

# MN2

Lorenzo De Sio

## **Elettori in movimento**

Nuove tecniche di inferenza ecologica  
per lo studio dei flussi elettorali



EDIZIONI POLISTAMPA

# Indice

Premessa.....	11
1. Flussi elettorali e inferenza ecologica.....	17
1.1. Il problema.....	20
1.2. La regressione ecologica.....	24
1.3. La fallacia ecologica.....	27
1.4. Il modello di Goodman.....	33
2. Lo studio dei flussi elettorali in Italia.....	39
2.1. Il comportamento elettorale a Milano.....	39
2.2. Le ricerche dell'Istituto Cataneo e il "boom" del modello.....	43
2.3. Prime incertezze e .. bolle di sapone.....	48
2.4. Critiche e perplessità.....	50
2.5. Alcuni spunti di riflessione.....	56
3. Oltre il modello di Goodman: dal metodo dei vincoli all'approccio EI.....	59
3.1. Il problema.....	60
3.2. Informazioni presenti nei dati.....	61
3.3. Il metodo dei vincoli.....	63
3.4. La tomografia.....	72
4. La stima dei parametri di interesse: un approccio, diverse tecniche.....	79
4.1. Stima di coefficienti distribuiti secondo una normale bivariata troncata.....	80
4.2. Minimizzazione della distanza dalle rette di vincolo.....	82
4.3. Criteri di confronto tra diverse tecniche di stima.....	84
4.4. Diverse tecniche di stima per un unico approccio.....	90
5. Oltre le variabili dicotomiche: la generalizzazione dell'approccio EI.....	95
5.1. La necessità dell'estensione a variabili politomiche.....	96

In copertina:  
Pieter Bruegel il vecchio, *Ballo di morge*, The Detroit Institute of Arts.

© 2008 Edizioni Polistampa  
Via Livorno 8/32 - 50142 Firenze  
Tel. 055 737871 (15 linee)  
info@polistampa.com - www.polistampa.com  
Sede legale: Via Santa Maria, 27/r - 50125 Firenze

L'editore è pienamente disponibile a soddisfare eventuali oneri derivanti da diritti di riproduzione per le immagini di cui non sia stato possibile reperire gli aventi diritto.

ISBN 978-88-596-0364-1

5.2. Un approccio a più stadi: uso multiplo della tecnica dicotomica EI.....	97
5.3. Ripensare EI in più dimensioni.....	105
6. Un esempio di applicazione empirica: il voto differenziato nelle elezioni politiche del 2001.....	119
6.1. Voto differenziato e differenziale di coalizione.....	120
6.2. L'analisi complessiva.....	127
7. Le mappe del voto differenziato.....	145
7.1. Il differenziale di coalizione.....	145
7.2. Il voto differenziato e le sue componenti.....	152
7.3. La struttura territoriale del voto differenziato.....	165
8. Conclusioni.....	169
Riferimenti bibliografici.....	173

## Indice delle tabelle

Tabella 1.1 – Definizione di matrice di flusso tra due dicotomie in un'unità ecologica.....	21
Tabella 1.2 – Incognite in una matrice di flusso tra due dicotomie in un'unità ecologica.....	22
Tabella 2.1 – Stima di matrice di flusso a Milano, 1972-75.....	41
Tabella 2.2 – Stima di matrice di flusso a Milano, 1972-75, corretta con RAS.....	42
Tabella 3.1 – Incognite in una matrice di flusso tra due dicotomie in un'unità ecologica.....	60
Tabella 3.2 – Definizione di coefficienti di flusso in un'unità ecologica.....	64
Tabella 4.1 – Caratteristiche di alcune soluzioni al problema dell'inferenza ecologica.....	91
Tabella 5.1 – Incognite in una matrice di flusso generale in un'unità ecologica.....	97
Tabella 5.2 – Dicotomizzazione artificiosa per la stima di un coefficiente.....	98
Tabella 5.3 – Coefficienti di flusso in una matrice 3x2 in un'unità ecologica.....	105
Tabella 5.4 – Coefficienti di flusso in una matrice 3x2 in un'unità ecologica.....	107
Tabella 5.5 – Coefficienti di flusso in una matrice mXn in un'unità ecologica.....	115
Tabella 6.1 – Definizione dei comportamenti di voto differenziato (% su elettori).....	130
Tabella 6.2 – Stima Eidis dei comportamenti di voto differenziato – totale nazionale (% su elettori).....	132
Tabella 6.3 – Confronto tra diverse stime del voto differenziato.....	136

## Indice delle figure

Figura 3.1 – Voto a sinistra in due elezioni nelle sezioni elettorali di un collegio .....	62
Figura 3.2 – Vincolo relativo ai valori dei coefficienti $b_j^{sS}$ e $b_j^{dS}$ in un'unità ecologica.....	67
Figura 3.3 – Vincolo relativo ai valori dei coefficienti $b_j^{sS}$ e $b_j^{dS}$ in due unità ecologiche .....	71
Figura 3.4 – Tomografia di un collegio elettorale .....	73
Figura 5.1 – Vincolo relativo ai valori dei coefficienti $b_j^{sS}$ , $b_j^{eS}$ e $b_j^{dS}$ in un'unità ecologica.....	108
Figura 6.1 – Differenziale di coalizione del centrosinistra 1994-2001 .....	123
Figura 6.2 – Differenziale di coalizione del centrodestra 1994-2001 .....	124
Figura 7.1 – Distacco Cdl-Ulivo nel proporzionale (% sugli elettori).....	146
Figura 7.2 – Distacco Cdl-Ulivo nel maggioritario (% sugli elettori).....	147
Figura 7.3 – Differenziale di coalizione per la Cdl (punti percentuali su elettori) .....	150
Figura 7.4 – Differenziale di coalizione per l'Ulivo (punti percentuali su elettori) .....	151
Figura 7.5 – Coerenza dell'elettorato proporzionale Cdl (% sui voti PR della Cdl) .....	154
Figura 7.6 – Coerenza dell'elettorato proporzionale Ulivo (% sui voti PR dell'Ulivo).....	155
Figura 7.7 – Conflitto subito dalla Cdl (% su elettori) .....	158
Figura 7.8 – Conflitto subito dall'Ulivo (% su elettori).....	159
Figura 7.9 – Dispersione subita dalla Cdl (% su elettori).....	161
Figura 7.10 – Dispersione subita dall'Ulivo (% su elettori).....	162
Figura 7.11 – Attrazione esercitata dalla Cdl (% su elettori) .....	164

## Premessa

Tra gli studiosi del comportamento elettorale, l'analisi degli *elettori in movimento* (anche conosciuta come analisi dei *flussi elettorali*) rappresenta da molto tempo un ambito importante. Tuttavia non si può dire che, in particolare in Italia, questo tipo di analisi abbia sempre goduto di un'accettazione serena da parte della comunità scientifica. In particolare nel periodo tra gli anni Ottanta e i primi Novanta – anni di grande fioritura di questo tipo di studi – la presenza di analisi di qualità metodologica eterogenea aveva finito per gettare una luce particolarmente negativa sull'analisi dei flussi elettorali, e sulle tecniche statistiche usate per condurla.

Rileggendo, dopo tanto tempo, quelle polemiche tuttavia sempre emerse tra le righe anche un'altra impressione: che uno dei motivi per cui le analisi di flusso vennero criticate con forza era perché fornivano un quadro dei comportamenti di voto in parte diverso dal paradigma della stabilità a cui si erano abituate generazioni di studiosi.

In particolare le analisi proposte dal gruppo di ricerca dell'Istituto Caritaneo alla fine degli anni Ottanta [Corbetta, Parisi, e Schadee 1988] avevano proposto un'interpretazione in parte inedita dei comportamenti di voto, arrivando a sostenere la presenza di una significativa mobilità elettorale anche in un periodo (un ciclo elettorale di quattro elezioni politiche dal 1968 al 1983) che la maggior parte degli studiosi considerava caratterizzato da una sostanziale stabilità del voto. Di qui molte critiche, perché è

evidente che l'idea di un elettorato ancora sostanzialmente “congelato” nelle appartenenze determinate da fratture di lungo periodo era difficile da conciliare con nuovi studi che evidenziasero invece la presenza – sotto l'apparente stabilità dei risultati aggregati – di una significativa mobilità individuale.

È difficile dire chi avesse ragione e chi torto in quel dibattito. Tuttavia per cercare di fare chiarezza è indispensabile fare una considerazione metodologica. Bisogna infatti sempre ricordare che i risultati aggregati del voto (ovvero le masse di voti raccolti a livello nazionale da ogni partito) non sono altro che la somma di tanti risultati locali, i quali a loro volta sono la somma di milioni di comportamenti individuali. Fin qui niente di strano: ma il punto è che l'aggregazione a livello nazionale dei risultati locali comporta inevitabilmente una significativa *soppressione della varianza*. Aree territoriali che magari si muovono significativamente in direzioni opposte finiscono per compensarsi a livello nazionale, generando una impressione di stabilità in parte ingannevole<sup>1</sup>. Lo stesso, su scala più piccola, può avvenire a livello regionale e locale, dando ancora una volta un'impressione di stabilità che nasconde un'effettiva mobilità degli elettori.

È chiaro che dal punto di vista dei rapporti di forza tra partiti quello che conta è essenzialmente il risultato nazionale. Tuttavia, già scendere al livello dei rapporti di forza *interni* ai partiti chiama in causa la distribuzione territoriale dei consensi.

C'è inoltre un secondo ordine di motivi per cui studiare gli elettori in movimento è utile per studiare la politica.

---

<sup>1</sup> A questo va aggiunto che a ogni elezione è presente un significativo ricambio della popolazione elettorale. L'Istat riporta, per gli ultimi anni, un numero annuo (sia di nascite che di decessi) di circa 560.000 unità. Arrivando per semplicità ai nuovi elettori la stessa numerosità dei nati in questi anni, il risultato è che – nell'arco di una legislatura – circa il 6% dell'intero elettorato viene sostituito da un altro 6% che è stato socializzato politicamente in condizioni completamente diverse (e pesa per circa il 7-8% sui voti validi).

Le elezioni non sono soltanto un meccanismo di selezione della classe politica. Nel linguaggio della teoria dei sistemi, le elezioni sono il principale *input* che il sistema politico riceve. Si tratta di una gigantesca raccolta periodica di informazioni relative alle preferenze politiche degli elettori. È in questa prospettiva che la mobilità elettorale è un'informazione chiave. Una classe politica che voglia tenere il polso dell'elettorato deve essere capace di non fermarsi all'apparente stabilità aggregata: occorre andare in maggiore profondità, alla ricerca delle direzioni che orientano una mobilità spesso complessa, ma molto raramente erratica.

Ed è proprio per queste finalità informative in senso lato che gli *elettori in movimento* esercitano un ruolo chiave per gli stessi regimi democratici. Se è vero che una delle proprietà fondamentali delle democrazie è la *responsiveness*, ovvero la capacità di rispondere alle richieste e aspettative dei governanti [Dahl 1971], è evidente che il ruolo degli elettori in movimento è cruciale. Gli elettori disposti a cambiare opinione (e scelta di voto) tra due elezioni sono quelli destinati a promuovere i successi di un'attività di governo, oppure a sanzionarne gli insuccessi, o infine ad approvare, con il loro consenso, la presentazione di progetti innovativi.

È questo il motivo per cui, nel sistema partitico italiano, ad oggi sostanzialmente bipolare e caratterizzato da alternanza, il ruolo degli elettori in movimento è diventato ancora più rilevante che in passato. A maggior ragione per la necessità di analizzare la risposta degli elettori alle grandi trasformazioni in atto nel sistema partitico e nei partiti stessi<sup>2</sup>.

Il problema, e veniamo al contributo specifico di questo volume, è che alla crescente importanza dello studio degli elettori in movimento corrispondono strumenti tecnici ancora sostanzialmente identici a quelli di molti anni fa, e afflitti da problemi rilevanti. Alcune delle critiche degli anni Ottanta alle analisi di flusso

---

<sup>2</sup> E in questo, soprattutto in riferimento alla prima parte della transizione italiana, caratterizzata dal sistema elettorale misto maggioritario-proporzionale, diventa ad esempio cruciale esaminare la presenza e la consistenza dei comportamenti di *voto differenziale* tra le due aree competitive.

erano fondate: la tecnica comunemente usata per stimare i flussi di voto da dati aggregati, ovvero il c.d. “modello di Goodman” si basa infatti su ipotesi teoriche abbastanza forti, in assenza delle quali è destinata a produrre risultati poco accurati. Il problema è che l'applicabilità di queste ipotesi ai dati studiati deve essere valutata soggettivamente dal ricercatore, senza la disponibilità di precisi strumenti diagnostici. Come uscirne?

Si tratta di un dibattito che non si è svolto solo in Italia, e nel quale sono stati proposti alcuni approcci alternativi. Tra questi, alla fine degli anni '90 ne è emerso uno particolarmente promettente, ovvero l'approccio EI (*ecological inference*), proposto dallo studioso americano Gary King [1997]. Si tratta di una proposta molto innovativa non solo in termini tecnico-statistici, ma anche in termini concettuali di fondo. Scopriremo infatti come i risultati elettorali, disaggregati a livello di piccole unità ecologiche come la sezione elettorale, contengano una quantità inaspettata di informazioni *deterministiche* (ovvero certe, non soggette a ipotesi) sui movimenti di voto; una quantità di informazioni ben superiore ai dati utilizzati dal modello di Goodman. Il risultato è che è possibile costruire delle procedure di stima in grado di utilizzare completamente queste informazioni, e quindi di giungere a stime dei flussi di voto decisamente più solide, in quanto appoggiate su una quantità di informazioni notevolmente maggiore. Tuttavia il modello proposto da King ha una grave limitazione: si limita ad analizzare la tabulazione incrociata di due dicotomie. È quindi inutilizzabile per qualunque applicazione reale di studio dei flussi di voto in un sistema multipartitico.

E qui giunge il contributo originale di questo volume: non solo la tecnica di stima, ma il vero e proprio approccio EI vengono infatti ripensati dalle fondamenta e generalizzati in termini multidimensionali, estendendoli al caso generale di matrici di flusso di qualunque dimensione. Contestualmente a questa estensione concettuale, viene anche sviluppata una specifica procedura di stima, in grado di tradurlo in uno strumento effettivamente utilizzabile per l'analisi dei flussi di voto.

La logica conclusione è quindi quella di utilizzare questo strumento per analizzare un concreto problema di stima di flussi: i comportamenti di *voto differenziato* tra maggioritario e proporzionale nelle elezioni politiche del 2001, prendendo in esame non un semplice campione di sezioni elettorali, ma le sezioni relative a *tutto il territorio nazionale*, suddivise in analisi locali indipendenti, a livello del singolo collegio elettorale. Il risultato è quindi anche un'analisi della *distribuzione geografica* dei vari comportamenti di voto differenziato.

Il volume è così strutturato. I primi due capitoli sono dedicati a una ricostruzione storica dello studio dei flussi elettorali. Il primo si occupa delle origini dell'*inferenza ecologica*, ovvero dello studio di comportamenti individuali attraverso dati aggregati, a partire dai primi esempi storici fino al modello di Goodman. Il secondo è invece dedicato all'Italia: si ricostruisce l'introduzione del modello in Italia, lo sviluppo di varie ricerche, e infine l'accesso dibattito metodologico.

La seconda parte del volume è invece dedicata all'approccio EI. Nel capitolo 3 viene presentato l'approccio nei suoi termini concettuali. Nel capitolo 4 vengono presentate invece le relative procedure di stima: non solo, come vedremo, quella proposta da King, ma anche alcune alternative più semplici. Nel capitolo 5 viene infine presentata la generalizzazione dell'approccio EI a matrici di flusso di qualunque dimensione, con le relative procedure di stima.

La terza parte del volume è infine dedicata all'applicazione della tecnica EI, generalizzata, alla stima dei comportamenti di voto differenziato nelle elezioni politiche del 2001 alla Camera. Il capitolo 6 presenta i risultati complessivi a livello nazionale delle stime, ottenute aggregando le stime locali ottenute in ogni collegio elettorale. Il capitolo 7, basandosi sulle stesse stime locali, presenta invece la distribuzione geografica dei vari comportamenti di voto differenziato sotto forma di mappe.

Come molti lavori di ricerca, questo volume – anche se è frutto di uno sforzo individuale – non sarebbe stato possibile senza molte altre persone. Tra queste ce ne sono alcune che voglio espressamente ringraziare. Anzitutto Roberto D’Alimonte, per il ruolo insostituibile avuto nell’incoraggiare e sostenere questa ricerca, nonché per aver messo a disposizione l’eccezionale *data set* che ha reso possibili le elaborazioni dell’ultima parte. Hans M. A. Schadee ha avuto una straordinaria disponibilità nel seguire e rivedere il lavoro, e ha fornito consigli e spunti preziosi per il suo sviluppo, così come – successivamente – Piergiorgio Corbetta. Andrea Colasanti ha dato alcuni spunti indispensabili per la parte più strettamente matematica dell’analisi; Alessandro Chiaromonte mi ha inoltre dato un costante e prezioso supporto. Non posso inoltre non menzionare Alberto Marradi, che non ha avuto un coinvolgimento diretto in questo lavoro, ma mi ha fornito una lezione metodologica senza la quale non avrei immaginato e condotto questa ricerca. Tutte queste persone hanno fornito spunti preziosi: ovviamente la responsabilità di come tali spunti siano poi stati messi in pratica, con tutti gli errori e le omissioni, ricade interamente su di me. Infine, un ultimo ringraziamento è quello a Mariella, senza la quale non solo questa ricerca, ma tutto il percorso da cui è partita non sarebbe stato possibile.

## 1. Flussi elettorali e inferenza ecologica

All’interno degli studi elettorali in Italia negli ultimi decenni, la stima dei *flussi elettorali* ha rappresentato un tema di particolare interesse. Con questo termine si designa di norma una stima delle quote di elettori che hanno cambiato voto tra due elezioni, comprensiva di informazioni sulle differenti quote in “entrata” e in “uscita” tra i singoli partiti nell’elezione precedente e i singoli partiti nell’elezione successiva. Concretamente, una tavola riassuntiva dei flussi elettorali è quindi una tabella di contingenza in cui vengono riassunti i movimenti di elettori da ciascun partito presente in una consultazione elettorale svoltasi al momento  $t_1$ , a ciascun partito presente in una consultazione successiva svoltasi al momento  $t_2$  (ovviamente con  $t_2 > t_1$ ).

In questo senso, l’analisi della mobilità elettorale tra due elezioni successive ha conosciuto una grande attenzione in particolare nel periodo successivo al c.d. “terremoto elettorale” del 1975-76; in quella fase lo studio della accresciuta mobilità degli elettori andava di pari passo con il riconoscimento stesso dell’esistenza di una quota di elettorato che iniziava a prendere in considerazione la possibilità di cambiare voto da un’elezione all’altra, in relazione all’indebolirsi dei tradizionali vincoli di appartenenza che agivano da potenti stabilizzatori del voto [Parisi e Pasquino 1977]. È così che, anche in relazione alla disponibilità di nuovi strumenti statistici, a partire dalla fine degli anni Settanta il tema dei flussi elettorali ha conosciuto una fase di grande attenzione, anche in senso



metodologico, con l'utilizzo di procedure di stima basate su dati aggregati<sup>1</sup>.

Negli anni successivi il mutamento congiunto del sistema elettorale e del sistema partitico ha in qualche modo ridefinito la gerarchia delle priorità nella ricerca elettorale, portando leggermente in secondo piano il semplice confronto diacronico tra due elezioni. Inoltre l'emersione di alcuni dubbi sull'affidabilità delle tecniche di stima dei flussi elettorali in base a dati aggregati ha ulteriormente scoraggiato l'uso di queste tecniche. A questo va aggiunto che proprio in quegli anni iniziavano alcuni programmi di ricerca destinati a costruire anche in Italia indagini regolari mediante lo strumento del *survey*, che permettevano lo studio dei flussi elettorali direttamente su base individuale [Segatti 1997].

In realtà però non si può dire che la stima dei flussi elettorali mediante dati aggregati – che è l'oggetto principale di questo studio – sia stata completamente abbandonata. Anzitutto essa ha continuato ad essere utilizzata – a volte anche in abbinamento con dati di sondaggio – per la stima del movimento di voto tra due elezioni [Natale 1997; 2000; 2002; De Sio 2007a]. Ma è stata in particolare la riforma elettorale del 1993, con l'introduzione – alla Camera – di un sistema misto maggioritario-proporzionale con due voti indipendenti [D'Alimonte e Chiaromonte 1995; Chiaromonte 2005] a creare l'opportunità per una nuova applicazione delle tecniche di cui sopra, in condizioni nuove e particolarmente interessanti.

Si tratta della nascita di applicazioni di tipo *sincronio*, ovvero in cui non si confrontano più due elezioni in tempi diversi, ma voti diversi (ad esempio, voto maggioritario e proporzionale) espressi dagli stessi elettori nello stesso giorno<sup>2</sup>.

Non che tali analisi non fossero possibili già in passato, ad esempio tra Camera e Senato. Il punto è che in un sistema fortemente proporzionale il confronto era di scarso interesse per la si-

multudine dei risultati tra le due Camere. Viceversa, le elezioni che seguono la riforma elettorale del 1993 metteranno in evidenza significative discrepanze tra voto maggioritario e proporzionale, tali da giustificare un confronto sistematico tra i due risultati, in grado di fornire informazioni molto interessanti<sup>3</sup>. A questo va aggiunto che queste applicazioni di tipo sincronico non presentano i problemi di ricambio della popolazione elettorale che invece affliggono le applicazioni diacroniche.

Infine va menzionato che nelle elezioni più recenti sono divenuti disponibili i risultati elettorali disaggregati a livello di sezione *per l'intero territorio nazionale*. In questo senso è divenuto possibile produrre stime di flusso su larga scala (dividendo il territorio nazionale in unità territoriali, all'interno delle quali effettuare stime separate dei flussi), capaci oltretutto di mostrare l'eterogeneità territoriale dei fenomeni studiati, senza dover ricorrere a un campione di sezioni.

Ce n'è abbastanza per giustificare la necessità di una riflessione sulla stima dei flussi elettorali (diacronici o sincronici) in base a dati aggregati. Tanto più alla luce di alcune recenti proposte innovative [King 1997] volte proprio a superare alcuni dei limiti tradizionali di queste tecniche di stima. Si tratta dell'oggetto di questo volume. A partire da una ricognizione storica sull'origine del problema dell'inferenza ecologica, e sulla storia dei flussi elettorali in Italia, saranno esaminate alcune tecniche innovative, e di queste verrà proposta una generalizzazione in grado di renderle utilizzabili per la stima dei flussi elettorali. Infine questa nuova tecnica di stima verrà applicata a un caso empirico.

Veniamo a questo punto a esporre in termini generali il problema della stima dei flussi elettorali.

---

<sup>1</sup> Per una rassegna vedi il capitolo 2.

<sup>2</sup> Ricolfi e Testa parlano non a caso di “flussi istantanei” [2002, 236].

<sup>3</sup> Per una rassegna sul *noto differenziale* vedi il capitolo 6.

## 1.1. Il problema

Il problema fondamentale nell'analisi dei movimenti di flusso elettorale è che tali movimenti (che corrispondono a elettori che cambiano scelta di voto) non sono direttamente rilevabili, in quanto l'unico dato disponibile è il semplice *saldo*, negativo o positivo, di un partito rispetto al risultato ottenuto nell'elezione precedente; saldi di entità minima o anche nulli possono in realtà nascondere trasferimenti di voti anche molto complessi e che coinvolgono più partiti [Mannheimer e Micheli 1976; Barbagli et al. 1979; Corbetta e Schadee 1984].

In relazione al problema della non diretta rilevabilità di questi flussi, sono emersi storicamente due possibili approcci.

Il primo è quello che prevede la ricostruzione dei movimenti individuali di voto mediante interviste a campione, con i problemi noti relativi alle procedure di campionamento (effettiva *casualità* e *rappresentatività* del campione) e alla veridicità delle risposte degli intervistati.

Il secondo consiste invece nella stima, mediante procedure statistiche, di movimenti di voto *medi* in un ambito ecologico *aggregato* (collegio elettorale, comune, provincia, ecc.) a partire dai risultati elettorali osservati in unità ecologiche a un livello di aggregazione inferiore (in genere, il più piccolo livello di aggregazione disponibile: in Italia, prevalentemente, la sezione elettorale). Queste lavoro intende concentrarsi su quest'ultimo approccio.

Punto di partenza di questo percorso è la decisione di operare inferenze su comportamenti individuali a partire esclusivamente da dati ecologici. È per questo motivo che King [1997] designa questo approccio operativo con il termine “inferenza ecologica”.

Tabella 1.1 – Definizione di matrice di flusso tra due dicotomie in un'unità ecologica

	$I_2$		
	Sinistra	Destra	Totale
$I_1$			
sinistra	$N_i^{SS}$	$N_i^{SD}$	$N_i^S$
destra	$N_i^{DS}$	$N_i^{DD}$	$N_i^D$
Totale	$N_i^S$	$N_i^D$	$N_i$

Da un punto di vista analitico, il nostro problema di inferenza ecologica può essere espresso nei seguenti termini [King 1997, 28]: date due variabili categoriali, si dispone di  $n$  tabulazioni incrociate di queste due variabili relative alle  $n$  unità ecologiche<sup>4</sup> considerate, dove ciascuna delle  $n$  tabulazioni incrociate si riferisce a frequenze di variabili individuali nell'ambito della singola unità ecologica. Conoscendo i marginali di riga e colonna in ciascuna delle  $n$  tabelle, si vuole ottenere una stima dei valori di cella per l'ambito aggregato.

In pratica, il punto di partenza del nostro problema di inferenza ecologica potrebbe essere una tabella come la Tabella 1.1, relativa al confronto tra il comportamento di voto in due elezioni  $I_1$  e  $I_2$ . La tabella dell'esempio si riferisce a una singola unità ecologica (sezione elettorale) denominata con l'indice  $i$ .

La notazione utilizzata è il più possibile semplice: con  $N_i^{SS}$  intendendo, nella sezione elettorale  $i$ , il numero di elettori che ha votato sinistra<sup>5</sup> in entrambe le elezioni; con  $N_i^{SD}$  il numero di elettori che ha votato sinistra alla prima elezione e Destra alla seconda, e così via.

<sup>4</sup> Da qui in poi, con il termine *unità ecologica* verrà indicata la singola unità del livello inferiore di aggregazione. Viceversa, l'aggregato territoriale di queste unità ecologiche verrà designato col termine “ambito aggregato” o più semplicemente “aggregato”. Data questa terminologia, l’inferenza ecologica ha come obiettivo una stima relativa all’“ambito aggregato” a partire da dati relativi alle singole “unità ecologiche”.

<sup>5</sup> Per convenzione indico con lettere minuscole la consultazione precedente, con maiuscole la successiva.

Tabella 1.2 – Inognuite in una matrice di Jussio tra due dicotomie in un'unità ecologica

	$t_2$		
$t_1$	Sinistra	Destra	Totale
sinistra	?	?	$N_i^s$
destra	?	?	$N_i^d$
Totale	$N_i^s$	$N_i^d$	$N_i$

Di conseguenza, riguardo ai totali marginali di colonna:  $N_i^s$  indica il numero di elettori che ha votato Sinistra nella seconda elezione,  $N_i^d$  il numero di elettori che ha votato Destra nella seconda elezione. Per i marginali di riga:  $N_i^s$  sono i voti a sinistra nella prima elezione,  $N_i^d$  i voti a destra nella prima elezione. Infine,  $N_i$  esprime com'è ovvio il numero complessivo di elettori nella sezione elettorale  $i^s$ .

Definita la notazione necessaria, è opportuno chiarire immediatamente la situazione reale, presentata nella Tabella 1.2 [vedi anche Ricolfi 1993, 139]:

Di ogni sezione elettorale conosciamo infatti soltanto i marginali di riga e di colonna, ovvero, nel nostro esempio concreto, i voti ottenuti da ciascun partito in ciascuna elezione.

Vediamo di esprimere i valori di cella (ciò che vogliamo stimare) in funzione dei marginali di riga e colonna (ciò che già conosciamo). Punto di partenza è la Tabella 1.1, che ci permette di enunciare le seguenti identità:

$$\left. \begin{aligned} N_i^s &= N_i^{ss} + N_i^{ds} \\ N_i^d &= N_i^{sd} + N_i^{dd} \end{aligned} \right\} \quad (1.1)$$

<sup>6</sup> L'esempio immagina ovviamente due elezioni in cui i partiti siano gli stessi, la popolazione sia identica, e non vi siano astenuti o voti non validi.

Fino ad ora abbiamo visto le quantità da determinare in valori assoluti. Possiamo però, ad esempio, invece che riferirci al *numero* di elettori di sinistra che confermano il loro voto, alla *relativa porzione*:

Ad esempio, la proporzione di elettori di sinistra che confermano il loro voto anche nella seconda elezione può essere espressa come:

$$b_i^{ss} = \frac{N_i^{ss}}{N_i^s} \quad (1.2)$$

da cui:

$$N_i^{ss} = b_i^{ss} N_i^s \quad (1.3)$$

andando a sostituire la (1.3) nelle identità (1.1) otteniamo le seguenti nuove identità<sup>7</sup>:

$$\left. \begin{aligned} N_i^s &= b_i^{ss} N_i^s + b_i^{ds} N_i^d \\ N_i^d &= b_i^{sd} N_i^s + b_i^{dd} N_i^d \end{aligned} \right\} \quad (1.4)$$

Le identità (1.4) sono di semplice interpretazione: esprimono, nel primo caso, il totale dei voti Sinistra a  $t_2$  come la proporzione di elettori sinistra a  $t_1$  che ha votato Sinistra a  $t_2$ , moltiplicata per il totale degli elettori sinistra a  $t_1$ , sommata alla proporzione di elettori destra a  $t_1$  che ha votato Sinistra a  $t_2$  moltiplicata per il totale degli elettori destra a  $t_1$ . Nel secondo caso, esprimono la stessa relazione per quanto riguarda chi ha votato Destra a  $t_2$ .

<sup>7</sup> vedi anche Coricelli e Weber [1993, 200].

Le (1.4) contengono tuttavia, nella loro semplicità, l'enunciazione completa del problema in questione. Oggetto della stima sono, in generale, coefficienti  $b_i^{ab}$ , che esprimono la proporzione di individui che, presentando la modalità  $a$  sulla variabile  $X$ , presentano la modalità  $B$  sulla variabile  $Y$ , e così via.

Ma le (1.4) esprimono anche l'impossibilità di individuare i coefficienti disponendo soltanto delle informazioni relative a una sezione elettorale: siamo infatti di fronte a un sistema di due equazioni con quattro incognite, non risolvibile per via analitica. Di qui la necessità di effettuare una *stima*.

## 1.2. La regressione ecologica

Il problema così come è stato appena formulato non è nuovo né recente nell'ambito delle scienze sociali. Achen e Shively [1995] mettono in evidenza l'importanza cruciale dei dati aggregati per la ricerca quantitativa, sia in Europa che negli Stati Uniti, nel periodo a cavallo tra secolo XIX e XX, prima dell'avvento della metodologia del *survey*.

Un esempio interessante di questa tradizione è quello del gruppo di ricerca diretto da William Ogburn all'Università di Chicago negli anni '10 e '20. E proprio di questo gruppo è il lavoro [Ogburn e Goltra 1919] che di fatto segna la comparsa di tecniche di regressione ecologica in una rivista di scienza politica.

Lo studio riguardava il comportamento elettorale femminile nello stato dell'Oregon, nel quale era stato recentemente introdotto il suffragio femminile. La necessità di far fronte con delle procedure di stima all'impossibilità di rilevare empiricamente il comportamento elettorale delle donne era così sintetizzata dagli autori:

[...] even though the method of voting makes it impossible to count women's votes, one wonders if there is not some indirect method of solving the problem. The height of a waterfall is not

measured by dropping a line from the top to the bottom, nor is the distance from the earth to the sun measured by a rod and chain. [Ogburn e Goltra 1919, 414]<sup>8</sup>

La soluzione ipotizzata dai due fu quella di calcolare coefficienti di correlazione a livello aggregato, per inferire la presenza o meno di correlazione a livello individuale. Nella fattispecie, misurarono la correlazione, attraverso le varie sezioni<sup>9</sup> elettorali della città di Portland, tra la percentuale di donne elettrici e il consenso ottenuto da alcuni *referenda*.

Il ragionamento metodologico era che fosse possibile affermare che le donne avessero una tendenza maggiore degli uomini a bocciare i quesiti referendari, se in sezioni con alte percentuali di donne al voto si osservavano percentuali più alte di "no" rispetto a sezioni con basse percentuali di donne<sup>10</sup>. Va aggiunto che i due studiosi erano già allora consapevoli di alcuni ordini di problemi che potevano affliggere questo tipo di stime: da un lato, la possibilità di correlazione spuria, che li portò a scontare dai risultati ottenuti una sorta di "indice di conservatorismo" della sezione elettorale, calcolato in base ad altri dati<sup>11</sup>; dall'altro, l'intuizione che una correlazione anche forte a livello aggregato potesse nascondere un'assenza di correlazione o addirittura una correlazione di segno opposto a livello individuale. I due autori sembrano consapevoli di questo potenziale problema:

<sup>8</sup> "[...] anche se il metodo di votazione rende impossibile conteggiare i voti delle donne, viene da chiedersi se non esista un modo indiretto di risolvere il problema. L'altezza di una cascata non si misura tracciando una linea dalla sommità al fondo, né si misura con il metro la distanza tra la Terra e il Sole."

<sup>9</sup> Traduco così il termine *precinct*, che in realtà è decisamente più ampio della tipica sezione elettorale italiana, per evitare il termine ristretto che creerebbe ambiguità rispetto al termine *district* che indica invece un aggregato simile al collegio elettorale.

<sup>10</sup> Appunto la definizione di correlazione.

<sup>11</sup> I risultati della precedente tornata referendaria, svoltasi senza il suffragio femminile.

It is also theoretically possible to gerrymander the precincts in such a way that there may be a negative correlative even though men and women each distribute their votes 50 to 50 on a given measure, but it is thought that no such gerrymandered situation exists in the precincts studied. [Ogburn e Goltra 1919, 415, nota 1]<sup>12</sup>

Dal passo citato emerge tuttavia la necessità di una valutazione soggettiva da parte del ricercatore, in grado di escludere un problema simile. Si tratta del punto debole della procedura, come emergerà successivamente.

Una prima soluzione al problema di riempire i valori di cella della Tabella 1.2 fu dunque quella di calcolare la correlazione tra le due variabili (eventualmente depurare da potenziali correlazioni spurie) utilizzando, come unità di analisi, delle unità ecologiche (le sezioni elettorali) piuttosto che direttamente gli individui<sup>13</sup>.

A detta di Achen e Shively [1995], questo approccio basato su coefficienti di correlazione rappresenta, per un periodo che dura più o meno fino all'inizio degli anni '50, una sorta di paradigma per quanto riguarda l'inferenza ecologica. Durante questo periodo, l'estensione al livello individuale di modelli costruiti a partire dal calcolo dei coefficienti di correlazione tra variabili su unità aggregate è stato considerato non solo un metodo lecito, ma ampiamente utilizzato nell'ambito della ricerca sociale americana.

---

<sup>12</sup> "È anche teoricamente possibile ridisegnare ad arte le sezioni elettorali in una maniera tale che esista una correlazione negativa anche quando uomini e donne distribuiscono i loro voti 50 e 50 su un dato quesito, ma si può ritenere che una tale situazione di gerrymandering non esista nelle sezioni studiate."

<sup>13</sup> La correlazione ecologica in realtà non permette di individuare direttamente i valori di cella della Tabella 1.2, quanto di calcolare un coefficiente che sintetizzi la forza della relazione tra le due variabili dicotomiche

### 1.3. La fallacia ecologica

Questa situazione di "paradigma" viene di fatto interrotta da Robinson [1950], che tematizza in maniera esplicita alcuni problemi che affliggono la correlazione ecologica.

Prima di occuparci direttamente della sostanza del contributo di Robinson, può tuttavia essere interessante soffermarci brevemente sui motivi del successo del suo articolo, e sugli effetti che esso ebbe sugli studiosi americani.

Va sottolineato che contributi critici verso la correlazione ecologica erano già comparsi, anche molto prima di quello di Robinson<sup>14</sup>. Tuttavia, è innegabile come soltanto dopo l'apparizione di quest'ultimo la consapevolezza dell'esistenza di problemi connessi all'inferenza ecologica si diffonda e dispieghi i suoi effetti anche nella pratica della ricerca sociale. King usa a questo proposito toni abbastanza netti:

the use of aggregate data by political scientists, quantitative historians, sociologists, and others declined relative to use of other forms of data; scholars began to avoid using aggregate data to address whole classes of important research questions. [King 1997, 5]<sup>15</sup>

Altrettanto decisi Achen e Shively, che parlano di "collapse of aggregate data analysis [...] and its replacement by individual

---

<sup>14</sup> Micheli [1976, 432] ricorda come già Thorndike nel 1939 avesse individuato il problema, ma senza l'eco suscitata dall'articolo di Robinson. Un insieme più ampio di contributi che individuarono il "paradosso dell'inferenza ecologica" prima di Robinson è fornito da King [1997, 4].

<sup>15</sup> "l'uso di dati aggregati da parte di studiosi di scienza politica, storici quantitativi, sociologi e altri conobbe un declino rispetto all'uso di altri tipi di dati; gli studiosi iniziarono a evitare l'uso di dati aggregati per affrontare interi ambiti di importanti problemi oggetto di ricerca"

survey analysis as the dominant method of quantitative social research” [Achen e Shively 1995, 5]<sup>16</sup>.

Ciò che incuriosisce è come l’articolo di Robinson abbia potuto avere un impatto così notevole nell’ambito della ricerca sociale, quando contributi precedenti che sostenevano posizioni simili erano passati praticamente sotto silenzio.

La risposta va quasi certamente ricercata nella disponibilità, emersa nel frattempo, di nuovi strumenti alternativi, in particolare della metodologia del *survey*, che si afferma proprio in quegli anni. L’articolo di Robinson sarebbe quindi apparso in un periodo in cui finalmente la comunità scientifica disponeva di nuovi strumenti, potendo quindi finalmente ammettere e anzi approvare con decisione critiche al metodo precedente che prima non potevano essere accettate pena il mettere in crisi l’equilibrio metodologico della disciplina<sup>17</sup>:

Over the decade following the Second World War, survey methodology swept the social sciences. [...] the “problem” of cross-level inference appears to have arisen simultaneously with this shift to survey research [...] What is interesting is that this problem was noted at least three times in the early twentieth century, but was ignored. Bernstein in particular reads almost exactly like the later literature of Robinson and Goodman. A cynic might conclude that social scientists tend to ignore logical problems and contradictions in their methods if they do not see anything to be done about them. In contrast, when W. S. Robinson rediscovered the problem in 1950, at a time when survey research was emerging as an alternative to aggregate data

analysis, his article fit researchers’ predispositions and had an enormous impact. [Achen e Shively 1995, 7-8]<sup>18</sup>

È così che l’articolo di Robinson del 1950 rappresenta lo spartiacque che segna di fatto la nascita del “paradosso dell’inferenza ecologica”, ovvero della problematizzazione del rapporto tra modelli costruiti in base a relazioni osservate al livello aggregato e i corrispondenti modelli al livello individuale. Ma cosa affermava esattamente Robinson, e in base a quali osservazioni?

Punto di partenza del suo studio [Robinson 1950] erano dati di censimento che rendevano disponibili, tra gli altri, i dati relativi ad analfabetismo e razza, tanto al livello individuale quanto a vari livelli di aggregazione. L’operazione di Robinson fu concettualmente molto semplice: calcolare il coefficiente di correlazione tra le due variabili dicotomiche “essere nero” e “essere analfabeta” prima a livello individuale (l’unico realmente sensato dal punto di vista sostantivo, visto che si tratta di proprietà individuali) e successivamente a livello aggregato (quindi correlando percentuali di neri e percentuali di analfabetismo nelle singole aggregazioni), utilizzando livelli di aggregazione crescenti.

Il risultato di una simile elaborazione fu notevole: il coefficiente di correlazione  $r$  di Pearson<sup>19</sup> assumeva, al livello individuale, un

---

<sup>16</sup> “collasso dell’analisi di dati aggregati [...] e la sua sostituzione con l’analisi di dati individuali di survey come metodo prevalente della ricerca sociale quantitativa”

<sup>17</sup> Un meccanismo simile a quello classicamente ipotizzato da Kuhn [1962].

---

<sup>18</sup> “Nel decennio che seguì la seconda guerra mondiale, la metodologia del survey sconvolse le scienze sociali [...] il “problema” dell’inferenza ecologica sembra essere sorto contemporaneamente a questo spostamento verso il survey [...] Ciò che è interessante è che questo problema era stato evidenziato almeno tre volte all’inizio del ventesimo secolo, ma nell’indifferenza generale. Bernstein in particolare anticipa quasi esattamente le successive affermazioni di Robinson e Goodman. Un cinico potrebbe dedurne che le scienze sociali tendono a ignorare i problemi logici e le contraddizioni nelle loro tecniche se non vedono una soluzione a portata di mano. Al contrario, quando W. S. Robinson riscoprì il problema nel 1950, in un’epoca in cui la ricerca basata su survey stava emergendo come un’alternativa all’analisi di dati aggregati, il suo articolo calzò a pennello sulle predisposizioni dei ricercatori ed ebbe un enorme impatto.”

valore di 0,203, mentre, se calcolato usando come unità aggregate i singoli stati degli Stati Uniti, il valore ottenuto era di 0,773. Infine, se calcolato usando come unità aggregate nove macrozone geografiche in cui era suddiviso il territorio degli Stati Uniti, esso assumeva un valore di 0,946.

Non solo. Passando a incrociare l'alfabetismo non più con la razza, ma con l'«essere immigrati» (nati negli USA *vs.* nati all'estero) Robinson ottenne risultati ancor più spettacolari: a livello individuale la correlazione era di 0,118, mentre a livello aggregato il valore di  $r$  era rispettivamente di  $-0,526$  a livello dei singoli stati, e di  $-0,619$  a livello delle nove macrozone geografiche.

Il contributo di Robinson non si limitò tuttavia a mostrare due casi in cui la regressione ecologica forniva stime paradossali. Partendo dall'individuazione della relazione esistente tra coefficienti

---

<sup>19</sup> Tale coefficiente si può applicare a un incrocio di due dicotomie interpretando concettualmente le due variabili dicotomiche come variabili cardinali (es. colore nero=1, bianco=0). È un tale procedimento concettuale che permette di calcolare  $r$  di Pearson nella tabella di Robinson, coefficiente che nel caso di due dicotomie è calcolabile con una formula semplificata e conosciuto anche come  $\varphi$ . Una tale operazione concettuale ha però spesso un prezzo che ricorda Marradi:

Alcuni [...] attribuiscono [...] un'interpretazione cardinale, adottando l'intervallo tra 0 e 1 come unità di misura [...] applicano alle dicotomie tutte le tecniche che presuppongono variabili di livello cardinale. Quest'attribuzione [...] non è certo ortodossa ma è tollerabile se i marginali di ciascuna delle dicotomie così trattate sono equilibrati. Se invece sono squilibrati, si producono distorsioni [...] così come provoca distorsioni la distribuzione asimmetrica di una variabile cardinale. [Marradi 1995, 16]

E questo è proprio il caso della tabulazione incrociata di Robinson, in cui una sola cella, quella dei bianchi alfabetizzati, conteneva l'88% dei casi, delineando così una situazione di vera e propria *corner correlation*. D'altronde una simile situazione offre un'interpretazione ancora più favorevole a Robinson, visto che, in situazioni di *corner correlation*,  $\varphi$  tende a sovrastimare la correlazione. A ulteriore riprova di questo fatto, il calcolo del coefficiente alternativo  $\tau$ , anch'esso tendente a sovrastimare situazioni di *corner correlation*, fornisce un valore di 0,046.

di correlazione ecologica, coefficienti di correlazione individuale e numero e dimensioni delle unità di aggregazione, mostrò come, dato il tipo di relazione matematica, fosse ipotizzabile che nella grande maggioranza dei casi la correlazione ecologica avrebbe fornito una notevole sovrastima della relazione a livello individuale, e in altri casi avrebbe addirittura stimato una relazione di segno opposto [Robinson 1950, 355-357]<sup>20</sup>.

D'altronde l'autore appariva pienamente consapevole dell'importanza del problema individuato, tanto da concludere il suo articolo in modo molto netto:

The relation between ecological and individual correlations which is discussed in this paper provides a definite answer as to whether ecological correlations can validly be used as substitutes for individual correlations. They cannot. While it is theoretically possible for the two to be equal, the conditions under which this can happen are far removed from those ordinarily encountered in data. From a practical standpoint, therefore, the only reasonable assumption is that an ecological correlation is almost certainly not equal to its corresponding individual correlation. I am aware that this conclusion has serious consequences, and that its effect appears wholly negative because it throws serious doubt upon the validity of a number of important studies made in recent years. The purpose of this paper will have been accomplished, however, if it prevents the future computation of meaningless correlations and stimulates the study of similar problems with the use of

---

<sup>20</sup> Mitchell [1976, 432-433] riporta un esempio, costruito ad arte, in cui la correlazione individuale assume un valore di  $r=-0,5$ , mentre la correlazione ecologica assume un valore di  $r=1$ . Pintaldi [2000], nel riassumere il dibattito tecnico sulla spiegazione in termini matematici del fenomeno evidenziato da Robinson, ricorda come la letteratura successiva abbia affermato che il paradosso della “fallacia ecologica” possa essere ricondotto, a seconda dei casi, a uno di quattro tipi di paradossi fondamentali dell'analisi multivariata.

meaningful correlations between the properties of individuals.  
[Robinson 1950, 357]<sup>21</sup>

Con Robinson assistiamo dunque alla concettualizzazione (e dunque alla “nascita” del problema) della *ecological fallacy*. La rilevanza del problema è testimoniata non soltanto dalle conseguenze suscitate dalla sua enunciazione nell’ambito della pratica della ricerca empirica, ma dai filoni di dibattito epistemologico e metodologico che ne sono scaturiti. A questo proposito Pinaroli [2000] individua diverse direzioni di indagine aperte dalla problematizzazione della fallacia ecologica. Tra le altre, da un lato la riflessione sui meccanismi *sostantivi* che portano a diversi risultati a diversi livelli di aggregazione, ovvero “sulla relazione tra l’individuo e il contesto sociale, ossia su come questo ultimo influenzi e condizioni l’agire individuale”; dall’altro riguardo al “ruolo da assegnare ai dati ecologici nella ricerca, una volta assodato che essi non sono l’equivalente dei dati individuali”. A questo proposito, le posizioni fondamentali nel dibattito sono essenzialmente tre. La prima è quella per cui i dati ecologici presentano un valore in sé, in quanto forniscono delle informazioni differenti e complementari ai dati individuali. La seconda, viceversa, considera i dati ecologici rilevanti, in quanto a certe condizioni possono essere impiegati per

---

<sup>21</sup> “La relazione tra correlazioni ecologica e individuale presentata in questo articolo fornisce una risposta precisa riguardo al fatto che le correlazioni ecologiche possano essere validamente utilizzate come sostituto delle correlazioni individuali. Non possono. Anche se è teoricamente possibile che le due siano uguali, le condizioni a cui ciò può accadere sono molto lontane da quelle incontrate correntemente nei dati. Da un punto di vista pratico, dunque, l’unico assunto ragionevole è che una correlazione ecologica è quasi sicuramente diversa dalla corrispondente correlazione individuale. Sono consapevole che questa conclusione ha conseguenze gravi, e che il suo effetto appare totalmente negativo in quanto pone forti dubbi sulla validità di un gran numero di importanti studi pubblicati negli ultimi anni. Lo scopo di questo articolo sarà stato raggiunto, in ogni caso, se impedirà il futuro calcolo di correlazioni senza senso e stimolerà lo studio di problemi simili con l’uso di correlazioni sensate tra le proprietà degli individui.”

stimare i dati individuali. Infine la terza e più radicale posizione afferma che i dati ecologici non hanno valore euristico in quanto non permettono *tout court* la stima di dati individuali [Pinaroli 2000, 87].

#### 1.4. Il modello di Goodman

È precisamente nel solco della seconda posizione metodologica suesposta, ovvero con l’intento di accertare se non esistano particolari condizioni sotto le quali possa continuare a essere possibile operare inferenze ecologiche valide, magari con tecniche diverse dalle precedenti, che Goodman [1953] propone una prima ipotesi di lavoro.

Obiettivo in questo caso non è il calcolo di un singolo coefficiente di correlazione, quanto direttamente la stima, in una tabulazione incrociata tra due dicotomie, dei valori di cella di una tabella simile in tutto e per tutto alla Tabella 1.2. Nel caso di Goodman le due dicotomie sono espresse dalla suddivisione della popolazione rispettivamente nei gruppi N e W (rispetto alla razza), e I e L (rispetto all’alfabetizzazione)<sup>22</sup>. Obiettivo è la stima dei coefficienti  $p$  (definito da Goodman “average probability that a person who is in group N will also be in I” [Goodman 1953, 663]<sup>23</sup>) e  $r$  (similmente, la probabilità di essere analfabeta, ma questa volta riferita ai bianchi).

Ciò premesso, estraendo dalla popolazione appartenente al gruppo N un campione di  $n$  individui, varrà che [Goodman 1953, 66-4]

$$E\{n_i\} = np$$

---

<sup>22</sup> Iniziali rispettivamente di Negro, White, Illiterate, Literate.

<sup>23</sup> “Probabilità media che una persona che è nel gruppo N sia anche in I”.



ovvero che  $E\{u\}$  (il numero atteso di individui alfabeti all'interno del campione di neri) sarà uguale a  $p$  per la numerosità del campione di neri.

Allo stesso modo varrà per il campione relativo ai bianchi, ovvero composto di  $w$  individui appartenenti al gruppo  $W$ :

$$E\{v\} = wr$$

Esaminando il campione complessivo, avremo dunque:

$$E\{u+v\} = np+wr$$

dove  $E\{u+v\}$  è dunque il numero atteso complessivo di alfabeti all'interno del campione estratto.

Dividendo entrambi i membri per  $n+w$  (la numerosità complessiva del campione), otteniamo che la proporzione attesa di alfabeti è

$$E\{Y\} = Xp + (1-X)r \quad (1.5)$$

dove  $X$  è  $\frac{n}{n+w}$ , ovvero la proporzione di neri all'interno del campione complessivo.

Ma la (1.5) può anche essere scritta come

$$E\{Y\} = r + (p-r)X$$

mostrando così la relazione lineare tra  $X$  e  $E\{Y\}$ , caratterizzata da intercetta  $r$  e pendenza  $p-r$ .

Identificando tradizionalmente questi due parametri con  $a$  e  $b$ , avremo:

$$a = r$$

$$b = p - r$$

da cui ovviamente:

$$E\{Y\} = a + bX \quad (1.6)$$

Occorre sottolineare che fino a qui non è stata posta alcuna ulteriore ipotesi: ci siamo limitati a esprimere delle semplici identità, tutt'al più definendo dei parametri alternativi per esprimere con una notazione diversa le stesse identità.

Abbiamo visto come, per conoscere una stima dei coefficienti di cella  $p$  e  $r$ , sia sufficiente stimare i parametri  $a$  e  $b$  della (1.6) relativi a un campione di individui. Ma come è possibile giungere alla necessaria stima di  $a$  e  $b$ ? Qui entrano in gioco le ipotesi di Goodman.

Se, infatti, si ipotizza che, fatte  $n$  osservazioni su  $n$  campioni,  $a$  e  $b$  non differiscano da campione a campione<sup>24</sup> a meno di un piccolo errore statistico non correlato con la variabile indipendente, e

<sup>24</sup> Non è casuale l'uso costante da parte di Goodman del termine *sample* (campione), nonostante che il suo articolo fosse esplicitamente in risposta a Robnson [1950], che non parla mai di campioni, ma sempre, nel concreto, di unità ecologiche. Soltanto ipotizzando una situazione in cui su una stessa popolazione si possono effettuare  $n$  osservazioni, assemblando gli  $n$  campioni con tecniche di randomizzazione applicate scrupolosamente e che non rischino di sovra- o sottorappresentare determinate aree geografiche, si può pensare che l'ipotesi di costanza dei parametri attraverso i campioni sia accettabile.

che la relazione tra  $X$  e  $Y$  attraverso i campioni sia lineare<sup>25</sup>, diventa possibile utilizzare le  $n$  osservazioni per stimare i parametri  $a$  e  $b$  della (1.6) mediante la tecnica della regressione lineare basata sui minimi quadrati.

A queste ipotesi fondamentali riguardo alla costanza dei parametri  $a$  e  $b$  e alla linearità della relazione tra  $X$  e  $Y$  attraverso i campioni, Goodman ne aggiunge una terza, che riguarda la possibilità, oltre che di stimare  $a$  e  $b$ , di ottenere una valutazione della correttezza delle stime effettuate utilizzando le tecniche standard relative alla regressione lineare (test delle ipotesi, intervalli di confidenza): si tratta della condizione che i valori di  $Y$  siano approssimativamente distribuiti secondo una distribuzione normale con la stessa varianza per ogni valore di  $X$  [Goodman 1953, 664]<sup>26</sup>.

L'articolo di Goodman si conclude infine con un'avvertenza significativa, che denota l'intuizione di alcuni dei problemi destinati ad affliggere l'uso del modello:

---

<sup>25</sup> Il passaggio dal derivare la (1.6) tramite identità a livello del singolo campione, a considerarla valida per i diversi valori di  $X$  e  $Y$  è un passaggio decisamente impegnativo. In particolare, il punto fondamentale è che per accettare questo passaggio occorre porre l'ipotesi di linearità della relazione tra  $X$  e  $Y$  attraverso diversi campioni, che è cosa ben diversa dal trascrivere la relazione contabile tra  $X$  e  $Y$  a livello di singolo campione. Ciò che è fondamentale è, ancora una volta, il ritenere di poter considerare delle unità ecologiche come se fossero dei campioni relativi a una popolazione, assumendo che ogni unità ecologica sia egualmente rappresentativa dell'aggregato, e che dunque i coefficienti da stimare assumano lo stesso valore in ciascuna unità ecologica (valore uguale a quello stimato per l'aggregato). È probabilmente questa l'ipotesi più irrealistica del modello: le unità ecologiche non sono campioni, i valori di cella (cioè i coefficienti che legano tra loro  $X$  e  $Y$ ) possono variare notevolmente da un'unità all'altra, e oltretutto – molto probabilmente – in maniera in parte correlata con  $X$ . È molto probabilmente questa l'origine di molte stime assurde fornite dal modello. Per una discussione più approfondita delle tre ipotesi fondamentali del modello di Goodman (costanza dei parametri; linearità; omoschedasticità) vedi Pintaldi [2000, 101].

<sup>26</sup> Si tratta dell'ipotesi di omoschedasticità. A questo proposito vedi anche Goodman [1959, 614], King [1997, 65], Ricolfi [1993, 144-146].

It is possible that both estimates of  $p$  and  $r$  obtained by the regression approach may not be between 0 and 1. For example, we may find from the ecological data that the regression approach leads to the estimates .2 and -.05 for  $p$  and  $r$  respectively. If this result was obtained we would be suspicious of our assumptions [...] although it is possible to obtain such estimates even if these assumptions are satisfied. If these assumptions were satisfied, then we might estimate the value of  $r$  as 0 rather than -.05. We would then determine a new least squares estimate for  $p$  using the equation  $E\{Y\} = Xp$ . [Goodman 1953, 664]<sup>27</sup>

Dal brano riportato emerge la necessità di grande attenzione nei confronti di eventuali risultati inaccettabili. In prima battuta, la presenza di risultati inaccettabili dovrebbe portare a rigettare la ipotesi poste dal modello; soltanto quando, a giudizio del ricercatore, le ipotesi siano soddisfatte è ammessa la possibilità di modificare il risultato e “ripulire” il tutto effettuando una nuova stima basata su queste informazioni “corrette”.

Goodman riprenderà in mano il modello alcuni anni dopo [Goodman 1959], confermando in parte alcuni dubbi su di esso: da un lato sottolineerà nuovamente l'opportunità dell'applicazione del modello quando si possa desumere da altri dati che l'ipotesi di costanza dei parametri attraverso le unità eco-

---

<sup>27</sup> È possibile che le stime sia di  $p$  che di  $r$  ottenute con l'approccio della regressione non siano comprese tra 0 e 1. Ad esempio, potremmo trovare, dai dati ecologici, che l'approccio della regressione porta alle stime 0,2 e -0,05 rispettivamente per  $p$  e  $r$ . Se si ottiene questo risultato, potremmo dubitare dalla validità delle nostre ipotesi [...], anche se è tuttavia possibile ottenere tali stime anche se le ipotesi sono soddisfatte. Se le ipotesi fossero soddisfatte, potremmo stimare il valore di  $r$  in 0 invece che -0,05. Da qui potremmo determinare una nuova stima dei minimi quadrati per  $p$  utilizzando l'equazione  $E\{Y\} = Xp$ . In quest'ultima equazione è infatti assente il termine  $(1-X)r$  visto che  $r$  è stimato 0. Un tale approccio è alla base delle tecniche di “aggiustamento” dei risultati inaccettabili introdotte successivamente, come ad esempio l'algoritmo RAS proposto da Micheli [1976].

logiche<sup>28</sup> non è soddisfatta, dall'altro insisterà sulla assoluta necessità di controllare l'ipotesi di linearità della relazione tra  $X$  e  $Y$  attraverso le unità ecologiche. Inoltre Goodman propone un insieme di tecniche integrative e alternative al modello, utilizzabili in caso di dubbi sull'applicabilità delle ipotesi. Tra queste, la possibilità di invertire la relazione tra  $X$  e  $Y$ , nel caso che i coefficienti della relazione invertita offrano la costanza ipotizzata, e la possibilità di controllo della plausibilità dei coefficienti tramite il c.d. "metodo dei vincoli"<sup>29</sup>.

Il quadro che emerge dai due articoli in cui Goodman espone il suo modello è quello dunque, nella stessa visione del suo autore, di una netta delimitazione dei casi in cui è possibile continuare a fare inferenza ecologica senza incorrere nelle distorsioni denunciate precedentemente da Robinson. Tutto ciò a patto di porre ipotesi ben chiare e, tuttavia, in assenza di tecniche davvero affidabili per il controllo dell'accuratezza delle stime. Gli strumenti per la diagnostica degli errori sono infatti quelli tradizionali della regressione lineare, che offrono sì una valutazione dell'accuratezza delle stime, ma a patto che le ipotesi richieste dalla regressione siano comunque soddisfatte. Tuttavia quest'ultimo punto è spesso molto difficile da accettare. È per questo che il modello di Goodman pone il ricercatore nella posizione di poter utilizzare i risultati dell'elaborazione in una situazione di incertezza maggiore, e con margini di discrezionalità non trascurabili.

<sup>28</sup> In questo articolo del 1959 non si parla più di campioni, ma di unità ecologiche.

<sup>29</sup> Si tratta di una tecnica introdotta da Duncan e Davis [1953] che, a partire dai marginali di riga e colonna di ogni tabella relativa a un'unità aggregata, permette di determinare il valore minimo e massimo possibili di ciascun coefficiente di cella. Come vedremo, tale tecnica è alla base dell'approccio EI.

## 2. Lo studio dei flussi elettorali in Italia

Nonostante la presenza degli interrogativi visti nel capitolo precedente, il modello di Goodman si è successivamente di fatto imposto, almeno nell'ambito degli studiosi italiani su cui adesso ci concentriamo, come la tecnica standard per la stima dei flussi elettorali a partire da dati aggregati.

I primi e più importanti contributi che si avvalgono di questo modello appaiono nella seconda metà degli anni Settanta; in entrambi l'apparato metodologico-tecnico è frutto del lavoro di Giuseppe Micheli.

### 2.1. Il comportamento elettorale a Milano

Ad aprire la nostra rassegna è un articolo di Renato Mannheim e Giuseppe Micheli, dal titolo *Il comportamento elettorale a Milano*, che compare nel 1976 sulla Rassegna Italiana di Sociologia. Si tratta del primo importante esempio in Italia di stima di flussi elettorali a partire da dati aggregati basata sul modello di Goodman. L'analisi prende in esame le consultazioni del 1972, 1975 e 1976 nelle venti circoscrizioni in cui è suddiviso il comune di Milano, interpretando i risultati sia in base alla composizione occupazionale delle zone che in base al risultato della precedente elezione politica [Mannheimer e Micheli 1976].

Fulcro dell'analisi è proprio l'utilizzo del modello di Goodman per stimare: nel primo caso, il diverso comportamento elettorale

dei diversi gruppi occupazionali (aumento o diminuzione, tra le consultazioni, della propensione di un determinato gruppo occupazionale a votare per un determinato partito); nel secondo caso, la vera e propria *matrice di flusso*, ovvero l'insieme dei movimenti di voto, tra le consultazioni, da ciascun partito verso ciascun altro.

I risultati vengono presentati non però in forma tabellare, fornendo direttamente dei coefficienti, ma sotto forma di diagrammi, nei quali i flussi sono simboleggiati da frecce di tre diversi spessori a seconda della “gradazione di intensità dei flussi” [Mannheimer e Micheli 1976, 633].

Il perché di questa soluzione, nonostante il fatto che il modello di Goodman fornisca soluzioni numeriche, è chiarito nell'appendice del saggio, dedicata a una breve esposizione metodologica del modello. In essa i due autori riportano per intero la matrice di flusso tra le elezioni del 1972 e del 1975 stimata in prima battuta dal modello di Goodman. La tabella, qui riportata nella Tabella 2.1, ci mostra immediatamente la presenza di quei problemi già previsti dall'autore del modello, e in seguito messi in evidenza dai suoi critici.

Nella tabella infatti sono presenti in abbondanza coefficienti inaccettabili, non tanto dal punto di vista sostantivo (ovvero, magari, un passaggio di voti tra due partiti considerato di segno o intensità poco plausibile) quanto dal semplice punto di vista logico-matematico: la tabella ci dice infatti che tra il 1972 e il 1975, a Milano, il 179,9% degli elettori degli “altri partiti di sinistra”<sup>30</sup> è passato al PCI, mentre il -41,1% è passato al PSI, e così via. In altre parole, il modello di Goodman, come d'altronde già previsto e anticipato molti anni prima dal suo autore, può fornire risultati inaccettabili.

<sup>30</sup> Il gruppo comprendeva “PDUP, MPL, Manifesto, e nel 1976 DP” [Mannheimer e Micheli 1976, 622].

Tabella 2.1 – *Sinistra di matrice di flusso a Milano, 1972-75*

	1975					Tot.
	PCI	PSI	SIN	CEN	DC	
1972 PCI	88,1	4,7	1,5	3,7	0,4	1,6
PSI	13,4	93,2	1,8	10,1	-22,7	4,2
SIN	179,9	-41,1	18,6	-74,7	61,8	-44,5
CEN	4,9	5,2	0,3	67,7	12,9	9,0
DC	15,6	8,2	4,0	4,5	66,8	0,9
MSI	-28,6	-6,0	11,7	11,1	55,9	55,9

[Mannheimer e Micheli 1976, 637]

A questo proposito, occorre ricordare che l'opinione di Goodman riguardo a simili “inconvenienti” è che essi rivelano, prevalentemente, che le ipotesi del modello semplicemente non si applicano ai dati analizzati. Sol tanto in alcuni casi, quando i coefficienti siano fuori dal *range* accettabile per una frazione minima, è possibile “forzarli” ad un valore accettabile e rieseguire la stima<sup>31</sup>. Questo ad esempio sembra difficilmente il caso della Tabella 2.1, dove appare molto arduo giustificare la forzatura a 0 (e la successiva accettazione) di un coefficiente che, in un possibile *range* 0-100, assume il valore di -74,7.

D'altronde il fatto che risultati assurdi siano dovuti all'inadeguatezza delle ipotesi è colto dagli stessi Mannheimer e Micheli:

Emerge [...] un grave limite del modello in esame, limite peraltro discendente dalla irrealità delle ipotesi sottostanti [...]. Le stime ottenute possono cioè essere più o meno lontane dal valore reale delle percentuali che ad esse corrispondono. L'errore insito nelle stime fa sì che esse possano anche non soddisfare i vincoli a cui deve soddisfare una percentuale: essere cioè non negative e non superiori a 100. [1976, 637]

<sup>31</sup> Goodman [1953, 664] fa l'esempio di un coefficiente -0,05 che viene forzato a 0. Vedi anche la nota 27 del capitolo precedente.

Tabella 2.2 – Stima di matrice di flusso a Milano, 1972-75, corretta con RAS

	1975					Tot.
	PCI	PSI	SIN	CEN	DC	
1972 PCI	88,5	4,6	1,8	3,0	0,4	1,7
PSI	11,3	76,5	1,8	7,0	0	3,4
SIN	68,9	0	8,2	0	22,9	0
CEN	5,6	5,9	0,5	63,8	14,2	10,0
DC	16,2	8,3	5,1	3,8	65,7	0,9
MSI	0	0	11,0	6,8	40,8	41,4
						100

[Mannheimer e Micheli 1976, 638]

La scelta degli autori, in linea con l'approccio delineato nel corrispondente saggio metodologico di Micheli dello stesso anno [Micheli 1976] è, invece, per non dover rinunciare alle stime, di utilizzare un successivo procedimento di calcolo per "trasformare" la matrice ottenuta con l'applicazione del modello di Goodman in una *matrice stocastica*, ovvero in cui i coefficienti rispettino il vincolo essenziale di essere compresi tra 0 e 1<sup>32</sup>.

Il procedimento in questione è l'algoritmo RAS<sup>33</sup>, che fornisce come risultato una nuova matrice che rispetti i vincoli di avere coefficienti compresi tra 0 e 1, e che sia tuttavia, la "meno distante" dalla matrice originaria. La procedura è adatta allo scopo, ma il problema (ovvero, la centralità di ciò che viene *prima* dell'algoritmo) rimane, ed è ben evidenziato dallo stesso Micheli:

È chiaro comunque che questo come altri algoritmi diventerà tanto più fattore di distorsione quanto minore è l'affidabilità della stima dei parametri delle regressioni lineari. [Micheli 1976, 444]

<sup>32</sup> Occorre notare che Micheli sembra aver chiare le implicazioni metodologiche di una simile scelta: "...troviamo stime negative di frequenze, un non senso che, se pur resta come motivo sufficiente a invalidare la sostanza delle stime, può tuttavia essere corretto, applicando [...] l'algoritmo RAS [...]". [Micheli 1976, 446].

<sup>33</sup> Vedi Micheli [1976, 442-444]; Corbetta e Schadee [1984, 205-224].

Comunque sia, il risultato della correzione è una matrice corretta in modo da non avere più coefficienti inaccettabili (Tabella 2.2).

È su questa tabella, riportata in appendice nell'articolo in questione, che si basano i diagrammi con cui Micheli e Mannheimer espongono i risultati della loro analisi. L'esclusione dal corpo dell'articolo della vera e propria matrice di transizione ottenuta è così motivata dagli autori:

Ben consci dei limiti sostanziali del metodo adoperato, le considerazioni non veteranno sulle stime puntuali ottenibili con esso, ma solamente sulle linee di tendenza [...] da esse emergenti. [Mannheimer e Micheli 1976, 630, nota 11]

Possiamo quindi dire che, alla comparsa dei primi importanti lavori italiani che utilizzano il modello di Goodman, appare già sufficientemente chiaro il problema dell'applicabilità delle ipotesi sottostanti al modello. Tuttavia non si può criticare la scelta dei due autori di utilizzare comunque lo strumento, cercando di utilizzare tutte le cautele metodologiche. Da un lato perché il tema era di grande interesse, in una fase in cui il fenomeno dell'aumentata mobilità elettorale richiedeva urgentemente di essere analizzato (siamo all'indomani delle elezioni politiche del 1975); dall'altro perché il basso numero (e la relativa grandezza geografica) delle unità ecologiche analizzate rendeva in qualche modo prevedibile la presenza di distorsioni.

## 2.2. Le ricerche dell'Istituto Cattaneo e il "boom" del modello

Così il modello di Goodman inizia la propria storia di successo tra gli studiosi italiani – Anastasi et al. [1989, 61] parlano di "vero e proprio boom del modello negli anni ottanta" – con gli articoli di Micheli e Mannheimer citati precedentemente, ma soprattutto con la successiva adozione del modello da parte dell'Istituto Cattaneo

per un'ampia ricerca sul comportamento elettorale in Italia. Il lavoro dei ricercatori del Catraneo porterà alla pubblicazione di numerosi contributi<sup>34</sup>, tra i quali a mio parere occorre menzionarne due in particolare.

Il primo di questi contributi [Barbagli et al. 1979] si pone fatto come il primo lavoro in cui la strategia di ricerca del Cattaneo, legata al modello di Goodman, si delinea in maniera chiara e strutturata. A tre anni dal pionieristico lavoro di Micheli, su molti dei temi in campo è stata fatta una scelta decisa.

Il primo punto è relativo alla scelta della sezione elettorale come unica unità ecologica effettivamente analizzabile (unità più ampie renderebbero le ipotesi inapplicabili).

Il secondo punto è invece relativo all'abbandono della prudenza dei lavori precedenti nel citare coefficienti numerici (Micheli e Mannheim si limitavano a parlare di "linee di tendenza", consoci che in molti casi solo le forzature dell'algoritmo RAS rendevano presentabili risultati altrimenti inaccettabili); si tratta di una scelta che acquista senso in relazione al primo punto, nella consapevolezza che con dati a livello di sezione le distorsioni sono destinate a essere decisamente più contenute.

Il terzo punto è relativo alla consapevolezza che stime errate indicano perlomeno un errore di specificazione del modello e dunque richiedono almeno il tentativo di introdurre variabili aggiuntive "di contesto".

Il quarto punto è infine relativo alla tematizzazione del problema del ricambio nella popolazione elettorale, anche se l'analisi in realtà è basata esclusivamente sui votanti e non sull'intera popolazione di elettori.

Le soluzioni e le scelte presenti in questo primo volume vengono, nell'arco di cinque anni, sistematizzate, ordinate e infine presentate in forma organica nel volume [Corbetta e Schadee

1984] che rappresenta esplicitamente l'approfondimento metodologico relativo alla ricerca.

Il volume di Corbetta e Schadee rappresenta un vero e proprio punto di riferimento per tutte le analisi successive da parte di studiosi italiani che si avvalgono del modello di Goodman. In particolare, vengono trattati in maniera ampia e approfondita tutti i principali temi relativi al modello già posti inizialmente da Micheli, così come altri relativi all'utilizzo concreto del modello per la stima dei flussi elettorali in Italia.

Corbetta e Schadee si soffermano in una lunga e dettagliata esposizione tecnica delle caratteristiche del modello, così come delle ipotesi sottostanti ad esso e del loro significato dal punto di vista statistico e sostantivo. Viene inoltre preso in esame il problema della scelta del livello di aggregazione territoriale su cui effettuare l'analisi, ribadendo la scelta della sezione elettorale come ambito di riferimento. Vengono proposti approcci per la soluzione di problemi di fondamentale importanza riguardanti le variazioni di popolazione tra due consultazioni.

Nei termini di una successiva ricostruzione:

il libro, per l'accuratezza e la precisione con cui vengono appunto esposti i problemi di ordine metodologico e statistico, si presenta come un vero e proprio manuale. Anche per questo il lavoro diventa immediatamente un importante punto di riferimento per tutti gli utenti del modello stesso. [Anastasi et al. 1989, 65]

Il volume d'altronde non nasconde la necessità di un uso accorto del modello, e di un controllo accurato dell'effettiva applicabilità delle ipotesi. Non a caso ben due capitoli dei sette comporgono il volume sono infatti dedicati ai possibili approcci da seguire nel caso che il modello fornisca coefficienti privi di senso. In questo gli autori riprendono di fatto la posizione originaria di Leo Goodman<sup>35</sup>.

---

<sup>34</sup> Per una dettagliata rassegna di questi e altri contributi di studiosi italiani che si occupano del modello di Goodman o lo utilizzano per l'analisi, vedi Anastasi et al. [1989].

---

<sup>35</sup> Goodman [1953, 664]. Vedi anche la nota 27 del capitolo precedente.

È [...] possibile che tali stime forniscano dei coefficienti negativi o maggiori di 1: situazioni entrambe inaccettabili. La comparsa di tali coefficienti mostra che alcune delle assunzioni fatte durante il processo di costruzione del modello non sono in realtà verificare. [Corbetta e Schadee 1984, 183]

Di fronte a questo problema, i due studiosi ipotizzano essenzialmente un *errore di specificazione* [184] del modello: in sostanza, se siamo in presenza di un errore sistematico, ciò avviene a causa dell'effetto di variabili "terze" che influenzano la relazione. La strada proposta è dunque quella di considerare nel modello altre variabili addizionali che spieghino l'errore di stima, in maniera da giungere questa volta a stime corrette. Tuttavia, poco oltre, sono gli stessi autori ad affermare che

la strategia qui presentata per trattare i coefficienti negativi mediante l'uso di variabili addizionali, appare soddisfacente in teoria; ma dal punto di vista pratico presenta non pochi problemi. Normalmente non si conoscono quali sono le variabili addizionali lasciate fuori dalle equazioni, e non sempre si hanno a disposizione le informazioni per individuarle. In pratica si è spesso limitati ad un insieme molto ristretto di possibilità. Inoltre l'uso di molte variabili addizionali porta a modelli nei quali i coefficienti si comportano in maniera assai complessa, il che può rendere l'analisi successiva e l'interpretazione dei risultati assai difficili. [Corbetta e Schadee 1984, 198]

Di fatto, l'unica soluzione facilmente applicabile al problema dei coefficienti inaccettabili rimane quella proposta otto anni prima da Micheli: l'applicazione dell'algoritmo RAS per trasformare una matrice che contiene coefficienti inaccettabili in una matrice che non li contiene più, sostanzialmente eliminando dal modello una certa quantità di informazioni. Tuttavia c'è un'importante differenza con l'approccio di Micheli: gli autori propongono infatti un coefficiente che permetta di misurare la quantità di informa-

zioni eliminate per far "tornare" i risultati: si tratta dell'indice  $VR^{36}$  [Corbetta e Schadee 1984, 56], calcolato in base alle differenze tra coefficienti prima e dopo l'applicazione dell'algoritmo. In questo modo si ha una misura di quanto la matrice corretta differisca da quella originaria 0, a seconda del punto di vista, di quanto la matrice originaria presentasse coefficienti inaccettabili. A questo proposito, gli autori propongono il seguente criterio per l'interpretazione del coefficiente VR:

Valori del coefficiente  $VR^{multi}$  superiori a 0,15 [...] ci sembrano, in generale, inaccettabili; ma ogni valore di  $VR^{multi}$  superiore a 0,10 deve essere guardato con un certo sospetto. [Corbetta e Schadee 1984, 57]

Il fatto di misurare direttamente la forza con cui l'algoritmo RAS *impone* risultati corretti appare una cautela sufficiente per abbandonare le remore con cui Micheli presentava i risultati, ovvero in forma solo qualitativa (le famose frecce di tre possibili intensità): Corbetta e Schadee, seguendo la scelta operata anni prima [Barbagli et al. 1979], passano ad utilizzare direttamente i coefficienti numerici stimati. Ciò, pur se ulteriormente giustificato dal fatto che le elaborazioni sono basate su sezioni elettorali invece che su quartieri, apre forse la strada, in studi successivi di altri autori, a interpretazioni a volte discutibili, basate su cifre e decimali che, abbiamo visto, possono anche nascere da forzature arbitrarie di coefficienti di cui in prima istanza è stato stimato un valore inaccettabile.

In ogni caso, a partire dalla pubblicazione del volume, fioriscono numerose in Italia le analisi sostantive basate sul modello di Goodman<sup>37</sup>. Anzitutto il gruppo dell'Istituto Cattaneo prosegue il

<sup>36</sup> Una novità anche rispetto al precedente lavoro del Cattaneo [Barbagli et al. 1979].

<sup>37</sup> Oltre ai contributi qui citati, una rassegna più completa dei lavori basati sul modello di Goodman, per tutti gli anni Ottanta, è presente in Anastasi et al. [1989].

proprio lavoro in un successivo studio [Corbetta, Parisi, e Schadee 1988] in cui gli autori analizzano un lungo ciclo elettorale in alcune città italiane, giungendo a una classificazione delle elezioni in base ai *pattern* di mobilità elettorale. Ma accanto a questo gruppo anche vari altri studiosi (tra cui Paolo Natale, Roberto Biorcio, Stefano Draghi, Gianni Statera, con i rispettivi gruppi di ricerca) pubblicano lavori basati sul modello. Tuttavia, nonostante la fioritura di ricerche, il consenso sul modello nasconde alcune incertezze destinate presto ad emergere.

### 2.3. Prime incertezze e ...bolle di sapone

Una prima fase di controversia è quella della c.d. “guerra dei flussi”. Si tratta della polemica, esplosa a metà degli anni Ottanta, tra Stefano Draghi e Gianni Statera [Draghi 1985; Statera 1985; 1987] riguardo a stime relative al rapporto tra elezioni politiche del 1983 e referendum sulla scala mobile del 1985<sup>38</sup>. E anche in relazione a questa controversia che Biorcio e Natale, nel 1987, ritornano sul problema dell'applicabilità del modello di Goodman:

L'attendibilità dei risultati ottenuti con questo metodo è però ancora in discussione fra gli studiosi del comportamento elettorale, e non sono infrequenti le polemiche originate da qualche stima della dinamica elettorale che appaia “sorprendente”. Poiché il modello di Goodman [...] risulta tuttora uno strumento difficilmente sostituibile nello studio della mobilità elettorale, riteniamo

---

<sup>38</sup> I gruppi dei due ricercatori ottennero risultati decisamente discordanti, e non c'è da stupirsi, visto che Statera stesso precisò che l'analisi era stata condotta su dati “disaggregati per provincia” (sic). Date le caratteristiche del modello, una simile analisi non è concepibile utilizzando la provincia come unità ecologica, e la matrice non corretta avrà sicuramente presentato svantaggi coefficienti inaccettabili [Anastasi et al. 1989, 67-73]. Sulla scelta della sezione elettorale come, di fatto, unico livello possibile a cui effettuare l'analisi, vedi Barbagli et al. [1979]; Corbetta e Schadee [1984].

poco ragionevole accantonarlo senza ulteriori verifiche e approfondimenti. [Biorcio e Natale 1987, 44-45]

È in questo clima di incertezza che appaiono ulteriori contributi critici nei confronti del modello di Goodman, ma soprattutto sulla possibilità stessa di effettuare stime diacroniche in presenza di un significativo ricambio della popolazione elettorale. In particolare, è Gangemi a rilevare alcuni aspetti problematici del modello e a progettare una ricerca volta a controllare empiricamente la correttezza delle stime. Il frutto di questo lavoro è il volume [Anastasi et al. 1989] dal significativo titolo *Guerra dei flussi o bolle di sapone?*, che ha il notevole merito di individuare finalmente delle possibilità di controllo empirico delle stime fornite dal modello di Goodman, e di procedere poi concretamente al confronto tra stime e dati reali. Dall'analisi emergono tre elementi principali: da un lato, la conferma dell'irrealità di fondo delle ipotesi poste dal modello, in particolare per quello che riguarda l'identità dell'elettorato tra due elezioni successive<sup>39</sup> e la costanza dei parametri attraverso il territorio; in secondo luogo, e si tratta del problema più rilevante, il fatto che spesso i valori stimati sono lontani da quelli reali *pur in assenza di coefficienti inaccettabili*.

D'altronde il lavoro del gruppo di ricerca di Gangemi si inserisce in un clima di acceso dibattito sul modello: nel 1988 e nel 1989 la Società Italiana di Studi Elettorali organizza due seminari dedicati alla discussione del modello di Goodman; le relazioni a quest'ultimo seminario, che vede presenti tutti i principali protagonisti del dibattito, vengono successivamente pubblicate in un

---

<sup>39</sup> Da una meticolosa analisi delle liste elettorali, gli autori individuano come la dinamica del ricambio della popolazione elettorale sia estremamente complessa (a causa non soltanto di decessi e nuovi elettori, facilmente rilevabili, ma soprattutto dello stato delle liste elettorali, frequentemente non aggiornate e che quindi mascherano numerosi trasferimenti), e soprattutto molto più complessa di quanto immaginato teoricamente per avere i presupposti per poter effettuare l'analisi. Questo problema non riguarda ovviamente il solo modello di Goodman, ma qualunque altra tecnica, compresa la tecnica EI, quando utilizzata per effettuare analisi diacroniche.



volume [Mannheimer 1993]<sup>40</sup> molto efficace nel mostrarci il punto della discussione a cavallo tra anni Ottanta e Novanta.

## 2.4. Critiche e perplessità

Per capire come i lavori presentati nel 1989 delineino una situazione mutata rispetto anche solo a pochi anni prima, può essere utile riferire un'osservazione del gruppo di Gangemi:

[in] un confronto tra metodologi sul modello di Goodman [...] tenutosi a Trento nell'ottobre 1985, molte delle comunicazioni discusse, all'interno del gruppo di lavoro di metodologia, presentavano i risultati ottenuti attraverso il modello e non una discussione dei principi metodologici sui quali il modello era basato. [Anastasi et al. 1989, 8]

Gli autori parlano del 1985. Viceversa, nel seminario di quattro anni dopo emerge una ormai diffusa consapevolezza della rilevanza dei problemi che affliggono le stime ottenute con il modello di Goodman.

La prima posizione da presentare è quella degli autori del vero e proprio "manuale" del 1984. Corbetta e Parisi [1993] si limitano prevalentemente a rilevare come le elaborazioni che hanno fornito risultati fuorvianti siano quasi sempre derivate dall'applicazione del modello senza i necessari presupposti: in particolare criticano fortemente la possibilità di effettuare inferenze sull'ambito nazionale a partire da presunte "sezioni campione", ammettendo tuttavia la fondatezza delle critiche relative all'impossibilità di postulare l'identità fisica dell'elettorato tra elezioni diverse. La replica tecnica alle critiche rivolte al modello viene invece lasciata al contributo di Schadee [1993], che risponde alla maggior parte delle critiche (ri-

volte in particolare da Gangemi e dal suo gruppo) riformulandole e sistematizzandole in termini teorici, ma soprattutto difendendo non tanto l'uso in generale del modello di Goodman per stimare i flussi elettorali, quanto l'utilità (effettivamente indiscutibile) di proporre quello strumento, in un determinato momento della disciplina e in risposta a una pressante esigenza di ricerca, come punto di partenza per successivi sviluppi.

Biorcio [1993] e Natale [1993], anzitutto insistono nel confermare la bontà del modello in base al fatto che fornisce delle stime corrette, quando vengono utilizzati dei dati simulati in modo da rispettare le ipotesi del modello. Partendo da quest'analisi, sottolineano come il problema principale sia costituito dall'*errore sistematico di aggregazione*, ovvero della correlazione presente, attraverso le unità ecologiche, tra il coefficiente di transizione stimato e la variabile indipendente. Mediante l'uso di dati simulati, i due autori mostrano come il problema dei coefficienti inaccettabili possa essere attribuito in gran parte a questo tipo di distorsione. Per affrontare il problema propongono due tipi di approcci. Il primo è in realtà quello già suggerito nel "manuale" del 1984: l'introduzione di variabili di contesto aggiuntive all'interno del modello, che riescano a spiegare l'effetto di correlazione geografica. Tuttavia, Natale in particolare concorda di fatto con l'osservazione di Schadee di cinque anni prima (il fatto che nella maggior parte dei casi reali sia praticamente impossibile individuare delle variabili che permettano di dar conto in maniera soddisfacente degli effetti del contesto) affermando che occorre cercare nuove variabili di contesto migliori di quelle al momento disponibili. Il secondo approccio è invece quello di individuare sottogruppi di sezioni elettorali sufficientemente omogenei<sup>41</sup> da poter supporre l'indispensabile costanza dei parametri. L'analisi andrebbe quindi effettuata separatamente su ciascun sottogruppo.

Ricolfi [1993] assume una posizione di fatto intermedia. La sua esposizione si apre con una rassegna sistematica dei principali

<sup>40</sup> Il volume, pubblicato nel 1993, raccoglie "i testi aggiornati e rivisti di relazioni originariamente presentate al seminario [...] svoltosi presso la Fondazione Feltrinelli il 16 ottobre 1989".

<sup>41</sup> Mediante la tecnica della cluster analysis.

problemi che affliggono la stima dei flussi elettorali mediante il modello di Goodman, individuando, qui d'accordo con Biorcio e Natale, l'errore sistematico di aggregazione come problema di gran lunga più importante. La sua contrarietà all'utilizzo dell'algoritmo RAS, in quanto "forzante", e all'introduzione di variabili di contesto, in quanto sostanzialmente impraticabile dal punto di vista empirico, lo porta a proporre da un lato una tecnica integrativa del modello di Goodman, dall'altro un'alternativa decisamente più radicale che ne prescinde completamente. La prima possibilità è quella di introdurre nel modello di Goodman non tanto direttamente delle variabili di contesto, quanto delle variabili, ottenute tramite tecniche di *multidimensional scaling*, che esprimano la "posizione" della singola sezione all'interno di uno "spazio elettorale": multidimensionale definito in base ai risultati nelle singole sezioni, con l'obiettivo di ottenere degli indici che sintetizzino in maniera più efficace gli effetti di contesto complessivi. La seconda possibilità prevede invece, semplicemente, l'abbandono del modello di Goodman e la sua sostituzione con il c.d. "metodo dei genelli", una tecnica basata su presupposti completamente diversi e che permetterebbe addirittura la stima della matrice di transizione di ogni singola sezione elettorale.

Il resto degli interventi si pone invece su una posizione radicalmente critica nei confronti della procedura. Corielli e Weber [1993] propongono ben due modelli, quello multinomiale di Hawkes [Hawkes 1969] e quello di Brown e Payne [Brown e Payne 1986], che possono fungere da alternativa a quello di Goodman. Il gruppo di Gangemi [Anastasi et al. 1993] riprende le critiche già formulate nel volume precedente, in particolare sottolineando come la proposta stessa di una pluralità di tecniche alternative, così come la comune ammissione del permanere dei problemi che rendono le stime di correttezza decisamente opinabile, sia la dimostrazione della natura *convenzionalista* del modello in contrasto con l'uso *realista* che ne viene fatto:

il modello è indissolubilmente legato al nome di Goodman, alla sua personalità e alla sua concezione della logica che era ed è di-

chiaramente convenzionalista [...] il ricorso al "modello di Goodman" è l'equivalente logico di una moneta lanciata in aria per prendere una decisione tra varie alternative. Il lancio, e quindi l'applicazione del modello, non serve per convincere, ma solo per decidere. [Anastasi et al. 1993, 52]

Su posizioni simili è Comero [1993], che presenta gli esiti della replicazione a Monza del controllo empirico effettuato precedentemente in Sicilia dal gruppo di Gangemi. A questo proposito occorre notare che molte delle conclusioni di Gangemi erano state criticate dai sostenitori del modello in quanto relative a un ambito geografico particolare e a un numero troppo esiguo di sezioni elettorali. Le stesse elaborazioni, condotte a Monza su un numero di sezioni molto più ampio, portano tuttavia Comero alle stesse conclusioni del gruppo di ricerca siciliano:

Si ritiene di concludere questa ricerca senza indicare soluzioni ai tanti problemi sollevati, e nemmeno si intende formulare un proprio ricettario di norme. È inutile, e forse anche dannoso, in questa fase produrre "regole" euristiche che avrebbero come conseguenza la scoperta di risultati, i flussi elettorali, che potrebbero essere impiegati a sostegno di una qualsiasi ipotesi di mobilità elettorale nell'analisi politologica, con i rischi che si sono appena evidenziati. [...] I risultati si sono dimostrati tutt'altro che stabili e univoci a conferma della tesi sostenuta da Gangemi di una sostanziale inadeguatezza del modello anche sotto una visione convenzionalista, in quanto "non esistono situazioni empiriche nelle quali siano riscontrabili le condizioni ideali richieste per il buon funzionamento del modello". Se esistono tali difficoltà per giungere convenzionalmente a un uso calibrato del modello, si può immaginare come un suo impiego in chiave realista sia semplicemente insostenibile. [1993, 269-271]

A confermare le perplessità verso il modello giungono i due interventi di Giorgio Galli e di Giuseppe Micheli, che tredici anni prima lo aveva introdotto in Italia. Galli [1993] arriva a negare addirittura la rilevanza dell'oggetto di studio, sottolineando che occorrerebbe concentrarsi sui *saliti* piuttosto che sui *flussi*, visto che

l'aspetto di maggiore rilevanza sono in realtà gli *ex-ante* prodotti dal fenomeno elettorale sul sistema politico: "E' veramente tanto importante analizzare i flussi, quando è il saldo attivo che decide degli sviluppi politici?" [37]. L'intervento di Micheli [1993], che si pone in chiave prevalentemente epistemologica, contiene un riferimento al fenomeno del "ciclo di vita" dei modelli. L'autore si riferisce al mutare dell'equilibrio tra la "razionalità interna" del modello, ovvero la sua coerenza interna e la razionalità con cui da determinate premesse si giunge ad altrettanto determinate conclusioni, e la "razionalità esterna", ovvero la capacità del modello di aderire al mondo, dunque di offrire capacità esplicativa dei fenomeni reali. È proprio il rapporto dialettico tra queste due razionalità, secondo Micheli, a determinare il ciclo di vita del modello: la crescente pressione della "razionalità esterna" obbliga a un "affinamento di intelligibilità", ovvero a un aumento dell'aderenza al mondo che tuttavia comporta una ridefinizione della razionalità interna. Ridefinizione che, oltre una certa soglia, porta inevitabilmente al "collasso della razionalità interna" dovuto all'implosione delle contraddizioni interne al modello [29].

La riflessione di Micheli ci offre in effetti una chiave di lettura diversa sul proliferare di tecniche di "aggiustamento", volte a rendere plausibili i risultati di un modello dalle ipotesi forse non applicabili empiricamente. È proprio la proposta di nuove e più raffinate tecniche di correzione del modello, anche da parte dei suoi più decisi sostenitori, che forse suggerisce come, già all'inizio degli anni Novanta, il ciclo di vita della stima dei flussi elettorali mediante il modello di Goodman presso gli studiosi italiani possa considerarsi in fase di declino.

Non a caso il volume del 1993 rappresenta di fatto, con la sua pluralità ed eterogeneità di voci, l'ultimo contributo rilevante a questo dibattito metodologico. L'impressione è che gli studiosi abbiano in seguito deciso di prendere alla lettera la posizione espressa da Schadee nel suo intervento, volta a riconoscere i problemi sul campo e a immaginare possibili scenari di evoluzione:

Il futuro delle stime dei "flussi elettorali". Per quelli che non hanno mai avuto fiducia nella procedura o nei suoi risultati [...] possono dire che [i dati] non soddisfano le richieste di nessuna delle procedure di stima finora suggerite a causa della variazione di popolazione [...] possono obiettare che l'ipotesi sul comportamento elettorale alla base dei metodi di stima è troppo semplice o sbagliata [...] infine possono sostenere che, in presenza di procedure di stima diverse che sembrano dare risultati diversi, non dovrebbe essere usata alcuna stima di flusso elettorale in argomenti reali finché questi problemi non siano chiariti e finché non si raggiunga un certo consenso.

Per coloro che vogliono continuare a lavorare con dati elettorali aggregati nella speranza di stimare i flussi elettorali e di costruire su di essi qualche modello più complesso [...] l'argomento riguardante la variazione dell'elettorato deve essere chiarificato e su di esso dovranno esser fatti controlli empirici [...] l'esistenza di metodi di stima diversi che portano a risultati diversi richiederà ulteriori ricerche [...] si può obiettare che il MMD [il "manuale" del 1984], che proponeva una procedura relativamente standard per stimare i flussi elettorali, è ormai superato nei suoi aspetti tecnici. Io credo che possa ancora essere utile come punto di partenza per approcci più complessi [Schadee 1993, 318-320]

La visione di Schadee appare decisamente realistica. Il boom italiano della stima dei flussi elettorali mediante il modello di Goodman sembra trascorso, forse anche per l'emergere di nuovi e più urgenti temi di ricerca in una situazione politica ormai in grande evoluzione (siamo all'inizio degli anni '90). Ma la fine del "boom" non significa che lo strumento sparisca dalla "cassetta degli attrezzi" degli studiosi italiani. Significa però che sicuramente è maturata la necessità di una consapevolezza dei limiti del modello, e parallelamente della possibilità di rivolgersi a delle alternative. Non a caso nel 1993 Corbetta e Schadee, autori meno di dieci anni prima del celebre "manuale", pubblicano un articolo [Corbetta e Schadee 1993] in cui ipotizzano di affrontare il problema della stima dei flussi mediante l'approccio log-lineare.

## 2.5. Alcuni spunti di riflessione

Prima di proseguire il nostro percorso, che ci porterà ad esaminare un approccio alternativo al modello di Goodman, occorre tuttavia fermarci brevemente, a suggerire alcuni spunti di riflessione su determinati aspetti che sono emersi dal dibattito fin qui presentato.

L'impressione generale è che in realtà a decretare la fine della grande popolarità del modello di Goodman presso gli studiosi italiani siano stati fattori non tanto legati alla procedura in se stessa<sup>42</sup>, quanto più che altro alla qualità dei dati utilizzati per le analisi.

Le critiche alla procedura infatti possono essere suddivise essenzialmente in due categorie: da un lato quelle che mettevano in evidenza la frequenza di risultati inaccettabili (i famosi coefficienti negativi o maggiori di 1), dall'altro quelle volte a mostrare la sostanziale instabilità della popolazione elettorale, capace di invalidare qualunque stima di flussi diacronici.

Questo secondo filone di critica risulta già palesemente slegato dalla procedura utilizzata; se è impossibile sostenere che la popolazione elettorale è più o meno stabile tra due elezioni, sarà impossibile con qualunque procedura stimare i flussi elettorali. Di fatto si tratta di una critica alla possibilità stessa di stimare i flussi elettorali tra due elezioni in base a dati aggregati.

Ma, da un certo punto di vista, lo stesso primo filone di critica è in una qualche misura legato alla qualità e al tipo di dati utilizzati. Occorre infatti notare come gli esempi in cui il modello di Good-  
man forniva risultati peggiori (alta frequenza di coefficienti inaccettabili) fossero spesso basati non su sezioni elettorali, ma su unità ecologiche più ampie. Ciò è vero già per l'analisi di Mannheim e Micheli [1976], basata su quartieri, ma anche in misura ben più grave per quella di Statera [1985], basata addirittura su province e non a caso fortemente bersagliata durante la c.d. "guerra dei flus-

si". E non si può negare che tali episodi abbiano gettato sulla procedura un'ombra che ha probabilmente avuto un ruolo importante nel suo successivo declino.

Spiace notare come, nella articolata formulazione originaria del "modello di Goodman nel contesto italiano" [Corbetta e Schadee 1984], fosse già stato previsto un insieme di regole operative relative all'ambito ecologico da utilizzare (necessariamente non più ampio della sezione elettorale) e alle precauzioni necessarie a limitare il più possibile i problemi di instabilità della popolazione elettorale. Se tutte le analisi successive basate sul modello fossero state condotte osservando scrupolosamente queste regole, probabilmente il "declino" di quest'ultimo sarebbe avvenuto in tempi e con modalità diverse.

Questa consapevolezza è assolutamente necessaria nel momento in cui si esaminano approcci nuovi. È infatti indispensabile ricordare che nessuna procedura, per quanto raffinata, è in grado di fornire risultati corretti a partire da dati di cattiva qualità, o a partire da tipi di dati esplicitamente non raccomandati per l'uso con quella procedura.

---

<sup>42</sup> Intendendo la procedura nei due stadi: regressione lineare + algoritmo RAS.

### 3. Oltre il modello di Goodman: dal metodo dei vincoli all'approccio EI

Nel capitolo precedente abbiamo ricostruito brevemente il dibattito svolto in Italia sulla stima dei flussi elettorali in base a dati aggregati mediante il modello di Goodman. Da quel dibattito emergono in maniera piuttosto netta due posizioni principali tra gli studiosi italiani: da un lato, chi ritiene che l'impatto di tali problemi possa essere circoscritto e controllato, quando non addirittura annullato, mediante l'impiego di tecniche opportune; dall'altro, chi invece ritiene che valga la pena di volgersi all'esplorazione di tecniche alternative. Questo secondo approccio è stato quello seguito da altri studiosi all'estero, e in particolare dall'americano Gary King, autore di un volume [King 1997] nel quale è stato presentato l'innovativo punto di vista che fa da base a questo capitolo. Si tratta dell'approccio che King battezza EI, con esplicito riferimento al termine generale *inferenza ecologica*. Vedremo come EI possa essere definito tanto un' *approccio* che una *tecnica di stima* vera e propria. Con la prima accezione (esaminata in questo capitolo) ci riferiamo a un approccio metodologico generale, che può essere compatibile con varie tecniche di stima; con la seconda (che vedremo nel capitolo successivo) alle specifica tecnica di stima proposta dallo studioso americano. Iniziamo quindi dalla presentazione dell'approccio EI.

Tabella 3.1 – Incognite in una matrice di Jasso tra due dicotomie in un'unità ecologica

	$t_2$		
	Sinistra	Destra	Totale
$t_1$			
sinistra	?	?	$N_i^s$
destra	?	?	$N_i^d$
Totale	$N_i^s$	$N_i^d$	$N_i$

### 3.1. Il problema

Come abbiamo ricordato all'inizio del capitolo precedente, l'impossibilità di conoscere direttamente i dati relativi al fenomeno che ci interessa (i flussi elettorali) ci obbliga ad adottare delle tecniche di *stimazione*.

Per comodità, ripetiamo la definizione del problema presentata nel capitolo precedente. All'interno di un ambito aggregato (ad esempio un collegio uninominale, o un comune), siamo in possesso di  $n$  tabulazioni incrociate come la Tabella 3.1<sup>1</sup>, ciascuna relativa a una unità ecologica (ad esempio una sezione elettorale)<sup>2</sup>.

Il problema può essere così riassunto: vorremmo conoscere i valori assoluti di ciascuna cella, dove vediamo invece un punto interrogativo. Se ciò non è possibile per ogni tabella, cioè per ogni sezione elettorale, vorremmo conoscere questi valori almeno a livello aggregato.

<sup>1</sup> Ricordo che, per convenzione, indico con lettere minuscole le modalità relative all'elezione precedente, e con maiuscole quelle relative all'elezione successiva.

<sup>2</sup> Sullo sfondo rimane il problema, presente in ogni analisi diachronica basata su dati aggregati, dell'identità dell'elettorato attraverso le due consultazioni elettorali. Sul problema dell'identità fisica dell'elettorato tra consultazioni diverse e su una proposta di alcuni approcci per tentare di neutralizzare gli effetti vedi Corbetta e Schadee [1984]; viceversa per un approccio al problema decisamente critico, che mette in evidenza come al di là dei saldi nel numero di iscritti a votare tra due elezioni vi siano movimenti di registrazione di elettori molto più complessi e difficili da controllare, vedi Anastasi et al. [1989].

Di fronte alla situazione presentata, che testimonia una *insufficiente quantità di informazioni* disponibili, qualunque procedimento dovrà, inevitabilmente, ricorrere all'*inserimento di informazioni aggiuntive*, in maniera da mettere insieme una quantità di informazioni sufficiente a dare una risposta al nostro interrogativo.

### 3.2. Informazioni presenti nei dati

Poniamo d'ora in poi l'esempio di voler stimare, relativamente a un determinato collegio elettorale, la quota di elettori che ha votato Sinistra in due elezioni successive, disponendo esclusivamente della registrazione, per ogni singola sezione elettorale, dei voti alla Sinistra in entrambe le consultazioni.

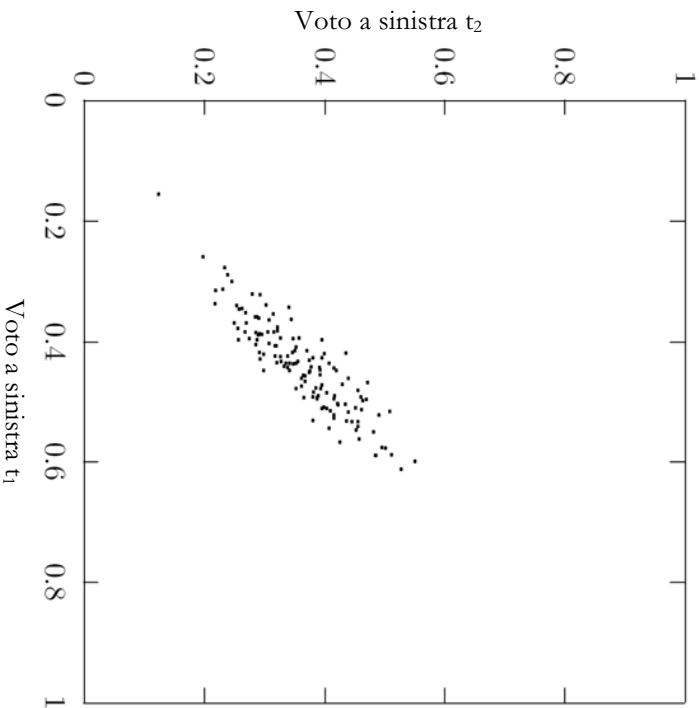
Una prima ricognizione sulle informazioni disponibili porta a una rappresentazione grafica come quella della Figura 3.1.

Si tratta di un semplice grafico a dispersione in cui vengono visualizzati i valori, in ciascuna sezione elettorale, delle due variabili di interesse<sup>3</sup>. Ogni sezione è rappresentata da un punto di ascissa corrispondente alla quota di voto a sinistra nell'elezione  $t_1$ , e di ordinata corrispondente alla quota di voto a sinistra nell'elezione  $t_2$ . Dal punto di vista del contenuto informativo, questo grafico è equivalente a un elenco in cui, per ciascuna sezione elettorale, vengono riportati i valori delle due variabili.

Occorre notare come sia il tradizionale approccio basato sulla "regressione ecologica" che il successivo modello di Goodman condividano un aspetto fondamentale: le informazioni empiriche utilizzate sono solo ed esclusivamente quelle rappresentate nella figura.

<sup>3</sup> Tutti gli esempi di questo capitolo usano risultati di sezione elettorale relativi alle elezioni politiche 1994 e 1996 (Camera maggioritaria) nel collegio uninominale Firenze 1, usando come "sinistra" il voto ai Progressisti nel 1994 e all'Ulivo nel 1996, e come "destra" tutti gli altri.

Figura 3.1—Voto a sinistra in due elezioni nelle sezioni elettorali di un collegio



Partendo dal presupposto che ogni stima deve basarsi su un insieme di informazioni di origine empirica (i dati disponibili, ancorché insufficienti) e su un altro insieme di informazioni di origine teorica (le ipotesi necessarie a colmare l'insufficienza informativa), per i due modelli tradizionali le informazioni di origine empirica si fermano qui.

La riflessione di King [1997] parte invece proprio da una domanda che si situa in questa fase dell'analisi: se i dati stessi non contengano altre e più ricche informazioni *deterministiche* che permettano di sapere di più anche senza introdurre subito delle ipotesi teoriche.

Una possibilità viene individuata nell'applicazione del c.d. “metodo dei vincoli”. Si tratta di una tecnica proposta originariamente da Duncan e Davis [1953] e successivamente riportata anche dallo stesso Goodman [1959] come tecnica alternativa da utilizzare nel caso non vi fossero i presupposti per l'applicazione del suo modello.

### 3.3. Il metodo dei vincoli

Vediamo di capire la logica di funzionamento del metodo dei vincoli basandoci sugli esempi finora utilizzati. Un utile punto di partenza può essere la prima delle identità 1.4 (vedi il capitolo 1):

$$N_i^S = b_i^{SS} N_i^S + b_i^{dS} N_i^d$$

Ancora una volta, questa identità è di semplice interpretazione: nella sezione elettorale  $i$ , il totale di voti alla Sinistra nella seconda elezione ( $N_i^S$ ) corrisponde alla somma di due componenti: il voto alla sinistra nella prima elezione ( $N_i^S$ ) moltiplicato per il coefficiente di flusso da sinistra nella prima elezione a Sinistra nella seconda ( $b_i^{SS}$ ), più il voto alla destra nella prima elezione ( $N_i^d$ ) moltiplicato per il coefficiente di flusso da destra (nella prima elezione) a Sinistra (nella seconda) ( $b_i^{dS}$ ).

Passiamo adesso da valori assoluti a proporzioni, definendo  $X$  la proporzione di voti alla sinistra nella prima elezione e  $Y$  la proporzione di voti alla Sinistra nella seconda elezione:

$$\left. \begin{aligned} X &= \frac{N_i^S}{N_i} \\ Y &= \frac{N_i^S}{N_i} \end{aligned} \right\}$$

Trattandosi di proporzioni (comprese tra 0 e 1) e visto che abbiamo a che fare con dicotomie, avremo anche:

$$1 - X = \frac{N_i^d}{N_i} \\ 1 - Y = \frac{N_i^p}{N_i}$$

e dunque possiamo riformulare il problema in termini di proporzioni e non più di voti assoluti (Tabella 3.2).

Tabella 3.2 – Definizione di coefficienti di flusso in un'unità ecologica

	$b_j$		
$T_j$	Sinistra	Destra	Totale
sinistra	$b_j^{ss} X_j$	$(1 - b_j^{ss}) X_j$	$X_j$
destra	$b_j^{ds} (1 - X_j)$	$(1 - b_j^{ds}) (1 - X_j)$	$(1 - X_j)$
Totale	$Y_j$	$(1 - Y_j)$	1

Trattandosi di due dicotomie, i coefficienti e i valori relativi alla colonna Destra  $b_j$  sono completamente determinati se si conoscono quelli relativi alla colonna Sinistra  $t_j$ . Di conseguenza, l'intera tabella è determinata dalle relazioni della prima colonna; relazioni che si sintetizzano nella seguente identità:

$$Y_j = b_j^{ss} X_j + b_j^{ds} (1 - X_j) \quad (3.1)$$

Apparentemente, la 3.1 ci appare una semplice tautologia. Tuttavia, non possiamo non notare come essa ci mostri l'esistenza di una relazione che lega i marginali di riga e di colonna con i possibili valori che possono assumere i coefficienti di cella. Per capire meglio, proviamo a fare degli esempi.

Il primo caso può essere quello di una sezione in cui la Sinistra abbia ottenuto il 100% dei voti in entrambe le elezioni. In questo caso, sostituendo  $X$  e  $Y$  nella (3.1), otteniamo:

$$1 = b_i^{ss} \cdot 1 + b_i^{ds} \cdot (1 - 1) \\ \Rightarrow 1 = b_i^{pU} + b_i^{ds} \cdot 0 \\ \Rightarrow b_i^{ss} = 1$$

Siamo di fronte a un caso in cui non solo abbiamo qualche idea sui possibili valori assunti dai coefficienti, ma li conosciamo con certezza<sup>4</sup>. E per conoscerli con certezza non è necessario che anche la seconda consultazione si concluda all'unanimità: è sufficiente la prima. Poniamo infatti un caso di sezione in cui la Sinistra sia scesa dal 100% della prima elezione all'80% della seconda:

$$0,8 = b_i^{ss} \cdot 1 + b_i^{ds} \cdot (1 - 1) \\ \Rightarrow 0,8 = b_i^{ss} + b_i^{ds} \cdot 0 \\ \Rightarrow b_i^{ss} = 0,8$$

In questo caso è chiaro che solo l'80% dell'originario voto a Sinistra si è confermato nella seconda elezione.

Ovviamente in questi due esempi abbiamo avuto a che fare con casi in cui il voto nella prima elezione è completamente omogeneo. Non appena perdiamo l'omogeneità nella prima consultazione, perdiamo anche la possibilità di determinare automaticamente la coppia di coefficienti che vogliamo stimare. Ciò non si-

<sup>4</sup> Sostituendo nella (3.1) il valore di  $b_i^{ss}$  così ottenuto, si ottiene anche quello di  $b_i^{ds}$ , che è ovviamente indeterminato, visto che non esistono elettori di Destra né nella prima né nella seconda consultazione.



gnifica tuttavia che venga a mancare qualunque informazione riguardo ai possibili valori di questi coefficienti.

Facciamo questa volta un esempio un po' più sensato: una sezione elettorale in cui la Sinistra sia passata, tra le due elezioni, dall'80% al 60%. In questo caso abbiamo che:

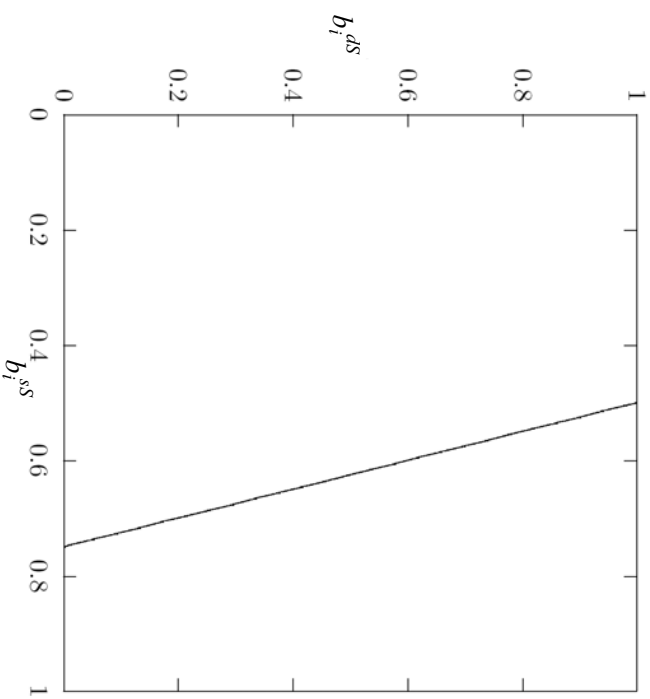
$$\begin{aligned} 0,6 &= b_i^{ss} \cdot 0,8 + b_i^{ds} \cdot (1 - 0,8) \\ \Rightarrow 0,6 &= b_i^{ss} \cdot 0,8 + b_i^{ds} \cdot 0,2 \\ \Rightarrow b_i^{ds} &= \frac{0,6}{0,2} - \frac{0,8}{0,2} b_i^{ss} \\ \Rightarrow b_i^{ds} &= 3 - 4 \cdot b_i^{ss} \end{aligned} \quad (3.2)$$

In questo caso non conosciamo esattamente nessuno dei due coefficienti oggetto della nostra stima. Tuttavia conosciamo perfettamente la relazione *deterministica* che li lega. Trattandosi di una relazione lineare, possiamo tracciare in un grafico la retta corrispondente.

La visione della rappresentazione grafica della relazione che lega i nostri coefficienti (Figura 3.2) ci dà un'idea decisamente più chiara della situazione. Di  $b_i^{ss}$  e  $b_i^{ds}$  non possiamo conoscere il valore esatto. Tuttavia abbiamo comunque delle informazioni molto rilevanti.

Su  $b_i^{ds}$ , ovvero sul passaggio a Sinistra di chi aveva votato a Destra nella prima elezione, non sappiamo alcunché: questo coefficiente può assumere qualunque valore tra 0 e 1 (percorrendo la retta di vincolo, infatti, si attraversa l'intero range 0-1 sull'asse delle ordinate). Non è un caso: la scarsa numerosità di questi elettori (il 20% nella prima elezione) fa sì che siano difficili da rintracciare nella seconda. Potrebbero essere anche passati tutti a Sinistra, o rimasti tutti a Destra: entrambi i risultati sarebbero compatibili con i dati a nostra disposizione.

Figura 3.2 – Vincolo relativo ai valori dei coefficienti  $b_i^{ss}$  e  $b_i^{ds}$  in un'unità ecologica



Disponiamo invece di informazioni molto più precise su  $b_i^{ss}$ , ovvero sulla proporzione di elettori di Sinistra che hanno confermato la loro scelta in entrambe le elezioni. I valori massimi e minimi che può assumere questo coefficiente sono infatti rispettivamente di 0,75 e 0,5 (percorrendo la stessa retta di vincolo, sull'asse delle ascisse si oscilla al massimo tra questi due valori).

Ovviamente i due coefficienti non sono liberi di variare *indipendentemente* nell'intervallo individuato per ciascun coefficiente, ma continuano a essere legati tra loro dalla (3.2). E questo ci permette di chiarire il significato sostantivo delle combinazioni di coefficienti.

Il primo caso limite (l'estremo inferiore destro della retta) è quello in cui  $b_i^{ds}=0$  (nessun trasferimento di voti da Destra a Sinistra). In questo caso, sostituendo  $b_i^{ds}=0$  nella (3.2), otteniamo un

valore di  $b_i^{dS}$  pari a 0,75. Il significato è molto semplice: se la Sinistra non convince nessun vecchio elettore di Destra, tutti i suoi voti sono da attribuire al vecchio voto di Sinistra. Tuttavia la Sinistra vede un calo dall'80% al 60%; dunque, su quattro elettori che avevano votato Sinistra alla prima elezione, solo tre confermano la loro scelta: precisamente lo 0,75.

Il secondo caso (l'estremo superiore sinistro) è viceversa quello in cui la Sinistra convince, alla seconda elezione, *mtt* ( $b_i^{dS}=1$ ) i vecchi elettori di Destra. Ma se c'è una così massiccia attrazione, e tuttavia la Sinistra ottiene un risultato peggiore, significa che c'è stata una massiccia defezione da parte dei vecchi elettori di Sinistra. E infatti, sostituendo  $b_i^{dS}=1$  nella (3.2), otteniamo precisamente 0,5: ha votato a Sinistra solo la metà dei vecchi elettori di sinistra.

La rappresentazione grafica ci ha permesso di comprendere meglio il significato sostantivo dei "vincoli" a cui sono legati i coefficienti. Tuttavia è, per fortuna, possibile derivare delle espressioni generiche che ci forniscono i vincoli presenti sui coefficienti di cella senza dover ogni volta tracciare il diagramma corrispondente. Dati  $X_i$  e  $Y_i$ , si dimostra [King 1997, 302-303] che i coefficienti devono essere compresi nei seguenti intervalli:

$$\begin{aligned} \max \left( 0, \frac{Y_i - (1 - X_i)}{X_i} \right) &\leq b_i^{pU} \leq \min \left( \frac{Y_i}{X_i}, 1 \right) \\ \max \left( 0, \frac{Y_i - X_i}{1 - X_i} \right) &\leq b_i^{qU} \leq \min \left( \frac{Y_i}{1 - X_i}, 1 \right) \end{aligned} \quad (3.3)$$

Ovviamente, le espressioni 3.3 forniscono gli stessi risultati ottenuti precedentemente col metodo grafico.

Occorre a questo punto soffermarsi brevemente per sottolineare come, fino ad adesso, non sia stato compiuto alcun passo che prevedesse stime, ipotesi, teorie o informazioni di tipo probabilistico: i vincoli a cui devono sottostare i coefficienti che cerchiamo

costituiscono un'informazione di tipo assolutamente *deterministica*. La retta visibile nella Figura 3.2 *non è*, ad esempio, "una retta in cui sia massima la probabilità che si trovino i coefficienti?"; è invece la retta su cui *obbligatoriamente* si devono trovare i coefficienti, in quanto se non si trovassero su quella retta sarebbero incompatibili con i marginali di riga e colonna, che non sono affatto stimati, ma costituiscono i dati osservati.

Ciò premesso va detto che, e l'abbiamo visto precedentemente, il "metodo dei vincoli" non costituisce certo una novità: è lo stesso Goodman [1959] a proporre, riportando il procedimento presentato anni prima da Duncan e Davis [1953].

È qui tuttavia che viene introdotto da King un passaggio aggiuntivo. Goodman infatti (ma così anche Duncan e Davis) presentava il metodo dei vincoli come una soluzione di ripiego in circostanze in cui le ipotesi necessarie al suo modello non fossero applicabili. In tale situazione, suggeriva il metodo dei vincoli come una possibilità aggiuntiva per ottenere, se non direttamente una stima puntuale del coefficiente cercato, almeno una precisazione della gamma di valori in cui quel coefficiente poteva variare. In sostanza, *il metodo dei vincoli veniva applicato a livello aggregato*. Le relazioni (3.3) interessavano nella loro validità a livello aggregato, e i valori  $X$  e  $Y$  erano i marginali di riga e colonna relativi all'unica tabella relativa all'aggregazione (nel nostro caso, il collegio elettorale). Tuttavia l'applicazione del "metodo dei vincoli" a livello aggregato è di utilità abbastanza relativa, visto che è alquanto probabile che l'aggregazione porti all'elisione reciproca proprio delle unità ecologiche più omogenee al loro interno, che generano vincoli molto stretti almeno per uno dei coefficienti e sono quindi più ricche di contenuto informativo. Il risultato è che nella maggior parte dei casi si ottengono vincoli molto ampi, di fatto inutilizzabili.

La soluzione proposta da King è invece quella di *applicare il metodo dei vincoli a livello di unità ecologiche* (nel nostro caso, le sezioni) in maniera da aumentare al massimo la quantità di informazione deterministica estratta dai dati, per poi cercare una tecnica opportuna che permetta di sintetizzare questa ricchezza informativa in un coefficiente o in una distribuzione relativi all'intero aggregato.

Con l'aiuto di una rappresentazione grafica, è abbastanza semplice capire l'approccio di King. Immaginiamo di aggiungere nello stesso diagramma più rette di vincolo, ciascuna relativa a una diversa sezione elettorale. Nel caso delle prime due sezioni elettorali (questa volta si tratta di sezioni reali), avremmo una rappresentazione come la Figura 3.3.

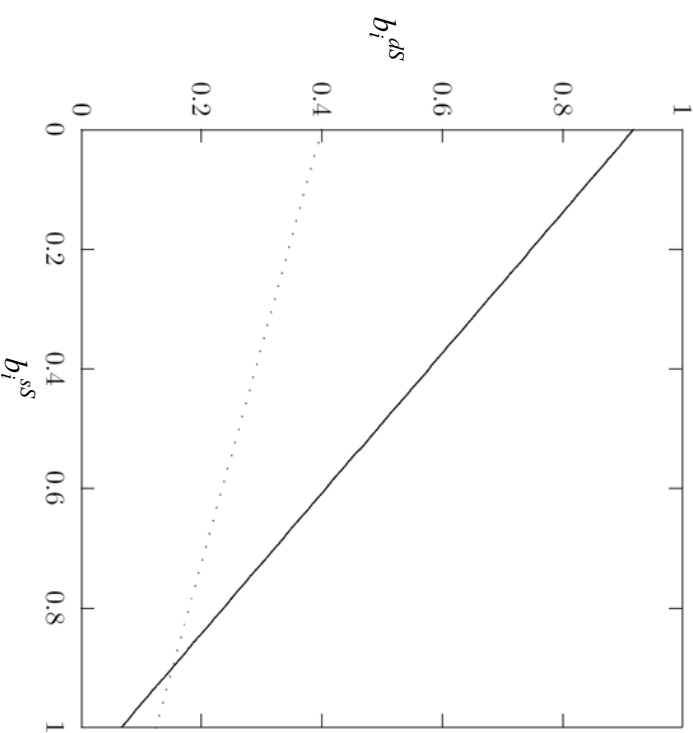
Come si può vedere, la sezione il cui vincolo è disegnato con una linea continua non è particolarmente informativa per nessuno dei due coefficienti; viceversa la sezione tratteggiata è notevolmente informativa per  $b_i^{ds}$ . Ci dice (in maniera, come visto precedentemente, assolutamente *deterministica*) qualcosa di molto preciso sul comportamento, in quella sezione, di chi aveva votato destra alla prima elezione: una percentuale di questi elettori compresa tra circa il 12% e circa il 40% ha infatti deciso di votare a Sinistra nella seconda.

A questo punto (basandoci per ora su sole due sezioni) possiamo dire di aver estratto dai nostri dati, senza ancora porre alcuna ipotesi, la massima quantità di informazione *deterministica* disponibile: tuttavia, almeno per le due sezioni che stiamo esaminando, questa informazione sui vincoli non è in grado di dirci molto da sola.

E qui entra in gioco l'idea di King, che per ora svolgiamo sotto forma di ipotesi molto blanda. Supponiamo infatti, semplicemente, che *i comportamenti elettorali delle varie sezioni di un collegio abbiano una qualche somiglianza tra di loro.*

Se accettiamo questa semplice ipotesi, possiamo compiere un salto concettuale notevole. Possiamo infatti, all'interno della retta di vincolo relativa a una singola sezione elettorale, *non considerare più egualmente probabile* ciascuna combinazione di coefficienti di flusso (che corrisponde a un punto sulla retta di vincolo), *ma assumere a ciascun punto della retta di vincolo una probabilità diversa in relazione alle rette di vincolo osservate nelle altre sezioni elettorali.* In altre parole, e con riferimento alla Figura 3.3, possiamo considerare più probabili quelle combinazioni di coefficienti che, in ogni sezione elettorale, sono, all'interno della retta di vincolo, più vicine alle rette di vincolo delle altre sezioni.

Figura 3.3 – Vincolo relativo ai valori dei coefficienti  $b_i^{ds}$  e  $b_i^{ss}$  in due unità ecologiche



Nel nostro caso, guardando una singola sezione non possiamo fare nessuna ipotesi su quale regione della retta abbia maggiore probabilità di individuare i reali coefficienti di flusso; guardandone due, e ipotizzando che in entrambe si abbia un comportamento almeno simile, ritreremo più probabili i punti collocati sul lato destro delle due rette, dove queste due si intersecano.

Occorre precisare che un'intersezione tra due rette di vincolo relative a una coppia qualsiasi di sezioni elettorali può benissimo *non esistere*. Non stiamo infatti ponendo l'ipotesi (di fatto equivalente a quella di Goodman) che esista un'unica coppia di coefficienti che individua il comportamento elettorale che si è verificato

in tutte le sezioni elettorali<sup>5</sup>: stiamo semplicemente ponendo l'ipotesi che, esaminando due o più rette relative a più sezioni elettorali, sia possibile considerare più probabili quelle combinazioni di coefficienti individuate dalle regioni di ciascuna retta più vicine alle altre.

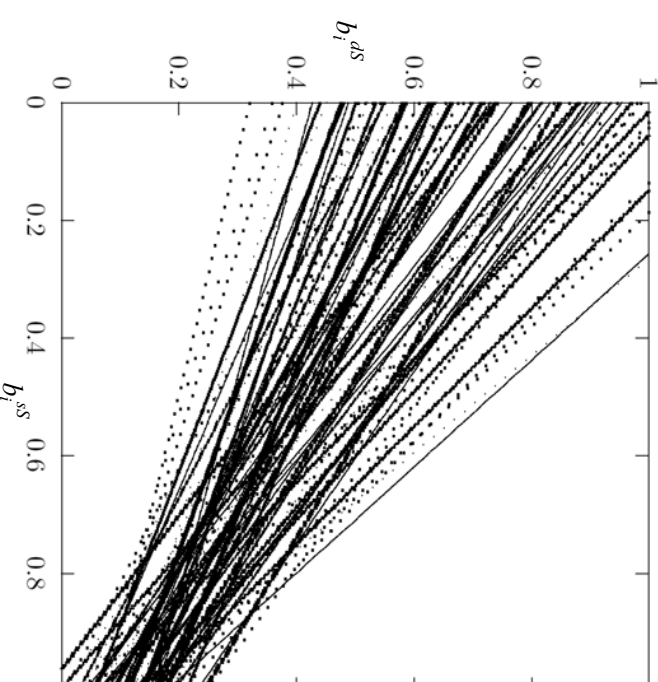
### 3.4. La tomografia

Ovviamente finora abbiamo considerato soltanto due sezioni elettorali, e dunque l'informazione aggiuntiva che ci porta a considerare più probabili le regioni più a destra delle due rette è poca, e di scarsa affidabilità. Lo sviluppo naturale, a questo punto, del nostro procedimento è quello di *tracciare in un unico diagramma le rette di vincolo relative a tutte le sezioni elettorali del collegio in esame*. Questo diagramma viene chiamato da King *tomografia* (per motivi che vedremo in seguito); l'esempio relativo ai nostri dati è visibile nella Figura 3.4.

Un diagramma come quello della figura rappresenta una efficace sintesi grafica di tutte le informazioni *deterministiche* relative ai possibili valori dei coefficienti che intendiamo stimare. Come abbiamo visto, ciascuna retta rappresenta il luogo dei possibili valori dei coefficienti di flusso in una singola sezione elettorale: osservare le diverse rette ci permette immediatamente di farci un'idea riguardo all'eterogeneità dei vincoli (e quindi dei marginali di riga e colonna) tra le varie sezioni elettorali dell'ambito aggregato in esame.

<sup>5</sup> Il caso in cui non solo esista con certezza un'intersezione tra due rette di vincolo qualunque, ma tutte le rette di vincolo si intersechino nello stesso punto corrisponde precisamente all'ipotesi "forte" di Goodman, ovvero che i coefficienti siano esattamente gli stessi in tutte le sezioni elettorali. Groffman e Merrill [2002, 17; appendice 1] dimostrano come in questo caso i coefficienti stimati mediante l'approccio di King coincidano esattamente con quelli stimati mediante il modello di Goodman.

Figura 3.4 – Tomografia di un collegio elettorale



Il principio fondamentale su cui si basa l'approccio EI è in sostanza quello di inserire sì delle ipotesi teoriche, ma a partire da un contenuto di informazione decisamente più ricco, che è quello dice almeno due cose. La prima è che *non esiste* un'unica intersezione di tutti i vincoli. In altre parole, non esiste un'unica combinazione di coefficienti che sia compatibile con i vincoli dei marginali in *tutte le sezioni elettorali*: in altre parole, i coefficienti di flusso devono per forza essere leggermente diversi da una sezione all'altra. Tuttavia la tomografia ci fornisce anche una seconda informazione: le rette di vincolo non convergono perfettamente, ma tuttavia mostrano comunque un certo grado di convergenza verso un punto che si trova abbastanza a destra (circa 0,8 di ascissa) e in basso (circa 0,2 di ordinata). L'idea è che, se affermassimo che, in

media, l'80% degli elettori di sinistra della prima elezione ha confermato il voto nella seconda, e che circa il 20% degli elettori di destra si è spostato a sinistra, faremmo un'affermazione che sarebbe per lo meno abbastanza *compatible* con i vincoli osservati.

È chiaro però che, perché si possa parlare di coefficienti comuni a tutte le unità ecologiche, le rette di vincolo devono presentare almeno un certo grado di convergenza. Se infatti, ad esempio, avessimo due sole sezioni elettorali, con vincoli tali da generare due rette di vincolo *parallele*, non saremmo in grado di fare nessuna congettura sui possibili valori dei coefficienti. Di qui veniamo al criterio fondamentale perché si possa applicare l'approccio EI: la pendenza delle rette di vincolo relative a ciascuna unità ecologica deve essere *il più possibile eterogenea* all'interno del ambito aggregato.

Questo requisito è complementare con quello, enunciato precedentemente, di avere sezioni elettorali molto omogenee al loro interno, in quanto fonti di vincoli più stretti e quindi di maggior contenuto informativo. Di conseguenza, lo scenario ideale è quello in cui abbiamo la *massima omogeneità all'interno di ciascuna unità ecologica*, abbinata però alla *massima eterogeneità tra le varie unità ecologiche all'interno dell'unità aggregata*.

In sostanza, sarà molto più ricco di informazioni un collegio elettorale che presenti numerosi "feudi" in cui ciascun partito gode di un consenso molto alto, piuttosto che un collegio in cui in ciascuna sezione elettorale è riprodotta più o meno, senza grosse variazioni, la proporzione di consensi registrata a livello aggregato.

Questa necessità di eterogeneità tra unità ecologiche come condizione per effettuare analisi corrette è all'origine di un parallelismo presentato dallo stesso King [1997, 112-117]: quello con la diagnostica medica per immagini, e in particolare con tecniche come la TAC (tomografia assiale computerizzata) e la PET (tomografia a emissione di positroni). Tali tecniche diagnostiche si pongono l'obiettivo di produrre immagini dell'interno del corpo irraggiandolo con radiazioni di vario tipo, e esaminando gli angoli con cui ciascun raggio viene deviato dai tessuti; in base a un'analisi statistica di queste deviazioni è poi possibile ricostruire l'immagine

dell'interno della parte esaminata. È facile comprendere come, per ottenere un'immagine tridimensionale accurata, sia indispensabile che la parte esaminata venga irraggiata da un'insieme di angolazioni il più possibile eterogenee.

Il caso della PET presenta similitudini ancora maggiori con la nostra situazione: nella PET, infatti, a lasciare tracce da analizzare statisticamente non sono raggi la cui angolazione è scelta dal ricercatore, ma coppie di particelle elementari che si sprigionano dal corpo del paziente che ha precedentemente ingerito un marcatore radioattivo. Tali coppie di particelle escono dal corpo del paziente in direzioni sempre opposte tra loro, ma con un'angolazione rispetto al resto del corpo (e agli strumenti di rilevazione) che è di fatto casuale ed è totalmente fuori dal controllo del ricercatore.

Questa considerazione non è casuale: è proprio l'analogia tra le tecniche utilizzate dalla diagnostica per immagini e l'approccio di King alla stima dei coefficienti di transizione ad aver suggerito a quest'ultimo l'adozione del termine *tomografico*<sup>6</sup> per indicare diagrammi come quello visto in Figura 3.4.

Abbiamo visto dunque come una condizione fondamentale sia che le rette di vincolo presenti nella tomografia relativa all'ambito aggregato abbiano una pendenza il più possibile eterogenea. Ma da che cosa è determinata la pendenza della retta di vincolo?

La (3.1) ci fornisce l'identità valida in ogni sezione:

$$Y_i = b_i^{SS} X_i + b_i^{DS} (1 - X_i)$$

in realtà, come possiamo facilmente vedere dalle tomografie precedenti, le rette di vincolo sono sì relative a questa relazione lineare, ma riformulata in maniera da tracciare  $b_i^{DS}$  in funzione di  $b_i^{SS}$ .

<sup>6</sup> Tomography plot nell'originale inglese.

Con alcuni passaggi, otteniamo dunque la stessa relazione nella formulazione che ci interessa:

$$\begin{aligned} Y_i &= b_i^{ss} X_i + b_i^{ds} (1 - X_i) \\ \Rightarrow b_i^{ds} (1 - X_i) &= Y_i - b_i^{ss} X_i \\ \Rightarrow b_i^{ds} &= -\frac{X_i}{1 - X_i} b_i^{ss} + \frac{Y_i}{1 - X_i} \end{aligned} \quad (3.4)$$

Esaminando la 3.4 notiamo d'altronde che la pendenza della retta di vincolo corrisponde a

$$-\frac{X_i}{1 - X_i}$$

Anzitutto, considerando che  $X_i$  è compreso tra 0 e 1, la pendenza sarà sempre e comunque negativa. In secondo luogo, tale pendenza dipende esclusivamente da  $X_i$ . Ciò significa che l'eterogeneità richiesta nella pendenza delle rette di vincolo corrisponde, nei dati, all'eterogeneità della variabile  $X_i$ , attraverso le varie sezioni elettorali. Nella nostra interpretazione elettorale,  $X_i$  esprime i consensi ottenuti nella prima elezione dal partito o dallo schieramento che stiamo analizzando: di conseguenza potremo affermare, presupponendo che l'eterogeneità del collegio nella prima elezione sia più o meno simile a quella nella seconda elezione, che otterremo stime tanto più accurate dei coefficienti di flusso quanto più sarà eterogeneo il nostro collegio elettorale riguardo ai consensi ricevuti (alla prima elezione) dal partito o schieramento che stiamo analizzando.

In questo senso, occorre sottolineare come, a differenza del modello di Goodman, l'approccio di King finora visto offre strumenti diagnostici per capire immediatamente *se è opportuno procedere alla stima*. Mentre per il modello di Goodman l'unico modo per

rendersi conto che le ipotesi erano inapplicabili era quello di ottenere coefficienti inaccettabili (inferiori a 0 o superiori a 1), ma non esisteva modo per rendersi conto se coefficienti formalmente corretti fossero in realtà errati dal punto di vista sostanziale, qui siamo in presenza di uno strumento che permette immediatamente di rilevare se i dati offrono le caratteristiche necessarie ad effettuare la stima dei coefficienti di flusso<sup>7</sup>. In sostanza, la tomografia offre uno strumento diretto per capire se è possibile applicare l'approccio EI, o se viceversa la varianza tra sezioni elettorali non è sufficiente a generare una tomografia informativa. In questo senso, disporre di un criterio di applicabilità rispetto ai dati in esame sembra un deciso progresso rispetto al modello di Goodman.

---

<sup>7</sup> La valutazione dell'eterogeneità delle pendenze delle rette di vincolo è non a caso alla radice dei principali coefficienti diagnostici forniti dalla procedura di stima EI.

## 4. La stima dei parametri di interesse: un approccio, diverse tecniche

Nel capitolo precedente abbiamo quindi visto la struttura generale dell'approccio EI; questo capitolo è invece dedicato alle specifiche procedure di stima.

Ricapitolando brevemente, abbiamo osservato come l'approccio EI preveda anzitutto il calcolo dei vincoli a cui sono soggetti i coefficienti di flusso in ciascuna unità ecologica; in seguito, l'ipotesi generica che le informazioni relative a ogni unità ecologica possano concorrere alla stima dei coefficienti relativi all'ambito aggregato, e che quest'insieme di informazioni sia utilmente rappresentato mediante una *tomografia*.

Il prossimo punto è quindi l'utilizzo di tecniche per stimare, a partire dalle informazioni sui vincoli sintetizzate nella tomografia, eventuali distribuzioni o valori puntuali dei coefficienti di flusso a livello aggregato.

Questo punto è stato volutamente tenuto separato per un motivo molto semplice. Dopo l'apparizione del volume di King, sono apparsi successivamente altri contributi (in particolare Grofman e Merrill [2002; 2004]) che condividono l'approccio, ma propongono, per la stima finale dei parametri, tecniche diverse da quella proposta originariamente da King [1997], in particolare per ridurre la mole dei calcoli.

Viene delineandosi quindi un vero e proprio approccio metodologico che propone una soluzione al problema dell'inferenza ecologica mediante l'aggregazione delle informazioni derivanti

dall'applicazione del metodo dei vincoli a livello di singola unità ecologica. Questo approccio è di fatto generale e compatibile con una molteplicità di tecniche per la stima effettiva dei parametri.

Veniamo ora a presentare alcune delle tecniche proposte per affrontare l'ultimo stadio del procedimento, ovvero la stima vera e propria dei parametri. Per ovvi motivi, la prima tecnica è quella proposta da King [1997].

#### 4.1. Stima di coefficienti distribuiti secondo una normale bivariata troncata

Una volta estratto dai dati (e messo sul campo sotto forma di tomografia), come abbiamo visto, il massimo di informazione relativa ai possibili valori dei coefficienti di transizione, effettuare una stima di possibili valori “medi” di questi coefficienti relativi all'ambito aggregato implica la scelta di un *modello* (e dunque la formulazione di *ipotesi*) riguardo alla distribuzione dei valori dei coefficienti tra le varie sezioni elettorali.

##### *Ipotesi sulla distribuzione dei coefficienti*

L'ipotesi posta a questo punto da King è che i due coefficienti ( $b_j^{sS}$  e  $b_j^{aU}$  nel nostro esempio) si distribuiscano secondo una *distribuzione normale bivariata troncata*.

La differenza tra la distribuzione normale bivariata e la sua versione troncata è semplice: nel primo caso la funzione di densità di probabilità ha dominio nell'intero piano  $[-\infty; +\infty] \times [-\infty; +\infty]$ ; nel secondo il dominio è invece ristretto al quadrato  $[0; 1] \times [0; 1]$ . Nel nostro caso è indispensabile riferirsi a quest'ultima versione della distribuzione normale bivariata, in conseguenza del fatto che valori dei coefficienti esterni al quadrato suddetto non sono accettabili.

##### *Procedure di stima*

Una volta accettata questa ipotesi, si procede alla stima dei parametri che caratterizzano la normale bivariata. Il risultato viene ottenuto tramite una stima di massima verosimiglianza, che ricorre

a una procedura di simulazione di possibili combinazioni di parametri. Il prodotto finale è quindi la vera e propria distribuzione dei coefficienti (normale bivariata troncata) che presenta la massima verosimiglianza rispetto ai dati.

##### *Densità di probabilità a livello di sezione elettorale*

Conseguenza di disporre dell'intera distribuzione di probabilità dei coefficienti nel quadrato  $[0; 1] \times [0; 1]$  è che, “tagliando” dal volume sottostante alla superficie della distribuzione una “fetta” in corrispondenza di ciascuna retta di vincolo, possiamo ottenere  $n$  distribuzioni monovariate di probabilità relative ai possibili valori dei due coefficienti in ciascuna sezione elettorale. In sostanza, se le rette di vincolo della Figura 3.4 definivano un luogo dove, in una singola sezione elettorale, *obbligatoriamente* si dovevano situare i due coefficienti, le distribuzioni monovariate così individuate ci dicono che la probabilità di individuare i coefficienti reali di quella sezione non è la stessa per tutti i punti della retta di vincolo, ma si distribuisce secondo la suddetta distribuzione di probabilità monovariata.

Individuata in ogni sezione elettorale la coppia di coefficienti che massimizza la probabilità definita dalla distribuzione appena “ritagliata”, la media ponderata<sup>1</sup> delle coppie di coefficienti di ciascuna sezione costituisce la coppia di coefficienti a livello aggregato.

##### *Costi e limitazioni della procedura*

La procedura messa a punto presenta alcuni costi e alcune limitazioni.

Anzitutto, la reperibilità del software. La procedura è stata implementata utilizzando il software statistico commerciale GAUSS, di scarsa diffusione rispetto ad altri sistemi come SPSS, SAS o STATA. A oggi non risultano implementazioni di questo algoritmo per questi ultimi sistemi.

<sup>1</sup> In base alla numerosità della sezione.



Di fatto, lo strumento più utilizzato (e quello a cui più spesso fa riferimento la letteratura) è una versione ridotta del software (EzI), resa disponibile pubblicamente dallo stesso autore, che non richiede la presenza del software GAUSS, è liberamente utilizzabile, ma che presenta una flessibilità decisamente minore. In particolare, non rendendo disponibile un vero e proprio linguaggio di programmazione, non rende possibili elaborazioni di tipo *batch*.

## 4.2. Minimizzazione della distanza dalle rette di vincolo

Problemi di questo tipo, uniti ad altre considerazioni che vedremo successivamente, hanno suggerito ad altri (e in particolare a Grofman e Merrill [2002; 2004]) la possibilità di mettere a punto tecniche di stima alternative che, tenuta ferma l'impostazione del problema, ovvero la stima dei coefficienti a partire dalle informazioni sintetizzate nella tomografia<sup>2</sup>, effettuino la stima con procedure più semplici dal punto di vista computazionale, cercando di sacrificare il meno possibile sul piano dell'accuratezza.

Nelle parole degli stessi autori:

Our primary focus in this paper is on offering three “new” methods of ecological inference that are more “quick and dirty” in nature than the sophisticated statistical models offered by King and his colleagues because they are very easy to estimate, e.g., using just an EXCEL spreadsheet. We will demonstrate that each of these “quick and dirty” methods [...] have many of the same

<sup>2</sup> I due autori [Grofman e Merrill 2002; 2004] propongono in realtà ben tre modelli alternativi per la stima dei coefficienti, ma due di essi non operano la massimizzazione in quello che viene definito “spazio di fase ( $\beta^u, \beta^w$ )”, che è quello in cui King definisce che debba essere tracciata la tomografia. Per questo motivo, unitamente al fatto che gli altri offrono minore accuratezza nelle stesse elaborazioni di prova effettuate dagli autori, qui sarà accennato soltanto al primo dei tre modelli.

nice properties as the methods proposed by King [...] [Grofman e Merrill 2002, 5]<sup>3</sup>

In sostanza, partendo come sempre dalla tomografia, Grofman e Merrill semplificano di molto il problema, ponendosi semplicemente l'obiettivo di individuare il punto della tomografia “meno distante” da tutte le rette di vincolo. In termini geometrici, anzitutto viene assunto che questo punto “meno distante” si trovi sulla retta di vincolo dell'aggregato (una normale retta di vincolo, solo che viene tracciata in base ai marginali di riga e colonna dell'unica tabella relativa all'intero aggregato)<sup>4</sup>. Su questa retta, si individua poi il punto che minimizza la somma dei quadrati delle distanze da ciascuna retta di vincolo.

A partire da questa soluzione, è inoltre possibile ottenere anche i valori delle  $n$  coppie di coefficienti relative alle  $n$  sezioni elettorali, semplicemente individuando per ogni retta di vincolo il punto meno distante dalla soluzione aggregata trovata precedentemente.

Tale soluzione è in prima battuta intuitiva, tuttavia pone alcuni problemi. Come è noto, la distanza tra un punto e una retta si misura tracciando la perpendicolare alla retta che passa per il punto. Per misurare quindi la distanza tra un punto  $P$  della retta di vincolo aggregata  $R$  e una retta di vincolo  $r$  di una determinata sezione elettorale occorrerebbe dunque tracciare la perpendicolare a  $r$  pas-

<sup>3</sup> “Il nostro obiettivo principale in questo contributo è quello di offrire tre “nuovi” metodi di inferenza ecologica di natura più brutale rispetto ai sofisticati modelli statistici offerti da King e dai suoi colleghi, per il motivo che sono molto semplici da stimare, ad esempio utilizzando un semplice foglio elettronico EXCEL. Dimosteremo che ciascuno di questi metodi più brutali [...] ha molte delle “buone” proprietà dei metodi proposti da King [...]”.

<sup>4</sup> In effetti i due autori dimostrano che, se esiste un'intersezione di tutte le rette di vincolo, questa si trova nelle coordinate individuate dal modello di Goodman (vedi nota 6), e dunque, ovviamente, anche sulla retta di vincolo dell'aggregato. Il fatto è che, nel caso (di fatto, sempre) in cui un'intersezione non esista, occorrerebbe dimostrare che il punto che dista meno da tutte le rette di vincolo si trova sulla retta di vincolo aggregata.

sante per  $P$ . Il problema è che l'intersezione tra  $R$  e la perpendicolare appena tracciata può cadere all'esterno del quadrato  $[0;1] \times [0;1]$ , dunque essere priva di senso. In questo caso, gli autori suggeriscono di assumere come distanza quella tra il punto  $P$  e il primo punto "l'ecito" della retta  $R$ . La procedura così salva la possibilità di fornire risultati accettabili, ma c'è da sospettare che si giunga a una sottostima dei coefficienti quando sono molto vicini a 1, o a una sovrastima quando sono prossimi a 0 (nel nostro caso, probabilmente, a una generale sottostima della fedeltà di partito o di schieramento).

### 4.3. Criteri di confronto tra diverse tecniche di stima

Un ulteriore contributo, decisamente importante, fornito da Grofman e Merrill è però quello relativo alla definizione di un insieme di criteri sotto cui è possibile confrontare le diverse soluzioni fino ad oggi proposte al problema dell'inferenza ecologica. La precisazione di questi criteri, oltre a indicare con chiarezza le direzioni in cui si dovrà orientare l'evoluzione delle tecniche di stima, ci offre l'occasione per riflettere meglio sull'innovatività delle intuizioni di King e sull'importanza delle differenze con gli approcci precedenti.

I due autori aprono le loro considerazioni con la necessità di individuare i requisiti di una qualunque proposta di *soluzione al problema dell'inferenza ecologica*:

A solution to the problem of ecological inference is a set of points,  $(\beta_i^b, \beta_i^m)$ <sup>5</sup>, one from each ecological unit, each of which is

<sup>5</sup> I due autori utilizzano la notazione originaria di King: i due coefficienti corrispondono a quelli che in questo lavoro sono stati chiamati  $b_i^{s^s}$  e  $b_i^{d^s}$ .

feasible, i.e., each of which satisfies the Duncan-Davis feasibility constraints for that precinct. [Grofman e Merrill 2002, 43]<sup>6</sup>

In questa definizione sta, a mio parere, tutta la rilevanza del contributo offerto dal lavoro di King. Soluzioni al problema sono infatti, secondo la definizione, soltanto quelle che sono in grado di fornire non solo un'unica stima a livello aggregato, ma anche  $n$  stime quante sono le  $n$  unità ecologiche di livello inferiore su cui si è basata la stima. Non solo: tanto la stima aggregata quanto le stime delle unità ecologiche devono sottostare ai vincoli definiti dai marginali di riga e di colonna, in base al metodo dei vincoli.

È interessante notare come, alla luce di questa definizione, il modello di Goodman non costituisca una soluzione valida. Non solo questo modello, che d'altronde non fa uso delle informazioni derivanti dall'applicazione del metodo dei vincoli, *non* fornisce alcuna soluzione relativa alla singola unità ecologica ma soltanto la soluzione aggregata; ma oltretutto non rispetta i vincoli *neanche per la soluzione aggregata*. Infatti, se il modello può addirittura fornire soluzioni fuori dall'intervallo  $[0;1]$ , è chiaro che le soluzioni a maggior ragione non rispetteranno il più restrittivo vincolo<sup>7</sup> di Duncan e Davis.

Definito a quali fatitipie debbano essere applicati i criteri di valutazione, gli autori procedono a esporre i dodici criteri in questione: *accettabilità, completezza, coerenza, sensibilità ai dati, plausibilità sostanziale, parsimonia, replicabilità attraverso le elaborazioni, caratterizzabilità dei parametri, diagnosticabilità, espandibilità, spiegabilità, calcolabilità*.

<sup>6</sup> "Una soluzione al problema dell'inferenza ecologica è un insieme di punti  $(\beta_i^b, \beta_i^m)$ , uno da ciascuna unità ecologica, ciascuno dei quali è accettabile, ovvero soddisfa i vincoli di accettabilità di Duncan-Davis per quell'unità ecologica."

<sup>7</sup> Anche se, nella maggior parte dei casi, i vincoli calcolati a livello aggregato sono decisamente più ampi e meno informativi di quelli a livello di unità ecologica. È questo probabilmente uno dei motivi dello scarso successo della proposta originaria di Duncan e Davis del metodo dei vincoli, che si applicava a livello aggregato.

Occorre precisare come questi dodici criteri siano stati concepiti nell'esigenza di confrontare tra loro il modello di Goodman, la procedura EI/MLE di King, i tre approcci proposti nello stesso lavoro da Grofman e Merrill (dei quali solo uno è qui esaminato) e infine il modello di Freedman et al. [1991], qui non presentato. Di conseguenza alcuni dei criteri, che di fatto erano necessari per discriminare rispetto ad alcune caratteristiche di quest'ultimo modello, non risultano particolarmente discriminanti tra le soluzioni qui esaminate.

### Accettabilità

Con questo termine definiamo il criterio fondamentale enunciato precedentemente nella stessa definizione di soluzione al problema dell'inferenza ecologica: il fatto di *fornire delle soluzioni che rispettino i vincoli calcolati secondo il metodo di Duncan e Davis*. Esempio illustre di possibile soluzione che però non rispetta questo criterio è proprio il modello di Goodman.

### Completezza

Con il termine definiamo il criterio in base al quale la soluzione deve fornire stime *tanto al livello aggregato quanto per ciascuna unità ecologica del livello inferiore*. Ancora una volta, è il modello di Goodman a non rispettare questo criterio.

### Coerenza empirica

Con questo criterio si intende richiedere una sorta di controllo aggiuntivo di coerenza rispetto ai dati empirici, assicurando che le stime a livello di unità ecologica siano compatibili con le stime che si potrebbero effettuare disponendo dei soli dati aggregati. Dal punto di vista formale, si richiede che *l'aggregazione*<sup>8</sup>, opportunamente ponderata, delle stime a livello di unità ecologica fornisca una *stima a livello aggregato compatibile con il vincolo aggregato*.

---

<sup>8</sup> Di fatto la media.

Questi tre criteri costituiscono le caratteristiche fondamentali richieste a una soluzione, e rappresentano di fatto i motivi per cui il modello di Goodman, sotto la definizione restrittiva adottata, non costituisce una soluzione al problema dell'inferenza ecologica.

Una volta individuati i metodi che soddisfano questi criteri fondamentali, i criteri che seguono si pongono come criteri per *preferire uno dei metodi* in relazione agli obiettivi di ricerca e alle caratteristiche ritenute di maggiore importanza.

### Sensibilità ai dati

Con questo criterio definiamo la *necessità che la tecnica non imponga a priori una delle soluzioni rispetto alle altre possibili, a prescindere dai dati*. Il criterio nasce in realtà come discriminante rispetto al modello di Freedman, che "forza" a priori la soluzione in maniera notevole, definendo la soluzione come l'intersezione tra ciascuna rete di vincolo e la bisettrice del primo quadrante, *a prescindere dai dati*. Conseguenza di ciò è che anche in *data set* prodotti artificialmente in maniera da prevedere un'unica intersezione di tutte le reti di vincolo (il caso in cui anche il modello di Goodman individuava perfettamente i coefficienti corretti), l'approccio di Freedman fornisce una soluzione notevolmente lontana dalla realtà.

### Plausibilità sostantiva

Il criterio corrisponde dunque alla *necessità di non porre ulteriori ipotesi*, tanto più se poco plausibili dal punto di vista sostantivo. Si tratta di un criterio strettamente collegato al precedente, anche nella sua esigenza di discriminare rispetto al modello di Freedman. Assunto fondamentale di quel modello è che non esista differenza di comportamento tra i due gruppi definiti dalla variabile  $X$  (nel nostro caso, elettori di sinistra *vs.* di destra, nella prima elezione). Di fatto, l'assunto corrisponde all'attendersi due coefficienti di

flusso uguali per i due gruppi<sup>9</sup>, il che nella maggior parte dei casi è assurdo.

### **Parsimonia**

Il criterio riguarda la complessità della procedura, non tanto dal punto di vista delle risorse di calcolo necessarie, quanto dal punto di vista di *quanti e quanto complessi elementi e stadi successivi di elaborazione siano previsti*. Da questo punto di vista, la stima MLE di King è decisamente inferiore a altri metodi più semplici.

### **Replicabilità attraverso le elaborazioni**

Con questo criterio si definisce *il fatto che ripetute applicazioni della procedura agli stessi dati forniscano sempre gli stessi risultati*. La rilevanza di questo criterio è di fatto legata alla stima di massima verosimiglianza di King: si tratta infatti dell'unico approccio che effettua la stima tramite simulazione, osservando i risultati ottenuti da diverse combinazioni di parametri. Il risultato è che, eseguendo ripetutamente la procedura con gli stessi dati, i risultati possono differire leggermente tra le varie esecuzioni. In pratica, l'alto numero di estrazioni casuali imposto dalla procedura per ciascuna stima fa sì che le differenze tra stime successive siano di fatto trascurabili, come precisano gli stessi autori.

### **Caratterizzabilità dei parametri**

Con questo criterio si definisce *la possibilità di ottenere stime affidabili degli intervalli di confidenza relativi ai coefficienti stimati*. Dunque, a prescindere dal fatto che la tecnica in questione fornisca sempre una soluzione, sapere quanto è affidabile la stima fornita.

---

<sup>9</sup> La bisettrice del primo quadrante, inclinata a 45°, definita come soluzione dal modello di Freedman, è infatti il luogo in cui i due coefficienti di transizione  $b_i^{sS}$  e  $b_j^{dS}$  sono sempre uguali tra loro.

### **Diagnosticità**

Con questo termine definiamo *l'esistenza (e facilità di esecuzione) di metodi per il controllo della plausibilità degli assunti della procedura utilizzata*. In particolare, quasi tutte le soluzioni proposte presuppongono un semplice assunto di *unimodalità* della distribuzione dei parametri. Tuttavia, la stima MLE di King è la sola a fornire uno strumento per il controllo di eventuali multimodalità, tramite un modello di stima non parametrico che fornisce una rappresentazione grafica tridimensionale capace di rivelare in maniera abbastanza affidabile l'eventuale multimodalità nella distribuzione dei parametri.

### **Espandibilità**

Si tratta della *possibilità di espandere il dominio del modello per poter analizzare problemi di analisi dei dati di maggiore complessità*, ad esempio tramite l'introduzione di variabili di contesto.

### **Spiegabilità**

Il criterio è relativo alla *facilità di spiegazione* della procedura, tendenzialmente a un pubblico privo di conoscenze statistiche approfondite. Tale esigenza è particolarmente sentita negli Stati Uniti, in considerazione del fatto che, in passato, il modello di Goodman ha svolto un ruolo importante in numerose dispute legali legate alle discriminazioni razziali, in relazione alle previsioni del Voting Rights Act. Da questo punto di vista, uno dei motivi del successo del modello di Goodman in quell'ambito è senza dubbio da individuare nella sua semplicità, che ha permesso che venisse facilmente spiegato a un pubblico di non addetti ai lavori, specialmente nel corso del processo, e che di conseguenza i risultati da esso forniti venissero accettati come veri e propri elementi di prova in sede dibattimentale. In relazione a queste esigenze, è anche facile comprendere la rilevanza del criterio; e forse è possibile individuare i motivi che stanno dietro alla scelta di Grofman e Merrill di cercare tecniche di stima più semplici e "spiegabili" rispetto alla stima MLE di King.

## Calcolabilità

Con questo criterio ci si riferisce infine alla *facilità di generare l'algoritmo (procedimento) necessario al calcolo della soluzione*. Non ci si riferisce di fatto alla mole di calcoli necessaria: possono essere necessari molti calcoli come ripetizione di una procedura estremamente semplice. Tuttavia, il criterio è a mio parere rilevante per la stima MLE di King, in quanto le procedure necessarie sono, nelle attuali implementazioni, di non facilissimo reperimento e uso, e difficili da implementare *ex novo*.

Definiti questi criteri fondamentali, Grofman e Merrill forniscono una loro valutazione di alcune soluzioni alla luce dei criteri. Dalla tabella originaria, qui riportata nella Tabella 4.1, sono state eliminate le soluzioni cui non si fa cenno in questo lavoro.

### 4.4. Diverse tecniche di stima per un unico approccio

In questo capitolo e nel precedente abbiamo quindi presentato sia l'approccio che la tecnica EI, veniamo quindi a una ricapitolazione delle caratteristiche fondamentali di questa proposta operativa.

Punto di partenza dell'intero approccio è l'ipotesi che sia possibile investigare in qualche modo relazioni tra variabili individuali attraverso l'analisi delle loro manifestazioni collettive. È l'approccio dell'*inferenza ecologica* (da cui l'acronimo EI), in riferimento al fatto che a partire da informazioni rilevate a livello ecologico si faccia inferenza su comportamenti individuali. In questo il punto di vista di King si colloca nella tradizione storica che vede prima la nascita della regressione ecologica<sup>10</sup>, e in seguito l'utilizzo del modello di Goodman.

Tabella 4.1 – *Caratteristiche di alcune soluzioni al problema dell'inferenza ecologica*

	modello di Grofman	MLE di King	Grofman e Merrill	Freedman
accettabilità	NO	SI	SI	SI
completezza	NO	SI	SI	SI
coerenza	NO	SI	SI	SI
sensibilità ai dati	SI	SI	SI	NO
plausibilità sostanziale	SI, spesso	SI, spesso	SI, spesso	NO quasi sempre
parsimonia	Alta	Bassa	Media	Alta
replicabilità attraverso le elaborazioni	SI	SI, di fatto	SI	SI
caratterizzabilità dei parametri	SI	SI	SI, in parte	NO, in quanto <i>a priori</i>
diagnosticabilità	SI, limitatamente	SI	NO, ma potrebbe utilizzare alcune idee di King	NO, in quanto <i>a priori</i>
espandibilità	SI	SI, ma per ora limitata	NO, ma si potrebbe concepire	NO
spiegabilità	Alta	Molto bassa	Bassa	Alta
calcolabilità	Alta	Bassa	Alta	Alta

[Grofman e Merrill 2002, 52]

Dal punto di vista tecnico-procedurale, tuttavia, l'approccio di King si differenzia notevolmente da quello dei suoi predecessori. Il principale elemento che caratterizza l'approccio EI è infatti l'applicazione del "metodo dei vincoli" di Duncan e Davis [1953] non a livello aggregato, come concepito originariamente dai suoi autori, ma a livello di singola unità ecologica.

L'applicazione del metodo su ciascuna unità ecologica fornisce di conseguenza una quantità notevole di informazioni *anche* sull'aggregato: in qualche modo ogni unità ecologica, con i suoi vincoli diversi dalle altre – ma che tendenzialmente fanno riferimento al medesimo fenomeno rilevato nelle diverse unità ecologi-

<sup>10</sup> Vedi il capitolo 1.

che che compongono l'aggregato<sup>11</sup> – rappresenta dunque un tassello del *processo* che il ricercatore si propone di risolvere.

Le informazioni sui vincoli contenute in ciascuna unità ecologica possono essere rappresentate tutte insieme in un diagramma chiamato *tomografia*, che fornisce informazioni dettagliate per ciascuna unità ecologica, ma al tempo stesso una chiara e netta visione d'insieme.

Definito questo nucleo fondamentale che rappresenta il “cuore”, decisamente innovativo, dell'approccio EI, è possibile pervenire alle stime vere e proprie dei valori di cella utilizzando una varietà di procedure di stima.

Una prima procedura per la stima dei coefficienti è stata introdotta dallo stesso King. Aspetto, a mio parere, estremamente importante dell'approccio EI è che tuttavia, successivamente, altri autori hanno, come visto precedentemente, proposto altre tecniche di stima, tutte ugualmente compatibili con l'approccio EI.

In questo, a mio parere, sta il nucleo centrale e l'innovatività della proposta. Tanto maggiore in quanto l'approccio è talmente generale da essere compatibile con un numero potenzialmente infinito di tecniche di stima da applicare sui dati forniti dal “metodo dei vincoli” applicato a livello di unità ecologiche.

A partire da questo punto fermo, si apre infatti un filone di ricerca riguardo al miglioramento dei modelli (e nella fase di stima abbiamo effettivamente modelli veri e propri con le relative ipotesi, ma da applicare a dati decisamente più ricchi e con ipotesi meno restrittive) e delle tecniche di stima dei valori di cella, nonché all'introduzione di nuove tecniche per rendere più veloci ed accessibili elaborazioni che oggi risultano lunghe e spesso di complessa progettazione.

Altro ambito che apre ampi spazi di innovazione è quello dell'estensione del metodo a variabili non dicotomiche: campo che presenta problemi di notevole complessità, ma per il quale già a-

nesso si possono concepire possibili soluzioni. Ad esso è dedicato il capitolo che segue, che costituisce l'innovazione originale di questo lavoro.

---

<sup>11</sup> Ciò che lega le unità ecologiche tra loro è, di fatto, una semplice ipotesi di unimodalità riguardo ai coefficienti di transizione da stimare.

## 5. Oltre le variabili dicotomiche: la generalizzazione dell'approccio EI

Come abbiamo visto nei due capitoli precedenti, non solo la tecnica EI, ma lo stesso approccio generale su cui si basa, ovvero l'applicazione del metodo dei vincoli su unità ecologiche, nasce e si sviluppa come strumento di indagine relativo alle relazioni tra due variabili dicotomiche.

Non si tratta solo di un punto di partenza, o di una semplice esigenza di semplicità. Esistono dei problemi sostantivi che hanno ispirato la messa a punto della tecnica, e che sono intrinsecamente dicotomici: tra essi il fenomeno del *racial bloc voting* negli Stati Uniti, la cui analisi si basa sullo studio della relazione tra le due dicotomie “bianchi *m.* nerî” e “voto *m.* non-voto”. La necessità di tecniche affidabili per l'analisi di questo specifico fenomeno è particolarmente sentita, in relazione anche al numero di dispute legali riguardo ad esso. Dispute che in numerose occasioni hanno visto ammesse, tra gli elementi probatori della causa, stime basate sulle tecniche allora disponibili, e in particolare sul modello di Goodman [King 1997; Grofman e Merrill 2004; King, Rosen, e Tanner 2004].

Accanto a una forte motivazione sostantiva vanno tuttavia evidenziati gli inevitabili aspetti tecnici: la relazione tra due dicotomie costituisce il caso più semplice, e in ogni caso un punto di partenza obbligato per poi tentare di affrontare relazioni tra variabili categoriali a più di due modalità. Considerazioni simili valgono d'altronde per la storia stessa dell'inferenza ecologica: i contributi

visiti nel capitolo 1, a partire dalla “correlazione ecologica” di Ogburn, passando per l’articolo critico di Robinson, fino all’enunciazione originaria del modello di Goodman, si riferiscono invariabilmente alla relazione tra due variabili dicotomiche. La stessa tecnica, vista nel capitolo 2, usata in Italia per la stima dei flussi elettorali a partire dalla fine degli anni ’70, assume come punto di partenza il modello di Goodman originario: solo in seguito avvenne la trasposizione al modello generale lineare, e dunque la riconduzione a un problema di stima di coefficienti di  $n$  equazioni lineari, risolto mediante gli strumenti del calcolo matriciale.

### 5.1. La necessità dell’estensione a variabili politomiche

Ciò premesso, appare chiaramente necessario trovare un sistema per estendere a relazioni tra due variabili categoriali *politomiche* l’approccio EI, e successivamente anche una delle relative tecniche di stima.

È banale ricordarne i motivi. Sarebbe sufficiente riflettere sulla necessità di analisi proprie dei sistemi multipartitici, ma soprattutto sul fatto che le tecniche di inferenza ecologica non si riferiscono solo ad applicazioni elettorali, ma possono potenzialmente essere usate per analizzare qualunque tipo di relazione tra variabili categoriali. Inoltre, anche per i sistemi bipartitici, l’analisi delle effettive scelte di voto (oltre la semplice scelta di recarsi a votare) richiede di andare oltre il confronto tra due dicotomie.

In questo caso il problema si esprime facilmente nei termini dell’esempio presentato nella Tabella 5.1.

È chiara la maggiore complessità del problema. Anche calcolando i coefficienti della colonna Altri  $t_2$  in maniera residuale a partire dagli altri coefficienti stimati, i coefficienti incogniti da stimare sono 12, invece che 2.

Tabella 5.1 – *Incognite in una matrice di flusso generale in un’unità ecologica*

$t_1$	$t_2$				Totale
	Sinistra	Centro	Destra	Altri	
sinistra	?	?	?	?	$N_i^s$
centro	?	?	?	?	$N_i^c$
destra	?	?	?	?	$N_i^d$
altri	?	?	?	?	$N_i^a$
Totale	$N_i^s$	$N_i^c$	$N_i^d$	$N_i^a$	$N_i$

Le vie da seguire per affrontare il problema possono essere essenzialmente due. La prima è di ripensare dalle fondamenta l’approccio EI (che nasce sul confronto tra due dicotomie), cercando di riformularlo in termini generali a partire dai suoi principi di base. La seconda, senza dubbio più semplice, è quella di utilizzare in qualche modo sia l’approccio che la tecnica di stima EI (dicotomici), con qualche tipo di procedura a più stadi. Vediamo anzitutto in maggior dettaglio quest’ultima possibilità.

### 5.2. Un approccio a più stadi: uso multiplo della tecnica dicotomica EI

Abbiamo visto che la soluzione più semplice, in prima battuta, è quella di usare la procedura EI esistente. In questo caso l’unica possibilità è quella di applicarla ripetutamente ai vari coefficienti della matrice di flussi da stimare. In termini sostantivi questo corrisponde alla scelta di isolare la relazione definita da ciascun coefficiente di cella, *analizzandola separatamente dalle altre come se fosse una relazione tra due variabili dicotomiche*.

Seguendo l’esempio della Tabella 5.1, possiamo stimare il primo coefficiente in alto a sinistra (ovvero la quota di elettori che votano Sinistra in entrambe le elezioni) immaginando di trasformare le due variabili in due dicotomie: “sinistra  $t_1$  vs. non-sinistra  $t_1$ ” e “Sinistra  $t_2$  vs. Non-Sinistra  $t_2$ ”. In questo modo la Tabella 5.1 si riduce alla Tabella 5.2.



Tabella 5.2 – Dicotomizzazione artificiosa per la stima di un coefficiente

	$t_2$		
$t_1$	Sinistra	Non-Sinistra	Totale
sinistra	?		$N_i^s$
non-sinistra			$N_i^n$
Totale	$N_i^s$	$N_i^n$	$N_i$

Il coefficiente che ci interessa è ovviamente quello evidenziato con il punto interrogativo. Ottenuto questo coefficiente, si procede riproducendo lo stesso procedimento per tutte le celle da stimare, operando ogni volta una dicotomizzazione artificiosa rispetto alla coppia di modalità che stiamo esaminando.

Questo procedimento permette di utilizzare in modo molto semplice<sup>1</sup> con variabili categoriali *politomiche* tecniche nate originariamente per analizzare relazioni tra due dicotomie.

Il punto fondamentale è però che, in questo caso, la stima di ciascun coefficiente avviene *independentemente da tutti gli altri*, e in uno stadio separato della procedura. E per questo motivo che si tratta inevitabilmente di un approccio *a più stadi*.

### 5.2.1. Incompatibilità reciproca dei coefficienti

Questo ha conseguenze importanti. Infatti in questo modo la stima di ciascuno dei coefficienti avviene in una condizione di totale “isolamento informativo” rispetto al resto della matrice. Degli altri coefficienti, appena stimati o ancora da stimare, non si sa alcunché; di fatto non si sa neanche quante modalità abbiano realmente le due variabili in gioco, visto che ciascuna di esse viene ricondotta artificiosamente a una semplice dicotomia.

Consanguenezza abbastanza ovvia di questa carenza informativa è che i coefficienti così stimati in generale *non* sono compatibili tra loro [Ricolfi 1993, 174]. In sostanza, la somma dei coefficienti di una stessa riga o colonna non è quasi mai uguale al totale di riga o di colonna, anche se può esservi molto vicina.

<sup>1</sup> Anche se, nella pratica, abbastanza macchinoso.

Questo tipo di incoerenza potrebbe ricordare i problemi visti in precedenza riguardo al modello di Goodman. In realtà non è esattamente così: il problema fondamentale del modello di Goodman non è infatti che esso possa stimare coefficienti incompatibili tra loro, ma che possa fornire stime di coefficienti *inaccettabili anche se esaminate singolarmente*. Esempi classici sono affermazioni del tipo “il 120% dei neri vota repubblicano” oppure “il -12% dei bianchi lavora come domestico”: si tratta di valori privi di senso già se valutati singolarmente. In questo caso invece la situazione non è così grave: il problema è semplicemente di incompatibilità dei coefficienti con i totali di riga o di colonna.

Detto questo, il problema dell’incompatibilità dei coefficienti può forse essere considerato meno grave, ma non può essere ignorato. Già in passato sono state concepite [Ricolfi 1993]<sup>2</sup> alcune strategie per limitarne gli effetti.

#### Stima residuale di alcuni coefficienti

Il primo di questi approcci prevede che vengano stimati tutti i coefficienti necessari eccetto quelli appartenenti all’ultima riga e all’ultima colonna, che vengono determinati in maniera residuale, ovvero sottraendo dal totale di riga o di colonna la somma dei coefficienti già stimati. Gli appunti che si possono muovere a questo approccio sono essenzialmente due.

Il primo (individuato dallo stesso Ricolfi) è che il metodo non è simmetrico, ovvero non agisce nello stesso modo su tutti i coefficienti: i coefficienti appartenenti all’ultima riga e all’ultima colonna vengono infatti trattati in maniera completamente diversa dagli altri. Ciò può essere considerato nella maggior parte dei casi tollerabile, visto che si può assegnare (e sovente si assegna) all’ultima modalità un significato residuale anche dal punto di vista sostantivo, raccogliendovi tutti i casi che non rientrano nelle altre, più definite dal punto di vista concettuale. Bisogna però notare come, ad

<sup>2</sup> Ricolfi propone queste strategie relativamente al “metodo dei gemelli” da lui proposto, anch’esso relativo a due variabili dicotomiche.

esempio in ambito elettorale, la modalità residuale non sempre è quella caratterizzata da minor significato sostantivo: nel caso elettorale, spesso è proprio la categoria residuale del “non voto” (astenuti, voti non validi, liste minori) a essere oggetto invece di attenzione notevole.

Il secondo appunto è invece relativo al fatto che la semplice stima residuale dell'ultimo coefficiente della riga (o della colonna) può non essere sufficiente a risolvere il problema. Possono infatti comunemente verificarsi casi in cui la somma dei coefficienti già stimati superi il valore di 1 anche prima di raggiungere l'ultima modalità della riga o della colonna. In questo caso la correzione richiederebbe di stimare un coefficiente di transizione negativo, che è inaccettabile.

### Stima con vincoli residuali successivi

Questa procedura, che è come vedremo proposta anche dallo stesso King, prevede invece una stima a stadi successivi in cui ogni stadio di stima “tiene conto” di tutti i coefficienti già stimati negli stadi precedenti.

Una volta stimato il primo coefficiente della riga, questo coefficiente viene posto come vincolo per le stime successive, che possono “muoversi” soltanto nello spazio residuale lasciato dalla stima precedente. Se, ad esempio, il primo coefficiente stimato è di 0,4, le stime successive vengono effettuate come se ci stesse riferendo non all'intera popolazione (1), ma soltanto al residuo di cui ancora non si è stimato il comportamento, in questo caso 0,6. Di conseguenza, le stime successive si muovono in un ambito sempre più ristretto. L'ultimo coefficiente viene infine, ovviamente, determinato in maniera residuale.

Rispetto all'approccio precedentemente esposto si ottiene un vantaggio: la stima mediante vincoli residuali successivi limita l'ambito di stima gradualmente ad ogni passaggio e non soltanto per l'ultimo coefficiente, rendendo impossibili per definizione i rischi di stime “eccessive” da correggere successivamente con altri coefficienti negativi.

Tuttavia c'è un prezzo da pagare: l'aumento notevole dell'asimmetria del procedimento. Se infatti l'approccio precedente comportava che tutti i coefficienti venissero stimati in condizioni identiche, ad eccezione della riga e della colonna residuale, in questo caso l'asimmetria riguarda tutti i coefficienti stimati.

Dal momento, infatti, che la stima di ciascun coefficiente costituisce un vincolo per i successivi, la conseguenza è che *cambiando l'ordine in cui vengono stimati i coefficienti, otterremo quasi certamente risultati diversi per tutti i coefficienti*. Si aprirebbe di conseguenza un ulteriore ambito di problemi, legati ai criteri per la scelta dell'ordine in cui stimare i coefficienti. A questo proposito, è lo stesso King che, pur ritenendo che l'ordine in cui vengono effettuate le stime non sia particolarmente rilevante, suggerisce di legare l'ordine delle dicotomizzazioni successive a un qualche significato sostantivo del problema, possibilmente interpretabile anch'esso come successione di scelte dicotomiche:

It is most convenient if the dichotomization parallels some substantive organization of the problem at hand or even a model for the individual's decision process. This will make it easier to interpret the diagnostic procedures and graphics. However, the ordering of the dichotomies is not necessarily material to the method: so long as one is careful to get good estimates of the early stages, the later stages will be properly estimated. [King 1997, 153]<sup>3</sup>

Un ultimo problema di questa procedura è che può essere applicata solo all'interno di una singola riga, senza poter vincolare con i suoi risultati le righe successive. In altre parole, si ottiene che

---

<sup>3</sup> “La condizione ideale è che la dicotomizzazione rifletta una qualche struttura sostantiva del problema in esame, o addirittura un modello per il processo individuale di decisione. Ciò rende più facile l'interpretazione delle procedure e dei grafici per la diagnosi dell'inaccuratezza delle stime. In ogni caso, l'ordine delle dicotomie non è necessariamente rilevante per il metodo: una volta posta la necessaria attenzione ad avere stime accurate dei primi stadi, anche i successivi saranno stimati correttamente.”

la somma dei coefficienti di ciascuna riga corrisponde al totale, ma non si è assolutamente certi che ciò valga anche per ciascuna colonna.

### **Stima senza vincoli, con applicazione successiva di algoritmi di correzione**

Quest'ultimo approccio è quello che tenta di coniugare la massima libertà nella stima del singolo coefficiente con l'esigenza di ottenere stime compatibili tra loro, evitando in più i problemi di asimmetria che affliggono i due approcci visti precedentemente.

Si tratta, in breve, di procedere anzitutto alla stima non vincolata di tutti i coefficienti della matrice, compresi quelli appartenenti all'ultima riga e colonna.

Otterremo come risultato una matrice in cui, come visto all'inizio del paragrafo, in generale le somme (per riga e colonna) dei valori di cella stimati saranno diverse dai totali di riga e colonna osservati empiricamente.

A questo punto si può però applicare alla matrice un algoritmo di correzione<sup>4</sup> che modifichi ciascuno dei coefficienti di cella della minima quantità necessaria a farli risultare compatibili con tutti i marginali di riga e colonna, suddividendo simmetricamente la correzione su tutti i coefficienti della matrice.

Il vantaggio di un simile approccio è che, in sostanza, non ri-sente di entrambi i difetti che affliggono invece gli altri due visti precedentemente. Il risultato di ottenere somme di celle coerenti con tutti i marginali è garantito per definizione; l'algoritmo inoltre è completamente simmetrico, visto che tutti i coefficienti hanno lo stesso status, e vengono corretti in misura maggiore o minore soltanto in relazione al fatto che appartengano a una riga o colonna i cui valori di cella sono incompatibili coi marginali effettivi più delle altre.

---

<sup>4</sup> È probabile che Ricolfi [1993, 174] pensasse all'algoritmo RAS, che appare assolutamente adatto allo scopo, e che viene abitualmente usato relativamente al modello di Goodman (vedi cap. 2).

Occorre infine notare che l'uso che verrebbe fatto dell'algoritmo di correzione potrebbe configurarsi come meno distrativo, rispetto all'uso che tradizionalmente ne viene fatto per correggere le matrici stimate dal modello di Goodman.

In quest'ultimo caso, infatti, esistono dei coefficienti inaccettabili che vengono *in primo luogo* ricondotti all'intervallo [0;1]. È sulla matrice così "corretta" che viene applicato l'algoritmo RAS, che si occupa semplicemente di "spalmare" su tutta la matrice lo squilibrio prodotto originariamente con la forzatura dei coefficienti inaccettabili.

Nel nostro caso, invece, l'algoritmo avrebbe esclusivamente il compito di operare correzioni minime su una matrice che è composta di coefficienti sempre e comunque accettabili se presi singolarmente, e in cui l'incompatibilità tra coefficienti è da imputarsi solo all'"isolamento informativo" di ciascuna stima rispetto alle altre.

### **Vincoli residuali successivi con imputazione multipla**

Si tratta di una proposta formulata dallo stesso King [1997, 151-153], nel cercare di rispondere alla necessità di estendere la tecnica EI all'analisi di variabili categoriali *politomiche*. La scelta che opera è, di fatto, quella di combinare due degli approcci individuati da Ricolfi, introducendo anche qualche interessante caratteristica aggiuntiva.

La proposta di King si avvale sia della stima mediante vincoli successivi che della stima residuale dell'ultima colonna. King fa infatti riferimento alla stima dei coefficienti di una tabella 2x3: la prima colonna viene stimata liberamente, la seconda assumendo come vincolo i risultati della stima precedente, infine l'ultima in modo residuale.

La caratteristica aggiuntiva interessante riguarda gli strumenti diagnostici presenti nella stima MLE per valutare l'accuratezza delle stime. Per ogni coefficiente, infatti, la procedura MLE di King fornisce abitualmente anche i relativi intervalli di confidenza. Volendo effettuare questa valutazione su una stima effettuata "in seconda battuta", è ovvio constatare come la stima di secondo grado

assuma come vincolo deterministico un qualcosa che deterministico non è, in quanto è a sua volta una stima. L'esigenza quindi è quella di "trasmettere" nel risultato, in una qualche forma, l'informazione relativa al fatto che l'ultima stima è basata a sua volta su una stima soggetta a incertezza.

Il procedimento adottato è semplice ma efficace. Come visto nel capitolo precedente, la stima del coefficiente avviene mediante simulazione, ovvero confrontando i risultati in termini di massimizzazione della verosimiglianza ottenuti con ripetuti valori casuali. La chiave è che i vincoli relativi a ciascun coefficiente, che nel problema originario costituiscono un punto fermo in quanto valori osservati empiricamente (quindi privi di varianza), vengono invece anch'essi *simulati* secondo la varianza ottenuta nella stima di primo grado. Rendendo variabile attraverso le simulazioni un valore che precedentemente era fisso (in quanto osservato) ha come conseguenza proprio il risultato desiderato: la stima viene effettuata comunque, ma risulta affetta da un grado maggiore di incertezza dovuto al fatto che una delle costanti originarie è in realtà variabile. Il risultato è che gli stessi indici diagnostici (intervalli di confidenza, ecc.) riflettono in maniera chiara la minore accuratezza delle stime con il progredire degli stadi, dando un quadro chiaro del percorso di analisi che si sta compiendo. A questo proposito, si comprende la raccomandazione dell'autore di far corrispondere all'ordine delle dicotomizzazioni un significato sostantivo.

Il limite principale di questa proposta è che, come visto poco sopra, è possibile garantire la correttezza del risultato all'interno di una riga, ma non all'interno di tutta la colonna. È questo il motivo per cui, non a caso, King circoscrive l'ambito di applicabilità della sua proposta a tabelle  $2 \times n$  e non  $m \times n$ .

Tabella 5.3 – Coefficienti di flusso in una matrice  $3 \times 2$  in un'unità ecologica (valori assoluti)

	$t_2$		
$t_1$	Sinistra	Destra	Totale
sinistra	$b_i^{sS} N_i^s$	$(1 - b_i^{sS}) N_i^s$	$N_i^s$
centro	$b_i^{cS} N_i^c$	$(1 - b_i^{cS}) N_i^c$	$N_i^c$
destra	$b_i^{dS} N_i^d$	$(1 - b_i^{dS}) N_i^d$	$N_i^d$
Totale	$N_i^S$	$N_i^D$	$N_i$

### 5.3. Ripensare EI in più dimensioni

Come abbiamo visto, è senza dubbio possibile usare, senza troppe difficoltà, la tecnica EI per stimare tabelle di qualunque dimensione: il problema è che la maggior parte delle procedure appena esaminate presenta problemi notevoli. Questo perché è l'approccio di fondo a essere stato messo a punto per studiare la relazione tra due dicotomie.

Per superare queste limitazioni occorre quindi porsi l'obiettivo, anzitutto, di una procedura di stima che, a partire dai dati conosciuti della tabella<sup>5</sup> *analizzati nel loro complesso*, fornisca una stima di *tutti* i coefficienti di cella ricercati. È evidente che questa procedura deve essere capace di tener conto *contemporaneamente* di tutte le informazioni disponibili. Ma per fare questo è indispensabile ripensare dalle fondamenta l'approccio EI vero e proprio, e cercare di riformularlo in termini generali.

#### 5.3.1. L'approccio

È possibile generalizzare la tecnica EI per lavorare non più solo con dicotomie? Proviamo anzitutto a fare un esempio semplice, in cui vogliamo lavorare con una tabella  $3 \times 2$ . Potremmo riferirci, ad esempio, a due elezioni caratterizzate da offerta partitica diversa, con un partito di centro presente nella prima elezione ma non

<sup>5</sup> Nel nostro caso, i marginali di riga e di colonna.

nella seconda. Utilizzando la notazione vista in precedenza, possiamo esprimere il problema nei termini della Tabella 5.3.

Ripetendo il procedimento del capitolo 3, passiamo adesso da valori assoluti a proporzioni, definendo  $X_i$  la proporzione di voti a sinistra nell'elezione  $t_1$ ,  $X_2$  la proporzione di voti al centro nell'elezione  $t_1$  e  $Y$  la proporzione di voti a Sinistra nell'elezione  $t_2$ :

$$X_1 = \frac{N_i^s}{N_i}$$

$$X_2 = \frac{N_i^c}{N_i}$$

$$Y = \frac{N_i^s}{N_i}$$

trattandosi di proporzioni (comprese tra 0 e 1), avremo anche:

$$\frac{N_i^d}{N_i} = 1 - (X_1 + X_2)$$

$$\frac{N_i^D}{N_i} = 1 - Y$$

e dunque possiamo riscrivere la Tabella 5.3 in termini di proporzioni e non più di voti assoluti, ottenendo la Tabella 5.4.

Tabella 5.4 – Coefficienti di flusso in una matrice 3x2 in un'unità ecologica (proporzioni)

	$t_1$	$t_2$	Totale
$t_1$			
sinistra	$b_i^{ss} X_{1i}$	$(1 - b_i^{ss}) X_{1i}$	$X_{1i}$
centro	$b_i^{cs} X_{2i}$	$(1 - b_i^{cs}) X_{2i}$	$X_{2i}$
destra	$b_i^{ds} (1 - (X_{1i} + X_{2i}))$	$(1 - b_i^{ds}) (1 - (X_{1i} + X_{2i}))$	$1 - (X_{1i} + X_{2i})$
Totale	$Y_i$	$(1 - Y_i)$	1

Abbiamo proceduto come negli esempi del capitolo 3, con la differenza che stavolta le righe sono 3. Tuttavia le colonne sono sempre 2, e i marginali di colonna sono conosciuti: ciò significa che per determinare tutta la tabella è sufficiente determinare la prima colonna. Ma le relazioni della prima colonna si possono sintetizzare nella seguente identità:

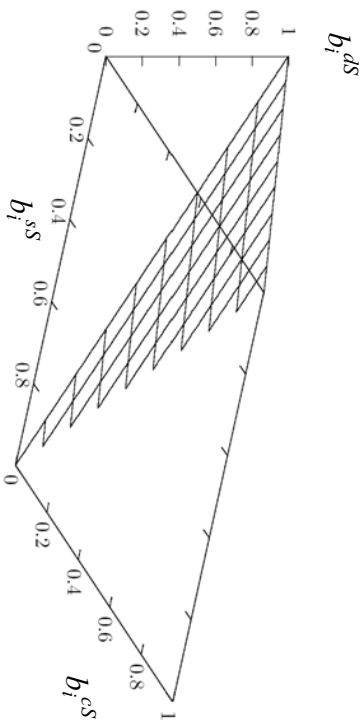
$$Y_i = b_i^{ss} X_{1i} + b_i^{cs} X_{2i} + b_i^{ds} (1 - (X_{1i} + X_{2i})) \quad (5.1)$$

Analogamente all'esempio del capitolo precedente, la 5.1 esprime il *vincolo* a cui devono sottostare i tre coefficienti  $b_i^{ss}$ ,  $b_i^{cs}$  e  $b_i^{ds}$  per essere compatibili con i marginali di riga e colonna, che sono osservati empiricamente.

Nell'esempio del capitolo precedente avevamo tracciato un piano cartesiano in cui sull'asse delle ascisse avevamo i possibili valori di  $b_i^{ss}$  e sulle ordinate quelli di  $b_i^{ds}$ . In quel caso il vincolo corrispondeva a una retta, che in un caso di esempio, avevamo disegnato nella **Errore. L'origine riferimento non è stata trovata.**

In questo caso i coefficienti da stimare sono tre: di conseguenza non potremo più accontentarci di un piano, ma dovremo muoverci in uno spazio a tre dimensioni. Sui tre assi cartesiani avremo i valori rispettivi dei tre coefficienti da stimare, e il vincolo espresso nella (5.1) corrisponderà a un piano, visibile nella Figura 5.1.

Figura 5.1 – Vincolo relativo ai valori dei coefficienti  $b_i^{s^s}$ ,  $b_i^{s^c}$  e  $b_i^{s^s}$  in un'unità ecologica



Analogamente all'esempio in due dimensioni, la rappresentazione grafica del vincolo ha un significato preciso. Tenuto conto che ogni punto dello spazio a tre dimensioni rappresenta, con le sue coordinate, una combinazione di tre coefficienti di cella (gli assi cartesiani sono infatti contrassegnati con i tre diversi coefficienti), il piano della Figura 5.1 restringe l'insieme dei punti accettabili in base ai marginali di riga e colonna osservati. Le possibili terne di coefficienti devono infatti rispettare la relazione (5.1); dal punto di vista grafico, ciò corrisponde al fatto che debbano appartenere al piano rappresentato nella Figura 5.1, che chiameremo, per analogia con il capitolo precedente, *piano di vincolo*<sup>6</sup>.

Abbiamo dunque tracciato il piano di vincolo per una unità ecologica. Ripetendo lo stesso procedimento per le  $n$  sezioni elettorali considerate, otterremo  $n$  piani di vincolo (uno per ogni sezione

<sup>6</sup> Di fatto ogni punto rappresenta, dal punto di vista metodologico, un diverso modello di flusso elettorale, visto che i coefficienti rappresentati dalle sue coordinate individuano completamente i flussi da un partito all'altro. Il significato del fatto che un punto appartenga o meno al piano di vincolo è precisamente che i punti del piano rappresentano modelli compatibili con i dati osservati (i marginali di riga e colonna, ovvero il materiale empirico a disposizione), mentre quelli esterni al piano corrispondono a modelli incompatibili con i dati.

elettorale) che, tracciati nel medesimo diagramma, costituiranno, in analogia con l'esempio in due dimensioni, la *tomografia* relativa al collegio elettorale che stiamo esaminando.

L'unica differenza è che si tratterà di una tomografia tridimensionale, in cui non tracciamo *rette* ma *piani di vincolo*, e che purtroppo non è visualizzabile in modo comprensibile. Ciò che è importante è che il procedimento è esattamente identico a quello dell'approccio EI in due dimensioni (ovvero per due coefficienti).

Ma il punto più interessante è che il procedimento qui esposto può essere senza grossi problemi esteso e generalizzato a  $m$  dimensioni (necessità di stimare  $m$  coefficienti). Nel caso di una tabella  $m \times 2$ , invece del caso  $3 \times 2$  appena visto, saremo infatti *commune* in grado di esprimere il vincolo che lega i coefficienti con un'espressione analoga alla (5.1):

$$Y_i = b_i^{1U} X_{1i} + b_i^{2U} X_{2i} + \dots + b_i^{(m-1)U} X_{(m-1)i} + b_i^{mU} (1 - (X_{1i} + X_{2i} + \dots + X_{(m-1)i})) \quad (5.2)$$

Esisteranno ovviamente  $n$  vincoli come quello espresso dalla 5.2, ciascuno relativo a una unità ecologica, così come il vincolo relativo all'aggregato nel suo complesso. Dal punto di vista geometrico, ciascuno di questi vincoli corrisponde a un *iperpiano di vincolo*<sup>7</sup> in un spazio a  $m$  dimensioni. L'unica difficoltà è che, a maggior ragione, non è possibile disegnare una tomografia comprensibile, in quanto si tratterebbe di rappresentare (su carta o su uno

<sup>7</sup> L'iperpiano è infatti il luogo geometrico dei punti che soddisfano una relazione lineare della forma (5.2). La retta è semplicemente un caso particolare di iperpiano in uno spazio a due dimensioni, così come il piano è un caso particolare di iperpiano in uno spazio a tre dimensioni.

schermo, quindi su una superficie a 2 dimensioni)  $n$  iperpiani in uno spazio  $m$ -dimensionale<sup>8</sup>.

### 5.3.2. La stima: minimizzazione della distanza

Tornando al nostro caso di esempio tridimensionale, il problema che si pone a questo punto è quello delle tecniche di stima. King stesso immagina che si possa lavorare con distribuzioni normali  $n$ -variate troncate, ma sconsiglia quest'approccio per motivi di onerosità computazionale. Tuttavia possiamo chiederci se non sia possibile adattare qualcuna delle tecniche di stima alternative viste nel capitolo precedente.

La tecnica che appare più adatta è, a questo proposito, la minimizzazione delle distanze quadrate proposta da Grofman e Merrill [2002; 2004], già vista in precedenza. Come anticipato, si tratta, rispetto alla tecnica proposta da King, di una tecnica meno raffinata dal punto di vista statistico, ma decisamente più semplice dal punto di vista dell'algoritmo impiegato.

Ricapitoliamo brevemente questa soluzione, nella sua versione originaria in due dimensioni: una volta tracciata la tomografia, ovvero l'insieme di segmenti di retta di vincolo relativi ciascuno a una sezione elettorale, si prende in esame la retta di vincolo aggregata, ovvero quella tracciata utilizzando come parametri non i marginali di riga e colonna di una singola unità ecologica, ma quelli aggregati dell'intero aggregato.

La retta di vincolo aggregata rappresenta l'insieme dei possibili modelli candidati a soluzione (ogni punto rappresenta, con le sue coordinate, una coppia di coefficienti di flusso, dunque un modello).

A questo punto, per ogni punto di questa retta si sommano le distanze (al quadrato) da ciascuna delle rette di vincolo che costi-

---

<sup>8</sup> Il procedimento appena visto per l'estensione diretta a  $m$  dimensioni della tecnica EI è applicabile a casi di tabelle  $m \times 2$ . Per un'estensione al caso generico  $m \times n$ , vedi oltre.

tuiscono la tomografia, il punto ottimale è quello che ha la minore somma di distanze complessiva.

Il procedimento può essere adattato senza modifiche all'esempio con tre coefficienti (ovvero in tre dimensioni). Una volta tracciati tutti i piani di vincolo relativi alle  $n$  unità ecologiche, potremo senza problemi tracciare anche il *piano di vincolo aggregato*.

I punti appartenenti a questo piano saranno i candidati a soluzione: calcolando per ciascuno di questi punti la distanza dal piano di vincolo di ogni unità ecologica, otterremo la somma di distanze per quel punto. Il punto con la somma di distanze minima è la nostra stima: le sue tre coordinate (siamo in uno spazio tridimensionale) costituiscono, tra i modelli compatibili con i dati osservati, il modello più verosimile.

### Soluzione per via analitica: vantaggi e limitazioni

Il procedimento così presentato può apparire laborioso: in realtà gli autori lo risolvono per via analitica, giungendo a una formula generale che fornisce *direttamente* la combinazione di coefficienti che minimizza la somma di distanze<sup>9</sup>.

La procedura è quindi di applicazione semplicissima, ed è elegante in termini matematici. Tuttavia presenta alcune limitazioni, soprattutto per quanto riguarda l'estensione a dimensioni superio-

---

<sup>9</sup> Il procedimento analitico utilizzato da Grofman e Merrill non è particolarmente complesso: Dato un punto  $P$  di ascissa  $b_1$  e ordinata  $b_2$  (i due coefficienti) appartenente alla retta di vincolo aggregata, e una retta di vincolo  $R_i$  di un'unità ecologica data, è possibile individuare la retta  $S_i$  perpendicolare a  $R_i$  che passa per il punto  $P$ . Individuata  $S_i$ , è possibile calcolare il punto  $Q_i$  in cui essa interseca  $R_i$ : il segmento  $PQ_i$  è appunto la distanza tra il punto  $P$  e la singola retta di vincolo  $R_i$ , ed è esprimibile come funzione delle coordinate di  $P$  e dei parametri della retta  $R_i$ . Di conseguenza è possibile esprimere la somma delle distanze da tutte le rette di vincolo con un'unica funzione  $f$  che è la somma delle singole funzioni di distanza così ottenute. Per individuare il punto in cui tale funzione  $f$  è minima, è sufficiente imporre la condizione che tutte le sue derivate parziali siano uguali a 0 per ottenere la soluzione. Applicando il teorema di Lagrange, Grofman e Merrill impongono inoltre che la soluzione si trovi all'interno della retta di vincolo aggregato.

ri. Da un lato, infatti, occorrerebbe risolvere un problema analitico diverso a seconda delle dimensioni della tabella da stimare; in secondo luogo (e si tratta del problema più grave), non è certo che esista necessariamente una soluzione analitica, già a partire da tabelle  $3 \times 2$ .

### Soluzione per via numerica

Tuttavia il problema può essere affrontato con un approccio numerico. Si tratta in effetti, di eseguire effettivamente tutti i passi della procedura delineate precedentemente. Vediamo quest'alternativa nel semplice caso bidimensionale.

Il problema è il seguente: tra i punti appartenenti alla retta di vincolo aggregata (i possibili modelli), individuare quello che presenta la somma di distanze minore da tutte le rette di vincolo delle singole unità ecologiche.

Iniziamo con l'individuare i punti che costituiscono la retta di vincolo aggregata  $\mathcal{A}$ , in maniera da poterli esaminare singolarmente. Il metodo è quello di prendere in esame tutti i punti del piano cartesiano compresi nell'intervallo  $[0;1] \times [0;1]$ . Per ciascuno di essi, si controlla che le sue coordinate verificano l'equazione della retta di vincolo aggregata  $\mathcal{A}$ . Se è così, abbiamo trovato un punto della nostra retta di vincolo aggregata. Al termine della nostra ricerca avremo un insieme  $a$  costituito da tutti e soli i punti appartenenti alla retta  $\mathcal{A}$  compresi nell'intervallo  $[0;1] \times [0;1]$ .

A questo punto, occorre prendere in esame tutti i punti appartenenti all'insieme  $a$  e calcolare, per ciascuno di essi, la somma delle distanze da ciascuna retta di vincolo. Per ogni punto  $P$  appartenente ad  $a$ , possiamo calcolare facilmente la distanza dalla retta di vincolo  $r_i$  relativa a una determinata unità ecologica, ricorrendo a una formula generale<sup>10</sup>. Ripetendo il procedimento per tutte le  $m$  rette di vincolo delle unità ecologiche, potremo calcolare la somma delle distanze dal punto  $P$  a ciascuna delle rette  $r$ .

<sup>10</sup> Lo stesso vale in spazi a più di due dimensioni, ovvero nel caso del piano o dell'iperpiano.

Il procedimento visto per un punto  $P$  appartenente alla retta di vincolo aggregata  $\mathcal{A}$  va poi ripetuto per tutti i punti dell'insieme  $a$ : il punto con la minor somma di distanze sarà la nostra miglior stima, secondo il criterio della distanza minima.

La procedura fin qui esposta presuppone però che noi siamo in grado di costruire degli insiemi che contengono tutti i punti appartenenti a una certa retta<sup>11</sup> compresi nell'intervallo  $[0;1] \times [0;1]$ . Siccome anche un intervallo di dimensioni finite contiene infiniti punti, dovremo necessariamente esplorare l'intervallo utilizzando una *griglia*, caratterizzata da un certo livello di approssimazione (precisione)<sup>12</sup>.

Per fare un esempio, se noi vogliamo costruire l'insieme  $a$ <sup>13</sup> con una precisione di 0,001, dovremo esaminare, per un esempio bidimensionale, circa un milione di punti.

0,001 è infatti un millesimo: per esaminare tutti i punti del quadrato  $[0;1] \times [0;1]$  dobbiamo dunque esaminare  $1001 \times 1001$  combinazioni. I calcoli aumentano se abbiamo una tabella  $3 \times 2$ , dunque siamo in tre dimensioni: i punti da esaminare diventano  $1001 \times 1001 \times 1001$ , ovvero circa un miliardo.

In generale, la lunghezza dell'elaborazione cresce esponenzialmente con il numero di dimensioni considerate (ovvero il numero di righe  $m$  di una tabella  $m \times 2$ ), e linearmente con il grado di precisione richiesto. Tuttavia, per contenere la mole dei calcoli, si può ipotizzare di usare approssimazioni anche grossolane. Oppure si

<sup>11</sup> Il vincolo è infatti costituito da una retta del piano cartesiano nel caso di due coefficienti. Nel caso di tre, corrisponde a un piano in uno spazio tridimensionale. Infine, per tabelle con più di tre righe, a quello che è genericamente conosciuto come iperpiano in uno spazio  $m$ -dimensionale, dove  $m$  è il numero di righe della tabella di contingenza.

<sup>12</sup> La scelta di un livello di precisione ci permette di passare da un numero infinito a un numero finito di punti. È infatti evidente che nell'intervallo tra 0 e 1 ci sono infiniti punti; tuttavia se decidiamo di esaminare la retta con una precisione di 0,1 i punti da esaminare divergono soltanto undici: 0, 0,1, 0,2, ... 0,9, 1.

<sup>13</sup> Ovvero l'insieme dei punti appartenenti alla retta di vincolo aggregato  $\mathcal{A}$  nell'intervallo  $[0;1] \times [0;1]$ .



possono combinare varie fasi di analisi: una prima fase rapida, e seguita con una approssimazione grossolana, con finalità esplorativa alla ricerca dei valori approssimativi della soluzione; una seconda fase con precisione maggiore, ottenuta esplorando intervalli più ristretti per ciascun coefficiente, scelti in base ai risultati dell'elaborazione precedente.

### 5.3.3. Da matrici $m \times 2$ a matrici $m \times n$

Fin qui abbiamo in realtà esaminato il caso  $m \times 2$ : tuttavia è possibile generalizzare la procedura appena esposta anche al caso generale di una tabella di contingenza di  $m$  righe per  $n$  colonne.

La differenza con il caso  $m \times 2$  è abbastanza netta. In quel caso infatti era sufficiente determinare una sola colonna per determinare automaticamente tutti i coefficienti di cella (visto che i marginali di riga sono conosciuti). Di conseguenza i vincoli che legavano tra loro i coefficienti di cella si riducevano a quelli esistenti all'interno della prima colonna, che vengono appunto sintetizzati nella relazione (5.1). Per delineare, tra tutti i punti possibili, l'insieme più ristretto dei possibili candidati a soluzione (ovvero l'insieme  $d$ ) sapevamo che potevano entrare a far parte di questo insieme soltanto i punti che soddisfacevano la relazione.

Nel caso a più colonne facciamo un ragionamento analogo, ma abbiamo a che fare con una situazione più complessa.

I coefficienti di cella non sono più, infatti, legati da *una sola relazione*, ma da *moltiplici relazioni* che, affinché un punto appartenga a un vincolo, devono essere *tutte* soddisfatte. La situazione appare chiara nella Tabella 5.5, dove — per semplicità — i coefficienti non sono più indicati con le iniziali dei partiti, ma con semplici numeri.

Esaminando contemporaneamente tutti e 16 i coefficienti, non facciamo altro che esaminare modelli caratterizzati da 16 coefficienti diversi, dunque, nella nostra metafora geometrica, punti in uno spazio a 16 dimensioni. Ciascun punto è definito dunque da 16 coordinate diverse, espresse dai coefficienti  $b_1 \dots b_{16}$ .

Tabella 5.5 — Coefficienti di flusso in una matrice  $m \times n$  in un'unità ecologica (proporzioni)

$i_1$	$i_2$				Totale
	Sinistra	Centro	Destra	Altri	
sinistra	$b_1^1 X_{11}$	$b_1^5 X_{11}$	$b_1^9 X_{11}$	$b_1^{13} X_{11}$	$X_{11}$
centro	$b_1^2 X_{21}$	$b_1^6 X_{21}$	$b_1^{10} X_{21}$	$b_1^{14} X_{21}$	$X_{21}$
destra	$b_1^3 X_{31}$	$b_1^7 X_{31}$	$b_1^{11} X_{31}$	$b_1^{15} X_{31}$	$X_{31}$
altri	$b_1^4 X_{41}$	$b_1^8 X_{41}$	$b_1^{12} X_{41}$	$b_1^{16} X_{41}$	$X_{41}$
Totale	$Y_{11}$	$Y_{21}$	$Y_{31}$	$Y_{41}$	1

Ogni punto, perché rispetti il vincolo relativo all'unità ecologica  $i^1$ , dovrà soddisfare contemporaneamente *tutti* i vincoli determinati dalle righe e colonne della Tabella 5.5.

Per quanto riguarda i vincoli di colonna, questi sono tutti della forma:

$$Y_{1i} = b_1^1 X_{1i} + b_1^2 X_{2i} + b_1^3 X_{3i} + b_1^4 X_{4i}$$

I vincoli di riga sono invece della forma:

$$X_{1i} = b_i^1 X_{1i} + b_i^5 X_{1i} + b_i^9 X_{1i} + b_i^{13} X_{1i}$$

dove, dividendo per il totale di colonna ( $X_{1i}$  in questo caso), si ottiene la forma semplificata:

$$1 = b_i^1 + b_i^5 + b_i^9 + b_i^{13}$$

Possiamo dunque elencare tutti i vincoli a cui deve sottostare una combinazione di coefficienti  $b_i^1 \dots b_i^{12}$  (ovvero un modello) per appartenere al vincolo relativo all'unità ecologica  $i$  (ovvero per

<sup>14</sup> E dunque appartenere ai modelli compatibili con i marginali di riga e colonna.

essere compatibile con i marginali di riga e colonna). Ricordando che  $Y_{1i} \dots Y_{ni}$  e  $X_{1i} \dots X_{mi}$  sono noti in quanto osservati empiricamente, avremo:

$$\left. \begin{aligned} Y_{1i} &= b_1^1 X_{1i} + b_1^2 X_{2i} + b_1^3 X_{3i} + b_1^4 X_{4i} \\ Y_{2i} &= b_2^5 X_{1i} + b_2^6 X_{2i} + b_2^7 X_{3i} + b_2^8 X_{4i} \\ Y_{3i} &= b_3^9 X_{1i} + b_3^{10} X_{2i} + b_3^{11} X_{3i} + b_3^{12} X_{4i} \\ Y_{4i} &= b_4^{13} X_{1i} + b_4^{14} X_{2i} + b_4^{15} X_{3i} + b_4^{16} X_{4i} \\ 1 &= b_i^1 + b_i^5 + b_i^9 + b_i^{13} \\ 1 &= b_i^2 + b_i^6 + b_i^{10} + b_i^{14} \\ 1 &= b_i^3 + b_i^7 + b_i^{11} + b_i^{15} \\ 1 &= b_i^4 + b_i^8 + b_i^{12} + b_i^{16} \end{aligned} \right\}$$

La procedura vista precedentemente non viene in nessun modo modificata o invalidata dal fatto che adesso il vincolo da rispettare non è più costituito da una sola relazione, ma da un insieme di più relazioni. Semplicemente, la costruzione dell'insieme  $a$  richiederà una mole di calcoli decisamente maggiore rispetto al caso precedente.

Se prima, infatti, si trattava di esaminare i punti di un intervallo in uno spazio a *tre* dimensioni, verificando che le coordinate del punto soddisfacessero *una* relazione, in questo caso si tratta di esaminare i punti di uno spazio a *sebbi* dimensioni, controllando che le coordinate del punto verifichino contemporaneamente *otto* relazioni. Tuttavia ciò pone esclusivamente problemi di onerosità

del calcolo numerico e non introduce nuove difficoltà concettuali<sup>15</sup>.

Il risultato di una simile procedura, a qualunque complessità dimensionale venga applicata, è dunque quello che ci eravamo proposti all'inizio: fornire una stima dell'intera matrice di flusso in un solo stadio di elaborazione. I vincoli presenti devono essere infatti rispettati *tutti* contemporaneamente: è impossibile per definizione ottenere coefficienti che contraddicano i marginali conosciuti, in quanto vengono ammesse come candidate a soluzione soltanto le combinazioni di coefficienti (ovvero i modelli) compatibili con tutti i vincoli definiti dai marginali di riga e colonna osservati empiricamente. Tra tutte queste combinazioni viene infine scelta quella che presenta la minima distanza dagli insiemi di modelli possibili in ciascuna unità ecologica.

Un simile risultato purtroppo presenta costi non indifferenti; in particolare, l'onerosità dei calcoli è notevole. Tuttavia l'elaborazione non è impossibile. La procedura descritta in questo capitolo infatti è stata implementata come procedura informatica, con il nome di EHDIS<sup>16</sup>.

Il capitolo che segue presenta appunto un esempio di applicazione di questa procedura a un caso empirico: la stima dei comportamenti di "voto differenziato" tra arena maggioritaria e proporzionale alla Camera nelle elezioni politiche del 2001.

<sup>15</sup> La complessità di elaborazione diventa tuttavia molto maggiore per calcolare la distanza da ciascun punto del vincolo aggregato a ciascun vincolo ecologico. I vincoli in questo spazio non sono più degli iperpiani: di conseguenza non può essere utilizzata la formula analitica per la distanza di un punto da un iperpiano. Occorre quindi costruire degli insiemi di punti, uno per ciascun vincolo ecologico, analoghi all'insieme  $\alpha$  per il vincolo aggregato, e poi – confrontando le distanze tra questi insiemi di punti – ottenere la distanza di un punto da ciascun vincolo ecologico.

<sup>16</sup> Ecological Inference / Direct Search, a sottolineare che la procedura opera nell'ambito dell'approccio EI, ma ricercando la soluzione tramite ricerca diretta, e dunque non per via analitica ma per via numerica.

## **6. Un esempio di applicazione empirica: il voto differenziato nelle elezioni politiche del 2001**

Nel capitolo precedente sono stati presentati alcuni possibili approcci per estendere l'uso della tecnica EI a tabelle di dimensioni maggiori di  $2 \times 2$ . Infine è stata presentata una proposta originale: l'estensione dell'approccio delle minime distanze [Grofman e Merrill 2002; 2004] a tabelle  $m \times 2$  e  $m \times n$  mediante individuazione della soluzione non più per via analitica ma per via numerica. Non resta quindi che proporre un primo esempio di applicazione della procedura a dati empirici.

Il caso che ho scelto di analizzare è quello del “voto differenziato” (tra maggioritario e proporzionale alla Camera) nelle elezioni politiche del 2001. I motivi di questa scelta sono essenzialmente due. Il primo è relativo alla preferenza per un'analisi sincronica piuttosto che diacronica; questo perché in questo modo si eliminano le distorsioni dovute al ricambio della popolazione elettorale tra due elezioni [Anastasi et al. 1989; vedi anche il capitolo 2]. Accettata la scelta sincronica, le possibilità per un'analisi su larga scala (ovvero su tutto il territorio nazionale suddiviso in sotto-analisi locali, invece che su un campione di sezioni) si restringono essenzialmente a due: il confronto Camera-Senato 2006 oppure quello proporzionale-maggioritario alla Camera nel 2001. La scelta è caduta su quest'ultimo caso essenzialmente perché le analisi esistenti mettono in evidenza una presenza più consistente di comporta-

menti differenziati<sup>1</sup>. Veniamo quindi anzitutto a inquadrare in generale il fenomeno del voto differenziato nelle elezioni svoltesi con la legge Mattarella.

## 6.1. Voto differenziato e differenziale di coalizione

Come è noto, dalla riforma del 1993 a quella del 2005 [D'Alimonte 2007] in Italia è stato in vigore un sistema elettorale misto maggioritario-proporzionale. Per quanto riguarda la Camera dei Deputati, essa veniva eletta per il 75% dei membri in collegi uninominali con sistema *plurality*, per il restante 25% con un sistema proporzionale a liste bloccate, basato su circoscrizioni regionali o sub-regionali. È importante sottolineare che il voto per queste due arene competitive veniva espresso in due schede separate, prive di qualunque legame tra loro. Di conseguenza era possibile senza problemi che gli elettori votassero in modo differenziato tra le due arene [D'Alimonte e Chiaromonte 1995].

La riforma del 1993 non era stata ispirata e disegnata dai partiti dell'epoca. Al contrario, fu essenzialmente subita da questi ultimi, in quanto resa praticamente obbligata dall'esito del referendum elettorale del 1993 [D'Alimonte e Bartolini 1995a]. È così che, immediatamente dopo l'approvazione della riforma, i partiti italiani intraprendono a tappe forzate quello che possiamo inquadrare ex post come un processo di *coordinamento strategico*<sup>2</sup> [Cox 1997]. L'esame dei vincoli e delle opportunità offerte dal nuovo sistema elettorale porta rapidamente i soggetti politici a ridefinire radical-

mente le proprie strategie competitive in vista delle elezioni del 1994, dando rapidamente vita a vere e proprie coalizioni politico-elettorali [Di Virgilio 1995].

La struttura delle coalizioni del 1994 è tuttavia ancora acerba e risente di un insufficiente livello di coordinamento strategico sia da parte dei partiti che degli elettori. Ad esempio, errori di valutazione del proprio bacino elettorale portano il Patto per l'Italia alla pericolosa scelta di correre autonomamente, ottenendo appena 4 collegi uninominali con circa il 16% di voti nazionali; per un altro verso, l'ipotesi che il "voto strategico" sia ancora poco diffuso tra gli elettori suggerisce a Forza Italia di coalizzarsi separatamente con la Lega Nord e con Alleanza Nazionale in diverse aree del paese<sup>3</sup>.

Tuttavia il processo di apprendimento si può dire che procederà abbastanza speditamente nelle elezioni successive. Da un lato con la creazione dell'Ulivo e con il grande successo della Lega nel 1996 (che corre da sola sapendo che il suo radicamento territoriale è un punto di forza nel nuovo sistema elettorale); dall'altro con gli scarsi consensi ottenuti dai "terzi poli" nel 2001, a testimonianza di un apprendimento anche da parte degli elettori, che non danno credito a operazioni che ritengono potenzialmente fallimentari come il "Patto per l'Italia" del 1994. Il risultato è un esito, sia pure con molti distinguo, di sostanziale bipolarizzazione del sistema partitico [Bartolini, Chiaromonte, e D'Alimonte 2002].

Se nelle sue grandi linee la ristrutturazione del sistema partitico italiano tra 1994 e 2001 sembra essere caratterizzata, nelle strategie dei partiti e nei comportamenti degli elettori, da un livello crescente di apprendimento delle regole elettorali e dunque di successo nel coordinamento strategico, tuttavia non si può negare l'esistenza di fenomeni meno chiari. Uno di questi è quello che ri-

---

<sup>1</sup> Riguardo al voto differenziato maggioritario-proporzionale nel 2001 vedi più oltre; per quanto riguarda il voto differenziato Camera-Senato nel 2006 alcune analisi (sia su dati aggregati che di sondaggio) ne mostrano una rilevanza tutto sommato contenuta [De Sio 2006c; Scappini e Tuorto 2006; De Sio 2007c].

<sup>2</sup> Il concetto di *coordinamento strategico* rappresenta di fatto un'estensione anche alle élites partitiche di quegli *effetti psicologici del sistema elettorale* che Duverger [1950] ipotizzava per gli elettori.

<sup>3</sup> Con il termine "voto strategico" si intende un comportamento di voto in cui l'elettore rinuncia alla scelta di voto preferita (voto *sincero*) se questa ha scarse probabilità di successo, operando viceversa per una scelta subottimale ma con maggiori probabilità di successo [Farquharson 1969; Felsenthal 1990; D'Alimonte e Chiaromonte 1995; Cox 1997].

tengo opportuno chiamare, in modo neutro, “differenziale di coalizione”, ovvero la sistematica presenza di una differenza tra i voti raccolti da una coalizione nell’arena maggioritaria e quelli raccolti dalla somma dei suoi partiti nell’arena proporzionale.

Nella prima elezione, quella del 1994 che aveva visto tre coalizioni, la coalizione di sinistra (i Progressisti) registra un differenziale negativo, mentre quella di centrodestra (nella sua duplice incarnazione di Polo della Libertà al Nord e Polo del Buon governo al Sud) presenta un differenziale positivo, ottenendo più voti nel maggioritario che nel proporzionale.

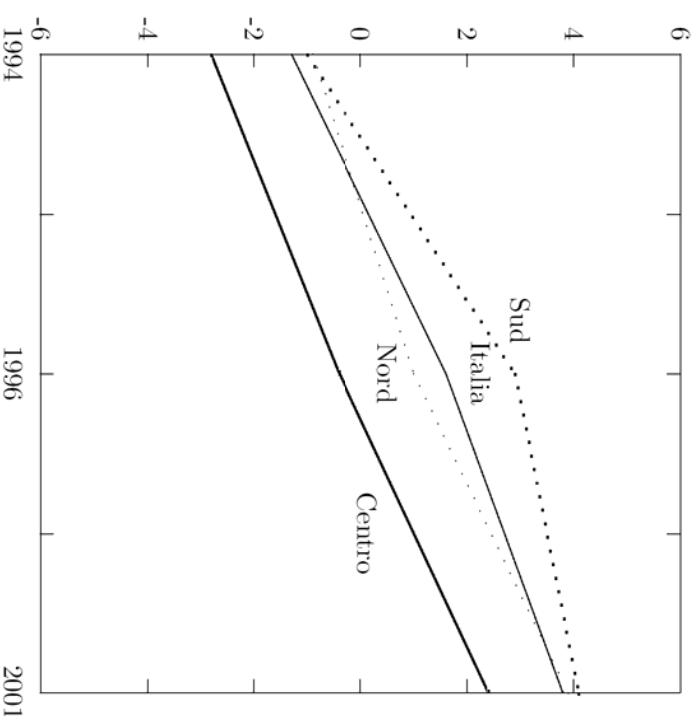
Nelle elezioni del 1996 e in quelle del 2001, più nettamente bipolari<sup>4</sup>, la situazione invece si rovescia: è l’Ulivo a presentare un differenziale positivo (a livello nazionale di +1,6 punti nel 1996 e +3,8 nel 2001), mentre la Cdl presenta un differenziale negativo (-3,7 punti nel 1996 e -3,6 nel 2001). In termini numerici non si tratta di un fenomeno di poco conto<sup>5</sup>. L’andamento del differenziale per le due coalizioni è visibile nella Figura 6.1 e nella Figura 6.2.

Da dove viene questa differenza? Le spiegazioni che sono state proposte sono essenzialmente classificabili in due categorie, a seconda che prendano come riferimento principale il voto proporzionale o quello maggioritario.

<sup>4</sup> Ricordando tuttavia che nel 1996 la Lega corse da sola sia nel maggioritario che nel proporzionale, ottenendo circa il 10% di voti validi in entrambe le arene, e che nel 2001 le terze forze di rilievo nazionale (Di Pietro, Bonino, Democrazia Europea e Fiamma Tricolore) ottennero ancora circa il 9%, contro lo 0,5% di tutte le terze forze nel 2006.

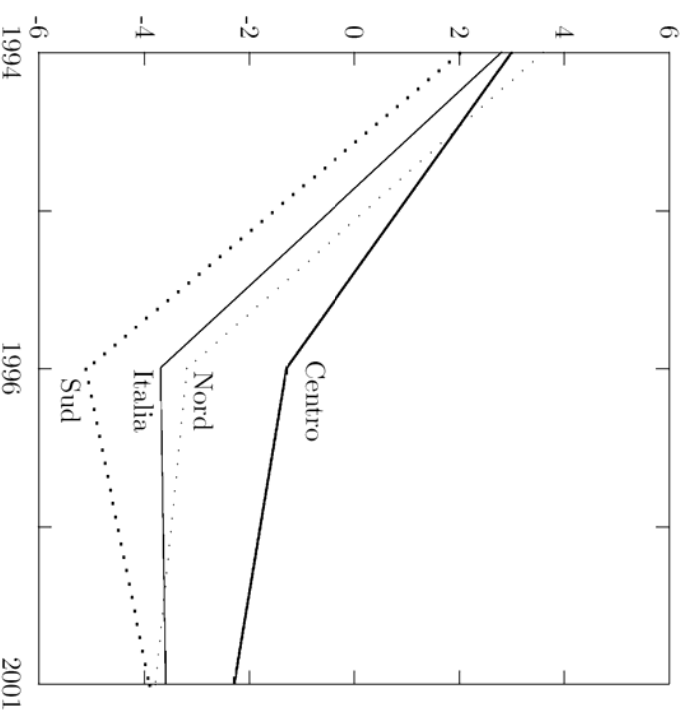
<sup>5</sup> Questa differenza non è necessariamente sorprendente per un sistema elettorale misto: nel panorama internazionale sono frequenti i casi di differenze anche molto maggiori [Chiaromonte 2005]. Tuttavia si tratta in generale di casi in cui il livello di strutturazione del sistema partitico è molto basso, e forte la presenza di candidature locali. Nel caso italiano siamo invece in presenza di una sostanziale nazionalizzazione della competizione, con scarsa o nulla rilevanza di candidati “independenti”.

Figura 6.1 – Differenziale di coalizione del centrosinistra 1994-2001



Alla prima categoria appartiene l’ipotesi del “rendimento coalizionale” [D’Alimonte e Bartolini 1997], proposta per la prima volta relativamente alle elezioni del 1996. All’origine di questa ipotesi c’è un modello esplicativo che combina due approcci teorici distinti. Da un lato l’analisi storica dello sviluppo del sistema partitico della prima Repubblica, costruito intorno a partiti ben strutturati e dall’elettorato fortemente ideologizzato e compartimentato su fratture di matrice storica; dall’altro il rifinimento ai modelli di scelta razionale. I due approcci convergerebbero nell’attribuire al voto *proporzionale* il ruolo di preferenza “vera” e “sincera” dell’elettore [D’Alimonte e Bartolini 1997, 243; Bartolini e D’Alimonte 2002a, 235].

Figura 6.2 – Differenziale di coalizione del centrodestra 1994-2001



Per il primo punto di vista, ciò si giustificerebbe per il condizionamento storico relativo a un rapporto di tipo ideologico con la politica, che dunque determinerebbe un attaccamento al partito, dotato di un profilo ideologico e identitario decisamente più netto rispetto a quello evanescente della coalizione. Per il secondo, il ventaglio più ampio e articolato di scelte offerto dal voto proporzionale permetterebbe una rivelazione più fedele delle preferenze dell'elettore. Il differenziale di coalizione si spiega quindi come "rendimento coalizionale": una sorta di "bonus" che permette a una coalizione di andare oltre il reale confine della proprio bacino elettorale "sincero", che va misurato comunque nella competizione proporzionale.

Riguardo alle ipotesi esplicative, gli autori propongono – per spiegare *in negativo* il rendimento negativo della CdL nel 1996 – i tre possibili effetti *candidato*, *coalizione* o *leadership*. Il deficit di voti maggioritari rispetto alla somma dei partiti nel proporzionale sarebbe dovuto nel primo caso al rifiuto di un candidato di collegio non gradito; nel secondo, al rifiuto delle scelte coalizionali del partito preferito, che però viene votato nel proporzionale; il terzo, infine, al non gradimento del leader nazionale della coalizione a cui appartiene il proprio partito. Il *benchmark* del bacino elettorale di riferimento delle coalizioni sarebbe tuttavia sempre il voto proporzionale. In direzione in parte simile vanno anche i contributi di Pappalardo [1996; 2000] che ipotizza un effetto negativo dei candidati dei Progressisti nel 1994, e viceversa un vero e proprio effetto *leadership*, stavolta di segno positivo, per spiegare il successo dell'Ulivo nel 1996.

Questa posizione viene capovolta nell'approccio che vede invece il primato del voto maggioritario. Lo scenario proposto suggerisce l'importanza del vero e proprio crollo del sistema partitico nei primi anni '90, che porta alla nascita di nuovi partiti o a radicali cambiamenti nei partiti "sopravvissuti", in un clima oltretutto di forte diffidenza nei confronti dell'istituzione-partito. Sarebbe anche questo clima, nonché l'emersione di una dinamica bipolare competitiva, ad alimentare una concezione del rapporto con la politica meno ideologizzato, più orientato ad alcuni valori di fondo piuttosto che a una complessa e strutturata ideologia, e dunque imperniato essenzialmente sulle meno impegnative identità di coalizione piuttosto che su quelle di partito [Natale 2002].

A questo si deve aggiungere che già dal 1994 il sistema dei media fornisce una rappresentazione della competizione che è essenzialmente bipolare ("o di qua o di là"), in questo sottolineando il cambiamento di strategia elettorale che vede le coalizioni, durante la campagna, enfatizzare molto di più l'identità di coalizione ("noi contro di loro") che le molteplici identità di partito. Né è staccata da questa interpretazione la meccanica vera e propria del sistema elettorale: le elezioni nel "Mattarellum" vengono decise nell'arena

maggioritaria, che assegna, in base al voto alla coalizione, il 75% dei seggi.

Infine, in termini di processo decisionale astratto, è possibile rovesciare il ragionamento visto precedentemente, sostenendo viceversa che un'offerta politica più semplice (tendenzialmente con due sole opzioni) produce un processo di scelta più coerente e meno soggetto ad errori o incertezze, fornendo dunque una rappresentazione paradossalmente più fedele delle "vere" preferenze politiche dell'elettore. Seguendo questo approccio, sarebbe quindi il voto *maggioritario* a rappresentare il punto di riferimento principale per l'analisi: l'espressione più fedele dell'orientamento politico prevalente dell'elettore.

In termini più strettamente contabili, i due filoni propongono spiegazioni in parte divergenti di quali sarebbero i comportamenti individuali di voto che al livello aggregato causerebbero il differenziale di coalizione osservato.

Barolini e D'Alimonte [2002a] spiegano il fenomeno, nel caso del 2001, mediante l'ipotesi del "centro scambiatore?". In sostanza, il nocciolo della spiegazione del divario tra voti maggioritari e proporzionali per le due coalizioni consisterebbe in due flussi di voto da sinistra verso destra, che riguarderebbero bacini di elettori separati. Da un lato avremmo una quota di elettori proporzionali di centrodestra che nel maggioritario avrebbero scelto candidati dei terzi poli; parallelamente, una quota di elettori proporzionali dei terzi poli avrebbe votato nel maggioritario per l'Ulivo.

A suffragio di questa ipotesi vanno i dati aggregati relativi alla correlazione tra numero di candidati presenti nelle competizioni di collegio e differenziale positivo o negativo delle due coalizioni. Il dato centrale è la totale assenza di correlazione per l'Ulivo (a testimonianza di una apparente impermeabilità dell'Ulivo verso i terzi poli), ma soprattutto la forte correlazione col differenziale negativo della Cdl, che perde di più tra maggioritario e proporzionale quanti più candidati indipendenti sono presenti, e non perde praticamente niente quando ha contro il solo candidato dell'Ulivo [Barolini e D'Alimonte 2002a, 237]. Ciò confermerebbe una vul-

nerabilità del centrodestra verso i terzi poli ma non direttamente verso l'Ulivo.

Il filone che invece suggerisce la crescente importanza dell'identità di coalizione sottolinea viceversa, soprattutto per l'Ulivo, la maggiore stabilità del voto maggioritario rispetto a quello proporzionale [Natale 2002; Ricolfi e Testa 2002]. In questo senso, il voto proporzionale a terze forze o addirittura non valido avrebbe un carattere sostanzialmente secondario, o di protesta, rispetto a una stabile identità di coalizione che orienterebbe il voto nel maggioritario. A suffragio di questa ipotesi vengono presentati dati di sondaggio, nonché alcune stime di flussi di voto basate su un campione di sezioni elettorali del 2001.

E' difficile dire quale dei due filoni di ipotesi abbia la migliore capacità esplicativa rispetto al fenomeno del voto differenziato, sia in termini teorici che empirici. Analisi successive di flusso basate su dati aggregati hanno tuttavia evidenziato la caratteristica fondamentale dell'*eterogeneità*: il voto differenziato sembra essere la somma (e in parte anche l'elisione reciproca) di numerosi fenomeni di tipo e segno diverso, dovuti a meccanismi individuali di tipo e segno eterogeneo, che quindi lasciano intendere anche una corrispondente eterogeneità territoriale. In questo faccio riferimento a una precedente analisi [De Sio 2006a] condotta utilizzando il modello di Goodman, e che quindi ci offre, come vedremo in seguito, anche la possibilità di un confronto tra i risultati delle due tecniche di stima.

L'obiettivo dell'analisi è quindi quello di accertare la reale consistenza dei vari comportamenti di voto differenziato, in modo da comprendere l'origine del differenziale tra maggioritario e proporzionale osservato per le due coalizioni.

## 6.2. L'analisi complessiva

Veniamo quindi all'analisi vera e propria. Il primo problema da affrontare è la formulazione del problema in termini di una tabulazione incrociata sufficientemente semplice da poter essere stima-

ta con la procedura Eidi, pesante dal punto di vista computazionale. A questo proposito, ho quindi ritenuto opportuno riprendere la classificazione di Ricolfi e Testa [2002], che suddivide l'elettorato essenzialmente in tre bacini: i due corrispondenti alle coalizioni principali, e un terzo che viene definito "residuale", che comprende i voti alle terze forze, ma anche i voti non validi e l'astensione.

Questa classificazione è valida sia per l'arena maggioritaria che per quella proporzionale; ovviamente nel caso proporzionale non facciamo riferimento alle coalizioni, ma alla somma dei partiti che le compongono.

Ciò premesso, è possibile distinguere tre meccanismi individuali di voto differenziato, che sono all'origine del differenziale di coalizione:

- *confitto*: l'elettore vota nel proporzionale un partito di una coalizione, e nel maggioritario la coalizione avversaria;
- *attrazione*: l'elettore nel proporzionale si astiene o vota un partito "residuale", ma nel maggioritario vota una delle due coalizioni;
- *dispersione*: l'elettore vota nel proporzionale un partito di una delle due coalizioni, ma nel maggioritario si astiene o vota una coalizione "residuale"<sup>6</sup>.

Dunque un primo oggetto di riscontro empirico può essere la consistenza, per ciascuna coalizione, di questi tre effetti. Premesso che l'Ulivo nel 2001 ha un differenziale di coalizione fortemente positivo, esso è dovuto più al *confitto* di alcuni elettori della Cdl (che passano all'Ulivo nel maggioritario) o all'effetto di *attrazione* dell'Ulivo sugli elettori proporzionali dell'area residuale? O forse il differenziale è stato anche attenuato da un effetto di *dispersione* (u-

---

<sup>6</sup> Faccio riferimento alla terminologia che ho utilizzato nell'analisi citata precedentemente [De Sio 2006a]. Ricolfi e Testa [2002] parlano rispettivamente di *trattamento incrociato*, *lealtà di coalizione* e *lealtà di partito*.

lvisiti proporzionali che nel maggioritario si disperdono verso l'area residuale)? Domande analoghe valgono per la Casa delle Libertà (anche se in questo caso occorre spiegare un differenziale negativo). Si tratta di interrogativi posti in maniera simile anche da Bartolini e D'Alimonte [2002b], che tuttavia ricorrono a una classificazione diversa dei "bacini elettorali". In particolare, il voto alle terze forze viene considerato separatamente dall'astensione, dividendo quindi il bacino "residuale" in due bacini separati (scelta che qui non è possibile seguire per le difficoltà di calcolo che ne scaturirebbero). Un'ulteriore differenza è che questi due autori considerano scarsamente rilevanti i "tradimenti incrociati", come vedremo in seguito.

Assumendo quindi la tripartizione dell'elettorato appena vista<sup>7</sup>, i flussi che concorrono a determinare il segno e l'intensità dei differenziali di coalizione possono essere espressi come valori di cella della Tabella 6.1.

I marginali presenti nella tabella si riferiscono al totale nazionale del risultato delle elezioni del 2001 alla Camera. I totali di riga esprimono la percentuale (sul totale degli elettori) ottenuta nel proporzionale dalla somma dei partiti di ciascuna coalizione e dell'area residuale (incluse le astensioni); i totali di colonna esprimono viceversa le percentuali nazionali ottenute nel maggioritario dalle coalizioni e dall'area residuale. Oggetto della stima sono dunque i valori di cella della Tabella 6.1.

Le celle sulla diagonale esprimono la quota di elettori che ha confermato nell'arena maggioritaria il proprio voto proporzionale. Se tutti gli elettori si comportassero così, le coalizioni otterrebbero nel maggioritario gli stessi voti del proporzionale, e non si avrebbe differenziale di coalizione.

---

<sup>7</sup> Da qui in avanti, i voti di Rifondazione Comunista (che correva autonomamente nel proporzionale, ma non presentava propri candidati nel maggioritario) vengono considerati all'interno dell'Ulivo.



Tabella 6.1 – Definizione dei comportamenti di voto differenziato (% su elettori)

Proporzionale	Maggioritario			Totale
	CdL	Area residuale	Ulivo	
CdL		dispersione	conflitto	37,2
Area residuale		attrazione	attrazione	32,3
Ulivo		conflitto	dispersione	30,5
<i>Totale</i>	34,2	32,9	32,9	100

Passando ai comportamenti di voto differenziato, le celle mediane rispettivamente della prima e dell'ultima riga identificano il fenomeno della *dispersione*: si tratta di quegli elettori che nel proporzionale votano all'interno delle due coalizioni, ma nel maggioritario si astengono o scelgono una terza forza.

Riguardo alla riga centrale, la prima e l'ultima cella identificano invece il fenomeno speculare dell'*attrazione*: elettori che nel proporzionale si astengono o votano terze forze, ma nel maggioritario scelgono una delle due coalizioni.

Infine le celle degli angoli in alto a destra e in basso a sinistra configurano il vero e proprio *confitto*, ovvero il comportamento di chi nel proporzionale sceglie il partito di una coalizione, e nel maggioritario vota la coalizione avversaria. È un comportamento di voto che agisce due volte, e sui differenziali di entrambe le coalizioni: in negativo per la coalizione votata nel proporzionale; in positivo per la coalizione votata nel maggioritario.

Obiettivo dell'analisi è dunque quello di stimare i singoli valori di cella in maniera da quantificare la consistenza dei diversi effetti che concorrono a formare i differenziali delle due coalizioni.

Veniamo a questo punto all'organizzazione complessiva dell'analisi. Essendo disponibili dati di sezione per tutto il territorio nazionale<sup>8</sup>, è decisamente interessante procedere a un'analisi su larga scala, ovvero analizzando le oltre 60.000 sezioni. Tuttavia una stima unica a livello nazionale ovviamente non ha senso. Questo perché, anche se l'approccio EI ha ipotesi meno impegnative del modello di Goodman, è comunque vero che si assume perlomeno l'unimodalità dei coefficienti da stimare, nella loro distribuzione su tutte le unità ecologiche incluse nell'aggregato da analizzare. La strategia più corretta è quindi di suddividere l'analisi dell'intero territorio nazionale in numerose sottoanalisi su aggregati locali di piccole dimensioni, dove si possa ipotizzare senza grossi problemi l'unimodalità dei parametri. La scelta cade quindi, comprensibilmente, sull'analisi per collegio elettorale. I vecchi collegi uninominali della Camera sono anzitutto collegati sostanzialmente ai fattori di competizione locale (i candidati di collegio). Inoltre hanno la caratteristica desiderabile di essere uniformi in termini di numero di elettori (intorno ai 100.000 per ogni collegio).

A partire dalle oltre 60.000 sezioni elettorali presenti nei dati forniti dal Ministero degli Interni, è stata stimata, mediante la procedura Eidis, una tabella analoga alla Tabella 6.1 per ciascun collegio elettorale<sup>9</sup>. Le tabelle relative a ciascun collegio sono state infine aggregate<sup>10</sup> in un'unica tabella riassuntiva nazionale (Tabella 6.2).

<sup>8</sup> Forniti dal Ministero dell'Interno.

<sup>9</sup> È stata esclusa la Valle d'Aosta, che elegge il proprio unico deputato in un collegio uninominale, nonché altri 8 collegi per incongruenze nei dati di sezione; le tabelle stimate sono quindi 466.

<sup>10</sup> L'aggregazione è avvenuta come segue: i valori di cella in ciascun collegio elettorale sono stati moltiplicati per la popolazione elettorale del collegio, ottenendo così un numero "virtuale" di elettori presenti nella cella in questione; sommando gli elettori virtuali presenti in ogni cella di ciascun collegio si ottiene la numerosità della cella al livello di aggregazione nazionale, da cui è possibile calcolare la percentuale sul totale nazionale.

Tabella 6.2 – Stima Eidis dei comportamenti di voto differenziato – totale nazionale (‰ su elettori)

Proporzionale	Maggioritario			Totale
	CdI	Area residuale	Ulivo	
CdI	30 ( $\pm 2$ )	3 ( $\pm 2$ )	4 ( $\pm 2$ )	37,1
Area residuale	2 ( $\pm 2$ )	29 ( $\pm 2$ )	2 ( $\pm 2$ )	32,7
Ulivo	2 ( $\pm 2$ )	1 ( $\pm 2$ )	27 ( $\pm 2$ )	30,1
Totale	34,1	33,2	32,6	100

Va anzitutto precisato che, per ogni cella, viene riportato anche un *range* relativo all'approssimazione della stima. Non si tratta di un intervallo di confidenza, ma della semplice approssimazione dovuta alla *previsione* scelta in fase di stima. Come visto alla fine del capitolo precedente, per contenere i tempi di calcolo è infatti necessario contenere la precisione della stima. In questo caso la precisione è del 10% rispetto al totale di riga: in sostanza, i coefficienti di cella possono assumere un valore compreso tra 0 e 1 in proporzione al totale di riga, ma con “scarti” di 0,1. In questo modo il valore ottimale<sup>11</sup> del coefficiente potrebbe trovarsi a metà strada tra due “scarti”: di qui l'intervallo di approssimazione di  $\pm 5\%$  rispetto al totale di riga, che si traduce in un'approssimazione di due punti percentuali sul totale degli elettori<sup>12</sup>. E questo, tra l'altro, il

<sup>11</sup> Cioè quello che fa parte di un modello che minimizza la distanza dai vincoli dovuti ai marginali.

<sup>12</sup> In tutte e tre le righe il 5% del totale di riga è compreso tra 1,5 e 1,85, quindi viene arrotondato a 2.

motivo per cui le percentuali sono riportate intere, senza decimale<sup>13</sup>.

### 6.2.1. Confronto con altre stime

Prima di commentare sostanzialmente i risultati, vale la pena di confrontare le stime ottenute con altre stime relative allo stesso fenomeno. Le possibilità di confronto sono essenzialmente tre: su dati aggregati, i dati provenienti dall'analisi citata precedentemente, effettuata però mediante il modello di Goodman [De Sio 2006a]; su dati individuali, le stime derivanti da un panel Abacus [Ricolfi e Testa 2002], e infine i dati dell'indagine postelettorale Itanes 2001.

L'eterogeneità di fonti dà l'idea del vero problema di fondo dell'inferenza ecologica, e del motivo per cui le tecniche di inferenza ecologica sono comunque utili, anche in presenza di dati individuali. Il problema infatti è che ovviamente anche i dati individuali sono, come vedremo anche in questo caso, soggetti agli errori di misurazione più disparati; di conseguenza semplicemente non è disponibile un dato “vero”, rispetto al quale si possano testare le stime in base a dati aggregati, perché lo stesso *survey* è soggetto a numerose fonti di errore<sup>14</sup>.

Nel caso specifico, esistono specifici fattori di distorsione relativi a tutte e tre le fonti dati utilizzate per il confronto. La prima fonte (modello di Goodman) è stata stimata sugli stessi dati usati

<sup>13</sup> Inoltre i marginali di riga e colonna non sono stati calcolati aggregando i voti effettivi, ma sommando tra loro i voti “virtuali” stimati dalla procedura in ogni cella. A causa degli arrotondamenti dovuti all'aggregazione, esiste quindi una lieve differenza tra i marginali reali e quelli calcolati in base alle stime, che ammonta allo 0,4% per tutta la tabella. Nella tabella sono stati riportati questi ultimi per permettere di confrontare tra loro i valori di cella senza incongruenze nei calcoli.

<sup>14</sup> Da questo punto di vista, l'unica possibilità teorica di controllo sistematico dell'accuratezza di stima di diverse procedure di inferenza ecologica può avvenire su dati simulati; e non a caso è l'approccio usato tra gli altri da Biorcio e Natale [1987]. Il problema è che però gli stessi dati simulati devono essere simulati secondo un modello, che necessariamente non è neutrale rispetto alla tecnica di stima che si intende sottoporre a controllo.

per la stima EIDIS, ma con la significativa differenza che le stime del modello di Goodman erano state originariamente effettuate non su una matrice di flusso  $3 \times 3$ , ma su una matrice relativa all'intera offerta politica disaggregata, quindi  $7 \times 15$  (i sette simboli di coalizione del maggioritario, e i quindici partiti presenti nel proporzionale). Il passaggio alla matrice  $3 \times 3$  è stato effettuato aggregando i valori di cella della matrice nazionale  $7 \times 15$ , invece che effettuando una nuova stima su una matrice  $3 \times 3$ . È noto [Corbetta e Schadee 1984] che in casi come questi le stime sono destinate ad essere comunque in parte diverse.

Per quanto riguarda i dati di sondaggio, non si deve pensare che le distorsioni siano meno rilevanti. Partendo dall'indagine posteleitoriale Itanes, possiamo ipotizzare almeno tre fonti di distorsione. La prima è l'errore di rappresentatività, dovuto sia alla procedura di campionamento che all'errore di misurazione della scelta di voto<sup>15</sup>: il risultato è che, anche ricorrendo a una ponderazione sul voto maggioritario, la distribuzione dei marginali relativi al voto proporzionale non corrisponde all'effettivo risultato registrato nel proporzionale<sup>16</sup>. La seconda è relativa alla nota sottorappresentazione dell'astensione nelle indagini campionarie, dovuta alla perdurante percezione di non accettabilità sociale dell'astensione elettorale<sup>17</sup>. Considerazioni simili valgono per la terza forma di di-

<sup>15</sup> Dove agiscono ovviamente una pluralità di cause: non veridicità o reticenza a dichiarare il voto non distribuite uniformemente tra tutti i partiti; difficoltà a ricordare il voto, ecc.

<sup>16</sup> In questi casi è possibile applicare l'algoritmo RAS anche a una tabella derivante da dati di sondaggio, per renderla compatibile con i marginali "veri". In questo caso tuttavia ciò si sarebbe tradotto in una sottostima dell'errore di misurazione del *margin*, visto che la procedura EIDIS non necessita di correzione.

<sup>17</sup> Per ovviare a questa distorsione, si è posta l'ipotesi, abbastanza forte, che, di fatto, le indagini campionarie si riferissero a un campione composto interamente di votanti. Di conseguenza, per passare a percentuali su elettori, si sono moltiplicati tutti i valori di cella per la percentuale di votanti alle elezioni del 2001 (circa l'81%), e gli astenuti sono stati inseriti nella cella centrale. L'operazione è stata effettuata anche relativamente al panel Abacus.

storsione, ovvero una probabile reticenza a dichiarare comportamenti differenziati di voto. Su questo non esiste una letteratura specifica; tuttavia sappiamo che in generale, l'intervistato ha la tendenza a razionalizzare e fornire una coerenza artificiosa alle risposte relative ad atteggiamenti politici, in considerazione del fatto che di norma la coerenza è considerata un valore socialmente desiderabile. Da questo punto di vista, non è peregrino pensare che anche i comportamenti di voto differenziato possano essere sotto-stimati da indagini campionarie. Le stesse considerazioni valgono ovviamente anche per il panel Abacus.

Fatte queste premesse, la Tabella 6.3 riporta i risultati del confronto tra le varie stime dei comportamenti di voto differenziato.

Vediamo quindi brevemente le principali differenze tra le stime, seguendo l'ordine dei vari tipi di comportamento di voto. La prima categoria è quella che possiamo chiamare della "coerenza", relativa cioè agli elettori che votano per la stessa coalizione nel maggioritario e nel proporzionale. In generale le stime si possono considerare abbastanza simili per tutte e tre le aree considerate, e in tutte le stime esaminate. Il bacino coerente della Cdl va da un minimo del 30% (Eidits) a un massimo del 33,1% (panel Abacus); l'intervallo di variazione è molto simile per l'Ulivo (tra il 27 e il 29,2%) e per l'area residuale (tra il 28,4 e il 30,6%). In generale le stime fornite dai due sondaggi hanno valori leggermente più alti rispetto a quelle calcolate in base ai dati aggregati; non è però possibile dire se ciò sia dovuto a una sovrastima della coerenza da parte dei sondaggi (a causa di quell'ipotesi di *social desirability* vista precedentemente) o a una sottostima da parte delle procedure di stima su dati aggregati.

Passiamo a questo punto ai comportamenti di voto differenziato. Il primo che esaminiamo è relativo all'area del *confitto*, ovvero degli elettori che votano coalizioni opposte nelle due aree. Qui osserviamo una variabilità che, in termini di percentuali complessive di elettori, è simile a quanto appena visto a proposito della *coerenza*. Per entrambe le possibilità di *confitto* (proporzionale Ulivo con maggioritario Cdl, e il viceversa) osserviamo infatti differenze di stima intorno ai 2-3 punti percentuali.

Tabella 6.3 – Confronto tra diverse stime del voto differenziato

	Dati aggregati		Dati individuali		
	Eidís	Goodman <sup>a</sup>	Ianes (N=2449)	Abacus <sup>b</sup> (N=n.d.)	Media
<i>Coerenza</i>					
Cdl 30 (±2)		32,3	32,5	33,1	32,0
Ulivo 27 (±2)		27,4	29,2	28,3	27,9
Area residuale 29 (±2)		28,4	30,6	30,0	29,4
<i>Conflitto</i>					
PR Ulivo e MG Cdl 2 (±2)		0,5	0,5	0	0,7
PR Cdl e MG Ulivo 4 (±2)		1,8	0,8	2,1	2,2
<i>Attrazione su area residuale</i>					
Cdl 2 (±2)		0,9	1,1	1,0	1,3
Ulivo 2 (±2)		3,3	2,6	2,1	2,5
<i>Dispersione verso area residuale</i>					
Cdl 3 (±2)		2,5	1,1	1,8	2,2
Ulivo 1 (±2)		1,7	1,2	1,4	1,5

<sup>a</sup>De Sio 2006a].

<sup>b</sup>Ricoli e Testa 2002].

Il punto è che il fenomeno è molto meno rilevante numericamente: di conseguenza una differenza di due punti è molto più importante. Così la quota di elettorato che sceglie Ulivo nel proporzionale e Cdl nel maggioritario varia dall'inesistenza nel panel Abacus, al 2% nella stima Eidís; e il comportamento opposto varia dallo 0,8% dell'indagine Ianes al 4% della stima Eidís. In sostanza siamo di fronte al fatto, messo d'altronde in luce da tutta la letteratura vista nel capitolo 2, che stimare flussi di consistenza molto ridotta è inevitabilmente soggetto a errori di stima molto rilevanti, in grado di precludere qualunque interpretazione affidabile.

Tuttavia le stime non sono completamente prive di informazioni. Ciò che appare abbastanza simile in tutte e quattro le fonti utilizzate è infatti il *rapporto* tra Cdl e Ulivo. È quest'ultima coalizione ad essere premiata nel maggioritario dai comportamenti di

conflitto: la combinazione “proporzionale Cdl e maggioritario Ulivo” è maggiore del suo opposto in tutte e quattro le stime presentate. Ciò si traduce in una netta differenza tra le medie dei due fenomeni: il meccanismo che premia l'Ulivo al maggioritario viene stimato in media per un 2,2% dell'elettorato, a fronte di uno 0,7% che segue il comportamento opposto. Va notato che la stima Eidís fornisce i risultati più alti, stimando rispettivamente una consistenza del 4 e del 2%.

Il secondo comportamento di voto differenziato è invece quello dell'*attrazione* corrisponde a elettori che nel proporzionale votano l'area residuale, ma vengono “attratti” dalle due coalizioni nel maggioritario. Qui le differenze tra stime sono leggermente più contenute: la stima è compresa tra 0,9 e 2% per la Cdl, e tra 2 e 3,3% per l'Ulivo. Ma ancora una volta il rapporto tra coalizioni appare abbastanza univoco, con una preponderanza dell'Ulivo rispetto alla Cdl, anche se più leggera rispetto a quanto visto a proposito del *conflitto*. Fa però eccezione la stima Eidís, che riporta un valore del 2% per entrambe le coalizioni. Tuttavia va osservato che il fatto che i valori siano uguali è dovuto alla scarsa precisione della stima: i valori sono presentati non a caso interi, in quanto la precisione è di ±2 punti percentuali.

Veniamo infine al terzo comportamento di voto differenziato, ovvero alla *dispersione*. Si tratta di elettori che nel proporzionale votano una delle due coalizioni, ma nel maggioritario si “dispersiono” verso l'area residuale. In questo caso le differenze tra le stime sono ancora contenute: per la Cdl si va da un minimo dell'1,1 a un massimo del 3%; per l'Ulivo le stime sono comprese tra l'1 e l'1,7%. Ancora una volta però la struttura sembra abbastanza simile per tutte le stime presentate: la dispersione del voto proporzionale è maggiore per la Cdl. Fa eccezione l'indagine Ianes, che tuttavia presenta due valori praticamente uguali per le due coalizioni.

Quali conclusioni possiamo trarre da questo confronto tra stime? Al di là degli specifici valori quantitativi, tutte e quattro le fonti sembrano accreditare un'analisi del fenomeno abbastanza simile. Ovviamente i valori divergono in maniera più marcata quando si esaminano i fenomeni numericamente meno consistenti.

ti; tuttavia anche in quel caso il confronto tra varie celle sembra permettere in generale un'interpretazione sostanzialmente univoca. In questo senso la procedura Eidis sembra permettere, già in questa fase di prototipo, risultati abbastanza simili a quelli del modello di Goodman, tuttavia con una maggiore coerenza di base in termini teorici. Sappiamo infatti che la procedura produce questi risultati a partire da un quantitativo di informazione maggiore, in quanto tiene conto dei vincoli dei marginali di tutte le sezioni elettorali, e perciò non richiede correzioni e aggiustamenti come quelli che sono indispensabili per rendere compatibili i risultati del modello di Goodman.

Tuttavia la tabella vista precedentemente mette anche in evidenza la limitazione della procedura: soprattutto per esaminare i fenomeni di minore consistenza numerica, è indispensabile poter lavorare a livelli di precisione maggiore. D'altronde in questa analisi l'obiettivo era semplicemente di ottenere delle prime stime, per verificare la plausibilità dei risultati della procedura, e senza la necessità di ottimizzare le prestazioni per un uso con precisioni maggiori. Da questo punto di vista l'esito può dirsi decisamente soddisfacente in quanto, considerato l'intervallo di approssimazione utilizzato, le stime appaiono assolutamente compatibili con quelle effettuate con altre procedure. Inoltre la procedura di stima Eidis si presta a ulteriori ottimizzazioni in due direzioni. Da un lato in termini di velocità di calcolo e semplificazione degli algoritmi; dall'altro in termini di ricorsività. È infatti possibile concepire di effettuare un primo stadio di analisi nell'intero intervallo [0;1] per tutti i coefficienti; una volta individuata una prima soluzione, restringere l'intervallo di ricerca per ogni coefficiente, ed effettuare in quell'intervallo ristretto una ricerca con una precisione maggiore. Questo procedimento può essere ripetuto più volte, fino a raggiungere la precisione desiderata, senza costi proibitivi in termini computazionali.

Viste queste considerazioni sulla procedura, vale la pena a questo punto di passare all'analisi sostantiva, utilizzando i risultati della procedura Eidis per esaminare il bilancio complessivo del differenziale di coalizione.

## 6.2.2. Alle origini del differenziale di coalizione

Il punto di partenza dell'analisi sostantiva di questo capitolo era il fenomeno del differenziale di coalizione. Abbiamo visto come questo differenziale aggregato sia, in termini di comportamenti individuali, il frutto della somma (e della reciproca compensazione) di diversi meccanismi individuali, di cui abbiamo potuto stimare la consistenza complessiva. Veniamo quindi a ricomporre il quadro, ricapitolando in forma di bilancio i vari effetti e evidenziando come si è costruito nel 2001 il differenziale negativo della CdI e quello positivo dell'Ulivo.

Il dato fondamentale è che il distacco tra le due coalizioni, di sette punti percentuali<sup>18</sup> nel proporzionale, si riduce a un punto e mezzo nel maggioritario. I 5,5 punti di recupero dell'Ulivo nel maggioritario sono il risultato del differenziale negativo della CdI (-3%) e di quello positivo dell'Ulivo (+2,5%). Abbiamo visto però come questi saldi finali siano solo la punta di un iceberg: si tratta infatti solo del saldo complessivo di un articolato movimento di flussi in entrata e in uscita che arriva a coinvolgere quasi il 15% dell'elettorato<sup>19</sup>. Questa quota dell'elettorato ha attuato comportamenti di voto differenziato, a loro volta scomponibili nelle tre dinamiche presentate precedentemente:

- un 6% di elettori in *confitto*: ovvero che nel proporzionale votano il partito di una coalizione, ma nel maggioritario scelgono la coalizione avversaria;
- un 4% di elettori nell'area della *dispersione*: nel proporzionale votano il partito di una coalizione, ma nel maggioritario si disperdono nell'area residuale;
- un 4% di elettori nell'area dell'*attrazione*: nel proporzionale votano nell'area residuale, ma nel maggioritario sono attratti da una delle due coalizioni principali.

<sup>18</sup> Sugli elettori.

<sup>19</sup> Si tratta anche qui di una percentuale sugli elettori: in termini di voti validi la percentuale corrisponde circa a un 17-18%.

Tabella 6.4 – Bilancio dei comportamenti di voto differenziato (% su elettori)

	CdL	Ulivo
<i>Entrate</i>		
conflitto a proprio vantaggio	+2	+4
attrazione	+2	+2
<i>Totale entrate</i>	+4	+6
<i>Uscite</i>		
conflitto a proprio svantaggio	-4	-2
dispersione	-3	-1
<i>Totale uscite</i>	-7	-3
<i>Saldo</i>	-3	+3

Volendo distinguere gli effetti delle tre dinamiche di comportamento sulle due principali coalizioni, otteniamo una sorta di bilancio dei comportamenti di voto differenziato, con le relative entrate e uscite (Tabella 6.4).

I risultati principali che scaturiscono da questa analisi (peraltro molto simili a quelli stimati in base ad altre fonti) sono quindi essenzialmente due.

Da un lato, la dinamica del voto differenziato risulta di estensione rilevante. Ad adottare comportamenti di voto che differenziano scelta proporzionale e maggioritaria è infatti circa il 15% degli elettori, corrispondente al 17-18% dei voti validi.

Dall'altro lato, questa dinamica risulta anche decisamente complessa. Come messo in evidenza da altre analisi, il saldo finale è composto in realtà da numerosi flussi che coinvolgono le due principali coalizioni e l'area residuale sia in entrata che in uscita, delineando un quadro non scontato. In particolare, a colpire è la presenza e la rilevanza del fenomeno del *conflitto*, non a caso considerato trascurabile in alcune delle proposte teoriche viste in precedenza. Considerando la media di tutte le fonti di stima che abbiamo considerato, la consistenza dell'area in conflitto appare invece molto simile a quella degli altri due fenomeni di voto diffe-

renziato<sup>20</sup>. In più il fatto che gli scambi diretti tra le due coalizioni siano altrettanto rilevanti di quelli con l'area residuale mette in dubbio eventuali assunti su una distribuzione spaziale perfettamente unidimensionale delle due coalizioni, divise da un'area residuale. In realtà, nonostante la presenza di terze forze intermedie, i passaggi diretti tra coalizioni appaiono presenti, e di una certa consistenza numerica.

Ovviamente, il limite dell'analisi basata su dati aggregati è quello di non poter esplorare le *motivazioni* dei diversi comportamenti di voto differenziato, con particolare riferimento al tipo di voto considerato più rilevante dall'elettore<sup>21</sup>. Dal punto di vista del risultato elettorale finale, una tale distinzione non è particolarmente rilevante<sup>22</sup>, ma dal punto di vista dell'interpretazione politica sostantiva dei risultati elettorali lo è senza dubbio. In particolare, se si tenta di dare un'interpretazione sostantiva del fenomeno del differenziale di coalizione.

Quando una coalizione prende più voti dei propri partiti, ciò infatti può voler dire almeno due cose. Da un lato, che la coalizione "pesca" voti anche fuori del proprio tradizionale bacino elettorale, che si suppone di poter misurare in base ai voti proporzionali. Ma, provando a cambiare punto di vista, si può ipotizzare che la misura reale del bacino elettorale della coalizione sia proprio il voto maggioritario, piuttosto che quello proporzionale<sup>23</sup>. Sotto que-

<sup>20</sup> La presenza e la rilevanza del fenomeno del *conflitto* va ovviamente in direzione opposta all'ipotesi del "centro scambiatore".

<sup>21</sup> Anche se l'influenza di alcune variabili di contesto può essere studiata mediante un'analisi multivariata sui coefficienti stimati; si tratta della seconda parte dell'analisi citata precedentemente [De Sio 2006a].

<sup>22</sup> Il voto espresso da questi due ipotetici elettori è infatti identico, sia sulla scheda proporzionale che su quella maggioritaria.

<sup>23</sup> Una simile ipotesi, almeno per quanto riguarda l'Ulivo, è forse meno azzardata di quanto si possa ritenere. Ad esempio confrontando i totali nazionali di voti maggioritari e proporzionali riportati dall'Ulivo nel 1996 e nel 2001 (vedi ad esempio Bartolini e D'Almondo [2002, 231]) si nota un numero di voti maggioritari abbastanza stabile tra le due consultazioni, mentre a variare sensibilmente è il numero di voti proporzionali.

sta ipotesi diventa accettabile che, per una quota più o meno significativa di elettori (non necessariamente distribuita omogeneamente tra le coalizioni), sia il voto maggioritario a essere quello più importante, quello che esprime in maniera più fedele le proprie preferenze politiche. In questa direzione sembrano andare i risultati delle domande relative a questi temi all'interno dell'indagine Itanes 2001. Alla domanda su quale fosse il voto in cui l'intervistato si riconosceva maggiormente, il 27% rispondeva "maggioritario", il 35% "proporzionale", il 38% "tutti e due allo stesso modo" [Maraffi 2002, 325].

L'accettazione di un simile presupposto deve portarci dunque a considerare che il concetto di "rendimento coalizionale" potrebbe essere anche rovesciato. È infatti possibile che un differenziale di coalizione fortemente positivo non indichi tanto la capacità della coalizione di attrarre voti di altri elettori che "sinceramente" apparterebbero ad altre aree politiche, quanto, paradossalmente, l'incapacità dei partiti che compongono quella coalizione di attrarre sulle loro liste nell'arena proporzionale i voti "sinceri" ottenuti nell'arena maggioritaria.

Questa interpretazione è stata proposta storicamente a proposito del centrosinistra [Ricoli e Testa 2002]. La plausibilità di questa interpretazione può essere difesa prendendo in esame la successiva evoluzione dell'Ulivo. Già nelle elezioni europee del 2004 la lista comune dell'Ulivo, pur se in un'elezione completamente proporzionale, ha riportato, in alcune aree del paese, risultati significativamente migliori della somma di Ds e Margherita<sup>24</sup>. Lo stesso è d'altronde avvenuto nelle successive elezioni politiche del 2006, in cui la lista comune dell'Ulivo alla Camera ha ottenuto, questa volta in tutta Italia, risultati migliori della somma di Ds e Margherita al Senato. Da questo punto di vista, anche la costituzione del Partito Democratico sarebbe stata difficile da concepire senza che avesse iniziato a manifestarsi la disponibilità degli elettori a pre-

miare un soggetto politico di dimensioni maggiori dei partiti originari.

Tornando al 2001, è chiaro che l'eventuale presenza di una vera e propria identità di coalizione può essere indagato nel modo migliore solo attraverso dati di sondaggio [Maraffi 2002]. Tuttavia l'analisi su dati aggregati è in grado di offrire informazioni molto ricche in una direzione diversa da quella del *survey*: la possibilità di analizzare i fenomeni nella loro eterogeneità geografica. In questo caso è la natura stessa dell'analisi, disaggregata per collegio elettorale per i motivi tecnici visti precedentemente, a offrirci già i dati per un'analisi dettagliata a livello territoriale. Per motivi di spazio non è ovviamente possibile presentare tutte le tabelle relative ai vari collegi elettorali: tuttavia molte informazioni possono essere condensate in forma cartografica. A questo ulteriore prodotto dell'analisi è dedicato l'ultimo capitolo.

<sup>24</sup> Se confrontata con le liste separate presentate nelle contemporanee elezioni amministrative.

## 7. Le mappe del voto differenziato

Veniamo dunque a presentare un importante, ulteriore prodotto reso possibile dall'analisi basata su dati aggregati: la distribuzione geografica dei vari tipi di comportamento di voto differenziato.

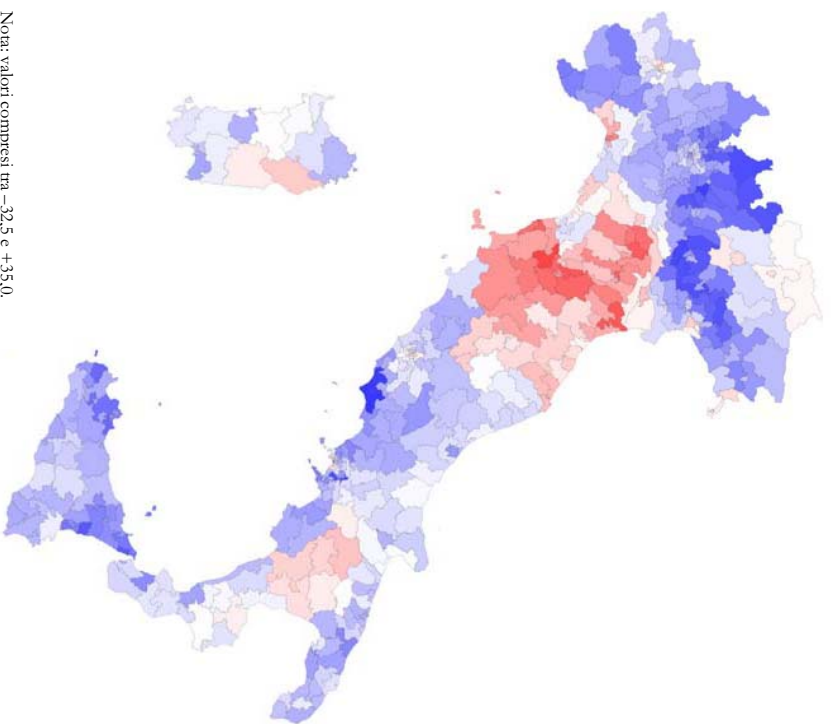
In prima battuta è necessario fornire anzitutto un inquadramento generale, in termini territoriali, del fenomeno aggregato del differenziale di coalizione. Qual è la distribuzione territoriale di questo fenomeno? E soprattutto: esistono marcate eterogeneità territoriali, tali ad esempio da lasciar supporre che il differenziale positivo dell'Ulivo a livello nazionale non sia tale in tutta Italia? Vale lo stesso per la CdL? La risposta a questi interrogativi viene dai dati presentati nel prossimo paragrafo.

### 7.1. Il differenziale di coalizione

Per inquadrare al meglio il fenomeno del differenziale di coalizione, il punto di partenza obbligato è anzitutto quello di riepilogare brevemente i rapporti di forza tra le coalizioni su base territoriale, sia relativamente al maggioritario che al proporzionale. Le due distribuzioni territoriali sono presentate rispettivamente nella Figura 7.1 per il proporzionale e nella Figura 7.2 per il maggioritario.



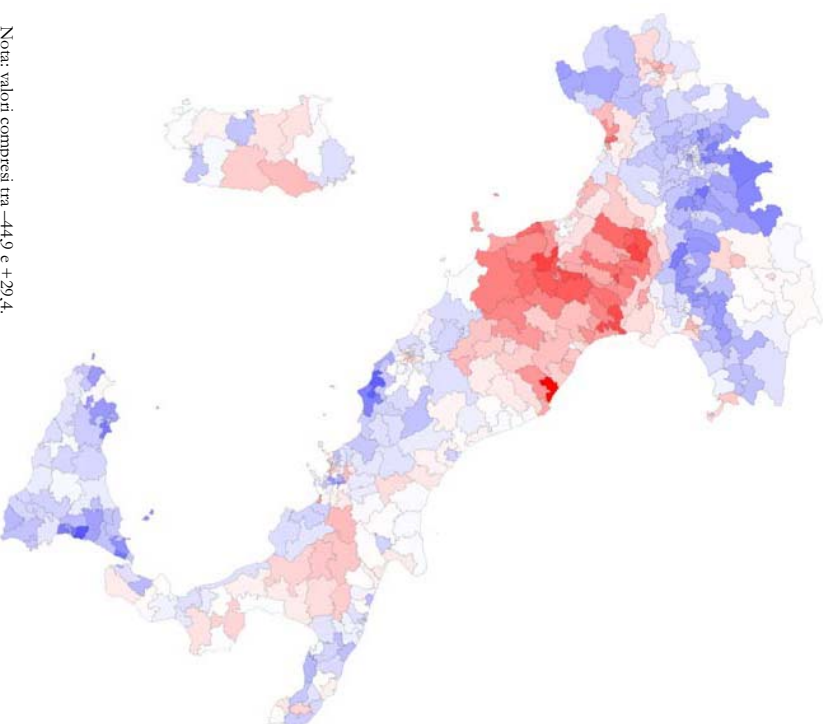
Figura 7.1 – Distacco Cdl-Ulivo nel proporzionale (% sugli elettori)



Nota: valori compresi tra -32,5 e +35,0.

In entrambe le figure la convenzione cromatica indica con gradazioni di blu il grado di prevalenza della Cdl, e con gradazioni di rosso la prevalenza dell'Ulivo. Entrambe le mappe utilizzano la stessa scala cromatica assoluta: le tonalità delle due mappe possono quindi essere confrontate direttamente.

Figura 7.2 – Distacco Cdl-Ulivo nel maggioritario (% sugli elettori)



Nota: valori compresi tra -44,9 e +29,4.

Osservando i distacchi proporzionali fra le due coalizioni<sup>1</sup> si rivedono anzitutto i tratti della geografia elettorale del 2001 [Chiaromonte 2002; Diamanti 2003].

La prevalenza complessiva della Cdl nel proporzionale appare nettamente anche in termini geografici. I partiti della coalizione di

---

<sup>1</sup> Nelle due mappe, la tonalità più scura di ciascun colore corrisponde a un distacco del 45%.

centrosinistra<sup>2</sup> dispongono di un bacino di consensi superiore alla Cdl soltanto in alcune aree ben definite. L'insediamento principale dei partiti dell'Ulivo è prevedibilmente costituito dalla tradizionale Zona rossa<sup>3</sup>, tuttavia non nella sua interezza. Al di fuori di quest'area, i partiti dell'Ulivo superano quelli della Cdl solo in pochi collegi urbani, sia in alcune città del Nord che a Roma e Napoli; infine in pochi collegi sparsi tra Basilicata, Calabria e Sardegna. In tutto il resto d'Italia i partiti del centrodestra vantano una prevalenza abbastanza netta; prevalenza che si fa ancora più marcata nella fascia pedemontana del Nord-Est e in alcune aree della Sicilia.

Il confronto con il maggioritario rivela invece un quadro in parte diverso. La distribuzione geografica conferma la presenza di un differenziale di coalizione. Infatti il quadro complessivo, anche da un semplice confronto cromatico generale, appare sensibilmente più competitivo, ricordando che l'uniformità delle scale cromatiche permette un confronto diretto.

Il confronto tra arena proporzionale e maggioritaria vede quindi, abbastanza uniformemente su tutto il territorio nazionale, l'Ulivo più competitivo nel maggioritario rispetto al proporzionale. Tuttavia occorre osservare come la maggiore competitività del centrosinistra (complessivamente uniforme) sia comunque insufficiente, in molte aree del paese, a creare un contesto competitivo nei collegi uninominali.

I collegi che possiamo definire competitivi sono infatti quelli dove il distacco tra le due coalizioni è minimo [Bartolini e D'Alimonte 2002b]. Nella Figura 7.2 sono quelli individuati dalle tonalità

---

<sup>2</sup> Come già visto nel capitolo precedente, i voti di Rifondazione Comunista (che correva autonomamente nel proporzionale) vengono considerati all'interno dell'Ulivo.

<sup>3</sup> Facendo riferimento alla classificazione di Barbagli e Corbetta [1980], a sua volta una rettificazione su confini regionali della proposta originaria, della fine degli anni '60, del gruppo di ricerca dell'Istituto Cattaneo diretto da Giorgio Galli [Galli et al. 1968]. Vedi anche Cartocci [1987; 1990].

tà più tenui di rosso o di blu, vicine al bianco: questo indica che una coalizione ha prevalso sull'avversaria con percentuali minime.

Ciò premesso, vediamo in quali aree geografiche vi sono più collegi competitivi. Al Nord la miglior performance maggioritaria dell'Ulivo sembra comunque insufficiente a mettere in discussione il primato della Cdl. Pur se con distacchi minori rispetto al proporzionale, l'Ulivo vede quasi sempre confermato il proprio svantaggio. Una situazione simile è quella del Lazio e della Sicilia. È senza dubbio vero che anche in queste roccaforti del centrodestra l'Ulivo riesce a fare sensibilmente meglio nel maggioritario, e l'effetto è chiaramente visibile dalle tonalità azzurre meno nette; tuttavia si tratta di un risultato chiaramente insufficiente per definire i collegi come competitivi.

Il differenziale positivo dell'Ulivo tra maggioritario e proporzionale ha invece effetti rilevanti nel resto d'Italia, anche per quello che riguarda il grado di competitività dei collegi uninominali. In particolare nel Centro-Sud, viene infatti a delinearsi un'estesa area di competitività che comprende Abruzzo, Molise, e parte di Campania, Puglia, Calabria e Sardegna. In tutte queste aree geografiche la maggior parte dei collegi, per l'una e l'altra coalizione, viene aggiudicata con distacchi molto contenuti.

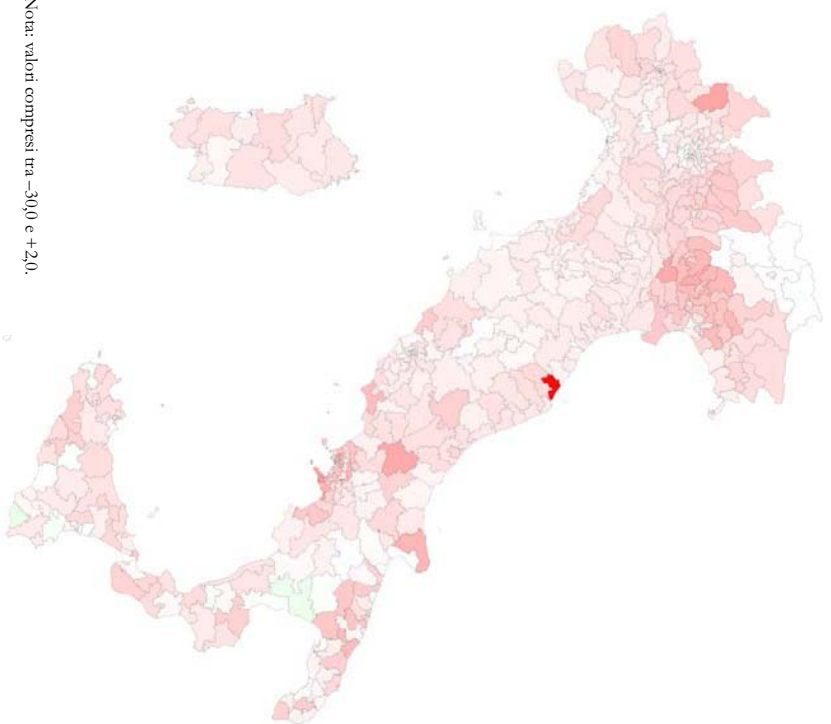
D'altronde, la distribuzione geografica inevitabilmente rispetta il risultato complessivo, e quindi l'importanza del differenziale di coalizione. Come abbiamo visto precedentemente (Tabella 6.2), il vantaggio della Cdl nel proporzionale, di sette punti percentuali<sup>4</sup>, si riduce a circa un punto e mezzo nell'arena maggioritaria.

Il confronto della prevalenza nel proporzionale e nel maggioritario sembrerebbe quindi suggerire una distribuzione territoriale abbastanza uniforme del differenziale di coalizione. E in effetti questo dato è assolutamente confermato dalle distribuzioni territoriali del differenziale di coalizione vero e proprio, mostrate separatamente per entrambe le coalizioni.

---

<sup>4</sup> Sugli elettori.

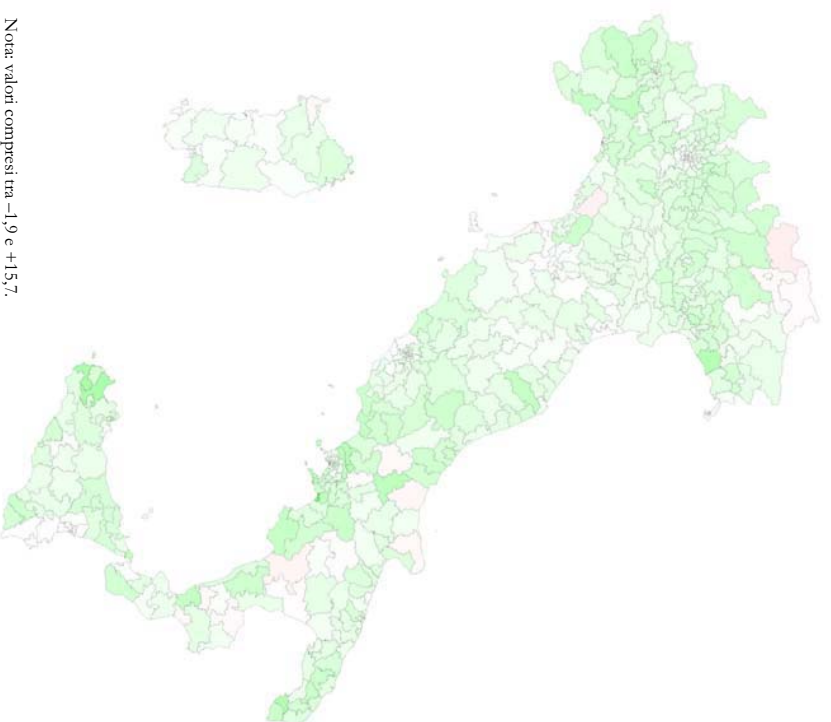
Figura 7.3 – Differenziale di coalizione per la Cdl (punti percentuali su elettori)



Nota: valori compresi tra -30,0 e +2,0.

Queste due distribuzioni sono presentate per la Cdl e per l'Ulivo rispettivamente nella Figura 7.3 e nella Figura 7.4. In questo caso la convenzione cromatica è ovviamente diversa dalla precedente: tonalità di verde indicano un differenziale di coalizione positivo (più voti nel maggioritario che nel proporzionale); tonalità di rosso indicano invece un differenziale negativo.

Figura 7.4 – Differenziale di coalizione per l'Ulivo (punti percentuali su elettori)



Nota: valori compresi tra -1,9 e +15,7.

Il dato che colpisce maggiormente è la notevole omogeneità territoriale del differenziale di coalizione (positivo o negativo) della Cdl e dell'Ulivo<sup>5</sup>. Il centrosinistra presenta un differenziale positivo quasi in tutti i collegi, così come la Cdl ha un differenziale sistematicamente negativo. In questo senso, la distribuzione terri-

<sup>5</sup> In queste due mappe la tonalità più scura corrisponde a un rendimento, positivo o negativo, del 31%.

riale sembra confermare la scarsa rilevanza di un effetto delle candidature locali, in virtù invece di fattori esplicativi che vanno decisamente ricercati a livello nazionale. Questo non toglie ovviamente che alcune particolarità locali emergano. Da un lato una particolare difficoltà della CdL a tradurre i voti proporzionali in maggioritari nel Veneto (anche se si tratta di un'area in cui il primato maggioritario della CdL non è in discussione). Dall'altro, e non sorprendentemente, il differenziale positivo dell'Ulivo si attenua molto nelle aree di insediamento storico dei partiti del centrosinistra; questo evidentemente a testimonianza della permanenza di un solido bacino elettorale, capace di orientare il voto in modo compatto in entrambe le arene competitive.

Abbiamo quindi visto sinteticamente la distribuzione geografica dell'esito *complessivo*, ovvero del fenomeno rilevabile come *differenziale di coalizione*. Passiamo a questo punto alla presentazione dei risultati veri e propri dell'analisi, ovvero alla distribuzione geografica dei comportamenti di voto differenziato, stimati dalla procedura Eidis su dati aggregati.

## 7.2. Il voto differenziato e le sue componenti

Come abbiamo visto precedentemente, il differenziale di coalizione osservato, sia a livello nazionale che locale, è il frutto di una pluralità di comportamenti di voto differenziato. L'esito complessivo è quindi il risultato della somma (e dell'elisione reciproca) di fenomeni di tipo diverso, che abbiamo classificato come *confitto*, *attrazione* e *dispersione* (vedi il capitolo precedente). Ovviamente la consistenza complessiva di questi comportamenti ha un'influenza sulle dimensioni dell'area della *coerenza*, ovvero di quegli elettori che viceversa votano in modo coerente nel maggioritario e nel proporzionale. Iniziamo quindi dalla distribuzione territoriale di questo primo fenomeno.

### 7.2.1. Coerenza

Con il termine *coerenza* abbiamo definito il comportamento di quegli elettori che scelgono nel proporzionale il partito di una coalizione, e nel maggioritario votano la stessa coalizione. Si tratta ovviamente del comportamento di voto nettamente prevalente, sia per le due coalizioni principali che per l'area residuale.

Riguardo alle due coalizioni, potremmo dire che l'area della coerenza può fornire una misura di un bacino stabile del proprio elettorato, una sorta di "zoccolo duro" di elettori che perlomeno non presentano tratti di ambiguità tra le due arene competitive<sup>6</sup>. Le stime fornite dalla procedura Eidis a livello locale permettono quindi di fornire una stima della distribuzione territoriale complessiva della consistenza di questi due bacini elettorali.

Tuttavia il dato viene presentato sotto forma di percentuale sul totale dei voti *proporzionali* di ciascuna coalizione. Ciò che quindi viene visualizzato è quindi, in realtà, il tasso di coerenza al maggioritario dell'elettorato proporzionale delle due coalizioni, ovvero la capacità delle due coalizioni di vedere confermato al maggioritario il voto dei propri elettori proporzionali. Le due distribuzioni sono presentate, per la CdL e per l'Ulivo, rispettivamente nella Figura 7.5 e nella Figura 7.6.

Il primo dato che si osserva è la complessiva maggiore indisciplina nel maggioritario degli elettori proporzionali della CDL<sup>7</sup>. A livello nazionale, la procedura stima un tasso di conferma (in percentuale sui voti proporzionali) dell'81% della CdL, contro l'89% dell'Ulivo (vedi anche la Tabella 6.2).

<sup>6</sup> In questo senso, ad esempio, le stime territoriali di questi due bacini, ottenute dall'analisi precedentemente citata [De Sio 2006a], sono state poi usate per il confronto diacronico con il 2006, allo scopo di parametrizzare l'effetto del cambiamento di sistema elettorale tra 2001 e 2006 [De Sio 2007a].

<sup>7</sup> Non a caso già ampiamente sottolineata in letteratura. Vedi a questo proposito Bartolini e Dall'Amonte [2002].

Figura 7.5 – Coerenza dell'elettorato proporzionale Cdl (% sui voti PR della Cdl)



Nota: valori compresi tra 0 e 100.

Il fenomeno è abbastanza chiaramente visibile anche in termini di distribuzione territoriale complessiva: le tonalità di grigio relative alla Cdl sono infatti leggermente più chiare di quelle dell'Ulivo<sup>8</sup>. Un secondo punto rilevante è invece relativo all'Ulivo.

<sup>8</sup> Anche in questo caso la scala cromatica è la medesima.

Figura 7.6 – Coerenza dell'elettorato proporzionale Ulivo (% sui voti PR dell'Ulivo)



Nota: valori compresi tra 0 e 100.

La Figura 7.6 mette infatti in evidenza i tassi di coerenza notevolmente più alti registrati in particolare nella Zona rossa. Il dato è pienamente in linea con quanto visto precedentemente a proposito del differenziale di coalizione (la cui positività, per l'Ulivo, si attenuava proprio in quest'area del paese).

In sostanza, la differenza netta tra l'Ulivo e la Cdl è che nel 2001 l'Ulivo dispone di un nucleo di consensi territorialmente de-

finito, che si comporta in termini di voto come se fosse ancora presente un vero e proprio “insediamento subculturale?”. Questo per tre aspetti: l'assenza di competitività (il predominio dell'Ulivo in quest'area è netto); il differenziale di coalizione paradossalmente basso; l'alto tasso di coerenza degli elettori di centrosinistra tra proporzionale e maggioritario.

Questo non vale invece per la Cdl. La mappa della coerenza di coalizione degli elettori proporzionali della Cdl rivela non solo una coerenza complessiva più bassa, ma anche l'assenza di un insediamento territoriale vero e proprio. Anche le aree in cui il centro-destra straripante non sono caratterizzate da una maggiore coerenza rispetto alle altre. Questo vale sia per la Sicilia che per la fascia pedemontana del Nord-Est, dove la fedeltà alla coalizione degli elettori proporzionali della Casa delle Libertà è addirittura inferiore alla media nazionale. In questo senso l'impressione è che, rispetto alla *dimensione* della coalizione, la coesione del proprio elettorato abbia rappresentato per la Cdl, nel 2001, una risorsa tutto sommato secondaria.

Passiamo a questo punto al voto speculare della coerenza rispetto al voto proporzionale: veniamo quindi a esaminare la distribuzione territoriale dei comportamenti di voto differenziato.

### 7.2.2. *Conflitto*

Il primo fenomeno che prendiamo in esame è quello più insospettato sul piano numerico, e tutto sommato più curioso dal punto di vista sostanziale: il *conflitto*, ovvero il comportamento di quegli elettori che si dividono tra le due coalizioni nell'arena proporzionale e in quella maggioritaria. Diversamente dalle mappe precedenti, in questo caso la consistenza di questa area di elettorato viene rappresentata in termini di percentuale rispetto all'*intero elettorato*. Di conseguenza valuteremo proprio la stima della consistenza dell'area del conflitto relativa a Cdl e Ulivo. Le due distribuzioni

territoriali di questa variabile sono presentate rispettivamente nella Figura 7.7 e nella Figura 7.8<sup>9</sup>.

Le due mappe esprimono quindi la consistenza territoriale dell'area di *conflitto* relativa alle due coalizioni. Le due dinamiche (speculari per Cdl e Ulivo) sono valutate in termini dell'effetto che hanno sull'arena che, nella legge Mattarella, era la più importante per il risultato elettorale, ovvero quella maggioritaria. Di conseguenza il conflitto relativo alla Cdl (che ha cioè un risultato negativo per quest'ultima) è quello relativo agli elettori proporzionali di questa coalizione che nel maggioritario votano per l'Ulivo. La stessa considerazione vale specularmente per l'Ulivo.

Le due mappe ci mostrano con nettezza, a partire dalla tonalità cromatica complessiva, la differente rilevanza quantitativa del fenomeno per le due coalizioni. Va ricordato, in relazione alle stime viste nel capitolo precedente, che il *conflitto* subito dalla Cdl è quantitativamente circa il doppio (in percentuale sugli elettori) rispetto a quello subito dall'Ulivo.

Ciò premesso, la configurazione territoriale del fenomeno presenta alcune caratteristiche interessanti. Anzitutto, esiste un'ampia area di sovrapposizione delle aree di conflitto delle due coalizioni. Non è facile identificare un denominatore comune per le aree dove il conflitto è più rilevante. Un tratto in comune può però essere osservato nei contesti urbani: per entrambe le coalizioni, le città, più o meno a Nord come a Sud, hanno tassi di conflitto maggiori.

---

<sup>9</sup> Queste mappe, e tutte quelle che seguono, sono diseguate secondo le seguenti convenzioni: la variabile rappresentata è la percentuale sul totale degli elettori. La scala cromatica va dallo 0% (bianco) al 20% (nero). In alcuni casi, valori leggermente superiori al 20% sono stati rappresentati ugualmente in nero, mantenendo la scala cromatica limitata al 20% per evitare di rendere meno distinguibili gli altri valori. Essendo la scala cromatica identica per tutte le mappe che seguono, è possibile comparare cromaticamente dati di mappe diverse.

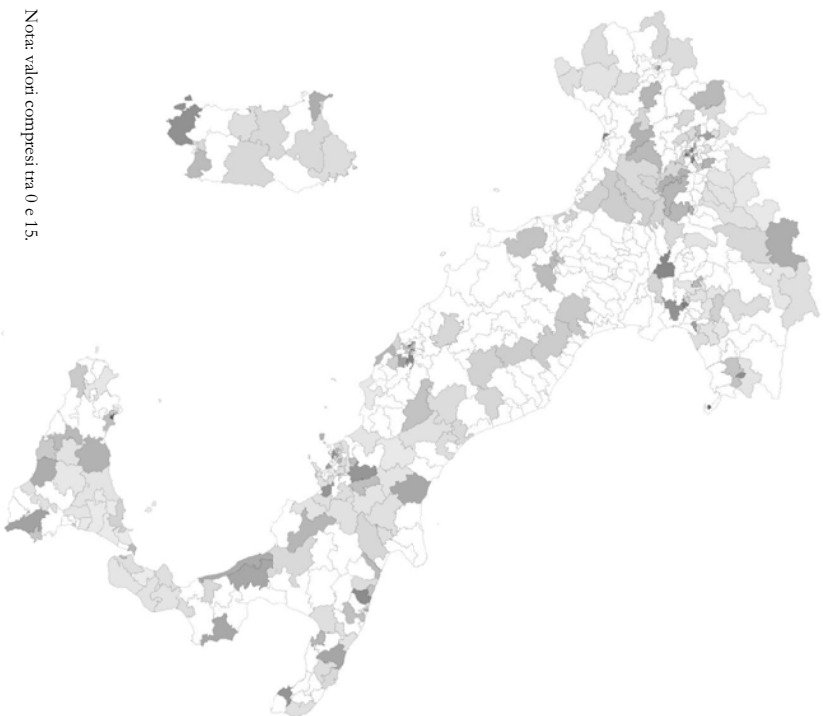
Figura 7.7 – Conflitto subito dalla Cdl (% su elettori)



Nota: valori compresi tra 0 e 15.

Tuttavia, tra le due mappe esistono anche significative differenze: quella relativa al conflitto per la Cdl mostra non solo un fenomeno più intenso, ma anche più esteso, visto che coinvolge un numero sensibilmente maggiore di collegi uninominali. Inoltre occorre dire che per, il centrosinistra, ricompare l'interpretazione che abbiamo già suggerito in precedenza: questa coalizione dispone di un insediamento geografico dove, non a caso, anche i tassi di conflitto sono molto bassi se non nulli.

Figura 7.8 – Conflitto subito dall'Ulivo (% su elettori)



Nota: valori compresi tra 0 e 15.

Ancora una volta non è così per la Cdl. La forza elettorale del centrodestra e il suo dominio anche nella competizione maggioritaria non appaiono correlate con l'assenza di conflitto: quest'ultimo è presente e rilevante anche in alcune zone chiave per il successo della Casa delle Libertà: la fascia pedemontana veneta, la Lombardia orientale, la Sicilia, il Lazio meridionale.

In ogni caso, il tratto dominante del fenomeno è una maggiore consistenza quantitativa a vantaggio del centrosinistra rispetto a quanto avviene per il centrodestra; e questo rapporto sembra pre-

sentare una certa uniformità territoriale. Vedremo come un andamento abbastanza simile venga registrato anche per la seconda categoria di voto differenziato.

### 7.2.3. Dispersione

Con questo termine abbiamo individuato il comportamento di quegli elettori che scelgono nel proporzionale un partito di una delle due coalizioni, ma nel maggioritario si astengono o votano una terza forza. Si tratta di un comportamento che quindi abbassa il differenziale della coalizione interessata, anche se – diversamente dal conflitto vero e proprio – non incrementa quello della coalizione avversaria.

Anche in questo caso la stima Eridis a livello nazionale indica un'incidenza del fenomeno sulla Cdl nettamente più marcata rispetto all'Ulivo. Questo dato è chiaramente confermato dalla distribuzione territoriale del fenomeno, visualizzata per le due coalizioni rispettivamente nella Figura 7.9 e nella Figura 7.10.

Ancora una volta i dati sono presentati in percentuale rispetto al totale degli elettori: si tratta quindi della consistenza complessiva di quella quota di elettorato che, dopo aver aderito a una delle due coalizioni nel proporzionale, nel maggioritario invece si “disperde” sia verso le terze forze che verso il non voto.

Prendiamo anzitutto in esame la Cdl. Un primo punto da sottolineare è che la distribuzione territoriale del fenomeno presenta una configurazione in parte diversa rispetto a quanto abbiamo visto a proposito del conflitto vero e proprio. In quel caso infatti si trattava di un fenomeno abbastanza diffuso a livello nazionale, e complessivamente uniforme. Nel caso della dispersione il *pattern* sembra invece notevolmente diverso. Il fenomeno appare molto più frastagliato, ma soprattutto particolarmente concentrato in senso territoriale. La dispersione dell'elettorato proporzionale della Cdl è praticamente assente in molti collegi; in altri invece raggiunge valori anche molto alti; il tutto, tuttavia, in assenza di una regolarità geografica particolarmente intelligibile, se si eccettua un accentuarsi del fenomeno nel Nord-Est.

Figura 7.9 – Dispersione subita dalla Cdl (% su elettori)



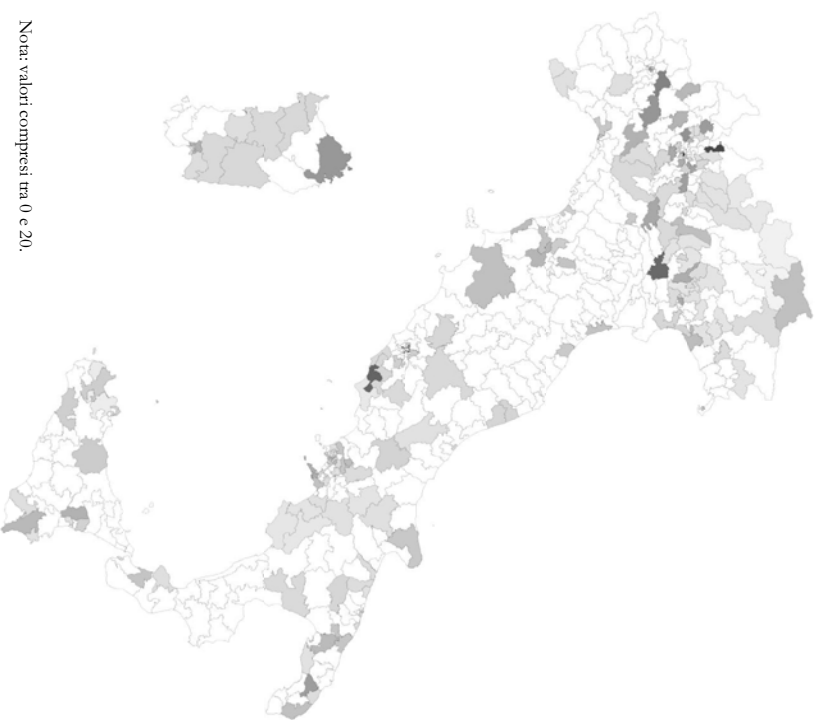
Nota: valori compresi tra 0 e 30.

In questo senso l'ipotesi esplicativa più probabile è che la dispersione abbia costituito, per l'elettore proporzionale del centro-destra, un tipo di defezione maggiormente associato a fattori locali, ovvero alla presenza di candidati di collegio sgraditi.

Per quanto riguarda il centrosinistra, il quadro sembra invece diverso. Anzitutto il fenomeno è, nel complesso, significativamente meno rilevante.



Figura 7.10 – Dispersione subita dall'Ulivo (% su elettori)



Nota: valori compresi tra 0 e 20.

In termini di distribuzione territoriale, sembrano però emergere comunque alcune regolarità geografiche. Escludendo il Sud che presenta un profilo più frastagliato, il resto del paese sembra invece maggiormente leggibile. Il Nord si conferma punto debole per la performance maggioritaria dell'Ulivo. Già nell'analisi del conflitto avevamo visto come questo fenomeno affliggesse l'Ulivo in misura maggiore nell'Italia settentrionale. Lo stesso accade per la dispersione: in questa area del paese il centrosinistra ha difficoltà a

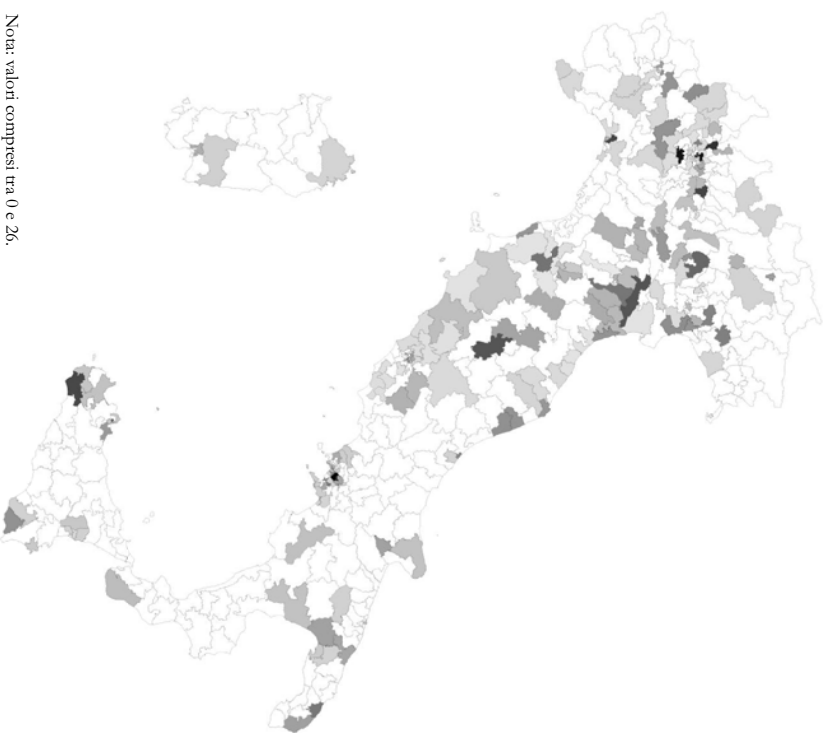
portare nel maggioritario tutto il proprio elettorato proporzionale (anche se magari è comunque in grado di attrarre una parte dell'elettorato proporzionale della Cdl). Per quanto riguarda infine la Zona rossa, esce ancora una volta confermata l'interpretazione che vede in quest'area una notevole coesione dell'elettorato di centrosinistra: è qui (come avevamo osservato per il conflitto) che anche la dispersione è praticamente assente.

#### 7.2.4. Attrazione

Passiamo a questo punto alla presentazione dell'ultima dinamica di comportamento differenziato rilevante per il differenziale di coalizione. Si tratta dell'*attrazione*, ovvero del comportamento di chi nel proporzionale si astiene o vota un terzo polo, ma nel maggioritario sceglie una delle due coalizioni maggiori. Questa volta il fenomeno (che riguarda a livello nazionale il 4% degli elettori) premia nella stessa misura le due coalizioni. La distribuzione territoriale della consistenza di quest'area per le due coalizioni è presentata rispettivamente nella Figura 7.11 e nella Figura 7.12.

Le due mappe sono in parte simili dal punto di vista della distribuzione geografica del fenomeno, ma non paiono mostrare una qualche logica territoriale chiara a livello nazionale. In questo caso ulteriori analisi potrebbero forse concentrare maggiormente l'attenzione sulla presenza di dinamiche a livello regionale (ad esempio nel caso della Campania) per poter distinguere con maggiore chiarezza eventuali tendenze significative. Inoltre un inevitabile motivo di difficoltà di lettura è costituito dal fatto che nell'"area residuale" sono inclusi fenomeni politicamente eterogenei. Tra questi il voto alle terze forze (alcune delle quali, prima delle elezioni, erano anche ritenute potenzialmente competitive nel maggioritario), il voto ai candidati minori, le schede bianche e nulle e le astensioni. Non ci si può quindi meravigliare che i risultati possano non mostrare dinamiche territoriali particolarmente nette.

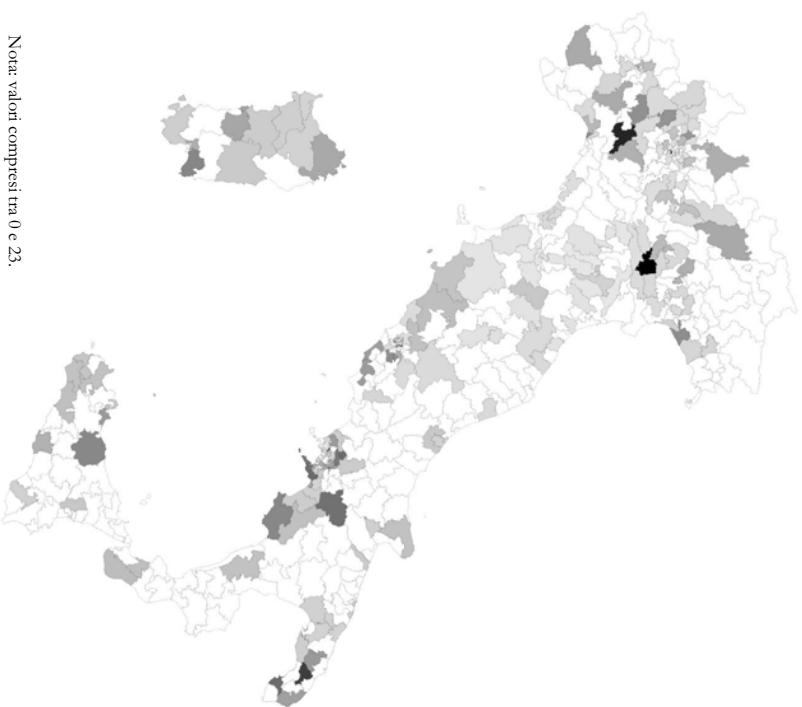
Figura 7.11 – Attrazione esercitata dalla Cdl (% su elettori)



Note: valori compresi tra 0 e 26.

In questo senso l'approccio forse più utile potrebbe essere quello di effettuare ulteriori analisi, disaggregando le terze forze dall'area del non voto: è probabile che un'analisi a questo livello di dettaglio metterebbe in evidenza regolarità territoriali più leggibili.

Figura 7.12 – Attrazione esercitata dall'Ulivo (% su elettori)



Note: valori compresi tra 0 e 23.

### 7.3. La struttura territoriale del voto differenziato

Abbiamo aperto questo capitolo ritenendo che, una volta presentata la consistenza nazionale dei vari comportamenti di voto differenziato, potesse risultare utile utilizzare le stesse informazioni disaggregate nella loro distribuzione territoriale. Si trattava infatti anche solo dell'opportunità di rispondere alla domanda di base relativa all'uniformità dei fenomeni. Ad esempio avremmo potuto

attenderci che il differenziale negativo della CdL non fosse tale sull'intero territorio nazionale, e che il saldo complessivo negativo risultasse da diverse dinamiche territoriali. In questo senso la risposta che ci forniscono le distribuzioni territoriali esaminate è per lo meno duplice.

Da un lato, infatti, le prime mappe esaminate, quelle relative al risultato complessivo del differenziale di coalizione, danno una risposta molto netta: il differenziale di coalizione nel suo complesso è territorialmente uniforme. In tutta Italia la CdL prende meno voti nel maggioritario rispetto al proporzionale; in tutta Italia l'Ulivo presenta invece un differenziale positivo. I collegi che fanno eccezione sono pochissimi.

La seconda parte della risposta deriva invece dall'esame delle stime vere e proprie prodotte dalla procedura di stima Eidis per tutti i collegi uninominali. E qui abbiamo rivisto la caratteristica fondamentale dell'*eterogeneità* dei fenomeni di voto differenziato che sono alla radice del differenziale di coalizione osservato. I vari tipi di comportamento differenziato, esaminati oltretutto separatamente per le due coalizioni, presentano infatti caratteristiche diverse. In alcuni casi sono presenti chiare regolarità territoriali; in altre sembra evidente la presenza di fattori locali, legati a particolari contesti o, presumibilmente, a specifiche candidature; in altri ancora, la costruzione di categorie di classificazione troppo ampie (combinando fenomeni diversi tra loro) produce una distribuzione territoriale apparentemente priva di regolarità intelligibili, impossibile da decifrare.

Tuttavia alcune conclusioni sostantive sembrano emergere in modo netto attraverso tutti i fenomeni osservati. La principale di esse è la fondamentale asimmetria tra le due coalizioni in termini di presenza di un particolare radicamento territoriale. La regolarità geografica che infatti emerge con maggior forza è quella della Zona rossa, relativamente all'Ulivo. La presenza di una specificità territoriale in quel caso è evidenziata da tutte le dinamiche esaminate: differenziale di coalizione "paradossalmente" basso; tasso di coerenza molto alto; tassi molto bassi di dispersione e di conflitto. Viceversa la Casa delle Libertà non sembra presentare regolarità ge-

ografiche particolarmente marcate: in questo senso è inevitabile immaginare che questo fenomeno rappresenti l'ennesima conferma delle diverse basi di radicamento sociale e territoriale delle due coalizioni.

In conseguenza di queste interpretazioni aggiuntive estratte dai dati, la conclusione di questo capitolo non può essere che implicitamente metodologica, in difesa dell'utilità delle stime in base a dati aggregati. Nel capitolo precedente abbiamo confrontato i risultati nazionali della stima Eidis anche con dati di *survey*. In quel caso vediamo come il prodotto finale di un'indagine campionaria non può che essere un semplice dato nazionale. È così possibile stimare la consistenza dei vari comportamenti di voto differenziato, ma l'analisi si ferma; e le specificità territoriali (di particolare importanza in un paese come l'Italia) non possono essere indagate in alcun modo. Viceversa l'analisi su dati aggregati (su larga scala, ovvero includendo l'intero territorio nazionale mediante numerose analisi locali) permette di accedere a una messe di informazioni di ricchezza di gran lunga maggiore; in questo modo la stessa consistenza nazionale di un determinato fenomeno può essere interpretata in modo decisamente più comprensibile.

## 8. Conclusioni

Questo volume si è aperto con un problema. Negli ultimi decenni infatti le tecniche tradizionali di inferenza ecologica (e in particolare il modello di Goodman) sono state infatti oggetto di numerose critiche, in particolare riguardo all'irrealisticità di alcuni degli assunti, destinata ad avere come conseguenza la produzione di stime non corrette o inaffidabili.

Il tentativo di questo volume era dunque quello di portare un contributo in questa direzione, prendendo in esame l'approccio EI, e soprattutto cercando di andare oltre le limitazioni che lo affliggono, rendendolo adatto anche ad analizzare i problemi tipici di un sistema multipartitico.

Da questo punto di vista si può dire che il traguardo sia stato complessivamente raggiunto. L'approccio EI può essere esteso con successo a tabelle di dimensioni maggiori di  $2 \times 2$ , aprendo quindi all'inferenza ecologica la possibilità di analizzare intere nuove classi di problemi, tra cui ovviamente quelli relativi all'analisi dei flussi elettorali, sia diacronici che sincronici. Non solo: la conferma dell'estensione teorica dell'approccio è stata fornita dalla effettiva realizzazione di una procedura informatica di stima, in grado di essere utilizzata con successo per rispondere un concreto quesito di ricerca. Infine, punto più importante, i risultati forniti dalla procedura sono assolutamente plausibili. Senza la necessità di algoritmi di correzione o di ulteriori ponderazioni (necessarie ad esempio per i dati di *smry*) la procedura Eidis ha prodotto stime estremamente simili a quelle del modello di Goodman

e a quelle dedotte da dati di sondaggio: il tutto incorporando una quantità molto maggiore di informazioni e senza produrre stime inaccettabili. E la plausibilità dei risultati sostantivi ottenuti sembra d'altronde confermata dalla sensatezza delle distribuzioni geografiche osservate.

Tuttavia il catalogo dei problemi è ampio. La procedura di stima è complessa, e quindi necessariamente impiega notevoli risorse di calcolo e richiede tempi lunghi<sup>1</sup>. E di conseguenza questo si riflette sui problemi che è possibile analizzare. Un'analisi seria dei flussi elettorali dovrebbe prevedere perlomeno matrici  $4 \times 4$ , e certamente anche di dimensioni sensibilmente maggiori. Si tratta tuttavia di una crescita dimensionale non semplice da gestire, in termini di capacità di calcolo.

Da questo punto di vista, occorre però osservare che ci sono ancora notevoli spazi per una semplificazione e ottimizzazione della procedura di stima. Inoltre, ulteriori margini di crescita sono offerte, oltre che dalla continua disponibilità di processori di velocità crescente, delle possibilità offerte dal calcolo distribuito, anche senza fare ricorso a costose infrastrutture. Il risultato è che in futuro sarà verosimilmente possibile stimare senza difficoltà matrici di flusso di dimensioni maggiori, e con precisione sempre più accurata.

In termini metodologici più generali, è abbastanza evidente che l'approccio EI, ovvero l'applicazione del metodo dei vincoli a livello di unità ecologiche, rappresenta per l'inferenza ecologica un fondamento decisamente più solido rispetto alle ipotesi del modello di Goodman. I vincoli intrinseci nei dati ecologici disponibili costituiscono infatti una grande quantità di informazione *deterministica*. Questa informazione rimaneva completamente inutilizzata nelle stime basate sul modello di Goodman; viceversa l'approccio EI ne fa un uso efficiente. Di qui la maggiore solidità (e la automatica accettabilità) delle stime prodotte mediante questo approccio.

<sup>1</sup> Non lunghissimi in realtà: le elaborazioni relative ai 466 collegi elettorali (con precisione 0,1) sono state completate in circa 12 ore di elaborazione, su un normale PC con processore Pentium IV a 2,4 GhZ.

Non solo: una grande differenza tra l'approccio EI e il modello di Goodman è costituita dal fatto che l'analisi della tomografia permette, in via preliminare, di determinare con maggiore affidabilità *se* i dati sono analizzabili con la tecnica EI; in questo senso siamo di fronte a una concreta possibilità di valutazione dell'applicabilità degli assunti del modello. Si tratta di una questione cruciale, visto che l'assenza di questa caratteristica è stata alla base della maggior parte delle critiche rivolte al modello di Goodman.

Di conseguenza l'introduzione di questo approccio all'inferenza ecologica, e la sua generalizzazione oltre il confronto di due dicotomie, possono essere visti come elementi utili in generale per l'evoluzione dello studio dei flussi elettorali. Elementi che si affiancano ad altre evoluzioni importanti: da un lato la più accessibile disponibilità di risultati elettorali disaggregati a livello di sezione su tutto il territorio nazionale; dall'altro la maggior facilità di accesso alle capacità di calcolo necessarie ad utilizzare questo tipo di procedure. Insomma, potrebbero esserci le condizioni per una ripresa, metodologicamente consapevole, dello sviluppo di analisi sistematiche della mobilità elettorale.

La necessità conseguenza di questo processo è il recupero della grande capacità informativa delle analisi basate su dati aggregati. Ovviamente non si tratta di una contrapposizione tra diversi approcci di analisi: le indagini campionarie sono uno strumento centrale nello studio degli atteggiamenti politici e dei comportamenti di voto; ambito che oggi sarebbe semplicemente impossibile da immaginare senza lo strumento del *survey*.

Tuttavia è inevitabile constatare che, soprattutto in una realtà territorialmente complessa come quella italiana, le conclusioni necessariamente generali delle indagini campionarie spesso non sono in grado di cogliere la complessità e l'eterogeneità territoriale dei fenomeni in esame. In più, non va dimenticato che le analisi generali a livello nazionale hanno sempre la conseguenza di sopprimere una parte rilevante di varianza: "baciini elettorali" apparentemente immobili a livello nazionale spesso nascondono profondi cambiamenti a livello locale.

## Riferimenti bibliografici

In questo senso quindi si può dire che lo strumento dell'analisi su dati ecologici è destinato a promuovere anche il suo stesso oggetto di studio: soltanto dotandosi di strumenti in grado di rilevare gli *elettori in movimento* è possibile mostrarne la consistenza numerica e indagarne i comportamenti. Da questo punto di vista, non ci si può meravigliare che le prime analisi sistematiche sulla mobilità elettorale in Italia abbiano creato un certo rumore, rivelando la presenza di una brulicante mobilità, dove ci si aspettava stabilità e continuità.

In questo senso la disponibilità di questi strumenti mette a disposizione un'ulteriore possibilità nell'analisi della ormai quindi-cennale transizione italiana. È possibile che, ancora una volta, nuovi strumenti rivelino – dove ci si aspettava un elettorato già stabilizzato in nuove appartenenze partitiche – la sorprendente presenza di elettori in movimento.

- Achen, C., e W. P. Shively. 1995. *Cross-Level Inference*, Chicago, University of Chicago Press.
- Achen, C. H. 1975. Mass Political Attitudes and the Survey Response. *American Political Science Review* 69 (4), pp. 1218-1231.
- Agosta, A. 2006. Database. Come leggere la partecipazione elettorale. *Il Mulino* LV (3), pp. 465-468.
- Allsop, D., e H. F. Weisberg. 1988. Measuring Change in Party Identification in an Election Campaign. *American Journal of Political Science* 32 (4), pp. 996-1017.
- Anastasi, A., G. Gangemi, R. Pavisic, e V. Tomaselli. 1989. *Guerra dei flussi o bolle di sapone? Ricerca empirica e riflessioni sul modello di Goodman per la stima dei flussi elettorali*, Acireale, Bonanno.
- . 1993. Guerre dei flussi o bolle di sapone? In *Quale mobilità elettorale? Tendenze e modelli. La discussione metodologica sui flussi elettorali*, a cura di R. Mannheim. Milano, Franco Angeli, pp. 41-58.
- Baldassari, D. 2005. *La semplice arte di votare: le scorciatoie cognitive degli elettori italiani*, Bologna, Il Mulino.
- Baldassari, D., e H. Schadee. 2004. Il fascino della coalizione. Come e perché le alleanze elettorali influenzano il modo in cui gli elettori interpretano la politica. *Rivista Italiana di Scienza Politica* (2), pp. 249-276.

- . 2007. Sinistra e destra: la dimensione ideologica tra Prima e Seconda repubblica. In *Gli italiani e la politica*, a cura di M. Maraffi. Bologna, Il Mulino, pp. 105-130.
- Barbagli, M., e P. Corbetta. 1980. L'elettorato, l'organizzazione del Pei e i movimenti. *Il Mulino* XXIX (3), pp. 467-490.
- Barbagli, M., P. G. Corbetta, A. Parisi, e H. M. A. Schadee. 1979. *Fluidità elettorale e classi sociali in Italia*. Bologna, Il Mulino.
- Barisone, M. 2001. Elettori indecisi, elettori fluttuanti: che volto hanno i "bilancieri" del voto? i casi italiano e francese. *Rivista Italiana di Scienza Politica* XXXI (1), pp. 73-108.
- . 2002. Interesse per la politica, appartenenza di coalizione e giudizio sui leader: gli effetti della campagna elettorale. In *Dall'Ulivo al governo Berlusconi*, a cura di G. Pasquino. Bologna, Il Mulino, pp. 139-179.
- Bartolini, S. 2000. *The political mobilization of the European left, 1860-1980 : the class cleavage*. Cambridge, Cambridge University Press.
- Bartolini, S., A. Chiaramonte, e R. D'Alimonte. 2002. Maggioritario finalmente? Il bilancio di tre prove. In *Maggioritario finalmente? La transizione elettorale 1994-2001*, a cura di R. D'Alimonte e S. Bartolini. Bologna, Il Mulino, pp. 363-379.
- Bartolini, S., e R. D'Alimonte. 2002. La maggioranza ritrovata. La competizione nei collegi uninominali. In *Maggioritario finalmente? La transizione elettorale 1994-2001*, a cura di R. D'Alimonte e S. Bartolini. Bologna, Il Mulino, pp. 199-248.
- Bartolini, S., e P. Mair. 1990. *Identity, competition, and electoral availability : the stabilisation of European electorates 1885-1985*. Cambridge, Cambridge University Press.
- Bellucci, P. 2006. All'origine della polarità del governo in Italia, 1994-2006. *Rivista Italiana di Scienza Politica* XXXVI (3), pp. 479-503.
- Berelson, B. R., P. F. Lazarsfeld, e W. N. McPhee. 1954. *Voting: a study of opinion formation in a presidential campaign*. Chicago, University of Chicago Press.
- Biorcio, R. 1993. Lo studio dei flussi elettorali nella ricerca politica: possibilità e limiti di una metodologia. In *Quale mobilità elettorale? Tendenze e modelli. La discussione metodologica sui flussi elettorali*, a cura di R. Mannheim. Milano, Franco Angeli, pp. 79-106.
- Biorcio, R., e P. Narale. 1987. Mobilità e fedeltà elettorale degli anni ottanta. Un'analisi comparata su dati aggregati e di survey. *Quaderni dell'Osservatorio Elettorale* 18, pp. 43-88.
- . 1989. La mobilità elettorale degli anni Ottanta. *Rivista Italiana di Scienza Politica* XIX (3), pp. 385-430.
- Black, D. 1948. On the Rationale of Group Decision-making. *The Journal of Political Economy* 56 (1), pp. 23-34.
- . 1958. *The theory of committees and elections*. Cambridge, University Press.
- Blalock, H. M. 1964. *Causal Inferences in Non-Experimental Research*. Chapel Hill, University of North Carolina Press.
- Born, R. 1994. Split-Ticket Voters, Divided Government, and Fiorina's Policy-Balancing Model. *Legislative Studies Quarterly* 19 (1), pp. 95-115.
- Brown, P. J., e C. D. Payne. 1986. Aggregate Data, Ecological Regression, and Voting Transitions. *Journal of the American Statistical Association* 81 (394), pp. 452-460.
- Budge, I., I. Crewe, e D. Farlie. 1976. *Party identification and beyond : representations of voting and party competition*. London, Wiley.
- Budge, I., e D. Farlie. 1977. *Voting and Party Competition*. London, Wiley.
- Caciagli, M., e P. G. Corbetta, (a cura di) 2002. *Le ragioni dell'elettore: perché ha vinto il centro-destra nelle elezioni italiane del 2001*. Bologna, Il Mulino.
- Campbell, A., P. E. Converse, W. E. Miller, e D. Stokes. 1960. *The American voter*. New York, Wiley.
- . 1966. *Elections and the political order*. New York, Wiley.
- Campus, D. 2000. *L'elettore pigro. Informazione politica e scelte di voto*. Bologna, Il Mulino.
- Caramani, D. 2004. *The Nationalization of Politics*. Cambridge, Cambridge University Press.

- Cartocci, R. 1987. Otto risposte a un problema: la divisione dell'Italia in zone politicamente omogenee. *Polis* I (3), pp. 481-514.
- . 1990. *Elettori in Italia. Riflessioni sulle vicende elettorali degli anni Ottanta*, Bologna, Il Mulino.
- Catellani, P., e P. Milesi. 2006. Identificazioni: dal partito alla coalizione. In *Sinistra e destra. Le radici psicologiche della differenza politica*, a cura di P. Catellani e P. Corbetta. Bologna, Il Mulino, pp. 73-89.
- Chiaromonte, A. 2002. Il voto proporzionale: verso la nazionalizzazione della competizione? In *Maggioritario finalmente? La transizione elettorale 1994-2001*, a cura di R. D'Alimonte e S. Bartolini. Bologna, Il Mulino, pp. 165-198.
- . 2005. *Tra maggioritario e proporzionale. L'universo dei sistemi elettorali misti*, Bologna, Il Mulino.
- . 2007. Il nuovo sistema partitico italiano tra bipolarismo e frammentazione. In *Proporzionale ma non solo. Le elezioni politiche del 2006*, a cura di R. D'Alimonte e A. Chiaromonte. Bologna, Il Mulino, pp. 369-407.
- Chiaromonte, A., e R. D'Alimonte, (a cura di) 2000. *Il maggioritario regionale. Le elezioni del 16 aprile 2000*. Bologna: Il Mulino.
- Clarke, H. D. 2004. *Political choice in Britain*, Oxford ; New York, Oxford University Press.
- Comero, D. 1993. Il modello dei flussi elettorali nell'analisi del voto: ragioni e limiti di un approccio convenzionalista. In *Quale mobilità elettorale? Tendenze e modelli. La discussione metodologica sui flussi elettorali*, a cura di R. Mannheim. Milano, Franco Angeli, pp. 229-272.
- Converse, P. E. 1964. The nature of belief systems in mass publics. In *Ideology and Discontent*, a cura di D. Apter. New York, Free Press, pp. 206-261.
- . 1969. Of Time and Partisan Stability. *Comparative Political Studies* 2, pp. 139-171.
- Corbetta, P., e A. Parisi. 1993. Sull'applicabilità dei metodi di stima dei flussi elettorali a livello nazionale. In *Quale mobilità elettorale? Tendenze e modelli. La discussione metodologica sui flussi elettorali*, a cura di R. Mannheim. Milano, Franco Angeli, pp. 59-77.
- Corbetta, P., A. Parisi, e H. Schadee. 1988. *Elezioni in Italia: struttura e tipologia delle consultazioni politiche*, Bologna, Il Mulino.
- Corbetta, P., e H. M. A. Schadee. 1993. L'analisi delle tavole di movimento elettorale tramite la modellistica log-lineare. *Sociologia e ricerca sociale* (40), pp. 77-100.
- Corbetta, P. G. 2002. Le generazioni politiche. In *Le ragioni dell'elettore*, a cura di M. Caciagli e P. G. Corbetta. Bologna, Il Mulino, pp. 79-112.
- Corbetta, P. G., e H. M. A. Schadee. 1984. *Metodi e modelli di analisi dei dati elettorali*, Bologna, Il Mulino.
- Corielli, F., e M. Weber. 1993. A proposito di flussi elettorali: qualche critica e una proposta. In *Quale mobilità elettorale? Tendenze e modelli. La discussione metodologica sui flussi elettorali*, a cura di R. Mannheim. Milano, Franco Angeli, pp. 197-228.
- Cox, G. W. 1997. *Making votes count : strategic coordination in the world's electoral systems*, Cambridge, Cambridge University Press; tr. it. *I voti che contano*, Bologna, Il Mulino, 2005.
- D'Alimonte, R. 2007. Il nuovo sistema elettorale. Dal collegio uninominale al premio di maggioranza. In *Proporzionale ma non solo. Le elezioni politiche del 2006*, a cura di R. D'Alimonte e A. Chiaromonte. Bologna, Il Mulino, pp. 51-88.
- D'Alimonte, R., e S. Bartolini. 1995a. Il sistema partitico italiano: una transizione difficile. In *Maggioritario ma non troppo*, a cura di S. Bartolini e R. D'Alimonte. Bologna, Il Mulino, pp. 429-466.
- . 1995b. La competizione maggioritaria: le origini elettorali del parlamento diviso. In *Maggioritario ma non troppo*, a cura di S. Bartolini e R. D'Alimonte. Bologna, Il Mulino, pp. 317-372.
- . 1997. Come perdere una maggioranza: la competizione nei collegi uninominali. In *Maggioritario per caso*, a cura di R. D'Alimonte e S. Bartolini. Bologna, Il Mulino, pp. 237-283.



- , (a cura di) 2002. *Maggioritario finalmente? La transizione elettorale 1994-2001*. Bologna: Il Mulino.
- D'Alimonte, R., e A. Chiaramonte. 1995. Il nuovo sistema elettorale italiano: le opportunità e le scelte. In *Maggioritario ma non troppo*, a cura di S. Bartolini e R. D'Alimonte. Bologna, Il Mulino, pp. 37-81.
- . 2006. Proporzionale ma non solo. La riforma elettorale della Casa delle libertà. *Il Mulino* (1), pp. 34-45.
- Dahl, R. A. 1971. *Pluralism: participation and opposition*, New Haven, Yale University Press.
- Dalton, R. 1984. Cognitive Mobilization and Partisan Dealignment in Advanced Industrial Democracies. *Journal of Politics* 46, pp. 264-284.
- Dalton, R., e M. P. Wattenberg. 2000. *Parties Without Partisans: Political Change in Advanced Industrial Democracies*, Oxford, Oxford University Press.
- De Sio, L. 2003. A Proposal for Extending King's El Method to m x n Tables, mimeo.
- . 2006a. C'era una volta la doppia scheda. Voto differenziato e rendimento di coalizione nelle elezioni politiche del 2001. *Polis* 20 (3), pp. 373-404.
- . 2006b. Dove stanno davvero gli elettori fluttuanti? *Rivista Italiana di Scienza Politica* 36 (3), pp. 393-414.
- . 2006c. Elettori «convertiti», elettori «traghettati». In *Dov'è la vittoria? Il voto del 2006 raccontato dagli italiani*, a cura di I-TANES. Bologna, Il Mulino, pp. 61-76.
- . 2007a. Al di là delle apparenze. Il risultato delle elezioni. In *Proporzionale ma non solo. Le elezioni politiche del 2006*, a cura di R. D'Alimonte e A. Chiaramonte. Bologna, Il Mulino, pp. 243-304.
- . 2007b. Il rapporto tra gli italiani e i partiti: declino o transizione? In *Gli italiani e la politica*, a cura di M. Maraffi. Bologna, Il Mulino, pp. 131-156.
- . 2007c. Movimento elettorale e voto diviso nelle elezioni politiche del 2006 in Toscana. *Quaderni dell'Osservatorio Elettorale* (58), pp. 7-40.
- . 2008. Are Less Involved Voters the Key to Win Elections? *Comparative Political Studies* 41 (2), pp. 217-241.
- Delli Carpini, M. X., e S. Keeter. 1996. *What Americans know about politics and why it matters*, New Haven, Yale University Press.
- Di Virgilio, A. 1995. Dai partiti ai poli: la politica delle alleanze. In *Maggioritario ma non troppo*, a cura di S. Bartolini e R. D'Alimonte. Bologna, Il Mulino, pp. 177-232.
- Diamanti, I. 2003. *Bianco, rosso, verde... e azzurro. Mappe e colori dell'Italia politica*, Bologna, Il Mulino.
- Diamanti, I., e R. Mannheim. 2002. Le basi sociali del voto: la frattura che attraversa i ceti medi. In *Le regioni dell'eletto: perchè ha vinto il centro-destra nelle elezioni italiane del 2001*, a cura di M. Caciagli e P. G. Corbetta. Bologna, Il Mulino, pp. 139-163.
- Downs, A. 1957. *An economic theory of democracy*, New York, Harper; tr. it. *Teoria economica della democrazia*, Bologna, Il Mulino, 1990.
- Draghi, S. 1985. Vero e falso nei flussi elettorali. *Rinascita*, 29 giugno 1985, pp. 12-14.
- Duncan, O. D., e B. Davis. 1953. An Alternative to Ecological Correlation. *American Sociological Review* (18), pp. 665-666.
- Duverger, M. 1950. *Les partis politiques*, Paris, Colin; tr.it. *I partiti politici*, Milano, Comunità, 1970.
- Fabrizio, D., P. Feltrin, e L. Marcone. 2006. La decisività del voto nelle regioni meridionali: un'analisi degli orientamenti elettorali negli ultimi dieci anni. *Palena* (2), pp. 9-35.
- Farquharson, R. 1969. *Theory of voting*, New Haven, Yale University Press.
- Felsenthal, D. S. 1990. *Topics in social choice: sophisticated voting, efficacy, and proportional representation*, New York, Praeger.
- Freedman, D. 1999. From Association to Causation: Some Remarks on the History of Statistics. *Statistical Science* 14 (3), pp. 243-258.

- Freedman, D. A., S. P. Klein, J. Sacks, C. A. Smyth, e C. G. Everett. 1991. Ecological Regression and Voting Rights. *Evaluation Review* 15 (6), pp. 673.
- Fuchs, D., e H.-D. Klingemann. 1989. The Left-Right Schema. In *Continuities in Political Action*, a cura di M. Kent Jennings e J. W. Van Deth. Berlin, de Gruyter, pp. 203-234.
- Galli, G. 1966. *Il bipartitismo imperfetto; comunisti e democristiani in Italia*. Bologna, Società editrice Il Mulino.
- . 1993. I flussi e i saldi. In *Quale mobilità elettorale? Tendenze e modelli. La discussione metodologica sui flussi elettorali*, a cura di R. Mannheim. Milano, Franco Angeli, pp. 33-39.
- Galli, G., V. Capecci, V. Cioni Polacchini, e G. Sivini. 1968. *Il comportamento elettorale in Italia: Una indagine ecologica sulle elezioni in Italia fra il 1946 e il 1963*. Bologna, Il Mulino.
- Goldthorpe, J. H. 2000. *On sociology : numbers, narratives, and the integration of research and theory*. Oxford, Oxford University Press.
- Goodman, L. A. 1953. Ecological Regressions and Behavior of Individuals. *American Sociological Review* 18, pp. 663-664.
- . 1959. Some Alternatives to Ecological Correlation. *American Journal of Sociology* 64, pp. 610-625.
- Grilli di Cortona, P. 2007. *Il cambiamento politico in Italia. Dalla Prima alla Seconda Repubblica*. Roma, Carocci.
- Grofman, B., e S. Merrill. 2002. Ecological Regression and Ecological Inference. Paper presentato a Ecological Inference Conference, 17-18 giugno 2002, Harvard University, Cambridge, MA.
- . 2004. Ecological Regression and Ecological Inference. In *Ecological Inference: New Methodological Strategies*, a cura di G. King, O. Rosen e M. Tanner. Cambridge, Cambridge University Press, pp. 123-143.
- Gunther, R., e J. R. Montero. 2001. The anchors of partisanship: a comparative analysis of voting behavior in four Southern European democracies. In *Parties, politics, and democracy in the new Southern Europe*, a cura di N. P. Diamandouras e R. Gunther. Baltimore, Johns Hopkins University Press.
- Hawkes, A. G. 1969. An Approach to the Analysis of Electoral Swing. *Journal of the Royal Statistical Society. Series A (General)* 132 (1), pp. 68-79.
- Hedström, P. 2004. Generative Models and Explanatory Research: On the Sociology of Aage B. Sørensen. In *Research in Social Stratification and Mobility - Volume 21 - Inequality: Structures, Dynamics, and Mechanisms: Essays in Honor of Aage B. Sørensen*, a cura di A. L. Kalleberg, S. L. Morgan, J. Myles e R. A. Rosenfeld.
- Hedström, P., e R. Swedberg, (a cura di) 1998. *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.
- Hotelling, H. 1929. Stability in Competition. *The Economic Journal* 39 (153), pp. 41-57.
- Inglehart, R. 1990. *Culture shift in advanced industrial society*. Princeton, Princeton University Press; tr. it. *Valori e cultura politica*, Padova, Liviana-Petini, 1993.
- Ianes. 2001. *Perché ha vinto il centro-destra*. Bologna, Il Mulino.
- Ianes, (a cura di) 2006a. *Dov'è la vittoria? Il voto del 2006 raccontato dagli italiani*. Bologna: Il Mulino.
- , (a cura di) 2006b. *Sinistra e destra. Le radici psicologiche della differenza politica*. Bologna: Il Mulino.
- King, G. 1997. *A solution to the ecological inference problem: reconstructing individual behavior from aggregate data*. Princeton, NJ, Princeton University Press.
- King, G., O. Rosen, e M. A. Tanner. 2004. *Ecological Inference: New Methodological Strategies*. Cambridge, Cambridge University Press.
- Kuhn, T. 1962. *The structure of scientific revolutions*. Chicago, University of Chicago Press.
- Lazarfeld, P. F. 1955. Interpretation of statistical relations as a research operation. In *The Language of Social Research*, a cura di P. F. Lazarfeld e M. Rosenberg. New York, Free Press.

- Lazarsfeld, P. F., B. Berelson, e H. Gaudet. 1944. *The People's Choice: How the Voter Makes up His Mind in a Presidential Campaign*, New York, Columbia University Press.
- Legnante, G., e P. Segatti. 2001. L'astensionista intermittente, ovvero quando decidere di votare o meno è lieve come una piuma. *Polis* (2), pp. 181-202.
- Lipset, S. M., e S. Rokkan. 1967. *Party systems and voter alignments: cross-national perspectives*, New York, Free Press.
- Mannheimer, R., (a cura di) 1993. *Quale mobilità elettorale? Tendenze e modelli. La discussione metodologica sui flussi elettorali*. Milano: Franco Angeli.
- Mannheimer, R., e G. Micheli. 1976. Il comportamento elettorale a Milano. *Rassegna Italiana di Sociologia* XVII, pp. 619-639.
- Mannheimer, R., e G. Sani. 1987. *Il mercato elettorale. Identikit dell'elettore italiano*, Bologna, Il Mulino.
- . 2001. *La conquistista degli astenuti*, Bologna, Il Mulino.
- Maraffi, M. 2002. Per che cosa si è votato il 13 maggio? Le mappe cognitive degli elettori italiani. In *Le ragioni dell'elettore*, a cura di M. Caciagli e P. G. Corbetta. Bologna, Il Mulino, pp. 301-338.
- . 2006. Nella selva della politica: partiti, coalizioni e altri animali. In *Dov'è la vittoria? Il voto del 2006 raccontato dagli italiani*, a cura di ITANES. Bologna, Il Mulino, pp. 197-208.
- Marradi, A. 1994. Referenti, pensiero e linguaggio: una questione rilevante per gli indicatori. *Sociologia e ricerca sociale* (43), pp. 137-191.
- . 1995. Appunti di analisi bivariata, mimeo.
- Mattei, F., e J. S. Howes. 2000. Competing Explanations of Split-Ticker Voting in American National Elections. *American Politics Research* 28 (3), pp. 379-407.
- Micheli, G. 1976. Il comportamento individuale nell'analisi sociologica del dato aggregato. *Il giornale degli economisti ed annali di economia* XXV, pp. 429-448.
- . 1993. Disaggregare il comportamento elettorale: modelli di cosa, modelli per che cosa. In *Quale mobilità elettorale? Tendenze e modelli. La discussione metodologica sui flussi elettorali*, a cura di R. Mannheimer. Milano, Franco Angeli, pp. 17-32.
- Miller, W. E., e J. M. Shanks. 1996. *The new American voter*, Cambridge, Mass., Harvard University Press.
- Morlino, L. 1991. *Costruire la democrazia: gruppi e partiti in Italia*, Bologna, Il Mulino.
- Natale, P. 1987. Dati attendibili, ecco come. *L'Unità*, 21/6/1987, pp. 6.
- . 1993. Lo studio empirico dei flussi elettorali: problemi metodologici e applicativi. In *Quale mobilità elettorale? Tendenze e modelli. La discussione metodologica sui flussi elettorali*, a cura di R. Mannheimer. Milano, Franco Angeli, pp. 107-135.
- . 1997. Mutamento e stabilità nel voto degli italiani. In *Maggioritario per caso. Le elezioni politiche del 1996*, a cura di R. D'Alimonte e S. Bartolini. Bologna, Il Mulino, pp. 207-235.
- . 2000. Il comportamento elettorale: la "fedeltà leggera". In *Il maggioritario regionale. Le elezioni del 16 aprile 2000*, a cura di A. Chiaromonte e R. D'Alimonte. Bologna, Il Mulino, pp. 157-174.
- . 2002. Una fedeltà leggera: i movimenti di voto nella "seconda repubblica". In *Maggioritario finalmente? La transizione elettorale 1994-2001*, a cura di R. D'Alimonte e S. Bartolini. Bologna, Il Mulino, pp. 283-317.
- Neumann, S. 1956. *Modern political parties*, Chicago, University of Chicago Press.
- Nie, N. H., S. Verba, e J. R. Petrocik. 1976. *The changing American voter*, Cambridge, Harvard University Press.
- Ogburn, W., e I. Goltra. 1919. How Women Vote: A Study of an Election in Portland, Oregon. *Political Science Quarterly* (34), pp. 413-433.
- Pappalardo, A. 1996. Dal pluralismo polarizzato al pluralismo moderato. Il modello di Sartori e la transizione italiana. *Rivista Italiana di Scienza Politica* XXVI (1), pp. 103-145.
- . 2000. Il sistema partitico italiano tra crisi e trasformazione. In *Condannata al successo? L'Italia nell'Europa integrata*, a cura di G. Di Palma. Bologna, Il Mulino, pp. 79-111.

- Parisi, A., e G. Pasquino. 1977. Relazioni partiti-elettori e tipi di voto. In *Continuità e mutamento elettorale in Italia: le elezioni del 20 giugno 1976 e il sistema politico italiano*, a cura di A. Parisi e G. Pasquino. Bologna, Il mulino, pp. 215-249.
- Pinaraldi, F. 2000. Alla ricerca dell'inferenza perduta: la relazione tra dati ecologici e dati individuali. *Sociologia e ricerca sociale* 63, pp. 86-135.
- Rabinowitz, G., e S. E. Macdonald. 1989. A Directional Theory of Issue Voting. *American Political Science Review* (83), pp. 93-121.
- Ricolfi, L. 1993. La stima dei flussi elettorali. Oltre il modello standard. In *Quale mobilità elettorale? Tendenze e modelli. La discussione metodologica sui flussi elettorali*, a cura di R. Manheimer. Milano, Franco Angeli, pp. 137-196.
- Ricolfi, L., e S. Testa. 2002. Il mito del valore aggiunto dell'Ulivo. *Il Mulino* 400 (2), pp. 232-242.
- Robertson, D. B. 1976. *A Theory of Party Competition*, J. Wiley New York.
- Robinson, W. S. 1950. Ecological Correlation and the Behavior of Individuals. *American Sociological Review* 15, pp. 351-357.
- Rokkan, S. 1970. *Citizens, Elections, Parties*, Oslo, Universitetsforlaget.
- Rose, R., e I. McAllister. 1986. *Voters begin to choose: from closed class to open elections in Britain*, London ; Beverly Hills, Sage Publications.
- Sani, G. 2005. Contrattilità elettorale 2005: flussi, mobilitazione, ricambio. *Palena* (2), pp. 87-90.
- Sani, G., e P. Segatti. 2002. Fratture sociali, orientamenti politici e voto: ieri e oggi. In *Maggioritario finalmente? La transizione elettorale 1994-2001*, a cura di R. D'Alimonte e S. Bartolini. Bologna, Il Mulino, pp. 249-281.
- Särtilk, B., e I. Crewe. 1983. *Decade of dealignment: the Conservative victory of 1979 and electoral trends in the 1970's*, Cambridge, Cambridge University Press.
- Sartori, G. 1976. *Parties and Party Systems*, Cambridge, Cambridge University Press.
- Scappini, E., e D. Tuorto. 2006. Database: I risultati difformi del voto alla Camera e al Senato. *Il Mulino* LV (3), pp. 461-464.
- Schadee, H. M. A. 1993. Problemi attuali nell'analisi dei flussi elettorali. In *Quale mobilità elettorale? Tendenze e modelli. La discussione metodologica sui flussi elettorali*, a cura di R. Manheimer. Milano, Franco Angeli, pp. 273-322.
- Schadee, H. M. A., e P. Segatti. 2002. Informazione politica, spazio elettorale ed elettori in movimento. In *Le ragioni dell'elettore. Perché ha vinto il centro-destra nelle elezioni italiane del 2001*, a cura di M. Caciagli e P. Corbetta. Bologna, Il Mulino, pp. 339-369.
- Schattschneider, E. E. 1942. *Party government*, New York, Farrar and Rinehart.
- Segatti, P. 1997. Un centro instabile eppure fermo. Mutamento e continuità nel movimento elettorale. In *A domanda risponde. Il cambiamento del voto degli italiani nelle elezioni del 1994 e del 1996*, a cura di P. Corbetta e A. Parisi. Bologna, Il Mulino, pp. 215-259.
- Simon, H. A. 1957. *Models of Man. Social and Rational*, New York, Wiley.
- Smithies, A. 1941. Optimum Location in Spatial Competition. *Journal of Political Economy* 49 (3), pp. 423-439.
- Sørensen, A. B. 1998. Theoretical mechanisms and the empirical study of social processes. In *Social Mechanisms: An Analytical Approach to Social Theory*, a cura di P. Hedström e R. Swedberg. Cambridge, Cambridge University Press, pp. 238-266.
- Sraffa, P. 1926. The Laws of Returns under Competitive Conditions. *The Economic Journal* 36 (144), pp. 535-550.
- Statera, G. 1985. Le basi sociali del voto al referendum. *Prospettiva sindacale* XVI (56), pp. 36-44.
- , (a cura di) 1987. *Le basi sociali dei poli elettorali*. Milano: Franco Angeli.

- Simsom, J. A. 1975. Belief Systems: Constraint, Complexity, and the 1972 Election. *American Journal of Political Science* 19, pp. 393-417.
- Stokes, D. 1966. Spatial Models of Party Competition. In *Elections and the Political Order*, a cura di A. Campbell, P. E. Converse, W. E. Miller e D. Stokes. New York, Wiley, pp. 161-179.
- Stokes, D. E. 1963. Spatial Models of Party Competition. *American Political Science Review* 57, pp. 368-377.
- . 1992. Valence Politics. In *Electoral Politics*, a cura di D. Kavanagh. Oxford, Clarendon Press.
- Tagepera, R. 2005a. Beyond Regression: The Need for Logical Models. Paper presentato a Third Conference of the Belgian Political Science Association -French community, 29-30 April 2005, Liège.
- . 2005b. Predictive vs. Postdictive Models. Paper presentato a Third ECPR Conference, 8-10 September 2005, Budapest.
- Thomassen, J. 1976. Party identification as a cross-national concept: Its meaning in the Netherlands. *Party Identification and Beyond*, pp. 63-80.
- . 2005. *The European voter: a comparative study of modern democracies*, Oxford, Oxford University Press.
- Venturino, F. 2002. Il voto differenziale nelle elezioni per la Camera dei Deputati del 1996 e del 2001. Un'analisi descrittiva e causale. *Quaderni dell'Osservatorio Elettorale* 47, pp. 5-31.
- Zaller, J. 1992. *The nature and origins of mass opinion*, Cambridge, Cambridge University Press.
- . 2004. Floating Voters in U.S. Presidential Elections, 1948-2000. In *Studies in Public Opinion*, a cura di W. Saris e P. M. Sniderman. Princeton, Princeton University Press, pp. 166-214.