Problem Set 3

QTM 200: Applied Regression Analysis

Due: February 17, 2020

Instructions

- Please show your work! You may lose points by simply writing in the answer. If the problem requires you to execute commands in `R`, please include the code you used to get your answers. Please also include the `.R` file that contains your code. If you are not sure if work needs to be shown for a particular problem, please ask.
- Your homework should be submitted electronically on the course GitHub page in `.pdf` form.
- This problem set is due at the beginning of class on Monday, February 17, 2020. No late assignments will be accepted.
- Total available points for this homework is 100.

In this problem set, you will run several regressions and create an add variable plot (see the lecture slides) in `R` using the `incumbents_subset.csv` dataset. Include all of your code.

Question 1 (20 points)

We are interested in knowing how the difference in campaign spending between incumbent and challenger affects the incumbent's vote share.

1. Run a regression where the outcome variable is `voteshare` and the explanatory variable is `difflog`.

```
#Q1
#import dataset
i <- read.csv("~/Documents/GitHub/QTM200Spring2020/problem_sets/PS3/
    incumbents_subset.csv")
summary(i)
#Build linear regression model
lm1 <- lm(i$difflog~i$voteshare, data=i)
summary(lm1)
```

2. Make a scatterplot of the two variables and add the regression line.

```
#Make a scatterplot and add the regression line
plot(i$difflog~i$voteshare, main="voteshare vs difflog",xlab = "difflog",
    ylab = "voteshare")
abline(lm1)
```

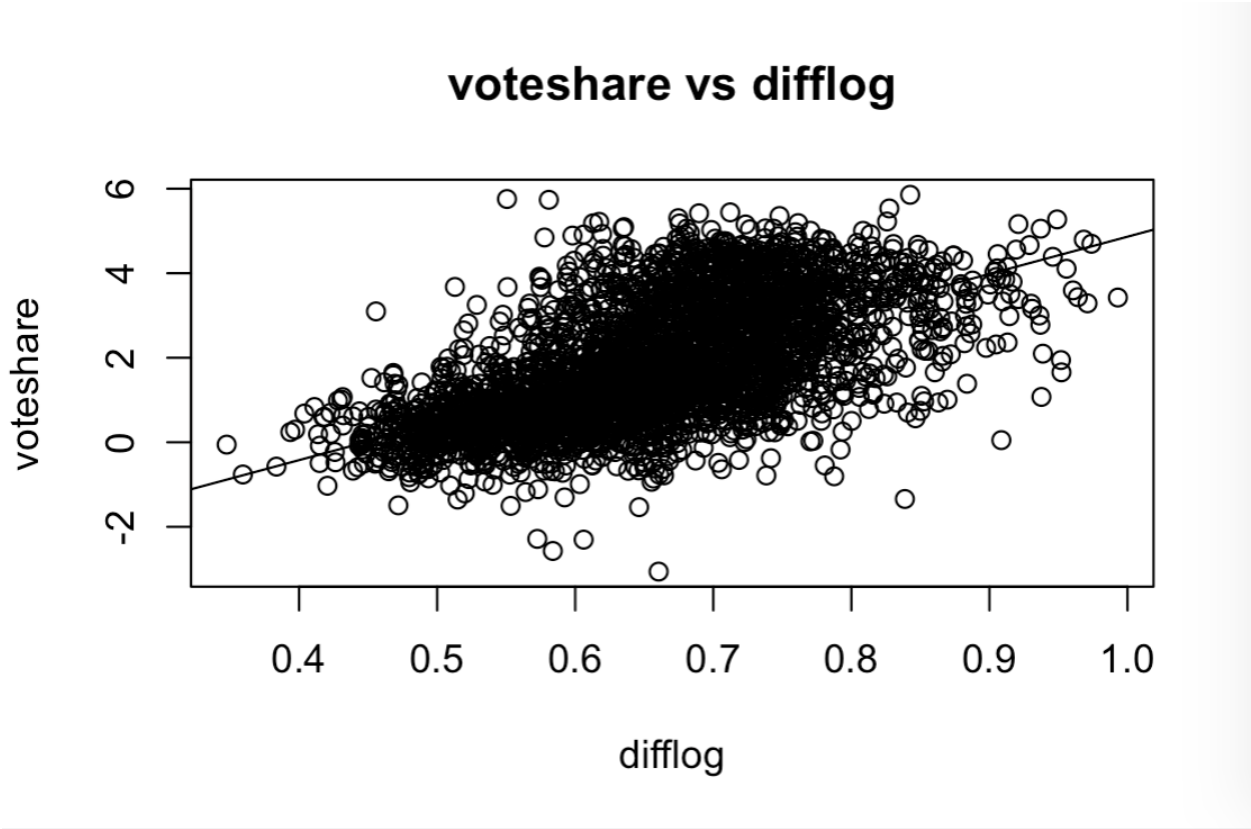

Figure 1: voteshare vs difflog.

3. Save the residuals of the model in a separate object.

```
#Save the residuals of the model in a separate object
re1 <- residuals(lm1)
```

4. Write the prediction equation.

```
#Write the prediction equation
#Coefficients:
#Estimate Std. Error t value Pr(>|t|)
#(Intercept) 0.579031   0.002251  257.19   <2e-16 ***
#difflog     0.041666   0.000968   43.04   <2e-16 ***
#Residual standard error: 0.07867 on 3191 degrees of freedom
#Multiple R-squared:  0.3673, Adjusted R-squared:  0.3671
#F-statistic:  1853 on 1 and 3191 DF,  p-value: < 2.2e-16
#y = 0.579 + 0.042x1 + sigma1 (Where epsilon is 0.079^2)
```

Question 2 (20 points)

We are interested in knowing how the difference between incumbent and challenger's spending and the vote share of the presidential candidate of the incumbent's party are related.

1. Run a regression where the outcome variable is `presvote` and the explanatory variable is `difflog`.

```
#Q2
#Build linear regression model
lm2 <- lm(i$presvote~i$difflog)
```

2. Make a scatterplot of the two variables and add the regression line.

```
#Make a scatterplot and add the regression line
plot(i$presvote~i$difflog, main = "presvote vs. difflog", xlab = "difflog
    ", ylab = "prevote")
abline(lm2)
```

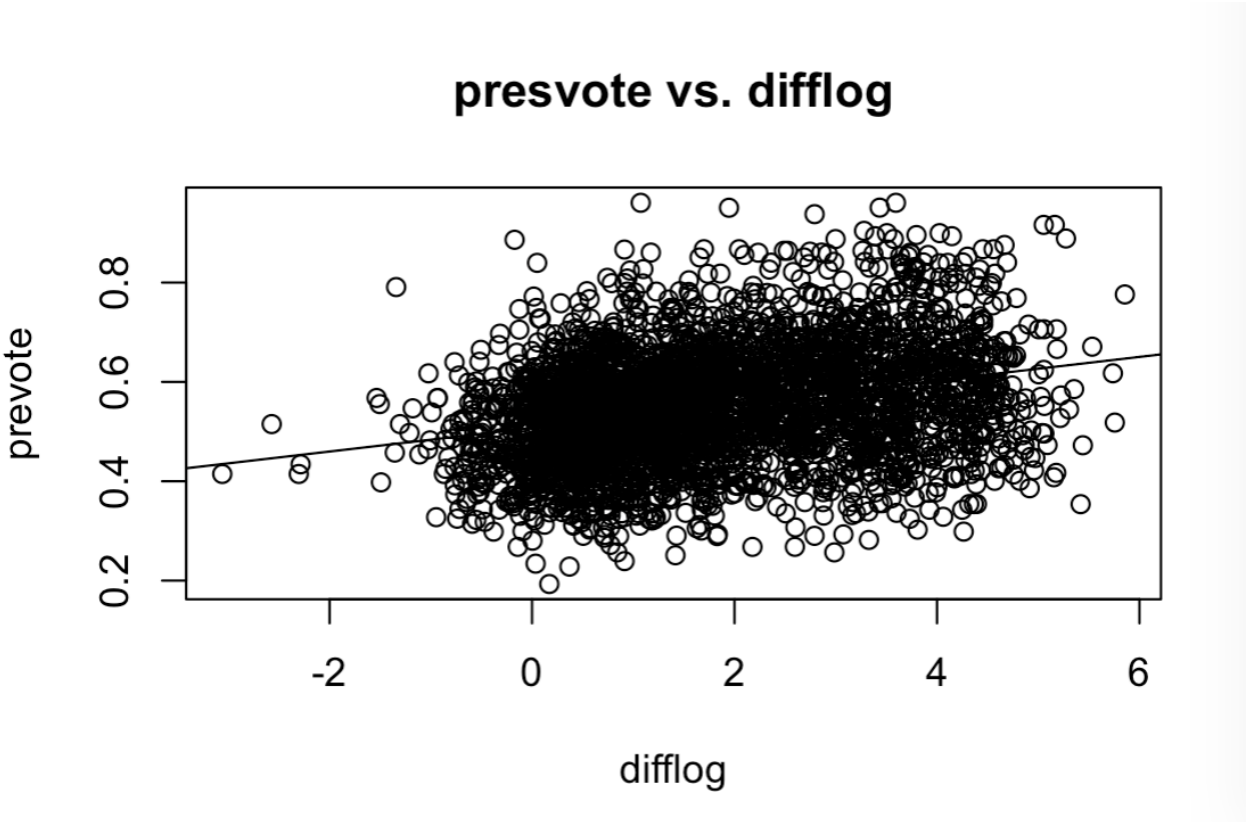

Figure 2: prevote vs difflog.

3. Save the residuals of the model in a separate object.

```r
#Save the residuals of the model in a separate object
re2 <- residuals(lm2)
```

4. Write the prediction equation.

```r
#Write the prediction equation
#Coefficients:
#Estimate Std. Error t value Pr(>|t|)
#(Intercept) 0.507583   0.003161  160.60   <2e-16 ***
#difflog     0.023837   0.001359   17.54   <2e-16 ***
#Residual standard error: 0.1104 on 3191 degrees of freedom
#Multiple R-squared:  0.08795,  Adjusted R-squared:  0.08767
#F-statistic: 307.7 on 1 and 3191 DF,  p-value: < 2.2e-16
#y = 0.508 + 0.024x1 + sigma1 (Where epsilon is 0.110^2)
```

Question 3 (20 points)

We are interested in knowing how the vote share of the presidential candidate of the incumbent's party is associated with the incumbent's electoral success.

1. Run a regression where the outcome variable is `voteshare` and the explanatory variable is `presvote`.

```
#Q3
#Build linear regression model
lm3 <- lm(i$voteshare~i$presvote)
```

2. Make a scatterplot of the two variables and add the regression line.

```
#Make a scatterplot and add the regression line
plot(i$voteshare~i$presvote, main = "voteshare vs presvote ", xlab = "
    presvote", ylab = "voteshare")
abline(lm3)
```

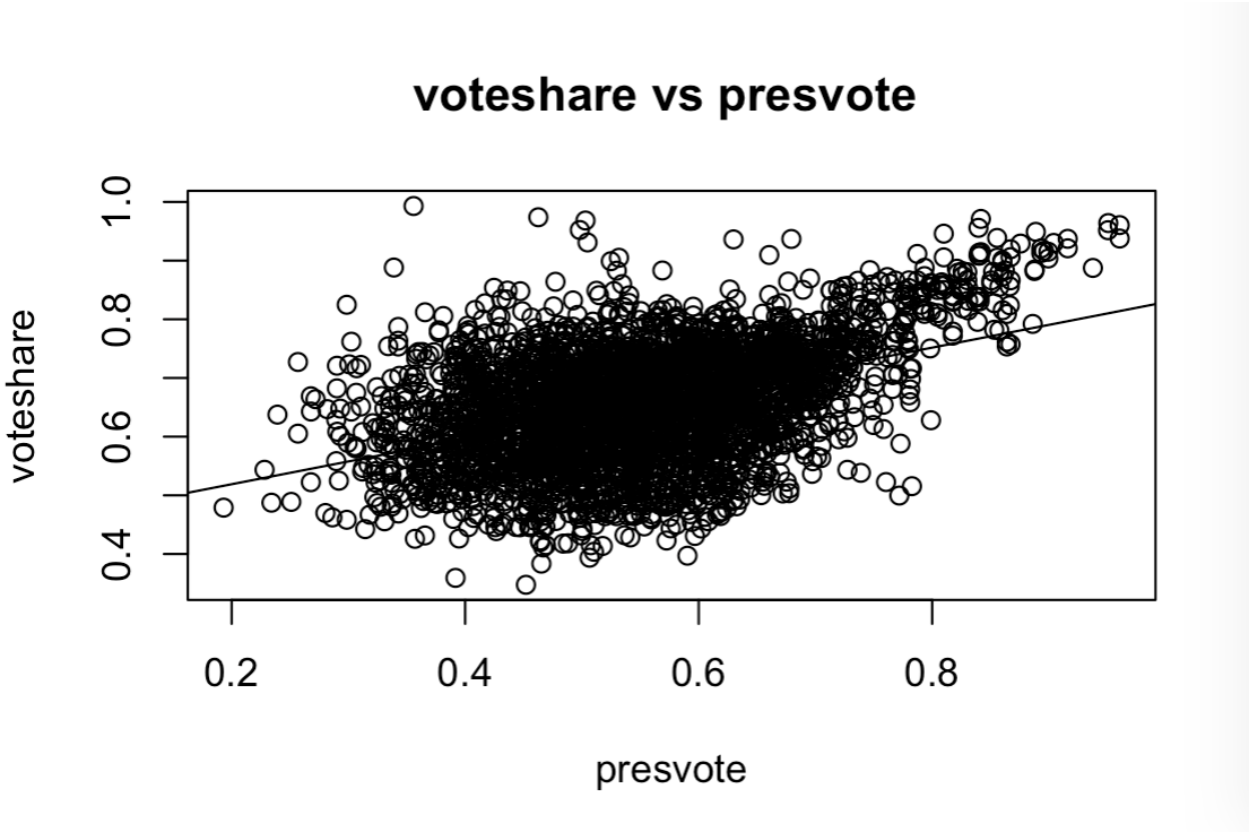

Figure 3: voteshare vs prevote.

3. Write the prediction equation.

```
#Write the prediction equation
#Coefficients:
#Estimate Std. Error t value Pr(>|t|)
#(Intercept) 0.441330   0.007599   58.08   <2e-16 ***
#prevote     0.388018   0.013493   28.76   <2e-16 ***
#Residual standard error: 0.08815 on 3191 degrees of freedom
#Multiple R-squared:  0.2058, Adjusted R-squared:  0.2056
#F-statistic:   827 on 1 and 3191 DF,  p-value: < 2.2e-16
#y = 0.441 + 0.388x1 + sigma1 (Where epsilon is 0.088^2)
```

Question 4 (20 points)

The residuals from part (a) tell us how much of the variation in `voteshare` is *not* explained by the difference in spending between incumbent and challenger. The residuals in part (b) tell us how much of the variation in `presvote` is *not* explained by the difference in spending between incumbent and challenger in the district.

1. Run a regression where the outcome variable is the residuals from Question 1 and the explanatory variable is the residuals from Question 2.

```
#Q4
#Build linear regression model
lm4 <- lm(re1~re2)
summary(lm4)
```

2. Make a scatterplot of the two residuals and add the regression line.

```
#Make a scatterplot and add the regression line
plot(re1~re2, main = "re1 vs re2", xlab = "re1", ylab = "re2")
abline(lm4)
```

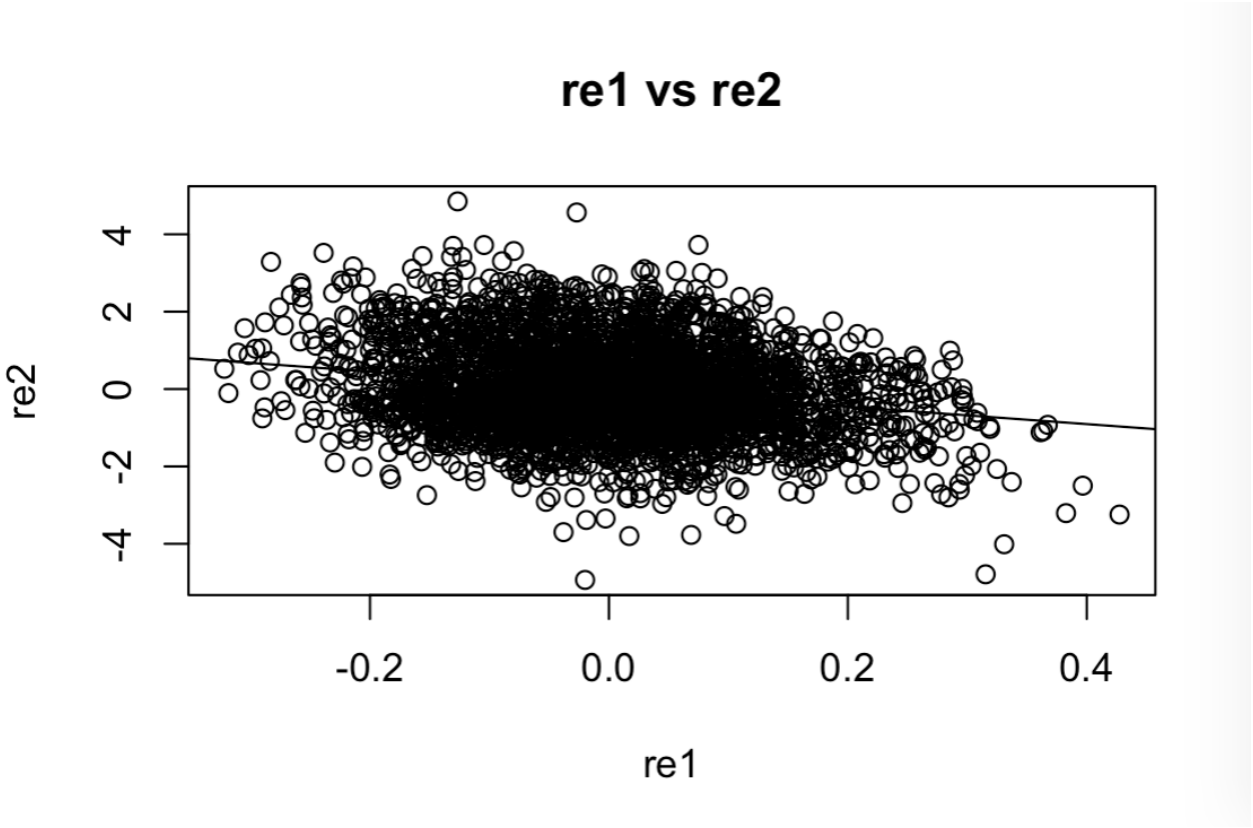

Figure 4: re1 vs re2.

3. Write the prediction equation.

```
#Write the prediction equation
#Coefficients:
#Estimate Std. Error t value Pr(>|t|)
#(Intercept) -4.860e-18  1.299e-03    0.00        1
#re2          2.569e-01  1.176e-02   21.84   <2e-16 ***
#Residual standard error: 0.07338 on 3191 degrees of freedom
#Multiple R-squared:   0.13,  Adjusted R-squared:  0.1298
#F-statistic:   477 on 1 and 3191 DF,  p-value: < 2.2e-16
#y = -4.860e-18 + 2.569e-01x1 + sigma1 (Where epsilon is 0.073^2)
```

Question 5 (20 points)

What if the incumbent's vote share is affected by both the president's popularity and the difference in spending between incumbent and challenger?

1. Run a regression where the outcome variable is the incumbent's `voteshare` and the explanatory variables are `difflog` and `presvote`.

```
#Q5
#Build linear regression model
Xa <- i$difflog
Xb <- i$presvote
Y <- i$voteshare
lm5 <- lm(Y~Xa+Xb)
```

2. Write the prediction equation.

```
#Write the prediction equation
summary(lm5)
#Coefficients:
#Estimate Std. Error t value Pr(>|t|)
#(Intercept) 0.4486442  0.0063297   70.88   <2e-16 ***
#Xa          0.0355431  0.0009455   37.59   <2e-16 ***
#Xb          0.2568770  0.0117637   21.84   <2e-16 ***
#Residual standard error: 0.07339 on 3190 degrees of freedom
#Multiple R-squared:  0.4496, Adjusted R-squared:  0.4493
#F-statistic:  1303 on 2 and 3190 DF,  p-value: < 2.2e-16
#y = 0.449 + 0.0355x1 + 0.257x2 + sigma1 (Where epsilon is 0.073^2)
```

3. What is it in this output that is identical to the output in Question 4? Why do you think this is the case?