Lecture 3

Difference-in-differences

Policy Evaluation
Nova SBE - Universidade Nova de Lisboa / IGC
Pedro C. Vicente
http://www.pedrovicente.org/

Introduction

• In a panel setting, Difference-in-differences (DD) estimation implies measuring outcomes and covariates for both participants and non-participants in pre- and post-intervention periods

• DD compares treatment and comparison groups in terms of outcome changes over time:

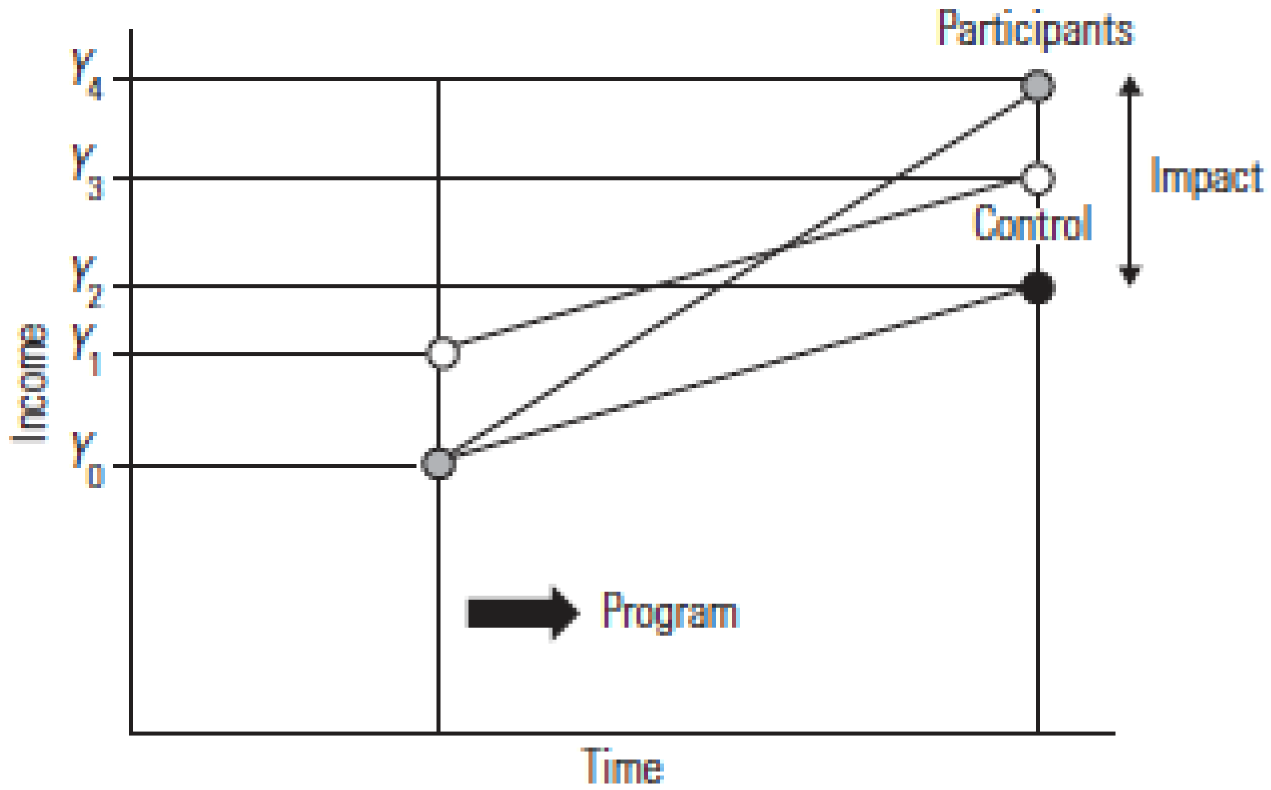
$$DD = E(Y_1^T - Y_0^T \mid T_1 = 1) - E(Y_1^C - Y_0^C \mid T_1 = 0)$$

where T_1=1 denotes treatment at t=1, T_1=0 otherwise; Y_t^T,Y_t^C are potential outcomes at time t

- Unlike PSM as we studied it, the DD estimator allows for unobserved heterogeneity (i.e., the unobserved difference in mean counterfactual outcomes between treated and control units) leading to selection bias
- For example, one may want to account for factors unobserved by the researcher, such as differences in innate ability across treated and control subjects
- DD assumes this unobserved heterogeneity is time invariant, so the bias cancels out through time differencing
    - This is weaker than the strict exogeneity assumption from before, as the assumed counterfactual is a time-difference

$$E(Y_1^T - Y_0^T \mid T_1 = 0)$$

Figure 5.1 An Example of DD



$$DD = (Y_4 - Y_0) - (Y_3 - Y_1) = Y_4 - Y_2$$

$$Assumption : Y_3 - Y_2 = Y_1 - Y_0$$

Theory

• The DD estimate can be calculated within a regression framework:

$$Y_{it} = \alpha + \beta T_{i1} t + \rho T_{i1} + \gamma t + \varepsilon_{it}$$

where the coefficient β on the interaction between the post-program treatment variable and time gives the DD effect of the program

• Let's use the DD definition (suppressing subscript *i*) to interpret the regression coefficients:

$$E(Y_1^T - Y_0^T \mid T_1 = 1) = (\alpha + \beta + \rho + \gamma) - (\alpha + \rho) = \beta + \gamma$$

$$E(Y_1^C - Y_0^C \mid T_1 = 0) = (\alpha + \gamma) - \alpha = \gamma$$

$$E(Y_0^T \mid T_1 = 1) - E(Y_0^C \mid T_1 = 0) = (\alpha + \rho) - \alpha = \rho$$

$$E(Y_1^T \mid T_1 = 1) - E(Y_1^C \mid T_1 = 0) = (\alpha + \beta + \rho + \gamma) - (\alpha + \gamma) = \beta + \rho$$

$$E(Y_0^C \mid T_1 = 0) = \alpha$$

$$DD = E(Y_1^T - Y_0^T \mid T_1 = 1) - E(Y_1^C - Y_0^C \mid T_1 = 0) = \beta$$

- For the DD regression estimator to be interpreted correctly, the following conditions must hold:
  - The model in the regression equation is correctly specified (namely the additive structure is correct)
  - The error term is uncorrelated with the other variables in the equation

$$\text{cov}(e_{it}, T_{i1}) = 0$$

$$\text{cov}(e_{it}, t) = 0$$

$$\text{cov}(e_{it}, T_{i1}t) = 0$$

The last assumption is the most critical, also named as the parallel-trend assumption, which imposes that unobserved characteristics affecting the outcome do not vary over time with treatment status

• We can add individual fixed effects in the above regression, which control for unobserved time-invariant heterogeneity

• And we can also control for heterogeneity in observed characteristics over time

$$Y_{it} = \alpha + \beta T_i t + \rho T_i + \gamma t + \delta X_{it} + \eta_i + \varepsilon_{it}$$

where X_{it} is a vector of potentially time-variant observed characteristics, and η_i is an individual fixed effect

• Note that time-invariant terms are dropped with individual fixed effects, namely T_i alone and potentially parts of X_{it}

- Differencing both sides of the previous equation over time (*t* and *t*-1):

$$Y_{it} - Y_{it-1} = (\alpha + \beta T_i t + \rho T_i + \gamma t + \delta X_{it} + \eta_i + \varepsilon_{it}) -$$
$$-(\alpha + \beta T_i (t-1) + \rho T_i + \gamma (t-1) + \delta X_{it-1} + \eta_i + \varepsilon_{it-1}) =$$
$$\Delta Y_{it} = \beta T_i + \gamma + \delta \Delta X_{it} + \Delta \varepsilon_{it}$$

- The preceding two-period model can be generalized with multiple time periods, which is called the panel fixed-effects model:

$$Y_{it} = \varphi T_{it} + \delta X_{it} + \eta_i + \mu_t + \varepsilon_{it} \Leftrightarrow$$
$$\Leftrightarrow Y_{it} - Y_{it-1} = (\varphi T_{it} + \delta X_{it} + \eta_i + \mu_t + \varepsilon_{it}) -$$
$$-(\varphi T_{it-1} + \delta X_{it-1} + \eta_i + \mu_{t-1} + \varepsilon_{it-1}) =$$
$$\Delta Y_{it} = \Delta \mu_t + \varphi \Delta T_{it} + \delta \Delta X_{it} + \Delta \varepsilon_{it}$$

- We can always run OLS on the time differenced equation and estimate the unbiased effect of the program (analogously to using fixed effects)
- Still, we are assuming that the treatment is not correlated with the error term, i.e., with time-variant unobservables
    - This is still a strong assumption in some applications
    - However it may be reasonable in others
    - And actually, it can be the best identifying assumption in many applications

- Bertrand, Duflo, and Mullainathan (QJE, 2004) draw attention to the possibility of serial correlation in the generalized DD model:
    - OLS produces inconsistent standard errors
    - They generate placebo interventions in an American panel and find that conventional standard errors severely understate the standard deviation of the estimators: they find a significant effect at the 5 percent level for up to 45 percent of the placebo interventions
    - Parametric assumptions on the autocorrelation process do not perform well
    - Solutions: Block bootstrap (taking into account autocorrelation) and collapsing of time into pre and post work well; Huber-White clustered standard errors (at the state-level) do not work well with a low number of states

Alternative DD models

- Controlling for initial conditions
  - This is appropriate in the case where public investments depend on initial preprogram local area conditions; without these initial conditions the DD treatment effect could be biased
- PSM with DD
  - Provided that rich data on treatment and control areas exist, PSM can be combined with DD to better match treatment and control units on pre-program characteristics
  - The propensity score can be used to match participant and control units in the pre-program year
  - The treatment impact is calculated across participant and non-participant units within the common support

- For two time periods, the DD estimate using PSM would be given by:

$$DD_{PSM} = \frac{1}{N_T} \sum_{i \in T} \left[(Y_{i2}^T - Y_{i1}^T) - \sum_{j \in C} \omega(i,j)(Y_{j2}^C - Y_{j1}^C) \right]$$

where $\omega(i,j)$ is the weight using the PSM approach given to the jth control area matched to treatment area i

- Triple-difference methods
  - E.g., adjusting for differential time trends

$$DDD = [E(Y_1^T - Y_0^T \mid T=1) - E(Y_1^C - Y_0^C \mid T=0)] -$$

$$- [E(Y_0^T - Y_{-1}^T \mid T=1) - E(Y_0^C - Y_{-1}^C \mid T=0)]$$

Example: Building Schools in Indonesia

- **Duflo (AER, 2001)**
  - This paper exploits a dramatic change in policy to evaluate the effect of building schools on education and earnings in Indonesia
  - Between 1973-74 and 1978-79 more than 61,000 primary school buildings (an average of two schools per 1,000 children aged 5-14 in 1971) were built
  - Enrollment rates among children aged 7-12 were 69 percent in 1973; these increased to 84 percent (males) and 82 percent (females) in 1978

- The identification strategy of the paper uses the fact that exposure to the school construction program varied by region and date of birth
    - Substantial variations existed in program intensity across regions, due to the government's efforts to allocate more schools to regions where initial enrollment was low
    - Difference-in-difference: education should be higher for younger (in school when program was launched) vs. older cohorts for all regions; but this difference should be larger in the regions that received more schools

Duflo (AER, 2001) - results

- The estimates suggest that the construction of primary schools led to an increase in education and earnings for children aged 2-6 in 1974 (for each school constructed per 1,000 children in their region of birth)
    - received 0.12 to 0.19 more years of education
    - received 1.5 to 2.7 percent higher wages
- Robustness: test exogeneity of interaction of age cohort with region by checking whether the program is correlated with the outcomes (educational attainment and wages) for older cohorts

(1) $$S_{ijk} = c_1 + \alpha_{1j} + \beta_{1k} + (P_j T_i)\gamma_1 + (\mathbf{C_j} T_i)\boldsymbol{\delta_1} + \varepsilon_{ijk}$$

where S_{ijk} is the education of individual i born in region j in year k, T_i is a dummy indicating whether the individual belongs to the "young" cohort in the subsample, c_1 is a constant, β_{1k} is a cohort of birth fixed effect, α_{1j} is a district of birth fixed effect, P_j denotes the intensity of the program in the region of birth, and $\mathbf{C_j}$ is a vector of region-specific variables.

TABLE 3—MEANS OF EDUCATION AND LOG(WAGE) BY COHORT AND LEVEL OF PROGRAM CELLS

	Years of education			Log(wages)		
	Level of program in region of birth			Level of program in region of birth		
	High (1)	Low (2)	Difference (3)	High (4)	Low (5)	Difference (6)
Panel A: Experiment of Interest						
Aged 2 to 6 in 1974	8.49	9.76	−1.27	6.61	6.73	−0.12
	(0.043)	(0.037)	(0.057)	(0.0078)	(0.0064)	(0.010)
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Difference	0.47	0.36	0.12	−0.26	−0.29	0.026
	(0.070)	(0.038)	(0.089)	(0.011)	(0.0096)	(0.015)
Panel B: Control Experiment						
Aged 12 to 17 in 1974	8.02	9.40	−1.39	6.87	7.02	−0.15
	(0.053)	(0.042)	(0.067)	(0.0085)	(0.0069)	(0.011)
Aged 18 to 24 in 1974	7.70	9.12	−1.42	6.92	7.08	−0.16
	(0.059)	(0.044)	(0.072)	(0.0097)	(0.0076)	(0.012)
Difference	0.32	0.28	0.034	0.056	0.063	0.0070
	(0.080)	(0.061)	(0.098)	(0.013)	(0.010)	(0.016)

Notes: The sample is made of the individuals who earn a wage. Standard errors are in parentheses.

TABLE 4—EFFECT OF THE PROGRAM ON EDUCATION AND WAGES: COEFFICIENTS OF THE INTERACTIONS BETWEEN COHORT DUMMIES AND THE NUMBER OF SCHOOLS CONSTRUCTED PER 1,000 CHILDREN IN THE REGION OF BIRTH

		Dependent variable					
		Years of education			Log(hourly wage)		
	Observations	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Experiment of Interest: Individuals Aged 2 to 6 or 12 to 17 in 1974 (Youngest cohort: Individuals ages 2 to 6 in 1974)							
Whole sample	78,470	0.124	0.15	0.188			
		(0.0250)	(0.0260)	(0.0289)			
Sample of wage earners	31,061	0.196	0.199	0.259	0.0147	0.0172	0.0270
		(0.0424)	(0.0429)	(0.0499)	(0.00729)	(0.00737)	(0.00850)
Panel B: Control Experiment: Individuals Aged 12 to 24 in 1974 (Youngest cohort: Individuals ages 12 to 17 in 1974)							
Whole sample	78,488	0.0093	0.0176	0.0075			
		(0.0260)	(0.0271)	(0.0297)			
Sample of wage earners	30,225	0.012	0.024	0.079	0.0031	0.00399	0.0144
		(0.0474)	(0.0481)	(0.0555)	(0.00798)	(0.00809)	(0.00915)
Control variables:							
Year of birth*enrollment rate in 1971		No	Yes	Yes	No	Yes	Yes
Year of birth*water and sanitation program		No	No	Yes	No	No	Yes

Notes: All specifications include region of birth dummies, year of birth dummies, and interactions between the year of birth dummies and the number of children in the region of birth (in 1971). The number of observations listed applies to the specification in columns (1) and (4). Standard errors are in parentheses.

What have we learnt?

- Difference-in-differences relaxes the assumption of selection on observables, since it proposes parallel-trends and allows for selection on time-invariant unobservables
- However, selection on time-variant unobservables is plausible in many applications
- Alternative DD models:
    - DD with PSM
    - Triple-difference

Problem Set 3

Exercise: Follow the instructions in chapter 14 of Khandker et al (2009). You should produce a do file and a log file, which should be commented to show that you understood the results. These should be emailed to the grader (Matilde Grácio): only one email per group, please.

Note: We should be able to run the do file on a computer given the original datafile and produce the raw log file.

Due date: Tuesday, March 8.