Sara Bolf March 4, 2019
ECON 209

The Influence of the Government Shutdown on Approval Rate

Introduction

Recently, an inability of Donald Trump and the 115th United States Congress to agree on an appropriations bill for the 2019 fiscal year resulted in the shutdown of the federal government, lasting from December 22, 2018 to January 25, 2019. The main reason for this disagreement was that the appropriations bill did not include funding for the U.S.-Mexico Border Wall. The government shutdown was unpopular with a significant number of Americans, many feeling that the border wall was not important enough to justify the shutdown. Relatedly, the main purpose of this report is to further study the influence of the government shutdown on Trump's approval rate. Specifically, in this paper, I attempt to answer if the government shutdown had an effect on Trump's approval rate, and, if so, the extent of this effect. In order for my conclusion to be significant, I consider both the negative and positive effects. In other words, I take into account both the government shutdown and the end of the government shutdown as factors that may influence approval rate.

I decide on six variables to use in a regression analysis. These variables are approval rate, shutdown, post shutdown, number of days of the shutdown, number of days after the shutdown, and the stock market. This data is gathered from November 14, 2018 to February 18, 2019. The dependent variable is Trump's job approval rate, found by polling adult citizens. Each approval rate is computed by a certain pollster, and this polling occurs over a number of days. The independent variables are, thus, shutdown, post shutdown, number of days of the shutdown, number of days after the shutdown, and the stock market. Shutdown is a categorical variable, a 1 indicating that the time period in which a given poll was conducted is during the government shutdown, and a 0 indicating that this time period was computed either before or after the government shutdown. Post Shutdown is also a categorical variable, signifying if given dates fall after (1), or before (0), the government shutdown ended. The number of days of the shutdown is the number of days for which the shutdown has been going on, and is equal to zero if the given dates are before or after the shutdown. The number of days after the shutdown is the number of days after the shutdown ended. If the dates fall before the end of the shutdown, this variable is equal to 0. The stock market variable represents the Dow Jones Industrial Average daily index.

For this multiple linear regression analysis, my purpose is twofold. The first purpose is to find how well the model fits the data, and the second is to find how the government shutdown affects Trump's approval rate. To be more specific, the second purpose includes both the effect of the government shutdown and the effect of the government shutdown ending. These two effects are defined by two variables both. The effect of the government shutdown is defined in terms of the shutdown variable and the variable indicating the number of days of the shutdown, while the effect of the government shutdown ending is defined in terms of the post shutdown variable and the variable indicating the number of days after the shutdown. These four variables are the variables of interest for this analysis. The stock market variable is the control variable, as it is not directly relevant to the conclusion of this report, and is mainly included to control for any possible omitted variable bias.

To observe the fit of the model, I focus on the adjusted R-squared, the fitted vs. observed plot, and the 95% p-value for the F-statistic. These three statistics allow me to make a relatively sound conclusion on how well the data fits the model. To discuss the individual effects of the four variables of interest, I observe the coefficients of these variables. For example, the effect of the government shutdown on the first day of the shutdown is explained by $\beta_1 + \beta_3$, where β_1 is the effect of the government shutdown on the approval rate and β_3 is the effect of another day of the government shutdown on the approval rate. The effect of the end of the government shutdown on the first day after the government shutdown ended is similarly explained by $\beta_2 + \beta_4$. To find the significance of the effect of these variables on approval rate, I observe the 95% p-value for the F statistic for each variable, as well as the 95% confidence interval. Finally, to make sure that my analysis is valid, I discuss the key least squares assumptions, and whether they are satisfied in my model.

The results of the analysis imply that the model fits the data, but this fit is weak, with a low adjusted R-squared. With regards to individual effects of the variables, all of the variables of interests have significant effects on approval rate, as their p-values are all less than .05. However, the control variable does not have a significant effect on approval rate, instead having a p-value close to 1. The key least square assumptions are relatively satisfied, but there are some slight issues which may cause bias in my model.

I go into more detail about this analysis in the following sections. The next section focuses on my data, describing their sources and details, as well as providing summary statistics and descriptive graphs. The third section is an econometric analysis of the data. In this section, I complete an analysis of a linear least squares model in order to find how well the model fits the data, as well as the individual influence of the independent variables on approval rate. The fourth and final section emphasizes the four least squares assumptions, verifying that they are satisfied.

Data Description

As previously mentioned, the data set for this report includes six variables: approval rate, shutdown, post shutdown, number of days of shutdown, number of days after shutdown, and stock market. This data set was created from two different data sets, accounting for the variables approval rate and stock market, with the data for the variables shutdown, post shutdown, number of days of shutdown, and number of days after shutdown provided by me. This data ranges in date from November 14, 2018 to February 18, 2019. This date range was chosen as it encompasses around 3 months, one month before the shutdown, the shutdown, and one month after the shutdown. By doing this, I hope to lessen external effects on the approval rate.

The dependent variable, approval rate, describes the job approval rate for Donald Trump. This data was taken from *FiveThirtyEight's* Trump Approval Ratings dataset. This dataset originally has 22 variables, but only 4 were relevant to me. The first relevant variable was the approval rate, the dependent variable of my final model. The second relevant variable was the subgroup, which is the group of the population that was polled. I focused on polls with all adults as the polling group. The variables startdate and enddate were also relevant, as they gave the date the polling started and the date it ended, providing a date range for each approval rate. These two variables are important for calculating the other variables in the final data set, since a date range is required for their computation, as will be detailed later. Additionally, date ranges which included dates both during the shutdown and not during the shutdown were removed for simplicity. After making these alterations, I was left with 401 values for my approval rate variable.

The variable shutdown is a categorical variable, indicating whether the date range for a given approval rate was during the government shutdown (1) or not (0). This variable was computed based on the date ranges mentioned previously, and is equal to 0 if the date range falls before 12/22/18 or after 1/25/19, and equal to 1 if it falls between those two dates. The variable post shutdown is defined similarly, with a '1' representing that the date range for a given approval rate was after the government shutdown ended, and a '0' representing otherwise. Therefore, the variable is equal to 1 if the date range falls after 1/25/19, and 0 if if falls before.

In order to compute the data for the variable for the number of days of shutdown, it is necessary to take the average number of days for each date range, as the data set does not have singular dates. Therefore, as an example, if the date range is from 12/23/18 to 12/25/18, the number of days of the government shutdown would be equal to (2+4)/3 = 6. (assuming 12/22 would be day 1 of the shutdown, 12/23 day 2, and so on.) For any date range not between 12/22/18 and 1/25/19, the number of days of shutdown is equal to 0. Similarly, the variable for the number of days after the shutdown is computed using the average of days for each date range that occurs after 1/25/19. For the date ranges before 1/25/19, the number of days after the government shutdown is equal to 0.

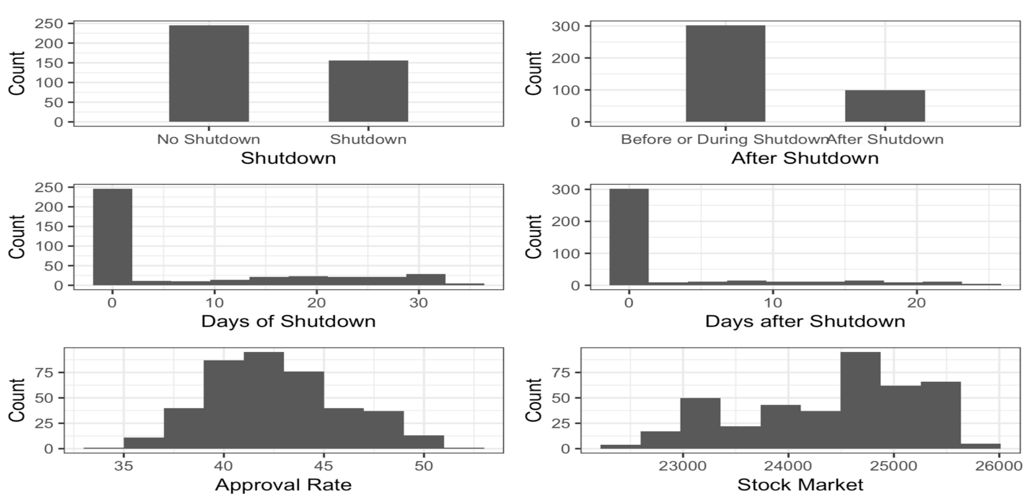
The final independent variable, stock market, is the Dow Jones Industrial (DJIA) daily index, provided by FRED Economic Data. The final data for this variable was again calculated using averages, but the method differs from the previous computations. In simple terms, I took the average of the DJIA daily indexes for the week before a given start date. So, if the start date for a date range was 12/10, I found the average of the DJIA indexes from 12/3 to 12/9. By taking DJIA data from the week prior to the polling instead of the week of, it is more likely that the DJIA index will actually have a relationship with Trump's approval rate. This is because the state of the economy reflected in the approval rate would most likely correlate to its state before the polls are conducted, rather than during. It is also important to note that the DJIA daily index does not include data for every day, so a week's worth of data actually only includes 4 or 5 days.

As can be seen in table 1, the stock market variable has a larger median than mean, implying that it is left skewed. The other five variables all seem to be right skewed, as their means are greater than their medians. However, the difference between the mean and median for the approval rate is not significantly large, which means that the approval rate may be normally distributed. The histograms and bar plots in figure 1 support these assumptions. For both the shutdown and post shutdown variables, there are significantly more '0' values than '1's', resulting in the data being skewed right. Similarly, for the number of days of the shutdown and the number of days after the shutdown, both variables are skewed right due to the large amount of '0's' present in the data. The approval rate is slightly right skewed, with a center around 43, which agrees with the mean and median being around 42.9. The stock market variable is left skewed, with most of the data being above 24,000. Notably, only one of the variables, approval rate, is close to being normally distributed, while the rest are significantly skewed. For the four variables of interest – shutdown, post shutdown, days of shutdown, and days after shutdown – this is because of the nature of the data. Both shutdown and post shutdown are categorical variables, and thus can only be equal to '0' or '1'. Since I gathered data over three months, and the shutdown and the period after the shutdown both only lasted around a month, both variables contain mostly '0's'. The days of shutdown and the days after shutdown are also affected by the time span of the data, and therefore most of the values are equal to '0'.

Table 1. Summary Statistics

	Approval Rate	Shutdown	After Shutdown	Number of Days of Shutdown	Number of Days after Shutdown	Stock Market
Min.	34.000	0.000	0.000	0.000	0.000	22493.90
1st Qu.	40.600	0.000	0.000	0.000	0.000	23839.16
Median	42.900	0.000	0.000	0.000	0.000	24558.43
Mean	42.961	0.389	0.247	7.852	3.183	24390.58
3rd Qu.	45.000	1.000	0.000	17.000	0.000	25126.54
Max.	52.000	1.000	1.000	34.500	24.500	25911.83

Figure 1. Histograms and Bar Plots of Variables.



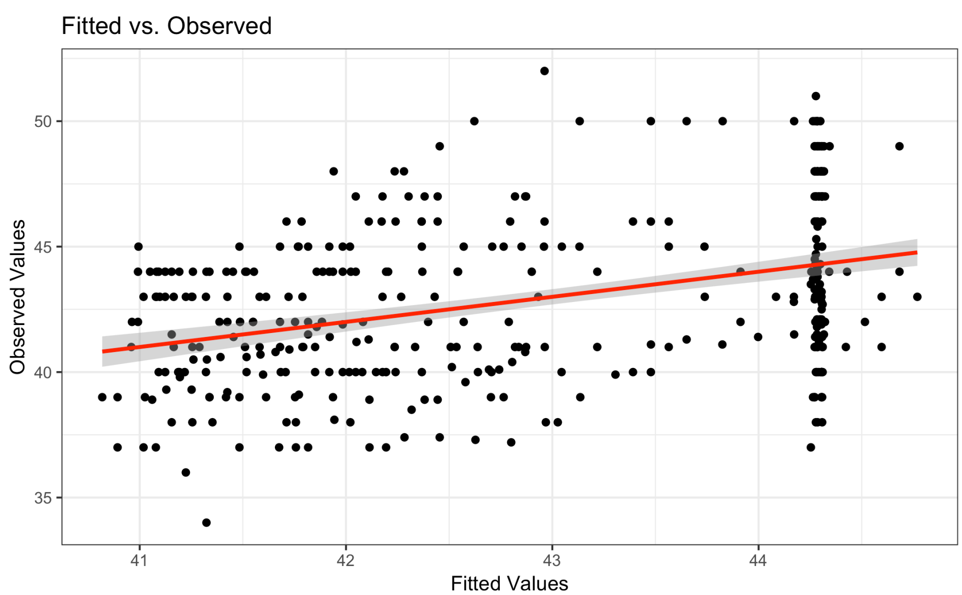
Econometric Analysis

Using this data, I fit a multivariate linear least squares model, defined as

$$\hat{y} = \mathbf{45.2318 - 1.2353x_1 - 3.7456x_2 - .06314x_3 + .17327x_4 - .0000378x_5}$$

where $\mathbf{x_1}$ = shutdown, $\mathbf{x_2}$ = post shutdown, $\mathbf{x_3}$ = days of shutdown, $\mathbf{x_4}$ = days after shutdown, and $\mathbf{x_5}$ = stock market. In order to identify how well my data fits this model, I first look at the adjusted R-squared value, or the coefficient of determination. The adjusted coefficient of determination is .1294, meaning that only 12.94% of the variation in approval rate is explained by the model. This implies that the model may not be the best fit for the data, as most of the variation in approval rate is not predictable from the independent variables. The 95% p-value for the F-statistic is highly significant, being equal to 1.263×10^{-11}. This means that the hypothesis that all the coefficients are equal to zero is rejected, signifying that at least one independent variable affects the approval rate. As can be seen in the fitted vs. observed plot in figure 2, even there is a positive linear relationship between the observed values and the predicted values, this relationship is very weak. Overall, although the model seems to fit the data, it is definitely not the best fit possible.

Figure 2. Fitted vs. Observed Plot



The effect of individual variables can be observed by looking at the 95% p-values of the F-statistics, presented in the ANOVA table below. All the variables of interest – that is, the shutdown, post shutdown, days of shutdown, and days after shutdown variables - have p-values of less than .05, and therefore are all statistically significant. Hence, I can conclude that all the variables of interest affect the approval rate, with the shutdown variable having the strongest effect and days of shutdown the weakest. The control variable, the stock market, is not statistically significant, with a p-value of .946. When looking at the effect of individual variables on the approval rate, I want to focus on the variables of interest, as they have a relevant effect on approval rate. The effect of the shutdown evaluated at a given day is defined as $-$ **1.2353 - .06314**$\mathbf{x_{3i}}$, where x_{3i} is the number of days of the shutdown on day i. Therefore, the effect of the shutdown on 12/22 is -1.29844, meaning that Trump's approval rate decreased by 1.3%. Correspondingly, the effect of the end of the shutdown evaluated at a given day is defined as $\mathbf{-3.7456_i + .17327x_{4i}}$, where x_{4i} is the number of days after the shutdown ended on day i. If the approval rate was measured on 1/26, it is expected to decrease by 2.01%. Notably, the effect of the end of the shutdown is less negative than the effect of the shutdown, which supports the assumption that the shutdown had an adverse effect on approval rate, and that approval rates started to increase after the shutdown ended. The effect of the variables of interest can also be seen in table 3, which presents the 95% confidence interval for each coefficient in the model.

Table 2. ANOVA

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Shutdown	1	324.5171002	324.5171002	33.5270735	0.0000000
After Shutdown	1	135.2888464	135.2888464	13.9771959	0.0002124
Days of Shutdown	1	49.3412441	49.3412441	5.0976282	0.0245038
Days after Shutdown	1	114.5736197	114.5736197	11.8370285	0.0006426
Stock Market	1	0.0442606	0.0442606	0.0045727	0.9461208
Residuals	395	3823.3058017	9.6792552	NA	NA

Table 3. 95% Confidence Intervals

	2.5 %	97.5 %
(Intercept)	17.877	72.587
Shutdown	-4.340	1.870
After Shutdown	-5.344	-2.147
Days of Shutdown	-0.148	0.022
Days after Shutdown	0.063	0.284
Stock Market	-0.001	0.001

Discussion of the Key Assumptions

There are four least squares assumptions for the regression model. The first is that the conditional mean of the error term given the included independent variables is zero. This is expressed as $E(u_i \mid X_{1i}, X_{2i}, \dots, X_{ki}) = 0$, where u_i is the error term, and k is the number of regressors in the model. The second assumption is that all variables in the model are independent and identically distributed (i.i.d.), or that $(X_{1i}, X_{2i}, \dots, X_{ki}, Yi)$ are all i.i.d. The third assumption states that large outliers are rare, and thus all variables have finite fourth moments. The final assumption is that there is not perfect multicollinearity, meaning that none of the regressors are linear functions of other regressors. In order for my prior analysis to be valid, these four assumptions need to be satisfied.

Since I used R's linear model (lm) function to create my model, the conditional mean of the error term equals zero, and the first assumption is satisfied. This is because the constant, 45.2318, forces the expectation of the error term to equal zero. This can be seen by taking the mean of the residuals, which equals 1.1756×10^{-16}. In a more theoretical sense, there is a possibility that omitted variable bias, which occurs due to the first assumption failing, is present in the model. I attempted to lessen this possibility by including the stock market as a control variable, but other factors still affect Trump's approval rate, such as scandals, legislation, and unemployment. I check the second assumption by looking at my data set, and concluding whether each variable is i.i.d. While all the variables do seem to have identical distribution, there may be a problem with how the data was collected with respect to the independence of the variables. For example, the variable stock market consists of the DJIA index recorded daily, and is therefore considered panel data. Hence, the data for the stock market variable may not be distributed independently. The approval rate variable could also have this problem, as it also consists of data measured over time. However, since the approval rates are recorded by a number of different pollsters, the possible dependence of this variable may be mitigated. This could affect my analysis, since if the data is not i.i.d., the estimators for the coefficients will not be the best linear unbiased estimators (BLUE). The third assumption can be confirmed by finding the fourth moment for each variable. I did this in R using the kurtosis function from the moments package. All the fourth moments were found to be finite, and thus I can conclude that my results are not affected by large outliers. The fourth assumption of no perfect multicollinearity is automatically satisfied, as any variable that is an exact linear function of another variable is removed by the statistical software when the regression in run. However, imperfect multicollinearity may still be present, which can be checked with Variance Inflation Factors (VIF). If a variable has a VIF greater than 10, then its multicollinearity is high. Table 3 presents the VIFs for each independent variable in the model. The VIF for the shutdown variable is greater than ten, which implies that the shutdown variable may be highly correlated with other independent variables. In general, none of the variables have noticeably small VIFs (5 is often used as a cutoff for multicollinearity also), which tells me that multicollinearity probably has a significant effect on my model. By looking at table 4, I can see which variables are strongly correlated with each other. Shutdown and days of shutdown are strongly correlated, as are after shutdown and days after shutdown. Interestingly, shutdown and stock market are also highly correlated. It is likely that one or more of the regression coefficients have been incorrectly estimated due to this multicollinearity. Overall, the key assumptions are generally satisfied, but there are obvious independence and multicollinearity issues in the data set, resulting in possible bias and inaccurate estimators.

Table 4. VIF

	VIF
Shutdown	24.567
After Shutdown	5.093
Days of Shutdown	9.739
Days after Shutdown	5.295
Stock Market	8.676

Table 5. Correlation

	Approval Rate	Shutdown	After Shutdown	Number of Days of Shutdown	Number of Days after Shutdown	Stock Market
approve	1.000	-0.270	-0.032	-0.288	0.051	0.209
sd	-0.270	1.000	-0.457	0.878	-0.399	-0.843
postsd	-0.032	-0.457	1.000	-0.401	0.872	0.416
sddays	-0.288	0.878	-0.401	1.000	-0.350	-0.559
daysafter	0.051	-0.399	0.872	-0.350	1.000	0.449
stockmarket	0.209	-0.843	0.416	-0.559	0.449	1.000

GitHub URL

- https://github.com/skb7/Econ209Project

References

- Rakich, N. (2019, January 04). The Public Blamed Trump For The Shutdown - But That May Be Changing. Retrieved from https://fivethirtyeight.com/features/the-public-blamed-trump-for-the-shutdown-but-that-may-be-changing/
- S&P Dow Jones Indices LLC, Dow Jones Industrial Average [DJIA], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DJIA.
- Silver, N. (2019, March 02) How Popular Is Donald Trump? Retrieved from https://projects.fivethirtyeight.com/trump-approval-ratings/
    - https://projects.fivethirtyeight.com/trump-approval-data/approval_polllist.csv