Problem Set 5

Emanuel Mejía

04/09/2024

Vietnam Draft Lottery

Observational esteimate

Suppose that you had not run an experiment. Estimate the "effect" of each year of education on income as an observational researcher might, by just running a regression of years of education on income (in R-ish, `income ~ years_education`). What does this naive regression suggest?

```
model_observational <- lm(income ~ years_education, data = d)

stargazer(
  model_observational,
  type = 'latex', header=F,
  title="Years of Education on Income",
  dep.var.labels = "Income",
  column.labels = c("Naive Regression"),
  order="Constant",
  covariate.labels = c("(Intercept)","Years of Education")
)
```

Table 1: Years of Education on Income

	Dependent variable:
	Income Naive Regression
(Intercept)	$-23{,}354.640^{***}$
	(1,252.740)
Years of Education	$5{,}750.480^{***}$
	(83.340)
Observations	19,567
R^2	0.196
Adjusted R^2	0.196
Residual Std. Error	26,592.180 (df = 19565)
F Statistic	$4{,}761.015^{***}$ (df = 1; 19565)
Note:	*p<0.1; **p<0.05; ***p<0.01

Answer: … This model suggests a very statistically significant and positive relationship between years of education and income. For every additional year of education we can expect, on average, an increase of 5750.48 in income.

Evaluating observational estimate

Continue to suppose that we did not run the experiment, but that we saw the result that you noted in part 1. Tell a concrete story about why you don't believe that observational result tells you anything causal.

Answer: … There may be a lot of reasons why this statement, while proving correlation, doesn't necessarily imply causation. Perhaps those with more years of education had more money to begin with, well accommodated families, better contacts or live in a metropolitan area, and those ommited variables could cause having higher incomes in the present while not necessarily just because having studied longer.

Natural experiment effect on education

Now, let's get to using the natural experiment. Define "having a high-ranked draft number" as having a draft number between 1-80. For the remaining 285 days of the year, consider them having a "low-ranked" draft number). Create a variable in your dataset called `high_draft` that indicates whether each person has a high-ranked draft number or not. Using a regression, estimate the effect of having a high-ranked draft number on years of education obtained. Report the estimate and a correctly computed standard error. (*Hint: How is the assignment to having a draft number conducted? Does random assignment happen at the individual level? Or, at some higher level?)

```
d[, high_draft := 0]
d[draft_number <= 80, high_draft := 1]

model_education <- lm(years_education ~ high_draft, data = d)
se_edu_cluster <- sqrt(diag(vcovCL(model_education, cluster = d[ , draft_number])))

coefs_edu <- coeftest(
  model_education,
  vcov = vcovCL(
    model_education,
    cluster = d[ , draft_number]
    )
  )

stargazer(
  model_education, model_education,
  type = 'latex', header=F,
  se = list(NULL, se_edu_cluster),
  title="High Ranked Draft on Years of Education",
  dep.var.labels = "Years of Education",
  column.labels = c("Vanilla SE", "Clustered SE"),
  order="Constant",
  covariate.labels = c("(Intercept)","High Ranked Draft")
)
```

Answer: … The effect of having a high-rank draft on years of education is positive and statistically significant at a 0.1% level. On average, having a high-rank draft implies 2.13 years of study with a standard error of 0.04, clustered at a draft number level since we know that the treatment was randomized at a birth date level (equivalent to draft number).

Table 2: High Ranked Draft on Years of Education

	Dependent variable:	
	Years of Education	
	Vanilla SE	Clustered SE
	(1)	(2)
(Intercept)	14.434***	14.434***
	(0.017)	(0.018)
High Ranked Draft	2.126***	2.126***
	(0.038)	(0.038)
Observations	19,567	19,567
R^2	0.138	0.138
Adjusted R^2	0.138	0.138
Residual Std. Error (df = 19565)	2.117	2.117
F Statistic (df = 1; 19565)	3,145.132***	3,145.132***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Natural experiment effect on income

Using linear regression, estimate the effect of having a high-ranked draft number on income. Report the estimate and the correct standard error.

```r
model_income <- lm(income ~ high_draft, data = d)
se_inc_cluster <- sqrt(diag(vcovCL(model_income, cluster = d[ , draft_number])))

coefs_inc <- coeftest(
  model_income,
  vcov = vcovCL(
    model_income,
    cluster = d[ , draft_number]
    )
  )

stargazer(
  model_income, model_income,
  type = 'latex', header=F,
  se = list(NULL, se_inc_cluster),
  title="High Ranked Draft on Income",
  dep.var.labels = "Income",
  column.labels = c("Vanilla SE", "Clustered SE"),
  order="Constant",
  covariate.labels = c("(Intercept)","High Ranked Draft")
)
```

Answer: ... The effect of having a high-rank draft on income is also positive and statistically significant at a 0.1% level. On average, having a high-rank draft implies an increase of 6637.55 dollars on income with a standard error of 511.9, once again clustered at a draft number level.

Table 3: High Ranked Draft on Income

	Dependent variable:	
	Income	
	Vanilla SE	Clustered SE
	(1)	(2)
(Intercept)	60,761.890***	60,761.890***
	(235.917)	(244.361)
High Ranked Draft	6,637.554***	6,637.554***
	(528.703)	(511.899)
Observations	19,567	19,567
R^2	0.008	0.008
Adjusted R^2	0.008	0.008
Residual Std. Error (df = 19565)	29,532.970	29,532.970
F Statistic (df = 1; 19565)	157.613***	157.613***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Instrumental variables estimate of education on income

Now, estimate the Instrumental Variables regression to estimate the effect of education on income. To do so, use `AER::ivreg`. After you evaluate your code, write a narrative description about what you learn.

```
model_iv <- ivreg(income ~ years_education | high_draft, data = d)
se_iv_cluster <- sqrt(diag(vcovCL(model_iv, cluster = d[ , draft_number])))

coefs_iv <- coeftest(
  model_iv,
  vcov = vcovCL(
    model_iv,
    cluster = d[ , draft_number]
    )
  )

stargazer(
  model_iv, model_iv,
  type = 'latex', header=F,
  se = list(NULL, se_iv_cluster),
  title="Education on Income. IV - High Rank Draft",
  dep.var.labels = "Income",
  column.labels = c("Vanilla SE", "Clustered SE"),
  order="Constant",
  covariate.labels = c("(Intercept)","Years of Education")
)
```

Answer: … When using the high ranked draft numbers as our instrumental variable we still have a statistically significant result of education on income, in this case, for each year of study we expect an increase of 3122.44 dollars in income. Although this effect is way smaller than the one we estimated with our first naive regression. Besides what we already said about a possible OVB, we know that in this case we would like to compare apples to apples and out treatment assignment method (draft numbers) doesn't

Table 4: Education on Income. IV - High Rank Draft

	Dependent variable:	
	Income	
	Vanilla SE	Clustered SE
	(1)	(2)
(Intercept)	15,691.580***	15,691.580***
	(3,416.372)	(3,371.412)
Years of Education	3,122.444***	3,122.444***
	(229.567)	(225.884)
Observations	19,567	19,567
R^2	0.155	0.155
Adjusted R^2	0.155	0.155
Residual Std. Error (df = 19565)	27,259.570	27,259.570
Note:	*p<0.1; **p<0.05; ***p<0.01	

necessarily imply receiving the treatment (years of education), so we need to adjust to measure among "compliers" and that's why the effect is actually smaller when analyzing it properly.

Evaluating the exclusion restriction

Give one reason this requirement might not be satisfied in this context. In what ways might having a high draft rank affect individuals' income **other** than nudging them to attend more school?

Answer: . . . In this specific case we have the *following process*¨: FIRST we're randomizing by the number obtained of the Vietnam Draft Lottery, THEN we're "treating" through the number of education years and FINALLY measuring the income as our outcome. The issue is that our randomization might affect the outcome in some other ways besides the years of education. For instance, we know about potential mental disorders for those who were drafted and actually went to Vietnam which could reduce the income and our positive effect for education might be underestimated. On the other hand it could be the other way around, perhaps those who were drafted and went to war ended up with other set of skills that prove to be valuable in the job market, or perhaps receiving some stipend for veterans besides their job which might increase their outcome and therefore overestimating our effect.

Differential attrition

Conduct a test for the presence of differential attrition by treatment condition. That is, conduct a formal test of the hypothesis that the "high-ranked draft number" treatment has no effect on whether we observe a person's income. **(Note, that an earning of $0 *actually* means they didn't earn any money – i.e. earning $0 does not mean that their data wasn't measured. Let's be really, really specific: If you write a model that looks anything like, `lm(income == 0 ~ .)` you've gone the wrong direction.)**

```
# count people by each draft number
count_x_day <- d[, .(people = .N), keyby = .(draft_number, high_draft)]

# test for difference in means
```

```r
att_test <- t.test(
  count_x_day[high_draft == 1, people],
  count_x_day[high_draft == 0, people]
  )
count_means <- att_test$estimate

att_test
```

```
##
##  Welch Two Sample t-test
##
## data:  count_x_day[high_draft == 1, people] and count_x_day[high_draft == 0, people]
## t = -6.6358, df = 123.31, p-value = 9.121e-10
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.160996 -4.410934
## sample estimates:
## mean of x mean of y
##  48.70000  54.98596
```

Answer: … When looking at the data, the attrition source might not be obvious since there are no specific missing values, and even when we have some rows with 0 income, these are actual earnings. Although, when counting how many people we have for each draft number, we see that on average there are 48.7 people for high ranked draft numbers, while 55 on average for low ranked draft numbers. By running a t-test we REJECT the null hypothesis H_0 : *Difference in means is equal to 0.* Meaning we have a statistically significant result telling us that there's less people recorded for high ranked draft numbers, which could be a source of attrition since these numbers are randomly assigned by birth date and there's no reason to believe there should be less people with birthdays among high ranked numbers. It seems more likely that we have differential attrition.

Evaluate differential attrition

Tell a concrete story about what could be leading to the result in part 7. How might this differential attrition create bias in the estimates of a causal effect?

Answer: … There may be many reasons why this attrition is not random. Those who were more likely to be drafted, seem to have a reason for not being measurable, and unfortunately one plausible explanation is that those who were drafted (and didn't manage to get into college), perhaps didn't return at all from Vietnam. And this could be leading to bias in our estimated effect because we can't know what the study years or income for those missing people would have been, and moreover, we already said our randomization could be not following the exclusion restriction so these not recorded results could be subject to other features moving income besides years of study.

Think about Treatment Effects

Throughout this course we have focused on the average treatment effect. *Why* we are concerned about the average treatment effect. What is the relationship between an ATE, and some individuals' potential outcomes? Make the strongest case you can for why this is a *good* measure.

The real average treatment effect is, in most experiments, just a fictional value since we can't really observe a subject's outcome after receiving the treatment and after not receiving it, at the very same time. Moreover this is just one number to represent all of the possible treatment effects that vary among many subjects. And let's face it, besides the ones with academic purposes, most people will be interested in an experiment to determine exactly how it's going to affect them specifically, and they don't want to know an average, they want to know the exact effect it will have on them (or their neighborhood, school, state, country, etc). So we might say, why bother so much in obtaining a number that we can't know for sure and that doesn't even tell people how a specific experiment will work on them.

First, if the experiment is run properly we know that the estimator for ATE will be unbiased, and therefore even when it's not the exact and utopian ATE, it should be a value that we can be confident that the only reason it could be different would be plain luck, and therefore we have the closest value we can get.

On the other hand and what's most important to the final users, it is because of that very same variability among subjects that we cannot really say an exact effect value for any subject that has not actually been tested. Our best guess is the ATE, since it holds onto statistical properties to ensure this is the best quantity to explain the researched effect, the one closest to every effect value or, in other words, the one with lowest error. And someone might say, but hey! what if I give my information to compare myself only to similar people (or neighborhoods, school, ...) and therefore get a more exact estimate; but unless the experiment was intended that way from the beginning (for a specific type of study subjects), every other feature should not affect the treatment effect because the experiment is really designed to know what the effect of the treatment is in a subject's outcome, not any other variable's effect on the outcome.

So, if you're planning on using the results of a specific experiment, the ATE is the best quantity to describe the result you should expect, and if this is not enough you might get a range around this number (called confidence interval) to determine how scattered the effects could be.

Optional Online advertising natural experiment.

Cross table of total_ads and treatment_ads

A. Run a crosstab – which in R is `table` – of `total_ads` and `treatment_ads` to sanity check that the distribution of impressions looks as it should. After you write your code, write a few narrative sentences about whether this distribution looks reasonable. Why does it look like this? (No computation required here, just a brief verbal response.)

```
cross_tab <- table(d$total_ads, d$treatment_ads)

cross_tab
```

```
##
##         0     1     2     3     4     5     6
##   0 61182     0     0     0     0     0     0
##   1 36754 37215     0     0     0     0     0
##   2 21143 42036 20965     0     0     0     0
##   3 10683 32073 32314 10726     0     0     0
##   4  5044 20003 30432 20223  5115     0     0
##   5  2045 10563 20970 20793 10293  2131     0
##   6   729  4437 10977 14771 11147  4486   750
```

Answer: ... From this crosstab we can see a reasonable distribution of treatment according to the proposed randomization process. First we know that there can't be more treatment ads than total ads for any person, therefore we have only zeroes to the right of the diagonal. Then we know that there's the same probability to receive the treatment or to receive the placebo, and every time we browse again we get the same 50/50 chances (ads are not persistent across subjects). So it is consistent that every row (total ads) will follow a binomial distribution with success probability of 50% where the number of "trials" would be the total ads and the number of "success" would be the treatment ads. The observed pattern seems to follow this description.

Placebo test

A colleague of yours proposes to estimate the following model: `d[ , lm(week1 ~ tretment_ads)]` You are suspicious. Run a placebo test with `week0` purchases as the outcome and report the results. Since treatment is applied in week 1, and `week0` is purchases in week 0, *should* there be an relationship? Did the placebo test "succeed" or "fail"? Why do you say so?

```
model_colleague <- d[ , lm(week1 ~ treatment_ads)]
model_colleague_placebo <- d[ , lm(week0 ~ treatment_ads)]
summary(model_colleague)
```

```
##
## Call:
## lm(formula = week1 ~ treatment_ads)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.409 -2.213 -1.615  2.388  8.285
##
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.614684   0.005995  269.34   <2e-16 ***
## treatment_ads 0.299113   0.003138   95.32   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.781 on 499998 degrees of freedom
## Multiple R-squared:  0.01785,    Adjusted R-squared:  0.01785
## F-statistic:  9086 on 1 and 499998 DF,  p-value: < 2.2e-16
```

```r
summary(model_colleague_placebo)
```

```
##
## Call:
## lm(formula = week0 ~ treatment_ads)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.248 -2.196 -1.670  2.430  8.330
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.669685   0.006027   277.0   <2e-16 ***
## treatment_ads 0.263099   0.003155    83.4   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.796 on 499998 degrees of freedom
## Multiple R-squared:  0.01372,    Adjusted R-squared:  0.01372
## F-statistic:  6955 on 1 and 499998 DF,  p-value: < 2.2e-16
```

Answer: … This model results in a positive and statistically significant effect of 0.2631 on purchase revenues in week 0 for those who received the treatment ads. This is not consistent, since week 0 was before receiving any treatment we should observe no significant difference by receiving any amount of treatment ads, therefore this placebo test failed.

What has gone wrong?

Here's the tip off: the placebo test suggests that there is something wrong with our experiment (i.e. the randomization isn't working) or our data analysis. We suggest looking for a problem with the data analysis. Do you see something that might be spoiling the "randomness" of the treatment variable? (Hint: it should be present in the cross-tab that you wrote in the first part of this question.) How can you improve your analysis to address this problem? Why does the placebo test turn out the way it does? What one thing needs to be done to analyze the data correctly? Please provide a brief explanation of why, not just what needs to be done.

Answer: … As we said when looking at the crosstab there doesn't seem to be any issues regarding the randomization process. The main issue is that we have different groups of people, from those who didn't receive any ad to those who received 6 ads and for instance (for instance, the only way someone could receive 6 treatment ads is by being in the group who received 6 total ads). People who received more adds in general might be systematically different from those who received none, for some reason there are people who browse a lot more than others, and this might not be the only difference between them, perhaps online purchasing habit different between them as well. So one plausible solution is to block between groups and analyze the results after removing the effect of pertaining to a certain group.

Conduct proposed solution and re-evaluate placebo test

Implement the procedure you propose from part 3, run the placebo test for the Week 0 data again, and report the results. (This placebo test should pass; if it does not, re-evaluate your strategy before wasting time proceeding.) How can you tell this this has fixed the problem? Is it possible, even though this test now passes, that there is still some other problem?

```r
model_passes_placebo <- d[ , lm(week0 ~ treatment_ads + as.factor(total_ads))]
summary(model_passes_placebo)
```

```
##
## Call:
## lm(formula = week0 ~ treatment_ads + as.factor(total_ads))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.847 -2.084 -1.616  2.458  7.818
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)            1.306783   0.011234 116.328   <2e-16 ***
## treatment_ads         -0.002287   0.004629  -0.494    0.621
## as.factor(total_ads)1  0.311564   0.015362  20.281   <2e-16 ***
## as.factor(total_ads)2  0.551112   0.015469  35.627   <2e-16 ***
## as.factor(total_ads)3  0.779854   0.016264  47.948   <2e-16 ***
## as.factor(total_ads)4  1.003116   0.017545  57.174   <2e-16 ***
## as.factor(total_ads)5  1.236092   0.019372  63.810   <2e-16 ***
## as.factor(total_ads)6  1.539865   0.021981  70.053   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.779 on 499992 degrees of freedom
## Multiple R-squared:  0.02563,    Adjusted R-squared:  0.02562
## F-statistic:  1879 on 7 and 499992 DF,  p-value: < 2.2e-16
```

Answer: … By removing the block effect we can see that we've fixed the issue we previously had an this placebo test is now consistent with the proposed methodology, there seems to be no effect of the amount of treatment ads received on the purchases before the experiment started. At this moment there seems to be no additional issue or violation, although this is no necessarily a guarantee.

Estimate treatment effect with proposed solution

Now estimate the causal effect of each ad exposure on purchases during Week 1. You should use the same technique that passed the placebo test in part 4. Describe how, if at all, the treatment estimate that your model produces changes from the estimate that your colleague produced.

```r
model_causal <- d[ , lm(week1 ~ treatment_ads + as.factor(total_ads))]
summary(model_causal)
```

```
##
## Call:
## lm(formula = week1 ~ treatment_ads + as.factor(total_ads))
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.003 -2.102 -1.552  2.447  8.110
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1.295932   0.011185  115.87   <2e-16 ***
## treatment_ads         0.056337   0.004609   12.22   <2e-16 ***
## as.factor(total_ads)1 0.256600   0.015295   16.78   <2e-16 ***
## as.factor(total_ads)2 0.487427   0.015402   31.65   <2e-16 ***
## as.factor(total_ads)3 0.693409   0.016194   42.82   <2e-16 ***
## as.factor(total_ads)4 0.919351   0.017469   52.63   <2e-16 ***
## as.factor(total_ads)5 1.136115   0.019287   58.91   <2e-16 ***
## as.factor(total_ads)6 1.368843   0.021886   62.55   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.767 on 499992 degrees of freedom
## Multiple R-squared:  0.02783,    Adjusted R-squared:  0.02782
## F-statistic:  2045 on 7 and 499992 DF,  p-value: < 2.2e-16
```

Answer: … We find now a statistically significant effect for purchases in week one of 0.0563 on income per each treatment ad received. This is a much smaller result than the 0.2991 effect found with the first model.

Defend you method

Upon seeing these results, the colleague who proposed the specification that did not pass the placeo test challenges your results – they make the campaign look less successful! Write a short paragraph (i.e. 4-6 sentences) that argues for why your estimation strategy is better positioned to estimate a causal effect.

Answer: … Even when both results are statistically significant and it seems like the proposed strategy is less successful since it results in less revenue per printed ad, it actually gives a better estimate of the treatment effect. The result of modeling without taking into account the blocking by total ads received would be comparing different people to each other, whose decisions about purchasing are different for other reasons besides the treatment effect and that's why we can observe a significant effect even before the experiment begins. If we want to know only about the effect of the ads we need to remove those systemic differences, and even when the result seems smaller, we can say that it is only due to the treatment.

Intertemporal substitution?

One concern raised by David Reiley is that advertisements might just shift *when* people purchase something – rather than increasing the total amount they purchase. Given the data that you have available to you, can you propose a method of evaluating this concern? Estimate the model that you propose, and describe your findings.

```
d[, sales := week0 + week1 + week2 + week3 + week4 +
    week5 + week7 + week8 + week9 + week10]
```

```r
model_overall <- d[ , lm(sales ~ treatment_ads + as.factor(total_ads))]
summary(model_overall)
```

```
##
## Call:
## lm(formula = sales ~ treatment_ads + as.factor(total_ads))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -30.457  -7.279  -0.715   6.543  63.374
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           16.45742    0.04208 391.082   <2e-16 ***
## treatment_ads          0.01038    0.01734   0.598     0.55
## as.factor(total_ads)1  2.66567    0.05755  46.321   <2e-16 ***
## as.factor(total_ads)2  4.91695    0.05795  84.852   <2e-16 ***
## as.factor(total_ads)3  7.14129    0.06093 117.209   <2e-16 ***
## as.factor(total_ads)4  9.20127    0.06572 139.998   <2e-16 ***
## as.factor(total_ads)5 11.33788    0.07257 156.239   <2e-16 ***
## as.factor(total_ads)6 13.93707    0.08234 169.254   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.41 on 499992 degrees of freedom
## Multiple R-squared:  0.1383, Adjusted R-squared:  0.1383
## F-statistic: 1.146e+04 on 7 and 499992 DF,  p-value: < 2.2e-16
```

Answer: ... When looking at the treatment effect of ads on overall sales, the effect while still positive, is diluted and not significant. This might be a sign of shifting purchase dates rather than actually having an effect on increasing sales.

Weekly effects

If you look at purchases in each week – one regression estimated for each outcome from week 1 through week 10 (that's 10 regression in a row) – what is the relationship between treatment ads and purchases in each of those weeks. This is now ranging into exploring data with models – how many have we run in this question alone!? – so consider whether a plot might help make whatever relationship exists more clear.

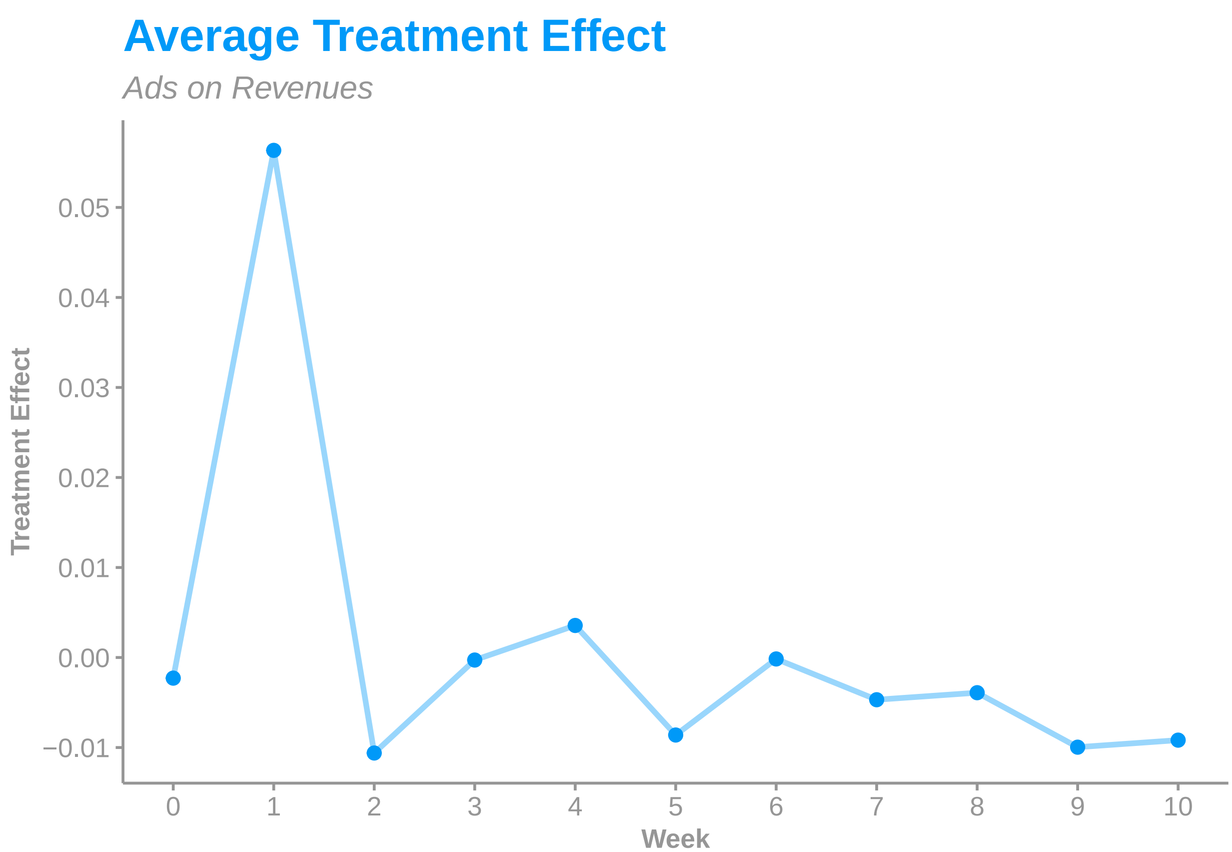
```r
eff_vec <- c()
eff_vec[1] <- coeftest(d[ , lm(week0 ~ treatment_ads + as.factor(total_ads))])[2]
eff_vec[2] <- coeftest(d[ , lm(week1 ~ treatment_ads + as.factor(total_ads))])[2]
eff_vec[3] <- coeftest(d[ , lm(week2 ~ treatment_ads + as.factor(total_ads))])[2]
eff_vec[4] <- coeftest(d[ , lm(week3 ~ treatment_ads + as.factor(total_ads))])[2]
eff_vec[5] <- coeftest(d[ , lm(week4 ~ treatment_ads + as.factor(total_ads))])[2]
eff_vec[6] <- coeftest(d[ , lm(week5 ~ treatment_ads + as.factor(total_ads))])[2]
eff_vec[7] <- coeftest(d[ , lm(week6 ~ treatment_ads + as.factor(total_ads))])[2]
eff_vec[8] <- coeftest(d[ , lm(week7 ~ treatment_ads + as.factor(total_ads))])[2]
eff_vec[9] <- coeftest(d[ , lm(week8 ~ treatment_ads + as.factor(total_ads))])[2]
eff_vec[10] <- coeftest(d[ , lm(week9 ~ treatment_ads + as.factor(total_ads))])[2]
eff_vec[11] <- coeftest(d[ , lm(week10 ~ treatment_ads + as.factor(total_ads))])[2]
```

```r
se_vec <- c()
se_vec[1] <- coeftest(d[ , lm(week0 ~ treatment_ads + as.factor(total_ads))])[10]
se_vec[2] <- coeftest(d[ , lm(week1 ~ treatment_ads + as.factor(total_ads))])[10]
se_vec[3] <- coeftest(d[ , lm(week2 ~ treatment_ads + as.factor(total_ads))])[10]
se_vec[4] <- coeftest(d[ , lm(week3 ~ treatment_ads + as.factor(total_ads))])[10]
se_vec[5] <- coeftest(d[ , lm(week4 ~ treatment_ads + as.factor(total_ads))])[10]
se_vec[6] <- coeftest(d[ , lm(week5 ~ treatment_ads + as.factor(total_ads))])[10]
se_vec[7] <- coeftest(d[ , lm(week6 ~ treatment_ads + as.factor(total_ads))])[10]
se_vec[8] <- coeftest(d[ , lm(week7 ~ treatment_ads + as.factor(total_ads))])[10]
se_vec[9] <- coeftest(d[ , lm(week8 ~ treatment_ads + as.factor(total_ads))])[10]
se_vec[10] <- coeftest(d[ , lm(week9 ~ treatment_ads + as.factor(total_ads))])[10]
se_vec[11] <- coeftest(d[ , lm(week10 ~ treatment_ads + as.factor(total_ads))])[10]


treat_eff <- data.table(
  week = 0:10,
  treat_effect = round(eff_vec, 5),
  standard_error = round(se_vec, 5)
)

treat_eff
```

```
##      week treat_effect standard_error
##     <int>        <num>          <num>
##  1:     0     -0.00229        0.00463
##  2:     1      0.05634        0.00461
##  3:     2     -0.01061        0.00544
##  4:     3     -0.00028        0.00545
##  5:     4      0.00356        0.00544
##  6:     5     -0.00861        0.00545
##  7:     6     -0.00016        0.00545
##  8:     7     -0.00469        0.00545
##  9:     8     -0.00391        0.00544
## 10:     9     -0.00996        0.00544
## 11:    10     -0.00918        0.00544
```

```r
# Basic line plot with points
ggplot(data=treat_eff, aes(x=week, y=treat_effect, group=1)) +
  geom_line(color="#0099F8",
            alpha = 0.4,
            size = 1)+
  geom_point(color="#0099F8",
             size = 2) +
  labs(
    title = "Average Treatment Effect",
    subtitle = "Ads on Revenues",
    x = "Week",
    y = "Treatment Effect"
  ) +
  theme_classic() +
  theme(
    plot.title = element_text(color = "#0099F8",
                              size = 17,
```

```r
                                face = "bold"),
    plot.subtitle = element_text(color="#969696",
                                     size = 12,
                                     face = "italic"),
    axis.title = element_text(color = "#969696",
                                size = 10,
                                face = "bold"),
    axis.text = element_text(color = "#969696", size = 10),
    axis.line = element_line(color = "#969696"),
    axis.ticks = element_line(color = "#969696")
  ) +
  scale_x_continuous(breaks=seq(0,10,1)) +
  scale_y_continuous(breaks=seq(-0.01,0.05,0.01))
```



Answer: . . . Besides week 1, which is the week when the experiment took place, every other week has treatment effects near zero or slightly negative. Although, when looking at the effects table we can notice that these aren't statistically significant for any other week besides week 1 (week 2 might be on the threshold of significance but there are no other cases).

Evaluating what is happening in the data

I. What might explain this pattern in your data. Stay curious when you're writing models! But, also be clear that we're fitting a **lot** of models and making up a theory/explanation after the fact.

Answer: . . . This could be indeed a shifting effect and that's why we have negative effect right after the treatment week. Although, this effect is not as large like to cancel the increase in sales experienced during the experiment week. And even when it seems like there are many negative values in the weeks after, these are too small and not significant at all like to say this is a true difference. So, after analyzing all this it seems like there is indeed an increase in sales but only during times when people are exposed to the ads. If we wanted to know if this effect is persistent while sending adds during several weeks, we should do a larger experiment to see if the effect is washed down as time passes.