THE WORLD BANK

IMPACT EVALUATION

# Technical Session VI
# Matching Techniques

Beijing, China

2009

Human Development Network

Spanish Impact Evaluation Fund

# When can we use matching?

- What if the assignment to the treatment is done not randomly, but on the basis of observables??
  - *This is when matching methods come in! Matching methods allow you to construct comparison groups when the asignment to the treatment is done on the basis of observable variables.*

- Warning: matching STILL does not allow to control for selection bias that arises when the assignment to the treatment is done on the basis of non-observables.

- Intuition: the comparison group needs to be as similar as possible to the treatment group, in terms of the observables before the start of the treatment.
- The method assumes there are no remaining unobservable differences between treatment and comparison groups.

# Key question….

What is

The effect of treatment on the treated

When the assignment to the treatment

Is based on observable variables?

# Unconfoundedness and selection on observables

- Let $X$ denote a matrix in which each row is a vector of pre-treatment observable variables for individual $i$.

- Definition Unconfoundedness

  *Assignment to treatment is unconfounded given pre-treatment variables X if*

  $$Y_1, Y_0 \perp D \mid X$$

- Note: unconfoundedness is equivalent to saying that:
  - within each cell defined by $X$: treatment is random
  - the selection into treatment depends only on the observables $X$

# Average effects of treatment on the treated assuming unconfoundedness given X

- Intuition:
  - Estimate the treatment effect within each cell defined by X
  - Take the average over the different cells

- Math: in your handouts in Annex 1

# Strategy for estimating average effect of treatment on the treated – selectión on observables

- Unconfoundedness suggests the following strategy for the estimation of the average treatment effect $\delta$
  - stratify the data into cells defined by each particular value of $X$
  - within each cell (i.e. conditioning on $X$) compute the difference between the average outcomes of the treated and the controls
  - average these differences with respect to the distribution of $X$ in the population of treated units.

- Is this strategy feasible?

# Is our strategy feasible?
# …. the dimensionality problem

- This may not be feasible when…
  - the sample is small,
  - the set of covariates is large
  - many of the covariates have many values, or are continuous

- This is what we call… the dimensionality problem

# The dimensionality problem

- Examples:
  - How many cells do we have with 2 binary *X* variables? And with 3 binary *X* variables? And with K binary *X* variables?
  - How about if we have 2 variables that take on 7 values each?

- As the number of cells grows, we'll get "lack of commom support"
  - cells containing only treated observations
  - cells containing only controls.

# An alternative to solve the dimensionality problem

- Rosenbaum and Rubin (1983) propose an equivalent and feasible estimation strategy based on the concept of *Propensity Score*

  - The propensity score allows to convert the multidimensional setup of matching into a one-dimentional setup

  - In that way, it allows to reduce the dimensionality problem.

# Matching based on the Propensity Score

- **Definition** Propensity Score: *The propensity score is the conditional probability of receiving the treatment given the pre-treatment variables:*

$$p(X) = Pr\{D = 1|X\} = E_X\{D|X\}$$

- **Lemma 1:** *If* $p(X)$ *is the propensity score, then* $D \perp X \mid p(X)$

  "Given the propensity score, the pre-treatment variables are balanced between beneficiaries and non- beneficiaries"

- **Lemma 2:** $Y_1, Y_0 \perp D \mid X \Rightarrow Y_1, Y_0 \perp D \mid p(X)$

  "Suppose that assignment to treatment is unconfounded given the pre-treatment variables X. Then assignment to treatment is unconfounded given the propensity score p(X)."

# Does the propensity score approach solve the dimensionality problem?

- Yes!

- The *balancing property* of the propensity score (Lemma 1) ensures that:
  - observations with the same propensity score have the same distribution of observable covariates independently of treatment status;
  - for a given propensity score: assignment to treatment is "random" and therefore treatment and control units are observationally identical on average.

# Implementation of the estimation strategy

This suggests the following strategy for the estimation of the average treatment effect $\delta$

- Step 1: Estimate the propensity score
  - Eg. With a logit function, see Annex 3
  - This step is necessary because the "true" propensity score is unknown and therefore the propensity score has to be estimated.

- Step 2: Estimate the average treatment effect given the propensity score

# When is propensity score matching appropriate?

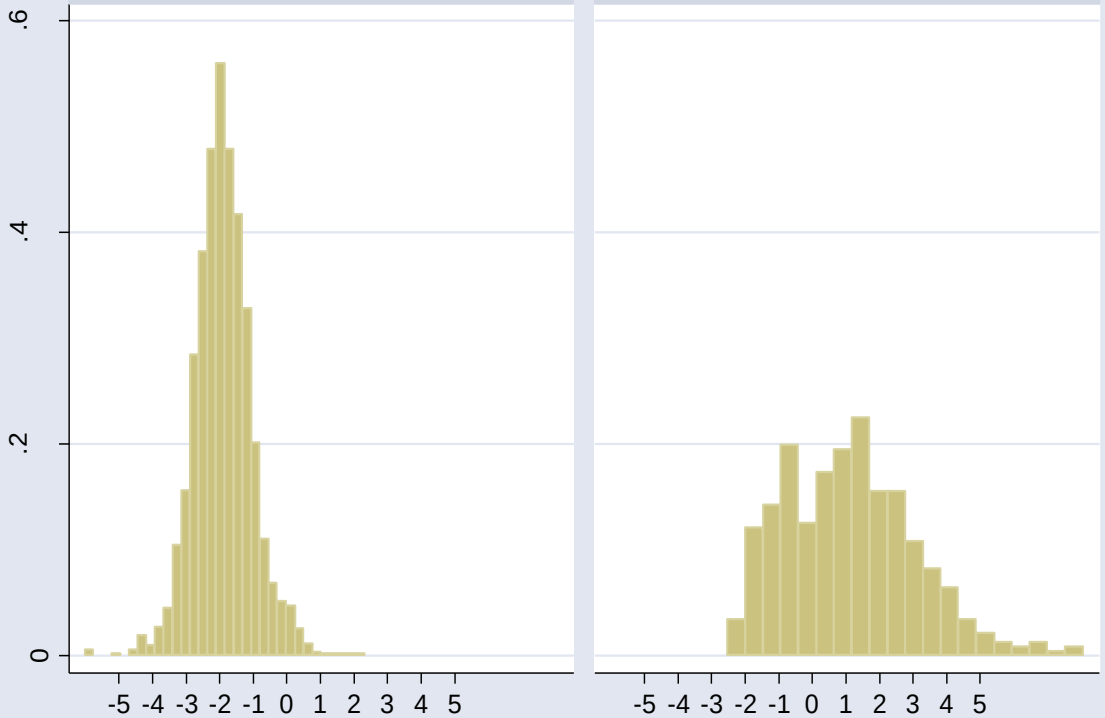
- Idea behind Propensity score matching: estimation of treatment effects requires a careful matching of treated and controls.
- If treated and controls are very different in terms of observables this matching is not sufficiently close and reliable or it may even be impossible.
- The comparison of the estimated propensity scores across treated and controls provides a useful diagnostic tool to evaluate how similar are treated and controls, and therefore how reliable is the estimation strategy.

# So you want propensity scores to be the "same" for treatments and controls…

- The range of variation of propensity scores should be the same for treated and controls.
  - count how many controls have a propensity score lower than the minimum or higher than the maximum of the propensity scores of the treated.
  - And vice versa
- Frequency of propensity scores is the same for treated and control.
  - draw histograms of the estimated propensity scores for the treated and controls.
  - The bins correspond to the blocks constructed for the estimation of propensity scores.

# An example of common support issues

**Figure A1:**
**Propensity Scores For EiC Phase 1 and non-EiC schools**




Source Machin, McNally, Meghir,
**EXCELLENCE IN CITIES: EVALUATION OF AN EDUCATION POLICY IN DISADVANTAGED AREAS**

# Implementation of the estimation strategy

Remember we're discussing a strategy for the estimation of the average treatment effect on the treated, called $\delta$

- **Step 1: Estimate the propensity score (see Annex 3)**
- **Step 2: Estimate the average treatment effect given the propensity score**
  - match treated and controls with exactly the same (estimated) propensity score
  - compute the effect of treatment for each value of the (estimated) propensity score
  - obtain the average of these conditional effects

# Step 2: Estimate the average treatment effect given the propensity score

- This is unfeasible in practice because it is rare to find two units with exactly the same propensity score.
- The closest we can get to an exact matching is to match each treated unit with the *nearest* control in terms of propensity score
- "Nearest" can be defined in many ways. These different ways then correspondent to different ways of doing matching….
  - Stratification on the Score;
  - Nearest neighbor matching on the Score;
  - Radius matching on the Score;
  - Kernel matching on the Score;
  - Weighting on the basis of the Score.

# References


- Dehejia, R.H. and S. Wahba (1999), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs", *Journal of the American Statistical Association*, 94, 448, 1053-1062.
- Dehejia, R.H. and S. Wahba (1996), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs", Harvard University, Mimeo.
- Hahn, Jinyong (1998), " ON the role of the propensity score in efficient semiparamentric estimation of average treatment effects", *Econometrica*, 66,2,315-331.
- Heckman, James J. H. Ichimura, and P. Todd (1998), " Matching as an econometric evaluation estimator ", *Review of Economic Studies*, 65, 261-294.
- Hirano, K., G.W. Imbens and G. Ridder (2000), "Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score", mimeo.
- Rosenbaum, P.R. and D.B. Rubin (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika* 70, 1, 41–55.
- Vinha, K. (2006) A primer on Propensity Score Matching Estimators" Documento CEDE 2006-13, Universidad de los Andes

# Annex 1: Average effects of treatment on the treated assuming unconfoundedness given X

If we are willing to assume unconfoundedness:

$$E_i\left\{Y_0\left(u_i\right)\mid D_i=0, X\right\}=E_i\left\{Y_0\left(u_i\right)\mid D_i=1, X\right\}=E_i\left\{Y_0\left(u_i\right)\mid X\right\}$$

$$E_i\left\{Y_1\left(u_i\right)\mid D_i=0, X\right\}=E_i\left\{Y_1\left(u_i\right)\mid D_i=1, X\right\}=E_i\left\{Y_1\left(u_i\right)\mid X\right\}$$

Using these expressions, we can define for each cell defined by $X$

$\delta_X$ =average treatment effect on the treated in cell defined by $X$

$$=E_i\{\Delta_i\mid D_i=1, X\}$$

$$=E_i\left\{Y_1\left(u_i\right)-Y_0\left(u_i\right)\mid D_i=1, X\right\}$$

$$=\underbrace{E_i\left\{Y_1\left(u_i\right)\mid D_i=1, X\right\}}_{\text{can measure sample analog}}-\underbrace{E\left\{Y_0\left(u_i\right)\mid D_i=1, X\right\}}_{\text{can NOT measure sample analog}}$$

$$=E_i\left\{Y_1\left(u_i\right)\mid D_i=1, X\right\}-\underbrace{E_i\left\{Y_0\left(u_i\right)\mid D_i=0, X\right\}}_{\text{can measure sample analog}}$$

# Annex 1: Average effects of treatment on the treated assuming unconfoundedness given X

Now what is the relation between

$\delta$ "average treatment effect on the treated"... and....

$\delta_X$ "average treatment effect on the treated within cell defined by $X$ "?

$\delta$ = average treatment effect on the treated

$= E_i \{\Delta_i \mid D_i = 1\}$

$\Downarrow$ by the law of iterated expectations

$= E_i \{E_X [\Delta_i \mid D_i = 1, X]\}$

$= E_X \{E_i [\Delta_i \mid D_i = 1, X]\}$

$= E_X \{\delta_X\}$

$= E_X$ {average treatment effect on the treated within cell defined by X}

# Annex 2: Average effects of treatment and the propensity score

So let's match treatments and controls

on the basis of the propensity score p(X) instead of X.

$$E_i\left\{Y_0(u_i)\mid D_i=0, p(X_i)\right\}=E_i\left\{Y_0(u_i)\mid D_i=1, p(X_i)\right\}=E_i\left\{Y_0(u_i)\mid p(X_i)\right\}$$

$$E_i\left\{Y_1(u_i)\mid D_i=0, p(X_i)\right\}=E_i\left\{Y_1(u_i)\mid D_i=1, p(X_i)\right\}=E_i\left\{Y_1(u_i)\mid p(X_i)\right\}$$

Using these expressions, we can define f cell defined by $p(X)$

$\delta_{p(X)}$ =average treatment effect on the treated in cell defined by $p(X)$

$$=E_i\{\Delta_i\mid D_i=1, p(X)\}$$

$$=E_i\left\{Y_1(u_i)-Y_0(u_i)\mid D_i=1, p(X)\right\}$$

$$=\underbrace{E_i\left\{Y_1(u_i)\mid D_i=1, p(X)\right\}}_{\text{can measure sample analog}}-\underbrace{E\left\{Y_0(u_i)\mid D_i=1, p(X)\right\}}_{\text{can NOT measure sample analog}}$$

$$=E_i\left\{Y_1(u_i)\mid D_i=1, p(X)\right\}-\underbrace{E_i\left\{Y_0(u_i)\mid D_i=0, p(X)\right\}}_{\text{can measure sample analog}}$$

# Annex 2: Average effects of treatment and the propensity score

Now what is the relation between

$\delta$ "average treatment effect on the treated"... and....

$\delta_{p(X)}$ "average treatment effect on the treated within cell defined by $p(X)$"?

$\delta$ = average treatment effect on the treated

$= E_i\{\Delta_i \mid D_i = 1\}$

$\Downarrow$ by the law of iterated expectations

$= E_i\left\{E_{p(X)}\left[\Delta_i \mid D_i = 1, p(X)\right]\right\}$

$= E_{p(X)}\left\{E_i\left[\Delta_i \mid D_i = 1, p(X)\right]\right\}$

$= E_{p(X)}\left\{\delta_{p(X)}\right\}$

$= E_{p(X)}\{\text{treatment effect on the treated within cell defined by } p(X)\}$

# Annex 3: Estimation of the propensity score

- Any standard probability model can be used to estimate the propensity score, e.g. a logit model:

$$Pr\{D_i \mid X_i\} = \frac{e^{\lambda h(X_i)}}{1 + e^{\lambda h(X_i)}} \qquad (16)$$

where $h(X_i)$ is a function of covariates with linear and higher order terms.

# Estimation of the propensity score

- Which higher order terms do you include in $h(X_i)$?
  - This is determined solely by the need to obtain an estimate of the propensity score that satisfies the *balancing property*.
- The specification of $h(X_i)$
  - is more parsimonious than the full set of interactions between observables X
  - Though not too parsimonious: it still needs to satisfy the *balancing property*
- Note: the estimation of the propensity scores does not need a behavioral interpretation.

# An algorithm for estimating the propensity score

1. Start with a parsimonious logit or probit function to estimate the score.
2. Sort the data according to the estimated propensity score (from lowest to highest).
3. Stratify all observations in blocks such that in each block the estimated propensity scores for the treated and the controls are not statistically different:
    a. start with five blocks of equal score range {0 - 0.2, ..., 0.8 - 1}
    b. test whether the means of the scores for the treated and the controls are statistically different in each block
    c. if yes, increase the number of blocks and test again
    d. if no, go to next step.

# An algorithm for estimating the propensity score (continued)

4. Test that the *balancing property* holds in all blocks for all covariates:
   a) for each covariate, test whether the means (and possibly higher order moments) for the treated and for the controls are statistically different in all blocks;
   b) if one covariate is not balanced in one block, split the block and test again within each finer block;
   c) if one covariate is not balanced in all blocks, modify the logit estimation of the propensity score adding more interaction and higher order terms and then test again.

Note that in all this procedure the outcome has no role

Use the STATA program pscore.ado

Download at http://www.iue.it/Personal/Ichino/Welcome.html