

n. 9/2010

La nuova rilevazione trimestrale del fatturato di alcune divisioni dell'informazione e comunicazione

A. Cirianni

DOCUMENTI ISTAT

n. 9/2010

**La nuova rilevazione trimestrale del fatturato
di alcune divisioni dell'informazione
e comunicazione**

A. Cirianni()*

(*) ISTAT – Direzione Centrale delle statistiche economiche congiunturali

Contributi e Documenti Istat

Istituto Nazionale di Statistica
Servizio Editoria – Centro stampa
Via Tuscolana, 1788 - 00173

La nuova rilevazione trimestrale del fatturato di alcune divisioni dell'informazione e comunicazione

Alfredo Cirianni, ISTAT, Direzione Centrale delle statistiche congiunturali, Servizio statistiche sulle attività dei servizi

Sommario: Il presente lavoro illustra la progettazione della nuova rilevazione trimestrale di fatturato della sezione J (informazione e comunicazione), che soddisfa le esigenze informative del Regolamento Comunitario delle statistiche congiunturali, con le modifiche apportate dall'introduzione della nuova classificazione delle attività economiche NACE REV 2. La complessa operazione di migrazione NACE è avvenuta in concomitanza con il cambio base, riferito all'anno 2005. La fase iniziale dello studio progettuale ha riguardato l'analisi della popolazione obiettivo al fine di scegliere il disegno campionario e la stratificazione più adatta a misurare la variabile obiettivo della rilevazione, costituita dalla stima campionaria degli indici di fatturato a base fissa. La fase successiva ha riguardato la scelta della strategia campionaria ed in seguito viene illustrata la tecnica di rilevazione impiegata nella fase della raccolta dati. L'ultima fase della progettazione riguarda la metodologia di controllo e di correzione degli errori non campionari. Il lavoro si conclude mettendo in evidenza alcuni problemi metodologici che sono stati incontrati e che potrebbero costituire i possibili sviluppi futuri per migliorare la qualità delle stime attualmente prodotte.

Parole chiave: panel, stimatore quoziente, mancate risposte, *outlier*, *backcasting*, *profiling*

Le collane esistenti presso l'ISTAT - Contributi e Documenti - costituiscono strumenti per promuovere e valorizzare l'attività di ricerca e per diffondere i risultati degli studi svolti, in materia di statistica ufficiale, all'interno dell'ISTAT e del Sistan, o da studiosi esterni.

I lavori pubblicati Contributi Istat vengono fatti circolare allo scopo di suscitare la discussione attorno ai risultati preliminari di ricerca in corso.

I Documenti Istat hanno lo scopo di fornire indicazioni circa le linee, i progressi ed i miglioramenti di prodotto e di processo che caratterizzano l'attività dell'Istituto.

I lavori pubblicati riflettono esclusivamente le opinioni degli autori e non impegnano la responsabilità dell'Istituto.

Indice

1. Introduzione.....	9
2. Analisi descrittiva della popolazione di riferimento, espressa in ATECO 2007, per lo studio del disegno campionario e della stratificazione.....	10
3. La strategia campionaria.....	14
3.1 Il disegno campionario.....	14
3.2 La scelta dello stimatore.....	18
3.3 Il sistema di ponderazione.....	21
4. La fase della raccolta dei dati.....	22
5. La metodologia di controllo e correzione degli errori non campionari.....	24
5.1 La casistica empirica degli errori non campionari.....	24
5.2 Il trattamento delle mancate risposte.....	25
5.3 Il trattamento degli outlier.....	29
5.4 L'effetto del processo di controllo e correzione sulle stime finali.....	31
6. Osservazioni conclusive e possibili sviluppi futuri.....	34
Appendice A – La procedura di ricostruzione retrospettiva delle serie storiche...	36
Appendice B – Il trattamento dei casi di <i>profiling</i>	39
Bibliografia.....	41

1. Introduzione^(*)

Il panorama delle statistiche congiunturali dell'attività dei servizi, e in particolare quelli relativi all'andamento del fatturato, è stato caratterizzato nel 2009 da notevoli cambiamenti, quale risultato del sovrapporsi di due ampi cambiamenti del sistema di riferimento delle statistiche economiche. Gli indicatori di fatturato, come tutte le altre statistiche congiunturali, sono state interessate, in Italia e più in generale nell'Ue, dal periodico cambiamento della base di riferimento, che è stata aggiornata al 2005. Come nel passato, l'aggiornamento della base implica l'adeguamento del sistema di ponderazione degli indici e un insieme articolato di operazioni volte ad adeguare gli indicatori ai mutamenti strutturali della sottostante popolazione di imprese alla cui attività essi si riferiscono. Tale processo, già di per sé occasione importante di mutamento metodologico e miglioramento qualitativo, si è intrecciato nel caso specifico delle basi 2005, con un altro mutamento di grande rilevanza del sistema di misurazione: l'introduzione di una nuova classificazione delle attività economiche (l'Ateco 2007, versione italiana della NACE REV 2) che modifica in profondità, per il settore dei servizi, la precedente nomenclatura Ateco 2002 (a sua volta corrispondente alla NACE REV 1.1).

La migrazione tra le due classificazioni ha comportato notevoli cambiamenti nella definizione dei settori dei servizi misurati dagli indicatori trimestrali del fatturato. Peraltro questi cambiamenti sono stati recepiti dal Regolamento Comunitario sulle statistiche di breve termine (definito comunemente regolamento STS) che definisce e omogenizza per i paesi dell'Ue le caratteristiche di questi come di tutti gli altri indicatori congiunturali. Nello specifico, sulla base della nuova ATECO, gli indicatori trimestrali di fatturato individuati dal regolamento STS (ISTAT, 2009¹; EUROSTAT, 2005²) sono divenuti 31, a fronte dei 25 richiesti in corrispondenza della vecchia classificazione.

Un'importante operazione introdotta dalla nuova classificazione è quella di aver creato una nuova sezione (che corrisponde al primo livello di disaggregazione della ATECO) che raggruppa tutti i servizi di informazione e comunicazione. Tale aggregato, la cui sigla è la J (sezione J è l'espressione sintetica che spesso si userà nel seguito) comprende appunto, tutte quelle attività che riguardano il trattamento dell'informazione, in particolare attraverso le nuove tecnologie IT, e che quindi costituiscono un insieme di grande dinamismo e rilevanza nelle economie avanzate. Nello specifico, la sezione J è composta dalle seguenti sei divisioni di attività economica:

- 1) editoria;
- 2) produzione cinematografica;
- 3) programmazione radio-televisiva;
- 4) telecomunicazioni;
- 5) servizi informatici e attività connesse;
- 6) servizi all'informazione e agenzie di stampa.

L'impatto della migrazione ATECO è stato molto rilevante all'interno di ciascuna divisione economica. In particolare, la divisione della nuova editoria è composta da una parte dell'editoria tradizionale proveniente dal settore dell'industria (per essa veniva prodotto un indice mensile con una differente tecnica e metodologia di indagine) e dall'edizione di *software*, proveniente dall'informatica. I settori del cinema e del *broadcasting* erano esclusi dal campo di osservazione dell'industria e servizi e non erano sottoposti a Regolamenti comunitari. Le telecomunicazioni non hanno subito sostanziali modifiche mentre le divisioni dei servizi informatici e dei servizi all'informazione provengono prevalentemente ma non esclusivamente³ dalla vecchia informatica: il processo di migrazione è stato particolarmente complesso per mancanza di corrispondenza biunivoca tra vecchia e nuova ATECO.

^(*) Si ringrazia il dottor Gian Paolo Oneto, direttore delle statistiche congiunturali, per i preziosi e fondamentali suggerimenti apportati nella stesura del documento.

¹ ISTAT *Nota informativa delle rilevazioni trimestrali di fatturato dei servizi*, comunicato stampa del 30 giugno 2009

² EUROSTAT *Regolamento n. 1158/2005*

³ In particolare, il vecchio gruppo della manutenzione di apparecchiature informatiche è migrato nell'industria, l'edizione di software nell'editoria e sono state incluse le agenzie di stampa nella divisione dei servizi all'informazione.

La complessa operazione di migrazione ATECO è avvenuta in concomitanza con il cambio della base⁴ ed ha comportato le seguenti operazioni:

- 1) riclassificazione dell'archivio ASIA in ATECO 2007 che permette di misurare la popolazione obiettivo sulla quale disegnare ed estrarre i campioni;
- 2) costruzione del sistema di ponderazione degli indici, basato sulle statistiche strutturali, in ATECO 2007;
- 3) progettazione ed estrazione dei campioni delle rilevazioni;
- 4) ricostruzione delle serie storiche retrospettive degli indici di fatturato.

Il presente lavoro focalizza l'attenzione sulla tecnica e la metodologia adottata per disegnare la nuova rilevazione trimestrale di fatturato di alcune divisioni economiche della sezione 'informazione e comunicazione' (editoria, servizi informatici e attività connesse, servizi all'informazione e agenzie di stampa, telecomunicazioni).

Nella prima parte, viene descritta la popolazione di riferimento, in ATECO 2007, al fine di definire la stratificazione e il disegno campionario più adatto a misurare il fenomeno economico.

Nella seconda parte, viene illustrata l'implementazione delle diverse fasi della strategia campionaria: scelta del disegno, dello stimatore e del sistema di ponderazione utilizzato per la stima degli indici di fatturato a base fissa.

La terza parte è dedicata ai principali aspetti della fase di raccolta dati: spedizione, struttura del questionario, strategia di prevenzione della mancata risposta totale, modalità di acquisizione dei dati.

La quarta parte riguarda il trattamento degli errori non campionari.

Il documento viene concluso con due appendici relative alla ricostruzione retrospettiva delle serie storiche ed al trattamento dei casi di *profiling* che interessano, in particolare, questi settori e che hanno un notevole impatto sulle stime longitudinali di fatturato.

2. Analisi descrittiva della popolazione di riferimento, espressa in ATECO 2007, per lo studio del disegno campionario e della stratificazione

Al fine di individuare una strategia campionaria, che assicuri una buona rappresentatività dei comportamenti della popolazione mantenendo contenuta la dimensione campionaria, è importante effettuare una analisi che colga le caratteristiche più rilevanti della popolazione oggetto di indagine.

Il primo obiettivo dell'analisi della popolazione è quello di individuare le variabili e le modalità di stratificazione da utilizzare nei passi successivi del disegno campionario. La stratificazione ha la finalità di partizionare la popolazione in modo da individuare strati omogenei al proprio interno (minimizzazione della varianza *within*) e nettamente distinti tra loro (massimizzazione della varianza *between*) in base alla variabile obiettivo dell'indagine, che, per queste rilevazioni, corrisponde al fatturato. Un possibile metodo di stratificazione è quello di utilizzare la *cluster analysis*, applicata sulle variabili fatturato e addetti, al fine di individuare strati omogenei e nettamente distinti tra loro (Ceccato, 2000⁵). Questo metodo, tuttavia, è più adatto per stime *cross section* in quanto, nel caso di stime longitudinali basate su panel ruotati, comporterebbe, in occasione del *refreshment*, una nuova identificazione dei *clusters* non confrontabili con quelli precedentemente utilizzati. In altre parole, la stratificazione, basata sulla *cluster analysis*, non è stabile nel tempo in quanto le modifiche della distribuzione delle variabili obiettivo, che si verificano in occasione dell'aggiornamento del panel dovuto al rilascio dell'archivio aggiornato, potrebbero comportare l'impossibilità di confrontare nel tempo *cluster* composti da differenti imprese, che, anche se presentano il requisito della longitudinalità negli anni, possono subire variazioni significative del fatturato e degli addetti che ne determinano una diversa allocazione nel tempo nei differenti *cluster*, dovuta alla mancanza di omogeneità con le unità che appartenevano al precedente

⁴ Il Regolamento sulle statistiche congiunturali chiede la produzione di numeri indice di fatturato a base fissa, con cambiamento della base ogni cinque anni per tener conto delle modifiche strutturali avvenute nei singoli settori. Gli indici finora prodotti in base 2000 sono stati aggiornati alla base 2005.

⁵ Ceccato F. (2000) *Studio progettuale della rilevazione trimestrale di fatturato del commercio e manutenzione auto*, documento interno ISTAT

gruppo. Di conseguenza, i frequenti cambiamenti di strato non permettono alle imprese interessate di contribuire alla stima degli indici elementari in quanto esse perdono il requisito della longitudinalità.

Per questo motivo, il metodo di stratificazione impiegato, correntemente utilizzato in ISTAT, è quello di definire la stratificazione in base alla classe dimensionale e all'attività economica in quanto si ipotizza che queste variabili economiche strutturali, che possiedono il requisito della stabilità nel tempo (Lavallée, 1997⁶), permettano di individuare strati omogenei al proprio interno e nettamente distinti tra loro in base alla variabile obiettivo della rilevazione (il fatturato). La classe dimensionale può essere definita in termini di addetti o fatturato. Si preferisce utilizzare il numero di addetti in quanto è una variabile molto più stabile nel tempo rispetto al fatturato, che, risentendo della dinamica congiunturale dei prezzi, potrebbe rilevare variazioni che, di fatto, non corrispondono effettivamente ad una variazione della dimensione aziendale. La scelta delle modalità delle classi di addetti dipende dalla struttura della popolazione e vengono selezionate in base alla loro significatività in termini di fatturato. Le modalità delle attività economiche sono scelte con criterio analogo.

Il secondo obiettivo dello studio della popolazione è quello di decidere il disegno campionario più appropriato⁷. In particolare, un disegno campionario casuale stratificato è ritenuto più adatto a settori caratterizzati da una moltitudine di piccole e medie imprese con caratteristiche strutturali simili mentre un campionamento *cut off* si adatta bene a settori oligopolistici con basso numero di imprese ed elevata concentrazione del fatturato. In questo ultimo caso, la parte non osservata del campione si considera trascurabile, ipotizzando che le piccole imprese in termini di fatturato hanno lo stesso comportamento delle grandi o che, anche se avessero una dinamica differente, il loro contributo alla stima dell'indice non risulterebbe statisticamente significativo.

Il terzo obiettivo dell'analisi della popolazione è quello di decidere gli strati censuari, in base al criterio della numerosità delle imprese e della quota di fatturato rappresentata. Se uno strato è costituito da un basso numero di imprese con alta quota di fatturato, esso viene censito.

In alternativa, la stratificazione e la scelta degli strati censuari potrebbe essere effettuata mediante metodi statistici molto più complessi (Lavallée P. et al, 1988⁸). In particolare, gli strati censuari potrebbero anche essere selezionati in base all'alta frazione di campionamento (ad esempio l'80 per cento). Questo metodo, tuttavia, tiene conto dell'allocatione delle unità degli strati che, secondo l'algoritmo di *Bethel* utilizzato ufficialmente dall'ISTAT, prevede un'alta frazione di campionamento nel caso di alta eterogeneità delle variabili obiettivo all'interno degli strati oltre che della numerosità della popolazione nello strato. Nel caso degli indici trimestrali di fatturato dei servizi, richiesti dal Regolamento STS, gli indici elementari di strato vengono ponderati con sistema di pesi di fatturato, riferito all'anno base, ed è, quindi, molto più importante tenere conto della quota di fatturato dello strato piuttosto che della frazione di campionamento.

La disponibilità dell'archivio ASIA, relativo all'anno 2005 in ATECO 2007 (ISTAT, 2008⁹), ha permesso di studiare la popolazione finita della sezione 'informazione e comunicazione' per le singole divisioni al fine di scegliere il disegno campionario più appropriato e, all'interno di ciascun dominio di pubblicazione, di partizionare la popolazione in strati omogenei.

La stratificazione teorica della sezione J, nei 6 domini di stima, è effettuata a livello di gruppo o di classe di attività economica a secondo delle singole divisioni e a livello di classe di addetti nelle tre modalità *standard* utilizzate per la vecchia informatica (1-19; 20-99; oltre 100 addetti), che sono risultate particolarmente efficaci anche per la sezione J.

⁶ Lavallée, P (1997), *Theory and application of enterprise panel surveys*, Course notes sponsored by the TES Institute, Luxembourg

⁷ Il regolamento Comunitario sulle statistiche congiunturali (STS) non impone nessuna tecnica di campionamento per la produzione degli indicatori trimestrali di fatturato degli altri servizi. Il manuale metodologico, invece, sottolinea l'importanza di aggiornare i panel con il rilascio dell'archivio (operazione di *refreshment*) e di ricorrere a campioni stratificati per ridurre la dimensione dei panel. La scelta della tecnica di campionamento, quindi, è lasciata alla discrezione dei ricercatori dei paesi membri.

⁸ Lavallée P. e Hidiroglou M. (1988) *On the stratification of skewed population*, Survey methodology, 14, pag. 33-44

⁹ ISTAT (2008): *Criteria for updating the business register*, in "Back casting STS time series according to the NACE REV 2 classification", GRANT per EUROSTAT

Il settore dell'editoria (tavola 1) è stato stratificato a livello di gruppo di attività economica (editoria tradizionale ed edizione di *software*) e di tre modalità di classe di addetti.

Le imprese con oltre 100 addetti coprono il 62,5 per cento del fatturato complessivo e sono 61, di cui 60 operanti nel settore dell'editoria tradizionale e solo una nel settore dell'edizione di *software*. La strategia è di censire tali imprese.

Tavola 1: Numero di imprese e composizione del fatturato per gruppo di attività economica e classe di addetti – settore dell'editoria (valori assoluti per numero di imprese e valori percentuali per composizione del fatturato)

Gruppo di attività economica	Classe di addetti			Totale
	1-19	20-99	Oltre 100	
<i>Numero di imprese</i>				
58.1 – Editoria tradizionale	6136	168	60	6364
58.2 - Edizione di <i>software</i>	364	27	1	392
Totale imprese	6500	195	61	6756
<i>Composizione percentuale del fatturato</i>				
58.1 – Editoria tradizionale	21,1	12,3	62,5	95,8
58.2 – Edizione di <i>software</i>	2,0	2,2	0,03	4,2
Totale fatturato	23,0	14,4	62,5	100

Fonte: ASIA, anno 2005

Il settore delle telecomunicazioni (tavola 2), composto da 5.316 imprese, è stato stratificato a livello di gruppo di attività economica (telecomunicazioni fisse; telecomunicazioni mobili; telecomunicazioni satellitari; altre attività di telecomunicazioni) e per classe di addetti. Le grandi imprese sono 29 e le medie sono 55; la maggior parte delle imprese (5.209) sono collocate nel settore residuale “altre attività di telecomunicazioni” e sono soprattutto di piccole dimensioni.

Le imprese con oltre 100 addetti coprono il 93,1 per cento del fatturato mentre le medie imprese il 4,7 per cento. L'ipotesi è di adottare per questo settore oligopolistico, con elevata concentrazione del fatturato in possesso di poche imprese, un campionamento di tipo *cut off*, stratificato a livello di gruppo di attività economica in quanto i gruppi delle telecomunicazioni fisse e mobili coprono la maggior parte del fatturato del settore (rispettivamente 47,2 per cento per telecomunicazioni fisse e 45,9 per cento per le telecomunicazioni mobili). Una ulteriore ipotesi sarebbe quella di ignorare la classe dimensionale in quanto si potrebbero verificare casi di imprese con elevato fatturato e pochi addetti che è rilevante includere nel campione.

Tavola 2: Distribuzione delle imprese e del fatturato per strato nella divisione delle telecomunicazioni (valori assoluti per numero di imprese e valori percentuali per fatturato)

Gruppo di attività economica	Classe di addetti			Totale
	1-19	20-99	Oltre 100	
<i>Numero di imprese</i>				
61.1 – telecomunicazioni fisse	58	20	14	92
61.2 – telecomunicazioni mobili	1	1	8	10
61.3 – telecomunicazioni satellitari	0	2	3	5
61.9 – altre attività di telecomunicazioni	5173	32	4	5209
Totale imprese	5232	55	29	5316
<i>Composizione percentuale del fatturato</i>				
61.1 – telecomunicazioni fisse	0,2	3,2	43,9	47,2
61.2 – telecomunicazioni mobili	0,0	0,0	45,9	45,9
61.3 – telecomunicazioni satellitari	0,0	0,1	2,3	2,4
61.9 - altre attività di telecomunicazioni	2,0	1,4	1,1	4,5
Totale fatturato	2,2	4,7	93,1	100

Fonte: ASIA, anno 2005

Tavola 3: Distribuzione delle imprese e del fatturato a livello di strato nella divisione “servizi informatici e attività connesse” (valori assoluti per numero di imprese e valori percentuali per composizione del fatturato)

Classe di attività economica	Classe di addetti			Totale
	1-19	20-99	Oltre 100	
<i>Numero di imprese</i>				
62.01 – Produzione di <i>software</i> non connesso all'edizione	19137	606	175	19918
62.02 - Consulenza nel settore delle tecnologie dell'informatica	9827	160	23	10010
62.03 – Gestione di strutture informatizzate	3026	94	9	3129
62.09 - Altre attività dei servizi connesse alle tecnologie dell'informatica	5742	41	1	5784
Totale imprese	37732	901	208	38841
<i>Composizione del fatturato</i>				
62.01 - Produzione di <i>software</i> non connesso all'edizione	15,5	12,9	44,1	72,4
62.02 - Consulenza nel settore delle tecnologie dell'informatica	5,9	2,7	3,8	12,4
62.03 – Gestione di strutture informatizzate	3,8	1,6	3,9	9,2
62.09 - Altre attività dei servizi connesse alle tecnologie dell'informatica	4,7	1,0	0,2	6,0
Totale fatturato	29,9	18,2	51,9	100

Fonte: ASIA, anno 2005

La divisione dei servizi informatici e attività connesse (tavola 3), costituita da 38.841 imprese, è partizionata per classe di attività economica (produzione di *software* non connesso all'edizione; consulenza nel settore delle tecnologie dell'informatica; gestione di strutture informatizzate; altre attività dei servizi connesse alle tecnologie dell'informatica) e per classe di addetti. Le grandi imprese sono 208. Data la struttura del settore, si può ipotizzare di censire le grandi imprese, che coprono il 51,9 per cento del fatturato complessivo, e di adottare per le piccole e medie imprese, che hanno un peso non trascurabile, uno schema campionario di tipo casuale stratificato.

Anche la popolazione dei servizi di informazione e agenzie di stampa (tavola 4), composta da 37.229 imprese, è stata partizionata a livello di classe economica (elaborazione dei dati, *hosting* e attività connesse; portali *web*; attività delle agenzie di stampa; altre attività dei servizi di informazione n.c.a.) e di classe di addetti.

Le grandi imprese sono 69 e le medie imprese sono 460; la maggior parte delle imprese (36.540 unità) è collocata nella classe “elaborazione dei dati, *hosting* e attività connesse” (che peraltro comprende anche l'attività delle banche dati).

Tavola 4: Distribuzione delle imprese e del fatturato per strato della divisione “attività dei servizi d'informazione e agenzie di stampa” (valori assoluti per numero di imprese e valori percentuali per composizione del fatturato)

Classe di attività economica	Classe di addetti			Totale
	1-19	20-99	Oltre 100	
<i>Numero di imprese</i>				
63.11- elaborazione dei dati, <i>hosting</i> e attività connesse	36026	450	64	36540
63.12 - Portali <i>web</i>	651	8	4	663
63.91 - Attività delle agenzie di stampa	15	0	0	15
63.99 - Altre attività dei servizi di informazione n.c.a.	8	2	1	11
Totale imprese	36700	460	69	37229
<i>Composizione del fatturato</i>				
63.11- elaborazione dei dati, <i>hosting</i> e attività connesse	54,6	13,8	28,4	96,8
63.12 - Portali <i>web</i>	0,9	0,5	1,3	2,7
63.91 - Attività delle agenzie di stampa	0,0	0,0	0,0	0,0
63.99 - Altre attività dei servizi di informazione n.c.a.	0,2	0,1	0,2	0,4
Totale fatturato	55,7	14,4	29,9	100,0

Fonte: ASIA, anno 2005

Le grandi imprese coprono il 29,9 per cento del fatturato e le medie imprese il 14,4 per cento: la maggior parte del fatturato (55,7 per cento) è coperto dalle piccole unità. Anche in questo caso sembra necessario procedere attraverso uno schema di campionamento che consideri sia le medie sia le piccole imprese, censendo le unità con oltre 100 addetti.

Per entrambe le divisioni economiche (J62 e J63), che provenivano dalla vecchia informatica, il comportamento delle piccole imprese era molto differente da quello delle grandi imprese in quanto le piccole unità lavoravano su commessa per le grandi imprese. Anche per questo motivo, oltre che per il peso in termini di fatturato, si è ritenuto di includerle nel campione perché non poteva essere considerata realistica l'ipotesi che avessero la stessa dinamica delle grandi imprese censite.

3. La strategia campionaria

La strategia campionaria consiste nella scelta del disegno e dello stimatore impiegato per il calcolo degli indici elementari, a livello di strato, e, nel caso specifico delle rilevazioni trimestrali di fatturato dei servizi, anche nella scelta del sistema di ponderazione per il calcolo dell'indice di dominio a base fissa. Il sistema di ponderazione, nel caso degli indici a base fissa calcolati su campioni stratificati, implicitamente, permette di effettuare un riporto all'universo per la stima della dinamica della variabile di interesse sulla base dell'ipotesi che la struttura produttiva della popolazione finita, in termini di demografia e cambiamenti strutturali, muti molto lentamente nel tempo.

3.1. Il disegno campionario

La progettazione del disegno campionario, adottato per queste quattro divisioni economiche, può essere così articolata:

- 1) scelta del disegno (probabilistico o *cut off*), in base all'analisi della popolazione obiettivo effettuata nel paragrafo precedente;
- 2) progettazione del campione probabilistico;
- 3) progettazione del campione *cut off*, adottato per il settore oligopolistico delle telecomunicazioni.

Con riferimento al punto 1, si ricorda che il campionamento probabilistico è stato adottato per i servizi informatici e all'informazione e per l'editoria mentre il campionamento *cut off* è stato impiegato per il settore oligopolistico delle telecomunicazioni.

In particolare, per quanto concerne il settore dell'editoria, nella migrazione dall'industria ai servizi, si è deciso di utilizzare un disegno casuale stratificato, che risulta coerente con le sue caratteristiche strutturali: alta numerosità di piccole e medie imprese con caratteristiche simili e forte concentrazione del fatturato nello strato censuario delle unità con 100 addetti. Tale scelta permette di tenere conto, oltre che del gruppo dell'edizione di *software*, anche delle piccole imprese (appartenenti alla classe di addetti 1-19) che assumono un peso non trascurabile in termini di fatturato. E' da notare che all'interno dell'indagine mensile sul fatturato industriale in base 2000, questo settore era trattato, coerentemente con l'approccio utilizzato per tutti i gruppi dell'industria, con una tecnica di campionamento *cut-off*. Per assicurare una maggiore continuità, la strategia adottata è stata quella di conservare nel nuovo campione le unità provenienti dal panel della rilevazione mensile del fatturato e ordinativi dell'industria.

Anche per le divisioni dei servizi informatici e servizi all'informazione, sono state incluse nel nuovo campione le unità che già venivano intervistate nella rilevazione trimestrale della vecchia informatica e che risultavano eleggibili dopo l'operazione di riclassificazione ATECO.

Per le telecomunicazioni, l'approccio di campionamento *cut off* è stato adottato a livello di gruppo di attività economica. Una parte delle imprese così selezionate già apparteneva al vecchio campione della rilevazione trimestrale delle telecomunicazioni; quest'ultimo comprendeva anche delle unità che nel tempo non sono più risultate significative e che, quindi, sono state escluse.

Dopo aver riassunto i motivi che hanno indotto a utilizzare disegni differenti per le 4 divisioni economiche della J, viene illustrata la progettazione del campionamento probabilistico (punto 2) e di quello *cut off* (punto 3).

Il campione casuale stratificato dell'editoria, dei servizi informatici e attività connesse e dei servizi all'informazione e agenzie di stampa è stato disegnato applicando il metodo di *Bethel*¹⁰ mediante il software MAUSS (Falorsi *et al.*, 1998¹¹) utilizzando come variabili di interesse il fatturato e gli addetti.

Prima di giungere alla scelta della dimensione campionaria e dell'allocazione delle unità all'interno degli strati si è operato un preliminare pre-trattamento dei dati di fatturato, presenti nell'archivio ASIA 2005, al fine di evitare che l'errore campionario commesso sia troppo influenzato dalle unità presenti nelle code della distribuzione, che si ipotizza siano affette da errori di compilazione o di processo¹².

I metodi impiegati per il pre-trattamento dei dati sono stati differenti per la popolazione longitudinale (ovvero le imprese presenti nell'universo sia nell'anno 2005 che in quello 2004) e per quella costituita dalle imprese nuove nate o che, anche per effetto della riclassificazione, provengono da settori diversi rispetto a quelli indagati.

Tavola 5: Dimensione del campione teorico, numerosità della popolazione e numero di unità da censire per divisione economica – Editoria, informatica e servizi all'informazione (valori assoluti)

Divisione	Numerosità campione	Numerosità Popolazione	Unità censuarie
58 – Editoria	278	6756	61
62 – Servizi informatici	565	38841	208
63 – Servizi di informazione	746	37229	71
Totale imprese	1589	82826	340

Fonte: ASIA provvisorio, anno 2005

La popolazione longitudinale, costituita dalle unità presenti in ASIA nel biennio 2004-2005, è stata sottoposta a una procedura di individuazione degli *outlier*, basata sul metodo dei residui *studentizzati*¹³ utilizzando il volume d'affari del 2005 come variabile dipendente e quello del 2004 come variabile ausiliaria. Gli *outlier* individuati da tale modello sono risultati pari allo 0,9 per cento delle unità indagate.

Le imprese nuove entrate sono pari al 14 per cento delle unità della popolazione. Per tali unità l'individuazione delle imprese anomale è basata sull'analisi della distribuzione del livello di fatturato. Le imprese, da escludere dai parametri per l'applicazione del metodo di *Bethel*, sono state individuate applicando il metodo di *Hidiroglou-Berthelot* (HB); le soglie di accettazione sono state rilassate fino a individuare una quota contenuta di *outlier* (5,1 per cento): il parametro 'c' dell'algoritmo di HB è stato posto pari a 7 per definire sia l'estremo inferiore che l'estremo superiore (*Hidiroglou M.A., Berthelot J.M., 1986*¹⁴).

¹⁰ Il metodo di *Bethel* è una estensione al caso multivariato del metodo di *Neyman*, che, nella fase di allocazione ottimale delle unità all'interno degli strati, considera l'incidenza della varianza delle variabili obiettivo e del numero delle imprese dello strato rispetto a quello del dominio di pubblicazione, sul quale viene calcolato l'errore campionario desiderato, misurato in termini di coefficiente di variazione.

¹¹ Falorsi P.D., Ballin M., De Vitiis C., Scepi G. (1998) *Principi e metodi del software generalizzato per la definizione del disegno di campionamento nelle indagini sulle imprese condotte dall'ISTAT*, Statistica applicata, vol. 10 n. 2, pag. 235-257

¹² La distribuzione del volume d'affari della popolazione è nota e, quindi, il pre-trattamento dei dati, finalizzato all'individuazione e trattamento degli *outlier*, ha come unico obiettivo l'individuazione di unità anomale non rappresentative in quanto affette da errori di misura e di processo. Siccome si tratta di una variabile di fonte amministrativa, non si può verificare se l'unità sospetta è effettivamente affetta da errori di compilazione e, quindi, si può solo ipotizzare che queste unità siano, con alta probabilità, errate e, incidendo notevolmente sulla dimensione campionaria, è preferibile escluderle nella fase della definizione della dimensione del campione teorico.

¹³ Il residuo è dato dalla differenza tra valore effettivo e quello stimato da modello di regressione. Se il modello è ben specificato (ossia ha un elevato R² e un basso *standard error* relativo al coefficiente di regressione), le unità con un elevato valore del residuo sono quelle che hanno un comportamento anomalo rispetto a quello delle imprese omogenee appartenenti allo stesso strato. Per rendere confrontabili le unità dello strato, è necessario ricorrere ai residui *studentizzati* che permettono di usare una unità di misura omogenea per tutte le imprese. I testi consigliano di considerare anomale le unità con valore assoluto dei residui *studentizzati* superiore a 2 o 5.

¹⁴ Hidiroglou M.A., Berthelot J.M. (1986), *Statistical editing and imputation for periodic Business Surveys*, Survey Methodology, 12, pag. 73-84. Il metodo, che verrà spiegato in modo approfondito nel paragrafo relativo

Tavola 6: Schema del campionamento probabilistico dell'editoria, dei servizi informatici e all'informazione (valori assoluti)^(a)

Strato	Asia 2005	Campione teorico di bethel (A)	Inf 08 (B)	Editoria (C)	Tlc (D)	ing. (E)	Somma dei panel esistenti (F=B+C+D+E)	Nuove entrate (G=A-F)	sovrapmpiona te (H=F-A)	Campione effettivo (A+H)
1_581	6136	171	0	5	0	0	5	166	0	171
1_582	364	8	5	0	0	0	5	3	0	8
2_581	168	33	0	15	0	0	15	18	0	33
2_582	27	5	4	0	1	0	5	0	0	5
3_581	60	60	0	35	0	0	35	25	0	60
3_582	1	1	1	0	0	0	1	0	0	1
1_6201	19137	185	136	0	2	0	138	47	0	185
1_6202	9827	48	111	0	2	0	113	0	65	113
1_6203	3026	23	26	0	0	0	26	0	3	26
1_6209	5742	34	39	0	0	1	40	0	6	40
2_6201	606	45	83	0	2	0	85	0	40	85
2_6202	160	9	24	0	1	1	26	0	17	26
2_6203	94	6	13	0	0	0	13	0	7	13
2_6209	41	7	8	0	0	0	8	0	1	8
3_6201	175	175	131	1	0	0	132	43	0	175
3_6202	23	23	17	0	0	0	17	6	0	23
3_6203	9	9	6	0	1	0	7	2	0	9
3_6209	1	1	1	0	0	0	1	0	0	1
1_6311	36026	565	269	0	3	1	273	292	0	565
1_6312	651	11	8	0	0	0	8	3	0	11
1_6391	15	2	0	0	0	0	0	2	0	2
1_6399	8	3	1	0	0	0	1	2	0	3
2_6311	450	92	83	0	0	1	84	8	0	92
2_6312	8	2	2	0	1	0	3	0	1	3
2_6399	2	2	0	0	0	0	0	2	0	2
3_6311	64	64	50	0	0	0	50	14	0	64
3_6312	4	4	3	0	0	0	3	1	0	4
3_6399	1	1	0	0	0	0	0	1	0	1
Totale	82.826	1.589	1.021	56	13	4	1.094	635	140	1.729

Fonte: ISTAT, archivio ASIA, rilevazione mensile del fatturato e ordinativi dell'industria e rilevazione trimestrale del fatturato degli altri servizi

(a) La prima colonna indica lo strato definito dalla combinazione della classe di addetti (le tre modalità – 1,2,3 – indicano rispettivamente le piccole, medie e grandi imprese) e dell'attività economica (le modalità sono espresse in base al codice ATECO 2007). La seconda colonna indica la numerosità della popolazione misurata dall'archivio ASIA, riferito all'anno 2005. La terza colonna indica la numerosità del campione teorico definito dall'applicazione di Bethel e in bordo oscurato le imprese censuarie. Le colonne successive indicano la numerosità dei panel esistenti dell'informatica (INF), dell'editoria, delle telecomunicazioni (TLC) e dell'ingrosso (ING) e della loro somma per strato. La colonna delle imprese nuove entrate indica la differenza tra campione teorico e somma delle imprese appartenenti ai panel esistenti. Le imprese sovrapcampionate sono quelle caratterizzate da una numerosità delle imprese dei panel esistenti maggiore di quelle richieste dal campione teorico e in bordo tratteggiato sono indicati gli strati sovrapcampionati. Il campione effettivo è la somma delle imprese del campione teorico e di quello delle imprese sovrapcampionate.

all'individuazione e al trattamento degli *outlier*, si basa sostanzialmente sull'individuazione di una differenza interquantilica, applicata sul valore del fatturato, opportunamente simmetrizzato. L'intervallo di accettazione può cambiare in base ad un parametro c , che può essere differente per definire l'estremo inferiore e superiore al di fuori del quale considerare anomale le unità. Tale parametro viene posto uguale per entrambe le code e l'intervallo viene definito a livello di strato.

Il pre-trattamento dei dati ha permesso di ottenere una dimensione campionaria teorica contenuta (tavola 5) pari a 1.589 imprese, con un errore di stima desiderato sul fatturato, misurato in termini di coefficiente di variazione, pari al 5 per cento per ciascuna divisione economica.

La numerosità del campione teorico (tavola 5) è pari a 278 imprese su un totale di 6.756 unità per la divisione dell'editoria, con 61 imprese da censire. Il campione teorico è pari a 565 imprese su una popolazione di 38.841 unità per la divisione dei servizi informatici, con 208 imprese da censire. Il campione teorico è pari a 746 imprese su una popolazione di 37.229 unità della divisione dei servizi all'informazione e agenzie di stampa, con 71 imprese da censire.

Nella definizione del campione effettivo la scelta è stata quella di includere le unità eleggibili che appartenevano ai panel delle rilevazioni trimestrali di fatturato dell'informatica, del commercio all'ingrosso (per quanto riguarda le imprese che erano classificate nel commercio all'ingrosso di apparecchiature informatiche), delle telecomunicazioni (in quanto settore affine all'informatica) e della rilevazione mensile del fatturato e degli ordinativi dell'industria (comprensiva dell'editoria tradizionale).

La strategia di includere imprese già presenti nei panel ed appartenenti a strati campionari, su cui si discuterà nel paragrafo relativo alla fase della raccolta delle informazioni, è motivata dal fatto che tali imprese hanno un differente comportamento di risposta rispetto alle imprese "nuove entrate" e ciò assicura elevati tassi di risposta.

Le imprese "nuove entrate", quindi, sono costituite dalla differenza, a livello di strato, tra campione teorico e le imprese appartenenti ai 4 panel delle predette rilevazioni. Qualora la somma delle imprese dei quattro panel è risultata superiore al campione teorico, la strategia è stata quella di sovracampionare tali strati. Le unità nuove entrate sono estratte casualmente nel caso di strati campionari e in modo deterministico nel caso degli strati censuari. Nel processo di selezione si escludono dalla popolazione le imprese con fatturato "fiscale" mancante o nullo; quelle con indirizzo all'estero, CAP generico o assenza di localizzazione e con data di cessazione successiva al 2005, già rilevata nell'archivio ASIA.

Nella tavola 6¹⁵ sono riassunti i risultati del processo di costruzione del campione effettivo, il quale è composto da 1.729 imprese, con 140 imprese sovracampionate (evidenziati con bordo tratteggiato sono gli strati sovracampionati) e 635 imprese "nuove entrate". Gli strati evidenziati in bordo oscurato nel campione teorico di *Bethel* sono stati scelti come censuari.

A regime, il panel verrà aggiornato con il rilascio dell'archivio ASIA riferito all'anno più recente, per tenere conto dell'effetto della demografia e dei cambiamenti strutturali. La sostituzione delle imprese del panel avrà la finalità di fronteggiare il problema dell'attrito dovuto alla caduta delle risposte ed allo stato di non eleggibilità di alcune imprese. Da un punto di vista metodologico, la rotazione dovrebbe comportare, invece, la rotazione casuale o in base all'anno di ingresso nel panel di una quota di imprese rispondenti, appartenenti a strati campionari, al fine di ridurre il fastidio statistico sulle unità¹⁶ (*Cranksham et al.*¹⁷). Poiché, in base all'evidenza empirica, tale approccio può causare significative cadute dei tassi di risposta, il problema della gestione di un piano di rotazione resta aperto.

Dopo aver descritto la progettazione del campionamento probabilistico (punto 2), vengono illustrate le fasi che hanno permesso l'implementazione del campionamento *cut off* (punto 3) riferita al solo settore delle telecomunicazioni.

Questo campionamento ragionato di tipo *cut off* è stato calcolato sul volume di affari presente in ASIA.

La popolazione viene ripartita in quattro sotto-popolazioni, relative ai gruppi di attività economica ottenendo i risultati mostrati nella tavola 7.

¹⁵ Nella tavola 6, lo strato è identificato con due variabili concatenate: la prima indica la classe di addetti nelle 3 modalità (1,2,3) corrispondenti alle piccole, medie e grandi imprese. La seconda variabile indica il codice di attività economica. I settori dai quali provengono le imprese dei panel esistenti sono identificati con una sigla nel caso dell'informatica (INF), delle telecomunicazioni (TLC) e del commercio all'ingrosso (ING).

¹⁶ In teoria, la rotazione potrebbe essere motivata anche dal fatto che le unità rispondenti, presenti nel panel da molto tempo, potrebbero stancarsi di rispondere e, quindi, la probabilità di risposta dipenderebbe anche dal numero di anni in cui l'impresa è presente nel panel. Nella prassi operativa, le imprese che sono presenti nel panel da molto tempo e chiedono di essere ruotate vengono sostituite.

¹⁷ Crankshaw M., Kujawal L., Stamas G. *Recent experiences in survey coordination and sample rotation within monthly business establishment surveys*, atti relativi al *Joint statistical meetings – section on survey research methods*

Le imprese del campione stratificato a livello di gruppo delle telecomunicazioni sono 97, di cui 34 appartenevano al vecchio panel. La copertura in termini di fatturato, per i gruppi delle telecomunicazioni fisse e mobili, è spiegata in gran parte dalla componente panel.

Tavola 7: Dimensione campionaria e copertura di fatturato fiscale per gruppo di attività economica – telecomunicazioni (valori assoluti)

Gruppo di attività economica	Numerosità teorica	Numero di imprese panel	Copertura teorica in termini di fatturato	Copertura in termini di fatturato della componente panel
61.1 – Telecomunicazioni fisse	11	7	95,6	88,8
61.2 – Telecomunicazioni mobili	5	4	98,2	63,2
61.3 – Telecomunicazioni satellitari	3	1	97	8,3
61.9 – Altre attività di telecomunicazioni	78	22	80	12,8
Totale imprese	97	34	-	

3.2. La scelta dello stimatore

La fase successiva della strategia campionaria prevede la scelta dello stimatore più adatto a stimare l'indice elementare di fatturato dello strato.

In questo contesto, pertanto, si parte dallo stimatore ottimale di *Horvitz-Thompson*, previsto per il fatturato totale nel caso di estrazione casuale semplice, ed in seguito si propone lo stimatore ottimale per il calcolo dell'indice a base fissa e se ne discutono i limiti, dovuti all'attuale impossibilità di tenere conto della demografia di impresa per le stime infrannuali relative al periodo più recente. In seguito, si illustra il metodo di aggiornamento degli indici e le ipotesi sottostanti, connesse con le strategie di *refreshment* e sostituzione annuale delle imprese non più eleggibili e non rispondenti appartenenti al panel. La sostituzione annuale ha l'obiettivo di fronteggiare il fenomeno dell'attrito (caduta di risposte) mentre il *refreshment* risponde all'esigenza di considerare la demografia e i cambiamenti strutturali avvenuti considerando la popolazione più aggiornata. Il metodo adottato per il calcolo dell'indice permette di includere anche le unità nate dopo l'anno base, che hanno maggiore probabilità di essere eleggibili. L'ultimo aspetto che viene illustrato riguarda le modifiche allo stimatore, dovute all'imputazione delle mancate risposte.

Lo stimatore ottimale è quello di *Horvitz-Thompson* (*Patterson*¹⁸), applicato nel caso di disegno campionario casuale stratificato, che prevede la stessa probabilità di inclusione (rapporto tra numero di imprese del campione e quello della popolazione) per ogni unità dello strato estratta casualmente.

Tale stimatore, a livello di strato elementare, è il seguente:

$$\hat{F}_{tot} = \sum_{i=1}^{n_s} F_i * \frac{N_s}{n_s} \quad (1)$$

Il fatturato totale stimato (\hat{F}_{tot}) dello strato è pari alla sommatoria del fatturato delle imprese del campione (F_i), moltiplicato per il coefficiente di riporto all'universo, dato dal rapporto tra il numero di imprese della popolazione (N_s) e quelle del campione (n_s).

Lo stimatore applicato per il calcolo dell'indice di strato è quindi il seguente:

$$I_s = \frac{\hat{F}_t}{\hat{F}_b} = \frac{\bar{F}_t}{\bar{F}_b} * \frac{N_t}{N_b} \quad (2)$$

¹⁸ Patterson *The Horvitz Thompson estimator*, articolo tratto dal link www.math.umt.edu/patterson/549/Horvitz.pdf

L'indice di strato viene decomposto in due componenti: la prima indica il rapporto tra le medie campionarie del fatturato espresse al tempo corrente (\bar{F}_t) ed all'anno base (\bar{F}_b)¹⁹ mentre la seconda componente indica la demografia di impresa, data dal rapporto tra numero di imprese presenti nella popolazione al tempo corrente (N_t) ed all'anno base (N_b).

Nella prassi operativa, non essendo disponibile l'archivio ASIA aggiornato per il periodo di riferimento dell'indice con la stessa tempestività richiesta per un indicatore congiunturale, non è possibile stimare la componente demografica con il rilascio delle stime degli indici e, quindi, si ipotizza che la popolazione non subisca variazioni rispetto all'anno base, nel senso che la natalità compensa esattamente la mortalità delle imprese. Si ricorda, infatti, che l'archivio annuale ASIA viene rilasciato dopo circa 15 mesi dalla fine del periodo di riferimento, come richiesto da uno specifico Regolamento Comunitario.

L'indice, inoltre, viene calcolato con il seguente metodo:

$$I_s = \frac{\bar{F}_{t-4}}{\bar{F}_b} * \frac{\bar{F}_t}{\bar{F}_{t-4}} \quad (3)$$

L'indice viene decomposto in due componenti: la prima è costituita dall'indice definitivo pubblicato relativo allo stesso trimestre dell'anno precedente (tempo t-4) mentre la seconda è il rapporto tra le medie campionarie del fatturato al tempo corrente ed al tempo t-4.

Le due componenti coincidono con l'indice calcolato sulla base solo se si utilizzano le stesse imprese con gli stessi valori per il calcolo della media del fatturato al tempo t-4. Nella prassi operativa è impossibile che ciò si verifichi: l'effetto delle ritardate (*late respondents*) o mancate risposte si sovrappone a quello dovuto al *refreshment* ed alla sostituzione annuale del panel, che inserisce nuove imprese nel calcolo dell'indice di strato.

Il *refreshment*, tuttavia, permette di aggiornare il campione teorico con l'archivio ASIA più aggiornato per tenere conto della demografia e dei cambiamenti strutturali avvenuti.

In termini formali, si ha:

$$F_{t-4}^{T-1} = \sum_{i=1}^{n_s^{R-T-1}} F_i^{R;T-1} \quad (4)$$

$$F_{t-4}^T = \sum_{i=1}^{n_s^{R;(T-1)}} F_i^R + \sum_{i=1}^{n_s^{LR}} F_i^{LR} + \sum_{i=1}^{n_s^{RETT}} F_i^{RETT} + \sum_{i=1}^{n_s^{nuove}} F_i^{nuove} - \sum_{i=1}^{n_s^{non_eleggibili}} F_i^{non_eleggibili} \quad (5)$$

La differenza tra il fatturato relativo al tempo (t-4), calcolato nell'anno T (5) rispetto a quello calcolato al tempo T-1 (4), è dovuta al contributo fornito dalle imprese ritardatarie (F_i^{LR}), da quelle nuove entrate (F_i^{nuove}), dalle unità che hanno rettificato i dati (F_i^{RETT}) e sottraendo il fatturato delle imprese che nel frattempo sono cessate o che hanno cambiato strato dopo la stima calcolata all'anno T-1 ($F_i^{non_eleggibili}$).

Lo stimatore impiegato nella rilevazione trimestrale di fatturato è uno stimatore quoziente (dato dal rapporto tra le medie campionarie del fatturato al tempo corrente ed allo stesso trimestre dell'anno precedente), che è distorto rispetto a quello ottimale di *Horvitz-Thompson* ma dovrebbe garantire una maggiore qualità dello stimatore grazie all'elevata correlazione tra il fatturato calcolato sulle stesse unità in due periodi differenti che ne riduce la varianza (Gismondi R., 1996²⁰).

Un possibile problema è quello relativo alla natura ibrida di questo indice proposto che non rientra, da un punto di vista teorico, nella casistica degli indici a base fissa e non è un indice a catena in quanto non viene aggiornato il sistema di ponderazione con il rilascio delle statistiche economiche strutturali più recenti: gli indici a base variabile non sono raccomandati dal manuale metodologico dell'EUROSTAT

¹⁹ Il fatturato dell'anno base è pari alla media campionaria calcolata sul fatturato medio annuo di ciascuna impresa.

²⁰ Gismondi R. (1996) *Riflessioni sul problema della stima del livello e della variazione tramite l'uso di informazioni complementari*, Statistica Applicata, vol. 8, 2, pag. 431-453, Rocco Curto Editore, Napoli

perché ritenuti troppo costosi. La metodologia di calcolo dell'indice a base fissa, tuttavia, implica che sia mantenuta costante la base e il problema dell'attrito può essere fronteggiato solo ricorrendo all'archivio dell'anno base (popolazione di riferimento sulla quale è stato stimato il sistema di ponderazione ed estratto il campione). L'inclusione di imprese inizialmente non incluse nel panel pone, peraltro problemi di fastidio statistico a causa del complesso questionario retrospettivo da somministrare inizialmente alle imprese nuove entrate. Una ulteriore ipotesi sottostante agli indici a base fissa è quella di riferirsi ad una popolazione statica, che considera irrilevanti fenomeni demografici e cambiamenti strutturali.

L'ultimo problema operativo legato alla scelta dello stimatore è dovuto al fenomeno delle mancate risposte.

Lo stimatore finale sarà composto da una parte osservata (per la quale si applica lo stimatore quoziente) ed una non osservata (per la quale si usa lo stimatore di regressione senza intercetta per l'integrazione delle mancate risposte).

Lo stimatore è quindi il seguente:

$$\hat{R}_{t;t-4} = \frac{\sum_{i=1}^{n_s^{obs}} F_{i;t}}{\sum_{i=1}^{n_s^{obs}} F_{i;t-4}} + \frac{\sum_{i=1}^{n_s^{nobs}} F_{i;t}}{\sum_{i=1}^{n_s^{nobs}} F_{i;t-4}} = \frac{\sum_{i=1}^{n_s^{obs}} F_{i;t}}{\sum_{i=1}^{n_s^{obs}} F_{i;t-4}} + \frac{\sum_{i=1}^{n_s^{nobs}} \hat{\beta}^* F_{i;t-4}}{\sum_{i=1}^{n_s^{nobs}} F_{i;t-4}} \quad (6)$$

La parte non osservata ("Not observed", nobs) viene stimata mediante modello di regressione²¹, applicato sulle unità rispondenti non anomale appartenenti allo stesso strato. Il coefficiente di regressione così ottenuto ($\hat{\beta}$) viene moltiplicato per la variabile ausiliaria, costituita dal fatturato dell'unità non rispondente riferita allo stesso trimestre dell'anno precedente²².

L'applicazione di tale metodo non permette di ricostruire il campione teorico in quanto le imprese senza variabile ausiliaria non vengono imputate.

Inoltre, l'impiego dello stimatore di regressione senza intercetta è coerente con la strategia campionaria (scelta dello stimatore quoziente) solo nel caso in cui la varianza dello stimatore è proporzionale alla variabile ausiliaria (Luizi *et al.*, 2007²³).

Una ultima importante operazione che viene effettuata sull'indice di strato è quella della post-stratificazione, nel senso che si tiene conto delle informazioni raccolte durante l'indagine per aggiornare le variabili di classificazione (classe di addetti e modalità di attività economica) in modo da rendere più omogenei gli strati nella fase di calcolo degli indici.

Nel caso del campionamento *cut off*, applicato per la divisione delle telecomunicazioni, l'indice elementare di strato viene calcolato in modo analogo con il metodo del quoziente (6), anche se per tale settore viene ignorata la parte non osservata del campione in quanto la presenza di imprese dominanti non permette di applicare modelli di imputazione basati sull'ipotesi MAR (Cirianni *et al.*, 2008²⁴). In altre parole, le imprese dominanti hanno caratteristiche strutturali nettamente diverse da quelle piccole e, quindi, hanno un comportamento che non può essere assimilato alle unità di minori dimensioni che operano nello stesso settore.

²¹ Il modello di regressione, a livello di strato, ha come variabile dipendente il fatturato dell'anno corrente e come variabile ausiliaria il fatturato dello stesso trimestre dell'anno precedente. Il modello è senza intercetta in quanto sono eventi rari quelli caratterizzati da un fatturato nullo al tempo t-4 (per temporanea inattività dell'impresa) e positivi al tempo corrente. Il coefficiente di regressione nel modello senza intercetta coincide esattamente con lo stimatore quoziente in quanto è il rapporto tra la media campionaria del fatturato del tempo corrente e quello dello stesso trimestre dell'anno precedente.

²² La formula (5) considera il rapporto tra totali che, nel caso di rilevazioni basate su panel, coincide con il rapporto tra medie campionarie in quanto le imprese considerate al numeratore sono le stesse di quelle del denominatore.

²³ Luizi, Di Zio, Guarnera, Manzari, De Waal, Pannekoek, Hoogland, Tempelman, Hulliger, Kilchmann (2007) *Recommended practices for editing and imputation in cross sectional business survey*, pag. 61

²⁴ Cirianni A., Coppola S., Panizon F. (2008) *Le indagini sul fatturato degli altri servizi: i metodi di controllo e correzione*, CONTRIBUTI ISTAT N. 13/2008, pag. 87-105

Per quanto riguarda la parte non osservata della popolazione, si ipotizza che le piccole imprese si comportino come quelle presenti nel campione o che, pur mostrando una dinamica differente, forniscono un contributo trascurabile alla stima dell'indice.

3.3. Il sistema di ponderazione

Gli indici elementari a base fissa di fatturato, a livello di strato, vengono riportati all'universo mediante il sistema di ponderazione, proveniente dalle statistiche strutturali sulle imprese (Structural Business Statistics, in breve SBS), al fine di ottenere l'indice del dominio di pubblicazione.

$$I_D = \sum_{s=1}^k I_s * W_s \quad (7)$$

dove I_D indica l'indice di dominio; I_s l'indice di strato (per s che va da 1 a K strati) e W_s i pesi di fatturato.

Tavola 8: Confronto tra il sistema di ponderazione SBS e ASIA per le divisioni economiche indagate – Anno base 2005 (valori percentuali sul fatturato)

attività economica	SBS	ASIA
	Editoria	
581	97,7	95,8
582	2,3	4,2
58	100	100
	Telecomunicazioni	
611	46,4	47,2
612	46,4	45,9
613	2,4	2,4
619	4,7	4,5
61	100	100
	Servizi informatici	
6201	75,1	72,4
6202	13,2	12,4
6203	8,2	9,2
6209	3,4	6,0
62	100	100
	servizi all'informazione	
6311	96,0	95,0
6312	2,2	2,7
6391	1,5	1,9
6399	0,3	0,4
63	100	100

Fonte: ISTAT, statistiche annuali sui conti economici delle imprese (SBS) e archivio ASIA, anno 2005

Il sistema di ponderazione, in base a quanto richiesto dal Regolamento sulle statistiche congiunturali, viene aggiornato ogni cinque anni per tenere conto della demografia e delle modifiche strutturali avvenute nella popolazione.

La costruzione del sistema di ponderazione è basata sulle statistiche strutturali e, in occasione della migrazione ATECO, è stata particolarmente complessa la stima del fatturato totale in ATECO 2007 in

quanto le statistiche strutturali, relative all'anno 2005, erano state disegnate in base a rilevazioni²⁵ che facevano riferimento alla struttura della popolazione espressa nella classificazione ATECO 2002.

La disponibilità dell'archivio ASIA riclassificato, relativo all'anno 2005, ha permesso di definire i domini e la nuova struttura della popolazione alla quale le stime derivanti dalle rilevazioni sui conti economici delle imprese sono state riportate (ISTAT, 2008²⁶).

La metodologia applicata ha permesso di ottenere stime del fatturato totale fino al livello di dettaglio della quarta cifra ATECO.

Per i campioni probabilistici, stratificati anche per classe di addetti, il sistema di ponderazione a livello di classe dimensionale è ottenuto mediante il volume d'affari totale, calcolato sulle imprese eleggibili della popolazione obiettivo, rappresentata da ASIA e riferita all'anno base 2005.

Una operazione di controllo sul sistema dei pesi è stata effettuata confrontando le stime aggregate provenienti dalla fonte ufficiale SBS con i totali noti, ricavati dal fatturato fiscale presente nell'archivio ASIA.

Per ciascun dominio di stima (tavola 8), il risultato del confronto mostra che i dati aggregati sono molto simili e le lievi differenze non sono considerate rilevanti per la stima degli indici di dominio.

4. La fase della raccolta dei dati

La raccolta dei dati consiste nelle seguenti fasi:

- impostazione del questionario;
- spedizione dei modelli;
- acquisizione dei dati;
- politica di prevenzione della mancata risposta totale;
- analisi del comportamento di risposta.

I questionari richiedono il fatturato al netto dell'IVA e il numero medio delle persone occupate. I dati sono richiesti solo per il trimestre di riferimento per le imprese che hanno sempre compilato il questionario e anche per i trimestri retrospettivi per le unità che non hanno ancora risposto o che vengono intervistate per la prima volta, in occasione del *refreshment* e della sostituzione annuale. A tutte le imprese, in ogni caso, viene fornita la possibilità di rettificare i dati precedentemente compilati e di giustificare ampie variazioni del fatturato rispetto ai periodi precedenti. La sezione anagrafica del questionario contiene informazioni pre-compilate relative all'impresa che possono, tuttavia, essere aggiornate dall'impresa. Di particolare interesse è l'informazione relativa al referente dell'impresa per stabilire un eventuale contatto in occasione della revisione dei modelli o del sollecito telefonico mirato e all'indirizzo e-mail, fondamentale per la gestione dei solleciti per posta elettronica.

Le definizioni riportate nel questionario relative al fatturato ed al numero di persone occupate sono in linea con quanto prescritto dal Regolamento Comunitario sulle definizioni statistiche (EUROSTAT, 2001²⁷), con l'unica eccezione che il fatturato segue il criterio della competenza economica e non quello di cassa, in modo da non considerare eventuali fatture che fanno riferimento a servizi erogati nei trimestri precedenti ma incassati nel trimestre di riferimento (caso di sovrastima del fatturato corrente) o a servizi di competenza del trimestre corrente che verranno incassati nei periodi successivi (caso di sottostima del fatturato corrente).

La spedizione dei questionari avviene mediante POSTEL.

La modalità di acquisizione dei dati può essere via web, via fax o via posta. Le imprese, nella lettera di presentazione, vengono invitate a rispondere preferibilmente via web, in alternativa per fax e solo in ultima analisi per posta.

²⁵ Le statistiche strutturali ISTAT sono basate su una rilevazione campionaria per le piccole e medie imprese (appartenenti alla classe di addetti 1-99) e una rilevazione censuaria per le grandi imprese (100 e più addetti).

²⁶ ISTAT (2008) *The estimation of the new weights 2005 in NACE REV 2*, pag. 18-27, estratto da *Back casting STS time series according to the NACE REV 2 classification*, rapporto trasmesso all'EUROSTAT

²⁷ EUROSTAT *Regolamento comunitario sulle definizioni delle variabili statistiche n. 588/2001*

La tavola 9 mostra che le imprese ricorrono prevalentemente al fax anche se, nel periodo considerato, è aumentata notevolmente la quota di imprese che ricorre al web, passando dal 34,5 per cento del primo trimestre 2007 al 43,4 per cento del primo trimestre 2009, grazie anche al progetto di reingegnerizzazione dei processi produttivi delle indagini trimestrali di fatturato degli altri servizi (Frongia et al., 2007²⁸).

La politica di prevenzione della mancata risposta totale comincia subito dopo la scadenza fissata in fase di spedizione e consiste in un sollecito per posta elettronica (per le unità in possesso di indirizzo e-mail), in un sollecito postale ed in un sollecito telefonico mirato al fine di ottenere elevati tassi di copertura in termini di fatturato.

Il sollecito per posta elettronica è particolarmente efficace e tempestivo in questo settore in quanto il 47,5 per cento delle unità del campione possiede un indirizzo e-mail. Il sollecito postale, tuttavia, non può essere eliminato in quanto la restante quota di non rispondenti resta elevata.

Tavola 9: Frequenza degli arrivi in base alla modalità di acquisizione – settori editoria, servizi informatici e servizi all'informazione. Periodo primo trimestre 2007 – primo trimestre 2009 (valori percentuali)²⁹

modalità di arrivo	2007				2008				2009
	I	II	III	IV	I	II	III	IV	I
fax	63,5	60,0	60,4	61,2	62,7	63,4	60,4	55,2	51,0
Posta	2,0	5,4	4,1	2,2	1,0	0,9	1,3	3,2	5,6
Web	34,5	34,6	35,5	36,6	36,3	35,7	38,3	41,6	43,4
Totale	100								

Tavola 10: Tassi di risposta delle imprese in base alla tipologia di ingresso nel panel (nuove entrate o provenienti da rilevazioni congiunturali in ATECO 2002) – settori editoria, servizi informatici e servizi all'informazione. Periodo primo trimestre 2007 – primo trimestre 2009 (valori percentuali)

tipologia di imprese	2007				2008				2009
	I	II	III	IV	I	II	III	IV	I(a)
Nuove	65,7	65,5	65,1	65,3	66,2	63,4	59,9	52,1	44,1
Panel	93,8	91,4	89,1	83,4	85,4	84,3	83,4	79,5	72,3

(a) il tasso di risposta del primo trimestre 2009 è più basso in quanto riferito alla stima provvisoria a 90 giorni mentre per altre occasioni di indagine si fa riferimento ai dati retrospettivi che sono stati raccolti in un lungo periodo di tempo superiore anche a quello riferito alla stima definitiva a 180 giorni.

La politica di prevenzione della mancata risposta è stata rivista in occasione del rilascio delle stime anticipate (a 60 giorni) all'EUROSTAT, in quanto l'evidenza empirica aveva mostrato come la principale causa della distorsione delle stime era dovuta all'effetto dei rispondenti ritardatari influenti che confutavano l'ipotesi *missing at random* (Panizon et al., 2007³⁰).

L'analisi del comportamento di risposta delle imprese mostra che i tassi di risposta delle imprese panel (tavola 10), provenienti dalle rilevazioni già esistenti in vecchia ATECO, sono sistematicamente più elevati di quelli delle imprese 'nuove entrate' ed è quindi molto differente, come sottolineato nei paragrafi precedenti, il comportamento di risposta delle imprese e questo fattore è stato finora considerato determinante nella strategia di rotazione annuale delle unità rispondenti appartenenti al panel.

²⁸ Frongia D., Martinelli R., Panizon F., Querini B., Stanco A. (2007) *Il nuovo sistema informatico altri servizi. Progetto di reingegnerizzazione dei processi produttivi delle indagini trimestrali di fatturato degli altri servizi*. CONTRIBUTI ISTAT N. 12/2007

²⁹ Il dato relativo agli arrivi via posta è sottostimato in quanto, per problemi di tempistica, gli operatori addetti all'inserimento potrebbero non inserire l'informazione sulla modalità di arrivo per posta o fax e, in questo caso, il sistema, in automatico, attribuisce il fax come modalità di arrivo.

³⁰ Panizon, Cirianni, Congia, Garozzo, Giorgi *L'evoluzione dei processi produttivi per la gestione ed il controllo dei tassi di risposta*, Rivista di statistica ufficiale n 2-3/2007, pag. 5-31, edizione Franco Angeli

5. La metodologia di controllo e correzione degli errori non campionari

La metodologia adottata per il controllo e correzione degli errori non campionari presenta strategie comuni al settore oligopolistico delle telecomunicazioni e a quelli probabilistici dell'editoria, dei servizi informatici e dei servizi all'informazione per alcune tipologie di errore e aspetti differenti, in base al settore considerato, per il trattamento delle mancate risposte e degli *outlier*.

5.1. La casistica empirica degli errori non campionari

La casistica dei principali errori non campionari, adottata a livello europeo (*Bergdahl et al*, 1997³¹), è stata particolarmente utile durante la fase di correzione degli errori in quanto ha permesso di applicare questo schema teorico alla realtà operativa, individuando anche casi di studio che, una volta generalizzati, hanno permesso di stabilire alcune regole empiriche di correzione degli errori.

La casistica empirica dei principali errori non campionari è la seguente:

- Errori di processo: tale tipologia di errore è costituita dagli errori di digitazione, di acquisizione dei dati via web, e da quelli causati da fax non perfettamente leggibili;
- Errori di misura: tali errori sono quelli di compilazione del questionario da parte delle imprese. Gli errori più frequenti sono quelli di inserimento del valore del fatturato riferito non solo al trimestre corrente ma anche a quelli precedenti; errori dovuti a errata interpretazione della definizione del fatturato, come ad esempio l'inserimento di ricavi accessori o di ricavi straordinari e occasionali per l'attività tipica; imputazione del fatturato in base al criterio di cassa anziché quello di competenza economica. In questo caso, verrebbe inserita anche una parte del ricavo che si riferisce a trimestri precedenti ma che vengono fatturati con ritardo;
- Errori di classificazione: gli errori di classificazione possono essere dovuti a imprese che cambiano strato (definito da classe di addetti e modalità di attività economica) oppure da imprese appartenenti a strati campionari con elevato livello del fatturato per addetto che non permette di considerarli omogenei con le imprese dello stesso strato. Nella rilevazione dell'informatica, questo tipo di errore di classificazione potrebbe essere dovuto a operazioni di *outsourcing*, al fatto che alcune attività aziendali vengono svolte da altre imprese del gruppo di appartenenza, all'impiego di lavoratori atipici – interinali e CO.CO.PRO – esclusi dalla definizione statistica di addetti;
- Errori di lista: questa tipologia di errore è dovuta ad errati indirizzi presenti nell'archivio ASIA e per tali unità si cerca di recuperare l'indirizzo tramite PAGINE BIANCHE, visure camerali e internet. Le imprese con indirizzo cambiato vengono nuovamente intervistate ed escluse dal panel solo nel caso in cui la ricerca dell'indirizzo non è andata a buon fine.

Queste tipologie di errore vengono corrette nella fase di raccolta mediante ricontatto telefonico, sia per il settore delle telecomunicazioni che per quelli caratterizzati da campioni probabilistici, anche se sono differenti le soglie di accettazione, misurate in termini di variazioni tendenziali, per l'identificazione dei probabili errori non campionari da sottoporre alla tecnica del *follow up*. La decisione di adottare soglie di accettazione differenti nella fase di *microediting* è motivata dal fatto che le grandi imprese del settore delle telecomunicazioni, caratterizzate da elevati livelli di fatturato, possono presentare probabili errori con soglie più basse rispetto a settori probabilistici dove alte variazioni per contenuti livelli di fatturato sono molto frequenti. Un'altra ragione è dovuta al fatto che il settore oligopolistico delle telecomunicazioni è caratterizzato da un basso numero di imprese (tutte influenti sulla stima dell'indice) mentre i settori probabilistici hanno una elevata numerosità campionaria e questo, per rispettare il criterio della

³¹ Bergdahl, Black, Bowater, Chambers, Davies, Draper, Elvers, Full, Holmes, Lundqvist, Lundstrom, Nordberg, Perry, Pont, Prestwood, Richardson, Skinner, Smith, Underwood, Williams (1997) *Model quality report in business statistics*, EUROSTAT, pag. 82-160 e pag. 54-60

tempestività, induce a rilassare la soglia per ridurre i tempi del ricontatto, focalizzando l'attenzione solo sui casi di alta variazione di imprese con significativi livelli di fatturato.

5.2. Il trattamento delle mancate risposte

L'integrazione delle mancate risposte totali (che nel caso delle rilevazioni trimestrali di fatturato coincide con il metodo dell'imputazione per la presenza di una sola variabile obiettivo) ha lo scopo di ricostruire il campione teorico, così come è stato disegnato (EUROSTAT, 2002³²).

L'imputazione ha le seguenti caratteristiche generali:

- a) segue l'ipotesi MAR (*Missing At Random*), ossia che le imprese non rispondenti si comportano allo stesso modo delle unità rispondenti che non cadono nelle code della distribuzione (in altre parole, un dato anomalo di un rispondente viene ignorato per evitare di generare un errore sistematico sulle stime finali) e che appartengono allo stesso strato (e quindi sono omogenee per caratteristiche strutturali). La revisione delle stime negli altri servizi è dovuta all'effetto di rispondenti ritardatari che confutano l'ipotesi e che potrebbero essere inseriti nelle liste del sollecito mirato, qualora essi si verificassero con regolarità nel tempo. Un altro motivo della revisione è dovuto al fatto che i rispondenti ritardatari vengono inclusi nell'insieme dei rispondenti sui quali viene calcolata la stima delle mancate risposte;
- b) la variabilità intrinseca nell'ambito della base di microdati utilizzata per calcolare le stime elementari si riduce, in quanto l'imputazione dei dati mancanti si basa sull'applicazione della variazione tendenziale media calcolata sui rispondenti. La verifica empirica dell'effetto dell'imputazione sulla varianza dello stimatore quoziente comporta una preliminare linearizzazione della varianza al fine di individuare un metodo di approssimazione della varianza dello stimatore (Banca d'Italia, 2007³³; Cocchi D.³⁴);
- c) la stima delle mancate risposte con il metodo di regressione senza intercetta è coerente con la scelta dello stimatore nell'ipotesi che la varianza dell'imputazione sia proporzionale con la variabile ausiliaria impiegata. In altre parole, eventuali differenze tra stime sui rispondenti e stime con imputazione dei dati mancanti potrebbero essere attribuite anche alla confutazione dell'ipotesi sulla varianza dell'imputazione, anche se questo complesso problema metodologico non viene affrontato in questo lavoro;
- d) in caso di stima diretta (assenza di imputazione), i risultati possono differire in quanto i pesi delle imprese del campione, espressi in termini di fatturato al tempo (t-4), non sono gli stessi per tutte le unità e, quindi, una impresa imputata contribuisce alla stima dell'indice elementare mentre una impresa non imputata non viene inclusa nelle stime. Inoltre tra dati grezzi e stimati c'è anche la differenza dovuta al fatto che le imprese anomale rispondenti sono incluse negli indici grezzi ma escluse dall'imputazione. Per quanto concerne il problema dell'inferenza, la stima delle mancate risposte permette la ricostruzione del campione teorico ma potrebbe rettificare l'errore di stima atteso, misurato in termini di coefficiente di variazione, mentre la stima diretta implica, per l'inferenza, l'ipotesi che il comportamento dei non rispondenti non incida sulla stima dell'errore di stima riferito al campione teorico;
- e) Nel disegno casuale stratificato, le imprese hanno la stessa probabilità di inclusione se appartenenti a strati campionari mentre sono tutte incluse nel caso di strati censuari. La probabilità di inclusione non tiene conto dell'asimmetria della distribuzione empirica del fatturato all'interno dello strato, a differenza del campionamento *cut off* che comunque resta valido solo per settori oligopolistici. La presenza dell'asimmetria del fatturato potrebbe essere una delle possibili cause della confutazione dell'ipotesi MAR in quanto non si riuscirebbero a individuare strati omogenei di imprese.

³² EUROSTAT (2002) *Methodology of short term business statistics*, pag. 33-35

³³ Banca D'Italia (2007) *La misurazione dell'errore di stima nelle statistiche sui tassi di interesse bancari*, Supplementi al bollettino statistico. Note Metodologiche. Anno XVII Numero 36 – 14 giugno 2007

³⁴ Cocchi Daniela “*La stima per quoziente*”, capitolo 4, estratto da “*Teoria dei campioni. Corso base*”, dispensa pubblicata su internet all'indirizzo /www2.stat.unibo.it/trivisano/default_files/files%20camp/dispensa%20Cocchi.pdf

In base a queste considerazioni generali, sono state scelte le seguenti strategie:

- 1) la stima diretta è stata impiegata per il settore delle telecomunicazioni in quanto la presenza di imprese dominanti confuta l'ipotesi MAR, nel senso che le piccole unità rispondenti, presenti nel campione, hanno di solito una dinamica congiunturale differente rispetto alle imprese dominanti. Per i settori caratterizzati da campioni probabilistici, il metodo di regressione senza intercetta, applicato a livello di strato, è stato scelto per l'imputazione delle mancate risposte e, come già sottolineato, è in linea con la strategia campionaria sotto certe ipotesi sulla varianza dell'imputazione. Questo metodo è stato scelto in quanto permette, mediante i residui studentizzati, di individuare un contenuto numero di *outlier* sull'insieme dei rispondenti da ignorare per l'imputazione delle mancate risposte, nell'ipotesi che non ci siano anomalie nella parte non osservata del campione;
- 2) nel caso di strati con poche osservazioni, si collassano gli strati con quelli adiacenti per classe dimensionale. Tale meccanismo è applicato anche per la stima degli strati vuoti;
- 3) la stima avviene utilizzando i dati storici della stessa variabile in quanto fortemente correlati con la variabile obiettivo (stima della variazione tendenziale). Solo per la stima della base, si usa il volume d'affari come informazione ausiliaria, dopo aver verificato empiricamente che il fatturato dell'indagine è fortemente correlato al volume d'affari.

Un aspetto di particolare interesse, che viene approfondito in questo paragrafo, riguarda l'impatto dell'imputazione delle mancate risposte sulla precisione delle stime intervallari, sottostanti alle stime campionarie puntuali correntemente diffuse mediante comunicato stampa.

Il problema, come sottolineato in precedenza, deriva dal fatto che l'imputazione impiegata applica lo stesso coefficiente di regressione, stimato a livello di strato sulle unità rispondenti non anomale, alle unità non rispondenti in possesso della variabile ausiliaria e questo introduce una forma di distorsione riducendo forzatamente la varianza della stima e, quindi, l'intervallo di confidenza riferito alla stima quoziente.

L'obiettivo, quindi, è quello di misurare empiricamente l'effetto dell'imputazione sulla varianza delle stime campionarie prodotte.

Lo stimatore quoziente \hat{R} , dato dal rapporto tra le medie campionarie del fatturato calcolate al tempo t ed al tempo $(t-4)$, è una funzione non lineare nelle due variabili e la non linearità rende difficile la stima analitica della varianza. Per risolvere il predetto problema metodologico (Banca d'Italia, 2007³⁵; Cocchi D³⁶), si usa la tecnica della linearizzazione della varianza, mediante l'applicazione del *delta method* al fine di trovare approssimazioni della varianza dello stimatore quoziente, basate sul polinomio di *Taylor*.

Un primo metodo (che nel seguito sarà definito metodo A) calcola la varianza approssimata dello stimatore quoziente a livello di strato ed in seguito la varianza del dominio di pubblicazione come media ponderata delle varianze di strato (con il sistema di ponderazione degli indici di fatturato). Questo metodo è coerente con la strategia campionaria (campioni probabilistici), adottata per i settori dell'editoria, dei servizi informatici e dei servizi all'informazione e agenzie di stampa per il calcolo degli indici a base fissa.

In termini formali, il metodo A per la stima della varianza dello stimatore quoziente a livello di strato è il seguente:

$$VAR(\hat{R}_s) \approx \frac{1 - \frac{n_s}{N_s^b}}{n_s * F_{t-4}^{-2}} * \frac{\sum_{i=1}^{n_s} (F_{i,t} - r_s * F_{i,t-4})^2}{n_s - 1} \quad (8)$$

³⁵ Idem nota 22.

³⁶ Idem nota 23

Dove $VAR(R_s)$ indica la varianza approssimata dello stimatore quoziente, n_s il numero di imprese del campione dello strato s , N_s^b il numero di imprese dello strato nella popolazione espresso al tempo base

nel quale è stato estratto il campione, \bar{F}_{t-4} il fatturato medio al quadrato espresso al tempo (t-4), F_{it} e F_{it-4} il fatturato della i -esima impresa calcolato al tempo t e al tempo (t-4) e, infine, r_s indica il rapporto tra le medie campionarie del fatturato al tempo corrente ed allo stesso trimestre dell'anno precedente dello strato s -esimo.

Per coerenza con il metodo degli indici a base fissa, la varianza approssimata dello stimatore quoziente degli indici del singolo dominio di pubblicazione è calcolata come:

$$VAR(R_D) \approx \sum_{s=1}^k VAR(R_s) * W_s \quad (9)$$

Dove $VAR(R_D)$ indica la varianza del dominio di pubblicazione e W_s il sistema di ponderazione relativo all'anno base.

Un secondo approccio utilizzabile in questo contesto empirico impiega il *combined ratio estimator* (Saxena et al, 1995³⁷) per la stima della varianza dello stimatore quoziente nel caso di campioni stratificati (nel seguito metodo B).

Tavola 11: Differenza tra le varianze delle stime quoziente dei rispondenti e con imputazione delle mancate risposte. Metodo A. Divisioni dell'editoria, dei servizi informatici, dei servizi all'informazione e agenzie di stampa. Periodo primo trimestre 2007 – primo trimestre 2009

Tipologia di stime	2007				2008				2009
	I	II	III	IV	I	II	III	IV	I
	<i>58 - editoria</i>								
Rispondenti	0,000003692	0,000003752	0,000002097	0,000001796	0,000004648	0,000003204	0,000000881	0,000000887	0,000017569
Con imputazione	0,000003692	0,000003753	0,000002103	0,000001796	0,000004632	0,000003198	0,000000882	0,000000854	0,000010505
Differenza	0,00000	-0,000000001	-0,000000001	0,00000	0,000000002	0,000000001	-0,000000001	0,000000003	0,00000706
	<i>62 - servizi informatici</i>								
Rispondenti	0,000003406	0,000000587	0,000003028	0,000000931	0,000001188	0,000006662	0,000001048	0,000001243	0,000001487
con imputazione	0,000003219	0,000000561	0,000002826	0,000000854	0,000001052	0,000005606	0,000001407	0,000001002	0,00000102
Differenza	0,000000187	0,000000026	0,000000202	0,000000077	0,000000136	0,000001056	-0,00000036	0,000000241	0,000000467
	<i>63 - servizi all'informazione e agenzie di stampa</i>								
Rispondenti	0,000042807	0,000032821	0,000044451	0,000030586	0,000013018	0,000009334	0,000025145	0,00001627	0,000080596
con imputazione	0,000041679	0,000028446	0,000038775	0,000023311	0,000011541	0,000008289	0,000023622	0,00001518	0,000063471
Differenza	0,000001128	0,000004375	0,000005676	0,000007275	0,000001477	0,000001045	0,000001523	0,00000109	0,000017125

Il metodo B calcola lo stimatore quoziente del dominio di pubblicazione nel seguente modo:

$$\hat{R}_D = \frac{\sum_{S=1}^k \sum_{i=1}^{n_s} r_{si} * w_{si}}{\sum_{s=1}^k \sum_{i=1}^{n_s} w_{si}} \quad (10)$$

Dove w_{si} è il fatturato relativo al tempo (t-4) della i -esima impresa appartenente allo strato s -esimo. Il sistema dei pesi, quindi, è riferito alle unità del campione e non della popolazione ed inoltre è più aggiornato rispetto alla base.

³⁷ Sanjay Kumar Saxena, A.K. Nigam e N.D. Shukla *Variance estimation for combined ratio estimator*, The Indian Journal of statistics, 1995, Volume 57, Serie B, pag. 85-92

I due metodi, implicitamente, fanno riferimento al metodo degli indici a base fissa (metodo A) ed al metodo degli indici a base variabile (metodo B).

La formula approssimata della varianza del dominio di pubblicazione è la seguente:

$$VAR(R_D) \approx \frac{\sum_{s=1}^k n_s * (1 - \frac{n_s}{N_s^b}) \sum_{s=1}^k \sum_{i=1}^{n_s} (F_{si;t} - r_s * F_{is;t-4})^2}{(\sum_{s=1}^k \sum_{i=1}^{n_s} F_{si;t-4}) \sum_{s=1}^k n_s - 1} \quad (11)$$

L'idea di considerare due metodi differenti per la stima della varianza del quoziente permette di verificare non solo se l'imputazione delle mancate risposte ha un impatto significativo (confrontando i risultati ottenuti sull'insieme dei rispondenti rispetto a quelli che includono anche i dati imputati) sulla variabilità delle stime e, implicitamente, sulle stime per intervalli ma anche di esaminare se, attribuendo un differente peso alle imprese imputate (riferito all'anno base per il metodo A e al tempo t-4 per il metodo B), si ottengono risultati differenti in termini di varianza.

L'evidenza empirica mostra che, impiegando sia il metodo A sia il metodo B, l'imputazione riduce lievemente la varianza delle stime (tavola 11 e 12). L'effetto è comunque molto contenuto sia perché la quota di fatturato stimata è limitata, grazie agli elevati tassi di copertura ottenuti (tavola 13) sia perché esiste una frazione di imprese non rispondenti che non vengono imputate per assenza della variabile ausiliaria, costituita dal fatturato dello stesso trimestre dell'anno precedente.

Tavola 12: Differenza tra le varianze delle stime quoziente dei rispondenti e con imputazione delle mancate risposte. Metodo B. Divisioni dell'editoria, dei servizi informatici, dei servizi all'informazione e agenzie di stampa. Periodo primo trimestre 2007 – primo trimestre 2009

Tipologia di stime	2007				2008				2009
	I	II	III	IV	I	II	III	IV	I
<i>58 – editoria</i>									
rispondenti	0,000036047	0,000049968	0,000155	0,000002531	0,000192	0,000079823	0,000007478	0,000007005	0,000112
con imputazione	0,000036047	0,000049967	0,000153	0,000002531	0,000188	0,00007912	0,000007622	0,00000997	0,000077921
differenza	0,000000	0,000000001	0,000002	0,000000	0,000004	0,000001	-0,0000001	-0,000003	0,000034
<i>62 – servizi informatici</i>									
rispondenti	0,000042775	0,00000921	0,000046802	0,000005287	0,000019033	0,000128	0,000016015	0,000010804	0,000002102
con imputazione	0,000042135	0,000009076	0,000046291	0,000005226	0,000018672	0,000126	0,000024441	0,000010708	0,000002016
differenza	0,0000006	0,0000001	0,0000005	0,0000001	0,0000004	0,0000020	-0,0000084	0,0000001	0,0000001
<i>63 – servizi all'informazione e agenzie di stampa</i>									
rispondenti	0,00012	0,000058176	0,000076273	0,00003732	0,000026367	0,000002046	0,000047088	0,000051671	0,00005105
con imputazione	0,000119	0,000047805	0,000067488	0,000026949	0,000023261	0,000001996	0,000046697	0,000051187	0,000043498
differenza	0,000001	0,000010	0,000009	0,000010	0,000003	0,00000005	0,00000039	0,00000048	0,00000755

Tavola 13: Tassi di copertura in termini di fatturato – editoria, servizi informatici e servizi all'informazione e agenzie di stampa. Periodo 2005-2009 (valori percentuali)

Divisione di attività economica	2005	2006				2007				2008				2009
		I	II	III	IV	I	II	III	IV	I	II	III	IV	I
58 – editoria	86,2	87,6	87,6	87,6	83,6	88,7	88,2	85,3	85,3	84,5	83,4	83,2	76,3	71,8
62 – servizi informatici	76,2	94,1	90,5	90,4	89,2	92,1	91,5	90,7	86,6	88,8	89,0	87,9	80,9	77,2
63 – servizi all'informazione	77,1	94,1	92,2	91,4	91,9	94,9	91,0	93,3	89,7	92,0	89,5	90,1	85,9	82,8

Va anche notato che la varianza delle stime, calcolata sui soli rispondenti o considerando anche l'imputazione delle mancate risposte, è comunque bassa, assicurando una elevata qualità delle stime. Tale risultato è principalmente dovuto al ricorso dei panel per le stime longitudinali: l'impiego delle

stesse imprese nel calcolo del quoziente riduce considerevolmente la varianza delle stime grazie alla elevata correlazione tra il fatturato al tempo (t-4) e quello al tempo t. Questo risultato non si sarebbe ottenuto nel caso di campioni indipendenti³⁸.

5.3. Il trattamento degli outlier

Un importante problema metodologico, che è stato affrontato nell'ambito delle rilevazioni trimestrali di fatturato degli altri servizi, è quello derivante dalla presenza di *outlier*. Questi possono essere definiti, in modo standard, come valori che si posizionano sulle code della distribuzione della variabile obiettivo, producendo un impatto significativo sulla stima finale e, quindi, mettendone a rischio la robustezza (Rizzi, 1992³⁹).

In letteratura (Chambers, 1986⁴⁰), gli *outlier* vengono distinti in:

- Outlier non rappresentativi: si tratta di valori supposti unici nella popolazione in quanto affetti da errori di misura, di processo o di classificazione o caratterizzati da *break* strutturali;
- Outlier rappresentativi: si tratta di unità con valori anomali, non affetti da errori non campionari, che possono essere rappresentative di altre unità non osservate presenti nella popolazione. Per questa tipologia di *outlier*, si pone il problema del loro trattamento in fase di stima in quanto, rientrando nella casistica degli errori campionari, il differente peso campionario che gli viene attribuito modifica in modo sostanziale le stime, non rendendole robuste.

Nel caso delle rilevazioni trimestrali di fatturato degli altri servizi caratterizzate da campioni stratificati, l'attenzione è stata posta sull'individuazione delle unità più influenti sulle stime, ricorrendo al metodo del *macroediting* e, in particolare, al metodo aggregato univariato (Barcaroli et al, 1998⁴¹).

Il *macroediting* consiste nelle seguenti operazioni:

- Individuazione degli strati influenti sulla stima dell'indice dello specifico dominio mediante una funzione di punteggio macro;
- Individuazione delle imprese influenti, all'interno degli strati selezionati, mediante una funzione di punteggio micro.

La funzione di punteggio macro (FP1) è data dal prodotto tra la variazione tendenziale del fatturato dello strato (ΔF_s) moltiplicata per il peso dello strato (W_s) nel dominio, riferito all'anno base.

$$FP1 = \Delta F_s * W_s \quad (12)$$

Gli strati selezionati sono quelli che presentano un valore della funzione punteggio macro superiore a una soglia prefissata. Esiste un margine di soggettività sulla scelta della soglia critica e tale decisione è di fondamentale importanza in quanto, come verrà mostrato in seguito, esiste un *trade off* tra numero degli strati selezionati e numero di *outlier* influenti che dovrebbero essere sottoposti a verifica, mediante ritorno sulle unità e ricontatto telefonico. Tale *trade off* è cruciale nelle statistiche congiunturali, le quali devono soprattutto rispettare il requisito qualitativo della tempestività.

Una volta selezionati gli strati influenti, per ciascuno strato viene calcolata la seguente funzione punteggio micro (FP2):

³⁸ Nella stima della varianza del rapporto, non si è tenuto conto della componente addizionale della varianza dovuta esclusivamente all'imprecisione implicita in ogni processo di imputazione.

³⁹ Rizzi A. (1992) *Robustezza*, pag. 351-380, capitolo tratto da *Inferenza statistica*, edizione UTET

⁴⁰ Chambers R.L. (1986) *Outlier robust finite population estimation*, pag. 1063-1069, Journal of the American Statistical association, vol 81 n 396

⁴¹ Barcaroli G., Luzi O. e Ceccarelli C. (1998) *Il macroediting: tecniche di correzione interattiva di variabili quantitative guidate dall'analisi degli aggregati. Il caso del sistema dei conti delle imprese*. ISTAT, quaderni di ricerca, n 1/1998, pag 3-58

$$FP2 = \Delta F_i * W_{i;t-4} \quad (13)$$

che, essendo il prodotto tra la variazione tendenziale del fatturato della singola impresa (ΔF_i) e il suo peso nel campione espresso al tempo t-4 ($W_{i;t-4}$), esprime esattamente il contributo di ciascuna unità alla stima dell'indice di strato.

Le unità influenti sono quelle che cadono nelle code della distribuzione della funzione di controllo micro, individuate con il metodo grafico del *box plot*, e che appartengono agli strati influenti, individuati mediante la funzione di punteggio macro.

Una impresa, quindi, può risultare influente per alto peso o elevata variazione ed è necessario separare la componente variazione per individuare il sottoinsieme delle unità influenti e anomale. Tale obiettivo viene conseguito applicando il metodo di *Hidiroglou-Berthelot*⁴², che, a partire da una trasformazione lineare che rende simmetrica la variazione tendenziale del fatturato, definisce una soglia di accettazione, al di fuori della quale cadono le unità anomale:

$$A = (q_{0,5} - c_{inf} * (q_{0,5} - q_{0,25}); q_{0,5} + c_{sup} * (q_{0,75} - q_{0,5})) = (A_{inf}; A_{sup}) \quad (14)$$

dove $q_{0,5}$ esprime il valore mediano, $q_{0,25}$ il primo quartile e $q_{0,75}$ il terzo quartile della variabile obiettivo; "c" è un parametro arbitrario (che, in linea di principio, potrebbe anche essere assegnato in modo differente per l'estremo inferiore e superiore). La scelta del valore del parametro "c" nell'algoritmo di *Hidiroglou-Berthelot* è lasciata alla discrezionalità dell'esperto dell'indagine e riveste una notevole importanza nella determinazione della quota di *outlier* identificati (Cirianni et al, 2006-2007⁴³).

Nella procedura correntemente utilizzata, il parametro "c" è posto di solito uguale a 1,5 e con tale soglia risultano caratterizzati da valori anomali influenti circa il 3-5 per cento delle imprese rispondenti.

Una volta individuati gli *outlier* influenti, si pone il problema del loro trattamento. Questo problema, in letteratura (Lee, 1995⁴⁴), viene affrontato mediante la correzione degli *outlier* non rappresentativi e la rettifica del peso campionario di quelli rappresentativi, anche se i metodi proposti risentono anch'essi dell'uso di parametri soggettivi in base ai quali modificare i pesi campionari.

In altre parole, non conoscendo la distribuzione della variabile obiettivo nella popolazione di riferimento, si pone il problema di capire se la distribuzione del campione, individuando nelle code gli *outlier* influenti rappresentativi, è la stessa di quella della popolazione oppure se è preferibile formulare alcune ipotesi sulla parte non osservata della popolazione.

Nel caso degli strati censuari, invece, poichè si conosce la distribuzione della variabile di interesse nella popolazione, vengono individuati e corretti solo gli errori non campionari influenti.

Con riferimento alla produzione di indici di fatturato a base fissa, la rettifica dei pesi campionari consiste nella modifica del sistema di ponderazione impiegato per calcolare l'indice di dominio.

In particolare, per le rilevazioni trimestrali dell'editoria, dei servizi informatici e dei servizi all'informazione e agenzie di stampa, il metodo impiegato è stato quello di autorappresentare gli *outlier*, supponendo che siano unici nella popolazione e, quindi, che il loro peso in termini di fatturato non debba essere attribuito ad altre unità non osservate presenti nella popolazione.

Il trattamento viene effettuato, per ciascun strato influente, nel seguente modo:

$$I_s = I_{normali} W_{normali} + I_{outlier} W_{outlier} \quad (15)$$

⁴² Hidiroglou M.A., Berthelot J.M. (1986), *Statistical editing and imputation for periodic Business Surveys*, Survey Methodology, 12, pag. 73-84

⁴³ Panizon F., Cirianni A. *The problem of outliers in short term surveys on turnover of other services*, atti del Convegno Intermedio SIS 2007, Venezia, 6-8 giugno 2007. Cirianni A. e Gismondi R. (2006) *Identificazione e trattamento statistico delle osservazioni anomale nel settore degli altri servizi*, paragrafo 4.4. tratto dalla direttiva ISTAT, TRAC16 "Sperimentazione di stime anticipate per specifici indicatori congiunturali, finalizzata al rilascio in produzione delle relative metodologie"

⁴⁴ LEE (1995) *Outliers in business surveys*, in Cox, Binder, Chianappa, Christianson, Colledge, Kott *Business survey methods*, John Wiley & Sons, New York, pag. 503-523

dove:

$$I_{normali} = \frac{\sum_{i=1}^{n(normali)} F_{i,t}}{\sum_{i=1}^{n(normali)} F_{i,t-4}} \quad (16)$$

$$I_{outlier} = \frac{\sum_{i=1}^{n(outlier)} F_{i,t}}{\sum_{i=1}^{n(outlier)} F_{i,t-4}} \quad (17)$$

$$W_{normali} = \frac{\sum_{i=1}^{N(s)-outlier} F_i}{\sum_{i=1}^{N_s} F_i} \quad (18)$$

$$W_{outlier} = \frac{\sum_{i=1}^{n(outlier)} F_i}{\sum_{i=1}^{N(s)} F_i} \quad (19)$$

L'indice di strato (I_s) è quindi calcolato come media ponderata degli indici calcolati sulle imprese normali e su quelle anomale influenti, con pesi, relativi all'anno base, dati rispettivamente dalla quota del fatturato totale della popolazione escludendo la quota di competenza delle imprese *outlier* ($W_{normali}$) e dalla quota del fatturato riferita solo alle imprese anomale ($W_{outlier}$).

Il controllo finale consiste nell'analisi degli scostamenti delle stime, a livello di dominio, nelle diverse fasi di *editing* (rispondenti, con l'imputazione delle mancate risposte, con l'esclusione e con l'autorappresentazione degli *outlier*).

In caso di scostamenti significativi, si analizzano le cause ritornando sulle unità, mediante l'analisi grafica (*box plot* e *scatter plot*) a livello di strato e il ricontatto telefonico, per la correzione degli eventuali errori influenti delle unità posizionate sulle code della distribuzione (cosiddetti *outlier* non rappresentativi).

Il metodo, come illustrato nel paragrafo successivo, permette di rendere più robuste le stime rispetto a quello dell'esclusione.

5.4. L'effetto del processo di controllo e correzione sulle stime finali

La ricostruzione micro delle serie storiche retrospettive, riferita al periodo 2006-2009, per i settori per i quali si utilizza un disegno probabilistico è avvenuta con metodi di controllo e correzione differenti per quanto riguarda la stima degli indici del 2006 in base 2005=100 e per quanto riguarda gli indici del periodo 2007-2009.

Per il periodo 2006, gli indici trimestrali sono stati calcolati impiegando il fatturato medio annuo del 2005 per la stima della base e, in caso di mancata risposta, è stata utilizzata l'informazione ausiliaria proveniente dal volume d'affari di ASIA, dopo aver verificato sull'insieme dei rispondenti che il fatturato dell'indagine era fortemente correlato a tale variabile. Le unità che presentavano anomalie rispetto all'anno base, individuate con il metodo dei residui *studentizzati* applicato al modello di

regressione che mette in relazione (a livello di strato) il fatturato dei singoli trimestri del 2006 con quello dell'anno base, sono state escluse dalla stima.

Per il periodo 2007-2009, gli indici sono stati stimati con il metodo *standard* descritto in precedenza e le mancate risposte sono state imputate con metodo di regressione. Gli *outlier* influenti sono stati definiti come le unità che cadono nelle code della distribuzione della variabile obiettivo (variazione tendenziale opportunamente simmetrizzata) e che risultano influenti in base al metodo del *macroediting*. Gli *outlier* influenti sono stati trattati con il metodo dell'autorappresentazione.

L'obiettivo del presente paragrafo è quello di mostrare l'impatto del processo di *editing* sulle stime retrospettive.

Con riferimento alla divisione dell'editoria (tavola 14), le stime del 2006 risentono molto più dell'effetto degli *outlier* (individuati sia sull'insieme dei rispondenti che su quello che integra le mancate risposte della base con l'informazione ausiliaria proveniente dal volume d'affari di ASIA) piuttosto che dell'integrazione con informazione ausiliaria.

Nel settore dei servizi informatici (tavola 15), invece, è evidente, in alcuni trimestri, anche l'impatto sulle stime dovuto all'integrazione delle mancate risposte con l'informazione ausiliaria costituita dal volume d'affari in quanto in questa divisione è significativo l'effetto delle *wave respondents* (tavola 13), che sono le imprese che hanno risposto nel 2006 e non nell'anno base 2005.

Il settore dei servizi all'informazione e agenzie di stampa (tavola 16) presenta risultati analoghi a quello dell'editoria e, quindi, è maggiore l'impatto dovuto agli *outlier*.

Con riferimento al periodo 2007-2009, l'analisi delle stime riferite al settore dell'editoria (tavola 17) mostra che l'effetto dell'imputazione ha un impatto contenuto sulle stime mentre il trattamento degli *outlier* influenti è rilevante, soprattutto considerando che in molte replicazioni del processo l'autorappresentazione modifica i risultati ottenuti con l'esclusione e ciò può essere attribuito a strati influenti, dove la combinazione tra la variazione e il peso degli *outlier* nello strato è significativo. In altri casi, il tipo di trattamento non incide significativamente sulle stime anche se si discosta rispetto alle stime grezze, trattandosi comunque di strati influenti sull'indice di dominio.

Tavola 14: *Analisi delle stime delle variazioni tendenziali nelle diverse fasi di editing – Periodo 2006. Divisione 'editoria' (valori percentuali)*

Periodo	fasi del processo di <i>editing</i>			
	dati solo rispondenti	dati solo rispondenti senza <i>outlier</i>	Dati con integrazione delle mancate risposte relative alla base con volume d'affari di ASIA	Dati con integrazione delle mancate risposte relative alla base con volume d'affari di ASIA e esclusione <i>outlier</i> trimestrali
I	1,8	6,4	1,8	6,3
II	16,2	7,5	14,9	7,1
III	-1,8	-7,7	-3,0	-8,0
IV	17,9	14,4	17,3	14,3

Per i servizi informatici (tavola 18), l'effetto dell'imputazione e dell'autorappresentazione non è rilevante mentre l'esclusione degli *outlier*, riguardando strati influenti, ha un impatto sulle stime.

Per la divisione dei servizi all'informazione e agenzie di stampa (tavola 19), l'impatto dell'imputazione non è significativo mentre l'autorappresentazione, in alcune replicazioni, ha un effetto rilevante.

In conclusione, il processo di controllo e correzione ha un impatto molto maggiore per le stime del 2006 rispetto a quelle relative al periodo 2007-2009 e ciò è dovuto principalmente ai diversi metodi applicati per l'individuazione degli *outlier* e per l'integrazione delle mancate risposte. Per il 2006, gli *outlier* da modello usano come variabile obiettivo il fatturato del trimestre corrente su quello della base e, quindi, pur essendo i modelli di regressione ben specificati, gli *outlier* esclusi (in ogni caso di numero limitato) potrebbero anche intercettare imprese caratterizzate da forte stagionalità. In particolare, le variazioni tendenziali del 2006 sono state calcolate facendo riferimento al metodo micro per gli indici del 2006 ed al metodo macro per gli indici del 2005. Gli indici del 2005, a loro volta, erano stati calcolati aggiornando gli indici del 2001 in base 2000=100 con le variazioni tendenziali e, quindi, ipotizzando

implicitamente che non ci siano state modifiche rilevanti nella stagionalità rispetto al vecchio anno base 2000. La presenza, quindi, di eventuali imprese influenti con forte stagionalità potrebbe creare un *break* nella ricostruzione delle serie storiche e per tale motivo sono state escluse.

Tavola 15: *Analisi delle stime delle variazioni tendenziali nelle diverse fasi di editing – Periodo 2006. Divisione ‘servizi informatici e attività connesse’ (valori percentuali)*

Periodo	fasi del processo di <i>editing</i>			
	dati solo rispondenti	dati solo rispondenti senza <i>outlier</i>	Dati con integrazione delle mancate risposte relative alla base con volume d'affari di ASIA	Dati con integrazione delle mancate risposte relative alla base con volume d'affari di ASIA e esclusione <i>outlier</i> trimestrali
I	-3,0	-6,8	-7,1	-1,1
II	-5,2	-3,3	-11,6	-4,3
III	16,6	11,3	8,3	8,2
IV	19,4	5,7	8,9	4,1

Tavola 16: *Analisi delle stime delle variazioni tendenziali nelle diverse fasi di editing – Periodo 2006. Divisione ‘servizi all’informazione e agenzie di stampa’ (valori percentuali)*

Periodo	fasi del processo di <i>editing</i>			
	dati solo rispondenti	dati solo rispondenti senza <i>outlier</i>	Dati con integrazione delle mancate risposte relative alla base con volume d'affari di ASIA	Dati con integrazione delle mancate risposte relative alla base con volume d'affari di ASIA e esclusione <i>outlier</i> trimestrali
I	2,5	7,6	3,1	7,6
II	2,4	6,0	3,0	5,8
III	-8,3	-9,7	-9,2	-10,6
IV	0,2	3,9	-1,4	2,7

Tavola 17: *Analisi delle stime delle variazioni tendenziali nelle diverse fasi di editing – Periodo 2007-2009. Divisione ‘editoria’ (valori percentuali)*

Periodo		dati grezzi (rispondenti)		esclusione	trattamento <i>outlier</i> autorappresentazione
		dati grezzi	dati con stima MRT		
2007	I	1,4	1,4	-1,9	-0,7
	II	-2,6	-2,6	1,0	1,3
	III	0,2	-0,3	-0,1	-0,2
	IV	-1,5	-1,5	-0,4	-2,2
2008	I	-4,3	-4,8	-1,6	-3,3
	II	-8,8	-8,1	-3,1	-3,3
	III	-6,4	-6,5	-2,8	-3,6
	IV	-10,1	-10,8	-7,8	-9,5
2009	I	-5,2	-5,4	-5,1	-5,8

Questo problema mostra un limite del metodo adottato per il calcolo degli indici in quanto le ipotesi sottostanti sulla stagionalità delle serie storiche sono molto restrittive e il processo di *editing*, intercettando unità anomale a causa della forte stagionalità, rischia, nella fase del trattamento, di introdurre una distorsione nella ricostruzione delle serie storiche.

Per le stime 2007-2009, il metodo di imputazione non modifica sostanzialmente le stime mentre il trattamento degli *outlier* influenti con il metodo dell'autorappresentazione (includendo le imprese anomale nelle stime ma rettificandone il peso) le rende più robuste (ossia più vicine ai dati grezzi calcolati sui rispondenti) e, quindi, l'impatto del processo di *editing* sulle stime è più contenuto. L'effetto sulle stime, tuttavia, dipende anche dalle caratteristiche delle unità anomale, sia per quanto riguarda la componente variazione, il peso all'interno dello strato ed il peso dello strato nel dominio di stima.

Il metodo impiegato, nel caso di significative divergenze delle stime nelle diverse fasi di *editing*, prevede comunque un'analisi dettagliata a livello di strato e un ritorno sulle unità influenti, mediante metodo grafico del *box plot*, al fine di correggere eventuali *outlier* non rappresentativi.

Tavola 18: *Analisi delle stime delle variazioni tendenziali nelle diverse fasi di editing – Periodo 2007-2009. Divisione 'servizi informatici e attività connesse' (valori percentuali)*

Periodo		dati grezzi (rispondenti)	dati con stima MRT	Esclusione	trattamento outlier autorappresentazione
2007	I	-2,1	-2,1	1,2	1,2
	II	8,8	8,6	7,9	7,9
	III	3,6	3,2	2,3	2,3
	IV	7,7	7,7	5,7	5,7
2008	I	-1,7	-1,6	-0,7	-0,7
	II	3,6	3,5	2,7	2,7
	III	3,8	3,7	0,8	1,0
	IV	-4,7	-6,0	-3,6	-3,5
2009	I	-1,7	-2,8	-5,6	-5,5

Tavola 19: *Analisi delle stime delle variazioni tendenziali nelle diverse fasi di editing – Periodo 2007-2009. Divisione 'servizi all'informazione e agenzie di stampa' (valori percentuali)*

Periodo		dati grezzi (rispondenti)	dati con stima MRT	esclusione	trattamento outlier autorappresentazione
2007	I	3,0	2,9	1,0	2,8
	II	3,0	2,7	1,7	2,4
	III	3,6	3,5	3,8	3,6
	IV	10,1	9,2	4,1	6,6
2008	I	1,8	1,9	1,7	1,8
	II	7,8	7,5	4,7	6,0
	III	8,1	7,8	5,3	5,3
	IV	-4,6	-4,5	-2,9	-4,6
2009	I	1,2	0,3	-2,2	-1,8

6. Osservazioni conclusive e possibili sviluppi futuri

Il presente documento illustra la tecnica e la metodologia adottata per la nuova rilevazione trimestrale di fatturato della sezione J 'informazione e comunicazione', così come è stata definita dalla nuova classificazione delle attività economiche ATECO 2007.

Il documento illustra la progettazione e la messa a regime della nuova rilevazione e degli indicatori di fatturato che da essa derivano, analizzando nel dettaglio le più importanti fasi del processo:

- analisi descrittiva della popolazione obiettivo al fine di individuare il disegno campionario e la stratificazione più idonea a misurare la variazione del fatturato;
- la definizione della strategia campionaria da adottare per i diversi domini settoriali;
- il trattamento degli errori non campionari;

- sono invece affrontati in specifici approfondimenti, posti in appendice al documento, due problematiche che completano il quadro dei lavori effettuati per lo sviluppo della nuova indagine: la ricostruzione retrospettiva delle serie storiche relative al periodo precedente la nuova base; lo studio dei casi di *profiling* delle imprese. In particolare, si tratta del trattamento dei break strutturali, causati da trasformazioni societarie, che hanno la finalità di ricostruire il profilo di impresa e, quindi, l'unità longitudinale.

Alcuni problemi metodologici, emersi in fase di progettazione, sarebbero risolvibili solo se fosse praticabile l'uso di un archivio statistico amministrativo infrannuale, rilasciato in tempi utili per la produzione degli indici trimestrali di fatturato.

Si tratta della correzione dell'indice elementare di strato per tenere conto della demografia di impresa; dell'analisi delle trasformazioni societarie per tenere conto del loro impatto sui cambiamenti strutturali e demografici della popolazione obiettivo.

I problemi connessi alla strategia campionaria riguardano la scelta iniziale del disegno, la strategia di rotazione annuale dei rispondenti, la qualità delle stime prodotte e la scelta della dimensione campionaria.

La scelta del disegno campionario ottimale, per settori caratterizzati da mercati concorrenziali imperfetti (e, quindi, da forte asimmetria del fatturato all'interno degli strati) è quella di adottare il campionamento a probabilità variabili in base alla dimensione in termini di fatturato, ma questo disegno è più adatto a stime *cross section*, basate su campioni indipendenti, in quanto per stime panel, in occasione del *refreshment*, ciò probabilmente comporterebbe una forte riduzione della componente panel, dovuta al fatto che la probabilità di estrazione, risentendo della distribuzione del fatturato, genererebbe una differente selezione di imprese e una conseguente elevata quota di nuove entrate.

La rotazione annuale dei rispondenti finora non è stata adottata in quanto è differente il comportamento di risposta delle imprese panel rispetto a quelle nuove entrate. Tuttavia, al fine di ridurre il fastidio statistico, una bassa quota di rispondenti per strati con bassa frazione di campionamento potrebbe essere adottata, anche se il requisito qualitativo delle elevate coperture in termini di fatturato dovrebbe essere privilegiato.

Per quanto riguarda la qualità delle stime, gli aspetti trattati in questo lavoro riguardano la distorsione dovuta alle mancate risposte, la robustezza delle stime e la precisione delle stime intervallari, includendo anche i dati imputati.

La massiccia politica di prevenzione della mancata risposta totale per la stima anticipata a 60 giorni ha permesso di evitare che ci sia una elevata revisione a 90 e 180 giorni, per effetto dei rispondenti ritardatari influenti che confutano l'ipotesi MAR. Il meccanismo della non risposta, tuttavia, potrebbe generare distorsioni anche a livello di inferenza statistica qualora l'allocazione effettiva delle unità fosse molto differente da quella teorica e, quindi, dall'errore campionario desiderato. Solo una analisi della coerenza tra statistiche congiunturali e strutturali permetterebbe di misurare la distorsione da mancata risposta.

Il problema della robustezza delle stime, dovuto alla presenza di *outlier* influenti rappresentativi, è stato mitigato ricorrendo all'autorappresentazione dei dati anomali non affetti da errori non campionari, anche se restano margini di eccessiva discrezionalità nella scelta dei parametri impiegati in fase di individuazione e trattamento degli *outlier*.

La precisione delle stime intervallari, basati sullo stimatore quoziente in assenza e presenza di dati imputati, mostra come il ricorso al panel per stime longitudinali garantisce una bassa varianza e un attendibile intervallo di confidenza della stima intervallare.

La dimensione campionaria, infine, è calcolata sul livello di fatturato, mentre, per le indagini future, si potrebbe usare come variabile obiettivo la variazione del fatturato per la popolazione longitudinale che, riguardando le stesse unità, permetterebbe di ottenere una varianza più contenuta e, quindi, una dimensione campionaria ridotta per panel rappresentativi. Resta il problema della popolazione delle nuove entrate, costituita non solo dalle neonate ma anche dalle unità già esistenti che sono entrate a far parte della popolazione obiettivo. Per questa sottopopolazione, si dovrebbe calcolare un campione rappresentativo per il livello di fatturato in quanto questa tipologia di imprese entra a far parte della popolazione nell'anno corrente e non esisteva nell'anno precedente. Tali imprese, in base alla teoria

economica del ciclo di vita aziendale, hanno una dinamica del fatturato differente rispetto alle unità longitudinali e, quindi, dovrebbero appartenere a strati distinti con una struttura implicita di ponderazione espressa all'anno corrente e non a quello base. In altre parole, la decisione di tenere conto delle imprese nuove entrate implica necessariamente il passaggio da un sistema di indici a base fissa ad uno a base variabile e tale sistema non è raccomandato dall'EUROSTAT perché ritenuto troppo costoso.

Appendice A – La procedura di ricostruzione retrospettiva delle serie storiche

La ricostruzione retrospettiva delle serie storiche, riferita al periodo 2000-2005, è avvenuta in modo differente per gli indici relativi alle quattro divisioni economiche considerate, a causa delle differenti caratteristiche della matrice di corrispondenza tra la vecchia e la nuova classificazione delle attività economiche per le attività appartenenti a questi settori (ISTAT, 2008⁴⁵; Centraal Bureau voor de statistiek, 2008⁴⁶).

Per il settore delle telecomunicazioni, a livello di dominio di pubblicazione esiste una corrispondenza biunivoca tra la vecchia e nuova classificazione ma, poiché la componente panel del nuovo campione copriva la maggior parte del fatturato della popolazione, si è utilizzato l'approccio micro dal primo trimestre 2005 mentre lo slittamento della base, mediante coefficiente di raccordo⁴⁷, è stato utilizzato dal 2000 al 2004.

Per il settore dell'editoria, per il periodo 2000-2005, è stata trimestralizzata e raccordata la vecchia serie storica dell'editoria tradizionale in quanto il peso dell'edizione di *software* e di alcune categorie economiche della vecchia editoria, riclassificate altrove, era irrilevante.

Per i servizi informatici (divisione 62) e i servizi di informazione e agenzie di stampa (divisione 63), la ricostruzione è stata molto più complessa in quanto lo studio della tabella di corrispondenza (raccordo teorico) aveva messo in luce che la vecchia divisione dell'informatica (divisione 72) in parte rientrava in alcune classi della divisione 62, in parte in quelle della 63 (dove rientrano anche le agenzie di stampa, precedentemente fuori dal dominio 'industria e servizi') e in parte sono migrate nell'industria (Gismondi *et al.*, 2008⁴⁸).

Per tali motivi, una possibile matrice di transizione tra vecchia e nuova classificazione poteva essere costruita solo mettendo in relazione i vecchi gruppi economici dell'informatica (per i quali venivano prodotti indici elementari) e le nuove classi economiche. Le classi economiche dovevano poi essere ponderate in base al sistema SBS, riferito all'anno 2005.

I problemi metodologici ed empirici che emergono dall'impiego delle matrici di transizione sono i seguenti:

- 1) l'effetto della riclassificazione, dovuto all'uso di informazioni aggiuntive (provenienti dagli studi di settore del Ministero delle Finanze) nella codifica delle imprese in ATECO 2007, viene neutralizzato ignorando le imprese che provengono da altri domini di stima (fuori campo di osservazione); all'opposto, le unità che non rientrano nello schema teorico di raccordo tra vecchia e nuova ATECO ma che appartengono al dominio della vecchia informatica vengono prese in considerazione;
- 2) le matrici di transizione permettono di ottenere gli indici in nuova ATECO in base al legame esistente tra le due classificazioni (ATECO 2002 e ATECO 2007) che può essere misurato in termini di imprese o di fatturato fiscale. Si è preferito impiegare il fatturato fiscale in quanto è un totale noto più vicino al sistema di ponderazione SBS, sul quale si ottengono normalmente gli indici. Il numero di imprese, all'opposto, potrebbe generare risultati molto differenti in

⁴⁵ ISTAT (2008) *Tavola di raccordo tra ATECO 2007 e ATECO 2002*

⁴⁶ Centraal Bureau voor de statistiek (2008) *Theory on the producing of historical time series for short term business statistics in NACE rev 2 with an application in the industrial turnover index in the Netherlands (1995-2008)*

⁴⁷ Il coefficiente di raccordo è dato dal prodotto tra l'indice trimestrale in base 2000=100 e il rapporto tra l'indice annuale del 2005 in nuova base (per definizione pari a 100) e l'indice annuale del 2005 in base 2000.

⁴⁸ Gismondi, Cirianni, Panizon, (2008) *Computer programming, consultancy and related services: turnover index, workshop on backcasting, Lussemburgo, 13-14 ottobre 2008*

quanto si possono verificare casi di forte concentrazione in termini di imprese che, tuttavia, hanno un basso peso in termini di volume d'affari.

In alternativa, si sarebbe potuto impiegare il metodo micro, usando il campione teorico estratto dall'archivio ASIA in ATECO 2007 del 2005, che sarebbe stato proiettato all'indietro fino all'anno 2000. Questo metodo, tuttavia, per motivi legati alla forte componente demografica che caratterizza il settore, sarebbe stato applicato su un campione di imprese non rappresentativo della popolazione.

La matrice di transizione riferita ai servizi informatici (tavola 20) mostra come una quota, che varia dal 2 al 5 per cento a secondo delle classi economiche, proviene da settori esterni all'attuale campo di osservazione (FCO) e quindi viene ignorata nella ricostruzione retrospettiva delle stime mentre le celle evidenziate in bordo oscurato indicano i casi in cui il raccordo teorico è rispettato.

Gli indici di classe in nuova ATECO, per il periodo 2000-2005, sono quindi ottenuti come media ponderata degli indici di gruppo della vecchia informatica (a tre cifre ATECO, cioè i gruppi da 721 a 726) slittati alla base 2005=100:

$$I_{6201} = \sum_{k=1}^6 I_{72.k} * W_{6201,72.k} \quad (20)$$

Tavola 20: Matrice di transizione in termini di volume d'affari tra vecchia e nuova classificazione – servizi informatici (valori percentuali)

ATECO 2002	ATECO 2007			
	6201	6202	6203	6209
721	0,3	6,8	0,2	0,3
722	91,3	73,0	13,7	7,6
723	3,6	2,8	10,6	0,3
724	0,1	0,1	0,7	0,2
725	0,1	0,7	0,6	2,5
726	2,5	11,2	71,1	84,2
FCO	2,2	5,4	3,1	4,9
Totale	100	100	100	100

Fonte: ASIA, anno 2005

Infine, l'indice aggregato della divisione 62 è ottenuto come media ponderata degli indici di classe, usando il sistema dei pesi SBS riferiti all'anno 2005 sia per il periodo precedente a tale anno che per quello successivo:

$$I_{62} = \sum_{k=1}^9 I_{620.k} * W_{620.k} \quad (21)$$

La decisione di utilizzare il sistema dei pesi SBS dell'anno base 2005 è stata motivata dal fatto che non esisteva un sistema di ponderazione in nuova ATECO precedente al nuovo anno base.

In realtà, il ribasamento, se non ci fosse stata la migrazione ATECO, aveva l'obiettivo, come prescritto dal manuale metodologico EUROSTAT, di aggiornare gli indici con i cambiamenti avvenuti nella struttura economica della popolazione obiettivo, creando, eventualmente, un *break* naturale nella serie storica ricostruita.

Per la divisione 63 (tavola 21), è stato possibile ricostruire solo il gruppo 63.1, che pesa il 98 per cento sulla divisione e, in particolare, la classe 63.11 pesa il 95 per cento. Anche in questo caso non si è tenuto conto delle imprese che provenivano da altri settori mentre è stata considerata la ricomposizione interna della divisione, dovuta all'effetto empirico della riclassificazione, che è particolarmente

significativa nella classe 63.12 (le celle con bordo tratteggiato indicano i casi in cui non è rispettato il raccordo teorico mentre quelle con bordo oscurato hanno lo stesso significato della tavola 20).

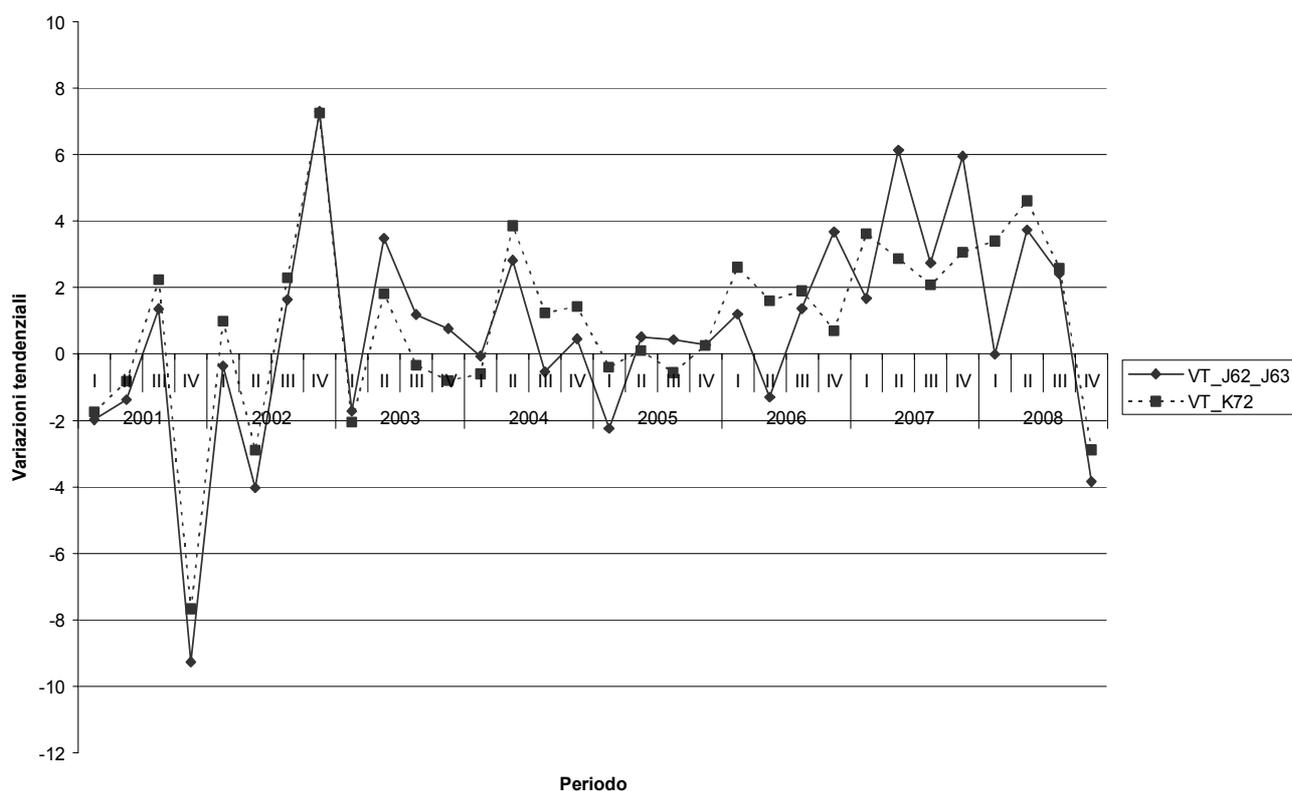
Tale classe economica ha un peso sull'indice complessivo molto basso (pari al 2,2 per cento della divisione 63 e allo 0,9 per cento della sottosezione JC⁴⁹) e, quindi, l'effetto della riclassificazione empirica ha un impatto trascurabile.

Tavola 21: Matrice di transizione in termini di volume d'affari tra vecchia e nuova classificazione – servizi all'informazione e agenzie di stampa (valori percentuali)

ATECO 2002	ATECO 2007	
	6311	6312
721	0,1	0,6
722	3,0	11,4
723	74,6	5,7
724	7,1	5,7
725	0,0	0,0
726	10,2	66,3
FCO	5,0	10,3
Totale	100	100

Fonte: ASIA, anno 2005

Figura 1: Confronto tra le variazioni tendenziali di fatturato della vecchia e nuova informatica – Periodo 2001-2008



La media ponderata tra gli indici delle divisioni 62 e 63 (con pesi rispettivamente del 70 e del 30 per cento) costituisce la nuova informatica (JC), che può essere confrontata con la vecchia informatica (K72), ritenendo trascurabili i pesi delle agenzie di stampa incluse nella nuova informatica, dell'edizione

⁴⁹ La sottosezione JC è composta dalle divisioni 62 e 63 ed è confrontabile, con un certo grado di approssimazione, con la vecchia informatica.

di *software* (confluita nell'editoria) e quello della manutenzione e riparazione di apparecchiature informatiche (confluita nell'industria) che appartenevano alla vecchia informatica.

Il grafico 1 mostra che esiste una elevata correlazione tra la serie storica delle variazioni tendenziali di fatturato della vecchia e nuova informatica.

Appendice B – Il trattamento dei casi di *profiling*

Uno dei principali problemi metodologici con cui si devono confrontare le rilevazioni longitudinali di fatturato degli altri servizi è relativo ai casi di *profiling*, definiti come la ricostruzione dei profili di imprese per tenere conto del *break* strutturale dell'attività, e quindi del volume d'affari, di singole imprese del panel che sono coinvolte da trasformazioni societarie. Queste rendono non omogeneo, e, quindi, non più significativo dal punto di vista economico il confronto del fatturato registrato al tempo corrente con quello dei periodi precedenti: ciò significa, in termini statistici, che l'impresa perde il requisito fondamentale di unità longitudinale (Lavallée, 1998⁵⁰).

In generale, le trasformazioni societarie, da un punto di vista statistico, possono essere trattate in quattro differenti modi (Lavallée, 1998⁵¹). Il primo metodo consiste nella scelta di una variabile strutturale (come ad esempio il fatturato) per la definizione dell'impresa prevalente alla quale attribuire il requisito di unità longitudinale dopo l'operazione straordinaria. Questo criterio funziona bene nei casi in cui, adottando una soglia di fatturato molto elevata, l'operazione di fusione o di scissione societaria è ininfluenza sulle stime longitudinali. Il secondo metodo consiste nel mantenere la struttura iniziale delle imprese in modo che l'impresa che nasce al tempo corrente viene ricostruita all'indietro facendo la somma dei fatturati delle unità che hanno partecipato all'operazione straordinaria. Questo metodo, di fatto, ipotizza che il *break* non sia avvenuto al tempo corrente ma nei periodi precedenti e quindi viene mantenuta inalterata la struttura iniziale dell'impresa, creando una unità longitudinale fittizia dopo l'evento della trasformazione societaria. Il terzo metodo è il caso opposto del secondo in quanto si tratta di una unità, ad esempio nuova entrata nel panel per l'appartenenza allo strato censuario delle grandi imprese. Tale impresa, alla quale vengono chiesti dati retrospettivi, non esisteva nel passato e, quindi, viene costruito il suo profilo come se fosse già grande anche nel passato, considerando tutte le unità della popolazione (presenti e assenti nel panel) che sono confluite in essa. Questo metodo è molto importante nella prassi operativa per cogliere la dinamica di unità che sono diventate nel tempo molto rilevanti. Il quarto metodo mantiene la struttura iniziale dell'impresa per le scissioni (secondo metodo) e la struttura finale per le concentrazioni.

Qui di seguito si esaminano, in forma schematica, le principali tipologie giuridiche di trasformazioni societarie (Buffelli, Sirtoli, 2004⁵²) e le regole empiriche di trattamento nel caso delle rilevazioni trimestrali di fatturato dei servizi.

I casi contemplati dalla normativa giuridica possono rientrare in quattro grandi gruppi.

Il primo gruppo riguarda casi in cui l'impresa mantiene di fatto la continuità longitudinale: si tratta di trasformazioni societarie con continuità longitudinale e di compravendita societaria. In entrambi i casi, da un punto di vista giuridico, l'impresa cambia codice fiscale identificativo ma l'impresa A che cessa continua a vivere nell'impresa B, con diversa forma giuridica nel primo caso e diverso proprietario nel secondo. Da un punto di vista statistico e operativo, l'impresa B mantiene per la serie storica retrospettiva i dati dell'impresa A.

Il secondo gruppo riguarda le concentrazioni societarie, denominate fusioni. La fusione per incorporazione viene valutata operativamente in modo diverso a seconda se le imprese coinvolte nell'operazione straordinaria appartengono o meno al campione esistente. Il primo caso è di una impresa A che viene incorporata in una impresa B presente nel campione. L'impresa B continua ad esistere e l'impresa A cessa. Se per l'impresa B il *break* strutturale è consistente, l'impresa viene

⁵⁰ Lavallée P., Struijs P. (1998) *Treatment of enterprise restructuring in business panels*, Assemblée annuale de la SSC, giugno 1998

⁵¹ Lavallée P., Struijs P. (1998) *Treatment of enterprise restructuring in business panels*, Assemblée annuale de la SSC, giugno 1998

⁵² Buffelli G, Sirtoli M. (2004) *Le operazioni straordinarie delle società*, edizione Giuffrè

momentaneamente sospesa dai calcoli e continuata a intervistare fino a quando torna ad essere una unità longitudinale. In altre parole, l'impresa A diventa dominante nell'impresa B, in termini di fatturato, e, anziché ricostruire indietro la serie storica come se il break fosse avvenuto l'anno precedente e quindi si ricostruirebbe l'unità longitudinale, si considera che il break crea una mancanza di memoria con il passato e quindi l'unità viene intervistata e sospesa fino a quando torna a soddisfare il requisito della longitudinalità. Il secondo caso è di una impresa A che viene incorporata in una impresa B non presente nel campione. In questo caso, l'impresa A cessa e l'impresa B viene intervistata come nuova entrata. Un ultimo caso più complesso riguarda l'incorporazione di una impresa A in una impresa B, che appartiene allo stesso dominio ma ad un altro settore economico (definito dal gruppo o dalla classe economica). In questo caso, la soluzione migliore è di intervistare l'impresa B come unità funzionale (come consigliato negli studi pilota del Regolamento Comunitario) al fine di mantenere le serie storiche delle due diverse modalità economiche coerenti con il sistema di ponderazione dell'anno base. La fusione con nascita di una nuova impresa è la seconda operazione societaria prevista dalla legislazione corrente e consiste nella fusione in C delle imprese A e B. Le imprese A e B cessano e l'impresa C va intervistata come nuova entrata considerando per l'anno precedente anche i valori del fatturato di A e B (secondo metodo di trattamento secondo EUROSTAT).

Il terzo gruppo riguarda le scorporazioni societarie. La normativa prevede due casi: la scissione di impresa e l'affitto o cessione di ramo di azienda. La scissione di impresa avviene quando l'impresa A viene divisa in B e C. L'impresa A cessa e le imprese B e C, se appartenenti al campo di osservazione, vengono intervistate ma sospese dai calcoli fino a quando non diventano unità longitudinali. L'affitto o cessione di ramo di azienda si ha quando l'impresa A affitta o cede il ramo di azienda all'impresa B. Se il ramo di azienda appartiene al campo di osservazione come attività prevalente dell'impresa B, l'impresa B viene intervistata e, se presente un *break* strutturale, viene momentaneamente sospesa dai calcoli fino a quando diventa una unità longitudinale. Bisogna valutare se l'impresa A, senza quel ramo di azienda, continua ad essere nel campo di osservazione e, subendo un *break*, viene intervistata ma momentaneamente sospesa dai calcoli.

Il quarto gruppo riguarda l'amministrazione controllata delle grandi imprese in crisi. Questo è un problema che potrebbe riguardare soprattutto le imprese dominanti dei settori oligopolistici in quanto in questo caso il *break* riguarda il settore e il piano industriale di ristrutturazione non ha solo effetti sull'impresa dominante ma anche su quelle appartenenti al mercato. In tali casi la eventuale perdita della quota di mercato dell'impresa dominante potrebbe essere non solo ripartita tra quelle presenti nel panel ma anche su imprese non osservate che diventerebbero statisticamente significative oppure potrebbe generare un aumento della natalità che, attualmente, è difficilmente misurabile per mancanza di un archivio aggiornato con il periodo di riferimento degli indici trimestrali. In questi casi non è possibile la ricostruzione micro ma viene ricostruito il settore a livello macro, perdendo la caratteristica del panel per la misurazione delle stime longitudinali.

In conclusione, rispetto ai quattro metodi proposti dall'EUROSTAT e descritti da Lavallée, non si cerca sempre di ricostruire l'unità longitudinale, ma spesso si preferisce sospendere l'unità coinvolta, che nasce dalle trasformazioni societarie, fino a quando torna ad essere una unità longitudinale.

Bibliografia

- Banca D'Italia *La misurazione dell'errore di stima nelle statistiche sui tassi di interesse bancari*, Supplementi al bollettino statistico. Note Metodologiche. Anno XVII Numero 36 – 14 giugno 2007
- Barcaroli G., Luzi O. e Ceccarelli C. *Il macroediting: tecniche di correzione interattiva di variabili quantitative guidata dall'analisi degli aggregati. Il caso del sistema dei conti delle imprese*. ISTAT, quaderni di ricerca, n 1/1998, pag 3-58
- Bergdahl, Black, Bowater, Chambers, Davies, Draper, Elvers, Full, Holmes, Lundqvist, Lundstrom, Nordberg, Perry, Pont, Prestwood, Richardson, Skinner, Smith, Underwood, Williams *Model quality report in business statistics*, EUROSTAT, pag. 82-160 e pag. 54-60, 1997
- Buffelli G, Sistoli M. *Le operazioni straordinarie delle società*, edizione Giuffrè, 2004
- Ceccato F. (2000) *Studio progettuale della rilevazione trimestrale di fatturato del commercio e manutenzione auto*, documento interno ISTAT
- Chambers R.L. *Outlier robust finite population estimation*, pag. 1063-1069, Journal of the American Statistical association, vol 81 n 396, 1986
- Crankshaw M., Kujawal L., Stamas G. *Recent experiences in survey coordination and sample rotation within monthly business establishment surveys*, atti relativi al *Joint statistical meetings – section on survey research methods*
- Cirianni A. e Gismondi R. *Identificazione e trattamento statistico delle osservazioni anomale nel settore degli altri servizi*, paragrafo 4.4. tratto dalla direttiva ISTAT, TRAC16 “*Sperimentazione di stime anticipate per specifici indicatori congiunturali, finalizzata al rilascio in produzione delle relative metodologie*”, 2006
- Cirianni A., Coppola S., Panizon F. *Le indagini sul fatturato degli altri servizi: i metodi di controllo e correzione*, CONTRIBUTI ISTAT N. 13/2008, pag. 87-105, 2008
- Cocchi D. “*La stima per quoziente*”, capitolo 4, estratto da “*Teoria dei campioni. Corso base*”, dispensa pubblicata su internet all'indirizzo
[/www2.stat.unibo.it/trivisano/default_files/files%20camp/dispensa%20Cocchi.pdf](http://www2.stat.unibo.it/trivisano/default_files/files%20camp/dispensa%20Cocchi.pdf)
- EUROSTAT *Methodology of short term business statistics*, pag. 33-35, 2002
- EUROSTAT *Regolamento n. 1983/2006* che integra il *Regolamento sulle statistiche congiunturali STS n. 1165/98* e il *Regolamento n. 1503/2006* sul rilascio delle stime anticipate a 60 giorni
- EUROSTAT *Regolamento comunitario sulle definizioni delle variabili statistiche n. 588/2001*
- Falorsi P.D., Ballin M., De Vitiis C., Scepi G. *Principi e metodi del software generalizzato per la definizione del disegno di campionamento nelle indagini sulle imprese condotte dall'ISTAT*, Statistica applicata, vol. 10 n. 2, pag. 235-257, 1998
- Frongia D., Martinelli R., Panizon F., Querini B., Stanco A. *Il nuovo sistema informatico altri servizi. Progetto di reingegnerizzazione dei processi produttivi delle indagini trimestrali di fatturato degli altri servizi*. CONTRIBUTI ISTAT N. 12/2007, 2007
- Gismondi R. *Riflessioni sul problema della stima del livello e della variazione tramite l'uso di informazioni complementari*, Statistica Applicata, vol. 8, 2, pag. 431-453, Rocco Curto Editore, Napoli, 1996
- Gismondi, Cirianni, Panizon, *Computer programming, consultancy and related services: turnover index*, workshop on backcasting, Lussemburgo, 13-14 ottobre 2008
- Hidiroglou M.A., Berthelot J.M., *Statistical editing and imputation for periodic Business Surveys*, Survey Methodology, 12, pag. 73-84, 1986
- ISTAT *Nota informativa delle rilevazioni trimestrali di fatturato dei servizi*, comunicato stampa del 30 giugno 2009
- ISTAT *Criteria for updating the business register*, in “*Back casting STS time series according to the NACE REV 2 classification*”, GRANT per EUROSTAT, 2008
- ISTAT *The estimation of the new weights 2005 in NACE REV 2*”, pag. 18-27, estratto da *Back casting STS time series according to the NACE REV 2 classification*, rapporto trasmesso all'EUROSTAT, 2008
- ISTAT *Tavola di raccordo tra ATECO 2007 e ATECO 2002*, anno 2008
- Lavallée, P (1997), *Theory and application of enterprise panel surveys*, Course notes sponsored by the TES Institute, Luxembourg
- Lavallée P., Struijs P. (1998) *Treatment of enterprise restructuring in business panels*, Assemblée annuelle de la SSC, giugno 1998

- Lavallée P. e Hidioglou M. (1988) *On the stratification of skewed population*, Survey methodology, 14, pag. 33-44
- LEE *Outliers in business surveys*, in Cox, Binder, Chianappa, Christianson, Colledge, Kott *Business survey methods*, John Wiley & Sons, New York, pag. 503-523, 1995
- Luzi, Di Zio, Guarnera, Manzari, De Waal, Pannekoek, Hoogland, Tempelman, Hulliger, Kilchmann *Recommended practices for editing and imputation in cross sectional business survey*, pag. 61, 2007
- Panizon, Cirianni, Congia, Garozzo, Giorgi *L'evoluzione dei processi produttivi per la gestione ed il controllo dei tassi di risposta*, Rivista di statistica ufficiale n 2-3/2007, pag. 5-31, edizione Franco Angeli, 2007
- Panizon F. , Cirianni A *The problem of outliers in short term surveys on turnover of other services*, atti del convegno intermedio SIS, 2007
- Patterson *The Horvitz Thompson estimator*,
articolo tratto dal link www.math.umt.edu/patterson/549/Horvitz.pdf
- Rizzi A. *Robustezza*, pag. 351-380, capitolo tratto da *Inferenza statistica*, edizione UTET, 1992
- Sanjay Kumar Saxena, A.K. Nigam e N.D. Shukla *Variance estimation for combined ratio estimator*, The Indian Journal of statistics, 1995, Volume 57, Serie B, pag. 85-92

Documenti ISTAT(*)

- 1/2006 – Roberta Roncati – www.istat.it (versione 3.0) *Il nuovo piano di navigazione*
- 2/2006 – Maura Seri e Annamaria Urbano – *Sistema Informativo Territoriale sulla Giustizia: la sezione sui confronti internazionali*
- 3/2006 – Giovanna Brancato, Riccardo Carbini e Concetta Pellegrini – *SIQual: il sistema informativo sulla qualità per gli utenti esterni*
- 4/2006 – Concetta Pellegrini – *Soluzioni tecnologiche a supporto dello sviluppo di sistemi informativi sulla qualità: l'esperienza SIDI*
- 5/2006 – Maurizio Lucarelli – *Una valutazione critica dei modelli di accesso remoto nella comunicazione di informazione statistica*
- 6/2006 – Natale Renato Fazio – *La ricostruzione storica delle statistiche del commercio con l'estero per gli anni 1970-1990*
- 7/2006 – Emilia D'Acunto – *L'evoluzione delle statistiche ufficiali sugli indici dei prezzi al consumo*
- 8/2006 – Ugo Guarnera, Orietta Luzi e Stefano Salvi – *Indagine struttura e produzioni delle aziende agricole: la nuova procedura di controllo e correzione automatica per le variabili su superfici aziendali e consistenza degli allevamenti*
- 9/2006 – Maurizio Lucarelli – *La regionalizzazione del Laboratorio ADELE: un'ipotesi di sistema distribuito per l'accesso ai dati elementari*
- 10/2006 – Alessandra Bugio, Claudia De Vitiis, Stefano Falorsi, Lidia Gargiulo, Emilio Gianicolo e Alessandro Pallara – *La stima di indicatori per domini sub-regionali con i dati dell'indagine: condizioni di salute e ricorso ai servizi sanitari*
- 11/2006 – Sonia Vittozzi, Paola Giacchè, Achille Zuchegna, Piero Crivelli, Patrizia Collesi, Valerio Tiberi, Alexia Sasso, Maurizio Bonsignori, Giuseppe Stassi e Giovanni A. Barbieri – *Progetto di articolazione della produzione editoriale in collane e settori*
- 12/2006 – Alessandra Coli, Francesca Tartamella, Giuseppe Sacco, Ivan Faiella, Marcello D'Orazio, Marco Di Zio, Mauro Scanu, Isabella Siciliani, Sara Colombini e Alessandra Masi – *La costruzione di un Archivio di microdati sulle famiglie italiane ottenuto integrando l'indagine ISTAT sui consumi delle famiglie italiane e l'Indagine Banca d'Italia sui bilanci delle famiglie italiane*
- 13/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: rilevazione Intrastat*
- 14/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: rilevazione Extrastat*
- 15/2006 – Ersilia Di Pietro – *Le statistiche del commercio estero dell'Istat: comparazione tra rilevazione Intrastat ed Extrastat*
- 16/2006 – Fabio M. Rapiti – *Short term statistics quality Reporting: the LCI National Quality Report 2004*
- 17/2006 – Giampiero Siesto, Franco Branchi, Cristina Casciano, Tiziana Di Francescantonio, Piero Demetrio Falorsi, Salvatore Filiberti, Gianfranco Marsigliesi, Umberto Sansone, Ennio Santi, Roberto Sanzo e Alessandro Zeli – *Valutazione delle possibilità di uso di dati fiscali a supporto della rilevazione PMI*
- 18/2006 – Mauro Albani – *La nuova procedura per il trattamento dei dati dell'indagine Istat sulla criminalità*
- 19/2006 – Alessandra Capobianchi – *Review dei sistemi di accesso remoto: schematizzazione e analisi comparativa*
- 20/2006 – Francesco Altarocca – *Gli strumenti informatici nella raccolta dei dati di indagini statistiche: il caso della Rilevazione sperimentale delle tecnologie informatiche e della comunicazione nelle Pubbliche Amministrazioni locali*
- 1/2007 – Giuseppe Stassi – *La politica editoriale dell'Istat nel periodo 1996-2004: collane, settori, modalità di diffusione*
- 2/2007 – Daniela Ichim – *Microdata anonymisation of the Community Innovation Survey data: a density based clustering approach for risk assessment*
- 3/2007 – Ugo Guarnera, Orietta Luzi e Irene Tommasi – *La nuova procedura di controllo e correzione degli errori e delle mancate risposte parziali nell'indagine sui Risultati Economici delle Aziende Agricole (REA)*
- 4/2007 – Vincenzo Spinelli – *Processo di Acquisizione e Trattamento Informatico degli Archivi relativi al Modello di Dichiarazione 770*
- 5/2007 – Anna Di Carlo, Maria Picci, Laura Posta, Michaela Raffone, Giuseppe Stassi e Fiorella Tortora – *La progettazione dei Censimenti generali 2010-2011: 1 - Analisi, valutazione e proposte in merito ad atti di normazione e finanziamento*
- 6/2007 – Silvia Bruzzone, Atonia Manzari, Marilena Pappagallo e Alessandra Reale – *Indagine sulle Cause di Morte: Nuova procedura automatica per il controllo e la correzione delle variabili demo-sociali*
- 7/2007 – Maura Giacommo, Carlo Vaccari e Monica Scannapico – *Indagine sulle Scelte Tecnologiche degli Istituti Nazionali di Statistica*
- 8/2007 – Lamberto Pizzicannella – *Sviluppo del processo di acquisizione e trattamento informatico degli archivi relativi al modello di dichiarazione 770. Anni 2004 – 2005*
- 9/2007 – Damiano Abbadini, Lorenzo Cassata, Fabrizio Martire, Alessandra Reale, Giuseppina Ruocco e Donatella Zindato – *La progettazione dei Censimenti generali 2010-2011 2 - Analisi comparativa di esperienze censuarie estere e valutazione di applicabilità di metodi e tecniche ai censimenti italiani*
- 10/2007 – Marco Fortini, Gerardo Gallo, Evelina Paluzzi, Alessandra Reale e Angela Silvestrini – *La progettazione dei censimenti generali 2010 – 2011 3 – Criticità di processo e di prodotto nel 14° Censimento generale della popolazione e delle abitazioni: aspetti rilevanti per la progettazione del 15° Censimento*
- 11/2007 – Domenico Adamo, Damiana Cardoni, Valeria Greco, Silvia Montecolle, Sante Orsini, Alessandro Ortensi e Miria Savioli – *Strategie di correzione del questionario sulla qualità della vita dell'infanzia e dell'adolescenza. Indagine multiscopo sulle famiglie. Aspetti della vita quotidiana 2005*
- 12/2007 – Carlo Nappi – *Manuale per la preparazione di originali "ready to print"*
- 1/2008 – Franco Lorenzini – *Indagine sulle unità locali delle imprese: la flessibilità organizzativa e il ruolo degli uffici regionali come strategia per la riduzione del disturbo statistico e il raggiungimento di elevati tassi di risposta*
- 2/2008 – Elisa Berntsen, Simone De Angelis, Simona Mastroluca – *La progettazione dei Censimenti generali 2010-2011 4-L'uso dei dati censuari del 2000-2001: alcune evidenze empiriche*
- 3/2008 – Marina Peci – *Progetto SCQ -Scuola Conoscenza Qualità-Statistica e Studenti*
- 4/2008 – Giampiero Siesto, Franco Branchi, Cristina Casciano, Tiziana Di Francescantonio, Piero Demetrio Falorsi, Salvatore Filiberti, Gianfranco Marsigliesi, Umberto Sansone, Ennio Santi, Roberto Sanzo e Alessandro Zeli – *Messa a regime dell'uso dei dati fiscali (Modelli UNICO) per l'integrazione delle mancate risposte e la riduzione del numero delle unità campione della rilevazione PMI*
- 5/2008 – Giovanni Seri e Maurizio Lucarelli – *A.D.ELE. Il laboratorio per l'Analisi dei Dati ELEmentari. Monitoraggio dell'attività Anni 2004-2007*

* ultimi cinque anni

- 6/2008 – Francesco Altarocca – *Strumenti informatici innovativi nella conduzione di indagini statistiche*
- 1/2009 – Silvia Dardanelli, Simona Mastroluca, Alessandro Sasso e Mariangela Verrascina – *La progettazione dei censimenti generali 2010 – 2011 5 - Novità di regolamentazione internazionale per il 15° Censimento generale della popolazione e delle abitazioni*
- 2/2009 – Rossana Balestrino e Alberto Gaucci – *Tecniche di cattura dati nei processi di produzione statistica*
- 3/2009 – Barbara Fiocco – *Le “misure” dell’Italia nell’Annuario Statistico Italiano*
- 4/2009 – Daniela Pagliuca, Raffaella Cianchetta, Marco Broccoli, Teresa Buglielli, Roberto Di Giuseppe e Diego Zardetto – *L’Osservatorio Tecnologico per i Software generalizzati (OTS) nel 2008*
- 5/2009 – Silvia Losco – *Il riuso informatico nelle Pubbliche Amministrazioni: normativa e prime esperienze in Istat*
- 6/2009 – Fabio Crescenzi Marco Fortini, Gerardo Gallo e Andrea Mancini – *La progettazione dei censimenti generali 2010 – 2011 6 - Linee generali di impostazione metodologica, tecnica e organizzativa del 15° Censimento generale della popolazione*
- 7/2009 – Silvia Losco – *Gli standard informatici dell’Istat*
- 8/2009 – Alfredo Roncaccia e Roberto Iannaccone – *L’indagine sulle Opere Pubbliche dalla costituzione dell’Istituto Centrale di Statistica ai giorni nostri*
- 9/2009 – Ugo Guarnera, Orietta Luzi e Massimo Greco – *La procedura automatica di controllo e correzione dell’indagine SPA 2007: aggiornamenti e integrazioni*
- 10/2009 – Francesca Brait, Claudia De Vitiis, Roberto Petrillo, Monica Russo, Massimo Strozza e Paola Ungaro – *L’indagine sui dottori di ricerca: un’esperienza pilota*
- 11/2009 – Marco Grazzi, Roberto Sanzo, Angelo Secchi e Alessandro Zeli – *ISTAT - MICRO.3 A new integrated system of business micro-data 1989–04*
- 12/2009 – Daniela Ichim – *Italian Farm Structure Survey: SDC Methodology for an MFR Dissemination*
- 1/2010 – Patrizia Grossi, Silvio Stoppoloni – *La progettazione dei censimenti generali 2010 – 2011 Informazione e formazione ai comuni sulle innovazioni di metodi e tecniche per il 15° Censimento generale della popolazione e delle abitazioni*
- 2/2010 – Donatella Tuzi – *L’individuazione di dati anomali nella Rilevazione Oros: un approccio in serie storiche basato su “Tramo for Errors”*
- 3/2010 – Lidia Brondi – *L’evoluzione del Sistema Statistico Nazionale nel periodo 1992-2008 attraverso l’analisi statistica delle schede identificative dei progetti Psn*
- 4/2010 – Andrea Brancatello, Luisa Ciardelli, Susi Osti, Raffaella Succi, Marina Attili e Claudia Iaccarino – *Il processo produttivo delle statistiche demografiche: le procedure e i controlli di qualità*
- 5/2010 – Maria Carla Congia e Fabio Massimo Rapiti – *Quality assessment and reporting in a short-term business survey based on administrative data*
- 6/2010 – Gerardo Gallo, Evelina Paluzzi, Angela Silvestrini e Paola Francesca Cortese – *Il confronto tra anagrafe e censimento 2001 nel Comune di Roma*
- 7/2010 – Roberto Nardecchia, Roberto Sanzo e Alessandro Zeli – *La costruzione di un panel retrospettivo di micro-dati per le imprese italiane con 20 addetti ed oltre dal 1998 al 2004*
- 8/2010 – Roberto Bartoli – *Strumenti a supporto del processo di informatizzazione degli archivi amministrativi comunali dell’anagrafe e dello stato civile ai fini delle statistiche demografiche*
- 9/2010 – Alfredo Cirianni – *La nuova rilevazione trimestrale del fatturato di alcune divisioni dell’informazione e comunicazione*