

**L'impatto della demografia delle imprese
nel calcolo di numeri indici: il caso
del comparto commerciale al dettaglio in Italia**

Roberto Gismondi (*)

(*) *ISTAT - Servizio SCO*

Sommario

Nell'ambito di un vasto programma di rinnovamento qualitativo, l'ISTAT ha intrapreso da anni un processo di revisione metodologica di tutte le rilevazioni condotte correntemente, molte delle quali si riferiscono al mondo delle imprese e sono finalizzate alla diffusione di indicatori di *output*. E' nell'ambito di una di queste rilevazioni, l'indagine mensile sulle vendite, che va inquadrata la necessità di approfondire lo studio della dinamica longitudinale degli operatori commerciali al dettaglio in sede fissa, oggetto primario della presente trattazione. Ciò è dovuto sia alla rilevanza di tale settore nell'economia nazionale (circa un terzo delle imprese italiane svolgono una attività prevalente di tipo commerciale, sia essa al minuto od all'ingrosso), sia al fatto che l'attuale indice del valore delle vendite al dettaglio, calcolato e diffuso ogni mese dall'ISTAT, richiederebbe, a rigore, la conoscenza mensile circa la reale consistenza del numero di operatori commerciali *attivi*, suddivisi per forma distributiva, area geografica, classe di addetti e tipologia dei prodotti venduti. Attualmente si opera in un contesto informativo più limitato, di cui si cercheranno di delineare le cause, in cui informazioni esaustive circa la consistenza degli operatori commerciali sono disponibili solo con cadenza annuale. In tale ottica l'articolo contiene delle proposte metodologiche per poter stimare la consistenza degli operatori commerciali, proponendone un confronto sia teorico che empirico.

Abstract

In the last years, ISTAT started a renewal process concerning all its current surveys, many of which refer to enterprises and are finalised at producing output indicators. In particular, the sample survey on retail trade sales was renewed as well, and it's now possible to calculate and disseminate monthly index numbers with a shorter delay than the recent past, in a frame characterised by a higher qualitative level. Notwithstanding that, we must realise that in many practical cases - and in the retail trade survey too - the correct calculation of an index number turns out to be difficult because of the lack of knowledge on enterprises' death, start-up and transformation process. When monthly indexes are needed, just yearly data about births and deaths of enterprises are often available, so that both significance and reliability of the index itself can be seriously affected. The aim of the paper is to propose some methods to estimate the amount of the existing retail trade operators (enterprise or outlets), in order to correct the indexes concerning the value of sales concerning the only operators included in the panel observed every month. These methods are fully explained and compared, with an empirical exercise at the end, and they should improve quality of indexes, even if in the short-term they could produce only rough estimates on the true amount of retailers.

1. Il legame tra il calcolo di numeri indici e la natimortalità¹ delle imprese

Nell'ambito delle rilevazioni a carattere economico, il calcolo di numeri indici rappresenta da molti anni uno strumento ampiamente utilizzato per misurare la dinamica longitudinale di una generica variabile quantitativa², misurabile su determinate unità statistiche (ad es., imprese). Ciò nonostante, quando si opera in un contesto campionario – quindi nella maggioranza delle situazioni correnti – è necessario soffermarsi su un particolare aspetto, non sempre oggetto delle dovute attenzioni in letteratura e nell'operatività corrente, costituito dalla corretta valutazione degli indici relativi alla variazione, rispetto al periodo scelto come base, del numero di unità attive nel dominio analizzato. Tale aspetto è determinante per poter interpretare l'indice calcolato e diffuso sulla base di una rilevazione parziale come un effettivo *indice di variazione dell'ammontare complessivo* della variabile studiata – ossia riferito all'intero universo di riferimento – e non piuttosto come un *indice relativo alla variazione media* di tale variabile sulle sole unità osservate nel campione disponibile.

In effetti, l'ipotesi che un campione ripetuto nel tempo possa considerarsi rappresentativo dell'intero universo è realistica solo nel breve periodo, e soprattutto in contesti caratterizzati da forte concentrazione imprenditoriale, bassa natimortalità delle imprese e stabilità longitudinale delle caratteristiche concorrenziali e tipologiche del comparto analizzato³. E' evidente come tali assunzioni siano poco realistiche soprattutto con riferimento al settore dei servizi, notoriamente caratterizzato da una forte polverizzazione e da una elevata turbolenza interna, fattori che rendono ardua la pianificazione di indicatori congiunturali di *output* per il comparto.

Più specificamente, in questo contesto si circoscriverà l'attenzione all'indagine mensile sulle vendite al dettaglio condotta dall'ISTAT, al momento l'unica indagine congiunturale nell'ambito del terziario di mercato in grado di produrre indici di valore. L'indagine è finalizzata al calcolo ed alla diffusione di indici del valore delle vendite al dettaglio di beni con un ritardo di circa 60 giorni rispetto al mese di riferimento e con riferimento a 5 dei 7 gruppi di attività economica contemplati nella divisione 52 della classificazione ATECO '91⁴. Come noto, da gennaio 1997 tale indagine è stata completamente rinnovata, sebbene tra i problemi tuttora aperti vada segnalato proprio la mancanza di un indicatore congiunturale (quantomeno trimestrale se non mensile) circa la natimortalità delle imprese, che rappresentano le unità di rilevazione. Rimandando a ISTAT (1998.2 e 1998.3) per ulteriori dettagli metodologici, in questo contesto ci si limiterà a

¹ Sebbene in letteratura si ricorra anche alla terminologia *demografia* delle imprese, l'etimologia del termine ne renderebbe preferibile il riferimento alle sole persone.

² Si rimanda, ad esempio, al noto saggio di Fisher (1992) o a Diewert (1995).

³ La riponderazione delle unità campionarie al fine di renderle rappresentative in funzione della dinamica dell'universo sembra concretamente attuabile solo con riferimento ad indagini campionarie annuali, dato che è difficile poter disporre a cadenza infraannuale di aggiornamenti dell'universo stesso (ad esempio, nell'indagine ISTAT sulle piccole e medie imprese si utilizza il metodo della ponderazione vincolata. Si rimanda al paragrafo 2 per ulteriori dettagli sui problemi legati all'aggiornamento dell'universo di riferimento).

⁴ All'interno della divisione 52 relativa al commercio al dettaglio si rilevano i gruppi 52.1 (vendite despecializzate), 52.2 (prodotti alimentari), 52.3 (prodotti farmaceutici e cosmesi), 52.4 (altri prodotti) e la classe 52.61 (vendite per corrispondenza), che spiegano circa il 95% del fatturato della divisione. Restano così escluse dal campo di osservazione le vendite al di fuori dei negozi diverse da quelle per corrispondenza, le vendite di beni usati e le riparazioni.

riassumere i termini del problema, ovviamente generalizzabile a qualunque altro contesto in cui si debbano calcolare indici di variazione basati su una osservazione parziale dell'universo di riferimento.

Si supponga dunque di voler stimare, tramite una rilevazione parziale, l'indice di variazione tra un tempo t ed un certo periodo b scelto come base, dell'ammontare complessivo di una variabile quantitativa Y relativa ad un universo di riferimento U . Mantenendo la trattazione nei limiti della più ampia generalità - indipendentemente dalla periodicità delle variazioni da calcolare (mensile, trimestrale, annuale) - con un'unica identità si potrà esprimere la stima del suddetto indice di variazione in funzione delle diverse componenti elementari *che è indispensabile calcolare* per poterne quantomeno accettare la *validità formale*. Supponendo, per semplicità di non considerare alcuna stratificazione delle unità (che nel prosieguo verranno indicate generalmente con il termine operatori, mentre nei casi più specifici ci si riferirà direttamente ad imprese o punti di vendita⁵) si possono introdurre queste definizioni:

N_t = numero degli operatori commerciali attivi al tempo t ;
 V_t = valore medio delle vendite per operatore commerciale al tempo t .

Consegue che l'indice di variazione *dell'ammontare complessivo* (cioè riferito all'intero universo di riferimento) della variabile studiata Y tra il periodo t ed il periodo base b sarà scrivibile nella forma seguente:

$$I_t = \frac{V_t N_t}{V_b N_b} = \left(\frac{V_t}{V_b} \right) \left(\frac{N_t}{N_b} \right) = (I_t^*)(I_t^{**}) \quad [1.1]$$

dove:

$$I_t^* = \left(\frac{V_t}{V_b} \right) \quad \text{e} \quad I_t^{**} = \left(\frac{N_t}{N_b} \right).$$

La prima componente esprime l'indice di variazione tra i due periodi prescelti *dell'ammontare medio* di Y per ogni impresa di U .

In genere, sul piano operativo tale stima si basa su un campione di unità di U , strutturato prevalentemente nella forma di *panel*⁶ (oppure è disponibile sulla base di altre fonti, quali dati amministrativi, archivi, precedenti indagini, ma questi casi sono meno frequenti ed esulano dal contesto di riferimento) e verrà anche indicata come *indice di tipo panel*. Si ribadisce come tale indice esprime la variazione del valore *medio* (e non *complessivo*) di Y calcolabile sulle unità osservate ed è dunque una stima campionaria della variazione «vera», per ipotesi incognita ed ovviamente nota solo qualora fosse osservabile l'intero universo oggetto d'indagine. Sulla qualità di tale componente influiscono direttamente due fattori: l'eventuale distorsione sistematica e la varianza campionaria. Nei prossimi paragrafi i simboli introdotti verranno sempre riferiti alle sole unità campionarie

⁵ Si noti come, a differenza del più generico termine di «operatore», i concetti di impresa e, in particolare, di punto vendita facciano uno stretto riferimento alla «fisicità» ed alla identificabilità spaziale del luogo in cui avvengono le transazioni commerciali, in parziale controtendenza rispetto all'attuale, rapida diffusione di forme di vendita telematiche basate su mercati «virtuali».

⁶ Con tale termine si farà riferimento ad un piano di osservazioni ripetute nel tempo su un sottoinsieme di unità di U eventualmente soggette a rotazione graduale. Per ulteriori dettagli si rimanda a Gismondi (1998.3) e ISTAT (1998.2).

effettivamente disponibili al tempo t , sebbene per evitare ambiguità quando necessario si introdurrà un pedice “c” identificativo delle sole misurazioni relative al campione.

La seconda componente identifica l'indice di natimortalità degli operatori tra i tempi t e b e riguarda l'effetto legato alla nascite, alle cessazioni ed alle trasformazioni societarie. E' una componente fondamentale soprattutto per la sua natura moltiplicativa ed è particolarmente importante nel caso di universi molto variabili, potendo rappresentare la principale fonte di errore nella stima di variazioni. Ad esempio, la mancata valutazione in sede di stima di un tasso medio di decrescita del numero di imprese operanti in un settore economico pari al 2% annuo comporterebbe, dopo cinque anni (tale è l'intervallo massimo che generalmente intercorre tra una misurazione ed il periodo scelto come base) ed a parità delle altre condizioni una sovrastima di poco inferiore al 10% della variazione dell'ammontare complessivo della variabile studiata.

Si noti anche come la scelta dell'unità di rilevazione - ad esempio la scelta tra impresa ed unità locale - *coeteris paribus* potrebbe avvenire proprio privilegiando tra le due quella la cui dinamica longitudinale risulti meno forte, sempre che la scelta non incida sulla qualità della prima componente.

L'indice dell'ammontare complessivo delle vendite e l'indice di tipo *panel* coincidono se $I_t^{**} = 1$; altrimenti, se si introduce la generica formula di Laspeyres⁷ per esprimere l'indice sintetico ottenibile per media aritmetica ponderata degli indici relativi allo strato *i*-mo, con pesi w_i , dove $w_i = \frac{V_{bi} N_{bi}}{\sum_i V_{bi} N_{bi}}$, la condizione di uguaglianza equivale a:

$$I_t = \sum_i I_{ti} w_i = \sum_i I_{ti}^* I_{ti}^{**} w_i .$$

Riducendo la trattazione a due soli strati indicabili con «1» e «2» si avrà che il calcolo dell'indice di natimortalità risulterà superfluo se:

$$\frac{V_{t1}}{V_{t2}} = \frac{(N_{t2} - N_{b2})}{(N_{b1} - N_{t1})} . \quad [1.2]$$

Ad esempio, se gli operatori dello strato 2 fossero aumentati di 10 unità e quelli dello strato 1 fossero diminuiti di 100, il rapporto a destra della precedente uguaglianza sarebbe pari a 0,10 e la stima di tale effetto di natimortalità sarebbe ininfluenza ai fini del calcolo dell'indice dell'ammontare complessivo se al tempo t il valore medio delle vendite degli operatori dello strato 2 fosse 10 volte più grande di quello relativo allo strato 1.

E' il caso che potrebbe presentarsi quando la chiusura di cn punti di vendita di tipo tradizionale è compensata dall'apertura di n punti di vendita della grande distribuzione, qualora quest'ultimi avessero un volume d'affari pari a c volte quello relativo ai punti di vendita tradizionali⁸.

In realtà questa ipotesi potrebbe risultare realistica solo nel caso di consumi finali approssimativamente costanti ed in economie ancora non completamente evolute riguardo al grado di penetrazione delle strutture commerciali moderne. Recenti studi (Bernardi, 1997)

⁷ Su tale aspetto si veda anche EUROSTAT (1998.2).

⁸ Per una definizione di grande distribuzione si rimanda nuovamente ad ISTAT (1998.3). Ci si limita a segnalare che in realtà non tutti i punti di vendita non appartenenti alla grande distribuzione sono considerabili necessariamente «tradizionali».

hanno evidenziato come esista un chiaro processo di sostituzione tra grande distribuzione ed imprese medio-piccole, soprattutto nel comparto alimentare, in cui mediamente negli anni '90 ad ogni apertura di una nuova grande superficie moderna ha corrisposto la chiusura di circa 53 punti di vendita tradizionali, sebbene poche evidenze siano disponibili circa la dinamica dei relativi ammontari medi delle vendite⁹.

La seguente tabella 1.1 contiene una verifica empirica di tale ipotesi riferita ai primi anni '90. Nelle colonne A e B sono riportati gli indici delle vendite calcolati sulla base della vecchia indagine, mentre le differenze calcolate nelle colonne C e D si basano sugli ammontari riportati nelle successive tabelle 2.1 e 2.2. Nella colonna E è calcolato il membro di sinistra della relazione [1.2], mentre nella colonna F i rapporti tra gli ammontari medi delle vendite per impresa sono ottenuti moltiplicando gli omologhi rapporti riferiti agli indici delle colonne A e B per 32,6, che rappresenta il rapporto, riferito al 1990, tra il valore medio delle vendite delle imprese della grande distribuzione e quelle tradizionali. Come è evidente, la validità di un effetto di sostituzione come quello implicato dalla [1.2] non risulta assolutamente confermata, dato che ad un andamento in costante (sebbene contenuta) decrescita di F corrisponde una dinamica assai più irregolare di E.

Alla luce di tale evidenza, a titolo esemplificativo si è tentato di ricalcolare gli indici delle vendite diffusi dall'ISTAT nei primi anni '90 cercando di incorporare una stima a posteriori dell'indice di natimortalità delle imprese nello stesso periodo.

In questo caso la stima della natimortalità si è basata sulla dinamica delle autorizzazioni d'esercizio, apparsa più regolare di quella relativa ad altri indicatori relativi allo stesso fenomeno¹⁰. I risultati sono riportati nel grafico 1.1, da cui emerge con chiarezza come dal 1991 al 1995 l'effetto della mortalità imprenditoriale nel commercio al dettaglio abbia comportato, rispetto al 1990, una sovrastima dell'indice del valore delle vendite di oltre sei punti (l'indice «corretto» risulterebbe pari a 113,6 rispetto al 119,8 medio del 1995), e solo nel biennio 1991-1992 tale effetto si mantiene entro limiti trascurabili. Tale evidenza suggerisce anche come possa risultare opportuno operare cambiamenti dell'anno base più frequenti di quelli generalmente previsti sia in sede nazionale che internazionale (si opera generalmente su base quinquennale), ed in ogni caso circoscrivere l'ampiezza dell'arco temporale entro cui operare dei confronti longitudinali.

Tabella 1.1 - Verifica dell'esistenza di un effetto di sostituzione (1991-1996)

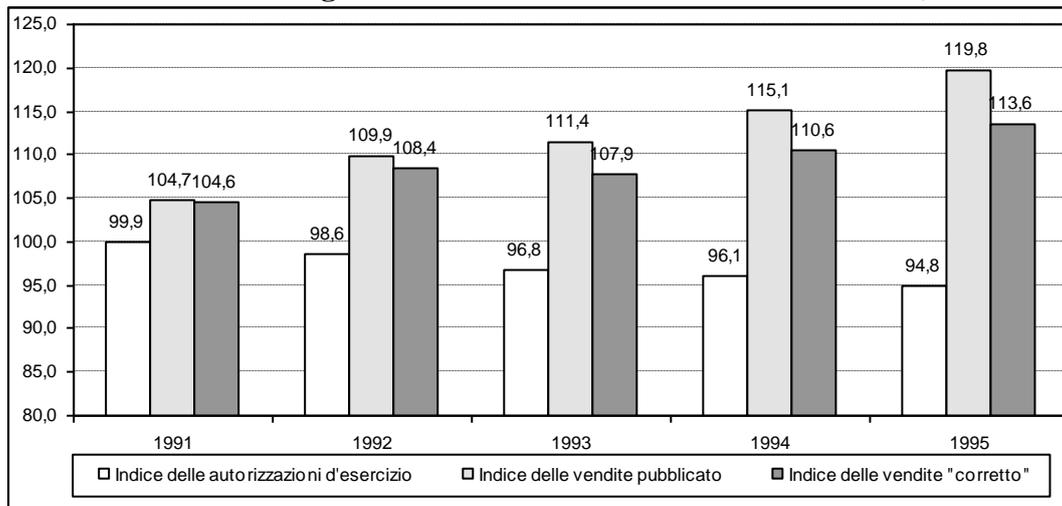
Anno	A (1) Imprese tradizionali	B (2) Grande distribuzione	C (N _{b1} -N _{p1})	D (N _{t2} -N _{b2})	E (D/C)x100	F(*) (A/B)x(32,6) x100
1991	103,6	110,0	-3.470	5	-0,14	2,89
1992	108,3	118,3	16.668	225	1,35	2,81
1993	109,3	122,4	67.636	441	0,65	2,74
1994	113,1	125,9	132.143	737	0,56	2,76
1995	117,1	133,7	140.838	1.362	0,97	2,69
1996	118,4	139,8	174.789	1.848	1,06	2,60

Fonte: elaborazioni su dati ISTAT e MICA (tabelle 2.1 e 2.2).

⁹ Per altri dettagli circa il tasso di sostituzione tra grande distribuzione e commercio tradizionale attualmente in essere in Italia si rimanda a ISTAT (1998.3 e 1998.4).

¹⁰ Confrontare oltre il paragrafo 2.

Grafico 1.1 – Stima degli indici del valore delle vendite «corretti» (1990-1995)



Fonte: elaborazioni su dati ISTAT.

In chiave internazionale, il Regolamento dell'Unione Europea sugli Indicatori Congiunturali (EUROSTAT, 1998.1), in vigore da luglio 1998, con riferimento al modulo «C» relativo al commercio al dettaglio non raccomanda particolari accorgimenti circa la tecnica di calcolo degli indici del fatturato delle imprese commerciali, nè è data particolare enfasi al problema dello studio della dinamica congiunturale (quantomeno trimestrale) del numero delle imprese attive. E' peraltro previsto uno studio pilota relativo alla possibilità di implementare una rilevazione congiunturale sulle creazioni e cessazioni d'impresa. E' evidente come sussista una certa tautologia in tale approccio: si relega a studio pilota l'esame della dinamica congiunturale sulle creazioni e cessazioni d'impresa e si richiede, al contempo, una certificazione di qualità per gli indici prodotti mensilmente, il che rappresenta un obiettivo non realizzabile, a meno di non disporre già di un buon indicatore congiunturale sulla dinamica longitudinale delle imprese¹¹ o di disporre di rilevazioni esaustive.

Con riferimento all'indice del valore delle vendite al dettaglio diffuso mensilmente dall'ISTAT sulla base della nuova indagine e già richiamato in precedenza, la procedura di calcolo consente di pervenire ad una stima congiunta del prodotto tra la prima e la seconda componente a destra della [1.1], sebbene verifiche a posteriori abbiano evidenziato come tale tecnica riduca, *senza annullarlo*, l'effetto distorto indotto dalla attuale non misurabilità del termine I_t^{**} .

Una proposta operativa utilizzabile per correggere a posteriori gli indici stimati e diffusi mensilmente senza una adeguata valutazione di tale termine sarà illustrata tra breve: essa consente di pervenire quantomeno ad una stima indiretta dell'effetto indotto dagli indici di natalità *di strato*, supponendo realisticamente di avere a disposizione solo un indice di natalità degli operatori I_t^{**} riferito ad uno strato *più aggregato* (ad esempio, il totale del commercio al dettaglio, o al più le singole quote alimentari e non alimentari).

¹¹ Tale problema si presenta anche con riferimento al modulo "D" del Regolamento Congiunturale, che prevede la produzione di indici trimestrali di fatturato per 25 settori del terziario, commercio al dettaglio escluso.

Se si fa riferimento all'*i-mo* indice di strato, si desidera ricavare dei nuovi indici I_{ti} che risultino il meno possibile diversi dagli indici I_{ti}^* effettivamente calcolati e diffusi mensilmente senza tenere in debito conto l'effetto di natimortalità I_{ti}^{**} , cercando cioè di minimizzare la funzione:

$$\Phi^2 = \sum_i (I_{ti} - I_{ti}^*)^2,$$

con il vincolo che la media ponderata dei nuovi indici di strato riproduca l'indice marginale corretto stimato a posteriori¹², ossia:

$$\sum_i I_{ti} w_i = I_t, \quad [1.3]$$

laddove varrà invece la relazione:

$$\sum_i I_{ti}^* w_i = I_t^*.$$

Il ricorso ad una simile funzione obiettivo da minimizzare deriva dalla necessità di ridurre il più possibile lo scarto tra dati diffusi preliminarmente e dati ricalcolati per tenere conto degli indici di natimortalità, di cui si suppone effettivamente calcolabile *a posteriori* solo quello relativo al complesso degli strati. Ciò in quanto uno degli obiettivi generalmente richiesti ad un organismo produttore di statistiche ufficiali consiste nel pianificare un sistema produttivo tale da ridurre al minimo lo scarto tra dati «preliminari» e dati «finali»¹³. Omettendo lo sviluppo algebrico, con il metodo dei moltiplicatori di Lagrange si ottiene facilmente la soluzione:

$$I_{ti} = I_{ti}^* + \frac{w_i (I_t - I_t^*)}{\sum_i (w_i)^2}. \quad [1.4]$$

Quindi per ogni strato i il nuovo indice “corretto” si otterrà sommando all'indice calcolato senza tenere conto della dinamica delle imprese un fattore positivo (negativo) se l'indice I_{ti} è maggiore (minore) dell'indice I_{ti}^* . In una situazione, come quella che sta caratterizzando gli anni '90, di progressiva flessione del numero di operatori commerciali al minuto tale fattore sarà dunque in prevalenza negativo, il che potrebbe comportare il problema della negatività di qualcuno dei nuovi indici; tale metodo dovrebbe comunque risultare preferibile al più immediato criterio del riproporzionamento, che peraltro non garantisce necessariamente il soddisfacimento della condizione [1.3].

Con queste premesse, il presente lavoro ha come obiettivo principale quello di proporre alcuni metodi di stima dell'indice di natimortalità degli operatori commerciali in chiave congiunturale (mensile o trimestrale), nell'ottica di illustrare e confrontare tecniche effettivamente implementabili in pratica, ossia basate su eventuali variabili ausiliarie

¹² Se, ad esempio, si potesse stimare a posteriori l'indice corretto per il totale dei settori, con tale procedura si potrebbero ricalcolare dei nuovi indici per i settori “alimentare” e “non alimentare”, tali che la loro media ponderata riproduca proprio l'indice corretto per il totale dei settori.

¹³ In ogni caso, la raccomandazione di base da seguire qualora si opti per la diffusione sia di dati provvisori che di dati definitivi consiste nel prevedere confronti longitudinali tra coppie di indici provvisori e coppie di indici definitivi *distintamente*, onde evitare di ridurre, anziché di aumentare, il reale contenuto informativo di un tale schema di diffusione.

effettivamente disponibili. Si è quindi cercato sia di selezionare, tra i metodi esistenti più o meno noti, quelli applicabili (anche) in chiave congiunturale, sia di presentare alcune proposte originali, schematizzando la presentazione supponendo che siano disponibili serie storiche brevi o lunghe, con l'ausilio o meno di variabili ausiliarie disponibili in chiave congiunturale (paragrafo 4). Si è inoltre deciso di fornire maggiori dettagli metodologici nei casi in cui la maggiore complessità della trattazione lo richiedesse ai fini di una più corretta comprensione delle tecniche proposte (ciò avviene in particolare nel paragrafo 4.4). Nel paragrafo 5 è poi proposto un metodo generalizzato per stimare le frequenze intermedie (o di secondo ordine), supponendo che siano note un certo numero di frequenze marginali: come già visto, si può presentare il problema di dover stimare la consistenza al tempo t degli operatori commerciali a vendita prevalente di prodotti «alimentari» e «non alimentari», supponendo nota la consistenza riferita allo stesso periodo del totale degli operatori. Infine, nel paragrafo 6 è riportata una applicazione comparativa delle varie tecniche, da cui sostanzialmente emerge la preferibilità delle procedure dotate di maggiori proprietà matematiche. Ricordando che nel prosieguo i termini di campione e *panel* saranno usati in modo interscambiabile, prima di procedere è necessario riepilogare, nel paragrafo seguente, le principali fonti informative attualmente disponibili circa la dinamica longitudinale degli operatori commerciali in Italia, ed illustrare alcuni, semplici accorgimenti per tentare di attenuare la mancata valutazione dell'indice di natimortalità nel calcolo di un indici dell'ammontare complessivo di una data variabile quantitativa.

2. Attuali fonti informative sulla natimortalità delle imprese commerciali in Italia

Nel complesso, in Italia sembrano essere (o essere state) essenzialmente cinque le tipologie di fonti informative da cui attualmente possono essere tratte informazioni circa lo *stock*, la nascita e la cessazione degli operatori commerciali:

1. l'archivio ASIA dell'ISTAT;
2. il Registro Ditte delle Camere di Commercio;
3. le autorizzazioni comunali di esercizio;
4. la rilevazione del Ministero del Commercio, dell'Industria e dell'Artigianato (MICA);
5. le rilevazioni e/o le stime elaborate da istituti specializzati.

A partire dal 1995 l'ISTAT ha dato il via al progetto ASIA, finalizzato alla creazione ed all'aggiornamento di un "Archivio Statistico delle Imprese Attive". Con tale iniziativa l'Italia è in grado di ottemperare a quanto richiesto da un apposito Regolamento dell'Unione Europea relativo agli archivi delle imprese. L'archivio, basato sull'unità economica «impresa», è ottenuto sulla base della sintesi di diversi archivi amministrativi: Anagrafe Tributaria, Registri Ditte delle Camere di Commercio (da cui derivano le informazioni circa la nascita, la cessazione e le modifiche delle imprese commerciali), archivi previdenziali INPS ed INAIL, archivio SEAT¹⁴. Il primo impianto è stato realizzato alla fine del 1996 e l'entrata a regime è stata realizzata all'inizio del 1998.

Il limite principale dell'archivio ASIA, peraltro esaustivo e dettagliato fino alle 5 cifre della classificazione nazionale ATECO '91, è quello di essere attualmente disponibile in una

¹⁴ In seguito si aggiungeranno archivi settoriali come quelli ABI (credito), ANIA (assicurazioni), PRA (autoveicoli).

versione «provvisoria», aggiornata con un ritardo biennale rispetto al periodo di riferimento delle informazioni. Sebbene il suo utilizzo in chiave congiunturale sembra al momento irrealistico, esso può rappresentare comunque un valido *benchmark* annuale per validare stime delle dinamiche infraannuali derivanti da altre fonti, caratterizzate generalmente da un profilo qualitativo inferiore.

Il Registro Ditte delle Camere di Commercio¹⁵ è, come già visto, uno degli archivi utilizzato per la gestione dell'archivio ASIA, ma pur essendo ipotizzabile il ricorso in chiave infraannuale alle informazioni in esso contenute va ricordato quanto segue:

- una gran parte delle imprese giuridicamente nuove rappresentano, in realtà, soltanto trasformazioni di unità produttive preesistenti;
- una quota abbastanza rilevante di nuove imprese registrate non inizia l'attività: l'atto amministrativo dell'iscrizione non rappresenta, infatti, condizione sufficiente per iniziare effettivamente l'attività produttiva;
- sussistono tuttora problemi classificatori che rendono problematico un monitoraggio capillare del comparto commerciale al dettaglio.

Per i primi due motivi, soprattutto qualora l'oggetto specifico di interesse sia la formazione di nuove imprese, i dati del Registro Ditte sono dunque probabilmente sovrastimati (Garofoli, 1994). Dato che la gestione di ASIA prevede un trattamento statistico delle informazioni in grado di assegnare ad ogni impresa una probabilità di «attività», il solo ricorso al Registro Ditte rischia di configurarsi come un duplicazione informativa peraltro di livello qualitativo probabilmente inferiore, almeno per le finalità del contesto ivi trattato.

Riguardo alle autorizzazioni comunali di esercizio per il commercio fisso al minuto, tale rilevazione è stata sospesa dall'ISTAT nel 1996, e gli ultimi dati disponibili si arrestano al 1994, con un dettaglio provinciale basato su 29 categorie merceologiche raggruppate in 4 grandi comparti (alimentari, tessuti, mobili e materiali per la casa, altri prodotti). In ogni caso, già prima dell'entrata in vigore della nuova normativa sulle autorizzazioni commerciali¹⁶ si doveva tenere presente che:

- dato che l'autorizzazione alla vendita di beni diversi richiedeva, in genere, l'acquisto di altrettante «tabelle merceologiche», l'acquisto di una pluralità di autorizzazioni all'esercizio di vendita poteva corrispondere all'inizio dell'attività di una sola impresa;
- il livello di dettaglio con cui risultavano disponibili le informazioni sulla natalità degli operatori commerciali era comunque insufficiente per le applicazioni di carattere congiunturale discusse in questo contesto.

¹⁵ Le Camere di Commercio hanno reso operativo, a partire dal 1990, un sistema informativo sulla rete distributiva noto con l'acronimo Siredi. Questo sistema raccoglie, grazie ai Comuni, le informazioni contenute nei modelli attraverso i quali avviene il rilascio dell'autorizzazione commerciale (per ulteriori dettagli si rimanda a INDIS, 1998).

¹⁶ Sulla base di un recente Decreto Legislativo, è stato abolito l'obbligo di acquistare una o più licenze per lo svolgimento di attività di vendita al dettaglio e, implicitamente, anche l'obbligo di segnalare alle Camere di Commercio l'inizio di una nuova attività. Il rilascio delle autorizzazioni d'esercizio spetta alle Regioni per quanto riguarda le superfici superiori ai 2.500 metri quadrati ed ai Comuni per le restanti, con una serie di facilitazioni burocratiche che favorirebbero il passaggio dalla piccola superficie (non oltre 200 metri quadrati) alla media e grande. Questa decentralizzazione dell'apparato di controllo amministrativo e statistico finisce con il rendere ancora più ardua la misurazione della natalità delle imprese commerciali.

Ciò nonostante le evidenze empiriche mostrano come sussista una buona correlazione tra dinamica delle autorizzazioni e delle imprese attive.

Il Ministero del Commercio, dell'Industria e dell'Artigianato (MICA) diffonde statistiche annuali anch'esse basate sui dati camerali, ma arricchite da una analisi molto approfondita sulla grande distribuzione e da alcuni correttivi che cercano di ridurre gli effetti di sovrastima succitati. Ciò nonostante il dato ministeriale, diffuso a cadenza annuale, è disponibile solo dopo circa 10 mesi dalla fine dell'anno di riferimento, e basandosi sull'entità economica «punto di vendita» (anzichè sull'impresa) non consente un facile raccordo con la classificazione ATECO, per cui le sole informazioni concretamente utilizzabili per le nostre finalità si limitano alla distinzione tra la vendita prevalente di generi alimentari o non alimentari. Per quanto riguarda, poi, il calcolo del numero complessivo dei punti di vendita attivi nel comparto al dettaglio, le valutazioni relative agli anni 1995 e 1996 sembrano affette da una sensibile sovrastima delle cessazioni, come anche evidenziato dal confronto con le analoghe valutazioni elaborate da Movimprese (cfr. ISTAT, 1998.4, 144).

Infine, le rilevazioni e gli studi condotti da istituti specializzati (in particolare CESCO, Gruppo CLAS, ISCOM, FAID) presentano l'ovvio limite della frequente parzialità del campo di osservazione, del ricorso a definizioni almeno parzialmente soggettive - non sempre raccordabili a quelle ufficiali - e della frequente carenza di note metodologiche che aiutino l'utilizzatore ad orientarsi al momento della scelta della fonte ritenuta più idonea per le proprie finalità. D'altra parte, sono soprattutto questi organismi a stimolare l'attenzione verso il tentativo di legare la dinamica delle imprese commerciali a possibili variabili esplicative, come i comportamenti di spesa, l'evoluzione strategica dell'offerta ed i principali indicatori macroeconomici.

Riassumendo, l'attuale panorama informativo consente, peraltro con un certo ritardo, di poter monitorare in chiave congiunturale la consistenza degli operatori commerciali attivi con un livello di dettaglio assai ridotto, spesso limitato alla sola distinzione tra prodotti alimentari e non, con una attenzione agli aspetti territoriali e competitivi scarsa o nulla.

Dalla tabella 2.1 emerge una sensibile differenza tra le dinamiche longitudinali evidenziate dai dati sulle autorizzazioni rispetto ai dati del MICA sui punti di vendita soprattutto a partire dal 1993 e, per i generi non alimentari, dal 1994. Di conseguenza nel 1995 si registra una flessione complessiva rispetto al 1990 pari al 2,7% per le autorizzazioni ed al 10,4% per i punti di vendita; inoltre, mentre il comparto non alimentare risulta in sia pur lieve crescita in termini di autorizzazioni (+0,3%), in termini di punti di vendita subisce un forte calo (8,8%), di poco inferiore alla media generale. Sebbene la forte crisi del biennio 1993-94 porti a ritenere più indicativo il segnale di flessione derivante dai dati MICA piuttosto che la sostanziale stabilità delle autorizzazioni (molti operatori, senza privarsi delle autorizzazioni, potrebbero aver sospeso l'attività), il tasso di decrescita dei punti di vendita sembra oggettivamente eccessivo, e risulterebbe ancora più evidente se si analizzassero i dati relativi al 1996, qui omessi¹⁷.

¹⁷ Dopo la diffusione dei dati relativi al 1996, il MICA ha attualmente sospeso la diffusione di tali statistiche, a causa di problemi metodologici inerenti all'esatto conteggio del numero dei punti di vendita effettivamente operativi, che nei primi anni '90 potrebbe essere stato sottostimato a causa di una forte sovrastima delle cessazioni.

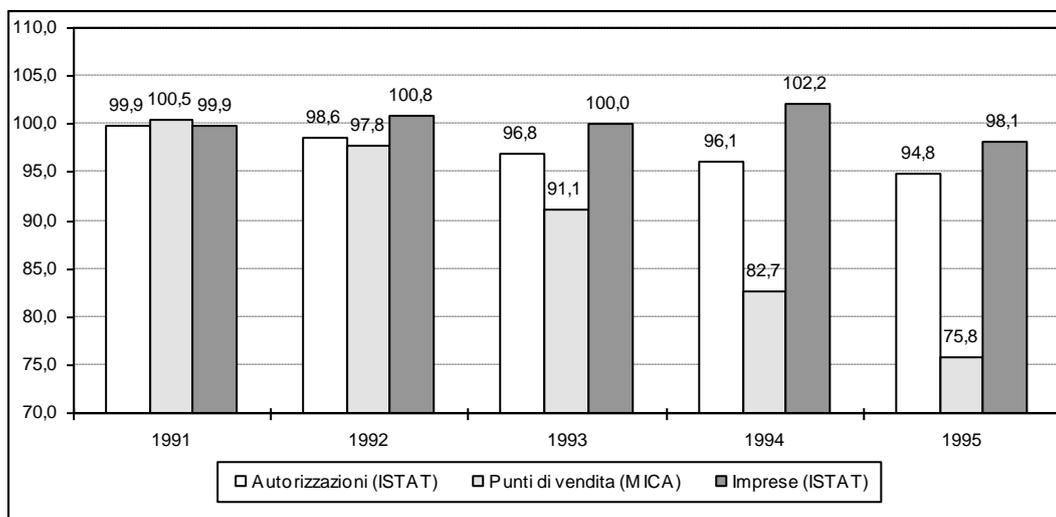
Tabella 2.1 - Confronto tra la dinamica delle autorizzazioni commerciali ed il numero dei punti di vendita (1990-1995)

Anno	AUTORIZZAZIONI (ISTAT)			PUNTI DI VENDITA (MICA)		
	Alimentari	Non alimentari	Totale	Alimentari	Non alimentari	Totale
Ammontari assoluti						
1990	373.942	675.730	1.049.672	277.961	480.610	758.571
1991	365.828	682.721	1.048.549	273.869	488.177	762.046
1992	352.400	683.001	1.035.401	263.686	478.442	742.128
1993	340.881	675.717	1.016.598	243.854	447.522	691.376
1994	332.406	676.103	1.008.509	220.856	406.309	627.165
1995	322.841	672.691	995.532	203.085	372.145	575.230
Numeri indici (1990=100)						
1990	100,0	100,0	100,0	100,0	100,0	100,0
1991	97,8	101,0	99,9	98,5	101,6	100,5
1992	94,2	101,1	98,6	94,9	99,5	97,8
1993	91,2	100,0	96,8	87,7	93,1	91,1
1994	88,9	100,1	96,1	79,5	84,5	82,7
1995	86,3	99,6	94,8	73,1	77,4	75,8
Media	91,7	100,3	97,3	86,7	91,2	89,6

Fonte: elaborazioni su dati ISTAT e MICA. Il dato ISTAT del 1995 è stimato.

La sola evidenza comune alle due fonti è la conferma del processo di concentrazione del comparto distributivo, particolarmente forte nel settore alimentare e dovuto soprattutto al ristagno dei consumi reali ed alla crescente concorrenzialità interna.

Grafico 2.1 - Indici di variazione (1990=100) delle autorizzazioni, dei punti di vendita e delle imprese (1991-1995)



Fonte: elaborazioni su dati ISTAT e MICA. Il dato ISTAT del 1995 sulle imprese è stimato.

Se si osserva il grafico 2.1, in cui rispetto alla tabella 2.1 viene aggiunto il dato ISTAT relativo al numero delle imprese attive nel comparto commerciale al dettaglio, sussiste un chiaro allineamento tra i dati sulle autorizzazioni e quelli sulle imprese attive, ad eccezione del 1995, sebbene tale discrepanza possa derivare almeno in parte dal fatto che il primo dei due dati è stato oggetto di stima.

Una maggiore convergenza sembra invece caratterizzare gli indici di variazione della tabella 2.2, relativi a tre forme di vendita della grande distribuzione. A fronte di una sostanziale coincidenza dei dati relativi ai supermercati, le principali divergenze sono imputabili alla minore consistenza nel 1990 di grandi magazzini ed ipermercati secondo le statistiche del MICA, che adotta generalmente criteri più rigidi della FAID per accertare l'effettiva operatività delle grandi superfici moderne. Tale evidenza conferma come, al momento, solo la dinamica della grande distribuzione sia misurabile con una certa precisione indipendentemente dalla fonte, mentre la grande maggioranza di operatori «tradizionali» sembra sfuggire ad un conteggio sufficientemente affidabile.

Tabella 2.2 - Confronto tra la dinamica dei punti di vendita della grande distribuzione rilevati dalla FAID e dal MICA (1990-1997)

Anno	DATI MICA				DATI FAID			
	Grandi magazzini	Supermercati	Ipermercati	Totale	Grandi magazzini	Supermercati	Ipermercati	Totale
Ammontari assoluti								
1990	942	3.399	150	4.491	937	3.371	114	4.422
1991	849	3.465	182	4.496	957	3.551	134	4.642
1992	838	3.696	182	4.716	935	3.766	165	4.866
1993	823	3.906	203	4.932	931	3.978	188	5.097
1994	820	4.198	210	5.228	924	4.253	206	5.383
1995	841	4.787	225	5.853	943	4.779	220	5.942
1996	902	5.207	230	6.339	1.007	5.231	247	6.485
1997	904	5.449	240	6.593	1.011	5.462	268	6.741
Numeri indici (1990=100)								
1990	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
1991	90,1	101,9	121,3	100,1	102,1	105,3	117,5	105,0
1992	89,0	108,7	121,3	105,0	99,8	111,7	144,7	110,0
1993	87,4	114,9	135,3	109,8	99,4	118,0	164,9	115,3
1994	87,0	123,5	140,0	116,4	98,6	126,2	180,7	121,7
1995	89,3	140,8	150,0	130,3	100,6	141,8	193,0	134,4
1996	95,8	153,2	153,3	141,1	107,5	155,2	216,7	146,7
1997	96,0	160,3	160,0	146,8	107,9	162,0	235,1	152,4
Media	90,6	129,1	140,2	121,4	102,3	131,5	178,9	126,5

Fonte: elaborazioni su dati MICA e FAID.

Utilizzando i dati finora introdotti, è possibile verificare, sia pure con le dovute approssimazioni¹⁸, l'entità di una eventuale correzione a posteriori degli indici medi annui dell'ammontare complessivo delle vendite per il periodo 1991-1996, qualora si tenesse conto del «vero» effetto di natimortalità (per ipotesi noto *solo* con riferimento al totale dei prodotti) espresso dagli indici I_t^{**} calcolabili sfruttando la dinamica delle autorizzazioni d'esercizio così come riportata nella tabella 2.1.

Tabella 2.3 – Stima degli indici delle vendite “corretti” [1.1] con il criterio [1.3]

Anno	Indici delle vendite I^*			Pesi			Indici di natimortalità I^{**}		
	Totale	Non		Totale	Non		Totale	Non	
		Alimentari	alimentari		Alimentari	alimentari		Alimentari	alimentari
1990	100,0	100,0	100,0	1,000	0,354	0,646	100,0	100,0	100,0
1991	104,7	103,9	105,1	1,000	0,350	0,650	99,9	97,8	101,0
1992	109,9	111,4	108,8	1,000	0,406	0,594	98,6	94,2	101,1
1993	111,4	116,1	108,0	1,000	0,415	0,585	96,8	91,2	100,0
1994	115,1	123,1	109,4	1,000	0,410	0,590	96,1	88,9	100,1
1995	119,8	130,0	112,4	1,000	0,419	0,581	94,8	86,3	99,6
1996	121,8	135,6	111,9	1,000	0,415	0,585	93,6	85,9	97,9

Anno	Stima indici “corretti” $I=(I^*)(I^{**})$			Indici riproporzionati			Differenza tra indici I e I^*		
	Totale	Non		Totale	Non		Totale	Non	
		Alimentari	alimentari		Alimentari	alimentari		Alimentari	alimentari
1990	100,0	100,0	100,0	100,0	100,0	100,0			
1991	104,6	103,8	104,9	104,6	103,8	104,9	-0,1	-0,1	-0,1
1992	108,4	110,5	107,0	108,4	109,9	107,3	-1,5	-1,0	-1,8
1993	107,9	113,8	103,8	107,9	112,5	104,6	-3,5	-2,3	-4,2
1994	110,6	120,2	104,1	110,6	118,3	105,1	-4,5	-2,9	-5,4
1995	113,6	126,0	105,0	113,6	123,3	106,6	-6,2	-4,0	-7,4
1996	114,0	130,5	102,6	114,0	127,0	104,7	-7,8	-5,1	-9,3

Nota: gli indici di natimortalità si riferiscono alle autorizzazioni della tabella 2.1, con l'aggiunta del dato '96.

Nella tabella 2.3 sono riportati gli indici di tipo *panel* I_t^* calcolati e diffusi correntemente dall'ISTAT, i pesi ad essi associati e gli indici di natimortalità I_t^{**} , posti uguale agli indici di variazione del numero di autorizzazioni commerciali riportati nella tabella 2.1. Con tali informazioni è possibile stimare gli indici corretti relativi al totale dei prodotti moltiplicando I_t^* per I_t^{**} (fino al 1995 tali indici sono stati già utilizzati nel grafico 1.1) e successivamente, tramite la formula [1.1], gli indici corretti relativi ai comparti alimentare e non alimentare. In chiave comparativa sono riportati anche i corrispondenti indici ottenibili

¹⁸ In effetti lo schema di calcolo dell'indice sintetico relativo al totale dei prodotti si basa su una stratificazione più complessa della semplice distinzione tra comparto alimentare e non alimentare. Ciò implica che nella schematizzazione della tabella 2.3 i pesi associati ai comparti suddetti varino di anno in anno.

con un semplice riproporzionamento. Secondo tali valutazioni nel periodo 1991-1996 l'indice dell'ammontare complessivo delle vendite sarebbe stato sovrastimato di 7,8 punti, e l'eventuale correzione a posteriori degli indici relativi ai prodotti alimentari e non alimentari finalizzata a minimizzare lo scarto medio tra indici già diffusi ed indici corretti implicherebbe sovrastime pari, rispettivamente, a 5,1 ed a 9,3 punti.

3. Metodi per attenuare la mancata valutazione dell'indice di natimortalità

Un'impresa commerciale al dettaglio può operare con una molteplicità di punti di vendita¹⁹, per cui si può introdurre questa nuova definizione:

P_t = numero dei punti di vendita operativi al tempo t .

Sulla base delle formalizzazioni introdotte nel primo paragrafo, l'indice dell'ammontare complessivo delle vendite sarà anche scrivibile in questa forma:

$$I_t = (I_t^*)(I_t^{**}) = \left(\frac{V_t}{V_b}\right)\left(\frac{N_t}{N_b}\right) = \left(\frac{V_t}{N_t}\right)\left(\frac{N_t}{N_b}\right)^2 = (I_t^m)(I_t^{**})^2 \quad [3.1]$$

e quindi l'indice dell'ammontare complessivo è pari al prodotto tra l'indice dell'ammontare medio per impresa (I_t^m) ed il quadrato dell'indice relativo alla dinamica delle imprese. Supponendo di poter disporre della dinamica del totale dei punti di vendita tra i tempi t e b , e di continuare a riferire le diverse grandezze all'intero universo, il ricorso al calcolo di un indice dell'ammontare medio per punto di vendita comporterebbe l'equivalenza con il calcolo dell'indice dell'ammontare complessivo delle vendite se valesse la relazione:

$$I_t = \left(\frac{V_t}{P_t}\right) \left(\frac{V_b}{P_b}\right) \quad \text{ossia, sulla base della [3.1], se:} \quad \left(\frac{P_t}{P_b}\right) = \frac{1}{I_t^{**}} \quad [3.2]$$

Occorrerebbe, quindi, che la dinamica longitudinale dei punti di vendita fosse pari al reciproco della dinamica delle imprese, ossia che ad un aumento (una flessione) del numero dei punti di vendita delle imprese rispondenti corrisponda una flessione (un aumento) del numero delle imprese attive nell'universo. Altrimenti, se valesse la relazione:

$$\left(\frac{P_t}{P_b}\right) = I_t^{**} \quad [3.3]$$

¹⁹ Sulla base dell'ultimo Censimento dell'Industria e dei Servizi, il numero medio di unità locali per ogni impresa commerciale al dettaglio era pari a 1,1, ma esistono imprese della grande distribuzione con oltre 100 punti di vendita.

l'indice dell'ammontare complessivo delle vendite potrebbe essere calcolato moltiplicando I_t^* per l'indice di variazione dei punti di vendita (P_t/P_b).

Il riferimento ai punti di vendita è importante, perchè fino a dicembre 1996 l'ISTAT ha calcolato e diffuso, con riferimento alla vecchia indagine ed alle imprese medio-grandi (con almeno 10 addetti) due serie storiche mensili di indici delle vendite: gli indici I_{tc}^* dell'ammontare medio riferito alle unità del *panel* e gli omologhi indici per punto di vendita calcolati sui dati campionari forniti dalle imprese del *panel* stesso e dati dalla relazione:

$$\frac{\left(\frac{V_{tc}}{P_{tc}}\right)}{\left(\frac{V_{bc}}{P_{bc}}\right)} \quad [3.4]$$

L'opinione, diffusa tra molti utilizzatori di statistiche sul commercio interno, che il calcolo di un indice per punto di vendita possa in qualche forma incorporare l'effetto della dinamica delle imprese e quindi risultare più preciso di un indice relativo al solo ammontare del *panel* è in effetti valida - al netto delle ovvie valutazioni addizionali circa l'errore campionario dovuto alla natura parziale della rilevazione dei dati necessari per i calcoli - solo se valesse la relazione [3.2], ossia se la dinamica longitudinale dei punti di vendita delle imprese del *panel* fosse almeno approssimativamente uguale al reciproco della dinamica longitudinale delle imprese dell'universo. Tale ipotesi appare nel complesso irrealistica, perchè non sembra plausibile supporre che la selezione di un campione di imprese rappresentativo dell'universo in termini di dinamica delle vendite (e di addetti) implichi una associazione inversa tra dinamica dei punti di vendita ed il complesso delle imprese.

Una verifica empirica è possibile notando come il rapporto tra l'indice delle vendite I_{tc}^* ed il suddetto indice medio per punto di vendita espresso dalla [3.4] - entrambi riferiti al *panel* - sia proprio pari al rapporto (P_{tc}/P_{bc}), per cui supponendo che tale rapporto possa approssimare il rapporto (P_t/P_b) riferito all'universo sarà possibile verificare o meno la validità della relazione [3.2].

Nella tabella 3.1 sono stati riportati gli indici del valore delle vendite complessivi (colonna A) e per punto di vendita (colonna B) per le imprese con almeno 10 addetti e per gli anni dal 1990 al 1995, nonchè (colonna C) gli indici relativi al numero dei punti di vendita della grande distribuzione (supermercati, grandi magazzini ed ipermercati) la cui dinamica dovrebbe approssimare bene la dinamica la dinamica dei punti di vendita delle imprese medio-grandi²⁰. La stima di tale dinamica tramite il rapporto (moltiplicato per 100) tra le colonne B e A fornisce i dati della colonna E, che differiscono in maniera evidente da quelli "veri" della colonna C. Sembra peraltro esistere una chiara connessione diretta tra la dinamica dell'indice delle vendite di tipo *panel* (colonna A) e quella dei punti di vendita della grande distribuzione (colonna C), che sarà approfondita nel paragrafo seguente.

E' soprattutto a causa della ambiguità interpretativa che ha caratterizzato nel recente

²⁰ In realtà possono esistere imprese della grande distribuzione con meno di dieci addetti, concentrate essenzialmente nella fascia da 6 a 9 addetti. Sulla base di quanto riportato in ISTAT (1998.2), nel 1995 le imprese della grande distribuzione con meno di 10 addetti erano circa il 35% del totale, con un peso in termini di fatturato pari a circa il 17%. D'altra parte, non essendo, disponibili indici delle vendite per punto vendita per le imprese con meno di 10 addetti la verifica empirica della tabella 3.1 sconta una parziale incongruenza degli aggregati confrontati, sebbene l'evidenza dei risultati sembri difficilmente confutabile.

passato gli indici per punto di vendita che l'ISTAT dal 1997 non ha ancora diffuso tale tipo di informazione con riferimento alla nuova indagine sulle vendite. La verifica del legame tra le dinamiche longitudinali del numero di punti di vendita e di imprese nel triennio 1996-1998 dovrebbe comunque consentire di calcolare, nel 1999, sia gli indici per punto vendita che una stima degli indici dell'ammontare complessivo delle vendite, in cui l'indice è stimabile con una delle tecniche riportate in seguito.

Tabella 3.1 - Stima della dinamica delle imprese tramite gli indici per punto vendita

Anno	Indici (1990=100)					
	A	B	C	D	E	F
	Indice delle vendite tipo <i>panel</i>	Indice delle vendite per punto vendita	Imprese della G.D.	A/B	(1/D)x100	C-E
1990	100,0	100,0	100,0			
1991	110,0	108,0	105,0	1,02	98,2	6,8
1992	118,4	112,6	110,0	1,05	95,1	14,9
1993	122,4	115,9	115,3	1,06	94,7	20,6
1994	126,0	116,8	121,7	1,08	92,7	29,0
1995	133,9	122,2	134,4	1,10	91,3	43,1

Fonte: elaborazioni su dati ISTAT e MICA.

In effetti, l'informazione sulla dinamica dei punti di vendita disponibile tramite il *panel* osservato mensilmente può *comunque* essere sfruttata per la stima della dinamica longitudinale delle imprese, sulla base di opportuni stimatori basati sulla relazione empirica che lega i due aggregati. Se si suppone valida l'ipotesi [3.3] un primo stimatore del tutto intuitivo di I_t^{**} è dato da:

$$\left(\frac{P_{tc}}{P_{bc}} \right), \quad [3.5]$$

mentre uno stimatore che sfrutta l'errore di stima commesso al tempo ($t-k$) utilizzando lo stimatore [3.5] è dato dalla relazione:

$$I_t^{**} = \left(\frac{P_{tc}}{P_{t-k,c}} \right) I_{t-k}^{**}. \quad [3.6]$$

Il ricorso a metodi regressivi - ragionevolmente possibile solo in presenza di serie storiche sufficientemente lunghe ed ipotizzando che il legame intrinseco²¹ tra la struttura del *panel* e quella dell'intero universo si mantenga stabile nel tempo - rappresenta un caso particolare tra quelli discussi nel paragrafo 4.2 seguente.

²¹ Rendendosi necessarie progressive rotazioni delle unità del campione mensile dell'indagine sulle vendite al dettaglio (circa un terzo ogni anno) potrebbero insorgere discontinuità longitudinali indesiderate, anche qualora le principali caratteristiche strutturali delle unità sostitutive (settore di attività, dimensione, localizzazione geografica) fossero del tutto analoghe a quelli delle unità sostituite. Su tale aspetto si veda Gismondi (1998.3).

4. Tecniche di stima delle numerosità marginali

In questo paragrafo saranno illustrate alcune tecniche di stima del numero di operatori commerciali al dettaglio attivi al tempo t , dove il termine di “operatore” si riferirà in prevalenza al concetto di impresa, anche se molte delle procedure di cui in seguito sono facilmente adattabili al caso in cui si faccia riferimento al numero dei punti di vendita.

Nel contesto si parlerà di numerosità “marginali”, ossia di grandi aggregati sintetici. Ad esempio, il totale generale degli operatori attivi potrebbe rappresentare un vincolo marginale per la stima degli operatori alimentari e non alimentari, secondo quanto formalizzato nel paragrafo 5. Tale opzione è in realtà condizionale, perchè sulla base delle informazioni effettivamente disponibili tutte le procedure qui descritte sono applicabili per la stima di aggregati intermedi, che per somma consentono di riprodurre l’ammontare totale degli operatori.

4.1 Disponibilità di variabili ausiliarie e serie storiche brevi

Utilizzando le stesse simbologie del paragrafo precedente, si definisca questa nuova variabile:

C_t = valore a prezzi correnti dei consumi di beni²² al tempo t .

Consegue che sarà sempre valida questa identità:

$$C_t = (1 + \alpha_t)(V_t N_t),$$

dove i consumi saranno generalmente superiori al valore complessivo delle spese transitate per il canale commerciale al dettaglio (ammontare nella seconda parentesi), secondo un parametro $\alpha_t \geq 0$ dipendente dal tempo t . Se nel breve periodo tale parametro si suppone costante, si avrà che, per $k \geq b$:

$$\frac{C_t}{C_{t-k}} = \frac{(V_t N_t)}{(V_{t-k} N_{t-k})} = \frac{I_t}{I_{t-k}}$$

e quindi una stima del rapporto I_t^{**} / I_{t-k}^{**} tra i due indici relativi alla dinamica delle imprese tra i tempi t e $(t-k)$ sarà dato da:

$$\frac{\left(\frac{C_t}{C_{t-k}} \right)}{\left(\frac{I_t^*}{I_{t-k}^*} \right)} \quad [4.1.1]$$

dove ovviamente al denominatore si utilizzeranno le corrispondenti stime campionarie degli

²² Le spese per consumi di beni includono diverse voci non considerate dall’indagine sulle vendite, tra cui: le vendite effettuate nei punti vendita di imprese che svolgono un’attività prevalente diversa dal commercio al dettaglio, o che comunque appartengono ad imprese estere; le spese non transitanti per il canale al dettaglio; gli autoconsumi. Non è inoltre possibile scorporare completamente dalle spese per consumi alcune tipologie di beni non considerate dall’indagine sulle vendite (autoveicoli, carburanti, accessori per auto, combustibili).

indici a rapporto. Se è noto l'indice I_{t-k}^{**} relativo ad un periodo di partenza, moltiplicando tale indice per la stima [4.1.1] si otterranno via via tutti gli altri indici I_t^{**} necessari per il corretto calcolo dell'ammontare complessivo delle vendite.

E' poi immediato constatare come un criterio analogo è applicabile in presenza di una qualsivoglia variabile ausiliaria X nota, purchè ragionevolmente correlata alle vendite al dettaglio, con il parametro α libero da vincoli.

Un metodo alternativo consiste nel considerare il semplice rapporto C_t/V_t come una prima stima grezza del numero di operatori in essere al tempo t . Avendo a disposizione almeno una coppia di periodi di osservazione, uno stimatore più preciso di tale numerosità sarà ricavabile sulla base della relazione seguente, in cui si sfrutta la conoscenza della relazione tra numero degli operatori in essere e rapporto tra consumi e vendite medie per operatore in un periodo precedente:

$$\hat{N}_t = N_{t-1} \frac{\left(\frac{C_t}{V_t}\right)}{\left(\frac{C_{t-1}}{V_{t-1}}\right)} = N_{t-1} \frac{\left(\frac{C_t}{I_t^*}\right)}{\left(\frac{C_{t-1}}{I_{t-1}^*}\right)}. \quad [4.1.2]$$

Tale procedura, per quanto parzialmente euristica e meno caratterizzata da particolari proprietà statistiche, può risultare utile qualora non si disponga di una serie storica particolarmente lunga che consenta l'applicazione di criteri di stima più sofisticati.

Per un concreto utilizzo delle precedenti tecniche di stima si rammenta che l'ISTAT diffonde i Conti Economici Trimestrali con un ritardo di circa tre mesi dalla fine dell'ultimo trimestre di riferimento²³.

4.2 Disponibilità di variabili ausiliarie e serie storiche lunghe

Questa situazione implica di poter disporre di serie storiche di lunghezza tale da garantire una sufficiente precisione delle stime dei parametri necessarie per poter a sua volta stimare il numero degli operatori commerciali attivi al tempo t o la variazione di tale numero rispetto ad un periodo ad esso precedente.

In generale, la recente letteratura concernente la natimortalità delle imprese ha approfondito maggiormente il tema della natalità che quello della mortalità delle imprese, come evidenziato, ad esempio, in Guesnier (1994). Inoltre il campo di applicazione dei modelli proposti è stato soprattutto quello dell'industria, assai più di rado quello dei servizi. Per il caso italiano uno studio di ampia portata è quello di Garofoli (1994), in cui si propongono modelli regressivi basati soprattutto su variabili demografiche (popolazione residente, popolazione attiva, tasso di disoccupazione, quota dei lavoratori autonomi sul totale degli occupati).

In questo contesto si supporrà di non poter considerare metodi regressivi che sfruttano tali variabili esplicative, perchè non disponibili in chiave congiunturale, così come, più in generale, sembrano difficilmente implementabili modelli basati su variabili macroeconomiche (O'Farrell, 1994), anche se il secondo dei modelli enunciati in seguito farà comunque

²³ La diffusione avviene su supporto informatico, unitamente ad altre variabili macroeconomiche utilizzabili per implementare alcune delle procedure descritte nel paragrafo 4.2.

riferimento formale a questa situazione.

Poichè, quindi, la maggioranza dei modelli proposti in letteratura si basa su variabili endogene esse stesse non disponibili al momento di dover effettuare la stima, si è ritenuto opportuno proporre modelli certamente più semplicistici, ma maggiormente peculiari rispetto allo specifico comparto economico analizzato e, soprattutto, applicabili concretamente e quindi più utili per le finalità esposte nella premessa. Nel dettaglio, saranno illustrati (e sperimentati nel paragrafo 6) i seguenti quattro metodi di stima di tipo regressivo.

1) Modello regressivo generale

Un modello generale di regressione è dato ovviamente dalla relazione:

$$f(N_t) = \mathbf{a}'\mathbf{x} + \varepsilon_t \quad [4.2.1]$$

con ε residuo casuale, \mathbf{a} vettore colonna di coefficienti da determinare e \mathbf{x} vettore colonna di variabili esplicative correlabili alla funzione f del numero di operatori attivi al tempo t . In pratica si avrà che $f(N_t)=N_t$ o $f(N_t)=N_t/N_{t-1}$.

2) Primo modello basato sui dati del campione (modello “base”)

In questo primo modello si rinuncia ad utilizzare variabili esogene rispetto all’indagine sulle vendite, e si stima la variazione del numero degli operatori attivi intercorsa tra i tempi t e $(t-k)$ regredendo tale variabile sull’omologa variazione relativa a *tutte* le altre variabili quantitative osservate nel campione a cadenza mensile, ossia il numero di addetti, il numero dei punti di vendita e l’ammontare medio delle vendite. Se si definisce:

A_t = numero di addetti delle imprese commerciali al dettaglio al tempo t

per l’universo delle imprese il modello è quindi dato dalla relazione:

$$\frac{N_t}{N_{t-k}} = a_0 + a_1 \frac{P_t}{P_{t-k}} + a_2 \frac{A_t}{A_{t-k}} + a_3 \frac{V_t}{V_{t-k}} + \varepsilon_t \quad [4.2.2]$$

ed il modello relativo ai dati campionari è scrivibile aggiungendo il pedice “c” con riferimento alle variabili coinvolte²⁴. Nel caso di modelli mensili (trimestrali) la variabile V relativa al valore delle vendite dovrebbe essere preventivamente destagionalizzata, a meno che la scelta del periodo ritardato $(t-k)$ corrisponda allo stesso mese (trimestre) dell’anno precedente. Tale metodo è ovviamente interpretabile come un caso particolare del modello [4.2.1]. In pratica il suo utilizzo sarà possibile solo da quando sarà disponibile una serie storica sufficientemente lunga basata sui dati della nuova indagine sulle vendite, dato che con l’indagine precedente non venivano calcolati indici dell’occupazione.

3) Secondo modello basato sui dati del campione (di “trasferimento”)

²⁴ Tale accorgimento caratterizza anche i modelli 3 e 4 seguenti.

Questo modello si basa sull'ipotesi che la variazione relativa del numero degli operatori occorsa tra i tempi $(t-k)$ e t dipenda, oltre che da una costante, dalla differenza tra il tasso di variazione del valore medio delle vendite per gli operatori della grande distribuzione (pedice G) ed il tasso di variazione del valore medio delle vendite per gli operatori non appartenenti alla grande distribuzione (pedice \bar{G}), secondo una logica di "trasferimento" delle spese effettuate nei secondi a favore dei primi. In simboli, con riferimento all'universo:

$$\frac{N_t}{N_{t-k}} = a_0 + a_1 \left(\frac{V_{tG}}{V_{t-k,G}} - \frac{V_{t\bar{G}}}{V_{t-k,\bar{G}}} \right) + \varepsilon_t \quad [4.2.3]$$

Un valore maggiore (minore) di zero della quantità in parentesi tonde - ossia un tasso di crescita delle vendite medie per operatore della grande distribuzione superiore (inferiore) al tasso di crescita degli operatori non appartenenti alla grande distribuzione - dovrebbe accompagnarsi, in un periodo di sostanziale stabilità dei consumi reali come quello attuale, ad una flessione (ad un aumento) del tasso di crescita complessivo degli operatori, dovuto essenzialmente alla flessione (all'aumento) del numero degli operatori attivi nel comparto \bar{G} , per cui è lecito ipotizzare $a_1 < 0$ (> 0).

La validità del modello dipende strettamente dal ritardo con cui gli operatori si adeguano alle fluttuazioni della domanda (che gradatamente tende sempre di più a privilegiare le formule distributive moderne, soprattutto nel comparto alimentare), per cui è plausibile ipotizzare una versione della [4.2.3] in cui i valori medi delle vendite a numeratore siano riferiti ad un periodo ritardato rispetto a t . E' altresì possibile aggiungere la variabile esplicativa di tale modello a quelle già presenti nel modello [4.2.2].

4) Modello "di stabilità".

Si tratta di un modello misto, ossia basato su una variabile osservata nel campione ed una variabile macroeconomica. Trae origine dall'ipotesi che la variazione del valore medio delle vendite per operatore dipenda, oltre che da una costante, dal rapporto tra il valore di una variabile X nota, ad esempio i consumi, ed il numero degli operatori commerciali. Implicitamente si suppone che le spese delle famiglie transitino essenzialmente per il canale distributivo al dettaglio e che il numero degli operatori dipenda non tanto dal grado di diffusione delle grandi superfici, quanto dalle fluttuazioni delle abitudini di spesa della clientela finale. In simboli si avrà il modello seguente:

$$V_t = \alpha + \beta \frac{X_t}{N_t} + \varepsilon_t$$

con ε residuo casuale e α e β coefficienti da determinare. Dalla precedente relazione deriva la stima finale:

$$N_t \cong \frac{\beta X_t}{V_t - \alpha} \quad [4.2.4]$$

4.3 Indisponibilità di variabili ausiliarie e serie storiche ragionevolmente lunghe

Il metodo proposto in questo paragrafo è, ovviamente, solo uno dei possibili criteri che

cercano di sfruttare una serie storica a fini previsivi. Tale criterio si caratterizza per la sua semplicità ed il fatto di non richiedere, per poter essere implementato, la conoscenza di particolari informazioni ausiliarie rispetto a quanto generalmente disponibile da rilevazioni amministrative correnti e sulla base di un *panel* osservato periodicamente e rappresentativo - direttamente od indirettamente²⁵ - dell'intero universo degli operatori oggetto di interesse. Siano date queste definizioni:

- B_t = numero degli operatori nati tra i tempi $(t-1)$ e t ;
- D_t = numero degli operatori cessati tra i tempi $(t-1)$ e t ;
- n_t = numero degli operatori in essere al tempo t ed appartenenti ad un campione osservato;
- d_t = numero degli operatori cessati tra i tempi $(t-1)$ e t ed appartenenti ad un campione osservato al tempo $(t-1)$.

E' allora chiaro che varranno le due relazioni seguenti:

$$N_1 = N_0 + B_1 - D_1 \quad \text{e} \quad N_2 = N_1 + B_2 - D_2$$

e che si potrà scrivere:

$$\begin{aligned} N_2 &= N_1 \left(1 - \frac{D_2}{N_1}\right) + B_2 = N_1 \left(1 - \frac{D_2}{N_1}\right) + (N_2 - N_1 + D_2) = N_1 \left(1 - \frac{D_2}{N_1}\right) + \alpha (N_1 - N_0 + D_1) = \\ &= N_1 \left(1 - \frac{D_2}{N_1}\right) + \alpha \left[N_1 - N_0 \left(1 - \frac{D_1}{N_0}\right) \right] \end{aligned}$$

con $\alpha = \frac{B_2}{B_1}$. Uno stimatore di N_2 sarà allora dato da:

$$T_2 = N_1 \left(1 - \frac{d_2}{n_1}\right) + \hat{\alpha} \left[N_1 - N_0 \left(1 - \frac{d_1}{n_0}\right) \right].$$

mentre uno stimatore di N_t è dato da:

$$T_t = N_{t-1} \left(1 - \frac{d_t}{n_{t-1}}\right) + \hat{\alpha} \left[N_{t-1} - N_{t-2} \left(1 - \frac{d_{t-1}}{n_{t-2}}\right) \right].$$

Più in generale, se si suppone di disporre, al tempo t , di una serie storica composta da $(k-2)$ coppie di osservazioni, con $k > 2$, relative al numero di operatori in essere ed alla loro stima fornita dal precedente stimatore. La scelta dello stimatore $\hat{\alpha}$ è data dalla soluzione del problema di minimizzazione della funzione di perdita:

²⁵ Se, ad esempio, si dispone di un *panel* di imprese, esse forniranno informazioni *dirette* sulla dinamica longitudinale del numero di imprese attive ed *indirette* sulla dinamica longitudinale delle unità locali.

$$\sum_{h=t-k}^{t-3} (T_{h+2} - N_{h+2})^2.$$

Non è difficile ricavare la soluzione ottimale, data dalla relazione:

$$\hat{\alpha} = \frac{\sum_{h=t-k}^{t-3} N_{h+2} - \sum_{h=t-k}^{t-3} N_{h+1} \left(1 - \frac{d_{h+2}}{n_{h+1}}\right)}{\sum_{h=t-k}^{t-3} N_{h+1} - \sum_{h=t-k}^{t-3} N_h \left(1 - \frac{d_{h+1}}{n_h}\right)} \quad [4.3.1]$$

che può essere utilizzata per stimare il numero di operatori al tempo t .

Tale criterio, pur godendo di buone proprietà matematiche, può rivelarsi scarsamente preciso soprattutto nei contesti, come quello italiano, in cui alla dinamica - di per se complessa - del ciclo economico si sovrappongono fattori parzialmente esogeni come la variabilità delle normative amministrative - citata nel paragrafo 1 - e la difficoltà nel registrare le nuove aperture effettive (che non siano, cioè, la semplice continuazione sotto altra ragione sociale di un'impresa preesistente). La stima del parametro ottimale secondo la [4.3.1] risulta pertanto scarsamente elastica rispetto alla mutevolezza della congiuntura e rischia di incorporare troppo lentamente eventuali modificazioni strutturali del sistema produttivo.

Una ipotesi fondamentale ai fini della bontà della stima basata sulla [4.3.1] è che la probabilità di cessazione delle imprese incluse nel *panel* sia almeno approssimativamente uguale alla probabilità di cessazione di una generica impresa dell'universo. Tale ipotesi potrebbe non essere verificata soprattutto se i rispondenti effettivi del *panel* rappresentano una autoselezione rispetto all'universo originario, nel qual caso la loro probabilità di cessazione potrebbe posizionarsi su un livello medio più basso.

4.4 Indisponibilità di variabili ausiliarie e serie storiche brevi

4.4.1 Stima del livello

Si supponga di considerare due periodi consecutivi, indicati con i pedici 1 e 2, con riferimento ai quali è disponibile un sottoinsieme di misurazioni relative ad un particolare strato della popolazione di riferimento tramite il quale è possibile verificare l'eventuale stato di attività delle unità oggetto di studio. Quindi, se si fosse interessati alla stima del numero di imprese attive ai tempi 1 e 2 relativamente alla divisione 52 (commercio al dettaglio), si potrebbe supporre di conoscere l'ammontare di imprese attive in un particolare gruppo (ad esempio, il gruppo 52.2 delle imprese specializzate con vendita prevalente di alimentari e bevande), oppure delle imprese attive in una certa regione, o di un *panel* la cui mortalità interna - in termini di imprese che rimangono attive nel tempo - possa comunque essere utile per stimare l'ammontare complessivo delle imprese in essere²⁶. In ogni caso il sottoinsieme di osservazioni si supporrà disponibile in tempi assai ridotti rispetto a quelli necessari per poter stabilire l'effettiva consistenza di imprese attive nell'intero universo, per cui con riferimento ad esso si parlerà di osservazioni *preliminari*.

Riassumendo, si supporrà che tramite una misurazione preliminare, basata su un

²⁶ Quindi anche, indirettamente, delle imprese nuove nate.

sottoinsieme del totale delle unità statistiche da osservare, indicabile con il pedice P, siano disponibili gli ammontari preliminari N_{1P} e N_{2P} , riferiti rispettivamente ai tempi 1 e 2; solo successivamente - e comunque non più in tempo utile per eventuali elaborazioni quali il calcolo di indici dell'ammontare complessivo delle vendite - si potranno conoscere gli esatti ammontari finali della variabile oggetto di studio in entrambi i tempi, ossia N_1 e N_2 ²⁷.

Supponendo di trovarci al tempo 2, e di disporre sia delle stime preliminari che degli ammontari finali relativi al tempo 1, nonché della stima preliminare relativa allo stesso tempo 2, il problema consiste nel determinare uno stimatore dell'ammontare finale incognito relativo al tempo 2 che risulti "ottimale", sulla base delle informazioni disponibili, nonché delle ipotesi e delle specificazioni formali che saranno illustrate nel prosieguo.

Lo stimatore del livello corrente N_2 incognito sarà scrivibile in questa forma:

$$N_L = N_{2P} + \alpha(N_1 - N_{1P}) \quad [4.4.1.1]$$

dipendente dal parametro incognito α . In particolare, se $\alpha = 0$ si ha $N_L = N_{2P}$, ossia uno stimatore che utilizza solo le informazioni derivate dalle osservazioni preliminari, mentre se $\alpha = 1$ si ha $N_L = N_{2P} + (N_1 - N_{1P})$, e le eventuali distorsioni dei due stimatori dipendono dai momenti primi delle grandezze coinvolte. Il problema consiste nella determinazione di un valore *ottimale* per α che minimizzi la media quadratica dell'errore (*MQE*) sotto condizioni non particolarmente restrittive. In questo contesto gli operatori di media e varianza saranno valutati in funzione della variabilità delle osservazioni misurabili sulle medesime unità del campione su cui è valutato l'ammontare finale Y , supponendo di poter replicare tali osservazioni più volte nello stesso istante di riferimento: secondo lo stesso approccio utilizzato da Rao, Srinath e Quenneville (1989), la variabilità non riguarda dunque un disegno campionario, ma il modello di *superpopolazione* sottostante alle modalità di risposta delle singole unità. Si supponrà poi valida, d'ora in avanti, l'ipotesi seguente:

ipotesi 1:
$$E(N_2 - N_{2P}) = \gamma E(N_1 - N_{1P}) = \gamma \delta.$$

In altri termini, si suppone che al tempo 2 la sottostima (relativa al modello di risposta) dell'ammontare preliminare rispetto all'ammontare totale sia in media proporzionale all'omologa sottostima relativa al tempo 1. Cambiando la coppia di periodi successivi considerata (t e $t+1$) cambierà, in generale, anche il valore del parametro, per cui in effetti $\gamma = \gamma_t$ e la costanza di γ va dunque riferita, in assenza di ulteriori ipotesi, solo alla suddetta coppia di periodi in esame²⁸. *In primis*, γ verrà supposto libero da vincoli, anche se si vedrà in seguito come risulti auspicabile (e plausibile) l'ipotesi $0 \leq \gamma \leq 1$. Posto poi che le

²⁷ Ovviamente un caso particolare della situazione descritta è quello in cui si stia operando un censimento degli operatori attivi con riferimento ad un dato periodo, e dopo un certo intervallo di tempo intercorso dall'inizio della rilevazione risulterà ritornato un certo contingente di risposte "preliminari", generalmente relativo ad un *panel* naturale di rispondenti "rapidi".

²⁸ Ad esempio, se sulla precisione delle stime preliminari infraannuali influisse la stagionalità del fenomeno osservato, l'implementazione della stessa tecnica previsionale potrebbe condurre, a parità delle restanti condizioni, ad una distorsione in valore assoluto più elevata in alcuni periodi dell'anno per il solo effetto del diverso ordine di grandezza del fenomeno studiato. Il verificarsi di una diminuzione nel tempo del valore (assoluto) assunto dal parametro γ potrebbe interpretarsi come un progressivo miglioramento della precisione delle stime preliminari.

distorsioni dei valori finali N_1 e N_2 siano nulle²⁹, ossia che $B(N_1) = B(N_2) = 0$, segue dalla [4.4.1.1] che:

$$E(N_L) = E(N_{2P}) + \alpha E(N_1 - N_{1P}) = E(N_{2P}) + \alpha \delta$$

cioè:

$$|B(N_L)| = |E(N_{2P}) - E(N_2) + \alpha \delta| = |-\gamma \delta + \alpha \delta| = |\alpha - \gamma| \delta \quad [4.4.1.2]$$

e quindi la posizione $\alpha = \gamma$ minimizzerebbe, rendendola nulla, la distorsione dello stimatore del livello [4.4.1.1].

Privilegiando la scelta del valore ottimale di α nell'ottica della minimizzazione dell'errore quadratico medio, occorre ricavarne l'espressione formale e procedere alla sua minimizzazione con il metodo dei moltiplicatori di Lagrange³⁰. Calcolando l'espressione formale della varianza della [4.4.1.1] e ricordando che, come noto, $MQE(N_L) = V(N_L) + [B(N_L)]^2$, dove V è l'operatore varianza, si ottiene dopo alcuni passaggi:

$$\begin{aligned} MQE(N_L) &= V(N_{2P}) + \alpha^2 V(N_1 - N_{1P}) + 2\alpha COV[N_{2P}(N_1 - N_{1P})] = \\ &= V(N_{2P}) + \alpha^2 V(N_1) - \alpha^2 V(N_{1P}) + 2\alpha^2 COV(N_1; N_{1P}) + \\ &\quad \underline{+ 2\alpha COV(N_1; N_{2P}) - 2\alpha COV(N_{1P}; N_{2P})} + (\alpha - \gamma)^2 \delta^2 . \end{aligned}$$

Una semplificazione della formulazione precedente è garantita dalla seguente:

ipotesi 2: se $N_1 = N_{1P} + Z_1$ \rightarrow $COV(N_{1P}; Z_1) = COV(N_{2P}; Z_1) = 0$.

In altri termini si suppone di poter scomporre ogni ammontare incognito nella somma di due componenti, date dall'osservazione preliminare N_P e dalla componente non osservabile preliminarmente Z , mutuamente incorrelate indipendentemente dal periodo di riferimento (1 o 2). Conseguentemente si avrà che:

$$COV(N_{1P}; N_1) = V(N_{1P}) \quad \text{e} \quad COV(N_{2P}; N_1) = COV(N_{2P}; N_{1P}).$$

E' allora evidente come le due covarianze sottolineate nella precedente espressione dell'errore quadratico medio dello stimatore del livello N_L si annulleranno e si avrà più semplicemente:

$$MQE(N_L) = V(N_{2P}) + \alpha^2 V(N_1) - \alpha^2 V(N_{1P}) + (\alpha - \gamma)^2 \delta^2$$

da cui derivano facilmente gli errori quadratici medi corrispondenti ai casi particolari $\alpha=1$ e

²⁹ Il simbolo B indica la distorsione (BIAS).

³⁰ In ogni caso, si ribadisce come una stima di α basata solo sul confronto tra due periodi successivi invece che su un arco temporale più lungo comporta, in genere, una perdita di efficienza. D'altra parte, circa la frequente indisponibilità di serie storiche sufficientemente lunghe e comparabili si è già fatta menzione nella premessa: i motivi sono in genere strettamente inerenti al particolare campo di interesse (finalità stesse della rilevazione, cambiamenti nella tecnica di raccolta dei dati, variazione dei domini di studio, aspetti definitivi, eventuali slittamenti di base, cambiamenti nella tecnica di correzione dei dati anomali, ecc.). Comunque, occorre sempre valutare preliminarmente il *trade/off* esistente tra maggiore semplicità di implementazione di uno stimatore e la relativa perdita di precisione.

$\alpha=0$. La derivata prima della precedente espressione è pari a:

$$2\alpha V(N_1) - 2\alpha V(N_{1P}) + 2(\alpha - \gamma)\delta^2$$

che uguagliata a zero fornisce questa soluzione:

$$\alpha_L^* = \frac{\gamma \delta^2}{V(N_1) - V(N_{1P}) + \delta^2}.$$

Infine, non è difficile ricavare la conseguente formula per l'errore quadratico medio dello stimatore del livello:

$$MQE(N_L^*) = V(N_{2P}) + \gamma \delta^2 (\gamma - \alpha_L^*).$$

Si noti come, in pratica, risulti frequente il ricorso ad uno stimatore di più immediata implementazione basato sul metodo del quoziente³¹, dato da:

$$N_L = \frac{N_{2P}}{N_{1P}} N_1,$$

ottenibile ponendo nella [4.4.1.1] $\alpha_L = N_{2P}/N_{1P}$; una versione lievemente più sofisticata è ottenibile correggendo il precedente stimatore per tenere conto dell'errore di stima occorso tra i tempi 1 e 0, ossia calcolando:

$$N_L = \frac{N_{2P}}{N_{1P}} N_1 \frac{\left(\frac{N_1}{N_0}\right)}{\left(\frac{N_{1P}}{N_{0P}}\right)} \quad \text{ottenibile ponendo} \quad \alpha_L = \frac{N_{2P} \left[\left(\frac{N_1}{N_{1P}}\right)^2 \frac{N_{0P}}{N_0} - 1 \right]}{(N_1 - N_{1P})}.$$

4.4.2 Stima della variazione

Sulla base delle formalizzazioni e delle varie ipotesi introdotte nel paragrafo precedente si può così definire lo stimatore della variazione tra i due tempi, ossia della quantità incognita ($N_2 - N_1$):

$$N_V = N_L - N_1 \quad [4.4.2.1]$$

$$E(N_V) = E(N_L) - E(N_1) = E(N_{2P}) + \alpha\delta - E(N_1)$$

e quindi la distorsione di tale stimatore sarà data da:

³¹ Il ricorso a questa tecnica presuppone, in genere, una forte rappresentatività delle osservazioni preliminari rispetto al totale incognito, o comunque un legame longitudinale molto stabile tra le due componenti, il che è presumibilmente verificabile solo nel breve periodo. D'altra parte, l'uso del valore ottimale del parametro α non comporta necessariamente una distorsione minore rispetto a quella dello stimatore quoziente, peraltro utilizzato dall'INSEE (l'istituto di statistica francese) per la stima degli indici provvisori delle vendite commerciali al dettaglio.

$$|B(N_V)| = |E(N_V) - E(N_2) + E(N_1)| = |\alpha - \gamma| \delta. \quad [4.4.2.2]$$

Quindi lo stimatore della variazione tra i due tempi avrà la stessa distorsione dello stimatore del livello ed anche in questo caso la distorsione sarà nulla qualora $\alpha = \gamma$. Il calcolo dell'espressione formale della varianza dello stimatore [4.4.2.1], ottenibile sfruttando l'identità:

$$VAR(N_V) = VAR(N_L) + VAR(N_I) - 2COV(N_L, N_I)$$

è più complesso rispetto a quello relativo alla stima del livello, ma basandosi su ipotesi del tutto analoghe a quelle già introdotte viene per semplicità omissa. Dopo aver aggiunto a tale espressione della varianza il quadrato della distorsione e dopo aver calcolato ed uguagliato a zero la corrispondente derivata prima si ottiene, infine, il parametro α che rende minimo l'errore quadratico medio dello stimatore della variazione, dato da:

$$\alpha_V^* = 1 - \frac{(1-\gamma)\delta^2}{V(N_I) - V(N_{1P}) + \delta^2} = \alpha_L^* + \frac{V(N_I) - V(N_{1P})}{V(N_I) - V(N_{1P}) + \delta^2},$$

che quindi *non coincide* con il corrispondente peso ottimale per la stima del livello. La formula dell'errore sarà infine data da:

$$MQE(N_V^*) = [V(N_{2P}) + V(N_I) - 2COV(N_{1P}, N_{2P}) + \gamma^2 \delta^2] - [V(N_I) - V(N_{1P}) + \gamma \delta^2] \alpha_V^*.$$

Dalle espressioni formali degli errori quadratici medi dei due stimatori ottimali si deduce che, supposto $\gamma \cong 1$, se il quadrato della distorsione δ risultasse molto più elevato della differenza tra le varianze degli ammontari finale e preliminare del tempo 1, allora probabilmente $\alpha \cong 0$ costituirebbe la scelta ottimale in quanto oltre a rendere nulla la distorsione minimizzerebbe anche l'errore quadratico medio, sia per la stima del livello che della variazione. D'altra parte sia la crescita della distorsione δ che della varianza dell'ammontare incognito al tempo 1 relativo ai non rispondenti aumentano l'imprecisione di entrambe le stime.

Sulla base dei precedenti risultati va notato come la validità dell'ipotesi $0 \leq \alpha \leq 1$, sia per la stima del livello che della variazione equivale alle seguenti disuguaglianze algebriche:

$$0 \leq \alpha_L^* \leq 1 \leftrightarrow 0 \leq \gamma \leq 1 + \frac{V(N_I) - V(N_{1P})}{\delta^2} \quad 0 \leq \alpha_V^* \leq 1 \leftrightarrow -\frac{V(N_I) - V(N_{1P})}{\delta^2} \leq \gamma \leq 1$$

il cui soddisfacimento simultaneo è assicurato dall'ipotesi $0 \leq \gamma \leq 1$. Ciò equivale a supporre che, ad esempio, nel tempo le distorsioni relative alle stime preliminari mantengano lo stesso segno e tendano via via a ridursi, oppure che, indipendentemente dalla qualità delle stime preliminari, l'ammontare degli operatori attivi presenti un *trend* decrescente³².

³² I risultati di questo paragrafo e del precedente prescindono dal tipo di operatore analizzato, essendo applicabili sia alle imprese che ai punti di vendita. Se ad esempio si disponesse di una stima dell'ammontare dei punti vendita e si desiderasse una stima anche per l'ammontare delle imprese in attività, ponendo IM = numero di imprese attive, PV = numero dei punti vendita, si potranno stimare i parametri del modello $IM_t = a + b PV_t + \varepsilon_t$

4.4.3 Stima del livello con due periodi di osservazione precedenti

In questo caso si suppone di trovarci al tempo 2, di disporre sia delle stime preliminari che degli ammontari finali relativi ai tempi 1 e 0, nonché della stima preliminare relativa allo stesso tempo 2, e di dover determinare uno stimatore dell'ammontare finale incognito relativo al tempo 2 che risulti "ottimale", sulla base delle informazioni disponibili, nonché delle ipotesi e delle specificazioni formali seguenti. Lo stimatore del livello corrente N_2 incognito sarà scrivibile in questa forma:

$$N_L = N_{2P} + \alpha(N_1 - N_{1P}) + (1 - \alpha)(N_0 - N_{0P}) \quad [4.4.3.1]$$

ed occorre determinare il valore di α che minimizzi la media quadratica dell'errore sotto condizioni non particolarmente restrittive. Si supponrà valida l'ipotesi seguente:

ipotesi 1:
$$E(N_2 - N_{2P}) = \gamma E(N_1 - N_{1P}) = \gamma[\gamma E(N_0 - N_{0P})] = \gamma^2 \delta .$$

In altri termini, si suppone che al tempo 2 la sottostima (relativa al modello di risposta) dell'ammontare preliminare rispetto all'ammontare totale sia in media proporzionale all'omologa sottostima relativa al tempo 1 secondo lo stesso coefficiente di proporzionalità relativo al tempo 1. Posto poi che le distorsioni dei valori finali N_1 , N_2 e N_3 siano nulle, ossia che $B(N_1) = B(N_2) = B(N_3) = 0$, segue dalla [4.4.3.1] che:

$$E(N_L) = E(N_{2P}) + \alpha E(N_1 - N_{1P}) + (1 - \alpha)E(N_0 - N_{0P}) = E(N_{2P}) + \alpha\gamma\delta + (1 - \alpha)\delta$$

cioè:

$$|B(N_L)| = |E(N_{2P}) - E(N_2) + \alpha\gamma\delta + (1 - \alpha)\delta| = |-\gamma^2\delta + \alpha\gamma\delta + (1 - \alpha)\delta| = |-\gamma^2 + \alpha\gamma + (1 - \alpha)|\delta$$

Analogamente a quanto visto in precedenza si può supporre che:

ipotesi 2:

se $N_1 = N_{1P} + Z_1$	→	$COV(N_{1P}; Z_1) = COV(N_{2P}; Z_1) = 0.$
se $N_0 = N_{0P} + Z_0$	→	$COV(N_{0P}; Z_0) = COV(N_{1P}; Z_0) = 0.$

Conseguentemente si avrà che:

$$\begin{aligned} COV(N_{1P}; N_1) &= V(N_{1P}) & \text{e} & & COV(N_{2P}; N_1) &= COV(N_{2P}; N_{1P}); \\ COV(N_{0P}; N_0) &= V(N_{0P}) & \text{e} & & COV(N_{1P}; N_0) &= COV(N_{1P}; N_{0P}). \end{aligned}$$

Calcolando l'espressione formale dell'errore quadratico medio dello stimatore [4.4.3.1] si ottiene dopo alcuni passaggi:

$$\begin{aligned} MQE(N_L) &= \\ &= V(N_{2P}) + \alpha^2 V(N_1) - \alpha^2 V(N_{1P}) + (1 - \alpha)^2 V(N_0) - (1 - \alpha)^2 V(N_{0P}) + [-\gamma^2 + \alpha(\gamma - 1) + 1]^2 \delta^2. \end{aligned}$$

sulla base di serie storiche o basi di dati *cross-section*, generalmente disponibili dalle rilevazioni statistiche ufficiali.

La derivata prima della precedente espressione è pari a:

$$2\alpha V(N_1) - 2\alpha V(N_{1P}) - 2(1-\alpha)V(N_0) + 2(1-\alpha)V(N_{0P}) + 2(\gamma-1)[- \gamma^2 + \alpha(\gamma-1) + 1]\delta^2$$

che uguagliata a zero fornisce questa soluzione:

$$\alpha_L^{**} = \frac{[V(N_0) - V(N_{0P})] + (1-\gamma)^2(1+\gamma)\delta^2}{[V(N_1) - V(N_{1P})] + [V(N_0) - V(N_{0P})] + (1-\gamma)^2\delta^2}.$$

La principale differenza rispetto all'espressione ottimale precedente è che, in questo caso, se si verifica $\gamma \approx 1$ la stima dell'ammontare finale diventa *indipendente* dalla distorsione δ - a differenza di quanto visto con la situazione del paragrafo 4.4.3 - per cui la disponibilità di almeno due periodi di osservazione precedenti a quello in esame, che delinea una situazione piuttosto realistica, semplificherebbe notevolmente l'implementazione dello stimatore ottimale del livello³³. L'espressione dell'errore quadratico medio, omessa per semplicità, resterebbe comunque dipendente dalla distorsione.

4.4.4 Stima della variazione con due periodi di osservazione precedenti

Dalla espressione [4.4.2.1] dello stimatore della variazione si ha in questo caso che:

$$E(N_V) = E(N_L) - E(N_1) = E(N_{2P}) + \alpha\gamma\delta + (1-\alpha)\delta - E(N_1)$$

e quindi la distorsione di tale stimatore sarà data da:

$$|B(N_L)| = |E(N_V) - E(N_2) + E(N_1)| = |-\gamma^2 + \alpha\gamma + (1-\alpha)|\delta, \quad [4.4.4.1]$$

e risulterà dunque uguale alla distorsione dello stimatore del corrispondente stimatore del livello. Sfruttando poi le ipotesi 1 e 2 del paragrafo precedente, e ricordando che si è implicitamente supposto che la componente Z_t non osservabile preliminarmente sia incorrelata con ogni altra variabile riferita ad un tempo ($t \pm h$), si avrà:

$$\begin{aligned} MQE(N_L) &= \\ &= V(N_{2P}) + \alpha^2 V(N_1) - \alpha^2 V(N_{1P}) + (1-\alpha)^2 V(N_0) - (1-\alpha)^2 V(N_{0P}) + V(N_1) + \\ &\quad - 2COV(N_{1P}; N_{2P}) - 2\alpha V(N_1) + 2\alpha V(N_{1P}) + [-\gamma^2 + \alpha(\gamma-1) + 1]^2 \delta^2. \end{aligned}$$

Calcolando la derivata prima della precedente espressione ed uguagliando a zero si ricava infine la soluzione ottimale, diversa da quella relativa alla stima del livello:

³³ Se si disponesse di k periodi di osservazione precedenti a quello in esame la soluzione ottimale assumerebbe una forma più complessa, non esplicitabile direttamente tramite la risoluzione analitica. Per ulteriori dettagli su questo punto si rimanda, ad esempio, ad Harvey (1984).

$$\alpha_V^{**} = \frac{[V(N_1) - V(N_{1P})] + [V(N_0) - V(N_{0P})] + (1 - \gamma)^2 (1 + \gamma) \delta^2}{[V(N_1) - V(N_{1P})] + [V(N_0) - V(N_{0P})] + (1 - \gamma)^2 \delta^2}.$$

Come nel precedente caso relativo alla stima del livello, l'ipotesi $\gamma \cong 1$ implica l'indipendenza del parametro α dalla distorsione, e peraltro in questo caso (o nel caso di distorsione pressochè nulla) il peso da assegnare alla stima preliminare ritardata di un solo periodo tenderebbe ad uno, il che renderebbe superflua la disponibilità di coppie di osservazioni ritardate per oltre un periodo.

Il calcolo di tutti i coefficienti incogniti relativi alle espressioni degli stimatori ottimali e dei relativi errori visti nei quattro casi esaminati comporta una stima della distorsione, del parametro di scala e delle varianze degli ammontari preliminari e finali coinvolti nel modello adottato. Pur rimandando a Gismondi (1998.1) per maggiori dettagli su tale problema, alcune osservazioni parzialmente euristiche sembrano comunque necessarie.

In primo luogo, la scelta di porre $\alpha = 1$ sia per la stima del livello che della variazione - che renderebbe superflua la stima di qualsivoglia parametro - giustifica la sua applicabilità in quanto permette l'annullamento delle relative distorsioni nel caso in cui $\gamma \cong 1$, situazione probabilmente frequente nell'analisi di flussi animati da un *trend* scarsamente dinamico e da stabilità dei livelli medi degli ammontari in esame.

Altrimenti, le tecniche di stima più semplici da seguire nella pratica consisteranno:

1. in un primo caso, nell'assumere la distorsione, la varianza ed il parametro γ al tempo 2 uguali a quelle stimabili al tempo 1 (nel caso di stime trimestrali, si potrebbe fare riferimento ai valori dello stesso trimestre dell'anno precedente).
2. In un secondo caso, ricordando la prima delle ipotesi introdotte nel paragrafo 4.4.1, si potrebbe stimare il parametro di scala come rapporto tra gli ammontari (disponibili) degli scarti tra stime preliminari e finali ai tempi $(t-1)$ e $(t-2)$; supponendo la stabilità del parametro nel tempo, potrebbe anche essere utilizzata una media aritmetica ponderata dei precedenti rapporti, includendo fino a k coppie di periodi precedenti, con pesi proporzionali alla somma degli ammontari (preliminari, finali od entrambi) relativi ad ogni coppia di periodi. Per la stima della varianza si potrà supporre:

$$V(N_2) - V(N_{2P}) \cong \gamma^2 [V(N_1) - V(N_{1P})].$$

Inoltre, dovendo optare, nella pratica, tra la scelta ottimale del parametro α in funzione del livello o della variazione, è probabilmente preferibile ottimizzare, in un'ottica congiunturale, la stima della variazione, con una conseguente perdita assoluta di efficienza, la cui entità è valutata in Gismondi (1998.1).

5. Stima delle numerosità di secondo ordine

Come si è accennato in precedenza, spesso si opera in un contesto in cui risulta al più disponibile la numerosità degli operatori commerciali complessivamente in essere, senza alcun dettaglio settoriale. Tale numerosità N_t può essere assimilata alla frequenza marginale di una tabella in cui sulle righe sono riportati i vari tempi di riferimento e sulle colonne gli strati in cui gli operatori commerciali vengono classificati, dipendenti dal contesto della

ricerca.

Seguendo Pastacaldi e Vitali (1986), una formula generalizzata per stimare le frequenze intermedie N_{ti} di tale tabella, con i pedice di strato variabile da 1 a k , fissata la riga di riferimento t -ma e la corrispondente numerosità marginale, è data da:

$$T_{ti} = \frac{\pi_i^\beta (N_{t-1,i})^\alpha}{\sum_{i=1}^k \pi_i^\beta (N_{t-1,i})^\alpha} N_t$$

dove α e β sono due coefficienti da stimare, mentre i termini π_i sono dei “pesi” da attribuire alle frequenze, supposte note, riferite al periodo precedente, su cui si tornerà tra breve.

I due coefficienti suddetti rappresentano rispettivamente, al netto del termine di normalizzazione al denominatore, l’elasticità del numero di operatori al tempo t rispetto al corrispondente numero di operatori al tempo $(t-1)$, mentre ogni peso è assimilabile alla probabilità che la consistenza al tempo t sia superiore rispetto a quella del periodo $(t-1)$.

Va notato come la ricerca dei valori ottimali per α e β con l’imposizione di vincoli marginali comporta, in assenza di ulteriori ipotesi semplificatrici, il ricorso a metodi di soluzione iterativi: ne costituisce un noto esempio il criterio di massimizzazione dell’entropia, proposto da Wilson e Bennett (1985, 210-222). Viene altrove verificato (Gismondi, 1997) come l’uso di una funzione di perdita di tipo tradizionale, basata cioè sulla somma dei quadrati degli scarti tra i veri valori incogniti e le loro stime, condurrebbe ad una soluzione lineare nei valori riferiti al periodo precedente a quello osservato, ma non necessariamente soddisfacente la condizione di non negatività delle stime.

In tal senso, appare evidente fin d’ora come la bontà della procedura possa dipendere in modo determinante dalla corretta valutazione dei coefficienti suddetti, facilitata dalla conoscenza di una o più variabili correlabili con i flussi incogniti³⁴.

La formula precedente ammette diversi casi particolari di interesse.

1. Caso in cui $\alpha = 0$.

Si verifica immediatamente che la soluzione ottimale si riduce a:

$$T_{ti} = \frac{\pi_i^\beta}{\sum_{i=1}^k \pi_i^\beta} N_t$$

e quindi dipende dai valori noti del periodo precedente solo per il tramite dei pesi π_i . Se anche $\beta = 0$ consegue che:

$$T_{ti} = \frac{N_t}{k}$$

per cui ogni frequenza intermedia risulterebbe pari ad una costante e dunque del tutto indipendente dai valori del periodo precedente, a cui viene in tal modo assegnata una rilevanza informativa nulla.

³⁴ Su questo aspetto si rimanda all’applicazione proposta nel paragrafo 6.

2. Caso in cui $\alpha = 1$ e $\pi_i = \pi$ per ogni strato i .

In questo caso risulterà che:

$$T_{ti} = \left(\frac{N_{t-1,i}}{N_{t-1}} \right) N_t$$

e quindi le frequenze di strato sono stimate per semplice riproporzionamento delle frequenze del periodo precedente in funzione del rapporto tra frequenze marginali di colonna ai tempi t e $(t-1)$.

3. Caso in cui $\pi_i = \gamma N_{t-1,i}$, con γ parametro tale che $0 \leq \pi_i \leq 1$ per ogni origine i .

Ne deriva che:

$$T_{ti} = \frac{(N_{t-1,i})^{\alpha+\beta}}{\sum_{i=1}^k (N_{t-1,i})^{\alpha+\beta}} N_t$$

e dunque in questo caso le frequenze di strato sono stimabili in base ad un semplice riproporzionamento, rispetto alle frequenze marginali note, delle frequenze del periodo precedente, elevate ad una opportuna potenza. Se $\alpha + \beta = 1$ si riottiene la soluzione particolare del caso 2.

Per quanto riguarda la determinazione dei pesi π_i , a parità dei coefficienti α e β essi contribuiscono ad esaltare o smorzare l'effetto indotto dalla distribuzione della frequenza marginale tra i vari strati relativa al periodo precedente $(t-1)$.

In tal senso la scelta può avvenire tramite criteri soggettivi dipendenti dalle conoscenze a priori che si hanno sul fenomeno, oppure sulla base di previsioni circa l'andamento longitudinale del numero di operatori di ogni strato ottenibili con altri criteri di stima.. In concreto nel paragrafo seguente si utilizzeranno le posizioni:

$$\pi_i = \text{costante} \quad [5.1]$$

$$\pi_i = \frac{\left(\frac{N_{t-1,i} + N_{t-2,i}}{N_{t-1,i}} \right)}{\sum_{i=1}^k \left(\frac{N_{t-1,i} + N_{t-2,i}}{N_{t-1,i}} \right)}, \quad [5.2]$$

che corrispondono, se $\alpha = \beta = 1$, al precedente caso 2) ed allo stimatore:

$$T_{ti} = \left(\frac{N_{t-1,i} + N_{t-2,i}}{N_{t-1} + N_{t-2}} \right) N_t.$$

Un criterio generale per la scelta dei pesi consiste nel supporre una dipendenza funzionale tra ogni peso ed alcune variabili esogene di cui si ipotizza di poter disporre. Un possibile modello è il seguente:

$$\pi_i = f(I_{ti}^*; I_{t-1,i}^{**}; I_{t-1}^{**}),$$

per cui il peso π_i è posto in relazione con l'indice del valore medio delle vendite al tempo t (indice *panel*) e con gli indici del valore complessivo delle vendite del periodo precedente, sia complessivo che riferito allo stesso strato i considerato. Se si ipotizza un certo ristagno dei consumi reali (come sta accadendo in Italia ed in molti stati europei da alcuni anni), il peso dovrebbe risultare correlato inversamente con la prima variabile esplicativa e direttamente con la seconda e la terza.

6. Una applicazione

Sulla base delle elaborazioni delle Camere di Commercio relative alle autorizzazioni d'esercizio si dispone, per gli anni dal 1990 al 1996, di una stima degli indici di natimortalità delle imprese commerciali al dettaglio in sede fissa, come riportato nella tabella 2.3. Si suppone di voler stimare la consistenza di tali punti di vendita nell'anno A supponendo di conoscerne la consistenza nell'anno $(A-1)$ e, in alcuni casi, nell'anno $(A-2)$.

A tale fine sono stati confrontati tra loro i metodi di stima elencati nella tabella 6.2, sulla base dei dati disponibili riassunti nella precedente tabella 6.1. Va premesso che una verifica preliminare ha evidenziato come, in generale, la stima della consistenza del numero totale degli operatori risulterebbe più precisa se ottenuta per somma delle due stime parziali riferite ai comparti alimentare e non alimentare, sebbene tale aspetto non verrà approfondito in questo contesto, in cui sembra realistico poter disporre, in sede di calcolo della diffusione degli indici, di una stima dell'indice di natimortalità solo per il complesso delle imprese³⁵.

Tabella 6.1 – Dati di base per la comparazione di diversi metodi di stima

Anno	Autorizzazioni			Indici delle vendite I*					
	Totale	Alimentari	Indici totale	Totale	Tradizionali	Grande distribuzione (G.D.)	G.D. meno tradizionali	Indice dei consumi	Indice prezzi al consumo
1990	1.049.672	373.942	100,0	100,0	100,0	100,0	0,0	100,0	80,4
1991	1.048.549	365.828	99,9	104,7	103,6	110,0	6,4	109,6	85,0
1992	1.035.401	352.400	98,6	109,9	108,3	118,3	10,0	116,6	88,2
1993	1.016.598	340.881	96,8	111,4	109,3	122,4	13,1	119,7	91,6
1994	1.008.509	332.406	96,1	115,1	113,1	125,9	12,8	126,1	95,1
1995	995.532	322.841	94,8	119,8	117,1	133,7	16,6	134,1	100,0
1996	982.491	321.047	93,6	121,8	118,4	139,8	21,4	141,7	103,7

Fonte: elaborazioni su dati ISTAT.

³⁵ Per dettagli vedere Gismondi (1998.2).

Il primo criterio (metodo del quoziente), a cui non si è fatto riferimento esplicito nei paragrafi precedenti è il più semplice tra gli stimatori ipotizzabili supponendo di non disporre di variabili ausiliarie, è dato dalla relazione:

$$N_t = \left(\frac{N_{t-1}}{N_{t-2}} \right) N_{t-1}$$

e risulta, per ragioni comprensibili, assai impreciso.

Per l'applicazione del metodo regressivo 7, sono state usate le variabili esplicative: indice delle vendite di tipo *panel*, indice dei consumi ed indice dei prezzi (queste due ultime, esogene al campione, sono riconducibili al modello 4.2.1), mentre il metodo 8 ha aggiunto a tali variabili la differenza tra gli indici delle vendite di tipo *panel* della grande distribuzione e delle imprese tradizionali (effetto di sostituzione), su cui si basa interamente il metodo [4.2.3]. In proposito si rammenta che lo schema originario del modello [4.2.2] non è implementabile per la attuale indisponibilità di alcune variabili.

Tabella 6.2 – Stimatori ed errori medi di stima a confronto

Cod.	Tipo di stimatore	Errore medio di stima	Posizione in graduatoria
1	Metodo del quoziente	1,67	10
2	Metodo del quoziente corretto	1,23	8
3	Metodo basato sui consumi [4.1.1]	6,26	14
4	Metodo basato sui consumi [4.1.1] corretto	2,02	11
5	Rapporto tra consumi e valore medio delle vendite (C/V)	13,33	15
6	Metodo basato sui consumi [4.1.2]	3,52	13
7	Metodo regressivo [4.2.1]-[4.2.2]	0,21	2
8	Metodo regressivo [4.2.1]-[4.2.2] con effetto di sostituzione	0,12	1
9	Metodo regressivo di sostituzione [4.2.3]	0,73	5
10	Modello di stabilità [4.2.4]	1,21	9
11	Modello di stabilità [4.2.4] corretto	2,12	12
12	Metodo ottimale [4.3.1]	1,01	7
13	Metodo ottimale [4.3.1] corretto	0,80	6
14	Stimatore ottimale del livello	0,48	3
15	Stimatore ottimale della variazione	0,49	4

Fonte:elaborazioni su dati ISTAT.

Il modello utilizzato per la stima 9 (modello di sostituzione tra grande distribuzione ed imprese tradizionali) è dato dalla relazione:

$$\frac{N_t}{N_{1990}} = a_0 + a_1 \left(\frac{V_t^G}{V_{1990}^G} - \frac{V_t^{\bar{G}}}{V_{1990}^{\bar{G}}} \right) + \varepsilon_t$$

con un R^2 corretto pari a 0,9383 ed un buon livello di significatività (0,00093) ha fornito questa stima:

$$\frac{N_t}{N_{1990}} = 1,0253 - 0,0044 \left(\frac{V_t^G}{V_{1990}^G} - \frac{V_t^{\bar{G}}}{V_{1990}^{\bar{G}}} \right) + \varepsilon_t.$$

Il modello utilizzato per la stima 8 è invece dato dalla relazione:

$$\frac{N_t}{N_{1990}} = a_0 + a_1 \frac{V_t}{V_{1990}} + a_2 \frac{C_t}{C_{1990}} + a_3 \frac{PRE_t}{PRE_{1990}} + a_4 \left(\frac{V_t^G}{V_{1990}^G} - \frac{V_t^{\bar{G}}}{V_{1990}^{\bar{G}}} \right) + \varepsilon_t,$$

che in pratica congloba il modello 9, con un valore di R^2 corretto pari a 0,9744 ed un modesto livello di significatività, pari a 0,107, dovuto essenzialmente alla brevità della serie storica disponibile. Il modello ha fornito questa stima³⁶:

$$\frac{N_t}{N_{1990}} = 1,3132 - 0,0018 \frac{V_t}{V_{1990}} + 0,0029 \frac{C_t}{C_{1990}} - 0,0051 \frac{PRE_t}{PRE_{1990}} - 0,0019 \left(\frac{V_t^G}{V_{1990}^G} - \frac{V_t^{\bar{G}}}{V_{1990}^{\bar{G}}} \right) + \varepsilon_t$$

da cui emergono chiaramente la correlazione positiva con la dinamica dei consumi (coefficiente positivo a_2) e lo spiazzamento dei punti di vendita derivato dall'espansione delle vendite della grande distribuzione (coefficiente negativo a_4).

Per l'applicazione del modello di stabilità [4.2.4] (metodo 10 nella tabella 6.2) la variabile X è stata posta uguale ai consumi, e si è ottenuto un R^2 corretto pari a 0,9806 con un livello di significatività pari a 0,0001.

Riguardo al calcolo degli stimatori ottimali del livello e della variazione (con una osservazione ritardata, metodi 12 e 14 della tabella 6.2) introdotti nel paragrafo 4.4, si è supposto di disporre preliminarmente di una stima dell'indice di natimortalità delle imprese alimentari, sfruttato per stimare l'indice di natimortalità complessivo. I due criteri conducono a risultati coincidenti dato che i relativi valori ottimali di α sono praticamente identici (0,99601 per il livello e 0,99648 per la variazione); nel dettaglio, i parametri γ (pari a 0,99648) e δ (675.344) sono stati ottenuti per media degli omologhi valori ottenibili sulla della serie storica analizzata³⁷, mentre le varianze necessarie per implementare gli stimatori sono state calcolate sulle serie storiche dei valori N_1 (totale delle autorizzazioni) e N_{1P} (autorizzazioni alimentari).

Infine, come è evidente dalla stessa tabella 6.2, in molti casi si è verificata l'opportunità di "correggere" gli stimatori proposti sulla base dell'errore di stima registrato con riferimento all'anno precedente, applicando un semplice correttivo di tipo "quoziente".

L'indicatore di qualità della procedura di stima utilizzato è dato dalla media degli scarti percentuali, presi in valore assoluto, tra valori veri e stimati, ossia:

³⁶ Si ribadisce come la brevità della serie storica influisce sensibilmente sulla significatività statistica del modello. I livelli di significatività dei parametri stimati sono dunque omessi, perché in questo contesto la loro interpretazione risulterebbe problematica.

³⁷ In altri termini la distorsione è stata stimata per media degli scostamenti tra gli ammontari finali e preliminari (il totale delle autorizzazioni meno le sole autorizzazioni alimentari) relativi agli anni della serie storica analizzata, mentre il parametro di scala γ è stato ottenuto calcolando la media dei rapporti tra gli scostamenti suddetti di due anni consecutivi.

$$Err(\hat{N}_t) = \sum_{t=1991}^{1996} \left| \frac{\hat{N}_t - N_t}{6N_t} \right| 100$$

che risulterà quindi tanto maggiore quanto più elevata sarà la distanza media intercorrente tra valori effettivi e stimati. In alcuni casi il numero degli anni oggetto di stima è inferiore a 6, con ovvi adattamenti della formula precedente.

Nel complesso sono stati ottenuti questi risultati:

- se si escludono il criterio 5 (rapporto tra i consumi ed il valore medio delle vendite per impresa), il criterio 3 (formula 4.1.1 basata sui consumi) e l'ulteriore metodo [4.1.2] basato sui consumi (criterio 6), tutti i restanti criteri forniscono risultati soddisfacenti, con un errore di stima inferiore al 2% in valore assoluto (o di poco superiore nel caso del criterio 11, che peraltro corregge l'originario criterio 10 peggiorandone le *performance*).
- Lo stimatore più preciso è quello basato sul criterio 8, ossia uno dei metodi regressivi commentati poc'anzi, caratterizzato da un errore medio di stima dello 0,12%, seguito dal criterio 7 – basato su un altro metodo regressivo – con un errore medio dello 0,21%. Al terzo ed al quarto posto troviamo gli stimatori ottimali del livello e della variazione (criteri 14 e 15), con un errore medio pari, rispettivamente, allo 0,48% ed allo 0,49%.
- Tra i rimanenti, occupano una posizione intermedia gli altri metodi regressivi ed il metodo ottimale [4.3.1], sia nella sua forma originaria che corretto. Quindi, nel complesso si può concludere che gli errori più bassi si riscontrano per i metodi caratterizzati dalle proprietà matematiche maggiori, mentre tra i criteri più “euristici” il modello di stabilità [4.2.4] ed il metodo del quoziente corretto sembrano i più affidabili (gli errori medi sono pari, rispettivamente, all'1,21% ed all'1,23%).
- In generale, pur esistendo una chiara connessione tra la dinamica dei consumi e quella delle imprese commerciali, è possibile sfruttare tale legame in sede di stima della dinamica di quest'ultime solo ricorrendo a modelli regressivi.

Ovviamente tutti questi risultati dovrebbero essere verificati alla luce della disponibilità di serie storiche più lunghe di quelle effettivamente utilizzabili in questa sede.

Con riferimento al criterio della stima ottimale del livello e della variazione, nella tabella 6.3 sono riportati i risultati di una applicazione addizionale rispetto a quella già commentata. In questo caso si fa riferimento a dati trimestrali sul numero dei punti di vendita relativi al biennio 1995-1996, stimati sulla base delle informazioni diffuse con tale cadenza da Movimprese³⁸. Le osservazioni preliminari in tale contesto si riferiscono alla dinamica dei punti di vendita registrata trimestralmente nel campione dell'indagine sulle vendite condotta dall'ISTAT. Anche in questo caso i valori di α corrispondenti alla stima ottimale del livello e della variazione sono praticamente identici (0,99810 e 0,99815), e l'errore medio di stima è addirittura più basso rispetto a quello registrato con gli stimatori 14 e 15 della tabella 6,1 visti in precedenza (rispettivamente 0,397 e 0,395), nonostante la quota relativa di osservazioni preliminari rispetto al totale sia molto più modesta (meno del 4%). Tale risultato è quello più confortante, dato che implica la piena utilizzabilità dei dati campionari relativi alla dinamica

³⁸ Tali dati sono diffusi gratuitamente tramite Internet.

dei punti di vendita per poter correggere in tempo reale la stima degli indici dell'ammontare complessivo delle vendite.

Riguardo agli altri parametri necessari per implementare gli stimatori, ricavati con una procedura simile a quella vista per l'omologo caso precedente, si è ottenuto $\delta=549.614$, $\gamma=0,99815$, $V(N_1)=15.080.549$, $V(N_{1P})=114.377$.

Tabella 6.3 – Stimatori ottimali del livello e della variazione per dati trimestrali

Anno	Trimestre	Punti vendita attivi	Punti vendita campione	Stima del livello	Errore di stima %	Stima della variazione	Errore di stima %
1995	1	574.262	21.732				
	2	575.169	21.944	573.423	-0,30	573.450	-0,30
	3	574.792	22.271	574.443	-0,06	574.471	-0,06
	4	576.698	22.053	573.523	-0,55	573.550	-0,55
1996	1	568.693	22.324	575.913	1,27	575.941	1,27
	2	568.638	21.658	566.987	-0,29	567.014	-0,29
	3	566.790	21.489	567.428	0,11	567.455	0,12
	4	566.657	21.319	565.582	-0,19	565.609	-0,18
Media					0,397		0,395

Fonte: elaborazioni su dati ISTAT.

Per quanto concerne la stima delle numerosità di secondo ordine, supposte note le numerosità marginali complessive, si è utilizzato il modello generale del paragrafo 5 con $\alpha=\beta=1$ e con i pesi π selezionati con i criteri [5.1] e [5.2]. In questo caso per gli operatori «di secondo ordine» (tabella 6.4) ci si è riferiti alla consistenza alimentare e non alimentare, supposta nota la consistenza totale.

Tabella 6.4 – Stima della consistenza dei punti vendita alimentari e non alimentari

Anno	Valori veri			Stime con pesi dati dalla [5.1]				Stime con pesi dati dalla [5.2]			
	Totale	Non		Stime		Errori %		Stime		Errori %	
		Alimentari	Alimentari	Non alimentari	Non alimentari	Alimentari	Non alimentari	Alimentari	Non alimentari	Alimentari	Non alimentari
1990	1.049.672	373.942	675.730								
1991	1.048.549	365.828	682.721	373.542	675.007	2,11	-1,13	373.542	675.007	2,11	-1,13
1992	1.035.401	352.400	683.001	361.241	674.160	2,51	-1,29	365.051	670.350	3,59	-1,85
1993	1.016.598	340.881	675.717	346.000	670.598	1,50	-0,76	350.368	666.230	2,78	-1,40
1994	1.008.509	332.406	676.103	338.169	670.340	1,73	-0,85	340.731	667.778	2,50	-1,23
1995	995.532	322.841	672.691	328.129	667.403	1,64	-0,79	330.984	664.548	2,52	-1,21
1996	982.491	321.047	661.444	318.612	663.879	-0,76	0,37	321.238	661.253	0,06	-0,03
Media						1,71	0,86			2,26	1,14

I risultati sono parzialmente soddisfacenti, dato che il principio ispiratore della tecnica di stima è comunque un criterio di tipo quoziente che, come visto nel caso della stima del totale degli operatori, fornisce risultati buoni, sebbene non particolarmente esaltanti: l'errore medio è risultato pari all'1,71% per il comparto alimentare ed allo 0,86% per il non alimentare utilizzando pesi pari a 0,5 (quindi ininfluenti ai fini dei calcoli). La scelta dei pesi secondo la [5.2] non consente di pervenire a stime più precise, dato che l'errore medio di stima sale al 2,26% per il comparto alimentare ed all'1,14% per il non alimentare: ciò in quanto tale scelta finisce con enfatizzare eccessivamente le tendenze alla flessione che, pur essendo evidenti, non presentano picchi annuali particolarmente elevati.

Riferimenti bibliografici

- BARTLETT M.S. (1966), *An Introduction to Stochastic Processes*, London.
- BERNARDI A. (1997), "Dinamica delle vendite e dei punti vendita nel settore distributivo: analisi e prospettive", documento presentato nel seminario *La nuova indagine sulle vendite al dettaglio*, febbraio 1997, Istat, Roma.
- BIANCO M.-SESTITO M. (1992), "Entrata nel mercato e crescita delle imprese: una prima analisi del caso italiano", *paper* presentato alla *Conference on Birth and Start-up of Small Firms*, Università L.Bocconi, Centro studi sull'imprenditorialità F.Cicogna, 18-19 giugno 1992, Milano.
- CAMAGNI R. (1979), "La morfologia spaziale dei mercati al dettaglio", *Commercio*, anno I, 2, 63-105, Cescom, Milano.
- CONTINI B.-REVELLI R. (1986), "Natalità e mortalità delle imprese italiane: risultati preliminari e nuove prospettive di ricerca", *L'Industria*, VII, 2, aprile-giugno, 195-232.
- DIWERT E. (1995), "Axiomatic and Economic Approaches to Elementary Price Indexes", *Discussion Paper*, 1, University of British Columbia, Vancouver, Canada.
- EUROSTAT (1998.1), *The Draft Regulation on Short-Term Indicators*, Eurostat, Luxembourg.
- EUROSTAT (1998.2), *Definition of Variables in the Context of the STS Regulation*, room document D3/EBT/dec98/9en, Eurostat, Luxembourg.
- FAID (1998), "Evoluzione della società, dell'economia e dei consumi: internazionalizzazione e razionalizzazione della distribuzione italiana", *Notizie FAID*, Febbraio, Milano.
- FISHER I. (1922), *The Making of Index Numbers*, First Edition, Boston.
- INDIS (1998), *L'evoluzione del commercio attraverso i dati Siredi 1996/97*, Profili, 55, Unioncamere, Roma.
- GAROFOLI G. (1994), "La formazione di nuove imprese in Italia: analisi dei differenziali territoriali", in *Istituto Guglielmo Tagliacarne: Formazione di nuove imprese: un'analisi comparata a livello internazionale*, 62-99, Franco Angeli, Milano.
- GIRARDI U. (1997), "Nuove regole per la pianificazione commerciale", *Disciplina del commercio*, anno IV, 3, 1053-1067, Indis, Roma.
- GISMONDI R. (1997) "Un metodo generalizzato per la stima di flussi origine-destinazione", *Documenti*, 4, Istat, Roma.

- GISMONDI R. (1998.1) "Un metodo per la stima rapida di ammontari tramite risposte preliminari in presenza di errori di misura", *Rassegna del lavoro dell'ISCO*, Anno XV, 1, 73-107, Isco, Roma.
- GISMONDI R. (1998.2), "Alcuni criteri per la stima della dinamica demografica dei punti di vendita al dettaglio in Italia", *paper* presentato alla *Joint IASS/IAOS Conference on Statistics for Economic and Social Development*, INEGI, Aguascalientes, Mexico.
- GISMONDI R. (1998.3) "Modificazioni nella struttura delle imprese e conseguenze sulle tecniche di stima", mimeo predisposto per la *Commissione campioni*, Istat, Roma.
- GUESNIER B. (1994), "La creazione di imprese in Francia", in *Istituto Guglielmo Tagliacarne: Formazione di nuove imprese: un'analisi comparata a livello internazionale*, 100-139, Franco Angeli, Milano.
- INDIS (1998), *L'evoluzione del commercio attraverso i dati Siredi 1996/97*, Profili, 55, Unioncamere, Roma.
- ISTAT (Anni vari), *Statistiche del commercio interno*, Annuario, Istat, Roma.
- ISTAT (1996), *La media e grande impresa in Italia dal 1991 al 1994*, Argomenti, 4, Istat, Roma.
- ISTAT (1998.1), *Conti economici nazionali*, Informazioni, 73, Istat, Roma.
- ISTAT (1998.2), *La nuova indagine sulle vendite al dettaglio: aspetto metodologici e contenuti innovativi*, "Metode e norme", 3, Istat, Roma.
- ISTAT (1998.3), *Gli indici delle vendite al dettaglio nel 1997*, Informazioni, 85, Istat, Roma.
- ISTAT (1998.4), *Come cambia il commercio*, Argomenti, 13, Istat, Roma.
- MINISTERO DEL COMMERCIO, DELL'INDUSTRIA E DELL'ARTIGIANATO (1997), *Caratteri strutturali del sistema distributivo italiano al 1/1/1996*, Roma, Ministero del Commercio, dell'Industria e dell'Artigianato.
- MORRILL R.-GAILE G.L.-THRALL G.I. (1988), *Spatial diffusion*, Sage Publications, London.
- O'FARRELL P. (1994), "L'imprenditorialità in Irlanda: una rassegna dell'evidenza empirica", in *Istituto Guglielmo Tagliacarne: Formazione di nuove imprese: un'analisi comparata a livello internazionale*, 158-188, Franco Angeli, Milano.
- O'SULLIVAN P. (1984), *Economia e territorio*, Il Mulino, Bologna.
- PASTACALDI A.-VITALI L. (1986), "Modelli di interazione spaziale e di scelta del punto di vendita: il caso dei servizi nelle aree urbane", *Atti della XXXIII riunione scientifica SIS*, Cacucci, Bari.
- SANTARELLI E. (1997), "La relazione tra dimensione iniziale, sopravvivenza e crescita delle imprese nel settore turistico in Italia", *Statistica*, anno LVII, 1, 125-138, Bologna.
- SANTORO M.T. (1997), "Crescita e sopravvivenza delle imprese: metodi statistici a confronto", *Documenti*, 9, Istat, Roma.
- SMITH J.M. (1975), *L'ecologia e i suoi modelli*, Edizioni scientifiche e tecniche, Mondadori, Verona.
- SOLINAS G. (1992), "Le imprese nuove nate e la legge di Gibrat", *Economia & Lavoro*, vol.26, 1, 1992, 117-140, Roma.
- WILSON A.G.-BENNETT R.J. (1985), *Mathematical Methods in Human Geography and Planning*, John Wiley & Sons, Great Britain.