How does government shutdown affect Donald Trump's popularity?

Bohan Wu, Alejandro Torrebiarte, Eli Smith

Rice University, ECON 209

How does government shutdown affect Trump's popularity?

Introduction

In this research, we answer the following empirical question: How does the government shutdown affect Trump's popularity? The government shutdown of 2018-2019 lasts 35 days from Dec 22nd, 2018 to Jan 25th, 2019. We used poll data from FiveThirtyEight and ran regression on various independent variables in order to see their relationship with Trump's approval rating during the period. We used SLR and MLR to build regression models on the data we were given as well as outside data from dow jones. All of these models were constructed on RStudio. We also used diagnostics plots to check that assumptions were correct. We concluded that when we ran the single regressions, all of the independent variables were statistically significant, except for the stock market index, meaning that they did have a significant effect on Trump's approval rating. Also, when we ran a multilinear regression,we found that that the F-statistic was statistically significant.

link to github:

https://github.com/bohanwu424/Rice-ECON-209-Empirical-Project-Spring-2019.git

I. Data Description

In order to get the data that we used for our regression and econometrics analysis, we downloaded the data from the FiveThirtyEight.com poll accumulator on "How Popular Is Donald Trump?". The dataset contains 7400 observations of poll data from dozens of poll companies, but the ones who carried most weight were those from: YouGov, Ipsos, HarrisX, SurveyUSA, Pew Research Center and Marist College among others. Each poll is weighted differently based on its reliability and how accurate it is. They are also adjusted by a few percentage points to account for the quality, recency, sample size and partisan leans in the polls. However, we chose to only analyze only the data that came from polls provided by YouGov because it is the only one that is updated on a daily basis, and there wouldn't be that much of a significant difference between YouGov and other poll companies accounted for in FiveThirtyEight. The time frame that we decided to use was from 6/1/2018 to 2/20/2019.

Each poll has a starting date and an end date. To run the regression, we need to assign each poll to a single date. To do so, we create a new parameter "midpoint date" by calculating the average between starting date and end date of each poll. "midpoint date" is designated as the single-valued date of each poll.

The dependent or outcome variable that we are analyzing in this project is the adjust approval rate for Trump's presidency. The data for these results comes from the aggregated polls that are found on FiveThirtyEight.

There are four independent variables that we decided to analyze in order to see their effects on Trump's approval rate. The first one is a dummy variable for the government shutdown. The second one is the number of days **during** the government shutdown. The third is the number of days **after** the government shutdown. Finally, the last independent variable is the value from the Dow Jones index. However, the data for the Dow Jones index that we downloaded did not go as far back as the data from the poll, which means that there was a difference in the number of observations in dow jones data and yougov poll data. To adjust for this, we added some blank variables (NA) in those spaces for the Dow Jones index data in order for us to be able to run the SLR and a MLR on Trump's approval rate and the stock value of the Dow Jones index. In the data these blank spots are seen as "NA", which means that data wasn't available for those days. For more information/statistics on the data, refer to the first page of the code, where a complete summary of the data is available.

II. Econometrics Analysis

To understand the magnitude of the effect of government shutdown on Trump's approval rate, we first construct 4 SLR models using 4 different independent variables:x.dummie, x.sd,x.end,and djindex. Among the four, djindex is used as a control variable. After analyzing all the SLR models, we create a MLR model that incorporate all 4 independent variables.

First, let's look at each SLR model one by one.

SLR 1). Let x.dummy and approvalrate.dummy denote dummy variable shutdown and adjusted approval rate.[1]x.dummy equals 1 when there is a government shutdown, equals 0 otherwise. approval.rate.dummy will be the adjusted approval rates from 6/30/2018 to 2/20/2019. The SLR model for estimating approvalrate.dummy from x.dummie has the following specification:

[1] Without further specification, here approval rate by default means the adjusted approval rate provided by Yougov.

How does government shutdown affect Trump's popularity?

$$approvalrate.dummy = \beta_0 + \beta_1 \cdot x.dummie + \varepsilon$$

β_0 is the Trump's average approval rate when there is no government shutdown. β_1 is the true effect of government shutdown on Trump's approval rate. ε is the error term. We will use OLS estimators $\widehat{\beta_0}$ and $\widehat{\beta_1}$ to estimate β_0 and β_1. Here we assume three least square assumptions are satisfied. Later we will run diagnostics to determine whether our assumptions are valid.

The summary statistics of this model and the datapoint plot are provided below:

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   41.6210     0.1097 379.544   <2e-16 ***
## x.dummie      -0.5790     0.2752  -2.104   0.0364 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.597 on 250 degrees of freedom
## Multiple R-squared:  0.01739,	Adjusted R-squared:  0.01346
## F-statistic: 4.425 on 1 and 250 DF,  p-value: 0.03641
```

According to the summary statistics, $\widehat{\beta_1} = -0.579\%$. i.e. When there is a government shutdown, Trump's approval rate is expected to decrease by 0.579%. The hypothesis testing for the effect of government shutdown is $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$. The standard error of the effect $\widehat{\beta_1}$ is 0.2752. We can use $\widehat{SE}_{\beta_1}$ to estimate the standard deviation of $\widehat{\beta_1} : \sigma_{\widehat{\beta}_1}$. With t-statistics $t^{dummy} = \frac{\widehat{\beta}_1 - 0}{\widehat{SE}_{\beta_1}} = \frac{-0.579-0}{0.2752} = -2.104$,the p-value for the test is

$p - value = 2\Phi(-\left|t^{dummy}\right| = 2\Phi(-2.104) = 0.0364$.

Thus, the p-value is less than 0.05, so we can reject the null hypothesis that there is no effect of government shutdown on a 5% significance level.

When there is no government shutdown, the sample average of Trump's approval rate is $\widehat{\beta_0} = 41.621\%$. When there is a government shutdown, the sample average of Trump's approval rate is $\widehat{\beta_0} + \widehat{\beta_1} = 41.621\% + (-.579\%) = 41.042\%$

The low adjusted R-squared(0.01346) means that there exist other factors affect Trump's approval rate that are not explained by the model.Therefore, we construct SLR model (2) to analyze the relationship between Trump's approval rate and the length of government shutdown.

SLR 2). Let x.sd and approvalrate.sdd denote the number of days during the shutdown and the approval rates. x.sd is calculated by subtracting the date of each poll during the shutdown period(12/22/2018~01/25/2019) by the beginning date 12/22/2018. approval.rate.sdd will be the same adjusted approval rates as in the last model. The SLR model for estimating approval rates from x.sdd has the following specification:

$$approvalrate.sdd_i = \beta_{0,i} + \beta_{2,i} \cdot x.sd_i + \varepsilon_i$$

β_0 is the Trump's average approval rate when the length of shutdown=0. β_2 is the true effect of each additional day of shutdown on Trump's approval rate. ε is the error term. We will OLS estimators $\widehat{\beta}_0$ and $\widehat{\beta}_2$ to estimate β_0 and β_2. Samely, we assume three least square assumptions are satisfied. The summary statistics of this model are provided below:

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 41.64216    0.10565 394.165  < 2e-16 ***
## x.sd        -0.04363    0.01373  -3.178  0.00167 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.579 on 250 degrees of freedom
## Multiple R-squared:  0.03884,    Adjusted R-squared:  0.035
## F-statistic:  10.1 on 1 and 250 DF,  p-value: 0.001667
```

According to the summary statistics, $\widehat{\beta}_2 = -0.04363\%$. i.e. With each additional day of government shutdown, Trump's approval rate is expected to drop by 0.04363%. The hypothesis testing for the effect of length of government shutdown is $H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 \neq 0$. The standard error of the effect $\widehat{\beta}_2$ is 0.01373. We can use $\widehat{SE}_{\beta_2}$ to estimate the population standard deviation of $\widehat{\beta}_2$: $\sigma_{\widehat{\beta}_2}$. With t-statistics $t^{length} = \frac{\widehat{\beta}_2 - 0}{\widehat{SE}_{\beta_2}} = \frac{-0.04363-0}{0.01373} = -3.178$,the p-value for the test is $p - value = 2\Phi(-\left|t^{length}\right| = 2\Phi(-3.178) = 0.00167$.

Thus, the p-value is less than 0.01, so we can reject the null hypothesis that there is no effect of government shutdown on a 1% significance level.

When there is no government shutdown, the sample average of Trump's approval rate is $\widehat{\beta}_0 = 41.642\%$, which should be the same as $\widehat{\beta}_0$ in model 1. When shutdown lasts x days, the Trump's estimated approval rate is

$$\widehat{approvalrate.sdd}_i = \widehat{\beta}_0 + \widehat{\beta}_2 \cdot x = 41.642\% + (-0.04363\%) \cdot x$$

The low adjusted R-squared(0.03884) means that there exist other factors affect Trump's approval rate that are not explained by the model. Therefore, we construct SLR model (3) that analyzes the relationship between Trump's approval rate and the number of days after the government shutdown.

SLR 3). Let x.end and y.end denote the number of days during the shutdown and the approval rates. x.sd is calculated by subtracting the date of each poll after the shutdown period(01/24/2019~02/25/2019) by the beginning date 01/24/2019. y.end will be the same adjusted approval rates as in the last models. The SLR model for estimating approval rates from x.end has the following specification:

$$y.end_i = \beta_{0,i} + \beta_{3,,i} \cdot x.end_i + \varepsilon_i$$

β_0 is the Trump's average approval rate before the end of shutdown. β_3 is the true effect of each additional day after end of shutdown on Trump's approval rate. ε is the error term. We use OLS estimators $\widehat{\beta_0}$ and $\widehat{\beta_3}$ to estimate β_0 and β_3. Samely, we assume all four least square assumptions are satisfied. The summary statistics of this model are provided below:

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 41.46366    0.10541  393.37   <2e-16 ***
## x.end        0.04239    0.02038    2.08   0.0386 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.597 on 250 degrees of freedom
## Multiple R-squared:  0.017,  Adjusted R-squared:  0.01307
## F-statistic: 4.325 on 1 and 250 DF,  p-value: 0.03858
```

According to the summary statistics, $\widehat{\beta_3} = 0.04239\%$. i.e. With each additional day after the government shutdown, Trump's approval rate is expected to increase by 0.04239%. The hypothesis testing for the effect of length of government shutdown is $H_0 : \beta_3 = 0$ vs. $H_1 : \beta_3 \neq 0$. The standard error of the effect $\widehat{\beta_3}$ is 0.02038. We can use $\widehat{SE}_{\beta_3}$ to estimate the population standard deviation of $\widehat{\beta_3}$: $\sigma_{\widehat{\beta}_3}$. With t-statistics $t^{end} = \frac{\widehat{\beta}_3 - 0}{\widehat{SE}_{\beta_3}} = \frac{0.04239-0}{0.02038} = 2.08$,the p-value for the test is $p - value = 2\Phi(-\left|t^{length}\right| = 2\Phi(-2.08) = 0.0386$.

Thus, the p-value is less than 0.05, so we can reject the null hypothesis that there is no effect of government shutdown on a 5% significance level.

Before the end of government shutdown, the sample average of Trump's approval rate is

How does government shutdown affect Trump's popularity?

$\widehat{\beta}_0 = 41.464\%$. When shutdown ends for x days, the Trump's estimated approval rate is

$\widehat{approvalrate.sdd}_i = \widehat{\beta}_0 + \widehat{\beta}_3 \cdot x = 41.464\% + (0.04239\%) \cdot x$

The low adjusted R-squared(0.01307) means that there exist other factors affect Trump's approval rate that are not explained by the model. Therefore, we construct SLR model (4) to analyze the relationship between Trump's approval and economy's performance.

SLR 4). Let djindex denote the dow jones daily index . y.end will be the same adjusted approval rates as in the last models. The SLR model for estimating approval rates from djindex has the following specification:

$$y.end_i = \beta_{0,i} + \beta_{4,,i} \cdot djindex_i + \varepsilon_i$$

β_0 here carries no particular interest to our analysis. β_4 is the true effect of a per-unit increase in dow jones index in Trump's approval rate. ε is the error term. We use OLS estimators $\widehat{\beta}_0$ and $\widehat{\beta}_4$ to estimate β_0 and β_4 . Samely, we assume all four least square assumptions are satisfied. The summary statistics of this model are provided below:

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.321e+01  3.162e+00  13.666   <2e-16 ***
## djindex     -6.261e-05  1.257e-04  -0.498    0.619
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.602 on 188 degrees of freedom
##   (62 observations deleted due to missingness)
## Multiple R-squared:  0.001318,   Adjusted R-squared:  -0.003994
## F-statistic: 0.2481 on 1 and 188 DF,  p-value: 0.619
```

According to the summary statistics, $\widehat{\beta}_4 =- 6.2161 \cdot 10^{(-5)}$. i.e. With each unit increase in dow jones index, Trump's approval rate is expected to increase by $- 6.2161 \cdot 10^{(-5)}$. The hypothesis testing for the effect of length of government shutdown is $H_0 : \beta_4 = 0$ vs. $H_1 : \beta_4 \neq 0$. The standard error of the effect $\widehat{\beta}_4$ is $- 1.257 \cdot 10^{(-4)}$. We can use $\widehat{SE}_{\beta_4}$ to estimate the population standard deviation of $\widehat{\beta}_4$: $\sigma_{\widehat{\beta}_4}$. With t-statistics $t^{djindex} = \frac{\widehat{\beta}_4 - 0}{\widehat{SE}_{\beta_4}} = \frac{-6.2161 \cdot 10^{(-5)} - 0}{-1.257 \cdot 10^{(-4)}} =- 0.498$,the p-value for the test is $p - value = 2\Phi(-\left|t^{length}\right| = 2\Phi(- 0.498) = 0.619 >> 0.05$.

Thus, $\widehat{\beta}_4$ is not statistically significant. But since we are using dow jones index as a control variable, we can ignore $\widehat{\beta}_4$ and focus on whether or not the MLR including djindex will give us conditional mean independence of the error term with respect to the other three variables, i.e. $E[\varepsilon_i | x.dummy_i, x.sd_i, x.end_i, djindex_i] = E[\varepsilon_i | djindex_i]$

Finally we construct the MLR model using the following specification:

$Adj\ Approval\ Rate_i = \beta_{0,i} + \beta_{1,i} \cdot x.dummie_i + \beta_{2,i} \cdot x.sd_i + \beta_{3,i} \cdot x.end_i + \beta_{4,i} \cdot djindex_i + \varepsilon_i$

$\beta_{0,i}$ carries no particular analytical interest. $\beta_{1,i}$ is the true effect of government shutdown or not on Trump's approval rate, ceteris paribus. $\beta_{2,i}$ is the true effect of of each additional shutdown day on Trump's approval rate, ceteris paribus. $\beta_{3,i}$ is the true effect of of each additional day after the shutdown on Trump's approval rate, ceteris paribus. $\beta_{4,i}$ is the true effect of of a per unit increase in dow jones index on Trump's approval rate, ceteris paribus. We use OLS estimators to estimate each of the $\beta's$. Assuming the four least square assumptions hold, the summary statistics are provided below:

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 48.0721486  4.2547611  11.298   <2e-16 ***
## x.dummie    -0.2408390  0.8070364  -0.298    0.766
## x.sd        -0.0403601  0.0332131  -1.215    0.226
## x.end        0.0362826  0.0245805   1.476    0.142
## djindex     -0.0002527  0.0001677  -1.507    0.134
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.568 on 185 degrees of freedom
##   (62 observations deleted due to missingness)
## Multiple R-squared:  0.05892,    Adjusted R-squared:  0.03857
## F-statistic: 2.896 on 4 and 185 DF,  p-value: 0.02344
```

Individually,the hypothesis testing for each of the β_j is $H_0 : \beta_j = 0$ vs. $H_1 : \beta_j \neq 0$ where $j \in \{1, 2, 3, 4\}$.

The mean of estimators $\left[\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3, \widehat{\beta}_4\right] = [48.072, -0.24, -0.04, 0.03, -0.0002527]$. The standard error for $\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3, \widehat{\beta}_4$ is 4.25, 0.8,0.0332,0.0246,0.001677. From the mean and standard errors, we can derive the p-values for each estimator. It turns out all individual p-values are larger than 5%. So none of the estimators is statistically significant from 0.

However, the failures to show statistical significance of individual estimators does not imply that our model has no predicting power. The hypothesis testing that all the slopes are 0 is $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ vs. $H_1 : \beta_1 \neq 0\ or\ \beta_2 \neq 0\ or\ \beta_3 \neq 0\ or\ \beta_4 \neq 0$. Because our F-statistics=0.02344<0.05, H_0 is rejected on a 5% significance level, which means our overall model has some predicting power.

III. Discussion of the key assumptions[2]

In summary,We conduct four univariate regressions using only yougov data and included a multivariate regression that accounted for the stock market changes during our observation times. For our method of univariate analysis, there are three required main assumptions prevent any bias in our conclusions. The first Least Squares assumption is that the conditional distribution of our error term, given X, has a mean of zero. $E[\varepsilon|X] = 0$ .The first assumption is particularly important in a real-world scenario such as our empirical project because it is impossible to account for every single factor that occurs in real life which could have an effect on our dependent variable. One issue that can arise if our data does not fit this assumption is the possibility of Omitted-Variable bias. To test whether we have omitted variables that are influencing our dependent variable, we can look at Residual vs Fitted plot. The mean of our Residuals remains zero when the fitted value increases, so it does not exhibit any patterns that might indicate OVB. However, the Q-Q plot of all our SLR models are piecewise, rather than 45 degree line like we should expect if $\varepsilon_i \sim N(0, \sigma^2{}_i)$. Thus our error term is not normally distributed, but the first assumption still holds because $E(\varepsilon_i|x_i) = 0$ .However, the Q-Q plot for our SLR models show no significant deviation from a straight line. Thus in MLR model, the error term follows a normal distribution $\varepsilon_i \sim N(0, \sigma^2{}_i)$.

The second Least Squares assumption is much more easily incorporated into real-world. This assumption states that (Xi, Yi) i=1, ..., n are independently and identically distributed (i.i.d.). We can be sure that our Trump approval data is i.i.d. easily due to the method with which that data was collected. The polling sites that we used (e.g.yougov) are reputable and renowned for their random sampling methods, which makes a guarantee to researchers that the data can be treated as i.i.d. Our stock market analysis, however, cannot be assumed to be i.i.d. Time series data, such as stock market levels, has the key feature that observations falling close to each other cannot be considered independent, and as such will violate the i.i.d. assumption. This can introduce slight

[2] For this section, please refer to the diagnostic plots in the attached R code

bias in that model's predictions, as high or low stock values on one day will heavily influence the next day's performance, rather than truly independent data.

The third Least Squares assumption is that large outliers are unlikely. Without this assumption, the regression and predictions will be heavily influenced by a few data points and can produce inaccurate or misleading results. To verify that our dataset has no large outliers, we conducted an Outlier-Leverage plot, which helps us to visualize which points could potentially be bad outliers. The rule for checking leverage point is the leverage $h_{ii} > 4/n$. In our case with n=252, the rule is $h_{ii} > 0.159$. The rule for checking an outlier is $|r_i| > 2$ where r_i is the standardized residual. A bad outlier is a point that is both an outlier and a leverage point. This plot reveals that we have a few outliers and a few leverage points but no bad leverage point, so our model can be assumed to have no large outliers that could violate the third assumption of Least Squares.

The fourth Least Squares assumption only applies to our multivariate plot, and it states that there must be no perfect multicollinearity. Our statistical software can automatically alert us to the situations of perfect multicollinearity, and no perfect multicollinearity was alerted.