# UNIVERSITA' DEGLI STUDI DI PADOVA

FACOLTA' DI SCIENZE STATISTICHE

CORSO DI LAUREA SPECIALISTICA IN SCIENZE STATISTICHE, ECONOMICHE, FINANZIARIE E AZIENDALI

TESI DI LAUREA

## LA VALUTAZIONE DELL'IMPATTO DI UNA POLITICA: CONFRONTO FRA METODI SPERIMENTALI E NON SPERIMENTALI NEL CASO DELLA NATIONAL SUPPORTED WORK DEMONSTRATION

**Relatore**: Ch.mo Prof. Anna Giraldo

**Laureanda:** Manuela Massaro

**ANNO ACCADEMICO 2005/06**

# INDICE

# Premessa

Il lavoro si propone di confrontare stime sperimentali e non sperimentali dell'impatto di una politica; la valutazione d'impatto serve sostanzialmente a stabilire se, ed eventualmente in che misura, l'attuazione di una politica ha prodotto variazioni in uno o più aspetti di una data situazione rispetto a ciò che si sarebbe osservato in assenza della politica. In questo caso la politica presa in considerazione è la *National Supported Work Demonstration* (*NSWD*), un intervento attuato negli Stati Uniti nella seconda metà degli anni settanta con lo scopo di valutare l'efficacia di un particolare approccio all'inserimento nel mercato del lavoro di soggetti che soffrono di forme estreme di emarginazione sociale.

Il lavoro dunque richiama nel primo capitolo nozioni, definizioni e alcune problematiche che delineano quella che è una valutazione d'impatto condotta sia in ambito sperimentale che non sperimentale.

Il secondo capitolo approfondisce l'approccio non sperimentale presentando un particolare metodo statistico, ossia il *matching* tramite i *propensity score*, per stimare l'impatto di una politica.

A seguire vengono illustrati gli scopi, le caratteristiche nonché le stime sperimentali dell'impatto della *National Supported Work Demonstration*; infine tali stime vengono impiegate come *benchmark* per valutare l'impatto della stessa politica con metodi non sperimentali.

# Capitolo 1

# LA VALUTAZIONE DELL'IMPATTO: DEFINIZIONE, PROBLEMATICHE E METODI

## 1.1 Premessa

I dibattiti pubblici stanno rivolgendo sempre di più la loro attenzione all'importanza e all'utilità della valutazione, in particolare a quella valutazione che mira ad analizzare e ad interpretare i provvedimenti attuati in contesti caratterizzati da problematiche di tipo sociale.

Nell'opinione pubblica e nei contesti politici si sta affermando infatti la consapevolezza che, oltre alla progettazione e all'implementazione di particolari iniziative o eventi, è indispensabile prevedere anche un'adeguata opera di valutazione di quest'ultimi visto che il sempre crescente contenimento della spesa pubblica necessita che le risorse vadano usate in modo razionale e mirato.

L'importanza e l'utilità della valutazione, intesa come verifica dei risultati raggiunti di un programma definito nei suoi obiettivi, tempi e risorse, risulta quindi essere centrale per orientare tale processo di razionalizzazione.

## 1.2 Tipologie di politiche

La classificazione parte dalla distinzione tra politiche attive e politiche passive (Martini, 1997). In particolare:

- *politiche attive*: mirano a rimuovere le cause del problema che si va ad affrontare (siano esse di natura sociale, economica, culturale,…) mediante l'uso di risorse pubbliche o interventi di tipo regolativo. Ad esempio nel caso di politiche a favore dei disoccupati le politiche attive hanno lo scopo di agevolare il ritorno al lavoro dei disoccupati (coinvolgimento in attività di formazione o riqualificazione professionale, incentivazione alla creazione di lavoro autonomo,…);
- *politiche passive*: intendono alleviare le conseguenze del disagio in questione, agendo dunque, non sui comportamenti degli individui, ma direttamente sulle loro condizioni di vita. Riferendoci poi alla politiche a favore dei disoccupati, come esempio di politica passiva basta pensare alla cassa integrazione.

Nel caso delle politiche attive, riferite ai disoccupati, è quindi necessario chiedersi se il disoccupato è riuscito grazie alla politica a trovare lavoro in meno tempo e se questo lavoro è meglio remunerato mentre nel caso delle politiche passive il risultato è implicito nel trasferimento di reddito.

In quest'ultimo caso però non bisogna sottovalutare eventuali effetti non voluti come il fatto che il trasferimento di reddito disincentivi il disoccupato a cercarsi un nuovo lavoro.

In generale tra le due realtà appena presentate troviamo comunque molteplici situazioni intermedie in cui l'intervento prevede l'attuazione di azioni sia di tipo attivo che passivo.

Un'ulteriore distinzione viene fatta tra politiche a regime e politiche pilota: questa distinzione nasce dal fatto che ci si chiede a che livello di sviluppo di una politica è utile condurre la valutazione.

Nel caso di interventi pilota, l'intervento è concepito e realizzato con lo scopo di valutarne effetti e fattibilità: si vuole cioè capire se l'intervento può funzionare, anticipare eventuali problematiche nella sua implementazione, individuare opportunità e verificare se è in grado di produrre, almeno su scala ridotta, gli effetti desiderati.

Nel caso invece di una valutazione di interventi a regime si è consapevoli che la valutazione si rivolge ad un intervento programmato e realizzato per l'intera popolazione o territorio; qui la fase di progettazione può considerarsi conclusa e si vuole verificare se le strutture amministrative e organizzative sono state in grado di condurre la politica in modo positivo ossia se l'implementazione della politica risponde agli obiettivi originari e se è in grado di incidere sul fenomeno nel suo complesso.

Da tutto ciò è chiaro che il ricorso a interventi pilota preliminarmente alla realizzazione su larga scala presenta il notevole vantaggio di sperimentare con un costo ridotto per la collettività soluzioni nuove evitando anche potenziali errori derivanti dell'implementazione su ampia scala di soluzioni

inefficaci. Dall'altra parte però bisogna ricordare che non sempre risulta facile e immediato generalizzare ad un contesto globale quanto si è appreso in un contesto limitato e particolare.

## 1.3 La valutazione delle politiche

La prima domanda da porsi preliminarmente a qualsiasi progetto di valutazione è: "perché si vuole valutare?", bisogna cioè capire le intenzioni che muovono il valutatore (Martini, 1997). In risposta a questa domanda ci si può muovere in due diverse direzioni:

1. valutazione come *apprendimento*: lo scopo è di migliorare la conoscenza di come l'intervento opera, di quali effetti produce, di come viene percepito dagli operatori e dagli utenti e di come eventualmente interagisce con altre politiche. La *policy analysis* deve permettere dunque di apprendere elementi utili per decisioni di riforma, mantenimento, estensione o eliminazione dell'intervento;
2. valutazione come *controllo*: l'attività analitica si rivolge al concreto operare delle unità che gestiscono un intervento; lo scopo è di verificare come tali unità operative utilizzano le risorse loro assegnate, per migliorarne eventualmente la performance, e come gli utenti diretti di ciascuna unità sono serviti.

Una volta posto il problema del motivo per cui si vuole valutare bisogna domandarsi "cosa si vuole valutare?": a questo proposito si ritiene che la distinzione più importante riguardi:

1. valutazione del *processo*: questo termine riassume tutti gli aspetti interni di un'attività pubblica, dall'acquisizione degli *inputs* produttivi, alle regole che determinano il loro utilizzo, fino al momento della trasformazione degli *inputs* in prestazioni, erogazioni, prodotti;
2. valutazione dei *risultati*: questi si riferiscono a tutti quegli aspetti di un intervento pubblico che si riflettono sull'ambiente esterno, dai prodotti dell'attività pubblica utilizzati esternamente (*outputs*), alle modifiche dell'ambiente che l'attività pubblica intende produrre (*effects*).

### 1.3.1 La valutazione dell'impatto delle politiche

Valutare l'efficacia di un intervento significa sostanzialmente stabilire un legame di causa-effetto tra la realizzazione dell'intervento e le modifiche osservate nel fenomeno che l'intervento intende modificare (se dunque l'intervento costa ma non produce effetti e gli eventuali cambiamenti osservati sono da attribuirsi ad altri fattori, è plausibile concludere che esso non è costo-efficace): si tratta cioè di identificare l'*impatto* dell'intervento sul fenomeno in oggetto. D'ora in poi useremo infatti la dizione *impatto* della politica e non efficacia della politica dal momento che la parola impatto enfatizza il fatto che tra gli aspetti del fenomeno modificati dalla politica possono esservi sia quelli che la politica si proponeva esplicitamente di modificare sia altri non considerati e desiderabili, mentre con efficacia si tende sempre ad attribuire una connotazione positiva alla questione.

La valutazione d'impatto serve dunque a stabilire se, ed eventualmente in che misura, l'attuazione di una politica ha prodotto variazioni in uno o più aspetti di una data situazione rispetto a ciò che si sarebbe osservato in assenza della politica. Quanto detto implica perciò il confronto di un evento fattuale con un evento controfattuale dove:

- *evento fattuale*: ciò che si è osservato in seguito all'attuazione della politica;
- *evento controfattuale*: ciò che si sarebbe osservato se la politica non fosse stata attuata.

Concludendo le fasi fondamentali di un disegno di valutazione possono essere così riassunte:

1. definizione delle domande a cui la valutazione vuole rispondere;
2. individuazione della metodologia di analisi più appropriata per dare una risposta alle domande poste;
3. definizione del fabbisogno informativo richiesto dall'analisi;
4. ricognizione delle fonti disponibili, dei metodi e dei costi per la raccolta dei dati;
5. predisposizione di un piano per la disseminazione e l'utilizzo dei risultati ottenuti.

Infine una valutazione d'impatto è praticabile quando:

1. è ben definita la politica come insieme di azioni intraprese con un certo fine;
2. è ben definita la popolazione obiettivo della politica;

3. una o più caratteristiche dei componenti la popolazione obiettivo possono essere modificate dalla politica e tali caratteristiche devono essere chiare anche in assenza della politica e devono essere misurabili.

### 1.3.2 Minacce alla validità interna ed esterna di una valutazione

Affrontiamo a questo punto la questione che riguarda l'esistenza di eventuali minacce alla validità interna ed esterna di una valutazione.

In particolare una corretta valutazione d'impatto deve vantare una:

- *validità interna*: cioè deve permettere di individuare correttamente l'impatto che la politica ha sulle unità coinvolte nel disegno;
- *validità esterna*: quando i risultati possono considerarsi rappresentativi di una popolazione più vasta che non riguarda i soli individui coinvolti nel disegno.

A volte però tali requisiti vengono compromessi da molteplici fattori. Se ci riferiamo alla validità interna, quando si teme che l'equivalenza fra beneficiari e non beneficiari sia venuta meno è chiaro che si teme anche che l'impatto della politica non sia stato stimato correttamente. C'è anche la possibilità che qualcos'altro rispetto al trattamento sia responsabile di tutto o di una parte del cambiamento osservato e ancora c'è l'eventualità che il trattamento venga "contaminato" da qualche determinante e che ciò influisca sull'impatto (basti pensare ad una errata

somministrazione del trattamento, a effetti non previsti nella somministrazione che causano problemi nella variabile obiettivo o a un diverso comportamento del gruppo dei non trattati proprio perché consapevoli della presenza del trattamento).

Quando invece viene compromessa la possibilità di generalizzare i risultati ottenuti dalla valutazione sperimentale si parla di mancanza di validità esterna: in questo caso le stime possono anche essere corrette ma scarsamente utili in quanto difficilmente estendibili ad un popolazione di individui più grande e non direttamente coinvolta nella politica.

Lo strumento della valutazione sperimentale va perciò applicato solo dopo un'attenta verifica delle eventuali problematiche che potrebbero sorgere e delle sue probabilità di successo.

## 1.4 Eventi fattuali e controfattuali

La valutazione d'impatto sappiamo essere utilizzata per stabilire se il "miglioramento" osservato tra i beneficiari di una politica sia da attribuirsi alla loro esposizione alla politica e non ad altre determinanti che influiscono su di loro indipendentemente dalla politica; con "miglioramento" quindi si vuole intendere il cambiamento nella direzione desiderata di quei comportamenti o condizioni dei beneficiari su cui la politica punta per raggiungere i propri obiettivi.

Procediamo dunque definendo con $Y$ la caratteristica delle unità appartenenti alla popolazione $P$ rispetto alla quale si conduce la valutazione d'impatto. $Y$ corrisponde alla variabile obiettivo ed è

definita per ogni unità, sia in caso di esposizione alla politica sia in caso di mancata esposizione.

Definiamo poi con $Y_i^T$ il valore assunto dalla variabile obiettivo per l'i-esima unità se questa fosse esposta alla politica mentre con $Y_i^{NT}$ il valore assunto dalla variabile obiettivo se questa non fosse stata esposta alla politica (dove con *T* intenderemo i *trattati* mentre con *NT* i *non trattati*).

L'impatto della politica per l'i-esima unità risulta essere:

$$\alpha_i = Y_i^T - Y_i^{NT}$$

A questo punto l'effettivo status esposto/non esposto alla politica, definito con la variabile *I*, determinerà quale dei due esiti potenziali $Y_i^T$ e $Y_i^{NT}$ diventerà l'evento fattuale e controfattuale, ricordando che con evento fattuale si intende ciò che si è osservato in seguito alla realizzazione di una politica e viceversa con evento controfattuale si intende ciò che si sarebbe osservato se la politica non fosse stata realizzata.

I possibili casi diventano quindi:

| | ***I = 0 (non esposti)*** | ***I = 1 (esposti)*** |
|---|---|---|
| ***$Y_i$ Trattati*** | evento controfattuale | evento fattuale |
| ***$Y_i$ Non Trattati*** | evento fattuale | evento controfattuale |

E' evidente a questo punto che l'impatto per l'i-esima unità $\alpha_i$ presuppone la differenza tra due variabili una delle quali non è osservabile in quanto si riferisce all'evento controfattuale e dunque l'impatto stesso risulta essere non osservabile: quanto

appena detto costituisce quello che viene chiamato *the fundamental problem of causal inference*.

## 1.5 *The fundamental problem of causal inference*

Tale problema abbiamo detto essere rappresentato dal fatto che l'impatto per i singoli individui $\alpha_i$ non è osservabile e dunque non è calcolabile: tuttavia per riuscire a ricavare una stima dell'impatto si può spostare l'attenzione dai singoli $\alpha_i$ ad alcune caratteristiche della distribuzione di $\alpha$ nella popolazione di riferimento; infatti sotto opportune condizioni, alcune caratteristiche, come l'impatto medio o l'impatto medio sugli esposti, sono identificabili.

Una soluzione che rende possibile tale identificazione è quella di procedere, come vedremo meglio più avanti, con un disegno di tipo sperimentale ossia tramite randomizzazione che assegna i soggetti ai gruppi dei trattati e dei non trattati in maniera del tutto casuale: ciò garantisce che gli individui assegnati ai due gruppi presentino aspetti pressoché uguali e se si riscontrano delle differenze alla conclusione dell'intervento, tali differenze sono da imputare all'intervento stesso.

Come anticipato si vuole andare a valutare l'impatto medio nella popolazione $E(\alpha)$ o tra gli esposti alla politica $E(\alpha|I=1)$.

A questo punto si può procedere confrontando il risultato ottenuto dagli esposti $Y^T|I=1$ con quello ottenuto dai non esposti $Y^{NT}|I=0$ oppure il risultato ottenuto dagli esposti dopo l'effettiva

esposizione alla politica $Y_{t_2}^{T}\left|I=1\right.$ con il risultato ottenuto dagli stessi nel periodo precedente all'esposizione $Y_{t_1}^{NT}\left|I=1\right.$ (dove $t_1 < t_2$).

Il punto critico per ogni valutazione d'impatto riguarda il grado di appropriatezza dei confronti sopra citati: ci si chiede infatti se il risultato dei non esposti o il risultato degli esposti prima dell'esposizione "approssimino bene" l'evento controfattuale cioè il risultato che gli esposti avrebbero ottenuto in caso di mancata esposizione: ed è proprio su questa questione che la randomizzazione gioca un ruolo fondamentale in quanto fa sì che si riesca ad approssimare il risultato controfattuale degli esposti con il risultato fattuale dei non esposti o degli esposti prima dell'effettiva esposizione.

Vediamo ora come procedere analiticamente: se prendiamo in esame il primo confronto citato prima, $Y^{T}\left|I=1\right.$ e $Y^{NT}\left|I=0\right.$, possiamo determinare le seguenti grandezze:

$$E\left(Y^{T}\left|I=1\right.\right) \text{ e } E\left(Y^{NT}\left|I=0\right.\right)$$

La determinazione dell'impatto medio sugli esposti però presuppone un termine non osservabile e dunque l'impatto medio stesso risulta non identificabile, infatti:

$$E\left(\alpha\left|I=1\right.\right)=E\left(Y^{T}-Y^{NT}\left|I=1\right.\right)=E\left(Y^{T}\left|I=1\right.\right)-E\left(Y^{NT}\left|I=1\right.\right)$$

dato che in generale:

$$E\left(Y^{NT}\left|I=1\right.\right)\neq E\left(Y^{NT}\left|I=0\right.\right)$$

ossia in assenza della politica esposti e non esposti non è detto che ottengano mediamente gli stessi risultati.

Possiamo però calcolare la seguente quantità:

$$E\left(Y^T \middle| I=1\right) - E\left(Y^{NT} \middle| I=0\right) + E\left(Y^{NT} \middle| I=1\right) - E\left(Y^{NT} \middle| I=1\right) =$$

$$= E\left(\alpha | I=1\right) + \left[E\left(Y^{NT} \middle| I=1\right) - E\left(Y^{NT} \middle| I=0\right)\right]$$

dove il secondo termine tra parentesi rappresenta la differenza media tra esposti e non esposti che si sarebbe osservata se anche gli esposti non fossero stati esposti nonostante fossero stati selezionati (tale differenza viene definita *selection bias* o distorsione da selezione). Tale quantità impedisce alla quantità calcolabile:

$$E\left(Y^T \middle| I=1\right) - E\left(Y^{NT} \middle| I=0\right)$$

di coincidere con il parametro di interesse:

$$E\left(\alpha | I=1\right) = E\left(Y^T - Y^{NT} \middle| I=1\right)$$

Ciò è dovuto fondamentalmente al processo di selezione in quanto i soggetti coinvolti nella politica ($I = 1$) per motivi dovuti al processo di selezione otterrebbero dei risultati mediamente diversi dai soggetti non coinvolti ($I = 0$) anche se non prendessero parte alla politica.

Se confrontiamo poi $Y_{t_1}^{NT} \middle| I=1$ e $Y_{t_2}^{T} \middle| I=1$ le grandezze che si ricavano risultano essere:

$$E\left(Y_{t_1}^{NT} \middle| I=1\right) \text{ e } E\left(Y_{t_2}^{T} \middle| I=1\right)$$

Come nel caso precedente però l'impatto medio sugli esposti non è identificabile dato che:

$$E\left(\alpha_{t_2} \middle| I=1\right) = E\left(Y_{t_2}^{T} \middle| I=1\right) - E\left(Y_{t_2}^{NT} \middle| I=1\right)$$

Infatti non si tiene conto che in generale:

$$E\left(Y_{t_2}^{NT} \middle| I=1\right) \neq E\left(Y_{t_1}^{NT} \middle| I=1\right)$$

ossia che in assenza della politica non è detto che gli esposti al tempo $t_2$ (post-politica) ottengano gli stessi risultati medi del tempo $t_1$ (pre-politica).

Procediamo allora calcolando la seguente quantità:

$$E\left(Y_{t_2}^{T}\middle|I=1\right)-E\left(Y_{t_1}^{NT}\middle|I=1\right)+E\left(Y_{t_2}^{NT}\middle|I=1\right)-E\left(Y_{t_2}^{NT}\middle|I=1\right)=$$

$$=E\left(\alpha_{t_2}\middle|I=1\right)+\left[E\left(Y_{t_2}^{NT}\middle|I=1\right)-E\left(Y_{t_1}^{NT}\middle|I=1\right)\right]$$

dove il secondo termine tra parentesi indica l'evoluzione media nel tempo del risultato da non esposti per gli esposti.

Riassumendo si può affermare che, in entrambi i casi appena presentati, le differenze osservate non indicano esclusivamente l'impatto della politica: rivelano pure quelle differenze che si sarebbero comunque osservate anche in assenza della politica. Perciò:

miglioramento osservato

=

impatto della politica

+

differenze osservate anche in assenza della politica

(situazione controfattuale)

Il fatto che un miglioramento sia osservato tra i beneficiari non implica per forza che l'impatto sia positivo; nel caso estremo in cui tutto il miglioramento sia dovuto all'effetto di altre determinanti l'impatto dell'intervento sarebbe pari a zero.

In pratica si procede come abbiamo visto sopra trovando una valida approssimazione per la situazione controfattuale in modo da ottenere, per differenza, una stima dell'impatto.

## 1.6 Il ruolo del processo di selezione

Quanto detto finora comporta che per ottenere una stima della situazione controfattuale vengano osservati comportamenti o condizioni di coloro che non prenderanno parte alla politica e per fare ciò è necessario creare un cosiddetto gruppo di controllo. L'idea che sta alla base è che, se gli esposti e non sono simili nelle condizioni di partenza (cioè in assenza dell'intervento), quello che si osserverà tra i non esposti approssimerà ciò che si sarebbe osservato tra gli esposti se l'intervento non fosse stato attuato.

A questo punto la possibilità di stimare l'impatto di un intervento dipende sostanzialmente da tre condizioni:

1. le osservazioni sul cambiamento devono essere disponibili per i beneficiari dell'intervento: tale condizione risulta essere facile da definire dato che i beneficiari, per definizione, sono individuabili dato che hanno ricevuto qualche tipo di prestazione o servizio;
2. le osservazioni sul cambiamento devono essere disponibili per i non beneficiari (gruppo di controllo);
3. beneficiari e non beneficiari devono essere "simili" in assenza dell'intervento.

L'ultima condizione è sicuramente la più cruciale per poter interpretare la differenza osservata tra i due gruppi come una

stima dell'impatto dell'intervento: infatti se i due gruppi sono diversi anche in assenza dell'intervento il confronto tra i due gruppi comporterebbe una stima distorta che porterebbe a sottostimare o a sovrastimare l'impatto dell'intervento.
Tale distorsione o *selection bias* dipende per lo più da come avviene il processo di selezione attraverso cui alcuni soggetti sono esposti mentre altri non lo sono.

### 1.6.1 Il metodo sperimentale

Un primo modo per eliminare il problema della distorsione è quello di intervenire manipolando direttamente il processo di selezione che determina quello che poi sarà il gruppo degli esposti e dei non esposti alla politica: si procede cioè selezionando in modo casuale i soggetti esposti alla politica (gruppo di trattamento) e quelli non esposti (gruppo di controllo); in questo modo si garantisce che i due gruppi siano equivalenti in termini di condizioni di partenza.
Dal punto di vista analitico sappiamo che vogliamo ottenere $E(\alpha)$ e disponiamo di informazioni su $Y^T \middle| I=1$ e $Y^{NT} \middle| I=0$. Sia $F_T(\cdot|\cdot)$ la funzione di ripartizione di $Y^T|\cdot$ e $F_{NT}(\cdot|\cdot)$ la funzione di ripartizione di $Y^{NT}|\cdot$. Il parametro di interesse è pari a:

$$E[\alpha] = E[Y^T] - E[Y^{NT}] = \int y f_T(u)du - \int y f_{NT}(u)du = \int y f F_T(y) - \int y dF_{NT}(y)$$

Le informazioni disponibili permettono però il calcolo di:

$\int y dF_T(y|I=1)$ e $\int y dF_{NT}(y|I=0)$

ma in generale:

$$\int ydF_T(y) \neq \int ydF_T(y|I=1)$$

$$\int ydF_{NT}(y) \neq \int ydF_{NT}(y|I=0)$$

dato che:

$$F_T(y) \neq F_T(y|I=1)$$

$$F_{NT}(y) \neq F_{NT}(y|I=0)$$

A questo punto condizione necessaria e sufficiente perché $F_T(\cdot) = F_T(\cdot|I=1)$ e $F_{NT}(\cdot) = F_{NT}(\cdot|I=0)$ è l'indipendenza stocastica tra $(Y^T, Y^{NT})$ e $I$: $(Y^T, Y^{NT}) \perp I$

Il processo di selezione delle unità esposte e non esposte alla politica e il processo che determina i valori di $Y^T$ e $Y^{NT}$ devono essere perciò indipendenti e se tale condizione vale, allora la distribuzione di $Y^T$ tra gli esposti coincide con la distribuzione di $Y^T$ nella popolazione e analogamente vale per $Y^{NT}$:

$$F_T(y|I=1) = F_T(y)$$

$$F_{NT}(y|I=0) = F_{NT}(y)$$

E dunque:

$$E(Y^T|I=1) = E(Y^T)$$

$$E(Y^{NT}|I=0) = E(Y^{NT})$$

In precedenza si era affermato che il confronto tra $Y^T|I=1$ e $Y^{NT}|I=0$ non tiene conto che $E(Y^{NT}|I=1) \neq E(Y^{NT}|I=0)$ ma se $(Y^T, Y^{NT}) \perp I$, le medie coincidono perché esposti e non esposti alla politica sono rappresentati dalla stessa popolazione.

Giunti a questo punto è necessario però chiedersi se è sempre praticabile questo disegno sperimentale: l'assunto fondamentale che sta alla base di questo processo di selezione è che lo status esposto/non esposto alla politica sia determinato univocamente dalla randomizzazione ma tale approccio nella realtà non è sempre applicabile agli interventi in quanto il tentativo di escludere alcuni individui dalla fruizione di un intervento si scontra spesso con ostacoli di natura etica, legale e organizzativa (basti pensare ad esempio ad un trattamento medico).

Bisogna comunque riconoscere all'approccio sperimentale, oltre al fatto di garantire l'equivalenza dei due gruppi, il vantaggio di produrre automaticamente un campione di non esposti. La selezione dei non beneficiari, infatti, rappresenta un problema pratico non indifferente in un disegno non sperimentale e può addirittura costituire, visti gli elevati costi da sostenere, un ostacolo alla realizzazione della valutazione a prescindere dalla problematica relativa all'eliminazione del *selection bias*.

### 1.6.2 Il metodo non sperimentale

Questo tipo di approccio, che tratteremo in maniera più approfondita nel capitolo seguente, è applicabile ad un'ampia gamma di situazioni e ovviamente saremo in presenza di una selezione non casuale nei casi in cui l'assegnazione al gruppo degli esposti e non esposti non sarà determinata univocamente mediante randomizzazione.

In questo caso valgono le seguenti disuguaglianze:

$$F_T(y) \neq F_T(y|I=1)$$

$$F_{NT}(y) \neq F_{NT}(y|I=0)$$

in quanto esposti e non esposti sono diversi rispetto alla variabile obiettivo anche in assenza della politica.

L'idea di fondo sta perciò nel fatto di voler correggere mediante l'uso di tecniche statistiche le differenze di partenza che emergono "naturalmente": si vuole cioè essere in grado di ricostruire la situazione controfattuale utilizzando l'esperienza di un gruppo di individui che non è stato coinvolto nell'intervento.

Questo tipo di approccio presenta numerose varianti a seconda del tipo di informazioni che sono disponibili sugli esposti e non esposti.

# Capitolo 2

# LA VALUTAZIONE DELL'IMPATTO CON METODI NON SPERIMENTALI

## 2.1 Disegni non sperimentali

Quando l'assegnazione al gruppo di trattamento e al gruppo di controllo viene realizzata tramite randomizzazione si è in presenza di una situazione "sperimentale", quasi da laboratorio e l'impatto medio viene stimato semplicemente confrontando la media della variabile obiettivo sui trattati con la media della variabile obiettivo sui non trattati.

In realtà però la valutazione di un intervento tramite un disegno sperimentale non sempre è realizzabile e dunque si ricorre all'impiego di metodi non sperimentali.

Caratteristica rilevante di tali metodi è senza dubbio la possibilità di poterli applicare ad una serie di situazioni molto più ampia rispetto all'approccio sperimentale e l'idea di fondo della valutazione non sperimentale è quella di ricostruire la situazione controfattuale servendosi dell'esperienza di un gruppo di individui che non è stato coinvolto nell'intervento.

La differenza sta nel processo di selezione che non è manipolabile dal valutatore, ossia non permette di assegnare casualmente i

soggetti ai gruppi di trattamento e controllo, ma è spontaneo cioè frutto delle decisioni degli individui o degli operatori: in particolare si parlerà di autoselezione se la decisione dell'esposizione o meno alla politica è presa direttamente dalle unità appartenenti alla popolazione di riferimento, mentre si parlerà eteroselezione se la decisione è presa da qualche autorità responsabile della politica (Martini, 1997).

Diversi approcci sono disponibili, tra questi citiamo brevemente:

- *Regression discontinuity design*: in questo caso la selezione dei beneficiari viene effettuata dagli operatori del servizio interamente sulla base delle caratteristiche osservabili dei richiedenti, ad esempio mediante la compilazione di graduatorie. Tutti i partecipanti infatti vengono sottoposti ad un pre-test e sulla base dei valori assunti dalla variabile osservata viene scelto un valore soglia e i due gruppi vengono formati da coloro che si situano al di sopra (o al di sotto) di questa soglia;
- *Difference-in-differences*: quando si hanno a disposizione dati longitudinali sui non partecipanti nel periodo precedente e successivo allo svolgimento della politica e è noto che la differenza *before-after* in assenza di trattamento è la stessa per il gruppo dei non trattati, allora si può ricorrere a questa tecnica che fornisce una stima del controfattuale come previsione sulla base dell'osservazione precedente;
- *Matching attraverso i propensity score*: la peculiarità di questa tecnica è quella di disporre di un insieme di variabili

esplicative $X$ tali che in ogni stato definito da $X$ la distribuzione della variabile risultato controfattuale è la stessa della distribuzione della variabile risultato osservata nei non partecipanti all'intervento; per ulteriori dettagli relativi a questo approccio rimandiamo comunque al paragrafo successivo.

## 2.2 Metodo del matching attraverso i propensity score

Quando l'assegnazione al trattamento non avviene in modo casuale e si vuole eliminare il *selection bias*, si può ricorrere a questa tecnica che ha come caratteristica fondamentale quella di disporre di un insieme di variabili esplicative $X$ tali che in ogni strato definito da $X$ la distribuzione della variabile risultato controfattuale $F\left(Y^{NT} \middle| I=1\right)$ è la stessa della distribuzione della variabile risultato osservata nei non partecipanti $F\left(Y^{NT} \middle| I=0\right)$.

Il metodo del *matching* assume dunque che esistano un insieme di variabili $X$ per cui:

$$\left(Y^{T}, Y^{NT}\right) \perp I \middle| X \qquad (1)$$

Di conseguenza si avrà che:

$$F\left(Y^{NT} \middle| I=1, X\right) = F\left(Y^{NT} \middle| I=0, X\right) = F\left(Y^{NT} \middle| X\right)$$

$$F\left(Y^{T} \middle| I=1, X\right) = F\left(Y^{T} \middle| I=0, X\right) = F\left(Y^{T} \middle| X\right)$$

Il principio intuitivo di questo metodo è quello per cui si è in grado di aggiustare le differenze tra i partecipanti e i non partecipanti usando i regressori disponibili.

Se l'assunzione (1) è valida possiamo usare i non partecipanti per misurare quanto i partecipanti avrebbero guadagnato se non avessero partecipato, condizionatamente al valore di *X*. Si cerca dunque di abbinare un esposto ad un non esposto con le stesse caratteristiche: per assicurare ciò però è necessario assumere che ci siano partecipanti e non partecipanti per ogni *X* per il quale vogliamo un confronto. Formalmente si richiede che:

$$0 < P(I = 1 | x \in X) < 1 \qquad (2)$$

Questa condizione assicura che la distribuzione in (1) sia definita per ogni *X* che la soddisfa.

L'insieme delle condizioni (1) e (2) viene definito come "*strong ignorability*" dello stato *I* per le variabili risultato $(Y^T, Y^{NT})$, condizionatamente alle caratteristiche *X*.

Sotto tale condizione il *matching* produce un gruppo di confronto per il gruppo dei trattati che assomiglia ad un gruppo di controllo sperimentale dato che la distribuzione del risultato controfattuale $Y^{NT}$ per i partecipanti è la stessa della distribuzione osservata $Y^{NT}$ per il gruppo di confronto e si ha che:

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

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

L'assunto di "*strong ignorability*" consente poi di specificare l'impatto sugli esposti nel seguente modo:

$$E(Y^T - Y^{NT} | I = 1) = \int E(Y^T | I = 1, x) dF_{x|I=1} - \int E(Y^{NT} | I = 0, x) dF_{x|I=1}$$

Un sistema per procedere alla stima di tale quantità può essere quello di supporre che *X* sia un insieme di variabili categoriali; definiamo innanzitutto i due gruppi rispettivamente $n_E$ e $n_{NE}$ con:

$E = \{esposti\}$ e $NE = \{non\_esposti\}$

Estraiamo da *NE* un sottoinsieme $NE^M$ in modo tale che la distribuzione di *X* tra i non esposti inclusi in $NE^M$ coincida con la distribuzione di *X* tra gli esposti.

Operativamente si procede come segue:

1. si estrae senza reinserimento una unità da *E* e si osserva il suo *X*. Siano $x_1^E$ le caratteristiche del primo estratto e $y_1^T$ il suo risultato da esposto;
2. si estrae senza reinserimento una unità da *NE* le cui caratteristiche $x_{(1)}^{NE}$ soddisfano la condizione: $x_{(1)}^{NE} = x_1^E$. Il risultato da non esposto è $y_{(1)}^{NT}$;
3. si replica la sequenza 1) e 2) fino all'esaurimento dell'insieme *E*.

Sia ora:

$$M = \begin{cases} 1 & \text{se l'i-esimo non esposto è stato incluso in } NE^M \\ 0 & \text{altrimenti} \end{cases}$$

Allora, per costruzione, le due funzioni di ripartizione empiriche: $\hat{F}_{X|I=1}$ e $\hat{F}_{X|I=0,M=1}$ coincidono.

La quantità:

$$\frac{1}{n^E}\sum_{i=1}^{n^E}\left(y_i^T - y_{(i)}^{NT}\right)$$

è una stima consistente di $E\left(Y^T - Y^{NT} \middle| I=1\right)$

Ottenuta da:

$$\int E\left(Y^T \middle| I=1, x\right) dF_{x|I=1} - \int E\left(Y^{NT} \middle| I=0, x\right) dF_{x|I=1}$$

sostituendo $F_{X|I=1}$ con $\widehat{F}_{X|I=1}$ nel primo integrale e con $\widehat{F}_{X|I=0,M=1}$ nel secondo integrale.

E' doveroso precisare che il requisito fondamentale per il quale si può procedere con questo metodo è che le variabili *X* siano tutte osservabili. Nella pratica però gli individui sono caratterizzati anche da altre variabili non osservate che definiremo con *U*.

E' necessario imporre che, condizionatamente a *X*, lo status esposti/non esposti e/o la variabile risposta non dipendano da *U*:

$$P\left(I \middle| X, U\right) = P\left(I \middle| U\right)$$

$$F\left(Y^{NT} \middle| X, U\right) = F\left(Y^{NT} \middle| X\right)$$

$$F\left(Y^{T} \middle| X, U\right) = F\left(Y^{T} \middle| X\right)$$

Infatti se le variabili non osservabili *U* sono incorrelate con *I* e/o i risultati potenziali, attraverso il matching, si riesce a stimare correttamente l'impatto.

Quanto detto finora comporta che l'effettiva realizzazione del *matching* avviene in modo esatto, oltre che per il fatto che sia soddisfatta la condizione di "*strong ignorability*", solo se *X* è una variabile discreta.

Se *X* è una variabile continua diventa nulla la probabilità di trovare un non esposto con caratteristiche di *X* esattamente uguali a quelle dell'esposto, qualsiasi sia la numerosità dei gruppi. Se ciò avviene si ricorre a qualche criterio di "vicinanza" ossia si calcola la distanza tra due unità con caratteristiche $X_1$ e $X_2$: $d(X_1, X_2)$.

L'abbinamento migliore risulta poi essere quello per cui: $d\left(X_1^E, X_2^{NE}\right) = \min$.

### 2.2.1 I propensity score

Un modo per ovviare al calcolo delle distanze quando si è in presenza di variabili continue è quello di calcolare i cosiddetti *propensity score* che corrispondono a:

$e(X) = P(I = 1|X)$

Un risultato fondamentale legato a tale quantità è quello per cui se si afferma che $(Y^T, Y^{NT}) \perp I|X$ allora vale anche la condizione: $(Y^T, Y^{NT}) \perp I|e(X)$

Ciò significa che è sufficiente che la composizione dei due gruppi secondo $e(X)$ sia la stessa.

Tale risultato è assai rilevante soprattutto per il fatto che se la dimensione di $X$ è $k$, variabile da caso a caso, la dimensione di $e(X)$ è 1 qualsiasi sia $k$.

Con la condizione di "*strong ignorability*" poi varrà che:

$E\left(Y^T \middle| I = 1, e(X)\right) = E\left(Y^T \middle| e(X)\right)$

$E\left(Y^{NT} \middle| I = 0, e(X)\right) = E\left(Y^{NT} \middle| e(X)\right)$

Se si pensa quindi a quanto detto finora, con i *propensity score* la procedura di *matching* abbina all'i-esimo esposto con caratteristiche $x_i^E$ un non esposto con caratteristiche $x_{(i)}^{NT}$ tali che: $e\left(x_i^E\right) = e\left(x_{(i)}^{NE}\right)$.

Dunque la stima dell'impatto medio sui trattati diverrà:

$$E\left[Y^T - Y^{NT} \middle| I=1\right] = \int E\left[Y^T - Y^{NT} \middle| I=1, e(X)\right] dF\left(e(X) \middle| I=1\right) =$$
$$= \int \left(E\left[Y^T \middle| I=1, e(X)\right] - E\left[Y^{NT} \middle| I=1, e(X)\right]\right) dF\left(e(X) \middle| I=1\right) =$$
$$= \int \left(E\left[Y^T \middle| I=1, e(X)\right] - E\left[Y^{NT} \middle| I=0, e(X)\right]\right) dF\left(e(X) \middle| I=1\right) =$$
$$= \int E\left[Y^T \middle| I=1, e(X)\right] dF\left(e(X) \middle| I=1\right) - \int E\left[Y^{NT} \middle| I=0, e(X)\right] dF\left(e(X) \middle| I=1\right)$$

Oltre alla proprietà vista all'inizio del paragrafo vale anche la proprietà per cui $X \perp I | e(X)$: conseguenza di ciò è il fatto che osservazioni con lo stesso *propensity score* hanno la stessa distribuzione delle caratteristiche osservabili indipendentemente dal fatto che siano state trattate o meno. Quindi per un dato *propensity score* l'assegnazione al trattamento è del tutto casuale.

Operativamente per stimare i *propensity score* si può seguire la seguente procedura:

1. dato un campione formato da esposti e non esposti sui quali vengono rilevate *k* variabili esplicative *X*, si modella la probabilità che un generico individuo appartenga o meno al gruppo degli esposti condizionatamente a *X*: $P(I_i = 1 | X_i) = F(h(X_i))$ dove $F(\cdot)$ è una funzione di ripartizione normale o logistica e $h(X_i)$ è una funzione delle covariate con termini lineari e di ordine superiore;
2. si divide il campione in *q* intervalli del *propensity score* di uguali dimensioni e si verifica che in ciascun intervallo il *propensity score* medio di trattati e non trattati non differisca;
3. se la verifica fallisce in qualche intervallo, si divide a metà l'intervallo e si verifica nuovamente tale uguaglianza;

4. si continua fino a che, in tutti gli intervalli, il *propensity score* medio di trattati e non trattati non differisce.

Stimare i *propensity score* non è però sufficiente per ottenere una stima dell'impatto in quanto la probabilità di osservare due unità con esattamente lo stesso valore del *propensity score* è pari a zero essendo $e(X)$ una variabile continua.

Per far fronte a ciò sono stati proposti vari metodi (*nearest neighbor matching*, *radius matching*, *stratification matching*, …) la cui logica comune è quella di trovare una unità tra i trattati e una tra i non trattati che abbiano valori vicini del *propensity score*.

# Capitolo 3

# IL CASO DELLA NATIONAL SUPPORTED WORK DEMONSTRATION

## 3.1 Caratteristiche e scopi dell'intervento

La *National Supported Work Demonstration* (*NSWD*) è un intervento attuato negli Stati Uniti nella seconda metà degli anni settanta, precisamente tra il 1975 e il 1978, al fine di valutare l'efficacia di un particolare approccio all'inserimento nel mercato del lavoro di soggetti che soffrono di forme estreme di emarginazione sociale.

In particolare i soggetti destinatari di tale politica possono essere ricondotti a quattro gruppi:

- madri non sposate che non hanno lavorato più di tre mesi negli ultimi sei mesi e che da più di tre anni percepiscono il sussidio di povertà;
- giovani *drop-outs* della scuola dell'obbligo, cioè giovani tra i 17 e i 20 anni senza il diploma dell'obbligo con eventuali precedenti penali;
- ex tossicodipendenti;
- ex carcerati.

L'obiettivo di quello che viene definito *supported work* è di fornire a tali soggetti un'esperienza lavorativa reale ma in un ambiente

"protetto" e fortemente strutturato con un aumento graduale dell'intensità dello sforzo lavorativo: il tutto viene realizzato sotto una stretta e attenta supervisione dei *policy-makers* e attraverso l'inserimento in squadre di lavoro composte da individui con problemi simili e con un retribuzione uguale o di poco superiore al salario minimo. La durata di questo lavoro "protetto" è di circa un anno, dopodiché si punta all'inserimento nel mercato del lavoro regolare.

Tale politica, visto il periodo di realizzazione, è sicuramente una tra le più costose politiche attive di reinserimento sociale; l'investimento richiesto per ogni partecipante si aggirava infatti tra i 5.000 e gli 8.000 dollari. Gli elevati costi di tale politica hanno indotto il governo federale americano a valutare l'impatto della politica su scala ridotta prima di estenderla su larga scala.

Il costo complessivo della *demonstration*, intesa qui come intervento pilota, è di quasi 82 milioni di dollari di cui 11 per la valutazione e i restanti 70 milioni per la realizzazione dell'intervento vero e proprio (locali, equipaggiamento, retribuzione dei partecipanti e dei supervisori). Circa 50 milioni vengono forniti dai finanziatori nazionali (in particolare dai Ministeri del lavoro, della giustizia, dell'istruzione, della sanità e servizi sociali, e dello sviluppo urbano e poi da alcune organizzazioni private tra cui la *Ford Fundation*, che è tutt'oggi uno dei principali finanziatori privati di valutazioni di questo tipo) mentre la differenza viene coperta mediante fondi messi a disposizione dai governi locali, con la vendita dei beni e servizi

prodotti dai partecipanti e con lo storno da parte dei sussidi di povertà che molti dei partecipanti avrebbero percepito se non fossero stati coinvolti nell'esperimento.

Per il disegno e la realizzazione della *NSWD* viene addirittura costituita una nuova organizzazione *non-profit*, la *Manpower Demonstration Research Corporation* (*MDRC*) dal momento che nella metà degli anni settanta negli Stati Uniti non esisteva ancora un'organizzazione con le caratteristiche adeguate per svolgere una valutazione di questo tipo.

La *NSWD* si svolge in 14 centri urbani, dislocati in altrettanti Stati e in ciascun sito viene costituita un'organizzazione *non-profit* con il compito di gestire l'intervento mentre la *MDRC* coordina le operazioni a livello nazionale e la gestione della valutazione.

Le attività lavorative svolte dai partecipanti riguardano per la metà i servizi (assistenza negli asili, lavoro presso distributori di benzina, tipografie,…) un quarto nel settore edilizio e un quasi 10% nel settore manifatturiero.

L'intervento, che si attua nell'arco di quattro anni, coinvolge circa 10.000 individui che prendono parte al lavoro cosiddetto "protetto" per circa sette mesi ciascuno (Martini, 1997).

L'obiettivo della valutazione di tale politica e quello di rispondere ai seguenti quesiti:

- Il *supported work* aumenta la partecipazione al lavoro degli esposti e diminuisce la loro dipendenza dai sussidi pubblici?
- Quali gruppi traggono maggior beneficio da tale politica?
- Qual è il rapporto tra costi sostenuti e benefici ottenuti?

- Quali strategie di implementazione funzionano meglio?
- Quali caratteristiche della politica hanno maggiore impatto sul comportamento e sulla *performance* dei partecipanti?

L'ampio spettro di questi interrogativi va ovviamente oltre alla semplice individuazione dell'impatto diretto dell'intervento, bensì copre anche questioni legate all'implementazione della politica stessa: il nostro interesse sarà comunque focalizzato all'analisi dell'impatto che riguarda soprattutto il primo dei quesiti sopra elencati.

## 3.2 Il processo di selezione

La valutazione d'impatto viene condotta tramite assegnazione casuale al gruppo degli esposti e al gruppo dei non esposti: nei siti in cui la *demonstration* è condotta una persona su due che fa domanda per partecipare a questa esperienza viene esclusa dalla partecipazione mediante sorteggio e va a formare quello che sarà il gruppo di controllo. Questa assegnazione attraverso randomizzazione non pone particolare problemi dal punto di vista etico e legale in quanto il prendere parte al *supported work* non rappresenta alcun tipo di diritto per chi fa domanda: se non vi fosse infatti la volontà esplicita di condurre l'esperimento, la politica stessa non avrebbe luogo.

Le uniche difficoltà incontrate riguardano gli operatori incaricati di selezionare le unità che, oltre a non trarre nessun vantaggio dall'essere coinvolti dall'esperimento, devono reclutare dapprima

un numero doppio di partecipanti e poi comunicare alla metà di questi che sono stati scartati mediante sorteggio.

Tale processo di selezione fa sì che i membri del gruppo di controllo siano del tutto simili, in termini di caratteristiche osservabili e non, ai membri del gruppo che partecipa all'intervento e dunque le differenze osservate sono da imputare esclusivamente all'esposizione all'intervento. Dati i limitati fondi disponibili però, rispetto ai circa 10.000 partecipanti al *supported work*, vengono coinvolti nella valutazione dell'impatto solamente 6.600 soggetti, 3.200 dei quali assegnati al gruppo di trattamento mentre i restanti 3.400 al gruppo di controllo.

A prescindere dal gruppo, tutti i soggetti vengono intervistati in cinque occasioni successive partendo con un'intervista immediatamente precedente all'assegnazione casuale, seguita da un'intervista periodica ogni nove mesi: di queste è importante notare che la seconda è quella che viene svolta durante il periodo di lavoro "protetto" mentre l'ultima si svolge a circa tre anni di distanza dall'inizio dell'intervento.

In realtà non tutti i soggetti coinvolti nel progetto riescono a completare il ciclo di interviste previste: all'ultima intervista, ad esempio, risposero il 72% dei trattati e il 68% dei non trattati. Questo fenomeno di perdita di unità campionarie nel tempo viene definito *attrition*: se tale perdita avviene in maniera non casuale si rischia di non avere più a disposizione informazioni riguardanti soggetti "particolari" e dunque rilevanti ai fine delle stime;

tuttavia l'assegnazione casuale ai due gruppi rende tale effetto poco influente sulle successive analisi (LaLonde, 1986).

Le informazioni raccolte si riferiscono ovviamente alle variazioni delle condizioni e dei comportamenti che l'intervento vuole modificare ossia la partecipazione al lavoro, il reddito percepito, lo stato di povertà, gli eventuali sussidi percepiti, i possibili problemi con la giustizia e l'assunzione di stupefacenti.

Nella tabella 1 che segue presentiamo le caratteristiche principali dei soggetti trattati e non trattati divisi per maschi e femmine.

**Tabella 1** – Medie campionarie e deviazioni standard dei redditi pre-trattamento e di altre caratteristiche per le femmine (*AFDC Participants*) e per i maschi (*Male Participants*)

| | Full National Supported Work Sample | | | |
|---|---|---|---|---|
| | AFDC Participants | | Male Participants | |
| Variable | Treatments | Controls | Treatments | Controls |
| Age | 33.37 | 33.63 | 24.49 | 23.99 |
| | (7.43) | (7.18) | (6.58) | (6.54) |
| Years of School | 10.30 | 10.27 | 10.17 | 10.17 |
| | (1.92) | (2.00) | (1.75) | (1.76) |
| Proportion | .70 | .69 | .79 | .80 |
| High School Dropouts | (.46) | (.46) | (.41) | (.40) |
| Proportion Married | .02 | .04 | .14 | .13 |
| | (.15) | (.20) | (.35) | (.35) |
| Proportion Black | .84 | .82 | .76 | .75 |
| | (.37) | (.39) | (.43) | (.43) |
| Proportion Hispanic | .12 | .13 | .12 | .14 |
| | (.32) | (.33) | (.33) | (.35) |
| Real Earnings | $393 | $395 | 1472 | 1558 |
| 1 year Before | (1,203) | (1,149) | (2656) | (2961) |
| Training | [43] | [41] | [58] | [63] |
| Real Earnings | $854 | $894 | 2860 | 3030 |
| 2 years Before | (2,087) | (2,240) | (4729) | (5293) |
| Training | [74] | [79] | [104] | [113] |
| Hours Worked | 90 | 92 | 278 | 274 |
| 1 year Before | (251) | (253) | (466) | (458) |
| Training | [9] | [9] | [10] | [10] |
| Hours Worked | 186 | 188 | 458 | 469 |
| 2 years Before | (434) | (450) | (654) | (689) |
| Training | [15] | [16] | [14] | [15] |
| Month of Assignment | −12.26 | −12.30 | −16.08 | −15.91 |
| (Jan. 78 = 0) | (4.30) | (4.23) | (5.97) | (5.89) |
| Number of Observations | 800 | 802 | 2083 | 2193 |

*Note:* The numbers shown in parentheses are the standard deviations and those in the square brackets are the standard errors.

Con *AFDC participants* si intendono le madri non sposate ossia il primi dei quattro gruppi citati prima mentre in *male participants* rientrano le altre tre categorie.

Come era prevedibile, la randomizzazione fa sì che i valori assunti nei gruppi di trattamento e controllo siano pressoché gli stessi tanto che nessuna differenza risulta essere statisticamente significativa.

## 3.3 I risultati sperimentali ottenuti

Vista la molteplicità di informazioni raccolte durante le interviste (partecipazione al lavoro, reddito, stato di povertà, eventuali sussidi,…), l'ente che si è occupato dell'implementazione e valutazione della politica ha ritenuto opportuno valutare gli effetti di quest'ultima su diversi aspetti. Tra questi presentiamo di seguito quelli riferiti agli effetti registrati sulle ore lavorate, sui redditi citando brevemente anche una riflessione sui costi-benefici derivanti dalla partecipazione alla politica.

### 3.3.1 L'impatto in termini di ore lavorate

Un primo risultato interessante da mettere in risalto è l'impatto in termini di ore mensili lavorate registrato in particolare per due dei gruppi visti sopra: le giovani madri non sposate che vivono del sussidio di povertà e per i giovani *drop-outs* della scuola dell'obbligo.

L'interesse non è rivolto solo alle stime dell'impatto ma anche al fatto che, specialmente in questi due casi, i risultati ottenuti

mostrano che senza un gruppo di controllo capita di attribuire alla politica miglioramenti che in realtà sono dovuti a processi di cambiamento indipendenti dalla politica.

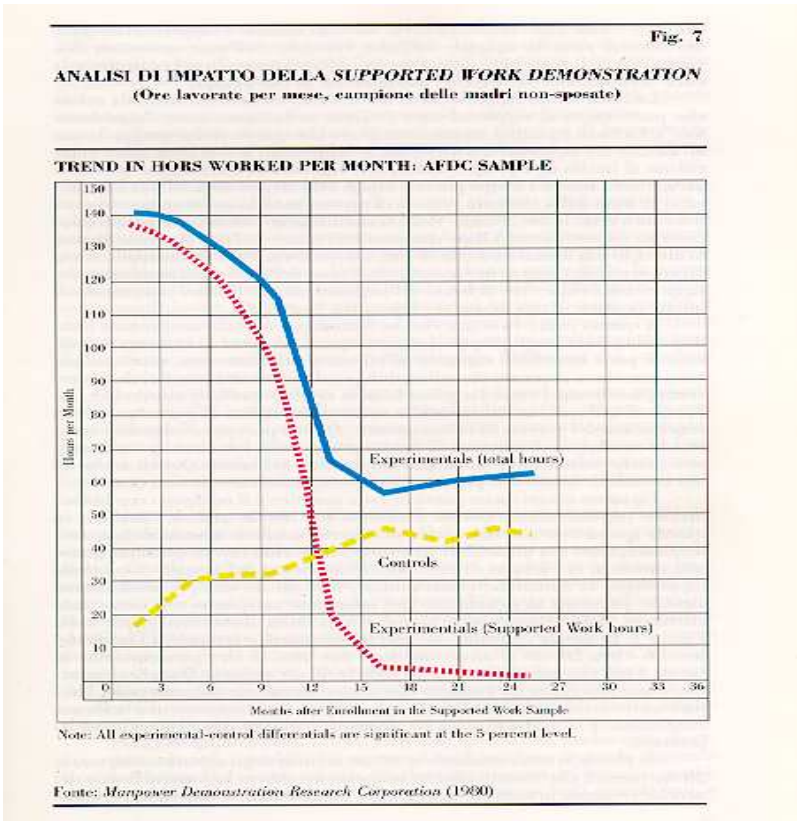
Dal punto di vista statistico, per stimare l'impatto si ricorre all'uso di medie campionarie ossia al numero medio di ore mensili lavorate, e in particolare possiamo definire con $h_t^T$ e $h_t^{NT}$ i due risultati potenziali al t-esimo mese dove:

$$\bar{h}_t^T \approx N\left( E\left[h_t^T \middle| P'\right], \frac{1}{n^T} \operatorname{var}\left[h_t^T \middle| P'\right]\right)$$

$$\bar{h}_t^{NT} \approx N\left( E\left[h_t^{NT} \middle| P'\right], \frac{1}{n^{NT}} \operatorname{var}\left[h_t^{NT} \middle| P'\right]\right)$$

Procediamo ora con l'analisi dei risultati ottenuti. Riportiamo di seguito il grafico riferito al gruppo delle madri non sposate:

**Figura 1**: Impatto in termini di ore lavorate per le madri non sposate




la linea continua di colore blu si riferisce al totale delle ore lavorate in media al mese da coloro che partecipano alla politica ossia dal gruppo degli esposti (definiti nel grafico come "*experimental*"); la linea a puntini di colore rosso si riferisce invece

al numero medio di ore lavorate durante il periodo di *supported work* sempre dal gruppo degli esposti e infine la linea tratteggiata gialla riguarda le ore di lavoro medie al mese effettuate dal gruppo di controllo, ossia l'evento controfattuale per cui le madri non sposate non partecipano al lavoro "protetto".

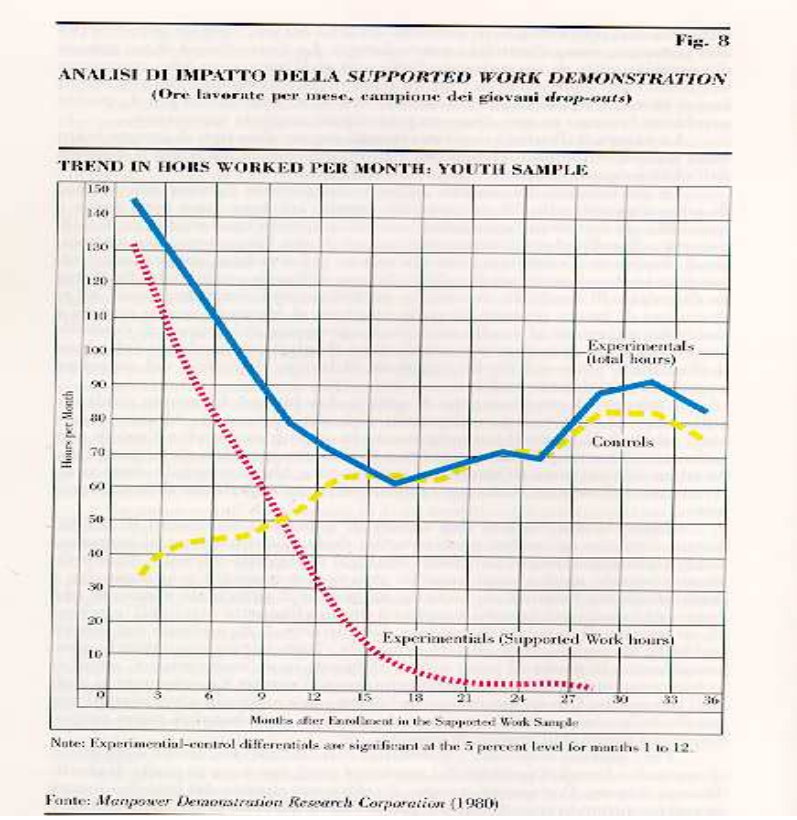
All'inizio del periodo le linee rossa e blu quasi coincidono al livello di 140 ore mensili, che vuol dire che la maggior parte dei partecipanti lavora a tempo pieno e svolge una minima attività lavorativa al di fuori della struttura protetta. Dopo circa 12 mesi il progetto di *supported work* termina e dunque le ore lavorate (rappresentate dalla linea rossa) tendono a 0 mentre le ore totali lavorate tendono a stabilizzarsi sulle 60 ore mensili (cioè circa 15 ore settimanali) dopo due anni dall'inizio dell'esperimento. Il salto da 140 a sole 60 ore lavorate inizialmente potrebbe essere interpretato come un fallimento della politica però bisogna ricordare che quelle 60 ore sono interamente lavorate al di fuori della struttura protetta, da persone le cui condizioni di partenza li mettevano ai margini del mercato del lavoro, il che rappresenta lo scopo fondamentale della politica ossia facilitare l'inserimento di questi soggetti in una situazione lavorativa "normale".

Procedendo nell'analisi, la situazione controfattuale è rappresentata dal gruppo di controllo che descrive ciò che sarebbe presumibilmente accaduto ai beneficiari della politica se questa non fosse mai stata realizzata. Dal grafico risulta che all'inizio del periodo le ore lavorate sono in media 20 ma dopo due anni balzano addirittura a 40: quindi se misurato come differenza tra

risultato osservato e situazione controfattuale, l'impatto netto della partecipazione al lavoro "protetto" ad un anno circa dalla sua conclusione è in media di circa 20 ore mensili, cioè un aumento del 50%.

Se invece si mettessero a confronto le 60 ore mensili degli *experimentals* con le 20 ore mensili che costoro presumibilmente avrebbero lavorato all'inizio del periodo (essendo in tutto simili al gruppo di controllo) si concluderebbe che l'impatto rilevato segnala un aumento da 20 a 60 ore, ossia un aumento del 200%, e perciò si sovrastimerebbe l'efficacia della politica. Concludendo ad un anno dal termine del periodo di lavoro "protetto" i partecipanti lavorano circa il 50% in più di quanto avrebbero lavorato se non avessero preso parte a questa politica.

**Figura 2**: Impatto in termini di ore lavorate per i giovani *drop-outs*




Passiamo ora ad analizzare i risultati che si riferiscono ai giovani *drop-outs* della scuola dell'obbligo.

Le ore lavorate dal gruppo di controllo all'inizio delle rilevazioni sono all'incirca

pari a 30 mentre tre anni dopo sono salite quasi a 80: il fatto interessante in questo caso è che chi ha partecipato al periodo di lavoro "protetto", una volta che questo si è concluso, segue lo stesso percorso di chi non vi ha preso parte. Dunque in assenza di gruppo di controllo si sarebbe portati ad affermare che la politica ha avuto un notevole successo ma in realtà l'aumento netto nella partecipazione al lavoro è tale per cui l'efficacia del *supported work* in questo caso è essenzialmente nulla (Martini, 1997).

### 3.3.2 L'impatto in termini di reddito

Concentriamoci ora sulle stime sperimentali dei redditi facendo riferimento alla suddivisione impiegata su LaLonde (1986): femmine e maschi vengono infatti analizzati separatamente (le femmine vengono riportate nella tabella 2 mentre i maschi nella tabella 3).

In particolare focalizziamo l'attenzione sulle prime due colonne delle tabelle che fanno riferimento alle stime sperimentali (le stime riportate nelle colonne successive verranno utili nel prossimo paragrafo quando si accennerà alle stime non sperimentali).

Riportiamo dopo ciascuna tabella anche un grafico riepilogativo dell'andamento dei redditi annuali per le femmine e i maschi trattati e non trattati.

**Tabella 2** – Redditi annuali per i gruppi di trattamento e controllo sperimentali e per gli otto gruppi di controllo non sperimentali ricavati da PSID e CPS-SSA nel caso delle **femmine**

| | | | Comparison Group[a,b] | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Year | Treat-ments | Controls | *PSID*-1 | *PSID*-2 | *PSID*-3 | *PSID*-4 | *CPS-SSA*-1 | *CPS-SSA*-2 | *CPS-SSA*-3 | *CPS-SSA*-4 |
| 1975 | $895 | $877 | 7,303 | 2,327 | 937 | 6,654 | 7,788 | 3,748 | 4,575 | 2,049 |
| | (81) | (90) | (317) | (286) | (189) | (428) | (63) | (250) | (135) | (333) |
| 1976 | $1,794 | $646 | 7,442 | 2,697 | 665 | 6,770 | 8,547 | 4,774 | 3,800 | 2,036 |
| | (99) | (63) | (327) | (317) | (157) | (463) | (65) | (302) | (128) | (337) |
| 1977 | $6,143 | $1,518 | 7,983 | 3,219 | 891 | 7,213 | 8,562 | 4,851 | 5,277 | 2,844 |
| | (140) | (112) | (335) | (376) | (229) | (484) | (68) | (317) | (153) | (450) |
| 1978 | $4,526 | $2,885 | 8,146 | 3,636 | 1,631 | 7,564 | 8,518 | 5,343 | 5,665 | 3,700 |
| | (270) | (244) | (339) | (421) | (381) | (480) | (72) | (365) | (166) | (593) |
| 1979 | $4,670 | $3,819 | 8,016 | 3,569 | 1,602 | 7,482 | 8,023 | 5,343 | 5,782 | 3,733 |
| | (226) | (208) | (334) | (381) | (334) | (462) | (73) | (371) | (170) | (543) |
| Number of Observations | 600 | 585 | 595 | 173 | 118 | 255 | 11,132 | 241 | 1,594 | 87 |

[a] The Comparison Groups are defined as follows: *PSID*-1: All female household heads continuously from 1975 through 1979, who were between 20 and 55-years-old and did not classify themselves as retired in 1975; *PSID*-2: Selects from the *PSID*-1 group all women who received AFDC in 1975; *PSID*-3: Selects from the *PSID*-2 all women who were not working when surveyed in 1976; *PSID*-4: Selects from the *PSID*-1 group all women with children, none of whom are less than 5-years-old; *CPS-SSA* – 1: All females from Westat *CPS-SSA* sample; *CPS-SSA*-2: Selects from *CPS-SSA*-1 all females who received AFDC in 1975; *CPS-SSA*-3: Selects from *CPS-SSA*-1 all females who were not working in the spring of 1976; *CPS-SSA*-4: Selects from *CPS-SSA*-2 all females who were not working in the spring of 1976.

[b] All earnings are expressed in 1982 dollars. The numbers in parentheses are the standard errors. For the NSW treatments and controls, the number of observations refer only to 1975 and 1979. In the other years there are fewer observations, especially in 1978. At the time of the resurvey in 1979, treatments had been out of Supported Work for an average of 20 months.

**Figura 3** – Andamento dei redditi per le femmine nel periodo 1975-79

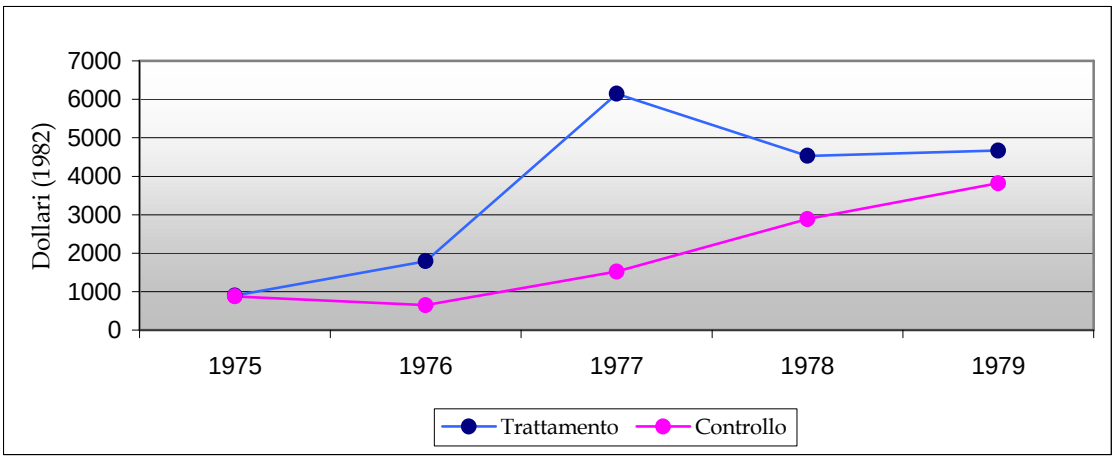


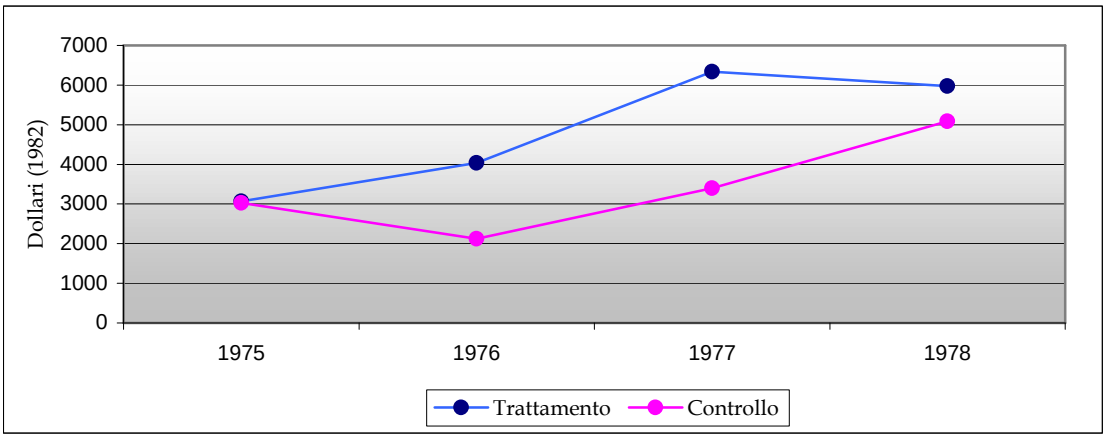
**Tabella 3** – Redditi annuali per i gruppi di trattamento e controllo sperimentali e per i sei gruppi di controllo non sperimentali ricavati da PSID e CPS-SSA nel caso dei maschi

| | | | Comparison Group[a,b] | | | | | |
|---|---|---|---|---|---|---|---|---|
| Year | Treatments | Controls | *PSID*-1 | *PSID*-2 | *PSID*-3 | *CPS-SSA*-1 | *CPS-SSA*-2 | *CPS-SSA*-3 |
| 1975 | $3,066 | $3,027 | 19,056[a] | 7,569 | 2,611 | 13,650 | 7,387 | 2,729 |
| | (283) | (252) | (272) | (568) | (492) | (73) | (206) | (197) |
| 1976 | $4,035 | $2,121 | 20,267 | 6,152 | 3,191 | 14,579 | 6,390 | 3,863 |
| | (215) | (163) | (296) | (601) | (609) | (75) | (187) | (267) |
| 1977 | $6,335 | $3,403 | 20,898 | 7,985 | 3,981 | 15,046 | 9,305 | 6,399 |
| | (376) | (228) | (296) | (621) | (594) | (76) | (225) | (398) |
| 1978 | $5,976 | $5,090 | 21,542 | 9,996 | 5,279 | 14,846 | 10,071 | 7,277 |
| | (402) | (227) | (311) | (703) | (686) | (76) | (241) | (431) |
| Number of Observations | 297 | 425 | 2,493 | 253 | 128 | 15,992 | 1,283 | 305 |

[a]The Comparison Groups are defined as follows: *PSID*-1: All male household heads continuously from 1975 through 1978, who were less than 55-years-old and did not classify themselves as retired in 1975; *PSID*-2: Selects from the *PSID*-1 group all men who were not working when surveyed in the spring of 1976; *PSID*-3: Selects from the *PSID*-1 group all men who were not working when surveyed in either spring of 1975 or 1976; *CPS-SSA*-1: All males based on Westat's criteria, except those over 55-years-old; *CPS-SSA*-2: Selects from *CPS-SSA*-1 all males who were not working when surveyed in March 1976; *CPS-SSA*-3: Selects from the *CPS-SSA*-1 unemployed males in 1976 whose income in 1975 was below the poverty level.

[b]All earnings are expressed in 1982 dollars. The numbers in parentheses are the standard errors. The number of observations refer only to 1975 and 1978. In the other years there are fewer observations. The sample of treatments is smaller than the sample of controls because treatments still in Supported Work as of January 1978 are excluded from the sample, and in the young high school target group there were by design more controls than treatments.

**Figura 4** – Andamento dei redditi per i maschi nel periodo 1975-78



Tenuto conto del processo di selezione casuale, i redditi sono chiaramente gli stessi prima dell'inizio dell'intervento, tendendo poi a divergere durante e immediatamente dopo la conclusione del periodo di lavoro "protetto": in particolare quelli del gruppo dei non trattati si mantengono su un livello nettamente inferiore rispetto a quello dei trattati. L'ultimo anno riporta quelli che sono i valori riferiti agli effetti della politica di più lungo periodo: sottraiamo al reddito annuale post-intervento dei trattati quello del gruppo di controllo sperimentale e otterremo, a circa tre anni di distanza dall'inizio del periodo di "lavoro protetto", nel caso delle femmine un impatto netto di $851 ($4.670 - $3.819) mentre per i maschi di $886 ($5.976 - $5.090).

Interessante è a questo punto riportare una breve riflessione su quelli che sono stati i costi-benefici della partecipazione al *supported work*. Tale analisi può essere condotta solo dopo che è stata effettuata l'analisi dell'impatto in quanto il principale elemento che entra nel calcolo dei benefici è proprio l'impatto netto dell'intervento in termini di aumento di reddito, diminuzione di dipendenza da sussidi pubblici e diminuzione di attività criminale. Tutte queste dimensioni poi vanno monetizzate, in modo da poter essere aggregate in un'unica misura, attualizzate al presente e confrontate con i costi dell'intervento.

Come abbiamo già affermato la realizzazione del *supported work* per le giovani madri non sposate si rivela un buon investimento in quanto il valore attualizzato dei benefici netti varia da 2.700 a 9.000 dollari; la stessa conclusione non si può trarre per i giovani

*drop-outs* della scuola dell'obbligo dato che i benefici prodotti dall'inserimento nel lavoro "protetto" non sono in grado di giustificarne i costi: il valore attualizzato dei benefici netti in questo caso è negativo e varia tra – 4.000 e – 250 dollari (Martini, 1997).

### 3.4 I risultati non sperimentali ottenuti con stimatori *difference-in-differences*

La valutazione dell'impatto condotta in ambito sperimentale può a questo punto essere utilizzata come *benchmark* per confrontare l'impatto della stessa politica ottenuto con metodi non sperimentali mediante una modellizzazione di tipo econometrico; in particolare analizzeremo e confronteremo quelli che sono i risultati in termini di reddito per il gruppo di trattamento e per i gruppi di controllo opportunamente scelti.

E importante ricordare che, adottando tecniche non sperimentali, il gruppo di controllo preso come riferimento nella valutazione non avrà presumibilmente le stesse caratteristiche osservabili del gruppo di trattamento: ogni valutazione non sperimentale deve tenere ben presente tale inconveniente, tuttavia se il modello econometrico usato è correttamente specificato le stime non sperimentali che si generano dovrebbero pressoché coincidere con quelle generate nel caso sperimentale.

Il primo passo da compiere è dunque quello di determinare un adeguato gruppo di controllo: in realtà ne verranno considerati molteplici distinti per maschi e femmine e tratti da *Panel Study of*

*Income Dynamics* (PSID) e da *Current Population Survey-Social Security Administration File* (CPS-SSA).

Il primo (PSID) riguarda un'indagine nazionale di tipo longitudinale che coinvolge circa 8.000 famiglie americane che dal 1968 vengono monitorate per raccogliere informazioni di tipo economico, sanitario e sociale. Il secondo (CPS) invece si riferisce ad un indagine, condotta sempre negli Stati Uniti, sulle forze lavoro abbinata ad un insieme di informazioni relative alla situazione socio-sanitaria, lavorativa e fiscale dei cittadini americani (SSA).

La composizione per le femmine può essere così riassunta:

- PSID-1: comprende tutte le femmine capofamiglia nel periodo 1975-1979, tra i 20 e i 55 anni che non risultano ancora essere andate in pensione nel 1975;
- PSID-2: ricavato da PSID-1, comprende solo le femmine che hanno ricevuto il sussidio di povertà (AFDC) nel 1975;
- PSID-3: ricavato da PSID-2, comprende solo le femmine che nel 1976 risultavano disoccupate;
- PSID-4: ricavato da PSID-1, comprende solo le femmine con figli di età superiore ai 5 anni;
- CPS-SSA-1: include tutte le femmine con meno di 55 anni;
- CPS-SSA-2: determinato da CPS-SSA-1, comprende solo le femmine che hanno percepito il sussidio di povertà (AFDC) nel 1975;

- CPS-SSA-3: ricavato da CPS-SSA-1, comprende tutte le femmine che risultavano disoccupate nella primavera del 1976;
- CPS-SSA-4: ricavato questa volta da CPS-SSA-2, include sempre tutte le femmine che risultavano disoccupate nella primavera del 1976.

La scelta e la composizione dei gruppi di confronto per i maschi invece è così definita:

- PSID-1: comprende tutte i maschi capofamiglia nel periodo 1975-1978, con meno di 55 anni che non risultano ancora essere andati in pensione nel 1975;
- PSID-2: ricavato da PSID-1, comprende solo i maschi che dichiaravano di essere disoccupati nella primavera del 1976;
- PSID-3: ricavato da PSID-1, comprende solo i maschi disoccupati sia nella primavera del 1975 che in quella del 1976;
- CPS-SSA-1: include tutti i maschi con meno di 55 anni;
- CPS-SSA-2: determinato da CPS-SSA-1, comprende solo i maschi disoccupati nel marzo del 1976;
- CPS-SSA-3: ricavato da CPS-SSA-1, comprende i maschi disoccupati nel 1976 i cui redditi nel 1975 risultavano sotto la soglia di povertà.

Rimandiamo a questo punto alle due tabelle precedentemente riportate che riassumono inoltre i redditi annuali per le femmine

(tabella 2) e per i maschi (tabella 3) distinti per tutti i gruppi di controllo appena descritti.

La prima evidenza significativa che si nota riguarda la numerosità campionaria che soprattutto nel caso di CPS-SSA-1 è molto distante da quella del gruppo di confronto sperimentale (infatti per le femmine è di 11.132 contro le 585 osservazioni sperimentali mentre per i maschi è di 15.992 contro le 425 osservazioni sperimentali).

Oltre dunque alla numerosità delle unità osservate, il motivo per il quale si ricavano tutti questi sottocampioni è evidente osservando i valori assunti dai redditi annuali nelle due tabelle: le differenze con il caso sperimentale sono piuttosto marcate specialmente per i gruppi con più osservazioni. Se guardiamo i redditi registrati nell'ultimo anno vedremo che per le femmine il gruppo PSID-1 riporta un reddito di $8.016 mentre il gruppo CPS-SSA-1 riporta $8.023 contro i $3.819 del gruppo di controllo sperimentale. La stessa cosa vale per l'ultimo anno ricavato per i maschi: i valori per PSID-1 e CPS-SSA-1 sono rispettivamente $21.542 e $14.846 contro i $5.090 del gruppo di controllo sperimentale. Il progressivo ricorso ad altri sottocampioni permette dunque di ridurre queste differenze in termini di reddito fra i gruppi di trattamento e controllo a costo però di una riduzione della numerosità campionaria: nell'ultimo anno per le femmine PSID-2 ($3.569) e CPS-SSA-4 ($3.733) sono quelli più prossimi al valore del gruppo di controllo sperimentale ($3.819). Per i maschi invece sono i

gruppi PSID-3 ($5.279) e CPS-SSA-3 ($7.227) che meglio si accostano ai valori sperimentali.

Una volta determinati i gruppi di confronto procediamo alla specificazione del modello.

Ipotizziamo che per ogni soggetto *i* il reddito al tempo *t* dipenda una variabile *D* che indica la partecipazione o meno alla politica al tempo *s+1*, da un vettore di caratteristiche individuali osservabili *X* (età, formazione scolastica e razza) e da una componente non osservabile che include un termine d'errore serialmente correlato $\varepsilon$ (vedi (2)), delle caratteristiche *b* relative solo al soggetto *i* e delle caratteristiche *n*, relative solo al tempo *t* ossia:

$$y_{it} = \delta D_i + \beta X_{it} + b_i + n_t + \varepsilon_{it} \qquad (1)$$

$$\varepsilon_{it} = \rho \varepsilon_{it-1} + \nu_{it} \qquad (2)$$

In particolare la variabile che indica la partecipazione alla politica:

$D_i = 1$ se $d_{is} > 0$ mentre $D_i = 0$ se $d_{is} < 0$

dove:

$$d_{is} = y_{is} + \gamma Z_{is} + \eta_{is} \qquad (3)$$

perciò la variabile $d_{is}$ dipenderà da un vettore di caratteristiche osservabili *Z*, dal reddito al tempo corrente e da una componente d'errore $\eta$.

Una restrizione da imporre per la corretta specificazione del modello è che le componenti non osservabili nell'equazione del reddito (1) e della partecipazione alla politica (3) siano incorrelate:

$$(b_i + n_t + \varepsilon_{it}) \perp \eta_{is} \qquad (4)$$

Uno stimatore *difference-in-differences* abbiamo detto essere in grado di fornire una buona stima dell'impatto se esistono delle informazioni longitudinali sui partecipanti e non partecipanti nel periodo precedente e successivo alla realizzazione della politica e le differenze *prima-dopo* in assenza di trattamento sono le stesse per i due gruppi. Infatti se è vale tale assunzione questo metodo è il punto di partenza per valutare programmi in cui si abbia uno o più gruppi di controllo e informazioni sull'andamento della variabile obiettivo prima dell'implementazione del programma da valutare: si tratta perciò di procedere, come abbiamo visto, con una modellazione delle dinamiche del fenomeno la cui stima permette poi il confronto tra i gruppi di trattamento e controllo.

**Tabella 4** – Stime dell'impatto per il gruppo sperimentale e gli otto gruppi ricavati da PSID e CPS-SSA nel caso delle femmine

| Name of Comparison Group[d] | Comparison Group Earnings Growth 1975–79 (1) | NSW Treatment Earnings Less Comparison Group Earnings: Pre-Training Year, 1975 Unadjusted (2) | NSW Treatment Earnings Less Comparison Group Earnings: Pre-Training Year, 1975 Adjusted[c] (3) | NSW Treatment Earnings Less Comparison Group Earnings: Post-Training Year, 1979 Unadjusted (4) | NSW Treatment Earnings Less Comparison Group Earnings: Post-Training Year, 1979 Adjusted[c] (5) | Difference in Differences: Difference in Earnings Growth 1975–79 Treatments Less Comparisons: Without Age (6) | Difference in Differences: Difference in Earnings Growth 1975–79 Treatments Less Comparisons: With Age (7) | Unrestricted Difference in Differences: Quasi Difference in Earnings Growth 1975–79: Unadjusted (8) | Unrestricted Difference in Differences: Quasi Difference in Earnings Growth 1975–79: Adjusted[c] (9) | Controlling for All Observed Variables and Pre-Training Earnings: Without AFDC (10) | Controlling for All Observed Variables and Pre-Training Earnings: With AFDC (11) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2,942 | −17 | −22 | 851 | 861 | 833 | 883 | 843 | 864 | 854 | – |
| Controls | (220) | (122) | (122) | (307) | (306) | (323) | (323) | (308) | (306) | (312) | |
| *PSID*-1 | 713 | −6,443 | −4,882 | −3,357 | −2,143 | 3,097 | 2,657 | 1746 | 1,354 | 1664 | 2,097 |
| | (210) | (326) | (336) | (403) | (425) | (317) | (333) | (357) | (380) | (409) | (491) |
| *PSID*-2 | 1,242 | −1,467 | −1,515 | 1,090 | 870 | 2,568 | 2,392 | 1,764 | 1,535 | 1,826 | – |
| | (314) | (216) | (224) | (468) | (484) | (473) | (481) | (472) | (487) | (537) | |
| *PSID*-3 | 665 | −77 | −100 | 3,057 | 2,915 | 3,145 | 3,020 | 3,070 | 2,930 | 2,919 | – |
| | (351) | (202) | (208) | (532) | (543) | (557) | (563) | (531) | (543) | (592) | |
| *PSID*-4 | 928 | −5,694 | −4,976 | −2,822 | −2,268 | 2,883 | 2,655 | 1,184 | 950 | 1,406 | 2,146 |
| | (311) | (306) | (323) | (460) | (491) | (417) | (434) | (483) | (503) | (542) | (652) |
| *CPS-SSA*-1 | 233 | −6,928 | −5,813 | −3,363 | −2,650 | 3,578 | 3,501 | 1,214 | 1,127 | 536 | 1,041 |
| | (64) | (272) | (309) | (320) | (365) | (280) | (282) | (272) | (309) | (349) | (503) |
| *CPS-SSA*-2 | 1,595 | −2,888 | −2,332 | −683 | −240 | 2,215 | 2,068 | 447 | 620 | 665 | – |
| | (360) | (204) | (256) | (428) | (536) | (438) | (446) | (468) | (554) | (651) | |
| *CPS-SSA*-3 | 1,207 | −3,715 | −3,150 | −1,122 | −812 | 2,603 | 2,615 | 814 | 784 | −99 | 1,246 |
| | (166) | (226) | (325) | (311) | (452) | (307) | (328) | (305) | (429) | (481) | (720) |
| *CPS-SSA*-4 | 1,684 | −1,189 | −780 | 926 | 756 | 2,126 | 1,833 | 1,222 | 952 | 827 | – |
| | (524) | (249) | (283) | (630) | (716) | (654) | (663) | (637) | (717) | (814) | |

[a] The columns above present the estimated training effect for each econometric model and comparison group. The dependent variable is earnings in 1979. Based on the experimental data, an unbiased estimate of the impact of training presented in col. 4 is $851. The first three columns present the difference between each comparison group's 1975 and 1979 earnings and the difference between the pre-training earnings of each comparison group and the NSW treatments.

[b] Estimates are in 1982 dollars. The numbers in parentheses are the standard errors.

[c] The exogenous variables used in the regression adjusted equations are age, age squared, years of schooling, high school dropout status, and race.

[d] See Table 2 for definitions of the comparison groups.

**Tabella 5** – Stime dell'impatto per il gruppo sperimentale e i sei gruppi ricavati da PSID e CPS-SSA nel caso dei maschi

| Name of Comparison Group[d] | Comparison Group Earnings Growth 1975–78 (1) | NSW Treatment Earnings Less Comparison Group Earnings: Pre-Training Year, 1975 Unadjusted (2) | Pre-Training Year, 1975 Adjusted[c] (3) | Post-Training Year, 1978 Unadjusted (4) | Post-Training Year, 1978 Adjusted[c] (5) | Difference in Differences: Difference in Earnings Growth 1975–78 Treatments Less Comparisons Without Age (6) | With Age (7) | Unrestricted Difference in Differences: Quasi Difference in Earnings Growth 1975–78 Unadjusted (8) | Adjusted[c] (9) | Controlling for All Observed Variables and Pre-Training Earnings (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| Controls | $2,063 | $39 | $−21 | $886 | $798 | $847 | $856 | $897 | $802 | $662 |
| | (325) | (383) | (378) | (476) | (472) | (560) | (558) | (467) | (467) | (506) |
| *PSID*-1 | $2,043 | −$15,997 | −$7,624 | −$15,578 | −$8,067 | $425 | −$749 | −$2,380 | −$2,119 | −$1,228 |
| | (237) | (795) | (851) | (913) | (990) | (650) | (692) | (680) | (746) | (896) |
| *PSID*-2 | $6,071 | −$4,503 | −$3,669 | −$4,020 | −$3,482 | $484 | −$650 | −$1,364 | −$1,694 | −$792 |
| | (637) | (608) | (757) | (781) | (935) | (738) | (850) | (729) | (878) | (1024) |
| *PSID*-3 | ($3,322 | ($455 | $455 | $697 | −$509 | $242 | −$1,325 | $629 | −$552 | $397 |
| | (780) | (539) | (704) | (760) | (967) | (884) | (1078) | (757) | (967) | (1103) |
| *CPS-SSA*-1 | $1,196 | −$10,585 | −$4,654 | −$8,870 | −$4,416 | $1,714 | $195 | −$1,543 | −$1,102 | −$805 |
| | (61) | (539) | (509) | (562) | (557) | (452) | (441) | (426) | (450) | (484) |
| *CPS-SSA*-2 | $2,684 | −$4,321 | −$1,824 | −$4,095 | −$1,675 | $226 | −$488 | −$1,850 | −$782 | −$319 |
| | (229) | (450) | (535) | (537) | (672) | (539) | (530) | (497) | (621) | (761) |
| *CPS-SSA*-3 | $4,548 | $337 | $878 | −$1,300 | $224 | −$1,637 | −$1,388 | −$1,396 | $17 | $1,466 |
| | (409) | (343) | (447) | (590) | (766) | (631) | (655) | (582) | (761) | (984) |

[a] The columns above present the estimated training effect for each econometric model and comparison group. The dependent variable is earnings in 1978. Based on the experimental data an unbiased estimate of the impact of training presented in col. 4 is $886. The first three columns present the difference between each comparison group's 1975 and 1978 earnings and the difference between the pre-training earnings of each comparison group and the NSW treatments.
[b] Estimates are in 1982 dollars. The numbers in parentheses are the standard errors.
[c] The exogenous variables used in the regression adjusted equations are age, age squared, years of schooling, high school dropout status, and race.
[d] See Table 3 for definitions of the comparison groups.

Le tabelle 4 e 5 appena riportate presentano le diverse specificazioni adottate e i risultati suddivisi per maschi e femmine. La prima colonna delle due tabelle riporta semplicemente la crescita in termini di salario tra il 1975 e il 1978 (per i maschi) o 1979 (per le femmine) registrata per ogni gruppo di controllo; le due colonne successive riportano la differenza tra i salari di chi ha preso parte alla *NSWD* e i rispettivi gruppi di controllo nel 1975, dunque prima dell'attuazione della politica: ovviamente ci si deve aspettare che le differenze pre-trattamento siano molto piccole (le migliori in questo senso sono quelle sperimentali!).

La distinzione tra *unadjusted* e *adjusted* si riferisce al fatto che nel primo caso non viene incluso nell'equazione che determina il

salario il vettore di caratteristiche $X$ mentre nel secondo sì (le variabili incluse sono: *età, età*$^2$*, anni di scuola, stato scolastico, razza*).
Le colonne (4) e (5) invece presentano le differenze post-intervento fra i salari dei trattati e i rispettivi non trattati (la variabile obiettivo è dunque il reddito nel 1979 per le femmine e il reddito nel 1978 per i maschi). Prendendo il caso *unadjusted* vediamo che i gruppi che più si avvicinano al rispettivo valore sperimentale sono per le femmine PSID-2 ($1.090) e CPS-SSA-4 ($926) e PSID-3 ($697) per i maschi.
Continuando nella colonna (6) si riporta la differenza fra il valore della crescita salariale avvenuta tra il 1975 e il 1979 (1978 nel caso dei maschi) per i trattati e il rispettivo valore registrato per i non trattati (praticamente il valore riportato nella prima colonna). In questo caso si procede anche ad una stima che controlli la componente relativa all'età dell'i-esimo soggetto (vedi colonna (7)) dal momento che i soggetti trattati potrebbero ottenere sistematicamente una crescita salariale maggiore e più veloce rispetto ai non trattati per il fatto di essere più giovani. Tale componente in questo caso comporta che la variabile *dummy* relativa alla partecipazione alla politica sia correlata con la componente d'errore presente nell'equazione del salario. Dunque differenziando l'equazione si elimina questo effetto che potrebbe portare a delle stime distorte:

$$y_{it} - y_{is} = \delta D_i + \beta \cdot age_i + (\eta_t - \eta_s) + \varepsilon_{it} - \varepsilon_{is} \qquad (5)$$

In ogni caso comunque le stime ottenute sono parecchio lontane dai valori sperimentali di riferimento.

Le due colonne successive riportano invece sempre la differenza salariale riferita al periodo 1975-1979 (1798 per i maschi) tra trattati e non tenendo constante dapprima solo il livello salariale pre-intervento (colonna (8)) e poi anche il vettore di indicatori demografici (colonna (9)).

Nell'ultima colonna (ultime due nel caso delle femmine visto che si tiene conto del fatto che si abbia ricevuto o meno il sussidio di povertà) delle tabelle viene ricalcolato l'effetto della politica tenendo sotto controllo tutte le variabili osservabili (oltre alla variabili indicate prima vengono incluse altre variabili che indicano lo stato occupazionale nel 1976, se si percepiva appunto il sussidio di povertà AFDC nel 1975 (solo nel caso delle femmine), se si era sposati, se si risiedeva in un'area metropolitana con più di 100.000 abitanti e il numero di figli).

In riferimento alle femmine il gruppo CPS-SSA-4 ($827, vedi colonna (10)) bene si avvicina ai valori sperimentali mentre per i maschi un buon accostamento si ha con il gruppo PSID-3 ($629, vedi colonna (8)).

Sono state dunque testate diverse specificazioni ed è chiaro che la stima dell'impatto risulta essere notevolmente sensibile alle variabili incluse nell'equazione del salario e alla composizione del gruppo di controllo: com'era prevedibile man mano che la composizione del gruppo di confronto si restringe i risultati meglio si accostano a quelli che erano i risultati sperimentali visto

che i due gruppi tenderanno ad avere caratteristiche sempre più prossime.

E inoltre chiaro che nel caso delle femmine i valori sono più grandi rispetto al caso sperimentale, viceversa vale per i risultati per i maschi: tuttavia prima di prendere più seriamente tali stime sarebbe importante che gli stimatori basati sul modello specificato fossero consistenti tenuto conto dei dati relativi ai salari pre-intervento. Confrontare i valori dei redditi prima dell'inizio della politica può effettivamente non bastare; prendiamo ad esempio le stime riportate per le femmine per il gruppo PSID-3: i valori pre-trattamento sono prossimi a quelli sperimentali ma se andiamo a vedere la stima post-trattamento vediamo che questa è notevolmente distante dal *benchmark* sperimentale (abbiamo oltre $2.000 di differenza!).

Ricordiamo che le stime sperimentali usate qui come valore *benchmark* corrispondono a $851 nel caso delle femmine e $886 nel caso dei maschi (LaLonde, 1986).

## 3.5 I risultati non sperimentali ottenuti con i *propensity score*

I risultati ricavati mediante l'approccio sperimentale sono stati nuovamente impiegati da alcuni autori come *benchmark* per interpretare l'impatto della stessa politica ottenuto però questa volta in via non sperimentale mediante l'impiego dei *propensity score*.

Nelle analisi presentate precedentemente vengono analizzati separatamente i maschi e le femmine esposti alla *NSWD*: nelle

discussioni che seguiranno l'attenzione sarà rivolta invece specialmente ai maschi in quanto le stime ottenute risultano essere le più interessanti.

Come già precedentemente detto, la *NSWD* viene realizzata tra il 1975 e il 1978: questo lungo periodo presuppone però che gli individui che prendono parte subito al progetto abbiano presumibilmente caratteristiche diverse da coloro che entrano più tardi ( è il cosiddetto "*cohort phenomenon*"). Oltre a ciò bisogna tenere presente che, nelle analisi condotte con questo tipo di approccio non sperimentale, i dati relativi ai redditi sono da considerare espressi in anni, dunque si deve procedere ad un restringimento del numero degli esposti da considerare: saranno considerati solo coloro che avranno preso parte e concluso la politica tra il gennaio 1976 e il dicembre 1977 in modo così da tenere come variabile preintervento il reddito percepito nel 1975 (RE75). Il numero di osservazioni si riduce passando a 297 nel caso dei trattati e a 425 nel caso dei non trattati rispetto ai 2083 maschi trattati e ai 2193 maschi non trattati.

Utile alle analisi sarebbe anche avere un'ulteriore variabile preintervento riferita al reddito percepito nel 1974 (RE74): non per tutti coloro che hanno preso parte al programma tale informazione è disponibile perciò se si tiene conto anche di questo il campione di osservazioni si riduce ulteriormente comprendendone 185 per i trattati e 260 per i non trattati (Dehejia Wahba, 1999).

Nelle prime quattro righe della tabella che segue vengono riportate le caratteristiche preintervento dei vari gruppi appena descritti.

**Tabella 6 –** Medie campionarie delle caratteristiche dei gruppi di trattamento e di controllo

| | *No. of observations* | *Age* | *Education* | *Black* | *Hispanic* | *No degree* | *Married* | *RE74 (U.S. $)* | *RE75 (U.S. $)* |
|---|---|---|---|---|---|---|---|---|---|
| NSW/Lalonde:[a] | | | | | | | | | |
| Treated | 297 | 24.63 | 10.38 | .80 | .09 | .73 | .17 | | 3,066 |
| | | (.32) | (.09) | (.02) | (.01) | (.02) | (.02) | | (236) |
| Control | 425 | 24.45 | 10.19 | .80 | .11 | .81 | .16 | | 3,026 |
| | | (.32) | (.08) | (.02) | (.02) | (.02) | (.02) | | (252) |
| RE74 subset:[b] | | | | | | | | | |
| Treated | 185 | 25.81 | 10.35 | .84 | .059 | .71 | .19 | 2,096 | 1,532 |
| | | (.35) | (.10) | (.02) | (.01) | (.02) | (.02) | (237) | (156) |
| Control | 260 | 25.05 | 10.09 | .83 | .1 | .83 | .15 | 2,107 | 1,267 |
| | | (.34) | (.08) | (.02) | (.02) | (.02) | (.02) | (276) | (151) |
| Comparison groups:[c] | | | | | | | | | |
| PSID-1 | 2,490 | 34.85 | 12.11 | .25 | .032 | .31 | .87 | 19,429 | 19,063 |
| | | [.78] | [.23] | [.03] | [.01] | [.04] | [.03] | [991] | [1,002] |
| PSID-2 | 253 | 36.10 | 10.77 | .39 | .067 | .49 | .74 | 11,027 | 7,569 |
| | | [1.00] | [.27] | [.04] | [.02] | [.05] | [.04] | [853] | [695] |
| PSID-3 | 128 | 38.25 | 10.30 | .45 | .18 | .51 | .70 | 5,566 | 2,611 |
| | | [1.17] | [.29] | [.05] | [.03] | [.05] | [.05] | (686) | [499] |
| CPS-1 | 15,992 | 33.22 | 12.02 | .07 | .07 | .29 | .71 | 14,016 | 13,650 |
| | | [.81] | [.21] | [.02] | [.02] | [.03] | [.03] | [705] | [682] |
| CPS-2 | 2,369 | 28.25 | 11.24 | .11 | .08 | .45 | .46 | 8,728 | 7,397 |
| | | [.87] | [.19] | [.02] | [.02] | [.04] | [.04] | [667] | [600] |
| CPS-3 | 429 | 28.03 | 10.23 | .21 | .14 | .60 | .51 | 5,619 | 2,467 |
| | | [.87] | [.23] | [.03] | [.03] | [.04] | [.04] | [552] | [288] |

NOTE: Standard errors are in parentheses. Standard error on difference in means with RE74 subset/treated is given in brackets. Age = age in years; Education = number of years of schooling; Black = 1 if black, 0 otherwise; Hispanic = 1 if Hispanic, 0 otherwise; No degree = 1 if no high school degree, 0 otherwise; Married = 1 if married, 0 otherwise; RE*x* = earnings in calendar year 19*x*.

[a] NSW sample as constructed by Lalonde (1986).

[b] The subset of the Lalonde sample for which RE74 is available.

[c] Definition of comparison groups (Lalonde 1986):

La differenza più spiccata è quella relativa alla variabile RE75 ma ciò è dovuto oltre che a quello che è stato definito come *cohort phenomenon*, al fatto che la maggior parte dei soggetti era disoccupata prima della partecipazione alla politica.

Confrontando poi nei due campioni i valori assunti dalle variabili nel caso dei trattati e non trattati, si noterà che sostanzialmente non ci sono differenze significative.

Spostandoci ora al caso non sperimentale, per ottenere delle stime dell'impatto della politica è necessario disporre di un gruppo di confronto: prendiamo come riferimento quelli precedentemente

presentati per i maschi ossia PSID-1, PSID-2, PSID-3, CPS-SSA-1, CPS-SSA-2, CPS-SSA-3.

Anche in questo caso, il progressivo ricorso a diversi sottocampioni permette di dimostrare quanto i valori assunti dalle variabili pre-intervento "migliorino", avvicinandosi ai valori sperimentali, man mano che i campioni si restringono.

Consideriamo ora le due successive tabelle:

**Tabella 7** – Caratteristiche e stime dell'impatto per i gruppi sperimentale e CPS-SSA

| Control Sample | No. of Observations | Mean Propensity Score[A] | Age | School | Black | Hispanic | No Degree | Married | RE74 | RE75 | U74 | U75 | Treatment Effect (Diff. in Means) | Regression Treatment Effect |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NSW | 185 | 0.37 | 25.82 | 10.35 | 0.84 | 0.06 | 0.71 | 0.19 | 2095 | 1532 | 0.29 | 0.40 | 1794[B] | 1672[C] |
| | | | | | | | | | | | | | (633) | (638) |
| Full CPS | 15992 | 0.01 | 33.23 | 12.03 | 0.07 | 0.07 | 0.30 | 0.71 | 14017 | 13651 | 0.88 | 0.89 | −8498 | 1066 |
| | | (0.02)[D] | (0.53) | (0.15) | (0.03) | (0.02) | (0.03) | (0.03) | (367) | (248) | (0.03) | (0.04) | (583)[E] | (554) |
| Without replacement: | | | | | | | | | | | | | | |
| Random | 185 | 0.32 | 25.26 | 10.30 | 0.84 | 0.06 | 0.65 | 0.22 | 2305 | 1687 | 0.37 | 0.51 | 1559 | 1651 |
| | | (0.03) | (0.79) | (0.23) | (0.04) | (0.03) | (0.05) | (0.04) | (495) | (341) | (0.05) | (0.05) | (733) | (709) |
| Low to high | 185 | 0.32 | 25.23 | 10.28 | 0.84 | 0.06 | 0.66 | 0.22 | 2286 | 1687 | 0.37 | 0.51 | 1605 | 1681 |
| | | (0.03) | (0.79) | (0.23) | (0.04) | (0.03) | (0.05) | (0.04) | (495) | (341) | (0.05) | (0.05) | (730) | (704) |
| High to low | 185 | 0.32 | 25.26 | 10.30 | 0.84 | 0.06 | 0.65 | 0.22 | 2305 | 1687 | 0.37 | 0.51 | 1559 | 1651 |
| | | (0.03) | (0.79) | (0.23) | (0.04) | (0.03) | (0.05) | (0.04) | (495) | (341) | (0.05) | (0.05) | (733) | (709) |
| With replacement: | | | | | | | | | | | | | | |
| Nearest neighbor | 119 | 0.37 | 25.36 | 10.31 | 0.84 | 0.06 | 0.69 | 0.17 | 2407 | 1516 | 0.35 | 0.49 | 1360 | 1375 |
| | | (0.03) | (1.04) | (0.31) | (0.06) | (0.04) | (0.07) | (0.06) | (727) | (506) | (0.07) | (0.07) | (913) | (907) |
| Caliper, $\delta = 0.00001$ | 325 | 0.37 | 25.26 | 10.31 | 0.84 | 0.07 | 0.69 | 0.17 | 2424 | 1509 | 0.36 | 0.50 | 1119 | 1142 |
| | | (0.03) | (1.03) | (0.30) | (0.06) | (0.04) | (0.07) | (0.06) | (845) | (647) | (0.06) | (0.06) | (875) | (874) |
| Caliper, $\delta = 0.00005$ | 1043 | 0.37 | 25.29 | 10.28 | 0.84 | 0.07 | 0.69 | 0.17 | 2305 | 1523 | 0.35 | 0.49 | 1158 | 1139 |
| | | (0.02) | (1.03) | (0.32) | (0.05) | (0.04) | (0.06) | (0.06) | (877) | (675) | (0.06) | (0.60) | (852) | (851) |
| Caliper, $\delta = 0.0001$ | 1731 | 0.37 | 25.19 | 10.36 | 0.84 | 0.07 | 0.69 | 0.17 | 2213 | 1545 | 0.34 | 0.50 | 1122 | 1119 |
| | | (0.02) | (1.03) | (0.31) | (0.05) | (0.04) | (0.06) | (0.06) | (890) | (701) | (0.06) | (0.06) | (850) | (843) |

Variables: Age, age of participant; School, number of school years; Black, 1 if black, 0 otherwise; Hisp, 1 if Hispanic, 0 otherwise; No degree, 1 if participant had no school degrees, 0 otherwise; Married, 1 if married, 0 otherwise; RE74, real earnings (1982US$) in 1974; RE75, real earnings (1982US$) in 1975; U74, 1 if unemployed in 1974, 0 otherwise; U75, 1 if unemployed in 1975, 0 otherwise; and RE78, real earnings (1982US$) in 1978.

(A) The propensity score is estimated using a logit of treatment status on: Age, $Age^2$, $Age^3$, School, $School^2$, Married, No degree, Black, Hisp, RE74, RE75, U74, U75, School • RE74.

(B) The treatment effect for the NSW sample is estimated using the experimental control group.

(C) The regression treatment effect controls for all covariates linearly. For matching with replacement, weighted least squares is used, where treatment units are weighted at 1 and the weight for a control is the number of times it is matched to a treatment unit.

(D) The standard error applies to the difference in means between the matched and the NSW sample, except in the last two columns, where the standard error applies to the treatment effect.

(E) Standard errors for the treatment effect and regression treatment effect are computed using a bootstrap with 500 replications.

**Tabella 8** – Caratteristiche e stime dell'impatto per i gruppi sperimentale e PSID

| Control Sample | No. of Observations | Mean Propensity Score[A] | Age | School | Black | Hispanic | No Degree | Married | RE74 US$ | RE75 US$ | U74 | U75 | Treatment Effect (Diff. in Means) | Regression Treatment Effect |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NSW | 185 | 0.37 | 25.82 | 10.35 | 0.84 | 0.06 | 0.71 | 0.19 | 2095 | 1532 | 0.29 | 0.40 | 1794[B] | 1672[C] |
| | | | | | | | | | | | | | (633) | (638) |
| Full PSID | 2490 | 0.02 | 34.85 | 12.12 | 0.25 | 0.03 | 0.31 | 0.87 | 19429 | 19063 | 0.10 | 0.09 | −15205 | 4 |
| | | (0.02)[D] | (0.57) | (0.16) | (0.03) | (0.02) | (0.03) | (0.03) | (449) | (361) | (0.04) | (0.03) | (657)[E] | (1014) |
| Without replacement: | | | | | | | | | | | | | | |
| Random | 185 | 0.25 | 29.17 | 10.30 | 0.68 | 0.07 | 0.60 | 0.52 | 4659 | 3263 | 0.40 | 0.40 | −916 | 77 |
| | | (0.03) | (0.90) | (0.25) | (0.04) | (0.03) | (0.05) | (0.05) | (554) | (361) | (0.05) | (0.05) | (1035) | (983) |
| Low to high | 185 | 0.25 | 29.17 | 10.30 | 0.68 | 0.07 | 0.60 | 0.52 | 4659 | 3263 | 0.40 | 0.40 | −916 | 77 |
| | | (0.03) | (0.90) | (0.25) | (0.04) | (0.03) | (0.05) | (0.05) | (554) | (361) | (0.05) | (0.05) | (1135) | (983) |
| High to low | 185 | 0.25 | 29.17 | 10.30 | 0.68 | 0.07 | 0.60 | 0.52 | 4659 | 3263 | 0.40 | 0.40 | −916 | 77 |
| | | (0.03) | (0.90) | (0.25) | (0.04) | (0.03) | (0.05) | (0.05) | (554) | (361) | (0.05) | (0.05) | (1135) | (983) |
| With replacement: | | | | | | | | | | | | | | |
| Nearest Neighbor | 56 | 0.70 | 24.81 | 10.72 | 0.78 | 0.09 | 0.53 | 0.14 | 2206 | 1801 | 0.54 | 0.69 | 1890 | 2315 |
| | | (0.07) | (1.78) | (0.54) | (0.11) | (0.05) | (0.12) | (0.11) | (1248) | (963) | (0.11) | (0.11) | (1202) | (1131) |
| Caliper, $\delta = 0.00001$ | 85 | 0.70 | 24.85 | 10.72 | 0.78 | 0.09 | 0.53 | 0.13 | 2216 | 1819 | 0.54 | 0.69 | 1893 | 2327 |
| | | (0.08) | (1.80) | (0.56) | (0.12) | (0.05) | (0.12) | (0.12) | (1859) | (1896) | (0.10) | (0.11) | (1198) | (1129) |
| Caliper, $\delta = 0.00005$ | 193 | 0.70 | 24.83 | 10.72 | 0.78 | 0.09 | 0.53 | 0.14 | 2247 | 1778 | 0.54 | 0.69 | 1928 | 2349 |
| | | (0.06) | (2.17) | (0.60) | (0.11) | (0.04) | (0.11) | (0.10) | (1983) | (1869) | (0.09) | (0.09) | (1196) | (1121) |
| Caliper, $\delta = 0.0001$ | 337 | 0.70 | 24.92 | 10.73 | 0.78 | 0.09 | 0.53 | 0.14 | 2228 | 1763 | 0.54 | 0.70 | 1973 | 2411 |
| | | (0.05) | (2.30) | (0.67) | (0.11) | (0.04) | (0.11) | (0.09) | (1965) | (1777) | (0.07) | (0.08) | (1191) | (1122) |
| Caliper, $\delta = 0.001$ | 2021 | 0.70 | 24.98 | 10.74 | 0.79 | 0.09 | 0.53 | 0.13 | 2398 | 1882 | 0.53 | 0.69 | 1824 | 2333 |
| | | (0.03) | (2.37) | (0.70) | (0.09) | (0.04) | (0.10) | (0.07) | (2950) | (2943) | (0.06) | (0.06) | (1187) | (1101) |

(A) The propensity score is estimated using a logit of treatment status on: Age, Age$^2$, School, School$^2$, Married, No degree, Black, Hisp, RE74, RE74$^2$, RE75, RE75$^2$, U74, U75, U74 • Hisp.
(B) The treatment effect for the NSW sample is estimated using the experimental control group.
(C) The regression treatment effect controls for all covariates linearly. For matching with replacement, weighted least squares is used, where treatment units are weighted at 1 and the weight for a control is the number of times it is matched to a treatment unit.
(D) The standard error applies to the difference in means between the matched and the NSW sample, except in the last two columns, where the standard error applies to the treatment effect.
(E) Standard errors for the treatment effect and regression treatment effect are computed using a bootstrap with 500 replications.

Osservando le prime due righe di ciascuna tabella si possono notare le pesanti differenze, in termini di età, composizione etnica, formazione scolastica e redditi, fra i gruppi di controllo sperimentale e i gruppi di confronto CPS-SSA e PSID (dove con *full* CPS e *full* PSID si intendono i campioni definiti prima con CPS-1 e PSID-1).

Le stime cruciali sono comunque riportate nelle ultime due colonne: in particolare nella penultima l'impatto è ottenuto come

una semplice differenza fra medie mentre nell'ultima si condizionano le precedenti medie ad un insieme di variabili osservabili.

A conferma di quanto detto sulle vistose differenze in termini di caratteristiche osservabili tra i valori sperimentali e i due gruppi di confronto prendiamo come valore di riferimento \$1.794, ottenuto con la randomizzazione: i valori assunti dagli altri due gruppi di controllo non sperimentali sono completamente divergenti (\$-8.498 per CPS-1 e \$-15.205 per PSID-1).

Apriamo a questo punto una breve digressione per riprendere le fasi salienti della stima dei *propensity score*: si è proceduto innanzitutto ad una stima separata per i diversi gruppi considerati, quello sperimentale, PSID e CPS.

In particolare si è fatto ricorso ad un modello probabilistico di tipo logit per cui: $e(X)=\frac{e^{\lambda h(X_i)}}{1+e^{\lambda h(X_i)}}$

dove $h(X_i)$ è una funzione delle variabili pre-intervento con termini lineari e di ordine superiore (i termini inclusi sono *età, età*$^2$, *scuola, scuola*$^2$, *sposati, no diplomati, neri, ispanici, RE74, RE74*$^2$, *Re75, RE75*$^2$, *U74, U75 e U74*ispanici*); poi si procede con una stratificazione delle unità in base al valore assunto dal corrispondente *propensity score* in modo da porre in uno stesso strato unità trattate e non con un *propensity score* "vicino" (statisticamente la differenza non deve essere significativa). Se in ogni strato che è stato definito non compaiono differenze

significative fra trattati e non trattati, allora la specificazione usata risulta essere valida e l'intera procedura non deve essere ripetuta.
Tornando ora ai valori riportati nelle due precedenti tavole, le cose migliorano se prendiamo in considerazione appunto i *propensity score*; vengono utilizzati metodi di *matching* sia con reinserimento che senza reinserimento che riassumiamo brevemente:

- *Matching* senza reinserimento: - *Random*
  - *Low to high*
  - *High to low*
- *Matching* con reinserimento: - *Nearest neighbour*
  - *Caliper*

dove nel caso senza reinserimento conta principalmente l'ordine con cui si fanno gli abbinamenti (siano essi casuali, dal più piccolo *propensity score* al più alto o dal più alto *propensity score* al più piccolo); nel caso con reinserimento, l'ordine ha minor valenza, sottolineiamo comunque che le diverse specificazioni di $\delta$ indicano il raggio entro cui devono posizionarsi gli abbinamenti
Le stime dell'impatto nel caso CPS variano tra \$1.559 e \$1.605 se si parla di *matching* senza reinserimento e tra \$1.119 e \$1.360 se siamo nel caso di *matching* con reinserimento; le stime che meglio si accostano a quelle ottenute tramite l'approccio sperimentale sembrano dunque essere quelle ottenute tramite il *matching* senza reinserimento.
Le cose appaiono un po' diverse nel caso del gruppo PSID: le differenze più evidenti (e statisticamente significative) si osservano in termini di redditi pre-trattamento nel caso di

*matching* senza reinserimento. Conseguenza di ciò è il fatto che le stime dell'impatto variano fra $-916 e $77, valori molti distanti da quello che è il *benchmark* sperimentale. Un miglioramento appare nel caso di *matching* con reinserimento: le stime variano tra $1.824 e $1.973. (Dehejia Wahba, 2002).

I due istogrammi che seguono riportano la distribuzione dei *propensity score* stimati per i due gruppi di controllo non sperimentali. E noto che la bontà d'adattamento è direttamente collegata a quello che potremmo definire il cosiddetto supporto comune: infatti più le caratteristiche dei trattati saranno simili a quelle dei non trattati, più ampio sarà il supporto comune e migliore sarà la bontà d'adattamento.

**Figura 5** – Istogramma dei *propensity score* stimati per i gruppi sperimentale e CPS-SSA

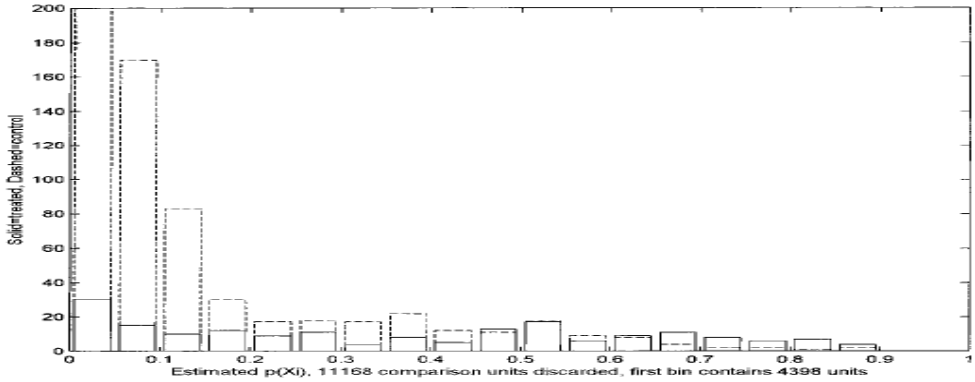
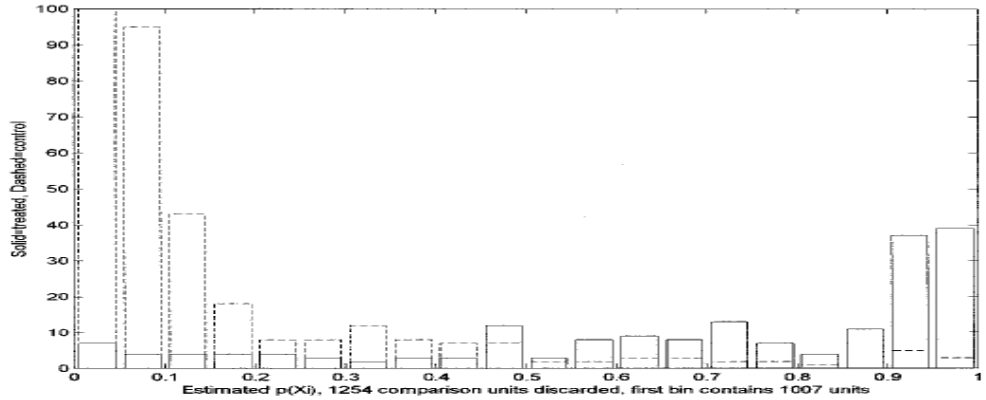



**Figura 6** – Istogramma dei *propensity score* stimati per i gruppi sperimentale e PSID



Da sottolineare che gli istogrammi non contengono quelle unità per cui il *propensity score* stimato nei gruppi di controllo è inferiore a quello stimato nel gruppo di trattamento: dunque sono state scartate 11.168 unità da CPS-1 e 1.254 da PSID-1.

A parte le unità scartate la maggior parte delle unità riferite ai gruppi di controllo sono contenute nelle prime sezioni dei digrammi (4.398 per CPS e 1.007 per PSID); le successive sezioni presentano un numero pressoché uguale di unità e addirittura nel caso PSID le ultime sezioni denotano una maggioranza di unità trattate rispetto a quelle non trattate. Quest'ultima osservazione spiega perché le stime ricavate dal *matching* senza reinserimento riportate nella tabella precedente risultavano problematiche: il *matching* con reinserimento funzione meglio quando ci sono poche unità del NSW che si accostano a quelle del PSID.

E evidente comunque che poche unità facenti parte dei gruppi di controllo sono comparabili a quelle del gruppo di trattamento. Osserviamo ora i seguenti grafici:

**Figura 7** – *Propensity score* per le unità trattate e non trattate (*matching* senza reinserimento, caso *random*)

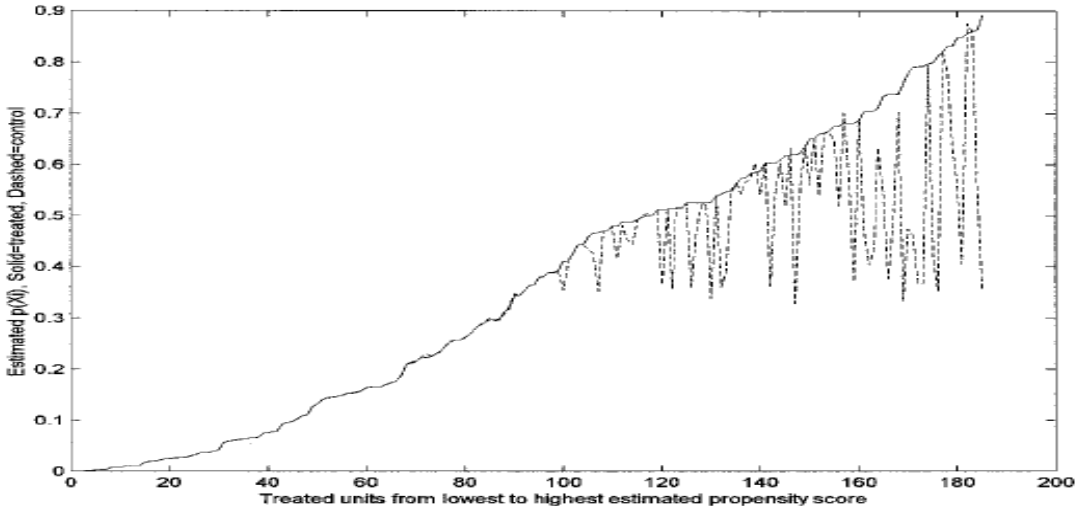


**Figura 8** – *Propensity score* per le unità trattate e non trattate (*matching* senza reinserimento, caso *lowest to highest*)

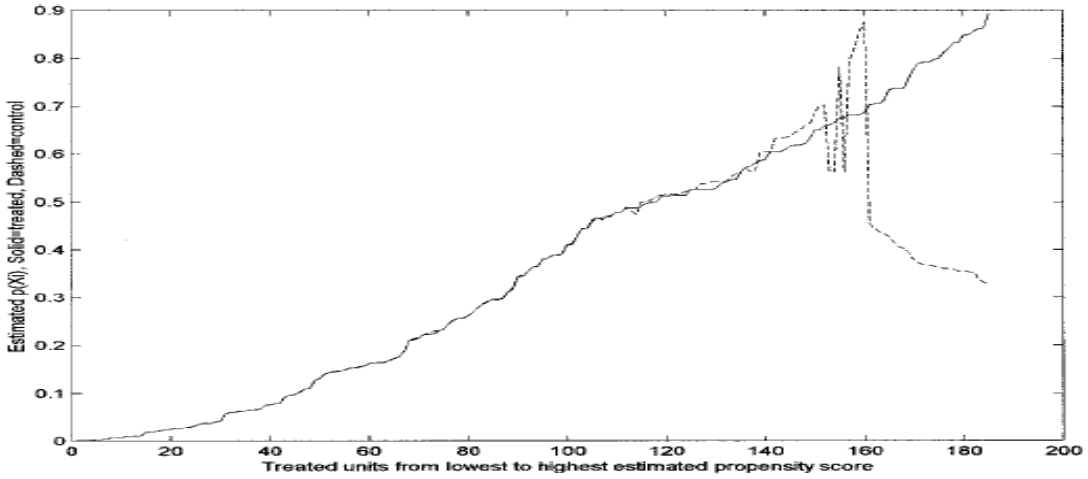



**Figura 9** – *Propensity score* per le unità trattate e non trattate (*matching* senza reinserimento, caso *highest to lowest*)

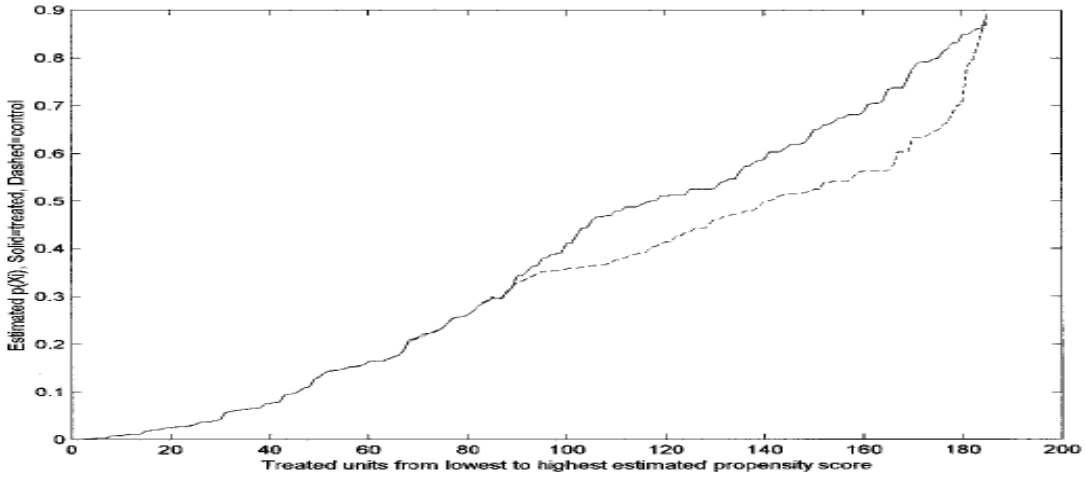
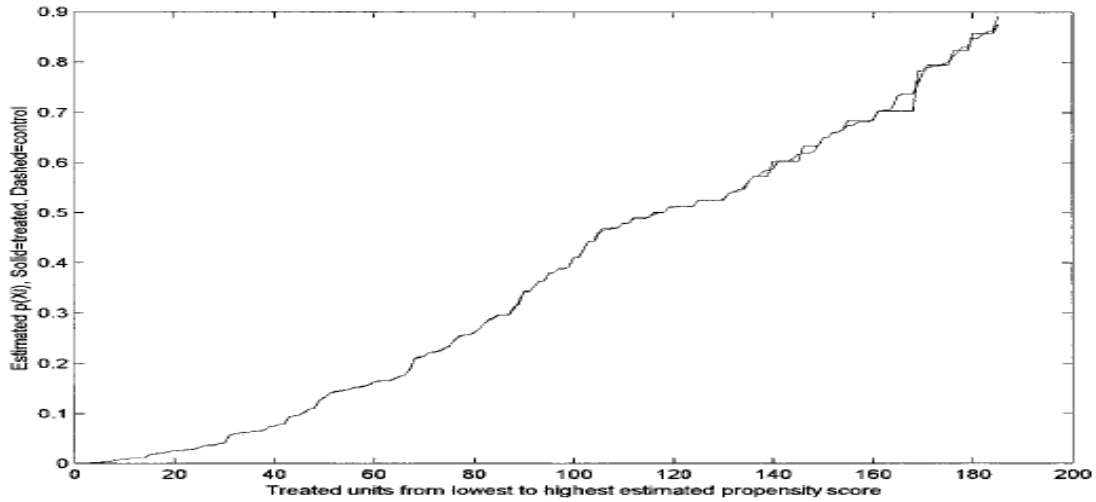



**Figura 10** – *Propensity score* per le unità trattate e non trattate (*matching* con reinserimento, caso *nearest neighbour*)



Nell'asse orizzontale vengono riportate le unità trattate mentre in quello verticale i valori dei *propensity score* stimati per i gruppi di trattamento e controllo (precisiamo che si riferiscono solo al caso CPS dato che nel caso PSID le conclusioni sono pressoché le stesse) e si vuole valutare la bontà degli abbinamenti tra i due gruppi.
In presenza di *matching* senza reinserimento (caso *random*) c'è un buon accostamento per le prime 100 unità visto che le due linee si sovrappongono bene, poi c'è un peggioramento reso evidente dalle pesanti oscillazioni casuali presenti nel grafico: tipico di questo approccio è che le prime unità scelte casualmente vengono ben abbinate fra loro poi man mano che si prosegue gli abbinamenti hanno una perdita in termini di bontà.
Sempre nel caso di *matching* senza reinserimento, questa volta però si opta per la variante *lowest to highest*, per le prime unità l'accostamento appare veramente positivo poi verso la 140° unità c'è un peggioramento dovuto al fatto che le unità fra i non trattati con un alto *propensity score* sono terminate e dunque si abbinano con un molto più basso valore del *propensity score*. Stesso discorso vale per il *matching* nella variante *highest to lowest*: la qualità di abbinamento comincia a decrescere dal momento in cui le unità trattate raggiungono un valore del *propensity score* attorno a 0,4.
Appare evidente a questo punto che la soluzione migliore sembra essere quella di procedere con un *matching* con reinserimento: entrambe le linee nell'ultima figura mantengono lo stesso andamento per ogni unità considerata.

Riprendiamo ora quello che è l'assunto fondamentale che sta alla base del concetto stesso di *propensity score* ossia il fatto che, condizionatamente al valore del *propensity score*, i redditi sono indipendenti dall'assegnazione o meno al gruppo di trattamento. In pratica confronteremo ora i redditi del gruppo di trattamento sperimentale con i due distinti gruppi di controllo impiegando i *propensity score*. Vediamo quindi i seguenti grafici:

**Figura 11** – Stima della distorsione nel caso CPS-SSA

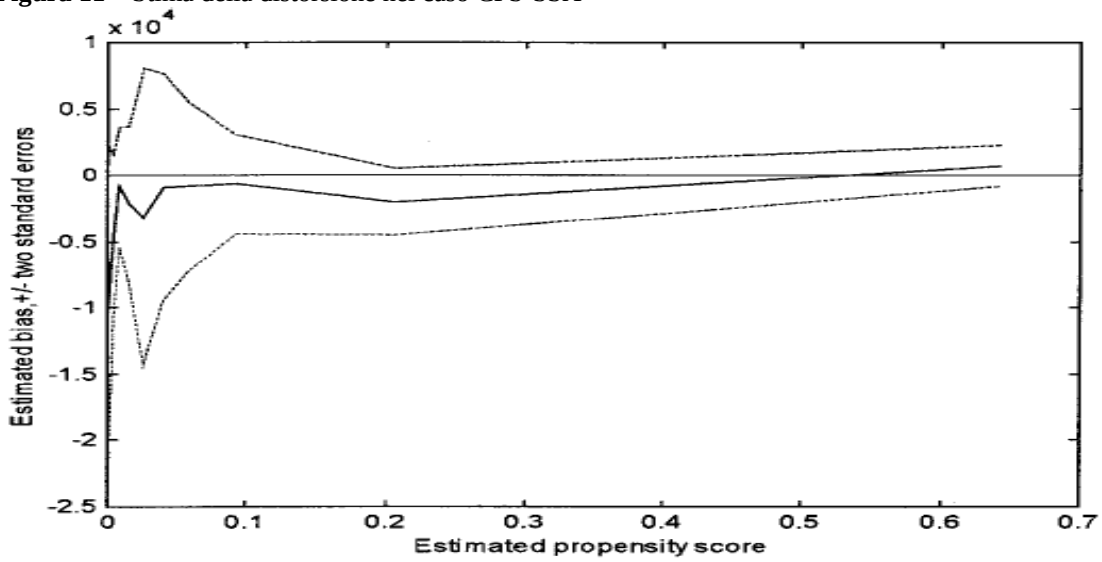
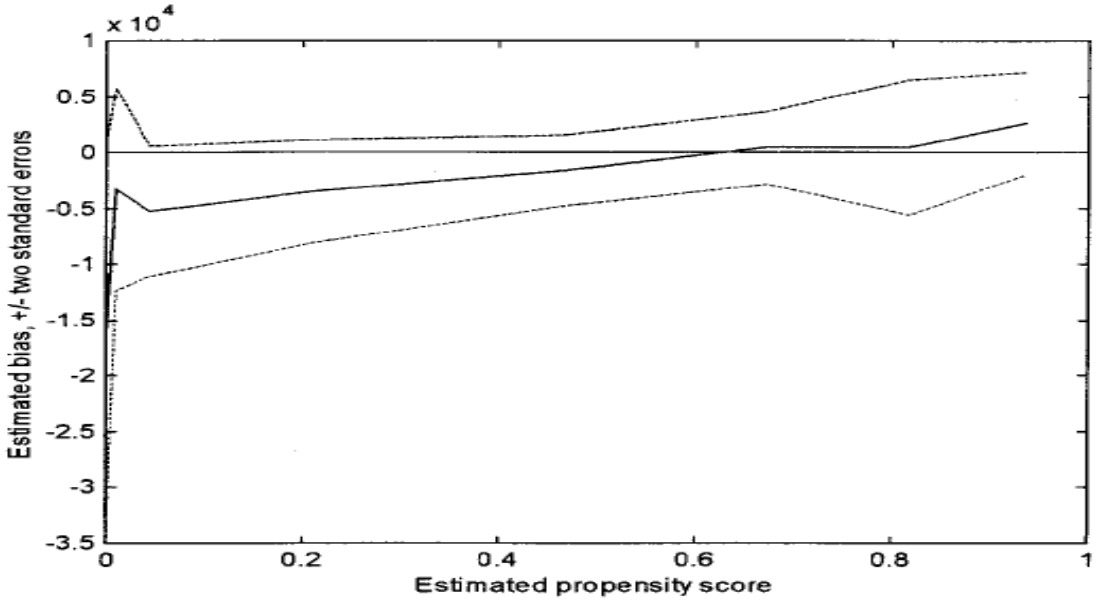



**Figura 12** – Stima della distorsione nel caso PSID



La parte più problematica riguarda i valori più bassi dei *propensity score*: ciò è dovuto presumibilmente al fatto che chi fa parte dei gruppi CPS-SSA e PSID ha un reddito superiore rispetto a chi rientra nel gruppo sperimentale, tuttavia nessuna di queste differenza risulta essere statisticamente significativa.

Concludiamo riportando una nota relativa alla stima dei *propensity score*: abbiamo detto che, nel modello logit, $h(X_i)$ è una funzione delle variabili pre-intervento con termini lineari e di ordine superiore ma quanto conta la scelta della specificazione delle variabili ai fini delle stime? La tabella che segue confronta la reale specificazione usata per stimare l'effetto della politica con altre eventuali varianti ottenute eliminando a mano a mano diversi termini.

Le differenze non sono particolarmente marcate specialmente se rapportate alle stime sperimentali: tuttavia non sempre si hanno a disposizione come *benchmark* le stime sperimentali, dunque è preferibile optare per una cosiddetta *full-specification*.

**Tabella 9** – Stime dell'impatto ottenute con diverse specificazioni

| Specification | Number of Observations | Difference-in-Means Treatment Effect (Standard Error)[B] | Regression Treatment Effect[A] (Standard Error)[B] |
|---|---|---|---|
| CPS | | | |
| Full specification | 119 | 1360 (633) | 1375 (638) |
| Dropping interactions and cubes | 124 | 1037 (1005) | 1109 (966) |
| Dropping indicators: | 142 | 1874 (911) | 1529 (928) |
| Dropping squares | 134 | 1637 (944) | 1705 (965) |
| PSID | | | |
| Full specification | 56 | 1890 (1202) | 2315 (1131) |
| Dropping interactions and cubes | 61 | 1004 (2412) | 1729 (3621) |
| Dropping indicators: | 65 | 1845 (1720) | 1592 (1624) |
| Dropping squares | 69 | 1428 (1126) | 1400 (1157) |

For all specifications other than the full specifications, some covariates are not balanced across the treatment and comparison groups.
(A) The regression treatment effect controls for all covariates linearly. Weighted least squares is used where treatment units are weighted at 1 and the weight for a control is the number of times it is matched to a treatment unit.
(B) Standard errors for the treatment effect and regression treatment effect are computed using a bootstrap with 500 replications.

## 3.6 Considerazioni conclusive

Partendo da quelli che sono stati i risultati sperimentali possiamo sostanzialmente affermare che c'è stato un impatto positivo in termini di reddito al termine della *National Supported Work Demonstration* specialmente per il gruppo delle giovani madri che percepivano il sussidio di povertà. Questo valutazione positiva ha spinto molti alla fine degli anni settanta a raccomandare l'applicazione di questo tipo di intervento su larga scala, come tentativo di soluzione al cronico problema dell'emarginazione dal mercato del lavoro da parte di soggetti fortemente disagiati.

La nostra attenzione si è concentrata soprattutto sulle questioni legate al reddito ma è chiaro che generalmente vi sono diversi parametri di interesse in una valutazione di una politica.

Può succedere perciò che stimatori adatti a stimare alcuni parametri siano poco adatti a stimarne altri: tuttavia per valutare appieno l'impatto di una politica devono essere valutati tutti quegli aspetti per cui può essere riscontrato un effetto, sia esso positivo o negativo.

Il vantaggio di poter disporre di stime sperimentali dell'impatto fa sì che queste possano essere utilizzate come *benchmark* per valutare le performance di stimatori non sperimentali: infatti il miglior modo di valutare la bontà di stime non sperimentali è proprio quello di confrontarle con quelle sperimentali quando questo sono disponibili.

In entrambi gli approcci non sperimentali affrontati la prima evidenza significativa riguarda la scelta e la composizione del

gruppo di controllo: è chiaro che questo deve avere caratteristiche osservabili pre-intervento molto simili al gruppo di trattamento sperimentale e può essere individuato attraverso degli archivi amministrativi, censimenti o indagini campionarie. Le alternative selezionate nel caso della *NSWD* hanno evidenziato il fatto che man mano che si diventa più selettivi i valori riferiti alle caratteristiche osservabili tendono a convergere sempre di più fornendo così successivamente delle stime dell'impatto post-intervento sempre più in linea con il *benchmark* sperimentale a danno però della numerosità campionaria che viene via via notevolmente ridotta.

La questione appena citata riguarda dunque entrambi i metodi adottati ma bisogna ricordare che altri aspetti, specifici dei due metodi, vanno tenuti in considerazione.

Nel caso in cui si è proceduto alla determinazione del reddito attraverso un modello econometrico è parso evidente quanto sia rilevante la specificazione sia dell'equazione che determina il reddito sia di quella che determina la partecipazione alla politica: infatti la stima dell'impatto appare notevolmente sensibile alle possibili restrizioni imposte nelle equazioni.

Per quel che riguarda invece la tecnica del *matching* attraverso i *propensity score*, un'altra questione determinante da tenere presente è sicuramente la scelta tra *matching* con reinserimento e *matching* senza reinserimento: solitamente quello con reinserimento appare migliore per il semplice fatto che l'abbinamento tra trattati e non trattati con valori dei rispettivi *propensity score* "vicini" risulta

essere più facile in quanto non si deve tener conto dell'ordine con cui si fanno gli abbinamenti.

Possiamo concludere affermando che la via migliore di procedere, per ottenere la stima dell'impatto di una politica, rimane senza dubbio quella sperimentale, tuttavia se ciò non fosse facilmente realizzabile si possono adottare delle tecniche statistico-econometriche per giungere alla stima dell'impatto. In particolare la scelta della stimatore appropriato dipenderà dal problema che la politica intende affrontare, dai dati disponibili e dalla domande alle quali, chi conduce una valutazione dell'impatto, vuole rispondere.

# Riferimenti bibliografici


- Dehejia R.H., Wahba S. (1999), Causal effects in nonexperimental studies: reevaluating the evaluation of training programs, *Journal of the American Statistical Society*, 94, 1053-1062;

- Dehejia R.H., Wahba S. (2002), Propensity score matching methods for nonexperimental casual studies, *The Review of Economics and Statistics*, 84(1), 151-161;

- LaLonde R.J. (1986), Evaluating the econometric evaluations of training programs with experimental data, *The American Economic Review*, 76(4), 604-620;

- Martini A. (1997), Valutazione dell'efficacia di interventi pubblici contro la povertà: questioni di metodo e studi di caso. *Collana della Commissione di Indagine sulla povertà e l'Emarginazione Sociale*, Roma, Presidenza del Consiglio dei Ministri.