents to Stay at Home](#), New York Times

⁹ [COVID-19 US State Policy Database](#), Boston University

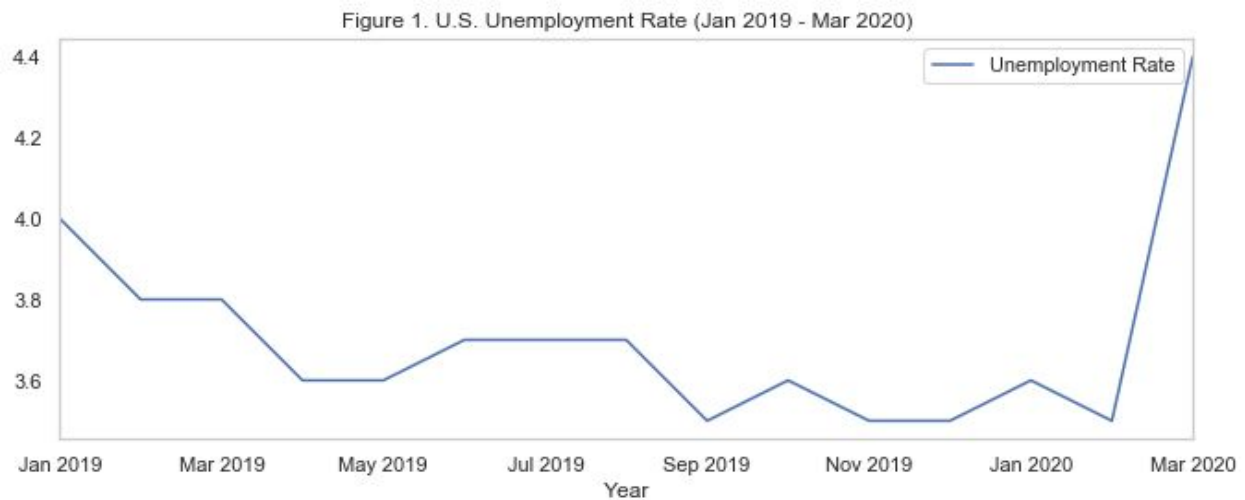
¹⁰ [Economic Research Service](#), United States Department of Agriculture

¹¹ [State Political Parties](#), Kaiser Family Foundation

¹² [County-Level Election Results](#), compiled from The Guardian and townhall.com

Unemployment Rate

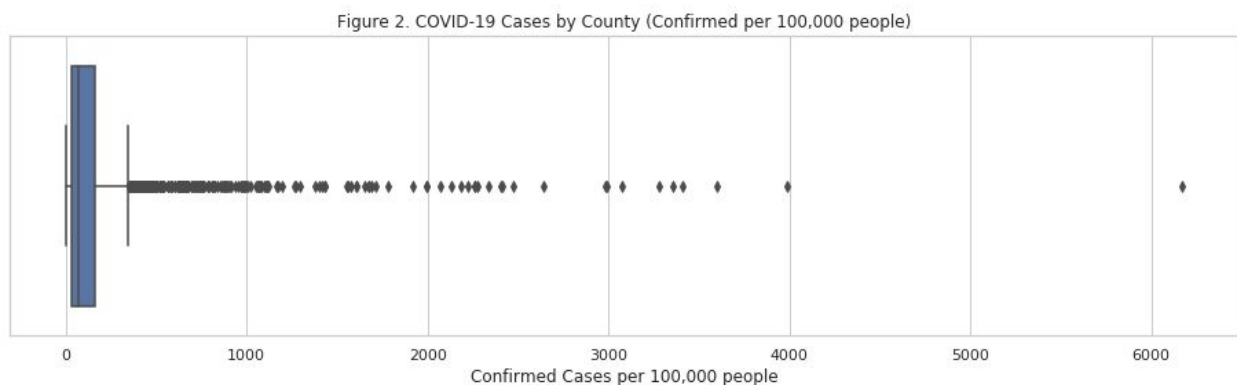
Nationally, the United States unemployment rate in March 2020 reached 4.4%, ranging from 1.0% to 20.7% at the county level with a median of 4.0%. This trend shows a sharp increase as national unemployment was 3.5% just a month before in February 2020 and 3.8% in March of 2019 (See Figure 1).



The National Bureau of Labor Statistics releases unemployment data monthly and is currently published through March 2020. We will begin building our initial model framework for using this available data and then update our data sets to include April 2020 data when available in our final model and analysis.

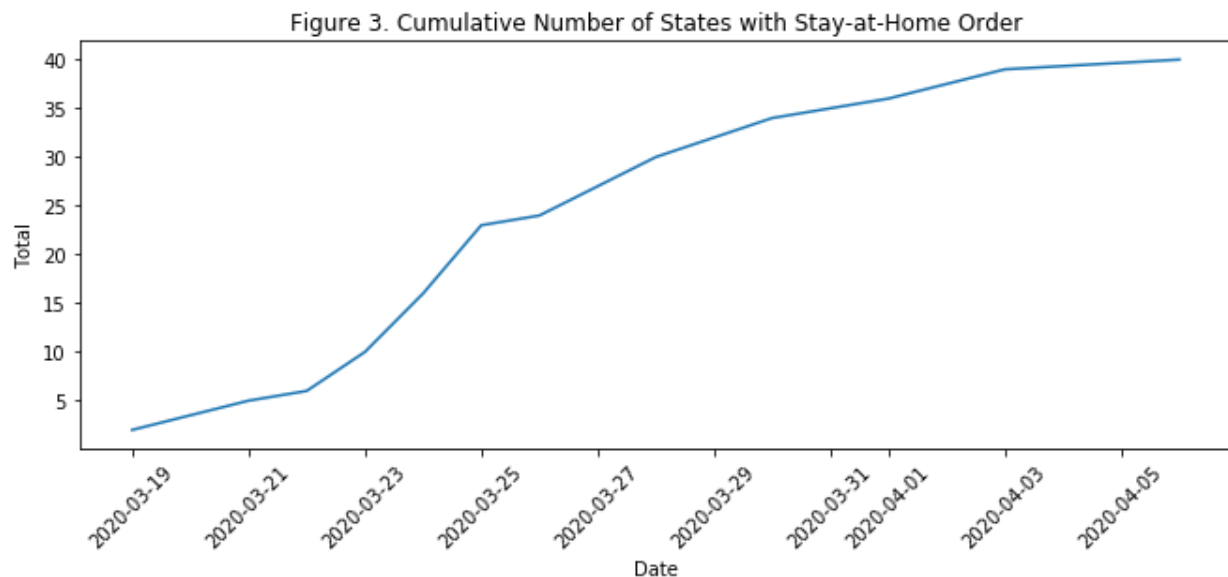
COVID-19 Cases

The boxplot in Figure 2 shows the number of confirmed cases per 100,000 residents by county, calculated from data made available by Johns Hopkins University. County level data shows an average of 162.37 cases for every 100,000 people, ranging from a minimum of 0 cases to a maximum of 6,174.15. Note that this distribution is heavily skewed; compared to the maximum of 6,174.15, the 75th percentile cutoff is just 159.73 cases per 100,000 people.



State-Level Restrictions

Figure 3 illustrates the cumulative number of states that closed essential businesses or implemented some type of sort of shelter-in-place or stay-at-home order by date; if a state implemented both on different days, the earlier date is used. These policies began in California as early as March 19 and spiked during the latter weeks of March with 34 states having issued some sort of guidance by the end of the month. In several cities and counties, local level restrictions may have been implemented earlier which we will capture using the New York Times dataset in our model. Given the heterogeneity in the types of orders issued, we will need to decide how exactly to model government response.



Machine Learning

Because we have a continuous dependent variable (unemployment rate), we will use regression to model this problem. Each county will be an observation. Our models will enable learning as it examines the impact of various factors on the change in unemployment rate.

Each county will be an observation, though some factors that are reported at the state level will be the same for each county in that state. There are 3,143 counties and parishes in the United States, providing a sufficient number of data points to train and test a model. The exact nature of the model is something that we will explore throughout the development process. Some considerations that we have are including proper controls in order to be able to compare counties (for example, using rates instead of absolute numbers to account for differences in population size) and properly representing non-numerical variables (such as the date a stay at home order was instituted) as values that can be used in regression.

Evaluation

We will use cross validation to evaluate the models. We will use standard evaluation techniques, such as mean squared error (MSE), and we will retain a testing set in order to validate the models after training them. We may use further evaluation methods as we cover them in class.

Ethics

One ethical consideration is the potential that some audiences might use our models as justification for lifting stay at home orders too early, and sacrificing public health in the name of economic strength and stability. This would be possible if our models show that early stay at home orders have a large effect on the rise in unemployment.

The range of control variables, such as demographics, that we plan to include will help to mitigate bias. We will look at the change in unemployment rate as well, rather than the absolute unemployment rate, which will control for some counties entering the pandemic with a higher unemployment rate. Even with these controls, there may still be bias present in our models. The pandemic has not affected everyone equally, and in particular has had a harsher impact on communities of color.¹³ Any policy response should keep this in mind. Reopening the economy always means making tradeoffs between economic health and public health, but those tradeoffs are not equal across every community.

¹³ [*'A Terrible Price': The Deadly Racial Disparities of Covid-19 in America*](#), New York Times