Applied Microeconometrics I

Lecture 9: Differences-in-differences (continued)

Cristina Bratu

Aalto University

October 3, 2023

`Lecture Slides`

What did we do last time?

- Difference-in-differences
- Two groups: Treatment and control
- Two periods: $t = 1, 2$

$$Y_{it} = \alpha_i + \lambda_t + \rho D_{it} + \epsilon_{it}$$

  where $i = \{T, C\}$, $t = \{1, 2\}$, $D_{T,1} = D_{C,1} = D_{C,2} = 0$, and $D_{T,2} = 1$

- OLS estimation of ρ gives the treatment effect

What did we do last time?

- Difference-in-differences is a special case of estimation with panel data

$$Y_{it} = \alpha_i + \lambda_t + \rho D_{it} + X_{it}'\beta + \epsilon_{i,t}$$

  - panel data: repeated observations on units over time
  - with a large number of units, it can get computationally burdensome to add dummies for each unit, because it means estimating a coefficient on each of those dummies
  - if we are not interested in estimating those coefficients (in most cases), we can "eliminate" α_i either by converting the data into deviations from i-specific means or by differencing over t
  - this is equivalent to controlling/adjusting for α_i, but without having to estimate the FEs
    - all done "in the background" in Stata; commands such as *areg* or *reghdfe* can *absorb* a large number of fixed effects

What did we do last time?

- Example: Card and Krueger (1994) on the effect of minimum wages on employment
- Treatment group: fast-food restaurants in New Jersey (NJ raises the minimum wage in $t = 1$)
- Control group: fast-food restaurants in Pennsylvania (Pennsylvania doesn't change its minimum wage)

$$L_{it} = \alpha_i + \lambda_t + \rho D_{it} + \epsilon_{it}$$
$$\rho = [E(L_{NJ,1}) - E(L_{NJ,0})] - [E(L_{PA,1}) - E(L_{PA,0})]$$

where $E(\cdot)$ is the average L_{it} value conditional on $(\alpha_i, \lambda_t, D_{it})$

What did we do last time?

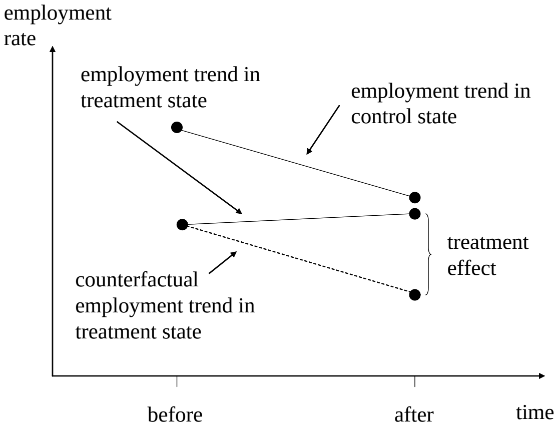


Figure 5.2.1: Causal effects in the differences-in-differences model

Table 5.2.1: Average employment per store before and after the New Jersey minimum wage increase

Variable		PA (i)	NJ (ii)	Difference, NJ-PA (iii)
1.	FTE employment before,	23.33	20.44	-2.89
	all available observations	(1.35)	(0.51)	(1.44)
2.	FTE employment after,	21.17	21.03	-0.14
	all available observations	(0.94)	(0.52)	(1.07)
3.	Change in mean FTE	-2.16	0.59	2.76
	employment	(1.25)	(0.54)	(1.36)

Notes: Adapted from Card and Krueger (1994), Table 3. The table reports average full-time equivalent (FTE) employment at restaurants in Pennsylvania and New Jersey before and after a minimum wage increase in New Jersey. The sample consists of all stores with data on employment. Employment at six closed stores is set to zero. Employment at four temporarily closed stores is treated as missing. Standard errors are reported in parentheses

Simple Regression Diff-in-Diff

Another way to write the same difference-in-differences equation:

$$Y_{ist} = \alpha + \gamma TREAT_s + \lambda AFTER_t + \rho(AFTER_t * TREAT_s) + \varepsilon_{ist}$$

- Y_{ist} is the number of full-time employees working in establishment i, located in state $s \in$ {NJ,PA}, in period $t \in$ {Feb 1992, Nov 1992}
- $TREAT_s$: dummy variable equal to 1 when s={NJ}
- $AFTER_t$: dummy variable equal to 1 when t={Nov 1992}
- $TREAT_s * AFTER_t$ interaction term that takes value one when s={NJ} & t={Nov 1992}

Simple Regression DD: interpreting coefficients

$$Y_{ist} = \alpha + \gamma TREAT_s + \lambda AFTER_t + \rho(AFTER_t * TREAT_s) + \varepsilon_{ist}$$

- α: average Y in non-treated group (PA) in the pre-treatment period
- γ: difference in Y between treatment group (NJ) and control group (PA) in the pre-treatment period
- λ: ΔY in the control group between the pre-treatment and the treatment period
- ρ: ΔY in the treatment group between the pre-treatment and the treatment period, relative to the ΔY in the control group $\rightarrow$ captures effect of the policy!

```
.
. xi: reg EMPTOT i.NEWJERSEY*i.AFTER, cluster(ID)
i.NEWJERSEY       _INEWJERSEY_0-1     (naturally coded; _INEWJERSEY_0 omitted)
i.AFTER           _IAFTER_0-1         (naturally coded; _IAFTER_0 omitted)
i.NEW~Y*i.AFTER   _INEWXAFT_#_#       (coded as above)

Linear regression                                      Number of obs =     794
                                                       F(  3,   409) =    1.80
                                                       Prob > F      =  0.1462
                                                       R-squared     =  0.0074
                                                       Root MSE      =  9.4056

                                   (Std. Err. adjusted for 410 clusters in ID)
```

EMPTOT	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
_INEWJERSEY_1	-2.891761	1.439546	-2.01	0.045	-5.721593	-.0619281
_IAFTER_1	-2.165584	1.218025	-1.78	0.076	-4.559954	.2287855
_INEWXAFT_1_1	2.753606	1.306607	2.11	0.036	.1851025	5.322109
_cons	23.33117	1.346536	17.33	0.000	20.68417	25.97816

Regression DD: Including controls

- Recall that the time-invariant factors at the state level are taken care of with the state fixed effects.
- Sometimes you may obtain identification only conditional on adding certain time-varying covariates, X_{st}, which vary at the state-year level.
  - for example, if something else happens in the two periods which also affects employment and has nothing to do with the minimum wage, you should control for it so as to be able to disentangle the actual effect of the minimum wage
  - identification becomes conditional on covariates
- You can also include controls that vary at the unit-level, X_{ist}, but these don't matter for identification; they may help to obtain more precise estimates.
- Whatever you do, do not include *bad controls*, i.e. covariates that may themselves be outcome variables of the treatment.

Results of the CK Study

- Wages increased by 10% in NJ, remained constant PA
- ... but employment rose in NJ and decreased in PA
- The diff-in-diff estimate suggests that **the rise in the minimum wage *increased* employment**
- Result robust to alternative specifications and to an alternative control group (workers with salaries above the minimum salary)

Reactions to the CK Study

- Angus Deaton: "*The reception accorded to Princeton faculty by their colleagues in other institutions is what might be expected by the friends and defenders of child-molesters*"
- James Buchanan in the Wall Street Journal:
  "*no self-respecting economist would claim that increases in the minimum wage increase employment. Such a claim, if seriously advanced, becomes equivalent to a denial that there is even minimum scientific content in economics, and that, in consequence, economists can do nothing but write as advocates for ideological interests. Fortunately, only a handful of economists are willing to throw over the teaching of two centuries; we have not yet become a bevy of camp-following whores*"

See Angus Deaton's "Letters from America" for more:
www.princeton.edu/~deaton/downloads/letterfromamerica_oct1996.html

Reactions to the CK Study

- Neumark and Wascher (2000, AER)
  - CK data has a lot of measurement error
  - data provided by Employment Policies Institute reveal that the minimum wage rise did decrease employment
- Card and Krueger (2000, AER)
  - administrative data from Bureau of Labor Statistics confirm the key findings of the 1994 paper
  - "*calls into question the representativeness of the sample assembled by Berman, Neumark and Wascher*"

See John Schimtt's "Cooked to Order" for more:
www.prospect.org/cs/articles?article=cooked_to_order

Employment in New Jersey and Pennsylvania fast-food restaurants, Oct. 1991 - Sept. 1997 (Card and Krueger, 2000)

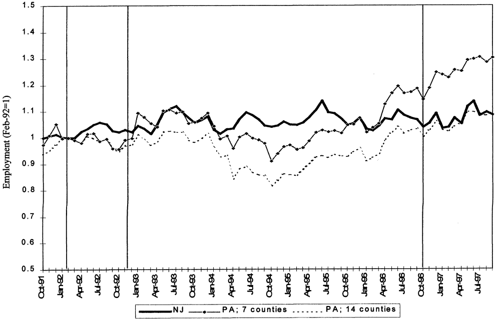


FIGURE 2. EMPLOYMENT IN NEW JERSEY AND PENNSYLVANIA FAST-FOOD RESTAURANTS, OCTOBER 1991 TO SEPTEMBER 1997

Note: Vertical lines indicate dates of original Card-Krueger survey and the October 1996 federal minimum-wage increase.
Source: Authors' calculations based on BLS ES-202 data.

- Some potential methodological concerns:
  1. The authors do not examine how the trends evolved in the past. Information from future trends suggests that they are not parallel.
  2. At the end of the day, we only have two observations. Possible common shocks may affect the treatment or the control group.
  3. Other policies?
  4. Note also the tension between having observations that are geographically close and the potential existence of an impact of the treatment on the control group.

Probing DD assumptions

- The crucial assumption in DD set up is that the control group provides information about how the treatment group would have evolved in the absence of treatment (parallel trends).
- However, it is a fundamentally untestable assumption, because it is based on a counterfactual we cannot observe.
- The best thing we can do is to come up with suggestive evidence that makes the parallel trends assumption more plausible.
- With more than two periods this can be investigated in several ways...
    1. Illustrate graphically that the average outcomes evolved similarly in the years before the policy was implemented
    2. Run placebo tests: does the placebo policy introduced in t-1, t-2, etc. have any significant impact?
    3. Include group-specific trends

Dynamic treatment effects

- The traditional 2×2 design collapses all the periods before treatment into one "pre" period, and all the periods after treatment into one "post" period.
- But often effects may take a while to show, or certain treatments can become more or less effective over time.
- A dynamic DiD allows for the effects to differ in each period.

$$Y_{ist} = \alpha_s + \lambda_t + \sum_{j=-m}^{q} \rho_j D_{s,t+j} + \varepsilon_{ist}$$

- where D_{st} is an indicator for whether the treatment got switched on in year t.
- Note we now have a "time to treatment" variable, which is equal to the original time variable minus the treatment period.
- Note: you need a reference time period (usually $t = -1$).

Multiple treatment groups, multiple treatment periods

- DiD setup with multiple groups in the treated category (the traditional setup allows for this as well).
- In addition, the groups get treated at different times (e.g. different states introduce minimum wages in different years).
- Depending on the setup, you may have a pure control group, e.g. states that never introduce minimum wages; or all the states in your data *eventually* get treated.
- MANY papers do this.
- Recent advances in econometric theory (e.g. Sun and Abraham, 2020; Borusyak and Jaravel, 2017; Callaway and Sant'Anna, 2020; Goodman-Bacon, 2019; Imai and Kim, 2020; Strezhnev, 2018; Athey and Imbens, 2018; de Chaisemartin and D'Haultfœuille, 2020) suggest that this kind of DiD setup often do not provide valid estimates of the causal effects of interest.

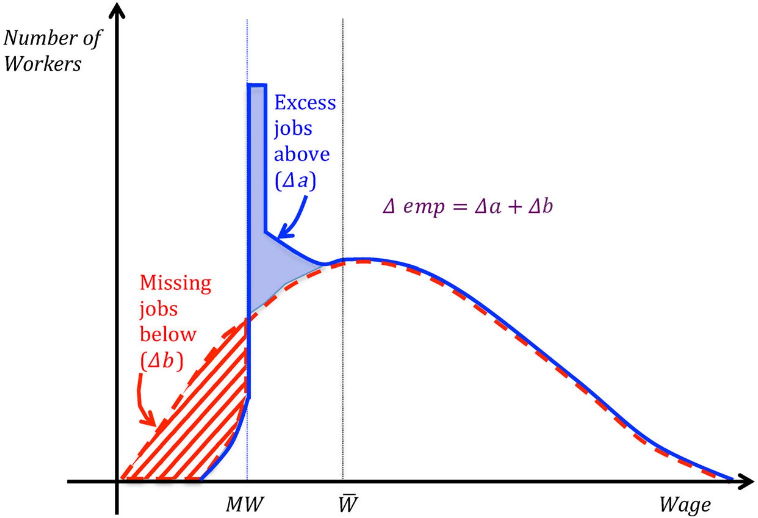
Example: Cengiz et al (2019): The effect of minimum wage on low-wage jobs

- Cengiz et al (2019) is a recent addition to the literature on the effects of minimum wages
- Like Card and Krueger (1994), it is a diff-in-diff paper but modern in its approach
- Exploit 138 state-level changes in minimum wages between 1979 and 2016 to identify the effect
- Instead of a particular sector (fast-food restaurants), focus on the effect on the total employment

Example: Cengiz et al (2019): The effect of minimum wage on low-wage jobs

- Diff-in-diff design to estimate the impact of minimum wage increases on the entire distribution of wages
- Zoom on the bottom of the distribution to estimate the impact on employment and wages of affected workers
- Idea:
  - An increase in minimum wages will directly affect jobs that were paying less than the minimum wage before the raise
  - Some of these jobs will disappear as a result of the raise
  - Some of these jobs will increase wages and show up as "excess jobs" at and above the minimum wages
  - At the upper tail of the wage distribution we shouldn't see any effects of minimum wages

The impact of minimum wages on the frequency distribution of wages



Example: Cengiz et al (2019): The effect of minimum wage on low-wage jobs

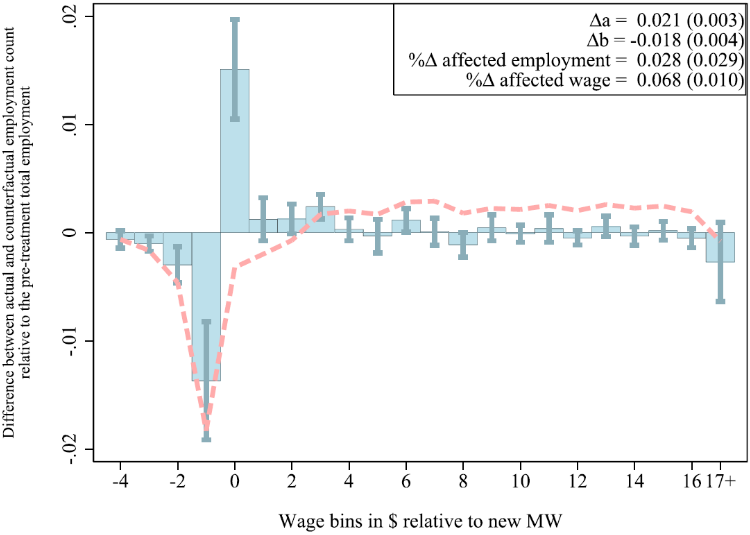
- Estimate employment changes in bins of the wage distribution relative to the minimum wage for three years prior to and five years following an event

$$\frac{E_{sjt}}{N_{st}} = \sum_{\tau=-3}^{4} \sum_{k=-4}^{17} \alpha_{\tau k} I_{sjt}^{\tau k} + \mu_{sj} + \rho_{jt} + \Omega_{sjt} + u_{sjt}$$

  where E_{sjt} is the employment in wage bin j in state s at a quarter t, N_{st} is the population in s at t

- The treatment dummy $I_{sjt}^{\tau k}$ is equal to one if the minimum wage was increased τ years from date t for bin j that falls between k and $k+1$ dollars of the new minimum wages
- State-bin effects μ_{sj}, period-bin effects ρ_{jt}

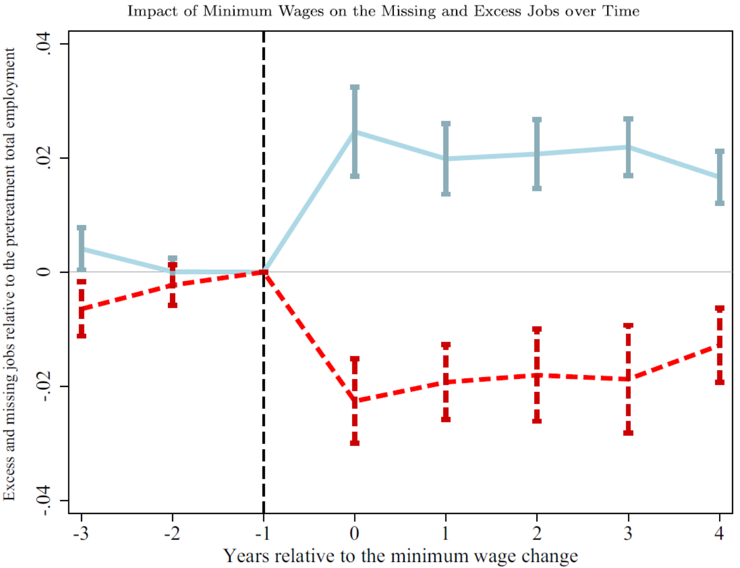
The impact of minimum wages on the wage distribution



Example: Cengiz et al (2019): The effect of minimum wage on low-wage jobs

- The authors have data on states several years before and after the increase in minimum wage change
- This allows the to estimate the "lead"and "lag" effects to assess pre-existing trends
- If employment below and above the new minimum wage diverge already before the raise, parallel trends assumption could be violated
- No evidence of diverging trends before the raise

The impact of minimum wages on the wage distribution



- So far we have used difference-in-differences to identify causal effects using data on units over time.
- However, instead of states, we may have demographic groups, some of which are affected by a policy and others are not.
- Similarly, instead of time, we could group data by cohort or other types of characteristics.
- We can also exploit this kind of variation in a difference-in-differences style strategy.

Example: Boeri and Jimeno (2005): The effects of employment protection

- Boeri and Juan Jimeno (2005) studied the effect of employment protection on the probability of being dismissed in Italy.
- Under Italian labour law, in the 1990's the degree of employment protection of workers on permanent contracts depended on firm size:
  - in firms with more than 15 employees the workers on permanent contract were covered by the most restrictive employment protection legislation in Europe
  - in firms with fewer than 15 employees the workers could be fired much more easily
- For workers on temporary contracts, however, the strictness of employment protection did not depend on firm size at all.

Example: Boeri and Jimeno (2005): The effects of employment protection

- Treatment: EPL coverage
- "Time" variable: firm size
- "State" variable: workers on permanent/temporary contracts

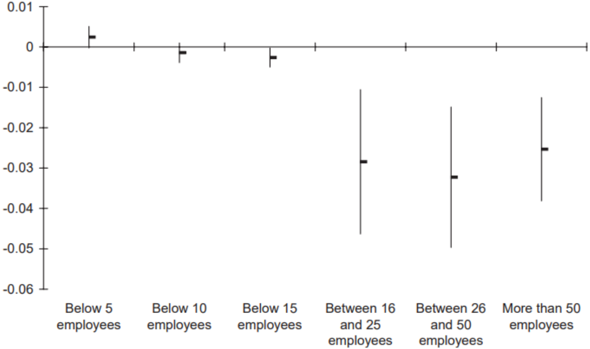


Fig. 7. Effect of EPL coverage on the dismissal rate.
Note: The figure plots the point estimates of the marginal effect of permanent contract on the probability of dismissal and the corresponding 95% confidence band.

A note on standard errors

- In DiD applications, we work with panel data, i.e. with data where observations on the unit are repeated over time.
- Such data is generally *serially correlated*, i.e. the values of variables for nearby periods are likely to be similar.
- Standard errors are also likely to be serially correlated.
- If we ignore serial correlation, we run the risk of exaggerating the precision of our estimates.
  - OLS standard errors assume that all observations are independent realizations (the data come from random samples).
- Solution: **clustered standard errors**.
- By clustering, we assume that clusters are randomly sampled, without requiring units within clusters to be randomly sampled.
- Rule of thumb: cluster at the level at which your treatment is assigned. The number of clusters should be sufficiently high.

Final comments on diff-and-diff:

- Identification again relies on assumptions that cannot be tested.
- But at least we can show that in the past trends were parallel.
- Discuss explicitly why it is a good assumption to believe that the timing of the treatment/policy was as good as random.
- Discuss explicitly the existence of alternative policies that might contemporaneously affect the treatment or the control group.
- Discuss the possibility that the control group is affected by the treatment.

Example

State	Wage group	t=1	t=2
	$wg_{1,low}$	$empl_{1,low,1}$	$empl_{1,low,2}$
s_1	$wg_{1,medium}$	$empl_{1,medium,1}$	$empl_{1,medium,2}$
	$wg_{1,high}$	$empl_{1,high,1}$	$empl_{1,high,2}$
	$wg_{2,low}$	$empl_{2,low,1}$	$empl_{2,low,2}$
s_2	$wg_{2,medium}$	$empl_{2,medium,1}$	$empl_{2,medium,2}$
	$wg_{2,high}$	$empl_{2,high,1}$	$empl_{2,high,2}$

Minimum wage increases in s=1 at t=2

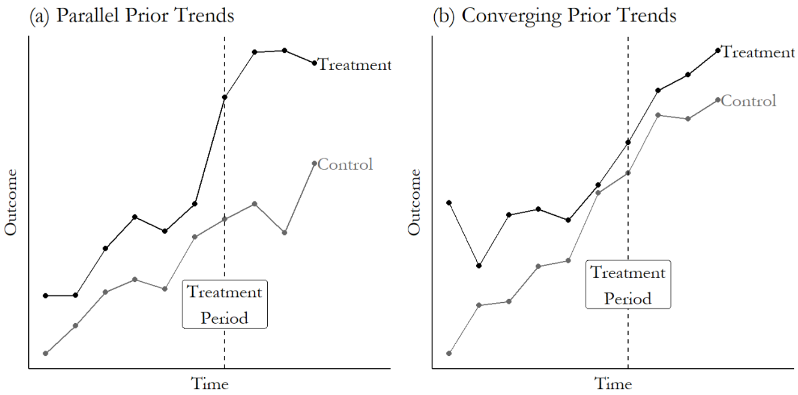
Increases wages in $wg_{1,low}$ at t=2,

Employment may decrease in $wg_{1,low}$ and increase in $wg_{1,medium}$

No effect: $(empl_{1,high,2}-empl_{1,high,1})-(empl_{2,high,2}-empl_{2,high,1})$

Net effect: $[(empl_{1,medium,2}-empl_{1,medium,1})-(empl_{2,medium,2}-empl_{2,medium,1})]-[(empl_{1,low,2}-empl_{1,low,1})-(empl_{2,low,2}-empl_{2,low,1})]$

Parallel trends assumption



Placebo tests

Let's go back to the Card and Krueger (1994) example, with the treatment implemented in New Jersey in April 1992. With additional data before the implementation, we could:

- use the data only from before April 1992
- choose a few different periods and pretend that the treatment was applied at that time
- estimate the DiD using those fake treatment dates
- if we estimate a non-zero DiD effect at those fake treatment dates, we should be worried about the parallel trends assumption
- if differences between the treated and the control group do not exactly cancel out at the fake treatment dates, hard to believe they would cancel out at the real treatment time

Group-specific trends

The effect of the minimum legal drinking age (MLDA) on mortality rates; treated = Allatsea; control = Alabaster

FIGURE 5.5
A spurious MLDA effect in states where trends are not parallel

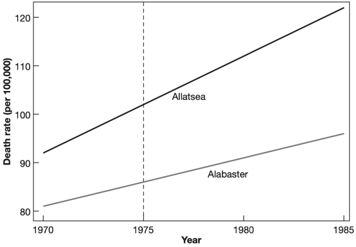
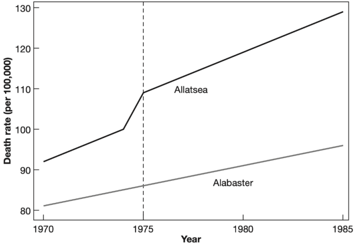


FIGURE 5.6
A real MLDA effect, visible even though trends are not parallel



- left graph: the DD estimate will be confounded by state-level trends that diverge already before the policy
- right graph: control for trends, add state dummies interacted with a linear time variable
- if the deviation from trend induced by the causal effect is sharp, this strategy can work
- controlling for trends may result in controlling away some of the treatment effect, especially if effects get stronger over time