European Commission

Econometric analysis and counterfactual studies in the context of IA practices

Giulia Santangelo

http://crie.jrc.ec.europa.eu

CRIE

Centre for Research on Impact Evaluation

DG EMPL - DG JRC

Joint Research Centre

CRIE

- Centre for Research on Impact Evaluation
  - Joint DG EMPL-DG JRC initiative, established in June 2013
- Motto
  - Support to MS and DG EMPL to set up the necessary arrangements for carrying out Counterfactual Impact Evaluations (CIE) of interventions funded by the European Social Fund (ESF)
- Team
  - Researchers with a background in Economics and Statistics
  - Experience in labour economics, theoretical and applied econometrics, impact evaluation methods
- Why did DG JRC and DG EMPL decide to put money and invest in this field?... I come back **later** on this

Presentation Outline

The purpose of my talk

- Not really on Ex-ante Impact Assessment (IA)...more on Ex-post Impact Evaluation (CIE)
- However, we have a common interest: we want to know about the Impact
- Cross breeding of ideas useful

European Commission

Key messages

- Ex-post CIE can be used for Ex-ante IA
- The use of Ex-post CIE for Ex-ante IA implies
  - Promotion of a culture of Evaluation
  - Staff in charge of IA to be knowledgable on CIE methods
- Ex-post CIE must be planned well in advance in the Ex-ante IA

Our common interest: Impact

- We need to quantify the effect of a policy
- Why? Evidence based policy making needed to
  - Improve the quality and effectiveness of public policies
  - Promote accountability
- Quantitative methods
  - Macroeconomic models
  - Impact evaluation based on counterfactuals
- ...Important to make the distinction between Impact and Expected Impact
- CIE is about quantifying the impact per se of a policy

Presentation Outline

1. Talk: messages
2. **Macroeconomic models**
3. Counterfactual Impact Evaluations - CIE
4. A crash introduction to CIE methods
5. Randomised experiments
6. Quasi-experimental methods
7. Conclusion

Macroeconomic models

Pros

- Holistic representation of an economic system
- Ex ante and ex post evaluations: possible to evaluate the intended effect of the program at its design stage
- Possibility to test various scenarios

Cons

- Theoretically driven: modeler's perspective
  - Different models = different conclusions for a given policy
  - Calibration issue
- Not causal inference
- Complex for non experts

Presentation Outline

CIE: one tool among others

- Impact evaluation: particular type of evaluation
  - Seek to answer cause-and-effect questions
- One tool among others
  - Must be associated with a theory-based impact evaluation
  - Must be complemented by monitoring/process evaluation

Ingredients for a counterfactual-based evaluation

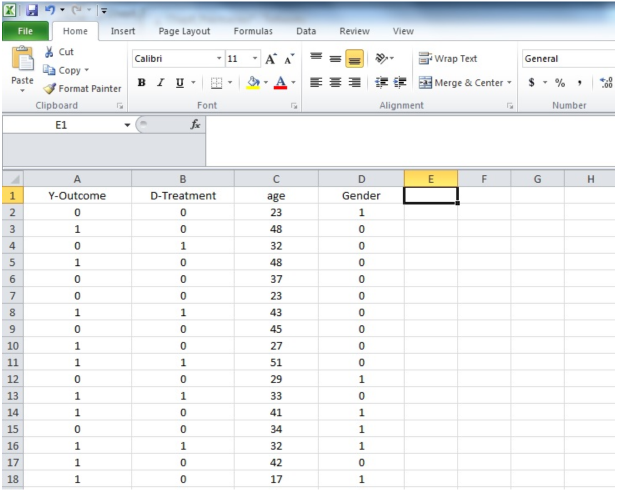
- Well defined intervention targeted at a well defined population
  - Some units i are exposed to the policy intervention (treated unit) others are not (untreated/control units)

$$
\begin{aligned}
D_i &= 1 \text{ if treated} \\
D_i &= 0 \text{ if untreated}
\end{aligned}
$$

- Expected effect: change of the outcome Y_i for the units i exposed to the intervention

$$
\begin{aligned}
Y_i &= Y_{1i} \text{ if treated} \\
Y_i &= Y_{0i} \text{ otherwise}
\end{aligned}
$$

What your dataset would look like



	A	B	C	D	E	F	G	H
1	Y-Outcome	D-Treatment	age	Gender				
2	0	0	23	1				
3	1	0	48	0				
4	0	1	32	0				
5	1	0	48	0				
6	0	0	37	0				
7	0	0	23	0				
8	1	1	43	0				
9	0	0	45	0				
10	1	0	27	0				
11	1	1	51	0				
12	0	0	29	1				
13	1	1	33	0				
14	1	0	41	1				
15	0	0	34	1				
16	1	1	32	1				
17	1	0	42	0				
18	1	0	17	1				

Ingredients for a counterfactual-based evaluation

- The treatment effect for individual i is defined by $Y_{1i} - Y_{0i}$
- Problem: only one potential outcome is observed which implies that we cannot measure $Y_{1i} - Y_{0i}$
- Consequence: we must build a counterfactual to be able to estimate the impact of the treatment
  - Counterfactual: if we observe Y_{1i} in the treatment group, we need to estimate Y_{0i}, the "value of the outcome variable for individual i in the treatment group had it been assigned to the control group instead"

Counterfactual-based program evaluation

In Practice

- Treated units: average outcome $\text{ave}\,(Y_{1i})$
- Control units used for the counterfactual, average outcome $\widehat{ave(Y_{0i})}$
- Impact of the intervention: $\text{ave}\,(Y_{1i}) - \widehat{ave(Y_{0i})}$
- Data: it is necessary to have information on treated and untreated units (administrative data/surveys)

Counterfactual-based program evaluation

Pros and Cons

- Measure the causal effect of an intervention: crucial for informed policy dialogue
- Challenge: identifying a convincing counterfactual
- Examines ”the effect of the cause” but does not explain ”the causes of an effect”
- After program implementation
- Can be only applied to some specific specific policies/interventions

Corollary 1

- Ex-post CIE can be used for Ex-ante IA
  - When available, ex-post evaluations of similar or related policies should support the ex-ante evaluation process, and in particular help identifying the predicted impact(s) of the policy proposal under scrutiny (e.g logic of intervention)

European Commission

Example

- Evaluation: Do hospitals make people healthier?
- The National Health Interview Survey (NHIS) includes the following 2 questions
  - "During the past 12 months, was the respondent a patient in a hospital overnight?"Can be used to identify recent hospital visitors.
  - "Would you say your health in general is excellent (1), very good (2), good (3), fair (4), poor (5)?". Answers to this question define health status (1 = excellent health to 5 = poor health)
- Impact of hospitals? ... we might be tempted to compare the health status of those who have been to the hospital to the health of those who have not been to the hospital.

Results based on a naive counterfactual

Group	Sample Size	Mean health status	Std. Error
Hospital	7774	2.79	0.014
No Hospital	90049	2.07	0.003

- The difference in the means is 0.71, a large and highly significant contrast in favor of the non-hospitalized.
- Taken at face value, this result suggests that going to the hospital makes people sicker.
- Are just hospitals full of other sick people who might infect us, and dangerous machines and chemicals that might hurt us?

Selection Bias

- Using individuals who were not hospitalized to measure the counterfactual is wrong
  - Selection bias: people who go to the hospital are probably less healthy to begin with. Moreover, even after hospitalization, people who have sought medical care are not as healthy, on average, as those who never get hospitalized in the first place (though they may well be better than they otherwise would have been).
- To measure the effect of the hospitalization, we need to find a good counterfactual, i.e. a proxy for the health status of those hospitalized, had they not been hospitalized

Corollary 2

- Staff need to be trained so as to understand how to read studies based on CIE and to judge about the ”quality” of the counterfactual used for the estimation of the impact
- CIE is essentially based on econometrics/statistics
  - Quantitative background needed

Why CRIE?

- Coming back to the previous question: Why CRIE?
- DG EMPL's concern :
  - Difficult economic times: pressing need for the EU to demonstrate the achievements of the European Social Fund (ESF) and other financial instruments
  - 2014-2020 programming period: common and ESF-specific provisions: involve a shift toward impact evaluation culture
- CRIE: to develop a capacity in house on CIE and help DG EMPL and MS when it is question of quantifying the impact of an intervention.

Presentation Outline

How to construct the counterfactual?

- Experimental approach
- Quasi-experimental methods
  - Difference in Differences (DiD)
  - Regression Discontinuity Design (RDD)
  - Propensity Score Matching (PSM)
  - Instrumental Variables (IV)
- Other methods exist

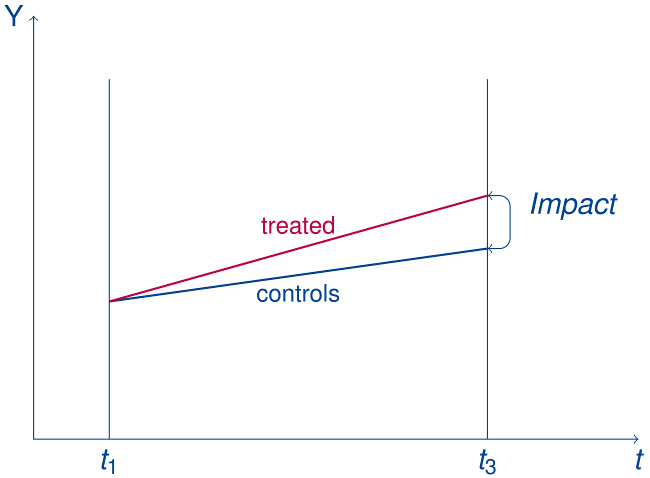
European Commission

Presentation Outline

Randomised experiments - the gold standard

- An essential objective of CIE methods is to identify situations where we can assume that the selection bias does not exist or find ways to correct for it
- Experimental approach = gold standard among CIE methods because the randomisation solves the selection bias
  - Since treatment and control groups are statistically equivalent, differences in outcomes between the groups can be attributed to the intervention.
  - Random assignment into treatment from the pool of experiment participants ensures that the treatment and control groups are "statistically" equivalent.

A graphical representation



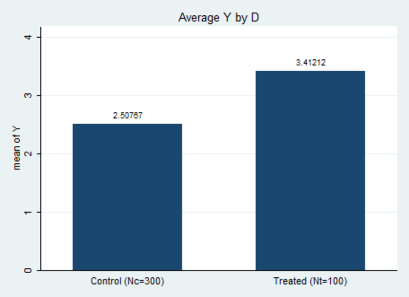
Practical problems with experiments

- **Ethical/legal issues**
  Randomised design can be politically unfeasible because it is hard to justify discrimination between individuals who can benefit and cannot, especially in poor areas.
  Counter-arguments:
  - Uncertainty about the results
  - Long-run benefits for the society
  - In the presence of limited resources, random assignment is fair and transparent
  - Some methods of randomisation can attenuate these issues: ”Randomized Promotion”

Job search assistance example

- Characteristics of the intervention
  - Job-search assistance and counseling.
  - Information is collected for $N = 400$ unemployed
  - Sampled individuals are randomly assigned to a treatment and control group.
  - Variable of interest Y is the change in income before-after (continuous).
- What we observe
  - Y_i, the change in log wage
  - D_i, = 1 if in the job-search assistance intervention, otherwise = 0
  - $X_i = age_i$, in the range [15,65]

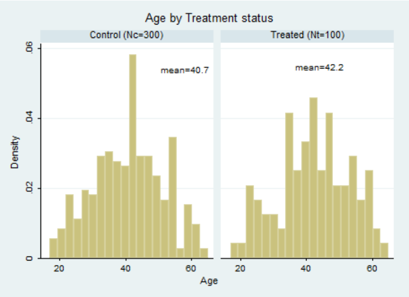
Example - perfect randomisation



$$Y_i = \alpha + \rho D_i + u_i$$

Y	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
D	.9044507	.0563235	16.06	0.000	.793722	1.015179
_cons	2.50767	.0281617	89.05	0.000	2.452305	2.563034

Example - perfect randomisation



$$H_0 : \text{ave}(age_i|D_i = 1) = \text{ave}(age_i|D_i = 0)$$

$$t = \left(\widehat{\text{ave}}(age_i|D_i = 1) - \widehat{\text{ave}}(age_i|D_i = 0)\right) / \sqrt{\left(\frac{\widehat{\text{var}}(age_i|D_i = 1)}{n_1} + \frac{\widehat{\text{var}}(age_i|D_i = 0)}{n_0}\right)} = 0.48$$

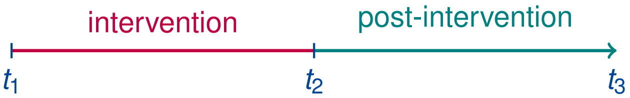
$|t| < 1.96$ implies H_0 cannot be rejected ($\alpha = 0.05$) $\Longrightarrow$ data supports perfect randomisation

Presentation Outline

Quasi-experimental methods

- When randomization not feasible, possibility to rely on statistical methods that attempt to mimic a randomization with the objective to solve the selection bias issue
  - From now on, participation to the program is not random. i.e., participation is based on some criteria or/and is voluntary
  - As a consequence, treated and controlled units differ along observable and unobservable characteristics: selection bias

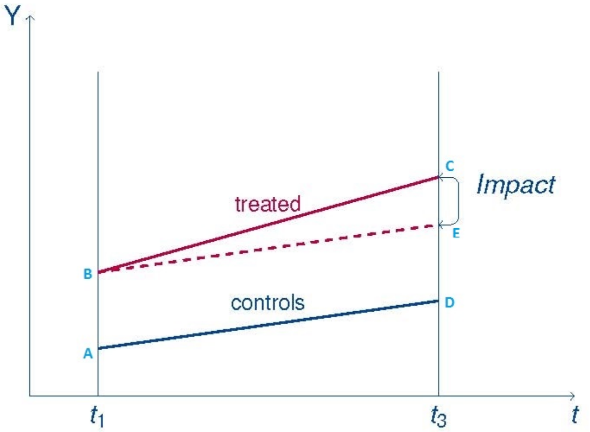
Differences in differences



$$\begin{pmatrix} Y_i \\ X_i \end{pmatrix}_{t_1} \qquad \begin{pmatrix} Y_i \\ X_i \\ D_i \end{pmatrix}_{t_3}$$

- Explore the time dimension of the data to define the counterfactual
  - Beneficiaries and non beneficiaries of the program are observed before and after the implementation of the policy

A graphical representation



Estimation

- The difference in the outcome variable before and after treatment in the control group is subtracted from the difference in outcome before and after treatment in the treated group

$$Impact = [CD] - [AB]$$

- Valid if the counterfactual trend is identical for treated and non treated groups:common trend assumption

Example: firm-simulation class

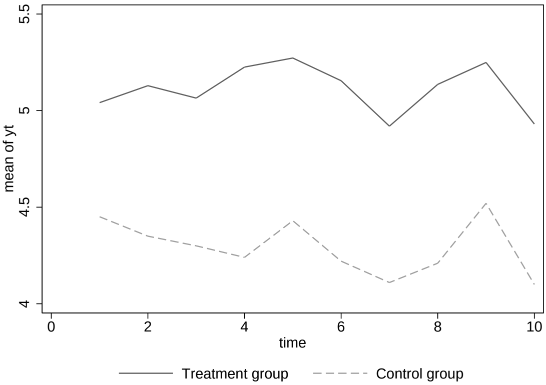
- Description. A 1-year firm-simulation class is implemented on some bachelor students. The outcome variable is a score on a entrepreneurial attitude and skills standardized test.
- Data. We observe the score before and after the intervention for participants and non-participants along with some individual characteristics.
- Impact. What is the effect of the intervention on the entrepreneurial skills of participants? Do they increase or remain unchanged?

Results

Simple difference-in-differences

	Treatment status		
	D=0; N=232	D=1; N=168	Difference
Time t_1	4.58	5.57	0.99
Time t_3	8.61	22.42	13.81
Change $t_3 - t_1$	4.03	16.85	12.82

Check the common trend assumption: Graphical check



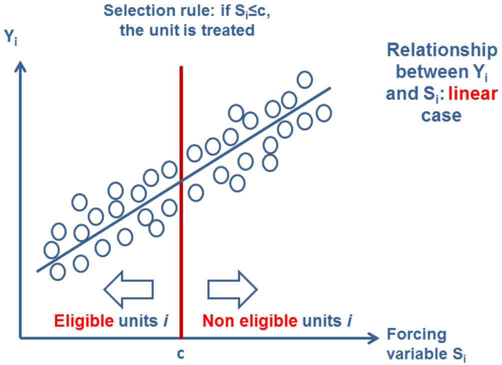
Regression Discontinuity Design

Eligibility for a program determined by a rule treatment assignment based on the value of a continuous variable S_i called the forcing variable : **Treatment:**

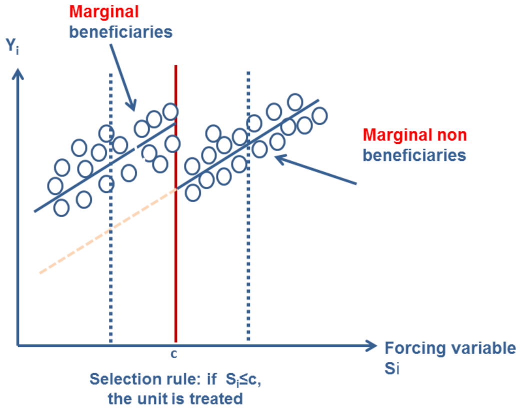
$$
\begin{aligned}
D_i &= 1 \ \mathit{if} \ S_i \leq c \\
D_i &= 0 \ \mathit{if} \ S_i > c
\end{aligned}
$$

European Commission

Without the intervention, linear case

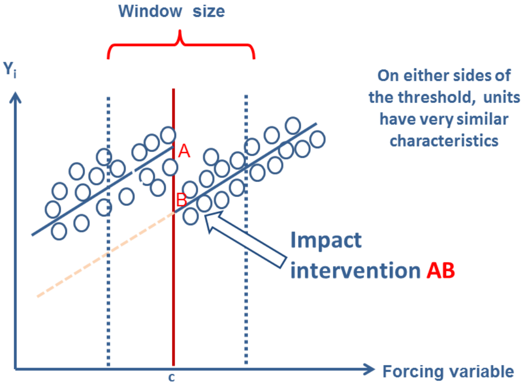


After the intervention



European Commission

Marginal beneficiaries and non beneficiaries



On either sides of the threshold, units have very similar characteristics

Selection rule: if $S_i \leq c$, the unit is treated

Impact in a snapshot

- Eligibility rule: On either sides of **c**, individuals have very **similar characteristics**, but some are treated and others are not
- Counterfactual: units above the cut-off who did not participate, i.e **marginal non beneficiaries**
- Effect of the intervention: **difference** in the average **performance** between **marginal beneficiaries** and **marginal non beneficiaries**

Example: Hunt-for-job support for non natives

Characteristics of the intervention

- To tackle unemployment among the non native population, unemployment offices in Spain put in place in 2010 hunt-for-job support.
- Only unemployed non-natives with a tertiary education, aged 40 or less can participate to this program.
- The duration of the support is 6 months.

Example: Hunt-for-job support for non natives

What we observe

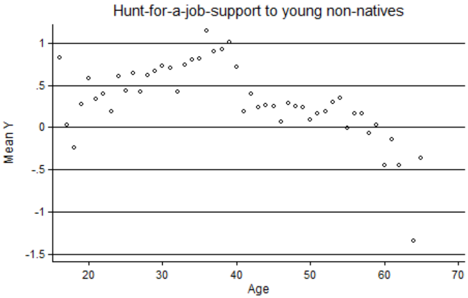
- Y_i, the average difference in monthly log income between 2012 and 2010 at the individual level
- D_i, =1 if in the support program, =0 otherwise
- S_i, the forcing variable, i.e. the age of the individual
- $c = 40$

What we want to measure

- Impact of the intervention on Y_i

Graphic interpretation

Impact of the intervention corresponds to the jump in the value of Y_i at $c = 40$



Training intervention: estimation of the impact

Participating in the hunt-for-a job support increases the monthly income by 66 percent

```
Linear regression                                      Number of obs =     400
                                                       F(  3,   396) =   56.64
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3122
                                                       Root MSE      =  .47317

------------------------------------------------------------------------------
             |               Robust
           Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           D |   .6634251   .0856094     7.75   0.000     .4951194    .8317309
         age |   .1181823   .0164631     7.18   0.000     .0858164    .1505482
        age2 |   -.001377   .0001916    -7.19   0.000    -.0017536   -.0010004
       _cons |  -2.268112   .3747427    -6.05   0.000    -3.004846   -1.531378
------------------------------------------------------------------------------
```

Corollary 3

- Ex-post CIEs need to be planned well in advance
  - CIE necessitates specific data arrangements which should be outlined in the IA
  - Advanced planning is required in order to first identify and then collect data on outcome indicators but also contextual information on both the treated and control groups, before and after the policy intervention.
  - Impact evaluation is also time and resource demanding. This should be clearly spelled out in the IA.

Presentation Outline

Conclusion

- If you are interested in assessing the impact of EU policy: CIE are useful, even for ex-ante IA