

21/1999

DOCUMENTI ISTAT

n. 21/1999

Acquisizione e trattamento dei dati nella nuova indagine sui consumi

*L. D' Aurizio, D. Grassi, S. Macchia,
A. Manzari, e N. Pannuzi*

ISTITUTO NAZIONALE DI STATISTICA - ROMA	
Piano	II
	001422
	/DOC
Inv.	3479
BIBLIOTECA	

ISTAT

SISTEMA STATISTICO NAZIONALE
ISTITUTO NAZIONALE DI STATISTICA

Le collane esistenti presso l'ISTAT - *Rivista di Statistica Ufficiale*, *Contributi ISTAT* e *Documenti ISTAT* - 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, del Sistan, o da studiosi esterni.

Per la *Rivista di Statistica Ufficiale* edita dalla casa editrice FrancoAngeli, il Comitato di redazione provvede ad ottemperare agli obblighi previsti dall'art. 1 del D.L.L. 31.8.45, n. 660. I *Contributi ISTAT* vengono fatti circolare allo scopo di suscitare la discussione attorno a 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. Gli studi pubblicati nelle diverse collane riflettono esclusivamente le opinioni degli autori e non impegnano in alcun modo la responsabilità dell'Istituto.

Il Comitato di redazione esamina le proposte di lavori da pubblicare nelle tre collane sopra indicate. Quelli pubblicati nei *Contributi ISTAT* e nei *Documenti ISTAT* sono valutati preventivamente dai dirigenti dell'Istituto, mentre quelli proposti per la pubblicazione della *Rivista di Statistica Ufficiale* sono subordinati al giudizio di referee esterni.

Al fine di definire un quadro di coerenza e di criteri armonizzati, il Comitato di redazione sottopone all'attenzione dei referee esterni gli elementi sulla base dei quali condurre la valutazione dei lavori per la pubblicazione nella *Rivista di Statistica Ufficiale*.

- I criteri guida sono:
- l'innovazione e la sistemazione ragionata dei contributi sul tema;
 - l'accuratezza nel trattamento dei dati;
 - la completezza della bibliografia citata;
 - l'accuratezza nella redazione del testo;
 - l'accuratezza nell'impostazione di tabelle e grafici.

I criteri di innovazione e di sistemazione ragionata possono essere valutati:

- rispetto allo strumento teorico con cui si analizza un tema;
- rispetto al risultato che costituisce l'obiettivo del lavoro;
- rispetto alle possibili diverse interpretazioni che lo studio rende possibile sul tema;
- rispetto ad un filone tematico o scientifico poco diffuso, sul quale si ritiene utile sollecitare un dibattito e fornire informazione scientifica;
- rispetto ad una sintesi ragionata e ad un'analisi critica di lavori di altri, sui quali ed attraverso i quali si individua lo status-quo della ricerca su un tema specifico.

Il Comitato di redazione chiede al referee esterno di argomentare, entro 3 mesi, il giudizio in un testo, delle dimensioni ritenute più opportune, contenente i commenti e le osservazioni generali e specifiche da trasmettere all'autore e inoltre, di compilare la seguente scheda di valutazione sintetica.

1. Potenzialità innovative e di sistematizzazione del lavoro:	Eccellenti <input type="radio"/>	Buone <input type="radio"/>	Sufficienti <input type="radio"/>	Insufficienti <input type="radio"/>
2. Accuratezza nel trattamento dei dati:	Eccellente <input type="radio"/>	Buona <input type="radio"/>	Sufficiente <input type="radio"/>	Insufficiente <input type="radio"/>
3. Completezza della bibliografia citata:	Eccellente <input type="radio"/>	Buona <input type="radio"/>	Sufficiente <input type="radio"/>	Insufficiente <input type="radio"/>
4. Accuratezza nella redazione del testo:	Eccellente <input type="radio"/>	Buona <input type="radio"/>	Sufficiente <input type="radio"/>	Insufficiente <input type="radio"/>
5. Accuratezza nell'impostazione di tabelle e grafici:	Eccellente <input type="radio"/>	Buona <input type="radio"/>	Sufficiente <input type="radio"/>	Insufficiente <input type="radio"/>
6. Il lavoro è:				
	Pubblicabile <input type="radio"/>			
	Pubblicabile dopo che vengano apportate le revisioni suggerite <input type="radio"/>			
	Non immediatamente pubblicabile, da modificare secondo le indicazioni suggerite e probabilmente pubblicabile <input type="radio"/>			
	Inaccettabile per la pubblicazione e non suscettibile di modificazioni che potrebbero renderlo pubblicabile <input type="radio"/>			

I Contributi ISTAT ed i Documenti ISTAT sono diffusi a cura degli autori. La *Rivista di Statistica Ufficiale* è distribuita per abbonamento dalla FrancoAngeli s.r.l. - Viale Monza 106, 20127 Milano - tel. 02/2827651 - casella postale 17175, 20100 Milano.

Il costo dell'abbonamento 1999 è di L. 90.000 per l'Italia e di L. 140.000 per l'estero da versare sul c.c.p. 17562208 intestato a FrancoAngeli s.r.l., Milano.

Direttore responsabile: Paolo Garonna
Comitato di redazione: Giulio Barcaroli Vittoria Buratta Roberta Crialesi Piero Crivelli Enrico D'Elia
Donatella Fazio Angela Ferruzza Paolo Garonna Mauro Masselli Isabella Pierantoni
Fabio Rapiti Umberto Triacca
Segretari di redazione: Donatella Fazio per la Rivista di Statistica Ufficiale
Umberto Triacca per Contributi ISTAT e Documenti ISTAT
Segreteria: Carlo Deli e Antonio Trobia

n. 21/1999

Acquisizione e trattamento dei dati nella nuova indagine sui consumi

L. D' Aurizio(*), D. Grassi(**), S. Macchia(*),
A. Manzari,* e N. Pannuzi(**)

(*) ISTAT- Servizio Studi Metodologici
(**) ISTAT- Servizio Condizioni Economiche delle Famiglie

001.422/DOC

Sommario

Nel 1997 l'indagine sui consumi delle famiglie italiane è stata profondamente ristrutturata al fine di garantire un più elevato livello di qualità delle stime; ciò ha però comportato problemi connessi alla discontinuità delle serie. E' stato quindi costituito presso l'Istat un progetto interarea con il compito di analizzare le scelte metodologiche effettuate e verificarne l'effetto sulle stime. Questo documento è frutto del lavoro svolto nell'ambito del citato progetto, limitatamente agli aspetti di acquisizione e trattamento dei dati. E' strutturato in due parti: nella prima sono descritte le innovazioni introdotte nell'indagine e nella seconda i possibili effetti distorsivi sulle stime. In particolare, nel primo capitolo vengono descritti gli aspetti relativi alla rete di rilevazione, al disegno dei questionari, all'adozione della tecnica CADI ed al trattamento delle variabili quantitative e qualitative. Nel secondo sono analizzati gli effetti sulle stime inerenti gli errori di risposta, quelli di registrazione e l'impatto delle procedure di controllo e correzione.

Abstract

The survey on Italian households expenditures was deeply restructured in 1997 with the purpose of obtaining a higher quality level. This has caused problems concerning discontinuities in the series of the estimates; for this reason a working group has been constituted with the aim of analysing the methodological changes and their impact on estimates. This paper reports the results of one year activities in this working group in so far as the aspects of the data capture and their processing are concerned. It is organised in two sections: the first one describes the innovations introduced in the survey, while the second considers their possible effects in terms of distortions of estimates. In particular, in the first section the aspects related to data collection network, questionnaire design, adoption of CADI technique and treatment of qualitative and quantitative variables are described. In the second one the effects on estimates are analysed with particular regard to response and registration errors and to the impact of editing and imputation procedures.

Introduzione¹

Il passaggio dalla vecchia alla nuova indagine sui consumi delle famiglie, avvenuto in forma compiuta nel 1997, ha comportato profonde trasformazioni metodologiche, organizzative e di processo nella conduzione dell'indagine stessa, riguardanti:

- il contenuto ed il dettaglio delle informazioni rilevate;
- i periodi di riferimento dei consumi effettuati;
- il disegno del campione;
- la rete di rilevazione;
- le modalità di registrazione dei dati;
- il sistema di controllo e correzione degli errori;
- il sistema di riporto all'universo.

Il fine ultimo per il quale si è proceduto ad una così radicale ristrutturazione dell'indagine è stato quello di migliorare la *qualità* dei dati rilevati, e conseguentemente delle stime degli aggregati di interesse prodotti (spesa totale e media, sia complessive che per i singoli capitoli e le voci elementari di spesa), qualità da intendere in primo luogo come *accuratezza*, cioè vicinanza dei dati e delle stime ai valori *veri* nella popolazione.

Tale passaggio ha prodotto rilevanti discontinuità nelle serie delle stime relative al totale delle spese effettuate dalle famiglie, sia a livello complessivo che nei diversi capitoli e voci elementari, discontinuità che nella quasi generalità dei casi sono costituite da incrementi notevoli dei livelli.

Al fine di analizzare sia le scelte metodologiche effettuate per ogni fase del processo di produzione dell'indagine, sia il corrispondente impatto in termini di differenze tra le stime degli aggregati ottenibili a partire dai dati rilevati con la vecchia e la nuova indagine, è stato istituito un progetto interarea ("La nuova indagine sui consumi delle famiglie: metodologia e analisi di qualità") con l'ordine di servizio n.195/97. Il relativo gruppo, costituito da ricercatori e collaboratori tecnici di diverse realtà dell'Istituto, ha iniziato i suoi lavori il 1.12.1997 e li ha conclusi il 30.11.1998.

Il presente lavoro riporta i contributi forniti da vari componenti del progetto limitatamente alle fasi di **acquisizione e trattamento dei dati** ed è articolato in due parti: nella prima si procede ad illustrare le innovazioni introdotte nell'impianto dell'indagine (*attività di ricognizione e documentazione*), nella seconda si procede ad analizzare gli elementi innovativi potenzialmente responsabili di sistematiche distorsioni nelle stime (*attività di analisi*).

¹ Nell'ambito del presente lavoro i paragrafi 1.1, 1.2.2.1, 1.3.1.2, 1.4.1, 1.4.2, 1.4.3, 1.4.5, 1.4.6 e 1.4.9, sono opera di D. Grassi, i paragrafi 1.3.1, 1.3.1.1, 1.3.2 e 2.2 sono opera di S. Macchia, i paragrafi 1.4.4, 1.4.8, 2.1, 2.3 e 2.4 sono opera di A. Manzari, i paragrafi 1.2.1, 1.2.2.2, 1.2.3, 1.2.4, 1.4.7.1, 1.4.7.2 e 1.4.7.4.2 e 1.4.10 sono opera di N. Pannuzi, i paragrafi 1.4.7.3, 1.4.7.4 e 1.4.7.4.1 sono opera di L. D'Aurizio.

1 Documentazione delle scelte di innovazione metodologica ed organizzativa

1.1 La rete di rilevazione

1.1.1 La riorganizzazione della rete nella nuova indagine sui consumi delle famiglie

Con la ristrutturazione dell'indagine sui consumi delle famiglie è stato rivisto tutto l'impianto della rilevazione nel tentativo di migliorarne la qualità globale ed in particolare di controllare l'insieme delle cause che in genere incrementano l'errore di tipo non campionario. L'obiettivo è stato quello di arrivare ad una indagine in cui il processo di produzione dei dati sia strettamente monitorato ed analizzato in tutte le sue fasi e nel quale il controllo e la validazione dei dati deve avvenire sempre più in prossimità della fonte dell'informazione e sempre minori dovranno essere gli interventi a posteriori sul dato registrato.

La rilevazione sul campo, fase fondamentale del processo di produzione dei dati, non è di facile controllo soprattutto a causa dell'autonomia dei Comuni nella scelta e nella formazione dei rilevatori, i quali è possibile che adottino procedure diverse da quelle prescritte. Il miglioramento della rete di rilevazione è stato dunque assunto, in fase di revisione dell'indagine, come obiettivo prioritario di tutta l'operazione di ristrutturazione. Ci si è così mossi verso una razionalizzazione di tutte le operazioni di rilevazione sul campo. In tale ambito è stato riorganizzato tutto l'insieme delle procedure di controllo della rete attraverso monitoraggi sistematici che hanno previsto tra l'altro contatti frequenti con gli organi periferici (UU.RR. Istat ed Uffici di Statistica dei Comuni), reinterviste telefoniche alle famiglie per il controllo dell'operato dei rilevatori, l'istituzione di un numero verde di ausilio alle famiglie.

Standardizzando quanto più possibile tutte le operazioni connesse con la rilevazione del dato, si è tentato di individuare quei casi in cui la rilevazione viene effettuata secondo modalità non corrette, di contenere la mancata risposta parziale e di far fronte alla mancata risposta totale.

Sforzi particolari sono stati compiuti nel tentativo di aumentare il grado di coinvolgimento di tutti coloro che intervengono nella fase di rilevazione ed in modo specifico degli addetti comunali e dei rilevatori attraverso una idonea attività di formazione e di ausilio.

Si è tentato inoltre di stimolare alla partecipazione gli intervistati fornendo informazioni, aiuto e rassicurazioni dirette sull'importanza dell'indagine e la salvaguardia dell'informazione da loro fornita.

Al fine di prevenire e correggere eventuali errori di impostazione, di omissione, di reticenza o di altro tipo, commessi da coloro che fanno parte della rete periferica di rilevazione, si è messa in atto una strategia complessa di supporto e di controllo. In particolare attraverso tale strategia si sono voluti controllare tre aspetti fondamentali per la qualità di una indagine:

1. le modalità di effettuazione della rilevazione
2. la mancata risposta parziale (casuale e sistematica)
3. la mancata risposta totale (casuale e sistematica).

In modo più specifico le azioni intraprese possono essere schematizzate come segue:

- ⇒ **Azioni di supporto**
 1. disponibilità al contatto telefonico e diretto (attraverso personale specializzato) con tutti gli addetti comunali che si rivolgono al nostro ufficio
 2. attivazione di un numero verde (gestito da personale appositamente formato) per rassicurare ed aiutare le famiglie coinvolte nell'indagine
- ⇒ **Azioni di prevenzione**
 1. enfaticizzazione della fase di istruzione ai rilevatori
 2. tentativo di aumentare il grado di coinvolgimento degli operatori e delle famiglie
 3. interventi sul questionario (domande filtro, cartoncino del reddito, cartoncino dei beni durevoli, ecc.) per ridurre la mancata risposta parziale
 4. pubblicizzazione dei controlli sull'operato dei rilevatori
- ⇒ **Azioni di controllo**
 1. monitoraggio dei tempi di avvenuta spedizione del materiale
 2. controllo quantitativo immediato (rispondenza tra campione teorico ed effettivo, correttezza dei codici identificativi dei modelli, rispetto dei periodi di rilevazione, rispetto del passo di estrazione, ecc.); in questa fase, in presenza di errori o situazioni poco chiare, vengono attivati contatti tempestivi con i Comuni ed effettuati interventi immediati sul materiale cartaceo pervenuto
 3. revisione qualitativa del materiale pervenuto attraverso regole e procedure standardizzate e personale specializzato (vengono utilizzate regole standard di revisione estremamente particolareggiate e precise²; il criterio di base è comunque quello della salvaguardia della spesa riportata dalle famiglie); in caso di necessità (valori outlier, mancanti o incompatibili) sono possibili nuovi contatti con i rilevatori o con le famiglie stesse
 4. resoconti mensili agli uffici regionali sull'andamento dell'indagine nei comuni di competenza
 5. controlli telefonici, effettuati da società esterne, sull'operato dei rilevatori attraverso reintervista di un sottocampione delle famiglie mensilmente intervistate; la reintervista telefonica non ripropone i quesiti dell'indagine, ma effettua unicamente controlli sulla modalità di svolgimento della rilevazione e sulla disponibilità della famiglia a partecipare all'indagine

Gli accorgimenti messi in atto mostrano effetti positivi sui risultati di indagine sia in termini di frequenze di spesa sia in termini di generale collaborazione da parte delle famiglie e degli Uffici di Statistica dei Comuni.

1.1.2 La mancata risposta totale

La numerosità campionaria effettiva dell'indagine (sulla quale si basano le stime) è sempre inferiore a quella teorica programmata. Ciò si verifica per tre motivi fondamentali:

² Ceccarelli, Grassi, Panuzi (1997)

1. alcuni Comuni campione non vogliono e/o non possono partecipare alla rilevazione (caduta di 1° stadio);
2. non si riesce ad intervistare le famiglie selezionate dai Comuni (caduta di 2° stadio);
3. la scarsa qualità delle risposte fornite dalle famiglie che hanno partecipato all'indagine, non permette l'utilizzo dei questionari ricevuti dall'Istat che vengono in tali casi annullati (caduta di 2° stadio).

Sia la caduta delle unità di primo stadio (UPS) che quella delle unità di secondo stadio (USS) possono avere effetti distorcenti in fase di stima dei principali parametri di indagine ed è pertanto necessario valutare, per entrambi i casi, l'entità del problema e le eventuali possibilità di intervento.

1.1.2.1 La caduta delle UPS

L'«effetto di rinnovamento» dovuto al totale svecchiamento dell'indagine oltre agli accorgimenti adottati sulla rete di rilevazione hanno contribuito a ridurre notevolmente la caduta delle UPS (tabella 1).

Tabella 1: Percentuale di cadute dei comuni e delle relative famiglie

	Comuni	Famiglie
Vecchia indagine		
1996	5,36	4,71
1° trim. 1997	6,74	5,88
Nuova indagine		
1997	4,56	2,93

I principali interventi adottati per ottenere tale risultato sono identificabili nell'opera di sensibilizzazione dei comuni, nel maggiore coinvolgimento degli UU.RR. con conseguente incremento della loro presenza sul territorio e nell'istituzione, a livello centrale, di «Responsabili di Regione» che hanno il compito di seguire i comuni di competenza in tutte le fasi della rilevazione.

Nel momento in cui si viene a conoscenza della caduta di una UPS i possibili interventi vanno distinti a seconda che l'unità sia «non autorappresentativa» (NAR) o «autorappresentativa» (AR) ed in base al fatto che la caduta debba considerarsi definitiva piuttosto che temporanea.

Se infatti per entrambe le tipologie Comunali è possibile intervenire tentando di trasformare una caduta definitiva in una caduta temporanea, recuperando cioè il Comune che non può o non vuole più eseguire l'indagine, solo per i comuni NAR è possibile, secondo modalità precise, procedere ad eventuale sostituzione.

In tale ambito al momento dell'estrazione dei comuni campione, viene effettuata una prima ricognizione, con il coinvolgimento degli UU.RR., al fine di verificare la presenza di eventuali situazioni che possono influenzare negativamente la buona riuscita dell'indagine (ad esempio presenza contemporanea di più indagini Istat, ristrutturazione dell'Ufficio di statistica, carenza di personale, ecc.). In seguito a questa operazione è possibile, all'inizio del ciclo di indagine, per i comuni del gruppo NAR, procedere ad alcune sostituzioni nell'ambito dello strato di appartenenza, dopo aver tentato di rimuovere gli eventuali ostacoli di partecipazione.

Per i comuni AR, non essendo possibile nessuna sostituzione, l'attività di organizzazione e di persuasione a collaborare all'indagine diventa un punto cruciale ed è proprio in tal senso che è stata indirizzata l'attività dell'Unità Operativa e degli UU.RR.

1.1.2.2 La caduta delle USS

La presenza di unità non rispondenti in un'indagine statistica ed in particolare nel caso di un'indagine campionaria, determina consistenti distorsioni nelle stime.

In letteratura esiste una varietà di metodi per il trattamento della mancata risposta totale; nella pratica si ricorre in generale a due procedure:

1. nel caso in cui si ipotizzi che il fenomeno di mancata risposta sia di origine casuale si adotta il sovracampionamento (sotto l'assunzione che i rispondenti si distribuiscano in modo analogo ai non rispondenti);
2. nel caso invece che la mancata risposta non venga considerata casuale o si debba comunque garantire la numerosità campionaria, si ricorre ad una sostituzione delle unità rispondenti.

Nell'indagine sui consumi delle famiglie viene adottata questa seconda procedura nell'ambito di unità selezionate casualmente dalle liste anagrafiche.

Il processo di sostituzione genera necessariamente distorsioni nelle stime connesse con le diverse caratteristiche dei rispondenti, dei non rispondenti e delle famiglie sostitutive. Per tentare di eliminare di tale distorsione, nel processo di sostituzione si richiede che le famiglie sostitutive presentino caratteristiche simili a quelle delle famiglie sostituite.

I nominativi selezionati dalle liste anagrafiche, che costituiscono la base di campionamento, sono associati unicamente alla numerosità familiare e alla sezione di censimento di appartenenza.

Queste due variabili costituiscono dunque le uniche informazioni di cui si può tenere conto in fase di sostituzione. È tuttavia noto che queste sono entrambe correlate con la spesa media mensile familiare. Si è allora stabilito che il responsabile comunale nel momento in cui si accinge ad effettuare una sostituzione debba selezionare le famiglie sostitutive tentando in primo luogo di riproporre la stessa numerosità familiare ed in secondo luogo una zona di residenza simile.

1.1.3 Modalità di raccolta delle informazioni

La selezione delle famiglie viene effettuata in modo casuale sulla base delle liste anagrafiche comunali. Nella vecchia indagine, per garantire la possibilità di sostituzione delle famiglie cadute, era prevista l'estrazione di un campione maggiorato del 50% rispetto alla numerosità teorica prevista. Nella nuova indagine, ogni responsabile comunale compila due liste di nominativi la prima delle quali rappresenta l'elenco base, la seconda, l'elenco suppletivo. Ad ogni caduta dell'elenco base viene associata una sostituzione ovvero una famiglia dell'elenco suppletivo che presenti caratteristiche uguali o simili a quelle della famiglia-base da sostituire. Tali passaggi vengono documentati attraverso un modello specifico (mod. W/404 e mod. BF/5 rispettivamente nella vecchia e nella nuova indagine) compilato dal responsabile comunale, sul quale il rilevatore deve annotare oltre alle caratteristiche delle famiglie (base e riserva) anche il motivo di mancata risposta.

Le informazioni in essi contenute possono essere utilizzate al fine di descrivere il fenomeno della non risposta, di valutare eventuali variazioni distributive imputabili alle differenti strategie di indagine e di individuare possibili interventi atti a ridurre il fenomeno stesso.

I modelli W/404 e BF/5 così come predisposti inducono tuttavia ad una sottostima della mancata risposta in quanto sono stati ideati per valutare le cadute attribuibili a famiglie dell'elenco base, le uniche che ricevono la lettera di preavviso del Presidente.

1.1.4 Alcuni risultati

Attualmente i dati disponibili riguardano, per quanto riguarda la vecchia indagine, il 1995 ed il 1996, e per quanto riguarda la nuova il 1997.

Nel seguito viene riportato il tasso complessivo di mancata risposta, calcolato come rapporto tra il numero delle famiglie sostitutive ed il numero complessivo delle famiglie effettivamente rilevate.

Regione	Vecchia indagine		Nuova indagine
	1995	1996	1997
Piemonte	14,5	12,7	18,2
Valle d'Aosta	4,4	10,2	15,3
Lombardia	14,6	16,1	22,8
Trentino A.A.	10,7	12,3	15,9
Veneto	6,6	8,6	14,0
Friuli V. Giulia	19,9	16,0	23,4
Liguria	9,8	6,5	14,2
Emilia Romagna	12,7	12,5	17,5
Toscana	9,5	10,8	16,2
Umbria	5,6	6,0	13,8
Marche	2,9	3,0	10,9
Lazio	7,2	7,0	7,8
Abruzzo	4,0	7,0	7,9
Molise	7,5	6,5	12,7
Campania	4,5	5,2	15,5
Puglia	3,8	4,9	11,3
Basilicata	1,6	2,7	7,8
Calabria	9,4	8,5	16,0
Sicilia	6,3	8,3	14,1
Sardegna	7,9	10,0	11,9
ITALIA	8,7	9,2	14,9

Tabella 1 - Famiglie base sostituite sul totale delle intervistate - Vecchia e nuova indagine

Nel passaggio dalla vecchia alla nuova indagine, si osserva un incremento complessivo di mancata risposta, in alcuni casi anche piuttosto consistente (Liguria, Umbria, Marche, Molise, Campania, Puglia, Basilicata, Calabria).

La diminuzione delle interviste false riscontrata nel passaggio dalla vecchia alla nuova indagine (dal 28% al 16% circa), insieme al fatto che i livelli di non risposta riscontrati si avvicinano al quelli osservati negli altri paesi europei, permette tuttavia di interpretare i dati riportati nella tabella 2 come un miglioramento della qualità sul lavoro sul campo.

1.2 Il modello di rilevazione³

La nuova indagine sui consumi delle famiglie è frutto di una radicale trasformazione del processo di indagine, in particolare di un profondo cambiamento nella struttura e nel contenuto dei modelli di rilevazione: libretto degli acquisti, taccuino degli autoconsumi e riepilogo delle spese.

Tutti e tre i questionari erano già utilizzati per la vecchia indagine sui consumi, ma sostanzialmente cambiati risultano sia i contenuti sia l'allocazione delle varie voci di spesa che costituiscono l'obiettivo primario dell'indagine.

1.2.1 Libretto degli acquisti

Il libretto degli acquisti rappresenta lo strumento per rilevare spese per consumi di frequenza elevata (poco meno che giornaliera), spesso associata anche ad un'entità molto ridotta. Il libretto viene giornalmente auto-compilato dai componenti della famiglia e si riferisce ad un periodo di sette giorni. Le famiglie da intervistare mensilmente nel singolo comune sono suddivise in due sottogruppi che compilano il libretto rispettivamente nel primo e nel secondo periodo di riferimento. Detti periodi, della durata di sette giorni, vengono selezionati *a priori* affinché tutti i comuni della stessa regione abbiano gli stessi periodi di riferimento ed affinché tutti i giorni del mese di rilevazione risultino coperti a livello nazionale. La sostanziale differenza che caratterizza il libretto della nuova indagine rispetto a quello della vecchia è la lunghezza del periodo di riferimento (precedentemente di dieci giorni) ed il fatto che le quantità alimentari consumate non sono più oggetto di rilevazione. Inoltre il confronto tra i libretti delle due indagini evidenzia un sostanziale cambiamento delle voci di spesa rilevate, in quanto alcune di queste sono diventate oggetto di intervista finale, mentre altre hanno subito il trasferimento opposto.

In particolare si può fare una distinzione tra i generi alimentari e i non alimentari.

I generi di consumo alimentari sono rilevati totalmente tramite libretto per entrambe le indagini. Le voci che presentano esattamente le stesse specifiche sono le seguenti:

Riso, Zucchero, Gelati, Carne di vitello e vitellone, Carne di manzo, Carne di maiale, Pollame, Olio di oliva, Olio di semi, Burro, Margarina, Yogurt, Pomodori freschi, Pomodori in scatola, Patate, Legumi freschi, Legumi conservati, Ortaggi freschi, Ortaggi conservati, Agrumi.

Per le altre voci si osserva in generale un aumento delle specifiche o un diverso accorpamento delle varie voci (si veda allegato 1).

In particolare, nella vecchia indagine alcune delle voci alimentari venivano rilevate tramite risposta aperta; queste non venivano cioè pre-stampate sul libretto destinato alla famiglia, ma venivano eventualmente da questa annotate in uno degli spazi lasciati in bianco, previsti per ciascuna categoria di beni; tali voci venivano successivamente codificate dal rilevatore e trascritte nello spazio predisposto del riepilogo delle spese. Risultano rilevate con tale tecnica le spese per consumo relative a: Farina, Carne equina, Carne ovina e caprina, Conigli, tacchini, quaglie, fagiani, etc., Carne in scatola, Lardo, strutto e simili, Frutta secca, Frutta conservata, Succhi di frutta, Altre bevande non alcoliche.

³ Lucia Fadioni ha curato le elaborazioni sui dati di base necessari allo sviluppo dell'analisi

Tale diverso metodo di rilevazione risulta avere un impatto decisamente non trascurabile sulla frequenza di spesa piuttosto che sull'entità della spesa stessa. Nella nuova indagine, infatti, le suddette voci sono pre-stampate nel libretto mantenendo le stesse specifiche.

In tab.1 si riporta il confronto tra le frequenze di spesa per la vecchia e la nuova indagine relativo a tutti i beni del libretto che presentano le stesse specifiche.

Voci	Vecchia	Nuova
Farina	17,7	45,3
Carne equina	3,8	7,7
Carne ovina	3,9	11,4
Conigli, tacchino, etc.	11,0	28,3
Lardo	0,6	3,2
Frutta secca	5,8	16,2
Frutta conservata	2,8	5,7
Riso	52,9	46,7
Zucchero	65,0	66,3
Gelati	14,8	19,8
Carne di vitello, vitellone	67,3	62,3
Carne manzo	50,3	40,1
Carne suina	45,9	43,5
Pollame	68,2	63,4
Olio di oliva	56,2	55,8
Olio di semi	39,6	28,7
Burro	45,3	45,0
Margarina	12,6	13,7
Yogurt	39,2	48,1
Uova	73,3	70,5
Pomodori freschi	50,6	53,8
Pomodori in scatola	41,1	36,8
Patate	61,1	56,8
Legumi freschi o surgelati	32,5	38,5
Legumi secchi o conservati	30,1	25,8
Ortaggi freschi o surgelati	79,5	84,1
Ortaggi secchi o conservati	24,3	19,2
Agrumi	82,4	82,3

Tab. 1 Percentuale di famiglie che hanno effettuato la spesa di alcuni generi alimentari rilevati tramite libretto-1 trim. 1997

Le voci direttamente confrontabili mostrano in generale percentuali di acquisto molto prossime a quelle della vecchia indagine. Spesso risultano inferiori in quanto a) la durata del libretto è diversa ed b) il maggior dettaglio presente per le altre voci del libretto nella nuova indagine, ha comportato anche una diversa scelta ed allocazione della spesa effettuata.

Per le spese non alimentari rilevate tramite libretto, le voci che presentano la stessa specifica si riducono alle seguenti: Tabacchi, Scope, guanti, fiammiferi, Barbieri e parrucchieri, Giornali, riviste e fumetti, Quaderni e cancelleria, Benzina, Gasolio, Gettoni telefonici, Francobolli e spese postali, Pellicole fotografiche, Dischi e video, Mangimi ed altre spese per gli animali domestici, Piante e fiori, Giocattoli, Totocalcio, Lavanderia e tintoria (ancora una

volta per le altre voci si rimanda all'appendice 1). Le differenze in termini di frequenza di spesa tra la vecchia e la nuova indagine risultano molto ridotte, evidenziando quindi anche un deciso effetto stanchezza nella compilazione del vecchio libretto; va inoltre ricordato che nella vecchia indagine le voci relative al libretto degli acquisti venivano riportate dal rilevatore in un'unica sezione del riepilogo delle spese, operazione che comportava una perdita di voci in media stimata intorno al 4%.

Voci	Vecchia	Nuova
Tabacchi	39,2	40,2
Scope, Guanti, Gomma, fiammiferi	33,2	30,3
Barbiere, parrucchiere	44,4	55,6
Giornali, riviste, fumetti	68,5	69,8
Quaderni, matite, cancelleria	23,0	23,4
Benzina per auto	72,3	73,5
Gasoli ed altri	10,0	11,7
Gettoni telefonici	26,8	25,8
Spese postali	19,9	19,5
Pellicole, Spese sviluppo e stampa	11,6	11,0
Dischi, video	14,2	14,9
Mangimi ad altro per animali domestici	17,4	21,0
Piante e fiori	29,2	26,0
Giocattoli	12,3	15,1
Totocalcio	33,5	33,0
Lavanderia	31,2	35,2

Tab. 2 Percentuale di famiglie che hanno effettuato la spesa per alcuni generi non alimentari rilevati tramite libretto.

1.2.2 Riepilogo delle spese

Il riepilogo delle spese è il questionario che il rilevatore compila durante l'intervista finale alla famiglia. Attraverso tale questionario vengono rilevate le caratteristiche socio-demografiche, abitative o di altro genere delle famiglie intervistate oltre ad alcune spese riguardanti beni e servizi di entità più elevata rispetto a quelle del libretto, spesso con una frequenza ridotta. Il periodo di riferimento infatti è mensile nella maggior parte dei casi, ma anche bimestrale, trimestrale od addirittura annuale.

Si possono quindi distinguere nell'ambito del questionario due gruppi di variabili:

- le variabili qualitative
- le variabili quantitative.

1.2.2.1 Variabili qualitative

Il vecchio ed il nuovo questionario mostrano alcune differenze sostanziali nella rilevazione delle variabili qualitative.

In particolare con il nuovo questionario vengono modificate sia la sezione del questionario relativa alle notizie sui componenti della famiglia, sia quella relativa alle notizie sull'abitazione e sul possesso dei beni durevoli.

In dettaglio per quanto riguarda le notizie sui componenti familiari vengono apportati i seguenti cambiamenti:

- la rilevazione della variabile età si riferisce attualmente all'anno di nascita e non più agli anni compiuti;
- è stata introdotta la rilevazione dello stato civile;
- il titolo di studio offre un maggior dettaglio includendo anche specifiche riguardo a titoli di ordine superiore;
- la condizione professionale, la branca di attività economica e la posizione nella professione presentano un maggior dettaglio;
- è stata soppressa una sezione delle notizie individuali che raccoglieva informazioni riguardo a:
 - ricerca di lavoro
 - ore lavorate
 - pratica di sport
 - pasti fuori casa (attualmente si rileva unicamente la spesa nel "libretto degli acquisti");
- alcune voci che prima venivano rilevate a livello individuale sono ora rilevate a livello familiare:
 - mezzo principale di sostentamento (con diversa specifica delle voci)
 - vacanze (per le quali cambia anche la definizione stessa della variabile, attualmente si rilevano le vacanze con almeno un pernottamento contro le 4 notti della vecchia indagine)
 - reddito (composizione percentuale).

Per quanto concerne l'abitazione vengono rilevate pressoché le stesse variabili eccezion fatta per il tipo di contratto nel caso dell'affitto, precedentemente non rilevato.

Riguardo al possesso di beni durevoli non esistono sostanziali differenze tra le voci, mentre muta radicalmente la loro posizione nel questionario. Infatti mentre nel nuovo questionario i singoli beni vengono raggruppati in "capitoli di spesa" secondo le caratteristiche dei beni stessi, in precedenza questi venivano elencati in un unico riquadro.

L'analisi della non-risposta parziale è sicuramente quella più adeguata per verificare l'effetto di tali cambiamenti. È risultata tuttavia inapplicabile in quanto non si dispone del file grezzo (senza modifiche dovute alla revisione manuale) dei dati per la vecchia indagine. Non è stato dunque possibile verificare l'esistenza o meno dell'effetto dei cambiamenti descritti sulla frequenza della mancata risposta.

Le uniche verifiche possibili hanno riguardato i file dei dati puliti, per quali si sono messe a confronto le distribuzioni di alcune variabili nell'intento di stabilire se per le principali caratteristiche strutturali esiste comunque uniformità di base tra le due indagini. Ovviamente le differenze eventualmente riscontrate possono essere attribuite in parte anche all'effetto dei piani di correzione dei dati. Tuttavia, per le variabili messe a confronto, gli interventi di correzione ed imputazione sono molto contenuti, il che rende quindi significativi i risultati ottenuti anche in tale contesto.

Per il primo trimestre 1997 sono state dunque elaborate le distribuzioni di quelle variabili che risultano confrontabili nelle due indagini.

Nella nuova indagine risultano meno frequenti le famiglie numerose (5 e più componenti). Rispetto alla vecchia indagine si incrementa lievemente la percentuale di abitazioni fornite di telefono (che passa dall'89,8% al 92,4) e di box per auto, risulta inoltre una distribuzione relativa alla tipologia di riscaldamento dell'abitazione maggiormente concentrata nella modalità "riscaldamento autonomo (60,5% contro il 57,3 della vecchia indagine).

Per quanto concerne il numero di stanze, con la nuova indagine si evidenzia una distribuzione lievemente più spostata sulla coda destra.

Numero comp.	Vecchia indagine	Nuova indagine
1	20,5	22,8
2	24,7	25,5
3	22,1	24,0
4	21,4	20,2
5	8,8	6,0
6+	2,5	1,5
Totale (=100)	20.120.277	21.449.447

Tab. 3: Famiglie secondo il numero di componenti

Numero comp.	Vecchia indagine	Nuova indagine
1	0,4	0,5
2	8,5	7,2
3	24,6	20,8
4	31,8	33,3
5	19,3	22,8
6+	15,4	15,4
Totale (=100)	20.120.277	21.449.447

Tab. 4: Famiglie secondo il numero di stanze dell'abitazione

Servizio	Vecchia indagine	Nuova indagine
WC	99,7	99,8
Bagno	99,2	99,8
Acqua potabile	98,9	98,5
Acqua calda	99,1	99,1
Riscaldamento	92,1	92,8
Risc. autonomo	57,3	60,5
Telefono	89,8	92,4
Box auto	48,7	52,1

Tab. 5: Percentuale di famiglie che possiedono alcuni servizi dell'abitazione

E' interessante notare che nel caso del possesso dei beni durevoli, la percentuale di famiglie che godono del bene è sistematicamente maggiore per la nuova indagine rispetto alla vecchia.

Servizio	Vecchia indagine	Nuova Indagine
Frigorifero	97,8	98,9
Lavastoviglie	23,4	27,3
Lavatrice	93,9	95,1
Aspirapolvere	52,5	69,0
Condizionatore d'aria	4,5	5,8
Telefono cellulare	13,4	13,5
Videoregistratore	54,4	60,4
Personal Computer	13,5	17,6

Tab. 6: Percentuale di famiglie che possiedono di alcuni beni durevoli

1.2.2.2 Variabili quantitative

Ancora una volta il confronto tra la nuova e la vecchia indagine mette in evidenza come la gran parte delle voci presentino nella nuova indagine un maggior dettaglio. Su un totale di 276 voci per la nuova e 192 per la vecchia, solo 10 risultano essere direttamente confrontabili, sia in termini di specifiche sia in termini di riferimento temporale (mensile, tab.3).

Appare evidente come, ad eccezione dell'affitto e del fitto figurativo che rimangono pressoché costanti, le altre frequenze nella nuova indagine siano più elevate rispetto a quelle della vecchia, indicando forse una maggiore attenzione del rilevatore ed una maggiore facilità di compilazione del nuovo questionario rispetto al vecchio (dovuta anche ad una diversa veste grafica).

Voci	Vecchia	Nuova
Affitto e subaffitto	22,9	20,7
Fitto figurativo	77,1	79,2
Quadri, specchi	1,6	2,3
Tappeti e moquettes	1,5	1,8
Macchine per cucire	0,1	0,2
Altri app. elettrici	3,6	5,7
Ricoveri in ospedale	0,2	0,9
Ricoveri in case di riposo	0,2	0,1
Lubrificanti per auto	6,4	13,2

Tab.3 Percentuale di famiglie che hanno effettuato la spesa per alcuni beni alimentari rilevati tramite riepilogo.

Di seguito si riportano le voci che, pur avendo la stessa specifica, hanno subito un cambiamento nel riferimento temporale; il primo gruppo di beni si riferisce a quelli che pur avendo lo stesso riferimento temporale (trimestrale), sono leggermente cambiate nella metodologia di rilevazione: nella vecchia indagine si chiedeva l'importo e la data di acquisto entro gli ultimi 12 mesi e venivano successivamente selezionate quelle relative agli ultimi tre

mesi; nella nuova si chiede direttamente l'acquisto nell'ultimo trimestre. Il secondo gruppo si riferisce ai beni passati da un importo mensile ad uno trimestrale.

Voci	Vecchia	Nuova
Frigoriferi	0,7	0,8
Lavastoviglie	0,3	0,4
Lavatrice	1,3	0,9
Condizionatori	0,07	0,1
Macchine per cucire	0,04	0,2
Telefono cellulare	1,7	1,8
Acqua	38,8	61,8
Gas da rete	61,2	62,4
Gas in bombole	16,2	29,7
Kerosene	3,3	8,1
Carbone e legna	4,4	9,8
Riscaldamento	12,13	18,9

Tab.4 Percentuale di famiglie che hanno effettuato la spesa per alcuni beni non alimentari rilevati tramite riepilogo (con riferimento mensile per la vecchia indagine e trimestrale per la nuova)

Appare abbastanza evidente come il diverso periodo di riferimento abbia un impatto ancora una volta non trascurabile sulle frequenze di acquisto; più lungo è il periodo, maggiore è la frequenza rilevata, soprattutto per spese di carattere più regolare (spese per abitazione). Le spese di carattere più raro ed eccezionale, mostrano una frequenza estremamente ridotta da non rendere significativo il confronto proposto.

1.2.3 Taccuino degli autoconsumi

Il modello relativo agli autoconsumi, anch'esso autocompilato dalla famiglia stessa, non presenta sostanziali cambiamenti sia nella struttura sia nel tipo di spese rilevate; essenzialmente nella nuova si ha un maggiore dettaglio ed una maggiore accuratezza nella definizione e nella scelta delle voci pre-stampate.

In conclusione la nuova indagine presenta una sostanziale modifica nella struttura e nell'aspetto dei modelli di rilevazione utilizzati; in particolare sono stati modificati sia le specifiche delle voci considerate sia il riferimento temporale, per facilitare il ricordo e per catturare una maggiore frequenza di spese rare. Estremamente difficoltoso risulta il confronto con la vecchia indagine, in quanto alle differenze in termini di modello si aggiungono quelle dovute alla revisione del piano di check, che ha comportato una modifica dei criteri e della metodologia di correzione dei dati errati/ anomali. Mentre per la nuova indagine si è potuto procedere alla stima dell'impatto del processo di correzione, nella vecchia tale studio si è rilevato estremamente complesso essendo le modifiche apportate durante la fase di revisione manuale estremamente influenti ed assolutamente non ricostruibili.

1.2.4 Appendice - Voci di spesa per il questionario della vecchia e nuova indagine

VECCHIA INDAGINE	Rif.	NUOVA INDAGINE	Rif.
	Temp.		Temp.
Pane	10 gg	Pane, grissini, crackers	7 gg.
Grissini, fette biscottate, crackers	10 gg	Biscotti e fette biscottate	7 gg.
Biscotti e pasticceria (esclusi gelati, cioccolato, caramelle e simili)	10 gg	Pasticceria e dolci	7 gg.
Pasta comune, speciale	10 gg	Pasta comune, all'uovo, preparata e ripiena	7 gg.
Pasta preparata	10 gg		
Riso e creme di riso	10 gg	Riso	7 gg.
Farina	10 gg	Farina	7 gg.
		Altri cereali e prodotti preparati (mais, pizze, torte salate, pop-corn alimenti dietetici e per bambini)	7 gg.
Vitello e vitellone	10 gg	Carne di vitello e di vitellone	7 gg.
Manzo, altri bovini adulti	10 gg	Carne di manzo	7 gg.
Carne di maiale	10 gg	Carne di maiale	7 gg.
Carne equina	10 gg	Carne di cavallo	7 gg.
Carne ovina e caprina (di pecora, agnello, castrato)	10 gg	Carne di pecora e di agnello	7 gg.
Pollame	10 gg	Pollame	7 gg.
Conigli, tacchini, quaglie e fagiani (compresi quelli congelati)	10 gg	Conigli, tacchini, selvaggina	7 gg.
Salumi	10 gg	Salsicce, pancetta, prosciutto, würstel, ecc.	7 gg.
Carne in scatola (compresi gli omogeneizzati, escluso il prosciutto in scatola)	10 gg	Carni preparate e in scatola (alimenti dietetici)	7 gg.
	10 gg	Altre carni (da specificare)	7 gg.
Pesce fresco, congelato e simili (compresi molluschi e crostacei)	10 gg	Pesce fresco o congelato	7 gg.
		Crostacei ed altri frutti di mare (freschi o surgelati)	7 gg.
Pesce conservato (secco, in scatola, omogeneizzato, ecc.)	10 gg	Pesce conservato, secco, affumicato	7 gg.
		Altri prodotti a base di pesce (pasta di acciughe, creme a base di pesce, caviale, ecc.)	7 gg.
Latte (anche condensato ed in polvere)	10 gg	Latte intero o scremato	7 gg.
		Latte in polvere o concentrato	7 gg.
Yogurt	10 gg	Yogurt	7 gg.
Panna	10 gg		
Formaggi freschi	10 gg	Formaggi freschi e stagionati	7 gg.
Formaggi stagionati	10 gg		
Uova	10 gg	Uova	7 gg.
		Altri prodotti derivati dal latte: (bevande e dessert a base di latte, creme fresche, panna, ecc.)	7 gg.
Olio d'oliva	10 gg	Olio d'oliva	7 gg.
VECCHIA INDAGINE	Rif.	NUOVA INDAGINE	Rif.
	Temp.		Temp.
Olio di semi (di arachidi, di girasole, ecc.)	10 gg	Olio di semi (di girasole, arachidi, soia, ecc.) di sansa	7 gg.

Burro (fresco, salato o in scatola)	10 gg	Burro	7 gg.
Margarina	10 gg	Margarina ed altri grassi vegetali	7 gg.
Lardo, strutto e simili	10 gg	Lardo, strutto e altri grassi animali	7 gg.
Agrumi (arance, limoni, mandarini, cedri)	10 gg	Arance, limoni, pompelmi, kiwi, ecc.	7 gg.
Mele, pere ed altra frutta fresca (anche esotica)	10 gg	Banane	7 gg.
		Mele	7 gg.
		Pere	7 gg.
		Frutta con nocciolo (pesche, ciliege, avocado)	7 gg.
		Uva e fragole	7 gg.
		Altra frutta fresca (meloni, ananas, ecc.)	7 gg.
Frutta secca	10 gg	Frutta secca (noci, castagne, prugne)	7 gg.
Pomodori freschi	10 gg	Pomodori freschi	7 gg.
Legumi freschi o surgelati	10 gg	Legumi freschi o surgelati (fagioli, fave, piselli)	7 gg.
Ortaggi freschi o surgelati	10 gg	Ortaggi freschi o surgelati (insalata, cavoli, bieta, ecc.)	7 gg.
Patate (anche fritte o surgelate)	10 gg	Patate o patatine fritte (anche surgelate)	7 gg.
Zucchero	10 gg	Zucchero	7 gg.
Marmellate, gelatine, mieie, cacao, confetti e simili (cioccolato, budini, ecc.)	10 gg	Confetture, marmellate, miele, cioccolato, ecc.	7 gg.
Gelati	10 gg	Gelati	7 gg.
Caffè, the, surrogati (orzo, camomilla, tisane, ecc.)	10 gg	Caffè	7 gg.
		Tè, orzo, cacao, camomilla, ecc.	7 gg.
Frutta conservata (surgelata, congelata, in scatola, omogeneizzata, sotto spirito, ecc.)	10 gg	Frutta conservata o surgelata	7 gg.
Pomodori in scatola e conserve (esclusi i succhi di pomodoro)	10 gg	Pomodori in scatola e conserve	7 gg.
Legumi secchi o conservati	10 gg	Legumi secchi o conservati (fagioli, ceci, lenticchie, ecc.)	7 gg.
Ortaggi secchi o conservati	10 gg	Ortaggi secchi o conservati (carciofini, funghi)	7 gg.
		Altri prodotti a base di legumi e ortaggi	7 gg.
Altri generi (sale, aceto, dadi da brodo, spezie)	10 gg	Sale, spezie, salse e condimenti	7 gg.
		Zuppe, preparati, lievito	7 gg.
Acqua minerale comprese acque toniche e brillanti)	10 gg	Acqua minerale	7 gg.
Succhi di frutta, di pomodoro	10 gg	Succhi di frutta, di pomodoro, ecc.	7 gg.
Altre bevande non alcoliche (limonate, chinotti, ecc.)	10 gg	Chinotti, cole e altre bevande non alcoliche	7 gg.
Liquori (compresi acquavite, alcool per liquori)	10 gg	Liquori, champagne, aperitivi alcolici, ecc.	7 gg.
Vino (compresi vini spumante, champagne, aperitivi a base di vino, marsala, porto, ecc.)	10 gg		7 gg.
		Vino	7 gg.
Birra	10 gg	Birra	7 gg.
Sigarette, sigari, tabacco	10 gg	Sigarette, sigari, tabacco (esclusi accendini e fiammiferi)	7 gg.
VECCHIA INDAGINE	Rif.	NUOVA INDAGINE	Rif.

	Temp.		Temp.
Abiti, cappotti, soprabiti, impermeabili (abiti interi, completi, ecc., compresi capi per neonati e bambini)	mens.		
		Abbigliamento per uomo: giacche, pantaloni, cappotti, impermeabili, camicie, maglioni	mens.
		Abbigliamento per donna: giacche, pantaloni, cappotti, tailleur, gonne, maglioni	mens.
		Abbigliamento per ragazzi, bambini e neonati	mens.
Maglieria e tessuti a maglia per abbigliamento a maglia (golf, gilet, vestiti, ecc. a maglia, compresi capi per neonati e bambini)	mens.		
Articoli di abbigliamento per lo sport e il tempo libero (Giubbotti, jeans, pantaloni, tute, abbigliamento da sci, tennis, ecc.)	mens.		
Biancheria personale, camicie e corsetteria (camicie, camicette, sottovesti, mutandine, ecc.; escluse quelle in maglia, compresi capi per neonati e bambini)	mens.		
		Biancheria intima (uomo, donna, ragazzi, bambini)	mens.
Pellicce ed articoli di pellicceria (pellicce intere, stole, boleri, colli di pelliccia ed altri articoli di pellicceria)	mens.	Pellicce e articoli di pellicceria, montoni	mens.
Tessuti per confezione su misura (tessuti a metraggio per indumenti da uomo, donna e ragazzo, esclusi i tessuti a maglia)	mens.	Tessuti per confezione su misura	mens.
Confezione di articoli di abbigliamento (confezione di capi di vestiario ed articoli di maglieria [da parte di sarti, maglieriste, modiste, ecc.] con materia prima fornita dalla famiglia)	mens.	Spese per confezionare e riadattare abiti nuovi e usati	
Maglieria intima (maglie, canottiere, slip, ecc. a maglia, compresi capi per neonati e bambini)	mens.		
Calze, fazzoletti merceria, cappelli	10 gg.		
		Bottoni, fili, passamaneria e gomitolini di lana	7 gg.
		Altri articoli ed accessori di abbigliamento (scarpe, cappelli, cravatte, guanti, cinture, ecc.)	mens.
Manutenzione e riparazione (di abiti, pellicce ed altri generi di abbigliamento)	mens.	Riparazioni di capi di abbigliamento	7 gg.
Calzature (scarpe, scarponi, pantofole e simili; comprese le calzature per uso sportivo)	mens.	Calzature da uomo	mens.
		Calzature da donna	mens.
		Calzature da bambino o neonato	mens.
Calzolaio		Riparazione di calzature	7 gg.
Affitto abitazione principale	mens.	Affitto abitazione principale	mens.
Fitto figurativo abitazione principale	mens.	Fitto figurativo abitazione principale	mens.
Affitto abitazione secondaria	mens.	Affitto abitazione secondaria	mens.

Fitto figurativo abitazione secondaria	mens.	Fitto figurativo abitazione secondaria	mens.
		Uso gratuito	mens.
VECCHIA INDAGINE	Rif.	NUOVA INDAGINE	Rif.
	Temp.		Temp.
Muratori e impiantisti abitazione principale	10 gg.	Tinteggiatura e carta da parati	3 m.
		Riparazione impianto idrico, sanitario, riscaldamento	3 m.
		Riparazione impianto elettrico	3 m.
		Riparazione e manutenzione di porte, infissi e pavimenti	3 m.
		Altri lavori	3 m.
		Rifacimenti esterni (facciate, tetti, scale, ecc.)	3 m.
		Rifacimenti interni (pavimenti, pareti, soffitti, ecc.)	3 m.
		Rifacimento impianto idrico e/o sanitario	3 m.
		Sostituzione di porte, finestre, infissi	3 m.
		Rifacimento o installazione di impianto di riscaldamento, rifacimento	3 m.
		Altri lavori	3 m.
Muratori e impiantisti abitazione secondaria	10 gg.	Tinteggiatura e carta da parati	3 m.
		Riparazione impianto idrico, sanitario, riscaldamento	3 m.
		Riparazione impianto elettrico	3 m.
		Riparazione e manutenzione di porte, infissi e pavimenti	3 m.
		Altri lavori	3 m.
		Rifacimenti esterni (facciate, tetti, scale, ecc.)	3 m.
		Rifacimenti interni (pavimenti, pareti, soffitti, ecc.)	3 m.
		Rifacimento impianto idrico e/o sanitario	3 m.
		Sostituzione di porte, finestre, infissi	3 m.
		Rifacimento o installazione di impianto di riscaldamento, rifacimento	3 m.
		Altri lavori	3 m.
Spese condominiali	mens.	Condominio	3 m.
Spese condominiali abitazione secondaria	mens.	Condominio abitazione secondaria	3 m.
Acqua (abitazione principale)	mens.	Acqua (abitazione principale)	3 m.
Acqua (abitazioni secondarie)	mens.	Acqua (abitazioni secondarie)	3 m.
Energia elettrica	u.p.	Energia elettrica	u.p.
Gas da rete	mens.	Gas da rete	u.p.
Gas in bombole o in contenitori esterni	mens.	Gas in bombole o contenitori esterni	3 m.
Kerosene, gasolio ed altri combustibili liquidi	mens.	Kerosene, gasolio ed altri combustibili liquidi	3 m.
Carbone e legna	mens.	Carbone e legna	3 m.
Riscaldamento da impianto centralizzato	mens.	Riscaldamento centralizzato	3 m.
Energia elettrica	mens.	Energia elettrica	u.p.*
Gas da rete	mens.	Gas da rete	u.p.*
Gas in bombole o in contenitori esterni	mens.	Gas in bombole o contenitori esterni	3 m.

Kerosene, gasolio ed altri combustibili liquidi	mens.	Kerosene, gasolio ed altri combustibili liquidi	3 m.
Carbone e legna	mens.	Carbone e legna	3 m.
Riscaldamento da impianto centralizzato	mens.	Riscaldamento centralizzato	3 m.
VECCHIA INDAGINE	Rif.	NUOVA INDAGINE	Rif.
	Temp.		Temp.
Mobili in legno	mens.	Arredamenti completi: cucine, soggiorni, ecc.	mens.
Mobili in metallo	mens.	Mobili singoli: tavoli, sedie, ecc.	mens.
Mobili in plastica	mens.	Mobili ed accessori per il bagno (esc. tappetini)	mens.
	mens.	Mobili per esterno	mens.
<i>Tappeti e moquette (guide, stuoini, scendiletto, ecc.)</i>	mens.	<i>Tappeti, moquette</i>	mens.
Falegname	10 gg.	Riparazione di mobili e articoli di arredamento	7 gg.
Coperte, lenzuola e tessuti da arredamento (coperte in lana e miste, cretonne, velluto e simili; tende; federe)	mens.	Coperte, lenzuola, asciugamani, tovaglie, tende, materassi, tessuti di arredamento	mens.
Materassi (di lana, crine, a molle, in gommapiuma e simili)	mens.		
Asciugamani, strofinacci	10 gg.		
Manutenzione e riparazione di biancheria per la casa	mens.	Riparazione biancheria per la casa	7 gg.
Cucine, forni a microonde, stufe e scaldabagni	3 m.-u.a.	Cucine elettriche (miste, forni elettrici, forni a microonde)	3 m.
		Cucine non elettriche	3 m.
		Stufe, scaldabagni, cappe aspiranti	3 m.
Frigoriferi, congelatori, surgelatori e combinati	3 m.-u.a.	Frigorifero, congelatori, o combinati	3 m.
Lavastoviglie	3 m.-u.a.	Lavastoviglie	3 m.
Lavatrici	3 m.-u.a.	Lavatrici	3 m.
Lucidatrici, aspirapolvere e battitappeto	3 m.-u.a.	Apparecchi per la pulizia (aspirapolvere, battitappeti)	3 m.
Condizionatori d'aria	3 m.-u.a.	Condizionatori d'aria	3 m.
<i>Macchine per cucire e per maglieria</i>	mens.	<i>Macchine per cucire, per maglieria</i>	3 m.
Piccoli apparecchi elettrici (frullatori, macinacaffè, tostapane, ferri da stiro, ventilatori, ecc.)	mens.		
	mens.	Frullatori, macinacaffè, tostapane, ferri da stiro, ventilatori, ecc.	mens.
Riparazione di elettrodomestici (escluso TV) ed altri apparecchi per la casa	10 gg.	Riparazione di grossi elettrodomestici [escluso TV]	7 gg.
		Riparazione di piccoli elettrodomestici (frullatori, ecc.)	7 gg.
		Grossi utensili per la casa e per il giardino (seghe e trapani elettrici, tagliaerba)	mens.
Pentole, padelle, posate (in metallo)	mens.	Posate (servizi completi o pezzi singoli, anche in argento)	mens.
Piatti, tazze, ecc. (in ceramica)	mens.	Bicchieri, bottiglie, piatti, tazze ecc.	mens.
Bicchieri, bottiglie, ecc. (in vetro, cristallo, vetro temperato)	mens.		

	mens.	Utensili da cucina e per la casa (pentole, padelle, ecc.)	mens.
<i>Quadri, soprammobili, specchi</i>	mens.	<i>Quadri, soprammobili, specchi</i>	mens.
Lampadari, lampade (lampadari e lumi da tavolo in vetro, metallo, ceramica e legno; pendoli: materiale elettrico [fili, lampadine, spine, trasformatori, ecc.])	mens.		
	mens.	Lampade, lampadari, abat-jour	mens.
VECCHIA INDAGINE	Rif.	NUOVA INDAGINE	Rif.
	Temp.		Temp.
Altri beni per la casa (frullini, ferri da stiro, bilance, ecc. non elettrici),	mens.	Altri oggetti per la casa: assi da stiro, bilance per alimenti, ecc.	mens.
Bicchieri, piatti, salviette, posate (in carta o plastica)	mens.	Tovaglioli, piatti e bicchieri di carta e di plastica	7 gg.
		Carta per cucina, contenitori di alluminio, ecc.	7 gg.
		Piccoli utensili e accessori (forbici, batterie)	7 gg.
		Aghi, uncinetti, spille di sicurezza	7 gg.
		Scope, guanti di gomma, fiammiferi	7 gg.
Sapone da bucato, detersivi, smacchiatori	10 gg.	Detersivi, cere per mobili, insetticidi, ecc.	7 gg.
Insetticidi, disinfettanti, vernici	10 gg.		
<i>Scope, guanti, fiammiferi, articoli da giardinaggio</i>	10 gg.	<i>Altro</i>	7 gg.
Lavanderia e tintoria	10 gg.	Lavanderia e tintoria	7 gg.
Servizi domestici (retribuzione del personale di servizio, compresi gli oneri previdenziali a carico della famiglia)	mens.	Domestici, giardinieri, baby sitter	mens.
		Parenti, vicini, persone alla pari	mens.
		Assistenza per disabili e per anziani non autosufficienti	mens.
Medicinali	mens.	Medicinali	7 gg.
Occhiali da vista, lenti a contatto, siringhe, apparecchi ortopedici e altro materiale sanitario (apparecchi per l'udito, sedie e carrozzine per invalidi, inalatori, lampade al quarzo, protesi, siringhe, aghi per siringhe, ecc.)	mens.	Occhiali da vista e lenti a contatto	mens.
		Protesi	mens.
		Sedie e carrozzine per invalidi	mens.
		Termometri, siringhe, calze elastiche	7 gg.
		Noleggio di attrezzature sanitarie	7 gg.
Onorari per medici, infermieri e simili	mens.	Medico generico o specialista (a domicilio, in ambulatorio, in clinica od ospedale)	mens.
		Dentista	mens.
		Infermieri, fisioterapisti, ginnastica correttiva	mens.
		Cure termali	mens.
Radiografie, analisi, esami ed altri accertamenti diagnostici	mens.	Analisi cliniche (del sangue, delle urine)	mens.

		Esami radiologici, ecografici, mens. elettrocardiogramma	
<i>Ricoveri in ospedali, cliniche e simili</i>	<i>mens.</i>	<i>Ricoveri in cliniche ed ospedali</i>	<i>mens.</i>
<i>Ricoveri in casa di riposo, istituti di assistenza e simili</i>	<i>mens.</i>	<i>Ricoveri in case di riposo, istituti di assistenza e simili</i>	<i>mens.</i>
Altre spese (fanghi, massaggi, saune terapeutiche, ecc.)	mens.		
		Pronto soccorso	mens.
Acquisto di automobili nuove	3 m-u.a.	Automobili	3 m.
Acquisto di automobili usate	3 m-u.a.		
Acquisto di scooters, moto e motorini	3 m-u.a.	Motocicletta	3 m.
		Ciclomotori, scooter, motorini	3 m.
VECCHIA INDAGINE	Rif.	NUOVA INDAGINE	Rif.
	Temp.		Temp.
Acquisto di roulotte rimorchi, camper	3 m-u.a.	Camper, roulotte, rimorchi	3 m.
Acquisto di bicicletta	3 m-u.a.	Bicicletta	3 m.
Pneumatici e camere d'aria	mens.	Pezzi di ricambi ed accessori (pneumatici, candele, batterie)	mens.
Meccanico e carrozziere (compresi ricambi ed accessori)	10 gg.	Manutenzione e riparazione (meccanico, carrozziere, elettrauto)	mens.
<i>Oli ed altri lubrificanti per veicoli</i>	<i>mens.</i>	<i>Olio ed altri lubrificanti</i>	<i>mens.</i>
<i>Benzina per auto e moto</i>	<i>10 gg.</i>	<i>Benzina per auto e moto</i>	<i>7 gg.</i>
<i>Gasolio e altri combustibili per auto e moto</i>	<i>10 gg.</i>	<i>Gasolio e altri combustibili per auto e moto</i>	<i>7 gg.</i>
Assicurazione per auto ed altri veicoli	mens.	Assicurazione per auto ed altri veicoli	1 anno
Garage per auto e moto	mens.	Custodia in garage privato o affitto per box o posto macchina	mens.
		Assicurazione natanti e rimessaggio natanti, caravan e roulotte	3 m.
Altre spese per auto e moto (lavaggio e ingrassaggio; pedaggi autostradali; soccorsi stradali [traino]; lezioni di guida, ecc.)	mens.	Lezioni di guida	3 m.
		Pedaggi autostradali, parcheggi, lavaggio e ingrassaggio	7 gg.
Biglietti, abbonamenti per autobus, tram, corriere, parcheggi	10 gg.	Biglietti e abbonamenti per bus, metro e tram	7 gg.
		Taxi	7 gg.
		Biglietti e abbonamenti per pullman o corriere per collegamenti extraurbani	mens.
Biglietti e abbonamenti ferroviari	mens.	Biglietti e abbonamenti ferroviari	mens.
Biglietti per aerei	mens.	Biglietti per aerei	mens.
Biglietti e abbonamenti per vaporetto, traghetti ed altri mezzi marittimi e lacuali	mens.	Biglietti e abbonamenti per vaporetto, traghetti	mens.
Altre spese di trasporto (traslochi, trasporto e deposito bagagli, noleggio di auto senza autista, di roulotte, di scooters e simili)	mens.	Traslochi, trasporto oggetti, noleggio mezzi senza autista	3 m.
Acquisto di telefoni, segreterie telefoniche	mens.	Apparecchio telefonico	3 m.
		Telefono cellulare	3 m.
Acquisto di telefoni cellulari	3 m-u.a.	Segreteria telefonica	3 m.
		Fax	3 m.
		Riparazione di apparecchi telefonici	7 gg.

Telefono (importo mensile della bolletta)	3 m-u.a.	Telefono (bolletta) per abitazione principale	u.p.
		Telefono (bolletta) per abitazione secondaria	u.p.
<i>Gettoni telefonici</i>	<i>10 gg.</i>	<i>Schede e gettoni telefonici</i>	<i>7 gg.</i>
<i>Francobolli e altre spese postali</i>	<i>10 gg.</i>	<i>Francobolli ed altre spese postali</i>	<i>7 gg.</i>
Televisore in bianco e nero	3 m-u.a.	Televisore	3 m.
Televisore a colori	3 m-u.a.		
Videoregistratori, telecamere	3 m-u.a.	Videoregistratori	3 m.
Radio e radiogrammofoni	3 m-u.a.	Radio, microfoni, cuffie	
Giradischi, registratori, lettori compact disc	3 m-u.a.	Impianti hi-fi	3 m.
Personal computer e periferiche	3 m-u.a.	Personal computer, stampanti	3 m.
Barche, canotti e spese connesse (acquisto di imbarcazioni di qualsiasi tipo, materiale e potenza, comprese le spese per accessori, gestione e simili)	3 m-u.a.		
VECCHIA INDAGINE	Rif.	NUOVA INDAGINE	Rif.
	Temp.		Temp.
		Barca, gommone	3 m.
		Canotto, canoa, windsurf	3 m.
Strumenti musicali (pianoforti, chitarre, batterie, fisarmoniche, amplificatori per strumenti, ecc. [compreso il loro affitto])	mens.	Strumenti musicali	3 m.
Macchine fotografiche, cineprese, proiettori, ottica (comprese moviole, ed altri apparecchi fotocinematografici; occhiali da sole, esclusi quelli da vista)	mens.	Macchine fotografiche, videocamere, proiettori, ingranditori, microscopi	mens.
Macchine da scrivere	3 m-u.a.		
		Calcolatrici, macchine da scrivere, accessori per personal computer	mens.
Articoli sportivi e da campeggio (tavoli da biliardo e da ping pong, racchette da tennis, articoli da caccia, da pesca, sci, tende, ombrelloni, sacchi a pelo, ecc.; compreso l'affitto di articoli sportivi e da campeggio)	mens.	Articoli sportivi, da campeggio e svaghi all'aria aperta	mens.
Animali domestici (cani, gatti uccellini, tartarughe, ecc. per esclusivo uso della famiglia)	mens.	Animali domestici (acquisto)	mens.
Altre spese (piettrine per accendini, carte da gioco, scacchi e calcolatrici)	mens.	Altre spese (carte da gioco, scacchi, ecc.)	mens.
<i>Pellicole e loro sviluppo</i>	<i>10 gg.</i>	<i>Pellicole fotografiche e loro sviluppo e stampa</i>	<i>7 gg.</i>
<i>Dischi, fonocassette, nastri magnetici, videocassette, floppy</i>	<i>10 gg.</i>	<i>Dischi, cassette, videocassette (compreso noleggio)</i>	<i>7 gg.</i>
<i>Piante e fiore</i>	<i>10 gg.</i>	<i>Piante e fiori</i>	<i>7 gg.</i>
<i>Mangimi ed altre spese per animali domestici</i>	<i>10 gg.</i>	<i>Mangimi e altre spese per animali domestici</i>	<i>7 gg.</i>
		Altro	7 gg.
<i>Giocattoli</i>	<i>10 gg.</i>	<i>Giocattoli e giochi</i>	<i>7 gg.</i>

Manutenzione e riparazione di macchine fotografiche, da scrivere, giradischi, radio, TV, accendini, ecc. (compresi i pezzi di ricambio e gli accessori)	mens.	Riparazione di radio, TV, personal computer	7 gg.
Bricolage, hobby (materiale e utensili per lavorare il legno, il metallo, il vetro, la ceramica, ecc.)	mens.	Bricolage, hobby	mens.
Sport: frequenza a piscine, palestre	10 gg.	Sport: frequenza a piscine, palestre, campi da tennis	mens.
Fiere, musei, manifestazioni varie	10 gg.	Abbonamenti a manifestazioni sportive	mens.
		Divertimenti: corsi di pittura, di ballo	mens.
		Biglietti per musei, manifestazioni sportive e varie	7 gg.
Cinema, teatri, discoteche, spettacoli sportivi	10 gg.	Biglietti per cinema, teatro, concerti	7 gg.
		Abbonamenti a concerti, teatri, circoli culturali	mens.
Abbonamento radio e televisione (compresi gli abbonamenti per apparecchi installati in abitazioni di villeggiature e su automezzi.)	mens.	Abbonamento radio o TV, Pay-TV, Internet	3 m.
		Altro	7 gg.
Totocalcio, lotto ed altri concorsi	10 gg.	Totocalcio, lotto e altri concorsi	7 gg.
Giornali, riviste e fumetti	10 gg.	Abbonamenti a giornale, riviste, dispense, enciclopedie	3 m.
VECCHIA INDAGINE	Rif.	NUOVA INDAGINE	Rif.
	Temp.		Temp.
		Giornali, riviste, fumetti	7 gg.
Libri	10 gg.	Libri non scolastici	7 gg.
Tasse scolastiche (per iscrizioni a scuole di ogni ordine e grado [elementari, medie, professionali, università, ecc.])	mens.	Tasse scolastiche	mens.
Altre spese per l'istruzione (rette, ivi comprese eventuali spese di riscaldamento, trasporto, ecc.; ripetizioni, lezioni private; escluse tasse scolastiche)	mens.		
		Rette e simili (corsi di lingue, personal computer)	mens.
		Trasporto scolastico con scuolabus	mens.
		Lezioni private e ripetizioni	mens.
		Spese per alloggio e mantenimento	mens.
		Libri scolastici	mens.
Barbiere e parrucchiere	10 gg.	Barbiere, parrucchiere istituto di bellezza	7 gg.
Saponette, dentifrici, articoli di profumeria, carta igienica	10 gg.	Saponette, dentifrici, articoli da profumeria (shampoo, deodoranti, ecc.)	7 gg.
		Fazzoletti di carta, pannolini per neonati, carta igienica, ovatta, ecc.	7 gg.
Argenteria, gioielli, orologi e simili (orologi da tasca e da polso, argenteria, e gioielleria di qualsiasi tipo, compresi la bigiotteria e gli orologi da parete)	mens.	Argenteria, gioielleria, orologi	
		Bigiotteria	mens.
		Apparecchi elettrici per la cura della persona	mens.

Borse e articoli in pelle (borse, borsette, valigie, portafogli, portachiavi, ed altri articoli in pelle, esclusi i guanti)	mens.	Borse, valigie ed altri articoli da viaggio	mens.
Articoli personali (pipe, bastoni, ombrelli, articoli per bambini, quali carrozzine, boxes, borse e valigie non in pelle, ecc.)	mens.	Altri articoli personali (articoli per fumatori, ombrelli, occhiali da sole)	mens.
		Articoli per l'infanzia (carrozzine ed articoli simili)	mens.
		Altro	mens.
Quaderni e cancelleria		Quaderni, cancelleria, fotocopie, fax	7 gg.
Gite e viaggi turistici a prezzo tutto compreso	mens.	All'Estero Gite o viaggi organizzati tutto compreso	
Alberghi pensione e simili (vitto e alloggio [escluse le spese per viaggi di affari e di lavoro]; spese per camping)	mens.	All'Estero Pensione completa (in albergo, residence, villaggi turistici, camping)	mens.
		All'Estero pernottamento (in albergo, residence, villaggi turistici, camping, case private)	mens.
		In Italia Gite o viaggi organizzati tutto compreso	mens.
		In Italia Pensione completa (in albergo, residence, villaggi turistici, camping)	mens.
		In Italia Pernottamento (in albergo, residence, villaggi turistici, camping, case private)	mens.
Bar, pasticceria, chioschi, ecc.	10 gg.	Bar, pasticceria, chioschi, ecc.	7 gg.
Ristorante, trattoria, tavole calde, ecc.	10 gg.	Ristorante, trattoria, tavole calde, ecc.	7 gg.
VECCHIA INDAGINE	Rif.	NUOVA INDAGINE	Rif.
	Temp.		Temp.
Mense aziendali, scolastiche, ecc.	10 gg.	Mense aziendali, scolastiche, ecc.	7 gg.
Assicurazione sulla persona (assicurazioni sulla vita, invalidità, malattie e infortuni)	mens.	Assicurazione malattie (escluse trattenute di legge)	
		Assicurazione vita, rendite vitalizie	3 m.
Altre assicurazioni (assicurazione contro il furto, incendio e tutte le altre assicurazioni private [escluse le assicurazioni sulla vita e per veicoli])	mens.	Assicurazione abitazione principale (furto, incendio, danni, ecc.)	3 m.
		Assicurazione abitazione secondaria (furto, incendio, danni, ecc.)	3 m.
Onorari per liberi professionisti (esclusi i medici) (spese per avvocati, consulenti fiscali, patrimoniali, ecc.)	mens.	Onorari per commercialisti, consulenti fiscali	3 m.
		Onorari per avvocati, notai, architetti	3 m.
Culto (spese per cerimonie anche civili, per il culto, ecc.)	mens.	Cerimonie religiose	3 m.
		Altro	3 m.
		Spese per altri servizi (avvisi sui giornali, certificati, spese funebri, ecc.)	3 m.

Legenda

mens. = mensile

3 m. = trimestrale

u.p. = ultimo pagamento

3 m.-u.a. = si chiede l'acquisto nell'ultimo anno e l'ammontare solo se acquistato entro gli ultimi 3 mesi.

1.3 L'acquisizione dei dati

Nel seguito verranno descritte le fasi concernenti l'acquisizione, il controllo e la correzione dei dati. La filosofia seguita nel disegno e nell'implementazione delle relative è stata quella di sottoporre i dati al massimo numero di controlli possibili, il più possibile coincidenti in entrambe le fasi di acquisizione e revisione, onde garantire la minimizzazione della componente non campionaria degli errori.

1.3.1 Caratteristiche dell'acquisizione in modalità CADI

Come sempre avviene nelle rilevazioni di indagini statistiche, anche i dati rilevati dall'indagine sui Consumi delle famiglie italiane possono contenere errori sia nelle variabili qualitative sia nelle variabili quantitative di spesa.

Gli errori possono verificarsi in sede di compilazione ed in fase di registrazione dei modelli. Per minimizzare la fonte di errori dovuta alla registrazione è stato deciso di ricorrere all'acquisizione controllata dei dati in modalità CADI (*Computer Assisted Data Input*). Tale tecnica consiste nell'utilizzare in fase di registrazione un apposito software che segnali la presenza di errori nei dati rilevati.

Nella pratica quindi la rilevazione è stata effettuata tradizionalmente, tramite questionario cartaceo, mentre la registrazione è avvenuta utilizzando un software appositamente sviluppato che ha effettuato un controllo preliminare sui dati rilevati.

Le successive fasi di controllo e correzione sono state implementate in modo da attuare una *corrispondenza quasi perfetta tra i controlli della procedura di data editing e quelli della procedura di acquisizione*.

L'acquisizione controllata in modalità CADI consente di inserire controlli sui valori delle variabili sia qualitative sia quantitative.

In fase di acquisizione dati, i possibili errori vengono classificati come:

- errori di *range* ossia inserimento di un valore esterno al campo di definizione della variabile;
- errori di *rotta* riguardanti il cammino da seguire nella somministrazione delle domande di un questionario;
- errori di *coerenza* tra quanto viene risposto ad un quesito rispetto a quanto è stato dichiarato precedentemente in un altro quesito.

Un software di controllo consente di:

- inibire la registrazione di un dato errato, oppure
- inserire il dato errato previa segnalazione di errore.

In termini tecnici la prima gestione degli errori è definita di tipo *hard* mentre la seconda è definita di tipo *soft*.

Qualora il software di controllo venisse usato direttamente in fase di rilevazione dei dati, l'impostazione, relativamente agli errori di range ed a quelli di rotta, dovrebbe essere di tipo *hard*, in quanto se l'errore fosse dipeso dal rilevatore, questo potrebbe correggerlo immediatamente oppure, se fosse dipeso dall'intervistato, gli sarebbe data la possibilità di fornire di nuovo la risposta al quesito. Gli errori di coerenza invece dovrebbero essere trattati sempre in modalità *hard*, qualora implicino regole che non possono essere assolutamente violate, oppure in modalità *soft* qualora contrastino regole non così rigide.

Per chiarire con degli esempi, si pensi alla regola: «due coniugi non possono avere lo stesso sesso». L'eventuale errore dovrebbe essere trattato secondo la modalità *hard*; al contrario un valore di spesa che si discosti non significativamente da un certo valore stimato dovrebbe essere trattato in modalità *soft*.

Applicando il CADI, invece, il trattamento degli errori non può essere così rigido, in quanto il dato errato è stato ormai acquisito in fase di rilevazione e non si può quindi inibirne sempre in modo assoluto la registrazione. Possono quindi essere seguite due strade:

- gestire le regole sempre in modalità *soft*, fornendo quindi all'operatore una messaggistica di errore; questi dovrà verificare se l'inserimento di un dato errato sia imputabile ad un errore di digitazione oppure se si tratti di un vero e proprio errore di compilazione del modello. Nel primo caso l'errore deve essere rimosso, nel secondo, invece, l'iter è diverso a seconda che chi inserisce i dati sia soltanto un operatore di data-entry, che non ha quindi competenza per correggere l'errore ed è tenuto a digitare fedelmente quanto riportato sul modello cartaceo, oppure che si tratti di una persona incaricata di effettuare revisione, che può invece intervenire sui dati rilevati;
- stabilire a priori un «modus operandi» sulle risposte date ad alcuni quesiti, che si traduce di fatto nel trattamento delle relative regole di rotta o di coerenza in modalità *hard*. Si può, per esempio, decidere di inserire le rotte nel questionario elettronico, gestendole così in modalità *hard*, dando però indicazioni specifiche a chi effettua la registrazione.

Per chiarire con un esempio quanto detto, si pensi ad un quesito del tipo:

Quali delle seguenti spese ha sostenuto la sua famiglia nel corso degli ultimi tre mesi?

Tipo di spesa	SI = 1	No = 2	Lire
Assicurazione malattie	<input type="checkbox"/> 1	<input type="checkbox"/> 2	
Assicurazione abitazione principale	<input type="checkbox"/> 1	<input type="checkbox"/> 2	

Qualora il rispondente abbia risposto «NO» ad una voce di spesa, il software di controllo non consentirà di rispondere al corrispondente quesito sulle Lire. Se però sul modello cartaceo risulta che è stato inserito un valore in corrispondenza delle Lire, si dovrà dare a chi effettua il data-entry l'indicazione di privilegiare quest'ultima risposta e di forzare quindi «SI» nella precedente risposta.

In questo caso infatti si considera evidente che «l'errore» è stato commesso nella risposta SI/NO, che viene quindi già corretta in questa fase. Tale strada viene abitualmente prescelta quando chi effettua la registrazione non è l'operatore di data-entry, ma il revisore esperto dell'indagine.

Il software di acquisizione dati utilizzato per l'applicazione del CADI sull'indagine sui Consumi delle famiglie è il BLAISE, sviluppato dall'Istituto nazionale di statistica olandese (Statistics Netherlands) e già ampiamente diffuso in ambito comunitario. Tale sistema opera in ambiente DOS su personal computer. Lo sviluppo dell'applicazione per la citata indagine è stato curato dal Servizio Studi Metodologici, e successivamente modificato dal Servizio CEF. La sintassi di Blaise consente, una volta definita la struttura del questionario e i quesiti che ne fanno parte, di associarvi tutti i già citati tipi di regole.

Nella pratica un programma Blaise è composto da due sezioni principali: nella prima, identificata dalla parola riservata FIELDS, sono elencate le variabili e le loro caratteristiche; nella seconda, identificata dalla parola riservata RULES, sono riportate le regole che rilevano gli errori presenti nei valori delle variabili e attivano la messaggistica che compare sullo schermo. Nell'ambito di quest'ultima sezione le regole sono definite di tipo CHECK (gestione *hard* dell'errore) o di tipo SIGNAL (gestione *soft* dell'errore).

Una data regola viene valutata dal sistema non appena terminato l'inserimento del valore dell'ultima variabile coinvolta. Se la regola non è rispettata, il sistema propone all'operatore un predefinito messaggio di errore unitamente all'elenco delle variabili coinvolte e dei valori registrati in esse. Se la regola è di tipo CHECK l'operatore è obbligato a rimuovere l'errore per poter continuare l'inserimento dei dati. Se la regola è di tipo SIGNAL l'operatore può scegliere tra rimuovere l'errore e ignorarlo continuando l'inserimento dei dati.

Nell'ambito dell'applicazione del CADI all'indagine sui Consumi delle famiglie sono state prese in considerazione tutte le già citate tipologie di errore; tuttavia la filosofia relativa alla gestione delle regole seguita per l'indagine sui consumi delle famiglie è stata diversa per i dati del 1996, rilevati nell'ambito dell'indagine sperimentale, e quelli del 1997 dell'indagine a regime.

Nel 1996, infatti, fu concordato di consentire all'operatore la sola rimozione degli errori di digitazione e di procedere alla correzione degli errori effettivamente presenti sul questionario (errori di compilazione) in una fase successiva mediante procedure automatiche.

Dal 1997, invece, il trattamento degli errori è stato diverso a seconda della tipologia di regole. Nella pratica, il loro trattamento è stato il seguente:

- le *rotte* già previste sul modello cartaceo sono state trattate in modalità *hard*;
- i *quesiti di tipo tabellare*, che prevedono il riempimento di una colonna subordinatamente al fatto che alla colonna precedente sia stato risposto «SI», sono stati trattati in modalità *hard*;
- le regole di *coerenza tra quesiti di modelli diversi* sono state trattate in modalità *hard*;
- le regole di *coerenza tra quesiti nell'ambito dello stesso modello* sono state trattate in modalità *soft*;
- i controlli di *quadratura tra gli importi parziali rispetto al totale* sono stati trattati in modalità *soft*;

I controlli sulle variabili quantitative di spesa rispetto a limiti minimi e massimi precedentemente stimati sono stati trattati in modalità *soft*.

1.3.1.1 Caratteristiche tecniche della procedura di acquisizione CADI sui Consumi delle famiglie

L'applicazione sviluppata richiama tre moduli, uno per ciascun modello dell'indagine (Riepilogo, Libretto e Taccuino).

I dati di ciascuna famiglia campione sono memorizzati su un unico record, evitando possibili problemi di accoppiamento. Una opportuna procedura di scarico consente l'esportazione dei dati dall'archivio Blaise in un file formato ASCII.

Le operazioni di registrazione sono attualmente effettuate per la gran parte presso il servizio CEF/A e per la restante parte presso il servizio adibito alla registrazione dati, DINF/C. Il servizio CEF/A utilizza il pacchetto di acquisizione sia per la registrazione che come supporto alla revisione, mentre il servizio DINF/C procede alla sola registrazione di modelli

preventivamente revisionati dal servizio CEF/A. I dati acquisiti presso le diverse postazioni di lavoro sono successivamente accorpati e trasferiti in ambiente mainframe VM-CMS (in prospettiva in ambiente UNIX). In ambiente VM-CMS un predisposto programma SAS provvede ad eseguire il compattamento delle spese del libretto dal dettaglio quotidiano al riepilogo settimanale.

1.3.1.2 Criteri per la definizione dei limiti di spesa per l'acquisizione controllata in BLAISE

Nel definire i limiti si è tenuto conto della distribuzione della spesa per i diversi generi e delle diverse tipologie di errore che con il programma di acquisizione si tenta di controllare e ridurre.

L'acquisizione dei dati attraverso un programma controllato, soprattutto se effettuata da personale esperto, consente di evidenziare e di intervenire in modo tempestivo su errori di trascrizione da parte della famiglia e/o del rilevatore e sugli errori di digitazione.

Le tipologie più frequenti di errore sono:

⇒ digitazione di uno zero in più o in meno;

⇒ digitazione di due o più cifre riferite a codici diversi nell'ambito di uno stesso codice;

⇒ slineamento

Lo *slineamento* può essere controllato solo in quei casi in cui la successione dei generi è tale per cui le spese debbono differire nettamente tra loro. In questo caso infatti i limiti da inserire sono nettamente differenti e l'errore viene evidenziato con maggiore facilità. Tuttavia, ai fini della sintesi dell'informazione, lo slineamento non genera gravi distorsioni in quanto, in generale si tende ad accorpare le spese nell'ambito di una stessa sezione del questionario.

La *digitazione di due o più cifre su uno stesso codice* viene in generale ben controllata da un programma di questo genere in quanto il risultato di una digitazione di questo tipo è quasi certamente esterna ai limiti di spesa imposti.

Il controllo sulla *digitazione di uno zero in più o in meno* risulta invece più complessa soprattutto in quei casi in cui le spese hanno un *range* di variazione molto elevato. Per quanto possibile è necessario, al fine di controllare un errore di questo tipo, prevedere limiti di spesa in cui il minimo e il massimo differiscano di una sola cifra (ad es. 100.000-1.000.000). Qualora il limite debba essere più ampio è bene comunque tentare di minimizzare tale possibilità di errore non includendo tra le spese accettate il limite superiore e quello inferiore (10.000 e 1.000.000 sono esclusi, mentre sono accettati 10.001 e 999.999). Nel determinare i limiti si è dunque tenuto conto di quanto esposto sopra costruendo dei limiti relativamente stretti nell'ambito di quanto evidenziato dalle distribuzioni di spesa dell'anno precedente (minimi e massimi realmente osservati, 1°, 5°, 95° e 99° percentile). Si è teso in questo modo a controllare i dati anomali nell'ambito di limiti di spesa ammissibili.

1.3.2 Prima valutazione sull'applicazione della procedura di acquisizione CADI

Gli aspetti da considerare per effettuare un'analisi «Costi-Benefici» dell'applicazione di tale tecnica, rispetto alla tradizionale modalità di acquisizione PAPI, sono i seguenti:

1. tempi di sviluppo del software di acquisizione;
2. effetti sulla fase di data-entry;
3. qualità dei dati registrati;
4. altri effetti sulla gestione dell'indagine.

Relativamente all'impatto sulla qualità dei dati, si rimanda al paragrafo 2.2; invece, per quanto riguarda gli altri tre elementi, dall'analisi effettuata nel 1996 (quando si era iniziato a registrare i dati dell'indagine sperimentale) è emerso quanto segue:

1. I tempi di sviluppo dell'applicazione sono stati giudicati più che accettabili, considerando che si è trattato della prima esperienza di utilizzo di questo pacchetto e che si è proceduto in modo «training on the job».
2. dal punto di vista della velocità di registrazione, gli operatori del centro di acquisizione di Via Tuscolana hanno valutato un allungamento dei tempi, stimabile in circa il 25%, rispetto alla tradizionale modalità di registrazione (non controllata). Ciò è d'altronde nella natura di tale metodologia, la cui corretta applicazione deve proprio consistere nel massimizzare l'accuratezza della registrazione e non certo la velocità della stessa. Dal punto di vista dell'impatto sugli operatori di data-entry, indubbiamente questo metodo modifica l'usuale modo di lavorare, imponendo la continua verifica della messaggistica sullo schermo del computer. Tuttavia gli operatori stessi hanno considerato positivamente questa innovazione, in quanto rende di fatto meno meccanica la fase di digitazione. Bisogna inoltre considerare che il software allora sviluppato, trattando tutte le regole di errore in modalità *soft*, costringeva a percorrere comunque tutti i quesiti del modello, mentre il successivo inserimento delle regole di rotta dovrebbe aver comportato una certa riduzione dei tempi di registrazione;
3. dal punto di vista della gestione dell'indagine nel suo complesso, l'allungamento dei tempi di registrazione è stato più che compensato dal supporto fornito dallo strumento alla fase di revisione, e dall'aver già effettuato in fase di disegno e sviluppo dell'applicazione per l'acquisizione dei dati lo sforzo concettuale di definizione delle regole di compatibilità che vengono poi acquisite dalla successiva fase di controllo e correzione.

1.4 Il trattamento dei dati

1.4.1 Il processo di trattamento dei dati

Il trattamento dei dati rilevati nell'ambito della nuova indagine sui consumi delle famiglie è piuttosto complessa ed articolata. Per ciascuno dei segmenti di cui si compone, si è mirato alla standardizzazione delle azioni al fine di eliminare, per quanto possibile, l'elemento soggettivo da tale processo.

Definito il piano delle compatibilità del questionario, individuando e descrivendo il complesso delle regole (riguardanti sia le variabili quantitative che quelle qualitative), si è stabilito il tipo di azione da compiere nel momento in cui venga accertata una determinata incompatibilità.

Sono stati previsti in sequenza sia interventi di tipo manuale che informatico qui di seguito brevemente descritti.

- 1) Revisione quantitativa preliminare e accertamento della corrispondenza tra diversi questionari relativi ad una singola unità di secondo stadio - Questo tipo di revisione consente di individuare repentinamente eventuali problematiche connesse con la rilevazione sul campo e di intervenire tempestivamente presso gli organi periferici prima di procedere con le successive fasi della revisione.

2) Revisione manuale qualitativa dei modelli - A tal fine è stato predisposto un manuale in cui sono raccolti i criteri standard che guidano l'azione dei revisori uniformando il processo⁴. Il principio alla base di tali criteri è quello della salvaguardia del dato riportato dalla famiglia o dall'intervistatore sui questionari. Laddove si riscontri un'incompatibilità tra due o più variabili del questionario, si procede ricontattando la famiglia stessa. Qualora l'incompatibilità non possa essere eliminata, si procede all'immissione del dato così come rilevato per poi trattare l'incompatibilità informaticamente.

3) Acquisizione controllata - Trattandosi di acquisizione controllata effettuata dagli stessi revisori esperti, tale fase è strettamente connessa con quella di revisione. In alcuni casi infatti le incongruenze di tipo formale vengono risolte attraverso i controlli inseriti nel programma di immissione (BLAISE), altre vengono segnalate e trattate opportunamente, secondo criteri standard dal revisore.

4) Controllo e correzione dei dati - Viene utilizzata una combinazione di programmi deterministici e probabilistici che individuano e correggono dapprima le incongruenze relative alle variabili qualitative e, successivamente, quelle relative alle variabili quantitative. In particolare per le variabili qualitative si utilizzano, in successione, un programma di tipo deterministico ed il sistema SCIA per il controllo delle coerenze e l'imputazione automatica dei dati con metodo probabilistico.

Il programma deterministico ha lo scopo evidenziare e correggere, sulla base delle regole di revisione manuale e di quelle inserite in BLAISE per l'immissione controllata, le incongruenze che ancora persistono dopo l'immissione dei dati. Il BLAISE infatti in alcuni casi, pur segnalando le incongruenze, permette di tralasciare la correzione nel momento in cui venga verificata la corrispondenza del dato digitato con quello riportato sul questionario. Si evita in questo modo che i singoli digitatori effettuino valutazioni di merito, preferendo intervenire in una fase successiva in modo uniforme e standardizzato attraverso programmi specifici.

Dopo aver verificato l'effettiva rispondenza dei dati alle regole di revisione ed immissione, si procede all'imputazione probabilistica delle variabili qualitative tramite SCIA. Sono state previste due fasi di correzione: nell'ambito della prima vengono corrette le variabili socio-demografiche individuali, con la seconda si imputano le restanti variabili qualitative del questionario. Successivamente vengono corrette le variabili quantitative (spese). In questo caso l'imputazione di voci mancanti o incongruenti, avviene con metodo da donatore (RIDA) laddove le frequenze lo consentono, attraverso l'impiego di massimi e minimi randomizzati quando le frequenze di spesa sono basse e quindi non consentono l'individuazione di un donatore.

1.4.2 Il trattamento delle variabili qualitative

Le variabili qualitative rilevate con l'indagine sui consumi delle famiglie possono essere distinte in tre gruppi:

1. le variabili identificative del modello;
2. le variabili relative alle notizie sui componenti della famiglia;
3. le variabili relative alle caratteristiche dell'abitazione, al possesso di beni durevoli e alle abitudini di spesa.

⁴ C. Ceccarelli, D. Grassi, N. Pannuzi (1997)

Queste vengono trattate diversamente nell'ambito della complessa fase di controllo e correzione dei dati. Tutte, infatti, vengono sottoposte ad una revisione manuale preliminare secondo le modalità standard incluse nel manuale di istruzione ai revisori, successivamente tuttavia il trattamento informatico è differenziato. I tre gruppi di variabili verranno dunque trattate separatamente.

1.4.3 I codici identificativi del modello

Ogni modello viene individuato univocamente attraverso una chiave di 23 posizioni contenente i codici identificativi del modello ed in particolare le seguenti informazioni:

- Provincia
- Comune
- Numero generale progressivo della famiglia
- Codice intervistatore
- Numero d'ordine della famiglia
- Elenco base o suppletivo
- Anno
- Periodo

In fase di revisione manuale si verifica in primo luogo l'esattezza di tali codici e la rispondenza di questi con quelli riportati sulla documentazione fornita dal comune (elenchi base e suppletivi delle famiglie, modelli BF/5 destinati al rilevatore con le caratteristiche anagrafiche delle famiglie da intervistare nel dato periodo). In alcuni casi per la ricostruzione della chiave esatta si rende necessario ricontattare il comune stesso.

Particolare attenzione viene posta, soprattutto nei primi mesi dell'anno, alla verifica del passo d'estrazione, valutando cioè se il comune abbia rispettato le regole di selezione delle famiglie dalla lista precedentemente predisposta.

Si procede inoltre a verificare il corretto accoppiamento dei diversi questionari relativi ad un'unica famiglia (BF/1, BF/2 ed eventualmente BF/1A). Successivamente ulteriori controlli vengono effettuati in fase di acquisizione. L'acquisizione in BLAISE consente in questo caso due tipi di controllo:

1. a livello di singolo PC non è possibile inserire modelli aventi le stesse variabili identificative;
2. per le variabili provincia, comune e mese di rilevazione non possono essere inserite combinazioni non previste dal piano di campionamento.

Ulteriori verifiche riguardo ai codici identificativi vengono effettuate sul file mensile completo al fine di individuare errori incorsi in fase di registrazione e accorpamento dei dati provenienti dai singoli PC in lavorazione. In particolare, attraverso programmi SAS, viene effettuato dapprima un controllo quantitativo per verificare la coincidenza tra numero di questionari pervenuti e numero di questionari immessi. Nel caso di mancata corrispondenza tra i due quantitativi si risale al modello cartaceo procedendo, laddove se ne verificasse l'occorrenza, all'immissione dei modelli mancanti. Successivamente si valutano singolarmente le anomalie (duplicazioni, chiavi vuote, ecc.) procedendo ad eliminarle nel modo più opportuno.

1.4.4 Notizie sui componenti della famiglia

Il controllo e la correzione degli errori casuali presenti nelle variabili relative alle notizie sui componenti della famiglia e nelle variabili relative alla abitazione e alle abitudini di spesa è effettuato mediante il software SCIA e da alcuni programmi sviluppati in linguaggio SAS.

Il processo di correzione è articolato in applicazioni sequenziali distinte del sistema SCIA: vengono controllate e corrette prima le variabili relative ai componenti della famiglia e successivamente quelle relative alla abitazione e alle abitudini di spesa. L'esecuzione di applicazioni sequenziali distinte non comporta alcuna perdita di ottimalità poiché le regole di incompatibilità definite per le diverse fasi non presentano variabili in comune.

Nel presente paragrafo viene illustrato il processo di correzione delle variabili relative alle notizie sui componenti della famiglia per i dati dell'indagine a regime 1997. La descrizione della procedura di correzione per le variabili relative alla abitazione e alle abitudini di spesa è demandata al paragrafo 1.4.6.

La procedura di controllo per le variabili relative alle notizie sui componenti della famiglia è basata sulla verifica simultanea delle:

- relazioni di incompatibilità tra valori di variabili relative a differenti componenti di una stessa famiglia (vincoli inter-componenti);
- relazioni di incompatibilità tra valori di variabili relative a un singolo componente (vincoli intra-componente).

La fase di correzione consiste nell'attribuire alle variabili con valori "errati" (inconsistenti rispetto alle regole di incompatibilità o mancanti laddove richiesti) i valori prelevati dalle corrispondenti variabili in unità con risposte "esatte" (*imputazione "da donatore"*) oppure prelevati dalle distribuzioni marginali delle variabili stesse (*imputazione forzata*).

La considerazione simultanea di entrambe le tipologie di vincoli comporta l'organizzazione dei dati in record familiari contenenti le variabili di tutti i componenti relativi ad una stessa famiglia e quindi nella denominazione delle variabili in modo che siano identificabili i numeri d'ordine dei componenti a cui si riferiscono. Questo approccio consente di determinare l'insieme minimo di variabili da correggere, nel rispetto del principio del minimo cambiamento, con la certezza di poter effettuare la correzione ovvero liberare il record dagli errori. Tale approccio garantisce l'ottimalità del controllo (e localizzazione dell'errore), ma potrebbe non garantire sempre l'ottimalità della procedura di imputazione. L'utilizzo del record familiare comporta, infatti, la gestione *posizionale* delle variabili relative ai diversi componenti e quindi le caratteristiche individuali del generico *i-esimo* componente vengono corrette prelevando i valori dalle corrispondenti variabili del componente *i-esimo* (che occupa cioè la medesima posizione) nel record donatore (imputazione da donatore) o in base alle distribuzioni marginali definite dalla posizione (imputazione forzata). Tale procedura potrebbe non essere appropriata laddove si verificano l'uno o l'altra delle seguenti circostanze:

1. esigua dimensione del serbatoio dei donatori (tale situazione si verifica soprattutto quando si debbano correggere variabili relative agli ultimi componenti di famiglie molto numerose);
2. mancata corrispondenza tra la "posizione" nella famiglia (coniuge, figlio, ascendente, etc. della persona di riferimento) del componente con valori errati e la "posizione" nella famiglia del componente utilizzato come donatore. Tale corrispondenza è sempre realizzata solo per la persona di riferimento in quanto solo la sua posizione nel record è fissa essendo rigorosamente registrata come primo componente.

A prima vista il secondo problema potrebbe essere risolto associando a tutte le variabili la variabile ausiliaria di accoppiamento: "relazione con la persona di riferimento". Questa soluzione non è stata però attuata poiché (i) le variabili ausiliarie di accoppiamento sono considerate dal sistema solo nel caso in cui l'imputazione sia di tipo sequenziale e (ii) il sistema consente di definire una sola variabile ausiliaria di accoppiamento per una data variabile negando così la possibilità di preservare altre relazioni doppie statisticamente più significative. In ogni caso permanerebbe il problema della scarsa numerosità degli ultimi componenti di famiglie molto numerose.

Per evitare imputazioni anomale o poco probabili si è provveduto ad operare per due fasi successive utilizzando due distinte applicazioni sequenziali del sistema SCIA. Tali applicazioni (CN971A e CN971B) effettuano:

- il controllo di tutte le relazioni di incompatibilità tra le variabili, la corrispondente localizzazione degli errori e la correzione delle variabili relative alle caratteristiche demografiche individuali (*relazione con la persona di riferimento, sesso, anno di nascita, stato civile*); nel seguito si indicheranno tali variabili come *variabili demografiche*;
- la correzione delle variabili relative alle altre caratteristiche individuali (*titolo di studio, condizione professionale, branca di attività economica prevalente, posizione nella professione*). Nel seguito si indicheranno tali variabili come *variabili socio-economiche*.

Nella applicazione CN971A si procede come se il passo di correzione fosse unico, ovvero si opera con il record familiare, si definiscono le regole di incompatibilità relative ad entrambe le tipologie di vincoli (inter-componenti e intra-componente) e al complesso delle variabili relative alle notizie sulla famiglia (demografiche e socio-economiche), si definiscono altresì le eventuali fissità e le eventuali variabili ausiliarie di accoppiamento (solo per le variabili demografiche). Il sistema procede ad individuare i record errati, a localizzare gli errori e ad attribuire alle variabili selezionate i nuovi valori. I valori imputati per le variabili demografiche sono considerati validi mentre i valori attribuiti alle variabili socio-economiche vengono "rimossi" e sostituiti con quelli originari. Ciascun record familiare viene frammentato in tanti record individuali quanti sono i componenti della famiglia e si procede ad eseguire l'applicazione CN971B.

Il piano di incompatibilità definito per l'applicazione CN971B è costituito dalle regole ove è presente almeno una variabile socio-economica, e poiché le variabili demografiche sono considerate fisse, la fase di localizzazione degli errori ripropone per la correzione le stesse variabili socio-economiche individuate nella corrispondente fase nell'applicazione CN971A. Si procede ad imputare i nuovi valori per le variabili socio-economiche. Da questa descrizione si evince che:

- le variabili demografiche sono corrette utilizzando l'intero record familiare: l'uso del record familiare è necessario per verificare le relazioni tra le caratteristiche demografiche di differenti componenti (ad esempio "la persona di riferimento ed il suo coniuge devono avere sesso diverso"); la correzione delle variabili demografiche è effettuata tenendo in considerazione anche le informazioni presenti nelle variabili socio-economiche, ovvero le eventuali incompatibilità tra i due gruppi di variabili;
- le variabili socio-economiche sono corrette utilizzando record individuali ottenuti dalla frammentazione per componente del record familiare originario: il record donatore è così selezionato dall'intero serbatoio dei componenti e non solo dal serbatoio definito dalla posizione nel record del componente errato; la correzione delle variabili socio-

economiche è effettuata tenendo fissi i valori delle variabili demografiche in quanto corrette nella fase precedente.

Le manipolazioni dei dati, necessarie per eseguire le applicazioni del sistema SCIA, sono effettuate mediante esecuzione di alcuni programmi scritti in linguaggio SAS. L'intera procedura di controllo e correzione per le variabili relative alle notizie sui componenti della famiglia è articolata nelle seguenti fasi:

1. Copia dei valori delle variabili socio-economiche, calcolo delle variabili *età* in anni per ciascun componente la famiglia e predisposizione del file da sottomettere al sistema SCIA per la correzione delle variabili demografiche.
2. Esecuzione dell'applicazione di SCIA per la correzione delle variabili demografiche.
3. Ripristino dei valori originari delle variabili socio-economiche, calcolo di alcune variabili aggiuntive quali la variabile di stratificazione *ripartizione geografica* e la variabile di accoppiamento ausiliario *sesso età* e predisposizione del file dei record individuali da sottomettere al sistema SCIA per la correzione delle variabili socio-economiche.
4. Esecuzione dell'applicazione di SCIA per la correzione delle variabili socio-economiche.
5. Esecuzione del programma che ricostruisce il file dei record familiari con tutte le variabili: le variabili demografiche e socio-economiche sono prelevate dal file di dati corretto da SCIA mentre le variabili relative alla abitazione e alle abitudini di spesa sono prelevate dal file dei dati originari.

1.4.4.1 Correzione delle variabili demografiche

Il controllo e la correzione delle variabili demografiche è effettuata utilizzando i record familiari, ovvero record che presentano i valori di tutte le variabili rilevabili per tutti i componenti la stessa famiglia. La posizione delle variabili nei record è fissa: vengono prima le variabili relative al primo componente, poi quelle relative al secondo componente e così via. E' stata considerata una dimensione massima della famiglia pari a dodici componenti e quindi i record familiari, che sono di lunghezza fissa, prevedono al massimo l'inserimento dei dati relativi a dodici componenti. Per ogni famiglia sono presenti solo i valori delle variabili relative ai componenti dichiarati mentre le variabili relative a componenti non dichiarati presentano valori mancanti.

Il controllo, ed eventuale imputazione, dell'anno di nascita è effettuato operando sull'età in anni (differenza tra *anno di rilevazione* e *anno di nascita*) del componente. Quindi, prima di eseguire i moduli di SCIA si procede a calcolare l'età in anni per ciascun componente della famiglia. La riconversione da *età* in *anno di nascita* è effettuata alla fine dell'intero processo di correzione mediante esecuzione di un predisposto programma SAS.

Le regole di incompatibilità utilizzate per la correzione delle variabili demografiche sono riportate nella Appendice A del paragrafo. Ciascuna regola presente in Appendice A si riferisce al generico componente la famiglia quindi nel piano di incompatibilità ogni regola è specificata per ogni possibile componente (relazioni intra-componente) o per ogni combinazione di componenti (relazioni inter-componenti), dando luogo ad un totale di 455 regole esplicite.

Nel piano di incompatibilità utilizzato per la correzione delle variabili demografiche sono presenti oltre alle regole composte dalle sole variabili demografiche anche regole composte dalle sole variabili socio-economiche e regole composte da variabili appartenenti ad entrambi i gruppi. Il motivo scaturisce dall'obiettivo che questa applicazione si prefigge: verificare la

presenza di errori nei dati (indipendentemente dal gruppo di appartenenza delle variabili o dalla tipologia inter-componenti o intra-componente), localizzare correttamente le variabili da correggere e imputare i nuovi valori per le sole variabili demografiche.

Dal punto di vista delle unità di rilevazione soggette al controllo è possibile distinguere tra:

- regole inter-componenti, che esprimono situazioni di non correttezza dei record causate dalle incompatibilità tra valori di variabili relative a differenti componenti di una stessa famiglia (regole 1-7);
- regole intra-componente, che esprimono situazioni di non correttezza dei record causate dalle incompatibilità tra valori di variabili relative a un singolo componente (regole 8-32);

Dal punto di vista della natura delle regole è possibile distinguere tra:

- regole strutturali, che derivano direttamente dalla struttura del questionario; esse esprimono situazioni di non correttezza dei record causate dal mancato rispetto delle norme di compilazione (rotte) imposte nella somministrazione delle domande (regole 19-22, 29-32);
- regole sostanziali, che derivano da considerazioni di tipo statistico o da conoscenze specifiche a priori sulle relazioni socio-culturali esistenti per le variabili rilevate; esse esprimono situazioni di non correttezza dei record causate da combinazioni di risposte non plausibili o inconsistenti (regole 1-18, 23-28).

1.4.4.2 Correzione delle variabili socio-economiche

Il controllo e la correzione delle variabili socio-economiche è effettuata utilizzando i record individuali ovvero record che presentano i valori di tutte le variabili rilevabili per un solo componente. I record individuali sono ottenuti dalla frammentazione del record familiare considerando esclusivamente i componenti dichiarati, quindi, i dati relativi ad una famiglia composta da *i* componenti sono ripartiti su *i* record individuali.

Nei record individuali sono presenti i valori originari delle variabili socio-economiche da correggere con la applicazione in esame e i valori delle variabili demografiche *età* e *sex* risultanti dalla procedura di controllo e correzione CN971A. Queste due variabili demografiche sono riportate in quanto coinvolte in regole di incompatibilità con le variabili socio-economiche sottoposte a correzione. Sono inoltre presenti i valori di due nuove variabili:

- la *ripartizione geografica* ottenuta accorpando le provincie di rilevazione nelle cinque macro ripartizioni (nord-ovest, nord-est, centro, sud, isole) utilizzata per la stratificazione del serbatoio dei donatori;
- una variabile composta ottenuta dalle combinazioni delle modalità della variabile *sex* con le classi di età <15, 15-29, 30-49, 50-64, ≥65 utilizzata come variabile ausiliaria di accoppiamento. Tale variabile sarà denominata nel seguito *sex_età*.

I record sono ordinati secondo i valori della variabile di stratificazione.

Le regole di incompatibilità utilizzate per la correzione delle variabili socio-economiche sono quelle in cui è presente almeno una variabile di questo tipo ovvero le regole 12-32 riportate in Appendice A (sono state escluse le regole che coinvolgono esclusivamente variabili demografiche). Poiché l'applicazione opera con i record individuali è stato sufficiente specificare solo 21 regole esplicite. Le regole utilizzate in questa applicazione sono un sottoinsieme di quelle utilizzate nell'applicazione precedente. Il motivo scaturisce dall'obiettivo che questa applicazione si prefigge: imputare i nuovi valori per le sole variabili

socio-economiche. Il piano di incompatibilità è caratterizzato dalla presenza di sole regole intra-componente essendo quelle inter-componenti specificate esclusivamente per le variabili demografiche. Sono invece presenti entrambe le tipologie strutturali e sostanziali.

1.4.5 La tipologia familiare

Per quanto concerne la tipologia familiare, dall'analisi dei risultati effettuata sui dati puliti sono emerse alcune incongruenze evidenziabili più avanti e ricollegabili con il metodo di correzione utilizzato.

Per comprendere a pieno tale aspetto è necessario tenere presenti alcuni aspetti.

In particolare:

- Le variabili che vengono coinvolte nella definizione della tipologia familiare sono:
 1. relazione di parentela
 2. sesso
 3. età
 4. stato civile.

Dalla diversa combinazione delle modalità assunte da queste variabili per i singoli individui che compongono la famiglia si arriva a ricostruire la nuova variabile «tipologia familiare».

- Riguardo alle quattro variabili considerate, le regole scritte in SCIA consentono di controllare e correggere *alcune* incongruenze riguardanti i singoli individui e di imputare i dati mancanti evitando di ingenerare *alcune* incompatibilità.
- Le informazioni individuali relative ai singoli componenti la famiglia vengono registrate su uno stesso record (il record familiare) in sequenza, così come vengono riportate sul questionario, a prescindere dalla relazione che questi dichiarano di avere con la Persona di Riferimento (PR).
In fase di controllo e correzione gli individui vengono considerati secondo la posizione che occupano nel record.
La correzione viene effettuata scegliendo donatori tra gli individui che occupano nel record la medesima posizione dell'individuo che deve essere corretto.
- La posizione non presenta alcuna corrispondenza formale con la relazione di parentela del singolo componente se non nel caso della PR che è sempre collocata in prima posizione.
- La correzione per posizione con metodo da donatore implica quindi che la donazione può essere effettuata anche tra componenti con caratteristiche del tutto diverse tra loro.
- Come è noto inoltre non è possibile considerare nel loro complesso le relazioni tra gli individui ed in particolare non è possibile mettere in relazione le età dei singoli componenti.
- Da quanto detto emerge come la variabile «tipologia familiare» non si possa considerare «controllata».
- Gli esempi allegati, di casi realmente riscontrati, mostrano l'evidenza di quanto affermato. SCIA infatti non prevede la correzione di alcuni errori evidenti che emergono da una analisi complessiva delle relazioni tra i componenti la famiglia. Infatti sui dati finali permangono situazioni non plausibili quali ascendenti-celibati (non è impossibile ma è poco plausibile), ascendenti con età inferiore alla PR o solo di poco superiore e, comunque, «tipologie familiari» non plausibili. In uno degli esempi riportati, SCIA imputa dati mancanti creando anomalie di questo tipo.

ESEMPI

1. SCIA imputa stato civile celibe ad un ascendente di 76 anni. Ciò avviene perché l'ascendente si trova in 5° posizione che generalmente è occupata da figli (a questa famiglia viene attribuito un peso annuale pari a 671,24).
2. SCIA accetta che un ascendente sia più giovane del discendente poiché non esiste controllo incrociato tra le età dei componenti (a questa famiglia viene attribuito un peso annuale pari a 652,52).
3. SCIA accetta che in una famiglia di 7 componenti (2 coniugi, 3 figli, 2 ascendenti) i due ascendenti, di sesso opposto, di 78 e 76 anni rispettivamente, siano celibi (a questa famiglia viene attribuito un peso annuale pari a 816,26).
4. SCIA accetta che un ascendente di 78 anni sia celibe (a questa famiglia viene attribuito un peso annuale pari a 1092,79).

In conclusione, sebbene l'aggiunta di una regola che valuti l'incompatibilità tra gli ascendenti e lo stato civile potrebbe portare all'eliminazione di alcune di queste problematiche, non si verrebbe comunque ad una soluzione complessiva e sostanziale. Né si può ipotizzare che un'eventuale stratificazione secondo la relazione di parentela in sede di correzione possa ritenersi risolutiva, in quanto questa potrebbe ingenerare problemi connessi con l'esiguità del serbatoio-donatori.

Per un appropriato trattamento della «tipologia familiare» potrebbe dunque essere di ausilio sviluppare una nuova tecnica di correzione che possa tenere conto delle necessità specifiche.

1.4.6 Le caratteristiche dell'abitazione, il possesso di beni durevoli e le abitudini di spesa

Anche per queste variabili il controllo e l'eventuale correzione avvengono attraverso successivi passaggi.

In primo luogo vengono controllate dai revisori esperti che procedono, sulla base delle istruzioni generali loro fornite, ad eliminare le principali incongruenze dal supporto cartaceo e quindi in fase di digitazione. L'idea di fondo è, come già detto, quella di ridurre l'apporto soggettivo in fase di revisione e correzione dei dati. In questo senso i revisori dovranno individuare le incongruenze predisponendo la fase di imputazione automatica, senza inserire alcuna informazione autonomamente.

Ad esempio, riguardo alla tipologia dell'abitazione (Villa o villino, Signorile, Civile, economica e popolare, Rurale, Impropria), quando questa risulti impropria devono risultare mancanti tutte le informazioni successive fino al quesito 31 (come indicato sul questionario). Qualora le informazioni siano presenti il revisore deve eliminare l'indicazione di abitazione impropria non barrando alcun altro codice. Se l'anno di costruzione dell'abitazione è superiore a quello di occupazione il revisore deve procedere cancellando l'anno di costruzione (ritenuto meno attendibile).

In questa fase, il revisore, constata l'incongruenza, potrà procedere, quando lo riterrà opportuno, a ricontattare l'intervistatore o la famiglia stessa ed apportando le eventuali correzioni.

Sia la revisione che la digitazione dovranno avvenire salvaguardando massimamente il dato riportato sul questionario. Per tale motivo il programma di acquisizione controllata, segnalerà alcune incongruenze permettendo tuttavia di eludere la segnalazione una volta constatato che il dato immesso corrisponde effettivamente con quello annotato sui questionari.

Tutte le informazioni mancanti verranno imputate successivamente in modo automatico dopo aver controllato (con SASDET) che le regole di revisione siano state effettivamente rispettate. In particolare si controllano e correggono in modo deterministico:

- 1) la tipologia dell'abitazione fissandola come impropria quando mancano contemporaneamente tutte le informazioni successive, negli altri casi si interviene in modo probabilistico;
- 2) l'anno di costruzione e l'anno di occupazione dell'abitazione nel solo caso in cui almeno uno dei due sia presente;
- 3) la superficie dell'abitazione con il numero di stanze nel caso in cui almeno una delle due sia presente (si fissa un rapporto standard tra le due variabili);
- 4) il servizio di acqua calda verrà dichiarato disponibile quando risulta indicata una modalità riguardante il tipo di impianto di riscaldamento dell'acqua;
- 5) la presenza di elettrodomestici e di energia elettrica (e quindi di bolletta);
- 6) la compatibilità tra servizio telefonico, apparecchio telefonico e bolletta telefonica;
- 7) alcune informazioni riguardanti la seconda abitazione;

Con SCIA vengono controllate e corrette tutte le altre incompatibilità (vedi Appendice B), per cui ad esempio nel caso di tipologia dell'abitazione mancante ed in presenza delle altre informazioni riguardanti l'abitazione, SCIA imputerà con metodo probabilistico una tipologia diversa da quella impropria. Nel caso invece che sia stata indicata la presenza di acqua calda ma non il tipo di impianto di riscaldamento acqua, SCIA imputerà il tipo di impianto. SCIA inoltre controllerà ed eventualmente correggerà tutte le incompatibilità relative al titolo di godimento dell'abitazione, al proprietario dell'abitazione e al tipo di contratto. Si è inoltre stabilito che l'eventuale imputazione degli elettrodomestici debba avvenire solo in presenza di energia elettrica. Infine SCIA controlla la dichiarazione delle modalità di pagamento.

L'imputazione avviene, oltre che nel rispetto delle regole di compatibilità (e quindi nell'ipotesi di minimo cambiamento), utilizzando come variabile di strato una variabile ottenuta come combinazione della tipologia familiare e della zona geografica di residenza. In particolare per le due variabili la disaggregazione presa in considerazione è la seguente:

Zona geografica di residenza	Nord / Centro / Sud
Tipologia familiare	2 componenti entrambi anziani
	2 componenti adulti
	2-4 componenti con almeno un giovanissimo
	3-4 componenti adulti o anziani
	5+ componenti.

Come sintesi delle due si ottengono 15 modalità di stratificazione. In ciascun caso le frequenze rilevate sono sufficientemente elevate per consentire l'uso di tale variabile come variabile di stratificazione. Per la corretta imputazione dei valori mancanti o incongruenti sono state stabilite le fissità delle variabili per cui ad esempio, fissata per ovvie ragioni la fissità della tipologia familiare a 9, la tipologia dell'abitazione presenta una fissità superiore rispetto ai servizi dell'abitazione stessa mentre per le altre variabili non è stata stabilita alcuna fissità.

1.4.7 Il trattamento delle variabili quantitative

Il trattamento delle variabili quantitative (essenzialmente di spesa) della nuova indagine sui consumi delle famiglie è costituito da due momenti fondamentali: il controllo e la correzione. La fase di controllo è quella relativa alla definizione ed all'individuazione dei valori mancanti

e/o anomali, che quindi non possono essere accettati così come provenienti dalla rilevazione, mentre la fase di correzione consiste nella sostituzione (imputazione) dei valori individuati come inaccettabili con altri provenienti dalle corrispondenti variabili di osservazioni (unità donatrici) che presentano valori (risposte) completi e corretti. Il software utilizzato per questa ultima fase è RIDA (Ricostruzione delle Informazioni con Distribuzione Automatica).

1.4.7.1 Individuazione dei valori mancanti

L'individuazione dei valori mancanti per le variabili quantitative di spesa si basa sulle relazioni esistenti tra le variabili stesse e le variabili qualitative di controllo ad esse connesse. L'introduzione di variabili indicatrici di presenza/ assenza di valore, relative ad ogni variabile di spesa, consente di risolvere le eventuali incompatibilità con i valori assunti dalle variabili qualitative all'interno del sistema SCIA. La prima parte del processo verifica quindi la congruenza tra l'informazione presente nella variabile indicatrice (sottoposta al trattamento tramite SCIA) e il valore registrato nella corrispondente variabile quantitativa. L'informazione relativa alla variabile indicatrice è la base di partenza e pertanto, compatibilmente ad essa, il valore della variabile quantitativa può:

- rimanere inalterato;
- risultare mancante e quindi essere sottoposto ad imputazione
- essere eliminato.

Con tale processo vengono anche risolti in modo ottimale tutti i problemi connessi al legame tra variabili quantitative e qualitative tra loro collegate.

Nel caso dell'indagine sui consumi delle famiglie, sono individuati come valori mancanti quelli relativi alle variabili quantitative di spesa in presenza:

- di un dichiarato codice identificativo del bene (ad. es. il codice di un bene regalato senza il corrispondente valore)
- dell'indicazione di avvenuta spesa mediante la corrispondente variabile qualitativa di controllo (domande filtro che precedono la sezione relativa alle singole voci, ed es. Nel corso dell'ultimo mese la sua famiglia ha sostenuto spese per la custodia di veicoli? SI o NO).
- le spese relative all'abitazione (energia elettrica, gas da rete e telefono, relativi all'ultimo pagamento) nel caso in cui la abitazione risulti dotata dei relativi servizi.

1.4.7.2 Individuazione dei valori anomali

Nel caso specifico dell'indagine sui consumi delle famiglie, viene definito valore anomalo il valore esterno ad una pre-definita regione di accettazione.

I limiti di accettazione, per ciascuna voce di spesa presente all'interno del questionario, sono stati definiti avvalendosi di alcune informazioni sia esterne sia interne disponibili al momento dell'indagine.

In particolare:

- si è tenuto conto dell'informazione dei prezzi minimi e massimi sul territorio nazionale forniti dal Servizio Prezzi per vari beni e servizi oggetto di indagine
 - si è tenuto conto della distribuzione campionaria delle spese relative alle singole voci;
- si sono individuati i limiti inferiori e superiori della regione di accettazione, massimizzando la probabilità di escludere i soli outliers tramite la tecnica D-Maso (Distance Measurement

Algorithm for selection of Outlier). Il metodo si basa sul principio che la probabilità per un valore di essere outlier è maggiore per il più estremo tra due valori (contigui) prossimi ad una coda della distribuzione se l'intervallo tra essi è significativamente superiore alla gran parte degli altri intervalli. Il meno estremo tra i due valori è assunto come limite (superiore od inferiore) per l'intervallo di accettazione. Tutte le informazioni disponibili sono state utilizzate per l'individuazione dei limiti, in quanto nessuna delle tecniche, singolarmente utilizzata, poteva fornire un'informazione soddisfacente. Anche la tecnica D-Maso, infatti, oltre a presentare una variabilità stagionale, se calcolata a livello trimestrale, presenta problemi di applicabilità per distribuzioni basate su un ristretto numero di osservazioni. Pertanto i limiti sono stati definiti in maniera estremamente empirica, avendo sempre l'accuratezza di definire gli intervalli in termini di plausibilità ed avendo ben presente il grafico delle frequenze relative con cui le singole variabili vanno fuori limite; questo è infatti un indicatore potente della necessità di aggiornare i limiti utilizzati: un numero eccessivo di fuori limiti (anche in termini percentuali) può indicare infatti l'eliminazione di un determinato comportamento di spesa piuttosto che l'individuazione di valori errati.

Controllo preventivo dell'effetto di RIDA tramite un indice di distanza dall'intervallo di accettazione

L'adozione di intervalli di accettazione per l'individuazione dei valori anomali consente di segnalare i valori grezzi (da correggere) responsabili delle maggiori variazioni rispetto ai valori imputati. Tramite la correzione con RIDA, il valore segnalato come anomalo sarà sostituito con un valore certamente interno all'intervallo di accettazione.

La distanza tra donatore e unità affetta da errore è calcolata rispetto alla possibile natura delle variabili di matching:

- *Variabili qualitative sconnesse*: $d=0$, se le due unità presentano la stessa modalità, $d=1$ altrimenti;
- *Variabili qualitative ordinate*: siano m le modalità della variabile, $d=0$ se la modalità è la stessa, $d=1$ se sono adiacenti, $d=2$ se tra esse c'è una modalità e così via fino al valore massimo pari a $m-1$; d è poi normalizzata tra 0 ed 1 dividendo il suo valore per $m-1$.
- *Variabili qualitative ordinate di tipo telescopico*: sono le variabili con una struttura gerarchica, che indica un gruppo primario, un sotto-gruppo e così via fino ad arrivare ad un sottogruppo minimale. Pertanto $d=0$ se la modalità è identica per entrambe le osservazioni, $d=1$ se le modalità appartengono allo stesso gruppo minimale, 2 se appartengono a due sottogruppi minimali ma sono contenute nello stesso sottogruppo di livello immediatamente superiore a quello minimale. Se, $K-1$ è il numero di livelli definiti, d varierà tra 0 e K , che risulta quindi la distanza massima possibile, ancora una volta d viene poi normalizzata.
- *Variabili quantitative*: d è definita come il valore assoluto della differenza tra i valori assunti dalle due unità.

E' possibile ottenere un'informazione sulle potenzialità perturbative da parte di RIDA sul singolo valore grezzo mediante il seguente indice:

$$I_2 = \frac{GR - L}{L} \times 100$$

Dove L è l'estremo superiore dell'intervallo di accettazione della generica voce di spesa presa in considerazione nel caso in cui il valore grezzo sia superiore ad esso, ed è l'estremo

inferiore nel caso in cui il valore grezzo sia inferiore. Tale indice misura l'entità di quanto il valore grezzo esterno all'intervallo di accettazione si discosta dall'estremo più vicino. Assume quindi valori positivi nel caso in cui il valore grezzo sia superiore al limite superiore e valori negativi nel caso in cui il valore grezzo sia inferiore al limite inferiore. Prima di passare al processo di correzione vero e proprio, per ogni variabile è stata analizzata la frequenza con cui i dati grezzi si discostavano in misura eccessiva rispetto ai limiti di accettazione. Successivamente si è deciso quali variabili correggere con Rida e quali invece con altri metodi di correzione, tenendo anche conto della numerosità campionaria delle osservazioni candidate come donatori per l'imputazione.

1.4.7.3 *Correzione delle variabili quantitative*

Alcuni dei beni considerati non hanno permesso l'imputazione tramite Rida a causa della esigua numerosità dei possibili donatori e si è pertanto adottata la seguente tecnica deterministica:

- se il valore dichiarato supera il valore massimo dell'intervallo, esso viene sostituito con il valore massimo dell'intervallo meno una componente casualmente generata compresa nel 10% del valore del limite;

se il valore dichiarato supera il minimo, esso è rimpiazzato con il minimo più la componente casuale. Tale tecnica viene adottata per tutte le voci di spesa che presentano una frequenza di acquisto molto ridotta.

1.4.7.4 *Procedura di correzione delle variabili di spesa*

Le variabili quantitative da sottoporre a controllo sono circa 280 per la nuova indagine sui consumi. Si è proceduto ad un raggruppamento in circa 20 sottogruppi, omogenei rispetto alla scheda parametro utilizzato. Si sono quindi definite due tipi di variabili ausiliarie: quelle di strato (i possibili donatori sono solo quelli che presentano le stesse modalità del record da imputare rispetto a tali variabili) e quelle di matching (rispetto a queste si definisce una funzione di distanza e si sceglie il donatore a distanza minima, vedi appendice)

L'unica variabile di strato considerata è la ripartizione geografica (nord-ovest, nord-est, centro, sud, isole); tale stratificazione risulta abbastanza ampia da non impoverire il serbatoio dei donatori e riesce a tener conto della differenziata struttura dei prezzi a livello nazionale.

Per ogni variabile soggetta a correzione vengono preselezionati i possibili donatori; se le variabili di matching sono di tipo quantitativo, non possono presentarsi valori mancanti.

Per la maggior parte delle variabili si è scelta la classe di reddito come variabile di matching; la scelta di tale variabile è stata dettata dalle seguenti considerazioni:

- si tratta di una variabile puramente ausiliaria nella ricostruzione di altre informazioni;
- si ritiene, sulla base anche delle precedenti indagini sul reddito, che tale classe sia sistematicamente sottostimata e che una siffatta distorsione non dovrebbe influenzare il suo utilizzo come variabile di matching nel processo di ricostruzione delle voci di spesa sulla base della distanza minima.

Ai fini del calcolo della distanza, la classe di reddito è da considerare quantitativa e come tale non può presentare valori mancanti. Tale variabile presenta tuttavia dei valori mancanti dovute non ad errori di rilevazione ma ad un atteggiamento di reticenza da parte dell'intervistato. Trattandosi di una variabile estremamente informativa sull'atteggiamento della famiglia, si è utilizzata la stessa Rida per la ricostruzione dei valori mancanti, usando il maggior numero possibile di variabili di matching, in modo da garantire una selezione molto fine di un donatore vicino all'unità affetta da errore. Ancora una volta la scelta della ripartizione come variabile di strato è fondamentale dato il suo legame con i livelli di reddito. Altra variabile che appare potenzialmente interessante come variabile di matching è il numero di percettori all'interno della famiglia; nonostante la variabile venga direttamente rilevata nel questionario, essa appare scarsamente affidabile e si è pertanto costruita una nuova variabile sulla base della condizione professionale dichiarata per ciascun componente (si considerano percettori coloro che si dichiarano occupati, inabili, ritirati dal lavoro ed in altra condizione).

1.4.7.4.1 *Criteri di scelta delle variabili di matching e di strato*⁵

Come già accennato, la sola variabile di strato utilizzata è la ripartizione geografica; le variabili di matching sono invece state definite rispetto a ciascuna variabile oggetto di correzione, sulla base dell'associazione che queste presentavano con la variabile di interesse. È stata anche definita una tipologia familiare molto disaggregata che tiene conto del numero di componenti per fascia d'età presenti all'interno della famiglia. Un esempio di scheda parametro è quella relativa al canone mensile di affitto, dove alla consueta variabile di strato costituita dalla ripartizione geografica, si hanno la classe di reddito, la posizione professionale della persona di riferimento, il tipo di abitazione, il tipo di fabbricato, la superficie ed il numero di percettori come variabili di matching per la selezione del donatore.

Per il resto delle voci e per ogni record errato il sistema Rida provvede a sostituire ciascun valore considerato errato con il valore per la stessa variabile del record del donatore prescelto. La scelta del donatore risulta fondamentale per il processo di correzione, ed è necessario definire:

- la funzione di distanza da utilizzare
- l'insieme di variabili da inserire nella funzione di distanza (variabili di matching)
- il criterio di ponderazione con cui pesare il contributo di ciascuna variabile alla funzione di distanza
- l'insieme delle variabili rispetto alle quali i possibili donatori devono avere le stesse caratteristiche dell'unità affetta da errore (variabili di strato)

Tutto ciò viene definito tramite opportune schede parametro. La definizione della scheda parametro ha richiesto un'analisi delle relazioni esistenti tra le variabili da correggere e le variabili potenzialmente utilizzabili come variabili di matching. Teoricamente è possibile definire una scheda parametro per ciascuna variabile da correggere, spesso però gruppi di variabili possono essere corrette tramite una stessa scheda parametro.

⁵ Lucia Fadioni ha curato le elaborazioni sui dati di base necessari allo sviluppo dell'analisi

1.4.7.4.2 Confronto tra le variabili quantitative prima e dopo l'applicazione di RIDA⁶

Il confronto tra le distribuzioni di ciascuna variabile prima e dopo la correzione tramite RIDA

$$I_1 = \frac{\sum_{i=1}^n (O_i - I_i)}{\sum_{i=1}^n I_i}$$

è fondamentale per capire l'impatto del processo sulla distribuzione della singola variabile. Si sono pertanto considerati i seguenti indici:

dove N è il numero delle osservazioni, O è il valore del dato originario, I quello ottenuto dal processo di correzione (O=I se il dato non è stato corretto).

$$I_2 = \frac{\sum_{i=1}^n (O_i - I_i)^2}{\sum_{i=1}^n I_i^2}$$

L'indice I₁ fornisce una misura delle variazioni avvenute ed indica anche se queste hanno globalmente sbilanciato la situazione iniziale e in quale senso, L'indice I₂ ha la stessa struttura di I₁ ma, essendo quadratico, non fornisce indicazioni sulla direzione della correzione, mentre è più sensibile alle perturbazioni dei dati lontani dalla media.

$$I_3 = \frac{1}{n} \sum_{i=1}^n |O_i - I_i|$$

L'ultimo indicatore utilizzato è rappresentato dalla variazione media avvenuta nella singola unità in seguito alla correzione. Di seguito si riportano le variazioni maggiormente significative per alcune delle voci oggetto di indagine, tenuto conto del fatto che sulla spesa media totale, la variazione percentuale è dello 0,3% e che circa il 15% dei records presentano almeno un valore da correggere.

Variabili	Pre	Post.	Variazione %
Lardo e strutto	456	355	-22.1
Rifacimento impianti	14.855	12634	-14.95
Acqua 2° abitazione	690	534	-22.6
Rifacimenti interni 2° ab.	4.695	3398	-27.64
Altri lavori 2° abitazione	2400	1521	-36.64
Scaldabagno	848	756	-10.90
Quadri, soprammobili	5519	4479	-18.85
Lampade	7935	6151	-22.48
Riparazioni piccoli elettr.	4218	3682	-12.72
Altre spese per la casa	2483	2014	-18.89
Dentista	51648	44388	-14.06
Olio e lubrificanti	9638	8636	-10.40
Franco bolli	4745	3966	-16.42
Abbonamenti sportivi	3049	2456	-19.43
Pellicole foto	15904	13761	-13.48
Argenteria	16103	13615	-15.45
Altre spese personali	2436	2177	-10.62
Quaderni	11070	8914	-19.48
Spese professionali	13719	11445	-16.58

Tab.1 Variabili quantitative con una variazione pre-post correzione negativa superiore al 10%-(I trimestre 1997)

⁶ Lucia Fadioni ha curato le elaborazioni sui dati di base necessari allo sviluppo dell'analisi

	Pre-rida	Post-rida	Variazione %
Proprietà 2° abitazione	37407	43983	+17,58
Tappeti, moquette	4718	5239	+11,03
Riparazioni mobili	5895	6493	+10,14
Assicurazione natanti	825	1297	+57,27
Taxi	3102	3571	+15,15
Assicurazione 2° abitazione	592	1487	+151,31
Spesa per traslochi	697	1610	+130,95
Altre spese	3834	4441	+15,83
Altre spese eccezionali	2623	3708	+41,35

Tab. 2 Variabili quantitative che presentano una variazione pre-post correzione positiva superiore al 10%- (I° trimestre 1997)

1.4.8 Appendice A: regole di compatibilità per il controllo delle variabili qualitative relative ai componenti della famiglia

- 1 le variabili relative ai componenti dichiarati non devono avere valori mancanti: il numero dei componenti della famiglia è una variabile sempre presente e non imputabile poiché è pari al numero di righe compilate nella tabella a pag. 5 del modello "Riepilogo delle spese familiari"; il sistema provvede ad imputare il valore per le variabili ove si verificano mancate risposte parziali;
- 2 esistenza della persona di riferimento: è espressa dalla regola di fuori dominio per la variabile relazione con la persona di riferimento del componente 1
- 3 unicità della persona di riferimento: è espressa dalle regole di fuori dominio per le variabili relazione con la persona di riferimento dei componenti 2-12.
- 4 la persona di riferimento può avere un solo coniuge;
- 5 la persona di riferimento e il suo coniuge devono avere sesso diverso;
- 6 il coniuge della persona di riferimento deve essere coniugato se la persona di riferimento è coniugata;
- 7 il coniuge della persona di riferimento non deve essere coniugato se la persona di riferimento non è coniugata;
- 8 l'ascendente della persona di riferimento deve avere almeno 35 anni;
- 9 il componente coniuge/convivente della persona di riferimento deve avere almeno 16 anni;
- 10 il componente non celibe deve avere almeno 16 anni;
- 11 il componente divorziato deve avere almeno 19 anni;
- 12 il componente con il dottorato deve avere almeno 23 anni;
- 13 il componente con la laurea deve avere almeno 21 anni;
- 14 il componente con il diploma universitario o laurea breve deve avere almeno 20 anni;
- 15 il componente con il diploma che permette l'accesso all'università deve avere almeno 17 anni;
- 16 il componente con la qualifica o licenza che non permette l'accesso all'università deve avere almeno 15 anni;
- 17 il componente con la licenza media deve avere almeno 13 anni;
- 18 il componente con la licenza elementare deve avere almeno 10 anni;
- 19 per il componente con età ≤14 anni non deve essere rilevata la condizione professionale;
- 20 per il componente con età ≤14 anni non deve essere rilevata la branca di attività economica prevalente;
- 21 per il componente con età ≤14 anni non deve essere rilevata la posizione nella professione;

- 22 per il componente con età ≥ 15 anni non deve essere mancante la condizione professionale;
- 23 componente con condizione professionale = casalinga non deve essere di sesso maschile
- 24 componente con condizione professionale = studente non deve avere età ≥ 35 anni
- 25 il componente con condizione professionale = pensionato non deve avere età < 40 anni;
- 26 il componente con condizione professionale = in servizio di leva o servizio civile sostitutivo non deve essere di sesso femminile;
- 27 il componente con condizione professionale = in servizio di leva o servizio civile sostitutivo non deve avere età < 18 anni o età > 30 anni;
- 28 il componente con età ≥ 65 anni non deve avere condizione professionale disoccupato o in cerca di prima occupazione;
- 29 per il componente con condizione professionale \neq occupato/disoccupato non deve essere rilevata la branca di attività economica prevalente;
- 30 per il componente con condizione professionale \neq occupato non deve essere rilevata la posizione nella professione;
- 31 per il componente con condizione professionale = occupato/disoccupato non deve essere mancante la branca di attività economica prevalente;
- 32 per il componente con condizione professionale = occupato non deve essere mancante la posizione nella professione.

1.4.9 Appendice B: regole di compatibilità per il controllo delle variabili qualitative relative alle caratteristiche dell'abitazione, al possesso di beni durevoli e alle abitudini di spesa.

SCIA2

Regole sull'abitazione	
Abitazione non impropria	➔ Devono essere indicati i servizi dell'abitazione
Abitazione impropria	➔ Non ci devono essere informazioni sull'abitazione
Acqua calda si	➔ Va indicato il sistema di riscaldamento dell'acqua
Acqua calda no	➔ Non va indicato il sistema di riscaldamento dell'acqua
Riscaldamento dell'abitazione si	➔ Tipo riscaldamento e combustibile si
Riscaldamento dell'abitazione no	➔ Tipo riscaldamento e combustibile no
Combustibile si	➔ Tipo riscaldamento si
Abitazione in affitto	➔ Ci deve essere l'indicazione sul proprietario e sul tipo di contratto
Abitazione non in affitto	➔ Non ci deve essere l'indicazione sul proprietario e sul tipo di contratto
Elettrodomestici ed energia elettrica	
Si imputano gli elettrodomestici solo in presenza di energia elettrica	
Possesso dell'automobile	
Si possiede l'automobile	➔ Il numero di auto posseduto deve essere indicato
Modalità di pagamento	
Deve essere indicata almeno una modalità di pagamento	
Sono ammesse al massimo due modalità di pagamento	

FISSITÀ E VARIABILI DI STRATIFICAZIONE

Nell'effettuare le correzioni e le imputazioni si tiene conto di diversi *gradi di fissità* tra le variabili. In particolare:

Il tipo dell'abitazione ha una fissità superiore a quella di tutte le altre variabili per evitare di creare abitazioni improprie. Viene inoltre utilizzata come *variabile di strato* una nuova variabile ottenuta come combinazione della tipologia familiare (5 tipologie) e della residenza geografica (2 ripartizioni).

1.4.10 Individuazione e trattamento di record anomali

(versione preliminare)

La necessità di individuare e successivamente trattare i *records* anomali nasce nell'ambito di una ricerca di coerenza anche in termini di micro-dato.

I record anomali non vengono individuati e trattati come record errati, ma come record che, pur presentando valori e comportamenti di spesa "lontani" o dissimili dal collettivo, sono comunque esatti.

La presenza di tali record influenza la stima dal punto di vista sia della distorsione sia della varianza, ma la pratica, diffusa in passato, che comportava la loro eliminazione è stata abbandonata in quanto essi apportano comunque informazioni utili.

Prima del trattamento di un record anomalo, il problema, peraltro non trascurabile, è la sua individuazione.

Nel caso specifico della nuova indagine sui consumi delle famiglie, è stata utilizzata una *cluster analysis*, sulla base dei principali 19 capitoli di spesa (Pane e cereali, Carne, Pesce, Latte, formaggi e uova, Oli e grassi, Patate, frutta e ortaggi, Zucchero, caffè e cacao, Bevande, Tabacco, Abbigliamento e calzature, Abitazione, Combustibili ed energia elettrica, Arredamenti, elettrodomestici e servizi per la casa, Servizi sanitari e spese per la salute, Trasporti, Comunicazioni, Istruzione, Tempo libero e cultura, altri beni e servizi), della zona di residenza (5 ripartizione) e del numero di componenti.

I capitoli di spesa sono stati utilizzati in termini di rapporto rispetto alla spesa totale sostenuta e sembrano soddisfacentemente caratterizzare i comportamenti di consumo familiare.

Attraverso la individuazione di cluster disgiunti tramite l'algoritmo leader di Hartigan e quello delle K-medie di Mac Queen, si ottiene una partizione efficace anche in presenza di un campione particolarmente ampio.

Il metodo di inizializzazione utilizzato (*Anderberg*, centroidi più vicini), che determina la selezione dei *seeds* iniziali (si minimizza la somma delle distanze quadratiche tra le medie dei *cluster*), risulta inoltre particolarmente adatto per individuare osservazioni anomale nei cluster di numerosità ridotta.

Le osservazioni individuate come anomale per l'indagine sui Consumi delle Famiglie sono pertanto quelle che presentano frequenze particolarmente ridotte (da 1 a 4 osservazioni) e vengono specificatamente trattati in fase di stima.

Una volta individuati, il trattamento di record anomali può essere effettuato tramite (Rinaldelli, 1995):

- 1) modifica dei valori assunti dalle unità anomale (winsorization)
- 2) riduzione dei pesi associati alle unità anomale
- 3) l'uso di stimatori robusti.

La prima soluzione si basa essenzialmente sulla modifica dei valori delle specifiche variabili che rendono anomala l'unità osservata, riportandoli a livelli di "normalità". Il vantaggio di tali stimatori è che permettono di trattare separatamente le singole variabili rilevate e risultano quindi particolarmente efficienti quando un record è anomalo solo rispetto ad alcune di esse. L'approccio di fatto risulta di limitata applicazione a causa della difficoltà di utilizzazione nei disegni campionari complessi (come ad esempio quello dell'indagine sui consumi delle famiglie).

Il trattamento tramite stimatori robusti è quello relativamente più recente e consiste nell'uso di metodi di stima poco sensibili alle osservazioni anomale; il notevole vantaggio che questi presentano è quello di non necessitare dell'individuazione a priori dei record anomali stessi.

Tali metodi hanno però trovato scarsa applicazione in quanto generalmente meno efficienti degli altri stimatori, sono distorti, non sono additivi e soprattutto sono difficilmente implementabili.

In particolare la loro utilizzazione risulta improponibile in un contesto di diffusione del micro-dato all'esterno dell'istituto nazionale di statistica.

Il metodo utilizzato è pertanto quello relativo alla riduzione dei pesi attribuiti alle osservazioni anomale, che presenta il vantaggio di effettuare la modifica solo in termini di

peso e non in termini di valore della variabili rilevate e che permette, successivamente, l'uso di stimatori classici.

Per la determinazione dei pesi, si è seguito l'approccio proposto da Hidioglou e Srinath (Journal of the American Statistical Population); che mette a confronto quattro diversi possibili stimatori.

Sia

Y_1, Y_2, \dots, Y_N il carattere osservato nella popolazione di interesse di N unità

T= numero di unità anomale nella popolazione

n= campione estratto

t= numero di unità anomale nel campione

Il primo stimatore proposto assegna peso unitario alle osservazioni anomale e un peso pari all'inverso del rapporto tra numero di osservazioni anomale nel campione e nella popolazione (supponendo che queste coincidano con quelle osservate):

$$\hat{Y}_1 = \sum_{i=1}^t y_i + \frac{N-t}{n-t} \sum_{i=t+1}^n y_i$$

Il secondo stimatore assegna un peso alle unità anomale che dipende dal numero e dal valore assunto dalle unità anomale nel campione.

$$\hat{Y}_2 = \frac{N}{n} y_i - \frac{Nt}{2n} \left(\frac{n-t}{n} \right) \cdot \left(\sum_{i=1}^t \frac{y_i}{t} - \sum_{i=t+1}^n \frac{y_i}{n-t} \right)$$

Il terzo stimatore

$$\hat{Y}_3 = r \sum_{i=1}^t y_i + \frac{N-rt}{n-t} \sum_{i=t+1}^n y_i \quad (1 \leq t \leq n)$$

non è altro che una generalizzazione del primo stimatore in quanto assegna un peso ottimale r alle osservazioni anomale, scelto in modo da minimizzare l'errore quadratico medio (MSE).

Infine il quarto risulta definito come

$$\hat{Y}_4 = \frac{T}{t} \sum_{i=1}^t y_i - \frac{N-T}{n-t} \sum_{i=t+1}^n y_i$$

s

Il primo stimatore proposto (\hat{Y}_1) risulta vantaggioso se la frazione di campionamento e il numero di valori anomali osservati sono in numero ridotto; il secondo (\hat{Y}_2) è invece utilizzabile quando sia f (n/N) sia T risultano abbastanza elevati ed infine \hat{Y}_4 sarebbe lo stimatore da usare se T fosse noto a priori.

$$r = \frac{NT(N-T-n+t)C_v^2 + (n-t)(N-T)T^2(\delta-1)^2}{(n-t)(N-T)[(T-t)C_\mu^2\delta^2 + tT(\delta-1)^2] + tT(N-T-n+t)C_v^2}$$

Infine, data la realizzazione campionaria

dove:

δ rappresenta il rapporto tra la media delle unità anomale e quella delle non anomale

C_μ coefficiente di variazione delle unità anomale

C_v coefficiente di variazione delle unità non anomale,

è il miglior stimatore in termini di efficienza.

Pertanto, nell'ambito dell'indagine sui consumi delle famiglie, si sono determinati i valori ottimali per r e lo stimatore \hat{Y}_3 è quello utilizzato per le stime finali.

2 I possibili effetti sulle stime derivanti dal trattamento degli errori non campionari

2.1 Gli errori di risposta

Per errori di risposta nell'indagine sui consumi intendiamo quegli errori generati dai seguenti comportamenti dei rispondenti:

- vengono omesse spese effettuate nel periodo osservato;
- le spese effettuate nel periodo osservato vengono indicate, ma con un valore inferiore al reale;
- vengono indicate spese in realtà non avvenute nel periodo osservato;
- vengono indicate spese avvenute nel periodo osservato, ma con un valore superiore al reale.

È evidente che il primo ed il secondo comportamento danno luogo a sottostime nei livelli dei consumi, mentre l'effetto opposto è generato dal terzo e quarto comportamento.

Le cause principali che possono determinare tali comportamenti sono:

- (i) quelle legate alle modalità dell'intervista riepilogativa effettuata a fine periodo, e che possiamo sintetizzare nella naturale tendenza dei rispondenti a non ricordare con precisione le spese effettuate;
- (ii) quelle legate alle modalità di compilazione del diario delle spese fra le quali ad esempio la difficoltà ad orientarsi in un questionario con molte voci fra le quali cercare quella corretta e con pagine uguali per ogni giorno di rilevazione.

Nel primo caso (riepilogo a fine periodo), se è l'evento stesso della spesa ad essere dimenticato, ciò automaticamente determina sottostime nel livello dei consumi. Nell'opposta direzione, possono essere erroneamente associati al periodo osservato eventi di spesa in realtà avvenuti prima di tale periodo (effetto *telescoping*), il che ovviamente determina sovrastime.

Se al contrario i problemi riguardano l'entità della spesa, l'effetto memoria può determinare sia sovrastime che sottostime. L'effetto è tanto maggiore, quanto più ampio è il periodo osservato (nell'indagine vengono presi in considerazione periodi di dodici mesi, tre mesi ed un mese); è tanto minore, quanto più significativa è l'entità delle spese effettuate nel periodo in esame. Ciò spiega perché mediante riepilogo vengono rilevate spese a frequenza non elevata e di entità significativa.

2.1.1 Gli errori di risposta nella compilazione del diario⁷

Per converso, mediante compilazione di un diario giornaliero vengono rilevate le spese caratterizzate da frequenza medio-alta, e non elevate. In questo caso, non sussistono, o sono limitati, problemi legati alla memoria, in special modo se il diario è compilato correttamente giorno per giorno. Possono però darsi altre cause generatrici di errori.

In letteratura⁸, si osserva come nella quasi totalità delle indagini sui consumi condotti nei diversi paesi l'andamento delle frequenze degli eventi di spesa registrati nei diversi giorni sottoposti ad osservazione (da un minimo di una settimana, a 14 giorni, a periodi superiori), rilevi di norma:

- un picco nel primo giorno di rilevazione;
- un andamento decrescente negli ultimi giorni di rilevazione.

Le possibili cause che spiegano il primo fenomeno sono così indicate:

- (i) la negligenza del rispondente a girare le pagine del libretto;
- (ii) un ridotto interesse dopo il primo giorno;
- (iii) il cosiddetto *telescoping esterno*, fenomeno per il quale spese effettuate immediatamente prima del periodo osservato vengono erroneamente ad esso associate;
- (iv) un effetto novità indotto dalla rilevazione che porta i rispondenti ad effettuare effettivamente più spese, generalmente nel primo giorno del diario;
- (v) una compilazione del diario effettuata al momento del ritorno dell'intervistatore, che porta ad associare le spese ricordate nel primo giorno utile.

Di tali cause, la terza e la quarta sono quelle determinanti sovrastime nei livelli delle spese, la prima e la quinta sono «neutrali», mentre la seconda determina sottostime.

L'andamento decrescente negli ultimi giorni del diario è universalmente spiegato con il cosiddetto «effetto stanchezza», con l'attitudine cioè dei rispondenti a diminuire la scrupolosità di registrazione delle spese col passare dei giorni. L'effetto di tale comportamento è in una generale sottostima delle spese relative alle voci contenute nel diario.

Questo elemento era presente nella vecchia indagine, come si evince da uno studio⁹ compiuto appositamente. In studi di tale tipo, il primo problema da superare è quello relativo allo

⁷ Giorgio Della Rocca e Lucia Fadoni hanno curato le elaborazioni sui dati di base necessari allo sviluppo dell'analisi

⁸ Cfr. «Expenditure diary surveys and their associated errors» di A.Silberstein e S.Scott, in *Measurement Errors in Survey*, 1991 John Wiley and Sons

⁹ Cfr. «Il pattern di registrazione nell'indagine sui consumi», di C.Brasili e M.R.Ferrante, CON PRI Rapporto di ricerca n.15, Febbraio 1995 Università di Bologna

scorporo dell'effetto del giorno della settimana: ad esempio, se il primo giorno è un sabato è chiaro che questo ha un effetto di per sé sul picco di frequenze che vi si registra. Nella vecchia indagine, per ogni mese la sequenza dei giorni della settimana nel periodo di osservazione, data la cadenza decadale, è sempre la stessa per tutte le famiglie che appartengono alla stessa decade. Tenendo conto di questo fatto, è possibile mettere a confronto il primo giorno con l'ottavo, il secondo con il nono ed il terzo con il decimo (nel caso dell'ultima decade, anche il quarto con l'undicesimo). Quello che emerge è che «...in quasi tutti i casi il dato medio è maggiore nella prima parte della decade... a conferma dell'ipotesi di una diminuzione delle registrazioni all'aumentare del numero dei giorni di rilevazione, e quindi della presenza di un effetto fatica». Dallo studio effettuato non si evince altrettanto chiaramente la presenza sistematica di un picco nel primo giorno.

Situazione quasi speculare nella nuova indagine: nella quasi generalità delle 109 voci contenute nel libretto settimanale, si assiste ad un andamento delle frequenze degli eventi di spesa caratterizzato da un picco normalmente molto significativo nel primo giorno, e da un andamento quasi costante negli altri sei giorni, con leggere gobbe negative nel secondo giorno, e positive nel quarto e quinto.

Sempre parlando in termini generali, non è più presente un effetto «stanchezza», come se lo scopo che ci si era prefissi riducendo il numero dei giorni osservati da dieci a sette sia stato effettivamente raggiunto.

Resta da spiegare il picco nel primo giorno.

Se le informazioni indipendenti a nostra disposizione possono essere considerate affidabili dobbiamo escludere la seconda spiegazione tra quelle possibili (un ridotto interesse dopo il primo giorno): in tal caso, la frequenza riscontrata nel primo giorno dovrebbe essere quella «vera» per tutti i giorni della settimana, il che dovrebbe portare a supporre la presenza di una fortissima sottostima dei livelli di spesa, mentre in realtà ci troviamo a dover spiegare una situazione opposta.

Le spiegazioni 1 e 5 sono entrambe possibili, così come quelli 3 e 4. Questi ultimi, come è ovvio, sono fattori che generano sovrastime, mentre i primi sono neutrali. Quali e quanto siano i reali elementi responsabili del picco nel primo giorno è cosa che è possibile stabilire solo con un'apposita indagine di controllo, in cui le famiglie siano particolarmente motivate ad una compilazione del diario la più possibile fedele, grazie all'intervento di rilevatori professionali che, ad esempio, oltre a spiegare dettagliatamente le modalità corrette di registrazione delle spese, intervengano durante la settimana di rilevazione, con visite e/o telefonate. Se dalle interviste così effettuate si dovesse rilevare la scomparsa (o la significativa attenuazione) del picco nel primo giorno, e contemporaneamente una diminuzione nel livello delle stime, potremmo concludere che le cause determinanti tale picco sono, in tutto in parte, quelle che generano sovrastime.

Qui, a titolo di valutazione dell'entità massima di sovrastima imputabile ai fattori che abbiamo descritto, riportiamo l'abbattimento che si otterrebbe nelle stime rinunciando a considerare quanto registrato nel primo giorno del diario. Nella tabella 1, per ogni capitolo di spesa riportiamo le medie campionarie della spesa mensile calcolate considerando tutti i giorni della settimana e moltiplicando l'importo delle relative spese per (30/7), e, in parallelo, le medie campionarie calcolate considerando l'importo delle spese effettuate tra il secondo ed il settimo giorno, moltiplicato per (30/6). In media, il decremento che si ottiene è nell'ordine del 10%.

AGGREGATO	SETTE GIORNI (SPESA TOTALE CAMPIONE) (1)	SEI GIORNI (SPESA TOTALE CAMPIONE) (2)	Abbattimento % spesa totale ((2-1)/2)*100
Pane e cereali	162.142.857	149.896.215	-7,6
Gelati, dolci ecc	127.152.180	113.739.097	-10,5
Carni e salumi	287.420.571	249.875.547	-13,1
Pesce	93.225.646	92.524.756	-0,8
Oli e grassi	62.005.506	51.124.537	-17,5
Latte formaggi e uova	165.762.366	150.405.154	-9,3
Legumi e ortaggi	101.742.332	93.341.072	-8,3
Frutta	99.486.257	89.126.827	-10,4
Bevande	105.491.418	94.217.002	-10,7
Pasti e con. fuori casa	176.313.260	171.143.108	-2,9
Tabacchi	63.873.577	59.975.787	-6,1
Art. merceria	14.976.650	13.597.575	-9,2
Ogg. per la casa	98.816.569	87.292.776	-11,7
Igiene personale	169.271.227	151.357.019	-10,6
Giornali, libri, canc.ria	69.232.260	64.127.306	-7,4
Telefoni e spese postali	24.917.030	22.180.673	-11,0
Trasporti	362.165.617	315.847.111	-12,8
Salute	133.697.188	115.608.171	-13,5
Spese varie	85.468.387	77.497.630	-9,3
Giocattoli, giochi e spettacoli	83.491.685	72.137.461	-13,6
Manutenzione e riparazione	120.317.984	104.209.486	-13,4
TOTALE	2.606.970.567	2.339.224.310	-10,3

Tabella 1 - Spesa totale campionaria calcolata con e senza il primo giorno

2.1.2 Gli errori di risposta dovuti ad ambiguità nei quesiti

E' doveroso menzionare un altro elemento che può pesantemente determinare distorsioni negative o positive: la mancata comprensione della definizione della voce.

Consideriamo a titolo di esempio due fra le voci che sembrano essere affette da una notevole sovrastima, stando almeno a dati esterni con cui è stato possibile avere un riscontro: *benzina per auto e moto, e medicinali*.

Per quanto riguarda la voce «Benzina per auto e moto», sapendo che l'intento dell'indagine è quello di stimare i consumi *finali* delle famiglie, dobbiamo escludere quelle spese per benzina effettuate nell'ambito di un'attività lavorativa (si pensi all'uso delle automobili, anche furgonate, da parte di distributori, rappresentanti, medici ecc.). Nel wording della domanda tale avvertenza non compare, né è presente nel resto del questionario o nelle istruzioni per la rilevazione. E' lecito supporre che almeno parte della sospetta sovrastima della spesa totale per la benzina sia dovuta a questo elemento.

Se consideriamo invece la domanda relativa alla voce «Medicinali» il wording è il seguente: «Medicinali (*costo totale o ticket*)». Può accadere che il rispondente intenda il quesito nel senso che, qualora abbia acquistato dei farmaci con prescrizione SSN, sia indifferente indicare

il costo pieno o il ticket: ad esempio, se non ricorda l'importo, può prendere la confezione ed indicare il prezzo stampato su di essa. E' chiaro che un comportamento di questo tipo è responsabile di distorsioni positive nella stima della spesa per medicinali.

Anche in questo caso, entrambe le ipotesi sono da comprovare mediante un'indagine di controllo che verifichi l'impatto sulla stima modificando adeguatamente l'impostazione del quesito e sensibilizzando i rispondenti sulla corretta comprensione delle voci.

2.2 Gli errori di registrazione: valutazione mediante registrazione di controllo

Al fine di quantificare il peso degli errori di registrazione sulle stime, si è provveduto ad effettuare una doppia registrazione di 500 modelli relativi alla rilevazione di gennaio 1997; per poter inoltre valutare l'impatto della registrazione controllata (effettuata con il sistema Blaise), la registrazione di controllo è stata effettuata in modalità "tradizionale", ossia utilizzando il sistema di acquisizione che normalmente viene adottato dal centro di registrazione interna ISTAT, che non prevede la segnalazione delle diverse tipologie di errore, tipica del CADI. I casi di divergenza, quindi, sono stati sottoposti a riconciliazione: si sono in pratica analizzati i dati riportati sui modelli cartacei per verificare se coincidessero con quelli registrati in Blaise oppure con quelli registrati senza controlli.

Le distribuzioni delle variabili divergenti secondo le due modalità di registrazione (che da ora in poi sono denominate rispettivamente "Reg. Blaise" e "Reg. Tradizionale") e secondo il tipo sono riportate nelle tabelle 1 e 2.

Si precisa che le percentuali sono calcolate su due totali diversi: il totale di variabili registrate tramite la "Reg. Blaise" e quello registrato con la "Reg. Tradizionale"; i due totali divergono perché con la "Reg. Blaise", per alcune variabili (soprattutto qualitative), è stato imputato il valore corretto, sulla base di controlli di coerenza con altri quesiti. Per meglio comprendere il perché della divergenza tra questi due totali, si riporta il caso della regola, considerata nella "Reg. Blaise", in base alla quale, se l'individuo ha meno di 15 anni, non gli vengono richieste la condizione professionale e la posizione nella professione. Qualora i dati del modello cartaceo contraddicano tale regola, la "Reg. Blaise" impone che sia rispettata, quindi, se l'età è minore di 15 anni, viene inibita la registrazione delle altre due variabili; ciò non avviene con la "Reg. Tradizionale".

Altri casi analoghi a questo, sono relativi, per esempio:

- ai controlli di coerenza tra la dichiarazione sull'effettuazione di spese (SI/NO) e la presenza o assenza dell'importo di spesa relativo;
- all'importo corretto del canone RAI.

Var. divergenti	Totale var. Reg. Blaise	Percentuale sul Totale var. Reg. Blaise	Totale var. Reg. Tradiz	Percentuale su Totale var. Reg. Tradiz
2.747	119.340	2,30%	118.870	2,31%

Tabella 1 - Variabili divergenti secondo le modalità di registrazione

Var. divergenti qualitative	Percentuale sul totale delle divergenti	Var. divergenti quantitative	Percentuale sul totale delle divergenti
1.174	42,74	1.573	57,26

Tabella 2 - Variabili divergenti secondo il tipo (qualitativo/quantitativo)

Le principali cause di divergenza riscontrate in fase di riconciliazione sono le seguenti:

- errori di registrazione sui singoli caratteri;
- errori dovuti a slineamenti;
- caratteri dubbi per problemi di interpretazione di quanto scritto sul modello (calligrafia non chiara);
- anomalie dovute all'aspetto tipografico dei modelli. Ci si riferisce in particolare a due casi per i quali il valore registrato non è chiaramente associabile ad una singola variabile (nel settore "Cura e igiene personale", i codici 9101 e 9102 e nel settore "Trasporti" i codici 6307 e 6211);
- imputazione del valore corretto nel caso della "Reg. Blaise" (soprattutto per alcune variabili qualitative).

Nelle tabelle 3 e 4 sono riportati i risultati dell'analisi effettuata, distinti per variabili qualitative e variabili quantitative.

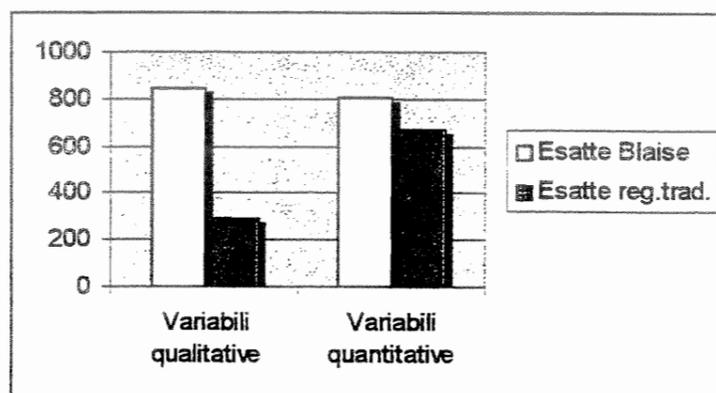
Per ciascuna tipologia di variabile, si definiscono "corretti" i valori coincidenti con quelli scritti sul modello. Vengono inoltre assimilati ai valori corretti quelli derivanti dall'imputazione del valore corretto da parte della Reg. Blaise, seppur non coincidenti con quanto scritto sul modello.

Totale variabili qualitative divergenti	Esatte Reg. Blaise		Esatte Reg. Tradiz	Dubbie
	Totale	di cui imputazione del valore corretto		
1.174	848	616	291	35
	72,23%		24,79%	2,98%

Tabella 3 - Risultati delle riconciliazioni per le variabili qualitative

Totale variabili quantitative divergenti	Esatte Reg. Blaise	Esatte Reg. Tradiz	Dubbie
1.573	809	667	97
	51,43%	42,40%	6,17%

Tabella 4 - Risultati delle riconciliazioni per le variabili quantitative



Relativamente alle variabili quantitative, si riporta infine un'analisi delle divergenze dovute a slineamenti.

Totale variabili quantitative divergenti	di cui dovute a slineamenti		Esatte Reg.Blaise	Esatte Reg.Tradiz
1.573	252	16.02%	123 48,81%	129 51,19%

Tabella 5 – Divergenze dovute a slineamenti

Considerazioni conclusive

Da quanto emerso dallo studio descritto, deve essere tratta una prima considerazione generale: il peso derivante dagli errori di registrazione è di per sé notevolmente basso (2.3%).

Relativamente invece all'impatto della registrazione controllata, rispetto a quella tradizionale, emerge quanto segue:

- *Il grande vantaggio della registrazione controllata* (mediante la quale è possibile inserire controlli stringenti sulle possibili modalità di risposta e sulla coerenza rispetto ad altri quesiti connessi) è evidente per le variabili qualitative: basti vedere l'elevato numero di divergenze tra le due modalità di registrazione dovuto all'imputazione dei valori corretti.
- *Relativamente alle variabili quantitative*, invece, considerando la numerosità di divergenze dovute a casi dubbi, le due modalità di registrazione non apportano significative differenze in termini di qualità. Ciò è imputabile al fatto che i valori minimi e massimi di spesa previsti dal piano di registrazione non risultano essere molto stringenti e non possono quindi contribuire significativamente alla prevenzione sull'errore di registrazione (che è spesso limitato ad una sola cifra).
- Ne consegue che la registrazione controllata non ha effetti inibitori sui casi di slineamento, che si verificano più spesso nell'ambito delle elencazioni di valori di spesa. Questa affermazione infatti, può essere smentita soltanto nei casi nei quali lo slineamento non avvenga soltanto tra una riga e l'altra, ma sia più ampio e si finisca quindi ad un certo

punto per registrare l'importo di una variabile in un campo con caratteristiche assolutamente diverse da quelle previste. Questo si è verificato, nella nostra analisi, nel caso di uno slineamento di pagina in un Libretto che ha comportato l'inserimento di 64 variabili errate; tale slineamento, infatti, è avvenuto nel corso della registrazione tradizionale.

2.3 Impatto della procedura di controllo e correzione

L'effetto prodotto dalla procedura di controllo e correzione sulla spesa per consumi è stato misurato ponendo a confronto la spesa media mensile per famiglia calcolata sui valori rilevati (grezzi) con la corrispondente spesa media mensile per famiglia calcolata sui valori validati (puliti). Per agevolare il confronto è stata calcolata la variazione relativa percentuale della spesa media ottenuta dai valori puliti rispetto alla spesa media ottenuta dai valori grezzi. In Tabella 1 sono riportate le variazioni registrate per ciascuno dei 50 gruppi omogenei di voci di spesa utilizzati dalla contabilità nazionale. Le variazioni sono state calcolate sia sui valori provenienti dalle famiglie campionate sia sulle stime prodotte a livello nazionale ovvero sui valori ottenuti moltiplicando la spesa di ciascuna famiglia per il corrispondente peso di riporto all'universo. I dati analizzati si riferiscono al I trimestre 1997.

Dall'esame della Tabella 1 si rileva che la procedura di controllo e correzione ha un impatto leggermente depressivo sulla spesa media mensile totale, si registra infatti una riduzione dello -0.48% per la spesa media campionaria e dello -0.61% per la stima della spesa media. La macro disaggregazione in consumi alimentari e non alimentari evidenzia un abbattimento leggermente superiore per i consumi alimentari rispetto a quelli non alimentari. Una maggiore variabilità dell'effetto prodotto dalla procedura di controllo e correzione è rilevabile analizzando l'entità delle variazioni per ciascun gruppo omogeneo di voci di spesa: insieme alla maggioranza dei gruppi con variazione contenuta si registrano alcuni gruppi (12, 19, 20, 22, 25, 27, 28, 43, 50) con variazioni superiori al 5%. Si rileva a tale riguardo che, ad eccezione delle variazioni relative ai gruppi 43 e 50, tutte le altre variazioni di una certa entità hanno segno negativo ovvero producono un abbattimento del valore grezzo. Inoltre per i gruppi 43 (Barbiere, parrucchiere, istituti di bellezza) e 50 (Altri servizi non classificati altrove), che subiscono un discreto incremento relativo del valore della spesa media per effetto della procedura di controllo e correzione, si ricorda che:

- nel passaggio dalla vecchia alla nuova indagine si registra per il gruppo 43 un incremento della stima del 57% (I trimestre nuova indagine rispetto al I trimestre vecchia indagine);
- nel passaggio dalla vecchia alla nuova indagine si registra per il gruppo 50 una riduzione della stima del 77% (I trimestre nuova indagine rispetto al I trimestre vecchia indagine);
- per entrambi i gruppi il contributo alla spesa complessiva è molto modesto.

In conclusione si può affermare che la procedura di controllo e correzione non può essere considerato un fattore responsabile dell'incremento registrato nelle stime dei consumi nel passaggio dalla vecchia alla nuova indagine.

Gruppo Omogeneo Di Voci Di Spesa	Campione			Universo		
	Grezzi	Puliti	(P-G)/G*100	Grezzi	Puliti	(P-G)/G*100
1 Pane e cereali	155,820	154,702	-0.72	150,758	149,435	-0.88
2 Carne	233,017	229,793	-1.38	223,842	221,378	-1.10
3 Pesce	72,009	72,196	0.26	68,406	68,649	0.36
4 Latte, formaggi, uova	130,105	129,233	-0.67	124,805	123,913	-0.71
5 Olii e grassi	52,030	50,140	-3.63	50,099	48,569	-3.05
6 Frutta e ortaggi	130,881	130,063	-0.62	127,744	126,807	-0.73
7 Patate	10,003	10,005	0.02	9,737	9,741	0.04
8 Zucchero	10,778	10,620	-1.47	10,364	10,167	-1.90
9 Caffè, the, cacao	48,765	48,227	-1.10	47,535	47,040	-1.04
10 Altri generi alimentari	28,021	27,957	-0.23	27,030	26,932	-0.36
11 Bevande analcoliche	41,329	40,624	-1.71	41,121	40,355	-1.86
12 Bevande alcoliche	46,448	43,103	-7.20	46,199	41,889	-9.33
13 Tabacco	49,311	49,058	-0.51	47,151	46,674	-1.01
14 Vestiario (comprese le riparazioni)	221,913	220,782	-0.51	210,688	209,567	-0.53
15 Calzature (comprese le riparazioni)	60,198	59,951	-0.41	56,556	56,536	-0.04
16 Affitti etc.	804,541	827,378	2.84	808,527	832,589	2.98
17 Acqua	18,416	18,234	-0.99	17,636	17,438	-1.12
18 Combustibili, energia elettrica	240,434	240,376	-0.02	235,449	235,438	0.00
19 Mobili, access., tappeti, rivest., ripar.	161,874	156,377	-3.40	152,517	143,597	-5.85
20 Biancheria ed altri art. di arr., ripar.	26,924	22,099	-17.92	26,848	19,758	-26.41
21 Elettrodom. E app. per la casa, ripar.	43,465	44,342	2.02	42,451	42,863	0.97
22 Cristallerie, vasell., Utensil. per casa	46,416	41,647	-10.27	45,480	40,351	-11.28
23 Articoli non durevoli per la casa	60,331	59,274	-1.75	58,791	57,713	-1.83
24 Servizi per la casa (meno serv. Dom)	27,670	27,756	0.31	27,710	27,903	0.70
25 Servizi domestici	23,051	21,453	-6.93	25,655	23,610	-7.97
26 Prodotti medicinali	92,748	91,617	-1.22	92,830	91,133	-1.83
27 Apparecchi e materiale terapeutico	29,306	28,046	-4.30	29,294	27,769	-5.20
28 Servizi medici ed infermieristici	77,410	74,093	-4.28	79,985	75,611	-5.47
29 Servizi sanitari e spese per la salute	25,553	24,767	-3.08	22,528	22,325	-0.90
30 Acquisto di mezzi di trasporto	160,960	158,906	-1.28	157,727	159,433	1.08
31 Pneumatici, camere d'aria, ric., ripar.	91,426	88,356	-3.36	84,247	80,995	-3.86
32 Carburanti e lubrificanti	259,261	254,241	-1.94	246,666	240,012	-2.70
33 Altre spese	94,453	96,408	2.07	92,747	93,726	1.06
34 Acquisto servizi di trasporto	50,979	52,106	2.21	50,179	51,090	1.82
35 Comunicazioni	91,397	89,356	-2.23	89,753	86,554	-3.56
36 App. radio tv., giradischi	16,453	16,460	0.04	15,718	15,725	0.04
37 Appar. foto. strum. music. imbarcazioni	14,021	13,675	-2.47	13,826	13,681	-1.05
38 Altri articoli ricreativi	92,255	90,297	-2.12	89,968	89,140	-0.92
39 Pezzi di ricambio. Access. e riparaz.	18,382	17,565	-4.44	17,112	16,674	-2.56
40 Servizi ricreativi, spettacolo e cult.	106,374	106,512	0.13	105,879	105,796	-0.08
41 Libri, giornali e periodici	69,807	67,952	-2.66	64,531	64,102	-0.66
42 Istruzione	54,595	54,774	0.33	48,477	48,655	0.37
43 Barbieri, parrucch., Istitut. di bellezza	73,839	79,706	7.95	73,102	78,072	6.80
44 Beni per l'igiene della persona	50,800	50,303	-0.98	49,417	49,085	-0.67
45 Bigiotteria e gioielleria	13,236	13,377	1.07	14,105	14,240	0.96
46 Altri articoli personali	11,906	11,950	0.37	12,042	12,085	0.36
47 Articoli da cancelleria e disegno	8,076	8,001	-0.93	7,192	7,103	-1.24
48 Alberghi e pubblici esercizi	186,791	185,095	-0.91	187,456	185,325	-1.14
48 Servizi finanziari non class. altrove	66,629	68,034	2.11	62,478	63,824	2.15
50 Altri servizi non classificati altrove	25,232	26,969	6.88	23,709	25,975	9.56

continua

continua

Spesa totale alimentare	959,204	946,661	-1.31	927,642	914,875	-1.38
Spesa totale non alimentare	3,566,433	3,557,302	-0.26	3,486,424	3,472,168	-0.41
Spesa totale	4,525,638	4,503,963	-0.48	4,414,067	4,387,042	-0.61

Tabella 1 - Spesa media mensile familiare (campionaria e stima) e impatto relativo percentuale della procedura di editing - Nuova indagine I trimestre 1997

2.4 Simulazione dell'adozione di limiti pro-capite e differenziati per tipologia familiare

La nuova procedura di individuazione dei valori anomali dei consumi prevede:

- la definizione, per ciascun bene, di un intervallo di accettazione e quindi l'individuazione dei limiti inferiore e superiore;
- il confronto tra il valore registrato con i suddetti limiti.

I valori che risultano esterni all'intervallo di accettazione sono sostituiti mediante imputazione.

I limiti dell'intervallo di accettazione adottati sono limiti assoluti ovvero limiti utilizzati per validare la spesa della generica famiglia, indipendentemente dal numero dei suoi componenti e della tipologia di appartenenza.

L'utilizzo di limiti non adeguati alla composizione della famiglia potrebbe contribuire a spiegare la distorsione per eccesso dei valori di alcune voci di spesa. Per le famiglie con ridotto numero di componenti i limiti attualmente utilizzati potrebbero essere troppo elevati causando una mancata individuazione di valori anomali e quindi la non modifica degli stessi, con conseguente possibile effetto di sovrastima dei consumi. In altre parole la forma plurimodale della distribuzione della spesa, (causata dalla differente numerosità e composizione delle famiglie) potrebbe rendere non idonea l'adozione di un unico limite superiore (assoluto) per l'individuazione dei valori anomali.

Il presente paragrafo riporta i risultati di alcune procedure di controllo alternative simulate al fine di valutare il contributo alla distorsione per eccesso derivante dall'impiego di una procedura che utilizza un unico limite superiore assoluto per ciascuna voce di spesa e che, quindi, potrebbe essere considerata inadeguata.

2.4.1 Metodo

L'effetto prodotto sui valori totali dei consumi (aggregazione sulle famiglie che compongono il campione) dall'adozione di una procedura che utilizza un unico limite superiore assoluto è stato confrontato con gli effetti ottenuti da due procedure alternative che utilizzano rispettivamente limiti superiori pro-capite e limiti per tipologia familiare.

Nella prima simulazione (d'ora in avanti *95° Pro-capite*) si è provveduto a:

1. selezionare alcuni beni privi di effetto scorta (in presenza di effetto scorta i limiti potrebbero essere associati alle possibilità di scorta piuttosto che alla composizione della famiglia) il cui consumo sia fortemente correlato al numero di componenti la famiglia;
2. per ciascun bene selezionato al passo 1 sono stati calcolati i valori pro-capite dividendo il valore registrato per il numero di componenti la famiglia; si è inteso in tal modo depurare l'influenza della numerosità della famiglia sull'ammontare della spesa sostenuta;

3. per ciascun bene selezionato al passo 1, sono stati definiti i nuovi limiti pro-capite; per semplicità i limiti pro-capite considerati sono stati i 95° percentili delle rispettive distribuzioni;
4. i valori anomali sono stati individuati confrontando il valore pro-capite con i nuovi limiti pro-capite;
5. per ciascun valore anomalo individuato, è stato imputato il nuovo valore. Il nuovo valore imputato è stato il prodotto del valore medio pro-capite per il numero di componenti (non si è interessati in questa simulazione a mantenere inalterata la variabilità dei valori imputati bensì a valutare la variazione dell'ammontare complessivo della spesa).

La seconda simulazione (d'ora in avanti *95° Tipologia*) è stata effettuata al fine di valutare l'effetto prodotto sui valori dei consumi dall'adozione di una procedura di editing che utilizzi limiti assoluti diversificati in base alla tipologia familiare cui la famiglia appartiene. La tipologia familiare considerata è stata quella composta da 12 modalità ed utilizzata dal sottogruppo (Fortini/Pallara) incaricato di analizzare un metodo di stima alternativo che utilizzi oltre all'ammontare della spesa anche l'informazione sulle frequenze di spesa. In questa simulazione i limiti superiori utilizzati sono stati i 95° percentili delle distribuzioni dei valori assoluti condizionate alla tipologia familiare di appartenenza della famiglia mentre i nuovi valori imputati sono stati i corrispondenti valori medi.

Per confrontare gli effetti prodotti sui valori dei consumi dall'utilizzo di limiti pro-capite e per tipologia con l'effetto prodotto dall'utilizzo di limiti assoluti è opportuno procedere in assenza di elementi confondenti (quali ad esempio la metodologia utilizzata per l'individuazione dei limiti).

Il confronto con l'effetto prodotto dall'utilizzo dell'attuale procedura non è stato considerato idoneo poiché in entrambe le simulazioni sopradescritte le variazioni rilevate nei valori dei consumi a seguito dell'utilizzo del 95° percentile come limite superiore potrebbero scontare l'impiego di una procedura troppo stringente (l'esclusione del 5% della distribuzione).

Gli effetti prodotti dall'utilizzo di limiti pro-capite e per tipologia sono stati quindi confrontati con l'effetto prodotto da una corrispondente procedura di editing che utilizza limiti assoluti. A tal fine è stata simulata una ulteriore procedura di editing (d'ora in avanti *95° Assoluti*) che adotta lo stesso criterio definito al precedente punto 3 ovvero utilizza come limiti superiori i 95° percentili delle distribuzioni dei valori assoluti. I nuovi valori imputati sono stati, in questa simulazione, i valori medi.

In tutte le tre simulazioni si è provveduto a modificare esclusivamente i valori anomali elevati considerando, appunto, nuovi limiti superiori per gli intervalli di accettazione. In nessun caso sono stati modificati i limiti inferiori degli intervalli di accettazione, pertanto le imputazioni che aumentano il valore grezzo sono esattamente quelle effettuate con l'attuale procedura di correzione (che utilizza un solo limite inferiore assoluto).

2.4.2 Risultati

I dati utilizzati sono relativi ad un insieme di 5585 famiglie intervistate nel primo trimestre 1997 con la nuova indagine. I valori sottoposti ad editing sono le spese settimanali per le seguenti voci di spesa: medicinali (5105), pane (1101), carne di vitello (1201), latte (1401). Tutti i risultati riportati si riferiscono a stime prodotte a livello nazionale ovvero i valori totali dei consumi sono stati ottenuti aggregando sulle 5585 famiglie i valori settimanali dei consumi di ciascuna famiglia moltiplicati per i corrispondenti pesi di riporto all'universo (coefficienti di espansione).

La Tabella 1 riporta, per ciascuna voce di spesa, i limiti inferiore e superiore utilizzati dall'attuale procedura, i valori totali dei consumi relativi agli insiemi di dati grezzi (Grezzi) e validati (Validati), il valore dell'indice di variazione relativa percentuale I

$$I = \frac{\sum_{i=1}^n (X_{vi} - X_{gi})}{\sum_{i=1}^n X_{gi}} \times 100$$

dove:

- n è il numero di famiglie considerate;
- X_v e X_g sono, rispettivamente, i valori della variabile X presenti nei generici file, *validato* e *grezzo*, messi a confronto.

Le ultime due colonne riportano il numero di imputazioni effettuate distinguendo tra imputazioni che aumentano il valore grezzo e imputazioni che diminuiscono il valore grezzo.

La Tabella 2 riporta per ciascuna voce di spesa, i valori dei 95° percentili delle distribuzioni dei valori assoluti e dei valori pro-capite utilizzati come limiti superiori rispettivamente dalle procedure *95° Assoluti* e *95° Pro-capite*.

La Tabella 3 riporta i valori dei 95° percentili delle distribuzioni dei valori assoluti condizionate per tipologia familiare utilizzati come limiti superiori dalla procedura *95° Tipologia*.

La Tabella 4 riporta i valori totali dei consumi relativi agli insiemi di dati:

- grezzi (Grezzi),
- validati con il 95° percentile della distribuzione dei valori assoluti ovvero utilizzando un limite assoluto (*Validati 95° Assoluti*),
- validati con il 95° percentile della distribuzione dei valori pro-capite ovvero utilizzando un limite pro-capite (*Validati 95° Pro-capite*),
- validati con il 95° percentile della distribuzione dei valori assoluti condizionate alla tipologia di appartenenza della famiglia ovvero utilizzando un limite per tipologia (*Validati 95° Tipologia*).

Le imputazioni sono state effettuate utilizzando la media aritmetica.

Le ultime tre colonne riportano i valori degli indici di variazione relativa percentuale I_{ass} , I_{pc} e I_{tipo} corrispondenti alle tre procedure utilizzate. Se le imputazioni sono effettuate utilizzando il valore mediano al posto della media aritmetica (il valore mediano è meno influenzato dalla presenza dei valori anomali rispetto alla media aritmetica) si ottengono i risultati riportati in Tabella 4 bis.

Dall'esame della Tabella 4 si evince che, come atteso, tutte le tre procedure di editing riducono le spese totali grezze dei consumi per tutti i beni considerati. L'entità della riduzione risulta, per tutte le voci analizzate, più elevata con l'utilizzo di limiti assoluti. Considerazioni analoghe valgono per i risultati riportati in Tabella 4 bis.

La Tabella 5 riporta, per ciascuna voce di spesa, il numero di imputazioni effettuate nell'ambito delle tre procedure di editing simulate, distinguendo tra imputazioni che aumentano il valore grezzo e imputazioni che diminuiscono il valore grezzo.

Dalla Tabella 5 si evince che il numero di imputazioni risulta dello stesso ordine di grandezza per tutte le tre procedure di editing simulate. L'utilizzo di limiti per tipologia modifica i valori considerati in misura complessivamente inferiore (833) rispetto alle modifiche indotte dalle altre due (856 e 871).

In conclusione si può affermare che non vi sono elementi per considerare l'utilizzo di limiti superiori unici e assoluti (per l'individuazione dei valori anomali elevati) come un fattore responsabile della sovrastima dei valori dei consumi.

Voce di spesa	Limite inferiore	Limite superiore	Grezzi	Validati	I	Imputazioni aumentanti il valore grezzo	Imputazioni diminuanti il valore grezzo
Medicinali	3,000	300,000	464,743	455,750	-1.94	22	6
Pane	800	50,000	261,143	257,619	-1.35	4	37
Carne di vitello	2,000	200,000	288,027	274,558	-4.68	10	35
Latte	800	50,000	165,341	162,087	-1.97	6	20

Tabella 1 - Limiti utilizzati dalla procedura Attuale, Spese totali (milioni), Indici di variazione relativa percentuale tra Grezzi e Validati e numero di imputazioni

Voce di spesa	Procedura 95° Assoluti	Procedura 95° Pro-capite
Medicinali	113,000	50,000
Pane	27,070	10,750
Carne di vitello	57,000	22,750
Latte	21,600	8,400

Tabella 2 - Limiti superiori utilizzati dalle procedure 95° Assoluti e 95° Pro-capite

Tipologia familiare	Voci di spesa			
	Farmaci	Pane	Carne di vitello	Latte
Persona sola < 35 anni	132,000	18,000	74,000	14,500
Persona sola 35-64 anni	98,100	16,000	38,000	16,050
Persona sola > 64 anni	93,000	14,600	30,000	13,450
Coppia senza figli P. R. <35 anni	99,000	20,050	53,000	15,400
Coppia senza figli P. R. 35-64 anni	105,000	21,400	56,000	15,500
Coppia senza figli P. R. > 64 anni	124,000	21,500	53,000	16,700
Coppia con 1 figlio	115,900	26,400	58,000	21,000
Coppia con 2 figli	112,000	28,650	60,000	24,300
Coppia con 3 e più figli	110,000	35,700	60,000	29,600
Monogenitore con figli	115,000	27,300	57,000	17,100
Altre famiglie con anziano (>64 anni)	130,000	30,770	52,000	22,500
Altre tipologie	157,100	31,715	41,500	19,900

Tabella 3 - Limiti superiori utilizzati dalla procedura 95° Tipologia

Voce spesa	di	Grezzi	Validati 95° Assoluti	Validati 95° Pro-capite	Validati 95° Tipologia	I _{ass}	I _{pc}	I _{tipo}
Medicinali		464,743	378,604	382,213	380,345	-18.53	-17.76	-18.16
Pane		261,143	241,696	242,346	242,656	-7.45	-7.20	-7.08
Carne vitello		288,027	252,188	253,416	253,209	-12.44	-12.02	-12.09
Latte		165,341	148,740	151,457	149,640	-10.04	-8.40	-9.50

Tabella 4 - Spese totali (milioni) e Indici di variazione relativa percentuale tra valori grezzi e validati con le procedure Assoluti, Pro-capite e Tipologia (imputazioni con la media aritmetica)

Voce spesa	di	Grezzi	Validati 95° Assoluti	Validati 95° Pro-capite	Validati 95° Tipologia	I _{ass}	I _{pc}	I _{tipo}
Medicinali		464,743	369,648	373,806	371,784	-20.46	-19.57	-20.00
Pane		261,143	240,559	241,030	241,444	-7.88	-7.70	-7.54
Carne vitello		288,027	249,065	250,152	250,051	-13.53	-13.15	-13.18
Latte		165,341	147,507	150,463	148,618	-10.79	-9.00	-10.11

Tabella 4 bis - Spese totali (milioni) e Indici di variazione relativa percentuale tra valori grezzi e validati con le procedure Assoluti, Pro-capite e Tipologia (imputazioni con il valore mediano)

Voce di spesa	Imputazioni aumentanti il valore grezzo	Imputazioni diminuanti il valore grezzo		
		Procedura 95° Assoluti	Procedura 95° Pro-capite	Procedura 95° Tipologia
	*			
Medicinali	22	170	169	164
Pane	4	274	291	266
Carne vitello	10	168	174	164
Latte	6	244	237	239

Tabella 5 - Numero di imputazioni effettuate

* il numero di imputazioni che aumentano il valore grezzo è identico per tutte le procedure.