



Propensity Score Matching

Racha Ramadan

racha.ramadan@feps.edu.eg

**Training on Applied Micro-Econometrics and Public
Policy Evaluation**

2024

What would have happened with the participants if they had not participated to the program?!

Average Treatment Effect?



- To answer such question, we need to compare the situation of the beneficiaries of a given program (treatment group) to non-beneficiaries (counterfactual group).
- In a context of non-randomized, non-experimental design, treatment and control groups are likely to have different distributions for non-treatment variables (NORC, 2012), this may yield to selection bias.

What to do in case of non-randomization?!

Propensity Score Matching (PSM)

- PSM consists of finding a proper counterfactual group by matching a participant to a non-participant with similar pre-intervention characteristics (X).
- PSM compares treatment and control observations that are similar (along observable characteristics). i.e. We only keep control observations that are "*close*" to the treatment observations we want to match.
- This method allows mimicking randomized assignment.

Main Question to be asked:

Which characteristics does one use and what weight to put on each of them?

□ Basic Theorem of PSM

- PSM assumes that eliminating selection based on observables will reduce overall bias.
- Main Question to be asked:
Which characteristics should we use and what weight to put on each of them?
- Rosenbaum and Rubin (1983) showed that instead of having to match on a multitude of dimensions in a vector of observable characteristics X , it is only necessary to match on a single dimension $P(X)$, which is **the propensity score**.
- A treatment and control observations that are "*close*" in the propensity score space, will also be "*close*" along the various elements X .

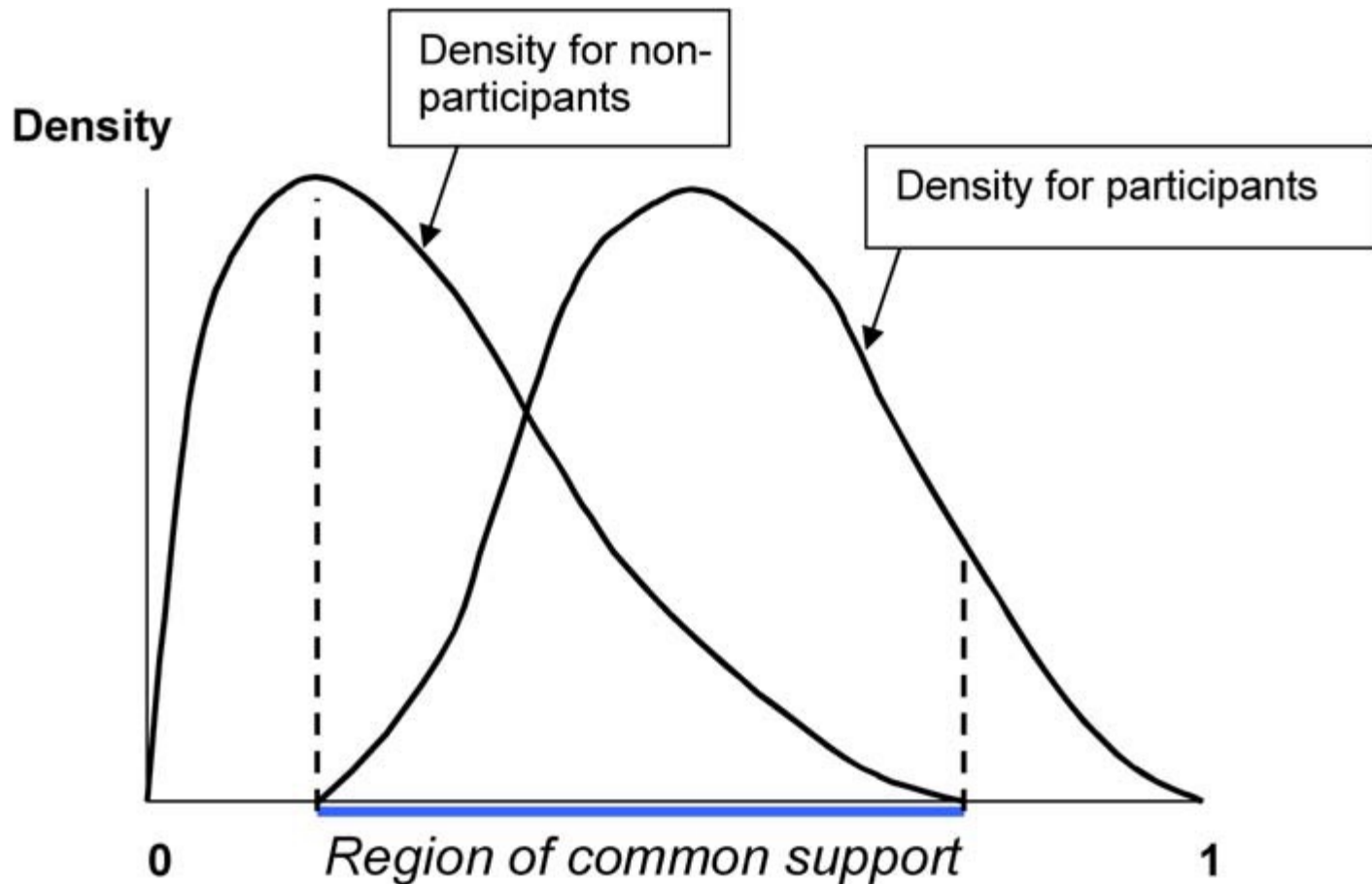
A Propensity Score = $P(X) = \Pr(T=1 | X)$

- where X is a vector of pre-exposure characteristics.
- $T=1$ means the observation is treated.
- $P(X)$ be estimated using probit or logit
- Treatment units are matched to control units with similar values of $P(X)$.

□ Steps:

1. Use predicted probabilities of participation from a probit or logit model
2. Matching is done based on the distance between each treatment observation and control observations in its neighborhood, using the propensity score as a metric of distance.
3. The outcome of each treatment is compared to that of its matched to detect whether the treatment has a statistically significant effect.

□ Region of Common Support



Region of common support for propensity score between participants and non-participants must be large enough to find an adequate comparison group



Estimates of treatment impact using PSM will depend on:

- the variables included in the equation used to estimate propensity score
- the specification of that equation.

What variables go in X?

- Any variable that can affect program placement, including the pre-intervention level of the outcome variable.
- Can be obtained through interviews.
- If one cannot find good observable variables to explain who participates and who doesn't, we will do a poor job eliminating selection bias.

How to decide which observations are close matches?

- As propensity score $P(X)$ is a continuous variable, no two observations will have the exact estimated score.
- There are many ways to undertake the matching:
 - A. Simple Matching Methods
 - B. More complicated Matching Methods.

A. Simple Matching Methods

One-to-one matching	<ul style="list-style-type: none">- Match each treatment observation with “nearest” control observation.- We can impose a maximum distance, through a “caliper”.- If the closest neighbor is far away. A caliper is a tolerance level on the maximum propensity score distance)
Nearest n-neighbors	<ul style="list-style-type: none">- Uses the average of the nearest n-neighbors among comparison observations.- We can an also impose a caliper

B. More complicated Matching Methods

Kernel matching <i>(KM)</i>	<ul style="list-style-type: none">- KM is non-parametric matching estimator that uses weighted averages of all individuals in the control group to construct the counterfactual outcome.-Weights depend on the distance between each individual from the control group and the participant observation for which the counterfactual is estimated.-When applying KM one has to choose the kernel function and the bandwidth parameter

□ Average Treatment Effect

- The impact of a given program is measured by:

$$ATE = \frac{1}{N_T} \sum_{j=1}^{N_T} (y_j^T - y_j^C)$$

Where Y_j^C is the value of outcome of the control matched observation.

- **Nearest n-neighbors:**

$$ATE = \frac{1}{N_T} \sum_{i=1}^{N_T} (Y_i^T - \sum_{j \in C(i)} w_{ij} Y_j^C)$$

Where:

- $C(i) = \min_j \|p_i - p_j\|$

- $w_{ij} = \frac{1}{N_i^C}$ if $j \in C(i)$

- N_i^C = Number of controls matched with observation i (treated)

• Kernel Matching Method

$$ATE = \frac{1}{N_T} \sum_{i=1}^{N_T} \left(Y_i^T - \frac{\sum_{j \in C(i)} Y_j^C G\left(\frac{p_j - p_i}{h_n}\right)}{\sum_{k \in C(i)} G\left(\frac{p_k - p_j}{h_n}\right)} \right)$$

Where:

○ $G(\cdot)$ is a kernel function and h_n is a bandwidth parameter

○ $\frac{\sum_{j \in C(i)} Y_j^C G\left(\frac{p_j - p_i}{h_n}\right)}{\sum_{k \in C(i)} G\left(\frac{p_k - p_j}{h_n}\right)}$: is a consistent estimator of the counterfactual

outcome Y_{oi}



Notes on Estimating Propensity Score

- Although effect of treatment on outcome is non-parametric, estimation of propensity score itself depends on a parametric specification (usually logit or probit).
- There are many modeling decisions that can affect the results, including:
 1. Specification of propensity score equation: *What variables to include? How many interactions to include?*
 2. Matching method; caliper or bandwidth to use.
- **Sensitivity Analysis:** test sensitivity of results to these decisions. If results are not robust to these changes, this should raise a question mark about their reliability.



Testing how good the match is!

- Need to test the extent to which observable characteristics are balanced in the matched sample, using a t-test of the difference in means of the covariates across matched samples.

None of the test should be statistically significant

- Should examine the effect of the matching method on the area of common support.
- The closer we try to make the match, the more likely that some treatment observation won't find matches, which reduce the efficiency of the model.



Standard Errors

- As propensity scores are predicted, the estimated variance of the treatment effect should also include the variance due to the estimation of the propensity score, the imputation of the common support, and possibly also the order in which treated individuals are matched.
- This yields to more variation than the normal sampling variation.
- Possible Solution: bootstrapping to estimate the correct standard errors.



PSM compared to Social Experiments

- In social experiments **Propensity Score= constant**. i.e. everyone has same probability of participating.
- PSM tries to match treatment and control groups based on them having an equal probability of participation (using the probability predicted from observables).
- But probability of participation can also be affected by unobservable characteristics raising concerns about remaining selection bias in PSM methods.

Propensity Score Weighting

Propensity Score Weighting

- Propensity scores can also be used to weight observations; like a sample weight.
- Specifically use inverse probability of treatment weighting (IPTW).
- This approach can be used in contexts where models are non-linear. Example: Propensity score weights in survival analysis (Austin, 2011)

□ Creating propensity score weights

- Create propensity scores for matching:
 - $P(X) = \Pr (T_i=1 \mid X_i)$ where X is a vector of pre-exposure characteristics
 - Predict probabilities of participation using logit or probit= \hat{P}_i

- Use propensity scores to create weights (w_i):

$$w_i = \frac{T_i}{\hat{p}_i} + \frac{1-T_i}{1-\hat{p}_i}$$

- If treated: $w_i = \frac{1}{\hat{p}_i}$
- If not treated: $w_i = \frac{1}{1-\hat{p}_i}$
- High weights for those who were unlikely to receive treatment but did and those who were likely to receive treatment but did not.

□ Estimates of treatment effects

- There are a number of ways to use IPTW directly to estimate treatment effects
- Can estimate average treatment effect (ATE) as:

$$ATE = \frac{1}{N} \left(\sum_1^N \frac{T_i Y_i}{\hat{p}_i} - \sum_1^N \frac{(1 - T_i) Y_i}{(1 - \hat{p}_i)} \right)$$



Checking weights

- When weights are applied to treatment and control groups, they should then have statistically indistinguishable group means
- Weights may be very inaccurate/unstable for subjects who are
 - very unlikely to get treatment and did
 - very likely to get treatment and did not
- Can stabilize weights by multiplying weights by mean probability of treatment received (T or C) propensity score.
- Can also trim (use region of common support)

Application

The influence of Bolsa Familia conditional cash transfer program on child labor in Brazil

Paloma Santana Moreira Pais, Felipe de Figueiredo Silva
and Evandro Camargos Teixeira

International Journal of Social Economics

Vol. 44 No. 2, 2017

Application: Motivation

- The insertion of children into the job market early perpetuates the cycle of poverty and low schooling within the society.
- Poverty is the most cited cause of child labor in the literature.
- Researchers have been studying the effect of several policies (as CCT) on child labor to reduce poverty in the long run and to increase the human capital of children in underprivileged families.
- They contribute to the literature by evaluating the influence of the Bolsa Familia CCT program on the time allocation of child labor.
- Child labor is represented by the time children spent on work

Application: Data and Methodology

- They define a set of variables that determine the participation in the Bolsa Familia program
- They set a control group based on these key variables to create a test sample.
- National Household Sample Survey (PNAD, 2006): Sample of families who are suitable for receiving BF transfer. They had individuals aged between 0 and 17 years and who had a net income per capita (excluding income from government transfers) of up to US\$139.53.

Application: X's?

- The variables qualifying participation in BF program and the demographic characteristics used in the Logit model include:
 - ✓ Children's Characteristics: age (in years) and a dummy for the color (1 if the child is black and 0 for the others)
 - ✓ Parents' characteristics and family characteristics: father's schooling and mother's schooling (in years); the net income per capita.
 - ✓ Other dummies: for families headed by women, for families living in rural areas and for families living in metropolitan regions.
 - ✓ Number of children aged 0 to 5 years in the family, and the number of children of 6 to 17 years.
- The labor response variable was the number of hours of child labor (*hours_inf*).

Results(1): Estimation of PSM

	Coefficient	SE	Z	p-value	Confidence interval	
School Attendance	0.419767	0.037709	11.13	0.000000	0.345858	0.493675
Mother's schooling	-0.051696	0.003991	-12.9	0.000000	-0.05952	-0.04388
Father's schooling	-0.003636	0.004091	-0.89	0.374000	-0.01165	0.004383
Age	-0.004092	0.004002	-1.02	0.307000	-0.01194	0.003751
Number of children 1	0.067546	0.018778	3.60	0.000000	0.030742	0.104351
Number of children 2	0.308758	0.012028	25.67	0.000000	0.285183	0.332333
Head_female	-0.200266	0.195703	-1.02	0.306000	-0.58384	0.183305
Rural	0.483943	0.034456	14.05	0.000000	0.416410	0.551477
Metropolis	-0.605106	0.036018	-16.8	0.000000	-0.67570	-0.53451
Family income	-0.002654	0.000185	-14.3	0.000000	-0.00302	-0.00229
Black	0.037256	0.059414	0.63	0.531000	-0.07919	0.153704
Constant	0.150244	0.062205	2.42	0.016000	0.028323	0.272164

Table II.
Estimate of the Logit model for the group of families of Brazil with net income per capita of the transfers up to R\$300.00, in 2006

Source: Results of the work

- The father's schooling, the age of the child, the fact the family is headed by a female, and the color of the family (black) are not relevant variables in determining the probability of whether the families would be beneficiaries.
- Higher school attendance increases the probability of a family participating in this program. While higher income per capita level decreases the probability of a family participating in it.
- Highly educated mother or when the family lives in a metropolitan (urban) area, the probability of participating in the program decreases
- When a household has a greater number of children or is located in a rural location, the probability of participation increases

Results (2): Estimation of Treatment Effect

- They contrast the treatment group with a control group of similar-characteristic non- participants

Response variable	Treatment	Control	Estimator of the ATT (nearest neighbor)
Hours_inf	11,869	5,250	0.525***

Note: ***Significant at 1 percent
Source: Results of the work

Table V.
Effect of the average treatment upon the treated one for the variable hours of child labor (hours_inf) for Brazil, in 2006

Participation in BF program increased the children's and teenagers' time allotment for labor by 52.5%.

References

- Slides of Ragui Assaad and Caroline Krafft.
- Becker, O. and Ichino, A. (2002) “Estimation of Average Treatment Effects based on Propensity Score”. The Stata Journal 2, Number 4, pp. 358-377
- Caliendo, M. and Kopeinig, S. (2005) “Some Practical Guidance for the Implementation of Propensity Score Matching”. IZA DP No. 1588.
- Khandker, S.R., Koolwal, G.B. and Samad, H.A. (2009) “Handbook on Impact Evaluation Quantitative Methods and Practices”. World Bank.
- Pais, P.S., Silva, F. and Camargos, E. (2017) “The influence of Bolsa Familia conditional cash transfer program on child labor in Brazil”. International Journal of Social Economics Vol. 44 No. 2. pp. 206-221.
- Trujilo, A.J., PORTILLO, J.E and VERNON, J.A. (2005) “The Impact of Subsidized Health Insurance for the Poor: Evaluating the Colombian Experience Using Propensity Score Matching”. International Journal of Health Care Finance and Economics.
- Wooldridge, J. (2002) “Econometric Analysis of Cross Section and Panel Data”. The MIT Press. Cambridge , Massachusstes. London, England.