

TREATMENT EVALUATION METHODS :

matching methods, difference in
differences, regression discontinuity
design

Micheline Goedhuys
UNU-MERIT

DEIP-Montevideo March, 30 - April, 3 2009

Structure of presentation

- 1. The evaluation problem
- 2. Randomisation
- 3. Regression Discontinuity Design
- 4. PSM
- 5. DID

- Concluding remarks



1. The evaluation problem

- To measure the impact of a program, the evaluator is interested in the counterfactual question:
what would have happened to the beneficiaries ,...
if they had not had access to the program (eg. RD subsidy, ...)
- This is however not observed, unknown.
- We can only observe the performance of non-beneficiaries and compare it to the performance of beneficiaries.
- This comparison however is not sufficient to tell us the impact of the program, it presents rather correlations, no causality



1. The evaluation problem

- Why not?
- Because there may be a range of characteristics that affect both the possibility of participating in treatment AND performing well on the performance indicators
- Examples:

In RD subsidy programme aimed at raising private RD: Size, age, exporting, solvency... affecting RD expenditures and application for grant or subsidy

Ravaillon (2001): PROSCOL programme aimed at reducing poverty: poverty indicators as eligibility criteria (#household members, education HHH, dwelling)... affecting participation in PROSCOL program and years of schooling



1. The evaluation problem

- This means, 'being in the treatment group or not' is not the result of a random draw, but there is a selection into a specific group, along both *observable* and *non-observable* characteristics
- The effect of selection has to be taken into account if one wants to measure the impact of the program on the performance of the firms!!
- More formally....



1. The evaluation problem

Define:

Y^T = the average expenses in innovation by a firm in a specific year if the firm participates in the TDF and
 Y^C = the average expenses by the *same* firm if it does not participate to the program.

- Measuring the program impact requires a measurement of the difference ($Y^T - Y^C$) which is the effect of having participated in the program for firm i .
- This requires knowledge of the counterfactual outcome Y^C which is not empirically observable since a firm can not be observed simultaneously as a participant and as a non-participant.



1. The evaluation problem

- by comparing data on participating and non-participating firms, we can evaluate an average effect of program participation

$$\Delta = E[Y_{it}^T | D_i = 1] - E[Y_{it}^C | D_i = 0]$$

- Subtracting and adding $E[Y^C | D=1]$

$$\begin{aligned}\Delta &= E[Y_{it}^T | D_i = 1] - E[Y_{it}^C | D_i = 0] - E[Y_{it}^C | D_i = 1] + E[Y_{it}^C | D_i = 1] \\ &= E[Y_{it}^T - Y_{it}^C | D_i = 1] - E[Y_{it}^C | D_i = 0] + E[Y_{it}^C | D_i = 1]\end{aligned}$$



1. The evaluation problem

- Only if there is no selection bias, the average effect of program participation Δ will give an unbiased estimate of the program impact ($Y_i^T - Y_i^C$)
- There is no selection bias, if participating and non-participating firms are similar with respect to dimensions that are likely to affect both the outcome and participation in the treatment



2. The evaluation problem avoided: Randomisation

- Incorporating randomized evaluation in programme design
- The observations (firms, individuals) are randomly divided in two groups: treatment and control group
- Random assignment of treatment (participation in the program) would imply that there are no pre-existing differences between the treated and non-treated firms, selection bias is zero



- $ATT = (Y_i^T - Y_i^C) = E[Y_i^T | D=1] - E[Y_i^C | D=0]$

- In regression terms:

- $Y_i = \alpha + \beta D + \varepsilon_i$

where the OLS estimate of β is representing the ATT

- Hard to implement for certain types of policy instruments



3. Regression discontinuity design

- Applicable if probability of participation is a discontinuous function of one or more observable variables.
- Eg. Eligibility for micro credit - <one acre of land
- Assumption: within some range e of X^* , selection bias is zero
- $E[Y_i^C | D=1, X^*-e < X < X^*+e] = E[Y_i^C | D=0, X^*-e < X < X^*+e]$
- X is observable variable
- X^* is threshold for participation
- → use the observations just below the threshold as controls for those just above the threshold
- Assignment rules need to be strictly implemented!



4. PSM: Controlling for selection bias

Controlling for pre-existing differences

- Develop a statistically robust control group of non-beneficiaries
- identify comparable participating and non-participating firms, conditional on a set of observable variables X ,
- i.o.w.: control for the pre-existing observable differences
- using econometric techniques:
e.g. propensity score matching



4. Propensity score matching (PSM)

- If there is only *one dimension* that affects both treatment and outcome, it would be relatively simple to find pairs of matching firms :
eg. size of the firm (participation in R&D subsidy)
eg. Size of household head (PROSCOL)
- When treatment and outcome are determined by a *multidimensional* vector of characteristics (size, industry, location...), this becomes problematic.
- Find pairs of observations that have equal or similar *probability* of being treated



4. PSM: idea

- Using *probit* or *logit* analysis on the whole sample of beneficiaries and non-beneficiaries, we calculate the probability (P) or propensity that an individual observation participates in a program, the *propensity score* :
- $P(D=1)=F(X)$
X= vector of observable characteristics
- Purpose: to find for each participant (D=1) at least one program non-participant that has equal/very similar chance of being participant, which is then selected into the control group
- It reduces the multidimensional problem of several matching criteria to one single measure of *distance*



4. PSM : Measures of proximity

- One-to-one matching: Compare outcome of firm i to the outcome of the most observably similar unit j .

. *nearest neighbour:*

$$|p_i - p_j| = \min_{k \in \{D=0\}} \{|p_i - p_k|\}$$

all are matched, sometimes poorly

. *caliper matching:*

$$\delta \geq |p_i - p_j| = \min_{k \in \{D=0\}} \{|p_i - p_k|\}$$

some may remain unmatched but quality of matching improves



4. PSM : Measures of proximity

One-to-several: compare the outcome of treated unit i to the mean outcome of a few control units

- *radius* matching: match to all control units falling within a radius r from p_i

$$| p_i - p_j | < r$$

- *Nearest x neighbours*



4. PSM : Measures of proximity

- Weighting methods: compare the outcome of firm i to a weighted average outcome of more (possibly all) non-treated units, where the weight given to non-treated unit j is in proportion to the proximity of their propensity scores (Gaussian, Epanechnikov)

$$\hat{y}_i = \sum_{j \in \{D=0\}} w_{ij} y_j$$

- Common support: treated units whose p_i is larger than the largest p_j in the non-treated pool are left unmatched



STEPS

1. Create control sample
2. Pool control and treated sample
3. Logit, create predicted values of the probability for participation
4. Exclude some controls (too low propensity scores) and maybe some treated (prop. scores too high to find control)
5. Match, using a criterion
6. Calculate difference in outcome for treated and the (weighted average) outcome of control(s) = program gain
7. Calculate mean program gain (stratified)



4. PSM : example

- Estimating the impact (Average effect of Treatment on Treated):

$$ATT = E[E(Y1 | D = 1, p(x)) - E(Y0 | D = 0, p(x)) | D=1]$$

Y is the impact variable

$D = \{0,1\}$ is a dummy variable for the participation in the program,

x is a vector of pre-treatment characteristics

$p(x)$ is the propensity score.



Example

Impact of ADTEN (Brazil) on (private) R&D intensity

Single difference in 2000

$[(RD/sales\ 2000\ beneficiaries - RD/sales\ 2000\ control)]$ after PSM

92 observations each

- beneficiaries 1.18%
- Control group 0.52%
- Difference: 0.66%
- positive and significant impact, net of subsidy



5. Difference in difference (DID)

The treated and control group of firms may also differ in *non-observable* characteristics, eg management skills.

- If panel data are available (data of pre-treatment and post-treatment time periods) the impact of unobservable differences and time shocks can be neutralised by taking the difference-in-differences of the impact variable.

$$ATT = [E(Y_1 | D = 1) - E(Y_0 | D = 1)] - [E(Y_1 | D = 0) - E(Y_0 | D = 0)]$$



- Important assumption: unobservables do not change over time;
- in absence of the treatment, the two groups would have followed parallel trends
- In case of DID, the impact variable is a growth rate.
- DID can be combined with PSM

$$ATT = [E(Y_1 | D = 1, p(x)) - E(Y_0 | D = 1, p(x))] - [E(Y_1 | D = 0, p(x)) - E(Y_0 | D = 0, p(x))]$$



STEPS

- Need for a baseline survey of participants and non-participants
- Need for at least one follow up survey, highly comparable to baseline survey
- Calculate mean difference between the 'after' and 'before' values of the outcome indicator, for treatment and control group
- Calculate the difference between these mean differences = program impact



Example of results

Impact of FONTAR-ANR (Argentina)
on (public+private) R&D intensity (=R&D
expenditures/sales)

Difference in difference with PSM

37 observations each

[(RDint. after ANR beneficiaries - RD/sales before ANR ben.) -
RD/sales after ANR control - RD/Sales before ANR control]]

- Beneficiaries $(0.20 - 0.08) = 0.12$
- Control group $(0.15 - 0.22) = -0.07$
- DID 0.19

positive and significant impact, GROSS of subsidy



Concluding remarks

A good evaluation design start by a clear rationale and short, medium and long run expected outcomes Y;

The collection of primary data on the programs' beneficiaries and on a group of comparable non-beneficiaries are essential, periodic data collection, baseline and follow up provide more correct insights;

The evaluator has a wide range of tools available, each providing different results. Some convergence in results along different techniques is desirable.



Concluding remarks

- The data needs of this type of evaluation are evident
- The merger and accessibility of several data sources create unprecedented opportunities for the evaluation and monitoring of policy instruments

Thank you!

