MQE: Economic Inference from Data: Module 5: Difference in Differences

Claire Duquennois

6/9/2020

Causal Inference with non-random assignment:

Randomizing treatment is not always possible:

- ▶ the program or policy has already happened

- ▶ randomization in unfeasible

- ▶ randomizing treatment would be unethical (ex: randomizing exposure to pollutants)

Causal Inference with non-random assignment:

With no randomized control trial we have to assume that the treatment was not randomly assigned:

- ▶ treatment will depend on observable and/or unobservable characteristics

- ▶ their are important differences between our treated and untreated units that we cannot control for

- ▶ Leaving these variables out in the error term will cause OVB.

Differences-in-differences is a way of getting around non-random assignment of a treatment.

DID: Example and intuition[1]

2002: Craigslist opens a new section called "erotic services" in San Francisco:

- ▶ mostly used by sex workers to advertise and solicit clients.

- ▶ Sex workers claimed it made them safer, because they could solicit indoors from their computers and learn more about the men contacting them.

- ▶ Activists and law enforcement worried that it was facilitating sex trafficking and increasing violence against women.

Which was it? Was erotic services (ERS) making women safer, or was it placing them in harm's way?

---

[1]This section is borrowed from scunning.com

SOS Empiricist!

This is an empirical question: What is the effect of ERS on female safety?

- ▶ The fundamental problem of causal inference strikes again!
- ▶ We can't know what effect it had because we are missing the data for the counterfactual:

$$E[\tau] = E[\underbrace{Y_{SF,2003}(D_{SF,2003} = 1)}_{\text{observed}}] - E[\underbrace{Y_{SF,2003}(D_{SF,2003} = 0)}_{\text{unobserved}}]$$

In 2003 only the first occurs, and the second is a counterfactual. So how do we proceed?

Diff-in-Diff to the rescue:

The standard differences-in-differences strategy (DiD):

- ▶ Define the intervention, $D=$ the availability of the Craigslist ERS webpage.

- ▶ We want to know the causal effect, τ of D on $Y=$ female murders.

Can we just compare SF murders in, say, 2003 with some other city, like Pittsburgh?

Differencing A:

....City....	..........Outcome............
San Francisco	$Y_{SF,2003} = \alpha_{SF} + \tau$
Pittsburgh	$Y_{P,2003} = \alpha_P$

- α_{SF} is a San Francisco fixed effect
- α_P is a Pittsburgh fixed effect.

If make a simple comparison between Pittsburgh and SF:

$$\tilde{\tau} = Y_{SF,2003} - Y_{P,2003} = \alpha_{SF} + \tau - \alpha_P.$$

Differencing A:

The simple difference is biased because of the difference in the underlying murder rates between the two cities:

$$\tilde{\tau} - \tau = \alpha_{SF} - \alpha_P.$$

Differencing B:

What if we compare SF to itself? Say in 2003 to 2001?

.....City....	..Time..	.........Outcome.............
San Francisco	Before	$Y_{SF,2001} = \alpha_{SF}$
	After	$Y_{SF,2003} = \alpha_{SF} + \lambda_{03} + \tau$

Again, this doesn't lead to an unbiased estimate of τ since:

$$\tilde{\tau} = \alpha_{SF} + \lambda_{03} + \tau - \alpha_{SF} = \lambda_{03} + \tau$$

We eliminated the city fixed effect but not the changes in the murder rate over time which will bias my estimate:

$$\tilde{\tau} - \tau = \lambda_{03}$$

How can I identify and control for these time effects?

Differencing A+B= Diff-in-Diff

Combining the two approaches to eliminate both the city effects and the time effects:

. . . City. . .	..Time..	. . . . .Outcome.	1st Diff	2nd Diff
SF	Before	$Y_{SF,2001} = \alpha_{SF}$		
	After	$Y_{SF,2003} = \alpha_{SF} + \lambda_{03} + \tau$	$\lambda_{03} + \tau$	
				τ
Pittsburgh	Before	$Y_{P,2001} = \alpha_P$		
	After	$Y_{P,2003} = \alpha_P + \lambda_{03}$	λ_{03}	

The idea

Sometimes treatment and control group outcomes move in parallel in the absence of treatment.

When they do, the divergence of a post-treatment path from the trend established by a comparison group may signal a treatment effect.

The mechanics

Difference-in-differences can be implemented as follows:

1) Compute the difference in the mean outcome variable Y in the post treatment period ($t = 1$) and the before treatment period ($t = 0$) for the control group C:

$$\bar{Y}_{C,1} - \bar{Y}_{C,0} = \Delta \bar{Y}_C$$

$\Rightarrow$ allows us to cancel out the control group fixed effect and identify the time fixed effect since

$$\bar{Y}_{C,1} - \bar{Y}_{C,0} = \alpha_C + \lambda_1 - \alpha_C = \lambda_1 = \Delta \bar{Y}_C$$

The mechanics

2) Compute the difference in the mean outcome variable Y in the post treatment period ($t = 1$) and the before treatment period ($t = 0$) for the treated group T:

$$\bar{Y}_{T,1} - \bar{Y}_{T,0} = \Delta \bar{Y}_T$$

which allows us to cancel out the treated group fixed effect

$$\bar{Y}_{T,1} - \bar{Y}_{T,0} = \alpha_T + \lambda_1 + \tau - \alpha_T = \lambda_1 + \tau = \Delta \bar{Y}_T$$

The mechanics

3) Treatment impact is then measured by the difference-in-differences:

$$(\bar{Y}_{T,1} - \bar{Y}_{T,0}) - (\bar{Y}_{C,1} - \bar{Y}_{C,0}) = (\Delta \bar{Y}_T - \Delta \bar{Y}_C)$$

since by comparing the differences we can cancel out the time fixed effect and isolate the treatment effect of interest:

$$\Delta \bar{Y}_T - \Delta \bar{Y}_C = \lambda_1 + \tau - \lambda_1 = \tau$$

DID Regressions:

This can be done in a regression framework:

How would you specify this regression?

$\Rightarrow$ Top Hat

DID Regressions:

This can be done in a regression framework:

$$Y_{it} = \beta_0 + \beta_1 Post_t + \beta_2 GetsTreat_i + \beta_3 Post_t \times GetsTreat_i + u_{it}$$

- ▶ $Post_t$ is an indicator for the post treatment period,
- ▶ $GetsTreat_i$ is an indicator for observations in the treatment group that eventually gets treated.

Which coefficient identifies the treatment effect?

$\Rightarrow$ Top Hat

DID Regressions:

β_3 is the difference-in-differences estimator:

▶ estimates the differential impact of being in the post treatment period if you are in the treated group since

$$E[Y_{C,1}] - E[Y_{C,0}] = (\beta_0 + \beta_1) - (\beta_0) = \beta_1$$
$$E[Y_{T,1}] - E[Y_{T,0}] = (\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_2) = \beta_1 + \beta_3$$

$$(E[Y_{T,1}] - E[Y_{T,0}]) - (E[Y_{C,1}] - E[Y_{C,0}]) = \beta_3$$

A simulation:

Suppose you are a principal of a school:

► ten 4th grade classrooms of 30 students each.

► Starting in 2001 school year, teachers can enroll their class in the scholastic book club

► 4 of your fourth grade teachers opted to enroll.

You are interested in estimating the effect of participation in the book club on 4th grade reading scores.

A simulation:

```
set.seed(6000)
scores<-as.data.frame(rep(c(1,2,3,4,5,6,7,8,9,10),times=30))
names(scores)<-c("class")
scores <- fastDummies::dummy_cols(scores, select_columns = "class")

scores$error<-rnorm(300, mean=0, sd=10)

#suppose teachers in the better performing classes (classes, 7,8,9,10) select to participate in the book cl
scores$treat<-0
scores$treat[scores$class%in%c(7,8,9,10)]<-1

tau<-10

#the data generating process
scores$read4<-(85+tau*scores$treat
                +(-10)*scores$class_1+(-15)*scores$class_2+(-5)*scores$class_3
                +(-8)*scores$class_4+(-7)*scores$class_5+(-13)*scores$class_6
                +(11)*scores$class_7+(8)*scores$class_8+(10)*scores$class_9
                +(12)*scores$class_10
                +scores$error)

scores$year<-"2001"
scores01<-scores
rm(scores)
```

A simulation:

```r
scores<-as.data.frame(rep(c(1,2,3,4,5,6,7,8,9,10),times=30))
names(scores)<-c("class")
scores <- fastDummies::dummy_cols(scores, select_columns = "class")

scores$error<-rnorm(300, mean=0, sd=10)

scores$treat<-0
scores$treat[scores$class%in%c(7,8,9,10)]<-1

#the data generating process
scores$read4<-(78
             +(-10)*scores$class_1+(-15)*scores$class_2+(-5)*scores$class_3
             +(-8)*scores$class_4+(-7)*scores$class_5+(-13)*scores$class_6
             +(11)*scores$class_7+(8)*scores$class_8+(10)*scores$class_9
             +(12)*scores$class_10
             +scores$error)

scores$year<-"2000"
scores00<-scores
rm(scores)

scores<-rbind(scores01, scores00)
```

A simulation:

```
regnodid<-felm(read4~treat,scores[scores$year=="2001",])

scores$post<-0
scores$post[scores$year=="2001"]<-1
regdid<-felm(read4~post+treat+post*treat,scores)

regdidfe<-felm(read4~post+treat+post*treat|class,scores)

## Warning in chol.default(mat, pivot = TRUE, tol = tol): the matrix is either
## rank-deficient or indefinite
```

A simulation:

```
stargazer(regnodid, regdid,regdidfe, type="latex", header=FALSE,
         add.lines = list(c("Class FE", "No", "No", "Yes")), omit.stat = c("ser","rsq", "adj.rsq"))
```

Table 4

	Dependent variable:		
	read4		
	(1)	(2)	(3)
post		7.612***	7.612***
		(1.073)	(1.023)
treat	27.482***	18.387***	
	(1.141)	(1.200)	
post:treat		9.095***	9.095***
		(1.697)	(1.617)
Constant	76.775***	69.163***	
	(0.721)	(0.759)	
Class FE	No	No	Yes
Observations	300	600	600
Note:		*p<0.1; **p<0.05; ***p<0.01	

Identifying Assumption:

Key assumption:

the difference between before and after in the comparison group is a good counterfactual for the treatment group.

- ▶ the trend in outcomes of the comparison group is what we would have observed in the treatment group absent the treatment

Identifying Assumption: Parallel Trends

Absent treatment, the outcome of the treated group would have followed a trend that was parallel to that of the control group
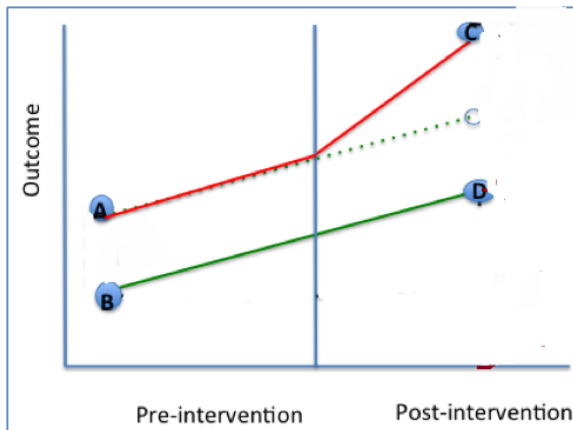

Figure 1: Green: Never Treated, Red: Treated

The Parallel Trends Assumption:

- ▶ We are treating the dashed green line as the counterfactual for the treated group

- ▶ Any deviation from this counterfactual is attributed to the treatment effect

- ▶ This assumption is straightforward but fundamentally **untestable**, because we will never actually observe this counterfactual of what would have happened to the treated group had they not been treated.

DID visually:

$$Y_{it} = \beta_0 + \beta_1 Post_t + \beta_2 GetsTreat_i + \beta_3 Post_t \times GetsTreat_i + u_{it}$$

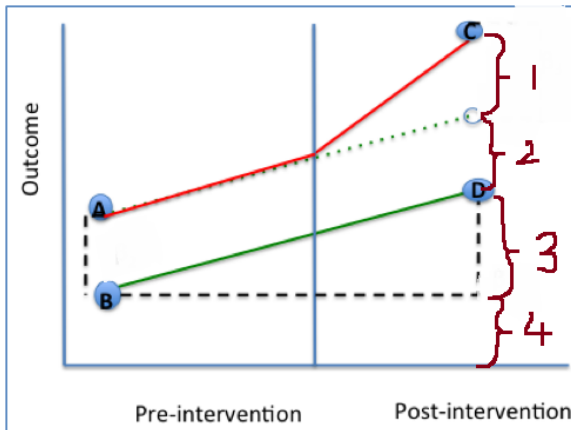Which number corresponds to which coefficient? ⇒ Top Hat

Figure 2: Green: Never Treated, Red: Treated

Identifying Assumption:

$$Y_{it} = \beta_0 + \beta_1 Post_t + \beta_2 GetsTreat_i + \beta_3 Post_t \times GetsTreat_i + u_{it}$$
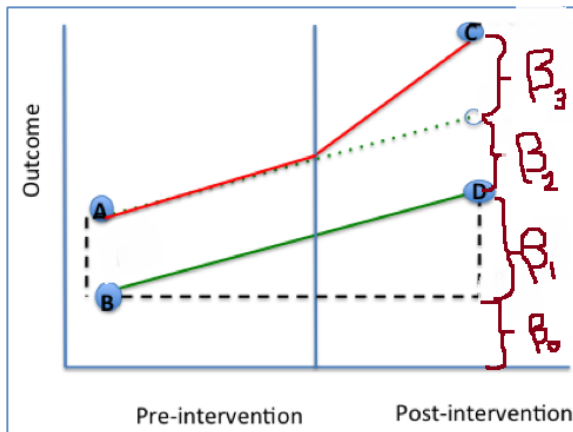

Figure 3: Green: Never Treated, Red: Treated

Problems with the Parallel Trends assumption

The parallel trends assumption is a fairly big assumption in many circumstances:

- ▶ policymakers may select treatment and control based on differences in the anticipated effects of treatment, or pre-existing differences in outcomes

In this case, the parallel trends assumption does not hold.

Example: Ashenfelter dips

Individuals who are "Treated" by job training programs are often individuals who just experienced a "dip" in earnings due to a job loss. When they get rehired their earnings increase substantially.

- ▶ If I compare their change in earnings to the change experienced by people who did not sign up for job training, I will see a large differential increase in earnings associated with program participation.

- ▶ This is due to who selects into training, not the causal effect of training.

Verifying Parallel trends

How can we check to see if the parallel trends assumption is likely to hold?

- ▶ Fundamentally untestable assumption

- ▶ use deduction to check this assumptions validity.

$\Rightarrow$ if the pre-treatment trends were parallel between the two groups, then wouldn't it stand to reason that the post-treatment trends would have been too?

Note: parallel pre-treatment trends does not **prove** that the assumption holds. But it does give some confidence that it does (absent some unobserved group specific time shock).

Verifying Parallel trends

Including leads into the DiD model is an easy way to check the pre-treatment trends. (Lags can also be included to see if treatment effect change over time).

Suppose treatment occurs right after period $t = 0$, estimate

$$Y_{it} = \beta_0 + \beta_2 GetsTreat_i + \sum_{t=-n}^{m} \beta_{3t} Period_t \times GetsTreat_i + \lambda_t + u_{it}.$$

If parallel trends holds:

▶ $E[\hat{\beta}_{3,t<1}] = 0$ since there is no treatment in these time periods.

▶ If $\tau \neq 0$, $E[\hat{\beta}_{3,t>0}] \neq 0$.

These results are typically best presented graphically.

Graphing DID estimates

Two main types of graphs

- ▶ plots of the mean outcome for both the treatment and control for several periods before and after treatment,

- ▶ and/or a graph that plots the $\hat{\beta}_{3t}$ estimates from the specification above.

Simulation: With leads and lags

```r
#simulating the data
set.seed(1999)
scoresbase<-as.data.frame(rep(c(1,2,3,4,5,6,7,8,9,10),times=30))
names(scoresbase)<-c("class")
scoresbase <- fastDummies::dummy_cols(scoresbase, select_columns = "class")

#suppose the better teachers (classes, 7,8,9,10)
#select to participate in the book club program
scoresbase$treat<-0
scoresbase$treat[scoresbase$class%in%c(7,8,9,10)]<-1
```

Simulation: With leads and lags

```r
#Gnerating simulated data for each year using a for loop
yr<-c(1995,1996,1997,1998,1999,2000,2001,2002,2003,2004,2005)
tauyr<-c(0,0,0,0,0,0,10,10,10,10,10)
yrfe<-c(72,77,75,79,81,79,83,77,82,84,81)

#loop to generate the data
for(i in 1:11){
  name<-paste("scores", yr[i], sep="_")
  scores<-scoresbase
  scores$error<-rnorm(300, mean=0, sd=10)
  tau<-tauyr[i]
  yearfe<-yrfe[i]
  #the data generating process
  scores$read4<-(yearfe+tau*scores$treat+scores$error
            +(-10)*scores$class_1+(-15)*scores$class_2+(-5)*scores$class_3
            +(-8)*scores$class_4+(-7)*scores$class_5+(-13)*scores$class_6
            +(11)*scores$class_7+(8)*scores$class_8+(10)*scores$class_9
            +(12)*scores$class_10)

  scores$year<-yr[i]
  assign(name, scores)
  rm(scores)
}

allscores<-rbind(scores_1995,scores_1996,scores_1997,scores_1998,scores_1999,
            scores_2000,scores_2001,scores_2002,scores_2003,scores_2004,scores_2005)
```

Simulation: With leads and lags

```
allscores$post<-0
allscores$post[allscores$year%in%c(2001,2002,2003,2004,2005)]<-1

allscores <- fastDummies::dummy_cols(allscores, select_columns = "year")

regdidall2<-felm(read4~treat
                +year_1995*treat+year_1996*treat+year_1997*treat+year_1998*treat
                +year_1999*treat+year_2001*treat+year_2002*treat+year_2003*treat
                +year_2004*treat+year_2005*treat,
                allscores)
```

Simulation: With leads and lags

`stargazer( regdidall2, type="latex",no.space=TRUE, header=FALSE, single.row=TRUE, omit.stat = "all")`

Table 5

	Dependent variable:
	read4
treat	18.774*** (1.207)
year_1995	−5.842*** (1.080)
year_1996	−2.945*** (1.080)
year_1997	−1.696 (1.080)
year_1998	0.027 (1.080)
year_1999	2.636** (1.080)
year_2001	5.125*** (1.080)
year_2002	−2.235** (1.080)
year_2003	1.849* (1.080)
year_2004	4.430*** (1.080)
year_2005	2.533** (1.080)
treat:year_1995	0.431 (1.707)
treat:year_1996	1.427 (1.707)
treat:year_1997	−2.476 (1.707)
treat:year_1998	0.568 (1.707)
treat:year_1999	−0.573 (1.707)
treat:year_2001	10.204*** (1.707)
treat:year_2002	12.290*** (1.707)
treat:year_2003	11.188*** (1.707)
treat:year_2004	11.186*** (1.707)
treat:year_2005	10.217*** (1.707)
Constant	69.542*** (0.764)

Note:	*p<0.1; **p<0.05; ***p<0.01

Simulation: With leads and lags

```r
#start with plot of group means

#calculateing the mean score for each year by treatment status
grp_mean<-allscores%>%
    group_by(year,treat)%>%
    dplyr::summarize(groupmean = mean(read4, na.rm=TRUE))


## `summarise()` regrouping output by 'year' (override with `.groups` argument)

grp_mean$treat<-as.factor(grp_mean$treat)

#difference in means plot
didmeans<-ggplot(grp_mean, aes(year, groupmean, group=treat, color = treat)) +
    stat_summary(geom = 'line') +
    geom_vline(xintercept = 2000) +
    theme_minimal()
```

Simulation: With leads and lags

Simulation: With leads and lags

```
#plot of differences coefficients

res<-coef(summary(regdidall2))
res<-as.data.frame(res)

res<-res[13:22,]

a<-c(0,0,0,0)

res<-rbind(res,a)

year<-c(1995,1996,1997,1998,1999,2001,2002,2003,2004,2005,2000)
res<-cbind(res,year)
res$ci<-1.96*res$`Std. Error`

names(res)<-c("Estimate","se", "t",  "p", "year", "ci")
# Use 95% confidence interval instead of SEM
didplot2<-ggplot(res, aes(x=year, y=Estimate)) +
    geom_errorbar(aes(ymin=Estimate-ci, ymax=Estimate+ci),width=.1) +
    geom_vline(xintercept = 2000)+
      geom_hline(yintercept = 0)+
       geom_point()
```
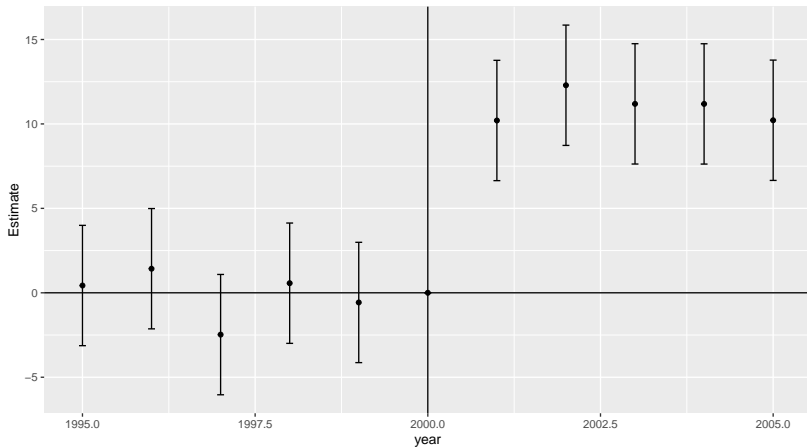
Simulation: With leads and Lags

A note on Inference: Clustering

What if the variable of interest only varies at a "group" level

- ex: at the state or class level
- ⇒ observations are related with each other within certain groups
- There is less variation then we actually think

Conventional standard errors will often severely understate the standard deviation of the estimators and lead to inference errors.

Solution: cluster standard errors at the group level.

A note on Inference: Clustering

Clustering standard errors can be done in R using the felm function:

```
regdidall3<-felm(read4~treat
                 +year_1995*treat+year_1996*treat+year_1997*treat+year_1998*treat
                 +year_1999*treat+year_2001*treat+year_2002*treat+year_2003*treat
                 +year_2004*treat+year_2005*treat
                 |0
                 |0
                 |class,
                 allscores)
```

A note on Inference: Clustering

`stargazer( regdidall2,regdidall3, type="latex",no.space=TRUE, single.row=TRUE, omit.stat = "all", header=`

Table 6

	Dependent variable:	
	read4	
	(1)	(2)
treat	18.774*** (1.207)	18.774*** (2.268)
year_1995	−5.842*** (1.080)	−5.842*** (0.879)
year_1996	−2.945*** (1.080)	−2.945*** (1.070)
year_1997	−1.696 (1.080)	−1.696 (1.032)
year_1998	0.027 (1.080)	0.027 (0.257)
year_1999	2.636** (1.080)	2.636** (1.237)
year_2001	5.125*** (1.080)	5.125*** (0.941)
year_2002	−2.235** (1.080)	−2.235 (1.582)
year_2003	1.849* (1.080)	1.849*** (0.654)
year_2004	4.430*** (1.080)	4.430*** (1.206)
year_2005	2.533** (1.080)	2.533*** (0.746)
treat:year_1995	0.431 (1.707)	0.431 (1.768)
treat:year_1996	1.427 (1.707)	1.427 (2.108)
treat:year_1997	−2.476 (1.707)	−2.476* (1.287)
treat:year_1998	0.568 (1.707)	0.568 (1.570)
treat:year_1999	−0.573 (1.707)	−0.573 (1.477)
treat:year_2001	10.204*** (1.707)	10.204*** (1.602)
treat:year_2002	12.290*** (1.707)	12.290*** (2.310)
treat:year_2003	11.188*** (1.707)	11.188*** (1.824)
treat:year_2004	11.186*** (1.707)	11.186*** (2.216)
treat:year_2005	10.217*** (1.707)	10.217*** (1.510)
Constant	69.542*** (0.764)	69.542*** (1.965)

Note: *p<0.1; **p<0.05; ***p<0.01

Falsification Tests: Alternative outcomes

In addition to checking for parallel pre-trends, you can also check and see if there is a pattern of parallel trends for an alternative outcome measure that should not be affected by treatment.

For the ERS example,

▶ estimate the same model with manslaughter or male murders as the outcome variable.

▶ these variables would be affected by the underlying violent crime in the city, but should be unaffected by ERS.

Falsification Tests: Triple differences

Adding another layer of differencing to the estimator (a triple-diff estimator) can make DID results more compelling.

In our simulated example:

- ▶ Suppose a subset of students already had access to the book club through an after school program.

- ▶ We can use the students in the after school program as an additional "control" group, since their performance should not change when the book club is introduced to the classrooms.

Falsification Tests: Triple differences

Let $\bar{Y}_{a,g,t}$ represent the mean reading score of students

- ▶ in group g (control or treatment classes)

- ▶ in year t

- ▶ that are ($a = AS$) or are not ($a = NAS$) in the after school program.

The triple difference estimator is then

$$[(\bar{Y}_{T,1,AS} - \bar{Y}_{T,1,NAS}) - (\bar{Y}_{T,0,AS} - \bar{Y}_{T,0,NAS})] - [(\bar{Y}_{C,1,AS} - \bar{Y}_{C,1,NAS}) - (\bar{Y}_{C,0,AS} - \bar{Y}_{C,0,NAS})]$$

$\Rightarrow$ we compare the evolution of the gap between the after school kids and the others in the treated classrooms to the evolution of the gap between after school kids and the others in the control classrooms.

Falsification Tests: Triple differences

Triple difference estimation allows us to relax some assumptions:

- ▶ We no longer need to assume that outcomes for treated and control would evolve similarly in expectation-

- ▶ we now only need to assume that, to the extent that outcomes evolve differently in treated and control classes, the difference affects participants and non-participants in the after school program similarly.

Triple differences

Triple diff regressions: Lots of interation terms!

- ▶ Put in an indicator for every main effect and interaction up to, but not including, the level at which the treatment varies.

- ▶ We include the main effects for time, class and after-school program participation as well as all possible two way interactions between each of these indicators.

The regression looks like this:

$$
\begin{aligned}
Y_{i,a,g,t} =& \alpha_0 + \alpha_1 GetT_g + \alpha_2 AS_a + \alpha_3 Post_t \\
& + \alpha_4 (GetT \times Post)_{gt} + \alpha_5 (GetT \times AS)_{ga} + \alpha_6 (Post \times AS)_{ta} \\
& + \alpha_7 (GetT \times Post \times AS)_{agt} + u_{igta}
\end{aligned}
$$

Triple differences

Where will we see the treatment effect?

- Students who are not in the after school program ($AS = 0$)
- When the the treated classes ($GetT = 1$) adopt the book club
- When treatment is implemented $Post = 1$,

$\Rightarrow$ we would expect the treatment effect to manifest as:

- $\alpha_4 > 0$
- $\alpha_4 + \alpha_7 = 0$ (since the after school kids experience no change in the post period).

Triple differences: Simulation

I simulate a new set of reading scores that based on a DGP that includes the after school program effects.

```r
set.seed(123456)
scores<-as.data.frame(rep(c(1,2,3,4,5,6,7,8,9,10),times=30))
names(scores)<-c("class")
scores <- fastDummies::dummy_cols(scores, select_columns = "class")

scores$error<-rnorm(300, mean=0, sd=10)

#a random indicator for participation in the afterschool program
scores$aftsch<-rbinom(300,1,0.5)

#suppose the best teachers (classes, 7,8,9,10)
#select into participating in the book club program
scores$treat<-0
scores$treat[scores$class%in%c(7,8,9,10)]<-1

#creating an indicator for students who are become treated
#and are not in the after school program
scores$treatnotaftsch<-0
scores$treatnotaftsch[scores$treat==1 & scores$aftsch==0]<-1

#the treatment effect
tau<-10

#the data generating process
scores$read4<-(85+13*scores$aftsch+tau*scores$treatnotaftsch+scores$error
              +(-10)*scores$class_1+(-15)*scores$class_2+(-5)*scores$class_3
              +(-8)*scores$class_4+(-3)*scores$class_5+(3)*scores$class_6
              +(5)*scores$class_7+(8)*scores$class_8+(10)*scores$class_9+(12)*scores$class_10)

scores$year<-"2001"
```

Triple differences: Simulation

```r
scores<-as.data.frame(rep(c(1,2,3,4,5,6,7,8,9,10),times=30))
names(scores)<-c("class")
scores <- fastDummies::dummy_cols(scores, select_columns = "class")

scores$error<-rnorm(300, mean=0, sd=10)

#a random indicator for participation in the afterschool program
scores$aftsch<-rbinom(300,1,0.5)

#suppose the best teachers (classes, 7,8,9,10)
#select into participating in the book club program
scores$treat<-0
scores$treat[scores$class%in%c(7,8,9,10)]<-1

#creating an indicator for students who are become treated
#and are not in the after school program
scores$treatnotaftsch<-0
scores$treatnotaftsch[scores$treat==1 & scores$aftsch==0]<-1

#the data generating process
scores$read4<-(78+18*scores$aftsch+scores$error
              +(-10)*scores$class_1+(-15)*scores$class_2+(-5)*scores$class_3
              +(-8)*scores$class_4+(-3)*scores$class_5+(3)*scores$class_6
              +(5)*scores$class_7+(8)*scores$class_8+(10)*scores$class_9
              +(12)*scores$class_10)

scores$year<-"2000"

scores00<-scores
rm(scores)

scores<-rbind(scores01, scores00)
```

Triple differences: Simulation

```
scores$post<-0
scores$post[scores$year=="2001"]<-1

regdid3d<-felm(read4~post+treat+post*treat|0|0|class,scores)

regdid3dinter<-felm(read4~post+treat+aftsch
                    +post*treat+treat*aftsch+post*aftsch
                    +post*aftsch*treat
                    |0|0|class,
                    scores)
```

Triple differences: Simulation

```
stargazer( regdid3d,regdid3dinter, type="latex",no.space=TRUE,  omit.stat = "all", header=FALSE)
```

Table 7

	Dependent variable:	
	read4	
	(1)	(2)
post	4.868***	7.850***
	(1.197)	(1.907)
treat	15.815***	14.287***
	(3.709)	(2.639)
aftsch		19.479***
		(1.764)
post:treat	4.640**	11.045***
	(2.071)	(2.323)
treat:aftsch		−0.585
		(2.469)
post:aftsch		−8.220***
		(2.609)
post:treat:aftsch		−9.339***
		(3.129)
Constant	80.400***	71.851***
	(3.353)	(2.446)
Note:	*p<0.1; **p<0.05; ***p<0.01	

Triple differences: Simulation

The simple DID specification underestimates of the true treatment effect.**Why?**

- ▶ the coefficient gives the **ATE** on the treated classes in the post period. Since 50% of students were already treated through the after school program the ATE is halved.

- ▶ Recall: the average treatment effect can hide vast differences in treatment effects across groups.

- ▶ To reveal these, we need to add additional interaction terms.

With the triple difference specification:

- ▶ $E[\hat{\alpha}_4] = \tau = 10$,

- ▶ and $E[\hat{\alpha}_7] = -\tau = -10$ (since the "new" treatment does not affect the after school participants).

Triple differences: As Robustness test

The triple difference shows that the jump in test scores in 2001 is driven by students who just gained access to the book club.

Students **in the same classes**, whose access to the book club did not change, did not experience this jump in scores.

Their performance continues to parallel that of the control group.

This is convincing because

- ▶ it provides support for the parallel trends assumption and

- ▶ it helps rule out that something else changed in the classes that select into treatment that could explain the shift in scores.

Generalized DID: Staggered events

Sometimes we have a "pure control" that never gets treated and can serve as a counterfactual to treatment.

Other times we do not have a true control, but a treatment is rolled out over a number of months/years across various treatment units.

Researchers have used the staggered nature of the roll out to identify the causal effects.

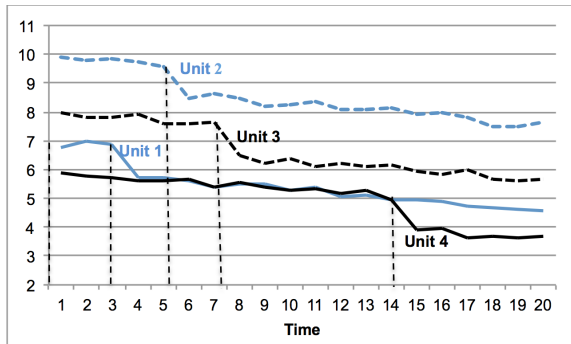# Generalized DID: Staggered events

Figure 4: Staggered roll out

Unit 1 is treated at time 3, unit 2 at time 5, unit 3 at time 7, etc.

Untreated units at time t serve as the control group for the units that are being treated at time t.

Generalized DID: Staggered events

In a regression framework, we would estimate a two-way FE DID,

$$y_{it} = \beta_0 + \beta D_{it} + \gamma_i + \phi_t + u_{it}$$

The key assumption:

- ▶ changes in the control group are a good counterfactual for the change in the treatment group

- ▶ test these assumptions are very similar to the methods we have already discussed.

The main difference:

- ▶ treatment does not happen at a fixed point in time.

- ▶ often need to generate a new set of "relative to event time" indicators that are centered around the treatment time of a particular unit.

Staggered events: WARNING

Our understanding of staggered event DID estimations is evolving.

Several recent new papers reassessing these estimation strategies:

- Goodman-Bacon (2018, 2019):
    - This "general" DD estimator is a weighted average of all possible two-group/two-period DD estimators in the data: treated units act as both controls and treatment depending on the situation
    - How the weighting is done is not clearly theoretically justified
    - Estimates are potentially biased when effects change over time.

Exercise caution when using this type of approach as our understanding of these estimators, and how to correct for their problems is rapidly evolving.