Difference-in-Differences Models

Applied Econometrics for Researchers, PhD

Vera Rocha, CBS-SI, vr.si@cbs.dk
Alba Marino, UniME, alba.marino@unime.it

7th December 2022

Agenda for today

1. What are difference-in-differences models?
2. Key «ingredients»
3. A textbook example (also using Stata)
4. Testing assumptions
5. Potential extensions on evaluation models

Key readings:

- Angrist and Pischke, «Mostly Harmless Econometrics: An Empiricist's Companion», **Chapter 5**
- Or: Scott Cunningham, «The Mixtape», Chapter 9, Difference-in-Differences

Matching methods vs. Difference-in-Differences

	Matching Models (PSM)	Difference-in-Differences (DiD)
When	We need to assess the effect of a "treatment" (e.g., choice, policy)	We need to assess the effect of a "treatment" (e.g., choice, policy)
Problem	T & C groups are very different ("selection on observables"); cross-section, no panel data	Requires data "before" and "after" the treatment + "treated" & "control" units. Is the "treatment" exogenous (e.g. natural experiment?)
Stata commands	*teffects psmatch, tebalance, teffects overlap*	Regular OLS regression or panel regression (though there are automatic Stata commands you can explore)
Key tests	Balancing and overlapping conditions (quality checks after PSM)	Parallel trends assumption (before the "intervention"), i.e. T and C's outcomes trajectories are "parallel"
Attention!	T & C only matched on **observable** characteristics. If unobservables matter, PSM does not provide causal effects → IV or panel DiD	If parallel trends assumption is violated, we may need to construct matched samples first and then run a DiD regression
First stage	Probit predicting assignment to "treatment" (X)	Obtain matched samples of treated and control units (via PSM/CEM etc)

Example 1: Equal but Inequitable

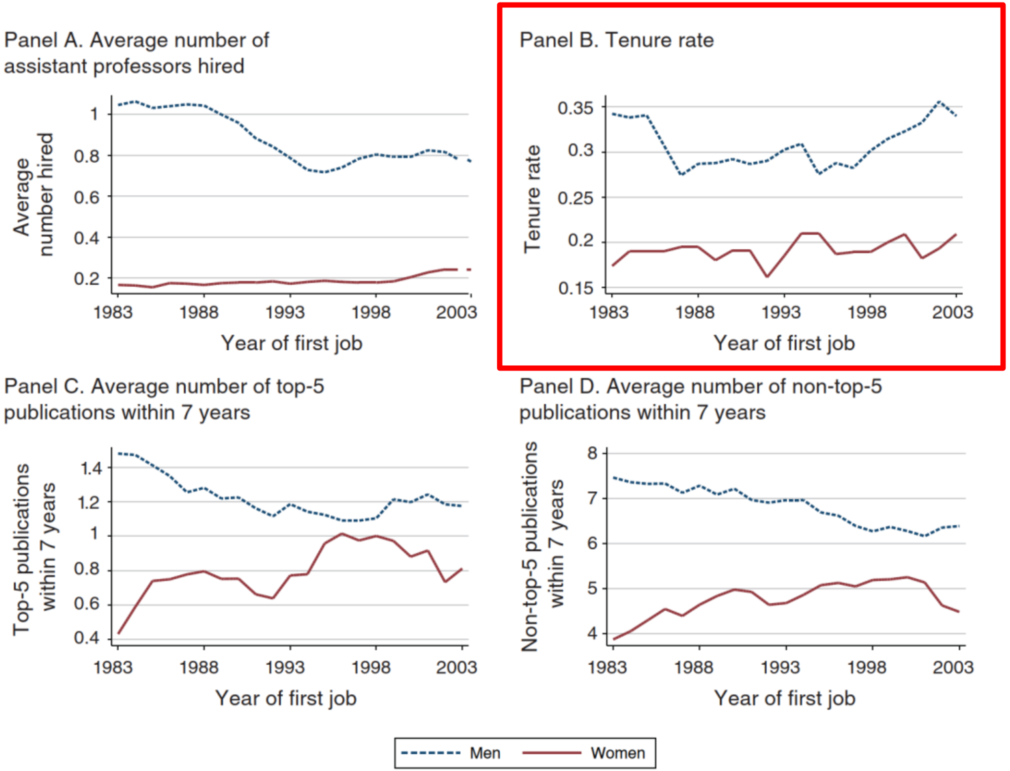


FIGURE 2. TRENDS IN TOP-50 ECONOMICS DEPARTMENTS

Tenure clock stopping policies:

Assistant professors are allowed to stop their tenure clock for 1 year after childbirth or adoption. No research is expected during this time.

Do **gender-neutral** clock stopping (GNCS) policies level the playing field in terms of tenure rates (i.e. do the effects differ for M and W and affect the gender gap)?

Top-50 Econ departments in the US; some universities adopted them (at different points in time), others didn't.

Antecol, H., Bedard, K., & Stearns, J. (2018). Equal but inequitable: Who benefits from gender-neutral tenure clock stopping policies?. *American Economic Review*, *108*(9), 2420-41.

Example 1: Equal but Inequitable

	Total effects (1)	Male-female (2)
Panel A. Policy effects years 0–3		
Men FOCS	−0.008	−0.181
	(0.067)	(0.140)
Women FOCS	0.172	
	(0.140)	
Men GNCS	0.051	0.068
	(0.079)	(0.145)
Women GNCS	−0.017	
	(0.107)	
Panel B. Policy effects years 4+		
Men FOCS	0.002	−0.047
	(0.075)	(0.128)
Women FOCS	0.049	
	(0.101)	
Men GNCS	0.176	0.370
	(0.083)	(0.146)
Women GNCS	−0.194	
	(0.106)	
Sample size	1,392	

DV = 1 if the individual makes tenure, 0 if not. GNCS = Gender Neutral tenure clock stopping policy/FOCS = Female only. Second column shows difference in the male and female coefficients for each policy type. Std. errors clustered at policy-university level. Examples of controls: time-varying university characteristics, PhD rank, having done a postdoc… Gender-specific university FE included.

a) Men whose first job was at a top-50 university with a gender-neutral tenure clock stopping policy in place for 4+ years have a **17.6 percentage point tenure rate advantage** over men at the same university prior to the implementation of any policy (diff: male T & C)

b) Women, in turn, are **19.4 percentage points less likely to get tenure** relative to other women hired by the same university prior to the clock stopping policy (diff: female T & C)

c) This **increased the gender gap between men and women by 37 percentage points**: women are even less likely to get tenure compared to men after the introduction of GNCS policies. (diff-in-diff)

Antecol, H., Bedard, K., & Stearns, J. (2018). Equal but inequitable: Who benefits from gender-neutral tenure clock stopping policies?. *American Economic Review*, *108*(9), 2420-41.

Example 2: Inventor death and innovation

If the collaboration between two patent inventors were to exogenously end, would this have a significant and long-lasting impact on the career, compensation, and patents of co-inventors? Or are co-inventors easily substituted for, beyond short-term disruption of ongoing work?

Data

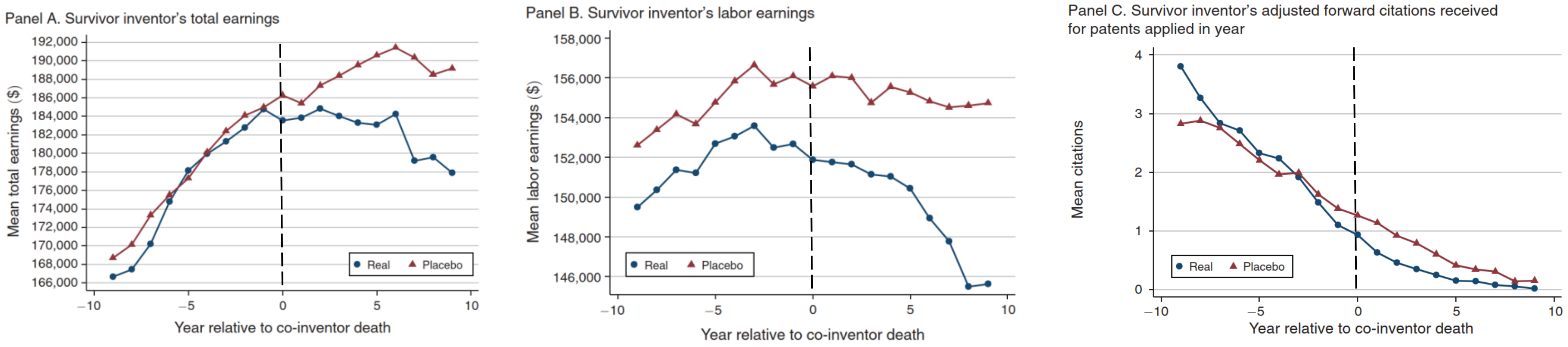
- USPTO patents data and Treasury administrative tax data
- Some inventors die suddenly before or at the age of 60 (4,714 inventors): <u>exogenous shock in collaborative networks</u>
- Compare inventors whose co-inventors did not pass away but who are otherwise similar to inventors who experienced the premature death of a co-inventor **(i.e. matched sample of inventors with and without loss of a team member – treated vs control groups)**

Jaravel, X., Petkova, N., & Bell, A. (2018). Team-specific capital and innovation. *American Economic Review*, *108*(4-5), 1034-73.

Example 2: Inventor death and innovation

Ending a collaboration causes a large and long-lasting decline in an inventor's:

- labor earnings (−3.8 percent after 8 years)
- total earnings (−4 percent after 8 years)
- citation-weighted patents (−15 percent after 8 years)



Note: "control inventors" are exactly matched to "real inventors" in age, year, and total nr of patent applications at the time of (real/control) death to secure parallel trends prior to the "shock"

Jaravel, X., Petkova, N., & Bell, A. (2018). Team-specific capital and innovation. *American Economic Review*, *108*(4-5), 1034-73.

Example 3: Why Marathons Can Be Deadly

Large marathons frequently involve widespread road closures and infrastructure disruptions, which may create delays in emergency care for individuals with acute medical conditions who live in proximity to marathon routes (“treated” by this exogenous shock).

Data

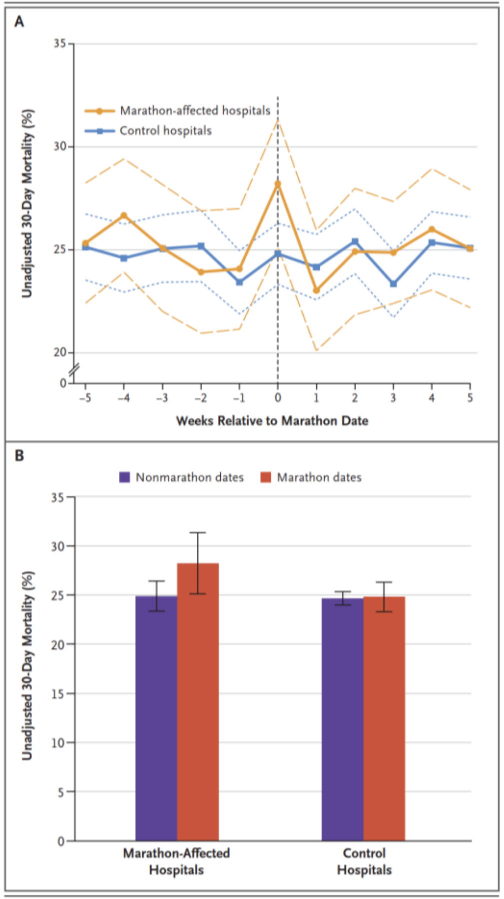
- Hospitalizations for acute myocardial infarction or cardiac arrest (age ≥ 65)
- 11 U.S. cities that hosted marathons (2002-2012)
- Mortality of those hospitalized the day of the marathon vs. those hospitalized
  - in the same week day but 5 weeks bef/after the marathon
  - In the same day but in surrounding ZIP codes unaffected by the marathon

Podcast available in Freakonomics M.D.

Summary here: https://time.com/4736467/marathons-running-heart-attack/

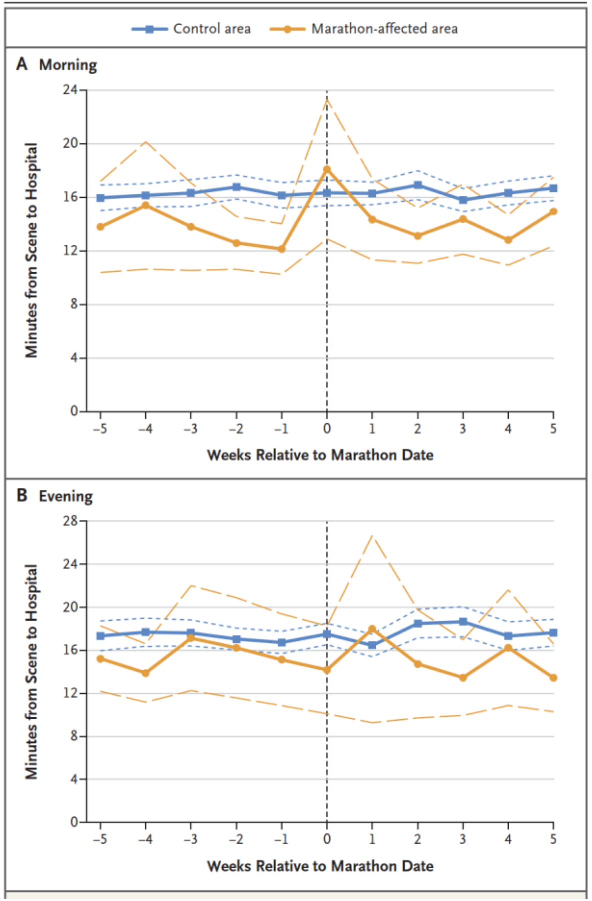
Jena, A. B., Mann, N. C., Wedlund, L. N., & Olenski, A. (2017). Delays in emergency care and mortality during major US marathons. *New England Journal of Medicine*, *376*(15), 1441-1450.

Example 3: Why Marathons Can Be Deadly



People who were admitted to marathon-affected hospitals on marathon dates had:

- longer ambulance transport times before noon (4.4 minutes longer)
- higher 30-day mortality than beneficiaries who were hospitalized on non-marathon dates
- higher 30-day mortality than those who were hospitalized on the same day as the marathon but in unaffected surrounding ZIP code areas



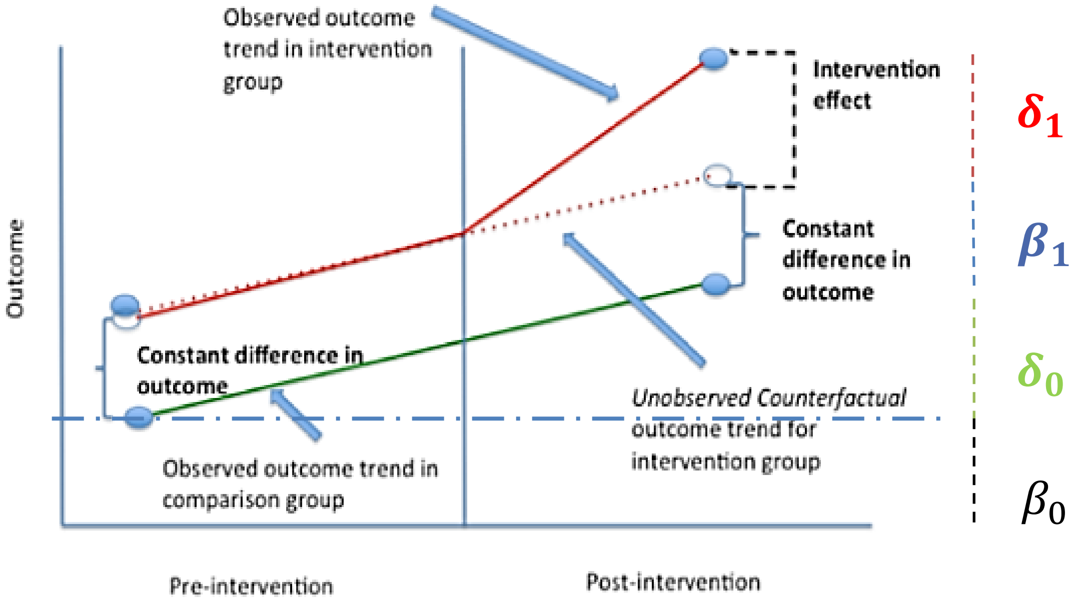
Jena, A. B., Mann, N. C., Wedlund, L. N., & Olenski, A. (2017). Delays in emergency care and mortality during major US marathons. *New England Journal of Medicine*, *376*(15), 1441-1450.

«Parallel worlds»: key ingredients

Time	The "**treatment**" – i.e., policy change, intervention, event – takes place at a certain point in time (pooled cross-section or panel data)
Policy change or treatment	We identify (*at least one*) **before** and **after period** with respect to the treatment
Comparison groups	**Treated group** receives the intervention or is subject to the policy change only in the post-period **Control group** is not affected by the treatment
Fixed factors	Assume that important factors associated with the outcome Y are fixed during the pre- and post-periods
Time invariant factors	If observed, we can control for those factors that could affect trends and vary over time (parallel trends or constant bias)

More formally, with a textbook example:

$$y = \beta_0 + \boldsymbol{\beta_1} dTreated + \boldsymbol{\delta_0} dPost + \boldsymbol{\delta_1} dTreated \cdot dPost + controls + u$$



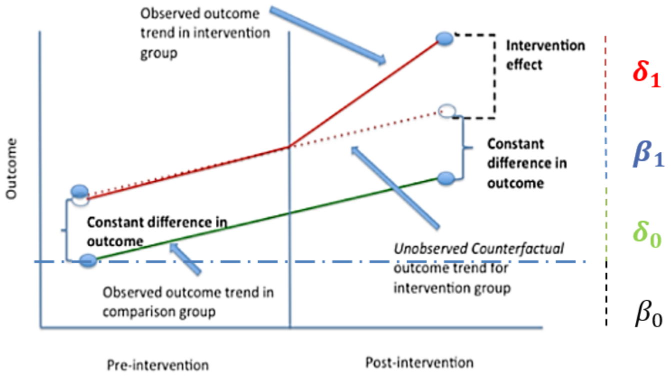
More formally, with a textbook example:

$$y = \beta_0 + \boldsymbol{\beta_1} dTreated + \boldsymbol{\delta_0} dPost + \boldsymbol{\delta_1} dTreated \cdot dPost + controls + u$$

	Before	After	After-Before
Control	β_0	$\beta_0 + \delta_0$	δ_0
Treated	$\beta_0 + \beta_1$	$\beta_0 + \delta_0 + \beta_1 + \delta_1$	$\delta_0 + \delta_1$
Treated-Control	β_1	$\beta_1 + \delta_1$	δ_1

$\boldsymbol{\delta_1}$ is the "average treatment effect" (ATE)

$$\widehat{\delta_1} = (\bar{y}_{2,T} - \bar{y}_{2,C}) - (\bar{y}_{1,T} - \bar{y}_{1,C})$$



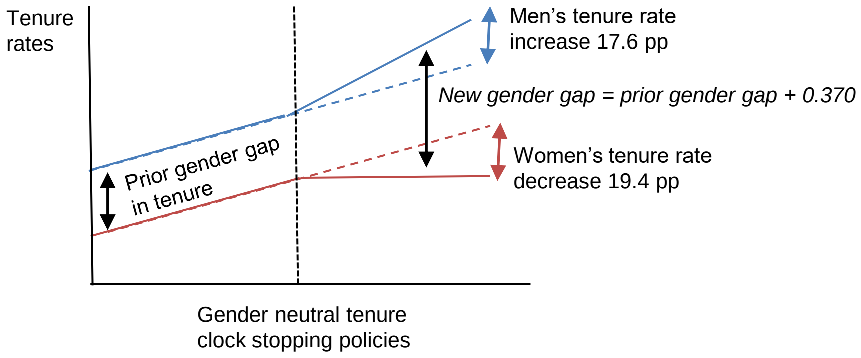
Recalling the tenure-gender example :

Recall slide 5

	After policy change in tenure rates
Men	0.176
Women	-0.194
Gender gap	0.370

$\boldsymbol{\delta_1}$ is the "average treatment effect" (ATE)

$$\widehat{\delta_1} = 0.176 - (-0.194) = 0.370$$



DiD Assumptions

COMMON TRENDS OR CONSTANT BIAS	Both treatment and control groups would have the same trends over time (possibly conditioning for other factors) if the intervention had not happened. Treated and control groups are not equivalent, but their difference remains constant over time.
RANDOM ASSIGNMENT INTO TREATMENT	The policy/intervention must ”naturally” affect certain group of subjects (firms, persons, families), but not all. Alternatively, the treatment may affect all but its effects may differ across groups (e.g. slide 5). Manipulation is the key: “No causation without (explicit, natural, or whatever) manipulation”
EXOGENEITY	The covariates X are not influenced by the treatment.
SUTVA	No interference (spillovers/externalities) and variation in treatment among the groups.

Example

- Introduction of a minimum wage policy (Card and Krueger, 1994)
- In a nutshell: in theory, a firm makes hiring decisions based on wages and the contribution of employees to revenue
- In a perfectly competitive market, a higher minimum wage implies that firms will demand fewer workers (or hours worked)
- Thus, a policy that helps those who can get jobs at the higher wage may harm some other workers, who won't find employment because of the higher minimum wage

Authors use a dramatic change in the New Jersey state minimum wage to test whether this is true.

Example

- In 1992, New Jersey raised the state minimum wage by about 19% from $4.25 to $5.05
- Card and Krueger (1994) obtained data from February 1992 (“before” or pre-period) and November 1992 (“after” or post-period) from fast-food chains*, which usually pay minimum wages (TREATED)
- They collected data from similar fast-food restaurants in eastern Pennsylvania – just across the Delaware river –, which did not change the minimum wage ($4.25) (CONTROL)

We have the key elements of a basic two-period DiD:

- observations over time (both pre- and post-treatment)
- exogeneous and random treatment
- policy change applies only to treated group in the post-period
- the control group is not exposed to the experiment during the time

* KFC, Burger King, Roy Roger’s, Wendy’s

Data structure

For each store (id), we have two observations (balanced two-year panel). Some are treated, some are in the control group; both have 2 periods

```
.
. list id fte treated post in 1/10, sep(2) nolabel
```

	id	fte	treated	post
1.	1	16	1	0
2.	1	20	1	1
3.	2	10	1	0
4.	2	7.5	1	1
5.	3	6	1	0
6.	3	4	1	1
7.	4	10	1	0
8.	4	5	1	1
9.	5	5	1	0
10.	5	10	1	1

```
.
. d id treated post fte chain
```

variable name	storage type	display format	value label	variable label
id	float	%9.0g		Restaurant ID
treated	float	%9.0g	treated	NJ = 1; PA = 0
post	float	%9.0g	post	Feb.92 = 0; Nov. 1992 = 1
fte	float	%9.0g		Output: Full Time Employees
chain	float	%9.0g		Burger King = 1; KFC = 2; Roys = 3; Wendy's = 4

We want to estimate the **causal effect of *x (treated)* on *y (fte)***, keeping other things equal.

Interpretation of the results

- Restaurants in PA employed on average 10 FTEs before the policy change ($\boldsymbol{\beta_0}$)
- Restaurants in NJ employed on average 2.6 fewer FTEs than PA before the policy change ($\boldsymbol{\beta_1}$) – significant at 5%
- There is no average significant effect (at the 5% significance level) for restaurants in PA on number of FTEs after the introduction of the policy ($\boldsymbol{\delta_0}$)
- The average number of FTEs in NJ increased by 3,44 FTEs units after the policy change with respect to PA – in other words, due to the increase in minimum wages ($\boldsymbol{\delta_1}$)

$$y = \beta_0 + \boldsymbol{\beta_1} dTreated + \boldsymbol{\delta_0} dTime + \boldsymbol{\delta_1} dTreated \cdot dTime + controls + u$$

```
. reg fte i.treated##i.post, robust

Linear regression                               Number of obs     =        784
                                                F(3, 780)         =       1.56
                                                Prob > F          =     0.1970
                                                R-squared         =     0.0084
                                                Root MSE          =     8.3213

------------------------------------------------------------------------------
             |               Robust
         fte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |
         NJ  |    -2.6006   1.319187    -1.97   0.049    -5.190177   -.0110226
      1.post |  -2.743421   1.578217    -1.74   0.083    -5.841476     .354634
             |
treated#post |
       NJ#1  |   3.442788   1.700103     2.03   0.043     .1054694    6.780107
             |
       _cons |   10.31579   1.239793     8.32   0.000     7.882063    12.74952
------------------------------------------------------------------------------
```

	Before	After	After-Before
Control	β_0	$\beta_0 + \delta_0$	δ_0
Treated	$\beta_0 + \beta_1$	$\beta_0 + \delta_0 + \beta_1 + \delta_1$	$\delta_0 + \delta_1$
Treated-Control	β_1	$\beta_1 + \delta_1$	δ_1

Estimate and compare means

```
.
. * Treated
. sum fte if treated ==1 & post==1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         fte |        316    8.414557    7.870619          0         40

.
. scalar y_tpost = r(mean)

.
. sum fte if treated ==1 & post==0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         fte |        316     7.71519    8.004734          0         60

.
. scalar y_tpre = r(mean)

.
. * Control
. sum fte if treated ==0 & post==1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         fte |         76    7.572368    8.548179          0         35

.
. scalar y_cpost = r(mean)

.
. sum fte if treated ==0 & post==0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         fte |         76    10.31579    10.85229          0         50

.
. scalar y_cpre = r(mean)
```

$\boldsymbol{\delta_1}$ is the "average treatment effect" (ATE)

- $\widehat{\delta_1} = (\bar{y}_{2,T} - \bar{y}_{2,C}) - (\bar{y}_{1,T} - \bar{y}_{1,C})$

```
.
. * DiD estimator
. di y_tpost - y_tpre - (y_cpost - y_cpre)
3.4427881
```

Adding controls

- Adding control variables reduces the residual variance, which in turn lowers the standard error of the regression estimates.
- If the treatment is really random, the point estimate should not change by adding more controls. Often this is seen as a robustness check on the claim of random assignment.
- Here, we are comparing the same restaurants before and after, so chain type couldn’t affect the difference-in-differences because it is not changing over time.

```
. reg fte i.treated##i.post, robust

Linear regression                               Number of obs     =        784
                                                F(3, 780)         =       1.56
                                                Prob > F          =     0.1970
                                                R-squared         =     0.0084
                                                Root MSE          =     8.3213

------------------------------------------------------------------------------
             |               Robust
         fte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |
         NJ  |    -2.6006   1.319187    -1.97   0.049    -5.190177   -.0110226
      1.post |  -2.743421   1.578217    -1.74   0.083    -5.841476     .354634
             |
treated#post |
       NJ#1  |   3.442788   1.700103     2.03   0.043     .1054694    6.780107
             |
       _cons |   10.31579   1.239793     8.32   0.000     7.882063    12.74952
------------------------------------------------------------------------------
```

```
. reg fte i.treated##i.post i.chain, robust

Linear regression                               Number of obs     =        784
                                                F(6, 777)         =      12.19
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0607
                                                Root MSE          =     8.1144

------------------------------------------------------------------------------
             |               Robust
         fte |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |
         NJ  |   -2.30713   1.258303    -1.83   0.067    -4.777207    .1629459
      1.post |  -2.743421   1.509528    -1.82   0.070    -5.706658    .2198161
             |
treated#post |
       NJ#1  |   3.442788   1.632124     2.11   0.035     .2388929    6.646683
             |
       chain |
        KFC  |  -5.123389   .6552841    -7.82   0.000    -6.409726   -3.837052
       Roys  |  -1.690295   .7306062    -2.31   0.021    -3.124491   -.2560989
    Wendy's  |   -1.00196    1.05185    -0.95   0.341    -3.066764    1.062845
             |
       _cons |   11.67423   1.289992     9.05   0.000     9.141944    14.20651
------------------------------------------------------------------------------
```

Adding unit and time FEs

So far, we have seen the standard 2x2 DiD regression model:

$$y = \beta_0 + \beta_1 dTreated + \delta_0 dPost + \delta_1 dTreated \cdot dPost + u$$

We can extend the model adding unit (λ_i) and time (μ_t) fixed effects

$$y = \beta_0 + \beta_1 dTreated + \delta_0 dPost + \delta_1 dTreated \cdot dPost + \lambda_i + \mu_t + u$$

λ_i captures time-invariant characteristics, including those that could affect self-selection (or assignment) into the treatment/program

μ_t identifies time-variant characteristics, regardless of the group the individual belongs to

If the treatment is really random, the point estimate should not change by adding FEs.

Adding control variables

We can also control for variables that we think could affect the evolution of the trends between treatment and control groups (remember, variables that affect the difference between trends):

$$y = \beta_0 + \beta_1 dTreated + \delta_0 dPost + \delta_1 dTreated \cdot dPost + \boldsymbol{\theta} \mathrm{X}_{it} + \lambda_i + \mu_t + u$$

Control variables can significantly affect the "treatment effect"!

If some variables in X do not change over time, they won't affect the DiD estimator

Checking assumptions I

Recall slide 14

COMMON TRENDS OR CONSTANT BIAS

RANDOM ASSIGNMENT INTO TREATMENT

EXOGENEITY

SUTVA

There are some “exclusion restrictions” that you need to discuss:

For example, exogeneity or the SUTVA assumption

When is the SUTVA assumption violated?

Whenever it is not possible to argue that the effect of the treatment affects only treated individuals (i.e., $\nexists$ externalities or spillovers). For example, there are interactions between individuals in the treatment and in the control group.

What about exogeneity?

The classical problem of OLS biased estimates due to various sources of endogeneity potentially occurring together so that sometimes it is not easy to distinguish between them.

Checking assumptions II

Other assumptions can be tested with data, or at least you can see if the data seems consistent with an assumption, even though it may not guarantee that the assumption is valid

- Graphically showing that trends are parallel is a good first step, but we can test the assumption as well
- We will see adjusted and unadjusted plots – sometimes we may need to adjust for factors that affect the difference in trends

DiD vs. Experiments

In order to mimic a **randomized controlled trial,** we need a ”**natural” or ”quasi-experiment”** (exogeneous and unforeseen!) to reveal the **causal effect of the ”policy”**

But ... What if the selection into treatment is not random?

We can relax this assumption
with **Conditional Independence Assumption**

Or in other words:

once you control for X, being treated is **“as good as random”**

Recap so far

- To address your research question, consider the randomized case (i.e. the “ideal experiment”) as an hypothetical benchmark.

- This is the experiment you would like to run, if you could!

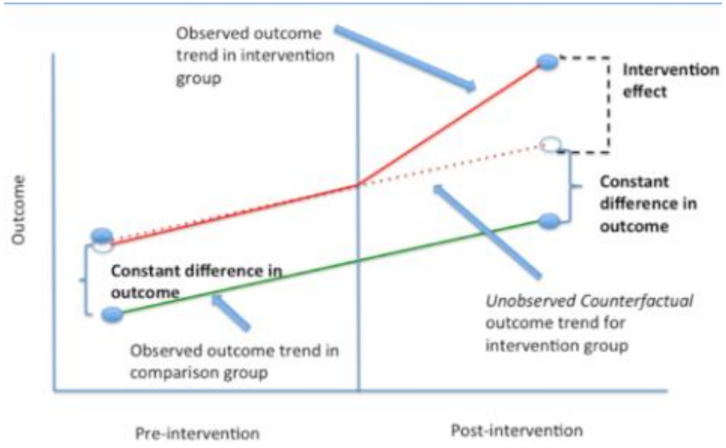
- Think about your problem in terms of the potential outcome framework and then check if the assignment to the treatment is any close to the ideal experiment (it must be independent of the potential outcomes).

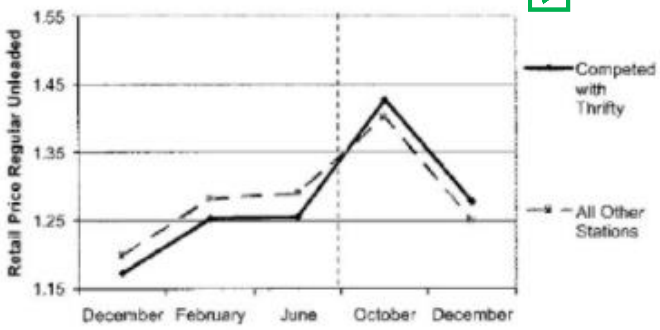
- Analyze the source of variation of the assignment to treatment. This is crucial and it must be exogenous (even conditional on observables, CIA).

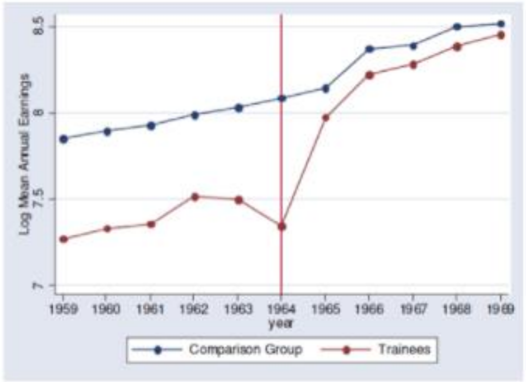
Recall slides 11 and 14

Parallel trends assumption in DiD

$$y = \beta_0 + \delta_0 dTime + \beta_1 dTreated + \delta_1 dTreated \cdot dTime + controls + u$$







In the absence of treatment, the difference between T and C is constant over time. Visual inspection can be useful.

Stata exercise: Prison rates

We use the dataset ”prison” (from J. Wooldridge).

Assume a policy change lowering prison sentences for some states (state = 17-51), but not for other states.

Assume that this policy started in 1986.

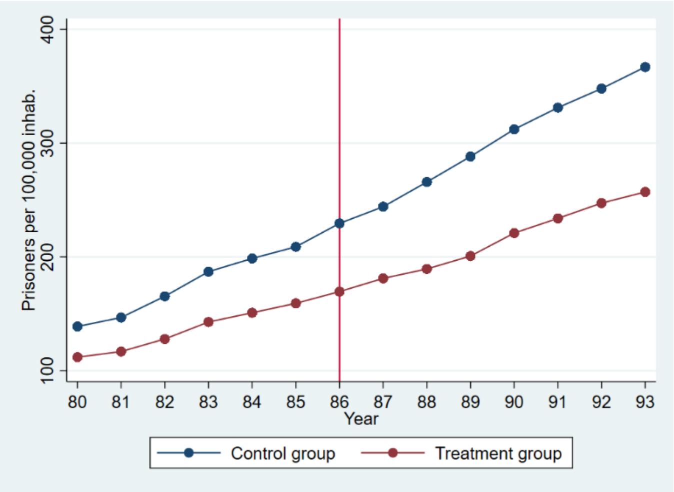
```
. d pris treated post did state year

              storage   display    value
variable name   type    format     label      variable label
-----------------------------------------------------------------------------------------
pris            float   %9.0g                 prison pop. per 100,000
treated         float   %9.0g                 0 = Control states; 1 = Treated States
post            float   %9.0g                 0 = Before treatment; 1 = After Treatment
did             float   %9.0g                 DiD estimator
state           byte    %9.0g                 alphabetical; DC = 9
year            byte    %9.0g                 80 to 93
```

Parallel trends: plot average values

```
bysort year: egen pris0=mean(pris) if treated==0
lab var pris0 "Control group"
bysort year: egen pris1=mean(pris) if treated==1
lab var pris1 "Treatment group"

twoway (connected pris0 pris1 year), xline(86) ytitle("Prisoners per 100,000 inhab.") ///
xtitle("Year") xlab(80(1)93)
graph export "trends1.png", as(png) replace
```



Parallel trends: same plot with *margins*

We can also plot the same graph using the *margins* command.

First, we would run a saturated model with dummy variables for each year, another dummy for “treated”, and each of their interactions.

```
reg pris i.year##i.treated
```

This approach is very helpful when we relax the assumption and we want to test for parallel trends conditional on observables!

Source	SS	df	MS
Model	2886138.58	27	106894.021
Residual	9671831.57	686	14098.8798
Total	12557970.1	713	17612.8614

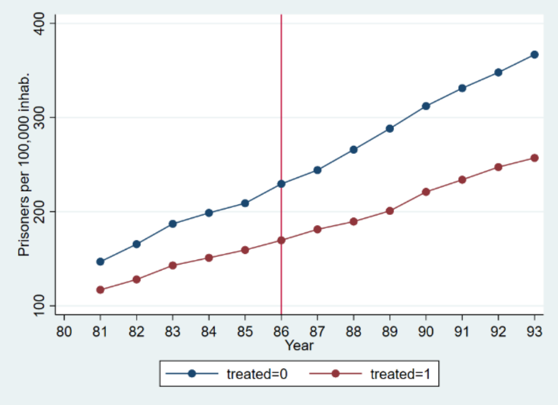
Number of obs	=	714
F(27, 686)	=	7.58
Prob > F	=	0.0000
R-squared	=	0.2298
Adj R-squared	=	0.1995
Root MSE	=	118.74

pris	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
year						
81	7.902258	41.98047	0.19	0.851	-74.52338	90.3279
82	26.54346	41.98047	0.63	0.527	-55.88218	108.9691
83	48.14177	41.98047	1.15	0.252	-34.28386	130.5674
84	59.75509	41.98047	1.42	0.155	-22.67054	142.1807
85	69.97154	41.98047	1.67	0.096	-12.4541	152.3972
86	90.56271	41.98047	2.16	0.031	8.137071	172.9883
87	105.2	41.98047	2.51	0.012	22.77434	187.6256
88	126.9169	41.98047	3.02	0.003	44.49122	209.3425
89	149.3067	41.98047	3.56	0.000	66.88103	231.7323
90	173.1955	41.98047	4.13	0.000	90.7699	255.6212
91	192.2164	41.98047	4.58	0.000	109.7907	274.642
92	208.9624	41.98047	4.98	0.000	126.5368	291.388
93	227.9274	41.98047	5.43	0.000	145.5017	310.353
1.treated	-26.95774	35.83302	-0.75	0.452	-97.3133	43.39781
year#treated						
81 1	-2.929324	50.67554	-0.06	0.954	-102.4271	96.56845
82 1	-10.55925	50.67554	-0.21	0.835	-110.057	88.93853
83 1	-17.24801	50.67554	-0.34	0.734	-116.7458	82.24977
84 1	-20.72339	50.67554	-0.41	0.683	-120.2212	78.77439
85 1	-22.66707	50.67554	-0.45	0.655	-122.1648	76.83071
86 1	-32.96124	50.67554	-0.65	0.516	-132.459	66.53654
87 1	-35.97616	50.67554	-0.71	0.478	-135.4739	63.52162
88 1	-49.44072	50.67554	-0.98	0.330	-148.9385	50.05706
89 1	-60.45866	50.67554	-1.19	0.233	-159.9564	39.03912
90 1	-64.1597	50.67554	-1.27	0.206	-163.6575	35.33808
91 1	-70.29631	50.67554	-1.39	0.166	-169.7941	29.20147
92 1	-73.59571	50.67554	-1.45	0.147	-173.0935	25.90206
93 1	-82.8252	50.67554	-1.63	0.103	-182.323	16.67258
_cons	138.9538	29.68468	4.68	0.000	80.67003	197.2375

Parallel trends: same plot with *margins*

```
margins treated, at(year=(81(1)93)) vsquish

marginsplot, noci xline(86) xlab(80(1)93) ytitle("Prisoners per 100,000 inhab.")
xtitle("Year") title("")
```




```
Adjusted predictions                            Number of obs     =        714
Model VCE    : OLS

Expression   : Linear prediction, predict()
1._at        : year            =          81
2._at        : year            =          82
3._at        : year            =          83
4._at        : year            =          84
5._at        : year            =          85
6._at        : year            =          86
7._at        : year            =          87
8._at        : year            =          88
9._at        : year            =          89
10._at       : year            =          90
11._at       : year            =          91
12._at       : year            =          92
13._at       : year            =          93
```

	Margin	Delta-method Std. Err.	t	P>|t|	[95% Conf.	Interval]
_at#treated						
1 0	146.856	29.68468	4.95	0.000	88.57228	205.1397
1 1	116.9689	20.0705	5.83	0.000	77.56195	156.3759
2 0	165.4972	29.68468	5.58	0.000	107.2135	223.7809
2 1	127.9802	20.0705	6.38	0.000	88.57324	167.3872
3 0	187.0955	29.68468	6.30	0.000	128.8118	245.3793
3 1	142.8898	20.0705	7.12	0.000	103.4828	182.2968
4 0	198.7088	29.68468	6.69	0.000	140.4251	256.9926
4 1	151.0277	20.0705	7.52	0.000	111.6207	190.4347
5 0	208.9253	29.68468	7.04	0.000	150.6416	267.209
5 1	159.3005	20.0705	7.94	0.000	119.8935	198.7075
6 0	229.5165	29.68468	7.73	0.000	171.2327	287.8002
6 1	169.5975	20.0705	8.45	0.000	130.1905	209.0045
7 0	244.1537	29.68468	8.22	0.000	185.87	302.4375
7 1	181.2198	20.0705	9.03	0.000	141.8128	220.6268
8 0	265.8706	29.68468	8.96	0.000	207.5869	324.1543
8 1	189.4721	20.0705	9.44	0.000	150.0652	228.8791
9 0	288.2604	29.68468	9.71	0.000	229.9767	346.5441
9 1	200.844	20.0705	10.01	0.000	161.437	240.251
10 0	312.1493	29.68468	10.52	0.000	253.8656	370.433
10 1	221.0319	20.0705	11.01	0.000	181.6249	260.4388
11 0	331.1701	29.68468	11.16	0.000	272.8864	389.4539
11 1	233.9161	20.0705	11.65	0.000	194.5091	273.3231
12 0	347.9162	29.68468	11.72	0.000	289.6324	406.1999
12 1	247.3627	20.0705	12.32	0.000	207.9557	286.7697
13 0	366.8811	29.68468	12.36	0.000	308.5974	425.1648
13 1	257.0982	20.0705	12.81	0.000	217.6912	296.5052

It looks the same as the previous one!

A fully saturated regression model

- In control states, prison rates increase over time (i.year)
- In treated states, the passage of time has a positive, but lower effect on prison rates (sum i.year & i.treated#i.year)
- In other words, prison rates increase in all states (*magnitude of coefficients*) but they are smaller in treated than in control states (*sign of coefficients*)
- **Note that here 1980 is the reference year, so comparing pre- and post-policy is quite difficult**

```
.
. * A) Fully saturated model:
. reg pris i.treated##i.year, cluster (state

Linear regression                               Number of obs     =        714
                                                F(27, 50)         =     150.08
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2298
                                                Root MSE          =     118.74

                                  (Std. Err. adjusted for 51 clusters in state)
------------------------------------------------------------------------------
             |               Robust
        pris |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   1.treated |  -26.95774   23.03849    -1.17   0.248    -73.2319    19.31642
             |
        year |
         81  |   7.902258   2.572847     3.07   0.003    2.734542    13.06997
         82  |   26.54346   4.055156     6.55   0.000    18.39844    34.68848
         83  |   48.14177   7.273863     6.62   0.000    33.53179    62.75176
         84  |   59.75509   8.602639     6.95   0.000    42.47618      77.034
         85  |   69.97154   11.12025     6.29   0.000    47.63585    92.30722
         86  |   90.56271   18.27964     4.95   0.000    53.84698    127.2784
         87  |      105.2   19.39936     5.42   0.000    66.23521    144.1648
         88  |   126.9169   26.52769     4.78   0.000    73.63443    180.1993
         89  |   149.3067   35.73909     4.18   0.000     77.5226    221.0907
         90  |   173.1955    34.7398     4.99   0.000    103.4186    242.9725
         91  |   192.2164   37.49753     5.13   0.000    116.9004    267.5324
         92  |   208.9624   41.13044     5.08   0.000    126.3495    291.5753
         93  |   227.9274   45.63096     5.00   0.000    136.2749    319.5798
             |
treated#year |
       1 81  |  -2.929324   3.026349    -0.97   0.338   -9.007925    3.149277
       1 82  |  -10.55925   4.564066    -2.31   0.025   -19.72644   -1.392048
       1 83  |  -17.24801   8.315161    -2.07   0.043    -33.9495   -.5465155
       1 84  |  -20.72339   10.23496    -2.02   0.048   -41.28092   -.1658684
       1 85  |  -22.66707   12.78809    -1.77   0.082    -48.3527    3.01856
       1 86  |  -32.96124   19.38873    -1.70   0.095   -71.90466    5.98218
       1 87  |  -35.97616    21.1049    -1.70   0.094   -78.36659    6.414272
       1 88  |  -49.44072   27.81833    -1.78   0.082   -105.3155    6.434047
       1 89  |  -60.45866   36.97279    -1.64   0.108   -134.7207    13.80338
       1 90  |   -64.1597   36.27455    -1.77   0.083   -137.0193    8.699881
       1 91  |  -70.29631   39.17234    -1.79   0.079   -148.9763    8.383664
       1 92  |  -73.59571   43.01121    -1.71   0.093   -159.9863    12.79485
       1 93  |   -82.8252   47.50028    -1.74   0.087   -178.2323    12.58192
             |
       _cons |   138.9538   20.97328     6.63   0.000    96.82768    181.0798
------------------------------------------------------------------------------
```

Testing pre-trends

```
. reg pris i.treated##i.years_pre if year>= 83 & year <= 89 , cluster(state)

Linear regression                               Number of obs     =        357
                                                F(7, 50)          =      13.52
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1007
                                                Root MSE          =     108.57

                                  (Std. Err. adjusted for 51 clusters in state)
-------------------------------------------------------------------------------
              |               Robust
         pris |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+----------------------------------------------------------------
    1.treated |  -71.66694   45.43946    -1.58   0.121    -162.9348     19.6009
              |
    years_pre |
            1 |  -48.02502   14.97244    -3.21   0.002    -78.09804   -17.95199
            2 |  -58.24146   17.17109    -3.39   0.001    -92.73061   -23.75231
            3 |  -69.85478   18.29084    -3.82   0.000     -106.593   -33.11654
              |
treated#years_pre |
          1 1 |   22.04213   15.35685     1.44   0.157    -8.803008    52.88726
          1 2 |    23.9858   17.65608     1.36   0.180    -11.47747    59.44907
          1 3 |   27.46119   19.03627     1.44   0.155    -10.77429    65.69667
              |
        _cons |   256.9503   42.93482     5.98   0.000     170.7132    343.1874
-------------------------------------------------------------------------------
```

- Compare trends 3 years pre-policy for treated and control states
- Using 86 (year of the policy introduction) as a baseline, control states had lower prison rates in the years just before the treatment (*i.year_pre*)
- But this trend was not significantly different for treated states (*i.treated#i.years_pre)*

There is no significantly different pre-trend for the treated group

Testing pre-trends, adding FEs

```
. xtset state year
       panel variable:  state (strongly balanced)
        time variable:  year, 80 to 93
                delta:  1 unit

. xtreg pris i.treated##i.years_pre if year>= 83 & year <= 89 , fe cluster(state)
note: 1.treated omitted because of collinearity

Fixed-effects (within) regression               Number of obs      =       357
Group variable: state                           Number of groups   =        51

R-sq:                                           Obs per group:
     within  = 0.3785                                         min =          7
     between = 0.0699                                         avg =        7.0
     overall = 0.0185                                         max =          7

                                                F(6,50)            =     15.51
corr(u_i, Xb)   = -0.0660                       Prob > F           =    0.0000

                                  (Std. Err. adjusted for 51 clusters in state)
------------------------------------------------------------------------------
             |               Robust
        pris |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   1.treated |          0  (omitted)
             |
   years_pre |
          1  |  -48.02502   14.95103    -3.21   0.002    -78.05505   -17.99498
          2  |  -58.24146   17.14654    -3.40   0.001    -92.68131   -23.80162
          3  |  -69.85478   18.26469    -3.82   0.000    -106.5405   -33.16906
             |
treated#years_pre |
        1 1  |   22.04213   15.33489     1.44   0.157    -8.758912    52.84317
        1 2  |    23.9858   17.63083     1.36   0.180    -11.42677    59.39838
        1 3  |   27.46119   19.00906     1.44   0.155    -10.71963      65.642
             |
       _cons |   207.7671   2.548821    81.51   0.000     202.6477    212.8866
-------------+----------------------------------------------------------------
     sigma_u |  109.85421
     sigma_e |  30.799141
         rho |  .92712446   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Same result as in the previous example:

There is no significantly different pre-trend for the treated group

Simple DiD estimation

```
reg pris i.treated##i.post, cluster (state)
```

```
Linear regression                               Number of obs     =        714
                                                F(3, 50)          =      44.58
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1775
                                                Root MSE          =     120.61

                                    (Std. Err. adjusted for 51 clusters in state)
```

pris	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.treated	-39.31225	26.77102	-1.47	0.148	-93.08343	14.45893
1.post	123.9003	26.98897	4.59	0.000	69.69136	178.1092
treated#post						
1 1	-46.35971	27.96324	-1.66	0.104	-102.5255	9.806117
_cons	174.3394	24.38248	7.15	0.000	125.3658	223.3131

How do you interpret the results?

```
diff pris, treated(treated) period(post) cluster(state)
```

```
.
. diff pris, treated(treated) period(post) cluster(state)
```

DIFFERENCE-IN-DIFFERENCES ESTIMATION RESULTS

Number of observations in the DIFF-IN-DIFF: 714

	Before	After	
Control:	96	128	224
Treated:	210	280	490
	306	408	

Outcome var.	pris	S. Err.	|t|	P>|t|
Before				
Control	174.339			
Treated	135.027			
Diff (T-C)	-39.312	26.771	-1.47	0.148
After				
Control	298.240			
Treated	212.568			
Diff (T-C)	-85.672	52.324	1.64	0.108
Diff-in-Diff	-46.360	27.963	1.66	0.104

R-square: 0.18

* Means and Standard Errors are estimated by linear regression

**Clustered Std. Errors

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

* *diff* is a user-written command to be installed

Testing before/after trends

```
. reg pris i.treated##i.years_post i.treated##i.years_pre if year>= 83 & year <= 89 , cluster (state)

Linear regression                               Number of obs     =        357
                                                F(13, 50)         =      13.42
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1117
                                                Root MSE          =     108.85

                                   (Std. Err. adjusted for 51 clusters in state)
```

pris	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.treated	-59.91898	39.08622	-1.53	0.132	-138.426	18.588
years_post						
1	14.63727	2.218082	6.60	0.000	10.18212	19.09242
2	36.35415	8.914252	4.08	0.000	18.44935	54.25895
3	58.74396	18.72665	3.14	0.003	21.13037	96.35755
treated#years_post						
1 1	-3.01492	3.389226	-0.89	0.378	-9.82238	3.79254
1 2	-16.47948	9.481491	-1.74	0.088	-35.52362	2.564654
1 3	-27.49742	19.31209	-1.42	0.161	-66.2869	11.29206
years_pre						
1	-20.59117	8.149227	-2.53	0.015	-36.95938	-4.222969
2	-30.80762	10.58	-2.91	0.005	-52.05818	-9.557056
3	-42.42093	11.9499	-3.55	0.001	-66.42301	-18.41886
treated#years_pre						
1 1	10.29417	8.334959	1.24	0.223	-6.447086	27.03543
1 2	12.23785	10.86806	1.13	0.266	-9.591288	34.06698
1 3	15.71323	12.44591	1.26	0.213	-9.285112	40.71158
_cons	229.5165	36.6686	6.26	0.000	155.8654	303.1675

- Now compare 3 years pre- and post-policy
- Compared to 86, control states had lower prison rates in the years just before the treatment (*i.year_pre*)
- But this trend was not significantly different for treated groups (*i.treated#i.years_pre)*
- After the treatment, prison rates increase in control states (*i.years_post)*
- This increase was smaller (but not significantly different) for treated states (*i.treated#i.years_post)*

Testing before/after trends with FEs

```
. xtset state year
       panel variable:  state (strongly balanced)
        time variable:  year, 80 to 93
                delta:  1 unit

.
. xtreg pris i.treated##i.years_post i.treated##i.years_pre if year>= 83 & year <= 89 , fe cluster(state)
note: 1.treated omitted because of collinearity

Fixed-effects (within) regression               Number of obs      =       357
Group variable: state                           Number of groups   =        51

R-sq:                                           Obs per group:
     within  = 0.4874                                         min =         7
     between = 0.0699                                         avg =       7.0
     overall = 0.0514                                         max =         7

                                                F(12,50)           =     14.06
corr(u_i, Xb)  = 0.0060                         Prob > F           =    0.0000

                                 (Std. Err. adjusted for 51 clusters in state)
------------------------------------------------------------------------------
             |               Robust
        pris |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   1.treated |          0  (omitted)
             |
  years_post |
          1  |   14.63727   2.214856     6.61   0.000      10.1886    19.08594
          2  |   36.35415   8.901286     4.08   0.000     18.47539    54.23291
          3  |   58.74396   18.69941     3.14   0.003     21.18508    96.30284
             |
treated#years_post |
        1 1  |   -3.01492   3.384296    -0.89   0.377    -9.812478    3.782639
        1 2  |  -16.47948   9.467699    -1.74   0.088    -35.49591    2.536953
        1 3  |  -27.49742     19.284    -1.43   0.160    -66.23047    11.23563
             |
   years_pre |
          1  |  -20.59117   8.137373    -2.53   0.015    -36.93557   -4.246777
          2  |  -30.80762   10.56461    -2.92   0.005    -52.02727   -9.587966
          3  |  -42.42093   11.93251    -3.56   0.001    -66.38809   -18.45377
             |
treated#years_pre |
        1 1  |   10.29417   8.322835     1.24   0.222    -6.422735    27.01108
        1 2  |   12.23785   10.85225     1.13   0.265    -9.559537    34.03523
        1 3  |   15.71323   12.42781     1.26   0.212    -9.248751    40.67521
             |
       _cons |   188.3956   1.030672   182.79   0.000     186.3254    190.4658
-------------+----------------------------------------------------------------
     sigma_u |  108.29803
     sigma_e |  28.252839
         rho |  .93627821   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Same result as in the previous example:

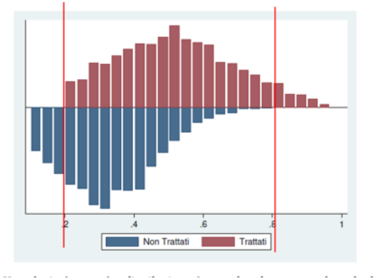
There is no significant pre-trend, but there is no significant effect of the policy either

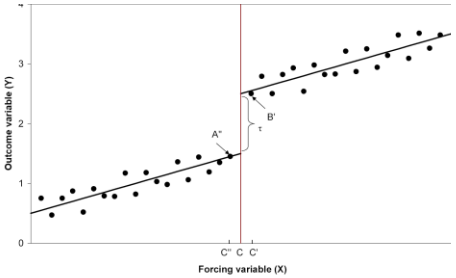
Recap

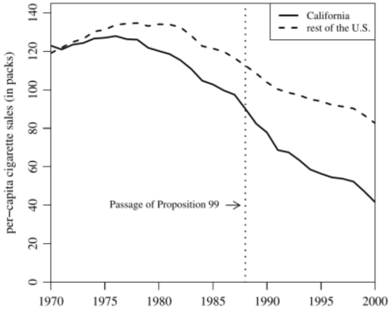
- Trends could be non-linear. Maybe the best fitting model is a quadratic trends model or other functional form
- Remember that the difference between the groups may not be parallel in the raw, unadjusted data, but they could become parallel after “holding” other variables constant or after “taking into account” the effect of other variables (in other words, the trends could become parallel conditional on other covariates)
- This is a common situation. The parallel trends test may fail with raw data (unadjusted) but it could pass when we control for covariates
- “Passing” here means that we do not reject the null hypothesis that trends are parallel

Extensions on DiD – related approaches

	MATCHING AND DID	REGRESSION DISCONTINUITY DESIGN	SYNTHETIC CONTROL GROUP
Assumption violated	Parallel Trends Random Selection into Treatment	Random Selection into Treatment	Parallel Trends Random Selection into Treatment
Main idea	*Identification of untreated units (control group) similar in several respects to treated units, before the treatment, as counterfactual.* *Based CIA: conditional on observables, the difference between the two groups is only the exposure to the treatment*	*Assignment to treatment depends on (a set of) variables (i.e., forcing variables) satisfying a set of known conditions.* *The effect of the treatment is estimated by the discontinuity of the outcome variable at the cut-off (also, geographic).*	*Since only a few units are treated, the counterfactual group is built as a weighted average of other untreated units.*

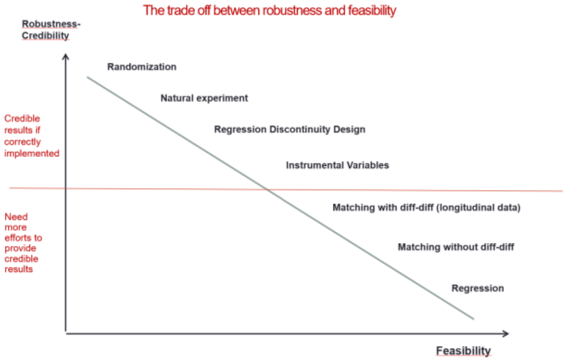






A Checklist for DiD Practitioners (to start)

- Is everyone treated at the same time?
- Are you sure about the validity of the parallel trends assumption?
- Do you have a large number of treated and untreated clusters sampled from a super-population?



Heterogeneity Robust Estimators for Staggered Timing

Package	Software	Description
did, csdid	R, Stata	Implements Callaway and Sant'Anna (2021)
did2s	R, Stata	Implements Gardner (2021), Borusyak et al (2021), Sun and Abraham (2020), Callaway and Sant'Anna (2021), Roth and Sant'Anna (2021)
didimputation, did_imputation	R, Stata	Implements Borusyak, Jaravel, and Spiess (2021)
DIDmultiplegt, did_multiplegt	R, Stata	Implements de Chaisemartin and D'Haultfoeuille (2020)
eventstudyinteract	Stata	Implements Sun and Abraham (2020)
flexpaneldid	Stata	Implements Dettmann (2020), based on Heckman et al (1998)
fixest	R	Implements Sun and Abraham (2020)
stackedev	Stata	Implements stacking approach in Cengiz et al (2019)
staggered	R	Implements Roth and Sant'Anna (2021), Callaway and Sant'Anna (2020), and Sun and Abraham (2020)
xtevent	Stata	Implements Freyaldenhoven et al (2019)

DiD with Covariates

Package	Software	Description
DRDID, drdid	R, Stata	Implements Sant'Anna and Zhao (2020)

Diagnostics for TWFE with Staggered Timing

Package	Software	Description
bacondecomp	R, Stata	Diagnostics from Goodman-Bacon (2021)
TwoWayFEWeights	R	Diagnostics from de Chaisemartin and D'Haultfoeuille (2020)

Diagnostic / Sensitivity for Violations of Parallel Trends

Package	Software	Description
honestDiD	R	Implements Rambachan and Roth (2021)
pretrends	R	Diagnostics from Roth (2021)

More on: https://asjadnaqvi.github.io/DiD/

Roth, J., Sant'Anna, P. H., Bilinski, A., & Poe, J. (2022). What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. arXiv preprint arXiv:2201.01194.

Extensions on DiD – complex settings

	MULTIPLE TIME PERIODS	MULTIPLE GROUPS	DYNAMIC (STAGGERED) TREATMENT
Main takeaways	Adding μ_t would control for ”time trends” and $dPost$ could be more general because it could accommodate different timing of treatment for some units Interacting $\gamma_s * \mu_t$(state-specific time trends) is often described as a robustness check: the DiD estimator shouldn’t change	Adding λ_i captures time-invariant group/individual characteristics, including those that could affect self-selection (or assignment) into the treatment/program.	Not everyone is treated at the same time Standard “static” TWFE models may not represent a straightforward weighted average of unit-level treatment effects when treatment effects are allowed to be heterogeneous across time or units.

Some references

Lechner, M., & Miquel, R. (2010). Identification of the effects of dynamic treatments by sequential conditional independence assumptions. *Empirical Economics*, *39*(1), 111-137.

Callaway, B., & Sant’Anna, P. H. (2021). Difference-in-differences with multiple time periods. Journal of Econometrics, 225(2), 200-230.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics, 225(2), 254-277.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates?. The Quarterly journal of economics, 119(1), 249-275.

Dasgupta, S., & Žaldokas, A. (2019). Anticollusion enforcement: Justice for consumers and equity for firms. The Review of Financial Studies, 32(7), 2587-2624

Roth, J., Sant'Anna, P. H., Bilinski, A., & Poe, J. (2022). What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. arXiv preprint arXiv:2201.01194.

It’s over! Good luck with the exam!

And remember the **motivate** command in Stata, if needed!

```
. motivate
'If you are going through hell, keep going.'
                                         Winston Churchill
```

Otherwise, there is always **demotivate** ☺

```
. demotivate
Best not to dwell on what your R-using colleagues really think of you.
```