From Diff in Diff to Panel Data

Econ 140, Section 11

Jonathan Old

Roadmap

1. Recap: DiD
2. Making DiD more general: Panel Data estimation
3. Practice question

Recap: DiD

How to get the causal effect of a treatment: DiD

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

We can do several comparisons:

- Comparison 3: Compare treated to untreated group, before and after the intervention. **Differences in differences!**

How to get the causal effect of a treatment: DiD

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

We can do several comparisons:

- Comparison 3: Compare treated to untreated group, before and after the intervention. **Differences in differences!**
- Estimated treatment effect?
  $(6-6)-(5-4)=(6-5)-(6-4)=-1$

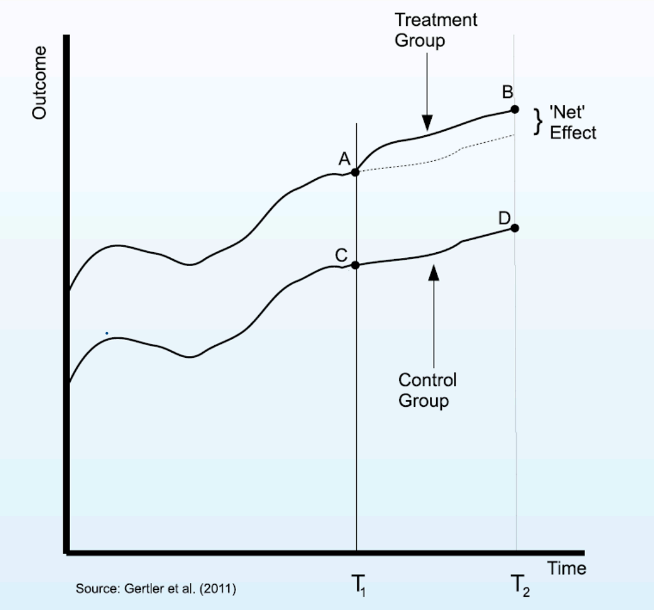
How to get the causal effect of a treatment: DiD

	2020	2022
Free Mental Health: Treated	6	6
No Free Mental Health: Untreated	4	5

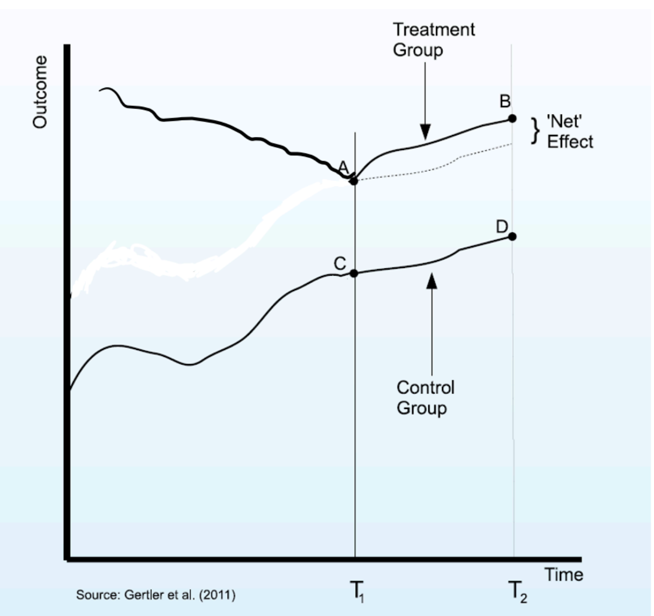
We can do several comparisons:

- Comparison 3: Compare treated to untreated group, before and after the intervention. **Differences in differences!**
- Estimated treatment effect?
  $(6-6)-(5-4)=(6-5)-(6-4)=-1$
- Identifying assumption: Parallel trends: Without the treatment, the average increase in the outcome of the treated would have been the same as the average increase in the outcome of the untreated.

Parallel trends assumption



(Potential) violation of a parallel trends assumption



Estimating DiD with regressions

We can set up a simple linear regression to estimate a DiD model:

$$Y_{it} = \alpha + \beta \text{ Treated }_i + \gamma \text{ Post }_t + \delta \text{ Treated }_i \cdot \text{ Post }_t + u_{it}$$

	2020	2022
Free Mental Health: Treated	$\alpha + \beta$	$\alpha + \beta + \gamma + \delta$
No Free Mental Health: Untreated	α	$\alpha + \gamma$

Any questions?

… Remember – Every question is useful!

Making DiD more general: Panel Data estimation

Limitations of the 2x2 DiD framework

In the simple 2x2 framework, we estimated:

$$Y_{it} = \alpha + \beta \text{ Treated }_i + \gamma \text{ Post }_t + \delta \text{ Treated }_i \cdot \text{ Post }_t + u_{it}$$

How can we extend this framework?

A panel dataset may look like this:

	country	iso2c	iso3c	year	oda_per_cap	perc_below_pov_line	gdp_per_cap	inflation
316	Chile	CL	CHL	2015	3.0402700	0.4	13569.9478	4.34877353
317	Chile	CL	CHL	2016	9.8673522	NA	13785.6883	3.78619356
318	Chile	CL	CHL	2017	3.8277324	0.3	15045.5277	2.18271847
319	Chile	CL	CHL	2018	NA	NA	15795.7085	2.43488981
320	Chile	CL	CHL	2019	NA	NA	14631.9469	2.55754476
321	Chile	CL	CHL	2020	NA	0.7	13094.4595	3.04549085
322	Chile	CL	CHL	2021	NA	NA	16265.0960	4.52456838
323	China	CN	CHN	2015	−0.2219718	1.2	8016.4314	1.43702381
324	China	CN	CHN	2016	−0.5702808	0.8	8094.3634	2.00000182
325	China	CN	CHN	2017	−0.7090169	0.7	8816.9869	1.59313600
326	China	CN	CHN	2018	−0.5027731	0.4	9905.3420	2.07479040
327	China	CN	CHN	2019	−0.4323084	0.1	10143.8382	2.89923416
328	China	CN	CHN	2020	−0.4059315	NA	10408.6698	2.41942189
329	China	CN	CHN	2021	−0.4002945	NA	12556.3331	0.98101514
330	Colombia	CO	COL	2015	28.7760584	4.9	6228.4263	4.98983116
331	Colombia	CO	COL	2016	23.2331299	4.9	5938.4639	7.51346025
332	Colombia	CO	COL	2017	17.5871894	4.3	6450.3196	4.31431326
333	Colombia	CO	COL	2018	36.1416353	4.5	6782.0379	3.24056933
334	Colombia	CO	COL	2019	17.4085507	5.3	6438.0602	3.52301933

It may also look like this:

	Crime_2006	Crime_2007	Crime_2008	Crime_2009	Crime_2010	Crime_2011	Crime_2012	Crime_2013	Crime_2014	Crime_2015	Crime_2016	Crime_2017
New York	0,74075276	0,80312483	0,57854938	0,58480987	0,20097276	0,80826596	0,97070048	0,02437202	0,07831943	0,84572914	0,75749919	0,98948363
Los Angeles	0,81840795	0,63440595	0,75200179	0,45218138	0,54532883	0,0015679	0,39512876	0,41847691	0,66467241	0,01946605	0,55885996	0,27521867
Chicago	0,81682982	0,00089416	0,74575534	0,5521186	0,60461758	0,34694645	0,47874246	0,15925374	0,34667588	0,83944993	0,44825934	0,58746291
Houston	0,05885616	0,79349661	0,18033028	0,51317119	0,79642654	0,50269029	0,56157808	0,77173283	0,60797558	0,40461519	0,785332	0,6477959
Phoenix	0,2568483	0,28097691	0,82722175	0,11994755	0,884851	0,83800072	0,08978678	0,21613254	0,97306225	0,57291156	0,27341507	0,20641572
Philadelphia	0,86364857	0,30478317	0,07818559	0,82930161	0,28324875	0,60204952	0,4467311	0,1165021	0,15597767	0,36430802	0,38034388	0,37091781
San Antonio	0,98256294	0,13215208	0,55220195	0,25274349	0,22117177	0,63049228	0,74109619	0,17937274	0,50914785	0,68773119	0,17854372	0,50634586
San Diego	0,26052609	0,90551179	0,96145068	0,04932108	0,55859179	0,57753748	0,38254621	0,36728436	0,39429268	0,40312699	0,55506545	0,51665312
Dallas	0,85045327	0,33482019	0,15361924	0,74231968	0,55762722	0,05057305	0,57960707	0,71774087	0,63565644	0,04305477	0,06451316	0,27719973
San Jose	0,75452737	0,21462214	0,24820183	0,2727016	0,00127307	0,36822105	0,32304996	0,9116018	0,86006177	0,05904204	0,1124391	0,83247218
Austin	0,45398495	0,27082977	0,89920786	0,82480029	0,52877813	0,9452614	0,82578514	0,828743	0,20606871	0,4439492	0,38857348	0,8154967
Jacksonville	0,18638302	0,66744344	0,38392491	0,83800597	0,30374371	0,16785163	0,53614039	0,15558203	0,14260844	0,51748284	0,07054862	0,70962828
Fort Worth	0,13254183	0,78355507	0,83321159	0,33460775	0,00772913	0,78127289	0,61026409	0,25552213	0,27373821	0,78588972	0,80784366	0,1307398
Columbus	0,21457262	0,59542339	0,65347292	0,16532587	0,35580394	0,7113844	0,72036005	0,39323372	0,73022336	0,32649945	0,95442814	0,44923343
Indianapolis	0,97993863	0,2017115	0,58883837	0,54914541	0,22495244	0,58047281	0,01984661	0,88891756	0,34920826	0,326581	0,11407985	0,76592867
Charlotte	0,13043704	0,44544929	0,13581143	0,32176217	0,06451777	0,60708331	0,24805805	0,91489751	0,39053247	0,36409354	0,94249222	0,12812588

Interpreting FE regressions

Suppose we collect yearly data on firms' profits and the gender of the firms' CEO, for a sample of 1,000 firms over 20 years.

We can estimate this model with firm and year fixed effects:

$$\text{Profits}_{it} = \beta_0 + \beta_1 \text{Female CEO}_{it} + \alpha_i + \delta_t + \varepsilon_{it}$$

How do we interpret β_1?

Interpreting FE regressions

Suppose we collect yearly data on firms' profits and the gender of the firms' CEO, for a sample of 1,000 firms over 20 years.

We can estimate this model with firm and year fixed effects:

$$\text{Profits}_{it} = \beta_0 + \beta_1 \text{Female CEO}_{it} + \alpha_i + \delta_t + \varepsilon_{it}$$

How do we interpret β_1?

The same way as in an OLS regression!

β_1 describes how having a female CEO is associated with profits, holding constant all firm-specific factors that do not change over time, and all time-specific factors that affect all firms equally. β_1 is the effect within each firm (and year!), controlling for the firm-specific and year-specific fixed effects.

Visualization

See separate gif!

Limitations of fixed effects

- Imagine we want to estimate the effect of annual income on happiness and we have a panel following 1,000 individuals over 10 years. You think there are many omitted variables and include individual-level fixed effects. You are **also** interested in the effect of parental income on happiness. What's the problem?

Limitations of fixed effects

- Imagine we want to estimate the effect of annual income on happiness and we have a panel following 1,000 individuals over 10 years. You think there are many omitted variables and include individual-level fixed effects. You are **also** interested in the effect of parental income on happiness. What's the problem?
- We want to estimate the effect of democracy on economic growth in a panel of 180 countries over 60 years. We can include country-fixed effects to account for any effect coming from different countries having a different climate, terrain, access to the sea, culture, etc. Are you satisfied with this approach?

Limitations of fixed effects

- Imagine we want to estimate the effect of annual income on happiness and we have a panel following 1,000 individuals over 10 years. You think there are many omitted variables and include individual-level fixed effects. You are **also** interested in the effect of parental income on happiness. What's the problem?
- We want to estimate the effect of democracy on economic growth in a panel of 180 countries over 60 years. We can include country-fixed effects to account for any effect coming from different countries having a different climate, terrain, access to the sea, culture, etc. Are you satisfied with this approach?
- We want to estimate the effect of education on wages and think of using individual-level fixed effects. What can go wrong?

Any questions?

… Remember – Every question is useful!

Practice question

Did Welfare Reform Increase Employment? In the 1990s, in an effort known as 'Welfare Reform' the United States implemented a series of policies that aimed to reduce welfare dependency by incentivizing work and raising the hassle of being on welfare. These reforms were implemented at the state level between 1993 and 1997. In this exercise, you will investigate whether welfare reform successfully raised employment rates.

The file welfare.csv contains information on individuals in the U.S. between 1990 and 2005. The key variables are

- employed: dummy equal to one if individual is employed
- reform: dummy equal to one if the individual is living in a state that has welfare reform in place in that year
- hsgrad: dummy equal to one if individual has a HS degree
- white: dummy equal to one if individual is white
- woman: dummy equal to one if individual identifies as a woman
- avg_benefit: the average welfare benefit (between 1990 and 2005) for a family of three in that individual's state

Estimate a regression of (1) employed against reform (2) employed against reform, hsgrad, white, and woman and interpret the coefficient on reform in both regressions.

Suggest a variable that varies across states but plausibly varies little-or not at all-over time and that could cause omitted variable bias in regression (2).

Estimate regression (2) using state fixed effects. How does the coefficient change? What does this suggest about the ommitted variable bias in part (b)? Which omitted variables do we account for with state FE?

Estimate regression (2) using time fixed effects. Interpret the interpret the coefficient on reform. How does the coefficient change? What does this suggest about the ommitted variable bias in part (b)? Which omitted variables do we account for with time FE?

Estimate regression (2) using time AND state fixed effects. Interpret the interpret the coefficient on reform. How does the coefficient change? What does this suggest about the ommitted variable bias in part (b)? Which omitted variables do we account for with state and time FE?

Which specification (no FE, state FE, time FE, or state + time FE) is most credible and why?

See you on R!