Capacity Building workshop on Impact Evaluation of Employment Programs

Quasi-Experimental Designs
Part 2: Difference-in-Difference and Matching

Celine Ferre, Gdańsk, February 22, 2017
Quasi-experimental methods (require more assumptions)

IE Methods Toolbox

- Randomized Assignment
- Regression Discontinuity Design
- Difference-in-Differences
- Matching
Randomized Assignment

Regression Discontinuity Design

Difference-in-Differences

Matching

IE Methods Toolbox
DIFFERENCE-IN-DIFFERENCES
(DIFF-IN-DIFF)
## Difference-in-differences (Diff-in-diff)

Y = Probability of being employed  
\( P = \) Youth training program

\[
\text{Diff-in-Diff: Impact} = (Y_{T1} - Y_{T0}) - (Y_{C1} - Y_{C0})
\]

<table>
<thead>
<tr>
<th></th>
<th>Enrolled (T)</th>
<th>Not Enrolled (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>After (1)</td>
<td>0.74</td>
<td>0.81</td>
</tr>
<tr>
<td>Before (0)</td>
<td>0.60</td>
<td>0.78</td>
</tr>
<tr>
<td>Difference</td>
<td>+0.14</td>
<td>+0.03</td>
</tr>
</tbody>
</table>

\[= 0.11\]
Impact = \( (A-B)-(C-D) = (A-C)-(B-D) \)

Probability of being employed

Similar trends before the program
## Example of Progresa

<table>
<thead>
<tr>
<th></th>
<th>Enrolled</th>
<th>Not Enrolled</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Follow-up (t=1)</strong></td>
<td>268.75</td>
<td>290</td>
<td>-21.25</td>
</tr>
<tr>
<td>Consumption (Y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Baseline (t=0)</strong></td>
<td>233.47</td>
<td>281.74</td>
<td>-48.27</td>
</tr>
<tr>
<td>Consumption (Y)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Difference</strong></td>
<td>35.28</td>
<td>8.26</td>
<td>27.02</td>
</tr>
</tbody>
</table>

### Estimated Impact on Consumption (Y)

<table>
<thead>
<tr>
<th>Regression Type</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Regression</td>
<td>27.06**</td>
</tr>
<tr>
<td>Multivariate Linear Regression</td>
<td>25.53**</td>
</tr>
</tbody>
</table>

*Note: If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**).*
### Impact of Progresa on Consumption (Y)

<table>
<thead>
<tr>
<th>Case</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1: Before &amp; After</td>
<td>34.28**</td>
</tr>
<tr>
<td>Case 2: Enrolled &amp; Not Enrolled</td>
<td>-4.15</td>
</tr>
<tr>
<td>Case 3: Randomized Assignment</td>
<td>29.75**</td>
</tr>
<tr>
<td>Case 4: Discontinuity Design</td>
<td>30.58**</td>
</tr>
<tr>
<td><strong>Case 5: Difference-in-Differences</strong></td>
<td>25.53**</td>
</tr>
</tbody>
</table>

**Note:** If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**).
Example: New Deal for Young People [United Kingdom] Program

WHAT? Program for 18-24 y.o. who have been claiming Jobseeker’s Allowance (JSA) for 6 months or more. It provides opportunities to work, get new skills and/or get work experience in the voluntary and environmental sectors of the economy.


WHY? Help the young unemployed into work and increase their employability.

Method/Eligibility

- Individuals who are between 18 and 24 years old, and who have been registered unemployed for more than 6 months are the treatment group.
- Individuals who are between 30 and 39 years old, and who have been unemployed for more than 6 months are the control group.

Data

Joint Unemployment and Vacancies Operating System, (JUVOS), and covers the period up to February 2001, 32 months after the start of the national program.
Problem I: Assume Equal Trends

• Diff-in-Diff only valid if both groups had similar trends before the program.
• The change in observed outcomes for those not enrolled would have been the same for those that are enrolled.
• What if attendance for those enrolled would have increased by more than those not enrolled in any case?

VIOLATION OF EQUAL TRENDS!
Same Trend

![Graph showing probability of being employed over time with points labeled B=0.60, C=0.81, D=0.78, A=0.74. Text reads: Similar trends before the program.](image)
Different Trend

- **B=0.60**
- **C=0.81**
- **D=0.78**
- **A=0.74**

Different trends before the program

Diff-in-Diff **cannot** measure the impact of the program
What if an event affects only one group?

Case 1: Training program

- Only **highly motivated people participated** in the program
- Only candidates who are **expected to be successful** in the training are enrolled by the placement officers

**DiD overestimates the effect of the program**

Case 2: Grants to set up a business with the UK

- Treatment group = small businesses working with the UK
- Control group = small domestic businesses
- Only the treatment group is affected by the BREXIT

**DiD underestimates the effect of the program**
To test this, at least 3 observations in time are needed:

- 2 observations before
- 1 observation after.
Problem 2: Changes in group composition over time

- Diff-in-Diff requires that we follow *the same* types of people over time

For example, all the *skilled people drop out* of a training program, because they don’t need the training

So average training outcomes for those in the program is *lower* at the end of the program

→ **DiD underestimates the effect of the program**

For example, all the *unskilled people drop out* of a training program, because they cannot travel to the training location

→ **DiD overestimates the effect of the program**
Keep in Mind

Difference-in-Differences

Combines *Enrolled & Not Enrolled* with *Before & After*.

**Slope:** Generate counterfactual for change in outcome

**FUNDAMENTAL ASSUMPTION**
Trends –slopes- are the same in treatments and comparisons

To test this, at least 3 observations in time are needed:
- 2 observations before
- 1 observation after.
Exercise:
Use DiD to evaluate the impact of a retraining program for unskilled youth

What is the impact of the program on unskilled youth?

<table>
<thead>
<tr>
<th>Year</th>
<th>Unskilled youth (15-24)</th>
<th>Skilled youth (15-24)</th>
<th>Unskilled prime-age (25-44)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>65</td>
<td>83</td>
<td>72</td>
</tr>
<tr>
<td>2010</td>
<td>58</td>
<td>75</td>
<td>62</td>
</tr>
<tr>
<td>2013</td>
<td>60</td>
<td>77</td>
<td>63</td>
</tr>
</tbody>
</table>
Poland:
When can you use Difference-in-Difference?

Examples/discussion

• What are the programs in Poland that may have used Difference-in-difference (DiD)?

• Can you think about programs in Poland that could use Difference-in-difference (DiD)?

• How would you define participants and non-participants?

• Could there be some issues with Difference-in-difference (DiD)?
IE Methods Toolbox

- Randomized Assignment
- Regression Discontinuity Design
- Difference-in-Differences
- Matching
Matching

• The group that enrolled is, on average, different from the group that did not enroll.
• However, some individuals are similar.
• So, can match similar individuals with each other.
Group Exercise

Can everyone stand up?
Try to compare outcomes for similar people

<table>
<thead>
<tr>
<th></th>
<th>NOT ENROLLED</th>
<th>ENROLLED</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO EDUCATION</td>
<td>5:3</td>
<td>10:0</td>
</tr>
<tr>
<td>PRIMARY EDUCATION</td>
<td>6:4</td>
<td>12:0</td>
</tr>
<tr>
<td>SECONDARY EDUCATION</td>
<td>8:2</td>
<td>16:0</td>
</tr>
<tr>
<td>TERTIARY EDUCATION</td>
<td>10:0</td>
<td>20:0</td>
</tr>
</tbody>
</table>

Try to compare outcomes for similar people.
More Complicated in Practice

- Match on all observable characteristics (e.g. income, gender, education...)
- Comparison group: non-participants with similar characteristics
- Create one aggregate Propensity Score to match:
  - Compute everyone’s probability of participating, based on their observable characteristics
  - Choose matches that have the same probability of participation as the treatments
Try to compare outcomes for similar people

- **NOT ENROLLED**
  - NO EDUCATION
  - PRIMARY EDUCATION
  - SECONDARY EDUCATION
  - TERTIARY EDUCATION

- **ENROLLED**

The diagram illustrates comparisons between individuals who are not enrolled and those who are enrolled in different levels of education.
Problem Two: Can Only Match on Observables

MATCHING DOES NOT OVERCOME SELECTION PROBLEM!

What if we can’t collect data on people characteristics that are relevant for program participation and outcomes?
## Case 6: Progresa Matching (*P-Score*)

<table>
<thead>
<tr>
<th>Baseline Characteristics</th>
<th>Estimated Coefficient Probit Regression, Prob Enrolled=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head’s age (years)</td>
<td>-0.022**</td>
</tr>
<tr>
<td>Spouse’s age (years)</td>
<td>-0.017**</td>
</tr>
<tr>
<td>Head’s education (years)</td>
<td>-0.059**</td>
</tr>
<tr>
<td>Spouse’s education (years)</td>
<td>-0.03**</td>
</tr>
<tr>
<td>Head is female=1</td>
<td>-0.067</td>
</tr>
<tr>
<td>Indigenous=1</td>
<td>0.345**</td>
</tr>
<tr>
<td>Number of household members</td>
<td>0.216**</td>
</tr>
<tr>
<td>Dirt floor=1</td>
<td>0.676**</td>
</tr>
<tr>
<td>Bathroom=1</td>
<td>-0.197**</td>
</tr>
<tr>
<td>Hectares of Land</td>
<td>-0.042**</td>
</tr>
<tr>
<td>Distance to Hospital (km)</td>
<td>0.001*</td>
</tr>
<tr>
<td>Constant</td>
<td>0.664**</td>
</tr>
</tbody>
</table>

*Note: If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**)*. 
Case 6: Progresa Matching (P-Score)

<table>
<thead>
<tr>
<th>Estimated Impact on Consumption (Y)</th>
<th>7.06+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multivariate Linear Regression</td>
<td></td>
</tr>
</tbody>
</table>

Note: If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**). If significant at 10% level, we label impact with +
## Progresa Policy Recommendation?

### Impact of Progresa on Consumption (Y)

<table>
<thead>
<tr>
<th>Case</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1: Before &amp; After</td>
<td>34.28**</td>
</tr>
<tr>
<td>Case 2: Enrolled &amp; Not Enrolled</td>
<td>-4.15</td>
</tr>
<tr>
<td>Case 3: Randomized Assignment</td>
<td>29.75**</td>
</tr>
<tr>
<td>Case 4: Discontinuity Design</td>
<td>30.58**</td>
</tr>
<tr>
<td>Case 5: Difference-in-Differences</td>
<td>25.53**</td>
</tr>
<tr>
<td><strong>Case 6: Propensity Score Matching</strong></td>
<td>7.06+</td>
</tr>
</tbody>
</table>

*Note:* If the effect is statistically significant at the 1% significance level, we label the estimated impact with 2 stars (**).
Keep in Mind

Matching

Requires large samples and good quality data.

Matching at baseline can be very useful:
- Know the assignment rule and match based on it
- combine with other techniques (i.e. diff-in-diff)

Ex-post matching is risky:
- If there is no baseline, be careful!
- matching on endogenous ex-post variables gives bad results.
Example: Active Labor Market Programs
[Poland]

Programs

WHAT? Three ALMP programs to combat unemployment in Poland:
(i) training: for workers with redundant or no skills → skills in high demand
(ii) “intervention works”: wages subsidies in the amount of the unemployment benefit
(iii) public works: jobs created by local government/municipalities


Method/Eligibility

- Individuals who declare that they participated in one of the three programs are the treatment group
- Individuals who did not participate, but have similar characteristics before the program took place are the control group

Data

Special supplement of the Polish Labor force Survey (LFS) of 1996 which includes retrospective data

Results

(i) Training = performs well, (ii) “intervention works” = none for women, negative for men, (iii) public works = negative for men
Poland: When can you use Matching?

Examples/discussion

• What are the programs in Poland that may have used some sort of matching techniques?

• Can you think about programs in Poland that could use Propensity Score Matching (PSM)?

• How would you define participants and non-participants?

• Could there be some issues with Propensity Score Matching (PSM)?
Test
Q1: Which of the following is the main assumption associated with Difference in Difference estimates?

A. In the absence of the program, the treatment and control groups will experience the same trend in outcome indicators over time

B. In the absence of the program, the treatment and control groups will experience different trends in outcome indicators over time

C. Treatment and control groups experience different shocks that affect outcome indicators (rainfall, drought, etc.)
Q2: You are evaluating a school management reform program that targets poor school. You decide to perform a diff-diff, comparing target schools with schools that did not receive the program.

Over the same period government deployed more teachers to poor areas. Would this over-estimate or under-estimate the program?

A. Over-estimate
B. Under-estimate
C. Neither
Q3: What is the biggest short-coming of propensity score matching?

A. Cannot match on observables characteristics
B. Cannot match on unobservables characteristics
C. Different trends between treatment and comparison groups.
Thank you!