

Part 8: Policy Evaluation

Difference in Difference

Chris Conlon

April 6, 2020

Applied Econometrics

Motivation: Recap Matching

Matching estimators had some advantages:

- Limited assumptions on **functional forms**
- We could do nearest neighbor matching and use kernels to compute treatment effects

Matching estimators had some drawbacks:

- Treated patients were “matched” to control patients based only on **observable characteristics**
 - Ignored **selection on unobservables**.
- Relied on **cross sectional** variation to construct a control group.

Motivation

IV estimators resolve some of those issues but

- Good IV are in short supply!

Often (in this course at least) we have access to **panel data**.

- What if we could use panel data to control for **unobserved heterogeneity** within a treated individual/group?

Difference in Difference estimators are like the opposite of matching

- **Strong** assumptions on **functional form**
- but... allow for **unobservable heterogeneity** in outcomes.

A Famous Example: Card and Krueger (AER 1994)

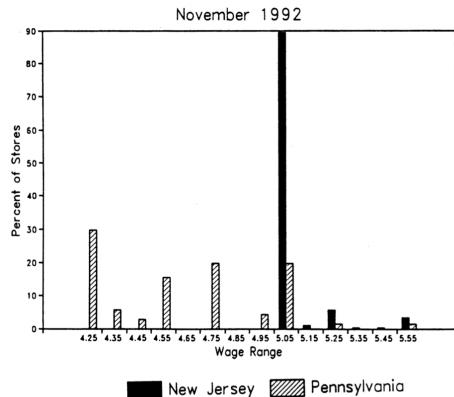
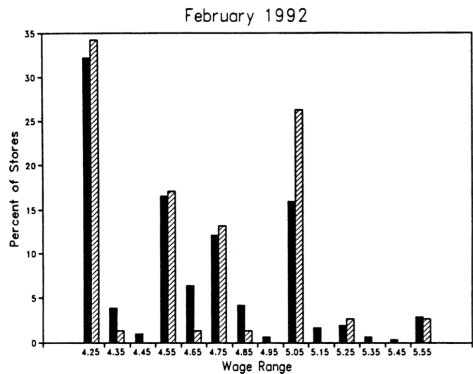
A Famous Example: Card and Krueger (AER 1994)

- On April 1, 1992 NJ raises its minimum wage from \$4.25 \rightarrow \$5.05 per hour.
- Question: Econ 101 predicts this will **reduce demand for low wage workers**
 - Focus on fast food restaurants (since they pay min wage)
 - Focus on starting wage (avoid tenure, high turnover)
- Survey 410 restaurants in NJ (treated group) and eastern PA (control group).
- Idea: Compare **change** in wages in *NJ* to *PA*: $\Delta_{DD} = \Delta_{NJ} - \Delta_{PA}$
 - Wave 1: February 15-March 4, 1992
 - Wave 2: November 5 - December 31, 1992

Balance Table: Covariates

Variable	Stores in:		<i>t</i> ^a
	NJ	PA	
1. <i>Distribution of Store Types (percentages):</i>			
a. Burger King	41.1	44.3	-0.5
b. KFC	20.5	15.2	1.2
c. Roy Rogers	24.8	21.5	0.6
d. Wendy's	13.6	19.0	-1.1
e. Company-owned	34.1	35.4	-0.2
2. <i>Means in Wave 1:</i>			
a. FTE employment	20.4 (0.51)	23.3 (1.35)	-2.0
b. Percentage full-time employees	32.8 (1.3)	35.0 (2.7)	-0.7
c. Starting wage	4.61 (0.02)	4.63 (0.04)	-0.4
d. Wage = \$4.25 (percentage)	30.5 (2.5)	32.9 (5.3)	-0.4
e. Price of full meal	3.35 (0.04)	3.04 (0.07)	4.0
f. Hours open (weekday)	14.4 (0.2)	14.5 (0.3)	-0.3
g. Recruiting bonus	23.6 (2.3)	29.1 (5.1)	-1.0
3. <i>Means in Wave 2:</i>			
a. FTE employment	21.0 (0.52)	21.2 (0.94)	-0.2
b. Percentage full-time employees	35.9 (1.4)	30.4 (2.8)	1.8
c. Starting wage	5.08 (0.01)	4.62 (0.04)	10.8
d. Wage = \$4.25 (percentage)	0.0	25.3 (4.9)	—
e. Wage = \$5.05 (percentage)	85.2 (2.0)	1.3 (1.3)	36.1
f. Price of full meal	3.41 (0.04)	3.03 (0.07)	5.0
g. Hours open (weekday)	14.4 (0.2)	14.7 (0.3)	-0.8
h. Recruiting bonus	20.3 (2.3)	23.4 (4.9)	-0.6

Distribution of Wages



Differences in Wages : 2 x 2 Table

TABLE 3—AVERAGE EMPLOYMENT PER STORE BEFORE AND AFTER THE RISE
IN NEW JERSEY MINIMUM WAGE

Variable	Stores by state			Stores in New Jersey ^a			Differences within NJ ^b	
	PA (i)	NJ (ii)	Difference, NJ – PA (iii)	Wage = \$4.25 (iv)	Wage = \$4.26–\$4.99 (v)	Wage ≥ \$5.00 (vi)	Low– high (vii)	Midrange– high (viii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	–2.89 (1.44)	19.56 (0.77)	20.08 (0.84)	22.25 (1.14)	–2.69 (1.37)	–2.17 (1.41)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	–0.14 (1.07)	20.88 (1.01)	20.96 (0.76)	20.21 (1.03)	0.67 (1.44)	0.75 (1.27)
3. Change in mean FTE employment	–2.16 (1.25)	0.59 (0.54)	2.76 (1.36)	1.32 (0.95)	0.87 (0.84)	–2.04 (1.14)	3.36 (1.48)	2.91 (1.41)
4. Change in mean FTE employment, balanced sample of stores ^c	–2.28 (1.25)	0.47 (0.48)	2.75 (1.34)	1.21 (0.82)	0.71 (0.69)	–2.16 (1.01)	3.36 (1.30)	2.87 (1.22)
5. Change in mean FTE employment, setting FTE at temporarily closed stores to 0 ^d	–2.28 (1.25)	0.23 (0.49)	2.51 (1.35)	0.90 (0.87)	0.49 (0.69)	–2.39 (1.02)	3.29 (1.34)	2.88 (1.23)

Notes: Standard errors are shown in parentheses. The sample consists of all stores with available data on employment. FTE (full-time-equivalent) employment counts each part-time worker as half a full-time worker. Employment at six closed stores is set to zero. Employment at four temporarily closed stores is treated as missing.

^aStores in New Jersey were classified by whether starting wage in wave 1 equals \$4.25 per hour ($N = 101$), is between \$4.26 and \$4.99 per hour ($N = 140$), or is \$5.00 per hour or higher ($N = 73$).

^bDifference in employment between low-wage (\$4.25 per hour) and high-wage ($\geq \$5.00$ per hour) stores; and difference in employment between midrange (\$4.26–\$4.99 per hour) and high-wage stores.

^cSubset of stores with available employment data in wave 1 and wave 2.

^dIn this row only, wave-2 employment at four temporarily closed stores is set to 0. Employment changes are based on the subset of stores with available employment data in wave 1 and wave 2.

Outcome Equation

- Differences lack any covariates (different fast food chains).
- Also $\Delta_{PA} < 0$ and $\Delta_{NJ} > 0$ (!)
- Recall i denotes stores, $t \in 1, 2$. Run the following regression:

$$Y_{it} = \beta X_{it} + \alpha \cdot [i \in NJ] + \gamma \cdot \text{After}_t + \delta \cdot NJ_i \times \text{After}_t + u_i$$

$$Y_{it} = \beta X_{it} + \alpha \cdot [\text{wage gap}_i] + \gamma \cdot \text{After}_t + \delta \cdot \text{wage gap}_i \times \text{After}_t + u_i$$

- α is mean difference between NJ and PA
- γ is mean difference between period 1 and 2
- δ is the parameter of interest, the **difference in difference**
- $\text{wage gap}_i = [\min \text{wage}_{i,2} - w_{i1}]_+ = \max\{0, \min \text{wage}_{i,2} - w_{i1}\}$.
(How much do you need to raise $t = 1$ wages to achieve minimum wage in $t = 2$?)

Differences in Wages

TABLE 4—REDUCED-FORM MODELS FOR CHANGE IN EMPLOYMENT

Independent variable	Model				
	(i)	(ii)	(iii)	(iv)	(v)
1. New Jersey dummy	2.33 (1.19)	2.30 (1.20)	—	—	—
2. Initial wage gap ^a	—	—	15.65 (6.08)	14.92 (6.21)	11.91 (7.39)
3. Controls for chain and ownership ^b	no	yes	no	yes	yes
4. Controls for region ^c	no	no	no	no	yes
5. Standard error of regression	8.79	8.78	8.76	8.76	8.75
6. Probability value for controls ^d	—	0.34	—	0.44	0.40

Notes: Standard errors are given in parentheses. The sample consists of 357 stores with available data on employment and starting wages in waves 1 and 2. The dependent variable in all models is change in FTE employment. The mean and standard deviation of the dependent variable are -0.237 and 8.825 , respectively. All models include an unrestricted constant (not reported).

^aProportional increase in starting wage necessary to raise starting wage to new minimum rate. For stores in Pennsylvania the wage gap is 0.

^bThree dummy variables for chain type and whether or not the store is company-owned are included.

^cDummy variables for two regions of New Jersey and two regions of eastern Pennsylvania are included.

^dProbability value of joint F test for exclusion of all control variables

A More General Method

Difference in Difference: General Approach

Potential outcome in period 1 = $Y_{i1}(0)$

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

	Treatment	Control
Before	$Y_{i1}(0)$	$Y_{i1}(0)$
After	$Y_{i2}(1)$	$Y_{i2}(0)$

We can write the outcome as:

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

Difference in Difference: General Approach

Consider the first difference $\Delta Y_{it} = Y_{i2} - Y_{i1}$:

$$\Delta Y_{it} = T_{i2} (Y_{i2}(1) - Y_{i2}(0)) + Y_{i2}(0) - Y_{i1}(0)$$

For treated group (first difference):

$$E[\Delta Y_{it} | T_{i2} = 1] = \overbrace{E[Y_{i2}(1) - Y_{i2}(0) | T_{i2} = 1]}^{ATT} + \overbrace{E[Y_{i2}(0) - Y_{i1}(0) | T_{i2} = 1]}^{\gamma(1)}$$

For control group (second difference):

$$E[\Delta Y_{it} | T_{i2} = 0] = \overbrace{E[Y_{i2}(0) - Y_{i1}(0) | T_{i2} = 0]}^{\gamma(0)}$$

The DiD (difference in difference) estimator

$$\Delta_{DD} = E[\Delta Y_{it} | T_{i2} = 1] - E[\Delta Y_{it} | T_{i2} = 0]$$

Difference in Difference: Parallel Trends

- If $\gamma(1) = \gamma(0)$ then the DiD estimator cancels and we are left with the $\Delta_{DD} = \text{ATT}$.
- This is the **parallel trends assumption**

$$E[Y_{i2}(0) - Y_{i1}(0)|T_{i2} = 0] = E[Y_{i2}(0) - Y_{i1}(0)|T_{i2} = 1]$$

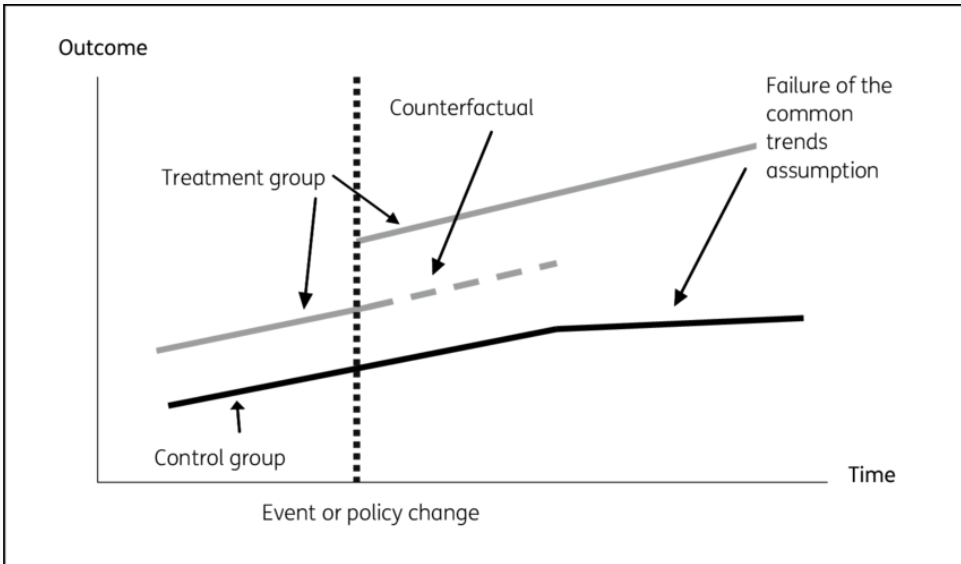
- Absent the treatment effect, both treatment and control would evolve identically over time.
- But, treatment and control groups can start from very different places...

$$E[Y_{it}(0)|T_{i2} = 1] \neq E[Y_{it}(0)|T_{i2} = 0], t = 1, 2$$

- And have selection on treatment effects...

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

Parallel Trends



Difference in Differences: Limitations

1. Functional form restrictions

- **Parallel trends** assumes that absent treatment that we add $\gamma_2 - \gamma_1$ to each unit
- Because this is **additive** it is not invariant to transformations $f(Y_{it})$ (ie: taking logs)

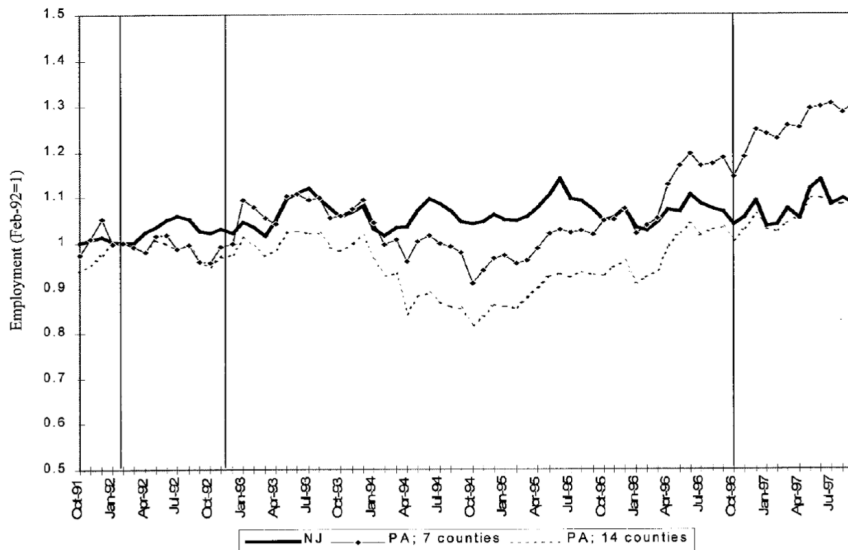
2. Parallel Trend Assumption is **not testable**

- Best we can hope is that it looks similar in the pre-period

3. Compositional Effects: the treatment may affect who is in each group

- Restaurants could close in NJ and open nearby in PA to avoid minimum wage.
- A good job training program may lead to migration, etc.
- One approach: redefine the population so that it doesn't endogenously respond to treatment
 - Recover something, but probably not ATT anymore...

Checking Pre-Trend: Card Krueger (2000)



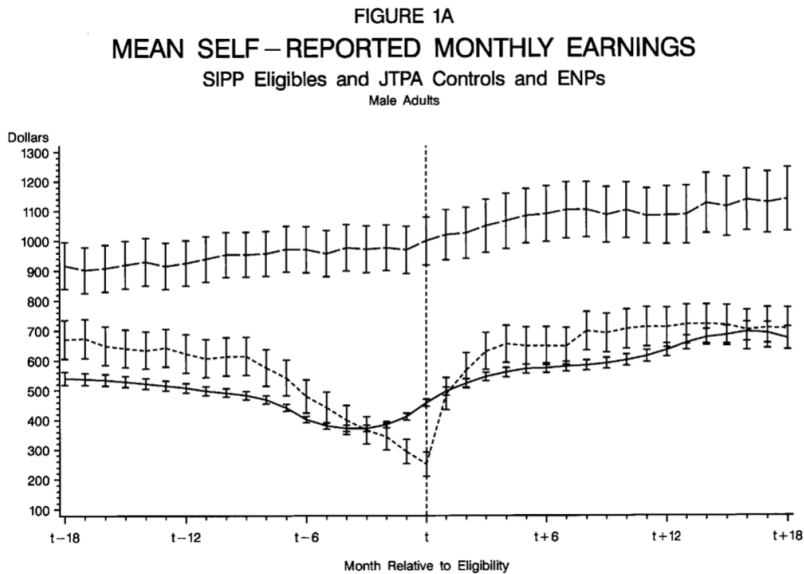
Difference in Differences

Just like in Card and Kruger, we can write as a regression equation:

$$Y_{it} = \alpha_i + \gamma_t + \beta X_{it} + \delta_i T_{it} + u_{it}$$

- Suppose we wish to evaluate a training program for those with low earnings. Let the threshold for eligibility be B .
- We have a panel of individuals and those with low earnings qualify for training, forming the treatment group.
- Those with higher earnings form the control group.
- Now the low earning group is low for two reasons
 1. They have low permanent earnings (α_i is low) - this is accounted for by diff in diffs.
 2. They have a negative transitory shock (u_{i1} is low) - this is not accounted for by diff in diffs.

The “Ashenfelter Dip” (Heckman and Smith 2000)



Difference in Differences

- #2 above violates the assumption $E[Y_{i2}(0) - Y_{i1}(0)|T] = E[Y_{i2}(0) - Y_{i1}(0)]$.
- To see why note that those participating into the program are such that $Y_{i0}(0) < B$. Assume for simplicity that the shocks u are *iid*. Hence $u_{i1} < B - \alpha_i - \gamma_1$. This implies:

$$E[Y_{i2}(0) - Y_{i1}(0)|T = 1] = \gamma_2 = \gamma_1 - E[u_{i1}|u_{i1} < B - \alpha_i - \gamma_1]$$

For the control group:

$$E[Y_{i2}(0) - Y_{i1}(0)|T = 1] = \gamma_2 = \gamma_1 - E[u_{i1}|u_{i1} > B - \alpha_i - \gamma_1]$$

$$\begin{aligned} E[Y_{i2}(0) - Y_{i1}(0)|T = 1] - E[Y_{i2}(0) - Y_{i1}(0)|T = 0] = \\ E[u_{i1}|u_{i1} > B - \alpha_i - \gamma_1] - E[u_{i1}|u_{i1} < B - \alpha_i - \gamma_1] > 0 \end{aligned}$$

- This is effectively regression to the mean: those unlucky enough to have a bad shock recover and show greater growth relative to those with a good shock. The nature of the bias depends on the stochastic properties of the shocks and how individuals select into training.

Difference in Differences

- The assumption on growth of the non-treatment outcome being independent of assignment to treatment may be violated, but it may still be true conditional on X .
- Consider the assumption

$$E[Y_{i2}(0) - Y_{i1}(0)|X, T] = E[Y_{i2}(0) - Y_{i1}(0)|X]$$

- This is just matching assumption on a redefined variable, namely the growth in the outcomes. In its simplest form the approach is implemented by running the regression

$$Y_{it} = \alpha_i + \gamma_t + \delta_i T_{it} + \beta'_t X_i + u_{it}$$

which allows for differential trends in the non-treatment growth depending on X_i . More generally one can implement propensity score matching on the growth of outcome variable when panel data is available.

Variants

Difference in Difference in Difference

The triple difference is also a thing:

- Suppose that we have: before/after, treated-state/untreated-state, treated-group (men)/ untreated-group women.
- We can compute two D-i-D here: $\Delta_{DDD} = \Delta_{DD,state} - \Delta_{DD,gender}$
- Literally difference, the difference in differences estimators.
- As a regression: interpret the triple-interaction term (make sure to control for ALL double interactions).

What does the panel regression give?

In practice we often have **multiple treatments**

$$Y_{it} = \alpha_i + \gamma_t + \beta X_{it} + \delta_i T_{it} + u_{it}$$

$$Y_{it} = \alpha_{i0} + \alpha_i \times t + \gamma_t + \beta X_{it} + \delta_i T_{it} + u_{it}$$

- Allow time dummies and/or group specific trends (w some collinearity)
 - States roll out tax/policy changes at different times.
 - Natural disasters hit different places at different times.
 - Judges hand down rulings in different jurisdictions.
- Previously treated observations can act as **controls** through γ_t .
- Recent work by Goodman-Bacon (2019) says that $\hat{\delta}$ gives a variance weighted average of all possible 2×2 diff-in-diff designs.

Panel Regression DiD (Conlon and Rao 2014/2019)

In some of my recent work with Nirupama Rao

- Looked at **price posting laws** which required that alcohol wholesalers post prices 30 days in advance of selling, and prohibited short term price cuts.
- Look at apparent consumption to NIAAA (ethanol gallons per capita)
- States got rid of laws at different times (often because they lost court cases).
- All specifications have state and time FE
 - Lots of variation across states in consumption of beer vs. wine vs. spirits
 - Lots of variation across time in consumption of beer vs. wine vs. spirits.
 - Trends in consumption also vary across states.
- Which states do we learn from?
 - States without policy changes mostly inform βX_{it} and γ_t

Variation in Timing (Conlon and Rao 2014/2019)

Table 2: States with Post and Hold Laws

	Wine	Beer	Spirits
Connecticut	Y	Y	Y
Delaware	End 1999	End 1999	End 1999
Georgia	N	Y	Y
Idaho	Y	Y	N
Maine	Y	Y	N
Maryland	End 2004	End 2004	End 2004
Massachusetts	End 1998	End 1998	End 1998
Michigan	Y	Y	Y
Missouri	Y	N	Y
Nebraska	End 1984	N	End 1984
New Jersey	Y	Y	Y
New York	Y	Y	Y
Oklahoma	End 1990	End 1990	Y
Pennsylvania	N	End 1990	N
South Dakota	Y	N	Y
Tennessee	N	Y	N
Washington	End 2008	End 2008	N
West Virginia	N	N	Y

Panel Regression DiD (Conlon and Rao 2014/2019)

Table 3: Post and Hold Laws and State Alcohol Consumption

	(All)	(All)	(All)	(PH only)	(PH NE)
Wine					
PH	-0.0545*** (0.0183)	-0.0623*** (0.0183)	-0.0229 (0.0192)	-0.0345* (0.0190)	-0.00430 (0.0340)
R^2	0.965	0.966	0.984	0.986	0.984
Beer					
PH	-0.0155 (0.0113)	-0.0283*** (0.0107)	-0.0242** (0.0095)	-0.0201** (0.0081)	-0.0276** (0.0129)
R^2	0.891	0.905	0.969	0.960	0.991
Spirits					
PH	-0.00702 (0.0175)	-0.0423** (0.0168)	-0.0787*** (0.0180)	-0.0854*** (0.0187)	-0.0979*** (0.0278)
R^2	0.950	0.955	0.982	0.976	0.986
Year FE	Y	Y	Y	Y	Y
State FE	Y	Y	Y	Y	Y
Demog. Controls	N	Y	Y	Y	Y
State Time Trends	N	N	Y	Y	Y
PH States	N	N	N	Y	N
NE States	N	N	N	N	Y
Observations	1,428	1,428	1,428	513	243

Note: The table presents coefficients from regression (4). The outcome of interest is the log of apparent consumption per capita, where consumption is in ethanol equivalent gallons and the relevant population is state residents age 14 and older. Column 1 only includes state and time fixed effects. Column 2 adds state demographic controls and column 3 adds state-specific time trends. Column 4 limits the sample to states that have had PH laws. Column 5 restricts the sample further to only northeastern states that once had PH

Difference in Differences with Repeated Cross Sections

- Suppose we do not have available panel data but just a random sample from the relevant population in a pre-treatment and a post-treatment period. We can still use difference in differences.
- First consider a simple case where $E[Y_{i2}(0) - Y_{i1}(0)|T] = E[Y_{i2}(0) - Y_{i1}(0)]$.
- We need to modify slightly the assumption to

$$\begin{aligned} E[Y_{i2}(0)|\text{Group receiving training}] - E[Y_{i1}(0)|\text{Group receiving training in the next period}] \\ = E[Y_{i2}(0) - Y_{i1}(0)] \end{aligned}$$

which requires, in addition to the original independence assumption that conditioned on particular individuals that population we will be sampling from does not change composition.

- We can then obtain immediately an estimator for ATT as

$$\begin{aligned} E[\beta_i|T_{i2} = 1] = E[Y_{i2}|\text{Group receiving training}] - E[Y_{i1}|\text{Group receiving training next period}] \\ - \{E[Y_{i2}|\text{Non-trainees}] - E[Y_{i1}|\text{Group not receiving training next period}]\} \end{aligned}$$

Difference in Differences with Repeated Cross Sections

- More generally we need an assumption of conditional independence of the form

$$\begin{aligned} E[Y_{i2}(0)|X, \text{Group receiving training}] - E[Y_{i1}(0)|X, \text{Group receiving training next period}] \\ = E[Y_{i2}(0)|X] - E[Y_{i1}(0)|X] \end{aligned}$$

- Under this assumption (and some auxiliary parametric assumptions) we can obtain an estimate of the effect of treatment on the treated by the regression

$$Y_{it} = \alpha_g + \gamma_t + \gamma T_{it} + \beta' X_{it} + u_{it}$$

Difference in Differences with Repeated Cross Sections

- More generally we can first run the regression

$$Y_{it} = \alpha_g + \gamma_t + \delta(X_{it})T_{it} + \beta'X_{it} + u_{it}$$

where α_g is a dummy for the treatment of comparison group, and $\delta(X_{it})$ can be parameterized as $\delta(X_{it}) = \delta'X_{it}$. The ATT can then be estimated as the average of $\delta'X_{it}$ over the (empirical) distribution of X .

- A non parametric alternative is offered by Blundell, Dias, Meghir and van Reenen (2004).

Difference in Differences and Selection on Unobservables

- Suppose we relax the assumption of *no selection* on unobservables.
- Instead we can start by assuming that

$$E[Y_{i2}(0)|X, Z] - E[Y_{i1}(0)|X, Z] = E[Y_{i2}(0)|X] - E[Y_{i1}(0)|X]$$

where Z is an instrument which determines training eligibility say but does not determine outcomes in the non-training state. Take Z as binary (1,0).

- Non-Compliance: not all members of the eligible group ($Z = 1$) will take up training and some of those ineligible ($Z = 0$) may obtain training by other means.
- A difference in differences approach based on grouping by Z will estimate the impact of being allocated to the eligible group, but not the impact of training itself.

Difference in Differences and Selection on Unobservables

- Now suppose we still wish to estimate the impact of training on those being trained (rather than just the effect of being eligible)
- This becomes an IV problem and following up from the discussion of LATE we need stronger assumptions
 - Independence: for $Z = a$, $\{Y_{i2}(0) - Y_{i1}(0), Y_{i2}(1) - Y_{i1}(1), T(Z = a)\}$ is independent of Z .
 - Monotonicity $T_i(1) \geq T_i(0) \forall i$
- In this case LATE is defined by

$$\frac{E(\Delta Y_{it} | Z_{it} = 1) - E(\Delta Y_{it} | Z_{it} = 0)}{Pr(T_{it} = 1 | Z_{it} = 1) - Pr(T_{it} = 1 | Z_{it} = 0)}$$

assuming that the probability of training in the first period is zero.

Changes in Changes: Dealing w Nonlinearity

- Athey and Imbens (2006) develop a model robust to nonlinearity complaints
- Combines nonparametrics with DiD.
- Works with **quantile treatment effects** and limits selection on unobservables
- Assume that your relative location in distribution is invariant to difference.

What if we can combine the benefits of matching with DiD approaches?