

Testing Narratives: Would Trump have won in a COVID-free world?

December 10, 2020

As discussed in my blog post about electoral shocks and COVID-19, my model used polling and economic data as proxies for the impact of COVID-19 on the election. Given the importance of economic fundamentals in determining election outcomes, it is not unreasonable to think that the pandemic's toll on the economy must have hurt Trump's re-election bid. Considering all that we have learned over the past few months, it seems that Trump's strong (pre-pandemic) economic record and his incumbent advantage could have reasonably carried him to another 4 years in the White House.

Several paths existed for a Trump victory in the Electoral College, but they all failed to come to fruition on Election Night. Donald Trump failed to secure Arizona, Georgia, and Wisconsin in 2020, despite having won all three of them in 2016. These are just a few of several states that were hit relatively hard by the pandemic and flipped from red to blue in 2020. In an alternate universe without COVID-19, would Donald Trump have won the 2020 election? In other words, did Trump lose because of COVID-19, or would Biden still have defeated the incumbent president in a COVID-free world?

How will I test the effects of COVID on the election?

Lack of experimental data Of course, I cannot run a randomized experiment with treatment and control groups to determine the causal impact of COVID-19 on Trump's vote share. Ideally, we would compare Donald Trump's vote share in areas hit by COVID-19 to areas untouched by the pandemic. However, the widespread nature of the pandemic makes natural experiments nearly impossible in this scenario. Because we do not have access to the ideal experimental data, I will examine differences in Trump's vote share across areas with varied COVID rates. To determine the exact regression formula, I looked to research the only comparable event to the COVID-19 pandemic in recent history: the 1918 Spanish influenza pandemic.

Drawing inspiration from previous research In *Democracy For Realists*, Achen and Bartels examined whether the states and cities hit hardest by the pandemic responded differently at the polls.¹ They found that the Spanish flu had little to no impact at the polls, but the national dialogue surrounding the pandemic looks quite different now than it did a century ago. Relative to the magnitude of the pandemic, the Spanish flu received little public attention, which contrasts greatly with how COVID-19 has dominated nearly every facet of life in 2020.

While Achen and Bartels focused on gubernatorial races during the 1918 midterms, I plan on applying the underlying structure of their regressions to the 2020 data. Achen and Bartels centered their analysis around excess flu deaths and vote share in the previous election. Similar to their analysis, I plan on running a simple regression that maps Donald Trump's 2020 vote share as a function of his 2016 vote share and COVID cases or deaths as a percentage of the state or county population.² Controlling for previous election results in these regressions will help to isolate the impact of COVID-19 on results between the two elections.³

¹[Achen and Bartels, 2017] Achen, C. H. and Bartels, L. M. (2017). *Democracy for realists: Why elections do not produce responsive government*

²Achen and Bartels also included a dummy variable that indicated whether the specified gubernatorial candidate was a Democratic incumbent. Since Donald Trump ran in both 2016 and 2020, I did not include an indicator for incumbency and instead just focused on his vote share for both races.

³This is still far from perfect since it does not account for any demographic changes within that state/county between 2016 and 2020. However, it does a much better job than if we simply looked at 2020 vote share and COVID numbers.

Regressions

Data sources My initial regression mapped Trump's 2020 state-level vote share from his 2016 vote share and total COVID deaths⁴ as a proportion of the state's population. Then, I extended the same underlying regression to explore COVID *cases* rather than deaths and *county-level* voting and then repeated these regressions at the county-level. These regressions use COVID, population, and voting data from several sources:

- **COVID-19 cases and deaths:** The state-level COVID data came from the CDC COVID Data Tracker, and the county-level COVID counts came from the GitHub of the New York Times.
- **Population:** I used 2019 population estimates from the US Census Bureau to calculate the COVID cases and/or deaths as a percentage of the total population within the relevant state or county. While 2020 population data would have been ideal, this was the most recent data I could find from a reliable source.
- **Voting:** Voting data is generally available from many sources, but this spreadsheet contains the vote counts I used for the county-level regressions. Of course, I used state-level counts for my state regressions.

State-level results The COVID deaths coefficient in the state-level model had a relatively large p-value; the same regression with cases instead of deaths yielded a slightly smaller p-value, but still statistically insignificant. Both of these coefficients were positive, but the large p-values make it difficult to draw any conclusions from the regression. Perhaps the pandemic was unlikely to sway voters in heavily partisan states but was more of a deciding factor in battleground states?

Sure enough, the state-level regressions yielded more significant coefficients for the COVID terms when focusing solely on battleground states. With a p-value of 0.086, the coefficient for COVID cases in battleground states yields the most significance. When paired with a significance level of $\alpha = 0.10$, we have sufficient evidence to conclude that a 1% increase in a battleground state's case count as a percentage of the population is associated with an approximate increase of 0.57% of Trump's 2020 two-party vote share within that state.

Comparing State-Level Models

<i>Predictors</i>	All States and COVID Deaths		All States and COVID Cases		Battleground States and COVID Deaths		Battleground States and COVID Cases	
	<i>Estimates</i>	<i>p</i>	<i>Estimates</i>	<i>p</i>	<i>Estimates</i>	<i>p</i>	<i>Estimates</i>	<i>p</i>
(Intercept)	0.98	0.497	1.36	0.283	2.04	0.647	5.69	0.202
Trump's 2016 Two-Party Vote Share	0.95	<0.001	0.93	<0.001	0.90	<0.001	0.83	<0.001
Deaths as Percent of Population	3.37	0.644			26.03	0.141		
Cases as Percent of Population			0.27	0.277			0.57	0.086
Observations	50		50		16		16	
R ² / R ² adjusted	0.972 / 0.971		0.973 / 0.972		0.905 / 0.890		0.911 / 0.897	

The significance of the slope coefficients depends on how you select your α value. A p-value of $\alpha = 0.05$ deems the results insignificant, but the regression still indicates the *possibility* of a positive association between COVID cases and Trump's 2020 vote share.

⁴I used total COVID cases and deaths as of Election Day.

County-level results Next, I wanted to extend the analysis one step further and run the same regressions with county-level COVID metrics and vote shares. Again, all of the coefficients indicated a positive relationship between Donald Trump's 2020 vote share and an increase in COVID cases or deaths as a percentage of the county's population. This time, all of the slope coefficients yielded significant⁵ p-values at an $\alpha = 0.001$ significance level, confirming the positive association:

Comparing County-Level Models

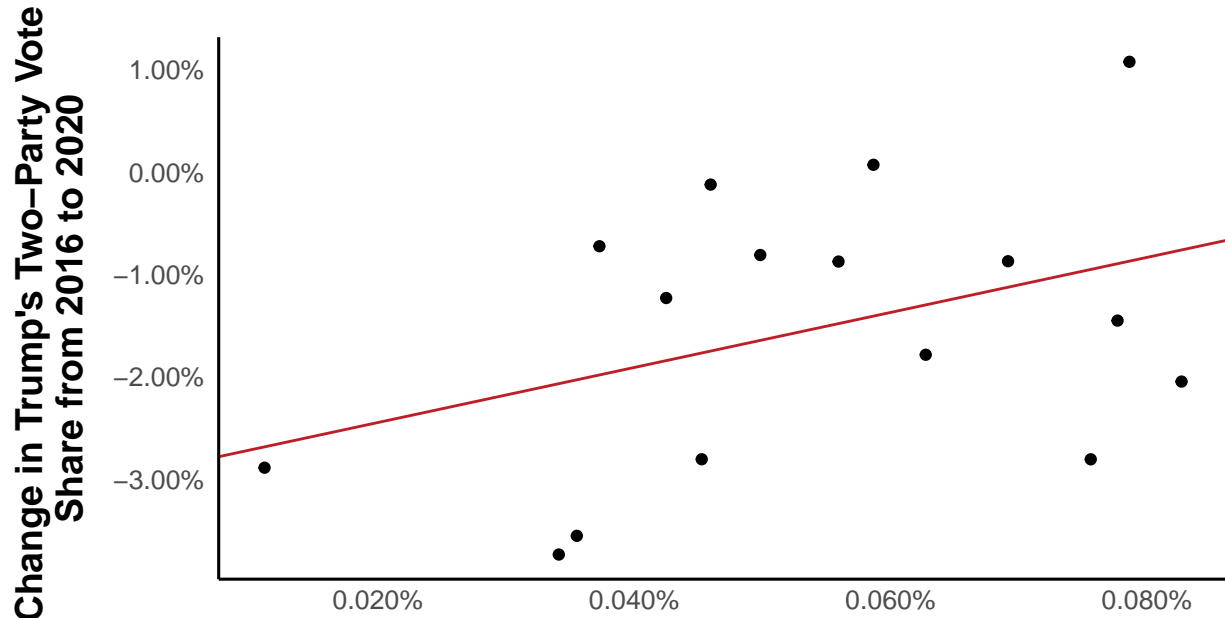
<i>Predictors</i>	All Counties and COVID Deaths		All Counties and COVID Cases		Counties in Battleground States and COVID Deaths		Counties in Battleground States and COVID Cases	
	<i>Estimates</i>	<i>p</i>	<i>Estimates</i>	<i>p</i>	<i>Estimates</i>	<i>p</i>	<i>Estimates</i>	<i>p</i>
(Intercept)	-1.40	<0.001	-1.19	<0.001	0.74	0.055	1.19	0.004
Trump's 2016 Two-Party Vote Share	1.00	<0.001	1.00	<0.001	0.97	<0.001	0.97	<0.001
Deaths as Percent of Population	11.20	<0.001			14.89	<0.001		
Cases as Percent of Population			0.30	<0.001			0.28	<0.001
Observations	2978		2978		1252		1252	
R^2 / R^2 adjusted	0.975 / 0.974		0.974 / 0.974		0.964 / 0.964		0.961 / 0.961	

The below plots provide a clearer visualization of this positive association between COVID-19 and Trump's 2020 vote share within battleground states. The regressions control for 2016 vote share when regressing the 2020 vote share from COVID deaths. To illustrate the multivariate relationship between the variables in these plots, the y-axis displays the difference between 2020 and 2016 vote share. Controlling⁶ for 2016 vote share in the regression isolates the impact of COVID on 2020's vote share since it looks specifically at the changes in voting patterns rather than absolute vote share in 2020:

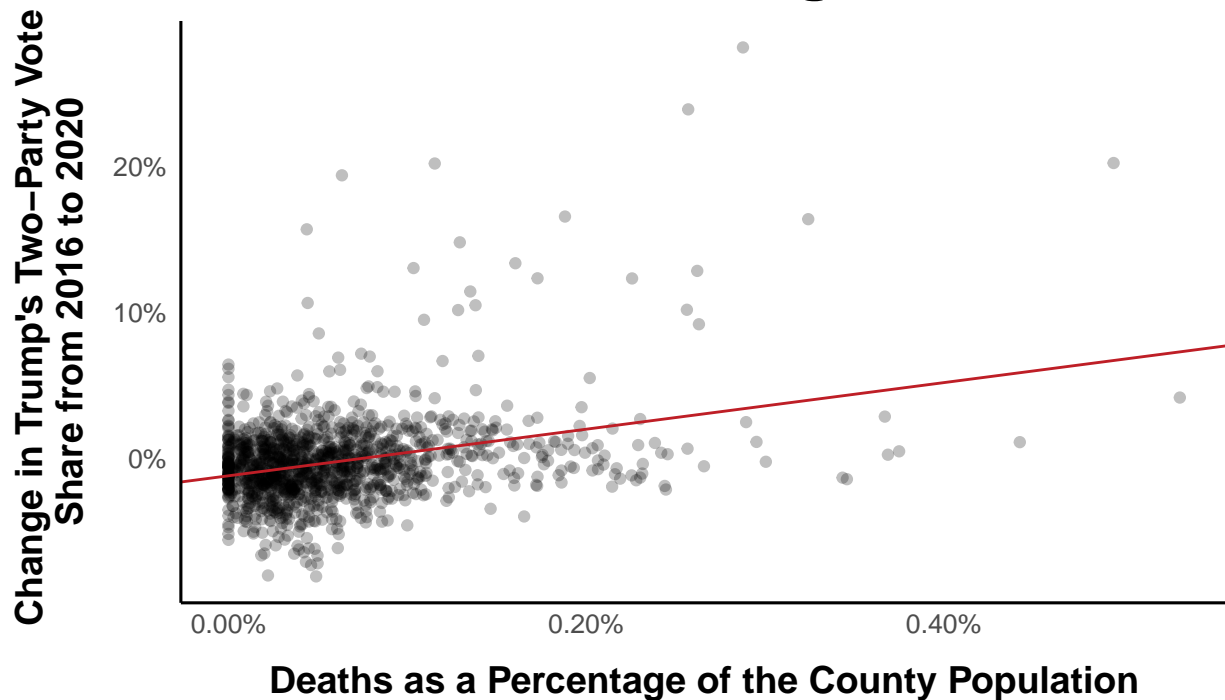
⁵Increasing the sample size generally results in a smaller p-value if the null hypothesis is false. In this case, our null hypothesis is that COVID-19 has no association with Trump's vote share, and since the p-values are significant, we can conclude that we have sufficient evidence of a positive association between Donald Trump's 2020 vote share and COVID-19 cases/deaths.

⁶This is still far from perfect since it does not account for any demographic changes within that state/county between 2016 and 2020. However, it does a much better job than if we simply looked at 2020 vote share and COVID numbers.

Republican Vote Share and COVID-19 Deaths in Battleground States



County-Level Vote Share and COVID-19 Deaths within Battleground States



Forecasting a hypothetical, COVID-free 2020

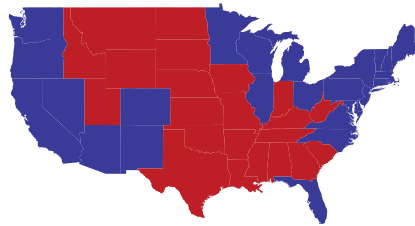
Taking an interest in the results of the regression, I decided to take a more nuanced look at the implications of these findings. The regressions take very crude measures of COVID numbers and previous vote share, without considering possible confounding variables. My final forecast drew from a mixture of demographic variables, economic metrics, incumbency status, and polling numbers to produce a probabilistic forecast for the 2020 election. While the forecast did not match the election results exactly, it did match the outcomes fairly closely, so it would not hurt to examine what happens without any measured impact of COVID-19.

COVID-19 bled into the polling and economic data used for the predictions, so I took steps to try to erase or minimize any impact of COVID-19 on these metrics:

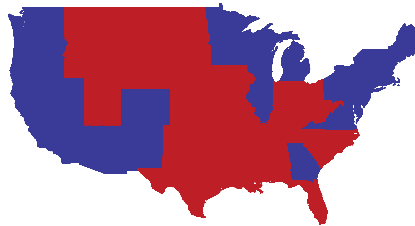
- I treated the election as if Trump was running for re-election off of his 2019 economy by using 2019's Q1 GDP in the prediction.
- For polling, I crafted my predictions with state-level polls from at least 15 weeks⁷ before the election, as opposed to focusing on the 4 weeks preceding the election.

I used a very similar⁸ model equation to that from my final forecast. In this hypothetical, pandemic-free world, Trump lost both the Electoral College and the national two-party popular vote by an even larger margin than what panned out on the actual election day and what was predicted in my final prediction:

Forecasted COVID-Free Results



Actual Election Results



Type	Biden EV	Trump EV	Biden Two-Party Vote	Trump Two-Party Vote
COVID-Free Forecast	349	186	53.3 %	46.7 %
Actual Results	306	232	51.3%	46.9%

⁷While this includes some data from after COVID-19 came to the United States, I had to expand the window of time to get a large enough sample size of polls for each state. The model would not run if I restricted the polling window to before the pandemic came to the US.

⁸I added an interaction term to the model from my original forecast for this iteration. In retrospect, it did not make sense to not include it in the first place since the state of the economy likely has opposite effects for incumbent and non-incumbent candidates. The final equation used to predict \hat{y} , the probability of voting for the Democratic or Republican candidate within each state, in this updated model is

$$\hat{y} = \text{incumbent} * \text{gdp_growth_qt} + \text{avg_state_poll} + \text{prev_dem_margin} + \text{black_change} + \text{age20_change} + \text{age65_change}$$

Do these results support the narrative? No, actually the opposite.

The pervasive nature of COVID-19 eliminates the possibility of a natural experiment, making a clear test of the impact of COVID-19 on the presidential election nearly impossible. The above regressions and hypothetical forecast do, however, indicate that COVID likely did not hurt—and could have possibly helped—Donald Trump’s 2020 re-election bid. The regression controls for Donald Trump’s performance in the 2016 election, but it fails to account for the possibility of demographic and socioeconomic changes that could have occurred between 2016 and 2020. While the above tests do not allow us to conclude that COVID-19 *caused* Trump to perform better in the 2020 election, these preliminary measures do support the likelihood of a *positive association* between COVID numbers and Donald Trump’s vote share.

While these findings seem somewhat counterintuitive, others have also suggested that COVID may have boosted Trump’s support in the 2020 election. Perhaps lockdowns following local outbreaks angered voters who support the idea of small government? This hypothesis does not seem out of reach given the protests and kidnapping plots that emerged in response to Michigan’s crackdown on COVID cases earlier this year.

The possibility of a positive association between Trump’s vote share and the impact of COVID-19 warrants further exploration. While I certainly hope that the near future does not hold any additional global pandemics, any lessons from this election will help to enhance our general understanding of how voters respond to large-scale crises.
