

Difference-in-Differences

Magne Mogstad

Outline

This week we will talk about the use of repeated cross sections or panel data

Today we focus on difference-in-differences (DiD) and extensions

- ▶ How to apply and interpret basic DiD models:
 - ▶ Set up in potential outcome framework
 - ▶ Key assumptions for identification
- ▶ Extensions of DiD model
 - ▶ multiple periods and groups
 - ▶ different or weaker identifying assumptions
- ▶ Issues of inference

Goal: Know how to motivate, assess and check your DiD model

Definitions and Terminology

Definitions

- ▶ A cross-section of data has one dimension, the unit i
- ▶ **Panel data** has both a unit i and a time period t

Benefits

- ▶ A panel is a repeated cross-section, but not conversely
- ▶ The difference is that the same units are observed multiple times
- ▶ This allows us to consider richer models with dynamics

Definitions and Terminology

Balance

- ▶ A panel is **balanced** if every unit i is observed in all time periods t
- ▶ Balance is common in theoretical discussions, but not in actual data
- ▶ Common to balance the panel artificially by throwing away data
 - Analysis is now conditional on balance - this may be unattractive
- ▶ Otherwise assume missing at random, or model the attrition process

Repeated Cross-Section vs Panel Data In a DID

- ▶ Difference between panel and repeated cross sections can be semantic
- ▶ Often, everything data is observed or aggregated to state/time
- ▶ So, could view this as a panel of states over time
- ▶ The key is the level at which the treatment varies, individual vs. state

The level of aggregation

- ▶ More generally, suppose that $X_{it} = W_{kt}$ for all i in group $G_{it} = k$
- ▶ Regressing Y_{it} on X_{it} is equivalent (check it!) to regressing

$$\frac{\sum_{i=1}^N Y_{it} \mathbb{1}[G_{it} = k]}{\sum_{i=1}^N \mathbb{1}[G_{it} = k]} \equiv \bar{Y}_{kt} \text{ onto } W_{kt} \text{ weighted by } N_{kt} = \sum_{i=1}^N \mathbb{1}[G_{it} = k]$$

Simple DiD setup: 2 periods and 2 groups

2 periods:

- ▶ *before* the intervention, $t = 1$
- ▶ *after* the intervention, $t = 2$

2 groups:

- ▶ before:
 - ▶ neither group is treated by the intervention, $D_{i1} = 0$
- ▶ after:
 - ▶ the *treatment group* is treated by the intervention, $D_{i2} = 1$
 - ▶ the *control group* is unaffected by the intervention, $D_{i2} = 0$

DiD: Potential outcomes

For each individual i , we imagine a what-if-scenario:

$$\text{Potential outcome in period 1} = Y_{i1}^0$$

$$\text{Potential outcome in period 2} = \begin{cases} Y_{i2}^1 & \text{if } D_{i2} = 1 \\ Y_{i2}^0 & \text{if } D_{i2} = 0 \end{cases}$$

	Treatment	Control
Before	Y_{i1}^0	Y_{i1}^0
After	Y_{i2}^1	Y_{i2}^0

The *observed outcome* in period t :

$$Y_{it} = D_{it} Y_{it}^1 + (1 - D_{it}) Y_{it}^0 = D_{it}(Y_{it}^1 - Y_{it}^0) + Y_{it}^0$$

DiD estimand

The change in individual i 's observed outcome over time:

$$Y_{i2} - Y_{i1} = D_{i2}(Y_{i2}^1 - Y_{i2}^0) + Y_{i2}^0 - Y_{i1}^0$$

First difference: After-before for treatment group

$$E(Y_{i2} - Y_{i1} | D_{i2} = 1) = E(Y_{i2}^1 - Y_{i2}^0 | D_{i2} = 1) + E(Y_{i2}^0 - Y_{i1}^0 | D_{i2} = 1)$$

Second difference: After-before for control group

$$E(Y_{i2} - Y_{i1} | D_{i2} = 0) = E(Y_{i2}^0 - Y_{i1}^0 | D_{i2} = 0)$$

DiD estimand

The change in individual i 's observed outcome over time:

$$Y_{i2} - Y_{i1} = D_{i2}(Y_{i2}^1 - Y_{i2}^0) + Y_{i2}^0 - Y_{i1}^0$$

First difference: After-before for treatment group

$$E(Y_{i2} - Y_{i1} | D_{i2} = 1) = E(Y_{i2}^1 - Y_{i2}^0 | D_{i2} = 1) + E(Y_{i2}^0 - Y_{i1}^0 | D_{i2} = 1)$$

Second difference: After-before for control group

$$E(Y_{i2} - Y_{i1} | D_{i2} = 0) = E(Y_{i2}^0 - Y_{i1}^0 | D_{i2} = 0)$$

Difference-in-differences (DiD):

$$E(Y_{i2} - Y_{i1} | D_{i2} = 1) - E(Y_{i2} - Y_{i1} | D_{i2} = 0)$$

DiD assumption

DiD identifies ATT:

$$E(Y_{i2}^1 - Y_{i2}^0 | D_{i2} = 1) = E(Y_{i2} - Y_{i1} | D_{i2} = 1) - E(Y_{i2} - Y_{i1} | D_{i2} = 0)$$

under the following assumption:

Common trend in the absence of intervention

$$E(Y_{i2}^0 - Y_{i1}^0 | D_{i2} = 1) = E(Y_{i2}^0 - Y_{i1}^0 | D_{i2} = 0)$$

⇒ no selection on the change in non-treatment outcome level

DiD assumption

DiD identifies ATT:

$$E(Y_{i2}^1 - Y_{i2}^0 | D_{i2} = 1) = E(Y_{i2} - Y_{i1} | D_{i2} = 1) - E(Y_{i2} - Y_{i1} | D_{i2} = 0)$$

under the following assumption:

Common trend in the absence of intervention

$$E(Y_{i2}^0 - Y_{i1}^0 | D_{i2} = 1) = E(Y_{i2}^0 - Y_{i1}^0 | D_{i2} = 0)$$

⇒ no selection on the change in non-treatment outcome level

Common trend assumption allows for:

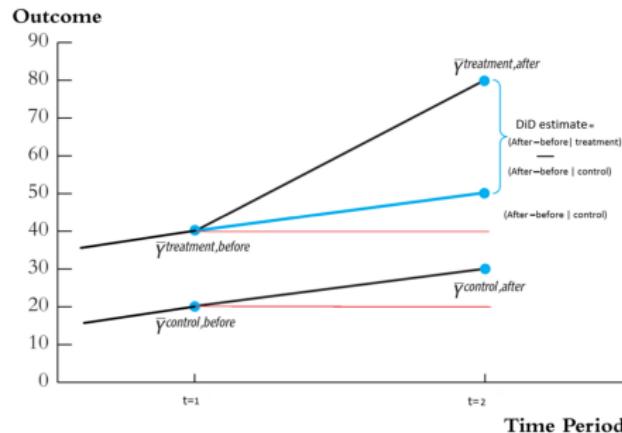
- ▶ selection on non-treatment levels:

$$E(Y_{it}^0 | D_{i2} = 1) \neq E(Y_{it}^0 | D_{i2} = 0), t = 1, 2$$

- ▶ selection on gains:

$$E(Y_{i2}^1 - Y_{i2}^0 | D_{i2} = 1) \neq E(Y_{i2}^1 - Y_{i2}^0 | D_{i2} = 0)$$

DiD assumption: Graphical representation



First difference:

- ▶ eliminates unobserved factors at the treatment level

Second difference:

- ▶ eliminates unobserved factors at the time level

First differences vs. DiD

Aim: Identify

$$E(Y_{i2}^1 - Y_{i2}^0 | D_{i2} = 1)$$

How: Estimate

$$E(Y_{i2}^0 | D_{i2} = 1) = E(Y_{i1}^0 | D_{i2} = 1) + E(Y_{i2}^0 - Y_{i1}^0 | D_{i2} = 1)$$

First differences vs. DiD

Aim: Identify

$$E(Y_{i2}^1 - Y_{i2}^0 | D_{i2} = 1)$$

How: Estimate

$$E(Y_{i2}^0 | D_{i2} = 1) = E(Y_{i1}^0 | D_{i2} = 1) + E(Y_{i2}^0 - Y_{i1}^0 | D_{i2} = 1)$$

DiD model assumes:

$$E(Y_{i2}^0 - Y_{i1}^0 | D_{i2} = 1) = E(Y_{i2}^0 - Y_{i1}^0 | D_{i2} = 0)$$

- ▶ this rules out idiosyncratic temporary shocks

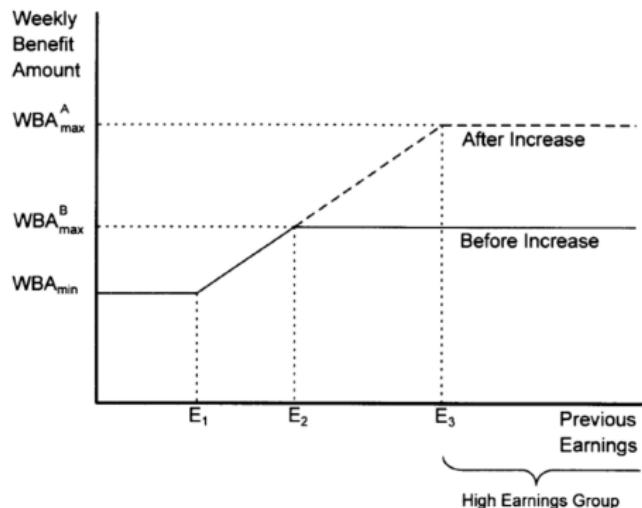
First differences model assumes:

$$E(Y_{i2}^0 - Y_{i1}^0 | D_{i2} = 1) = 0$$

- ▶ this rules out changes due to business cycle, life cycle behavior, etc.

Example: Disability Payments

Meyer, Viscusi and Durbin (AER, 1995) consider the effect of an increase in disability payments on time out of work:



Example: Disability Payments

Data

```
. d  
  
Contains data from injury.dta  
obs: 7,150  
vars: 30 14 Nov 2009 19:16  
size: 471,900 (99.8% of memory free)  
  
-----  
variable name  storage  display      value  
           type    format     label  
-----  
durat        float    %9.0g      duration of benefits  
afchnge      byte     %9.0g      =1 if after change in benefits  
highearn     byte     %9.0g      =1 if high earner  
male         byte     %9.0g      =1 if male  
married       byte     %9.0g      =1 if married  
hosp          byte     %9.0g      =1 if inj. required hosp. stay  
indust        byte     %9.0g      industry  
injtype       byte     %9.0g      type of injury  
age           byte     %9.0g      age at time of injury  
prewage       float    %9.0g      previous weekly wage, 1982 $  
totmed        float    %9.0g      total med. costs, 1982 $  
injdes        int      %9.0g      4 digit injury description  
benefit       float    %9.0g      real dollar value of benefit  
ky            byte     %9.0g      =1 for kentucky  
mi            byte     %9.0g      =1 for michigan  
-----  
Sorted by:
```

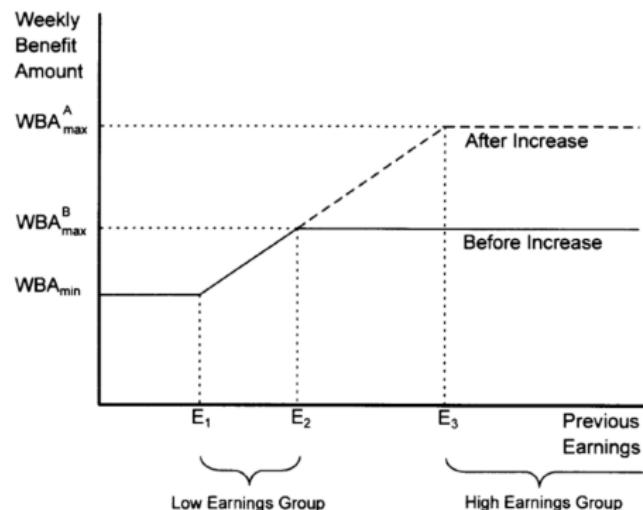
Example: Disability Payments

First differences by hand

```
. table mi afc if highearn==1, c(m dur) row  
  
-----  
| =1 if after change  
=1 for | in benefits  
michigan | 0 1  
-----+  
0 | 11.1766 12.89363  
1 | 14.77929 19.43379  
|  
Total | 11.76155 13.93152  
-----  
  
. di 12.89363 - 11.1766  
1.71703  
  
. di 19.43379 - 14.77929  
4.6545  
  
. di 13.93152 - 11.76155  
2.16997
```

Example: Disability Payments

Introducing a control group



Example: Disability Payments

DiD by hand

```
. table mi afc highearn, c(m dur) row

-----  
           | =1 if high earner and =1 if after change  
           |                                in benefits  
=1 for   | ----- 0 ----- 1 -----  
michigan |          0          1          0          1  
-----+-----  
      0 |  6.271554  7.037328  11.1766  12.89363  
      1 | 10.95883  13.65094  14.77929  19.43379  
      |  
    Total |  7.475044  8.611526  11.76155  13.93152  
-----  
  
. di (12.89363 - 11.1766) - (7.037328-6.271554)  
.951256  
  
. di (19.43379 - 14.77929) - (13.65094 -10.95883)  
1.96239  
  
. di (13.93152 - 11.76155) - (8.611526-7.475044)  
1.033488
```

DID as regression

The following regression

$$Y_{it} = \beta_0 + \beta_1 treat_i + \beta_2 after_t + \beta_3 treat_i * after_t + \epsilon_{it}$$

delivers our DID estimate:

	Treatment	Control	Difference
Before	$\beta_0 + \beta_1$	β_0	β_1
After	$\beta_0 + \beta_1 + \beta_2 + \beta_3$	$\beta_0 + \beta_2$	$\beta_1 + \beta_3$
Difference	$\beta_2 + \beta_3$	β_2	β_3

Advantages of regression:

- ▶ convenient way to obtain s.e.'s
- ▶ easy to add covariates
 - ▶ control for confounding trends
 - ▶ reduce residual variance (may increase precision of β)

Disability Payments

DID as regression

```
. reg durat afchng e highearn afhigh
```

Source	SS	df	MS	Number of obs	=	7150
Model	44343.4979	3	14781.166	F(3, 7146)	=	24.88
Residual	4245982.66	7146	594.176136	Prob > F	=	0.0000
Total	4290326.16	7149	600.129551	R-squared	=	0.0103
				Adj R-squared	=	0.0099
				Root MSE	=	24.376

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
durat	1.136483	.7453242	1.52	0.127	-.3245728 2.597539
afchng e	4.286505	.8140426	5.27	0.000	2.690741 5.88227
highearn	1.033489	1.178865	0.88	0.381	-1.277435 3.344414
afhigh	7.475044	.5089333	14.69	0.000	6.477384 8.472704
_cons					

Revisiting LaLonde

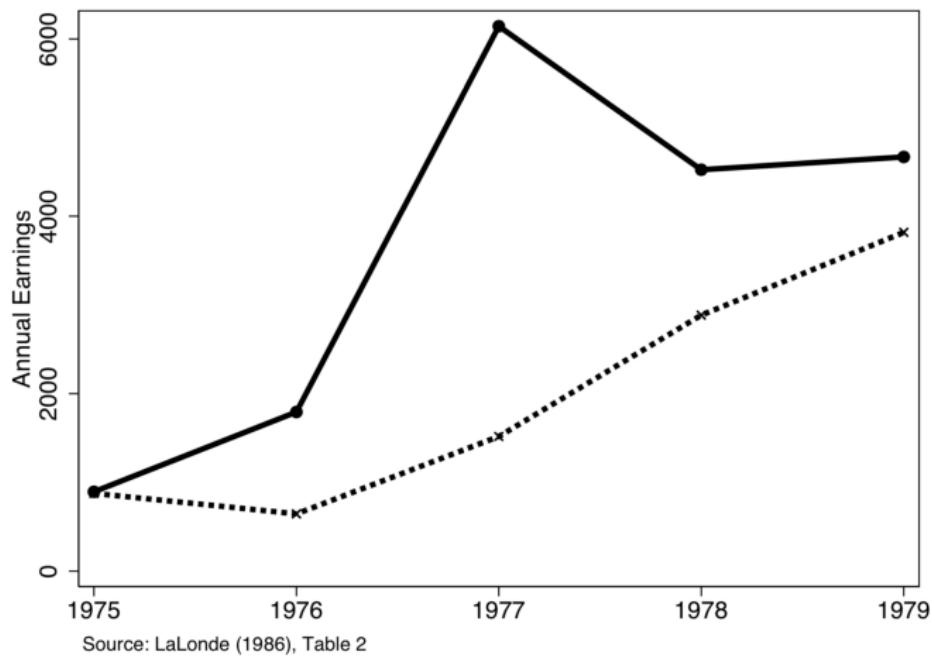
NSW was a randomized experiment that assigned people to training positions

- ▶ mid 1970s
 - ▶ AFDC women, ex-drug addicts, ex-criminal offenders, high school drop-outs of both sexes
 - ▶ guaranteed a job for 9 to 18 months
- ▶ Baseline earnings and demographics (1975)
- ▶ Post treatment earnings and demographics (1979)

LaLonde (1986) compared experimental with non-experimental estimates

- ▶ previously, we considered matching estimators
- ▶ now we consider difference-in-differences

Annual Earnings - Females



Source: LaLonde (1986), Table 2

Choice of comparison group - Females

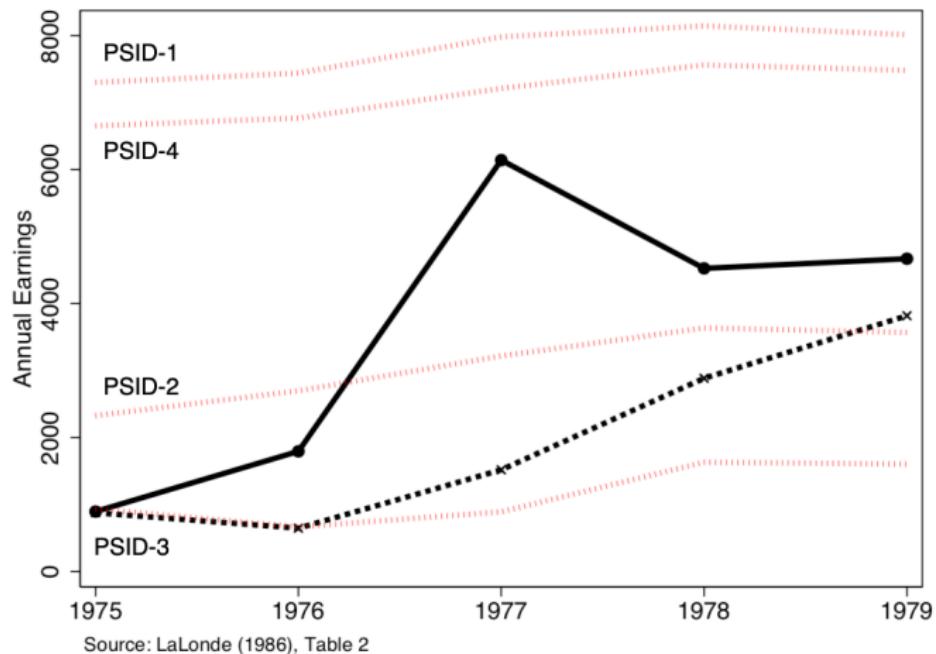
PSID

1. All female household heads aged 20-55
2. (1) + received AFDC in 1975
3. (2) + not working in 1976
4. (1) + no children below 5

CPS SSA

1. matched sample
2. + received AFDC in 1975
3. (1) + not working in 1976
4. (2) + not working in 1976

Annual Earnings - Females + PSID Comparisons



Source: LaLonde (1986), Table 2

Treatment effects - Females

Name of Comparison Group ^d	Comparison Group Earnings Growth 1975-79 (1)	NSW Treatment Earnings Less Comparison Group Earnings				Difference in Differences: Difference in Earnings Growth 1975-79 Treatments Less Comparisons		Unrestricted Difference in Differences: Quasi Difference in Earnings Growth 1975-79		Controlling for All Observed Variables and Pre-Training Earnings	
		Pre-Training Year, 1975		Post-Training Year, 1979		Without Age (6)	With Age (7)	Unadjusted (8)	Adjusted ^c (9)	Without AFDC (10)	With AFDC (11)
		Unadjusted (2)	Adjusted ^c (3)	Unadjusted (4)	Adjusted ^c (5)						
Controls	2,942 (220)	-17 (122)	-22 (122)	851 (307)	861 (306)	833 (323)	883 (323)	843 (308)	864 (306)	854 (312)	-
<i>PSID-1</i>	713 (210)	-6,443 (326)	-4,882 (336)	-3,357 (403)	-2,143 (425)	3,097 (317)	2,657 (333)	1,746 (357)	1,354 (380)	1,664 (409)	2,097 (491)
<i>PSID-2</i>	1,242 (314)	-1,467 (216)	-1,515 (224)	1,090 (468)	870 (484)	2,568 (473)	2,392 (481)	1,764 (472)	1,535 (487)	1,826 (537)	-
<i>PSID-3</i>	665 (351)	-77 (202)	-100 (208)	3,057 (532)	2,915 (543)	3,145 (557)	3,020 (563)	3,070 (531)	2,930 (543)	2,919 (592)	-
<i>PSID-4</i>	928 (311)	-5,694 (306)	-4,976 (323)	-2,822 (460)	-2,268 (491)	2,883 (417)	2,655 (434)	1,184 (483)	950 (503)	1,406 (542)	2,146 (652)
<i>CPS-SSA-1</i>	233 (64)	-6,928 (272)	-5,813 (309)	-3,363 (320)	-2,650 (365)	3,578 (280)	3,501 (282)	1,214 (272)	1,127 (309)	536 (349)	1,041 (503)
<i>CPS-SSA-2</i>	1,595 (360)	-2,888 (204)	-2,332 (256)	-683 (428)	-240 (536)	2,215 (438)	2,068 (446)	447 (468)	620 (554)	665 (651)	-
<i>CPS-SSA-3</i>	1,207 (166)	-3,715 (226)	-3,150 (325)	-1,122 (311)	-812 (452)	2,603 (307)	2,615 (328)	814 (305)	784 (429)	-99 (481)	1,246 (720)
<i>CPS-SSA-4</i>	1,684 (524)	-1,189 (249)	-780 (283)	926 (630)	756 (716)	2,126 (654)	1,833 (663)	1,222 (637)	952 (717)	827 (814)	-

DID: Card & Krueger (AER, 1994)

What is the effect of increase in minimum wage on employment?

Prediction economic theory: A rise in the minimum wage leads perfectly competitive employers to cut employment.

Card and Krueger investigate effect of increase in minimum wage from \$4.25 to \$ 5.05 in New Jersey on April 1, 1992.

Data on 410 fast-food restaurants (Burger King, Wendy's,...):

- ▶ in New Jersey (**treatment group**)
- ▶ and Pennsylvania (**control group**)
- ▶ in February/March 1992 (**before**)
- ▶ and in November/December 1992 (**after**)

DID: two groups & two time periods

Example: Card & Krueger (AER, 1994)

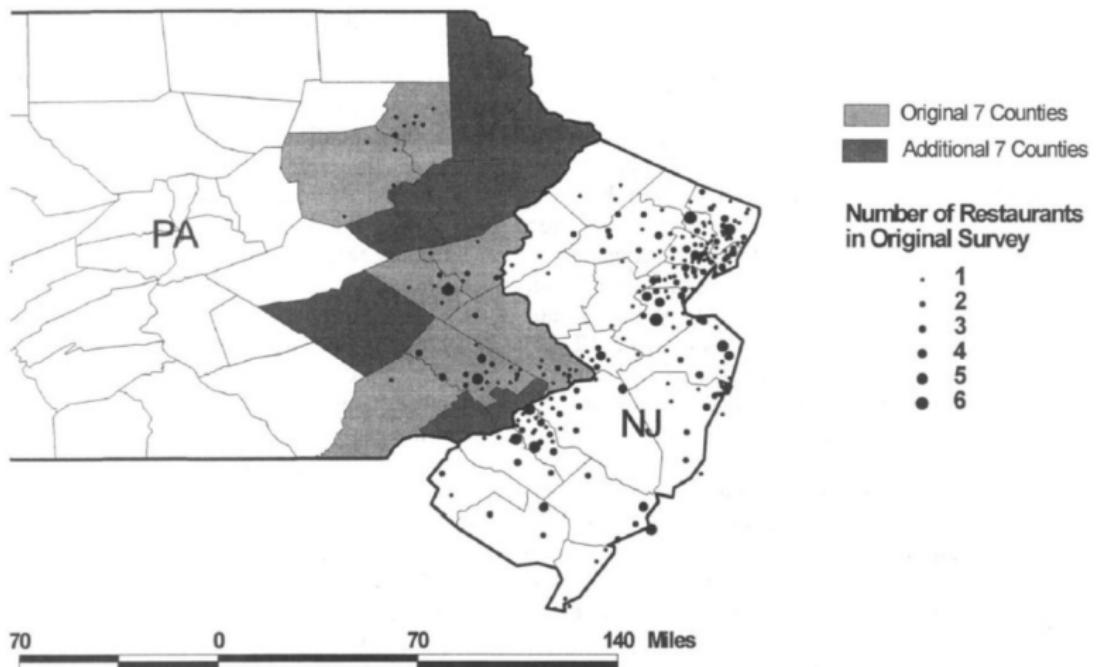


FIGURE 1. AREAS OF NEW JERSEY AND PENNSYLVANIA COVERED BY ORIGINAL SURVEY AND BLS DATA

DID: two groups & two periods

Example: Card & Krueger (AER, 1994)

```
. bys state time : sum emptot

-> state = 0, time = 1

      Variable |       Obs        Mean    Std. Dev.     Min     Max
-----+-----emptot |       77    23.33117   11.85628    7.5    70.5

-> state = 0, time = 2

      Variable |       Obs        Mean    Std. Dev.     Min     Max
-----+-----emptot |       77    21.16558   8.276732     0    43.5

-> state = 1, time = 1

      Variable |       Obs        Mean    Std. Dev.     Min     Max
-----+-----emptot |      321    20.43941   9.106239     5     85

-> state = 1, time = 2

      Variable |       Obs        Mean    Std. Dev.     Min     Max
-----+-----emptot |      319    21.02743   9.293024     0    60.5
```

DID: two groups & two time periods

Example: Card & Krueger (AER, 1994)

Outcome Y_{igt} is employment in restaurant i in state g at time t :

	Before: t=0	After: t=1
New Jersey (g=A)	$E[Y_{iA0}] = 20.4$	$E[Y_{iA1}] = 21.0$
Pennsylvania (g=B)	$E[Y_{iB0}] = 23.3$	$E[Y_{iB1}] = 21.2$

$$\hat{\beta}^{DID} = (21.0 - 20.4) - (21.2 - 23.3) = 2.7$$

Counter-intuitive result: Employment increased as consequence of increase in minimum wage (significant at 5% level)

Note: small change in NJ, but downward trend in PA

Common trend assumption: In absence of intervention employment in NJ would have had same downward trend as PA

DID: regression set-up

DID-estimator can be obtained by estimating following equation with OLS

$$Y_{igt} = \beta \cdot D_{gt} + \alpha_g + \lambda_t + \varepsilon_{igt}$$

with α_g and λ_t as group and time dummies

Advantages of specifying DiD in a regression equation:

- ▶ possible to extend to multiple groups & multiple time periods
 - ▶ $D_{gt} = 1$ if group g received treatment in period t and $D_{gt} = 0$ otherwise
- ▶ easy to add time-varying covariates and obtain standard errors
- ▶ treatment variable can be continuous
(at least under constant effects)
- ▶ treatment can be implemented at different groups at different times
(at least under constant effects)

Main assumption: In absence of intervention treatment and control groups would have **common trends**

DID regression

Example: Card & Krueger (AER, 1994)

```
. g treat = (state==1) * (time==2)
```

```
. reg emptot treat state time
```

Source	SS	df	MS	Number of obs	=	794
Model	521.116464	3	173.705488	F(3, 790)	=	1.96
Residual	69887.878	790	88.4656683	Prob > F	=	0.1180
Total	70408.9944	793	88.7881392	R-squared	=	0.0074
				Adj R-squared	=	0.0036
				Root MSE	=	9.4056
emptot	Coef.	Std. Err.	t	P> t	[95% Conf.	Interval
treat	2.753606	1.688409	1.63	0.103	-.560693	6.067905
state	-2.891761	1.193524	-2.42	0.016	-5.234614	-.5489079
time	-2.165584	1.515853	-1.43	0.154	-5.14116	.8099912
_cons	25.49675	2.396774	10.64	0.000	20.79196	30.20155

Threats to identification

Common trend

Recall the common trend assumption:

$$E(Y_{i2}^0 - Y_{i1}^0 | D_{i2} = 1) = E(Y_{i2}^0 - Y_{i1}^0 | D_{i2} = 0)$$

The assumption

- ▶ is functional form dependent
- ▶ may fail because of
 - ▶ group specific time trends
 - ▶ composition effects

The art of DID lies in the choice of an appropriate control group

- ▶ common pre-intervention trends are reassuring
 - ▶ Requires 2+ pre-treatment periods

1) Compositional changes

Composition Effects

The treatment under consideration may affect the composition of the treatment and control groups

- ▶ Example: a state lowers welfare benefits
 - ▶ this may induce poor families to move to other states
- ▶ Potential solution: (re)define group such that it is not affected by treatment
 - ▶ for example pre-treatment state of residence
- ▶ We no longer estimate ATT since some families move away from the treatment
 - ▶ we now estimate the so-called *intention to treat*

2) The nonlinearity critique of DID

- ▶ DiD suffers from noninvariance to transformations of the outcome variable
 - ▶ Why? Common trend implies that growth in outcome of a group doesn't depend on its level
- ▶ In other words, Identification depends on the transformation used
 - The same DID strategy cannot apply to both wages and log wages
- ▶ Is this a real concern?
 - ▶ The justification for wages vs log wages should be different
 - ▶ How/why would you argue for common trend in logs vs levels?

The Changes in Changes Model

Overview

- ▶ Athey and Imbens (2006) develop a model that is immune to this critique
- ▶ The model isolates the nonparametric aspects of the DID strategy
- ▶ They describe the argument as “changes in changes”(CiC)
- ▶ Idea is to compare quantile mappings to infer Y_0 for treated
→ Requires “rank invariance” - restriction on unobserved heterogeneity
- ▶ CiC model also allows estimation of QTE

DiD and quantile treatment effects (QTE)

Athey and Imbens (2006) showed how to use non-linear DiD method (called CiC) to estimate QTE:

- ▶ Quantile DiD
 - ▶ Common trend in outcome at a given quantile
- ▶ Changes in changes (CiC)
 - ▶ Common trend in outcome at the same quantile value

Both allow estimation of counterfactual outcome distribution

- ▶ in the absence of intervention

CiC is invariant wrt to monotonic transformations of outcome

- ▶ allows common trend in absolute and relative outcome

DiD and quantile treatment effects (con't)

Recall that standard DiD identify

$$E(Y_{i2}^1 - Y_{i2}^0 | D_{i2} = 1)$$

by estimating the counterfactual mean outcome:

$$E(Y_{i2}^0 | D_{i2} = 1) = E(Y_{i1}^0 | D_{i2} = 1) + E(Y_{i2}^0 - Y_{i1}^0 | D_{i2} = 0)$$

DiD and quantile treatment effects (con't)

Recall that standard DiD identify

$$E(Y_{i2}^1 - Y_{i2}^0 | D_{i2} = 1)$$

by estimating the counterfactual mean outcome:

$$E(Y_{i2}^0 | D_{i2} = 1) = E(Y_{i1}^0 | D_{i2} = 1) + E(Y_{i2}^0 - Y_{i1}^0 | D_{i2} = 0)$$

Non-linear DD methods rely on a similar idea for quantiles

Below, for $t = 1, 2$ let

- ▶ $F_t(Y)$ be the distribution of Y in the treatment group
- ▶ $G_t(Y)$ be the distribution of Y in the control group

Quantile DID

Three steps:

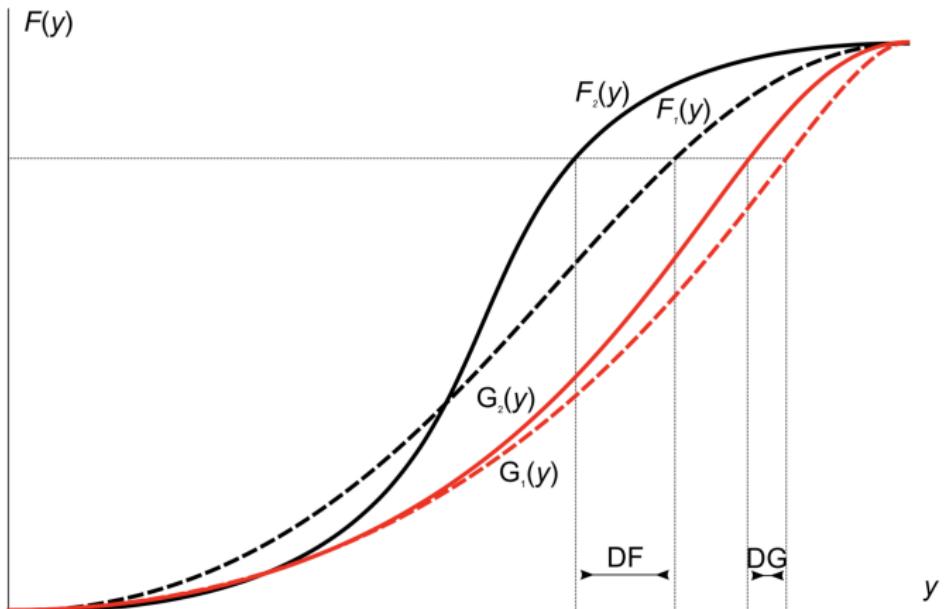
1. Fix the quantile of Y in the pre-reform outcome distribution of the treatment group, $F_1(Y) = \tau$
2. Counterfactual post-reform outcome at that quantile in the treatment group:

$$\begin{aligned} k^{QDID}(\tau) &= F_1^{-1}(\tau) + \Delta^{QDID} \\ &= F_1^{-1}(F_1(y)) + (G_2^{-1}[F_1(y)] - G_1^{-1}[F_1(y)]) \end{aligned}$$

3. QTE estimate at quantile τ is then

$$F_2^{-1}(\tau) - k^{QDID}(\tau)$$

DID: Graphical representation



Three steps:

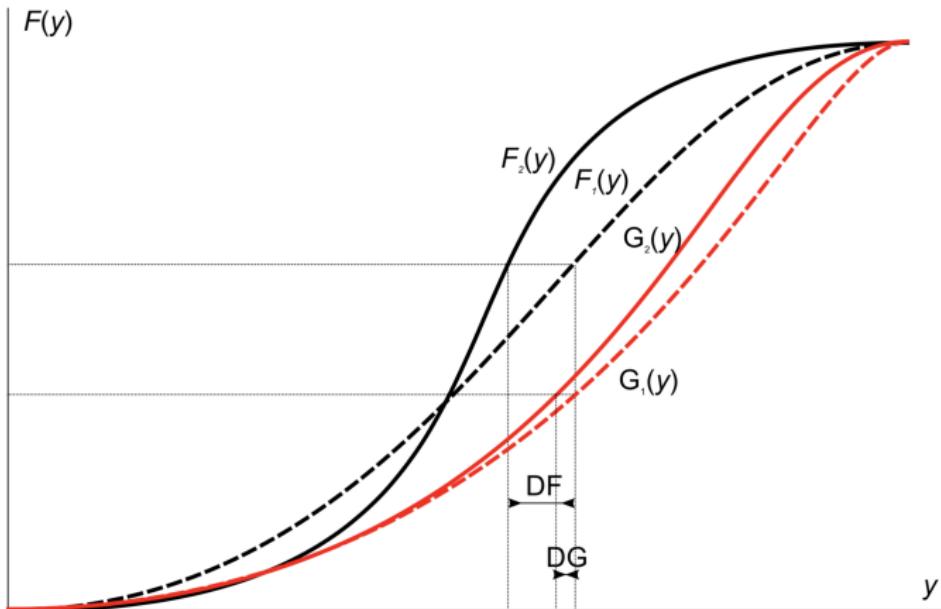
1. Fix the outcome level y , giving the quantiles in the two groups pre-reform, $F_1(y)$ and $G_1(y)$
2. Counterfactual post-reform outcome at y in the treatment group:

$$\begin{aligned}k^{CIC}(y) &= F_1^{-1}[F_1(y)] + \Delta^{CIC} \\&= y + (G_2^{-1}[G_1(y)] - G_1^{-1}[G_1(y)]) \\&= G_2^{-1}[G_1(y)]\end{aligned}$$

3. The CIC-estimate at level y is then

$$F_2^{-1}(F_1(y)) - k^{CIC}(y)$$

CIC: Graphical representation



3) Differential underlying trends

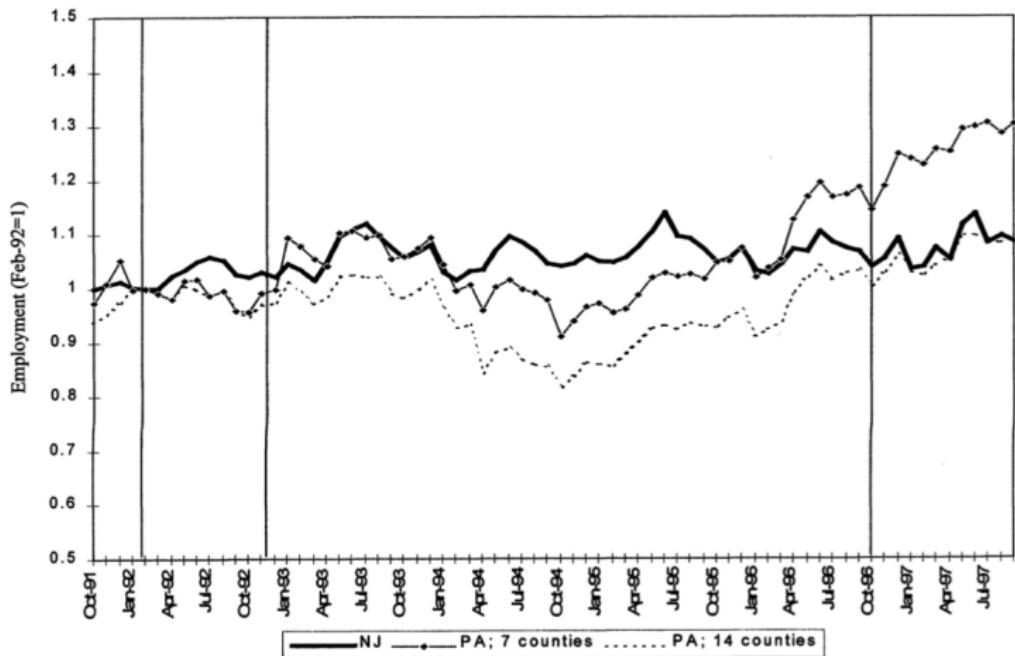
Key identifying assumption: **common trend assumption:**

In absence of the intervention the treatment and control group(s) should have common trends in the outcome variable

- ▶ assumption is formally untestable
- ▶ yet common pre-intervention trends would probably increase our (or at least my) confidence in common trend assumption

DID: Common Trend Assumption

Card & Krueger (2000), data on employment from October 1991-1997:



DID: Common Trend Assumption

Ashenfelter's Dip

Well known reason for violation of common trend assumption:
“Ashenfelter’s Dip”

Ashenfelter (1978) noted that enrollment in a training program is more likely if temporary dip in earnings occurs just before start of program

As a consequence earnings growth after enrollment likely different for participants even without treatment

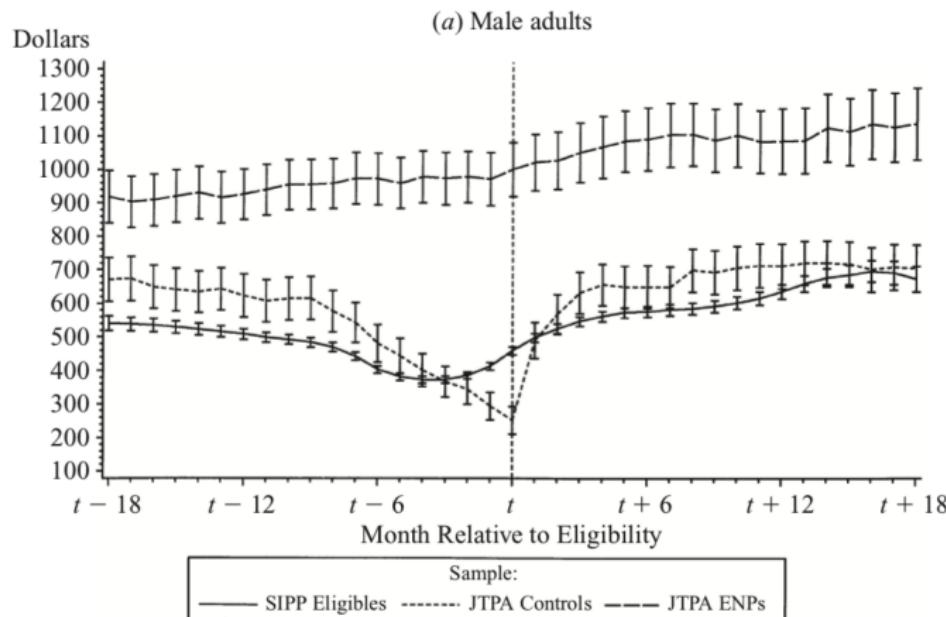
Heckman & Smith (1999) investigate earnings growth for randomized-out participants of Job Training Partnership Act program.

They show that randomized-out participants show larger earnings growth than nonparticipants

Due to this violation of common trend assumption DID estimator overestimates effect of treatment.

DID: Common Trend Assumption

Ashenfelter's Dip



Note: Figure from Heckman and Smith (EJ, 1999)

DID: Common Trend Assumption

What people do when common trend assumption is unlikely to hold:

- ▶ Including time varying covariates and/or group specific time trends
- ▶ Difference-in-difference-in-differences
- ▶ Difference-in-differences + Instrumental variable approach
- ▶ Semiparametric DiD (Abadie, 2005)
 - ▶ Common trend conditional on covariates
- ▶ Matching on pre-trends (e.g. Blundell et al., 2001)
 - ▶ DiD for controls with similar pre-trends / levels
- ▶ Synthetic control group approach (Abadie, 2010)

But first, let's talk about inference.

DID: standard errors

Correlation between residuals within a state-time period

$$Y_{igt} = \beta \cdot D_{gt} + \alpha_g + \lambda_t + \varepsilon_{igt} \quad \text{for } g = 1, \dots, G \quad \text{and } t = 1, \dots, T$$

Observations in group g at time t unlikely independent:

$$E[\varepsilon_{igt}\varepsilon_{jgt}] \neq 0$$

Possible solutions when $E[\varepsilon_{igt}\varepsilon_{jgt}] \neq 0$

parametric: assume $\varepsilon_{igt} = v_{gt} + \epsilon_{igt}$ (and use WLS)

use cluster-robust s.e.'s (robust to any correlation structure and heteroskedasticity)

block bootstrap (resample group-time periods instead of individual observations)

use group-time period averages: $\bar{Y}_{gt} = \beta \cdot D_{gt} + \alpha_g + \lambda_t + \bar{\varepsilon}_{gt}$

DID: standard errors

Correlation between residuals within a group-time period

Group-time period random shocks cause correlation between residuals within a group-time period

- ▶ bad news for DID with 2 groups and 2 time periods

Consider minimum wage example but now with state-time period random shocks v_{gt}

$$Y_{igt} = \beta \cdot D_{gt} + \alpha_g + \lambda_t + v_{gt} + \epsilon_{igt}, \quad \text{with } E[v_{gt}] = E[\epsilon_{igt}] = 0$$

DID estimator inconsistent:

$$\begin{aligned} plim(\hat{\beta}) &= (E[Y_{iA1}] - E[Y_{iA0}]) - (E[Y_{iB1}] - E[Y_{iB0}]) \\ &= \beta + (v_{A1} - v_{A0}) - (v_{B1} - v_{B0}) \neq \beta \end{aligned}$$

DID: standard errors

Serial correlation

DID with many groups and more than 2 time periods:

- ▶ most common solution when $E[\varepsilon_{igt}\varepsilon_{jgt}] \neq 0$: cluster by group \times time

$T > 2$ there is likely a serial correlation problem: $E[v_{gt}v_{gs}] \neq 0$

Bertrand, Duflo and Mullainathan (QJE, 2004) investigate the consequences of ignoring serial correlation in DID

- ▶ randomly generate placebo laws in state-level data on female wages (50 states, 21 time periods)
- ▶ compute DID estimates and s.e.'s and find an "effect" significant at 5% level in 45% of the placebo interventions!

Three factors make serial correlation important issue for DID

- ▶ DID often based on long time series
- ▶ common used Y_{igt} typically highly positively serially correlated
- ▶ treatment variable D_{gt} highly serially correlated

DID: standard errors

Serial correlation

BDM perform Monte Carlo simulations to investigate how several estimation techniques deal with serial correlation:

1. Parametric corrections (assume AR(1) process

$$u_{gt} = \rho u_{g,t-1} + \epsilon_{gt}$$

- ▶ performs poorly: with short time series $\hat{\rho}$ is biased downward

2. Nonparametric block bootstrap (entire groups are resampled):

- ▶ performs well when number of groups is reasonably large ($>=50$)

3. Aggregate data into one pre- and one post-intervention period:

- ▶ performs well also with small number of groups (with small G adjust t-statistics, Donald & Lang; REStat 2007)

4. Use empirical variance-covariance matrix (assuming homoskedasticity and common autocorrelation across states)

- ▶ performs well with enough groups

5. Cluster s.e.'s at level of the group (instead of group \times time)

- ▶ performs well when number of groups is reasonably large (≥ 50)

DID: standard errors

Example: Effect of changes in minimum school leaving age on years of education

$$Y_{igt} = \pi D_{gt} + \alpha_g + \lambda_t + \varepsilon_{igt}$$

- ▶ Y_{igt} is years of education of individual i who lived in state g in year t when he was 14 years old
- ▶ D_{gt} minimum school leaving age in state g in year t
- ▶ data on 46,154 individuals, 44 states & 61 years (1915–1975)

	(1)	(2)	(3)
Minimum school leaving age (D_{gt})	0.19*** (0.008)	0.19*** (0.011)	0.19*** (0.058)
F-test ($H_0 : \pi = 0$)	621.7	322.1	10.7
Clustering of s.e.'s	None	State × Year	State

s.e. in column (3) nearly 8 times higher than s.e. in column (1) !

DID: standard errors

Small number of groups

Cluster-robust estimate of variance-covariance matrix:

$$\hat{W} = (V'V)^{-1} \left(\sum_{g=1}^G \hat{u}_g \hat{u}_g' \right) (V'V)^{-1}$$

with V the matrix with year dummies, state dummies and treatment dummy and \hat{u}_g the vector of pooled OLS residuals (multiplied by vector of independent variables) for group g

- ▶ Summation in equation for \hat{W} is over groups; Cluster-robust se's therefore consistent for number of groups going to infinity
- ▶ Often heard at seminars: With less than (about) 50 groups se's are incorrect (most often too small!)

Clustered Standard Errors with Few Clusters

Few clusters

- ▶ Clustering is an asymptotic argument
- ▶ So something needs to go to infinity....
- ▶ Is the number of clusters “large enough”?
- ▶ Hard to tell but many DiD papers with clusters have very few clusters

Possible solutions?

- ▶ Wild bootstrap - widely used based on simulation evidence
→ But requires strong assumptions to work (Canay, Santos, Shaikh 2018)
- ▶ Permutation inference (requires other assumptions) - Canay et al (2017)
- ▶ Small sample approximations derived under parametric assumptions
- ▶ Good survey by Cameron and Miller (2015) for all gory details

DID: Extensions

A few extensions of DiD used when common trend assumption is unlikely to hold:

- ▶ Including time varying covariates and/or group specific time trends
- ▶ Difference-in-difference-in-differences
- ▶ Difference-in-differences + Instrumental variable approach
- ▶ Synthetic control group approach (not really an extension; conceptually different)

Time varying covariates and/or group specific time trends

Common trend assumption can be relaxed by including time varying covariates (X_{gt}) and group specific (linear) time trends ($\mu_g \cdot t$)

$$Y_{igt} = \beta \cdot D_{gt} + X'_{gt}\pi + \mu_g \cdot t + \alpha_g + \lambda_t + \varepsilon_{igt}$$

We need at least three time periods to estimate model with group specific time trends

Besley & Burgess (2004) study effect of labor market regulation on manufacturing performance in Indian states

Data on labor regulation come from state amendments to the Industrial Disputes Act of 1947.

- ▶ Y_{gt} : log manufacturing output per capita in state g in year t
- ▶ D_{gt} : Labor regulation:
 - ▶ = 1 if amendments to Industrial Disputes Act were pro-worker
 - ▶ = 0 if amendments to Industrial Disputes Act were neutral
 - ▶ = -1 if amendments to Industrial Disputes Act were pro-employer

Time varying covariates and/or group specific time trends

With state specific time trends identification comes from whether law changes lead to deviations from pre-existing state-specific trends.

Table 5.2.3: Effect of labor regulation on the performance of firms in Indian states

	(1)	(2)	(3)	(4)
Labor regulation (lagged)	-0.186 (.0641)	-0.185 (.0507)	-0.104 (.039)	0.0002 (.02)
Log development expenditure per capita	0.240 (.1277)	0.184 (.1187)		0.241 (.1057)
Log installed electricity capacity per capita	0.089 (.0605)	0.082 (.0543)		0.023 (.0333)
Log state population	0.720 (.96)	0.310 (1.1923)		-1.419 (2.3262)
Congress majority		-0.0009 (.01)	0.020 (.0096)	
Hard left majority			-0.050 (.0168)	-0.007 (.0091)
Janata majority			0.008 (.0235)	-0.020 (.0333)
Regional majority			0.006 (.0086)	0.026 (.0234)
State-specific trends	NO	NO	NO	YES
Adjusted R-squared	0.93	0.93	0.94	0.95

Apparently labor regulation in India increased in states where output was declining anyway.

Difference-in-difference-in-differences

Sometimes a third difference might work when you suspect violation of common trend

For example: a state implements change in health care policy for people 65 and older

- ▶ Possible DID 1: data on health in treatment state before and after, for people ≥ 65 & for people < 65 (control group)
 - ▶ violation common trend: different trends between old & young people
- ▶ Possible DID 2: data on health before and after, for people ≥ 65 in treatment state & in neighboring state (control group)
 - ▶ violation common trend: two states might have different trends

Difference-in-difference-in-differences

- ▶ Difference-in-difference-in-differences:
 - ▶ DID in treated state with people <65 as control group
 - ▶ DID for same age contrast, but in control state
 - ▶ Difference-in-Difference-in-Differences:

$$DID_{\text{treatment state}} - DID_{\text{control state}}$$

- ▶ Identification is now coming from differences in the differential trends between the age groups in the treated versus controls states
 - ▶ In other words, would have had the same DiD in both states, in the absence of the treatment
- ▶ Regression DIDID: treatment effect is coefficient on triple interaction (with levels and double interactions as controls)

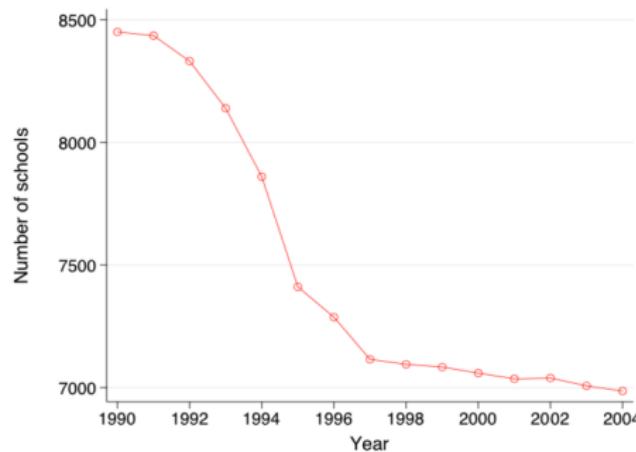
Difference-in-differences + instrumental variable(s)

Possible solution when treatment variable is potentially related to group specific trends:

- ▶ find variable that affects treatment but is (arguably) unrelated to group specific trends (instrumental variable)

Example: Effect of supply of schools on pupils test scores
(combination of choice and scale effects)

See Haan et al. (2018)



Difference-in-differences + instrumental variable(s)

What is the effect of the supply of schools on pupil test scores?

Problem: municipalities with many schools are different from municipalities with fewer schools

Possible solution: investigate *changes* in number of schools within municipalities over time (municipality fixed effects)

But change in number of schools can be related to change in (unobserved) municipality characteristics (violation common trend assumption)

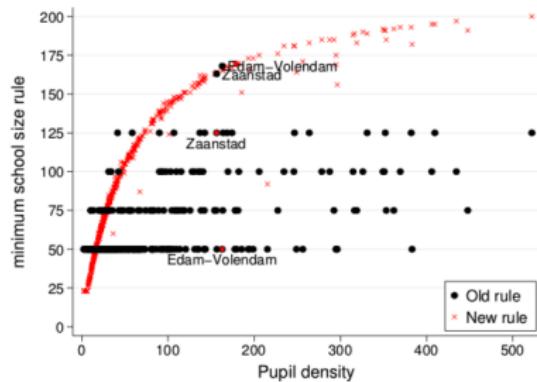
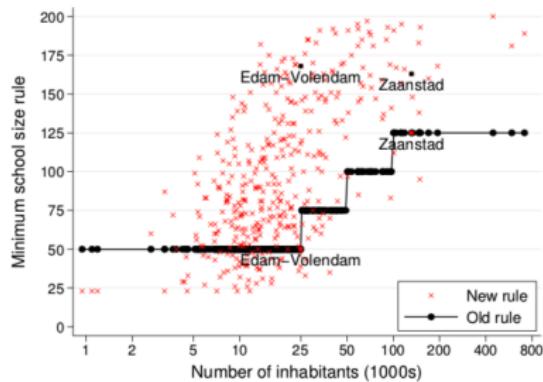
DID+IV approach:

- ▶ Include municipality and year fixed effects (DID)
- ▶ AND isolate change in number of schools which is due to a reform (instrumental variable approach)

Difference-in-differences + instrumental variable(s)

School only receives funding if number of pupils is above minimum required school size.

On January 1, 1994 a new law was enforced that changed minimum school size rule.



On average minimum required school size increased and many schools merged in order to comply with new rule

Difference-in-differences + instrumental variable(s)

$$Y_{imt} = \alpha + \delta \cdot \ln(s_{mt}) + \alpha_m + \lambda_t + \varepsilon_{imt}$$

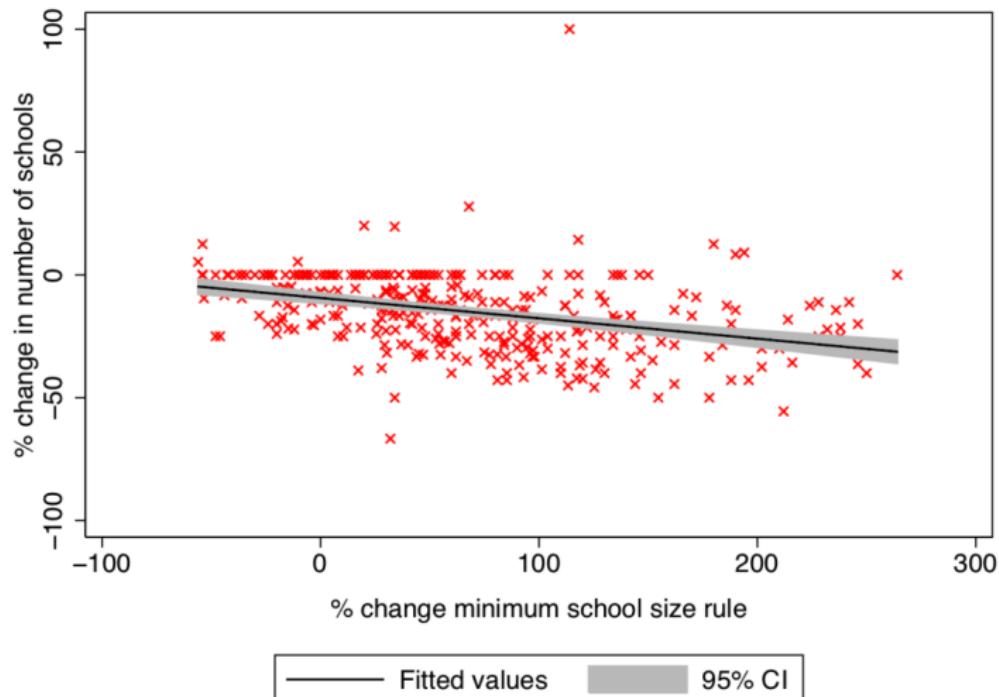
- ▶ Y_{imt} : test score pupil i living in municipality m in year t
- ▶ s_{mt} : number of schools in municipality m in year t
- ▶ α_m : municipality fixed effect
- ▶ λ_t : year fixed effects

First stage:

$$\ln(s_{mt}) = \gamma \cdot \ln(z_{mt}) + \eta_m + \tau_t + \nu_{imt}$$

- ▶ z_{mt} : minimum required school size based on the number of inhabitants and pupil density in pre-reform year
- ▶ Identifying assumption: Change in minimum required school size and change in residual pupil achievement are mean independent.

Difference-in-differences + instrumental variable(s)



Difference-in-differences + instrumental variable(s)

$T = 2$: Cohort of pupils that finished school before reform (1992) and one cohort that started & finished school after reform (2003)

Dependent variable: standardized score on nationwide test at end of primary school

	DID	DID+IV
ln(# of schools in municipality)	-0.08 (0.06)	-0.26** (0.11)
Partial F-stat first stage		91.4
municipality fixed effects	yes	yes
year fixed effect	yes	yes
# municipalities	345	345
# observations	182509	182509

Note: S.e.'s are clustered at the municipality level. Controls: ln(# pupils), ln(# inhabitants), municipality share of ethnic minority pupils, # of people on DI, and # of jobs in the municipality.

Synthetic Controls: Motivation

Overview

- ▶ Panel data, one treated and many untreated units, many time periods
- ▶ Suppose common trends is unappealing for all untreated units
- ▶ Construct a weighted average of untreated units to use for counterfactuals
- ▶ Weighted average is chosen to match pre-period outcomes/covariates
- ▶ Idea due to Abadie and Gardeazabal (2003), Abadie et al (2010, 2015)
- ▶ Increasingly used but concerns about identification and inference remain

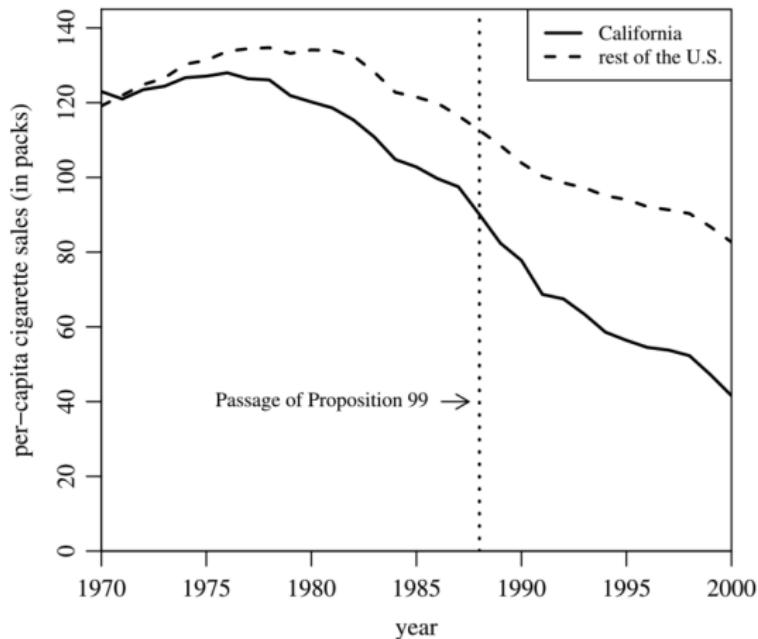
Synthetic Controls: Motivation

Empirical example (Abadie, Diamond, Hainmueller 2010)

- ▶ What was the effect of California's anti-tobacco law on cigarette sales?
- ▶ Proposition 99, passed in 1988, one of the first (modern) laws of its kind
- ▶ California already had sharply declining smoking rates before 1988
- ▶ Probably one reason the law was passed - common trends will not hold
- ▶ No other US state had the same trend - so combine several of them

Synthetic Control Group approach

Abadie, Diamond and Hainmueller (2010) apply synthetic control approach to estimate effect of a large scale tobacco control program implemented in California in 1988.



Synthetic Control Group approach

Combination of control groups often better comparison for treatment group than any single control group alone.

Synthetic control: weighted average of available control groups

- ▶ suppose: T time periods, a treatment group $g = G$ treated in final period and $G - 1$ control groups
- ▶ estimated outcome for treatment group at $t = T$ in case of no treatment: $\sum_g^{G-1} \lambda_g \cdot \bar{Y}_{gT}$ with weights λ_g chosen to minimize:

$$\left\| \begin{array}{c} \bar{Y}_{G1} - \sum_g^{G-1} \lambda_g \cdot \bar{Y}_{g1} \\ \vdots \\ \bar{Y}_{G,T-1} - \sum_g^{G-1} \lambda_g \cdot \bar{Y}_{g,T-1} \end{array} \right\|_V$$

- ▶ weights λ_g are chosen to minimize the difference between treatment and control group prior to the intervention
- ▶ diagonal V is chosen to put more weight on variables with large predictive power for the outcome of interest

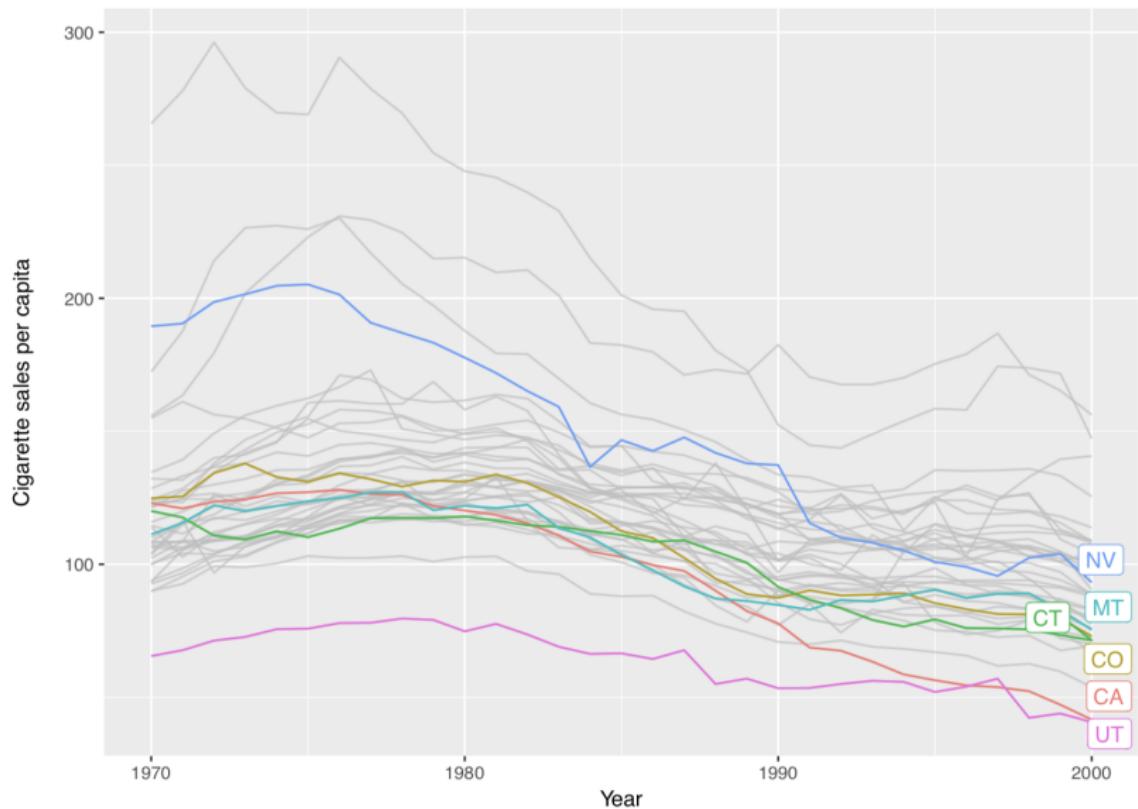
The Composition of Synthetic California

Table 2. State weights in the synthetic California

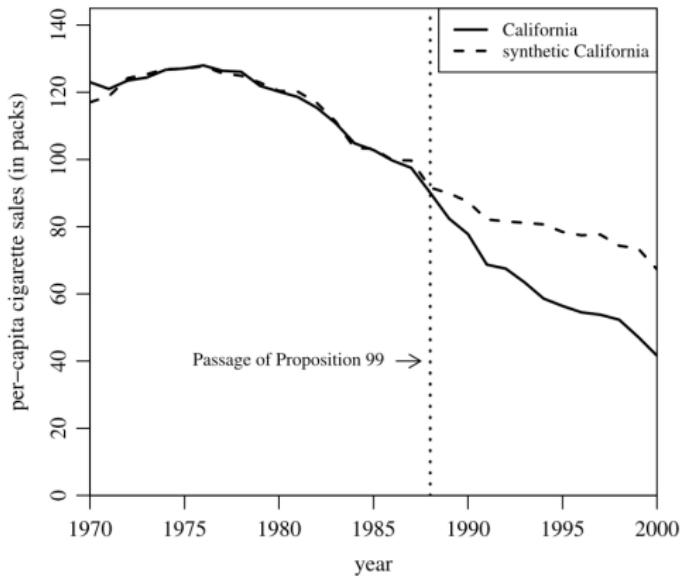
State	Weight	State	Weight
Alabama	0	Montana	0.199
Alaska	–	Nebraska	0
Arizona	–	Nevada	0.234
Arkansas	0	New Hampshire	0
Colorado	0.164	New Jersey	–
Connecticut	0.069	New Mexico	0
Delaware	0	New York	–
District of Columbia	–	North Carolina	0
Florida	–	North Dakota	0
Georgia	0	Ohio	0
Hawaii	–	Oklahoma	0
Idaho	0	Oregon	–
Illinois	0	Pennsylvania	0
Indiana	0	Rhode Island	0
Iowa	0	South Carolina	0
Kansas	0	South Dakota	0
Kentucky	0	Tennessee	0
Louisiana	0	Texas	0
Maine	0	Utah	0.334
Maryland	–	Vermont	0
Massachusetts	–	Virginia	0
Michigan	–	Washington	–
Minnesota	0	West Virginia	0
Mississippi	0	Wisconsin	0
Missouri	0	Wyoming	0

- states were removed because they passed similar laws in post-period

Series Used in Synthetic California



Synthetic California Pre-Trends



Variables used to construct Synthetic California

- ▶ Cigarette price, log per capita income, percentage aged 15-24, per capita beer consumption (all averaged over 1980-1988)
- ▶ Lagged outcome (per capita sales) in 1975, 1980, 1988

Estimated Treatment Effects

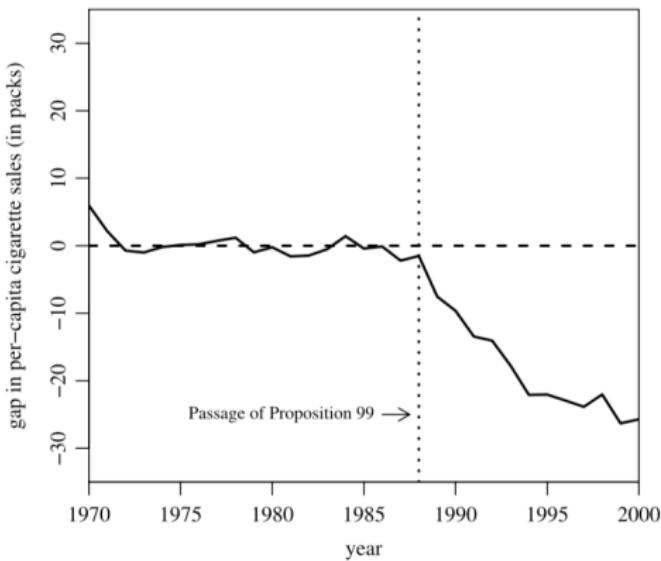


Figure 3. Per-capita cigarette sales gap between California and synthetic California.

- ▶ Subtract synthetic California from real California in post-period
- ▶ 25% reduction (20 packs) averaged over the entire post-period
- ▶ Considerably larger estimate than in previous papers

Assumptions Behind Synthetic Controls

- ▶ The motivation of synthetic controls is **weakening common trends**
 - Otherwise one could just use standard DID methods
- ▶ However clearly we need a model to justify the weighting procedure
 - Matching on pre-periods need not say anything about the post-period
- ▶ They use a factor model for untreated outcomes
- ▶ What restrictions does the factor structure place on potential outcomes?
- ▶ What restrictions on potential outcome functions allow you to construct a good control from many bad controls?

Criticisms of Synthetic Controls

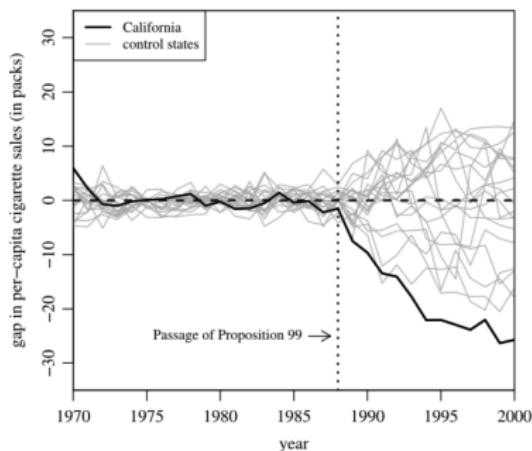
Statistical inference

- ▶ Major conceptual difficulties with statistical inference (not all new)
- ▶ Abadie et al (2010) only argue for (approximate) unbiasedness
- ▶ Eliminating variance (consistency) depends on sampling assumptions
- ▶ Abadie et al (2010) use a permutation test - unclear if justified
- ▶ Very active line of research in thinking about proper procedures

Synthetic Control Group approach

To assess significance of estimates: conduct series of placebo studies

In each iteration treatment is assigned to one of the control groups and “effect” is estimated by synthetic control group approach



Picture based on 19 control states (excl states with $MSPE > 2MSPE_{Cal}$); probability of estimating effect as large as California-effect under random permutation of intervention is 5%

Event Study

- ▶ In the standard DID, the treated units all receive treatment at the same time
- ▶ In an **event study** units receive treatment at different times
 - However being treated is still an **absorbing state**
- ▶ Sometimes there is a pure control group (untreated all t), sometimes not
 - In the former case, one may be (more) worried about comparability between treatment and controls
 - In the latter case, one may be worried about leads and lags in responses

Event Study Notation

Notation

- ▶ Time runs from $t = 0, \dots, T$ and treatment is $D_{it} \in \{0, 1\}$
- ▶ E_i denotes i's first treatment (event) date, possibly $E_i = \infty$
→ Since treatment is absorbing, $D_{it} = 1$ if and only if $t \geq E_i$
- ▶ E_i fully determines $D_i \equiv (D_{i0}, \dots, D_{iT})$ - reduces possible sequences
- ▶ $Y_{it}(e)$ - potential outcomes for $e \in \{0, 1, \dots, T, \infty\}$
→ $Y_{it} = Y_{it}(E_i)$ is what we actually observe at each t
- ▶ Note **the contrast** with the standard DID setup since D_i is not binary:

$$E_i = e \iff D_i = (0, \dots, 0, \overbrace{1}^{t=e}, \dots, 1) \text{ vs } E_i = \infty \iff D_i = (0, \dots, 0)$$

Baseline Assumptions for an Event Study

Common trends

- ▶ Analogous assumption with a control group is for any $t \geq e > s$:

$$\underbrace{\mathbb{E}[Y_{it}(\infty)|E_i = e]}_{\text{desired counterfactual}} = \underbrace{\mathbb{E}[Y_{is}(\infty)|E_i = e]}_{\text{not data}} + \underbrace{\mathbb{E}[Y_{it}(\infty) - Y_{is}(\infty)|E_i = \infty]}_{\text{observed change in the data}}$$

No anticipation condition

- ▶ Extra assumption: $Y_{is}(e) = Y_{is}(\infty)$ for all untreated periods ($s < e$)
 $\implies \mathbb{E}[Y_{is}(\infty)|E_i = e] = \mathbb{E}[Y_{is}|E_i = e]$ and the counterfactual are identified
- ▶ Also allows one to replace a pure control group with any $e' > t$
- ▶ Keep in mind “no anticipation” is a strong statement about behavior
→ It is basically embedded in common trends for a DID as well

Nonparametric Identification in an Event Study

Result

- ▶ Using the common trends and no anticipation assumptions,

$$\begin{aligned}ATE_t(e) &\equiv \mathbb{E}[Y_{it}(e) - Y_{it}(\infty)|E_i = e] \\&= (\mathbb{E}[Y_{it}|E_i = e] - \mathbb{E}[Y_{is}|E_i = e]) \\&\quad + (\mathbb{E}[Y_{it}|E_i = e'] - \mathbb{E}[Y_{is}|E_i = e']) \text{ where } e' > t \geq e > s\end{aligned}$$

- ▶ $ATE_t(e)$: average effect on Y_{it} of being treated at time e (vs never)
- ▶ Many parameters: $ATE_t(e)$ for all $e = 1, \dots, T, t > e$
- ▶ Multiple ways to estimate it: Any $s < e$ will do
- ▶ Could also aggregate across $s < e$ by averaging

Typical regression model

Two-way fixed effects (“static” specification)

- ▶ Regress Y_{it} on unit dummies, time dummies, $D_{it} \equiv \mathbb{1}[t \geq E_i]$
- ▶ Same specification as in DID with multiple groups/times
- ▶ Coefficient on D_{it} (call it β_{fe}) - the object of interest

Interpretation (Abraham and Sun 2018)

$$\beta_{fe} \equiv \sum_{e=0}^E \sum_{t=e}^T \omega_t(e) ATE_t(e)$$

- ▶ The weights are identified, sum to one, but can be negative if causal effects are heterogeneous
- ▶ Specification implicitly makes contrasts we wouldn't nonparametrically

Common Event Study Implementation 2

Adding leads and lags to the event date ("dynamic" specification)

- ▶ Regress Y_{it} on unit dummies, time dummies, and $\{R_{it}^l\}_{l \in \mathcal{L}}$
- ▶ $R_{it}^l \equiv \mathbb{1}[l = t - E_i]$ - indicator for l periods since event
- ▶ \mathcal{L} has leads ($l < 0$) and lags ($l \geq 0$) (collinearity caveat)
- ▶ Coefficients on $\{R_{it}^l\}_{l \in \mathcal{L}}$ are treated as the objects of interest
- ▶ For $l > 0$ capture **dynamic treatment effects**
- ▶ For $l < 0$ checks on the validity of the design

Interpretation (Abraham and Sun 2018)

- ▶ Similar weighting results; lead and lag weights are different
Lag weights can be negative: some lead weights must be negative
- ▶ Problem: pre-trends maybe not detectable with lead estimates
Alternatively, zero pre-trends might be estimated spuriously

What's a solution

- ▶ Problem stems from the regression specifications
→ No problem estimating $ATE_t(e)$ directly
- ▶ So why not estimate them directly, then construct whatever weighted average you care about?
- ▶ Which reminds us....
 - ▶ Do not start with the estimator and hope it gives something interpretable
 - ▶ Start with a clear parameter of interest and then choose estimator
 - ▶ Heterogeneity in effects complicates the interpretation of estimates from commonly used estimators
 - ▶ Can you really give the results a causal interpretation? What's the (policy) question they answer?

Example of event study: Income effects and labor supply

Labor supply responses to taxation are of fundamental importance for income tax policy

- ▶ efficiency costs and optimal tax formulas

Ideally, we would like to know compensated elasticities of different people

- ▶ E.g. across income distribution

Saez (2010) (claims to) develop method of using bunching at kinks to estimate the compensated income elasticity

- ▶ Argues compensated elasticities are small, especially for high income households
- ▶ Method requires strong assumptions about counterfactual earnings density
- ▶ Assumes no income effect (quasi-linear utility) so comp and uncomp el. are the same

In an ongoing project, we aim to estimate compensated elasticities across income distribution, allowing income effects to vary across

Example of event study: Income effects and labor supply

Data

- ▶ Sample: all individuals winning \$600+ from a U.S. lottery between 2001 and 2016
- ▶ Outcomes of interest: employment/LFP and earnings for 1999 to 2015

Summary of design

- ▶ Lottery participation (and winning) may not be exogenous
- ▶ However, conditional on winning, **timing** of win may be
 1. difficult for an agent to control
 - ⇒ limited scope for anticipation effects
 2. independent of any pre-existing trends in LFP/earnings
 - ⇒ parallel trends in outcomes

To illustrate, focus on \$1M+ winners and their employment

Breaking the Problem into Cohorts

Define a cohort as a group of individuals i that win their \$1M+ in a given year e (the e cohort)

A family of parameters of interest specific to cohort e are the cohort-specific ATTs:

$$CATT_t(e) \equiv E[Y_{it}(E_i) - \underbrace{Y_{it}(\infty)}_{\text{potential outcome if treated at } t=\infty} | E_i = e]$$

These $CATT_t(e)$ parameters will serve as building blocks.

Walkthrough with Two Cohorts

Suppose we only have two cohorts of winners — 2004 and 2016

Given the data are 1999 to 2015, we can hope to identify ATTs specific to the 2004 cohort up to

- ▶ five years before winning ($2004 - 1999 = -5$)
- ▶ eleven years after winning ($2015 - 2004 = 11$)

How? Let's illustrate for the effect one year after winning

$$\begin{aligned} CATT_{2005}(2004) &\equiv E [Y_{i,2005}(2004) - Y_{i,2005}(\infty) \mid E_i = 2004] \\ &= E [Y_{i,2005}(2004) - Y_{i,2005}(\infty) \mid E_i = 2004] \\ &\quad + \underbrace{E [Y_{i,2003}(\infty) - Y_{i,2003}(\infty) \mid E_i = 2004]}_{=0} \\ &= \underbrace{E [Y_{i,2005}(2004) - Y_{i,2003}(\infty) \mid E_i = 2004]}_{(A)} \\ &\quad - \underbrace{E [Y_{i,2005}(\infty) - Y_{i,2003}(\infty) \mid E_i = 2004]}_{(B)} \end{aligned}$$

Walkthrough with Two Cohorts (continued)

Both (A) and (B) are functions of potential outcomes that we wish to map to observed quantities.

Using our no anticipation assumption, for all $t < e$, $Y_{it} = Y_{it}(\infty)$

- ▶ i.e, during pre-treatment years, observed outcome = potential outcome in the untreated state

We can then rewrite (A) in terms of observable quantities - a “long” difference in means for the 2004 winners:

$$\begin{aligned}(A) &\equiv E [Y_{i,2005}(2004) - Y_{i,2003}(\infty) | E_i = 2004] \\&= E [Y_{i,2005}(2004) - \textcolor{red}{Y_{i,2003}(2004)} | E_i = 2004] \\&= E [Y_{i,2005}(2004)] - E [Y_{i,2003}(2004)]\end{aligned}$$

Walkthrough with Two Cohorts (continued)

Returning to (B), we will use both the no anticipation and parallel trends assumptions

Recall, in this context, parallel trends is mean independence of outcome growth wrt time of event, e.g., $\forall t \neq s$ and $e \neq e'$

$$E [Y_{it}(\infty) - Y_{is}(\infty) | E_i = e] = E [Y_{it}(\infty) - Y_{is}(\infty) | E_i = e']$$

With this, we can substitute in (B) using the later-treated (2016) cohort:

$$\begin{aligned}(B) &\equiv E [Y_{i,2005}(\infty) - Y_{i,2003}(\infty) | E_i = 2004] \\ &= E [Y_{i,2005}(\infty) - Y_{i,2003}(\infty) | E_i = 2016]\end{aligned}$$

Finally, as we did with (A), the no anticipation assumption implies

$$\begin{aligned}&E [Y_{i,2005}(\infty) - Y_{i,2003}(\infty) | E_i = 2016] \\ &= E [Y_{i,2005}(2016) - Y_{i,2003}(2016) | E_i = 2016] \\ &= E [Y_{i,2005}(2016)] - E [Y_{i,2003}(2016)]\end{aligned}$$

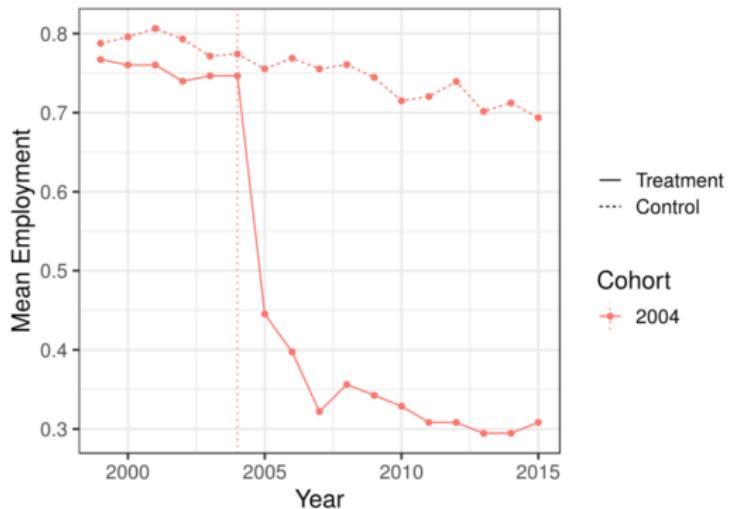
Walkthrough with Two Cohorts (continued)

To recap, we have now shown that $CATT_{2005}(2004)$ is identified with a difference-in-differences using only observed quantities:

$$\begin{aligned}CATT_{2005}(2004) &\equiv E [Y_{i,2005}(2004) - Y_{i,2005}(\infty) \mid E_i = 2004] \\&= E [Y_{i,2005}(2004)] - E [Y_{i,2003}(2004)] \\&\quad - (E [Y_{i,2005}(2016)] - E [Y_{i,2003}(2016)])\end{aligned}$$

Using the 2016 cohort, we can apply this approach to all other observed years

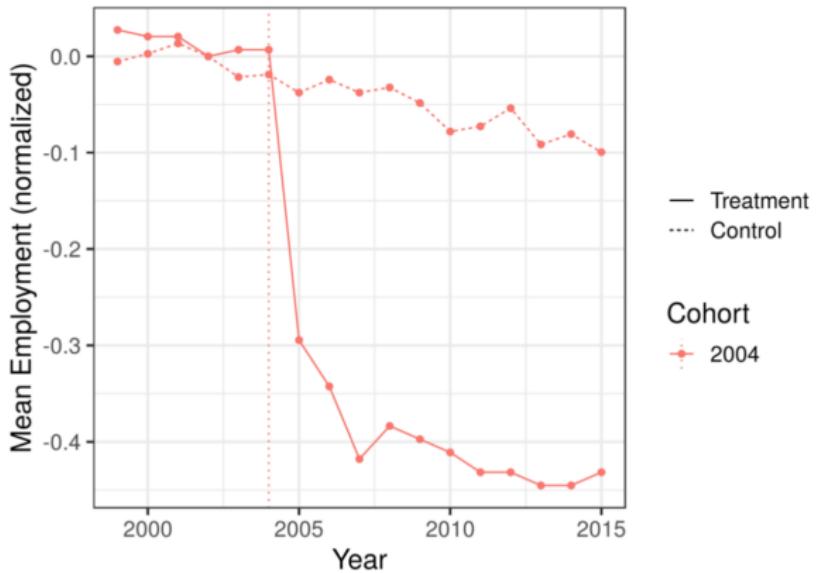
Mean Employment of 2004 and 2016 Winners



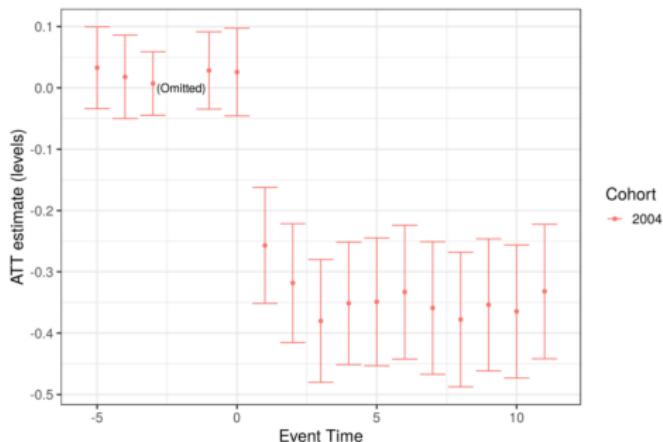
In years prior to the win year (2004), employment of 2004 (treated) and 2016 winners (control) behaves similarly

In years after 2004, change in outcomes of 2016 relative to chosen pre-period year serve as counterfactual growth

Subtracting off the 2002 mean for each group



Employment CATTs for the 2004 Winners over time

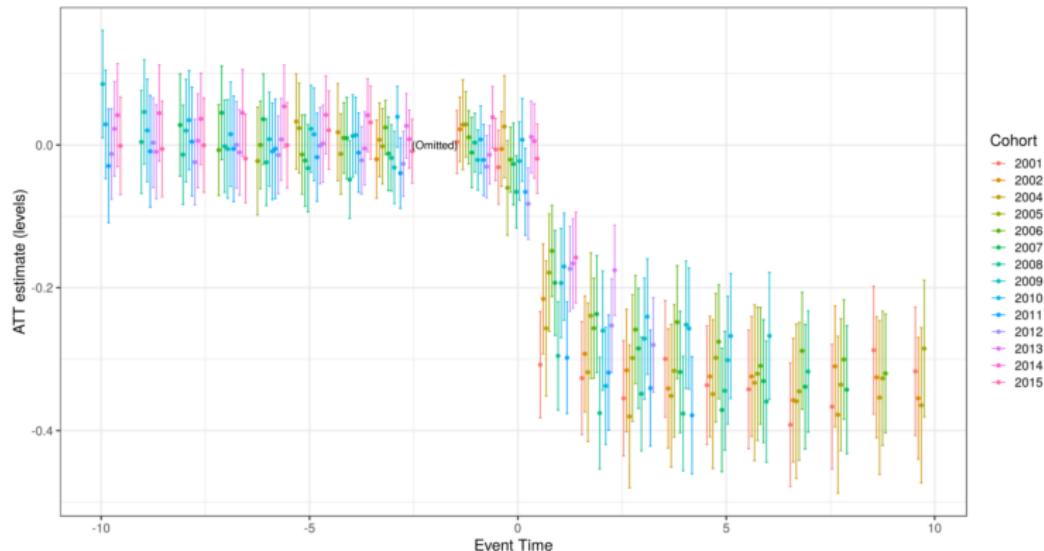


Here, the x-axis has now mapped calendar time for 2004 winners into event time — just a translation

Conditional on no anticipation, do not reject parallel trends in pre-treatment period

- ▶ Parallel trends in post-period is untestable, but hope that pre-period is informative

2001-2015 Employment CATTs (2016 as Control Group)



As before, x-axis is time relative to each cohort's win year

- ▶ moving to the right from 0 will filter to earlier win years
- ▶ moving to the left from 0 will filter to later win years

Aggregating CATTs

We have focused on identifying and estimating all CATTs (given constraints of the data)

What if instead of cohort-specific questions, we wish to answer questions such as:

1. what is the average effect of lottery winning on employment N years after winning?
2. how different is effect of lottery winning on employment in 2008-2010 (recession years)?

To address such questions, we can use the estimated CATTs as building blocks with suitable weights

Example: Average +1 Effects

Question: What is the average effect of lottery winning on employment one year after winning?

Example: Average +1 Effects

Step 1: Identify relevant CATTs to address question

- ▶ $CATT_{2002}(2001), CATT_{2003}(2002), \dots, CATT_{2015}(2014)$
- ▶ 14 in total; $\mathbb{E} = \{2001, 2002, \dots, 2014\}$

Step 2: Calculate weighted average $\sum_{e \in \mathbb{E}} \omega_e CATT_{e+1}(e)$

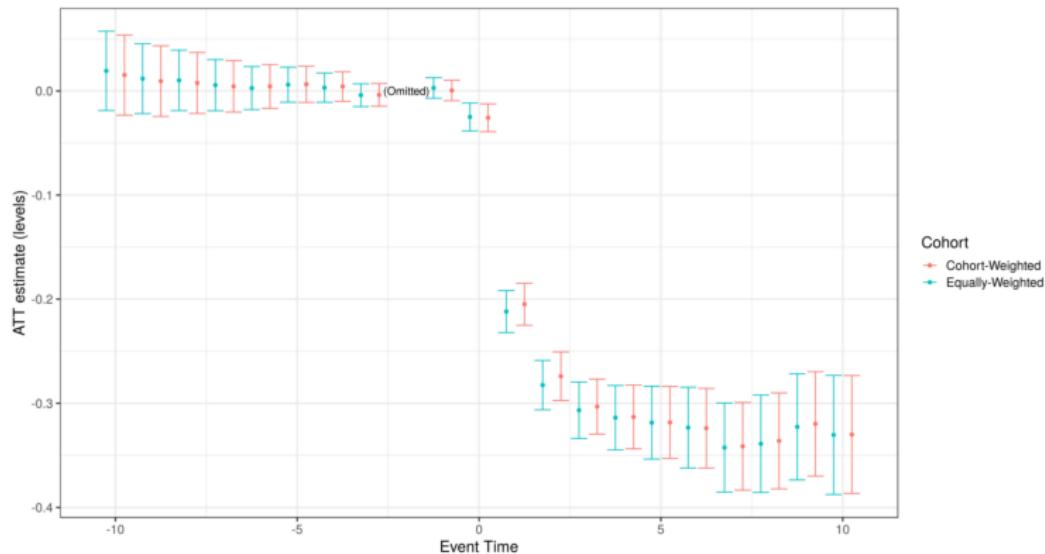
- ▶ Equally-weighted average: $\omega_e = \frac{1}{|\mathbb{E}|}$
- ▶ Cohort-size-weighted average: $\omega_e = \frac{N_e}{\sum_{e \in \mathbb{E}} N_e}$ where N_e is the size of cohort e

No reason to limit yourself to either of above – with CATTs in hand, chose weights to answer economic question of interest

Confidence intervals for above averages are straightforward

- ▶ delta method / bootstrap

Average CATTs for Employment



In answer to our question, +1 average effect is approx. 20pp

- ▶ both illustrated weighting approaches give similar estimates

Effect grows up to +5 and appears sustained afterwards

Revisiting the Choice of Control Group

In the prior exercises, we always held fixed that the 2016 winners were the control group

But this ignores all of the intermediate years where, assuming no anticipation, other cohorts can also serve as controls

Using the Largest Control Group Possible

Recall our worked-out example for $CATT_{2005}(2004)$:

$$\begin{aligned}CATT_{2005}(2004) &\equiv E [Y_{i,2005}(2004) - Y_{i,2005}(\infty) | E = 2004] \\&= E [Y_{i,2005}(2004)] - E [Y_{i,2003}(2004)] \\&\quad - E [Y_{i,2005}(2016)] - E [Y_{i,2003}(2016)]\end{aligned}$$

But under the same maintained assumptions, could instead do:

$$\begin{aligned}CATT_{2005}(2004) &\equiv E [Y_{i,2005}(2004) - Y_{i,2005}(\infty) | E_i = 2004] \\&= E [Y_{i,2005}(2004)] - E [Y_{i,2003}(2004)] \\&\quad - E [Y_{i,2005}(e') | e' \geq 2006] - E [Y_{i,2003}(e') | e' \geq 2006]\end{aligned}$$

Using the Largest Control Group Possible

Similarly, for the +2 effect for 2004 winners:

$$\begin{aligned} CATT_{2006}(2004) &\equiv E [Y_{i,2006}(2004) - Y_{i,2006}(\infty) | E_i = 2004] \\ &= E [Y_{i,2006}(2004)] - E [Y_{i,2004}(2004)] \\ &\quad - E [Y_{i,2006}(e') | e' \geq 2007] - E [Y_{i,2003}(e') | e' \geq 2007] \end{aligned}$$

Under no anticipation, many choices on which pre-period to use

Still, key moving parts are always:

1. For a given cohort (e.g., 2004 winners), the control group must be treated later (e.g., 2005+)
2. For a given effect (+2 for 2004 winners), the control group must be treated later than the corresponding calendar time (e.g., 2007+)