ECON-C4200 - Econometrics II: Capstone

Lecture 3: Difference in difference regression

Otto Toivanen

Learning outcomes

- At the end of lectures 3 & 4, you

1. understand what a difference-in-difference (DiD) estimator is
2. how it identifies the effect of the treatment
3. what the identifying assumptions are
4. what the economic content of the identifying assumptions are
5. how to implement DiD
6. what an **A**verage **T**reatment **E**ffect (ATE) is.

Difference estimator

- Let's assume a discrete treatment (=all those who get the treatment get the same treatment):

  $T_i = 1$ if individual i gets the treatment.

  $T_i = 0$ if individual i does not get the treatment.

- Notice how - unlike in perhaps medicine - giving the same treatment in a social science context can be difficult to ensure (or even to define).

Randomized Control Trial (RCT)

- In economics:
  1. The researcher decides on what the experiment is.
  2. The researcher decides what the population of interest is.
  3. The researcher draws a random sample.
  4. Individuals in the random sample are randomly allocated into control and treatment groups.

Difference estimator

$$Y_i = \beta_0 + \beta_1 T_i + \epsilon_i$$

Q1: what is the interpretation of β_1?

Q2: is there an omitted variable problem?

Q3: what if individuals truly randomized and the researcher observes other characteristics besides Y_i, T_i?

RTC estimator

Q4: is there any reason to include control variables (W_i)?

1. Efficiency - $corr(T_i, W_i) = 0$ by design.
2. Control for randomization: if β_1 without controls $\neq \beta_1$ with controls ($=$ in a statistically (& economically) significant way), then RCT has failed.

RCT and the Difference estimator

- Treatment: $\mathbb{E}[Y|T=1] = \beta_0 + \beta_1 \times 1$
- Control: $\mathbb{E}[Y|T=0] = \beta_0 + \beta_1 \times 0$
- Difference: $\mathbb{E}[\Delta Y] = \beta_1$
- This is why a t-test on the difference in Y between treatment and control groups often sufficient.

So what is DiD?

Imagine the researcher has 2 consecutive observations / individual:

1. Before experiment (period $t = 1$).
2. After experiment (period $t = 2$).

So what is DiD?

Two possibilities:

1. Everybody gets the treatment $\rightarrow$ **event study**.
2. Some get the treatment (**treatment group**), some don't (**control group**) $\rightarrow$ **Difference-in-difference** (DiD) setup.

- Notice: the use of the term "event study" fluctuates somewhat.
- We concentrate on DiD.

Examples of DiD setups

Bloom, N., Liang, J., Roberts, J. & Ying, Z. J. (2015). Does working from home work? Evidence from a Chinese experiment. *The Quarterly Journal of Economics*, *130*(1), 165–218

1. Effect of work from home (WFH) on productivity.
2. Difference in change of performance between those workers that shift to WFH and those who stay in office.

Examples of DiD setups

Gil, R. (2015). Does vertical integration decrease prices? Evidence from the paramount antitrust case of 1948. *American Economic Journal: Economic Policy*, *7*(2), 162–91

1. Effect of vertical integration (VI) on downstream prices (=theatre tickets).
2. Difference in change of ticket prices between VI and non-VI theatres due to removing VI.

Examples of DiD setups

Aghion, P., Akcigit, U., Hyytinen, A. & Toivanen, O. (2022). A year older, a year wiser (and farther from frontier): Invention rents and human capital depreciation. *Review of Economics and Statistics, forthcoming*

1. Effect of invention on wages of (co-)workers.
2. Difference in change of wages before and after invention for individuals in inventing and non-inventing firms.
3. Differential effects by co-worker type.

Regression framework

Write generally

$$Y_{it} = \alpha_0 + \beta_{group} T_i + \beta_{period} D_{after,t} + \beta_1 T_i D_{after,t} + \epsilon_{it}$$

- $T_i = 1$ if in treatment group, 0 if in control group. Notice missing t-index.
- $T_i D_{after,t} = T_i \times D_{after,t}$.
- Notice t-index. It allows us to keep track of whether an observation is **before** or **after** the introduction of the treatment. $t \in \{1, 2\}$.
- $D_{after,t} = 1$ if period after, 0 otherwise. Notice missing i-index.
- $\alpha_0 =$ common constant. Notice that with individual level data, we can replace α_0 with α_i, but then cannot identify β_{group} any more / do not need it any more.
- $\epsilon_{it} =$ error term / unobservables.

Regression framework - control group

For control group $T_i = 0$ rewrite:

$$Y_{it} = \alpha_0 + \beta_{period} D_{after,t} + \beta_1 T_i D_{after,t} + \epsilon_{it}$$

$$= \alpha_0 + \beta_{period} D_{after,t} + \epsilon_{it}$$

Regression framework - control group

$$Y_{it} = \alpha_0 + \beta_{period} D_{after,t} + \epsilon_{it}$$

1. After: $Y_{i2} = \alpha_0 + \beta_{period} + \epsilon_{i2}$
2. Before: $Y_{i1} = \alpha_0 + \epsilon_{i1}$
3. Diff: $\Delta Y_i = \beta_{period} + \epsilon_{i2} - \epsilon_{i1}$

Regression framework - treatment group

For treatment group $T_i = 1$ rewrite:

$$Y_{it} = \alpha_0 + \beta_{group} T_i + \beta_{period} D_{after,t} + \beta_1 T_i D_{after,t} + \epsilon_{it}$$

$$= \alpha_0 + \beta_{group} + \beta_{period} D_{after,t} + \beta_1 D_{after,t} + \epsilon_{it}$$

Regression framework - treatment group

$$Y_{it} = \alpha_0 + \beta_{group} + \beta_{period} D_{after,t} + \beta_1 D_{after,t} + \epsilon_{it}$$

1. After: $Y_{i2} = \alpha_0 + \beta_{group} + \beta_{period} + \beta_1 + \epsilon_{i2}$
2. Before: $Y_{i1} = \alpha_0 + \beta_{group} + \epsilon_{i1}$
3. Diff: $\Delta Y_i = \beta_{period} + \beta_1 + \epsilon_{i2} - \epsilon_{i1}$

- Notice how we could replace $\alpha_0 + \beta_{group}$ with α_i if we have individual level data. α_i, too, would vanish in the differencing over time in both the control and the treatment group.

DiD

1. Treatment group: $\mathbb{E}[\Delta Y] = \beta_{period} + \beta_1$
2. Control group: $\mathbb{E}[\Delta Y] = \beta_{period}$
3. DiD: $\mathbb{E}[\Delta\Delta Y] = \beta_1$

- β_1 is the **A**verage **T**reatment **E**ffect, (**ATE**), as it measures the average change in Y_{it} due to the treatment.
- An RCT also delivers (an estimate of) ATE.

Diff vs. DiD

- DiD needs data over at least 2 periods.
- DiD allows for individual specific constants if you have data on the same individuals before and after.
- $\rightarrow$ DiD doesn't necessitate randomization.

1. Identifying assumption #1: **common trends**: The outcome variable would have developed similarly in the treatment group as it did in the control group, had the treatment group not received the treatment.
2. Identifying assumption #2: $\mathbb{E}[\epsilon_{it}|\boldsymbol{X}_{it},\ T_i,\ D_{after,t},\ \alpha_i] = 0$.

Diff vs. DiD - Identifying assumption #1

Common trends.

1. When would this be violated?
2. Technically, $\beta_{period|control} \neq \beta_{period|treatment}$.
3. Call $\beta_{period|treatment} - \beta_{period|control} = \Delta\beta_{period}$.

Diff vs. DiD - Identifying assumption #1

1. Treatment group: $\mathbb{E}[\Delta Y] = Y_2 - Y_1 = \beta_1 + \beta_{period|treatment}$
2. Control group: $\mathbb{E}[\Delta Y] = Y_2 - Y_1 = \beta_{period|control}$
3. DiD: $\mathbb{E}[\Delta\Delta Y] = \beta_1 + \Delta\beta_{period} \neq \beta_1$ if and only if $\Delta\beta_{period} \neq 0$; in other words, if and only if $\beta_{period|treatment} \neq \beta_{period|control}$.

Diff vs. DiD - Identifying assumption #1

- Substantively?
- Example #1: Bloom et al., 2015
  - Those that know their productivity is (permanently) declining decide to work from home (or office).
- Example #2: Gil, 2015
  - Think of the effect of hiring a new CEO on firm performance. Firm observes performance is (permanently) declining compared to peers, and therefore hires a new CEO.
- Example #3: Aghion et al., 2018
  - The treatment firms are in growing markets where within-firm human capital important. The trend growth of wages therefore different from that of control group firms.

Diff vs. DiD - Identifying assumption #2

- Identifying assumption #2: $\mathbb{E}[\epsilon_{it}|\boldsymbol{X}_{it},\ T_i,\ D_{after,it},\ \alpha_i] = 0$.

$\rightarrow$ selection into treatment can depend on individual specific "things" that are constant over the periods.

- Even 2 period DiD allows control variables.
- Controls may be more important than in an RCT to reduce variation & to remove omitted variable bias.

Diff vs. DiD - Identifying assumption #2

- Identifying assumption #2: $\mathbb{E}[\epsilon_{it}|\boldsymbol{X}_{it},\ T_i,\ D_{after,it},\ \alpha_i] = 0$.
- When would this be violated?
- Technically, the "shock" in the 1^{st} period leads somebody to (not) choose the treatment.

Diff vs. DiD - Identifying assumption #2

- Substantively?
- Example #1: Bloom et al., 2015
  - Those that know their productivity was (temporarily) lower decide to work from home (or office).
- Example #2: Gil, 2015
  - Think again of the effect of hiring a new CEO on firm performance. Firm observes a shock to performance compared to peers, and therefore hires a new CEO.
- Example #3: Aghion et al., 2018
  - Inventions usually do not come as a surprise. The firm may change wages in anticipation of the invention.
  - Notice AAHT take this into account by dividing the treatment period into pre- and post-invention. The problem may still remain.

What data for more than 2 periods?

- More data always a plus.
- Makes distinction between (differential) trends and temporary shocks clearer.
- Can allow for more flexible models (e.g. introduction of time/period dummies; testing of common trends using treatment group - time period dummy interactions).
- BUT: notice that one stretches what α_i captures.
- Remember: Even 2-period DiD allows control variables.

A complication on all causal estimators

- What have we assumed about the effect of treatment on the control group?
- That there is none.

= "no general equilibrium effects".

= "**S**table **U**nit **T**reatment **V**alue **A**ssumption (SUTVA)".

A complication on all causal estimators

- When is this an issue?
- In a lab, think of infectious diseases.
- Regarding human behavior, think of interactions (markets).
- Important but difficult topic. We will neglect it, as is all too often done in the literature, too.

SUTVA

- Example #1: a merger affects the prices of all firms (products) in the market, not just those of the merging parties.
- Example #2: a wholesale education reform (think of the Finnish reform making secondary education compulsory) affects the wages of not only those whose education changes because of the reforms, but also of those who compete with them in the job market.
- Example #3: a regulatory reform affects the prices of all pharmaceuticals based on the same molecule. Kortelainen et al. (2023) study reforms in Nordic pharmaceutical markets and define prices both at the
  1. market-level and the
  2. package-level.

ATT: Main outcome variables

	Part I		Part II		Part III	
	Finland 2003 VGS → GS	Finland 2009 GS → IRP	Denmark 2000 IRP → ERP	Denmark 2005 ERP → IRP	Norway 2005 GS → SP-IRP	Sweden 2009 GS-IRP → Auction-IRP
Average Expenditure	-0.03	-0.13*	-0.05*	0.04	-0.21*	-0.27*
	[-0.07, 0.01]	[-0.18, -0.08]	[-0.09, -0.01]	[-0.01, 0.09]	[-0.29, -0.12]	[-0.34, -0.20]
Number of Product Names	0.01	0.04	-0.02	-0.01	-0.01	0.04
	[-0.03, 0.05]	[-0.02, 0.10]	[-0.06, 0.02]	[-0.05, 0.03]	[-0.15, 0.15]	[-0.00, 0.09]
Average Price	-0.04	-0.05	-0.07*	0.07*	-0.10	-0.04
	[-0.12, 0.04]	[-0.09, -0.00]	[-0.12, -0.01]	[0.02, 0.12]	[-0.18, -0.00]	[-0.11, 0.04]
Number of Doses	0.01	0.04*	0.00	0.07*	0.04	0.12*
	[-0.04, 0.07]	[0.01, 0.07]	[-0.04, 0.04]	[0.03, 0.12]	[-0.00, 0.09]	[0.02, 0.22]
Wholesale Price	-0.05	-0.10*	-0.09*	0.05	-0.11*	-0.06*
	[-0.11, 0.02]	[-0.14, -0.07]	[-0.13, -0.05]	[-0.02, 0.12]	[-0.20, -0.01]	[-0.11, -0.01]

[1] Estimator: Two-way fixed effects and Callaway and Sant'Anna (2020). Outcome data source: DLI-MI (1999–2013), Farmastat (2004–2013), Fimea (1999–2012), IQVIA MIDAS Quarterly Sales and IQVIA MIDAS (2007–2013).
[2] * = statistically significant at the 95% confidence level. 10000 replications for ATC-5 wild bootstrapped standard errors.

Kortelainen, M., Markkanen, J., Siikanen, M. & Toivanen, O. (2023). *The effects of price regulation on pharmaceutical expenditure and availability* [Unpublished manuscript].

Choosing the comparison group

- A large body of literature has demonstrated that key to success in using DiD (more generally, in identifying causal effects) is the choice of the control group.
- Control group observation units should be "as similar" to treatment group observation units.

$\rightarrow$ **conditional DiD**.

- Conditional = first choose carefully which observation units to include in the control group.
- When done correctly, this helps a great deal.

Choosing the comparison group

- Execution of conditional DiD:

1. Choose some key characteristics.
2. Choose treatment group observation unit #1.
3. Go through potential control group observation units and choose a unit / units that are as similar as the treatment group observation unit #1. Many different technical solutions to implement this.
4. Repeat for all treatment group observation units.

Choosing the comparison group

- Aghion, P., Akcigit, U., Hyytinen, A. & Toivanen, O. (2018). On the returns to invention within firms: Evidence from finland. *AEA Papers and Proceedings*, *108*, 208–12
- We study what happens to wages of individuals after invention.
- We split individuals in a firm into 4 groups:

1. Inventors
2. Entrepreneurs
3. White-collar workers
4. Blue-collar workers

Treatment group

- Those in the treatment group work in the same firm as the inventor in the year of the patent application.

Control group

- Those in the control group:

1. Never work in a firm that invents.
2. Have the same socioeconomic status (excl. inventors)
3. Are similar to an inventor in terms of
   i. Education (MSc or not)
   ii. Age ($< 30, 31 - 40, 41 - 50; > 50$)
   iii. Quintiles of firm size
   iv. IQ ($< 50^{th}$ percentile, $51^{st} - 80^{th}$, $81^{st} - 90^{th}$, > 90).

Results

Table 1—Returns Estimation

Variables	Inventor	Entrepreneur	White-collar	Blue-collar
Treated × pre	0.0417	–0.0153	0.00567	–0.0107
	(0.0133)	(0.0825)	(0.00402)	(0.00504)
Treated × post	0.0511	0.279	0.0208	0.0227
	(0.0162)	(0.0902)	(0.00463)	(0.00556)
Observations	93,939	13,372	1,320,370	916,811
R^2	0.329	0.180	0.347	0.256
Number of individuals	8,185	1,123	107,986	87,288
Dependent variable	ln wage	ln wage	ln wage	ln wage
Age fixed effects	Yes	Yes	Yes	Yes
Calendar year fixed effects	Yes	Yes	Yes	Yes
Treatment year fixed effects	Yes	Yes	Yes	Yes
Individual fixed effects	Yes	Yes	Yes	Yes
Age × calendar year fixed effects	No	No	No	No
Pre-treatment effects	Yes	Yes	Yes	Yes
Sample	Base	Base	Base	Base

DiD-issues we ignore but which should be considered

- Different units get the treatment at different times (at the extreme, all units eventually get the treatment, leading to an **event study** setting).
- Different units get a different treatment. Example: Finnish cost subsidies to firms during the COVID-19 crises vary from 2 000€ to 500 000€.
- The effect of the treatment is different for different treatment units, possibly conditional on observables.
- DID methods have developed rapidly in the last few years re all these issues. It is now well understood that the base two-way FE DID may produce biased results in settings that are even a little more complicated than the textbook setting.