Evaluation of Public Policy

Dott.ssa Rossella Iraci Capuccinello

Università degli Studi di Ferrara and Lancaster University,
r.iracicapuccinello@lancaster.ac.uk

a.a. 2018-2019

Causal Effect

We face 2 possible problems:

- Fundamental evaluation problem $\Longrightarrow$ no counterfactual
- Selection bias

Propensity Score Matching

Average Treatment Effect on the Treated

$$Y_i \equiv Y_i(D_1) = \begin{cases} Y_i(0) & \text{if } D_i = 0 \\ Y_i(1) & \text{if } D_i = 1 \end{cases}$$

Using this notation the treatment effect can be described as

$$\tau_i = Y_i(1) - Y_i(0)$$

The parameter that we estimate in our analysis is the average treatment effect on the treated "ATT", which is defined as

$$\tau_{ATT} = E(\tau|D=1) = E(Y(1)|D=1) - E(Y(0)|D=1).$$

However, we cannot observe a counterfactual $E(Y(0)|D=1)$

PSM and selection bias

Taking the mean outcome of non-treated - $E(Y(0)|D=0)$ - as an approximation would not solve our problem as it would give a biased estimate of the ATT:

$$E(Y(1)|D=1) - E(Y(0)|D=0) =$$

$$\tau_{ATT} + E(Y(0)|D=1) - E(Y(0)|D=0)$$

We would in fact estimate an effect equal to $\tau_{ATT}+$ selection bias.
$\Longrightarrow$ we have to define a mechanism that allows to describe the process of assignment into treatment.

Propensity Score Matching

- Models participation into treatment
- Non-parametric technique
- Allows to select a group of non-treated individuals similar to the treated ones in all the relevant pre-treatment characteristics (X) $\Longrightarrow$ overcomes selection bias
- If this is so, the difference in outcomes can be attributed exclusively to treatment.
- 2 assumptions: Conditional Independence Assumption and Common Support

Conditional Independence Assumption

Conditional Independence Assumption or Unconfoundedness states that conditional on a set of pre-treatment observable variables X, potential outcomes are independent of assignment into treatment:

$$Y(1), Y(0) \perp D | X \qquad (unconfoundedness) \tag{1}$$

That is, given X the non-treated outcomes are what the treated outcomes would have been had they not been treated.
Problem: Curse of Dimensionality
Solution: Conditioning on a *balancing score* would allow to achieve equivalent results and at the same time to pass from a multi-dimensional setting to a one-dimensional one.

Conditional Independence Assumption 2

The propensity score (PS) is one of the possible balancing scores and corresponds to the conditional probability of receiving the treatment given the pre-treatment variables:

$$p(X) = Pr(D = 1|X) = E(D|X) \tag{2}$$

Rosenbaum and Rubin (1983) show that the CIA remains valid if controlling for p(X) instead of X, that is conditionally on p(X) the treatment and potential outcomes are independent:

$$Y(1), Y(0) \perp D|p(X) \qquad (\textit{unconfoundedness given PS}). \tag{3}$$

Common Support

The second key assumption about treatment assignment is the *overlap or common support* condition:

$$0 < P(D = 1|X) < 1 \qquad (\textit{overlap}). \tag{4}$$

If we are going to estimate the counterfactual for a given individual by someone matched to that individual, there has to be at least one similar individual in the counterfactual state. For every single value of X the probability of finding a treated and a control individual must be greater then 0 (Heckman, LaLonde and Smith, 1999).

Strong and weak ignorability

The combination of the two assumptions is called Strong Ignorability and it ensures that both ATE and ATT are defined. However, we are only interested in estimating the ATT. In this case we only need a weak ignorability.

- The first one is called unconfoundedness for controls:

$$Y(0) \perp D | p(X) \qquad (\textit{unconfoundedness for controls given PS}). \tag{5}$$

- The second is the weak overlap condition → the probability of being treated conditional on the propensity score has to be lower than one:

$$P(D = 1 | p(X)) < 1 \qquad (\textit{weak overlap given PS}). \tag{6}$$

Estimation Procedure - Example: Effect of FE College mergers on dropout

- Data for 2002-03. 387,253 students 52,360 of whom enrolled in FE colleges merged from 1997-1998 onwards
- We estimate the probability of enrolling in a merged institution (propensity score) on the basis of a set of pre-treatment characteristics
- We then use the estimated propensity score to estimate the ATT using different types of matching algorithms
- We check for the satisfaction of the common support condition and for the quality of our matching
- We assess the plausibility of the CIA and the sensibility of our estimates to the existence of hidden bias through the use of three different techniques

Propensity Score Estimation

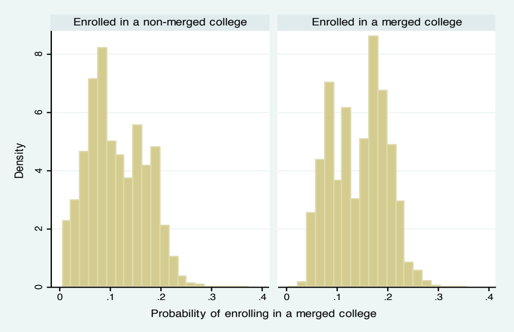
- The variables included in the estimation have to be chosen on the basis of the existence of a well known relationship with the outcome of interest (dropout probability) but also for their capacity to predict treatment
- Used the leave one-out cross validation method (starting with a minimal model, gradually increasing the number of covariates and checking the mean squared error)
- The information about treated and untreated individuals comes from the same questionnaire/form

Propensity Score Estimation - 2

Covariates	*dF/dx*	*(s.e.)*
Age > 20	0.025 * **	(0.002)
Male	−0.008 * **	(0.001)
Disability	0.006 * **	(0.002)
Bangladeshi	−0.063 * **	(0.003)
Black african	−0.042 * **	(0.002)
Black caribbean	−0.019 * **	(0.003)
Black other	0.021 * **	(0.003)
Chinese	0.053 * **	(0.005)
Indian	0.036 * **	(0.003)
Pakistani	0.008 * **	(0.002)
Asian other	−0.016 * **	(0.003)
Ethnic other	−0.009 * **	(0.003)
No qualification	0.022 * **	(0.004)
Qualif. < *level 1*	0.066 * **	(0.011)
Qualif. level 1	0.009 * *	(0.003)
Qualif. level 2	−0.037 * **	(0.003)
Qualif. level 4-5	0.045 * **	(0.014)
Qualif. unknown	0.045 * **	(0.003)
No. collleges in LLSC	−0.007 * **	(0.000)

Common Support

This assumption implies that both the students enrolled in a merged institution and the students enrolled in a non-merged one, should have a positive probability of enrolling in a merged college as shown by their propensity score.



Common Support, Min-Max

- We find the minima and maxima of the propensity score distribution for both treated [0.0191, 0.3433] and untreated individuals [0.0057,0.3572]
- We define our region of common support by selecting the highest of the two minima and the lowest of the two maxima [0.0191, 0.3433]
- The region of common support corresponds to the interval showing the distribution of the propensity score for the treated individuals
- We can conclude that we have perfect overlap at least for the estimation of the ATT

Choice of Matching Algorithm

All matching estimators can be seen as a special case of the following where the weights W_{ij} take different forms:

$$\tau_{ATT} = \sum_{i \in T} (Y_i - \sum_{j \in C} W_{ij} Y_j) \, w_i \tag{7}$$

T and C indicate respectively the treatment and control individuals, W_{ij} denotes the weights assigned to the control individuals when matching with the treated ones and w_i represents a re-weighting needed to re-build the outcome distribution for the treated.

Average Treatment Effect on the Treated

Matching Algorithm	*ATT*	*Stand. Error*	*Stand. Bias*	*(S.E.)*
Unmatched	−0.009 * **	0.001	9.160 * **	(0.660)
Nearest Neighbor, with replacement	−0.006	0.034	2.284 * **	(0.101)
Nearest Neighbor, no replacement	−0.022 * **	0.002	3.085 * **	(0.091)
Caliper=0.005, with replacement	−0.006	0.034	2.284 * **	(0.101)
Caliper=0.005, no replacement	−0.022 * **	0.002	3.085 * **	(0.091)
Caliper=0.05, with replacement	−0.006	0.034	2.284 * **	(0.101)
Caliper=0.05, no replacement	−0.022 * **	0.002	3.085 * **	(0.091)
Caliper=0.1, with replacement	−0.006	0.034	2.284 * **	(0.101)
Caliper=0.1, no replacement	−0.022 * **	0.002	3.085 * **	(0.091)
Multiple neighbors, N=15	−0.028 * **	0.009	2.073 * **	(0.088)
Multiple neighbors, N=10	−0.032 * **	0.011	2.076 * **	(0.072)
Multiple neighbors, N=5	−0.040 * **	0.015	1.800 * **	(0.068)
Radius, caliper=0.05	−0.016 * **	0.001	2.121 * **	(0.119)
Radius, caliper=0.005	−0.020 * **	0.001	0.984 * **	(0.034)

Matching Quality

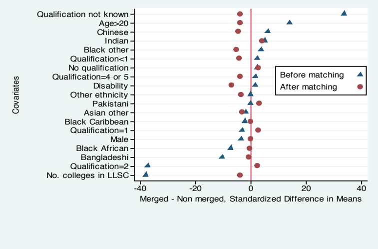
The standardized bias is computed applying the following formula:

$$SB = 100 \frac{(\bar{x}_{non-treated} - \bar{x}_{treated})}{\sqrt{0.5(s^2_{non-treated} - s^2_{treated})}} \tag{8}$$

$\bar{x}_{non-treated}$ and $s^2_{non-treated}$ are, respectively, the mean and the variance for the students enrolled in a college which has not been merged and $\bar{x}_{treated}$ and $s^2_{treated}$ are the mean and variance for students enrolled in a merged college.
A Standardised Difference in Means lower than 10 in absolute value is usually considered as an indication of a good balance.

Matching Quality

Obtaining a good covariates balance implies that the marginal distribution of each covariate is very similar for treated and untreated individuals.
We plot the standardized difference in means for each covariate before and after matching.



When using the pseudo-R^2 method we can see that it decreases from a value of 0.045 for the unmatched sample to a value of 0.002 or 0.003 for the matched one.

Multiple Control Groups (Rosenbaum, 1987)

We have created two control groups dividing the untreated students in two categories:

- The ones living in an LLSCs where at least one merged institution exists
- The ones whose LLSC doesn't have any merged institution.

We have, then, defined a treatment variable equal to 1 if the control student lives in an LLSC where there is at least one merged college.
We then estimate the ATT of being an "entitled not treated" on the probability of dropping out.
We find a zero effect implying that the CIA is plausible.

Rosenbaum bounds

This strategy assesses how strongly an unobserved factor would have to influence the treatment probability in order to bring our effect to 0.

Matching Algorithm		$e^\gamma = 1$	$e^\gamma = 1.25$	$e^\gamma = 1.5$	$e^\gamma = 1.75$	$e^\gamma = 2$
N. N., with repl.	Q^+_{MH}	4.204***	6.827***	9.031***	10.952***	12.667***
	Q^-_{MH}	4.204***	1.632*	0.367	2.137**	3.68***
N. N., no repl.	Q^+_{MH}	13.637***	23.928***	32.501***	39.911***	46.478***
	Q^-_{MH}	13.637***	3.476***	4.776***	11.793***	17.915***
Caliper=0.005, with repl.	Q^+_{MH}	4.204***	6.827***	9.031***	10.952***	12.667***
	Q^-_{MH}	4.204***	1.632*	0.367	2.137**	12.667***
Caliper=0.005, no repl.	Q^+_{MH}	13.637***	23.928***	32.501***	39.911***	46.478***
	Q^-_{MH}	13.637***	3.476***	4.776***	11.793***	17.915***
Caliper=0.05, with repl.	Q^+_{MH}	4.204***	6.827***	9.031***	10.952***	12.667***
	Q^-_{MH}	4.204***	1.632*	0.367	2.137**	12.667***
Caliper=0.05, no repl.	Q^+_{MH}	13.637***	23.928***	32.501***	39.911***	46.478***
	Q^-_{MH}	13.637***	3.476***	4.776***	11.793***	17.915***
Caliper=0.1, with repl.	Q^+_{MH}	4.204***	6.827***	9.031***	10.952***	12.667***
	Q^-_{MH}	4.204***	1.632*	0.367	2.137**	12.667***
Caliper=0.1, no repl.	Q^+_{MH}	13.637***	23.928***	32.501***	39.911***	46.478***
	Q^-_{MH}	13.637***	3.476***	4.776***	11.793***	17.915***
M.N., N=15	Q^+_{MH}	17.897***	25.922***	32.680***	38.579***	43.855***
	Q^-_{MH}	17.897***	10.057***	3.729***	1.572**	6.194***
M.N., N=10	Q^+_{MH}	16.211***	23.366***	29.401***	34.676***	39.398***
	Q^-_{MH}	16.211***	9.231***	3.601***	1.106	5.216***
M.N., N=5	Q^+_{MH}	13.748***	19.484***	24.337***	28.588***	32.401***
	Q^-_{MH}	13.748***	8.167***	3.678***	0.056	3.326***
Radius, caliper=0.05	Q^+_{MH}	6.733***	19.580***	30.228***	39.395***	47.488***
	Q^-_{MH}	6.733***	6.021***	16.500***	25.458***	33.334***
Radius, caliper=0.005	Q^+_{MH}	6.781***	19.627***	30.276***	39.443***	47.538***
	Q^-_{MH}	6.781***	5.972***	16.451***	25.408***	33.282***

Confounder variable approach

- We assumed the existence of a confounder variable distributed as the students' prior attainment (level 2).
- We computed the ATT and repeated the matching estimation 100 times.
- The resulting ATT is indicating that enrolling in a merged college reduces the probability of dropping out by 2.1%
- We can conclude that our estimates are not sensitive to the existence of a confounder distributed as the students prior attainment (level 2).

Usefull References

- Caliendo & Kopeinig "Some Practical Guidance for the Implementation of Propensity Score Matching", *Journal of Economic Surveys*, 2008, vol. 22, issue 1, 31-72
- Becker & Caliendo "Sensitivity analysis for average treatment effects" *The Stata Journal*, 2007, vol.7, issue 1
- R. H. Dehejia and S. Wahba. Propensity score matching methods for non-experimental causal studies. *Review of Economics and Statistics*, 84(1):151-161, 2002.
- J. J. Heckman, H. Ichimura, and P. E. Todd. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, 64(4, Special Issue):605-654, 1997.

Usefull References

- P. R. Rosenbaum. The role of a second control group in an observational study (with discussion). Statistical Science, 2:292-316, 1987.
- A. Ichino, F. Mealli, and T. Nannicini. From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity? *Journal of Applied Econometrics*, 23(3):305-327, 2008.

Natural Experiments

- Angrist and Krueger (Handbook of Labor Economics, 1999)
  - " ...the 'experimentalist' approach, ... puts front and center the problem of identifying causal effects from specific events or situations."
- "Events or situations" = natural experiments
  - Generate exogenous variation in certain variables that would otherwise be endogenous in the behavioral relationship of interest.
  - Natural experiment occurs when one group gets a treatment and another (very similar) group does not - (almost) like a real experiment
  - A natural experiment helps to better model the post-treatment counterfactual.

Difference-in-differences

- Suppose we randomly assign treatment to some units (or nature assigns treatment 'as if' by random assignment).
- To estimate the treatment effect, we could just compare the treated units before and after treatment.
- However, we might pick up the effects of other factors that changed around the time of treatment.
- Therefore, we use a control group to 'difference out' these confounding factors and isolate the treatment effect.

Difference in differences - method

We need to find 4 groups:

1. treated - before treatment
2. treated - after treatment
3. untreated - before treatment
4. untreated - after treatment

Take the mean value of each group's outcome:

	Treatment group	**Control group**
Before	$\overline{Y}_B(1)$	$\overline{Y}_B(0)$
After	$\overline{Y}_A(1)$	$\overline{Y}_A(0)$

and then calculate the "difference-in differences" of the means:
Treatment effect $=$ DD$= (\overline{Y}_A(1) - \overline{Y}_B(1)) - (\overline{Y}_A(0) - \overline{Y}_B(0))$

Difference in differences

The difference in differences, DD, should be the same as the value of β_3 in the regression

$$Y = \beta_0 + \beta_1 T + \beta_2 B + \beta_3 T \times B$$

- where $B = 0$ before treatment and $B = 1$ after treatment.
- where $T = 0$ if the person is in the control group and $T = 1$ if the person is in the treatment group.

- Advantage: estimator is insensible to change in the global state of the economy.
- Disadvantage: common trend assumptions, i.e. trends that may affect participants and nonparticipants are identical.

Eissa and Liebman (1996)

- Tax reform 1986: expansion of earned income tax credits (EITC) $\Longrightarrow$ financial incentive to take low wage jobs, but only for those with children.
- Sample: single women
  - single childless women
  - single women with at least 1 child.
- Effects of the reform on participation rates of these two groups.

Figure: Participation rates of single women

	Pre-TRA86	Post-TRA86	Difference	$\hat{\alpha}$
Treated group	0.729 (0.004)	0.753 (0.004)	0.024 (0.006)	
Control group	0.952 (0.001)	0.952 (0.001)	0.000 (0.002)	0.024 (0.006)

Standard errors in parentheses.
Source: Eissa and Liebman (1996, table 2).

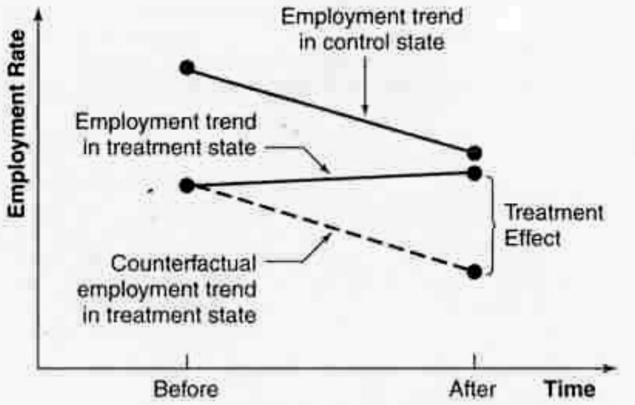
Card and Krueger (AER 1994)

- Min wage rose in New Jersey in April '92 (from \$4.25 to \$5.05 i.e. by about 20%) but not in Pennsylvania.
- Collected data on employment (from telephone surveys) in fast food restaurants in February '92 in NJ and in Eastern Pa
- and again in November '92.
- Effects of the policy on full time employment (FTE) .

Average employment in fast food restaurants before and after the New Jersey minimum wage increase

Variable	PA (i)	NJ (ii)	Difference, NJ − PA (iii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (.51)	−2.89 (1.44)
2. FTE employment after, all available observations	21.17 (.94)	21.03 (.52)	−.14 (1.07)
3. Change in mean FTE employment	−2.16 (1.25)	.59 (.54)	2.76 (1.36)

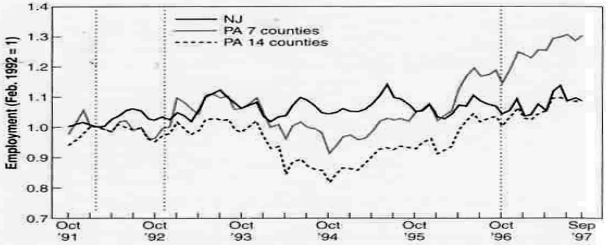
Notes: Adapted from Card and Krueger (1994), table 3. The table reports average full-time-equivalent (FTE) employment at restaurants in Pennsylvania and New Jersey before and after a minimum wage increase in New Jersey. The sample consists of all restaurants with data on employment. Employment at six closed restaurants is set to zero. Employment at four temporarily closed restaurants is treated as missing. Standard errors are reported in parentheses.



Causal effects in the DD model.

Is common trends reasonable?

- Berman, Neumark and Wascher (2000) get employment data from payroll records for Pa and NJ
- FTE trends differ even when the min wage is constant
- So Pa is not a good control group.



Employment in New Jersey and Pennsylvania fast food restaurants, October 1991 to September 1997 (from Card and Krueger 2000). Vertical lines indicate dates of the original Card and Krueger (1994) survey and the October 1996 federal minimum wage increase.

Card and Krueger reply

- Card and Krueger (2011) repeat their analysis using revised administrative data and alternative estimation methods.
- They also study the 1996 increase in the minimum wage in PA (in NJ the rise was not binding) and they find a modest effect on unemployment.
- They provide new evidence in support of their first study and conclude that the increase in NJ min-wage in April 1992 probably had no effect on total employment in NJ's fast-food industry.
- Although there may have been individual restaurants where employment rose or fell in response to the higher min-wage.

Value and Limits of Natural Experiments

- Simplicity of NE
- Rigorous identification of consequences of a particular event, if NE properly conducted.
- Limit: Many NE in the literature would not really count as experiments
  - failure of common trends assumption.
- Each NE is a particular event by definition, so its consequences requires a theory for generalisation in other context.

References

- Angrist, J.D. and Krueger, A.(1991), *Does Compulsory Schooling Attendance Affect Schooling and Earnings?*, Quarterly Journal of Economics.
- Angrist, J.D. and Krueger, A. (1999) *Empirical Strategies in Labor Economics*, Chapter 23 in O. Ashenfelter and D. Card, eds. The Handbook of Labor Economics, Volume III, North Holland.
- Angrist, J and Pischke, J.S. (2009) *Mostly Harmless econometrics*, Princeton University Press.
- Card, D. and and Krueger, A. (2011) *Wages, School Quality, and Employment Demand*, Oxford University Press Inc., New York.