Metodi Statistici per l'Analisi Socio-Economica
Parte II
Introduzione alla valutazione di un intervento

Lezione 6

Metodi di matching (abbinamento)

non è noto/non molto chiaro il meccanismo di assegnazione al trattamento

Idea di base:

Per ogni trattato, abbinare tramite selezione (match) la migliore unità per fare il confronto (anche da altra fonte di dati)

Come?

La selezione (il match) avviene sulla base della somiglianza delle caratteristiche osservate

Problema:

Presenza di caratteristiche non osservabili che influenzano la partecipazione al trattamento: distorsione da selezione!

Può essere utilizzato nell'ambito di quasi qualsiasi regola di partecipazione ad un intervento, a patto che esista un gruppo di non partecipanti

Metodi di matching (abbinamento)

Obiettivo:

creare un gruppo di controllo *ex post* di soggetti non trattati "simili" nelle caratteristiche osservabili ai trattati

dato un opportuno insieme di caratteristiche individuali X tali che, in ogni strato definito da X, la distribuzione della variabile risultato controfattuale degli esposti sia la stessa della variabile risultato fattuale dei non esposti.

selezionato il gruppo di controllo, l'effetto del trattamento è pari a:

$$E[\alpha \mid I=1] = E[Y^{T} \mid I=1,X] - E[Y^{NT} \mid I=0,X]$$

= differenza tra le medie della variabile-risultato Y nel gruppo dei trattati e nel gruppo dei non trattati ***abbinati***

Metodi di matching (abbinamento)

Ipotesi di fondo: è possibile compensare le differenze tra trattati e non trattati facendo uso delle caratteristiche X disponibili. Dato X, non ci sono differenze per Y^{NT} tra trattati e non trattati

Assunzione: le variabili X cui ci si condiziona soddisfano la condizione di indipendenza condizionata

(*Conditional Independence Assumption* - CIA):

$$(Y^T, Y^{NT}) \perp I \mid X \qquad (1)$$

- a parità di X *osservate*, assegnazione al trattamento (o non) è casuale:
- come conseguenza si ha che:

$$F(Y^{NT} \mid I=1, X) = F(Y^{NT} \mid I=0, X) = F(Y^{NT} \mid X)$$

$$F(Y^{T} \mid I=1, X) = F(Y^{T} \mid I=0, X) = F(Y^{T} \mid X)$$

Se assunzione (1) verificata: possiamo usare i non trattati per identificare il risultato controfattuale dei trattati.

Metodi di matching *vs* altri metodi

Matching e Metodo sperimentale (Ms):

- in entrambi i casi: stima effetto = differenza tra 2 gruppi
- Ms: gruppo controllo pre-trattamento, gruppi resi simili dall'assegnazione casuale sia per caratteristiche osservabili che non osservabili (*gruppi equivalenti*)
- Matching: gruppo controllo post-trattamento, gruppi resi simili solo per caratteristiche osservabili (*gruppi non equivalenti*)

Matching e analisi di regressione:

- in entrambi i casi: assunzione CIA
- diverso utilizzo dei dati disponibili:
  - **regressione** ipotizza la relazione (forma funzionale) tra variabile risultato e var. X (modello parametrico), stimata su tutti i dati (anche quindi con non trattati molto diversi dai trattati)
  - **Matching**: approccio non parametrico che usa (abbina) solo i soggetti più confrontabili tra trattati e non trattati (abbinamento usa solo osservazioni con ***supporto comune*** (*common support*)
  - **Matching**: meno restrittivo perché non implica forma funzionale ma più restrittivo perché impone che ci siano unità trattate e non trattate con caratteristiche simili.

    Matching emula un esperimento casuale rimpiazzando la randomizzazione con il condizionamento ad un insieme di variabili X.

Matching operativamente

Come effettuare il matching ?

- in presenza di varie caratteristiche X
  - abbinamento esatto: poco praticabile
  - uso del *propensity score*

Propensity score

- insieme caratteristiche X (effetto delle X) è riassunto dal *ps* = P(I=1 | X)
  - Si dimostra che (Rosenbaum e Rubin, 1983)

$E(Y^{NT} | P(X), I=1) = E(Y^{NT} | P(X), I=0) = E(Y^{NT} | P(X))$

  - e, inoltre:

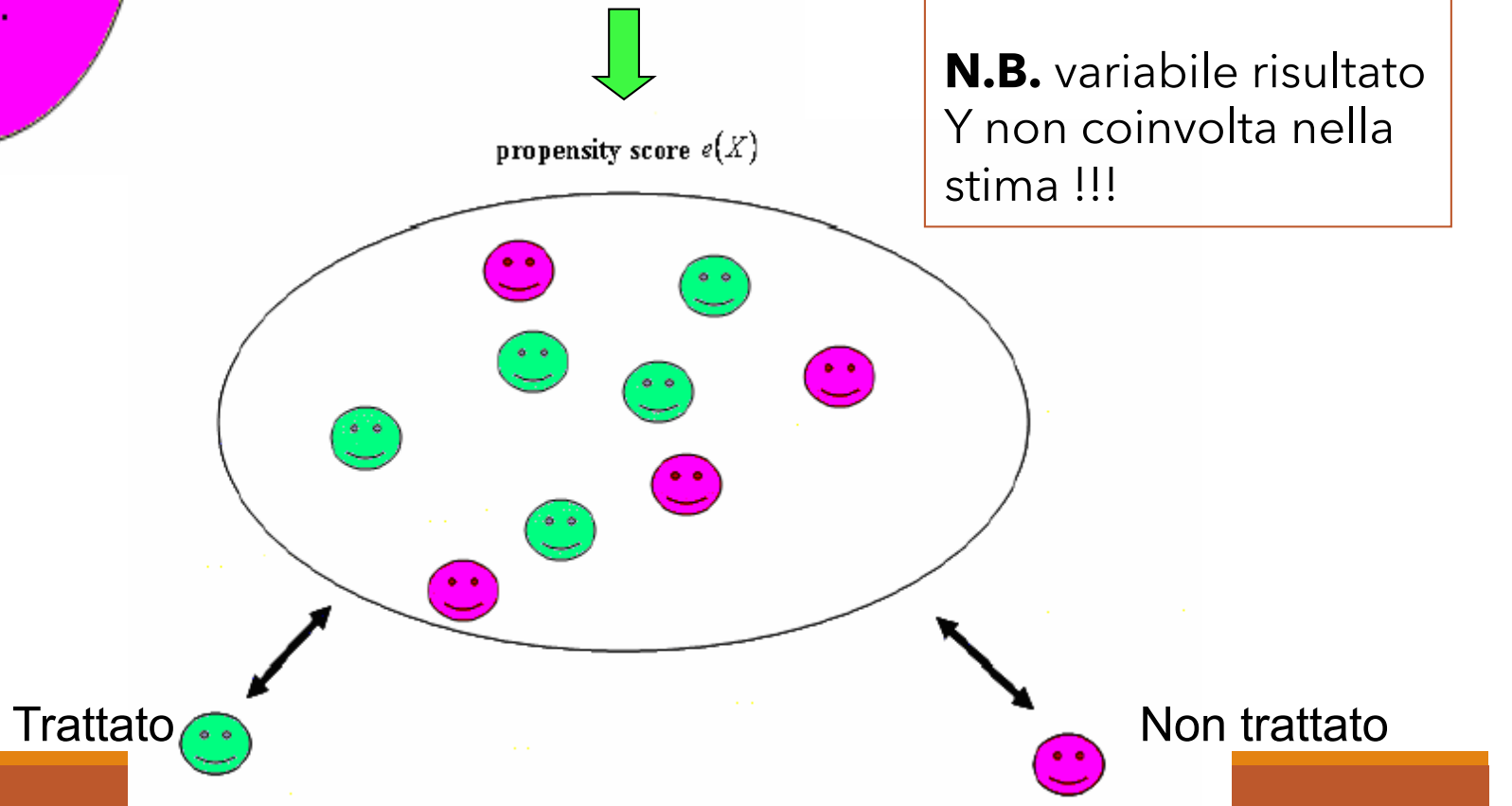
$0 < P(I=1 | X) < 1$ per tutte le unità

per nessun valore di X si trovano solo trattati o solo non trattati
(P(I=1| X non è mai uguale a 1o non è mai uguale a 0)

⟹ esistenza di un supporto comune

- se supporto comune (valori X in comune), *ps* = Prob "essere trattato" racchiude tutta l'informazione rilevante ipotizzata dalla CIA:
  - non è necessario usare direttamente le X per l'abbinamento, l'abbinamento è fatto solo rispetto al propensity score

Caratteristiche osservabili X e propensity score



Propensity score = probabilità condizionata di ricevere il trattamento dato un insieme (collezione) di covariate.

N.B. variabile risultato Y non coinvolta nella stima !!!

Costruzione del *Propensity score*

Stima della probabilità di "essere esposti al trattamento" date le var. osservabili X

(che possono influenzare tale probabilità)

- tutti i dati disponibili (trattati e non trattati)
- individuazione modello probabilistico per P(T=1) = *f* (X) (variabile dipendente è ora T =1/X)

Modello *logistico (logit)* (anche altri, es. modello *probit)*

$$P(T=1/X)=\frac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}}$$

Se i soggetti con stesso *propensity score* sono suddivisi nei due gruppi trattati e non trattati - i gruppi saranno approssimativamente bilanciati rispetto alle variabili utilizzate per la stima

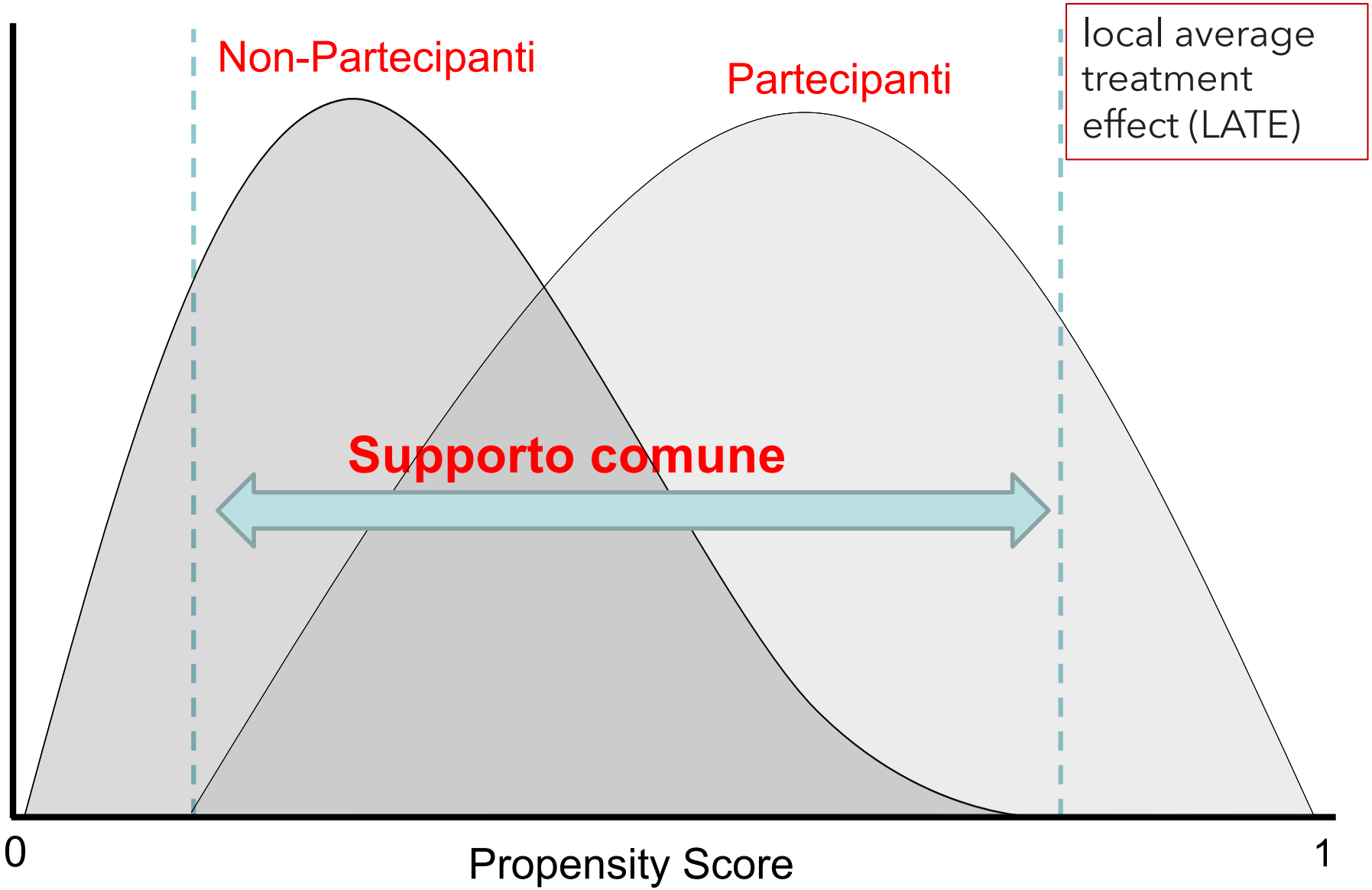
$\hat{p}$	Treatment	Control
0.9	$Y_{0.9}^{1}$	$Y_{0.9}^{0}$
0.7		
0.5		
0.3		
0.1		

Costruzione del *Propensity score* e stima effetto

dati rappresentativi e confrontabili su non-partecipanti e partecipanti

1. unione dei due campioni e stima di un modello logit (o probit) per la partecipazione all'intervento
2. restrizione dei campioni per assicurare il supporto comune (altrimenti fonte di distorsione in studi osservazionali)
3. per ogni partecipante trovare un campione di non-partecipanti che hanno *propensity scores* simili
4. confrontare le variabili risultato. La differenza è la stima dell'effetto dovuto all'intervento per quella osservazione
5. calcolare la media degli effetti individuali per ottenere l'effetto medio totale

Distribuzione dei propensity scores e supporto comune



Distribuzione dei propensity scores e supporto comune

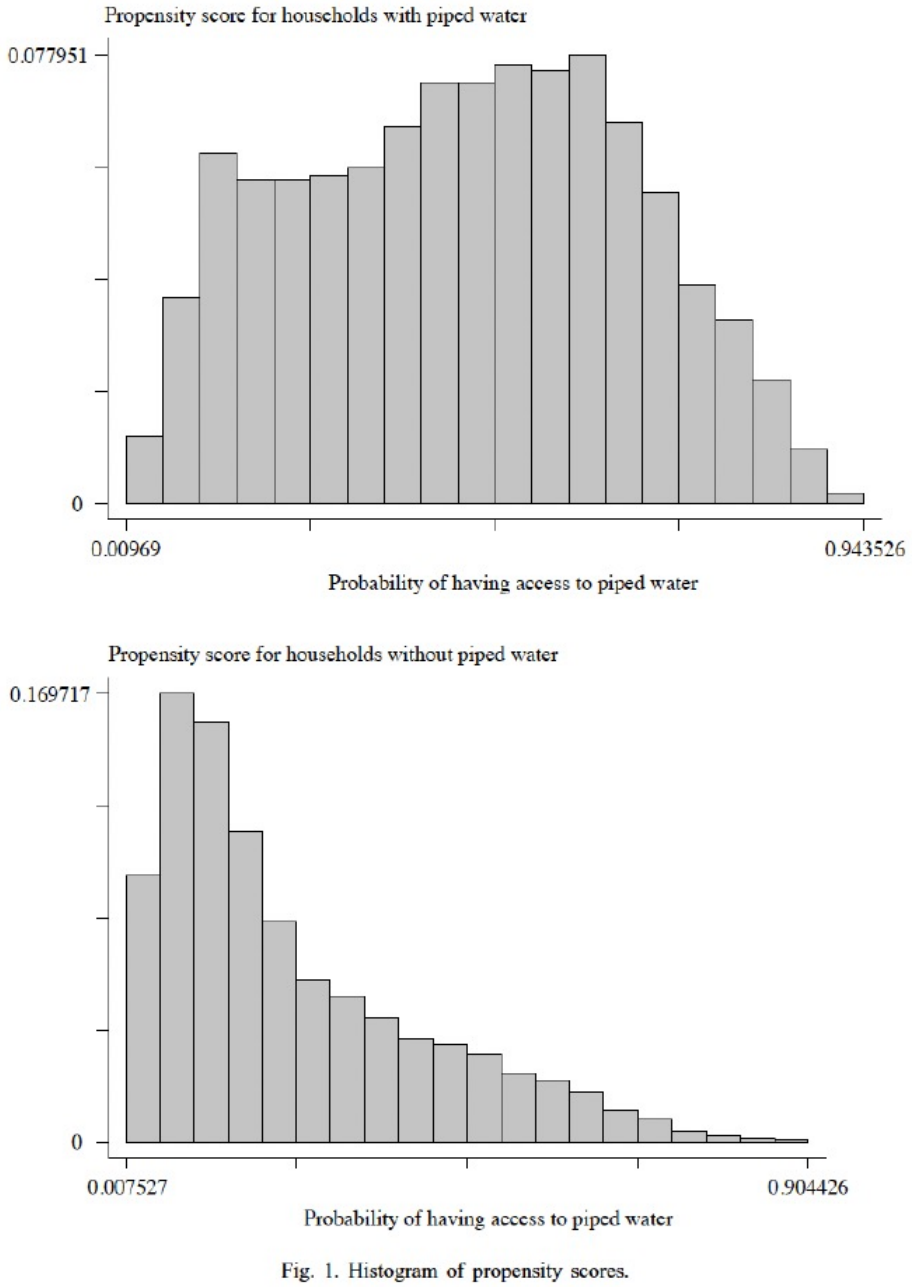


Fig. 1. Histogram of propensity scores.

Dati: ampia indagine cross-section nell'India rurale, 1993-94.

Campione: 33,000 famiglie rurali da 1765 villaggi, in 16 stati indiani:

- 9,000 famiglie con acqua corrente
- 24,000 famiglie senza

Con l'indagine raccolte informazioni dettagliate su stato di salute dei componenti la famiglia, condizioni socio-economiche e proprietà familiari

Perdita di 650 famiglie trattate per le quali non è stato trovato un abbinamento adeguato

Jalan J., M. Ravallion (2003) Does piped water reduce diarrhea for children in rural India? *Journal of Econometrics,* 112 pp. 153 – 173

Selezione dell'abbinato

- Sia $NE_{(i)}$ l'insieme dei non esposti tra i quali viene scelto quello da abbinare all'i-esimo esposto.

- $NE_{(1)} = NE$
- $NE_{(i)}$ è costituito da tutti i non esposti che non sono stati abbinati ai primi i-1 esposti.
- al crescere di i, l'insieme $NE_{(i)}$ si riduce:

  il non esposto da abbinare al 1^ esposto viene scelto in un insieme più ricco di quello entro il quale viene scelto il non esposto da abbinare all'ultimo esposto.

  Il problema è trascurabile se la numerosità di NE è molto superiore a quella di E.

- possibile soluzione: estrazione con reinserimento da NE. Ciò comporta che un certo non esposto può risultare abbinato a più esposti. Se ne deve tener conto nel calcolo degli "standard errors" delle stime.

Selezione dell'abbinato: senza o con reinserimento?

- *estrazione senza reinserimento* dall'insieme dei controlli: alla fine si possono ottenere abbinamenti di qualità peggiore
- *estrazione con reinserimento* la qualità dell'abbinamento resta elevata ma aumenta la varianza della stima

- varianza della stima della media in un campionamento casuale semplice (CCS)

  (popolazione di N unità dalla quale è estratto un campione di *n* unità)

  - CCS con reinserimento: $\text{var}(\bar{y}) = \frac{\sigma^2}{n}$
  - CCS senza reinserimento $\text{var}(\bar{y}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$

Tecniche di matching

- stima del *propensity score* ⟹ primo passo per ottenere la stima dell'impatto medio.

  Si deve poi procedere all'abbinamento di soggetti non esposti ad ogni soggetto esposto sulla base di *ps*(X) *[=stima di P(T/X)]*.

- problema: essendo *ps*(X) una variabile continua, potremo non trovare due soggetti che presentano esattamente lo stesso valore del *propensity score*.

- vari metodi proposti in letteratura per risolvere il problema.

  I più usati sono:

  - *nearest neighbor matching*
  - *radius matching*
  - *stratification matching*

- logica seguita:

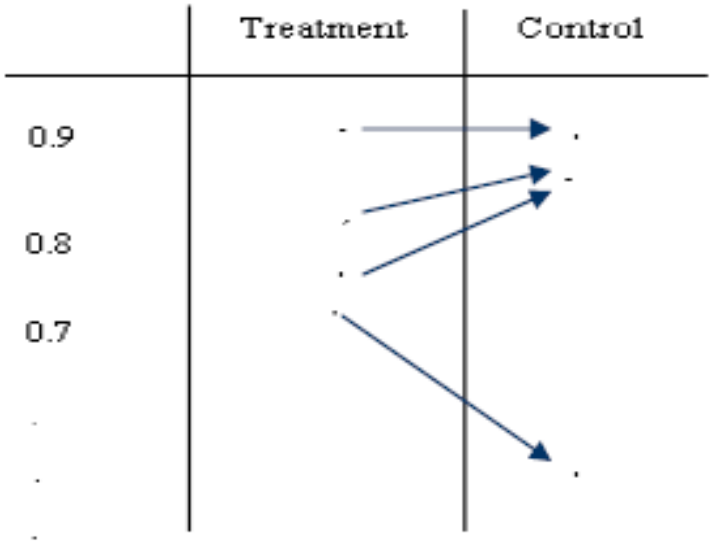
  trovare un soggetto tra i non trattati che presenti un valore di *ps*(X) simile a quello del soggetto esposto cui deve essere abbinato

Nearest neighbor (NN) matching

- Per ogni soggetto trattato si cerca il non trattato che presenta il *propensity score* più vicino.
- Operativamente:
  - T = insieme dei trattati
  - NT = insieme dei non trattati
  - Y_i^T, Y_j^{NT} i risultati per i trattati e non trattati rispettivamente
  - *C(i)* l'insieme di non trattati abbinati al trattato *i* con *propensity score* pari a ps_i.
  - insieme *C(i)* è tale che:

$$C(i) = \min_j \left| ps_i - ps_j \right|$$

- Solitamente, *C(i)* è formato da un solo elemento.
  Il caso di abbinamenti multipli è piuttosto raro, soprattutto nel caso
  in cui X includa variabili continue

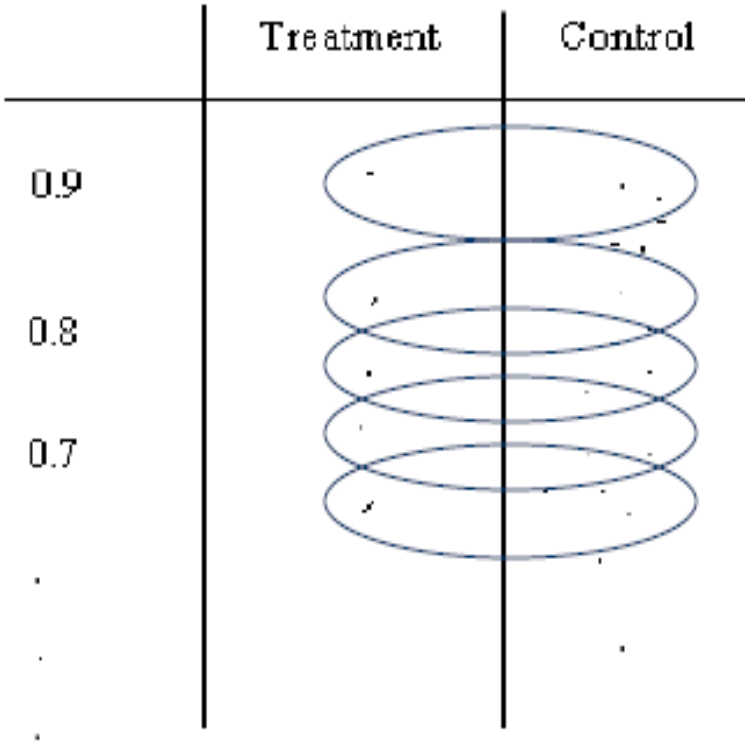


Radius matching (o caliper matching)

- NN matching: si rischia di abbinare a unità trattate anche unita NT con *ps* molto distante (se il più vicino tra i disponibili!)
- Nel radius matching l'insieme *C(i)* è definito nel seguente modo:

$$C(i) = \{ps_j : |ps_i - ps_j| < r\}$$

- l'insieme *C(i)* è formato da tutti i non trattati con *propensity score* che cade entro un raggio *r* da ps_i.
- Caliper = abbinamento con unità più vicina entro il raggio *r*.



Stima effetto in R/C M

- In entrambi i metodi la stima dell'impatto è data da:

$$\alpha = \frac{1}{N^{TR}} \sum_{i=1}^{N^{TR}} (Y_i^T - \bar{Y}_{(i)}^C)$$

dove

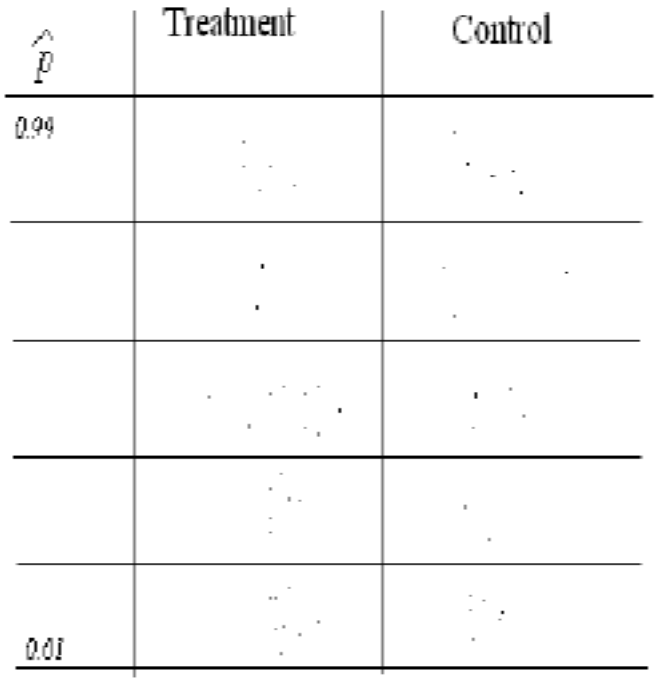
N^{TR} è il numero di trattati che hanno trovato abbinamento all'interno del radius/caliper

$\bar{Y}_{(i)}^C$ è il valore medio dei controlli abbinati all'*i*-esimo trattato (caliper: 1 sola unità)

Stratification matching

- il campo di variazione del *propensity score è suddiviso in intervalli (strati)* tali che in ogni strato trattati e non trattati hanno lo stesso *propensity score* medio
- la stima dell'effetto è calcolata prima entro ogni strato
- la stima dell'effetto complessivo è poi ottenuta come media ponderata degli effetti dentro gli strati

(ponderazione rispetto
a numero di casi nello
strato,
maggiore peso agli strati in
cui si concentra il maggior
numero di trattati,
suddivisione strati:
in genere usando i quintili)



Stima SM

- stima dell'impatto in ogni strato:

$$\alpha_q = \frac{\sum_{i \in I(q)} Y_i^T}{N_q^T} - \frac{\sum_{i \in I(q)} Y_i^{NT}}{N_q^{NT}}$$

- stima dell'impatto medio:

$$\alpha = \sum_{q=1}^{Q} \alpha_q \frac{\sum_{i \in I(q)} I_i}{\sum_{\forall i} I_i}$$

dove il peso di ogni strato è dato dalla corrispondente frazione di trattati e Q è il numero di strati

Considerazioni su tecniche di matching

- Uno dei problemi dello stratification matching (SM) è che scarta osservazioni negli strati in cui trattati o non trattati sono assenti.
- In questo caso il nearest neighbor matching (NNM) presenta sia vantaggi che svantaggi. Da un lato, a tutti i trattati risulta abbinato un non trattato, mentre nel SM se un trattato è l'unico appartenente ad uno strato non avrà abbinati.
- Dall'altro, alcuni degli abbinamenti trovati con il NNM possono risultare di scarsa qualità: infatti il non esposto più vicino ad un certo esposto può avere un *ps(X)* molto diverso.
- Il radius matching (RM) offre una soluzione alternativa in quanto ogni trattato è abbinato a un non trattato il cui *propensity score* cade in un predefinito intorno del *propensity score* del trattato. Ma se tale intorno è troppo piccolo è possibile che non contenga non trattati.
- D'altro canto più piccolo è l'intorno, migliore è la qualità dell'abbinamento.
- Questi metodi risolvono in modo diverso il trade-off tra qualità e quantità degli abbinamenti. Nessuno dei tre è a priori migliore degli altri.
  Provarli tutti è il modo migliore per verificare la robustezza delle stime!

HISP: Matching (2 scenari diversi per *ps*)

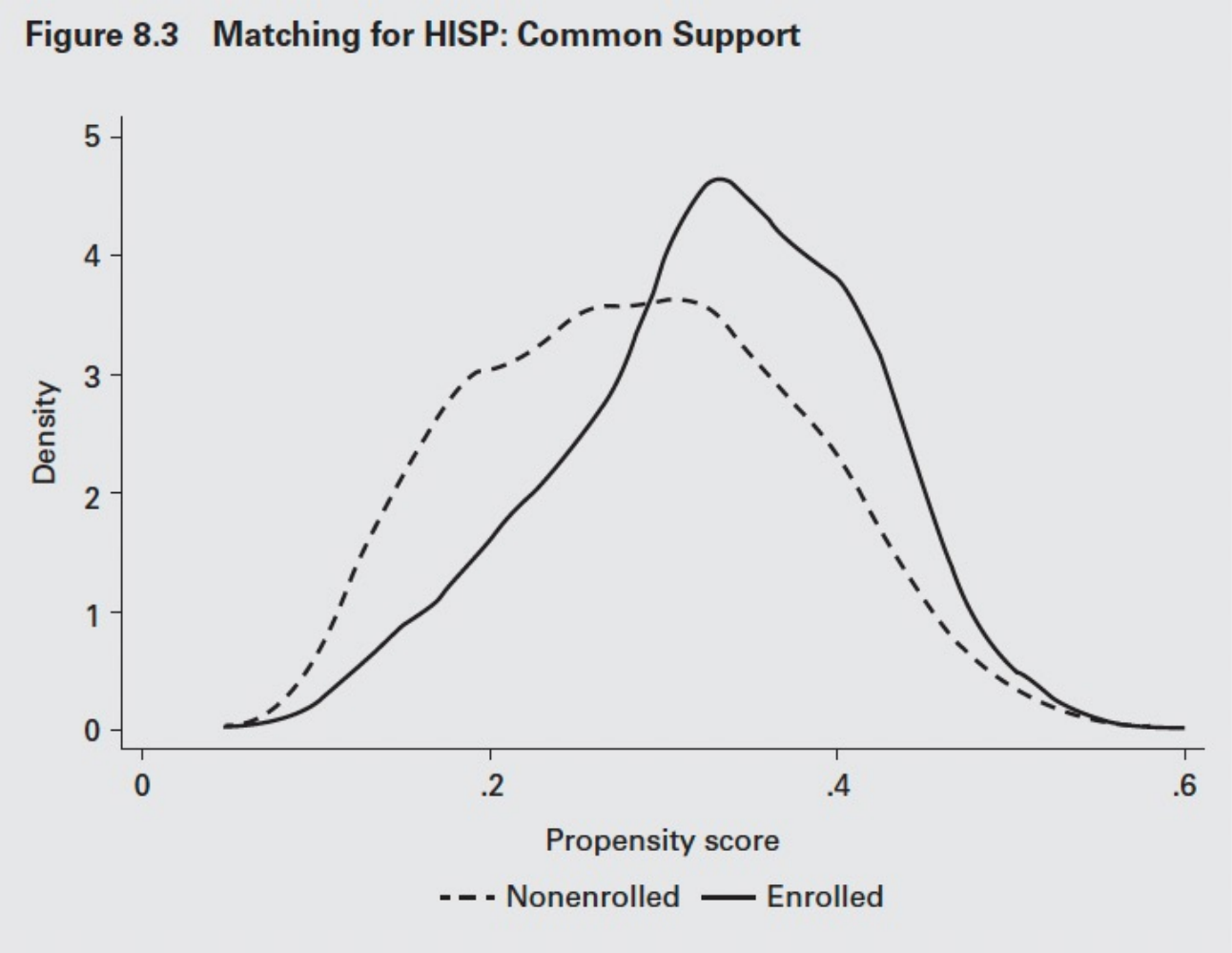
Table 8.1 Estimating the Propensity Score Based on Baseline Observed Characteristics

Dependent variable: Enrolled = 1	Full set of explanatory variables	Limited set of explanatory variables
Explanatory variables: Baseline observed characteristics	**Coefficient**	**Coefficient**
Head of household's age (years)	–0.013**	–0.021**
Spouse's age (years)	–0.008**	–0.041**
Head of household's education (years)	–0.022**	
Spouse's education (years)	–0.016*	
Head of household is female =1	–0.020	
Indigenous = 1	0.161**	
Number of household members	0.119**	
Dirt floor = 1	0.376**	
Bathroom = 1	–0.124**	
Hectares of land	–0.028**	
Distance to hospital (km)	0.002**	
Constant	–0.497**	0.554**

Note: Probit regression. The dependent variable is 1 if the household enrolled in HISP, and 0 otherwise. The coefficients represent the contribution of each listed explanatory variable to the probability that a household enrolled in HISP.

Significance level: * = 5 percent, ** = 1 percent.

HISP: Supporto comune (*ps* con più variabili)



Il supporto comune si estende su tutta la distribuzione del *ps*: nessuna delle famiglie partecipanti cade al di fuori dell'area di supporto comune.

Per ogni famiglia partecipante si trova un match con una famiglia non partecipante

HISP: Stima effetto con nearest neighbor matching

Table 8.2 Evaluating HISP: Matching on Baseline Characteristics and Comparison of Means

	Enrolled	Matched comparison	Difference
Household health expenditures (US$)	7.84	17.79 (using full set of explanatory variables)	−9.95
		19.9 (using limited set of explanatory variables)	−11.35

Note: This table compares mean household health expenditures for enrolled households and matched comparison households.

Table 8.3 Evaluating HISP: Matching on Baseline Characteristics and Regression Analysis

	Linear regression (Matching on full set of explanatory variables)	Linear regression (Matching on limited set of explanatory variables)
Estimated impact on household health expenditures (US$)	−9.95** (0.24)	−11.35** (0.22)

Note: Standard errors are in parentheses. Significance level: ** = 1 percent.

HISP: Matching e DID combinato

Table 8.4 Evaluating HISP: Difference-in-Differences Combined with Matching on Baseline Characteristics

		Enrolled	Matched comparison using full set of explanatory variables	Difference
Household health expenditures (US$)	Follow-up	7.84	17.79	–9.95
	Baseline	14.49	15.03	0.54
				Matched difference-in-differences = –9.41**
				(0.19)

Note: Standard error is in parentheses and was calculated using linear regression. Significance level: ** = 1 percent.

Se osservazioni su variabile outcome disponibili prima dell'intervento

Domande:

A. What are the basic assumptions required to accept these results based on the matching method?

B. Why are the results from the matching method different if you use the full versus the limited set of explanatory variables?

C. What happens when you compare the result from the matching method with the result from randomized assignment? Why do you think the results are so different for matching on a limited set of explanatory variables? Why is the result more similar when matching on a full set of explanatory variables?

D. Based on the result from the matching method, should HISP be scaled up nationally?

HISP rand = -10.14

Limiti del Propensity Score

- Sono richiesti preferibilmente grandi campioni
- Sovrapposizione (supporto comune) sostanziale
- Matching ex-post rischioso
  (alcune caratteristiche possono essere influenzate dal programma)
- Controllo solo rispetto a caratteristiche osservabili
  (che si assume siano misurate senza errore)