

I
L
A
I
R
E
T
A
M

Il Formez nel 2001 ha rilanciato e potenziato l'attività editoriale con una nuova veste grafica e con contenuti fortemente collegati alle attività di formazione, assistenza e consulenza.

La produzione editoriale si articola in tre collane:

“Quaderni”, momenti di riflessione teorica e culturale su temi di grande rilievo per la Pubblica Amministrazione e per lo sviluppo locale; “Strumenti”, guide, manuali e modulistica di supporto agli operatori pubblici e privati; “Azioni di Sistema per la Pubblica Amministrazione” che presentano i risultati dei progetti realizzati all'interno delle Azioni di Sistema del Fondo Sociale Europeo per sostenere i processi di riforma della P.A.

Nel 2005 sono state riattivate le collane “Materiali” e “Materiali Azioni di Sistema per la P.A.” che contengono una documentazione didattica immediata e preziosa su attività di formazione e assistenza dell'Istituto.

Esse rappresentano anche un momento di verifica e uno strumento utile ad attivare rapporti di partenariato e di scambio tra le esperienze realizzate in vari contesti.

Pertanto possono essere la base per futuri approfondimenti e sono comunque un prodotto finito di formazione permanente per tutti gli utenti.

La presente ricerca inquadra la valutazione degli effetti delle politiche, intesa come analisi controfattuale, distinguendola dalle altre pratiche “valutative” adottate nelle P.A.

Essa contribuisce a delineare, nella prima parte, i principali metodi finalizzati a comprendere se la politica, intesa come terapia ad un problema collettivo, ha agito nel senso desiderato e in che misura ha prodotto benefici sui destinatari potenziali.

Nella seconda parte vengono analizzati sei casi di valutazione di effetti delle politiche promosse da diverse tipologie di amministrazioni pubbliche.

In conclusione vengono proposte, quali utili consigli operativi, delle “linee guida” per le P.A. che intendono intraprendere nuove pratiche valutative tuttora scarsamente utilizzate in ambiente pubblico.

Angelo Raffaele Dinardo
Vicepresidente Formez



**VALUTARE GLI EFFETTI DELLE
POLITICHE PUBBLICHE**
**METODI E APPLICAZIONI
AL CASO ITALIANO**

Questo volume è stato redatto nell'ambito del Progetto Governance,
Linea Osservatorio

Coordinamento del progetto
Valeria de Magistris, *Dirigente Formez*
Giuseppe Gioioso, *Formez*

L'indagine è stata realizzata da
ProgettoValutazione, sotto la direzione di Alberto Martini

Il volume è stato redatto da
Alberto Martini, Luca Mo Costabella, Marco Sisti, con la collaborazione
di Barbara Romano

Organizzazione editoriale
Paola Pezzuto

INDICE

Premessa, <i>di Giuseppe Gioioso</i>	7
--	---

Introduzione

Valutazione e governance pubblica, <i>di Marco Sisti</i>	11
--	----

PRIMA PARTE: METODI PER LA VALUTAZIONE DEGLI EFFETTI DELLE POLITICHE.....19

Capitolo 1. Cinque modi di interpretare la “valutazione”, <i>di Alberto Martini</i>	21
---	----

Capitolo 2. L’approccio controfattuale alla valutazione degli effetti delle politiche, <i>di Alberto Martini e Marco Sisti</i>	33
---	----

Capitolo 3. Il metodo sperimentale: un punto di riferimento per la valutazione degli effetti, <i>di Alberto Martini</i>	51
--	----

Capitolo 4. I confronti spazio-temporali con dati non-sperimentali e le minacce alla loro validità, <i>di Alberto Martini e Marco Sisti</i>	77
--	----

Capitolo 5. La regressione e il <i>matching</i> statistico come strumenti per ridurre le differenze di partenza, <i>di Alberto Martini</i>	95
---	----

Capitolo 6. I metodi basati sulla discontinuità nel trattamento, <i>di Alberto Martini</i>	117
---	-----

Appendice: Dettagli sul <i>matching</i> statistico.....	147
---	-----

SECONDA PARTE: APPLICAZIONI AL CASO ITALIANO.....151

Capitolo 7. Gli effetti dei finanziamenti della legge 488 sulla performance delle imprese e sullo sviluppo locale, <i>di Alberto Martini</i>	157
---	-----

Capitolo 8. Gli effetti della legge regionale sarda 15/94 sulla performance delle imprese, <i>di Luca Mo Costabella</i>	185
--	-----

Capitolo 9. Gli effetti della legge 27/93 della Regione Toscana sull’imprenditoria giovanile, <i>di Marco Sisti</i>	199
--	-----

Capitolo 10. Gli effetti del lavoro interinale sulla probabilità di trovare un impiego stabile, <i>di Luca Mo Costabella</i>	227
---	-----

Capitolo 11. Gli effetti dell'indennità di mobilità sulla probabilità di rioccupazione dei lavoratori licenziati, <i>di Luca Mo Costabella</i>	243
Capitolo 12. Gli effetti delle borse di studio sulla regolarità degli studi universitari, <i>di Luca Mo Costabella</i>	265
Conclusioni: Qualche suggerimento (e cautela) per chi intenda valutare gli effetti di una politica, <i>di Alberto Martini e Marco Sisti</i>	285
Bibliografia.....	303

PREMESSA

di Giuseppe Gioioso

La presente indagine si inserisce nelle analisi* sviluppate nell'ambito della Linea Osservatorio del Progetto Governance sulle nuove modalità di azione e strumenti coerenti con il nuovo paradigma della buona governance pubblica.

Ma qual è il legame tra la valutazione degli effetti delle politiche, oggetto del presente volume, e le innovazioni negli assetti istituzionali e gestionali delle PA riconducibili agli approcci della good public governance?

Per spiegarlo occorre richiamare le indicazioni dettate dalla Commissione Europea nel suo Libro Bianco, secondo cui le azioni delle PA devono essere improntate a principi di apertura, trasparenza, partecipazione, responsabilità, efficacia e coerenza. Ossia, detta in modo colorito ed un po' retorico, devono abbassare i propri ponti levatoi per aprirsi al confronto, al dialogo, alla partecipazione degli stakeholder e dell'intera comunità, sin dalla fase dell'individuazione, alla gestione e alla valutazione delle public policy.

La governance, infatti, delinea un passaggio da un modello di autorità, basato sul principio di legalità che definisce una ripartizione rigida e gerarchica dei poteri e delle competenze, ad uno, basato sul principio di sussidiarietà, in cui, più che la ripartizione formale, diviene centrale il risultato finale delle azioni pubbliche e le modalità e le forme più adeguate per il suo raggiungimento.

La forte flessibilità del modello non riguarda soltanto i rapporti tra le PA, ma anche tra queste ed il privato; non solo attraverso l'esternalizzazione di funzioni e servizi pubblici, o la costituzione di partnership pubblico/privato, ma anche e soprattutto attraverso l'arricchimento del processo decisionale mediante la partecipazione dei soggetti destinatari delle politiche pubbliche.

**La public governance in Europa* (a cura di V. de Magistris) - Collana Quaderni Formez n.30 (7 voll); *Nuovi soggetti della governance esterna* (a cura di V. de Magistris e G. Gioioso) Collana Quaderni Formez n. 31; *Governance regionale e agenzie di sviluppo* (a cura di V. de Magistris) Collana Materiali Formez ; *Nuovi profili di accountability nelle PA italiane: Teoria e strumenti* (a cura di V. de Magistris e G. Gioioso) Collana Quaderni Formez n. 40 (2 voll + CD rom); *La pianificazione strategica partecipata in Italia* (a cura di G. Gioioso) Collana Materiali Formez; "La governance nei servizi sociali", report non pubblicato.

Si delinea quindi un sistema multilivello estremamente complesso di competenze formali ed informali in cui i pubblici poteri non possono più invocare esclusivamente una legittimazione “formale” derivante dalla legge, ma sono chiamati a conquistare una “sostanziale” che viene a dipendere in larga parte dalla capacità di fissare obiettivi condivisi di politiche pubbliche, di assicurarne l’implementazione amministrativa, di mediare, in vista della concreta decisione, tra una vasta gamma di interessi pubblici e privati, di adempiere ai nuovi compiti con competenza tecnica e con cognizioni scientifiche appropriate. In tale scenario, alle PA è richiesta una profonda riconfigurazione di ruolo e una reinterpretazione intelligente della propria ragion d’essere. Una PA che non fa più tutto autonomamente, ma delega, esternalizza e fa partecipare altri soggetti delle proprie funzioni ed attività deve necessariamente riuscire a sviluppare un ruolo di regia sapiente, di indirizzo e coordinamento di network, che presuppone una capacità di valutare gli effetti delle proprie azioni e di apprendere dall’esperienza dei risultati, al fine di acquisire una maggiore autorevolezza ed una più ampia legittimazione “sostanziale” verso i propri stakeholder e verso l’intera comunità amministrata.

L’aumento della complessità che questo nuovo modo di operare comporta, richiede l’utilizzo di un adeguato set di strumentazione orientato a fornire alla PA e ai suoi stakeholder una maggiore consapevolezza delle proprie azioni e dei relativi risultati. Attraverso la conoscenza degli effetti delle politiche, si produce una più forte legittimazione nell’ambito di processi decisionali inclusivi e disarticolati. La legittimazione dipende dunque dall’efficacia dell’azione amministrativa. Da qui si evidenzia l’estrema rilevanza dell’affermazione delle pratiche valutative e della valutazione intesa come apprendimento.

Tuttavia, se tutti sono concordi sulla necessità di valutare l’efficacia, si registra ancora, in pratica ed in letteratura, una forte confusione metodologica sul concetto stesso e sulle modalità più adeguate per la sua misurazione. Molto spesso, infatti, si dà all’efficacia un significato notevolmente restrittivo, qualificandola cioè, come semplice capacità di realizzazione degli obiettivi, per la cui misurazione sono sufficienti semplici indicatori che quantificano la percentuale di quelli realizzati rispetto a quelli programmati. Per tale via un’amministrazione può qualificarsi efficace se realizza in tutto o in buona parte

quanto promesso. Seppur utile e necessaria, un tale tipo di valutazione, più che al principio di efficacia, risponde più propriamente ad altri due principi della riforma amministrativa e della buona governance pubblica: quello di responsabilità e quello relativo di accountability. Entrambi si basano sul rapporto di delega e quindi di attribuzioni di poteri da parte del delegante, a cui fa da contrappeso l'assunzione di responsabilità e il dovere di rendere conto dei risultati raggiunti da parte del delegato. Tali rapporti sono alla base delle democrazie rappresentative e delle nuove logiche di gestione delle PA ed investono le relazioni tra cittadini\elettori e politici\eletti, tra legislativo ed esecutivo, tra esecutivo e dirigenza, ecc., sui cui strumenti di resa del conto (accountability) verso l'alto (controllo strategico) e verso il basso (bilancio sociale e bilancio di mandato) si è già ampiamente trattato nell'ambito di una recente indagine del Formez*.

In questo volume si fa riferimento ad un significato diverso del concetto di efficacia che richiama necessariamente quello di politica pubblica, giacchè un'amministrazione efficace non è quella che realizza semplicemente programmi e progetti, ma quella che tramite politiche pubbliche, intese come terapie, dà risposta a problemi e bisogni collettivi. In tal senso un'amministrazione efficace è quella che riesce ad elaborare, prescrivere una cura, una terapia, ossia una politica pubblica, in grado di risolvere o quantomeno dare sollievo ad una patologia generale, intesa come problema o bisogno collettivo su cui si decide di intervenire. Su tale presupposto la valutazione dell'efficacia richiede una misurazione degli effetti netti della politica pubblica che rifugga dalla pratica superficiale di coloro che tendono ad accettare l'assunto del *post hoc ergo propter hoc*, determinando quale effetto della politica la semplice differenza tra le osservazioni post e pre-intervento. La metodologia di valutazione degli effetti proposta invece nel volume si basa sull'approccio controfattuale che "complica", ma rende più rigorosa la stima degli effetti, tentando di stabilire relazioni causali tra fenomeni complessi e depurando la semplice differenza sopraccitata della stima di un termine ignoto, relativo alla situazione senza intervento. Tale tipo di valutazione pur essendo eminentemente *ex-post*, è orientata al futuro in termini di utilizzo. Nonostante

* Nuovi profili di accountability nelle PA italiane, cit.

la sua natura retrospettiva, è fondamentalmente prospettica nell'utilizzo, nel senso che la sua ambizione di fondo è quella di riorientare il disegno delle future politiche verso quelle forme di intervento che siano risultate più efficaci in passato. In questo senso, se realizzata in modo sistematico e rigoroso, la valutazione degli effetti delle politiche può essere un importante strumento per una *buona governance* pubblica.

Sulla base di tali considerazioni lo scopo del volume è stato quello di provare a fare innanzitutto chiarezza metodologica su ciò che si intende per valutazione nella PA, analizzando in dettaglio, ma in modo sempre chiaro e comprensibile i metodi per la valutazione degli effetti delle politiche pubbliche, proponendo sia esemplificazioni, che analisi di casi di pratiche valutative e terminando con una serie di utili consigli alle pubbliche amministrazioni che intendono approcciare la valutazione come apprendimento.

INTRODUZIONE

VALUTAZIONE E GOVERNANCE PUBBLICA

di Marco Sisti

Non è raro trovare nei discorsi dei politici, nei documenti ufficiali, o negli editoriali degli opinionisti, il richiamo alla necessità di *valutare le conseguenze prodotte dall'azione pubblica sulla collettività*. Secondo questo sentire Comune, conoscere in che misura si è stati efficaci nell'approntare soluzioni ai numerosi problemi presenti nell'agenda politica è un ingrediente per “ben governare”. Valutare gli effetti delle politiche e governare sembrerebbero quindi due attività strettamente connesse: senza la realizzazione della prima verrebbe meno una delle condizioni per far bene la seconda.

Sfortunatamente non sempre a tali affermazioni di principio seguono decisioni ed azioni coerenti; e in Italia la valutazione, in particolare quella degli effetti, resta sovente allo stato di pura invocazione. In molti casi (forse nella maggior parte) essa si manifesta sotto forma di *slogan*, di accattivante formula retorica, di nobile e volenterosa aspirazione che sembra destinata (verrebbe da dire condannata) a rimanere tale.

Queste considerazioni iniziali, che a qualcuno potranno apparire pessimistiche e ad altri semplicemente realistiche, non devono però scoraggiare chi si accinga a leggere questo volume. Non è ovviamente questo il fine per cui le abbiamo formulate. Guardare in faccia la realtà ed evidenziarne gli aspetti negativi è solo un modo per iniziare a cambiarla. D'altra parte, il fatto che la valutazione degli effetti abbia avuto, fino a questo momento, una scarsa applicazione nel nostro Paese non significa necessariamente che non ne possa avere una maggiore in futuro. O che non possa, da qui a qualche anno, consolidarsi e diffondersi nelle pubbliche amministrazioni italiane come una prassi abituale di lavoro.

Questo volume è anzi motivato proprio da questa convinzione: cioè che sia *davvero* possibile cercare di rispondere, in modo intellettualmente onesto, a domande sugli effetti delle politiche e, soprattutto, che tali risposte possano *davvero* arricchire il processo decisionale pubblico. Con la conseguenza di rendere più informate, e quindi anche più consapevoli, le successive scelte di *policy*. Vogliamo perciò essere fiduciosi

sul fatto che la valutazione possa, in un futuro non troppo lontano, diventare realmente uno strumento utilizzabile (ed utilizzato) da chi decide.

Affinché ciò possa concretizzarsi, occorre però non assegnare alla valutazione missioni troppo ambiziose o compiti al di fuori della sua portata. Da un lato è necessario essere pienamente consci delle concrete potenzialità di questo particolare strumento d'analisi, senza nascondere i notevoli limiti e le grandi difficoltà d'impiego; dall'altro, è necessario tentare di individuare e far funzionare i giusti incentivi e i meccanismi istituzionali, che possano rendere vantaggioso ed utile, per i decisori pubblici, servirsi della valutazione. Proprio su questo versante recenti innovazioni amministrative e nuovi modi di concepire la gestione e il “governo della cosa pubblica” sembrano aprire il campo ad opportunità, fino a qualche anno fa, impensabili.

Recentemente il tema dell'efficacia delle politiche e della sua valutazione si è accompagnato a quello più ampio della governance pubblica. Testimone qualificato di questa tendenza è il Libro Bianco dedicato dall'Unione Europea proprio a questo argomento. In tale documento l'efficacia delle politiche viene citata tra i cinque principi ispiratori di una buona governance (assieme ad apertura, partecipazione, responsabilità e coerenza). Lo stesso documento contiene anche un riferimento esplicito all'importanza della valutazione. Il Libro Bianco sottolinea infatti come *“le politiche dell'Unione Europea devono essere efficaci e tempestive, producendo i risultati richiesti in base a obiettivi chiari, alla valutazione del loro impatto futuro e, ove possibile, delle esperienze acquisite in passato”*.

Ma quale collegamento esiste tra il concetto di governance pubblica e quello di valutazione? Perché tali termini vengono così spesso associati? Perché alla maturazione dell'una dovrebbe (o potrebbe) seguire anche un maggior impiego dell'altra? La risposta a queste domande non è semplice, né scontata; tanto più che i due concetti in esame non hanno una definizione chiara ed univoca, e a volte sono usati semplicemente come parole d'ordine, a metà tra la metafora e il *wishful thinking*, con un valore più evocativo che sostanziale.

Alla base di tale associazione crediamo vi sia l'idea che un sistema pluralistico di governo (di governance, appunto), fondato non più sulla esclusiva volontà e sulla forza

di una sola autorità centrale, ma sulla forte interazione e sulla partecipazione di più attori sociali¹, abbia bisogno di nuovi strumenti cognitivi, che permettano una maggiore conoscenza dei fatti e, nello stesso tempo, una più solida legittimazione delle scelte compiute.

In questa prospettiva, la valutazione può aiutare le diverse parti in gioco ad avere una più profonda capacità d'analisi dei problemi, nonché delle soluzioni adottate per darvi risposta. Essa permette (o dovrebbe permettere) di portare, sul tavolo delle decisioni, argomentazioni basate non solo sulla difesa di valori e di interessi particolari, ma anche sull'osservazione rigorosa della realtà e, quindi, sul portato dell'esperienza passata. Detto in altri termini, la presenza di più voci in seno al processo decisionale apre anche uno spazio inatteso per la costruzione di *evidence-based policies*. Quasi paradossalmente, un modello di governance basato su una concezione fortemente pluralistica della decisione pubblica, può creare i giusti incentivi per un maggiore impiego di strumenti e metodi, riconducibili ad una visione più "razionalistica" delle politiche pubbliche.

Essere informati su ciò che ha funzionato delle precedenti politiche e avere consapevolezza di ciò che invece si è rivelato un fallimento, consente infatti ai singoli attori di essere più forti e più credibili nel momento in cui avanzano proposte di modifica o ricette alternative. Sapere "come sono andate le cose" è una risorsa aggiuntiva che può essere opportunamente spesa sul banco delle trattative e delle negoziazioni.

Se il processo di formulazione delle scelte pubbliche si complica, se i passaggi decisionali si moltiplicano, se le arene di *policy* si allargano e il numero dei soggetti "politici" chiamati in causa sulle singole questioni aumenta, crescono parallelamente le occasioni per promuovere ed utilizzare studi di valutazione e per verificare l'efficacia delle soluzioni proposte. L'*audience* interessata a questo tipo di analisi diventa infatti

¹ In termini generali con il termine *governance* (utilizzato spesso in opposizione a quello di *government*) viene fatto riferimento ad un nuovo modo di governare, diverso dal tradizionale modello gerarchico nel quale l'autorità statale esercita un controllo sovrano sul popolo e sulla "società civile". In questo nuovo sistema si amplia il numero di soggetti "non governativi" che partecipano attivamente e in modo esplicito alla formulazione e all'implementazione delle politiche pubbliche. Per un approfondimento sul tema, si veda Mayntz R., "La teoria della *governance*: sfide e prospettive", *Rivista Italiana di Scienza della Politica*, a. XXIX, n. 1, 1999, pp. 3-22.

sempre più consistente e cresce di conseguenza l'attenzione sui risultati conoscitivi che la valutazione è in grado di produrre.

Una simile evoluzione potrebbe condurre la valutazione al di là della sua attuale dimensione di “adempimento burocratico” (come viene perlopiù concepita e vissuta in molte realtà amministrative) e farla finalmente diventare, anche in Italia, uno strumento per alimentare l'attenzione e la riflessione collettiva intorno alle politiche pubbliche. Gli esiti delle valutazioni potrebbero così trovare nuovi canali d'espressione e divenire oggetto di discussione e di approfondimento al di fuori degli uffici degli enti pubblici, dove oggi sembrano essere confinati.

La sfida (mai completamente vinta) della valutazione degli effetti

Le politiche pubbliche hanno sempre, per loro natura, esiti incerti. Per quanto siano ben disegnate e concepite, esse intervengono in contesti economici e sociali complessi e suscettibili di continui cambiamenti e trasformazioni impreviste. Appare quindi evidente quanto sia importante apprendere se una politica pubblica è stata capace di produrre gli effetti desiderati, e se, nel contempo, è stata in grado di evitare (o non provocare) effetti sgraditi. È fondamentale in questa esigenza conoscitiva che ha la sua ragion d'essere la valutazione degli effetti, oggetto di trattazione in questo volume.

Riuscire ad attribuire meriti o demeriti ad una politica pubblica è però un'operazione sempre molto complessa e, come vedremo, una sfida cognitiva mai completamente vinta. La difficoltà maggiore sta nel fatto che molti altri fenomeni, indipendenti dalle attività messe in atto dalla pubblica amministrazione, influenzano più o meno direttamente i comportamenti e le condizioni che la politica intende modificare. Senza la necessaria accortezza nell'interpretazione delle informazioni a nostra disposizione, si rischia di sovrastimare, o viceversa di sottostimare, il contributo dato dalla politica stessa.

Capire quale ricaduta abbiano avuto le politiche attuate può quindi diventare un *input* importante nel processo decisionale pubblico, purché vengano rispettati alcuni principi essenziali. In particolare, tali valutazioni devono essere effettuate utilizzando metodologie d'analisi rigorose che siano in grado di produrre risultati credibili. A volte

l'impiego di tale rigore metodologico dovrà persino condurre il valutatore a riconoscere la non-valutabilità di una politica e dei suoi effetti. Ammettere i limiti degli strumenti e dei modelli d'analisi impiegati non deve essere considerato come un sintomo di debolezza intellettuale, quanto di serietà e correttezza professionale.

Un messaggio infatti deve essere chiaro: non di tutte le politiche è possibile valutare gli effetti (anche se magari potrebbe essere desiderabile farlo); e anche nel caso in cui sia possibile, i risultati ai quali si giunge sono sempre (anche se in misura ogni volta diversa) discutibili, parziali e controversi. Ciò dipende essenzialmente dalle molteplici, e mai del tutto eliminabili, difficoltà cognitive che il valutatore si trova dinanzi nella costruzione del suo percorso d'analisi. Chi ignora, o non chiarisce fino in fondo, la natura e la portata di tali difficoltà, rende di solito un pessimo servizio al suo committente.

Il presente volume, allo scopo di mostrare come tali difficoltà possano essere messe a fuoco ed affrontate, presenta una rassegna di casi in cui pubbliche amministrazioni italiane hanno condotto o commissionato una valutazione degli effetti di una specifica politica pubblica. Tale rassegna è preceduta da un'esposizione, condotta a fini divulgativi con l'impiego di un linguaggio non tecnico, dell'approccio metodologico a cui queste esperienze fanno principalmente riferimento: il cosiddetto "approccio controfattuale".

Lo scopo di questa operazione non è di illustrare come in generale si debba "fare valutazione" nella pubblica amministrazione. Molto più modestamente, il nostro scopo è quello di offrire un punto di riferimento concettuale a quelle amministrazioni che si interrogano su *quali ricadute* le politiche da esse attuate abbiano sul problema collettivo che le motiva, e di quale entità siano tali ricadute. Sottolineare fin dall'inizio la natura circoscritta di questa operazione è fondamentale per prevenire possibili equivoci: il termine "valutazione" è infatti usato all'interno della pubblica amministrazione con una molteplicità di significati e di intenti conoscitivi. Significati e intenti tutti perfettamente legittimi, ma nella gran parte dei casi molto distanti dal tema specifico di questo volume.

Quella che qui definiamo come “valutazione degli effetti di una politica, condotta con un approccio controfattuale” è infatti un tipo di valutazione particolare, almeno per due motivi. Innanzitutto, perché pone al centro dell’attenzione il problema dell’*attribuzione di causalità*, cioè dell’identificazione del *contributo netto* dato da un’azione pubblica ai cambiamenti eventualmente osservati tra i destinatari dell’azione. In questo approccio l’analisi ha dunque lo scopo di “isolare” tale contributo dall’influenza di altri fattori che insistono sullo stesso fenomeno.

In secondo luogo, la peculiarità di questo approccio sta nel tentativo di ancorare la misurazione di tale contributo netto ai dati empirici ottenibili da fonti esistenti o raccolti *ad hoc*, riconoscendo perciò esplicitamente i limiti alla valutabilità delle politiche connessi alla disponibilità di dati. Questa scelta costringe il valutatore a mettere sempre in chiaro *quali effetti sono stimabili con i dati disponibili mediante l’applicazione di quali metodi*. Tale doveroso ancoraggio alla base empirica disponibile, coniugato con l’obiettivo conoscitivo di identificare legami di causa-effetto tra azione e risultati, rende questo tipo di valutazione assai diversa dalle molte altre forme presenti nella pubblica amministrazione.

La struttura del volume

Il volume è strutturato come segue. Il capitolo 1 è dedicato ad una lettura critica dei principali utilizzi del termine valutazione nella pubblica amministrazione italiana. Lo scopo è di mettere in evidenza le peculiari finalità conoscitive dell’approccio valutativo che sarà poi oggetto d’approfondimento nel resto del volume e di collocarlo in seno al più ampio panorama di approcci e utilizzi della valutazione.

Il capitolo 2 è volto ad illustrare l’essenza di quello che abbiamo qui chiamato “valutazione degli effetti delle politiche pubbliche mediante un approccio controfattuale”. In particolare, viene offerta una definizione rigorosa del concetto stesso di “effetto” e di quello di “situazione controfattuale”, che come vedremo costituisce un elemento centrale per comprendere la logica su cui si basa questo approccio.

I capitoli successivi (dal 3 al 6) costituiscono il cuore della sezione metodologica e spiegano più in dettaglio diversi metodi attraverso i quali i dati empirici disponibili

possono essere elaborati per ricavare stime credibili degli effetti delle politiche. I metodi descritti in questi capitoli trovano poi un'ulteriore presentazione negli studi di caso analizzati nella sezione applicativa del lavoro (capitoli dal 7 al 12). Attraverso la ricostruzione di alcune esperienze di valutazione, condotte negli ultimi anni in Italia, si cercherà di chiarirne i risvolti metodologici e i principali risultati.

L'ultimo capitolo presenta alcune considerazioni conclusive, sotto forma di raccomandazioni alle pubbliche amministrazioni che intendano investire risorse in questa forma di conoscenza.

Il contenuto di questo volume è frutto delle riflessioni maturate negli ultimi anni dallo staff di ProgettoValutazione². La sezione metodologica trae ispirazione dal materiale didattico utilizzato nell'ambito del Master in Analisi delle Politiche Pubbliche di Torino e presso altri corsi universitari in metodi di valutazione delle politiche. Gli autori ringraziano tutti coloro che, a vario titolo, hanno contribuito alla stesura del lavoro.

² ProgettoValutazione è il braccio operativo dell'Associazione per lo Sviluppo della Valutazione e l'Analisi delle Politiche Pubbliche (ASVAPP), un'organizzazione non-profit nata nel 1998 con lo scopo di contribuire alla diffusione della valutazione delle politiche in Italia. Sono soci dell'ASVAPP la Compagnia di San Paolo di Torino, la Fondazione Cassa di Risparmio di Torino, il CSI-Piemonte e l'IRES-Piemonte.

PRIMA PARTE

METODI PER LA VALUTAZIONE DEGLI EFFETTI DELLE POLITICHE

CAPITOLO 1

CINQUE MODI DI INTERPRETARE LA “VALUTAZIONE”

di Alberto Martini

Definizioni e classificazioni non mancano nella letteratura in materia di valutazione. Anche negli Stati Uniti, dove la valutazione è nata e conta ormai su un’esperienza quarantennale di lavori e di studi, ogni manuale di nuova pubblicazione si apre con il tentativo di fornire una risposta originale alla più classica delle domande: “*what is evaluation?*”.

Questa necessità di proporre sempre nuove e più puntuali definizioni dipende in parte dalla natura multidisciplinare di questa pratica professionale, nella quale coesistono persone con retroterra disciplinari e culturali molto diversi tra loro: sociologi, economisti, statistici, politologi, psicologi, aziendalisti ed urbanisti. In parte dipende anche dal fatto che molti dei termini impiegati in questo campo (tra cui *in primis* il termine valutazione) hanno un significato molto esteso, che muta a seconda dei contesti e delle situazioni: questo provoca a volte una certa difficoltà di comunicazione e di comprensione reciproca tra gli stessi addetti ai lavori.

Non rientra nei nostri scopi fornire una rassegna esaustiva di tali definizioni e classificazioni, per cui, nello sviluppare questo argomento di natura epistemologica, opereremo in modo molto selettivo e certamente parziale. In queste prime pagine ci accontentiamo infatti di offrire un sintetico quadro d’insieme, all’interno del quale potremo collocare la trattazione del tema, molto più circoscritto, che ci siamo proposti.

Cominceremo riportando le definizioni molto generali di valutazione date da tre autori italiani all’interno di altrettante accreditate pubblicazioni sul tema. Successivamente presenteremo una classificazione di quelli che, a nostro parere, sono gli utilizzi principali del termine “valutazione” all’interno della pubblica amministrazione italiana.

1.1 Tre definizioni di valutazione a confronto

Vediamo, quindi, le tre definizioni di valutazione dalle quali abbiamo deciso di iniziare la nostra trattazione:

*“La valutazione è un’attività cognitiva rivolta a fornire un giudizio su di un’azione, intenzionalmente svolta o che si intende svolgere, destinata a produrre effetti esterni, che si fonda su un’attività di ricerca delle scienze sociali e che segue procedure rigorose e codificabili”*³

*“Valutare significa analizzare se un’azione intrapresa per uno scopo corrispondente ad un interesse collettivo abbia ottenuto gli effetti desiderati o altri, ed esprimere un giudizio sullo scostamento che normalmente si verifica”*⁴

*“La valutazione è principalmente (...) un’attività di ricerca sociale applicata (...) avente come scopo la riduzione della complessità decisionale, attraverso l’analisi degli effetti diretti e indiretti, attesi e non attesi, voluti e non voluti, dell’azione, compresi quelli non riconducibili ad aspetti materiali (...)”*⁵

Non sorprendentemente, tutte e tre le definizioni sottolineano la natura prettamente analitica della valutazione, pur usando termini tra loro diversi (“attività cognitiva”, “attività di ricerca delle scienze sociali”, “valutare significa analizzare”, “attività di ricerca sociale applicata”). Si presuppone cioè che per valutare sia necessario l’impiego (non arbitrario) di appropriati metodi e strumenti d’analisi, mutuati dall’esperienza delle scienze sociali.

Meno scontato è il fatto che tutti e tre gli autori pongano gli “effetti dell’azione” al centro della valutazione, seppure con una diversa enfasi e con diverse sottolineature di cosa sia da intendersi per “effetto” da valutare: Mauro Palumbo fa riferimento agli “effetti esterni” come scopo ultimo dell’azione da valutare; Nicoletta Stame distingue

³ Palumbo M., *Il processo di valutazione*, FrancoAngeli, 2001, p. 59.

⁴ Stame N., *L’esperienza della valutazione*, SEAM, Roma, 1998, p. 9.

⁵ Bezzi C., *Il disegno della ricerca valutativa*, FrancoAngeli, Milano, 2003, p. 60.

tra effetti desiderati ed “altri”; mentre Claudio Bezzi introduce ben quattro distinzioni tra differenti tipi di effetti, diretti e indiretti, attesi e non attesi, voluti e non voluti, riconducibili o meno ad aspetti materiali.

I tre autori concordano anche sul fatto che l’oggetto del giudizio dato dal valutatore sia un’azione, pur non usando mai il termine politica pubblica: Mauro Palumbo sottolinea la possibilità che l’azione da valutare non sia ancora avvenuta (“*un’azione, intenzionalmente svolta o che si intende svolgere*”); Nicoletta Stame si avvicina di più alla nozione di politica pubblica (“*un’azione intrapresa per uno scopo corrispondente ad un interesse collettivo*”), che presenteremo nel prossimo capitolo; Claudio Bezzi preferisce invece rimanere sul generale (“*l’analisi degli effetti ... dell’azione*”).

Volendo trovare quindi un minimo comune denominatore tra le tre definizioni appena proposte, potremmo dire che per valutazione si intende “*un’attività analitica che utilizza i metodi delle scienze sociali per giudicare gli effetti prodotti da un’azione pubblica*”.

Come vedremo a partire dal prossimo capitolo, tale definizione sintetica si adatta quasi alla perfezione all’accezione ristretta di valutazione utilizzata nel resto di questo volume, mentre non rende giustizia dell’ampiezza di significati con cui questo termine viene utilizzato nella pubblica amministrazione italiana. Abbandonando quindi il tentativo (a nostro parere vano) di trovare un’unica definizione, prendiamo in esame quelli che a nostro parere sono i cinque *usi prevalenti* di questo termine.

Per correttezza avvertiamo il lettore che non troverà traccia della nostra classificazione nella letteratura valutativa italiana o europea, dove le classificazioni tendono a seguire altri schemi concettuali. Ad esempio, con riferimento alla *scansione temporale* dell’attività valutativa, si usa sovente distinguere tra valutazione *ex-ante*, *in itinere* ed *ex-post*. Oppure con riferimento ai *criteri di giudizio utilizzati*, si distingue tra valutazione della coerenza, della rilevanza, dell’efficacia, dell’efficienza, della sostenibilità, ecc. O ancora, con riferimento al *grado di autonomia e di distanza* di coloro che sono chiamati a pronunciare il giudizio, si parla di valutazione interna, esterna o indipendente. Rimandiamo ai tre testi citati per l’approfondimento di queste (ed altre) distinzioni.

La nostra classificazione si basa sul tentativo di “distillare” le finalità conoscitive *prevalenti* nel variegato insieme di attività analitiche chiamate in Italia “valutazione”. Quindi si tratta di stilizzazioni (a volte molto estreme) di quello che *a noi* pare essere l’essenza della pratica valutativa nel nostro Paese (ma le cose probabilmente non cambierebbero poi molto se si estendesse lo sguardo all’esperienza europea).

Nella nostra lettura, la valutazione viene impiegata per soddisfare *cinque grandi tipologie di bisogni conoscitivi*. Essa viene infatti utilizzata: (i) come strumento per allocare le risorse agli usi più *meritevoli*; (ii) come strumento di *controllo della performance* nell’ambito di organizzazioni; (iii) come veicolo per *rendere conto* delle realizzazioni effettuate in un certo ambito di azione pubblica; (iv) come strumento di analisi critica dei processi di *attuazione* di una politica; e infine (v) come stima degli *effetti* prodotti da una politica. Esaminiamo ora in dettaglio ciascuna di queste finalità conoscitive, cercando di metterne in evidenza le caratteristiche essenziali.

La valutazione come strumento per allocare le risorse agli usi più meritevoli

In questo caso la finalità conoscitiva che la valutazione si pone è quella di dare un giudizio comparativo su diversi soggetti al fine di individuare i più *meritevoli*.

Questa stilizzazione richiama una forma quasi primordiale di valutazione, quella che tutti incontriamo, fin dai primi anni di vita, nella valutazione scolastica, cioè nel “dare e ricevere voti”. Nel corso dell’attività della pubblica amministrazione si manifesta spesso l’esigenza di “dare i voti” a una molteplicità di oggetti (persone, azioni, progetti, eventi), per stabilire quali siano i migliori, solitamente allo scopo di meglio allocare risorse scarse, ma anche di stimolare comportamenti di emulazione premiando “chi se lo merita”.

Le forme concrete di valutazione che si richiamano all’esigenza conoscitiva di “individuare i migliori” vanno dalle procedure di valutazione (*alias* selezione) dei progetti, alla valutazione del personale, ai metodi per valutare la qualità della ricerca scientifica. Dietro a questo tipo di valutazione vi è sempre la necessità di *operare una scelta* tra opzioni alternative. Sia che la scelta riguardi dirigenti pubblici ai quali riconoscere la retribuzione accessoria che ne premi la maggior produttività; sia che riguardi la decisione di investire risorse nella costruzione di un’opera pubblica piuttosto

che un'altra; sia che riguardi l'allocazione di finanziamenti tra differenti progetti di ricerca posti in concorrenza tra di loro.

In tutti questi casi la conoscenza che si produce è un giudizio comparativo su qualche tipo di s/oggetto, o di azione, o di proposta progettuale, che vengono in questo modo "valutati": questo tipo di valutazione quindi è fondamentalmente un modo per formalizzare e proceduralizzare (e in conclusione legittimare) il processo attraverso il quale sono riconosciuti dei meriti su base comparativa.

La valutazione in questo caso richiede tre passaggi fondamentali, che presi insieme la distinguono dalle altre forme di valutazione:

- a. la determinazione dei *criteri*, cioè le dimensioni lungo le quali gli N s/oggetti devono essere giudicati: in questa fase vanno anche determinati il tipo di punteggio e il *peso* da assegnare a ciascun criterio;
- b. l'esame degli N s/oggetti, condotto da *esperti* che esprimono un *giudizio* (solitamente sotto forma di punteggio o di metrica equivalente) su ciascun s/oggetto secondo ciascun criterio;
- c. l'*aggregazione* dei giudizi così espressi usando una esplicita procedura, che può andare dalla semplice somma ponderata dei punteggi normalizzati all'uso di sofisticate tecniche di analisi multicriteri. I giudizi aggregati consentono così di identificare i "migliori" e di escludere i peggiori, e in taluni casi di formulare "graduatorie".

Al di là di questi passaggi comuni, le forme concrete che assume questo tipo di valutazione variano molto da caso a caso. In particolare, il contenuto del secondo passaggio, che richiede il ricorso ad esperti, dipende molto dalla natura degli strumenti analitici che tali esperti useranno per emettere il loro giudizio: si va dalla mera espressione di giudizi soggettivi, legittimati dal prestigio professionale che accompagna l'esperto e dalle conoscenze specialistiche che lo caratterizzano, fino all'uso di sofisticate e complesse tecniche analitiche, quali l'analisi costi-benefici, quando, ad esempio, si debba giudicare il merito relativo di progetti alternativi di carattere infrastrutturale o logistico.

Nonostante questa eterogeneità di strumenti, il processo di valutazione è guidato sempre dalla medesima finalità conoscitiva: esprimere giudizi comparativi del tipo “A è migliore di B” secondo un insieme dato di criteri.

La valutazione come controllo della performance di organizzazioni

L'esigenza conoscitiva prevalente per questa forma di valutazione è capire “quanto bene” un'organizzazione pubblica (o una parte o un aspetto di essa) stia funzionando. Si parla a questo proposito di “analisi della performance”, intendendo con questo ultimo termine l'insieme delle caratteristiche desiderabili dell'operare di un'organizzazione (i costi, la qualità delle prestazioni, i volumi di attività).

Questa forma di valutazione consiste, nella sua essenza estrema, nel confronto tra *misurazioni* di ciò che l'organizzazione ha prodotto (e/o di come lo ha prodotto) e i corrispondenti *valori desiderati*, che rappresentano ciò che l'organizzazione *dovrebbe* produrre (e/o come dovrebbe produrlo). Si ricorre spesso al termine “indicatore” per riferirsi alle singole dimensioni della performance da valutare mediante il confronto tra valori osservati e valori desiderati.

Le fonti da cui possono provenire i valori desiderati, o standard, variano molto secondo il contesto. Ad un estremo possono essere semplicemente i valori osservati in passato per lo stesso indicatore, usati come termine di paragone per verificare l'avvenuto miglioramento o per rilevare un eventuale peggioramento; possono essere valori-obiettivo (*target*) fissati dal *management* per stimolare direttamente le diverse unità o parti dell'organizzazione a raggiungere prestazioni migliori; possono essere standard qualitativi fissati, a livello centrale, da organismi di natura professionale o da autorità pubbliche; possono infine essere valori osservati per gli stessi indicatori presso organizzazioni simili, che servono da *benchmark*.

Le fasi che caratterizzano in generale questa tipologia di valutazione sono cinque:

- a. l'individuazione delle *dimensioni* e aspetti dell'agire dell'organizzazione che si ritiene utile tenere sotto controllo, sotto l'assunto che siano suscettibili di miglioramento;

- b. la definizione degli *indicatori* che servono per cogliere la performance dell'organizzazione lungo ciascuna dimensione e aspetto;
- c. l'individuazione dei *valori desiderati (standard)* appropriati per ciascun indicatore, cioè dei punti di riferimento che consentano di separare una performance buona da una meno buona;
- d. la raccolta dei *dati* sulla performance dell'organizzazione, per ricavare il valore di ciascun indicatore e confrontarlo con il rispettivo valore desiderato;
- e. l'interpretazione delle *deviazioni* della performance osservata dagli standard prefissati, in modo da poter considerare quali *azioni correttive* intraprendere per correggere le performance inadeguate.

Numerosi approfondimenti sarebbero necessari per superare l'impressione di superficialità che probabilmente questa generica lista di passaggi ingenera nel lettore, ma tali approfondimenti ci porterebbero decisamente fuori strada. Qui ci preme sottolineare come sia il terzo passaggio, *l'individuazione dei valori desiderati*, a costituire la vera sfida che deve affrontare chi si accinge ad operare questo tipo di valutazione. Una sfida, ci sia consentita una nota critica, non sempre raccolta nella pratica valutativa che in Italia, direttamente o indirettamente, si richiama alla valutazione di performance e che generalmente tende a concentrarsi sul secondo passaggio, relativo alla *definizione degli indicatori*⁶.

Esempi di concrete attività che si richiamano a questo schema concettuale, pur nelle loro diversità, sono il *controllo di gestione*, la *certificazione di qualità*, ormai diffusa ampiamente anche in ambito pubblico, e alcune forme di *accreditamento*. Anche strumenti quali le indagini di *soddisfazione dell'utenza* possono essere ricondotte a questa tipologia di valutazione: in questo caso il confronto tra valori osservati e valori desiderati è effettuato direttamente dall'utente, confrontando il servizio che ha ricevuto

⁶ Per un approfondimento si veda Martini A., Sisti M., "Indicatori o analisi di performance? Implicazioni dell'esperienza statunitense di *performance measurement*", *Rivista trimestrale di Scienza dell'Amministrazione*, 2002.

(e la qualità che ha percepito) con le proprie aspettative, che in questo particolare ambito applicativo fungono da standard.

La valutazione come veicolo per rendere conto delle realizzazioni effettuate

Alla base di questa terza forma di valutazione vi è l'esigenza di descrivere e misurare le *realizzazioni* prodotte da specifiche forme di azione pubblica, al fine di poterne *rendere conto* agli *stakeholder* più rilevanti. Questa forma di valutazione confina direttamente con la precedente, da cui però si distingue nettamente per almeno tre ragioni.

In primo luogo, l'analisi di performance presta (o dovrebbe prestare) grande attenzione all'individuazione degli scostamenti da standard prefissati; l'idea è di cogliere, attraverso rilevazioni e confronti molto puntuali, operati su singole dimensioni operative, se e in che misura l'organizzazione stia lavorando bene e quanto essa sia distante da una performance ritenuta desiderabile. Nella valutazione come "veicolo per rendere conto" questa impostazione basata fortemente sulla comparazione è meno presente, mentre permane una maggiore enfasi descrittiva. Piuttosto che a misure specifiche e di dettaglio, viene dato molto più rilievo all'osservazione di grandezze aggregate, che offrono un'immagine d'insieme dello sforzo compiuto da quella amministrazione per farsi carico dei problemi collettivi che ne motivano l'esistenza.

In secondo luogo, l'analisi di performance è utilizzata prevalentemente per misurare e valutare le prestazioni di singole unità organizzative. Al contrario, l'ambito di utilizzo privilegiato della valutazione come strumento di rendicontazione non è quello della singola unità organizzativa, bensì di organizzazioni complesse e multifunzionali, oppure di grandi programmi di intervento pubblico.

Infine, la differenza più rilevante riguarda la ricaduta decisionale dei due tipi di valutazione. Se nel primo caso i risultati dell'analisi sono utilizzati per assumere decisioni di natura gestionale e, quindi, per intervenire direttamente sulle realtà organizzative che mostrano performance carenti o inadeguate, le informazioni prodotte dal secondo tipo di valutazione hanno una ricaduta decisionale molto più sfumata ed assumono anche, se non soprattutto, una valenza "comunicativa". Lo scopo di tali informazioni consiste infatti nell'offrire *a soggetti esterni* un'idea complessiva delle strategie d'intervento adottate dalle amministrazioni; delle motivazioni che stanno

dietro a tali strategie; delle attività realizzate durante il periodo cui fa riferimento l'analisi; dei risultati che sono stati conseguiti al termine di quel periodo. I termini chiave in questo caso sono *trasparenza*, *responsabilità* e *legittimazione*: attraverso questa forma di comunicazione le amministrazioni pubbliche cercano di essere più trasparenti e rispondono ai loro portatori di interesse delle scelte assunte e dei comportamenti adottati, al fine di recuperare, o consolidare, la loro legittimazione ad operare.

A nostro parere gli esempi più rilevanti che cadono in questa categoria sono, da un lato, il complesso processo di “monitoraggio e valutazione” che accompagna l'erogazione dei Fondi Strutturali europei, e dall'altro il vasto movimento che cerca di favorire una maggior “*accountability*” delle pubbliche amministrazioni italiane, in particolar modo nei confronti dei cittadini, mediante strumenti quali il *bilancio di mandato* e il *bilancio sociale*⁷.

A molti potrà sembrare riduttivo caratterizzare sia la valutazione dei Fondi Strutturali, sia i bilanci sociali/di mandato come semplici strumenti per “rendere conto di realizzazioni”. Siamo consapevoli del fatto che entrambe queste forme di valutazione reclamano finalità conoscitive ben più ambiziose, che sconfinano ampiamente in quelle che noi assegniamo alle due tipologie illustrate successivamente. Spesso infatti si afferma, da un lato, di voler individuare le criticità dei processi di attuazione delle politiche al fine di correggerle in corso d'opera; dall'altro, si fa riferimento all'esigenza di valutare gli effetti, il contributo, il valore aggiunto o la ricaduta delle politiche sui problemi collettivi che le motivano.

Non intendendo entrare in polemica né con il potente apparato che gestisce e valuta i Fondi Strutturali, né con il variegato movimento che ha dato vita ai bilanci sociali e di mandato, ci limiteremo pertanto a ribadire la nostra impressione: che la peculiarità (e l'utilità) di questi strumenti stia soprattutto nel *raccontare in maniera sintetica* (con tabelle, grafici o semplicemente a parole) ai propri finanziatori da un lato, e ai propri contribuenti/utenti/elettori dall'altro, cosa ne è stato delle risorse messe a disposizione

⁷ Per approfondimenti si veda de Magistris V., Gioioso G. (a cura di), *Nuovi profili di accountability nelle P.A. italiane: Teoria e strumenti*, Collana Quaderni Formez, n. 40; Tanese A. (a cura di), *Rendere conto ai cittadini – Il bilancio sociale nelle amministrazioni pubbliche*, Edizioni Scientifiche Italiane, Roma, 2004.

di quell'Autorità di Gestione, piuttosto che di quel Sindaco o di quella Giunta regionale. Senza la pretesa che questo *racconto* riesca davvero a individuare criticità da un lato ed effetti dall'altro.

La valutazione come analisi critica dei processi di attuazione delle politiche

In questo caso l'esigenza conoscitiva alla base della valutazione nasce dalla consapevolezza che i processi attuativi delle politiche (e di qualsiasi azione pubblica) sono imprevedibili e poco controllabili dall'alto. Qui il modo di procedere del valutatore è quello di ricostruire, prevalentemente mediante strumenti di indagine qualitativa, i processi di attuazione delle politiche per capire come e per quali ragioni si discostano dal disegno originario. Lo scopo quindi non è giudicare la performance, né rendere conto di cosa si è fatto: lo scopo del valutatore è piuttosto quello di "entrare nei processi" con cui si dà concreta attuazione ad una politica pubblica, interloquendo con gli attori di questi processi e tentando di mettere in luce le criticità, le incongruenze, gli ostacoli che si frappongono sulla strada della sua attuazione.

Ogni politica pubblica deve infatti passare dallo stato di dettato legislativo a quello di pratica amministrativa, dallo stato in cui vengono enunciate buone intenzioni a quello in cui le risorse disponibili sono effettivamente utilizzate, servizi reali sono davvero erogati, obblighi o divieti sono effettivamente imposti. Un'espressione inglese molto efficace descrive questo come il passaggio *from policy fiction to policy facts*. La *fiction* qui è il disegno astratto della politica prodotto dai *policy maker*, siano essi legislatori, membri degli esecutivi o livelli apicali delle strutture pubbliche. Per diventare *policy facts*, tale disegno astratto deve fare i conti con le volontà, gli interessi e le risorse di tutti quegli attori che devono incorporare il dettato e gli obiettivi della politica pubblica nella loro attività quotidiana (all'interno o all'esterno della pubblica amministrazione).

Non raramente il destino di una politica pubblica è quello di restare "lettera morta" o di essere attuata in maniera profondamente diversa rispetto allo spirito del suo disegno originario. Si parla in questo caso di "*implementation deficit*", un fenomeno ricorrente che si incontra sotto tutte le latitudini e in tutti i sistemi politici. Una vasta letteratura (in lingua inglese soprattutto) si occupa di studiare, documentare e spiegare le difficoltà che si incontrano sul cammino della *policy implementation*, ovvero della concreta

realizzazione di una politica pubblica. Questa letteratura è conosciuta con diversi nomi, tra cui *implementation research* è forse il più comune.

L'*implementation research* “cerca di descrivere e spiegare quanto la concreta attuazione della politica si avvicina davvero alle intenzioni originali dei policy maker; perché esistono delle discrepanze, o vi sono state delle modificazioni in corso d’opera; e quali conseguenze, se ve ne sono state, certe modalità attuative hanno prodotto sulle successive decisioni di policy. Essa spesso si concentra sui dettagli del processo realizzativo: le dinamiche e le relazioni esistenti tra le diverse fasi d’attuazione, il contesto organizzativo all’interno del quale la politica viene implementata, le modalità di selezione e di trattamento dei destinatari/clienti dell’intervento pubblico, la conduzione e la gestione da parte dei funzionari delle singole attività operative”⁸.

Per spiegare l’oggetto di questo tipo di valutazione si ricorre spesso all’immagine metaforica della *black box*. Il processo di trasformazione delle idee di *policy* in azioni e pratiche amministrative concrete è una sorta di “scatola nera”, che l’analista deve aprire, portando alla luce tutti gli elementi in essa contenuti. Solo così facendo si riesce a capire *perché* e *come* le “grandi speranze” affidate alle politiche pubbliche diventano delusioni, se non addirittura completi fallimenti⁹.

La valutazione come stima degli effetti delle politiche

La finalità conoscitiva di questa ultima forma di valutazione è quella di capire “*what works*”, cioè quali strumenti di intervento pubblico sono efficaci nel determinare i cambiamenti desiderati nei fenomeni collettivi ritenuti problematici.

Pur trattandosi di una forma di valutazione eminentemente *ex post*, nel senso che esamina politiche realizzate in passato per determinare gli effetti che esse hanno prodotto, in termini di utilizzo essa è ovviamente indirizzata al futuro. Nonostante la sua natura retrospettiva, essa è fondamentalmente prospettica: la sua ambizione di fondo è di riorientare il disegno delle politiche verso quelle forme di intervento che siano risultate più efficaci in passato. Ad una così grande ambizione fa da contrappeso una

⁸ Holcomb P., Nightingale D., *Theoretical and Substantive Dimensions in Process and Implementation Studies of Welfare*, The Urban Institute, Washington, 1999.

⁹ Pressman J., Wildavsky A., *Implementation - How great expectations in Washington are dashed in Oakland*, University of California Press, Berkeley, 1973.

difficoltà altrettanto grande: quella di isolare l'effetto delle politiche dalla miriade di altri fattori che influiscono sui fenomeni collettivi.

Per attribuire “meriti” (o demeriti) ad una politica pubblica occorre infatti riuscire ad identificare quei cambiamenti che si sarebbero prodotti comunque anche in assenza di intervento. Ad esempio, prima di attribuire il merito del calo di immigrazione clandestina ad un certo provvedimento legislativo (o di gridare al fallimento del medesimo provvedimento nel caso in cui si osservi un aumento dell'immigrazione) occorre capire cosa sarebbe presumibilmente successo all'immigrazione clandestina in assenza di provvedimento: sarebbe calata comunque? Sarebbe aumentata ancora di più? L'approccio controfattuale, oggetto del prossimo capitolo, pone al centro dell'attenzione esattamente questo tipo di questioni metodologiche: per stabilire che effetto ha avuto una politica pubblica, occorre in qualche modo ricostruire ciò che sarebbe successo in assenza di intervento. A questa questione volgiamo ora l'attenzione.

CAPITOLO 2

L'APPROCCIO CONTROFATTUALE ALLA VALUTAZIONE DEGLI EFFETTI DELLE POLITICHE

di Alberto Martini e Marco Sisti

In questo volume parliamo di politica pubblica assegnandole il significato di “risposta ad un particolare problema collettivo”. Tuttavia, il termine “politica pubblica” è utilizzato nella letteratura specialistica e nel parlar comune con una varietà di significati. Anche lasciando da parte la “Politica” e concentrandosi esclusivamente sulla politica intesa come *policy*, restano comunque possibili diverse accezioni e sfumature¹⁰. Come afferma una nota studiosa di questo tema, “*lo studio delle politiche pubbliche è caratterizzato da una curiosa anomalia: oltre che dalle metodologie, i ricercatori sono divisi dalla definizione stessa dell’oggetto dell’indagine*”¹¹.

Una rassegna del dibattito epistemologico relativo all’analisi delle politiche pubbliche andrebbe al di là degli scopi di questo volume¹². Ciò che ci preme sottolineare è il fatto che le politiche pubbliche “*non sono fenomeni che si autodefiniscono, bensì categorie analitiche, i cui contenuti sono definiti dall’analista più che dallo stesso policy maker*”¹³. Si tratta perciò di costrutti analitici, di interpretazioni della realtà operate dallo studioso che, almeno in una certa misura, “fabbrica” il proprio oggetto d’indagine a partire da una serie di elementi osservabili (leggi, comportamenti amministrativi,

¹⁰ In italiano esiste un solo termine, *politica*, per definire due concetti assai diversi tra loro che gli anglosassoni tengono separati anche sul piano terminologico grazie all’uso dei vocaboli *politics* e *policy*. Per evitare fraintendimenti chiariamo velocemente quali sono le differenze tra queste due accezioni di “politica”. La politica nel senso di *politics* può essere definita come “la competizione esistente tra gruppi e/o tra individui per la conquista e il mantenimento delle risorse necessarie all’esercizio del potere”. Chi studia la politica sotto questa prospettiva è portato a considerare quelle componenti della fenomenologia della politica correlate essenzialmente con la dimensione del potere: il suo perseguimento, la sua legittimazione, la sua gestione ed organizzazione. La politica nel senso di *policy*, spesso tradotta con l’espressione “politica pubblica”, può essere definita come “l’insieme di azioni messe in atto da attori pubblici al fine di affrontare un problema collettivo”.

¹¹ Regonini G., *Lo studio delle politiche pubbliche*, in Panebianco A. (a cura di), *L’analisi della politica*, il Mulino, Bologna, 1989, pp. 491-516.

¹² Per approfondire questo tema consultare le voci “*Politica e politiche*” e “*Politica pubblica*” contenute in Capano G., Giuliani M. (a cura di), *Dizionario di politiche pubbliche*, La Nuova Italia Scientifica, Roma, 1996, pp. 306-323.

¹³ Hecl H., “Review Article: Policy Analysis”, in *British Journal of Political Science*, vol.2, n.1, 1972 pp. 83-108.

atteggiamenti dei destinatari,...), nonché dalle proprie sensibilità e competenze metodologiche. Le caratteristiche e i contenuti di tale oggetto d'indagine possono perciò variare a seconda dell'approccio analitico adottato. Da queste considerazioni discende l'esigenza, ogni qual volta ci si accinga a parlare di politiche pubbliche, di offrire una propria definizione di tale concetto, e di non dare per scontato che ne esista una universalmente condivisa.

2.1 La metafora della terapia

Ai fini della nostra trattazione è opportuno dare del concetto di politica pubblica una lettura piuttosto restrittiva, pur sapendo che non si tratta della sola interpretazione possibile, né probabilmente della più esaustiva. Definiamo la (singola) politica pubblica come lo *sforzo consapevole, condotto da parte di attori pubblici, di “curare” un certo problema collettivo*.

Per “problema collettivo” intendiamo una situazione socialmente percepita come insoddisfacente e quindi meritevole di cambiamento (ed eventualmente di un intervento pubblico). La lista di esempi di situazioni considerate insoddisfacenti socialmente (cioè da un numero sufficientemente ampio di membri della collettività) è, purtroppo, assai lunga. Dall'inquinamento nei centri urbani alla diffusione delle droghe pesanti, dal bisogno di assistenza degli anziani soli alla scarsa competitività delle piccole e medie imprese, dalla sicurezza sui luoghi di lavoro all'emarginazione dei disabili, i problemi per cui si invoca l'intervento pubblico sotto forma di destinazione di risorse e/o di imposizione di regole sono potenzialmente infiniti. Quando ad un problema fa seguito un intervento consapevole per affrontarlo (anche solo marginalmente o simbolicamente), noi parliamo di “politica pubblica”.

Quello che adottiamo in questo lavoro è quindi un modo particolare di *leggere* le politiche pubbliche, motivato da un particolare obiettivo cognitivo: valutarne il successo, e in particolare misurarne gli effetti, in modo rigoroso, mentre non ci interessa qui analizzare i processi sociali e decisionali che danno vita a quelle stesse politiche.

La nostra lettura interpreta ciascuna politica pubblica come una “terapia”, cioè come un trattamento somministrato a certi soggetti per curare (o in alcuni casi prevenire) un

determinato problema. Si tratta ovviamente di una metafora, ma tale metafora è utile perché ci consente di mettere a fuoco gli aspetti di una politica pubblica che più sono rilevanti per valutarne il successo, mentre ci permette di lasciare in ombra altri aspetti meno importanti ai nostri fini. Questi aspetti sono viceversa essenziali per *altri* scopi: per capire, ad esempio, come la politica pubblica è stata concettualizzata a partire dalle percezioni e dai punti di vista dei diversi attori proponenti; come e perché è entrata nell'agenda dei *policy maker*; oppure in che modo ha guadagnato il consenso.

La metafora della terapia ci aiuta a pensare al successo della politica pubblica come capacità di alleviare una particolare patologia che si manifesta in determinati pazienti. La politica pubblica da valutare è vista come insieme di azioni dirette ad un determinato segmento della popolazione per modificarne in un senso desiderato uno o più *comportamenti o condizioni* (quei comportamenti o condizioni connotati appunto come “problema”). Il “paziente” può essere la collettività nel suo complesso, oppure un insieme di aree geografiche, o determinati gruppi o categorie di soggetti, quali individui, famiglie, imprese, enti territoriali. Riportiamo nella Tabella 2.1 alcuni esempi di politica pubblica letta come terapia. Indichiamo quindi il problema, la terapia e una possibile misura del successo della politica.

Tabella 2.1: Alcuni esempi di lettura di una politica pubblica come terapia

La patologia (il problema collettivo)	La terapia (la politica pubblica)	Una possibile misura del successo della politica
Disoccupazione in seguito a licenziamento collettivo	Sgravi fiscali per chi assume lavoratori licenziati	Riduzione della durata della disoccupazione
Incidenti stradali causati da comportamenti scorretti alla guida	Introduzione della patente a punti	Riduzione del numero di decessi in seguito a incidente stradale
Infortuni sul lavoro	Norme di sicurezza sul luogo di lavoro	Diminuzione del tasso di infortuni
Emarginazione degli anziani non autosufficienti	Servizio di assistenza domiciliare	Riduzione della frequenza dei ricoveri in casa di riposo
Disoccupazione tra i giovani senza titolo di studio	Formazione professionale	Aumento della probabilità di trovare un lavoro qualificato
Difficoltà di accesso al credito da parte dei giovani che vogliono “mettersi in proprio”	Credito agevolato per l'imprenditoria giovanile	Aumento delle imprese create da giovani

Una politica pubblica può ovviamente servirsi di una molteplicità di strumenti “terapeutici”: può utilizzare un trasferimento mirato di risorse, un’imposizione di obblighi o divieti, un’erogazione di servizi a determinati utenti, un incentivo che vuole favorire o inibire un particolare comportamento, una campagna di comunicazione che tenta di modificare un certo atteggiamento. E naturalmente può essere anche un *mix* più o meno complesso di questi diversi strumenti. Chi si propone di valutare una politica pubblica deve essere in grado di decifrare tale complessità e di riconoscere la particolare terapia che sta alla base della politica nei suoi contenuti essenziali, in modo da stabilire *che tipo* di trattamento è stato effettivamente somministrato, secondo *quale logica* quel trattamento dovrebbe indurre un certo tipo di cambiamento, *chi* tra i potenziali beneficiari lo ha ricevuto ed eventualmente *con quale intensità*.

Le azioni in cui la politica consiste possono essere indirizzate direttamente ai soggetti che manifestano il comportamento o la condizione da modificare (ad esempio, l’erogazione di servizi di formazione professionale direttamente agli individui senza lavoro), oppure ad altri soggetti la cui risposta ci si attende influisca positivamente sui “portatori” del problema (incentivi dati alle imprese perché investano nella formazione dei loro dipendenti).

Effetti inattesi e indesiderati

Come sappiamo, ogni terapia può produrre effetti collaterali più o meno gravi. Allo stesso modo ogni politica pubblica può avere *effetti inattesi o indesiderati*. Le due cose, si badi bene, non coincidono. Da un lato vi possono essere effetti non attesi che hanno però risvolti positivi; dall’altro, la politica può anche produrre conseguenze negative che, sebbene paventate da chi ne aveva promosso l’attuazione, non è stato possibile evitare. Un esempio di questo ultimo tipo di effetti può essere tratto dalle politiche di sostegno al reddito di persone indigenti. In tali politiche è prevedibile che parte di coloro che ricevono (o che potrebbero ricevere) il sussidio scelga di non lavorare e di rimanere legata per tempi anche lunghi al sistema assistenziale pubblico. In questi casi l’erogazione del sussidio diventa un disincentivo ad entrare, o a restare, nel mercato del lavoro e produce un possibile effetto indesiderato (spesso definito come “trappola della povertà”). Se, come in questo caso, l’effetto indesiderato è stato previsto, la

strumentazione analitica presentata in questo lavoro rimane valida ed adottabile, e serve a verificarne l'esistenza e a quantificarne l'entità. Al contrario, se gli effetti sono inattesi, siano essi desiderabili o meno, l'apparato analitico presentato in questo lavoro mostrerà la corda e sarà necessario ricorrere ad altri approcci, di tipo più esplorativo e meno "confermativo".

Alcune puntualizzazioni sono utili per chiarire il senso della lettura delle politiche pubbliche come "terapie". Innanzitutto, questa lettura non ha alcun intento *prescrittivo*. Non si vuole sostenere che le politiche pubbliche *debbano, o possano*, risolvere i problemi collettivi che ne motivano l'adozione; e ancor meno auspicare che le politiche risolvano *determinati* problemi collettivi e non altri; e nemmeno affermare che l'uso di *determinati* strumenti di *policy* sia più opportuno o meritevole di altri possibili. Le politiche pubbliche sono ciò che sono, o meglio sono quello che le porta ad essere il complesso processo di formazione del consenso e di mediazione tra interessi in conflitto nella società. Qui si sostiene soltanto come sia *utile* (nella prospettiva di valutarne il successo con gli strumenti illustrati in questo lavoro) *leggere* le politiche pubbliche come trattamento mirato ad affrontare un problema che si manifesta sotto forma di comportamenti o condizioni ritenuti socialmente insoddisfacenti. Chi opera questa lettura deve perciò mantenere nei confronti del suo oggetto d'indagine una sorta di *distaccato scetticismo* che lo porti a riconoscere gli elementi terapeutici contenuti nella politica in esame, senza però ritenere a priori che tale terapia sia necessariamente efficace o addirittura quella vincente.

La seconda precisazione consiste nel riconoscere come questa particolare prospettiva non sia del tutto neutrale e anzi porti a privilegiare la valutazione di certi tipi di politiche rispetto ad altre. Non c'è dubbio alcuno infatti che alcune tipologie di politiche pubbliche si prestano meglio di altre ad essere interpretate come terapie o trattamenti. Il nostro approccio tende quindi a privilegiare l'analisi di alcune politiche pubbliche. In particolare, esso tende a privilegiare quelle politiche che perseguono *obiettivi di cambiamento ben definiti* e quelle politiche che *non hanno una natura meramente simbolica* (questi due aspetti sono legati, ma non coincidono). Vale la pena soffermarsi su queste ultime considerazioni, per quanto a prima vista possano sembrare ovvie.

Gli “obiettivi” di una politica pubblica

Innanzitutto, le politiche pubbliche sono tanto più *leggibili* nell’ottica della terapia quanto più è possibile attribuire loro *obiettivi di cambiamento* circoscritti e ben definiti (diminuire la mortalità sulle strade; ridurre gli incendi boschivi; aumentare la competitività delle piccole imprese; incrementare la probabilità di occupazione dei lavoratori in mobilità), piuttosto che obiettivi sfocati o generici (incentivare lo sviluppo locale; lottare contro la criminalità; favorire le pari opportunità; combattere l’emarginazione sociale). Una più diretta leggibilità delle politiche come terapia implica anche che queste politiche sono più facilmente *valutabili*, almeno nel senso inteso in questo lavoro.

Tuttavia, assai raramente le finalità di una politica pubblica sono espresse in modo chiaro e diretto in qualche documento ufficiale. Molto più spesso, per facilitare il passaggio della legge che dà impulso alla politica o per non legarsi le mani con impegni troppo stringenti, i *policy maker* preferiscono esprimere gli obiettivi in termini più ambigui e indefiniti. Ad esempio, raramente sono utilizzati verbi “forti” come *aumentare, accrescere, diminuire* o *ridurre* per esprimere l’obiettivo di cambiamento di una politica pubblica; è più probabile che siano impiegati verbi più “deboli” come *incoraggiare, aiutare, favorire, promuovere* o *contrastare*, che lasciano qualche margine di vaghezza in più sulla direzione desiderata di cambiamento.

In modo analogo si evita di individuare chiaramente una sola dimensione, ben riconoscibile, sulla quale la politica pubblica potrebbe produrre i suoi effetti (ad es. percentuale di occupati, tasso d’inquinamento, numero di reati), ma si preferisce fare riferimento a fenomeni compositi, non immediatamente misurabili e un po’ “astratti” (sviluppo locale, sostenibilità ambientale, sicurezza). Questo però non dovrebbe spaventare l’analista/valutatore; dovrebbe piuttosto renderlo consapevole che uno dei suoi compiti principali consiste nel “*collaborare con gli sponsor della valutazione, i responsabili della politica, e gli altri stakeholder (coloro che hanno un qualche interesse, una qualche posta – stake – nel gioco di policy) per meglio specificare i reali obiettivi della politica e trasformare finalità espresse in modo ambiguo, ampio,*

eccessivamente idealizzato e retorico, in proposizioni chiare, concrete ed esplicite, il cui verificarsi sia realisticamente e direttamente osservabile, e quindi valutabile”¹⁴.

La seconda dimensione lungo la quale si gioca la rilevanza pratica della nostra lettura dipende dalla natura più o meno *simbolica* della politica stessa. È evidente come alcune politiche pubbliche abbiano una finalità prevalentemente simbolica, cioè non siano motivate dalla volontà di affrontare realmente un problema ma solo da quella di mostrare che chi promuove o sostiene quella politica “ha a cuore il problema”. Questo è vero sia nel caso di politiche che sono soltanto “annunciate”, sia di politiche che abbiano una loro concreta attuazione; in tutti e due i casi ciò che caratterizza questo tipo di politiche è una *evidente sproporzione* tra i risultati che si dichiara di voler raggiungere e le risorse messe in campo. Dietro una politica simbolica si nasconde quindi l’obiettivo del *policy maker* di guadagnare visibilità pubblica proponendo soluzioni (spesso) fittizie a problemi (spesso) irrisolvibili. Nel caso di una politica meramente simbolica, l’applicazione dei metodi qui proposti per valutarne il successo appare del tutto fuori luogo.

La minore o maggiore plausibilità della lettura che proponiamo ha in altre parole conseguenze importanti per l’applicabilità degli strumenti proposti. Quanto più appare difficile mettere a fuoco (anche a posteriori, una volta che sia attuata) su quali comportamenti o condizioni la politica cerchi di incidere, tanto meno è rilevante il paradigma analitico illustrato in questo lavoro. Il che non toglie che possano esistere altri metodi, di natura non quantitativa, adatti a “valutare” il successo di politiche i cui obiettivi siano enunciati in modo più sfumato o generico: di questi metodi non ci occupiamo in questo lavoro.

2.2 La domanda di valutazione

Riprendendo la metafora della terapia, partiamo dal momento in cui la cura è stata somministrata. Forse il trattamento non è stato esattamente quello previsto in origine, forse non è stato somministrato nelle condizioni ottimali, tuttavia i pazienti hanno

¹⁴ Rossi P., Freeman H., Lipsey M., *Evaluation: A Systematic Approach*, Sage Publications, Beverly Hills, 1999, VI ed., pag. 94.

ricevuto, o sono stati sottoposti, a “qualcosa”: risorse sono state spese, obblighi e divieti sono stati imposti, servizi sono stati erogati, incentivi sono stati resi disponibili, messaggi sono stati lanciati.

A questo punto ci si chiede: gli strumenti adottati hanno funzionato? La terapia è stata efficace, nel senso che ha prodotto gli effetti desiderati? *Queste sono le domande valutative di cui ci occuperemo in questo lavoro.* Come abbiamo più volte sottolineato, questo non vuol dire che siano le uniche domande valutative possibili e forse neppure le più importanti.

Il passo preliminare per formulare una corretta e completa domanda valutativa rispetto agli effetti di una politica è definire *su cosa* si cerca l'effetto della politica e, in secondo luogo, *di che cosa* si sta ricercando l'effetto. Entrambi questi passi si concretizzano nell'identificazione di “variabili”: la prima sarà chiamata variabile-risultato, la seconda variabile-trattamento. La formulazione della domanda valutativa presuppone inoltre una certa chiarezza concettuale rispetto a “*che cosa si intende per effetto*”.

Nelle pagine seguenti tenteremo di proporre una definizione quanto più possibile rigorosa del concetto di effetto, concentrandoci su quelle che riteniamo essere le sue dimensioni cruciali; su tale definizione si imporrà tutto l'apparato concettuale sviluppato nei prossimi capitoli. Una volta definite tali dimensioni e formulato così un prototipo di domanda valutativa, si tratterà di individuare quali sono i metodi disponibili per ottenere una valida risposta a tale domanda. Si passa cioè ad affrontare il problema del *come* si stabilisce se l'effetto c'è e (se c'è) *quanto* esso sia grande. Ma prima vediamo quali sono gli ingredienti necessari per formulare una corretta domanda valutativa sugli effetti delle politiche pubbliche.

Effetto su cosa? Specificare la variabile-risultato

Proseguiamo con la metafora della politica pubblica come terapia. Intendiamo misurarne il successo in termini di capacità di cura o di prevenzione, cioè in termini di capacità di modificare nella direzione voluta i comportamenti o le condizioni che costituiscono il problema da cui la politica trae motivazione. La metafora della terapia ci aiuta a focalizzare l'attenzione sul fatto che è assolutamente necessario, come pre-condizione per valutare il successo o l'insuccesso di una politica, identificare con

sufficiente precisione le dimensioni problematiche da modificare. Per poter dire se un miglioramento c'è stato (che è preliminare a poter concludere “il miglioramento è dovuto o meno alla terapia”) occorre definire il miglioramento *rispetto a che cosa*.

Quindi i comportamenti/condizioni che rappresentano il problema vanno *tradotti* in dimensioni che siano – in linea di principio almeno – *osservabili e misurabili*. Questo è il primo passo che si deve compiere sulla strada della valutazione delle politiche: occorre chiedersi quali sono le caratteristiche osservabili della popolazione su cui la politica interviene, che meglio sembrano rappresentare i comportamenti o le condizioni che la politica intende modificare.

La risposta a questa domanda non è né semplice, né univoca, e spesso non la si può dare una volta per tutte, ma occorre procedere per approssimazioni successive. Tuttavia questo è, con tutte le sue incertezze, un *passaggio obbligato*. Detto in termini più semplici e brutali: se l'obiettivo della politica è modificare qualcosa, questo qualcosa deve essere definito in termini tali da poter essere *misurabile*. Si tratta di individuare delle “variabili-risultato” (in inglese *outcome variables*) che rappresentino, nel modo migliore possibile, il problema che la politica intende affrontare. Si è usato il plurale non a caso: un singolo problema di *policy* può infatti essere rappresentato da più variabili-risultato.

Per fare un esempio concreto, poniamo che una certa politica pubblica abbia l'obiettivo (generico) di combattere la disoccupazione tra i giovani, e più nello specifico quello di ridurre i tempi di ricerca di lavoro dopo il completamento del ciclo formativo. Le variabili-risultato utili a rappresentare il problema e quindi il potenziale miglioramento rispetto al problema possono essere molte. Prendiamone tre, tra le molte possibili:

- a) la percentuale di occupati ad una certa distanza temporale dalla conclusione del ciclo formativo;
- b) la durata media della ricerca di lavoro necessaria a trovare la prima occupazione;
- c) il numero di settimane di lavoro svolte in un certo periodo successivo alla conclusione del ciclo formativo.

La scelta tra queste variabili-risultato dipende da un complesso di considerazioni, in parte legate al contenuto stesso della *policy*, in parte di carattere più strettamente

tecnico-statistico. L'importante è che si giunga a definire *almeno una variabile-risultato*, ma non necessariamente una soltanto: la valutazione può essere fatta rispetto a più variabili-risultato. Come vedremo, i problemi metodologici che si pongono per determinare l'*effetto della politica* su una variabile-risultato si pongono anche rispetto a tutte le altre variabili. Nel caso si ritenga opportuno prendere in considerazione più variabili, lo stesso tipo di analisi condotta rispetto ad una di esse deve essere replicata anche rispetto alle altre.

A questo punto si impone una precisazione, per fugare un facile equivoco: essere in grado di indicare una caratteristica della popolazione come variabile-risultato della politica è cosa *ben diversa* dall'essere in grado di stimare l'effetto della politica su quella variabile. La prima azione non produce la seconda, è solo una delle sue pre-condizioni: per stimare un effetto occorre prima stabilire effetto *su cosa*. Che poi questo effetto ci sia (quanto sia grande, se sia statisticamente significativo) è, come si suol dire, tutto un altro paio di maniche.

Questo è un punto su cui è assai facile fare confusione, indotti anche da certa letteratura in materia, nella quale quella che noi chiamiamo scelta della (o delle) variabili-risultato viene definita come "scelta degli indicatori". In effetti, la parola indicatore può essere considerata sotto molti aspetti un sinonimo di variabile-risultato. In questo contesto preferiamo però quest'ultimo termine, in primo luogo perché sottolinea in modo netto che si tratta di una variabile (cioè una caratteristica misurabile per le singole unità di una popolazione) e non di una qualche statistica aggregata riferita alla popolazione. In secondo luogo, perché sottolinea come questa variabile sia usata come "risultato", cioè come rappresentazione di ciò che la politica vuole migliorare. Naturalmente è possibile anche utilizzare il termine "indicatore", purché si accetti il fatto che, di per sé, tale indicatore non indica l'effetto della politica. Una certa letteratura tenta invece di attribuire proprio tale significato agli indicatori, etichettandoli, a nostro avviso artificialmente, come "indicatori di impatto".

Effetto di cosa? Specificare la variabile-trattamento

Una volta stabilito su quale (o quali) variabile-risultato vanno ricercati gli effetti della politica, è necessario individuare operativamente il “trattamento” in cui concretamente consiste l’intervento pubblico che si intende valutare. Occorre cioè stabilire esattamente *di cosa* si stia cercando l’effetto. La tentazione più immediata è quella di dire che si sta cercando l’effetto “della politica pubblica”, ma questo ovviamente non basta. Tranne che in poche eccezioni, le politiche pubbliche si presentano come un *mix di interventi di natura differente, realizzati spesso in tempi diversi*. In quanto tale, una politica è raramente valutabile nel suo complesso, almeno secondo l’approccio d’analisi presentato in questo lavoro.

È necessario perciò semplificare le cose e individuare una *variabile-trattamento*, che nella sua forma più semplice sarà una variabile dicotomica (Sì/No) e rappresenterà l’assenza-presenza del trattamento, o meglio l’esposizione o non-esposizione dei membri della popolazione al trattamento. La partecipazione o meno ad un corso di formazione professionale, l’esposizione o meno ad una campagna pubblicitaria contro il fumo, l’incasso o meno di un contributo a fondo perduto per l’avvio di una nuova impresa sono alcuni concreti esempi di variabili-trattamento. Ciò che ci preme sottolineare è che la politica, per essere valutata, deve essere rappresentabile mediante una variabile che, appunto, presenti delle *variazioni osservabili* nel tempo oppure nello spazio, cioè tra soggetti.

Gli effetti di una politica costante e universale, cioè che sottopone tutti allo stesso trattamento e che non subisce alcun tipo di variazione nel tempo, non sono valutabili con i metodi illustrati in questo corso (e riteniamo difficilmente valutabili in generale). Tuttavia è assai difficile trovare una politica pubblica che sia costante (nel tempo) e universalmente applicabile a tutti. La storia insegna che le politiche pubbliche sono sempre soggette a qualche tipo di perturbazione o di modifica nel tempo (nuove regole vengono introdotte, nuovi sussidi o servizi erogati,...); quindi ci sarà sempre un periodo “pre” in cui la politica di fatto non esisteva, o non era ancora stata attuata, ed un periodo “post”, durante il quale ci si aspetta di raccogliere i frutti di quanto di nuovo la politica ha introdotto.

L'*entrata in vigore* di una nuova politica (o la sua *cessazione*) è la prima e più evidente fonte di variazione del trattamento; e resta la più importante fonte di variazione per quelle politiche che consistono nell'imposizione di obblighi o divieti *erga omnes*, come ad esempio l'obbligo di indossare le cinture di sicurezza o il divieto di fumare nei locali pubblici.

Ma questa non è certo l'unica fonte di possibile variazione nel trattamento. Molte altre politiche, pur restando sempre uguali a loro stesse e non essendo soggette a drastiche innovazioni, offrono la possibilità di distinguere tra un "prima" e un "dopo" il trattamento. Si tratta delle politiche che consistono nell'erogazione di servizi o benefici per un determinato periodo di tempo a determinati soggetti: ad esempio, la formazione professionale o il sussidio di disoccupazione. Poiché ci sarà sempre un momento in cui l'utente comincia a usufruire del servizio o il beneficiario a ricevere la prestazione, sarà sempre possibile definire un momento precedente, in cui il servizio non era ancora stato fruito; analogamente, sarà possibile definire un periodo successivo alla fruizione del servizio o alla ricezione del beneficio.

Altrettanto importanti fonti di variazioni nel trattamento sono quelle legate al fatto che, in un certo periodo temporale, sono esposti alla politica solo alcuni individui e non altri. In questo caso la mancata esposizione può dipendere da decisioni personali dell'individuo oppure da decisioni collettive, legate cioè al disegno della politica stessa. Alcuni individui possono semplicemente decidere di non usufruire mai di un certo servizio o beneficio. In questo caso si dice che la mancata esposizione alla politica è frutto di "autoselezione". Ad esempio, alcuni imprenditori possono decidere di non fare domanda per un certo incentivo a sostegno degli investimenti, pur avendone pieno diritto, perché non lo considerano conveniente.

In molti altri casi la mancata esposizione alla politica è dovuta ad un'esclusione determinata in modo consapevole da coloro che hanno disegnato la politica. Per restare al nostro esempio, alcune categorie di imprenditori possono essere escluse dall'agevolazione, ed altre incluse, sulla base di certe caratteristiche osservabili (dimensionali, geografiche, settoriali) che determinano i requisiti d'ammissione al

contributo. È quindi il disegno stesso della politica ad imporre un processo di selezione dei destinatari.

Come si è detto, nella gran parte degli esempi e dei casi utilizzati in questo lavoro la variabile-trattamento sarà una variabile dicotomica: il trattamento c'è o non c'è (variazione nel tempo), oppure ad esso sono esposti o non sono esposti i membri della popolazione (variazione nello spazio). Naturalmente si tratta di una semplificazione, in quanto la variabile-trattamento potrebbe anche *non* essere una variabile dicotomica e potrebbe presentarsi con più di due valori, corrispondenti a differenti gradi di intensità nel trattamento (“dosaggio”). La scelta di considerare in questa sede soltanto il caso della variabile dicotomica nasce dalla considerazione che per una buona parte delle politiche pubbliche il trattamento può in effetti essere rappresentato in questo modo; ma dipende anche dal fatto che è più semplice utilizzare questo caso per spiegare in termini intuitivi l'approccio analitico utilizzato in questo rapporto. La gran parte dei ragionamenti fatti per questo particolare caso si estendono comunque anche al caso in cui il trattamento sia somministrato con un diverso “dosaggio” da caso a caso.

2.3 Una definizione rigorosa di “effetto”

Una volta messo in evidenza *su cosa* si ricercano gli effetti e *di cosa* si ricercano gli effetti, resta da spiegare cosa precisamente si intende per effetto di una politica. All'interno del paradigma concettuale illustrato in questo lavoro, *l'effetto di una politica si identifica nel cambiamento nella variabile-risultato attribuibile in senso causale alla variabile-trattamento*.

Da ciò consegue che l'effetto della politica che desideriamo stimare per valutarne il successo *non* sarà semplicemente il *livello* della variabile-risultato *osservato* tra i soggetti esposti alla politica, come si legge in certa letteratura dedicata agli indicatori. Ad esempio, il tasso di occupazione osservato dopo la frequenza di un corso di formazione *non* è l'effetto del corso di formazione.

Come abbiamo già sottolineato, tale confusione concettuale nasce spesso dalla scelta di attribuire alla variabile-risultato il nome di “indicatore d'impatto”, come se il problema della stima degli effetti potesse essere ricondotto semplicemente alla definizione

dell'indicatore. Secondo questa logica (errata), una volta *misurato* il valore per l'indicatore prescelto, si otterrebbe automaticamente una stima dell'impatto (o dell'effetto) della politica. In questa logica scompare totalmente la nozione di "cambiamento attribuibile" che invece abbiamo posto al centro della nostra definizione e che è fondamentale per comprendere la logica della valutazione degli effetti.

Quindi l'effetto *non* è il valore della variabile-risultato osservato dopo l'attuazione della politica, neppure se quest'ultimo è definito come "indicatore di impatto". Si tratta di un errore banale, ma nondimeno presente in molti documenti che trattano di valutazione di politiche e programmi.

Ma esiste anche un secondo tipo di errore, piuttosto comune, quello che porta a considerare l'effetto della politica come il semplice *cambiamento osservato* nella variabile-risultato tra i soggetti esposti alla politica. Ma l'effetto non è la differenza tra il valore della variabile-risultato osservato *dopo* l'attuazione della politica e quello osservato *prima* dell'attuazione della politica. Ciò che manca alla nozione di effetto appena enunciata è una precisa attribuzione di causalità: il cambiamento osservato tra prima e dopo l'intervento potrebbe infatti essersi verificato comunque, per cause non riconducibili all'attuazione della politica. Ad esempio l'aumento della percentuale di occupati osservato dopo l'erogazione di un corso di formazione non è necessariamente effetto del corso, ma potrebbe essere spiegato dalle mutate condizioni nel mercato del lavoro locale.

Per poter essere chiamato effetto, il cambiamento osservato deve essere attribuibile *in senso causale* alla variabile-trattamento. Facendo riferimento all'esempio appena fatto, dovremmo essere sicuri che l'aumento di occupazione registrato tra i partecipanti al corso non si sarebbe verificato *comunque*, cioè anche senza il corso di formazione. L'effetto rappresenta perciò il "contributo netto" dato dalla politica al cambiamento osservato nella variabile-risultato, cioè quella parte del cambiamento causato interamente dall'intervento pubblico.

Questa attribuzione di causalità rappresenta il tratto distintivo dell'approccio alla valutazione degli effetti presentato in questo volume. Quindi è importante elaborare questa definizione di effetto, rendendo più esplicita l'idea di attribuzione di causalità.

Definiamo “effetto di una politica” la *differenza tra il valore osservato nella variabile-risultato dopo l’attuazione della politica e il valore che si sarebbe osservato senza la politica.*

Come è facilmente comprensibile, mentre il primo termine di questa differenza si riferisce ad un valore osservabile, il secondo termine fa riferimento ad un valore puramente ipotetico, non osservabile per definizione. Infatti, poiché la politica è stata attuata, ciò che sarebbe accaduto in sua assenza non può essere osservato. Questo valore ipotetico è definito in gergo valore, o situazione, *controfattuale*. L’idea di controfattuale è fondamentale per comprendere i metodi analitici illustrati all’interno di questo lavoro, tanto che l’approccio che li accomuna può essere sinteticamente definito come “paradigma controfattuale”.

L’effetto di una politica è *definito* come *differenza* tra due valori della variabile-risultato. Ma di questi due valori:

- uno è *osservabile*, tra i soggetti esposti alla politica, *dopo* l’esposizione (valore fattuale)
- l’altro è un valore *ipotetico* e si riferisce a ciò che *si sarebbe* osservato tra gli stessi soggetti, nello stesso momento, se costoro non fossero stati esposti alla politica (valore controfattuale)

È utile ribadire come tale differenza non sia direttamente calcolabile, perché mentre uno dei termini della differenza è direttamente osservabile (almeno in linea di principio), il secondo termine *non lo è mai*, in quanto è l’attuazione stessa della politica a renderne impossibile l’osservazione. Sono possibili solo due eventi: o un insieme di soggetti è esposto alla politica (e quindi non possiamo osservare il controfattuale); oppure non vi è esposto (e quindi non possiamo osservare alcun valore dopo l’esposizione alla politica, dato che tale esposizione non c’è stata).

Questo è un dilemma che non riguarda solo la valutazione delle politiche ma in genere tutta quella parte delle scienze sociali che tenta di inferire rapporti di causalità

dall'osservazione di fenomeni. Holland (1986) definisce la non osservabilità del controfattuale come “*the fundamental problem of causal inference*”¹⁵.

In linea di principio, questo dilemma non ha soluzione. Se l'effetto di una politica è la differenza tra ciò che è successo *dopo* e ciò che sarebbe successo *senza*, l'effetto non sarà mai osservabile e la valutazione dell'effetto della politica sarà a rigore impossibile. C'è qualcosa di profondamente vero in questa conclusione “pessimistica”: comunque andranno le cose, non saremo mai *assolutamente certi* di quale effetto una politica abbia avuto una volta attuata. Tuttavia non è la certezza assoluta che stiamo perseguendo: ci accontentiamo di una ragionevole e credibile approssimazione. E tale approssimazione dell'effetto di una politica potrà essere ottenuta sostituendo al valore controfattuale un valore che sia osservabile e allo stesso tempo approssimi, nel modo più *credibile* possibile, ciò che sarebbe successo ai soggetti esposti alla politica se non lo fossero stati.

Le strategie per la valutazione degli effetti di una politica sono quindi, nella loro essenza, strategie per approssimare il controfattuale con qualche valore *credibile*, utilizzando le informazioni disponibili. *Una valutazione degli effetti sarà tanto più plausibile quanto più sarà credibile la strategia che abbiamo adottato per approssimare il controfattuale.*

Da questo punto di vista la valutazione delle politiche non produce certezze incrollabili, ma può produrre *argomentazioni*, che possono comunque arricchire il processo decisionale pubblico. Occorre essere consapevoli del fatto che l'uso, anche estremamente rigoroso, di numeri e statistiche non genera di per sé risultati incontrovertibili, al massimo mette a disposizione dei *policy maker* e della collettività, ragionamenti più solidi e credibili a favore, o contro, determinate ricette di *policy*. Ma si tratterà pur sempre di argomentazioni, più o meno plausibili, e non di verità scientifiche.

Nei prossimi cinque capitoli esamineremo una serie di possibili disegni di valutazione degli effetti di politiche pubbliche: tali disegni sono molto diversi tra loro, in termini di complessità analitica, di ricchezza delle informazioni che richiedono, di difficoltà di

¹⁵ Holland P., “Statistics and causal inference”, *Journal of the American Statistical Association*, 1986, 81: 945-960.

applicazione e di plausibilità delle stime che producono¹⁶. Però tutti questi disegni sono accomunati da un tema di fondo: per valutare un effetto occorre trovare un modo per ricostruire la situazione controfattuale. Cominceremo con il modo in linea di principio più credibile per operare tale ricostruzione: l'esperimento con randomizzazione.

¹⁶ I riferimenti per eventuali approfondimenti sono molteplici. La letteratura di taglio econometrico è magistralmente riassunta da Heckman J., LaLonde R. e Smith J., "The Economics and Econometrics of Active Labor Market Programs", in Orley Ashenfelter e David Card (a cura di), *Handbook of Labor Economics*, pp. 1865-2097, North-Holland, Amsterdam, 1999. Per una rassegna leggermente meno tecnica si veda Blundell R., Costa Dias M., "Alternative approaches to evaluation in empirical microeconomics", *Portuguese Economic Journal*, 2002, Vol. 1, No. 2, pp. 91-115. Per un impianto più sociologico e meno formalizzato si veda Mohr L.B., *Impact Analysis for Program Evaluation* (2nd edition), Sage Publications, Beverly Hill, 1995 e Shadish W., Cook T. e Campbell D., *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton-Mifflin, Boston, 2002.

CAPITOLO 3

IL METODO SPERIMENTALE: UN PUNTO DI RIFERIMENTO PER LA VALUTAZIONE DEGLI EFFETTI

di Alberto Martini

Nel capitolo precedente abbiamo utilizzato la metafora della terapia per richiamare l'attenzione su due punti-chiave della valutazione degli effetti di una politica: la necessità di definire con precisione la condizione che la politica pubblica vuole modificare (effetto *su cosa*), e quella di chiarire di quale trattamento si vuole analizzare l'effetto (effetto *di cosa*). In questo capitolo la metafora della terapia si dimostra ancora utile per illustrare l'idea centrale dell'approccio *sperimentale* alla valutazione.

Testare l'efficacia dei farmaci

L'ispirazione del metodo sperimentale per valutare gli effetti delle politiche deriva infatti dall'ambito medico-farmacologico, dove è applicato per testare l'efficacia dei nuovi farmaci, mediante i cosiddetti *clinical trials*. Un gruppo di pazienti affetti da una certa patologia viene suddiviso in due gruppi mediante randomizzazione, cioè sorteggio: ai membri di uno dei due gruppi viene somministrato il farmaco da sperimentare (gruppo dei trattati o "gruppo sperimentale"), mentre ai membri dell'altro viene somministrato un *placebo*, cioè una sostanza inerte, che ha le stesse caratteristiche organolettiche del farmaco ma non può avere alcun effetto diretto sulla patologia¹⁷. Questo secondo gruppo viene definito "gruppo di controllo".

Perché è necessario un gruppo di controllo? Il gruppo di controllo serve, per riprendere il linguaggio del capitolo precedente, ad approssimare la situazione controfattuale. Più precisamente, grazie alla randomizzazione, il decorso della patologia che si osserva tra i membri del gruppo di controllo sarà presumibilmente simile al decorso che la patologia

¹⁷ La somministrazione del *placebo* un effetto diretto può comunque averlo, nel senso di ingenerare nel paziente la convinzione di venire effettivamente curato e quindi di modificare il suo atteggiamento nei confronti della malattia (ad esempio convincendolo che guarirà). Si parla infatti di "effetto *placebo*". Questo particolare fenomeno è poco probabile che si verifichi nell'ambito di esperimenti sociali, in quanto il gruppo di controllo di solito non è "cieco", cioè i suoi membri sanno di non essere esposti al particolare intervento pubblico che viene valutato, sanno di non ricevere il servizio, sussidio o agevolazione prevista dall'intervento.

avrebbe avuto tra i membri del gruppo dei trattati *qualora essi non avessero ricevuto il farmaco*. Questa espressione è equivalente alla definizione di controfattuale. Una differenza significativa (nella direzione desiderata) tra il decorso osservato nella patologia per il gruppo sperimentale e il decorso osservato per il gruppo di controllo rivela che il farmaco è efficace, cioè ottiene un effetto nella direzione desiderata¹⁸.

Ovviamente, i membri del gruppo di controllo non sono *identici* ai membri del gruppo sperimentale. L'esperimento "ideale" dovrebbe prendere due gruppi di pazienti assolutamente identici e sottoporne uno al trattamento e l'altro no. Quando si opera con esseri viventi o organismi complessi che si differenziano lungo moltissime dimensioni, creare due gruppi perfettamente identici è virtualmente impossibile.

Tuttavia la perfetta identità non è un requisito indispensabile: è sufficiente che i due gruppi siano *statisticamente equivalenti*, cioè abbiano la stessa distribuzione di tutte le caratteristiche (ad esempio, la stessa media, la stessa varianza, gli stessi decili), eccezion fatta per l'inevitabile errore campionario, che però tende ad essere "piccolo" quando il campione utilizzato per generare i due gruppi è sufficientemente "grande". Per ottenere questa equivalenza in senso statistico è comunque necessario che la selezione dei due gruppi sia operata in modo assolutamente casuale, mediante quella randomizzazione (dall'inglese *random*, casuale) o assegnazione casuale (in inglese *random assignment*).

Si parla infatti di *randomized clinical trials*: nella letteratura sulla valutazione si trova spesso il termine, con lo stesso acronimo *RCT* ma un significato più generale, di *randomized controlled trials*. La parola "*controlled*" sottolinea come il ricercatore (o il valutatore) abbia completo controllo sul processo di selezione, cioè su quel processo che determina quali soggetti sono sottoposti al trattamento e quali no. Il ricercatore non esercita questo controllo scegliendo i due gruppi secondo le sue preferenze, simpatie o altri criteri mirati, ma designandoli mediante randomizzazione, cioè a caso.

¹⁸ Per riprendere una questione sollevata nel capitolo precedente, il farmaco può anche avere effetti collaterali indesiderati: ma anche questi andranno stimati "per differenza" rispetto al gruppo di controllo, che presumibilmente non li subirà. La logica controfattuale vale per la stima di effetti desiderati come per quelli indesiderati.

La randomizzazione

Lasciando da parte le questioni di natura tecnica sulla dimensione che deve avere il campione complessivo di soggetti prima della randomizzazione (supponiamo che il campione sia molto grande), concentriamoci sulle conseguenze della randomizzazione. Essa garantisce che le caratteristiche dei due gruppi siano le stesse, sia per le caratteristiche *osservabili*, sia per le caratteristiche *non osservabili* dei membri dei due gruppi.

Questo è un punto su cui vale la pena soffermarsi. I metodi non-sperimentali presentati nei prossimi capitoli fanno ricorso, in un modo o nell'altro, alle caratteristiche *osservabili* dei soggetti trattati e non-trattati e si basano su qualche *assunto arbitrario* che riguarda le caratteristiche *non osservabili* dei soggetti trattati e non-trattati (ad esempio, sull'assunto che le eventuali differenze tra i due gruppi nelle caratteristiche non osservabili *non* influenzino la variabile-risultato).

Di quelle tecniche (e di quegli assunti arbitrari) non abbiamo bisogno, in linea di principio almeno, nel caso dei *randomized controlled trials*. La completa equivalenza pre-trattamento tra i due gruppi rende plausibile attribuire in senso *causale* al trattamento le differenze nella variabile-risultato che si osservano tra i due gruppi *dopo* il trattamento. In altre parole, se trattati e non-trattati sono simili “in tutto e per tutto” prima del trattamento, resta solo il trattamento, ricevuto dagli uni e non dagli altri, a spiegare le differenze che (eventualmente) si osservino *dopo* il trattamento.

Quindi *le differenze post-trattamento* nella variabile-risultato “rivelano” in modo credibile l'effetto che il trattamento ha avuto *in media* sui soggetti trattati. E questo perché l'esperienza avuta dopo la randomizzazione dal gruppo di controllo rappresenta un'approssimazione plausibile dell'esperienza che il gruppo sperimentale avrebbe avuto se non fosse stato sottoposto al trattamento.

Questa logica non si applica ovviamente solo alla verifica dell'efficacia dei farmaci. La logica sperimentale è stata applicata con successo anche in psicologia e nella scienza della formazione, dove il trattamento può consistere in un tipo nuovo di tecnica didattica o in un cambiamento dei programmi scolastici e la variabile-risultato può essere una misura della performance degli studenti ottenuta mediante test standardizzati.

Negli Stati Uniti, a partire dagli anni Sessanta, il metodo sperimentale è stato ampiamente utilizzato per valutare gli effetti di specifiche politiche pubbliche, soprattutto nei settori della formazione professionale e dei servizi all'impiego, delle politiche attive del lavoro, delle politiche abitative per le famiglie a basso reddito, dei programmi di assistenza economica (*welfare*) e dell'assistenza sanitaria pubblica. Esiste un'ampia letteratura, a cui si rimanda¹⁹, che passa in rassegna questa esperienza, peraltro quasi esclusivamente confinata agli Stati Uniti, con qualche sporadica applicazione nel Regno Unito, in Canada e nei Paesi dell'Europa settentrionale. In Italia e nei Paesi dell'Europa continentale l'utilizzo del metodo sperimentale nell'ambito della valutazione delle politiche è sostanzialmente sconosciuto.

Nonostante la forte concentrazione geografica del suo utilizzo (e, come vedremo, i limiti pratici alla sua adozione), dal punto di vista metodologico il metodo sperimentale rappresenta un punto di riferimento importante per tutta la valutazione degli effetti delle politiche pubbliche. Questo almeno per due ragioni. A livello didattico, l'illustrazione dei risultati ottenuti da un esperimento con randomizzazione aiuta ad apprezzare l'importanza pratica del controfattuale e fa toccare con mano cosa concretamente significhi stimare l'effetto di una politica come differenza tra un valore osservato e un valore controfattuale. Ad un livello più metodologico, le stime degli effetti prodotte da un esperimento sociale rappresentano un *benchmark*, un termine di paragone, per le stime ottenute mediante metodi non-sperimentali applicati alla stessa politica pubblica: un particolare metodo non-sperimentale è considerato accettabile se riesce a produrre stime degli effetti simili a quelle prodotte dall'esperimento stesso.

Questi due utilizzi (quello didattico e quello metodologico) offrono una motivazione per esaminare in dettaglio il disegno e i risultati di una delle valutazioni sperimentali condotte negli Stati Uniti negli ultimi quarant'anni. Si tratta della *National Supported Work Demonstration*, condotta nella seconda metà degli anni Settanta per valutare gli effetti di un particolare tipo di approccio al reinserimento lavorativo di persone

¹⁹ Si vedano in particolare il saggio metodologico di Orr L., *Social Experiments: Evaluating Public Programs with Experimental Methods*, Sage Publications, Beverly Hill, 1999; la rassegna di Greenberg D. e Shroder M., *Digest of Social Experiments* The Urban Institute Press, Washington, 1997 e il saggio di Greenberg D., Linksz D., Mandel M., *Social Experimentation and Public Policymaking*, The Urban Institute Press, Washington, 2003.

appartenenti alle cosiddette “fasce deboli”. Le indicazioni di *policy* che questa valutazione produsse a suo tempo sono oggi ormai di limitato interesse: tuttavia, la NSWSD è ancora attuale e ampiamente citata nella letteratura in quanto i dati da essa generati sono stati utilizzati in due importanti studi di tipo metodologico.

Il primo è uno studio pubblicato nel 1986 da Robert LaLonde per dimostrare quanto povera fosse la capacità dei metodi non-sperimentali disponibili a quell’epoca di riprodurre le stime degli effetti ottenuti dall’esperimento della NSWSD²⁰. Lo studio di LaLonde fu all’epoca molto influente nel determinare da un lato un maggior ricorso agli esperimenti sociali e dall’altro a stimolare la comunità degli statistici e degli econometrici a sviluppare tecniche più robuste per valutare gli effetti delle politiche con metodi non-sperimentali²¹. Il secondo studio è invece pubblicato da Rajeev Dehejia e Sadek Wahba nel 1999²²: in esso i due studiosi stimano gli effetti della NSWSD utilizzando una tecnica introdotta negli studi di valutazione a partire dalla metà degli anni Novanta, il *matching* statistico²³. Utilizzando questa tecnica, i due studiosi raggiungono la conclusione che il suo utilizzo produce risultati molto simili a quelli prodotti dall’esperimento, contribuendo così a sfatare l’idea della superiorità assoluta dei metodi sperimentali.

3.1 Disegno e risultati della National Supported Work Demonstration

Lo scopo della *National Supported Work Demonstration* è quello di valutare l’efficacia del *supported work*, un approccio nuovo per inserire o reinserire al lavoro i soggetti che soffrono di forme severe di emarginazione sociale. Il *supported work* consiste nel fornire a tali soggetti un’esperienza lavorativa “vera”, ma in un *ambiente protetto*, con un aumento molto graduale dell’intensità dello sforzo di lavoro, sotto supervisione di personale specializzato e con un inserimento in squadre di lavoro composte di individui con simili problemi, con una retribuzione garantita pari almeno al salario minimo

²⁰ LaLonde R., “Evaluating the Econometric Evaluations of Employment and Training Programs with Experimental Data”, *American Economic Review*, 1986.

²¹ Uno degli studiosi più impegnati e più creativi su questo fronte è, fin dalla metà degli anni Ottanta, James Heckman, dell’Università di Chicago, a cui nel 2000 è stato conferito il premio Nobel per l’economia.

²² Dehejia R., S. Wahba, “Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs”, *Journal of the American Statistical Association*, 1999.

²³ Il *matching* statistico è illustrato nel capitolo 5.

stabilito per legge. L'esperienza di *supported work* è destinata a durare all'incirca un anno, dopodiché si tenta l'inserimento nel mercato del lavoro regolare.

Quattro gruppi distinti vengono scelti come *target* per testare l'efficacia di questa politica attiva del lavoro: madri *single* che da più di tre anni ricevono il sussidio di povertà; giovani *drop-out* della scuola dell'obbligo, molti dei quali con precedenti penali, ex-tossicodipendenti ed ex-carcerati²⁴.

L'idea del *supported work* non è nuova, anzi è mutuata da varie esperienze di "lavoro protetto" condotte in Paesi europei quali il Regno Unito, la Svezia e i Paesi Bassi, dove sono state usate soprattutto per l'inserimento lavorativo di portatori di *handicap*. L'idea di estendere il *supported work* a soggetti diversi dai portatori di *handicap* appare ad alcuni affascinante e promettente, mentre ci sono tra i *policy maker* forti dubbi sull'efficacia di tale tipo di intervento quando sia applicato alle forme di esclusione sociale diffuse nei grandi centri urbani americani. Questa incertezza motiva l'idea di valutare l'efficacia del *supported work* mediante un intervento realizzato su scala ridotta, in modo da evitare di impegnare ingenti risorse su una politica dai risultati ancora incerti.

Così come realizzato nell'ambito della *NSWD*, si tratta infatti di un intervento che richiede un investimento notevole di risorse per partecipante, tra i 5.000 e gli 8.000 dollari, cifre che farebbero del *supported work* la più costosa politica attiva di reinserimento lavorativo esistente all'epoca. Prima di lanciare un simile intervento su larga scala, il governo federale americano decide di valutarne l'efficacia mediante uno studio-pilota (in inglese *demonstration*), per verificare se l'alto costo è giustificabile quando sia rapportato ai benefici che produce. Questi benefici attesi sono definiti in termini di aumento della partecipazione al lavoro, riduzione della dipendenza dai sussidi pubblici e diminuzione di comportamenti anti-sociali, quali l'uso di droga e l'attività criminale. Nel linguaggio introdotto nel capitolo precedente, questi obbiettivi rappresentano la base per definire le variabili-risultato da utilizzare nella valutazione degli effetti della politica.

²⁴ Il disegno e i risultati di questo esperimento sono illustrati in Manpower Demonstration Research Corporation, *Summary and Findings of the National Supported Work Demonstration*, 1980, su cui è basata la descrizione seguente.

Per il disegno e la realizzazione della *NSWD* viene costituita una nuova organizzazione *non-profit*, la *Manpower Demonstration Research Corporation (MDRC)*, in quanto alla metà degli anni Settanta manca ancora negli Stati Uniti un'organizzazione di ricerca con le caratteristiche adeguate per svolgere una valutazione di questa portata²⁵. Il finanziamento per la *NSWD* viene fornito da un *pool* di Ministeri federali (del lavoro, della giustizia, dell'istruzione, della sanità e dello sviluppo urbano) e da alcune organizzazioni private, tra cui soprattutto la *Ford Foundation*, che rimane oggi uno dei principali finanziatori privati di valutazioni di questo tipo. Il costo totale della *demonstration* è di oltre 80 milioni di dollari, di cui 10 milioni per la valutazione e i restanti 70 milioni per la realizzazione dell'intervento vero e proprio (locali, equipaggiamento, retribuzione dei partecipanti e dei supervisori). Circa 50 milioni di dollari vengono forniti dai finanziatori nazionali, mentre la differenza viene coperta mediante fondi messi a disposizione dai governi locali, tramite la vendita dei beni e servizi prodotti dai partecipanti e con lo storno di parte dei sussidi di povertà che molti dei partecipanti avrebbero ottenuto se non fossero stati coinvolti nell'esperimento.

La *NSWD* viene realizzata in 14 centri urbani, localizzati in altrettanti Stati. In ciascun sito viene incaricata o costituita *ad hoc* un'organizzazione *non-profit* con il compito di gestire l'intervento, mentre *MDRC* mantiene il coordinamento delle operazioni a livello nazionale e la gestione della valutazione, compresa la supervisione del *random assignment*. Le organizzazioni locali hanno l'incarico di gestire "laboratori di lavoro protetto". I partecipanti sono inviati a questi laboratori da assistenti sociali, centri di disintossicazione, *welfare offices* e altre strutture di assistenza pubblica.

Le attività lavorative in cui i partecipanti sono coinvolti sono per la metà nei servizi, dalla manutenzione di edifici all'assistenza negli asili-nido, un quarto nel settore edilizio, e quasi un 10 per cento nel settore manifatturiero. I partecipanti vengono regolarmente retribuiti, con una retribuzione di poco superiore al salario minimo stabilito dalla legislazione federale. L'intervento dura in complesso quasi quattro anni, dal 1975 al 1979, e coinvolge circa 10.000 individui, che partecipano al *supported work* per una durata media di circa sette mesi e mezzo, rispetto ai 12 previsti come standard.

²⁵ *MDRC* resta tuttora l'organizzazione *leader* nel settore della valutazioni di interventi-pilota nel campo delle politiche sociali. Si veda il sito www.mdrc.org.

L'aspetto che più ci interessa di questa *demonstration* è il fatto che è accompagnata, fin dalla fase iniziale di progettazione, da una valutazione della sua fattibilità e della sua efficacia. La finalità principale dell'intera *demonstration* è infatti di tipo conoscitivo, e più precisamente quella di fornire una risposta a questo tipo di interrogativi:

- qual è l'efficacia del *supported work* nell'aumentare la partecipazione al lavoro di persone ai margini del mercato del lavoro e nel ridurre la loro dipendenza dai sussidi pubblici?
- quale tipologia di persone svantaggiate trae maggior beneficio da questo intervento?
- quanto costa operare un simile tipo di intervento?
- qual è il rapporto tra costi sostenuti e benefici ottenuti?

Come si vede, si tratta di un ampio spettro di interrogativi, che va al di là del semplice effetto diretto dell'intervento e copre anche questioni legate all'implementazione, cioè a come l'intervento viene concretamente realizzato. Ma in tutti i casi si tratta di interrogativi che scaturiscono da una genuina incertezza sull'efficacia e sulla fattibilità di un intervento di questo tipo: le risposte a questi interrogativi sono destinate ad informare il processo decisionale pubblico qualora si preveda di realizzare l'intervento su larga scala.

La valutazione degli effetti riguarda soprattutto il primo degli interrogativi elencati sopra. Per rispondere a questo interrogativo è stato implementato il disegno sperimentale con gruppo di controllo: in 10 dei 14 siti in cui è attuata la *demonstration*, una percentuale delle persone inviate dai servizi sociali ai laboratori viene esclusa dalla partecipazione mediante sorteggio. Coloro che sono esclusi dal *supported work* rappresentano il gruppo di controllo. Da un punto di vista etico e legale, tale esclusione non è un problema insormontabile, in quanto la partecipazione non rappresenta alcun tipo di diritto per coloro che fanno domanda. Anzi, se non ci fosse stata l'intenzione di condurre la valutazione sperimentale, l'intervento stesso non sarebbe stato realizzato (se non altro perché sarebbero mancati i finanziamenti messi a disposizione dal governo federale e dalle fondazioni) e quindi nessuno avrebbe usufruito di questa esperienza.

Circa 6.600 individui vengono coinvolti nel processo di randomizzazione e di essi circa 3.200 sono assegnati al gruppo di trattamento e 3.400 al gruppo di controllo. Ai primi viene consentito di partecipare al *supported work*, mentre i secondi vengono esclusi. Solo un terzo dei 10.000 partecipanti ai lavoratori viene quindi utilizzato per la valutazione degli effetti (come gruppo sperimentale). Questo perché i fondi allocati per la valutazione consentono di condurre interviste (ripetute nel tempo ogni nove mesi) solamente su circa 6.500 individui, mentre i 10.000 partecipanti, suddivisi su 14 siti, sono considerati il minimo indispensabile per poter trarre lezioni utili sulle sfide organizzative che pone l'implementazione di un intervento sociale di questo tipo.

I 6.600 individui appartenenti al gruppo sperimentale o al gruppo di controllo vengono sottoposti a cinque *round* successivi di interviste, cominciando con un'intervista immediatamente prima della randomizzazione, seguita da un'intervista ogni nove mesi: l'ultima si svolge a circa tre anni di distanza dall'inizio dell'esperimento. Queste interviste vengono utilizzate per rilevare come variano nel tempo i comportamenti e le condizioni che l'intervento vuole modificare, cioè la partecipazione al lavoro, il reddito da lavoro, lo stato di povertà, la percezione di sussidi pubblici, nonché l'uso di stupefacenti e la commissione di reati.

Ci limiteremo qui, a scopo esemplificativo, ad esaminare l'effetto del *supported work* sulla partecipazione al lavoro e sul reddito da lavoro per due delle quattro sottopopolazioni coinvolte nell'esperimento, le madri non sposate che vivono del sussidio di povertà e i giovani *drop-out* della scuola dell'obbligo. Questi risultati sono interessanti non solo per le stime che presentano, ma anche perché mostrano chiaramente che senza un gruppo di controllo è facile attribuire all'intervento miglioramenti che sono invece dovuti a processi di cambiamento e maturazione indipendenti dall'intervento.

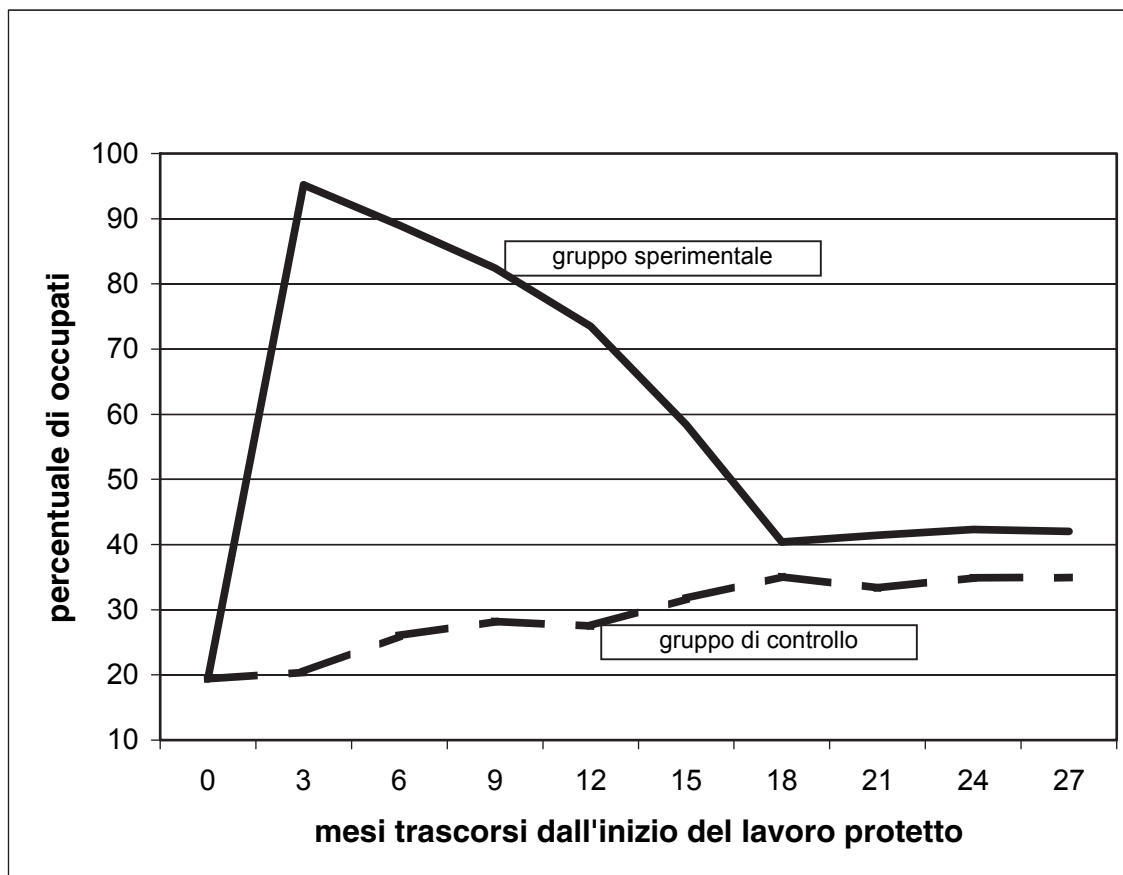
La Figura 3.1 presenta uno dei principali risultati della valutazione per il gruppo delle madri non sposate che ricevono il sussidio di povertà definito *Aid to Families with Dependent Children (AFDC)* da almeno 3 anni. La variabile-risultato rispetto a cui si intende misurare l'effetto della politica è rappresentata in questo caso dal tasso di occupazione, definito come la percentuale di individui con un lavoro in un certo mese.

Questo tasso è riportato sull'asse verticale, mentre sull'asse orizzontale troviamo il numero di mesi trascorsi dal momento della randomizzazione.

La linea continua in alto rappresenta sinteticamente l'esperienza lavorativa delle circa 700 madri *single* che hanno partecipato al laboratorio (in sei distinte città) e che hanno risposto alle interviste. Il loro tasso di occupazione precedente all'esperienza di lavoro protetto è di circa il 20 per cento, che sottolinea lo stato di emarginazione dal lavoro di questo gruppo (le ore lavorate alla settimana dalla piccola percentuale che lavora del tutto sono in media circa 20). La partecipazione al *supported work* si manifesta immediatamente con un salto verso l'alto del tasso di occupazione, che dopo tre mesi sfiora il 100 per cento. A queste donne è stato offerto un lavoro, per giunta in un contesto relativamente facile e poco stressante e quasi tutte lo hanno accettato.

Questa percentuale di occupate decresce rapidamente con il passare dei mesi (la durata media della permanenza nel laboratorio delle madri *single* di 9 mesi e mezzo): scende al 75 per cento dopo 12 mesi e al 40 per cento dopo 18 mesi, quando praticamente nessuna di queste donne è più coinvolta in attività di lavoro protetto. Quindi quel 40 per cento di occupate, che cresce ancora leggermente nei nove mesi successivi, rappresenta occupazione regolare, al di fuori della struttura protetta, il che rappresenta l'obiettivo di fondo dell'esperimento: facilitare l'inserimento di queste persone in una situazione lavorativa "normale".

Figura 3.1: Tasso di occupazione delle madri *single* con sussidio di povertà da più di 3 anni



Fonte: Manpower Demonstration Research Corporation (1980)

La valutazione potrebbe concludersi con questo risultato, che fa registrare più di un raddoppio del tasso di occupazione di questo gruppo, dal 20 per cento prima dell'esperienza di lavoro protetto al 42 per cento all'ultima rilevazione a 27 mesi dall'inizio dell'esperimento. Quindi un netto successo dell'intervento pubblico, che è riuscito con pochi mesi di occupazione sussidiata a raddoppiare la partecipazione al mercato del lavoro di questo gruppo socialmente emarginato.

Una tale conclusione tuttavia non sarebbe coerente con la definizione di "effetto" di una politica data nel capitolo precedente. Abbiamo infatti definito effetto come la differenza tra il valore della variabile-risultato (la percentuale di occupati, in questo caso) osservato *dopo* l'attuazione della politica pubblica e il valore che si sarebbe osservato *in sua assenza*. Ma la nozione intuitiva di successo della politica appena utilizzata è basata su un confronto tra il valore della variabile-risultato osservato *dopo* l'attuazione della

politica pubblica e il valore osservato *prima*. Il confronto pre-post non rivela l'effetto della politica, a meno che non si voglia ipotizzare che la situazione osservata prima dell'intervento (il 20 per cento di occupati) sarebbe rimasta immutata due anni e mezzo dopo.

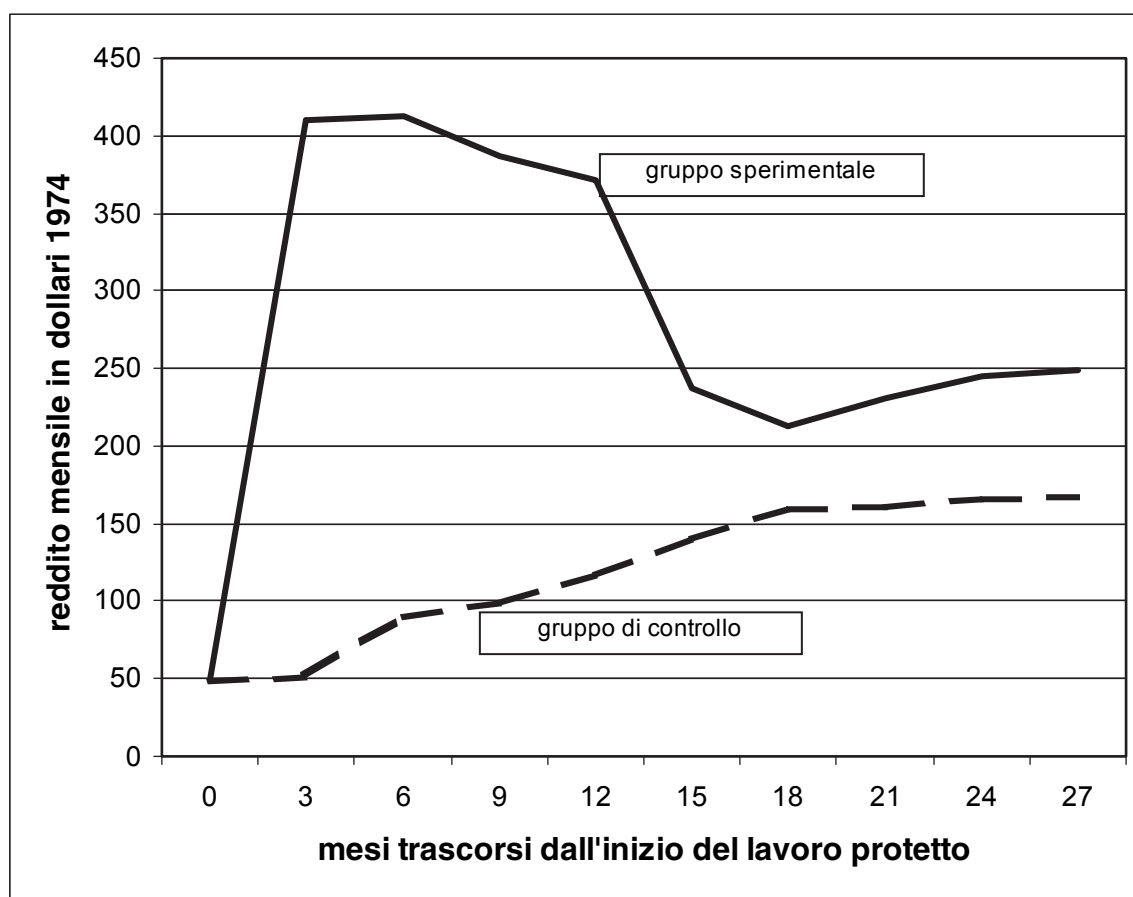
Non è tuttavia necessario fare questa ipotesi (peraltro poco plausibile), grazie alla presenza del gruppo di controllo. Si tratta di un numero analogo di madri *single*, con caratteristiche in tutto e per tutto simili alle partecipanti al lavoro protetto, che però, dopo essere state inviate dai servizi sociali al laboratorio del *supported work* con la prospettiva di ottenere un lavoro, se lo sono visto rifiutare a causa del sorteggio sfavorevole. Si può argomentare che questa convocazione seguita da una delusione abbia rappresentato di per sé un "trattamento", che può avere alterato il loro percorso lavorativo: ciò appare tuttavia alquanto improbabile, e senza dubbio rappresenta un evento marginale se confrontato con l'esperienza di lavoro a tempo pieno fatta per parecchi mesi dai membri del gruppo sperimentale.

Tornando alla Figura 3.1, la linea tratteggiata rappresenta il tasso di occupazione dei "controlli", registrato durante interviste del tutto identiche a quelle condotte con i membri del gruppo sperimentale. Come si può facilmente vedere, pur in assenza di lavoro protetto, il tasso di occupazione dei controlli è aumentato progressivamente durante i 27 mesi del periodo di osservazione, fino ad arrivare al 35 per cento. La differenza tra trattati e controlli al 27° mese dalla randomizzazione è di soli 7 punti percentuali (42% – 35%).

Se siamo disposti ad accettare questa differenza come una stima plausibile dell'effetto dell'intervento, dobbiamo concludere che la partecipazione al lavoro protetto ha fatto aumentare l'occupazione di chi vi ha partecipato di circa il 20 per cento rispetto al livello pre-intervento: un risultato positivo, ma nettamente inferiore a quello che avremmo se considerassimo la differenza pre-post come effetto dell'intervento. In altre parole, di quei 22 punti percentuali di incremento nel tasso di occupazione fatto registrare dal gruppo sperimentale, 15 punti di aumento si sarebbero osservati comunque, come suggerisce l'esperienza del gruppo di controllo, mentre solo i 7 punti addizionali sono plausibilmente da attribuirsi alla politica pubblica che stiamo valutando.

Si può obiettare a questo intero ragionamento che l'occupazione *tout court* non è l'unico obiettivo di una politica di reinserimento lavorativo, soprattutto quando essa mira a recuperare situazioni di dipendenza da sussidi pubblici e quindi di povertà. Il reddito da lavoro è una misura migliore di successo, in quanto incorpora più dimensioni della performance lavorativa di una persona: non solo il fatto di avere un lavoro, ma anche la continuità dell'occupazione, le ore lavorate settimanalmente e la retribuzione oraria. L'analisi svolta sul tasso di occupazione può essere replicata, seguendo la stessa logica, utilizzando il reddito da lavoro mensile come variabile-risultato. Questa analisi è mostrata graficamente nella Figura 3.2.

Figura 3.2: Reddito da lavoro delle madri *single* con sussidio di povertà da più di 3 anni



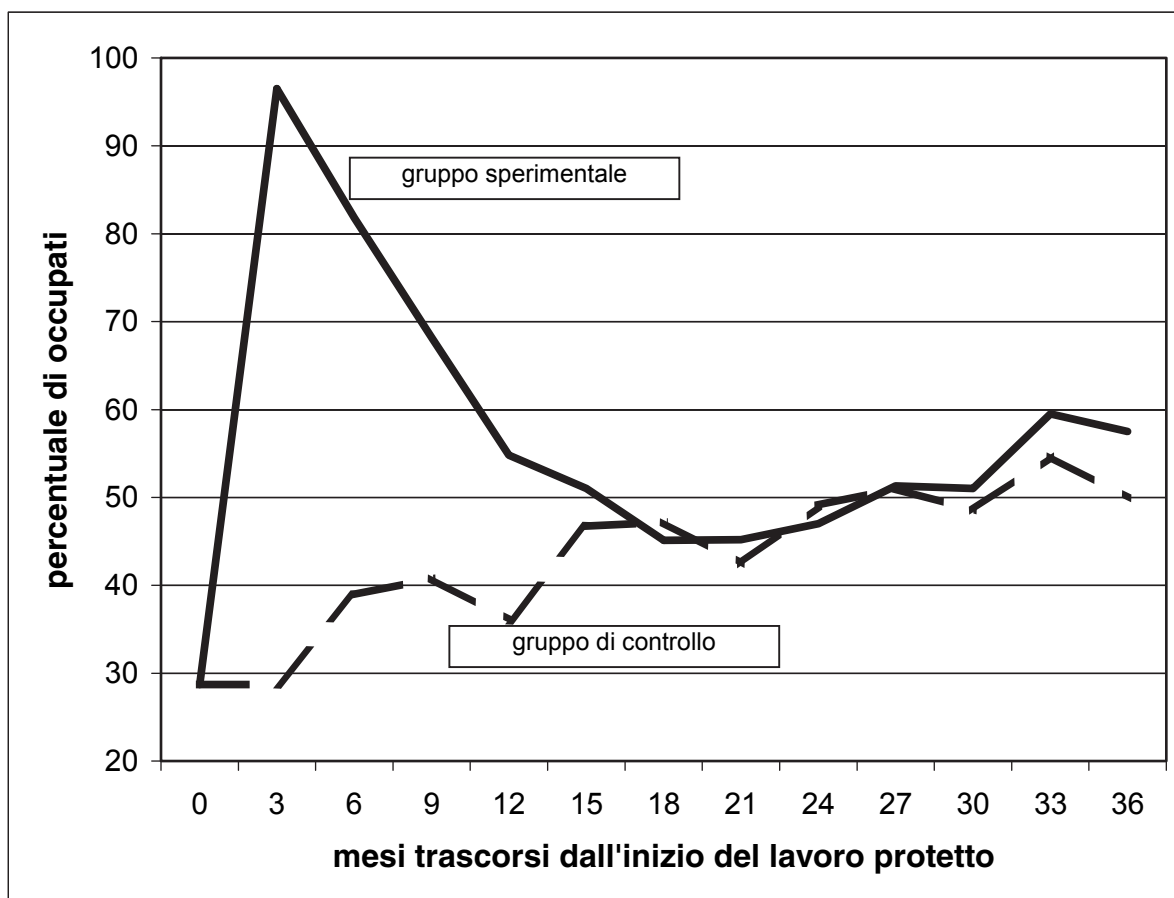
Fonte: Manpower Demonstration Research Corporation (1980)

Il reddito medio mensile (considerando come zero il reddito di chi non lavora) di entrambi i gruppi è di circa 50 dollari al mese all'inizio dell'esperimento. (Questa

uguaglianza, ricordiamolo, è conseguenza diretta della randomizzazione, che crea due gruppi equivalenti). Dopo 27 mesi, la retribuzione del gruppo sperimentale sale a 250 dollari mensili, un aumento proporzionalmente superiore a quello fatto registrare dal tasso di occupazione, il che implica che vi siano anche stati miglioramenti sul lato delle ore lavorate e della retribuzione oraria. I membri del gruppo di controllo non sono stati tuttavia inerti, facendo anzi registrare alla fine del periodo di osservazione una retribuzione media di circa 170 dollari mensili. L'effetto netto sulla retribuzione media mensile derivante dall'esperienza di *supported work* è quindi un aumento del 47 per cento: un risultato considerevole, ma nettamente inferiore all'aumento di quasi il 500 per cento che avremmo ottenuto considerando l'aumento osservato per il solo gruppo sperimentale.

Oltre che per diverse variabili-risultato, l'analisi degli effetti di *supported work* è stata condotta per le altre categorie di soggetti coinvolti nell'esperimento. La Figura 3.3 presenta i risultati relativi al tasso di occupazione osservato per i giovani *drop-out* della scuola dell'obbligo (per la precisione, coloro che abbandonano la scuola dopo il 10° anno di istruzione, l'equivalente del nostro secondo anno delle scuole superiori).

Figura 3.3: Tasso di occupazione per i giovani drop-out della scuola dell'obbligo



Fonte: Manpower Demonstration Research Corporation (1980)

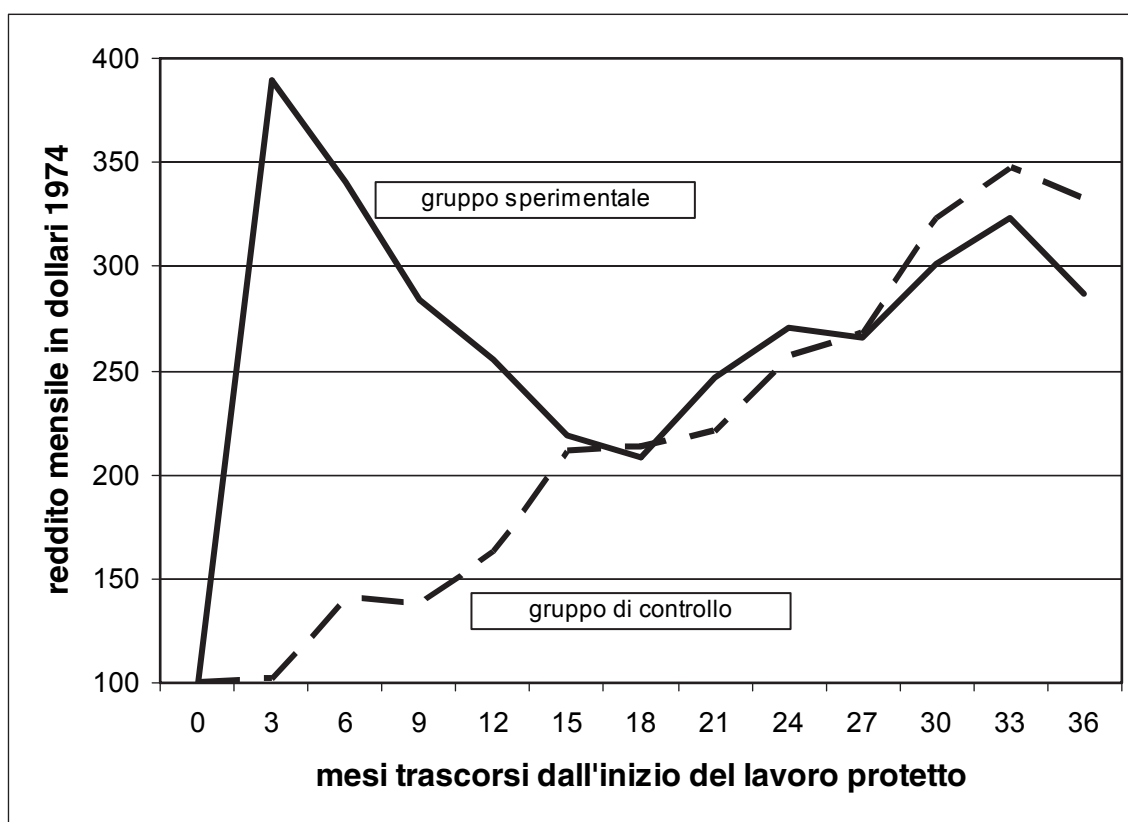
Il gruppo dei circa 1200 giovani *drop-out*, come i gruppi di ex-detenuiti e ex-tossicodipendenti, è stato osservato per 36 mesi invece di 27 e quindi intervistato una volta di più rispetto al gruppo delle donne *single*. L'andamento dell'occupazione del gruppo sperimentale mostra un netto aumento del tasso di occupazione, seguito da un'altrettanto rapido declino. Dal 18° mese in poi si osserva un lento aumento per i 18 mesi successivi, che porta il tasso di occupazione dal 45 a quasi il 60 per cento.

Tuttavia, dal 18° mese in poi l'andamento del gruppo di controllo è fondamentalmente identico a quello del gruppo sperimentale: nella logica controfattuale, questo implica che l'effetto dell'intervento sia sostanzialmente nullo. Quindi in questo caso non si è solo ridotta drasticamente la stima dell'effetto della politica, ma si è annullata. Questa

politica di reinserimento, quando applicata ai giovani, chiaramente non aiuta a migliorare la loro posizione nel mercato del lavoro.

Questo verdetto negativo è confermato anche dall'analisi dell'andamento del reddito, rispetto al quale il gruppo di controllo sembra addirittura superare il gruppo sperimentale negli ultimi mesi del periodo di osservazione, come mostrato nella Figura 3.4.

Figura 3.4: Reddito da lavoro dei giovani drop-out della scuola dell'obbligo



Fonte: Manpower Demonstration Research Corporation (1980)

L'effetto di un intervento di questo tipo può anche essere stimato su altre dimensioni, utilizzando cioè altre variabili-risultato. Alcune di esse sono misurabili, come il coinvolgimento in attività criminali o l'uso di stupefacenti. Vi sono poi dimensioni non facilmente quantificabili, quali l'autostima e la capacità di interagire socialmente, che è possibile siano positivamente influenzate dalla partecipazione ad un'esperienza di lavoro protetto. Tuttavia, queste dimensioni non sono state misurate durante le

interviste. Se lo fossero state, per determinare l'effetto dell'intervento su di esse sarebbe stato necessario un tipo di analisi del tutto analogo a quello condotto per il tasso di occupazione e per il reddito: calcolare la differenza tra le variabili-risultato osservate nei due gruppi.

Partendo dalle stime degli effetti dell'intervento, i valutatori hanno condotto un'analisi costi-benefici della partecipazione al *supported work*, separatamente per le quattro tipologie di soggetti coinvolti. Va sottolineato come l'analisi costi-benefici possa essere condotta solo dopo che è stata effettuata l'analisi degli effetti, in quanto il principale elemento che entra nel calcolo sul lato dei benefici è proprio l'effetto netto dell'intervento, in termini di aumento di reddito e diminuzione di dipendenza dai sussidi pubblici. Tutte queste dimensioni vanno monetizzate, in modo da poter essere aggregate in un'unica misura, attualizzate al presente mediante un opportuno tasso di sconto e confrontate ai costi dell'intervento.

Per i giovani *drop-out*, i ricercatori concludono che, anche sotto diversi scenari, i benefici prodotti dal *supported work* non sono in grado di giustificarne il costo. Per questo gruppo, il valore attualizzato dei benefici netti è negativo, variando tra - 4.000 e - 250 dollari a seconda degli assunti fatti, e in nessun caso diventa positivo. Questo risultato è determinato in larga parte da quanto illustrato nella Figura 3.4, cioè dal fatto che la partecipazione al *supported work* non *causa* alcun aumento del reddito da lavoro di questi giovani.

Un risultato diverso si ottiene dall'analisi costi-benefici applicata al gruppo delle madri non sposate. In questo caso il valore attualizzato dei benefici netti varia da 2.700 a 9.000 dollari. La realizzazione del *supported work* per questo particolare gruppo si rivela quindi un buon investimento dal punto di vista della società.

Questa conclusione ovviamente è difficile da generalizzare dall'ambito ristretto di un intervento-pilota a quello di un intervento realizzato su vasta scala. È possibile che l'estensione di un intervento del genere ad una scala più ampia incontri difficoltà organizzative che ne riducano l'efficacia; è possibile che la sua capacità di penetrazione nella platea dei potenziali beneficiari si riveli molto limitata; è possibile infine che il successo ottenuto nell'aumentare la partecipazione al lavoro di un piccolo gruppo di

individui non sia replicabile su larga scala, a causa della limitata capacità da parte del mercato del lavoro di assorbire questo tipo di lavoratori svantaggiati. Restano comunque lezioni importanti da questo esperimento: ad esempio, il fatto che, per i giovani, il coinvolgimento in un'esperienza di questo tipo non pare produrre benefici sufficienti a giustificarne i costi.

I risultati positivi della *NSWD* spingono molti alla fine degli anni Settanta a raccomandare l'applicazione di questo tipo di intervento su larga scala, come tentativo di soluzione al cronico problema dell'emarginazione dal mercato del lavoro da parte delle madri che dipendono per anni dal sussidio di povertà. Queste raccomandazioni si scontrano però con un brusco cambiamento di clima, seguito all'elezione del presidente Reagan nel 1980. Uno dei cavalli di battaglia della piattaforma di Reagan è l'abolizione di ogni forma di *subsidized employment* (cioè di lavoro il cui costo è almeno in parte coperto da un contributo pubblico), con l'obiettivo di porre fine a forme di abuso e di spreco che sono emerse durante l'amministrazione Carter, ricevendo vasta eco sulla stampa e durante la campagna elettorale. Il *supported work* non è parte di tali forme di abuso, ma diventa vittima innocente di tale pubblicità negativa e viene sostanzialmente abbandonato.

3.2 I limiti del metodo sperimentale

Sulla base di quanto detto finora, il metodo sperimentale può apparire in una luce del tutto positiva, capace di dare una traduzione pura e immediata della nozione di valutazione dell'effetto di una politica come differenza tra osservato e controfattuale. In questa immediatezza e purezza concettuale sono individuabili, infatti, la forza e il fascino di questo metodo. È venuto però il momento di metterne in luce i numerosi limiti.

A scopo di premessa, diremo che da molti anni è in corso un acceso dibattito tra fautori e detrattori del metodo sperimentale, dibattito che spesso assume i toni di una polemica ideologica. Un esempio di opinione decisamente negativa lo offre Claudio Bezzi²⁶.

²⁶ Bezzi C., *Il disegno della ricerca valutativa*, FrancoAngeli, Milano, 2003.

“Benché il disegno sperimentale sia contemplato in tutta la letteratura metodologica, non ci si può nascondere la sua difficile realizzabilità di fatto quando si lavora nell'ambito delle scienze sociali. La ricerca sperimentale (...) presuppone infatti il controllo ferreo della clausola *ceteris paribus*, realizzabile sostanzialmente (chiudendo almeno un occhio) in piccoli gruppi e con esperimenti di laboratorio, di cui è lecito dubitare la riproducibilità in un contesto *reale*. (...) I disegni sperimentali sono essenzialmente dei modelli di studio, finanziati e realizzati probabilmente in ambiente universitario, e non dei modelli operativi finanziati e realizzati da committenti impegnati in un programma *reale*” (pag. 336-338).

“Difficile realizzabilità”, “di cui è lecito dubitare la riproducibilità”, “modelli di studio realizzati probabilmente in ambiente universitario”. Di parere diametralmente opposto è invece Larry Orr, uno dei massimi protagonisti di questa esperienza negli Stati Uniti, che scrive²⁷:

“Nei 30 anni successivi all'introduzione della sperimentazione sociale, questo metodo è stato ampiamente accettato come il “*gold standard*” della valutazione delle politiche (...) questo insieme di metodi consolidati è frutto del grande numero di esperimenti realizzati durante questo periodo (...) molto poco su questi metodi è stato pubblicato nelle riviste accademiche” (pag. 2).

Quindi metodi consolidati e ampiamente accettati, frutto del gran numero di esperimenti realizzati, ma svolti al di fuori dell'ambiente accademico, che in effetti ha prestato relativamente poca attenzione a questo fenomeno e si è concentrato più sullo sviluppo di metodi non-sperimentali.

Un giudizio più equidistante lo esprimono Rossi, Freeman e Lipsey nel loro classico trattato, giunto ormai alla sesta edizione²⁸:

“L'esperimento con randomizzazione è il più solido disegno di ricerca per valutare l'impatto netto degli interventi. Gli esperimenti con randomizzazione hanno limiti di applicazione, tra cui quello di essere utilizzabili solo per programmi a copertura parziale. Inoltre, il loro uso è ulteriormente limitato da problemi pratici relativi all'ottenimento della collaborazione tra i vari attori e a i tempi e ai costi necessari a condurli” (pag. 279).

²⁷ Orr L., *Social Experiments: Evaluating Public Programs with Experimental Methods*, Sage Publications, Beverly Hill, 1999.

²⁸ Rossi P., Freeman R., Lipsey M., *Evaluation: A Systematic Approach*, Sage Publications, Beverly Hill, 2003.

Il “più solido disegno di ricerca” e “limiti di applicazione”: queste due affermazioni non sono in contraddizione. In teoria, il disegno sperimentale è il modo più plausibile per derivare stime credibili (e facilmente comunicabili) dell’effetto di una politica sui suoi destinatari. In concreto, le difficoltà nell’applicarlo sono tali da limitarne la rilevanza pratica, persino negli Stati Uniti, dove questo approccio ha raggiunto il massimo sviluppo.

Un primo limite è quello menzionato da Rossi e i suoi colleghi: l’applicabilità del metodo sperimentale è limitata alle *politiche non universali*, cioè a copertura parziale. Tutti gli interventi universali (quali l’istruzione elementare, gli interventi di tutela ambientale, le regole sulla circolazione stradale) o comunque tutti quegli interventi pubblici in cui è impossibile definire una fruizione individuale, non si prestano per definizione all’individuazione di un gruppo di controllo e quindi alla randomizzazione.

Questa affermazione va tuttavia qualificata, in quanto è rilevante più a livello di singoli individui che a livello di aree geografiche: in molti casi in cui il singolo individuo non può essere escluso da un intervento per la natura collettiva di quest’ultimo, la singola comunità potrebbe invece esserne esclusa. Nell’esempio che introdurremo nel prossimo capitolo, gli interventi di recupero urbano sono attuati a livello di singolo quartiere: il singolo abitante del quartiere non può essere escluso dall’intervento di recupero urbano, ma l’intero quartiere sì. Non si può randomizzare a livello di singolo cittadino, ma (almeno in teoria) si può randomizzare a livello di singolo quartiere.

Anche tenuto conto di questa precisazione, resta comunque il fatto che, per una lunga serie di motivi tecnici, la creazione di un gruppo di controllo mediante randomizzazione non sia fattibile in un ampio spettro di politiche pubbliche. Questo è un fatto ampiamente riconosciuto anche dai più appassionati fautori degli esperimenti, ma che non detrae dall’utilità del metodo *dove esso sia tecnicamente fattibile*: laddove cioè la politica pubblica si concretizza in un intervento ben definito nel tempo e nello spazio, con dei destinatari chiaramente definibili (a livello individuale o di gruppo o di comunità). Nel campo dell’istruzione, della formazione al lavoro, della sanità, dell’assistenza sociale, delle politiche del lavoro e delle politiche industriali, numerosi

sono gli esempi di interventi pubblici che consentirebbero l'utilizzo del metodo sperimentale.

Tuttavia, anche dove il metodo sperimentale è *tecnicamente* fattibile, esso finisce spesso per non essere fattibile per ragioni di tipo etico, legale o politico. Il profilo etico è il più delicato: molti infatti ritengono eticamente inaccettabile escludere alcuni individui dalla fruizione di un servizio, da cui potrebbero (forse) trarre qualche beneficio, al solo scopo di accumulare conoscenza collettiva sull'efficacia del servizio stesso. Altri ritengono questo sacrificio individuale accettabile, in virtù dell'incerta utilità stessa del servizio: se non esistesse questo dubbio sull'efficacia, non ci sarebbe ragione di proporre la valutazione degli effetti. Questo dilemma si pone anche nel caso dei *clinical trials*: con le dovute cautele, in questo ambito il dilemma è stato però risolto in favore dell'accettabilità del metodo sperimentale. Nel caso degli esperimenti sociali, soprattutto in Europa, la bilancia sembra pendere nella direzione opposta.

Su questa avversione al metodo sperimentale in ambito sociale pesa tuttavia non solo il mero profilo etico, ma anche un fatto più concreto: la tipica avversione alla randomizzazione da parte degli operatori dei servizi coinvolti. A questo alludono Rossi, Freeman e Lipsey quando dicono “il loro uso è ulteriormente limitato da problemi pratici relativi all'ottenimento della collaborazione tra i vari attori”. Anche negli Stati Uniti, la difficoltà pratica più rilevante nel realizzare un esperimento è ottenere la collaborazione dagli operatori dei servizi, a cui si chiede di escludere a caso una quota dei loro utenti: questo genera ogni sorta di malumori e lamentele che ricadono *in primis* sugli operatori stessi. Si chiede quindi a qualcuno di sopportare un costo immediato notevole (almeno in termini di stress lavorativo) in nome di una “causa” dai confini abbastanza sfocati e che, per di più, trova la sua motivazione nel *dubbio* sull'efficacia del servizio, un dubbio che gli operatori raramente condividono in quanto identificano la propria professionalità con il successo stesso del servizio.

Alle difficoltà di questo tipo si aggiungono spesso vincoli di tipo legale, quando l'erogazione del servizio (o del sussidio o dell'agevolazione) sia prevista dalla legislazione vigente. Questo non è un ostacolo del tutto insormontabile, almeno negli Stati Uniti, grazie alla emanazione di apposite esenzioni legislative (*waiver*) che

consentono di differenziare l'erogazione dei servizi se ciò è fatto a scopo di valutazione. Piuttosto che legale, il problema però diventa "politico". Anche se legalmente possibile, diventa politicamente difficile difendere l'esclusione di utenti dalla fruizione di servizi ormai andati a regime e considerati quindi come un diritto acquisito da parte degli utenti. L'incertezza sulla loro utilità ed efficacia, se da un lato motiva la valutazione, dall'altro non serve a placare il cittadino escluso dal servizio: dal suo punto di vista, resta comunque una possibilità che il servizio sia utile, e questa possibilità gli verrebbe del tutto negata dall'esclusione *tout court* dalla fruizione.

Questo complesso di problemi etici, legali e politici diventa più facile da superare quando la randomizzazione sia utilizzata nell'ambito di interventi-pilota implementati su scala ridotta (*demonstration*), cioè di interventi che sono concepiti e attuati con lo scopo specifico di valutarne la fattibilità e l'efficacia. Ad esempio, in un caso come la *NSWD*, problemi di tipo etico non si pongono neppure: l'esclusione dei 3.400 "controlli" è più che giustificata dall'inclusione di 10.000 beneficiari: se non fosse stato per la valutazione, neppure i beneficiari ammessi avrebbero usufruito del *supported work*. Inoltre, in situazioni del genere ottenere la collaborazione degli operatori sociali nella gestione dell'esperimento è molto più facile, in quanto è chiaro fin dall'inizio che la randomizzazione è parte integrante della politica stessa. Infine, problemi legali non se ne pongono perché interventi di questo tipo o non sono previsti da alcuna legge o sono espressamente autorizzati tramite *waiver* legislativi.

La maggioranza delle valutazioni sperimentali condotte negli Stati Uniti negli ultimi quarant'anni sono state infatti effettuate nell'ambito di *demonstrations*. Esistono eccezioni importanti, quali la valutazione del *Job Training Partnership Act (JTPA)*, il sistema federale di formazione e servizi all'impiego per le persone economicamente svantaggiate, valutazione effettuata mediante un esperimento con randomizzazione nei primi anni Novanta²⁹. Sedici delle circa 600 unità locali che erogavano questi servizi di supporto all'impiego furono scelte (dopo molto difficoltà) per l'esperimento e circa 20.000 utenti di queste 16 unità sottoposti a randomizzazione: circa 6.000 andarono a formare il gruppo di controllo e non venne loro consentito di usufruire dei servizi

²⁹ I risultati sono illustrati in dettaglio in Orr *et al.*, *Does Training for the Disadvantaged Work? Evidence from the National JTPA Study*, The Urban Institute Press, Washington, 1996.

all'impiego per un periodo di 18 mesi. Per dimensione ed estensione territoriale si è trattato del più rilevante esperimento con randomizzazione attuato su una politica pubblica già a regime.

L'analisi dei dati di *JTPA*, effettuata mediante il semplice confronto tra gruppo sperimentale e gruppo di controllo, ha mostrato un risultato analogo a quello della *NSWD*: un effetto positivo dei servizi sul reddito da lavoro degli adulti (più per le donne che per gli uomini) e un effetto nullo per i giovani (in particolare per i maschi). Questi risultati hanno avuto una importante ricaduta a livello decisionale, diversamente dal caso della *NSWD*: come dice il rapporto finale della valutazione di *JTPA*, “la chiarezza e il rigore metodologico di questo studio lo hanno reso un fattore importante nel determinare quale componente di *JTPA* è sopravvissuta ai tagli operati dalla legge finanziaria del '95”.

Il diverso utilizzo dei risultati della valutazione di *JTPA* e della *NSWD* è dovuta in parte al diverso clima politico prevalente nel momento in cui i risultati diventano disponibili, ma in parte anche alla loro diversa capacità di produrre risultati “generalizzabili”. Quello della scarsa generalizzabilità dei risultati è infatti un altro limite importante della valutazione sperimentale, in particolare quando sia fatta nell'ambito (ad essa peraltro più congeniale) degli interventi-pilota.

Una distinzione classica a questo proposito è quella tra validità interna e validità esterna. La randomizzazione serve a garantire la *validità interna* delle stime, cioè il fatto che le stime degli effetti riflettano realmente il contributo netto dell'intervento alla modifica della situazione che si sarebbe verificata in assenza di intervento. Da questo punto di vista la sperimentazione con randomizzazione ha le carte in regola, e tranne pochi critici, tutti le riconoscono questa superiorità.

Dove la sperimentazione mostra la corda è sul lato della *validità esterna*, cioè della generalizzabilità delle stime: in particolare quando ottenute nell'ambito di interventi realizzati per un periodo limitato di tempo e su scala ridotta, queste stime sono difficili da generalizzare alla politica pubblica che deve andare a regime ed essere implementata su scala vasta. In particolare, gli effetti stimati su scala ridotta non tengono conto dei vincoli di tipo macro che diventano evidenti quando l'intervento è generalizzato.

Inoltre, l'implementazione su scala ridotta è sempre più agevole dal punto di vista organizzativo dell'implementazione su scala vasta.

Esiste quindi un *trade-off* tra i due tipi di validità, interna ed esterna. Si possono ottenere stime del tutto corrette degli effetti dell'intervento nell'ambito ristretto in cui sono ottenute, ma difficili da generalizzare al di là di questo ambito: questo è in una certa misura il caso di *NSWD*. Viceversa, si possono ottenere stime più dubbie dal punto di vista della validità interna, per le maggiori difficoltà di gestione dell'esperimento in un contesto "reale", ma più generalizzabili alla politica, in quanto ottenute randomizzando utenti "reali" in un contesto "reale". Questo è il caso della valutazione di *JTPA*.

Va, infine, menzionata un'altra difficoltà nell'utilizzo della valutazione sperimentale: la difficoltà di mantenere l'*integrità* dell'esperimento, in particolare quella di mantenere nel tempo la separazione tra gruppo di controllo e gruppo sperimentale. L'essenza della randomizzazione è la divisione del campione in due gruppi, di cui uno è esposto alla politica e uno no. Al fine di ottenere stime corrette dell'effetto, è indispensabile che gruppo sperimentale e gruppo di controllo *restino tali* per l'intera durata dell'esperimento: cioè è importante evitare che i membri del gruppo dei trattati smettano di ricevere il trattamento da un lato, e dall'altro che i membri del gruppo di controllo ricevano un trattamento simile da qualche altra amministrazione.

Questa integrità è relativamente più facile da ottenere nel caso dei *clinical trials*, anche grazie al fatto che il gruppo di controllo è "cieco" e che l'assunzione di un farmaco in gran parte dei casi non è un'attività impegnativa o alternativa rispetto ad altre. Il problema è invece più serio nel caso degli esperimenti sociali: i membri di ciascun gruppo fanno infatti perfettamente a che regime appartengono e la fruizione di un servizio è un'attività spesso impegnativa a cui alcuni membri del gruppo sperimentale finiscono talvolta per sottrarsi.

Emergono quindi due tipi di problemi endemici nella gestione degli esperimenti, che possono mettere in discussione persino la validità interna delle stime ottenute. Da un lato il problema della non partecipazione o partecipazione incompleta da parte di coloro che sono stati *assegnati* al gruppo di controllo (il cosiddetto problema dei *no-show*,

letteralmente “quelli che non si presentano”); dall’altro il problema dei *cross-over*, cioè di quei membri del gruppo di controllo che riescono a usufruire egualmente del servizio o di un servizio analogo³⁰. Esistono tecniche statistiche per correggere queste distorsioni: ma si tratta di tecniche che si basano su assunti non verificabili, simili a quelli inerenti ai metodi non-sperimentali.

³⁰ La lista delle possibili critiche al metodo sperimentale non finisce qui. Come si è detto in precedenza, il dibattito tra sostenitori e detrattori della sperimentazione sociale è tuttora in corso, senza vincitori né vinti. Se il lettore vuole approfondire i termini di questo dibattito, due utili riferimenti sono i seguenti: (i) James Heckman e Jeffrey Smith, con “Assessing the Case for Social Experiments” *Journal of Economic Perspectives*, 1995, offrono una critica implacabile agli argomenti di solito portati a favore dell’uso della randomizzazione come metodo di valutazione. (ii) Ray Pawson e Nick Tilley, nel saggio *Realistic Evaluation*, Sage Publications, 1997, offrono una critica del metodo sperimentale molto diversa rispetto alla precedente. La loro critica è rivolta infatti a tutta la valutazione quantitativa degli effetti delle politiche, perché, nelle loro parole “*experimentalists have pursued too single-mindedly the question of whether a program works, at the expense of knowing why it works*”. Secondo i due autori, sapere se una politica ha avuto un certo effetto è secondario, quello che conta veramente è sapere *perché* ha (o non ha) un effetto. Questa tesi lascia un dubbio irrisolto: se non si sa neppure se una politica un effetto l’ha avuto oppure no, come si fa orientare la riflessione sul *perché* ce l’ha o non ce l’ha?

CAPITOLO 4

I CONFRONTI SPAZIO-TEMPORALI CON DATI NON- SPERIMENTALI E LE MINACCE ALLA LORO VALIDITÀ

di Alberto Martini e Marco Sisti

Quello sperimentale è ritenuto il metodo in grado di produrre le stime più valide e credibili degli effetti di una politica pubblica, almeno dal punto di vista della validità interna. Per ottenere tale risultato conoscitivo è necessario però che sia soddisfatta una condizione non da poco: che il valutatore abbia la possibilità di *manipolare* il processo di selezione, e quindi di determinare mediante la randomizzazione chi è ammesso al servizio, o al sussidio, o all'agevolazione e chi viceversa ne è escluso. La possibilità di operare tale manipolazione è tuttavia necessariamente limitata per ragioni pratiche, legali o politiche, anche in contesti favorevoli culturalmente alla sperimentazione sociale, quale quello statunitense.

Quando il metodo sperimentale non è praticabile e quindi il processo di selezione non è manipolabile dal valutatore, a quest'ultimo non resta che utilizzare gli esiti dei processi che avvengono "naturalmente" nel corso dell'attuazione di una politica pubblica, frutto delle decisioni dei potenziali beneficiari della politica pubblica oppure di coloro che la disegnano e la gestiscono. In altre parole, il valutatore deve utilizzare uno dei metodi di valutazione definiti collettivamente come *non-sperimentali*³¹.

Va sottolineato come, anche in una situazione non-sperimentale, l'effetto di una politica è *definito* come differenza tra una situazione fattuale e una controfattuale. Ma ora la situazione controfattuale non è più approssimabile mediante un gruppo di controllo scelto con una procedura di randomizzazione. Ora il valutatore non può manipolare nulla, si limita ad osservare quello che succede: il controfattuale andrà approssimato *osservando* cosa succede ad altri soggetti e/o in altri periodi di tempo. I dati generati nel

³¹ Parte della letteratura parla di metodi "quasi-sperimentali": noi preferiamo il termine non-sperimentali, che è prevalente nella letteratura di tipo statistico-economico e sottolinea con più forza la differenza con i metodi sperimentali. I valutatori di formazione sociologica o psicologica tendono a usare il termine quasi-sperimentale, in omaggio alla terminologia introdotta quarant'anni fa da Donald Campbell, autore del saggio fondamentale con Julian Stanley, *Experimental and Quasi-experimental Designs for Research*, Rand McNally, Chicago, 1966.

corso dell'attuazione di una politica sono detti “dati osservazionali” (dall'inglese *observational data*).

Un caso stilizzato di politica pubblica

L'illustrazione dei metodi non-sperimentali sarà esemplificata nei prossimi capitoli mediante un caso stilizzato: consideriamo una ipotetica politica che distribuisca finanziamenti per rivitalizzare quartieri urbani degradati, realizzando un complesso di interventi mirati a far diventare i quartieri più vivibili e sicuri. Chiameremo questa politica pubblica Urban Renewal Policy, che abbrevieremo con URP. Supponiamo che i finanziamenti URP non siano disponibili per tutte le città e per tutti i quartieri degradati, bensì vengano concessi ad un numero massimo di quartieri scelti in quelle città che abbiano fatto domanda e che abbiano certi requisiti.

Prima di procedere ad esaminare i metodi utilizzabili per stimare gli effetti di una politica di questo tipo con dati non-sperimentali, dobbiamo rispondere alle due domande preliminari: effetto su cosa ed effetto di cosa?

Effetto su cosa? La risposta a questa prima domanda dipende da quali sono gli obiettivi che la politica in questione si pone e da come vengono interpretati ai fini della valutazione. Poniamo che l'URP abbia, tra gli altri, l'obiettivo di rendere più vivibili i quartieri riducendo gli atti di vandalismo contro edifici pubblici e privati. Supponiamo anche, per facilità di trattazione, che questa sia una variabile facilmente e univocamente misurabile. La variabile-risultato per questo esercizio sarà il “tasso di vandalismo”, misurato come “numero di atti di vandalismo contro edifici pubblici e privati commessi e denunciati in un anno, per 10.000 abitanti”, cioè il numero di atti di vandalismo diviso in numero dei residenti del quartiere e moltiplicato per 10.000. Si tratta quindi, a tutti gli effetti, di un “indicatore”. Probabilmente qualcuno lo indicherebbe come “indicatore di impatto”: per noi invece è solo una delle possibili variabili-risultato rispetto alla quale vorremmo stimare l'effetto di URP.

Omettiamo qui le discussioni che si potrebbero fare su come definire “vandalismo”, su come si possano rilevare e misurare gli “atti di vandalismo”, su quali siano le cause del vandalismo nei quartieri, su quali siano i meccanismi che possono portare una politica di recupero urbano a ridurre gli atti di vandalismo. Sono tutte questioni molto

interessanti che dovrebbero essere affrontate se ci accingessimo a lavorare sul serio sulla valutazione di tale politica. Qui però, supponendo che tali questioni siano state già risolte in sede preliminare, ci interessa risolvere solamente il seguente problema conoscitivo: come si fa a stabilire in modo credibile se la politica URP abbia effettivamente prodotto una riduzione degli atti di vandalismo nei quartieri in cui è stata attuata?

I ragionamenti che faremo nelle prossime pagine sarebbero da replicare tali e quali se invece del tasso di vandalismo utilizzassimo una variabile-risultato diversa, ad esempio il “numero di furti negli alloggi”, oppure la “percentuale di esercizi commerciali che chiudono”, o la “percentuale di tossicodipendenti residenti nel quartiere che sono in cura presso il SERT”.

Effetto di cosa? La politica pubblica che ci immaginiamo dietro l’etichetta URP si articola in un complesso di interventi e misure, dal miglioramento dei trasporti pubblici, agli interventi di restauro e recupero di edifici degradati, all’erogazione di corsi di formazione professionale: ciascuna di tali misure potrebbe essere in linea di principio essere valutata misurandone gli effetti sulle appropriate variabili-risultato. Un approccio così analitico spesso non è fattibile, innanzitutto per la mole di informazioni che richiederebbe da parte degli attuatori della politica. Nel nostro caso decidiamo di usare un approccio “globale” e di considerare come variabile-trattamento l’insieme di URP attuata (o meno) in un determinato quartiere. Un quartiere, in un certo anno, sarà o completamente esposto alla politica o completamente non esposto. Il programma URP o è attivo per il quartiere *i*-esimo oppure non lo è. Nel nostro esempio non sono possibili situazioni intermedie: in questo modo la variabile-trattamento si presenta come dicotomica e varia sia negli anni, sia tra i quartieri e, quindi, sia nel tempo, sia nello spazio.

Supponiamo per concretezza che URP sia stata realizzata tra il 2001 e il 2004 in 50 quartieri di altrettante città, vincitrici di un bando di gara. Questi sono i quartieri esposti alla politica pubblica, o quartieri trattati, o più semplicemente “quartieri URP”. Supponiamo inoltre di disporre di dati su altri 250 quartieri degradati di altrettante città. Questi saranno i quartieri non esposti alla politica, o non-trattati, o semplicemente “quartieri NonURP”. Successivamente introdurremo una distinzione all’interno dei 250

quartieri NonURP, tra i 100 che hanno partecipato al bando di gara ma non sono risultati vincitori e i 150 che non hanno neppure partecipato alla gara. I primi avranno in comune con i quartieri URP la decisione di partecipare alla gara e quindi per essi sarà disponibile il punteggio ottenuto, che consentirà l'applicazione di uno specifico metodo di valutazione. Fino ad allora, l'unica distinzione rilevante è quella tra i 50 quartieri esposti alla politica di recupero urbano e i 250 non esposti. Per entrambi i gruppi possediamo dati su alcune caratteristiche socio-economiche del quartiere, misurate nel 2000, prima dell'attuazione della politica, e sul tasso di vandalismo, misurato tutti gli anni dal 1996 al 2004.

La domanda di valutazione è quindi la seguente: "L'attuazione di URP tra il 2001 e il 2004 ha ridotto il tasso di vandalismo nei quartieri in cui è avvenuta?". Nella domanda non viene esplicitato alcun termine di paragone. Se ce ne fosse uno, esso a rigore dovrebbe essere "...rispetto a quello che si sarebbe osservato in assenza di URP", ovvero rispetto a quella situazione che abbiamo definito come situazione controfattuale.

Poniamo per un attimo di avere a disposizione solamente i dati sul tasso di vandalismo nei 50 quartieri URP nel corso del 2004. Questa informazione è palesemente insufficiente per una valutazione degli effetti della politica, in quanto costituisce solamente la base per rappresentare la situazione "fattuale". Utilizzando questi dati, apprendiamo che il tasso di vandalismo nel 2004 nei 50 quartieri è in media 66,4 atti ogni 10.000 abitanti (quindi circa un atto di vandalismo ogni 200 abitanti), e che varia da un minimo di 39 ad un massimo di 98. Con questi dati quale effetto di URP possiamo valutare?

Nessuno, ovviamente. Per valutare l'effetto che ci interessa dobbiamo fare dei confronti con "altri dati" che ci consentano di approssimare plausibilmente la situazione controfattuale, cioè il tasso di vandalismo che si sarebbe osservato nei 50 quartieri se URP non fosse stato attuato. I confronti che faremo saranno perciò sempre tra quel 66,4 e qualche altro termine, che avrà via via il compito di approssimare la situazione controfattuale. Quindi quel 66,4 non è un dato inutile, ma richiede un termine di paragone per poter generare una stima dell'effetto.

4.1 Il confronto prima-dopo e la “distorsione da dinamica spontanea”

Il tipo di confronto che più comunemente si tende a fare per misurare l’effetto di una politica è quello tra il valore della variabile-risultato osservato *dopo* l’attuazione della politica e quello osservato *prima*. Muovendosi in quest’ottica, nel nostro caso si potrebbe tentare una valutazione degli effetti di URP sul vandalismo calcolando la differenza tra la media dei tassi osservati nel 2004 nei 50 quartieri in cui URP è stata attuata e la media dei tassi osservati nel 2000 *negli stessi quartieri*. Seguendo questa logica, una *diminuzione* del tasso di vandalismo implicherebbe che la politica è stata *efficace* e l’effetto di URP coinciderebbe semplicemente con la diminuzione nel tasso di vandalismo.

La potenziale fallacia di questo approccio è nota fin da tempi antichi, fin da quando è stata messa in discussione la validità della massima latina “*post hoc ergo propter hoc*” (“dopo di ciò *quindi* a causa di ciò”), che erroneamente fa coincidere sequenzialità e causalità. Eppure quello di definire come effetto di una politica la semplice differenza “tra prima e dopo” è un errore ancora molto comune.

Il modo più immediato per evidenziare la fallacia di tale definizione di effetto di una politica pubblica è quello di considerare la situazione in cui la differenza tra prima e dopo vada nella direzione “sbagliata”, cioè nella direzione contraria a quella incorporata negli obiettivi della politica. Questa è precisamente la situazione che abbiamo per la politica URP attuata nei 50 quartieri: i dati nella tabella seguente ci mostrano un aumento medio, tra il 2000 e il 2004, di 3,47 atti vandalici ogni 10.000 abitanti. Se accettassimo la definizione di effetto come differenza prima-dopo, dovremmo concludere che URP *fa aumentare* il tasso di vandalismo.

Tabella 4.1: Confronto prima-dopo nei quartieri URP

Anno	Numero di quartieri	Media del tasso di vandalismo
2004	50	66,37
2000	50	62,90
Differenza		+ 3,47

Questo risultato paradossale indurrebbe chiunque a mettere in discussione l'equiparazione di "effetto" con "differenza prima-dopo". Ma tale messa in discussione deve essere più generale e non solo indotta dal verificarsi di situazioni paradossali: in generale, tale differenza *non* coincide con l'effetto, a meno che siamo disposti ad utilizzare come controfattuale la situazione osservata prima dell'attuazione della politica: quindi ad ipotizzare che, in assenza di politica pubblica, *nulla sarebbe cambiato*.

La principale minaccia alla validità di tale modo di stimare l'effetto di una politica è infatti la possibilità che il fenomeno su cui la politica vuole incidere (e quindi la variabile-risultato) abbia una sua *dinamica spontanea*. Con questo termine intendiamo la possibilità che la variabile-risultato sia soggetta ad altre influenze che ne determinano, tra il periodo "prima" e il periodo "dopo", una dinamica, che definiamo spontanea nel senso di indipendente dall'implementazione della politica. Nella letteratura valutativa anglosassone si parla di "*threat of history*" o "*maturation*", ma a noi pare che il termine "distorsione dovuta alla dinamica spontanea" catturi meglio l'idea che vogliamo comunicare.

Nel nostro caso specifico, è plausibile che il livello di vandalismo osservato nel 2000 sia rimasto costante e quindi rappresenti il livello che si sarebbe osservato nel 2004 *se URP non ci fosse stato*? All'opposto, nel nostro caso è abbastanza plausibile che ci sia una dinamica spontanea: è ragionevole immaginare che tante cose succedano nei 50 quartieri URP (e nelle città che li contengono) tra il 2000 e il 2004, cose che (a) possono avere un'influenza sul vandalismo; e (b) avrebbero avuto questa influenza anche se URP non ci fosse stato. (Entrambi i requisiti sono necessari perché la dinamica spontanea sia una minaccia alla validità delle stime ottenute con la differenza prima-dopo).

Escludere che ci sia una dinamica spontanea nel livello di vandalismo tra il 2000 e il 2004 ci porterebbe logicamente alla conclusione paradossale di cui sopra: il vandalismo è aumentato tra il 2000 e il 2004 e quindi, *se questa differenza va interpretata come effetto*, questo aumento nel vandalismo sarebbe *colpa di URP*. Nel nostro caso la scarsa plausibilità della stima ottenuta ci viene suggerita da una conclusione controintuitiva: è

difficile credere che un intervento di recupero urbano addirittura *faccia peggiorare* la situazione in termini di vandalismo.

Ma non sempre la minaccia alla validità delle nostre stime ci viene rivelata così chiaramente da conclusioni paradossali e controintuitive. Facendo confronti prima-dopo, dobbiamo *sempre* stare in guardia contro la minaccia rappresentata dalla dinamica spontanea del fenomeno. Valgono *sempre* le cautele seguenti, a seconda che il cambiamento osservato vada nella direzione desiderata o nella direzione opposta:

- un cambiamento prima-dopo *nella direzione desiderata* non necessariamente è *merito* della politica pubblica: la dinamica spontanea del fenomeno avrebbe potuto fare migliorare il fenomeno anche senza la politica;
- un cambiamento prima-dopo nella direzione *opposta* a quella desiderata non necessariamente è *colpa* della politica pubblica: il peggioramento potrebbe esserci stato comunque (anzi, potrebbe essere stato ancora più grande, quindi un peggioramento osservato può *nascondere* un effetto nella direzione desiderata).

Torniamo ad URP e al suo “effetto” sul tasso di vandalismo. Alla luce di quanto appena argomentato, i dati sul vandalismo nei 50 quartieri osservati prima e dopo l’attuazione della politica non sono sufficienti per stimare in modo *credibile* l’effetto della politica sul tasso di vandalismo. Abbiamo quindi bisogno di un disegno diverso di valutazione, e questo disegno inevitabilmente richiederà *più dati* di quelli che abbiamo finora utilizzato.

Una direzione possibile è quella di utilizzare le informazioni relative ai 250 quartieri che non sono stati esposti al programma URP (ovvero non hanno ricevuto i finanziamenti necessari a realizzare gli interventi): esploriamo se questi “nuovi dati” riescono a fornirci una stima più credibile degli effetti del programma sul tasso di vandalismo, realizzando un confronto spaziale anziché uno temporale.

4.2 Il confronto con-senza e la “distorsione da differenze di partenza”

L’idea è di approssimare ciò che sarebbe successo al vandalismo nei quartieri URP in assenza di intervento, utilizzando questa volta il livello di vandalismo osservato nei 250 quartieri che *non* hanno ricevuto il finanziamento URP (che chiameremo quartieri NonURP). Quindi per approssimare la situazione controfattuale in questo caso si usa la media del tasso di vandalismo osservato nei quartieri NonURP nel 2004. La tabella seguente contiene le informazioni necessarie ad implementare questa stima dell’effetto di URP.

Tabella 4.2: Confronto tra quartieri URP e quartieri NonURP nel 2004

Quartieri	Numero di quartieri	Media del tasso di vandalismo
URP 2004	50	66,37
NonURP 2004	250	57,50
Differenza		+ 8,87

Nel 2004 il vandalismo è maggiore nei quartieri URP di quasi 9 atti vandalici per 10.000 abitanti, che in termini percentuali rappresenta circa il 15% in più rispetto ai quartieri NonURP. Quale *interpretazione* è possibile assegnare ad una tale differenza? Notiamo che la differenza è nuovamente *positiva*: quindi, come già accadeva per il confronto prima-dopo, dovremmo concludere che il programma URP *fa aumentare* il tasso di vandalismo.

La domanda ora diventa: questa differenza “con-senza” è una stima plausibile dell’effetto di URP? Di nuovo, per *interpretare* questa differenza come *effetto* di URP sul vandalismo, dobbiamo ritenere che il livello di vandalismo osservato nei quartieri NonURP nel 2004 sia una buona approssimazione della situazione controfattuale. Ma ciò è sensato *se e solo se* le condizioni di partenza dei due gruppi di quartieri nel 2000 (sia il livello di vandalismo, sia le altre caratteristiche socio-economiche che lo determinano) fossero le stesse. Ma questo è molto poco plausibile: anzi è probabile che i 50 quartieri ammessi al finanziamento fossero già nel 2000 sistematicamente diversi dai

250 non ammessi, sia in termini di livello di vandalismo, sia di caratteristiche associate con il vandalismo.

È possibile che entrambe le decisioni, sia quella di partecipare alla gara (da parte delle singole città) sia quella di ammissione ai finanziamenti (da parte della commissione), siano state influenzate dalle condizioni comparativamente peggiori in cui si trovano questi quartieri dal punto di vista sociale. Se, quindi, i quartieri URP avessero mostrato già nel 2000 un tasso di vandalismo più alto di quello presente nei quartieri NonURP, non ci sentiremmo di “dare la colpa” ad URP se nel 2004 continuassimo a trovare un più alto livello di vandalismo nei quartieri finanziati.

Questo esempio stilizzato serve ad introdurre un principio più generale: le *differenze post-trattamento* tra trattati e non-trattati *non* rivelano l’effetto della politica, a meno che si possa dimostrare (o argomentare plausibilmente) che le condizioni di partenza siano uguali nei due gruppi (come nel caso dell’esperimento con randomizzazione). La minaccia alla validità dell’effetto viene dalle *differenze (nelle condizioni) di partenza*: invece, nel caso di confronto prima-dopo, era la *dinamica spontanea del fenomeno* a rappresentare la principale minaccia alla validità delle stime.

Quindi, nella loro forma più semplice, entrambe le strategie (confronto prima-dopo con le sole unità trattate; confronto con-senza nel solo periodo post-intervento) sono in linea di principio molto *deboli* come modi per valutare gli effetti di una politica pubblica. La ragione per cui è utile esporle è perché ciascuna di esse è la “capostipite” di una serie di sviluppi e approfondimenti.

Il *confronto prima-dopo per le unità trattate* è il capofila di quello che nella letteratura valutativa viene talvolta definito come “*pre-post one group design*”, o in psicologia come “*pretest-posttest design*”. La caratteristica principale di questa famiglia di approcci è di non usare alcun gruppo di unità non-trattate, ma di riferirsi solo alle unità trattate. Sarà la scelta obbligata per valutare gli effetti di politiche universali (ad esempio l’imposizione di obblighi o divieti generalizzati, oppure la conduzione di campagne di informazione di massa) oppure quando informazioni sulle unità non-trattate semplicemente non sono disponibili.

Il *confronto post-intervento tra unità trattate e non-trattate* è il capofila di quello che nella letteratura valutativa viene spesso definito come *comparison group design*. Le forme che questo disegno di base assume sono molteplici e si differenziano a seconda della strategia utilizzata per eliminare la distorsione causata dalle differenze di partenza. Questa letteratura definisce tale distorsione come *selection bias* o “distorsione da selezione”: questo termine sottolinea come le differenze di partenza siano indotte dal processo di selezione che separa trattati e non-trattati.

4.3 Il metodo differenza-nelle-differenze e la “distorsione da trend divergenti”

Una strategia di valutazione molto utilizzata combina i due approcci “semplici” appena descritti, mettendo cioè assieme il confronto tra unità trattate e non-trattate con il confronto prima-dopo. Questo è l’approccio che la letteratura sulla valutazione chiama *comparative-change design* e gli statistici chiamano più semplicemente *difference-in-differences*, o differenza-nelle-differenze³². Questo metodo ha un’importante valenza dimostrativa, per cui vale la pena vederlo applicato numericamente al caso di URP: l’illustrazione di questo metodo fa infatti toccare con mano come combinando più dati si ottengano risultati più plausibili, ma allo stesso tempo come non ci si liberi mai dalla minaccia di qualche distorsione residua.

Finora i nostri tentativi di valutazione degli effetti di URP si sono serviti di tre insiemi di dati, riassunti ciascuno dalla rispettiva media:

1. La media dei tassi di vandalismo nei quartieri URP nel 2004 (media di 50 osservazioni).
2. La media dei tassi di vandalismo nei quartieri URP nel 2000 (media di 50 osservazioni).

³² Per una trattazione grafica simile a quella utilizzata qui si veda Moffit R., “Program evaluation with non-experimental data”, *Evaluation Review*, 1991, 15: 291-314. Due interessanti applicazioni del metodo *difference-in-differences* sono Card D., Krueger A., *Myth and Measurement: The New Economics of the Minimum Wage*, Princeton University Press, Princeton, 1995 e Card D. “The Impact of the Mariel Boatlift on the Miami Labor Market”, *Industrial and Labor Relations Review*, 1990, 43(2), 245-257.

3. La media dei tassi di vandalismo nei quartieri NonURP nel 2004 (media di 250 osservazioni).

Aggiungiamo ora una quarta (e cruciale) osservazione:

4. La media dei tassi di vandalismo nei quartieri NonURP nel 2000 (media di 250 osservazioni).

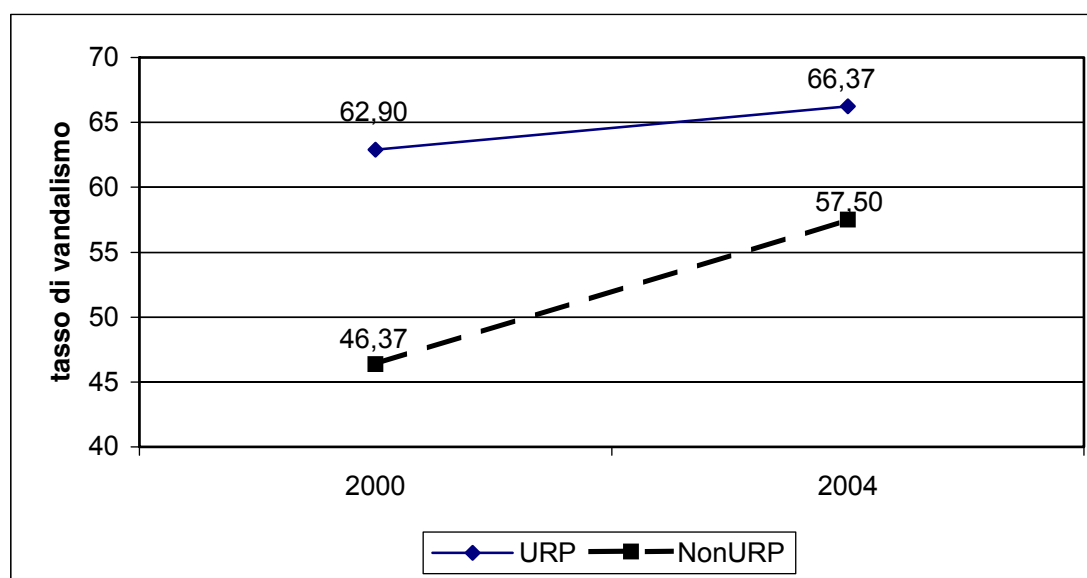
L'ultima sembrerà a prima vista un'aggiunta di poco conto, ma non è così. Questa quarta osservazione ci consente di fare dei confronti più "intelligenti": soprattutto ci consente di fare un'approssimazione del controfattuale più plausibile rispetto alle due fatte in precedenza. La tabella seguente mostra i quattro valori appena menzionati:

Tabella 4.3: Il tasso di vandalismo nei due anni nei due gruppi di quartieri

Tipo di quartiere	Numero di quartieri	Media del tasso di vandalismo osservato nel	
		2000	2004
URP	50	62,90	66,37
NonURP	250	46,37	57,50

La Figura 4.1 mostra i quattro valori sotto forma di grafico: i due segmenti illustrano visivamente i cambiamenti nel tasso di vandalismo tra il 2000 e il 2004, nei quartieri URP e nei quartieri NonURP, rispettivamente.

Figura 4.1: Il tasso di vandalismo in due anni nei due gruppi di quartieri



In che modo il nuovo dato a disposizione, relativo ai quartieri NonURP nel 2000, cambia le cose? Soprattutto, come arriviamo alla costruzione della situazione controfattuale che, come sappiamo, è un passo necessario per giungere alla stima dell'effetto di una politica?

Cercheremo di dare una risposta a queste domande proponendo due possibili percorsi logici. Per giungere alla stima dell'effetto della politica è infatti possibile procedere con il ragionamento in due modi diversi ma, come si vedrà, la conclusione alla quale si perviene è la stessa.

Primo ragionamento: *l'effetto è ancora una differenza, ma in questo caso una differenza tra due differenze*. A questo scopo riproduciamo la tabella precedente aggiungendo le differenze in verticale e in orizzontale:

Tabella 4.4: Differenze nel tempo e differenze tra i due gruppi di quartieri

Tipo di quartiere	Media del tasso di vandalismo nel		Differenze (2004 - 2000)
	2000	2004	
URP	62,90	66,37	+ 3,47
NonURP	46,37	57,50	+ 11,13
Differenze (U - NU)	+ 16,53	+ 8,87	- 7,66

Consideriamo prima le colonne della tabella e le *differenze in verticale*. Abbiamo così le differenze tra quartieri trattati e non-trattati calcolate rispettivamente nel 2000 (+ 16,53 = la differenza iniziale) e nel 2004 (+ 8,87 = la differenza finale, quella che nel paragrafo 4.2 avevamo, poco verosimilmente, provato ad interpretare come effetto di URP). Nel 2004, la differenza nel vandalismo tra i due gruppi di quartieri è ancora positiva, *ma si è ridotta* rispetto alla differenza inizialmente osservata nel 2000: questa *riduzione delle differenze iniziali* è un plausibile indizio del fatto che la politica funziona nella direzione giusta, quella di ridurre il tasso di vandalismo.

Alla luce di questo ragionamento, possiamo interpretare come effetto della politica la *differenza nelle differenze*, e più precisamente la differenza finale meno la differenza iniziale ($8,87 - 16,53 = - 7,66$). La stima risultante è *negativa*, il che implica che URP sembra ora *ridurre* il tasso di vandalismo. Invece che interpretare tutta la differenza

osservata nel 2004 come effetto (come abbiamo fatto nel paragrafo 4.2), l'abbiamo prima "depurata" della differenza iniziale. Il che non ci porta ad un risultato sicuramente "vero", ma certamente ad uno *più plausibile* di quello ottenuto sulla base delle sole differenze osservate nel 2004. Quantomeno, questo *escamotage* rappresenta un modo per eliminare quella distorsione che abbiamo prima attribuito alle differenze di partenza³³.

Prendiamo ora la stessa tabella e leggiamola *in orizzontale*, calcolando le differenze tra il 2004 e il 2000, per ciascuno dei due gruppi di quartieri. Come già sappiamo, il vandalismo è *aumentato* nei quartieri URP di 3,47 atti vandalici (aumento che nel paragrafo 4.1 avevamo provato, poco verosimilmente, ad interpretare come effetto di URP) e nei quartieri NonURP di un più sostanziale 11,13. Quest'ultimo dato suggerisce che, anche in assenza di politica di recupero urbano, il vandalismo aumenta comunque. Se questo aumento viene interpretato come "dinamica spontanea", andrà sottratto all'aumento osservato nei quartieri URP. Ma sottraendo 11,13 da 3,47 otteniamo nuovamente -7,66, cioè la stessa stima dell'effetto ottenuta nel paragrafo precedente.

Questo risultato può essere visto intuitivamente nella Figura 1. I due segmenti hanno pendenze diverse, e poiché le pendenze rappresentano la dinamica temporale del fenomeno, la pendenza più accentuata dei quartieri NonURP implica la presenza di una forte dinamica spontanea positiva: questo conferma quanto poco plausibile sarebbe *attribuire* il significato di effetto all'aumento osservato nei soli quartieri URP. Quello che dobbiamo fare è invece guardare al "cambiamento comparato", che è l'idea del *comparative-change design*.

Quale che sia l'ordine con cui calcoliamo la doppia differenza, l'effetto calcolato mediante la differenza-nelle-differenze non cambia.

Secondo ragionamento: *l'effetto è la differenza tra osservato e controfattuale*. La discussione nei paragrafi precedenti ha in un certo senso tradito la nostra definizione di

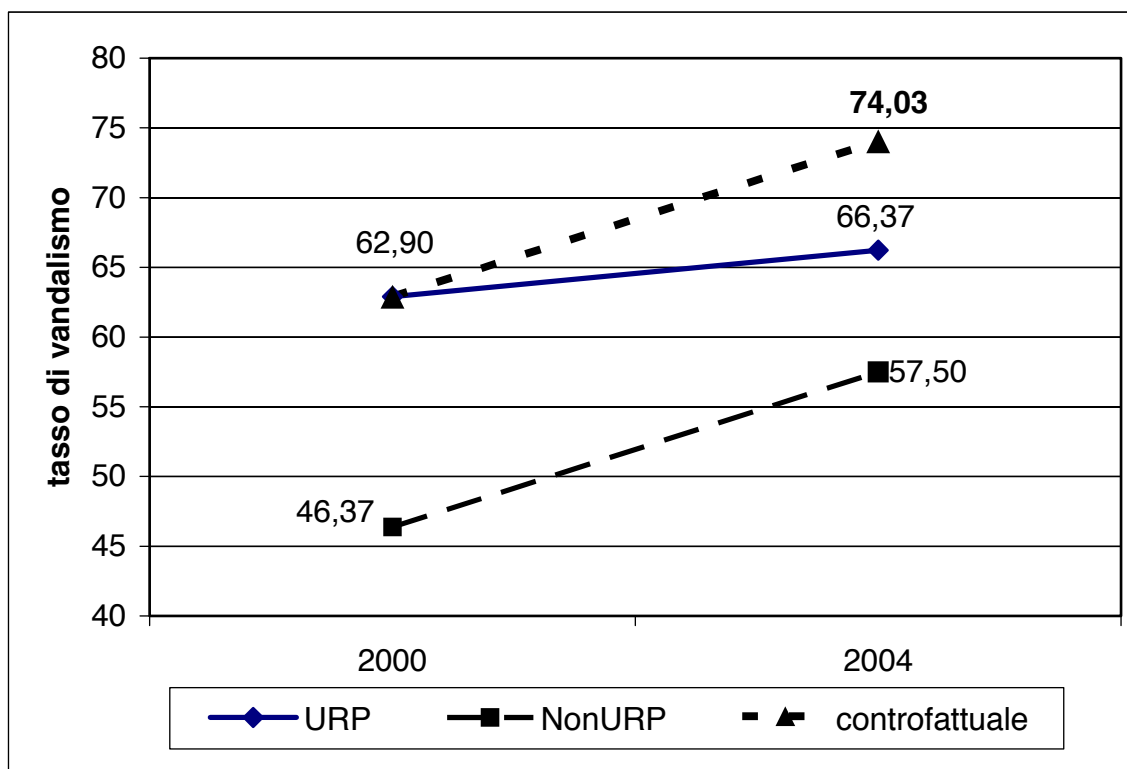
³³ Le differenze sono calcolate in termini assoluti, anziché in termini percentuali, per facilitare l'esposizione, non perché questo modo di procedere sia più corretto. Si potrebbero calcolare le differenze in percentuale e poi calcolare le differenze tra percentuali. Numericamente i risultati sarebbero diversi. Ad esempio, applicando al livello osservato nei quartieri URP nel 2000 l'aumento *percentuale* osservato nei quartieri NonURP tra il 2000 e il 2004, si ottiene un effetto stimato di -11,63 atti vandalici. Risultato diverso numericamente, ma analogo in termini logici.

effetto, perché non ha fatto nulla per derivare una stima del controfattuale. Le stime poco plausibili di effetto derivate nei paragrafi 4.1 e 4.2 utilizzavano come controfattuale rispettivamente il valore “URP 2000” (nel caso di differenza prima-dopo) o il valore “NonURP 2004” (nel caso di differenza con-senza), valori che venivano sottratti a turno da “URP 2004”, il valore fattuale. Ma quale controfattuale abbiamo implicitamente usato per derivare la stima di 7,66 atti di vandalismo *in meno*?

Abbiamo implicitamente utilizzato una combinazione delle altre tre medie, secondo il ragionamento seguente. Dove ci saremmo aspettati di trovare il tasso di vandalismo nei quartieri URP nel 2004 se URP non ci fosse stato, ma in questi quartieri la dinamica del tasso di vandalismo tra il 2000 e il 2004 fosse stata la stessa osservata nei quartieri NonURP? Se questo ultimo assunto fosse valido (torneremo su questo punto) ci saremmo aspettati di trovare 74,03 atti vandalici in media ($= 57,50 + 16,53$). Applicando a questo punto la definizione effetto = valore osservato – valore controfattuale, otteniamo $66,37 - 74,03 = -7,66$, lo stesso risultato di prima. Il controfattuale è più alto del livello osservato, quindi la stima dell’effetto è negativa (Figura 4.2)³⁴.

³⁴ Se si fosse fatto il calcolo delle differenze in percentuale, invece che in termini assoluti, il controfattuale stimato sarebbe pari a 78,0 atti vandalici, che, sottratto a 66,37, produce una stima dell’effetto pari a -11,63.

Figura 4.2: La derivazione del controfattuale con il metodo differenza-nelle-differenze



Il grafico indica la “dinamica controfattuale” con una linea a puntini. Questa linea è tracciata *parallelamente* a quella tratteggiata della “dinamica spontanea” osservata nei quartieri NonURP. Questo parallelismo è un’ipotesi innocua? O almeno, è verosimile?

Il *parallelismo* tra la linea della dinamica controfattuale e quella della dinamica spontanea nasconde in realtà un assunto che può risultare in certi casi poco plausibile. E se questo assunto non è accettabile, perde di robustezza anche la stima dell’effetto che abbiamo appena (e così faticosamente) ottenuto. Qual è, in termini di sostanza e non puramente geometrici, questo assunto? L’assunto è che le differenze di partenza tra i due gruppi di quartieri siano contenute interamente nelle differenze osservate tra *i livelli* iniziali del tasso di vandalismo, e *non ci siano differenze nei trend* del vandalismo nei due gruppi di quartieri.

I dati a nostra disposizione ci dicono che il vandalismo era più alto già nel 2000 nei quartieri URP (e può essere questa una delle ragioni per cui il finanziamento è stato concesso a questi 50 quartieri). Ma nulla esclude che i quartieri poi diventati URP avessero *anche* un trend di crescita più accentuato negli anni precedenti al 2000. Se questo fosse vero, il controfattuale dovrebbe essere posto *più in alto* di quanto

l'abbiamo posto nella Figura 4.2 e quindi l'effetto di URP sul vandalismo sarebbe ancora *maggiore* di quello appena stimato. Oppure, a livelli più alti potrebbero corrispondere trend meno crescenti: in questo secondo caso la nostra ipotesi di trend simili contribuirebbe a produrre una sovrastima dell'effetto di URP sul vandalismo. Tale effetto sarebbe ancora *minore* di quello che abbiamo stimato. Quale che sia la direzione della distorsione, essa dipende dalle differenze nei trend. Per questo la stima ottenuta con il metodo differenza-nelle-differenze può essere affetta da “distorsione da trend divergenti”.

Come si fa a sciogliere questo dilemma (differenze nei trend pre-2000 o unicamente differenze nei livelli del 2000?) da cui dipende la plausibilità della stima $-7,66$ come effetto di URP sul vandalismo? La risposta è sempre la stessa: *occorre avere a disposizione più dati*. In questo caso, al minimo, ci servono le medie dei tassi di vandalismo nei due gruppi di quartieri nel 1996 (per simmetria con la distanza quadriennale tra 2004 e 2000). Tra il 1996 e il 2000 possiamo osservare il trend nei tassi di vandalismo nei due gruppi di quartieri essendo sicuri che nessuno dei due trend possa essere influenzato dalla politica in esame (che tra il 1996 e il 2000 non esisteva ancora). Aggiungiamo perciò una colonna alla Tabella 4.3 e due ulteriori punti al grafico corrispondente, ottenendo:

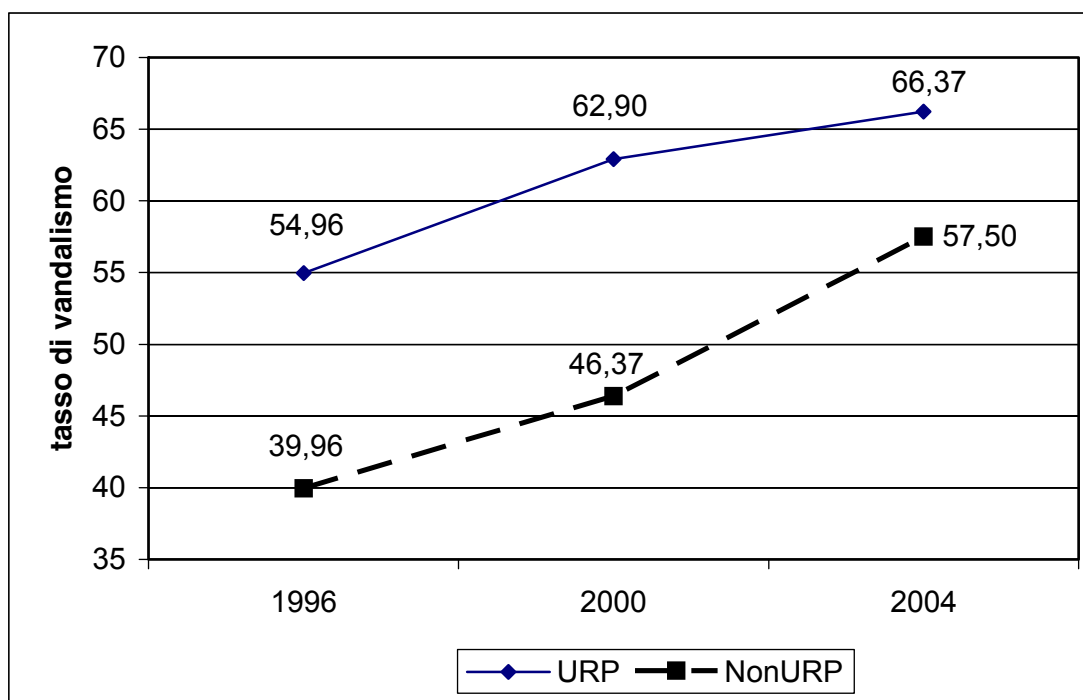
Tabella 4.5: Il tasso di vandalismo in tre anni diversi nei due gruppi di quartieri

Tipo di quartiere	Numero di quartieri	Media del tasso di vandalismo osservato nel		
		1996	2000	2004
URP	50	54,96	62,90	66,37
NonURP	250	39,96	46,37	57,50

In che modo da queste sei osservazioni è possibile ricavare la stima dell'effetto di URP? Usiamo prima l'intuizione, esaminando graficamente l'andamento del fenomeno nei due gruppi di quartieri. La differenza nei tassi di vandalismo si è *ampliata* tra il 1996 e il 2000: quindi non solo i quartieri URP avevano dei livelli più alti nel 2000, ma anche dei trend più accentuati. Quindi URP ha invertito questa tendenza alla divaricazione tra i due gruppi di quartieri.

Per ottenere una stima quantitativa, occorre sviluppare il ragionamento fatto in precedenza. Con quattro osservazioni l'effetto poteva essere stimato dalla seguente sottrazione: “differenza post *nei livelli*” – “differenza pre *nei livelli*”, che era una stima valida sotto l'assunto che non ci fossero differenze tra i trend a distorcere le stime ottenute.

Figura 4.3: Il tasso di vandalismo in tre anni diversi nei due gruppi di quartieri



Adesso invece possiamo, con sei osservazioni, anche calcolare le differenze-pre *nei trend*: quindi possiamo *sottrarle* dalla stima dell'effetto ottenuta con quattro osservazioni. Se le differenze-pre nei trend sono zero, la stima non cambierà. Viceversa, se la dinamica sarà più forte nei quartieri URP, il differenziale positivo andrà ad *aggiungersi* alla stima dell'effetto ottenuta con quattro osservazioni. Se viceversa la dinamica sarà più debole nei quartieri URP, questo ridurrà la stima dell'effetto già ottenuta. La formula è illustrata nella Tabella 4.6.

Tabella 4.6: Formula della differenza-nelle-differenze corretta per le differenze nei trend pre-intervento

Effetto =	$(U_{2004} - N_{2004})$	$-(U_{2000} - N_{2000})$	$-[(U_{2000} - U_{1996}) - (N_{2000} - N_{1996})]$
	differenza post nei livelli	- differenza pre nei livelli	- differenza pre nei trend
- 9,19	8,87	- (+ 16,53)	- [+ 1,53]

Lo scenario che si ottiene è il primo dei due descritti. La differenza post nei livelli indicava un effetto (poco plausibile) di + 8,87 atti di vandalismo. La correzione ottenuta sottraendo le differenze pre nei livelli portava la stima a - 7,66. Ora questa stima viene ulteriormente corretta, nella stessa direzione. Infatti, come già si intuiva dalla figura, il trend in aumento nel periodo 1996-2000 era *maggiore* nei quartieri che poi sarebbero diventati URP.

Questo trend addizionale (+ 1,53) va sottratto alla stima di - 7,66, producendo una stima “finale” di 9,19 atti vandalici *in meno* ogni 10.000 abitanti. Rispetto al punto di partenza dei quartieri URP nel 2000 (62,90 atti per 10.000 abitanti) rappresenta una riduzione del 14,6%. (Un risultato quantitativamente notevole, ma non dimentichiamo che questi numeri sono fittizi: quello che importa è il metodo rigoroso e trasparente con cui queste stime sono state ottenute).

Siamo quindi giunti al termine del nostro percorso e siamo riusciti, con i dati a nostra disposizione, ad ottenere una stima (abbastanza) plausibile dell’effetto di URP, utilizzando il metodo delle differenza nelle differenze. Facendo questo ci siamo accorti come la plausibilità delle stime non dipende tanto dal modello d’analisi adottato, quanto piuttosto dalla ricchezza dei dati a nostra disposizione. Dati più ricchi ci hanno infatti consentito di adottare ipotesi meno stringenti e più verosimili sull’evoluzione del fenomeno nei due gruppi di quartieri.

CAPITOLO 5

LA REGRESSIONE E IL *MATCHING* STATISTICO COME STRUMENTI PER RIDURRE LE DIFFERENZE DI PARTENZA

di Alberto Martini

Per stimare l'effetto di una politica, fino a questo punto abbiamo fatto esclusivamente ricorso al calcolo di *differenze tra medie* (nel tempo, nello spazio o in entrambe le dimensioni). Nel capitolo precedente abbiamo evidenziato come queste semplici differenze tra medie “rivelino” (e quindi consentano di stimare in modo plausibile) l'effetto di una politica solo se sono soddisfatte particolari condizioni: ad esempio, l'assenza di dinamica spontanea nel caso di confronti prima-dopo, l'assenza di differenze di partenza nel caso del confronto tra trattati e non-trattati, nonché l'assenza di differenze nei trend pre-intervento nel caso del doppio confronto (differenza-nelle-differenze).

Riprendiamo il caso del confronto tra trattati e non-trattati osservati solo dopo il trattamento. La differenza osservata tra le due medie può essere scritta come la somma di due componenti:

$$(1) \quad \text{differenza osservata tra trattati e non-trattati nella variabile-risultato} = \\ \text{effetto della politica} + \text{differenze di partenza}$$

Questa relazione evidenzia come la differenza osservata tra trattati e non-trattati “rivelino” l'effetto della politica *se e solo se* le differenze di partenza sono *nulle*³⁵. Ciò si verifica (quasi) sicuramente quando trattati e non-trattati sono scelti mediante assegnazione casuale, cioè quando si è effettuato un esperimento con randomizzazione. Se trattati e non-trattati si sono invece autoselezionati o sono stati assegnati al trattamento in un modo comunque non casuale, come avviene nella grande maggioranza dei casi, le differenze di partenza tipicamente *non* sono nulle; la loro presenza fa sì che la differenza osservata tra trattati e non-trattati sia una stima *distorta* dell'effetto della

³⁵ Allo stesso modo, quando si considera la differenza tra medie *nel tempo*, si può scrivere l'equivalente della (1) come:

$$\text{differenza pre-post} = \text{effetto della politica} + \text{dinamica spontanea}$$

In questo caso la differenza pre-post “rivela” l'effetto della politica solo se non c'è dinamica spontanea.

politica, come si vede chiaramente dall'espressione (1). Questa distorsione è anche detta *da selezione*, perché le differenze di partenza sono dovute al modo con cui i soggetti vengono selezionati o si autoselezionano per ricevere o non ricevere il trattamento (il termine inglese per questo tipo di distorsione è *selection bias*).

La situazione a cui fa riferimento l'espressione (1) si verifica di frequente nella valutazione delle politiche e sarà centrale nella trattazione di questo capitolo: è la situazione in cui è possibile solo fare confronti tra trattati e non-trattati e non è possibile invece effettuare anche confronti nel tempo. Più precisamente, questa situazione si presenta quando, da un lato la politica pubblica è di tipo non-universale, cioè consiste in prestazioni (servizi, sussidi, agevolazioni) a cui solo alcuni soggetti sono ammessi (o decidono di fruirne): è quindi possibile distinguere i soggetti trattati dai non-trattati; e dall'altro, non esistono dati sulla variabile-risultato nel periodo pre-intervento, o perché tali dati non sono stati raccolti, oppure perché non è possibile logicamente definire la variabile-risultato per più periodi di tempo³⁶.

La strategia di valutazione presentata nella prima parte di questo capitolo parte dalla seguente considerazione: se i due gruppi sono diversi già prima del trattamento, questa diversità si tradurrà in diversità nelle loro caratteristiche demografiche, economiche, sociali, *osservabili prima* del trattamento: in particolare, ci interessano le caratteristiche che a loro volta influenzano la variabile-risultato.

Tornando all'esempio del capitolo precedente, per come si sono (o sono stati) selezionati, i quartieri URP potrebbero essere sistematicamente diversi dai quartieri NonURP in termini di una serie di caratteristiche socio-economiche, che a loro volta possono favorire il fenomeno del vandalismo: ad esempio, i primi potrebbero avere una popolazione più svantaggiata economicamente e socialmente, in termini di reddito, accesso al lavoro, disgregazione familiare, livello di istruzione, scarsità di reti sociali, e così via.

³⁶ Si pensi, ad esempio, al problema di stimare l'effetto di ricevere un'indennità sulla durata della disoccupazione seguente a licenziamento collettivo. Alcuni lavoratori hanno diritto all'indennità di disoccupazione per un certo periodo di tempo dopo il licenziamento, altri no. Il confronto può essere fatto tra indennizzati e non indennizzati, ma non tra durata pre-indennità e durata post-indennità, perché la durata della disoccupazione dopo il licenziamento è una sola. In generale vale la regola che, per fare confronti intertemporali, la variabile-risultato deve essere misurabile ad intervalli regolari e ripetuti nel tempo (come nel caso del tasso di vandalismo).

Riscriviamo l'espressione (1) mettendo in evidenza il fatto che le differenze di partenza si traducono in diversità nelle caratteristiche che influenzano la variabile-risultato:

$$(2) \quad \text{differenza osservata tra trattati e non-trattati nella variabile-risultato} = \text{effetto della politica} + \text{differenze di partenza nelle caratteristiche che influenzano la variabile-risultato}$$

Se le caratteristiche che influenzano la variabile-risultato fossero davvero *tutte* osservabili, si aprirebbe una strada nuova per valutare gli effetti della politica: l'utilizzo di quei metodi statistici, quali l'*analisi di regressione*, che consentono di stimare l'effetto della variabile-trattamento sulla variabile-risultato *a parità di altre condizioni*, cioè a parità di caratteristiche *osservabili*. Tali metodi possono essere pensati come modi per *depurare* la differenza osservata tra trattati e non-trattati della componente imputabile alle differenze di partenza nelle caratteristiche osservabili.

L'applicazione di questi metodi, a condizione che le caratteristiche che influenzano la variabile-risultato siano *tutte* osservabili, consente di riscrivere l'espressione (2) come:

$$(3) \quad \text{differenza trattati non-trattati} \text{ depurata delle differenze nelle caratteristiche osservabili} = \text{effetto della politica}$$

Se il metodo statistico adottato è davvero in grado di depurare il confronto tra trattati e non-trattati dalle loro differenze nelle caratteristiche osservabili, la differenza *depurata* rappresenterà plausibilmente l'effetto della politica. Va tuttavia sottolineato come la plausibilità di questa stima dell'effetto non dipenda dal particolare metodo statistico adottato, quanto piuttosto dalla *ricchezza dei dati* che si hanno a disposizione per la stima. Nelle righe precedenti abbiamo supposto che *tutte* le caratteristiche che influenzano la variabile-risultato siano osservabili dal valutatore. Più realistica è la situazione in cui alcune delle caratteristiche siano osservabili e altre non lo siano. Un modo più realistico di scrivere l'espressione (2) è separare le differenze di partenza in due componenti, una dovuta alle caratteristiche osservabili e una dovuta alle caratteristiche non osservabili dei soggetti a cui la politica si rivolge:

$$(4) \quad \text{differenza osservata tra trattati e non-trattati nella variabile-risultato} = \text{effetto della politica} + \text{differenze di partenza nelle caratteristiche osservabili} + \text{differenze di partenza nelle caratteristiche **non** osservabili}$$

L'applicabilità dei metodi statistici descritti in questo capitolo dipende dall'assunto che le differenze di partenza dovute a caratteristiche *non* osservabili, una volta *tenuto conto* di quelle osservabili, *siano nulle o quasi nulle*. Se le differenze di partenza dovute a caratteristiche non osservabili sono davvero nulle o quasi, i metodi illustrati di seguito sono in grado di produrre stime corrette dell'effetto di una politica pubblica.

Nella prima parte di questo capitolo illustreremo come questo risultato possa essere ottenuto utilizzando lo strumento della *regressione multipla*. Prima di arrivare a questo, faremo un passaggio intermedio: mostreremo come le stime ottenute nel capitolo precedente mediante differenza tra medie possano essere *esattamente* replicate utilizzando un modello di regressione *semplice* (che utilizza cioè una sola variabile esplicativa).

Una volta dimostrata tale equivalenza tra regressione *semplice* e differenze tra medie, esploreremo i vantaggi che offre la regressione *multipla*. L'uso del modello di regressione multipla offre infatti un modo per *depurare* le stime degli effetti della politica dall'influenza delle *caratteristiche (o variabili) osservabili*.

5.1 L'equivalenza tra regressione semplice e differenza tra medie

Proviamo innanzitutto a replicare, utilizzando il modello di regressione semplice, la differenza nel tasso vandalismo tra quartieri URP e NonURP nel 2004. Ricordiamo che nel capitolo precedente abbiamo giudicato quest'ultima una stima poco plausibile dell'effetto di URP, in quanto ignora completamente l'esistenza di differenze di partenza tra i due gruppi di quartieri: infatti utilizza implicitamente come controfattuale la media del tasso di vandalismo osservata nei quartieri NonURP nel 2004.

Stimiamo ora una regressione semplice utilizzando come variabile dipendente il tasso di vandalismo osservato nel 2004 e come *unica* variabile esplicativa la variabile-trattamento, che assume valore uguale a 1 se il quartiere è URP, uguale a zero se NonURP. Quello che otteniamo come coefficiente stimato per la variabile-trattamento è un valore *identico* alla differenza tra medie calcolata nel capitolo precedente. Questo non è un caso, ma è una diretta conseguenza del fatto che una regressione in cui l'unica variabile esplicativa sia una variabile dicotomica (che assume solo valori uguali a

zero e 1) produce un coefficiente stimato il cui valore corrisponde esattamente alla differenza tra le medie della variabile dipendente calcolate per i due gruppi.

In termini più formalizzati, consideriamo il seguente modello di regressione lineare:

$$(5) \quad Y_i = \alpha + \beta T_i + \varepsilon_i$$

dove:

Y_i rappresenta il tasso di vandalismo osservato nel quartiere i-esimo nel 2004;

T_i rappresenta la variabile-trattamento, = 1 per i quartieri URP, = 0 per i quartieri NonURP;

ε_i è il termine di errore (con media zero), che rappresenta l'influenza di tutte le variabili non osservabili³⁷.

I coefficienti da stimare sono α e β : vedremo ora come, in questo modello, la stima di α coincida con la media del tasso di vandalismo nei quartieri NonURP nel 2004, mentre la stima di β coincida con la differenza del tasso di vandalismo tra i due gruppi di quartieri.

Questo concetto può essere formalizzato ricorrendo al simbolo $E(\bullet)$, che significa “valore atteso”, cioè “media di”, e a quello di $E(\bullet|T)$, che significa “media condizionata al particolare valore di T”. Quindi la media del tasso di vandalismo condizionata al fatto che il quartiere è NonURP (cioè $T_i = 0$), sarà:

$$(6) \quad E(Y_i | T_i = 0) = \alpha + \beta \cdot 0 + E(\varepsilon_i | T_i = 0)$$

Tuttavia, $E(\varepsilon_i | T_i = 0) = 0$ per via dell'assunto che le variabili non osservabili hanno media zero, indipendentemente dal fatto che il quartiere sia di tipo URP o NonURP (il che corrisponde all'assunto che non ci siano differenze tra i due gruppi di quartieri nelle variabili non osservabili). Quindi la (6) può essere riscritta come:

$$(6') \quad E(Y_i | T_i = 0) = \alpha$$

Poiché, per lo stesso assunto, $E(\varepsilon_i | T_i = 1) = 0$, dalla (5) avremo la seguente media del tasso di vandalismo, condizionata al fatto che il quartiere è URP:

³⁷ Va notato che in questo modello tutte le variabili diverse dal trattamento sono considerate non osservabili, poiché l'unica variabile esplicativa inclusa nel modello è la variabile-trattamento T.

$$(6'') \quad E(Y_i | T_i = 1) = \alpha + \beta$$

Per differenza tra le due espressioni precedenti, si ricava che:

$$(6''') \quad \beta = E(Y_i | T_i = 1) - E(Y_i | T_i = 0)$$

Quello appena illustrato è un modo formale per mostrare che β rappresenta l'effetto di T su Y e coincide con la *differenza* tra le medie del tasso di vandalismo nei due gruppi di quartieri. Va ricordato però come questa interpretazione di β come “effetto” di URP dipenda dall'assunto che le altre caratteristiche che influenzano la variabile-risultato non differiscano sistematicamente tra i due gruppi di quartieri, e cioè che le differenze di partenza siano *nulle*.

Questo assunto sarebbe molto più plausibile nel caso in cui i quartieri fossero assegnati al programma URP in modo casuale, cioè se la politica fosse realizzata sotto forma di un esperimento con randomizzazione. In quel caso $E(\varepsilon_i | T_i = 0) = E(\varepsilon_i | T_i = 1) = 0$ per costruzione, non per assunto. Se il trattamento fosse assegnato a caso, il gruppo di controllo e gruppo di trattamento avrebbero in media le stesse caratteristiche, siano esse osservabili o non osservabili.

Ma questo non è il nostro caso: i 50 quartieri trattati da URP si sono prima auto-selezionati (nel senso che non tutti i quartieri esistenti fanno domanda per i finanziamenti) e sono poi stati etero-selezionati (non tutti quelli che fanno domanda sono ammessi). Come abbiamo già visto, l'esistenza di un meccanismo non casuale di selezione rende probabile che le caratteristiche dei quartieri, rappresentate nel modello (5) dal termine ε_i , varino in modo sistematico tra i due gruppi di quartieri. Quindi l'interpretazione di β come effetto di URP dipende dalla plausibilità dell'assunto che, tra i due gruppi di quartieri, le differenze nelle caratteristiche *non* osservabili siano nulle.

Nella discussione fatta finora, β ha rappresentato sia il coefficiente della retta di regressione sia la differenza tra due medie. Questo perché, lo ripetiamo, il coefficiente di una variabile dicotomica inserita come unica variabile esplicativa in un modello di regressione equivale ad una differenza tra medie. Questa proprietà può essere verificata

nel nostro caso stimando la regressione (5) con un apposito programma di analisi statistica. I risultati di tale stima sono rappresentati nella Tabella 5.1.

Tabella 5.1: Stime della regressione (5)³⁸

Variabile	Simbolo	Coefficiente	Errore Standard	t-statistico	<i>p-value</i>
Costante	α	57,50	0,906	63,4	0,000
T = trattamento	β	8,870	2,221	3,99	0,000

Ritroviamo come stima del coefficiente β lo stesso valore che abbiamo ottenuto nel capitolo precedente calcolando la differenza tra medie, cioè + 8,87. Parimenti, la stima dell'intercetta è uguale alla media del tasso di vandalismo nei quartieri NonURP nel 2004, che è appunto 57,5 atti vandalici ogni 10.000 residenti. Nella Tabella 5.1 troviamo che il t-statistico del coefficiente angolare è 3,99, quindi rifiutiamo l'ipotesi $\beta = 0$ con una probabilità superiore al 99%. La regressione ci conferma che, statisticamente parlando, l'“effetto” di URP è positivo e significativamente diverso da zero. Ma questo risultato nulla aggiunge alla sostanza del problema: soprattutto non modifica la sua (assai scarsa) plausibilità.

Allora perché derivare lo stesso effetto con due metodi diversi, se è poco plausibile? Lo scopo di questo esercizio è ricordare che il fatto di utilizzare la regressione, con la sua pretesa di maggiore scientificità, non cambia la *plausibilità* della stima ottenuta. Se la stima di un effetto è poco plausibile perché non si è tenuto conto delle differenze di partenza, essa resta egualmente poco plausibile anche se è ottenuta mediante una regressione. Cambia la tecnica di stima, ma i dati sono gli stessi e quindi anche l'informazione che abbiamo a disposizione è la stessa.

³⁸ In questo capitolo non presentiamo il valore dell' R^2 delle regressioni perché non è rilevante per il particolare tipo di analisi che stiamo conducendo: non stiamo infatti cercando di spiegare la variabilità della variabile-risultato, quanto piuttosto stiamo cercando di capire quanto la variabile-trattamento (e solo quella) influisce sulla variabile-risultato. Il nostro scopo è fare inferenza causale sull'effetto di URP, non spiegare il fenomeno del vandalismo. Quindi l'unica stima di interesse è quella del coefficiente β . Si ricordi inoltre che i dati utilizzati per queste stime sono fittizi.

5.2 La regressione multipla³⁹

L'analisi di regressione presenta però un notevole vantaggio rispetto alla differenza tra medie: allungare la lista delle variabili esplicative, non fermandosi cioè alla sola variabile-trattamento, rappresenta un modo per eliminare la distorsione dovuta dalle differenze di partenza nelle caratteristiche *osservabili*.

Riprendiamo il caso di URP. È plausibile che i due gruppi di quartieri differiscano in termini di caratteristiche socio-economiche osservabili, che a loro volta possono influire sulla frequenza degli atti di vandalismo. Consideriamo due caratteristiche, o variabili, che supponiamo siano state misurate per ciascuno dei 300 quartieri nel 2000: il tasso di disoccupazione e il tasso di immigrazione.

Ipotizziamo che queste due variabili influenzino (positivamente) il tasso di vandalismo e che siano distribuite diversamente tra i quartieri URP e NonURP. Più precisamente, facciamo le seguenti due ipotesi. La prima è che alta disoccupazione e alta immigrazione contribuiscano a peggiorare il clima di convivenza sociale in un quartiere e con questo ad aumentare gli atti di vandalismo⁴⁰. La seconda è che, nel selezionare i quartieri a cui assegnare i finanziamenti di URP, sia stata data priorità a quelli con più alta disoccupazione e immigrazione (oltre ad altri criteri di preferenza). Queste due ipotesi creano la classica situazione di “correlazione spuria”: vandalismo e finanziamenti URP sono *positivamente* correlati non perché URP causi un aumento nel tasso di vandalismo, ma perché i quartieri *selezionati* per URP tendono ad avere *in partenza* più alta disoccupazione e immigrazione. Quindi una parte delle differenze nel tasso di vandalismo osservate nel 2004 tra i due gruppi di quartieri è spiegata dalle differenze di partenza nelle condizioni socio-economiche, rappresentate dalla disoccupazione e dall'immigrazione. Ignorare queste differenze di partenza rende la semplice differenza tra i tassi di vandalismo una stima *distorta* dell'effetto di URP. In questo caso distorta verso i valori positivi.

³⁹ La trattazione della regressione fatta in questo volume è estremamente semplificata: per trattazioni più approfondite del modello di regressione si veda Johnston J. (1993), *Econometrica* (3a edizione), FrancoAngeli, Milano, 1993 e in inglese Greene W., *Econometric Analysis* (5th edition), Prentice-Hall, New Jersey, 2003.

⁴⁰ Si tratta di un'ipotesi che ha una funzione puramente didattica e nessuna valenza ideologica.

La regressione multipla è un modo per rimediare a questa situazione: ad ottenere cioè stime dell'effetto di una variabile di interesse (in questo caso, la variabile-trattamento) *al netto* dell'influenza delle altre variabili (dette collettivamente variabili di controllo). Nel nostro caso, questa stima “netta” di β si ottiene aggiungendo all'equazione (5) le due variabili di controllo a disposizione, il tasso di disoccupazione e quello di immigrazione, ottenendo:

$$(7) \quad Y_i = \alpha + \beta T_i + \theta \text{IMMIGRAZ}_i + \lambda \text{DISOCCUP}_i + \varepsilon_i$$

Va notato come i valori di immigrazione e disoccupazione devono essere osservati nel 2000, cioè al momento della selezione dei quartieri e comunque prima dell'implementazione della politica, perché quest'ultima potrebbe influenzarne il valore: se fossero osservate nel 2004, le differenze di immigrazione e disoccupazione non rappresenterebbero più differenze di *partenza*. Quello che vogliamo fare è depurare la stima dell'effetto di URP dalle differenze di partenza, per cui le variabili di controllo vanno misurate prima dell'intervento.

Un modo più intuitivo per interpretare il modello (7) è quello di riscriverlo nel modo seguente, portando le variabili di controllo sul lato sinistro dell'equazione, moltiplicate per il corrispettivo coefficiente:

$$(8) \quad Y_i - \theta \text{IMMIGRAZ}_i - \lambda \text{DISOCCUP}_i = \alpha + \beta T_i + \varepsilon_i$$

Si può pensare ciò che sta sul lato sinistro della (8) come al tasso di vandalismo “depurato” dell'influenza che su di esso hanno l'immigrazione e la disoccupazione: il modello (8) è infatti l'applicazione al contesto della regressione dell'espressione (3), che qui riproduciamo:

differenza trattati non-trattati depurata delle differenze nelle caratteristiche osservabili = effetto della politica

Il modello (8) mette in evidenza come il coefficiente β sia ancora *interpretabile* come differenza tra trattati e non-trattati, come avveniva nel modello (5): ora, però, è una differenza nei tassi di vandalismo *al netto* delle differenze di partenza in termini di immigrazione e disoccupazione. Le stime del modello (7) sono le seguenti:

Tabella 5.2: Stime della regressione (7)

Variabile	Simbolo	Coefficiente	Errore Standard	t-statistico	p-value
Costante	α	- 23,0	2,460	-9,35	0,000
T = trattamento	β	- 0,164	1,075	-0,15	0,879
IMMIGRAZ	θ	2,71	0,130	20,85	0,000
DISOCCUP	λ	3,68	0,131	28,09	0,000

La stima di β (cioè l'effetto stimato di URP sul vandalismo) è diventata ora negativa, ma non è statisticamente significativa (infatti il suo *p-value* è 0,88, cioè la probabilità che l'ipotesi $\beta=0$ sia vera è quasi il 90 per cento). Gli effetti della immigrazione e della disoccupazione sul vandalismo sono invece positivi (come ci si attendeva) e statisticamente diversi da zero. La nostra aspettativa ne esce quindi confermata: eliminare con un metodo statistico appropriato le differenze di partenza fra i quartieri ha fatto cambiare drasticamente la stima dell'effetto di URP sul vandalismo. Invece che una stima di + 8,87 abbiamo ora una stima (non significativa) di - 0,164 (cioè praticamente uguale a zero). Le stime del modello (7) implicano che URP non fa diminuire il tasso di vandalismo, ma non lo fa neppure aumentare.

È questa la stima *vera* dell'effetto di URP sul tasso di vandalismo? Non lo sappiamo. La plausibilità di questa stima è condizionata alla plausibilità dell'assunto che le differenze di partenza tra i due gruppi di quartieri siano dovute solamente alle differenze in termini di disoccupazione e immigrazione. È tuttavia possibile, anzi probabile, che permangano differenze di partenza tra i due gruppi di quartieri che non sono catturate dalle due variabili di controllo utilizzate: è possibile cioè che con questo modello non si sia eliminata del tutto la distorsione da selezione.

Questo problema può essere affrontato *allungando la lista* delle variabili incluse che possono rappresentare differenze di partenza rilevanti: ma per quanto sia lunga questa lista, non saremo mai completamente sicuri di avere catturato tutte le differenze causate dal processo di selezione. Nonostante la sua apparenza di scientificità, neppure l'analisi di regressione può fornire risposte certe e inconfutabili sugli effetti di una politica. Il vantaggio della regressione è quello di eliminare parte della distorsione dovuta a

differenze di partenza, ma il suo (ovvio) limite è quello di limitarsi alle caratteristiche osservabili.

5.3 La stima differenza-nelle-differenze ottenuta mediante la regressione

Nel capitolo precedente abbiamo esplorato un metodo importante per ridurre le differenze di partenza quando siano disponibili osservazioni pre-trattamento sulla variabile-risultato: sottrarre, alla differenza tra trattati e non-trattati osservata *dopo* che la politica è stata attuata, la differenza tra trattati e non-trattati osservata *prima* che la politica sia stata attuata. Questa è l'essenza del metodo differenza-nelle-differenze. Anche questo metodo soffre peraltro di potenziali distorsioni, che sono evidenziabili con un'espressione parallela alla (1):

$$(9) \quad \text{differenza-nelle-differenze} = \text{effetto della politica} + \text{differenze nei trend pre-intervento}$$

La stima differenza-nelle-differenze “rivela” l'effetto della politica solo se *non* ci sono, tra trattati e non-trattati, *differenze di partenza negli andamenti temporali* della variabile-risultato. Se queste differenze invece esistono (e non possono essere eliminate con “altri dati”, come è stato fatto alla fine del capitolo precedente), quella ottenuta con la differenza-nelle-differenze sarà una stima ancora *distorta* dell'effetto della politica. Il metodo differenza-nelle-differenze elimina le differenze di partenza nei *livelli* della variabile-risultato, ma resta vulnerabile alle differenze di partenza nei *trend* della stessa variabile.

Tuttavia, è possibile anche in questo caso utilizzare la regressione multipla per tenere conto (almeno in parte) delle differenze di partenza nei trend. Vediamo però prima come la stima differenza-nelle-differenze ottenuta nel capitolo precedente può essere replicata mediante una regressione semplice. A questo fine, occorre specificare un modello in cui la variabile dipendente è espressa come differenza, per l'*i*-esimo quartiere, tra il valore del tasso di vandalismo osservato nel 2004 e quello osservato nel 2000. Indicheremo questa versione “differenziata” della variabile-risultato come ΔY_i . Ogni valore di ΔY_i incorpora quindi due valori della variabile-risultato, uno pre e uno post.

Questo modello può essere scritto come segue.

$$(10) \quad \Delta Y_i = \alpha + \beta T_i + \varepsilon_i$$

in cui la variabile-trattamento T è l'unica variabile esplicativa. Essendo la variabile dipendente espressa come differenza tra due anni e non più come livello, il significato dei coefficienti α e β cambierà radicalmente. Vediamo più in dettaglio come vanno interpretati questi coefficienti, utilizzando alcune manipolazioni algebriche. Il significato dell'intercetta α sarà il seguente:

$$(11) \quad \alpha = E(\Delta Y_i | T_i = 0) = E(Y_{i,2004} - Y_{i,2000} | T_i = 0)$$

Ora α stima la “dinamica spontanea”, cioè il cambiamento osservato nei quartieri NonURP tra il 2000 e il 2004, mentre

$$(11') \quad \alpha + \beta = E(\Delta Y_i | T_i = 1) = E(Y_{i,2004} - Y_{i,2000} | T_i = 1)$$

stima il cambiamento osservato per i quartieri URP tra il 2000 e il 2004. Per differenza si ricava che:

$$(11'') \quad \beta = E(Y_{i,2004} - Y_{i,2000} | T_i = 1) - E(Y_{i,2004} - Y_{i,2000} | T_i = 0)$$

Il lato destro dell'espressione (11'') rappresenta null'altro che la “differenza-nelle-differenze” vista al capitolo precedente. Come si nota dalla Tabella 5.3, i coefficienti stimati dell'equazione (10) confermano ancora una volta l'equivalenza tra la regressione semplice e la differenza tra medie.

Tabella 5.3: Stime della regressione (10)

Variabile	Simbolo	Coefficiente	Errore standard	t-statistico	p-value
Costante	α	11,13	0,122	91,40	0,000
T = trattamento	β	- 7,66	0,298	-25,68	0,000

Otteniamo nuovamente la stima di un effetto di URP pari a 7,66 atti di vandalismo *in meno*. L'unico altro coefficiente stimato è l'intercetta, che in questo modello non rappresenta un livello iniziale, bensì la “dinamica spontanea” del fenomeno (cioè la crescita del tasso di vandalismo osservata tra il 2000 e il 2004 nei quartieri NonURP).

La stima dell'effetto di URP ottenuta con questo modello dipende però sempre dall'assunto di “parallelismo”, cioè di assenza di differenze nei trend tra i due gruppi di quartieri. L'aggiunta di variabili esplicative al modello (10) serve a “catturare” possibili differenze nei trend correlate con le caratteristiche osservabili: nel nostro caso, la speranza è che le differenze tra i quartieri nel tasso di disoccupazione e di immigrazione nel 2000 servano a depurare la stima dell'effetto di URP dalle distorsioni dovute alle differenze nei trend. Il modello di regressione che serve a questo scopo è il seguente:

$$(12) \Delta Y_i = \alpha + \beta T_i + \theta \text{IMMIGRAZ}_i + \lambda \text{DISOCCUP}_i + \varepsilon_i$$

La Tabella 5.4 contiene le stime del modello (12). Il tasso di immigrazione osservato nel 2000 non sembra influenzare il trend di crescita del tasso di vandalismo tra il 2000 e il 2004: infatti il coefficiente di questa variabile ha un *p-value* molto elevato. Conclusione diversa si raggiunge per il tasso di disoccupazione, il cui coefficiente è positivo e significativamente diverso da zero: questo significa che i quartieri con più alta disoccupazione tendono ad essere quelli che presentano un maggiore trend di crescita nel tasso di vandalismo. Poiché i quartieri URP tendono ad avere un livello di disoccupazione più elevato, parte della crescita del vandalismo in questi quartieri è spiegato dal loro alto tasso di disoccupazione.

Tabella 5.4: Stime della regressione (12)

Variabile	Simbolo	Coefficiente	Errore Standard	t-statistico	<i>p-value</i>
Costante	α	8,72	0,702	12,43	0,000
T= trattamento	β	- 8,10	0,307	-26,42	0,000
IMMIGRAZ	θ	0,766	0,038	0,740	0,460
DISOCCUP	λ	0,164	0,037	4,420	0,000

La stima dell'effetto di URP generata da questo modello risulta essere leggermente superiore (8,10 atti vandalici *in meno*) a quella prodotta dal modello differenza-nelle-differenze senza variabili di controllo (- 7,66). Aver incluso il tasso di disoccupazione come variabile di controllo ha “depurato” la stima dell'effetto di URP da una (piccola) distorsione verso valori positivi.

5.4 Il *matching* statistico

Torniamo ora alla situazione descritta all'inizio del capitolo, quella in cui abbiamo accesso ai dati sulla variabile-risultato *solamente* per il periodo post-intervento, cioè per il 2004: invece per il 2000 supponiamo come prima di avere informazioni sulle caratteristiche di ciascun quartiere (più precisamente i tassi di disoccupazione e di immigrazione) ma non sulla variabile-risultato.

Esploriamo ora un metodo alternativo a quello della regressione multipla, il *matching* (o abbinamento) statistico. La differenza tra i due metodi è nel modo in cui essi utilizzano i dati a disposizione, non nei dati a cui fanno ricorso, che sono (in buona parte) gli stessi. Con la regressione, imponiamo sui dati un modello *parametrico*, cioè supponiamo di sapere quale forma abbia la relazione tra la variabile dipendente e le variabili di controllo: nel caso esaminato, abbiamo supposto una relazione lineare. Potremmo rendere il modello meno lineare (mediante l'aggiunta di interazioni o di funzioni polinomiali delle variabili di controllo), ma comunque faremmo sempre qualche assunto arbitrario sulla forma della relazione tra le variabili.

Il metodo del *matching* statistico si basa su un'idea diversa, che trae ispirazione dal metodo sperimentale: l'idea è quella di creare un gruppo di controllo *ex post*, scegliendo tra i non-trattati quel sottogruppo che sia il più simile possibile al gruppo dei trattati in termini di caratteristiche osservabili. La procedura adottata per scegliere il gruppo di controllo *ex post* tenta nuovamente di eliminare le differenze di partenza che il processo di selezione ha generato tra i due gruppi, cioè lo stesso obiettivo perseguito dall'analisi di regressione.

Una volta scelto il gruppo di controllo *ex post*, la stima dell'effetto si otterrà semplicemente calcolando la differenza tra la media della variabile-risultato nel gruppo dei trattati e nel gruppo dei controlli abbinati. Questo è quindi un approccio *non-parametrico*, nel senso che non richiede di imporre una precisa forma alla relazione tra variabile-risultato e variabili di controllo.

Notiamo come questa strategia si colleghi immediatamente all'idea di controfattuale: potremmo dire che la media della variabile-risultato delle unità non-trattate *abbinate* rappresenta la nostra migliore stima del controfattuale. Come prevediamo si sarebbero

comportate le unità trattate *in assenza di trattamento*? Come le unità non-trattate che sono più simili alle trattate. Va tuttavia sottolineato come anche questo metodo si basi sullo *stesso assunto della regressione*, cioè che le differenze di partenza tra le caratteristiche *non* osservabili di trattati e non-trattati siano nulle o “ignorabili”, una volta che si sia tenuto conto delle caratteristiche osservabili.

Le procedure di *matching* sono più d’una: prima di esaminarne alcune in dettaglio, soffermiamoci su un’altra differenza tra il *matching* e la regressione, una differenza che motiva ulteriormente l’utilizzo del primo rispetto alla seconda. La differenza consiste in questo: la regressione utilizza tutte le osservazioni disponibili, e quindi anche le osservazioni di soggetti non-trattati che siano molto diversi dai soggetti trattati: queste differenze sono oscurate dalla procedura di stima utilizzata dalla regressione, che è una sorta di “scatola nera”. Viceversa, il *matching* mette in evidenza se vi sono unità non-trattate molto diverse dalle unità trattate e tende a non usarle per l’abbinamento e quindi per la stima dell’effetto. In altre parole, il *matching* per sua natura utilizza solo i soggetti *più confrontabili*.

Facciamo un esempio per sviluppare questo punto. Poniamo che i 50 quartieri URP siano caratterizzati da tassi di disoccupazione medio-alti, mentre tutti i 250 quartieri NonURP abbiano al contrario tassi medio-bassi. Ciò non costituisce un problema per la regressione. È comunque possibile stimare il modello (7) e la retta cercherà di interpolare al meglio i dati disponibili: la relazione tra vandalismo e disoccupazione sarà stimata con i dati NonURP dove la disoccupazione è molto bassa (e non ci sono dati URP) e con i dati URP dove la disoccupazione è molto alta (e non ci sono dati NonURP).

Il *matching* funziona diversamente: avendo necessità di confrontare unità simili, esso semplicemente non può essere utilizzato a meno che unità trattate e unità non-trattate abbiano un *supporto comune*, cioè esistano unità trattate e unità non-trattate con valori simili nelle variabili di controllo. Nel nostro esempio, nel gruppo di controllo *ex post* finiranno solo i quartieri NonURP che tendono ad avere un tasso di disoccupazione medio (e, a seconda della procedura adottata, solo i quartieri URP con disoccupazione media saranno utilizzati).

In sintesi: il metodo del *matching* non richiede un assunto (spesso difficile da giustificare) sulla forma della relazione tra variabile-risultato e variabili di controllo, perché si basa proprio sull'abbinamento tra i valori delle variabili di controllo stesse. In questo senso il *matching* è meno restrittivo della regressione. D'altro canto, esso impone che esistano unità trattate e non-trattate che abbiano caratteristiche simili, pena l'impossibilità di abbinarle adeguatamente. Quindi da questo punto di vista il *matching* è più restrittivo della regressione rispetto ai dati che si possono utilizzare.

5.5 Come si realizza in pratica il *matching* statistico

La particolare procedura di *matching* dipende da due scelte: (a) la scelta della misura di distanza tra unità (per definire *la somiglianza o la diversità* delle unità tra loro); e (b) la scelta della tipologia di abbinamento (il criterio per scegliere *quante unità abbinare e come*, sulla base della distanza).

La misura di distanza oggi più comunemente usata si basa sul cosiddetto *propensity score* (indice di propensione). Il *propensity score* di una unità (trattata o non-trattata) è la probabilità che l'unità venga assegnata al trattamento *dato le sue caratteristiche osservabili* prima del trattamento (ignorando quindi il fatto che sia stata realmente trattata o non-trattata). Il vantaggio principale nell'utilizzo del *propensity score* sta nella possibilità di ridurre a una il numero delle variabili da utilizzare per calcolare le differenze tra unità, riducendo sensibilmente la mole e il tempo di lavoro, e semplificando la lettura dei risultati.

Come si ottiene questo *propensity score*? Il primo passo è quello di stimare, utilizzando *tutte* le unità trattate e non-trattate disponibili, un modello probabilistico detto *regressione logistica*, che mette in relazione il trattamento T (che qui diventa la variabile dipendente) con le caratteristiche osservabili dei soggetti. La relazione funzionale tra la probabilità di trattamento e le variabili esplicative è data dalla funzione logistica, che garantisce che la probabilità stimata cada nell'intervallo tra zero e uno⁴¹:

⁴¹ In alternativa è anche utilizzabile un modello *probit*. Per una trattazione di questi modelli si veda Fabbris L., *Statistica multivariata. Analisi esplorativa dei dati*, McGraw-Hill, Milano, 1997.

$$(13) \quad P(T = 1) = \frac{e^{(\alpha + \beta_1 X_1 + \dots + \beta_m X_m)}}{(1 + e^{(\alpha + \beta_1 X_1 + \dots + \beta_m X_m)})}$$

Una volta stimato, tale modello permetterà, per ogni unità, di *predire* la probabilità di trattamento in funzione del valore delle sue variabili esplicative: cioè produrrà, per ogni unità, il suo *propensity score* (che non sarà altro che un numero tra zero e uno). Quelle unità le cui caratteristiche implicano un’alta probabilità di trattamento (nel nostro caso, un’alta probabilità di ricevere i finanziamenti URP) avranno un *propensity score* più vicino a 1 di quelle unità che hanno una bassa probabilità di trattamento. Va ribadito che il *propensity score* si calcola per *tutte* le unità (trattate e non-trattate), usando *esclusivamente* informazioni pre-trattamento.

Una volta che disponiamo del *propensity score* per tutte le unità, possiamo utilizzare come distanza tra due unità la differenza tra i loro *propensity score* (che indichiamo con p_i nel caso dell’unità trattata e p_j nel caso della non-trattata):

$$d_{ij} = | p_i - p_j |$$

Una volta definita la distanza tra le unità, è possibile procedere all’abbinamento di unità trattate e non-trattate quindi alla costruzione del gruppo di controllo *ex post*.

Esistono diverse metodologie per eseguire l’abbinamento. A scopo esemplificativo ne illustriamo quattro, che sono utilizzate in alcuni dei casi illustrati nella seconda parte del volume⁴². Quella che segue è una trattazione fatta in termini intuitivi, rimandando all’Appendice 1 la trattazione più formalizzata.

L’abbinamento sull’unità più vicina (*Nearest neighbour matching*) rappresenta il metodo più semplice e intuitivo di condurre l’abbinamento, in quanto consiste semplicemente nell’abbinare ad ogni unità trattata quella *particolare* unità non-trattata che ha il *propensity score* più vicino numericamente. Quindi il “gruppo dei controlli” (così chiameremo per semplicità l’insieme della unità non-trattate che sono abbinate ad una singola unità trattata) è rappresentato da *una sola* unità non-trattata.

⁴² Per una trattazione più approfondita di queste procedure, si veda Sasha Becker e Andrea Ichino, “*Estimation of average treatment effects based on propensity scores*”, *The Stata Journal*, 2002. Per una trattazione altamente formalizzata del *matching* si veda Guido Imbens, “*Semiparametric Estimation of Average Treatment Effects under Exogeneity: A Review*” *Review of Economics and Statistics*, 86, 2004, pp. 4-29.

La selezione delle unità di controllo viene solitamente effettuata *con ripetizione (o reimmissione)*: quindi è possibile assegnare la stessa unità non-trattata ripetutamente a più unità trattate. Il numero di unità non-trattate inserite nel gruppo di controllo *ex post* può essere inferiore a quello delle unità trattate.

Una volta effettuato l'abbinamento, la stima dell'effetto della politica è semplicemente ricavata tra la media della variabile-risultato tra i trattati e la media calcolata tra i controlli abbinati. Un limite di questo metodo è che è possibile abbinare ad alcune unità trattate delle unità non-trattate con un *propensity score* molto distante, pur essendo il più vicino tra quelli disponibili. Questo limite è superabile stabilendo una distanza minima che deve essere rispettata tra i due *propensity score* per considerare le due unità abbinabili. Questa accortezza è la caratteristica distintiva del secondo metodo detto "Abbinamento in un raggio" (*Radius matching*). Con questo metodo ad ogni unità trattata sono abbinata tutte le unità di controllo il cui *propensity score* ha una distanza minore o uguale a un certo "raggio" δ (che sarà tendenzialmente un numero molto vicino a zero: per esempio 0,01). Tralasciando per un attimo la questione della scelta di δ , notiamo subito due aspetti che differenziano il risultato di questa procedura rispetto alla precedente: (a) alcune unità trattate potrebbero venire *scartate* in quanto non si trova *nessuna* unità non-trattata con un *propensity score* che cade nell'intervallo $p_i \pm \delta$; (b) *più di una unità non-trattata* può venire abbinata ad una singola unità trattata, quando viceversa si trovano *più unità* non-trattate con un *propensity score* che cade nell'intervallo $p_i \pm \delta$.

La scelta del raggio δ va quindi fatta considerando che, se il raggio è molto piccolo, si perderanno alcune unità trattate ma si avrà il vantaggio di effettuare un confronto tra unità molto simili. Al contrario, un raggio ampio permetterà di aumentare il numero dei controlli, col rischio però che questi siano meno simili all'unità trattata di quanto si creda. Una volta effettuato l'abbinamento, la stima dell'effetto è derivata in modo simile al metodo precedente, con l'unica differenza che, se se si abbinano più unità non-trattate alla stessa unità trattata perché il loro *propensity score* sta all'interno dell'intervallo $p_i \pm \delta$, si calcola la media aritmetica della loro variabile-risultato. Lo svantaggio potenziale del *radius matching* è la possibile perdita di unità trattate, come vedremo

nell'applicazione ai dati di URP. Questo svantaggio tuttavia è compensato dal fatto di garantire che il confronto sia condotto tra unità trattate e non-trattate molto simili tra loro.

Il terzo metodo, detto “Abbinamento con stratificazione” (*Stratification matching*) consiste nel suddividere il campo di variazione del *propensity score* in intervalli (o strati) tali che, all'interno di ciascun intervallo, le unità trattate e le unità non-trattate abbiano lo stesso valore medio del *propensity score*. Ad esempio, il primo strato potrà contenere tutte le unità con *propensity score* inferiore a 0,2; il secondo strato conterrà le unità con *propensity score* tra 0,2 e 0,3; e così via....)⁴³.

Una volta ottenuta la stratificazione che soddisfa questa proprietà, per il calcolo dell'effetto della politica si procede in due passi successivi. Prima si calcola, all'interno di ogni strato, la differenza tra le medie della variabile-risultato calcolate separatamente tra le unità trattate e tra le unità di controllo, ottenendo quindi un effetto limitato al particolare strato. L'effetto complessivo è calcolato come media delle differenze nei vari strati, ponderate con il numero di unità trattate presenti (si dà quindi maggiore peso agli strati in cui si concentrano maggiormente le unità trattate). Un possibile svantaggio del metodo della stratificazione è il fatto che le osservazioni negli strati in cui non compaiono unità trattate o unità non-trattate non possono venire utilizzate.

Il quarto e ultimo metodo detto “Abbinamento con la funzione kernel” (*Kernel matching*) a rigore non rappresenta una vera e propria procedura di abbinamento, in quanto ad ogni unità trattata sono “abbinate” *tutte* le unità non-trattate, pesate in modo inversamente proporzionale alla distanza del loro *propensity score* da quello dell'unità trattata. Il principale vantaggio del metodo di abbinamento basato sulla funzione *kernel* è il fatto di utilizzare tutta l'informazione disponibile, in quanto sia tutte le unità trattate sia tutte le unità non-trattate sono incluse nella stima.

⁴³ Una proprietà che il *propensity score* deve avere in generale, anche con altre procedure di *matching*, è quella del *bilanciamento*: all'interno di ogni strato (definito in modo analogo a quanto fatto per la procedura di stratificazione) non ci devono essere differenze significative tra unità trattate e unità di controllo nelle variabili esplicative osservate.

5.6 L'applicazione del *matching* statistico al caso URP

Cominciamo stimando sui dati URP una regressione logistica, che mette in relazione la partecipazione al programma URP con i tassi di immigrazione e di disoccupazione. Il modello stimato con due sole variabili esplicative avrà la seguente forma:

$$(13') \quad P(T = 1) = \frac{e^{(\alpha + \theta IMMIGR_i + \lambda DISOCC_i)}}{(1 + e^{(\alpha + \theta IMMIGR_i + \lambda DISOCC_i)})}$$

La tabella seguente contiene le stime dei coefficienti della regressione logistica rappresentata dall'equazione (13').

Tabella 5.5: Stime del modello logistico (13')

Variabile	Simbolo	Coefficiente	Errore standard	t-statistico	p-value
Costante	α	-7,53	1,320	-5,72	0,000
IMMIGRAZ	θ	0,109	0,061	1,80	0,072
DISOCCUP	λ	0,325	0,062	5,24	0,000

Le stime dei coefficienti possono essere utilizzate per calcolare il *propensity score* per ciascuna unità (trattata e non-trattata). Il *propensity score* sarà prodotto dalla seguente formula:

$$(14) \quad p_i = \frac{e^{(-7,53 + 0,109 IMMIGR_i + 0,325 DISOCC_i)}}{(1 + e^{(-7,53 + 0,109 IMMIGR_i + 0,325 DISOCC_i)})}$$

È chiaro dalla formula (14) come per valori alti di immigrazione e disoccupazione il *propensity score* tenda a uno, mentre per valori bassi tenda a zero, in quanto entrambi i coefficienti sono positivi. L'intensità della relazione tra la probabilità di ottenere i finanziamenti e la caratteristica del quartiere dipende dal valore del coefficiente stimato, che è più grande per il tasso di disoccupazione che per il tasso di immigrazione.

La Tabella 5.6 mostra, a scopo puramente esemplificativo, il *propensity score* stimato per alcuni quartieri URP e NonURP.

Tabella 5.6. *Propensity score* stimato per alcuni quartieri

Identificativo del quartiere	Tasso di disoccupazione	Tasso di immigrazione	Trattamento ricevuto	Valore del <i>propensity score</i>
270	16,2	7,7	URP	0,1953
15	16,4	8,2	NonURP	0,2148
295	7,7	16,2	URP	0,0372
131	8,2	16,4	NonURP	0,0444
258	8,9	13,3	URP	0,0399
134	9,6	14,9	NonURP	0,0586
281	17,1	15,9	URP	0,2867
128	12,3	15,0	NonURP	0,1315

È evidente come quartieri con tassi di immigrazione e (soprattutto) disoccupazione simili tendano ad avere *propensity score* simili. È altrettanto evidente (si osservino le prime quattro righe) come unità con tassi simili ma “scambiati” (uno con immigrazione alta e disoccupazione bassa, l’altro con immigrazione bassa e disoccupazione alta) non abbiano *propensity score* simili e non siano dunque facilmente abbinabili per il confronto (al contrario di quanto si sarebbe fatto, in pratica, utilizzando il modello di regressione lineare).

Una volta stimato il *propensity score*, è possibile procedere alla stima dell’effetto con le procedure di *matching* descritte più sopra. La Tabella 5.7 riporta gli effetti stimati ottenuti utilizzando la singola procedura, i relativi errori standard, nonché il numero di unità trattate e di controlli utilizzati nell’abbinamento.

Tabella 5.7: Stime dell’effetto di URP ottenute con diverse procedure di *matching*

Procedura di <i>matching</i>	No. di unità trattate utilizzate	No. di controlli utilizzati	Effetto stimato	Errore standard	t-statistico
1. Sull’unità più vicina	50	30	– 7,75	2,15	– 3,60
2. In un raggio ($\delta=0,01$)	20	24	– 8,51	5,47	– 1,55
3. Con stratificazione	50	250	– 5,33	1,78	– 3,00
4. Con funzione <i>kernel</i>	50	250	– 4,91	2,23	– 2,21

La prima riga della Tabella 5.7 contiene i risultati dell'abbinamento sull'unità più vicina. Come previsto, a fronte di 50 unità trattate, l'abbinamento è stato fatto utilizzando solo 30 unità non-trattate, quelle che possiedono un *propensity score* più simile a quello delle unità trattate. L'effetto stimato è di $-7,75$ atti vandalici ed è statisticamente significativo: una stima molto diversa da quella ottenuta stimando il modello di regressione (7), che era di $-0,164$ atti vandalici (cioè praticamente zero).

La seconda riga contiene i risultati dell'abbinamento in un raggio (posto pari a 0,01). Come previsto, sono state utilizzate solo una parte delle unità trattate (solamente 20 su 50), che sono state abbinate a 24 unità non-trattate. L'effetto stimato non varia di molto dal precedente, attestandosi su $-8,51$ atti vandalici, ma la ridotta dimensione del campione utilizzato fa sì che la stima non sia statisticamente significativa.

I risultati del metodo della stratificazione sono mostrati nella terza riga. Come anticipato, tutte e 300 le osservazioni sono state utilizzate. L'effetto stimato è significativo ma inferiore a quello stimato con gli altri due metodi e con la regressione, attestandosi sui $-5,33$ atti vandalici.

Infine, la quarta riga mostra le stime ottenute con il metodo *kernel*. L'effetto stimato è simile al precedente ed è statisticamente significativo. Per definizione, sono state utilizzate tutte 50 le unità trattate e le 250 unità non-trattate, seppure assegnando alle unità non-trattate con pesi via via più piccoli, man mano che il loro *propensity score* si allontana da quello di ciascuna unità trattata.

CAPITOLO 6

I METODI BASATI SULLA DISCONTINUITÀ NEL TRATTAMENTO

di Alberto Martini

Si incontrano spesso situazioni in cui l'esposizione o meno ad un trattamento (sia esso un servizio, un sussidio, un'agevolazione, persino un'imposta) è determinata da una precisa *regola di ammissione*, anziché da un processo di autoselezione. I metodi di valutazione discussi nella prima parte di questo capitolo si applicano appunto nelle situazioni in cui la separazione tra trattati e non-trattati è determinata da una regola di ammissione che sia riconducibile allo schema seguente: viene ammesso al trattamento chi si posiziona al di sopra (o, a seconda dei casi, al di sotto) di un determinato *valore-soglia* su una certa *variabile*, mentre chi sta sotto (o sopra) viene escluso. Un esempio di tale variabile può essere l'età. L'esenzione dal pagamento del ticket sui medicinali può essere condizionata al compimento di un certo anno di età da parte del paziente: ad esempio, chi è oltre i 60 anni ne è esentato, mentre chi è più giovane è tenuto a pagarlo. In un altro esempio, l'indennità di mobilità in seguito a licenziamento collettivo viene corrisposta al lavoratore licenziato per la durata di un anno se il lavoratore ha meno di 40 anni, per due anni se è sopra i 40, per tre anni se è sopra i 50 (qui il "trattamento" è rappresentato dagli anni aggiuntivi di indennità).

Queste regole di ammissione, usate di frequente in ambito pubblico, creano una *discontinuità* netta nel trattamento, per cui chi è posizionato su un lato della soglia si trova in un regime di trattamento diverso rispetto a chi è immediatamente sull'altro lato, pur avendo valori molto simili. I metodi discussi in questo capitolo si basano sull'idea che quando si confrontano individui situati *subito sotto* e *subito sopra* la soglia, ci si trovi in una situazione molto simile alla randomizzazione, per cui l'effetto del trattamento può essere stimato confrontando direttamente i risultati ottenuti dai due gruppi degli ammessi e degli esclusi intorno alla soglia, senza incorrere nelle distorsioni tipiche delle situazioni in cui c'è autoselezione.

Alcuni definiscono questa situazione, intermedia tra esperimento e autoselezione, come "esperimento naturale". Nella letteratura valutativa per "esperimento naturale" si

intende la situazione in cui l'esposizione ad una politica pubblica è determinata da un meccanismo "esogeno", che gli individui coinvolti non controllano completamente. Alcuni soggetti sono esposti alla politica e altri no, ma non sono loro a decidere: almeno in parte, la decisione viene loro imposta. Oppure sono esposti alla politica dopo o prima di una certa data, ma non sono loro a decidere la tempistica dell'esposizione. Quindi si tratta di "esperimenti" perché non c'è autoselezione: ma sono "naturali", e quindi diversi da quelli "randomizzati", in quanto non c'è manipolazione da parte del valutatore su chi è esposto e chi no al trattamento.

Un caso frequente di esperimento naturale è quello connesso all'uso di *graduatorie* per determinare chi può accedere o meno ad un certo servizio, o ricevere un sussidio, o beneficiare di un'agevolazione. Tutti coloro che "fanno domanda", per definizione, vogliono accedervi, ma non per tutti è possibile, a causa della scarsità di risorse: ogni soggetto che fa richiesta riceve un *punteggio* (a sua volta determinato da "qualcuno" sulla base delle sue caratteristiche osservabili) e viene ammesso al servizio/sussidio solo chi è sopra un certo *punteggio-soglia* (predeterminato o stabilito in base alle risorse disponibili). Attorno alla soglia di ammissione si crea quindi una discontinuità, una frattura netta, nell'esposizione alla politica. È proprio questa frattura netta che permette di stimare l'effetto della politica in modo più plausibile che in altre situazioni non-sperimentali.

Per illustrare concretamente questo approccio ritorniamo al caso di URP visto nei capitoli precedenti. Supponiamo che la selezione dei quartieri per l'ammissione ai finanziamenti URP sia stata fatta sulla base di una gara, una supposizione non lontana da quello che succede nella realtà. Supponiamo inoltre che alla gara abbiano partecipato solo 150 delle 300 città considerate nei capitoli precedenti. A questo primo stadio si tratta di un processo di autoselezione, che però non porta direttamente all'esposizione alla politica, ma solo alla partecipazione alla gara. Da questo punto in poi, per i 150 partecipanti alla gara, l'accesso ai finanziamenti URP non dipende da una scelta, bensì dal punteggio deciso da una commissione giudicatrice formata da esperti. Quindi ciò che separa i 50 vincitori dai 100 perdenti non è più un processo di autoselezione, bensì una regola di ammissione basata su un punteggio e una soglia.

Supponiamo poi che il punteggio in base al quale è stilata la graduatoria finale sia assegnato dalla commissione giudicatrice tenendo conto di due insiemi di fattori: da una parte quelle caratteristiche del quartiere che rappresentano il “bisogno” di azioni di recupero urbano, tra le quali anche il tasso di vandalismo; dall’altra il giudizio dato dalla commissione stessa sulla bontà del progetto presentato, che rappresenta il lato del “merito”. Questo meccanismo fa sì che il punteggio sia, anche se non perfettamente, correlato con le condizioni di degrado del quartiere: a parità di bontà del progetto presentato, i quartieri *più degradati* riceveranno un punteggio *più alto*, i quartieri meno degradati un punteggio più basso.

Poniamo che la soglia critica sia un punteggio di 71,5: questo valore è stato scelto in modo che i primi 50 quartieri ricevano i finanziamenti, mentre i successivi 100 quartieri in graduatoria non ricevano finanziamenti. La seguente tabella riporta minimo, massimo e media dei punteggi ottenuti, separatamente per i quartieri URP e NonURP. I punteggi ottenuti dai 100 quartieri NonURP vanno da 59,8 a 71,4, quelli ricevuti dai 50 quartieri URP da 71,5 a 85,2.

Tabella 6.1: Punteggi ottenuti dai quartieri partecipanti alla gara URP

	N	minimo	massimo	medio
NonURP	100	59,8	71,4	65,7
URP	50	71,5	85,2	75,5
Totale	150	59,8	85,2	69,0

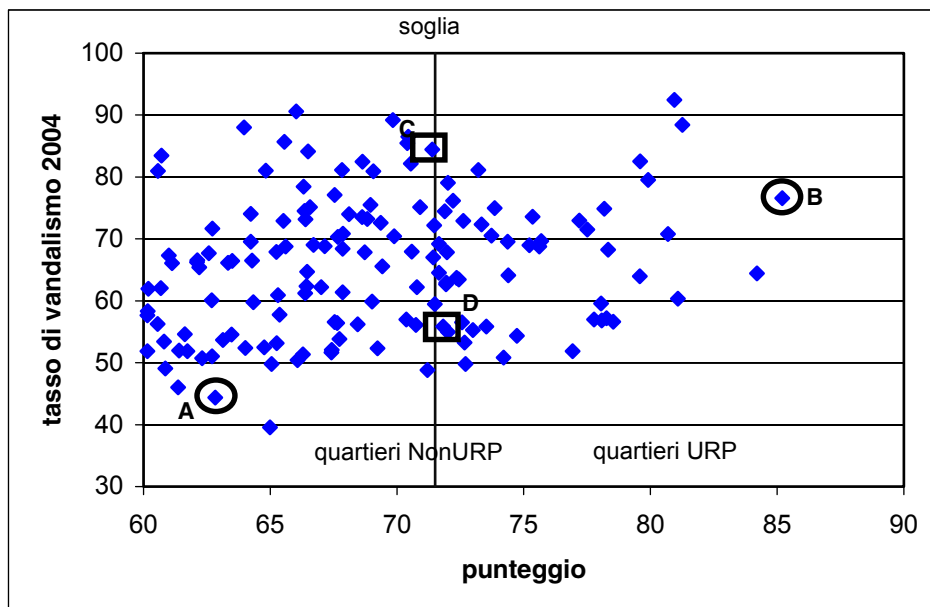
Occorre ribadire come ci si trovi in una situazione diversa da quelle viste nei capitoli precedenti. Qui abbiamo una *chiara regola di ammissione* al trattamento. I due gruppi di quartieri possono differire lungo un’infinità di dimensioni, osservabili e non osservabili, possono essere soggetti a dinamiche diverse, ma alla fine ciò che ne determina l’esposizione al trattamento è *una e una sola variabile*, il punteggio. Quindi non è il “caso” a determinare l’esposizione alla politica (come negli esperimenti con randomizzazione), ma neppure la decisione dei singoli (come nei casi di autoselezione): è piuttosto una regola che si riassume in una variabile, osservabile anche da parte del valutatore.

Va richiamato un altro fatto importante: in base ai criteri utilizzati dalla commissione, il punteggio ottenuto sarà correlato *positivamente* con il tasso di vandalismo. Ciò significa che i quartieri alti in graduatoria avranno “per costruzione” un tasso di vandalismo mediamente più alto nel 2000 rispetto ai quartieri bassi in graduatoria. E nella misura in cui il tasso di vandalismo è correlato nel tempo, questi quartieri tenderanno ad avere un tasso di vandalismo mediamente più alto anche nel 2004: questo *non* per effetto di URP, ma come conseguenza diretta della procedura di selezione.

Il grafico seguente mostra sull’asse orizzontale il punteggio ottenuto dai 150 quartieri partecipanti alla gara nel 2000, e sull’asse verticale il tasso di vandalismo nel 2004. La linea verticale al centro del grafico indica la soglia di ammissione: i quartieri a destra della soglia sono ammessi al programma URP, quelli a sinistra della soglia non lo sono. Concentriamo l’attenzione sui quartieri *attorno alla soglia*, cioè i “primi esclusi” e gli “ultimi inclusi”: i due gruppi sono *molto simili* in termini di punteggio, ma sono *diversi* in termini di trattamento. Quindi nelle *immediate* vicinanze della soglia si determina una situazione paragonabile alla randomizzazione: differenze minime (e quindi praticamente casuali) in termini di punteggio determinano grosse differenze in termini di trattamento, anzi determinano l’inclusione di qualcuno e l’esclusione di qualcun altro⁴⁴.

⁴⁴ E vale anche la seguente considerazione, all’apparenza paradossale: quanto più il punteggio è determinato in modo imperfetto, cioè la graduatoria è “fatta male”, tanto più la situazione che si crea diventa simile alla randomizzazione, quindi preferibile per valutare gli effetti della politica.

Figura 6.1: Relazione tra tasso di vandalismo nel 2004 e punteggio assegnato



Nel grafico si nota una (seppur lieve) correlazione positiva tra il punteggio e il tasso di vandalismo nel 2004: i quartieri con punteggi molto alti (sulla destra del grafico) tendono anche ad avere tassi di vandalismo crescenti. Se calcoliamo la media del tasso di vandalismo in ciascuno dei due gruppi di quartieri (i 50 inclusi e i 100 esclusi), otteniamo un tasso di vandalismo medio *più alto* tra i quartieri URP rispetto ai quartieri NonURP: tuttavia, non siamo nuovamente in grado di sapere in che misura questa differenza sia dovuta alle differenze di partenza (qui riassunte dalle differenze di punteggio), e quindi in che misura queste ultime “mascherino” un possibile effetto (di segno negativo) di URP sul vandalismo.

6.1 Il confronto attorno al punto di discontinuità

Procediamo ora con un ragionamento diverso. Confrontiamo la situazione dei quartieri attorno alla soglia con quella dei quartieri distanti dalla soglia di ammissione. Prendiamo prima i quartieri contrassegnati nel grafico dalle lettere A e B, collocati lontano dalla soglia. Questi due quartieri sono all'estremo opposto della graduatoria e quindi differiscono molto in termini di punteggio, oltre ad essere uno finanziato da URP e l'altro no. Quindi le forti differenze tra di loro in termini di vandalismo osservato

dopo l'attuazione di URP sono dovute a forti differenze di partenza: potrebbe essere che A *non* si trovasse in una situazione di degrado socio-economico già nel 2000 e anche per questo abbia ricevuto un punteggio basso, mentre B fosse molto degradato già nel 2000 e anche in ragione di questo fatto sia stato ammesso al finanziamento. In altre parole, A non sarebbe un buon "controfattuale" per B.

Ben diversa è la situazione immediatamente attorno alla soglia. Prendiamo i due quartieri contrassegnati con le lettere C e D. I loro punteggi sono quasi identici (71,4 e 72,0), quindi le loro differenze nel tasso di vandalismo non possono essere spiegate dalla loro posizione in graduatoria, che è quasi identica. È quindi possibile che tali differenze siano, almeno in parte, effetto della politica URP. Un effetto che andrebbe, si noti bene, nella direzione sperata: infatti D (quartiere URP) ha un tasso di vandalismo molto inferiore a C (quartiere NonURP). Detto diversamente, C potrebbe essere un buon controfattuale di D: cioè potrebbe rappresentare cosa sarebbe successo a D se non fosse stato esposto all'intervento di riqualificazione urbana.

Ovviamente due osservazioni non provano nulla, la posizione relativa di C e D potrebbe essere dovuta al caso. Per avere una stima statisticamente significativa dell'effetto della politica, occorre confrontare la *media* del tasso di vandalismo tra un numero molto più grande di quartieri vicini alla soglia. Ma qual è un numero sufficiente e quanto "vicino" alla soglia è sufficientemente vicino?

Non esiste alcuna regola precisa, ma esiste un chiaro *trade-off*: *quanto più larga* sarà la banda di osservazioni attorno alla soglia, tanto più le differenze di partenza/punteggio torneranno a distorcere il confronto, come nel caso di A e B; *quanto più stretta* la banda, tanto più piccola la dimensione campionaria e quindi tanto minore la precisione delle stime.

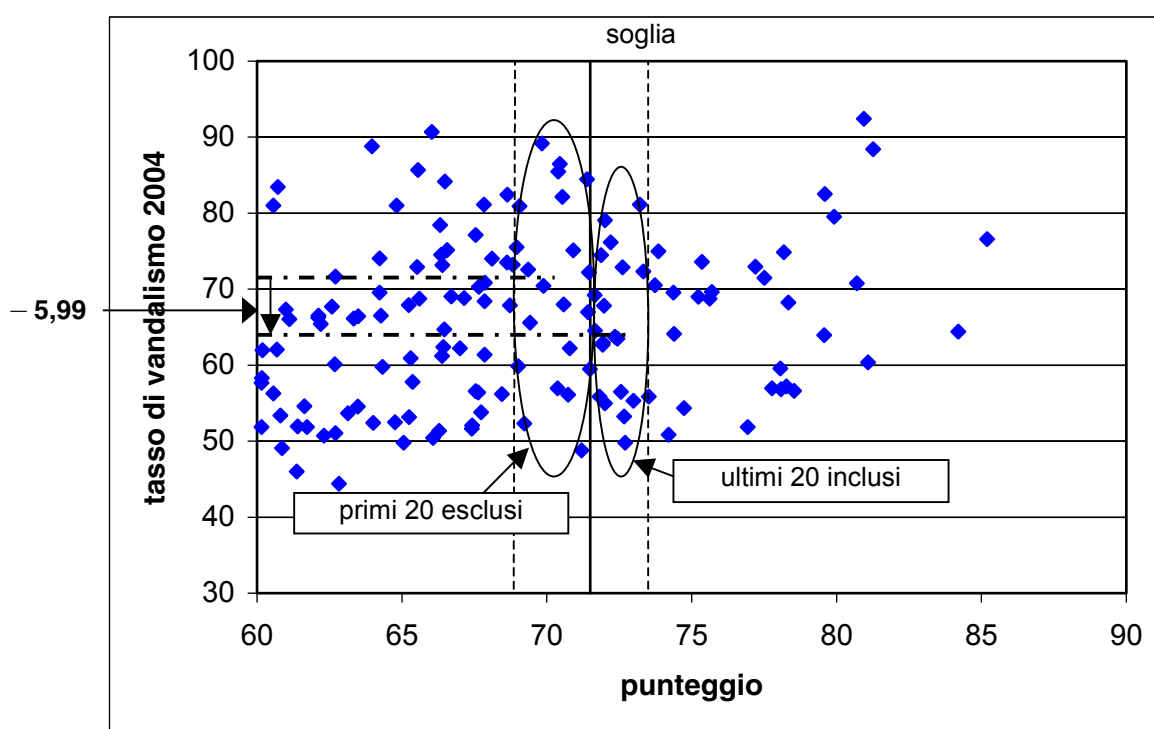
Nel grafico seguente mostriamo una banda che comprende da un lato i *primi 20 esclusi* e dall'altra gli *ultimi 20 inclusi*. I risultati numerici sono mostrati nella parte sinistra della Tabella 6.2. La media del tasso di vandalismo è ora più *bassa* nei 20 quartieri URP ultimi inclusi, con una differenza in negativo di - 5,99 atti vandalici. Tuttavia, questa differenza non risulta statisticamente significativa: la scelta di una banda ristretta, per avere due campioni il più possibile simili in termini di punteggio, ha determinato una

bassa precisione delle stime, che può migliorare soltanto ampliando la larghezza della banda – incorrendo però sempre più nel rischio di distorsione.

Tabella 6.2: Differenze nel tasso di vandalismo attorno alla soglia e tra tutti i quartieri

	20 quartieri sopra e 20 sotto la soglia			Tutti i 150 quartieri che partecipano alla gara		
	N	media	dev. st.	N	Media	dev. st.
URP	20	64,58	10,56	50	66,37	10,49
NonURP	20	70,57	12,00	100	65,53	12,63
differenza		- 5,99			+ 0,84	

Figura 6.2: Il confronto attorno al punto di discontinuità



Il metodo che abbiamo utilizzato per derivare la stima di $-5,99$ atti di vandalismo viene definito come “*confronto attorno al punto di discontinuità*”. Questo metodo ha un vantaggio e uno svantaggio. Il vantaggio è quello di essere semplice e intuitivo: non ha bisogno di un grosso apparato statistico, ma solo della differenza tra medie (come nel caso dell’esperimento “classico”). Lo svantaggio è l’arbitrarietà della scelta dei due gruppi da confrontare, arbitrarietà che può far pensare che il risultato ottenuto sia stato

costruito ad arte. Perché non ci siamo limitati alle prime 10 osservazioni sopra e 10 sotto la soglia? Perché non andare fino a 40 sopra e 40 sotto? Non c'è una risposta univoca a queste domande. Scegliere 40 e 40 avrebbe potuto garantire una maggiore precisione nella stima della differenza, ma l'avrebbe anche maggiormente "distorta", inducendo forti differenze di punteggio tra i due gruppi posti a confronto.

Notiamo quello che succede se invece consideriamo *tutti* i 150 quartieri che partecipano alla gara (parte destra della Tabella 6.2): la differenza nei tassi di vandalismo tra i quartieri URP e NonURP torna ad essere positiva, anche se vicina allo zero. Questo risultato è la combinazione di due spinte opposte: l'effetto di URP, che tende ad essere negativo, e l'effetto delle differenze di partenza, che tende ad avere segno positivo. Se ci teniamo vicini alla soglia, prevale l'effetto negativo di URP. Se allarghiamo la banda e ci allontaniamo dalla soglia, tende a prevalere sempre più l'effetto delle differenze di partenza.

6.2 Il regression-discontinuity design

Il metodo appena utilizzato sarebbe convincente e plausibile come metodo di valutazione degli effetti di una politica, se non fosse per la necessità di limitarsi a poche osservazioni attorno alla soglia, il che può generare stime imprecise. C'è un'alternativa a questa scelta: usare *tutte* le osservazioni disponibili, eliminando però, mediante un'appropriata tecnica statistica quale la regressione multipla, l'influenza delle differenze nel punteggio. Questo approccio è detto *regression-discontinuity design*⁴⁵.

Deve essere chiara la natura del *trade-off* che ci sta di fronte: nel paragrafo precedente *non* avevamo bisogno di un modello statistico ma allo stesso tempo *non* potevamo usare tutti i dati, pena non riuscire a separare l'effetto della politica da quello delle differenze di partenza; mentre ora usiamo *tutti* i dati, anche quelli distanti dalla soglia, ma

⁴⁵ Il riferimento classico sul *regression-discontinuity design* è Trochim W., *Research Design for Program Evaluation: The Regression-Discontinuity Approach*. Sage Publications, Beverly Hill, 1984. Di recente c'è stata una rinnovata attenzione per questo metodo da parte degli econometrici. Per una trattazione molto formalizzata si veda Hahn J., Todd P., Van der Klaauw W., "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design", *Econometrica*, Vol. 69, No. 1, pp. 201-209, 2001. Per una trattazione più accessibile Battistin E., Rettore E., "Another Look at the Regression Discontinuity Design", di prossima pubblicazione sul *Journal of Econometrics*.

dobbiamo avvalerci di un modello statistico al fine di separare effetto di URP e differenze di partenza.

Procediamo per gradi, partendo da una regressione che stima solo la relazione tra punteggio e tasso di vandalismo osservato dopo l'attuazione di URP:

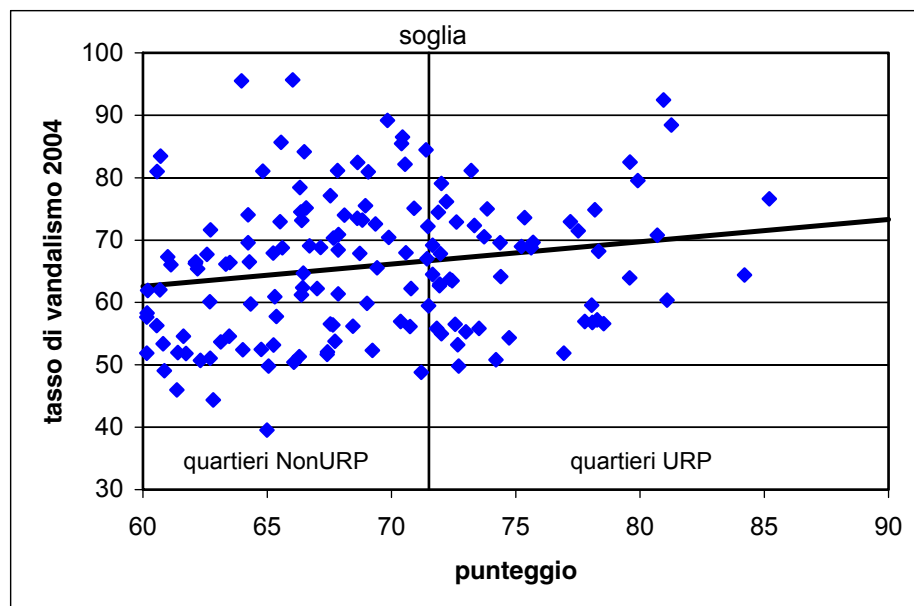
$$(1) \quad Y_i = \alpha + \varphi \text{ PUNTEGGIO}_i + \varepsilon_i$$

I risultati della stima sono riportati nella Tabella 6.3 e illustrati nel grafico seguente.

Tabella 6.3: Stime della regressione (1)

Variabile	Simbolo	Coefficiente	Errore Standard	t-statistico	<i>p-value</i>
Costante	α	40,95	11,53	3,55	0,001
Punteggio	φ	0,360	0,166	2,16	0,032

Figura 6.3: Rappresentazione grafica della regressione (1)



La pendenza positiva della retta di regressione (0,36 atti vandalici in più ogni punto addizionale in graduatoria) conferma la relazione *positiva* tra punteggio e tasso di vandalismo che avevamo intuito ispezionando visivamente la Figura 6.1. Questa

relazione positiva però non ci dice nulla sull'effetto di URP sul tasso di vandalismo, in quanto non fa altro che riflettere il modo con cui i quartieri sono stati selezionati dalla commissione. Infatti, con la regressione (1) non è possibile stimare l'effetto della politica. Il modello di regressione che ci serve è quello in cui siano incluse sia la variabile-trattamento sia il punteggio in graduatoria, di modo che l'effetto del trattamento venga stimato “a parità” di punteggio. Il modello seguente non è ancora il modello finale, ma ci serve come passaggio intermedio.

$$(2) \quad Y_i = \alpha + \beta T_i + \varphi \text{PUNTEGGIO}_i + \varepsilon_i$$

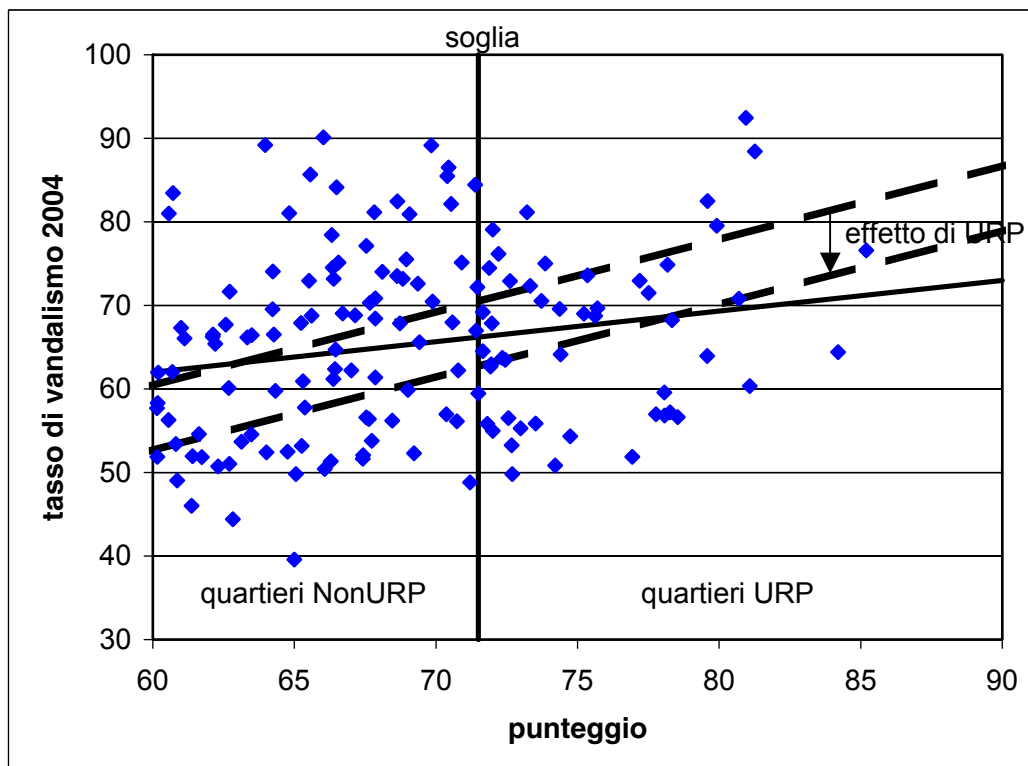
Questo è un tipo di modello che ci è familiare: una regressione lineare con due variabili esplicative, una continua e una dicotomica. Il modello produce le seguenti stime.

Tabella 6.4: Stime della regressione (2)

Variabile	Simbolo	Coefficiente	Errore Standard	t-statistico	<i>p-value</i>
Costante	α	8,16	18,28	0,45	0,656
T = trattamento	β	-7,80	3,41	-2,29	0,023
Punteggio	φ	0,874	0,278	3,14	0,002

I risultati sono illustrati graficamente nella Figura 6.4. La *pendenza* delle due rette parallele tratteggiate rappresenta la relazione tra punteggio e tasso di vandalismo, per i trattati e i non-trattati, rispettivamente. La *distanza verticale* tra le due rette tratteggiate è data dal coefficiente della variabile-trattamento, che è negativo: ciò indica un effetto nella direzione voluta, nel senso cioè di una riduzione del vandalismo. In questo caso la stima dell'effetto è una riduzione di quasi 8 atti vandalici. Notare che, con il modello (2), l'effetto stimato è *costante*: quale che sia il punteggio iniziale, questa specificazione del modello implica che URP sposti il tasso di vandalismo in basso di 8 punti percentuali.

Figura 6.4: Rappresentazione grafica della regressione (2)



Notiamo anche come la pendenza delle due rette tratteggiate sia più “ripida” di quella della retta corrispondente al modello (1), che è riprodotta nella Figura 6.4 con la linea continua. Il modello (2) stima 0,87 atti vandalici in più in media per ogni punto in più nella graduatoria, mentre il modello (1) solo 0,36 in più, segno del fatto che quest’ultima stima era distorta verso il basso a causa dell’omissione della variabile-trattamento. Questa omissione tendeva ad “appiattire” la retta di regressione. Il modello (2) ottiene due stime separate per l’effetto del punteggio sul vandalismo (che è positivo) e per quello dei finanziamenti URP (che ha segno negativo). Il modello (1) “mescolava” le due cose, producendo una pendenza più “piatta”.

Occorre sottolineare un punto importante: *non* sono necessarie altre variabili di controllo nel modello, il punteggio è l’unica variabile di controllo di cui abbiamo bisogno. Questo è il punto di forza di questo metodo, ma è anche il più difficile da comprendere appieno. Il processo di attribuzione dei punteggi può essere molto complesso, può dipendere da molti fattori, alcuni dei quali possono essere poco

“trasparenti”, cioè non osservabili dal valutatore. Ma la scelta finale di chi ammettere e chi non ammettere al finanziamento URP dipende dal fatto che il punteggio superi o meno una certa soglia. Se questo punteggio, che è l’unico criterio di selezione, è osservabile dal valutatore e può essere usato come variabile esplicativa, il gioco è (quasi) fatto: *a parità di punteggio*, le stime dell’effetto della politica non sono più distorte a causa delle differenze di partenza.

Il modello (2) assume che l’effetto di URP sia costante, non dipenda cioè dal livello del punteggio. Quest’ultima è una restrizione non necessaria, che può essere eliminata stimando il modello seguente:

$$(3) \quad Y_i = \alpha + \beta T_i + \varphi \text{PUNTEGGIO}_i + \delta T_i * \text{PUNTEGGIO}_i + \varepsilon_i$$

Il modello (3) aggiunge nella regressione un termine di *interazione* (cioè il prodotto) tra il punteggio e la variabile-trattamento. Le stime ottenute sono presentate nella tabella seguente.

Tabella 6.5: Stime della regressione (3)

Variabile	Simbolo	Coefficiente	Errore Standard	t-statistico	p-value
Costante	α	5,56	22,92	0,24	0,809
T = trattamento	β	0,122	41,93	0,00	0,998
Punteggio	φ	0,913	0,348	2,62	0,010
Interazione	δ	-0,110	0,581	-0,19	0,850

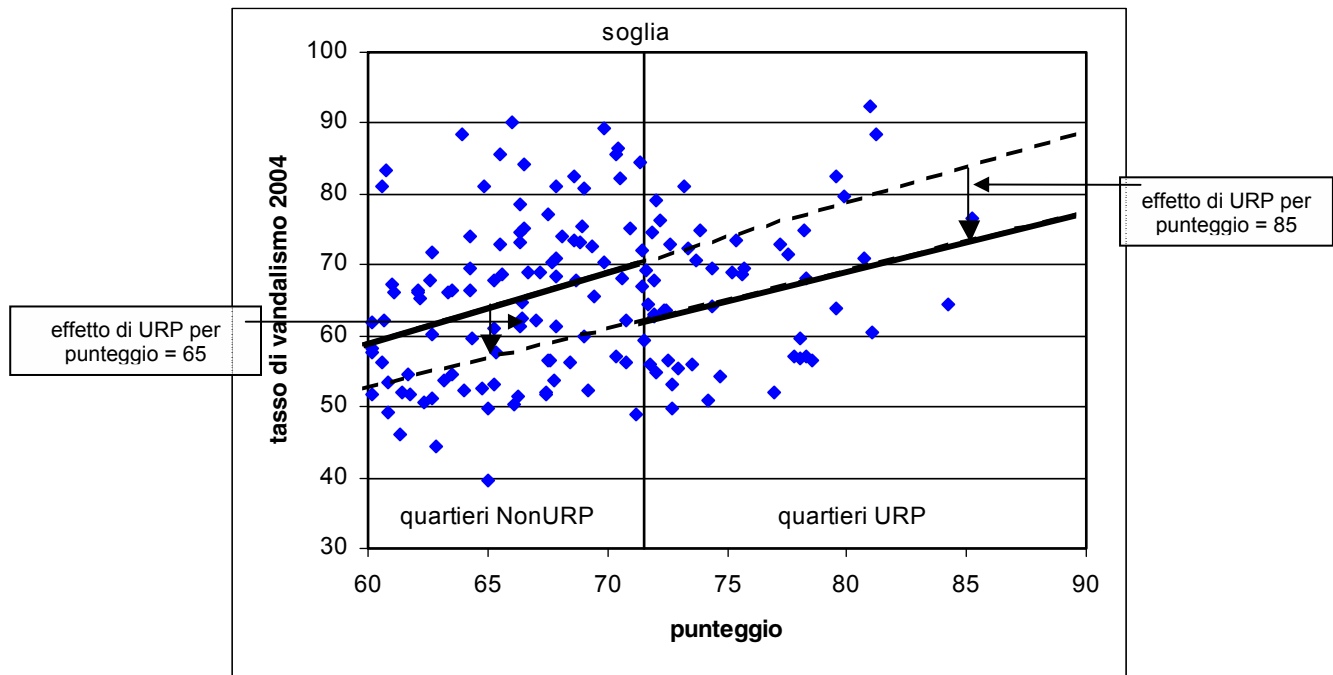
Queste stime sono più difficili da interpretare, ma ci aiutano a meglio comprendere come questa regressione sfrutti la “discontinuità” per derivare le stime dell’effetto della politica. Riscriviamo l’equazione (3) nel modo seguente:

$$(4) \quad Y_i = (\alpha + \beta T_i) + (\varphi + \delta T_i) * \text{PUNTEGGIO}_i + \varepsilon_i$$

Non abbiamo fatto altro che raggruppare diversamente i termini dell’equazione (3). Questo modo di scrivere l’equazione (4) mette in evidenza come *sia il valore dell’intercetta, sia la pendenza*, siano diverse a seconda che $T = 0$ o $T = 1$. (Quindi le stesse stime si otterrebbero stimando due regressioni separatamente, una per i quartieri

URP e una per i NonURP). Per $T = 0$ (quartieri NonURP) abbiamo una retta con intercetta uguale a α (5,56) e pendenza uguale a φ (0,913). È la retta *più in alto* nel grafico seguente.

Figura 6.5: Rappresentazione grafica delle regressioni (3) e (4)



La retta più in alto è rappresentata con una linea continua quando passa tra i quartieri NonURP e con una linea tratteggiata quando passa tra i quartieri URP. Questo per sottolineare che questa retta è stimata in realtà *solo* usando i dati sui quartieri NonURP: quando passa tra i quartieri URP, è una semplice proiezione lineare, non supportata da dati reali.

La retta più in basso nella Figura 6.4 rappresenta invece la relazione tra vandalismo e punteggio per i quartieri URP ($T = 1$). La sua intercetta è infatti uguale ad $\alpha + \beta$ ($5,56 + 0,12 = 5,68$) e la sua pendenza è uguale a $\varphi + \delta$ ($0,91 - 0,11 = 0,80$). Quindi la retta che vale per $T = 1$ ha una pendenza *più bassa* della retta che vale per $T = 0$.

Qual è adesso la stima dell'effetto di URP? Come nel grafico precedente, è la distanza verticale tra le due rette. Però adesso tale distanza verticale *non* è costante, ma varia al variare del punteggio. Questo è il primo caso in cui otteniamo una stima dell'effetto di

una politica che non è costante, cioè non è rappresentabile da un solo parametro. Con questo modello l'effetto di URP varia al variare del punteggio ed è calcolabile con la seguente formula:

$$(5) \quad \text{effetto di URP} = \beta + \delta * \text{PUNTEGGIO} = 0,12 - 0,11 * \text{PUNTEGGIO}$$

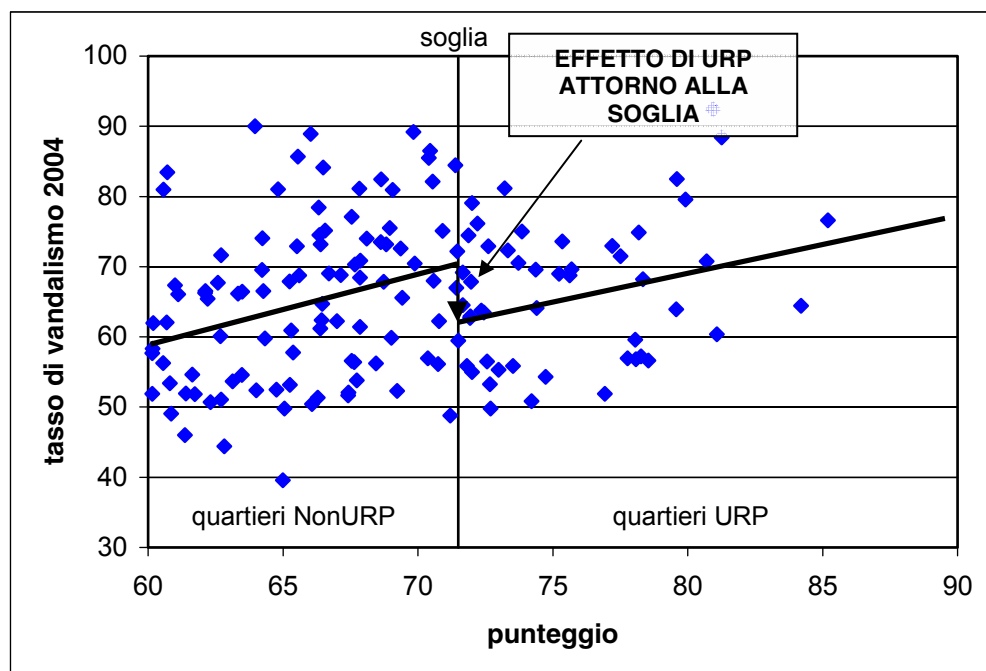
Dal grafico (e dalla formula) vediamo che l'effetto della politica è, all'interno dell'intervallo empiricamente rilevante (cioè per valori del punteggio superiori a 60), sempre negativo ed è *più forte per i punteggi più alti* in graduatoria. La freccia sulla destra del grafico rappresenta l'effetto per un quartiere che ha avuto un punteggio di 85 (effetto pari a $-9,23$ atti vandalici). La freccia sulla sinistra del grafico rappresenta l'effetto per un quartiere che ha avuto un punteggio di 65 (pari a $-7,03$ atti vandalici).

Questa è una buona notizia per URP: in termini di riduzione del tasso di vandalismo, i quartieri URP hanno guadagnato dall'essere inclusi più di quanto i quartieri NonURP abbiano perso dall'essere esclusi. Per i quartieri con punteggi molto bassi (cioè quartieri che hanno meno bisogno di riqualificazione urbana) il guadagno in termini di riduzione nel vandalismo sarebbe stato minore. Per i quartieri molto bisognosi, la mancata riduzione nel tasso di vandalismo (se la politica non fosse stata attuata) sarebbe stata invece più sostanziale.

Questa conclusione è peraltro (almeno in parte) una forzatura. Infatti, man mano che ci allontaniamo dalla soglia (in entrambe le direzioni), le stime dell'effetto diventano progressivamente meno credibili. Questo perché le porzioni tratteggiate delle due rette nel grafico precedente sono delle *proiezioni lineari*, non supportate da alcun dato. Nella retta più in alto, la proiezione rappresenta il vandalismo che si *sarebbe osservato* nei quartieri URP se la politica non fosse stata attuata *neppure* per loro. Nella retta più in basso, la proiezione rappresenta il vandalismo che si *sarebbe osservato* nei quartieri NonURP se la politica fosse stata attuata *anche* per loro. Ma entrambe queste proiezioni diventano *meno credibili* quanto più ci si allontana dai dati da cui sono estrapolate. Per questa ragione nella Figura 6.6 abbiamo eliminato queste proiezioni, tracciando le rette di regressione che avremmo ottenuto stimando il modello (2) separatamente per i quartieri URP e NonURP.

La stima più credibile dell'effetto è quella che si calcola proprio *attorno alla soglia*, cioè per un punteggio uguale a 71,5, poiché tale stima non dipende da proiezioni. Tale stima è data da $\beta + \delta \cdot 71,5 = 0,12 - 0,11 \cdot 71,5 = -7,75$. Quindi il modello (3) ci dice che l'effetto di URP (misurato attorno al punto di discontinuità) è una riduzione del tasso di vandalismo di circa 8 punti percentuali. Questa stima è più credibile perché non richiede proiezioni, ma si basa su “dati reali”. Nello stesso tempo, questa stima a rigore vale solo “localmente”: rappresenta in modo credibile l'effetto che la politica ha avuto sui quartieri “marginali”, quelli che avevano un punteggio tale da collocarsi vicino alla soglia di ammissione. Per i quartieri molto degradati, all'estrema destra nel grafico, l'incertezza sulla dimensione dell'effetto è maggiore.

Figura 6.6: Rappresentazione grafica della stima attorno alla soglia



Un'ultima considerazione: il *regression-discontinuity design* produce stime più credibili dei normali metodi non-sperimentali, però anche anch'esso soffre di una limitata applicabilità: non sono infatti frequenti le situazioni in cui l'esposizione o meno ad una politica è determinata da una semplice graduatoria o comunque da un criterio quantitativo direttamente osservabile dal valutatore. E questo è il requisito essenziale

per l'applicazione del *regression-discontinuity design*. Questo non toglie che la ricerca di discontinuità nell'attuazione di una politica e quindi di esperimenti naturali sia una strategia che vale sempre la pena di *esplorare*, prima di ricorrere a più sofisticati (e meno “robusti”) metodi non-sperimentali. Le discontinuità spesso non sono evidenti come nel caso estremo della graduatoria, ma tuttavia sono presenti.

6.3 I metodi che sfruttano le discontinuità temporali nel trattamento

Esaminiamo ora un insieme di metodi che sfruttano un tipo di “esperimento naturale” diverso da quello, appena visto, della graduatoria: mentre nel caso della graduatoria la discontinuità si creava tra individui, nel caso che presentiamo qui la discontinuità si crea *nel tempo* per la stessa popolazione di destinatari. Un esperimento naturale di questo tipo si verifica nelle situazioni in cui una politica pubblica *universale* entra in vigore *da una certa data in poi*: un nuovo obbligo o divieto, un inasprimento di una sanzione o la sua attenuazione, una riforma di un ordinamento o di una organizzazione, e più in generale l'entrata in vigore di un nuovo insieme di regole che hanno come obiettivo (o come possibile conseguenza) quello di modificare i comportamenti dei soggetti a cui sono imposte. Va notato che ciò che importa dal punto di vista della valutazione è il fatto che si crei una *discontinuità* al momento dell'entrata in vigore della politica, e che tale entrata in vigore non dipenda dalla libera scelta dei singoli individui a cui la regola si applica: in questo senso ci troviamo di fronte ad un esperimento naturale.

L'esempio che useremo è una politica pubblica reale (abbandoniamo così l'esempio fittizio di URP), come reali sono i dati utilizzati: l'*introduzione della patente a punti*, avvenuta in Italia nel luglio del 2003. Lo scopo di questa norma è modificare i comportamenti degli automobilisti in modo da indurli ad un maggiore rispetto del codice della strada, con il fine ultimo di produrre un miglioramento della sicurezza sulle strade e la diminuzione del numero e della gravità degli incidenti stradali.

Va sottolineato che in casi come questo non è possibile distinguere tra trattati e non-trattati: *tutti* gli automobilisti italiani sono immediatamente soggetti al regime di sanzioni connesso con la patente a punti. Se pochi modificano il loro comportamento alla guida, incorrendo nella prevista sottrazione di punti, ne dedurremo che il

trattamento introdotto dalla politica non è stato efficace: ma il trattamento si applica istantaneamente a tutti nel momento in cui la disposizione entra in vigore. Quindi per valutare gli effetti di una politica di questo tipo non esiste la possibilità di ricorrere ad alcun gruppo di controllo, neanche di tipo non-sperimentale. Possiamo però sfruttare la discontinuità rappresentata dalla “brusca” entrata in vigore della politica.

Ma dovremmo innanzitutto chiederci: effetto su cosa? Alla luce di quanto detto, non sarebbe sensato utilizzare come variabile-risultato né il numero di infrazioni che hanno dato luogo ad una perdita di punti, né il totale dei punti sottratti agli automobilisti, essendo queste variabili relative all’implementazione del “trattamento”. Le variabili-risultato che utilizzeremo sono invece il *numero di incidenti stradali* e il *numero di morti in seguito ad incidente stradale*. Su queste due variabili intendiamo misurare l’effetto dell’introduzione della patente a punti, che rappresenta la variabile-trattamento, uguale a zero nei mesi precedenti al luglio 2003 ed uguale ad uno da luglio 2003 in poi.

Il numero di incidenti (e il numero di quelli mortali) dipende da molti fattori, tra cui il numero di veicoli circolanti, l’intensità con cui i veicoli circolanti vengono utilizzati, l’età e le caratteristiche tecniche dei veicoli circolanti, lo stato di manutenzione delle strade, l’adeguatezza della segnaletica, le condizioni atmosferiche e, certamente non da ultimo, il comportamento degli automobilisti: determinato a sua volta da molti fattori, tra cui senza dubbio il tipo di sanzioni a cui sono soggetti in caso di violazione del codice della strada. La patente a punti intende incidere sull’ultimo di questa lunga lista di fattori, ma nulla può fare per influenzare gli altri elencati (e i molti altri che non abbiamo elencato). Entrambe le variabili-risultato sono quindi potenzialmente soggette ad una “dinamica spontanea”, dovuta ad un complesso di fattori estranei alla patente a punti.

L’idea basilare del metodo di valutazione illustrato in questa sezione (metodo che nelle sue applicazioni più sofisticate va sotto il nome di *analisi delle serie storiche interrotte*) è molto semplice: 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 stimare l’effetto come differenza tra l’andamento osservato (fattuale) e l’andamento predetto (controfattuale). In questo modo la situazione

controfattuale viene ricostruita usando la *proiezione di una serie storica*, proiezione realizzata mediante un'appropriata tecnica statistica. Un requisito per l'applicazione di questo metodo è quindi la disponibilità della “storia” della variabile-risultato per un periodo precedente all'introduzione della politica e sufficientemente *lungo* da poterne *modellare* l'andamento temporale. La complessità del metodo statistico con cui viene effettuata la proiezione dipende dalla complessità e lunghezza della serie storica: qui utilizzeremo serie storiche molto semplici e quindi metodi molto semplici, che non vanno al di là del modello di regressione lineare⁴⁶.

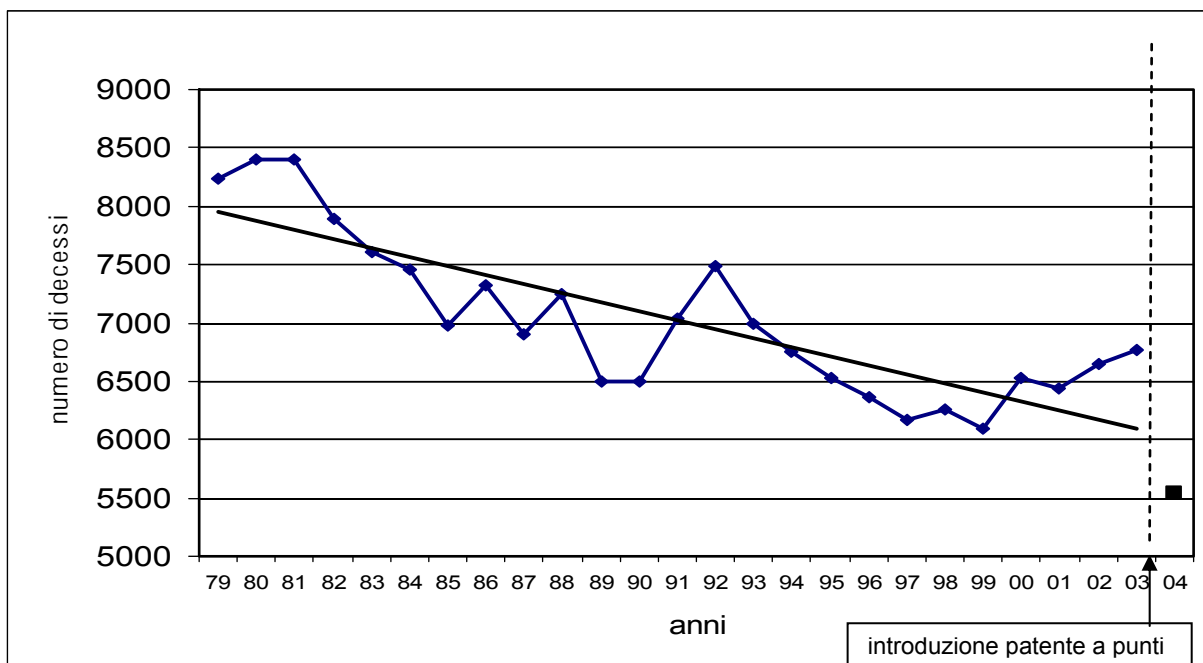
L'effetto della patente a punti sul numero di morti in seguito ad incidente stradale

Iniziamo considerando come variabile-risultato il numero di morti in seguito ad incidente stradale. È disponibile una serie storica (di fonte ISTAT) per questa variabile a partire dal 1979 al 2004: sono disponibili le osservazioni mensili, che abbiamo aggregato in osservazioni annuali, per semplificarne l'andamento⁴⁷. Il percorso temporale di questa variabile nei 25 anni precedenti all'entrata in vigore della patente a punti, illustrato nella Figura 6.7, mostra chiaramente un trend lineare *decrescente*: dagli oltre ottomila morti all'anno nel 1979 si è scesi negli ultimi anni tra i seimila e i settemila, una riduzione di circa un quarto. Evidentemente, nonostante il contemporaneo aumento del numero di veicoli circolanti e (come vedremo) del numero di incidenti, altri fattori (ad esempio l'uso di dispositivi di sicurezza quali cinture o *air bag*) hanno progressivamente ridotto la gravità degli incidenti in termini di conseguenze sulla vita delle persone coinvolte. Il dato relativo al 2004 è mostrato con un piccolo quadrato a destra della linea tratteggiata.

⁴⁶ Per approfondimenti si veda il testo classico McDowall D., McCleary R., Meidinger, R., Hay R., *Interrupted Time Series Analysis*, Sage Publications, Beverly Hill, 1980 o la parte dedicata a questo tema in Shadish W., Cook T., Campbell D., *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton-Mifflin, Boston, 2002.

⁴⁷ Per la precisione, al fine di allineare i dati con l'introduzione della patente a punti, avvenuta a metà 2003, abbiamo costruito il dato per il t-esimo anno prendendo gli ultimi sei mesi dell'anno t-1 e i primi sei mesi dell'anno t. Ad esempio, il numero di morti relativo al 2004 è in realtà la somma dei numeri mensili dal luglio del 2003 al giugno del 2004. Per questa analisi quindi restano inutilizzati i dati disponibili per gli ultimi sei mesi del 2004 e per i primi sei mesi del 1978.

Figura 6.7: Andamento dei decessi in seguito ad incidente stradale 1979-2004, con trend lineare



Fonte: Nostre elaborazioni su dati ISTAT

L'andamento temporale del fenomeno è rappresentato nel grafico dalla linea frastagliata a sinistra della data di entrata in vigore della patente a punti. Attorno al trend lineare decrescente ci sono variazioni "casuali", che rappresentano degli "*shock*" occasionali. Questo trend lineare può essere stimato mediante la seguente regressione sulle 25 osservazioni annuali pre-patente a punti:

$$(6) \quad Y_t = \alpha + \gamma \cdot t + \varepsilon_t \quad t = 1, \dots, 25$$

dove Y_t è il numero di morti nell'anno t -esimo, mentre la variabile esplicativa t è espressa come un semplice numero progressivo indicante l'anno, cioè 1,2,3,4,..., fino a 25. Si ottengono così le stime dei coefficienti mostrate nella Tabella 6.6, che a loro volta consentono di tracciare la linea interpolante mostrata nella Figura 6.7.

Tabella 6.6: Stime della regressione (6)

Variabile	Simbolo	Coefficiente	Errore Standard	t-statistico	p-value
Costante	α	8020,7	161,4	49,6	0,000
t = tempo	γ	- 76,89	10,86	7,08	0,000
		N = 25	$R^2 = 0,68$		

Fonte: Nostre elaborazioni su dati ISTAT

Qual è dunque l'effetto dell'introduzione della patente a punti? Nel grafico si nota chiaramente come il numero di decessi relativi al 2004 (post-patente a punti) sia situato *sotto* il trend storico, qualora questo venga proiettato in avanti di un anno. Seguendo questa logica, il valore controfattuale (il numero di decessi che ci saremmo aspettati nel 2004 se il trend fosse proseguito come nei 25 anni precedenti) si può calcolare come:

$$\hat{Y}_{2004} = 8021 - 76,89 * t = 8021 - 76,89 * 26 = 6022$$

Ci saremmo aspettati, sulla base del trend storico, circa 6000 morti per incidente stradale nel 2004; il valore osservato nel 2004 è di 5538 decessi; quindi l'effetto stimato sarà:

$$\text{Effetto della patente a punti nel 2004} = Y_{2004} - \hat{Y}_{2004} = 5538 - 6022 = -484$$

cioè la patente a punti avrebbe *causato una riduzione di circa 500 morti nel primo anno di vigenza* (in termini percentuali, una riduzione dell'8%). Un modo equivalente di ottenere la stessa stima, con l'aggiunta del relativo errore standard, è quello di aggiungere alla regressione la variabile-trattamento e di stimare la regressione su 26 osservazioni, includendo cioè il 2004.

$$(7) \quad Y_t = \alpha + \beta * (\text{Patente a punti})_t + \gamma * t + \varepsilon_t \quad t = 1, \dots, 26$$

dove $(\text{Patente a punti})_t$ rappresenta la variabile-trattamento, uguale a zero negli anni precedenti al 2004 ed uguale ad 1 per il solo 2004. I coefficienti stimati sono riportati nella Tabella 6.7.

Tabella 6.7: Stime della regressione (7)

Variabile	Simbolo	Coefficiente	Errore Standard	t-statistico	p-value
Costante	α	8020,7	161,4	49,6	0,000
Patente a punti	β	- 483,6	423,5	1,14	0,256
t = tempo	γ	- 76,89	10,86	7,08	0,000
N =26			$R^2 = 0,73$		

Fonte: Nostre elaborazioni su dati ISTAT

Il coefficiente β (stimato con una sola osservazione aggiuntiva) rappresenta la deviazione dal trend 1979-2003 del dato relativo al 2004. Il t-statistico implica però che tale deviazione non è significativamente diversa da zero.

Abbiamo ottenuto una stima corretta dell'effetto della patente a punti sul numero di decessi sulle strade? La correttezza di questa stima dipende da quanto correttamente la proiezione al 2004 del trend precedente approssima la situazione controfattuale. È possibile immaginare altri valori controfattuali, basati su argomentazioni diverse. Ad esempio, si può sostenere che la migliore predizione del numero di decessi che sarebbero avvenuti nel 2004 è l'ultimo valore osservato prima dell'entrata in vigore della patente a punti. Questa idea ripropone la semplice analisi pre-post vista nel capitolo 4, basata sull'assunto di assenza di dinamica spontanea. In questo caso l'effetto stimato sarebbe di $6767 - 5538 = -1229$ decessi (meno 18%). Anziché l'ultimo valore, si può usare come controfattuale la media degli ultimi anni, ad esempio degli ultimi quattro. In questo caso l'effetto della patente a punti sarebbe di $6595 - 5538 = -1057$ decessi (meno 16%). Oppure si può argomentare che, poiché negli ultimi cinque anni precedenti all'introduzione della patente a punti (a partire dal 1999) il trend è stato sempre crescente, quello è il trend da proiettare in avanti per derivare il controfattuale. Compiuta questa operazione, si ottiene un effetto della patente a punti pari a $6933 - 5538 = -1395$ decessi (meno 20%).

Un risultato intermedio tra quelli appena illustrati lo si ottiene permettendo al trend di avere un andamento non lineare (cioè non rappresentabile da una linea retta). Si può stimare la seguente retta di regressione:

$$(8) \quad Y_t = \alpha + \beta * (\text{Patente a punti})_t + \gamma * t + \delta * t^2 + \varepsilon_t \quad t = 1, \dots, 26$$

dove t^2 rappresenta il tempo al quadrato. I coefficienti stimati sono riportati nella Tabella 6.8.

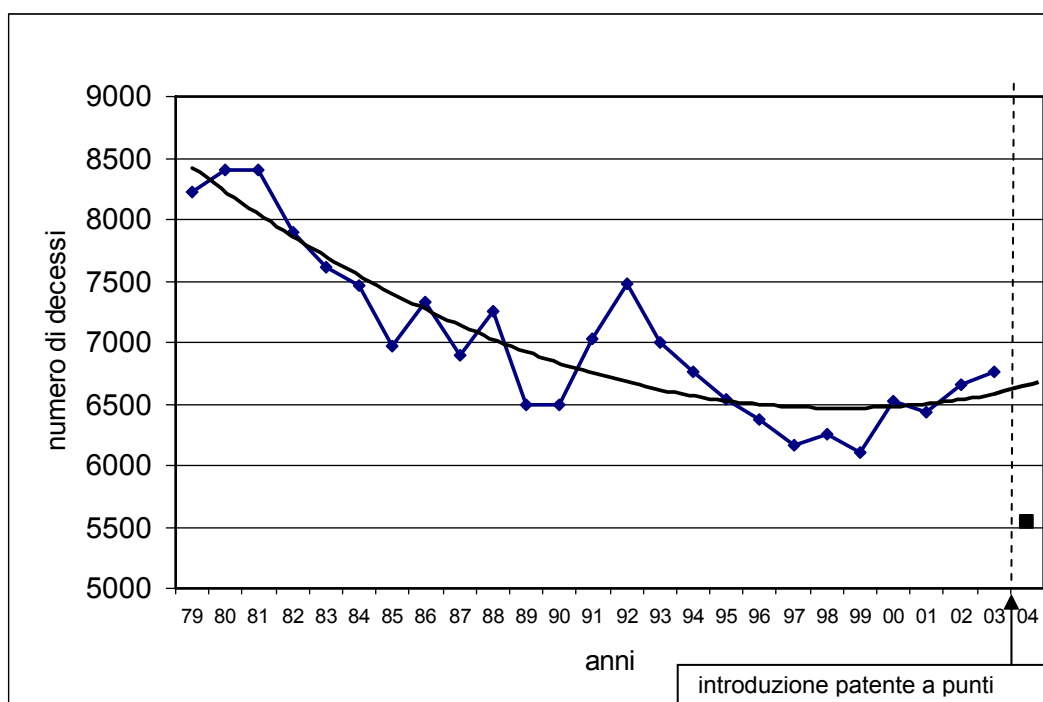
Tabella 6.8: Stime della regressione (8)

Variabile	Simbolo	Coefficiente	Errore Standard	t-statistico	<i>p-value</i>
Costante	α	8631,0	199,4	43,27	0,000
Patente a punti	β	-1093,8	365,4	-2,99	0,007
t = tempo	γ	-212,5	35,3	-6,01	0,000
t^2 = tempo al quadrato	δ	5,216	1,32	3,95	0,001
N =26			$R^2 = 0,84$		

Fonte: Nostre elaborazioni su dati ISTAT

I valori stimati dei coefficienti implicano un andamento temporale a forma di U: come illustrato nella Figura 6.8, per i primi anni il trend è decrescente e poi lentamente cambia direzione, diventando crescente a partire dal 2001. La proiezione di questo trend al 2004 implica un numero di decessi atteso pari a 6632. Se questo fosse il controfattuale, l'effetto della patente a punti sarebbe di $6632 - 5538 = -1094$ decessi (meno 16,5%). È da notare come la stima dell'effetto possa essere letta direttamente nella Tabella 6.8, prendendo il valore stimato del coefficiente β , il cui t-statistico ora ci dice che la stima è significativamente diversa da zero. Rispetto al modello lineare, l'R-quadro è migliorato, indicando un miglior adattamento ai dati.

Figura 6.8: Andamento dei decessi in seguito ad incidente stradale 1979-2004, con trend non lineare



Fonte: Nostre elaborazioni su dati ISTAT

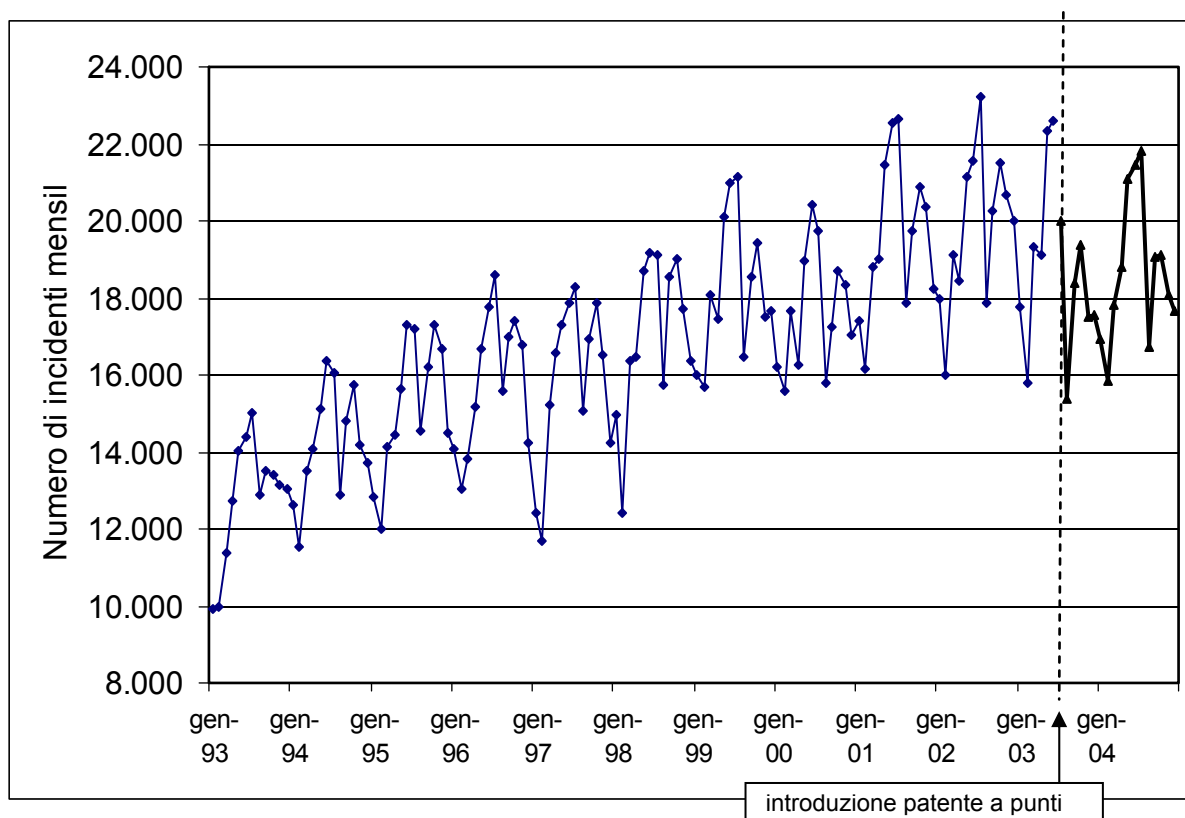
Non sorprendentemente, nozioni diverse di controfattuale producono stime diverse dell'effetto della politica. La più "conservativa" è la stima ottenuta assumendo che il trend lineare negativo sarebbe proseguito come nei 25 anni precedenti. Una stima forse più plausibile è quella che riconosce quella che pare essere un'inversione di tendenza nel trend avvenuta negli ultimi anni. Ma questa non è una garanzia di verità, è solo un'argomentazione migliore. Quello che possiamo concludere è che il numero di morti sulle strade evitati grazie alla patente a punti nei primi dodici mesi di applicazione si colloca probabilmente tra i 500 e i 1.400, con una riduzione in percentuale che sta tra l'8 e il 20%, con un valore più plausibile che si colloca attorno ai 1.100 decessi evitati. Ovviamente, queste conclusioni possono cambiare una volta che si aggiungano altre osservazioni relative al periodo post-riforma: sia perché nuovi dati rendono meno soggetta a shock temporanei la stima del valore fattuale, sia perché nel frattempo l'effetto vero della patente a punti potrebbe essere cambiato, attenuandosi o aumentando nel tempo.

L'effetto della patente a punti sul numero di incidenti stradali

Prendiamo ora in considerazione la seconda variabile-risultato, il numero di incidenti, i cui dati sono disponibili (da fonte ISTAT) in maniera omogenea solo a partire dai primi anni Novanta, a causa di un cambiamento delle modalità di registrazione avvenute all'inizio del decennio. Per questa ragione, e per cambiare tipo di modellazione statistica, per questa variabile-risultato utilizzeremo i dati mensili per il periodo gennaio 1993-dicembre 2004: quindi 126 mesi precedenti l'introduzione della patente a punti e 18 mesi successivi, per un totale di 144 osservazioni mensili.

Come si nota facilmente dalla Figura 6.9, l'andamento mensile degli incidenti stradali è affetto da una forte stagionalità, con valori più elevati tra giugno e luglio e valori più bassi tra gennaio e febbraio. A parte la stagionalità, si nota anche un trend generale *crescente*, almeno fino all'entrata in vigore della patente a punti.

Figura 6.9: Andamento mensile degli incidenti stradali, gennaio 1993 – dicembre 2004



Fonte: Nostre elaborazioni su dati ISTAT

Come nel caso dei decessi sulle strade, anche in questo caso sono possibili diverse ricostruzioni di ciò che sarebbe successo al numero di incidenti dopo il luglio 2003 in assenza di patente a punti. L'approccio più semplice è nuovamente quello di un mero confronto pre-post, avendo l'accortezza di confrontare il dato mensile post-intervento con il dato dello stesso mese dell'anno precedente: questo è un modo per tenere conto della stagionalità (ma non del trend crescente). Utilizzando per tale confronto i mesi di agosto 2003 (fattuale) e agosto 2002 (controfattuale, valido nell'ipotesi di assenza di dinamica spontanea), l'effetto della patente a punti sul dato mensile sarebbe uguale a $15.389 - 17.872 = -2.483$, cioè una riduzione del 14%.

Un confronto leggermente più credibile, ma sempre in un'ottica pre-post, consisterebbe nel confrontare il numero di incidenti nei 12 mesi precedenti e nei 12 mesi successivi all'entrata in vigore della patente a punti: in questo caso la stima dell'effetto sarebbe $220.240 - 240.466 = -20.226$, una riduzione dell' 8%.

Proviamo ora a ricorrere ad una modellazione ispirata alla logica dell'analisi delle serie storiche interrotte, utilizzando però sempre un modello di regressione lineare. Abbiamo bisogno di un modello che colga sia la componente stagionale sia il trend storico, che stimeremo con i soli 126 dati mensili relativi al periodo precedente all'introduzione della patente a punti:

$$(9) \quad Y_t = \alpha + \gamma * t + \sum_{m=1}^{11} \delta_m M_m + \varepsilon_t \quad t = 1, \dots, 126$$

dove le variabili M_m sono 11 variabili indicanti il mese dell'anno. Si tratta di 11 variabili dicotomiche, uguali a 1 se il mese è il mese m-esimo, uguali a zero altrimenti. Ciascuno dei coefficienti δ_m rappresenta la deviazione del numero atteso di incidenti per il mese m-esimo rispetto al dodicesimo mese escluso dalla lista, che arbitrariamente è stato individuato nel mese di dicembre. I coefficienti stimati del modello sono contenuti nella Tabella 6.9.

Tabella 6.9: Stime della regressione (9)

Variabile	Simbolo	Coefficiente	Errore Standard	t-statistico	p-value
Costante	α	11.938,2	269,3	44,32	0,000
t = tempo	γ	59,99	1,859	32,27	0,000
Gennaio	δ_1	-839,64	331,4	-2,53	0,013
Febbraio	δ_2	-2038,6	331,3	-6,15	0,000
Marzo	δ_3	417,01	331,3	1,26	0,211
Aprile	δ_4	563,47	331,3	1,70	0,092
Maggio	δ_5	2475,6	331,3	7,47	0,000
Giugno	δ_6	3278,0	331,3	9,89	0,000
Luglio	δ_7	3508,4	339,2	10,34	0,000
Agosto	δ_8	-189,8	339,1	-0,56	0,577
Settembre	δ_9	1562,7	339,1	4,61	0,000
Ottobre	δ_{10}	2350,1	339,1	6,93	0,000
Novembre	δ_{11}	1349,0	339,1	3,98	0,000
Dicembre	--	<i>Categoria omessa (mese di riferimento)</i>			
N =126			R ² = 0,936		

Fonte: Nostre elaborazioni su dati ISTAT

Dalla tabella emerge come il trend di crescita sia positivo e pari a circa 60 incidenti in più al mese (cioè circa 700 in più all'anno). Le differenze tra i vari mesi dell'anno sono molto forti: prendendo dicembre come mese di riferimento, vediamo come in giugno e luglio avvengano mediamente più di 3.000 incidenti in più al mese (circa il 25% in più), e a febbraio (che è anche il mese più corto) circa il 16 percento in meno.

Come si possono utilizzare queste grandezze per stimare l'effetto della patente a punti sul numero di incidenti? In modo analogo a quanto fatto per l'effetto sul numero di decessi, con l'importante differenza che ora abbiamo 18 osservazioni (mensili) post-intervento, anziché solo un'osservazione (annuale). Il controfattuale va quindi calcolato per ciascuno dei 18 mesi post-introduzione della patente a punti, sulla base della formula:

$$\hat{Y}_t = 11938 + 59,99 * t + \delta_m M_t \quad t = 127, \dots, 144$$

La Tabella 6.10 riporta i 18 valori osservati, i 18 valori predetti sulla base della formula precedente e le 18 stime dell'effetto della patente a punti.

Tabella 6.10: Stime dell'effetto sul numero di incidenti stradali nei 18 mesi successivi all'introduzione della patente a punti

Mese	Valore osservato	Valore predetto	Effetto stimato	Effetto in %
Luglio 2003	20.023	23.065,3	- 3.042,3	-13,2%
Agosto	15.389	19.427,0	- 4.038,0	-20,8%
Settembre	18.401	21.239,6	- 2.838,6	-13,4%
Ottobre	19.353	22.086,9	- 2.733,9	-12,4%
Novembre	17.483	21.145,8	- 3.662,8	-17,3%
Dicembre 2003	17.558	19.856,8	- 2.298,8	-11,6%
Gennaio 2004	16.935	19.077,2	- 2.142,2	-11,2%
Febbraio	15.869	17.938,2	- 2.069,2	-11,5%
Marzo	17.844	20.453,8	- 2.609,8	-12,8%
Aprile	18.825	20.660,3	- 1.835,3	-8,9%
Maggio	21.089	22.632,4	- 1.543,4	-6,8%
Giugno	21.471	23.494,8	- 2.023,8	-8,6%
Luglio 2004	21.834	23.785,2	- 1.951,2	-8,2%
Agosto	16.738	20.146,9	- 3.408,9	-16,9%
Settembre	19.061	21.959,5	- 2.898,5	-13,2%
Ottobre	19.125	22.806,8	- 3.681,8	-16,1%
Novembre	18.088	21.865,7	- 3.777,7	-17,3%
Dicembre 2004	17.674	20.576,7	- 2.902,7	-14,1%
Totale per i 18 mesi	332.760	382.218,7	- 49.458,7	-12,9%

Fonte: Nostre elaborazioni su dati ISTAT

Nei 18 mesi successivi alla sua introduzione, la patente a punti avrebbe ridotto il numero totale degli incidenti di quasi 50.000 unità, circa il 13 % rispetto a quanto ci si poteva aspettare sulla base dell'andamento stagionale e del trend storico di crescita. Tale effetto sembra ridursi nei primi mesi del 2004, ma poi torna nella seconda metà dell'anno vicino ai valori stimati nella seconda metà del 2003, segno che a distanza di più di un anno dalla sua introduzione la patente a punti continua ad esercitare un effetto deterrente.

Concludiamo stimando un modello di regressione che ottiene direttamente gli stessi risultati presentati nella Tabella 6.10. Questo modello va stimato su tutte le 144 osservazioni mensili che abbiamo a disposizione.

$$(10) \quad Y_t = \alpha + \sum_{k=1}^{18} \beta_k P_k + \gamma * t + \sum_{m=1}^{11} \delta_m M_m + \varepsilon_t \quad t = 1, 144$$

dove i P_k sono 18 variabili dicotomiche che rappresentano ciascuna uno dei mesi post-intervento. Le stime sono presentate nella tabella seguente (sono omesse le stime ottenute per gli effetti stagionali δ).

Tabella 6.11: Stime della regressione (10)

Variabile	Simbolo	Coefficiente	Errore Standard	t-statistico	p-value
Costante	α	11.938,2	269,3	44,32	0,000
t = tempo	γ	59,99	1,859	32,27	0,000
Luglio 2003	β_1	-3042,3	804,6	-3,78	0,000
Agosto	β_2	-4038,0	804,6	-5,02	0,000
Settembre	β_3	-2838,6	804,6	-3,53	0,001
Ottobre	β_4	-2733,9	804,6	-3,40	0,001
Novembre	β_5	-3662,8	804,6	-4,55	0,000
Dicembre 2003	β_6	-2298,8	804,6	-2,86	0,005
Gennaio 2004	β_7	-2142,1	803,2	-2,67	0,009
Febbraio	β_8	-2069,1	803,2	-2,58	0,011
Marzo	β_9	-2609,7	803,2	-3,25	0,002
Aprile	β_{10}	-1835,2	803,2	-2,28	0,024
Maggio	β_{11}	-1543,4	803,2	-1,92	0,057
Giugno	β_{12}	-2023,7	803,2	-2,52	0,013
Luglio 2004	β_{13}	-1951,1	808,3	-2,41	0,017
Agosto	β_{14}	-3408,8	808,3	-4,22	0,000
Settembre	β_{15}	-2898,4	808,3	-3,59	0,000
Ottobre	β_{16}	-3681,7	808,3	-4,55	0,000
Novembre	β_{17}	-3777,6	808,3	-4,67	0,000
Dicembre 2004	β_{18}	-2902,6	808,3	-3,59	0,000
<i>Effetti stagionali</i>	δ	<i>Omessi in quanto simili a quelli presentati nella Tabella 6.9</i>			
N =144			$R^2 = 0,942$		

Fonte: Nostre elaborazioni su dati ISTAT

I 18 coefficienti β (stimati con una sola osservazione ciascuno) rappresentano la deviazione dal trend 1993-2003 del numero di incidenti stradali relativo a ciascun mese post-intervento. Tutte e 18 le deviazioni risultano diverse da zero ai consueti livelli di significatività. Questo risultato suggerisce come la patente abbia effettivamente contribuito a “spostare” verso il basso l’andamento degli incidenti stradali in Italia.

Questo interessante risultato è stato ottenuto mediante l’utilizzo di un metodo statistico relativamente semplice. Questa semplicità ci è stata concessa dal fatto di avere a che

fare con serie storiche dalla struttura temporale relativamente semplice: appare abbastanza evidente che le principali *determinanti* delle serie storiche utilizzate siano solo un trend pressoché lineare e una componente stagionale.

Per serie storiche più complesse, di cui non sia chiaro a priori quali sono le componenti, occorre ricorrere ai più sofisticati modelli ARIMA (*Auto Regressive Integrated Moving Average*), che hanno lo scopo specifico di “far parlare i dati”, rivelando empiricamente le varie componenti della serie storica. Una volta stimati, i coefficienti di questi modelli possono essere utilizzati per proiettare in avanti la serie storica, come abbiamo fatto con il modello di regressione. Alcune versioni dei modelli ARIMA consentono di incorporare una variabile che rappresenta il periodo in cui è in vigore l'intervento, di cui si vuole stimare l'effetto sulla serie storica.

Appendice

Dettagli sul *matching* statistico

Questa appendice rappresenta alcuni dettagli sulla stima dell'effetto ottenuta mediante *matching* statistico, omessi dal testo principale per facilitarne la lettura.

ABBINAMENTO SULL'UNITÀ PIÙ VICINA (NEAREST NEIGHBOUR MATCHING)

La stima dell'effetto della politica è ricavata dalla media delle differenze tra le coppie di unità abbinate (che è equivalente alla differenza tra la media dei trattati e la media dei controlli abbinate):

$$\text{effetto stimato} = \sum_i (Y_i^T - Y_{(i)}^C) / N^T$$

dove:

Y_i^T rappresenta il valore della variabile-risultato dell'*i*-esima unità trattata;

$Y_{(i)}^C$ rappresenta il valore della variabile-risultato dell'unità di controllo (con *propensity score* più vicino) abbinata all'*i*-esima unità trattata;

N^T rappresenta il numero di unità trattate presenti nel campione.

ABBINAMENTO IN UN RAGGIO (RADIUS MATCHING)

Una volta effettuato l'abbinamento, la stima dell'effetto è derivata in modo simile al metodo precedente:

$$\text{effetto stimato} = \sum_i (Y_i^T - \bar{Y}_{(i)}^C) / N^{TR}$$

dove:

Y_i^T rappresenta il valore della variabile-risultato dell'*i*-esima unità trattata;

$\bar{Y}_{(i)}^C$ rappresenta il valore *medio* della variabile-risultato per le unità di controllo abbinate all'*i*-esima unità trattata;

N^{TR} rappresenta il numero di unità trattate che hanno trovato almeno un abbinamento all'interno del raggio ($N^{TR} \leq N^T$).

STRATIFICATION MATCHING (“ABBINAMENTO CON STRATIFICAZIONE”)

Il calcolo dell'effetto della politica richiede due passi successivi. Prima si calcola, all'interno di ogni strato, la differenza tra le medie della variabile-risultato calcolate separatamente tra le unità trattate e tra le unità di controllo, ottenendo quindi un effetto limitato al particolare strato k:

$$\Delta_k = \sum_i Y_i^T / N_k^T - \sum_i Y_i^C / N_k^C$$

dove N_k^T e N_k^C sono, rispettivamente, il numero di unità trattate e non-trattate presenti nello strato k. L'effetto complessivo è calcolato come media delle differenze nei vari strati, ponderate con il numero di unità trattate presenti (si dà quindi maggiore peso agli strati in cui si concentrano maggiormente le unità trattate).

$$\text{effetto stimato} = \sum_k \Delta_k N_k^T / N^T$$

KERNEL MATCHING (“ABBINAMENTO CON LA FUNZIONE KERNEL”)

La stima dell'effetto è data dalla seguente formula:

$$\text{effetto stimato con “kernel”} = \sum_i (Y_i^T - \sum_{j=1}^{N^C} w_{ij} Y_j^C) / N^T$$

dove:

Y_i^T rappresenta il valore della variabile-risultato dell'i-esima unità trattata;

Y_j^C rappresenta il valore della variabile-risultato della j-esima unità non-trattata;

N^T rappresenta il numero di unità trattate;

N^C rappresenta il numero di unità non-trattate (controlli);

$\sum_{j=1}^{N^C} w_{ij} Y_j^C$ rappresenta la media ponderata della variabile-risultato di tutte le unità non-trattate, da calcolarsi per ciascuna i-esima unità trattata.

I pesi w_{ij} sono costruiti in modo che i loro valori decrescano rapidamente al crescere della distanza tra il *propensity score* di ciascuna unità non-trattata da quello dell'unità trattata per la quale si sta costruendo il valore controfattuale. Il massimo valore del peso si raggiunge quando $p_i = p_j$, cioè quando la j -esima unità non-trattata ha lo stesso *propensity score* dell'unità trattata i -esima. All'aumentare della distanza $|p_i - p_j|$ il valore del peso decresce esponenzialmente (sulla base di una funzione matematica detta *kernel*) fino praticamente ad annullarsi per le unità con *propensity score* molto diverso.

SECONDA PARTE

**APPLICAZIONI
AL CASO ITALIANO**

La scelta dei casi

In questa sezione proponiamo sei differenti casi di valutazione di effetti di politiche pubbliche. La selezione di questi casi è stata guidata principalmente da tre criteri: (1) lo studio di valutazione doveva avere come punto di partenza una *domanda di natura controfattuale*, per rispondere alla quale si richiedeva l'applicazione di uno dei metodi quantitativi presentati nelle pagine precedenti; (2) gli studi, tutti realizzati e pubblicati in Italia, dovevano riguardare politiche pubbliche afferenti a *diversi settori d'intervento* e a *differenti livelli di governo*; (3) ogni studio doveva partire da una *precisa e ben individuata committenza pubblica*.

Come si desume dal seguente quadro sinottico, che riporta in sintesi i tratti caratteristici di ogni esperienza di valutazione presentata, tale selezione non ha dato i frutti sperati in termini di diversificazione delle politiche considerate (la maggior parte dei casi riguarda politiche di aiuto alle imprese e politiche del lavoro) e dei livelli di governo (quattro su sei sono state commissionate da enti statali). Ciò dipende dal fatto che la valutazione degli effetti è un approccio relativamente ancora giovane in Italia, ancora poco diffuso, se non in una comunità abbastanza ristretta di economisti e statistici impegnati nella realizzazione di questo tipo di analisi. Tutti gli studi sono stati infatti realizzati negli ultimi cinque anni ed è raro trovare in Italia una valutazione degli effetti risalente ad un periodo precedente. Questa osservazione si presta però ad una lettura ottimistica: se il recente fiorire di valutazioni degli effetti rappresenta un tendenziale segno di interesse, è probabile allora che nei prossimi anni questo tipo di studi sia sempre più conosciuto ed apprezzato, e possa dunque trovare uno sviluppo ed un impiego maggiore anche presso le amministrazioni di diversa natura.

Un'ampia varietà è invece assicurata sulle questioni e gli aspetti di natura metodologica. Ogni caso vede l'applicazione di uno o più metodi non-sperimentali. Le tecniche e gli strumenti statistici adottati per affrontare tali questioni variano di caso in caso, adattandosi alla particolarità delle domande di valutazione e alla disponibilità di informazioni. In alcuni casi, per realizzare la valutazione, gli autori hanno impiegato dati derivati da archivi amministrativi preesistenti; in altri, sono state realizzate indagini

ad hoc che hanno consentito di utilizzare informazioni generalmente più ricche ed attendibili.

Sebbene vi sia un certo assortimento anche rispetto alle variabili-risultato considerate (dalle decisioni di investimento delle imprese finanziate alla probabilità di sopravvivenza delle imprese di nuova creazione; dalle durate di disoccupazione per lavoratori in mobilità ai tassi di abbandono delle matricole universitarie), prevale nei sei casi l'interesse per gli aspetti legati all'occupazione.

I sei casi a confronto

	1. Gli effetti dei finanziamenti della legge 488 sulla performance delle imprese e sullo sviluppo locale	2. Gli effetti dei finanziamenti della legge regionale sarda 15/94 sulla performance delle imprese	3. Gli effetti della politica della Regione Toscana a sostegno dell'imprenditoria giovanile	4. Gli effetti del lavoro interinale sulla probabilità di trovare un impiego stabile	5. Gli effetti dell'indennità di mobilità sulla probabilità di rioccupazione dei lavoratori licenziati	6. Gli effetti delle borse di studio sulla regolarità degli studi universitari
Settore d'intervento	Politica di aiuto alle imprese	Politica di aiuto alle imprese	Politica di aiuto alle imprese	Politica del lavoro	Politica del lavoro	Politica per il diritto allo studio
Anno dello studio	Primo studio: 2000 Secondo studio: 2003	2005	Primo studio: 2000 Secondo studio: 2001 Terzo studio: 2005	2003	2004	2004
Principale committente	Ministero delle Attività Produttive	Regione Sardegna	Consiglio regionale della Toscana	Ministero del Lavoro e Politiche Sociali	Ministero del Lavoro e Politiche Sociali	Comitato nazionale per la Valutazione del Sistema Universitario
Oggetto di valutazione	I. Gli effetti incentivanti sugli investimenti II. Gli effetti sulle performance aziendali III. Gli effetti sull'occupazione territoriale	I. Gli effetti sul numero di occupati II. Gli effetti sul fatturato	I. Gli effetti sulla probabilità di sopravvivenza delle imprese II. Gli effetti sulla decisione di mettersi in proprio	I. Gli effetti sulla probabilità di trovare un lavoro a tempo indeterminato	I. Gli effetti del regime di mobilità sulla durata della disoccupazione II. Gli effetti sul numero di esami sostenuti	I. Gli effetti sui tassi di abbandono II. Gli effetti sul numero di esami sostenuti
Metodi impiegati	1. Domande dirette ai beneficiari 2. Regressione 3. Differenza-nelle-differenze	1. Differenza-nelle-differenze 2. <i>Regression-Discontinuity Design</i>	1. Regressione 2. <i>Matching</i> statistico 3. Domande dirette ai beneficiari	1. <i>Matching</i> statistico	1. <i>Matching</i> statistico 2. <i>Regression-Discontinuity Design</i>	1. <i>Regression - Discontinuity Design</i> 2. Regressione 3. <i>Matching</i> statistico

CAPITOLO 7

GLI EFFETTI DEI FINANZIAMENTI DELLA LEGGE 488 SULLA PERFORMANCE DELLE IMPRESE E SULLO SVILUPPO LOCALE

di Alberto Martini

La legge 488/92 nasce dopo la conclusione dell'intervento straordinario nel Mezzogiorno, con lo scopo di sostenere lo sviluppo economico nelle aree depresse del Paese tramite la concessione ed erogazione di agevolazioni in favore delle attività produttive. La rilevanza di questa politica, in termini di entità di importi erogati e di investimenti effettuati, ha motivato lo stesso legislatore nazionale a interrogarsi sulla sua efficacia rispetto agli obiettivi che essa si poneva. Il Regolamento attuativo della legge 488 contiene infatti un raro esempio di esplicita *clausola valutativa*, sotto forma di un esplicito mandato assegnato allo stesso Ministero dell'Industria a svolgere attività di “*valutazione dell'efficacia degli interventi rispetto allo sviluppo economico delle aree interessate*” (art. 1-bis, comma 1 punto a).

A questo mandato parlamentare, il Ministero dell'Industria, poi diventato Ministero delle Attività Produttive, ha risposto svolgendo una serie di studi di valutazione sugli effetti dei finanziamenti erogati. L'attenzione del valutatore si è rivolta sia agli effetti sulla performance delle aziende beneficiarie delle agevolazioni sia a quelli sullo sviluppo economico delle aree geografiche in cui le aziende beneficiarie operano.

In questo capitolo riassumeremo la metodologia e i risultati raggiunti da una pluralità di studi valutativi. Il primo in ordine temporale è quello condotto dallo stesso Ministero dell'Industria con la collaborazione dell'Istituto Tagliacarne e pubblicato nel giugno del 2000 con il titolo di “Indagine sulla legge 488/92”. Il secondo è stato condotto dal Ministero delle Attività Produttive e pubblicato nel 2003 con il titolo “Valutazione della legge 488/92: gli effetti occupazionali sul territorio”. Infine, utilizzando gli stessi dati alla base del rapporto del Ministero dell'Industria, due ricercatori, Guido Pellegrini

dell'Università di Bologna e Carla Carlucci del Ministero del Tesoro, hanno condotto degli approfondimenti di analisi, apparsi in due pubblicazioni distinte⁴⁸.

7.1 La logica e gli obiettivi della legge 488

La novità principale della politica di intervento introdotta dalla legge 488/92, rispetto ad altri sistemi agevolativi, è il meccanismo di assegnazione degli incentivi attraverso bando di gara, che rimpiazza la distribuzione “a pioggia” dei contributi. Vengono incentivati i progetti che risultano più meritevoli quanto a qualità dell'intervento e che, contemporaneamente, richiedono meno in termini di contributo rispetto al valore dell'investimento totale.

Lo strumento di incentivazione della legge 488 è il contributo a fondo perduto a fronte di programmi di investimento pluriennali volti *“alla realizzazione di nuove attività produttive ovvero all'incremento della capacità produttiva e dell'occupazione, all'aumento della produttività, al miglioramento delle condizioni ecologiche legate ai processi produttivi, all'aggiornamento tecnologico, al rinnovo, alla riorganizzazione, alla diversificazione della produzione, alla modifica dei cicli produttivi, alla ripresa dell'attività, al cambiamento della localizzazione degli impianti di unità produttive esistenti”* (art. 3).

Destinatarie delle agevolazioni sono le imprese dei settori manifatturiero, commerciale, turistico e dei servizi, situate nelle aree del Mezzogiorno in ritardo di sviluppo (aree Obiettivo 1), nelle aree del Centro-Nord a declino industriale (aree Obiettivo 2), nelle zone del Centro-Nord rurali svantaggiate (aree Obiettivo 5b) e in tutte le altre aree depresse ammesse agli aiuti di competenza regionale.

Gli attori coinvolti nel processo di erogazione dei contributi sono, oltre alle imprese richiedenti:

- il Ministero dell'Industria (ora delle Attività Produttive), che pubblica i bandi stabilendo le date di apertura e di chiusura; ripartisce i fondi disponibili per le

⁴⁸ “Gli effetti della legge 488/92: una valutazione dell'impatto occupazionale sulle imprese agevolate” pubblicato nel 2003 in “Rivista Italiana degli Economisti” e “*Non parametric analysis of employment effects of public subsidies to capital accumulation: the case of Law 488/92 in Italy*”, relazione presentata al XIX convegno AIEL, 2004.

agevolazioni alle Regioni; forma le graduatorie e trasferisce i fondi alle Banche concessionarie;

- le Banche concessionarie, che raccolgono le domande, si occupano di condurre l'istruttoria, formulano il giudizio di ammissibilità, trasmettono i risultati al Ministero per la formazione delle graduatorie ed erogano i contributi alle imprese ammesse ai finanziamenti;
- le Regioni, che indicano i settori produttivi e le aree di intervento prioritari.

Entro la data di scadenza del bando relativo al proprio settore produttivo, le imprese sono tenute a presentare domanda corredata da un piano di investimento articolato su più anni (da due a quattro) contenente il dettaglio delle spese che intendono sostenere, la descrizione dell'impresa nei suoi aspetti organizzativi e finanziari e un'analisi del mercato di riferimento. Le domande, compilate su apposito software predisposto dal Ministero, vengono consegnate alle Banche concessionarie che si occupano della fase istruttoria, a conclusione della quale formulano un giudizio di ammissibilità alle graduatorie e attribuiscono un punteggio calcolato sulla base di cinque criteri (illustrati più avanti). Entro 90 giorni dalla data di chiusura del bando, la Banca concessionaria trasmette gli esiti delle istruttorie al Ministero, il quale forma le graduatorie entro un mese.

Per ogni bando vengono create due graduatorie nazionali (una per il Centro-Nord e l'altra per il Sud) nelle quali confluiscono i progetti di importo superiore ai 26 milioni di euro, mentre per i progetti di importo inferiore le graduatorie sono formate su base regionale. Ogni Regione ha una graduatoria ordinaria (con riserve per piccole e medie imprese) e una speciale alla quale può essere riservata una quota consistente delle risorse (fino al 50%) e nella quale (a partire dal 1999) vengono inserite solo le aziende di determinate aree o categorie produttive considerate "critiche" per lo sviluppo economico.

I criteri, in base ai quali le Banche concessionarie attribuiscono i punteggi utilizzati poi dal Ministero per stilare le graduatorie, sono i seguenti: l'occupazione attesa, il capitale

proprio immesso, la percentuale di contributo richiesto, il parametro ambientale e le preferenze regionali.

OCCUPAZIONE ATTESA. È un indicatore calcolato rapportando al capitale investito la differenza tra il numero di addetti attesi “a regime” (cioè dopo l’attuazione dell’investimento) e il numero di addetti 12 mesi prima della richiesta di finanziamento. Questo indicatore, quindi, assumerà valori maggiori laddove maggiore sarà l’occupazione che si intende creare per unità di capitale investito: ciò al fine di evitare che i contributi vadano a finanziare interventi *capital intensive*, quando uno degli obiettivi della politica è l’aumento dell’occupazione.

CAPITALE PROPRIO IMMESSO. Questo indicatore tiene conto solamente dei nuovi mezzi propri investiti nell’impresa, quindi aumenti di capitale sociale e conferimenti in conto capitale versati appositamente in riferimento all’iniziativa. Il valore che l’indicatore assume è dato dal rapporto tra il capitale proprio immesso e l’investimento complessivo. Questo criterio premia le imprese che si assumono una quota maggiore di rischio e che, quindi, implicitamente dichiarano la loro fiducia nella buona riuscita del progetto.

PERCENTUALE DI CONTRIBUTO RICHIESTO. È il rapporto tra l’importo richiesto dall’impresa e l’entità massima dell’agevolazione concedibile. Gli importi massimi concedibili sono stabiliti dall’Unione Europea e dipendono dalle dimensioni dell’impresa (privilegiando quelle minori) e dall’ubicazione dell’unità produttiva (privilegiando le aree maggiormente svantaggiate). L’indicatore assumerà valori tanto più alti, quanto più bassa sarà l’entità dell’agevolazione richiesta. L’uso di questo criterio intende massimizzare il numero degli investimenti agevolati dato l’ammontare di fondi disponibili e scoraggiare l’imprenditore dal richiedere finanziamenti superiori rispetto a quelli strettamente necessari per attivare l’iniziativa.

PARAMETRO AMBIENTALE. È un parametro che incorpora un concetto complesso quale “l’attenzione dimostrata dall’impresa richiedente nei confronti delle tematiche ambientali”. Viene ricavato attraverso le risposte date, in sede di compilazione della richiesta, a domande riguardanti la situazione *pre* e la situazione *post* intervento relativamente alla quantità di rifiuti prodotti e ai consumi di acqua e energia provenienti da fonti rinnovabili.

PARAMETRO REGIONALE. È dato dalla somma dei punteggi prestabiliti dalle Regioni in base a priorità settoriali o territoriali, che tengono conto di tre aspetti: l'ubicazione, il settore di attività e la tipologia di investimento.

Una volta formate le graduatorie, i contributi vengono erogati partendo dal progetto più meritevole fino ad esaurimento dei fondi disponibili. I progetti ammessi in graduatoria ma non finanziati per mancanza di disponibilità vengono automaticamente inseriti in quella relativa al bando successivo. La prima *tranche* di contributo, che può anche essere pari al 50% dello stesso, viene versata entro due mesi dalla pubblicazione delle graduatorie. Le imprese, pertanto, possono avere la disponibilità del denaro sei-otto mesi dopo la presentazione della domanda. Si tratta di tempi di erogazione piuttosto veloci se paragonati a quelli di altre leggi di incentivazione.

Gli incentivi erogati in base alla legge 488 sono cresciuti fino a diventare il principale strumento di agevolazione a favore delle imprese situate in aree economicamente svantaggiate. Dal 1996 al 2003 sono stati pubblicati 13 distinti bandi, di cui 6 riguardanti le imprese del settore manifatturiero. Con questi 13 bandi sono state concesse agevolazioni per 17 miliardi di euro (circa il 30% del totale degli aiuti alle imprese), di cui l'88% destinati ad imprese situate nel Mezzogiorno. I progetti approvati sono stati circa 31.000, di cui due terzi al Sud. A causa delle selettività della procedura di ammissione, meno del 50% delle richieste di finanziamento presentate sono state approvate.

7.2 La domanda di valutazione

La complessità di questa politica pubblica fa sì che le domande che si possono legittimamente porre riguardo ai suoi effetti siano numerose: tale pluralità di possibili domande si è riflessa nelle diverse impostazioni adottate dai rapporti di valutazione esaminati in questo capitolo. Il primo effetto rispetto al quale è legittimo interrogarsi è quello *sulla propensione delle imprese ad investire*, e quindi sull'*addizionalità* degli investimenti agevolati. Un secondo effetto da valutare è quello sulla performance delle imprese beneficiarie delle agevolazioni nel periodo successivo alla realizzazione dell'investimento agevolato. Infine, è legittimo chiedersi quali effetti le agevolazioni

abbiano avuto sui livelli occupazionali delle aree depresse, il cui sviluppo rappresenta il principale obiettivo di questa politica.

L'effetto incentivante delle agevolazioni

Una domanda che ci si pone (o ci si dovrebbe porre) sulle politiche di agevolazione degli investimenti privati riguarda l'*addizionalità* degli investimenti agevolati, cioè l'effetto incentivante delle agevolazioni. Nel nostro caso, la domanda di valutazione può essere formulata nel modo seguente: "L'accesso ai contributi della legge 488 ha spinto le imprese a realizzare progetti di investimento che altrimenti non avrebbero effettuato?" In altre parole, il contributo ha indotto l'impresa ad investire di più o ha semplicemente consentito all'impresa di ridurre il costo di un investimento che avrebbe effettuato comunque, sostituendo le risorse pubbliche ai mezzi propri o all'indebitamento sul mercato del credito?

Lo studio commissionato nel 2000 dall'allora Ministero dell'Industria ("Indagine sulla legge 488/92") si pone innanzitutto questa domanda di valutazione e tenta di ottenerne risposta ricorrendo allo strumento della *beneficiary survey*, cioè un'indagine condotta su un campione di imprese beneficiarie delle agevolazioni. Agli imprenditori intervistati si pongono direttamente quesiti riguardanti l'effetto incentivante delle agevolazioni.

L'effetto sulla performance delle aziende agevolate

Al di là della sua capacità di stimolare nuovi investimenti, ci si attende che l'agevolazione abbia effetti positivi su altre dimensioni della performance aziendale: effetti sul fatturato, sui livelli di occupazione, sulla redditività e sulla stessa probabilità di sopravvivenza dell'impresa sul mercato. Il già citato rapporto del Ministero dell'Industria del 2000 pone esplicitamente questo tipo di domanda di valutazione (che, per quanto attiene all'occupazione, è anche al centro degli studi condotti da Pellegrini e Carlucci).

Le dimensioni della performance aziendale prese in considerazione dallo studio del Ministero sono quattro: il fatturato, l'occupazione, il margine operativo lordo (MOL) rapportato al fatturato e la redditività del capitale investito (ROI). In tutti questi casi, la domanda di valutazione può essere ricondotta allo schema seguente: "Avere ricevuto

l'agevolazione della legge 488 ha fatto aumentare l'occupazione (il fatturato, la redditività) tra le aziende beneficiarie rispetto a quanto sarebbe successo nelle stesse aziende se esse non avessero ricevuto le agevolazioni?"

In questo caso non viene utilizzata la strategia di valutazione basata sulla *beneficiary survey*. Gli effetti vengono invece stimati confrontando la performance delle aziende beneficiarie con quella delle aziende non beneficiarie: quest'ultima, se sono soddisfatte certe condizioni, può servire ad approssimare ciò che sarebbe successo alle aziende beneficiarie in assenza di agevolazioni.

L'effetto sull'occupazione a livello territoriale

Una diversa domanda di valutazione è quella al centro dello studio pubblicato dal Ministero delle Attività Produttive nel 2003, che sposta l'attenzione sugli effetti occupazionali a livello di territorio. Nonostante ci si concentri anche in questo caso sulla variabile occupazione, la logica della valutazione è radicalmente diversa da quella appena vista: non sono in discussione gli effetti della 488 sulla crescita dell'occupazione *nelle aziende agevolate*, bensì gli effetti che le agevolazioni hanno sulla crescita dell'occupazione *complessiva nelle aree territoriali* in cui le imprese stesse sono localizzate (più precisamente, l'effetto che l'occupazione creata nelle aziende agevolate dalla 488 ha sull'occupazione a livello territoriale).

Va notato come questa domanda di valutazione possa ottenere risposte di segno diverso rispetto alle precedenti: è possibile immaginare che le agevolazioni producano un effetto *positivo* sulla propensione delle aziende ad investire e/o ne producano uno *positivo* sull'occupazione nelle aziende agevolate: questi effetti positivi non eliminano tuttavia la possibilità che l'effetto sull'occupazione a livello territoriale risulti *nullo*. È il classico argomento dello *spiazzamento*: le iniziative agevolate possono spiazzare analoghe iniziative che sullo stesso territorio avrebbero realizzato altre imprese, con la conseguenza di produrre un beneficio netto ridotto o addirittura nullo a livello territoriale.

D'altro canto, è possibile che siano all'opera anche meccanismi di segno opposto, cioè tali per cui l'effetto complessivo a livello di territorio si riveli *superiore* alla somma degli effetti a favore delle singole imprese agevolate: questo nel caso in cui gli

investimenti realizzati grazie alle agevolazioni generino sul territorio degli *spill over* positivi e la presenza di nuova accumulazione di capitale generi a sua volta nuova domanda e offerta e quindi ulteriori investimenti, con una ricaduta positiva sullo sviluppo e sull'occupazione a livello locale.

L'esistenza di questo tipo di meccanismi "virtuosi" è suggerita da alcune teorie economiche che analizzano le cause dell'agglomerazione e della dispersione delle imprese sul territorio. Lo sviluppo economico, secondo queste teorie, è da attribuirsi prevalentemente all'agglomerazione sul territorio di un certo settore o di una determinata catena produttiva. In Italia tali agglomerazioni sono state razionalizzate ricorrendo al concetto di "distretto industriale", con cui si sono spiegati i differenziali di crescita a livello territoriale. La predizione che questa teoria fa con riferimento alle agevolazioni della legge 488 è che esse producano esternalità prevalentemente positive, determinando quindi un effetto netto positivo sull'occupazione a livello territoriale.

Nella realtà è possibile che entrambe le esternalità, positive e negative, coesistano. Ciò che la valutazione condotta dal Ministero delle Attività Produttive tenta di stimare è la risultante netta di queste due forze contrastanti. La domanda di valutazione quindi diventa: "Qual è l'effetto netto delle agevolazioni sulla crescita dell'occupazione nelle aree territoriali in cui sono concentrate le imprese agevolate dalla legge 488?". Rispondere a questa domanda richiede un approccio diverso, in cui non si confrontano imprese agevolate e non agevolate, bensì aree a maggiore o minore concentrazione di imprese agevolate.

7.3 Le principali questioni metodologiche e le soluzioni adottate

Le questioni metodologiche che si pongono per la valutazione degli effetti della legge 488 dipendono in parte dalla specifica domanda valutativa, e in generale riproducono quelle tipicamente incontrate per la valutazione degli effetti di una politica pubblica.

Il metodo utilizzato per stimare l'effetto incentivante delle agevolazioni

Partiamo dalla questione dell'addizionalità: l'obiettivo della valutazione è capire se e in che misura le imprese avrebbero effettuato *comunque* gli investimenti proposti anche se

non avessero ricevuto i contributi in conto capitale della legge 488. Se tale quantità risulta pari a quella degli investimenti effettuati, l'effetto incentivante è nullo.

Una strategia possibile è quella di confrontare la quantità di investimenti effettuati da imprese che hanno ricevuto le agevolazioni con quella effettuata da imprese non agevolate. Questa tuttavia non sarebbe una strategia del tutto soddisfacente: per essere coerenti con il quesito valutativo posto, è importante determinare se le imprese agevolate avrebbero effettuato o meno *quei* progetti di investimento *proposti* per il finanziamento, non tanto se avrebbero investito di più *in generale*. Quindi non si può ottenere una risposta a questo quesito specifico mediante un semplice confronto tra imprese agevolate e *generiche* imprese non agevolate, poiché per queste ultime non esiste un "progetto proposto".

Il confronto va invece condotto tra le imprese che hanno avuto un progetto finanziato e quelle che hanno presentato uno specifico progetto, ma non hanno ottenuto il finanziamento. Il meccanismo attuativo della legge 488 consente in linea di principio questo tipo di confronto, in quanto un numero consistente di imprese che entrano in graduatoria non vengono finanziate per esaurimento dei fondi su quello specifico bando. Ad esempio, per il primo e secondo bando della legge 488 (relativi all'industria) sono stati ammessi e inclusi in graduatoria rispettivamente 7.518 e 6.208 progetti di investimento, di cui 1.125 e 1.871 (il 15% e il 30 %) non sono stati finanziati per esaurimento dei fondi.

Per perseguire questa strategia di valutazione sarebbe necessario selezionare un campione di imprese agevolate e uno "di controllo", composto da imprese ammesse in graduatoria, ma non finanziate: dopodiché ad entrambe andrebbe chiesto quale esito abbia avuto il progetto di investimento per cui il finanziamento è stato richiesto. Se le imprese agevolate e le imprese escluse rivelassero la stessa propensione a realizzare i progetti presentati, ciò sarebbe una chiara indicazione di una scarsa addizionalità degli investimenti agevolati. Questo approccio richiederebbe quindi l'effettuazione di un'indagine *ad hoc* sulle imprese agevolate e sulle richiedenti non agevolate.

La necessità di svolgere indagini dirette presso le imprese è pressante per la valutazione dell'effetto incentivante. Mentre per altre dimensioni della performance aziendale, quali

il fatturato o l'occupazione, è possibile ricorrere a dati di bilancio o altre fonti amministrative da cui trarre le informazioni (sia riguardo alle imprese agevolate, sia a quelle non agevolate), nel caso si voglia verificare l'effettuazione o meno di un dato progetto di investimento, l'unica fonte di informazione resta un'indagine diretta presso le imprese.

Seguendo in parte questa esigenza, il Ministero dell'Industria, con il supporto dell'Istituto Tagliacarne, ha effettuato un'indagine diretta sulle imprese coinvolte nel primo e nel secondo bando della legge 488. Tuttavia, nel disegno di questa indagine, è stata fatta la scelta di privilegiare la raccolta di informazioni presso le imprese agevolate, includendo solo un piccolo campione di imprese non agevolate. Quindi l'effetto incentivante è stato valutato non mediante un confronto nei tassi di completamento degli investimenti tra imprese agevolate e quelle escluse per mancanza di fondi, bensì ricorrendo a domande dirette rivolte agli imprenditori beneficiari delle agevolazioni (illustrate in dettaglio più avanti).

L'universo di riferimento era costituito dalle 2.291 imprese che, finanziate sui bandi del 1996 e del 1997, avessero concluso l'intervento agevolato entro il settembre 1999⁴⁹. Da questo universo di riferimento è stato selezionato un campione di 1.391 imprese, stratificate in base alla localizzazione (secondo le macroaree Centro-Nord e Sud) e alla dimensione aziendale (suddivisa in tre classi: piccola, media e grande impresa)⁵⁰.

Queste imprese sono state interpellate in parte nel 1999 e in parte nel 2000, utilizzando un questionario somministrato con una tecnica di rilevazione mista, che prevede l'intervista diretta per una parte delle domande, mentre la parte riguardante i dati di bilancio può essere autocompilata e restituita in un secondo momento.

Il questionario si articola in quattro parti: la prima tesa a rilevare informazioni di carattere generale, la seconda informazioni riguardanti il mercato di riferimento dell'impresa, la terza gli investimenti effettuati e l'accesso al credito, la quarta i risultati

⁴⁹ Questa scelta non è innocua: essa infatti impedisce di utilizzare questo campione per stimare la probabilità di realizzare l'investimento da parte delle imprese agevolate, e quindi di confrontarlo con l'analoga probabilità da parte delle imprese ammesse ma non finanziate.

⁵⁰ Il rapporto del Ministero menziona l'aggiunta nel campione di circa 300 imprese non agevolate che hanno realizzato comunque l'investimento, ma non contiene indicazioni precise circa l'utilizzo dei dati raccolti presso queste imprese.

economici. Le domande contenute nella prima e nella seconda parte hanno natura prettamente descrittiva, mentre quelle utilizzate per la valutazione soggettiva degli effetti incentivanti della 488 si trovano nella terza sezione.

I seguenti sono tre esempi di domande incluse nel questionario con lo scopo di rilevare la percezione degli effetti della legge 488. Come accennato, queste specifiche domande sugli effetti sono incluse in una lunga batteria di domande che riguardano soprattutto aspetti descrittivi dell'attività aziendale.

Box 7.1

D.34	<i>In seguito agli investimenti effettuati con il ricorso alla legge 488, ritiene che gli impianti e i processi produttivi della sua impresa possono considerarsi, rispetto all'attuale livello medio tecnologico:</i>	
	- all'avanguardia	☐ 1
	- abbastanza avanzati	☐ 2
	- allineati all'attuale livello tecnologico	☐ 3
	- già un po' superati	☐ 4
	- decisamente superati	☐ 5

Questa prima domanda *non* tenta esplicitamente un'attribuzione di causalità tra il livello tecnologico di impianti e processi e gli investimenti effettuati: l'espressione "in seguito agli investimenti effettuati" resta ambigua, potendosi intendere come "grazie agli investimenti" o anche solo come "dopo gli investimenti". Questa ambiguità è frequente in questo tipo di questionari e rende più difficile l'interpretazione delle risposte. Ovviamente, quanto più le risposte saranno positive (cioè sono scelte le modalità di risposta "all'avanguardia" o "abbastanza avanzati") tanto più questo getterà una luce positiva sulla legge 488. A rigore però non sappiamo se l'essere all'avanguardia tecnologicamente sia *merito* della legge 488 o sia una *caratteristica* delle aziende che hanno fatto ricorso alla legge 488.

La domanda seguente va invece al cuore del problema dell'effetto incentivante: all'imprenditore si chiede infatti di rivelare se la possibilità di accedere ai finanziamenti della legge 488 sia stata determinante o meno nella decisione di realizzare gli investimenti.

Box 7.2

D.35 Con riferimento agli investimenti realizzati ricorrendo alla legge 488, la decisione di attivarli quanto è stata condizionata dalla possibilità di accedere a tali forme di finanziamento? (una sola risposta)

- in nessuna maniera: si sarebbe comunque investito lo stesso ammontare ☐ 1
- in qualche misura: si sarebbe comunque investito ma con un ammontare inferiore ☐ 2
- in maniera significativa: sarebbe stato radicalmente modificato il piano di investimenti ☐ 3
- in maniera determinante: si sarebbero annullati gli investimenti ☐ 4
- non indica ☐ 5

Questa domanda contiene una richiesta esplicita di attribuzione di causalità, e quindi di individuazione di effetti. L'effetto percepito è massimo se il rispondente afferma che la decisione di investire è stata condizionata *in maniera determinante* dalla possibilità di ricevere i finanziamenti, mentre è nullo quando l'imprenditore rivela che l'investimento sarebbe stato *comunque effettuato nello stesso ammontare*.

Uno schema concettuale analogo è adottato dalla domanda seguente, che riguarda la localizzazione degli investimenti.

Box 7.3

D.37 La scelta della localizzazione degli investimenti effettuati ricorrendo alla legge 488 è stata condizionata dalla possibilità di accedere a tali forme di finanziamento? (una sola risposta)

- in nessuna maniera: si sarebbe comunque investito nella stessa area ☐ 1
- in maniera significativa: si sarebbe forse investito in altra area ☐ 2
- in maniera determinante: si sarebbe certamente investito in altra area ☐ 3

Al rispondente viene chiesto di rivelare in che misura la scelta di localizzare gli investimenti sia stata condizionata dalla possibilità di accedere alle agevolazioni previste dalla legge 488. Le risposte possibili vanno dall'implicare l'assenza di effetto ("si sarebbe investito nella stessa area"), ad implicarne uno decisamente positivo ("si sarebbe certamente investito in altra area").

Rispetto ai metodi quantitativi illustrati nella parte metodologica di questo rapporto, un vantaggio importante di questo approccio "diretto" consiste nell'evitare di dover individuare un "gruppo di controllo" di imprese non beneficiarie. L'effetto viene infatti stimato rilevando le percezioni che di esso hanno i beneficiari diretti della politica: con

tutti i limiti che tale metodologia porta con sé. L'utilizzo di questo tipo di domande può suscitare forti perplessità, sia per il forte onere cognitivo imposto al rispondente, sia per la presenza di possibili incentivi a distorcere le risposte verso un effetto positivo. Come vedremo nella sezione successiva, illustrando i risultati dell'indagine condotta dall'Istituto Tagliacarne, questo secondo motivo di perplessità sembra non sussistere nel caso della legge 488.

Il metodo utilizzato per stimare l'effetto sulla performance delle aziende agevolate

Se le domande dirette rivolte agli imprenditori suscitano perplessità quando siano utilizzate per valutare l'effetto incentivante, ancora maggiori perplessità accompagnerebbe il loro uso nel caso in cui si volessero rilevare gli effetti delle agevolazioni su complesse dimensioni della performance aziendale, quali il fatturato o la redditività. Ottenere dall'imprenditore un'opinione su quanto la legge 488 abbia influito sulle sue scelte di investimento è difficile ma non impossibile: richiede che il rispondente ricostruisca un solo processo decisionale interno, quello che ha portato l'azienda a voler realizzare quel nuovo impianto o quel particolare acquisto di macchinari. Un onere cognitivo nettamente più contenuto rispetto a quello che sarebbe imposto agli imprenditori da eventuali domande sugli effetti che la legge 488 può avere avuto sulle singole dimensioni della performance aziendale.

Per questo tipo di valutazione è più opportuno ricorrere, anziché a domande dirette inserite in una *beneficiary survey*, a *confronti quantitativi* tra la performance tra imprese beneficiarie e non beneficiarie dei contributi, utilizzando gli opportuni correttivi per depurare il confronto da possibili “differenze di partenza” tra i due gruppi. La seconda parte del rapporto del Ministero dell'Industria, intitolata “Analisi degli effetti degli incentivi”, ricorre appunto a diverse varianti di questo fondamentale approccio quantitativo, illustrato in forma elementare nei capitoli 4, 5 e 6 della sezione metodologica.

Il confronto viene effettuato tra le imprese ammesse in graduatoria e poi finanziate, e le imprese ammesse in graduatoria, ma non finanziate per esaurimento delle risorse. Il fatto che entrambi i gruppi di imprese abbiano in comune la volontà di usufruire di

finanziamenti pubblici per realizzare progetti di investimento presumibilmente elimina alcune delle differenze di partenza che invece si avrebbero tra imprese finanziate e imprese che non hanno neppure fatto domanda per i finanziamenti. Questo approccio elimina alcune differenze di partenza, ma non tutte: resta il fatto che le imprese finanziate sono più alte in graduatoria delle imprese non finanziate, e quindi tendono ad essere “migliori”. Sono pertanto necessarie opportune tecniche statistiche per tentare di eliminare le differenze residue e poter così interpretare le differenze di performance tra i due gruppi come *effetto* del finanziamento ricevuto.

Ricorrendo a questo tipo di approccio, lo studio del Ministero dell’Industria tenta di stimare l’effetto dei finanziamenti della 488 su quattro distinte dimensioni della performance aziendale: il fatturato, l’occupazione, il margine operativo lordo (MOL) e la redditività del capitale investito (ROI). Più precisamente, lo studio intende stimare l’effetto della 488 sul *tasso di crescita* di queste variabili nel tempo. Questo richiede di avere a disposizione due osservazioni annuali per ciascuna impresa e per ciascuna variabile analizzata: una relativa ad un anno precedente all’erogazione del contributo e una relativa ad un anno successivo. Poiché lo studio utilizza le imprese ammesse alle graduatorie del primo e del secondo bando, relativi al 1996 e al 1997, gli anni utilizzati per calcolare il tasso di crescita sono il 1995 e il 1998.

Il reperimento di informazioni affidabili su fatturato, occupazione, MOL e ROI per tutte le imprese è una grossa sfida, che i valutatori affrontano utilizzando le migliori banche dati esistenti, ma tenendosi pronti a scendere agli inevitabili compromessi dettati dal contenuto di queste ultime. Qui si riassumono le tappe principali del lungo processo necessario alla costruzione della banca-dati utilizzata nell’analisi, rimandando al Rapporto del Ministero per i dettagli.

a. Le informazioni sulle caratteristiche fondamentali delle aziende sono ricavabili dalla banca-dati sulla 488 del Ministero, utilizzata per la formulazione delle graduatorie e per la gestione del processo di erogazione. L’unità di osservazione di questa banca-dati è tuttavia il singolo “intervento”, non l’impresa, in quanto una stessa impresa può fare domanda per il finanziamento di più di un intervento. L’universo di partenza è rappresentato da 13.554 interventi, ammessi in graduatoria nei primi due bandi. Per 11.608 (l’86%) di questi, ricorrendo anche ad integrazioni dall’Archivio Statistico delle

Imprese Attive (ASIA) dell'ISTAT, si sono ricavate le informazioni essenziali per l'analisi degli effetti sull'occupazione e il fatturato.

b. A questi 11.608 interventi, per i quali sono disponibili dati completi, corrispondono 7.350 imprese, di cui 5.932 sono imprese “agevolate” (nel senso che almeno uno degli interventi da esse presentati è stato finanziato) e 1.418 sono imprese “non agevolate” (nel senso che nessuna delle domande presentate, pur essendo entrate in graduatoria, è stata finanziata). Una parte dell'analisi degli effetti della 488 è condotta su questo campione di imprese.

c. Le informazioni disponibili nella banca-dati del Ministero non sono tuttavia sufficienti per tutte le analisi, ma vanno integrate con informazioni ricavate dai bilanci aziendali successivi al periodo in cui è stata ricevuta l'agevolazione. Questo impone però una perdita di circa il 30% delle imprese presenti in graduatoria, in quanto non costituite da società di capitali, le uniche obbligate per legge a presentare i bilanci civilistici.

d. Per acquisire i dati di bilancio delle società di capitali si è fatto ricorso a due fonti diverse. In prima battuta è stata utilizzata una banca-dati di bilanci civilistici disponibile sul mercato, denominata AIDA, contenente gli ultimi sei bilanci delle società di capitali con fatturato superiore a due miliardi di lire (circa 100.000 imprese). L'incrocio con questa banca-dati ha permesso di acquisire i dati di bilancio relativi al 1995 e al 1998 per un totale di 4.392 imprese incluse nella graduatoria dei primi due bandi. Per le rimanenti 2.298 imprese non trovate nella banca-dati AIDA è stato invece necessario ricorrere alla banca-dati dei bilanci d'impresa delle Camere di Commercio. Per ragioni di costo, sono stati estratti i bilanci di 405 imprese, a cui è stato assegnato un peso statistico che consente loro di rappresentare le 2.298 imprese.

Questa molteplicità di fonti informative ha fatto sì che fossero disponibili campioni diversi per le analisi. Da un lato, è stato utilizzato il *campione completo* di circa 7.000 imprese, i cui dati sono tratti in parte dalla banca-dati del Ministero, in parte dalla banca-dati AIDA e in parte da quella delle Camere di Commercio. Questa soluzione “mista” garantisce una notevole dimensione campionaria, ma a prezzo di una ridotta omogeneità nei dati. In particolare, può essere messa in dubbio la confrontabilità dei

dati ricavati da una fonte per il 1995 e da un'altra fonte per il 1998. Per ovviare a questo problema, i valutatori utilizzano anche un *campione ridotto* di circa 2.000 imprese, in cui tutte le informazioni sulle variabili-risultato sono tratte dalla banca-dati AIDA, sia per il 1995 sia per il 1998. Con questa soluzione si sacrifica la dimensione del campione (e la sua rappresentatività), ma si ottiene una migliore qualità dei dati.

Come si è anticipato, il confronto tra le imprese agevolate e quelle incluse in graduatoria ma non finanziate per mancanza di fondi, deve tenere conto delle differenze “di partenza” che sussistono tra i due gruppi e che non sono quindi imputabili alle agevolazioni: le imprese agevolate saranno sistematicamente diverse, se non altro in quanto superiori in graduatoria rispetto alle non agevolate. I valutatori adottano una pluralità di soluzioni statistiche per eliminare le differenze di partenza tra imprese agevolate e non agevolate, differenze che potrebbero distorcere le stime dell'effetto della 488 sulla performance aziendale, attribuendo alla legge 488 effetti che non ha realmente causato.

Le soluzioni adottate rappresentano tre varianti di un modello-base: quest'ultimo consiste in una regressione lineare che ha come variabile dipendente il tasso di crescita di ciascuna dimensione della performance aziendale (cioè, alternativamente, il tasso di crescita del fatturato, del numero di dipendenti, del margine operativo lordo e del ROI⁵¹).

La variabile-trattamento è quella che indica se l'impresa è stata agevolata o non agevolata: il coefficiente stimato di questa variabile rappresenta l'effetto dell'agevolazione sul tasso di crescita di ciascuna dimensione della performance aziendale. Le variabili di controllo (incluse in tutte le varianti del modello-base) sono le caratteristiche dimensionali (grande, media, piccola), geografiche (Centro-Nord e Mezzogiorno) e il settore economico in cui opera l'impresa.

Le tre varianti al modello-base si differenziano nella soluzione adottata per eliminare le differenze di partenza tra i due gruppi di imprese.

⁵¹ Per la precisione, la variabile dipendente è espressa come il logaritmo del tasso di crescita di queste variabili. L'uso del logaritmo riduce l'influenza dei valori estremi della variabile dipendente e consente di interpretare i coefficienti stimati come effetto percentuale della variabile esplicativa sulla variabile dipendente.

- a. Una prima variante aggiunge alle variabili esplicative la *probabilità stimata* che la singola impresa sia inclusa tra le agevolate. La probabilità è stimata come funzione dei tre indicatori utilizzati nella formazione delle graduatorie⁵². Questa probabilità può essere pensata come un'approssimazione del punteggio ricevuto nella fase di selezione. La sua inclusione nella regressione ha la funzione di correggere il fatto che le agevolate hanno *a priori* caratteristiche “migliori” delle non agevolate e quindi sono più alte in graduatoria.
- b. La seconda variante del modello-base restringe il confronto (cioè il campione utilizzato nella regressione) alle *imprese che si collocano in graduatoria nelle vicinanze della soglia di ammissione*: vengono, cioè, utilizzate le più alte in graduatoria tra le escluse e le più basse in graduatoria tra le incluse. Come illustrato nella sezione metodologica, attorno al punto di discontinuità si crea una situazione simile alla randomizzazione, in cui cioè le differenze di partenza sono minimizzate rispetto alla situazione in cui si usino tutti i soggetti inclusi nella graduatoria.
- c. La terza variante del modello usa una soluzione ancora più radicale per correggere le differenze di partenza: utilizza, tra le imprese non agevolate, solo quelle che hanno riportato, durante l'indagine campionaria effettuata dall'Istituto Tagliacarne, di avere comunque effettuato l'investimento. Questo fatto dovrebbe renderle più simili alle imprese agevolate e quindi dovrebbe far risaltare ancora di più l'effetto della agevolazione sulla performance aziendale.

Il metodo utilizzato per stimare l'effetto sullo sviluppo locale

I problemi metodologici che si pongono per stimare l'effetto della legge 488 sull'occupazione a livello locale sono diversi rispetto al caso della performance aziendale. Innanzitutto, l'unica variabile-risultato considerata è l'occupazione, ma non più definita a livello di singola azienda agevolata, bensì a livello di unità territoriale: si passa cioè da un livello micro ad uno macro. Si vuole stimare l'effetto della 488 sulla crescita dell'occupazione a livello di area.

⁵² L'occupazione attesa, la percentuale di capitale proprio e la percentuale richiesta del massimo dell'agevolazione.

Il primo nodo metodologico è quindi quello di scegliere l'unità territoriale adatta. La Provincia viene ritenuta un'unità troppo grande ed eterogenea per cogliere correttamente gli effetti della legge 488, mentre il Comune un'unità troppo piccola ed eterogenea per dimensione. La scelta ricade su un'unità intermedia, il Sistema Locale del Lavoro (SLL), una ripartizione del territorio definita per finalità di tipo statistico e non amministrativo.

L'ISTAT ha individuato in Italia 784 SLL, di cui 365 appartenenti al Mezzogiorno. La legge 488 è stata applicata in modo non omogeneo sul territorio nazionale, e quindi anche nei diversi SLL. Considerando i primi cinque bandi 488 per l'industria manifatturiera, ben 223 SLL su 784 non contengono imprese che hanno usufruito della legge 488, e persino nel Mezzogiorno 33 su 365 non sono stati toccati dalle agevolazioni della 488 (mentre nel Centro-Nord quelli non toccati sono 190 su 419). Anche tra le aree che includono imprese agevolate, l'intensità dell'intervento della 488 varia considerevolmente da area ad area.

La variazione nel numero degli occupati a livello di SLL (la variabile *sulla quale* si ricerca l'effetto della 488) può essere agevolmente ricostruita a partire dai censimenti dell'industria e dei servizi, disponibili per gli anni 1981, 1991, 1996 e 2001. La valutazione si concentra sulla variazione dell'occupazione tra il 1996 e il 2001. La domanda valutativa quindi diventa "in che misura tale variazione è spiegata dall'intensità dell'intervento della 488 nello stesso periodo"? Ad un'intensità maggiore dell'intervento corrispondono variazioni positive maggiori dell'occupazione?

La seconda scelta metodologica fondamentale è la specificazione della variabile-trattamento (cioè la variabile *della quale* si ricerca l'effetto), quella che è stata appena definita come "intensità dell'intervento della 488". A differenza del rapporto del Ministero dell'Industria, in cui il trattamento a livello di singola impresa è specificato come variabile dicotomica (agevolata/non agevolata), qui i valutatori vogliono tenere conto dell'*intensità* del trattamento che la singola zona ha ricevuto tramite la legge 488. Viene quindi scelta, per rappresentare tale intensità, la variabile "numero di posti di lavoro direttamente creati nelle imprese beneficiarie e documentati in sede di rendicontazione degli investimenti effettuati".

Va ricordato che l’“occupazione attesa” (cioè il numero di posti di lavoro che l’azienda dichiara di voler creare in seguito agli investimenti proposti) è uno degli indicatori utilizzati per determinare il punteggio con cui l’azienda entra nella graduatoria ministeriale. Una volta concluso il progetto, in sede di rendicontazione, l’impresa è tenuta a riferire al Ministero il numero di posti di lavoro *effettivamente creati*. Da queste autodichiarazioni risulta che i progetti completati nel periodo 1996-2001 e relativi ai primi sei bandi “industria” hanno creato 122.646 occupati, di cui 86.218 al Sud: definiremo questi come “occupati-488”.

Questi “occupati-488” sono distribuiti in modo non omogeneo tra i 784 SLL presenti sul territorio nazionale: alcuni SLL non avranno nessun “occupato-488” nel periodo considerato, perché non sono presenti sul loro territorio imprese che hanno concluso progetti finanziati dalla 488, mentre in altri SLL, soprattutto al Sud, saranno invece presenti numeri consistenti di “occupati-488” .

Lo scopo del modello è quello di stimare quanto la variazione dell’occupazione totale registrata nei SLL tra il 1996 e il 2001 (variabile dipendente) sia influenzata dal numero di “occupati-488” attivati nello stesso periodo dalle imprese localizzate nei rispettivi SLL (variabile-trattamento), a parità di altre caratteristiche. In termini formali, il modello di regressione stimato ha la seguente struttura:

$$(1) OCC_{i,2001} - OCC_{i,1996} = \text{costante} + \alpha Occ488_i + \gamma (OCC_{i,1996} - OCC_{i,1991}) + \beta X_i + \varepsilon_i$$

dove:

- la variabile $OCC_{i,anno}$ indica il numero di occupati osservati in ciascun i-esimo SLL nell’anno indicato;
- la variabile $Occ488_i$ indica il numero cumulato di posti di lavoro rendicontato in ciascun i-esimo SLL dalle imprese agevolate dalla legge 488.

La variabile dipendente (la variazione dell’occupazione osservata in ciascun SLL tra il 2001 e il 1996) è quindi spiegata da tre fattori: (i) dall’occupazione direttamente creata dalla legge 488 tra il 1997 e il 2001; (ii) dalla crescita occupazionale osservata nei 5 anni precedenti, cioè tra il 1996 e il 1991, quando la legge 488 non era ancora in vigore;

(iii) da una serie di caratteristiche osservabili di ciascun SLL, e precisamente: il tasso di crescita della popolazione, la superficie del SLL, o il fatto che il SLL sia prevalentemente urbano o ad alta specializzazione produttiva, caratteristiche che possono influenzare sia il ricorso alla legge 488 sia la dinamica dell'occupazione.

Il modello (1) è una variante del modello differenza-nelle-differenze ampiamente utilizzato nella letteratura valutativa: la prima differenza è quella *tra prima e dopo*, cioè tra il 1996, quando non esistevano imprese che avessero concluso investimenti finanziati dalla legge 488, e il 2001, quando invece la politica è stata attuata; la seconda differenza è quella *tra località "trattate"*, nelle quali cioè esistono (più) occupati attivati grazie alla legge 488, e *località "non-trattate"*, dove non ne esistono (o ne esistono di meno).

Il coefficiente di interesse dal punto di vista della valutazione degli effetti è α , il coefficiente che moltiplica la variabile $Occ488_i$, cioè il numero cumulato di posti di lavoro rendicontato in ciascun i -esimo SLL dalle imprese agevolate dalla legge 488. Se la stima del coefficiente α risultasse significativamente più grande di uno, ciò significherebbe che ad un posto di lavoro creato direttamente nelle imprese finanziate dalla legge 488 corrisponde un incremento maggiore di un'unità nella crescita dell'occupazione sul territorio.

Numerosi sono i problemi tecnici che si pongono per la stima corretta di questo modello di regressione. La stima del modello (1) mediante i minimi quadrati ordinari infatti presenta delle possibilità di distorsione, in quanto è presente sul lato destro dell'equazione anche la variabile dipendente "ritardata". In questi casi occorre ricorrere a metodi di stima più sofisticati che tengano conto di tale deviazione dal modello di regressione classico. Gli autori del rapporto ricorrono a due di questi metodi, quello delle variabili strumentali e quello definito come *Generalized Method of Moments*. I risultati, discussi nella prossima sezione, confermano la sensibilità delle stime al metodo prescelto.

7.4 I risultati della valutazione

I risultati della valutazione sono presentati distinguendo a seconda dell'effetto che si intendeva stimare: rispettivamente, l'effetto incentivante, l'effetto sulla performance delle aziende agevolate e quello sullo sviluppo locale. Va sottolineato come si tratti di tre aspetti del problema molto diversi tra loro, per cui non deve stupire l'eterogeneità nei risultati ottenuti. A questa inoltre contribuiscono le forti differenze nei metodi di valutazione utilizzati per rispondere alle diverse domande: differenze peraltro ampiamente giustificate dalla disponibilità dei dati e dalla natura del problema analizzato.

I risultati relativi all'effetto incentivante

I risultati della valutazione dell'addizionalità dell'investimento agevolato sono riportati nella sezione 2.4 del rapporto del Ministero dell'Industria del 2000, opportunamente intitolata "L'efficacia dell'intervento come rilevata dall'impresa". Il titolo di questa sezione riflette infatti la natura dei risultati presentati, che si basano interamente sulla *beneficiary survey* condotta dall'Istituto Tagliacarne tra il 1999 e il 2000 su un campione di circa 1400 imprese agevolate.

Qui commenteremo solo le risposte date a uno dei quesiti sottoposti agli imprenditori, quello che riguarda direttamente la questione dell'addizionalità dell'investimento. Il questionario conteneva molti altri quesiti riguardanti la performance dell'azienda agevolata, ma in questi non si ravvisa il tentativo di valutare gli effetti delle agevolazioni, bensì quello di descrivere le caratteristiche delle imprese agevolate.

La domanda è quella riportata per intero nel Box 7.2: all'imprenditore si chiede in quale misura la decisione di attivare gli investimenti realizzati ricorrendo alla legge 488 sia stata condizionata dalla possibilità di accedere a tali forme di finanziamento. La Tabella 7.1 contiene la distribuzione di frequenza delle risposte date dagli intervistati.

Le risposte degli imprenditori fugano la preoccupazione che da parte loro prevalga un atteggiamento di "acquiescenza", cioè che le risposte siano distorte sistematicamente verso il riconoscimento di ruolo positivo svolto dalle agevolazioni nell'influenzare le decisioni di investimento. Se tale acquiescenza fosse diffusa, essa distorcerebbe le

risposte verso un giudizio largamente positivo sull'effetto dell'intervento della legge 488.

Ciò tuttavia non succede nel nostro caso: la stragrande maggioranza degli imprenditori intervistati ammette candidamente come la decisione di attivare gli investimenti *non* sia stata influenzata in maniera determinante e neppure significativa dalla possibilità di accedere ai finanziamenti: a livello nazionale, ben l'80% indica che o si sarebbe investito lo stesso ammontare o si sarebbe comunque investito anche se in misura inferiore. Solo il 15% indica un sicuro effetto positivo della legge 488, e solo il 4% arriva a dire che l'effetto della legge è stato determinante e che in sua assenza si sarebbero annullati gli investimenti.

Tabella 7.1: Stime dell'effetto incentivante delle agevolazioni

Risposte alla domanda "Con riferimento agli investimenti realizzati ricorrendo alla legge 488, la decisione di attivarli quanto è stata condizionata dalla possibilità di accedere a tali forme di finanziamento?"	Frequenza percentuale delle risposte		
	Italia	Centro-Nord	Mezzogiorno
- in nessuna maniera: si sarebbe comunque investito lo stesso ammontare	41	49	27
- in qualche misura: si sarebbe comunque investito ma con un ammontare inferiore	39	36	44
- in maniera significativa: sarebbe stato radicalmente modificato il piano di investimenti	11	7	17
- in maniera determinante: si sarebbero annullati gli investimenti	4	1	9
- non indica	5	7	3
Totale	100%	100%	100%

Fonte: Ministero dell'Industria (2000)

L'effetto incentivante appare leggermente più forte nel Mezzogiorno, in linea con la maggiore generosità dei contributi erogati al Sud: la percentuale di coloro che non riportano un ruolo determinante o significativo della legge 488 scende al 71%, con una presenza più ridotta rispetto al Centro-Nord di coloro che ammettono che le agevolazioni non hanno alterato per nulla le loro decisioni di investimento (il 27% al Sud contro quasi il 50% al Centro-Nord). Al Sud la percentuale di coloro che reclamano un ruolo positivo per la legge 488 sale al 26%, contro un mero 8% al Centro-Nord.

Nonostante il maggiore effetto incentivante che la legge 488 sembra avere al Sud, *emerge chiaramente da questa analisi come l'effetto incentivante delle agevolazioni sia piuttosto limitato*: a detta degli stessi imprenditori, nella maggioranza dei casi l'investimento sarebbe stato realizzato comunque. Anche se le risposte fossero sistematicamente distorte verso l'acquiescenza, una volta tenuto conto di tale distorsione, si otterrebbe un risultato ancora più accentuato di quello ottenuto.

I risultati relativi agli effetti sulla performance delle imprese agevolate

Le stime degli effetti della legge 488 sulla performance delle imprese agevolate variano considerevolmente a seconda della variabile-risultato considerata e della particolare soluzione utilizzata per eliminare le differenze di partenza tra imprese agevolate e non agevolate. La Tabella 7.2 riassume i risultati che nel rapporto del Ministero sono riportati in quattro tabelle separate e di non agevole lettura. Ciascuna colonna rappresenta una diversa dimensione della performance aziendale.

Poiché ogni variabile dipendente è inserita nel modello sotto forma di tasso di crescita (più precisamente, come logaritmo del tasso di crescita), i valori che appaiono nel corpo della tabella vanno interpretati come “differenza percentuale tra imprese agevolate e non agevolate nel tasso di crescita della particolare variabile-risultato considerata”. Ad esempio, il primo modello utilizzato indica un tasso di crescita del fatturato del 18% più elevato tra le imprese agevolate rispetto a quelle non agevolate.

La Tabella 7.2 mostra una certa instabilità nelle stime passando da un modello all'altro per la stessa variabile-risultato. Tuttavia, i risultati positivi sono quasi tutti statisticamente significativi. L'indicazione generale che se ne può trarre è quindi di tipo prevalentemente qualitativo: *le agevolazioni hanno un effetto positivo sulla performance delle aziende agevolate, ma è difficile quantificare credibilmente questo effetto*. Più in particolare:

- *L'effetto sul tasso di crescita del fatturato* è positivo e statisticamente significativo in tutti i modelli stimati: nei due modelli che mostrano risultati più stabili (il primo e il terzo), l'effetto varia tra il 12 e il 18%.
- *L'effetto sul tasso di crescita dell'occupazione* è significativo in tre dei quattro modelli e varia in un intervallo abbastanza limitato, tra il 15 e il 23%. Anche in questo

caso è ragionevole concludere che l'avere ricevuto un'agevolazione della legge 488 aumenta il tasso di crescita dell'occupazione.

▪ Meno stabili sono i risultati per quanto riguarda il tasso di crescita della redditività, misurata mediante il MOL sul fatturato e il ROI: due su quattro stime non sono significative e la stima puntuale in qualche caso è addirittura negativa. Tuttavia, anche in questo caso i due modelli più stabili producono stime positive.

Tabella 7.2: Le stime degli effetti delle agevolazioni sulla performance aziendale

Fonte dei dati	Modello utilizzato per eliminare le differenze di partenza tra imprese agevolate e non agevolate	Effetto sul tasso di crescita derivante dall'essere un'impresa agevolata			
		Fatturato	Occupazione	MOL/fatturato	ROI
Solo dati di fonte AIDA	Campione limitato alle imprese attorno al punto di discontinuità	18%	15%	25%	39%
	Gruppo di controllo costituito da imprese non agevolate che hanno investito	47%	21%	- 8%	11%
	Probabilità di essere agevolata inclusa nel modello	12%	17%	12%	17%
Anche dati di fonte CERVED	Probabilità di essere agevolata inclusa nel modello	23%	23%	- 4%	- 4%

Nota: I valori in grassetto indicano stime statisticamente significative al 90% di confidenza
Fonte: Ministero dell'Industria (2000)

Il rapporto del Ministero conclude infatti che “*i risultati dell'analisi, pur se da considerare con la dovuta cautela ... portano ad indurre che l'effetto dell'agevolazione è positivo e significativo sulle variabili di crescita considerate, ovvero fatturato e dipendenti. L'effetto sulla redditività è incerto: questo può derivare dal fatto che l'agevolazione può attivare anche progetti con redditività minore della media. Le imprese agevolate, comunque, mantengono una redditività complessiva superiore a quella del campione di controllo...*”.

I risultati positivi che le agevolazioni hanno sull'occupazione sono stati replicati, anche se in parte attenuati, da Guido Pellegrini e Carla Carlucci in una serie di analisi econometriche. Nel primo studio, i due ricercatori specificano una varietà di modelli parametrici per eliminare le differenze tra i due gruppi di imprese, agevolate e non agevolate. I modelli adottati sono di natura più generale rispetto a quelli utilizzati per

ottenere le stime presentate nel rapporto del Ministero, e consentono di controllare un maggiore ventaglio di fonti di eterogeneità.

Tra le innovazioni introdotte c'è la possibilità di tenere in considerazione la variabilità connessa a trend di crescita pre-intervento differenti da impresa a impresa; le stime d'impatto sono ottenute in quel caso con un modello di regressione più articolato (ma affine per logica ai modelli precedentemente descritti), per la cui stima si utilizzano, oltre alle informazioni sul numero di dipendenti nel 1998 (post-trattamento) e nel 1995 (pre-trattamento), anche delle informazioni aggiuntive relative al 1994, allo scopo di osservare le dinamiche occupazionali precedenti all'implementazione degli interventi. Le stime degli effetti della legge 488 sul tasso di crescita occupazionale nelle imprese agevolate mostrano una contenuta variabilità da modello a modello (dal 3% al 15%), ma rimangono positive e significative.

Nel secondo studio, gli autori verificano la robustezza dei risultati ottenuti replicando le stime dell'effetto sull'occupazione per mezzo di tecniche non parametriche. Una prima tecnica consiste nel semplice confronto della crescita media dell'occupazione fra le imprese agevolate e non agevolate posizionate in un intorno del punto di discontinuità della graduatoria (metodo illustrato al capitolo 6 della sezione metodologica).

Parallelamente, si stima l'impatto ricorrendo a diverse tecniche di *matching*, confrontando cioè la crescita di ogni impresa beneficiaria con quella di una o più imprese non beneficiarie ad essa simili (metodo illustrato al capitolo 5 della sezione metodologica). Le tecniche differiscono tra loro per il numero di abbinamenti effettuati (gli autori ricorrono a un abbinamento tra ogni beneficiaria e la non beneficiaria maggiormente simile, o in alternativa a un abbinamento tra blocchi di imprese appartenenti a strati con caratteristiche affini) e per il criterio adottato per valutare la somiglianza (confronto diretto su dimensione, Regione e settore di attività, oppure confronto su *propensity score*, cioè probabilità stimata di beneficiare dei finanziamenti).

Indipendentemente dalla tecnica utilizzata, i risultati suggeriscono un effetto positivo sul tasso di crescita dell'occupazione, per quanto più contenuto rispetto alle stime parametriche (circa il 10% per il confronto intorno al punto di discontinuità, tra il 3% e il 6% per le varie tecniche di *matching*), e sembrano confermare, seppure attenuandole,

le evidenze ottenute in precedenza circa un impatto positivo degli incentivi sull'aumento dell'occupazione nelle aziende agevolate.

I risultati relativi all'effetto sull'occupazione a livello territoriale

Concludiamo questa rassegna degli effetti della legge 488 riassumendo le stime riportate nel rapporto del Ministero delle Attività Produttive del 2003. Come anticipato nella sezione precedente, i ricercatori stimano tre versioni dello stesso modello, basandosi su diverse soluzioni al problema della presenza di una variabile dipendente ritardata tra le variabili esplicative. Inoltre, il modello viene stimato separatamente per il Mezzogiorno e per il Centro-Nord. La Tabella 7.3 riporta le stime dell'effetto che un posto di lavoro in più creato dalla legge 488 avrebbe sulla crescita dell'occupazione a livello locale.

Risulta subito evidente la differenza tra il Mezzogiorno e il Centro-Nord: per il primo, l'effetto stimato è sempre positivo e significativo, mentre per il secondo risulta negativo in due casi, sostanzialmente zero nel terzo caso. La stima ottenuta dal terzo modello per il Mezzogiorno implica che ogni occupato in più creato in un'impresa agevolata dalla legge 488 attiverrebbe quasi 4 occupati in più a livello territoriale, mentre al Centro-Nord un occupato 488 in più sarebbe associato alla perdita di quasi 2 posti di lavoro nell'area di riferimento.

Tabella 7.3: Le stime degli effetti sull'occupazione a livello territoriale

Modello econometrico utilizzato	Stima dell'effetto di un occupato 488 sulla crescita dell'occupazione a livello locale	
	Mezzogiorno	Centro-Nord
Minimi quadrati ordinari	0,66	- 2,18
Variabili strumentali	1,91	- 0,16
Generalized Method of Moments	3,75	- 1,79

Nota: I valori in grassetto indicano stime statisticamente significative al 95% di confidenza
Fonte: Ministero delle Attività Produttive (2003)

Tuttavia, avvertono gli autori del rapporto, è possibile che il modello non abbia eliminato del tutto la presenza di effetti di selezione, tali per cui la relazione positiva al Sud tra occupati 488 e dinamica occupazionale sia dovuta al meccanismo di selezione

delle aree in cui si collocano gli investimenti: ipotesi plausibile soprattutto per il Mezzogiorno, dove può essere la dinamica occupazionale positiva ad attirare gli investimenti, e non gli investimenti a influenzare la dinamica occupazionale. Per eliminare questa potenziale distorsione, gli autori stimano una serie di varianti del terzo modello, ma ottengono una stima quasi inalterata del coefficiente della variabile relativa all'occupazione 488. Gli autori concludono che *“appare quindi confermata l'ipotesi che al Sud gli effetti di attivazione di occupazione nell'area, generati dalla 488, sono positivi: non si registra un effetto spiazzamento, ma al contrario un forte effetto moltiplicativo... legato probabilmente alla creazione di possibilità di sviluppo a livello locale”*.

CAPITOLO 8

GLI EFFETTI DELLA LEGGE REGIONALE SARDA 15/94

SULLA PERFORMANCE DELLE IMPRESE

di Luca Mo Costabella

La legge regionale sarda 15 del 1994 può essere considerata una versione locale della legge 488, discussa nel capitolo precedente. Simile è l'obiettivo generale, quello di migliorare le condizioni economiche di zone svantaggiate, simili sono le soluzioni adottate, che consistono in agevolazioni alle imprese che intendono investire sul territorio.

Anche nel caso di tale legge, l'importanza dello strumento e la sua visibilità hanno stimolato l'esigenza di approfondire con uno studio *ad hoc* l'efficacia delle agevolazioni erogate, ma, a differenza della legge 488, non è dalla sede legislativa che proviene il primo impulso alla valutazione. In questo caso la spinta è venuta direttamente dall'Assessorato all'Industria della Regione Sardegna, che ha commissionato all'Osservatorio Economico della Sardegna uno studio con lo scopo specifico di valutare gli effetti, su una serie di dimensioni relative alla performance delle imprese sarde, delle agevolazioni erogate nei primi due bandi della legge 15.

Il rapporto tecnico che risponde a questo mandato dell'amministrazione regionale e rappresenta la fonte informativa per questo capitolo, è stato completato nel dicembre 2005 a cura di Alessandra Mura, Igor Lobascio e Cristina Persico dell'Osservatorio economico della Sardegna, con il titolo "*Valutazione degli incentivi della Regione Sardegna: indagine sugli effetti della Legge regionale n. 15/1994*".

8.1 La logica della legge 15

L'inadeguatezza delle infrastrutture e dei servizi, la bassa qualità delle risorse umane, la difficoltà incontrata dalle imprese per accedere al credito, sono tra i problemi caratteristici delle aree che evidenziano un ritardo nello sviluppo economico. Nel caso sardo, le complicazioni aumentano a causa della natura insulare del territorio, che rende

più difficili gli scambi con l'esterno, non permettendo alle imprese di sfruttare in modo adeguato le economie di scala.

La legge 15 intende contribuire a migliorare questa situazione di ritardo facilitando lo sviluppo delle imprese sul territorio, in particolare riducendo i costi per l'acquisizione di capitale. La legge prevede varie tipologie di agevolazione (contributi in conto capitale, contributi in conto interessi e partecipazioni delle società finanziarie controllate dalla Regione nel capitale sociale delle imprese), ammissibili per sostenere investimenti che abbiano uno dei seguenti scopi: costruzione di nuovi impianti, ampliamento di impianti produttivi preesistenti, ammodernamenti e ristrutturazioni, riconversione della produzione, riattivazione di insediamenti produttivi inattivi, trasferimento di impianti.

Sono potenziali beneficiarie tutte le piccole e medie imprese industriali che rientrano nei settori ammissibili. La definizione di piccole e medie imprese utilizzata non era inizialmente allineata con quella della normativa nazionale e comunitaria. La Direttiva di attuazione del 1996 della legge 15, che ha regolamentato il primo bando (relativo alle domande presentate tra il dicembre 1996 e il febbraio 1998, nel seguito I Bando) considerava piccole e medie imprese tutte quelle con investimenti fissi non superiori a 43 miliardi di lire al netto degli ammortamenti.

Solo a partire dal secondo bando (che ha coperto il periodo da maggio 1999 a settembre 1999), attuato secondo le nuove Direttive del 1998, tale definizione si è conformata a quella comunitaria, circoscrivendo le piccole e medie imprese a quelle con un numero di addetti non superiore a 250, un fatturato annuo non superiore ai 40 milioni di euro (o un totale di bilancio non superiore ai 27 milioni) e possedute per una quota non superiore al 25% da altre imprese non rientranti tra le PMI.

Il sistema di selezione delle domande di agevolazione segue i meccanismi adottati dalla legge 488. È previsto un sistema d'asta, secondo il quale vengono privilegiate le domande che eccellono rispetto ad una serie di criteri; tra questi spiccano l'ammontare dell'agevolazione richiesta e i risultati aziendali previsti. L'utilizzo di tali criteri cautele le istituzioni da due tipologie di rischio: da un lato finanziare imprese che anche senza agevolazione avrebbero effettuato gli investimenti, dall'altro concedere agevolazioni di entità superiore a quelle realmente necessarie agli investimenti.

Anche rispetto ai criteri di selezione esiste una forte differenza tra il I Bando e i successivi, differenza che si rifletterà anche sui metodi adottabili per la valutazione. Nel primo periodo di attuazione, la legge definisce quattro indicatori di priorità sulla cui base è costruita una graduatoria: (a) *occupazione attesa*: è lo stesso indicatore usato per la selezione delle agevolazioni della legge 488, ed è dato dal rapporto tra investimento e occupazione aggiuntiva; (b) *capitale proprio immesso*: anche questo deriva concettualmente dalla legge 488, ed è costituito dalla percentuale di capitale proprio che l'azienda intende investire rispetto al totale dell'investimento previsto (tale rapporto non deve essere inferiore al 30%, una soglia più restrittiva di quella prevista dalla 488, che impone un minimo del 25%); (c) *valorizzazione delle risorse locali*: si intende premiare le imprese la cui attività si appoggia maggiormente alle risorse derivanti dalla produzione locale: il criterio è calcolato come percentuale di materie prime e semilavorati locali sul totale di quelli utilizzati nel processo produttivo; (d) *esportazioni verso mercati extra-isolani*: è calcolato come rapporto tra le esportazioni e le vendite totali, ed è volto a premiare le attività che, con l'esportazione, cercano di superare le barriere dovute alla natura insulare del territorio.

A partire dal II Bando, i criteri di selezione utilizzati per attuare la legge 15 si allineano a quelli della legge 488, acquisendone le modalità di formazione del punteggio, che risulta dunque basato sui cinque indicatori già descritti nel capitolo 7.

Un'ulteriore importante differenza tra il I Bando e i successivi riguarda la modalità di gestione delle domande di agevolazione escluse per mancanza di fondi. Nella sua prima formulazione, la legge 15 prevede una valutazione trimestrale delle domande di agevolazione: le domande escluse per mancanza di fondi vengono rinviate al trimestre successivo. Il risultato di questa prassi è che le imprese partecipanti al I Bando e inserite in graduatoria sono state *tutte* agevolate, mentre ciò non è avvenuto nei Bandi successivi.

8.2 La domanda di valutazione

Pur variando da un bando all'altro i criteri utilizzati per formare le graduatorie, una serie di punti restano in comune. Quello più evidente riguarda l'attenzione all'occupazione

prevista: tale indicatore assume una rilevanza variabile da bando a bando, ma non viene mai eliminato. Ciò conferma la tendenza, in linea con quella della legge 488, a favorire quelle imprese che, a parità di agevolazione richiesta, promettono un maggiore aumento dell'occupazione. Quindi l'occupazione resta anche in questo caso una dimensione privilegiata su cui valutare gli effetti degli interventi.

Una seconda grandezza d'interesse su cui i valutatori concentrano l'attenzione è il valore della produzione. Per quanto l'obiettivo di accrescere la produzione non sia esplicitato nella legge e nelle direttive di attuazione quanto quello relativo all'occupazione, gli autori dello studio identificano in esso una dimensione legata al conseguimento della finalità generale della legge 15, cioè lo sviluppo economico dei territori in cui risiedono le imprese finanziate. La grandezza utilizzata per approssimare il valore della produzione è la somma del fatturato e della variazione delle rimanenze, preferita al valore aggiunto in virtù della migliore copertura informativa nelle banche dati utilizzate.

La variabile-trattamento utilizzata in alcuni casi è semplicemente la variabile dicotomica che distingue tra imprese che non hanno ricevuto le agevolazioni e imprese che le hanno ricevute. In altri è invece una variabile continua, rappresentante l'importo erogato ad ogni impresa. In tali casi, l'effetto stimato va letto come variazione nella variabile-risultato attribuibile a una variazione unitaria dell'importo erogato⁵³.

Dato questo impianto, le domande valutative a cui gli autori cercano di dare risposta sono le seguenti: (i) qual è la differenza tra il numero medio di occupati nelle imprese dopo che queste hanno ricevuto le agevolazioni della legge 15 e il numero medio di occupati che le stesse imprese avrebbero avuto se non avessero ricevuto le agevolazioni? (ii) qual è la differenza tra la produzione media⁵⁴ delle imprese dopo che queste hanno ricevuto le agevolazioni della legge 15 e la produzione media che le stesse imprese avrebbero avuto se non avessero ricevuto le agevolazioni?

La valutazione dell'effetto delle agevolazioni è condotta separatamente per il I Bando e il II Bando, relativi rispettivamente ai periodi dicembre 1996 - febbraio 1998 e maggio

⁵³ Ad esempio può rappresentare, a seconda dell'unità di misura adottata, l'aumento nell'occupazione ogni 100.000 euro erogati, oppure ogni 1.000 euro.

⁵⁴ Calcolata come rapporto tra fatturato e costo del lavoro.

1999 - settembre 1999. L'effetto è stimato misurando le variabili-risultato a tre anni di distanza da tali periodi, cioè nel 2002 per il I Bando e nel 2003 per il II Bando, quando tutte le imprese agevolate dovrebbero aver terminato gli investimenti previsti.

8.3 Le principali questioni metodologiche e le soluzioni adottate

La strategia principale utilizzata per la valutazione è il confronto tra trattati e non-trattati, cioè tra imprese che hanno e non hanno ricevuto agevolazioni tramite la legge 15. Le imprese di “controllo” vengono estratte comunque tra le imprese che hanno fatto richiesta di finanziamento. Affinché il confronto dei livelli di occupazione e di produzione tra imprese agevolate e non agevolate fornisca una stima non distorta dell'effetto, è necessario che queste siano simili nelle loro caratteristiche di partenza; solo in tal caso, le eventuali differenze nelle grandezze considerate si possono addebitare alla diversa esposizione alla politica. I due gruppi di imprese sono però caratterizzati da differenze sistematiche, connesse alle caratteristiche che hanno determinato l'assegnazione delle agevolazioni alle prime e non alle seconde.

Le imprese di controllo possono essere entrate in graduatoria e poi essere state escluse dal finanziamento per il loro basso punteggio, nel qual caso risulterebbero sistematicamente differenti dalle agevolate rispetto agli indicatori che compongono il punteggio. Una situazione diversa è quella che si crea quando le imprese, pur avendo un punteggio tale da garantire loro accesso alle agevolazioni, non ne abbiano invece usufruito per altri motivi: queste imprese sarebbero comunque sistematicamente diverse dalle imprese agevolate.

Il metodo utilizzato per stimare gli effetti del I Bando

Per i peculiari metodi adottati per la selezione, nel I Bando non si dispone di un gruppo di imprese escluse per esaurimento dei fondi e per via del punteggio troppo basso. Non potendo sfruttare il punteggio come discriminante tra trattati e non-trattati, le differenze di partenza vanno eliminate seguendo altre strade.

Delle 353 imprese richiedenti e ammesse all'istruttoria del I Bando, solo 149 hanno ricevuto l'agevolazione. Si possono quindi utilizzare le 204 imprese non agevolate come base per individuare uno o più gruppi di confronto. Va però esercitata molta

cautela nel condurre questa operazione. Non solo le 204 imprese non agevolate sono sistematicamente diverse dalle 149 agevolate, ma anche al loro interno è possibile riscontrare differenze determinate sia dallo stato di avanzamento dell'iter amministrativo, sia dal comportamento dell'impresa che in qualsiasi momento può decidere di abbandonare lo stesso iter, rinunciando al finanziamento. I valutatori sfruttano la distinzione tra le diverse caratteristiche che hanno determinato la selezione (o auto-selezione) per individuare quattro differenti gruppi di confronto, ognuno caratterizzato da particolari punti di somiglianza o di differenza rispetto al gruppo di trattamento.

Un primo gruppo (G1) è costituito dalle 100 imprese la cui pratica è stata *archiviata per decorrenza dei termini* necessari per l'invio della documentazione richiesta. Questo primo gruppo di confronto presenta le maggiori differenze rispetto alle 149 imprese agevolate in ordine allo stato di avanzamento dell'iter amministrativo. Lasciando trascorrere il tempo utile per l'invio della documentazione, queste 100 imprese manifestano la volontà di rinunciare al finanziamento: a sua volta, questa volontà di rinuncia può implicare l'esistenza di differenze di partenza tra le imprese rinunciatricie e le imprese agevolate.

Un secondo gruppo (G2) è costituito dalle 29 imprese che, pur avendo terminato l'iter amministrativo, non hanno ricevuto gli incentivi perché l'istruttoria ha dato *esito negativo*. La differenza fondamentale tra queste 29 imprese e le 149 agevolate consiste, dunque, nella minore credibilità del progetto finanziario presentato, che potrebbe implicare forti differenze di partenza.

È possibile inoltre individuare un terzo gruppo (G3) di 61 imprese che, dopo aver terminato l'istruttoria con esito positivo e ottenuto il decreto di concessione, hanno *rinunciato* al finanziamento. In alcuni casi, tale rinuncia è esplicitamente riconducibile alla decisione di accedere a finanziamenti alternativi, in altri no. In tutti i casi il comportamento che ha determinato la rinuncia o la revoca del contributo differenzia sistematicamente queste imprese dalle 149 che hanno ricevuto le agevolazioni, e costituisce un'altra fonte di distorsione che è necessario tentare di eliminare.

Un ultimo gruppo (G4) è costituito da 14 imprese alle quali è stato *revocato* il finanziamento per non aver completato l'investimento entro i termini stabiliti dalla legge.

Gli autori ricorrono al metodo differenza-nelle-differenze, illustrato nel capitolo 4, per ottenere stime al netto delle differenze di partenza. Disponendo di un'osservazione sulla variabile-risultato precedente al periodo di erogazione dei contributi, la differenza tra valori pre-agevolazione e post-agevolazione (nel numero di occupati e nel fatturato) calcolata tra imprese agevolate e non agevolate dà una possibile stima dell'effetto degli incentivi della legge 15.

Come spiegato nel capitolo 4, tale stima dell'effetto è corretta se vale la condizione (detta "di parallelismo") che le differenze tra imprese agevolate e non agevolate osservate *prima* delle agevolazioni avrebbero dovuto riprodursi tali e quali *in assenza* delle agevolazioni. Detto diversamente, il cambiamento osservato nelle imprese non-avvolate si sarebbe verificato in egual misura anche tra le agevolate, se queste non avessero ricevuto il finanziamento. Nel caso in cui questa condizione non sia plausibile, è necessario disporre di osservazioni in più momenti precedenti alle agevolazioni.

I valutatori hanno a disposizione osservazioni sulle variabili-risultato per ogni impresa fino all'anno 2002. Sarebbe dunque possibile, in via teorica, produrre stime con il metodo differenza-nelle-differenze utilizzando i dati del 2002, del 1999 e del 1996. In pratica, le osservazioni sul 1996 sono utilizzabili solo per le stime degli effetti sull'occupazione. Per quanto riguarda invece la stima dell'effetto sulla produzione, l'insufficiente copertura nelle banche dati costringe gli autori ad adottare un metodo differenza-nelle-differenze basato sulle sole osservazioni del 1999 e del 2002.

Il metodo utilizzato per stimare gli effetti del II Bando

Il meccanismo di selezione operante per II Bando è diverso, in quanto *non tutte* le imprese incluse in graduatoria hanno ricevuto le agevolazioni. Delle 216 imprese che hanno ricevuto un punteggio, 109 sono risultate assegnatarie di un finanziamento e le altre 107 (le più basse in graduatoria) sono state escluse a causa dell'esaurimento dei fondi.

È perciò nota a priori l'esistenza di un'unica grandezza, cioè il punteggio ricevuto, che ha determinato l'assegnazione delle agevolazioni ad alcune imprese e non ad altre. Gli autori producono stime degli effetti seguendo la strategia del *regression discontinuity design*, illustrata in dettaglio nel capitolo 6. In sintesi, si stima un modello di regressione in cui la variabile-risultato è funzione di due sole variabili esplicative: il punteggio e una variabile dicotomica che distingue tra imprese agevolate e non agevolate. Va ricordato tuttavia che il *regression discontinuity design* fornisce stime per loro natura non generalizzabili a soggetti (in questo caso imprese) con punteggi molto distanti dalla soglia che discrimina tra trattati e non-trattati.

La procedura di stima appena descritta è modificata per tenere conto di una complicazione, il fatto cioè che non tutti i “vincitori” del bando usufruiscono dei finanziamenti. Infatti, solo 68 delle 109 imprese con i punteggi più alti hanno ricevuto il finanziamento previsto. Tra le altre 41, tutte rinunciatarie, 11 hanno ricevuto forme alternative di incentivi derivanti da altre leggi (ad esempio dalla legge 488). Ipotizzando che l'effetto di questi altri incentivi sia complessivamente equivalente a quello della legge 15, queste 11 imprese sono state assimilate alle altre 68 finanziate. Per quanto riguarda le 30 rimanenti, 20 non hanno sicuramente ricevuto alcun finanziamento, mentre delle ultime 10 non si conosce la reale condizione, per cui sono state escluse dall'analisi.

Una situazione di non conformità tra risultati della graduatoria e fruizione di agevolazioni si verifica anche per le 107 imprese con punteggi più bassi: 75 sono state effettivamente escluse dai finanziamenti, mentre 32 sono state beneficiarie di erogazioni da fonti alternative, dalla legge 488 a bandi successivi della stessa legge 15, il cui effetto è supposto essere uguale a quello del II Bando.

La presenza di una quota di imprese “non obbedienti” alla condizione loro assegnata comporta una stima distorta dell'effetto della politica. Nel caso in cui si cercasse di stimare l'effetto semplicemente utilizzando come variabile-trattamento *la condizione assegnata dalla graduatoria* alle singole imprese (cioè una variabile dicotomica che discrimina tra le 109 imprese con i punteggi più alti e le 107 coi punteggi più bassi), si commetterebbe l'errore di considerare finanziate quelle 20 imprese con punteggio alto

che non hanno ricevuto alcun finanziamento, e di considerare non finanziate quelle 32 imprese con punteggi bassi che hanno ricevuto agevolazioni da fonti alternative. Se invece si ricorresse all'*effettiva condizione* come variabile-trattamento, la stima ottenuta con il *regression-discontinuity design* non eliminerebbe tutte le differenze di partenza, ma soffrirebbe della presenza di altre differenze nelle caratteristiche non osservabili che hanno portato alcune imprese a “disobbedire” alla condizione loro assegnata.

I dati utilizzati

I dati utilizzati per stimare l'effetto della legge 15 provengono principalmente da due fonti. Per le informazioni sull'occupazione, gli autori ricorrono all'Archivio Statistico Regionale delle Imprese (ASTRI), un archivio costruito dall'Osservatorio Economico della Sardegna. Lo stesso Osservatorio cura il Nuovo Archivio dei Bilanci (NAB), che costituisce la seconda fonte di informazioni, dalla quale si ricavano i dati sul fatturato delle imprese.

I due archivi mostrano tuttavia problemi di copertura rispetto alle esigenze informative poste dalla valutazione. L'archivio ASTRI raccoglie informazioni solo sulle imprese attive. Questo costituisce un vincolo al reperimento dei dati necessari, dal momento che nell'archivio della legge 15 è presente un grande numero di imprese in via di costituzione. Inoltre, le imprese dell'archivio della legge 15 non sono sempre identificabili con una chiave univoca (codice fiscale) che permetta di collegare le informazioni dei due archivi; questo secondo limite è più evidente per le imprese non agevolate. Per quanto riguarda il NAB, oltre al problema delle imprese in via di costituzione, vi sono due ulteriori problemi. In primo luogo, non tutte le imprese adempiono all'obbligo di depositare i bilanci. Inoltre, l'archivio è stato costruito retroattivamente e presenta forti limiti di copertura per gli anni precedenti al 2000.

I problemi sopra descritti costringono ad eliminare dall'analisi un certo numero di imprese, per le quali non risultano disponibili i dati necessari. Per la stima dell'effetto del I Bando sull'occupazione, si possono utilizzare informazioni solo su 110 imprese delle 149 agevolate. La carenza di osservazioni è molto più evidente per i gruppi di confronto. Per quanto riguarda il gruppo delle imprese la cui pratica è stata archiviata per mancanza di documentazione (G1) si perde completamente l'informazione, mentre

per gli altri tre gruppi di controllo la numerosità dei campioni è rispettivamente di 14, 24 e 10 imprese. La disponibilità dei dati sul fatturato è ancora minore: si hanno informazioni solo su 82 imprese agevolate e su un esiguo numero di imprese di controllo (12 per il gruppo G1 e 24 per il gruppo G3).

Considerazioni simili valgono per il II Bando: la numerosità del gruppo di imprese effettivamente agevolate è di 48 unità per la valutazione dell'effetto sull'occupazione, mentre per la valutazione sul fatturato si dispone di 54 osservazioni. Riguardo al gruppo di confronto, gli autori possono utilizzare rispettivamente 27 e 35 osservazioni. La scarsa numerosità dei gruppi utilizzati costituisce un forte vincolo all'analisi: l'impossibilità di tenere in debito conto la variabilità campionaria delle variabili-risultato comporta infatti il rischio di ottenere stime statisticamente non significative.

8.4 I risultati della valutazione

Le stime relative al I Bando

Le stime dell'effetto della legge 15 sul numero di dipendenti, ottenute con il metodo differenza-nelle-differenze (ricorrendo ai dati sul 1996), sono presentate nella Tabella 8.1.

Tabella 8.1: Effetto sull'occupazione stimato con il metodo differenza-nelle-differenze

Imprese trattate disponibili = 110	Gruppo di confronto			
	G1 Archivate	G2 Esito negativo	G3 Rinunciarie	G4 Revocate
Imprese non-trattate disponibili	Nessuna	14	24	10
Stime dell'effetto sul numero di occupati	--	+ 0,83	- 0,60	- 1,42

Nota: I valori in grassetto indicano stime statisticamente significative al 95% di confidenza
Fonte: Mura A., Lobascio I., Persico C. (2005)

L'effetto stimato è positivo nel caso in cui si ricorra al gruppo di controllo costituito dalle imprese escluse per via di un esito negativo dell'istruttoria (G2), mentre negli altri due casi esso è negativo. Tuttavia, in nessuno dei casi la stima ottenuta è statisticamente significativa, per cui non è possibile affermare che l'effetto sia diverso da zero. Come

ipotizzato in precedenza, la scarsa numerosità dei gruppi coinvolti non permette di ottenere stime molto precise e conduce a un'inevitabile indeterminatezza dei risultati.

Anche per quanto riguarda il fatturato, il confronto tra imprese agevolate e imprese non agevolate non evidenzia risultati significativi. Le stime differenza-nelle-differenze (in questo caso ottenute senza ricorrere ai dati sul 1996) sono presentate nella Tabella 8.2. Le stime dell'effetto sono ancora negative, tuttavia la scarsa numerosità campionaria rende anche in questo caso i risultati stimati non significativi. In sintesi, anche rispetto al valore della produzione non è possibile individuare alcun effetto delle agevolazioni.

Tabella 8.2: Effetto sul fatturato stimato con il metodo differenza-nelle-differenze

Imprese trattate disponibili = 82	Gruppo di confronto			
	G1 Archivate	G2 Esito negativo	G3 Rinunciarie	G4 Revocate
Imprese non-trattate disponibili	12	Nessuna	24	Nessuna
Stime dell'effetto sul fatturato (in euro)	- 2.508	--	- 1.967	--

Nota: I valori in grassetto indicano stime statisticamente significative al 95% di confidenza
Fonte: Mura A., Lobascio I., Persico C. (2005)

Le stime relative al II Bando

Potendo sfruttare per il II Bando anche i dati sul punteggio ricevuto dalle imprese al momento della richiesta di finanziamento, gli autori sono in grado di produrre rispetto al I Bando una maggiore varietà di stime, alcune delle quali sono illustrate qui di seguito.

Come abbiamo visto, le imprese escluse dal II Bando, ma agevolate attraverso altre fonti sono da considerare come trattate: poiché tuttavia per alcune di queste le agevolazioni sono avvenute in precedenza (ad esempio, nel 1999 invece che nel 2000 per le imprese agevolate già dal I Bando), per rendere confrontabili le variazioni nelle variabili-risultato viene operato un riallineamento temporale delle osservazioni. In questo modo, le variazioni vengono osservate per tutti a partire da un "anno zero" in cui hanno ricevuto le agevolazioni. Le variabili-risultato sono poi osservate a uno, due e tre

anni di distanza, termine entro il quale l'investimento per cui sono stati erogati gli incentivi dovrebbe essere stato concluso.

Le stime del modello *regression-discontinuity design* sono ottenute per questo bando ricorrendo a una variabile-trattamento continua, rappresentante l'ammontare dell'erogazione concessa alle imprese richiedenti. La stima d'effetto perciò non va più letta come semplice differenza media tra trattati e non-trattati, ma come misura della crescita della variabile-risultato commisurata alla crescita dell'importo erogato. Nella Tabella 8.3 sono riassunti i risultati delle stime dell'effetto sul numero di dipendenti.

Tabella 8.3: Effetto sull'occupazione stimato con il *regression-discontinuity design*

	Alla fine del primo anno	Alla fine del secondo anno	Alla fine del terzo anno
Stime dell'effetto sul numero di occupati (ogni 100.000 euro di agevolazioni)	0,429	0,038	1,847

Nota: I valori in grassetto indicano stime statisticamente significative al 95% di confidenza

Fonte: Mura A., Lobascio I., Persico C. (2005)

L'effetto stimato è decisamente variabile a seconda del tempo trascorso tra il momento dell'erogazione e l'osservazione dei risultati: è praticamente zero a due anni di distanza, mentre nelle altre due occasioni esso sembra aumentare il numero di dipendenti rispettivamente di una unità ogni circa 200.000 euro erogati e di una unità ogni 50.000 euro. Tuttavia, le stime non sono statisticamente significative.

Le stesse conclusioni si traggono per quanto riguarda l'effetto sul fatturato delle imprese, i cui risultati sono sintetizzati nella Tabella 8.4. Le stime seguono un andamento simile a quello visto per l'occupazione: l'effetto sembra essere maggiore alla fine dell'investimento (per cui si stima un aumento del fatturato di un euro circa per ogni euro erogato), mentre a un anno dall'erogazione degli incentivi esso risulta più contenuto, ed è decisamente più basso al secondo anno (addirittura negativo). Ma sia nel caso della stima negativa al secondo anno che nei casi di stima positiva per le altre due occasioni, la variabilità dei risultati è troppo alta, per cui il vero effetto potrebbe essere tanto positivo quanto negativo. In conclusione, anche in questo caso non si può dire nulla sui reali effetti della legge 15.

Tabella 8.4: Effetto sul fatturato stimato con il *regression-discontinuity design*

	Alla fine del primo anno	Alla fine del secondo anno	Alla fine del terzo anno
Stime dell'effetto sul fatturato	0,686	-0,778	1,171

Nota: I valori in grassetto indicano stime statisticamente significative al 95% di confidenza
Fonte: Mura A., Lobascio I., Persico C., (2005)

Con i dati a disposizione non si è quindi in grado di evidenziare con chiarezza un effetto dei contributi della legge 15 sulla performance delle imprese. Le stime prodotte, per entrambi i bandi considerati, non sono significativamente diverse da zero. L'impossibilità di trarre conclusioni significative riguarda tutti i tentativi di analisi, anche quelli non riprodotti qui, nei quali gli autori hanno provato a variare sia le metodologie utilizzate sia i possibili gruppi di confronto. Si rimanda al testo del rapporto per un'illustrazione esaustiva dei risultati ottenuti.

CAPITOLO 9

GLI EFFETTI DELLA LEGGE 27/93 DELLA REGIONE TOSCANA SULL'IMPRENDITORIA GIOVANILE

di Marco Sisti

A partire dalla seconda metà degli anni Ottanta, sono stati adottati numerosi programmi d'intervento pubblico rivolti a giovani che intendono avviare un'attività imprenditoriale. Gli strumenti di sostegno all'imprenditorialità giovanile previsti nell'ambito di tali programmi variano sia nella natura che nell'intensità dell'aiuto offerto: si va dalla concessione di "prestiti d'onore", alla fornitura diretta di servizi di tutoraggio ed assistenza; dall'assegnazione di agevolazioni fiscali e/o creditizie, in alcuni casi anche di cospicua entità, fino all'erogazione gratuita di corsi di formazione in economia aziendale e management.

Recentemente a questi strumenti più tradizionali si sono aggiunti i cosiddetti "incubatori d'impresa" che, oltre ad offrire un pacchetto variabile di servizi di consulenza, formazione e finanza, mettono a disposizione degli aspiranti imprenditori anche spazi fisici e strutture logistiche condivise.

Al di là delle forti differenze che li contraddistinguono, tutti questi strumenti sono accomunati da una medesima finalità di fondo: stimolare la nascita di "imprese giovani" ed aiutarle a superare i diversi ostacoli della delicata fase d'avvio delle loro attività.

L'esperienza di valutazione illustrata in questo capitolo ha per oggetto una politica regionale, attuata in Toscana fin dal 1993, che si pone esattamente questo scopo: sostenere l'imprenditoria giovanile (più precisamente le imprese costituite da meno di 6 mesi, da persone con età inferiore ai 35 anni), attraverso la concessione di agevolazioni al credito e di garanzie sussidiarie su operazioni di finanziamento.

Gli studi che, a partire dal 1999, si sono occupati di questo intervento regionale hanno cercato di dare risposta a domande di valutazione relative a differenti aspetti della politica. In questa sede focalizzeremo l'attenzione su due domande di natura controfattuale, al centro di un corposo lavoro pubblicato dall'IRPET (Istituto Regionale

per la Programmazione Economica della Toscana) nel 2001⁵⁵ e curato da Fabrizia Mealli e Roberto Pagni. La prima domanda indaga l'effetto degli aiuti regionali su una particolare dimensione della performance aziendale delle imprese agevolate, ovvero la loro capacità di sopravvivenza. In altre parole l'analisi cerca di capire se, e in che misura, le agevolazioni e le garanzie concesse dalla Regione sono riuscite ad aumentare le probabilità di sopravvivenza delle imprese neonate condotte da giovani. Su questo specifico aspetto prenderemo in esame altri due studi, uno precedente⁵⁶ e l'altro successivo⁵⁷ alla valutazione condotta dall'IRPET. Il confronto tra questi tre lavori ci consentirà di comprendere come i diversi ricercatori che si sono susseguiti nello studio della politica, in possesso di una mole sempre maggiore di informazioni, abbiano potuto rendere più sofisticato il disegno di valutazione.

La seconda domanda (affrontata in modo soltanto marginale all'interno della ricerca dell'IRPET) mira a comprendere se, e in che misura, i giovani imprenditori siano stati influenzati dall'esistenza dell'agevolazione regionale nella loro decisione di avviare un'attività imprenditoriale.

9.1 La logica della politica toscana di aiuto all'imprenditoria giovanile

La politica toscana di aiuto all'imprenditoria giovanile è stata formalmente istituita dalla legge regionale 26 aprile 1993 n. 27, "Agevolazioni per la creazione di nuove imprese a sostegno dell'imprenditoria giovanile". La legge in questione prevede l'erogazione di contributi finanziari, nonché l'eventuale concessione di garanzie sussidiarie a favore di imprese, costituite da non più di sei mesi da giovani di età non superiore ai 35 anni e che realizzino investimenti all'interno del territorio regionale. Dispone, inoltre, che siano realizzati interventi di formazione a favore dei giovani imprenditori che ne facciano richiesta. La finalità dichiarata esplicitamente all'interno della legge consiste

⁵⁵ Mealli F., Pagni R., *Analisi e valutazione delle politiche per le nuove imprese – Il caso della L.R. Toscana n. 27/93*, FrancoAngeli, Milano, 2001.

⁵⁶ Bondonio D., Martini A., *Using Event History Analysis to Evaluate the Impact of Investment Subsidies Targeted to Youth-Owned Firms*, in Società Italiana di Statistica, Atti del Convegno "Processi e metodi statistici di valutazione", Roma 4-6 giugno 2001.

⁵⁷ Conforti D. (2004), *Metodi statistici per l'analisi di politiche attive per il lavoro: la legge regionale della Toscana n. 27/93*, Tesi di Dottorato in Statistica Applicata, Università degli Studi di Firenze.

nel “*promuovere la costituzione e l’avvio di nuove imprese al fine di sviluppare l’imprenditoria giovanile*”.

La logica che sta dietro a questo tipo di politica è, nei suoi tratti essenziali, piuttosto semplice. L’intervento si basa sull’assunto che le persone più giovani, che intendono intraprendere un’attività imprenditoriale, incontrino delle barriere (in termini di scarsa disponibilità di capitale iniziale necessario all’avvio dell’impresa, di particolari difficoltà di accesso al mercato del credito e di mancanza di formazione specifica) che ne scoraggiano l’iniziativa. Questa situazione impedisce di fatto, o comunque rende assai difficile, la nascita di nuove imprese condotte da giovani.

A questo assunto di fondo si affianca l’ipotesi che i giovani, che potrebbero trarre beneficio da un eventuale abbattimento, o da una parziale riduzione, di tali barriere, hanno comunque la capacità, se messi nelle giuste condizioni favorevoli, di dar vita a progetti imprenditoriali validi e competitivi e, quindi, di creare imprese tanto solide quanto quelle già presenti nel mercato.

Dunque, seguendo la logica della politica, sebbene i giovani costituiscano una categoria “svantaggiata” tra gli aspiranti imprenditori, a causa delle elevate barriere esistenti in entrata, essi sono nello stesso tempo portatori di opportunità non sfruttate e di potenzialità inesprese. Di conseguenza l’esistenza di barriere, o disincentivi “naturali”, alla libera iniziativa giovanile fa sì che il numero di imprese costituite da giovani sia di molto inferiore a quanto il decisore pubblico (in questo caso l’amministrazione regionale) ritiene desiderabile. Da ciò discende la decisione di incentivare la nascita di imprese “giovani” attraverso l’impiego di strumenti *ad hoc*, per supplire a ciò che possiamo considerare una sorta di “fallimento del mercato”.

Per poter affermare che una politica di questo tipo, una volta adottata, abbia avuto successo, occorre verificare che si siano realizzati due eventi distinti. In primo luogo, l’intervento deve essere stato capace di stimolare la nascita di un numero di imprese “giovani” superiore a quanto sarebbe avvenuto se l’intervento non fosse esistito. Occorre cioè verificare che gli aiuti concessi abbiano davvero indotto i giovani con idee e volontà imprenditoriali a mettersi in proprio e a costituire impresa. In secondo luogo, le imprese, create da giovani che si sono avvalsi degli aiuti regionali per superare le

difficoltà presenti nella fase di *start up* della loro attività, devono mostrare performance aziendali (in termini di sopravvivenza sul mercato, incremento occupazionale, aumento del fatturato, ecc.) perlomeno simili a quelle ottenute dalle imprese nate spontaneamente.

Se così non fosse, le risorse pubbliche utilizzate per dar vita a questa politica finanzierebbero la nascita di imprese fallimentari in partenza. In questo sfortunato caso, gli aiuti regionali concessi ad imprese strutturalmente deboli, piuttosto che contrastare un'anomalia indesiderata del mercato, agirebbero come meccanismo di disinnesco dei virtuosi processi selettivi operanti "in natura". Tali processi selettivi, lasciati agire liberamente, avrebbero infatti il pregio di riconoscere e premiare le idee e le capacità imprenditoriali migliori, quelle cioè con maggiori probabilità di restare con successo sul mercato.

Prima di esaminare più in dettaglio le domande che stanno alla base degli studi di valutazione condotti su questa politica, descriviamo brevemente quali sono le specifiche modalità procedurali e i concreti strumenti d'aiuto adottati dalla Regione Toscana per intervenire a sostegno dell'imprenditoria giovanile.

Iniziamo con la descrizione delle caratteristiche della popolazione *target* dell'intervento, ovvero dei requisiti d'ammissibilità imposti dalla Regione. Sono riconosciute come possibili beneficiarie delle agevolazioni regionali tutte le imprese con meno di 95 dipendenti e con capitale sociale inferiore a 5 miliardi di lire. Come già sottolineato, tali imprese non possono avere più di sei mesi di vita alla data di richiesta del contributo e il titolare (se si tratta di imprese individuali) deve avere meno di 35 anni di età. Per le società tale requisito deve essere posseduto dal legale rappresentante, nonché da almeno la metà dei soci che detengano come minimo il 51% del capitale sociale o dei voti nell'assemblea dei soci. Per quanto riguarda invece la natura degli aiuti concessi, la legge prevede l'erogazione di contributi in conto capitale per un massimo del 10% dell'investimento e in conto interessi per finanziamenti fino ad un massimo del 65%

dell'investimento previsto, inizialmente con un abbattimento della quota interessi fino ad un massimo di 6 punti percentuali⁵⁸.

Un ruolo centrale nell'attuazione della politica è giocato da FidiToscana Spa, una società finanziaria che conta tra i suoi azionisti numerosi enti pubblici locali (tra i quali la stessa Regione Toscana) e diverse banche operanti sul territorio regionale.

FidiToscana è stata incaricata dalla Regione Toscana dello svolgimento dell'istruttoria delle domande presentate dalle imprese, della stesura delle graduatorie delle imprese ammesse a finanziamento, nonché dell'erogazione dei contributi concessi. Inoltre, essa fornisce alle stesse imprese richiedenti le eventuali garanzie sussidiarie, necessarie per l'accesso al credito bancario, e cura le pratiche per il recupero dei finanziamenti concessi a soggetti che abbiano successivamente perduto i requisiti d'ammissibilità.

Di seguito riportiamo una descrizione sintetica dell'iter amministrativo che scandisce l'attuazione della politica e il ruolo svolto dai singoli attori coinvolti:

- la Regione Toscana stabilisce i requisiti d'ammissibilità, le spese d'investimento ammissibili e i valori massimi raggiungibili dagli aiuti regionali;
- l'impresa neonata, condotta da uno o più giovani sotto i 35 anni, entro sei mesi dalla sua costituzione presenta domanda di finanziamento presso gli uffici della FidiToscana;
- FidiToscana svolge l'istruttoria; ogni tre mesi pubblica una graduatoria delle imprese ammesse ed invia alla banca di competenza una relazione scritta, che accompagna la decisione d'ammissione e con la quale eventualmente concede garanzie sussidiarie;
- l'istituto di credito perfeziona il prestito all'impresa beneficiaria del contributo regionale e ne dà comunicazione a FidiToscana che finalmente eroga l'agevolazione regionale a favore di quell'impresa.

⁵⁸ Nel tempo la quota di abbattimento degli oneri per interessi è stata progressivamente ridotta, fino a raggiungere il valore di un solo punto percentuale. Tale riduzione è stata decisa al fine di adeguare il contributo regionale al forte ribasso del tasso d'interesse di mercato, registrato a partire dalla seconda metà degli anni Novanta.

9.2 Le domande di valutazione

Per una serie di coincidenze, a partire dalla fine degli anni Novanta, diversi enti di ricerca sono stati impegnati nella realizzazione di studi di valutazione che avevano lo scopo di analizzare il processo d'attuazione e i risultati di questa particolare politica regionale.

Il primo studio, in ordine di tempo, nasceva da un progetto del Consiglio regionale della Toscana, che aveva il fine di sperimentare il possibile utilizzo di metodi di valutazione *ex post* all'interno delle assemblee legislative⁵⁹. Qualche tempo dopo l'avvio di questo primo studio, l'IRPET ha ricevuto, sempre dal Consiglio regionale, l'incarico di svolgere un'indagine sui diversi strumenti d'aiuto utilizzati dalla Regione per sostenere la nascita di nuove imprese. In questa occasione, anche dopo alcuni scambi con gli autori del lavoro precedente, nasceva l'idea di usare la politica istituita con la legge 27 come principale banco di prova per testare l'applicazione di alcuni differenti approcci valutativi, tra i quali assumeva particolare rilevanza quello finalizzato a stimare gli effetti ottenuti dall'intervento regionale. L'indagine svolta dall'IRPET si concludeva sottolineando la necessità di proseguire il lavoro di valutazione degli effetti nel corso degli anni successivi; questo al fine di eliminare i limiti più evidenti dei precedenti studi e di consolidare i risultati analitici ottenuti. Il testimone è stato raccolto da un giovane ricercatore, che ha fatto di questa politica regionale il principale oggetto della sua tesi di dottorato, discussa all'inizio del 2005.

Sebbene i tre studi siano nati da motivazioni e in contesti differenti, hanno tutti quanti affrontato una medesima domanda di valutazione, ispirata alla logica controfattuale, e possono essere letti a posteriori come tentativi successivi di superare i limiti metodologici ed informativi insiti negli studi precedenti.

Gli effetti sulla sopravvivenza delle imprese

Quando si intendono valutare gli effetti di una politica pubblica, uno dei primi problemi che ci si pone riguarda la scelta della variabile-risultato da tenere sotto osservazione.

⁵⁹ In seguito, a partire dal 2001, tale progetto si è evoluto ed ha assunto una dimensione interregionale con la denominazione di *CAPiRe* (Controllo delle Assemblee sulle Politiche e gli Interventi Regionali). Attualmente partecipano al progetto anche i Consigli Regionali dell'Emilia-Romagna, della Lombardia e del Piemonte. Per ulteriori informazioni si veda: www.capire.org.

Nel caso di politiche di aiuto alle imprese, sembra ragionevole aspettarsi che, in seguito alla concessione dei contributi pubblici, le imprese beneficiarie abbiano migliorato la loro performance aziendale o, nel caso di una congiuntura economica particolarmente sfavorevole, siano almeno riuscite a limitare i danni. La performance di un'impresa è però un concetto multi-dimensionale, soggetto cioè ad essere rappresentato, anche quantitativamente, attraverso l'uso di diverse variabili. Si può dire infatti che un'impresa ha registrato una buona performance, quando si osserva la presenza di diversi indizi positivi: ad esempio quando incrementa il suo fatturato, oppure quando aumenta il numero di persone che lavorano al suo interno, o ancora quando cresce la redditività del capitale investito, o la quota di produzione esportata all'estero. In alcuni casi, in particolare quando i beneficiari della politica di aiuto sono imprese costituite da pochi mesi, prive quindi di una lunga esperienza aziendale e, probabilmente, con pochi margini di crescita nel breve periodo su altre dimensioni (fatturato, occupazione, produttività, ecc.), la performance può essere misurata semplicemente osservando la capacità della stessa azienda di sopravvivere sul mercato.

Naturalmente si tratta di una definizione di performance ridotta ai minimi termini, che rischia cioè di non cogliere l'intero insieme di effetti che l'agevolazione può (o nelle intenzioni dei *policy maker* dovrebbe) provocare sulle attività aziendali. Nondimeno, nel caso di politiche indirizzate a sostenere imprese giovani nella critica fase di avvio delle loro attività (politiche quindi che hanno come target imprese ad elevato rischio di cessazione), concentrare l'attenzione sulla loro capacità di resistere sul mercato appare una scelta certamente ragionevole ed appropriata⁶⁰.

La domanda di valutazione, al centro dei tre studi considerati in questo capitolo, può essere dunque così tradotta: *“in che misura le agevolazioni regionali hanno aumentato la probabilità di sopravvivenza delle imprese beneficiarie, rispetto a quanto sarebbe accaduto se tali agevolazioni non fossero state loro concesse?”*.

⁶⁰ Vi sono stati altri studi su politiche adottate in Italia a sostegno all'imprenditoria giovanile, che si sono concentrati sulla probabilità di sopravvivenza delle imprese. A questo proposito si veda l'analisi della legge 44 svolta in Battistin E., Gavosto A., Rettore E., “Why do subsidised firms survive longer? An evaluation of a program promoting youth entrepreneurship in Italy”, in Lechner M., Pfeiffer (a cura di), *Econometric Evaluation of Active Labour Market Policies*, Physica/Springer-Verlag, Heidelberg, 2001.

Le strategie di analisi adottate nei tre studi per rispondere a tale domanda di valutazione sono molto simili e si basano tutte sull'idea di confrontare la durata in vita delle imprese beneficiarie con quella osservata su un gruppo di imprese non beneficiarie. La differenza tra le durate in vita di questi due gruppi di imprese, sotto certe condizioni e con alcune cautele, può essere interpretata come l'effetto della politica di aiuto sulla capacità di sopravvivenza.

Il lavoro dell'IRPET non si ferma però al tentativo di dare una risposta alla domanda sulla sopravvivenza delle imprese; sebbene in misura sfumata, tale lavoro indaga anche la capacità della politica di incentivare la nascita di nuove imprese condotte da giovani. Poiché si tratta di una domanda di valutazione assai rilevante, in quanto la sua risposta fornisce un ulteriore elemento di giudizio sul successo della politica, sarà anche questa oggetto di trattazione in questo capitolo.

Gli effetti sulla scelta di “mettersi in proprio”

Il sostegno dato dalla Regione alle imprese non ha soltanto la finalità di migliorare la loro performance aziendale o, più nello specifico, di aumentarne le probabilità di sopravvivenza. La bassa natalità nelle imprese condotte da giovani e il progressivo processo d'invecchiamento nell'imprenditoria toscana sembrano essere due dei principali fenomeni che hanno motivato l'adozione di questa politica d'aiuto. Rendendo più facile, o relativamente meno costoso, per i giovani “mettersi in proprio” si intende incentivarne l'iniziativa imprenditoriale e spostare le decisioni di coloro che sono incerti o dubbiosi.

In questo caso la domanda può essere espressa nei seguenti termini: *“in che misura l'esistenza dei contributi regionali ha influenzato i giovani imprenditori nella loro scelta di costituire una nuova impresa?”*. Rispondendo a questa domanda si intende scoprire se i contributi regionali hanno spinto i giovani indecisi a “mettersi in proprio”, oppure hanno semplicemente sortito l'effetto di abbassare i costi di *start up* per imprese che sarebbero nate comunque, anche in assenza di un aiuto da parte della Regione. Si tratta di un caso particolare della categoria più generale di quesiti valutativi volti ad

indagare l'*addizionalità* prodotta dalla politica⁶¹ (ovvero il numero di imprese giovani, tra quelle sovvenzionate, nate grazie all'esistenza del contributo). La ricerca dell'IRPET ha tentato di rispondere a questa domanda, conducendo un'indagine diretta su un campione di imprese beneficiarie dei contributi. Data l'impossibilità di disporre di un gruppo di confronto che consentisse di ricostruire la situazione controfattuale, si è chiesto direttamente agli imprenditori "agevolati" di stabilire in che misura l'esistenza del contributo regionale abbia pesato sulla loro scelta di avviare l'impresa.

9.3 Le principali questioni metodologiche e le soluzioni adottate

La scelta di una particolare strategia di valutazione dipende essenzialmente dal tipo di domanda a cui si intende rispondere e dalla disponibilità delle informazioni esistenti (o dalla concreta possibilità di rilevarne di nuove). In questo paragrafo ci proponiamo di mettere in evidenza le motivazioni alla base delle scelte metodologiche adottate nei diversi studi. Iniziamo con le soluzioni messe in campo per rispondere alla prima domanda di valutazione, relativa agli effetti delle agevolazioni sulle probabilità di sopravvivenza delle imprese.

Un semplice confronto tra imprese beneficiarie e non beneficiarie

Una prima strategia per rispondere ad una domanda del tipo "*in che misura le agevolazioni regionali hanno aumentato la probabilità di sopravvivenza delle imprese beneficiarie, rispetto a quanto sarebbe accaduto se tali agevolazioni non fossero state loro concesse?*", consiste nel confrontare il tasso di mortalità delle imprese beneficiarie dei contributi regionali nate in un certo periodo con il tasso di mortalità di imprese *non* beneficiarie nate nello stesso periodo. Il dato relativo alle imprese non beneficiarie in questo caso viene assunto come una stima della situazione controfattuale. La differenza tra i valori assunti dai due tassi di mortalità è attribuita al fatto di aver ricevuto il contributo regionale.

Il disegno di questa analisi è molto semplice: in un determinato momento X nel tempo si contano quante imprese sono nate in un periodo di durata definita, ad esempio

⁶¹ Cfr. par. 7.2.

nell'anno precedente, e quante tra quelle stesse imprese, nello stesso periodo, sono cessate. Il rapporto tra questi due numeri ci indica il tasso di mortalità delle imprese considerate.

Si tratta di una versione molto semplificata di una tipologia di analisi, detta “di durata”, impiegata soprattutto in campo medico per studiare i tempi di sopravvivenza degli individui dopo un intervento chirurgico oppure dopo la diagnosi di una grave malattia. La stessa logica può comunque essere applicata ad altri campi d'interesse, ad esempio per studiare la durata della disoccupazione in particolari categorie di soggetti, o la durata delle unioni matrimoniali, o ancora la durata della condizione di libertà negli ex-detenuti. In ogni caso ciò che si osserva è l'intervallo di tempo che un soggetto trascorre in un particolare *status* (come essere disoccupato, o essere sposato, oppure essere fuori dal carcere). Poiché l'osservazione avviene in un determinato istante nel tempo, le informazioni disponibili sulle unità osservate (che siano disoccupati, matrimoni, ex-detenuti o, come in questo caso, imprese) si limitano a registrare se tali unità hanno mantenuto o meno lo *status* d'interesse, fino a quel particolare istante; nulla ci dicono su quanto accade dopo quel momento. Si parla perciò di “dati censurati a destra”, in quanto essi ci informano del fatto che la durata del fenomeno posto sotto osservazione è maggiore, per alcune delle unità osservate, nel periodo considerato, ma non ci consentono di stabilire di quanto essa sia maggiore.

Per realizzare il confronto sopra indicato occorre avere almeno quattro quantità:

1. il numero di imprese beneficiarie nate nel periodo di osservazione;
2. il numero di imprese beneficiarie (di cui al punto 1) cessate nel periodo di osservazione;
3. il numero di imprese nate nello stesso periodo che, pur avendone i requisiti, non hanno ricevuto alcun contributo;
4. il numero di imprese (di cui al punto 3) cessate nel periodo di osservazione.

Il rapporto tra il secondo e il primo numero determina il valore del tasso di mortalità delle imprese che hanno ricevuto il finanziamento, in un certo periodo di tempo. Il rapporto tra il quarto e il terzo numero lo determina invece per le imprese che, pur

avendone i requisiti, non hanno ricevuto alcun contributo. Sul confronto tra questi due valori si basa il semplice disegno di valutazione.

Il valutatore deve, quindi, individuare le fonti informative che permettano di ricostruire la storia delle imprese, dalla loro nascita fino alla loro eventuale cessazione. Per implementare anche questo semplice disegno valutativo occorre infatti avere a disposizione delle banche dati che contengano informazioni sulle singole imprese (come minimo data di nascita e di cessazione e il fatto che abbiano ricevuto o meno il contributo regionale). In questo caso sono state impiegate due differenti fonti, entrambe di natura amministrativa:

- 1) l'archivio FidiToscana che raccoglie minime informazioni anagrafiche sulle imprese che hanno presentato domanda di finanziamento;
- 2) la banca-dati InfoCamere che raccoglie informazioni sulle imprese iscritte nel Registro delle imprese presso le singole Camere di Commercio della Toscana (almeno in teoria tale database dovrebbe contenere l'universo di imprese esistenti in Regione).

Il primo archivio viene usato da FidiToscana per gestire le pratiche di ammissione ai finanziamenti regionali e le procedure di pagamento; ai fini dell'analisi valutativa, esso serve per individuare quali sono le imprese beneficiarie del contributo. Il secondo archivio, che contiene la data d'avvio attività e l'eventuale data di cessazione di tutte le imprese toscane, consente invece di calcolare la durata della loro sopravvivenza.

Il primo studio che ha affrontato la valutazione di questa politica ha utilizzato le banche dati InfoCamere con riferimento alle imprese nate in Toscana in un periodo che va dal 1° gennaio 1994 fino al 31 dicembre 1998. Per ogni impresa si è in possesso delle seguenti informazioni: (i) data di inizio dell'attività (usata come approssimazione della data di costituzione); (ii) settore di appartenenza; (iii) Provincia in cui essa è localizzata; (iv) eventuale data di cessazione; (v) numero di addetti; ed, infine, (vi) sesso ed (vii) età dell'imprenditore. Per ogni impresa, grazie alle informazioni derivanti dall'archivio FidiToscana, si è poi in grado di sapere se è stata ammessa o meno al contributo della legge 27, nonché l'ammontare del contributo erogato. L'analisi si limita alle sole ditte individuali poiché soltanto per tale categoria si è in grado di avere informazioni sul

sesso e sull'età dell'imprenditore⁶², caratteristiche discriminanti per poter presentare domanda. Sono perciò escluse dall'analisi tutte le imprese che operano in forma societaria. Il lavoro dell'IRPET, realizzato a qualche mese di distanza dal primo studio considerato, utilizza le stesse fonti informative (InfoCamere e FidiToscana), ma restringe l'analisi alle imprese nate dal 1° gennaio 1996 al 31 dicembre 1998.

La Tabella 9.1 riporta alcune statistiche descrittive contenute negli archivi di InfoCamere. Le imprese beneficiarie delle legge sono una piccola percentuale di tutte le imprese create da giovani imprenditori tra il 1994 e il 1998, di cui rappresentano soltanto l'1,8%. La distribuzione per età sembra essere simile sia per le imprese con contributo che per quelle senza, con una media ed una percentuale di imprenditori con età inferiore ai 30 anni molto simili.

Tabella 9.1: Descrizione delle imprese individuali nate nel periodo 94-99

	Imprese giovani finanziate dalla legge 27	Imprese giovani non finanziate dalla legge 27	Imprese costituite da non giovani (over 35)
Numero di imprese	706	39.192	26.699
Età media dell'imprenditore	27,2	27,8	45,6
% di imprenditori di 30 anni o più giovani	68.3	74.7	-
Tasso di mortalità	9,3%	25,3%	20,6%

Fonte: Bondonio D., Martini A. (2001)

Il tasso di mortalità dei tre gruppi di imprese è riportato nell'ultima riga: soltanto il 9,3% delle imprese della legge 27 ha cessato la propria attività tra il 1994 e il 1999, mentre il valore sale fino al 20,6% per le imprese costituite da imprenditori più anziani e fino al 25,3% per le imprese giovani non beneficiarie del contributo. La differenza tra il valore osservato per le imprese giovani beneficiarie e quello osservato per le imprese giovani che non hanno ricevuto il contributo (ma che avrebbero avuto i requisiti per farne richiesta) è dunque pari a circa 16 punti percentuali. Questa differenza è una stima plausibile dell'effetto dell'agevolazione sulla capacità di sopravvivenza delle imprese? È possibile affermare che il contributo ha avuto un effetto positivo sulla sopravvivenza

⁶² Il sesso e l'età dell'imprenditore sono informazioni ottenute attraverso la decifrazione del codice fiscale del titolare che fortunatamente era uno dei campi inseriti nell'archivio InfoCamere.

delle imprese, facendo ridurre dal 25,3 al 9,3% il tasso di mortalità delle imprese giovani?

Il lettore più accorto avrà già anticipato la risposta: interpretare questa differenza come effetto del contributo regionale è un'operazione molto azzardata. Tale interpretazione si basa infatti sull'ipotesi che i due gruppi di imprese siano simili per una serie di caratteristiche (come il settore produttivo, la localizzazione territoriale, la capacità imprenditoriale ed altre), che potrebbero ragionevolmente incidere sulla loro capacità di sopravvivenza. Solo prendendo per buona tale ipotesi di somiglianza (ovvero assumendo uguali condizioni di partenza), si può attribuire la causa della differenza osservata tra i tassi di mortalità dei due gruppi al fatto di aver ricevuto, o meno, il contributo regionale. Le informazioni riportate in tabella non consentono di accertare la veridicità di tale ipotesi, ma il buon senso suggerisce di dubitarne. È assai probabile infatti che i due gruppi si siano selezionati, o siano stati selezionati, in quanto erano in possesso di qualità diverse. Se si intende giungere ad un disegno di valutazione più credibile, è necessario perciò complicare ulteriormente l'analisi, prendendo in considerazione quei fattori (almeno quelli osservabili) che sembrano influenzare il fenomeno della mortalità delle imprese.

Prima di procedere, occorre aggiungere un'ulteriore notazione rispetto alla variabile-risultato scelta. Usare semplicemente il tasso di mortalità, cioè contare il numero di cessazioni in un determinato momento nel tempo, può essere fuorviante. Infatti anche se i due gruppi presentano un eguale tasso di mortalità, ipotizziamo il 15% dopo quattro anni, non è detto che non vi siano differenze nella loro capacità di sopravvivenza. Può accadere, ad esempio, che le cessazioni osservate nel gruppo delle non beneficiarie siano avvenute già a partire dal primo anno; mentre le cessazioni osservate nel gruppo delle beneficiarie siano avvenute a partire soltanto dal quarto anno. In questo caso, a parità del valore assunto dal tasso di mortalità (15%), i tempi di sopravvivenza delle imprese nei due gruppi sarebbero ben diversi: le aziende beneficiarie dimostrerebbero, nei primi tre anni di attività, una capacità di resistenza ben maggiore rispetto alle altre. I disegni di valutazione illustrati nei paragrafi successivi incorporano questa consapevolezza confrontando l'effettiva durata in vita delle imprese e tenendo perciò

conto della loro data di costituzione. In questa impostazione, più corta è la durata osservata in un gruppo, più alto è il rischio di fallimento.

L'uso del modello a rischi proporzionali di Cox

Sia nel primo che nel secondo studio si sfruttano le informazioni contenute nell'archivio InfoCamere per verificare la misura in cui alcuni fattori esterni all'intervento condizionano la sopravvivenza delle imprese. Per far ciò, entrambi i lavori effettuano un'analisi multivariata, adottando un particolare modello di regressione, conosciuto con il nome di *modello a rischi proporzionali di Cox*. Tale modello permette di determinare l'influenza di una serie di variabili osservate (ovvero, secondo la terminologia usata nel capitolo 5, delle variabili di controllo) sulla probabilità di fallimento dell'impresa, specificando una funzione di rischio. In altri termini, esso consente di confrontare la diversa capacità di sopravvivenza dei due gruppi di imprese, a parità di caratteristiche osservabili.

Nella costruzione del modello i ricercatori possono compiere scelte diverse rispetto alle variabili di controllo da inserire e nel modo in cui esse vengono misurate. Lo studio dell'IRPET ha impiegato quali variabili di controllo: l'anno di iscrizione dell'impresa, la Provincia di appartenenza, la localizzazione della sede in un Comune capoluogo di Provincia, il settore produttivo di riferimento (distinguendo per settore edile, industriale, servizi alle persone e servizi alle imprese), il genere e l'età dell'imprenditore, il numero di persone che lavorano nell'impresa.

L'uso del modello di Cox impone di compiere una precisa scelta riguardo alla forma funzionale con cui le variabili esplicative compaiono nel modello di regressione. Si suppone cioè di conoscere che tipo di relazione esista tra la variabile-risultato, in questo caso il rischio di fallimento, e le variabili di controllo impiegate. Per tentare di superare questo limite, lo studio dell'IRPET percorre altre strade, ad esempio impiega un metodo che i ricercatori definiscono del *matching esatto*.

L'uso del matching

Secondo questo metodo ad ogni impresa beneficiaria viene associata un'impresa gemella non beneficiaria, che presenta quindi caratteristiche perfettamente uguali alla

prima: per questo motivo si può parlare di *matching*, o abbinamento, esatto. Si tratta in realtà di un caso particolare dell'abbinamento sull'unità più vicina illustrato al capitolo 5. L'impiego di questa procedura è possibile solo nel caso in cui il numero dei non-trattati sia molto elevato rispetto a quello dei trattati: ciò rende infatti molto più probabile individuare nei due gruppi unità perfettamente simili. In altri casi meno fortunati, come si è visto nel capitolo 5, si possono seguire differenti procedure di *matching*.

Ad ogni impresa trattata corrisponde un'impresa non-trattata, che ha (i) lo stesso numero di addetti; (ii) appartiene alla stessa Provincia ed (iii) è collocata o meno nel capoluogo; (iv) è stata iscritta alla Camera di Commercio nello stesso anno; (v) opera nello stesso settore di attività; ha un titolare (vi) della stessa età e (vii) dello stesso sesso. Soltanto per 39 imprese, su un gruppo di 522 beneficiarie prese a riferimento nello studio, non è stato possibile individuare un'impresa identica e non finanziata; a ciascuna di queste imprese senza "gemelle naturali" i ricercatori hanno deciso di abbinare un'impresa con le stesse caratteristiche, ma collocata nella Provincia più vicina. Al termine di tale operazione si è ottenuto quindi un gruppo di controllo identico in termini di caratteristiche osservabili a quello delle aziende beneficiarie; ciò che si osserva per tale gruppo può costituire una approssimazione plausibile di ciò che si sarebbe osservato per le imprese beneficiarie se non avessero ricevuto il finanziamento della legge 27. La differenza dei tempi di sopravvivenza osservati nei due gruppi può essere perciò interpretata come effetto del contributo regionale.

Resta però ancora aperto un problema. Sia nel caso della regressione multipla, sia in quello del *matching*, la stima dell'effetto della politica sulla sopravvivenza delle imprese è tanto più plausibile quanto più si è disposti a credere che le variabili di controllo impiegate esauriscano le possibili differenze tra aziende beneficiarie e non beneficiarie. In realtà i due gruppi di imprese possono differire rispetto a molte altre caratteristiche, che non sono descritte dalle informazioni contenute nei database amministrativi di InfoCamere e che, quindi, non sono state utilizzate dai ricercatori nelle loro procedure di stima. In sostanza, esistono buone probabilità che, nonostante gli sforzi fatti, i due gruppi di imprese siano tutt'altro che simili: le aziende beneficiarie della legge 27 potrebbero essere condotte da imprenditori più preparati, più dinamici,

più competenti. Questo, e non il contributo regionale, potrebbe spiegare la differenza osservata in termini di sopravvivenza. Per risolvere questo problema l'IRPET nel 2001 ha deciso di realizzare un'indagine diretta su un campione di imprese nate in Toscana nel 1999, che rilevasse alcune delle informazioni “mancanti”, quali ad esempio il livello di studi o le esperienze lavorative dell'imprenditore. Ciò ha consentito di avere una base di informazioni molto più ricca che è servita, a qualche anno di distanza, per realizzare un confronto più plausibile tra la capacità di sopravvivenza delle imprese beneficiarie della legge 27 e quella delle imprese non beneficiarie.

L'indagine diretta sulle imprese nate nel 1999

Il limite delle valutazioni presentate fino a questo momento è la scarsità di informazioni relative a fattori che possono avere una forte incidenza sulla performance delle imprese e che potrebbero spiegare eventuali differenze riscontrate tra il gruppo delle imprese beneficiarie e quello delle imprese non beneficiarie. Il passo avanti fatto dall'IRPET è stato di realizzare un'indagine diretta su un campione di imprese, in modo tale da ottenere le informazioni assenti e ritenute (dai ricercatori) significative. Non solo questo ha consentito di estendere il numero di informazioni relative ad ogni singola impresa, ma ha permesso di realizzare un confronto su tutte le possibili tipologie di imprese beneficiarie (mentre i casi precedenti si limitavano a considerare le ditte individuali, le uniche per le quali fosse possibile stabilire sesso ed età dell'imprenditore).

L'indagine è stata realizzata con il metodo CATI (*Computer Assisted Telephone Interview*) nel gennaio del 2001. La scelta di intervistare un campione di imprese nate nel 1999 è dovuta al fatto che era l'anno più recente per il quale fosse disponibile l'archivio InfoCamere. L'esigenza di avere un campione nel quale entrassero anche le imprese con vita più corta imponeva infatti di effettuare le interviste alla distanza più breve possibile dalla nascita; più si aspettava a realizzare l'indagine e maggiore sarebbe stato il numero di imprese cessate (e quindi non più intervistabili). Grazie ai dati a disposizione i ricercatori sapevano che dopo un anno e mezzo (corrispondente alla distanza temporale media tra la nascita e il mese di svolgimento dell'indagine) più del 90% delle imprese risultava essere ancora in vita.

Per individuare le imprese da far rientrare nel campione si è scelta la tecnica del campionamento stratificato proporzionale, utilizzando, ai fini della stratificazione, principalmente le seguenti caratteristiche: la natura giuridica dell'impresa, l'età dell'imprenditore e la presentazione della domanda per la legge 27⁶³. Al termine delle interviste, scartando i casi di imprese non rintracciate o già cessate, e i casi di rifiuti da parte dell'imprenditore di rispondere (la partecipazione all'intervista da parte dell'imprenditore è avvenuta su base volontaria), il campione di imprese è risultato composto da 1067 imprese, di cui 215 avevano fatto domanda per ottenere contributi regionali della legge 27.

Il questionario telefonico ha permesso di verificare le informazioni già in possesso dei ricercatori, in quanto derivanti dall'archivio InfoCamere, ma di scarsa attendibilità, come ad esempio il numero di addetti per impresa o la stessa data di costituzione, nonché di ottenere alcune importanti informazioni sulla storia personale dell'imprenditore e sulle caratteristiche dell'impresa, non contenute in alcuna banca-dati già esistente. Come è comprensibile, l'indagine aveva certamente un alto valore conoscitivo in sé in quanto permetteva di esplorare in profondità le caratteristiche delle imprese neonate in Toscana con dati originali ed inediti, ma essa costituiva anche il primo passo verso un disegno di valutazione degli effetti più solido dei precedenti, in quanto permetteva di ridurre le (eventuali) differenze di partenza tra il gruppo di imprese beneficiarie della legge 27 e il gruppo di imprese non beneficiarie, che si fosse usato come gruppo di confronto.

L'utilizzo dei dati raccolti nel corso dell'indagine allo scopo di valutare gli effetti sulla sopravvivenza delle imprese della legge 27 è avvenuto qualche anno dopo da parte di Davide Conforti nell'ambito di una tesi di dottorato presso l'Università di Firenze. Sfruttando i dati contenuti nell'archivio InfoCamere fino al 31/12/2002, è stato possibile confrontare le durate in vita delle imprese nate nel 1999 ed intervistate nel 2001. Il confronto tra imprese beneficiarie e non beneficiarie ha potuto tener conto di tutte le caratteristiche ormai note sulla storia dell'impresa e dell'imprenditore, in quanto documentate dall'indagine dell'IRPET. L'abbinamento in questo caso è avvenuto

⁶³ Si è scelto di inserire nel campione tutte le imprese nate nel 1999 che avevano ricevuto il contributo della legge 27; non tutte quante però al momento della rilevazione erano rintracciabili o ancora in vita.

utilizzando in vario modo il *propensity score*, ovvero la stima della probabilità che l'unità venga assegnata al trattamento date le sue caratteristiche osservabili prima dell'intervento. Il ricercatore ha analizzato le differenze nella stima degli effetti, derivanti dall'applicazione di diversi metodi di *matching* con uso del *propensity score*, in particolare l'abbinamento sull'unità più vicina e l'abbinamento con stratificazione. L'analisi è stata ripetuta più volte, impiegando dapprima tutte le variabili disponibili (ovvero le risposte date al questionario dall'imprenditore), e successivamente soltanto quelle di natura "oggettiva" (ovvero escludendo tutte le variabili riferite alle considerazioni personali dell'imprenditore): questo al fine di evidenziare eventuali discrepanze derivanti dall'impiego di set di variabili differenti. Particolare attenzione è stata data dal ricercatore al confronto tra due sottoinsiemi di imprese, le ditte individuali da un lato, e le società dall'altro: i risultati consentono di verificare l'esistenza di effetti diversi, in termini di sopravvivenza, su differenti tipi di impresa.

Un limite di questa valutazione consiste nel considerare soltanto la coorte di imprese nate nel 1999, con una conseguente riduzione nella numerosità di unità osservate e successivamente confrontate. Ciò che viene guadagnato in *profondità*, grazie all'acquisizione di informazioni più puntuali e specifiche (nonché più attendibili) su ogni singola impresa intervistata, si perde in *ampiezza* del confronto (restringendo l'universo di riferimento): nelle analisi illustrate in precedenza i confronti sfruttavano poche informazioni relative a tutte le imprese nate in un periodo di quattro anni: in questo ultimo caso si hanno a disposizione un numero maggiore di informazioni relative però ad un campione di poche imprese nate in un solo anno. D'altra parte il dilemma tra profondità ed ampiezza dell'informazione si presenta spesso a chi si occupa di valutazione.

La stima dell'addizionalità attraverso domande dirette ai beneficiari

L'ultima questione metodologica da affrontare riguarda la domanda sull'addizionalità formulata nel paragrafo 9.2: "*in che misura l'esistenza dei contributi regionali ha influenzato i giovani imprenditori nella loro scelta di costituire una nuova impresa?*". Si tratta in verità di un quesito poco enfatizzato nella ricerca pubblicata dall'IRPET, ma riteniamo opportuno, soprattutto a scopo illustrativo, dedicare un paragrafo anche a

questa parte dello studio, in quanto è stato impiegato un metodo di analisi del tutto diverso dagli altri esaminati fino ad ora. Con questo metodo non si usano più le osservazioni delle imprese non beneficiarie per ricostruire la situazione controfattuale. Le unità di interesse non sono più le imprese, quanto piuttosto i giovani, così come la variabile-risultato non è più costituita dalla performance dell'azienda, espressa sotto forma di sopravvivenza, bensì dalle decisioni assunte dagli stessi giovani di costituire impresa.

Dato questo obiettivo conoscitivo, è chiaro come realizzare un confronto tra trattati e non-trattati diventi particolarmente difficile, se non impossibile: il gruppo dei trattati dovrebbe essere infatti costituito da giovani in grado di fare domanda per ottenere il contributo regionale; il gruppo di confronto dovrebbe essere composto da giovani aventi esattamente le stesse caratteristiche del primo gruppo, senza però la possibilità di ottenere il contributo. Per operare un simile confronto sarebbe necessario addirittura cambiare alcune modalità d'attuazione della politica (ad esempio il contributo dovrebbe essere concesso agli imprenditori prima di costituire l'impresa e non alle imprese già costituite) e selezionare i beneficiari in modo casuale tra coloro che hanno fatto domanda, in modo da avere uguali condizioni di partenza tra trattati e non-trattati (così come avviene nelle valutazioni sperimentali).

Poiché queste strade sono precluse, l'unico modo percorribile per rispondere alla domanda sull'addizionalità consiste nel chiederlo ai diretti interessati: i giovani imprenditori beneficiari del contributo regionale. Che cosa avrebbero fatto se non avessero ricevuto il contributo? Avrebbero ugualmente costituito impresa? Lo sforzo di ricostruire la situazione controfattuale in disegni valutativi di questo genere si sposta perciò dall'analista al soggetto intervistato: è il giovane imprenditore, in questo caso, a cui viene chiesto di stabilire se avrebbe ugualmente costituito impresa, nel caso in cui non avesse avuto a disposizione l'incentivo regionale. Come già illustrato nel paragrafo 7.3, l'uso di questo metodo di valutazione può suscitare diverse obiezioni, soprattutto sulla capacità (e sulla buona volontà) dell'imprenditore di immaginare quale realisticamente sarebbe stata la sua decisione in assenza di intervento pubblico. L'impiego di tale metodo rappresenta un tentativo (eroico) di ricostruire la situazione controfattuale facendo ricorso ad una sola osservazione sui beneficiari.

Box 9.1

Qu.11 Qual è stata l'influenza dell'agevolazione pubblica sulla decisione di avviare l'impresa, in una scala da 0 a 4?

- | | |
|--------------|----------------------------|
| - per niente | ☐ 0 |
| - poco | ☐ 1 |
| - abbastanza | ☐ 2 |
| - molto | ☐ 3 |
| - moltissimo | ☐ 4 |

Il gruppo di ricerca dell'IRPET ha inserito nel questionario somministrato al campione di imprese nate nel 1999 due domande finalizzate ad indagare proprio questo aspetto. Le domande sono state rivolte a tutti gli imprenditori che avevano dichiarato di aver ricevuto contributi pubblici, indipendentemente da quale ne fosse la fonte (legge 27 o altre). La prima domanda, riportata nel Box 9.1, si riferisce al peso avuto dall'agevolazione pubblica nella fondamentale decisione di avviare l'impresa. Come si può notare la formulazione delle modalità di risposta è tale da non permettere un'attribuzione di causalità esclusiva da parte dell'imprenditore all'incentivo pubblico: al massimo l'imprenditore può affermare che l'influenza del contributo sulla sua decisione di avviare l'impresa corrisponde a "moltissimo", ma non se tale influenza sia stata determinante o meno.

La seconda domanda, riportata nel Box 9.2, è invece relativa all'influenza dell'agevolazione sull'ammontare di investimento realizzato. Anche in questo caso il gruppo di ricerca ha rinunciato a porre una domanda in termini più cogenti (come accade invece nell'indagine sulle imprese della legge 488 illustrata nel capitolo 7) e si accontenta di ottenere risposte più sfumate, ma anche di difficile interpretazione a fini valutativi.

Box 9.2

Qu.12 Qual è stata l'influenza dell'agevolazione pubblica sull'entità dell'investimento effettuato, in una scala da 0 a 4?

- | | |
|--------------|----------------------------|
| - per niente | ☐ 0 |
| - poco | ☐ 1 |
| - abbastanza | ☐ 2 |
| - molto | ☐ 3 |
| - moltissimo | ☐ 4 |

Questo metodo è molto più semplice dei precedenti per quanto riguarda il trattamento dei dati: una volta ottenute le risposte è infatti sufficiente calcolarne le distribuzioni di frequenza per ogni domanda e darne un'interpretazione. Ammesso che si ritengano affidabili le risposte date, tutti coloro che hanno scelto la modalità 0 (nessuna influenza) rappresentano una stima della mancanza di addizionalità (ovvero coloro che avrebbero comunque avviato impresa anche in assenza di intervento pubblico), mentre coloro che hanno scelto le modalità di risposta da 1 a 4 risultano essere stati in qualche misura influenzati (anche se non sappiamo se avrebbero o meno costituito impresa in assenza di intervento). Non sono perciò necessari sofisticati modelli statistici. La difficoltà maggiore sta nel cercare di formulare domande e modalità di risposta che siano chiare e facilmente interpretabili.

9.4 I risultati della valutazione

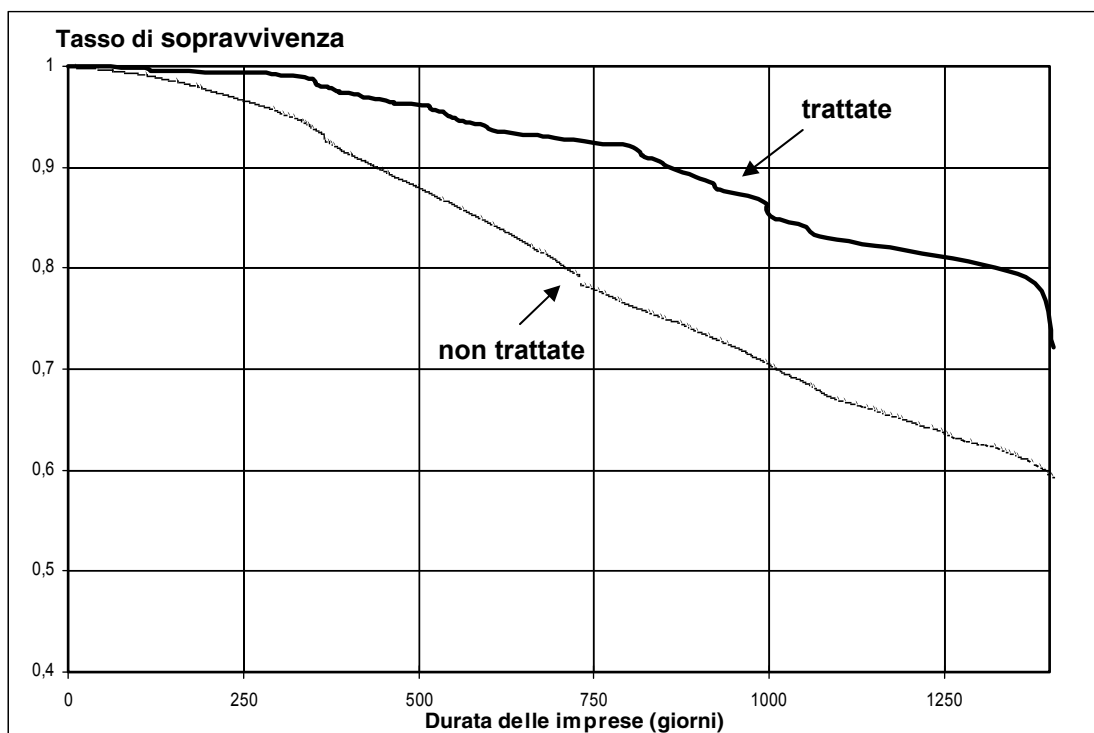
I risultati sono presentati distinguendoli a seconda del tipo di effetto che si intendeva stimare e delle fonti informative impiegate. Inizieremo esponendo gli esiti dell'analisi condotta sulla sopravvivenza delle imprese e concluderemo con le risposte date dagli imprenditori alle domande sull'addizionalità.

Gli effetti sulla sopravvivenza stimati usando solo le informazioni esistenti

Il grafico riportato nella Figura 9.1 mostra l'effetto del finanziamento relativo alla legge 27 sulla probabilità di sopravvivenza delle imprese: a partire dal modello di Cox stimato, sono state calcolate le funzioni di sopravvivenza medie, rispetto alla distribuzione delle caratteristiche conosciute delle imprese beneficiarie, in presenza ed in assenza di trattamento. La linea continua descrive la curva di sopravvivenza per le imprese trattate, mentre la linea tratteggiata la curva di sopravvivenza stimata per le imprese non-trattate. La differenza tra le due appare evidente nel grafico. Si osserva come in media un'impresa beneficiaria abbia una probabilità di sopravvivenza, intorno ai 730 giorni (due anni), pari a circa 0,93, contro una probabilità media di un'impresa non beneficiaria del finanziamento di circa 0,80: ciò significa che, secondo le stime prodotte dal modello, le imprese beneficiarie hanno un rischio di cessazione molto

minore di quelle non beneficiarie, pari a circa il 35% di quello delle imprese non beneficiarie⁶⁴.

Figura 9.1: Sopravvivenza media di imprese trattate e non-trattate (modello di Cox)



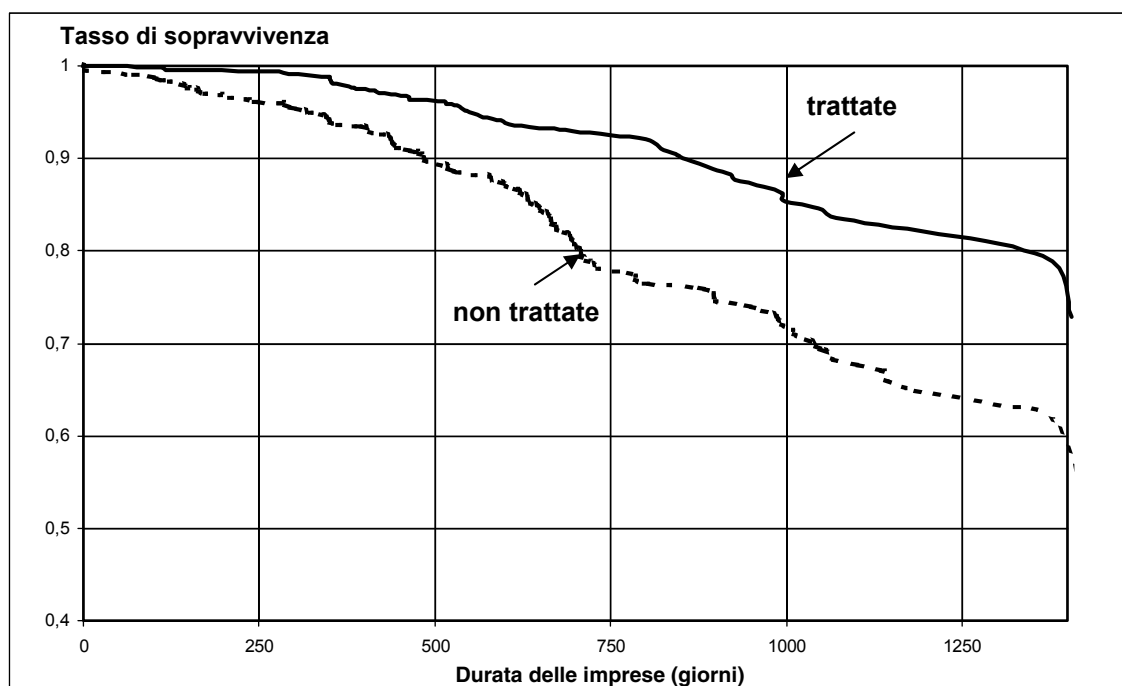
Fonte: Mealli, F., Pagni, R. (2001)

Anche applicando la procedura del *matching esatto*, in modo da non imporre una particolare forma funzionale alle variabili esplicative che compaiono nel modello di regressione, il risultato non cambia e il finanziamento sembra comunque avere un discreto effetto positivo sulla capacità di sopravvivenza delle imprese. Il grafico riportato nella Figura 9.4 mostra le curve di sopravvivenza per i due gruppi di imprese: la performance delle beneficiarie è rappresentata sempre dalla linea continua mentre quella delle “gemelle” è rappresentata dalla linea tratteggiata. Ancora una volta le differenze tra le due curve risultano essere significative e molto simili a quelle presentate nel grafico precedente. In media un’impresa beneficiaria aumenta la sua probabilità di sopravvivenza dopo 1095 giorni (tre anni) di circa 0,15. A quella scadenza le imprese beneficiarie hanno un tasso di sopravvivenza pari a 0,83 contro lo

⁶⁴ Ad una conclusione molto simile giungeva anche la valutazione precedente, che ugualmente utilizzava il modello a rischi proporzionali di Cox. Si veda Bondonio e Martini (2001).

0,68 registrato per le imprese non beneficiarie; nell'arco di quattro anni le imprese beneficiarie hanno una durata media stimata di 153 giorni maggiore rispetto a quella osservata sulle imprese non beneficiarie.

Figura 9.2: Confronto nella sopravvivenza media tra imprese trattate e non-trattate (matching esatto)



Fonte: Mealli F., Pagni R. (2001)

Come abbiamo avuto modo di spiegare nel paragrafo precedente, i confronti illustrati in queste pagine presentano una debolezza di fondo: i modelli statistici sono in grado di controllare le differenze esistenti per le variabili note (in quanto tratte dall'archivio InfoCamere), ma non sono in grado di tener conto dell'influenza del processo di (auto)selezione che caratterizza l'intervento e che probabilmente condiziona la performance delle imprese. Possono esservi "differenze invisibili" all'occhio del ricercatore (in termini di dinamismo, livello culturale, lunghezza e tipo di esperienza professionale precedente) nelle caratteristiche di partenza degli imprenditori che potrebbero spiegare la diversa durata riscontrata nei due gruppi di imprese. I risultati ottenuti ci confortano quindi sul fatto che i contributi della legge 27 non sono stati sprecati su imprese senza speranza, in quanto le aziende beneficiarie dimostrano un'ottima capacità di sopravvivenza, ma non rappresentano una stima degli effetti

esente da critiche. Nel prossimo paragrafo vengono perciò illustrati i risultati di uno studio che tenta di superare tali debolezze.

Gli effetti sulla sopravvivenza stimati con le informazioni dell'indagine campionaria

Ricordiamo come in questo caso le unità di osservazione considerate nel confronto provengano da un campione di poco più di 1000 imprese nate nel 1999 ed intervistate nel 2001. La fonte informativa di riferimento non è più l'archivio amministrativo InfoCamere, ma una banca-dati costruita a fini valutativi in seguito alla realizzazione di un'indagine campionaria. All'interno del campione di riferimento viene individuato un sotto-campione definito come "insieme delle imprese eleggibili", che raccoglie le imprese che rispettano i requisiti di ammissibilità definiti nella normativa. Il numero complessivo delle eleggibili è 634, di cui 157 sono beneficiarie.

L'autore presenta nella Tabella 9.2 le differenze riscontrate sulla variabile-risultato, espressa come *probabilità di cessazione*, ad una certa distanza dalla data di costituzione. I confronti sono effettuati prendendo a riferimento cinque possibili distanze temporali (riportate per riga), a partire da un anno fino a tre anni. Le imprese sono invece distinte in beneficiarie della legge 27 e non beneficiarie; il confronto prende soltanto in considerazione il fatto che le imprese abbiano ricevuto o meno il contributo regionale e non tiene conto di eventuali differenze nei due gruppi. Naturalmente la stima dell'effetto riportata nell'ultima colonna, e derivata dalla differenza tra i valori delle precedenti due colonne, è poco plausibile, a causa delle diverse condizioni di partenza nei due gruppi. Da tale confronto si desume comunque una performance peggiore, in termini di sopravvivenza, per le imprese non beneficiarie, a partire dai due anni di distanza dal contributo.

Tabella 9.2: Confronto nelle probabilità di cessazione tra le imprese toscane nate nel 1999

Distanza dalla data di costituzione	Beneficiarie	Non beneficiarie	Stima dell'effetto
<i>Dopo un anno</i>	0,013	0,006	+0,007
<i>Dopo un anno e mezzo</i>	0,019	0,021	-0,003
<i>Dopo due anni</i>	0,019	0,057	-0,038
<i>Dopo due anni e mezzo</i>	0,038	0,086	-0,048
<i>Dopo tre anni</i>	0,064	0,128	-0,064

Fonte: Conforti D. (2004)

La Tabella 9.3 riporta le stime dell'effetto del contributo regionale sulla sopravvivenza delle imprese. Questa volta le stime derivano da confronti effettuati a parità di condizioni, grazie all'applicazione di due differenti metodi di stima. Il metodo presentato nella prima colonna (abbinamento sull'unità più vicina) impone di associare ad ogni unità trattata quella che ha il *propensity score* più vicino; il metodo presentato nella seconda colonna (abbinamento con stratificazione) prevede di confrontare imprese trattate e non-trattate con un *propensity score* che cade nello stesso intervallo di valori.

Tabella 9.3: Stima degli effetti sulla probabilità di cessazione (metodi di *matching* a confronto)

Distanza dalla data di costituzione	Stima dell'effetto (<i>matching</i> sull'unità più vicina)	Stima dell'effetto (<i>matching</i> con stratificazione)
<i>Dopo un anno</i>	+0,006	+0,010
<i>Dopo un anno e mezzo</i>	+0,013	+0,010
<i>Dopo due anni</i>	-0,032	-0,040
<i>Dopo due anni e mezzo</i>	-0,076	-0,054
<i>Dopo tre anni</i>	-0,166	-0,076

Fonte: Conforti D. (2004)

I risultati ottenuti con l'applicazione del primo metodo suggeriscono l'esistenza di un effetto positivo molto accentuato, a partire dai due anni e mezzo di distanza dalla data di costituzione; tra le beneficiarie e le non beneficiarie con caratteristiche simili viene registrata una differenza in termini di probabilità di cessazione di 0,16 dopo i tre anni di vita. Nel caso del metodo stratificato, l'effetto appare sempre positivo, ma molto meno marcato (0,076 dopo tre anni), meno della metà della stima precedente; in ogni caso l'entità dell'effetto stimato cresce all'aumentare dell'intervallo di tempo considerato. Le stime calcolate con metodi differenti possono dunque rappresentare il limite inferiore e superiore di un intervallo di valori all'interno del quale collocare una stima plausibile dell'effetto dell'intervento.

La Tabella 9.4 presenta infine le stime dell'effetto del contributo regionale calcolate con gli stessi due metodi illustrati sopra per due differenti sottogruppi di imprese: le società e le ditte individuali. Tali confronti sono utili per comprendere se l'effetto si distribuisce diversamente tra le due tipologie di imprese e, dunque, per stabilire se questa particolare politica di aiuto è più efficace quando viene rivolta alle società piuttosto che alle ditte individuali.

Tabella 9.4: Stima degli effetti sulla probabilità di cessazione per tipo di imprese

	Società		Ditte individuali	
	unità più vicina	stratificazione	unità più vicina	stratificazione
<i>Dopo un anno</i>	+0,027	+0,026	0,000	-0,002
<i>Dopo un anno e mezzo</i>	+0,040	+0,039	0,000	-0,004
<i>Dopo due anni</i>	+0,040	+0,030	-0,061	-0,066
<i>Dopo due anni e mezzo</i>	+0,053	+0,035	-0,049	-0,090
<i>Dopo tre anni</i>	0,000	-0,031	-0,098	-0,084

Fonte: Conforti D. (2004)

Le stime calcolate per i due gruppi utilizzando i due diversi metodi sono leggermente difforni nei risultati, ma mettono chiaramente in risalto alcune tendenze generali. Nel caso delle società la politica sembra non ridurre la probabilità di cessazione, mentre nel caso delle ditte individuali l'intervento sembra ridurre in modo cospicuo il rischio di cessazione, in particolare dopo due anni. Questo punto viene espresso nelle conclusioni dello studio: *“l'analisi condotta su società e ditte individuali dà luogo a risultati molto diversi tra loro. Se per queste ultime, infatti, tutti i metodi adottati rilevano la totale inefficacia del trattamento, per le prime si riscontra un impatto sicuramente positivo dell'intervento sulle probabilità di cessazione”*.

Gli effetti sull'addizionalità

Per concludere, proponiamo un'analisi delle risposte offerte dagli imprenditori alle domande sull'addizionalità contenute nell'indagine dell'IRPET. Nel rapporto conclusivo dell'indagine questi risultati sono presentati in un paragrafo intitolato “l'autovalutazione dell'impatto”, proprio per sottolineare il fatto che sono gli stessi imprenditori a “valutare” gli effetti del contributo pubblico sulla loro decisione di avviare l'impresa. Tali domande sono state sottoposte sia alle imprese che avevano ricevuto i contributi della legge 27, sia ad imprese che avevano ricevuto agevolazioni di altra natura; questo permette di rilevare eventuali differenze tra le diverse tipologie di contributo.

La Tabella 9.5 riporta le distribuzioni di frequenza delle risposte date dagli imprenditori. Come abbiamo già avuto modo di sottolineare tali risposte non consentono in realtà di stabilire per quale percentuale di imprenditori il contributo sia stato determinante nella scelta di costituire impresa; esse avrebbero però permesso di stabilire per quale

percentuale di imprenditori il contributo è stato del tutto irrilevante. Tale percentuale è data dalla quota di coloro che hanno risposto “per niente” alla domanda sull’addizionalità. Gli autori della ricerca presentano però i risultati aggregando le diverse modalità di risposta in tre categorie: (i) poco o per niente, (ii) abbastanza, (iii) molto e moltissimo. Questa scelta non consente di comprendere appieno da quanti imprenditori il contributo non sia stato percepito come un reale incentivo ad investire, ma come una sorta di “dono” da parte dell’amministrazione pubblica. Le modalità di presentazione dei risultati finiscono perciò per ridurre il portato conoscitivo dell’analisi.

Tabella 9.5: Risposte alle domande sull’addizionalità

	Influenza su decisione di avvio dell’impresa			Influenza su entità dell’investimento		
	Poco o niente	Abbastanza	Molto o moltissimo	Poco o niente	Abbastanza	Molto o moltissimo
Legge 27	45,9%	33,6%	20,5%	44,3%	34,4%	21,3%
No legge 27	54,6%	24,4%	21,1%	49,9%	31,4%	18,8%

Fonte: Mealli F., Pagni R. (2001)

Due le principali conclusioni che si possono trarre dall’analisi delle risposte. In primo luogo, si rileva su entrambe le domande un’alta percentuale di risposte “negative” (poco o per niente), nel senso cioè di risposte che negano o minimizzano l’addizionalità della politica: ciò lascia intravedere una certa “buona fede” da parte degli intervistati, in quanto essi non hanno alcuna convenienza a sottostimare l’influenza della politica (anzi, in vista di un eventuale rifinanziamento dalla Regione, avrebbero la convenienza contraria). In secondo luogo, si registra una certa differenza tra le risposte date dai beneficiari della legge 27 rispetto a coloro che hanno usufruito di altre agevolazioni pubbliche; sembra infatti che l’influenza dei contributi sia più alta per i primi che non per i secondi. Ciò significa che la legge 27 sembra aver avuto, almeno nelle percezioni degli imprenditori, una forza incentivante maggiore.

CAPITOLO 10

GLI EFFETTI DEL LAVORO INTERINALE SULLA PROBABILITÀ DI TROVARE UN IMPIEGO STABILE

di Luca Mo Costabella

L'introduzione del lavoro interinale rappresenta uno tra i più importanti provvedimenti presi nell'ultimo decennio per rendere maggiormente flessibile il mercato del lavoro italiano. Valenza che rende il lavoro interinale oggetto di vivaci controversie. Secondo alcuni, questa nuova forma contrattuale ha portato ovvi benefici alle imprese senza danneggiare i lavoratori, offrendo anzi loro nuove opportunità di occupazione. Secondo altri, invece, l'introduzione del lavoro interinale è stato il principale veicolo attraverso cui si è aumentata la precarizzazione del lavoro in Italia.

Anche messa da parte la natura prevalentemente ideologica della contrapposizione, resta un quesito importante – e in linea di principio empiricamente risolvibile – sugli effetti di questa politica di flessibilizzazione del mercato del lavoro: un lavoratore che ricorra al lavoro interinale vede ridotte o aumentate le sue *chance* di trovare successivamente un'occupazione a tempo indeterminato?

Nell'ambito di un'iniziativa più ampia mirata ad applicare strumenti di monitoraggio e valutazione a questioni di politica del lavoro, nel 2001 il Ministero del Lavoro e delle Politiche Sociali bandisce una gara per l'affidamento di una valutazione degli effetti del lavoro interinale sulla probabilità di trovare un impiego stabile. Lo studio viene affidato all'Istituto Universitario Europeo e all'Università di Firenze. Il rapporto finale della valutazione è stato realizzato nel 2003 da Andrea Ichino, Fabrizia Mealli e Tommaso Nannicini con il titolo "*Il lavoro interinale in Italia. Trappola del precariato o trampolino verso un impiego stabile?*". Questo capitolo è basato sui risultati presentati in tale rapporto.

10.1 La logica del lavoro interinale

La logica del lavoro interinale si basa innanzitutto sul bisogno di flessibilità da parte delle imprese, ostacolate nella loro possibilità di reagire a repentini cambiamenti esterni

dalla rigidità dei contratti di assunzione (in particolare dalle difficoltà di licenziamento). Le politiche destinate a flessibilizzare il mercato del lavoro possono essere da un lato volte all'ammorbidimento delle regole del lavoro "standard", diminuendone la rigidità; dall'altro possono invece lasciare inalterata la regolamentazione del lavoro a tempo indeterminato, introducendo forme contrattuali più flessibili o facilitando il ricorso a quelle già esistenti. In Italia il processo di flessibilizzazione è stato caratterizzato prevalentemente da politiche del secondo tipo. La regolamentazione del lavoro standard è stata lasciata inalterata, mentre è stata introdotta o resa più accessibile una vasta gamma di contratti meno rigidi quali, ad esempio, i contratti di formazione e lavoro, i contratti *part-time* e i contratti a tempo determinato.

Il lavoro interinale è stato introdotto in Italia con la legge 196 del 1997, nota come "pacchetto Treu". La caratteristica maggiormente innovativa di questa forma contrattuale è la presenza di un intermediario privato (l'impresa *fornitrice* di lavoro temporaneo, cioè l'agenzia interinale) tra il lavoratore e l'impresa (detta in questo contesto impresa *utilizzatrice*). L'impresa fornitrice svolge opera di selezione e reclutamento della manodopera e invia all'impresa utilizzatrice che ne faccia richiesta uno o più lavoratori adatti a svolgere le mansioni richieste. Non esiste quindi un rapporto contrattuale tra impresa utilizzatrice e lavoratore, perché tutti i rapporti formali sono filtrati dall'impresa fornitrice. Essa stipula un contratto con l'impresa utilizzatrice, con il quale si impegna a fornirle la manodopera richiesta, e un contratto di assunzione con il lavoratore da inviare presso l'impresa utilizzatrice (in "missione").

Dal punto di vista del lavoratore, con l'interinale non si introducono particolari novità. Egli è assunto dall'impresa fornitrice con un normale contratto, quasi sempre a tempo determinato, di durata variabile da pochi giorni a oltre un anno; inoltre deve percepire almeno lo stesso salario spettante ai pari livello assunti dall'impresa utilizzatrice a tempo indeterminato e gode obbligatoriamente della stessa copertura assicurativa e previdenziale.

L'innovazione apportata dall'interinale è maggiormente visibile dal lato delle imprese che utilizzano il lavoratore. L'impresa utilizzatrice è sollevata da tutti gli oneri di gestione del contratto di assunzione, delegati all'impresa fornitrice in quanto ufficiale

datore di lavoro. Inoltre, essa può fruire di un servizio di preselezione del personale nonché di un reperimento tempestivo della manodopera necessaria. Tali vantaggi sono bilanciati dall'ovvio aumento dei costi connesso al compenso per i servizi dell'agenzia interinale (variabile in linea di massima dal 20% al 30% del salario del lavoratore).

In modo simile a quanto fatto per altre forme contrattuali flessibili, il lavoro interinale è stato inizialmente introdotto con una regolamentazione tesa a ostacolarne l'utilizzo per il reclutamento sistematico della manodopera. E, in modo simile alle altre forme contrattuali, le limitazioni inizialmente previste sono state ammorbidite o quasi eliminate successivamente. La legge istitutiva prevedeva che il ricorso all'interinale fosse limitato ai casi di emergenza (ad esempio per fare fronte a picchi occasionali di lavoro o assetti produttivi imprevisi, o per sostituire lavoratori temporaneamente assenti), e stabiliva precisi limiti quantitativi, tra cui un numero massimo di missioni a cui le imprese potevano ricorrere annualmente. Tutte queste limitazioni non hanno impedito che il lavoro interinale incontrasse subito un grande successo: nei primi anni esso ha fatto registrare una crescita impressionante, che solo dal 2002 è rallentata, a seguito dell'entrata nella fase di assestamento.

Stando alle stime di Confinterim, dal 1999 al 2002 il numero di missioni svolte è triplicato, passando da 195.000 a oltre 615.000. Alla crescita dell'utilizzo dell'interinale si è accompagnata la diffusione delle agenzie sul territorio: nel 1999 si contavano circa 600 agenzie, mentre a fine 2002 queste erano oltre 2200. Attualmente le agenzie interinali coprono il territorio nazionale in modo quasi capillare, soprattutto nell'Italia settentrionale, dove ha sede più del 70% delle filiali.

Nonostante il forte successo, l'interinale rimane tuttavia poco rappresentativo del lavoro totale svolto in Italia. Se le missioni costituiscono oltre il 15% del totale annuo delle assunzioni, la percentuale dello stock di occupati totali rappresentato dagli occupati interinali non arriva all'1%, in linea con le evidenze provenienti dagli altri Paesi in cui l'interinale è stato introdotto.

Il lavoro interinale si è configurato, soprattutto nei primi tempi, come un fenomeno diffuso prevalentemente in ambito industriale e in particolare nell'industria metalmeccanica. Ne è conseguito un maggiore coinvolgimento di lavoratori maschi, che

costituiscono il 60% circa del totale. I lavoratori interinali sono mediamente giovani, con un'età molto inferiore a quella sia degli occupati a tempo indeterminato, sia degli occupati a termine. La composizione per età ricorda più che altro quella dei disoccupati, il che costituisce un indizio circa il carattere spesso transitorio delle missioni, che rappresentano per molti una possibilità di lavoro occasionale (ad esempio da conciliarsi con gli impegni scolastici) o di primo approccio al mondo del lavoro.

Riguardo alle modalità di utilizzo da parte delle imprese, rivestono particolare importanza descrittiva i motivi di ricorso all'interinale e la durata delle missioni. I due terzi delle missioni non vanno oltre il mese, un terzo addirittura non supera la settimana, mentre sono abbastanza rari i casi di missioni molto lunghe. Queste ultime sono in genere indirizzate allo svolgimento di mansioni di livello più alto, e sono infatti caratterizzate da salari decisamente maggiori. La durata delle missioni dipende da molti fattori, tra cui il tipo di lavoro da svolgere, la volontà dell'individuo e le intenzioni dell'impresa.

A questo proposito va citata una delle particolari motivazioni che spingono le imprese a ricorrere all'interinale. I vincoli previsti dalla legge, che ammettevano il ricorso all'interinale solo nei casi in cui particolari condizioni di emergenza rendessero necessaria una maggiore flessibilità, hanno avuto in realtà una pura valenza formale. L'interinale è invece spesso utilizzato per assunzioni in situazioni ordinarie, ed è molto ricorrente il caso (seppure ufficialmente non ammissibile, ma di fatto ben accetto) in cui si ricorre ad esso per prendere temporaneamente in prova un lavoratore, con la prospettiva di assumerlo a tempo indeterminato se la prova dà esito positivo. Relativamente a questi casi, esiste evidenza empirica riguardo al legame tra intenzioni dell'impresa e durata della missione⁶⁵, che mostra come quest'ultima influenzi la probabilità che il lavoratore sia successivamente assunto a tempo indeterminato dall'impresa utilizzatrice (in particolare, la probabilità sembra essere maggiore per le missioni di durata medio-lunga).

Negli anni l'interinale ha parzialmente modificato la sua identità, portando ad una progressiva apertura ad altri settori. Pur rimanendo uno strumento utilizzato soprattutto

⁶⁵ Si veda Montanino A., Sestito P., "Le molte funzioni del lavoro interinale in Italia: da strumento di flessibilità a contratto di prova", *Rivista di Politica Economica*, marzo-aprile 2003.

nel manifatturiero, è progressivamente aumentata la sua diffusione nel terziario. A questo allargamento della sfera degli ambiti produttivi è corrisposta una variazione nelle caratteristiche dei lavoratori coinvolti, tra i quali è aumentata la percentuale di donne. Per quanto riguarda l'età, si è incrementato il peso delle classi meno giovani. Questa evidenza, unita all'aumento delle missioni di durata medio-lunga (sempre connesse a mansioni di consulenza o con alti contenuti professionali) e delle persone che lavorano ripetutamente con l'interinale, implica che questa nuova forma contrattuale abbia progressivamente assunto un ruolo più ampio rispetto alle intenzioni iniziali: nonostante le modifiche citate non siano state tali da stravolgere l'identità dell'interinale, esse hanno comunque contribuito ad allontanarlo dal ruolo esclusivo di "lavoro per una volta".

10.2 La domanda di valutazione

L'effetto sulla storia lavorativa successiva derivante dal fatto di essere passati per il canale dell'interinale è altamente incerto, per via delle potenzialità sia positive (legate alla possibilità che l'esperienza interinale favorisca l'inserimento in azienda) sia negative (legate alla possibilità che essa contribuisca a relegare il lavoratore in uno stato di precarietà). La valutazione che esaminiamo cerca di risolvere questa incertezza, stimando l'effetto del ricorso al lavoro interinale sulla probabilità di trovare un lavoro stabile.

Nonostante la stabilità possa essere definita in vari modi, nello studio in questione si considera stabile, secondo la concezione più usuale, un contratto a tempo indeterminato. Obiettivo della valutazione è quindi capire se lo svolgimento di una missione interinale fa aumentare o diminuire la probabilità di trovare successivamente un lavoro a tempo indeterminato.

Il segno di questo effetto non è a priori scontato. Una tesi è che, mantenendo la rigidità del lavoro "standard" e aumentando le possibilità di assunzione con contratti flessibili, si fa gravare tutto il peso della flessibilità sui lavoratori senza contratto a tempo indeterminato, costringendoli a una storia lavorativa segnata da precarietà e incertezza. Secondo questa tesi, il lavoratore interinale rischia di non uscire mai dal "tunnel" del

lavoro precario, riuscendo solo a inanellare una serie di esperienze lavorative temporanee, oppure di sperimentare un periodo di incertezza eccessivamente lungo prima di riuscire a stabilizzare la propria condizione.

La tesi opposta è che il lavoro interinale funga da efficace trampolino per l'ingresso nel lavoro stabile. Le sue caratteristiche, volte a una facilitazione dell'incontro tra domanda e offerta di lavoro, possono favorire l'ingresso nel mercato del lavoro di alcune particolari categorie di lavoratori, ad esempio i giovani in cerca di prima occupazione. La missione interinale può inoltre permettere a un lavoratore di aumentare la propria esperienza o di aggiornare le competenze in determinate mansioni. È anche possibile che le imprese utilizzatrici facciano ricorso al lavoro interinale per provare un lavoratore per un breve periodo, con la prospettiva di inserirlo successivamente nell'organico in modo stabile.

Gli autori della valutazione riassumono in due categorie i potenziali motivi che possono determinare, a seguito di una missione interinale, un aumento o viceversa un calo delle *chance* di trovare un lavoro a tempo indeterminato.

Sul lato dei motivi positivi, rendersi disponibile a svolgere mansioni provvisorie dà luogo ad un meccanismo di "segnalazione": per i lavoratori più produttivi, ciò equivale a rendersi disponibili per essere osservati ed eventualmente valutati per un successivo inserimento stabile nell'organico dell'impresa. Inoltre, oltre ad aumentare l'esperienza e le competenze lavorative, il lavoratore in missione interinale può allargare la propria rete di contatti sociali e acquisire informazioni su altre possibilità lavorative.

I motivi negativi sono speculari ai precedenti. Il meccanismo di "segnalazione" potrebbe essere inverso, evidenziando il lavoratore che accetta la missione interinale come impossibilitato ad entrare nel mercato del lavoro attraverso canali primari, dunque meno abile di altri. Infine, il lavoratore accumula abilità e competenze inferiori a quelle che avrebbe guadagnato con altre esperienze, a causa della reticenza del datore di lavoro a formare un lavoratore destinato a non rimanere nell'organico dell'impresa.

Identificare il contributo di questi singoli meccanismi all'effetto complessivo del lavoro interinale è molto difficile, ma non rientra direttamente negli obiettivi della valutazione. Qualunque sia l'entità delle varie componenti che concorrono al risultato, ciò che

interessa in primo luogo è sapere se in generale lo svolgimento di una missione interinale comporta o meno un aumento nella probabilità di trovare successivamente un lavoro a tempo indeterminato. Solo quando si sia stabilito in modo credibile il segno di questa relazione si potrà procedere a indagare quali meccanismi la producono.

Il quesito al centro della valutazione può essere quindi formulato nel modo seguente: si vuole sapere qual è la differenza tra la probabilità che un lavoratore trovi un'occupazione a tempo indeterminato dopo aver effettuato missioni di lavoro interinale e la probabilità che lo stesso lavoratore avrebbe avuto nel caso in cui non avesse svolto tali missioni. Questa differenza, se positiva, indica che svolgere un lavoro interinale ha *aumentato* le sue *chance* occupazionali successive e le ha *diminuite* nel caso in cui sia negativa.

10.3 Le principali questioni metodologiche e le soluzioni adottate

La strategia di fondo utilizzata dai valutatori è quella di confrontare gli esiti occupazionali osservati per un gruppo di lavoratori che abbiano lavorato con l'interinale, con gli esiti occupazionali di un altro gruppo di lavoratori che *non* abbiano svolto missioni interinali. Tuttavia, affinché la differenza nelle probabilità dei due gruppi si possa considerare *effetto* del lavoro interinale, è necessario che sia soddisfatta una condizione: il gruppo dei “non interinali” deve fornire una buona approssimazione del controfattuale: bisogna cioè essere disposti a credere che, se i lavoratori interinali non avessero svolto missioni, essi avrebbero avuto la stessa probabilità di occupazione a tempo indeterminato osservata per i lavoratori non interinali. Perché i risultati di questo confronto indichino realmente un effetto e non solo differenze di partenza, è dunque necessario tenere debitamente conto di queste differenze iniziali ricorrendo a tecniche statistiche *ad hoc*.

Il primo passo del disegno della valutazione è consistito nel delimitare l'analisi dal punto di vista territoriale e temporale. L'analisi ha come periodo di riferimento il biennio 2001-2002: si tratta di anni in cui l'interinale ha già superato lo stadio di novità e, sebbene sia ancora in forte espansione, funziona già a pieno regime.

Per quanto riguarda la dimensione territoriale, si sono scelte nove Province di due Regioni, Toscana e Sicilia. La scelta di concentrare l'attenzione sul Centro-Sud è dettata da due ragioni. In primo luogo, gli autori considerano più interessante l'efficacia dell'interinale in tali zone, caratterizzate da condizioni economiche e opportunità lavorative decisamente diverse da quelle dell'Italia settentrionale.

La seconda ragione è di carattere maggiormente metodologico. Mentre nel Nord le filiali delle imprese fornitrici mostrano una diffusione ormai capillare, nel Centro e nel Sud esse sono ancora poco presenti e tendono a collocarsi in modo da essere vicine alle zone con maggiore concentrazione di imprese. Supponendo di conoscere la distanza dell'abitazione di un lavoratore dalla più vicina agenzia interinale è dunque possibile avere una misura della facilità con cui il lavoratore poteva accedere al lavoro interinale. Queste informazioni sono state sfruttate in fase d'analisi per correggere le stime degli effetti, tenendo cioè conto delle differenze preesistenti tra gli individui in termini di opportunità di accesso all'interinale. La scelta delle Province all'interno della Regione è stata effettuata in modo da disporre di dati su realtà territoriali con strutture produttive e mercato del lavoro simili, ma con una diversa diffusione delle agenzie interinali.

Il disegno della valutazione prevede la costruzione di due gruppi di lavoratori da porre a confronto. Il gruppo dei trattati è costituito da lavoratori residenti nelle nove Province considerate che abbiano svolto una missione di lavoro interinale nel primo semestre del 2001. Il gruppo di confronto è costituito da un campione di residenti nelle stesse nove Province che all'inizio del 2001 non fossero impegnati in una missione interinale né occupati a tempo indeterminato (erano cioè non occupati oppure occupati con contratti a termine).

L'individuazione del gruppo dei trattati avviene grazie alla collaborazione di una importante agenzia interinale, Manpower, che all'epoca ha una quota di mercato del 30% circa in Toscana e del 45% circa in Sicilia. Questa agenzia mette a disposizione la propria banca-dati per l'estrazione di tutti i nominativi dei propri lavoratori che hanno svolto una missione nel periodo d'interesse. Dagli archivi vengono estratti complessivamente 1.173 lavoratori, che si tenta successivamente di contattare per via telefonica allo scopo di ottenere una serie di informazioni utili per l'analisi. Di questi,

alcuni vengono successivamente eliminati perché, pur avendo svolto una missione nel territorio di riferimento nella prima metà del 2001, erano residenti altrove, e dunque inseriti in contesti economici differenti da quelli osservati. Un'altra porzione dei 1.173 lavoratori viene esclusa dal campione in quanto non reperibile o non disponibile a rispondere alle domande poste. Alla fine, si può contare su un campione di 553 lavoratori interinali disponibili per l'analisi.

Il gruppo di confronto viene invece individuato mediante una procedura necessariamente più complessa e costosa. Si comincia con un campionamento casuale della popolazione residente nelle nove Province, a partire dagli elenchi telefonici. I soggetti da includere nel gruppo di controllo vengono individuati contattando telefonicamente le famiglie residenti estratte casualmente dagli elenchi, verificando la presenza nel nucleo familiare di una persona tra i 18 e i 40 anni (fascia d'età che assicura una buona confrontabilità con i trattati) che al 1° gennaio 2001 non sia occupata a tempo indeterminato e non sia impegnata in una missione interinale. Se nella famiglia contattata esiste una persona con tali caratteristiche, questa entra nel gruppo di confronto. Nel caso in cui la persona dichiara di avere svolto una missione interinale nel periodo di riferimento (il che è possibile, se la persona si è rivolta a un'agenzia diversa da quella contattata dagli autori), questa viene inserita nel gruppo di trattamento.

Al termine delle interviste telefoniche si dispone di informazioni su 3.724 individui non occupati stabilmente né impegnati in una missione interinale nel gennaio del 2001. Da questi i valutatori decidono successivamente di eliminare tutti gli studenti (1.259) e gli altri soggetti genericamente non appartenenti alle forze di lavoro (596), in quanto entrambi i gruppi hanno alcune caratteristiche (la storia lavorativa in particolare) troppo differenti dai lavoratori interinali.

Alla fine delle operazioni di pulizia dei dati, volte a eliminare tutti i soggetti che presentano informazioni incomplete o incongruenti, il gruppo di confronto è composto da 1.519 individui disoccupati o occupati con contratti atipici (ma non interinali) nel gennaio 2001, mentre il gruppo dei trattati è composto da 511 lavoratori impegnati in una missione interinale nel primo semestre del 2001.

Una volta operata la selezione dei due gruppi, restano ancora da raccogliere due tipi di informazioni cruciali: lo *stato occupazionale* ad una certa distanza dal primo semestre 2001 (cioè la variabile-risultato) e le informazioni necessarie a rendere *più confrontabili* i due gruppi di lavoratori. A questo scopo i membri di entrambi i gruppi vengono intervistati telefonicamente nella seconda metà del 2002, a circa 18 mesi dal periodo in cui per il gruppo dei trattati si è svolta la missione interinale. Durante l'intervista si chiede loro se sono al momento occupati a tempo indeterminato: la risposta determina quindi il valore della variabile-risultato per ciascuno dei soggetti sotto osservazione.

Se i due gruppi osservati fossero perfettamente omogenei tra loro, il semplice confronto tra la quota di occupati nel gruppo degli interinali e la quota di occupati nel gruppo dei non interinali fornirebbe una stima attendibile dell'effetto dell'aver compiuto una missione interinale. La scelta di eliminare dal gruppo di confronto certe categorie di individui (occupati a tempo indeterminato già nel 2001, non forze di lavoro, soggetti in età non compresa tra i 18 e i 40 anni), le cui caratteristiche sono decisamente contrastanti con quelle solitamente osservabili nei lavoratori interinali, esclude che tra trattati e controlli esistano drastiche differenze tali da precludere ogni tentativo di confronto. Tuttavia i gruppi mantengono ancora delle differenze, che si tenta di eliminare sfruttando la tecnica del *matching* statistico, illustrata al capitolo 5 della sezione metodologica.

Come si è detto in quella sede, l'ipotesi di base di questo metodo è che le differenze di partenza (cioè prima dell'eventuale missione interinale) nella probabilità di trovare un lavoro stabile tra i due gruppi siano dovute ad altrettante differenze in una serie di caratteristiche *osservabili*. Se la probabilità di trovare un lavoro stabile dipende dal genere o dall'età del lavoratore o dal suo titolo di studio, e rispetto a queste caratteristiche i due gruppi sono diversi (ad esempio i lavoratori interinali mostrano una maggiore quota di diplomati e in media sono più giovani rispetto ai "controlli"), si deve abbinare ad ogni lavoratore interinale uno o più individui di controllo che mostrino caratteristiche molto simili alle sue. Avendo ora l'interinale e i controlli ad esso abbinati le stesse caratteristiche osservabili, ed avendo ipotizzato che queste siano le uniche in grado di influenzare la probabilità di trovare un lavoro a tempo indeterminato, tra questi individui non dovrebbero più sussistere differenze di partenza in tale probabilità. L'aver

svolto una missione è l'unica caratteristica che distingue il trattato dai controlli ad esso abbinati, per cui eventuali differenze a posteriori sono interpretabili plausibilmente come l'effetto di aver svolto la missione.

Durante l'intervista telefonica effettuata a fine 2002, oltre alla condizione occupazionale, ai membri di entrambi i gruppi viene sottoposta una batteria di domande per raccogliere tutte le informazioni di possibile utilità per l'abbinamento statistico. Al termine dell'indagine si hanno quindi a disposizione, per interinali e controlli, informazioni riguardo a:

- caratteristiche demografiche: sesso, età, nazionalità;
- istruzione: titolo di studio più alto, informazioni sui titoli di studio (anno di conseguimento, voto), eventuale partecipazione a corsi di formazione;
- caratteristiche sulla famiglia di provenienza, in particolare riguardo al padre: età, titolo di studio, condizione occupazionale e informazioni sulla storia lavorativa;
- informazioni sulla storia lavorativa dell'intervistato, fra cui: percentuale di tempo passata senza occupazione dal termine degli studi al presente, numero di lavori svolti, condizione occupazionale prevalente e sue caratteristiche.

Una volta raccolte tutte le informazioni, è necessario stabilire un criterio per quantificare la somiglianza tra i lavoratori interinali e i controlli da abbinare. Gli autori ricorrono alla tecnica del *propensity score*, discussa in dettaglio al capitolo 5, dove sono illustrate quattro tecniche alternative. Ad ogni alternativa sono associati determinati vantaggi e svantaggi, e in certi casi la scelta del metodo di *matching* ha un'influenza sulle stime dell'effetto. Per tutelarsi da questa fonte di incertezza, è buona pratica replicare l'analisi ricorrendo a diversi metodi, in modo da poter saggiare la robustezza dei risultati rispetto alla scelta del metodo. Nello studio in questione gli autori replicano la stima degli effetti con il *matching sull'unità più vicina* e con il *matching con la funzione kernel*. Per necessità di sintesi, saranno presentati i risultati solo per il primo metodo.

A seguito della procedura di *matching*, sono stati abbinati 511 interinali con 305 controlli⁶⁶.

⁶⁶ Il numero dei controlli è inferiore al numero dei trattati perché i primi sono selezionati *con reinserimento*, ammettendo cioè la possibilità di associare lo stesso controllo a più lavoratori interinali.

A questo punto, per ottenere la stima dell'effetto dell'interinale è sufficiente fare la differenza tra la quota che ha trovato un lavoro a tempo indeterminato tra i 511 lavoratori interinali e l'analoga quota osservata tra i 305 controlli abbinati.

10.4 I risultati della valutazione

Il risultato principale della valutazione è molto semplice da illustrare, alla luce dell'impianto metodologico appena delineato. La stima dell'effetto è data dalla differenza tra la probabilità di impiego degli interinali e dei controlli. Tra i 511 individui che hanno svolto una missione di lavoro interinale nel 2001, il 27,8% risulta aver trovato un lavoro a tempo indeterminato entro la seconda metà del 2002. Tra i "controlli" abbinati, che si utilizzano per stimare cosa sarebbe successo agli interinali se non avessero svolto la missione, solo il 13,8% ha avuto un simile successo lavorativo.

La principale conclusione dell'analisi quindi è estremamente semplice: sotto l'ipotesi che le caratteristiche osservate permettano di cogliere tutte le differenze sistematiche tra i due gruppi, si stima che una missione interinale accresce di 14 punti percentuali la probabilità di trovare un lavoro stabile, e quindi la *raddoppia*, poiché la stima della situazione in assenza di interinale è circa del 14%. *Questi risultati suggeriscono che svolgere missioni di lavoro interinale non è una trappola come da molti paventato, anzi facilita la transizione verso un lavoro stabile.*

Gli autori della valutazione non si limitano tuttavia a produrre questa stima complessiva dell'effetto del lavoro interinale. Accanto all'interrogativo complessivo, gli autori pongono alcune domande di valutazione più mirate, chiedendosi in particolare quale sia l'effetto dell'interinale per alcuni specifici sottogruppi di individui. È possibile infatti che i 14 punti percentuali stimati siano semplicemente una media, attorno a cui ci potrebbe essere molta variabilità. Il procedimento per la stima dell'effetto in ciascun sottogruppo è il medesimo mostrato in precedenza: semplicemente, l'abbinamento è effettuato tra i soli interinali e controlli ad esso appartenenti.

I valutatori provano a dare risposta ai seguenti interrogativi:

- a. qual è l'effetto dell'interinale in Sicilia rispetto all'effetto in Toscana?

- b. come cambiano le stime dell'effetto se si distingue, all'interno del gruppo dei controlli, tra coloro che si sono iscritti ad un'agenzia interinale ma non sono andati in missione nel periodo d'interesse e soggetti mai iscritti ad un'agenzia interinale, quindi tra controlli "potenzialmente interinali" e controlli "mai interinali"?
- c. qual è l'effetto dell'interinale rispetto ad altri contratti di lavoro atipico (tutti i contratti a termine differenti dall'interinale, per esempio il lavoro a tempo determinato, i contratti di collaborazione e i contratti di formazione e lavoro)?

Le risposte a questi interrogativi sono riassunte nella Tabella 10.1.

Tabella 10.1: Stime dell'effetto di una esperienza di lavoro interinale per alcuni sottogruppi

	Intero campione	Toscana	Sicilia
Differenza tra interinali e tutti i controlli	+ 0,14	+ 0,22	+ 0,03
Differenza tra interinali e i "mai interinali"	+ 0,10	+ 0,18	- 0,03
Differenza tra interinali e i "potenziali interinali"	+ 0,07	+ 0,20	+ 0,16
Differenza tra interinali e occupati atipici nel gennaio 2001	- 0,01	+ 0,14	- 0,33

Nota: I valori in grassetto indicano stime statisticamente significative al 95% di confidenza
Fonte: Ichino A., Mealli F., Nannicini T. (2003)

La suddivisione territoriale dell'analisi mette in evidenza differenze molto marcate tra gli effetti stimati in Toscana e in Sicilia. L'effetto medio stimato di 14 punti percentuali (riportato nella prima riga della Tabella 10.1) è in verità una media tra un effetto ancora più forte in Toscana, dove i lavoratori interinali sembrano aumentare di 22 punti percentuali la loro probabilità di occupazione stabile, e di un effetto praticamente nullo in Sicilia (la stima riportata nella tabella è di un aumento di 3 punti percentuali, ma non è significativamente diversa da zero).

Gli autori ipotizzano che la causa vada ricercata nella forte differenza tra i due contesti economici. La Sicilia sembra presentare particolari difficoltà nell'offrire lavoro a tempo indeterminato, ottenibile in linea di massima solo grazie a concorsi pubblici. A parziale

conferma dell'ipotesi vanno citati i risultati di un'altra analisi svolta parallelamente dagli autori: se si sposta l'obiettivo dell'analisi dalla probabilità di trovare un'occupazione a tempo indeterminato alla probabilità di trovare un'occupazione qualsiasi, le stime mostrano un effetto decisamente positivo anche in Sicilia (+27%, mentre in Toscana +15%).

Il secondo interrogativo posto dagli autori riguarda strettamente la composizione del gruppo di controllo: ci si chiede quali risultati si otterrebbero considerando separatamente i "mai interinali" e i "potenziali interinali". Nell'analisi appena vista ci si assicura che i controlli non abbiano svolto una missione interinale nel gennaio del 2001: ciò non toglie tuttavia che essi possano aver svolto una missione in un momento successivo, antecedente all'autunno del 2002. Per quanto, in un'eventualità del genere, i controlli in questione abbiano avuto a disposizione meno di 18 mesi per mettere a frutto l'esperienza interinale, è evidente che anche essi avrebbero tratto vantaggio, in misura forse minore, dei benefici della missione, e di conseguenza le stime sarebbero affette da una distorsione.

Per eliminare questo rischio gli autori ripetono l'analisi con due differenti gruppi di controllo: i "mai interinali", che hanno dichiarato di non essersi mai rivolti ad un'agenzia interinale (e dunque non possono avere mai svolto missioni), e i "potenziali interinali", che invece si sono rivolti a un'agenzia interinale, pur non avendo svolto missioni nel gennaio 2001. I risultati, presentati rispettivamente nella seconda e nella terza riga della Tabella 10.1, mostrano come l'effetto sia maggiormente evidente nel confronto con i "mai interinali". I lavoratori interinali hanno, rispetto ai "mai interinali", una probabilità di lavoro a tempo indeterminato superiore del 10%. L'aumento è del 18% se si considerano solo i lavoratori toscani mentre, come in precedenza, in Sicilia non si evidenzia nessun particolare effetto. Per quanto riguarda il confronto con i "potenziali interinali", le stime restano positive ma diventano non significative.

Infine gli autori si domandano se l'interinale aumenta la probabilità di trovare un lavoro stabile rispetto a quella che si sarebbe avuta lavorando con altri contratti atipici. Il gruppo di controllo utilizzato in precedenza raccoglie individui che nel gennaio del 2001 erano disoccupati oppure occupati con contratti atipici. C'è da aspettarsi che gli

esiti occupazionali dei disoccupati siano diversi da quelli degli occupati a termine: questi ultimi potrebbero aver aumentato le proprie opportunità grazie a benefici, simili a quelli dell'interinale, derivanti dallo svolgimento di un lavoro temporaneo. Se questo è vero, i risultati della stima non dovrebbero essere insensibili alla scelta di inserire tra i controlli solo gli occupati a termine o anche i disoccupati. Gli autori ottengono perciò delle stime aggiuntive, utilizzando come controlli i soli occupati nel gennaio 2001, riportate nella quarta riga della Tabella 10.1.

L'effetto dell'interinale rispetto al lavoro con un altro contratto atipico sembra essere nullo. La stima sul campione totale è prossima a zero (-1%) e non significativa; le stime regionali sono di entità decisamente maggiore, e con segni inversi (+14% per la Toscana, -33% per la Sicilia), ma in entrambi i casi le stime non sono significativamente diverse da zero. In conclusione, sembra che una missione interinale aumenti la probabilità di trovare un lavoro a tempo indeterminato nella stessa misura in cui la aumenta un'altra esperienza di lavoro atipico.

Al termine dell'esposizione dei risultati, va fatta una considerazione sull'utilità del metodo del *matching* in questa analisi degli effetti dell'interinale. Ritornando alla prima stima discussa, si evidenziava un aumento di 14 punti percentuali nella probabilità di successo dei lavoratori interinali, ottenuta come differenza tra il 27,8% del campione di interinali e il 13,8% dei controlli abbinati. Gli autori rendono noto che la totalità dei 1.519 lavoratori del gruppo di controllo mostrava una probabilità di occupazione a tempo indeterminato vicina al 14%, sostanzialmente uguale a quella del sottogruppo dei soli abbinati. Pure facendo a meno del *matching*, e confrontando direttamente tutti i trattati con tutti i controlli, si sarebbero dunque ottenuti risultati molto simili. Ciò non significa che l'affinamento del metodo sia stato inutile, perché a priori non si poteva sapere in che misura le differenze di partenza potessero distorcere le stime.

CAPITOLO 11

GLI EFFETTI DELL'INDENNITÀ DI MOBILITÀ SULLA PROBABILITÀ DI RIOCCUPAZIONE DEI LAVORATORI LICENZIATI

di Luca Mo Costabella

Il sistema previdenziale italiano include vari ammortizzatori sociali volti a sostenere il reddito dei lavoratori che, per vari motivi, si trovino a dover affrontare un periodo di disoccupazione in seguito ad un licenziamento collettivo. Fino ai primi anni Novanta tali strumenti erano visti prevalentemente come politica “passiva”, improntata cioè più ad una logica assistenziale che di sostegno all’occupazione (o rioccupazione). In questo ambito la Cassa Integrazione Guadagni Straordinaria era lo strumento largamente più diffuso e rispondeva alla finalità di sostenere il reddito dei lavoratori in esubero nelle aziende in crisi.

La nascita dell’istituto della “mobilità” all’inizio degli anni Novanta sottintende una maggior enfasi sugli strumenti di politica attiva del lavoro. Questo istituto prevede, accanto al sostegno al reddito, forme di incentivo alla rioccupazione del lavoratore licenziato, sotto forma di sussidi marginali all’occupazione a beneficio delle aziende che assumono lavoratori in mobilità. La logica non è esclusivamente quella di sostenere il reddito, ma anche quella di promuovere l’occupazione. Le due componenti, attiva e passiva, coesistendo nel nuovo istituto, possono avere effetti di segno diverso sulla probabilità di rioccupazione del lavoratore licenziato: mentre la componente attiva potrebbe ridurre la durata della disoccupazione, la componente passiva potrebbe prolungarla.

Lo studio di valutazione esaminato in questo capitolo si propone di stimare l’effetto netto che sulla probabilità di occupazione hanno la componente passiva e la componente attiva dell’istituto della mobilità. Il lavoro è stato commissionato nel 2002 dal Ministero del Lavoro e delle Politiche Sociali all’Agenzia Veneto Lavoro ed eseguito da un team congiunto di ricercatori dell’Agenzia e del Dipartimento di Statistica dell’Università di Padova. In particolare, la parte dello studio che riguarda più direttamente la valutazione

degli effetti è stata condotta da un gruppo di statistici economici dell'Università di Padova (Enrico Rettore, Adriano Paggiaro e Ugo Trivellato), che già in passato avevano lavorato su questo tema. Quindi la commessa del Ministero del Lavoro si è raccordata ed ha rafforzato un filone di ricerca valutativa già attivo per iniziativa autonoma di ricercatori universitari.

Il rapporto finale di valutazione è stato pubblicato nel 2004 con il titolo *“Interazione fra sussidi passivi e incentivi al reimpiego: provenienze ed esiti di lavoratori iscritti nelle liste di mobilità”* a cura dell'Agenzia Veneto Lavoro.

11.1 La logica dell'istituto della mobilità

L'istituto della mobilità nasce nel 1991 con lo scopo di regolamentare in modo organico la complessa materia della gestione delle eccedenze di personale, affiancando e sostituendo in parte lo strumento della Cassa Integrazione Guadagni Straordinaria. La mobilità, introdotta dalla legge 223/91 e integrata dalla legge 236/93, combina una componente passiva di sostegno del reddito dei lavoratori licenziati con una componente attiva di sussidi marginali all'occupazione a beneficio dei datori di lavoro che assumono i lavoratori posti in mobilità.

Due sono i casi in cui il legislatore ha previsto il ricorso alla mobilità. La mobilità può essere attivata innanzitutto dalle imprese che occupano più di quindici dipendenti, in seguito al licenziamento collettivo (definito come licenziamento di almeno cinque dipendenti nell'arco di 120 giorni), in conseguenza di una riduzione o trasformazione di attività o di lavoro, o in caso di cessazione di attività. La legge consente inoltre il ricorso alla mobilità alle imprese che abbiano già in atto sospensioni dal lavoro con intervento della Cassa Integrazione Guadagni Straordinaria, nel caso in cui l'imprenditore ritenga di non poter attuare il risanamento o la ristrutturazione necessari al superamento della Cassa.

La procedura è in entrambi i casi sostanzialmente identica. Innanzitutto l'imprenditore deve informare le rappresentanze sindacali aziendali e i sindacati maggiormente rappresentativi dei motivi che impediscono l'adozione di strumenti alternativi al licenziamento. La legge ha fissato alcuni criteri per l'individuazione dei dipendenti in

esuberano (esigenze tecnico-produttive; carichi di famiglia; anzianità di servizio presso l'azienda): tuttavia imprenditore e sindacati possono raggiungere un accordo che individui, tra l'altro, i criteri di scelta dei lavoratori da licenziare in maniera diversa da quelli indicati dalla legge.

I nominativi dei lavoratori licenziati a seguito delle procedure di mobilità vengono comunicati direttamente dall'azienda interessata alla Direzione regionale del lavoro, che compila un'unica *lista regionale* di tutti i lavoratori collocati in mobilità, che viene approvata dalla Commissione regionale tripartita e comunicata agli organi competenti, l'INPS e i Centri per l'Impiego.

La natura dei benefici a favore dei lavoratori licenziati dipende da una serie di parametri, riassumibili in tre ordini di disposizioni: la *durata massima* della permanenza in lista di mobilità; i *benefici per i lavoratori* in mobilità; e i *benefici per le imprese* che assumono lavoratori in mobilità.

La *durata massima* della permanenza in lista dipende dall'età del lavoratore al momento del licenziamento. La durata massima di permanenza è pari a un anno per i lavoratori con meno di 40 anni, a due anni per i lavoratori tra 40 e 49 anni, a tre anni per i lavoratori con più di 49 anni al momento dell'iscrizione. La *ratio* di questa disposizione è evidentemente quella di compensare i lavoratori più anziani per la loro presumibilmente ridotta probabilità di rioccupazione. Alla stessa logica si ispira la disposizione secondo cui nelle Regioni del Mezzogiorno i periodi sopra indicati sono rispettivamente aumentati a 2, 3 e 4 anni per le tre fasce di età.

Il periodo di permanenza in lista può essere prolungato nel caso in cui il lavoratore venga assunto durante il periodo di mobilità con un contratto a tempo determinato, della durata massima di un anno, o *part-time*: in tali casi, l'iscrizione viene “congelata” per un periodo totale non superiore alla durata massima prevista. I lavoratori ultracinquantenni e con un'adeguata anzianità contributiva, possono invece sfruttare la cosiddetta “mobilità di accompagnamento alla pensione”, o “mobilità lunga”, che consente di allungare la permanenza in lista di mobilità, ed i relativi benefici, fino al pensionamento.

I *benefici a favore dei lavoratori* dipendono in modo cruciale dalla dimensione dell'impresa che li ha licenziati. I lavoratori licenziati da *imprese con più di 15 dipendenti* hanno diritto ad un sostegno al reddito, detto *indennità di mobilità*. Tale indennità, corrisposta per il periodo di iscrizione alle liste, è pari per i primi 12 mesi al 100% del trattamento della Cassa Integrazione Guadagni Straordinaria e all'80% per i periodi successivi⁶⁷. Inoltre, i lavoratori che intendano intraprendere una nuova attività di lavoro autonomo o associarsi in cooperativa, hanno il diritto a ottenere la corresponsione anticipata delle mensilità di mobilità non ancora percepite.

I lavoratori licenziati da *imprese con 15 dipendenti o meno* non percepiscono alcuna indennità, ma beneficiano della componente attiva della politica (di cui beneficiano anche i lavoratori licenziati dalle imprese più grandi). La componente attiva è rappresentata da benefici a favore dei datori di lavoro che assumano lavoratori in mobilità. Il principale consiste in una *riduzione degli oneri sociali*: l'impresa che assuma a tempo pieno e indeterminato un disoccupato in mobilità (con o senza indennità) usufruisce infatti per 18 mesi di una consistente riduzione degli oneri sociali per quel lavoratore. La riduzione degli oneri sociali si applica anche nel caso di assunzione a tempo determinato, per la durata massima di un anno. L'impresa può cumulare queste riduzioni assumendo un iscritto alle liste con un contratto a tempo determinato e trasformandolo successivamente in contratto a tempo indeterminato. Inoltre, l'impresa che assume a tempo indeterminato un lavoratore che percepisce l'indennità riceve un bonus pari al 50% dell'indennità residua che sarebbe spettata al lavoratore.

Il forte incentivo a favore delle imprese che assumono ha favorito comportamenti fraudolenti. In alcuni casi infatti si sono verificati licenziamenti fittizi con immediata riassunzione da parte della stessa impresa o di imprese collegate. Per scoraggiare questi comportamenti, con la legge 451/94 i benefici sono stati esclusi nel caso in cui l'assunzione riguardasse *“lavoratori che siano stati collocati in mobilità, nei sei mesi precedenti, da parte di impresa dello stesso o di diverso settore di attività che, al*

⁶⁷ L'indennità viene erogata comunque nei limiti di un importo massimo che è più volte variato nel corso degli anni (per il 2003 era pari a un netto di € 747, elevato a € 898 per i lavoratori che potessero far valere una retribuzione lorda mensile superiore a € 1.712).

momento del licenziamento, presenta assetti proprietari sostanzialmente coincidenti con quelli dell'impresa che assume ovvero risulta con quest'ultima in rapporto di collegamento o controllo”.

11.2 La domanda di valutazione

L'istituto della mobilità rappresenta uno degli strumenti più rilevanti nell'ambito delle politiche del lavoro nazionali: tuttavia, a distanza di oltre un decennio dalla sua introduzione, scarsa resta la conoscenza del rapporto tra l'iscrizione alle liste di mobilità e i successivi esiti occupazionali. Il dubbio principale riguarda l'efficacia della mobilità come strumento volto al reinserimento dei lavoratori licenziati, cioè l'effetto netto della componente attiva e passiva: se da un lato i benefici per le imprese potrebbero indurle ad assumere i lavoratori iscritti alle liste di mobilità, dall'altro l'indennità corrisposta agli iscritti per un lungo periodo di tempo potrebbe innescare una serie di comportamenti volti a ritardare il rientro nel mercato del lavoro regolare.

Gli “esiti occupazionali” di un periodo di disoccupazione possono ovviamente essere intesi in vario modo ed essere misurati, assecondando diversi obiettivi conoscitivi, con diverse grandezze. Ad esempio, l'esito occupazionale potrebbe essere declinato nella probabilità di occupazione a una certa distanza dal licenziamento o dal tempo necessario per trovare il nuovo impiego. Lo studio esaminato adotta la più semplice definizione di esito occupazionale: la proporzione di lavoratori occupati (nel settore privato) in ciascun mese successivo al licenziamento.

I diversi regimi di iscrizione alle liste possono generare diversi risultati negli esiti occupazionali, a prescindere da come questi vengano declinati. Il “trattamento” offerto dall'istituto della mobilità è legato, come si è visto, a due fattori: la dimensione dell'impresa licenziante e l'età del lavoratore al momento del licenziamento. Nel caso di licenziamento da imprese con più di 15 dipendenti è prevista l'indennità a favore del lavoratore e un beneficio a favore delle imprese che lo riassumono. Per i lavoratori provenienti da imprese con meno di 15 dipendenti è prevista unicamente la riduzione degli oneri sociali a favore delle aziende che assumono gli iscritti. L'età del lavoratore

incide invece sulla durata di iscrizione alle liste: maggiore è l'età del lavoratore, maggiore sarà il periodo di permanenza in lista.

La domanda generale riguardo agli effetti della mobilità sulla probabilità di rioccupazione va quindi declinata in interrogativi di maggiore dettaglio, volti a comprendere *come questi effetti varino da un regime di mobilità all'altro*.

Il tratto caratteristico della mobilità, che induce la prima di queste domande valutative “di dettaglio”, è la corresponsione dell'indennità per un periodo definito di tempo. Se da una parte questa indennità rappresenta un ammortizzatore sociale indispensabile per offrire al disoccupato un sostegno al reddito, dall'altro si tratta di capire in che misura essa rappresenti anche un disincentivo all'occupazione o quantomeno un incentivo ad una maggiore (ed eccessiva) selettività delle occasioni di rioccupazione.

Nel declinare la domanda valutativa in questo senso, ci si può chiedere *qual è la differenza tra gli esiti occupazionali dei lavoratori che percepiscono l'indennità e gli esiti che avrebbero sperimentato nel caso in cui tale indennità non fosse stata concessa*. In altre parole, la presenza dell'indennità in quanto tale *aumenta* la durata della disoccupazione e quindi riduce il tasso di occupazione?

Il secondo quesito “di dettaglio” ha a che fare con la *diversa durata dei periodi* in cui i lavoratori possono restare iscritti alle liste di mobilità (e quindi godere della combinazione di indennità e di incentivi all'assunzione). Esiste ampia evidenza a livello internazionale sul fatto che forme prolungate di sostegno al reddito siano disincentivanti nei confronti della rioccupazione. Nel caso italiano, chi è iscritto alle liste per un periodo più esteso di tempo (sia che fruisca dei soli incentivi all'assunzione, sia che, a maggior ragione, goda anche dell'indennità), può essere indotto a cercare un lavoro con minore intensità, o ancora a rifiutare un'offerta di lavoro in attesa che si presenti un'occasione migliore, stimolato dalla prospettiva di poter fruire ancora per lungo tempo dei benefici concessi.

In questo caso, la domanda può essere formulata come: *“qual è la differenza tra gli esiti occupazionali degli iscritti nelle liste per una certa durata massima e quelli che avrebbero sperimentato nel caso in cui la permanenza massima concessa fosse stata di*

durata minore?”. In altre parole, la maggiore durata dei benefici (di natura attiva e/o passiva) per alcuni lavoratori ne aumenta la durata della disoccupazione?

Il peso che questa politica occupa nello scenario del mercato del lavoro nazionale ha stimolato la produzione di altri studi empirici di valutazione della sua efficacia⁶⁸. Gli studi prodotti sulla mobilità sono accomunati dall'utilizzo delle medesime fonti informative (gli archivi regionali delle Liste di mobilità) e dalle medesime domande di valutazione: *quali sono gli effetti legati ad una più o meno generosa tutela dei soggetti posti in mobilità?* Pur utilizzando metodi di stima diversi e basandosi su differenti contesti, sia territoriali che temporali, questi studi conducono in genere al medesimo risultato: a trattamenti “più generosi” a favore dei soggetti posti in mobilità si associa una minore probabilità di transizione verso una occupazione. Il risultato di queste analisi suggerisce che il poter percepire più a lungo i benefici scoraggia la ricerca del lavoro da parte degli iscritti. Lo studio esaminato in dettaglio in questo capitolo si distingue dai precedenti per un utilizzo di informazioni empiriche di migliore qualità e soprattutto per l'uso di metodi analitici che corrispondono allo stato dell'arte in materia.

11.3 Le principali questioni metodologiche e le soluzioni adottate

Il maggiore problema metodologico è, come in altri casi, la difficoltà di costruire un valido gruppo di confronto per stimare l'effetto della mobilità. La mobilità è una politica di tipo universale: quasi tutti i lavoratori licenziati sono iscritti alle liste e hanno diritto a qualche beneficio. Essendo l'iscrizione nelle liste, almeno per i licenziati dalle imprese con più di 15 dipendenti, l'esito di un procedimento amministrativo quasi automatico, non è possibile trovare lavoratori licenziati che non siano per nulla esposti a questa politica pubblica.

⁶⁸ Tra questi, Brunello G., Miniaci R., “Fattore Milano? Un'esplorazione delle liste di mobilità in Lombardia”, in Borzaga C., Brunello G. (a cura di), *L'impatto delle politiche attive del lavoro in Italia*, Edizioni Scientifiche Italiane, Firenze, 1997; Franceschini F., Trivellato U., “L'impatto delle liste di mobilità nel Veneto: un sussidio di disoccupazione o uno strumento di politica attiva del lavoro?”, in Agenzia dell'Impiego del Veneto (a cura di), *Il mercato del lavoro veneto. Tendenze e politiche. Rapporto 1998*, FrancoAngeli, Milano, 1998; Caruso E., “Durata della disoccupazione e probabilità di reimpiego in presenza di incentivi all'assunzione: l'evidenza empirica della lista di mobilità della Regione Umbria”, *Politica Economica*, 2001.

Si potrebbe considerare di utilizzare come gruppo di confronto i lavoratori licenziati in un periodo precedente all'esistenza delle liste di mobilità: tuttavia, anche supponendo di avere informazioni adeguate per effettuare un confronto al netto delle differenti congiunture economiche, le stime sarebbero comunque distorte a causa della presenza di altri provvedimenti, precedenti all'istituto della mobilità, volti alla tutela dei lavoratori licenziati⁶⁹: per cui la differenza tra i due gruppi così costruiti fornirebbe una stima della differenza tra l'effetto della mobilità e l'effetto delle politiche precedenti.

Se è vero che la mobilità è una politica di tipo universale, è tuttavia anche vero che, come abbiamo visto, non tutti i beneficiari ricevono lo stesso trattamento: alcuni ricevono l'indennità, altri no (a seconda della dimensione dell'impresa che li ha licenziati); alcuni possono rimanere in lista per un anno, alcuni per due, alcuni per tre (a seconda dell'età del lavoratore). Gli autori della valutazione hanno scelto di sfruttare queste differenze di trattamento per stimarne gli effetti differenziali. Gli autori si chiedono perciò (i) quale sia l'effetto sulla probabilità di occupazione di un anno aggiuntivo di *permanenza nelle liste*; e (ii) quale sia l'effetto sulla probabilità di occupazione della *percezione dell'indennità*.

Per quanto riguarda la stima degli *effetti dell'indennità*, la valutazione può partire dal confronto tra due gruppi di iscritti alle liste, il primo composto da lavoratori con indennità, il secondo da lavoratori senza indennità. Dal punto di vista metodologico, la sfida è quella di eliminare le differenze sistematiche tra i due gruppi *legate alla dimensione dell'azienda* in cui hanno lavorato prima del licenziamento: coloro che ricevono l'indennità (i trattati) hanno lavorato in imprese sistematicamente più grandi di coloro che non la ricevono (i non-trattati).

Per stimare gli *effetti dell'anno aggiuntivo*, il punto di partenza può essere il confronto di un campione di licenziati con meno di 40 anni di età, per i quali è previsto un solo anno di permanenza in lista, con un campione di licenziati tra i 40 e i 49 anni, ai quali è concesso un anno aggiuntivo. (Sarebbe anche possibile sfruttare le informazioni sui soggetti con più di 50 anni di età, che possono rimanere in lista di mobilità per tre anni, ma per questi ultimi è prevista la possibilità di entrare nella "mobilità lunga", che

⁶⁹ Nel periodo precedente all'entrata in vigore della legge sulla mobilità, i lavoratori licenziati erano comunque coperti da (ancor più generosi) ammortizzatori sociali.

consiste nella fruizione dei benefici della mobilità fino al momento del pensionamento. Le forti peculiarità di questo particolare regime di mobilità hanno indotto gli autori a escludere tutti i soggetti con almeno 50 anni).

Anche in questo caso va risolto il problema delle differenze di partenza: i licenziati con età inferiore ai 40 anni hanno diritto a un solo anno di permanenza, mentre i licenziati con almeno 40 anni hanno diritto a un periodo doppio, per cui i trattati sono sistematicamente *più anziani* dei non-trattati (dove il “trattamento” qui è l’anno aggiuntivo di permanenza in lista).

In entrambi i casi, le eventuali differenze tra i due gruppi nel tasso di occupazione si potrebbero perciò addebitare tanto al differente regime di mobilità (e in tal caso costituirebbero l’effetto differenziale ricercato) quanto alla diversità nelle caratteristiche di partenza, che possono essere osservabili come pure inosservabili. Si rende dunque necessario il ricorso ad appropriate tecniche di stima che permettano di scindere le due componenti, in modo da quantificare l’effetto che interessa. Nelle due sezioni successive si spiegheranno separatamente le strategie adottate dagli autori nei due casi.

La strategia usata per stimare l’effetto dell’anno aggiuntivo di permanenza in lista

La sfida metodologica in questo caso è all’apparenza semplice: è noto che gli appartenenti ai due gruppi (trattati e non-trattati) hanno età sistematicamente diverse. Quindi il meccanismo di selezione è noto. Il vantaggio derivante dalla perfetta conoscenza del meccanismo di selezione è bilanciato dall’impossibilità di trovare un trattato e un non-trattato con le stesse caratteristiche, in quanto avranno sempre età differenti. Per riuscire a istituire un confronto tra individui che siano il più possibile simili, si può concentrare l’attenzione *intorno al punto di discontinuità*, secondo quanto discusso nel capitolo 6 della parte metodologica. Se non è possibile confrontare individui identici rispetto alla caratteristica che causa la selezione (cioè l’età), la scelta più sensata consiste nel confrontare quelli comunque più simili, tra i quali la differenza di età sia minima: i trattati con un’età appena sopra la soglia dei 40 anni (il *punto di discontinuità*) e i non-trattati con un’età appena sotto la soglia dei 40 anni.

Il limite del confronto attorno al punto di discontinuità è duplice: (i) si basa su campioni piccoli, perché non molti sono i lavoratori che hanno circa 40 anni; (ii) produce stime dell'effetto valide solo per gli individui che hanno circa 40 anni: avendo ottenuto le stime basandosi solo su soggetti con quell'età, è difficile generalizzare i risultati a soggetti molto diversi, per esempio ai trentenni o ai quasi cinquantenni.

Per ottenere stime più generalizzabili, gli autori utilizzano un disegno di valutazione inconsueto. Ignorando il vantaggio che deriva dalla conoscenza del meccanismo di selezione, essi ricorrono a una strategia basata sul *matching* statistico (discusso al capitolo 5), abbinando cioè ogni individuo trattato con un individuo "di controllo" il più simile possibile rispetto a una serie di caratteristiche individuali osservate (esclusa l'età). La somiglianza tra le unità è misurata sulla base del *propensity score*, cioè la probabilità di essere assegnati al trattamento (rimanere in mobilità per un anno aggiuntivo) stimata sulla base delle caratteristiche individuali.

Questo modo di procedere ha senso sotto determinate ipotesi. La normativa, che prevede un anno aggiuntivo di mobilità per i meno giovani, è evidentemente volta a facilitare il reinserimento lavorativo dei più svantaggiati tra i licenziati. Se questa condizione di svantaggio fosse approssimabile ricorrendo ad altre caratteristiche differenti dall'età, un metodo di stima che confronti individui simili rispetto a tali caratteristiche sarebbe affidabile quanto un metodo che confronti individui con età simile.

In altre parole, questo metodo funziona (nel senso che fornisce stime corrette) se *condizionatamente al propensity score l'età non influenza la variabile-risultato*. Se ciò si verificasse, si sarebbe riusciti a eliminare le differenze di partenza senza condizionarsi all'età, e dunque si potrebbe stimare correttamente l'effetto dell'anno aggiuntivo con il *matching*. Come conseguenza, i risultati sarebbero generalizzabili ai licenziati di tutte le età.

La strategia per stimare l'effetto dell'indennità di mobilità

Similmente al caso della valutazione dell'effetto dell'anno aggiuntivo, sono noti i meccanismi che conducono all'assegnazione dell'indennità a un lavoratore licenziato: deve provenire da un'impresa con più di 15 dipendenti. In linea con la logica esposta

riguardo al caso precedente, sarebbe possibile condurre un *confronto intorno al punto di discontinuità*. Le differenze nei tassi di occupazione tra i licenziati da imprese con *poco più* di 15 dipendenti e quelli licenziati da imprese con 15 dipendenti o *poco meno* sarebbero, infatti, una buona approssimazione dell'effetto dell'indennità: la distorsione nelle stime dovuta alla diversa dimensione delle imprese di provenienza sarebbe contenuta, perché le stesse differenze tra le dimensioni delle imprese sarebbero contenute, stando tutte intorno alla soglia di 15 dipendenti.

Anche in questo caso esistono però dei limiti, questa volta molto più severi, che spingono gli autori a stimare l'effetto in un altro modo: (i) come in precedenza, i risultati di un confronto intorno al punto di discontinuità non sarebbero generalizzabili, dunque ci si dovrebbe accontentare di un impatto valido per i soli licenziati da imprese di *circa* 15 dipendenti; (ii) in questo caso specifico, il ricorso al confronto intorno al punto di discontinuità rischierebbe di non fornire una stima adeguata neppure per quel sottoinsieme di individui: la soglia di 15 dipendenti costituisce una discontinuità anche per una serie di altre normative di carattere contributivo, fiscale e di tutela dei lavoratori. Un confronto intorno a tale soglia non fornirebbe pertanto una stima dell'effetto dell'indennità di mobilità, bensì una stima dell'impatto di un più complesso insieme di trattamenti; (iii) il limite più critico è di natura strettamente pratica: nei dati utilizzati (descritti nella sezione successiva) non è presente il numero di dipendenti delle imprese licenzianti. Pertanto, il confronto intorno al punto di discontinuità non è attuabile a causa della mancanza delle informazioni necessarie.

La strategia seguita dagli autori è la stessa descritta in precedenza. Essi attuano un confronto tra soggetti con e senza indennità abbinati con una tecnica di *matching* sulla base della loro somiglianza rispetto a una serie di caratteristiche individuali, riassunte nel *propensity score*. I risultati sono corretti se, condizionatamente al *propensity score*, il numero di dipendenti dell'impresa licenziante non influisca sulla variabile-risultato.

I dati utilizzati

La banca-dati utilizzata per questa analisi è ottenuta fondendo due differenti basi di dati, allo scopo di ottenere il maggior numero possibile di informazioni sulle caratteristiche individuali degli iscritti, la loro storia lavorativa, i tipi di mobilità a cui sono assegnati e

le caratteristiche delle imprese licenzianti: l'archivio regionale veneto degli iscritti alle liste di mobilità e l'archivio amministrativo dei Centri per l'Impiego, denominato Netlabor.

Per ogni lavoratore licenziato, nell'archivio regionale vengono registrate una serie di caratteristiche demografiche, dati sull'ultimo episodio lavorativo e informazioni sull'iscrizione alle liste di mobilità (data di iscrizione, diritto o meno all'indennità). Il record individuale può successivamente essere aggiornato per segnalare l'uscita definitiva dell'iscritto dalle liste.

L'utilizzo dell'archivio Netlabor permette di arricchire notevolmente i dati dell'archivio regionale. Nell'archivio Netlabor di ogni Centro per l'Impiego sono registrate informazioni riguardanti tutti gli episodi di lavoro dipendente avvenuti all'interno della circoscrizione del Centro, tutte le imprese che abbiano comunicato al Centro almeno un avviamento e tutti i lavoratori oggetto di avviamento. Pertanto è possibile, almeno in via teorica, abbinare (utilizzando i codici fiscali) i record di iscrizione individuali presenti negli archivi regionali delle liste di mobilità con le rispettive informazioni presenti in Netlabor, ottenendo un nuovo archivio che, oltre alle citate informazioni presenti nelle liste, contenga la storia lavorativa dell'individuo (elenco degli episodi lavorativi, di cui si conoscono data di avviamento e di fine rapporto, tipo di contratto, qualifica, ragione sociale e settore di appartenenza dell'impresa). In sintesi, agli archivi di gestione delle liste è lasciato il compito di fornire le informazioni sull'entrata in mobilità degli iscritti, mentre da Netlabor si traggono tutte le informazioni sulla storia lavorativa e sulle caratteristiche individuali.

Il contesto territoriale di partenza è quello delle Province di Belluno, Treviso e Vicenza, selezionate per via della buona qualità dei dati dei loro Centri per l'Impiego. Dalle liste di mobilità di queste Province sono estratte inizialmente le informazioni di 14.872 lavoratori licenziati tra il 1995 e il 1999. I dati sugli iscritti alle tre liste provinciali sono abbinati a quelli degli archivi Netlabor dei rispettivi Centri per l'Impiego. Al termine della fase di abbinamento, durante la quale sono stati eliminati 2.111 individui non presenti in Netlabor oppure per i quali non era registrato un licenziamento prossimo alla data di entrata in mobilità, si dispone di un campione di 12.761 iscritti.

Successivamente si limita l'osservazione alle Province di Treviso e Vicenza, per le quali gli archivi Netlabor permettono l'osservazione delle storie lavorative per un anno in più (fino al 2001) rispetto a Belluno. L'osservazione è inoltre limitata ai soli iscritti nelle liste negli anni 1997 e 1998. Considerato che i dati Netlabor sugli avviamenti sono ritenuti affidabili dagli autori a partire dal 1995, le limitazioni territoriali e temporali adottate permettono di disporre, per ogni iscritto osservato, di due anni di storia lavorativa precedente all'entrata in lista e tre anni di storia successiva.

Dal campione rimanente sono infine esclusi gli iscritti con almeno 50 anni, per i quali è possibile transitare direttamente nel pensionamento, e un certo numero di lavoratori licenziati e riassunti secondo modalità che suggeriscono comportamenti fraudolenti da parte delle imprese coinvolte: la loro storia lavorativa sarebbe viziata e sistematicamente diversa dal comportamento dei lavoratori che devono effettivamente trovare un nuovo lavoro. L'analisi finale è condotta su un campione di 4.230 individui, le cui principali caratteristiche sono rappresentate nella Tabella 11.1.

Tabella 11.1: Caratteristiche del campione utilizzato per l'analisi

Età	Uomini		Donne		Totale
	Con indennità	Senza indennità	Con indennità	Senza indennità	
<40 anni	445	922	347	1.487	3.201
40-49 anni	276	369	141	243	1.029
Totale	721	1.291	488	1.730	4.230

Fonte: Agenzia Veneto Lavoro (2003)

Sfruttando le informazioni sugli episodi lavorativi, per ognuno dei 4.230 iscritti si costruisce una serie storica mensile di cinque anni (i 24 mesi precedenti al licenziamento e i 36 successivi) relativa al “tasso di occupazione”, calcolato come quota dei giorni di ogni mese spesi nella condizione di occupazione. I tassi dei 36 mesi successivi all'entrata in lista sono utilizzati come variabile-risultato. Invece, i tassi di occupazione dei 24 mesi precedenti al licenziamento entrano a far parte delle variabili di controllo, utilizzate per confrontare gli individui ed abbinare quelli più simili. Le variabili di controllo utilizzate sono l'età (solo nel caso della stima dell'impatto dell'indennità); Provincia di residenza e titolo di studio; qualifica e settore di attività

relativi all'ultimo episodio lavorativo precedente al licenziamento e la storia lavorativa precedente, rappresentata dai tassi di occupazione nei 24 mesi precedenti al licenziamento.

Nella lista delle variabili non compare il sesso. Ciò è dovuto al fatto che, per tenere in considerazione ed evidenziare le possibili differenze nei risultati, tutte le analisi sono condotte separatamente per maschi e femmine.

Utilizzando gli stessi abbinamenti, è possibile stimare l'effetto differenziale dell'indennità o dell'anno aggiuntivo su un'ampia gamma di variabili-risultato, cercando di fornire risposte a domande su vari aspetti del fenomeno in esame. Gli autori forniscono ad esempio stime dell'effetto, oltre che sui tassi di occupazione osservati nei tre anni successivi al licenziamento, anche sulla probabilità di trovare un lavoro a tempo indeterminato e sui tempi di attesa per trovare un primo lavoro. Nel seguito sono descritte solamente le stime dell'effetto sui tassi di occupazione.

11.4 I risultati della valutazione

I risultati della valutazione sono presentati separatamente per le due domande di valutazione: l'effetto dell'anno aggiuntivo di permanenza in lista e l'effetto dell'indennità.

L'effetto dell'anno aggiuntivo di permanenza in lista

Per stimare l'effetto dell'anno aggiuntivo, si comincia calcolando per ogni individuo il *propensity score*, in questo caso la probabilità di beneficiare di un anno aggiuntivo di permanenza in lista, stimata in funzione delle caratteristiche osservabili con un modello *logit*. Successivamente, ad ogni iscritto con due anni di permanenza in lista (trattato) è abbinato il lavoratore con un anno di permanenza in lista (quindi non-trattato dall'anno aggiuntivo) con il *propensity score* più vicino (scelto tra coloro che hanno un *propensity score* distante al massimo 0,01)⁷⁰. Al termine della procedura di *matching*, si ottengono 925 abbinamenti, corrispondenti all'89,9% degli individui trattati. Per il restante 10,1%

⁷⁰ Si tratta di una combinazione di due tecniche di abbinamento descritte al capitolo 5, abbinamento in un raggio e abbinamento sull'unità più vicina. Questo stringente criterio di *matching* cautela gli autori dal rischio di abbinare due individui che siano troppo differenti.

non è stato possibile trovare un iscritto senza anzianità abbastanza simile (con un *propensity score* distante non più di 0,01).

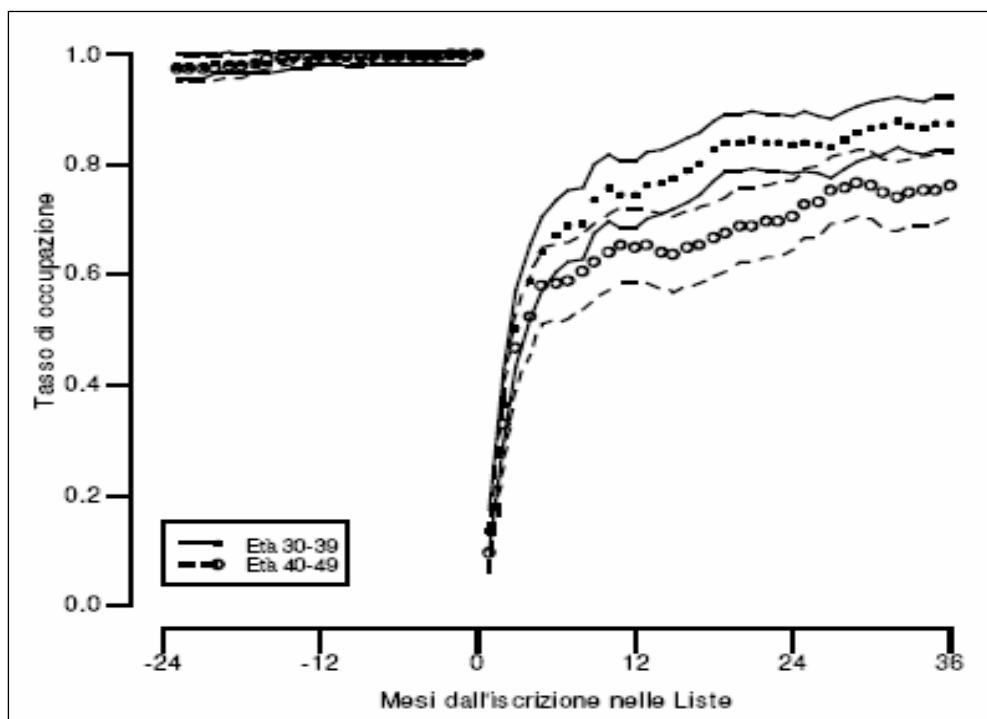
Prima di procedere alla stima dell'effetto dell'anno aggiuntivo, ricordando i dubbi sulla validità delle stime di *matching* formulati in fase di definizione della strategia d'analisi, gli autori cercano di verificare se, condizionatamente al *propensity score*, l'età influenzi il tasso di occupazione. Per questa verifica vengono posti a confronto i tassi degli iscritti in età 20-29 con quelli degli iscritti (abbinati con *propensity score*) in età compresa tra i 30-39 anni. Essendo entrambi sottoposti alla stessa durata di permanenza in lista (un anno), ed essendo simili rispetto al *propensity score*, se l'età non avesse alcun effetto non si dovrebbero osservare differenze nei tassi di occupazione, né prima né dopo il licenziamento. Se invece rimanessero delle differenze tra ventenni e trentenni, si dovrebbe concludere che con ogni probabilità rimangono differenze anche tra iscritti under 40 e over 40.

I risultati evidenziano la permanenza di differenze, che attribuiscono ai più giovani un maggiore tasso di occupazione, suggerendo il rischio di distorsione delle stime dell'effetto dell'anno aggiuntivo. Tali differenze sono più evidenti per le donne. Successive analisi evidenziano differenze più contenute (per quanto non totalmente assenti) se si restringe l'osservazione agli iscritti tra i 30 e i 39 anni. *Gli autori decidono quindi di stimare l'effetto abbinando i 40-49enni con anno aggiuntivo con i soli 30-39enni senza anno aggiuntivo.* La nuova procedura di *matching* produce 840 abbinamenti, pari all'81,6% degli esposti all'anno aggiuntivo.

Le prossime figure mostrano graficamente alcuni dei risultati presentati nel rapporto, a cui si rimanda per un'esposizione completa. Le figure vanno lette nel modo seguente. Sull'asse verticale è riportata la variabile-risultato, il tasso di occupazione; sull'asse orizzontale il tempo espresso in mesi: alla destra dello zero sono rappresentati i 36 mesi successivi al licenziamento (e iscrizione nella lista di mobilità); alla sinistra dello zero (con il segno negativo) i 24 mesi precedenti. Le linee rappresentate da "pallini" (pieni e vuoti) rappresentano *le stime* del tasso di occupazione mensile per i due gruppi messi a confronto (indicati nella legenda), mentre le due linee continue e le due linee

tratteggiate delimitano i rispettivi intervalli di confidenza al 95%. L'effetto stimato è quindi la distanza verticale tra le due linee di "pallini".

Figura 11.1: Stima dell'effetto dell'anno aggiuntivo sugli iscritti maschi con indennità (Confronto tra i tassi di occupazione dei gruppi abbinati mediante *propensity score*)



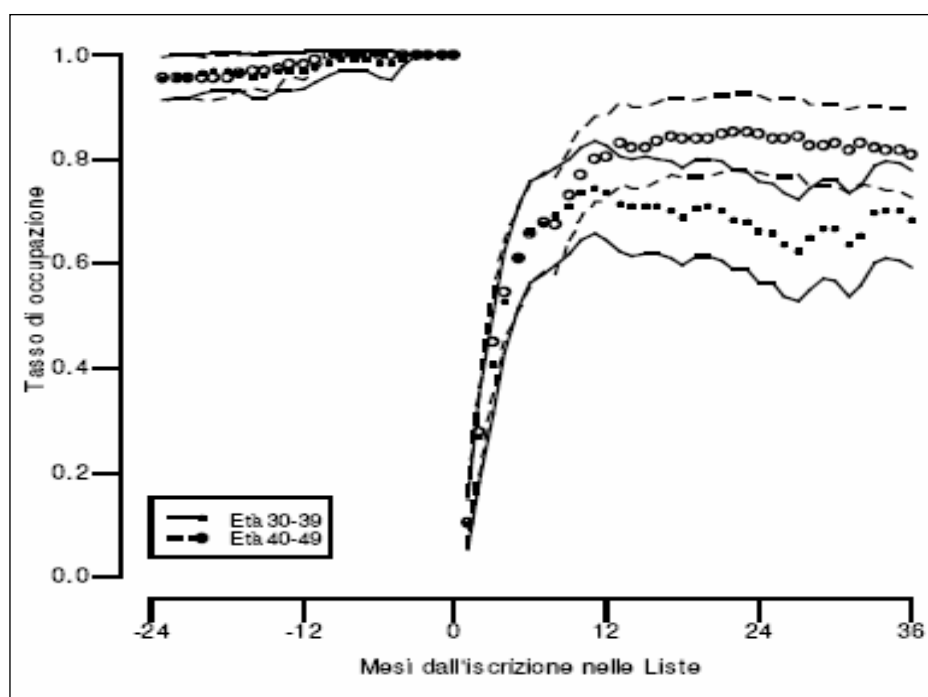
Fonte: Agenzia Veneto Lavoro (2003)

La Figura 11.1 mostra, limitatamente alla popolazione maschile, un effetto significativo dell'anno aggiuntivo di permanenza in lista di mobilità. Tra coloro che ricevono l'indennità, si stima un *effetto negativo* dell'anno aggiuntivo. I beneficiari dell'anno aggiuntivo (40-49enni) hanno un tasso di occupazione inferiore di 10-20 punti percentuali rispetto ai non beneficiari (30-39enni abbinati). Tale differenza è statisticamente significativa nel secondo e nel terzo anno successivi all'entrata in mobilità⁷¹.

⁷¹ Le storie lavorative precedenti dei due gruppi sono sempre sovrapposte. Questo perché i tassi di occupazione precedenti sono stati usati come variabili di *matching*, creando gruppi di iscritti con storie simili.

Tra i maschi non percettori di indennità, mostrati nella Figura 11.2, i risultati sono opposti. L'anno aggiuntivo di mobilità sembra in questo caso favorire le *chance* di rientro nel mercato del lavoro, con un aumento del tasso di occupazione del 20% circa⁷².

Figura 11.2: Stima dell'effetto dell'anno aggiuntivo sugli iscritti maschi senza indennità (Confronto tra i tassi di occupazione dei gruppi abbinati mediante *propensity score*)



Fonte: Agenzia Veneto Lavoro (2003)

Da questo doppio confronto emerge, quindi, il fatto che ricevere per un anno aggiuntivo un pacchetto di benefici che include l'indennità è associato ad un rientro *più lento* nel mercato del lavoro, mentre un anno aggiuntivo del solo incentivo alla rioccupazione è associato ad un rientro *più veloce*.

Il confronto relativo alla popolazione femminile (non mostrato in questa sintesi) evidenzia differenze della stessa natura osservata per gli uomini. L'effetto dell'anno aggiuntivo sembra essere negativo in presenza di indennità e positivo altrimenti. Le differenze sono però molto più contenute rispetto a quanto rilevato per gli uomini, oltre a non essere statisticamente significative.

⁷² A causa della bassa numerosità campionaria, però, le stime non mostrano significatività statistica.

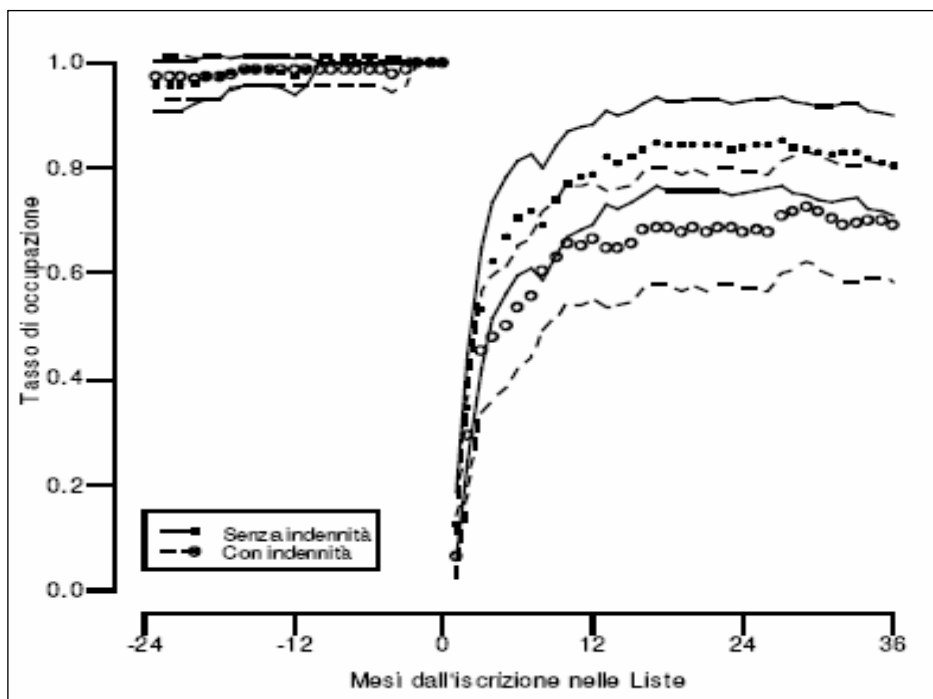
In conclusione, per gli iscritti senza indennità l'anno aggiuntivo sembra costituire (almeno per gli uomini) uno strumento utile a rientrare nel mercato del lavoro. Le evidenze sono opposte per i lavoratori che fruiscono dell'indennità. L'utilizzo congiunto delle evidenze suggerisce che la componente passiva della politica tende a prevalere sulla componente attiva.

L'effetto stimato dell'indennità

Per stimare l'effetto dell'indennità sul tasso di occupazione, si calcola per ogni individuo il *propensity score*, cioè la stima della probabilità di ricevere l'indennità in funzione delle variabili di controllo. Ad ogni iscritto con indennità è associato, tra tutti gli iscritti senza indennità con *propensity score* distante al massimo 0,01, quello con il *propensity score* più vicino.

In questo caso, tuttavia, l'abbinamento risulta più complicato di quello effettuato in precedenza. Gli autori sottolineano la forte differenza tra i *propensity score* dei trattati e dei controlli, che si traduce in una maggiore difficoltà nel reperire coppie di iscritti simili. Al termine della procedura di *matching* sono stati effettuati 676 abbinamenti, pari al 47,7% degli iscritti con indennità. Nelle figure successive si presentano le stime dell'effetto dell'indennità separatamente per genere ed età.

Figura 11.3: Stima dell'effetto dell'indennità sul tasso di occupazione degli uomini in età compresa tra i 40-49 anni (Confronto tra i tassi di occupazione dei gruppi abbinati mediante *propensity score*)

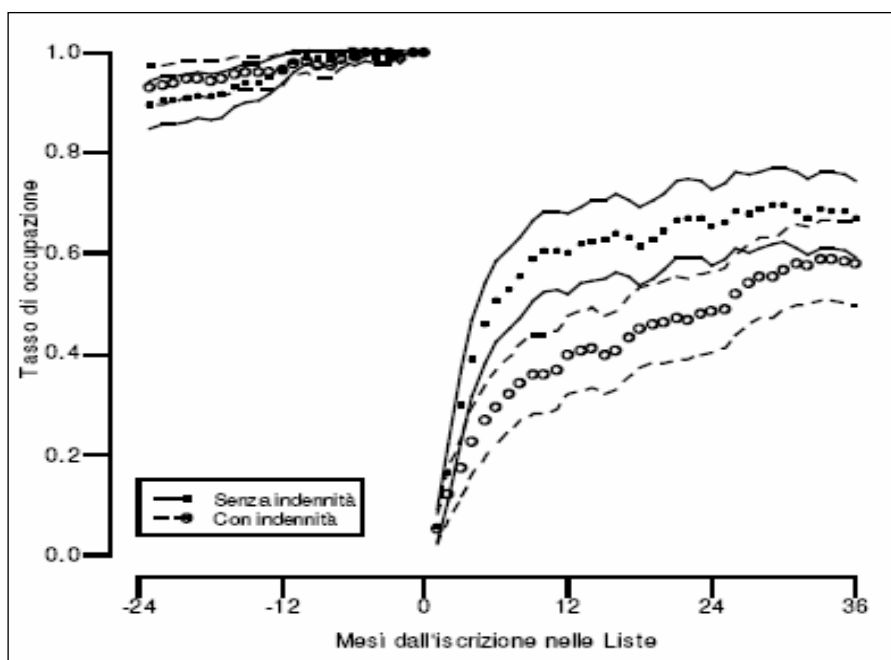


Fonte: Agenzia Veneto Lavoro (2003)

La Figura 11.3 illustra i risultati per gli uomini tra i 40 e i 49 anni di età ed evidenzia una stima dell'effetto palesemente negativa: i percettori dell'indennità hanno un tasso di occupazione inferiore di circa 20 punti percentuali rispetto ai loro pari senza indennità. La differenza negativa, che si manifesta a partire dal dodicesimo mese dall'entrata in lista, tende a ridursi nell'ultimo anno, ma tra percettori e non percettori resta un divario di 15 punti.

La Figura 11.4 presenta il caso delle donne tra i 40 e i 49 anni di età. Le stime d'impatto non sono molto differenti da quelle ottenute per gli uomini. Tra donne con e senza indennità c'è, a parità di condizioni, una differenza di 20 punti nel tasso di occupazione.

Figura 11.4: Stima dell'impatto dell'indennità sul tasso di occupazione delle donne in età compresa tra i 40-49 anni (Confronto tra i tassi di occupazione dei gruppi abbinati mediante propensity score)



Fonte: Agenzia Veneto Lavoro (2003)

Tale differenza è ben evidente già nel primo anno dopo il licenziamento. Anche in questo caso però i due gruppi si riavvicinano nel terzo anno: rimane una differenza di circa 10 punti percentuali, ma non statisticamente significativa.

L'analisi sulla popolazione più giovane fornisce risultati abbastanza diversi da quelli evidenziati per gli ultraquarantenni. Le donne più giovani mostrano, al pari delle donne con più di 40 anni, un effetto fin da subito negativo, ma di entità molto minore. La differenza tra donne con e senza indennità tende inoltre a riassorbirsi già alla fine del primo anno, suggerendo, per confronto con il caso delle donne meno giovani, che l'impatto negativo dell'indennità scompare al termine del periodo di erogazione.

Il caso degli uomini più giovani costituisce una particolare eccezione ai risultati negativi finora presentati. I trentenni percettori di indennità hanno un tasso di occupazione sistematicamente maggiore dei non percettori. Inoltre tale differenza non si riassorbe nel tempo, ma sembra aumentare fino all'ultimo anno, in cui raggiunge la sua massima entità (circa 15 punti percentuali). Riguardo a questa stima gli autori esprimono qualche

dubbio, notando come essa sia da addebitare non tanto a una maggiore crescita del tasso di occupazione per gli iscritti con indennità, ma piuttosto a un singolare arresto nel gruppo di controllo osservabile a partire dalla metà del primo anno. Il sospetto degli autori è che, in questo caso specifico, una considerevole quota di iscritti senza indennità si sia ricollocata con qualche tipo di lavoro autonomo, non osservabile perché non registrato negli archivi dei Centri per l'Impiego.

CAPITOLO 12

GLI EFFETTI DELLE BORSE DI STUDIO SULLA REGOLARITÀ DEGLI STUDI UNIVERSITARI

di Luca Mo Costabella

Le borse di studio rappresentano uno strumento per agevolare il diritto allo studio, soprattutto a livello universitario. L'obiettivo delle borse di studio è innanzitutto quello di ridurre il rischio che l'accesso all'istruzione universitaria sia ostacolato dalle condizioni economiche disagiate della famiglia di provenienza dello studente. Tuttavia, l'uguaglianza nelle opportunità di accesso non rappresenta l'unico obiettivo perseguito da questo intervento: la fruizione di una borsa di studio dovrebbe anche consentire a chi la riceve un percorso di studi più regolare e soprattutto una minore probabilità di abbandono degli studi.

La valutazione discussa in questo capitolo si focalizza su questo secondo aspetto: tenta cioè di valutare l'effetto che le borse di studio hanno sulla regolarità delle carriere degli studenti universitari. Lo studio è stato commissionato nel 2000 dal Comitato Nazionale per la Valutazione del Sistema Universitario all'Istituto Regionale per la Programmazione Economica della Toscana (IRPET), che ha operato in collaborazione con alcuni ricercatori del Dipartimento di Statistica dell'Università di Firenze. L'analisi empirica è stata condotta sugli iscritti di 11 atenei italiani⁷³ negli anni accademici 1998/99, 1999/00 e 2001/02 e fornisce evidenze riguardo all'influenza delle borse di studio su una varietà di dimensioni connesse alla carriera degli studenti.

Il rapporto di valutazione è stato pubblicato nel 2004 dall'IRPET con il titolo "*La valutazione di efficacia delle borse per il diritto allo studio*"⁷⁴.

⁷³ Università di Cagliari, Università della Calabria, Università di Camerino, Università di Catania, Università di Ferrara, Università di Firenze, Politecnico di Milano, Università di Padova, Università "La Sapienza" di Roma, Università di Salerno, Università di Trieste.

⁷⁴ Una descrizione più sintetica della ricerca si trova nel volume *L'efficacia delle politiche di sostegno agli studi universitari*, a cura di Biggeri L., Catalano G., Quaderni per il Comitato Nazionale della Valutazione del Sistema Universitario, in corso di pubblicazione, il Mulino, Bologna.

12.1 La logica delle borse di studio

L'istituzione delle borse di studio si fonda sull'ipotesi che un fondamentale ostacolo all'accesso e al successo universitario sia di natura economica. Pur tralasciando possibili aspetti disincentivanti costituiti dalle mancate opportunità di guadagno e dall'incertezza dei benefici economici derivanti dall'eventuale conseguimento della laurea, un individuo che intenda frequentare l'università (e/o la sua famiglia) deve comunque affrontare una serie di oneri, tra cui le tasse universitarie, l'acquisto di materiali di studio, le spese per gli spostamenti o per l'alloggio, nel caso in cui la sede universitaria sia distante dalla residenza.

In assenza di apposite fonti di sostegno, la frequenza degli studi universitari sarebbe dunque molto meno agevole per coloro che si collocano nelle fasce di reddito medio-basse. Le politiche volte all'eliminazione delle barriere economiche perseguono obiettivi di equità di due tipi. La prima è un'equità di tipo *orizzontale*, che concerne direttamente l'uguaglianza delle opportunità di accesso all'istruzione. La seconda è un'equità di tipo *verticale*, connessa alla futura distribuzione dei redditi: facilitando l'accesso all'università delle classi economicamente più svantaggiate si innalza il loro livello di istruzione e si migliora di conseguenza la loro condizione economica successiva.

L'accesso alle borse di studio è regolato da procedure variabili da ateneo ad ateneo, ma fondate sui medesimi criteri di massima, stabiliti per legge. Gli studenti che richiedono una borsa di studio sono sottoposti ad un vaglio di idoneità, dipendente da fattori di merito e di condizione economica. Il merito è approssimato dal numero di esami sostenuti negli anni precedenti o, per le matricole, dal voto di maturità⁷⁵. La situazione economica della famiglia dello studente è solitamente definita dal reddito e, in tempi recenti, da un indicatore più complesso, l'ISEE (Indicatore della Situazione Economica Equivalente), che viene calcolato tenendo conto del reddito, del patrimonio e della dimensione del nucleo familiare. In alcuni atenei il numero degli idonei risulta troppo elevato perché a tutti si possa erogare una borsa, rispettando i vincoli di bilancio. In

⁷⁵ Il criterio fondato sul voto di diploma è stato tuttavia utilizzato solo per gli anni precedenti al 2001. Negli anni successivi la selezione degli studenti al primo anno è stata basata esclusivamente sugli indicatori di condizione economica.

questo caso gli idonei sono inseriti in una graduatoria, sulla base della quale vengono erogate le borse di studio, fino ad esaurimento dei fondi.

Una certa eterogeneità esiste sul fronte dell'ammontare della borsa: se in alcuni casi le borse di studio hanno un ammontare identico per tutti gli studenti, o sono caratterizzate da una minima variabilità, esistono altri contesti in cui sono previsti differenti regimi di erogazione determinati da differenti condizioni economiche o dalla situazione abitativa dei beneficiari (se siano cioè residenti in sede, fuori sede o pendolari).

12.2 La domanda di valutazione

Come già argomentato in precedenza, favorire l'accesso all'università non è l'unico obiettivo perseguito da questo strumento di diritto allo studio. La disponibilità di una borsa di studio, una volta che lo studente che la riceve si sia iscritto all'università, dovrebbe anche favorire la sua regolare progressione negli studi.

È possibile infatti che su queste due distinte dimensioni (accesso all'università e regolarità degli studi) le borse di studio abbiano effetti di segno diverso. Una tra le ipotesi possibili è che le borse non influenzino in modo determinante l'accesso all'università (chi vuole iscriversi deciderebbe di farlo comunque), ma, una volta che lo studente si sia iscritto, il fatto di disporre di un reddito minimo gli consenta di utilizzare il proprio tempo in maniera più consona alle esigenze poste dagli studi universitari e conseguentemente di procedere in modo più regolare. All'opposto, è possibile che le borse di studio favoriscano l'accesso all'università senza però avere un'influenza positiva sulla regolarità degli studi. Essendo questi effetti, almeno potenzialmente, di segno e di intensità diversi, essi motivano domande di valutazione distinte.

La domanda di valutazione a cui lo studio tenta di dare risposta è centrata sulla dimensione della regolarità ed è così sintetizzabile: "La fruizione di una borsa di studio da parte degli studenti iscritti all'università influenza positivamente la loro regolarità negli studi, rispetto alla situazione in cui gli stessi studenti non avessero accesso alla borsa?".

Gli autori individuano quattro differenti dimensioni della regolarità degli studi, alle quali si possono associare diverse variabili-risultato:

- un primo aspetto è costituito dalla *persistenza* negli studi, che pone l'attenzione sulla possibilità che gli studenti abbandonino l'università prima del conseguimento della laurea. Per tenere conto di questo fenomeno, la variabile-risultato potrebbe essere la probabilità di abbandono entro un certo numero di anni dall'iscrizione;
- un secondo aspetto è rappresentato dalla *durata* degli studi ed è in sintesi rappresentabile dal numero di anni impiegati da uno studente per conseguire la laurea;
- la terza dimensione è rappresentata dalla *produttività* del tempo trascorso nella condizione di studente, dimensione che si può rappresentare con il numero di esami sostenuti da uno studente entro un certo periodo dall'iscrizione;
- l'ultima dimensione è quella della *qualità* degli studi, cioè la misura in cui si sono aumentate le conoscenze e le competenze dello studente, che si può (molto imperfettamente) approssimare con le votazioni ottenute negli esami.

Gli autori hanno scelto di concentrare l'attenzione sugli aspetti della persistenza e della produttività. La domanda valutativa è quindi circoscritta agli effetti che le borse di studio hanno sui tassi di abbandono dell'università e sul numero di esami sostenuti. L'esclusione delle votazioni riportate negli esami deriva dalla forte eterogeneità nelle procedure di raccolta e di calcolo dei vari atenei⁷⁶. La durata degli studi non può essere invece presa in considerazione perché i dati disponibili sulle carriere universitarie sono limitati ai primi due anni accademici.

In definitiva, le variabili-risultato sulle quali si intende stimare l'effetto della borsa di studio sono: (i) la probabilità di abbandonare l'università dopo il primo anno; (ii) la probabilità di non sostenere esami al primo anno; (iii) il rapporto tra il numero di esami sostenuti e il numero di esami previsti al secondo anno.

La prima nozione di effetto esplorata è quella "classica", cioè la differenza tra il valore della variabile-risultato per gli studenti con borsa di studio e il valore della variabile-risultato che si sarebbe osservato per gli stessi studenti se non avessero ricevuto la borsa. I valutatori non si fermano a questa nozione di effetto, ma si chiedono anche quale possa essere l'eventuale effetto differenziale connesso a diverse *modalità di*

⁷⁶ Gli archivi messi a disposizione dalle Università contengono dati relativi talvolta ad annualità, talvolta ad esami semestrali, e impediscono la creazione di una variabile sufficientemente omogenea e confrontabile.

percezione della borsa di studio. In particolare, le analisi sono volte a capire: (a) come vari l'effetto sulle diverse variabili-risultato al variare dell'importo della borsa, e (b) come vari l'effetto sul numero di esami sostenuti se la borsa di studio viene percepita per un solo anno (solo al primo oppure solo al secondo) oppure per due anni consecutivi. Questo secondo aspetto è indagato solo in relazione agli aspetti della produttività. L'effetto sulla probabilità di abbandono, infatti, viene osservato solo alla fine del primo anno e non è dunque influenzato dall'eventuale secondo anno di borsa.

12.3 Le principali questioni metodologiche e le soluzioni adottate

La strategia di base utilizzata per la valutazione degli effetti prevede un confronto fra un gruppo di trattati e un gruppo di non-trattati: si considerano trattati coloro che hanno ricevuto una borsa di studio (borsisti), mentre entrano nel gruppo di confronto coloro che non l'hanno ricevuta (non borsisti). Per una qualunque delle variabili-risultato d'interesse, un confronto tra borsisti e non borsisti successivo all'erogazione delle borse di studio evidenzierebbe differenze imputabili tanto alle borse di studio quanto a preesistenti differenze sistematiche tra i due gruppi. Tali differenze sistematiche derivano dal processo di selezione in virtù del quale rientrano nel gruppo dei beneficiari solo soggetti in possesso di specifiche caratteristiche.

Il primo passo del processo di selezione è rappresentato dalla presentazione della domanda per la concessione della borsa. Ciò consente di operare una prima importante distinzione tra richiedenti e non richiedenti. In seconda battuta, tra i richiedenti vengono individuati i beneficiari con una procedura distinta in due fasi:

- dapprima vengono selezionati gli studenti considerati *idonei* a ricevere la borsa di studio, in base ai criteri di merito e/o condizione economica descritti in precedenza;
- in un secondo momento gli idonei sono inseriti in una graduatoria, ancora fondata su fattori economici e di merito, sulla base della quale sono erogate le borse fino a esaurimento dei fondi.

I meccanismi di selezione e auto-selezione appena illustrati renderebbero il confronto tra borsisti e non borsisti viziato da differenze in termini di motivazioni (un borsista era a conoscenza della possibilità di ricevere una borsa e ha voluto presentare domanda per

ottenerla), di capacità e impegno (un borsista appartiene a una fascia di studenti “più bravi”, pertanto meritevoli di ricevere la borsa) e di possibilità economiche (un borsista appartiene a una fascia di studenti in condizioni economiche relativamente svantaggiate).

La conoscenza dei processi di selezione e delle differenze tra gruppi da questi introdotte, può essere sfruttata per selezionare dalla totalità dei non borsisti un gruppo di confronto *ad hoc*, in maniera tale da ridurre le fonti di distorsione nella stima dell’effetto. Gli autori individuano in particolare due possibilità:

- (a) un gruppo di confronto formato da *studenti che hanno presentato domanda*, sono stati ritenuti idonei, ma non hanno ricevuto la borsa perché i fondi a disposizione non erano sufficienti. In un tale caso, tra il gruppo dei trattati e il gruppo di confronto non esistono differenze legate alla motivazione, in quanto entrambi i gruppi hanno presentato domanda per ricevere la borsa di studio. Entrambi i gruppi sono idonei, dunque rispettano i requisiti minimi di merito e di condizione economica prefissati, tuttavia la differente posizione in graduatoria implica che i trattati siano sistematicamente più bravi e/o meno abbienti;
- (b) un gruppo di confronto formato da *studenti che non hanno presentato la domanda*, ma le cui condizioni economiche e il merito sarebbero stati tali da soddisfare i requisiti di idoneità. In questo caso, tra gruppo di trattamento e gruppo di confronto esistono sistematiche differenze di motivazione, che spiegano il differente comportamento rispetto alla presentazione della domanda.

I differenti rischi di distorsione da selezione connesse all’utilizzo di un gruppo di confronto del primo o del secondo tipo portano gli autori a ipotizzare il ricorso a due differenti metodologie di valutazione.

Nel caso in cui si ricorra a un gruppo di confronto di *richiedenti idonei ma esclusi dalla borsa*, il confronto è influenzato dalle differenze di merito e di condizione economica: gli idonei che presentano valori entro un certo limite sono esclusi dal beneficio della borsa, mentre quelli che la superano sono ammessi. Esiste perciò una *soglia* che discrimina i “trattati” dai “non-trattati”. Ma intorno alla soglia si trovano (da un lato) borsisti e (dall’altro) idonei non borsisti con caratteristiche molto simili. Per gli individui intorno a questo *punto di discontinuità* si può affermare, viste le differenze

minime che li contraddistinguono, che il processo di assegnazione della borsa sia stato quasi casuale, per cui si può operare un confronto non affetto (o meglio, affetto in misura minima) da distorsione da selezione.

Come visto nel capitolo 6 della parte metodologica, non esiste una sola tecnica per ottenere stime dell'effetto sfruttando tale discontinuità. La scelta degli autori è di ricorrere al *regression discontinuity design*, cioè un modello di regressione che, opportunamente specificato, consente di cogliere l'effetto della borsa di studio separatamente da quello delle caratteristiche che hanno determinato la graduatoria. Tale modello potrebbe essere specificato includendo tra le variabili esplicative solo il punteggio che ha determinato la graduatoria e la variabile che identifica il trattamento (una variabile dicotomica che identifica gli stati di borsista e non borsista), perché è noto che non esistono altri fattori che possano differenziare borsisti e non borsisti. Tuttavia spesso (e così fanno gli autori di questo studio) si inserisce nel modello anche una serie di caratteristiche individuali, ad esempio demografiche, allo scopo di non trascurare le eventuali differenze impreviste fra trattati e non-trattati.

L'utilizzo dell'informazione sui criteri di formazione della graduatoria comporta il vantaggio di poter eliminare con una certa sicurezza le fonti di distorsione, ma porta con sé un inevitabile svantaggio: ciò che accomuna tutte le tecniche di stima intorno al punto di discontinuità è che *i risultati non possono essere generalizzati a soggetti che siano troppo distanti dalla soglia*. Infatti, la soglia è l'unica "zona" della graduatoria in cui si osservano sia trattati sia non-trattati, e in cui si possano trarre informazioni generalizzabili ai due gruppi sulla dipendenza tra la variabile-risultato e il punteggio.

Se invece si utilizza un gruppo di controllo formato da *non richiedenti* che avrebbero avuto i requisiti di idoneità, ci si trova a dover fronteggiare un più ampio spettro di fonti di distorsione. Esistono da un lato differenze di motivazione, che hanno portato gli uni a richiedere una borsa di studio e gli altri a non farlo. Dall'altro, esistono anche differenze di merito e di condizione economica ma, a differenza del caso descritto, in questo non esiste un criterio "deterministico" (come in precedenza la graduatoria) che permetta di distinguere con certezza borsisti e non borsisti sulla base di caratteristiche ben definite. In questo caso, perciò, è noto che tra i due gruppi posti a confronto esistono differenze connesse sia a fattori noti e osservabili sia ad altri fattori non osservabili.

Anche in questo caso la tecnica utilizzata dai valutatori per stimare l'effetto è un modello di regressione. A differenza del caso precedente, è molto importante inserire nel modello il maggior numero possibile di caratteristiche individuali: non essendo espliciti tutti i fattori che influenzano la selezione, l'utilizzo di un grande numero di variabili esplicative, che possono essere indirettamente correlate con la probabilità di chiedere e ottenere una borsa, ci cautela dal rischio che, al netto delle variabili utilizzate, rimangano tra borsisti e non borsisti sistematiche differenze dovute a fattori non osservabili. Gli autori producono stime dell'effetto affiancando al modello di regressione anche una procedura di *matching*, discussa al capitolo 5. Come per la regressione, i risultati del *matching* sono attendibili se vale l'ipotesi di *selezione sulle osservabili*, cioè se le variabili esplicative prese in considerazione sono in grado di spiegare in modo esaustivo le differenze di partenza tra borsisti e non borsisti.

Gli autori propendono per un confronto che usi come gruppo di confronto i richiedenti idonei che non hanno percepito la borsa perché, per quanto i risultati siano difficilmente generalizzabili agli studenti con caratteristiche troppo differenti da quelle della soglia della graduatoria, l'esatta conoscenza del meccanismo di selezione dà una ragionevole sicurezza sull'assenza di selezione su altre variabili non osservabili. Tuttavia non è sempre possibile ricorrere agli idonei non borsisti, in quanto nel periodo d'interesse in molti atenei le erogazioni hanno riguardato la quasi totalità degli idonei, precludendo la possibilità di disporre di un gruppo di controllo abbastanza numeroso. Più in generale, da un ateneo all'altro esistono tali e tante differenze nelle regole di selezione, modalità di erogazione, importo delle borse e di disponibilità di dati che le analisi sono state svolte separatamente per ogni ateneo, ricorrendo volta per volta alla strategia di analisi più adatta alla situazione.

I dati utilizzati

La banca-dati utilizzata per lo studio è stata generata a partire dagli archivi anagrafici, delle iscrizioni e dei corsi, forniti dagli 11 atenei coinvolti, relativi agli studenti immatricolati negli anni accademici 1998/99, 1999/00 e 2001/02. L'anno accademico 2001/02 è il primo anno in cui sono stati attivati in modo generalizzato i corsi di laurea triennali del nuovo ordinamento: l'analisi è perciò condotta separatamente per anno

accademico, in modo da tener conto delle eventuali differenze tra le situazioni pre-riforma e post-riforma. L'anno accademico 2000/01 è stato invece escluso perché a cavallo del processo di riforma che ha portato all'istituzione delle nuove lauree. I differenti archivi sono stati sottoposti a un processo di omogeneizzazione, controllo e correzione dei dati al termine del quale si è giunti alla creazione di un unico archivio di riferimento. L'archivio contiene informazioni su circa 240.000 studenti, ai quali corrispondono oltre 500.000 iscrizioni, la cui suddivisione per ateneo è mostrata nella Tabella 12.1.

Tabella 12.1: Iscritti e iscrizioni per ateneo

Ateneo	Iscritti	Iscrizioni
Cagliari	16.193	34.625
Calabria	14.502	29.481
Camerino	2.960	5.667
Catania	28.736	59.862
Ferrara	7.329	15.444
Firenze	26.107	55.451
Milano (Politecnico)	19.180	43.730
Padova	29.019	64.094
Roma (La Sapienza)	67.358	135.879
Salerno	19.001	40.250
Trieste	9.829	21.680
TOTALE	240.214	506.163

Fonte: IRPET (2004)

L'archivio contiene, oltre ai dati anagrafici degli studenti relativi a ogni iscrizione, il numero di esami da essi sostenuti nei primi due anni accademici e una variabile dicotomica rappresentante il possesso o meno dei requisiti necessari per l'ottenimento di una borsa. L'evento "abbandono dell'università" è individuato osservando se un individuo, immatricolato nell'anno accademico t , si è successivamente iscritto o meno per il secondo anno del corso nell'anno accademico $t+1$. L'analisi sulla probabilità di abbandono è per questo motivo ristretta ai soli immatricolati entro l'anno accademico 1999/00: per osservare gli abbandoni degli immatricolati nel 2001/02 sarebbe stato infatti necessario disporre dei dati sulle iscrizioni del 2002/03. Lo stesso discorso vale per il numero di esami svolti nel secondo anno. In pratica la coorte degli iscritti

nell'anno 2001/02 è utilizzata solo per stimare l'impatto delle borse sulla probabilità di non sostenere esami al primo anno.

Gli autori segnalano una qualità e disponibilità dei dati necessari complessivamente soddisfacente. Le maggiori difficoltà sono connesse all'acquisizione dei dati sulle condizioni economiche degli studenti, dovute principalmente alle differenti modalità con cui i vari atenei hanno registrato le informazioni. Da un lato esistono problemi di completezza per gli atenei che raccolgono dati sulla condizione economica solo per quegli studenti che fanno richiesta di benefici connessi al diritto allo studio; dall'altro si hanno problemi di omogeneità, che derivano dai differenti strumenti adottati dai vari atenei per la rilevazione della condizione economica: alcuni non raccolgono informazioni sul patrimonio, in quanto la condizione economica è valutata sulla base del solo reddito, mentre altri basano la rilevazione direttamente sul valore dell'ISEE, per cui non si dispone dei dati separatamente per reddito e patrimonio.

La strategia di valutazione

Come già accennato, la forte eterogeneità fra gli atenei ha comportato la scelta di svolgere l'analisi separatamente per ogni singolo ateneo, ricorrendo di volta in volta alle più appropriate strategie di analisi in base alle modalità di selezione dei borsisti, di erogazione delle borse e alla disponibilità dei dati.

Il fattore maggiormente determinante nella scelta della tecnica è la possibilità di costruire un gruppo di confronto costituito da *richiedenti idonei ma non borsisti*. In tal caso le stime degli effetti al primo anno (effetto sulla probabilità di abbandonare l'università e sulla probabilità di non sostenere alcun esame) sono ottenute ricorrendo al *regression discontinuity design* e utilizzando un modello *logit* (un modello di regressione per variabili dipendenti dicotomiche, quali sono le variabili-risultato utilizzate), in cui l'effetto della borsa è stimato al netto di una serie di caratteristiche individuali.

Nei casi in cui invece non sia possibile disporre di un gruppo di confronto formato da richiedenti idonei, perché tutti o quasi tutti gli idonei hanno ricevuto una borsa di studio, si effettua una stima dell'effetto via *matching*, abbinando ogni percettore di una borsa di studio con il più simile fra tutti gli studenti che non ne hanno fatto richiesta ma che

avrebbero soddisfatto i requisiti di idoneità. Le caratteristiche individuali utilizzate per l'abbinamento, ricavabili dagli archivi universitari, sono: genere, voto di maturità, tipo di diploma, reddito familiare equivalente (cioè l'indicatore utilizzato per quantificare la condizione economica della famiglia), facoltà di appartenenza, anno accademico di immatricolazione, distanza in anni tra il conseguimento del diploma e l'iscrizione all'università, il grado di prossimità dello studente all'università (in sede, pendolare, fuori sede).

Le analisi sono condotte separatamente, oltre che per ateneo, anche per una serie di gruppi di individui, in modo da cogliere la variabilità degli effetti di una borsa di studio rispetto alle condizioni in cui essa viene erogata. Le variabili di stratificazione utilizzate sono:

- l'anno accademico di immatricolazione, per distinguere soprattutto tra l'effetto pre-riforma (negli anni 1998/99 e 1999/00) e quello post-riforma (nell'anno 2001/02);
 - la residenza dello studente, per cogliere i differenti effetti connessi a differenti prossimità alla sede universitaria, e per tenere conto della possibilità che i borsisti fuori sede hanno di beneficiare anche di un alloggio;
 - l'importo della borsa, per verificare la relazione tra effetti ed entità degli importi.
- La stratificazione sulla base dell'importo è applicata solo per alcuni atenei. Sono esclusi quelli il cui regolamento prevede l'erogazione di borse dall'importo scarsamente variabile, e quelli in cui l'importo dipende unicamente dalla residenza dello studente, già inserita tra le variabili di stratificazione.

Per quanto riguarda le stime degli effetti su variabili relative al secondo anno di università (effetto sulla quota di esami sostenuti sul totale previsto), l'analisi è condotta in tutti gli atenei, indipendentemente dal tipo di gruppo di confronto, ricorrendo a un modello di regressione, in cui la quota di esami sostenuti è messa in relazione con le stesse caratteristiche individuali utilizzate per le analisi sul primo anno. In questo caso, però, la variabile-trattamento è più articolata, prevedendo non solo la condizione di borsista/non borsista, ma anche il numero di anni di fruizione della borsa. La variabile-trattamento assume perciò un numero maggiore di modalità, corrispondenti a differenti regimi di trattamento. In questo modo è possibile per gli autori stimare separatamente

l'effetto di ricevere la borsa solo al primo anno, oppure solo al secondo, oppure in entrambi.

12.4 I risultati della valutazione

L'eterogeneità dei metodi utilizzati è pari a quella dei risultati ottenuti, che sono perciò esposti nel rapporto separatamente per ogni ateneo. Nel seguito si sintetizzano, a scopo esemplificativo, i risultati relativi a due degli undici atenei considerati, il Politecnico di Milano e l'Università di Catania, scelti perché valutati con tecniche diverse, rimandando al rapporto originale per i risultati completi.

Politecnico di Milano

Il Politecnico di Milano è uno degli atenei nei quali la quasi totalità dei richiedenti idonei beneficia di una borsa di studio, per cui la valutazione degli effetti delle borse di studio è condotta ponendo a confronto i borsisti con i non richiedenti che soddisfano i requisiti di idoneità.

Un primo confronto tra i due gruppi evidenzia esiti marcatamente diversi tra borsisti e non borsisti:

- tra i borsisti il tasso di abbandono oscilla, a seconda dell'anno accademico di riferimento, tra lo 0% e il 7,2%, mentre per i non-borsisti varia tra il 14,9% e il 15,6%;
- per quanto riguarda la produttività al primo anno, una quota di borsisti compresa tra lo 0% e il 6,4% non ha sostenuto esami alla fine del primo anno, mentre tra i non borsisti la quota varia tra l'11,6% e il 18,3%;
- riguardo alla produttività al secondo anno, l'1,9% dei borsisti sostiene meno di un quarto degli esami/crediti previsti, contro il 7,1% dei non borsisti. Tra i borsisti, il 63,9% svolge almeno i tre quarti degli esami, contro il 40,8% dei non borsisti.

Tali risultati, che sembrerebbero suggerire che la borsa riesce a migliorare diffusamente la regolarità degli studi, vanno però intesi come indicazioni puramente descrittive. Essi non tengono in nessuna considerazione le eventuali differenze di partenza tra i due gruppi, per cui i vari esiti potrebbero essere addebitabili a tali diversità e non (o in misura minore) alla borsa. A questo stadio dell'analisi non è neppure possibile escludere che, al contrario, le differenze di partenza operino nel senso di favorire i non borsisti,

nel qual caso l'effetto della borsa potrebbe essere addirittura sottostimato dalle differenze appena presentate.

Per le variabili-risultato relative al primo anno si ricavano stime dell'effetto al netto delle differenze di partenza ricorrendo al *matching*. Essendo l'importo delle borse erogate al Politecnico di Milano piuttosto variabile, le stime sono calcolate separatamente per le borse di importo basso e quelle di importo medio-alto. Nella Tabella 12.2 sono sintetizzati i risultati delle stime.

Tabella 12.2: Effetti della borsa al primo anno di università (Politecnico di Milano)

Effetto sulla probabilità di abbandono		
	Importo della borsa	
	Basso	Medio-alto
a.a. 1998/99 e 1999/00		
In sede	-10,5%	- 5,5%
Pendolari	- 8,5%	- 14,1%
Fuori sede	- 13,6%	- 13,8%
Effetto sulla probabilità di non sostenere esami il primo anno		
	Importo della borsa	
	Basso	Medio-alto
a.a. 1998/99 e 1999/00		
In sede	- 10,5%	- 13,3%
Pendolari	- 10,7%	- 15,1%
Fuori sede	- 9,4%	- 12,7%
a.a. 2001/02		
In sede	- 21,1%	-
Pendolari	- 8,3%	-
Fuori sede	- 14,5%	- 16,5%

Nota: I valori in grassetto indicano stime statisticamente significative al 90% di confidenza
Fonte: IRPET (2004)

Per quanto riguarda gli effetti sulla probabilità di abbandono al primo anno, le stime suggeriscono un effetto delle borse di studio di segno negativo (nel senso di ridurre tale probabilità). Con l'eccezione degli studenti in sede, le stime d'impatto sono significative e indicano un calo dagli 8 ai 14 punti percentuali nella probabilità di abbandonare l'università dopo il primo anno. Per le borse di importo basso si stima un impatto medio leggermente superiore rispetto a quello delle borse di importo medio-

alto, mentre l'osservazione dei risultati stratificati per residenza dei beneficiari produce risultati alterni e difficilmente interpretabili.

L'analisi dell'effetto sulla probabilità di non sostenere esami al primo anno fornisce, relativamente alle coorti degli immatricolati negli anni 1998/99 e 1999/00, evidenze molto simili a quelle ottenute per la probabilità di abbandono⁷⁷. Le stime sono negative e significative, anche questa volta con l'eccezione degli studenti residenti in sede. Le borse di studio sembrano abbassare di circa dieci punti percentuali il rischio di non sostenere nemmeno un esame nel primo anno, con un impatto leggermente superiore per le borse di importo più alto. L'analisi sugli immatricolati nell'anno 2001/02 fornisce invece minori evidenze: anche in questo caso le stime mostrano un effetto generalmente negativo, ma quasi mai significativo, con l'eccezione dell'effetto sugli studenti fuori sede per le borse di importo maggiore.

L'effetto di ricevere una borsa al secondo anno sulla quota di esami superata è stimato ricorrendo a un modello di regressione lineare, con cui si studia la dipendenza tra la variabile-risultato e le medesime variabili esplicative usate per il *matching*. Rispetto alle analisi sul primo anno, cambia non solo la variabile-risultato ma anche la variabile-trattamento. In questo caso si tiene conto sia dei differenti importi della borsa, sia della possibilità di aver fruito della borsa di studio per un solo anno o per due anni di fila. Una sintesi delle stime è mostrata nella Tabella 12.3.

Tabella 12.3: Effetto della borsa al 2° anno sulla quota di esami sostenuti (Politecnico di Milano)

	Borsa medio-alta (senza borsa al 1° anno)	Borsa bassa (senza borsa al 1° anno)	Borsa medio-alta (già con borsa al 1° anno)	Borsa bassa (già con borsa al 1° anno)
In sede	+4,7%	+6,1%	+4,0%	+3,8%
Pendolari	+3,5%	+3,8%	+1,3%	+0,8%
Fuori sede	+4,5%	+5,5%	+10,2%	+9,8%

Nota: I valori in grassetto indicano stime statisticamente significative al 90% di confidenza
Fonte: IRPET (2004)

⁷⁷ Le due dimensioni sono altamente correlate: la maggior parte degli studenti che non sostengono esami al primo anno finisce per non iscriversi al secondo anno.

I risultati mostrano un effetto delle borse al secondo anno positivo e significativo (per quanto contenuto) *per coloro che non ne hanno beneficiato al primo anno*. Ricevere una borsa di importo basso al secondo anno aumenta il numero di esami sostenuti di una quota variabile tra il 3,8% e il 6,1%, a seconda della residenza dello studente. Lo stesso vale per l'erogazione di una borsa di importo medio-alto, per quanto in tal caso l'effetto stimato sia lievemente inferiore (dal 3,5% al 4,7%).

Tra gli studenti *che hanno già percepito una borsa di studio al primo anno*, l'effetto di ricevere una borsa anche al secondo anno è positivo e significativo, ma nel solo caso degli studenti fuori sede, mentre per gli altri non ci sono evidenze significative circa l'efficacia della borsa. È importante rilevare che il modello di regressione utilizzato produce un effetto di ricevere una borsa al primo anno (sulla quota di esami sostenuti al secondo anno) fortemente negativo per gli stessi studenti fuori sede. Tale effetto, non riportato nella Tabella 12.3, è del 7,5%. Ciò significa che, per uno studente fuori sede, l'erogazione di una borsa *al primo ma non al secondo anno* è connessa a una peggiore produttività rispetto a chi non ha mai ricevuto una borsa. Considerando congiuntamente tale effetto negativo con quello positivo, mostrato nella Tabella 12.3, di ricevere una borsa al secondo anno (+ 10,2% per le borse alte e + 9,8% per le borse basse), si può concludere che l'effetto di due borse consecutive sulla produttività al secondo anno è positivo ma molto contenuto, del 2,5% circa.

Università di Catania

L'analisi condotta sull'Università di Catania è caratterizzata, rispetto a quella appena vista per il Politecnico di Milano, da due differenze. In primo luogo, l'erogazione delle borse di studio ha coperto, nei tre anni accademici considerati, molto meno del 50% delle richieste di studenti idonei, per cui la numerosità degli *idonei non borsisti* consente in questo caso di utilizzarli come gruppo di confronto. In secondo luogo, l'importo della borsa di studio dipende solamente dalla residenza dello studente, per cui le analisi non sono condotte separatamente per importo (in quanto già effettuate separatamente per studenti in sede, pendolari e fuori sede).

Un primo confronto, di carattere puramente descrittivo, tra il gruppo dei borsisti e il gruppo di controllo fornisce le seguenti indicazioni:

- il tasso di abbandono al primo anno non sembra molto differente: nel 1999/00 è stato del 17% in entrambi i gruppi, mentre nel 1998/99 è stato superiore per i borsisti (19% contro il 16% degli idonei non borsisti);
- la quota di borsisti che non sostengono esami al primo anno varia tra il 15% e il 16%, a seconda dell'anno accademico di immatricolazione. Nel gruppo di controllo tale quota è quasi identica, con l'unica rilevante eccezione del 2001/02, in cui essa è inferiore di tre punti percentuali (13% contro 16%);
- la quota di studenti che sostengono meno di un quarto degli esami previsti al secondo anno è piuttosto simile nei due gruppi (7% per i borsisti, 9% per i controlli); lo stesso vale per la quota di coloro che ne sostengono almeno i tre quarti (31% per i borsisti e 30% per i controlli).

Per poter ottenere delle stime depurate dalle differenze di partenza tra i due gruppi, gli autori ricorrono alle tecniche descritte in precedenza. Disponendo di un gruppo di confronto differente dai borsisti solo per la posizione in graduatoria, si utilizza il *regression discontinuity design*. L'effetto di una borsa al primo anno sulla probabilità di abbandono e sulla probabilità di non sostenere esami è stimato mediante una regressione logistica, mentre l'effetto sulla quota di esami sostenuti al secondo anno è stimato con un modello di regressione lineare.

Le variabili esplicative sono le stesse descritte in precedenza. Va ricordato che in questo caso gli effetti stimati sono soprattutto validi per gli individui vicini alla soglia che taglia in due la graduatoria e saranno tanto meno affidabili quanto più distanti dalla soglia saranno gli individui a cui li si vuole generalizzare. Nella Tabella 12.4 si presentano le stime degli effetti relativi al primo anno d'università.

Tabella 12.4: Effetti della borsa al primo anno di università (Università di Catania)

Effetto sulla probabilità di abbandono	
a.a. 1998/99 e 1999/00	
In sede	- 4%
Pendolari	- 10%
Fuori sede	- 6%
Effetto sulla probabilità di non sostenere esami il primo anno	
a.a. 1998/99 e 1999/00	
In sede	- 3%
Pendolari	- 5%
Fuori sede	- 6%
a.a. 2001/02	
In sede	+ 4%
Pendolari	+ 13%
Fuori sede	- 2%

Nota: I valori in grassetto indicano stime statisticamente significative al 90% di confidenza
 Fonte: IRPET (2004)

L'analisi produce una stima di un effetto negativo e significativo sulla probabilità di abbandonare dopo il primo anno, ma limitatamente a studenti pendolari e fuori sede. In merito alla probabilità di non sostenere esami al primo anno, gli effetti stimati sono ancora minori: si registra un effetto del 6% per gli studenti fuori sede immatricolati entro l'anno 1999/00, mentre per studenti in sede e pendolari l'effetto non è significativamente diverso da zero. Per quanto riguarda gli immatricolati nel 2001/02, le stime d'impatto sono non significative indipendentemente dalla residenza degli studenti.

Nella Tabella 12.5 si presenta una sintesi delle stime del modello di regressione utilizzato per valutare l'effetto sulla quota di esami sostenuti al secondo anno.

Tabella 12.5: Effetto della borsa al 2° anno sulla quota di esami sostenuti (Università di Catania)

	Borsa al 1° anno ma non al 2° anno	Borsa al 2° anno ma non al 1° anno	Borsa al 1° anno e al 2° anno
In sede	- 9,2%	+ 0,5%	+ 14,5%
Pendolari	- 6,1%	+ 0,7%	+ 1,0%
Fuori sede	- 5,3%	+ 3,4%	+ 7,8%

Nota: I valori in grassetto indicano stime statisticamente significative al 90% di confidenza
 Fonte: IRPET (2004)

L'effetto delle borse di studio sugli esami sostenuti al secondo anno sembra essere molto variabile a seconda dei beneficiari e del numero di volte in cui si è ottenuta la borsa. La percezione di una borsa di studio, sia essa al primo, al secondo, o in entrambi gli anni non sembra influenzare significativamente i risultati degli studenti pendolari. Per quanto riguarda le altre tipologie di studenti, si evidenzia una relazione negativa tra il numero di esami sostenuti e la percezione di una borsa di studio solo al primo anno.

La percezione di una borsa solo al secondo anno ha un'influenza abbastanza modesta: non modifica significativamente i risultati degli studenti in sede, mentre aumenta del 3% la quota di esami svolti dagli studenti fuori sede. Più evidenti sono i risultati nel caso in cui si eroghi una borsa sia al primo che al secondo anno: per gli studenti fuori sede la produttività aumenta quasi dell'8%, mentre per gli studenti in sede si registra un aumento del 14,5%.

Una sintesi dei risultati

L'osservazione congiunta dei risultati ottenuti sugli undici atenei coinvolti nella ricerca offre un quadro di non facile interpretazione. L'effetto stimato è generalmente nullo quando si ricorre a un gruppo di confronto formato da richiedenti idonei. C'è invece qualche evidenza in più nel caso in cui si utilizzi un gruppo di non richiedenti con i requisiti di idoneità, ma gli autori suggeriscono di interpretare questo dato con cautela: le differenze di motivazione tra trattati e controlli potrebbero in quei casi avere condotto a una sovrastima dell'effetto.

La stratificazione delle stime per una serie di variabili d'interesse (residenza degli studenti, importo delle borse, anno accademico di immatricolazione) ha arricchito il quadro informativo della ricerca, ma non è stato possibile individuare delle regolarità nelle stime in relazione ad alcuna di tali variabili.

Fatte queste considerazioni, gli autori formulano dubbi riguardo ad una reale efficacia delle borse nel sostenere la regolarità negli studi degli studenti privi di mezzi. La conclusione si accompagna a due considerazioni sulla politica in esame. La prima riguarda l'importo delle borse: sono di importo mediamente modesto, quindi, a parere degli autori, insufficienti a ridurre in modo sensibile i costi di mantenimento degli studi universitari. La seconda riguarda un aspetto fondamentale della politica del diritto allo

studio: la selezione sociale potrebbe aver luogo non tanto all'entrata o durante l'università, quanto piuttosto ai livelli scolastici inferiori. L'ipotesi avanzata è che le condizioni familiari (quelle socio-culturali, l'istruzione dei genitori in particolare) influenzino principalmente l'esito alle scuole medie inferiori, dal quale scaturisce la scelta del tipo di scuola superiore e successivamente, in congiunzione con l'effetto dei risultati alle superiori, la scelta di frequentare l'università.

La conclusione degli autori è che in un tale quadro, la parificazione delle opportunità di istruzione andrebbe forse perseguita non tanto con una politica come quella delle borse di studio universitarie, ma con un supporto, non solo economico, nel periodo fra le scuole medie e le superiori, in cui le condizioni sociali degli studenti esplicano in misura maggiore la loro influenza.

CONCLUSIONI

QUALCHE SUGGERIMENTO (E CAUTELA) PER CHI INTENDA VALUTARE GLI EFFETTI DI UNA POLITICA

di Alberto Martini e Marco Sisti

La comprensione dei legami causali tra interventi effettuati e cambiamenti osservati è un ingrediente importante di una moderna e consapevole *governance* pubblica. Individuare questi legami causali è tuttavia difficile, quando non impossibile: crediamo che sia importante sottolinearlo onestamente, per non creare aspettative eccessive.

Forte e diffusa è infatti la tentazione di avanzare pretese conoscitive sempre più ambiziose sugli effetti delle politiche pubbliche. Qual è l'effetto delle politiche per l'integrazione degli immigrati sul clima di convivenza sociale? Qual è l'effetto della liberalizzazione del mercato del lavoro sul tasso di disoccupazione? Qual è l'effetto di un certo assetto dell'organizzazione sanitaria sulla salute dei cittadini? Qual è l'effetto di quel determinato investimento sullo sviluppo locale di una certa area? Le domande sugli effetti sono facili da formulare, soprattutto quando non si ha consapevolezza di cosa sia necessario per ottenere risposte *plausibili* agli interrogativi posti. Un messaggio deve esser perciò chiaro: *gli strumenti analitici presentati in questo volume vanno utilizzati con cautela e con la consapevolezza dei loro limiti, oltre che delle loro potenzialità.*

Inoltre, non tutte le politiche pubbliche giustificano una valutazione degli effetti prodotti. Talvolta tali effetti sono scontati o banali, e non vale la pena spendere risorse per quantificare ciò che è ovvio. In altri casi, non è possibile delimitare con precisione quali siano i confini della politica pubblica di cui si indagano gli effetti, cioè non si riesce a trovare una risposta chiara alla domanda "effetto di cosa?". In altri casi ancora, non è possibile determinare esattamente cosa la politica pubblica voglia ottenere, a parte creare un "alone" di interesse per il problema collettivo in questione: cioè non si riesce a trovare una risposta chiara all'interrogativo "effetto su cosa?". *In tutti questi casi, è sconsigliabile affrontare l'impresa di valutare gli effetti ed è più opportuno orientarsi verso altri tipi di attività valutative.*

Nelle prossime pagine presenteremo alcune riflessioni su quella che chiameremo verifica della *valutabilità* delle politiche pubbliche. Tale verifica richiede la risposta ad una serie di interrogativi, discussi nei prossimi paragrafi.

A. Si intendono davvero valutare *gli effetti* della politica pubblica?

La prima operazione da compiere, e forse la più importante, è la verifica della *genuinità* della domanda di valutazione. Il primo passo cioè è chiedere, e chiedersi, se ci si vuole *realmente* impegnare nella valutazione degli effetti. Ogni disegno di valutazione dovrebbe partire da una risposta intellettualmente onesta a questa domanda.

Quando si inizia a pensare alla valutazione delle politiche pubbliche, è facile lasciarsi trasportare dall'entusiasmo verso le sfide cognitive più ambiziose. Crediamo di non sbagliare asserendo che non esiste amministrazione pubblica in Italia che non abbia affermato in qualche documento ufficiale di voler *valutare l'efficacia* o *misurare l'impatto* della sua azione e delle sue politiche. Questo significa forse che in Italia esiste un'ampia domanda di valutazione degli effetti, così come essa è stata descritta nelle pagine di questo volume? Se così fosse saremmo i primi a rallegrarcene, ma crediamo di poter avanzare qualche dubbio in proposito. Dietro tali affermazioni spesso si nasconde, oltre all'esercizio della retorica, una certa confusione concettuale.

C'è un uso disinvolto del termine *efficacia*. A rigore, ottenere gli effetti voluti è sinonimo di essere efficaci. In questa prospettiva, una politica pubblica è efficace quando riesce a causare i cambiamenti desiderati. In molti casi però i termini “efficacia” ed “efficace” sono usati in accezioni che non hanno nulla a che fare con la produzione di effetti, nel senso inteso qui di “conseguenze attribuibili in senso causale ad una particolare azione pubblica”.

Spesso il termine efficacia è usato nel significato di *qualità*, cioè di presenza di modalità e attributi ritenuti desiderabili. Ad esempio, un asilo nido può essere detto “efficace” se offre un servizio di buona qualità, se cioè rispetta certi standard e/o incontra il favore degli utenti. Naturalmente nella valutazione dell'efficacia del servizio di asilo nido non sono in discussione gli “effetti” di tale servizio (effetti che sarebbe difficile persino concettualizzare): si intende semplicemente sapere se tale servizio soddisfa le

aspettative dei genitori e i bisogni dei bambini. Le insegnanti sono attente e gentili? Il cibo è buono e sano? L'orario di apertura e di chiusura è adeguato alle esigenze dei genitori? Le risposte date a queste domande possono descrivere "l'efficacia" del servizio offerto, ma non mettono in discussione l'esistenza stessa del servizio, e quindi la sua capacità di produrre degli effetti positivi, come invece avviene quando ci poniamo nell'ottica controfattuale. Detto in altri termini, rispondendo a tali domande non vogliamo capire se il servizio di asilo nido "serve", ma solo se viene erogato nel modo migliore possibile.

Efficacia è usato in altri casi come sinonimo di *grado di copertura della domanda espressa* per un particolare servizio. Di nuovo, nel caso dell'asilo nido, il servizio può essere definito come pienamente efficace se non vi sono liste di attesa, cioè se tutti coloro che hanno bisogno dell'asilo nido sono stati serviti e nessuno è rimasto escluso. In questa logica, il Comune che ha servito il 75% delle famiglie richiedenti un posto all'asilo è stato più efficace del Comune che ne ha servito solo il 50%. Anche in questo caso, la verifica dell'efficacia non riguarda gli "effetti" dell'asilo nido, ma intende semplicemente verificare la diffusione del suo utilizzo tra le famiglie, secondo il semplice principio che maggiore è l'ampiezza della platea servita e migliore, o più "efficace", è il servizio offerto. In questa accezione dunque l'efficacia massima del servizio corrisponde all'assenza di razionamento nelle prestazioni erogate.

In altri casi ancora, l'efficacia viene definita con l'espressione più generica di *grado di conseguimento degli obiettivi*. In questo caso occorre intendersi sul significato da attribuire al termine "obiettivo". Se l'obiettivo di cui si vuole verificare il conseguimento consiste nell'indurre cambiamenti in una direzione desiderata, allora stiamo parlando di valutazione degli effetti. In questo senso, infatti, una politica che ha raggiunto l'obiettivo è una politica che ha prodotto gli effetti desiderati. Molto spesso però l'obiettivo da conseguire non riguarda la *produzione di cambiamenti*, bensì la *realizzazione di attività*. In questa logica, è efficace l'amministrazione o l'ente che riesce a rispettare gli impegni presi, espressi in termini di "cose da fare". Se, ad esempio, un certo Comune si è prefisso di riparare, entro sei mesi, dieci chilometri di strade cittadine, la verifica di quanto è stato "efficace" dipende dal numero di chilometri che ha effettivamente riparato al termine di quel periodo. Allo stesso modo, se una

Provincia aveva impegnato risorse per erogare cinquanta corsi di formazione professionale in un anno, la verifica della sua efficacia dipende dal numero di corsi che, dopo un anno, ha effettivamente erogato.

Per verificare la *capacità realizzativa* di un'amministrazione non è ovviamente necessario utilizzare il complesso apparato analitico illustrato in questo volume: le realizzazioni possono semplicemente essere "contate" e/o "descritte". In questi casi il problema conoscitivo da risolvere grazie all'attività di valutazione non riguarda l'attribuzione di causalità (si dà per scontato che la strada non sarebbe stata riparata senza l'intervento del Comune o che il corso di formazione non sarebbe stato erogato senza il finanziamento della Provincia), ma la descrizione quantitativa e qualitativa dello sforzo compiuto dalle amministrazioni per soddisfare gli impegni presi.

A queste considerazioni, infine, aggiungiamo un'ulteriore cautela riguardante la distinzione tra valutazione degli effetti e semplice descrizione (che potremmo definire monitoraggio, se non fosse anche questo un termine plurivalente) del cambiamento osservato in un fenomeno considerato problematico. Come si è argomentato nel capitolo 4, sequenzialità non implica in alcun modo causalità. Vi sono però alcune situazioni in cui può essere opportuno, o necessario, verificare l'evoluzione nel tempo di un problema, senza con questo pretendere di attribuire la responsabilità del cambiamento registrato ad una particolare politica pubblica. È diminuito l'uso di sostanze stupefacenti? È aumentata la fiducia nelle istituzioni da parte della collettività? Si è ridotta la disoccupazione? Le risposte a queste domande possono fornire utili indicazioni a chi deve disegnare una politica pubblica, in quanto descrive l'entità, e traccia i confini, dell'evoluzione del problema sul quale si intende intervenire. Ma tale descrizione ha poco a che fare con la valutazione degli effetti.

Talvolta anche la misurazione del cambiamento di un fenomeno, e il confronto del cambiamento riscontrato con il cambiamento previsto o auspicato, viene definita valutazione di efficacia. In questi casi la valutazione si basa sul fatto che, a priori, si sia imposto un *target* numerico che le amministrazioni devono cercare di raggiungere. Ad esempio, l'obiettivo delle politiche per l'innovazione e la ricerca da realizzare in Italia nei prossimi anni, potrebbe essere di provocare un incremento nel numero di nuovi

brevetti industriali che ogni anno sono registrati presso l'*European Patent Office*⁷⁸. Poniamo che a tale obiettivo sia data una dimensione numerica, un *target*: raggiungere, entro il 2010, la quota annuale di 150 nuovi brevetti registrati per milione di abitanti. Se attualmente sono registrati in media ogni anno circa 100 brevetti, è chiaro che ci si prefigge di aumentare tale indicatore del 50% nei prossimi quattro anni. Questo rappresenta dunque il *target* di cambiamento che le amministrazioni pubbliche devono prendere a riferimento come stimolo al miglioramento della loro azione. Anche in questo caso però, come nei precedenti, sarebbe inappropriato chiamare valutazione di efficacia il confronto tra il *target* di cambiamento (+50%) e il cambiamento che si osserva alla fine del periodo. Al più, si potrebbe parlare di “verifica dei progressi compiuti”. Ma nulla ci garantisce che i progressi compiuti (l'avvicinamento al *target* prefissato) siano stati *causati* dagli interventi pubblici adottati a sostegno della ricerca dalle varie amministrazioni nazionali e locali, e neppure che la mancanza di progressi (la distanza dal *target* prefissato) sia segno inequivocabile di un'assenza di effetti da parte di tali interventi.

Abbiamo dunque visto come sia possibile giocare con le parole della valutazione, utilizzando lo stesso termine per indicare cose diverse. *Il suggerimento è di non farsi confondere dall'ambiguità né affascinare dal richiamo evocativo dei molti termini impiegati (spesso con leggerezza) nel variegato mondo della valutazione.* Non sempre chi chiede (o chi si propone) di valutare l'efficacia di un'azione pubblica vuol davvero verificare (e quindi mettere in discussione) la capacità di quell'azione di produrre gli effetti desiderati. Molto spesso (oserei dire nella maggioranza dei casi) tale richiesta è motivata da altre esigenze conoscitive, che possono riguardare di volta in volta: la qualità della prestazione offerta; il grado di copertura di una platea di soggetti richiedenti un servizio; la realizzazione degli impegni assunti; il raggiungimento di *target* numerici relativi a fenomeni complessi; e probabilmente molto altro ancora.

Un buon committente di valutazione dovrebbe sempre riuscire a formulare il proprio mandato valutativo eliminando tutti i termini potenzialmente equivoci ed esprimendo cosa vuol sapere attraverso domande specifiche e facilmente comprensibili. Allo stesso

⁷⁸ Si tratta dell'organismo ufficiale che concede brevetti per invenzioni industriali, che hanno validità su 24 Paesi europei.

modo un buon valutatore dovrebbe rispondere a tale mandato senza ricorrere a facili scappatoie terminologiche, ma (eventualmente) cercando di condurre il committente a definire le domande nella forma più adeguata.

B. Su quale *variabile-risultato* si intende stimare l'effetto?

Una volta stabilito che l'intenzione è realmente quella di stimare gli effetti di una politica pubblica, occorre porre mano alla definizione delle dimensioni misurabili, o variabili-risultato, *su cui* vanno valutati tali effetti. L'obiettivo di questa operazione è giungere all'individuazione di una o più grandezze osservabili, che siano ragionevolmente collegate all'intervento adottato dalla pubblica amministrazione, resistendo alla tentazione di evocare entità troppo fumose o astratte.

Si consiglia a questo fine di redigere un elenco delle possibili variabili di interesse, che esprimano le dimensioni del problema sulle quali la politica vorrebbe incidere, mettendo a fianco di ogni variabile la direzione del cambiamento desiderato. Per ogni variabile individuata è necessario poi esplicitare i *meccanismi causali* che (idealmente) possono legare il cambiamento auspicato in tale variabile alle attività svolte nell'ambito dell'intervento in questione. Occorre cioè domandarsi: “È ragionevole attendersi un aumento (o una diminuzione) di questa variabile in seguito alla realizzazione di questa attività? In che modo dovrebbe avvenire tale aumento? Secondo quale specifico meccanismo?”. Se il legame individuato è debole e poco convincente, o la catena dei nessi causali ipotizzati troppo lunga, è consigliabile scartare la relativa variabile-risultato. Nel caso limite in cui sia impossibile individuarne altre, si dovrebbe prendere in considerazione l'opzione di rinunciare a valutare gli effetti di quella politica.

Quando la politica pubblica coinvolge una piccola parte della popolazione, può essere materialmente impossibile verificarne gli effetti ad un livello molto aggregato, ed occorre scegliere una variabile definita ad un livello più coerente con la portata dell'intervento. Un esempio può essere di aiuto. Se un Comune concede incentivi finanziari ai tassisti per indurli a sostituire le vecchie auto a benzina con vetture più ecologiche (a trazione elettrica, a metano o *bi-fuel*), è quantomeno azzardato tentare di valutare gli effetti di tale politica sul livello di inquinamento atmosferico presente in

città. In primo luogo, perché i tassisti costituiscono una percentuale molto piccola del totale degli automobilisti cittadini e, di conseguenza, incidono in misura probabilmente impercettibile sull'inquinamento prodotto dai gas di scarico delle automobili. In secondo luogo, i livelli di inquinamento atmosferico sono soggetti all'influenza di numerosi fattori non connessi al traffico cittadino. Il nesso logico che lega i contributi concessi ai tassisti alle variazioni nei livelli di inquinamento appare perciò *troppo esile* per diventare oggetto di una seria valutazione degli effetti.

Una variabile-risultato più plausibile, sulla quale verificare l'effetto di questa politica, potrebbe essere invece la percentuale di vetture ecologiche circolanti sul totale del parco taxi cittadino; in questo modo attraverso la valutazione si cercherebbe di comprendere se, e in che misura, la politica ha influenzato la propensione dei tassisti ad acquistare ed utilizzare una macchina a basso tasso d'inquinamento. Si tratterebbe di una variabile-risultato di rilevanza magari più modesta, e quindi forse meno accattivante, ma senza dubbio più strettamente connessa alla politica di incentivazione finanziaria di cui sono destinatari i tassisti.

Il suggerimento consiste dunque nell'essere molto realistici nell'individuazione delle variabili-risultato e di trattenersi dal cercare la prova scientifica di nessi causali poco credibili o indimostrabili. Le risorse sempre scarse per la valutazione vanno concentrate sulle dimensioni realisticamente "aggredibili" dall'intervento, rispetto alle quali la presenza o l'assenza di un effetto possa essere verificata con gli strumenti analitici a disposizione.

C. Qual è, precisamente, l'intervento di cui si vuole misurare l'effetto?

La regola generale è presto enunciata: un intervento è tanto più valutabile quanto più è possibile circoscriverne i confini spazio-temporali. *Dal punto di vista temporale*, è molto utile poter stabilire con una relativa precisione quando un intervento comincia e quando finisce, sia in generale, sia rispetto al singolo soggetto esposto all'intervento. Interventi che hanno confini temporali molto sfumati, perché hanno lunghi periodi di *phase-in* e di *phase-out*, o di cui è difficile ricostruire i periodi di vigenza o di fruizione effettiva, possono essere comunque molto utili ed efficaci, e la loro lenta o incerta

adozione può rispondere a precise esigenze organizzative: ciò non toglie che *valutarli* con i metodi quantitativi proposti in questo volume può essere molto problematico. Difficile diventa innanzitutto definire una fase pre-intervento e una fase post-intervento, il che riduce la possibilità di effettuare i confronti descritti al capitolo 4 (confronti che, anche laddove fattibili, sono comunque soggetti alle cautele interpretative ivi discusse).

Analoghe considerazioni vanno fatte *dal punto di vista dei confronti tra soggetti*. Una pre-condizione per effettuare confronti tra trattati e non-trattati è (ovviamente) che sia direttamente osservabile chi è stato trattato e chi non è stato trattato (nel caso più semplice di variabile-trattamento dicotomica), oppure che sia possibile misurare a quale intensità di trattamento ciascuno dei soggetti è stato esposto. Politiche pubbliche che non offrono un trattamento ben definito a individui ben definiti possono non di meno essere molto utili ed efficaci: solamente, la loro indeterminatezza in termini di esposizione effettiva renderà molto più difficile valutarne l'efficacia in modo plausibile.

Per affrontare talune situazioni di incertezza sul trattamento ricevuto sono stati ideati metodi statistici *ad hoc*: si è sviluppata una letteratura specializzata nell'analizzare gli effetti di interventi in presenza di quella che in inglese è definita come *partial compliance*, cioè aderenza incompleta al trattamento assegnato. Questi metodi sono tuttavia utilizzabili soprattutto negli esperimenti sociali con randomizzazione, ed in particolare in quelle situazioni in cui alcuni di coloro che sono assegnati al trattamento vi si sottraggono (i cosiddetti *no-show*), mentre alcuni di coloro che ne sono esclusi riescono ad usufruirne comunque (i cosiddetti *cross-over*).

Una situazione diversa, ma molto rilevante per il contesto italiano, è quella in cui la politica consiste nell'erogazione di un "pacchetto" di trattamenti di natura diversa, con un *mix* che varia da beneficiario a beneficiario. Si pensi a tutti quei casi in cui vengono erogati servizi "integrati" a soggetti portatori di una certa problematica. Ad esempio, un intervento di sostegno abitativo integrato con uno di assistenza economica e con il supporto dei servizi socio-sanitari. Oppure un'agevolazione concessa ad imprese di recente costituzione assieme all'offerta di servizi formativi e consulenziali.

In tutte le situazioni di servizi integrati, o di politiche multi-intervento, la valutazione quantitativa degli effetti diventa più difficile da condurre e i suoi risultati

potenzialmente meno interpretabili ed utilizzabili a scopo decisionale. I problemi di misurazione infatti aumentano: si tratta di rilevare più informazioni relative ai medesimi soggetti, in una situazione in cui diventa maggiore l'incidenza di casi per i quali si hanno dati incompleti. Ma sono soprattutto l'interpretazione e l'utilizzo dei risultati a divenire più problematici. Poniamo che la valutazione stabilisca che una certa politica, composta di più interventi tra loro integrati, abbia avuto nel suo complesso effetti positivi su una qualche variabile-risultato: non sapremo però a quale specifico intervento/i, di cui si compone il trattamento, tale effetto vada imputato. Tuttavia, è proprio questa l'informazione che tipicamente più serve a chi deve decidere sul *mix* di servizi da offrire: si vorrebbe sapere quale intervento è più efficace e per quali soggetti. La valutazione della politica nel suo complesso può dire solamente in che misura i soggetti a cui è stato offerto il "pacchetto" ne hanno beneficiato. In situazioni simili, anche la più corretta valutazione degli effetti può dare risultati poco utili dal punto di vista conoscitivo.

Queste situazioni vanno riconosciute con onestà: la complessità nell'offerta dei servizi pubblici, motivata da una varietà di ragioni (che possono andare dall'esigenza di maggior coordinamento espressa dagli operatori del settore, alla necessità di dare un ruolo attivo a servizi facenti capo a livelli di governo diversi), può essere di ostacolo alla valutazione degli effetti. Di fronte a questa tensione, riteniamo che sia la valutazione, almeno nelle forme delineate in questo volume, a dover farsi da parte, a favore eventualmente di altre forme di valutazione più qualitative e non basate sull'approccio controfattuale. Resta tuttavia un gran numero di situazioni in cui il trattamento erogato da una politica pubblica è sufficientemente semplice e omogeneo da rendere possibile la valutazione dei suoi effetti con i metodi quantitativi illustrati in questo volume.

Il suggerimento consiste nell'individuare e definire con chiarezza il trattamento di cui si vogliono conoscere gli effetti, lasciando da parte le situazioni di maggior complessità. Va infatti riconosciuta apertamente la difficoltà di valutare le politiche che prevedono più componenti di trattamento: se ciascuna componente non può essere chiaramente identificata, isolata e trattata come fosse un singolo intervento di cui si

tenti di stimare l'effetto specifico, l'effetto complessivo della politica rischia di essere di scarso interesse per i decisori.

D. Quale metodo di valutazione conviene utilizzare?

Così formulata, la domanda presuppone che sia sempre possibile scegliere tra una vasta pluralità di metodi, ma questa possibilità esiste più in teoria che in pratica. La scelta è tanto maggiore *quanto prima* le esigenze informative della valutazione vengono prese in considerazione nel processo di disegno e di attuazione della politica.

Innanzitutto, il metodo sperimentale è concepibile solo quando venga preso in considerazione in una fase preliminare al disegno di una politica e vi siano simultaneamente (i) la volontà esplicita, da parte dei proponenti della politica, di “sperimentarla”, cioè di testarne la fattibilità e l'efficacia mediante una realizzazione su scala ridotta, e (ii) la disponibilità ad attendere il tempo necessario alla realizzazione dell'esperimento prima di avere i risultati della valutazione e quindi essere in grado di estenderla ad una scala più ampia. Lungimiranza, pazienza e volontà di sperimentare da parte dei *policy maker* sono quindi i veri presupposti per l'applicabilità del metodo sperimentale: le difficoltà tecniche e le obiezioni di tipo etico sono in varia maniera superabili e in molte situazioni non sorgono nemmeno.

Il suggerimento quindi è di prendere in considerazione il metodo sperimentale per la valutazione degli effetti solo quando il soggetto titolare della politica pubblica non sia oggetto di forti pressioni a “realizzare tutto e subito”, quando il problema da affrontare non sia considerato un'emergenza, ma soprattutto quando esista una reale incertezza sulla capacità da parte di quella politica di produrre gli effetti desiderati (o esista un'opinione diffusa sui rischi che essa produca effetti indesiderati). Solo una reale incertezza giustifica i tempi di attesa imposti tipicamente dalla conduzione di esperimenti sociali.

Se queste condizioni (e altre meno importanti) non si verificano, il metodo sperimentale non è utilizzabile e occorre scegliere invece dall'ampio *menu* dei metodi non sperimentali. Anche in questo caso tuttavia la lungimiranza aiuta, nel senso che quanto

più presto si decide di condurre una valutazione nell'arco di vita di una politica pubblica, quanto più aperta resta la scelta del metodo.

Una questione importante riguarda la disponibilità di *dati pre-intervento* sulla variabile-risultato stessa e/o sulle caratteristiche dei soggetti trattati e non-trattati. I dati pre-intervento sono indispensabili per alcuni dei metodi illustrati nella prima parte di questo volume: senza dati pre-intervento il *menu* si restringe molto. Ad esempio, il metodo differenza-nelle-differenze discusso al capitolo 4 (e la sua estensione realizzata attraverso l'uso del modello di regressione, discussa nel capitolo 5) richiede che sia osservata la variabile-risultato in un periodo *precedente* all'esposizione all'intervento. Ma anche l'applicabilità del *matching* statistico (discusso al capitolo 5) dipende dalla disponibilità di informazioni sulle caratteristiche pre-intervento dei soggetti trattati e non-trattati.

Talvolta i dati pre-intervento possono essere reperiti in archivi statistici esistenti: in questo caso la tempestività non è cruciale, l'estrazione dei dati può essere effettuata ad intervento già avviato o addirittura concluso. Tuttavia, tali archivi spesso sono poveri di informazioni per la valutazione, perché non sono stati concepiti a questo scopo. Molto spesso si è infatti costretti a rinunciare all'utilizzo di archivi pre-esistenti e occorre raccogliere i dati necessari mediante discese sul campo: in questi casi, quanto prima si decide di raccogliere i dati pre-intervento, tanto maggiore sarà la loro qualità e attendibilità, e in molte situazioni dalla tempestività dipende la possibilità stessa di ottenere le informazioni necessarie.

Decidere con ritardo sulla raccolta dei dati pre-intervento può anche fare perdere l'opportunità di ottenerli in maniera facile e poco costosa. Basti pensare a quelle situazioni in cui queste informazioni possono essere ottenute nel momento in cui il soggetto *fa domanda* per ricevere un servizio o un'agevolazione: se la valutazione è già stata programmata, la richiesta delle informazioni pre-intervento può essere inclusa nel modulo di domanda stesso.

Il suggerimento è quindi molto semplice: includere il disegno di valutazione degli effetti quanto prima possibile nel ciclo di vita di una politica pubblica, in modo da non

perdere informazioni cruciali per l'utilizzabilità di tutti quei metodi che richiedono informazioni pre-intervento.

L'altro elemento fondamentale per l'applicabilità dei metodi illustrati in questo volume è la disponibilità di dati sui soggetti non-trattati, cioè su quegli individui che potrebbero essere beneficiari della politica ma, grazie a qualche processo di selezione, non lo sono. L'informazione sui non-trattati è decisiva per la grande maggioranza dei metodi discussi in questo volume, e ne rappresenta in effetti il vero tallone di Achille. Questo perché, al di fuori della valutazione degli effetti, disporre di informazione sui non beneficiari non è rilevante per l'erogazione della politica: nei casi migliori, infatti, i dati vengono raccolti soltanto sui beneficiari dell'intervento mediante l'adozione di sistemi di monitoraggio amministrativo. Non esiste ragione (a meno che non si intenda procedere con una valutazione di tipo controfattuale) per raccogliere informazioni su coloro che non sono stati raggiunti dalla politica.

Il caso della valutazione degli effetti del lavoro interinale illustrata nel capitolo 10 è emblematico da questo punto di vista: le agenzie di lavoro interinale raccolgono informazioni sui soggetti che usufruiscono di questa nuova forma contrattuale, ma ignorano tutti gli altri. Per avere un campione di non utilizzatori del lavoro interinale i valutatori sono dovuti ricorrere ad una indagine sulle famiglie, tentando di individuare i soggetti che potenzialmente avrebbero potuto usufruire del lavoro interinale ma non l'hanno fatto. Parimenti, per la valutazione della legge toscana sull'imprenditoria giovanile illustrata nel capitolo 9 ad un certo punto è stato necessario condurre un'indagine campionaria sulle aziende di recente costituzione, beneficiarie e non beneficiarie delle agevolazioni.

Il suggerimento è il seguente: occorre essere consapevoli del fatto che valutare gli effetti di una politica, con rare eccezioni, impone come condizione necessaria (anche se non sufficiente) la raccolta di informazioni su soggetti che non hanno beneficiato della politica. Tale raccolta può essere costosa e laboriosa, ma senza di essa la gran parte dei metodi esistenti non può essere applicata.

Alcune situazioni sono più facili di altre. Le informazioni sulle imprese non beneficiarie di un'agevolazione possono essere individuate interrogando archivi esistenti, come nel

caso della valutazione degli effetti della legge 488 presentata al capitolo 7. In quel caso era anche disponibile un elenco di imprese non beneficiarie, in quanto queste avevano comunque fatto domanda per le agevolazioni. In altri casi tale elenco non esiste, come nel caso della valutazione del lavoro interinale.

Una situazione particolarmente fortunata è quella in cui esista una graduatoria di soggetti che hanno fatto domanda per un servizio o un'agevolazione e in base alla graduatoria alcuni abbiano accesso alla prestazione e altri non vi abbiano accesso: ciò consente da un lato di semplificare la raccolta dei dati sui non beneficiari, dall'altro di utilizzare i metodi discussi nel capitolo 6, metodi che sfruttano la discontinuità nel trattamento attorno ad una precisa soglia di ammissione.

Il suggerimento consiste nel prestare attenzione a tutte le situazioni in cui l'ammissione ad una prestazione sia basata su una regola precisa: queste situazioni sono molto favorevoli alla valutazione degli effetti, basata sul confronto tra i soggetti situati attorno alla soglia di ammissione. Queste situazioni sono nella pubblica amministrazione abbastanza comuni, e vanno sfruttate con attenzione, perché possono rappresentare veri e propri "esperimenti naturali".

A questa serie di suggerimenti occorre aggiungere un'avvertenza di natura generale: se è vero che l'adozione di particolari metodi di valutazione può essere preclusa dalla mancanza di informazioni, ciò non giustifica affermazioni sommarie come: "è impossibile valutare gli effetti di questa politica perché non ci sono i dati". Se valutare gli effetti di una politica rappresenta davvero un interesse per i decisori pubblici e per la collettività, è quasi sempre possibile escogitare un modo per superare l'indisponibilità momentanea delle informazioni occorrenti: ad esempio, migliorando il sistema di monitoraggio esistente, oppure preparando specifiche indagini sulla platea dei beneficiari e dei non beneficiari. Si tratterà di avere a disposizione tempo e risorse da investire, affinché le nuove modalità di rilevazione delle informazioni entrino in funzione e diano i loro frutti.

In linea di principio, e nonostante tutte le cautele che abbiamo elencato, è perciò opportuno non rinunciare in maniera troppo frettolosa all'idea di valutare gli effetti di una politica pubblica: anche la "non valutabilità" di una politica, a causa di carenze

informative, deve essere sempre documentata ed argomentata, spiegando quali sono le principali difficoltà nell'uso delle informazioni esistenti e perché è impossibile (o troppo costoso) procedere con l'adozione dei metodi proposti in questo volume.

E. Chi è interessato ai risultati della valutazione?

Un passaggio essenziale nel disegnare una valutazione degli effetti consiste nel domandarsi quali sono gli attori che potrebbero utilizzarne i risultati, quali sono le loro motivazioni e che tipo di utilizzo essi potrebbero farne. Può apparire capzioso porsi una domanda di questo genere: la risposta più facile è che la valutazione abbia nel suo diretto committente, colui che paga per la realizzazione dello studio, il principale utilizzatore e che tale committente sia motivato dalla volontà di apprendere qualche lezione sulla capacità della politica analizzata di produrre gli effetti desiderati. In caso di mancanza di effetti, ci si aspetta che il committente intervenga per modificare la politica in questione. In realtà, le cose sono molto più complicate e assai raramente corrispondono alla situazione appena descritta.

Molte esperienze di valutazione non nascono in seguito ad un preciso e consapevole mandato da parte di un committente pubblico. Talvolta sembra anzi che la committenza sia quasi del tutto assente (o comunque molto distratta) e che gli studi più rigorosi da un punto di vista metodologico nascano sostanzialmente in ambito accademico. È solitamente un gruppo di ricercatori che, ispirato da curiosità scientifiche e/o da motivazioni professionali, decide in modo autonomo di realizzare una valutazione degli effetti. Crediamo di non sbagliare asserendo che, anche in alcuni dei casi riportati in questo volume, l'iniziativa di valutare gli effetti della politica non è realmente partita da un qualche decisore pubblico, ma è stata perlopiù promossa dagli stessi valutatori. Ciò non significa che non esista formalmente una committenza pubblica: essa però, come insegna gran parte dell'esperienza legata ai fondi strutturali europei, appare più interessata al lato burocratico e contabile della vicenda, cioè all'impegno e all'erogazione delle risorse dedicate all'attività di valutazione, che non agli aspetti di sostanza. In questo contesto, la presenza di un soggetto che sia portatore di un qualche interesse nei confronti degli esiti della valutazione, e che al contempo abbia un qualche ruolo attivo in seno al processo decisionale, è più l'eccezione che la regola.

Porsi il problema di riconoscere quali sono i possibili fruitori della valutazione e, soprattutto, curare il successivo utilizzo dei risultati, dovrebbe perciò rappresentare la principale preoccupazione di chi si sta accingendo a realizzare un'analisi degli effetti. In assenza di interlocutori attenti, l'intero processo di valutazione rischia di trasformarsi in un esercizio statistico fine a sé stesso.

Mentre per altri tipi di valutazione gli “utilizzatori” possono essere quasi sempre individuati nel *management* amministrativo di un ente pubblico (pensiamo, per fare qualche esempio, alla valutazione intesa come selezione di progetti o al controllo della performance di organizzazioni), è molto più difficile immaginare che gli esiti di una valutazione degli effetti possano avere soltanto una ricaduta interna agli uffici dell'amministrazione. Poiché questo tipo di valutazione intende verificare la capacità della politica di affrontare con successo un problema collettivo, i risultati dovrebbero avere una circolazione più ampia ed eventualmente alimentare un dibattito pubblico al quale partecipino tutti gli *stakeholder* della politica in questione. Capire se una politica funziona non è (o non dovrebbe essere) interesse esclusivo di qualche funzionario o manager pubblico, quanto piuttosto dell'intera comunità di soggetti impegnati a informare, discutere e riflettere sul problema collettivo che ha motivato la nascita dell'intervento pubblico.

Nelle esperienze più mature di valutazione degli effetti il committente non è un unico soggetto e, soprattutto, non è rappresentato unicamente dall'amministrazione responsabile di dare attuazione alla politica valutata (come invece accade quasi sempre in Italia). Molto spesso, come nel caso presentato nel capitolo 3 relativo alla *National Supported Work Demonstration*, la committenza è composta da rappresentanti di organizzazioni pubbliche (nell'esperienza di NSWDC i Ministeri) e di enti privati (essenzialmente Fondazioni, ma non solo), che uniscono gli sforzi e le risorse per tentare di apprendere qualcosa di nuovo sull'efficacia di una particolare ricetta di *policy*. La lontananza di alcune delle istituzioni committenti dalla concreta implementazione degli interventi permette loro di avere un atteggiamento più distaccato circa i possibili esiti delle valutazioni. Non temono troppo eventuali giudizi negativi in quanto la loro missione organizzativa non è legata al successo di un particolare approccio di *policy*, ma alla ricerca dell'approccio “migliore”. Non devono dimostrare la bontà di una

particolare azione pubblica, o difendere l'operato degli uffici; vogliono semplicemente conoscere quali sono gli interventi che hanno più successo per poterne sostenere l'attuazione con le risorse (finanziarie e non) a loro disposizione.

La lezione in questo caso è di non vedere nella valutazione degli effetti un mero procedimento capace di coinvolgere soltanto qualche funzionario pubblico in veste di committente, e di qualche ricercatore universitario in veste di valutatore. Se si intende evitare che gli esiti della valutazione cadano nel vuoto, oppure finiscano, non letti, nel cassetto di una scrivania di qualche funzionario, occorre lavorare per includere in seno al processo di valutazione soggetti portatori di interessi diversi, non direttamente coinvolti nella messa in opera della politica e quindi disposti a coltivare qualche dubbio sul suo successo e desiderosi di imparare dai risultati delle analisi. Da questo punto di vista la “valutabilità di una politica”, nel senso della probabilità che la sua valutazione possa produrre una qualche ricaduta decisionale, dipende anche dal clima di interesse e di attenzione che si riesce a costruire attorno all'evento valutativo.

Nell'esperienza italiana non solo è rara la partecipazione al processo valutativo di soggetti esterni (e quando si realizza, avviene solo a livello formale), ma gli stessi risultati delle valutazioni faticano ad uscire dagli uffici delle amministrazioni committenti. Anche quando i rapporti di valutazione sono fatti circolare, la loro divulgazione, a parte qualche rara eccezione, non viene curata con attenzione e i testi sono formulati per essere letti soltanto dagli addetti ai lavori.

L'ultimo suggerimento è dunque di non trascurare questo importante momento del processo di valutazione, credendolo secondario o accessorio, ma di definire fin dal primo momento quali saranno le modalità di diffusione dei risultati delle valutazioni e costringere gli stessi valutatori ad elaborare strategie comunicative adeguate.

In sintesi

Abbiamo sintetizzato dieci *regole di buon senso* che andrebbero prese in seria considerazione da chi si sta apprestando a valutare gli effetti di una politica pubblica.

1. Per giustificare le risorse dedicate alla valutazione, deve esistere un *ragionevole dubbio* sulla capacità di quella politica di produrre gli effetti desiderati. Politiche

simboliche da un lato, e politiche per cui esiste una condivisa “presunzione di efficacia” dall’altro, non sono buone candidate per l’uso di queste risorse.

2. Valutare l’efficacia non è sinonimo di valutare gli effetti. La prima accezione è molto più vasta della seconda. Solo la seconda ha l’obiettivo cognitivo di stabilire *relazioni di causa-effetto* tra interventi e cambiamenti.

3. Condurre una valutazione degli effetti può produrre il beneficio di costringere i *policy maker* a riconoscere quali obiettivi di cambiamento, al di là della retorica, la politica pubblica davvero persegue: questa *opera di chiarificazione* è necessaria per scegliere le variabili-risultato su cui condurre la valutazione degli effetti.

4. *La complessità* è un ostacolo alla valutazione degli effetti: questo non vuol dire che le politiche pubbliche devono evitare la complessità, suggerisce solo di evitare lo sforzo di valutare gli effetti di politiche “macedonia”, nelle quali sia difficile riconoscere l’intensità e il tipo di trattamento offerto ad ogni soggetto, o addirittura sia impossibile distinguere chi è stato trattato da chi non lo è stato. In sostanza, il trattamento deve essere individuato con precisione, pena la difficoltà di utilizzare i risultati a scopo decisionale.

5. Il *metodo ideale* di valutazione non esiste, e la scelta è spesso dettata dalle circostanze. Quanto prima si decide di condurre la valutazione, tanto maggiore sarà la possibilità di scegliere tra i metodi teoricamente possibili.

6. La *manca di dati* non è una ragione sufficiente per abbandonare l’idea di valutare gli effetti di una politica pubblica. Nel caso si decida di non impegnarsi nell’attività di valutazione a causa di insuperabili carenze informative, sarebbe opportuno spiegare quali sono le difficoltà insite nell’uso delle informazioni esistenti.

7. La valutazione degli effetti non può essere intesa come un procedimento amministrativo, utile soltanto a chi ha responsabilità di natura esecutiva o gestionale. I risultati della valutazione sono *patrimonio dell’intera comunità* interessata al problema collettivo che la politica oggetto di valutazione tenta di risolvere.

8. La fase della *divulgazione dei risultati* è tanto importante quanto quella di realizzazione dell'analisi. Se la valutazione è mal comunicata, e quindi poco capita o addirittura ignorata dai possibili utilizzatori, è come se non fosse mai stata realizzata.
9. La valutazione è destinata ad entrare in *processi decisionali* erratici, confusi e spesso poco reattivi anche alle sollecitazioni razionali più robuste. D'altra parte le decisioni non sono mai il frutto di un meccanismo perfetto e i risultati della valutazione devono sempre fare i conti con i valori e gli interessi rappresentati dagli attori politici e sociali.
10. La valutazione produce solo *argomentazioni*. Queste devono essere basate su evidenza empirica e su un metodo di analisi rigoroso, ma restano pur sempre argomentazioni, non verità assolute.

BIBLIOGRAFIA

Agenzia Veneto Lavoro (2004), "Interazione fra sussidi passivi e incentivi al reimpiego: provenienze ed esiti di lavoratori iscritti nelle liste di mobilità", Rapporto di ricerca per il Ministero del Lavoro e delle Politiche Sociali.

Battistin E., Gavosto A., Rettore E. (2001), "Why do subsidised firms survive longer? An evaluation of a program promoting youth entrepreneurship in Italy", in Lechner M., Pfeiffer (a cura di), *Econometric Evaluation of Active Labour Market Policies*, Physica/Springer-Verlag, Heidelberg.

Battistin E., Rettore E. (2006), "Another Look at the Regression Discontinuity Design", di prossima pubblicazione, *Journal of Econometrics*.

Becker S., Ichino A. (2002), "Estimation of average treatment effects based on propensity scores", *The Stata Journal*.

Bezzi C. (2003), *Il disegno della ricerca valutativa*, FrancoAngeli, Milano.

Biggeri L., Catalano G. (2006), *L'efficacia delle politiche di sostegno agli studi universitari*, Quaderni per il Comitato Nazionale della Valutazione del Sistema Universitario, in corso di pubblicazione, il Mulino, Bologna.

Blundell R., Costa Dias M. (2002), "Alternative approaches to evaluation in empirical microeconomics", *Portuguese Economic Journal*, Vol. 1, No. 2, 91-115.

Bondonio D., Martini A. (2001), "Using Event History Analysis to Evaluate the Impact of Investment Subsidies Targeted to Youth-Owned Firms", in Società Italiana di Statistica, Atti del Convegno "Processi e metodi statistici di valutazione", Roma 4-6 giugno 2001.

Brunello G., Miniaci R. (1997), "Fattore Milano? Un'esplorazione delle liste di mobilità in Lombardia", in Borzaga C., Brunello G. (a cura di), *L'impatto delle politiche attive del lavoro in Italia*, Edizioni Scientifiche Italiane, Firenze.

Campbell D., Stanley J. (1966), *Experimental and Quasi-experimental Designs for Research*, Rand McNally, Chicago.

Capano G., Giuliani M. (a cura di), (1996), *Dizionario di politiche pubbliche*, La Nuova Italia Scientifica, Roma.

Card D. (1990), "The Impact of the Mariel Boatlift on the Miami Labor Market", *Industrial and Labor Relations Review*, 43(2), 245-257.

Card D., Krueger A. (1995), *Myth and Measurement: The New Economics of the Minimum Wage*, Princeton University Press, Princeton NJ.

Caruso E. (2001), “Durata della disoccupazione e probabilità di reimpiego in presenza di incentivi all’assunzione: l’evidenza empirica della lista di mobilità della Regione Umbria”, *Politica Economica*.

Conforti D. (2004), *Metodi statistici per l’analisi di politiche attive per il lavoro: la legge regionale della Toscana n. 27/93*, Tesi di Dottorato in Statistica Applicata, Università degli Studi di Firenze.

Dehejia R., Wahba S. (1999), “Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs”, *Journal of the American Statistical Association*.

De Magistris V., Gioioso G. (a cura di) (2005), *Nuovi profili di accountability nelle P.A. italiane: Teoria e strumenti*, Collana Quaderni Formez n. 40.

Fabbris L. (1997), *Statistica multivariata. Analisi esplorativa dei dati*, McGraw-Hill, Milano.

Franceschini F., Trivellato U. (1998), “L’impatto delle liste di mobilità nel Veneto: un sussidio di disoccupazione o uno strumento di politica attiva del lavoro?”, in Agenzia dell’Impiego del Veneto (a cura di), *Il mercato del lavoro veneto. Tendenze e politiche. Rapporto 1998*, FrancoAngeli, Milano.

Greenberg D., Links D., Mandel M. (2003), *Social Experimentation and Public Policymaking*, The Urban Institute Press, Washington.

Greenberg D., Shroder M. (1997), *Digest of Social Experiments*, The Urban Institute Press, Washington.

Greene W. (2003), *Econometric Analysis*, Prentice-Hall, New Jersey.

Hahn J., Todd P., Van der Klaauw W. (2001), “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design”, *Econometrica*, Vol. 69, No. 1, 201-209.

Hecklo H. (1972), “Review Article: Policy Analysis”, in *British Journal of Political Science*, vol.2, n.1, 83-108.

Heckman J., LaLonde R., Smith J. (1999), “The Economics and Econometrics of Active Labor Market Programs”, in Ashenfelter O., Card D. (a cura di), *Handbook of Labor Economics*, 1865-2097, North-Holland, Amsterdam.

Heckman J., Smith J. (1995), “Assessing the Case for Social Experiments” *Journal of Economic Perspectives*.

Holland P. (1986), "Statistics and causal inference", *Journal of the American Statistical Association*. 81: 945-960.

Holcomb P., Nightingale D. (1999), *Theoretical and Substantive Dimensions in Process and Implementation Studies of Welfare*, The Urban Institute Press, Washington.

Ichino A., Mealli F., Nannicini T. (2003), "Il lavoro interinale in Italia. Trappola del precariato o trampolino verso un impiego stabile?", Rapporto di ricerca per il Ministero del Lavoro e delle Politiche Sociali.

Imbens G. (2004), "Semiparametric Estimation of Average Treatment Effects under Exogeneity: A Review", *Review of Economics and Statistics*.

IRPET (2004), "La valutazione di efficacia delle borse per il diritto allo studio", Rapporto di ricerca per il CNVSU, IRPET, Firenze.

Johnston J. (1993), *Econometrica*, FrancoAngeli, Milano.

LaLonde R. (1986), "Evaluating the Econometric Evaluations of Employment and Training Programs with Experimental Data", *American Economic Review*.

Manpower Demonstration Research Corporation (1980), *Summary and Findings of the National Supported Work Demonstration*, MDRC, New York.

Martini A., Sisti M. (2002), "Indicatori o analisi di performance? Implicazioni dell'esperienza statunitense di *performance measurement*", *Rivista trimestrale di Scienza dell'Amministrazione*.

Mayntz R., (1999), "La teoria della *governance*: sfide e prospettive", *Rivista Italiana di Scienza della Politica*, XXIX, n. 1, 3 - 22.

McDowall D., McCleary R., Meidinger, R., Hay R. (1980), *Interrupted Time Series Analysis*, Sage Publications, Beverly Hills, CA.

Mealli F., Pagni, R. (2001), *Analisi e valutazione delle politiche per le nuove imprese – Il caso della L.R. Toscana n. 27/93*, FrancoAngeli, Milano.

Mohr L.B. (1995), *Impact analysis for program evaluation*, Sage Publications, Beverly Hills, CA.

Moffit R. (1991), "Program evaluation with nonexperimental data", *Evaluation Review*, 15: 291-314.

Montanino A., Sestito P. (2003), "Le molte funzioni del lavoro interinale in Italia: da strumento di flessibilità a contratto di prova", *Rivista di Politica Economica*.

Mura A., Lobascio I., Persico C. (2005), “*Valutazione degli incentivi della Regione Sardegna: indagine sugli effetti della Legge regionale n. 15/1994*”, Rapporto Tecnico, Osservatorio economico della Sardegna, Cagliari.

Orr L. (1999), *Social Experiments: Evaluating Public Programs with Experimental Methods*, Sage Publications, Beverly Hills.

Orr et al. (1996), *Does Training for the Disadvantaged Work? Evidence from the National JTPA Study*, The Urban Institute Press, Washington.

Palumbo M. (2001), *Il processo di valutazione*, FrancoAngeli, Milano.

Pawson R., Tilley N. (1997), *Realistic Evaluation*, Sage Publications, London.

Pressman J., Wildavsky A. (1973), *Implementation - How great expectations in Washington are dashed in Oakland*, University of California Press, Berkeley, CA.

Regonini G. (1989), “*Lo studio delle politiche pubbliche*”, in Panebianco A. (a cura di), *L’analisi della politica*, il Mulino, Bologna.

Rossi P., Freeman H., Lipsey M. (1999), *Evaluation: A Systematic Approach*, Sage Publications, Beverly Hills, CA.

Shadish W., Cook T., Campbell D. (2002), *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Houghton-Mifflin, Boston.

Stame N. (1998), *L’esperienza della valutazione*, SEAM, Roma.

Tanese A. (a cura di) (2004), *Rendere conto ai cittadini – Il bilancio sociale nelle amministrazioni pubbliche*, Edizioni Scientifiche Italiane, Roma.

Trochim W. (1984), *Research Design for Program Evaluation: The Regression-Discontinuity Approach*. Sage Publications, Beverly Hills, CA.

Stampa Tipograf srl - Roma
Finito di stampare nel mese di aprile 2006
Pubblicazione non in vendita