Donald Trump Popularity Project

Applied Econometrics

Connor Ayubi, John Townley, Chris Zupan

February 2019

1. Table of Contents:

2. Executive Summary

At the end of every fiscal year, Congress and the President must reach an agreement on a series of federal budget proposals to be put into effect the following year. In 1976, a failure to do so by President Gerald Ford and the Democratic majority in Congress made the government inactive for a total of twelve days. Since this first modern "government shutdown," the everyday operations of the federal system have been brought to a halt on over twenty separate occasions under the watch of six of the seven Presidents in office since Ford.

It was not unprecedented that on December 22nd, 2018 the United States government was shutdown once again, this time due to President Trump's insistence on receiving funding for a border wall, and a recently flipped Democratic Congress's unwillingness to accept. This most recent government shutdown lasted for a total of 35 days, the longest in US history (previously 21 days in 1995). The effects have been felt across the country, with an estimated 800,000 federal employees being forced to work without pay, which according to salary data from the U.S. Office of Personnel Management, "cost U.S. taxpayers more than $86 million a day" in back pay promised to those workers (ThoughtCo, 2019).

It would seem intuitive that Trump's approval rating would take a hit due to the shutdown and its unparalleled negative effects, but Trump's controversial stint as President has been full of many surprises. In this report, we will search to definitively answer the question of what effect (if any) this shutdown has had on Donald Trump's popularity. We will do this by comparing the actual popularity data from the shutdown period to what we would expect his popularity to be if there had been no shutdown. We will also examine how any effects we might find compare to the effects of previous government shutdowns on the standing President's popularity, to give another measurement illustrating the level of impact the shutdown had uniquely on Trump.

Our ultimate goal for this project is to answer the empirical question: what effect (if any) did the most recent government shutdown have on President Trump's popularity. In order to do so, we collected data from FiveThirtyEight containing polling results on Americans' "approval" and "disapproval" of Donald Trump as President from multiple different pollster sources. After deciding that Donald Trump's Twitter activity would be another worthwhile variable to consider,

we then collected his Trump's "Tweets per Day" for the relevant dates. We were then able to run a multivariable regression where we tested the effects of the President Trump's tweeting and the length of the government shutdown on his popularity.

To further contextualize this analysis, we chose to follow this with comparisons of presidential popularity effects resulting from past shutdown. To do this, we collected data from Gallup Polls, which although it is not as well rounded as the FiveThirtyEight measures since it doesn't take it other sources polls into account, does maintain consistency for each president. This data was then cut down to focus only on approval rating numbers directly before, during, and after any government shutdown each President experienced. We discuss more specifics regarding the data in the following "data description" portion of our report.

Our methodology for determining the effect that the most recent government shutdown had on Trump's popularity was to first decide what potential variables could be correlated with his popularity, which could be used as secondary variables. We collected data for each of these variables, our methods for which we will discuss later, then compiled this data into a STATA friendly format. From here we tested what effects these variables had on Trump's popularity estimate, using data from FiveThirtyEight to measure the effects, and concluded that certain variables, such as the stock market, should not be included while others such as tweets, should be included in our regression. Moving forward we ran our regressions seeing the effect of the shutdown on President Trump's popularity as well as the effect post-shutdown, then ran our multivariable regression with length of the shutdown and tweets as regressors. We also analyzed the effect that government shutdowns had on the popularity of previous Presidents, which we then compared with the effect that this shutdown had on President Trump. We will provide additional details on our specific methods and our data in the sections that follow.

We will discuss specifics about our results more in the econometric analysis portion of our paper, from our regressions we conclude that the shutdown had a negative effect on President Trump's popularity. Similarly, while not having as pronounced of an effect as the length of the shutdown, our variable "Tweets per Day," had an overall negative correlation on President Trump's popularity, but a positive correlation when looked at during the shutdown period.

3. Data Description

For the purpose of this project our group obtained approval rating data for Trump, the dependent variable, directly from FiveThirtyEight, a website which analyzes politics and opinion poll analysis. Specifically, we got the poll data from this link: https://projects.fivethirtyeight.com/trump-approval-data/approval_polllist.csv. This downloaded data has measures that include the pollster, the grade of the pollster (based on how accurate the group has been historically), the population who was polled (All Adults, Voters, etc.), the weight of each poll (based on sample size, frequency of polls from pollster, grade of pollster, and how the pollster conducts surveys), approval ratings and disapproval ratings plus their adjusted counterparts, as well as the direct links to each poll.

From here we decided which specific measure should be used to operationally define "popularity." We decided to go with adjusted approval ratings rather than the original unweighted approval ratings in our regressions, as the adjusted approval ratings account for house effects. By house effects, it means that if a pollster has shown persistent bias when compared to the trend line in approval or disapproval ratings in the past, then FiveThirtyEight adjusts that pollster's data to account for this bias. We believe using the adjusted approval ratings provides better data to analyze, as FiveThirtyEight attempts to remove any bias from polling results in this metric.

With regards to our independent variables, we got this information from a variety of locations. While we each personally knew the length of the shutdown from seeing the news or our in-class discussion of the project, we also analyzed the effect that President Trump's tweets had on his popularity, specifically through John's research. John looked at President Trump's Twitter feed and counted how many times each day, during the thirty-five-day government shutdown as well as the two weeks before and after, President Trump tweeted or retweeted a post. For the purpose of this experiment, tweets and retweets were conglomerated together to form one singular variable "tweets." President Trump has a habit of posting a tweet, then in the following tweets retweeting his own tweet. For these retweets of his own tweets, the original tweet was attributed to the day he posted it, and the retweet of the initial tweet was attributed to

the day he retweeted it, although the timestamp on Twitter reads the day of the initial tweet. Using this data, we then ran a multivariable regression with President Trump's popularity, the length of the shutdown, and the frequency of tweets from President Trump. The "tweets" variable was then assigned to match up with the "shutdown" variable.

We also compared President Trump's popularity with the eleven previous Presidents, using data from Gallup. Though not as complete as the data we pulled from FiveThirtyEight for President Trump's popularity, Gallup still has a pollster grade of B according to FiveThirtyEight, which is why we felt comfortable comparing President Trump's popularity with previous Presidents using this data. We considered scaling these polling results in a similar fashion as the FiveThirtyEight data, but ultimately concluded that we could not make the scaler identical given our limited information in the Gallup data and related research.

In the dataset we downloaded from FiveThirtyEight, there are numerous polls in which the poll begins before the shutdown and ends after the shutdown ("A Polls"), as well as polls that began during the shutdown and ended after the government had reopened ("B Polls"). For the "A Polls," we counted each day that the government was shutdown during the polling period and used that for our value in the variable "shutdown." For any poll whose polling period completely resided within the government shutdown, we added the number of days polling was open and the number of days before polling opened that the government was shutdown for (as of the last day the poll was open). For the "B Polls," we took the number of days the poll was open after the government had reopened, and subtracted that from the total number of days that the government was shutdown, thirty-five.

The reason this method is valid is because the β_1 of "shutdown" and the β_1 of "post_shutdown" are extremely close to being mirrored effects. What this fact means is that in the few days after the government reopened, President Trump's adjusted approval rating picked back up at almost the exact same rate that his adjusted approval rating was dropping for the last few days of the shutdown. We made this adjustment because we did not want to neglect the polls that fell into the "B Polls" category, so this adjustment was made in order to include them in the paramount regression of "adjusted_approve" and "shutdown."

We also constructed an additional variable in order to measure the above β_1. We called this variable "post_shutdown", and it is simply the number of days the government had been reopened when the poll had ended. We created this variable to judge how President Trump's adjusted approval rating recovered after the government reopened. This variable also ties into our "B Polls" for the "shutdown" variable. For any "B Poll," the sum of "shutdown" and "post_shutdown" must be 35, the total number of days the government was shut down.

Another important caveat in our approach was that we did not want to assign tweets to days that fell into the "B Polls" because the value of our "shutdown" variable did not align with the days that the tweets were tweeted. The reason that they do not align is because of the adjustments we made to the "B Polls" data. This, unfortunately, means that these "B Polls" were not able to be taken into account when regressing "adjusted_approve" with "shutdown" and "tweets," but we thought this was valid to do because "tweets" is more of a secondary value, and the aforementioned regression of "adjusted_approve" and "shutdown" is the most fundamental regression to find the most practical results.

One other key independent variable we looked at was the effect of the stock market before, during, and after the government shutdown. We looked at stock market data during these periods and noticed no pronounced effect. In fact, the stock market overall was rising fairly consistently throughout these periods, with its lowest point in months coming directly before the shutdown on December 23rd with the market making an unrelated recovery ever since. Because of this, we thought it best not to include it as an independent variable when we ran our multivariable regression on variables that could have affected Trump's popularity.

While it was also not used, we created a variable entitled "dis_app," which is the adjusted approval rating subtracted from the adjusted disapproval rating. We were not able to find telling results with this variable as a measure of popularity, and we believe that it would double count some voters, so we disregarded it. The reason why we thought it would double count some voters is because when a person switched from approving to disapproving, the approval rating would go down and the disapproval rating would go up. This would cause the "dis_app" variable to change more than it should.

Sample data is provided on the following page:

1	president	subgroup	modeldate	startdate	enddate	pollster
4918	Donald Trump	Adults	2/12/19	12/14/18	12/20/18	Ipsos
4919	Donald Trump	Adults	2/12/19	12/15/18	12/21/18	Ipsos
4920	Donald Trump	Adults	2/12/19	12/16/18	12/22/18	Ipsos
4921	Donald Trump	Adults	2/12/19	12/17/18	12/20/18	American Resea

1	grade	samplesize	population	weight	influence	approve
4918	B+	2231	a	0.1530451	0	42.7
4919	B+	2237	a	0.1516252	0	42.7
4920	B+	2572	a	0.1722265	0	42
4921	C+	1100	a	1.148289	0	40

M	N	O	P	Q	R
disapprove	adjusted_approv	dis_app	adjusted_disappı	multiversions	tracking
52.7	42.17087	9.63415	51.80502		T
53.1	42.17087	10.03414	52.20501		T
53.2	41.47087	10.83415	52.30502		T
57	41.19181	13.53153	54.72334		

S	T	U	V	W	X	Y	Z
url	poll_id	question_id	createddate	timestamp	shutdown	post_shutdown	tweets
http://polling.reut	57178	92379	12/26/18	2/12/19 17:05	0		
http://polling.reut	57177	92378	12/26/18	2/12/19 17:05	0		
http://polling.reut	57176	92377	12/26/18	2/12/19 17:05	1		10
https://americanr	57161	92333	12/21/18	2/12/19 17:05	0		

4. Econometric Analysis

$$\widehat{Popularity} = 41.455 - .082 \times Shutdown$$

$$\widehat{Popularity} = 40.801 - .056 \times Shutdown + .005 \times Tweets$$

$$\widehat{Popularity} = 38.802 + .07 \times Post\_Shutdown$$

Above are our results for our econometric analysis where we have run multiple regressions seeing the effects of multiple variables on President Trump's popularity. In the first equation, we can see that the shutdown had a negative effect on the estimate of President Trump's popularity, with each day reducing his popularity by an average of 0.082 percentage points. The second equation involves our multivariable regression, where we can see the effects

of the length of the shutdown and tweets on President Trump's popularity. In this regression, we similarly see that the shutdown has a negative impact on his popularity, as his popularity estimate went down by an average of 0.056 percentage points each day, while, in an interesting finding, Trump's popularity estimate actually went up by 0.005 percentage points for each of his tweets. This fact is particularly notable as Trump had multiple days during the shutdown where he sent out 10+ tweets, including a day where he even sent out 40 tweets, which would have a decently-sized effect on his popularity estimate. Lastly, we can see that for every day post-shutdown, Trump's popularity estimate went up by an average of 0.07 percentage points, which closely mirrors the opposite of what happened in the first equation. Looking at the t-statistics for the independent variables, "shutdown" is clearly a huge factor in determining President Trump's adjusted approval rating because its t-statistic is almost -9. The other variables we looked at do not appear to be as statistically significant because they have t-statistics lower in magnitude than 1.96, but we determined this to be acceptable because they are secondary variables to the most important variable, "shutdown."

In addition to the regressions we ran on "shutdown," "tweets," and "post_shutdown," we also wanted to see how President Trump's shutdown compared to previous government shutdowns. For this regression, we gathered data from Gallup on former presidents. Once we had the data, we created a new "shutdown" variable in the form of a dummy variable. For this "shutdown" variable, 0 indicates that the government is not shutdown during the time of the poll, and 1 if the government is shutdown during the time of the poll. After running the regression, we came to the following conclusion:

$$\widehat{Popularity} = 50.482 - 4.191 \times Shutdown$$

After reaching the above equation, we compared President Trump's shutdown to this new equation defining past presidents' shutdowns. In order to do so, we multiplied President Trump's β_1 value from the regression of only "adjusted_approve" and "shutdown," -0.82, by the total number of days shutdown, 35. The purpose of doing so was to normalize our old continuous "shutdown" with the new dummy "shutdown" variable. Doing so gave us a value of -2.87. We immediately noticed that this value appears to be significantly lower than the -4.191 value from

previous presidents. The next step was to perform a hypothesis test at the 5% significance level to determine if this was indeed significant. Our null hypothesis was H_0: β_{1T}- β_{1P}=0, and our alternative hypothesis was H_1: β_{1T}- $\beta_{1P}\neq 0$. β_{1T} is the value given from President Trump's regression, -2.87. β_{1P} is the value given from the former presidents; regression, -4.191. We performed this hypothesis test by hand, using the standard method of hypothesis testing, and found a t-statistic of 0.9.

As we can see by this t-statistic, this is not statistically significant. When examining this result compared to the shutdowns of previous presidents, which lasted on average a much shorter time, they had more short term negative impact, while Donald Trump's had a more gradual negative impact. Because our t-statistic was not significant at this level, which means that the effects of Donald Trump's shutdown on his popularity are statistically similar to the shutdowns of previous previous presidents, and Donald Trump's shutdown was considerably longer, we conclude that Donald Trump has a more loyal fanbase than previous presidents.

5. Discussion of Key Assumptions

Since we are utilizing a multiple regression model, we have to take into account the corresponding four least squares assumptions. Before diving into more specifics, specifically if any of the assumptions are not satisfied in the empirical setting, the assumptions are listed below.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i,\ i = 1, \ldots n, \text{ where}$$

1. u_i has conditional mean zero given $X_{1i}, X_{2i}, \ldots X_{ki}$; that is,
   $$E(u_i \mid X_{1i}, X_{2i}, \ldots X_{ki}) = 0$$
2. $(X_{1i}, X_{2i}, \ldots X_{ki}, Y_i)$, i=1, …, n are identically and independently distributed (i.i.d.) draws from their joint distribution.
3. Large outliers are unlikely: $X_{1i}, X_{2i}, \ldots X_{ki}$ as well as Y_i have nonzero finite fourth moments.
4. There is no perfect multicollinearity.

To make our analysis valid, we need these four assumptions to hold. We will now analyze these assumptions and see if they hold for our multivariable regression model for President Trump's popularity.

Looking at the first assumption we can see that given our multivariable regression, which has President Trump's popularity as the dependent variable, then the length of the shutdown and tweets as the independent variable, the expectation of the error term given the independent variables will be zero, meaning that the factors that are affecting Trump's popularity (such as his tweets and the length of the shutdown) are not related. Now we can analyze assumption two. The second assumption is violated as the data is not independently and identically distributed for this project. We are utilizing time-series data where the polls are done over and over by the same pollsters over different periods of time. Despite assumption two not holding for our specific data, we believe our regression is still valid. We believe this as despite the lack of independent data for our dependent variable, that the polls that FiveThirtyEight utilizes still provides a good source of information for our regression. Since there are thousands of polls conducted that we use in our regression, we believe that despite assumption two not holding, that our resulting regression still provides valid data that we can use to compare the effects of the length of the shutdown and tweets on President Trump's popularity.

The third assumption of large outliers is not an issue with the pollster data. With simple random samples for the surveys and sample sizes that are not too small, the chance for an outlier is greatly decreased. Furthermore, since we are utilizing the adjusted approval ratings in our data, the polls that could have any bias or that are too extreme in their approval/disapproval ratings are automatically brought closer to the trend line, which means any bias or outliers in the original data are being mitigated by FiveThirtyEight's adjustments. Therefore, assumption three holds as there are no large outliers, which both we and FiveThirtyEight have looked for in the data. Similarly, we can say that the fourth moments of our variables are all finite and nonzero. Lastly, we look at our fourth assumption regarding perfect multicollinearity. Since tweets and the length of the government shutdown are not linear combinations of each other, then there is no perfect multicollinearity in our multivariable regression model.

In conclusion, we can see that three of the necessary four least squares assumptions for multivariable regressions hold. While only three of these four assumptions hold, we still believe, as mentioned previously, that our analysis is still valid.