PLSC 30600: Problem Set 3

Solutions

February 16, 2023

This problem set is due at **11:59 pm on Tuesday, February 28th**.

Please upload your solutions as a .pdf file saved as "Yourlastname_Yourfirstinitial_pset3.pdf". In addition, an electronic copy of your .Rmd file (saved as "Yourlastname_Yourfirstinitial_pset3.Rmd") must be submitted to the course website at the same time. We should be able to run your code without error messages. In addition to your solutions, please submit an annotated version of this `.rmd` file saved as "Yourlastname_Yourfirstinitial_pset3_feedback.rmd" and a corresponding PDF saved as "Yourlastname_Yourfirstinitial_pset3_feedback.pdf" noting the problems where you needed to consult the solutions and why along with any remaining questions or concerns about the material. In order to receive credit, homework submissions must be substantially started and all work must be shown. Late assignments will not be accepted. In total your submissions should consist of four files.

Problem 1

This problem will have you replicate and analyze the results from Moser and Voena's 2012 AER paper on the impact of the World War I "Trading with the Enemy Act" on U.S. domestic invention. The full citation is below

> Moser, P., & Voena, A. (2012). Compulsory licensing: Evidence from the trading with the enemy act. American Economic Review, 102(1), 396-427.

The premise of the study is to evaluate the effect that "compulsory licensing" policy – that is, policies that permit domestic firms to violate foreign patents and produce foreign inventions without needing to obtain a license from the owner of the foreign patent – have on domestic invention. Does access to foreign inventions make domestic firms more innovative? The authors leverage an exogenous event in U.S. licensing policy that arose from World War I – the 1917 "Trading with the Enemy Act" (TWEA) which permitted U.S. firms to violate patents owned by enemy-country firms. This had the consequence of effectively licensing all patents from German-owned firms to U.S. firms after 1918 (that is, from 1919 onward), allowing them to produce these inventions without paying for a license from the German-owned company.

The authors look specifically at domestic innovation and patent activity in the organic chemicals sector. They note that only some of the sub-classes of organic chemicals (as defined by the US Patent Office) received any compulsory licenses under the Trading with the Enemy Act while others did not. They leverage this variation in exposure to the "treatment" of compulsory licensing to implement a differences-in-differences design looking at domestic firm patent activity in each of these sub-classes (comparing sub-classes that were exposed to compulsory licensing to those that were unexposed).

The dataset is `chem_patents_maindataset.dta` – the code below will load it.

```
library(tidyverse)
# Read in the Moser and Voena (2012) dataset
chem <- haven::read_dta("chem_patents_maindataset.dta")
```

The unit of the dataset is the sub-class/year (471,120 observations) of 7248 US Patent and Trademark Office (USPTO) patent sub-classes over 65 years.

The relevant variables are

- `uspto_class` - USPTO Patent Sub-Class (unit)
- `grntyr` - Year of observation (year)
- `count_usa` - Count of patents granted to US-owned firms in the year
- `count_france` - Count of patents granted to French-owned firms in the year
- `count_for` - Count of patents granted to foreign-owned (non-US) firms in the year
- `treat` - Treatment indicator – Whether the patent sub-class received any German patents under TWEA (after 1918 when the policy went into effect) (Note that this is not an indicator for the overall treatment *group* (whether the unit *ever* received treatment) – it is only 1 after 1918 for units that receive treatment but is still 0 for those "treated" units prior to the initiation of treatment)

Part A

If you try to use a two-way fixed effects estimator on the dataset as it is, it will likely freeze up your computer as this is a *very large* dataset. We'll instead first aggregate the data in a way that will let you use a simple difference-in-differences estimator to estimate the treatment effect.

Generate a point estimate for the average treatment effect of receiving treatment on the average annual count of US patents using a difference-in-differences estimator (using all post-treatment (1919-1939) and pre-treatment (1875-1918) time periods). You should aggregate your data such that the outcome is the post-/pre-difference in the outcome (preferably using `tidyverse` functions like `group_by` and `summarize`) and each row is a USPTO patent sub-class (rather than a sub-class/year observation) and use a difference-in-means estimator with the differenced outcome. Again, if you use `lm_robust` or even `lm` with two-way fixed effects, your computer will likely freeze up as there are many FE parameters to estimate.

Provide a 95% robust confidence interval and interpret your point estimate. Do we reject the null of no treatment effect at the $\alpha = .05$ level?

First, aggregate the data

```
library(estimatr)

## Group by patent class, calculate average in pre-treatment and post-treatment periods, create an indi
chem_agg <- chem %>% group_by(uspto_class) %>% summarize(treated = max(treat), count_usa_post = mean(cou
                                                        count_usa_pre = mean(count_usa[grntyr >=1875 &
                                                        count_fra_pre = mean(count_france[grntyr >=1875
                                                        count_fra_post = mean(count_france[grntyr >= 19
                                                        count_for_pre = mean(count_for[grntyr >= 1875 &
                                                        count_for_post = mean(count_for[grntyr >= 1919


## Difference in outcomes
chem_agg$count_usa_diff <- chem_agg$count_usa_post - chem_agg$count_usa_pre

## DID estimator
lm_robust(count_usa_diff ~ treated, data=chem_agg)

##              Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)    0.386     0.0101   38.15 2.55e-290    0.366    0.406 7246
## treated        0.255     0.0377    6.77  1.35e-11    0.181    0.329 7246
```

2

We estimate that compulsory licensing raised the average number of patents granted to US firms by .25 per year (95% CI: [.18, .33]). We would reject the null of no ATE at $\alpha = .05$

Part B

A colleague suggests that you should instead just compare the average differences in the count of US patents in the post-1918 period between exposed and unexposed sub-classes to estimate the treatment effect. Based on what we observe in the pre-1919 period, is ignorability of the treatment likely to hold under this strategy? Discuss why or why not – what do you observe in the patent counts in the pre-treatment period between exposed and unexposed subclasses.

```
lm_robust(count_usa_pre ~ treated, data=chem_agg)
```

```
##             Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)    0.228    0.00685   33.26 7.10e-226    0.214    0.241 7246
## treated       -0.145    0.01887   -7.69  1.63e-14   -0.182   -0.108 7246
```

No, we see a difference in pre-treatment patent counts in the exposed vs. unexposed sub-classes - exposed sub-classes have, on average *fewer* patents in the pre-treatment period, suggesting that the sub-classes where compulsory licensing took place were substantively different from the ones where it did not (less existing innovation). Therefore ignorability is likely violated.

Part C

The authors implement a test of their identification assumptions by also estimating the effect (using the differences-in-differences design) of the Trading with the Enemy Act on patents granted by French firms, which the authors note "could not license enemy patents under the TWEA." Describe what sort of a diagnostic strategy this is. What do the authors expect to find if their parallel trends assumption holds?

Estimate the effect of TWEA exposure on the count of French firm patents using a difference-in-differences design and provide a a 95% robust confidence interval. Are the results consistent with what the authors expect if their design assumptions hold?

This is a placebo test. The authors expect to find no effect on French patenting if their design assumptions hold. They in fact find no effect on French patents in these sectors using the DiD estimator.

```
chem_agg$count_fra_diff <- chem_agg$count_fra_post - chem_agg$count_fra_pre
lm_robust(count_fra_diff ~ treated, data=chem_agg)
```

```
##             Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)  0.00194   0.000479   4.047 5.25e-05  0.00100  0.00288 7246
## treated     -0.00203   0.003852  -0.527 5.98e-01 -0.00958  0.00552 7246
```

They estimate that the effect of compulsory licensing on French patenting is -0.00203 with a 95% CI of $[-0.00958, 0.00552]$. They would fail to reject the null at $\alpha = .05$. This relatively precise null estimate, this is consistent with what the authors would expect under their design assumptions.

Part D

We might be concerned that there are differential trends in pre-treatment patenting between those sub-classes exposed to the treatment and those exposed to control. Estimate the difference in the trend in US patents between exposed and unexposed sub-classes from 1918 to 1917, 1916, 1915, and 1914 (four estimates in total: 1918-1917, 1918-1916, 1918-1915, 1918-1914). Provide a 95% robust confidence interval for each of these estimates and interpret your results. Do we reject the null that any of these differ from 0 (at $\alpha = .05$)? If

the outcome trends were evolving in parallel between the treated and control groups, what would we expect these estimates to be? What do your results suggest for the validity of the parallel trends assumption?

If the outcome trends were evolving in parallel between the treated and control groups, we would expect these estimates to be 0 as these are all pre-treatment "placebo" periods.

First we'll calculate average counts in each of the time periods

```
chem_parallel <- chem %>% group_by(uspto_class) %>% summarize(treated = max(treat), count_usa_1918 = mea
                                          count_usa_1917 = mean(count_usa[grntyr == 1917]),
                                          count_usa_1916 = mean(count_usa[grntyr == 1916]),
                                          count_usa_1915 = mean(count_usa[grntyr == 1915]),
                                          count_usa_1914 = mean(count_usa[grntyr == 1914]))
```

For the 1918 v. 1917 placebo:

```
chem_parallel$count_usa_1918_1917 <- chem_parallel$count_usa_1918 - chem_parallel$count_usa_1917
lm_robust(count_usa_1918_1917 ~ treated, data=chem_parallel)

##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)   -0.033     0.0121  -2.733   0.0063  -0.0566 -0.00932 7246
## treated        0.027     0.0446   0.605   0.5449  -0.0605  0.11456 7246
```

We obtain a diff-in-diff estimate of 0.027 with a 95% CI of $[-0.0605, 0.11456]$

For the 1918 v. 1916 placebo

```
chem_parallel$count_usa_1918_1916 <- chem_parallel$count_usa_1918 - chem_parallel$count_usa_1916
lm_robust(count_usa_1918_1916 ~ treated, data=chem_parallel)

##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)  -0.0488     0.0136   -3.58 0.000343  -0.0754  -0.0221 7246
## treated       0.0964     0.0368    2.62 0.008765   0.0243   0.1684 7246
```

We obtain a diff-in-diff estimate of 0.0964 with a 95% CI of $[0.0243, 0.1684]$

For the 1918 v. 1915 placebo

```
chem_parallel$count_usa_1918_1915 <- chem_parallel$count_usa_1918 - chem_parallel$count_usa_1915
lm_robust(count_usa_1918_1915 ~ treated, data=chem_parallel)

##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper   DF
## (Intercept) -0.00405     0.0134  -0.303   0.7621  -0.0303   0.0222 7246
## treated      0.06357     0.0344   1.850   0.0644  -0.0038   0.1310 7246
```

We obtain a diff-in-diff estimate of 0.06357 with a 95% CI of $[-0.0038, 0.1310]$

For the 1918 v. 1914 placebo

```
chem_parallel$count_usa_1918_1914 <- chem_parallel$count_usa_1918 - chem_parallel$count_usa_1914
lm_robust(count_usa_1918_1914 ~ treated, data=chem_parallel)

##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)   0.0505     0.0135   3.727 0.000195   0.0239   0.0770 7246
## treated      -0.0237     0.0395  -0.601 0.548183  -0.1011   0.0537 7246
```

We find some potential evidence of a parallel trends violation in the WWI period. Notably we reject the null that the diff-in-diff between 1918 and 1916 is 0 at $\alpha = .05$ - if our parallel trends assumption were to hold across all time periods, then this should be 0.

Part E

The authors adjust for covariates out of concern for possible parallel trends violations. One possible confounder that might be driving a parallel trends violation is the overall amount of foreign patenting in the sub-class and its change over time – reflecting general technological differences that might differ between the patent sub-classes. Since the treatment does not affect the amount of foreign patenting, this is a valid control.

Create a variable for the change between the post- and pre-treatment count of foreign patents in the USPTO subclass. Use the Abadie (2005) weighting estimator to estimate the ATT under a conditional parallel trends assumption (fit a model for the propensity score using this covariate). Do we reject the null of no treatment effect at the $\alpha = .05$ level? Compare your results to your estimate from Question A and discuss why they might differ.

Let's first calculate the change in foreign patenting.

```
# Change in foreign patenting
chem_agg$count_for_diff <- chem_agg$count_for_post - chem_agg$count_for_pre

# Histogram
hist(chem_agg$count_for_diff)
```

Histogram of chem_agg$count_for_diff
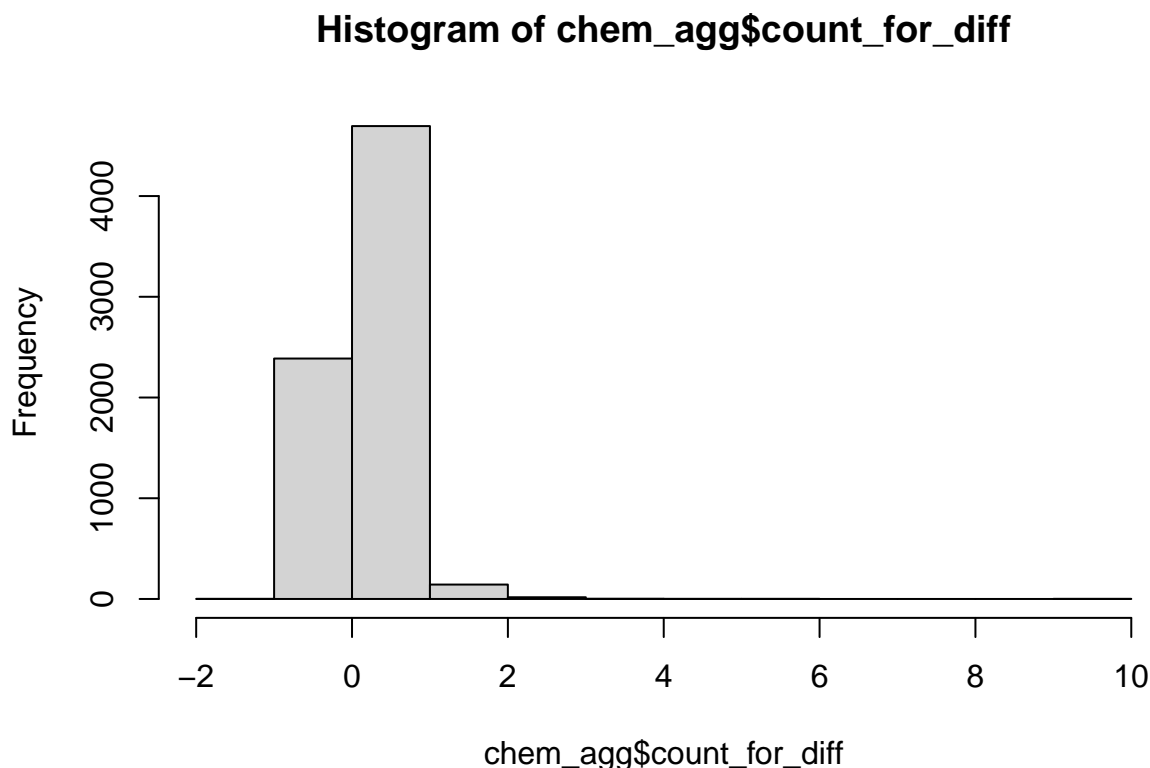
Now, taking a look at the distribution of this covariate, it looks like we might have some *very* notable outliers. Supposing that we fit our propensity score model with just treatment and outcome (assuming linearity in the link function), we'll end up with the following:

```
# Fit a propensity score model that breaks down terribly!
pscore_model  <- glm(treated ~ count_for_diff, data=chem_agg, family=binomial(link="logit"))

# Predict weights
chem_agg$e <- predict(pscore_model, type="response")
chem_agg$did_wt <- (1/mean(chem_agg$treated)) * ((chem_agg$treated - chem_agg$e)/(1-chem_agg$e))
```
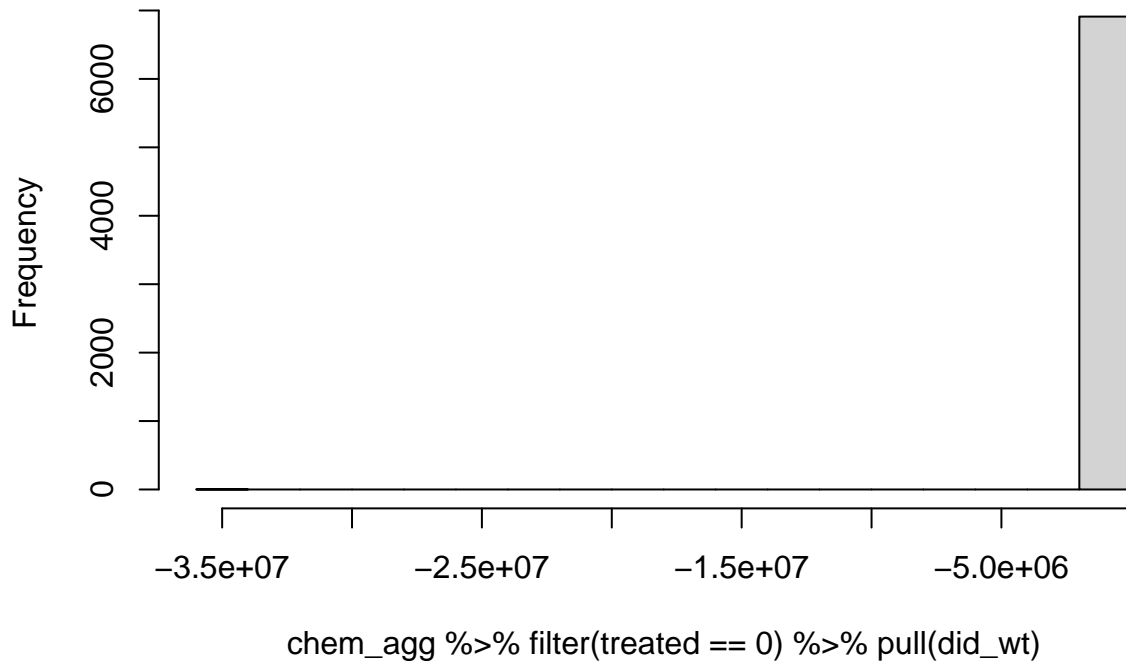
```
# Point estimate
mean(chem_agg$count_usa_diff*chem_agg$did_wt)
```

```
## [1] -26281
```

```
# Histogram of the weights
hist(chem_agg %>% filter(treated == 0) %>% pull(did_wt))
```

Histogram of chem_agg %>% filter(treated == 0) %>% pull(did_wt)


```
set.seed(60637)
niter <- 1000
boot_est <- rep(NA, niter)
for(i in 1:niter){
  boot_chem_agg <- chem_agg[sample(1:nrow(chem_agg), nrow(chem_agg), replace=T),]
  # Fit a propensity score model
  weight_model_boot <- glm(treated ~ count_for_diff, data=boot_chem_agg, family=binomial(link="logit"))

  boot_chem_agg$e <- predict(weight_model_boot, type="response")
  boot_chem_agg$did_wt <- (1/mean(boot_chem_agg$treated)) * ((boot_chem_agg$treated - boot_chem_agg$e)/

  # Point est
  boot_est[i] <- mean(boot_chem_agg$count_usa_diff*boot_chem_agg$did_wt)
}
#Bootstrap 95\% CI
quantile(boot_est, c(.025, .975))
```

```
##      2.5%      97.5%
## -2.29e+05 -1.89e-01
```

Our 95% confidence interval is kind of absurd: $[-229,000, -0.189]$. Perhaps we should change the treatment-covariate model?

This part is a bonus note on extreme weights as an important issue to be aware of

Something is going wrong with this estimator (crazy extreme negative value) - this seems to be a computation issue that's being driven by an extreme weight on one of the control observations (around 3.5×10^7)! This is a circumstance under which our IPTW estimators are going to behave in a rather crazy manner. This is a big red flag and suggests we should re-specify the propensity score model. Let's instead bin our variable and estimate a coefficient for each bin. We'll make 6 bins using the quantile function (`cut2` in `Hmisc` does quantile binning).

```r
# Make bins instead of just using the continuous variable
chem_agg$count_for_diff_bin <- Hmisc::cut2(chem_agg$count_for_diff, g=6)

# Fit a propensity score model that breaks down terribly!
pscore_model  <- glm(treated ~ as.factor(count_for_diff_bin), data=chem_agg, family=binomial(link="logi

# Predict weights
chem_agg$e <- predict(pscore_model, type="response")
chem_agg$did_wt <- (1/mean(chem_agg$treated)) * ((chem_agg$treated - chem_agg$e)/(1-chem_agg$e))

# Point estimate
mean(chem_agg$count_usa_diff*chem_agg$did_wt)
```
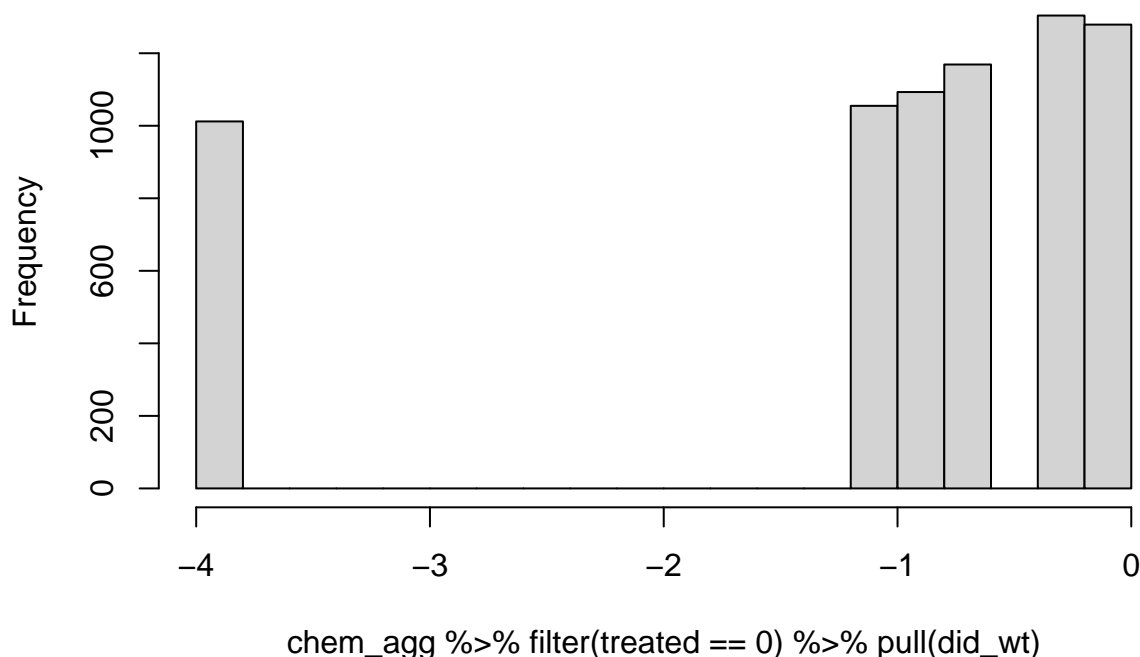
```
## [1] -0.0744
```

```r
# Histogram of the weights
hist(chem_agg %>% filter(treated == 0) %>% pull(did_wt))
```


Histogram of chem_agg %>% filter(treated == 0) %>% pull(did_wt)

We obtain a much more stable point estimate of $-.0744$ and as the histogram of the weights shows, there are no longer absurdly extreme values for the weights.

Bootstrapping for the confidence interval:

```
set.seed(60637)
niter <- 1000
boot_est <- rep(NA, niter)
for(i in 1:niter){
  boot_chem_agg <- chem_agg[sample(1:nrow(chem_agg), nrow(chem_agg), replace=T),]
  # Fit a propensity score model
  weight_model_boot <- glm(treated ~ as.factor(count_for_diff_bin), data=boot_chem_agg, family=binomial

  boot_chem_agg$e <- predict(weight_model_boot, type="response")
  boot_chem_agg$did_wt <- (1/mean(boot_chem_agg$treated)) * ((boot_chem_agg$treated - boot_chem_agg$e)/

  # Point est
  boot_est[i] <- mean(boot_chem_agg$count_usa_diff*boot_chem_agg$did_wt)
}
#Bootstrap 95\% CI
quantile(boot_est, c(.025, .975))
```

```
##    2.5%   97.5%
## -0.1506  0.0119
```

In this case, our confidence interval contains 0 and we would fail to reject the null of no ATT. Adjusting for the covariate (and assuming conditional parallel trends) appears to have eliminated the positive treatment effect estimate. Those sub-classes that were affected by compulsory licensing may have been those sorts of sub-classes that would generally experience an upward trend in patenting post-war.

Problem 2

Consider the standard instrumental variables set-up in the treatment non-compliance setting, Z_i denotes the assignment to treatment, D_i is the actual treatment that is taken and Y_i is the outcome of interest. Assume that all of the instrumental variables assumptions hold (treatment is ignorable, exclusion restriction, non-zero first stage, monotonicity/no-defiers). The specific example you'll examine in this problem is the JOBS II randomized trial which evaluated the effect of a job training program on reducing depression among those who recently experienced job loss (among other health outcomes). This specific dataset on 502 "high-risk" individuals from this experiment comes from:

> Little, Roderick J., and Linda HY Yau. "Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model." Psychological Methods 3.2 (1998): 147.

The code below will load the data into R

```
## Load the JOBS II dataset
jobs2 <- read_table2("wjobs.tab", na= ".")
## Make the treatment and instrument variable
jobs2$Z <- jobs2$Tx
jobs2$D <- as.integer(jobs2$Tx == 1 & jobs2$c1 == 1)
```

The relevant columns are

- `Z` - Assignment to the job training program
- `D` - Actual participation in job training
- `depress` - Change in depression from baseline (higher values = more depression)
- `risk` - Pre-treatment mental health risk score
- `educ` - Number of years of education completed
- `age` - Age
- `single` - Marital status: single

Part A

Examine the data and explain why monotonicity **is guaranteed** to hold for this particular design.

Compliance is one-sided. No units that receive assignment to control can take the treatment.

```
jobs2 %>% group_by(Z,D) %>% summarize(n())
```

```
## # A tibble: 3 x 3
## # Groups:   Z [2]
##       Z     D `n()`
##   <dbl> <int> <int>
## 1     0     0   167
## 2     1     0   152
## 3     1     1   183
```

Therefore there are only two possible compliance groups/principal strata - the never-takers and the compliers. The always-takers do not exist since all units assigned control take control and no defiers can exist either since it is impossible for a unit to take treatment when assigned control. Therefore monotonicity holds by default - either the instrument has a zero individual causal effect on treatment uptake (the never-takers) or it has a positive individual causal effect (the compliers).

Part B

Using the Wald estimator, estimate the local average treatment effect of participation in the job training program among the compliers. Assuming asymptotic normality, provide a 95% confidence interval and conduct a hypothesis test for the null of no LATE at $\alpha = .05$.

The Wald estimator with a binary instrument is equivalent to the 2SLS estimator with no additional covariates. We can calculate an estimate and CI using standard IV routines.

```
iv_robust(depress ~ D | Z, data=jobs2)
```

```
##             Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
## (Intercept)   -0.383     0.0623  -6.145 1.63e-09   -0.506   -0.261 500
## D             -0.136     0.1372  -0.994 3.21e-01   -0.406    0.133 500
```

We estimate that participating in the job training program, among the compliers, reduced depression scores by -0.136 points with a 95% CI of $[-0.406, 0.133]$. We would fail to reject the null of no LATE at $\alpha = .05$.

Part C

Estimate the first stage effect of assignment to training on participation in the program. Would you consider assignment to the job program to be a strong instrument for participation?

The first stage is a regression of taking the treatment (participation) on assignment to the program.

```
lm_robust(D ~ Z, data=jobs2)
```

```
##             Estimate Std. Error t value Pr(>|t|)  CI Lower  CI Upper  DF
## (Intercept) -3.22e-15   2.08e-16   -15.5 1.54e-44 -3.63e-15 -2.81e-15 500
## Z            5.46e-01   2.72e-02    20.1 4.50e-66  4.93e-01  6.00e-01 500
```

We estimate that assignment to treatment raises participation rates by .546 (95 percent CI: [.493, .600]). Assignment to treatment is a *very* strong instrument – by conventional first stage F-statistic diagnostics,

we obtain an F-statistic around $20.1^2 = 404$, far above commonly used thresholds for determining weak vs. strong instruments (even above 100). We would expect that the bias in two-stage least-squares even under our design assumptions will be very minimal as a result.

Note that the typical first-stage F-statistic diagnostics use the F-statistic/t-test under homoskedasticity, which in this case would be 200 - still an extremely strong instrument.

```
summary(lm(D~ Z, data=jobs2))
```

```
##
## Call:
## lm(formula = D ~ Z, data = jobs2)
##
## Residuals:
##    Min    1Q Median    3Q    Max
## -0.546 -0.546  0.000  0.454  0.454
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.14e-15   3.15e-02     0.0        1
## Z            5.46e-01   3.86e-02    14.2   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.408 on 500 degrees of freedom
## Multiple R-squared:  0.286,  Adjusted R-squared:  0.285
## F-statistic:  200 on 1 and 500 DF,  p-value: <2e-16
```

Part D

Which subset of your sample consists exclusively of never-takers? Suppose you were to take the mean of one of your covariates among this group – explain why this would identify the mean of the covariate among the never-takers.

All observations with $Z_i = 1$ and $D_i = 0$ are never-takers. Because compliance is one-sided, we know that any unit that is assigned to receive control will not take treatment. Therefore the observations assigned to take treatment that nevertheless take control *must* take control if assigned to control and therefore would be under control under either value of the instrument.

This one-sided non-compliance implies:

$$E[X_i|D_i = 0, Z_i = 1] = E[X_i|D_i(1) = D_i(0) = 0, Z_i = 1]$$

Because Z_i is independent of X_i and the potential outcomes $D_i(1), D_i(0)$, we can further write

$$E[X_i|D_i = 0, Z_i = 1] = E[X_i|D_i(1) = D_i(0) = 0]$$

Therefore the covariate mean among units with $D_i = 0$ and $Z_i = 1$ identifies the covariate mean among the never-takers under one-sided non-compliance.

Part E

Using the group you identified in D, examine the pre-treatment covariate averages among the never-treated respondents. Compare these to the overall means in the sample. Discuss whether there are any notable differences and what this would imply for the representativeness of the LATE you identified in Part B.

The thing we want to know is whether the never-takers are meaningfully different from the sample as a whole (and by extension whether they differ from the compliers who are the only other group in the data). If the covariate profiles of the compliance groups differ substantially, then that could imply that the LATE is not representative of the true full-sample ATE if there exists effect heterogeneity along these covariates.

We'll look at the four pre-treatment covariates `risk`, `educ`, `age`, and `single`

In the full sample

```
jobs2 %>% summarize(mean(risk), mean(educ), mean(age), mean(single))
```

```
## # A tibble: 1 x 4
##    `mean(risk)` `mean(educ)` `mean(age)` `mean(single)`
##           <dbl>        <dbl>       <dbl>          <dbl>
## 1          1.68         13.4        36.6          0.622
```

Among the assigned-treated take-control (only never-takers)

```
jobs2 %>% filter(D==0&Z==1) %>% summarize(mean(risk), mean(educ), mean(age), mean(single))
```

```
## # A tibble: 1 x 4
##    `mean(risk)` `mean(educ)` `mean(age)` `mean(single)`
##           <dbl>        <dbl>       <dbl>          <dbl>
## 1          1.68         12.9        33.3          0.632
```

On the four covariates, we do not see noticeable differences between the never-takers and the sample as a whole - these groups have roughly the same average risk score average education. About 63 percent are single. The never-takers appear to be slightly younger (33 years vs. 36 years) though this is a relatively small deviation. Just looking at the covariate means for these pre-treatment covariates, we don't find considerable evidence that the never-takers and compliers are particularly different on covariates that might moderate the treatment effect. Therefore, we don't find much to suggest the LATE is likely to significantly deviate from the sample ATE.

Problem 3

In this problem you will analyze the data from David Broockman's paper examining whether congressional incumbents help boost the vote shares for their party's presidential candidate in their respective congressional districts. The citation is here

> Broockman, David E. (2009) "Do Congressional Candidates Have Reverse Coattails? Evidence from a Regression Discontinuity Design," Political Analysis 17(4): 18–434.

Below is the code to import the dataset into R

```
### Load in the Broockman (2009) dataset
congress <- read_csv("congress.csv")
```

This paper is interested in examining whether there exist "reverse coattails" in US elections. The "coattail" effect refers to the observation that popular presidential candidates tend to boost the performance of their party's candidates in down-ballot races (e.g. U.S. House and Senate races). This paper investigates whether the *reverse* also might exist – whether congressional incumbents are able to boost the performance of co-partisan Presidential candidates in their districts.

The study analyzes the effect of congressional incumbency on district Presidential vote share using a Regression Discontinuity Design. The dataset consists of 8994 observations of congressional districts over time. For each observation, we observe the Democratic party congressional candidate's vote share margin of victory in the election at time t along with the Democratic party congressional candidate's vote share margin of victory in time $t + 1$ and the Democratic party presidential candidate's vote share margin of victory in time $t + 1$.

Candidates who have a margin of victory greater than 0 in time t go on to be the incumbent in time $t+1$, so the discontinuity in treatment assignment (incumbency) is at 0.

- `dv_c_t1` - Democratic candidate *congressional* vote-share margin of victory in time t
- `dv_c_t2` - Democratic candidate *congressional* vote-share margin of victory in time $t+1$
- `dv_p_t2` - Democratic candidate *presidential* vote-share margin of victory in time $t+1$
- `t2_is_midterm` - Whether the election at time $t+1$ is a midterm election (no Presidential election)

Part A

In words, what are the main assumptions that we would need to make in order to identify the causal effect of Democratic congressional incumbency in time $t+1$ on Democratic presidential vote share in time $t+1$ using a regression discontinuity design?

We know that all incumbents at time $t+1$ won their election in time t, so treatment is assigned via a discontinuity in the running variable of democratic margin of victory at time t. We must assume that, around the discontinuity of 0 margin-of-victory, the assignment of incumbency essentially approximates a locally-randomized experiment. The potential outcomes are assumed to be continuous around the discontinuity such that any "jump" in the observed outcome can be attributed to the change in treatment status induced by the discontinuity.

Part B

The study first replicates the incumbency findings in Lee (2008). Considering only years with Presidential elections (no midterm elections), estimate the effect of democratic incumbency on democratic party *congressional* vote share margin of victory in time $t+1$ using a local-linear RD estimator with a bandwidth of .25. Provide a 95% confidence interval and interpret your estimate.

```
# Indicator for democratic party victory
congress$dwin <- as.integer(congress$dv_c_t1 > 0)

# Run the regression on non-midterm elections w/in .25 of the cut-off
lm_robust(dv_c_t2 ~ dwin + dv_c_t1 + dwin*dv_c_t1,
          data= congress %>% filter(t2_is_midterm == 0&abs(dv_c_t1) < .25))
```

```
##              Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)   -0.0504    0.00422  -11.95  3.03e-32  -0.0587  -0.0422 3299
## dwin           0.0912    0.00633   14.39  1.36e-45   0.0787   0.1036 3299
## dv_c_t1        0.7324    0.03655   20.04  2.26e-84   0.6607   0.8040 3299
## dwin:dv_c_t1   0.1292    0.05227    2.47  1.35e-02   0.0267   0.2317 3299
```

We estimate that the effect of democratic incumbency on democratic congressional vote share in non-midterm elections is 0.0912 with a 95% confidence interval of $[0.0787, 0.1036]$. We would reject the null of no effect at the 0.05 level. This replicates the original finding for the incumbency effect in Lee (2008) – just barely winning a close election in time t has a strong effect on that party's vote share in time $t+1$.

Part C

Create a binned scatterplot for your analysis in Part B and overlay the local linear regressions above and below the discontinuity. Does the linear model seem like a reasonable approximation for the conditional expectation function above and below the discontinuity for the original bandwidth choice (.25)?

```
bin_scatter_reg <- ggplot(aes(x=dv_c_t1, y=dv_c_t2),
                          data=subset(congress, t2_is_midterm == 0&abs(dv_c_t1) <.25)) +
  stat_summary_bin(fun='mean', bins=50,
                   size=2, geom='point') +
  geom_vline(xintercept=0, col="red", lty=2) +
  geom_smooth(data=subset(congress, t2_is_midterm == 0&abs(dv_c_t1) <.25&dwin==1),
              formula= y ~ x, method="lm_robust") +
  geom_smooth(data=subset(congress, t2_is_midterm == 0&abs(dv_c_t1) <.25&dwin==0),
              formula= y ~ x, method="lm_robust", col="orange") +
  xlab("Democratic congressional vote share margin of victory, Election t") +
  ylab("Democratic congressional vote share margin of victory, Election t+1") +
  theme_bw()

bin_scatter_reg
```
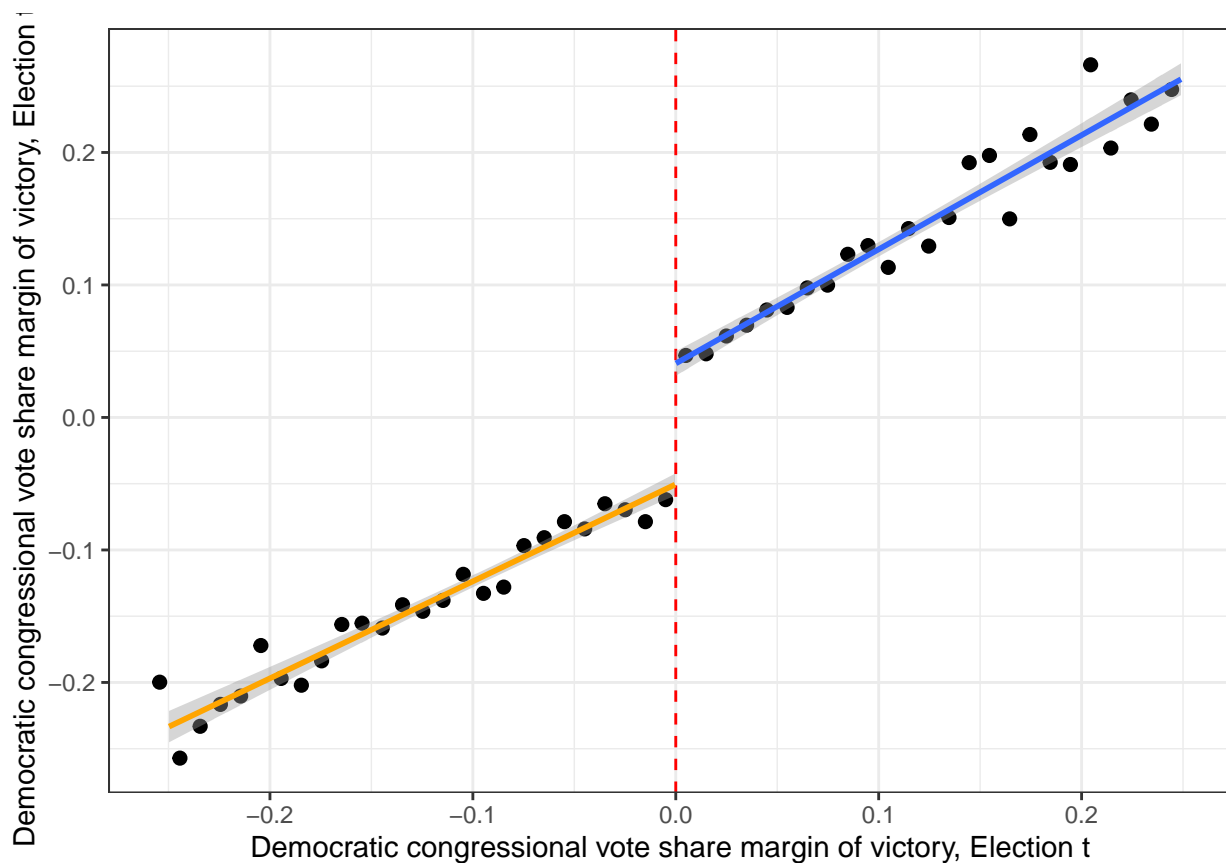

The linear approximation is quite good for observations within the bandwidth of .25 - the bins appear to be roughly scattered evenly above and below the regression line and there is no clear deviation that would be suggestive of obvious non-linearities (especially near the discontinuity).

Part D

Now estimate the effect of democratic congressional incumbency on democratic *presidential* candidate vote share margin of victory. Again, use a local linear regression with a bandwidth of .25. Provide a 95% confidence interval for your estimate and interpret your results. Provide a binned scatterplot with the local linear regressions overlaid on top and discuss whether the local linear models are a reasonable approximation to the conditional expectation functions.

```
# Run the regression on non-midterm elections w/in .25 of the cut-off
lm_robust(dv_p_t2 ~ dwin + dv_c_t1 + dwin*dv_c_t1, data= congress %>% filter(t2_is_midterm == 0&abs(dv_c
```
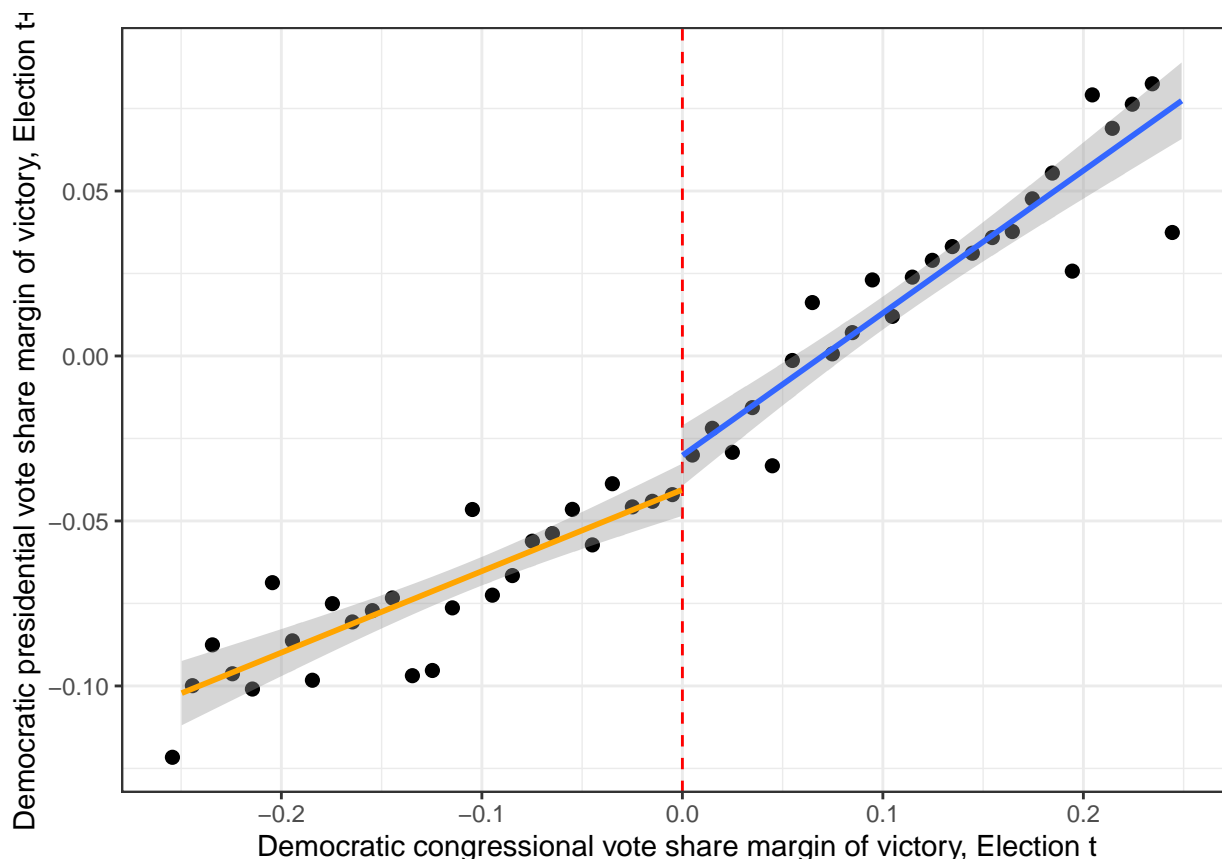
```
##              Estimate Std. Error t value  Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)  -0.0405    0.00404  -10.04  2.26e-23 -0.04845  -0.0326 3299
## dwin          0.0103    0.00617    1.68  9.39e-02 -0.00176   0.0224 3299
## dv_c_t1       0.2469    0.03143    7.85  5.40e-15  0.18524   0.3085 3299
## dwin:dv_c_t1  0.1850    0.04867    3.80  1.46e-04  0.08960   0.2804 3299
```

We estimate that the effect of democratic congressional incumbency on democratic presidential vote share in
the district is 0.01 with a 95% confidence interval of $[-0.00176, 0.0224]$. We would fail to reject the null of no
effect at the 0.05 level. In contrast to the large discontinuity detected for *congressional* vote share, we find
minimal evidence that this transfers over to *presidential* vote share.

```
bin_scatter_reg2 <- ggplot(aes(x=dv_c_t1, y=dv_p_t2), data=subset(congress, t2_is_midterm == 0&abs(dv_c
  stat_summary_bin(fun='mean', bins=50,
                   size=2, geom='point') +
  geom_vline(xintercept=0, col="red", lty=2) +
  geom_smooth(data=subset(congress, t2_is_midterm == 0&abs(dv_c_t1) <.25&dwin==1), formula= y ~ x, meth
  geom_smooth(data=subset(congress, t2_is_midterm == 0&abs(dv_c_t1) <.25&dwin==0), formula= y ~ x, meth
  xlab("Democratic congressional vote share margin of victory, Election t") +
  ylab("Democratic presidential vote share margin of victory, Election t+1") +
  theme_bw()
```

```
bin_scatter_reg2
```


Overall the linear approximation to the CEF is not too bad for the chosen bandwidth, though it appears that

14

there might be some non-linearities in the relationship as we look at margins of victory near .2 to .25.

Part E

Consider choosing a number of different bandwidths for estimating the RD effect on *presidential* vote share using a local linear regression. Estimate the effect using three different bandwidth choices: .05, .1, and .4. For each estimate, provide a 95% confidence interval and interpret your results. Discuss the extent to which the results differ (both the point estimates and standard errors).

Given your analyses, would you conclude that there is evidence of a "reverse coattails" effect?

```
## Smallest bandwidth - 0.05
lm_robust(dv_p_t2 ~ dwin + dv_c_t1 + dwin*dv_c_t1, data= congress %>% filter(t2_is_midterm == 0&abs(dv_
```

```
##                 Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
## (Intercept)      -0.0390    0.00861  -4.530 6.85e-06  -0.0559  -0.0221 751
## dwin              0.0112    0.01242   0.905 3.66e-01  -0.0131   0.0356 751
## dv_c_t1           0.2525    0.28437   0.888 3.75e-01  -0.3057   0.8108 751
## dwin:dv_c_t1     -0.1866    0.43315  -0.431 6.67e-01  -1.0369   0.6637 751
```

```
## Intermediate bandwidth - 0.1
lm_robust(dv_p_t2 ~ dwin + dv_c_t1 + dwin*dv_c_t1, data= congress %>% filter(t2_is_midterm == 0&abs(dv_
```

```
##                 Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)     -0.03691    0.00629  -5.865 5.54e-09  -0.0493  -0.0246 1475
## dwin            -0.00152    0.00908  -0.167 8.67e-01  -0.0193   0.0163 1475
## dv_c_t1          0.30240    0.11855   2.551 1.08e-02   0.0699   0.5349 1475
## dwin:dv_c_t1     0.29756    0.17240   1.726 8.46e-02  -0.0406   0.6357 1475
```

```
## Largest bandwidth - 0.40
lm_robust(dv_p_t2 ~ dwin + dv_c_t1 + dwin*dv_c_t1, data= congress %>% filter(t2_is_midterm == 0&abs(dv_
```

```
##                 Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper   DF
## (Intercept)     -0.03810    0.00385   -9.90 8.28e-23  -0.0456  -0.0306 3740
## dwin            -0.00993    0.00583   -1.70 8.87e-02  -0.0214   0.0015 3740
## dv_c_t1          0.27398    0.02734   10.02 2.45e-23   0.2204   0.3276 3740
## dwin:dv_c_t1     0.33044    0.03998    8.27 1.91e-16   0.2521   0.4088 3740
```

Using the smallest bandwidth of 0.05, we estimate an effect of 0.0112 with a 95% confidence interval of $[-0.0131, 0.0356]$. With the intermediate bandwidth of 0.1, we estimate an effect of -0.00152 with a 95% confidence interval of $[-0.0193, 0.0163]$. With the largest bandwidth, we estimate an effect of -0.00993 with a 95% confidence interval of $[-0.0193, 0.0163]$. In all four cases (including our original analysis in Question 4), we fail to reject the null of no effect at the $\alpha = .05$ level. All of the estimates are very close to 0. The most extreme point estimate comes from our 0.05 bandwidth, but in that case, the standard error is also much larger and our confidence interval still covers 0. As expected, as we increase the bandwidth, we reduce the standard error of our estimate (since we are using more observations). Whether this induces any additional biases depends on how well the linear approximation to the CEF holds for the larger bandwidths. In this case, the approximation *does* start to break down as we expand the bandwidth out to .4 (as shown in the figure below).

```
bin_scatter_reg3 <- ggplot(aes(x=dv_c_t1, y=dv_p_t2), data=subset(congress, t2_is_midterm == 0&abs(dv_c_
  stat_summary_bin(fun='mean', bins=50,
                  size=2, geom='point') +
  geom_vline(xintercept=0, col="red", lty=2) +
  geom_smooth(data=subset(congress, t2_is_midterm == 0&abs(dv_c_t1) <.4&dwin==1), formula= y ~ x, metho
  geom_smooth(data=subset(congress, t2_is_midterm == 0&abs(dv_c_t1) <.4&dwin==0), formula= y ~ x, metho
```

```
xlab("Democratic congressional vote share margin of victory, Election t") +
ylab("Democratic presidential vote share margin of victory, Election t+1") +
theme_bw()
```

bin_scatter_reg3
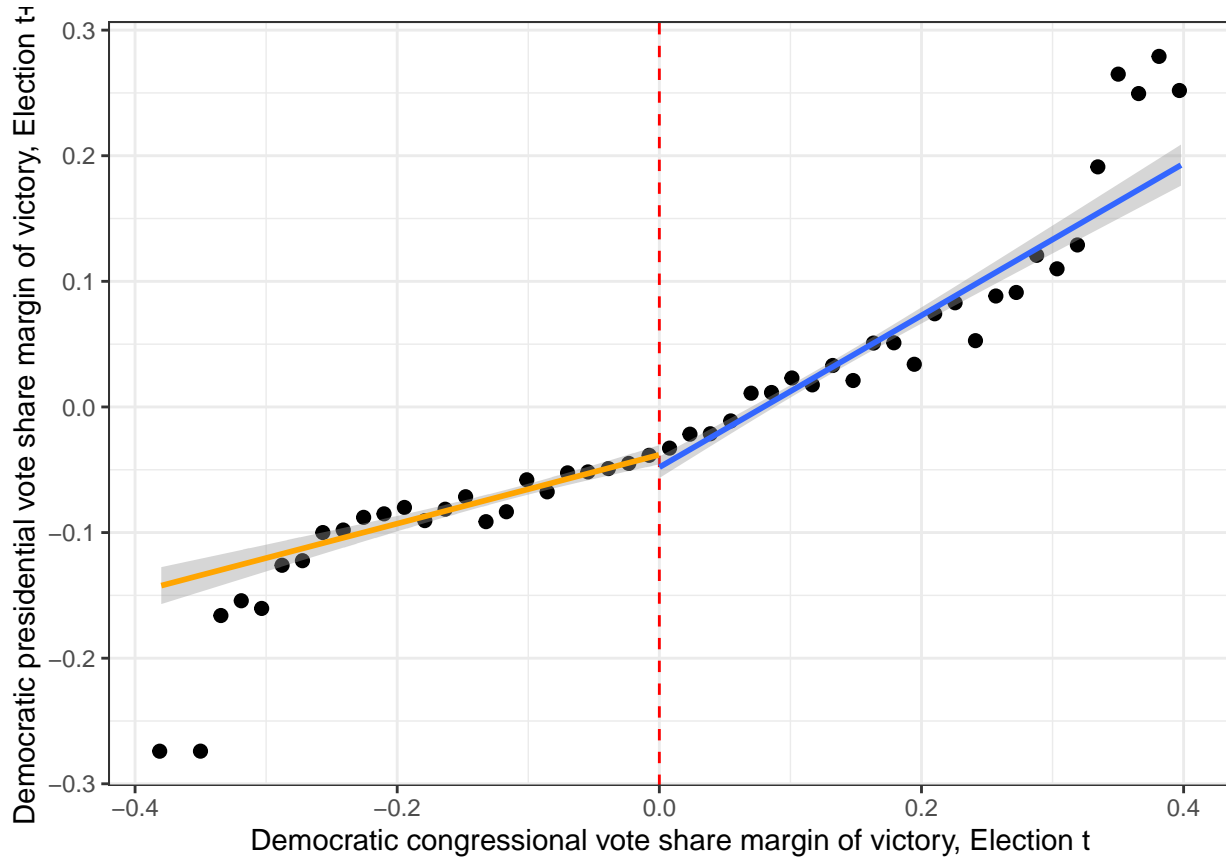
Overall, the results do not appear to meaningfully change depending on our choice of bandwidth. In all cases, we fail to detect evidence of an effect of democratic congressional incumbency on democratic presidential vote share despite replicating the original incumbency effect on democratic congressional vote share. The analysis does not provide evidence for a "reverse coattails" effect.