

**LA VALUTAZIONE DELLE POLITICHE DEL LAVORO IN PRESENZA DI
SELEZIONE: MIGLIORARE LA TEORIA, I METODI O I DATI?**

Enrico Rettore[°], Ugo Trivellato[°] e Alberto Martini*

[°]Dipartimento di Scienze Statistiche, Università di Padova

**Dipartimento di Politiche Pubbliche e Scelte Collettive, Università del Piemonte Orientale*

Padova, agosto 2002

LA VALUTAZIONE DELLE POLITICHE DEL LAVORO IN PRESENZA DI SELEZIONE: MIGLIORARE LA TEORIA, I METODI O I DATI?*

Enrico Rettore^o, Ugo Trivellato^o e Alberto Martini*

^o*Dipartimento di Scienze Statistiche, Università di Padova*

**Dipartimento di Politiche Pubbliche e Scelte Collettive, Università del Piemonte Orientale*

1. Introduzione

Gli esiti di un intervento sociale volto a modificare in un senso desiderato la condizione e/o il comportamento dei suoi destinatari sono spesso incerti. Incerta, in particolare, è l'attribuzione (in senso causale) all'intervento dei cambiamenti osservati nella condizione e/o comportamento in questione. Il miglioramento osservato è merito dell'intervento o si sarebbe verificato comunque? Il peggioramento osservato sarebbe stato più grave in assenza di intervento?

Questo tipo di incertezza è la principale motivazione per il ricorso a procedure di valutazione degli effetti di un intervento (o programma o politica: termini che useremo come sinonimi). L'obiettivo conoscitivo di tale valutazione – che, in accordo con la terminologia prevalente nella letteratura, chiameremo valutazione di *impatto* dell'intervento – consiste nello stabilire se la condizione/comportamento dei destinatari si sia modificata *grazie* all'intervento stesso.

È appena il caso di precisare che la valutazione d'impatto non esaurisce lo spettro degli obiettivi conoscitivi rilevanti per la valutazione di interventi sociali. In parecchie circostanze le questioni su cui la valutazione intende fare luce (detto altrimenti, le domande a cui si è interessati) attengono all'implementazione e alla gestione dell'intervento e ai modi per migliorarla¹. L'ambito per il quale le tematiche che affrontiamo in questo articolo sono rilevanti è, dunque, circoscritto. Ma, allo stesso tempo, è un ambito di indubbio interesse, anche per l'importanza crescente che hanno assunto interventi che si pongono obiettivi ambiziosi di mutamento di condizioni/comportamenti sociali, e quindi parecchio incerti nei loro esiti: valga, per tutte, l'esempio delle cosiddette 'politiche attive del lavoro'².

* I precedenti di questo articolo sono in due relazioni, presentate al Convegno SIS "Processi e metodi statistici di valutazione" (Roma, 4-6 giugno 2001) e al XVI Convegno nazionale di Economia del Lavoro (Firenze, 4-5 ottobre 2001). La ricerca è stata svolta in parte nell'ambito del progetto "Dinamiche e persistenze nel mercato del lavoro italiano ed effetti di politiche (basi di dati, misura, analisi)", cofinanziato dal MIUR. Siamo grati ai partecipanti ai due incontri, e in particolare ad Andrea Ichino e Guido Pellegrini, per commenti e suggerimenti.

¹ Nella polarizzazione di Martini (1997), pp. 8-9, questa diversa prospettiva di valutazione è chiamata "*valutazione come controllo*" usata come "*strumento di management pubblico*", ed è contrapposta alla "*valutazione come apprendimento*" utilizzabile come "*strumento per il disegno delle politiche pubbliche*".

² Per rassegne o raccolte di saggi sulla valutazione dell'impatto di politiche attive del lavoro vedi OECD (1991), Manski e Garfinkel (1992), Jensen e Madsen (1993), Calmfors (1994), Heckman, LaLonde e Smith (1999, sez. 10), Lechner e Pfeiffer (2001). Sulla valutazione di politiche attive del lavoro in Italia, vedi Samek Ludovici (1996) e Borzaga e Brunello (1997). Una riflessione su questioni di disegno e di valutazione delle politiche attive del lavoro nel nostro paese, incentrata sulle condizioni al contorno perché si possano condurre valutazioni d'impatto e si possa quindi imparare dalle stesse – in altre parole, sulla *policy* per la

In termini molto generali, stabilire quale sia l'impatto di un intervento sociale sui suoi destinatari richiede di confrontare la condizione/comportamento che i soggetti esposti all'intervento manifestano successivamente all'intervento con la condizione/comportamento che gli *stessi* soggetti avrebbero manifestato nello *stesso* periodo in assenza di esposizione all'intervento. Palesemente ciò non è possibile, perché non è dato di osservare ciò che sarebbe successo, in assenza dell'intervento, ai soggetti esposti – la *situazione controfattuale*. Ne segue che la questione centrale della valutazione di impatto consiste nell'ottenere una adeguata approssimazione di questa condizione/comportamento controfattuale.

La strategia di valutazione più comunemente utilizzata per stimare l'impatto di un intervento consiste nell'ottenere un'approssimazione della situazione controfattuale ricorrendo alla condizione/comportamento registrata per un conveniente gruppo di soggetti *non* esposti all'intervento, il cosiddetto *gruppo di confronto o di controllo*. L'ovvia minaccia alla validità di tale modo di procedere consiste nel fatto che i soggetti non esposti potrebbero essere sistematicamente diversi dai soggetti esposti per ragioni che non dipendono dal successo dell'intervento, bensì dal processo attraverso il quale gli uni hanno scelto di (o sono stati scelti per) essere esposti all'intervento e gli altri ne sono stati, invece, esclusi.

Questa 'differenza nelle condizioni di partenza' nella letteratura sulla valutazione prende il nome di *selection bias*. Si tratta di un *bias* nel senso che dà luogo a una distorsione dello stimatore dell'impatto ottenuto dal confronto tra gli esiti conseguiti, rispettivamente, dagli esposti e dai non esposti. Si tratta di *bias* dovuto a *selection* nel senso che è il processo di (auto)selezione dei soggetti nei due gruppi degli esposti e dei non esposti a renderli potenzialmente diversi in modo sistematico fin dall'origine, vale a dire già prima che l'intervento abbia luogo.

Se il processo di selezione avesse luogo mediante *assegnazione casuale* dei soggetti all'uno o all'altro dei due gruppi, il *selection bias* risulterebbe nullo *per costruzione*. Pertanto, l'impatto medio dell'intervento verrebbe stimato in modo esente da distorsione mediante la differenza tra i risultati medi ottenuti dai soggetti, rispettivamente, esposti e non esposti all'intervento. Da qui, appunto, viene la naturale attrattiva che esercita il ricorso all'esperimento con assegnazione casuale, come modo per risolvere il problema dell'(auto)selezione, eliminandolo alla radice.

Tuttavia, è assai raro che i soggetti esposti ad un intervento sociale siano selezionati casualmente. Questa evenienza è largamente circoscritta ai casi in cui un intervento sia messo in atto su scala ridotta sotto forma di progetto-pilota, con lo scopo specifico di valutarne l'efficacia – le cosiddette *demonstrations*, utilizzate soprattutto nell'esperienza statunitense. Casi di utilizzo di assegnazione casuale nell'ambito di interventi sociali a regime sono ancora più rari, anche nell'esperienza statunitense³.

Al di fuori di questi ambiti circoscritti, l'assegnazione dei soggetti ad un intervento sociale è frutto di un processo sistematico di selezione che non è controllabile e manipolabile dall'analista-valutatore. Tocca quindi all'analista il compito di ricostruire nei dettagli il processo di selezione dei soggetti esposti, al fine di identificare l'insieme delle caratteristiche rispetto alle quali essi risultano verosimilmente diversi in modo sistematico

valutazione –, è in Rettore e Trivellato (1999). Recentemente un ampio insieme di contributi, peraltro solo in parte in tema di valutazione di impatto, è stato presentato al Convegno SIS "Processi e metodi statistici di valutazione", Roma, 4-6 giugno 2001 (vedi Società Italiana di Statistica, 2001).

³ Sulle ragioni per cui è risultato di fatto impossibile preservare a questi '*social experiments*' le caratteristiche del disegno sperimentale, ci fermiamo brevemente nella sez. 3.

dai soggetti non esposti. La scelta dello stimatore dell'impatto dell'intervento poggia su tale analisi, nel senso che viene costruito in modo tale da (cercare di) rendere nullo la *selection bias* causato dalle differenze sistematiche fra i due gruppi rispetto alle variabili identificate dall'analisi del processo di selezione.

Negli ultimi due decenni si è accumulata sull'argomento un'imponente letteratura segnata da contributi di notevole rilievo analitico, grazie anche al fecondo intreccio fra approfondimenti metodologici e studi empirici di valutazione. La rigorosa ed esaustiva rassegna di Heckman, LaLonde e Smith (1999) – una monografia più che il capitolo di un *Handbook* – ne è forse la testimonianza più eloquente. In questo articolo riprendiamo e sviluppiamo tre affermazioni di Heckman, LaLonde e Smith (1999), che giudichiamo particolarmente rilevanti con riferimento allo stadio di maturazione cui è giunta la pratica della valutazione di interventi sociali nel nostro paese. Le tre affermazioni, leggermente riformulate, sono le seguenti.

- (a) I dati che sono solitamente disponibili per la valutazione di impatto di interventi sociali sono anche nella maggior parte dei casi strutturalmente insufficienti ad eliminare la *selection bias*. Per rimediare a tale insufficienza nella pratica della valutazione si tende a chiedere troppo ai metodi statistico-econometrici. La migliore soluzione consiste nel migliorare l'adeguatezza dei dati raccolti ai fini di valutare gli effetti degli interventi, non nello sviluppare metodi statistico-econometrici sempre più sofisticati.
- (b) È fondamentale che il gruppo di confronto sia davvero tale, cioè sia costituito da soggetti confrontabili con i soggetti esposti all'intervento. Molti studi osservazionali ottengono l'identificazione dell'impatto di un programma ricorrendo ad assunzioni arbitrarie sulle forme funzionali di certe funzioni di regressione, al fine di rendere confrontabili soggetti che in realtà confrontabili non sono. Da questo punto di vista, il pregio principale degli stimatori non parametrici dell'impatto consiste nel fatto che obbligano l'analista a confrontare soggetti confrontabili.
- (c) L'impatto degli interventi sociali può variare, e abitualmente varia, da soggetto a soggetto. Tale variabilità pone all'analista due problemi principali: la semplice identificazione dell'impatto medio è largamente insufficiente alla comprensione degli esiti del programma stesso; la stessa nozione di impatto medio è ambigua e dà luogo ad equivoci se non adeguatamente circostanziata, nel senso che stimatori diversi possono dar luogo a stime di impatti medi molto diverse solo perché si è in presenza di stime relative a sottopopolazioni diverse.

Muovendo da queste proposizioni, ci proponiamo di esporre in maniera piana i termini dello snodo valutazione d'impatto-*selection bias* e gli apporti che possono venire alla soluzione del problema da una chiara esplicitazione degli obiettivi dell'analisi empirica e dalla pertinente combinazione di dati adeguati, di metodi appropriati e – se necessario – di fondate e parsimoniose ipotesi di comportamento (quella che usualmente viene detta 'teoria economica'). Allo svolgimento delle argomentazioni affianchiamo, in chiave esemplificativa, riferimenti ad alcune esperienze italiane di valutazione.

Dopo aver richiamato, nella sezione 2, i termini essenziali del *selection bias*, discutiamo brevemente dell'impiego del metodo sperimentale a scopi di valutazione di politiche sociali (sezione 3). Nella sezione 4 tocchiamo un nodo cruciale per il controllo del *selection bias*: dati migliori o metodi più sofisticati? L'importanza di effettuare confronti appropriati fra soggetti esposti e soggetti non esposti all'intervento, senza ricorrere ad assunzioni parametriche sovente arbitrarie, è discussa nella sezione 5. Nella sezione successiva ci soffermiamo sulle complicazioni che sorgono quando l'obiettivo della valutazione non sia (o comunque non sia circoscritto a) l'impatto medio, ma riguardi

(anche) altre caratteristiche della distribuzione degli effetti. Segue, nella sezione 7, un commento conclusivo

2. Una presentazione essenziale del *selection bias*

Sia Y la condizione/comportamento dei soggetti – appartenenti ad una certa popolazione obiettivo – che l'intervento in questione intende modificare. Sia I la variabile binaria che identifica i soggetti esposti, $I=1$, e i soggetti non esposti, $I=0$, all'intervento.

Sia Y_i^T il modo in cui la condizione/comportamento si presenterebbe per l' i -esimo soggetto se venisse esposto all'intervento, e sia Y_i^{NT} il modo in cui la condizione/comportamento si presenterebbe se lo stesso soggetto non venisse esposto all'intervento. Denominiamo tali variabili *risultati potenziali* relativi all' i -esimo soggetto. Per l' i -esimo soggetto l'esposizione all'intervento *causa* una variazione nella condizione/comportamento in questione pari a:

$$\alpha_i = Y_i^T - Y_i^{NT}. \quad (1)$$

Denominiamo tale variazione *impatto* dell'intervento sull' i -esimo soggetto.

Manifestamente α_i non è osservabile, dato che per ognuno dei soggetti risulta osservabile solo la quantità:

$$Y_i = Y_i^T I_i - Y_i^{NT} (1 - I_i), \quad (2)$$

vale a dire l'uno oppure l'altro dei due risultati potenziali, a seconda del valore assunto da I_i . Da qui nasce, per l'appunto, quello che Heckman, LaLonde e Smith (1999, p. 1879) definiscono “*the fundamental evaluation problem*”⁴.

Stabilita la non identificabilità dell'impatto sui singoli soggetti, giocoforza ci si accontenta di studiare l'identificabilità di singoli aspetti della *distribuzione dell'impatto* nella popolazione. Fermiamo per ora l'attenzione sulla sua media.

Se l'assegnazione all'uno o all'altro dei due gruppi avesse luogo mediante *assegnazione casuale*, varrebbe per costruzione la seguente condizione di ortogonalità:

$$(Y^T, Y^{NT}) \perp I. \quad (3)$$

La conseguenza immediata della (3) è che le distribuzioni dei risultati osservati sui soggetti, rispettivamente, esposti e non esposti coincidono con le corrispondenti distribuzioni nella popolazione:

$$F_T(y|I=1) = F_T(y), \quad (4.1)$$

$$F_{NT}(y|I=0) = F_{NT}(y). \quad (4.2)$$

Ne discende che la differenza tra le medie dei risultati ottenuti dai due gruppi coincide con l'*impatto medio* dell'intervento *nella popolazione*:

$$E(Y^T|I=1) - E(Y^{NT}|I=0) = E(Y^T) - E(Y^{NT}) = E(Y^T - Y^{NT}). \quad (5)$$

Merita di essere notato che la seconda eguaglianza nella (5) vale grazie alla *linearità* dell'operatore valore medio. Il punto è tutt'altro che trascurabile, anche per le sue

⁴ Vale la pena di notare che l'osservabilità di uno soltanto dei due risultati potenziali ha implicazioni ben più ampie, di portata epistemologica generale. Seguendo Holland (1986, p. 947) possiamo ben dire, infatti, che essa costituisce “*the fundamental problem of causal inference*”.

conseguenze pratiche. Ad esempio, se l'aspetto di interesse della distribuzione dell'impatto fosse la sua varianza, la (3) non sarebbe di alcun aiuto perché consentirebbe solo di identificare le varianze dei due risultati potenziali nella popolazione, ma non la varianza della loro differenza:

$$\text{var}(Y^T|I=1) - \text{var}(Y^{NT}|I=0) = \text{var}(Y^T) - \text{var}(Y^{NT}) \neq \text{var}(Y^T - Y^{NT}) . \quad (6)$$

Se il processo che seleziona il gruppo degli esposti dalla popolazione *non* ha luogo mediante assegnazione casuale, la differenza tra le medie dei risultati ottenuti dai due gruppi in generale *non* corrisponde all'impatto medio nella popolazione:

$$\begin{aligned} E(Y^T|I=1) - E(Y^{NT}|I=0) &= \\ &= [E(Y^T|I=1) - E(Y^{NT}|I=1)] + [E(Y^{NT}|I=1) - E(Y^{NT}|I=0)] . \end{aligned} \quad (7)$$

La prima differenza entro parentesi al secondo membro della (7) è l'*impatto medio tra gli esposti*. In generale, esso *non* corrisponde all'impatto medio nella popolazione, proprio perché il valore di I non è determinato mediante assegnazione casuale. In altre parole, in presenza di variabilità dell'impatto non c'è ragione per attendersi che impatto medio tra gli esposti e impatto medio nella popolazione coincidano. D'altra parte, v'è da sottolineare che esso fornisce un'informazione tutt'altro che priva di importanza: in molti casi, anzi, il parametro di interesse è proprio l'impatto medio tra gli esposti.

La seconda differenza entro parentesi al secondo membro della (7) fornisce le differenze medie tra i due gruppi rispetto al risultato potenziale Y^{NT} , vale a dire le differenze medie che si sarebbero osservate tra i due gruppi *in assenza* dell'intervento. Palesemente, tale termine può risultare diverso da zero *solo* come conseguenza del modo in cui sono stati selezionati i due gruppi. Tale termine prende, per l'appunto, il nome di *selection bias*.

Il *selection bias* risulta diverso da zero nelle situazioni in cui il processo di selezione dà luogo ad un gruppo di soggetti esposti all'intervento sistematicamente diversi dai non esposti rispetto a certe caratteristiche X , al variare delle quali il risultato potenziale Y^{NT} assume mediamente valori diversi. Il *selection bias* è, dunque, non nullo se con riferimento ad un opportuno insieme di variabili X valgono simultaneamente le due condizioni:

$$F_X(x|I=1) \neq F_X(x|I=0) , \quad (8.1)$$

$$E(Y^{NT}|X) \neq E(Y^{NT}) , \quad (8.2)$$

F_X essendo la distribuzione di probabilità della variabile X .

Avendo a disposizione i risultati ottenuti da un gruppo di esposti e da un gruppo di non esposti, il tipico problema – ma, come avremo modo di argomentare, non l'unico interessante – affrontato in sede di valutazione dell'impatto di un intervento consiste nel depurare la differenza tra le medie (7) dal *selection bias*, in modo tale da mettere in luce l'impatto medio sugli esposti.

Nelle situazioni nelle quali i dati a disposizione sono di tipo sezionale, le soluzioni al problema posto dal *selection bias* sono essenzialmente due: (i) operare il confronto tra esposti e non esposti *a parità di valori assunti dalle variabili X* in (8); (ii) ricorrere a (almeno) una variabile che influenzi la probabilità di esposizione all'intervento e il cui valore sia assegnato in modo assimilabile ad una assegnazione casuale⁵.

⁵ La disponibilità di dati longitudinali – nel nostro caso di osservazioni ripetute su esposti e non esposti per periodi *precedenti e seguenti* l'intervento – consente l'applicazione di particolari metodi per eliminare il

Quanto alla prima soluzione, se l'analisi del processo di selezione identifica correttamente *tutte* le variabili X per le quali vale la (8) e se tali variabili *sono osservate* dall'analista, risulta possibile operare il confronto tra esposti e non esposti *a parità di X* , vale a dire rispetto a tutte le variabili che causano *selection bias*. Infatti, il condizionamento a X comporta la nullità del *selection bias*:

$$E(Y^{NT}|I=1, X) = E(Y^{NT}|I=0, X), \quad (9)$$

proprio perché le variabili alle quali ci si condiziona sono tutte, e sole, quelle rispetto alle quali i due gruppi presentano una diversa composizione (condizione 8.1) e al variare delle quali la media di Y^{NT} varia (condizione 8.2). Ne segue che la differenza tra esposti e non esposti *a parità di X* :

$$E(Y^T|I=1, X) - E(Y^{NT}|I=0, X) \quad (10)$$

corrisponde all'impatto medio sugli esposti in corrispondenza di *quel* particolare valore di X . La media di tale impatto calcolata rispetto a X fornisce l'impatto medio sull'insieme degli esposti.

Quanto alla seconda soluzione, se nel processo di selezione degli esposti interviene una variabile Z – che nel seguito per semplicità di esposizione assumiamo binaria – tale per cui al suo variare varia la probabilità di esposizione:

$$Pr(I=1|Z) \neq Pr(I=1) \quad (11.1)$$

e che assume valori in modo indipendente dal processo che genera i risultati potenziali:

$$(Y^T, Y^{NT}) \perp Z, \quad (11.2)$$

è intuitivamente chiaro che ne discende un assetto logicamente prossimo a quello nel quale I è assegnato casualmente. La prossimità è tanto più forte quanto più è forte l'influenza esercitata da Z su I (se Z determinasse univocamente I ne risulterebbe esattamente il caso in (3)). In effetti, un risultato classico dell'econometria mostra che valendo le (11) e limitatamente *alle situazioni in cui l'impatto non varia da soggetto a soggetto* lo stimatore delle variabili strumentali:

$$\hat{\alpha} = \sum_i (Y_i - \bar{Y}) Z_i / \sum_i (I_i - \bar{I}) Z_i \quad (12)$$

è consistente per l'impatto dell'intervento (vedi ad esempio Wooldridge, 1999, pp. 464-465)⁶.

selection bias (o per verificarne la presenza). In generale tali metodi sono preferibili proprio perché la più ricca informazione empirica a disposizione – estesa alle variazioni intraindividuali – permette di allentare assunti adottati nella specificazione del processo di selezione sulla base di dati *cross-section* (o di trattarli come ipotesi passibili di verifica). Vedi, ad esempio, Heckman, LaLonde e Smith (1999), sez. 7.6. *Mutatis mutandis*, tuttavia, la logica delineata nel testo principale per dati di tipo sezionale resta valida.

⁶ A conclusione di questi essenziali richiami su valutazione d'impatto e *selection bias*, vale la pena di notare che, qui e nel seguito dell'articolo, omettiamo di considerare altri aspetti potenzialmente rilevanti degli effetti di un intervento, riferiti da un lato ai non esposti – usualmente distinti in *substitution effects* e *displacement costs* – e dall'altro all'intero sistema in termini di effetti dinamici di equilibrio generale. Per una lucida rassegna di tali questioni, vedi Calmfors (1994). Per un recente esame di approcci di misura (anche) di tali effetti, incentrato sull'esperienza inglese di *targeted wage subsidies*, vedi Bell, Blundell e Van Reenen (1999). Ai fini di valutazione dell'impatto tra gli esposti, un'importante distinzione si impone fra effetti indiretti sui non esposti, del tipo *substitution effects* e simili, ed effetti dinamici di equilibrio generale. Ignorare questi ultimi, infatti, può portare alla errata identificazione dello stesso impatto medio sugli esposti, quando anch'esso sia influenzato da *feedback* indotti dal processo dinamico di equilibrio generale. Trascurare

3. Il metodo sperimentale: panacea per la valutazione o poco informativa ‘scatola nera’?

Visti i problemi di valutazione dell’impatto medio di un intervento che si incontrano quando ci si allontani dall’assegnazione casuale dei soggetti a uno o all’altro dei gruppi – gli esposti ed i non esposti, è naturale cercare di risolvere la questione *ex ante*, cioè a dire in sede di disegno stesso dell’intervento sociale, realizzandolo appunto in termini di ‘*social experiment*’.

L’uso di disegni sperimentali per valutare interventi sociali è stato quasi del tutto limitato agli Stati Uniti (Greenberg e Shroder, 1997). L’esperienza statunitense ha messo in luce molto bene sia le potenzialità sia i limiti della sperimentazione come metodo per valutare l’impatto degli interventi sociali. Stimolato da questa esperienza, si è aperto un importante dibattito sulla validità e sull’utilità dei risultati conoscitivi ottenuti mediante gli esperimenti sociali⁷.

I sostenitori del metodo sperimentale hanno buone frecce al loro arco. Il metodo evita gli oneri di raccolta di una larga massa di dati, quelli sulle variabili *X* responsabili del *selection bias* secondo la (8), estesa anche ai soggetti non esposti. Evita soprattutto di ricorrere a dubbie ipotesi comportamentali, e alle fragili assunzioni funzionali in cui si traducono, per controllare fattori non osservati responsabili del *selection bias*. In definitiva, il metodo sperimentale fornisce risultati robusti, e a poco prezzo: per gli analisti, le stime dell’impatto sono facili da calcolare; per gli utilizzatori, sono altrettanto facili da capire.

Gli scettici del metodo sperimentale hanno un arsenale altrettanto nutrito di argomentazioni a loro disposizione. Innanzitutto si possono sollevare le tradizionali obiezioni sugli ostacoli di ordine etico e pratico che si frappongono all’adozione di valutazioni sperimentali, soprattutto per interventi già a regime. La difficoltà non sta solo nel fatto che la decisione di negare, a un insieme di soggetti ammissibili, l’accesso ad un programma sociale esistente si scontra con una lunga serie di impedimenti di natura etica, legale e organizzativa. A ciò si aggiunge la circostanza che, innestandosi su un programma già in atto, il metodo sperimentale inevitabilmente ne altera, e in maniera profonda, le procedure operative. Quando ciò abbia successo (e non sia invece reso impraticabile dall’opposizione degli operatori locali del programma), il programma che viene valutato col metodo sperimentale finisce per essere diverso dal programma esistente che si intendeva valutare.

Queste obiezioni sono valide, e di notevole peso. Tuttavia, esse delineano limiti applicativi al metodo sperimentale, non una strutturale debolezza metodologica, che invece viene ad esso addebitata da altri critici. Soprattutto Heckman e i suoi collaboratori hanno formulato persuasive riserve sulla praticabilità e sulla portata conoscitiva del metodo sperimentale per la valutazione di interventi sociali.

i primi, invece, non preclude l’identificazione dell’impatto medio (o di altre caratteristiche di interesse della distribuzione dell’impatto) per gli esposti: semplicemente, si ignorano gli affetti sugli altri soggetti.

⁷ Per contributi improntati a punti di vista marcatamente diversi, vedi Burtless (1995) e Heckman e Smith (1995).

Quanto alla praticabilità, hanno argomentato come l'assegnazione casuale abbia essa stessa effetti comportamentali che non possono essere ignorati⁸. Anche nell'ipotesi (sovente eroica) che sia effettivamente possibile praticare l'assegnazione casuale dei soggetti ai due gruppi degli esposti e dei non esposti, in ogni esperimento svolto in un contesto sociale – dunque non in laboratorio – si hanno inevitabilmente fenomeni di uscita dal gruppo degli esposti (eventualmente dopo aver beneficiato in misura parziale del trattamento) e, all'opposto, di accesso di soggetti del gruppo dei non esposti a servizi parecchio prossimi a quelli somministrati agli esposti, offerti da altri enti e/o nell'ambito di altri programmi⁹. Ciò non significa che la stima di impatti medi di interesse diventi impossibile. Tali parametri restano identificabili, ma solo a prezzo di ulteriori assunzioni e condizioni. L'impiego del metodo sperimentale, dunque, difficilmente può prescindere da un'attenta considerazione degli effetti comportamentali che esso stesso induce, e in definitiva la stessa polarizzazione fra metodo sperimentale *theory-free* e metodi econometrici *model-based* è almeno in parte di maniera.

Lo stereotipo della superiorità del metodo sperimentale è, poi, ulteriormente messo in discussione quando le esigenze conoscitive di un esercizio di valutazione siano più articolate: in altre parole, quando i parametri di interesse riguardino aspetti distributivi degli effetti dell'intervento, e non solo gli effetti medi. Ci fermeremo più diffusamente sull'argomento nella sezione 6, quando discuteremo delle complicazioni derivanti dalla variabilità dell'impatto.

Infine, merita di essere richiamata la critica a cui il metodo sperimentale è stato sottoposto da un gruppo di teorici della valutazione inglesi (vedi in particolare Pawson e Tilley, 1997)¹⁰, per la povertà di risultati conoscitivi che essa produrrebbe, anche una volta superati tutti gli ostacoli pratici e concettuali elencati sopra (e quindi nella migliore delle ipotesi). Il metodo sperimentale ci dice 'che cosa funziona in media', ma non fornisce alcuna risposta a molte altre domande di reale interesse nel processo di *policy-making*, riassumibili nella frase: "Che cosa funziona, per chi, in quale contesto, e soprattutto perché?".

4. Migliorare l'adeguatezza dei dati (più che sviluppare metodi sofisticati)

4.1. Chiarimenti metodologici

⁸ Un'importante esperienza per le riflessioni critiche sul metodo sperimentale, a fini di valutazione di interventi sociali, è stata offerta dal *Job Training Partnership Act* statunitense. Vedi Orr *et al.* (1995) e Heckman e Smith (1995).

⁹ Nella letteratura statunitense, sovente ci si riferisce a questi due problemi, non peculiari alle valutazioni con metodo sperimentale ma indubbiamente esacerbati nelle stesse, rispettivamente con i termini di "*treatment group dropout*" e di "*control group substitution bias*" (vedi, ad esempio, Heckman, Lalonde e Smith, 1999, sezz. 5.2 e 5.3.2).

¹⁰ Si tratta di ricercatori di area sociologica, con marcati interessi epistemologici e con esperienze di ricerca empirica legate soprattutto al *social work*. A quanto è dato di vedere, essi hanno scarsa o addirittura nessuna familiarità con la letteratura economica e statistico-econometrica sulla valutazione: significativamente, essa neppure fa capolino nella pur corposa bibliografia del volume di Pawson e Tilley (1997). È interessante notare come, pur muovendo da un contesto di riferimenti teorici e di ricerca empirica affatto diverso, questi sociologi – che si autodefiniscono "*realist*" – svolgano una critica del metodo sperimentale che in parecchi punti echeggia, *mutatis mutandis*, quella di Heckman, e in altri la articola ulteriormente.

Tornando alla discussione delle tre proposizioni di Heckman, LaLonde e Smith (1999), supponiamo che l'analisi del processo di selezione abbia identificato in certe variabili X la fonte del *selection bias*. Detto in altre parole, esposti e non esposti otterrebbero, *in assenza dell'intervento*, risultati medi diversi perché, come esito del processo di selezione, i due insiemi sono composti in modo diverso rispetto a tali X rilevanti per il risultato potenziale Y^{NT} .

Nel caso in cui le variabili X siano tutte osservabili, un semplice modo per ottenere una stima dell'impatto medio sugli esposti consiste nell'abbinare ad ogni soggetto esposto un soggetto non esposto che presenta caratteristiche X simili e nel calcolare la media delle differenze tra i risultati osservati per le coppie di soggetti abbinati:

$$\hat{E}(Y^T - Y^{NT}|I=1) = \sum_i (Y_i^T - Y_{(i)}^{NT})/n, \quad (13)$$

dove n è pari al numero di soggetti esposti e $Y_{(i)}^{NT}$ è risultato ottenuto dal non esposto abbinato all' i -esimo esposto, sicché vale $X_i \cong X_{(i)}$. La (13) fornisce uno stimatore non parametrico dell'impatto medio – o meglio, una famiglia di stimatori, perché vi sono molteplici schemi di abbinamento che possono essere adottati.

Palesamente, il calcolo dello stimatore (13) richiede che per *tutti* i soggetti coinvolti nell'analisi siano osservabili *tutte* le variabili incluse in X . L'esperienza accumulata nella valutazione di impatto insegna che molto difficilmente le tipiche indagini correnti (e le tipiche basi di dati amministrativi) rendono disponibili le variabili necessarie allo scopo. Ne segue che all'analista, il quale può contare solo sulle informazioni rese disponibili da tali fonti, (gran) parte delle caratteristiche individuali che si congetturano essere responsabili del *selection bias* risultano *non osservabili*.

Da tale situazione è derivata una grande attenzione allo sviluppo di stimatori dell'impatto in presenza della cosiddetta *eterogeneità non osservabile* tra soggetti (ad esempio, negli anni '80 ha riscosso grande popolarità lo stimatore in due passi di Heckman, 1979). Ma vi è ormai un consolidato consenso sul fatto che tali stimatori danno luogo a stime dell'impatto poco plausibili, in quanto poggiano su assunti arbitrari circa la distribuzione delle variabili non osservabili¹¹.

L'ovvia – anche se assai più impegnativa – alternativa consiste nella *predisposizione del supporto informativo del quale la valutazione ha bisogno*. È inutile insistere con l'utilizzare dati la cui produzione non è stata progettata – e comunque non è idonea – a scopo valutativo. Se le indagini correnti e le fonti amministrative non rendono disponibili le variabili necessarie alla valutazione, occorre procedere alla loro rilevazione mediante un'indagine *ad hoc* tagliata sui bisogni della specifica valutazione. È nell'ambito della valutazione stessa dell'intervento che si deve provvedere a reperire *tutte* le informazioni necessarie alla valutazione. L'alternativa rischia di essere secca: la rinuncia al tentativo di stimare l'impatto dell'intervento.

4.2. Qualche evidenza dall'esperienza italiana

Se si considera lo stato della valutazione di interventi sociali nel nostro paese, emerge con chiarezza la necessità di un autentico salto di qualità. Occorre passare da valutazioni

¹¹ Inoltre, anche se tali assunti fossero corretti le proprietà degli stimatori che ne risultano sono decisamente scadenti (Copas e Li, 1997; Rotnitzky *et al.*, 2000).

artigianali condotte sfruttando, con ingegnosità acuita dal bisogno, quel po' di informazioni che si riesce a racimolare tra indagini correnti e archivi risultanti da processi amministrativi, a valutazioni pianificate per tempo (e adeguatamente finanziate), che si dotano del supporto informativo necessario alla soluzione del problema di valutazione in esame¹².

A monte di questa situazione – e a darne conto – stanno ritardi di indole generale, culturale e di 'politica della valutazione', sui quali ci siamo soffermati in una nota di un paio di anni fa e che ci hanno indotto a parlare di “*assenza di un approccio maturo al disegno e alla valutazione delle politiche*” (Rettore e Trivellato, 1999, p. 897)¹³. In questa sede non torniamo su tali aspetti. Né, d'altra parte, intendiamo certo disconoscere i decisi, corposi progressi – veri e propri segnali di discontinuità – testimoniati dai due rapporti di monitoraggio sulle politiche occupazionali e del lavoro prodotti dal Gruppo di Lavoro *ad hoc* del Ministero del Lavoro¹⁴.

Ma modificare lo stato delle cose è opera di lunga lena. Sinora gli avanzamenti ai quali abbiamo accennato sono riusciti al più a scalfire ritardi e inadeguatezze, in particolare per il punto sul quale vogliamo fermare l'attenzione: la carenza delle basi informative necessarie per poter svolgere adeguate valutazioni.

La ricognizione di recenti esperienze di valutazione di tre distinti interventi – le misure in favore dell'imprenditoria giovanile disposte con la legge 44/1986, le liste di mobilità introdotte con la legge 223/1991, alcuni programmi di formazione professionale – consente di suffragare in maniera convincente l'affermazione.

4.2.1. *Gli interventi di incentivazione dell'imprenditorialità giovanile*

Battistin, Gavosto e Rettore (2001) discutono degli interventi di incentivazione dell'imprenditorialità giovanile varati con la legge 44/1986. Si tratta di provvedimenti che, all'epoca della loro introduzione, hanno rappresentato un'innovazione di notevole portata in materia di interventi pubblici di sostegno alle imprese. Ridotto ai suoi termini essenziali, l'intervento nella sua formulazione originaria consiste in questo. La popolazione

¹² Una pianificazione tempestiva è essenziale, perché dopo la messa in atto dell'intervento possono venire meno le condizioni favorevoli (o le condizioni *tout court*) per la rilevazione. Per una discussione sulla relazione fra i tempi dell'intervento e i tempi della rilevazione, vedi Martini e Garibaldi (1993).

¹³ In termini non molto dissimili si sono recentemente espressi anche ricercatori dell'Isfol, l'istituto cui è affidata l'attività di monitoraggio e valutazione degli interventi cofinanziati dal Fondo Sociale Europeo. Scrivono Centra, De Vincenzi e Villante (2000, p. 17): “*Le analisi ... degli impatti della formazione professionale sui destinatari degli interventi, fino a poco tempo fa scontavano un'arretratezza culturale e un vuoto informativo abbastanza evidente soprattutto se messo a confronto con l'operato degli altri Paesi membri dell'UE*”.

¹⁴ Vedi Ministero del Lavoro e della Previdenza Sociale (2000) e (2001). I rapporti sono stati stimolati dal cosiddetto 'processo di Lussemburgo', attivato dall'UE nel 1997, dagli impegni comunitari connessi alla redazione dell'annuale *National Action Plan* (NAP) per l'occupazione dell'Italia. Soprattutto il Rapporto n. 2/2000, peraltro, ha assunto un autonomo rilievo conoscitivo rispetto al NAP per l'occupazione. Dei rapporti meritano di essere segnalati perlomeno due aspetti: (i) la chiarezza dell'impostazione, con la nitida distinzione fra documentazione statistica di quadro, monitoraggio delle politiche del lavoro e valutazione delle stesse; (ii) per l'appunto, la presentazione di alcuni contributi, ancora *in progress*, ma di sicuro interesse, di valutazione dell'impatto di alcune politiche (vedi i Box A3 “Iscrizione al collocamento e *chances* occupazionali” e B1 “Part-time e *chances* occupazionali future”). Altri segnali promettenti vengono da contributi, pur diseguali, maturati nell'ambito di altre strutture pubbliche di osservazione/monitoraggio del mercato del lavoro: vedi, tra gli altri, Centra, De Vincenzi e Villante (2000) per l'attività dell'Isfol e Ceccarelli (2000) per l'esperienza della Regione Valle d'Aosta.

ammissibile è costituita da giovani residenti nel Mezzogiorno intenzionati ad avviare un'impresa nella stessa area. I richiedenti devono presentare un *business plan* nel quale descrivono il loro progetto e ne discutono le prospettive. Un comitato di valutazione analizza le richieste e seleziona i progetti più promettenti. I progetti ammessi ricevono finanziamenti generosi in parte a fondo perduto in parte sotto forma di mutui agevolati, sia in conto capitale che in conto spese correnti: nel periodo considerato, mediamente gli ammessi hanno ricevuto 1,7 milioni di Euro. È inoltre previsto un periodo di formazione dei neo-imprenditori e una fase durante la quale la loro impresa viene supervisionata da esperti esterni.

La *ratio* dell'intervento pare consistere nella convinzione che nell'area in questione risiedano buoni imprenditori potenziali, in particolare tra i giovani, impossibilitati a dare corso ai loro progetti imprenditoriali a causa delle condizioni ambientali sfavorevoli. L'intervento si propone di far uscire allo scoperto questi buoni imprenditori potenziali e di aiutarli a superare le difficoltà che si frappongono alla realizzazione dei loro progetti.

All'innovazione nei contenuti introdotta da tale intervento si è accompagnata anche una pregevole opera di monitoraggio sulle imprese che hanno goduto dei benefici previsti dalla legge. Non altrettanto pregevole è stata l'opera di valutazione dei risultati ottenuti grazie alla legge. Nel corso degli anni infatti, si è accumulata l'evidenza che le imprese finanziate vivono molto più a lungo di imprese coetanee sorte spontaneamente e confrontabili quanto ad attività economica e ad area di insediamento. Tale evidenza è stata interpretata come la prova che le imprese selezionate sono di qualità comparativamente elevata. Ne è derivato un diffuso giudizio positivo sull'intervento, al punto che la Commissione Europea lo ha incluso nell'elenco delle *European best practice examples* di politiche attive del lavoro (European Commission, 1997, p. 44).

In effetti, Battistin, Gavosto e Rettore (2001) mostrano che la maggiore sopravvivenza delle imprese finanziate è interamente dovuta agli ingenti finanziamenti dei quali tali imprese hanno goduto. La Figura 1 riassume graficamente questa evidenza. Essa presenta le funzioni di sopravvivenza rispettivamente delle imprese finanziate e di un gruppo di confronto di imprese non finanziate, entrambe ancora in vita al 30^o mese. Il condizionamento alla sopravvivenza al 30^o mese – detto grossolanamente, la 'sterilizzazione' dei primi 30 mesi – è cruciale: il 30^o mese, infatti, segna il termine del periodo di percezione (della quasi totalità) dei finanziamenti da parte delle imprese finanziate, periodo durante il quale esse, proprio a motivo dei finanziamenti di cui godono, sostanzialmente *non possono cessare*. È immediato notare che le differenze osservate tra le due funzioni di sopravvivenza non risultano statisticamente significative ai convenzionali livelli di conduzione dei test. È pertanto insostenibile la tesi che le imprese ammesse al programma siano migliori delle altre.

Figura 1 circa qui

Lo studio appena menzionato appartiene a tutti gli effetti al gruppo degli studi di valutazione che abbiamo definito artigianali (con la connotazione anche positiva che il termine implica): condotti, cioè, utilizzando con acume informazioni racimolate da quel che è dato di reperire. Pur ribaltando l'opinione comune sui risultati ottenuti dalla legge, le evidenze sono tutt'altro che definitive. In particolare vanno segnalati due limiti, che

avrebbero potuto essere superati se la valutazione degli esiti della legge fosse stata preordinata.

Innanzitutto, la scelta del gruppo di confronto, costituito da imprese spontanee coetanee delle finanziate, insediate nelle stesse aree geografiche e operanti nelle stesse attività economiche, non è del tutto soddisfacente. Molto meglio si sarebbe potuto fare se fosse stato possibile avere accesso alla lista dei progetti di impresa *non* ammessi al programma e, contemporaneamente, fossero state disponibili informazioni sui criteri di selezione¹⁵.

In secondo luogo, l'utilizzo come variabile risultato della sola durata in vita delle imprese, pur consentendo di concludere in modo convincente che le imprese ammesse al programma non presentano livelli qualitativi superiori a quelli delle imprese spontanee, non consente analisi più fini del modo in cui tali livelli si determinano nelle due popolazioni di imprese (vedi la discussione in Mazzotta, 1999). La distinzione rilevante per il caso in questione è tra inefficienza *tecnica* e inefficienza *allocativa*: la prima causata da errori nella selezione degli ammessi e/o dall'insufficiente formazione loro impartita; la seconda (potenzialmente) indotta dal disegno degli incentivi. È possibile, infatti, che il sistema di incentivi previsto dalla legge sia stato tale da indurre le imprese ammesse al programma a domandare lavoro in misura eccedente il livello ottimo ai prezzi di mercato. Al termine del periodo di percezione dei benefici, vale a dire quando hanno iniziato ad acquisire i fattori produttivi ai prezzi di mercato, tali imprese potrebbero essersi trovate nell'impossibilità di ridurre la domanda di lavoro, a causa delle rigidità associate alla legislazione di protezione dell'impiego. In altre parole, *come risultato degli incentivi previsti dalla legge* le imprese esposte potrebbero essere risultate *allocativamente* inefficienti.

È chiaro che per il (ri)disegno del programma sarebbe della massima importanza distinguere tale tipo di inefficienza dall'inefficienza *tecnica*, conseguente a imperfezioni nel modo di selezionare gli ammessi. Ed è altrettanto chiaro che per lo svolgimento di una valutazione tale da consentire la distinzione, occorre ben altro che informazioni sommarie sulla durata in vita delle imprese.

4.2.2. Il programma 'liste di mobilità'

In via preliminare, è utile segnalare come le 'liste di mobilità' si prestino assai bene per mettere in luce le incertezze circa gli esiti di un intervento. Il programma, introdotto con la legge 223/1991, prevede che le imprese con più di 15 addetti possono collocare i lavoratori in eccedenza in liste di mobilità regionali. Il programma persegue obiettivi sia di mantenimento del reddito (tramite l'erogazione di un'indennità ai lavoratori iscritti alle liste) sia di reinserimento nel lavoro (tramite sgravi contributivi e il trasferimento di parte dell'indennità non ancora goduta dal beneficiario al nuovo datore di lavoro)¹⁶.

¹⁵ Si noti che tali informazioni sono, palesemente, disponibili presso l'agenzia responsabile dell'attuazione della politica, la quale svolge anche l'azione di monitoraggio della stessa. Ciò chiama in causa il tema, delicato e cruciale, dell'accesso alle informazioni necessarie alla valutazione, accesso che dovrebbe essere aperto ad una pluralità di analisti indipendenti. Per riflessioni sull'argomento incentrate sull'importanza di distinguere tra responsabili dell'attuazione e valutatori di una politica, e in particolare sulla necessità di evitare che i primi abbiano il monopolio (di parte) delle informazioni utili per esercizi di valutazione, rimandiamo a Rettore e Trivellato (1999), pp. 901-903.

¹⁶ Il programma è stato successivamente modificato con diverse disposizioni, fra le quali spiccano le leggi n. 236/1993 e 451/1994. Alle liste possono iscriversi anche i lavoratori licenziati da piccole imprese – fino a 15 addetti –, ma per essi l'iscrizione è volontaria; inoltre, tali lavoratori non godono dell'indennità.

L'incertezza in merito agli esiti della politica è di natura teorica, ma è anche connessa alle modalità operative con cui essa è stata specificata. Per quanto riguarda gli aspetti teorici, *coeteris paribus* l'erogazione dell'indennità aumenta il salario di riserva dei lavoratori iscritti alle liste e, per questa via, allunga la durata dell'episodio di disoccupazione. D'altra parte, sempre a parità di altre circostanze, i benefici a favore di chi assume lavoratori in mobilità aumentano la probabilità che tali lavoratori ricevano offerte di lavoro e riducono quindi la durata dell'episodio di disoccupazione. Il segno dell'effetto netto è, dunque, *a priori* incerto, e dipende dalla prevalenza dell'uno o dell'altro dei due effetti.

Quanto alle incertezze connesse alle modalità operative, non è scontato che l'entità dei benefici e l'articolazione adottata per gli stessi – la loro variabilità secondo l'età all'iscrizione alle liste e la dimensione dell'impresa, il modo in cui essi si estinguono progressivamente nel tempo – sia in un qualche senso la migliore¹⁷. È evidente, dunque, che si è di fronte a una politica i cui esiti sono problematici. Ne dovrebbe discendere una forte preoccupazione di valutarne empiricamente l'impatto, e una serrata discussione sulla modulazione dei benefici. E ancor prima – almeno così parrebbe ragionevole attendersi – l'impegno a predisporre una base informativa adeguata a questi scopi.

Niente di tutto ciò, tuttavia, è avvenuto. Gli esercizi di valutazione delle liste di mobilità di cui disponiamo sono ormai numerosi (vedi vari saggi raccolti in Borzaga e Brunello, 1997; Brunello e Miniaci, 1997; Caruso, 2001; Paggiaro e Trivellato, 2002). Ma tutti, o quasi¹⁸, sono stati forzatamente condotti sulla base dei soli archivi amministrativi regionali di gestione dell'ammissione alla politica in questione: parzialmente disomogenei da regione a regione, sistematicamente privi di alcune informazioni chiave, di mediocre qualità. Gli studi più avvertiti hanno cercato di fare fronte alle carenze nella base informativa in due modi: restringendo l'obiettivo alla sola valutazione degli effetti differenziali tra diversi 'regimi' della politica, segnatamente in relazione all'età del lavoratore; specificando modelli statistico-econometrici piuttosto flessibili.

Si tratta di scelte sensate, ma non esenti da inconvenienti. Per un verso, la portata conoscitiva dell'esercizio di valutazione ne viene fortemente circoscritta. Per un altro verso, i risultati restano inevitabilmente fragili. Anche ai fini della valutazione dell'effetto differenziale di un regime della politica rispetto all'altro (tipicamente, il regime per i lavoratori fra i 40 e i 49 anni, rispetto a quello per i lavoratori fino a 39 anni), vi sono due potenziali elementi di notevole debolezza.

(a) Innanzitutto, il periodo di permanenza nella lista è una 'scatola nera'. Gli archivi di gestione delle liste di mobilità, infatti, si limitano a registrare quando un lavoratore entra nella lista, se e quando viene assunto con un contratto a tempo indeterminato, oppure se e quando esce dalla lista per decorrenza del periodo massimo ammissibile. Anche assumendo che tali informazioni siano attendibili (ma vedremo tra poco che per parte di esse ciò non vale), si perde così traccia dei periodi di occupazione con un contratto a tempo determinato, i quali – è importante ricordarlo – comportano il

¹⁷ Ad esempio, si potrebbe argomentare che l'abbassamento dell'anzianità contributiva minima richiesta per accedere a un sussidio di disoccupazione (e, per i lavoratori ammessi alle liste, tale è l'indennità di mobilità) renderebbe più attraenti per i disoccupati le offerte di lavoro che danno scarse garanzie di stabilità e, per questa via, ridurrebbe la durata degli episodi di disoccupazione (così Boeri, 1997).

¹⁸ Un'eccezione è Caroleo *et al.* (1997). La procedura seguita per costruire il gruppo di confronto è, tuttavia, assai sommaria.

‘congelamento’ del beneficio ma non l’uscita del lavoratore dalla lista¹⁹. La conseguenza è che i periodi di permanenza nella lista essendo occupato a tempo determinato sono assimilati a quelli di permanenza come disoccupato. Episodi di permanenza nella lista di uguale durata possono dunque celare situazioni profondamente diverse – al limite, sola disoccupazione o solo lavoro a tempo determinato. Quanto il prezzo di questa semplificazione possa essere alto, in sede di valutazione degli effetti della politica²⁰, non necessita certo di essere sottolineato.

- (b) In secondo luogo, come abbiamo anticipato, la qualità degli archivi è assai mediocre, in particolare per quanto riguarda la registrazione delle uscite dalla lista a seguito di assunzione a tempo indeterminato. La ragione sta nella pluralità dei soggetti coinvolti, con compiti segmentati, nella gestione del programma, e nella corrispondente pluralità di archivi amministrativi che essi utilizzano. In particolare, gli archivi amministrativi regionali delle liste di mobilità sono collegati all’attività delle Commissioni regionali per l’impiego, le quali decidono dell’ammissione dei lavoratori al programma: gli archivi registrano, per l’appunto, gli esiti positivi di tali istruttorie, e sono dunque affidabili per la documentazione degli ingressi; il loro aggiornamento in tema di uscite è lasciato, invece, all’osservanza – da parte dei Centri per l’impiego – di disposizioni burocratiche, prive di implicazioni sostanziali tanto per i lavoratori che per le imprese, ed è dunque, non sorprendentemente, non tempestivo e incompleto²¹.

Ora, i risultati di uno studio-pilota di integrazione di un archivio di gestione delle liste di mobilità con gli archivi di ‘Netlabor’ (Paggiaro e Trivellato, 2001)²² forniscono nitide evidenze a sostegno dell’importanza di queste preoccupazioni. L’abbinamento con Netlabor consente di individuare un numero di transizioni all’occupazione a tempo indeterminato parecchio superiore a quello documentato dagli archivi delle liste: quasi il 50% in più. Inoltre, mostra che, nel periodo di iscrizione dei lavoratori alla lista, frequenza

¹⁹ Per la durata dell’occupazione a tempo determinato (o a tempo parziale), viene sospesa la corresponsione dell’indennità al lavoratore, ma insieme viene ‘fermato l’orologio’ che misura il tempo di permanenza nella lista, per un periodo pari, al massimo, al periodo cui il lavoratore ha diritto (detto altrimenti, il periodo di iscrizione può, al massimo, raddoppiare).

²⁰ L’analisi di valutazione, infatti, è condotta tramite modelli di sopravvivenza, forzatamente in uno spazio a due stati: lo stato di permanenza nella lista (al quale è assimilata la decadenza per raggiungimento della durata massima ammissibile, per appunto con durata censurata all’uscita); lo stato nuova occupazione a tempo indeterminato.

²¹ Gli archivi delle liste di mobilità servono, evidentemente, anche alle imprese che domandano lavoratori dalle stesse, per l’estrazione di elenchi di iscritti sui quali condurre la ricerca. Ma il fatto che le liste siano ridondanti – contengano, cioè, una frazione di lavoratori già occupati a tempo pieno – non determina per le imprese disagi tali da indurle a richiedere una più accurata gestione delle liste stesse. D’altra parte, gli altri soggetti coinvolti nella gestione del programma, cioè l’INPS (cui spetta l’erogazione dei benefici ai lavoratori e alle imprese) e il Ministero del Lavoro (cui spetta l’usuale funzione di ispezione e controllo), per i rispettivi compiti usano propri archivi, gestiti senza alcun collegamento sistematico con quelli delle liste di mobilità.

²² ‘Netlabor’ è il sistema informativo del collocamento e delle attività amministrative pubbliche in materia di lavoro disponibile, tra le altre regioni, per il Veneto. Lo studio è condotto su tre province del Veneto – Belluno, Treviso e Vicenza –, per il periodo 1 gennaio 1995-31 marzo 1999. Il vantaggio dell’integrazione degli archivi non è circoscritto alla ricostruzione delle storie lavorative degli iscritti per il periodo di loro permanenza nella lista. Esso si estende anche ad altri due aspetti: consente di collegare a ciascuno di questi segmenti di storia lavorativa quello precedente (importante per il controllo della distorsione da selezione, in sede di valutazione d’impatto del programma) e quello successivo (utile per disporre di elementi di giudizio sugli effetti di medio periodo del programma); permette di vagliare la coerenza, e indirettamente la qualità, delle informazioni presenti nelle due fonti.

e incidenza dei contratti di lavoro a tempo determinato sono tutt'altro che trascurabili: nel complesso, essi coprono oltre il 35% del periodo di permanenza nella lista. Infine, le distribuzioni della durata di permanenza nella lista variano in misura apprezzabile in dipendenza delle diverse tipologie di percorsi lavorativi sperimentati dai lavoratori iscritti. Nel complesso, queste evidenze suggeriscono che l'acquisizione di dati più adeguati allo scopo potrebbe mettere in discussione le conclusioni degli studi di valutazione condotti sinora.

4.3.3. Alcuni studi di caso su programmi di formazione professionale

La terza area di esperienze di valutazione che consideriamo attiene agli esiti occupazionali della formazione professionale. È questa un'area dove gli esercizi di valutazione sono stati relativamente numerosi, anche perché stimolati dalle attività di documentazione e monitoraggio richieste per gli interventi cofinanziati dal Fondo Sociale Europeo (FSE). Ci limitiamo qui a tre studi di caso, quelli di Croce e Montanino (1997), di Centra, De Vincenzi e Villante (2000) e di Laudisa (1998). Comune ai tre lavori è la variabile obiettivo utilizzata per gli esercizi di valutazione: la transizione all'occupazione di soggetti in cerca di occupazione al momento in cui inizia il corso di formazione.

Croce e Montanino (1997) si propongono di valutare gli effetti di corsi di formazione professionale per adulti ammessi a programmi di politica del lavoro – liste di mobilità, Cassa integrazione guadagni e Lavori socialmente utili – realizzati da un ente gestore nel Mezzogiorno nell'anno 1995. La sola informazione di cui dispongono è la lista dei partecipanti ai corsi; il grave ostacolo con il quale si scontrano è dunque l'assenza, *tout court*, di informazioni sulle caratteristiche e gli esiti professionali degli stessi. In altre parole, si trovano costretti a (cercare di) rimediare, a cose fatte, alla mancata pianificazione della valutazione. Lo fanno progettando e conducendo un'indagine di *follow-up* a corsi largamente conclusi, quindi inevitabilmente esposta a un tasso di risposta parecchio basso, che finisce per attestarsi sotto il 24%²³. Oltre a – e ancor più che – ridurre drasticamente la numerosità del campione, ciò ne mina la rappresentatività, per verosimili fenomeni di selezione non casuale dei rispondenti, e in definitiva compromette la credibilità dei risultati. (E gli autori, correttamente, utilizzano le evidenze empiriche più per formulare ipotesi che per stabilire conclusioni.) Quel che preme sottolineare è che il tutto si sarebbe potuto svolgere diversamente, in modo ben più proficuo, se i responsabili dei corsi si fossero posti tempestivamente l'obiettivo della valutazione dell'intervento, e avessero quindi affrontato le questioni della progettazione del processo di raccolta delle informazioni utili allo scopo – prima, durante e dopo lo svolgimento dei corsi .

Lo studio di Centra, De Vincenzi e Villante (2000) verte su programmi di formazione professionale realizzati da otto Regioni del Centro-Nord nell'anno 1996, nell'ambito dell'Obiettivo 3 del FSE. Esso si presta per documentare, innanzitutto, i vantaggi di un coordinamento *ex ante*, tempestivo, nella predisposizione della base informativa: nel caso in questione, il coordinamento è dell'Isfol, incaricato appunto di sovrintendere all'azione di monitoraggio e valutazione degli interventi del FSE in Italia. L'Isfol ha successo nell'indurre un certo numero di Regioni a condurre (i) indagini sui partecipanti ai corsi e (ii) indagini di *follow-up* su un campione di partecipanti a distanza di

²³ L'indagine di *follow-up* è stata svolta nell'ottobre 1996, 15 mesi dopo la conclusione dei corsi. Il tasso di risposta all'indagine è stato del 28,2%. Tenendo conto di incompletezze nell'indirizzario dei partecipanti e di risposte non utilizzabili, la frazione di risposte utilizzate rispetto alla popolazione obiettivo è risultata del 23,1%.

un anno dalla chiusura dell'attività formativa, tra la fine del 1997 e gli inizi del 1998. Le indagini sono progettate in maniera integrata e risultano perciò ragionevolmente omogenee, con evidenti vantaggi sulla qualità – e la comparabilità – dei dati raccolti. Per un altro verso, un tratto distintivo dell'iniziativa è la limitazione delle rilevazioni a un campione dei partecipanti ai corsi. Emerge dunque, la mancata raccolta di una parte cruciale delle informazioni necessarie per la valutazione: quelle su un appropriato gruppo di confronto.

Per ovviare a questa carenza, Centra, De Vincenzi e Villante (2000) costruiscono quello che chiamano “*gruppo di confronto Istat*” traendolo dal panel a due occasioni (aprile 1996-aprile 1997) dell'indagine sulle forze di lavoro, ovviamente relativo allo stesso insieme di regioni²⁴. La procedura che adottano mira, in sostanza, a identificare un gruppo (i) paragonabile ai formati, perché costituito da persone alla ricerca di occupazione (alla prima occasione) e con analoga distribuzione per genere, classe di età e livello di istruzione, ma (ii) che comprende persone contraddistinte dal fatto non aver partecipato ad attività di formazione professionale nell'arco di tempo che va da aprile 1996 ad aprile 1997²⁵. Che tale gruppo possa fungere da *appropriato* gruppo di confronto, vi è tuttavia ragione di dubitare. Infatti, non vediamo come la comune condizione iniziale di disoccupazione e l'analoga distribuzione rispetto a poche, scarse variabili socio-demografiche siano in grado di dare conto dei molteplici fattori di eterogeneità, rilevanti per gli esiti occupazionali, fattori che rendono i non formati Istat potenzialmente diversi in modo sistematico dai formati.

La Figura 2 esemplifica le riserve circa il modo di costruzione del gruppo di confronto. Essa mette in luce le caratteristiche comuni e l'eterogeneità non osservata nelle ipotetiche, schematiche storie di vita di due soggetti, l'uno tratto dal gruppo dei formati ($I=1$) e l'altro tratto dal gruppo di confronto Istat ($I=0$), *nel migliore dei casi possibili*: nel caso, cioè, di due persone dello stesso genere, della stessa età e dello stesso livello di istruzione. Alla domanda cruciale: “È stato identificato l'insieme delle caratteristiche rispetto alle quali i non partecipanti ai corsi risultano diversi in modo sistematico dai partecipanti?”, la Figura 2 suggerisce, in maniera stilizzata ma nitida, una risposta negativa. La ragione sta nel fatto che, nel processo di costruzione del gruppo di confronto Istat, viene completamente trascurata la storia lavorativa relativa al periodo di tempo che va dall'ingresso nel mercato del lavoro ad aprile '97 (rappresentata nella Figura 2 dalla linea tratteggiata): storia che per individui dello stesso genere, età – si badi: fino a 45 anni! – e grado di istruzione può essere parecchio differente. Ora, la storia lavorativa ha rilievo per gli esiti occupazionali sotto un duplice profilo: di per sé, per il processo di formazione di capitale umano che incorpora; e inoltre come *proxy* di variabili non osservabili (capacità non meramente cognitive, motivazioni, ecc.). La possibilità di aver eliminato – tramite il

²⁴ In realtà, per ovviare alla non piena sincronia con l'indagine di *follow-up* e al fatto che la stessa si estende su un certo arco di mesi – gli ultimi del 1997 e i primi del 1998 –, gli autori utilizzano la media di tre panel a due occasioni distanziate di un anno: rispettivamente aprile 1995-96, aprile 1996-97 e aprile 1997-98 (Centra, De Vincenzi e Villante, 2000, pp. 128-129). Per semplicità nel testo principale trascuriamo questa complicazione, irrilevante per le argomentazioni svolte.

²⁵ Sommarariamente, la procedura di costruzione del gruppo di confronto è la seguente: (i) al primo passo, viene estratto un sub-campione costituito dalle persone in età 14-45 anni in cerca di occupazione (alla prima occasione), che abbiano dichiarato di non avere seguito alcuna attività di formazione professionale nel mese precedente entrambe le occasioni di indagine; (ii) al secondo passo, viene ulteriormente selezionato un sotto-gruppo, che costituisce appunto il “*gruppo di confronto Istat*”, in modo che riproduca le distribuzioni di quello dei formati in termini di genere, classe di età e livello di istruzione.

gruppo di confronto Istat – il *selection bias* non è, in definitiva, credibile, e mina la validità delle stime dell’impatto prodotte dagli autori.

Figura 2 circa qui

All’opposto dei due studi di caso appena visti, Laudisa (1998) documenta una situazione fortunata nella quale, nonostante l’assenza di progettazione della valutazione dell’impatto, la rilevazione degli esiti occupazionali anche sui non esposti all’intervento, svolta *ex post*, ha consentito di ottenere risultati convincenti in merito all’impatto di un programma di formazione professionale. Si tratta di corsi di informatica gestionale organizzati dal Comune di Torino, riservati essenzialmente a disoccupati, che hanno avuto luogo tra ottobre 1995 e giugno 1996. Va notato esplicitamente come il buon esito di tale valutazione sia da attribuire in larga misura ad eventi fortuiti. Questo caso non rappresenta, quindi, una smentita della tesi qui sostenuta circa la necessità di predisporre le condizioni per poter disporre del supporto informativo adeguato per la valutazione d’impatto.

Gli eventi fortuiti ai quali ci riferiamo sono di due ordini. Innanzitutto, a causa della limitata disponibilità di posti, ai corsi in questione sono stati ammessi solo alcuni tra coloro che ne avevano fatto domanda, selezionati sulla base del punteggio ottenuto ad un test attitudinale. Come conseguenza immediata del razionamento, è stato possibile utilizzare come gruppo di confronto i soggetti esclusi dal corso. Lo stimatore che ne risulta, con le cautele di cui diremo nella sez. 5.2, consente di risolvere i problemi di *selection bias* in modo particolarmente semplice. Il punto è che nel contesto istituzionale italiano casi come questo rappresentano l’eccezione. Di norma non vi sono impedimenti all’ammissione ai corsi di tutti coloro che ne fanno richiesta. Pertanto, il gruppo di confronto deve essere selezionato nell’insieme molto più vasto costituito da coloro che non hanno preso parte al corso né ne hanno fatto richiesta, in modo tale da rendere possibile il ricorso all’uno o all’altro degli stimatori presentati nella sezione 2.

In secondo luogo, l’indagine di *follow up*, condotta nel mese di novembre 1997 su tutti i soggetti che hanno chiesto l’ammissione al corso – ammessi ed esclusi –, ha goduto di una collaborazione degli intervistati elevata, con un tasso di risposta pari a circa il 90%. È difficile immaginare che tale risultato sia replicabile in assenza di una adeguata organizzazione dell’indagine.

Questi due elementi – presenza di aspiranti esclusi (a seguito della formazione di una graduatoria per l’ammissione) e informazioni di *follow-up* sia sugli ammessi ai corsi che sugli esclusi – saranno per l’appunto utilizzati per illustrare, nella sez. 5.2, una robusta procedura di stima dell’impatto dei corsi.

5. Confrontare soggetti confrontabili

5.1. Stima dell’impatto mediante modelli parametrici di regressione vs. metodi di abbinamento

Veniamo ora all’esigenza, cruciale, di costruire in maniera appropriata il gruppo di confronto, in modo che sia costituito da soggetti confrontabili con i soggetti esposti all’intervento. A questo scopo, è istruttivo contrapporre stimatori dell’impatto basati su

modelli parametrici di regressione con stimatori non parametrici basati su metodi di abbinamento.

Ad analisi del processo di selezione compiuta e a lista delle variabili X responsabili del *selection bias* compilata, molto spesso si osserva il ricorso al seguente stimatore dell'impatto (vedi Bjorklund e Moffit, 1987). Specificate le regressioni di Y^T e Y^{NT} su X :

$$Y^T = \alpha^T + \beta^T X + \varepsilon^T \quad (14.1)$$

$$Y^{NT} = \alpha^{NT} + \beta^{NT} X + \varepsilon^{NT}, \quad (14.2)$$

utilizzando informazioni relative, rispettivamente, ai soggetti esposti e non esposti si ottiene una stima delle due medie condizionate $E\{Y^T|I=1, X\}$ e $E\{Y^{NT}|I=0, X\}$. Ma, per il modo in cui sono state scelte le variabili incluse in X , vale l'eguaglianza (9). Conseguentemente, la quantità:

$$\hat{\alpha}^T - \hat{\alpha}^{NT} + (\hat{\beta}^T - \hat{\beta}^{NT}) X \quad (15)$$

è una stima dell'impatto medio sui soggetti con caratteristiche X . La media della (15) calcolata rispetto alla distribuzione di X nell'insieme degli esposti fornisce quella che *prima facie* appare essere una stima dell'impatto medio sugli esposti.

Tale modo di procedere espone al rischio di scambiare per stime dell'impatto medio sugli esposti *basate sull'evidenza empirica* a disposizione – i risultati ottenuti da esposti e non esposti – stime che tali non sono, perché ottenute confrontando il risultato ottenuto dagli esposti con il risultato ottenuto da non esposti *diversi* dagli esposti rispetto alle caratteristiche X . Si consideri il seguente caso limite. Sia X_1 il supporto di $X|I=1$ e X_0 il supporto di $X|I=0$, vale a dire l'insieme dei valori di X osservati, rispettivamente, sugli esposti e sui non esposti. Poniamoci nella situazione in cui i due supporti sono disgiunti ($X_1 \cap X_0 = \emptyset$). In corrispondenza di $X \in X_1$ lo stimatore (15) ottiene una stima dell'impatto *estrapolando* il valore di $E(Y^{NT}|I=0, X)$ dalle regolarità osservate in X_0 , *non* confrontando i risultati ottenuti da soggetti esposti e da soggetti non esposti in corrispondenza a *quel* particolare valore di X , confronto questo ultimo irrealizzabile data l'assenza di elementi in comune tra i due insiemi X_1 e X_0 . Analogo problema sorge se $X \in X_0$.

Palesamente, si tratta di stime dell'impatto che poggiano, oltre che sull'evidenza empirica, sull'assunto che al di fuori dei rispettivi supporti X_1 e X_0 le due regressioni mantengano la stessa forma funzionale osservata in corrispondenza dei valori di X appartenenti ai due supporti. In questo senso, lo stimatore risulta fragile.

Al contrario lo stimatore (13) – il quale, si noti, non formula alcuna assunzione parametrica sulla regressione dei risultati potenziali su X – *obbliga* al confronto di soggetti effettivamente confrontabili rispetto a X . Nel caso limite appena esposto, il ricorso allo stimatore (13) obbligherebbe l'analista a rendersi conto che l'impatto dell'intervento è *intrinsecamente* non identificabile (se non ricorrendo agli assunti che giustificano l'estrapolazione). La Figura 3a presenta ipotetici risultati per una variabile obiettivo Y nel caso di supporti disgiunti per esposti e non esposti, e illustra chiaramente il punto.

Le situazioni che di norma si presentano sono meno estreme di quella illustrata appena sopra. Capita che X_1 e X_0 , pur non coincidenti, abbiano intersezione non vuota. Detta $X_{1 \cap 0}$ la loro intersezione, l'abbinamento tra soggetti esposti e soggetti non esposti potrà avere luogo *solo* per i soggetti che presentano valori di X appartenenti a $X_{1 \cap 0}$. Pertanto, lo stimatore (13) darà luogo ad una stima di $E(Y^T - Y^{NT}|I=1, X \in X_{1 \cap 0})$. Anche in questa situazione il ricorso allo stimatore (13) obbliga l'analista a toccare con mano che l'impatto medio effettivamente identificabile si riferisce ad una *particolare*

sottopopolazione, non all'intero insieme degli esposti. Analogamente al caso precedente, per un'esemplificazione torna utile guardare alla Figura 3b, la quale presenta ipotetici risultati per una variabile obiettivo Y nel caso di supporti con intersezione non vuota per esposti e non esposti. La distinzione tra l'impatto intrinsecamente identificabile, quello per la sottopopolazione con valori di X appartenenti a $X_{1\cap 0}$, e l'impatto per la cui identificazione è invece necessaria un'assunzione parametrica sulla regressione – e la sua estrapolazione –, quello per le sottopopolazioni con risultati potenziali su $X \notin X_{1\cap 0}$, emerge in maniera nitida.

La questione qui si intreccia con i problemi posti all'analista dalla variabilità dell'impatto, dei quali diremo più diffusamente nella sez. 6.1. Se vi fossero ragioni per ritenere che la variabilità dell'impatto tra soggetti è trascurabile, l'identificabilità dell'impatto medio sugli esposti appartenenti a $X_{1\cap 0}$ risolverebbe – è evidente – il problema dell'identificabilità dell'impatto medio *tout court*. Dato che solitamente così non è, l'impatto medio sui soggetti non appartenenti a $X_{1\cap 0}$ risulta identificabile solo grazie all'estrapolazione.

Figura 3 circa qui

Un caso particolare, che da un lato agevola di molto la realizzazione di un confronto fra soggetti confrontabili ma dall'altro circoscrive all'estremo la sottopopolazione per la quale risulta identificabile l'impatto medio, si ha quando, per costruzione, vi è un'unica variabile, *osservata* dal valutatore, che influenza la probabilità di (al limite, determina univocamente la) esposizione all'intervento. Ciò si dà, ad esempio, quando siano utilizzate graduatorie e i beneficiari siano scelti come coloro che si situano al di sopra (o al di sotto) di una certa soglia, e si conduca il confronto restringendosi ai soggetti con punteggi nell'intorno della soglia. In questi, e analoghi, casi di *regression discontinuity design* (Rubin, 1977; Trochim, 1984), per stimare l'impatto medio dell'intervento si può utilizzare uno stimatore prossimo a (13), sfruttando per l'appunto la discontinuità intorno alla soglia²⁶. Si noti come sia il particolare disegno del processo di selezione a definire in modo univoco quale sia la variabile rispetto alla quale esposti e non esposti differiscono e, contemporaneamente, come per effetto dello stesso disegno l'impatto medio risulti identificabile solo per i soggetti prossimi alla soglia di ammissione al/esclusione dal programma.

5.2. Valutazione di impatto di un programma di formazione professionale con ammissione in base a graduatoria

Nel panorama della letteratura italiana, un interessante esempio di utilizzazione innovativa di questo approccio è nello studio di caso di Laudisa (1998), brevemente presentato alla fine della sez. 4.3.3.

Come già anticipato, si tratta di corsi di informatica gestionale organizzati dal Comune di Torino nell'anno formativo 1995/96. Data la limitata disponibilità di posti, ai

²⁶ Hahn, Todd e Van der Klaauw (2001) hanno messo in luce che lo stimatore dell'impatto ottenuto in questo modo può essere interpretato come uno stimatore delle variabili strumentali pur essendo le sue proprietà asintotiche diverse da quelle dello stimatore delle variabili strumentali convenzionale.

corsi sono stati ammessi solo parte di coloro che ne avevano fatto richiesta. La selezione ha avuto luogo sulla base del punteggio ottenuto ad un test attitudinale, mirante ad accertare abilità logico-matematiche, somministrato ai richiedenti. Sono stati ammessi i soggetti meglio classificati nella graduatoria, in numero pari ai posti disponibili.

Nel novembre del 1997, a circa sedici mesi dalla conclusione dei corsi, *tutti* i soggetti che avevano partecipato al test di ammissione sono stati intervistati telefonicamente, ed è stata chiesta loro, tra le altre cose, la condizione occupazionale. La valutazione dell'intervento è stata svolta con riferimento alla variabile binaria occupato/in cerca di occupazione al momento dell'intervista.

Detto S il punteggio riportato al test di ammissione, è evidente che S (i) è una variabile che verosimilmente causa *selection bias* e, inoltre, (ii) è l'*unica* variabile che può causare *selection bias*.

Che S verosimilmente causi *selection bias* risulta chiaro dal fatto che vale la condizione (8.1) ed è plausibile che valga anche la condizione (8.2). Per costruzione, infatti, i soggetti esposti presentano sistematicamente valori di S più elevati dei soggetti non esposti (condizione (8.1)). D'altra parte, è ragionevole attendersi che la maggiore abilità degli esposti diagnosticata al test sia osservabile, e sia apprezzata, dai potenziali datori di lavoro, tanto da rendere gli esposti maggiormente 'occupabili' dei non esposti anche in assenza dei corsi (condizione (8.2)).

In secondo luogo, S è *per costruzione* l'unica variabile che può causare *selection bias*, dato che non vi è alcuna altra caratteristica dei richiedenti rilevante per la decisione di ammissione/esclusione.

Siamo pertanto nella situazione rappresentata dall'equazione (13), nella quale uno stimatore di tipo *matching* realizzato con riferimento alla variabile S risolve il problema di *selection bias*. Con l'ovvia limitazione, di cui abbiamo detto nella sezione 5.1 immediatamente precedente, che per costruzione l'abbinamento risulta possibile solo per gli esposti i quali presentano valori di S prossimi al valore soglia per l'ammissione: solo per tali soggetti, infatti, è possibile trovare soggetti non esposti somiglianti rispetto ad S .

La Figura 4a riassume i risultati dell'analisi. Vi è ben visibile il valore soglia di S che separa gli esposti dai non esposti, circa pari a 56. E altrettanto visibile è la marcata differenza tra esposti e non esposti nella probabilità di essere occupati a circa sedici mesi dalla conclusione dei corsi. Per quanto appena detto, è verosimile che almeno parte di tale differenza sia dovuta al *selection bias*.

Nella Figura 4a sono pure riportate le regressioni della variabile risultato calcolate separatamente per il gruppo degli esposti – la curva a destra del valore soglia – e per il gruppo dei non esposti – la curva a sinistra di tale valore. Essendo la variabile risultato binaria, le due regressioni forniscono la probabilità di essere occupato al variare di S separatamente per i due gruppi: $Pr(occupato|S, I)$. In linea con le aspettative, si nota una dipendenza molto forte tra la variabile risultato ed S : in entrambi i gruppi, al crescere del punteggio riportato al test cresce la probabilità di essere occupato al momento dell'intervista. Si noti come questa evidenza (associata alla dipendenza deterministica tra S ed I dovuta al disegno) provi l'esistenza di *selection bias*, fin qui solo congetturata.

L'impatto stimato è dato dalla differenza tra la probabilità di essere occupato calcolata per i soggetti esposti che presentano un valore di S pari a 56 e la probabilità di essere occupato per i soggetti non esposti che presentano lo stesso valore di S : $Pr(occupato|S=56, I=1) - Pr(occupato|S=56, I=0)$. Operando il confronto tra esposti e non esposti in questo modo, infatti, si eliminano tutte le differenze tra i due gruppi dovute al processo di selezione: pertanto, se tra i due gruppi si riscontrano differenze nella

probabilità di occupazione, queste non possono che essere dovute all'intervento. Nel caso in questione, la differenza è piccola (vedi ancora la Figura 4a), e risulta statisticamente non significativa. Si conclude, dunque, per l'assenza di impatto dell'intervento sulla probabilità di occupazione *limitatamente* alla sottopopolazione costituita dai soggetti al margine tra ammissione e non ammissione²⁷.

L'evidenza fornita dai risultati ottenuti da esposti e non esposti consente di concludere sull'impatto dell'intervento *solo* per una particolare sottopopolazione, i soggetti con punteggio al test prossimo a 56. Valutazioni circa l'impatto dell'intervento su soggetti diversi da questi possono essere ottenute soltanto estrapolando le due probabilità di occupazione al variare di S , stimate rispettivamente sugli esposti e sui non esposti nel modo rappresentato nella Figura 4b. Ad esempio, la valutazione dell'impatto del programma sui soggetti non esposti richiede il calcolo di $Pr(occupato|S, I=1)$ in corrispondenza di valori di S inferiori a 56. Ma in corrispondenza di tali valori di S , *per costruzione* non si osservano soggetti esposti. Pertanto, si può al più congetturare che se la curva $Pr(occupato|S, I=1)$ mantenesse per valori di S inferiori a 56 le stesse caratteristiche osservate per valori di S superiori a 56, allora l'impatto sui non esposti sarebbe pari alla differenza tra la curva stimata sui soggetti non esposti e la curva ottenuta estrapolando quella stimata sui soggetti esposti.

Nel caso in questione, si noti che la congettura basata sull'estrapolazione delle regolarità osservate rappresentata nella Figura 4b suggerisce che sono stati ammessi al programma i soggetti che non ne ricavano alcun beneficio, mentre sono stati esclusi proprio quelli che ne avrebbero ricavato un beneficio: gli aspiranti meno abili, con punteggio al test intorno a 5.

Figura 4 circa qui

6. Non ignorare la variabilità dell'impatto tra i beneficiari

6.1. Questioni di identificazione e di interpretazione

Se l'impatto è fortemente variabile da soggetto a soggetto, la semplice stima dell'impatto medio dà scarso conto degli effetti dell'intervento. Può essere fuorviante se induce a sopprimere un intervento i cui esiti sono mediamente nulli (o addirittura negativi), ma che risulta utile ad una frazione non trascurabile di soggetti. In tale situazione, l'intervento andrebbe indirizzato meglio, non soppresso. Può risultare scarsamente informativa se si riferisce a una piccola sottopopolazione, o comunque a una sottopopolazione della cui rappresentatività vi è ragione di dubitare.

Caratteristiche della distribuzione dell'impatto di potenziale interesse includono tipicamente:

²⁷ Lo studio di caso è ripreso in Battistin e Rettore (2002), che discutono i problemi posti dalla presenza di un certo numero di soggetti che non si conformano all'esito della selezione – il 18% degli esclusi ha avuto accesso a un altro programma di formazione – e propongono una procedura non parametrica per verificare la presenza di impatto. Le conclusioni cui giungono sono simili a quelle di Laudisa (1998).

- (a) la proporzione di esposti per i quali l'intervento è positivo: $Pr(\alpha_i > 0 | I_i = 1)$ ²⁸;
- (b) selezionati quantili della distribuzione dell'impatto;
- (c) la distribuzione dell'impatto per selezionati valori, diciamoli y_0 , nella condizione di base, corrispondente all'assenza di intervento: $F_\alpha(\alpha_i | I = 1, Y_i^{NT} = y_0)$;
- (d) l'incremento nella proporzione di soggetti che presentano risultati oltre una certa soglia, diciamola \bar{y} , dovuto all'intervento: $Pr(Y_i^T > \bar{y} | I = 1) - Pr(Y_i^{NT} > \bar{y} | I = 1)$.

Gli importanti interrogativi in tema di valutazione ai quali queste misure danno risposte sono evidenti. È forse il caso di aggiungere che la misura (c) è appropriata se si ha uno specifico interesse nel valutare l'impatto dell'intervento per beneficiari che si collocano, ad esempio, nella coda bassa della distribuzione della condizione di base²⁹. D'altra parte, la misura (d) fornisce gli elementi per rispondere ad un interrogativo del tipo: le distribuzioni dei risultati per gli esposti ad un intervento dominano (e, variando \bar{y} , in che misura) la distribuzione dei risultati che essi avrebbero conseguito se non vi avessero preso parte?

Sul fronte dell'identificazione dell'eterogeneità dell'impatto la ricerca è ancora nella sua fase pionieristica (vedi Heckman, Smith e Clements, 1997 e Imbens e Rubin, 1997). Per semplicità espositiva, presentiamo la questione utilizzando come misura dell'eterogeneità dell'impatto la sua varianza. Supponiamo che I sia determinato mediante assegnazione casuale. La (6) mostra che in tale caso è identificabile la differenza tra le varianze dei due risultati potenziali. Essendo:

$$Y^T = Y^{NT} + \alpha,$$

la differenza tra le due varianze risulta pari a:

$$var(Y^T) - var(Y^{NT}) = var(\alpha) + 2 cov(Y^{NT}, \alpha). \quad (16)$$

Dato che in generale la covarianza tra il risultato potenziale da non esposti e l'impatto è diversa da zero, $var(\alpha)$ risulta non identificabile.

Si noti esplicitamente come l'assegnazione casuale di I non consenta di risolvere il problema: l'assegnazione casuale assicura l'ortogonalità tra (Y^T, Y^{NT}) e I , non tra Y^{NT} e α . Heckman, Smith e Clements (1997) studiano il problema di identificabilità di $var(\alpha)$, e più in generale della distribuzione dell'impatto, in presenza di vari tipi di restrizioni sul processo generatore dei risultati potenziali e della variabile di trattamento.

Anche rinunciando all'identificazione dell'eterogeneità dell'impatto, il semplice fatto che vi sia ragione per ritenere l'impatto dell'intervento eterogeneo pone problemi tutt'altro che trascurabili per l'interpretazione dei risultati della stima dell'impatto medio. La questione si pone nei termini in parte già anticipati nella sezione 5. Stimatori alternativi dell'impatto medio in generale producono stime che si riferiscono a sottopopolazioni diverse. In assenza di eterogeneità dell'impatto, non sorge alcun problema: essendo definito in modo univoco il parametro da stimare, le stime risultano diverse solo per effetto della variabilità campionaria. Al contrario, se l'impatto è eterogeneo, le stime possono differire perché sono diversi gli impatti medi cui si riferiscono (oltre che, beninteso, per effetto della variabilità campionaria).

²⁸ Per semplicità, e senza perdita di generalità, assumiamo che Y , la variabile risultato, sia tale che l'intervento miri a modificarla in senso crescente. Tipicamente, ciò accade quando Y è il reddito oppure è la variabile binaria uguale a 1 (≡occupazione) o 0 (≡in cerca di occupazione).

²⁹ In generale, essa consente di analizzare come la distribuzione dell'impatto dipenda dalla condizione di base degli esposti.

A titolo di esempio, si consideri il caso di un intervento per il quale l'autorità responsabile ha previsto l'assegnazione casuale ai due gruppi degli esposti e dei non esposti. Sia A la variabile binaria che identifica i due gruppi come risultano dall'assegnazione casuale. Supponiamo che ad assegnazione avvenuta vi sia non perfetta conformità dei soggetti all'assegnazione, cioè a dire che parte dei soggetti destinati all'esposizione si sottragga all'intervento e che parte dei soggetti destinati al gruppo di controllo trovi invece il modo di essere esposta all'intervento. Sia I la variabile binaria che identifica l'effettivo stato dei soggetti rispetto all'intervento. Palesemente, ricorrendo a A è possibile ottenere una stima delle *variabili strumentali* dell'impatto. Tuttavia, a differenza del caso da libro di testo di cui si è detto nella sezione 2, Imbens e Angrist (1994) hanno mostrato che – se è soddisfatta una condizione aggiuntiva, plausibile in molte circostanze ma non scontata, detta di *monotonicità* – l'impatto medio stimato in questo modo si riferisce alla particolare sottopopolazione costituita dai soggetti che obbediscono all'esito dell'assegnazione casuale, i cosiddetti *compliers*³⁰. Si tratta dei soggetti che accetterebbero l'esito dell'assegnazione casuale sia se li destinasse al coinvolgimento nel programma sia li destinasse all'esclusione dal programma³¹.

Si noti che, in generale, i *compliers* sono un sottoinsieme *non rappresentativo* dell'intera popolazione. Verosimilmente, coloro che vanno contro l'esito dell'assegnazione casuale hanno qualche ragione speciale per farlo, la quale li rende diversi da coloro che si conformano a tale esito.

A complicare le cose, l'insieme dei *compliers* è *non osservabile*, per cui non si sanno nemmeno caratterizzare con precisione i soggetti cui si riferisce l'impatto medio identificabile. Infatti, risulta osservabile l'insieme costituito dai soggetti che non contraddicono l'esito dell'assegnazione casuale *effettivamente realizzatasi* – sono coloro per i quali vale l'eguaglianza $I=A$. Ma, tra questi, non si sa dire quali soggetti avrebbero obbedito anche se l'assegnazione casuale avesse dato luogo all'altro esito possibile.

È invece osservabile la *dimensione* dell'insieme dei *compliers*. Sempre se vale la condizione di monotonicità, essa risulta pari a $Pr(I=1|A=1) - Pr(I=1|A=0)$. Pertanto, si sa almeno dire se si tratta di una sottopopolazione sufficientemente grande da rendere in qualche modo interessante l'impatto medio che li riguarda.

Se l'impatto dello *stesso* intervento viene stimato trascurando A , analizzando il processo che determina I – cioè a dire, che dà luogo ai due gruppi degli esposti e dei non esposti – rilevando tutte le variabili rilevanti per il processo di selezione e infine ricorrendo allo stimatore (13), come si è già detto la stima ottenuta si riferisce all'impatto medio $E(Y^T - Y^{NT}|I=1, X \in X_{1 \cap 0})$. Dato che le due sottopopolazioni costituite rispettivamente dai *compliers* e dai soggetti appartenenti a $X_{1 \cap 0}$ sono in generale diverse, se l'impatto varia da soggetto a soggetto non c'è ragione per attendersi che i due impatti medi risultino eguali.

Il mancato riconoscimento del fatto che, come conseguenza dell'eterogeneità dell'impatto, è del tutto ovvio che stimatori diversi diano luogo a risultati diversi, può indurre scetticismo sull'utilità pratica della valutazione d'impatto – se stimatori diversi forniscono risultati diversi, a quale credere? Uno scetticismo, è il caso di ribadirlo, del tutto ingiustificato. La diversità dei risultati forniti da stimatori alternativi rappresenta una

³⁰ La condizione di monotonicità consiste essenzialmente nell'assenza dei cosiddetti *defiers*, cioè a dire di soggetti che andrebbero comunque contro l'esito dell'assegnazione casuale, vuoi se li destinasse all'esposizione al programma vuoi se ne indicasse l'esclusione dallo stesso.

³¹ In effetti, Imbens e Rubin (1997) hanno mostrato che, in tale situazione, per il gruppo dei *compliers* sono identificabili le distribuzioni di entrambi i risultati potenziali (ma non la loro distribuzione congiunta!).

ricchezza per l'analisi, non un suo difetto, nel senso che aiuta a comprendere quali sono i soggetti che traggono giovamento dall'intervento e quali no.

6.2. *Un altro studio di caso: i programmi regionali di formazione professionale cofinanziati dal FSE rivisitati*

Per illustrare i problemi che l'eterogeneità dell'impatto pone per l'interpretazione degli impatti medi, è istruttivo considerare un recente studio di Berliri e Pappalardo (2001). I due autori riaffrontano il problema della valutazione dell'impatto dei programmi regionali di formazione professionale cofinanziati dal FSE, di cui abbiamo discusso nella sez. 4.3.3. Il lavoro si colloca a valle di quello di Centra, De Vincenzi e Villante (2000) e ne costituisce, idealmente, un affinamento, nel senso che si propone di tenere esplicitamente conto della possibile presenza di variabili non osservabili nel processo di selezione. L'analisi empirica è condotta sui partecipanti ai corsi di due regioni – Lombardia ed Emilia Romagna – nell'anno 1997, e condivide in larga parte la logica della costruzione della base di dati di Centra, De Vincenzi e Villante (2000): col sub-campione dei formati affiancato da un sub-campione di (potenziali partecipanti) non formati tratto dal panel aprile 1997-aprile 1998 dell'indagine sulle forze di lavoro nelle due regioni in questione³².

Ridotto all'essenziale, il modello di Berliri e Pappalardo (2001) consiste nell'affiancare alla coppia di equazioni (14) un'ulteriore equazione che rappresenta la decisione di prendere/non prendere parte al corso di formazione:

$$I^* = \gamma_0 + \gamma_1 X + \gamma_2 Z + u, \quad (17)$$

dove I^* è una variabile continua *non osservabile*, cui corrisponde la variabile binaria *osservabile* I pari a 1 se I^* è positiva e pari a 0 altrimenti. In questa formulazione, l'esistenza di caratteristiche non osservabili dei soggetti, che causano *selection bias*, si traduce in una correlazione diversa da zero tra i disturbi (ε^T , ε^{NT}) nelle equazioni (14) e il disturbo u nell'equazione (17). A causa di tale correlazione l'applicazione del metodo dei minimi quadrati alle equazioni (14) dà luogo a stime inconsistenti dei parametri (vedi Heckman, 1979), e risulta quindi inconsistente anche la stima (15) dell'impatto medio sui soggetti con caratteristiche X .

Al contrario, la stima congiunta delle equazioni (14)-(17) tenendo esplicitamente conto della correlazione tra (ε^T , ε^{NT}) e u consente di calcolare in modo appropriato l'impatto medio sui soggetti con caratteristiche X . L'espressione che ne risulta è identica alla (15), salvo che le stime dei parametri che vi compaiono sono ottenute tramite stimatori consistenti, di sistema, di (14)-(17), e non applicando i minimi quadrati separatamente alle due regressioni (14).

In presenza di eterogeneità dell'impatto, questo modo di procedere dà luogo a due ordini di problemi, entrambi in qualche modo nascosti dalla specificazione parametrica del

³² Il sub-campione dei non formati, tratto dal panel di cui abbiamo appena detto, è costituito semplicemente dalle persone in età 14-45 anni in cerca di occupazione alla prima occasione, che non abbiano seguito alcuna attività di formazione professionale nel mese precedente entrambe le occasioni di indagine. Berliri e Pappalardo (2001) evitano dunque il secondo, discutibile passo della procedura di Centra, De Vincenzi e Villante (2000), mirato a costruire deterministicamente il "gruppo di confronto Istat", facendo riferimento a poche, scarse variabili socio-demografiche.

modello (14)-(17). Per semplicità, ragioniamo con riferimento all'impatto medio sui soggetti con caratteristiche X ³³.

- (a) Innanzitutto, vale quanto detto nella sez. 5.1: l'impatto medio sui soggetti con caratteristiche X è identificabile limitatamente ai soggetti che appartengono all'insieme $X_{1 \cap 0}$, non per l'intera popolazione. Nella presentazione dei risultati è pertanto della massima importanza caratterizzare, con riferimento alle caratteristiche X , la sottopopolazione cui si riferisce l'impatto ottenuto.
- (b) In secondo luogo, il modello (14)-(17) risulta identificabile *solo* in virtù della *restrizione di esclusione* della variabile Z dalle equazioni (14). Tale restrizione rende la procedura di stima adottata equivalente sul piano logico – anche se non sul piano operativo – ad una stima delle variabili strumentali: Z funge da variabile strumentale per la variabile esplicativa endogena I . L'analisi svolta da Imbens e Angrist (1994), di cui abbiamo detto nella sez. 6.1, chiarisce che (se è soddisfatta la condizione di monotonicità, cioè nella migliore delle ipotesi) uno stimatore delle variabili strumentali identifica l'impatto medio sui *compliers*.

A quanto si arguisce dal lavoro in questione, la restrizione di esclusione fondamentale consiste nel porre la variabile Z pari alla variabile binaria 'il soggetto ha/non ha interrotto un ciclo di studi in corso'³⁴. Pertanto, l'insieme dei *compliers* cui si riferisce l'impatto medio stimato è costituito solo dai soggetti la cui decisione di partecipazione al corso dipende dal fatto di avere/non avere interrotto un ciclo di studi in corso³⁵. Semplici calcoli basati sui risultati riportati nel lavoro, relativi all'equazione di selezione (17), consentono di concludere che i *compliers* rappresentano al più il 15% della popolazione³⁶.

In conclusione, se l'impatto dei corsi di formazione professionale è eterogeneo – com'è del tutto ragionevole attendersi –, l'impatto medio ottenuto da Berliri e Pappalardo (2001) *non* si riferisce all'intera popolazione ma ad un suo particolare sottoinsieme,

³³ Berliri e Pappalardo (2001) ottengono anche stime dell'impatto medio rispettivamente sugli esposti e sui non esposti. Tali impatti medi, peraltro, risultano identificabili solo grazie agli assunti parametrici formulati per il modello (14)-(17), in particolare alla specificazione *single index* dell'equazione di selezione e alla multinormalità dei disturbi. (La sola restrizione di esclusione della variabile Z dalle equazioni (14), infatti, non garantisce l'identificabilità dei due impatti medi in questione.) Ora, tali assunti sono del tutto arbitrari, sicché l'identificazione degli impatti medi per esposti e non esposti finisce per risultare vistosamente fragile. La restrizione di esclusione di Z dalle equazioni (14) è, invece, sufficiente per l'identificabilità dell'impatto medio sui soggetti con caratteristiche X . In questo senso, svolgiamo i nostri commenti con riguardo ai risultati più robusti tra quelli ottenuti dai due autori.

³⁴ In Berliri e Pappalardo (2001, pp. 7-8 e 21) si dice che sono state usate come variabili strumentali anche la posizione occupazionale dei genitori e il tasso di disoccupazione provinciale nell'anno di partecipazione al corso di formazione professionale. Che queste due variabili possano fungere da strumenti, è tuttavia dubbio. Quanto alla prima variabile, non è chiaro come sia stata ottenuta per i soggetti che non vivono più con la famiglia di origine (dal testo si desume, infatti, che il questionario dell'indagine di *follow-up* rileva la condizione occupazionale soltanto per i componenti della famiglia del formato). Quanto al tasso di disoccupazione provinciale nell'anno di partecipazione al corso, è verosimilmente molto correlato con i tassi di disoccupazione provinciale – distinti per genere – nell'anno cui si riferisce la variabile risultato (si tratta di due anni adiacenti, rispettivamente il 1997 e il 1998), variabili queste ultime che compaiono nelle equazioni (14).

³⁵ Si noti che, nella discussione che stiamo svolgendo, diamo per scontata l'accettabilità sia della restrizione di esclusione che della proprietà di monotonicità.

³⁶ In ipotesi di monotonicità, tale frazione è data dalla differenza $Pr(I=1|X,Z=1) - Pr(I=1|X,Z=0)$, dove $Z=0/1$ se il soggetto ha/non ha interrotto gli studi durante la scuola media superiore (la variabile 'interruzione di studi universitari', inclusa nell'equazione, non risulta significativa). Tale differenza palesemente dipende da X , ma utilizzando i risultati riportati nel lavoro è immediato mostrare che al massimo può valere 0,15.

costituito dai soggetti appartenenti all'insieme $X_{1\cap 0}$ e che si conformano all'assegnazione determinata dalla variabile strumentale Z . La mancata caratterizzazione dell'insieme $X_{1\cap 0}$ da un lato, l'oggettiva difficoltà ad identificare l'insieme dei *compliers* dall'altro e, infine, la ridotta dimensione del sottogruppo cui l'impatto medio si riferisce circoscrivono assai l'interesse del parametro stimato.

7. Un commento conclusivo

In questo articolo abbiamo trattato una serie di temi connessi alle difficoltà indotte dalla presenza di *selection bias* nella valutazione dell'impatto di interventi sociali, e di politiche attive del lavoro in particolare. Vi sono, peraltro, alcuni punti generali sufficientemente importanti – crediamo – da meritare di essere evidenziati in sede di notazioni conclusive.

All'origine del problema di valutazione dell'impatto di un intervento sta la difficoltà di discriminare le variazioni prodotte dall'intervento dalle variazioni dovute alle reazioni degli agenti – segnatamente al processo di (auto)selezione dei soggetti nei due gruppi rispettivamente degli esposti e dei non esposti. La soluzione al problema dell'identificabilità dell'impatto poggia su una attenta analisi del fenomeno – le condizioni o comportamenti dei membri della popolazione di interesse che l'intervento si propone di modificare – e delle interazioni tra fenomeno e intervento, volta ad individuare ed eliminare le minacce alla cosiddetta 'validità interna' della valutazione (detto altrimenti, le fonti di mutamenti che potrebbero essere scambiati per effetti dell'intervento e tali non sono), riassunte nel termine *selection bias*.

Le argomentazioni che siamo venuti svolgendo sul terreno analitico, e le evidenze in merito ad alcuni studi di caso che abbiamo presentato e commentato, suffragano in maniera convincente le tre affermazioni-chiave da cui abbiamo preso le mosse: (i) l'importanza di disporre di dati adeguati, e quindi di pianificare tempestivamente – in sede di disegno stesso dell'intervento – la rilevazione mirata delle informazioni richieste per la valutazione del suo impatto; (ii) l'esigenza di effettuare confronti appropriati fra soggetti esposti e soggetti non esposti all'intervento, cioè a dire di disporre di un gruppo di confronto costituito da soggetti per davvero confrontabili con gli esposti, evitando di ricorrere – per identificazione dell'impatto – ad assunzioni funzionali o parametriche sovente arbitrarie; (iii) la consapevolezza della eterogeneità dell'impatto degli interventi sociali, con le implicazioni che ne discendono sia su caratteristiche di potenziale interesse della distribuzione dell'impatto sia sui problemi, tutt'altro che trascurabili, di interpretazione di stime dell'impatto medio riferite, in realtà, a particolari sottopopolazioni.

Vi è poi un ulteriore, rilevante punto generale al quale le riflessioni svolte conducono. La validità dei risultati della valutazione d'impatto poggia in modo imprescindibile – abbiamo visto – sulla capacità dell'analista di proteggere la valutazione da tutte le minacce rilevanti. Ora, per quanto attente siano state l'analisi del problema, l'acquisizione delle informazioni e il disegno dei confronti fra soggetti esposti e non esposti, in molte circostanze è problematico escludere che vi sia una minaccia alla validità che non è stata diagnosticata (o che non è stata adeguatamente debellata). Ne discende che, in generale, una valutazione non stabilisce *definitivamente* se la politica ha o meno prodotto un impatto, e di quale entità. Ne fornisce piuttosto una plausibile stima, ferma restando la possibilità che qualche altro analista fornisca una stima significativamente diversa, più plausibile e meglio suffragata dall'evidenza empirica. In questa ottica, la valutazione di una specifica politica più che un'esperienza circoscritta ad opera di un

singolo analista è un processo che si dispiega nel tempo, animato dal confronto tra più analisti indipendenti l'uno dall'altro, che in un certo senso operano in competizione. Con Heckman e Smith (1995, p. 93) possiamo concludere rimarcando che “*evaluations build on cumulative knowledge*”.

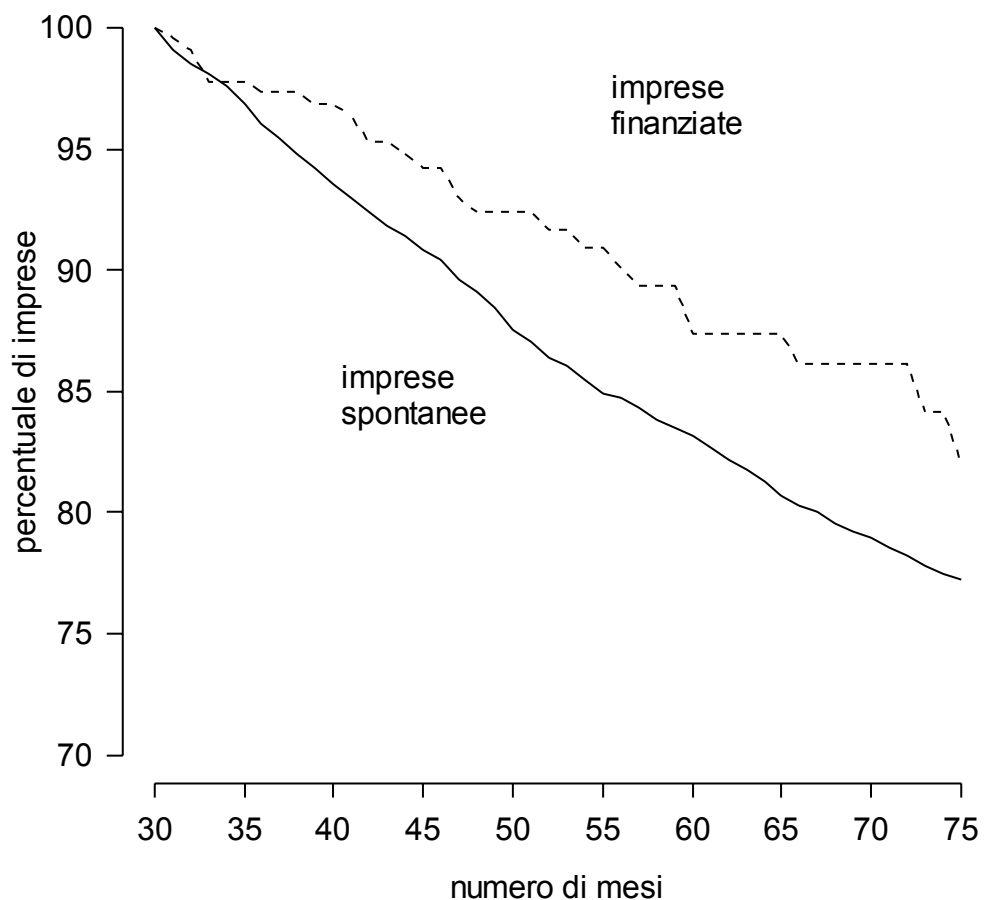
Riferimenti bibliografici

- Battistin E., A. Gavosto and E. Rettore (2001), “Why do subsidised firms survive longer? An evaluation of a program promoting youth entrepreneurship in Italy”, in M. Lechner and F. Pfeiffer (eds.), *Econometric evaluation of labour market policies*, Physica-Verlag, Heidelberg.
- Battistin E. and E. Rettore (2002), “Testing for the presence of a programme effect in a regression discontinuity design with non compliance”, *Journal of the Royal Statistical Society, A*, 165: 39-57.
- Bell B., R. Blundell and J. Van Reenen (1999), “Getting the unemployed back to work: an evaluation of the New Deal proposal”, *International Tax and Public Finance*, 6: 339-360.
- Berliri C. e C. Pappalardo (2001), “La valutazione di efficacia dei corsi di formazione regionale: effetti di autoselezione e stima degli impatti occupazionali netti”, comunicazione al XVI Convegno nazionale di Economia del Lavoro, Firenze, 4-5 ottobre 2001 (mimeo.).
- Bjorklund A. and R. Moffit (1987), “Estimation of wage gains and welfare gains in self-selection models”, *Review of Economics and Statistics*, 69: 42-49.
- Boeri T. (1997), “Un sussidio che crea nuovi posti”, *Il Sole-24 Ore*, 27 maggio 1997.
- Borzaga C. e G. Brunello (a cura di) (1997), *L'impatto delle politiche attive del lavoro in Italia*, Edizioni Scientifiche Italiane, Firenze.
- Brunello, G. and R. Miniaci (1997a). “Benefit transfers in Italy: an empirical study of mobility lists in the Milan area”, *Oxford Bulletin of Economics and Statistics*, 59: 329-347.
- Burtless G. (1995), “The case for randomized field trials in economic and policy research”, *Journal of Economic Perspectives*, 9 (2): 63-84.
- Calmfors L. (1994), “Active labour market policy and unemployment. A framework for the analysis of crucial design features”, *OECD Economic Studies*, 22 (1): 7-47.
- Caroleo F., P. Clarizia, P. Di Monte e N. O' Higgins (1997), “Liste di (im)mobilità? L'impatto della L. 223 sulla probabilità di rioccupazione. Studio di un caso: la Campania”, in C. Borzaga e G. Brunello (a cura di), *cit.*
- Caruso E. (2001), “Durata della disoccupazione e probabilità di impiego in presenza di incentivi all'assunzione: l'evidenza della lista di mobilità della regione Umbria”, *Politica Economica*, 17 (1): 73-95.
- Ceccarelli D. (2000), *Valutare le politiche del lavoro. La valutazione di impatto come fattore di programmazione: il caso della Valle d'Aosta*, Franco Angeli, Milano.

- Centra M., R. De Vincenzi e C. Villante (2000), *Formazione professionale e occupabilità. La valutazione dell'impatto degli interventi formativi sugli esiti occupazionali*, Isfol, Roma.
- Copas J.B. and H.G. Li (1997), "Inference for non-random samples", *Journal of the Royal Statistical Society, B*, 59: 55-95.
- Croce G. e A. Montanino (1997), "Formazione degli adulti e politiche del lavoro: quale impatto da interventi congiunti?", in L. Frey (a cura di), *Formazione e lavoro*, Quaderni di Economia del Lavoro n. 58, Franco Angeli, Milano.
- European Commission (1997), *Joint Employment Report*, Directorate General V, Bruxelles.
- Greenberg D. and M. Shroder (1997), *The digest of social experiments*, The Urban Institute Press, Washington, D.C..
- Hahn J, P. Todd and W. Van der Klaauw (2001), "Identification and estimation of treatment effects with a regression-discontinuity design", *Econometrica*, 69: 201-209.
- Heckman J. (1979), "Sample selection bias as a specification error", *Econometrica*, 47: 153-161.
- Heckman J., R. LaLonde and J. Smith (1999), 'The economics and econometrics of active labor market programs', in O. Ashenfelter e D. Card (eds.), *Handbook of Labor Economics – Vol. 3A*, North-Holland, Amsterdam.
- Heckman J. and J. Smith (1995), "Assessing the case for social experiments", *Journal of Economic Perspectives*, 9 (2): 85-110.
- Heckman J., J. Smith and N. Clements (1997), "Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts", *Review of Economic Studies*, 64: 487-535.
- Holland P. (1986), "Statistics and causal inference", *Journal of the American Statistical Association*, 81: 945-960.
- Imbens G. and J. Angrist (1994), "Identification and estimation of local average treatment effects", *Econometrica*, 62: 467-476.
- Imbens G. and D. Rubin (1997), "Estimating outcome distributions for compliers in instrumental variables models", *Review of Economic Studies*, 64: 555-574.
- Jensen K. and P. Madsen (eds.) (1993), *Measuring labour market measures - Evaluating the effects of active labour market policies*, Ministry of Labour, Copenhagen.
- Laudisa F. (1998), "Come valutare l'efficacia dei corsi di formazione professionale", Tesi di Master, COREP, Torino (mimeo.).
- Lechner M and F. Pfeiffer (eds.) (2001), *Econometric evaluation of labour market policies*, Physica-Verlag, Heidelberg..
- Manski C. and I. Garfinkel (eds.) (1992), *Evaluating welfare and training programs*, Harvard University Press, Cambridge (MA).
- Martini A. (1997), *Valutazione dell'efficacia di interventi pubblici contro la povertà: questioni di metodo e studi di casi*, Commissione di Indagine sulla Povertà e l'Emarginazione Sociale, Presidenza del Consiglio dei Ministri, Roma.

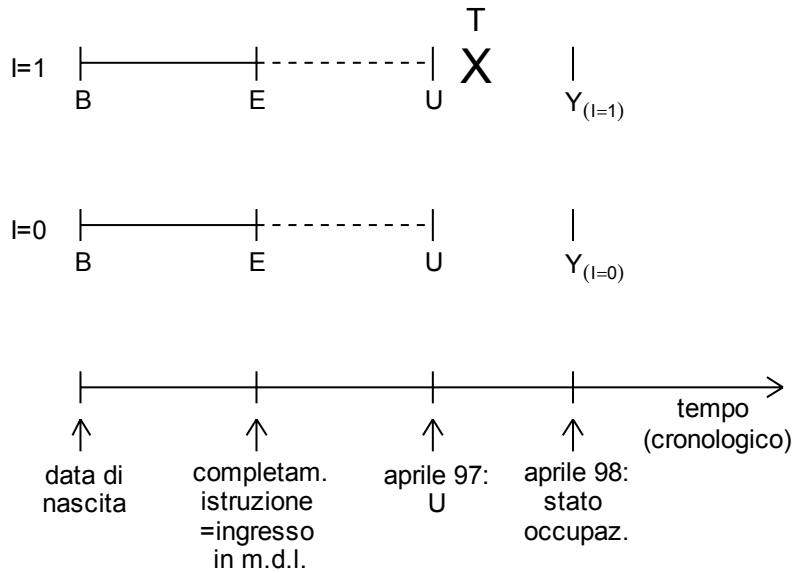
- Martini A. e P. Garibaldi (1993), "L'informazione statistica per il monitoraggio e la valutazione degli interventi di politica del lavoro", *Economia & Lavoro*, 27 (1): 3-22.
- Mazzotta F. (1999), "Efficienza tecnica delle nuove imprese nelle aree in crisi: un'analisi sulla legge 44", *L'Industria*, 20 (3): 511-536.
- Ministero del Lavoro e della Previdenza Sociale (2000), *Rapporto di monitoraggio sulle politiche occupazionale e del lavoro. N.1/1999*, Roma.
- Ministero del Lavoro e della Previdenza Sociale (2001), *Rapporto di monitoraggio sulle politiche occupazionale e del lavoro. N.2/2000*, Roma [con "Nota di aggiornamento" del febbraio 2001].
- OECD (1991), *Evaluating labour market and social programmes. The state of a complex art*, Paris.
- Orr. L., H. Bloom, S. Bell, F. Doolittle, W. Lin and G. Cave (1995), *Does training for the disadvantaged work? Evidence from the national JTPA study*, The Urban Institute Press, Washington, D.C..
- Paggiaro A. and U. Trivellato (2002), "Assessing the effects of the 'Mobility Lists' programme by flexible duration models", *Labour*, 16 (2), 2002, pp. 1-32.
- Paggiaro A. e U. Trivellato (2001), "Il monitoraggio e la valutazione delle 'Liste di mobilità': prime evidenze dalla integrazione con gli archivi Netlabor", in U. Trivellato (a cura di), *Servizi per l'impiego e ricerche sul lavoro. L'esperienza del Veneto*, Quaderni di Economia del Lavoro n. 72, Milano, Franco Angeli, 2001 pp. 121-141.
- Pawson R. and N. Tilley (1997), *Realistic evaluation*, Sage Publications, Beverly Hill (CA).
- Rettore E. e U. Trivellato (1999), "Come disegnare e valutare politiche attive del lavoro", *Il Mulino*, 48 (385): 891-904.
- Rotnitzky A., D.R. Cox, M. Bottai and J. Robins (2000), "Likelihood-based inference with singular information matrix", *Bernoulli*, 6(2): 243-284.
- Rubin D. (1974), "Estimating causal effects of treatments in randomized and non-randomized studies", *Journal of Educational Psychology*, 66: 688-701.
- Rubin D. (1977), "Assignment to treatment group on the basis of a covariate", *Journal of Educational Statistics*, 2: 4-58.
- Samek Lodovici E. (1996), 'La valutazione delle politiche attive del lavoro: l'esperienza internazionale e il caso italiano', *Economia & Lavoro*, 29 (1): 63-96.
- Società Italiana di Statistica (2001), *Processi e metodi statistici di valutazione*, Atti del convegno intermedio, Roma 4-6 giugno 2001, 2 voll. (mimeo.).
- Trochim W. (1984), *Research design for program evaluation: the regression-discontinuity approach*, Sage Publications, Beverly Hill (CA).
- Wooldridge J.M. (1999), *Introductory econometrics: a modern approach*, South-Western College Publishing, Australia.

Figura 1: *Valutazione degli interventi di incentivazione dell'imprenditorialità giovanile (legge n. 44/1986): confronto fra la sopravvivenza delle imprese finanziate e di quelle non finanziate**



* Le funzioni di sopravvivenza si riferiscono alle imprese ancora in vita al 30^ mese. Le differenze osservate tra le due funzioni di sopravvivenza non risultano statisticamente significative ai convenzionali livelli di conduzione dei test.
Fonte: nostre elaborazioni su dati di Battistin, Gavosto e Rettore (2001).

Figura 2: *Schema di riferimento per l'esame della valutazione dell'impatto di programmi di formazione professionale condotta da Centra, De Vincenzi e Villante (2000). Esempificazione delle caratteristiche comuni e dell'eterogeneità non osservata fra il gruppo dei formati (I=1) e il "gruppo di confronto Istat" (I=0) nel migliore dei casi possibili: due persone dello stesso genere, età e livello di istruzione**



* Eventi o stati: B = nascita; E = completamento dell'istruzione, con contemporaneo ingresso nel mercato del lavoro; U = disoccupazione all'aprile 1997; Y = stato occupazionale all'aprile 1998.

Figura 3: *Stima dell'impatto medio con modelli parametrici di regressione o, alternativamente, con metodi di abbinamento: illustrazione di ipotetici risultati per una variabile obiettivo Y, nel caso (a) di supporti disgiunti per esposti e non esposti e (b) di supporti con intersezione non vuota per esposti e non esposti**

Figura 3a: *Caso di supporti disgiunti*

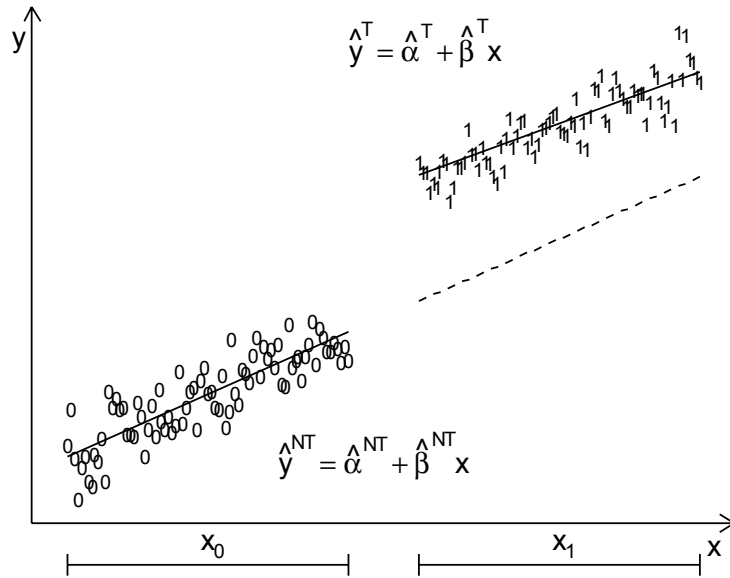
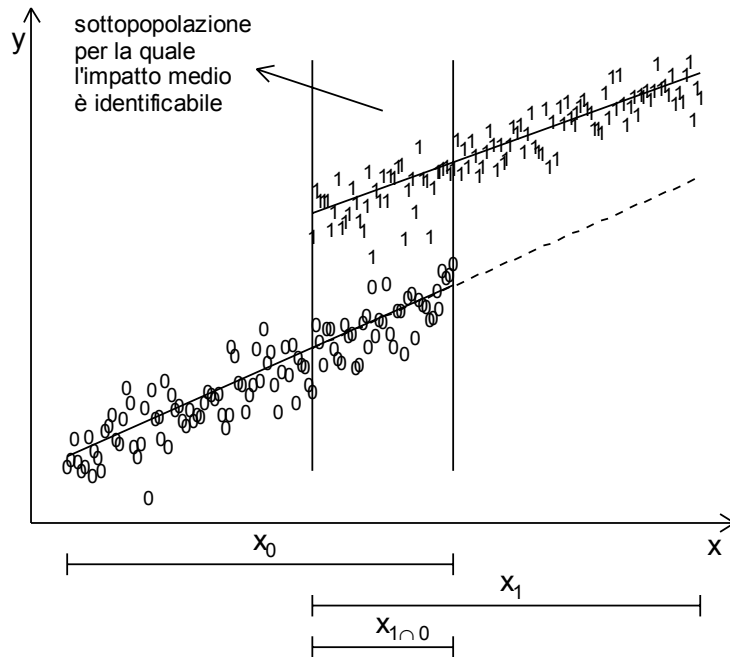


Figura 3b: *Caso di supporti con intersezione non vuota*



* X_0 è il supporto per i non esposti; X_1 è il supporto per gli esposti; nel caso (b), $X_1 \cap 0$ è l'intersezione non vuota dei due supporti.

Figura 4: *Valutazione dell'impatto di un programma di formazione professionale con ammissione in base alla graduatoria, e con indagine di "follow up" condotta su tutti i partecipanti al test di ammissione**

Figura 4a: *Stima dell'impatto medio in corrispondenza della soglia per l'ammissione*

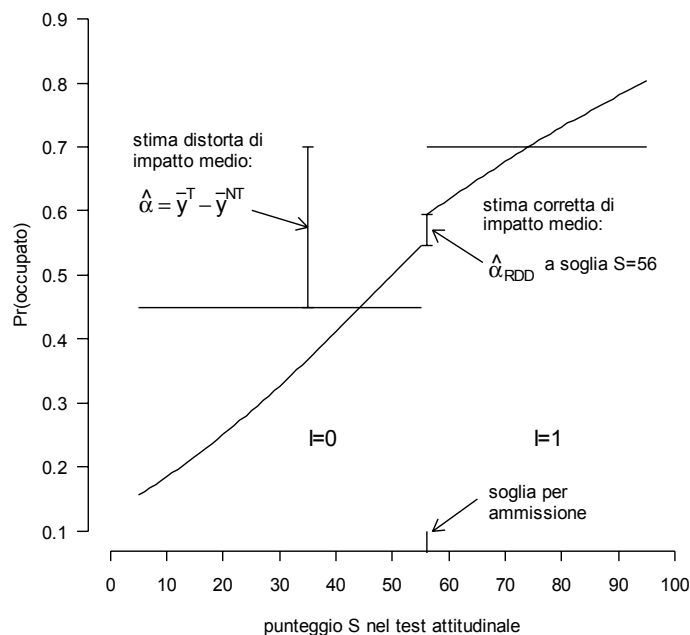
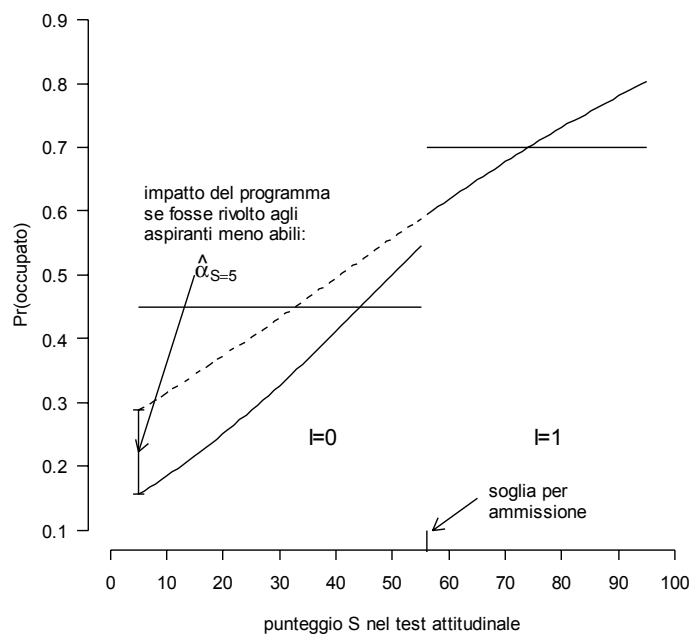


Figura 4b: *Stima dell'impatto medio per soggetti con punteggio $S=5$, condizionatamente all'estrapolazione della curva $Pr(\text{occupato}|S, I=1)$*



* $Pr(\text{occupato})$ = probabilità di essere occupato a 16 mesi dalla conclusione del corso.
Fonte: nostre elaborazioni da Laudisa (1998).