

Economia

**Temi avanzati
di economia e politica
della cooperazione
internazionale
allo sviluppo**

a cura di
**Mario Biggeri
Giovanni Caritani**

FrancoAngeli

A Elena e Daniela

INDICE

Prefazione , di <i>Franco Volpi</i>	pag.	15
--	------	----

PARTE 1

I TEMI TRASVERSALI DELLE POLITICHE DI COOPERAZIONE

1. IL PANORAMA DELLA COOPERAZIONE INTERNAZIONALE NEL XXI SECOLO , di <i>Giovanni Canitano e Mario Biggeri</i>	»	21
1.1. Nuove realtà del sistema di cooperazione internazionale allo sviluppo	»	21
1.2. Gli obiettivi e gli attori della cooperazione allo sviluppo	»	23
1.3. Gli strumenti di trasferimento e la misurazione dei flussi finanziari	»	28
1.4. Conclusioni	»	32
2. SVILUPPO UMANO SOSTENIBILE, CAPABILITY APPROACH E COOPERAZIONE INTERNAZIONALE , di <i>Mario Biggeri e Enrica Chiappero Martinetti</i>	»	37
2.1. Introduzione	»	37
2.2. Sviluppo umano sostenibile e <i>capability approach</i> : alcuni aspetti di definizione	»	39
2.3. Le implicazioni di <i>policy</i>	»	43
2.3.1. Sviluppo umano sostenibile e processi di cooperazione tra sistemi locali di sviluppo	»	45
2.3.2. Le sinergie tra le <i>capability</i> e l'attuazione delle <i>policy</i>	»	47
2.4. Le sinergie nei servizi sociali di base	»	52
2.5. Lo sviluppo umano nella cooperazione internazionale	»	56
2.5.1. BSS e cooperazione internazionale	»	56
2.5.2. Cooperazione internazionale e sviluppo umano sostenibile a livello locale	»	58
2.6. Conclusioni	»	60

3. LA RELAZIONE FRA DIRITTI UMANI E SVILUPPO UMANO NELL'ESPERIENZA CONCRETA DI COOPERAZIONE IN PAESI IN CONFLITTO , di <i>Alessandro Preti</i>	pag. 65
3.1. I diritti umani: alcune definizioni	» 65
3.2. Relazioni fra diritti umani e sviluppo	» 69
3.3. L'approccio dei diritti umani allo sviluppo (<i>Human Rights-Based Approach development – HRBA</i>)	» 72
3.4. Riflessioni su due casi di studio di paesi in conflitto e post-conflitto	» 75
3.4.1. Difendere i diritti umani nel paese dell'eterna impunità: partecipazione a una Missione di Verifica sui Diritti Umani (MINUGUA) e al Programma di Lotta contro la povertà (UNDP) in Guatemala	» 75
3.4.2. Possono lo sviluppo umano e i diritti umani contribuire alla costruzione della pace? La realizzazione di una Strategia territoriale per lo Sviluppo e la Pace in Colombia (UNDP)	» 79
3.5. Conclusioni critiche sul ruolo della cooperazione internazionale in contesti di conflitto	» 83
4. PARTECIPAZIONE E COOPERAZIONE INTERNAZIONALE: PER UNO SVILUPPO “DAL BASSO”? , di <i>Renato Libanora</i>	» 86
4.1. La partecipazione come paradigma innovativo nei campi dello sviluppo	» 87
4.2. La partecipazione come condizione di successo e legittimità dei progetti di sviluppo	» 90
4.3. Genesi di una pratica di indagine “dal basso”: l'osservazione partecipante	» 92
4.4. Dai PRAs ai PLAs: strumenti per la ricerca-azione partecipativa	» 95
4.5. Estensione delle metodologie partecipative al M&V e alla teoria delle <i>capability</i>	» 98
4.6. Conclusioni: lo sviluppo partecipativo come rischio da correre	» 101
5. AZIONE COLLETTIVA, CO-SVILUPPO E IMPRESE COOPERATIVE , di <i>Nicolò Bellanca, Mario Biggeri, Pasquale De Muro e Sara Vicari</i>	» 108
5.1. Introduzione	» 108
5.2. Le coordinate teoriche del co-sviluppo	» 109
5.3. Azione collettiva e azione comune nelle imprese cooperative	» 113
5.4. Azione collettiva, settore informale, associazioni di base ed <i>empowerment</i>	» 117
5.5. Il ruolo dell'impresa cooperativa nel co-sviluppo internazionale	» 120
5.6. Prospettive	» 125
6. MIGRAZIONI, RIMESSE E AIUTO ALLO SVILUPPO , di <i>Francesca Marchetta e Simone Bertoli</i>	» 129
6.1. Introduzione	» 129
6.2. Rimesse dei migranti e flussi di aiuto	» 130
6.3. Migrazione e aiuto: un rapporto di sostituzione?	» 132

6.4. Le possibili complementarità	pag.	136
6.5. Possibili incoerenze fra politiche migratorie e cooperazione	»	139
6.6. Conclusioni	»	142
7. AMBIENTE, POVERTÀ E AIUTO INTERNAZIONALE , di <i>Giovanni Canitano e Elisa Tici</i>	»	146
7.1. Introduzione	»	146
7.2. La complessa relazione tra povertà e degrado ambientale	»	149
7.2.1. Accesso a risorse naturali non degradate	»	150
7.2.2. Capacità dei poveri di sfruttare i benefici potenziali del capitale naturale a cui hanno accesso e capacità di adottare pratiche di gestione delle risorse naturali sostenibili	»	151
7.2.3. Grado di dipendenza dal capitale naturale	»	152
7.2.4. Capacità di adottare strategie difensive e adattive	»	152
7.2.5. Contesto istituzionale	»	153
7.3. La diversa percezione del ruolo della protezione ambientale	»	153
7.4. L'aiuto allo sviluppo nel settore ambientale	»	157
7.5. I meccanismi di erogazione degli aiuti ambientali	»	163
7.6. Le implicazioni strategiche del disallineamento degli obiettivi	»	166
7.7. Conclusioni	»	168
8. CONFLITTI, EMERGENZE E AIUTO UMANITARIO , di <i>Gianni Ruffini</i>	»	172
8.1. Politica ed etica umanitaria	»	173
8.2. Dalle emergenze complesse alle guerre umanitarie	»	174
8.3. La diversità degli attori	»	176
8.3.1. Le Nazioni Unite (UN)	»	177
8.3.2. La Croce Rossa	»	177
8.3.3. ECHO	»	178
8.3.4. I militari	»	178
8.3.5. Le ONG	»	178
8.3.6. Le agenzie governative	»	179
8.3.7. Le imprese private	»	179
8.4. La ridefinizione dell'umanitario	»	180
8.5. Aiuto e politica estera	»	182
8.6. Il problema delle risorse	»	185
8.7. Conclusioni	»	186
9. LA COOPERAZIONE INTERNAZIONALE SUD-SUD: IL SISTEMA DI COOPERAZIONE CINESE IN AFRICA E LA COOPERAZIONE TRA CUBA E VENEZUELA , di <i>Mario Biggeri, Fausto Montagna e Marco Sanfilippo</i>	»	189
9.1. La Cooperazione Sud-Sud	»	189
9.1.1. Radici storiche	»	189

9.1.2. La cooperazione Sud-Sud nel XXI secolo	pag.	192
9.2. La cooperazione sino-africana: prospettiva storica e analisi empirica	»	195
9.2.1. Introduzione	»	195
9.2.2. Il sistema di aiuto cinese	»	197
9.2.3. Una prospettiva storica	»	199
9.2.4. La cooperazione internazionale cinese in Africa	»	201
9.2.5. La cooperazione internazionale cinese in Africa: analisi delle determinati	»	206
9.2.6. La cooperazione sino-africana: conclusioni	»	211
9.3. La cooperazione cubano-venezuelana	»	212
9.3.1. Evoluzione della cooperazione cubano-venezuelana	»	213
9.3.2. La seconda fase della cooperazione cubano-venezuelana	»	214
9.3.3. Caratteristiche della cooperazione cubano-venezuelana	»	216
9.3.4. Commercio, investimenti e aiuto cubano-venezuelani	»	219
9.3.5. La cooperazione cubano-venezuelana: conclusioni	»	222
9.4. Conclusioni	»	224

PARTE 2 MECCANISMI E STRUMENTI DI EROGAZIONE DELL'AIUTO

10. INCENTIVI E GOVERNANCE NEL SISTEMA DI EROGAZIONE DEGLI AIUTI , di <i>Giovanni Canitano e Nicolò Bellanca</i>	»	231
10.1. Il complesso sistema di erogazione degli aiuti	»	231
10.1.1. La specificità del sistema di aiuti allo sviluppo e d'emergenza	»	233
10.1.2. Il sistema di erogazione degli aiuti come catena di rapporti di agenzia	»	234
10.2. Il funzionamento delle agenzie di cooperazione	»	235
10.2.1. Le inefficienze delle agenzie di cooperazione	»	235
10.2.2. Perché esistono le agenzie di cooperazione	»	237
10.2.3. Le agenzie di cooperazione in condizioni di elevata incertezza	»	239
10.3. Le interazioni tra <i>contractor</i> , agenzie e donatori	»	240
10.3.1. Il contesto della relazione e i rapporti tra le parti	»	241
10.3.2. Contrattazione e negoziazione nell'attuazione dei progetti	»	243
10.3.3. Il dilemma del funzionamento delle ONG negli interventi di emergenza	»	246
10.4. Il ruolo e i limiti della valutazione	»	249
10.5. Conclusioni	»	253
11. IL SUPPORTO DIRETTO AL BILANCIO DELLO STATO: ANALISI DI IMPATTO ATTRAVERSO LE MATRICI DI CONTABILITÀ SOCIALE (SAM) , di <i>Marco Missaglia e Lia Quartapelle</i>	»	257
11.1. Introduzione	»	257
11.2. Il supporto diretto al bilancio dello Stato	»	257

11.2.1. Cosa è il supporto diretto al bilancio dello Stato	pag.	257
11.2.2. Genesi di una modalità di aiuto	»	259
11.2.3. Effetti del supporto diretto al bilancio dello Stato	»	263
11.3. Il modello dei moltiplicatori SAM	»	266
11.3.1. Alle origini degli studi su sviluppo e cambiamento strutturale	»	266
11.3.2. L'analisi dei moltiplicatori	»	269
11.3.3. L'analisi dei moltiplicatori nel contesto di una SAM	»	273
11.3.4. L'analisi dei moltiplicatori e il sostegno al bilancio dello Stato	»	278
11.3.5. I modelli CGE	»	282
11.4. Back to the future?	»	284

12. IL *TARGETING*: UNO STRUMENTO PER MIGLIORARE

L'EFFICIENZA DELL'AIUTO , di <i>Donato Romano e Fabio Boncinelli</i>	»	288
12.1. Introduzione	»	288
12.2. I <i>tradeoff</i> del <i>targeting</i>	»	289
12.3. I costi e i benefici del <i>targeting</i>	»	293
12.3.1. I costi amministrativi	»	294
12.3.2. I costi privati	»	296
12.3.3. I costi incentivo	»	296
12.3.4. I costi sociali	»	297
12.3.5. I costi politici	»	298
12.3.6. I benefici	»	298
12.4. I diversi metodi di <i>targeting</i>	»	299
12.4.1. Test dei mezzi semplice (<i>means test</i>)	»	300
12.4.2. Test dei mezzi basato su proxy (<i>proxy means test</i>)	»	301
12.4.3. L'approccio comunitario (<i>community-based targeting</i>)	»	302
12.4.4. Il <i>targeting</i> geografico	»	304
12.4.5. Il <i>targeting</i> demografico	»	305
12.4.6. Il <i>targeting</i> basato sull'auto-selezione (<i>self-targeting</i>)	»	305
12.5. La valutazione dell'impatto del <i>targeting</i>	»	308
12.6. Conclusioni	»	311

13. ANALISI DI MONITORAGGIO CONTINUO: I *FEEDBACK LOOP* NELLE ORGANIZZAZIONI DELL'AIUTO

di <i>Enrico Testi e Marco Tognetti</i>	»	315
13.1. Introduzione	»	315
13.2. Il Monitoraggio e la Valutazione	»	317
13.3. Analisi di Monitoraggio Continuo	»	319
13.4. I <i>feedback loop</i> e il loro ruolo nelle organizzazioni dell'aiuto	»	325
13.5. Conclusioni	»	327

14. METODI STATISTICI PER LA VALUTAZIONE DI IMPATTO DEI PROGRAMMI DI AIUTO , di <i>Vincenzo Mauro</i>	pag.	331
14.1. Introduzione	»	331
14.2. Il modello causale di Rubin	»	333
14.2.1. Unità, trattamento, risultati potenziali	»	333
14.2.2. Effetti causali e distorsione da selezione	»	334
14.2.3. Il meccanismo d'assegnazione al trattamento	»	336
14.3. Metodi sperimentali	»	337
14.4. Metodi non sperimentali	»	340
14.4.1. Procedure di abbinamento	»	343
14.4.2. Il metodo difference in differences	»	347
14.4.3. Regression discontinuity design	»	348
14.4.4. L'uso di variabili strumentali	»	349
14.4.5. Altri metodi di stima	»	352
14.5. Riflessioni conclusive	»	353

PARTE 3 GLI AIUTI SETTORIALI

15. SVILUPPO UMANO LOCALE ATTRAVERSO CLUSTER DI PMI , di <i>Marco Bellandi, Mario Biggeri e Francesca Marchetta</i>	»	361
15.1. Introduzione	»	361
15.2. Sviluppo locale e <i>cluster</i> di PMI	»	363
15.3. Percorsi e qualità dello sviluppo locale: un modello interpretativo	»	366
15.4. Una sintesi dei programmi attuati dalle agenzie internazionali e dall'UE nei PVS	»	371
15.5. Implicazioni di <i>policy</i>	»	374
15.5.1. <i>Up-grading</i> dalla <i>dirt road</i>	»	376
15.5.2. <i>Up-grading</i> dalla <i>low road</i>	»	377
15.6. Conclusioni	»	378
16. STRATEGIE DI AIUTO ALLO SVILUPPO: WINDOWS OF OPPORTUNITY E BOOTSTRAPPING STRATEGIES , di <i>Mauro Lombardi e Marika Macchi</i>	»	382
16.1. Introduzione	»	382
16.2. Principi e criteri costitutivi della <i>Global Digital Age</i>	»	383
16.3. Aspetti fondamentali dei processi di innovazione e di crescita	»	384
16.4. Morfologia dei processi produttivi e le nuove "leve competitive"	»	386
16.4.1. Nuova morfologia dei processi produttivi	»	386
16.4.2. Le nuove leve competitive	»	388
16.5. Il nuovo pattern della dinamica innovativa: il modello <i>Open Source</i>	»	389
16.6. Nuovi paradigmi e regole di comportamento per le imprese ed economie emergenti	»	392

17. AIUTO, AGRICOLTURA E SVILUPPO RURALE , di <i>Francesco Burchi, Pasquale De Muro e Sara Turchetti</i>	pag.	399
17.1. Introduzione	»	399
17.2. Tendenze dell'aiuto internazionale al settore agricolo	»	399
17.2.1. Analisi Settoriale	»	401
17.2.2. Analisi per Beneficiario	»	402
17.2.3. Analisi per Donatore	»	403
17.3. Evoluzione delle teorie e delle politiche	»	406
17.3.1. Tra cambiamento strutturale e aiuto	»	406
17.3.2. L'epoca d'oro dell'aiuto all'agricoltura	»	407
17.3.3. La fine dell'epoca d'oro: i Programmi di Aggiustamento Strutturale	»	409
17.4. Evoluzione rispetto alle componenti	»	413
17.5. Conclusioni	»	415
18. AIUTO ALIMENTARE E AIUTO ALLA SICUREZZA ALIMENTARE , di <i>Francesco Burchi e Sara Turchetti</i>	»	418
18.1. Aiuto alimentare: definizione e obiettivi	»	418
18.2. Un quadro dell'aiuto alimentare	»	419
18.3. Categorie di aiuto alimentare	»	421
18.4. Aiuto alla sicurezza alimentare	»	427
18.5. Il dibattito sull'aiuto alimentare	»	429
18.6. Conclusioni	»	433
19. LA PROMOZIONE DELL'UGUAGLIANZA DI GENERE NELLE POLITICHE ESTERE DI SVILUPPO , di <i>Lisa Ann Richey</i>	»	435
19.1. Introduzione	»	435
19.2. Le donne come forza di sviluppo	»	436
19.2.1. Gli approcci Women in Development (WID)	»	437
19.2.2. L'approccio Gender and Development (GAD)	»	440
19.3. L'impatto del genere e dello sviluppo sull'aiuto estero	»	443
19.3.1. Mainstreaming di genere per l'adozione dell'approccio GAD	»	443
19.3.2. Le potenzialità del GAD per un cambiamento incrementale	»	444
19.3.3. Integrare il GAD nella politica dell'aiuto	»	449
19.4. Conclusioni	»	449
20. AIUTO INTERNAZIONALE VERSO I MINORI: EVOLUZIONE STORICA, APPROCCI, SETTORI E PROSPETTIVE , di <i>Mario Biggeri e Aesa Pighini</i>	»	457
20.1. Introduzione	»	457
20.2. Evoluzione storica	»	458
20.3. Approcci teorici e aspetti operativi	»	461
20.4. Settori di intervento	»	465

20.4.1. Salute	pag.	467
20.4.2. Istruzione	»	468
20.4.3. Emergenza e rifugiati	»	469
20.4.4. Child Protection (ex bambini soldato, lavoro minorile, traffico di minori ecc.)	»	469
20.4.5. HIV	»	470
20.5. Conclusioni	»	471

APPENDICI

APPENDICE 1

AIUTO INTERNAZIONALE VERSO I BSS, di <i>Enrico Testi</i>	»	477
--	---	-----

APPENDICE 2

L'INIZIATIVA ART, di <i>Andrea Ferrannini</i>	»	484
---	---	-----

APPENDICE 3

ANALISI DEL CAPITALE SOCIALE NEI PROGRAMMI E PROGETTI DI AIUTO, di <i>Claudio Fedi e Mario Biggeri</i>	»	491
--	---	-----

APPENDICE 4

UN CASO DI PARTNERSHIP DI AIUTO E SVILUPPO TRA COOPERATIVE DI PRODUZIONE E LAVORO, di <i>Dario Marmo</i>	»	500
--	---	-----

APPENDICE 5

AIUTO ED EMERGENZA, di <i>Enrico Testi</i>	»	505
--	---	-----

Autori	»	511
---------------	---	-----

Ringraziamenti	»	515
-----------------------	---	-----

14. METODI STATISTICI PER LA VALUTAZIONE DI IMPATTO DEI PROGRAMMI DI AIUTO

di *Vincenzo Mauro*

14.1. Introduzione

Come sottolineato nel capitolo introduttivo, tra le principali questioni riguardanti l'aiuto internazionale allo sviluppo perdura senza dubbio un nodo irrisolto: l'effettivo impatto dell'aiuto, sia esso macroeconomico o a livello di programma. Se l'impatto macroeconomico è stato ampiamente affrontato nella letteratura (si veda ad es. Canitano, 2005; McGillivray *et al.*, 2006; Rajan e Subramanian, 2005), l'analisi di impatto relativa ai programmi e progetti di aiuto è stata troppo spesso trascurata (Riddell, 2007). L'obiettivo di questo capitolo è dunque quello di introdurre i principali approcci e metodi utilizzabili per la valutazione d'impatto dei programmi di aiuto (politiche pubbliche).

Per politica pubblica¹ la cui definizione può essere sfumata², intenderemo *un piano di azioni decise e coordinate da uno o più attori al fine di affrontare un problema collettivo in modo razionale*.

Pur essendo questo un ambito sul quale in Italia è stata posta recentemente una certa enfasi, la letteratura al riguardo non sembra ancora sufficientemente consolidata, soprattutto per le difficoltà nel reperire e integrare i dati necessari alle analisi. Al di là dell'interessante dibattito internazionale tra le principali agenzie di aiuto internazionale (Biggeri e Volpi, 2007) la vasta letteratura sull'impatto macroeconomico dell'aiuto è quindi poco incisiva³, soprattutto in termini di consigli pratici a decisori pubblici (*policy makers*) e amministratori dell'aiuto per migliorarne l'efficacia in termini di sviluppo (White, 2007a, 2007b). Gli studi a livello di programma e di progetto possono invece portare sia a risultati più affidabili, sia a informazioni più rilevanti per i policy maker e i manager dell'aiuto (White, 2007b, p. 14).

¹ Il termine "politica", in italiano, viene solitamente seguito dall'aggettivo "pubblica" in modo da non confonderlo con la più comune definizione di politica. Questo problema non sussiste per gli anglosassoni, che distinguono tra *politics* e *policy*.

² Si veda, a tale riguardo Belligni (2003) o Birkland (2001).

³ Secondo Howard White (2007b, p. 1), gran parte delle analisi sull'impatto dell'aiuto è stata condotta a livello macro e gran parte di questa ha supportato la "Aid fatigue" o il pessimismo riguardo l'aiuto mostrando poco o nessun legame tra l'aiuto e la crescita. Tali conclusioni sono fondate su regressioni *cross-country* che hanno numerose debolezze quando vengono applicate a relazioni di aiuto-crescita (White, 2007b).

Per quel che riguarda le metodologie, il punto di riferimento resta l'ampia letteratura degli Stati Uniti, e in parte della Gran Bretagna, dove l'attenzione al problema della corretta valutazione delle politiche si è sviluppata già dagli anni Settanta⁴.

Non potendo certo analizzare in modo consono le varie scuole di pensiero e le metodologie che nel corso degli anni si sono sviluppate, ci limiteremo a descrivere alcuni aspetti chiave, comuni a quasi tutti gli approcci, lasciando al lettore la possibilità di approfondire, eventualmente, attraverso i riferimenti bibliografici proposti.

Una prima distinzione, utile per demarcare gli ambiti descritti in questo capitolo, è quella tra analisi di implementazione (*implementation analysis*) e analisi d'impatto (*impact analysis*). Con la prima espressione si fa riferimento all'analisi delle procedure riguardanti il modo in cui è stato effettuato l'intervento (dalla normativa che l'ha attivata fino all'attuazione finale), mentre con la seconda si richiama la stima dell'impatto dell'intervento su una o più variabili di interesse dei soggetti beneficiari. Anche se entrambe le fasi possono riguardare la stessa politica, restano concettualmente distinte⁵. In questo capitolo la nostra attenzione sarà rivolta alla seconda analisi, quella d'impatto.

Partendo dall'assunto che una politica pubblica si propone di modificare il tessuto sociale esistente per migliorare la situazione presente in un certo ambito, è bene sottolineare cosa intenderemo, in questo capitolo, per effetto netto (o impatto) di una politica. Dato un certo intervento di aiuto, considereremo il suo effetto netto, a livello di singola unità, come una funzione delle due situazioni (presumibilmente differenti) in cui la stessa unità si verrebbe a trovare nel caso in cui fosse beneficiaria dell'intervento, e nel caso in cui non lo fosse. L'aggettivo "netto" si riferisce a tutti i fattori indipendenti dall'intervento in analisi che potrebbero però avere influenza sulle due situazioni di cui sopra.

Si userà il termine "trattamento" come sinonimo di "intervento". Analogamente, le unità "trattate" saranno quelle che hanno ricevuto l'intervento oggetto d'analisi, contrapposte a un gruppo detto "di controllo", comprendente le unità che non hanno beneficiato del trattamento.

Lo scopo principale, comune a tutte le procedure di valutazione che verranno presentate in questo capitolo, è quello di fornire una stima attendibile, basata su varie assunzioni, dell'effetto causale di politiche su alcune variabili di interesse. Il quadro teorico di riferimento è quello dell'inferenza causale.

Nel paragrafo 2 si introduce un approccio all'inferenza causale decisamente consolidato, il Modello Causale di Rubin (Holland, 1986), e si stabiliscono alcuni concetti chiave sui quali si basa una valutazione di impatto. Ci si focalizza poi su un problema tipico di queste analisi, la distorsione da selezione (*selection bias*), che è inquadrato più formalmente introducendo il concetto di meccanismo di assegnazione. Nei paragrafi 3 e 4 si utilizza la definizione di meccanismo di assegnazione per attuare una fondamentale divisione tra approcci sperimentali e non sperimentali. Sono qui brevemente accennate anche le più consolidate

⁴ Per una rassegna dell'ampia letteratura degli ultimi decenni si veda Stame (1998, 2007) e Imbens e Woolridge (2008).

⁵ Una terza parte di analisi può riguardare l'individuazione della logica che ha portato il *policy maker* alla formulazione degli interventi.

tecniche di stima proprie di questi due ambiti, con richiami alla letteratura sia dal punto di vista metodologico che applicato ai programmi di aiuto. Infine, il paragrafo 5 trae alcune riflessioni conclusive.

14.2. Il modello causale di Rubin

Un approccio all'inferenza causale che si è progressivamente sviluppato e diffuso negli ultimi decenni è quello basato sul concetto di risultati potenziali (Neyman, 1923; Rubin, 1974, 1977, 1978, 1990). La struttura metodologica che caratterizza tale approccio, noto in letteratura anche con il nome di Modello causale di Rubin (Rubin Causal Model o RCM, Holland, 1986), si sviluppa a partire da alcuni concetti di base. Questo paragrafo introduce brevemente le nozioni fondamentali di questo approccio allo studio delle relazioni di causa-effetto.

14.2.1. Unità, trattamento, risultati potenziali

I concetti fondamentali su cui si basa l'approccio del RCM sono quelli di unità, trattamento, risultati potenziali.

Un'unità è un oggetto fisicamente ben definito (ad esempio una persona, un'impresa) in un particolare istante di tempo.

Un trattamento è una ben specificata azione cui l'unità può essere esposta o no. Possono essere considerati trattamenti un programma di aiuto per l'infanzia, una somma erogata, un'agevolazione fiscale, un corso di recupero scolastico, ecc. Ovviamente, il trattamento non è, per sua natura, solo dicotomico, ma può assumere anche più valori, variando anche in un continuo di valori: basti pensare all'erogazione di una somma in denaro. In questo capitolo ci limiteremo però all'analisi di trattamenti dicotomici (ricevere versus non ricevere un'azione), una semplificazione che si adatta piuttosto bene alla maggior parte dei programmi di aiuto⁶.

I due risultati potenziali (Rubin, 1974) associati all'applicazione di un trattamento a un'unità sono i due valori di una variabile scelta (normalmente la variabile sulla quale si è interessati a valutare l'impatto del trattamento) misurati in un certo istante (successivo all'applicazione del trattamento) nel caso in cui all'unità sia o non sia assegnato il trattamento.

Fissata una variabile risultato sulla quale si vuole misurare l'impatto del trattamento, il confronto tra i valori di questa variabile, misurata nelle due situazioni potenziali, può essere definito come l'effetto causale del trattamento.

Una differenza fondamentale tra questo e gli altri approcci all'inferenza causale, *in primis* i modelli a equazioni simultanee (Goldberger, 1972; Heckman, 1978), è data dal fatto che l'effetto causale che si vuole misurare è indipendente da qualunque modello probabilistico (detto anche *meccanismo di assegnazione*, Rubin 1978, si veda paragrafo 2.3.) che ha portato

⁶ Per estensioni metodologiche a trattamenti non dicotomici, si veda ad esempio Imbens e Hirano (2004).

all'assegnazione o meno dell'unità al trattamento. Una volta definiti i due risultati potenziali, infatti, ogni singola unità ha un ben definito valore della variabile di risultato in ognuna delle due possibili situazioni, e il trattamento ha dunque un ben definito effetto causale sull'unità, indipendente dal fatto che l'unità sia poi trattata o no.

Questa definizione di effetto causale, anche se intuitivamente molto semplice, nasconde in sé un problema indiscutibile: anche se a priori sono entrambi potenzialmente osservabili, i due risultati non sono in nessun modo osservabili congiuntamente. Questo problema, che da un articolo di Holland (1986) prende il nome, appunto, di un problema fondamentale dell'inferenza causale, apre la strada a una serie di metodologie che tentano, basandosi su differenti assunzioni, di aggirare l'ostacolo fornendo stime attendibili degli impatti medi del trattamento.

14.2.2. Effetti causali e distorsione da selezione

Formalmente, sia T_i l'indicatore di trattamento che vale uno se l' i -esima unità riceve una certa *trattamento* attivo (beneficia di uno specifico intervento d'aiuto) e zero se non lo riceve. Siano $Y(1)$ e $Y(0)$, rispettivamente, il valore della variabile risultato se l'unità riceve il trattamento attivo o quello di controllo. Nel paragrafo 2.1 abbiamo definito, per ogni unità, l'effetto causale del trattamento come un confronto⁷ tra i due risultati potenziali $Y(1)$ e $Y(0)$, sottolineando come l'effetto definito non dipende da quale dei due possibili valori della variabile risultato si sia effettivamente osservato.

Siccome il problema fondamentale dell'inferenza causale impone che, per ogni unità, possa essere osservato al massimo un solo valore della variabile risultato (quello corrispondente al trattamento a cui l'unità è effettivamente assegnata), il risultato potenziale sotto l'altra condizione di trattamento è sempre mancante (missing).

I valori osservati e mancanti della variabile risultato possono essere ora definiti in funzione dei risultati potenziali, $Y(0)$ e $Y(1)$, e dell'indicatore di trattamento, Z , attraverso le seguenti relazioni:

$$Y_{\text{oss}} = T \cdot Y(1) + (1 - T) \cdot Y(0)$$

e

$$Y_{\text{mis}} = (1 - T) \cdot Y(1) + T \cdot Y(0)$$

In generale, non è possibile ottenere informazioni sull'effetto causale di un trattamento per una singola unità, perché la valutazione dell'efficacia dell'intervento richiede il confronto tra i risultati potenziali.

⁷ Solitamente si considera la differenza $Y(1) - Y(0)$, o il rapporto $Y(1)/Y(0)$.

Un modo intuitivo per aggirare questo problema è quello di confrontare la stessa unità esponendola sia al trattamento che al controllo (disegni a misure ripetute). Per far ciò si deve però necessariamente agire in due istanti diversi, in contrasto con la definizione di unità vista in precedenza. In linea di principio non c'è nessuna garanzia che i due risultati ottenuti applicando o meno il trattamento in due diversi istanti $t=1$ e $t=2$ coincidano con i due risultati potenziali ottenuti in uno stesso istante, non fosse altro perché al tempo $t=2$ l'unità ha già subito una certa azione. Il valore della variabile risposta al tempo $t=2$ potrebbe cioè risentire, per esempio, anche dell'effetto del trattamento all'istante $t=1$.

D'altro canto, osservare invece più unità nello stesso istante, esponendone alcune al trattamento e altre no, complica ulteriormente la situazione. Se nel caso di una singola unità i due risultati potenziali dipendono solo dall'assegnazione al trattamento dell'unità stessa, nel caso di più unità questo non è, in generale, sicuro. Il risultato potenziale dell'unità i potrebbe dipendere, oltre che dal valore di T_i^8 , anche dai valori di T relativi alle altre unità, con la complicazione che i risultati potenziali per l'unità i sarebbero più di due. In questo contesto, inoltre, non si ha più la certezza che il trattamento, visto come azione che può essere effettuata o meno su una unità, sia lo stesso per tutto l'insieme di unità con $T=1$. Questo tipo di inferenza su effetti causali, basandosi sull'osservazione di più unità esposte a diversi trattamenti, necessita dunque di opportune ipotesi che limitino sia il numero di risultati potenziali che potrebbero essere generati dalla interdipendenza tra le unità, sia l'esistenza di differenti versioni del trattamento.

Un'assunzione usuale nell'ambito dell'inferenza causale basata su confronto tra differenti unità è l'"assunzione di stabilità", meglio nota con il nome di "Stable Unit Treatment Value Assumption" (SUTVA, Rubin, 1980; 1990). L'ipotesi SUTVA richiede che i risultati potenziali di ogni unità non siano influenzati dal trattamento assegnato alle altre unità (Cox, 1958), e che non esistano versioni alternative del trattamento.

Per fare un esempio relativo al contesto della valutazione di programmi di aiuto, l'assunzione di stabilità esclude, tra le altre cose, che vi siano effetti di *spillover* all'interno dei gruppi di unità definiti dallo status di trattamento: per esempio, i soggetti che non beneficiano del programma non devono trarre nessun beneficio dal fatto che altri soggetti siano stati aiutati.

Sotto questa fondamentale ipotesi (della quale si deve giustificare, ove necessario, l'assunzione), i risultati potenziali di tutte le unità oggetto di analisi possono essere rappresentati tramite una matrice $N \times 2$, dove ogni singola riga, avente due elementi, rappresenta i due risultati potenziali per l' i -esima unità.

Come sottolineato in precedenza, i risultati potenziali identificano gli effetti causali, che sono ben definiti indipendentemente dal trattamento a cui un'unità è esposta, e quindi indipendentemente da quale dei due risultati potenziali è effettivamente osservato. L'impossibilità di osservare contemporaneamente entrambi i risultati potenziali $Y(0)$ e $Y(1)$ per ogni unità sposta l'attenzione su quantità stimabili, ovvero, ad esempio, caratteristiche della distribuzione della differenza $Y(1)-Y(0)$ nella popolazione. Solitamente l'obiettivo della

⁸ Con T_i si è indicato il valore di T per l'unità i , cioè uno scalare, mentre con $T_{(i)}$ si indicherà un generico vettore T .

valutazione è espresso in termini del valore atteso, che prende il nome di effetto causale medio (Average Treatment Effect, ATE):

$$ATE = E[Y(1) - Y(0)] = E[Y(1)] - E[Y(0)]$$

Un'altra quantità di interesse è l'effetto medio sui trattati (Average Treatment Effect on the Treated, ATT):

$$ATT = E[Y(1) - Y(0) | T = 1] = E[Y(1) | T = 1] - E[Y(0) | T = 1]$$

Il problema centrale riguarda la possibilità di utilizzare l'informazione fornita dai trattati e dai non trattati per stimare gli effetti medi sopra definiti. Le strategie di stima tentano solitamente di ottenere un'adeguata approssimazione della situazione che si sarebbe osservata nel caso in cui l'unità fosse stata esposta al trattamento opposto. Questo modo di procedere espone a un rischio: che le unità non esposte al trattamento siano sistematicamente diverse da quelle trattate (delle quali dunque dovrebbero fornire una approssimazione) già prima della assegnazione al trattamento. Questo problema, generalmente noto in letteratura con il nome di *selection bias*, o distorsione da (auto)selezione, rappresenta una minaccia all'affidabilità delle stime in tutti gli studi non sperimentali (quelli cioè dove non si abbia un controllo diretto sull'assegnazione al trattamento. Una differenza *ex ante* tra i due gruppi potrebbe infatti essere erroneamente interpretata *ex post* come un effetto causale del trattamento.

Il *selection bias* nasce dal modo in cui i soggetti sono assegnati ai due gruppi dei trattati e non trattati. Per fare inferenza sugli effetti causali di interesse assume dunque un ruolo chiave il meccanismo di assegnazione, definito come il processo attraverso il quale le unità ricevono il trattamento a cui sono esposte.

14.2.3. Il meccanismo d'assegnazione al trattamento

Il meccanismo di assegnazione è il modo con cui ogni singola unità è assegnata o meno al trattamento. Formalmente, è definito come la distribuzione di probabilità di tutti i possibili modi di assegnare le unità al trattamento.

Dato l'insieme dei singoli vettori N-dimensionali $T_{(i)}$ a valori 0 e 1, ognuno dei quali rappresenta un possibile modo di allocare le N unità, il meccanismo di assegnazione è una funzione, scambiabile per riga, che associa a ogni vettore $T_{(i)}$ una probabilità in modo che valga la relazione

$$\sum_i p(T_{(i)} | X, Y(0), Y(1)) = 1$$

per qualunque x , $y(0)$ e $Y(1)$, dove si è definito X come una matrice $N \times k$ di variabili (covariate) precedenti l'assegnazione al trattamento (e dunque non influenzate da esso).

Se questa matrice di covariate da cui dipende l'assegnazione è osservata, diremo che il meccanismo di assegnazione è detto *ignorabile* (Rubin, 1974, 1977, 1978), mentre se riguarda covariate latenti, o valori di Y non osservati, il meccanismo si dice *non ignorabile*.

Concludiamo il paragrafo con un'ultima importante definizione, quella di *assegnazione probabilistica*. Un meccanismo d'assegnazione si dice *probabilistico* se per ogni i , X , $Y(0)$ e $Y(1)$, la probabilità di assegnazione è strettamente compresa tra 0 e 1, se si ha, cioè:

$$0 < p_i(X, Y(0), Y(1)) < 1$$

Quest'assunzione impone cioè il vincolo che ciascuna unità abbia una probabilità positiva p_i di essere trattata e che non esistano unità che abbiano la certezza di ricevere il trattamento.

Il concetto di meccanismo di assegnazione ci permette di discriminare tra due possibili situazioni di partenza nella valutazione di un programma di aiuto, a seconda che il meccanismo di assegnazione sia sotto il controllo dello sperimentatore o meno. I paragrafi 3 e 4 illustrano più in dettaglio queste due situazioni.

14.3. Metodi sperimentali

Un esempio banale di meccanismo di assegnazione è quello degli esperimenti randomizzati, in cui l'assegnazione al trattamento è sotto il controllo dello sperimentatore. Questo meccanismo è un meccanismo di assegnazione *ignorabile* e *probabilistico*.

Questo meccanismo di assegnazione è alla base della maggior parte degli studi in campo medico, soprattutto per la valutazione degli effetti dei farmaci. È piuttosto raro che una valutazione di una politica pubblica rientri in questo caso particolare, anche se negli ultimi anni si è osservato un crescente interesse verso questo tipo di approccio⁹. In ogni modo, siccome sotto questa assegnazione l'inferenza causale risulta essere piuttosto diretta, il caso degli esperimenti randomizzati rappresenta un importante punto di riferimento per qualunque analisi.

L'aspetto più importante di questo particolare meccanismo d'assegnazione è la sua indipendenza da qualunque forma di distorsione dovuta a variabili non osservate. In altre parole, nell'ambito qui analizzato dei trattamenti binari si può assumere che ogni differenza tra le caratteristiche medie del gruppo dei trattati e dei controlli sia unicamente dovuta al caso. Il problema del *selection bias* viene così eliminato per definizione, visto che l'assegnazione casuale, non dipendendo da nessuna delle caratteristiche intrinseche delle unità, mette al riparo lo sperimentatore dalle distorsioni sistematiche¹⁰.

Oltre a questa, gli esperimenti randomizzati sono caratterizzati da almeno altre due importanti caratteristiche.

⁹ Si veda, ad esempio, Schwartz e Bhushan (2005), Banerjee *et al.* (2007), Glewwe *et al.* (2006), o Banerjee e Duflo (2008) per un'analisi di punti di forza e limitazioni di questo approccio.

¹⁰ Per campioni non sufficientemente ampi, il gruppo dei trattati e quello dei controlli potrebbero però risultare sbilanciati (rispetto a qualche variabile rilevante) anche sotto ipotesi di randomizzazione.

Innanzitutto, la differenza tra le medie osservate dei due gruppi è uno stimatore non distorto per la media della differenza, a livello di singola unità, tra $Y(1)$ e $Y(0)$.

Formalmente si dimostra facilmente che:

$$E(\bar{Y}_1^{obs} - \bar{Y}_0^{obs}) = \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0))$$

dove con i due termini a sinistra si sono indicati i valori osservati della variabile risposta.

Inoltre, nei disegni randomizzati, le affermazioni su quanto fosse verosimile, a priori, osservare quel che si è rilevato, hanno un chiaro e corretto contenuto probabilistico (Rubin, 1974).

Oltre al vettore T e ai vettori dei risultati potenziali $Y(0)$ e $Y(1)$, si è introdotto, nel paragrafo 2.3 una matrice X che rappresenta alcune rilevanti caratteristiche dell'unità i . Il ruolo di tali variabili (anche dette covariate o variabili pre-trattamento) nell'ambito dell'inferenza causale è triplice.

Anzitutto, come in molte altre tipologie di studio, le covariate permettono di rendere i risultati delle analisi più precisi controllando alcune possibili fonti di variazione della variabile risultato.

Inoltre, spesso può essere di interesse valutare effetti causali per specifiche sottopopolazioni definite dai valori delle variabili precedenti il trattamento.

Infine, anche se questo è un aspetto legato più alle metodologie non sperimentali che saranno illustrate nel paragrafo 4, l'importanza delle covariate è principalmente legata alla loro potenziale relazione con il meccanismo di assegnazione al trattamento. Alcune assunzioni sul meccanismo di assegnazione possono apparire più ragionevoli se riferite, invece che all'intera popolazione, a sottopopolazioni omogenee rispetto ai valori di queste variabili.

Quando sono disponibili informazioni su alcune rilevanti caratteristiche delle unità cui è ragionevole imputare parte della variabilità che caratterizza le variabili risultato di interesse, è opportuno disegnare l'esperimento in modo tale che il trattamento sia casualmente assegnato all'interno di ogni sottopopolazione di unità individuata dai livelli delle covariate osservate.

Il meccanismo di assegnazione alla base di tale disegno, denominato in letteratura esperimento randomizzato "a blocchi", crea una situazione in cui la distribuzione del trattamento è nota all'interno di ogni sottoclasse definita dai valori delle variabili osservabili X che individuano ciascun "blocco". Si ha cioè, definendo W come la matrice $N \times b$ (con $b > k$) dei valori di tutte le covariate, sia latenti che osservabili, per ciascuna unità,

$$T \perp W | X$$

e

$$T \perp Y(0), Y(1) | X$$

Da tali condizioni di indipendenza condizionata si ricava che

$$ATE = E[Y(1) | T = 1, X] - E[Y(0) | T = 0, X]$$

La funzione a destra dell'uguaglianza è direttamente stimabile dai dati, quindi un esperimento randomizzato a blocchi permette l'identificazione puntuale degli effetti causali di interesse.

Questi scenari, seppur appena accennati, lasciano intuire con quale semplicità la randomizzazione del trattamento permetta di ottenere stime non distorte degli effetti causali di interesse. Gli studi sperimentali e, più in generale, gli esperimenti randomizzati a blocchi, costituiscono quindi le basi per la valutazione di effetti causali in situazioni più complesse, come gli studi basati su dati osservazionali (in presenza cioè di selezione endogena delle unità trattate, per le quali è ignota la probabilità di assegnazione al trattamento).

Nonostante abbia un'immediatezza e una forza concettuale innegabili, l'approccio sperimentale non è scevro da limiti che, in determinate circostanze, lo rendono meno opportuno di quel che potrebbe sembrare¹¹.

Innanzitutto, è evidente che, almeno in determinati settori, a causa di ostacoli di natura etica, pratica o di opportunità, risulta impossibile poter davvero procedere a una randomizzazione rigorosa delle unità.

Inoltre, per quel che riguarda la valutazione degli interventi (specialmente per quelli già a regime), l'implementazione di procedure complesse per tenere sotto controllo il meccanismo di assegnazione rischia di snaturare la politica stessa. Ne può derivare che, a causa di cambiamenti nelle procedure burocratiche dell'intervento, la politica di aiuto che si valuta risulti, alla fine, differente da quella che si intendeva valutare originariamente.

Infine, i risultati degli esperimenti randomizzati sono solitamente poco generalizzabili (in termini statistici, hanno una scarsa validità esterna). Per poter controllare adeguatamente il meccanismo di assegnazione si agisce di solito in ambiti limitati, sia temporalmente che come dimensione campionaria, e ciò rende più difficile generalizzare i risultati, specialmente nel caso di politiche di aiuto implementate su scale maggiori.

Per un'ampia rassegna dei lavori in ambito sperimentale (limitato inizialmente agli Stati Uniti e progressivamente esteso a Canada ed Europa settentrionale) si veda Greenberg, Links e Mandell (2003) e Greenberg e Shroder (1997). Duflo e Hanna (2006) propongono un'applicazione sperimentale agli insegnanti di scuole indiane incentivate al fine di ridurre l'assenteismo, mentre uno studio di maggior portata è quello di Schulz (2001), che analizza l'impatto di sussidi scolastici su circa due milioni di famiglie messicane.

La maggior parte dei dati disponibili per la valutazione d'impatto dei programmi di aiuto deriva da studi di tipo osservazionale, dove, in generale, non è noto il meccanismo che ha portato a osservare l'assegnazione al trattamento. Nel prossimo paragrafo vedremo come sotto specifiche ipotesi sia possibile riprodurre, almeno parzialmente, situazioni simili a quelle che si osservano nei disegni appena descritti.

¹¹ Per un approfondimento di questi aspetti, si veda Rossi *et al.* (2003).

14.4. Metodi non sperimentali

Quando il meccanismo di assegnazione che ha generato il vettore T non è noto, si è nell'ambito di uno studio non sperimentale¹².

Il problema principale, in questo ambito di analisi, è quello, già precedentemente descritto, del *selection bias*: le unità esposte al trattamento potrebbero essere diverse da quelle non esposte non per fattori puramente casuali, ma per caratteristiche intrinseche possibilmente correlate alla variabile risposta. Da ciò si deriva che, in generale

$$E(\bar{Y}_1^{obs} - \bar{Y}_0^{obs}) \neq \frac{1}{N} \sum_{i=1}^N (Y_i(1) - Y_i(0))$$

Per ovviare a questo problema intrinseco degli studi di tipo osservazionale, sono state sviluppate, negli ultimi decenni, diverse metodologie.

La maggior parte dei lavori sulla valutazione dell'impatto dei programmi di aiuto si colloca in un contesto osservazionale. Il confronto tra trattati e non trattati può essere allora condotto qualora:

- sia plausibile formulare appropriate ipotesi circa il meccanismo di assegnazione al trattamento;
- siano disponibili sufficienti dati statistici per condurre l'analisi e supportare le ipotesi;
- siano utilizzati appropriati strumenti statistici per tale confronto, che tengano conto delle differenze tra i due gruppi.

Un modo per risolvere il problema di identificabilità prima richiamato (l'impossibilità di osservare entrambi i risultati potenziali su uno stesso individuo) in ambito non sperimentale è quello di cercare di ricondurre lo studio osservazionale a un ipotetico esperimento randomizzato caratterizzato da probabilità di assegnazione al trattamento che variano in funzione dalle covariate (come, ad esempio, nell'esperimento randomizzato a blocchi).

A tal fine un'ulteriore assunzione deve essere fatta. Si tratta dell'assunzione di *assenza di confondimento* (Rosenbaum e Rubin, 1983a), la quale richiede che condizionatamente alle variabili osservate X precedenti il trattamento, l'assegnazione al trattamento sia indipendente dai risultati potenziali:

$$T \perp Y(0), Y(1) | X$$

L'assunzione di assenza di confondimento permette di far inferenza causale in ambito osservazionale utilizzando i risultati che caratterizzano uno studio randomizzato a blocchi. L'assunzione appena vista, infatti, implica che gli effetti causali medi introdotti nel paragrafo 2.2 possano essere identificati attraverso la stima della quantità:

¹² In letteratura si fa spesso riferimento a questo ambito parlando di metodi "quasi sperimentali", dal lavoro di Campbell e Stanley (1966).

$$E[Y(1)|T = 1, X] - E[Y(0)|T = 0, X]$$

Richiedere che il meccanismo di assegnazione soddisfi la condizione di assenza di confondimento significa assumere che anche se unità con caratteristiche diverse possono avere una diversa propensione a ricevere il trattamento (l'aiuto), e tali caratteristiche possono essere associate ai risultati potenziali, le unità che hanno le stesse caratteristiche osservabili hanno la stessa probabilità di essere trattate. I fattori (anche non osservabili) che fanno sì che alcune unità effettivamente vi partecipino, mentre altre no, assumono la stessa valenza del meccanismo di estrazione casuale di uno studio randomizzato, e si assumono dunque indipendenti dai risultati potenziali.

Assumere però, senza poterla verificare empiricamente, la conoscenza della totalità delle variabili che hanno portato una certa unità a essere assegnata o meno al trattamento può essere, in molte situazioni, un'ipotesi soggetta a critiche giustificate. In alcuni studi (Connors *et al.*, 1996) si è tentato di approfondire quali variabili potessero verosimilmente essere utilizzate per sostenere l'assenza di confondimento. Tuttavia tale assunzione, anche se non sempre sostenibile, data la potenziale presenza di caratteristiche non osservabili nei due gruppi, può essere comunque un buon punto di partenza per studi di valutazione in contesti osservazionali: ovviamente l'ipotesi sarà tanto più credibile e conforme alla realtà quanto più ricco è l'insieme di variabili osservabili di cui si può disporre.

Se da un lato, però, disporre di un ampio set di covariate è una condizione necessaria per poter sostenere ragionevolmente l'ipotesi di assenza di confondimento, dall'altro la stima degli effetti causali di interesse può non essere semplice, proprio a causa della presenza di un numero troppo elevato di variabili. All'aumentare delle colonne della matrice delle covariate X , cioè, è sempre più difficile trovare righe con identiche modalità.

In questo caso diventa indispensabile uno strumento che permetta di ridurre la dimensionalità del problema, senza perdere però i benefici dati da un insieme elevato di informazioni aggiuntive. Rosenbaum e Rubin (1983a) suggeriscono, a tal proposito, di utilizzare una funzione unidimensionale delle covariate, il cosiddetto *score* di propensione (o *propensity score*) $e(X)$, definito come la probabilità condizionata di ricevere il trattamento date le covariate osservate:

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

Rosenbaum e Rubin (1983a) introducono anche il concetto di *balancing score* $b(X)$, definito come una funzione di X tale che l'ipotesi di assenza di confondimento sia valida anche condizionatamente a questa funzione.

$$T \perp Y(0), Y(1) \mid b(X)$$

Il *propensity score*, in questo senso, può essere visto come il più grossolano dei *balancing score*, cioè come funzione di uno qualunque dei possibili *balancing score*. Sotto l'ipotesi di assenza di confondimento, dunque, condizionarsi al *propensity score* equivale a condizionarsi a tutto il set di covariate X . Valgono cioè le relazioni:

$$T \perp Y(0), Y(1) \mid e(X)$$

e

$$0 < P(T = 1|e(X)) < 1$$

Utilizzando una funzione univariata della i -esima riga della matrice X si risolve il problema, appena accennato, di trovare unità trattate e non trattate che abbiano le stesse modalità delle variabili X . Più precisamente, il problema riduce drasticamente la sua dimensionalità da k dimensioni (le colonne di X) a una sola (il valore del *propensity score*)¹³.

La proprietà di bilanciamento implica che unità con lo stesso valore del *propensity score* siano caratterizzate dalla stessa distribuzione delle variabili precedenti il trattamento. La seconda proprietà del *propensity score* afferma che se l'esposizione al trattamento all'interno delle celle definite dai valori delle covariate X può essere assimilabile a un esperimento randomizzato, allora può esserlo anche all'interno delle celle definite dai valori della variabile unidimensionale $e(X)$.

In altre parole, per un dato valore del *propensity score*, o più generalmente in un intorno di questo valore, l'aver ricevuto un trattamento è un evento casuale. La procedura di abbinamento delle unità in base al *propensity score* riproduce così le condizioni ideali di uno studio randomizzato a blocchi accennato nel paragrafo 3, bilanciando tutte le covariate osservate tra i gruppi di unità definiti dallo status di trattamento.

Se il *propensity score* è noto per ogni unità i , gli effetti ATE e ATT possono essere quindi stimati utilizzando, rispettivamente, le seguenti uguaglianze:

$$ATE = E_{e(X)}[E[Y(1)|T = 1, e(X)] - E[Y(0)|T = 0, e(X)]]$$

e

$$ATT = E_{e(X)|T=1}[E[Y(1)|T = 1, e(X)] - E[Y(0)|T = 1, e(X)]]$$

dove il simbolo a pedice del primo termine sta a indicare che il valore atteso è rispetto alla distribuzione di $e(X)$.

Condizionarsi al valore del *propensity score* (che, ricordiamo, non è influenzato per definizione dal trattamento, essendo una funzione di sole variabili misurate in un istante precedente il trattamento) bilancia tra i gruppi le covariate osservate X , ma in generale non garantisce un uguale bilanciamento (a differenza dei disegni randomizzati) in presenza di eventuali covariate latenti. Per poter dunque sostenere l'ipotesi di assenza di confondimento condizionatamente al valore del *propensity score*, si deve ipotizzare che non ci siano covariate non osservate che influenzano l'assegnazione al trattamento o che, se esistono, siano correlate con le X

¹³ In realtà il *propensity score*, è, per sua natura, una grandezza continua (se tra le variabili di X c'è almeno una variabile continua). Può quindi essere molto difficile ottenere unità con uguali valori di $e(X)$. Solitamente si ovvia a questo problema facendo ricorso a intorni dei valori dei *propensity score* osservati.

in modo da raggiungere un bilanciamento implicito come conseguenza del bilanciamento di queste ultime¹⁴ (si veda, a questo proposito, Rosenbaum, 1984).

In uno studio randomizzato il valore del *propensity score* è banalmente, per ogni unità, la probabilità assegnata a quella unità per far parte del gruppo delle unità trattate. In uno studio osservazionale, invece, la forma funzionale del *propensity score* non è nota, e deve essere dunque stimata attraverso un modello di probabilità. I modelli più utilizzati sono quello di regressione logistica e il modello probit. In tali casi,

$$P(T = 1|X) = F(h(X))$$

dove $F(\cdot)$ è la funzione di ripartizione logistica o normale, e $h(X)$ è una funzione delle covariate che deve essere specificata in modo tale da ottenere una stima del *propensity score* che soddisfi la proprietà di bilanciamento. In questa fase dell'analisi, l'obiettivo principale è, di solito, il bilanciamento delle covariate nei gruppi, mentre la bontà di adattamento del modello è un problema secondario.

Stimato il *propensity score* per ogni unità, e controllato che l'ipotesi di bilanciamento sia soddisfatta, si può procedere alla costruzione di un gruppo di controllo di unità del tutto simili alle unità beneficiarie, ma che non sono state trattate. La selezione del gruppo di controllo avviene in base al valore di $e(X)$ stimato.

14.4.1. Procedure di abbinamento

Le procedure di abbinamento (*matching*) delle unità (Rubin, 1973; Rubin e Thomas, 1992; Rubin e Thomas, 1996; Abadie e Imbens, 2002) consistono nell'abbinare ad ogni unità trattata uno o più unità non trattate. Queste procedure hanno lo scopo di ricostruire, per ogni unità, il risultato potenziale non osservato.

I vari metodi illustrati in questo paragrafo hanno quindi l'obiettivo comune di stimare effetti causali individuando specifiche unità che siano "simili" alle trattate, e ricostruendo verosimilmente una situazione che, per il problema fondamentale dell'inferenza causale, non si è potuta osservare.

Supponiamo di disporre, per una certa popolazione di numerosità N , della matrice X e dei due vettori N -dimensionali Y^{obs} e T , che rappresentano, rispettivamente, una matrice di covariate, la variabile risposta osservata e l'assegnazione al trattamento.

Se si è interessati, ad esempio, a un effetto medio sui trattati (ATT), l'obiettivo di una procedura di *matching* è quello di ricostruire, per ogni unità, il valore $Y(0)$ che si sarebbe osservato in assenza di trattamento.

Le principali tecniche che si basano sulla matrice X senza affidarsi al calcolo di una funzione *propensity score* sono quelle di *matching esatto*, *caliper* e *nearest neighbor*.

¹⁴ Si noti, però, che queste considerazioni varrebbero anche se si utilizzasse l'intera matrice X invece del solo *propensity score*.

Il *matching esatto* si basa sulla ricerca, per ogni unità trattata di un'unità di controllo con le stesse modalità delle variabili della matrice X . Anche in presenza di un ampio gruppo di unità di controllo, e di covariate discrete (per una variabile continua X_k la probabilità di trovare due unità con lo stesso valore di x_k è in linea di principio nulla), questa tecnica soffre di un inevitabile problema. Come accennato in precedenza, questo metodo ha il suo principale limite proprio nella ricchezza del dataset disponibile. All'aumentare delle k colonne della matrice X , infatti, la probabilità di trovare un'unità con le stesse realizzazioni delle k covariate si riduce drasticamente (per approfondimenti su questo aspetto si veda Rosenbaum e Rubin, 1983b; Rosenbaum, 1984).

Una tecnica che tenta di attenuare questa problematica è quella del *caliper matching*, del quale il *matching esatto* può essere considerato un caso particolare. In questo tipo di abbinamento si utilizzano unità di controllo con valori vicini a quelli delle unità trattate. Questa tecnica, pur rendendo meno marcato il problema accennato, introduce però una componente arbitraria, quella relativa alla soglia di tolleranza nella differenza tra modalità delle covariate.

Un'altra tecnica che non si basa sul calcolo di un *propensity score* (pur condividendo con esso la caratteristica di un tentativo di riduzione della dimensione delle covariate) è quella del *nearest neighbor*. L'idea di base è quella di ridurre la dimensionalità spesso troppo elevata del set di covariate tramite una funzione distanza. Scelta una metrica, si può infatti calcolare la distanza tra due vettori k -dimensionali di covariate e abbinare ad ogni unità trattata il controllo più vicino. Le metriche più utilizzate sono quelle basate sulla distanza euclidea, sulla distanza di Mahalanobis (Rubin, 1976; Rubin, 1980) o sulla diagonale della matrice di varianza e covarianza (Abadie e Imbens, 2002). Zhao (2004) propone due nuove metriche che includono informazioni sulla variabile di risposta e sull'indicatore di trattamento.

Procedure di abbinamento basate su propensity score

Vi sono poi numerose tecniche che si basano sul valore del *propensity score* stimato allo scopo di ricostruire, tramite risultati delle unità di controllo, i risultati potenziali non osservati delle unità trattate. Le tecniche principali sono la *stratificazione*, il *kernel matching*, il *radius matching* e il già accennato *nearest neighbor*.

La procedura di *stratificazione* si basa sulla costruzione di sottoinsiemi (detti anche "blocchi" o "strati") di unità, trattate e non, in modo che, in ciascuno di essi, il *propensity score* non sia, in media, differente nel gruppo di trattati e controlli (Dehejia e Wahba, 1999).

L'effetto causale è poi calcolato come una media ponderata, tra i vari blocchi, delle differenze medie tra unità trattate e di controllo. Tramite questa metodologia si tenta di riprodurre la replica di mini esperimenti randomizzati all'interno di ciascun blocco. Dehejia e Wahba (1999) dimostrano anche che 5 gruppi sono quasi sempre sufficienti a rimuovere il 95% della distorsione.

La metodologia del *kernel matching* utilizza, per ogni unità trattata, l'intero gruppo dei non trattati per ricostruire ciascun risultato potenziale non osservato. L'idea di fondo è quella di assegnare un peso alle unità non trattate in modo da privilegiare le unità più simili alle tratta-

te. La variabile utilizzata per misurare la distanza tra unità è il *propensity score*. Il sistema di pesi utilizzato è del tipo:

$$w(j) = \frac{K_j}{\sum_{j \in I_0} K_j}$$

con

$$K_j = \frac{1}{h} K\left(\frac{e(X_j) - e(X_i)}{h}\right)$$

dove $K(\cdot)$ è una determinata funzione di densità¹⁵, h è il parametro di banda, e I_0 è l'insieme delle unità di controllo non trattati.

I due punti deboli di questa metodologia risiedono nella scelta soggettiva del parametro di banda e della funzione *kernel*, anche se Dehejia e Wahba (2002) dimostrano che la scelta di quest'ultima non è fondamentale.

La tecnica del *nearest neighbor* è quasi la stessa quella descritta tra le metodologie che non utilizzano la matrice X , con la ricerca della(e) unità più vicina(e) effettuata utilizzando il *propensity score*. L'insieme A_i dei controlli abbinati è quindi

$$A_i = \{j \in I_0 : \text{Min}_j(\|e(X_i) - e(X_j)\|)\}$$

Questa procedura può prevedere la scelta di più di un vicino per ciascun trattato, col numero di vicini che deve essere comunque stabilito a priori. Questa tecnica espone al rischio, nel caso si abbiano unità trattate con valori del *propensity score* anomali, di abbinamenti con unità piuttosto distanti. Un altro limite evidenziato recentemente da un lavoro di Abadie e Imbens (2006) è relativo alla procedura *bootstrap* (Efron e Tibishirani, 1993) per il calcolo degli errori standard delle stime degli effetti ottenute da questa metodologia. A causa della irregolarità della procedura di *nearest neighbor*, le condizioni standard per la stima *bootstrap* non sono soddisfatte, e la stima della varianza diverge dalla vera varianza.

Nella metodologia del *radius matching* (abbinamento a raggio), a differenza di quanto appena visto per il *nearest neighbor*, il numero di controlli da impiegare per la ricostruzione di $Y(0)$ non viene stabilito a priori, ma si utilizzano tutte le unità il cui *propensity score* è compreso in un prefissato intervallo (Dehejia e Wahba, 2002). Le unità abbinare sono date da

$$A_i = \{j \in I_0 : e(X_i) - c \leq e(X_j) \leq e(X_i) + c\}$$

Uno dei vantaggi di questa metodologia è quello di non dover stabilire preventivamente il numero di controlli da utilizzare per ciascun trattato, ed evitare quindi pericolose forzature. Si può dunque utilizzare un maggior numero di controlli (nel caso questi siano caratterizzati

¹⁵ Nel caso, ad esempio, di *kernel* gaussiano, si ha: $K_j = \frac{1}{h} \left[\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \left(\frac{e(X_j) - e(X_i)}{h}\right)^2\right\} \right]$

da valori del *propensity score* molto vicini a quello del trattato) e un numero inferiore per le unità con valori del *propensity score* più estremi. Analogamente a quanto visto per il *kernel matching*, anche questa tecnica ha però il suo punto debole nella necessità di dover stabilire a priori, in modo soggettivo, il valore c del raggio.

Il fatto che esistano molte tecniche per l'abbinamento delle unità di controllo induce una domanda banale: qual è la tecnica più appropriata?

A livello asintotico queste scelte finiscono per abbinare tra loro solo unità esattamente uguali, e quindi forniscono le stesse stime, ma in campioni limitati scegliere un metodo piuttosto che un altro porta quasi sempre a una differenza, talvolta significativa. Per le proprietà asintotiche di questi stimatori si veda Abadie e Imbens (2006).

Posto che nessuna delle tecniche menzionate è a priori migliore delle altre, la tendenza della letteratura, in generale, è quella di scegliere, in campioni piccoli, metodi che permettano la reintroduzione delle unità di controllo (che quindi possono apparire più volte nell'insieme dei controlli utilizzato). Così facendo però, pur guadagnando in correttezza, si perde qualcosa in efficienza degli stimatori, e non è sempre semplice scegliere il giusto compromesso tra distorsione e variabilità delle stime. Se invece si dispone di un dataset di controlli sufficientemente ampio, la scelta più sensata sembra essere quella del *nearest neighbor*, che sfrutta al meglio la ricchezza delle unità di controllo per ottenere stime efficienti, mentre *kernel* e *radius matching* sembrano essere indicati per situazioni dove il dataset, seppur ricco, presenta delle asimmetrie.

Il metodo di stratificazione è particolarmente utile se c'è il sospetto di effetti non osservati nel *matching*: siccome la stratificazione raggruppa osservazioni simili, eventuali effetti latenti tendono ad attenuarsi.

Alcuni autori hanno tentato di riassumere alcune linee guida per la scelta del metodo di abbinamento. Alcune indicazioni si possono trovare in Zhao (2004), che propone sia un interessante confronto tra metodi di abbinamento basati su covariate e su *propensity score*, sia una serie di indicazioni, ricavate da simulazioni, su come scegliere tra i vari metodi e metriche. Baser (2006), propone un algoritmo per la scelta del metodo più appropriato basato su una serie di cinque criteri.

Questi metodi di stima basati su procedure di matching non presuppongono nessun modello per i dati. Ovviamente, modellare la distribuzione congiunta $P(Y(1), Y(0), X)$, apre la possibilità ad estensioni metodologiche per la stima dell'effetto causale di interesse. Oltre alla imponente letteratura econometrica sull'argomento (Goldberger, 1972; Heckman, 1978, solo per citare alcuni lavori chiave) Rubin (1973), Reinisch *et al.* (1996) e Rubin e Thomas (2000) dimostrano come si possano ottenere stime migliori condizionatamente all'uso di modelli probabilistici.

Tra le valutazioni d'impatto che utilizzano il matching basato su *propensity score*, ricordiamo il lavoro di Jalan e Glinskaya (2005), applicato ai programmi di educazione elementare in India, o quello dell'*Independent Evaluation Group* della World Bank (2005) relativamente al progetto di nutrizione integrata in Bangladesh. Più recentemente, Mauro *et al.* (2010) applicano questa la metodologia per la valutazione dell'impatto sulla disabilità dei *Community-Based Rehabilitation Programs* nel distretto di Mandya (India).

14.4.2. Il metodo *difference in differences*

Una metodologia largamente utilizzata per la stima di effetti causali è quella delle *difference in differences* (DID) (Ashenfelter, 1978; Boarnet e Bogart 1995; Papke 1994; Moffit 1991; Heckman e Robb 1985; Wooldridge 2002). Questa tecnica si basa sulla disponibilità di più osservazioni (dati panel) della variabile risultato Y antecedenti l'istante t di realizzazione del trattamento, che vengono sfruttate per depurare l'effetto stimato dalle differenze sistematiche tra gruppi di trattati e controlli tipici del *selection bias*.

Proviamo ad illustrare le basi di questa metodologia con un semplice esempio, in cui si dispone di due soli istanti temporali, t e $t-1$.

La quantità oggetto di interesse

$$(1) \text{ ATT} = E[Y_t(1) - Y_{t-1}|T = 1] - E[Y_t(0) - Y_{t-1}|T = 1]$$

non è direttamente stimabile perché il termine $Y_t(0)$, cioè il valore della variabile risultato nel caso in cui l'unità non avesse ricevuto il trattamento, non può essere osservato. È invece possibile stimare

$$E[Y_t(1) - Y_{t-1}|T = 1] - E[Y_t(0) - Y_{t-1}|T = 0]$$

che è uguale alla quantità (1) nel caso in cui il valore atteso della variazione della variabile risultato registrata nel gruppo delle unità di controllo coincida con la variazione controfattuale delle unità trattate nel medesimo periodo.

Meno formalmente, per ottenere una stima *difference in differences* si sottrae dalla differenza al tempo t la differenza¹⁶ al tempo $t-1$. Nell'ipotesi che tra i due gruppi ci siano solo differenze di livello e non di trend, la distorsione da autoselezione coincide con la differenza al tempo $t-1$ e può quindi essere rimossa.

Nel caso in cui si disponga di più osservazioni temporali ($t-1, t-2... t-k$), queste possono essere utilizzate per stimare il trend pretrattamento dei due gruppi per alleggerire l'ipotesi di uguaglianza del coefficiente angolare (parallelismo), inevitabile nell'esempio proposto.

È chiaro che, la bontà della stima DID dipende fortemente dalle assunzioni che si possono fare sulla differenza tra trend nei due gruppi. Come per tutte le metodologie che tentano di fornire stime degli effetti causali in ambito non sperimentale, questa assunzione non è testabile, a meno di disporre di altri dati.

Lo stimatore DID, ampiamente utilizzato nella letteratura econometrica, è stato utilizzato più recentemente anche nella valutazione di politiche pubbliche (ad esempio, Adorno *et al.*, 2007). Questo stimatore sfrutta la rilevazione delle unità in almeno due istanti temporali per ridurre la distorsione dovuta ad eventuali effetti invarianti nel tempo (si veda anche Deheja e Wahba, 1999; Smith e Todd, 2005). Recentemente (Athey e Imbens, 2006) ne è stata propo-

¹⁶ Si calcola cioè una differenza tra due quantità ottenute come singole differenze. Da qua il nome "differenza tra differenze".

sta un'interessante generalizzazione in ambito non parametrico che permette ai due gruppi (trattati e controlli) di avere differenti effetti dell'intervento. Elbers *et al.* (2009) propongono un'estensione dello stimatore DID applicata a trattamenti eterogenei nello studio di programmi di aiuto per l'istruzione nello Zambia, mentre Todd e Wolpin (2009) utilizzano questa tecnica per l'analisi di molteplici programmi d'aiuto in paesi in via di sviluppo.

14.4.3. Regression discontinuity design

In ambito econometrico (Heckman, 1978; Heckman e Hotz, 1989; Angrist *et al.*, 1996) sono stati tentati approcci che permettessero di evitare l'assunzione di assenza di confondimento introdotta nel paragrafo 4. Come si vedrà più avanti, in questo ambito vi è però un elemento di complicazione dovuto al fatto che è spesso necessario identificare alcune variabili di tipo strumentale, che non abbiano cioè effetto diretto sulla variabile risultato e che invece influenzino la probabilità di ricevere il trattamento.

Un'altra interessante situazione, che può consentire l'identificazione degli effetti del trattamento senza dover ipotizzare l'assenza di confondimento, si ha quando l'esposizione al trattamento varia in modo discontinuo come funzione di una certa variabile osservabile. Si può dunque supporre che in un intorno del punto di discontinuità le unità trattate e non siano, in qualche senso, "simili". È questo il caso del *Regression Discontinuity Design* (RDD) (Thistlethwaite e Campbell, 1960; Hahn *et al.*, 2000; Trochim, 2001), un disegno non sperimentale che sfrutta questa similarità tra unità per ipotizzare che, nel sottoinsieme di unità prossime alla soglia, l'assegnazione del trattamento sia avvenuta in modo casuale. Quest'approccio si divide in due tipologie leggermente differenti, a seconda della dipendenza dell'esposizione al trattamento alla variabile osservabile. Se la dipendenza è di tipo deterministico (nel senso che la variabile osservabile determina univocamente l'assegnazione ai trattati o ai controlli) si parla di *sharp* RDD, mentre se la variabile influenza solo la probabilità di ricevere il trattamento si parla di *fuzzy* RDD (Trochim, 1984).

Più formalmente, se S è la variabile osservabile dalla quale dipende, in qualsivoglia misura, l'assegnazione al trattamento, e s è il punto del dominio di S dove si ha la discontinuità nella probabilità di assegnazione al trattamento, allora si ha

$$\lim_{S \rightarrow s^+} P(T = 1) \neq \lim_{S \rightarrow s^-} P(T = 1)$$

Se questi due limiti esistono finiti, senza perdere in generalità, si può scrivere la precedente equazione come

$$\lim_{S \rightarrow s^+} P(T = 1) - \lim_{S \rightarrow s^-} P(T = 1) > a \quad 0 < a \leq 1$$

dove, se $a=1$ si ha il caso di *sharp* RDD e per $a<1$ di *fuzzy* RDD.

Restringendo l'analisi allo *sharp* RDD, il risultato notevole è che (sotto l'ipotesi che, nel caso a nessuna unità fosse stato assegnato il trattamento, non vi sarebbe stata discontinuità nel punto di soglia s), definito

$$\beta = Y(1) - Y(0)$$

come l'impatto per la singola unità, e

$$E(\beta) = E(Y(1)|T = 1) - E(Y(0)|T = 0)$$

come il suo valore atteso, si può identificare l'impatto medio dell'intervento su un sottogruppo di imprese trattate (appartenenti dunque a un intorno destro di s) come

$$E(\beta|S \in s_+) = E(Y|S \in s_+) - E(Y|S \in s_-)$$

dove si è indicato con s_+ (s_-) un intorno destro (sinistro) del punto di discontinuità s .

Intuitivamente, in un intorno di s valgono dunque le stesse condizioni di un disegno sperimentale (Cook e Campbell, 1979).

Un esempio tipico in cui questo tipo di approccio può essere utilizzato è quello in cui l'assegnazione del trattamento sia determinato in base a una graduatoria. Ad esempio, nel caso di una borsa di studio assegnata sulla base di un punteggio che deriva da un set di variabili, si può ipotizzare che gli studenti al di sotto della soglia minima per ottenere la borsa non presentino differenze sistematiche (precedenti il beneficio) da quelli appena sopra la soglia.

Un'interessante applicazione di questa metodologia a programmi di aiuto si ha in Pitt e Shaidur (1998), che analizzano programmi di microcredito alle famiglie in Bangladesh. Urquiola (2006) applica una stima RDD in scuole boliviane per misurare l'impatto di un maggior numero di insegnanti per alunno, mentre Edmonds (2004) la utilizza per analizzare l'incidenza del lavoro minorile in Sudafrica. Per una rassegna più ampia di lavori in ambito economico, si veda Lee e Lemieux (2009).

14.4.4. L'uso di variabili strumentali

L'approccio basato su variabili strumentali fa parte, nella letteratura econometrica, del più ampio ambito dei modelli ad equazioni strutturali (SEM), che prende le mosse dai lavori di Wright (1928 e 1934) di Haavelmo (1943, 1944), e, più recentemente, di Goldberger (1972) e Morgan (1990). I SEM, infatti, utilizzano spesso variabili (dette, appunto, strumentali) che sono incluse solo in alcune delle equazioni strutturali, e omesse in altre, con il risultato che la loro correlazione con alcune delle variabili risultato si esplicita solo attraverso l'effetto esercitato su variabili intermedie. Il semplice modello utilizzato da Angrist, Imbens e Rubin (AIR, 1996) per introdurre la possibilità di conciliare l'approccio IV con quello del RCM è del tipo

$$(2) \quad Y = \beta_0 + \beta_1 \cdot T_i + \epsilon_i$$

$$(3) \quad \bar{T}_i = \alpha_0 + \alpha_1 \cdot Z_i + u_i$$

con

$$T_i = 1 \text{ se } \bar{T}_i \geq 0 \quad e \quad T_i = 0 \text{ se } \bar{T}_i < 0$$

Con Y variabile risposta e T e Z dicotomiche a valori 0 e 1.

Nella letteratura econometrica si ipotizza solitamente che la variabile Z sia incorrelata con i termini di errore ε_i e u_i ; il suo effetto sulla variabile risposta Y è determinato dunque solo dall'effetto indiretto esercitato su T . Se il parametro α_1 nell'equazione (3) è diverso da 0 (cioè se non è nulla la correlazione tra T e Z), la Z è definita una variabile strumentale.

L'effetto causale di T su Y , cioè il parametro β_1 (o β_{IV} , Instrumental Variable) dell'equazione (2) può quindi essere stimato (in questo modello) tramite il rapporto delle covarianze tra Y e Z e tra T e Z

$$(4) \quad \hat{\beta}_{IV} = \frac{\widehat{cov}(Y, Z)}{\widehat{cov}(T, Z)}$$

Nel loro lavoro, AIR cercano di riportare al contesto dei risultati potenziali le assunzioni tipiche dell'approccio econometrico. Supponiamo di avere, ancora, una variabile Y , risultato, un trattamento T dicotomico e una variabile Z , anch'essa dicotomica, avente il ruolo di variabile strumentale.

Oltre all'ipotesi SUTVA (si veda paragrafo 2.1) e all'ipotesi di assegnazione casuale del valore (dicotomico) della variabile Z a tutte le unità, AIR focalizzano su altre due assunzioni fondamentali¹⁷ per poter definire in maniera formalmente corretta una variabile strumentale in ambito del RCM.

La prima assunzione, detta *vincolo di esclusione*, impone che ogni effetto di Z su Y passi attraverso un effetto di Z su T . Fissato il valore $T_i=t$, quindi, le quantità $Y(t, Z=0)$ e $Y(t, Z=1)$ sono uguali.

La seconda assunzione detta di *monotonicità forte* (Imbens e Angrist, 1994), assume che si abbia, per tutte le unità

$$T_i(Z=1) \geq T_i(Z=0) \quad \forall i$$

e che per almeno un'unità i sia $T_i(Z=1) > T_i(Z=0)$ ¹⁸

Se valgono queste due assunzioni, unitamente all'ipotesi SUTVA e all'ipotesi di assegnazione casuale dello strumento, allora Z è una variabile strumentale per l'effetto del trattamento T sulla variabile risposta Y .

Anche se, come si è visto, i programmi di aiuto devono quasi sempre tener conto di dati osservazionali, a volte può esistere un fattore esterno che influenza il processo di selezione (si pensi ad esempio a un programma di aiuto che viene pubblicizzato in modo differente in diverse aree geografiche). Se questo fattore esterno non ha influenza diretta sulla variabile di

¹⁷ In realtà, nel paper di AIR le assunzioni sono altre tre, ma due di esse possono esser riunite in una sola.

¹⁸ La seconda assunzione impone implicitamente che Z abbia un effetto causale su T non nullo. Che si abbia, cioè, $E[T_i(Z=1) - T_i(Z=0)] \neq 0$.

risultato, può essere utilizzato in alternativa ai metodi introdotti nei paragrafi precedenti per stimare un effetto causale.

Il fattore esterno (o “variabile strumentale” o “strumento”) viene utilizzato per scomporre la variabile di risultato in due parti: una relativa alle decisioni e alle preferenze dell’individuo (e quindi affetta da distorsione da autoselezione), e una dovuta al fattore stesso (e dunque più assimilabile a uno studio sperimentale). Un caso estremo di strumento è il caso visto nel paragrafo 3 relativo alla randomizzazione, tecnica che non solo influenza il processo decisionale, ma in linea teorica dovrebbe determinare l’assegnazione di ogni unità al gruppo di trattati e di controlli. In pratica, però, questo non accade, in quanto non sempre le unità (ad esempio, un gruppo di individui) obbediscono esattamente all’assegnazione imposta loro.

Aldilà di questo caso estremo, spesso si può considerare variabile strumentale come un incentivo (assegnato casualmente) alla partecipazione al trattamento. Un esempio nell’ambito dei programmi di aiuto potrebbe essere dato da una politica pubblicizzata maggiormente in determinate aree scelte a caso, e meno in altre. Le unità presenti nell’area pubblicizzata tenderanno a scegliere di essere trattate non solo per caratteristiche endogene (processo che distorcerebbe le stime finali se non adeguatamente considerato), ma anche per la spinta, questa assegnata casualmente, data dalla pubblicità. È ovvio che la variabile strumentale, oltre ad essere valida, cioè non correlata con i risultati potenziali, deve anche essere sufficientemente rilevante, cioè significativamente correlata¹⁹ con il trattamento. Nel caso in cui quest’ultima correlazione sia perfetta, ci si riconduce al caso di uno studio sperimentale accennato all’inizio di questo paragrafo.

Essendo impossibile riassumere in questo breve paragrafo l’imponente letteratura statistica ed econometrica sull’utilizzo di variabili strumentali, ci limitiamo ad introdurre uno stimatore IV adatto a trattamenti binari. Definita ancora Z la variabile strumentale che assume valore uno se lo strumento è assegnato all’unità i e zero altrimenti

$$\delta^{IV} = \frac{E(Y|Z=1) - E(Y|Z=0)}{P(T|Z=1) - P(T|Z=0)}$$

è la stima da variabili strumentali. Come si vede, essa deriva dal rapporto tra due differenze. Tornando all’esempio della politica pubblicizzata ($Z=1$) in certe aree e non pubblicizzata ($Z=0$) in altre, al numeratore abbiamo la differenza della variabile di risultato tra le due aree, mentre al denominatore la differenza tra la probabilità di esser trattati, sempre nelle due aree. Se T e Z coincidono (cioè se si è in uno studio sperimentale), il denominatore vale uno, e la stima si riduce alla differenza tra medie nel gruppo di trattati e controlli tipica degli studi basati su randomizzazione.

Nell’ambito dei risultati potenziali, l’approccio IV è stato analizzato da Angrist, Imbens e Rubin (1996) in ottica frequentista, e da Imbens e Rubin (1997) in ambito bayesiano, anche se precedenti lavori, come quelli di Hearst *et al.* (1986), Holland (1988), Permutt e Hebel (1989) Sommer e Zeger (1991), Imbens e Angrist (1994) avevano già tentato di combinare

¹⁹ Una delle difficoltà maggiori nell’ambito di stima delle variabili strumentali risiede proprio nel trovare uno strumento che soddisfi queste due condizioni di validità e rilevanza, dato che spesso la credibilità di una è inversamente proporzionale a quella dell’altra.

alcuni aspetti dei due ambiti (RCM e IV). Jeanty e Hitzhusen (2006) utilizzano lo stimatore IV applicato a dati panel per analizzare l'effetto di guerre civili e conflitti sui programmi alimentari nei paesi in via di sviluppo.

14.4.5. Altri metodi di stima

Oltre a quelli già visti, esistono altri metodi di stima per un effetto causale in ambito non sperimentale.

Un metodo largamente utilizzato per la sua semplicità è quello che utilizza il modello di regressione. Il punto di forza di questa tecnica risiede nel fatto che permette di controllare in modo relativamente semplice le variabili osservate che potrebbero essere sbilanciate all'interno dei due gruppi di unità trattate e di controllo²⁰. Ovviamente, nel caso vi siano variabili latenti che si distribuiscono in modo diverso nei due gruppi, il modello di regressione può essere soggetto alla distorsione da selezione, a meno di non ipotizzare che, condizionatamente al set di covariate usate come variabili esplicative, anche le variabili latenti siano bilanciate.

Nel caso di trattamento binario, utilizzare la variabile T come variabile esplicativa e interpretare il coefficiente angolare stimato del modello

$$(5) \quad Y = \beta_0 + \beta_1 \cdot T_i + \epsilon_i$$

come effetto netto del trattamento, equivale a calcolare la semplice differenza tra medie nei due gruppi definiti da T. Analogamente, esistono dei modelli di regressione multipla che, in modo più complesso di quello appena illustrato, riproducono i risultati ottenibili con lo stimatore DID introdotto nel paragrafo 4.2.

Altre tecniche di stima dell'impatto di una politica si possono basare sull'analisi in serie storiche, qualora ci sia la disponibilità di molti dati a livello temporale. Questa tecnica, detta "delle serie storiche interrotte", utilizza la letteratura statistica delle *time-series* per ricostruire la situazione controfattuale che si sarebbe osservata in assenza di trattamento.

Infine, un approccio talvolta utilizzato per ricostruire la situazione controfattuale, consiste nel rilevarla direttamente sui beneficiari, chiedendo loro di ricostruire soggettivamente la situazione in cui verosimilmente si sarebbero trovati nel caso in cui non avessero beneficiato del trattamento.

Pur essendo questa una tecnica particolarmente economica (non fosse altro che per non avere la necessità di rilevare dati su unità non trattate), le stime che se ne ricavano sono inevitabilmente soggette a grossi rischi di distorsione, in particolar modo quando si fa riferimento a periodi lontani nel tempo o di difficile ricostruzione. Questa tecnica, che si contrappone a quelle viste nei paragrafi precedenti perché non necessita di complesse elaborazioni statistiche, è solitamente utilizzata quando non è possibile condurre una raccolta dei dati rigorosa e l'unica strada percorribile sia dunque quella di utilizzare unicamente i dati ottenuti dai beneficiari.

²⁰ Spesso questa tecnica viene utilizzata anche nell'ambito di studi sperimentali, per eliminare differenze che potrebbero esistere nonostante la randomizzazione.

14.5. Riflessioni conclusive

In questo capitolo si è provato a delineare un quadro generale delle metodologie più comunemente adottate nella valutazione di politiche di aiuto. Come accennato nell'introduzione, le analisi presentate in questo capitolo non pretendono certo di offrire un quadro esaustivo della letteratura sull'argomento che, negli ultimi vent'anni, è cresciuta notevolmente, grazie anche alla convergenza tra approccio econometrico e statistico sotto il *framework*, ormai dominante, del RCM.

Vale infine la pena di ricordare alcuni punti chiave (comuni a quasi tutte le metodologie illustrate) messi in risalto da Heckman *et al.* (1999) riguardo alle procedure di valutazione delle politiche sociali.

Innanzitutto si deve prendere atto che, in molti casi, se i dati disponibili sono carenti è quasi impossibile poter identificare effetti causali credibili basandosi sulle metodologie statistiche. Nella valutazione dell'impatto di un intervento è verosimilmente più proficuo investire risorse per arricchire e migliorare i dati, piuttosto che chiedere ai metodi più di quanto possano dare, finendo per basare le proprie valutazioni su effetti identificati a costo di assunzioni poco credibili. Se ciò non è possibile, è più opportuno condurre un'analisi semplice, basata su ipotesi meno restrittive, piuttosto che chiedere alle metodologie di sopperire alla carenza dei dati.

In secondo luogo, laddove si utilizzi un gruppo di controllo, è fondamentale che i soggetti siano davvero "simili", e dunque confrontabili, con i soggetti trattati. Spesso vengono utilizzati modelli parametrici che, basandosi su forme funzionali arbitrarie, cercano di fare un raffronto tra soggetti aventi caratteristiche diverse. Sotto questo punto di vista gli stimatori di tipo non parametrico hanno almeno il vantaggio di doversi basare, per costruzione, su confronti tra unità effettivamente confrontabili.

Infine, l'impatto di un intervento sociale è spesso variabile tra soggetti diversi. Tra le varie problematiche che questa considerazione dischiude, ne sottolineiamo almeno due principali. La prima è relativa alle stime di impatti medi, che, almeno in presenza di alta variabilità, rendono i parametri identificati poco indicativi, esattamente come una semplice media aritmetica mal riassume una distribuzione molto variabile. La seconda osservazione, parzialmente legata alla prima, è legata alla possibilità di ottenere stime dell'effetto disomogenee fra loro in base allo stimatore scelto. In effetti, stimatori diversi, se non adeguatamente dettagliati, possono generare stime differenti solo perché relative a sottopopolazioni differenti.

Bibliografia

- Adams R. (2000), "Self-targeted Subsidies: The Distributional Impact of the Egyptian Food Subsidy System", *World Bank Policy Research Working Paper*, No. 2322, The World Bank, Washington D.C.
- Abadie A., Imbens G.W. (2002), "Simple and Bias Corrected Matching Estimators for Average Treatment Effects", *NBER working paper*.
- Abadie A., Imbens G.W. (2006), "Large Sample Properties of Matching Estimators for Average Treatment Effects", *Econometrica*, Vol. 74, No. 1, pp. 235-267.
- Adorno V., Bernini C., Pellegrini G. (2007), "The Impact of Capital Subsidies: New Estimations under Continuous Treatment", *Giornale degli Economisti*, Bocconi University, Vol. 66, No. 1, pp. 67-92.

- Angrist J.D., Imbens G.W., Rubin D.B. (1996), "Identification of Causal Effects using Instrumental Variables (with Discussion)", *Journal of the American Statistical Association*, Vol. 91, pp. 444-472.
- Ashenfelter O. (1978), "Estimating the Effect of Training Programs on Earnings", *Review of Economics and Statistics*, Vol. 60, pp. 47-57.
- Athey S., Imbens G.W. (2006), "Identification and Inference in Nonlinear Difference-in-Differences Models", *Econometrica*, Vol. 74, No. 2, pp. 431-497.
- Banerjee A., Duflo E., Cole S., Linden L. (2007), "Remedying Education: Evidence from Two Randomized Experiments in India," forthcoming, *Quarterly Journal of Economics*.
- Banerjee A., Duflo E. (2008), "The Experimental Approach to Development Economics", *NBER Working Paper*, No. 14467.
- Baser O. (2006), "Too much Ado about Propensity Score Models? Comparing Methods of Propensity Score Matching", *Value in Health*, Vol. 9, No. 6, pp. 377-385.
- Belligni S. (2003), *Cinque idee di politica*, Il Mulino, Bologna.
- Biggeri M., Volpi F. (2007), *Teoria e politica dell'aiuto allo sviluppo*, FrancoAngeli, Milano.
- Birkland T.A. (2001), *An Introduction to the Policy Process: Theories, Concepts, and Models of Public Policy Making*, M.E. Sharpe, Armonk, NY.
- Campbell D., Stanley J. (1966), *Experimental and Quasi-Experimental Design for Research*, Rand McNally, IL.
- Canitano G. (2005), "L'impatto dell'aiuto sulla crescita. Una rassegna della letteratura empirica cross-country", *Papers del Dottorato in Politica ed Economia dei Paesi in via di Sviluppo*, Università di Firenze, giugno, mimeo.
- Connors A.F., Speroff T., Dawson N.V., Harrell F.E., Wagner D. (1996), *The Effectiveness of Right Heart Catheterization in the Initial Care of Critically Ill Patients*. JAMA, 276 889-897
- Cook T.D., Campbell D.T. (1979), *Quasi-Experimentation: Design and Analysis for Field Settings*, Rand McNally, Chicago, Illinois.
- Cox D. R. (1958), *Planning of Experiment*, Wiley, New York.
- Davies R., Dart J. (2005), "The 'Most Significant Change' (MSC) Technique: A Guide to Its Use", *MandE*, Melbourne, Australia.
- Dehejia R., Wahba S. (1999), "Causal Effects in Non-experimental Studies: Re-evaluating the Evaluation of Training Programs", *Journal of the American Statistical Association*, Vol. 94, pp. 1053-1062.
- Dehejia R., Wahba S. (2002), "Propensity Score Matching Methods for Nonexperimental Causal Studies", *Review of Economics and Statistics*, Vol. 84, No. 1, pp. 151-161.
- Duflo E., Hanna R. (2006), "MonitoringWorks: Getting Teachers to Come to School", *NBER Working Paper*, No. 11880.
- Edmonds E. V. (2004), "Does Illiquidity Alter Child Labor and Schooling Decisions? Evidence from Household Responses to Anticipated Cash Transfers in South Africa", *NBER Working Paper*, No. 10265, National Bureau of Economic Research.
- Efron B., Tibshirani R. (1993), *An Introduction to the Bootstrap. Monograph in Statistics and Applied Probability*, Chapman and Hall, London.
- Elbers C., Gunning J.W. (2006), *Assessing Budget Support with Statistical Impact Evaluation*, Vrije Universiteit, Amsterdam.
- Elbers C., Gunning J.W., de Hoop K. (2009), "Assessing Sector-wide Programs with Statistical Impact Evaluation: A Methodological Proposal" *World Development*, Vol. 37, No. 2, pp. 513-520.
- Goldberger A. (1972), "Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations", mimeo, Madison, WI.
- Glewwe P., Park A., Zhao M. (2006) "The Impact of Eyeglasses on the Academic Performance of Primary School Students: Evidence from a Randomized Trial in Rural China", Conference Paper No. 6644, University of Minnesota, Center for International Food and Agricultural Policy.
- Greenberg D., Links D., Mandell M. (2003) *Social Experimentation and Public Policymaking*, The urban institute press, Washington D.C.
- Greenberg D., Morris S. (2005) "Large-Scale Social Experimentation in Britain. What Can and Cannot be Learnt from the Employment Retention and Advancement Demonstration?!", *Evaluation*, Vol. 11, No. 2.

- Greenberg D., Shroder M. (1997) *The Digest of Social Experiments*, The Urban Institute Press, Washington D.C.
- Haavelmo T. (1943), "The Statistical Implication of a System of Simultaneous Equations", *Econometrica*, Vol. 11, pp. 1-12.
- Haavelmo T. (1944), "The Probability Approach in Econometrics", *Econometrica*, Vol. 12, No. 1 (Supplement), pp. 15-24.
- Hahn J., Todd P., VanderKlaauw W. (2000), "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design", *Econometrica*, Vol. 69, No. 1, pp. 201-209.
- Hearst N., Newman T., Hulley S. (1986), "Delayed Effects of the Military Draft on Mortality: A Randomized Natural Experiment", *New England Journal of Medicine*, Vol. 3, No. 14, pp. 34-49.
- Heckman J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System", *Econometrica*, Vol. 46, pp. 931-959.
- Heckman J., Hotz J. (1989), "Alternative Methods for Evaluating the Impact of Training Programs", (with discussion), *Journal of the American Statistical Association*, Vol. 84, No. 804, pp. 862-874.
- Heckman J., Robb R. (1985), "Alternative Methods for Evaluating the Impact of Interventions", in Heckman, Singer (a cura di), *Longitudinal Analysis of Labor Market Data*, Cambridge University Press, Cambridge.
- Heckman J.J., Lalonde R., Smith J. (1999), "The Economics and Econometrics of Active Labor Market Programs", *Handbook of Labor Economics*, Vol. 3, North Holland, Amsterdam.
- Holland P. (1986), "Statistics and Causal Inference," (with discussion), *Journal of the American Statistical Association*, Vol. 81, pp. 945-970.
- Holland P. (1988), "Causal Inference, Path Analysis, and Recursive Structural Equations Models" (with discussion), *Sociological Methodology*, American Sociological Association, 449-493, Washington, D.C.
- Hopkins R. (1995), "Impact Assessment: Overview and Methods of Application", Oxfam/Novib, Oxford.
- Imbens G.W., Angrist J.D. (1994), "Identification and Estimation of Local Average Treatment Effects", *Econometrica*, Vol. 62, pp. 467-476.
- Imbens G.W., Hirano K. (2004), The Propensity Score with Continuous Treatment. Chapter for Missing data and Bayesian Method in Practice: Contributions by Donald Rubin Statistical Family.
- Imbens G.W., Rubin D.B. (1997), "Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance", *Annals of Statistics*, Vol. 25, pp. 305-327.
- Imbens G.W., Wooldridge J.M. (2009), "Recent Developments in the Econometrics of Program Evaluation", *Journal of Economic Literature*, Vol. 47, No. 1.
- Jalan J., Glinskaya E. (2005), "Improving Primary School Education in India: An Impact Assessment of DPEP I", World Bank, Washington, D.C.
- Jeanty P.W., Hitzhusen F. (2006), "Analyzing the Effects of Civil Wars and Violent Conflicts on Food Security in Developing Countries: An Instrumental Variable Panel Data Approach", Selected paper presented at the American Agricultural Economics Association meeting, Long Beach, CA.
- Lee D.S., Lemieux T. (2009), "Regression Discontinuity Designs in Economics", *NBER Working Paper*, No. 14723.
- Lichfield N. (1996), *Community Impact Evaluation*, UCL Press, London.
- Mauro V., Biggeri M., Trani J.F. (2010), "Analyzing the impact of CBR in Mandya District (India)", mimeo, Università di Firenze.
- Mayoux L., Chambers R. (2005), "Reversing The Paradigm: Quantification, Participatory Methods And Pro-Poor Impact Assessment", *Journal of International Development*, Vol. II, No. 17, March-April, pp. 271-298.
- McGillivray M., Feeny S., Hermes N., Lensink R. (2006), "Controversies over the impact of development aid: it works; it doesn't; it can, but that depends,...", *Journal of International Development*, Vol. 18, pp. 1031-1050.
- Morgan M. (1990), *The History of Econometric Ideas*. Cambridge University Press, Cambridge.
- Permutt T., Hebel, J. (1989), "Simultaneous-Equation Estimation in a Clinical Trial of the Effect of Smoking on Birth Weight", *Biometrics*, Vol. 45, pp. 411-422.

- Pitt M.M., Khandker S.R. (1998), "The Impact of Group-Based Credit Programs on Poor Households in Bangladesh: Does the Gender of Participants Matter?", *The Journal of Political Economy*, Vol. 106, No. 5, October, pp. 958-996.
- Rajan R.G., Subramanian A. (2005), "What Undermines Aid's Impact on Growth?", *NBER Working Papers*, No. 11657, National Bureau of Economic Research.
- Reinisch J., Sanders S., Mortensen E., Rubin D.B. (1996), "In Utero Exposure to Phenobarbital and Intelligence Deficits in Adult Men", *Journal of the American Medical Association*, Vol. 274, pp. 1518-1525.
- Riddell R. (2007), *Does foreign aid really work?*, Oxford University Press, Oxford.
- Rosenbaum P.R. (1984) "From Association to Causation in Observational Studies: The Role of Tests of Strongly Ignorable Treatment Assignment", *Journal of the American Statistical Association*, Vol. 79, pp. 41-48.
- Rosenbaum P., Rubin D. (1983a), "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, Vol. 70, pp. 41-55.
- Rosenbaum P., Rubin D. (1983b), "Assessing the Sensitivity to an Unobserved Binary Covariate in an Observational Study with Binary Outcome", *Journal of the Royal Statistical Society, Ser. B*, Vol. 45, pp. 212-218.
- Rossi P., Freeman H., Lipsey M. (2003), *Evaluation: a Systematic Approach*, SAGE, Beverly Hills, CA.
- Rubin D.B. (1973), "The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies", *Biometrics*, Vol. 29, No. 1, pp. 185-203.
- Rubin D.B. (1974), "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies", *Journal of Educational Psychology*, Vol. 66, pp. 688-701.
- Rubin D.B. (1976), "The Use of Matching and Regression Adjustment to Remove Bias in Observational Studies", *Biometrics*, Vol. 29, pp. 185-203.
- Rubin D.B. (1977), "Assignment to Treatment Group on the Basis of a Covariate", *Journal of Educational Statistics*, Vol. 2, No. 1, pp. 1-26.
- Rubin D.B. (1978), "Bayesian inference for causal effects: The Role of Randomization", *Annals of Statistics*, Vol. 6, pp. 34-58.
- Rubin D.B. (1980), "Bias Reduction Using Mahalanobis Metric Matching", *Biometrics*, Vol. 36, pp. 293-298.
- Rubin D.B. (1990), "Formal Modes of Statistical Inference for Causal Effects", *Journal of Statistical Planning and Inference*, Vol. 25, pp. 279-292.
- Rubin D.B., Thomas N. (1992), "Characterizing the Effect of Matching Using Linear Propensity Score Methods with Normal Distributions", *Biometrika*, Vol. 79, pp. 797-809.
- Rubin D.B., Thomas N. (1996), "Matching Using Estimated Propensity Score: Relating Theory to Practice", *Biometrics*, Vol. 52, pp. 249-264.
- Rubin D.B., Thomas N. (2000), "Combining Propensity Score Matching with Additional Adjustments for Prognostic Covariates", *Journal of the American Statistical Association*, Vol. 95, pp. 573-585.
- Schulz T. (2001), "School Subsidies for the Poor: Evaluating the Mexican Progresa Poverty Program", *Economic Growth Center Discussion Paper*, No. 834.
- Smith J., Todd P. (2005), "Does Matching Overcome, LaLonde's Critique of Nonexperimental Estimators?", *Journal of Econometrics*, Vol. 125, No. 1-2, pp. 305-353.
- Sommer A., Zeger S. (1991), "On Estimating Efficacy from Clinical Trials", *Statistics in Medicine*, Vol. 10, pp. 45-52.
- Stame N. (1998), *L'esperienza della valutazione*, SEAM, Roma.
- Stame N. (2007) *I classici della valutazione*, FrancoAngeli, Milano.
- Schwartz J.B., Bhushan I. (2006), "Cambodia: Using Contracting to Reduce Inequity in Primary Health Care Delivery", in Gwatkin D.R., Wagstaff A., Yazbeck A.S. (a cura di), *Reaching the Poor with Effective Health, Nutrition, and Population Services: What Works, What Doesn't, and Why*, The World Bank, Washington, D.C.
- Thistlewaite D., Campbell D. (1960), "Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment", *Journal of Educational Psychology*, Vol. 51, pp. 309-317.

- Todd P.E., Wolpin K.I. (2009), "Structural Estimation and Policy Evaluation in Developing Countries", *PIER Working Paper Archive*, No. 09-028, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.
- Trochim W. (1984), *Research Design for Program Evaluation; The Regression-discontinuity Design*, Sage Publications, Beverly Hills, CA.
- Trochim W. (2001), "Regression-Discontinuity Design", in Smelser N.J., Baltes P.B. (a cura di), *International Encyclopedia of the Social and Behavioral Sciences*, Vol. 19, pp. 12940-12945, Elsevier North-Holland, Amsterdam.
- Urquiola M. (2006), "Identifying Class Size Effects in Developing Countries: Evidence from Rural Bolivia", *Review of Economics and Statistics*, Vol. 88, pp. 171-177.
- White H. (2007a), "Evaluating Aid Impact: Approaches and Findings", in Lahiri S. (a cura di), *The Theory and Practice of Foreign Aid*, Vol. 1, Elsevier, New York.
- White H. (2007b), "Evaluating Aid Impact", *WIDER working paper*, No. 2007/75.
- Wooldridge J., (2002a), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, MA.
- World Bank (2005), *Maintaining Momentum to 2015? An Impact Evaluation of Interventions to Improve Maternal and Child Health and Nutrition Outcomes in Bangladesh*, Operations Evaluation Department, The World Bank, Washington D.C.
- Wright S. (1928), Appendix to: The Tari on Animal and Vegetable Oils, P.G. Wright, MacMillan, New York.
- Wright S. (1934), "The Method of Path Coefficients", *Annals of Mathematical Statistics*, Vol. 5, pp. 161-215.
- Zhao Z. (2004), "Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence", *Review of Economics and Statistics*, Vol. 86, No. 1, pp. 91-107.