UNEMPLOYMENT TRENDS AMIDST COVID-19 AND ITS GOVERNANCE

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

Executive Summary

This project examines the change in unemployment at the county level during the COVID-19 pandemic. The national unemployment rate rose from 3.5% in February 2020 to 14.7% in April 2020 as most states issued stay-at-home orders and closed all nonessential businesses in an effort to stem the spread of the virus. As the economic pressures mounted and increased, some argued that restrictions should be eased so that businesses would be better able to weather the downturn.

Using data on population demographics, political lean, and COVID-19 case trends and government response, this project uses models to predict the percent change in unemployment by county. A series of linear regression models are trained, using several types of regularization in addition to first and second degree polynomials. The best-performing model is found to be a first-degree polynomial with no regularization. In this model, dummy variables by state are more important predictors than most other features, suggesting that unobserved state-level features are better predictors in the change in unemployment than other county-level information.

Given that features related to COVID-19 were not important predictors, the length of stay-at-home orders does not appear to be strongly associated with the change in unemployment. If local and state governments wish to enact measures to mitigate the economic effects of the pandemic, they should look to policies that will strengthen the community's ability to respond to the virus.

Background and Overview

This project aims to model county-level unemployment in the United States using population demographics, political information, COVID-19 case trends, and government responses to the COVID-19 pandemic. As of May 30, 2020, the Centers for Disease Control and Prevention (CDC) reported 1,719,827 total cases in the United States, 5.9% of which have proven fatal.[1] In response to the spread of the virus, states, counties, and cities have enacted various restrictions on businesses, services, and personal behavior.

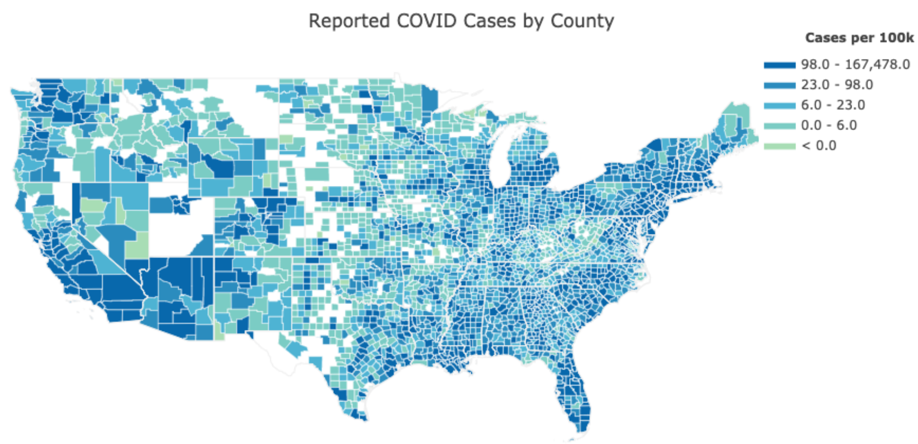

Figure 1. Reported COVI9-19 cases per 100,0000 people (as of April 30, 2020)

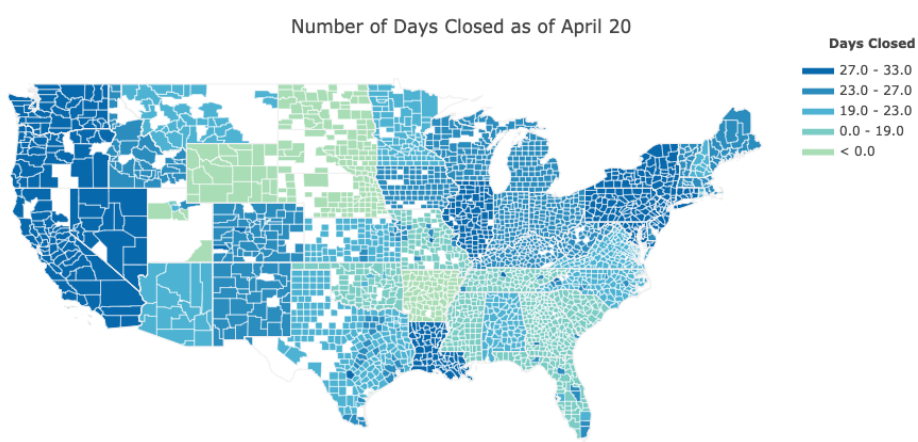

Figure 2. Number of days under a shelter-in-place, stay-at-home, or closure of non-essential business order as of April 20

[1] "Coronavirus Disease 2019 (COVID-19) in the U.S.," Centers for Disease Control and Prevention, May 29, 2020, https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html.

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 ballooned to 14.7% in April.[2] As stay-at-home orders continued causing stores, restaurants, and schools to remain closed, several states and cities saw protests[3] and a few business owners reopened in defiance of government mandates.[4] Federal social distancing guidelines expired April 30 and throughout May, and all states relaxed their orders to varying degrees and at different paces.[5]

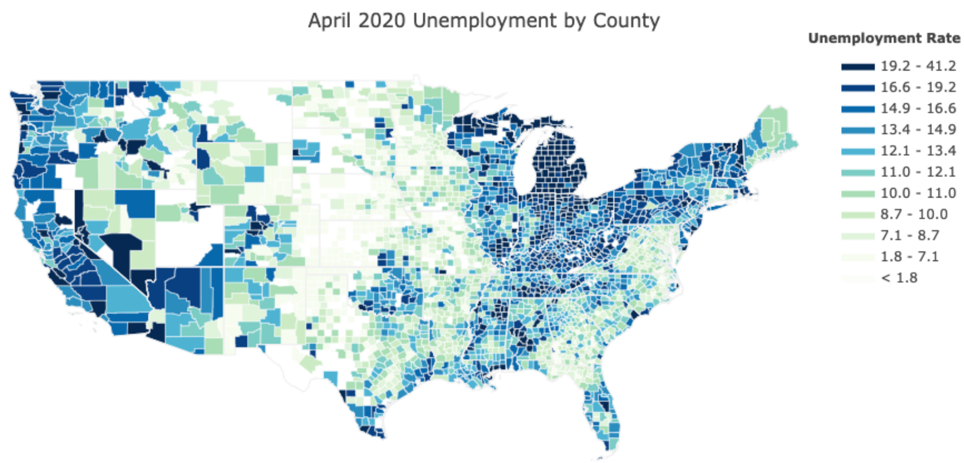

Figure 3. April 2020 Unemployment Rate by County

The economic toll has been far-reaching and there has been hope that the U.S. economy would bounce back as communities began to resume activity. However, early evidence and models point to a different story of enduring impact. A working paper from economists at the University of Chicago estimates that 42% of recent layoffs will result in permanent job loss.[6] Government policies alone have not guided behavior; there were declines in consumer spending that preceded business closures and stay-at-home orders.[7] A working paper by

[2] "The Employment Situation - April 2020" (Bureau of Labor Statistics, May 8, 2020), https://www.bls.gov/news.release/pdf/empsit.pdf.

[3] Abigail Censky, "Heavily Armed Protesters Gather Again At Michigan Capitol To Decry Stay-At-Home Order," *NPR*, May 14, 2020, https://www.npr.org/2020/05/14/855918852/heavily-armed-protesters-gather-again-at-michigans-capitol-denouncing-home-order.

[4] Manny Fernandez and David Montgomery, "Businesses Chafing Under Covid-19 Lockdowns Turn to Armed Defiance," *The New York Times*, May 13, 2020, sec. U.S., https://www.nytimes.com/2020/05/13/us/coronavirus-businesses-lockdown-guns.html.

[5] Sarah Mervosh et al., "See How All 50 States Are Reopening," *The New York Times*, June 1, 2020, sec. U.S., https://www.nytimes.com/interactive/2020/us/states-reopen-map-coronavirus.html.

[6] Jose Maria Barrero, Nick Bloom, and Steven J Davis, "COVID-19 Is Also a Reallocation Shock," Working Paper (University of Chicago: Becker Friedman Institute, May 5, 2020), https://bfi.uchicago.edu/wp-content/uploads/BFI_WP_202059.pdf.

[7] Emily Badger and Alicia Parlapiano, "Government Orders Alone Didn't Close the Economy. They Probably Can't Reopen It.," *The New York Times*, May 7, 2020, sec. The Upshot, https://www.nytimes.com/2020/05/07/upshot/pandemic-economy-government-orders.html.

economists at University of California, Berkeley suggests that stay-at-home orders may only account for about 25% of the rise in unemployment.[8]

With much of the attention focused on these state- and national-level trends in unemployment and government response, this project looks instead at the county level. It aims to determine the extent to which differences in government reaction and case outbreaks can predict the percent change in unemployment at the county level from February 2020 to April 2020.

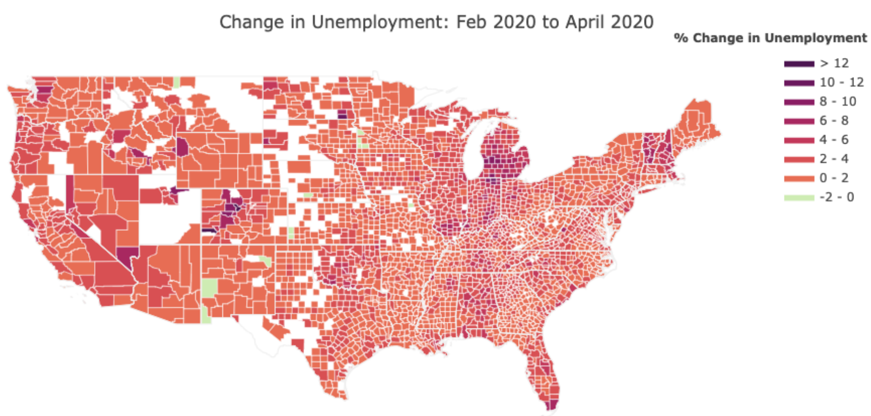

Figure 4. Percent change in unemployment from February 2020 to April 2020 (target model)

As pandemic and restriction management continues, worries about a second wave of infections loom and economic pains mount. State- and county-level policymakers could use these results to inform their actions based on the changes seen in demographically similar areas. In particular, these results could inform the level of economic stimulus or protections that may be necessary. A retrospective view of these results will communicate to infectious disease specialists and the general public the association between various policies and leadership with the economic impact during the time of COVID-19.

Data

Data were compiled from a variety of sources at both the county and state level, which together created a profile of each county that included both background characteristics and information related to COVID-19 conditions and response.

The outcome variable was the percent change county-level unemployment rate from February 2020 to April 2020, which came from the U.S. Bureau of Labor Statistics.[9] Percent change was chosen to account for previous levels of unemployment that existed in the county, thus highlighting the effects of the pandemic throughout March and April.

---

[8] ChaeWon Baek et al., “Unemployment Effects of Stay-at-Home Orders: Evidence from High Frequency Claims Data,” IRLE Working Paper (University of California Berkeley, April 23, 2020).

[9] “Local Area Unemployment Statistics Home Page,” accessed June 1, 2020, https://www.bls.gov/lau/.

Demographic data were primarily drawn from the 2018 American Community Survey[10] and are all measured at the county level. Total population and population counts by race were used to calculate the proportion of the population that is white, black, and Hispanic. Also calculated were the proportion of the population that does not have access to the internet, the proportion with at least a bachelor's degree, and the proportion employed in arts, entertainment, recreation, accomodation, and food services. The food services industry was especially important to consider given the large impact of business closures on tourism and restaurants.[11] The Census Bureau's Population Estimates Program provided population density by county.[12]

Data on COVID-19 case counts by county were drawn from the dataset maintained by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University.[13] The CSSE also provided data on the testing rate at the state level. Both case counts and testing rate are reported per 100,000 residents.

Political lean was included, as it may be an effective predictor of coronavirus response, both in terms of policy action and citizen response. A dataset by the Kaiser Family Foundation[14] provided the political affiliation of the state governor. The political lean by county was measured as the difference in the proportion of votes cast for Donald Trump and Hillary Clinton in the 2016 presidential election, using county-level election data from the MIT Election Lab.[15] Positive values indicate a greater share of votes were cast for Donald Trump.

Quantifying the policy response to the coronavirus was less straightforward. This feature combined two data sources: an article by the New York Times[16] tracking closures and a dataset on a variety of state-level policy responses created and maintained by a team of public health researchers at Boston University.[17] Prior versions of the New York Times article were

---

[10] "American Community Survey Data," The United States Census Bureau, accessed June 1, 2020, https://www.census.gov/programs-surveys/acs/data.html.

[11] Gita Gopinath, "Limiting the Economic Fallout of the Coronavirus with Large Targeted Policies," in *Mitigating the COVID Economic Crisis: Act Fast and Do Whatever It Takes* (London: CEPR Press, 2020), 41–47.

[12] "County Population Totals: 2010-2019," The United States Census Bureau, accessed June 1, 2020, https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html.

[13] JHU CSSE, *CSSEGISandData/COVID-19*, 2020, https://github.com/CSSEGISandData/COVID-19.

[14] "State Political Parties," KFF, January 6, 2020, https://www.kff.org/other/state-indicator/state-political-parties/.

[15] "County Presidential Election Returns 2000-2016," *MIT Election Data and Science Lab, Harvard Dataverse* (Harvard Dataverse, December 4, 2019), https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/VOQCHQ.

[16] Sarah Mervosh, Denise Lu, and Vanessa Swales, "See Which States and Cities Have Told Residents to Stay at Home," *The New York Times*, March 31, 2020, sec. U.S., https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html.

[17] "COVID-19 US State Policy Database (CUSP) - Google Sheets," accessed June 1, 2020, https://docs.google.com/spreadsheets/d/1zu9qEWI8PsOI_i8nI_S29HDGHIlp2lfVMsGxpQ5tvAQ/edit?fbclid=IwAR3hnDxiNpvTOdgcloBG4dBR2plG4y-q3Xz8yoIc9Bbhf8zlcVkhwxPuL1I#gid=973655443.

accessed using the WayBack Machine,[18] as the article was updated to reflect that some local stay-at-home orders were later superseded by state-level orders.

The differences between stay-at-home and shelter-in-place orders did not seem meaningful for the purposes of this analysis, nor were the differences between orders and suggestions. Any such declaration was treated the same under the umbrella term of a stay-at-home order. Similarly, in the context of business closures, there were differences among what businesses were considered essential, but all closures of nonessential businesses were treated the same. Some states may have closed nonessential businesses several days before implementing a stay-at-home order, while others did so simultaneously. For the purposes of this analysis, it was decided that either measure—a stay-at-home order or the closure of nonessential businesses—was likely to have similar impacts on the economy. If a measure went into effect at or later than 5 p.m., it was recorded as being enacted the following day.

This produced three dates for consideration: the implementation of a stay-at-home order for a county, the implementation of such an order for a state, or the closure of nonessential businesses for a state. Each county may have had up to three relevant dates, or zero if no measures were enacted there. The earliest of the relevant dates was used, and the number of days under that measure as of April 20 was calculated. This number would be zero if no measures were enacted in that county. April 20 was chosen instead of April 30 because some areas began lifting restrictions as early as April 24. If an order was enacted on the city level, that date was recorded as the date for the county if the city's population was at least 50% of its county's population.

Because the pandemic remains ongoing, this process involved a number of simplifications and also relied on data that may not have faced the level of scrutiny given to other sources. The creation of a comprehensive dataset of these policy measures, in particular at the local or county level, is outside the scope of this assignment and will be the work of policy researchers in the coming months. Constraints such as limited time and the continuously evolving nature of the situation presented challenges to obtaining a higher-quality dataset.

Models were run twice, once including one-hot encoded state indicators among the feature set and again without the states. Including the state is a recognition that county-level observations are not entirely independent due to regional economic impacts, and also implicitly captures unobserved state-level predictors.[19]

---

[18] "Internet Archive: Wayback Machine," accessed June 1, 2020, https://archive.org/web/.

[19] BC Thiede and SM Monnat, "The Great Recession and America's Geography of Unemployment.," *Demographic Research* 35 (September 27, 2016): 891–928, https://doi.org/10.4054/demres.2016.35.30.

Machine Learning and Solution Details

This is a prediction machine learning problem, attempting to predict change in unemployment. It is important to note that this analysis is not intended to predict future unemployment rates, which would require a time-series analysis. Instead, data were randomly split into training and testing sets to predict the percent change in unemployment for one time period: change from February 2020 to April 2020. Since percent change is a continuous variable, as were most of the features, regression models with ridge, lasso, and elastic net regularization were fitted, as well as a linear regression model with no regularization.

Appendix A includes a full table of the feature and target attributes, their sources, and the level at which they were reported. All continuous features were normalized. Non-normalized variables include state, which is a categorical feature that was one-hot encoded, and party affiliation for the state's governor, a binary indicator coded as 1 for a Republican governor and 0 for Democratic. Most features from the American Community Survey (White, Black, Hispanic, no internet access, bachelor's degree or higher, and workers in population, arts, recreation, entertainment, and food services) were transformed from counts to percentages by dividing by the relevant county total.

The election differential for the 2016 presidential election was calculated by subtracting the proportion of votes cast for Hillary Clinton from the proportion of votes cast for Donald Trump. As discussed, days with essential business closed or under stay-at-home orders were converted from the date each order was enacted to the number of days it had been in effect as of April 20 to curtail the influence of states that began reopening or loosening restrictions toward the end of April. The median income was transformed into log median income due to its right-skewed distribution. Scatter plots of each continuous feature against the target suggested that further transformations were unnecessary.

Using scikit-learn's Pipeline and GridSearchCV, a linear regression model without regularization was fit, as well as linear regression models with ridge, lasso, and elastic net regularization. Each model was fit with polynomial degrees of 1 and 2 and cross-validation with a parameter of 5. Visualizations of each feature plotted against the target did not suggest a need to include higher degrees. For the regularization models, the alpha parameters included a range values between 0.0001 and 1.0, inclusive. These models were fitted twice, once including the one-hot-encoded dummy variables for states, and then again omitting the state attributes.

A random forest regressor was also considered. However, such a model would need to exclude states because one hot encoding weakens the performance of tree-based models by introducing sparseness in the feature space. Given that this would make it difficult to properly compare across model types, the random forest regressor was not utilized and the results are not reported here.

Evaluation and Results

Using lowest mean squared error as the determining evaluation metric, the linear regression model without regularization and a polynomial degree of 1 performed best on the test data set, with a mean squared error of 0.4152 an adjusted R-squared value of 0.5516. (See Appendix B for evaluation metrics for additional models.)

Figure 5 shows the most important features of this model, with the top features and their coefficients highlighted in Table 1.

Figure 5. Feature importances from the best model: linear regression (degree 1) with no regularization, sorted by the absolute value of the coefficient. Appendix C includes a full list of these features and their coefficients.

Table 1. Most significant features

Feature	Coefficient
State: Hawaii	3.1816
State: Connecticut	-1.6835
State: Michigan	1.6751
State: District of Columbia*	-1.4694
State: Maryland	-1.3589

** In the ACS data, the destination for D.C. was included in the "state" field that was one-hot encoded.*

Notably, the one-hot encoded state features were more important than most other features. The importance of COVID-related features such as length of stay-at-home order was much lower. This suggests that variations in unemployment rate are associated more strongly with some factors inherent to states themselves that are not otherwise captured by the attributes in the feature set.

The models were trained again without state features to explore the difference in the predictive power of the resulting models. In the best model without state features (elastic net regularization degree 1, alpha = 0.01), the most important predictors were governor party and proportion of population working in arts, recreation, accommodation, or food services.

Policy Recommendations

Since government policy response was not a relatively important predictor in these models, this analysis does not suggest that the length of a stay at home order is as strongly associated with the rise in unemployment as other factors. While the testing rate was also a relatively weak predictor compared to other features, few states have achieved a robust testing capacity or implemented contact tracing programs.[20] This is true in early June and was certainly the case throughout March and April.

The models presented here do not suggest that an end to pandemic-related restrictions will be an effective means of stemming the rise in unemployment. Instead, states should focus on policies related to expanding health care capacity such as increasing the availability of tests and personal protective equipment. Doing so will bolster the ability of a community to curb and contain the spread of the virus, which will ultimately boost consumer confidence. Allaying this fear means that citizens will be less hesitant to resume ordinary economic patterns as restrictions lift. In the meantime, while some or most economic activity must remain restricted, states should focus on robust economic supports for their citizens. Doing so may help communities retain jobs and alleviate long-term economic shocks.

Ethics

Because this analysis looks at data compiled on a county level, it does not examine nuanced differences for how unemployment and other COVID-19 responses disproportionately impact low-income communities and communities of color. The range of demographic variables included as predictors in these models were included in response to that. Nonetheless, the models cannot escape bias in the data upon which they were trained. In particular, the official unemployment rate reflects a particular way of measuring unemployment and does not capture underemployment or the loss of economic activity outside formal employment.

An ethical consideration that arose prior to training the models was that, if the length of stay-at-home orders was an important predictor for the rise in unemployment, the models could be used as justification for avoiding or shortening such measures in the future. There

[20] Erik Ortiz, "As States Reopen, Contact Tracing Efforts Hobbled by Obstacles," *NBC News*, May 19, 2020, https://www.nbcnews.com/news/us-news/states-reopen-contact-tracing-efforts-hobbled-obstacles-n1210266.

would be no true way to mitigate this potential impact if the models supported a strong relationship between stay-at-home order length and rise in unemployment. At best, such results could come with an urging to pair stay-at-home orders with policies such as stimulus benefits that would soften the economic impact, but there would be no way to hold policymakers accountable for following through with such actions.

Limitations

The ongoing nature of the coronavirus epidemic posed several challenges. Measuring policy responses was also difficult because no county-level dataset existed already. Using two sources provided a check for this, although a more rigorously assembled county-level dataset would have been preferred.

The data that we hoped to use, county-level unemployment for April 2020, was not released until June 3, and even then the figures released were only provisional. Comparing the provisional rates initially released for March with updated figures released with the provisional April data showed that the difference was generally small, within 0.3. This provided confidence in using the provisional April data, although ideally these models would later be retrained with the updated, official figures.

The included features, while numerous, do not predict changes in unemployment as well as desired. Before the Bureau of Labor Statistics release of the April unemployment data, these models were trained using the percent change from February 2020 to March 2020. These models reported much lower mean squared error; the best model (ridge regularization degree 1, alpha = 0.1) had mean squared error of 0.187, compared to 0.458 for the same model using February to April. This further indicates that there are factors that occurred in April that are not captured in the feature set. Consumer and employer confidence could be important to consider, if good quantitative measures of this sort exist. These models do not capture when states or counties planned to lift restrictions, nor do they include data on the extent to which residents adhered to stay-at-home orders.

Clearly, further exploration is needed in this area to better understand the factors associated with changes in unemployment during the pandemic. Disentangling the factors associated with changes in unemployment will be an ongoing process and one that may be difficult to understand, especially while the pandemic is still unfolding. Preliminary investigations such as this will nonetheless serve to assist policymakers and economists in the short term.

Appendix A. Variables and Data Sources

Variable (API Variable Name)	Source	Level
Proportion of population, White (calculated from B02001_002E and B02001_001E)	2018 American Community Survey	County
Proportion of population, Black (calculated from B02001_003E and B02001_001E)	2018 American Community Survey	County
Proportion of population, Hispanic (calculated from B03001_003E and B02001_001E)	2018 American Community Survey	County
Proportion of population, no internet access (calculated from B28002_013E and B288011_001E)	2018 American Community Survey	County
Proportion of population, bachelor's degree or higher (calculated from B23006_001E, B23006_022E, B23006_023E, B23006_024E, and B23006_025E)	2018 American Community Survey	County
Proportion of population, arts, recreation, entertainment, accomodation, and food services (calculated from C24050_001E and C24050_012E)	2018 American Community Survey	County
Log median family income (calculated from B19013_001E)	2018 American Community Survey	County
Population density	2018 Census Bureau Population Estimates Program	County
COVID-19 case total, 4/30/20 (per 100k)	Johns Hopkins Center for Systems Science and Engineering	County
COVID-19 testing rate (per 100k)	Johns Hopkins Center for Systems Science and Engineering	State

Appendix A. Variables and Data Sources, continued

Variable (API Variable Name)	Source	Level
Political affiliation of governor	Kaiser Family Foundation	State
Difference in proportion of votes cast for Donald Trump and Hillary Clinton in 2016 presidential election	MIT Election Lab	County
Days with essential businesses closed or under stay-at-home order as of April 20, 2020	New York Times; spreadsheet maintained by a team at Boston University	Earliest of City, County, and State
% Change in Unemployment by Year (calculated from April 2019 and April 2020 data)	Bureau of Labor Statistics	County
% Change in Unemployment over COVID timespan (calculated from February 2020 and April 2020 data)	Bureau of Labor Statistics	County

Appendix B. Evaluation Metrics from Grid Search Models

For each grid search, the best fitted model object was added to a dictionary, along with the model's mean squared error, R-squared value, and adjusted R-squared value, rounded to the nearest ten-thousandth. The results from this dictionary are compiled in the table below, sorted by mean squared error. Models without the state indicators are shaded.

Model	Parameters	MSE	R^2	Adj R^2
No Regularization, with states	degree = 1	0.4512	0.6024	0.5516
Ridge, with states	degree = 1 alpha = 1.0	0.4582	0.5962	0.5447
Lasso, with states	degree = 1 alpha = 0.001	0.4613	0.5935	0.5416
Elastic Net, with states	degree = 1 alpha = 0.001	0.4648	0.5904	0.5381
Elastic Net, without states	degree = 1 alpha = 0.01	0.8147	0.2821	0.2652
Ridge, without states	degree = 1 alpha = 1.0	0.8149	0.2819	0.2649
Lasso, without states	degree = 1 alpha = 0.001	0.8152	0.2816	0.2647
No Regularization, without states	degree = 1	0.8152	0.2816	0.2647

Appendix C. Feature Importances from Best Model

Features are ranked by the absolute value of the coefficient and shaded for readability. One-hot encoded state variables are indicated by their state abbreviation.

Rank	Feature	Coefficient
1	HI	3.1816
2	CT	-1.6835
3	MI	1.6751
4	DC	-1.4694
5	MD	-1.3589
6	AL	1.2214
7	MN	-1.1248
8	AZ	-1.0369
9	VT	0.9938
10	IN	0.9415
11	OK	0.8034
12	WY	-0.7660
13	ME	-0.7453
14	MO	-0.6621
15	NH	0.6371
16	CA	-0.6190
17	KY	0.6137
18	WA	-0.5977
19	IL	0.5957
20	IA	-0.5797

Rank	Feature	Coefficient
21	NM	-0.5563
22	Governor Party (1 = Republican)	0.5362
23	WV	-0.5169
24	TN	0.4415
25	OH	-0.4250
26	PA	-0.4230
27	LA	0.4191
28	NY	-0.3941
29	VA	0.3923
30	MT	-0.3878
31	SC	0.3867
32	MS	0.3465
33	UT	0.3414
34	NV	0.3342
35	ND	0.3247
36	2016 Election Differential	-0.3242
37	Log Med. Income	0.3210
38	NC	0.3099
39	RI	0.2896
40	NE	-0.2604

Appendix C. Feature Importances from Best Model, continued

Rank	Feature	Coefficient
41	NJ	-0.2532
42	SD	-0.2158
43	CO	0.2155
44	% in Services	0.2056
45	ID	-0.2056
46	AR	0.1956
47	Days Closed	0.1948
48	MA	-0.1659
49	WI	-0.1408
50	% Black	-0.1379
51	GA	0.1341
52	% White	0.1281
53	% Bachelors +	-0.1199
54	% No Internet	-0.1177
55	TX	-0.1155

Rank	Feature	Coefficient
56	% Hispanic	-0.1068
57	OR	-0.0997
58	FL	0.0771
59	DE	-0.0476
60	Testing Rate	-0.0311
61	KS	-0.0208
62	Pop. Density	0.0117
63	COVID Cases	0.0000
64	AK	0.0000