



SISTEMA STATISTICO NAZIONALE
ISTITUTO NAZIONALE DI STATISTICA



L'indagine europea sui redditi e le condizioni di vita delle famiglie (Eu-Silc)



I settori

AMBIENTE E TERRITORIO		<i>Ambiente, territorio, climatologia</i>
POPOLAZIONE		<i>Popolazione, matrimoni, nascite, decessi, flussi migratori</i>
SANITÀ E PREVIDENZA		<i>Sanità, cause di morte, assistenza, previdenza sociale</i>
CULTURA		<i>Istruzione, cultura, elezioni, musei e istituzioni similari</i>
FAMIGLIA E SOCIETÀ		<i>Comportamenti delle famiglie (salute, letture, consumi, etc.)</i>
PUBBLICA AMMINISTRAZIONE		<i>Amministrazioni pubbliche, conti delle amministrazioni locali</i>
GIUSTIZIA E SICUREZZA		<i>Giustizia civile e penale, criminalità</i>
CONTI ECONOMICI		<i>Conti economici nazionali e territoriali</i>
LAVORO		<i>Occupati, disoccupati, conflitti di lavoro, retribuzioni</i>
PREZZI		<i>Indici dei prezzi alla produzione e al consumo</i>
AGRICOLTURA E ZOOTECNIA		<i>Agricoltura, zootecnia, foreste, caccia e pesca</i>
INDUSTRIA E SERVIZI		<i>Industria, costruzioni, commercio, turismo, trasporti e comunicazioni, credito</i>
COMMERCIO ESTERO		<i>Importazioni ed esportazioni per settore e Paese</i>

Alla produzione editoriale collocata nei 13 settori si affiancano le pubblicazioni periodiche dell'Istituto: Annuario statistico italiano, Bollettino mensile di statistica e Compendio statistico italiano. Il Rapporto annuale dell'Istat viene inviato a tutti gli abbonati anche ad un solo settore.



SISTEMA STATISTICO NAZIONALE
ISTITUTO NAZIONALE DI STATISTICA

**L'indagine europea sui redditi e le
condizioni di vita delle famiglie
(Eu-Silc)**

A cura di: Claudio Ceccarelli, Marco Di Marco, Claudia Rinaldelli

Per chiarimenti sul contenuto della pubblicazione rivolgersi a:
Istat - Servizio Condizioni economiche delle famiglie
Tel. 06 59524719
e-mail: clceccar@istat.it, madimarc@istat.it, rinaldel@istat.it

Eventuali rettifiche al volume saranno diffuse attraverso il sito
www.istat.it nella relativa pagina di presentazione.

L'indagine europea sui redditi e le condizioni di vita delle famiglie (Eu-Silc)

Metodi e Norme n. 37 - 2008

Istituto nazionale di statistica
Via Cesare Balbo, 16 - Roma

Coordinamento:
Servizio produzione editoriale

Supervisione tecnica editoriale:
Carlo Nappi

Stampa:
Istat - Centro stampa
Via Tuscolana, 1788 - 00173 Roma

Ottobre 2008 - copie 700

Si autorizza la riproduzione a fini non commerciali
e con citazione della fonte

Indice

Pagina

1. Il progetto Eu-Silc e la nuova indagine italiana su reddito e condizioni di vita	7
1.1 Il regolamento Eu-Silc	7
1.2 La progettazione della nuova indagine italiana	9
1.3 Obiettivi dell'indagine e contenuti informativi.	10
1.4 Strategie di rilevazione per i redditi	13
1.5 Redditi lordi, imposte, contributi sociali	17
1.6 Comparabilità internazionale	18
1.7 L'indagine longitudinale	19
2. Strategia di campionamento e precisione delle stime	21
2.1 Definizioni	21
2.2 Il disegno di rilevazione e il piano di campionamento	22
2.3 Strategia di ponderazione per le stime trasversali	32
2.4 Valutazione della precisione delle stime	43
3. Le strategie di rilevazione	49
3.1 La popolazione oggetto di indagine e l'unità di rilevazione	49
3.2 Le strategie di rilevazione	49
3.3 La struttura dei questionari	51
3.4 La gestione informatizzata del processo di indagine e delle famiglie campione (Sigif)	56
4. Procedure di controllo, correzione, imputazione e integrazione di dati da fonte amministrativa	57
4.1 Introduzione	57
4.2 La strategia di acquisizione dati (Blaise)	57
4.3 Il processo di correzione e imputazione delle informazioni	59
4.4 Correzioni e imputazioni delle informazioni sui redditi	66
4.5 L'integrazione dei dati campionari con fonti amministrative	87
5. I regolamenti europei e la documentazione sulla qualità	103
5.1 Introduzione	103
5.2 Il regolamento europeo n. 28/2004 e i <i>quality report</i>	103
5.3 Il contenuto dei <i>quality report</i>	105
5.4 I <i>quality report</i> prodotti da Eurostat	107
6. Comparabilità internazionale e problemi metodologici aperti	109
6.1 Comparabilità dei dati a livello internazionale	109
6.2 Metodologia per la stima degli affitti figurativi	115

	Pagina
6.3 Dai redditi netti ai redditi lordi: metodologia di stima e prospettive	132
Appendice - Le indagini sui redditi e le misure di benessere	143
A.1 Reddito ed indicatori di disagio economico e di deprivazione materiale	143
A.2 Disuguaglianza dei redditi e indicatori di povertà: confronti con l'indagine sui bilanci delle famiglie della Banca d'Italia	162
Bibliografia	175

Capitolo 1 - Il progetto Eu-Silc e la nuova indagine italiana su reddito e condizioni di vita

1.1 - Il regolamento Eu-Silc

Il Regolamento del Parlamento europeo, *Statistics on Income and Living Conditions* (n° 1177/2003, meglio noto come Eu-Silc)¹, risponde alla crescente domanda di informazioni da parte delle istituzioni nazionali ed europee, della comunità scientifica e degli stessi cittadini sulle condizioni di vita nei diversi paesi dell'Unione. Il progetto ha come obiettivo principale la produzione sistematica di statistiche comunitarie su reddito, povertà ed esclusione sociale, sia a livello trasversale che longitudinale, puntando all'armonizzazione di un insieme di indicatori statistici.

Il Regolamento precisa le responsabilità dei paesi membri e di Eurostat nel progetto e definisce un insieme di regole comuni al fine di migliorare la qualità, la comparabilità e la tempestività dei dati, oltre a promuovere una migliore integrazione delle nuove statistiche nei sistemi statistici nazionali.

Gli aspetti metodologici del nuovo strumento Eu-Silc sono stati sviluppati in cinque regolamenti della Commissione (*Sampling and tracing rules; Definitions, List of primary variables; Fieldwork aspect and imputation procedures; Intermediate and final quality reports*); inoltre, ogni anno, un nuovo regolamento definisce la lista di variabili target per un modulo *ad hoc* scelto tra una rosa di tematiche di interesse.² Il progetto ha previsto un periodo di transizione (fino al 2007) che consentisse agli Istituti nazionali di statistica di adattare i propri strumenti agli standard comuni con particolare riferimento a: i fitti imputati, i contributi sociali a carico dei datori di lavoro e le componenti del reddito lordo.

Il progetto Eu-Silc è stato lanciato nel 2003 su base sperimentale in sette paesi (Belgio, Norvegia, Grecia, Lussemburgo, Austria, Danimarca e Irlanda) anche con l'obiettivo di studiare l'impatto dei cambiamenti dello strumento nella serie degli indicatori di Laeken. Il lancio ufficiale di Eu-Silc si è avuto invece nel 2004 in tredici dei vecchi Stati membri, compresa l'Italia (non hanno partecipato Paesi Bassi, Germania e Regno Unito), e in dieci di quelli nuovi (eccetto Estonia), oltre che in Norvegia e Islanda. Nel 2005, Eu-Silc ha raggiunto la sua piena estensione con venticinque stati membri, più Norvegia e Islanda (Eu-Silc è in preparazione anche in Turchia, Romania, Bulgaria, Svizzera).

1.1.1 Le principali differenze tra Eu-Silc ed Echp

Il progetto ha sostituito l'*European Household Panel Survey* (Echp), condotto tra il 1994 e il 2001, e si differenzia da quest'ultimo, in primo luogo, per la base legale che lo contraddistingue. Il regolamento ha consentito di ancorare il nuovo strumento al Sistema

I paragrafi 1.1 e 1.2 sono a cura di Cristina Freguja, i paragrafi 1.3-1.7 sono a cura di Marco Di Marco

¹ Il Regolamento è stato pubblicato il 3 luglio 2003 (Official Journal n. 165).

² Per il primo anno di indagine non sono stati previsti moduli aggiuntivi, mentre nel 2005 è stato incluso un approfondimento su "Trasmissione intergenerazionale della povertà"; nel 2006 su "Partecipazione sociale"; nel 2007 su "Adeguatezza dell'abitazione". Il modulo per il 2008 riguarderà "L'indebitamento finanziario".

statistico nazionale, garantendo il rilascio di una serie di dati, secondo un preciso calendario³, anche in assenza di fondi della Commissione (che per i primi quattro anni contribuisce per circa 2/3 dei costi complessivi).

Per assicurare la comparabilità dei dati tra i paesi membri, il regolamento definisce alcune regole comuni (popolazione *target*, definizione delle variabili, dimensioni del campione, regole di inseguimento delle famiglie e dei loro componenti). Tuttavia, al contrario di Echip, che rappresentava uno strumento rigido e le cui articolazioni venivano stabilite in sede europea, il regolamento Eu-Silc ha lasciato ai singoli paesi alcuni margini di flessibilità rispetto all'impiego di differenti fonti di dati (indagine campionaria/archivi), al periodo di riferimento del reddito (fisso/mobile), alla modalità di raccolta delle informazioni sui redditi lordi (indagine/archivi/microsimulazione) e alla struttura dei questionari nazionali. Con particolare riferimento alla definizione del disegno campionario, ai metodi di imputazione e al calcolo degli stimatori, Eurostat, in ottemperanza all'articolo 1 del regolamento che riconosce la comparabilità come obiettivo fondamentale, ha prodotto una serie di raccomandazioni che sono state raccolte dalla maggior parte dei paesi.

Lo sviluppo di studi metodologici condotti in stretta collaborazione tra Stati Membri ed Eurostat sono oggetto dei lavori di una *task-force* europea, cui partecipa anche l'Italia, che ha aperto un lungo percorso di analisi e sperimentazioni. L'obiettivo della comparabilità si configura, infatti, come un processo di convergenza graduale che dovrà considerare, non solo le metodologie adottate dai diversi Istituti nazionali, ma anche le importanti differenze che caratterizzano i sistemi di *welfare* dei paesi membri e la loro evoluzione nel corso del tempo.

Il requisito della tempestività dei dati è stato riconosciuto come uno dei principali obiettivi del nuovo strumento di rilevazione. Pertanto, mentre per l'indagine panel europea, i dati *cross-sectional* e longitudinali venivano raccolti e trattati nello stesso momento, nel caso di Eu-Silc è stato previsto che i dati trasversali e longitudinali, che possono derivare da fonti separate, siano rilasciati secondo un diverso calendario.

Riguardo al disegno campionario, l'Italia ha seguito le raccomandazioni rivolte da Eurostat ai paesi che intendevano lanciare una nuova indagine, adottando un disegno che prevede, ogni anno, un campione trasversale, costituito da quattro gruppi di rotazione. A regime, ogni sottogruppo permane nella rilevazione per quattro anni, consentendo di ottenere sia stime trasversali che longitudinali.

Il disegno adottato dall'Istat è stato il frutto di un'attenta valutazione di differenti ipotesi elaborate tenendo conto dei vincoli imposti dal regolamento sotto il profilo della numerosità campionaria minima (stabilita in base alla precisione della stima della percentuale di famiglie povere a livello nazionale) e della durata minima del panel, ma anche della possibilità di soddisfare i requisiti relativi alle due componenti trasversale e longitudinale attraverso due rilevazioni indipendenti.

La scelta di un disegno integrato presenta una rilevante serie di vantaggi. Il *panel* ruotato (campione parzialmente sovrapposto tra due anni consecutivi) determina una riduzione dell'errore di campionamento delle stime trasversali di variazione (*net change*) e permette l'introduzione nel campione di nuovi sotto-gruppi di popolazione (ad esempio gli immigrati), consentendo una più ampia rappresentatività dei dati rispetto a un *panel* puro. Inoltre, la minore durata della componente longitudinale (quattro anni) limita il *response burden* e il conseguente *attrition*.

³ I dati *cross-sectional* relativi ai redditi dell'anno T vengono rilevati al tempo T+1, trasmessi ad Eurostat entro novembre dell'anno T+2 e resi disponibili da Eurostat alla comunità scientifica entro marzo T+3. La componente longitudinale viene invece trasmessa dai paesi membri entro marzo T+3 e messa a disposizione di ricercatori entro luglio T+3.

1.2 - La progettazione della nuova indagine italiana

In Italia, forse più che in altri Paesi, la rilevazione del reddito presenta numerosi aspetti problematici e per mettere a punto la nuova indagine, l'Istat ha progettato e condotto tra il 2000 e il 2003 tre indagini pilota che si sono affiancate a una serie di analisi sulle fonti esistenti (*Bilanci di famiglia* della Banca d'Italia e il *Panel europeo sulle famiglie*) e di sperimentazioni *ad hoc*.

La definizione di un'adeguata strategia di rilevazione dei redditi tramite intervista diretta si è avvalsa dei risultati di un'indagine pilota "*Reddito e condizioni di vita*", condotta dall'Istat nel 2002 su un campione di ottocento famiglie (seguita a una prima indagine pilota su duecento famiglie), che ha permesso di testare la disponibilità da parte dei cittadini a utilizzare i documenti fiscali per rispondere all'intervista.

Un versante di attività complementare all'analisi dei risultati delle rilevazioni sul campo ha impegnato il gruppo di ricerca nella realizzazione di sperimentazioni *ad hoc* attraverso metodologie di natura qualitativa. Per migliorare le tecniche di rilevazione del reddito in relazione ai possibili problemi psicologici, di memoria, di conoscenza e di motivazione dei rispondenti, è stata condotta, infatti, una serie di colloqui in profondità, con l'obiettivo di far emergere le propensioni, i condizionamenti, le resistenze comuni alle differenti tipologie di percettori. Ai colloqui in profondità è seguito uno studio - condotto con la tecnica del *focus group* e specificamente rivolto ai lavoratori autonomi - allo scopo di individuare, attraverso la discussione e le valutazioni emerse dai partecipanti, le strategie di approccio e di raccolta dei dati (contatto con la famiglia, tipo e sequenza delle domande, eccetera) più adeguate ed efficaci a massimizzare la disponibilità dei rispondenti e l'attendibilità delle loro risposte.

I risultati delle analisi, delle sperimentazioni e delle prime due indagini pilota hanno consentito l'affinamento degli strumenti per la rilevazione delle variabili di interesse, approdando alla messa a punto di nuovi questionari che, alla fine del 2003, sono stati sottoposti a verifica nell'ambito dell'indagine sulle *condizioni di vita* condotta dall'Istat su un campione di ottomila famiglie residenti in Italia. Questa sperimentazione ha rappresentato una sorta di "prova generale" dell'indagine Eu-Silc che ha consentito anche di definire le strategie per il controllo e la correzione delle variabili e per l'imputazione delle mancate risposte.

Va ricordato un altro importante fronte di sperimentazione che si è aperto con il *linkage* tra i dati campionari e quelli di fonte Agenzia delle entrate per la raccolta dei redditi lordi e la ricostruzione dei redditi netti da lavoro autonomo. L'esperimento, condotto dapprima sui dati dell'indagine Echn del 2001, è stato replicato in occasione della prima edizione dell'indagine Eu-Silc del 2004 (ampliato ai dati dell'archivio Inps sui pensionati). Per la seconda *wave* i dati amministrativi sono stati utilizzati, insieme a quelli raccolti tramite questionario⁴, anche per la ricostruzione dei redditi da lavoro dipendente. La strategia messa a punto presenta dunque da un punto di vista metodologico elementi di elevata complessità, ed è basata su un utilizzo sempre più importante delle fonti amministrative a supporto della validazione e correzione dei dati sia a livello micro, sia a livello macro. I dati amministrativi, dopo un primo periodo di consolidamento delle procedure di *linkage* e del sistema di validazione incrociata con i dati dell'indagine, divengono il perno centrale del processo di trattamento dei dati dalla fase di *check*, alla fase di calcolo dei coefficienti di riporto all'universo per arrivare all'integrazione con i dati dell'indagine.

Un requisito fondamentale per lo studio dei processi di esclusione sociale è l'osservazione dinamica delle principali transizioni rispetto al mercato del lavoro, ai cambiamenti che intervengono nella struttura della famiglia (morti, nascite, eccetera), ai flussi di entrata ed uscita

⁴ I questionari sono disponibili all'indirizzo http://www.istat.it/strumenti/rispondenti/indagini/famiglia_societaeusilc/.

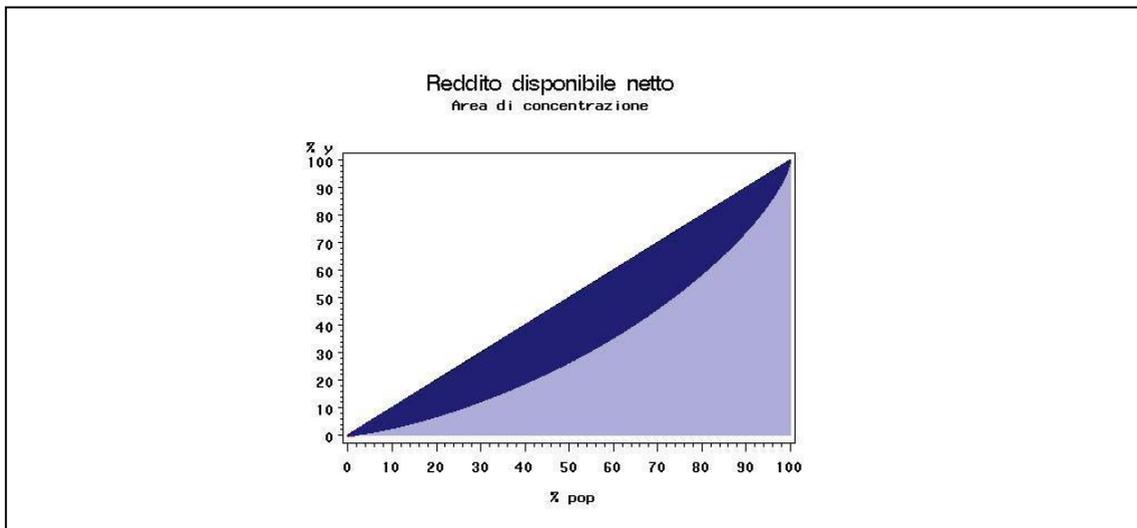
nello/dallo stato di povertà. La dimensione longitudinale dell'indagine Eu-Silc rappresenta quindi sul piano metodologico una sfida altrettanto importante e impegnativa di quella trasversale.

In questo ambito è stata messa a punto una complessa organizzazione per l'inseguimento delle famiglie e degli individui che garantisce il controllo puntuale e in tempo reale degli spostamenti sul territorio. Inoltre sono state definite le metodologie di controllo e correzione integrata dei dati campionari panel e *cross-section*, che tengono conto contestualmente anche dell'integrazione delle fonti amministrative, richiedendo l'implementazione di un piano di compatibilità delle informazioni longitudinali provenienti dalle due fonti di dati.

1.3 - Obiettivi dell'indagine e contenuti informativi

Le indagini sui redditi e le condizioni di vita hanno due tipi di utilizzo: scientifico e politico. Fra gli impieghi di carattere scientifico, i più importanti sono l'analisi della distribuzione dei redditi, della disuguaglianza e della povertà. La disponibilità di microdati consente infatti di rappresentare la distribuzione dei redditi di un paese attraverso grafici e indicatori.

Figura 1.1 - Area di concentrazione dei redditi e curva di Lorenz



Una rappresentazione grafica molto usata in letteratura è l'area di concentrazione (l'area più scura nella figura 1.1). Il confine inferiore dell'area di concentrazione è noto come curva di Lorenz e mostra quale percentuale di famiglie (asse orizzontale) guadagna una data frazione del reddito totale (mostrata sull'asse verticale). Il confine superiore dell'area di concentrazione è il segmento di equidistribuzione, che rappresenta il caso in cui tutte le famiglie hanno lo stesso reddito. L'area di concentrazione misura quindi la differenza fra la curva di Lorenz e il segmento di equidistribuzione ed è tanto minore quanto maggiore è l'eguaglianza della distribuzione dei redditi.

Alle analisi grafiche si accompagnano misure sintetiche del grado di disuguaglianza della distribuzione dei redditi. Fra gli indicatori noti in letteratura, il più utilizzato è l'indice di concentrazione del Gini, che assume valori compresi fra zero (perfetta eguaglianza: tutte le famiglie hanno lo stesso reddito) e uno (totale disuguaglianza: una sola famiglia possiede tutto il

reddito nazionale). Graficamente, l'indice di Gini misura il rapporto fra: (i) l'area di concentrazione e: (ii) l'area dell'intero triangolo al di sotto del segmento di equidistribuzione. L'indice di Gini può essere calcolato con varie formule, la più semplice delle quali è:

$$G = \frac{2 \text{cov}(y_i, r_i)}{\bar{y}}$$

dove: y_i è il reddito della singola unità statistica
 \bar{y} è il reddito medio

r_i è il rango frazionario, cioè il posto occupato dall'unità statistica nella distribuzione dei redditi dal più basso al più alto (diviso per il numero totale di unità statistiche).

Come è facile intuire, le indagini sui redditi e le condizioni di vita delle famiglie sono utilizzate nel dibattito di politica economica per valutare analiticamente lo stato delle famiglie in un paese. Ma, soprattutto, i dati sui redditi e le condizioni di vita possono essere usati per definire e valutare gli effetti sulle famiglie delle politiche economiche e sociali adottate in passato oppure programmate per il futuro. Per esempio, è possibile studiare gli effetti di una particolare riforma tributaria, descrivere chi sono i beneficiari di particolari trasferimenti pubblici eccetera.

1.3.1 Definizioni di reddito

Secondo la definizione prevalente nei manuali di economia, il reddito è semplicemente la somma dei consumi e dei risparmi:

$$Y = C + S$$

Si noti che il risparmio S è eguale alla variazione della ricchezza. Se in un periodo $Y > C$ (ovvero, se il reddito supera i consumi), il risparmio sarà positivo e si avrà un aumento di ricchezza. Quando invece $Y < C$, il risparmio sarà negativo e tale situazione corrisponderà ad una diminuzione di ricchezza. In quest'ultimo caso, in effetti, l'individuo finanzia i suoi consumi attraverso la vendita di beni e attività finanziarie oppure indebitandosi (in tutt'e due i casi, la ricchezza si riduce).

Da questa semplice premessa discende la definizione di reddito accettata internazionalmente, ed accolta dal progetto Eu-Silc nelle sue linee fondamentali: *“il reddito è la quantità massima di moneta che un individuo può spendere per consumi senza diminuire la propria ricchezza, cioè senza vendere parte del proprio patrimonio e senza fare nuovi debiti”*.

Nell'indagine Eu-Silc, il reddito viene osservato come un insieme di entrate ricavate da fonti diverse, secondo lo schema seguente:

1. Reddito guadagnato sul mercato

1.1 Redditi da lavoro

- dipendente
- autonomo

1.2 Redditi da capitale

- reale (affitti e rendite di terreni e fabbricati)
- finanziario (interessi, dividendi, utili)
- intellettuale (diritti d'autore)

2. Reddito da trasferimenti

2.1 Trasferimenti pubblici

- pensioni
- altri trasferimenti pubblici in denaro (per esempio assegni familiari)

2.2 Trasferimenti privati

- aiuti in denaro di familiari ed amici, assegni di ex-coniugi
- aiuti in denaro di istituzioni private (per esempio da associazioni religiose).

Accanto a queste componenti, misurate in moneta, si considerano anche altre risorse “non-monetarie” che concorrono al benessere familiare:

- salari in natura (*fringe benefits*): come l’uso gratuito di una abitazione, l’auto aziendale per usi privati, i buoni pasto (dal 2007), l’asilo nido aziendale (dal 2007);
- affitti imputati delle case occupate dai proprietari: che è pari al valore del servizio che queste abitazioni rendono a chi ne è proprietario. Per convenzione, è “come se” i proprietari affittassero la casa a sé stessi;
- autoconsumi (dal 2007): cioè il valore stimato dei beni che la famiglia ha eventualmente prodotto per il proprio consumo, come per esempio frutta, vino e ortaggi.

Sono invece escluse dalla definizione di reddito adottata per l’indagine Eu-Silc, per difficoltà di rilevazione e/o di stima del valore monetario corrispondente, alcune componenti che pure concorrono a determinare le condizioni economiche delle famiglie:

- trasferimenti pubblici in natura, come per esempio i servizi sanitari e scolastici forniti gratuitamente o a prezzi agevolati dalla pubblica amministrazione. È indiscutibile che tali beni, se utilizzati, contribuiscono al benessere delle famiglie. Tuttavia, è praticamente impossibile, in assenza di un mercato privato indipendente, stimare il valore corrispondente a questo tipo di benefici pubblici ricevuti in natura dalle famiglie. Anche quando esistono servizi privati in concorrenza di quelli pubblici, infatti, il loro prezzo di mercato è “residuale” rispetto alla politica di offerta dell’operatore pubblico. La valutazione al costo di produzione, a sua volta, ignora la qualità dei servizi e può non riflettere la disponibilità a pagare degli utenti;
- i beni e i servizi in natura ricevuti da parenti e amici (per esempio, la cura dei figli da parte di una parente non coabitante), per la difficoltà di valutarne sia la quantità, sia il valore “figurativo”;
- per difficoltà di rilevazione, sono anche escluse tutte quelle attività lavorative effettuate dai membri della famiglia in sostituzione di analoghi servizi di mercato, come per esempio la riparazione di elettrodomestici, la manutenzione di mobili eccetera. La difficoltà in questo caso riguarda la vasta gamma coperta dalla produzione domestica: in effetti, anche le pulizie di casa e la preparazione dei cibi sostituiscono servizi acquistabili altrimenti sul mercato.

Sono, infine, escluse alcune entrate eccezionali che sono considerate come variazioni “istantanee” della ricchezza:

- le vincite alla lotteria
- le eredità e le donazioni *una tantum*
- i guadagni in conto capitale, cioè gli aumenti del valore del patrimonio posseduto (case, terreni, gioielli, azioni ed altre attività finanziarie).

A partire dall’edizione 2007, l’indagine Eu-Silc includerà anche i redditi lordi individuali e familiari. Si tratta di una importante innovazione metodologica, che consentirà di valutare in che misura la disuguaglianza nella distribuzione dei redditi dipende dalle opportunità di mercato

(distribuzione primaria), per esempio dai livelli di occupazione e di salario, e quale sia l'effetto redistributivo delle imposte e dei trasferimenti pubblici.

Nel progetto Eu-Silc, il reddito lordo totale sarà eguale alla somma dei redditi netti, delle imposte personali sui redditi, delle imposte patrimoniali e dei contributi sociali a carico dei lavoratori.⁵

1.4 - Strategie di rilevazione per i redditi

La rilevazione campionaria dei redditi pone numerosi problemi, dovuti a due ordini di motivi:

- scarsa conoscenza da parte degli intervistati:
 - delle definizioni di reddito
 - degli importi esatti percepiti
- scarsa disponibilità a rispondere all'intervista:
 - per diffidenza (soprattutto timore di controlli fiscali)
 - per sfiducia nelle istituzioni e nell'utilità delle indagini statistiche.

Il primo problema è stato affrontato attraverso una formulazione il più possibile semplice e precisa del questionario, l'accurata formazione dei rilevatori e prevedendo, per chi non ricorda un importo esatto, la possibilità di dare risposte approssimate. Per superare il secondo problema, è stata determinante, almeno nell'esperienza "sul campo" del progetto Eu-Silc, la condivisione da parte delle famiglie dello scopo dell'indagine, insieme alla reputazione dell'Istat a garanzia dell'assoluta riservatezza delle informazioni raccolte.

La scelta fondamentale in merito alla strategia di rilevazione ha portato, per l'indagine Eu-Silc, ad un approccio sostanzialmente diverso da quello seguito per l'Indagine sui bilanci delle famiglie della Banca d'Italia. Nella strategia di rilevazione seguita dalla Banca d'Italia, l'organizzazione delle interviste prevede l'esclusione a priori delle famiglie che dichiarano di non voler rispondere a domande sui redditi. Quindi, la Banca d'Italia assegna una priorità assoluta all'obiettivo di ridurre al minimo le mancate risposte sui redditi, anche al costo di conseguire un minore tasso di adesione all'indagine.

La strategia dell'Istat, invece, ha come obiettivo prioritario quello di massimizzare l'adesione all'indagine senza, nel contempo, rinunciare al contenimento delle mancate risposte sui redditi. Quest'ultima strategia, più tradizionale e, in un certo senso, obbligata per un istituto nazionale di statistica, implica in effetti un maggior rischio potenziale di mancate risposte a singoli quesiti particolarmente "delicati".

In questo contesto, assumono un'importanza cruciale sia le modalità di contatto e di sensibilizzazione delle famiglie, sia le tecniche di rilevazione che consentono di ridurre al minimo le mancate risposte (in particolare il disegno del questionario e la formazione dei rilevatori), sia infine le metodologie di imputazione delle mancate risposte sui redditi.

Considerata la delicatezza dell'argomento dell'intervista, la partecipazione delle famiglie alle tre indagini pilota ed alla prima edizione dell'indagine è risultata sostanzialmente soddisfacente: il rapporto fra interviste realizzate e famiglie contattate indica un sostanziale successo della strategia di sensibilizzazione degli intervistati. Questo risultato è spiegato in larga misura dallo sforzo di far capire e condividere alle famiglie lo scopo dell'indagine.

Nel contatto con la famiglia, infatti, l'Istat ha cercato di comunicare nel modo più semplice possibile le ragioni scientifiche e di politica economica che giustificano la richiesta di partecipare ad un'indagine sui redditi e sulle condizioni di vita. Inoltre, è stata fondamentale la

⁵ Entro il 2007, sulla base di studi di fattibilità, si deciderà se considerare anche i contributi sociali a carico dei datori di lavoro.

garanzia di assoluta riservatezza sulle informazioni raccolte. Infine, la comprensibilità delle domande e la preparazione dei rilevatori hanno rafforzato la disponibilità iniziale delle famiglie.

A questo proposito, è ragionevole ipotizzare sulla base dell'esperienza dell'indagine Eu-Silc che l'associazione di domande riferite alle condizioni di vita ed ai problemi economici della famiglia, accanto a quelle "ragionieristiche" sui redditi percepiti, sia un fattore positivo di motivazione dei rispondenti. Nella maggioranza dei casi, in effetti, le famiglie hanno compreso l'importanza e lo scopo dell'indagine e collaborato attivamente all'intervista, consultando quando possibile i documenti a loro disposizione (come la busta paga o l'estratto delle condizioni di mutuo). Ciò ha comportato una ricaduta positiva anche per quanto riguarda il contenimento delle mancate risposte su singoli quesiti e la qualità delle informazioni raccolte.

1.4.1 Integrazione delle misure di reddito con informazioni sulle condizioni di vita

Tradizionalmente, il reddito monetario viene utilizzato per valutare il benessere delle famiglie e degli individui che ne fanno parte. Tuttavia, un recente filone di studi ha messo in luce che l'analisi delle condizioni delle famiglie sulla base di una misura soltanto monetaria a volte non consente di valutare aspetti importanti della qualità della vita. Per quanto importante, infatti, il reddito monetario è soltanto un mezzo che gli individui utilizzano per avere ciò di cui hanno bisogno e la qualità della vita dovrebbe essere misurata anche considerando i risultati concreti che le persone desiderano ottenere usando la moneta.

Come dicono gli economisti, la moneta non ha valore intrinseco. Insieme agli indicatori monetari si dovrebbero quindi utilizzare anche indicatori del benessere "reale", inteso quest'ultimo come l'insieme delle possibilità di avere beni e, inoltre, di fare o essere ciò che si considera necessario per star bene. La qualità della vita delle persone può essere valutata direttamente considerando le possibilità di fare, essere o avere ciò che si desidera (capacità) ed i risultati effettivamente ottenuti (funzionamenti). Uno degli obiettivi più importanti del progetto Eu-Silc è quello di raccogliere, accanto ai redditi, anche informazioni sulle condizioni di vita (abitazione, salute, lavoro eccetera) per consentire una più realistica rappresentazione multidimensionale della qualità della vita.

A questo scopo, l'indagine Eu-Silc considera un insieme ampio di rilevazioni delle condizioni di vita. A livello individuale, sono rilevate le condizioni lavorative, i livelli di istruzione, il grado di salute. A livello familiare, le caratteristiche della casa e della zona di abitazione ed una batteria di indicatori soggettivi sulle difficoltà economiche della famiglia. Altri indicatori non-monetari di disagio rilevano le conseguenze della scarsità di moneta. Si rileva, per esempio, se le famiglie possono permettersi di riscaldare adeguatamente l'abitazione, una settimana di vacanza fuori-casa almeno una volta l'anno e se, negli ultimi dodici mesi, ci sono stati momenti in cui la famiglia non ha avuto i soldi per pagare particolari beni e servizi (cibo, vestiti necessari, spese mediche, affitto, mutuo per la casa, bollette di luce, gas, telefono, spese scolastiche, spese per trasporti).

Per non appesantire troppo l'intervista, la serie di indicatori non monetari è prevalentemente orientata a rilevare le conseguenze della scarsità di risorse monetarie sulla qualità della vita. In effetti, la maggior parte delle variabili non monetarie rileva i funzionamenti (o, meglio, i mancati funzionamenti) che dipendono dalla disponibilità di moneta. Per esempio, si chiede alle famiglie se possono riscaldare adeguatamente l'abitazione, permettersi un livello di nutrizione adeguato, se riescono a sostenere senza difficoltà le spese scolastiche, ad affrontare spese impreviste, ad andare in vacanza almeno una settimana l'anno eccetera.

In questo contesto più ampio, la misura del reddito monetario può essere più correttamente considerata come un importante indicatore indiretto delle opportunità di trasformare le proprie capacità e risorse personali in potere d'acquisto.

La ricchezza informativa dell'indagine è anche la conseguenza dell'inserimento di batterie di domande aggiuntive rispetto a quelle previste dal Regolamento Eu-Silc. In alcuni casi è stato infatti necessario articolare in maggior dettaglio la variabile-obiettivo per ottenere una precisione sufficiente. Per esempio, l'esperienza delle indagini pilota ha suggerito di chiedere agli intervistati, oltre alla rata di mutuo pagata, anche le informazioni relative al capitale prestato, al tasso di interesse e alla durata del prestito.

Questo dettaglio di informazione consente di controllare con precisione la coerenza delle risposte e di definire procedure di imputazione più affidabili in caso di mancata risposta. Rispetto all'insieme delle variabili-obiettivo stabilite dal Regolamento Eu-Silc, inoltre, sono stati aggiunti ulteriori quesiti relativi alla qualità della vita la cui utilità era stata già verificata in altre indagini campionarie dell'Istat (in particolare, nelle indagini Multiscopo sulle famiglie).

1.4.2 Strategie di controllo e correzione delle informazioni

L'abbondanza di informazioni e la complessità di alcuni percorsi di intervista richiedono un accurato ed esteso controllo "incrociato" delle informazioni. I dati vengono sottoposti ad articolate procedure di identificazione e correzione degli errori. Nella fase di registrazione in formato digitale, una serie di controlli automatici elimina la maggior parte degli errori di digitazione, in particolare di quelli relativi agli importi monetari, chiedendo all'operatore di consultare con maggiore attenzione il questionario compilato in caso di valori anomali.

Successivamente, le informazioni vengono analizzate per individuare le possibili incoerenze logiche delle risposte. Il piano di controlli di compatibilità prevede la individuazione e la cancellazione dei dati incoerenti. In una fase successiva, le variabili qualitative vengono corrette deterministicamente nei casi in cui esistono, per lo stesso individuo, altre informazioni attendibili che consentono di ricostruire le risposte mancanti o errate. Quando questa operazione non risulta possibile, si ricorre a tecniche di imputazione probabilistica.

Nel caso delle variabili qualitative dell'indagine Eu-Silc, si è impiegato un algoritmo di ricerca del "donatore" più simile all'individuo che presenta informazioni mancanti o errate. La similitudine fra individui è valutata con riferimento alle principali caratteristiche individuali (età, sesso, titolo di studio eccetera) e alle variabili correlate con il dato da imputare. Per esempio, oltre ai dati anagrafici si considera anche la somiglianza nelle condizioni lavorative (settore, qualifica professionale, anni di anzianità eccetera) quando si deve imputare tramite donazione una mancata risposta relativa al tipo di contratto di lavoro (a tempo determinato o indeterminato).

Le informazioni dell'individuo più somigliante vengono in pratica "donate" (cioè, semplicemente, copiate) al posto dei valori errati o mancanti. Una caratteristica importante della procedura di imputazione da donatore adottata dall'Istat è che rispetta il piano di compatibilità, evitando di introdurre incoerenze nei dati.

Le mancate risposte sui redditi sono state imputate con i modelli di regressione multipla disponibili nel software Iveware dell'Università del Michigan. Progettato come modulo aggiuntivo di Sas, Iveware consente di trattare con relativa semplicità un insieme complesso di dati correlati fra loro nei casi in cui sarebbe troppo oneroso esplicitare un modello completo di relazioni multivariate. In sintesi, la tecnica consiste nella generazione, per ogni valore mancante di ciascuna variabile, di una predizione condizionale ai valori di tutte le altre variabili. In un primo passo preliminare, il modello imputa in sequenza i valori mancanti di ognuna delle variabili, iniziando da quella che ne presenta una minore percentuale.

Una volta che ne siano stati imputati i valori mancanti, ciascuna variabile entra come covariata nel processo di imputazione di tutte le altre. Nei passi successivi, dal secondo fino alla convergenza dei valori imputati, il modello è costituito da tante equazioni quante sono le

variabili da imputare. Quando è raggiunta la convergenza delle stime, ogni equazione predice i valori mancanti della variabile dipendente.

Il principale vantaggio del software Iweware consiste nella flessibilità d'uso: per ogni variabile da imputare è possibile sia selezionare le covariate rilevanti, sia definire il modello di regressione appropriato (lineare per le variabili continue, *logit* per le dicotomiche, *log-lineare* per le variabili discrete). Inoltre, è possibile vincolare le imputazioni al rispetto di vincoli di coerenza (per esempio, imputare soltanto un particolare sottinsieme di casi) e assegnare limiti superiori ed inferiori ai valori da imputare. Quest'ultima caratteristica è utile per trasformare in valori puntuali le cifre approssimate indicate da quegli intervistati che non ricordano l'importo esatto dei redditi. In questi casi, i valori puntuali sono imputati con Iweware all'interno di una banda centrata sul valore approssimato.

Per l'indagine Eu-Silc, gli effetti delle imputazioni vengono controllati attraverso il confronto dell'intero profilo distributivo, prima e dopo l'intervento di correzione (cioè tutti i percentili, attraverso la sovrapposizione dei grafici delle distribuzioni di frequenza semplici e cumulate). Inoltre, gli effetti delle imputazioni sono valutati comparando i principali parametri della distribuzione (media, mediana, primo e terzo quartile, minimo, massimo, *standard error*) disaggregati per sesso, età, area geografica, titolo di studio, condizione professionale ed altre covariate significative.

1.4.3 Integrazione di dati amministrativi e campionari: i redditi da lavoro autonomo

Il *record linkage* fra i dati campionari della prima edizione dell'indagine Eu-Silc e i dati dei modelli Unico e 730 ha consentito una importante innovazione metodologica per quanto riguarda la ricostruzione dei redditi netti da lavoro autonomo. Lo schema 1.1 illustra in modo semplificato il problema della rilevazione dei redditi netti autonomi: la somma delle due aree evidenziate corrisponde al reddito netto effettivo, che è l'obiettivo del progetto Eu-Silc (Y_{vero}).

Schema 1.1 – I redditi netti autonomi



$$Y_{\text{vero}} = \text{Reddito imponibile netto} + \text{Reddito esente}$$

Poiché il reddito esente da imposta non è necessariamente presente nelle dichiarazioni dei redditi, il reddito imponibile netto dell'Agenzia delle entrate può essere diverso dal reddito netto

effettivamente percepito. A loro volta, i dati campionari sui redditi autonomi possono essere affetti da *under-reporting*, cioè dalla tendenza a dichiarare meno di Y_{vero} . Per l'indagine Eu-Silc, la variabile reddito netto autonomo Y_{aut} è costruita in base alla seguente procedura:

(per ogni percettore di redditi da lavoro autonomo):

1. Si considera inizialmente il valore risultante dalla dichiarazione dei redditi: $Y_{\text{aut}} = Y_{\text{fisco}}$
2. Se il reddito campionario risulta maggiore, lo si sostituisce al reddito fiscale: $Y_{\text{aut}} = Y_{\text{camp}}$

In altre parole, la regola di scelta fra le due fonti di dati è $\max(Y_{\text{fisco}}, Y_{\text{camp}})$. L'obiettivo è di ridurre al minimo il grado di sottostima del reddito vero con i dati a disposizione, senza ricorrere a correzioni basate su ipotesi *ad hoc*. Il reddito netto autonomo Y_{vero} è infatti pari alla somma del reddito netto campionario Y_{camp} e del reddito non riportato nell'intervista Y_{under} . Nello stesso tempo, il reddito netto vero è anche pari alla somma del reddito netto imponibile Y_{fisco} e del reddito evaso/eroso Y_{es} :

$$[1] Y_{\text{camp}} + Y_{\text{under}} = Y_{\text{vero}} = Y_{\text{fisco}} + Y_{\text{es}}$$

Da cui segue immediatamente che:

- (A) $Y_{\text{fisco}} > Y_{\text{camp}} \Rightarrow Y_{\text{under}} > Y_{\text{es}}$
 (B) $Y_{\text{fisco}} < Y_{\text{camp}} \Rightarrow Y_{\text{under}} < Y_{\text{es}}$

Assumendo che l'*under-reporting* non sia mai negativo, cioè che il rispondente non riporti mai nell'intervista un reddito netto superiore al vero, la scelta del maggiore fra il dato campionario ed il dato dell'Agenzia delle entrate minimizza la sottostima del reddito che conseguirebbe all'utilizzo esclusivo di una sola delle due fonti. In altre parole, nei casi in cui $Y_{\text{fisco}} > Y_{\text{camp}}$ il dato tributario corregge, almeno parzialmente, l'*under-reporting* campionario mentre, nei casi in cui $Y_{\text{camp}} > Y_{\text{fisco}}$ il dato campionario corregge l'evasione/erosione presente nel dato tributario. La procedura, inoltre, migliora la comparabilità internazionale dei dati di reddito italiani.

1.5 - Redditi lordi, imposte, contributi sociali

A partire dal 2007, il progetto Eu-Silc prevede la disponibilità, accanto ai redditi netti, di microdati su redditi lordi, imposte e contributi sociali. Si tratta di un importante obiettivo, che consentirà agli studiosi e ai *policymakers* europei di valutare gli effetti delle politiche tributarie e sociali sulla distribuzione dei redditi. Il raggiungimento di tale ambizioso obiettivo costituirà un rilevante avanzamento dello stato dell'arte in materia di statistiche sulla distribuzione dei redditi. Attualmente, i progetti di ricerca internazionali sono basati sull'impiego di modelli di microsimulazione applicati ad indagini nazionali non armonizzate fra loro. Anche per quanto riguarda l'Italia, la produzione di microdati su redditi netti e lordi all'interno di una stessa indagine campionaria costituisce una novità significativa.

Per conseguire questo difficile obiettivo l'Istat sta sperimentando contemporaneamente due metodologie: la costruzione di un modello di microsimulazione delle imposte e dei contributi

sociali e la integrazione dei dati campionari con fonti amministrative. Per quanto riguarda le microsimulazioni, è già disponibile il modello SM2, costruito appositamente per l'indagine Eu-Silc da un *team* internazionale di ricerca coordinato da Vijay Verma.

Il principale vantaggio del modello SM2 è la sua estrema flessibilità. Il modello ha una struttura modulare e può essere facilmente integrato con sotto-sezioni "nazionali" che consentono di replicare le diverse legislazioni europee in materia di imposte, contributi e trasferimenti sociali. Inoltre, può trattare sia la conversione in redditi lordi di dati campionari raccolti al netto delle imposte e dei contributi sociali, sia il caso opposto di calcolo dei redditi netti a partire da dati lordi (campionari o amministrativi).

Il cuore del modello è una *routine* iterativa che, utilizzando i parametri della scheda di imposta e le aliquote contributive, stima le imposte sui redditi ed i contributi sociali e consente la costruzione delle variabili-obiettivo (lorde o nette) che non sono disponibili nei dati campionari e/o amministrativi. Inoltre, il modello può stimare separatamente sia le imposte anticipate (cioè trattenute alla fonte) sulle singole componenti di reddito, sia il saldo finale a credito o a debito.

La qualità delle stime ottenute con i modelli di microsimulazione dipende, ovviamente, dalla qualità dei dati di input e dalla accuratezza delle sotto-sezioni che incorporano i parametri della normativa tributaria e contributiva. In una prima ipotesi di lavoro, per l'indagine Eu-Silc si era immaginato di utilizzare il modello di microsimulazione nella maniera più tradizionale, ovvero per la conversione di redditi netti campionari in redditi lordi. Tuttavia, l'*under-reporting* nei dati campionari, la complessità della normativa tributaria e la disponibilità di dati amministrativi suggeriscono la possibilità di un utilizzo meno banale del modello.

Due sembrano, al momento, le prospettive metodologiche più promettenti. In primo luogo, il modello SM2 può essere integrato con le procedure di imputazione dei valori mancanti dei redditi netti e lordi (sia campionari sia amministrativi) per individuare eventuali anomalie nei dati amministrativi ed eliminare potenziali incoerenze generate dalla rigidità della sequenza:

correzione/imputazione dei dati campionari → controllo/integrazione dei dati amministrativi.
--

In secondo luogo, in presenza di dati sia amministrativi sia campionari, l'utilizzo del modello potrebbe rivelarsi prezioso per trattare nel modo più opportuno ed efficiente le numerose differenze che esistono nella definizione delle variabili del progetto Eu-Silc rispetto a quelle adottate nelle fonti amministrative.

1.6 - Comparabilità internazionale

Su proposta dell'Istat, nel Regolamento dell'indagine Eu-Silc la comparabilità internazionale dei dati è definita come un obiettivo fondamentale, da perseguire attraverso opportuni studi metodologici. Nella fase preliminare del progetto Eu-Silc, lo sforzo principale di Eurostat e dei paesi membri è stato concentrato nel definire accuratamente le variabili-obiettivo, eliminando per quanto possibile le ambiguità concettuali legate alle diverse terminologie nazionali.

Ulteriori problemi sono attualmente esaminati da una commissione di esperti, coordinata da Eurostat, cui partecipano alcuni paesi membri, compresa l'Italia. Fra le potenziali ragioni che potrebbero ridurre il grado di comparabilità internazionale dei redditi, quelle più urgenti e rilevanti riguardano:

- il fatto che alcuni paesi utilizzeranno dati di reddito campionari ed altri soltanto dati amministrativi. Se da un lato i redditi rilevati in interviste condotte nei vari paesi possono

essere affetti da *under-reporting* in misura diversa, dall'altro i dati di fonte amministrativa possono far riferimento a definizioni di reddito diverse da quelle accolte nel progetto Eu-Silc. L'Italia, che ha a disposizione dati sia campionari sia amministrativi, è impegnata a fornire un contributo particolarmente significativo su tali problemi;

- la definizione del reddito da lavoro autonomo, che nella versione attuale rischia di non essere sufficientemente comparabile a livello internazionale. A causa della eterogeneità delle fonti a disposizione dei diversi paesi, è stata ammessa la possibilità di rilevare il reddito autonomo sia come risultato economico dell'attività (profitto o perdita), sia come reddito imponibile nella dichiarazione dei redditi, sia infine come reddito impiegato per finanziare spese e risparmi personali e familiari. Nel caso dell'Italia, i tre concetti sopra richiamati non sono necessariamente coincidenti per tutti i lavoratori autonomi. Ai fini dell'analisi delle condizioni economiche effettive delle famiglie con redditi da lavoro autonomo, la definizione più adatta è quella che fa riferimento alle risorse monetarie che, grazie all'attività di lavoro autonomo, sono disponibili per le spese e per i risparmi della famiglia;
- la completezza della definizione di reddito disponibile e la possibilità di osservare redditi negativi. La definizione di reddito adottata dal progetto Eu-Silc, che riflette in buona sostanza le raccomandazioni del manuale di Canberra, include dal lato delle uscite sia i trasferimenti effettuati ad altre famiglie (per esempio gli assegni di mantenimento di un coniuge separato), sia le perdite dell'attività autonoma. Dal lato delle entrate, sono escluse poste che non sono considerate reddito, ma piuttosto come variazioni "istantanee" della ricchezza (per esempio: le vincite alla lotteria, i guadagni in conto capitale, il denaro preso a prestito). Poiché queste ultime possono essere impiegate per finanziare le uscite familiari correnti, ne risulta la possibilità di registrare redditi inferiori allo zero. Il problema cruciale, a questo proposito, consiste nel distinguere i casi "ammissibili", derivanti dall'esclusione di alcune fonti di risorse monetarie dalla definizione di reddito, da quelli in cui (a causa di *under-reporting* o evasione) non vengono osservati redditi compresi nella definizione.

1.7 - L'indagine longitudinale

Il disegno dell'indagine Eu-Silc integra una componente trasversale (*cross-section*) ed una longitudinale (*panel*) a rotazione quadriennale, secondo il seguente schema:

Prospetto 1.1 – Schema dell'indagine Eu-Silc

	2004	2005	2006	2007
1 ^a edizione	A B C D			
2 ^a edizione		B C D E		
3 ^a edizione			C D E F	
4 ^a edizione				D E F G

A partire dalla seconda edizione, un quarto del campione iniziale viene eliminato e sostituito: per esempio, con riferimento allo schema precedente, nella seconda edizione il gruppo di famiglie A viene sostituito dalle famiglie del gruppo E. Nella quarta edizione, sarà disponibile

un panel formato dalle famiglie che sono state intervistate in tutti e quattro gli anni (cioè, il gruppo **D** dello schema). Anche per l'aspetto longitudinale, l'obiettivo fondamentale è la minimizzazione dei tassi di *attrition*. A questo scopo, è stata predisposta una complessa organizzazione, centrata su un sistema informativo di gestione delle informazioni familiari (Sigif), allo scopo di rintracciare sul territorio nazionale gli individui e le famiglie che, fra un'edizione e l'altra dell'indagine, cambiano abitazione. Si rimanda al capitolo 2 per la descrizione dettagliata della componente longitudinale.

Capitolo 2 - Strategia di campionamento e precisione delle stime

2.1 - Definizioni

Popolazione di riferimento

La popolazione di riferimento è costituita da tutti i componenti delle famiglie residenti in Italia, anche se temporaneamente all'estero. Sono escluse le famiglie residenti in Italia che vivono abitualmente all'estero e i membri permanenti delle convivenze istituzionali (ospizi, brefotrofi, istituti religiosi, caserme, eccetera).

Unità di rilevazione

L'unità di rilevazione è la famiglia di fatto. Questa va intesa come un insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o da vincoli affettivi, coabitanti ed aventi dimora abituale nello stesso comune (anche se non residenti secondo l'anagrafe nello stesso domicilio).¹

Individuo campione, famiglia campione e famiglia longitudinale

Tutti gli individui appartenenti alle famiglie campione debbono essere intervistati a patto che abbiano compiuto 15 anni nell'anno di riferimento del reddito, ovvero l'anno solare precedente a quello dell'intervista.² Ogni individuo che appartiene alla famiglia campione intervistato nella prima *wave* diviene *individuo campione* e va intervistato anche nelle *wave* successive a meno che, nel frattempo, sia deceduto o si sia trasferito all'estero. La definizione di *famiglia longitudinale* deriva direttamente da quella di individuo campione. Da un punto di vista formale, l'unità di rilevazione appena definita si riferisce alla famiglia al tempo t_1 , cioè alla prima *wave* dell'indagine longitudinale. Dalla seconda *wave* in poi (da t_2 a t_4), le *famiglie campione* sono quelle composte da almeno un individuo campione. Dalla seconda *wave* in poi, tutti i componenti, a qualsiasi titolo, che si aggiungono alla famiglia campione sono considerati *individui coabitanti*. Tali individui continuano ad essere intervistati nelle *wave* successive solo se restano "agganciati" ad un individuo campione.

Periodicità e riferimento temporale

L'indagine viene svolta annualmente, in un periodo successivo alle dichiarazioni dei redditi in modo da dare la possibilità alle famiglie e agli individui di poter utilizzare le informazioni derivanti dalle proprie dichiarazioni fiscali.

Le notizie acquisite fanno riferimento a due periodi distinti: alcune alla data di indagine (anno t) e altre, principalmente quelle sul reddito, all'anno precedente ($t-1$).

Il paragrafo 2.2 è a cura di Stefano Falorsi, Claudia De Vitiis e Loredana Di Consiglio; i paragrafi 2.1 e 2.3 di Claudio Ceccarelli e Andrea Cutillo; il paragrafo 2.4 di Claudia Rinaldelli

¹ Nel caso dell'indagine Eu-Sile, il *personale di servizio o alla pari* e le persone che abitualmente non vivono più nella famiglia ma ne hanno fatto parte per almeno tre mesi nell'anno di riferimento del reddito vanno incluse nella rilevazione (fornendo soltanto le informazioni richieste nel registro familiare) ma non divengono individui campione e quindi non dovranno poi essere seguiti per gli altri anni del processo longitudinale. Gli ospiti sono comunque esclusi dalla rilevazione.

² In realtà Eurostat chiede di intervistare tutti gli individui che nell'anno di riferimento del reddito abbiano compiuto 16 anni. L'Italia ha scelto di intervistare anche i quindicenni per uniformare le definizioni a quelle di altre indagini sulle famiglie.

I riferimenti temporali delle notizie raccolte sono:

- *periodo dell'intervista (anno t)*. A questo vanno ricondotte le informazioni familiari e individuali che caratterizzano la condizione di vita attuali (come ad esempio, le caratteristiche dell'abitazione, il possesso dei beni durevoli, le condizioni di salute degli individui, l'istruzione, l'attuale condizione lavorativa, eccetera).
- *Ultimi dodici mesi*. A questi vanno ricondotte, ad esempio, le principali spese per l'abitazione.
- *Anno t-1*. A questo vanno ricondotte tutte le informazioni familiari ed individuali che caratterizzano la situazione economica della famiglia e degli individui (come ad esempio i mutui e i prestiti, i redditi, eccetera).

2.2 - Il disegno di rilevazione e il piano di campionamento

La progettazione del disegno campionario è stata guidata dai requisiti stabiliti da Eurostat relativamente ai diversi aspetti della rilevazione, ovvero: la tipologia di parametri che l'indagine deve produrre, la cadenza e il periodo di riferimento dei quesiti, nonché la precisione di alcune stime obiettivo di tipo *trasversale* e *longitudinale*. Oltre ai requisiti Eurostat, si sono aggiunte le esigenze di informazione per livelli territoriali sub-nazionali di specifico interesse dell'Istat.

La natura dei parametri da stimare rende necessario che la rilevazione sia costituita da una componente trasversale e una longitudinale. Più nello specifico, la natura longitudinale di alcuni parametri e i requisiti relativi alla cadenza e al periodo di riferimento, impongono la necessità di progettare panel di durata minima di quattro anni. L'articolazione delle indagini trasversale e longitudinale non rappresenta, tuttavia, un vincolo del Regolamento, anche se tale tema ha rappresentato oggetto di dibattito nell'ambito della *task-force* per la progettazione dell'indagine nell'ambito del quale vengono suggerite alcune ipotesi alternative adeguate ai differenti sistemi informativi di ciascun Paese.

Per la comparabilità delle stime dei diversi paesi in termini di precisione, Eurostat impone una numerosità campionaria minima sotto l'ipotesi di campionamento casuale semplice; la definizione della numerosità campionaria da realizzare, sulla base della quale viene pianificata l'indagine, deriva poi dalle ipotesi sul *design effect* connesso con il disegno di campionamento attuato dai diversi Istituti, nonché dai tassi di risposta attesi per la rilevazione. Nel caso di un'indagine come quella Eu-Silc, in cui la rilevazione ha anche natura longitudinale, la valutazione dei tassi di risposta richiede, inoltre, la specificazione di un andamento dell'*attrition*.³

2.2.1 Gli obiettivi conoscitivi

Dal momento che Eu-Silc è un'indagine ripetuta a cadenza annuale con la finalità di stimare sia parametri di carattere trasversale che longitudinale, la popolazione d'interesse appena definita, la natura di popolazione trasversale e longitudinale deve essere ulteriormente precisata.

La definizione di popolazione trasversale non pone particolari problemi, in quanto comunemente ci si riferisce alle unità esistenti ad una prefissata data di ciascun anno di rilevazione. Relativamente, invece, alla popolazione longitudinale, che è riferita a un intervallo di tempo, la definizione è molto più complessa, in quanto è necessario tenere in considerazione gli aspetti dinamici della popolazione, ossia tutti i fenomeni di entrata e uscita di unità

³ Negli studi longitudinali per i quali le informazioni sono raccolte ripetutamente nel tempo, un problema comune è l'*attrition*, ossia l'assenza di risposta dell'unità campionaria a partire da una occasione di indagine.

nell'intervallo di tempo considerato. Poiché non pienamente osservabili, tali popolazioni sono individuabili a partire dalle unità appartenenti a popolazioni trasversali relative a tempi prefissati, quali, ad esempio, l'istante iniziale e finale dell'intervallo considerato, oppure un istante intermedio.

Le variabili e i parametri d'interesse

L'indagine Eu-Silc deve fornire sia dati trasversali su reddito, povertà e esclusione sociale e condizioni di vita, sia dati longitudinali su reddito, lavoro e su alcuni indicatori non-monetari di esclusione sociale, tra cui anche indicatori legati alle condizioni di salute e sul ricorso ai servizi sanitari.

Indicando con t la generica occasione di indagine, i parametri obiettivo relativi alle variabili d'interesse possono essere classificati nelle seguenti tipologie (Duncan et al., 1989):

- a. parametri della popolazione a una data occasione di indagine t ;
- b. valore medio di parametri della popolazione in un dato intervallo temporale $(t, t+k)$ per $k \geq 1$;
- c. variazioni di parametri a livello aggregato o variazioni nette (*net change*);
- d. flussi o variazioni lorde (*gross change*);
- e. frequenze, ricorrenze e durate medie di eventi che si verificano in un determinato intervallo di tempo.

I parametri (a), (b), (c) sono parametri di tipo trasversale, ossia parametri di livello riferiti a un dato istante oppure ottenuti come media o differenza di parametri di livello riferiti a più occasioni di indagine. I parametri (d) ed (e) sono, invece, parametri di tipo longitudinale che richiedono per la loro costruzione, l'osservazione su ciascuna unità delle variabili d'interesse in più istanti di tempo dell'intervallo.

In particolare, al fine di rispondere ai requisiti Eurostat riguardanti le stime che Eu-Silc deve produrre, la progettazione del disegno è stata principalmente guidata dai seguenti parametri d'interesse:

- ad ogni occasione di indagine t (parametri di tipo a):
 1. la percentuale di individui poveri, ossia la percentuale di individui il cui reddito familiare *equivalente* è inferiore al 60 per cento del valore mediano della distribuzione nazionale del reddito;
 2. il reddito medio annuo familiare;
- tra due occasioni di indagine consecutive t e $t+1$:
 1. la variazione netta della percentuale di famiglie povere (parametro di tipo c);
 2. la variazione netta del reddito familiare medio (parametro di tipo c);
 3. la percentuale di individui che permangono nello stato di povertà per due o più occasioni di indagine (parametro di tipo e).

Ai parametri precedenti, occorre aggiungere gli obiettivi di interesse specifici dell'Istat, tra i quali ricordiamo le variazioni lorde tra due occasioni di indagine relativamente allo stato occupazionale (ad esempio, il numero di individui che passano dallo stato di occupato allo stato di disoccupato) e alla classe di reddito (stime già fornite con l'indagine *European community household panel*).

Poiché Eurostat ha la finalità di porre a confronto le condizioni di vita dei diversi paesi, l'unico dominio territoriale di studio relativo alle finalità europee è l'intero territorio nazionale. Tuttavia, per le finalità nazionali si è aggiunta l'esigenza di produrre stime anche a livello di ripartizione geografica e regione.

2.2.2 Il disegno di rilevazione

2.2.2.1 Articolazione delle indagini trasversale e longitudinale

Nella fase di progettazione del disegno di indagine è stato necessario tenere conto del seguente insieme di requisiti indicati da Eurostat:

- i. è richiesto che l'indagine presenti una componente trasversale e una componente longitudinale, pur assumendo maggiore importanza la componente trasversale, in quanto la priorità dell'indagine è fornire stime comparabili, tempestive e di alta qualità a livello trasversale;
- ii. le due componenti devono avere cadenza annuale, il periodo di riferimento dei quesiti riferiti al reddito deve essere l'anno solare precedente il momento dell'intervista, mentre per gli altri quesiti tale periodo di riferimento può essere variabile;
- iii. la componente longitudinale deve rilevare informazioni riferite a un arco temporale di durata di almeno quattro anni;
- iv. devono essere prodotti archivi di dati aggiornati annualmente, con le seguenti caratteristiche:
 - gli archivi di dati trasversali, che si riferiscono sia alle famiglie che agli individui che le compongono, devono contenere per ciascuna unità sia le variabili di tipo economico che quelle di tipo sociale;
 - gli archivi di dati longitudinali, riferiti agli individui, devono riportare, per ciascun individuo rilevato, i valori delle variabili di interesse osservati in almeno quattro occasioni di indagine (quella corrente e le tre precedenti) in cui detto individuo è stato intervistato;
- v. la componente trasversale ha la finalità di fornire stime riferite alle famiglie e agli individui che le compongono con riferimento a:
 - parametri di livello (tipo a) con cadenza annuale;
 - variazioni nette (parametri di tipo c), con principale interesse per le variazioni ad un anno di distanza;
- vi. la componente longitudinale ha lo scopo di produrre stime riferite agli individui con riferimento a:
 - flussi (parametri del tipo d) e durate medie (parametri di tipo e) per quanto riguarda la dinamica dei processi legati all'esclusione sociale e alla povertà;
- vii. la numerosità campionaria (in termini di unità finali) di ciascuna delle due componenti deve rispettare certi livelli minimi già stabiliti (approfondimenti nel paragrafo 2.2.4).

Ciò premesso, il primo aspetto importante per la specificazione della rilevazione è il fatto che l'interesse di Eu-Silc riguarda variabili di natura differente, ossia il reddito e le condizioni socio-economiche. Nella discussione della *task-force* si è ipotizzato che i dati di queste variabili possano provenire da fonti differenti, purché tra loro abbinabili. Ad esempio, la rilevazione sui redditi potrebbe provenire da archivi amministrativi qualora disponibili, mentre un'indagine campionaria potrebbe essere effettuata solo per le variabili del secondo tipo. L'utilizzo di archivi amministrativi è particolarmente indicato per i paesi con popolazione ridotta, nei quali sono disponibili archivi tempestivamente aggiornati e affidabili. Per quanto riguarda l'Italia, la scelta è stata quella di organizzare un'unica indagine per la rilevazione di tutte le variabili richieste da Eurostat.

Un altro aspetto rilevante è stato quello relativo alla scelta di un disegno di indagine che fosse in grado di produrre stime, corrette ed efficienti, sia dei parametri trasversali sia dei parametri longitudinali ricordati nel paragrafo precedente, prevedendo, pertanto, una componente trasversale e una componente longitudinale. Anche per quanto riguarda la relazione

tra le due componenti dell'indagine, Eurostat non obbligava ad un'unica soluzione da applicare in ogni contesto nazionale; in tal modo, nei casi in cui già esistessero indagini distinte e indipendenti in grado di fornire le informazioni richieste, non è stato necessario impiantare nuove rilevazioni.

Data l'assenza di vincoli, era, pertanto, possibile adottare uno dei due seguenti disegni alternativi:

- i. due rilevazioni separate: una prima rilevazione trasversale a campioni indipendenti o basata sul campionamento ruotato (l'opzione con campionamento ruotato era maggiormente auspicabile per migliorare la precisione delle stime di variazione netta), una seconda rilevazione longitudinale basata su un panel ruotato con o senza sovrapposizione;
- ii. un'unica rilevazione integrata, del tipo dell'indagine *Survey on Income and Program Participation* (SIPP), basata su un *panel* ruotato annualmente con sovrapposizione, mediante la quale ottenere sia le stime di natura trasversale sia quelle di tipo longitudinale.

La scelta italiana è stata di utilizzare un'indagine integrata per la componente longitudinale e trasversale, anche per ridurre i costi complessivi della rilevazione.

2.2.2.2 *Caratteristiche dell'indagine integrata e schema di rotazione*

E' stato definito un disegno di rilevazione di tipo *panel*, in base al quale le informazioni relative alle variabili di interesse sono raccolte sulle medesime unità campione in tempi differenti, secondo una logica di tipo *prospettico*. Si è scelto di utilizzare un *panel ripetuto* costituito da una serie di panel, ciascuno dei quali di durata fissa e sovrapposti tra loro, nel senso che due o più panel coprono parte dello stesso periodo temporale.

Questo schema è equivalente al *campionamento ruotato* (Duncan et al, 1989), in quanto, in entrambi, ciascun panel ha durata limitata nel tempo e due o più panel vengono seguiti nello stesso periodo di tempo. Tuttavia, i due disegni nascono da esigenze differenti: il campionamento ruotato è ampiamente usato per produrre stime di livello e stime di variazioni nette nelle indagini trasversali, mentre il panel ripetuto con sovrapposizione risponde all'obiettivo di ottenere stime di natura longitudinale. Nel caso dell'indagine Eu-Silc, che persegue entrambi gli obiettivi, la scelta di questo tipo di disegno è risultata la più ragionevole.

Da un punto di vista generale, gli aspetti da considerare nella definizione di uno schema di rotazione riguardano:

- la durata di ciascun panel;
- il numero di occasioni per ogni panel;
- l'introduzione di nuovi panel.

Nel caso specifico, la scelta da effettuare era limitata alla durata del panel; infatti, poiché è necessario effettuare una rilevazione ogni anno, il numero di occasioni coincide con la durata e l'introduzione dei nuovi panel deve essere annuale per mantenere invariata nel tempo la dimensione e la rappresentatività del campione.

Per la definizione della durata di un panel si noti che generalmente tanto più questa è lunga, tanto più è elevata la ricchezza di informazioni utilizzabili ai fini dell'analisi statistica. E' necessario, inoltre, tener conto dell'esistenza di una forte relazione tra la scelta effettuata per la definizione della durata del panel e la qualità delle stime che l'indagine produce. Infatti, ad esempio, quanto più elevata è la durata del panel, tanto maggiore è il numero di unità campionarie che – durante la vita dello stesso – passano dallo stato "*disoccupato*" allo stato "*occupato*" e, quindi, tanto più elevato è il livello di precisione della stima di flusso relativa alle unità che nella popolazione presentano il suddetto cambiamento di "*stato*". In ogni caso quando

si stabilisce la lunghezza di un panel, occorre tenere in considerazione anche il rischio di introdurre errori di tipo non campionario nelle stime. Tra tali errori i più rilevanti sono: l'*attrition*, il *condizionamento da panel* e l'*incompletezza delle transizioni e delle durate di eventi*. Dall'analisi di importanti indagini internazionali emerge che l'*attrition*, generalmente di non trascurabile entità, è sostanzialmente funzione del numero di occasioni (e, quindi nel caso specifico di Eu-Silc, funzione della durata del panel), mentre gli effetti dovuti al condizionamento da panel sono meno rilevanti.

Al fine di illustrare come si articola concretamente il campione, di seguito vengono illustrati due schemi di rotazione relativi rispettivamente a un panel di quattro anni (la durata minima necessaria dati i requisiti di Eurostat) e un panel di sei anni. Lo schema riportato nel prospetto 2.1 mostra come il campione relativo a ogni occasione d'indagine è costituito da quattro gruppi di rotazione (ognuno di dimensione pari a un quarto della numerosità campionaria complessiva n), ciascuno dei quali rimane nel campione per quattro anni consecutivi. Ad esempio, con riferimento all'occasione T+3, le unità del gruppo di rotazione D4 vengono intervistate per la quarta volta, le unità di E3 per la terza, le unità di F2 per la seconda e le unità di G1 entrano nel campione dell'anno T+3 per la prima volta. Per il panel di sei anni, illustrato nel prospetto 2.2, sono presenti sei gruppi di rotazione (ognuno di numerosità campionaria pari a un sesto del campione complessivo) che seguono uno schema analogo.

Prospetto 2.1 - Schema di rotazione del panel di 4 anni

	A	B	C	D	E	F	G	H	I
T	A4	B3	C2	D1					
T+1		B4	C3	D2	E1				
T+2			C4	D3	E2	F1			
T+3				D4	E3	F2	G1		
T+4					E4	F3	G2	H1	
T+5						F4	G3	H2	I1

Prospetto 2.2 - Schema di rotazione del panel di 6 anni

	A	B	C	D	E	F	G	H	I
T	A6	B5	C4	D3	E2	F1			
T+1		B6	C5	D4	E3	F2	G1		
T+2			C6	D5	E4	F3	G2	H1	
T+3				D6	E5	F4	G3	H2	I1
T+4					E6	F5	G4	H3	I2
T+5						F6	G5	H4	I3

Al fine di illustrare le differenti numerosità campionarie su cui sono basate le stime longitudinali riferite a periodi di due, tre o quattro anni, nei prospetti 2.3 e 2.4 sono evidenziate le sovrapposizioni e le rispettive numerosità campionarie a due, tre e quattro anni per i due schemi considerati.

Per il panel di durata pari a quattro anni, la sovrapposizione del campione in due anni consecutivi è pari a $3/4$ della numerosità campionaria complessiva n ; ad esempio, per i due anni T+1 e T+2 i gruppi di rotazione sovrapposti (evidenziati nello schema) sono C3, C4, D2, D3, E1 ed E2. La parte di campione sovrapposto a tre anni e a quattro anni è di numerosità rispettivamente pari a $2/4$ di n e $1/4$ di n . Analogamente nel panel di sei anni la dimensione dei campioni sovrapposti a due, tre e quattro anni è rispettivamente pari a $5/6$, $4/6$ e $3/6$ di n . Risulta

quindi evidente che la sovrapposizione, e quindi la dimensione campionaria su cui sono basate le stime longitudinali, è tanto maggiore quanto più lunga è la durata del panel; ad esempio, relativamente ad un periodo di quattro anni, la dimensione del campione sovrapposto risulta doppia nel panel di sei anni rispetto a quello di quattro anni. A tale circostanza si aggiunge il fatto che, in generale, con panel di durata maggiore si ottiene una maggiore precisione delle stime di longitudinali anche se, come già osservato, il panel di quattro anni consente di tenere maggiormente sotto controllo gli errori non campionari. In seguito a tali considerazioni oltre che ad aspetti di natura operativa, si è scelta una durata del panel di quattro anni, operando sulla definizione della numerosità campionaria n al fine di garantire la precisione richiesta delle stime longitudinali.

Prospetto 2.3 - Panel di quattro anni

	A	B	C	D	E	F	G	H	I
SOVRAPPOSIZIONE A DUE ANNI									
T	A4	B3	C2	D1					
T+1		B4	C3	D2	E1				
T+2			C4	D3	E2	F1			
T+3				D4	E3	F2	G1		
T+4					E4	F3	G2	H1	
T+5						F4	G3	H2	I1
SOVRAPPOSIZIONE A TRE ANNI									
T	A4	B3	C2	D1					
T+1		B4	C3	D2	E1				
T+2			C4	D3	E2	F1			
T+3				D4	E3	F2	G1		
T+4					E4	F3	G2	H1	
T+5						F4	G3	H2	I1
SOVRAPPOSIZIONE A QUATTRO ANNI									
T	A4	B3	C2	D1					
T+1		B4	C3	D2	E1				
T+2			C4	D3	E2	F1			
T+3				D4	E3	F2	G1		
T+4					E4	F3	G2	H1	
T+5						F4	G3	H2	I1

Prospetto 2.4 - Panel di sei anni

	A	B	C	D	E	F	G	H	I
SOVRAPPOSIZIONE A DUE ANNI									
T	A6	B5	C4	D3	E2	F1			
T+1		B6	C5	D4	E3	F2	G1		
T+2			C6	D5	E4	F3	G2	H1	
T+3				D6	E5	F4	G3	H2	I1
T+4					E6	F5	G4	H3	I2
T+5						F6	G5	H4	I3
SOVRAPPOSIZIONE A TRE ANNI									
T	A6	B5	C4	D3	E2	F1			
T+1		B6	C5	D4	E3	F2	G1		
T+2			C6	D5	E4	F3	G2	H1	
T+3				D6	E5	F4	G3	H2	I1
T+4					E6	F5	G4	H3	I2
T+5						F6	G5	H4	I3
SOVRAPPOSIZIONE A QUATTRO ANNI									
T	A6	B5	C4	D3	E2	F1			
T+1		B6	C5	D4	E3	F2	G1		
T+2			C6	D5	E4	F3	G2	H1	
T+3				D6	E5	F4	G3	H2	I1
T+4					E6	F5	G4	H3	I2
T+5						F6	G5	H4	I3

2.2.3 Il piano campionario

In questo paragrafo viene descritto il piano campionario del singolo panel. Poiché l'indagine è condotta mediante intervista diretta su un campione di famiglie selezionato dalle anagrafi comunali, il disegno di campionamento adottato segue uno schema standard a due stadi comuni-famiglie con stratificazione dei comuni in base alla dimensione demografica; lo schema, ormai consolidato, viene utilizzato per le principali indagini Istat sulle famiglie condotte mediante intervista faccia a faccia.

Vale la pena rilevare che anche le più importanti indagini panel sulle famiglie condotte da altri Paesi europei ed extra-europei si fondano su disegni di campionamento analoghi.

Il disegno di campionamento adottato nell'ambito delle regioni prevede un disegno a due stadi comuni-famiglie in cui i comuni vengono suddivisi in Ar e Nar; inoltre i comuni Nar sono stratificati in base alla dimensione demografica.

Le ragioni sottostanti la scelta di questo disegno sono determinate dall'esigenza di aumentare il livello di precisione delle stime attraverso la stratificazione dei comuni e dall'esigenza di conseguire vantaggi da un punto di vista organizzativo ed economico attraverso la selezione di un campione clusterizzato.

L'indagine integrata è basata su un campione di tipo panel ruotato, che segue lo schema riportato nel prospetto 2.1, in cui il campione relativo a ogni occasione d'indagine è costituito da quattro gruppi di rotazione, ciascuno dei quali rimane nel campione per quattro anni consecutivi.

2.2.3.1 Descrizione dello schema di selezione delle unità

Lo schema di selezione è di tipo complesso e si avvale di due differenti piani di campionamento. Nell'ambito di ognuno dei domini definiti dalla regione geografica i comuni sono suddivisi in due sottoinsiemi sulla base della popolazione residente:

- l'insieme dei comuni Auto rappresentativi (che indicheremo d'ora in avanti come comuni Ar) costituito dai comuni di maggiore dimensione demografica;
- l'insieme dei comuni Non auto rappresentativi (o Nar) costituito dai rimanenti comuni.

Nell'ambito dell'insieme dei comuni Ar, ciascun comune viene considerato come uno strato a se stante e viene adottato un disegno noto con il nome di campionamento a grappoli. Le unità primarie di campionamento sono rappresentate dalle famiglie anagrafiche, estratte in modo sistematico dall'anagrafe del comune stesso; per ogni famiglia anagrafica inclusa nel campione vengono rilevate le caratteristiche oggetto di indagine di tutti i componenti di fatto.

Nell'ambito dei comuni Nar viene adottato un disegno a due stadi con stratificazione delle unità primarie. Le unità primarie sono i comuni, le unità secondarie sono le famiglie anagrafiche, estratte sistematicamente come per i comuni Ar; per ogni famiglia anagrafica inclusa nel campione vengono rilevate le caratteristiche oggetto di indagine di tutti i componenti di fatto.

I comuni vengono selezionati con probabilità proporzionali alla loro dimensione demografica e senza reimmissione, mentre le famiglie vengono estratte con probabilità uguali e senza reimmissione.

Stratificazione e selezione delle unità campionarie

L'obiettivo della stratificazione è quello di formare gruppi (o strati) di unità caratterizzate, relativamente alle variabili oggetto d'indagine, da massima omogeneità interna agli strati e massima eterogeneità fra gli strati. Il raggiungimento di tale obiettivo si traduce in termini statistici in un guadagno nella precisione delle stime, ossia in una riduzione dell'errore campionario a parità di numerosità campionaria.

Nell'indagine in esame, i comuni vengono stratificati in base alla loro dimensione demografica e nel rispetto delle seguenti condizioni:

- autoponderazione del campione a livello regionale;
- selezione di quattro comuni campione nell'ambito di ciascuno strato definito sui comuni dell'insieme Nar;
- scelta di un numero minimo di famiglie da intervistare in ciascun comune campione; tale numero è stato posto pari a ventiquattro;
- formazione di strati aventi ampiezza approssimativamente costante in termini di popolazione residente.

Il procedimento di stratificazione, attuato all'interno di ogni regione geografica, si articola nelle seguenti fasi:

- ordinamento dei comuni del dominio in ordine decrescente secondo la loro dimensione demografica in termini di popolazione residente;
- determinazione di una soglia di popolazione per la definizione dei comuni Ar, mediante la relazione:

$${}_r\lambda = \frac{{}_r\bar{m} \cdot {}_r\delta}{{}_r f} \quad (1)$$

in cui per la generica regione geografica r si è indicato con: ${}_r\bar{m}$ il numero minimo di famiglie da intervistare in ciascun comune campione; ${}_r\delta$ il numero medio di componenti per famiglia; ${}_r f$ la frazione di campionamento;

- suddivisione di tutti i comuni nei due sottoinsiemi Ar e Nar: i comuni di dimensione superiore o uguale a ${}_r\lambda$ sono definiti come comuni Ar e i rimanenti come Nar;
- suddivisione dei comuni dell'insieme Nar in strati aventi dimensione, in termini di popolazione residente, approssimativamente costante e all'incirca pari a $4 \times {}_r\lambda$.

Effettuata la stratificazione, i comuni Ar sono inclusi con certezza nel campione; per quanto riguarda, invece, i comuni Nar, nell'ambito di ogni strato vengono estratti quattro comuni

campione con probabilità proporzionali alla dimensione demografica, mediante la procedura di selezione sistematica proposta da Madow.⁴

La selezione delle famiglie da intervistare in ogni comune campione viene effettuata dalla lista anagrafica di ciascun comune senza reimmissione e con probabilità uguali. In particolare, la tecnica di selezione è di tipo sistematico e, nell'ambito di ogni comune viene attuata attraverso le seguenti fasi:

1. vengono messi in sequenza i fogli delle famiglie dell'anagrafe del comune;
2. si calcola il passo di campionamento e_{hi} , come rapporto tra il numero delle famiglie residenti M_{hi} nel comune i dello strato h e il corrispondente numero di famiglie campione m_{hi} , $e_{hi} = M_{hi}/m_{hi}$;
3. si selezionano le m_{hi} famiglie che nella sequenza costruita al punto 1) occupano le seguenti posizioni : $1, 1+e_{hi}, 1+2e_{hi}, \dots, 1+(m_{hi}-1)e_{hi}$.

Per quanto concerne, infine, l'articolazione del campione di famiglie nei quattro gruppi di rotazione (si veda prospetto 2.1) si è proceduto come segue. Il campione di ciascun comune Ar è stato suddiviso in quattro sottocampioni di pari dimensione, ognuno assegnato casualmente a un gruppo di rotazione. Nella parte Nar , invece, l'assegnazione ai gruppi ha riguardato l'intero campione di famiglie di ogni comune campione; pertanto, ciascun comune estratto in ogni strato Nar è associato casualmente a un unico gruppo di rotazione.⁵

2.2.4 Definizione delle numerosità campionarie

2.2.4.1 La numerosità campionaria trasversale

Per quanto riguarda la definizione della numerosità campionaria per la componente trasversale, Eurostat ha fissato le numerosità minime (in termini di famiglie) dei campioni nazionali sulla base della precisione della stima della percentuale di famiglie povere a livello nazionale, nell'ipotesi di campionamento casuale semplice.

La definizione del numero minimo stabilito in generale da Eurostat è basato sui seguenti elementi:

- la percentuale di famiglie povere nei paesi EU varia nell'intervallo 5-25 per cento ed è stato pertanto considerato come valore di riferimento una percentuale intermedia pari a $p=15$ per cento;
- si è supposto un effetto del disegno pari a 1, ossia si è ipotizzato un disegno di campionamento casuale semplice di unità finali;
- si è definita la precisione della stima prefissando l'ampiezza dell'intervallo di confidenza della stima di p pari a 2 punti percentuali; a tale ampiezza corrisponde l'intervallo (14-16 per cento) e un errore relativo del 3,4 per cento.

In base a questi elementi, per il campione trasversale di ogni anno la numerosità minima risultante è di 5 mila famiglie. Tale valore è stato adattato tenendo conto del peso demografico dei vari paesi partecipanti all'indagine e, per l'Italia, è risultato pari a 7.250 famiglie; con quest'ultima numerosità l'errore relativo della stima di p risulta pari a 2,8 per cento.

Per quanto riguarda la precisione della stima del reddito medio familiare, che costituisce un altro importante elemento guida della progettazione di Eu-Silc, Eurostat ha stimato, sulla base

⁴ Madow, W.G. "On the theory of systematic sampling II", *Annals of Mathematical Statistics*, 20, (1949): 333-354.

⁵ La stratificazione ha prodotto alcuni strati Nar con un numero di comuni inferiore a quattro. In questi casi, si è optato per la scelta di due comuni campione, invece che quattro; il campione di famiglie estratto, in ciascuno di questi comuni, è stato suddiviso in due sottocampioni di pari dimensione, ognuno assegnato casualmente a un gruppo di rotazione.

dei dati dell' Echnp, che per l'Italia un campione di 7.250 famiglie produce una stima con un errore campionario atteso approssimativamente del 1 per cento.

Ogni paese ha dovuto stabilire la numerosità campionaria effettivamente da selezionare sulla base della valutazione dell'effetto del disegno (*deff*) relativo al disegno di campionamento prescelto, tenendo in considerazione l'impatto sull'efficienza delle stime prodotto dalla stratificazione, dal clustering e dalla ponderazione, e dei tassi attesi di mancata risposta totale.

Sia n_{eff} la numerosità campionaria minima o effettiva e sia n_{ach} la numerosità campionaria necessaria per garantire la precisione prefissata delle stime ottenuta tenendo conto del disegno di campionamento adottato; tra le due quantità sussiste la seguente relazione

$$n_{ach} = n_{eff} \cdot deff \quad (2)$$

in cui *deff* è il rapporto tra la varianza della stima basata sul disegno campionario effettivamente utilizzato e la varianza della stima basata su un ipotetico campione casuale semplice di uguale numerosità in termini di unità finali.

Per determinare il numero di famiglie campione da selezionare, n_{sel} , è necessario tenere conto, infine, del tasso di risposta atteso *R*; risulta quindi

$$n_{sel} = n_{ach} / R \quad (3)$$

Per la definizione finale della numerosità campionaria, ai vincoli derivanti dai requisiti Eurostat, è stato necessario aggiungere i vincoli sugli errori campionari delle principali stime di interesse a livello di diversi domini territoriali di stima (ripartizioni e regioni), stabiliti all'interno dall'Istat. La numerosità campionaria complessiva e la sua allocazione tra i diversi domini di stima è stata stabilita sulla base di una procedura basata sulla metodologia proposta nel lavoro di Bethel (1989), generalizzata per tenere conto sia della compresenza di più tipologie di domini territoriali di stima (Falorsi et al, 1998), sia della complessità del disegno campionario (Falorsi e Russo, 2001). Tale procedura ha consentito di definire la dimensione campionaria totale e l'allocazione tra i diversi domini territoriali sulla base dei vincoli imposti sugli errori campionari delle due principali stime di interesse, ossia la stima del numero di famiglie povere e la stima del reddito medio familiare. Le informazioni utilizzate a tale scopo sono state ottenute a partire dai dati dell'indagine Echnp, relativi a più anni di indagine. I vincoli sugli errori campionari delle stime sono stati imposti a livello nazionale e di ripartizione.

La valutazione dell'effetto del disegno, necessaria per lo studio della numerosità campionaria e dell'allocazione, è stata effettuata a livello di ripartizione geografica e dipende essenzialmente dai seguenti parametri

- la numerosità del campione di individui di ogni ripartizione n_{ach} ;
- il numero medio di interviste per comune \bar{m} ;
- la suddivisione della popolazione e del campione in A_r e N_{AR} , entro le ripartizioni.

L'espressione utilizzata per il calcolo del *deff* è la seguente

$$deff = \frac{n_{ach}}{N^2} \left\{ \frac{N_{AR}^2}{n_{ach,AR}} (1 + \rho_{AR}(\delta - 1)) + \frac{N_{NAR}^2}{n_{ach,NAR}} (1 + \rho_{NAR}(\bar{m} - 1)) \right\} \quad (4)$$

Il calcolo del *deff* è basato su un processo iterativo, essendo presenti nell'espressione precedente quantità che a loro volta dipendono dall'effetto del disegno.

La valutazione del *deff*, richiede inoltre la stima dei coefficienti di correlazione intraclassa ρ_{AR} e ρ_{NAR} che è stata effettuata sulla base dei dati dell' Echnp a livello di ripartizione.

La numerosità a livello di ripartizione, definita tenendo conto dell'effetto del disegno, è stata allocata tra le regioni seguendo un'ottica di compromesso tra un'allocazione uniforme e un'allocazione proporzionale alla popolazione delle regioni, tale da garantire sia l'affidabilità prefissata delle stime a livello nazionale e ripartizionale, sia quella delle stime a livello regionale.

Nel prospetto 2.5 sono riportate le numerosità campionarie e gli errori relativi attesi delle due stime di riferimento a livello di ripartizione geografica e nazionale.

Prospetto 2.5 - Allocazione del campione per ripartizione ed errori relativi delle stime

RIPARTIZIONE	Famiglie Popolazione	Famiglie Campione	Errore relativo % Reddito Medio	Errore relativo % Incidenza Povertà
Nord	10.583.085	12.513	1,5	4,3
Centro	4.226.377	6.320	1,7	4,3
Sud	7.197.453	6.668	2,2	2,8
Italia	22.006.915	25.501	1,1	2,1

Una volta fissata la numerosità regionale n_{ach} , per definire la numerosità del campione da selezionare è stato necessario valutare il tasso di risposta. I dati più recenti a disposizione erano relativi all'Indagine sui redditi e condizioni di vita condotta nel 2003; tali dati sono stati utilizzati per stimare il tasso di risposta separatamente per ripartizione e tre classi di dimensione comunale.

2.2.4.2 La numerosità campionaria longitudinale

La numerosità longitudinale è stata definita da Eurostat come il numero di famiglie intervistate con successo in ogni anno e di cui la maggior parte dei membri sono intervistati anche nell'anno successivo. Per la definizione di tale numerosità si è proceduto analogamente alla definizione della numerosità trasversale.

E' stata prefissata la precisione della stima del tasso di persistenza nella condizione di povertà, ossia la proporzione di popolazione che rimane in povertà per due o più anni. Considerando un valore di riferimento del parametro pari al 10 per cento, e fissando un'ampiezza dell'intervallo di confidenza della stima pari a 2 punti percentuali, corrispondente ad un errore relativo pari al 4,7 per cento, si ottiene una numerosità minima pari a 4mila famiglie. Tenendo, infine, conto della dimensione demografica nazionale, da tale numerosità discende una numerosità campionaria minima per l'Italia pari a 5.500 famiglie. Questo valore corrisponde approssimativamente al 75 per cento della numerosità campionaria definita per la componente trasversale. Si ricorda che lo schema di rotazione dell'indagine integrata, basata su panel ruotati quadriennali, garantisce tale condizione.

2.3 - Strategia di ponderazione per le stime trasversali

L'indagine, nella sua componente trasversale, deve produrre sia le stime riferite al numero di individui (o famiglie) che nella popolazione di riferimento possiedono una certa caratteristica, sia il livello di una quantità misurata sugli individui (o famiglie), come ad esempio il reddito.

La metodologia utilizzata, basata sull'uso di una famiglia di stimatori noti in letteratura come *calibration estimator* (stimatori di ponderazione vincolata), consente la determinazione di un unico coefficiente di riporto all'universo in grado di produrre stime coerenti a totali noti, desunti da fonti esterne, sia per individui sia per famiglia assegnando, cioè, lo stesso coefficiente di riporto all'universo a tutti gli individui della stessa famiglia (calibrazione integrata).

2.3.1 Gli elementi caratteristici del piano di campionamento

Riassumiamo brevemente per comodità espositiva le principali caratteristiche del piano di campionamento dettagliatamente illustrato nel paragrafo precedente.

Il piano di campionamento è stato progettato e realizzato in funzione delle principali stime che l'indagine deve produrre e dei domini di studio pianificati⁶. È basato su quattro campioni longitudinali indipendenti. Tali campioni sono sfasati nel tempo in modo che in ogni anno ci sia la chiusura del panel che arriva alla quarta *wave* e l'inizio di un nuovo panel. Ogni campione longitudinale è a due stadi con stratificazione delle unità di primo stadio, i comuni, mentre le unità di secondo stadio sono le famiglie estratte dalle anagrafi dei comuni campione.

La stratificazione delle unità di primo stadio, effettuata a livello regionale, è basata sulla dimensione demografica dei comuni e determina la suddivisione del territorio nazionale in 288 strati. All'interno di ciascuna regione sono presenti tre tipologie di strato:

- gli strati Ar (autorappresentativi), dove è presente un solo comune che entra quindi automaticamente in tutte e quattro i campioni longitudinali;
- gli strati Nar (non autorappresentativi) del primo tipo, che raggruppano pochi comuni di grandi dimensioni; da ognuno di questi strati sono estratti due comuni campione, ognuno dei quali è presente in due dei quattro campioni longitudinali;⁷
- gli strati Nar del secondo tipo, che raggruppano i comuni di minore dimensione demografica; da ogni strato di questo tipo sono estratti quattro comuni campione, ognuno dei quali è presente in un solo campione longitudinale.

Nel 2004, i primi quattro campioni longitudinali (contrassegnati nel prospetto 2.6 con C1, C2, C3 e C4) hanno partecipato all'indagine tutti per la prima volta. Per iniziare la rotazione, si è ipotizzato che il campione C1 fosse alla quarta, e quindi ultima, *wave* (w(4) nel prospetto 2.6), il campione C2 fosse alla terza *wave* (w(3)), il campione C3 alla seconda (w(2)). Quello indicato con C4 è il primo campione longitudinale che, iniziato nel 2004, proseguirà correttamente per quattro *wave*, come da disegno, e consentirà la realizzazione del primo file longitudinale completo (composto da w1, w2, w3, w4). Nel 2005, C5 è il nuovo campione longitudinale che entra nell'indagine e ha preso il posto di C1 terminato nel 2004. In genere, un nuovo campione

Prospetto 2.6 - Schema di campionamento adottato dall'indagine Eu-Silc

CAMPIONI	Anni						
	2004	2005	2006	2007	2008	2009	2010
C1	w(4)						
C2	w(3)	w(4)					
C3	w(2)	w(3)	w(4)				
C4	w1	w2	w3	w4			
C5		w1	w2	w3	w4		
C6			w1	w2	w3	w4	
C7				w1	w2	w3	w4

⁶ In particolare le principali tipologie di stime sono riferite alla percentuale di individui poveri e il reddito medio annuo familiare netto, per la componente *cross-section* e alla variazione della percentuale di famiglie povere e del reddito familiare medio e la percentuale di famiglie che permangono nello stato di povertà per due o più occasioni di indagine, per la componente longitudinale. I domini pianificati per l'indagine Eu-Silc rispetto ai quali sono calcolati i parametri relativi alla popolazione di riferimento sono: l'intero territorio nazionale, le cinque ripartizioni territoriali, le regioni (eccezion fatta per il Trentino-Alto Adige per il quale le province di Trento e Bolzano sono trattate separatamente).

⁷ All'interno di ogni strato Nar del primo tipo, ognuno dei due comuni estratti ha due campioni longitudinali. Ad esempio, nel 2004, il primo comune ha avuto i campioni C1 e C3 mentre l'altro i campioni C2 e C4; nel 2005, il primo comune ha avuto i campioni C3 e C5 mentre il secondo sempre i campioni C2 e C4.

longitudinale è composto dalle stesse unità di primo stadio⁸ (comuni) e da nuove unità di secondo stadio (famiglie).

Il campione *trasversale* è composto dall'unione dei quattro campioni longitudinali, ognuno per la sua specifica *wave*: in tal modo, ogni campione *trasversale* è composto da un quarto di famiglie che partecipano per la prima volta all'indagine, un quarto di famiglie che partecipano per la seconda volta, un quarto di famiglie che partecipano per la terza volta e infine un quarto di famiglie che partecipano per la quarta volta all'indagine.

Il campione trasversale d'avvio, relativo all'anno 2004, è composto complessivamente da circa 32mila famiglie, circa 8mila per ogni campione longitudinale.

Per l'anno 2005, la numerosità campionaria del campione trasversale si determina come somma delle seguenti quantità:

- numero di famiglie con individui campione rispondenti nella prima wave per i 3 campioni longitudinali (C2, C3 e C4);⁹
- 8mila famiglie nuove estratte appartenenti al nuovo campione longitudinale (C5).¹⁰

2.3.2 Le stime trasversali

La strategia adottata per la costruzione dei coefficienti di riporto all'universo si sviluppa attraverso le fasi tipiche utilizzate per la costruzione degli stimatori nelle varie indagini campionarie dell'Istituto. In particolare possiamo distinguere:

- 1) la determinazione della probabilità di inclusione di ogni unità statistica e del relativo peso base;
- 2) calcolo dei coefficienti di correzione per mancata risposta totale;
- 3) determinazione dei coefficienti di riporto all'universo finali vincolati ai totali noti desunti da fonti esterne all'indagine.

Come precedentemente accennato, per il calcolo dei coefficienti di riporto all'universo finali si ricorre agli stimatori di ponderazione vincolata che coincidono asintoticamente con lo stimatore di regressione generalizzato. Per campioni sufficientemente grandi, quindi, tali stimatori hanno approssimativamente le stesse proprietà, ovvero sono corretti, consistenti e con la stessa varianza campionaria. Tali stimatori sono utilizzati correntemente nelle indagini campionarie dell'Istituto.¹¹

La probabilità di inclusione e il peso base

Il principio su cui è basato ogni metodo di stima campionaria è che le unità appartenenti al campione rappresentino anche le unità della popolazione di riferimento che non sono incluse nel campione stesso. A tale scopo, ad ogni unità campionaria viene attribuito un peso, o coefficiente di riporto all'universo, che indica quante unità della popolazione sono rappresentate, rispettivamente, da ogni unità presente nel campione.

Senza perdere di generalità e per chiarire gli argomenti trattati, definiamo la seguente simbologia:

- U popolazione di riferimento oggetto di indagine;
- y_k valore della variabile Y assunto dalla k -esima osservazione della popolazione;

⁸ I comuni Nar del secondo tipo, quelli associati ad un solo campione longitudinale, possono essere sostituiti, per problemi organizzativi, quando il panel a cui appartengono è finito e riparte con nuove famiglie.

⁹ In questa maniera fanno parte del campione anche famiglie non estratte alla prima *wave* ma che abbiano al proprio interno un individuo campione che, uscito dalla famiglia di origine, abbia creato o sia confluito in una altra famiglia.

¹⁰ La procedura per la determinazione del campione negli anni successivi è analoga.

¹¹ Il software prodotto dall'Istat, denominato Genesees, richiama la metodologia illustrata da Deville, J.C. e Särndal, C.E. in *Calibration Estimation in Survey Sampling*, Journal of the American Statistical Association, Vol. 87, n.418, 1992.

π_k probabilità, assegnata da un generico disegno di campionamento, che l'unità k-esima ha di essere inclusa nel campione s .

Il totale di una generica variabile Y , calcolato sull'intera popolazione, assume la seguente forma:

$$Y = \sum_{k \in U} y_k \quad (5)$$

In linea teorica il disegno di campionamento assegna le probabilità di inclusione ad ogni unità del campione in modo tale che

$$\hat{Y} = \sum_{j \in s} y_j \frac{1}{\pi_j} \quad (6)$$

sia uno stimatore corretto della (5).

Nel caso del campione di Eu-Silc, come nel caso dei disegni di campionamento tradizionalmente utilizzati dall'Istat per le principali indagini su famiglie e individui, la probabilità di inclusione assume la seguente forma:

$$\frac{1}{\pi_j} = n_h \frac{P_h}{P_{hi}} \frac{M_{hi}}{m_{hi}} \quad (7)$$

dove :

h denota l'indice di strato;

i è l'indice di comune;

n_h indica il numero di comuni campione dello strato h ;

P_h indica il totale della popolazione residente nello strato h ;

P_{hi} il totale della popolazione residente nel comune i dello strato h ;

M_{hi} indica il totale di famiglie residenti nel comune i dello strato h ;

m_{hi} indica il numero di famiglie campione nel comune i dello strato h .

In tal senso, ogni famiglia estratta da un certo comune campione si vede assegnata dal disegno di campionamento la medesima probabilità di inclusione.¹² Pertanto la (6) può essere scritta come:

$$\hat{Y} = \sum_{h=1}^H \sum_{i=1}^{m_h} y_{hi} \frac{1}{\pi_{hi}} \quad (8)$$

La correzione per mancata risposta

Nel corso della fase di raccolta delle informazioni presso le famiglie che formano il campione, come accade per tutte le indagini statistiche, alcune di queste si trovano nell'impossibilità di partecipare all'indagine. Quando il meccanismo che genera la mancata partecipazione, e quindi la mancata risposta, è *ignorabile*, si assume l'ipotesi che il comportamento dei rispondenti sia del tutto simile a quello dei non rispondenti con conseguente riduzione casuale della numerosità del campione teorico di partenza e riduzione della precisione delle stime prodotte. In tale circostanza, il correttore per mancata risposta assume la forma dell'inverso della probabilità di risposta (δ_{hi}):

¹² Nel caso di Eu-Silc, in ogni campione longitudinale, il numero di comuni campione in ogni strato è pari a 1. Inoltre, l'espressione (7) diventa $1/\pi_{hi} = M_{hi}/m_{hi}$ nel caso dei comuni Ar che costituiscono uno strato a sé.

$$\frac{1}{\delta_{hi}} = \frac{m_{hi}}{m_{hi}^r} \quad (9)$$

in cui m_{hi}^r rappresenta il numero di famiglie rispondenti nel comune i dello strato h mentre il coefficiente di riporto all'universo corretto per mancata risposta risulta essere:

$$k_{hi} = \frac{1}{\pi_{hi}} \frac{1}{\delta_{hi}} = n_h \frac{P_h}{P_{hi}} \frac{M_{hi}}{m_{hi}} \frac{m_{hi}}{m_{hi}^r} = n_h \frac{P_h}{P_{hi}} \frac{M_{hi}}{m_{hi}^r} \quad (10)$$

Quando il meccanismo che genera la mancata risposta è *non ignorabile*, come nella maggior parte delle situazioni reali, l'ipotesi di uguaglianza di comportamento tra chi partecipa e chi non partecipa all'indagine viene meno ed è quindi necessario ricorrere ad altre ipotesi e strumenti in modo da poter ricondurre la problematica in un contesto di ignorabilità del meccanismo di mancata risposta.

Per l'indagine Eu-Silc si è pensato di adottare una strategia che ripercorre i criteri delle *celle di ponderazione* al fine di individuare delle sottopopolazioni dove è ipotizzabile una uguaglianza di comportamento di risposta tra coloro che hanno partecipato all'indagine e coloro che non hanno partecipato. La partizione delle famiglie campione in celle è stata ottenuta tramite segmentazione attraverso un albero di decisione basato sull'algoritmo del Chaid (*Chi-squared automatic interaction detection*, Kass, 1980) che divide il campione in sottogruppi secondo la relazione tra tasso di risposta e le variabili esplicative. La metodologia sottesa alle *celle di ponderazione* rientra nel novero della modellizzazione esplicita per la riduzione della distorsione della mancata risposta totale. In particolare, si determina la probabilità di risposta in funzione delle variabili che determinano le celle.

In Eu-Silc, utilizzando le informazioni sia da fonte anagrafica che da fonte fiscale, disponibili già in fase di definizione del campione da intervistare, è stata operata una suddivisione in celle omogenee secondo le variabili per le quali si è avuto un tasso di risposta differenziato. In particolare sono state utilizzate le seguenti informazioni:

- dimensione demografica del comune in classi;
- cittadinanza della persona di riferimento;
- regione di residenza;
- distribuzione delle famiglie per numero di componenti;
- distribuzione delle famiglie per tipologia di reddito percepito.¹³

Sia Z la partizione delle famiglie campione composta dagli elementi (celle) generati dalla segmentazione del campione. Per ogni cella z della partizione Z è stata calcolata la probabilità di risposta (η) come rapporto tra il numero di famiglie rispondenti e il numero di famiglie desunto dal campione teorico.

In simboli:

$$\eta_z = \frac{m_z^r}{m_z} \quad (11)$$

¹³ La difficoltà ad ottenere l'intervista è crescente al crescere dell'ampiezza demografica del comune di residenza e al diminuire del numero di componenti della famiglia perché spesso non si riesce proprio ad effettuare il contatto; le famiglie con persona di riferimento straniera hanno minori probabilità di partecipare all'indagine, in parte per l'alta mobilità degli stranieri sul territorio, in parte per le difficoltà o la diffidenza che possono avere persone di lingua e cultura differenti rispetto agli intervistatori. Si riscontrano inoltre notevoli differenze nella partecipazione all'indagine a seconda della regione di residenza.

L'inverso di η è il correttore per mancata risposta e ad ogni famiglia del comune i dello strato h , appartenente alla cella z viene moltiplicato il proprio peso base per tale correttore, pertanto il coefficiente di riporto all'universo corretto per mancata risposta totale diviene:

$${}_{hi}\psi_z = \frac{1}{\pi_{hi}} \frac{1}{\eta_z} = n_h \frac{P_h}{P_{hi}} \frac{M_{hi}}{m_{hi}} \frac{m_z}{m_z^r} \quad (12)$$

mentre la stima del totale espressa dalla (8) diviene:

$$\hat{Y} = \sum_{i=1}^{m^r} y_j \psi_j \quad (13)$$

dove m^r è il numero complessivo di famiglie rispondenti.

Dal processo di validazione sono emerse discrepanze tra la stima di aggregati di interesse e i dati provenienti da altre fonti usati come riferimento. Tali differenze sono dovute verosimilmente al diverso comportamento di risposta in funzione del livello e della tipologia di reddito. Tale fenomeno, già riscontrato in altre indagini sui redditi come quella condotta dalla Banca d'Italia¹⁴, porta a possibili distorsioni nella stima degli aggregati e delle distribuzioni di interesse.

La soluzione adottata per la prima *wave* di Eu-Silc, soluzione che possiamo definire transitoria, è stata quella di inserire un coefficiente di correzione ν calcolato utilizzando la metodologia degli stimatori di ponderazione vincolata e le informazioni desunte dalla Rilevazione continua sulle forze di lavoro. L'utilizzo di tali stimatori per limitare gli effetti dovuti alla mancata risposta totale rappresenta un esempio di modellizzazione implicita del processo di mancata risposta.

Nello specifico, l'informazione utilizzata è la condizione professionale della popolazione opportunamente riclassificata¹⁵ desunta dalla Rcfl del IV trimestre 2004.¹⁶ I totali noti applicati sono:

- distribuzione delle famiglie per numero di componenti;
- popolazione per sesso e classi d'età;¹⁷
- popolazione per condizione professionale.

L'introduzione dei totali noti riferiti alla popolazione per sesso e classi d'età e alla distribuzione delle famiglie per numero componenti sono serviti per dare "stabilità" ai correttori in funzione sia della costruzione delle celle di ponderazione sia della "calibrazione" successiva che porta ai coefficienti di riporto all'universo finali.¹⁸

¹⁴ Il Supplemento al Bollettino statistico della Banca d'Italia del gennaio 2006, a pagina 35 recita: "...La difficoltà a ottenere l'intervista è crescente al crescere del reddito e del titolo di studio del capofamiglia; minori difficoltà si incontrano con le famiglie con un ridotto numero di componenti, con capofamiglia pensionato o residenti in comuni di piccole dimensioni...".

¹⁵ Le modalità della condizione professionale utilizzata in questo contesto sono: 1) dirigente, quadro e impiegato; 2) operaio e assimilati; 3) imprenditore e libero professionista; 4) lavoratori in proprio, soci di cooperative e coadiuvante nell'azienda di un familiare; 5) persone con contratto di collaborazione coordinata e continuativa e prestatori d'opera occasionali; 6) persone in cerca di occupazione; 7) inattivi.

¹⁶ Il riferimento al IV trimestre 2004 della Rcfl, che coincide con il periodo di rilevazione dei dati di Eu-Silc, è perché in Eu-Silc la rilevazione dettagliata delle caratteristiche della condizione è riferita a tale periodo.

¹⁷ Le classi d'età utilizzate sono: 0-14, 15-24, 25-44, 45-64, 65 e più al netto delle convivenze istituzionali.

¹⁸ In genere, quando si operano calibrazioni successive è necessario inserire in ogni passo dei vincoli altamente correlati con i vincoli dell'ultimo passo. Trattandosi di procedure indipendenti, infatti, c'è il rischio che ciò che si intende imporre in un generico passo di calibrazione venga sostanzialmente annullato nel passo successivo. Nel nostro caso si è preferito inserire la distribuzione per sesso e classe d'età (più ampie di quelle finali) in modo da "indirizzare" la calibrazione utilizzata per correggere la mancata risposta nella direzione dei vincoli finali che, nel caso di Eu-Silc, hanno il solo obiettivo di aumentare l'accuratezza delle stime.

Con riferimento alla (10), per ogni famiglia campione del comune i appartenente allo strato h e per ogni cella di ponderazione z elemento della partizione Z , il correttore completo per mancata risposta assume la seguente forma:

$$h_i \psi_{zg}^* = h_i \psi_z \cdot v_g \quad (14)$$

dove l'indice g esprime la generica combinazione dei totali noti inseriti nel passo di calibrazione.

Merita un cenno di approfondimento il vincolo imposto sulle famiglie per numero componenti. Attualmente l'Istat non dispone di statistiche annuali provenienti da fonte anagrafica che riguardano la distribuzione delle famiglie per numero componenti. Eurostat, dal canto suo, ne suggerisce l'introduzione nel pacchetto di vincoli per la costruzione degli stimatori. Non avendo dati da fonte anagrafica, si è optato per l'inserimento della distribuzione delle famiglie per numero componenti, desunta sempre dalla Rcfl (in conformità di quanto fatto per i dati sull'occupazione), in questo passo piuttosto che nel passo finale di calibrazione proprio per la natura del dato considerato e per la transitorietà della soluzione adottata. E' infatti previsto un progetto per il rilascio annuale dei dati da fonte anagrafica sulla distribuzione delle famiglie per numero componenti.

I totali noti e i coefficienti di riporto all'universo finali

Per il calcolo dei coefficienti di riporto all'universo finali, come già evidenziato nel corso del lavoro, si adottano gli stimatori di ponderazione vincolata (*calibration estimator*) metodologia implementata in un software, denominato Genesees, utilizzato in numerose altre indagini dell'Istituto. La metodologia si basa sull'utilizzo di opportune informazioni ausiliarie, sintetizzate in totali noti, che, correlate con le variabili principali oggetto di indagine, hanno la funzione di aumentare l'accuratezza delle stime. La scelta delle informazioni ausiliarie, inoltre, è anche legata alla disponibilità periodica di tali fonti: proprio per questo Eurostat suggerisce ma non impone l'utilizzo delle informazioni da utilizzare per la costruzione delle stime finali.

In Eu-Silc, come in ogni indagine armonizzata a livello di Unione Europea, la strategia di ponderazione è "guidata" al fine di garantire la comparabilità delle stime di ogni paese membro, oltre che per l'aumento dell'accuratezza delle stime. In particolare, il documento Eu-Silc 134-rev/04 - *Cross-sectional weighting: first year each sub-sample* - suggerisce, tra l'altro, i totali noti da utilizzare per la costruzione dello stimatore di ponderazione vincolata.

Il campione è vincolato a:

- popolazione residente¹⁹ per ripartizione territoriale, sesso e classi d'età (0-15, 16-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75 e più) al 31 dicembre dell'anno di riferimento del reddito (anno t-1);
- il numero di famiglie residenti per regione riferite alla data di rilevazione (31 dicembre dell'anno t).

Per esigenze informative nazionali, le stime prodotte con il campione *trasversale* sono vincolate anche a:

- popolazione residente per regione, sesso e classi d'età (0-14, 15-24, 25-44, 45-64, 65 e più) al 31 dicembre dell'anno di riferimento del reddito (anno t-1);
- popolazione residente per ripartizione, sesso e classi d'età al 31 dicembre dell'anno di rilevazione (anno t);
- popolazione straniera maggiorenne residente per ripartizione e sesso al 31 dicembre dell'anno di riferimento del reddito (anno t-1);

¹⁹ Si intende al netto delle convivenze istituzionali.

- popolazione straniera maggiorenne residente per ripartizione e provenienza (UE e non UE) 31 dicembre dell'anno di riferimento del reddito (anno t-1).
- Popolazione residente per ripartizione e ampiezza demografica del comune all'anno di riferimento del reddito (t-1).

Il sistema di vincoli così congeniato offre la possibilità di fornire stime accurate anche a livello regionale che sono di notevole interesse e soddisfano anche importanti esigenze informative interne.

In secondo luogo, il vincolo sulla popolazione all'anno t-1 (suggerito da Eurostat) non è sufficiente a controllare la stima del numero di nati tra l'anno di riferimento del reddito (t-1) e il momento dell'effettuazione dell'indagine (anno t). Infatti, calcolando l'età rispetto alla data di riferimento del reddito, questa può assumere valore pari a “-1” in corrispondenza di tutti coloro che sono nati dopo il 31 dicembre dell'anno t-1. L'inserimento del vincolo aggiuntivo della popolazione per sesso e classe d'età riferita al momento di rilevazione (anno t), garantisce anche una certo livello di accuratezza delle stime del numero di nati desumibili da Eu-Silc. Anche se questa non è certamente la fonte ufficiale del numero di nati in un certo anno, l'informazione può essere utile per calcolare indicatori sulle condizioni socio-economiche di particolari sottoinsiemi di famiglie.

2.3.3 La valutazione della strategia di stima

Nell'indagine Eu-Silc italiana, la strategia di ponderazione si basa sulla possibilità di essere variata in funzione del tipo di mancata risposta che ogni anno può verificarsi, senza per questo stravolgere la natura e la struttura della calibrazione finale. Quest'ultima deve infatti restare quanto più possibile invariata per consentire la comparabilità delle stime prodotte sia nel tempo che nei confronti degli altri paesi europei.²⁰ Nella *wave* del 2004 è stata utilizzata una correzione per mancata risposta tramite la metodologia delle celle di ponderazione. Questa, come vedremo in seguito, non ha apportato migliorie rispetto alla “tradizionale” correzione per strato tali da giustificare l'utilizzo, probabilmente a causa della limitata disponibilità di informazioni sul campione estratto, tanto da rendere necessario un passo di calibrazione intermedio che tenesse conto di informazioni campionarie provenienti dall'indagine Rcf. Tuttavia, si sono comunque utilizzate le celle di ponderazione perché questa sarà la metodologia utilizzata dalla *wave* 2005, a partire dalla quale si potranno usare le informazioni collezionate nella *wave* precedente (per tre quarti del campione) al fine di formare delle celle più caratterizzate rispetto al processo che determina la mancata risposta.²¹ L'obiettivo è quello di determinare una partizione “ottimale” del campione tale che in ogni cella si possa assumere la casualità del meccanismo che ha generato la mancata risposta.

È il caso di evidenziare che l'intero processo di calcolo dei coefficienti di riporto, dalla correzione per mancata risposta ai due passi di calibrazione, ha riguardato separatamente il campione longitudinale C4 dagli altri. Si pone infatti la questione della coerenza tra le stime *cross-section* relative ad un certo anno e le stime ottenute dai campioni longitudinali, ed il campione C4 dell'indagine 2004 è la base di partenza per il primo campione longitudinale che arriverà a compimento all'edizione di indagine 2007. Nella seconda *wave* del suo panel, al netto della mancata risposta totale, deve quindi essere rappresentativo della popolazione iniziale al netto delle uscite da

²⁰ A meno di possibili miglioramenti, quale ad esempio la disponibilità annuale da fonte anagrafica della distribuzione delle famiglie per numero di componenti, o di problemi contingenti, quali ad esempio problemi di convergenza che richiedono l'accorpamento di alcuni vincoli.

²¹ Va anche accennata l'esistenza di un motivo tecnico: l'utilizzo dei correttori per mancata risposta tramite celle di ponderazione in Genesee ha permesso di diminuire l'errore campionario in particolari domini di stima rispetto agli errori ottenuti tramite correttori per strato.

tale popolazione. Le stime *trasversali*, ottenute dai quattro campioni insieme, devono essere così ragionevolmente coerenti con quelle ottenute dal solo campione C4: i pesi finali dell'anno di partenza rappresentano infatti la base per il calcolo dei pesi longitudinali.²²

Relativamente agli effetti della calibrazione intermedia, nel prospetto 2.7 sono confrontate le distribuzioni della popolazione di 15 anni e più ottenute per i quattro campioni longitudinali e per il campione *cross-section*, ottenute tramite la strategia adottata (metodo 3) con le distribuzioni ottenute tramite correzione per strato (metodo 1) e correzione per celle di ponderazione senza calibrazione intermedia (metodo 2), e con quella del quarto trimestre 2004 della Rcfl.²³

L'introduzione del passo di calibrazione intermedio permette di rispettare sostanzialmente in tutti e quattro i campioni longitudinali la distribuzione della popolazione di 15 anni e più per condizione professionale riclassificata. In realtà, anche le stime trasversali ottenute tramite i due diversi metodi di correzione della mancata risposta avrebbero permesso di rispettare, anche se non in maniera così forte, la distribuzione che deriva dalla Rcfl: in questi due casi, tuttavia, differenze più marcate si sono verificate considerando separatamente i quattro campioni longitudinali.²⁴ E' in effetti questa la ragione che rende opportuno l'utilizzo della calibrazione intermedia, che consente di rispettare sempre la distribuzione della popolazione per condizione professionale, distribuzione non di primario interesse per Eu-Silc ma sicuramente legata alle variabili oggetto di indagine, migliorando verosimilmente l'accuratezza delle stime.

Il coefficiente così ottenuto per la *wave* 2004 può quindi essere considerato come un coefficiente che considera la non ignorabilità del meccanismo che determina la mancata risposta, ed è il risultato congiunto di una modellizzazione esplicita (le celle di ponderazione) e di una modellizzazione implicita (la calibrazione intermedia) del processo che determina la mancata risposta.

Prospetto 2.7 - Distribuzione percentuale della popolazione di 15 anni e più per condizione nella professione, campione e metodo di ponderazione utilizzato

CONDIZIONE PROFESSIONALE	Metodo 3					Metodo 1 (campione trasversale)	Metodo 2 (campione trasversale)	Rcfl
	C1	C2	C3	C4	Campione trasversale			
Dirigenti e impiegati	16,0	16,5	16,2	16,4	16,3	16,3	16,2	16,8
Operai e assimilati	16,2	16,0	16,2	16,1	16,1	14,9	15,0	16,4
Imprenditori e liberi professionisti	3,3	3,1	3,2	3,2	3,2	3,6	3,6	3,2
Lavoratori in proprio	8,6	8,6	8,6	8,6	8,6	7,4	7,5	8,7
Co.co.co	1,0	1,1	1,1	1,0	1,1	2,1	2,1	1,0
In cerca di occupazione	4,1	4,2	4,0	4,2	4,1	4,4	4,4	4,1
Inattivi	50,8	50,5	50,7	50,5	50,6	51,2	51,2	49,8

È opportuno dare una valutazione sui diversi metodi di correzione della mancata risposta proposti: questo viene fatto sulla base del cambiamento che si ottiene tra l'insieme dei pesi diretti e quello dei pesi finali, scomponendo questa misura nel contributo introdotto nei diversi

²² Questo aspetto rappresenta parte di un problema più generale di coerenza delle stime, sia a livello micro che a livello macro, che deve essere affrontato nell'intero processo di trattamento dei dati.

²³ Si presentano le stime separatamente per tutti i campioni longitudinali per completezza di analisi, e poiché questi sono comunque estratti come campioni indipendenti tra di loro, fermo restando che la questione importante è la coerenza tra C4 e l'intero campione trasversale.

²⁴ Queste distribuzioni non sono presentate per motivi di spazio e di comprensibilità della tabella. Si consideri inoltre che i quattro campioni devono essere rappresentativi a livello di ripartizione territoriale (Nuts II level). Differenze ancora più marcate rispetto alla Rcfl si sono verificate tra le distribuzioni per ripartizione nei quattro sottocampioni.

passi. Per fare questo si è seguita l'impostazione di Dufour et al. (2001), secondo i quali la misura globale del cambiamento D si divide in quattro termini:

$$D = R_{01} + R_{12} + R_{int} + G \quad (15)$$

dove R_{01} misura i cambiamenti intercorsi nel passaggio dai pesi diretti ai pesi corretti per mancata risposta; R_{12} misura i cambiamenti nei pesi individuali che si verificano passando dai pesi intermedi ai pesi finali; R_{int} rappresenta una misura di interazione tra i due passi di cambiamento dei pesi. Il segno di questo valore indica se i due processi si stanno muovendo nella stessa direzione, quando positivo, o in direzioni opposte, quando negativo. Infine, il termine G rappresenta il cambiamento medio dovuto dalla mancata risposta totale; G può anche essere visto come la distorsione dovuta alla mancata risposta totale. Il valore assoluto di G, dovuto al livello della mancata risposta, sarà quindi uguale per ognuno dei metodi proposti.

Il prospetto 2.8 presenta il valore di D diviso per le sue componenti per ognuno dei tre metodi proposti, insieme al contributo percentuale di ciascuna componente sul totale del cambiamento. Come già detto, il metodo 1 rappresenta la correzione per mancata risposta per strato senza passo di calibrazione intermedio; il metodo 2 rappresenta la correzione tramite celle di ponderazione ma senza passo di calibrazione intermedio; il metodo 3 rappresenta l'utilizzo delle celle di ponderazione e del passo di calibrazione intermedio, ed è quindi il metodo effettivamente utilizzato per la *wave* 2004. Secondo gli autori precedentemente citati sono da preferire i metodi che migliorano la comprensione del meccanismo che determina la mancata risposta, e che aumentano quindi il contributo dovuto al primo passo (R_{01}), diminuendo al contempo il contributo percentuale della distorsione dovuta al fattore G.

Passando dalla correzione per strato a quella tramite celle di ponderazione le cose rimangono sostanzialmente invariate: diminuisce leggermente la distanza complessiva tra pesi base e pesi finali D. Il contributo dovuto alla correzione per mancata risposta passa invece dal 21,7 al 18,2 per cento mentre quello dovuto alla distorsione, misurata dal termine G, aumenta dal 22,8 al 24,5 per cento. Questo significa che le informazioni disponibili per creare le celle di ponderazione non sono allo stato sufficienti a creare una partizione ottimale del campione.²⁵ Si è scelto comunque di utilizzare questa procedura per i motivi precedentemente esposti, integrandola però con l'introduzione del primo passo di calibrazione. Introducendo questa calibrazione intermedia che tiene conto della distribuzione della popolazione per condizione professionale riclassificata e delle famiglie per numero di componenti, la misura globale del cambiamento aumenta in valore assoluto. Contrariamente a quanto si potrebbe pensare, questo non è necessariamente un aspetto negativo.

Prospetto 2.8 - Valore medio di D per metodo utilizzato per componente e contributo percentuale alla misura del cambiamento

METODO	D	R_{01}	R_{01}/D (%)	R_{12}	R_{12}/D (%)	R_{int}	R_{int}/D (%)	G	G/D (%)
Metodo1	0,415	0,090	21,74	0,232	55,98	-0,002	-0,50	0,094	22,79
Metodo2	0,386	0,070	18,21	0,217	56,20	0,004	1,08	0,094	24,50
Metodo3	0,470	0,169	35,99	0,219	46,73	-0,013	-2,85	0,094	20,13

²⁵ Si consideri che delle informazioni utilizzate per la creazione delle celle, la dimensione demografica del comune è già inserita in fase di determinazione degli strati, la cittadinanza è inserita nel insieme di vincoli finali, così come il numero di famiglie per regione. La nuova informazione introdotta si riduce quindi alla distribuzione delle famiglie estratte per numero componenti da anagrafe e la tipologia di reddito percepito da fonte fiscale.

Bisogna infatti considerare che il disegno di campionamento non solo è fatto in un momento in cui non sono disponibili le popolazioni aggiornate utilizzate nella calibrazione finale, ma è anche realizzato utilizzando una stratificazione che considera solamente la dimensione demografica del comune e non anche altre informazioni sulle famiglie legate sia alle variabili oggetto di indagine che al meccanismo che genera la mancata risposta. Il contributo dovuto alla correzione della mancata risposta diventa più importante, arrivando a spiegare il 36 per cento del totale del cambiamento, indicando come il metodo 3 sembra modellare meglio la mancata risposta. All'aumentare del contributo percentuale di R_{01} diminuisce il contributo di R_{12} , mostrando come le modifiche introdotte tramite poststratificazione divengono via via meno importanti all'aumentare della variazione introdotta dal trattamento della mancata risposta. Infine, la misura di G vede diminuire il proprio contributo relativo, facendo così diminuire l'importanza della distorsione dovuta alla mancata risposta globale. Complessivamente, poiché G è costante, R_{int} è molto vicino allo zero e R_{12} rimane all'incirca costante, la variazione complessiva di D è influenzata principalmente dal metodo per la correzione della mancata risposta utilizzato.

Resta da stabilire quanto il metodo adottato per ottenere una presumibile maggiore accuratezza della stima abbia influito sull'errore campionario. Tramite l'errore di campionamento è possibile dare una valutazione del grado di precisione delle stime ottenute da un'indagine campionaria.²⁶

Nei prospetti 2.9 e 2.10 sono riportati i valori medi, con relativi errori di campionamento percentuali in corsivo, del reddito individuale netto (per i 15enni e più) e del reddito da lavoro netto (tra quanti lavorano), a livello Italia e per le cinque ripartizioni, per i quattro campioni longitudinali e per l'intero campione trasversale.

Questo implica che nonostante si sia introdotta una maggiore dispersione tra il valore dei pesi (la misura D), si sono creati domini maggiormente omogenei al loro interno relativamente ai coefficienti di riporto all'universo. Lo stesso risultato è peraltro riscontrato nell'analisi empirica proposta da Dufour et al. (2001). Relativamente alla coerenza delle stime, gli errori che risultano dalla metodologia adottata indicano comunque l'esistenza di coerenza tra le stime prodotte per ognuno dei quattro campioni longitudinali con il campione trasversale.

Prospetto 2.9 - Reddito individuale netto per la popolazione di 15 anni e più ed errore relativo percentuale per ripartizione e campione utilizzato

RIPARTIZIONE	Campione				Campione trasversale
	C1	C2	C3	C4	
Nord-ovest	13.884	14.308	13.227	13.141	13.640
	<i>2,3</i>	<i>3,4</i>	<i>1,7</i>	<i>2,1</i>	<i>1,2</i>
Nord-est	13.331	13.336	13.324	13.582	13.393
	<i>1,8</i>	<i>1,8</i>	<i>1,9</i>	<i>2,2</i>	<i>0,9</i>
Centro	12.372	12.537	12.337	12.859	12.526
	<i>2,1</i>	<i>1,9</i>	<i>1,9</i>	<i>2,7</i>	<i>1,1</i>
Sud	8.783	9.078	8.764	9.172	8.949
	<i>2,7</i>	<i>3,0</i>	<i>2,8</i>	<i>2,7</i>	<i>1,3</i>
Isole	8.794	9.377	8.551	9.010	8.933
	<i>3,7</i>	<i>4,5</i>	<i>3,9</i>	<i>3,7</i>	<i>1,8</i>
Italia	11.710	11.990	11.495	11.771	11.742
	<i>1,1</i>	<i>1,4</i>	<i>1,0</i>	<i>1,2</i>	<i>0,6</i>

²⁶ La metodologia per il calcolo degli errori di campionamento è presentata nel paragrafo 2.4 del presente capitolo.

Prospetto 2.10 - Reddito individuale da lavoro netto (a) per la popolazione di 15 anni e più occupata ed errore relativo per ripartizione e campione utilizzato

RIPARTIZIONE	Campione				Campione Trasversale
	C1	C2	C3	C4	
Nord-ovest	17.024 3,2	18.190 5,1	16.578 2,5	16.537 2,9	17.086 2,0
Nord-est	16.705 2,6	16.840 2,6	16.456 2,7	16.725 2,9	16.681 1,2
Centro	16.828 3,0	16.313 3,1	16.119 2,8	16.497 3,8	16.438 1,6
Sud	13.767 3,9	14.319 3,6	13.549 4,7	13.838 4,2	13.868 2,1
Isole	14.244 5,2	15.732 6,8	13.946 6,7	15.060 6,0	14.740 3,6
Italia	16.019 1,5	16.538 2,1	15.616 1,5	15.894 1,6	16.016 0,9

(a) Il reddito netto da lavoro è stato calcolato per tutti gli individui con reddito da lavoro nell'anno 2003 ma che si dichiaravano occupati al momento della rilevazione, ovvero ottobre 2004.

Si osserva come il sistema di ponderazione utilizzato faccia sì che l'errore di campionamento resti assolutamente limitato relativamente alle distribuzioni marginali. Ovviamente, errori maggiori si registrano considerando i quattro campioni longitudinali (ognuno dei quali composto da circa seimila famiglie) distinti per ripartizione. Anche gli errori relativi ottenuti con gli altri metodi di correzione per mancata risposta sono piuttosto contenuti ed in linea con quelli presentati.

La metodologia adottata ha però permesso un miglioramento, anche se limitato, degli errori al diminuire della numerosità campionaria nei diversi domini di stima.

2.4 - Valutazione della precisione delle stime

2.4.1 Introduzione

Le principali statistiche che permettono di valutare la precisione delle stime prodotte da un'indagine campionaria sono gli errori di campionamento assoluto e relativo. Con riferimento alla generica stima \hat{Y}_d nel dominio d, la stima dell'errore assoluto di campionamento è espressa da:

$$\hat{\sigma}(\hat{Y}_d) = \sqrt{V\hat{a}r(\hat{Y}_d)} \quad (16)$$

e la stima dell'errore relativo di campionamento è definita come:

$$\hat{\varepsilon}(\hat{Y}_d) = \frac{\hat{\sigma}(\hat{Y}_d)}{\hat{Y}_d} \quad (17)$$

Nella (16) $V\hat{a}r(\hat{Y}_d)$ denota la stima della varianza campionaria di \hat{Y}_d . L'errore assoluto di campionamento permette inoltre di costruire l'intervallo di confidenza nel quale con una certa probabilità si trova il parametro oggetto di stima:

$$\Pr\{\hat{Y}_d - k\hat{\sigma}(\hat{Y}_d) \leq Y_d \leq \hat{Y}_d + k\hat{\sigma}(\hat{Y}_d)\} = P \quad (18)$$

Nella (18) il valore di k dipende dal valore fissato per la probabilità P; ad esempio, per P=0,95 si ha k=1,96.

In linea generale, l'indagine Eu-Silc fornisce le seguenti stime:

1. medie, ad esempio il reddito medio individuale netto;
2. frequenze, ad esempio il numero di famiglie per caratteristiche e titolo di godimento dell'abitazione;
3. gli indicatori di povertà relativa e disuguaglianza sociale.

Gli errori di campionamento delle medie e delle frequenze, stime correntemente diffuse dall'Istituto, sono calcolati mediante la metodologia *standard*²⁷ di stima della varianza campionaria utilizzando le procedure software²⁸ sviluppate in Istat a tale scopo.

La precisione degli indicatori di povertà relativa e disuguaglianza sociale è invece valutata impiegando metodologie più complesse; questi indicatori sono infatti espressi da funzioni non lineari e pertanto la metodologia *standard* non può essere applicata per stimare i corrispondenti errori di campionamento.

2.4.2 Gli indicatori di povertà relativa e disuguaglianza sociale (indicatori di Laeken)

Gli indicatori di povertà relativa e disuguaglianza sociale²⁹ sono stimati annualmente mediante i dati rilevati dal campione trasversale dell'indagine Eu-Silc³⁰ e trasmessi ad Eurostat con i corrispondenti errori di campionamento secondo quanto stabilito dal regolamento europeo che disciplina l'indagine.

Ripartiamo di seguito gli indicatori di povertà relativa e disuguaglianza sociale: la linea di povertà relativa, l'incidenza di povertà relativa, l'indice di disuguaglianza della distribuzione del reddito, il gender pay gap, l'intensità di povertà relativa, il rapporto interquintile; indicando con k l'indice dell'unità (individuo), y_k il valore della variabile Y sull'unità k , w_k il peso dell'unità k , \hat{Y}_β la stima del β -esimo quantile della distribuzione della variabile Y ($0 \leq \beta \leq 1$):

1. la linea di povertà relativa (at *Risk-of-poverty threshold*), definita come il 60 per cento del valore mediano della distribuzione del reddito³¹ a livello nazionale:

$$RPT = 0,60 \cdot \hat{Y}_{0,5} \quad (19)$$

2. l'incidenza di povertà relativa (at *Risk-of-poverty rate*), definita come la percentuale di individui con reddito inferiore alla linea di povertà relativa:

$$RPR = \frac{\sum_{k \in S} I_k w_k}{\sum_{k \in S} w_k} \times 100 \quad (20)$$

dove S denota il campione rilevato, I_k la variabile indicatrice così definita:

²⁷ Per *metodologia standard* si intende metodologia notoriamente diffusa e applicata; la letteratura sulla teoria del campionamento da popolazioni finite, fornisce le formule per il calcolo della varianza degli stimatori frequentemente utilizzati nei più importanti disegni di campionamento.

²⁸ La procedura Sgce (Sistema generalizzato di calcolo degli errori campionari) e il software Genesees (software generalizzato per il calcolo degli errori campionari nelle indagini); il software Genesees è disponibile sul sito dell'Istat.

²⁹ Gli indicatori di povertà relativa e disuguaglianza sociale sono noti anche come *indicatori di Laeken* poiché adottati ufficialmente durante il Consiglio europeo svoltosi a Laeken nel 2001.

³⁰ L'indagine Eu-Silc è basata su un disegno di campionamento a due stadi di selezione (comuni, famiglie) con stratificazione delle unità di primo stadio. È prevista una componente longitudinale; il campione di famiglie è infatti articolato in quattro gruppi di rotazione in modo che ogni famiglia sia rilevata per quattro anni consecutivi.

³¹ Col termine *reddito* si intende *reddito disponibile equivalente*.

$$I_k = \begin{cases} 1 & \text{se } y_k < RPT \\ 0 & \text{altrimenti} \end{cases} \quad (21)$$

3. *l'indice di disuguaglianza della distribuzione del reddito (inequality income distribution)*, misurato dall'indice di Gini secondo la formulazione proposta da Eurostat:

$$G = 100 \times \left\{ \frac{2^* \sum_{k=prima\ pers.}^{ultima\ pers.} \left(y_k * w_k * \sum_{prima\ pers.}^{pers.\ k} w_k \right) - \sum_{k=prima\ pers.}^{ultima\ pers.} y_k * w_k^2}{\left(\sum_{k=prima\ pers.}^{ultima\ pers.} w_k \right) * \sum_{k=prima\ pers.}^{ultima\ pers.} y_k * w_k} - 1 \right\} \quad (22)$$

4. Il *Gender pay gap (indice di disuguaglianza, tra maschi e femmine, della retribuzione lorda media oraria)*, definito come la differenza tra la retribuzione lorda media oraria dei maschi e delle femmine rispetto alla retribuzione lorda media oraria dei maschi (la popolazione consiste di tutti i lavoratori dipendenti in età 16-64 anni con almeno 15 ore di lavoro a settimana):

$$GPG = \frac{\frac{\sum_{k \in M} y_k w_k}{\sum_{k \in M} y_k} - \frac{\sum_{k \in F} y_k w_k}{\sum_{k \in F} y_k}}{\frac{\sum_{k \in M} y_k w_k}{\sum_{k \in M} y_k}} \times 100 \quad (23)$$

dove M denota l'insieme degli individui campione di sesso maschile lavoratori dipendenti con età compresa tra 16 e 64 anni con almeno 15 ore di lavoro settimanali, F l'analogo campione di sesso femminile

5. *l'intensità di povertà relativa (Relative median at risk-of-poverty gap)*, definita come la differenza tra la linea di povertà relativa e la mediana del reddito dei poveri rapportata alla linea di povertà relativa medesima:

$$RPG = \frac{RPT - \hat{Y}_{0.5}^{poor}}{RPT} \times 100 \quad (24)$$

dove RPT è la linea di povertà relativa espressa dalla (19) e $\hat{Y}_{0.5}^{poor}$ è il valore mediano della distribuzione del reddito degli individui poveri (individui con $I_k=1$, si veda formula (21))

6. *il rapporto interquintilico (Income quintile share ratio)*, definito come il rapporto tra il totale del reddito ricevuto dal 20 per cento della popolazione con reddito più elevato e il totale del reddito ricevuto dal 20 per cento della popolazione con reddito più basso:

$$QSR = \frac{\sum_{k \in T} y_k w_k / \sum_{k \in T} w_k}{\sum_{k \in L} y_k w_k / \sum_{k \in L} w_k} \quad (25)$$

dove T è l'insieme degli individui campione con $y_k > \hat{Y}_{0.80}$ e L è l'insieme degli individui campione con $y_k \leq \hat{Y}_{0.20}$.

2.4.3 Lo studio empirico per la scelta della metodologia di stima della varianza campionaria degli indicatori di Laeken

Per stimare la varianza campionaria di indicatori complessi, la letteratura propone due metodologie: la linearizzazione e il ricampionamento. Al fine di individuare la strategia metodologica più adeguata per calcolare gli errori campionari degli indicatori di Laeken³², Istat ha comparato le due differenti metodologie mediante uno studio empirico.

La linea di povertà relativa, l'incidenza di povertà relativa, l'indice di Gini e il gender pay gap sono stati linearizzati³³ mediante la metodologia delle *equazioni stimanti* e il metodo di Taylor–Woodruff e gli errori campionari degli indicatori di Laeken sono stati stimati, in via sperimentale, sia tramite la linearizzazione (ove possibile) sia tramite il ricampionamento mediante la tecnica delle replicazioni bilanciate ripetute (Brr)³⁴ applicata ai dati dell'indagine Echp.³⁵

Il prospetto 2.11 riporta i valori degli errori campionari relativi ottenuti dallo studio sperimentale; si osserva che le due differenti metodologie conducono a valori simili.

Il risultato ottenuto è importante in quanto le due metodologie messe a confronto sono differenti tra loro sia in termini concettuali che computazionali. La linearizzazione è infatti basata su un procedimento analitico mentre il ricampionamento su elaborazioni intensive. Sinteticamente, la linearizzazione consiste nel definire, per ogni stimatore complesso, una nuova variabile detta *linearizzata*, tale che la varianza campionaria dello stimatore lineare costruito in base alla variabile linearizzata sia asintoticamente uguale alla varianza campionaria dello stimatore complesso iniziale; la stima della varianza di uno stimatore complesso viene quindi ricondotta alla stima della varianza di uno stimatore lineare.

Prospetto 2.11 - Errore campionario relativo percentuale degli indicatori di Laeken

INDICATORE	Errore campionario relativo (%)	
	Linearizzazione	Brr
Linea di povertà relativa	1,72	1,73
Incidenza di povertà relativa	3,55	3,53
Indice di Gini	1,71	1,60
Gender Pay Gap	30,51	32,61
Intensità di povertà relativa	-	6,61
Rapporto interquintilico	-	2,88

La teoria della linearizzazione fornisce le metodologie per definire la variabile linearizzata; è possibile distinguere tra una metodologia classica applicabile in presenza di stimatori complessi *più regolari* (ad esempio, il metodo di Taylor–Woodruff applicabile nel caso di stimatori complessi espressi come combinazione lineare di totali) e una metodologia più

³² Inizialmente Eurostat non ha fornito indicazioni particolari sull'argomento.

³³ In questa fase non si è ottenuta una linearizzazione attendibile dell'intensità di povertà relativa e del rapporto interquintilico.

³⁴ Si è fatto uso di tale tecnica in quanto è risultata molto affidabile in precedenti esperienze di stima della varianza di indicatori non lineari; in particolare, la sua bontà è stata verificata nell'ambito del *Progetto per la produzione delle misure di povertà regionale* (stimate dall'indagine sui Consumi delle famiglie) dove si è constatato che i risultati ottenuti attraverso la linearizzazione sono prossimi a quelli ottenuti attraverso la tecnica Brr.

³⁵ I dati dell'indagine Eu-Silc non erano disponibili al momento dello studio sperimentale.

generale e recente, utile in presenza di stimatori particolarmente complessi non trattabili con la tecnica classica (ad esempio, la metodologia delle *equazioni stimanti*).

Il ricampionamento si basa invece su elaborazioni intensive; in particolare, la tecnica Brr consiste nel: a) formare un certo numero di sub-campioni con le unità del campione rilevato; ogni sub-campione estratto costituisce una replicazione; b) calcolare la stima dell'indicatore di interesse per ogni replicazione; c) stimare la varianza campionaria dell'indicatore di interesse mediante la variabilità delle stime calcolate attraverso le replicazioni.

Per eseguire lo studio comparativo qui descritto, Istat ha implementato la metodologia di linearizzazione e la tecnica Brr in ambiente Sas. Infatti, le procedure software di stima della Brr varianza campionaria già utilizzate in Istituto non elaborano automaticamente espressioni linearizzate e non fanno ricorso a tecniche di ricampionamento; inoltre, pacchetti software esterni esistenti sul mercato, non risolvono completamente le problematiche metodologiche e computazionali di stime complesse ottenute da indagini campionarie complesse e di grandi dimensioni come quelle eseguite dall'Istat. Lo scopo principale della procedura informatica dedicata alla linearizzazione è quello di predisporre la variabile linearizzata opportuna per ogni indicatore complesso e fare in modo che le variabili linearizzate ottenute siano gestite in maniera automatica dal software Genesee per la stima della varianza campionaria già utilizzato in Istituto. La seconda procedura informatica è stata invece sviluppata per implementare la versione per disegni complessi della tecnica Brr al fine di tenere in debito conto il disegno dell'indagine Eu-Silc.

2.4.4 Metodologia corrente di stima della varianza campionaria degli indicatori di Laeken

A seguito della difficoltà sull'argomento incontrata da alcuni paesi europei, Eurostat ha rivisto la propria posizione. Ha quindi raccomandato agli Istituti nazionali di statistica l'uso della linearizzazione piuttosto che il ricampionamento e ha fornito le linearizzazioni degli indicatori ai paesi in difficoltà.

Istat, avendo ottenuto le due linearizzazioni mancanti dell'*intensità di povertà relativa* e del *rapporto interquintilico*, ha quindi adottato ufficialmente l'approccio della linearizzazione per il calcolo corrente degli errori campionari degli indicatori di Laeken, tenendo in considerazione le seguenti motivazioni: a) la raccomandazione di Eurostat sull'argomento; b) i risultati ottenuti dallo studio empirico; c) il maggiore onere computazionale della tecnica Brr rispetto alla linearizzazione.

Gli errori di campionamento degli indicatori di Laeken sono quindi calcolati correntemente dal 2005 con la procedura software sviluppata per l'approccio di linearizzazione.

In base al regolamento europeo, le stime degli indicatori di Laeken sono trasmesse ad Eurostat con i corrispondenti errori di campionamento a livello nazionale fatta eccezione per l'*incidenza di povertà relativa* e per l'*intensità di povertà relativa* che sono stimate anche per sottopopolazioni di interesse. In particolare, gli errori di campionamento dell'*incidenza di povertà relativa* sono stimati anche per *genere*, *età*, *tipologia familiare*, *condizione professionale*, *titolo di godimento dell'abitazione* mentre quelli dell'*intensità di povertà relativa* sono stimati anche per *genere* ed *età*.

Capitolo 3 - Le strategie di rilevazione

3.1 - La popolazione oggetto di indagine e l'unità di rilevazione

Come già descritto nel capitolo 2, la popolazione di riferimento per le statistiche Eu-Silc è costituita da tutte le famiglie residenti nel territorio nazionale al momento della rilevazione, al netto delle persone che vivono in convivenze istituzionali. Le informazioni rilevate riguardano sia le famiglie sia gli individui che ne fanno parte, e devono essere collegabili tra loro (in entrambe le componenti trasversale e longitudinale), per consentire un'analisi pluridimensionale sia a livello di famiglia che di singolo individuo. Per l'indagine, si fa riferimento alla composizione di fatto della famiglia, anche se il registro anagrafico è utilizzato come lista di estrazione del campione.

3.2 - Le strategie di rilevazione

Il buon livello di tasso di risposta ottenuto nelle prime edizioni dell'indagine è il risultato di una approfondita messa a punto dell'intero impianto d'indagine, resa possibile dall'analisi critica dei risultati delle indagini pilota.¹ Il contributo delle indagini pilota si è espresso nel 2004 in termini sia di contenimento delle mancate risposte totali e parziali, che di affidabilità del dato rilevato e di perfezionamento della gestione del processo d'indagine, ed ha costituito un ottimo punto di partenza per la progettazione ed implementazione della versione definitiva dell'indagine Eu-Silc, nella componente sia trasversale che longitudinale. Le strategie di rilevazione che hanno avuto maggiore impatto sull'aumento del tasso di risposta, sono state quelle relative ai seguenti aspetti:

3.2.1 Sensibilizzazione delle famiglie

Prima della rilevazione, le famiglie ricevono due lettere di sensibilizzazione: una a firma del Presidente dell'Istat ed una, dai contenuti simili, a firma del sindaco del comune di residenza.

Anche dalla consolidata esperienza delle Indagini Multiscopo sulle famiglie, infatti, risulta che:

- 1) il duplice invio riduce la possibilità che una famiglia non venga raggiunta da alcuna lettera di preavviso, con conseguente difficoltà di coinvolgimento e quindi, verosimilmente, aumento del rischio di “caduta” (familiare e/o individuale), o comunque di non veridicità delle informazioni fornite, soprattutto se di tipo reddituale;
- 2) il tasso d'adesione all'indagine e la qualità delle risposte dipendono anche dal contenuto della lettera in merito ad aspetti cruciali quali:
 - a) un'esposizione sintetica, con terminologia chiara e semplice, degli obiettivi della rilevazione. Si è cercato di far comprendere al rispondente che le informazioni fornite avrebbero consentito di indirizzare interventi di livello europeo in campo socio-economico su alcuni dei problemi più “sentiti” dai cittadini, quali disoccupazione, povertà ed esclusione sociale. L'obiettivo era far sentire il rispondente protagonista di

Il capitolo è a cura di Beatrice Greco

¹ “Indagine pilota sul reddito e condizioni di vita – Anno 2002”, “Indagine europea sulle condizioni di vita delle famiglie – Anno 2003”, “Indagine trasversale sulle condizioni di vita – Anno 2003”.

un processo nel quale egli fosse inizialmente “fonte” dell’informazione e poi “destinatario” delle politiche; rassicurazioni in merito al trattamento in forma anonima ed aggregata delle risposte fornite, coerentemente con la normativa vigente in tema di segreto statistico. Al rispondente sono stati forniti gli estremi delle leggi in materia di tutela della riservatezza e del segreto statistico non solo nella lettera del presidente e del sindaco, ma anche sul retro dei modelli di rilevazione, in modo che potesse prenderne visione in qualunque momento dell’intervista. L’attenzione a tali aspetti ha avuto impatto notevole sul contenimento delle mancate risposte, soprattutto considerando che l’indagine mira a rilevare dettagliatamente variabili di reddito, percepite come “quesiti sensibili” dal rispondente, e dunque oggetto di reticenze, omissioni o dichiarazioni distorte;

- b) disponibilità di un “numero verde” a cui la famiglia potesse rivolgersi per dubbi o chiarimenti sull’indagine, sul rilevatore o altro;
- c) tempestività dell’invio di entrambe le lettere, recapitate nei giorni immediatamente precedenti l’inizio della rilevazione. Il rilevatore ha così potuto sfruttare al meglio, nell’approccio iniziale con la famiglia, gli effetti della sensibilizzazione tramite le lettere di preavviso, con conseguente incremento della probabilità di portare a termine con successo l’intervista.

3.2.2 *Formazione dei rilevatori e dei referenti presso gli uffici regionali*

La formazione dei rilevatori è stata progressivamente migliorata, tenendo conto delle esperienze già maturate, soprattutto grazie alla stretta collaborazione fra le unità operative dell’Istat, impegnate nel progetto Eu-Silc a livello centrale, e gli uffici regionali dell’Istat. Nei confronti degli uffici regionali si è agito per migliorare il livello di conoscenza ed approfondimento delle tematiche d’indagine ed aumentare il grado di coinvolgimento, secondo le rispettive competenze. Ne è risultata una continua collaborazione nella messa a punto dell’indagine, sia per quanto riguarda il processo sia per ciò che attiene al lavoro di predisposizione degli strumenti di rilevazione.²

L’investimento nella formazione e nell’integrazione degli uffici regionali nelle dinamiche d’indagine è stato ripagato dall’evidenza di “profili di rilevatore” più alti e qualificati, con conseguenti ripercussioni sui risultati d’indagine, in termini sia di quantità di interviste portate a termine che di qualità dei dati rilevati.³

Il raggiungimento di tali *performance* è stato possibile anche grazie ad alcune decisioni prese sulla base dell’esperienza delle indagini pilota: tra queste, ricordiamo la realizzazione di un apposito questionario per il rilevatore,⁴ aggiuntivo rispetto ai consueti modelli di rilevazione e agli appositi riquadri (*box* per l’intervistatore) posti alla fine dei questionari familiari ed individuali. Tale questionario, costituendo un canale diretto di comunicazione

² A tal fine, sono stati predisposti appositi questionari per gli uffici regionali che, unitamente ai *feedback* riportati in sede di riunione di istruzione con le unità operative centrali, hanno messo in evidenza alcuni punti critici sia nei modelli di rilevazione (in termini di chiarezza, sequenza o esaustività dei quesiti) sia nelle tecniche di gestione dell’intervista.

³ L’investimento nella formazione dei rilevatori è stato ancor più determinante per la gestione sul campo del *follow-up* degli individui campione, previsto dalla natura longitudinale dell’ Indagine sulle condizioni di vita del 2005 e delle *wave* successive. L’obiettivo, anche grazie all’evidenza tratta da elaborazioni su dati dell’indagine Echn (Panel Europeo delle Famiglie), è infatti quello di ottenere un buon livello di “fidelizzazione” tra rilevatore e famiglia, allo scopo di ridurre l’*attrition*.

⁴ Nel questionario per il rilevatore sono stati approfonditi i seguenti aspetti: difficoltà di comprensione dell’importanza dell’indagine da parte della famiglia; livello di comprensione ed eventuali segni di insofferenza in corrispondenza di quesiti sul reddito; motivi della difficoltà nel riferire importi monetari; tipologie di reddito per le quali l’intervistatore ha riscontrato maggiori difficoltà di rilevazione; reazione dei rispondenti alla possibilità di utilizzo di documenti fiscali/busta paga; difficoltà da parte del rispondente e del rilevatore nell’uso di documenti fiscali/busta paga.

ricercatore-rilevatore, è stato fonte di preziose informazioni ed osservazioni, che hanno consentito di:

- 1) perfezionare il questionario per intervista in termini di contenuti, comprensibilità della terminologia, struttura ed utilizzo dei modelli cartacei;
- 2) ristrutturare la guida per l'intervistatore attraverso:
 - a) un più forte coinvolgimento del rilevatore negli obiettivi dell'indagine fin dalle prime pagine, facendolo sentire "attore" fondamentale per la buona riuscita dell'indagine;
 - b) utilizzare un linguaggio quanto più diretto, chiaro e comprensibile possibile.

Particolarmente complesso è risultato coniugare la semplificazione della terminologia con il rigore definitorio, un problema che è stato affrontato attraverso:

- 3) l'approfondimento delle tematiche risultate di difficile comprensione sia per il rilevatore che per il rispondente (esempio: distinzione fra le varie tipologie di reddito o di pensione). È stato così possibile non solo fugare dubbi di interpretazione, ma anche incrementare il livello di padronanza delle tematiche d'indagine, soprattutto di quelle reddituali;
- 4) la ridefinizione e chiarificazione, illustrata da esempi, del modello comportamentale ed operativo che il rilevatore deve osservare per gestire al meglio l'intervista, dal primo contatto con la famiglia fino alla conclusione.

3.3 - La struttura dei questionari

A partire dalle indagini pilota, i modelli di rilevazione sono stati oggetto di numerose e sostanziali modifiche, resesi necessarie a seguito dell'introduzione di nuove variabili *target* da fornire ad Eurostat, nonché dell'analisi delle mancate risposte totali e parziali.

Ai fini dell'incremento del tasso di risposta ed in generale delle *performance* d'indagine, le scelte di maggior rilievo sono state quelle in merito a: tipologia dei singoli questionari e sequenza nella somministrazione degli stessi; posizionamento e sequenza delle tematiche (sezioni) e dei quesiti all'interno dei modelli di rilevazione. Entrambe verranno illustrate in dettaglio nei paragrafi seguenti.

3.3.1 Tipologia di questionari e sequenza nella somministrazione

L'impianto inizialmente utilizzato per la prima indagine pilota è stato ripreso dalla tradizione Multiscopo e prevedeva:

- 1) un modello "generale" (Silc/P1/Fam) contenente:
 - a) una scheda generale per la rilevazione, per ciascun componente della famiglia di fatto, di un insieme di variabili strutturali (sesso, età, titolo di studio, relazione di parentela, eccetera);
 - b) una serie di questionari individuali (non autocompilati) per i rispondenti di tutte le età, con informazioni di maggior dettaglio su dati anagrafici, istruzione, condizioni di salute, attività lavorativa, status professionale;⁵
 - c) un questionario familiare, somministrato ad un solo rispondente adulto per ciascuna famiglia, contenente quesiti su: casa e zona di abitazione, titolo di godimento dell'abitazione, redditi di tipo familiare (esempio: da terreni e fabbricati), situazione economica;

⁵ Nel caso che i componenti fossero più di cinque, è stato compilato un modello Silc/P1/Fam aggiuntivo, saltando la parte della scheda generale e del questionario familiare.

- 2) un modello (Silc/P1/Ind) contenente quesiti sui redditi individuali percepiti nel periodo di riferimento del reddito dai componenti di 15 anni o più, suddivisi per tipologia di reddito.

Per quanto riguarda il questionario sui redditi, già a partire dalla prima pilota si è deciso che il passaggio da una sezione all'altra non dovesse essere soltanto di tipo "sequenziale", con le singole sezioni distinte l'una dall'altra attraverso domande "filtro" poste ad inizio sezione. È stato infatti aggiunto, all'inizio del questionario sui redditi individuali, un "Prospetto riassuntivo dei redditi percepiti", nel quale il rilevatore annota quali tipi di reddito (da lavoro dipendente, autonomo, da pensioni eccetera) sono stati percepiti dal rispondente e quali sezioni del questionario devono essere compilate per ciascun tipo di reddito. Con un accorgimento tipografico, il prospetto costituisce una pagina a sé, che va tenuta aperta durante il resto dell'intervista. Consultando il prospetto, il rilevatore è così sempre in grado di ricordare quali sezioni somministrare all'intervistato. D'altra parte, vista la difficoltà nella rilevazione di variabili di reddito, i sì e i no alle domande del prospetto hanno consentito, in mancanza della compilazione delle rispettive sezioni, di ricostruire il tipo di redditi percepiti, sebbene non l'importo.

La struttura del prospetto ha subito nel tempo sostanziali modifiche, tuttavia è stato evidente fin dalla prima pilota che esso costituiva un buon punto di riferimento per raccogliere informazioni sui redditi percepiti (e quindi sui percettori), anche in presenza di mancate risposte parziali ad una o più delle sezioni corrispondenti.

Le variazioni più consistenti ai modelli di rilevazione sono intervenute nella progettazione e realizzazione della seconda indagine pilota, nel 2003, con ulteriori affinamenti nelle successive edizioni del 2004 e 2005. La sequenza finale dei questionari risulta modificata profondamente: la scheda generale è stata estratta dal modello "familiare" ed è diventata un "registro" a sé stante. Lo stesso è successo per il questionario familiare e il questionario individuale. Quest'ultimo contiene in un unico modello tutte le variabili (di reddito e non) riferite a ciascun componente di 15 anni o più (calcolando l'età al 31 dicembre dell'anno di riferimento del reddito e non più rispetto al momento dell'intervista).

Rispetto all'impianto precedente, inoltre, la struttura del modello "familiare" è rimasta sostanzialmente invariata. Quanto ai contenuti, nel questionario individuale sono stati introdotti quesiti sui redditi correnti da lavoro dipendente (retribuzione lorda e netta mensile), quesiti su tematiche "di contesto" (esempio: uso del personal computer, internet, letture, partecipazione sociale, eccetera). Ma, soprattutto, è stato migliorato il *wording*, sulla base dei suggerimenti forniti dai rilevatori e dell'analisi dei risultati della prima indagine pilota. Le classificazioni precedentemente utilizzate sono state riviste e corrette per renderle pienamente esaustive, nonché omogenee rispetto ad altre fonti (per esempio, è stata introdotta la classificazione Cp2001 al posto della Isco-88Com a due *digit*, che è quella richiesta da Eurostat ed è ottenibile, tramite ricodifica, a partire dalla prima). I quesiti sull'attività lavorativa sono stati allineati (quanto a sequenza, modalità e *wording*) a quelli dell'indagine Forze di Lavoro, sebbene con un livello di dettaglio inferiore, derivante dai diversi obiettivi delle due indagini.

Nella seconda indagine pilota, si è sperimentata la possibilità di rilevare i redditi individuali non solo tramite un questionario tradizionale, basato sulla memoria, ma anche, in alternativa, attraverso l'utilizzo dei modelli fiscali a disposizione del contribuente. A questo scopo, la predisposizione di due modelli separati ed alternativi tendeva a verificare se fosse possibile ottenere sui redditi tassi di risposta più elevati e informazioni più veritiere, cercando di orientare il rispondente all'uso di documenti fiscali piuttosto che di dichiarazioni a memoria. Tale sperimentazione, tuttavia, non ha prodotto i risultati attesi: una parte significativa dei rispondenti non aveva in casa i modelli delle dichiarazioni dei redditi o non voleva utilizzarli. Inoltre, la trascrizione degli importi dei modelli fiscali nel questionario di indagine aumentava in molti casi

il carico sul rispondente, a causa soprattutto della difficoltà di interpretazione dei modelli fiscali. Infine, la richiesta di utilizzare i modelli fiscali durante l'intervista ha generato in alcuni intervistati il timore di un controllo fiscale, non superabile con le usuali rassicurazioni. Per questi motivi, già dall'indagine del 2003 (terza indagine pilota), si è optato per una versione unificata del questionario sui redditi, inserendo tuttavia nel testo delle domande il suggerimento di consultare i documenti (busta paga, cedolino della pensione, riepilogo annuale del mutuo, modelli fiscali, eccetera).

3.3.2 Posizione e sequenza di tematiche (sezioni) e quesiti nei questionari

A partire dalla prima indagine pilota, sono state numerose le variazioni apportate alle "sezioni" del questionario (ossia agli argomenti affrontati) e, di conseguenza, ai quesiti in esse contenuti. I motivi di tali cambiamenti sono imputabili a sopravvenute esigenze espresse in sede Eurostat;⁶ e, soprattutto, all'accurata opera di sistematicizzazione delle aree tematiche, mirata a contenere le mancate risposte totali e parziali, attraverso il raggiungimento di alcuni obiettivi intermedi, quali la riduzione del carico sul rispondente, particolarmente alto in considerazione del fatto che l'indagine rileva variabili di reddito; la realizzazione di un buon questionario, inteso sia come strumento di misurazione dei fenomeni di interesse (attraverso il miglioramento della struttura e del contenuto), sia come strumento di rilevazione essenziale per un'efficace consultazione e gestione da parte del rilevatore.

Per raggiungere tali scopi, si è operato in più direzioni:

- 1) scelta per la non autocompilazione del questionario, con trascrizione diretta delle risposte da parte del rilevatore: tale scelta è stata motivata dalla complessità della gestione delle interviste e degli argomenti affrontati, e, soprattutto, dal fatto che il rilevatore potesse fungere da "tampone", prevenendo ed attutendo atteggiamenti di chiusura da parte dell'intervistato;
- 2) sequenza fra sezioni: le modifiche effettuate in tal senso, hanno avuto un notevole impatto sui tassi di risposta. Infatti, l'aver inserito in un momento strategico dell'intervista quesiti sui principali problemi economici della famiglia, ha reso più evidenti gli scopi generali dell'indagine e ragionevolmente reso più facile affrontare successivamente le domande "sensibili" sui redditi percepiti dai rispondenti. Ciò vale anche per il questionario individuale, strutturato in modo che il rispondente venisse "accompagnato" gradualmente da tematiche generali (dati anagrafici, istruzione) a tematiche più attinenti concetti di reddito (attività lavorativa, condizione nella professione), fino a quelle strettamente riferite al reddito.⁷
Un altro esempio di riduzione delle mancate risposte grazie al miglioramento della sequenza è rappresentato dai redditi da pensione. Nella prima pilota si era registrato un alto numero di missing values nella batteria sulle "pensioni", per lo più imputabile alla mancanza di un "percorso guidato" tra i vari tipi di pensione. Per tale motivo, già dalla seconda pilota, tutte le tipologie di pensione sono state accorpate in una sola sezione, in cui sono stati inseriti opportuni quesiti-filtro che consentono al rispondente di focalizzare l'attenzione su un solo tipo per volta;
- 3) posizione delle sezioni: è stata modificata per cogliere meglio alcuni fenomeni e, in alcuni casi, per evitare duplicazioni di redditi. Ad esempio: la sezione "redditi da terreni e fabbricati" nella prima indagine pilota era una sezione a sé stante ed era posizionata nel

⁶ Ad esempio, si pensi all'introduzione della sezione sul *child care* a partire dall'indagine del 2003.

⁷ La progressività nel trattamento delle tematiche è stata massimizzata nella formulazione del questionario individuale (Silc/04/Ind) dell'indagine definitiva del 2004 (e poi del 2005), che, come visto, raggruppa in un unico modello tutte le variabili riferite a ciascun componente di 15 anni o più.

questionario familiare. A partire dalla seconda pilota, per ottenere informazioni ed importi più precisi, ciascun componente della famiglia deve rispondere rispetto alla propria situazione: tale tematica è stata infatti spostata nel questionario individuale ed inserita nella sezione “altri redditi”,⁸

- 4) tipologia, posizione e sequenza dei quesiti all'interno delle sezioni: una volta individuata la posizione e la sequenza ottimale fra aree tematiche di interesse, sono state rivisti i quesiti di ciascuna sezione, alla luce delle esperienze man mano accumulate, dei risultati dei focus group, nonché, come visto, dei suggerimenti forniti dai rilevatori e dagli uffici regionali dell'Istat.

Le scelte hanno interessato molti aspetti, tra i quali:

- a) tipologia di domande (domande aperte/chiusure; livello di dettaglio delle modalità di risposta; eliminazione di quesiti⁹) e posizione dei quesiti all'interno della sezione, al fine di minimizzare le mancate risposte parziali e di evitare sottostime di alcuni fenomeni.¹⁰ In tema di redditi percepiti, le sezioni che hanno subito le modifiche più consistenti sono:

- redditi da lavoro autonomo: questa è la sezione che ha subito le variazioni più forti già dalla prima pilota, in quanto ha sempre mostrato tassi di mancata risposta piuttosto elevati (sia a livello di intera sezione, sia di singolo quesito) e rischi di importi non attendibili. In sintesi, queste sono state le strategie mirate al contenimento delle mancate risposte:
 - raggruppare in un'unica sezione tutti i redditi da lavoro autonomo, diversamente da quanto fatto nella prima pilota; il passaggio da una tipologia all'altra è stato guidato da domande-filtro, in modo che il rispondente non saltasse alcuna tipologia e si concentrasse di volta in volta sulla specifica voce di reddito autonomo. Ciò ha consentito di avere meno missing, importi più congrui ed evitare sottostime di alcune categorie di reddito autonomo (esempio: i redditi Co.Co.Co, che a volte sono confusi con i redditi da lavoro dipendente);
 - iniziare la sezione con un quesito sugli impieghi del reddito da lavoro autonomo, col duplice scopo di ottenere un graduale approccio all'argomento, e portare alla sua memoria gli usi più frequenti di tali importi, in modo da ottenere successivamente importi più precisi;
 - richiedere importi per cifre prestabilite nel caso in cui il rispondente non voglia o non sappia rispondere in modo puntuale. Ciò ha consentito una sorta di “recupero” dei casi (numerosi) che hanno preferito non riferire un ammontare preciso;
 - prevedere la possibilità di dichiarare le perdite subite e il relativo importo;
- redditi da attività finanziarie: tale sezione ha evidenziato fin dalla prima pilota tassi di risposta piuttosto bassi, imputabili alla natura “sensibile” del tema. Il rispondente infatti ha mostrato spesso molta difficoltà a dichiarare l'ammontare del proprio risparmio, per timore non solo di controlli ma anche del giudizio del rilevatore (ciò è ancor più plausibile per le piccole realtà comunali, dove il rilevatore è spesso ben conosciuto). I questionari, quindi, sono stati modificati nel tentativo di ottenere quante più informazioni possibili. Si è passati da una

⁸ Tale sezione è somministrata a tutti gli individui di 15 anni o più, anche quelli che non percepiscono redditi da lavoro o pensione.

⁹ Ad esempio, la maggior parte dei quesiti su tasse ed imposte sono stati eliminati dall'indagine del 2003 in poi.

¹⁰ Ad esempio, le domande sugli assegni familiari, a seguito degli scarsi risultati della pilota, sono stati inserite in ciascuna delle sezioni sui redditi, piuttosto che mantenute in un'unica sezione.

sequenza che anteponeva il quesito sul rendimento degli investimenti a quello sull'ammontare, ad una versione simmetrica, in cui si procede gradualmente: viene chiesta prima la composizione del portafoglio (per ciascun tipo di attività finanziaria, va scelta la percentuale predefinita più adeguata al caso), poi l'ammontare puntuale, e quindi, se non fornito, l'ammontare per cifre prestabilite (indicare quella più vicina alla situazione reale); solo a questo punto, si chiede il rendimento, prima puntuale e poi, eventualmente, per cifre prestabilite;

- b) selezione di riferimenti temporali quanto più omogenei possibile, in modo da "accompagnare" la memoria del rispondente verso una più efficace focalizzazione del periodo di riferimento e dei relativi importi percepiti.¹¹ Ad esempio: considerando un lavoratore dipendente, si passa gradualmente da temi di tipo "corrente" (attività lavorativa principale, redditi correnti da lavoro dipendente, informazioni sul lavoro svolto), a quesiti sui redditi percepiti nell'anno precedente (cioè nel periodo di riferimento del reddito);
- c) *wording*, calibrato in modo da garantire da una parte un'esposizione neutrale da parte del rilevatore, dall'altra una comprensione piena e "serena" da parte del rispondente, lontano da timori di controlli fiscali o giudizi.

3.3.3 Il modello Ril prestampato: modalità di individuazione della famiglia da intervistare

A partire dall'edizione del 2004, i rilevatori hanno individuato l'indirizzo di residenza di ciascuna famiglia-campione da intervistare servendosi di un modello cartaceo (Silc/Ril), contenente alcune informazioni prestampate, sulla base dei dati forniti dal comune in fase di estrazione.¹²

Il modello è stato strutturato, e successivamente migliorato, per poter disporre di uno strumento di misurazione più efficace ed efficiente, grazie al maggior livello di:

- 1) affidabilità dei contenuti grazie alla prestampa: dato il carattere longitudinale dell'indagine, si è ritenuto fondamentale avere assoluta certezza che alcune informazioni contenute nel modello Ril fossero esenti da errori. Tra queste: le chiavi identificative familiari ed individuali, essenziali per il *follow-up* negli anni successivi; il nominativo (nome e cognome) dei componenti così come risultante nell'anagrafe comunale, al fine di una più agevole identificazione da parte del comune in fase di successiva verifica della presenza in anagrafe;¹³
- 2) praticità di utilizzo da parte del rilevatore, attraverso:

¹¹ Le variabili di reddito richieste da Eurostat sono: di tipo corrente; rilevate nel periodo di riferimento del reddito; rilevate su periodi predefiniti ed antecedenti il momento dell'intervista (esempio: ultimi 12 mesi); quindi, notevole sforzo è stato fatto per ridurre al minimo la possibilità di dichiarazioni falsate da effetto memoria o salti di riferimenti temporali.

¹² Codice Istat della famiglia (attribuito dall'Istat in fase di validazione del file delle famiglie estratte, dunque prima della rilevazione); indirizzo della famiglia; codice Istat del componente (anch'esso attribuito dall'Istat); cognome e nome del componente. Se al momento della rilevazione, si osservano componenti (detti di fatto) dimoranti abitualmente presso l'abitazione ma non prestampati, il loro nome e cognome viene inserito a mano a partire dalla prima riga vuota del modello Ril, dopodiché l'intervista prosegue, come per gli altri, sul modello Reg ed eventualmente (in base all'età) su quello Ind.

¹³ A partire dal 2005, lo schema di campionamento e gli esiti di intervista degli anni precedenti determinano l' "entrata" o l' "uscita" di una famiglia dal campione nell'anno in corso (anno t). Per le sole famiglie da intervistare al tempo t, a metà anno, si richiede al comune di effettuare via *web* la verifica della presenza in anagrafe degli individui-campione rilevati, in t-1, presso ciascuna unità abitativa (definita dall'insieme di: via/piazza, numero civico e Cap) coinvolta nell'indagine dell'anno t-1 (ossia, quelle estratte dal comune il primo anno di vita della famiglia e quelle originate a seguito di *split*). Per ciascuno di questi individui, il comune dovrà dichiarare quale delle seguenti situazioni si è verificata: a) l'individuo risulta residente anagraficamente presso l'unità abitativa; b) l'individuo è deceduto; c) l'individuo si è trasferito. In quest'ultimo caso il comune dovrà inoltre dichiarare se l'individuo si è trasferito all'estero, o presso un'istituzione, o presso un altro indirizzo nello stesso comune o in altro comune italiano, dando gli estremi del nuovo indirizzo, se noti. Nel caso in cui l'individuo si sia trasferito all'estero o presso un'istituzione, non dovrà essere ricontattato in occasione dell'indagine al tempo t, in accordo con le direttive di Eurostat (regolamento Ce n. 1982/2003).

- a) l'uso di un solo modello (foglio) per ogni famiglia-campione, piuttosto che un elenco indistinto di famiglie;¹⁴
- b) scelte di carattere grafico, mirate a ridurre gli errori di compilazione ed ottimizzare i tempi di intervista. Il rilevatore, infatti, affiancando il modello Ril al modello Reg, può fornire le informazioni richieste sul Reg per ciascun componente riportato sul Ril senza rischio di slineamento fra individui (righe);
- c) facilità distribuzione delle interviste fra i rilevatori, sulla base di aree territoriali omogenee e/o in funzione del grado di mobilità geografica di ciascuno di essi.¹⁵

3.4 - La gestione informatizzata del processo di indagine e delle famiglie campione (Sigif)

Per la gestione del processo d'indagine dell'indagine sulle condizioni di vita delle famiglie, l'Istat ha progettato e sviluppato un sistema informativo integrato (Sigif - Sistema informativo gestione indagine sulle famiglie). Obiettivo di tale sistema è supportare le differenti indagini della Direzione dell'Istat per le indagini sulle famiglie, mantenendo costanti le strutture fisiche, gli standard e le logiche di utilizzo, pur adattando ogni singola componente *software*. Tale integrazione è stata realizzata contemporaneamente

- 1) fra componenti del sistema, ossia:
 - a) *web*, tramite il portale <https://nuovafdl.istat.it>, che i comuni usano prevalentemente come “nodo di scambio” con l'Istat in fase di acquisizione dei nominativi delle famiglie estratte (e dei relativi componenti) e, dal 2005, di verifica della presenza in anagrafe degli individui campione da sottoporre di nuovo ad intervista, ai fini di un più efficace *follow-up* degli stessi in fase di rilevazione;
 - b) Gefi (Gestione integrata delle indagini sulle famiglie) *software* che permette la registrazione dei dati del comune, delle famiglie estratte e dei relativi componenti in sostituzione dei tradizionali modelli cartacei e la successiva preparazione del *file* da inviare all'Istat per garantire l'omogeneità di tutte le informazioni provenienti dai differenti comuni;
 - c) Oracle, ad uso dell'Istat centrale e degli uffici regionali, con funzioni prettamente tecnico-operative e di processo;
- 2) fra le singole fasi del processo;¹⁶
- 3) fra le singole utenze (Istat centrale, uffici regionali, comuni campione), in stretta collaborazione fra loro.

¹⁴ È stato inoltre eliminato il rischio di errori di trascrizione dall'elenco a più fogli (osservati spesso nelle indagini pilota), particolarmente gravi soprattutto in considerazione della necessità di mantenere univoca, anno per anno, la chiave della famiglia e dei relativi componenti assegnata dall'Istat prima della rilevazione.

¹⁵ Tale aspetto ha assunto particolare rilievo nei caso di grandi comuni, che sono stati ulteriormente agevolati attraverso la fornitura, via *web*, di un file contenente l'elenco di tutte le famiglie-campione di propria competenza, ordinabile secondo alcune variabili a scelta del comune (esempio: sezione di censimento).

¹⁶ Le principali fasi d'indagine, incluse quelle di tipo più strettamente operativo, supportate dal Sigif, sono: definizione del campione, acquisizione e validazione dei nominativi delle famiglie estratte e dei relativi componenti, gestione della formazione agli enti periferici, monitoraggio della rilevazione, supporto al controllo quali-quantitativo del materiale di rilevazione inviato dai comuni a fine indagine. Dal 2005, è stato inoltre introdotto un ulteriore “modulo” (con componenti *web* e non) per l'attività di verifica della presenza in anagrafe degli individui-campione da reintervistare di anno in anno, e per la gestione del *follow-up* degli stessi sul territorio italiano.

Capitolo 4 - Procedure di controllo, correzione, imputazione e integrazione di dati da fonte amministrativa

4.1 Introduzione

L'indagine sulle condizioni di vita è attualmente realizzata con tecnica di tipo Papi (*Paper assisted personal interview*) da parte di intervistatori della rete di rilevazione comunale. Il questionario cartaceo utilizzato è costituito da quattro modelli di rilevazione:

- Il modello Ril contiene informazioni per l'identificazione della famiglia da intervistare e informazioni per identificare i componenti della famiglia stessa.
- Il registro familiare consente la registrazione dell'esito del contatto e dell'intervista alla famiglia. Inoltre contiene una scheda familiare in cui per ogni individuo sono riportate informazioni anagrafiche, la relazione di parentela con l'intestatario della scheda, lo stato civile, la condizione lavorativa prevalente nel periodo di riferimento del reddito e informazioni sulla cura dei bambini.
- Il questionario familiare è somministrato al componente della famiglia che conosce meglio la situazione della famiglia e contiene informazioni sulla casa, sulle spese della abitazione e sulla situazione economica della famiglia.
- Il questionario individuale è somministrato ad ogni individuo di almeno 15 anni presente in famiglia. Consente di rilevare informazioni sull'istruzione, sulla situazione lavorativa, sull'eventuale ultima attività lavorativa svolta, sulla storia lavorativa ed infine sulle varie tipologie di reddito percepite (da lavoro dipendente, da lavoro autonomo, da pensioni e da altri assegni sociali, da investimento di capitale, da trasferimenti privati eccetera).

L'acquisizione delle informazioni tramite questionario cartaceo aumenta la possibilità di errori non campionari in fase di rilevazione e rende quindi indispensabile una fase di controllo e correzione dei dati a posteriori. La complessità dei modelli di rilevazione, la tipologia delle variabili rilevate e la molteplicità delle relazioni tra le variabili rendono il processo di controllo e correzione estremamente complesso e articolato. Tale processo, infatti, ha richiesto lo sviluppo di tecniche di correzione ed imputazione mirate, segnatamente per le variabili qualitative (metodo di Fellegi-Holt) e per le variabili quantitative (metodo di Hidiroglou-Berthelot e di regressioni sequenziali multiple). Il processo di correzione ed imputazione si è avvalso dei dati fonte amministrativa (componenti di reddito: trasferimenti pensionistici e lavoro autonomo) opportunamente agganciati con la tecnica dell'*exact matching*.

4.2 La strategia di acquisizione dati (Blaise)

La prima fase della procedura di controllo e correzione dei dati è costituita dalla registrazione dei dati in modalità controllata.

Il *data entry* controllato dei dati rilevati è realizzato mediante un'applicazione sviluppata con il software Blaise, creato dall'Istituto di statistica olandese come sostegno alle indagini Capi, ma poi adattato anche alla sola registrazione dei dati.

Il paragrafo 4.2 è a cura di D. Quartuccio e R. Ricci, il paragrafo 4.3 di I. Siciliani, il paragrafo 4.4 di S. Vitaletti e il paragrafo 4.5 di P. Consolini. L'introduzione è frutto del lavoro congiunto degli autori.

La fase di registrazione controllata risulta essere di notevole importanza per la riduzione dell'errore non campionario che potrebbe verificarsi a causa di una non corretta immissione dei dati. La registrazione controllata, infatti, impedisce l'immissione di valori non ammissibili per una certa variabile (controlli di dominio) e consente di controllare i rimandi a specifici quesiti in funzione di risposte fornite a quesiti filtro (controlli di percorso).

Nell'applicazione Blaise relativa all'Indagine sulle condizioni di vita si è deciso di introdurre solo controlli di dominio e si è preferito non inserire controlli di percorso, a meno di poche eccezioni, in quanto la registrazione è affidata ad una ditta esterna e, nel caso in cui il rilevatore non abbia rispettato i percorsi del questionario, si è scelto di non lasciare a chi digita la possibilità di interpretare quanto scritto sul questionario. In particolare, per i quesiti inerenti a variabili qualitative è possibile solo la digitazione di un valore compreso nelle modalità di risposta previste; per i quesiti che prevedono come risposta una variabile quantitativa, l'applicazione per la registrazione provvede a segnalare se il valore inserito risulta esterno ad un intervallo di accettazione prestabilito a priori mediante la specificazione dei valori minimo e massimo. Tale segnalazione costringe l'addetto alla registrazione a verificare che il valore registrato corrisponda a quello riportato sul questionario cartaceo, ciò al fine di limitare al minimo gli errori non campionari dovuti alla fase di registrazione.

Tale scelta, ovviamente, ha portato un maggiore carico di lavoro per la successiva fase dell'indagine, cioè quella relativa al controllo della coerenza delle risposte e alla correzione ed imputazione dei dati mancanti.

I quattro modelli compilati per ciascuna famiglia intervistata sono registrati in Blaise secondo l'ordine di compilazione:

- Modello Silc/Ril;
- Modello Silc/Reg;
- Modello Silc/Fam;
- Modello Silc/Ind.

Grazie all'implementazione del Sigif (Sistema integrato di gestione delle indagini sulle famiglie), in fase di registrazione è possibile realizzare un controllo sulle chiavi identificative delle famiglie, in modo tale da non consentire l'immissione di chiavi familiari diverse da quelle del campione teorico delle famiglie. Inoltre, sono state utilizzate alcune informazioni strutturali relative agli individui per un ulteriore controllo con l'informazione effettivamente rilevata e per la riduzione di possibili errori non campionari, sia di rilevazione che di registrazione. Le variabili a livello individuale acquisite dal Sigif sono:

- nome e cognome;
- giorno, mese e anno di nascita;
- sesso

che, al momento della registrazione, compaiono come campi pre-compilati e immutabili in quanto si tratta di dati provenienti dalle anagrafi comunali e per questo ritenuti corretti.

Per poter procedere all'immissione dei dati di una specifica famiglia, è indispensabile digitare le tre variabili necessarie ad identificare univocamente la famiglia stessa: *codice provincia, codice comune e codice Istat della famiglia*. Questa ultima variabile viene reinserita all'inizio di ogni modello, al fine di avere la garanzia che i dati si riferiscano alla stessa famiglia. Per essa è attivato un controllo *hard*. Si tratta di una tipologia di errore grave che non può essere

annullato. Quindi, se il codice della famiglia che viene digitato non è uguale a quello inserito per gli altri modelli di rilevazione relativi alla stessa intervista, la digitazione si blocca.¹

I codici provincia e comune, una volta digitati all'inizio, appaiono sempre come valori pre-compilati e, quindi, non più modificabili.

Nel modello Silc/Ril sono pre-stampati i nomi e cognomi di tutti i componenti della famiglia anagrafica. Tuttavia, al momento dell'intervista, l'intervistatore può rilevare una situazione diversa da quella anagrafica per la presenza, per esempio, di "componenti di fatto". Per questo motivo l'applicazione Blaise, oltre a far comparire come pre-compilati e immutabili tutti i nomi delle persone anagraficamente residenti presso quella famiglia, dà la possibilità di inserire nuovi componenti a cui, automaticamente, viene dato il numero d'ordine immediatamente successivo.

Il secondo modello è il Silc/Reg che contiene, tra le altre, tutte le informazioni strutturali di ogni componente della famiglia. Come già detto, il sesso e la data di nascita di ogni componente sono informazioni pre-compilate e non modificabili. Nel caso di incoerenze con quanto riportato sul questionario cartaceo, prevale sempre l'informazione riportata a video, cioè quella presente nel Sigif. Vi è un'eccezione costituita dagli stranieri. Dato che le anagrafi a volte non hanno informazioni complete relative prevalentemente agli extracomunitari, si è preferito raccogliere direttamente in fase di rilevazione. Solo in questo caso, quindi, è possibile la digitazione di sesso e data di nascita.

Per il modello Silc/Fam, sono stati previsti controlli di dominio per le variabili di reddito e quelle relative alle spese per l'abitazione per le quali, grazie alle precedenti indagini pilota, sono stati stabiliti dei valori indicativi di minimo e massimo. Questo significa che al momento dell'inserimento di un valore che non rientra nel dominio indicato, si attiva un controllo *soft*.

All'interno del modello Silc/Ind, oltre ai controlli *soft* di dominio previsti per le variabili quantitative, sono stati inseriti alcuni controlli *hard*. Infatti, al momento della digitazione di data di nascita e sesso del componente che risponde al questionario, viene fatto un controllo con gli stessi dati già presenti nel modello Silc/Reg (dati provenienti dal Sigif). Se le due tipologie di dati non corrispondono si attiva un controllo che impedisce di proseguire. Tale scelta è stata fatta per essere sicuri che il modello individuale che si sta inserendo appartenga ad un componente di quella stessa famiglia. Inoltre, dato che il modello Sinc/Ind può essere compilato solo per le persone con almeno 15 anni, il software impedisce l'inserimento di eventuali modelli compilati per persone di età inferiore.

All'interno del modello Silc/Ind, inoltre, sono stati inclusi dei controlli di percorso al fine di rendere più rapida l'immissione dei dati. All'interno del questionario, prima delle sezioni relative alla rilevazione dei redditi, vi è un prospetto riassuntivo delle diverse tipologie di reddito che il rispondente dichiara di percepire e che risulta utile per proseguire poi a riempire le rispettive sezioni. Nel prospetto, dopo aver risposto negativamente ad una domanda sulla tipologia del reddito, al fine di non perdere eventuali informazioni comunque inserite nella rispettiva sezione, si chiede a chi digita di verificare se ci siano biffature in quella specifica sezione. In caso di risposta negativa, chi digita non vedrà apparire a video la relativa sezione.

4.3 Il processo di correzione e imputazione delle informazioni

4.3.1. Tipologia dei controlli

La fase di individuazione delle situazioni di errore precede quella in cui si interviene per la rimozione dello stesso tramite l'imputazione di valori a dati mancanti o incongruenti nelle variabili rilevate.

¹ L'errore *hard* si distingue da quello *soft*. Questa ultima tipologia di errore è una sorta di avvertimento sulla plausibilità di situazioni poco verosimili, ma non necessariamente errate. Se chi digita verifica che il valore è effettivamente quello riportato sul questionario cartaceo, il controllo può essere ignorato.

La fase di controllo è finalizzata ad individuare errori che si manifestano tramite:

- 1) non univocità o non coerenza delle chiavi identificative (errori sulle chiavi identificative);
- 2) violazione dei limiti imposti dal campo di variazione delle singole variabili (errori sui domini) o dei limiti definiti dall'intervallo di accettazione per i valori plausibili delle variabili quantitative (individuazione dei valori anomali);
- 3) violazione delle relazioni logico-formali stabilite dalle norme di compilazione dei modelli cartacei (errori sui percorsi di compilazione);
- 4) violazione delle relazioni intercorrenti fra le variabili rilevate (incoerenze tra valori di variabili diverse).

La fase di controllo delle chiavi identificative consiste nella verifica dell'univocità delle stesse e della coerenza delle chiavi tra file corrispondenti a diversi modelli di rilevazione. Essa è realizzata in parte durante la fase di registrazione dei dati, in parte con procedure specifiche eseguite successivamente alla registrazione dei dati.

Come già illustrato nel paragrafo precedente, il controllo sull'ammissibilità dei valori avviene già in fase di registrazione dei dati, mentre l'identificazione dei valori anomali sulle variabili quantitative dovuti a errori di compilazione dei questionari avviene tramite una procedura generalizzata fondata sul metodo di Hidiroglou-Berthelot.²

Il controllo sul rispetto delle norme formali di compilazione dei questionari (compilazione o meno di certi quesiti a seconda delle risposte fornite a domande filtro precedenti) avviene con una procedura generalizzata che consente di identificare tutti i record e le rispettive variabili per le quali si verifica un'errata compilazione nei percorsi.

La coerenza tra le risposte a domande diverse (anche di questionari differenti, es. familiare e individuale) è verificata mediante una procedura che rileva per ciascun record tutte le situazioni di incompatibilità riscontrate. Tali incoerenze riguardano, ad esempio, la presenza di determinate condizioni professionali e la mancanza di tipologie di redditi ad esse relative; la percezione di specifiche tipologie di pensioni o assegni sociali e la mancanza dei requisiti legali previsti; la presenza di importi di retribuzione lorda e netta incompatibili tra di loro o implausibili relativamente alle ore di lavoro svolte; tipologie di lavoro svolto non compatibili con la professione dichiarata; condizioni professionali non compatibili con l'età; il possesso di un titolo di studio non compatibile con l'età di conseguimento dello stesso e con eventuali tipi di corsi frequentati; spese per l'abitazione (in particolare mutuo e affitto) implausibili in relazione alla grandezza dell'abitazione (metri quadri e stanze).

4.3.2. Procedura di correzione ed imputazione delle informazioni

Caratteristica principale dell'indagine è la notevole presenza di variabili quantitative relative agli importi delle varie tipologie di reddito che i rispondenti riferiscono di aver ricevuto e agli importi delle spese per l'abitazione. Congiuntamente a tali variabili, vi sono, inoltre, informazioni di tipo qualitativo sulla situazione lavorativa, la condizione abitativa, la situazione economica della famiglia eccetera.

La strategia di correzione è stata impostata separando l'imputazione e la correzione delle variabili qualitative dall'imputazione delle variabili quantitative. In particolare, l'imputazione di valori mancanti sulle variabili di natura qualitativa, le incompatibilità logiche tra le variabili qualitative e quelle tra le variabili qualitative e la presenza o meno di valori nelle variabili quantitative sono prevalentemente trattate tramite l'applicazione della metodologia di Fellegi-

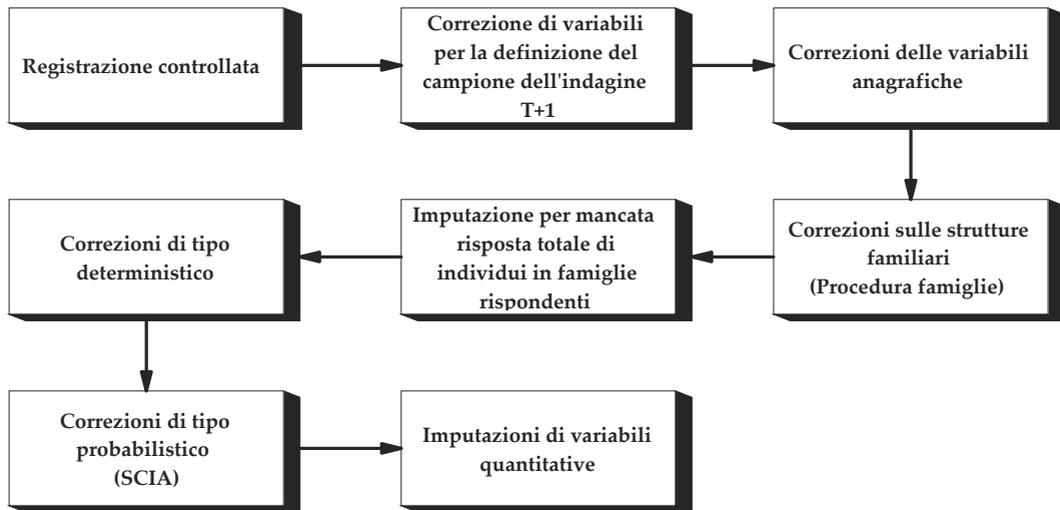
² Per un approfondimento si veda Hidiroglou M.A., Berthelot J.M., 1986 *Statistical Editing and Imputation for Periodical Business Survey*; Vitaletti S. (2005) *Correzioni e imputazioni delle informazioni sui redditi*.

Holt, ampiamente impiegata e consolidata all'interno dell'Istituto nell'ambito delle indagini sulle famiglie. I valori mancanti per le variabili di natura quantitativa sono stati imputati seguendo l'approccio Srmi (*Sequential regression multivariate imputation*) suggerito da Eurostat (2004) e basato su una serie di modelli di regressione multipla. Tale scelta garantisce, quindi, per le variabili quantitative e in particolare per quelle di reddito, una migliore comparabilità con i dati degli altri Paesi, imputati con l'utilizzo della medesima metodologia.

In termini più specifici, l'imputazione delle variabili qualitative e la rimozione delle incoerenze tra i valori delle variabili qualitative e la presenza o meno di valori nelle variabili quantitative precede l'imputazione delle variabili quantitative.

Schematicamente la procedura di correzione è articolata nei seguenti passi:

Schema 4.1 - Procedura di correzione



L'acquisizione dei dati dei questionari cartacei avviene, come detto, tramite un'applicazione che consente un *data entry* controllato, sia per quanto riguarda i valori ammissibili delle variabili, sia per quanto riguarda la digitazione delle chiavi identificative delle famiglie, in quanto non è realizzabile l'inserimento di chiavi familiari diverse da quelle esistenti nell'archivio delle famiglie campione, costruito a partire dalle informazioni estratte dalle anagrafi comunali.

Una volta eseguita la registrazione dei dati, la prima fase di correzione dà priorità alle variabili necessarie per la definizione del campione dell'occasione d'indagine successiva: trattandosi di un'indagine panel, infatti, sono state definite, da un regolamento europeo, apposite regole di inseguimento delle unità di rilevazione (*follow-up rules*). Esse sostanzialmente prevedono di re-intervistare le famiglie o meno a seconda dell'esito avuto all'intervista familiare e prevedono la re-intervista o meno degli individui nella stessa famiglia in funzione della presenza o dell'assenza in famiglia, per esempio a seguito di trasferimento/decesso. Per la necessità di definire in tempi rapidi il campione dell'occasione d'indagine successiva, la correzione delle variabili relative a tali informazioni, che avviene secondo un approccio deterministico in funzione di altre informazioni presenti nei questionari, è realizzata prima di ogni altra correzione.

4.3.3. *Correzioni delle variabili anagrafiche e strutture familiari*

Le eventuali incoerenze tra le informazioni anagrafiche individuali (data di nascita, sesso e luogo di nascita) presenti in più modelli di rilevazione sono corrette in modo deterministico attingendo alle informazioni contenute nell'archivio degli individui campione, costruito a partire dalle informazioni estratte dalle anagrafi comunali e inviate in forma elettronica e riservata all'Istat da parte dei comuni.

Le incoerenze tra l'età, il sesso, le relazioni di parentela tra i componenti della famiglia, lo stato civile, l'anno di matrimonio e l'eventuale stato civile precedente al matrimonio sono risolte con una procedura automatica denominata "Procedura Famiglie". Essa consente anche di determinare i nuclei all'interno della famiglia e, in base a questi, le tipologie familiari. Dal punto di vista metodologico, la correzione avviene dopo aver separato le famiglie monocomponenti dalle altre. Per le prime si riscontrano solo le eventuali incoerenze tra le informazioni rilevate sulla singola persona. Per le altre sono individuate le coppie possibili all'interno di ogni famiglia e sono corrette le caratteristiche dei componenti (relazioni di parentela con la persona di riferimento, stato civile, anno di matrimonio) in base alle possibili relazioni con tutte le coppie individuate, tenendo conto delle strutture familiari più frequenti.

4.3.4. *Imputazione per mancata risposta totale da parte di individui eleggibili*

Nell'indagine, oltre ad acquisire delle informazioni sulla situazione economica a livello di intera famiglia, vengono rilevate tramite il questionario individuale informazioni più dettagliate sulla situazione lavorativa e sul reddito dei singoli componenti che abbiano almeno 15 anni presenti in famiglia (diversi dal personale di servizio) o temporaneamente assenti.

Nel caso in cui una persona eleggibile per l'intervista individuale non risponda al questionario individuale, cioè nel caso di una mancata risposta totale a livello individuale in una famiglia rispondente, si procede all'imputazione totale del record mancante mediante una procedura da donatore.

Un approccio alternativo sarebbe stato l'aggiustamento dei coefficienti di riporto all'universo in modo tale da tener conto che per alcuni individui eleggibili sono disponibili solo le informazioni a livello individuale riportate nella scheda familiare. Questa soluzione avrebbe comportato la creazione di due differenti sistemi di pesi, uno per tutti gli individui e l'altro per gli individui con risposta al questionario individuale, con la possibilità che lo stesso individuo avrebbe potuto avere un coefficiente di riporto all'universo diverso a seconda della provenienza delle informazioni utilizzate (scheda familiare o questionario individuale). Inoltre, tale soluzione alternativa avrebbe avuto la necessità di individuare un metodo per la determinazione del fattore di aggiustamento del reddito familiare al fine di tener conto anche del reddito personale degli individui non rispondenti in famiglia.

Entrambe le alternative erano possibili secondo Eurostat (2004). Si è optato per l'imputazione totale del record in quanto si è ritenuto che la presenza di due sistemi di ponderazione differenti per gli stessi individui potesse creare maggiori difficoltà nell'utilizzo dei dati. Inoltre, allo stato attuale, non sono stati proposti da parte di Eurostat possibili metodi comuni per la determinazione del fattore di aggiustamento del reddito familiare, in presenza di individui di 15 anni o più non rispondenti.

L'imputazione da donatore avviene mediante l'utilizzo di una procedura generalizzata che all'interno di gruppi, originati dalle modalità di specifiche variabili di strato, seleziona i donatori, da cui prelevare i valori, tra i record completi aventi distanza minima dal record ricevente. La misura di distanza tra il donatore e il ricevente è basata sui valori assunti da variabili definite di *matching*. La scelta delle variabili di strato e *matching* è condizionata dalle informazioni comunque presenti per gli individui non rispondenti, fondamentalmente quelle riportate nella scheda familiare e quelle presenti o ricostruibili dal Questionario familiare. In particolare, sono

utilizzate la ripartizione di residenza, l'età e la condizione lavorativa principale nel periodo di riferimento del reddito, quali variabili di strato. Tale scelta comporta la definizione di 80 strati in cui suddividere donatori e riceventi. Il bacino dei donatori è definito tenendo conto della corretta compilazione della maggior parte dei quesiti, in particolare di quelli fondamentali e correlati con il tipo di reddito. Il donatore più simile all'interno dello strato è individuato sulla base di variabili relative al genere, alla percezione della famiglia sulla propria situazione economica, al reddito netto mensile familiare, al numero di componenti della famiglia, al titolo di godimento dell'abitazione, alla presenza di redditi da lavoro dipendente in famiglia, di redditi da lavoro autonomo e di redditi da pensione.

4.3.5. *Correzioni di tipo deterministico*

Come accennato in precedenza, le incompatibilità logiche tra le variabili di natura qualitativa e tra le variabili qualitative e quelle quantitative sono corrette mediante l'utilizzo della metodologia di Fellegi-Holt, basata sul principio del minimo cambiamento, che presuppone la casualità degli errori.

In alcuni casi, invece, si riscontra la presenza di errori di tipo sistematico, legati per lo più al mancato rispetto da parte dei rilevatori delle norme di compilazione del questionario; tale comportamento è facilitato anche dallo strumento di rilevazione adoperato (questionario cartaceo). In particolare, per la rilevazione degli importi delle variabili di reddito e delle altre variabili quantitative, si è spesso utilizzata una struttura tale per cui la risposta a un quesito filtro sulla percezione o meno di una data tipologia di reddito regoli la somministrazione di altri quesiti relativi all'importo e ad altre informazioni più dettagliate. Può accadere che il rilevatore registri l'importo o altre informazioni più specifiche, omettendo la compilazione del quesito filtro. Nella correzione basata sulla metodologia di Fellegi-Holt, la selezione casuale delle variabili da modificare in caso di soluzioni multiple, potrebbe portare anche ad un'eventuale correzione sulla variabile con l'informazione più puntuale, con relativa cancellazione dell'importo. Invece, la presenza dello stesso fa piuttosto presumere che l'errore di compilazione sia sul quesito filtro. Per tale ragione, al fine di ridurre il rischio di perdita di informazione durante la correzione di tipo probabilistico realizzata successivamente, basata sulla metodologia Fellegi-Holt, in presenza di situazioni di questa natura si è adottato un approccio deterministico di tipo "if-then" utilizzando alcune informazioni più puntuali per l'imputazione di variabili filtro.

Altre situazioni di errori corrette con un approccio di tipo deterministico sono quelle per cui, in presenza di ragionevoli informazioni di contesto, si può ipotizzare un'errata classificazione del tipo di reddito da parte del rispondente (esempio: pensioni da lavoro integrate al minimo considerate erroneamente pensioni sociali da alcuni rispondenti di età inferiore alla minima prevista per legge per poter beneficiare della pensione sociale). In tal caso la rimozione dell'errore implica lo spostamento di certi importi da una variabile ad un'altra.

Infine, sono realizzate deterministicamente anche le correzioni effettuate in seguito a controlli dei dati rilevati con le informazioni presenti per lo stesso rispondente in alcuni archivi amministrativi³ quale l'Anagrafe tributaria, disponibile presso l'Agenzia delle entrate del Ministero dell'economia, contenente i dati sulle dichiarazioni dei redditi dei contribuenti e il Casellario pensionistico, disponibile presso l'Inps, contenente informazioni sulle pensioni da lavoro, di reversibilità, sociali e di invalidità di tutti i beneficiari (ivi compresi i beneficiari di pensioni erogate da istituti di previdenza diversi dall'Inps). Tale correzione è resa possibile grazie all'abbinamento esatto dei record provenienti dai dati rilevati e da quelli di fonte amministrativa tramite i codici fiscali dei rispondenti. La correzione riguarda alcune tipologie di

³ Per un approfondimento si veda Consolini P. (2005), L'integrazione dei dati campionari con fonti amministrative.

reddito la cui rilevazione risulta più problematica: il reddito da lavoro autonomo, affetto da problemi di reticenza e le pensioni, per le quali sono frequenti errori di classificazione da parte del rispondente (per esempio il beneficiario di più di una tipologia di pensione, tende a classificare l'importo cumulato in un'unica tipologia).

4.3.6. *Correzioni e imputazioni di tipo probabilistico*

Come sopra accennato, la maggior parte delle incoerenze tra le risposte a quesiti diversi e quelle dovute ad un'errata compilazione dei percorsi sono trattate con la metodologia del minimo cambiamento di Fellegi-Holt, implementata nel sistema Scia (Sistema di controllo ed imputazione automatica) sviluppato presso l'Istituto (Istat, 2004).

Per la correzione delle variabili quantitative, che non sono trattabili in Scia, sono utilizzate delle variabili qualitative indicatrici della presenza o meno di un valore per la corrispondente variabile quantitativa. Dopo la fase di correzione probabilistica, il confronto tra i valori assunti dalle variabili indicatrici e la presenza o meno del valore nelle variabili quantitative consente di individuare i soggetti tenuti a rispondere su determinati quesiti di natura quantitativa. Se il valore non è stato indicato in fase di rilevazione, esso è considerato mancante e gli viene assegnato un valore possibile durante la successiva fase di imputazione delle variabili quantitative.

Per individuare le situazioni di errore e realizzare la correzione probabilistica è stato necessario definire le regole esplicite di incompatibilità tra i valori delle variabili. Esse sono di due tipi: gli *edit* formali, ovvero quelle regole che verificano l'errata compilazione del questionario in base ai salti definiti dalle variabili filtro, e gli *edit* sostanziali, ovvero quelle regole che verificano la mancata coerenza di combinazioni di risposta che scaturiscono da considerazioni di tipo statistico o conoscenze specifiche a priori sulle relazioni socio-economiche tra le variabili.

Per migliorare la qualità della correzione da donatore sono state definite delle variabili di strato, che servono per identificare i serbatoi dei donatori da cui attingere le informazioni da sostituire in un record errato. I donatori hanno per la variabile di stratificazione gli stessi valori osservati nel record ricevente.

La correzione probabilistica è realizzata separando le variabili familiari e quelle individuali.

La correzione ed imputazione delle variabili qualitative del questionario familiare avviene in un unico passo di correzione. La variabile di strato utilizzata è la regione di residenza della famiglia, a cui è stata attribuita fissità massima, vale a dire che il valore non è modificabile da parte del sistema Scia. Altra variabile a cui è assegnata la fissità massima è il numero di componenti della famiglia, in quanto corretta in una fase antecedente in modo deterministico.

La correzione delle variabili individuali avviene distintamente per gli individui eleggibili per l'intervista personale e per i non eleggibili. Per questi ultimi sono sottoposte a controllo e correzione le sole variabili della scheda familiare. Per i rimanenti, il piano di controllo e correzione riguarda non solo le informazioni della scheda familiare, ma anche quelle del questionario individuale.

Per gli individui con le sole informazioni della scheda familiare, la derivazione dell'insieme completo di regole di incompatibilità è stata possibile in un unico passo di correzione. La variabile di strato utilizzata per definire il serbatoio dei donatori è la ripartizione geografica. Le variabili età e sesso sono impostate a fissità massima, in quanto sono corrette in una fase antecedente in modo deterministico.

Per gli individui provvisti anche del questionario individuale, l'esecuzione in un unico passo di correzione che utilizza tutti gli edit formali e sostanziali non è stata possibile, data la complessità strutturale del questionario e i limiti computazionali dell'applicazione utilizzata.

Per questo motivo, la correzione dei dati è effettuata in diversi passi, suddividendo le regole in sottoinsiemi trattabili e aventi il minor numero di variabili in comune tra loro. A tal fine, si

procede dapprima con la correzione delle sezioni relative alla situazione lavorativa presente o passata. Le sezioni relative ai diversi tipi di redditi percepiti sono corrette successivamente a quelle sulla situazione lavorativa, in quanto presumibilmente l'intervistato risponde più facilmente alle domande sulla condizione professionale, mentre manifesta una certa reticenza a rispondere a domande sul reddito. Infine, si procede con la correzione delle rimanenti sezioni (ad esempio: istruzione, salute).

Il trattamento simultaneo delle sezioni sulla situazione lavorativa attuale e di quelle relative all'ultima attività svolta non è stato realizzabile a causa dell'impossibilità di generare l'insieme completo delle regole di incompatibilità, data la complessità strutturale di tali sezioni. Tuttavia, l'analisi dei percorsi di risposta da seguire nella compilazione del questionario ha consentito di individuare due gruppi di rispondenti, tenuti alla compilazione di sezioni alternative. Pertanto, il campione di individui è suddiviso in due sottoinsiemi: A) coloro che non hanno attualmente un'occupazione e hanno quindi compilato le sezioni relative all'ultima attività lavorativa svolta (sezione 7), ma non le sezioni sulla attività lavorativa attuale (sezione 5-6); B) coloro che, invece, sono in situazione occupazionale o in cassa integrazione guadagni per meno di tre mesi o con una retribuzione di almeno il 50 per cento e hanno compilato quindi le sezione 5-6 e non la sezione 7.

Nel caso in cui, per un'errata compilazione dei percorsi del questionario, siano compilate congiuntamente le sezione 5 e/o 6 e la sezione 7 non è possibile assegnare il rispondente a uno dei due gruppi A e B. Per realizzare tale assegnazione, preliminarmente al trattamento tramite Scia, è stata implementata una procedura per l'individuazione del percorso più verosimile per ciascun rispondente. Tale procedura si basa sull'analisi sia delle risposte fornite alle domande filtro sia della presenza/assenza di risposte ai quesiti interni alle sezioni stesse. Tali informazioni consentono di costruire delle variabili indicatrici della compilazione delle sezioni e di delineare quindi i percorsi di risposta forniti dai rispondenti. Per ciascun rispondente con percorso non ammissibile si procede ad individuare, sulla base di una funzione di distanza, il percorso ammissibile più vicino a quello effettuato e a sceglierne uno nel caso di soluzioni multiple. In quest'ultimo caso, l'assegnazione del percorso ammissibile avviene in modo casuale con probabilità pari alla frequenza relativa del percorso ammissibile di minima distanza osservata tra gli individui con percorso esatto appartenenti alla stessa categoria (stesso sesso e classe di età) dell'individuo da correggere.⁴

I due sottoinsiemi di individui sono sottoposti, separatamente, ad un ulteriore passo di correzione che risolve le incompatibilità tra le professioni svolte e il tipo di attività dell'unità locale in cui il rispondente lavora o lavorava.

Dopo questi due passi di correzioni paralleli, i due sottoinsiemi di individui sono aggregati per essere trattati congiuntamente nelle fasi di lavorazione relative alle sezioni sui redditi e alle altre sezioni.

In generale, nell'esecuzione di un determinato passo di correzione, le variabili incluse anche in un passo di correzione precedente sono trattate con fissità massima, vale a dire che non sono modificabili da parte del sistema.

Alcune variabili, inoltre, sono considerate più affidabili di altre e quindi è loro assegnata una fissità intermedia: nella fase di individuazione dell'insieme minimo di variabili da correggere, il sistema considera la variabile con fissità maggiore solo se non è in grado di individuare un insieme minimo con le variabili con fissità minore. Ad esempio, è assegnata una fissità intermedia alla condizione lavorativa principale nel periodo di riferimento del reddito, in quanto

⁴ Per un approfondimento si veda Bianchi G., Manzari A., Reale A., Siciliani I. (2004), Identificazione dei percorsi per i rispondenti dell'indagine trasversale sulle condizioni di vita Anno 2003.

essa risulta legata da specifiche regole di incompatibilità con le tipologie di reddito percepito, per le quali si potrebbe manifestare una maggiore reticenza nel rispondere. Le variabili anagrafiche sesso ed età hanno in tutti i passi di correzione fissità massima, in quanto corrette precedentemente. La variabile strato è costituita dalla ripartizione geografica.

4.4 Correzioni e imputazioni delle informazioni sui redditi

4.4.1. Premessa

Caratteristica peculiare dell'indagine Eu-Silc, rispetto ad altre indagini campionarie rivolte alle famiglie, è che una parte importante delle informazioni raccolte è di tipo quantitativo, con quesiti che riguardano direttamente gli importi dei redditi percepiti a livello individuale o familiare o delle spese inerenti l'abitazione di residenza. È richiesto, inoltre, che tali informazioni siano complete a livello di ogni singola unità di rilevazione, in quanto costituiscono la base per la costruzione di variabili sintetiche (le variabili target definite dal regolamento Eurostat) che riuniscono fonti di reddito diverse oppure aggregano le stesse informazioni a livello familiare.

Tali circostanze impongono l'adozione di strategie per l'individuazione/correzione dei dati anomali e per l'imputazione dei dati mancanti non usuali nel campo delle indagini sulle famiglie, dove è predominante l'impiego della metodologia Fellegi-Holt per l'imputazione dei valori mancanti e per la rimozione delle eventuali incompatibilità tra variabili di natura qualitativa.

In Eu-Silc, la metodologia Fellegi-Holt è stata utilizzata in modo particolare per definire, sulla base delle informazioni di contesto, l'insieme dei soggetti tenuti alla risposta ai diversi quesiti. La strategia adottata per il controllo e la correzione delle variabili quantitative dell'indagine è articolata in una serie di passi schematicamente sintetizzabili nel modo seguente:

- 1) individuazione ed eliminazione degli outlier;
- 2) correzioni deterministiche di alcuni errori sistematici;
- 3) imputazione dei valori mancanti.

I primi due passi, fondamentali per arrivare a disporre di informazioni il più possibile affidabili su cui basare il processo di imputazione dei valori mancanti, è stato affrontato con l'adozione di una procedura generalizzata per l'individuazione e la rimozione di eventuali dati anomali (*outlier*), che verrà esaminata più approfonditamente nel paragrafo 4.4.2, e con la realizzazione di alcune procedure ad hoc per l'individuazione e l'eliminazione di incoerenze tra i dati. In alcuni dei casi di incoerenza, come vedremo meglio nel paragrafo 4.4.4, si è potuto provvedere direttamente alla correzione dei valori anomali in modo deterministico, sulla scorta di altre informazioni, ritenute più affidabili, rilevate nel questionario. In altri casi, così come per gli *outlier*, si è preferito cancellare i dati portatori di incoerenze per sottoporli successivamente alle procedure di imputazione.

Per questo passaggio, l'ultimo precedente la verifica finale, l'approccio seguito è stato quello basato sulla costruzione di modelli di regressione delle variabili da imputare rispetto alle altre variabili disponibili. Secondo questo approccio, i valori da sostituire alle mancate risposte (o ai valori ritenuti non validi e di conseguenza eliminati) sono predetti dal modello in funzione delle informazioni possedute dal soggetto considerato.

Tale approccio è apparso preferibile a quello basato sulla tecnica del donatore, il quale consiste nell'imputazione dei valori prelevati dal soggetto più simile rispetto ad un insieme di variabili ritenute legate a quelle da imputare. Nel nostro caso, infatti, è presumibile che le mancate risposte riguardino in modo particolare classi di reddito specifiche, in particolare i redditi più elevati, nelle quali è più difficile reperire donatori in numero adeguato.

Nell'ambito delle tecniche di imputazione basate su modelli, la scelta è ricaduta sulla procedura Iweware, illustrata con maggiore dettaglio nel paragrafo 4.4.3, in quanto, raccomandata da Eurostat e già da questi utilizzata nell'indagine Echp, rende maggiormente praticabili i confronti con le stime prodotte da altri Paesi, ottenute a parità di metodologia. Occorre sottolineare, inoltre, la notevole semplicità di implementazione all'interno di procedure anche estremamente articolate e, soprattutto, la bontà dei risultati che, a nostro parere, ci ha consentito di ottenere.

4.4.2. L'individuazione e l'eliminazione degli outlier

Per ognuna delle variabili quantitative, i valori anomali che ci siamo proposti di eliminare sono stati sicuramente quelli così bassi o talmente elevati da apparire con molta probabilità non veri, dovuti ad errori di compilazione o di registrazione dei questionari. Soprattutto, però, ci siamo proposti di eliminare quei valori che, nonostante la loro scarsissima frequenza, per effetto della loro distanza dal centro della distribuzione, avrebbero portato a forti perturbazioni nei modelli di regressione costruiti ai fini dell'imputazione dei valori mancanti.

A questo fine, si è scelto di adottare una procedura generalizzata per il riconoscimento di valori *outlier*⁵ fondata sul metodo di Hidioglou-Berthelot⁶ il quale consiste, in primo luogo, in una trasformazione del carattere tale da rendere maggiormente simmetrica la sua distribuzione rispetto alla mediana (Me).

$$HB_Y = \begin{cases} \frac{Y - Me_Y}{Y} & 0 < Y < Me_Y \\ \frac{Me_Y - Y}{Me_Y} & Y \geq Me_Y \end{cases} \quad (1)$$

In un secondo momento vengono calcolate le soglie minima e massima al di fuori delle quali i valori registrati sono da considerare *outlier*. Tali soglie vengono determinate in funzione della distanza tra i quartili e il centro della distribuzione (la mediana) così da tenere conto della concentrazione del carattere:

$$HB_{\min}(K) = Me_{HB_Y} - K(Me_{HB_Y} - Q1_{HB_Y}) \quad HB_Y < Me_{HB_Y}$$

$$HB_{\max}(K) = Me_{HB_Y} + K(Q3_{HB_Y} - Me_{HB_Y}) \quad HB_Y \geq Me_{HB_Y} \quad (2)$$

⁵ L'implementazione del software e l'elaborazione dei risultati sono state curate da A. Rondinone.

⁶ Hidioglou M.A., Berthelot J.M. (1986), Statistical Editing and Imputation for Periodical Business Survey, Survey methodology, 12,1,73-83.

J.W. Hunt, J.S. Johnson, C.S. King (1999), Detecting outliers in the monthly retail trade survey using the Hidioglou-Berthelot method, Proceedings of the Survey Research Methods Section, American Statistical Association.

Considerando congiuntamente la (1) e la (2) le soglie possono scriversi nel modo seguente:

$$HB_{\min}(K) = \frac{Q1_Y Me_Y}{Q1_Y + K(Me_Y - Q1_Y)} \quad 0 < Y < Me_Y$$

$$HB_{\max}(K) = Me_Y + K(Q3_Y - Me_Y) \quad Y > Me_Y$$
(3)

La costante K, dopo numerose verifiche, è stata fissata pari a 40 per evidenziare e rimuovere soltanto i casi particolarmente rilevanti (Tavole 4.1 e 4.2).

Un possibile limite di questo metodo è costituito dal fatto che agisce secondo un ottica univariata.

Per alcune variabili, viceversa, è essenziale mettere in relazione il valore che è stato rilevato con altre informazioni che possano spiegarne la variabilità. È il caso, ad esempio, delle variabili familiari relative alle spese per l'abitazione, dove è stato posto al vaglio il rapporto con la superficie dell'abitazione e con il numero delle stanze. Analogamente sono stati valutati il rapporto tra reddito netto e lordo da lavoro dipendente e il rapporto tra i guadagni finanziari e i risparmi.

In questi casi, quando il rapporto in questione è stato trovato al di fuori delle soglie, si è optato per la rimozione del valore, tra i due, maggiormente sospetto di essere un *outlier* (in base alla distribuzione univariata). Nel caso in cui entrambi i valori siano risultati al di fuori di soglie calcolate in funzione di una costante K pari a 20, si è scelto di eliminare entrambi i valori.

La tavola 4.3 sintetizza l'incidenza del numero di casi eliminati a seguito dell'applicazione dei criteri precedentemente illustrati. Come si vede, l'incidenza di tali casistiche è risultata di entità trascurabile e ha riguardato soltanto casi particolarmente eclatanti. Soltanto per fare un esempio, è stato eliminato il caso di 15 euro di spese annuali di condominio per un'abitazione di 110 mq.

A partire dalla seconda tornata di indagine (Eu-Silc 2005), per i soggetti già intervistati l'anno precedente (circa il 75 per cento del campione totale) è stato inoltre possibile praticare il confronto tra le risposte fornite nelle due occasioni di indagine. In questi casi il metodo di Hidioglou-Berthelot è stato applicato alla distribuzione del rapporto Y_t/Y_{t-1} tra i valori assunti dalla stessa variabile nei due periodi, in modo del tutto analogo a quello utilizzato nel caso di rapporti tra variabili quantitative diverse: nel caso di rapporti al di fuori della soglia stabilita, si è scelto di considerare *outlier* il valore maggiormente decentrato nell'ambito della distribuzione univariata e di considerare *outlier* entrambi i valori quando sia l'uno che l'altro sono al di fuori di soglie calcolate in funzione di una costante K pari a 20.

Tavola 4.1 - Risultati dell'applicazione del metodo di Hidiroglou-Berthelot alle variabili quantitative individuali

VARIABILE	1° quartile	Mediana	3° quartile	Soglia minima	Soglia massima	Numero di casi totali	Sotto la soglia minima	Sopra la soglia massima	Totale casi eliminati	
									N.	%
Dipendenti- Retribuzione mensile lorda attuale	1.100	1.450	1.860	106	17.850	5.010	0	2	2	0,0
Dipendenti- Retribuzione mensile netta attuale	850	1.100	1.300	86	9.100	5.010	0	1	1	0,0
Dipendenti-Guadagno netto mensile	840	1.038	1.300	100	11.518	5.502	3	1	4	0,1
Dipendenti-Compensi aggiuntivi mensili	70	110	300	5	7.710	779	1	3	4	0,5
Dipendenti-Compensi aggiuntivi annuali	800	1.000	1.525	91	22.000	3.533	4	1	5	0,1
Dipendenti-Arretrati	400	838	1.350	19	21.338	207	1	0	1	0,5
Dipendenti- Liquidazione	600	1.500	5.000	25	141.500	366	3	4	7	1,9
Dipendenti-Assegni familiari	25	55	120	1	2.655	1.053	0	2	2	0,2
Dipendenti-Contributi volontari	28	50	120	1	2.850	409	1	3	4	1,0
Autonomi-Guadagno netto	2.000	8.000	15.000	66	288.000	2.047	1	0	1	0,0
Autonomi-Perdite	1.000	2.300	6.000	43	150.300	34	0	0	0	0,0
Autonomi-Diritti d'autore	500	1.000	3.000	24	81.000	10	0	0	0	0,0
Autonomi-Assegni familiari	15	33	150	1	4.713	27	0	1	1	3,7
Pensione sociale	227	300	350	22	2.300	208	0	0	0	0,0
Pensione da lavoro	450	700	1.000	30	12.700	3.787	0	0	0	0,0
Pensione di reversibilità	250	450	600	14	6.450	1.121	0	0	0	0,0
Pensione di invalidità	216	400	500	11	4.400	655	0	0	0	0,0
Pensione di accompagnamento	380	400	430	129	1.600	142	0	0	0	0,0
Assegni familiari pensionati	20	29	50	2	869	297	0	1	1	0,3
Pensione integrativa privata	400	480	568	53	4.000	207	0	0	0	0,0
Ammontare risparmi	3.000	7.500	15.490	123	327.100	5.660	13	5	18	0,3
Guadagni da attività finanziarie	100	300	900	4	24.300	2.816	14	4	18	0,6
Guadagno da terreni o fabbricati	1.266	2.591	5.000	60	98.951	986	6	0	6	0,6
Guadagno ipotetico da terreni o fabbricati	400	1.200	3.000	15	73.200	601	6	1	7	1,2
Imposta comunale sugli immobili (Ici)	80	150	290	4	5.750	7.766	3	4	7	0,1
Soldi a persone esterne alla famiglia	1.000	2.400	5.000	42	106.400	404	3	0	3	0,7
Soldi da persone esterne alla famiglia	700	2.400	4.800	24	98.400	414	1	0	1	0,2
Contributi volontari per pensione privata	80	110	195	7	3.510	1.259	2	5	7	0,6
Indennità di disoccupazione	300	500	740	18	10.100	260	0	0	0	0,0
Assegni per disoccupati	80	150	210	4	2.550	39	0	0	0	0,0
Cassa integrazione guadagni	413	625	800	29	7.625	69	0	0	0	0,0
Assegni familiari per Cig	50	60	180	7	4.860	14	0	0	0	0,0
Assegni o indennità	103	282	700	4	17.022	213	0	0	0	0,0
Borsa lavoro/ formazione	150	500	870	5	15.300	16	0	0	0	0,0
Borsa di studio	100	241	820	4	23.401	72	2	0	2	2,8

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2004

Tavola 4.2 - Risultati dell'applicazione del metodo di Hidiroglou-Berthelot alle variabili quantitative familiari

VARIABILE	1° quartile	Mediana	3° quartile	Soglia minima	Soglia massima	Numero di casi totali	Sotto la soglia minima	Sopra la soglia massima	Totale casi eliminati	
									N.	%
Guadagno da affitto di stanze	1.200	2.113	3.009	67	37.953	31	1	0	1	3,2
Condominio	300	600	1.218	15	25.300	2.924	5	3	8	0,3
Riscaldamento con condominio	300	525	800	17	11.525	531	1	1	2	0,4
Acqua con condominio	108	160	250	8	3.760	1.571	0	3	3	0,2
Gas con condominio	130	271	600	6	13.431	2.542	1	1	2	0,1
Gas	200	500	1.000	8	20.500	3.843	0	2	2	0,1
Riscaldamento	350	650	1.040	18	16.250	1.039	1	0	1	0,1
Legna, carbone	450	700	1.020	30	13.500	446	0	2	2	0,4
Acqua	95	150	240	6	3.750	4.485	2	1	3	0,1
Spurgo di pozzi	50	100	200	2	4.100	1.453	0	2	2	0,1
Elettricità	240	360	600	17	9.960	7.245	0	5	5	0,1
Nettezza urbana	100	140	200	8	2.540	7.409	1	4	5	0,1
Telefono fisso	300	400	600	28	8.400	5.729	0	2	2	0,0
Contributi pubblici per spese diverse dall'affitto	120	240	371	6	5.480	67	1	0	1	1,5
Spese di ristrutturazione	1.000	3.000	8.000	37	203.000	799	5	3	8	1,0
Trasferimenti	600	2.000	5.000	21	122.000	400	1	0	1	0,3
Stima dell'affitto	250	400	600	16	8.400	5.993	0	3	3	0,1
Affitto casa	150	256	365	9	4.616	1.416	2	1	3	0,2
Stima affitto al prezzo di mercato	250	350	500	21	6.350	423	0	0	0	0,0
Contributi pubblici per l'affitto	60	90	200	4	4.490	37	0	0	0	0,0
Ammontare mutuo	26.000	50.000	70.000	1.318	850.000	723	19	4	23	3,2
Rata annuale mutuo	3.000	4.800	6.800	192	84.800	723	5	1	6	0,8
Spesa per interessi del mutuo	700	1.549	2.900	31	55.589	240	1	0	1	0,4

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2004

Tavola 4.3 - Eliminazione di outlier tramite l'analisi di alcuni rapporti tra variabili

Variabile al numeratore Variabile 1	Variabile al denominatore Variabile 2	Casi eliminati	
		variabile 1	variabile 2
<i>Variabili Individuali</i>			
Reddito mensile lordo da lavoro dipendente	Reddito mensile netto da lavoro dipendente	4	3
Guadagni da attività finanziarie	Ammontare dei risparmi	2	0
<i>Variabili familiari</i>			
Affitto	N. di stanze	2	0
Mutuo	N. di stanze	1	0
Condominio	Metri quadri abitazione	1	0
Spesa per gas	Metri quadri abitazione	2	0
Spurgo pozzi, altri servizi	Metri quadri abitazione	1	0
Spesa per elettricità	Metri quadri abitazione	1	0
Spesa per nettezza urbana	Metri quadri abitazione	2	0
Contributi pubblici alle spese per la casa diverse dall'affitto	Metri quadri abitazione	1	0

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2004

4.4.3. La procedura di imputazione dei valori mancanti mediante la procedura Iweware

Le imputazioni dei valori mancanti sono state effettuate mediante il modulo Impute del pacchetto Iweware⁷ sviluppato nell'ambito del Survey methodology program presso il Centro di Ricerca sulle Indagini dell'Università del Michigan, Istituto per la Ricerca Sociale.

Il modulo in questione è costituito da una procedura generalizzata di imputazione multivariata in grado di trattare strutture di dati relativamente complesse sotto l'ipotesi Mar (*missing at random*).

A partire da una matrice di dati del tipo unità-variabili, la procedura imputa i valori mancanti di ciascuna unità condizionatamente a tutti gli altri valori osservati per quella unità, osservati o imputati. Ciò viene realizzato attraverso l'elaborazione di una sequenza di regressioni multiple, utilizzando modelli di regressione diversi a seconda delle tipologie di variabili oggetto di imputazione:

1. variabili continue, per le quali viene utilizzato il modello di regressione lineare classico. È necessario ricondurre alla normalità variabili che presentano distribuzioni di probabilità di tipo diverso (nel caso delle variabili di reddito, fortemente asimmetriche a sinistra, è stata considerata adeguata una trasformazione logaritmica). Dopo l'imputazione è poi necessario effettuare la trasformazione inversa;
2. variabili dicotomiche, per le quali viene utilizzato il modello logistico;
3. variabili politomiche, per le quali viene utilizzato il modello logistico polinomiale;
4. variabili di conteggio, per le quali viene utilizzato il modello di regressione di Poisson;
5. variabili di tipo misto, ovvero variabili continue con probabilità non nulla nell'origine. Queste vengono trattate in due passi: nel primo viene stabilito lo stato (0 o non 0) con l'applicazione di un modello di regressione logistica, in un secondo momento viene imputato un valore diverso da 0 ai valori mancanti mediante un modello di regressione lineare normale.

Il procedimento, di tipo iterativo, può essere schematizzato nel modo seguente.

Nella prima iterazione vengono imputati i valori mancanti di tutte le variabili a cominciare dalla variabile che ne presenta il minor numero. Per questa (che chiamiamo Y1) viene costruito il modello di regressione appropriato utilizzando come covariate la matrice X di tutte le variabili prive di valori mancanti. Le imputazioni sono generate dalla distribuzione predittiva a posteriori specificata dal modello di regressione, assumendo una distribuzione di probabilità a priori per i parametri di forma piatta (non informativa). Si passa quindi ad imputare la variabile Y2 (la successiva con minor numero di valori mancanti) sulla base della matrice (X, Y1) dove Y1 contiene anche valori imputati. Si continua con le variabili successive fino a riempire tutta la matrice dei dati.

Il processo di imputazione viene quindi ripetuto nelle iterazioni successive alla prima modificando l'insieme delle variabili indipendenti includendovi tutte le variabili con l'eccezione, ovviamente, di quella considerata come dipendente. Così, si costruisce il modello di regressione di Y1 su (X, Y2, Y3, ..., Yk), di Y2 su (X, Y1, Y3, ..., Yk) e così via.

Il ciclo termina una volta raggiunto il numero di iterazioni fissato (superiore o uguale a due) o al raggiungimento di risultati stabili per i valori imputati.

Nel processo di imputazione è possibile imporre delle restrizioni al campo di applicabilità di una variabile cosicché il modello di regressione possa essere adattato al sottoinsieme di unità appropriato. La variabile oggetto di restrizioni viene utilizzata come covariata per l'imputazione

⁷ T.E. Raghunathan, P.W. Solemberger, J. Van Hoewyk (2002), *IVeware: Imputation and Variance Estimation Software*, Survey Methodology Program, Survey Research Center, Institute for Social Research, University of Michigan.

delle altre variabili, accompagnata dalle variabili *dummy* derivate dalla definizione di applicabilità.

È anche possibile definire delle limitazioni ai valori che possono venire imputati alle variabili. In questo caso la procedura considera, per le variabili continue, la versione troncata del modello di regressione lineare classico o, per le variabili politomiche, la distribuzione multinomiale opportunamente condizionata.

Si è ritenuto di effettuare l'imputazione dei valori mancanti nelle diverse sezioni del questionario, omogenee per classi di rispondenti e relazioni tra le variabili, in modo separato considerando che la sovrapposizione di diverse sezioni, ovvero la presenza di individui con redditi provenienti da fonti diverse, risulta poco frequente. In questo modo si è potuta limitare la scelta delle variabili per la costruzione dei modelli di regressione tra quelle maggiormente correlate con le variabili oggetto di imputazione (nel Prospetto 4.1 vengono riportate le schede con il dettaglio delle variabili utilizzate nei modelli).

Le spese, i redditi e i trasferimenti a livello familiare sono stati trattati per primi in quanto contribuiscono a stabilire il tenore di vita della famiglia e quindi a spiegare il livello degli altri redditi. Le altre sezioni sono state affrontate partendo da quelle con minore incidenza di valori mancanti.

L'ordine che è stato seguito nel trattamento delle diverse sezioni è stato quindi il seguente:

1. Spese, redditi e trasferimenti a livello familiare
2. Redditi da lavoro dipendente
3. Redditi da pensione
4. Redditi da lavoro autonomo
5. Redditi finanziari
6. Redditi da fabbricati
7. Altri redditi.

Per ognuna delle sezioni nelle quali si è articolato il lavoro, sono stati effettuati confronti sui parametri distributivi (media, valore minimo, primo quartile, mediana, terzo quartile, massimo), prima e dopo gli interventi effettuati, in relazione alle principali caratteristiche di volta in volta considerate nei modelli.

Per alcune variabili (i redditi o le perdite da lavoro autonomo, l'ammontare dei risparmi a fine anno, i guadagni dai risparmi, i redditi da fabbricati) erano disponibili, in alternativa alle dichiarazioni degli importi esatti, le risposte per classi di importo. In tali casi è stata seguita una procedura in due passi:

1. dopo avere posto in classi di importo tutte le risposte puntuali disponibili, sono stati imputati da modello i valori di quelli che disponevano soltanto della risposta in classi, considerando quindi soltanto coloro che, in un modo o nell'altro, avessero fornito una risposta;
2. sono stati imputati da modello (con la stessa struttura di quello utilizzato nel primo passo) i valori di quelli che non avevano fornito risposta del tutto.

In questo modo si è voluta salvaguardare la variabilità delle risposte anche all'interno delle singole classi di importo, evitando l'appiattimento sul valore centrale della classe, utilizzando le informazioni di contesto opportunamente sintetizzate nei modelli. Nello stesso tempo, tale procedura ha permesso di risolvere il problema dell'attribuzione di un valore di classe alle classi estreme.

4.4.4 La correzione di alcuni errori sistematici

Esaurita la fase di correzione generalizzata degli *outlier*, contestualmente all'applicazione delle procedure di imputazione dei dati mancanti, sono state analizzate le risposte ai diversi quesiti per evidenziare eventuali anomalie da correggere mediante procedure ad hoc.

Tale analisi ha consentito di definire, in molti casi, gli estremi inferiore e superiore del campo di variazione delle variabili da imputare in funzione di altre informazioni presenti nel questionario.

In sintesi, le anomalie riscontrate hanno riguardato, sezione per sezione, i seguenti aspetti:

Schema 4.2 - Anomalie riscontrate

SEZIONE	NOMALIE TRATTATE CON METODO DETERMINISTICO	CORREZIONE ADOTTATA
Notizie sull'abitazione di residenza	Incoerenza tra anno di costruzione dell'immobile di residenza, anno a partire dal quale la famiglia vi abita ed età del componente più anziano	Sbiancamento casi anomali
Informazioni sui mutui che gravano sull'abitazione	Incoerenza tra le variabili del piano di ammortamento (importo del finanziamento, tasso di interesse, rata annuale, durata). Mancata risposta generalizzata sulle spese per interessi	Applicazione di un modello matematico per il calcolo del piano di ammortamento integrato con interventi di imputazione nei casi di risposta incompleta
Informazioni sul lavoro svolto	Inconsistenza tra l'età al primo lavoro, anni lavorati, anni di contribuzione ed età attuale	Sbiancamento casi anomali
Redditi correnti da lavoro dipendente	Reddito netto superiore al lordo	Inversione se il rapporto che ne deriva risulta congrui, o sbiancamento di entrambe i valori
Redditi finanziari	Totale delle percentuali di composizione del portafoglio finanziario diverso da 100	Quadratura del portafoglio. Applicazione dei tassi medi di rendimento differenziati per tipo di investimento agli stock di risparmio
	Guadagni finanziari non coerenti con l'ammontare dei risparmi e la composizione del portafoglio	Sbiancamento casi anomali
Altri redditi	Trasferimenti effettuati verso persone non coabitanti superiori alle somme di tutti i redditi percepiti	Sbiancamento casi anomali

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2004

4.4.5 I risultati del processo di correzione e imputazione

Si può concludere affermando che il processo di correzione e imputazione delle variabili quantitative dell'indagine, estremamente contenuto e ispirato da un generale criterio di prudenza, non ha dato luogo ad uno stravolgimento dei dati di partenza.

L'imputazione dei valori mancanti ha riguardato l'11,1 per cento dei casi relativi alle variabili individuali e il 4,1 per cento dei casi relativi alle variabili familiari mentre l'impatto delle procedure di correzione delle risposte ritenute errate è apparso globalmente ancora più modesto, con l'1,2 per cento dei casi nelle variabili individuali e lo 0,1 per cento dei casi in quelle a livello familiare.

Fanno eccezione i redditi da lavoro autonomo che hanno fatto registrare un'incidenza del 16,0 per cento di casi modificati.

Ciò che interessa maggiormente, ad ogni modo, è che l'applicazione delle procedure di imputazione e di correzione ai dati dell'indagine non ha determinato un impatto rilevante sulle distribuzioni dei redditi e delle altre variabili monetarie. Anche per il reddito da lavoro

autonomo, la variabile maggiormente coinvolta nelle correzioni, si può affermare che gli effetti siano stati comunque piuttosto moderati, con lo sgonfiamento della percentuale dei percettori dei redditi più bassi, redistribuita in modo uniforme sui redditi medio-alti. A questo ha necessariamente fatto seguito una rivalutazione delle posizioni media e mediana, cresciute rispettivamente del 20,0 per cento e del 15,2 per cento.

È utile sottolineare, infine, che la procedura adottata in Eu-Silc 2004, basata esclusivamente sui dati campionari, ha condotto a risultati non dissimili da quelli ottenuti nell'indagine Eu-Silc 2005, per la quale ci si è potuti avvalere dei dati di fonte amministrativa.

Nelle tavole seguenti vengono illustrati, con maggiore dettaglio, gli interventi effettuati nelle diverse fasi di correzione delle variabili quantitative (Tavole 4.4 e 4.5) e gli effetti sui loro principali parametri distributivi (Tavole 4.6 e 4.7).

Tavola 4.4 - Interventi sulle variabili individuali (in valore assoluto ed in percentuale)

VARIABILE	Valori invariati				Valori modificati		Valori mancanti imputati		Numero totale di individui	
	Importo puntuale		Importo in classi		N.	%	N.	%	N.	%
	N.	%	N.	%						
Dipendenti-Retribuzione mensile lorda attuale	4.468	89,2	-	-	37	0,7	505	10,1	5.010	100,0
Dipendenti-Retribuzione mensile netta attuale	4.802	95,8	-	-	32	0,6	176	3,5	5.010	100,0
Dipendenti-Guadagno netto mensile	5.035	91,5	-	-	-	-	467	8,5	5.502	100,0
Dipendenti-Compensi aggiuntivi mensili	738	94,7	-	-	-	-	41	5,3	779	100,0
Dipendenti-Compensi aggiuntivi annuali	3.403	96,3	-	-	-	-	130	3,7	3.533	100,0
Dipendenti-Arretrati	77	37,2	-	-	-	-	130	62,8	207	100,0
Dipendenti-Liquidazione	224	61,2	-	-	-	-	142	38,8	366	100,0
Dipendenti-Assegni familiari	1.003	95,3	-	-	-	-	16	1,5	1.052	100,0
Dipendenti-Contributi volontari	302	73,8	-	-	-	-	107	26,2	409	100,0
Autonomi-Guadagno netto	587	28,6	934	45,6	328	16	201	9,8	2.050	100,0
Autonomi-Perdite	10	32,3	21	67,7	-	-	-	-	31	100,0
Autonomi-Diritti d'autore	7	70	-	-	-	-	3	30	10	100,0
Autonomi-Assegni familiari	23	85,2	-	-	1	3,7	3	11,1	27	100,0
Pensione sociale	179	86,1	-	-	-	-	29	13,9	208	100,0
Pensione da lavoro	3.349	88,4	-	-	-	-	438	11,6	3.787	100,0
Pensione di reversibilità	1.069	95,4	-	-	-	-	52	4,6	1.121	100,0
Pensione di invalidità	641	97,9	-	-	-	-	14	2,1	655	100,0
Pensione di accompagnamento	99	69,7	-	-	-	-	43	30,3	142	100,0
Assegni familiari pensionati	271	91,2	-	-	-	-	26	8,8	297	100,0
Pensione integrativa privata	194	93,7	-	-	-	-	13	6,3	207	100,0
Ammontare risparmi	1.571	27,5	3.627	63,5	60	1,1	453	7,9	5.711	100,0
Guadagni da attività finanziarie	449	14,4	1.948	62,3	1	0	730	23,3	3.128	100,0
Guadagno da terreni o fabbricati	345	62,1	163	29,3	-	-	48	8,6	556	100,0
Guadagno ipotetico da terreni o fabbricati	534	88,9	-	-	-	-	67	11,1	601	100,0
Imposta comunale sugli immobili (Ici)	6.467	83,3	-	-	-	-	1.299	16,7	7.766	100,0
Soldi a persone esterne alla famiglia	293	78,6	-	-	21	5,6	59	15,8	373	100,0
Soldi da persone esterne alla famiglia	357	86,2	-	-	4	1	53	12,8	414	100,0
Contributi volontari per pensione privata	837	66,5	-	-	100	7,9	322	25,6	1.259	100,0
Indennità di disoccupazione	221	85	-	-	16	6,2	23	8,8	260	100,0
Assegni familiari per disoccupati	31	79,5	-	-	2	5,1	6	15,4	39	100,0
Cassa integrazione guadagni	58	84,1	-	-	-	-	11	15,9	69	100,0
Assegni familiari per Cig	13	92,9	-	-	-	-	1	7,1	14	100,0
Assegni o indennità	187	87,8	-	-	3	1,4	23	10,8	213	100,0
Borsa lavoro o per formazione	10	62,5	-	-	-	-	6	37,5	6	100,0
Borsa di studio	63	87,5	-	-	4	5,6	5	6,9	72	100,0
Totale	37.917	74,5	6.693	13,2	609	1,2	5.642	11,1	50.884	100,0

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2004

Tavola 4.5 - Interventi sulle variabili familiari (in valore assoluto ed in percentuale)

VARIABILE	Valori invariati				Valori modificati		Valori mancanti imputati		Numero totale di famiglie	
	Importo puntuale		Importo in classi		N.	%	N.	%	N.	%
	N.	%	N.	%						
Guadagno da affitto di stanze	23	74,2	-	-	1	3,2	7	22,6	31	100,0
Condominio	2.847	96,9	-	-	-	-	91	3,1	2.938	100,0
Riscaldamento con condominio	485	91	-	-	21	3,9	27	5,1		100,0
Acqua con condominio	1.577	99,8	-	-	-	-	3	0,2	1.580	100,0
Gas con condominio	2.496	97,8	-	-	-	-	57	2,2	2.553	100,0
Gas	3.809	98,9	-	-	-	-	44	1,1	3.853	100,0
Riscaldamento	1.028	98,9	-	-	-	-	11	1,1	1.039	100,0
Legna, carbone	406	90,8	-	-	9	2	32	7,2	447	100,0
Acqua	4.150	92,3	-	-	-	-	346	7,7	4.496	100,0
Spurgo di pozzi	1.285	88,2	-	-	-	-	172	11,8	1.457	100,0
Elettricità	7.178	98,8	-	-	-	-	90	1,2	7.268	100,0
Nettezza urbana	6.846	92,1	-	-	-	-	588	7,9	7.434	100,0
Telefono fisso	5.666	98,6	-	-	-	-	82	1,4	5.748	100,0
Contributi pubblici spese diverse dall'affitto	65	97	-	-	-	-	2	3	67	100,0
Spese di ristrutturazione	800	99,5	-	-	-	-	4	0,5	804	100,0
Trasferimenti	347	85,5	-	-	42	10,3	17	4,2	406	100,0
Stima dell'affitto	5.748	95,6	-	-	-	-	265	4,4	6.013	100,0
Affitto casa	1.385	97,5	-	-	-	-	36	2,5	1.421	100,0
Stima affitto al prezzo di mercato	406	95,8	-	-	-	-	18	4,2	424	100,0
Contributi pubblici per l'affitto	37	100	-	-	-	-			37	100,0
Ammontare mutuo	650	89,3	-	-	-	-	78	10,7	728	100,0
Rata annuale mutuo	685	94,1	-	-	-	-	43	5,9	728	100,0
Spesa per interessi mutuo	224	92,9	-	-	-	-	17	7,1	241	100,0
Totale	48.143	96,8	0	0	73	0,1	2.030	4,1	49.713	100,0

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2004

Tavola 4.6 - Effetti delle procedure di correzione e imputazione sulle variabili individuali

VARIABILE	Prima				Dopo			
	Numero di individui	Mediana	Media	Deviazione standard	Numero di individui	Mediana	Media	Deviazione standard
Dipendenti-Retribuzione mensile lorda attuale	4.505	1.450	1.586	845,6	5.010	1.450	1.581	844,0
Dipendenti-Retribuzione mensile netta attuale	4.834	1.100	1.139	502,6	5.010	1.100	1.135	503,1
Dipendenti-Guadagno netto mensile	5.035	1.033	1.124	592,5	5.502	1.007	1.117	589,7
Dipendenti-Compensi aggiuntivi mensili	738	110	332	627,4	779	120	336	621,6
Dipendenti-Compensi aggiuntivi annuali	3.403	1.000	1.558	2.125,6	3.533	1.000	1.554	2.100,8
Dipendenti-Arretrati	77	775	1.329	1.867,6	207	500	1.811	3.408,0
Dipendenti-Liquidazione	224	1.500	5.821	14.367,5	366	1.399	6.023	14.576,3
Dipendenti-Assegni familiari	1.019	55	100	144,7	1.052	54	90	98,3
Dipendenti-Contributi volontari	302	50	145	313,8	409	50	177	412,7
Autonomi-Guadagno netto	1.849	10.000	13.474	14.540,9	2.050	12.000	15.525	16.994,3
Autonomi-Perdite	31	4.000	10.526	22.083,3	31	4.000	10.526	22.083,3
Autonomi-Diritti d'autore	7	1.000	1.521	1.134,3	10	988	1.554	1.083,0
Autonomi-Assegni familiari	24	33	128	299,5	27	20	81	121,8
Pensione sociale	179	300	282	79,3	208	300	279	88,5
Pensione da lavoro	3.349	700	821	606,7	3.787	700	813	599,8
Pensione di reversibilità	1.069	450	483	369,3	1.121	450	484	376,2
Pensione di invalidità	641	400	403	274,6	655	400	406	287,8
Pensione di accompagnamento	99	420	454	107,2	142	420	475	121,1
Assegni familiari pensionati	271	28	50	66,8	297	30	52	70,0
Pensione integrativa privata	194	480	531	340,8	207	465	553	472,6
Ammontare risparmi	5.258	8.685	16.645	24.190,1	5.711	8.218	16.842	25.376,8
Guadagni da attività finanziarie	2.398	200	764	1.509,7	3.128	300	819	1.526,6
Guadagno da terreni o fabbricati	508	3.000	5.851	10.336,6	986	393	3.400	8.261,4
Guadagno ipotetico da terreni o fabbricati	534	1.200	2.677	4.678,8	534	1.200	2.677	4.678,8
Imposta comunale sugli immobili (Ici)	6.467	150	251	352,1	6.467	150	251	352,1
Soldi a persone esterne alla famiglia	314	2.400	4.165	5.990,0	392	2.000	3.153	3.483,1
Soldi da persone esterne alla famiglia	361	2.400	3.364	4.505,8	414	2.000	2.998	3.267,1
Contributi volontari per pensione privata	937	110	299	512,7	1.259	108	162	165,0
Indennità di disoccupazione	237	500	764	893,4	260	500	561	437,8
Assegni familiari per disoccupati	33	150	184	165,0	39	150	168	123,9
Cassa integrazione guadagni	58	625	640	260,6	69	630	693	356,1
Assegni familiari per Cig	13	60	104	87,4	14	65	102	84,2
Assegni o indennità	190	282	511	616,8	213	258	467	511,0
Borsa lavoro o per formazione	10	500	564	455,0	16	835	821	589,2
Borsa di studio	67	241	553	736,6	72	200	522	719,7

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2004

Tavola 4.7 - Effetti delle procedure di correzione e imputazione sulle variabili familiari

VARIABILE	Prima				Dopo			
	Numero di individui	Mediana	Media	Deviazione standard	Numero di individui	Mediana	Media	Deviazione standard
Guadagno da affitto di stanze	24	2.113	2.661	2.136,1	31	2.113	2.557	2.411,2
Condominio	2.847	600	878	892,9	2.938	600	879	896,0
Riscaldamento con condominio	506	525	597	414,6	533	515	600	406,7
Acqua con condominio	1.577	160	205	167,2	1.580	160	205	167,2
Gas con condominio	2.496	271	421	410,4	2.553	280	425	415,4
Gas	3.809	500	699	655,3	3.853	500	697	656,0
Riscaldamento	1.028	650	826	656,5	1.039	640	822	655,0
Legna, carbone	415	700	877	659,9	447	700	875	629,2
Acqua	4.150	150	190	165,8	4.496	150	188	164,2
Spurgo di pozzi	1.285	100	183	298,0	1.457	100	182	289,5
Elettricità	7.178	360	461	365,1	7.268	360	461	368,6
Nettezza urbana	6.846	140	166	114,5	7.434	140	165	114,0
Telefono fisso	5.666	400	472	312,2	5.748	400	471	311,5
Contributi pubblici spese diverse dall'affitto	65	247	292	254,8		247	362	652,1
Spese di ristrutturazione	800	3.000	7.897	13.781,1	804	3.000	7.876	13.752,2
Trasferimenti	389	2.000	3.560	5.098,4	406	2.240	4.712	8.201,0
Stima dell'affitto	5.748	400	473	349,9	6.013	400	476	349,8
Affitto casa	1.385	256	283	197,1	1.421	258	286	199,0
Stima affitto al prezzo di mercato	406	350	390	219,6	424	350	393	230,1
Contributi pubblici per l'affitto	37	90	190	264,9	37	90	190	264,9
Ammontare mutuo	650	50.000	54.411	37.591,5	650	50.000	54.411	37.591,5
Rata annuale mutuo	685	4.800	5.378	3.558,6	685	4.800	5.378	3.558,6
Spesa per interessi mutuo	224	1.558	2.080	1.931,0	224	1.558	2.080	1.931,0

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2004

Prospetto 4.1 - Le variabili utilizzate nei modelli per l'imputazione dei valori mancanti

Scheda I - Spese, redditi e trasferimenti a livello familiare

Variabili da imputare

- Spese per l'abitazione
 - Spese di condominio
 - Spese di riscaldamento non comprese nel condominio
 - Spese per la fornitura di acqua non comprese nel condominio
 - Spese per la fornitura di gas non comprese nel condominio
 - Spese per la fornitura di gas delle famiglie che non vivono in condominio
 - Spese di riscaldamento non comprese nel consumo di gas delle famiglie che non vivono in condominio
 - Spese di riscaldamento delle famiglie che non hanno spese di gas e che non vivono in condominio
 - Spese per la fornitura di acqua delle famiglie che non vivono in condominio
 - Spese per spurgo pozzi o fogne, cura del giardino, servizi di vigilanza, eccetera
 - Spese per la fornitura di elettricità (esclusa quella destinata a riscaldamento)
 - Spese per la nettezza urbana
 - Spese telefoniche
 - Spese per la ristrutturazione del palazzo o della casa di abitazione
 - Spese per l'affitto
 - Spese per affitto ipotetiche (considerata la situazione di mercato) delle famiglie che vivono in un'abitazione di proprietà
 - Spese per affitto al prezzo di mercato per le famiglie che ritengono di pagare un affitto a condizioni favorevoli
- Redditi e trasferimenti a livello familiare
 - Contributi pubblici per coprire in tutto o in parte le spese per la casa diverse dall'affitto
 - Trasferimenti ricevuti con regolarità dalla famiglia da persone non coabitanti
 - Contributi pubblici per coprire in tutto o in parte le spese di affitto
 - Reddito derivante dall'affitto di parti dell'abitazione di residenza

Altre variabili indipendenti

- Caratteristiche territoriali
 - Ripartizione geografica
 - Tipologia e dimensione del comune di residenza
 - Problemi nella zona in cui vive la famiglia: inquinamento, rumori dai vicini o dall'esterno, criminalità, violenza, vandalismo
- Caratteristiche della famiglia e dell'abitazione
 - Numero di componenti la famiglia
 - Tipologia dell'abitazione
 - Titolo di godimento dell'abitazione
 - Superficie dell'abitazione
 - Numero di stanze dell'abitazione
 - Presenza di: cucina abitabile, gabinetto interno all'abitazione, vasca o doccia, due o più bagni, cantina, solaio o soffitta, terrazza o balcone, giardino privato, orto, acqua calda, aria condizionata garage privato o posto auto coperto
 - Anno a partire dal quale la famiglia vive nell'abitazione

- Anno di costruzione dell'abitazione
- Problemi relativi all'abitazione: tetti, soffitti, porte, finestre o pavimenti danneggiati, umidità nei muri, nei pavimenti, nei soffitti o nelle fondamenta, scarsa illuminazione, distanza dagli altri familiari, distanza dal lavoro
- Disponibilità di beni durevoli: lavatrice, televisore a colori, personal computer, telefono cellulare, automobile privata, lavastoviglie, frigorifero, videocamera, antenna parabolica, impianto Hi-Fi, console per videogiochi, modem, accesso a internet, segreteria telefonica, fax, bicicletta, motorino, motocicletta
- Segnali di difficoltà economiche delle famiglie
 - Esistenza di debiti diversi dal mutuo e loro gravosità
 - Situazioni di difficoltà con ritardo nei pagamenti nel pagare l'affitto, il mutuo, le bollette, i debiti diversi dal mutuo
 - Situazioni di difficoltà a causa di mancanza di denaro per comprare il cibo necessario, comprare i vestiti di cui c'era bisogno, pagare le spese per la scuola, pagare le spese per i trasporti, pagare le tasse, fare fronte a spese impreviste
 - Grado di difficoltà nell'arrivare a fine mese con i redditi disponibili
 - Gravosità delle spese per l'affitto

Scheda 2 - Redditi da lavoro dipendente

Variabili da imputare

- Redditi correnti mensili da lavoro dipendente (considerati anche le eventuali ore di straordinario abituale)
 - Reddito netto mensile
 - Reddito lordo mensile
- Redditi netti da lavoro dipendente nell'anno precedente la rilevazione
 - Stipendio netto mensile
 - Compensi aggiuntivi medi su base mensile: straordinari, indennità per lavoro serale o notturno, a turni o in periodi festivi, eccetera
 - Compensi aggiuntivi su base annuale: altre mensilità (tredicesima, quattordicesima, eccetera), premi ed incentivi di produttività, di risultato, per progetti speciali, eccetera
 - Arretrati da lavoro
 - Liquidazioni da lavoro
 - Assegni familiari mensili
 - Contributi volontari mensili per un fondo pensione aziendale (pensione integrativa)

Altre variabili indipendenti

- Caratteristiche territoriali
 - Ripartizione geografica
 - Tipologia e dimensione del comune di residenza
 - Presenza di problemi nella zona in cui vive la famiglia (inquinamento, rumori o criminalità)
- Caratteristiche della famiglia e dell'abitazione
 - Numero di componenti la famiglia
 - Tipologia dell'abitazione
 - Titolo di godimento dell'abitazione

- Superficie dell'abitazione
- Numero di stanze dell'abitazione
- Disponibilità di più bagni nell'abitazione
- Disponibilità di beni durevoli (lavatrice, lavastoviglie, automobile, antenna parabolica, personal computer o accesso a internet)
- Grado di difficoltà nell'arrivare a fine mese con i redditi disponibili
- Caratteristiche individuali
 - Sesso
 - Età
 - Stato civile
 - Livello di istruzione
 - Condizioni generali di salute
- Caratteristiche della professione
 - Professione (prima cifra della codifica Cp2001)
 - Settore di attività (pubblico o privato)
 - Ore di lavoro settimanali
 - Svolgimento di più lavori contemporaneamente
 - Cambiamento di lavoro negli ultimi 12 mesi
 - Stabilità del rapporto di lavoro (a termine o a tempo indeterminato)
 - Esistenza di un contratto per i rapporti di lavoro a termine
 - Incarico di coordinare il lavoro di altri
 - Facoltà di prendere decisioni sullo stipendio o la promozione del personale coordinato
 - Posizione nella professione
 - Età in cui si è iniziato a lavorare
 - Numero di anni di lavoro
 - Numero di anni di contributi versati

Scheda 3 - Redditi da pensione

Variabili da imputare

- Redditi netti mensili da pensione nell'anno precedente la rilevazione
 - Pensione sociale
 - Pensione da lavoro (anzianità o vecchiaia)
 - Pensione di reversibilità
 - Pensione di invalidità
 - Assegno di accompagnamento o altri sussidi per l'invalidità
 - Assegni familiari
 - Pensione integrativa privata

Altre variabili indipendenti

- Caratteristiche territoriali
 - Ripartizione geografica
 - Tipologia e dimensione del comune di residenza
 - Presenza di problemi nella zona in cui vive la famiglia (inquinamento, rumori o criminalità)
- Caratteristiche della famiglia e dell'abitazione

- Numero di componenti la famiglia
- Tipologia dell'abitazione
- Titolo di godimento dell'abitazione
- Superficie dell'abitazione
- Numero di stanze dell'abitazione
- Disponibilità di beni durevoli (videoregistratore, impianto Hi-Fi)
- Grado di difficoltà nell'arrivare a fine mese con i redditi disponibili
- Spese annuali per l'abitazione sostenute dalla famiglia
 - Rate di mutuo
 - Spese di condominio
 - Spese per la fornitura di gas delle famiglie che non vivono in condominio
 - Spese per la fornitura di elettricità (esclusa quella destinata a riscaldamento)
 - Spese per la nettezza urbana
 - Spese telefoniche
 - Spese per l'affitto
 - Spese per l'affitto ipotetiche (considerata la situazione di mercato) delle famiglie che vivono in un'abitazione di proprietà
- Caratteristiche individuali
 - Sesso
 - Età
 - Stato civile
 - Livello di istruzione
 - Condizioni generali di salute
- Altre fonti di reddito
 - Reddito mensile da lavoro dipendente
 - Reddito annuale da lavoro autonomo
 - Redditi annuali da risparmi finanziari
- Caratteristiche del lavoro svolto attualmente
 - Professione (prima cifra della codifica Cp2001)
 - Tipo di lavoro svolto
 - Ore di lavoro settimanali
 - Posizione nella professione dei dipendenti
- Caratteristiche del lavoro svolto in passato
 - Professione (prima cifra della codifica Cp2001)
 - Tipo di lavoro svolto
 - Posizione nella professione dei dipendenti
 - Esistenza di un contratto per i rapporti di lavoro a termine
 - Incarico di coordinare il lavoro di altri
 - Età in cui si è iniziato a lavorare
 - Numero di anni di lavoro
 - Numero di anni di contributi versati

Scheda 4 - Redditi da lavoro autonomo

Variabili da imputare

- Redditi netti da lavoro autonomo nell'anno precedente la rilevazione
 - Guadagni personali da lavoro autonomo annuali

- Assegni familiari mensili

Altre variabili indipendenti

- Caratteristiche territoriali
 - Ripartizione geografica
 - Tipologia e dimensione del comune di residenza
- Caratteristiche della famiglia e dell'abitazione
 - Numero di componenti la famiglia
 - Tipologia dell'abitazione
 - Titolo di godimento dell'abitazione
 - Superficie dell'abitazione
 - Numero di stanze dell'abitazione
 - Disponibilità di più bagni nell'abitazione
 - Disponibilità di beni durevoli (lavatrice, lavastoviglie, automobile, antenna parabolica, personal computer o accesso a internet)
 - Grado di difficoltà nell'arrivare a fine mese con i redditi disponibili
- Spese annuali per l'abitazione sostenute dalla famiglia
 - Rata di mutuo
 - Spese di condominio
 - Spese per la fornitura di gas delle famiglie che non vivono in condominio
 - Spese per la fornitura di elettricità (esclusa quella destinata a riscaldamento)
 - Spese per la nettezza urbana
 - Spese telefoniche
 - Spese per l'affitto
 - Spese per l'affitto ipotetiche (considerata la situazione di mercato) delle famiglie che vivono in un'abitazione di proprietà
- Caratteristiche individuali
 - Sesso
 - Età
 - Stato civile
 - Livello di istruzione
 - Condizioni generali di salute
 - Persona responsabile dell'abitazione
 - Necessità di ricorrere al dentista ed eventuale motivo per il mancato ricorso
 - Necessità di ricorrere a visite specialistiche ed eventuale motivo per il mancato ricorso
- Caratteristiche della professione
 - Professione (prima cifra della codifica Cp2001)
 - Professione svolta in passato
 - Ramo di attività (codifica Ateco)
 - Ramo di attività svolta in passato
 - Ore di lavoro settimanali
 - Svolgimento di più lavori contemporaneamente
 - Cambiamento di lavoro negli ultimi 12 mesi
 - Tipologia di lavoro autonomo
 - Disponibilità di dipendenti
 - Tipologia di lavoro autonomo svolto in passato
 - Età in cui si è iniziato a lavorare

Variabili da imputare

- Redditi da attività finanziarie nell'anno precedente la rilevazione
 - Ammontare delle attività finanziarie a fine anno precedente la rilevazione
 - Guadagni netti dagli investimenti finanziari nell'anno precedente la rilevazione

Altre variabili indipendenti

- Caratteristiche territoriali
 - Ripartizione geografica
 - Tipologia e dimensione del comune di residenza
- Caratteristiche della famiglia e dell'abitazione
 - Numero di componenti la famiglia
 - Tipologia dell'abitazione
 - Titolo di godimento dell'abitazione
 - Superficie dell'abitazione
 - Numero di stanze dell'abitazione
 - Disponibilità di più bagni, aria condizionata o garage nell'abitazione
 - Disponibilità di beni durevoli (videocamera, antenna parabolica)
 - Grado di difficoltà nell'arrivare a fine mese con i redditi disponibili
- Composizione del portafoglio finanziario
 - Percentuale di risparmi investiti in c/c
 - Percentuale di risparmi investiti in libretti di risparmio, certificati di deposito, buoni fruttiferi
 - Percentuale di risparmi investiti in titoli di stato
 - Percentuale di risparmi investiti in azioni, obbligazioni, fondi comuni o altri strumenti finanziari
 - Percentuale di risparmi investiti in partecipazioni a società non quotate o in prestiti a cooperative
- Altre fonti di reddito
 - Reddito mensile da lavoro dipendente
 - Reddito annuale da lavoro autonomo
 - Reddito da pensione di lavoro
 - Redditi da fabbricati
 - Contributi volontari dei dipendenti
 - Contributi per pensioni integrative
- Spese annuali per l'abitazione sostenute dalla famiglia
 - Rate di mutuo
 - Spese di condominio
 - Spese per la fornitura di gas delle famiglie che non vivono in condominio
 - Spese per la fornitura di elettricità (esclusa quella destinata a riscaldamento)
 - Spese per la nettezza urbana
 - Spese telefoniche
 - Spese per l'affitto
 - Spese per l'affitto ipotetiche (considerata la situazione di mercato) delle famiglie che vivono in un'abitazione di proprietà
- Caratteristiche individuali
 - Sesso
 - Età

- Stato civile
- Livello di istruzione
- Fonte principale di reddito
- Condizione lavorativa attuale
- Condizione lavorativa in passato

Scheda 6 - Redditi da fabbricati

Variabili da imputare

- Redditi da fabbricati nell'anno precedente la rilevazione
 - Guadagni per affitto di terreni o abitazioni di proprietà
 - Guadagni ipotetici per affitto di terreni o abitazioni di proprietà tenuti a disposizione (affitti imputati)

Altre variabili indipendenti

- Caratteristiche territoriali
 - Ripartizione geografica
 - Tipologia e dimensione del comune di residenza
- Caratteristiche della famiglia e dell'abitazione
 - Numero di componenti la famiglia
 - Tipologia dell'abitazione
 - Titolo di godimento dell'abitazione
 - Superficie dell'abitazione
 - Numero di stanze dell'abitazione
 - Disponibilità di più bagni, aria condizionata o garage nell'abitazione
 - Disponibilità di beni durevoli (videocamera, antenna parabolica)
 - Grado di difficoltà nell'arrivare a fine mese con i redditi disponibili
- Composizione del portafoglio finanziario
 - Percentuale di risparmi investiti in c/c
 - Percentuale di risparmi investiti in libretti di risparmio, certificati di deposito, buoni fruttiferi
 - Percentuale di risparmi investiti in titoli di stato
 - Percentuale di risparmi investiti in azioni, obbligazioni, fondi comuni o altri strumenti finanziari
 - Percentuale di risparmi investiti in partecipazioni a società non quotate o in prestiti a cooperative
- Altre fonti di reddito
 - Reddito mensile da lavoro dipendente
 - Reddito annuale da lavoro autonomo
 - Reddito da pensione di lavoro
 - Redditi da fabbricati
 - Contributi volontari dei dipendenti
 - Contributi per pensioni integrative
- Spese annuali per l'abitazione sostenute dalla famiglia
 - Rate di mutuo
 - Spese di condominio
 - Spese per la fornitura di gas delle famiglie che non vivono in condominio
 - Spese per la fornitura di elettricità (esclusa quella destinata a riscaldamento)

- Spese per la nettezza urbana
- Spese telefoniche
- Spese per l'affitto
- Spese per l'affitto ipotetiche (considerata la situazione di mercato) delle famiglie che vivono in un'abitazione di proprietà
- Caratteristiche individuali
 - Sesso
 - Età
 - Stato civile
 - Livello di istruzione
 - Fonte principale di reddito
 - Condizione lavorativa attuale
 - Condizione lavorativa in passato
- Caratteristiche del lavoro svolto attualmente
 - Professione (prima cifra della codifica Cp2001)
 - Settore di attività economica
 - Ore di lavoro settimanali
- Caratteristiche del lavoro svolto in passato
 - Professione (prima cifra della codifica Cp2001)
 - Settore di attività economica
 - Numero di anni di lavoro
 - Numero di anni di contributi versati

Scheda 7 - Altri redditi

Variabili da imputare

- Redditi da fonti diverse nell'anno precedente la rilevazione
 - Diritti d'autore
 - Trasferimenti regolari a persone non coabitanti (assegno per l'ex-coniuge)
 - Trasferimenti regolari da persone non coabitanti (assegno dall'ex-coniuge)
 - Contributi volontari per una pensione integrativa privata
 - Indennità di disoccupazione, di mobilità, di prepensionamento
 - Assegni familiari che spettano ai lavoratori disoccupati
 - Cassa integrazione
 - Assegni famigliari che spettano ai lavoratori in cassa integrazione
 - Assegno per famiglie con 3 o più figli
 - Assegno di maternità per la nascita di figli
 - Minimo vitale, reddito minimo di inserimento o altri aiuti in denaro da enti pubblici per le famiglie in difficoltà economica
 - Borsa lavoro o assegno per la formazione o l'inserimento professionale
 - Borsa di studio

Altre variabili indipendenti

- Caratteristiche territoriali
 - Ripartizione geografica
 - Tipologia e dimensione del comune di residenza
- Caratteristiche della famiglia e dell'abitazione

- Numero di componenti la famiglia
- Tipologia dell'abitazione
- Titolo di godimento dell'abitazione
- Superficie dell'abitazione
- Numero di stanze dell'abitazione
- Disponibilità di più bagni nell'abitazione
- Disponibilità di beni durevoli (lavatrice, lavastoviglie, automobile, antenna parabolica, personal computer o accesso a internet)
- Grado di difficoltà nell'arrivare a fine mese con i redditi disponibili
- Caratteristiche individuali
 - Sesso
 - Età
 - Stato civile
 - Livello di istruzione
 - Condizioni generali di salute
- Caratteristiche della professione
 - Professione (prima cifra della codifica Cp2001)
 - Professione svolta in passato (prima cifra della codifica Cp2001)
 - Settore di attività (pubblico o privato)
 - Posizione nella professione attuale
 - Posizione nella professione in passato
 - Età in cui si è iniziato a lavorare
 - Numero di anni di lavoro
 - Numero di anni di contributi versati

4.5 L'integrazione dei dati campionari con fonti amministrative

4.5.1 Premessa

Il paragrafo che segue descrive lo stato dell'arte in tema di integrazione di dati campionari ed amministrativi relativo alla prima edizione 2004 dell'indagine Eu-Silc.

Il principale obiettivo del progetto Eu-Silc consiste nella rilevazione dei redditi individuali e familiari netti sia in complesso che in forma disaggregata per componente economica.⁸ La rilevazione "mediante intervista" delle componenti di reddito percepite dai soggetti e dalla famiglia nell'anno di riferimento solleva il problema della verifica della qualità del dato. Al riguardo, il problema è rappresentato dagli errori di risposta connessi a questo tipo di rilevazione. L'incapacità dell'intervistato di quantificare con precisione eventi distanti nel tempo o semplicemente di ricordarne l'esistenza (in linea teorica potrebbero trascorrere 12 mesi dal momento dell'intervista all'anno di riferimento dell'indagine), l'effetto di interazione intervistato-intervistatore, la reticenza a dichiarare il vero, costituiscono alcuni degli esempi di questi potenziali errori. Per altro verso, l'indagine ammette anche l'utilizzo di risposte proxy, con conseguente peggioramento della qualità dell'informazione fornita dall'intervistato per conto dell'unità di osservazione. Agli errori esposti si affiancano altre tipologie errore di natura non campionaria, quali errori di registrazione, di codifica, l'auto-selezione dei rispondenti, eccetera.

⁸ Le principali componenti di reddito sono: da lavoro dipendente, da lavoro autonomo, da capitale reale, da capitale finanziario, da trasferimenti distinti a loro volta in pensioni e altri sussidi.

Per limitare l'impatto degli errori di misurazione sui redditi riportati nel questionario di ciascun percettore e, più in generale, per migliorare la qualità dei dati, è stato avviato un progetto che prevede l'utilizzo di microdati di fonte fiscale (limitatamente a talune componenti reddituali) in fase di controllo e correzione dei dati.⁹ Affinché si possa implementare un simile piano di correzione si deve tener conto di alcuni requisiti di base (cfr. van der Laan P., 2000). Le unità statistiche devono essere definite allo stesso modo nelle diverse fonti di dati. Tutte le fonti devono fare riferimento alla stesso collettivo statistico (copertura). Il contenuto informativo della variabili della fonte "donatrice" deve essere identico a quello delle variabili "obiettivo" dell'indagine Eu-Silc (armonizzazione delle variabili e delle classificazioni).

La tecnica utilizzata per integrare le fonti fiscali ai dati dell'indagine campionaria è quella *exact matching* (abbinamento esatto) che permette di collegare informazioni relative alla stessa unità statistica da distinte fonti del dato. Esistono nella realtà diverse tipologie di abbinamento esatto, nel nostro caso ci riferiamo al semplice abbinamento "uno a uno", dove cioè ad ogni unità statistica corrisponde un solo record in ognuna delle fonti da integrare. L'abbinamento fra diverse fonti del dato avviene, nel nostro caso, tramite l'utilizzo di una "chiave di abbinamento" (*matching key*) che identifica univocamente l'unità statistica comune ai diversi giacimenti informativi. Si tratta dunque di un abbinamento di tipo deterministico realizzato attraverso la chiave del codice fiscale dell'individuo. Una volta effettuato l'aggancio con i dati amministrativi, il codice fiscale viene eliminato dalle basi di dati integrate, in ragione delle disposizioni in materia di protezione dei dati personali (cfr. Dlgs n.196/2003).

Le fonti fiscali deputate per il *matching* con i dati individuali dei redditi Eu-Silc sono: 1) il modello Cud per le categorie di redditi da lavoro, redditi da pensione e liquidazioni di fine rapporto; 2) il modello 730, per le medesime componenti di reddito da lavoro/pensione presenti nel Cud, per i redditi da fabbricati o da terreni e per i redditi da lavoro autonomo non derivanti da attività professionale; 3) il modello Unico persone fisiche, per le stesse componenti di reddito rilevate nel 730, con la differenza che i redditi da lavoro autonomo sono riferiti ad attività professionale (titolari di reddito d'impresa e redditi derivante dall'esercizio di arti e professioni), e per redditi derivanti da rapporti di collaborazione coordinata e continuativa, svolte anche per attività manuali ed operative, senza vincolo di subordinazione e d'impiego di mezzi organizzati, e con retribuzione periodica prestabilita (cariche di amministratore, sindaco o revisore di società, eccetera).

Rispetto alle variabili obiettivo del progetto Eu-Silc, i dati delle dichiarazioni dei redditi sono generalmente insufficienti e devono essere integrati con altre fonti. Le fonti fiscali esaminate non forniscono, infatti, alcuna informazione riguardo alle seguenti componenti di reddito:

- redditi soggetti a imposta sostitutiva (redditi da capitale finanziario);
- redditi esenti da Irpef (pensioni sociali, pensioni di guerra pensioni di invalidità, indennità di accompagnamento, rendite per infortunio permanente);
- altre prestazioni sociali in denaro di natura non pensionistica, con l'esclusione delle liquidazioni di fine rapporto ed equipollenti e degli assegni al nucleo familiare dei dipendenti del privato;
- assegni al nucleo familiare dei dipendenti pubblici e dei pensionati;
- le deduzioni illegali e legali dall'imponibile (evasione ed elusione fiscale).

Per quanto concerne i redditi soggetti ad imposta sostitutiva non si prevede, nel breve termine, di acquisire dati di fonte fiscale, ma verrà avviato uno studio di fattibilità per verificare

⁹ L'ipotesi sottostante è che l'informazione da fonte fiscale (e più in generale da fonte amministrativa) sia più accurata rispetto a quella rilevata nel questionario dell'indagine campionaria.

il contenuto informativo della fonte fiscale modello 770-Ordinario (quadro sf – redditi da capitale, compensi per l'avviamento commerciale e contributi pubblici e privati) e, al tempo stesso, per stimare i tempi di acquisizione del dato e i carichi di lavoro conseguenti al trattamento dei dati.

Riguardo alle prestazioni pensionistiche è stata utilizzata, a decorrere dall'indagine Eu-Silc 2004, la fonte Inps "Casellario centrale dei pensionati" congiuntamente alle fonti fiscali. Essa raccoglie informazioni non altrimenti rilevabili, quali la tipologia di pensione percepita e le pensioni esenti dalla tassazione personale sui redditi (Irpef). La fonte amministrativa in questione viene consolidata ai fini statistici¹⁰ per la rilevazione dei titolari di pensione e degli importi erogati alla data del 31 dicembre di ciascun anno.

In merito alle altre prestazioni sociali in denaro non pensionistiche,¹¹ verrà effettuata, nel breve-medio periodo, una ricognizione del contenuto informativo dell'archivio fiscale Inps – Gestione fiscale delle prestazioni non pensionistiche – già impiegato per la rilevazione Istat sui trattamenti monetari non pensionistici (indennità di disoccupazione, indennità di mobilità, parte delle integrazioni salariali dei cassintegrati).

Gli assegni al nucleo familiare dei dipendenti pubblici sono rilevati in aggregato nel Conto annuale della Ragioneria Generale dello Stato. Quest'ultima sta attualmente sviluppando un sistema informativo che raccoglierà, tra le altre voci, le informazioni individuali collegate ai cedolini degli stipendi (tra cui figurano gli assegni familiari) di una quota consistente dei dipendenti del pubblico impiego. Pertanto, non si esclude la possibilità di collegare in futuro questa fonte amministrativa ai dati dell'indagine campionaria.

4.5.2 Costruzione delle variabili reddito Eu-Silc a partire dalle fonti fiscali

Tra le variabili obiettivo di Eu-Silc sono particolarmente rilevanti: il reddito netto monetario da lavoro dipendente (PY010N), il reddito netto da lavoro autonomo (PY050N), le "prestazioni sociali nette in denaro" associate alle funzioni vecchiaia (PY100N), superstiti (PY110N) e invalidità (PY130N). Con riferimento a dati dell'indagine sulle condizioni di vita, anno 2003, queste tre componenti di reddito comprendono il 94,7 per cento del reddito totale individuale. La prima variabile target¹² è definita come l'insieme delle "componenti monetarie, al netto delle trattenute fiscali e previdenziali, erogate dai datori di lavoro a remunerazione delle prestazioni lavorative dei propri dipendenti.

La seconda variabile target è definita come l'ammontare dei guadagni conseguiti, durante il periodo di riferimento, dal soggetto (per sé o per conto dei propri familiari) in relazione all'attività di lavoratore autonomo. La remunerazione del lavoro autonomo è generalmente legata ai profitti derivanti dell'attività produzione di beni e servizi (o agli autoconsumi quando sono ritenuti parte dei profitti).

Le ultime tre sono definite come l'insieme delle prestazioni in denaro di protezione sociale rispettivamente contro i rischi legati alla funzione vecchiaia (PY100N), alla funzione superstiti (PY110N) e alla funzione invalidità (PY130N). Queste prestazioni in denaro possono essere periodiche e continuative (pensioni in senso stretto) oppure una tantum (liquidazioni).

Gli elementi che concorrono alla determinazione del reddito monetario da lavoro dipendente, definito da Eurostat, sono rappresentati nel contesto italiano da: paga base o minimo tabellare, provvigioni, aumenti periodici per anzianità (scatti di anzianità), superminimi (per merito del lavoratore o per onerosità del mansione svolta), premi aziendali (produzione,

¹⁰ Il processo di consolidamento è realizzato attraverso la collaborazione tra l'Istat e l'Ente previdenziale.

¹¹ Esse includono le indennità di disoccupazione e di mobilità, le integrazioni salariali dei cassintegrati, le indennità di malattia e di inabilità temporanea assoluta, le indennità di maternità.

¹² Secondo la definizione adottata dall'Eurostat (cfr. Eurostat 2004).

rendimento e fedeltà), compensi per lavoro straordinario, indennità compensative (connesse al rischio della mansione o alla funzione, alla reperibilità, alla sede disagiata, eccetera), gratifiche natalizie o mensilità aggiuntive (tredicesima, quattordicesima), patto di non concorrenza, mance, contributi spese per la frequenza asili nido dei figli del lavoratore. Tutti questi elementi, con la sola eccezione dei contributi per le rette dell'asilo,¹³ costituiscono base imponibile ai fini fiscali e sono pertanto rilevati nel modello Cud. In particolare, questi redditi figurano al punto 1 (redditi per i quali è possibile fruire della deduzione di cui art.10-bis e art.13 del Tuir) e al punto 2 (redditi per i quali è possibile fruire della deduzione di cui art.10-bis, comma 1) della Parte B - Sezione "Dati fiscali", segnatamente al lordo delle trattenute fiscali e al netto delle ritenute contributive. Per ottenere i redditi netti da lavoro dipendente (PY010N) occorre che dai redditi imponibili (punti 1 e 2) siano defalcati le ritenute Irpef (punto 12 del Cud 2004), l'addizionale regionale (punto 13 del Cud 2004) e l'addizionale comunale (punto 14 del Cud 2004).

Per quanto riguarda le prestazioni sociali in denaro aventi carattere periodico e continuativo (pensioni), vi è una parte esente da imposta (pensioni sociali, pensioni di guerra pensioni di invalidità, indennità di accompagnamento, rendite per infortunio permanente) e quindi non coperta dall'archivio dei modelli Cud. Mentre le prestazioni sociali erogate a tantum (liquidazioni) sono rilevabili in un'altra sezione: Trattamenti di fine rapporto e indennità equipollenti. I dati sui trattamenti pensionistici al lordo (tassabili) sono riportati negli stessi punti 1 e 2, Parte B - Sezione "Dati fiscali", occupati dai redditi di lavoro dipendente e assimilati. Le ritenute fiscali applicate ai titolari di pensione sono indicate, come per i lavoratori dipendenti, ai punti 12-13-14 della medesima sezione. Sebbene il Cud non preveda campi distinti per i redditi da lavoro e quelli da pensione, è ugualmente possibile tracciare alcune linee guida per la loro separazione. Così ad esempio, al punto 9 della Parte A - Sezione "Dati relativi al dipendente, pensionato o altro percettore" è specificata la modalità che contraddistingue il pensionato (lettera A o B) dal resto. Inoltre, ai punti 6 e 7 della Parte B - Sezione "Dati fiscali" è possibile rilevare con precisione il numero dei giorni di pensione, in corrispondenza dei quali spettano le deduzioni di cui all'art.10-bis, commi 2 e 3, del Tuir. L'identificazione del sostituto d'imposta erogatore di pensioni (tramite la ragione sociale o il codice fiscale) fornisce un altro criterio per individuare il titolare di reddito da pensione.

Il computo del "reddito netto da lavoro o da pensione" dal Cud, come differenza tra l'imponibile lordo e le ritenute fiscali applicate nell'anno (imposta provvisoria) dal sostituto d'imposta, non tiene conto di eventuali aggiustamenti in positivo in sede di dichiarazione dei redditi (modello 730) e pertanto potrebbe fornire un valore in difetto. Se il contribuente non possiede altri redditi nell'anno e fa valere il diritto ad eventuali deduzioni o detrazioni di imposta (per spese sanitarie, per interessi passivi legati a mutui ipotecari per acquisto dell'abitazione principale, premi per l'assicurazione vita, eccetera), in sede di conguaglio fiscale riceverà un rimborso fiscale. L'imposta definitiva in questo caso sarà inferiore a quella provvisoria (ritenuta) applicata dal datore di lavoro.

In realtà, l'aggiunta di eventuali altri redditi non tassati anticipatamente (ad esempio, i redditi da fabbricati o i redditi da terreni) o tassati su un'aliquota convenzionale (compensi di attività occasionale di lavoro autonomo) determina un aumento della base imponibile e produce un saldo d'imposta a debito (differenza tra l'imposta dovuta e le trattenute o le ritenute d'acconto), rendendo più difficile il calcolo del reddito netto di ciascuna componente di reddito. In questi casi per ottenere il reddito netto di ogni tipologia, si procede a calcolare il contributo di ciascuna componente rispetto al reddito complessivo e ad applicare pro-quota l'imposta

¹³ In base all'art.48 c.2, lett.f, del Tuir i contributi del datore lavoro a favore della generalità dei dipendenti o a relative categorie per frequenza di asili nido e di colonie climatiche, per borse di studio dei loro familiari non costituiscono base imponibile ai fini fiscali e contributivi.

definitiva (al netto delle detrazioni e deduzioni comuni a tutti i redditi). Quest'ultima operazione dovrà tener conto delle detrazioni specifiche per tipologia di reddito (per esempio la detrazione per lavoro dipendente) che si aggiungono alle detrazioni comuni, cioè indipendenti dalla fonte di reddito (per esempio la detrazione per spese sanitarie). Se il contribuente è registrato nell'archivio dei modelli Cud ma non in quello dei modelli 730 (o modello Unico persone fisiche), si può ragionevolmente ipotizzare che l'imposta provvisoria, applicata dal datore di lavoro nel Cud, coincida con quella definitiva e che il reddito netto desumibile da quest'ultima fonte sia corretto. Si ricorda, inoltre, che nel Cud è presente una sezione sui redditi derivanti dalle collaborazioni coordinate e continuative. Sebbene assimilati fiscalmente ai redditi da lavoro dipendente questi compensi sono classificati in Eu-Silc come redditi da lavoro autonomo.

Le componenti di reddito da lavoro e da pensione, sinora esaminate, sono rilevabili anche presso il modello 730, nel caso in cui percettore intenda assolvere all'obbligo di presentazione e di pagamento delle imposte, per i redditi che si aggiungono a quelli dichiarati nel Cud, ovvero ottenere in tempi rapidi rimborsi d'imposta spettanti (deduzioni e detrazioni). I redditi imponibili da lavoro dipendente o da pensione, di cui al punto 1 della Parte B - Sezione "Dati fiscali del Cud, sono ricompresi nel Quadro C - Sezione I "Redditi da lavoro dipendente e assimilati" del modello 730 e distinti in tre categorie: 1) redditi da pensione; 2) redditi da lavoro e assimilati; 3) compensi ai lavoratori socialmente utili che hanno raggiunto l'età pensionabile. I redditi imponibili da lavoro dipendente parzialmente deducibili, di cui al punto 2 della Parte B - Sezione "Dati fiscali del Cud, sono indicati nel Quadro C - Sezione I "Redditi da lavoro dipendente e assimilati per i quali non spettano le deduzioni". Il complesso delle ritenute applicate dal sostituto d'imposta per questi redditi è riportato nella Sezione III "Ritenute sui redditi da lavoro dipendente e addizionali regionali e comunali Irpef". Dal modello 730 si può ottenere una stima dei redditi netti da lavoro dipendente, defalcando le ritenute operate dal sostituto (imposta provvisoria) ai redditi imponibili da lavoro o da pensione, come per il Cud. Se il contribuente è presente contestualmente negli archivi dei modelli Cud e dei modelli 730, allora è necessario verificare la coerenza delle informazioni ivi riportate e procedere al calcolo esatto del reddito netto. Le informazioni raccolte nel prospetto di liquidazione del modello 730 forniscono tutti gli elementi per il computo di questo aggregato. Il modello 730 consente, tra l'altro, di rilevare altre variabili reddito obbiettivo, quali i redditi da capitale reale (affitto di terreni e/o fabbricati) e talune componenti del reddito da lavoro autonomo (compensi derivanti dall'esercizio di attività occasionali da lavoro autonomo).

Il modello Unico persone fisiche viene regolarmente utilizzato da percettori di compensi da lavoro autonomo, riferiti ad attività professionale, in sede di dichiarazione dei redditi. Esso comprende, tuttavia, un apposito Quadro – RC – sui redditi da lavoro dipendente e assimilati, per includere le dichiarazioni dei lavoratori autonomi che percepiscono contestualmente redditi da lavoro dipendente e assimilati. Il modello Unico, in talune circostanze¹⁴ può essere utilizzato dai percettori di soli redditi da lavoro dipendente e assimilati. Nel quadro RC dell'Unico persone fisiche si riporta, come da modello 730, la distinzione per tipologia di reddito: 1) reddito da lavoro dipendente; 2) reddito da pensione; 3) compensi ai lavoratori socialmente utili che hanno raggiunto l'età pensionabile. Il modello Unico, al pari del modello 730, classifica le tipologie di redditi (imponibile) da lavoro dipendente e assimilati sulla base alla possibilità di applicare o meno le

¹⁴ Nella fattispecie trattasi di: 1) lavoratori dipendenti che hanno cambiato datore di lavoro e sono in possesso di più certificazioni di lavoro dipendente o assimilati (Cud), dove l'imposta corrispondente al reddito complessivo supera il totale delle detrazioni spettanti (ovvero se non è richiesto il conguaglio fiscale all'ultimo datore di lavoro); 2) lavoratori dipendenti che hanno percepito direttamente dall'Inps o da altri Enti indennità e somme a titolo di integrazione salariale o ad altro titolo, se erroneamente non sono state effettuate le ritenute e se non ricorrono le condizioni di esonero; 3) lavoratori dipendenti che hanno percepito retribuzioni e/o redditi da privati non obbligati per legge ad effettuare ritenute d'acconto (per esempio collaboratori familiari, autisti ed altri addetti della casa).

specifiche deduzioni. Le ritenute applicate dai sostituti d'imposta sui redditi da lavoro dipendente sono indicate nel complesso, cosicché non vi è possibilità di associarle alle fonti di reddito. Il modello Unico permette inoltre di rilevare una serie di componenti di reddito da capitale reale (redditi da fabbricati e da terreni), alla stregua del modello 730. Tale modello costituisce il punto di riferimento per la validazione dei dati campionari sulle componenti di redditi da lavoro autonomo. Il reddito che deriva da lavoro autonomo è definito come quello di un'impresa: ricavi meno costi. Esso è quindi costituito dalla differenza tra i compensi percepiti e le spese sostenute per l'esercizio del proprio lavoro durante il periodo d'imposta. Gran parte di questi redditi sono anch'essi soggetti al sistema della ritenuta alla fonte a titolo di acconto. Seguendo la struttura del modello Unico i redditi da lavoro autonomo si compongono di: redditi derivanti dall'esercizio delle arti e professioni (Quadro RE); redditi dei titolari d'impresa in contabilità ordinaria (Quadro RF) e semplificata (Quadro RG); redditi di partecipazione in società di persone e assimilate (Quadro RH); redditi di allevamenti (Quadro RD), compensi derivanti da attività di lavoro autonomo, anche se svolte all'estero, non esercitate abitualmente (Quadro RL). La somma di questi redditi al netto delle ritenute d'acconto versate fornisce la stima del reddito netto da lavoro autonomo.

4.5.3 Strategie per l'integrazione a livello micro tra Eu-Silc e le fonti fiscali

Il processo di integrazione tra la fonte campionaria Eu-Silc e le fonti fiscali sopra menzionate si articola in dieci fasi che si susseguono con le seguente sequenza temporale:

1. selezione delle “variabili individuali generatrici del codice fiscale individuale” (nome, cognome, sesso, anno e luogo di nascita) e dei “codici fiscali individuali rilevati presso i Comuni” attraverso il registro familiare e individuale Sigif¹⁵ (Sistema integrato di gestione delle indagini sulle famiglie);
2. controllo dei codici fiscali individuali in Sigif (ed eventuale generazione dei codici fiscali inesatti o mancanti) e preparazione del file standard per lo scambio e la trasmissione dei dati con la Sogei;
3. preparazione del modulo di richiesta dati alla direzione DCAR dell'Istat, contenente la specifica degli archivi (modello fiscale e anno di riferimento) e dei quadri o delle sezioni da rilevare;
4. confronto fra il file standard rilasciato dalla Sogei (contenente le informazioni anagrafiche e fiscali delle persone fisiche della lista di campionamento) e il file dell'indagine Eu-Silc contenente le informazioni sulle unità di analisi intervistate, attraverso la chiave identificativa del soggetto;
5. predisposizione e aggiornamento del programma di lettura e conversione dei database fiscali dal formato “codice a serrare” rilasciato dalla Sogei ad un formato direttamente leggibile in SAS (variabili di lunghezza e posizione prefissata);
6. predisposizione del programma di identificazione e correzione dei record duplicati e di consolidamento delle informazioni relative allo stesso individuo riportate su record multipli;
7. predisposizione del programma di controllo e correzione degli outlier sui redditi dei dati fiscali;
8. utilizzo delle procedure di *merge* delle diverse fonti fiscali attraverso la chiave identificativa del record data dal “codice fiscale dell'individuo” e costituzione dei 5 archivi di riferimento¹⁶;

¹⁵ Questo registro fornisce la lista delle unità campionarie dell'indagine Eu-Silc. Esso comprende sia informazioni a livello individuale (nome, cognome, età, sesso, stato civile, codice fiscale) che informazioni a livello familiare (luogo e indirizzo di residenza).

¹⁶ I cinque archivi di riferimenti sono dati rispettivamente da: 1) archivio dei percettori che hanno solo dichiarazioni Cud; 2) archivio dei percettori di reddito con dichiarazioni Cud e modello 730; 3) archivio delle sole dichiarazioni 730; 4) archivio delle dichiarazioni Cud e modello Unico persone fisiche; 5) archivio delle sole dichiarazioni Unico persone fisiche;

9. analisi della coerenza dei valori reddituali dello stesso individuo riportati su più archivi fiscali (Cud e 730, Cud e Unico) e formulazione di ipotesi di lavoro in presenza di dati incoerenti;
10. analisi della coerenza fra i dati fiscali e dati dell'indagine Eu-Silc, formulazione di ipotesi di lavoro in presenza di dati incoerenti ed imputazione dei dati da fonte fiscale.

Durante la prima fase si analizza la sezione dell'archivio Sigif che include le variabili anagrafiche del percettore di reddito, utili alla generazione automatica del codice fiscale. Ai fini dell'implementazione della suddetta procedura occorre disporre di una lista aggiornata dei Codici dei Comuni d'Italia e degli Stati esteri da associare al luogo di nascita del percettore. Oltre queste informazioni Sigif fornisce i codici fiscali degli individui da intervistare, rilasciati dai Comuni presso cui essi risiedono. Tale fase è stata già realizzata per l'indagine Eu-Silc 2004.

Nella seconda fase si appronta il file standard per lo scambio delle informazioni con la Sogei. Tale fase prevede l'identificazione e l'attribuzione del codice fiscale della persona appartenente alla famiglia da intervistare. Relativamente all'indagine Eu-Silc 2004, si osserva come il registro Sigif riporti un totale di 78.837 individui appartenenti alle 31.998 famiglie della lista di campionamento dell'indagine. Di questi individui ben 20.242 presentano codici fiscali mancanti. La procedura di generazione automatica dei codici fiscali, implementata in Istat, consente di effettuare 19.158 imputazioni sui dati mancanti, fallendo in appena 1.084 casi a causa della indisponibilità delle informazioni di base. Viceversa, i restanti 58.595 individui riportano codici fiscali valorizzati in Sigif. Il file standard inviato alla Sogei contiene i dati individuali delle persone fisiche presenti nella lista campionaria di Eu-Silc 2004, compresi tanto del nominativo, età, genere e codice fiscale dell'individuo, quanto dell'indirizzo (via, numero civico, c.a.p.), del comune di residenza della famiglia di appartenenza. Si noti che questa procedura consente l'acquisizione di dati fiscali relativi anche a famiglie che rifiuteranno l'intervista, rendendo potenzialmente praticabili strategie di riponderazione basate sull'analisi della probabilità di non-risposta.

Nella terza fase si predispongono i moduli di richiesta dati da inviare alla direzione DCAR dell'Istat. In tale fase si definiscono gli archivi, le sezioni e i quadri da rilevare. Per essi la direzione DCAR inoltrerà formale richiesta all'Agenzia delle entrate, nel quadro di un'apposita convenzione per lo scambio dei dati tra i due Enti. Anche questa fase è stata già conclusa per l'indagine sopra menzionata;

La quarta fase prevede un'analisi del file standard "persone fisiche" rilasciato dalla Sogei, con particolare riguardo alla validazione dei codici fiscali attribuiti ai soggetti appartenenti alla lista campionaria Eu-Silc. Segnatamente per l'indagine 2004, la Sogei ha convalidato 72.616 codici fiscali (osservati o generati) di persone fisiche presenti nella lista campionaria di Sigif, ha sostituito 1.520 codici errati o mancanti ed, infine, ha individuato 4.701 individui come "non presenti all'anagrafe tributaria" e non altrimenti recuperabili (a causa di errori riportati nella sezione dati anagrafici di Sigif). Ne consegue che l'archivio rilasciato dalla Sogei garantisce una copertura del 94 per cento delle persone fisiche presenti nella lista campionaria Eu-Silc 2004.

La quinta fase, relativa alla migrazione dei dati fiscali da un formato compresso a formato leggibile in Sas, è stata testata, in via sperimentale, su un campione di modelli Cud, 730 e Unico persone fisiche dell'anno 2001 (anno di riferimento 2000) ed ha richiesto l'implementazione di un programma in Visual Basic per la conversione della struttura del file (codici a serrare). Successivamente il programma è stato applicato alle basi di dati fiscali rilasciate dall'Agenzia delle entrate: Unico persone fisiche e modello 730, periodo d'imposta 2003.

La sesta fase, riguardante l'eliminazione dei record duplicati e il consolidamento delle informazioni dello stesso individuo poste su più record, è stata anch'essa testata in via sperimentale sui dati dei modelli Cud anno 2001, utilizzati nella precedente fase. Essa ha richiesto l'impiego di programmi Sas, già utilizzati allo stesso scopo sui dati fiscali dei modelli 770 (che includono le

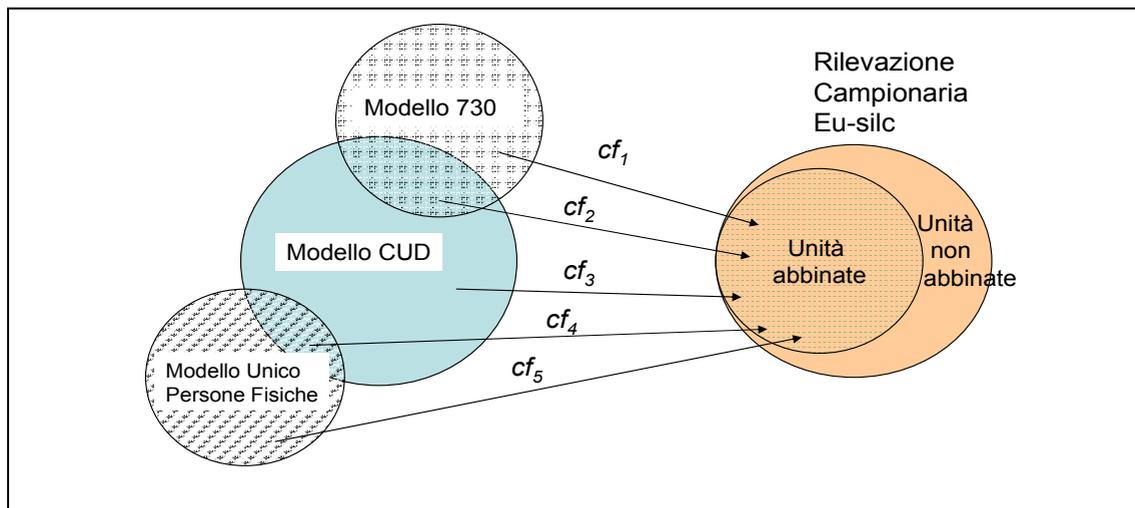
informazioni del Cud) dell'anno 2000.¹⁷ Al momento della stesura del presente documento, tale fase non era stata ancora implementata sui dati Cud anno 2004 per indisponibilità del relativo archivio. La fase in questione è stata tuttavia applicata successivamente.

La settima fase è strettamente legata alla precedente e prevede l'impiego di procedure per l'identificazione dei valori anomali e la relativa eliminazione/imputazione.

L'ottava fase, concernente il *merge* tra i dati fiscali, può essere schematizzata come in figura 4.1. In questa rappresentazione grafica si ipotizza, per semplicità, l'inesistenza di problemi di copertura tra i vari archivi fiscali. Pertanto l'archivio dei modelli Cud è esaustivo delle dichiarazioni riguardanti i percettori dei redditi da lavoro, dei redditi da pensione (tassabili) e delle liquidazione di fine rapporto. Similmente, l'archivio dei modelli 730 comprende il complesso delle dichiarazioni dei titolari di redditi, aggiuntivi o sostitutivi a quelli da lavoro dipendente e assimilati (da fabbricati, terreni, compensi di lavoro autonomo occasionali, eccetera), o di coloro che richiedono rimborsi d'imposta per detrazioni/deduzioni. Come si evince dalla figura 4.1, uno stesso individuo può essere contestualmente presente nell'archivio dei modelli Cud e dei modelli 730 (insiemi sovrapposti). Tale situazione si verifica quando i titolari di redditi da lavoro dipendente e assimilati dichiarano di possedere altri redditi oppure semplicemente richiedono il rimborso d'imposta. Le dichiarazioni 730 non si sovrappongono a quelle del Cud quando i titolari possiedono esclusivamente redditi diversi da quelli da lavoro dipendente e assimilati. Lo stesso ragionamento si applica, pur con le opportune differenze in termini di componenti di reddito coinvolte, nei confronti delle dichiarazioni dei modelli Cud e Unico persone fisiche. Viceversa le dichiarazioni 730 e Unico persone fisiche sono alternative e per definizione non si sovrappongono.

La presenza di talune tipologie di reddito esenti da imposizione fiscale o soggette ad imposta sostitutiva determina l'eventualità che alcune unità – individui – dell'indagine Eu-Silc non siano abbinata alle unità delle fonti fiscali (*unit not linked*).

Figura 4.1 - Struttura esemplificata del record-linkage tra i dati fiscali e i dati campionari Eu-Silc



Fonti: Istat, Indagine sulle condizioni di vita - Anno 2004 e Agenzia delle entrate, archivi fiscali (periodo d'imposta 2003)

¹⁷ Per un approfondimento si veda Consolini P. (2004), L'indagine sperimentale sull'archivio fiscale modello 770 anno 1999: analisi della qualità del dato e stime campionarie.

Nella nona fase si esamina il problema della coerenza fra i microdati rilevati su più fonti fiscali. Per effettuare tale analisi occorre verificare preliminarmente se le componenti di reddito sono definite allo stesso modo tra i vari archivi. Nel caso in cui le stesse componenti di reddito, riferite al medesimo soggetto, assumano valori diversi fra il modello Cud e il modello 730, si preferisce il dato riportato su quest'ultimo modello, in quanto esso risulta più affidabile e più aggiornato¹⁸ (rileva il complesso dei redditi da lavoro dipendente riferiti anche a più rapporti di lavoro, mentre i modelli Cud sono affetti da problemi di copertura). Stesso discorso si applica al caso della sovrapposizione dei microdati relativi all'archivio Cud e all'archivio Unico persone fisiche, dove si privilegia quest'ultimo per i motivi sopra menzionati.

Nella decima ed ultima fase si confrontano i valori delle componenti di reddito rilevate nell'indagine Eu-Silc con quelli desumibili dagli archivi fiscali consolidati. Anche in questo caso si ipotizza che le informazioni desumibili dagli archivi fiscali siano più affidabili rispetto a quelle rilevate nel questionario dell'indagine Eu-Silc. Tuttavia, se tale ipotesi è plausibile nei confronti delle componenti di reddito da lavoro dipendente, redditi da pensione tassabili e liquidazioni di fine rapporto, altrettanto non si può dire nei confronti della componente di reddito da lavoro autonomo. In questi casi, per il fenomeno dell'evasione/elusione fiscale sui redditi da lavoro autonomo, il dato campionario potrebbe rilevarsi più rispondente al vero di quanto non lo sia il dato fiscale. In realtà l'evasione/elusione fiscale non è esplicitamente menzionata nella definizione di reddito lordo/netto Eu-Silc, per la difficoltà di stabilire metodologie comparabili di stima e per l'arbitrarietà delle ipotesi di comportamento sottostanti ai modelli di imputazione del reddito evaso. Un problema collegato è quello della reticenza a dichiarare interamente i propri redditi nel corso di una intervista campionaria (*under-reporting*). In generale, non è detto che l'*under-reporting* sia eguale all'evasione/elusione fiscale e non vi sono quindi motivi per pensare che i redditi netti campionari siano approssimativamente uguali ai redditi netti che risultano dalla dichiarazione fiscale. Per ragioni simili a quelle relative all'evasione fiscale, nel Regolamento che definisce il reddito netto Eu-Silc l'*under-reporting* non è esplicitamente menzionato. Un'ipotesi di lavoro è quella di effettuare correzioni/imputazioni sui valori dell'indagine Eu-Silc, cioè di donare i valori dei redditi da lavoro autonomo da fonte fiscale, solo nei casi in cui i valori dei primi siano inferiori a quelli dei secondi ovvero risultino mancanti.

4.5.4 *Un'applicazione del processo d'integrazione fra dati campionari Eu-Silc e fonte amministrativa: i redditi da lavoro autonomo e i redditi da pensione*

L'indagine Eu-Silc 2004 costituisce il primo esempio in Italia di integrazione, a livello micro, fra dati campionari sui redditi delle famiglie e degli individui e le principali fonti amministrative. Tale lavoro pionieristico è stato realizzato, al momento, limitatamente alla componente di reddito netto da lavoro autonomo (PY050N) e, parzialmente, al reddito netto da pensione (PY100N, PY110N, PY130N).

Con riferimento al reddito netto da lavoro autonomo, è stata implementata una procedura di integrazione dei dati di fonte amministrativa, comprensiva di sei fasi così strutturate:

1. identificazione dei questionari che, al giudizio del rilevatore, riportano informazioni poco attendibili sui redditi (variabile Eu-Silc: "Parere"; modalità associate: poco "3", per niente "4");
2. cancellazione delle informazioni sui redditi poco attendibili degli autonomi;

¹⁸ In questa sede, per semplicità, si traslascia l'ipotesi secondo cui la differenza in valore tra dati dei due archivi sia attribuibile ad errori nella chiave identificativa del record (*matching key*) presenti in uno dei due archivi e, quindi, il confronto sia effettuato su unità statistiche differenti. L'eventualità che la variabile (o le variabili) di abbinamento sia affetta da errori suggerisce l'impiego di tecniche di *exact matching* di tipo probabilistico che prendano in esame un complesso di variabili di abbinamento.

3. imputazione del complesso dei valori mancanti (comprendenti sia dati mancanti nel questionario quanto quelli cancellati al passo 2.) sui redditi degli autonomi tramite il pacchetto statistico Iweware;
4. costruzione della variabile reddito netto da lavoro autonomo da fonte fiscale, al netto delle ritenute fiscali, e distinzione delle componenti relative a: i) compensi per l'attività di libero professionista o di collaboratore coordinato e continuativo; ii) redditi di impresa percepiti in qualità di titolare dell'attività o di coadiuvante. Tale suddivisione si rende necessaria in quanto la prima componente di reddito rappresenta l'effettiva remunerazione dell'attività lavorativa svolta dal professionista, mentre la seconda componente comprende al suo interno anche i compensi "fittizi" attribuiti ai coadiutori o soci di impresa "di comodo". Infatti, simili coadiutori o soci, sebbene figurino negli archivi fiscali come percettori di redditi di partecipazione in società di persone o assimilati, in realtà non svolgono alcun ruolo attivo nell'impresa. Essi sono inclusi nelle dichiarazioni dei redditi autonomi solo per ragioni di convenienza fiscale da parte del vero titolare o dei veri coadiutori;
5. abbinamento (deterministico) delle unità campionarie Eu-Silc con le unità appartenenti all'insieme dei due archivi fiscali, Unico persone fisiche e modello 730, attraverso la chiave del "codice fiscale". In particolare, dei 7.643 lavoratori autonomi rilevati in Eu-Silc 2004, 5.120 trovano un abbinamento con le persone fisiche presenti negli archivi fiscali, mentre i restanti 2.523 non si agganciano (*units not linked*). Dall'insieme delle persone abbinate, 3.782 dichiarano redditi da lavoro autonomo positivi al fisco. Infine, dalle fonti fiscali si individuano ben 1.178 percettori di reddito da lavoro autonomo che in Eu-Silc non rispondono alle sezione dei redditi autonomi;
6. costruzione delle "ipotesi di lavoro" sottostanti l'integrazione dei dati campionari con le informazione di fonte fiscale e applicazione delle relative regole di calcolo. L'ipotesi di base è che il valore vero sul reddito da lavoro autonomo sia potenzialmente sottostimato da ambedue le fonti di dati, sia per motivi di elusione/evasione fiscale che di reticenza a dichiarare il vero da parte dell'intervistato (*under-reporting*). Per la componente dei redditi derivanti da "attività di libera professione e collaborazioni coordinate" si è stabilito di considerare come "prossimo al vero" il valore massimo tra il dato dichiarato al fisco e il dato registrato in Eu-Silc (sia esso quello osservato o imputato tramite Iweware). Si precisa che le due componenti di reddito autonomo non sono rilevate separatamente in Eu-Silc, per cui in caso di compresenza delle due tipologie di reddito si procede a stimare una componente come percentuale del reddito autonomo dichiarato nell'indagine, in rapporto al valore registrato al fisco per l'altra tipologia. Per i "redditi di impresa" si utilizza un diverso criterio d'integrazione dei dati amministrativi rispetto a quanto sopra descritto. Si identificano, dapprima, i "lavoratori autonomi fittizi"; ovvero quei soggetti che in Eu-Silc dichiarano di "non aver mai lavorato in passato e di essere attualmente in condizione non professionale" e che, parallelamente, al fisco risultano come percettori di redditi di partecipazioni in società. Quando questi soggetti appartengono a famiglie in cui non vi sono lavoratori autonomi, si ipotizza realisticamente che essi percepiscano solo redditi da capitale, ovvero a remunerazione del capitale che loro mettono nell'attività senza prestare lavoro (nell'indagine Eu-Silc 2004 ne sono stati individuati 42 casi). Mentre, se essi vivono presso famiglie in cui vi sia almeno un lavoratore autonomo (che dichiara in Eu-Silc di essere autonomo o di aver lavorato nella settimana di riferimento o in precedenza), allora si attribuisce proporzionalmente il reddito dei primi (lavoratori autonomi fittizi) agli effettivi lavoratori autonomi. Nel caso dei percettori di reddito da impresa si effettua un confronto a livello familiare tra i redditi di impresa riportati nella fonte fiscale e quelli di fonte campionaria (al netto dell'eventuale componente di reddito da libera professione). Analogamente al caso dei liberi professionisti, si utilizza come "prossimo al vero", in questo caso però su base

familiare, il valore massimo dell'aggregato dei redditi di impresa nella famiglia tra le due fonti considerate. Tuttavia, anche quando su base familiare emerge una dominanza dei dati reddituali di fonte fiscale rispetto ai corrispettivi valori campionari, può accadere che a livello individuale tale ordinamento sui redditi non sia rispettato. In altri termini può succedere che taluni redditi autonomi individuali dichiarati in Eu-Silc risultino superiori rispetto ai dati della controparte fiscale, nonostante a livello familiare sia vero il contrario. Tale incongruenza rappresenta un spia dell'eventuale presenza di lavoratori autonomi fittizi. In queste circostanze si assume che la vera struttura dei redditi da impresa riferita a ciascun familiare sia quella osservata in Eu-Silc, mentre l'ammontare dei redditi di impresa su base familiare sia espresso dal valore più elevato, cioè quello fiscale. Viceversa, quando il valore aggregato dei redditi da impresa di fonte campionaria domina il corrispettivo dato di fonte fiscale, allora si ritiene che sia sempre corretto il dato individuale Eu-Silc. In analogia a quanto esposto in precedenza, è possibile che qualche valore individuale del reddito di impresa di fonte fiscale sia più elevato del corrispondente dato campionario, a dispetto dell'ordinamento osservato sull'aggregato famiglia. Questa circostanza fornisce anch'essa un segnale della presenza di percettori di redditi di impresa fittizi, e suggerisce pertanto di utilizzare come vera la struttura dei redditi individuali di fonte campionaria. Nel campione Eu-Silc 2004 sono stati individuati 52 lavoratori autonomi fittizi appartenenti a famiglie con almeno un lavoratore autonomo.

Se si confrontano gli importi del reddito da lavoro autonomo rilevati contestualmente nell'indagine campionaria e nelle fonti fiscali (5.120 osservazioni), si osserva come in 2.956 casi queste ultime riportino valori più elevati rispetto alla prima.

I principali risultati dell'integrazione fra dati campionari e fonti fiscali, relativamente alla sezione redditi autonomi, sono sintetizzati nella Tavola 4.8. Come si può notare dai dati riportati in seconda e quarta colonna, l'utilizzo integrato di informazioni di fonte fiscale produce, a livello nazionale, un incremento di circa 1,1 milioni di unità sulla stima del numero di percettori di reddito autonomo ed un aumento del 15 per cento del reddito netto medio di questa tipologia, rispetto ai dati della sola indagine Eu-Silc.

Tavola 4.8 - Confronto fra il numero di percettori e reddito medio netto da lavoro autonomo, prima e dopo l'integrazione delle fonti fiscali – Anno 2003

BASE DATI DI RIFERIMENTO	Percettori di reddito lavoro autonomo		Reddito netto medio da lavoro autonomo	
	Campione (a) (unità)	Stima del dato nazionale (b) (migliaia di unità)	Campione (a) (in euro)	Stima del dato nazionale (b) (in euro)
Dati Eu-Silc 2004 dopo l'imputazione Ivedere	7.643	6.890	13.806	13.991
Dati Unico_P.F./04 e Modello 730/04	6.714	-	15.431	-
Dati Eu-Silc 2004 integrati con (UPF/04+M730/04)	8.821	8.010	15.646	16.091

(a) dati non pesati.

(b) dati pesati.

Fonti: Indagine sulle condizioni di vita anno 2004, Modello 730/04 periodo d'imposta 2003, Modello UPF/04 (Unico persone fisiche periodo d'imposta 2003)

Per quanto concerne la tipologia di reddito da pensione, si stava sperimentando nella prima edizione Eu-Silc l'integrazione fra dati campionari e fonti amministrative attraverso lo sviluppo di diverse fasi operative. A differenza del caso dei redditi autonomi, in quest'ambito si era

ancora in una fase sperimentale, per cui erano in corso di affinamento le tecniche per il calcolo dei redditi al netto delle ritenute alla fonte. Il disporre di una fonte di *benchmark* in merito al “numero dei percettori”, quale il Casellario Centrale dei pensionati dell’Inps, avrebbe permesso la validazione completa delle stime campionarie sui pensionati. L’integrazione dei dati prevedeva l’utilizzo di due distinte fonti esterne: il Casellario pensionistico e gli archivi fiscali “Unico persone fisiche e modello 730”. Poiché gli archivi fiscali menzionati rilevavano solo parzialmente i pensionati, occorre applicare un modello di microsimulazione, per operare il passaggio al netto dell’imposta dei redditi lordi da pensione del Casellario.

I passi operativi per ottenere l’integrazione assumevano la seguente articolazione:

1. imputazione dei valori mancanti sui redditi da pensione dell’indagine Eu-Silc 2004 attraverso il package statistico Ivieware;
2. realizzazione di un primo abbinamento (deterministico) tra le unità campionarie Eu-Silc e le unità statistiche del Casellario attraverso la chiave del “codice fiscale del percettore”. In questa fase sono state abbinate 15.292 persone fisiche del campione Eu-Silc, mentre le altre 1.996 unità non hanno trovato un corrispettivo nell’archivio del Casellario (*units not linked*). In particolare, tra le 15.292 persone agganciate, 14.032 figurano in Eu-Silc come percettrici di pensioni, mentre le restanti 1.260 possiedono redditi diversi;
3. implementazione di un secondo abbinamento (deterministico) fra le 15.292 unità contemporaneamente presenti in Eu-Silc e nel Casellario e le “persone fisiche con redditi da pensione” presenti negli archivi fiscali Unico persone fisiche e modello 730, tramite la chiave del codice fiscale;
4. calcolo della componente reddito netto mensile da pensione del Casellario, mediante un modello semplificato di microsimulazione, distintamente per: pensioni da lavoro (vecchiaia e anzianità), pensioni di reversibilità o superstiti, pensioni di invalidità. Questa fase come detto in precedenza era in fase di implementazione;
5. calcolo della componente reddito netto mensile da pensione degli archivi fiscali Unico persone fisiche e modello 730. Per semplicità sono state considerate le sole pensioni al netto delle ritenute alla fonte che riportavano un numero di giorni utili per il calcolo delle deduzioni non inferiori a 323 (pari a dodici mensilità). Tale decisione è stata adottata per eliminare potenziali fonti di errore nel calcolo degli importi netti mensili delle pensioni, riscontrate prevalentemente in corrispondenza delle pensioni con periodicità inferiore agli 11 mesi;
6. assegnazione della struttura osservata nel Casellario delle pensioni erogate (distinte per tipologia) ai dati campionari Eu-Silc, con relativa attribuzione degli importi percepiti secondo criteri e regole in fase di studio. In ogni caso si prevede di confrontare i valori netti dei redditi da pensioni rilevati negli archivi fiscali (riportati su base mensile) con i redditi da pensione del Casellario pensionistico dopo l’applicazione di un modello di microsimulazione.

I risultati dell’integrazione dei dati campionari Eu-Silc con le fonti amministrative, relativamente ai percettori di reddito da pensione, sono illustrati nella Tavola 4.9. Come si può osservare il processo di integrazione consente di correggere le forti distorsioni nella distribuzione dei pensionati, per tipologia di prestazione, presenti in Eu-Silc. Il dato finale, contenente le integrazioni, sui percettori presenta dunque una palese similarità con la struttura osservata nella rilevazione totale del Casellario pensionistico (*benchmark*).

Tavola 4.9 - Confronto fra il numero di percettori residenti in Italia da fonte: a) Eu-Silc 2004 (stima campionaria), b) Eu-Silc 2004 integrata con dati amministrativi (stima campionaria), c) Casellario pensionistico (universo) – Anno 2003

BASE DATI DI RIFERIMENTO	Percettori di pensione residenti in Italia (dati assoluti in migliaia)				
	Vecchiaia e anzianità	Superstiti	Invalidità o Inabilità (a)	Sociale	Totale
Dati Eu-Silc 2004	10.672	3.375	2.210	1.021	14.668
Dati Eu-Silc 2004 integrati con dati fiscali e Casellario pensionistico	10.573	4.194	4.148	749	15.861
Casellario pensionistico (b)	10.471	4.602	4.523	743	15.726

(a) Sono esclusi i percettori di sola indennità di accompagnamento.

(b) Sono esclusi i percettori di pensione di età inferiore ai 14 anni per garantire la comparabilità con Eu-Silc.

Fonti: Istat, Indagine sulle condizioni di vita anno 2004, Elaborazioni su Casellario centrale dei pensionati anno 2003, Modello 730/04 e Modello UPF/04

4.5.5 Pianificazione del processo di integrazione dei dati di fonte amministrativa

Per pianificare l'attività di integrazione tra le fonti fiscali e la rilevazione Eu-Silc occorre verificare la compatibilità fra i tempi del processo dell'indagine e i tempi per l'acquisizione ed il trattamento dei dati fiscali. Le criticità del processo di integrazione tra fonti fiscali e dati campionari Eu-Silc riguardano principalmente i tempi di rilascio degli archivi fiscali da parte dell'Agenzia delle entrate e la standardizzazione dei protocolli di trasmissione. In merito ai tempi di rilascio, si può accertare un ritardo di circa 2 mesi nell'acquisizione dei dati delle dichiarazioni "730" e "Unico persone fisiche" anno 2004 (prevista per il 15 aprile e realizzata il 15 giugno), mentre per i dati "Cud" – anno 2004 – si è riscontrato un ritardo di sei mesi. A tali ritardi si aggiungono altre difficoltà connesse alla standardizzazione delle procedure di trasmissione e decriptazione dei dati.

Il prospetto 4.2 illustra sinteticamente i tempi di raccolta, registrazione, correzione, imputazione, validazione e consegna dei dati dell'indagine Eu-Silc relativi all'anno di indagine N¹⁹ (che raccoglie i redditi dell'anno N-1):

Prospetto 4.2 - Il flusso informativo di Eu-Silc e il processo di integrazione con le fonti amministrative

PERIODO DI RIFERIMENTO	PROCESSO INDAGINE Eu-Silc	PROCESSO DATI FISCALI
N-1	Preparazione del questionario	Autotassazione, ritenute sostituti
N	aprile-giugno	Estrazione del campione, stampa del questionario, formazione dei rilevatori
	settembre-dicembre	Interviste
N+1	gennaio-febbraio	Trasmissione dichiarazioni sostituti e percettori
	marzo	Istruttoria Istat richiesta dati
	marzo-aprile	Registrazione dati
	maggio-settembre	Correzione e imputazione dati qualitative
N+1	1° ottobre (a)	Imputazione redditi netti campionari (Iveware)
	30 marzo	Imputazione redditi lordi
N+2	30 marzo	Rilascio dati dall'Agenzie entrate
		Decriptazione, Lettura, Trattamento dati fiscali, record linkage, armonizzazione variabili, integrazione dati EuSilc
		Consegna dati trasversali ad Eurostat
		Consegna dati longitudinali ad Eurostat

(a) L'utilizzo dei dati amministrativi dell'Agenzia delle entrate consentirebbe, a norma dell'art. 10 del Regolamento Quadro, di spostare al 30 novembre di ciascun anno il termine per la consegna dei dati trasversali.

¹⁹ I redditi lordi saranno obbligatori solo a partire dall'edizione 2007, tuttavia, i paesi membri hanno assunto l'impegno di sviluppare opportune metodologie di stima o integrazione dei redditi netti campionari sin dal primo anno di indagine. In questa nota tecnica, ci si riferisce alla situazione a regime, quando cioè bisognerà consegnare i dati sui redditi sia netti sia lordi.

Per rispettare l'impegno di consegna dei dati trasversali, la richiesta Istat di invio dei dati delle dichiarazioni dei redditi dovrebbe partire entro la prima metà di gennaio dell'anno N+1 e dovrebbe essere soddisfatta al massimo entro aprile dello stesso anno. Si fa notare che l'articolo 10 del Regolamento Quadro di Eu-Silc consente un ritardo di consegna dei dati trasversali (alla fine del mese di novembre N+1) per quei Paesi che utilizzano dati amministrativi: nell'ipotesi di utilizzo dei dati di fonte tributaria, l'Istat potrebbe avanzare ad Eurostat la richiesta di ritardare la consegna, guadagnando due mesi di tempo. A proposito dei tempi di acquisizione dei dati tributari, occorre inoltre notare che la trasmissione telematica dei modelli Unico e 730 rende disponibili le informazioni contenute in tali modelli presso l'Agenzia delle entrate, prima che siano disponibili quelle relative ai contribuenti che hanno compilato soltanto il Cud. Quest'ultimo modello deve infatti essere "scorporato" dai modelli di dichiarazione dei sostituti di imposta e tale operazione richiede all'Agenzia delle entrate tempi di elaborazione più lunghi. Inoltre, la richiesta dei dati Cud relativi alle sezioni - Dati previdenziali Inps-Inpdap, comporta ulteriori carichi di lavoro per la Sogei (azienda deputata per conto dell'Agenzia delle entrate) con conseguenti ripercussioni sui tempi di consegna.

Poiché la categoria di contribuenti che hanno compilato soltanto il Cud (soprattutto lavoratori dipendenti e pensionati) è anche quella più facile da trattare nei modelli di microsimulazione, sembra opportuno chiedere all'Agenzia delle entrate di soddisfare la richiesta di trasmissione dati in due tempi, dando la massima priorità ai modelli Unico e 730. Infine, bisogna considerare che la scadenza per la consegna di dati longitudinali è posticipata al marzo dell'anno N+2, un arco di tempo che sembra offrire un margine abbastanza ampio per acquisire ed elaborare le informazioni tributarie. Per la componente longitudinale dell'indagine, quindi, l'utilizzo dei dati tributari non dovrebbe porre grosse difficoltà di tipo logistico e sembra la più indicata dal punto di vista della accuratezza.²⁰ Tuttavia, l'utilizzo dei dati fiscali per la sola componente longitudinale implica un rischio di incoerenza con i dati trasversali che occorre minimizzare.²¹

Qualora i tempi tecnici si rivelassero incompatibili con il rispetto della scadenza di consegna dei dati trasversali, è comunque possibile proporsi un utilizzo sub-ottimale delle informazioni dell'Agenzia delle entrate, selezionando i soli dati contenuti nel 730 e Unico persone fisiche. Infatti, il disegno campionario di Eu-Silc e la dilazione al 2007 dei redditi lordi suggeriscono l'opportunità di integrare le informazioni tributarie in un modello di microsimulazione dei dati trasversali anche con un ritardo temporale di un anno. Questa alternativa sarebbe comunque migliore rispetto alla costruzione di un modello di microsimulazione che ignori completamente le informazioni tributarie riferite all'anno precedente. In effetti, tre quarti delle famiglie campionarie vengono reintervistate l'anno successivo. Per queste famiglie, dunque, il modello di microsimulazione dei dati trasversali potrebbe utilizzare le informazioni dell'Agenzia delle entrate ritardate di un periodo. Più precisamente, le indagini dalla quarta edizione in poi disporrebbero di informazioni sulle variabili fiscali (ritardate di un anno rispetto a quelle necessarie) per tre quarti del campione, sulle stesse variabili ritardate di uno e due anni per metà del campione e, infine, sulle stesse variabili ritardate di uno, due e tre anni per un quarto del campione. Per il quarto di campione che, ogni anno, viene intervistato per la prima volta, si

²⁰ Naturalmente, ai fini dell'accuratezza occorrerà anche valutare con attenzione le differenze fra gli importi dichiarati nell'intervista e nelle dichiarazioni dei redditi, tenendo conto dei fenomeni (solo parzialmente sovrapposti) di reticenza durante l'intervista e di evasione/erosione fiscale. Per la costruzione delle variabili di reddito lordo si può adottare come regola generale (da validare nei casi particolari) di sommare ai redditi netti dichiarati nel corso dell'intervista le imposte ed i contributi sociali riportati nelle dichiarazioni fiscali. Infatti, per gli importi delle imposte pagate le dichiarazioni fiscali sono ovviamente la fonte più precisa ed attendibile.

²¹ E' anche per questa ragione che i dati tributari dovrebbero essere disponibili prima del processo di imputazione delle mancate risposte sui redditi netti.

dovrebbe utilizzare un modello di microsimulazione dei dati trasversali che incorpori opportune procedure di imputazione delle informazioni tributarie mancanti.²² Tali procedure di imputazione potrebbero infatti utilizzare l'informazione contenuta negli altri tre quarti del campione (in particolare, le relazioni fra le variabili fiscali e quelle campionarie).

In sintesi, si prospettano due possibili strategie di impiego dei dati dell'Agenzia delle entrate, a seconda della disponibilità dei dati stessi, dei tempi tecnici delle procedure di controllo e correzione e della possibilità di posticipare la scadenza per la parte trasversale al 30 novembre dell'anno N+1:

- utilizzo delle variabili tributarie misurate nello stesso periodo di riferimento del reddito sia per la componente longitudinale sia per la componente trasversale;
- utilizzo delle variabili tributarie misurate nello stesso periodo di riferimento del reddito per la componente longitudinale e, per la parte trasversale, delle stesse variabili ritardate nel contesto di un modello di microsimulazione integrato con procedure di correzione/imputazione dei dati mancanti.

La prima alternativa è certamente preferibile dal punto di vista metodologico (perché più accurata) e di efficienza (perché impiega meno risorse) ma espone al rischio di un'eccessiva dipendenza dai tempi di fornitura di dati esterni.²³ La seconda alternativa implica possibili scostamenti fra i dati trasversali microsimulati e quelli longitudinali ottenuti dall'Agenzia delle entrate in un tempo successivo, con possibili conseguenze negative sulla comunicazione dei risultati dell'indagine.

²² Questa possibilità è uno degli sviluppi previsti dei modelli di microsimulazione che saranno elaborati dai Paesi membri e da Eurostat. Un prototipo molto avanzato, applicato su dati del Panel europeo, è già stato realizzato da un gruppo di ricerca coordinato da V. Verma ed ha dato risultati che sono da considerare abbastanza soddisfacenti per un primo esperimento. Ai fini di Eu-Silc, tale modello dovrà essere integrato in modo più preciso per la parte che riguarda le normative tributarie e le altre caratteristiche specifiche dei singoli Paesi membri. A partire dal 2005, l'Istat collaborerà operativamente con Eurostat e con altri Paesi membri a questo scopo.

²³ I fattori non prevedibili che possono influire su questo rischio sono legati alla struttura dei modelli fiscali ed alla organizzazione delle modalità di acquisizione dei dati fiscali da parte dell'Agenzia delle entrate, che potrebbero variare nel tempo e determinare tempi di elaborazione e consegna all'Istat più lunghi.

Capitolo 5 - I regolamenti europei e la documentazione sulla qualità

5.1 - Introduzione

L'indagine Eu-Silc è disciplinata dal regolamento europeo n. 1177/2003 adottato dal Parlamento europeo e dal Consiglio del 16 giugno 2003.¹

Il regolamento in oggetto, noto come Regolamento quadro (*framework Regulation*), istituisce una struttura comune per la produzione sistematica di statistiche comunitarie sul reddito e sulle condizioni di vita, comprendente dati longitudinali e trasversali comparabili sul reddito, nonché sulla povertà e l'esclusione sociale a livello nazionale ed europeo; la comparabilità dei dati tra i Paesi Membri viene attuata attraverso metodologie e studi eseguiti in stretta collaborazione tra Eurostat e i Paesi stessi.

Gli aspetti più importanti disciplinati nel Regolamento quadro sono: gli scopi dell'indagine Eu-Silc, i riferimenti temporali (data di inizio della rilevazione, periodo di riferimento per la rilevazione del reddito, eccetera), le caratteristiche dei dati (trasversali e longitudinali), le variabili obbligatorie nella rilevazione, le unità di rilevazione (famiglie, persone), il disegno campionario e le regole di inseguimento (*tracing rules*), la dimensione del campione, i tempi e le modalità di trasmissione ad Eurostat dei dati rilevati; i tempi e le modalità di diffusione dei dati, l'eventuale permesso di accesso ai microdati per motivi scientifici, i contributi di finanziamento ai Paesi Membri per l'esecuzione dell'indagine, la produzione di studi e relazioni dettagliate sulla qualità dei dati e dell'indagine.

Gli aspetti citati nel Regolamento quadro sono approfonditi in regolamenti successivi più specifici (detti Regolamenti attuativi) e/o in documenti tecnici predisposti da Eurostat.

Il presente capitolo è dedicato alla descrizione del Regolamento attuativo n. 28/2004 della Commissione del 5 gennaio 2004² relativo alla produzione e ai contenuti delle relazioni sulla qualità (*quality report*) e sull'attuazione dell'indagine.

5.2 - Il Regolamento europeo n. 28/2004 e i *quality report*

Gli Istituti nazionali di statistica accompagnano la diffusione delle stime di un'indagine campionaria con adeguate relazioni contenenti informazioni ed indicatori per: a) mettere a conoscenza l'utente delle metodologie utilizzate nella rilevazione e nel processo dei dati; b) valutare la precisione delle stime fornite ai fini di una loro corretta utilizzazione.

Per l'indagine Eu-Silc, la produzione e la diffusione di tale documentazione è disciplinata dal Regolamento attuativo n. 28/2004 della Commissione del 5 gennaio, in un'ottica di comparabilità delle relazioni prodotte dai singoli Istituti.

Il Regolamento in oggetto stabilisce il numero, la tipologia, il contenuto dettagliato e i tempi di consegna delle relazioni sulla qualità (*quality report*) che gli Istituti nazionali di statistica devono trasmettere ad Eurostat.

Il capitolo è a cura di Claudia Rinaldelli

¹ Legge 165/1 Official Journal of the European Union 3.7.2003.

² Legge 5/42 Official Journal of the European Union 9.1.2004.

Entrando nel dettaglio, con riferimento alla rilevazione eseguita nell'anno N, è obbligatorio produrre due *quality report*, intermedio e finale, per la componente trasversale e longitudinale dell'indagine. Il *quality report intermedio* deve essere prodotto *solo* per la componente trasversale dell'indagine e trasmesso entro la fine dell'anno N+1 mentre il *quality report finale* deve essere prodotto sia per la componente trasversale che per quella longitudinale e trasmesso entro la fine dell'anno N+2. Il prospetto 5.1 riassume la tipologia dei *quality report* e i relativi tempi di consegna con riferimento alla rilevazione eseguita nell'anno N.

Prospetto 5.1 - Tipologia di *quality report* e anno di consegna (rilevazione dell'anno N)

QUALITY REPORT	Anno di consegna	
	Anno N+1	Anno N+2
Intermedio componente trasversale	SI	---
Finale componente trasversale	---	SI
Finale componente longitudinale	---	SI

Col termine *componente longitudinale*, Eurostat identifica non solo il sottocampione ruotato della durata di quattro anni, ma anche i sottocampioni ruotati della durata rispettivamente di tre e due anni che contengono i due anni più recenti della rilevazione; ad esempio, partendo dalla rilevazione dell'anno N, all'anno N+3 sono presenti tre componenti longitudinali rispettivamente di durata quattro, tre e due anni, ossia: (N, N+1, N+2, N+3), (N+1, N+2, N+3), (N+2, N+3).

Il *quality report* finale longitudinale va trasmesso per tutte le componenti longitudinali specificate e non solo per quella di durata massima (quattro anni). Tenendo conto di ciò e del fatto che lo schema di rotazione del disegno dell'indagine sarà a regime nell'anno 2007³, il numero e la tipologia dei *quality report* varierà annualmente nel periodo 2005-2009. Il prospetto 5.2 riassume i *quality report* che dovranno essere consegnati ad Eurostat nel suddetto periodo.

Prospetto 5.2 - Tipologia di *quality report* per anno di rilevazione e componente dell'indagine nel periodo 2005-2009

ANNO DI CONSEGNA	Tipologia di <i>quality report</i> per anno di rilevazione e componente dell'indagine
2005	Intermedio trasversale rilevazione 2004
2006	Intermedio trasversale rilevazione 2005, finale trasversale rilevazione 2004
2007	Intermedio trasversale rilevazione 2006, finale trasversale rilevazione 2005, finale longitudinale 2004-2005
2008	Intermedio trasversale rilevazione 2007, finale trasversale rilevazione 2006, finale longitudinale 2004-2005-2006, finale longitudinale 2005-2006
2009	Intermedio trasversale rilevazione 2008, finale trasversale rilevazione 2007, finale longitudinale 2004-2005-2006-2007, finale longitudinale 2005-2006-2007, finale longitudinale 2006-2007

³ L'Istat ha eseguito la prima rilevazione nell'anno 2004.

5.3 - Il contenuto dei *quality report*

Il Regolamento n. 28/2004 stabilisce in maniera dettagliata il contenuto dei *quality report* intermedio e finale. In particolare:

il *quality report intermedio trasversale* deve contenere:

- stime
la stima degli indicatori di Laeken (linea di povertà relativa, incidenza di povertà relativa, indice di disuguaglianza della distribuzione del reddito, gender pay gap, intensità di povertà relativa, rapporto interquintilico) e del reddito medio disponibile equivalente; ciascuna stima deve essere corredata dai corrispondenti errori di campionamento;
- disegno dell'indagine e di campionamento
descrizione di: disegno di campionamento adottato, unità campionarie ai vari stadi di selezione, criteri e variabili di stratificazione, criteri di definizione della dimensione campionaria e sua allocazione, criteri di selezione delle unità campionarie, schema di rotazione delle unità campionarie;
- sistemi di ponderazione
descrizione di: criteri di correzione dei pesi base per mancata risposta totale, criteri di calibrazione per il calcolo dei pesi finali e variabili utilizzate;
- sostituzioni
descrizione del criterio di selezione delle unità suppletive, analisi delle unità campionarie sostituite per dimensione della famiglia, età e genere, livello di istruzione, attività svolta, eccetera, in assoluto e comparata rispetto alle caratteristiche delle unità suppletive; questa sezione del *quality report* non viene compilata da Istat in quanto la sostituzione non è ammessa nell'indagine in oggetto;
- errori non campionari
descrizione di: liste di campionamento utilizzate in termini di qualità, aggiornamento e problemi di copertura incontrati, modalità di costruzione e verifica del questionario, numero di giorni di istruzione ai rilevatori, durata media dell'intervista, numero di famiglie rispondenti, numero di persone con più di 16 anni rispondenti, numero di famiglie/individui per mancata risposta parziale rispetto alla variabile reddito e ai suoi componenti (da lavoro, da pensione, eccetera), numero di famiglie/individui con valori imputati rispetto alla variabile reddito e ai suoi componenti (da lavoro, da pensione, eccetera);
- comparabilità
descrizione delle definizioni nazionali se utilizzate al posto di quelle stabilite da Eurostat con particolare riferimento alla popolazione oggetto di studio, definizione di famiglia privata e componente della famiglia, periodo di riferimento per la rilevazione del reddito e dei suoi componenti (da lavoro, da pensione, eccetera), eventuali fonti utilizzate per il reddito in alternativa alla rilevazione, durata totale delle operazioni di rilevazione;
- coerenza
descrizione degli eventuali confronti effettuati tra il reddito rilevato (e suoi componenti) e fonti esterne.

Il *quality report finale trasversale* deve contenere:

- stime
la stima del valore medio nazionale per ogni componente del reddito familiare e individuale (reddito da lavoro, da pensione, eccetera); la stima del valore medio del reddito disponibile equivalente per genere, età, dimensione della famiglia; ciascuna stima deve essere corredata dai corrispondenti errori di campionamento;

- procedura di imputazione
la descrizione della procedura di imputazione utilizzata, se diversa dai metodi proposti da Eurostat, le variabili imputate e la percentuale di imputazione sul numero totale di osservazioni del campione; il numero di osservazioni prima e dopo l'imputazione per ogni componente del reddito familiare e individuale (reddito da lavoro, da pensione, eccetera);
- fitti figurativi
la descrizione della procedura utilizzata per calcolare il fitto figurativo se diversa dal metodo proposto da Eurostat;
- autovettura aziendale
la descrizione della procedura utilizzata per imputare un valore all'uso privato dell'autovettura aziendale.

Il *quality report finale longitudinale* deve contenere:

- stime
la stima degli indicatori di Laeken calcolati sulla componente longitudinale dell'indagine Eu-Silc (persistenza dell'incidenza di povertà relativa); per ogni rilevazione (*wave*) della componente longitudinale, la stima del valore medio nazionale di ogni componente del reddito familiare e individuale (reddito da lavoro, da pensione, eccetera), la stima del valore medio del reddito disponibile equivalente per genere, età, dimensione della famiglia; ciascuna stima deve essere corredata dai corrispondenti errori di campionamento;
- disegno dell'indagine e di campionamento
per la prima rilevazione della componente longitudinale, la descrizione di: disegno di campionamento adottato, unità campionarie ai vari stadi di selezione, criteri e variabili di stratificazione, criteri di definizione della dimensione campionaria e sua allocazione, criteri di selezione delle unità campionarie, schema di rotazione delle unità campionarie;
- sistemi di ponderazione
per la prima rilevazione della componente longitudinale, la descrizione di: criteri di correzione dei pesi base per mancata risposta totale, criteri di calibrazione e variabili utilizzate; calcolo dei pesi longitudinali finali; per la seconda e successive rilevazioni della componente longitudinale, la descrizione di: criteri di correzione dei pesi base per mancata risposta totale, criteri di calibrazione e variabili utilizzate; calcolo dei pesi longitudinali finali; calcolo dei pesi trasversali finali;
- sostituzioni
per la prima rilevazione della componente longitudinale, la descrizione del criterio di selezione delle unità suppletive, analisi delle unità campionarie sostituite per dimensione della famiglia, età e genere, livello di istruzione, attività svolta eccetera, in assoluto e comparata rispetto alle caratteristiche delle unità suppletive; questa sezione del *quality report* non viene compilata da Istat in quanto la sostituzione non è ammessa nell'indagine in oggetto;
- errori non campionari
per la prima rilevazione della componente longitudinale, la descrizione di: liste di campionamento utilizzate in termini di qualità, aggiornamento e problemi di copertura incontrati; per ogni rilevazione della componente longitudinale, la descrizione di: modalità di costruzione e verifica del questionario, numero di giorni di istruzione ai rilevatori, durata media dell'intervista, numero delle famiglie rispondenti, numero delle persone con più di 16 anni rispondenti; per la seconda e successive rilevazioni della componente longitudinale, il tasso di risposta per rilevazione (famiglie, individui), il tasso di seguito longitudinale (famiglie e individui); per ogni rilevazione della componente longitudinale, la distribuzione delle famiglie per numero dei contatti, per stato di accettazione

dell'intervista, la percentuale di valori mancanti per ogni componente del reddito (reddito da lavoro, da pensione, eccetera) sia a livello familiare che personale, la percentuale di informazioni parziali per ogni componente del reddito (reddito da lavoro, da pensione, eccetera) sia a livello familiare che personale;

- procedura di imputazione
per ogni rilevazione della componente longitudinale, la descrizione della procedura di imputazione utilizzata, se diversa dai metodi proposti da Eurostat;
- fitti figurativi
per ogni rilevazione della componente longitudinale, la descrizione della procedura utilizzata per calcolare il fitto figurativo se diversa dal metodo proposto da Eurostat;
- autovettura aziendale
per ogni rilevazione della componente longitudinale, la descrizione della procedura utilizzata per imputare un valore all'uso privato dell'autovettura aziendale;
- comparabilità
per ogni rilevazione della componente longitudinale, l'eventuale descrizione delle definizioni nazionali se utilizzate al posto di quelle stabilite da Eurostat con particolare riferimento alla popolazione oggetto di studio, definizione di famiglia privata e componente della famiglia, periodo di riferimento per la rilevazione del reddito e dei suoi componenti (reddito da lavoro, da pensione, eccetera), eventuali fonti utilizzate per il reddito in alternativa alla rilevazione;
- regole di inseguimento
le regole di inseguimento (*tracing rules*) adottate se differenti da quelle stabilite da Eurostat;
- coerenza
per ogni rilevazione della componente longitudinale, la descrizione degli eventuali confronti effettuati tra il reddito rilevato (e suoi componenti) e fonti esterne.

5.4 - I *quality report* prodotti da Eurostat

Il Regolamento attuativo n. 28/2004 stabilisce il numero, la tipologia, il contenuto e i tempi di consegna dei *quality report* comparativi che devono essere prodotti da Eurostat.

Con riferimento alla rilevazione eseguita nell'anno N, Eurostat ha l'obbligo di produrre il *quality report* intermedio trasversale entro giugno dell'anno N+2 e il *quality report* finale comprensivo della componente trasversale e longitudinale entro giugno dell'anno N+3.

Tali *quality report* costituiscono un quadro comparativo delle relazioni sulla qualità prodotte dai singoli Istituti nazionali di statistica; contengono inoltre informazioni sulla tempestività e puntualità dei Paesi Membri quali ad esempio, il ritardo medio dei dati non trasmessi puntualmente, i motivi del ritardo, eccetera.

Capitolo 6 - Comparabilità internazionale e problemi metodologici aperti

6.1 - Comparabilità dei dati a livello internazionale

Il Regolamento Quadro del progetto Eu-Silc stabilisce, nel suo primo articolo, che la comparabilità internazionale dei dati è uno scopo fondamentale, da perseguire con opportuni studi metodologici, condotti in stretta collaborazione fra gli Stati membri ed Eurostat. In questo, il Regolamento Quadro riflette l'insufficienza delle *best practices* relative a questo importante aspetto. In effetti, il progetto Eu-Silc è caratterizzato sin dall'inizio da due scopi paralleli: (i) fornire un insieme di statistiche armonizzate sui redditi e sulle condizioni di vita sulla base delle migliori pratiche suggerite dall'esperienza precedente e: (ii) migliorare lo stato dell'arte per quanto riguarda la rilevazione e la misura dei redditi e delle condizioni di vita. Dato lo scopo generale del progetto, non sorprende che la comparabilità internazionale risulti essere una delle principali sfide metodologiche da affrontare.

Lo stesso concetto di comparabilità internazionale, tuttavia, non è immediatamente evidente. Secondo Verma,¹ la comparabilità dei dati di un'indagine:

“[...] può eludere una definizione precisa [...] significa che i dati (le osservazioni) per popolazioni differenti possono (in un senso statisticamente valido) essere legittimamente messi insieme (aggregati), confrontati fra loro (differenziati) ed essere interpretati (dotati di significato) [...]. La comparabilità è un concetto relativo: possiamo soltanto avere gradi relativi di comparabilità”.

La spiegazione precedente presenta il vantaggio di mettere a fuoco due aspetti cruciali della comparabilità. Il primo si riferisce alle molteplici dimensioni della comparabilità (“*di che cosa è fatta la comparabilità*”). La comparabilità è, allo stesso tempo, definita come una proprietà dei dati, delle statistiche usate per aggregare i dati e, inoltre, delle interpretazioni delle statistiche. Il secondo aspetto importante della comparabilità è il suo carattere relativo. Si possono avere soltanto misure ordinali della comparabilità, basate su valutazioni come: “l'insieme di dati (di statistiche, di interpretazioni) A è più adatto a confronti internazionali rispetto all'insieme B”. Si noti come tali giudizi ordinali siano di fatto l'unico modo per valutare il successo dei tentativi di produrre dati (indicatori, analisi) *armonizzati* a livello internazionale, essendo la comparabilità lo scopo ultimo dell'armonizzazione.

6.1.1 Comparabilità: una definizione

Il problema della comparabilità internazionale delle *statistiche* campionarie è stato esplorato a fondo soprattutto da Vijay Verma. In questo paragrafo, invece, l'attenzione sarà rivolta alla comparabilità dei redditi di *unità statistiche* di Paesi differenti (micro-comparabilità). La comparabilità a livello micro, cioè fra i redditi di tutte le coppie di famiglie/individui che vivono

Il paragrafo 6.1 è a cura di Marco Di Marco, il paragrafo 6.2 è a cura di Andrea Cutillo e Davide Di Laurea, il paragrafo 6.3 è a cura di Gabriella Donatiello

¹ Si veda Verma 2002.

in Paesi diversi o nello stesso Paese, è una condizione necessaria per poter condurre confronti *significativi* di benessere.

Spesso, quando i confronti internazionali sono condotti su dati aggregati, per esempio quelli di Contabilità Nazionale, la comparabilità delle unità statistiche viene di fatto sottovalutata (per meglio dire, non svolge necessariamente un ruolo centrale). Questa pratica può essere accettata per le descrizioni su grande scala che considerano ogni Paese come una economia equivalente ad un solo consumatore aggregato (*one consumer equivalent economy*). Tuttavia, in un contesto così altamente stilizzato, non è possibile dire nulla sulla disegualianza dei redditi dentro e fra i Paesi. Le analisi distributive richiedono in realtà la comparabilità fra unità statistiche. Un passaggio preliminare per il calcolo corretto della maggior parte dei più noti indici di disegualianza è infatti l'ordinamento delle unità statistiche dalla più povera alla più ricca (*Pen's parade*). Chiaramente, per ordinare le famiglie o gli individui di un Paese in questo modo, i redditi di tutte le coppie di unità statistiche devono essere confrontati fra loro.

Si può così distinguere fra micro-comparabilità *dentro* un singolo Paese ("tutte le coppie di unità statistiche del Paese A possono essere confrontate") e *fra* i Paesi ("qualsiasi unità statistica del Paese A può essere confrontata con qualsiasi unità statistica del Paese B"). Tutt'e due sono condizioni necessarie per la comparabilità internazionale, com'è facile intuire. La micro-comparabilità dentro un Paese, come si è già notato, è condizione necessaria (ma non sempre anche sufficiente) per costruire e confrontare fra loro gli indicatori di disegualianza "nazionali". La micro-comparabilità fra Paesi è invece richiesta soprattutto quando un gruppo di Paesi deve essere confrontato rispetto ad un parametro comune, come per esempio una linea di povertà europea.

Per definire più precisamente la nozione di comparabilità, può essere utile a questo punto distinguere fra (i) comparabilità dei *dati* e (ii) comparabilità delle *statistiche* e delle *interpretazioni/analisi* (comparabilità dei livelli di benessere, nel caso dei redditi). Mentre la prima è responsabilità soprattutto dei produttori di dati, le seconde riguardano principalmente le analisi a posteriori sui dati disponibili.

Ovviamente, le statistiche sul reddito di Paesi differenti (regioni, sottopopolazioni) possono essere confrontate ed interpretate chiaramente e correttamente solo quando sono calcolate su dati comparabili. Quindi, i due tipi di comparabilità possono essere ordinati gerarchicamente: la comparabilità dei dati di reddito può essere considerata come una condizione necessaria per la comparabilità dei livelli di benessere. Tuttavia, la condizione non è sempre anche sufficiente. Per esempio, se anche i redditi monetari fossero misurati senza errori, sorgerebbero ancora dei problemi nel confronto dei livelli di benessere delle famiglie (individui) che vivono in Paesi diversi. Il potere di acquisto dei redditi monetari dipende dai livelli dei prezzi e, pertanto, le parità di potere d'acquisto (Ppa) sono quindi un complemento importante delle analisi comparative delle condizioni economiche delle famiglie. E tuttavia questo tipo di problemi è sostanzialmente esterno al processo di produzione dei dati, poiché in effetti il bisogno di un insieme di Ppa non implica deviazioni dal requisito di base dell'accuratezza nella misura dei redditi monetari. Similmente, l'uso delle scale di equivalenza pone problemi di comparabilità che, almeno a prima vista, non richiedono una metodologia particolare nella misura dei redditi monetari.²

Al contrario, la sottostima dei redditi richiede miglioramenti nell'accuratezza delle misure e l'esistenza di componenti non-monetarie (auto-produzione, affitti imputati, trasferimenti sociali in natura) rende necessaria una definizione più completa di reddito. Gli sviluppi recenti nel

² Tuttavia, poiché non è possibile applicare le scale di equivalenza ai redditi negativi, la presenza di questi ultimi può suggerire l'adozione di una ricodifica dei redditi inferiori ad una certa soglia minima (*bottom coding*).

dibattito circa la natura multidimensionale della qualità della vita, infine, richiedono soltanto informazioni supplementari sulle condizioni di vita. In sintesi, si può dire che la comparabilità dei *dati* può essere una condizione sufficiente per la comparabilità dei livelli di benessere se il processo di produzione dei dati (e/o dati aggiuntivi esterni) include anche tutte le ulteriori informazioni che sono necessarie a questo fine.

Insieme all'accuratezza, naturalmente, per la comparabilità internazionale è anche necessaria la consistenza semantica delle definizioni di reddito (inclusa la completezza). Questo requisito "linguistico", apparentemente scontato, non è necessariamente garantito dalla traduzione nelle diverse lingue nazionali dei termini riferiti alle diverse componenti di reddito. I casi controversi riguardano sia situazioni in cui ad uno stesso termine (per esempio, "*family allowance*") di cui esistono traduzioni letterali accettate come non ambigue ("assegno familiare", "allocation familiales", eccetera) corrispondono nei diversi Paesi realtà economiche diverse, sia situazioni in cui un termine nazionale non ha corrispondenze, né linguistiche, né economiche, negli altri Paesi (per esempio, la Cassa Integrazione Guadagni italiana). Nel progetto Eu-Silc, la coerenza semantica è garantita dall'armonizzazione dei significati delle variabili-obiettivo comuni a tutti i Paesi, ciascuno dei quali ha poi ampia autonomia nel definire la migliore strategia di raccolta dei dati nazionali, anche per quanto riguarda gli usi linguistici più appropriati (per esempio, il *wording* dei questionari).

6.1.2 *Comparabilità come "accuratezza significativa"*

Una definizione di comparabilità potrebbe inizialmente essere espressa come segue:

La comparabilità dei dati di reddito è una proprietà (un insieme di proprietà) dei processi di produzione di dati (input, tecniche e risultati) che consente confronti significativi all'interno e fra i Paesi (regioni, sottogruppi), fra tutte le coppie di unità statistiche.

Tuttavia, la definizione appena proposta è circolare, poiché descrive la comparabilità in termini di confronti (comparazioni) significativi. Per eliminare la circolarità, il termine "confronto" deve essere spiegato. In matematica, gli elementi di un insieme sono comparabili se, per qualunque coppia di elementi x e y dell'insieme, esiste una relazione R tale che almeno una delle due dichiarazioni seguenti risulta vera:

$$x R y; \quad y R x$$

Ognuna delle due proposizioni precedenti (così come la loro unione logica) è un confronto. Per la variabile di reddito, una scelta ovvia per il confronto è la relazione "uguale o maggiore", che consente di ordinare le unità statistiche dalla più povera alla più ricca. Ma questo implica che, almeno nel caso del reddito, l'accuratezza aritmetica costituisce il nucleo della nozione di comparabilità. Insieme ad appropriate condizioni relative all'interpretazione corretta dei risultati ("quali confronti sono significativi?"), l'accuratezza è anche tutto quello che di cui abbiamo bisogno per la comparabilità dei livelli di benessere.

Nella sua ovvietà, si tratta di una conclusione confortante se si pensa che, a prima vista, la comparabilità internazionale si presenta come un elusivo concetto multidimensionale. Incidentalmente, si può notare che la vaghezza concettuale sparisce non appena la micro-comparabilità entra in gioco.

La richiesta di confronti *significativi* corrisponde al requisito di consistenza semantica e stabilisce un ponte fra la comparabilità dei dati e la comparabilità dei livelli di benessere: i

confronti possono essere interpretati correttamente (cioè, possono essere significativi) se sono adatti alle analisi del benessere.

Il requisito di completezza, incluso nella nozione di consistenza semantica, può essere precisato nel modo seguente:

Per tutte le coppie di unità statistiche, la relazione \geq fra i loro redditi può essere stabilita correttamente se la definizione di reddito è completa, cioè se include tutte le componenti necessarie per la comparabilità dei livelli di benessere.

Eventuali conflitti fra l'accuratezza aritmetica della misura del reddito e la completezza della definizione di reddito non dovrebbero essere decisi a scapito della comparabilità dei livelli di benessere, dato che questa decisione si tradurrebbe in un'accuratezza priva di significato. L'esclusione dal reddito di una componente "difficile da misurare" si riflette necessariamente in un'accuratezza inutile, perchè non significativa dal punto di vista dei confronti di benessere. Alla luce della definizione della comparabilità come "accuratezza significativa", si può in realtà sostenere che nessun conflitto può sorgere fra accuratezza e comparabilità significativa, poiché si tratta essenzialmente di modi differenti di esprimere lo stesso concetto. L'accuratezza aritmetica e la completezza della definizione di reddito devono essere perseguite simultaneamente.

La definizione della comparabilità dei redditi può essere sintetizzata come segue:

Per quanto riguarda la raccolta di dati sui redditi delle famiglie e/o degli individui, la comparabilità dei dati sui redditi richiede la comparabilità delle unità statistiche all'interno e fra i Paesi (micro-comparabilità). La micro-comparabilità, a sua volta, coincide essenzialmente con l'accuratezza significativa. Cioè, con l'accuratezza aritmetica insieme alla consistenza semantica delle definizioni di reddito adottate nei vari Paesi (regioni, sottogruppi). Se un grado accettabile di accuratezza significativa è raggiunto dai processi di produzione di dati, la micro-comparabilità è inoltre una condizione sufficiente per l'uso appropriato dei dati di reddito nei confronti dei livelli di benessere nei diversi Paesi (comparabilità dei livelli di benessere).

L'ultima frase non dovrebbe essere interpretata come una pretesa di sufficienza dei redditi osservati per le analisi del benessere. Significa piuttosto che le ulteriori correzioni ed aggiunte necessarie per i confronti di benessere (quali le Ppa, le scale di equivalenza, le condizioni di vita eccetera) non interessano direttamente il problema della osservazione del reddito monetario delle famiglie e/o degli individui. Quindi, per valutare (stabilire, migliorare) la comparabilità dei dati di reddito, è necessario e sufficiente dare un'occhiata da vicino ai processi di produzione dei dati, controllandone l'accuratezza significativa.

6.1.3 L'integrazione delle fonti per migliorare la comparabilità dei dati italiani: i redditi autonomi

In quanto segue, si concentra l'attenzione su uno dei problemi metodologici più complessi relativi alla comparabilità, vale a dire la sottostima di alcune componenti di reddito, quali i redditi da lavoro autonomo. Da un lato, l'inclusione delle componenti sottostimate riduce l'accuratezza e, pertanto, la comparabilità "aritmetica", basata sulla relazione "maggiore o eguale". D'altro lato, l'esclusione delle componenti di reddito sottostimate viola il principio di completezza e diminuisce la comparabilità "semantica" dei dati. In linea di principio, poiché riduce comunque la comparabilità, la sottostima dei redditi non dovrebbe essere affrontata né alterando la definizione di reddito né ammettendo misure meno accurate. In pratica, tuttavia,

molti ricercatori e produttori di dati sono costretti ad accettare un certo, e di solito sconosciuto, grado di sottostima. È importante notare, allo stadio attuale della discussione, che l'opinione secondo la quale dati influenzati da gradi simili di sottostima siano abbastanza comparabili fra loro è insostenibile, almeno nella misura in cui la micro-comparabilità è implicata.

In caso di sottostima dei redditi, alcuni confronti a livello aggregato possono forse essere effettuati applicando opportuni fattori di correzione, noti o stimati. Purtroppo, poiché il reddito non osservato non è uniformemente distribuito fra le unità statistiche, un'analogia semplificata non può essere applicata al livello micro. Soltanto se tutti i redditi in un Paese fossero sottostimati nella stessa misura (nota o approssimata) dell' α per cento, allora non sorgerebbero problemi di micro-comparabilità. Basterebbe in questo caso ipotetico applicare a tutti i redditi un fattore correttivo, simile ad un tasso di cambio, "come se" i redditi fossero stati misurati in un'altra valuta. In effetti, la sottostima dei redditi è un problema importante proprio perché impedisce la micro-comparabilità e, di conseguenza, anche la comparabilità dei livelli di benessere. Per quanto possa essere difficile, la sottostima dei redditi dovrebbe essere minimizzata comunque, considerando in ogni Paese il reddito vero (al livello micro) come il termine di riferimento per la comparabilità internazionale.

Per fare fronte agli ambiziosi obiettivi del progetto Eu-Silc, l'Istituto nazionale di statistica italiano ha implementato una strategia mista di raccolta dei dati, basata su interviste Papi ("carta e penna") e sull'integrazione di dati amministrativi con i risultati dell'indagine.

Una prima importante questione semantica riguarda proprio la definizione dei redditi da lavoro indipendente. Le definizioni economica, contabile e amministrativa dei redditi da lavoro indipendente non necessariamente coincidono e generano problemi di affidabilità e comparabilità delle diverse fonti. Inoltre, le definizioni hanno un'influenza sulla comprensione soggettiva del termine "reddito" da parte degli intervistati.

Le fonti alternative di microdati sui guadagni da lavoro indipendente non necessariamente contengono il reddito disponibile come tale. I dati campionari possono essere sottodichiarati (*under-reporting*). I dati amministrativi, relativi alle dichiarazioni dei redditi, non tengono conto, naturalmente, dell'evasione fiscale illecita e potrebbero inoltre non riportare tutte le deduzioni dal reddito imponibile autorizzate (erosione fiscale). I libri contabili, a loro volta, solitamente riportano le tasse pagate dall'azienda come entità giuridica ma non contengono informazioni sulle imposte personali dei soci.

Comunque, ignorando l'evasione fiscale, i profitti contabili al netto delle imposte sull'impresa possono essere considerati come una misura del reddito personale lordo. Ciò nonostante, potrebbero ancora essere differenti dal reddito imponibile personale. Infatti, l'amministrazione tributaria prevede deduzioni speciali per i profitti reinvestiti nell'attività, stipula eccezioni alle regole contabili per i costi di ammortamento eccetera. Alcune categorie di contribuenti (per esempio le imprese familiari, i coltivatori, le imprese di nuova formazione, eccetera) potrebbero essere soggette ad un regime di imposta preferenziale, caratterizzato da benefici speciali.

Un'altra questione semantica riguarda la ripartizione dei guadagni dei lavoratori indipendenti fra le categorie dei redditi da lavoro e quella dei redditi da capitale. A questo riguardo, le convenzioni contabili e quelle sottese alle norme tributarie non necessariamente sono le più adatte per l'analisi economica e, inoltre, possono anche ostacolare la comparabilità internazionale. Il sistema dei conti nazionali opportunamente riassume le due componenti nella categoria unica del "reddito misto", una convenzione che consente di analizzarle come ricompense per il lavoro, assistito spesso dal capitale di proprietà dell'individuo. Il manuale di

Canberra³ e la risoluzione dell'Oil sulle indagini sui redditi⁴ suggeriscono di escludere dai redditi da lavoro indipendente i profitti distribuiti a soci che non lavorano nell'impresa.

Data l'ambiguità della definizione dei redditi da lavoro indipendente nella normativa tributaria, nella versione italiana del progetto Eu-Silc la fonte amministrativa è stata usata con prudenza (sostanzialmente, per controllare e sostituire i redditi campionari sottodichiarati). Per evitare errori dovuti alle definizioni legali, quando i redditi di un individuo non occupato sono riportati nei dati tributari esclusivamente come redditi da impresa (che includono anche i guadagni di molti lavoratori autonomi), le informazioni non sono state usate (cioè non sono state confrontate con i redditi campionari, né sono state incluse nel dataset finale).

Nel progetto Eu-Silc, la procedura standard per misurare i redditi netti da lavoro autonomo richiede di osservare tramite intervista l'importo ricavato dall'attività indipendente soltanto quando i profitti contabili o il reddito imponibile non sono disponibili. Per la versione italiana di Eu-Silc, quando sia la fonte campionaria sia quella amministrativa lo riportano, il reddito da lavoro autonomo è invece determinato dal valore massimo fra: (i) il reddito (netto) da lavoro indipendente nella dichiarazione dei redditi e: (ii) il reddito (netto) da lavoro indipendente rilevato nell'indagine.

Questa differenza rispetto alla definizione standard è stata adottata per minimizzare la sottostima dovuta all'evasione/erosione nei dati amministrativi o alla sottodichiarazione nei dati campionari, a seconda di quale delle due sia maggiore. La procedura aumenta il grado di comparabilità internazionale dei dati italiani, nell'ipotesi che il reddito da lavoro indipendente nel Paese di riferimento non sia sottostimato.

Le due fonti di dati non coincidono perfettamente. Infatti, alcuni individui riportano redditi da lavoro indipendente soltanto in una delle due. È questo il caso di alcuni individui la cui condizione professionale al momento dell'intervista è diversa da quella del periodo di riferimento del reddito, ma anche di molti percettori di redditi secondari da lavoro indipendente (e/o di piccolo importo). I dati campionari includono come redditi da lavoro indipendente quelle piccole remunerazioni per servizi informali che sono frequentemente trascurate in sede di dichiarazione dei redditi, se non altro perché rientrano nell'area di esenzione totale (*no tax area*). Per esempio, i guadagni delle babysitter. D'altra parte, alcuni redditi secondari da lavoro autonomo riportati nelle dichiarazioni dei redditi possono essere ignorati durante l'intervista per alleggerire il carico di risposta.

Nel questionario, l'importo del reddito da lavoro indipendente viene chiesto dopo una domanda di controllo, che elenca gli usi personali dei guadagni (consumo e risparmio). Questa sequenza è stata concepita per suggerire all'intervistato un'interpretazione il più possibile vicina al concetto di "soldi tolti dall'attività" (*money drawn out*) richiesta dalla procedura standard. In effetti, per molti autonomi la traduzione letterale in italiano della domanda: "ha tolto soldi dall'attività per il suo uso personale?" può evocare l'idea di un comportamento illecito come, per esempio, ritirare i soldi dalla cassa senza annotare nei libri contabili i ricavi corrispondenti (evasione fiscale). Per quanto riguarda l'ammontare del reddito da lavoro indipendente, le istruzioni agli intervistatori raccomandano di spiegare all'intervistato che:

- l'attività di lavoro indipendente ha portato a:
- guadagni, se l'individuo o la sua famiglia hanno ricavato da questa soldi che sono stati usati per le spese familiari/personali, risparmiati, investiti nell'attività o in strumenti finanziari, in abitazioni o altri beni immobili;

³ Cfr. Canberra Group 2001.

⁴ Si veda Ilo 2003.

- una perdita, se l'individuo non ha ottenuto da questa soldi per pagare le spese familiari/personali (o per risparmiare/investire) e, inoltre, se ha usato redditi da altre fonti, soldi presi in prestito o venduto parte del patrimonio per pagare i costi dell'attività di lavoro indipendente.”

Il motivo per una tale definizione è abbastanza ovvio: se i guadagni positivi sono soldi “tolto dall'attività”, allora le perdite dovrebbero essere intese come soldi “messi nell'attività”.

Durante le indagini pilota, molti autonomi si sono trovati più a loro agio con la semplice logica della definizione precedente che con il concetto di reddito contabile (per non parlare del calcolo del reddito imponibile, un'operazione che è solitamente delegata al commercialista).

Per minimizzare la percentuale di mancate risposte alla domanda sul reddito, per quei rispondenti che non ricordano l'importo esatto del loro reddito da lavoro indipendente, una domanda supplementare chiede un importo approssimativo, da selezionare all'interno di una lista predeterminata. Agli intervistatori è stato ripetutamente raccomandato di non costringere le persone confuse o visibilmente infastidite a dare una risposta contro la loro volontà, poiché potrebbe verosimilmente trattarsi di risposte false. Come principio generale, le mancate risposte sono state preferite sempre a quelle false.

Inoltre, agli intervistatori è stato chiesto di fornire la loro propria valutazione, dopo l'intervista, sull'affidabilità dei redditi riportati. La strategia di rilevazione dei redditi da lavoro autonomo attraverso interviste personali punta in complesso a minimizzare gli errori e, allo stesso tempo, a favorire le procedure di imputazione dei valori mancanti. La messa a punto delle procedure di imputazione è stata facilitata, da un lato, dalle informazioni qualitative disponibili nell'indagine e, d'altro lato, dalla riduzione della distorsione dovuta all'inclusione di risposte inaffidabili fra i casi validi. Queste ultime sono state minimizzate dalla preferenza sistematica per le mancate rispetto alle false risposte e tramite la rimozione degli importi considerati altamente inaffidabili dagli intervistatori.

Rispetto all'uso esclusivo dei dati campionari, l'integrazione con i dati amministrativi ha aumentato sostanzialmente il numero di percettori e il reddito medio da lavoro autonomo. Fra gli individui per cui entrambe le fonti contengono redditi da lavoro indipendente, l'integrazione dei dati rivela che la sottostima è osservata più frequentemente nei dati tributari che nei dati di indagine.

6.2 - Metodologia per la stima degli affitti figurativi

Introduzione

Nell'ambito delle analisi sulla povertà, sulla disuguaglianza e sulla distribuzione dei redditi, la tematica dei fitti figurativi nasce dalla necessità di introdurre una nozione di reddito disponibile la più completa ed accurata possibile⁵. Secondo la letteratura, il fitto figurativo deve corrispondere al costo che una famiglia dovrebbe sostenere se avesse affittato ai prezzi vigenti sul mercato immobiliare un'unità abitativa equivalente, in termini di caratteristiche, a quella in cui effettivamente vive. Ciò vale tanto per i proprietari quanto per gli inquilini con affitti agevolati, inferiori ai prezzi di mercato.

Detto con il manuale di Canberra, che stabilisce a livello internazionale le linee guida per la definizione di reddito e ricchezza: “*In theory, imputed rent is the difference between the cost of*

⁵ Si ringraziano A.B. Atkinson, A. Brandolini, G. D'Alessio, A. Lemmi, A. Russo e N. Torelli per gli utili suggerimenti emersi dai nostri incontri.

renting one's living arrangements (in a competitive market) minus the cost actually incurred in owning the home (or renting it at a below market price)”⁶.

Si prenda in esame, a titolo esemplificativo, il caso di due famiglie identiche dal punto di vista delle entrate finanziarie e delle caratteristiche socio-demografiche ma che vivono l'una in una casa in affitto, l'altra in casa di proprietà. O ancora, il caso di due famiglie identiche che vivono entrambe in casa in affitto, con l'una che ha avuto accesso ad un affitto inferiore ai prezzi di mercato a differenza dell'altra. In questi casi portati ad esempio, la mancata considerazione della posta dei fitti figurativi porterebbe a considerarle uguali rispetto alla distribuzione dei redditi. Al fine di rendere comparabili le stime delle risorse familiari per i nuclei che stanno in affitto o in casa di proprietà è, quindi, necessario prendere in considerazione il flusso di servizi abitativi che si ottiene in virtù del proprio titolo di godimento.⁷

Già nel 1977 le Nazioni Unite raccomandavano l'inclusione dei fitti figurativi nelle statistiche sul reddito familiare; così pure l'Ilo⁸ ed il manuale di Canberra. Nel Sec95⁹ si fa riferimento a questa posta: come attività di produzione del settore delle famiglie, è compresa fra le produzioni di beni e servizi destinati all'autoconsumo; specularmente, la posta rientra nella spesa per consumi finali. Questo equivale ad ipotizzare che la famiglia stia, al contempo, producendo e consumando un flusso di servizi abitativi senza passare dal mercato. Seguendo questo orientamento, Eurostat ne raccomanda l'inserimento nei principali aggregati stimati tanto sulla base della rilevazione sui Consumi delle Famiglie (Hbs) quanto dell'indagine sul Reddito e le Condizioni di Vita (Eu-Silc). Allo stato attuale, comunque, Eurostat non include l'affitto figurativo nel reddito disponibile, ma questo verrà fatto solamente a partire dall'indagine 2007. Esistono, infatti, diverse metodologie per il calcolo dell'affitto figurativo, e si vuole dare tempo a quanti partecipano al progetto Eu-Silc di arrivare ad una soluzione condivisa che permetta la maggiore comparabilità possibile tra i diversi Paesi. Nel resto del capitolo verranno quindi proposti alcuni metodi per la stima di questa posta e verrà presentata, attraverso l'analisi di questi metodi, la proposta italiana per la metodologia da adottare.

6.2.1 Metodi di stima

6.2.1.1 User cost of capital

Le metodologie più largamente utilizzate, e coerenti con la teoria, sono quelle del *user cost of capital* e del *rental equivalence*.

Nel primo caso si tratta di stimare il tasso di rendimento implicito del capitale investito considerando il tasso di apprezzamento futuro sul mercato immobiliare. L'applicazione di questo metodo ha però degli aspetti controversi: innanzitutto, mostra un'estrema volatilità nel tempo, anche in funzione del tasso di interesse che viene usato per l'operazione di sconto. In secondo luogo, la stima del tasso di apprezzamento futuro del bene, ossia la predizione dell'andamento futuro del mercato immobiliare, pone non pochi problemi metodologici. Inoltre, la stessa Eurostat¹⁰ sconsiglia l'utilizzo di questo metodo:

“Although this method [user cost] may solve the problem which arises when the sub-sample of privately rented dwellings is too small, it has some important drawbacks for its implementation in the context of Eu-Silc and Hbs. The most important one is the macro-

⁶ Cfr. Canberra Group, op.cit.

⁷ È da citare un'ipotesi alternativa che consiste nel calcolare gli indicatori distributivi in base al reddito “spendibile”, al netto sia degli affitti realmente pagati dagli inquilini sia dei fitti imputati delle case abitate per tutti coloro che non pagano un affitto a prezzi di mercato. Questo metodo è attualmente in uso per le statistiche inglesi.

⁸ Cfr. Ilo, op. cit.

⁹ Sistema europeo dei conti 1995.

¹⁰ Si veda: Eurostat, Hbs and Eu-Silc imputed rent, Doc. Eu-Silc 155/05, 2005.

economic approach of some calculations which prevents from getting accurate imputations at micro-level".¹¹

6.2.1.2 *La stima del rental equivalence value*

Seguendo questo approccio, al fine di stimare il fitto imputato si deve valutare il flusso di servizi di un'unità abitativa il cui godimento non comporta transazioni di mercato, alla stessa stregua dei flussi corrispondenti osservabili sul mercato degli affitti.

Questo può avvenire tramite tecniche differenti: *i)* autovalutazione dei rispondenti (il cosiddetto affitto soggettivo); *ii)* metodo della stratificazione; *iii)* stima econometrica, in particolare tramite il metodo della regressione edonica, applicando o meno tecniche di correzione per la distorsione della selezione del campione (procedura di Heckman per la correzione del *sample selection bias*).

L'autovalutazione

Il metodo dell'autovalutazione consiste nella stima, da parte dei rispondenti ad indagine campionaria, del potenziale valore sul mercato degli affitti della casa nella quale vivono con titolo di godimento diverso dall'affitto a prezzi di mercato (proprietà, usufrutto, in uso gratuito o, infine, in affitto con canone inferiore ai prezzi di mercato). Questa tecnica è quella attualmente in uso in Istituto nell'indagine sui Consumi delle Famiglie.

Eurostat richiede agli stati membri di non usare l'affitto figurativo stimato tramite autovalutazione, principalmente per la natura intrinsecamente soggettiva della stima. I rispondenti potrebbero non conoscere in maniera adeguata le reali condizioni del mercato, la qual cosa può portare tanto ad una sottostima quanto ad una sovrastima dei "veri" affitti potenzialmente da pagare (ossia del "vero" flusso di servizi abitativi goduto).

La stratificazione

Con il metodo della stratificazione, dal punto di vista operativo, si tratterebbe di: *i)* definire gruppi di famiglie omogenei rispetto alle caratteristiche della casa e della zona di abitazione; *ii)* calcolare l'affitto medio/mediano per il sottoinsieme di ogni strato che vive in affitto a prezzi di mercato; *iii)* imputare questo valore a tutti coloro che, nello stesso gruppo, non pagano un affitto a prezzi di mercato.

Il metodo della stratificazione ha delle serie controindicazioni in situazioni, come ad esempio quella italiana, caratterizzate da un'alta incidenza di famiglie che vivono in case di proprietà ed un'ancora più alta percentuale di famiglie che, comunque, non vivono in case in affitto a prezzi di mercato. Difatti, il numero di caratteristiche rilevanti per la definizione degli strati porterebbe verosimilmente a frequenze di cella molto basse.¹²

Anche se la discussione tra i Paesi membri è ancora aperta, la proposta italiana consiste nel ricorrere a stime econometriche, tramite regressione edonica corretta con la procedura di Heckman per il *sample selection bias*, metodologia largamente coincidente con quella proposta da Eurostat ma con una fondamentale differenza, che vedremo in seguito, relativa agli effetti del periodo di occupazione della casa sull'affitto figurativo.

¹¹ Cfr. Eurostat 2005, op. cit. pag.10.

¹² La prescrizione di Eurostat è di non scendere sotto le 30 osservazioni per strato.

Regressione edonica

La regressione edonica¹³ si basa sull'assunzione - standard nella teoria micro-economica neoclassica- che il prezzo di un bene indivisibile, in presenza di mercati perfettamente concorrenziali, è dato dall'intersezione tra la domanda e l'offerta delle caratteristiche del bene. La domanda e l'offerta delle caratteristiche di un'abitazione non sono direttamente osservabili, ma si può considerare l'affitto come funzione di esse. In altri termini, il prezzo che emerge dal mercato degli affitti ha il contenuto informativo necessario per un'equa valutazione delle caratteristiche rilevanti dell'unità abitativa.

In pratica, il bene in questione è rappresentato dalle sue caratteristiche per le quali si assume che ognuna abbia un proprio mercato e che esista una sorta di effetto sostituzione tra una caratteristica e l'altra. Il modello può essere stimato tramite una regressione dei minimi quadrati ordinari (Ols) dove la variabile dipendente è il logaritmo dell'affitto di mercato e i parametri stimati per le caratteristiche del bene, che rappresentano l'elasticità al prezzo di ciascuno di essi, costituiscono il punto di incontro della domanda e offerta (non osservabili) di ciascuna caratteristica.

È importante notare che in questa maniera il prezzo dell'affitto è determinato solamente dalle caratteristiche del bene, e non dalle caratteristiche della famiglia in affitto: una famiglia benestante pagherà di più perché affitti una casa dotata di migliori caratteristiche, e non perché ha maggiori disponibilità. Inoltre, questa tecnica trascura l'uso dei prezzi come meccanismo di selezione della domanda, anche a parità di caratteristiche e di zona di ubicazione. Ossia, si esclude la possibilità che un locatore possa chiedere un prezzo superiore ai prezzi di mercato vigenti per una casa con determinate caratteristiche, anche a rischio di tenere la casa sfitta per qualche tempo, per discriminare la domanda potenziale, aumentando la probabilità di avere degli affittuari che soddisfano le sue aspettative.

C'è da constatare che l'utilizzo di questa metodologia risulta alquanto problematico nel caso in cui il mercato immobiliare risultasse segmentato e la distribuzione delle caratteristiche delle case in affitto fosse solo parzialmente sovrapponibile a quella delle case di proprietà; in questo caso le stime ottenute sulla base degli affitti realmente pagati potrebbe introdurre una distorsione quando usate per "predire" i fitti figurativi.

In ogni caso, tanto il metodo della stratificazione quanto la regressione edonica sono *information intensive*, ossia molto esigenti in termini di caratteristiche da rilevare.

Procedura di Heckman

La procedura a due stadi di Heckman¹⁴ permette di incorporare nel processo di stima tramite Ols la possibile distorsione da *sample selection*: l'equazione degli affitti è infatti stimata solamente su quanti sono in affitto ai prezzi di mercato, e questi potrebbero differire in maniera sistematica dai proprietari di immobili. La procedura consiste nella stima di un'equazione che stabilisce le determinanti dell'essere in affitto a prezzi di mercato o meno calcolando allo stesso tempo un parametro (λ) per ogni famiglia da inserire successivamente nell'equazione degli affitti come nuova variabile esplicativa; questo parametro cattura l'effetto di caratteristiche non osservabili che influenzano sia la decisione di stare in affitto sia il prezzo che si è disposti a pagare per un'abitazione con le caratteristiche di quella che si occupa.

¹³ Cfr. Rosen, 1974.

¹⁴ Cfr. Heckman 1979.

Il problema del tenure discount

Un ulteriore aspetto problematico, e trasversale rispetto alla scelta della tecnica più appropriata, è introdotto dal prendere in considerazione il *tenure discount*. Succede difatti che, anche a parità di caratteristiche abitative, i prezzi degli affitti reali possono essere anche significativamente diversi a seconda dell'anno in cui i contratti di locazione sono stati stipulati. In particolare, più lungo è il periodo di locazione di un'abitazione minore è il prezzo di affitto, a parità di altre condizioni. E ciò perché:

- i locatari scambiano il minore introito con la sicurezza del pagamento da parte di chi è affittuario da lungo tempo, come fosse una ricompensa implicita del rapporto fiduciario;
- i lavori di ristrutturazione vengono spesso fatti prima di affittare la casa, quindi al momento di ingresso la casa è in buone condizioni, ma queste peggiorano nel tempo.

La rilevanza del *tenure discount* è accentuata in un mercato come quello italiano che per lungo tempo è stato regolamentato. Malgrado la liberalizzazione delle tipologie contrattuali e la definitiva abrogazione, nel 1995, dell'equo canone per le nuove locazioni, risulta ancora una percentuale non trascurabile di famiglie la cui locazione è governata da questo tipo di contratto.

Il *tenure discount* è considerato da molte ricerche empiriche una caratteristica rilevante che influenza l'affitto pagato; tra gli altri, Miron¹⁵ sostiene che il *tenure discount* deve essere tenuto in considerazione nella stima degli affitti figurativi, poiché è una precisa condizione dell'abitazione che potrebbe modificare il prezzo del servizio abitativo ricevuto dalla famiglia. Questa è anche l'impostazione di Eurostat: il *tenure discount* deve essere considerato una determinante dell'affitto figurativo. Esso costituisce il secondo motivo per cui si richiede il non utilizzo del *self-assessment method* (stima soggettiva del *rental equivalence value*), che difficilmente potrebbe tenere conto di questo aspetto.

Nel manuale di Canberra questa condizione non è esplicitamente dichiarata. Il manuale recita infatti:

*“Finally, the service-yielding asset can be bought and sold on real estate markets. Hence, the value to the consumer is close to the market value of the service flow, since the owner could presumably sell the housing unit and rent it back from the new owner were it more profitable to do so. [...] imputed rent can therefore be valued to consumers at its market price”*¹⁶

Come si vede, vi è un'ambiguità sul fatto che il prezzo di mercato debba incorporare lo sconto temporale o se come prezzo di mercato si intenda quello che si andrebbe ad affrontare se si entrasse in quel preciso momento sul mercato degli affitti.

Una considerazione aggiuntiva riguarda il fatto che includendo il *tenure discount* tra le determinanti degli affitti figurativi viene alterata la comparabilità dei livelli di benessere: se due famiglie vivono in abitazioni identiche, quella che è proprietaria da più tempo risulterebbe meno ricca in virtù dell'effetto del maggior periodo di occupazione, anche se nei fatti questa differenza non avrebbe ragione di esistere. Infine, l'utilizzo di uno sconto temporale per i proprietari può portare a difficoltà di comparazione internazionale, nel caso in cui le distribuzioni del periodo di occupazione delle abitazioni da parte dei proprietari fossero sensibilmente differenti fra due o più Paesi. Se, ad esempio, le famiglie italiane avessero un periodo di occupazione mediamente più lungo rispetto a quelle di un altro Paese risulterebbero, a parità di altre condizioni, meno ricche rispetto a quelle del Paese di confronto.

Sulla questione dello sconto temporale torneremo diffusamente più avanti.

¹⁵ Cfr. Miron 1990.

¹⁶ Canberra Group, op. cit. pag. 64.

6.2.2 Applicazioni: principali risultati dall'indagine Eu-Silc

Una volta esaurita la discussione sul quadro teorico di riferimento, possono essere introdotti i risultati della prima fase della sperimentazione condotta sui dati del primo anno dell'indagine Eu-Silc 2004 per l'Italia. Le informazioni raccolte riguardano 24.204 famiglie. Il gruppo sul quale effettuare le stime soggettive e Ols è determinato in base alle domande sul titolo di godimento dell'abitazione e, tra quanti sono in affitto, al fatto se pagano o meno un prezzo inferiore ai prezzi di mercato (domande 1.34 e 2.5).¹⁷ Per esclusione, risultano come affittuari ai prezzi di mercato 2.940 famiglie.

Al fine di condurre la sperimentazione sui fitti figurativi sono state implementate le metodiche dell'approccio del *rental equivalence value*, con l'esclusione del metodo della stratificazione. Si è già fatto notare come questa tecnica abbia delle serie controindicazioni in situazioni come quella italiana. Stando ai dati Eu-Silc, una quota di famiglie residenti inferiore al 15 per cento dichiara di stare in affitto a prezzi di mercato; la numerosità è troppo bassa per evitare efficacemente il *trade-off* fra frequenze di cella non significative per alcune strati ed un livello di disaggregazione troppo approssimativo per essere realmente informativo.

Nel prossimo paragrafo il livello dei fitti soggettivi sarà raffrontato con le stime derivanti dall'applicazione della regressione edonica, senza e con la procedura di Heckman. Le stime ottenute tramite autovalutazione, lungi dall'essere una soluzione ottimale, sono comunque presentate perché è il metodo attualmente in uso in Istat; secondariamente, non si esclude che su questo metodo possa comunque cadere la scelta dei Paesi membri, poiché è quello che si presta a minori differenze dovute alle diverse realtà nazionali, con ricadute positive sulla comparabilità dei dati; in più, essendo l'unico metodo senza implicazioni "pratiche" se non la semplice domanda all'intervistato, è sicuramente il più semplice da utilizzare.

Prima di arrivare a questo confronto, presentiamo la formalizzazione dei modelli utilizzati.

Il modello della regressione edonica può essere scritto come:

$$\ln y_i = \beta_0 + \beta_{1k} \mathbf{X}_{ki}^a + \beta_2 T_i^a + u_i \quad i \text{ varia da } 1 \text{ a } n_i \quad (1)$$

dove y_i è l'affitto pagato, \mathbf{X}_{ki} è un vettore di caratteristiche della casa e della zona dove è situata l'abitazione, T_i è il periodo di occupazione della casa e u_i è il termine di errore. Il carattere a in apice sta ad indicare che il modello è applicato agli affittuari ai prezzi di mercato, di numerosità pari ad n_i . I parametri β , stimati sugli affittuari ai prezzi di mercato, sono successivamente moltiplicati alle caratteristiche del sottocampione di numerosità n_j dei proprietari¹⁸ per ottenere il corrispondente valore dell'affitto figurativo.

$$Y_{fig,j} = \exp(\beta_0 + \beta_{1k} \mathbf{X}_{kj}^p + \beta_2 T_i^p) \quad j \text{ varia da } 1 \text{ a } n_j \quad (2)$$

Secondo Eurostat, il modello descritto nell'equazione (1) potrebbe soffrire di un problema di distorsione da *sample selection*: l'affitto è infatti osservato solamente tra quanti pagano un affitto ai prezzi di mercato e costoro potrebbero differire in maniera sistematica e non osservabile dal resto del campione. Occorre quindi risolvere il problema della mancanza di conoscenza dei fattori che determinano la condizione di affittuario. La procedura a due passi di Heckman¹⁹ incorpora nel processo di stima la possibile esistenza di questa *selection bias*. In

¹⁷ La domanda 2.5 recita: "La sua famiglia paga un affitto inferiore al prezzo di mercato?". La domanda così posta può comunque portare ad una distorsione dovuta al fatto che la selezione è fatta in base alla valutazione soggettiva dell'intervistato.

¹⁸ D'ora in avanti indicheremo come proprietari quanti non sono in affitto ai prezzi di mercato, indipendentemente dal reale titolo di godimento dell'abitazione.

¹⁹ Cfr. Heckman, op. cit.

questo contesto, la procedura consiste nello stima di un modello *probit* con la condizione di affittuario come variabile risposta ed una serie di caratteristiche familiari come esplicative. Il coefficiente trovato tramite la *probit* per l'*inverse Mill's ratio* viene successivamente inserito nella stima dei minimi quadrati ordinari sugli affitti come variabile esplicativa.

Sia R_i^* la variabile latente che indica l'utilità indiretta di stare in affitto per la famiglia i .

$$R_i^* = \alpha_0 + \alpha_{1k} Z_{ki} + e_i \quad (3)$$

dove Z_i è un vettore di variabili esplicative, α è il vettore di parametri da stimare e e_i è il termine casuale. Poiché R^* non è osservabile, la variabile dipendente R assume valore 1 (la famiglia è in affitto ai prezzi di mercato) se $R^* \geq 0$ e assume valore 0 (la famiglia non è in affitto ai prezzi di mercato) se $R^* < 0$. Il valore di Y è osservabile solamente per quelle famiglie per le quali $R^* \geq 0$ e, quindi, il valore atteso di Y nell'equazione (1) è il valore atteso di Y condizionatamente al fatto che sia osservato ($R=1$). L'*inverse Mill's Ratio* (chiamato λ), stimato dalla *probit* dell'equazione (3), viene successivamente inserito nell'equazione Ols, che quindi diventa:

$$\ln y_i = \beta_0 + \beta_{1k} X_{ki} + \beta_2 T_i + \beta_3 \lambda_i + u_i \quad (4)$$

La procedura di Heckman richiede che Z_i includa almeno una variabile che influenzi la probabilità di essere in affitto ma che possa essere legittimamente esclusa da X_i nell'equazione (4). Nel contesto della regressione edonica, con il costo dell'affitto determinato dalle sole caratteristiche dell'abitazione, possiamo utilizzare le caratteristiche familiari come variabili strumentali che influenzino la condizione di affittuario o meno.

La condizione di affittuario

Come variabili esplicative sono state prese in considerazione informazioni sulla famiglia, sull'intestatario del foglio di famiglia e sulla zona di abitazione. Come variabili relative alla famiglia sono state utilizzate il numero di componenti, il numero di percettori di reddito e il quinto di reddito equivalente di appartenenza²⁰; come variabili relative alla persona di riferimento della famiglia sono considerate l'età, il titolo di studio, la condizione professionale, lo stato civile e la cittadinanza; come variabili relative alla zona di abitazione sono state utilizzate la regione e il tipo di comune di residenza (è possibile vedere le modalità delle variabili direttamente dal Prospetto 6.1).

Vediamo brevemente i risultati della *probit* che descrive la condizione abitativa delle famiglie. Più è alto il numero di componenti in famiglia, maggiore la probabilità di stare in affitto, mentre maggiore è il numero di percettori e minore è la probabilità di stare in affitto. Il quinto di reddito di appartenenza non sembra invece avere influenza sulla condizione abitativa, in parte perché l'effetto è colto dal numero di percettori di reddito, in parte perché avrebbe avuto maggiore significato una variabile sulla ricchezza posseduta dalla famiglia piuttosto che sul reddito dell'ultimo anno. Questo è confermato anche dal fatto che la probabilità di essere in affitto decresce al crescere dell'età della persona di riferimento, proprio perché la ricchezza accumulata tende ad aumentare nel corso del ciclo di vita delle persone.

²⁰ Il reddito equivalente è stato calcolato secondo la scala Ocese tramite la formula $\text{Reddito equivalente} = \text{Reddito disponibile} * (1 + 0,5 * (\text{numero di adulti} - 1) + 0,3 * \text{numero di bambini})$, dove i bambini sono quanti hanno un'età inferiore ai 14 anni.

Osservando i coefficienti delle altre variabili relative alla persona di riferimento, l'essere cittadino straniero esercita un'influenza molto forte sulle probabilità di stare in affitto, così come l'essere coniugato o convivente ha una forte influenza sulla probabilità di essere in casa di proprietà. Quanti si trovano nella condizione di operaio o di lavoratore autonomo (ad eccezione di imprenditori e liberi professionisti) presentano una maggiore probabilità degli altri di essere in affitto, così come accade per quanti hanno un basso titolo di studio.

Molto importante risulta il comune di domicilio della famiglia, con una probabilità di essere affittuari crescente al crescere dell'ampiezza demografica e del grado di urbanizzazione, così come la regione di residenza: Piemonte, Valle d'Aosta e Liguria risultano, a parità di altre condizioni, le regioni con maggiore probabilità di essere in affitto ai prezzi di mercato, mentre Toscana, Lazio e Sardegna sono quelle con le minori probabilità.

Prospetto 6.1 - Le determinanti della condizione di affittuario

VARIABILE	Parametro	Errore Standard
Costante	-1,1178*	0,1017
Numero componenti	0,0872*	0,0132
Numero percettori	-0,2216*	0,0207
Quinto di reddito equivalente (Rif: Primo quinto)		
<i>Secondo quinto</i>	0,0296	0,0341
<i>Terzo quinto</i>	0,0484	0,0361
<i>Quarto quinto</i>	0,0039	0,0390
<i>Quinto quinto</i>	-0,0355	0,0432
Età del p.r.	-0,0093*	0,0011
Straniero	1,0751*	0,0479
Stato civile del p.r. (Rif: coniugato/convivente)		
<i>Celibe / nubile</i>	0,3199*	0,0378
<i>Separato / divorziato</i>	0,4489*	0,0388
<i>Vedovo / a</i>	0,0944*	0,0399
Posizione nella professione del p.r. (Rif: dirigenti, quadri, impiegati)		
<i>Operaio o assimilato</i>	0,2437*	0,0398
<i>Imprenditori o liberi professionisti</i>	0,0435	0,0607
<i>Altro autonomo</i>	0,1671*	0,0433
<i>Disoccupato</i>	0,1216	0,0783
<i>Pensionato da lavoro</i>	0,0346	0,0461
<i>Altro</i>	0,1466*	0,0488
Titolo di studio (Rif: Nessuno / elementari)		
<i>Medie inferiori</i>	-0,0808*	0,0319
<i>Medie superiori</i>	-0,2491*	0,0351
<i>Universitario</i>	-0,3848*	0,0519

*Significativo al 5 per cento. **Significativo al 10 per cento.

Prospetto 6.1 segue - Le determinanti della condizione di affittuario

VARIABILE	Parametro	Errore Standard
Tipo di comune (Rif. Comune con meno di 2.000 abitanti)		
<i>Centro di area metropolitana</i>	0,8099*	0,0590
<i>Periferia di area metropolitana</i>	0,5615*	0,0628
<i>Abitanti in (2.001-10.000)</i>	0,3135*	0,0570
<i>Abitanti in (10.001-50.000)</i>	0,4936*	0,0565
<i>Più di 50.000 abitanti</i>	0,7404*	0,0589
Regione (Rif: Lombardia)		
<i>Piemonte</i>	0,2218*	0,0438
<i>Valle d'Aosta/Vallé d'Aoste</i>	0,3785**	0,2068
<i>Trentino-Alto Adige</i>	0,1288	0,0885
<i>Veneto</i>	-0,1704*	0,0494
<i>Friuli-Venezia Giulia</i>	-0,1483**	0,0830
<i>Liguria</i>	0,1492*	0,0610
<i>Emilia Romagna</i>	-0,1321*	0,0490
<i>Toscana</i>	-0,3179*	0,0558
<i>Umbria</i>	-0,1995*	0,1013
<i>Marche</i>	-0,0809	0,0785
<i>Lazio</i>	-0,2815*	0,0474
<i>Abruzzo</i>	-0,1105	0,0853
<i>Molise</i>	-0,2799	0,1917
<i>Campania</i>	0,0938*	0,0443
<i>Puglia</i>	-0,0546	0,0521
<i>Basilicata</i>	-0,0347	0,1197
<i>Calabria</i>	-0,1103	0,0700
<i>Sicilia</i>	-0,1589*	0,0484
<i>Sardegna</i>	-0,2288*	0,0785
Log-verosimiglianza	-8.179,61	
Numerosità del campione	24.204	

*Significativo al 5 per cento. **Significativo al 10 per cento.

Fonte: Indagine sulle condizioni di vita anno 2004

L'equazione degli affitti

Come variabili esplicative nell'Ols sugli affitti sono state utilizzate le caratteristiche dell'abitazione, le caratteristiche della zona di abitazione e il periodo di occupazione dell'abitazione.

Relativamente a quest'ultimo aspetto, come già osservato, Eurostat suggerisce di considerare lo sconto temporale per quanti sono in case di proprietà, anche se questa raccomandazione non è esplicitamente contenuta nel manuale di Canberra. Nei termini utilizzati in precedenza, la considerazione o meno del *tenure discount* per i proprietari tira in ballo la questione di quale debba essere il prezzo di mercato di riferimento per la valutazione dei flussi di servizi abitativi.

Quale che sia la posizione nei confronti dei proprietari, quello del *tenure discount* (o *seniority*) è un effetto realmente esistente sugli affittuari e, di conseguenza, il periodo di occupazione della casa va considerato tra le variabili esplicative della Ols. Diverso è il suo utilizzo a seconda che si decida o meno di sterilizzarne l'effetto sui proprietari. Nel Prospetto

6.3, le seconde due colonne riportano le stime ottenute considerando il periodo di occupazione come variabile continua; in questo caso nell'Ols è stato inserito anche il periodo di occupazione al quadrato, al fine di cogliere eventuali effetti non lineari sul livello dell'affitto. Successivamente, nel calcolo dell'affitto figurativo come in equazione (2) si impone che il periodo di occupazione per i proprietari sia pari a zero di modo da "sterilizzare" questo sconto temporale. Nel caso, invece, si voglia considerare l'esistenza di questo sconto anche per i proprietari, non è possibile utilizzare la *seniority* in modalità continua. Questo perché le distribuzioni del periodo di occupazione per le due sub-popolazioni sono molto differenti (Prospetto 6.2): ad esempio, la media e la mediana degli anni di occupazione della casa sono pari, rispettivamente, a 21 e 18 per i proprietari e 11 e 6 per gli affittuari. In questa maniera, si rischierebbe di assegnare alle famiglie proprietarie i valori stimati sulla base di una distribuzione significativamente diversa rispetto alla variabile "periodo di occupazione".

Prospetto 6.2 - Principali momenti della distribuzione del numero di anni di occupazione della casa

ANNI	Media	Mediana	I quintile	II quintile	III quintile	IV quintile
Proprietà	21	18	5	14	23	34
Affitto	11	6	2	4	9	20

Fonte: Indagine sulle condizioni di vita anno 2004

Per evitare questo effetto, le due sub-popolazioni sono state divise in decimi in base ai decili calcolati tramite il periodo di occupazione della casa e nel modello sono state inserite diverse variabili *dummy* ad indicare il decimo di appartenenza nella distribuzione della *seniority*: l'ipotesi è che se i proprietari fossero stati in affitto avrebbero avuto lo stesso *turn-over* delle abitazioni che si registra per gli affittuari.

Riassumendo, nel prospetto 6.3 la seconda e la quarta colonna riportano le stime corrette tramite selezione; la prima e la terza colonna presentano le stime Ols semplici. Inoltre, le prime due colonne considerano le *dummy* relative del periodo di occupazione dell'abitazione, laddove le seconde due presentano le stime ottenute considerando la *seniority* come variabile tempo di tipo continuo.

Prospetto 6.3 - Le stime Ols

	Distribuzione della <i>seniority</i> secondo il decimo di appartenenza		<i>Seniority</i> come variabile continua	
	Senza Heckman	Con Heckman	Senza Heckman	Con Heckman
Intercetta	4,13316* (0,13297)	4,04011* (0,13897)	4,70860* (0,13126)	4,65262* (0,13352)
<i>Villa, villino unifamiliare</i>	-0,04866 (0,03717)	-0,04939 (0,03715)	-0,04579 (0,03714)	-0,04614 (0,03711)
<i>Villa, villino plurifamiliare</i>	-0,10965* (0,03341)	-0,10518* (0,03344)	-0,10502* (0,03338)	-0,10050* (0,03341)
<i>App. in edificio con < 10 app.</i>	-0,04227** (0,02484)	-0,04192** (0,02482)	-0,03831 (0,02478)	-0,03759 (0,02476)
<i>Altro tipo</i>	-0,0485 (0,06453)	-0,05135 (0,0645)	-0,04162 (0,06449)	-0,04389 (0,06445)
MQ	0,00824* (0,00074)	0,00797* (0,00075)	0,00827* (0,00074)	0,00801* (0,00075)
MQ^2	-0,00001* (0)	-0,00001* (0)	-0,00001* (0)	-0,00001*(0)
Vasca o doccia	-0,06247 (0,07974)	-0,0685 (0,07973)	-0,07005 (0,0798)	-0,07633 (0,07979)
Bagno interno	0,25087* (0,10203)	0,24891* (0,10196)	0,25962* (0,10205)	0,25779* (0,10198)
Acqua calda	0,15003** (0,07834)	0,14113** (0,07837)	0,15330** (0,07829)	0,14447** (0,07834)
Terrazza o balcone	0,08924* (0,02451)	0,08504 (0,02457)	0,09009* (0,02447)	0,08571* (0,02453)
Giardino	0,04256 (0,03164)	0,04054 (0,03163)	0,04215 (0,03161)	0,03977 (0,03161)
Parti danneggiate	-0,14030* (0,02893)	-0,13784* (0,02893)	-0,14293* (0,02892)	-0,14069* (0,02892)
Umidità in muri, tetti	0,01012 (0,02609)	0,01343 (0,02611)	0,01388 (0,02604)	0,01722 (0,02607)
Scarsa luminosità	-0,06837* (0,02886)	-0,06904* (0,02884)	-0,06665* (0,02885)	-0,06732* (0,02883)
Inquinamento	0,00409 (0,02567)	0,00363 (0,02565)	0,00492 (0,02564)	0,00417 (0,02562)
Inquinamento acustico	0,03257 (0,02371)	0,03219 (0,02369)	0,03184 (0,02366)	0,03164 (0,02365)
Criminalità in zona	-0,10162* (0,02776)	-0,09869* (0,02777)	-0,10095* (0,02775)	-0,09810* (0,02776)
<i>Centro di area metropolitana</i>	0,65859* (0,05967)	0,69462* (0,06167)	0,65628* (0,05963)	0,69131* (0,0616)
<i>Periferia di area metropolitana</i>	0,57280* (0,06112)	0,59581* (0,0619)	0,56745* (0,06105)	0,58967* (0,0618)
<i>Abitanti in (2.001-10.000)</i>	0,15359* (0,05593)	0,16655* (0,05618)	0,14916* (0,05589)	0,16146* (0,05612)
<i>Abitanti in (10.001-50.000)</i>	0,37907* (0,05566)	0,39723* (0,05618)	0,37544* (0,05563)	0,39284* (0,05613)
<i>Più di 50.000 abitanti</i>	0,45659* (0,05765)	0,48670* (0,05909)	0,45188* (0,05763)	0,48059* (0,05899)
<i>Piemonte</i>	-0,10992* (0,03621)	-0,09988* (0,03645)	-0,11223* (0,03621)	-0,10268* (0,03644)
<i>Valle d'Aosta/Vallée d'Aoste</i>	0,08232 (0,17317)	0,09282 (0,1731)	0,07936 (0,17319)	0,08907 (0,17313)
<i>Trentino-Alto Adige</i>	0,16266* (0,07545)	0,17060* (0,07547)	0,16340* (0,07545)	0,17127* (0,07548)
<i>Veneto</i>	0,11380* (0,04512)	0,11220* (0,04509)	0,10766* (0,04498)	0,10650* (0,04495)
<i>Friuli-Venezia Giulia</i>	-0,09466 (0,07579)	-0,0974 (0,07574)	-0,09307 (0,07578)	-0,09617 (0,07574)
<i>Liguria</i>	-0,07037 (0,04983)	-0,06332 (0,04989)	-0,0709 (0,04978)	-0,06409 (0,04984)
<i>Emilia Romagna</i>	0,08824* (0,04348)	0,08487* (0,04348)	0,08404** (0,04344)	0,08105** (0,04343)
<i>Toscana</i>	0,04035 (0,05238)	0,03013 (0,05253)	0,04092 (0,05239)	0,03101 (0,05254)
<i>Umbria</i>	-0,15762* (0,08972)	-0,16103* (0,08966)	-0,17001* (0,08971)	-0,17356* (0,08966)
<i>Marche</i>	-0,17375* (0,07053)	-0,17074* (0,07049)	-0,17799* (0,07054)	-0,17521* (0,0705)
<i>Lazio</i>	-0,18022* (0,04176)	-0,18722* (0,04185)	-0,17982* (0,04171)	-0,18719* (0,04181)
<i>Abruzzo</i>	-0,40879* (0,07892)	-0,41012* (0,07886)	-0,41481* (0,07883)	-0,41706* (0,07878)
<i>Molise</i>	-0,47334* (0,19014)	-0,47671* (0,19)	-0,47943* (0,19014)	-0,48299* (0,19002)
<i>Campania</i>	-0,24761* (0,03804)	-0,24189* (0,0381)	-0,24581* (0,03794)	-0,24050* (0,03799)
<i>Puglia</i>	-0,49133* (0,04704)	-0,49413* (0,04702)	-0,49237* (0,04684)	-0,49581* (0,04684)
<i>Basilicata</i>	-0,58331* (0,11408)	-0,59021* (0,11404)	-0,57762* (0,11403)	-0,58506* (0,114)
<i>Calabria</i>	-0,71321* (0,06243)	-0,71120* (0,06239)	-0,70977* (0,06215)	-0,70793* (0,06211)
<i>Sicilia</i>	-0,72198* (0,04283)	-0,72364* (0,0428)	-0,73002* (0,04278)	-0,73228* (0,04276)
<i>Sardegna</i>	-0,48563* (0,07456)	-0,49226* (0,07456)	-0,48072* (0,0745)	-0,48705* (0,07451)
<i>1° decimo</i>	0,54733* (0,04026)	0,58373* (0,04326)	--	--
<i>2° decimo</i>	0,53257* (0,04372)	0,56508* (0,04595)	--	--
<i>3° decimo</i>	0,54326* (0,04521)	0,57382* (0,04711)	--	--
<i>4° decimo</i>	0,55815* (0,04543)	0,58613* (0,04702)	--	--
<i>5° decimo</i>	0,46441* (0,04307)	0,48694* (0,04415)	--	--
<i>6° decimo</i>	0,39037* (0,04638)	0,40780* (0,04696)	--	--
<i>7° decimo</i>	0,36672* (0,04227)	0,38574* (0,04305)	--	--
<i>8° decimo</i>	0,27942* (0,04488)	0,28870* (0,04503)	--	--
<i>9° decimo</i>	0,18161* (0,04389)	0,18705* (0,04392)	--	--
<i>Seniority</i>	--	--	-0,02065* (0,00193)	-0,1326 (0,00204)
<i>Seniority^2</i>	--	--	0,00016* (0,00004)	0,00018* (0,00004)
<i>Lambda</i>	--	0,05861* (0,02564)	--	0,05706* (0,02544)
n° oss	2.940	2.940	2.940	2.940
R ²	0,388	0,389	0,386	0,387

*Significativo al 5 per cento. **Significativo al 10 per cento. In parentesi gli errori standard.

Fonte: Indagine sulle condizioni di vita anno 2004

L'effetto delle variabili esplicative è assolutamente analogo nei differenti modelli; quanto andremo quindi a commentare è valido per ciascuno dei quattro modelli presentati. Si vede anzitutto come il parametro di selezione, λ , sia statisticamente significativo. Questo implica l'esistenza di un processo di selezione potenzialmente distorto all'interno dei dati; questo fatto va comunque analizzato in maniera più approfondita, e ci torneremo nel paragrafo successivo. Relativamente al tipo di abitazione nella quale si abita, si osserva come solamente l'abitare in una villa plurifamiliare si differenzi rispetto agli altri tipi di casa, con impatto negativo sul canone di locazione; è comunque da considerare che gran parte dell'effetto del tipo di abitazione covaria con la tipologia di comune di residenza. Rispetto alle dotazioni della casa, l'aver un terrazzo, così come la disponibilità di acqua calda e di bagno interno all'abitazione, aumenta considerevolmente il valore della casa e, quindi, del canone di locazione; viceversa, avere parti della casa danneggiate o infiltrazioni di umidità diminuisce questo valore. Come atteso, le metrature dell'abitazione hanno un effetto positivo ma non linearmente crescente sul prezzo dell'affitto.

Relativamente alle caratteristiche della zona di abitazione, la presenza di criminalità fa diminuire fortemente il valore degli affitti, mentre non sembrano avere effetto la presenza di inquinamento atmosferico ed acustico, tipica nelle grandi città, il cui effetto è colto dalla variabile relativa al tipo di comune. Proprio rispetto a quest'ultima, si vede come al crescere della dimensione demografica del comune di residenza ed all'avvicinarsi ad un'area metropolitana il livello del canone di locazione aumenta in maniera considerevole.

Rispetto alla regione di residenza, si osservano prezzi in forte decremento passando dal Nord al Mezzogiorno, con Trentino-Alto Adige, Friuli-Venezia Giulia ed Emilia Romagna che risultano le regioni più costose e Basilicata, Calabria e Sicilia le meno dispendiose.

Quanto all'esistenza di un effetto da *tenure discount*, dalla regressione sugli affittuari emerge chiaramente l'esistenza di questo sconto che ha un impatto negativo, anche se non linearmente decrescente, sull'affitto da pagare.

I risultati descritti sono stati ottenuti da modelli che presentano un buon grado di adattamento: l' R^2 presenta infatti un valore alto per un'applicazione a dati trasversali, pari all'incirca a 0,39 per ciascuno dei modelli.

6.2.3 Commenti ai risultati e approfondimenti

Una volta applicati i parametri alle caratteristiche delle abitazioni di proprietà (o in usufrutto, uso gratuito o in affitto con prezzo inferiore ai prezzi di mercato) come descritto nell'equazione (2), si osserva (Prospetto 6.4) come l'utilizzo dei diversi metodi porta a risultati differenti relativamente al valore sul mercato degli affitti delle abitazioni di proprietà. Prendendo come riferimento l'affitto soggettivo medio a livello nazionale, pari ad una stima di 530 euro, si vede che ciascuna delle quattro stime econometriche è inferiore a quella risultante dall'autovalutazione. Considerando al contempo il processo di selezione nelle due sub-popolazioni e lo sconto temporale anche per i proprietari, le stime si abbassano considerevolmente: sul territorio nazionale, la stima ottenuta tenendo conto di entrambi gli effetti risulta in media pari a 325 euro, del 38,7 per cento inferiore rispetto alla stima soggettiva; è inferiore del 19,6 per cento quando non si tiene conto di nessuno dei due. Questi andamenti sono confermati anche a livello ripartizionale.

Prospetto 6.4 - Le stime dei fitti figurativi secondo le varie tecniche del tipo rental equivalence value. Base: autovalutazione Italia=100.

FITTI FIGURATIVI	Italia	Nord-ovest	Nord-est	Centro	Sud	Isole
Autovalutazione	530	595	627	633	365	323
Modello che considera lo sconto temporale						
Ols semplice	360	391	457	397	266	219
Ols con procedura di Heckman	325	354	413	359	241	198
Modello che <i>sterilizza</i> lo sconto temporale						
Ols semplice	426	461	540	470	317	259
Ols con procedura di Heckman	390	422	495	431	291	237

Fonte: Indagine sulle condizioni di vita anno 2004

Si pone quindi il problema su quale procedura sia la più corretta: relativamente all'esistenza di uno sconto temporale, è da notare anzitutto come il periodo di occupazione debba essere incorporato nel processo di calcolo dei coefficienti della regressione con il logaritmo dell'affitto come variabile dipendente: questo sconto esiste ed è anche una delle determinanti più importanti del livello del canone di locazione. Non considerarlo porterebbe peraltro a risultati distorti relativamente alle altre esplicative utilizzate.

Per capire se invece debba essere considerato o "sterilizzato" nel calcolo dei fitti figurativi, ricorriamo nuovamente al manuale di Canberra:

"Hence, the value to the consumer is close to the market value of the service flow, since the owner could presumably sell the housing unit and rent it back from the new owner were it more profitable to do so".

La parte che abbiamo evidenziato in grassetto sembra indicare la strada: il proprietario vende la casa e la riaffitta nello stesso momento dal nuovo proprietario in un mercato i cui prezzi sono fissati sulla base delle condizioni di mercato vigenti. In altre parole, il contratto di locazione scatterebbe al momento dell'ipotetico ingresso nella casa nella nuova condizione di affittuario, al momento quindi dell'intervista. Se questa interpretazione fosse corretta, il periodo di occupazione della casa non andrebbe incorporato nella stima dell'affitto figurativo per i proprietari; il periodo di occupazione dovrebbe, quindi, essere posto uguale a 0.

Relativamente all'utilizzo della procedura a due stadi di Heckman, è da osservare come il parametro di selezione dovrebbe tenere conto di variabili non osservabili che influenzano sia la decisione di stare in affitto sia il prezzo che si è disposti a pagare per quella abitazione. Interpretando il risultato in maniera analoga alla stima del salario di riserva (ambito in cui la procedura è stata massicciamente utilizzata), l'affitto figurativo si può interpretare come una sorta di affitto di riserva: è il prezzo massimo che il proprietario sarebbe disposto a pagare per cambiare la propria condizione, ossia per accettare di diventare inquilino nella sua stessa casa. Nel diventare inquilino, il proprietario accetta di avere meno sicurezza e, quindi, chiede ragionevolmente un affitto inferiore rispetto al prezzo di mercato. In altre parole, mentre un inquilino può essere disposto ad accettare un determinato affitto se non ha alternative, per indurre un proprietario a cambiare status, cioè a rinunciare alla sicurezza della proprietà di un bene rifugio come è la casa, anche se potrebbe investire i soldi della vendita e ricavarne un

guadagno, è necessario offrire un canone di locazione minore.²¹ La procedura applicata ai nostri dati fornisce un valore dell'affitto figurativo all'incirca del 9 per cento inferiore rispetto alla Ols semplice.

Nella citazione precedente dal manuale di Canberra abbiamo evidenziato, questa volta in corsivo, anche un'altra parte. La frase sembra suggerire l'incorporazione del processo di selezione nel calcolo delle stime, poiché prende proprio in considerazione il fatto che debba esistere un vantaggio economico per convincere un proprietario a cambiare status.

Quanto appena detto, insieme ai diversi commenti presentati nel corso del paragrafo, porta a formalizzare la proposta italiana, che consiste nel ricorrere a stime econometriche, tramite regressione edonica corretta con la procedura di Heckman per il *sample selection bias*, sterilizzando i risultati ottenuti per i proprietari dallo sconto temporale che si ottiene in base al periodo di occupazione della casa. La metodologia coincide largamente con quella proposta da Eurostat, ma con la fondamentale differenza relativa agli effetti del periodo di occupazione della casa sull'affitto figurativo.

Nel Prospetto 6.5 vengono rappresentati i fitti soggettivi e le stime econometriche senza considerare lo sconto temporale sui proprietari. La regressione edonica è considerata invece sia con che senza la selezione di Heckman, e questo perché, ferma restando la nostra interpretazione, questo ultimo aspetto ci pare sicuramente più dubitabile rispetto alla sterilizzazione dello sconto temporale.

I valori sono stati trasformati in numeri indice per rendere visivamente più immediate le differenze riscontrate. Decomponendo per tipo di comune, si nota come i livelli imputati tramite tecniche econometriche mostrano una variabilità più contenuta rispetto agli affitti soggettivi. Ponendo come base l'autovalutazione distinta per tipo di comune (parte bassa del Prospetto 6.5), si nota che gli scostamenti relativi più ampi si registrano per i centri delle aree metropolitane.

Non è semplice dare conto della differenza fra fitti soggettivi e tecniche econometriche. In linea di principio la differenza può essere tanto dovuta ad una sovrastima da parte dei rispondenti nella valutazione soggettiva quanto ad una sottostima dovuta alle metodiche econometriche.

Per tentare di fare luce sulla questione abbiamo condotto un ulteriore esercizio comparativo. Abbiamo stimato, sulla popolazione delle famiglie con casa di proprietà, il modello di regressione edonica prendendo come variabile dipendente il logaritmo dei fitti soggettivi; il set di regressori è identico a quello utilizzato in precedenza (cfr. Prospetto 6.3, 1^a colonna). La comparazione viene effettuata sui parametri stimati dalle funzioni edoniche aventi come dipendente rispettivamente i fitti reali e quelli soggettivi: se i parametri fossero simili, nel segno, nel peso e nella significatività statistica, si potrebbe legittimamente concludere in favore dell'attendibilità della valutazione soggettiva. Nel caso contrario, che è quello che abbiamo riscontrato²² con i dati Eu-Silc, le spiegazioni possono essere diverse: in un verso si può concludere per l'errata valutazione del flusso dei servizi abitativi tramite fitti soggettivi, date le caratteristiche inserite fra i regressori; d'altro canto si può, altrettanto verosimilmente, pensare che la differenza stia nelle diverse distribuzioni delle caratteristiche delle abitazioni fra le due sub-popolazioni. In altri termini, il mercato immobiliare potrebbe essere segmentato: un'abitazione, data la propria dotazione di caratteristiche, non avrebbe la stessa probabilità di finire sul mercato delle locazioni e su quello delle compravendite.

²¹ Si noti comunque che un ragionamento del genere è valido solamente per quanti sono proprietari di casa, e si troverebbero nella condizione di potere scegliere tra proprietà e affitto, mentre non è valido per quanto sono in condizioni di usufrutto, uso gratuito o affitto inferiore ai prezzi di mercato. Per questi ultimi la procedura di Heckman non andrebbe probabilmente considerata. In questa fase dell'analisi non sembra comunque il caso di complicare ulteriormente lo studio.

²² I risultati non sono presentati per non appesantire troppo l'analisi.

Resta da dare una sommaria valutazione su quale effetto abbia la scelta della tecnica sulle misure di disuguaglianza. A tal fine, abbiamo selezionato due fra i cosiddetti indicatori di Laeken, cioè gli indicatori che Eurostat utilizza per fornire un quadro delle diverse realtà nazionali: il rapporto di concentrazione di Gini e la percentuale di famiglie sotto la linea di povertà, identificata come il 60 per cento della mediana della distribuzione del reddito familiare equivalente.

Prospetto 6.5 - Le stime degli affitti figurativi senza sconto temporale

TITOLO DI GODIMENTO DELL'ABITAZIONE	Metodo di stima		
	Autovalutazione	Regressione edonica	Edonica con Heckman
BASE: AFFITTO FIGURATIVO STIMATO CON LE DIFFERENTI TECNICHE =100			
Area metropolitana – Centro	142,6	123,9	125,4
Area metropolitana – Periferia	114,3	121,1	121,3
Più di 50.000 abitanti	106,6	108,0	108,5
Da 10.000 a 50.000 abitanti	89,2	98,6	98,2
Meno di 10.000 abitanti	82,3	80,5	80,0
Totale	100 (=530 €)	100 (=426 €)	100 (=390 €)
BASE: AUTOVALUTAZIONE PER TIPO DI COMUNE=100			
Area metropolitana – Centro	100 (=756 €)	69,8	64,7
Area metropolitana – Periferia	100 (=606 €)	85,1	78,1
Più di 50.000 abitanti	100 (=565 €)	81,4	74,9
Da 10.000 a 50.000 abitanti	100 (=473 €)	88,8	81,0
Meno di 10.000 abitanti	100 (=436 €)	78,7	71,6
Totale	100 (=530 €)	80,4	73,6

Fonte: Indagine sulle condizioni di vita anno 2004

Prospetto 6.6 - Principali indicatori di Laeken sulla distribuzione di reddito familiare senza e con fitti imputati (secondo differenti tecniche)

REDDITO FAMILIARE EQUIVALENTE	Senza affitto figurativo	Con affitto figurativo stimato tramite:		
		Autovalutazione	Regressione edonica	Edonica con Heckman
RAPPORTO DI CONCENTRAZIONE DI GINI				
Nord Ovest	0,308	0,285	0,279	0,281
Nord Est	0,299	0,269	0,262	0,264
Centro	0,310	0,285	0,277	0,28
Sud	0,343	0,32	0,313	0,315
Isole	0,349	0,324	0,322	0,324
Italia	0,328	0,308	0,301	0,302
PERCENTUALE DI FAMIGLIE A RISCHIO DI POVERTÀ				
Nord Ovest	11,6	9,2	8,6	8,7
Nord Est	10,3	6,9	5,9	6,0
Centro	13,8	10,0	9,9	10,0
Sud	31,6	33,0	31,3	31,0
Isole	35,4	36,5	36,3	36,1
Italia	18,7	17,0	16,2	16,2

Fonte: Indagine sulle condizioni di vita anno 2004

Dal Prospetto 6.6 si vede come l'inclusione degli affitti figurativi nel reddito familiare implichi una minore disuguaglianza, tanto a livello nazionale quanto per singola ripartizione nel Paese. Se è vero che nella parte destra (alta) della distribuzione del reddito familiare si osserva una maggiore incidenza di proprietari è altrettanto vero che questa posta figurativa, comunque calcolata, aumenta in maniera meno che proporzionale rispetto al reddito. L'utilizzo dei differenti metodi di computo dei fitti imputati non sembra invece portare a scostamenti di rilievo, detto che la disuguaglianza è lievemente più elevata con gli affitti soggettivi.

Disaggregando per i diversi titoli di godimento dell'abitazione (Prospetto 6.7), si nota che le incidenze delle famiglie sotto la linea di povertà cambiano considerevolmente, una volta inclusi gli affitti figurativi. Non risulta sorprendente la variazione notevole di tale incidenza per le famiglie che vivono in affitto, passando dal 26,7 quando non si consideri la posta figurativa ad un massimo del 35,1 con i fitti soggettivi. Viceversa, l'incidenza delle famiglie povere diminuisce sensibilmente per gli altri sotto gruppi.

Prospetto 6.7 - Percentuale di famiglie sotto la linea di povertà di Laeken sulla distribuzione di reddito familiare senza e con fitti imputati (secondo differenti tecniche)

REDDITO FAMILIARE EQUIVALENTE	Senza affitto figurativo	Con affitto figurativo stimato tramite:		
		Autovalutazione	Regressione edonica	Edonica con Heckman
PERCENTUALE DI FAMIGLIE A RISCHIO DI POVERTÀ				
Affitto	26,7	35,1	33,3	32,6
Proprietà	15,0	11,2	10,8	11,0
Usufrutto	22,8	12,9	11,9	11,9
Uso gratuito	27,1	23,2	22,0	22,2
Italia	18,7	17,0	16,2	16,2

Fonte: Indagine sulle condizioni di vita anno 2004

6.2.4 Conclusioni

In questo capitolo è stata presentata una sperimentazione sul calcolo dell'affitto figurativo dell'indagine Eu-Silc; questo, stimato per proprietari, usufruttuari e famiglie che vivono a titolo gratuito nella propria abitazione, oltre che per gli affittuari a prezzi inferiori ai prezzi di mercato, deve dare conto del flusso di servizi abitativi che si ottiene in virtù del proprio titolo di godimento e deve essere considerato come parte del reddito disponibile secondo le disposizioni di Nazioni Unite, Ilo e Eurostat. Per l'indagine Eu-Silc, Eurostat prevede l'inclusione di questa posta figurativa nel reddito disponibile a partire dall'indagine 2007, per dare tempo ai Paesi membri di arrivare ad una soluzione condivisa tra le diverse metodologie possibili. Si è quindi stimato questo flusso di beni e servizi tramite modello di regressione secondo la teoria della domanda edonica, per la quale il valore di un'abitazione è dato dall'intersezione tra domanda e offerta delle caratteristiche della casa stessa. Questo metodo è stato implementato come alternativa all'utilizzo del metodo dell'autovalutazione da parte dei rispondenti, che potrebbe essere troppo influenzato dalla mancata o parziale conoscenza del mercato degli immobili da parte di alcuni rispondenti. Tre sono le questioni affrontate:

- anzitutto, se la metodologia dell'autovalutazione da parte dei rispondenti porti effettivamente a una non corretta valutazione del valore del flusso dei servizi abitativi. I fitti soggettivi sono, in media, più elevati di quelli imputati tramite tecniche econometriche. Ma questo, di per sé, non indica quale sia il livello più vicino al valore "vero". La comparazione fra le stime edoniche sugli affitti realmente pagati per quanto attiene agli inquilini a prezzi di mercato e sui fitti soggettivi stimati dai proprietari fa emergere come le caratteristiche delle case abbiano impatti differenti. Non è chiaro però se questo sia dovuto a una errata valutazione dei proprietari sull'importanza delle caratteristiche della casa (cosa che escluderebbe l'utilizzo dell'autovalutazione) o piuttosto al fatto che ci si riferisce a due mercati di abitazioni segmentati e completamente differenti (fatto che, al contrario, non escluderebbe l'autovalutazione).
- Relativamente all'utilizzo della regressione edonica, si pone il problema del *tenure discount*: se lo sconto temporale che si osserva tra gli affittuari debba quindi essere considerato anche nella stima del calcolo figurativo o se, al contrario, quest'ultimo non ne debba tenere conto, come se il proprietario fosse entrato nella casa nella condizione di affittuario alla data dell'intervista.

- Infine, abbiamo verificato l'eventualità dell'utilizzo della procedura di Heckman per considerare l'esistenza di un processo di selezione sulla condizione abitativa. In questa maniera si considera non già il valore di mercato dei flussi abitativi dell'abitazione, quanto piuttosto - dati i prezzi prevalenti sul mercato degli affitti- il valore intrinseco che il proprietario assegna ai servizi abitativi goduti, ivi compreso il costo opportunità del cambiamento di status: una sorta di affitto di riserva che la famiglia accetterebbe per diventare affittuaria nella sua stessa casa, con la possibilità di investire in altro modo i propri soldi, ma scontando comunque una maggiore insicurezza.

Stando alla nostra interpretazione del manuale di Canberra, che definisce le linee guida per le metodologie sul calcolo del reddito disponibile delle famiglie, lo sconto temporale deve essere "sterilizzato", ma il processo di selezione va invece considerato per tenere conto del fatto che per convincere un proprietario a diventare affittuario occorre offrirgli un affitto inferiore ai prezzi di mercato. Questa soluzione si presta comunque a più di una critica, puntualmente presentate nei paragrafi precedenti, ma sembra comunque essere da preferire rispetto all'autovalutazione dell'affitto figurativo. Quest'ultimo sconta infatti l'errata conoscenza della reale situazione del mercato immobiliare da parte delle famiglie proprietarie che non hanno particolari motivi (quale ad esempio un cambio di casa previsto a breve) per conoscere questo mercato, soprattutto relativamente al mercato degli affitti.

In ogni caso, è importante ricordare come il metodo proposto sia molto esigente dal punto di vista informativo; allo stato attuale esiste un problema di sottospecificazione del modello dovuto proprio alle informazioni presenti sul questionario. Anzitutto, la selezione su quanti sono in affitto ai prezzi di mercato è fatta sulla valutazione dei rispondenti, ed è vincolata quindi ad un alto grado di soggettività. Sarebbe preferibile, ed è già stato previsto per il questionario dell'indagine relativa ai redditi 2006, avere l'informazione su chi sia il proprietario dell'immobile (Ente, privato, società, parente, eccetera) e sul tipo di contratto di locazione; questo permetterà di selezionare più precisamente, per la stima dei modelli, quanti sono in affitto da privati e con patti in deroga, quanti cioè sono presumibilmente affittuari ai prezzi di mercato. Dalla stessa data, saranno inoltre inserite informazioni aggiuntive sulle caratteristiche dell'abitazione, quali ad esempio la disponibilità di garage o box o l'esistenza dei doppi servizi, che sicuramente possono fornire informazioni rilevanti sul valore di mercato di un immobile.

6.3 - Dai redditi netti ai redditi lordi: metodologia di stima e prospettive

Il Regolamento Eu-Silc richiede ai Paesi membri di fornire dal 2007 i redditi lordi per tutte le componenti di reddito individuali e per tre variabili target di reddito familiare. Nel caso in cui i redditi lordi non siano disponibili dalle indagini o dalle fonti amministrative, per la conversione netti-lordi si possono imputare le tasse e i contributi sociali usando un modello di microsimulazione basato sul regime fiscale relativo al periodo di riferimento del reddito.

A questo scopo, l'Istat sta sperimentando l'integrazione di due metodologie che, congiuntamente, costituiscono un'importante innovazione metodologica in materia di misurazione dei redditi familiari ed individuali al lordo delle imposte e dei contributi sociali: (i) un modello di microsimulazione delle imposte e dei contributi sociali e: (ii) l'integrazione di dati campionari con dati amministrativi.

Prima del progetto Eu-Silc, per la conversione dei redditi netti in lordi sono stati sviluppati modelli di microsimulazione "separati", basati su indagini nazionali non armonizzate. Tali modelli vengono utilizzati tipicamente per analizzare la distribuzione dei redditi e della povertà e simulare l'impatto delle misure di politica economica e fiscale. È questo il caso, ad esempio, del

Tax-Benefits Model della *London School of Economics*, di Euromod e dei modelli di tasse e benefici impiegati dall'Oecd.

Se i modelli di microsimulazione “nazionali” complessi, pur essendo generalmente più precisi, non risultano sufficientemente flessibili e “trasportabili” ad altri Paesi, quelli più semplici, che considerano soltanto gli aspetti centrali dello schema di tassazione nazionale, non garantiscono necessariamente la comparabilità internazionale.

Il progetto Eu-Silc ha offerto un'occasione importante per disegnare un modello di microsimulazione flessibile e adattabile alle diverse realtà nazionali, che avrà oltretutto il vantaggio di poter partire da una base armonizzata di dati di input.

Per costruire le variabili obiettivo del progetto Eu-Silc, un gruppo di ricerca dell'Università di Siena coordinato da Vijay Verma ha predisposto il modello di microsimulazione Sm2, adottato dalla Commissione europea come procedura raccomandata nell'ambito del progetto Eu-Silc.

Il modello Sm2 è stato concepito come uno strumento in grado di fornire una procedura robusta di conversione dei redditi netti e lordi delle famiglie. Il modello, costruito ed utilizzato sperimentalmente sui microdati dell'indagine Echip²³ dell'Italia, Francia e Spagna, è caratterizzato da una struttura modulare e flessibile che consente l'utilizzo in Paesi che presentano diversi sistemi di tasse e benefici, diverse tipologie dei dati di reddito disponibili (netti, lordi, alcuni netti e altri lordi) ed anche diverse strategie di raccolta dei dati (da fonte campionaria e/o amministrativa).

Il modulo di conversione netto-lordo del modello Sm2 tende ad un livello intermedio di complessità, ossia è più preciso del metodo semplice applicato con l'Echip per stimare i redditi lordi partendo dalle componenti nette raccolte con l'indagine ma è, salvo estensioni decise dai ricercatori dei singoli Paesi, meno dettagliato rispetto ad un modello altamente specializzato come Euromod.

Euromod è sostanzialmente un modello costruito per rappresentare nel dettaglio e stimare gli effetti delle politiche sociali e fiscali sui redditi familiari nei Paesi membri dell'Unione. Il modello consente in effetti di simulare i trasferimenti sociali, completa le informazioni raccolte mediante le indagini campionarie e utilizza come dati di input i redditi lordi piuttosto che i netti.

Lo scopo immediato del progetto Eu-Silc è, in un certo senso, più limitato, poiché considera prioritariamente la costruzione delle variabili di reddito lordo da quelli netti (o viceversa) e solo come utile “sotto prodotto” l'eventuale stima di dati dettagliati sui singoli benefici e sulle singole tasse.

La procedura di conversione netti-lordi utilizzata in Euromod non è, inoltre, sufficientemente flessibile rispetto all'eterogeneità dei dati presenti in Eu-Silc ed il modello Sm2 risponde all'esigenza di disporre di uno strumento di conversione netti-lordi adattabile tanto ai diversi sistemi fiscali europei quanto alle diverse modalità di raccolta dei dati di input nei diversi Paesi.

L'implementazione del modello Sm2-Eu-Sil-Italia, sulla base del modello Sm2, ha richiesto preliminarmente la transizione dalla versione sperimentale applicata ai dati Echip a quella ai dati Eu-Silc e la costruzione del file di input e delle variabili ausiliarie sulla base delle informazioni disponibili nella nuova indagine Eu-Silc. Nei paragrafi seguenti viene descritta la struttura e le principali caratteristiche del modello Sm2 e le fasi di lavoro e le innovazioni metodologiche introdotte con la costruzione del modello Sm2-Eu-Silc Italia da parte dell'Istat.

²³ Panel europeo sulle famiglie.

6.3.1 *La struttura del Siena Microsimulation Model (Sm2)*

Partendo dai dati di reddito individuali e familiare, il modello è in grado di stimare, in base al sistema fiscale nazionale, tutte le informazioni per componente di reddito e di disaggregare gli importi lordi in tasse, contributi sociali e in reddito netto e reddito disponibile.²⁴

Caratteristica rilevante del modello è la presenza di un set di *routines* principali standardizzate, che rappresentano il cuore del modello e la struttura comune utilizzabile da tutti i diversi sistemi fiscali, in grado di gestire diversi tipi di dati di input e diversi sistemi di tassazione nazionale. Ci sono poi le *routines* specifiche contenenti i parametri della normativa fiscale nazionale, in grado di convertire i dati di input in forma standardizzata. Queste *routines* nazionali sono richiamate dal modulo centrale a cui forniscono i dati di input; il modello mantiene distinte la struttura centrale, comune ai diversi sistemi fiscali nazionali, e le *routines* specifiche per i diversi Paesi.

Sm2 può convertire in redditi lordi i dati campionari raccolti al netto delle imposte e dei contributi sociali, ma può anche fornire i redditi netti a partire dai dati lordi (campionari o amministrativi). Il modello può stimare separatamente sia le imposte anticipate (cioè trattenute alla fonte) sulle singole componenti di reddito, sia il saldo finale a credito o a debito.

Per la costruzione delle variabili obiettivo nette e lorde di Eu-Silc sono necessarie almeno cinque tipi di variabili: ossia variabili raccolte mediante l'indagine, variabili intermedie, variabili imputate, variabili del modello e variabili target. Poiché nel file di input non vi possono essere valori missing, è necessario imputare i valori mancanti usando le variabili ausiliarie o esogene. Le informazioni di reddito necessarie al modello Sm2 sono, quindi, in qualche modo raccolte, compilate o imputate. Il modello converte tali informazioni nello standard richiesto da Eu-Silc, sulla base di uno specifico schema di tassazione nazionale.

Il reddito familiare in Eu-Silc scaturisce dalla somma dei redditi individuali dei componenti della famiglia. La normativa fiscale per le diverse fonti di reddito può fare riferimento a differenti unità di analisi, come gli individui o la famiglia e può differire in base alle caratteristiche del percettore (alcuni possono essere esenti da tassazione e altri no, alcuni possono beneficiare di detrazioni e altri no). Il reddito percepito può, inoltre, essere tassato alla fonte, ricevuto lordo o tassato in sede di dichiarazione dei redditi.

Il programma di calcolo della tassazione e dei trasferimenti sociali di un modello di microsimulazione replica sostanzialmente le caratteristiche rilevanti di un sistema fiscale nazionale, sulla base di alcune assunzioni che semplificano necessariamente la complessità della realtà osservata.

6.3.1.1 *Principali assunzioni del modello e relazioni basilari tra reddito, imposte, contributi sociali*

Quando si utilizza un modello di microsimulazione occorre tenere presente che il grado di complessità del modello deriva sostanzialmente dal livello di dettaglio con cui è inserita la normativa fiscale e dalla necessità di un suo aggiornamento annuale, oltre che dalla presenza di differenti periodi di riferimento per il reddito, la tassazione, o per i saldi fiscali. Vi possono essere, inoltre, problemi di inconsistenza tra i redditi dichiarati nell'indagine e quelli calcolati con la microsimulazione.

²⁴ Si veda: Eurostat, Income in EU-SILC: Net/Gross/Net conversion. Report on common structure of the model; model description and application to the Echip data for France, Italy and Spain, prepared by V. Verma, G. Betti and co-researcher. EU-SILC 133/04, Luxembourg 2004.

Nel modello Sm2, considerando la situazione in cui una persona riceve reddito da un'unica fonte ed è tassata separatamente come singola "unità fiscale"²⁵ si hanno le seguenti relazioni:

1 Reddito lordo	$GG = G + SS(G)$
2 Contributi sociali	$S = S(G)$
3 Reddito lordo tassabile	$H = G - S$
4 Ritenuta alla fonte di tasse e contributi	$XS = H$
	$XST = H - T(H)$
	$XT = H + S(G) - T(H)$
5 Deduzioni	$D = D(H)$
6 Reddito netto imponibile	$Y = H - D$
7 Imposta dovuta	$W = W(Y)$
8 Detrazioni	$C = C(Y)$
9 Imposta pagata	$X = W - C$
10 Reddito netto	$N = H - X$

SS:	Contributi sociali datore di lavoro
XS	Ritenuta alla fonte dei contributi sociali
XST	Ritenuta alla fonte delle tasse e dei contributi
XT	Ritenuta alla fonte delle tasse

Il reddito lordo di un individuo o di una famiglia è uguale al reddito ricevuto prima di qualsiasi ritenuta alla fonte delle tasse e dei contributi sociali. Il modello distingue tra il reddito lordo GG che include anche i contributi sociali dei datori di lavoro (SS) e il reddito lordo G che include solo i contributi sociali dei lavoratori (SI). I contributi sociali gravano sui redditi da lavoro e comprendono i contributi dei datori di lavoro e dei lavoratori dipendenti e degli autonomi.

In Sm2, il reddito lordo tassabile (H) è uguale al reddito lordo meno i contributi sociali. Gli oneri deducibili si riferiscono a quella parte di reddito esente dalla tassazione, sono relativi a specifiche componenti di reddito e corrispondono ad una riduzione del reddito lordo tassabile.

Il reddito imponibile (Y) è ricavato dal reddito lordo tassabile meno le deduzioni. L'imposta lorda è calcolata inizialmente come funzione del reddito imponibile.

Le detrazioni d'imposta corrispondono ad una diminuzione dell'imposta dovuta, dipendono generalmente dalle caratteristiche dell'unità fiscale (lavoratore, pensionato, monogenitore, eccetera), oppure compensano particolari spese sostenute (mediche, istruzione, mutuo, eccetera). L'imposta netta pagata è uguale all'imposta lorda meno le detrazioni ed infine il reddito netto (N) è uguale al reddito lordo tassabile meno l'imposta pagata.

Se si considera una situazione più realistica in cui vi sono più fonti di reddito o più di un individuo nell'unità fiscale, la relazione tra redditi netti e lordi è esemplificata nella Figura 6.1.

In questo caso occorre distinguere essenzialmente tra le componenti di reddito che vanno sommate per tipologia e successivamente aggregate per individui nell'unità fiscale e le componenti che vanno trattate separatamente ai fini dell'imposizione fiscale e contributiva.

In Sm2, la relazione tra reddito lordo tassabile di una singola componente di reddito e reddito lordo è comunque semplice ed indipendente dalle altre componenti di reddito o dagli altri

²⁵ Con unità fiscale si intende l'individuo o la famiglia i cui redditi costituiscono l'imponibile per la tassazione e l'imposizione contributiva.

individui nell'unità fiscale. Questo vale anche per la relazione tra reddito lordo e reddito netto per le componenti che sono tassate separatamente ad un'aliquota costante o per le componenti di reddito che sono esenti da tassazione.

La relazione, invece, tra reddito lordo tassabile e reddito netto in caso di redditi aggregati per tipologia è meno diretta.

Nel caso in cui siano disponibili i redditi lordi, mediante l'aliquota della tassazione è possibile ottenere i redditi netti utilizzando una iterazione semplice per ogni tipologia di reddito.

Il procedimento inverso dai netti ai lordi è, invece, più complesso, in quanto implica un doppio *loop* iterativo. Il primo *loop* interno applica un valore definito del parametro "aliquota di tassazione" (R) ad ogni componente di reddito e per tutti gli individui della stessa unità fiscale. Il secondo *loop* iterativo ottiene un valore convergente del parametro R, comune a tutte le componenti di reddito.

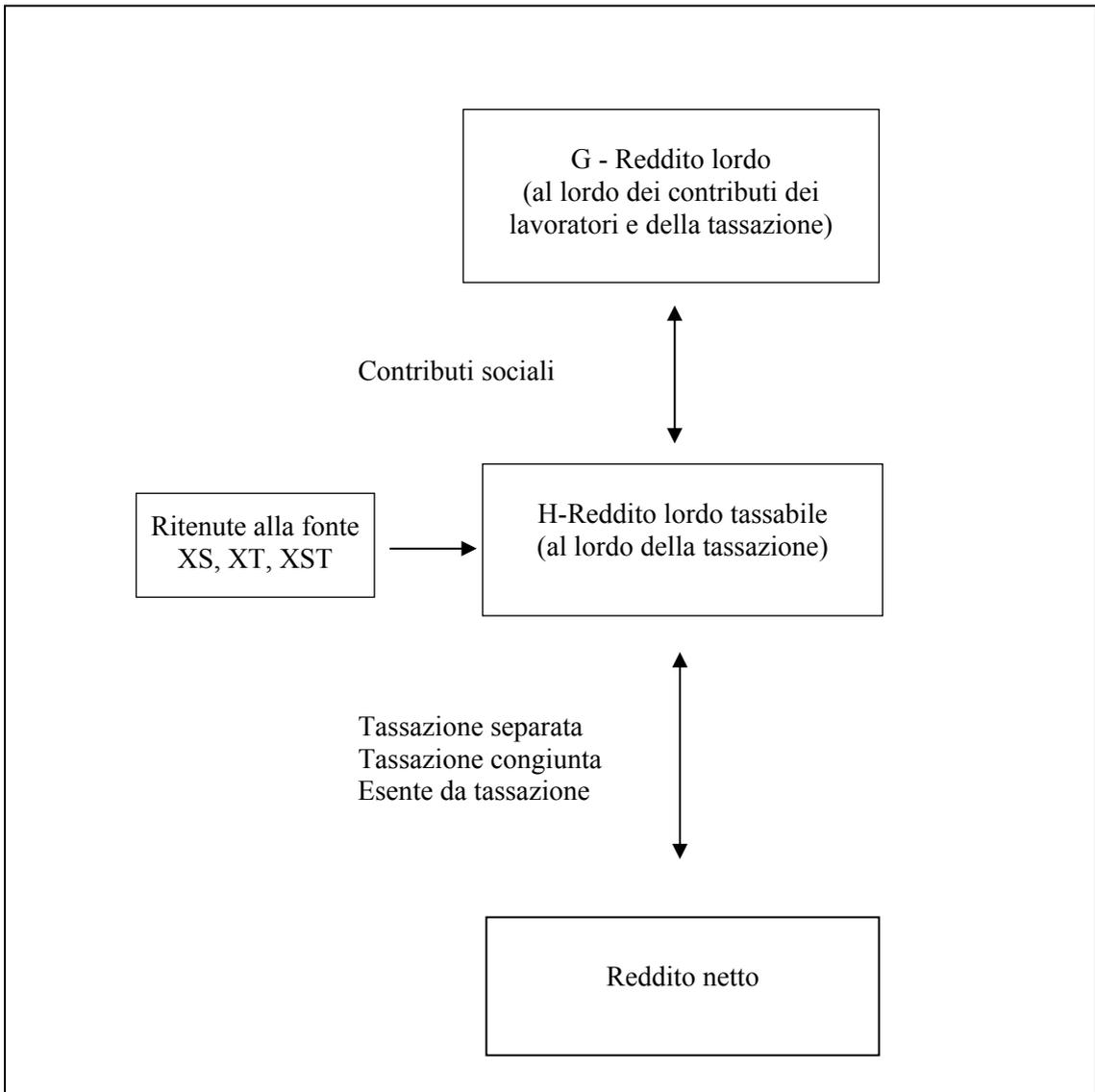
Il parametro R non si configura quindi come una semplice aliquota media applicata al reddito totale, ma piuttosto come un'aliquota specifica per ogni componente di reddito la quale consente di disaggregare la tassazione per tipologia di reddito.

Caratteristica peculiare di Sm2 è la possibilità di specificare in modo opportuno le deduzioni e le detrazioni in modo da utilizzare la stessa procedura per le diverse componenti di reddito e per i diversi schemi di tassazione.

Nel caso in cui una componente di reddito è esente da tassazione è sufficiente specificare che per la componente esente la deduzione è pari al reddito lordo tassabile, ciò consente di non considerare il contributo di questa componente nel totale del reddito sottoposto a tassazione e di non dover separare tale componente dagli altri redditi.

Allo stesso modo se una componente di reddito è tassata con un'aliquota costante, si specifica che la deduzione è pari al reddito lordo tassabile e la deduzione è considerata come una quantità negativa. In tal modo l'imponibile finale è automaticamente incrementato dell'ammontare appropriato.

Figura 6.1 - Relazione tra redditi lordi e netti



6.3.2 La realizzazione del modello Sm2 -Eu-Silc-Italia per la conversione dei redditi netti in lordi

Il file di input del modello Sm2 è il risultato di diversi programmi di calcolo necessari alla predisposizione della base dati e delle variabili ausiliarie. In Sm2 le variabili ausiliarie sono costruite sui dati Echp Udb Pdb²⁶ dell'anno 1998 ed identificano per i tre Paesi a cui è stato applicato (Italia, Francia e Spagna) il set di variabili ausiliarie necessarie in base alla normativa fiscale nazionale.

²⁶ Echp Users' Data Base (Udb) e Production Data base (Pdb).

Per quanto riguarda l'Italia, Sm2 applica ai dati Echp del 1998 il sistema di tassazione dell'anno 2003, in modo da includere la riforma dell'Irpef del 2003, con *routines* specifiche per il calcolo dei contributi sociali dei lavoratori dipendenti e autonomi e della tassazione Irpef a livello nazionale, regionale e comunale.

L'implementazione del modello Sm2-Eu-Silc-Italia ha richiesto preliminarmente la transizione dai dati Echp ai dati Eu-Silc e la costruzione del file di input e delle variabili ausiliarie sulla base delle informazioni disponibili nella nuova indagine, con diverse fasi di lavoro che possono essere sintetizzate come segue:

a) Preparazione del file di input del modello:

- sostituzione diretta, dove possibile, delle variabili Echp con le nuove variabili Eu-Silc;
- costruzione delle variabili ausiliarie del modello;
- costruzione della variabile obiettivo Py120n e completamento della variabile obiettivo Hy050n.

b) Modifica di alcune procedure di conversione netti-lordi per la stima:

- del reddito da lavoro autonomo;
- dell'Irap (imposta regionale sulle attività produttive)
- dei contributi sociali dei lavoratori parasubordinati.

c) Introduzione di alcune innovazioni di processo mediante il *record linkage* con i dati amministrativi.

6.3.2.1 La preparazione del file di input del modello

La preparazione del file di input, sulla base dei dati della prima indagine Eu-Silc disponibile (anno 2004, con periodo di riferimento dei redditi anno 2003), è stata effettuata procedendo dapprima con la sostituzione diretta, dove possibile, delle variabili target Eu-Silc alle corrispondenti variabili Echp. Successivamente si è proceduto a costruire le variabili target che nella prima wave di Eu-Silc non erano ancora state calcolate, a cui è seguita la costruzione delle variabili ausiliarie del modello.

Fra le variabili target necessarie al modello ma non disponibili nell'edizione 2004 dell'indagine vi era la variabile Py120n, con cui si calcola l'indennità di malattia e astensione facoltativa dei lavoratori dipendenti e il congedo parentale delle lavoratrici autonome. È stata quindi costruita la Py120n utilizzando, in caso di malattia dei lavoratori dipendenti, la retribuzione media giornaliera per le giornate indennizzabili dall'Inps e per la quota parte posta a carico dei datori di lavoro. Allo stesso modo è stata quantificata l'indennità prevista per la cura dei figli malati, nel caso dei lavoratori della Pubblica Amministrazione.

Per quanto riguarda il congedo parentale delle lavoratrici autonome (artigiane, commercianti e agricole), per le quali è prevista un'astensione massima di tre mesi nel primo anno di vita del bambino, non essendo disponibile l'informazione sui giorni di astensione usufruiti dalle lavoratrici è stato calcolato l'importo erogato dall'Inps, utilizzando l'indennità giornaliera convenzionale prevista dalla normativa, ed ipotizzando un utilizzo pieno del congedo (tre mesi) da parte delle lavoratrici che hanno dichiarato di avere ricevuto l'indennità. Per l'edizione 2005 dell'indagine Eu-Silc è stata modificata la domanda del questionario, in modo da ottenere i giorni di astensione.

Nell'edizione 2004 dell'indagine Eu-Silc, l'indennità di maternità spettante alle lavoratrici dipendenti non è stata scorporata dal reddito da lavoro delle stesse e non è stata inclusa nella variabile target Hy050n (trasferimenti sociali alle famiglie).²⁷ Per stimare l'importo delle prestazioni sociali includendo anche i lavoratori dipendenti è stata calcolata l'indennità di maternità corrisposta dall'Inps per le giornate indennizzabili, stimando separatamente anche l'integrazione della retribuzione da parte dei datori di lavoro prevista dalla maggior parte dei Ccnl. Per le lavoratrici autonome (artigiane, commercianti, agricole) è stata ricalcolata l'indennità di maternità sulla base della retribuzione giornaliera convenzionale dell'Inps ed è stata stimata l'indennità per le libere professioniste e le lavoratrici parasubordinate (i collaboratori coordinati e continuativi e i collaboratori a progetto).

Fra le numerose variabili ausiliarie necessarie al modello, particolare attenzione è stata riservata alla costruzione delle "tax units" all'interno delle famiglie, per stimare le detrazioni spettanti per coniuge, figli e parenti a carico, e all'individuazione dei settori di attività economica utile al calcolo dell'imposizione contributiva per il lavoro dipendente e autonomo.

Come è noto, il sistema fiscale nazionale è basato su un'imposizione a livello individuale.²⁸ A livello familiare vi possono essere due tipi di "unità fiscali":

- a) "la famiglia" che include il capofamiglia e le persone a carico; il reddito sottoposto a tassazione è soltanto quello del capofamiglia e i redditi delle persone a carico sono esenti da tassazione: vi è quindi un'unica *tax units*;
- b) "gli individui" che fanno parte della famiglia e che non sono a carico e che dichiarano un proprio reddito separatamente; ogni individuo forma una unità fiscale separata.

I membri della famiglia sono considerati "fiscalmente a carico" e danno diritto ad una detrazione se nel 2003 il loro reddito non ha superato la soglia di 2.840,51 euro. È stato quindi necessario ricostruire le famiglie rilevate dall'indagine per individuare:

- a) le famiglie nucleari con capofamiglia e i fiscalmente a carico;²⁹
- b) i membri isolati che sono parte della famiglia ma che dichiarano un proprio reddito.

L'individuazione delle *tax units* in Eu-Silc Italia si è avvalsa di uno strumento definito "procedura famiglie" utilizzato per la classificazione delle famiglie nelle indagini sociali dell'Istat, il quale consente di ricostruire le relazioni nei nuclei familiari e successivamente di definire, con opportune modifiche, i fiscalmente a carico.

Per il calcolo dei contributi sociali del lavoro dipendente e autonomo è necessario individuare il settore di attività economica del lavoratore e il numero di addetti dell'impresa sulla base dei raggruppamenti Nace. Sono state quindi utilizzate le maggiori informazioni disponibili in Eu-Silc rispetto a Echp e relative alla posizione lavorativa, al settore di appartenenza, al calendario retrospettivo della condizione professionale (numero di mesi) e al numero di addetti dell'unità locale dell'impresa.

Alcune variabili dell'indagine Eu-Silc, come i trasferimenti monetari non pensionistici alle famiglie, i trasferimenti privati tra le famiglie o i contributi per la casa, sono rilevate, inoltre, a

²⁷ La variabile Hy050n include gli assegni al nucleo familiare, l'assegno di sostegno che spetta alle famiglie con almeno tre figli minori, l'assegno per la nascita di un secondo figlio, l'indennità di maternità delle lavoratrici autonome e l'indennità di maternità per le madri non indennizzate erogata dallo stato o dai comuni.

²⁸ La scelta tra tassazione separata o congiunta, nel caso dei coniugi, riguarda solo la fase finale del pagamento delle tasse.

²⁹ Possono essere fiscalmente a carico:

- i figli indipendentemente dall'età;
- la moglie non divorziata non separata;
- la moglie anche divorziata ma se vive con il contribuente o se riceve alimenti non obbligatori;
- altre persone: parenti, nonni, figli e generi, nuore, nipoti solo se vivono con il contribuente oppure se ricevono trasferimenti interfamiliari.

livello familiare. Ai fini del calcolo dell'imponibile individuale, nel file di input del modello, gli importi dichiarati sono stati attribuiti al maggiore percettore di reddito all'interno della famiglia.

6.3.2.2 *Il sistema fiscale e contributivo nel modello Sm2- Eu-Silc Italia*

Dopo la preparazione del file di input è stata implementata una prima versione del modello Sm2-Eu-Silc, introducendo le variabili dell'indagine Eu-Silc 2004 come una delle modifiche rilevante rispetto alle assunzioni e alle procedure previste in Sm2.

La transizione dai dati Echn ai dati Eu-Silc ha richiesto, oltre allo sviluppo di nuove procedure per la costruzione del file di input, anche la modifica di alcune *routines* di conversione netti-lordi, in particolare per la stima del reddito da lavoro autonomo e per il calcolo dell'imposta regionale sulle attività produttive (Irap). Allo stesso modo è stato necessario modificare la procedura per stimare il reddito derivante da attività di collaborazioni coordinate e continuative.

Il modello Sm2- Eu-Silc presenta quattro *routines* specifiche, richiamate dal programma principale attraverso macro di sas, due delle quali stimano i contributi sociali pagati dai lavoratori dipendenti e dagli autonomi e le restanti due calcolano le imposte pagate e l'Irap dovuta dai lavoratori autonomi.

In Italia, la principale imposta sul reddito (Irpef) è calcolata applicando un'aliquota marginale progressiva a scaglioni di reddito crescenti. Nel modello, le diverse fonti di reddito sono raggruppate per tipologia e sottoposte alla tassazione sul reddito.

I redditi da lavoro (dipendente e autonomo) sono sottoposti al prelievo contributivo obbligatorio. I lavoratori autonomi oltre alla tassazione sul reddito pagano l'Irap. Nel modello la doppia tassazione degli autonomi è trattata considerando l'Irap come una detrazione negativa.

Il reddito da attività finanziarie è tassato con un'aliquota costante che varia in base all'ampiezza del capitale, ma indipendentemente dal reddito del contribuente. Non contribuisce all'imponibile totale e nel modello è considerato come una detrazione specifica negativa.

Il reddito da capitale reale è considerato in modo simile al reddito da attività finanziarie, ma è sottoposto alla tassazione Irpef.

I trasferimenti sociali alle famiglie (assegni familiari), alcune pensioni di invalidità (erogate dall'Inail), i contributi per la casa non sono sottoposti a tassazione e quindi nel modello non contribuiscono all'imponibile totale.

I trasferimenti per l'istruzione e le indennità di disoccupazione, al pari del reddito da lavoro dipendente, sono sottoposti invece a tassazione.

I contributi previdenziali e assistenziali obbligatori non sono sottoposti a tassazione Irpef, a differenza dei contributi previdenziali volontari. Il modello include i contributi sociali obbligatori per i lavoratori dipendenti, autonomi e dei lavoratori parasubordinati. Per i lavoratori dipendenti e per i datori di lavoro, l'aliquota contributiva varia in base al settore di appartenenza, al numero di addetti e alla posizione lavorativa del dipendente: Sm2 calcola le aliquote contributive per gli operai, gli impiegati e i dirigenti dell'industria e delle aziende artigiane.

Per quanto riguarda gli autonomi, il modello stima i contributi sociali pagati dai commercianti, dagli artigiani, dai lavoratori agricoli e dai liberi professionisti e da quelli iscritti alla gestione separata dell'Inps.

Il modello calcola anche i contributi sociali dei lavoratori parasubordinati, i collaboratori coordinati e continuativi (CoCoCo) e i collaboratori a progetto (CoCoPro), considerati autonomi, da un punto di vista fiscale, ma sottoposti ad una disciplina specifica con aliquote contributive, da parte del datore di lavoro e dei collaboratori, inferiori rispetto ai lavoratori dipendenti. L'indagine Echn non rilevava i lavoratori parasubordinati e in Sm2 questa categoria di lavoratori era imputata da fonti esterne. Le informazioni disponibili in Eu-Silc consentono di includere nel file di input di Sm2-Eu-Silc i dati sui parasubordinati rilevati dalla nuova indagine.

Oltre al calcolo della tassazione Irpef, il modello stima anche le addizionali regionali e comunali, mentre i pagamenti differiti (Tfr) e le liquidazioni non sono attualmente inclusi nel modello e non costituiscono montante retributivo.

Nel precedente modello Sm2, i dati necessari al calcolo delle deduzioni di reddito e della detrazioni d'imposta sono imputati con tecniche di regressione sulla base di fonti esterne. In particolare, le deduzioni dall'imponibile previste dalla normativa fiscale sono imputate utilizzando i dati contenuti nel Rapporto di Euromod per l'Italia (anno 2004), ossia come percentuali di spesa derivata dai dati di contabilità nazionale per scaglioni di reddito.

Le detrazioni di imposta per alcune spese sostenute dalle famiglie, le cui informazioni non sono correlate al reddito ma dipendono dalle preferenze e dai livelli di consumo (spese mediche, istruzione, eccetera), sono imputate in Sm2 sulla base dei dati dell'Indagine sui consumi delle famiglie dell'Istat e dell'Indagine sui Bilanci delle famiglie della Banca d'Italia. Il modello Sm2-Eu-Silc-Italia mantiene, in una prima versione, la procedura di imputazione delle deduzioni di Sm2 per poi modificarla successivamente ed integrarla, mediante procedure di *record linkage*, con i dati fiscali relativi agli oneri deducibili e detraibili.

Il modello Sm2-Eu-Silc-Italia, una volta richiamate le *routines* per la stima della tassazione e della contribuzione obbligatoria, costruisce i dati aggregati dalle variabili di base necessarie per la procedura di conversione netti-lordi, e genera due output finali contenenti i microdati delle variabili target nette e lorde di Eu-Silc e i valori medi delle stesse.

6.3.3 Le innovazioni nel modello Sm2-Eu-Silc-Italia

La qualità attesa dei risultati delle microsimulazioni e la loro comparabilità a livello internazionale dipendono dal livello di dettaglio con cui è inserita la normativa fiscale nel modello e soprattutto dalla qualità dei dati di input.

I dati rilevati dalla nuova indagine Eu-Silc rappresentano di per sé un'importante cambiamento rispetto ad Echp, evidente soprattutto nella fase di costruzione del file di input del modello e delle variabili ausiliarie. Il contributo fornito da Eu-Silc è particolarmente significativo per il calcolo dei contributi sociali dei lavoratori parasubordinati rilevati dall'indagine, invece che in base alla stima degli stessi rispetto alla popolazione di lavoratori dipendenti derivata da fonti esterne ed utilizzata in Sm2.

Anche il calcolo dei contributi sociali dei lavoratori dipendenti e autonomi beneficia della disponibilità di maggiori informazioni di Eu-Silc relative alla posizione lavorativa, al settore di appartenenza del lavoratore, al calendario retrospettivo della condizione professionale (numero di mesi) e al numero di addetti dell'unità locale dell'impresa.

In Eu-Silc, inoltre, è disponibile l'informazione sui contributi volontari versati a forme pensionistiche complementari, necessari al calcolo delle deduzioni Irpef e, a differenza di Echp, la nuova indagine rileva distintamente i giorni di malattia dei lavoratori dipendenti e le pensioni di invalidità con evidenti vantaggi per la stima dei trasferimenti sociali. L'indagine Eu-Silc fornisce anche i dati relativi all'imposta comunale sugli immobili (Ici), completando le informazioni necessarie alle *routines* del modello per il calcolo delle deduzioni e detrazioni d'imposta.

L'utilizzo congiunto dei dati campionari e amministrativi per la costruzione delle variabili di reddito dell'indagine Eu-Silc ha consentito, inoltre, di introdurre in Sm2-Eu-Silc-Italia alcune importanti innovazioni metodologiche.

I miglioramenti attesi nella stima dei redditi lordi delle variabili obiettivo individuali e familiari di Eu-Silc dipendono quindi dal cambiamento della base dati Eu-Silc rispetto ad Echp e soprattutto dalle innovazioni di processo dovute alle procedure di *record linkage* con i dati

amministrativi. In sintesi, rispetto al precedente Sm2, il modello Sm2-Eu-Silc-Italia si caratterizza per:

- a) miglioramento qualitativo della base dati:
 - maggiore disponibilità di informazioni;
 - migliore stima dei redditi e dei contributi sociali;

- b) innovazioni di processo:
 - integrazione di fonti campionarie ed amministrative per il calcolo degli oneri deducibili e detraibili;
 - comparazione e validazione dei risultati delle microsimulazioni con i redditi lordi fiscali.

La costruzione di un data set integrato e coerente di tutte le informazioni campionarie e amministrative disponibili³⁰ ha consentito di utilizzare i dati di fonte fiscale come input del modello di microsimulazione, in particolare per il calcolo degli oneri deducibili e detraibili.

Le fonti utilizzate derivano dalle dichiarazioni dei redditi modello 730 e modello Cud, dei lavoratori dipendenti e pensionati, e dal modello Unico Persone Fisiche utilizzato da tutti i contribuenti ed in particolare da chi dichiara redditi da lavoro autonomo. A queste fonti si aggiungono i dati derivati dal modello 770 (quadro SA dei redditi e delle ritenute per coloro che non devono presentare la dichiarazione), al netto dei dati Cud, e le informazioni sui pensionati derivate dal Casellario Pensionistico elaborato congiuntamente dall'Inps e dall'Istat.

A partire dall'indagine Eu-Silc 2007, l'Istat fornirà per la prima volta i redditi lordi delle famiglie residenti in Italia. La produzione di microdati sui redditi al netto e al lordo dell'imposizione fiscale e contributiva derivanti da una stessa indagine campionaria rappresenta una novità importante per il nostro Paese.

L'integrazione delle microsimulazioni con i dati amministrativi e le tecniche di imputazione dei dati mancanti può essere considerato come un avanzamento metodologico delle procedure di conversione dei redditi netti-lordi.

Rispetto all'utilizzo di dati di sola fonte amministrativa, il modello di microsimulazione consente di stimare i contributi sociali pagati dai datori di lavoro non presenti negli archivi fiscali attualmente utilizzati (variabile obiettivo Py030g) e di fornire una stima della tassazione sul reddito, incluso il credito e il debito fiscale, e del prelievo contributivo in Italia (variabile target Hy140n).

Mediante Sm2-Eu-Silc-Italia è possibile ottenere la microsimulazione delle tasse e dei contributi sociali per gli individui campionari non presenti negli archivi fiscali, a causa del fallimento delle procedure di "aggancio" con i dati del fisco, e parimenti per i componenti di "fatto" delle famiglie non presenti nel campione, in quanto aggregati successivamente all'estrazione del campione teorico. Il modello consente, quindi, di ampliare l'informazione disponibile includendo anche le famiglie e gli individui appartenenti alla cosiddetta "no tax area" i quali, per definizione, non sono presenti negli archivi amministrativi.

Infine, la disponibilità di dati amministrativi consentirà di comparare e validare i risultati delle microsimulazioni con i redditi lordi fiscali (al netto dei contributi sociali pagati dai datori di lavoro).

³⁰ Si veda Consolini P., *Integrazione di dati campionari EU-SILC con dati di fonte amministrativa*, in corso di pubblicazione su Collana Metodi e Norme Istat, Roma, anno 2008.

Appendice - Le indagini sui redditi e le misure di benessere

L'Appendice presenta i dati della terza indagine pilota "Indagine sulle condizioni di vita, Anno 2003". La terza indagine pilota assume particolare rilevanza poichè sulla base dei risultati ottenuti con la terza indagine pilota è stato possibile l'avvio e la realizzazione del progetto Eu-Silc.

Questo capitolo analizza in particolare i redditi percepiti nel 2002 dalle famiglie residenti in Italia, gli indicatori di disagio economico e di deprivazione materiale, la disuguaglianza nella distribuzione dei redditi familiari e alcuni indicatori di povertà. Segue, infine, un'analisi comparata con i dati relativi ai bilanci delle famiglie rilevati dalla Banca d'Italia nell'anno 2002.

A.1 - Reddito ed indicatori di disagio economico e di deprivazione materiale

Nell'appendice A.1 sono presentati alcuni dei risultati più rilevanti della terza indagine pilota (da ora in poi Icv2003): nella prima parte del lavoro sono analizzate le diverse componenti di reddito che concorrono alla formazione del reddito familiare ed individuale, per ripartizione geografica e genere del percettore, ed i redditi medi per tipologia e caratteristiche della persona di riferimento.

La seconda parte è dedicata agli indicatori di disagio economico e di deprivazione materiale delle famiglie. Essi fanno riferimento ai dati sui vincoli economici, sulla percezione della condizione economica familiare, sull'indisponibilità di beni durevoli, sulle condizioni abitative e sui problemi presenti nella zona di abitazione. L'analisi congiunta degli indicatori non monetari di deprivazione materiale e di reddito familiare equivalente consente di delineare un quadro di sintesi sulle condizioni di vita delle famiglie nel 2003.

A.1.1 Le fonti di reddito familiare e individuale

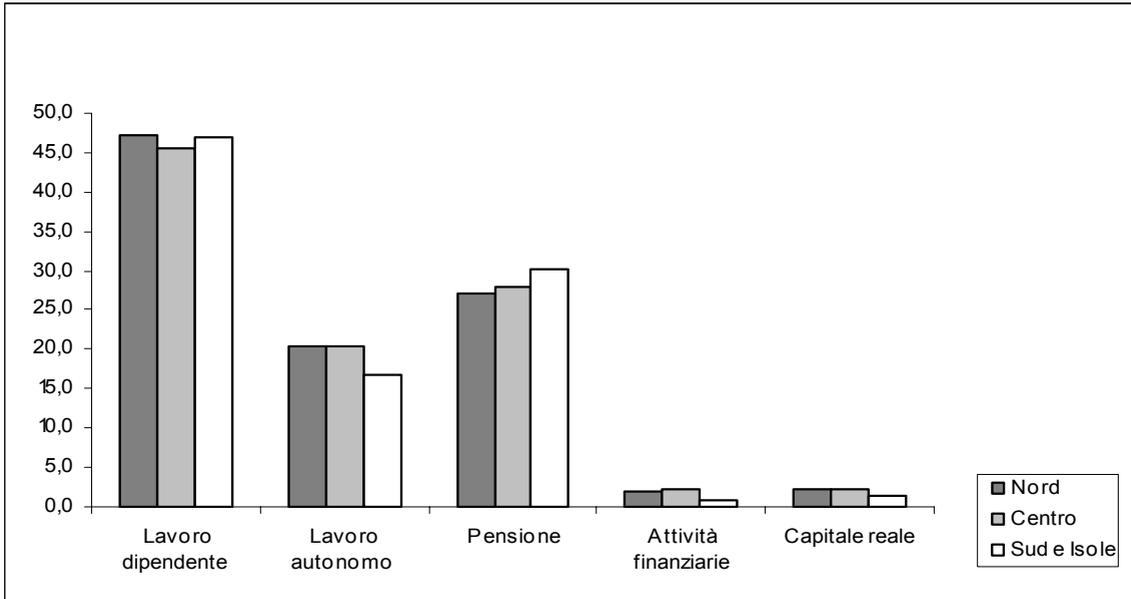
L'analisi che segue riguarda i redditi familiari al netto dei fitti imputati delle abitazioni occupate dai proprietari, oltre che delle imposte sul reddito e dei contributi previdenziali ed assistenziali. Queste componenti, come previsto dal Regolamento comunitario di Eu-Silc, saranno disponibili a partire dal 2007.

La figura A.1 descrive le principali componenti di reddito che concorrono alla formazione del reddito familiare totale nelle tre ripartizioni del Paese.¹ Appare evidente nella composizione del reddito familiare la prevalenza dei redditi da pensione e da lavoro dipendente, in particolare al Sud e nelle Isole, ed il ruolo secondario dei redditi da capitale reale e da attività finanziarie.

Il paragrafo A.1 è a cura di Gabriella Donatiello, il paragrafo A.2 è a cura di Isabella Siciliani

¹ Occorre considerare che, per l'anno 2002, i redditi da lavoro dipendente non includono il reddito derivante dall'utilizzo di macchine aziendali, i redditi da lavoro autonomo non includono le perdite dei lavoratori autonomi, i redditi da pensione includono le pensioni integrative. Il reddito da attività finanziarie risulta costituito dai guadagni delle somme investite in titoli, azioni, obbligazioni, libretti di risparmio e simili al netto dei costi dei servizi bancari ed il reddito da capitale reale non include i fitti imputati.

Figura A.1 - Composizione del reddito familiare totale per ripartizione – Anno 2002 (a) (valori percentuali)



Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003
 (a) Reddito al netto dei fitti imputati.

Le pensioni rappresentano la fonte primaria di reddito per le famiglie dal primo al terzo decile, con una quota che supera il 56 per cento del totale (Tavola A.1). Nel primo decile, le misure di sostegno economico alle famiglie in difficoltà, presenti nella componente altri redditi, costituiscono il 7,2 per cento del reddito totale. Nei decili superiori si osserva un aumento progressivo della quota di reddito fornito dal lavoro dipendente e dal lavoro autonomo. Nei primi otto decili i redditi da capitale reale (si passa dal 0,8 per cento nel primo decile al 4,0 per cento nel decimo) ed anche da attività finanziarie concorrono in modo marginale alla formazione del reddito familiare.²

² È tuttavia probabile che i rendimenti delle attività finanziarie siano stati solo parzialmente indicati dalle persone intervistate.

Tavola A.1 - Composizione del reddito familiare totale per decili – Anno 2002 (a) (valori percentuali)

DECILI	Lavoro Dipendente	Lavoro Autonomo	Pensione	Attività Finanziarie	Capitale Reale	Altro (b)
Primo	18,4	13,6	58,1	1,9	0,8	7,2
Secondo	21,6	11,2	62,6	0,9	1,1	2,6
Terzo	30,5	8,6	56,2	1,0	0,7	3,1
Quarto	38,6	8,9	48,3	0,9	0,8	2,5
Quinto	46,0	10,3	39,2	1,0	0,9	2,6
Sesto	47,6	13,6	33,5	1,8	1,2	2,4
Settimo	52,8	13,9	28,0	1,2	1,3	2,9
Ottavo	58,4	13,5	23,7	1,6	0,9	1,9
Nono	57,4	19,2	18,3	2,0	2,0	1,1
Decimo	42,5	34,0	16,4	2,3	4,0	0,7

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Reddito al netto dei fitti imputati.

(b) Reddito derivante da: assegni per i figli, assegni di maternità, minimo vitale, reddito minimo di inserimento, aiuti in denaro da enti pubblici e privati per le famiglie in difficoltà.

L'apporto delle diverse componenti di reddito alla formazione del reddito individuale rispecchia l'andamento già rilevato a livello familiare con una preponderanza del reddito da lavoro dipendente e da pensione (Tavola A.2). L'analisi per genere conferma che il 65,7 per cento del reddito totale è percepito dagli uomini (Tavola A.3). Se si considera la composizione del reddito individuale nelle tre ripartizioni nazionali, appare evidente che il divario fra la quota di reddito percepito dai maschi e dalle femmine si presenta, conformemente alle attese, più ampio nel Sud e nelle Isole rispetto alle altre due ripartizioni. Ciò vale per tutte le diverse fonti di reddito ad eccezione del capitale reale. In questo caso, la quota percepita dalle donne rappresenta il 51,2 per cento del totale nazionale: al Sud e alle Isole tale quota raggiunge il 52,8 per cento e al Nord arriva al 57 per cento del totale.

Tavola A.2 - Composizione del reddito individuale per ripartizione – Anno 2002 (a) (valori percentuali)

REDDITO	Nord	Centro	Sud e Isole	Italia
Lavoro dipendente	47,2	45,8	47,1	46,9
Lavoro autonomo	20,5	20,6	17,0	19,6
Pensione	27,3	28,0	30,3	28,2
Attività finanziarie	1,9	2,3	0,8	1,7
Capitale reale	2,3	2,1	1,3	2,0
Altro (b)	0,9	1,2	3,5	1,6

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Reddito al netto dei fitti imputati.

(b) Assegni per i figli, assegni di maternità, disoccupazione, minimo vitale, reddito minimo di inserimento, aiuti in denaro ricevuti o dati a persone esterne alla famiglia.

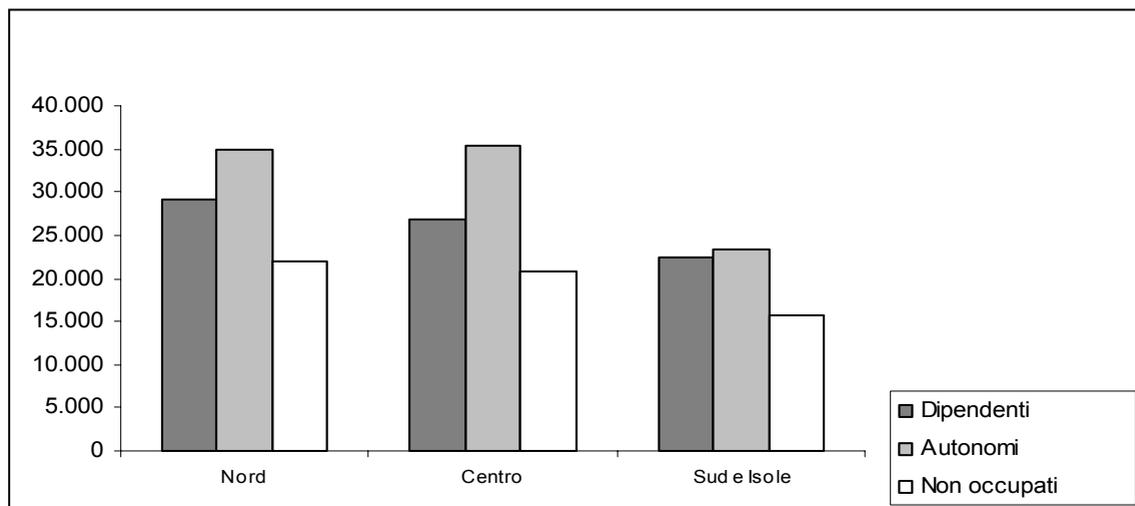
**Tavola A.3 - Composizione del reddito individuale per genere del percettore, tipologia e ripartizione
Anno 2002 (a) (valori percentuali)**

GENERE REDDITO	Nord	Centro	Sud e Isole	Italia
	LAVORO DIPENDENTE			
Maschio	61,8	60,7	67,9	63,1
Femmina	38,2	39,3	32,1	36,9
	LAVORO AUTONOMO			
Maschio	73,2	70,1	77,1	73,4
Femmina	26,8	29,9	22,9	26,6
	PENSIONE			
Maschio	54,4	56,2	57,4	55,6
Femmina	45,6	43,8	42,6	44,4
	ATTIVITÀ FINANZIARIE			
Maschio	56,4	57,0	58,9	56,9
Femmina	43,6	43,0	41,1	43,1
	CAPITALE REALE			
Maschio	43,0	65,7	47,2	48,8
Femmina	57,0	34,3	52,8	51,2
	REDDITO TOTALE			
Maschio	61,3	61,1	65,2	65,7
Femmina	38,7	38,9	34,8	34,3

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003
(a) Reddito al netto dei fitti imputati

L'analisi del reddito medio familiare in base alle caratteristiche socio-demografiche della persona di riferimento, indica livelli di reddito medio più elevati per coloro che dichiarano di essere lavoratori autonomi (Figura A.2), di essere in possesso di una laurea o di una formazione postlaurea (Figura A.3) e di essere nella fascia di età compresa fra i 51 e i 65 anni. Per altro verso, all'aumentare del numero dei componenti della famiglia cresce il reddito disponibile familiare (Figura A.5). In questo ultimo caso, il divario del Sud e delle Isole rispetto al reddito delle famiglie residenti al Nord e al Centro si presenta in modo più accentuato.

Figura A.2 - Reddito medio familiare per ripartizione e condizione professionale della persona di riferimento – Anno 2002 (a) (b) (valori in euro)

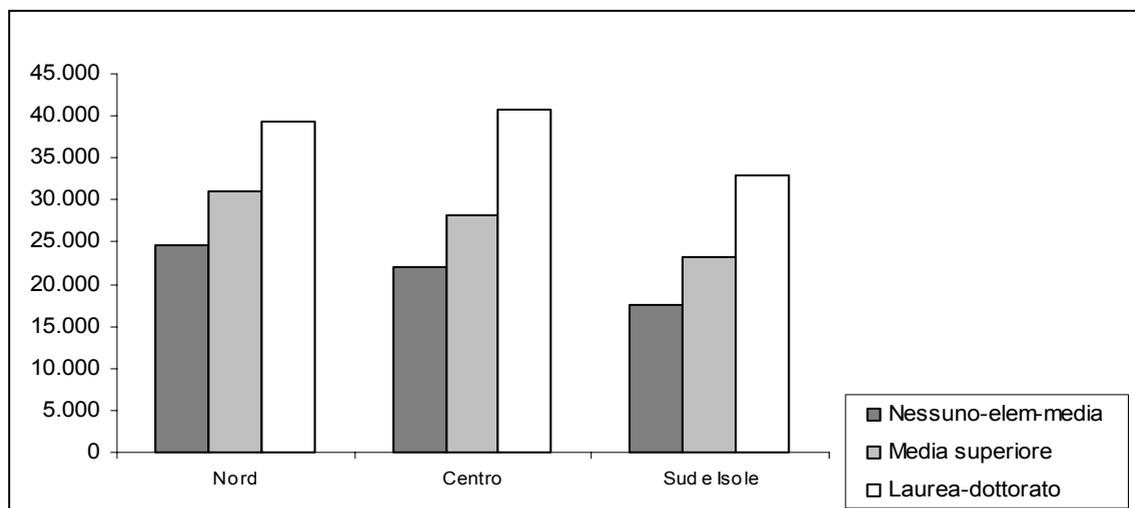


Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Reddito al netto dei fitti imputati.

(b) La persona di riferimento è l'intestatario del foglio di famiglia all'anagrafe.

Figura A.3 - Reddito medio familiare per ripartizione e titolo di studio della persona di riferimento Anno 2002 (a) (b) (valori in euro)

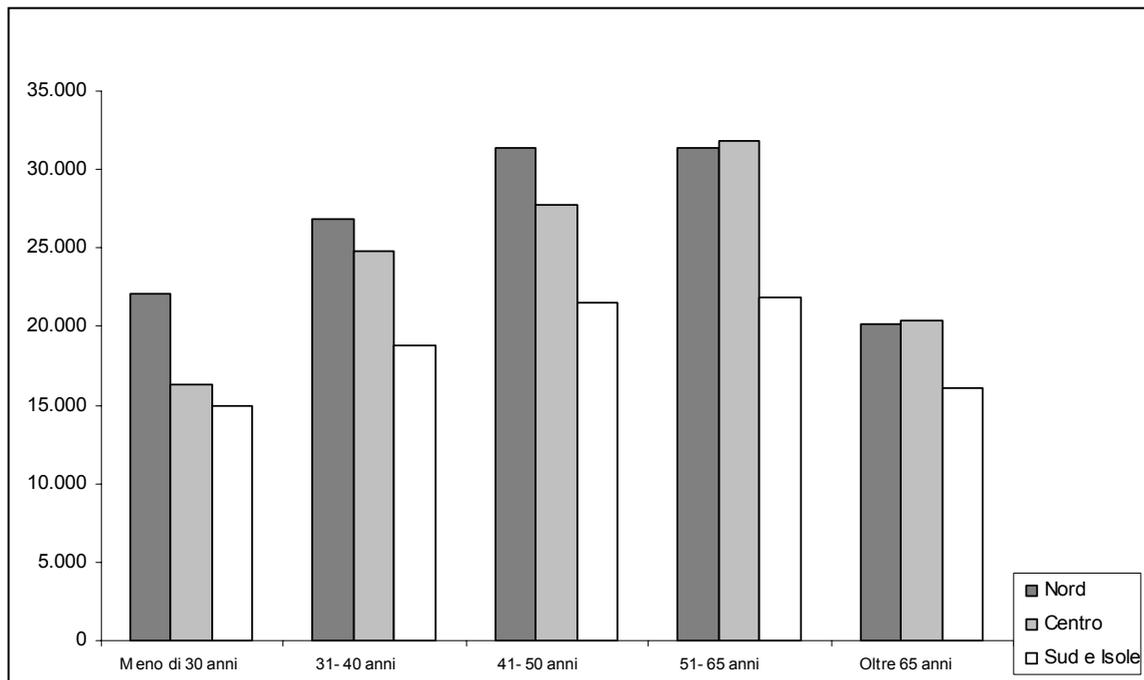


Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Reddito al netto dei fitti imputati.

(b) La persona di riferimento è l'intestatario del foglio di famiglia all'anagrafe.

Figura A.4 - Reddito medio familiare per età della persona di riferimento e per ripartizione – Anno 2002 (a) (b) (valori in euro)

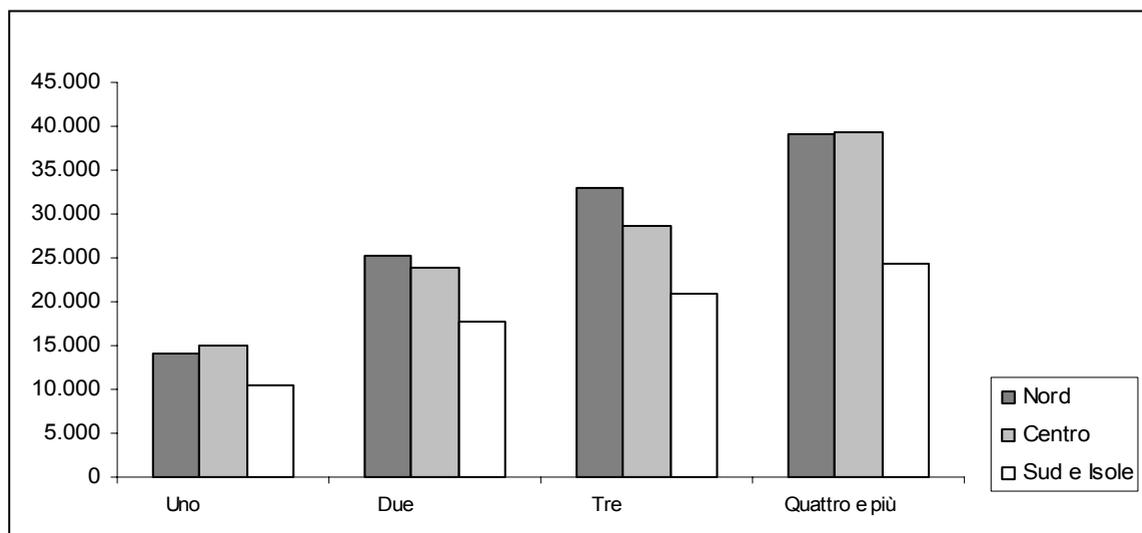


Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Reddito al netto dei fitti imputati.

(b) La persona di riferimento è l'instatario del foglio di famiglia all'anagrafe.

Figura A.5 - Reddito medio familiare per ripartizione e numero dei componenti della famiglia – Anno 2002 (a) (valori in euro)

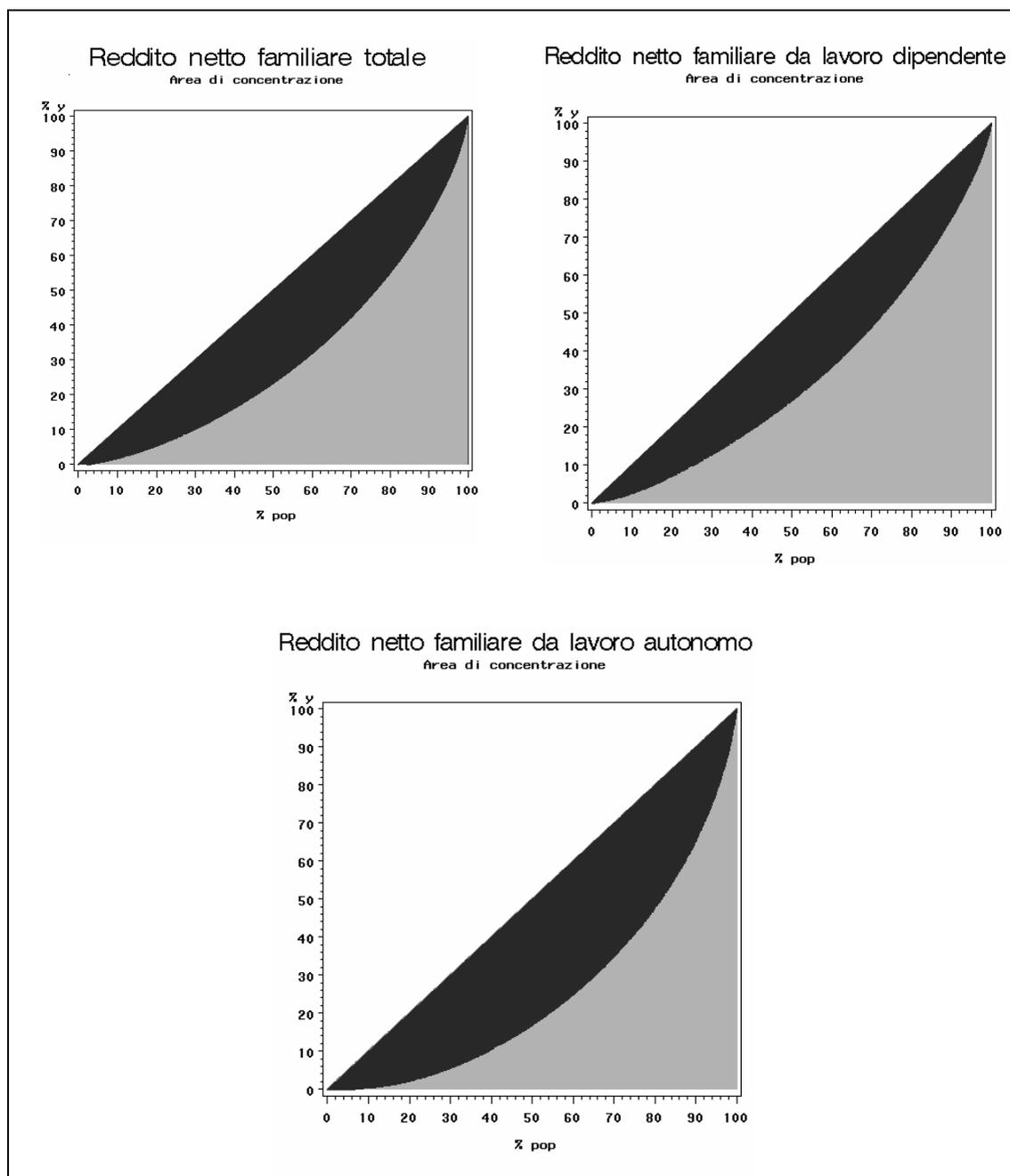


Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Reddito al netto dei fitti imputati.

La disuguaglianza dei redditi è rappresentata nella figura A.6 dove sono evidenziate le aree di concentrazione delle diverse fonti di reddito. L'area di concentrazione è compresa tra il segmento di equidistribuzione e la curva di concentrazione dei redditi (curva di Lorenz) ed è tanto più ampia quanto più diseguale è la distribuzione del reddito. L'asse orizzontale della Figura rappresenta la percentuale di famiglie ordinate in base al reddito crescente, mentre l'asse verticale indica la quota di reddito a loro attribuito. Se si considerano le diverse fonti di reddito familiare, appare evidente che il reddito netto familiare da attività finanziarie e da capitale reale presentino le distribuzioni più diseguali. Le distribuzioni dei redditi da lavoro dipendente e da pensione risultano, conformemente alle attese, meno diseguali.

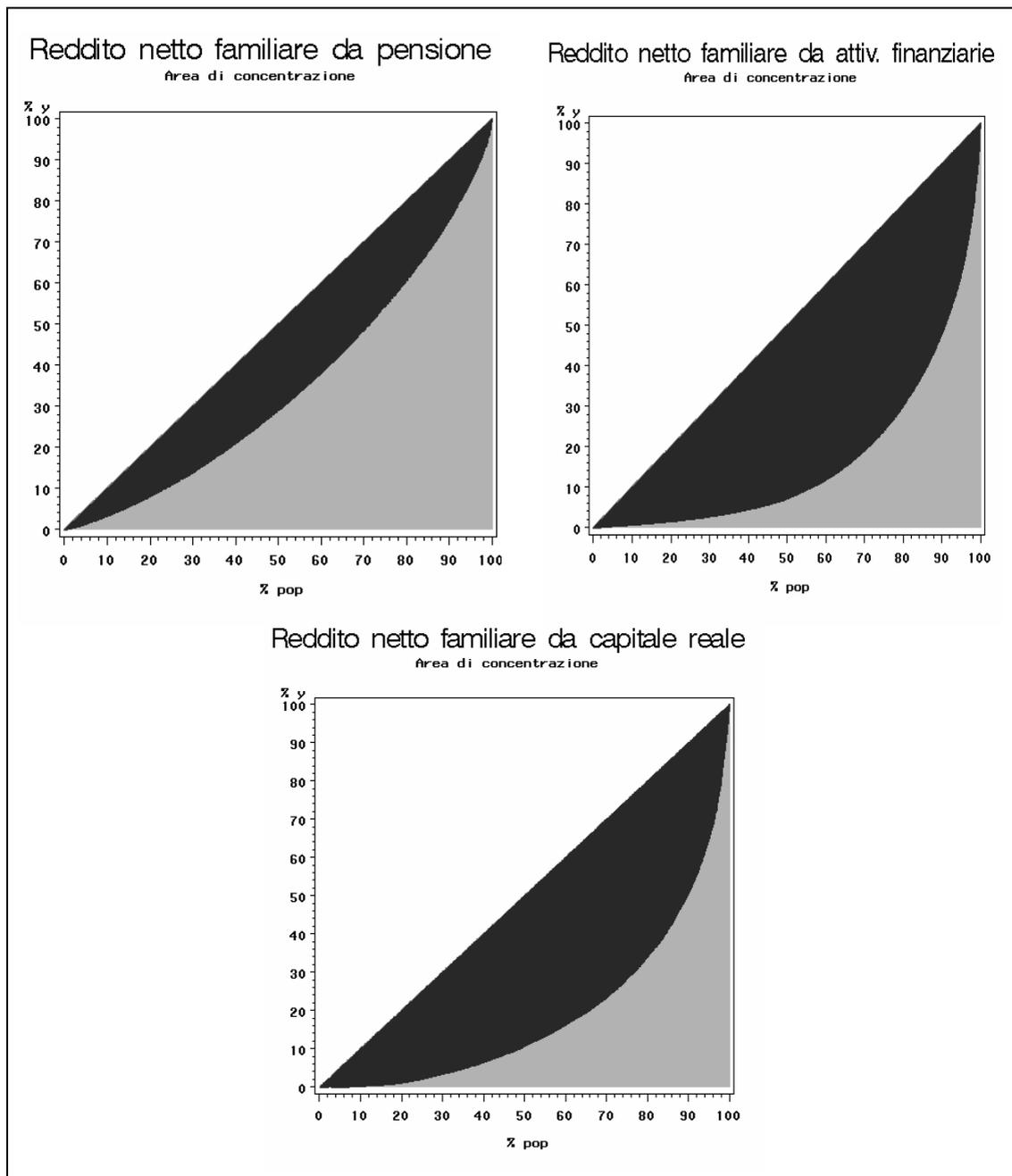
Figura A.6 - Area di concentrazione del reddito netto familiare per tipologia – Anno 2002 (a) (valori percentuali di popolazione e di reddito)



Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Reddito al netto dei fitti imputati

Figura A.6 segue - Area di concentrazione del reddito netto familiare per tipologia – Anno 2002 (a)
(valori percentuali di popolazione e di reddito)



Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Reddito al netto dei fitti imputati

A.1.2 Indicatori di disagio economico e di deprivazione materiale

L'indagine Icv2003 integra alla rilevazione del reddito numerose informazioni riguardanti aspetti fondamentali delle condizioni economiche delle famiglie, quali l'accesso al mercato del lavoro, lo stato di salute, le difficoltà nei pagamenti, la deprivazione materiale ed il disagio abitativo.³ I dati raccolti consentono, quindi, di approfondire aspetti cruciali della distribuzione del reddito, della povertà e dell'esclusione sociale attraverso l'analisi delle dimensioni monetarie e non monetarie delle condizioni di vita delle famiglie.

L'utilizzo congiunto delle misure monetarie e non monetarie rivela che le diverse forme di disagio e deprivazione tendono ad essere associate tra loro. Va sottolineato, tuttavia, che gli indicatori di disagio economico e di deprivazione materiale riflettono, in parte, la disponibilità di risorse finanziarie delle famiglie ed in parte le preferenze ed i modelli di consumo. Nell'analisi degli indicatori di deprivazione materiale va considerato un certo grado di reticenza delle famiglie ad ammettere una condizione di deprivazione dovuta a motivi economici. Numerosi fattori concorrono a formare le valutazioni soggettive dello stato di deprivazione da parte delle famiglie: fra questi il livello socio-economico, la percezione dello standard medio di vita con cui confrontarsi⁴ e non ultime le aspettative per l'immediato futuro. Le aspettative degli individui ricoprono, infatti, un ruolo non secondario se si considera che esse tendono ad aumentare con il livello del reddito e a diminuire con la povertà di lungo periodo. Alcuni studi hanno evidenziato che i poveri tendono ad accettare livelli di qualità della vita inferiori rispetto al resto della popolazione e quindi a sottovalutare il loro stato di deprivazione relativa.

La ricchezza di informazioni rilevate con l'indagine Icv2003 consente di confrontare indicatori non monetari di tipo oggettivo riguardanti l'indebitamento delle famiglie, il ritardo nei pagamenti e l'insufficienza di risorse per le necessità quotidiane con le valutazioni soggettive delle famiglie sulle difficoltà ad arrivare a fine mese e a sostenere il carico finanziario delle spese.

Analizzando i dati relativi agli arretrati nei pagamenti delle spese per la casa (Tavola A.4) risulta evidente che le maggiori difficoltà delle famiglie riguardano il pagamento dell'affitto (15,4 per cento) ed il pagamento delle bollette (8,8 per cento delle famiglie). Sono in particolare le famiglie residenti al Sud e nelle Isole a mostrare le percentuali in assoluto più elevate di arretrati per tutte le tipologie di pagamenti.

Le difficoltà delle famiglie ad affrontare le spese quotidiane sono riportate nella Tavola A.5. Le famiglie meno abbienti sperimentano difficoltà economiche soprattutto per affrontare eventuali imprevisti (14,5 per cento), per comprare i vestiti che necessitano (8,9 per cento), per pagare le tasse (8,7 per cento) ed anche per sostenere i costi in caso di malattia (8,2 per cento). Al Nord e al Centro le percentuali di famiglie in difficoltà risultano pressoché uguali per le diverse spese, mentre al Sud e nelle Isole ben il 26,5 per cento delle famiglie dichiara di non avere soldi sufficienti ad affrontare spese impreviste, il 17,2 per cento per pagare le tasse ed il 15,2 per cento per le malattie. Appare significativo il dato relativo al 2,3 per cento delle famiglie il cui stato di deprivazione è tale da non permettere loro di comprare il cibo. Queste famiglie costituiscono il 2,9 per cento delle famiglie residenti al Sud e nelle Isole e rappresenta il 2,2 per cento famiglie del Nord Italia.

³ I dati relativi alle condizioni di vita delle famiglie si riferiscono alla situazione rilevata al momento dell'intervista (anno 2003) ed i dati di reddito fanno riferimento all'anno 2002.

⁴ La definizione di uno standard di vita che possa essere considerato accettabile dalla maggiore parte della popolazione richiede un'analisi piuttosto complessa. Sarebbero necessarie, fra le altre, informazioni su ciò che è considerato un bene essenziale per poter costruire una gerarchia delle preferenze della popolazione. Per i beni durevoli, ad esempio, ciò viene approssimato dal grado di penetrazione del bene nella società.

Tavola A.4 - Famiglie che dichiarano, in alcuni periodi dell'anno, arretrati nei pagamenti per tipologia e ripartizione - Anno 2003 (per 100 famiglie)

PAGAMENTI	Nord	Centro	Sud e Isole	Italia
Bollette	4,3	7,8	16,1	8,8
Debiti diversi dal mutuo	1,9	3,7 (b)	8,0	4,3
Affitto (a)	10,6	15,7 (b)	22,4	15,4

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Per le sole famiglie che pagano l'affitto

(b) Stima corrispondente ad una numerosità campionaria compresa tra 20 e 49 unità.

Tavola A.5 - Famiglie che dichiarano, in alcuni periodi dell'anno, di non avere soldi per tipologia di spesa e ripartizione - Anno 2003 (per 100 famiglie)

SPESA	Nord	Centro	Sud e Isole	Italia
Cibo	2,2	..	2,9	2,3
Vestiti	6,7	5,7	14,1	8,9
Malattie	4,8	4,8	15,2	8,2
Scuola	1,4 (a)	1,9 (a)	5,6	2,9
Trasporti	2,3	2,1 (a)	7,0	3,8
Tasse	3,8	6,9	17,2	8,7
Imprevisti	8,7	8,7	26,5	14,5
Per almeno una delle precedenti spese	14,0	14,3	33,8	20,4

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Stima corrispondente ad una numerosità campionaria compresa tra 20 e 49 unità.

La percezione delle difficoltà e dei vincoli economici da parte delle famiglie nelle diverse ripartizioni è sintetizzata dalla tavola A.6. La percentuale di famiglie che dichiara di avere debiti diversi dal mutuo è più elevata al Nord (14,8 per cento), dove vi sono le famiglie con redditi medi più elevati e per le quali è più facile accedere al credito al consumo rispetto alle altre ripartizioni. Tuttavia al Sud e nelle Isole è presente una percentuale elevata di famiglie (29,6 per cento) che considera il pagamento dei debiti diversi dal mutuo un carico molto pesante. Allo stesso modo, le famiglie che al Sud e nelle Isole considerano molto difficile arrivare a fine mese hanno una percentuale quasi doppia (18,5 per cento) di quelle residenti al Nord (9,7 per cento). Per altro verso la percentuale di famiglie del Nord che dichiarano di arrivare con facilità a fine mese rappresenta il 12,6 per cento rispetto al 5,8 per cento delle famiglie del Sud e delle Isole. Anche il pagamento delle spese per la casa diverse dall'affitto è considerato un carico molto pesante dal 25,5 per cento delle famiglie residenti al Sud e nelle Isole rispetto al 21,9 per cento della media nazionale.

Tavola A.6 - Famiglie per ripartizione e giudizio sulla situazione economica familiare – Anno 2003
(per 100 famiglie)

SITUAZIONE ECONOMICA	Nord	Centro	Sud e Isole	Italia
Famiglie con debiti per la casa diversi dal mutuo	14,8	12,7	13,4	13,9
Carico dei debiti per la casa diversi dal mutuo				
Molto pesante	16,6	12,0 (a)	29,6	19,7
Abbastanza pesante	40,3	49,0	47,4	44,1
Poco, per niente pesante	43,1	39,1	23,0	36,2
Come arriva a fine mese				
Molto difficile	9,7	11,0	18,5	12,8
Difficile	15,9	17,4	22,0	18,2
Abbastanza difficile	27,0	27,4	27,5	27,2
Non troppo difficile	34,7	33,0	26,2	31,6
Facile, molto facile	12,6	11,2	5,8	10,1
Carico delle spese diverse da affitto				
Molto pesante	20,8	19,0	25,5	21,9
Abbastanza pesante	58,1	56,9	57,9	57,8
Poco pesante	18,4	19,7	15,2	17,7
Per niente pesante	2,7	4,4	1,4 (a)	2,6

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Stima corrispondente ad una numerosità campionaria compresa tra 20 e 49 unità.

L'inadeguatezza delle condizioni abitative rappresenta un aspetto importante della povertà e dell'esclusione sociale, così come riconosciuto al Consiglio Europeo di Laeken nel dicembre 2001, in occasione del quale è stata adottata una lista comune di indicatori per l'analisi della povertà. Conoscere, quindi, quante famiglie vivono in condizioni di disagio per i problemi inerenti l'alloggio o la zona di abitazione è ritenuto di fondamentale importanza per le politiche finalizzate all'individuazione dei gruppi a rischio di povertà. A tal fine, l'indagine Eu-Silc 2007 sarà dedicata all'approfondimento di alcuni aspetti cruciali del disagio abitativo, mediante l'inserimento di un modulo specifico di variabili.

L'indagine pilota sulle condizioni di vita del 2003 fornisce già un insieme di indicatori che consentono di quantificare il numero delle famiglie che dichiarano di avere problemi nell'abitazione e fra questi, l'umidità rappresenta il problema più frequente coinvolgendo il 19,4 per cento delle famiglie residenti (Tavola A.7). L'abitazione distante dai familiari, e quindi dalla rete familiare di sostegno, è considerato un problema dal 17 per cento delle famiglie, mentre al Sud e nelle Isole la percentuale raggiunge il 21,2 per cento del totale. L'abitazione è considerata, inoltre, distante dal lavoro, con ripercussioni sui tempi degli spostamenti quotidiani, dal 14,1 per cento delle famiglie.

La precarietà delle condizioni abitative è, tuttavia, particolarmente evidente per le famiglie che si trovano in affitto (Tavola A.8). Il 26,5 per cento di queste famiglie lamenta il problema della distanza dai familiari ed il 24,7 per cento l'umidità dell'abitazione. Inoltre più del 16 per cento delle famiglie dichiara di avere un'abitazione scarsamente illuminata, oppure distante dal lavoro o con tetti, soffitti, porte, finestre o pavimenti danneggiati (18,5 per cento delle famiglie).

Tavola A.7 - Famiglie che dichiarano problemi inerenti l'abitazione per tipologia e ripartizione - Anno 2003 (per 100 famiglie)

PROBLEMI	Nord	Centro	Sud e Isole	Italia
Tetti, soffitti, porte, finestre, pavimenti danneggiati	11,8	8,3	15,4	12,3
Umidità	18,5	14,8	23,5	19,4
Scarsa illuminazione	8,5	8,9	17,7	11,6
Abitazione distante da Familiari	15,0	15,0	21,2	17,0
Abitazione distante da lavoro	12,1	11,2	18,9	14,1

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

Tavola A.8 - Famiglie di affittuari che dichiarano problemi inerenti l'abitazione per tipologia e ripartizione - Anno 2003 (per 100 famiglie)

PROBLEMI	Nord	Centro	Sud e Isole	Italia
Tetti, soffitti, porte, finestre, pavimenti danneggiati	17,7	13,3 (a)	22,3	18,5
Umidità	23,0	26,0	26,4	24,7
Scarsa illuminazione	12,2	15,1 (a)	23,2	16,4
Abitazione distante da Familiari	26,0	28,6	26,1	26,5
Abitazione distante da lavoro	15,6	14,4 (a)	19,7	16,8

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Stima corrispondente ad una numerosità campionaria compresa tra 20 e 49 unità.

Considerando i problemi che riguardano la zona di abitazione (Tavola A.9), il 9,6 per cento delle famiglie considera il rumore molto presente: al Nord la percentuale raggiunge l'11,3 per cento e al Centro l'8,6 per cento. Anche l'inquinamento atmosferico è considerato molto presente dall'8,4 per cento delle famiglie e dall'11,6 per cento delle famiglie che vivono al Nord. La criminalità è percepita come un problema molto presente dal 4,1 per cento delle famiglie con una percentuale leggermente superiore al Sud e nelle Isole (4,6 per cento). Anche in questo caso, le famiglie che vivono in affitto lamentano una situazione di maggiore problematicità, con percentuali più elevate rispetto al resto della popolazione (Tavola A.10).

Tavola A.9 - Famiglie che dichiarano la presenza di problemi nella zona di abitazione per tipologia e ripartizione – Anno 2003 (per 100 famiglie)

PROBLEMI	Nord	Centro	Sud e Isole	Italia
Rumore				
Molto	11,3	8,6	7,8	9,6
Abbastanza	20,1	23,6	23,6	21,9
Poco	28,0	27,4	29,7	28,4
Per niente, non so	40,7	40,3	38,9	40,1
Inquinamento atmosferico				
Molto	11,6	7,3	4,5	8,4
Abbastanza	24,9	19,1	17,5	21,3
Poco	30,8	28,0	32,0	30,6
Per niente, non so	32,8	45,6	46,0	39,7
Criminalità				
Molto	3,9	3,4 (a)	4,6	4,1
Abbastanza	12,5	15,9	14,2	13,8
Poco	31,5	30,6	30,9	31,1
Per niente, non so	52,1	50,1	50,3	51,1

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Stima corrispondente ad una numerosità campionaria compresa tra 20 e 49 unità.

Tavola A.10 - Famiglie di affittuari che dichiarano la presenza di problemi nella zona di abitazione per tipologia e ripartizione – Anno 2003 (per 100 famiglie)

PROBLEMI	Nord	Centro	Sud e Isole	Italia
Rumore				
Molto	15,5	12,3 (a)	13,8	14,4
Abbastanza	26,2	31,9	31,4	28,9
Poco	28,6	32,7	30,6	30,0
Per niente, non so	29,7	23,1	24,2	26,7
Inquinamento atmosferico				
Molto	12,3	..	8,0 (a)	10,5
Abbastanza	27,1	24,3 (a)	27,6	26,8
Poco	32,1	30,4	34,2	32,5
Per niente, non so	28,6	35,2	30,2	30,3
Criminalità				
Molto	7,3 (a)	..	10,2	8,1
Abbastanza	14,6	27,6 (a)	21,5	19,2
Poco	31,4	29,3	32,5	31,4
Per niente, non so	46,7	36,7	35,9	41,3

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Stima corrispondente ad una numerosità campionaria compresa tra 20 e 49 unità.

Il possesso di beni durevoli da parte delle famiglie rappresenta uno degli indicatori di deprivazione materiale più utilizzati. Per scindere la componente delle preferenze dei

consumatori dalla indisponibilità dei beni causata dalla mancanza di risorse finanziarie, l'indagine Icv2003 chiede alle famiglie di indicare se la famiglia non può permettersi l'acquisto di determinati beni (Tavola A.11). È da sottolineare che la percentuale delle famiglie che dichiarano di non potersi permettere determinati beni, e che quindi dovrebbero rappresentare le famiglie più povere, è sempre inferiore alle famiglie che dichiarano di non possedere i beni per "altro motivo". Le percentuali più elevate di famiglie che dichiarano i motivi economici quale fattore dell'indisponibilità dei beni durevoli riguardano la lavastoviglie (10,4 per cento), la videocamera (10,3 per cento) e l'antenna parabolica (10,1 per cento). Il 7,0 per cento delle famiglie dichiara, inoltre, di non potersi permettere l'acquisto della moto, il 6,9 per cento del personal computer ed il 6,7 per cento delle famiglie di non poter accedere ad Internet.

Tavola A.11 - Famiglie che dichiarano di non possedere beni durevoli per motivo e tipologia- Anno 2003 (a) (per 100 famiglie)

BENI DUREVOLI	Nord	Centro	Sud e Isole	Italia
Lavastoviglie				
Non se lo possono permettere	6,9	6,7	17,9	10,4
Altro motivo	52,9	51,5	63,0	55,9
Videoregistratore				
Non se lo possono permettere	3,0	3,5 (b)	9,0	5,0
Altro motivo	30,8	31,1	34,0	31,9
Videocamera				
Non se lo possono permettere	7,1	7,2	16,8	10,3
Altro motivo	73,2	71,0	65,1	70,1
Impianto stereo				
Non se lo possono permettere	4,3	3,5 (b)	11,1	6,3
Altro motivo	45,4	43,9	45,1	45,0
Antenna parabolica				
Non se lo possono permettere	7,1	6,9	16,5	10,1
Altro motivo	72,1	70,7	64,4	69,4
Consolle, videogiochi				
Non se lo possono permettere	3,6	4,0 (b)	12,0	6,4
Altro motivo	80,7	78,0	73,7	77,9
Personal computer				
Non se lo possono permettere	5,1	4,1 (b)	11,2	6,9
Altro motivo	54,7	55,1	55,6	55,1
Modem				
Non se lo possono permettere	3,8	4,8	11,8	6,6
Altro motivo	66,7	65,7	67,8	66,9
Accesso ad Internet				
Non se lo possono permettere	4,0	5,1	11,6	6,7
Altro motivo	65,0	65,2	65,8	65,3
Segreteria telefonica				
Non se lo possono permettere	3,4	4,3 (b)	11,9	6,3
Altro motivo	80,3	83,5	82,9	81,8
Fax				
Non se lo possono permettere	3,2	3,5 (b)	11,5	5,9
Altro motivo	88,9	88,3	84,5	87,4

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Sono esclusi beni ad ampia diffusione come la lavatrice e il televisore a colori, in quanto posseduti da oltre il 90 per cento delle famiglie.

(b) Stima corrispondente ad una numerosità campionaria compresa tra 20 e 49 unità.

**Tavola A.11 segue - Famiglie che dichiarano di non possedere beni durevoli per motivo e tipologia
Anno 2003 (a) (per 100 famiglie)**

BENI DUREVOLI	Nord	Centro	Sud e Isole	Italia
Telefono cellulare				
Non se lo possono permettere	1,9	1,7 (b)	5,5	3,0
Altro motivo	21,0	17,7	24,6	21,5
Bicicletta				
Non se lo possono permettere	1,2 (b)	..	5,8	2,8
Altro motivo	27,2	51,7	58,8	42,4
Motorino				
Non se lo possono permettere	4,0	3,4 (b)	9,5	5,7
Altro motivo	76,9	71,9	71,3	74,1
Moto				
Non se lo possono permettere	4,9	4,5	11,8	7,0
Altro motivo	86,4	88,8	83,5	86,0
Automobile				
Non se lo possono permettere	2,7	2,3 (b)	5,2	3,4
Altro motivo	19,2	20,5	22,2	20,4

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Sono esclusi beni ad ampia diffusione come la lavatrice e il televisore a colori, in quanto posseduti da oltre il 90 per cento delle famiglie.

(b) Stima corrispondente ad una numerosità campionaria compresa tra 20 e 49 unità.

L'utilizzo congiunto dei dati di reddito e degli indicatori di disagio economico e di deprivazione materiale conferma che la mancanza di un adeguato livello di risorse finanziarie si riflette nella deprivazione materiale, nell'inadeguatezza delle condizioni abitative e nella percezione di vincoli economici stringenti da parte delle famiglie. Le tavole seguenti dimostrano chiaramente che le difficoltà risultano concentrate in particolare nel primo e nel secondo quinto di reddito familiare medio equivalente⁵, ossia tra le famiglie con i redditi più bassi. Ciò è evidente nel caso delle famiglie che dichiarano di avere arretrati nei pagamenti (Tavola A.12), in modo particolare per l'affitto (il 25,4 per cento delle famiglie del primo quinto rispetto al 15,4 per cento della media nazionale), e mancanza di risorse per alcune spese essenziali (Tavola A.13). La percentuale di famiglie che dichiarano di avere arretrati per almeno una delle tipologie di spesa considerata rappresenta il 20,4 per cento del totale e raggiunge il 33,6 per cento nel primo quinto.

⁵ Per confrontare le condizioni economiche delle famiglie con diversa numerosità è stata utilizzata la scala di equivalenza dell'Ocse modificata (primo adulto= 1,0; ogni altro adulto= 0,5; ogni bambino con età inferiore a 14 anni = 0,3).

Tavola A.12 - Famiglie che dichiarano, in alcuni periodi dell'anno, arretrati nei pagamenti per tipologia di pagamenti e per quinti di reddito familiare equivalente - Anno 2002 (a) (per 100 famiglie)

PAGAMENTI	Primo	Secondo	Terzo	Quarto	Quinto	Italia
Bollette	17,7	10,3	7,3	4,9	4,0	8,8
Debiti diversi da mutuo	7,2	4,3	4,5	3,1	2,1 (c)	4,3
Affitto (b)	25,4	17,4	10,4 (c)	9,5 (c)	..	15,4

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Reddito al netto dei fitti imputati

(b) Per le sole famiglie che pagano l'affitto

(c) Stima corrispondente ad una numerosità campionaria compresa tra 20 e 49 unità.

Tavola A.13 - Famiglie che dichiarano, in alcuni periodi dell'anno, di non avere soldi per tipologia di spesa e per quinti di reddito familiare equivalente - Anno 2002 (a) (per 100 famiglie)

SPESA	Primo	Secondo	Terzo	Quarto	Quinto	Italia
Cibo	5,1	3,3 (b)	1,4 (b)	2,3
Vestiti	16,7	10,6	7,5	6,4	3,4	8,9
Malattie	17,0	10,4	5,9	5,3	2,1 (b)	8,2
Scuola	4,3	3,4 (b)	3,1	2,1 (b)	1,5 (b)	2,9
Trasporti	7,9	4,3	3,1	2,5 (b)	..	3,8
Tasse	16,0	11,0	7,7	5,2	3,7	8,7
Imprevisti	24,5	18,1	12,9	10,7	6,1	14,5
Per almeno una delle precedenti spese	33,6	25,5	18,6	15,0	9,3	20,4

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Reddito al netto dei fitti imputati.

(b) Stima corrispondente ad una numerosità campionaria compresa tra 20 e 49 unità.

La percezione delle difficoltà economiche risulta quindi più elevata nelle famiglie del primo e del secondo quinto (Tavola A.14), dove però in media sussiste una percentuale più bassa di famiglie che dichiarano di avere debiti per la casa diversi dal mutuo, trattandosi presumibilmente di famiglie che hanno difficoltà ad accedere al credito al consumo (il 7,5 per cento delle famiglie del primo quinto rispetto al 13,9 per cento a livello nazionale). Le maggiori difficoltà ad arrivare a fine mese, a sopportare i debiti contratti per l'abitazione diversi dal mutuo e a sostenere le spese familiari diverse dall'affitto, sono più rilevanti per le famiglie del primo e del secondo quinto. Il 23,2 per cento delle famiglie del primo quinto e il 16,1 per cento delle famiglie del secondo quinto sostengono che arrivare a fine mese è per loro molto difficile, mentre soltanto il 4,4 per cento delle famiglie del primo quinto dichiara che le proprie disponibilità finanziarie consentono di arrivare facilmente o molto facilmente a fine mese. Allo stesso modo, il 31,3 per cento delle famiglie appartenenti al quinto di reddito più basso considera molto pesante sostenere le spese per la casa diverse dall'affitto rispetto al 21,9 per cento della media nazionale.

Tavola A.14 - Famiglie per giudizio sulla situazione economica per quinti di reddito familiare equivalente - Anno 2002 (a) (per 100 famiglie)

SITUAZIONE ECONOMICA	Primo	Secondo	Terzo	Quarto	Quinto	Totale
Famiglie con debiti per la casa diversi dal mutuo	7,5	11,0	17,0	16,5	17,8	13,9
Carico dei debiti per la casa diversi dal mutuo						
Molto, abbastanza pesante	79,4	76,2	61,1	65,2	51,1	63,9
Poco, per niente pesante	20,6 (b)	23,8 (b)	38,9	34,8	48,9	36,2
Come arriva a fine mese						
Molto difficile	23,2	16,1	12,2	8,2	4,2	12,8
Difficile	22,3	23,7	17,2	19,8	8,0	18,2
Abbastanza difficile	29,5	30,6	29,8	25,4	20,8	27,2
Non troppo difficile	20,5	25,6	32,9	36,3	42,8	31,6
Facile, molto facile	4,4	4,0	7,9	10,2	24,1	10,1
Carico delle spese diverse dall'affitto						
Molto pesante	31,3	25,8	20,2	19,4	13,0	21,9
Abbastanza pesante	54,0	58,5	59,0	60,1	57,5	57,8
Poco, per niente pesante	14,7	15,7	20,8	20,6	29,5	20,3

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Reddito al netto dei fitti imputati.

(b) Stima corrispondente ad una numerosità campionaria compresa tra 20 e 49 unità.

Allo stesso modo, le percentuali più elevate di famiglie (Tavola A.15) che dichiarano di avere umidità nelle abitazioni (25,3 per cento), alloggio distante dai familiari (22 per cento), scarsa illuminazione (18,1 per cento) e cattive condizioni abitative, con soffitti, porte, finestre o pavimenti danneggiati (17 per cento) rappresentano le famiglie più povere del primo quinto.

Tavola A.15 - Problemi relativi all'abitazione per tipologia e per quinti di reddito familiare equivalente - Anno 2002 (a) (per 100 famiglie)

PROBLEMI	Primo	Secondo	Terzo	Quarto	Quinto	Italia
Tetti, soffitti, porte, finestre, pavimenti danneggiati	17,0	13,9	11,8	10,0	8,6	12,3
Umidità	25,3	21,1	19,8	17,7	13,0	19,4
Scarsa illuminazione	18,1	13,9	9,7	8,9	7,2	11,6
Abitazione distante da familiari	22,0	19,9	15,9	14,4	12,8	17,0
Abitazione distante da lavoro	8,5	12,8	17,1	17,0	15,2	14,1

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Reddito al netto dei fitti imputati.

Per quanto riguarda la percezione dei problemi ambientali e sociali presenti nella zona di abitazione (rumore, inquinamento atmosferico, criminalità) è evidente che il disagio delle famiglie per questi aspetti (Tavola A.16) non è associato al livello del reddito, in quanto le famiglie dei quinti più bassi presentano percentuali non molto diverse dagli altri quinti per quasi tutte le modalità considerate.

Tavola A.16 - Famiglie che dichiarano la presenza di problemi nella zona di abitazione per tipologia e per quinti di reddito familiare equivalente - Anno 2002 (a) (per 100 famiglie)

PROBLEMI	Primo	Secondo	Terzo	Quarto	Quinto	Italia
Rumore						
Molto	10,2	8,2	9,2	10,3	10,2	9,6
Abbastanza	21,3	25,1	22,6	21,4	19,3	21,9
Poco	29,5	28,9	28,6	27,4	27,7	28,4
Per niente, non so	39,1	37,9	39,6	40,9	42,8	40,1
Inquinamento atmosferico						
Molto	7,9	6,9	9,0	9,3	9,0	8,4
Abbastanza	18,2	21,8	20,8	20,9	24,9	21,3
Poco	30,6	28,2	30,9	31,3	32,0	30,6
Per niente, non so	43,3	43,1	39,3	38,5	34,1	39,7
Criminalità						
Molto	5,9	3,6	4,3	3,1 (b)	3,4 (b)	4,1
Abbastanza	13,2	15,4	12,8	14,7	12,5	13,8
Poco	30,4	30,2	31,6	31,2	32,1	31,1
Per niente, non so	50,5	50,7	51,3	50,9	52,0	51,1

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Reddito al netto dei fitti imputati.

(b) Stima corrispondente ad una numerosità campionaria compresa tra 20 e 49 unità.

Il possesso dei beni durevoli risulta, invece, chiaramente dipendente dal livello del reddito e le relative informazioni rappresentano in qualche misura una gerarchia delle preferenze dei beni da parte delle famiglie (Tavola A.17). Conformemente alle attese, sono soprattutto le famiglie del primo e del secondo quinto a dichiarare di non potersi permettere alcuni beni per motivi economici, in particolare per quanto riguarda la lavastoviglie (rispettivamente 19,3 e 13 per cento), la videocamera (rispettivamente 17 e 12,8 per cento) e l'antenna parabolica (rispettivamente 16,3 e 11,9 per cento). Allo stesso modo risulta molto bassa la percentuale delle famiglie più ricche dell'ultimo quinto che motivano l'indisponibilità dei beni con la carenza di risorse.

Tavola A.17 - Famiglie che dichiarano di non possedere beni durevoli (a) perché non possono permetterselo, per tipologia di bene e per quinti di reddito familiare equivalente – Anno 2002 (b) (per 100 famiglie)

BENI DUREVOLI	Primo	Secondo	Terzo	Quarto	Quinto	Italia
Lavastoviglie	19,3	13,0	10,4	6,3	3,0 (c)	10,4
Videoregistratore	12,0	6,6	3,7	1,7 (c)	..	5,0
Videocamera	17,0	12,8	9,9	7,2	4,4	10,3
Impianto stereo	12,9	8,1	5,1	3,1	2,4 (c)	6,3
Antenna parabolica	16,3	11,9	9,9	7,3	4,9	10,1
Console, videogiochi	11,6	8,0	5,8	4,0	2,6 (c)	6,4
Personal computer	12,6	8,6	6,5	4,6	2,1 (c)	6,9
Modem	11,9	8,2	5,7	4,5	2,7 (c)	6,6
Accesso ad Internet	11,9	8,5	5,7	4,8	2,5 (c)	6,7
Segreteria telefonica	11,2	7,7	5,9	4,1	2,7 (c)	6,3
Fax	10,4	7,6	5,5	3,7	2,4 (c)	5,9
Telefono fisso	9,7	5,4	2,8 (c)	2,3 (c)	..	4,3
Telefono cellulare	7,9	3,6	2,3 (c)	3,0
Bicicletta	6,5	3,0 (c)	2,3 (c)	1,3 (c)	..	2,8
Motorino	10,2	6,7	5,6	3,3	2,5 (c)	5,7
Moto	10,8	8,2	7,3	4,8	4,1	7,0
Automobile	9,6	4,1	1,9 (c)	3,4

Fonte: Istat, Indagine sulle condizioni di vita - Anno 2003

(a) Sono esclusi beni ad ampia diffusione come la lavatrice e il televisore a colori, in quanto posseduti da oltre il 96 per cento delle famiglie.

(b) Reddito al netto dei fitti imputati.

(c) Stima corrispondente ad una numerosità campionaria compresa tra 20 e 49 unità.

A.1.3 - Alcune note conclusive

I risultati dell'Indagine pilota sulle condizioni di vita Anno 2003 consentono di analizzare le condizioni economiche e sociali delle famiglie con un approccio multidimensionale, in cui l'indicatore di reddito disponibile delle famiglie è affiancato da un insieme di indicatori non monetari riguardanti gli aspetti più rilevanti della povertà e dell'esclusione sociale. La scarsità di risorse finanziarie è, infatti, spesso associata a situazioni di svantaggio come l'insufficiente scolarizzazione, la disoccupazione, il disagio abitativo e la grave deprivazione materiale. L'indagine Icv2003, mediante un insieme di indicatori di tipo oggettivo e soggettivo, evidenzia i segnali di disagio economico nelle condizioni di vita delle famiglie, come ad esempio la capacità o meno di disporre di beni essenziali, a cui è possibile associare la percezione soggettiva delle difficoltà ad arrivare a fine mese e a sostenere il carico finanziario per le necessità quotidiane e per l'abitazione.

A.2 - Disuguaglianza dei redditi e indicatori di povertà: confronti con l'indagine sui bilanci delle famiglie della Banca d'Italia

A.2.1 - Introduzione

Nell'ambito dell'analisi della situazione economica e sociale di un Paese, è fondamentale non solo focalizzare l'attenzione sui livelli di benessere, ma anche sugli aspetti di equità e disuguaglianza nella distribuzione dei redditi. Le scelte politiche in materia fiscale o di riforma

dello stato sociale esercitano un significativo impatto redistributivo; per tali ragioni, l'analisi della distribuzione del reddito tra le famiglie e gli individui e della diffusione della povertà rappresentano strumenti indispensabili per indirizzare gli interventi di politica sociale ed economica.

Le analisi di seguito presentate hanno lo scopo di mettere in luce l'entità della disuguaglianza nella distribuzione dei redditi totali, così come emerge dall'Indagine sulle condizioni di vita dell'anno 2003 (Icv2003), che ha avuto tra i suoi principali obiettivi la rilevazione del reddito percepito dagli individui e dalle famiglie residenti in Italia. Nello stesso tempo viene realizzato un confronto con quanto risulta dall'indagine sui bilanci delle famiglie della Banca d'Italia (Bi) in relazione alla distribuzione dei redditi.

Nell'indagine sulle condizioni di vita dell'anno 2003, il reddito totale familiare è determinato come somma dei redditi percepiti da tutti i componenti della famiglia con almeno 15 anni e di alcune componenti percepite dall'intera famiglia (es. assegno al terzo figlio, minimo vitale, redditi derivanti dall'affitto di parti dell'abitazione eccetera). Il reddito complessivo di ciascun individuo comprende redditi da lavoro dipendente (salari e stipendi, compensi accessori, arretrati e liquidazioni), redditi da lavoro autonomo,⁶ redditi da pensioni (pensioni da lavoro, sociali, di reversibilità, di invalidità e integrative), altri assegni sociali (es. sussidi di disoccupazione, cassa integrazione guadagni, borse di studio), redditi derivanti dall'investimento in attività finanziarie o dall'affitto di abitazione secondarie, trasferimenti ricevuti regolarmente da altre famiglie (es. assegno di mantenimento). Sono sottratti eventuali trasferimenti dati regolarmente ad altre famiglie.

L'anno di riferimento per le informazioni di reddito è il 2002, che è l'anno a cui si riferisce anche l'indagine sui bilanci delle famiglie della Banca d'Italia.

Il concetto di reddito utilizzato nelle analisi che seguono è meramente monetario, sono escluse quindi eventuali integrazioni in natura quali i buoni pasto o l'uso per fini personali dell'auto aziendale, così come sono escluse componenti figurative di reddito (ad es. fitti figurativi per le abitazioni di proprietà). Il reddito rilevato dall'indagine della Banca d'Italia è stato depurato anch'esso dai fitti figurativi, al fine di rendere più simili le definizioni di reddito delle due indagini.

Nei paragrafi che seguono si esamina dapprima l'entità della disuguaglianza nella distribuzione del reddito familiare; segue quella nella distribuzione del reddito equivalente tra gli individui ed infine si focalizza l'attenzione sulla coda bassa della distribuzione dei redditi equivalenti, analizzando alcuni indicatori relativi alla diffusione e l'intensità della povertà.

A.2.2 - Reddito familiare: risultati e confronti

Dai dati dell'indagine sulle condizioni di vita dell'anno 2003, il reddito medio netto delle famiglie italiane risulta di 23.789 euro annuali⁷ nel 2002 (Tavola A.18).

Per le famiglie del Nord, il reddito medio risulta dell'11 per cento più elevato rispetto alla media nazionale (26.448 euro), mentre nel Sud e nelle Isole risulta di circa il 20 per cento inferiore (19.084 euro).

⁶ Per comparabilità con l'indagine della Bi, nei confronti non sono considerati i redditi negativi da lavoro autonomo, che non sono presenti nel campione di Bi.

⁷ In tale valore sono inclusi anche i redditi negativi (che riguardano lo 0,04 per cento delle famiglie) e i redditi nulli (che interessano 1,6 per cento delle famiglie)

Tavola A.18 - Reddito familiare medio annuale netto e percentili di reddito familiare annuale netto per ripartizione – Anno 2002 (valori in euro)

	Icv2003			
	Nord	Centro	Sud ed Isole	Italia
MEDIA	26.448	25.052	19.084	23.789
PERCENTILE				
5 °	5.954	4.800	2.435	4.563
10 °	8.450	6.890	5.500	6.825
20 °	11.732	10.308	8.060	10.166
30 °	14.808	13.650	10481	12.935
40 °	18.001	16.328	12.700	15.647
50 °	21.300	19.018	15.200	18.650
60 °	25.578	23.090	17.720	22.500
70 °	30.040	28.500	21.960	27.192
80 °	37.372	35.698	27.311	33.800
90 °	49.101	47.610	36.465	45.144
95 °	61.961	62.372	48.043	57.100

	Bi			
	Nord	Centro	Sud ed Isole	Italia
MEDIA	26.760	23.678	16.938	22.853
PERCENTILE				
5 °	7.146	6.643	3.469	5.550
10 °	9.719	8.237	5.460	7.146
20 °	12.600	10.993	7.355	10.191
30 °	15.793	13.695	9.750	12.899
40 °	19.045	16.073	11.700	15.256
50 °	22.750	19.175	13.601	18.449
60 °	26.236	23.064	16.000	22.418
70 °	31.006	26.980	19.660	26.433
80 °	36.860	32.470	24.563	32.366
90 °	46.632	42.188	32.895	42.381
95 °	59.237	54.675	41.876	53.136

Fonte: Istat, Indagine sulle Condizioni di vita - Anno 2003. Elaborazioni su dati Banca d'Italia, 'I Bilanci delle Famiglie Italiane nel 2002'

Sulla base dei dati della Banca d'Italia il reddito medio familiare risulta inferiore del 4 per cento (22.853 euro) rispetto ai risultati Icv2003. Nei dati di Bi, le differenze territoriali sono più marcate rispetto a quanto osservato dall'Icv2003: nelle regioni meridionali il reddito medio familiare è inferiore del 26 per cento rispetto al valore nazionale (16.938 euro) e nelle regioni settentrionali risulta superiore del 17 per cento (26.760 euro).

Considerando la distribuzione dei redditi nel suo insieme, risulta utile esaminare in prima istanza i livelli dei decili di reddito nelle due indagini. Il valore mediano risulta inferiore al valore medio in entrambe le indagini: in altri termini la distribuzione dei redditi è caratterizzata da un'asimmetria che vede una maggiore concentrazione delle famiglie nella parte sinistra della distribuzione. Sulla base dei dati dell'Icv2003, il 50 per cento delle famiglie italiane percepisce un reddito non superiore a 18.650 euro. Secondo i dati Bi, il valore mediano è lievemente inferiore (18.449 euro).

Osservando gli altri percentili, nell'Icv2003 emerge che il 20 per cento delle famiglie meno ricche percepisce un reddito non superiore a 10.166 euro, mentre il 20 per cento delle famiglie più ricche ha un reddito superiore a 33.800 euro. Solo un 5 per cento delle famiglie italiane ha un reddito superiore a 57.100 euro.

Dal confronto con Bi, si nota che fino al 20° percentile i livelli di reddito sono lievemente superiori per Bi. Dal 30° percentile in poi la situazione si inverte e spostandosi verso i livelli alti di reddito le differenze si acuiscono: il 20 per cento delle famiglie più ricche del campione di Bi percepisce un reddito superiore a 32.366 euro (-4 per cento rispetto al corrispondente valore di Icv2003). Il 90° percentile e il 95° percentile di Bi sono inferiori rispettivamente del 6 per cento e del 7 per cento rispetto a quelli dell'Icv2003. Queste prime osservazioni mettono quindi in luce che dall'Icv2003 emergono differenze nella distribuzione dei redditi maggiori rispetto a quelle evidenziate da Bi, soprattutto con riguardo alla coda destra della distribuzione dei redditi.

Tuttavia, considerando le differenze territoriali, ancora una volta i valori osservati per Bi indicano differenze più marcate rispetto a quanto risulta dall'Icv2003. In quest'ultima indagine per le famiglie del Nord il valore mediano è superiore del 14 per cento rispetto alla media nazionale, mentre per quelle meridionali risulta inferiore del 18 per cento. Gli stessi indicatori osservati per Bi, mostrano variazioni più forti (+23 per cento nel Nord, -26 per cento nel Sud ed Isole). Indicazioni simili si hanno anche per gli altri percentili.

Tavola A.19 - Quote di reddito familiare annuale netto per decile – Anno 2002 (valori percentuali)

	Icv2003	Bi
DECILE		
1°	1,6	2,1
2°	3,6	3,8
3°	4,8	5,0
4°	6,0	6,1
5°	7,2	7,3
6°	8,6	8,9
7°	10,4	10,7
8°	12,7	12,8
9°	16,3	16,1
10°	28,8	27,1

Fonte: Istat, Indagine sulle Condizioni di vita - Anno 2003. Elaborazioni su dati Banca d'Italia, 'I Bilanci delle Famiglie Italiane nel 2002'

Una conferma di una maggiore disuguaglianza nella distribuzione dei redditi rilevate nell'Icv2003 si ottiene dall'esame delle quote di reddito familiare percepito nelle varie classi delimitate dai decili, che mostrano differenze più forti tra le due indagini soprattutto nelle classi estreme. Nell'Icv2003, l'1,6 per cento del reddito totale è percepito dal 10 per cento delle famiglie più povere contro il 2,1 per cento di Bi. Il 28,8 per cento del reddito è percepito dal 10 per cento delle famiglie più ricche nell'Icv2003 contro il 27,1 per cento di Bi (Tavola A.19).

Nel complesso il 50 per cento delle famiglie più povere percepisce il 23,2 per cento del reddito totale, secondo l'Icv2003 e il 24,4 per cento secondo Bi.

Ricapitolando, questi confronti preliminari mettono in luce che nell'indagine Icv2003 si riscontrano maggiori disuguaglianze rispetto a Bi, soprattutto nelle code estreme della distribuzione dei redditi. In ogni caso per valutare la disuguaglianza complessiva è opportuno riferirsi ad indicatori sintetici. L'indice di concentrazione del Gini, misurato sui redditi familiari dell'Icv2003 risulta pari a 0,396 contro un valore di 0,373 stimato per Bi⁸ (Tavola A.20).

⁸ L'indice di concentrazione di Gini, misurato sui redditi familiari, inclusivi dei fitti figurativi, risulta pari a 0,359 (Banca d'Italia, 2004)

Un'ulteriore riprova di una minore disuguaglianza in Bi, si ha anche dall'esame della deviazione logaritmica media⁹ che risulta pari a 0,85 nell'Icv2003 contro un valore di 0,508 in Bi. Anche l'indice di Theil_1 (0,276 nell'Icv2003 contro 0,243 di Bi) e i tre indici di Atkinson con parametri 0,25, 0,50 e 0,75, pari rispettivamente a 0,068, 0,138 e 0,224 nell'Icv2003 contro 0,060, 0,119 e 0,186 di Bi confermano la presenza di una distribuzione dei redditi più diseguale nell'Icv2003.

Tavola A.20 - Indicatori della distribuzione dei redditi familiari netti annuali - Anno 2002

	Icv2003	Bi
Gini	0,396	0,373
Deviazione logaritmica media	0,850	0,508
Theil 1	0,276	0,243
CV	0,839	0,806
Atkinson25	0,068	0,060
Atkinson50	0,138	0,119
Atkinson75	0,224	0,186
SCOMPOSIZIONE DEVIAZIONE LOGARITMICA MEDIA PER RIPARTIZIONE		
NORD	0,661	0,265
CENTRO	0,882	0,393
SUD ED ISOLE	1,077	0,855
Dev log media nei gruppi	0,840	0,489
Dev log media tra gruppi	0,010	0,020
Dev log media totale	0,850	0,508
Dev log media nei gruppi (per cento)	98,8	96,1
Dev log media tra gruppi (per cento)	1,2	3,9
SCOMPOSIZIONE DEVIAZIONE LOGARITMICA MEDIA PER FONTE PRINCIPALE DI REDDITO FAMILIARE		
LAVORO DIPENDENTE	0,17165	0,16099
LAVORO AUTONOMO	0,83678	0,3203
PENSIONI	0,20949	0,18643
ALTRO	0,94539	0,83229
NESSUNA FONTE	0	0
Dev log media nei gruppi	0,31052	0,21443
Dev log media tra gruppi	0,53997	0,29401
Dev log media totale	0,85049	0,50844
Dev log media nei gruppi (per cento)	36,5	42,2
Dev log media tra gruppi (per cento)	63,5	57,8

Fonte: Istat, Indagine sulle Condizioni di vita - Anno 2003. Elaborazioni su dati Banca d'Italia, 'I Bilanci delle Famiglie Italiane nel 2002' N.B. redditi negativi esclusi e redditi nulli approssimati a 0,000001

Le ragioni di questa maggiore disuguaglianza non sono imputabili alle differenze territoriali. Infatti, come era emerso in precedenza dall'esame dei valori medi e dei percentili e come è confermato dalla scomposizione della deviazione logaritmica media per ripartizione, i dati di Bi mostrano una maggiore disuguaglianza tra le ripartizioni. Se, infatti, sui dati dell'Icv2003 risulta che la deviazione logaritmica media tra i gruppi è solo l'1,2 per cento di quella totale, sui dati Bi,

⁹ Questo indice e i successivi sono calcolati escludendo i redditi negativi e approssimando i redditi nulli a 0,000001.

le disuguaglianze tra i gruppi costituiscono invece il 3,9 per cento di quella complessivamente rilevata. Emerge invece un livello della disuguaglianza maggiore all'interno di ciascuna ripartizione, se si considerano i dati dell'Icv2003: in altri termini all'interno di ogni ripartizione si rilevano differenze più marcate tra i redditi più bassi e quelli più elevati.

Sembrano invece spiegare la maggior disuguaglianza rilevata in Icv2003, le differenze più elevate che si evidenziano tra le tipologie di reddito. Se si definiscono gruppi di famiglie sulla base del tipo di reddito prevalente percepito (da lavoro dipendente, da lavoro autonomo, da pensione e da altra fonte), si ha che la quota di deviazione logaritmica media attribuibile alle differenze tra i gruppi è il 63,5 per cento di quella totale per l'Icv2003 ed è pari al 57,8 per cento per Bi. Inoltre, per le famiglie che vivono prevalentemente di lavoro autonomo, nell'Icv2003 si riscontrano disuguaglianze maggiori rispetto a Bi: infatti per la prima indagine la deviazione logaritmica media osservata sul sottogruppo delle famiglie con reddito prevalentemente da lavoro autonomo è in linea con quella complessiva, mentre per Bi la deviazione logaritmica media è circa 2/3 di quella totale.

Va anche sottolineato che le famiglie in cui si percepiscono redditi da lavoro autonomo sono più numerose in Icv2003 (22,8 per cento) rispetto a Bi (18,4 per cento): ciò determina un incremento della disuguaglianza complessiva nell'intera popolazione, essendo il reddito da lavoro autonomo soggetto a maggiore variabilità rispetto alle altre tipologie di reddito.

È opportuno mettere in evidenza che per, quanto Icv2003 sia caratterizzata dall'obbligo di risposta a differenza di Bi, è stata ugualmente riscontrata una minore propensione da parte dei lavoratori autonomi di partecipare ad una rilevazione sui redditi rispetto alle altre categorie: di ciò si è tenuto conto nella costruzione dei coefficienti di riporto all'universo, adottando gli opportuni accorgimenti che consentissero di utilizzare informazioni provenienti da altre fonti (es. rilevazione sulle forze di lavoro) per la verifica della numerosità dei lavoratori autonomi. Queste considerazioni potrebbero giustificare il motivo per cui Icv2003 coglie meglio i percettori di reddito da lavoro autonomo e mostra conseguentemente una disuguaglianza complessiva tra i redditi più accentuata rispetto a Bi.

A.2.3 - Reddito equivalente: risultati e confronti

L'analisi del reddito familiare potrebbe indurre ad una sottostima delle reali differenze in termini di reddito disponibile per singolo individuo: si deve, infatti, tener presente che la dimensione familiare è maggiore nelle regioni meridionali rispetto ad altre aree del Paese. Per ovviare a ciò si fa usualmente ricorso al concetto di reddito equivalente.

Il concetto di "reddito familiare equivalente" si basa sulla constatazione che, all'aumentare del numero di componenti familiari, si creano all'interno della famiglia delle economie di scala tali per cui il reddito necessario per garantire un determinato tenore di vita aumenta in modo meno che proporzionale. Il reddito pro-capite non è dunque l'indicatore appropriato a considerare tali economie di scala: più opportuno risulterà invece il reddito calcolato tenendo conto che ogni individuo che si aggiunge alla famiglia avrà un peso minore di uno. È stata, dunque, calcolata la "numerosità equivalente della famiglia" utilizzando la scala di equivalenza modificata dell'Ocse, che attribuisce il peso pari a 1 al primo adulto della famiglia, il peso pari a 0,5 ad ogni individuo aggiuntivo di 14 anni o più e il peso di 0,3 ad ogni individuo di età inferiore ai 14 anni. Il reddito equivalente è dunque il reddito familiare rapportato alla numerosità equivalente della famiglia. L'unità di riferimento, cioè l'unità elementare per cui viene valutato il tenore di vita usando come proxy il reddito equivalente, è la persona. La distribuzione è quindi misurata tra gli individui, attribuendo a ciascuna persona il reddito equivalente della famiglia di appartenenza. Questa operazione presuppone l'ipotesi di distribuzione egualitaria dei redditi all'interno della famiglia e cioè che le entrate familiari siano messe in comune e ripartite in misura uguale tra tutti i componenti.

Tavola A.21 - Reddito equivalente medio annuale netto e percentili di reddito equivalente annuale netto per ripartizione – Anno 2002 (valori in euro)

	Icv2003			
	Nord	Centro	Sud ed Isole	Italia
MEDIA	16.340	15.569	10.446	14.072
PERCENTILE				
5 °	5.330	3.906	1.820	3.250
10 °	7.185	6.000	3.413	4.900
20 °	9.200	8.070	5.200	7.110
30 °	10.750	9.705	6.431	8.667
40 °	12.250	11.329	7.602	10.242
50 °	13.945	12.920	8.667	11.906
60 °	15.916	15.211	10.197	13.694
70 °	18.200	17.422	11.948	16.065
80 °	20.844	20.313	14.292	19.034
90 °	27.025	27.360	19.000	24.551
95 °	34.985	33.036	24.240	31.300
	Bi			
	Nord	Centro	Sud ed Isole	Italia
MEDIA	15.475	13.575	8.987	12.769
PERCENTILE				
5 °	5.780	4.762	1.789	3.467
10 °	6.894	6.154	3.329	4.980
20 °	8.830	7.877	4.852	6.560
30 °	10.400	9.107	5.833	8.000
40 °	11.869	10.311	6.685	9.344
50 °	13.457	11.758	7.791	10.936
60 °	15.115	13.073	8.877	12.675
70 °	17.059	14.988	10.162	14.665
80 °	19.938	17.731	12.424	17.212
90 °	25.057	22.603	15.882	21.947
95 °	31.953	27.381	20.607	27.584

Fonte: Istat, Indagine sulle Condizioni di Vita - Anno 2003. Elaborazioni su dati Banca d'Italia, 'I Bilanci delle Famiglie Italiane nel 2002'.

Tra l'altro, il reddito equivalente rappresenta la variabile di analisi per la maggior parte degli indicatori comunitari fissati nel Consiglio d'Europa di Laeken (dicembre 2001) per il monitoraggio della lotta alla povertà e alla esclusione sociale.

Il valore medio del reddito equivalente determinato sulla distribuzione degli individui è pari a 14.072 euro secondo l'Icv2003. Adottando la stessa scala di equivalenza, secondo i dati Bi il reddito equivalente per individuo è mediamente 12.769 euro (-9 per cento). Ancora una volta le disuguaglianze territoriali sono più marcate se si osservano i risultati di Bi (Tavola A.21).

Anche la distribuzione del reddito equivalente è caratterizzata da una maggiore concentrazione dei valori nella coda bassa e quindi il valore mediano è inferiore al valore medio. Il confronto tra valori mediani delle due indagini consente di confermare la presenza di un livello più basso per Bi: il 50 per cento degli individui ha a disposizione un reddito familiare equivalente non superiore a 11.906 euro secondo Icv2003 e 10.936 euro secondo Bi (-8 per cento).

Si noti che in termini di reddito equivalente le differenze tra l'Icv2003 e Bi sono più marcate rispetto a quanto riscontrato in precedenza considerando il reddito familiare. Ciò si giustifica in parte tenendo conto che la dimensione familiare media dei campioni delle due indagini in esame è diversa: in particolare il campione di Bi sembra essere caratterizzato da famiglie più numerose, la dimensione media familiare risulta pari a 2,64 componenti a fronte di 2,53 registrato nell'Icv2003 (differenze più forti si rilevano per le famiglie del Centro per le quali la dimensione media nel campione di Bi è 2,65 contro 2,34 dell'Icv2003).

Secondo i dati dell'Icv2003, l'indicatore di concentrazione del Gini determinato sul reddito equivalente è pari a 0,3498, con un valore più basso (0,3415) per il sottogruppo di individui fino a 15 anni (Tavola A.22). Si ricorda che il reddito equivalente è determinato dal reddito familiare, quindi sembrerebbe che le disuguaglianze siano meno forti in famiglie con minori.

Le disuguaglianze reddituali aumentano spostandosi dal Nord al Sud ed Isole: in quest'ultima ripartizione l'indice di concentrazione è pari a 0,36 contro lo 0,31 delle regioni settentrionali.

Tavola A.22 - Indicatore di concentrazione del Gini su reddito equivalente - Anno 2002

	Icv2003	Bi
SESSO		
Maschi	0,3503	0,3329
Femmine	0,3489	0,3344
ETA'		
Fino a 15 anni	0,3415	0,3487
16 anni o più	0,3502	0,3295
RIPARTIZIONE		
Nord	0,3110	0,2970
Centro	0,3460	0,3015
Sud ed Isole	0,3606	0,3384
Totale	0,3498	0,3337

Fonte: Istat, Indagine sulle Condizioni di vita - Anno 2003. Elaborazioni su dati Banca d'Italia, 'I Bilanci delle Famiglie Italiane nel 2002'.

Sulla base dei dati di Bi, l'indicatore di concentrazione del Gini determinato sul reddito equivalente è pari a 0,3337: si conferma quindi l'osservazione di una distribuzione dei redditi meno disuguale rispetto all'Icv2003. Da sottolineare che, in base a Bi, gli individui minori di 15 anni sono caratterizzati da una maggiore disuguaglianza dei redditi a loro disposizione (0,3487) rispetto al sottogruppo complementare dei maggiori di 15 anni (0,3295), a differenza di quanto osservato per l'Icv2003. Se si considerano le differenze a livello di ripartizione, è confermata la presenza di una minore disuguaglianza al Nord e di una maggiore disuguaglianza al Sud ed Isole.

Altro indicatore di disuguaglianza dei redditi equivalenti incluso tra i summenzionati indicatori di Laeken è il rapporto tra il reddito percepito dal 20 per cento della popolazione più ricco e quello percepito dal 20 per cento della popolazione più povero (S80_S20). Dai dati dell'Icv2003, risulta che il 20 per cento della popolazione più ricco percepisce 6,6 volte il reddito del 20 per cento più povero (Tavola A.23). Le differenze tra le code estreme della distribuzione dei redditi sono più marcate nel Sud ed Isole e meno tra individui anziani.

Ancora una volta, considerando i dati della Bi, emergono disuguaglianze minori: il rapporto tra il reddito del 20 per cento più ricco e quello del 20 per cento più povero è 5,8. In ogni caso le

regioni meridionali mostrano maggiori disuguaglianze e gli anziani sono caratterizzati da minori disuguaglianze.

Tavola A23 - Rapporto tra il reddito equivalente del 20 per cento della popolazione più ricca e il reddito equivalente del 20 per cento della popolazione più povera (S80_S20) - Anno 2002

	Icv2003	Bi
SESSO		
Maschi	6,7	5,8
Femmine	6,5	5,8
ETA'		
0-15	6,6	6,4
16-24	6,7	5,8
25-49	6,8	6,1
50-64	6,6	5,7
65+	4,9	4,3
RIPARTIZIONE		
Nord	4,9	4,5
Centro	6,2	4,7
Sud ed Isole	7,2	6,5
Totale	6,6	5,8

Fonte: Istat, Indagine sulle Condizioni di vita - Anno 2003. Elaborazioni su dati Banca d'Italia, 'I Bilanci delle Famiglie Italiane nel 2002'

A.2.4 - Indicatori di povertà: risultati e confronti

Tra gli indicatori comunitari fissati nel Consiglio d'Europa di Laeken del dicembre 2001, hanno particolare rilevanza quelli relativi alla diffusione e all'intensità della povertà. Il primo aspetto viene misurato con l'incidenza di povertà, vale a dire la percentuale di individui poveri, cioè con reddito equivalente al di sotto della linea di povertà. Il secondo fenomeno è misurato con il gap relativo mediano di povertà, ovvero la distanza relativa dalla linea di povertà del valore mediano del reddito equivalente dei poveri. La linea di povertà, secondo la definizione ufficiale europea, è pari al 60 per cento della mediana della distribuzione del reddito equivalente tra gli individui.

Nel 2002, secondo i dati dell'Icv2003, l'incidenza di povertà risulta 20,2 per cento con una linea di povertà pari a 7.143 euro annui per una famiglia di un componente (Tavola A.24). Le regioni particolarmente interessate dalla diffusione della povertà sono quelle meridionali, con una percentuale del 36,4 per cento della popolazione sotto la linea di povertà, circa 1,8 volte il valore nazionale. Nelle regioni del Nord, la povertà interessa meno del 10 per cento della popolazione.

Secondo quanto risulta dai dati della Bi, l'incidenza di povertà, determinata sulla base della metodologia ufficiale europea, risulta del 20 per cento a livello nazionale, con punte del 38,3 per cento nel Sud ed Isole (circa 2 volte il valore medio italiano). Rispetto all'Icv2003, i dati di Bi evidenziano una minore diffusione della povertà nel Nord e nel Centro e una maggiore diffusione nel Sud ed Isole.

Con riferimento al genere non si riscontrano differenze forti in entrambe le indagini. Si ricorda che l'incidenza di povertà è determinata utilizzando il reddito equivalente, identico per le

donne e gli uomini di una stessa famiglia. Il rischio di povertà leggermente più elevato per le donne si spiega con la maggior esposizione al rischio per le famiglie di donne sole, in particolare se anziane.

Tavola A.24 - Soglia di povertà e incidenza di povertà per età, sesso e ripartizione - Anno 2002

		Icv2003	Bi
SOGLIA DI POVERTA' ANNUALE (EURO)			
Persona sola		7143	6562
2 adulti con 2 bambini		15001	13779
INCIDENZA DI POVERTA'			
ETA'	SESSO		
0-15 anni	Maschi	24,6	29,0
	Femmine	25,2	29,0
	Totale	24,9	29,0
16-24 anni	Maschi	26,3	20,2
	Femmine	28,6	25,0
	Totale	27,4	22,4
25-49 anni	Maschi	18,1	19,3
	Femmine	20,2	20,1
	Totale	19,2	19,7
50-64 anni	Maschi	16,0	15,3
	Femmine	15,4	16,8
	Totale	15,7	16,0
65+ anni	Maschi	15,8	13,1
	Femmine	21,7	17,5
	Totale	19,3	15,6
Totale	Maschi	19,3	19,3
	Femmine	21,2	20,7
RIPARTIZIONE			
Nord		9,8	8,3
Centro		14,2	12,8
Sud ed Isole		36,4	38,3
Totale		20,2	20,0

Fonte: Istat, Indagine sulle Condizioni di vita - Anno 2003. Elaborazioni su dati Banca d'Italia, I Bilanci delle famiglie italiane nel 2002

I minori e i giovani fino a 24 anni sono maggiormente esposti alla condizione di povertà, indicazione questa del più elevato rischio di povertà per le famiglie con più figli a carico.

Tra le differenze tra le due indagini, emerge che in Bi il rischio di povertà per i minori di 15 anni è 1,5 volte quello medio (29 per cento contro 20 per cento dell'intera popolazione), mentre per l'Icv2003 è lievemente più basso (1,2 volte quello medio). Per le persone anziane con età superiore a 65 anni il rischio di povertà è più basso per entrambe le indagini, ma per Bi l'incidenza di povertà di questo sottogruppo di popolazione è più contenuta (15,6 per cento di Bi contro 19,3 per cento dell'Icv2003). Da sottolineare che, nell'ambito di questo sottogruppo di popolazione, le donne ultrasessantacinquenni secondo i dati Icv2003 risultano maggiormente esposte al rischio di povertà (21,7 per cento) della media della popolazione.

In relazione all'intensità della condizione di povertà, si rileva dai dati dell'Icv2003 un gap relativo mediano di povertà pari a 0,307: in altri termini la mediana del reddito equivalente della popolazione povera è inferiore alla linea di povertà di circa il 30 per cento (Tavola A.25).

I dati Bi mostrano un'intensità di povertà minore: il gap mediano di povertà risulta pari a 0,241. A rischio di una povertà più intensa sono soprattutto i minori di 15 anni e i residenti nelle regioni meridionali, coerentemente con quanto risulta dall'Icv2003.

Tavola A.25 - Gap relativo mediano di povertà per età, sesso e ripartizione - Anno 2002

		Icv2003	Bi
Totale		0,307	0,241
ETA'	SESSO		
0-15 anni	Maschi	0,336	0,260
	Femmine	0,314	0,274
16 anni o più	Maschi	0,316	0,238
	Femmine	0,281	0,236
RIPARTIZIONE			
Nord		0,265	0,163
Centro		0,333	0,214
Sud ed Isole		0,317	0,284

Fonte: Istat, Indagine sulle Condizioni di vita - Anno 2003. Elaborazioni su dati Banca d'Italia, 'I Bilanci delle Famiglie Italiane nel 2002'

A.2.5 - Nota tecnica sugli indicatori utilizzati

Di seguito sono riportate le modalità di calcolo dei principali indici di disuguaglianza e di povertà utilizzati nell'analisi.

Sia Y_i il reddito dello i -esimo individuo (o della i -esima famiglia), n il numero degli individui (o delle famiglie) e si supponga di ordinare i redditi in maniera non decrescente cioè

$$Y_1 \leq Y_2 \leq Y_3 \leq \dots \leq Y_n$$

L'indice di concentrazione di Gini è così determinato:

$$G = \frac{1}{n^2 \mu} \sum_{i=1}^n (2i - 1 - n) Y_i$$

dove

$$\mu = \frac{\sum_{i=1}^n Y_i}{n}$$

L'indice di concentrazione di Gini varia da un minimo di 0 nella situazione di perfetta uguaglianza ad un massimo di 1 nella situazione di massima disuguaglianza.

L'indice di Theil è definito come:

$$Theil\ 1 = \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{\mu} \ln\left(\frac{Y_i}{\mu}\right)$$

La deviazione logaritmica media è definita come:

$$Deviazione\ log\ media = \frac{1}{n} \sum_{i=1}^n \ln\left(\frac{\mu}{Y_i}\right)$$

Dati k gruppi, la deviazione logaritmica media può essere scomposta in due componenti: la deviazione logaritmica media nei gruppi e quella tra i gruppi. Sia MDL_j la deviazione logaritmica media relativa allo j-esimo gruppo e n_j la numerosità dello j-esimo gruppo.

Si ha:

$$Deviazione\ log\ media = \frac{1}{n} \sum_{i=1}^n \ln\left(\frac{\mu}{Y_i}\right) = \sum_{j=1}^k \frac{n_j}{n} MDL_j + \sum_{j=1}^k \frac{n_j}{n} \ln\left(\frac{\mu}{\mu_j}\right)$$

Il coefficiente di variazione è definito come:

$$cv = \frac{1}{\mu} \sqrt{\frac{\sum_{i=1}^n (Y_i - \mu)^2}{n}}$$

L'indice di Atkinson con parametro α è così definito:

$$Atkinson(\alpha) = 1 - \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{Y_i}{\mu}\right)^{1-\alpha} \right]^{\frac{1}{1-\alpha}} \quad \text{con } 0 \leq \alpha < 1$$

L'incidenza della povertà è definita nel modo seguente: sia il reddito equivalente Y_EQUIV_i ($i=1,2,\dots,n$) dello i-esimo individuo definito come rapporto tra il reddito familiare e la dimensione equivalente della famiglia ottenuta con l'applicazione della scala Ocese modificata, cioè:

$$\begin{aligned} \text{dim equivalente familiare} &= 1 + \\ &\quad (\text{num_individui} \geq 14\text{anni} - 1) * 0,5 + \\ &\quad (\text{num_individui} \leq 13\text{anni}) * 0,3 \end{aligned}$$

sia y_p la soglia di povertà:

$$y_p = 0,60 * \text{mediana}(Y_EQUIV)$$

definiamo la funzione:

$$h(Y_EQUIV_i | y_p) \begin{cases} 1 & \text{se } Y_EQUIV_i < y_p \\ 0 & \text{se } Y_EQUIV_i \geq y_p \end{cases}$$

L'incidenza di povertà sarà data da:

$$H = \frac{\sum_{i=1}^n h(Y_EQUIV_i | y_p)}{n}$$

Il gap relativo mediano di povertà è definito come segue:

$$\text{gap relativo mediano di povertà} = \frac{y_p - Me_pov}{y_p}$$

dove Me_pov = mediana reddito equivalente per i poveri

Bibliografia

Arèvalo R., *El mercado della vivienda in España*, Unpublished Ph.D. dissertation, Universidad Complutense de Madrid, 2001.

Arèvalo R., Ruiz-Castillo J., *The rental equivalence approach to nonrental housing in the Consumer Price Index: evidence from Spain*, Universidad Carlos III de Madrid, Departamento de Economía, Working Paper Series, N. 04-17, 2004.

Banca d'Italia, *I bilanci delle famiglie italiane nell'anno 2002*, Supplementi al Bollettino Statistico, Note metodologiche ed informazioni statistiche, Anno XIV, Numero 12, 2004.

Berger, Y.G., Skinner C.J., *Variance estimation for a low income proportion*, 2003, Applied Statistics, 52, 4, pp.457-468.

Bethel J., *Sample Allocation in Multivariate Surveys*, 1989, Survey Methodology, Vol. 15.

Betti G., Verma V., *Gross-net conversion issues*, Eurostat and Statistics Finland International Conference on "Comparative EU Statistics on Income and Living Conditions: Issues and Challenges", Helsinki, 6-7 November 2006.

Betti G., Donatiello G., Verma V., *The Siena Micro-Simulation Model (SM2) for net-gross conversion of EU-SILC (EU Statistics on Income and Living Condition) income variables: features and development*, Atti della Conferenza: 1ST General Conference of the International Microsimulation Association "Celebrating 50 Years of Microsimulation", Vienna 20-22 agosto 2007.

Bianchi G., Manzari A., Reale A., Siciliani I., *Identificazione dei percorsi per i rispondenti dell'indagine trasversale sulle condizioni di vita Anno 2003*, documento del "Gruppo di lavoro avente il compito di valutare l'applicabilità delle soluzioni tecniche e metodologiche adottate per il controllo e la correzione dei dati censuari 2001 della popolazione residente ai dati di altre indagini", 2004.

Binder D.A., Kovacevic M.S., *Estimating some measures of income inequality from survey data: an application of the estimating equation approach*, 1994, Survey Methodology, 21, pp. 137-145.

Bravi R., Moretti D., Rinaldelli C., *La valutazione dell'errore di campionamento degli indicatori europei di povertà e disuguaglianza*. Seminario Istat "Strategie metodologiche dell'indagine campionaria sui redditi e le condizioni di vita delle famiglie", Roma 5 Aprile 2006.

Canberra Group, Expert Group on Household Income Statistics, Final Report and Recommendations. Ottawa, 2001.

Coccia G., Pannuzi N., Rinaldelli C., *Poor and non poor households: the estimation from sample surveys*. Book of Short papers, CLADAG 2005, June 2005, MUP editore, pp.69-72.

Commission Regulation (EC) No 28/2004 of 5 January 2004 regarding the detailed content of intermediate and final quality report. (9.1.2004 L 5/42 Official Journal of the European Union).

Consolini P., *L'indagine sperimentale sull'archivio fiscale modd 770 anno 1999: analisi della qualità del dato e stime campionarie*, Contributi Istat, n. 29, Roma , anno 2004.

Consolini P., *L'integrazione dei dati campionari con fonti amministrative*. Seminario Istat "Dalla progettazione alla realizzazione della nuova indagine campionaria sui redditi e le condizioni di vita", Roma, 13 ottobre 2005.

Consolini P., *Integrazione di dati campionari EU-SILC con dati di fonte amministrativa*, .in corso di pubblicazione su Collana Metodi e Norme Istat, Roma, anno 2008.

Deville J.C., *Variance Estimation for complex statistics and estimators: linearization and residual techniques*, 1999, Survey Methodology, vol. 25 n.2, pp.193-203.

De Vitiis C., Di Consiglio L., Falorsi S., Pauselli C., Rinaldelli C., *La valutazione dell'errore di campionamento delle stime di povertà relativa*. Convegno Istat "Povertà Regionale ed Esclusione Sociale", Roma 17 Dicembre 2003.

Dougherty A., van Order R., *Inflation, Housing Costs and the Consumer Price Index*, 1982, American Economic Review, Vol. 72, n.1, pagg. 154-164.

Dufour J., Gagnon F., Morin Y., Renaud M., Särndal C.E., *A Better Understanding of Weight Transformation Through a Measure of Change*, Survey Methodology, June 2001, vol.27, N.1.

Duncan G., Kalton G., Kasprzyk D., Singh M.P., 1989, Panel Surveys, Wiley, New York.

Euromod, *Euromod: an integrated European benefits-tax model*, 2001, Final Report, edited by Sutherland, H., Euromod Working Paper n.EM9/01.

European Parliament and Council Regulation (EC) No 1177/2003 of 16 June 2003 concerning Community statistics on income and living conditions (EU-SILC). (3.7.2003 L 165/1 Official Journal of the European Union).

Eurostat, *Imputation of Income in the ECHP*, Doc.Pan 164/2002-12, Luxembourg: Eurostat, 2002.

Eurostat, *ECHP Udb, Construction of variables. From ECHP questions to UDB variables*, Doc.Pan 167/2003-12, Luxembourg: Eurostat, 2003.

Eurostat, *ECHP Udb Manual. European Community Household Panel, Longitudinal Users' Database. Waves 1 to 8 - Survey years 1994 to 2001*, Doc.Pan N 168/2003-12 Luxembourg: Eurostat, 2003.

- Eurostat, *Description of Target Variables: Cross-sectional and Longitudinal*, Doc. EU-SILC 065/2004, 2004.
- Eurostat, *Common cross-sectional EU indicators based on EU-SILC; the gender pay gap, Statistics on Income and Living Conditions (EU-SILC)*, Doc. EU-SILC 131/04, 2004.
- Eurostat, *Technical document on intermediate and final quality reports, Statistics on Income and Living Conditions (EU-SILC)*, Doc. EU-SILC 132/04, 2004.
- Eurostat, *Income in EU-SILC: Net/Gross/Net conversion. Report on common structure of the model; model description and application to the ECHP data for France, Italy and Spain*, Doc. EU-SILC 133/04, 2004.
- Eurostat, *Imputation Procedure*, Doc. EU-SILC 136/04, 2004.
- Eurostat, *HBS and EU-SILC imputed rent*, Doc. EU-SILC155/05, 2005.
- Falorsi S., Rinaldelli C., *Un software generalizzato per il calcolo delle stime e degli errori di campionamento*, 1998, *Statistica Applicata*, 10, 2, pp. 217-234.
- Falorsi P.D., Ballin M., De Vitiis C., Scepi G., *Principi e metodi del software generalizzato per la definizione del disegno di campionamento nelle indagini sulle imprese condotte dall'Istat*, 1998, *Statistica Applicata* Vol. 10, n.2.
- Falorsi P.D., Falorsi S., Russo A., Tranquilli G.B., *Indagini ripetute nel tempo: obiettivi e disegni di rilevazione*, 2001, *Rivista di Statistica Ufficiale*, Istat, n. 1/2001.
- Falorsi S., Rinaldelli C., Russo A., *Obiettivi conoscitivi e disegno di rilevazione delle principali indagini ripetute nel tempo condotte dall'Istat sugli individui e sulle famiglie*, 2001, *Rivista di Statistica Ufficiale*, Istat, n. 1/2001.
- Falorsi P.D., Falorsi S., Russo A., *Stimatori corretti di parametri longitudinali inerenti alle famiglie*, 2001, *Rivista di Statistica Ufficiale*, Istat, n. 2/2001.
- Falorsi S., Russo A., *Il disegno di rilevazione per indagini panel sulle famiglie*, 2001, *Rivista di Statistica Ufficiale*, Istat, n. 3/2001.
- Fellegi I.P., Holt D, *A systematic approach to edit and imputation*, 1976, *Journal of the American Statistical Association*, vol.71, pp.17-35.
- Gillingham R.F., *Measuring the Cost of Shelter for Homeowners: Theoretical and Empirical Considerations*, 1983, *The Review of Economics and Statistics*. Vol. 65, n. 2: 254-265.
- Heckman J., *Selection bias as a specification error*, 1979, *Econometrica*, Vol. 47 n.1: pagg. 153-161.
- Hidirouglou M.A., Berthelot J.M., *Statistical Editing and Imputation for Periodical Business Survey*, 1986, *Survey methodology*, 12, 1, 73-83.

Hoffmann J., Kurz C., *Rent indices for housing in West Germany 1985 to 1998*, 2004, Studies of Economic Research Centre of the Deutsche Bundesbank, Discussion Paper n° 08/2004.

ILO, *Resolution concerning household income and expenditure statistics*, 17th International Conference of Labour Statisticians, 2003.

Istat, *Panel Europeo sulle famiglie*, 2002, Collana Metodi e Norme n.15.

Istat, *Concord V. 1.0 Controllo e correzione dei dati, Manuale utente e aspetti metodologici*, 2004, Tecniche e Strumenti N. 1/2004.

Istat, *GENESEES v3.0 Funzione stime ed errori, Manuale Utente e Aspetti Metodologici*, 2005.

Kass G.V., *An explanatory technique for investigating large quantities of categorical data*, 1980, Applied Statistics, 29, 119-127.

Kish L., *Survey sampling*, 1965, Wiley and sons, New York.

Kovacevic M.S., Binder D.A., *Variance estimation for measures of income inequality and polarization – the estimating equations approach*, 1997, Journal Official Statistics, Vol. 13 N. 1, pp. 41-58.

McCarthy P.J., *Pseudoreplication: Further evaluation and application of the balanced Half-Sample Technique*, 1969, Vital and Health Statistics, Series 2 n.31, National Center for Health Statistics, Public Health Service, Washington, D.C.

Miron J.R., *Security of tenure, costly tenants and rent regulation*, 1990, Urban studies, Vol. 27, n. 2: 167-184.

Moretti D., Pauselli C., Rinaldelli C., *Il "Relative Poverty Gap" e il "Quintile Share Ratio": la stima dell'errore di campionamento secondo le Replicazioni Bilanciate Ripetute*, 2005, Statistica Applicata, 17, 2, pp.179-190.

Moretti D., Pauselli C., Rinaldelli C., *La stima della varianza campionaria di indicatori complessi di povertà e disuguaglianza*, 2005, Statistica Applicata, 17, 4, pp. 529-550.

Moretti D., Rinaldelli C., *Variance estimation for relative poverty measures and inequality indicators from complex sample surveys*, 2005, Atti del Quarto Convegno S.Co.2005, September 2005, Cluep editrice Padova, pp. 67-72.

Pauselli C., Rinaldelli C., *La valutazione dell'errore di campionamento delle stime di povertà relativa secondo la tecnica Replicazioni Bilanciate Ripetute*, 2003, Rivista di Statistica Ufficiale, 2/2003, pp.7-22.

Preston I., *Sampling distributions of relative poverty statistics*, 1996, Applied Statistics, 45, pp. 91-99.

Rao J.N.K., Wu C.F.J., *Resampling inference with complex survey data*, 1988, American Statistical Association, 83, 401, pp. 231-241.

Rao N.K., Shao J., *On balanced half-sample variance estimation in stratified random sampling*, 1996, Journal of American Statistical Association, vol. 91, n.433.

Rinaldelli C., *Statistiche complesse e software*, 2005, Statistica & Società, Anno III, n.2, 01.2005, pp.27-29.

Rinaldelli C., *Experiences of variance estimation for relative poverty measures and inequality indicators in official sample surveys*, 2006, Proceedings in Computational Statistics, COMPSTAT 2006, Italy, August 2006, pp. 1465-1472.

Rosen S., *Hedonic prices and implicit markets: product differentiation in perfect competition*, 1974, Journal of Political Economy, Vol. 82: 34-55.

Särndal C.E., Swensson B., Wretman J., *Model assisted survey sampling*, 1992, Spriger-Verlag, New York.

van der Laan P., *Integrating administrative registers and households surveys*, Netherlands Official Statistics, Vol. 15, Summer 2000.

Verma V., *Comparability in International Survey Statistics*, Keynote address to the International Conference on Improving Surveys, Copenhagen, 2002.

Verma V., Betti G., *Sampling error for measures of inequality and povert*, Book of Short papers, CLADAG 2005, June 2005, MUP editore, pp.175-178.

Vitaletti S., *Correzioni e imputazioni delle informazioni sui redditi*, Seminario Istat “Dalla progettazione alla realizzazione della nuova indagine campionaria sui redditi e le condizioni di vita”, Roma, 13 ottobre 2005.

Wolter K.M., *Introduction to variance estimation*, 1985, Spriger-Verlag, New York.

Metodi e Norme - Nuova serie - Volumi pubblicati

Anno 2000

6. *L'indice del costo della vita valevole ai fini dell'applicazione della scala mobile delle retribuzioni. Dalle origini alla cessazione (1945-97)*
7. *Le nuove stime dei consumi finali delle famiglie secondo il Sistema Europeo dei Conti SEC95*

Anno 2001

8. *La nuova indagine sulle cause di morte. La codifica automatica, il bridge coding e altri elementi innovativi*
9. *Il settore delle costruzioni in contabilità nazionale. I nuovi standard europei dettati dal SEC95*
10. *Indagini sociali telefoniche. Metodologia ed esperienze della statistica ufficiale*
11. *Elenco dei comuni al 31 maggio 2001* 
12. *Classificazione delle professioni* 

Anno 2002

13. *Le statistiche culturali in Europa*
14. *Gli investimenti lordi di contabilità nazionale dopo la revisione: nota metodologica*
15. *Panel Europeo sulle famiglie*

Anno 2003

16. *Metodi statistici per il record linkage*
17. *Metodologia e organizzazione dell'indagine multiscopo sulla domanda turistica "Viaggi e vacanze"*
18. *Classificazione delle attività economiche. Ateco 2002*

Anno 2004

19. *Inventario sulle fonti e metodi di calcolo per le valutazioni a prezzi costanti - Italia*
20. *Metodologia e tecniche di tutela della riservatezza nel rilascio di informazione statistica*
21. *Metodologia di stima degli aggregati di contabilità nazionale a prezzi correnti*
22. *Numeri indici dei prezzi alla produzione dei prodotti industriali venduti sul mercato interno - Base 2000=100*

Anno 2005

23. *I conti economici nazionali per settore istituzionale: le nuove stime secondo il Sec 95* 
24. *La rete di intervistatori Capi dell'Istat per la conduzione dell'indagine continua sulle Forze di Lavoro*
25. *Il monitoraggio del processo e la stima dell'errore nelle indagini telefoniche*
26. *Classificazione delle forme giuridiche delle unità legali*

Anno 2006

27. *Gli stranieri nella rilevazione continua sulle forze di lavoro*
28. *L'indagine campionaria sulle nascite: obiettivi, metodologia e organizzazione*
29. *Rilevazione mensile sull'occupazione, gli orari di lavoro e le retribuzioni nelle grandi imprese*
30. *La classificazione Istat dei titoli di studio italiani. Anno 2003* 
31. *Il sistema di indagini sociali multiscopo. Contenuti e metodologia delle indagini*
32. *La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*
33. *Il calcolo della spesa pubblica per la protezione dell'ambiente - Linee guida per riclassificare i rendiconti delle amministrazioni pubbliche*

Anno 2007

34. *Come si progetta il monitoraggio del lavoro sul campo di un'indagine sulle famiglie* 
35. *Istruzioni integrative per l'applicazione della Icd-10 nella codifica delle cause di morte* 

Anno 2008

36. *La progettazione e lo sviluppo informatico del sistema CAPI sulle forze di lavoro*
37. *L'indagine europea sui redditi e le condizioni di vita delle famiglie (Eu-Silc)*



***Produzione editoriale
&
Altri servizi***

La produzione editoriale

LE PUBBLICAZIONI A CARATTERE GENERALE

Annuario statistico italiano 2007
pp. XXIV+856+1 cd-rom; € 50,00
ISBN 978-88-458-1555-3

Atti - Ottava Conferenza Nazionale di Statistica
novembre 2006
pp. 160+1 cd-rom; € 22,00
ISBN 978-88-458-1569-0

Bollettino mensile di statistica
pp. 168 circa; € 15,00
ISSN 0021-3136

Compendio statistico italiano 2006
pp. 388; € 15,00
ISBN 978-88-458-1374-0

Italian Statistical Abstract 2005
pp. 376; € 15,00
ISBN 978-88-458-1316-0

Metodi e software per la codifica automatica e assistita dei dati
Tecniche e strumenti
n. 4, edizione 2007
pp. 184; € 17,00
ISBN 978-88-458-1544-7

Rapporto annuale
La situazione del Paese nel 2006
pp. XXXII+496; € 27,00
ISBN 978-88-458-1550-8
ISSN 1594-3135

Rapporto annuale
La situazione del Paese nel 2006
pp. XXXII+496+1 cd-rom; € 32,00
ISBN 978-88-458-1548-5
ISSN 1594-3135

LE NOVITÀ EDITORIALI A CARATTERE TEMATICO

AMBIENTE E TERRITORIO

Agrienvironmental indicators: methodologies, data needs and availability
Essays, n. 16/2006
pp. 140; € 14,00
ISBN 978-88-458-1360-3

Waste statistics on agriculture, forestry and fishing sectors
Essays, n. 17/2006
pp. 144; € 14,00
ISBN 978-88-458-1359-7

Water resources assessment and water use in agriculture
Essays, n. 18/2006
pp. 196; € 14,00
ISBN 978-88-458-1364-1

Atlante statistico dei comuni
Informazioni, n. 25, edizione 2006
pp. 68+1 cd-rom; € 14,00
ISBN 88-458-1362-2

Il calcolo della spesa pubblica per la protezione dell'ambiente (*)
Metodi e norme, n. 33, edizione 2006
pp. 368; € 26,00
ISBN 978-88-458-1380-1

POPOLAZIONE

Avere un figlio in Italia
Approfondimenti tematici dall'indagine campionaria sulle nascite
anno 2002
Informazioni, n. 32, edizione 2006
pp. 120+1 cd-rom; € 19,00
ISBN 88-458-1377-0

Decessi: caratteristiche demografiche e sociali (*)
anno 2002
Annuari, n. 11, edizione 2007
pp. 144; € 13,00
ISBN 978-88-458-1540-9

Matrimoni, separazioni e divorzi (*)
anno 2003
Annuari, n. 16, edizione 2006
pp. 164+1 cd-rom; € 22,00
ISBN 88-458-1371-1

Popolazione comunale per sesso, età e stato civile
anni 2002-2005
Informazioni, n. 29, edizione 2006
pp. 184+1 cd-rom; € 22,00
ISBN 88-458-1372-X

Popolazione e movimento anagrafico dei comuni
anno 2004
Annuari, n. 17, edizione 2006
pp. 280+1 cd-rom; € 27,00
ISBN 88-458-1373-8

SANITÀ E PREVIDENZA

I bilanci consuntivi degli enti previdenziali (*)
anno 2005
Informazioni, n. 9, edizione 2007
pp. 108+1 cd-rom; € 19,00
ISBN 978-88-458-1560-7

Cause di morte
anno 2002
Annuari, n. 18, edizione 2007
pp. 420; € 32,00
ISBN 978-88-458-1542-3

Decessi: caratteristiche demografiche e sociali (*)
anno 2002
Annuari, n. 11, edizione 2007
pp. 144; € 13,00
ISBN 978-88-458-1540-9

Dimissioni dagli istituti di cura per aborto spontaneo
anni 2002-2003
Informazioni, n. 20, edizione 2006
pp. 118; € 19,00
ISBN 88-458-1350-9



L'interruzione volontaria di gravidanza in Italia

anno 2003

Informazioni, n. 1, edizione 2006

pp. 118; € 14,00

ISBN 88-458-1314-2

Le organizzazioni di volontariato in Italia (*)

anno 2003

Informazioni, n. 27, edizione 2006

pp. 172+1 cd-rom; € 22,00

ISBN 88-458-1365-7

Statistiche della previdenza e dell'assistenza sociale (*)

I - I trattamenti pensionistici

anno 2005

Annuari, n. 9, edizione 2007

pp. 128+1 cd-rom; € 18,00

ISBN 978-88-458-1553-9

Statistiche della previdenza e dell'assistenza sociale (*)

II - I beneficiari delle prestazioni pensionistiche

anno 2004

Annuari, n. 8, edizione 2006

pp. 124+1 cd-rom; € 18,00

ISBN 978-88-458-1383-2

Stime preliminari della mortalità per causa nelle regioni italiane

anno 2004

Informazioni, n. 1, edizione 2007

pp. 116+1 cd-rom; € 19,00

ISBN 978-88-458-1381-8

Struttura e attività degli istituti di cura

anno 2003

Informazioni, n. 10, edizione 2006

pp. 126+1 cd-rom; € 19,00

ISBN 88-458-1331-2

CULTURA

La classificazione Istat dei titoli di studio italiani

anno 2003

Metodi e norme, n. 30, edizione 2006

pp. 96+1 cd-rom; € 15,00

ISBN 88-458-1340-1

I laureati e lo studio (*)

Inserimento professionale dei laureati
Indagine 2004

Informazioni, n. 3, edizione 2007

pp. 116+1 cd-rom; € 19,00

ISBN 978-88-458-1386-3

Statistiche culturali

anno 2005

Annuari, n. 45, edizione 2007

pp. 176+1 cd-rom; € 22,00

ISBN 978-88-458-1546-1

FAMIGLIA E SOCIETÀ

I consumi delle famiglie

anno 2004

Annuari, n. 11, edizione 2006

pp. 160+1 cd-rom; € 21,00

ISBN 88-458-1348-7

Diventare padri in Italia

Fecondità e figli

secondo un approccio di genere

Argomenti, n. 31, edizione 2006

pp. 344; € 27,00

ISBN 88-458-1355-X

Reddito e condizioni di vita

anno 2004

Informazioni, n. 31, edizione 2006

pp. 172+1 cd-rom; € 22,00

ISBN 978-88-458-1376-4

Il sistema di indagini sociali multiscopo

Metodi e norme, n. 31, edizione 2006

pp. 276; € 22,00

ISBN 88-458-1341-X

Strutture familiari e opinioni su famiglia e figli

anno 2003

Informazioni, n. 18, edizione 2006

pp. 208+1 cd-rom; € 27,00

ISBN 88-458-1347-9

L'uso del tempo

anni 2002-2003

Informazioni, n. 2, edizione 2007

pp. 236+1 cd-rom; € 27,00

ISBN 978-88-458-1385-6

I viaggi in Italia e all'estero nel 2004 (*)

Informazioni, n. 15, edizione 2006

pp. 94+1 cd-rom; € 15,00

ISBN 88-458-1339-8

La vita quotidiana nel 2005

Informazioni, n. 4, edizione 2007

pp. 284+1 cd-rom; € 27,00

ISBN 978-88-458-1387-0

PUBBLICA AMMINISTRAZIONE

I bilanci consuntivi degli enti previdenziali (*)

anno 2005

Informazioni, n. 9, edizione 2007

pp. 108+1 cd-rom; € 19,00

ISBN 978-88-458-1560-7

Il calcolo della spesa pubblica per la protezione dell'ambiente (*)

Metodi e norme, n. 33, edizione 2006

pp. 368; € 26,00

ISBN 978-88-458-1380-1

Le cooperative sociali in Italia

anno 2003

Informazioni, n. 30, edizione 2006

pp. 144+1 cd-rom; € 19,00

ISBN 978-88-458-1332-0

Finanza locale: entrate e spese dei bilanci consuntivi (comuni, province e regioni)

anni 2001 e 2002

Annuari, n. 10, edizione 2006

pp. 108+1 cd-rom; € 19,00

ISBN 978-88-458-1366-5

Le organizzazioni di volontariato in Italia (*)

anno 2003

Informazioni, n. 27, edizione 2006

pp. 172+1 cd-rom; € 22,00

ISBN 88-458-1365-7

La ricerca e sviluppo in Italia (*)

Consuntivo 2002

Previsioni 2003-2004

Informazioni, n. 2, edizione 2006

pp. 136; € 14,00

ISBN 88-458-1315-0

Statistiche della previdenza e dell'assistenza sociale (*)

I - I trattamenti pensionistici
anno 2005

Annuari, n. 9, edizione 2007

pp. 128+1 cd-rom; € 18,00

ISBN 978-88-458-1553-9

Statistiche della previdenza e dell'assistenza sociale (*)

II - I beneficiari delle prestazioni pensionistiche - Anno 2004

Annuari, n. 8, edizione 2006

pp. 124+1 cd-rom; € 18,00

ISBN 978-88-458-1383-2

Statistiche delle amministrazioni pubbliche

anno 2003

Annuari, n. 4, edizione 2006

pp. 416+1 cd-rom; € 37,00

ISBN 978-88-458-1384-9



GIUSTIZIA E SICUREZZA

Matrimoni, separazioni e divorzi (*)
anno 2003
Annuari, n. 16, edizione 2006
pp. 164+1 cd-rom; € 22,00
ISBN 88-458-1371-1

Ritardi della giustizia civile e ricadute sul sistema economico
Costi della giustizia civile rilevanti per il sistema delle attività produttive
Quaderni del MIPA, n. 13, ed. 2006
pp. 146; € 14,00
ISBN 88-458-1352-5

Statistiche giudiziarie civili
anno 2004
Annuari, n. 13, edizione 2006
pp. 248; € 22,00
ISBN 978-88-458-1368-9

Statistiche giudiziarie penali
anno 2004
Annuari, n. 13, edizione 2006
pp. 336+1 cd-rom; € 32,00
ISBN 88-458-1369-X

CONTI ECONOMICI

Il calcolo della spesa pubblica per la protezione dell'ambiente (*)
Metodi e norme, n. 33, edizione 2006
pp. 368; € 26,00
ISBN 978-88-458-1380-1

Contabilità nazionale
Conti economici nazionali
Anni 1970-2005
Annuari, n. 10, edizione 2007
pp. 760; € 45,00
ISBN 978-88-458-1549-2

LAVORO

Forze di lavoro
Media 2006
Annuari, n. 12, edizione 2007
pp. 208+1 cd-rom; € 27,00
ISBN 978-88-458-1557-7

I laureati e lo studio (*)
Inserimento professionale dei laureati
Indagine 2004
Informazioni, n. 3, edizione 2007
pp. 116+1 cd-rom; € 19,00
ISBN 978-88-458-1386-3

Lavoro e retribuzioni
anni 2001-2004
Annuari, n. 8, edizione 2007
pp. 196+1 cd-rom; € 22,00
ISBN 978-88-458-1552-2

Rilevazione mensile sull'occupazione, gli orari di lavoro e le retribuzioni nelle grandi imprese
Metodi e norme, n. 29, edizione 2006
pp. 92; € 10,00
ISBN 88-458-1328-2

La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione
Metodi e norme, n. 32, edizione 2006
pp. 208; € 22,00
ISBN 88-458-1357-6

Statistiche della previdenza e dell'assistenza sociale (*)
I - I trattamenti pensionistici
anno 2005
Annuari, n. 9, edizione 2007
pp. 128+1 cd-rom; € 18,00
ISBN 978-88-458-1553-9

Statistiche della previdenza e dell'assistenza sociale (*)
II - I beneficiari delle prestazioni pensionistiche - Anno 2004
Annuari, n. 8, edizione 2006
pp. 124+1 cd-rom; € 18,00
ISBN 978-88-458-1383-2

Gli stranieri nella rilevazione sulle forze di lavoro
Metodi e norme, n. 27, edizione 2006
pp. 76; € 9,00
ISBN 88-458-1323-1

PREZZI

Il valore della moneta in Italia dal 1861 al 2005
Informazioni, n. 21, edizione 2006
pp. 168; € 17,00
ISBN 88-458-1351-7

AGRICOLTURA E ZOOTECNIA

Statistiche dell'agricoltura
anni 2001-2002
Annuari, n. 49, edizione 2006
pp. 356; € 32,00
ISBN 88-458-1334-7

INDUSTRIA E SERVIZI

Conti economici delle imprese
anno 2003
Informazioni, n. 8, edizione 2007
pp. 138+1 cd-rom; € 19,00
ISBN 978-88-458-1556-0

Produzione e caratteristiche strutturali dell'industria siderurgica - Anni 2004-2005
Informazioni, n. 33, edizione 2006
pp. 84+1 cd-rom; € 15,00
ISBN 978-88-458-1378-8

La ricerca e sviluppo in Italia (*)
Consuntivo 2002
Previsioni 2003-2004
Informazioni, n. 2, edizione 2006
pp. 136; € 14,00
ISBN 88-458-1315-0

Statistica annuale della produzione industriale
anno 2003
Informazioni, n. 5, edizione 2006
pp. 52+1 cd-rom; € 15,00
ISBN 88-458-1322-3

Statistiche dei trasporti
anno 2004
Annuari, n. 5, edizione 2007
pp. 280; € 22,00
ISBN 978-88-458-1543-0

Statistiche dei trasporti marittimi
anni 2002-2004
Informazioni, n. 5, edizione 2007
pp. 56+1 cd-rom; € 15,00
ISBN 978-88-458-1545-4

Statistiche del trasporto aereo
anni 2004-2005
Informazioni, n. 6, edizione 2007
pp. 44+1 cd-rom; € 15,00
ISBN 978-88-458-1547-8

Statistiche sui permessi di costruire
anni 2003-2004
Informazioni, n. 7, edizione 2007
pp. 36+1 cd-rom; € 15,00
ISBN 978-88-458-1551-5

I viaggi in Italia e all'estero nel 2004 (*)
Informazioni, n. 15, edizione 2006
pp. 94+1 cd-rom; € 15,00
ISBN 88-458-1339-8

COMMERCIO ESTERO

Commercio estero e attività internazionali delle imprese 2006
1. Merci, servizi, investimenti diretti
2. Paesi, settori, regioni
+ *L'Italia nell'economia internazionale*
Rapporto ICE 2006-2007 + 1 cd-rom
+ Sintesi del Rapporto ICE
Annuari, n. 9, edizione 2007
pp. 376+440+536+48
€ 100,00 (in cofanetto)
ISBN 978-88-458-1561-4

**5° CENSIMENTO GENERALE DELL'AGRICOLTURA
22 OTTOBRE 2000**

Caratteristiche strutturali delle aziende agricole

Fascicolo nazionale; € 25,00
Fascicoli regionali; € 22,00
Fascicoli provinciali; € 22,00

Caratteristiche tipologiche delle aziende agricole

Fascicolo nazionale; € 20,00
Fascicoli regionali; € 14,00

Organizzazione e atti del 5° Censimento generale dell'agricoltura

pp. 152; € 14,00; ISBN 88-458-1287-1

APPROFONDIMENTI TEMATICI

La coltivazione della vite in Italia

Volume I - Caratteristiche generali
pp. 300; € 26,50; ISBN 88-458-1280-4
Volume II - Vitigni
pp. 248; € 22,00; ISBN 88-458-1281-2

La donna in agricoltura

pp. 316; € 14,00; ISBN 88-458-1284-7

Le imprese agricole

pp. 338; € 22,00; ISBN 88-458-1283-9

Le infrastrutture delle aziende agricole

pp. 150; € 11,50; ISBN 88-458-1279-0

La zootecnia in Italia

pp. 380; € 26,50; ISBN 88-458-1282-0

**14° CENSIMENTO GENERALE DELLA POPOLAZIONE E
DELLE ABITAZIONI - 21 OTTOBRE 2001**

Primi risultati

pp. 300+1 cd-rom; € 25,00; ISBN 88-458-0689-8

Popolazione legale

pp. 312+1 cd-rom; € 27,00; ISBN 88-458-1069-0

Per gli utenti che acquistano oltre 10 volumi dei **PRODOTTI CENSUARI** è previsto uno sconto del 25%.

Struttura demografica e familiare della popolazione residente - Italia

pp. 294+1 cd-rom; € 32,00; ISBN 88-458-1388-6

Struttura socioeconomica e condizione abitativa della popolazione residente - Italia

pp. 226+1 cd-rom; € 25,00; ISBN 88-458-1538-2

Popolazione residente e abitazioni nelle province italiane

Fascicoli provinciali; +1 cd-rom; € 19,00/27,00

Popolazione residente e abitazioni nelle regioni italiane

Fascicoli regionali; +1 cd-rom; € 27,00

Popolazione residente e abitazioni nei grandi comuni italiani

+1 cd-rom; € 17,00/19,00

I documenti - Conoscere il censimento

pp. 628+1 cd-rom; € 50,00; ISBN 88-458-1536-6

Il Piano di rilevazione e il Sistema di produzione - Conoscere il censimento

pp. 450; € 32,00; ISBN 88-458-1539-0

APPROFONDIMENTI TEMATICI

Gli stranieri in Italia: analisi dei dati censuari

pp. 248; € 20,00; ISBN 88-458-1537-4

**8° CENSIMENTO GENERALE DELL'INDUSTRIA E DEI
SERVIZI - 22 OTTOBRE 2001**

Imprese, istituzioni e unità locali

Fascicolo nazionale; € 31,50
Fascicoli regionali; € 22,00
Fascicoli provinciali; € 14,00

Distretti industriali e sistemi locali del lavoro 2001

pp.158+1 cd-rom; € 17,00; ISBN 88-458-1505-6

Atti

pp.262; € 11,50; ISBN 88-458-1278-2

Altri prodotti e servizi

ABBONAMENTI 2008

L'abbonamento consente di disporre di tutte le informazioni relative al settore tematico prescelto, diffuse attraverso le pubblicazioni edite nel 2008, accompagnate, ove previsto, da supporto informatico (floppy disk, cd-rom). Gli abbonati riceveranno per posta i prodotti che saranno via via pubblicati nel/i settore/i prescelto/i, editi nell'anno di sottoscrizione dell'abbonamento, ad esclusione dei volumi inseriti nelle collane: Tecniche e strumenti, Essays, Quaderni del Mipa e Censimenti. Oltre all'abbonamento ai singoli settori editoriali è prevista la modalità di abbonamento "Tutti i settori", che comprende tutta la produzione editoriale dell'Istituto edita nel 2008 compresi i volumi del Commercio estero. L'abbonamento all'area "Generale", infine, comprende 11 numeri del *Bollettino mensile di statistica* e l'*Annuario statistico italiano*. Tutti coloro che sottoscriveranno un abbonamento anche ad un solo settore riceveranno una copia del *Rapporto annuale*. Per meglio comprendere il sistema degli abbonamenti è possibile visionare, sul sito www.istat.it, l'elenco 2006 e l'elenco 2007 delle pubblicazioni inviate agli abbonati alle edizioni 2006 e 2007.

Per sottoscrivere gli abbonamenti si può utilizzare il modulo riportato nella pagina seguente.

WWW.ISTAT.IT

Nel sito Internet è possibile informarsi sulla produzione editoriale più recente, richiedere prodotti e servizi offerti dall'Istat, leggere e prelevare i comunicati stampa, accedere alle Banche Dati, collegarsi con altri siti nazionali e internazionali. Attraverso il catalogo on line della produzione editoriale è possibile acquisire tutte le informazioni relative ai prodotti a partire dalle edizioni 2000.

Ulteriori informazioni possono essere richieste a:

ISTAT - Direzione centrale per la diffusione dell'informazione statistica e il supporto alla produzione editoriale - PED/D
Via Cesare Balbo, 16 - 00184 ROMA - Tel. 0646733278/80 - Fax 0646733477 - e-mail: marketing@istat.it

Abbonamenti 2008

Inviare questo modulo via fax al numero 064673.3477 oppure spedire in busta chiusa a:
Istituto nazionale di statistica - DCDE - Commercializzazione e Attività promozionali - Via Cesare Balbo, 16 - 00184 Roma
Per ulteriori informazioni telefonare ai numeri 064673.3278/79/80

Desidero sottoscrivere i seguenti abbonamenti per l'anno 2008:

TIPOLOGIE DI ABBONAMENTO

PREZZI

	ITALIA Euro	ESTERO Euro
Generale (Bollettino mensile di statistica, Annuario statistico italiano)	<input type="checkbox"/> 180,00	<input type="checkbox"/> 200,00
Ambiente e territorio	<input type="checkbox"/> 70,00	<input type="checkbox"/> 80,00
Popolazione (escluso censimenti)	<input type="checkbox"/> 50,00	<input type="checkbox"/> 55,00
Sanità e previdenza	<input type="checkbox"/> 100,00	<input type="checkbox"/> 110,00
Cultura	<input type="checkbox"/> 40,00	<input type="checkbox"/> 45,00
Famiglia e società	<input type="checkbox"/> 140,00	<input type="checkbox"/> 150,00
Pubblica amministrazione	<input type="checkbox"/> 80,00	<input type="checkbox"/> 90,00
Giustizia e sicurezza	<input type="checkbox"/> 40,00	<input type="checkbox"/> 45,00
Conti economici	<input type="checkbox"/> 100,00	<input type="checkbox"/> 110,00
Lavoro	<input type="checkbox"/> 130,00	<input type="checkbox"/> 140,00
Prezzi	<input type="checkbox"/> 40,00	<input type="checkbox"/> 45,00
Agricoltura e zootecnia (escluso censimenti)	<input type="checkbox"/> 40,00	<input type="checkbox"/> 45,00
Industria e servizi (escluso censimenti)	<input type="checkbox"/> 40,00	<input type="checkbox"/> 45,00
Commercio estero	<input type="checkbox"/> 80,00	<input type="checkbox"/> 90,00
Tutti i settori (escluso censimenti)	<input type="checkbox"/> 900,00	<input type="checkbox"/> 1.000,00

Per un totale di _____
 Eventuale sconto (a) _____
Importo da pagare _____

Qualunque abbonamento comprende l'invio di una copia del Rapporto annuale.

(a) **Sconti e agevolazioni:** i soggetti pubblici e privati appartenenti al Sistan, gli **Enti pubblici**, le **Biblioteche** e le **Università** usufruiscono di uno sconto del **20% solo** se sottoscrivono l'abbonamento **direttamente con l'Istat.**

DATI PER LA FATTURAZIONE

Ente/Cognome e Nome _____
 Referente _____
 Codice fiscale * _____
 Partita IVA * _____
 Indirizzo _____ CAP _____ Città _____
 Prov. _____ Tel. _____ Fax _____ E-mail _____
 * è necessario compilare sia il codice fiscale che la partita IVA

DESTINATARIO DEI PRODOTTI (compilare solo se diverso dai DATI PER LA FATTURAZIONE)

Ente/Cognome e Nome _____
 Indirizzo _____ CAP _____ Città _____
 Prov. _____ Tel. _____ Fax _____ E-mail _____

MODALITÀ DI PAGAMENTO:

L'importo dovrà essere versato dall'acquirente, **dopo il ricevimento della fattura**, sul c/c postale n. 619007, oppure con bonifico bancario c/o la Banca Nazionale del Lavoro, **indicando con chiarezza il numero, la data della fattura e il codice cliente.** Per i versamenti tramite bonifico le coordinate bancarie sono: c/c n. 218050, ABI 01005.8, CAB 03382.9; via swift: B.N.L.I. IT RR, codice CIN K, codice anagrafico 63999228/j; IBAN IT64K0100503382000000218050.

INFORMATIVA - I dati da lei forniti saranno utilizzati per l'esecuzione dell'ordine e per l'invio, da parte dell'Istat, di promozioni commerciali, senza alcun impegno da parte sua. Il trattamento dei dati avverrà nell'assoluto rispetto della disciplina dettata dal d.lgs. 196/2003; essi non verranno utilizzati per finalità diverse da quelle indicate e saranno trattati esclusivamente dai dipendenti dell'Istituto incaricati. Il titolare dei dati è l'Istituto nazionale di statistica, Via Cesare Balbo, 16 - 00184 Roma - tel. 064673.3267, fax 064673.3477; responsabile del trattamento dei dati è il Direttore Centrale per la diffusione dell'informazione statistica e il supporto alla produzione editoriale, anche per quanto riguarda l'esercizio dei diritti dell'interessato di cui all'articolo 7 del d.lgs. 196/2003.

Data _____ Firma _____

PVABB08

PIÙ INFORMAZIONI. PIÙ VICINE A VOI.

I Centri d'Informazione Statistica

Per darvi più servizi e per esservi più vicino l'Istat ha aperto al pubblico una rete di Centri d'Informazione Statistica che copre l'intero territorio nazionale. Oltre alla vendita di prodotti informatici e pubblicazioni, i Centri rilasciano certificati sull'indice dei prezzi, offrono informazioni tramite collegamenti con le banche dati del Sistema statistico nazionale (Sistan) e dell'Eurostat (Ufficio di statistica della Comunità europea), forniscono elaborazioni statistiche "su misura" ed assistono i laureandi nella ricerca e selezione dei dati.

Presso i Centri d'Informazione Statistica, semplici cittadini, studenti, ricercatori, imprese e operatori della pubblica amministrazione troveranno assistenza qualificata e un facile accesso ai dati di cui hanno bisogno. D'ora in poi sarà più facile conoscere l'Istat e sarà più facile per tutti gli italiani conoscere l'Italia. Per gli orari di apertura al pubblico consultare il sito www.istat.it nella pagina "Prodotti e servizi".

ANCONA Via Castelfidardo, 4
Telefono 071/5013011 Fax 071/5013085

BARI Piazza Aldo Moro, 61
Telefono 080/5789317 Fax 080/5789335

BOLOGNA Galleria Cavour, 9
Telefono 051/6566111 Fax 051/6566185-182

BOLZANO Via Michael Gamper, 1
Telefono 0471/418400 Fax 0471/418419

CAGLIARI Via Firenze, 17
Telefono 070/34998700-1 Fax 070/34998732-3

CAMPOBASSO Via G. Mazzini, 129
Telefono 0874/604854-8 Fax 0874/604885-6

CATANZARO Viale Pio X, 116
Telefono 0961/507629 Fax 0961/741240

FIRENZE Via Santo Spirito, 14
Telefono 055/2393311 Fax 055/2393335

GENOVA Via San Vincenzo, 4
Telefono 010/5849701 Fax 010/5849742

MILANO Via Porlezza, 12
Telefono 02/806132214 Fax 02/806132205

NAPOLI Via G. Verdi, 18
Telefono 081/4930190 Fax 081/4930185

PALERMO Via G. B. Vaccarini, 1
Telefono 091/6751811 Fax 091/6751836

PERUGIA Via Cesare Balbo, 1
Telefono 075/5826411 Fax 075/5826484

PESCARA Via Caduta del Forte, 34
Telefono 085/44120511-2 Fax 085/4216516

POTENZA Via del Popolo, 4
Telefono 0971/377211 Fax 0971/36866

ROMA Via Cesare Balbo, 11/a
Telefono 06/46733102 Fax 06/46733101

TORINO Via Alessandro Volta, 3
Telefono 011/5166758-64-67 Fax 011/535800

TRENTO Via Brennero, 316
Telefono 0461/497801 Fax 0461/497813

TRIESTE Via Cesare Battisti, 18
Telefono 040/6702558 Fax 040/6702599

VENEZIA-MESTRE Corso del Popolo, 23
Telefono 041/5070811 Fax 041/5070835

La Biblioteca centrale

È la più ricca biblioteca italiana in materia di discipline statistiche e affini. Il suo patrimonio, composto da oltre 500.000 volumi e 2.700 periodici in corso, comprende fonti statistiche e socio-economiche, studi metodologici, pubblicazioni periodiche degli Istituti nazionali di statistica di tutto il mondo, degli Enti internazionali e dei principali Enti e Istituti italiani ed esteri. È collegata con le principali banche dati nazionali ed estere. Il catalogo informatizzato della biblioteca è liberamente consultabile in rete sul sito Web dell'Istat alla voce Biblioteca (www.istat.it).

Oltre all'assistenza qualificata che è resa all'utenza in sede, è attivo un servizio di ricerche bibliografiche e di dati statistici a distanza, con l'invio dei risultati per posta o via fax, cui i cittadini, gli studenti, i ricercatori e le imprese possono accedere.

È a disposizione dell'utenza una sala di consultazione al secondo piano

ROMA Via Cesare Balbo, 16 Telefono 06/4673.2380 Fax 06/4673.2617

E-mail: biblio@istat.it

Orario: Piano secondo da lunedì a venerdì 9.00 - 18.00

L'indagine europea sui redditi e le condizioni di vita delle famiglie (Eu-Silc)

Il volume presenta un panorama dell'indagine sui redditi e le condizioni di vita delle famiglie (Eu-Silc) realizzata secondo quanto stabilito dal regolamento comunitario 1177/2003. I capitoli illustrano gli aspetti metodologici e organizzativi più importanti dell'indagine quali la progettazione del disegno, le strategie di rilevazione e di elaborazione dei dati fino al calcolo degli errori campionari degli indicatori europei di povertà e disuguaglianza sociale. Oltre a quanto previsto dal regolamento, viene dedicata particolare attenzione alle soluzioni metodologiche e organizzative sviluppate dall'Istat quali la gestione informatizzata per seguire le famiglie campione per quattro anni consecutivi, l'integrazione dei dati rilevati con fonti amministrative, la metodologia di stima dei redditi lordi e le procedure complesse di calcolo degli errori di campionamento per gli indicatori stimati dall'indagine.

European Statistics on Income and Living Conditions (EU-SILC)

The book presents an overview of the survey on income and living conditions (EU-SILC) conducted in accordance with EC Regulation No. 1177/2003. The chapters show the most relevant methodological and organizational aspects of the survey. The sampling design, the survey methods, data processing and the calculation of the sampling errors of the European indicators of poverty and social inequality are described. In addition to what is required by the Regulation, Istat paid particular attention to the methodological and organizational solutions, for instance: the computerized management in tracing sampled households for four years, integration of surveyed data by administrative sources, methodology used for the estimation of gross income and the complex procedure of calculation of sampling errors of the survey.

ISBN 978-88-458-1596-6



9 788845 815966

1M01200803700000

€ 18,00