Applied Econometrics JEM007, IES
Lecture 3

DIFFERENCE-IN-DIFFERENCE

Eva Hromádková, 14.10.2010

Introduction

Why don't we just do experiments?

- Direct randomization often not feasible
- People in experiments behave differently than in real world situations

Introduction of public policy / law – similar to experiment

- Defined control and treatment group pre- and post-intervention

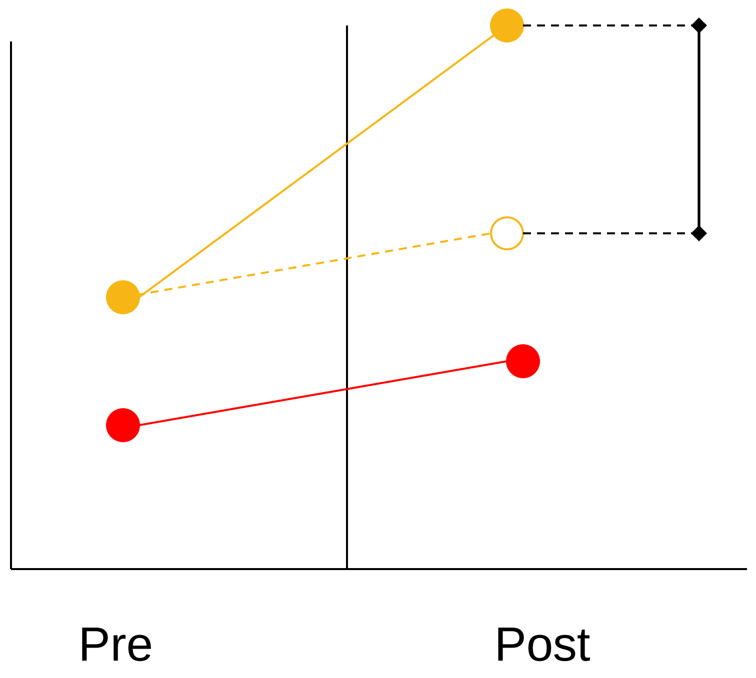
Ex.1: eligibility of individuals based on observable characteristics

Ex.2: different timing of policy implementation across countries / regions / groups within population

Diff-in-diff estimator

Intuition I

- DD estimator = comparison of outcomes for treated and control group before and after treatment



Effect of program difference-in-difference (taking into account pre-existing differences between T & C and general time trend).

Diff-in-diff estimator

Intuition II

- Baseline assumption:

  difference between treatment and control group is constant over time

  - D0: difference pre-treatment = normal diff
  - D1: difference post-treatment = normal diff + treatment effect
  - D1-D0 = treatment effect

Diff-in-diff estimator

Mathematics behind

- Explore a policy rule occurring at period k – denote periods before k as time t_0 and after as time t_1
- Follow individuals I before and after policy change
- treatment status $d_i = 1$ if $d_{it} = 1$ at t_1

$$y_{it} = \beta + \alpha_i d_{it} + u_{it}$$

$$\text{where } \mathrm{E}\left[u_{it} \middle| d_i, t\right] = \mathrm{E}\left[n_i \middle| d_i\right] + m_t.$$

  - Here, n_i is individual fixed effect (that can be correlated with treatment status) and m_t is aggregate macro shock, common to everyone

$$\mathrm{E}\left[y_{it} \middle| d_i, t\right] = \begin{cases} \beta + \mathrm{E}\left[\alpha_i \middle| d_i = 1\right] + \mathrm{E}\left[n_i \middle| d_i = 1\right] + m_t & \text{if } d_i = 1 \text{ and } t = t_1 \\ \beta + \mathrm{E}\left[n_i \middle| d_i\right] + m_t & \text{otherwise.} \end{cases}$$

Diff-in-diff estimator

Mathematics behind II

$$D_0 = \mathrm{E}[y_{it} \mid d_i = 1, t = t_0] - \mathrm{E}[y_{it} \mid d_i = 0, t = t_0] =$$

$$= \{\beta + \mathrm{E}[n_i \mid d_i = 1] + m_t\} - \{\beta + \mathrm{E}[n_i \mid d_i = 0] + m_t\} =$$

$$= \mathrm{E}[n_i \mid d_i = 1] - \mathrm{E}[n_i \mid d_i = 0]$$

$$D_1 = \mathrm{E}[y_{it} \mid d_i = 1, t = t_1] - \mathrm{E}[y_{it} \mid d_i = 0, t = t_1] =$$

$$= \{\beta + \mathrm{E}[\alpha_i \mid d_i = 1] + \mathrm{E}[n_i \mid d_i = 1] + m_t\} - \{\beta + \mathrm{E}[n_i \mid d_i = 0] + m_t\} =$$

$$= \mathrm{E}[\alpha_i \mid d_i = 1] + \mathrm{E}[n_i \mid d_i = 1] - \mathrm{E}[n_i \mid d_i = 0]$$

$$D_1 - D_0 = \mathrm{E}[\alpha_i \mid d_i = 1] \quad (ATT)$$

Assumptions (revised) – treatment and control group can be different in unobservables – these differences have to be constant(or predictable) over time

Diff-in-diff estimator

Implementation

- Panel data: same individuals in different periods

$$\Delta y_i = \beta_0 + \beta_1 d_i + \varepsilon_i$$

- Repeated cross-section data:

$$y_{it} = \beta_0 + \beta_1 d_i + \beta_2 T_t + \beta_3 d_i T_t + \varepsilon_{it}$$

where $T_t=1$ if $t = t_1$

What if we ran both specifications on panel data?

- Same coefficient estimates, different SE
- Second specification assumes independent observations, which is unlikely in case of panel
- Try clustering ☺

Diff-in-diff estimator

Implementation II

- Inclusion of other regressors – OK
  - ! CAREFUL – how you put them into equation
  - E.g. if X affects level of y => ΔX should be in the difference version (spec 1)
- Different trends for control and treatment group
  - If more than 2 periods available => you can test for it (visually, statistically) and adjust – e.g. put time effects into regression

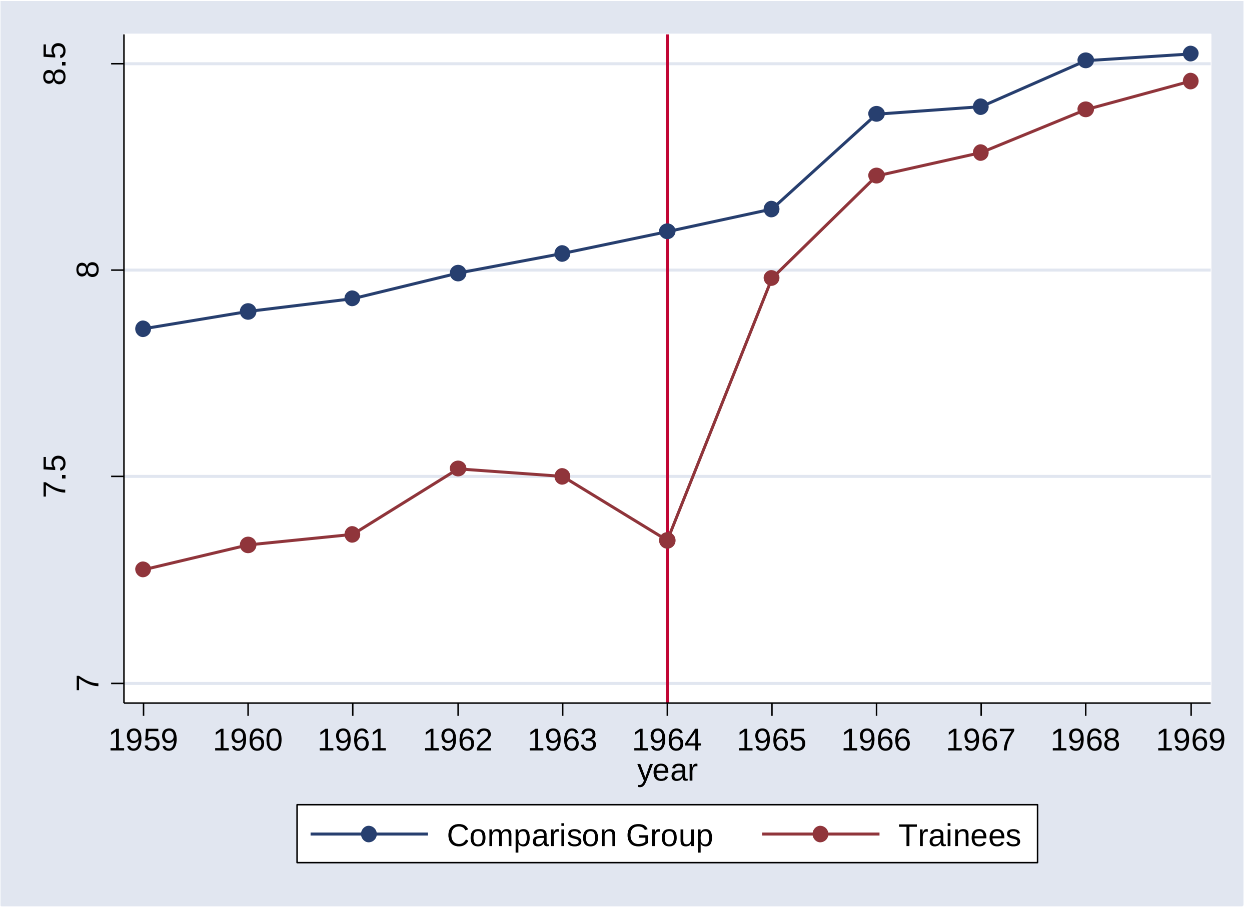
Diff-in-diff estimator

Issue A – Ashenfelter’s Dip

- `pre-program ***dip***', for participants
  - Related to the idea of *mean reversion*: individuals experience some idiosyncratic shock and enter program when things are especially bad
  - Would have improved anyway (reversion to the mean)
- Another issue may be if treatment is *selected* by participants - then only the worst off individuals elect the treatment =>not comparable to general effect of policy
- Ex: effect of government sponsored training on earnings

Diff-in-diff estimator

Issue A – Ashenfelter's Dip - example



Exercise: write DD specification to analyze the effect

Diff-in-diff estimator

Issue A – Ashenfelter's Dip - example

- Earnings for trainees very low in 1964 as training not working in that year – should ignore this year

=> Always understand how the policy works!!!

- Simple D-in-D approach would compare earnings in 1965 with 1963
- But earnings of trainees in 1963 seem to show a 'dip' – so D-in-D assumption probably not valid
- Probably because those who enter training are those who had a bad shock (e.g. job loss)

Diff-in-diff estimator

Issue B – Anticipation of policy step

- People anticipate the policy step and adjust to it
- Examples:
  - Tax reform: people shift taxable income to the next period to take advantage of lower marginal tax rate
  - Co-payments: people withdraw their recipes before the introduction of co-payments to lower costs

=> Could policy have been anticipated? What effect would it have on the behavior of people? In which direction could this affect your estimates?

Diff-in-diff estimator

Issue C – Macro trends

- Different macro trends [m_t] affecting treatment and control group
- Example – generation specific characteristics
  - Cohort specific shocks (e.g. born before/after 1989)
  - Different trends for unemployment of older/younger people

Example 1: Anti-malaria campaign

Malaria Eradication in the Americas (Bleakley, 2007)

Question: How much childhood exposure to malaria depresses labor productivity?

Data: Malaria Eradication campaign in

- Southern United States (1920's)
  - + Brazil, Colombia, Mexico (1950's)

Diff-in-Diff:

- birth cohorts - old vs. young people at the time of campaign
- regions with high vs. low incidence of malaria

Example 1: Anti-malaria campaign

Intuition

- Areas with high pre-treatment malaria will benefit more from malaria eradication
- Treatment group: Young people living in high pre-treatment malaria areas will benefit more than older people
  - older people might have partial immunity
- Comparison group: young and older people living in low pre-treatment malaria areas – natural evolution of income over cohorts (without malaria)

Example 1: Anti-malaria campaign

Empirical model

$$Y_{jkt} = \beta_k \, \mathrm{M}_j + \delta_k + X_j \, \Gamma_k + \nu_{jkt}$$

Y_{jkt} – average outcome (income) in area of birth j for cohort k at time t

M_j – pre-campaign malaria intensity in area of birth j

⇒ β_k – year-of-birth specific coefficient on malaria

X_j – state-of-birth controls (health and education related)

=> They have run this separately for each cohort and obtained β_k

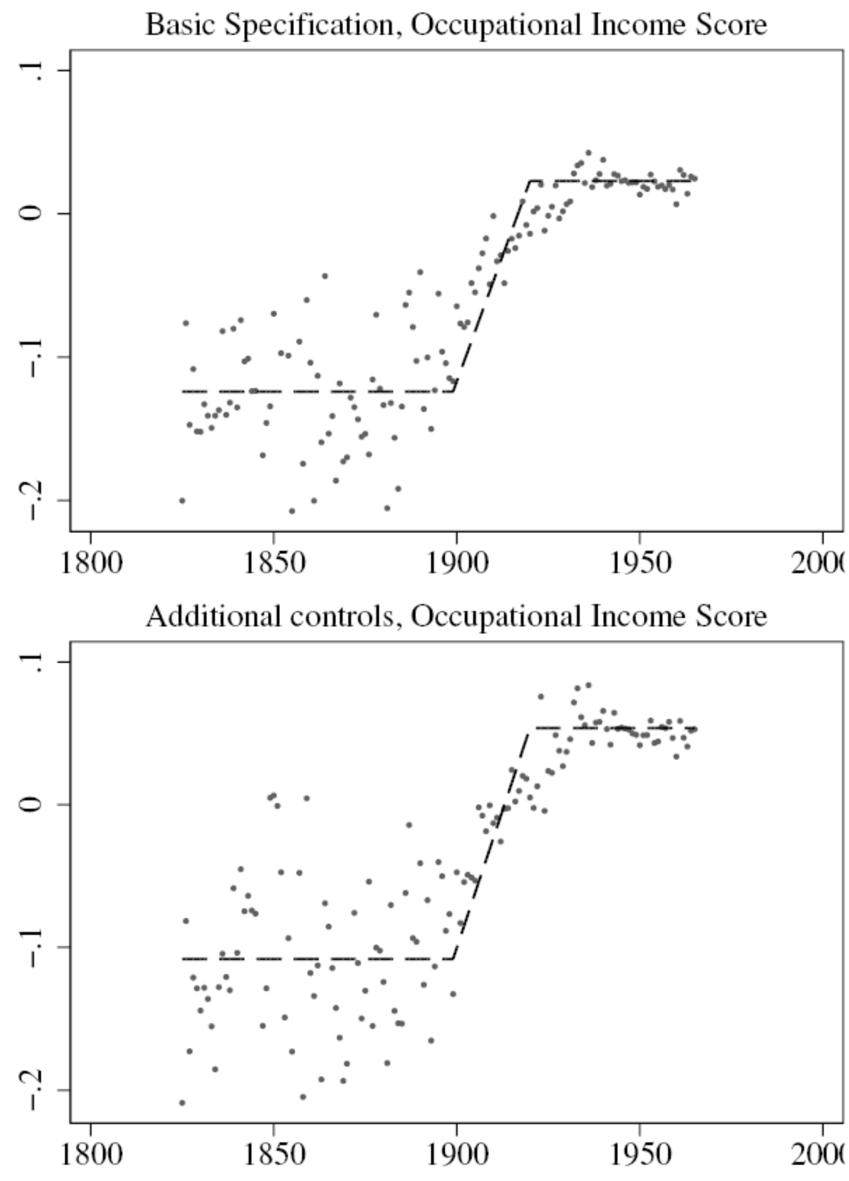
Example 1: Anti-malaria campaign

Results

Hypothesis about βk

(if exposure to malaria in younger age has effect) :

- For older cohorts (before 1900) – negative relationship between malaria intensity and outcomes
- For younger cohorts (after 1920) – relationship was purged by the effect of campaign
- In-between – decreasing strenght of the relationship (more and more exposure to campaign in the childhood



Example 2: D-in-D-in-D

Set-up

Implementation of (imaginary) health care policy, aiming at people of age 65 and older in country A

- Looking at effect on health outcomes (y)
- DD approach:
  - 2 periods (before x after);
  - control group age 55-65
- ? What problems do you see?

Example 2: D-in-D-in-D

Comparison groups

Let’s use elderly patients from the country B, where the health reform wasn’t introduced at all

3 dummies:

- Eligibility: d_i=1 if age of person i>65
- Time eligibility: T_t = 1 if time period t is AFTER
- Country identificator: A_i=1 if person i from country A

$$y_{it} = \beta_0 + \beta_1 d_i + \beta_2 A_i + \beta_3 d_i A_i + \\ + \delta_0 T_t + \delta_1 T_t A_i + \delta_2 d_i T_t + \delta_3 d_i T_t A_i + \varepsilon_{it}$$

Example 2: D-in-D-in-D

Interpretation of coefficient

$$\bar{\delta}_3 = (\bar{y}_{A,d=1,T=2} - \bar{y}_{A,d=1,T=1}) - (\bar{y}_{B,d=1,T=2} - \bar{y}_{B,d=1,T=1}) - (\bar{y}_{A,d=0,T=2} - \bar{y}_{A,d=0,T=1})$$

By including different control groups, we hope to control for different confounding factors

- Cohort specific
- State specific

Reality check – Bertrand et al. (2004)

- How much should we trust diff-in-diff estimates?

General specification of D-in-D model:

$$Y_{ist} = A_s + B_t + cX_{ist} + \beta I_{st} + \varepsilon_{ist}$$

A_s – state (group) fixed effect [dummies for each, -1]

B_t – time fixed effect effect [dummies for each, -1

X_{ist} – individual controls

I_{st} – indication whether policy has effect on state s at time t

- Usually cluster by year & state (group)
- Are standard errors OK?

Reality check – Bertrand et al. (2004)

How does DD perform on placebo laws?

- Take typical data used in DD estimations
  - CPS, women 25-50 with positive earnings, 50 years
- Assign randomly treated states and years of introduction
- “If hundreds of researchers analyzed the effects of various laws in the CPS, what fraction would find a significant effect even when laws have no effect?”
- Significant effect at 5% level should be found in … % of cases

Reality check – Bertrand et al. (2004)

Result: Bertrand et al. has found significant effect in 45% of cases!! (even after clustering)

Reason = serial (time) correlation problem

- Use of fairly long time-series (avg.16.5 periods)
- Dependent variables (e.g. income) are typically highly positively serially correlated
  - And not only AR(1)
- Treatment variable has small variation over time; usually 0 before and 1 after – think malaria

Reality check – Bertrand et al. (2004)

Solution:

- Block-bootstrapping: OK if large number of groups
- Aggregate data to 2 periods – before and after, for each group (small # of groups)
- Allow for unrestricted covariance over time within states – cluster on states!!! (EASY)