



Istituto per la Ricerca Valutativa sulle Politiche Pubbliche

La valutazione degli effetti di politiche pubbliche: paradigma e pratiche

Ugo Trivellato

IRVAPP WP 2009-01
June 2009

La valutazione degli effetti di politiche pubbliche: paradigma e pratiche

Ugo Trivellato

IRVAPP, Università di Padova & CSS

Working Paper No. 2009-01

Giugno 2009



Istituto per la ricerca valutativa sulle politiche pubbliche
Fondazione Bruno Kessler
Via S. Croce 77
38122 Trento
Italy

Tel.: +39 0461 210242
Fax: +39 0461 210240
Email: info@irvapp.it
Website: <http://irvapp.fbk.eu>

Le opinioni espresse sono degli autori e non riflettono necessariamente quelle dell'IRVAPP.

I *Working Papers* dell'IRVAPP presentano spesso i risultati di un lavoro preliminare e sono diffusi per incoraggiare la discussione. Nella citazione di un tale paper si prega di dar conto del suo carattere preliminare. L'eventuale versione rivista è disponibile nella *Reprint Series*, se pubblicata, o direttamente presso l'IRVAPP.

Corresponding author: Ugo Trivellato, IRVAPP - Istituto per la Ricerca Valutativa sulle Politiche Pubbliche, Via S. Croce 77 - 38122 Trento, Italy. E-mail: trivellato@irvapp.it

La valutazione degli effetti di politiche pubbliche: paradigma e pratiche*

Giugno 2009

Sommario

Molte questioni empiriche in economia e in altri ambiti sociali dipendono dagli effetti causali di politiche pubbliche. Negli ultimi decenni la ricerca sui metodi per valutare gli effetti di tali politiche ha conosciuto notevoli avanzamenti. Essa ha oggi raggiunto un livello di maturità che ne fa uno strumento fondamentale in molteplici aree della ricerca empirica nelle scienze sociali (istruzione, ricerca e sviluppo, economia del lavoro, organizzazione industriale, sviluppo urbano, *welfare* e povertà, epidemiologia, ecc.), sotto il profilo analitico così come per le pratiche di valutazione degli effetti di politiche. Questa piana rassegna presenta e discute alcuni dei recenti sviluppi, corredandoli di sintetici riferimenti storici. Il fuoco è sulla *logica* che presiede alla valutazione degli effetti di politiche, e sull'importanza di *pratiche* di disegno e realizzazione delle politiche che incorporino l'obiettivo di una credibile valutazione dei loro effetti. In questa seconda prospettiva, sono brevemente considerate esperienze di alcuni Paesi ed è discusso lo stato dell'arte in Italia.

Parole-chiave: valutazione degli effetti di politiche, causalità, analisi controfattuale, politiche del lavoro, incentivi alle imprese, istruzione, *welfare*

Classificazione JEL: C14, C21, I21, I38, J68, L52

* Versione riveduta e ampliata della *Lecture* annuale del Consiglio italiano per le Scienze Sociali (CSS), che ho tenuto all'Accademia Nazionale dei Lincei il 26 novembre 2008. Essa deve molto alla consuetudine di ricerca con Alberto Martini ed Enrico Rettore, e più recentemente con Erich Battistin e Adriano Paggiaro. Sono loro grato anche per aver consentito di trarre passi da scritti comuni. Per utili commenti su una versione preliminare ringrazio, oltre a loro, Antonio Schizzerotto e Alberto Zuliani, e per l'acuta discussione Pia Marconi e Sergio Ristuccia. Ho inoltre beneficiato di commenti e suggerimenti in occasione di un Seminario sul tema tenuto al Master "Parlamento e politiche pubbliche", LUISS Guido Carli, Roma, 13 marzo 2009, e della relazione invitata al Convegno dell'Associazione Nazionale di Epidemiologia "Valutazione delle prove e interventi di promozione della salute", Roma, 14-15 maggio 2009. Ai colleghi della Commissione di studio del CSS "La valutazione degli effetti di politiche pubbliche: metodi, pratiche, prospettive", attiva dalla fine del 2006, e a quanti hanno offerto il loro contributo in occasione delle audizioni e degli *Workshop* che la stessa ha organizzato nell'arco di due anni sono debitore di suggerimenti e stimoli critici. Ciò detto, diventa d'obbligo la precisazione che le opinioni espresse ed eventuali errori sono soltanto miei.

1. *Introduzione, con alcune precisazioni*

Accanto alle politiche macroeconomiche, negli ultimi decenni¹ hanno assunto un'importanza crescente politiche mirate a gruppi variamente ristretti di soggetti – individui, famiglie, imprese, che si propongono di modificare in un senso desiderato la condizione e/o il comportamento dei destinatari mediante l'applicazione di un trattamento: l'offerta di un servizio, l'erogazione di un incentivo, l'imposizione di un obbligo, o una qualche loro combinazione.

Ad esempio, è questa la tipologia nella quale si collocano gran parte delle politiche del lavoro: attività di orientamento dei Centri per l'impiego; interventi di formazione per l'avviamento al lavoro; sostegno del reddito dei disoccupati, accompagnato da azioni – e correlati obblighi per i beneficiari – che ne favoriscano il ritorno al lavoro; lavori sussidiati di pubblica utilità; incentivi per l'avvio di attività di lavoro autonomo o imprenditoriali; regolazione di nuovi rapporti di lavoro o modificazione di quelli esistenti; sussidi alle imprese per l'assunzione di lavoratori o per la trasformazione di contratti di lavoro a termine in contratti tempo indeterminato; ecc..

Politiche con caratteristiche analoghe si sono venute affermando in molti altri campi: istruzione, ricerca e sviluppo, sostegno alle imprese, lotta alla povertà, sanità – soprattutto in ambito epidemiologico, recupero urbano, e via dicendo.

La nozione di “politica pubblica” alla quale mi riferisco è, dunque, piuttosto ampia: un *intervento mirato a una popolazione determinata con l'intento di indurre un cambiamento in una condizione e/o in un comportamento*. In questa accezione di politica pubblica risultano definite tre dimensioni, essenziali sia per connotarla sia ai fini della sua valutazione: la popolazione-obiettivo, il trattamento, il risultato perseguito. Tale accezione riflette la nozione anglosassone di *programme*. In questo senso utilizzerò termini quali politica pubblica – o *tout court* politica, programma, intervento come sinonimi.

La domanda che è legittimo, anzi opportuno, porsi rispetto a ogni politica pubblica è: “L'intervento è stato efficace? Ha inciso positivamente sul problema che l'ha motivato?”². Rispondere in modo convincente alla domanda è tutt'altro che facile. Una risposta chiarificatrice è possibile soltanto se si dispone di un'evidenza empirica adeguata, ottenuta con strumenti analitici rigorosi: tali da consentire di separare i cambiamenti imputabili alla politica da quelli che si sarebbero verificati comunque; detto altrimenti, di *valutare gli effetti della politica*³.

¹ In tema di politiche di *welfare*, risale agli anni 1940 il “Beveridge Report” (Beveridge 1942). L'iniziale esperienza alla quale mi pare ragionevole riferirsi è, tuttavia, la “*war on poverty*” lanciata negli Stati Uniti dal Presidente Johnson a metà degli anni 1960, anche per l'impulso fondamentale che dette allo sviluppo, se non alla nascita, della *programme evaluation*.

² Nel mondo anglosassone questo tipo di incertezza è racchiuso nella classica domanda “*What works?*”.

³ Userò sistematicamente questa locuzione, e non gli usuali termini anglosassoni *impact evaluation* o *programme evaluation* – o le traduzioni letterali – per la polisemia, e quindi gli equivoci, con cui sono frequentemente usati nella letteratura italiana, e a ben vedere non solo italiana, sulla valutazione (vedi Martini 2008, sulla nebulosità terminologica e logica dei manuali, e delle pratiche, della Commissione dell'Unione Europea sull'argomento).

L'ostacolo che si incontra sta nel fatto che la valutazione degli effetti di una politica deve affrontare uno dei problemi cognitivi più difficili: l'*attribuzione causale*. Si tratta di attribuire a una specifica azione pubblica il merito – o meglio, la parte di merito che essa ha determinato – nei miglioramenti osservati, o nei peggioramenti evitati, del fenomeno che intende modificare.

Prima di procedere tornano utili alcune precisazioni. Innanzitutto, a questo scritto è chiesto di essere accessibile a un pubblico di lettori largo, con formazione e interessi disciplinari diversi; e inoltre di non trascurare l'interrelazione fra analisi teorica e suggerimenti operativi. Pertanto, eviterò argomentazioni formali e correrò consapevolmente il rischio di drastiche semplificazioni, appena mitigate da qualche lunga nota. Il proposito è trasmettere il messaggio essenziale: *la logica* che presiede alla valutazione degli effetti di politiche; l'importanza di *pratiche* di disegno e realizzazione delle politiche che incorporino l'obiettivo di una credibile valutazione dei loro effetti.

In secondo luogo, la nozione di politica pubblica che ho adottato è sì larga, ma certo non esaustiva. Non rientrano – o rientrano solo marginalmente – nel suo ambito, e dunque non considererò nel seguito, altri tipi di politiche. Ciò vale innanzitutto per le politiche di grandi infrastrutture e opere pubbliche, che hanno la caratteristica di sostanziale irreversibilità e richiedono valutazioni *ex ante* basate su analisi costi-benefici. Per un altro verso, ciò vale in generale per quelle che potremmo chiamare “macropolitiche” – monetarie, fiscali e di bilancio, industriali, ambientali, di modifica di forme generali di regolazione, ecc.: tipicamente esse consistono in un insieme variegato di interventi che sovente interessano popolazioni-obiettivo differenti e si dispiegano in tempi diversi (l'eccezione è costituita da bruschi cambiamenti di *policy regimes*, quale uno shock monetario).

Esula poi da questa accezione di politica pubblica la prestazione di servizi svolta in via corrente, ordinaria, dalle amministrazioni pubbliche. A queste attività manca, infatti, la caratteristica distintiva di intervento mirato a indurre un cambiamento in una popolazione ben determinata. Per esse non si pone tanto una questione di valutazione degli effetti, quanto piuttosto di misura della qualità e dei costi. In quest'ambito la nozione di politica pubblica qui accolta diventa rilevante nel caso di riforme, cioè a dire di discontinuità nell'entità e/o nelle modalità di erogazione di tali prestazioni.

In terzo luogo, va tenuto presente che il vaglio di una politica pubblica sovente è motivato anche, e forse in maggior misura, da un dubbio diverso da quello sulla sua efficacia. Quest'altra preoccupazione attiene alla conformità della messa in atto della politica a regole e direttive da parte dei soggetti che hanno il compito di attuarla. Da qui nasce l'esigenza di *monitorare il processo di attuazione* di una politica e i suoi esiti, di *rendere conto* dei progressi compiuti verso gli obiettivi prefissati.

Certo, per valutare gli effetti di una politica pubblica si deve fare riferimento alla politica realizzata, non a quella disegnata dal *policy maker*. Il monitoraggio del processo di attuazione ne è, dunque, un pre-requisito essenziale. E tanto più le azioni di monitoraggio non si limitano agli input e ai processi, ma si estendono agli esiti, tanto più forniscono un

sostrato conoscitivo utile per le risposte alla domanda di nostro primario interesse: se la politica sia stata efficace.

Ciò nonostante, è importante sottolineare le diverse implicazioni dei due tipi di incertezza che motivano la valutazione delle politiche pubbliche: quella appena richiamata motiva uno spettro di attività valutative che hanno per oggetto le decisioni riguardanti l'implementazione; la prima riguarda l'efficacia dell'azione pubblica e motiva la valutazione dei suoi effetti⁴.

Infine, tornando alla nozione di politica pubblica che utilizzo, è evidente che lo schema concettuale che la connota, scandito dalla terna "popolazione-trattamento-risultato", non è tipicamente utilizzato in sede di disegno della politica. Si tratta di una razionalizzazione *ex-post*, utile al fine di analizzarla e valutarla. È appena ovvio notare, infatti, come il disegno delle politiche pubbliche e la loro implementazione siano un processo molto meno lineare e razionale del percorso appena richiamato. Resta il fatto che, volendo porsi l'obiettivo di rispondere alla domanda "L'intervento è stato efficace?", lo schema "popolazione-trattamento-risultato" rappresenta un modo appropriato di impostare il problema.

L'articolo procede come segue. Nella sez. 2 presento e discuto brevemente il paradigma controfattuale per l'inferenza causale, contrastandolo col programma di ricerca avviato in ambito economico dalla Cowles Commission intorno alla metà del secolo scorso. Nella sez. 3, delinea le principali strategie che, alla luce del disegno della politica pubblica e dell'informazione disponibile, possono essere adottate per identificarne gli effetti. Nella sez. 4 passo in rassegna alcune buone pratiche di disegno, realizzazione e valutazione degli effetti di politiche pubbliche in Paesi sviluppati, e ne traggio indicazioni

⁴ Secondo una logica analoga va vista la fondamentale diversità, e insieme la possibile complementarità, fra valutazione degli effetti di politiche pubbliche e analisi dei risultati estesa all'intero bilancio pubblico, svolta tramite il *performance budgeting*. Esso consiste nell'introduzione di meccanismi di fissazione di obiettivi e di verifica dei risultati all'interno del processo di formazione del bilancio pubblico, che viene strutturato per funzioni-obiettivo invece che per capitoli di spesa. A ogni funzione-obiettivo corrisponde inevitabilmente un *mix* di politiche pubbliche eterogenee, segnatamente per quanto riguarda i trattamenti, nonché un insieme di prestazioni svolte in via ordinaria dall'amministrazione pubblica. Il *performance budgeting* tenta di quantificare il contributo prodotto congiuntamente da questi *mix* di politiche pubbliche nonché da questi insiemi di prestazioni ordinarie, a fronte delle risorse stanziare nel bilancio.

Tre sono gli elementi che rendono fundamentalmente diversi il *performance budgeting* e la valutazione degli effetti delle politiche. Del problema connesso all'aggregazione di politiche eterogenee ho appena detto. Un secondo elemento di diversità, o meglio di incompatibilità, sta nella compresenza, in pressoché ogni funzione-obiettivo del bilancio pubblico, di un insieme di prestazioni svolte in via ordinaria, rispetto alle quali strutturalmente non si pone una questione di valutazione degli effetti. Un terzo elemento di diversità tra i due strumenti è l'orizzonte temporale: annuale o comunque fisso per il *performance budgeting*; variabile e dipendente dal particolare problema e contesto nel caso della valutazione degli effetti.

È interessante notare che alcuni operatori e studiosi nell'area del *performance budgeting* mostrano consapevolezza delle diversità dei due strumenti (vedi soprattutto Mayne 2001). Essi propongono una prospettiva, detta *contribution analysis*, secondo cui il successo di una politica pubblica può essere *approssimativamente* valutato mediante la combinazione di una serie di elementi, tra i quali (i) un quadro logico che metta a fuoco la *ratio* della politica, cosa la motiva e quali obiettivi si pone; (ii) un sistema di controllo delle realizzazioni, per rendere conto delle "cose fatte"; (iii) una verifica dei progressi compiuti lungo le dimensioni sulle quali la politica vuole influire. E tutto ciò senza pretendere di giungere a stringenti attribuzioni di causalità, che sono affidate propriamente all'analisi degli effetti di politiche pubbliche.

di metodo e di merito. Nella sez. 5 percorro per rapidi tratti l'esperienza italiana sul tema. Alla sez. 6 affido scarse considerazioni conclusive.

2. *Il paradigma scientifico*

2.1. *L'analisi controfattuale*

Come ho anticipato, il problema cognitivo consiste nell'attribuire alla politica, *in senso causale*, i cambiamenti osservati nella condizione e/o nel comportamento dei destinatari. Gli interrogativi che si pongono sono del seguente tenore: il miglioramento osservato è merito della politica o si sarebbe verificato comunque? il peggioramento osservato sarebbe stato più grave in assenza della politica oppure no? E non sono interrogativi banali.

Vi sono molte ragioni che concorrono a rendere difficile la risposta.

- (a) Contemporaneamente alla messa in atto di una politica, i processi spontanei, tra i quali quelli che la politica intende modificare, continuano a fluire: e possono fluire in maniera diversa per destinatari e non destinatari della politica.
- (b) In generale, poi, i soggetti – persone, famiglie, imprese – destinatari della politica, e in molte circostanze anche i non destinatari, reagiscono alla messa in opera della politica comportandosi in modo diverso da come si sarebbero comportati in sua assenza (al di là, naturalmente, della reazione di interesse: la risposta dei destinatari al trattamento).
- (c) Ancora, molte politiche sono poste in essere a scale “locali” (regionali, provinciali, comunali, delle singole agenzie o centri che le erogano), e si caratterizzano quindi per i diversi contesti e per le specifiche modalità di realizzazione. A contesti diversi ed a protocolli applicativi diversi corrispondono in generale esiti diversi.

È dunque evidente che una definizione ingenua di effetto di un intervento come “differenza tra la situazione osservata dopo l'esposizione all'intervento e la situazione osservata prima dell'esposizione” ha un'altissima probabilità di fornire una risposta fallace. A fronte di un miglioramento osservato, l'obiezione è che esso avrebbe potuto verificarsi comunque, per cause diverse dall'intervento, e quindi l'effetto potrebbe essere nullo, o addirittura negativo, pur in presenza di un miglioramento. Seguendo la stessa logica, nel caso opposto di un peggioramento osservato l'obiezione è che le cose sarebbero potute andare ancora peggio: l'intervento può quindi avere un effetto positivo pur in presenza di un peggioramento.

Per evitare queste fallacie concettuali, è cruciale adottare la definizione di effetto secondo la logica *controfattuale*, ormai affermata nelle scienze sociali⁵: l'effetto di un intervento è la *differenza* tra quanto si osserva *in presenza* dell'intervento e quanto si

⁵ A titolo esemplificativo della diffusa accettazione del paradigma controfattuale – tra statistici, economisti, sociologi, politologi, psicologi, pedagogisti, studiosi di intelligenza artificiale, ecc., – si vedano i testi di Heckman, LaLonde e Smith (1999), Pearl (2000), Shadish, Cook e Campbell (2002), Lee (2005), Winship e Morgan (2007), Imbens e Rubin (2009). Presentazioni più sintetiche, variamente formalizzate, sono in Wooldridge (2002: 603-644), Rettore, Trivellato e Martini (2003), Blundell e Costa Dias (2008), Imbens e Wooldridge (2008) e Martini, Rettore e Trivellato (2009).

sarebbe osservato in *sua assenza*. Va subito notato che mentre il primo termine di questo confronto è *osservabile*, il secondo termine è *ipotetico*, non osservabile per definizione. Se l'intervento è stato attuato, per gli esposti all'intervento ciò che sarebbe accaduto in assenza dell'esposizione non può più essere osservato⁶. Questo risultato ipotetico è definito situazione o *risultato controfattuale*.

Ne discende che l'effetto di un intervento *a livello individuale* resta non osservabile per definizione, perché i due risultati – quello osservabile e quello ipotetico – non possono essere osservati entrambi⁷. Da qui nasce quello che Heckman, LaLonde e Smith (1999: 1879) definiscono «*il problema fondamentale della valutazione [degli effetti]*»⁸. Incidentalmente, vale peraltro la pena di notare che l'osservabilità di uno soltanto dei due risultati ha implicazioni ben più ampie, di portata epistemologica generale. Seguendo Holland (1986: 947) possiamo ben dire, infatti, che essa costituisce «*il problema fondamentale dell'inferenza causale*».

Per uscire da questa *impasse*, si considerano effetti causali medi, nella popolazione o in specifici sottogruppi. Del resto, sono tipicamente questi effetti medi l'oggetto di interesse, sia dell'analista sia del *policy maker*.

Sfruttando l'esistenza di una pluralità di soggetti, alcuni esposti e altri non esposti all'intervento, possiamo cercare di identificare effetti causali medi. Il paradigma controfattuale⁹ è la logica che accomuna i metodi analitici utilizzati per l'attribuzione di

⁶ Analogamente, una volta che l'intervento è stato attuato, per i non esposti all'intervento si osserva il risultato *in assenza* dell'intervento, mentre quanto si sarebbe osservato per gli stessi *se fossero stati esposti* all'intervento è *ipotetico*, non osservabile per definizione.

⁷ Per semplicità espositiva (che spero raggiunta), compio qui una grossolana forzatura rispetto al “modello dei risultati potenziali” sistematizzato da Rubin (1974). Dato un trattamento di tipo binario, egli postula che esistano, per ogni individuo *i*-esimo, due valori della variabile-risultato *Y*, detti *risultati potenziali*: Y_i^T nel caso che l'individuo sia esposto al trattamento e Y_i^{NT} nel caso che l'individuo non sia esposto al trattamento. Ciò consente di definire in modo rigoroso e generale la nozione di effetto causale del trattamento *T* sull'*i*-esimo individuo *come differenza tra i due risultati potenziali*: $\delta_i = Y_i^T - Y_i^{NT}$. Si tratta di una nozione generale di effetto, in quanto prescinde dal fatto che sia uno dei due risultati potenziali ad essersi realizzato.

⁸ Qui e nel seguito le traduzioni sono mie.

⁹ Sul piano filologico, l'uso del termine “controfattuale” per definire la logica di analisi della causalità affermata dalla metà degli anni '70 del secolo scorso è impreciso. Esso non rende giustizia al fatto che il tema centrale dell'econometria classica a partire dagli anni '40, segnatamente dall'attività della Cowles Commission, è stata la messa a punto di strumenti per rispondere alla domanda “*What if?*”, cioè per simulare controfattuali (tra i molti, vedi soprattutto Marschak 1953). È pur vero che nel contesto dei modelli elaborati nell'ambito della Cowles Commission l'interesse era sulla valutazione *ex-ante* di una politica nuova, mai messa in atto precedentemente; ma sempre di ricorso alla nozione di controfattuale si trattava (vedi le penetranti rassegne di Heckman 2000, 2001).

Sono tuttavia propenso a mantenere il termine “controfattuale” per connotare l'analisi della causalità in sede di valutazione *ex-post* degli effetti di politiche per varie ragioni. Il termine si è ormai affermato nella comunità scientifica (vedi, tra i molti esempi, il titolo del volume di Morgan e Winship 2007). Inoltre, serve a dare visibilità ad un aspetto logico centrale per la valutazione degli effetti di un intervento: l'effetto di un intervento è definito dal confronto tra ciò che accade a valle della messa in atto dell'intervento, il fattuale, e ciò che sarebbe accaduto se l'intervento non fosse stato realizzato, il controfattuale. Infine, altre denominazioni mi paiono meno persuasive. La «*general rubric of the “natural experiment” movement*», utilizzata da Heckman (2000: 51) sposta troppo l'enfasi sui metodi a scapito degli aspetti logici. D'altra parte, la dizione «*the Rubin causal model*», con cui spesso ci si riferisce al paradigma controfattuale (vedi per primo Holland 1986, e, ad esempio, Imbens e Wooldridge 2008) attribuendo la paternità della sua

causalità a un intervento¹⁰. Esso richiede di ricostruire credibilmente ciò che si sarebbe osservato sugli esposti all'intervento in assenza di loro esposizione allo stesso.

2.2. *Interrogarsi sugli effetti di una politica pubblica: due esempi stilizzati*

Per illustrare la logica con cui si formula una domanda di valutazione degli effetti, e si cerca di rispondervi, introduco due esempi di politiche pubbliche. Gli esempi si ricollegano a politiche adottate in Italia (e in vari altri Paesi), ma, per lo scopo appena detto, di proposito sono presentati in maniera stilizzata, senza i dettagli del caso. La prima politica riguarda l'offerta di servizi di orientamento al lavoro per lavoratori disoccupati. La seconda politica consiste nell'offerta di un incentivo monetario alle imprese per stabilizzare i propri lavoratori precari.

Per ciascuno dei due esempi, la tabella 1 riassume, sotto l'intestazione "Effetto", le risposte alla terna di domande su "trattamento-risultato-popolazione" che connotano una politica. Questa parte della tabella è in buona misura auto-esplicativa. Mi limito a corredarla di qualche commento.

La tabella delinea poi sinteticamente, sotto l'intestazione "Identificazione degli effetti", le strategie di analisi e i dati richiesti per affrontare tale tema. È con riferimento a questa seconda parte che vedrò di chiarire i passi salienti del percorso che porta all'interpretazione in senso causale dell'evidenza disponibile sulla realizzazione di una politica – quella che è chiamata appunto "identificazione degli effetti", e il ruolo decisivo che in questo percorso hanno da un lato il disegno della politica (e la sua implementazione) e dall'altro i dati disponibili. In questo itinerario, avrò anche modo di segnalare le difficoltà, i rischi di "non validità" che può incontrare un esercizio di valutazione.

2.2.1. *I servizi di orientamento al lavoro*

I servizi di orientamento ai disoccupati nella ricerca di lavoro sono un caso emblematico di offerta di un servizio complesso, sotto due profili. Innanzitutto, il trattamento consiste, potenzialmente, in un *mix* di misure "attive": colloquio iniziale di orientamento, servizio intensivo di *counselling*, servizio di supporto al *matching* con domande di lavoro. Sovente esse sono poi affiancate da una misura "passiva", un sostegno al reddito (tipicamente limitato nel tempo e decrescente), così come da prescrizioni (ad esempio, l'obbligo per il disoccupato di accettare un'offerta di lavoro congrua). In definitiva, siamo di fronte a un trattamento integrato, che consiste nella combinazione di più misure. Per semplicità, e senza rilevante perdita di generalità, possiamo dare una

formulazione in termini di risultati potenziali allo statistico statunitense Donald Rubin (Rubin 1974), è contestata. Altri, tra i quali Heckman (2000, 2008) e suoi coautori, attribuiscono, in maniera persuasiva, l'elaborazione del paradigma controfattuale a un filone di pensiero che data dagli anni '20 del secolo scorso, e coinvolge statistici ed economisti quali Neyman, Fisher, Roy, Quandt, Rubin e Heckman.

¹⁰ Non è superfluo aggiungere che la logica controfattuale ha preso piede anche nella ricerca storica (vedi, tra gli altri, Ferguson 1999, e Bunzl 2004). Una spiegazione controfattuale della causalità esiste anche in filosofia; essa ambisce peraltro a un modello-nozione di causalità più generale (vedi Lewis 1973, e Woodward 2003).

specificazione operativa del trattamento ridotta all'essenziale, tramite una variabile dicotomica che rappresenta la presenza-assenza del trattamento stesso. È quello che farò nel seguito. Occorre peraltro essere consapevoli che stiamo considerando un trattamento integrato, ed essere attenti alle implicazioni che ciò ha per l'interpretazione dei risultati e nella prospettiva di possibili affinamenti dell'analisi¹¹.

In secondo luogo, ai servizi di orientamento al lavoro accede una parte dei disoccupati, su base volontaria. Inoltre, al disegno e alla realizzazione dell'intervento, in particolare della sua componente "attiva" di servizi, concorrono operatori che sono portatori di professionalità e caratterizzati da convinzioni e motivazioni circa il modo di intervenire. Questi servizi sono quindi un terreno difficile dal punto di vista della valutazione dei loro effetti, perché si trovano all'intersezione di sollecitazioni potenzialmente contraddittorie: l'individuazione dei soggetti che più hanno bisogno di aiuto nella ricerca di lavoro e l'assegnazione del soggetto al servizio, o al *mix* di servizi, più adatto tra quelli disponibili. Il pericolo che corrono questi servizi è duplice: che siano i disoccupati più capaci a richiedere maggiormente i servizi; che gli operatori assegnino, o erogino direttamente, i servizi – se molteplici, i servizi migliori – ai soggetti migliori, per accresce la possibilità di successo nell'inserimento lavorativo (con le positive conseguenze che ciò può avere per l'operatore).

Le considerazioni appena fatte portano a riflettere sul terzo elemento-chiave della domanda di valutazione: come viene definita la popolazione-obiettivo della politica, cioè l'insieme dei soggetti ammissibili al trattamento? E qual è, poi, il meccanismo di assegnazione, che in sede di realizzazione della politica determina quali sono, di fatto, gli esposti al trattamento?

Due sono i punti che preme mettere in luce. In primo luogo, i criteri di ammissibilità, che definiscono la popolazione-obiettivo, sono in generale tali da indurre variazioni nelle variabile-trattamento, tra soggetti oppure nel tempo¹². Sono proprio queste *discontinuità* della politica pubblica, nella determinazione dei soggetti ammissibili o nel tempo che generano l'indispensabile variabilità per affrontare la valutazione dei suoi effetti.

In secondo luogo, è decisivo distinguere tra i criteri di ammissibilità, che sono un ingrediente del disegno della politica, e il meccanismo di assegnazione, che nel corso della messa in atto della politica determina *fattualmente* quali sono i *trattati*, e per differenza i *non-trattati*. Per molte politiche pubbliche è questo lo snodo fondamentale che va affrontato, e per quanto possibile risolto, per darne una credibile valutazione degli effetti. La questione è emersa con chiarezza in sede di presentazione dell'esempio dei servizi di orientamento alla ricerca di lavoro.

¹¹ Considerare trattamenti su più livelli o trattamenti continui o trattamenti integrati dei quali si intende scomporre l'effetto del *mix* di misure comporta analisi empiriche più esigenti quanto a dati richiesti, più articolate, talvolta più complesse, ma non altera la logica del problema.

¹² Una politica universale, omogenea e costante, che sottopone cioè tutti allo stesso trattamento e che non subisce alcun tipo di variazione nel tempo, non è valutabile con una logica controfattuale.

Tabella 1: *Una schematizzazione del processo di formulazione delle domande di valutazione e dell'identificazione degli effetti, con riferimento a due esempi stilizzati*

Due esempi stilizzati di politica pubblica	Effetto			Identificazione degli effetti: strategie di analisi e dati
	Di cosa:	Su cosa:	Per chi:	
	variabile trattamento	Variabile risultato	popolazione obiettivo	
Servizi di orientamento nella ricerca di lavoro	Servizio ricevuto	Durata della disoccupazione (dopo il trattamento)	A. Iscritti a liste di disoccupazione	<p>A. Alto rischio di <i>autoselezione</i> dei disoccupati nell'accedere al servizio e di <i>selezione</i> da parte del personale dei Centri per l'impiego nell'assegnazione dei migliori ai servizi (migliori).</p> <p>Necessità di <i>ricche basi di dati individuali</i> sugli iscritti alle liste di disoccupazione (caratteristiche socio-demografiche, scolarità/formazione, storia lavorativa precedente, ecc.), nonché <i>sugli operatori</i>, per eliminare la distorsione da selezione.</p> <p>Caso particolare di <i>confronto attorno alla soglia</i>: il servizio è razionato, e l'ammissibilità allo stesso è ristretta a disoccupati al di sopra di una data soglia di una variabile, fissata in maniera ben definita.</p> <p>B. Differenza fra medie di trattati e controlli.</p>
		Condizione occupazionale dopo t mesi	B. Esperimento (su un campione casuale di iscritti alle liste)	
Incentivi alle imprese per la stabilizzazione dei precari	Ammissibilità all'incentivo	Trasformazioni in contratti a tempo indeterminato	<p>Tutte le imprese da tempo t a $t+s$</p> <p>+ restrizione territoriale;</p> <p>+ restrizione lavoratori precari età $\geq x$ anni</p>	<p>Confronto tra periodi con/senza incentivi</p> <p>+ confronto tra aree territoriali con/senza incentivi</p> <p>+ confronto attorno alla soglia di età.</p> <p><i>Serie storiche di trasformazioni in contratti a tempo indeterminato, totale e con incentivi.</i></p> <p>Possibile <i>entry effect</i> complica l'analisi.</p> <p>- Utilizzo dei metodi segnalati – confronto tra periodi con/senza incentivi, confronto tra aree con/senza incentivi, confronto attorno alla soglia di età – <i>per la prima volta</i> in cui la politica viene introdotta.</p> <p>- Negli altri casi, servono notevoli affinamenti nei metodi e maggiori informazioni.</p>
		[Assunzioni immediate a tempo indeterminato]		

2.2.2. *Gli incentivi alle imprese per la stabilizzazione dei precari*

Questo intervento è un esempio di un'ampia gamma di politiche, caratterizzate dal tentativo di favorire comportamenti "desiderabili" da parte di agenti economici, offrendo loro un incentivo condizionato all'adozione di quei comportamenti.

Nel caso dell'incentivo alle imprese per stabilizzare i precari, l'intervento può essere universale e omogeneo a livello nazionale, e allora la discontinuità sarà soltanto temporale: prima del tempo t la politica non era in essere e non vi erano quindi imprese ammissibili; poi lo è da t a $t+s$ e in questo lasso di tempo tutte le imprese sono ammissibili. Oppure può essere adottato, per un dato arco temporale, in una regione e in altre no, e in questo caso la condizione di ammissibilità/non ammissibilità varia anche nella dimensione geografica.

Le ragioni di incertezza sull'efficacia di questo tipo di politica pubblica sono molteplici. Una risposta possibile all'offerta dell'incentivo è la mancata adesione da parte degli agenti economici. L'incentivo è inefficace perché ritenuto poco attraente da coloro che dovrebbero fruirne. In questo caso la valutazione degli effetti è banale: l'effetto semplicemente non c'è.

La risposta diametralmente opposta è la massiccia adesione e richiesta di incentivi. Quest'ultima non può tuttavia essere interpretata automaticamente come evidenza dell'efficacia dell'incentivo, perché i comportamenti in questione avrebbero potuto essere adottati in egual misura, o in parte, anche in assenza di incentivo. L'effetto, anche in caso di massiccia adesione, può essere nullo (nell'esempio, è possibile che nessun precario sia stato stabilizzato *grazie* all'incentivo) o all'opposto massimo (che tutti lo siano stati *grazie* all'incentivo), come sono possibili combinazioni intermedie. Una miope presunzione di efficacia porterebbe, invece, a ritenere che tutte le stabilizzazioni dei precari *finanziate* dall'incentivo siano *dovute* alla presenza dell'incentivo, escludendo la possibilità che la decisione di stabilizzare un lavoratore precario sia stata già presa per ragioni che prescindono dalla presenza dell'incentivo.

L'unico modo per stabilire il reale effetto degli incentivi è ricostruire cosa sarebbe accaduto in loro assenza: quanti contratti precari sarebbero stati trasformati in contratti a tempo indeterminato in assenza di incentivi?

2.3. *Ricostruire la situazione controfattuale*

2.3.1. *Distorsione da selezione*

È la regola, più che l'eccezione, che per gran parte delle politiche pubbliche, segnatamente quelle che si traducono nella fornitura di un servizio non obbligatorio, la coincidenza fra insieme degli ammissibili e insieme dei trattati non si realizzi: necessariamente nel caso in cui il servizio sia razionato; di fatto nella grande maggioranza degli altri casi. In generale, infatti, soltanto una parte dei soggetti ammissibili richiede il

servizio¹³. Ciò si traduce in un processo di autoselezione: i soggetti trattati in generale sono sistematicamente diversi dai non-trattati per ragioni che non dipendono dal successo dell'intervento, bensì dal precedente processo attraverso il quale gli uni hanno scelto di essere trattati e gli altri di non esserlo. Tale processo rimanda a “differenze delle condizioni di partenza” dei soggetti – caratteristiche ascrivibili, socio-culturali, economiche, motivazionali, ecc. – che da un lato rendono trattati e non-trattati diversi in modo sistematico fin dall'origine, vale a dire prima che l'intervento abbia luogo, e dall'altro influiscono sulla variabile-risultato.

Se vi è razionamento, o comunque una qualche discrezionalità nell'assegnazione al servizio da parte del personale incaricato dell'erogazione dei servizi, vi può essere poi un processo sistematico di selezione, che porta a scegliere alcuni soggetti perché siano esposti all'intervento, o comunque all'intervento migliore (o peggiore).

È immediato riconoscere che, ai fini della valutazione degli effetti di una politica, i due processi – autoselezione e selezione da parte degli operatori – sono equivalenti, e pongono perciò problemi analoghi. Ad essi usualmente ci si riferisce, globalmente, col termine *distorsione da selezione*¹⁴. Si tratta di distorsione, nel senso che questa “differenza nelle condizioni di partenza” – indotta da comportamenti e scelte degli ammissibili e/o del personale che eroga il servizio – impedisce di utilizzare come stima dell'effetto medio del trattamento la differenza osservata tra trattati e non-trattati. Si tratta di distorsione dovuta a selezione, perché dipende dalle differenze tra trattati e non-trattati che pre-esistono al trattamento e sono dovute al processo con cui i singoli individui si sono autoselezionati o sono stati selezionati.

C'è un'importante eccezione a questa regolarità empirica. Se il processo di selezione avesse luogo mediante *assegnazione casuale* dei soggetti all'uno o all'altro dei due gruppi, la distorsione da selezione risulterebbe nulla per costruzione. Pertanto, l'effetto medio dell'intervento verrebbe stimato in modo esente da distorsione mediante la differenza tra i risultati medi ottenuti dai soggetti, rispettivamente, esposti e non esposti all'intervento. Da qui viene la naturale attrattiva che esercita il ricorso agli esperimenti¹⁵, come modo per risolvere il problema della selezione eliminandolo alla radice. Tuttavia, per ragioni sia etiche che pratiche, l'utilizzo di esperimenti tende ad essere circoscritto ai casi in cui un intervento sia messo in atto come progetto-pilota, a scala ridotta, con lo scopo specifico di valutarne l'efficacia: le cosiddette *demonstrations* statunitensi, molto diffuse¹⁶, o i *pilots* inglesi.

D'altra parte, persuasive riserve sulla praticabilità del metodo sperimentale per la valutazione di interventi sociali sono state formulate soprattutto da Heckman e i suoi

¹³ Per un altro verso, può accadere che soggetti non ammissibili riescano, in qualche modo, ad accedere al servizio. Per semplicità, trascuro questa eventualità.

¹⁴ Il termine traduce l'inglese *selection bias*.

¹⁵ Uso questo termine invece della traduzione letterale dell'inglese *randomized experiments*, che in italiano è ridondante.

¹⁶ Per una rassegna vedi Greenberg e Shroder (1997).

coautori (Heckman e Smith 1995; Heckman 1996; Heckman, Lalonde e Smith 1999). Essi hanno argomentato come l'assegnazione casuale abbia essa stessa effetti comportamentali che non possono essere ignorati. Si tratta, del resto, di una consapevolezza maturata da tempo in ambito medico, dove per i *Randomized Clinical Trials* sono stati introdotti i disegni cosiddetti "cieco" e "doppio cieco": il primo per controllare l'effetto *placebo* associato al trattamento (tramite somministrazione di un *placebo* ai non-trattati); il secondo per neutralizzare la maggiore interazione medico-paziente che si ha quando quest'ultimo è un trattato (celando la condizione del paziente al medico). In generale, anche nell'ipotesi che sia possibile praticare l'assegnazione casuale dei soggetti ai due gruppi degli esposti e dei non esposti, in ogni esperimento svolto in un contesto sociale, dunque non in laboratorio, si hanno fenomeni di uscita dal gruppo degli esposti (eventualmente dopo aver beneficiato in misura parziale del trattamento) e, all'opposto, di accesso di soggetti del gruppo dei non esposti a servizi parecchio prossimi a quelli somministrati agli esposti, offerti da altri enti e/o nell'ambito di altri programmi¹⁷. Ciò non significa che la stima di effetti di interesse diventi impossibile. Tali parametri restano identificabili, ma solo a prezzo di ulteriori assunzioni e condizioni. L'impiego del metodo sperimentale, dunque, non può prescindere da un'attenta considerazione degli effetti comportamentali che esso stesso induce. Sulla scorta di una lucida rassegna degli elementi di forza e di debolezza degli esperimenti, anche un loro convinto assertore, Berk (2005: 19), conclude che «*gli esperimenti randomizzati poggiano su basi più complicate, oscure e fragili di quanto alcuni ricercatori comprendano. [...] Pertanto, gli esperimenti randomizzati non sono il "gold standard"»*. A suo avviso, rimangono comunque il «*bronze standard*».

Al di fuori dell'ambito di valutazioni condotte col metodo sperimentale, l'assegnazione dei soggetti a un intervento è frutto di un processo sistematico di selezione. Tale processo, si noti, non è controllabile o manipolabile dall'analista, e in generale gli è anche ignoto *a priori*. Sta quindi all'analista il compito di ricostruire il processo di selezione dei soggetti esposti, al fine di individuare l'insieme delle caratteristiche rispetto alle quali essi risultano verosimilmente diversi in modo sistematico dai soggetti non esposti. L'identificazione dell'effetto dell'intervento richiede, quindi, di (riuscire a) rendere nulla la distorsione da selezione.

2.3.2. *Uno strumento per analizzare relazioni di causa-effetto: il diagramma causale di Pearl*

Per chiarire le strategie che possono essere adottate a questo fine, conviene prendere le mosse da un diagramma causale dei processi di assegnazione al trattamento e di determinazione della variabile-risultato, schematizzato nel grafico 1. Il grafico è disegnato usando le parti più intuitive dell'apparato sviluppato da Pearl (2000) e dandone la spiegazione minimale. Ogni nodo del grafico rappresenta una variabile (casuale) osservabile; ogni freccia orientata sta a significare che la variabile all'origine causa la

¹⁷ Sovente ci si riferisce a questi due problemi, non peculiari alle valutazioni con metodo sperimentale ma indubbiamente esacerbati nelle stesse, rispettivamente con i termini di «*treatment group dropout*» e di «*control group substitution bias*» (Heckman, Lalonde e Smith 1999: 1903-1905 e 1907-1912).

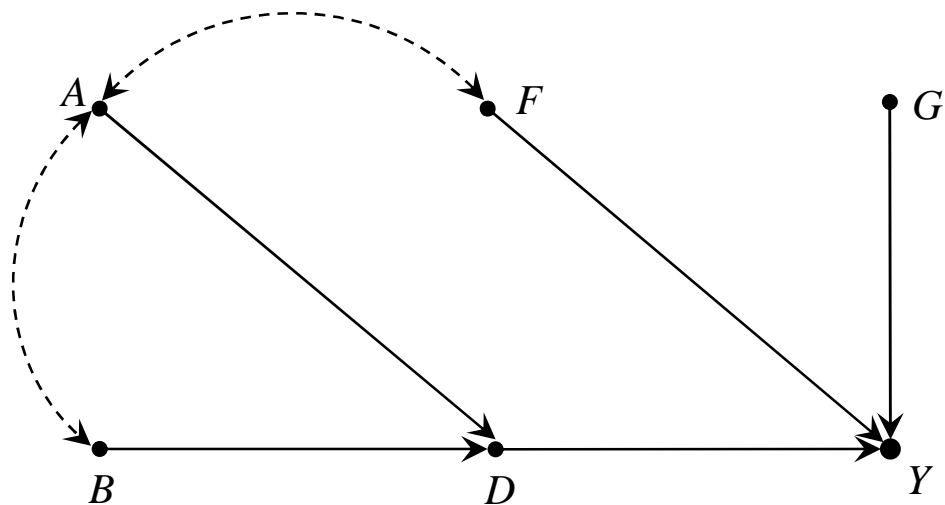


Grafico 1: *Diagramma causale di un modello di assegnazione al trattamento (D) e di determinazione della variabile-risultato (Y)*

Fonte: adattato da Morgan e Winship (2007: 25).

variabile al termine della freccia; ogni arco tratteggiato e bidirezionale rimanda all'esistenza di variabili comuni non osservate che causano entrambe le variabili terminali. Le relazioni sono, in generale, non lineari.

Nel grafico 1, D è una variabile binaria che descrive la presenza-assenza del trattamento, e il nostro obiettivo è stimare l'effetto di D su Y , la variabile-risultato. La questione che ci poniamo è la seguente: data la struttura delle relazioni causali rappresentata nel grafico, quali variabili dobbiamo osservare e come dobbiamo usarle per stimare l'effetto causale di D su Y ?

Per chiarire meglio la questione, è utile riferirsi ancora all'esempio dei servizi di orientamento alla ricerca di lavoro. In tal caso, D rappresenta l'aver ricevuto o meno il servizio di orientamento, e Y , poniamo, la condizione occupazionale – occupato oppure disoccupato – a distanza di sei mesi. Quanto alle altre variabili, si può ipotizzare che A sia un indicatore riassuntivo delle abilità e motivazioni dell'iscritto alla lista di disoccupazione, B la priorità data dall'operatore nell'ammetterlo al servizio, F l'appetibilità del disoccupato da parte dei datori lavoro, G la situazione generale del mercato del lavoro. Sulla base del grafico 1, A , B e F influenzano, direttamente o indirettamente, sia il processo di selezione sia la variabile-risultato, mentre ciò non si dà per G . Come appena detto, il nostro obiettivo è stimare l'effetto medio di D su Y ; a questo fine occorre, ed è sufficiente, guardare a quanto varia Y al variare di D controllando per A , B e F .

La strada percorsa per lungo tempo da economisti ed econometrici è stata quella di muovere da un tale grafico e specificare un modello completo dei processi di assegnazione

al trattamento e di determinazione della variabile-risultato. Tale modello consiste in un sistema di equazioni, delle quali si specificano la forma funzionale e le restrizioni necessarie per tener conto delle componenti non osservabili, nonché di altre informazioni *a priori* suggerite dalla teoria economica. L'insieme delle restrizioni imposte porta a un modello che può essere stimato. La stima di tutte le cause dell'effetto totale sulla variabile-risultato consente anche di stimare l'effetto causale (marginale) di D su Y , perché la considerazione contestuale delle altre cause realizza la condizione *ceteris paribus*.

Questa strategia ha conosciuto un decisivo avanzamento col programma di ricerca della Cowles Commission, intorno alla metà del secolo scorso¹⁸, ed è stata poi sviluppata dallo “*structural research program*” di orientamento microeconomico degli anni 1970 e '80. Ma, nelle parole di uno dei maggiori cultori e contributori, James Heckman, premio Nobel per l'economia nel 2000, «*dalla metà degli anni 1960 il programma di ricerca della Cowles fu largamente percepito come un successo intellettuale ma un fallimento empirico. [E ...] il documentato percorso dei risultati empirici dello “structural research program” non è certo impressionante*» (Heckman 2000: 48 e 76).

2.3.3. *Dai modelli parametrici ai metodi non parametrici: uso intensivo e rigoroso dei dati, e risultati robusti*

Prende così piede una nuova strategia. Essa tende a evitare le assunzioni su forme funzionali e restrizioni *a priori* tipicamente imposte ai modelli, poco credibili e sovente produttive di risultati discordanti, quindi non affidabili. Cerca invece trasparenza e robustezza nell'uso dell'evidenza empirica. È ancora Heckman, che, pur continuando a rivendicare l'attenzione a preferenze, vincoli e scelte degli agenti (insomma alla teoria economica, o meglio, generalizzando, alle teorie del comportamento sociale), in un monumentale saggio scritto con Lalonde e Smith sulla valutazione degli effetti di politiche attive del lavoro delinea in maniera persuasiva il nuovo programma di ricerca. Le due affermazioni basilari (vedi Heckman, Lalonde e Smith, 1999: 1868-1869) sono le seguenti.

(a) «*Dati migliori aiutano moltissimo. I dati utilizzabili dalla maggior parte degli analisti [per la valutazione degli effetti di interventi sociali] sono estremamente grezzi. Per rimediare alle insufficienze nei dati sottostanti, troppo è stato chiesto ai metodi econometrici. [...] La migliore soluzione al problema della valutazione degli effetti consiste nel migliorare la qualità dei dati sui quali sono condotte le valutazioni, non*

¹⁸ La nozione di un cambiamento *ceteris paribus*, cioè di quel che gli economisti intendono per effetto causale, per il caso di variazioni indipendenti si trova già in Marshall (1961: 36 [1860]), enunciata in questi termini: «*Si dice talvolta che le leggi dell'economia sono “ipotetiche”. Naturalmente, al pari di ogni altra scienza, l'economia studia gli effetti che saranno prodotti da certe cause, non in via assoluta, ma sotto la condizione che le altre cose siano uguali e che le cause possano svolgere i loro effetti indisturbate. Pressoché tutte le dottrine scientifiche, quando sono formulate in modo formale e accurato, conterranno per l'effetto qualche condizione che le altre cose siano uguali; l'azione delle cause di cui di tratta si suppone che sia isolata; alcuni effetti sono attribuiti ad esse, ma soltanto nell'ipotesi che non si permetta d'intervenire ad alcuna causa, salvo che a quelle specificamente permesse*». L'importante risultato dei ricercatori della Cowles Commission è stato di estendere la nozione di causa mantenendo la clausola *ceteris paribus*, ma rilassando la condizione di variazione indipendente e ammettendo cause correlate nel contesto di modelli “*well-posed*” (nel senso di Heckman, 2000: 47-48).

nello sviluppo di metodi econometrici formali per aggirare il problema di dati inadeguati».

- (b) «È importante confrontare soggetti confrontabili. Molte valutazioni non sperimentali identificano il parametro di interesse [\equiv l'effetto del programma] usando estrapolazioni basate su forme funzionali inappropriate [di funzioni di regressione o analoghe], imposte al fine di rendere confrontabili soggetti che in realtà confrontabili non sono. Il pregio principale dei metodi non parametrici per risolvere il problema della distorsione da selezione [e quindi stimare appropriatamente l'effetto di un programma] consiste nel fatto che, applicati rigorosamente, essi obbligano l'analista a confrontare soltanto soggetti confrontabili».

I rischi nell'uso di modelli che ricorrono ad assunzioni arbitrarie, «al fine di [... trattare come] confrontabili soggetti che in realtà confrontabili non sono», sono evidenziati nel semplice, ipotetico caso del grafico 2. Si assume che vi sia una sola variabile responsabile del processo di selezione, lì denotata X , e si stimano le equazioni di regressione lineare rispettivamente per trattati (T) e non-trattati (NT). Restando ancora all'esempio dei servizi di orientamento al lavoro, si può ipotizzare che X sia un indicatore riassuntivo delle abilità e motivazioni dell'iscritto alla lista di disoccupazione.

Conseguentemente, la differenza fra l'equazione stimata sui trattati, la retta in alto a destra nel grafico, e quella stimata sui non-trattati, la retta in basso a sinistra, è una stima dell'effetto medio sui soggetti con caratteristica X_i . La media calcolata rispetto alla distribuzione di X nell'insieme dei trattati fornisce quella che *prima facie* appare essere una stima dell'effetto medio sui trattati.

Tale modo di procedere espone al rischio di scambiare per stime dell'effetto medio sui trattati *basate sull'evidenza empirica* a disposizione – i risultati di trattati e non-trattati, stime che tali non sono, perché ottenute confrontando il risultato dei trattati con il risultato di non-trattati *diversi* dai trattati rispetto alla caratteristica X . Il grafico 2 mette in luce tale rischio presentando il caso limite in cui l'insieme dei valori di X osservati rispettivamente sui trattati (X_1) e sui non-trattati (X_0) è disgiunto. In questa situazione, in corrispondenza di ogni singolo valore di X osservato per i trattati, cioè in X_1 , lo stimatore di regressione ottiene una stima dell'effetto *estrapolando* il valore medio dalle regolarità osservate per i non-trattati, cioè in X_0 (nel grafico ciò viene evidenziato con la retta tratteggiata), *non* confrontando i risultati ottenuti da soggetti trattati e da non-trattati in corrispondenza di *quel* particolare valore di X , confronto quest'ultimo irrealizzabile data l'assenza di elementi in comune tra i due insiemi X_1 e X_0 .

Palesamente, si tratta di stime dell'effetto che poggiano, oltre che sull'evidenza empirica, sull'assunto che le due regressioni mantengano la stessa forma funzionale osservata al di fuori dei rispettivi insiemi, X_1 e X_0 . In questo senso, lo stimatore poggia su un'assunzione arbitraria ed è quindi fragile.

La ricostruzione credibile della situazione controfattuale consiste appunto nell'evitare questi rischi, tramite un uso intensivo, e nello stesso tempo rigoroso, di insiemi di dati adeguati.

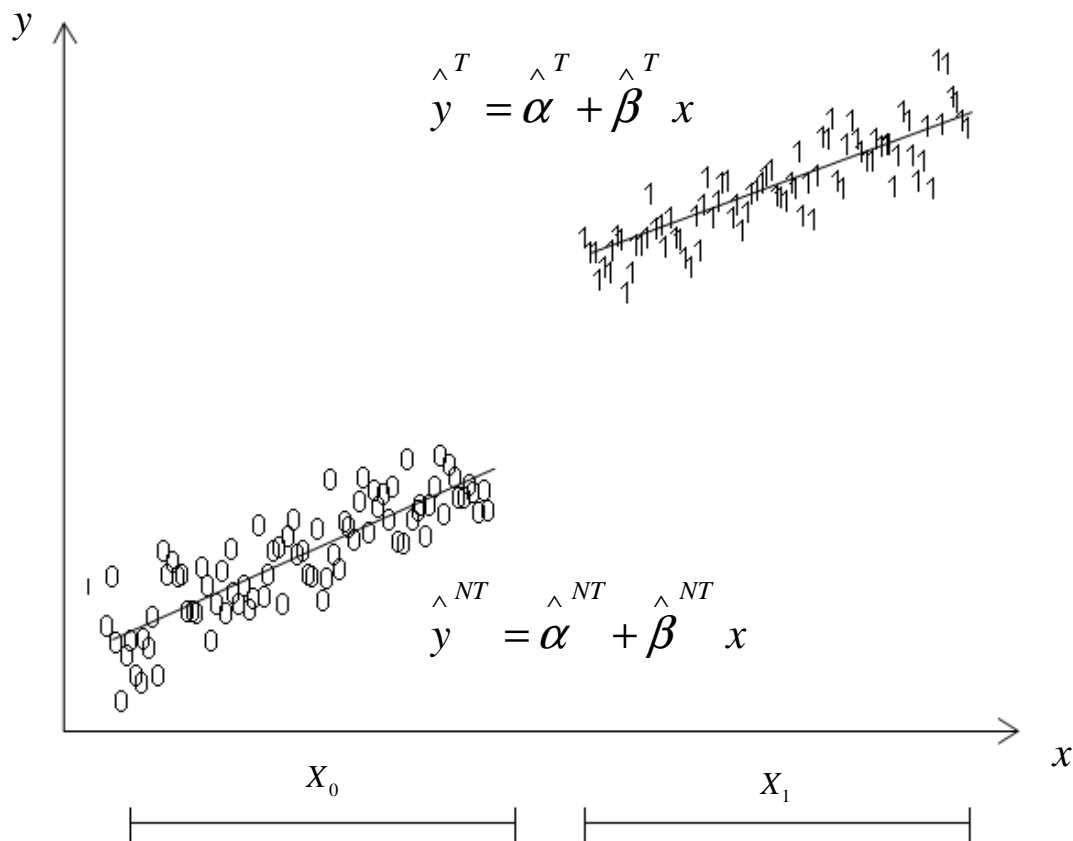


Grafico 2: *Stima dell'effetto medio con un modello di regressione lineare: ipotetico risultato per una variabile obiettivo Y, nel caso di insiemi X₁ e X₀ disgiunti per trattati e non-trattati*

Fonte: Rettore, Trivellato e Martini (2003: 327).

L'obiettivo consiste nel determinare un plausibile gruppo di confronto, i *controlli*, formato da soggetti non-trattati *equivalenti* ai trattati rispetto alle caratteristiche rilevanti per la variabile-risultato, salvo appunto il fatto che non sono stati esposti all'intervento. Ciò comporta di individuare *tutte le variabili X responsabili del processo di selezione* e di *reperire informazioni affidabili* sulle stesse, e costruire quindi il gruppo di controllo condizionandosi ad esse, eliminando così la distorsione da selezione. In un modo affatto diverso, e robusto, ci si riconduce così alla condizione *ceteris paribus*: trattati e controlli sono equivalenti perché sono stati "pareggiati" rispetto all'insieme delle variabili che, in assenza del trattamento, influiscono sulla variabile-risultato. Appropriati stimatori delle differenze fra trattati e controlli forniscono stime dell'effetto dell'intervento.

Nella sua versione più marcatamente empirica, questa strategia è visualizzata nel grafico 3 confrontato, per contrasto, con il grafico 1. Esso prescinde dalla specificazione dello stesso modello grafico (che in prima istanza abbiamo assunto come "vera", ma che,

pur nella sua generalità, può non essere tale), utilizza tutte le variabili osservate $\{A, B, F, G\}$ che causano Y , e rispetto ad esse realizza con opportuni metodi di “pareggiamento” la condizione di equivalenza appena detta. Tutti i sentieri che portano a Y (ovviamente, a meno di $D \rightarrow Y$) sono evidenziati con segmenti punteggiati, a significare che rispetto alle variabili dalle quali essi si dipartono si è realizzata la condizione di equivalenza. Marcata in nero resta soltanto la freccia orientata da D a Y , dove D , a seguito dell’operazione di “pareggiamento”, contrasta ora trattati e controlli (cioè non-trattati equivalenti ai primi), sicché la differenza fra le medie della variabile-risultato per i due gruppi è una stima dell’effetto dell’intervento.

È importante mettere in chiaro che questa strategia opera un consapevole ridimensionamento delle aspirazioni conoscitive. Non ambisce a individuare tutte le cause della variabile-risultato, e quindi a stimarne congiuntamente gli effetti. Mira, invece, alla *stima dell’effetto di una (sola) causa*: quella d’interesse, l’intervento. È altrettanto importante notare, peraltro, che tale stima ha due caratteristiche decisamente apprezzabili:

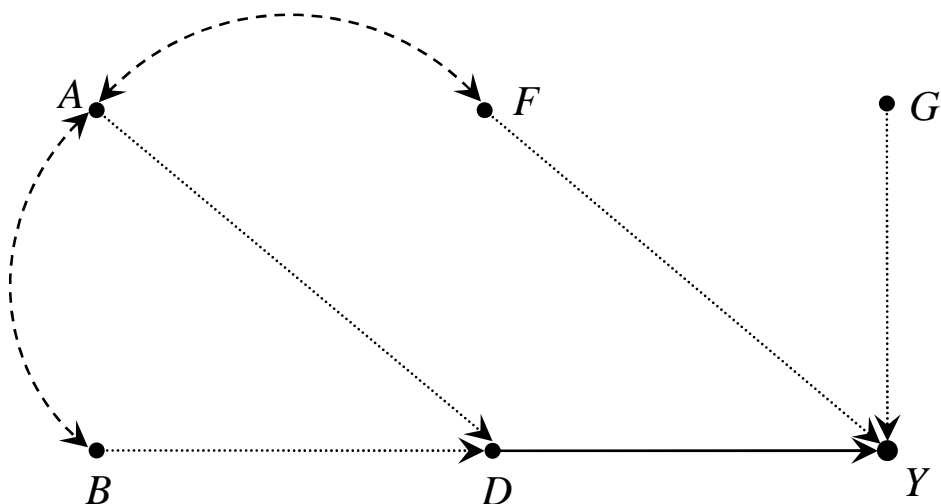


Grafico 3: *Diagramma causale di un modello di assegnazione al trattamento (D) e di determinazione della variabile-risultato (Y), dove i sentieri causali che portano a D e a Y (a meno di $D \rightarrow Y$) sono stati eliminati mediante condizionamento rispetto alle variabili $\{A, B, F, G\}$ ¹⁹*

¹⁹ Nella modellazione teorico-grafica di Pearl (2000), ciò equivale a bloccare tutti i sentieri che precedono quello (“back-door paths”) da D a Y . I due back-door paths che generano un’associazione addizionale non-causale fra D e Y sono (i) da D ad A a F a Y , e (ii) da D a B ad A a F a Y . Assumendo la specificazione del modello grafico come “vera”, è dunque sufficiente osservare e quindi condizionare rispetto a $\{A, B\}$ oppure osservare e quindi condizionare rispetto a $\{F\}$. Queste due strategie di condizionamento nel linguaggio figurato usato nel testo principale, di realizzazione dell’equivalenza fra trattati e controlli – sono sufficienti minimali, nel senso che l’analista può osservare e quindi condizionare rispetto a ogni sottoinsieme delle variabili osservate in $\{A, B, F, G\}$ che contenga $\{A, B\}$ oppure $\{F\}$. Palesemente, data la specificazione del modello grafico il condizionamento rispetto all’insieme delle variabili osservate $\{A, B, F, G\}$ è ridondante.

D'altra parte, esso è valido, quindi robusto, anche se la specificazione del modello grafico fosse errata, purché l'insieme $\{A, B, F, G\}$ includa tutte le variabili responsabili del processo di selezione.

(i) è robusta, perché non poggia su assunzioni arbitrarie²⁰; (ii) è facile da comunicare al *policy-maker*, e più in generale all'opinione pubblica.

3. *Le principali strategie di identificazione degli effetti di una politica*

In via generale, le varie strategie per identificare effetti medi di una politica pubblica mirano a realizzare *ex-post* la condizione *ceteris paribus*, che nel caso degli esperimenti è assicurata per costruzione. Trattati e non-trattati devono essere (resi) equivalenti in tutti gli aspetti rilevanti, cioè con riguardo alle caratteristiche osservabili e non osservabili che influiscono sia sulla (auto)assegnazione allo stato di trattato/non trattato sia sulla variabile risultato, eccetto che nella probabilità di essere esposti o non esposti al trattamento.

3.1. *La selezione nelle osservabili e il metodo dell'abbinamento*

Assumendo di osservare tutte le variabili X responsabili della distorsione da selezione, un metodo robusto per ottenere una stima dell'effetto medio sui trattati consiste nell'*abbinare* a ogni trattato un soggetto non-trattato che presenta le stesse caratteristiche X (sicché il suo risultato può essere utilizzato come stima del risultato, controfattuale, che il trattato abbinato avrebbe conseguito se non fosse stato esposto al trattamento), e nel calcolare la media delle differenze tra i risultati osservati per le coppie di soggetti abbinati.

Operativamente, sovente si utilizza un indicatore detto "propensione al trattamento", cioè la probabilità che il soggetto venga trattato in funzione delle sue caratteristiche osservabili²¹, il quale per ogni soggetto trattato e non-trattato riassume l'informazione sull'insieme delle variabili X . A ciascun trattato viene abbinato il non-trattato più simile in termini di propensione al trattamento, in generale entro un livello di somiglianza minima pre-definito. Si ha così uno stimatore non parametrico dell'effetto medio (o meglio una famiglia di stimatori, perché vi sono molteplici procedure di abbinamento che possono essere adottate).

Alla luce di queste chiarificazioni, torniamo ora alla tabella 1, segnatamente all'esempio dei servizi di orientamento al lavoro, emblematico di una verosimile distorsione da selezione. Vi è infatti un'alta probabilità di autoselezione dei disoccupati nell'accedere al servizio, nonché di selezione dei migliori nell'assegnazione al servizio, o ai servizi migliori, da parte degli operatori dei Centri per l'impiego. Le soluzioni che si possono prospettare sono due.

²⁰ Tale robustezza è rafforzata dal fatto che sovente si danno strumenti (i cosiddetti "tests di specificazione") per saggiare la bontà delle procedure di "pareggiamento", in altre parole per verificare che i controlli siano equivalenti ai trattati rispetto alle caratteristiche rilevanti per la variabile-risultato, a meno dell'effetto dell'intervento.

²¹ L'indicatore assume quindi un valore compreso fra 1 e 0.

Se l'intervento è realizzato su larga scala, lasciando ampia iniziativa ai singoli Centri per l'impiego, è necessario disporre di ricche basi di dati individuali (caratteristiche socio-demografiche, scolarità e formazione ulteriore, storia lavorativa, ecc.) sugli iscritti alle liste di disoccupazione – ammessi, aspiranti esclusi e non aspiranti, nonché sugli operatori che hanno provveduto ad assegnare o erogare i servizi, per (cercare di) ricostruire in maniera attendibile il processo di selezione e, quindi, stimare correttamente l'effetto dell'intervento. Palesemente, la disponibilità di dati adeguati, *ex-post* o meglio perché se ne è pianificata *ex-ante* la rilevazione proprio allo scopo di analizzare il processo di selezione e neutralizzarne gli effetti, riveste un'importanza decisiva.

Per l'ipotetico programma di servizi di orientamento al lavoro, il grafico 4 mostra le distribuzioni dell'indicatore che riassume l'informazione sulle variabili X – nel caso considerato è la propensione ad accedere al servizio di orientamento – rispettivamente per l'insieme dei trattati, in grigio, e per l'insieme dei non-trattati, in bianco.

L'evidenza fornita dal grafico 4 può essere riassunta in tre proposizioni: (i) per larga parte dei trattati che presentano un dato valore dell'indicatore corrisponde un adeguato sotto-insieme di non-trattati abbinabili; (ii) vi è un non trascurabile gruppo di trattati, quelli con valore dell'indicatore superiore a 0,7, ai quali non corrisponde alcun non-trattato; (iii) vi è un sotto-insieme, anch'esso considerevole, di non-trattati, quelli con

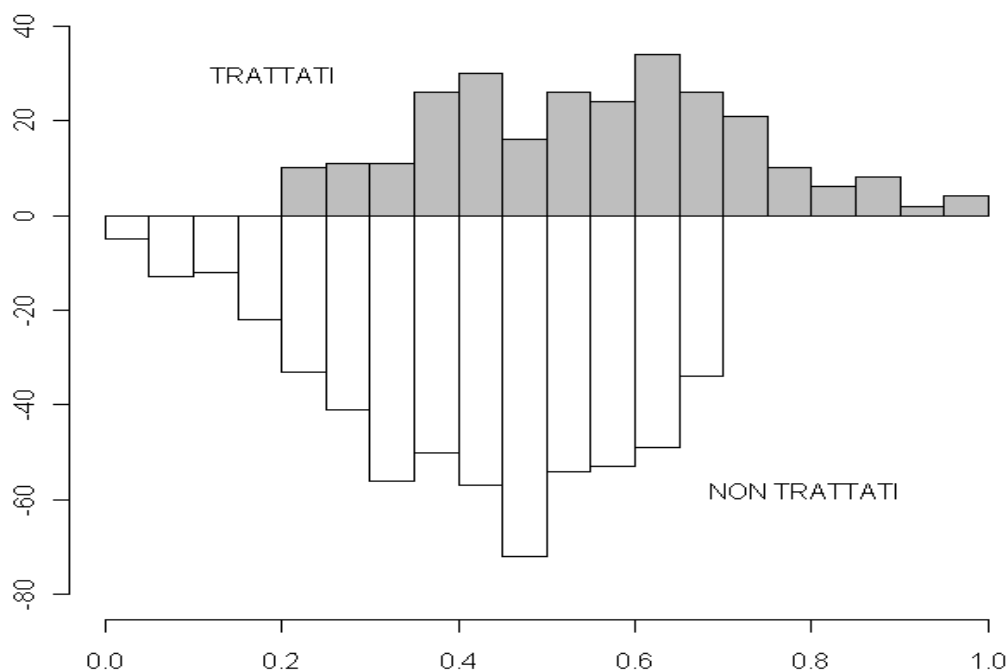


Grafico 4: *Ipotetico esempio di distribuzione dei trattati e dei non-trattati di un servizio di orientamento alla ricerca di lavoro sulla base di un indicatore che varia da 0 a*

1, e per ogni soggetto riassume l'informazione sull'insieme delle variabili X responsabili della processo di selezione.

valore dell'indicatore inferiore a 0,2, che presentano caratteristiche X dissimili dai trattati, e sono quindi inutilizzabili per l'abbinamento.

Dal grafico 4 discende un importante insegnamento di portata generale. Con procedure di abbinamento (che, è bene ribadirlo, utilizzino tutte le variabili X responsabili della distorsione da selezione) siamo in grado di ottenere una stima robusta dell'effetto medio del trattamento *limitatamente al sottoinsieme di trattati e non-trattati che giacciono nello spazio comune delle caratteristiche X* (ovvero dell'indicatore che le riassume). D'altra parte, il notevole vantaggio di uno stimatore basato sull'abbinamento sta nel fatto che *obbliga* l'analista a rendersi conto di ciò: che deve restringere il confronto, e la stima dell'effetto, ai soggetti effettivamente confrontabili; detto altrimenti, che l'effetto dell'intervento è *intrinsecamente* non identificabile per i trattati con caratteristiche X al di fuori dello spazio comune.

Restando all'esempio, nulla possiamo dire dell'effetto dell'intervento per i disoccupati "migliori", cioè con un insieme di caratteristiche – abilità e motivazioni personali, priorità data dall'operatore nell'ammetterli al servizio, appetibilità da parte dei datori di lavoro – che si traducono in un valore dell'indicatore superiore a 0,7, semplicemente perché non vi sono non-trattati con i quali poterli abbinare. La stima dell'effetto dell'intervento è possibile per tutti gli altri trattati.

Il grafico 4 fornisce, infine, un'ultima evidenza interessante. Vi è un sottogruppo di non-trattati, i "peggiori", con indicatore inferiore a 0,2, sui quali pure non possiamo dire alcunché circa l'eventuale effetto che su di essi potrebbe avere il servizio se fosse loro fornito. Per essi, infatti, non vi sono trattati con i quali poterli abbinare. Ciò segnala un aspetto importante nel merito. Il programma, al quale sono ammissibili tutti gli iscritti alle liste di disoccupazione, nei fatti è selettivo: ne sono esclusi i soggetti con caratteristiche peggiori, i più bisognosi di servizi di sostegno per (ri)accedere al lavoro. Dal processo di valutazione degli effetti del programma si apprende dunque un suo mal-funzionamento, di rilievo nella prospettiva di ridisegnarlo.

La seconda soluzione prospettabile, ma di fatto molto meno utilizzata, è il ricorso al *social experiment*. Se il servizio di orientamento al lavoro viene introdotto per la prima volta, o viene cambiato in maniera consistente, e c'è un genuino interesse del *policy-maker* a "imparare" su come funzioni e quale effetto abbia, si può optare per un intervento-pilota, realizzato su scala ridotta e disegnato appunto in termini di esperimento. Come mostrato nella sez. 2.3.1, i due gruppi di trattati e non-trattati sono in media equivalenti per costruzione, detto altrimenti i non-trattati costituiscono l'appropriato gruppo di controllo, e quindi la differenza in media nella variabile-risultato fra trattati e controlli dà la stima dell'effetto medio del servizio erogato.

3.2. Il confronto attorno alla soglia di discontinuità

Un'interessante strategia di identificazione, alla quale accenno anche nella tabella 1 con la dizione “confronto attorno alla soglia”, si ha per una classe particolare di programmi, i quali prevedano razionamento, e a questo fine abbiano prefissato la soglia di ammissibilità sulla base di un criterio ben definito – un dato valore di una variabile o una data posizione nella graduatoria dei candidati rispetto alla stessa.

Il grafico 5 fa ancora riferimento a un ipotetico programma di orientamento al lavoro dei disoccupati, riservato a coloro che abbiano ottenuto i punteggi più alti in un test attitudinale. La soglia, denotata S , è fissata pari a 56, e coloro che si collocano sopra tale soglia partecipano al programma, mentre coloro che ne stanno sotto sono esclusi. Per costruzione, vi è quindi una sola variabile *esogena* (non manipolabile né dai candidati né dai gestori del programma) e *osservata* dall'analista, che determina univocamente la partecipazione al programma. La soglia di ammissibilità rappresenta un “punto di discontinuità”, e attorno ad essa si determina un situazione simile alla randomizzazione: i trattati che stanno immediatamente al di sopra della soglia sono *equivalenti* ai non-trattati che ne stanno immediatamente al di sotto, rispetto sia alle variabili osservabili che a quelle non osservabili. In tal caso, il confronto tra trattati e non-trattati condotto attorno alla soglia di ammissibilità – più precisamente, la differenza nella media della variabile-risultato per i trattati che stanno immediatamente al di sopra della soglia e quella dei non-trattati immediatamente al di sotto della soglia – identifica l'effetto della politica.

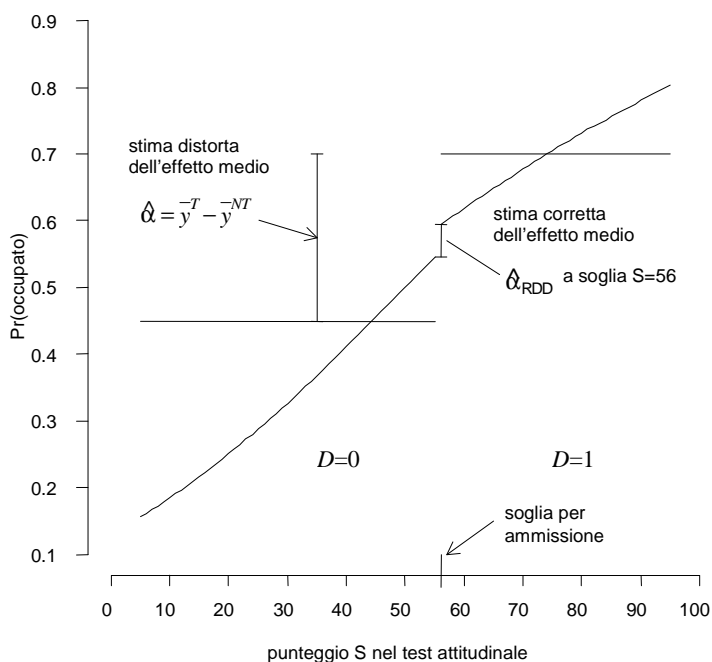
Questa strategia di valutazione è nota come *Regression Discontinuity Design* (RDD)²². Per essa si impone un importante *caveat*: se l'effetto della politica varia da soggetto a soggetto, in particolare varia in relazione alla lontananza del soggetto dalla soglia, la stima dell'effetto che si ottiene è circoscritta ai soggetti collocati nell'intorno alla soglia. Detto nel linguaggio proprio degli studi di valutazione, questa strategia garantisce la “validità interna” dell'effetto stimato, ma a prezzo di una limitata “validità esterna”.

In presenza di eterogeneità dell'effetto, il grafico 5a mette bene in luce come il RDD fornisca un stima corretta dell'effetto medio per la circoscritta sottopopolazione dei soggetti prossimi alla soglia di ammissione al/esclusione dal programma. Per un altro verso, evidenzia come una stima dell'effetto medio ottenuta come differenza fra i risultati medi rispettivamente di trattati e non trattati sarebbe affetta da una forte distorsione da selezione.

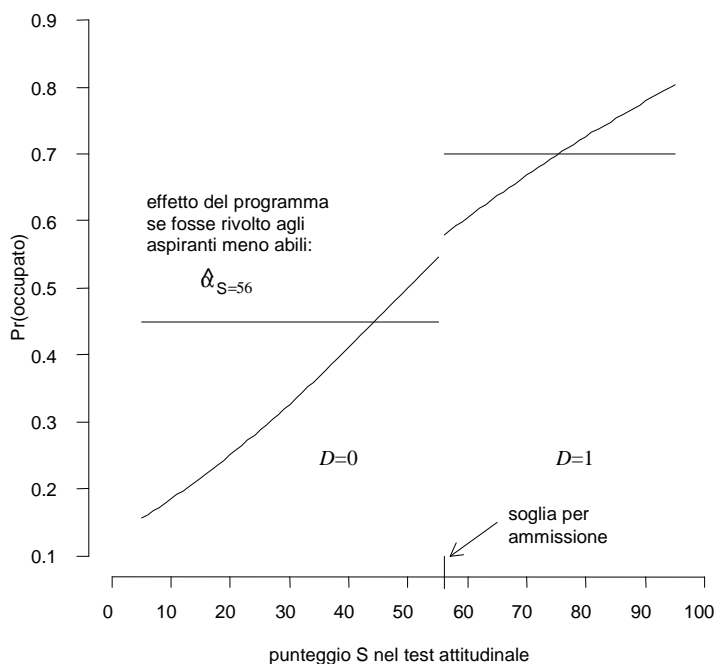
Valutazioni circa l'effetto del programma su soggetti diversi dalla circoscritta sottopopolazione nell'intorno della soglia possono essere ottenuti soltanto tramite estrapolazione delle due probabilità di occupazione al variare di S , *stimate* rispettivamente sui trattati e sui non-trattati nel modo rappresentato nel grafico 5b. Ad esempio, la valutazione dell'impatto del programma sui soggetti non-trattati richiede il calcolo di $Pr(occupato/S, D=1)$ in corrispondenza di valori di S inferiori a 56. Ma in corrispondenza di tali valori di S , per costruzione non si osservano soggetti trattati. Pertanto, si può al più *congetturare* che se la curva $Pr(occupato/S, D=1)$ mantenesse per valori di S inferiori a 56 le stesse caratteristiche osservate per valori di S superiori a 56, allora l'effetto sui non-

²² Vedi Trochim (1984) e Imbens e Lemieux (2008), e inoltre i testi generali citati nella nota 5.

trattati sarebbe pari alla differenza tra la curva stimata sui soggetti non-trattati e quella ottenuta *estrapolando* quella stimata sui soggetti trattati. Nel caso in questione, è



Graf. 5a: *Stima dell'effetto medio in corrispondenza della soglia per l'ammissione*



Graf. 5b: *Stima dell'impatto medio per soggetti con punteggio S=5, condizionatamente all'estrapolazione della curva $Pr(occupato|S, T)$*

Grafico 5: *Valutazione dell'effetto di un programma di orientamento al lavoro dei disoccupati, con ammissione in base alla graduatoria e con indagine di "follow*

up” condotta su tutti i partecipanti al test di ammissione. Variabile-risultato: probabilità di essere occupato a 6 mesi dalla conclusione del programma

Fonte: adattato da Rettore, Trivellato e Martini (2003).

interessante notare che la congettura basata sull’extrapolazione delle regolarità osservate rappresentata nel grafico 5b suggerisce che sono stati ammessi al programma i soggetti i quali ne ricavano un beneficio trascurabile, mentre sono stati esclusi proprio quelli che ne avrebbero ricavato un notevole beneficio: gli aspiranti meno abili, con punteggio al test intorno a 56.

3.3. *Distorsione da dinamica spontanea e confronto attorno alla soglia di discontinuità temporale*

Sinora ho considerato il problema della ricostruzione del controfattuale in presenza di distorsione da selezione, che si può manifestare, e in generale si manifesta – salvo il caso di esperimenti, quando si confrontano soggetti esposti e non esposti alla politica. Un analogo problema di potenziale distorsione si presenta quando si confrontano gli stessi soggetti *prima* che siano esposti alla politica e *dopo* l’esposizione.

Questo confronto si impone quanto più una politica ha caratteristiche universali, o comunque quando per la sua popolazione di destinatari ammissibilità e trattamento coincidano. È questo l’esempio degli incentivi alle imprese per stabilizzare i precari.

In questo contesto, sovente si utilizzano non tanto dati micro sui singoli beneficiari dell’intervento, quanto serie storiche di dati aggregati di grandezze-chiave: nell’esempio appena menzionato, le trasformazioni di contratti a tempo determinato in contratti a tempo indeterminato.

Supponiamo che, in una data area, il trattamento venga erogato in un dato periodo e fissiamo l’attenzione su due valori: l’uno per il periodo immediatamente precedente e l’altro per quello immediatamente successivo. Il primo si riferisce alla grandezza della variabile-risultato pre-trattamento – quindi in sua assenza, il secondo alla grandezza del fenomeno post-trattamento – quindi in sua presenza. È immediato riconoscere che la differenza osservata tra i due valori della variabile-risultato, riferiti rispettivamente al periodo post- e a quello pre-trattamento, è scomponibile in due grandezze distinte, anche se non osservabili: l’effetto causale del trattamento e “l’evoluzione naturale del fenomeno”. Quest’ultimo termine prende il nome di *distorsione da dinamica spontanea*: non perché ci sia alcunché di distorto nella dinamica spontanea, ma perché la sua presenza costringe a riconoscere che la differenza osservata pre-post in generale non coincide con l’effetto causale del trattamento sulla variabile-risultato.

Tale differenza identifica l’effetto solo nell’evenienza che in assenza dell’intervento il fenomeno non avrebbe manifestato alcuna dinamica. Un assunto, a ben vedere, eroico.

Questo assunto, in generale poco plausibile, può essere rilassato quando si disponga non solo del valore della grandezza nel periodo che precede immediatamente il trattamento, ma di una *serie storica* precedente l’intervento. In questo caso il controfattuale

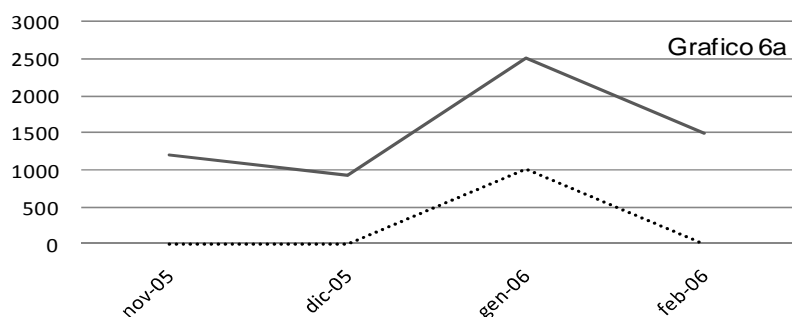
non è identificato dal valore della variabile-risultato nel periodo immediatamente precedente il trattamento, bensì dalla proiezione della storia pre-intervento del fenomeno, proiezione basata su un modello per la serie storica. Si può, cioè, utilizzare la storia pre-intervento della variabile-risultato per predire quale sarebbe stato il suo andamento nel periodo post-intervento *se l'intervento non fosse stato attuato*, e identificare l'effetto come differenza tra l'andamento osservato post-intervento (fattuale) e l'andamento predetto (controfattuale). Il requisito per l'utilizzo di questa strategia di identificazione è, dunque, la disponibilità della "storia" della variabile-risultato per un periodo precedente l'introduzione della politica, sufficientemente lungo da poterne modellare l'andamento temporale.

L'esempio degli incentivi per la stabilizzazione dei precari si presta per illustrare questa strategia. La variabile-risultato discende dal disegno stesso della politica, e consiste appunto nella trasformazione dei rapporti di lavoro precario in contratti a tempo indeterminato. Il confronto da fare è con un attendibile controfattuale: l'andamento predetto delle trasformazioni sulla base della serie storica pre-intervento. Assumiamo che nell'area in questione vi sia stato un intervento nel gennaio 2006, e che l'informazione sia quella contenuta nel grafico 6.

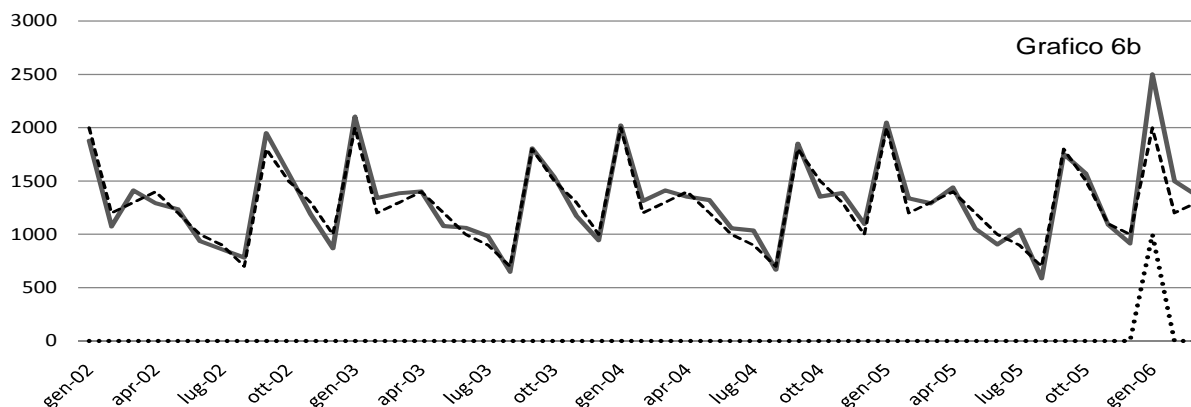
- (a) Il grafico 6a presenta la dinamica delle trasformazioni dei rapporti di lavoro precario in contratti a tempo indeterminato nella finestra temporale attorno al mese dell'intervento, da novembre 2005 a febbraio 2006. Il numero totale di trasformazioni è rappresentato con una spezzata continua, mentre il numero delle trasformazioni per le quali è stato erogato l'incentivo è rappresentato con una spezzata punteggiata. *Prima facie*, l'ispezione del grafico suggerisce un cospicuo effetto degli incentivi, dell'ordine di 1.000 contratti trasformati. Ma tale affermazione è palesemente esposta al rischio di fallacia di ogni confronto pre-post. Il fatto che il numero di trasformazioni aumenti in concomitanza con le erogazioni può essere erroneamente interpretato come evidenza di un effetto. Anzi, spesso è l'erogazione stessa degli incentivi ad essere interpretata come indicativa di un effetto.
- (b) Nel grafico 6b è rappresentata la serie storica delle trasformazioni dei contratti sufficientemente estesa all'indietro (con una spezzata continua), utilizzata per predire l'andamento delle trasformazioni contrattuali nella circoscrizione in questione nella finestra dell'intervento, cioè per ricostruire il controfattuale (rappresentato dalla spezzata tratteggiata). L'effetto a gennaio 2006, in concomitanza delle erogazioni, è molto ridimensionato e corrisponde alla deviazione tra quanto osservato e il controfattuale. L'effetto è stimato attorno ai 500 contratti trasformati.

La conclusione è presto tratta. La politica si è rivelata solo parzialmente efficace per la riduzione del precariato, perché le imprese hanno convertito in contratti di lavoro a tempo indeterminato anche una parte dei rapporti di lavoro precario che comunque intendevano trasformare (per i quali, cioè, si sono comportate allo stesso modo in cui si

sarebbero comportate se l'incentivo non fosse stato introdotto), semplicemente godendo degli incentivi²³.



Graf. 6a: Serie intorno a gennaio 2006.



Graf. 6b: Serie storiche lunghe

Grafico 6: Ipotetico esempio di un programma di incentivi per la stabilizzazione dei precari realizzato nel gennaio 2006

Legenda: spezzata continua: serie osservata; spezzata tratteggiata: serie prevista; spezzata punteggiata: trasformazioni con incentivi.

Non è poi da escludere che tale politica possa essere particolarmente insidiosa, nel merito e per la possibilità stessa di condurre una valutazione. Se già adottata in passato, e iterata con modeste varianti, essa potrebbe indurre alcune delle imprese ad assumere con

²³ Anche nell'esempio degli incentivi alle imprese per la stabilizzazione dei precari si può ricorrere all'identificazione dell'effetto "attorno alla soglia", tramite RDD, quando l'ammissibilità agli incentivi sia ristretta ai lavoratori precari al di sopra di una data età (come accennato nella tabella 1). Tale criterio, fissato amministrativamente in maniera ben definita e nota, nel breve periodo (quindi per la prima applicazione di un tale programma) è esogeno, perché le imprese non hanno modo di modificare le loro scelte di reclutamento in favore di lavoratori precari anziani. L'età che definisce la soglia di ammissibilità rappresenta, palesemente, il "punto di discontinuità".

rapporti di lavoro precari coloro che altrimenti avrebbero assunto subito con contratti a tempo indeterminato, per “catturare” poi gli incentivi trasformandoli in tali. Nel merito, nel breve periodo la politica produrrebbe, in parte, un risultato perverso, che rovescia, o attenua, l’effetto sperato: un maggior numero di lavoratori precari. Quanto alla sua valutazione, essa risulterebbe di molto complicata da questo fenomeno, denominato *entry effect*. I destinatari della politica non sono più un insieme “dato”; vi è una modificazione della platea degli ammissibili indotta dalla politica stessa. Gli adattamenti necessari nelle strategie di identificazione, sui quali per brevità non mi soffermo, sono sinteticamente richiamati nella tabella 1.

3.4. *Il doppio confronto spazio-temporale*

Quando la variabile-risultato che la politica si propone di influenzare è misurabile in momenti ripetuti nel tempo e si dispone inoltre di osservazioni anche su un insieme di non-trattati, si può applicare il metodo del doppio confronto: l’effetto della politica è rivelato confrontando la differenza tra prima e dopo l’attuazione della politica (primo confronto, *nel tempo*) tra trattati e non-trattati (secondo confronto, *tra soggetti*). Tipicamente, in tali contesti si hanno a disposizione dati longitudinali – cioè a dire osservazioni ripetute nel tempo su trattati e non-trattati per periodi *precedenti e seguenti* l’intervento, il che consente appunto l’applicazione di questo metodo – e di altri analoghi – per eliminare la distorsione da selezione (o per verificarne la presenza)²⁴.

L’esempio del programma di incentivi per la stabilizzazione dei precari, con una modesta variante, si presta bene per presentare, in maniera intuitiva, anche la logica di questo metodo di valutazione. Supponiamo che prossima all’area dove è stato realizzato il programma, chiamiamola A, vi sia un’area, chiamiamola B, caratterizzata da una struttura produttiva analoga, ma che *non* è stata interessata da alcuna politica di incentivi alle imprese per la stabilizzazione dei precari. Supponiamo poi che il *policy-maker* illuminato, affiancato da un analista-valutatore, si sia preoccupato che fossero raccolti i dati sui contratti a tempo determinato e la loro eventuale stabilizzazione in entrambe le aree per *almeno un periodo precedente* (meglio se più) e *per il periodo in cui è stato operante l’intervento* nell’area A. Avendo disponibili questi due insiemi di dati longitudinali, l’effetto medio dell’intervento del programma di incentivi per la stabilizzazione dei precari può essere ottenuto sottraendo al cambiamento osservato nella frazione delle trasformazioni dei contratti temporanei in contratti a tempo indeterminato nell’area A l’analogo cambiamento osservato nell’area B.

Il grafico 7 riporta sull’asse verticale la frazione di contratti temporanei stabilizzati, e su quello orizzontale i due periodi di osservazione, rispettivamente quello immediatamente precedente la realizzazione del programma, il dicembre 2005, e quello in cui lo stesso è stato operante, il gennaio 2006. L’andamento nel tempo è mostrato separatamente per l’area A, interessata dal programma – dunque quella dei trattati – e per l’area B, non interessata dal programma – dunque quella dei non-trattati. Come già anticipato, ipotizziamo inoltre che le due aree abbiano una struttura produttiva analoga – in

termini di distribuzione settoriale delle attività economiche, di struttura per età degli occupati, di prodotto per addetto, ecc., quindi relativamente stabile nel tempo.

²⁴ In generale tali metodi sono preferibili proprio perché la più ricca informazione empirica a disposizione – estesa alla variabilità sezionale fra individui e a quella intra-individuale nel tempo – permette di allentare assunti adottati nella specificazione del processo di selezione sulla base rispettivamente di dati *cross-section* e di serie storiche (o di trattarli come ipotesi passibili di verifica). *Mutatis mutandis*, tuttavia, la logica delineata nella sez. 3.1 per dati di tipo sezionale e nella sez. 3.3 per serie storiche aggregate, resta valida.

Il grafico 7 presenta le differenze fra gennaio 2006 e dicembre 2005 nella frazione di contratti temporanei trasformati per entrambi i gruppi di precari : la differenza osservata nell'area A mostra chiaramente una crescita più accentuata (oltre che avere livelli di partenza sensibilmente più bassi). Se si sottrae alla differenza osservata tra i precari dell'area A, trattati, quella osservata tra i precari nell'area B, non-trattati, si ottiene una stima dell'effetto della politica di incentivi per la stabilizzazione dei precari.

In generale, un tale metodo di stima dell'effetto di una politica, noto come “differenza nelle differenze”, è valido a condizione che valga il seguente assunto: l'evoluzione della variabile-risultato che avrebbero avuto i trattati se non lo fossero stati è la stessa che hanno avuto i non-trattati. Detto altrimenti, possono sì esserci differenze sistematiche tra trattati e non-trattati in caratteristiche – anche non osservabili – che determinano il livello della variabile-risultato prima dell'intervento, ma queste differenze in assenza dell'intervento si sarebbero mantenute costanti nel tempo.

L'assunto cruciale consiste quindi nell'escludere che ci siano differenze sistematiche tra i due gruppi, trattati e non-trattati, nel trend della variabile-risultato in assenza della politica. Secondo questo assunto la dinamica osservata per i non-trattati riproduce perfettamente quella che sarebbe stata la dinamica per i trattati in assenza di trattamento. Quindi la differenza fra le differenze prima-dopo rispettivamente di trattati e non-trattati identifica l'effetto causale della politica.

Tornando all'esemplificazione del grafico 7, questo assunto implica che il segmento tratteggiato, relativo alla differenza controfattuale prima-dopo della frazione di

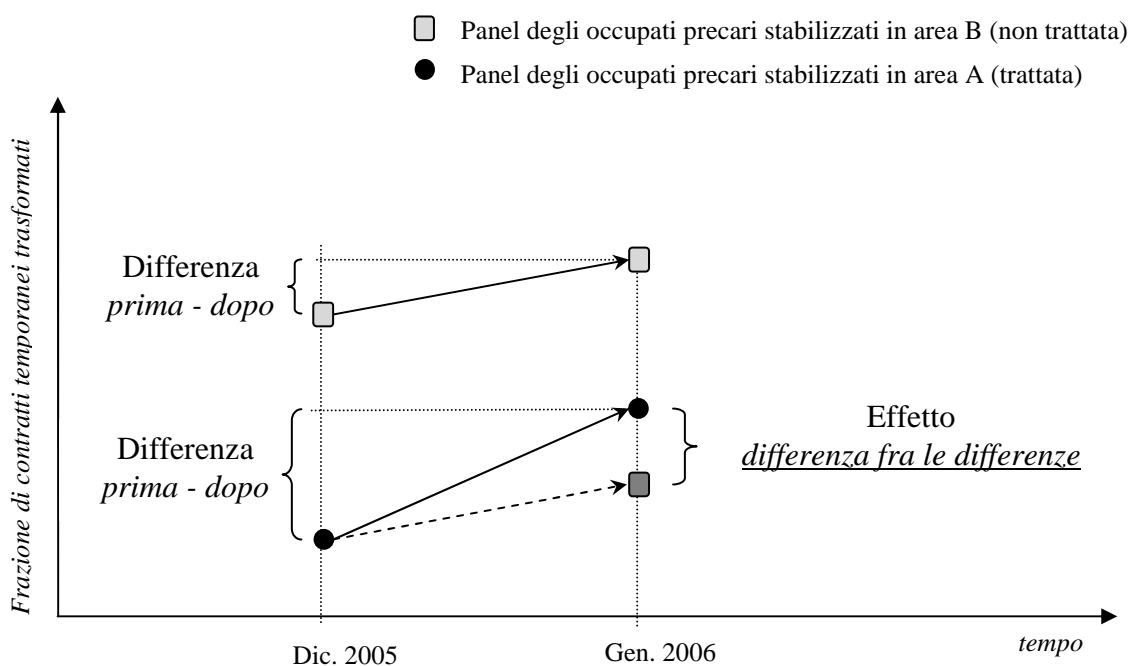


Grafico 7: Il doppio confronto spazio-temporale applicato all'esempio del programma di incentivi per la stabilizzazione dei precari nell'area A nel gennaio 2006

precari stabilizzati dell'area A, trattati, sia parallelo al segmento osservato della differenza fattuale prima-dopo dell'analoga frazione nell'area B, non-trattati²⁵.

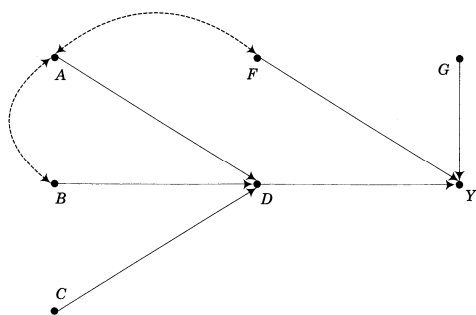
4. Uno scorcio a buone pratiche di altri Paesi

La valutazione di politiche pubbliche è pratica corrente in molti Paesi sviluppati. Gli Stati Uniti offrono il panorama più maturo, e a più larga scala, con le prime esperienze di valutazione di programmi di *welfare* e di politiche attive del lavoro che risalgono alla seconda metà degli anni 1960. Ma pratiche consolidate, sistematiche di valutazione degli effetti di politiche pubbliche si riscontrano ormai nei Paesi di cultura anglosassone e scandinava, in molti altri Paesi europei – prima fra tutte la Germania, in alcuni Paesi in via di sviluppo – tra questi spicca il Messico – anche per l'azione di organismi internazionali come la Banca Mondiale.

Ogni pretesa di rassegna, anche per un solo comparto di politiche pubbliche, sarebbe fuori luogo. Mi fermerò, brevemente, su quattro casi, che si prestano per illustrare uno spettro diversificato di *buone pratiche* di disegno e realizzazione di politiche pubbliche, accomunate dal fatto di includere come tassello essenziale la valutazione dei loro effetti. I quattro casi sono: l'attività della *What Works Clearing House* statunitense; l'insieme delle "riforme Harz" tedesche; una recente innovazione nelle politiche di avviamento al lavoro in Francia; il disegno di nuove politiche di *welfare* del governo inglese.

C'è un'ovvia discrezionalità nella scelta. Ma ad essa si accompagna una motivazione. I casi spaziano da meta-analisi su un ampio insieme di studi di valutazione

²⁵ Vi è un'ulteriore, importante strategia di identificazione degli effetti, che per esigenze di concisione e semplicità richiamo soltanto qui in nota. Essa assume che vi sia una variabile C la quale causa D , ma non ha alcun effetto su Y se non per il tramite del suo effetto su D (detto altrimenti, assume valori in modo indipendente dal processo che, tramite $\{A, B, F, G\}$, genera D e Y). Una tale situazione è esemplificata nel grafico che segue:



È intuitivamente chiaro che l'assetto che ne discende è prossimo a quello in cui D è assegnato casualmente, cioè a un esperimento. In sostanza, si è in presenza di uno shock esogeno sulla variabile trattamento e, per il solo tramite di questa, sulla variabile-risultato. Nel linguaggio statistico-econometrico, C è un valido strumento per D , e l'effetto di D su Y si può stimare adeguatamente tramite il rapporto delle relazioni fra C e Y e di quelle fra C e D (vedi, ad esempio, Imbens e Wooldridge, 2008). Si noti che, se il solo interesse dell'analista è l'effetto causale di D su Y , per questa strategia di identificazione non è necessario osservare $\{A, B, F, G\}$.

degli effetti di interventi in campo educativo; al sistematico impegno di valutazione degli effetti di un vasto piano di riforme della mercato del lavoro e della protezione sociale; alla drastica innovazione di una politica del lavoro, proprio nella prospettiva di valutarne gli effetti e di “imparare” da tale esperienza, in un Paese tradizionalmente freddo rispetto a queste tematiche; al disegno e alla proposta di nuove politiche di *welfare* in un Paese che viene da pratiche diffuse di realizzazione e valutazione degli effetti di politiche pubbliche. Offrono quindi un spettro sì selettivo, ma ampio, di contesti e di modalità di intervento.

4.1. *La What Works Clearing House*

Istituita nel 2002 per iniziativa dell'Institute of Education Sciences del Department of Education statunitense, la *What Works Clearinghouse* (WWC) è una fonte basilare e affidabile di evidenza scientifica sulla valutazione degli effetti di politiche educative, su “ciò che funziona” di tali politiche. Tra le sue attività primarie, infatti, la WWC (i) «accerta il rigore dell'evidenza scientifica sull'effettività di interventi (programmi, prodotti, pratiche e politiche), fornendo agli educatori gli strumenti per prendere decisioni informate», e a tal fine (ii) «sviluppa e realizza standards per rivedere e sintetizzare la ricerca educativa»²⁶.

È quest'ultimo l'aspetto sul quale vorrei soffermarmi, anche perché in chiave comparata con altre iniziative segnala un importante avanzamento. La pre-condizione è

l'esistenza di un largo insieme di studi di valutazione di singole politiche educative, ad ampia e a piccola scala, condotti con metodi osservazionali e sovente sperimentali. Come utilizzare questa evidenza, vasta, diversificata, talvolta contraddittoria, e sintetizzarla in maniera pregnante? Le abituali meta-analisi sono condotte in chiave narrativa, al più affiancando e commentando risultati quantitativi. Nell'ambito della WWC si utilizzano e si sviluppano, invece, procedure per strutturare rigorosamente il processo di comparazione e sintesi.

Si selezionano le ricerche che obbediscono ad adeguati standard scientifici; si guarda soltanto agli studi la cui dimensione campionaria consente di intercettare effetti anche piccoli, se esistono; la meta-analisi viene condotta su *clusters* omogenei di studi di politiche; la sintesi dei risultati avviene con tecniche, soprattutto grafiche, altamente informative. Insomma, il processo di apprendimento dalle molteplici esperienze di valutazione di effetti di interventi è strutturato secondo criteri chiari, convincenti²⁷. L'evidenza che emerge sull'efficacia, o meno, delle diverse tipologie di interventi acquisisce autorevolezza, e diventa un riferimento persuasivo per decidere quali mettere in atto su più larga scala.

4.2. Le "Riforme Harz" tedesche

Il caso tedesco è istruttivo innanzitutto per l'alto tasso di innovazione che ha introdotto nel processo di disegno, attuazione e valutazione degli effetti di politiche del lavoro. Non è azzardato dire che una decina di anni fa in Germania lo stato dell'arte su questi temi non era molto migliore che in Italia. Anche per l'acuirsi dei problemi di disoccupazione e di sostenibilità del *welfare* indotto dall'unificazione, alla fine degli anni 1990 la Germania si è trovata a misurarsi con esigenze di profonda riforma del sistema di protezione sociale e delle politiche del lavoro. E con un quadro complicato da forti disparità territoriali e da un assetto federale²⁸.

Ha affrontato la questione con una serie di riforme, note come "riforme Hartz I-IV" (dal nome del presidente dell'apposita Commissione istituita dal Governo federale), realizzate in quattro fasi tra il 2002 e il 2005.

Per tutti i principali programmi è stata prevista una valutazione sistematica dell'implementazione e degli effetti. Ciò ha richiesto un notevole miglioramento del sistema informativo, basato su archivi amministrativi, e una politica ragionevolmente liberale di accesso ai microdati da parte dei ricercatori. Inoltre, ed è questo un aspetto

²⁶ L'enfasi del sottolineato è aggiunta. Vedi <http://ies.ed.gov/ncee/wwc/>, anche per la documentazione sugli argomenti di prevalente interesse sostanziale della WWC.

²⁷ Vedi, ad esempio, Cottingham, Maynard e Stagner (2004, 2005).

²⁸ Su un diverso piano, attinente alle possibilità di ricerca empirica, pesavano poi severi vincoli all'utilizzo di basi di microdati per esigenze di tutela della *privacy*.

fondamentale, ha indotto una diffusa attività di valutazione da parte di istituti di ricerca e di gruppi di ricercatori, con un deciso aumento della qualità delle valutazioni. L'entità di questi progressi è emblematicamente documentata dal fatto che cominciano ad apparire studi sull'efficacia comparata di diversi programmi per diversi sottoinsiemi di destinatari²⁹. Infine, vi è stata una crescente attenzione ai risultati delle valutazioni nei mezzi di informazione e nel dibattito pubblico: tassello, questo, decisivo perché buone pratiche di valutazione degli effetti si affermino.

4.3. *Una recente innovazione nelle politiche di avviamento al lavoro in Francia*

Vengo ora a una recente esperienza avviata in Francia. Per comprenderne la portata, debbo prima richiamare quanto emerso nell'“*International Workshop on impact evaluation practice and prospects*”, organizzato dalla Commissione del CSS su “La valutazione degli effetti di politiche pubbliche” un anno e mezzo fa, con la partecipazione di sette esperti internazionali di alto livello, fra i quali un collega francese. Dai suoi argomentati interventi, e dalle discussioni, emerse con tutta evidenza che in Francia vi era un mediocre interesse dei *policy makers* e delle amministrazioni pubbliche per studi di valutazione degli effetti di politiche pubbliche, e di conseguenza uno loro scarso sviluppo. L'eccezione erano studi, di notevole livello ma di taglio prettamente accademico, condotti da ricercatori dell'INSEE (l'istituto statistico francese) e dal CREST-INSEE (il centro di ricerca dello stesso), anche perché soltanto questi ricercatori hanno accesso ai dati appropriati, prevalentemente di origine amministrativa³⁰.

Ebbene, in questo lasso di tempo vi sono state modifiche d'indole generale, che hanno portato da un lato a una maggiore interazione fra Ministero del Lavoro e Ministero dell'Economia e dall'altro al decollo della Paris School of Economics, con un fuoco (anche) sulla politiche pubbliche e lo sviluppo e un'importante partenariato con J-PAL (A.L. Jameel-Poverty Action Laboratory, MIT), del quale ospita la sezione J-PAL Europe.

In questo quadro, nel 2007 ha preso avvio un esperimento sociale su larga scala, che coinvolge più di 200.000 disoccupati di lunga durata e si propone di valutare gli effetti di tre diverse modalità di avviamento al lavoro. Il primo rapporto intermedio, che dà conto delle transizioni al lavoro a 6 e 9 mesi dall'avvio del programma, lo presenta con esemplare chiarezza: «*Questa valutazione degli effetti è stata realizzata utilizzando un*

²⁹ Vedi Jacobi e Kluve (2006), Biewen *et al.* (2007), Wunch e Lechner (2007), Caliendo (2008), Stephan e Pahnke (2008), Huber *et al.* (2009) e vari altri studi apparsi prevalentemente nella serie di Discussion Papers dell'IZA, sezione Evaluation of Labor Market Programs (<http://www.iza.org/>).

³⁰ Citando liberamente dalla minuta dello *Workshop*, emergono riscontri e giudizi di questo tipo: “I promotori di programmi (politici e alti burocrati) sono in generale riluttanti a richiedere valutazioni scientifiche degli effetti. [...] In Francia, la valutazione degli effetti non è mai stata usata come un pre-test dell'efficacia di una politica [...]. Ciò ha a che vedere anche con la generale avversione all'uso di esperimenti randomizzati per valutare interventi pubblici. [...] La consapevolezza [delle difficoltà concettuali e pratiche nel produrre inferenza causale in merito a politiche pubbliche] è decisamente insufficiente. L'amministrazione francese e l'élite politica hanno una cultura economica e metodologia molto debole. La cultura dominante viene dalla scienza politica non quantitativa. [...] Tuttavia, la percezione dell'utilità della valutazione degli effetti tra i *policy-makers*, i media e l'opinione pubblica] è progressivamente cambiata negli ultimi anni. L'atteggiamento generale è via via più favorevole”.

metodo sinora inedito in Francia. Questo metodo consiste nel distribuire a caso i potenziali beneficiari ammissibili in diversi gruppi: uno [circa 120.000 soggetti] si vede proporre l'accompagnamento classico dell' ANPE [Agence Nationale Pour l'Emploi]; l'altro [circa 40.000 soggetti] l'accompagnamento [individuale intensivo] del programma "Cap Vers l'Entreprise"; il terzo [ancora circa 40.000 soggetti] l'accompagnamento [rinforzato] assicurato da gruppi di operatori privati di collocamento» (Behaghel et al. 2008: ii-iii).

Lo scarto fra la diagnosi di appena un anno e mezzo fa e le realizzazioni maturate nel frattempo è notevole, per molti versi sorprendente. Esso segnala che un Paese – la Francia – al quale siamo prossimi sotto molti punti di vista, non ultima la tradizione politico-amministrativa, si è avviato, almeno in parte, su una strada nella quale pratiche rigorose della valutazione degli effetti di interventi prendono piede, anche nella versione di *social experiments* su larga scala³¹.

4.4. La proposta di nuove politiche di welfare del governo inglese

Conclusa la decennale esperienza del *New Deal*, il programma-quadro di politiche attive del lavoro introdotto nel Regno Unito dal Governo Blair nel 1998, il Governo Brown ha recentemente presentato, per una consultazione pubblica, la proposta di un nuovo programma di riforma del *welfare* sul lavoro, nel *Green Paper* "No one written off: Reforming welfare to reward responsibility" (Department for Work and Pensions 2008b).

Non siamo quindi di fronte a una politica pubblica *realizzata*, della quale valutare *ex-post* gli effetti. Ci misuriamo, invece, col *disegno di un nuovo, ambizioso programma*. Ma anche il modo con cui il programma è stato elaborato e viene proposto, e se ne prospetta la messa in atto e la valutazione è significativo: emblematico di una maniera di affrontare il tema *evidence-based*, aperta a una non rituale consultazione pubblica, lungimirante nel predisporre risorse e strumenti per una sistematica attività di monitoraggio dell'implementazione e di valutazione degli effetti³². L'idea-guida del progetto è riassunta in una parola: "condizionalità". Essa si richiama a un principio già enunciato nel Rapporto Beveridge del 1942: «Lo Stato dovrà offrire sicurezza per servizi e contributi monetari. Nell'organizzare la sicurezza, lo Stato non dovrà mortificare incentivi, opportunità, responsabilità»³³. È la prospettiva delle politiche attive del lavoro che nell'ultimo decennio si è venuta affermando in termini di "welfare to work", "mutual

³¹ La scala dell'intervento, circa 80mila disoccupati trattati (mentre non hanno rilievo i 120mila non-trattati, esposti cioè alle pratiche usuali), solleva qualche interrogativo su possibili effetti di equilibrio economico generale, che ne renderebbero problematica la valutazione degli effetti entro la logica di un *social experiment*. A ben vedere, i trattati sono poco più del 3,6% dei disoccupati, il che dovrebbe rendere questo rischio trascurabile. Il *caveat* resta peraltro giustificato per il fatto che l'esperimento è concentrato sui disoccupati di lunga durata in alcune regioni.

³² È banale aggiungere che tutto ciò non vale per assicurare il successo, cioè significativi effetti positivi, del programma. Aiuta molto, però. E in ogni caso garantisce che si potrà imparare, e in maniera scientificamente fondata, quindi condivisa.

³³ L'enfasi della sottolineatura è aggiunta.

obligations” e, nella forma più matura, “*flexicurity*” (vedi, ad esempio, Martin e Grubb 2001, OECD 2004: cap. 2, e Andersen e Svarer 2007)³⁴.

Nella prospettiva che qui interessa, ciò che va segnalato dell’esperienza inglese è in almeno tre tratti.

- (a) Il *Green Paper* è stato preceduto da un rapporto snello, presentato appunto come *Background research and discussion paper*. Innanzitutto, esso fa tesoro dell’esperienza del *New Deal*. È questo un aspetto che mi preme sottolineare. Il *Background research and discussion paper* muove dai risultati di preordinate valutazioni degli effetti del precedente programma³⁵, dalle revisioni e dai miglioramenti che esso ha conosciuto nel decennio, innescati dagli esiti di tali valutazioni, in un processo di continuo ridisegno degli interventi e di altrettanto sistematica attività di loro monitoraggio e valutazione. Da questa riflessione esso trae le linee-guida della nuova proposta, la sua filosofia, che enuncia nello stesso titolo: “*More support, higher expectations: The role of conditionality in improving employment outcomes*” (Department for Work and Pensions 2008a).
- (b) La consultazione sul *Green Paper* avviene in maniera strutturata, sulla base di una lista di domande che orienta e focalizza la discussione fra attori sociali, media e cittadini (Department for Work and Pensions 2008c).
- (c) Infine, tassello ultimo ma non certo meno importante, il *Green Paper* è accompagnato dall’*Impact Assessment* (Department for Work and Pensions 2008d): una rigorosa Analisi di Impatto della Regolamentazione (AIR, *ex-ante*)³⁶, che analizza i prevedibili costi, benefici ed effetti attesi delle principali opzioni presentate.

³⁴ Essa è enunciata in maniera chiara da James Purnell, Segretario di Stato per il Lavoro e le Pensioni del Regno Unito: «[Il programma di riforme] propone un sistema di benefici che premia la responsabilità, dà alle persone l’incentivo a fare la cosa giusta, e le incoraggia a cercare lavoro e ad acquisire le abilità di cui avranno bisogno nel futuro. Ma nello stesso tempo esso assicura che tale opportunità sia disponibile per ciascuno: cerca di porre fine alla presente ingiustizia, per cui troppe persone sono confinate a una vita di sussidi ed escluse da un aiuto per tornare al lavoro. In breve, è basato su un semplice scambio: più sostegno, più responsabilità» (Department for Work and Pensions 2008b: 7).

³⁵ Sulla valutazione degli effetti del *New Deal* vedi, tra i molti, Blundell *et al.* (2004), Van Renen (2004, 2005) e numerosi riferimenti bibliografici in Department for Work and Pensions 2008a).

³⁶ L’AIR, suggerita inizialmente dall’OECD e quindi dall’Unione Europea, è stata introdotta in Italia con la legge n. 50/1999 e successive disposizioni. L’obbligo di accompagnare una nuova regolazione, o riforma, con l’AIR è stato peraltro largamente disatteso. L’AIR è stata prodotta occasionalmente e spesso in maniera superficiale. Andranno vagliate attentamente, nei fatti, le potenziali novità introdotte recentemente dal DPCM n. 170/2008 sulla “disciplina attuativa dell’analisi dell’impatto della regolamentazione (AIR)” (Gazzetta Ufficiale n. 257 del 3 novembre 2008), e dalla Direttiva del Presidente del Consiglio dei Ministri del 10 settembre 2008 su “tempi e modalità di effettuazione dell’analisi tecnico-normativa (ATN)”, con allegata griglia metodologica per la stesura della pertinente relazione (Gazzetta Ufficiale n. 219 del 18 settembre 2008).

V’è da aggiungere che la legge n.246/2005, comma 4, prevede anche la «*verifica dell’impatto delle regolamentazione (VIR), [...] definita come] stima, anche periodica, [...] degli effetti prodotti da atti normativi*». Per quanto mi è noto, per la VIR l’analogia appropriata è alle grida di manzoniana memoria: semplicemente, non è mai stata applicata.

4.5. *Che cosa suggeriscono, nel metodo e nel merito, le esperienze più mature di valutazione degli effetti di politiche?*

4.5.1. *Sul disegno delle politiche e della valutazione dei loro effetti*

Tirando le fila dai casi considerati, e dalla letteratura, che suggerimenti possiamo trarre sugli aspetti di metodo: sul modo di disegnare le politiche pubbliche e di valutarne gli effetti?

Guardo in primo luogo, brevemente, agli aspetti istituzionali. È ormai diventata pratica abbastanza diffusa includere nelle leggi istitutive di politiche pubbliche le cosiddette “clausole valutative”, cioè a dire i punti chiave per lo svolgimento della valutazione: (i) gli obiettivi con riferimento ai quali avrà luogo; (ii) indicazioni circa il modo in cui dovrà essere svolta; (iii) lo stanziamento per lo svolgimento della valutazione³⁷.

Ma queste prescrizioni rischiano di restare inefficaci, o di dar luogo a meri adempimenti formali, se le pratiche di valutazione degli effetti delle politiche non poggiano su condizioni, culturali e istituzionali, perché esse possano essere introdotte e affermarsi. Il caso statunitense e tedesco (attuati) e quello inglese (progettato sulla scorta di un matura esperienza precedente) suggeriscono che tale sostrato può essere ricondotto a tre fattori basilari.

- (a) Il processo di (ri)definizione delle politiche è *evidence-based*. L’esame delle diverse possibili configurazioni di una politica pubblica non prescinde certo da opzioni di valore, né dagli interessi in gioco. Ma questo confronto di preferenze e di interessi avviene muovendo da, e restando ancorati a, un solido terreno comune: l’evidenza dei fatti, documentata in maniera adeguata, condivisa da tutti gli interlocutori.
- (b) Sul piano istituzionale, un contesto favorevole è offerto dal cosiddetto “*modello anglo-americano di ‘accountability’ democratica*”, in cui la funzione di controllo dell’esecutivo e di valutazione degli effetti di politiche pubbliche è svolta da agenzie indipendenti, “*watching dogs*” – così, letteralmente, è definito il loro ruolo in alcuni atti istitutivi – con elevata specializzazione, notevoli professionalità e reputazione di autonomia, che operano per conto della/e assemblea/e legislativa/e.

Decisamente meno efficace si mostra il cosiddetto “*modello continentale*”, che affida la funzione di controllo a magistrature indipendenti – le Corti dei Conti, equidistanti dai poteri esecutivo e legislativo: in generale con il controllo incentrato sulla conformità alla legge e al bilancio di previsione degli atti che danno luogo a entrate e spese pubbliche e sulla certificazione dei bilanci consuntivi; non sorprendentemente, con mediocre sensibilità e competenze inadeguate per la valutazione degli effetti di politiche³⁸.

³⁷ Raguagli su un paio di esperienze statunitensi di disegno e valutazione di politiche attive del lavoro, che risalgono a oltre 15 anni fa e documentano in maniera nitida questi tratti distintivi, sono in Rettore e Trivellato (1999).

³⁸ Vedi, tra i molti, Regonini (2009). Qui sfioro appena una tematica, quella degli assetti istituzionali e della loro rilevanza per l’*accountability* democratica, che è al cuore della riflessione di filosofi della politica e

(c) In buona parte in conseguenza dei due fattori appena segnalati, il processo di (ri)definizione delle politiche avviene nel quadro di un dibattito pubblico sufficientemente approfondito e focalizzato sulle “*policy issues*”, non prigioniero di un confronto deformato da schermi ideologici o comunque falsato da rappresentazioni propagandistiche. Alla qualità del dibattito pubblico concorre, in particolare, la capacità di università, *think tanks* e media di alimentarlo in maniera semplice ma ben documentata.

Poggiando su questo terreno, le buone pratiche si distinguono per molteplici aspetti, in parte propri dei diversi contesti nazionali, in parte generali. Di questi ultimi, cinque meritano di essere segnalati.

I. La domanda “quando occorre iniziare la valutazione degli effetti di una politica?” ha una risposta semplice e univoca: prima possibile. Meglio se prima ancora di cominciare ad attuare la politica. La contrapposizione tra valutazione *ex-ante* ed *ex-post* rischia di essere futile, se non addirittura ingannevole. Tutta la valutazione degli effetti è, infatti, banalmente *ex-post*, nel senso che giudica la presenza di effetti di interventi che *sono stati* realizzati. Al tempo stesso, però, è fondamentale che la valutazione degli effetti sia disegnata *ex-ante*, sin dall’avvio della riflessione sull’introduzione – o la riforma – della politica, con *interazione* fra il decisore e il valutatore. E ciò per due motivi, l’uno di rilievo generale, attinente al processo di “disegno della politica - sua messa in atto in maniera coerente - valutazione dei suoi effetti - ridisegno della politica nel quadro di una strategia del cambiamento”, l’altro funzionale allo studio di valutazione.

- Tra i vantaggi dell’adozione consapevole di una tale prospettiva di valutazione degli effetti vi è il fatto che induce le parti in causa (*policy makers*, interessi organizzati, opinione pubblica) ad uno sforzo di concretezza nella discussione pubblica. Stimola a chiarire cosa ci si aspetta dalla politica e quindi a stabilire *a priori*, in modo circostanziato, a quali grandezze si guarderà per stabilire se la politica è riuscita. Nel merito, ciò può contribuire ad affinare di molto il suo disegno.
- Questa *contestualità* è funzionale allo studio di valutazione, perché porta a individuare le informazioni necessarie per ricostruire la situazione controfattuale,

politologi. Mi limito a due notazioni rilevanti per la valutazione degli effetti di politiche pubbliche. (i) Negli USA, dagli anni 1990 il modello federale, che fa perno sul *Gouvernement Accountability Office*, è stato adottato a livello di molti Stati, ad oggi 41. Una recente rassegna delle funzioni svolte dai *Legislative Program Evaluation Offices* dei 41 Stati registra che la grandissima maggioranza (il 90%) svolge attività di “*perform audits/program evaluation/policy analyses*”, mentre meno della metà (il 46%) svolge “*financial/compliance audits*” (National Conference of State Legislatures, 2008). (ii) In diversi Paesi con assetto istituzionale ispirato al “modello continentale” è in corso un processo di revisione, che tende a potenziare le funzioni di controllo del legislativo, supportato da adeguati organi tecnici, sull’esecutivo. L’esempio più significativo è quello francese. La legge di revisione costituzionale n. 2008/724 “*de modernisation des institutions de la V^e République*” (JO n. 171 del 24 luglio 2008), che ha come finalità principale riequilibrare le istituzioni a seguito al rafforzamento del “presidenzialismo”, accresce il ruolo di controllo del Parlamento e all’art. 24 stabilisce che «*Le Parlement [...] contrôle l’action du Gouvernement. Il évalue les politiques publiques*» (l’enfasi del sottolineato è aggiunta).

e quindi a predisporre tempestivamente la loro rilevazione. In tal modo, effettuato l'intervento (o un ciclo dello stesso, se si tratta di una politica a ciclo ripetuto) si disporrà dei dati in tempi ragionevolmente brevi; e di dati tali da consentire una valutazione della politica "robusta". In definitiva, si potranno avere valutazioni degli effetti della politica credibili e ragionevolmente sollecite.

La distinzione futile tra valutazione *ex-ante* ed *ex-post* lascia quindi il posto alla distinzione tra valutazione condotta retrospettivamente, dopo che la politica è conclusa, e valutazione condotta prospetticamente, in parallelo all'attuazione della politica stessa³⁹.

- II. A fronte della complessità delle politiche pubbliche e della loro messa in atto in maniera decentrata, spesso diversificata, le buone pratiche si caratterizzano per la combinazione di *analisi di implementazione* – cioè monitoraggio della messa in atto della politica – e di *valutazione degli effetti*, condotte in modo rigoroso e coordinato, ma senza confusioni di ruoli. Tra l'altro, l'analisi di implementazione serve a chiarire quali siano i contenuti effettivi della politica, che è poi oggetto della valutazione.
- III. Per svolgere in maniera adeguata l'analisi di implementazione e l'analisi degli effetti di una politica è essenziale la disponibilità di un'adeguata *base informativa*. Buoni sistemi informativi, alimentati soprattutto da archivi amministrativi, sono un supporto fondamentale, anche se spesso deve essere integrato da tempestive indagini *ad hoc*.
- IV. Prima di introdurre una politica a regime, rivolta cioè alla generalità dei destinatari, sovente è opportuno mettere in atto un intervento su scala ridotta sotto forma di *progetto-pilota*, con lo specifico scopo di "imparare" sulle modalità di funzionamento dell'intervento e sui suoi effetti. Proprio perché condotto su piccola scala e svolto con finalità di apprendimento, il progetto-pilota può essere realizzato come *social experiment*. Il disegno sperimentale può riguardare non i destinatari dell'intervento, persone o famiglie o imprese, ma aree geografiche o ripartizioni amministrative (ad esempio, province, comuni, Centri per l'impiego).
- V. Spostando l'attenzione sugli esiti della valutazione degli effetti, è bene sgombrare il campo da attese miracolistiche: che la valutazione consenta di stabilire in modo univoco e in ogni circostanza se la politica abbia o meno prodotto gli effetti desiderati. Così non è. All'origine del problema vi sono le stesse incertezze che rendono necessaria la valutazione degli effetti di una politica, sulle quali mi sono soffermato nella sezione 2. Come ben illustrato dall'esperienza della *What Works Clearing House*, la valutazione di una specifica politica, più che un'operazione circoscritta svolta da un singolo analista eventualmente incaricato ufficialmente della cosa, è un

³⁹ Vi è un'ulteriore ragione per rifiutare la contrapposizione tra valutazione *ex-ante* ed *ex-post*. Dal punto di vista dell'utilizzo, tutta la valutazione degli effetti è infatti *ex-ante*, nel senso che trova la sua ragion d'essere nell'informare il disegno o il ridisegno di politiche che devono essere attuate. Nulla è più inutile di una valutazione degli effetti di una politica che non è più in vigore e che non si ha intenzione di continuare (se non per quel che potrebbe insegnare, per analogia, per altre politiche che ne condividano alcune rilevanti caratteristiche). La valutazione degli effetti è, infatti, una forma di apprendimento, non di rendicontazione.

processo che si dispiega nel tempo, animato dal confronto tra più analisti indipendenti l'un l'altro, sovente arricchito dall'esame di studi di valutazione di interventi analoghi condotti con disegni o in ambiti diversi. Come è stato detto con efficace sintesi, «*le valutazioni [degli effetti] si basano su una conoscenza cumulativa*» (Heckman e Smith 1995: 93). Tra le altre cose, ciò ha un'importante implicazione per le politiche che regolano (o meglio, dovrebbero regolare) il mercato della valutazione: *l'accessibilità delle informazioni necessarie alla valutazione a una pluralità di analisti*, in condizione di parità con l'eventuale valutatore ufficiale; condizione, questa, imprescindibile per la necessaria *replicabilità degli studi*.

4.5.2. *Sul merito degli effetti di politiche attive del lavoro*

Venendo ai risultati di studi di valutazione degli effetti di politiche pubbliche, mi soffermo, con uno sguardo a maglia larga, su un ambito che mi è familiare: quello delle politiche del lavoro. Rassegne estese a più Paesi sono in Heckman, LaLonde e Smith (1999: 2043-2080), Martin e Grubb (2001), OECD (2005: cap. 4), Kluve (2007) e soprattutto Card, Kluve e Weber (2009). Apprezzabili insiemi di studi, che già ho richiamato, si hanno poi per il *New Deal* inglese e per le “riforme Hartz” tedesche, e inoltre per politiche attive del lavoro nei paesi scandinavi e in Olanda.

Evito ogni tentazione, presuntuosa più che eroica, di ricapitolazioni attente alle specificità delle singole politiche e dei contesti nazionali. E, fidando soprattutto sui quadri comparativi di Heckman, LaLonde e Smith (1999) e di Card, Kluve e Weber (2009), azzardo sintesi di larga massima.

Per i programmi governativi di orientamento e di formazione al lavoro, Heckman, LaLonde e Smith (1999: 1868) registrano un'evidenza largamente prevalente: «*Come attualmente formati, questi programmi sono spesso inefficaci [...]. Per molti gruppi di partecipanti, i benefici sono modesti [...]. Per di più, molti programmi e iniziative non sono in grado di passare una verifica costi-benefici. [...]. Ma, allo stesso tempo, vi è una sostanziale eterogeneità negli effetti di questi programmi. Per alcuni gruppi essi appaiono in grado di generare benefici significativi, sia per i partecipanti che per la società*».

Le ragioni per questo insoddisfacente stato delle cose sono, verosimilmente, quattro. Le prime due sono prospettate ancora da Heckman, LaLonde e Smith (1999). Essi sostengono, in maniera convincente, che le spese pro-capite per i partecipanti sono usualmente piccole rispetto ai deficit di abilità e competenze che ai programmi viene chiesto di affrontare. Argomentano, inoltre, che una seconda ragione per cui i guadagni da tali programmi sono generalmente bassi sta nel fatto che i servizi offerti sono mirati ai soggetti non specializzati e meno capaci⁴⁰.

A mio avviso, altre due ragioni di indole generale sono riconducibili all'eterogeneità degli effetti da un lato (che pure Heckman e coautori menzionano tra le evidenze salienti), e al disegno delle politiche del lavoro dall'altro. Il fatto che un

⁴⁰ Questa argomentazione si collega all'importanza che, nelle sue ricerche sulla formazione di abilità nel ciclo di vita, Heckman assegna all'istruzione nella prima infanzia. Vedi Cuhna *et al.* (2006) e Heckman, Stixrud e Urzua (2006).

programma produca effetti anche marcatamente diversi su differenti sottoinsiemi di partecipanti segnala che un adeguato *targeting* di tale programma, mirandolo ai partecipanti per i quali produce benefici significativi, potrebbe migliorarne di molto la *performance*.

D'altra parte, la varietà di politiche attive del lavoro messe in atto in quest'ultimo decennio e il loro progressivo affinamento hanno mostrato come alcune siano mediamente piuttosto efficaci, e altre no. Lo stato dell'arte è migliorato molto, e il giudizio di fine anni 1990 sulla loro (scarsa) efficacia è oggi confermato solo in parte, come documenta in maniera persuasiva la meta-analisi di 97 studi di valutazione di diversi programmi svolta da Card, Kluve e Weber (2009). Si è imparato a selezionare le politiche e a disegnarle meglio, e si è imparato parecchio.

Così, vi sono evidenze che, seppure inevitabilmente riferite a specifici contesti (un Paese, un insieme di destinatari, un modo di porre in essere la politica, ecc.) sono nette, in sostanza conclusive, e *a priori* non scontate. Restando a due solo esempi, politiche di occupazione sussidiata si sono rivelate nella grandissima maggioranza dei casi inefficaci, mentre politiche di riavviamento al lavoro che combinino azioni mirate di formazione e *on-the-job-training*, orientamento e obblighi reciproci hanno sovente avuto effetti positivi⁴¹.

Ancora, vi sono casi in cui il risultato delle valutazioni non è univoco, ma restringe in misura non banale le conclusioni. L'insieme di evidenze empiriche accumulate sulle politiche di sostegno del reddito dei disoccupati ne è un esempio significativo⁴². Tra gli effetti (non desiderati) della politica vi è una diminuzione dell'offerta di lavoro, segnatamente per le donne, con un picco di ripresa dell'offerta all'avvicinarsi dell'esaurimento dell'indennità. L'entità di questi effetti varia nei diversi contesti e in relazione all'entità, alla durata, alla progressività della riduzione e alla "condizionalità" dell'indennità. La direzione e l'ordine di grandezza degli effetti appaiono tuttavia acquisiti con un robusto grado di consenso. Anche per le politiche di sostegno del reddito, poi, l'affiancamento di incentivi, azioni attive e "obblighi reciproci" si è mostrato efficace nell'indurre un più sollecito ritorno al lavoro dei disoccupati. In particolare, si sono avuti risultati soddisfacenti quando l'indennità di disoccupazione è stata accompagnata da un *occupational bonus* – cioè a dire, un trasferimento monetario aggiuntivo ai beneficiari dell'indennità che trovano un nuovo lavoro rapidamente e lo conservano per uno specifico periodo di tempo – oppure da combinazioni di servizi per migliorare la ricerca di lavoro e per rafforzare gli obblighi di ricerca/accettazione di un lavoro.

5. *L'esperienza italiana in tema di valutazione degli effetti di politiche*

⁴¹ Per ragguagli sull'argomento, vedi anche Martin e Grubb (2001), OECD (2005: cap. 4) e gli articoli riferiti alle "riforme Hartz".

⁴² Vedi la classica rassegna di Atkinson e Micklewright (1991), Meyer (1990, 1995) e recentemente Fredricksson e Söderström (2008).

5.1. *Il quadro di fondo*

La previsione di rapporti, in genere annuali, del Ministro al Parlamento (o al Comitato Interministeriale per la Programmazione Economica) sullo stato di attuazione di una politica o di una regolazione ha una lunga storia nel nostro Paese. Tuttavia, tali rapporti hanno finalità nettamente diverse dalla valutazione degli effetti dell'intervento. Al meglio, essi ne documentano il processo di realizzazione: le attività organizzative messe in atto, le risorse impiegate, i soggetti coinvolti, quelli che potremmo chiamare i risultati lordi (detto altrimenti, le differenze pre-post) dell'intervento, le difficoltà incontrate e gli intendimenti per farvi fronte. Forniscono dunque informazioni utili, ma non rispondono all'interrogativo proprio di uno studio di valutazione degli effetti. Per di più, progressivamente essi sono sovente scaduti a piatto adempimento burocratico – in qualche caso *tout court* hanno finito per non essere più presentati.

Una significativa novità si registra verso la fine degli anni 1990, e si collega alla riflessione e alle proposte della cosiddetta “Commissione Onofri”⁴³. Uno dei tratti distintivi della relazione finale della Commissione è l'intreccio fra le indicazioni di merito per la riforma del *welfare* e i suggerimenti sugli strumenti per attuarle. Ciò vale non soltanto per “*l'analisi delle compatibilità macroeconomiche della spesa sociale*”, il fuoco del mandato della Commissione⁴⁴, ma anche, in forma più sintetica, per le condizioni amministrative e le strategie di intervento richieste per realizzare le riforme proposte.

Le proposte, di merito e di metodo, della “Relazione Onofri” sono note, e non serve certo che le richiami. Un ampio insieme di contributi che ne discutono merito ed esiti è in un recente volume a cura di Guerzoni (2008), al quale rimando⁴⁵.

5.2. *Le modeste realizzazioni negli ultimi dieci anni*

I dieci anni che ci separano dalla Commissione Onofri sono stati segnati da un vivace dibattito pubblico e da vari interventi riformatori, segnatamente in materia di lavoro e di assistenza. Vanno segnalate, sinteticamente, innovazioni su tre direttrici.

- (a) In tema di strumenti indispensabili per avviare politiche improntate all' “universalismo selettivo”⁴⁶, è stato definito l'Indicatore della Situazione Economica Equivalente (ISEE). L'innovazione non è stata tuttavia seguita da un'attenta azione di vaglio e ridisegno dell'ISEE e di gestione del suo sistema informativo (Toso 2007).
- (b) In materia di assistenza, sono stati attuati alcuni interventi: da un lato l'assegno per il terzo figlio e l'assegno di maternità; dall'altro la “*sperimentazione*”, o meglio il

⁴³ Più propriamente, Commissione per l'analisi delle compatibilità macroeconomiche della *spesa sociale* (1997).

⁴⁴ A ben vedere un mandato ristretto, che fortunatamente la Commissione Onofri in buona parte disattese ampliando la prospettiva alla riforma del *welfare*.

⁴⁵ Traggio parte delle riflessioni che seguono da Trivellato (2008a), pubblicato in tale volume.

⁴⁶ È questo il felice ossimoro con cui Gorrieri (2002: 74-75) ha riassunto un'indicazione-chiave della Commissione, secondo la quale «*la riforma deve ispirarsi a una scelta equilibrata fra universalismo, quanto ai beneficiari, e selettività, quanto all'erogazione delle prestazioni*» (Commissione per l'analisi delle compatibilità macroeconomiche della spesa sociale (1997: 18).

progetto-pilota, del Reddito Minimo di Inserimento (RMI)⁴⁷, rinominato poi reddito di ultima istanza per sancirne in realtà la fine, almeno come intervento di contrasto della povertà generalizzato a livello nazionale.

- (c) In tema di politiche del lavoro e di connesse misure di *welfare*, c'è stata una serie di interventi normativi «di tipo marginale e incrementale»⁴⁸, concentrati sull'introduzione di nuovi contratti di lavoro variamente “flessibili” e su modifiche dei servizi per l'impiego. È stato meno toccato, invece, il fronte dei cosiddetti ammortizzatori sociali, se non per una maggiore generosità dell'indennità di disoccupazione, in termini di *replacement ratio* e di durata⁴⁹.

Già da questi scarni richiami appare evidente che l'originario disegno riformatore è stato disatteso: per la relativa pochezza degli interventi attuati, anche se alcuni concordano con i suggerimenti della Commissione; e soprattutto perché è stata abbandonata l'indicazione-chiave di un disegno organico e della connessa redistribuzione delle risorse fra i vari comparti di spesa per la protezione sociale.

Quanto si riscontra in termini di larga massima per il merito, non è indifferente per il metodo, segnatamente per sviluppi nel disegno di politiche di riforma e nella valutazione dei loro effetti. Pur con le cautele del caso, dalle esperienze di altri Paesi, incluse quelle richiamate poco fa⁵⁰, si trae infatti la fondata opinione che quanto più intenso, organico è il programma di riforme, tanto maggiori sono gli avanzamenti nei metodi e nelle pratiche di valutazione delle politiche.

Una riflessione si impone anche sulle innovazioni negli strumenti e nelle attività di valutazione di politiche pubbliche. Va detto che i suggerimenti della Commissione Onofri al riguardo appaiono ancora oggi largamente condivisibili. Ma essi rivelano anche manchevolezze. Una, in particolare, va menzionata.

L'interesse per il tema della valutazione degli effetti delle politiche è modesto. Alla fin fine, la convinzione della bontà delle proposte fa premio su un atteggiamento di apprendimento per prova ed errore nell'introdurre e ridisegnare al meglio le politiche. Certo, l'indicazione per la «*gradualità e [... la] sperimentazione*» è nitida, ma appare dettata dalla preoccupazione di mettere a punto meccanismi di controllo della spesa più che da un genuino interesse ad imparare, da sperimentazioni e valutazioni dei loro effetti, quali politiche – con il loro specifico disegno di incentivi e obblighi – “funzionario” e quali no.

⁴⁷ “*Sperimentazione*” è il termine usato nei documenti istitutivi del RMI, a partire dal DL n. 237/1998. Ma non si è trattato di un *social experiment* in senso proprio, bensì della selezione ragionata di 39 comuni sui quali è stato inizialmente condotto il progetto-pilota (estesi poi a oltre 270 per il biennio 2001-02), non affiancati da un insieme di comuni di controllo.

⁴⁸ Così Sestito (2002: 17), prima ancora che la scomposta produzione normativa avesse modo di dispiegarsi pienamente, qualificava i mutamenti intervenuti: con la “parzialità incrementale” che vale a sottolineare l'incompletezza delle riforme introdotte con riguardo agli aspetti istituzionali in generale e al sistema di *welfare* in particolare; con la “natura al margine” dei nuovi rapporti di lavoro, riservati ai nuovi entranti, che pone problemi equitativi e favorisce l'accentuazione della struttura duale del mercato del lavoro.

⁴⁹ Vedi Sestito (2002), Pirrone e Sestito (2006), Trivellato (2006) e Pirrone (2008).

⁵⁰ Ma è il caso di ricordare perlomeno la “*war on poverty*” dell'amministrazione Johnson nella seconda metà degli anni 1960, e il compiuto disegno della “*flexicurity*” danese negli anni 1990 (Andersen e Svarer 2007).

Ciò detto, sul piano dell'innovazione degli strumenti e delle attività di valutazione di politiche pubbliche, questi ultimi dieci anni ci consegnano modeste novità.

- (a) Per gli aspetti conoscitivi sul mercato del lavoro, ha avuto notevole rilievo l'operato del Gruppo di lavoro per il monitoraggio degli interventi di politica occupazionale e del lavoro, stimolato dall'azione dell'Unione Europea nel quadro della "strategia di Lisbona". La sua attività è stata decisiva nello stimolare, coordinare e organizzare al meglio, dati i limiti dell'informazione di base disponibile, la produzione statistica e l'analisi sulle politiche del lavoro (Ministero del Lavoro e delle Politiche Sociali 2005; Sestito 2008).
- (b) Un primo, fragile segnale nella direzione della valutazione degli effetti di politiche è venuto dalla "sperimentazione" del RMI. L'aspetto più innovativo è stata la presenza, nel decreto legislativo del giugno 1998 che l'ha disciplinata, di una circostanziata "clausola valutativa": la contestuale previsione di un'attività di valutazione; la destinazione a tale fine di una somma pari allo 0,3% dello stanziamento assegnato al RMI; l'affidamento dell'incarico a una società esterna all'amministrazione, con una procedura di bando⁵¹.
- (c) Recentemente, vi sono stati altri segnali positivi nella direzione di esercizi di valutazione di politiche che mostrano un coinvolgimento dei *policy-makers*. Per quanto mi è dato di conoscere, essi sono circoscritti a: (c1) alcune esperienze di committenza di valutazioni realizzate negli anni recenti, con affidamento tramite bando, dal Ministero del Lavoro e dal Ministero dell'Economia; (c2) studi di valutazione di politiche di aiuto alle imprese condotti da un gruppo di lavoro, integrato da esperti, del Ministero delle Attività produttive (ora dello Sviluppo economico); (c3) l'esperienza dell'Unità di valutazione degli investimenti pubblici (UVAL) e della rete dei Nuclei di valutazione degli investimenti pubblici (NUVV)⁵²; (c4) sporadici, ma apprezzabili studi di valutazione degli effetti di interventi svolti nell'ambito di alcune amministrazioni regionali⁵³; (c5) la pratica di introdurre "clausole valutative" nei testi di legge, che si sta diffondendo nei Consigli regionali coinvolti nel progetto CAPIRE (Controllo delle Assemblee sulle Politiche e gli Interventi Regionali)⁵⁴.

Il quadro complessivo resta, in ogni caso, deludente. Sul terreno degli strumenti e delle modalità di intervento, in particolare delle attività di monitoraggio e di valutazione,

⁵¹ I risultati di questa esperienza sono stati, peraltro, piuttosto modesti. Infatti, per il disegno stesso del progetto-pilota, privo di un campione di comuni di controllo, l'esito non è stato uno studio di valutazione degli effetti propriamente inteso. Per questa ragione, e per l'orizzonte temporale limitato, l'attenzione prevalente è stata dedicata agli aspetti di realizzazione della politica, mentre non si è indagato, se non marginalmente, sugli effetti prodotti. In sostanza, ciò di cui dispone è un discreto monitoraggio dell'intervento (vedi il rapporto di sintesi di IRS, Fondazione Zancan e CLES 2001). Inoltre, la documentazione raccolta non è stata messa a disposizione della comunità dei ricercatori, il che ha ovviamente impedito *replication studies*.

⁵² L'attività dell'UVAL e dei NUVV è stata, peraltro, prevalentemente focalizzata sulla valutazione *ex-ante* di progetti e *in itinere* ed *ex-post* di realizzazioni, più che sulla valutazione degli effetti.

⁵³ Vedi, ad esempio, le attività di valutazione nella Regione Sardegna, da ultimo Lobascio e Mura (2007).

⁵⁴ Vedi <http://www.capire.org>.

c'è un' enfasi impropria su norme e procedure, e una scarsa attenzione ai processi reali che vengono messi in atto e alla "qualità" degli stessi. Tale approccio è favorito dal permanere di un assetto istituzionale decisamente ispirato al "modello continentale" di *accountability* democratica, del quale ho già brevemente richiamato le debolezze nella funzione *lato sensu* di controllo (mentre la decisa evoluzione del sistema politico verso la logica bipolare, l'incremento, nei fatti, dei poteri dell'esecutivo e il crescente peso del suo Presidente richiederebbero una congrua ridefinizione dei rapporti fra esecutivo e Parlamento).

Sullo sfondo, permane poi una sorta di "presunzione di efficacia": la radicata tendenza, cioè, a presumere che una certa azione pubblica serva allo scopo, purché sia realizzata secondo le intenzioni e le disposizioni di chi l'ha concepita (il che porta, al più, a preoccuparsi della sua implementazione da parte delle amministrazioni pubbliche, essenzialmente in chiave di rispetto delle regole). Secondo questo modo di vedere l'azione pubblica, una volta eliminata ogni discrepanza tra progetto e realizzazione, l'intervento sarà inevitabilmente e pienamente efficace. Una versione estrema di questa visione è la nozione giuridica di efficacia di una legge, quando questa viene fatta coincidere con la pubblicazione nella Gazzetta Ufficiale, come se la legge agisse *ex opere operato*.

5.3. *L'emergere di segnali incoraggianti*

Fortunatamente, negli anni recenti cominciano ad emergere anche evidenze di segno opposto. Vi sono apprezzabili studi di valutazione degli effetti di politiche pubbliche condotti in larga misura per iniziativa di ricercatori. Nel quadro delle attività dell'apposita Commissione del Ciss, già ricordata, sono stati prodotti quattro dossier di schede sinottiche che documentano numerosi studi sugli effetti rispettivamente di politiche del lavoro, di interventi di sostegno alle imprese, di programmi di contrasto della povertà, di politiche educative⁵⁵. Parecchi studi sono svolti con metodi appropriati, e approdano a risultati robusti. E in alcune università e in uffici studi di importanti istituzioni vi sono ormai qualificati gruppi di ricerca.

Da parecchi anni opera poi, positivamente, PROVA (PROgetto VALutazione): un centro animato da Alberto Martini, promotore di iniziative tese a diffondere un uso rigoroso della valutazione nella pubblica amministrazione e nelle fondazioni. Esso fornisce stimolo e supporto al già menzionato progetto CAPIRE, che comincia a produrre esiti di un

⁵⁵ Vedi, nell'ordine, Trivellato e Zec (2008), Ercoli e Guelfi (2008), Silvestri e Mo Costabella (2008) e Schizzerotto (2008). Una sintetica ricognizione di studi di valutazione incentrati sulle politiche del lavoro è anche in Contini e Trivellato (2005: 57-69). Inoltre, nell'ambito delle attività istruttorie della Commissione di Indagine sul Lavoro Guelfi (2008) ha curato le sintesi di alcuni studi di caso particolarmente interessanti: per le diverse politiche considerate, per il rigore del metodo e per i risultati ai quali giungono. Gli studi vertono rispettivamente sul lavoro di somministrazione – prima più chiaramente chiamato interinale – (Ichino, Mealli e Nannicini 2005), sulle Liste di mobilità (Paggiaro, Rettore e Trivellato 2009), sul credito d'imposta per le assunzioni a tempo pieno (Cipollone e Guelfi 2006) e sui fondi strutturali per le aree in declino (Bondonio e Greenbaum 2006).

certo rilievo⁵⁶. Inoltre, da sette anni PROVA tiene una *Summer School* su “Metodi quantitativi per la valutazione delle politiche pubbliche”, che ha coinvolto un buon numero di ricercatori, professionisti e operatori pubblici.

Recentemente, gli incontri promossi dalla Commissione del Css hanno trovato una buona rispondenza. In particolare lo *Workshop* su “Problemi di valutazione degli effetti di politiche pubbliche in Italia”, organizzato con la Fondazione Bruno Kessler, a Trento nel marzo 2008, ha coinvolto in un serrato confronto numerosi ricercatori, accademici e non, impegnati in attività di valutazione. A ben vedere, può essere la prefigurazione di un *network*, per il quale ritengo sia utile che la Commissione, e il Css in prima persona, si pongano come nodo di riferimento e vi diano continuità.

Ancora, sulla traccia dell’attività avviata dalla Commissione e grazie all’impegno del collega Antonio Schizzerotto, nell’aprile 2008 si è costituito, presso la Fondazione Bruno Kessler, l’IRVAPP (Istituto per la Ricerca VALutativa sulle Politiche Pubbliche), con la specifica missione di garantire qualità scientifica e continuità temporale a studi di valutazione di politiche pubbliche e di intensificare le relazioni con analoghi istituti a livello europeo e nord-americano. All’IRVAPP, sorto su base associativa, partecipano anche il Css e l’Istituto Cattaneo.

Infine, nell’ottobre 2008, alla conclusione della sua attività, la Commissione di Indagine sul Lavoro ha approvato una raccomandazione su “*Analisi e proposte in tema di monitoraggio e valutazione degli effetti di politiche del lavoro*” (Trivellato 2008b), che recepisce i principali suggerimenti di mature esperienze straniere, nonché le riflessioni e le indicazioni emerse dall’attività della Commissione del Css.

6. *Quali conclusioni?*

Mirando a un sommario bilancio sugli strumenti di implementazione delle politiche pubbliche e sulla valutazione dei loro effetti in Italia, l’interrogativo ricalca uno abituale: quanto è pieno, o vuoto, il bicchiere?

Le indicazioni che vengono dalle esperienze dell’ultimo decennio, lo abbiamo appena visto, sono molteplici e di segno diverso. Promettenti, ma bisognose di consolidarsi, sul fronte della ricerca. Ancora opache, invece, salve eccezioni, sul fronte del processo politico e delle amministrazioni pubbliche.

Questo giudizio può forse sembrare troppo polarizzato. Esso trova, tuttavia, un riscontro difficilmente contestabile. In vari casi, le evidenze degli studi che ho richiamato mettono drasticamente in discussione l’efficacia delle politiche in questione. Dovrebbero quindi indurre una riflessione per il loro ridisegno. Sfortunatamente, queste evidenze continuano ancora a restare ai margini del dibattito pubblico, e trovano scarsa udienza tra politici e dirigenti pubblici.

⁵⁶ È dell’agosto 2008 lo Statuto della Regione Lombardia, che all’art. 14, comma 2, stabilisce che «*Il Consiglio esercita [...] la funzione di controllo delle leggi e la valutazione degli effetti delle politiche regionali*» (l’enfasi del sottolineato è aggiunta). Vedi Regione Lombardia (2008).

Un ultimo esempio, collegato alle recenti vicende sui cambiamenti nell'organizzazione della scuola elementare, è istruttivo. Tra l'altro, è stato stabilito che si passerà dal *team* di maestri al "maestro prevalente". L'interrogativo che l'innovazione, programmata ma non ancora attuata, solleva è ovvio: quale tra le due soluzioni, maestro prevalente *vs.* *team* di maestri, ha un effetto maggiore sulle curve di apprendimento dei ragazzi in giovane età? Un'amministrazione sensibile al problema avrebbe potuto utilizzare la riforma del 1990, che operò nella direzione opposta e portò dal maestro unico al *team* di maestri, per trovare elementi di riposta alla domanda, e per dare così una decente base conoscitiva – detto più apertamente, *un senso* – alle troppe parole che si sono lette e sentite in merito.

In ogni caso, per rispondere all'interrogativo su quanto il bicchiere sia pieno o vuoto, è utile ancorarsi a un metro di giudizio passabilmente oggettivo, quale può essere offerto dal confronto con le esperienze straniere. Esso segnala un grave ritardo dell'Italia. Non soltanto il divario nelle pratiche di disegno, monitoraggio e valutazione degli effetti delle politiche è notevole. Ancora più preoccupante è il fatto che *il divario tende a dilatarsi*. Il nostro Paese rischia una pericolosa deriva, proprio quando le società europee sono chiamate a sfide severe. La capacità di far fronte a tali sfide chiama in causa programmi su molteplici fronti: l'istruzione, l'innovazione e la competitività delle imprese, le politiche attive del lavoro e la sicurezza sociale, il contrasto alla povertà, l'ambiente. E chiama in causa la qualità di tali programmi; quindi, la capacità di valutarne credibilmente gli effetti per (ri)disegnarli in maniera efficace.

Quali conclusioni trarre? Dopo quanto sono venuto dicendo, non sorprenderò affermando che le conclusioni non sono permeate di ottimismo. Esse restano tuttavia aperte, problematiche.

Molto dipenderà dall'evoluzione che mostrerà la *domanda di valutazione*, da parte di *policy makers*, amministrazioni pubbliche e opinione pubblica. Questa crescita della domanda potrà essere favorita da una revisione dell'assetto istituzionale, che accentui l'alterità fra esecutivo e Parlamento e del Parlamento rafforzi le funzioni di controllo e di valutazione delle politiche. Essa potrà inoltre trovare stimolo nella capacità degli *opinion makers*, e degli uomini di cultura, di far crescere la sensibilità e l'interesse per la valutazione degli effetti delle politiche; detto altrimenti, la consapevolezza della sua importanza per l'*accountability* democratica.

Certo, è appena ovvio ricordare che tra le finalità dell'azione pubblica vi è anche il consenso e la legittimazione del decisore. L'assertività e la ricerca di evidenze apologetiche rispetto alle decisioni assunte sono, in qualche misura, connaturate all'azione pubblica. In questa prospettiva, il *policy maker* può guardare alla valutazione degli effetti di un intervento come a una potenziale minaccia, se considera la valutazione essenzialmente come uno strumento di giudizio sulle scelte che egli ha fatto, e non di apprendimento collettivo per migliorare (o confermare) l'intervento. Ma il proposito di corroborare le scelte fatte, o per l'opposizione di contrastarle, non si basa soltanto su apriorismi o sulle «*retoriche dell'intransigenza*» lucidamente tratteggiate da Hirschman (1991).

La novità, intellettuale e pratica, rappresentata dalla valutazione può essere un elemento che gioca a favore della legittimazione del *policy maker*, soprattutto se fa leva su una sua sensibilità per decisioni *evidence based* e si colloca nel quadro di un dibattito pubblico non dominato da discriminanti ideologiche. In proposito, è istruttivo il passo di un articolo della metà degli anni 1990 sullo stato della valutazione in Francia: «*La valutazione rappresenta un tentativo pragmatico di capire ciò che sta succedendo [e funziona], un'idea seducente e rassicurante in un periodo di insicurezza ideologica e quando gli attori politici appaiono sprovvisti di spiegazioni credibili*»⁵⁷.

È questo il percorso, faticoso ma promettente, che si prospetta? I segnali che poco fa ho richiamato inducono a confidare. Guardando alla “*long term persistence*” dell’insieme dei fattori culturali e istituzionali, quel che abitualmente chiamiamo capitale sociale⁵⁸, è ragionevole dubitare.

Bragantini (2008), in un intervento su *Il Sole-24 Ore*, ha sostenuto che “*La politica non discute del merito*” – questo il titolo dell’articolo. E ha scritto, con cruda chiarezza: «*È vero che da noi ci si accapiglia troppo, si decide poco e si realizza, in concreto, ancor meno; ma manca una parte di verità, essenziale. Noi ci accapigliamo sì, ma non discutiamo del merito dei temi; ci scanniamo su aspetti nominalistici, e rifiutiamo di entrare nel concreto, lasciato ai tecnici. Essi razionalizzeranno ‘ex-post’ la decisione, che così non è politica, solo avventata. Il ‘primato’ nuoce alla politica, se vuol decidere senza approfondire: allora prevale solo la presuntuosa superficialità*».

La speranza, per molti l’impegno, è perché questa diagnosi cominci a essere smentita dai fatti.

⁵⁷ Duran, Monnier e Smith (1995: 45-46). L’enfasi del sottolineato è aggiunta.

⁵⁸ Sulla persistenza di lungo periodo del capitale sociale vedi, tra i molti, il saggio di Guiso, Sapienza e Zingales (2008).

Riferimenti bibliografici

- Andersen, T.M. and Svarer, M. (2007) 'Flexicurity – Labour market performance in Denmark', *CESifo Economic Studies*, 53 (3): 389–429.
- Atkinson, A.B. and Micklewright, J. (1991) 'Unemployment compensation and labor market transitions: A critical review', *Journal of Economic Literature*, 29 (4): 1679-1727.
- Behagel, L., Crépon, B., Guitard, J. et Gurgand, M. (2008) *Evaluation d'impact de l'accompagnement de demandeurs d'emploi par les Opérateurs Privés de Placement et le programme Cap Vers l'Entreprise. Rapport intermédiaire*, Paris, J-PAL Europe, Paris School of Economics.
- Berk, R. (2005) *Randomized experiments as the Bronze Standard*, Paper 2005080201, Los Angeles, Department of Statistics Papers, UCLA.
- Beveridge, W. (1942) *Social insurance and allied services*, Report presented to Parliament by command of His Majesty, London, HMSO, CMND 6404 [noto come "Beveridge Report"].
- Biewen, M., Fitzenberger, B., Osikominu, A. and Walzer, M. (2007) *Which program for whom? Evidence on the comparative effectiveness of public sponsored training programs in Germany*, IZA Discussion Paper No. 2885, Bonn, IZA.
- Blundell, R. and Costa Dias, M. (2008) *Alternative approaches to evaluation in empirical microeconomics*, IZA Discussion Paper No. 3800, Bonn, IZA [forthcoming in *Journal of Human Resources*].
- Blundell, R., Costa Dias, M. Meghir, C. and Van Reenen, J. (2004) 'Evaluating the employment impact of a mandatory job search program', *Journal of the European Economic Association*, 2 (4): 569-606.
- Bondonio, D. and Greenbaum, R. (2006) 'Do business investment incentives promote employment in declining areas? Evidence from EU Objective 2 Regions', *European Urban and Regional Studies*, 13 (3): 225-244.
- Bragantini, S. (2008) 'La politica non discute del merito', *Il Sole – 24 Ore*, 24 ottobre 2008.
- Bunzl, M. (2004) 'Counterfactual history: A user's guide', *The American Historical Review*, 109 (3): 845-858.
- Caliendo, M. (2008) *Start-up subsidies in East Germany: Finally, a policy that works?*, IZA Discussion Paper No. 3360, Bonn, IZA.
- Card D., Kluve, J. and Weber, A. (2009) *Active labor market policy evaluations: A meta-analysis*, IZA Discussion Paper No. 4002, Bonn, IZA.

- Cipollone, P. and Guelfi, A. (2006) 'Financial support to permanent jobs: The Italian case', *Politica Economica*, 32 (1): 51-75.
- Commissione per l'analisi delle compatibilità macroeconomiche della spesa sociale (1997) *Relazione finale*, Roma [nota come "Commissione Onofri"] (<http://www.cermlab.it/documents/Relazione%20conclusiva%20Commissione%20Onofri.pdf>).
- Contini, B. e Trivellato, U. (2005) 'Dinamiche e persistenze nel mercato del lavoro italiano: una sintesi', in B. Contini e U. Trivellato (a cura di), *Eppur si muove. Dinamiche e persistenze nel mercato del lavoro italiano*, Bologna, Il Mulino: 13-84.
- Cottingham, P., Maynard, R.A. and Stagner, M. (2004) 'Synthesizing evidence on the impacts of programmes and policies in education, crime and justice, and social welfare: Practical recommendations based on 14 Test-Bed reviews', *Evaluation and Research in Education*, 18 (1 & 2): 28-53.
- Cottingham, P., Maynard, R.A. and Stagner, M. (2005) 'Generating and using evidence to guide public policy and practice: Lessons from the Campbell test-bed project', *Journal of Experimental Criminology*, 1(3): 279-294.
- Cunha, F., Heckman, J.J., Lochner, L. and Masterov, D. (2006) 'Interpreting the evidence on life cycle skill formation', in E. Hanushek and F. Welch (eds.), *Handbook of the Economics of Education*, 1, Amsterdam, Elsevier: 697-812.
- Department for Work and Pensions (2008a) *More support, higher expectations: The role of conditionality in improving employment outcomes. A background research and discussion paper*, London, Department for Work and Pensions (<http://www.dwp.gov.uk/welsh/welfarereform/noonewrittenoff>).
- Department for Work and Pensions (2008b) *No one written off: Reforming welfare to reward responsibility. Public consultation*, London, Department for Work and Pensions (<http://www.dwp.gov.uk/welsh/welfarereform/noonewrittenoff>).
- Department for Work and Pensions (2008c) *No one written off: Reforming welfare to reward responsibility. List of consultation questions*, London, Department for Work and Pensions (<http://www.dwp.gov.uk/welsh/welfarereform/noonewrittenoff>).
- Department for Work and Pensions (2008d) *Impact Assessment of "No one written off: Reforming the welfare state to reward responsibility"*, London, Department for Work and Pensions (<http://www.dwp.gov.uk/welsh/welfarereform/noonewrittenoff>).
- Duran, P., Monnier, E. and Smith, A. (1995) 'Evaluation à la française: Towards a new relationship between social science and public action', *Evaluation*, 1 (1): 45-63.
- Ercoli, R. e Guelfi A. (2008) *Schede sinottiche di studi di valutazione degli effetti di politiche di sostegno alle imprese*, Commissione del CSS "Valutazione degli effetti di politiche pubbliche", doc. n.15, Roma, CSS.

- Ferguson, N. (ed.) (1999) *Virtual history: Alternatives and counterfactuals*, New York, Basic Books.
- Fredriksson, P. and Söderström, M. (2008) *Do unemployment benefits increase unemployment? New evidence on an old question*, IZA Discussion Paper No. 3570, Bonn, IZA.
- Gorrieri, E. (2002) *Parti uguali fra disuguali*, Bologna, Il Mulino.
- Greenberg, D. and Shroder M. (1997) *The digest of social experiments*, Washington, D.C., The Urban Institute Press.
- Guerzoni, L. (a cura di) (2008) *La riforma del Welfare. Dieci anni dopo la "Commissione Onofri"*, Bologna, Il Mulino.
- Guiso L., Sapienza P. and Zingales, L. (2008) *Long term persistence*, Firenze, European Univerisity Institute (mimeo).
- Heckman, J.J. (1996) 'Randomization as an instrumental variable', *Review of Economics and Statistics*, 78 (2): 336-341.
- Heckman, J.J. (2000) 'Causal parameters and policy analysis: A twentieth century retrospective', *The Quarterly Journal of Economics*, 115 (1): 45-97.
- Heckman, J.J. (2001) 'Microdata, heterogeneity and the evaluation of public policy: Nobel Lecture', *Journal of Political Economy*, 109 (4): 673-748.
- Heckman, J.J. (2008) 'Econometric causality', *International Statistical Review*, 76 (1): 1-27.
- Heckman, J.J., LaLonde, R. and Smith, J. (1999) 'The economics and econometrics of active labor market programs', in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics – Vol. 3A*, Amsterdam, Elsevier.
- Heckman, J.J. and Smith, J. (1995) 'Assessing the case for social experiments', *Journal of Economic Perspectives*, 9 (2): 85-110.
- Heckman, J.J., Stixrud, J. and Urzua, S. (2006) 'The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior', *Journal of Labor Economics*, 24(3): 661-700.
- Hirschman, A.O. (1991) *The rhetoric of reaction. Perversity, futility, jeopardy*, Cambridge, MA, The Belknap Press of Harvard University Press.
- Holland, P. (1986) 'Statistics and causal inference', *Journal of the American Statistical Association*, 81 (396): 945-960.
- Huber M., Lechner M., Wunsch C. and Walter T. (2009), *Do German welfare-to work programmes reduce welfare and increase work?*, IZA Discussion Paper No. 4090, Bonn, IZA.

- Ichino, A., Mealli, F. and Nannicini, T. (2005) 'Temporary Work Agencies in Italy: A springboard to permanent employment?', *Giornale degli Economisti e Annali di Economia*, 64 (1): 1-27.
- Imbens, G. and Lemieux, T. (eds.) (2008) *The regression discontinuity design: Theory and applications*, special issue of *Journal of Econometrics*, 144 (2).
- Imbens, G. and Rubin, D. (2009) *Causal inference in statistics, and in the social and biomedical sciences*, New York, Cambridge University Press [in corso di pubblicazione].
- Imbens, G. and Wooldridge, J.M. (2008) *Recent developments in the econometrics of program evaluation*, IZA Discussion Paper No. 3640, Bonn, IZA.
- IRS, Fondazione Zancan e CLES (2001) *Valutazione della sperimentazione dell'istituto del Reddito Minimo di Inserimento*, 3 voll., Milano (mimeo).
- Jacobi, L. and Kluge, J. (2006) *Before and after the Hartz Reforms: The performance of active labour market policy in Germany*, IZA Discussion Paper No. 2100, Bonn, IZA.
- Kluge, J. et al. (2007) *Active labor market policies in Europe, Performance and perspective*, Berlin-Heidelberg, Springer-Verlag.
- Lee, M.-J. (2005) *Micro-econometrics for policy, programme, and treatment effects*, Oxford, Oxford University Press.
- Lewis, D. (1973) 'Causation', *Journal of Philosophy*, 70: 556-567.
- Lobascio, I. e Mura A. (2007) *Valutazione della Legge regionale n. 4 del 2006, art. 4. Stima dell'effetto dell'imposta sulla nautica sugli scali e sulle presenze di unità da diporto in Sardegna*, Cagliari, Regione Autonoma della Sardegna, Osservatorio Economico della Sardegna.
- Marschak, J. (1953) 'Economic measurements for policy and prediction', in W. Hood and T. Koopmans (eds.), *Studies in econometric method*, New York, John Wiley, pp. 1-26.
- Marshall, A. (1961 [1890]) *Principles of economics*, London, Macmillan.
- Martin, J.P. and Grubb, D. (2001) 'What works and for whom: A review of OECD countries' experience with active labour market policies', *Swedish Economic Policy Review*, 8 (1), pp. 9-56.
- Martini, A. (2008) *How counterfactuals got lost on the way to Brussels*, paper presented at the Symposium "Policy and programme evaluation in Europe: Cultures and prospects", Strasbourg, July 3-4, 2008 (mimeo).
- Martini, A., Rettore, E. e Trivellato, U. (2009) 'Valutare gli effetti delle politiche attive del lavoro: la logica controfattuale', in M. Cantalupi e M. Demurtas (a cura di), *Politiche di attivazione e 'performance' dei servizi per l'impiego. Esperienze e percorsi di*

- implementazione in Italia e in Europa*, Bologna, Il Mulino [in corso di pubblicazione].
- Mayne, J. (2001) 'Addressing attribution through contribution analysis; Using performance measures sensibly', *Canadian Journal of Programme Evaluation*, 16 (1): 1-24.
- Meyer B.D. (1990) 'Unemployment insurance and unemployment spells', *Econometrica*, 58 (4): 757-782.
- Meyer B.D. (1995) 'Lessons from the U.S. unemployment insurance experiments', *Journal of Economic Literature*, 33 (1): 91 -131.
- Ministero della Solidarietà Sociale (2007) *Relazione al Parlamento. Attuazione della sperimentazione del Reddito Minimo di Inserimento e risultati conseguiti*, Roma.
- Ministero del Lavoro e delle Politiche Sociali – Gruppo di Lavoro per il Monitoraggio degli Interventi di Politica Occupazionale e del Lavoro (2005) *Aggiornamento del quadro informativo sulle politiche del lavoro*, Roma (<http://www.welfare.gov.it/Statistiche/MercatoLavoro/default.htm>).
- National Conference of State Legislatures (2008) *Ensuring the public trust 2008. Program policy evaluation's role in serving State Legislatures*, (<http://www.oppaga.state.fl.us/reports/pdf/EPT2008.pdf>).
- OECD (2004) *Employment Outlook 2004*, Paris.
- OECD (2005) *Employment Outlook 2005*, Paris.
- Paggiaro, A., Rettore, E. and Trivellato, U. (2009) 'The effect of extending the duration of eligibility in an Italian labour market programme for dismissed workers', *Labour*, 23 (1): 37-66.
- Pearl, J. (2000) *Causality. Models, reasoning, and inference*, Cambridge, United Kingdom, Cambridge University Press.
- Pirrone, S. (a cura di) (2008) *Flessibilità e sicurezze. Il nuovo welfare dopo il Protocollo del 23 luglio*, Bologna, Il Mulino.
- Pirrone, S. e Sestito, P. (2006) *Disoccupati in Italia. Tra Stato, Regioni e cacciatori di teste*, Bologna, Il Mulino.
- Regione Lombardia (2008) *Statuto d'Autonomia della Lombardia. Legge regionale statutaria 30 agosto 2008, n. 1*, Milano.
- Regonini, G. (2009) *La valutazione delle politiche; una sfida per le Assemblee*, in Servizio Valutazione Processo Legislativo e Politiche Regionali (a cura di), *Atti del Convegno 'Fare leggi, rappresentare interessi, valutare politiche'*, Consiglio Regionale della Lombardia, Milano: 32-35.
- Rettore, E. e Trivellato, U. (1999) 'Come disegnare e valutare politiche attive del lavoro', *Il Mulino*, 48 (385): 891-904.

- Rettore, E., Trivellato, U. e Martini A. (2003) ‘La valutazione delle politiche del lavoro in presenza di selezione: migliorare la teoria, i metodi o i dati?’, *Politica Economica*, 19 (3): 301-341.
- Rubin, D.B (1974) ‘Estimating causal effects of treatments in randomized and non-randomized studies’, *Journal of Educational Psychology*, 66 (5): 688-701.
- Sestito, P. (2008) ‘Il monitoraggio e la valutazione delle politiche del lavoro in Italia: l’esperienza del gruppo di monitoraggio presso il Ministero del Lavoro’, in M. Cantalupi e M. Demurtas (a cura di), *Politiche di attivazione e ‘performance’ dei servizi per l’impiego. Esperienze e percorsi di implementazione in Italia e in Europa*, Bologna, Il Mulino [in corso di pubblicazione].
- Shadish, W., Cook, T. and Campbell, D. (2002) *Experimental and quasi-experimental designs for generalized causal inference*, Boston, Houghton-Mifflin.
- Schizzerotto, A. (2008) *Schede sinottiche di studi di valutazione degli effetti di politiche educative*, Commissione del CSS “Valutazione degli effetti di politiche pubbliche”, doc. n.17, Roma, CSS.
- Silvestri, P. e Mo Costabella, L. (2008) *Schede sinottiche di studi di valutazione degli effetti di politiche di contrasto della povertà e sostegno alle famiglie con figli*, Commissione del CSS “Valutazione degli effetti di politiche pubbliche”, doc. n.16, Roma, CSS.
- Stephan, G. and Pahnke, A. (2008) *The relative effectiveness of selected active labour market programmes and the common support problem*, IZA Discussion Paper No. 3767, Bonn, IZA.
- Toso, S. (2007) ‘L’ISEE alla prova dei fatti: uno strumento irrinunciabile, ma da riformare’, in L. Guerzoni (a cura di), *Le politiche di sostegno alle famiglie con figli. Il contesto e le proposte*, Bologna, Il Mulino: 299-315.
- Trivellato, U. (2006) ‘Trasformazioni del mercato del lavoro italiano, politiche del lavoro e disponibilità/fabbisogni informativi: un percorso a metà del guado’, *Economia e Società Regionale*, 94 (2); 5-28.
- Trivellato, U. (2008a) ‘L’implementazione e la valutazione degli effetti delle politiche: tra buoni propositi e pochezza delle pratiche’, in L. Guerzoni (a cura di) *La riforma del welfare. Dieci anni dopo la “Commissione Onofri”*, Bologna, Il Mulino: 149-162.
- Trivellato, U. (a cura di) (2008b) *Cambiamenti del lavoro, protezione sociale e politiche attive del lavoro. Rapporto tematico n. 10: Analisi e proposte in tema di valutazione degli effetti di politiche del lavoro*, Commissione di Indagine sul Lavoro, Roma, CNEL (<http://www.portalecnel.it/Portale/IndLavrapportiFinali.nsf/vwCapitoli?OpenView&Count=40>).

- Trivellato, U. e Zec, S. (2008) *Schede sinottiche di studi di valutazione degli effetti di politiche del lavoro*, Commissione del CSS “Valutazione degli effetti di politiche pubbliche”, doc. n.14, Roma, CSS.
- Trochim, W. (1984) *Research design for program evaluation: The regression-discontinuity approach*, Sage Publications, Beverly Hill (CA).
- Winship, C. and Morgan, S. (2007) *Counterfactuals and causal inference*, Cambridge, UK, Cambridge University Press.
- Woodward, J. (2003) *Making things happen: A theory of causal explanation*, New York, Oxford University Press.
- Wunch, C. and Lechner, M. (2007) *What did all the money do? On the general ineffectiveness of recent West German labour market programmes*, IZA Discussion Paper No. 2800, Bonn, IZA.
- Van Reenen, J. (2004) ‘Active labour market policies and the British New Deal for youth in context’, in D. Card, R. Blundell and R. Freeman (eds.), *Seeking a premier economy: The economic effects of British economic reforms, 1980-2000*, Chicago, University of Chicago Press: 461-496.
- Van Reenen, J. (2005) *No more skivvy schemes? Active labour market programmes and the British New Deal for the young unemployed in context*, London, Institute for Fiscal Studies.
- Wooldridge, J. (2002) *Econometric analysis of cross-section and panel data*, Cambridge, MA, MIT Press.