



SISTEMA STATISTICO NAZIONALE
ISTITUTO NAZIONALE DI STATISTICA

La metodologia di stima dei redditi lordi nell'indagine Eu-Silc

Indagine europea sui redditi e le condizioni di vita delle famiglie



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 simili</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>



SISTEMA STATISTICO NAZIONALE
ISTITUTO NAZIONALE DI STATISTICA

La metodologia di stima dei redditi lordi nell'indagine Eu-Silc

Indagine europea sui redditi e le condizioni di vita delle famiglie

A cura di: Gabriella Donatiello

Coordinamento redazionale: Maria Grazia Fiorentino

Per informazioni sul contenuto della pubblicazione
rivolgersi al Cont@ct Centre dell'Istat all'indirizzo:
<https://contact.istat.it/>

Eventuali rettifiche ai dati pubblicati saranno diffuse
all'indirizzo www.istat.it nella pagina di presentazione del volume

La metodologia di stima dei redditi
lordi nell'indagine Eu-Silc
*Indagine europea sui redditi e
le condizioni di vita delle famiglie*

Metodi e Norme • n. 49 - 2011

ISBN 978-88-458-1683-3

© 2011

Istituto nazionale di statistica
Via Cesare Balbo, 16 - Roma

Realizzazione: Istat, Servizio editoria

Stampato nel mese di luglio 2011 - copie 250
presso il Centro stampa dell'Istat
Via Tuscolana, 1788 - Roma

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

Indice

Capitolo 1 - Metodologie di stima dei redditi lordi	Pag. 7
1.1 - Introduzione.....	7
1.2 - I modelli di microsimulazione: una sintesi.....	9
1.3 - Le procedure di lordizzazione delle variabili di reddito nelle indagini campionarie	15
1.3.1 - <i>La conversione dei redditi dell'indagine Eu-Silc</i>	17
Capitolo 2 - Il modello di microsimulazione dell'Università di Siena (SM2) per la conversione dei redditi dell'indagine Eu-Silc	19
2.1 - Introduzione.....	19
2.2 - La struttura del Siena Microsimulation Model (SM2).....	20
2.3 - L'algoritmo di conversione lordi-netti in SM2.....	21
2.4 - La procedura iterativa per la conversione dei redditi netti in lordi.....	24
2.4.1 - <i>Le routines di conversione di SM2</i>	26
2.4.2 - <i>Le deduzioni e detrazioni specifiche: uno strumento per gestire la complessità</i>	28
2.5 - La sperimentazione e l'implementazione di SM2-Eu-Silc.....	29
Capitolo 3 - Strategie di rilevazione dei redditi in Eu-Silc: l'uso congiunto di fonti amministrative e campionaria	33
3.1 - Introduzione.....	33
3.2 - Le fasi del processo di integrazione dei dati di fonte diversa.....	34
3.3 - Generazione e validazione dei codici fiscali (presunti).....	35
3.4 - Variabili di reddito: armonizzazione del contenuto e delle definizioni tra le fonti del dato.....	39
3.4.1 - <i>I redditi da lavoro autonomo</i>	39
3.4.2 - <i>Redditi da lavoro dipendente</i>	45
3.4.3 - <i>Redditi da pensione</i>	49
3.5 - Analisi dell'incoerenza e riconciliazione delle componenti di reddito tra le fonti del dato.....	54
3.5.1 - <i>Riconciliazione del reddito da pensione tra Casellario, fonti fiscali e fonte campionaria</i>	54
3.5.2 - <i>Riconciliazione del reddito da lavoro tra fonti fiscali e campionaria</i> ...	55
3.5.3 - <i>Riconciliazione del reddito autonomo tra fonti fiscali e campionaria</i> ..	59
Capitolo 4 - La strategia di costruzione dei redditi lordi Eu-Silc: microsimulazione e integrazione dei dati campionari e amministrativi	63
4.1 - Introduzione.....	63
4.2 - La versione preliminare del modello SM2-Eu-Silc.....	66
4.2.1 <i>La parametrizzazione del sistema fiscale nazionale in SM2</i>	68
4.2.2 <i>Le innovazioni del modello SM2-Eu-Silc</i>	70
4.3 - La versione definitiva del modello SM2-Eu-Silc e l'uso congiunto di <i>record linkage</i> e microsimulazione per la costruzione dei redditi lordi.....	73

4.4 - La costruzione delle variabili obiettivo di reddito lordo familiare e.individuale.....	Pag. 74
Capitolo 5 - La valutazione della precisione delle stime.....	79
5.1 - Introduzione.....	79
5.2 - La valutazione della precisione delle stime dei redditi.....	80
5.3 - La documentazione sulla qualità secondo i regolamenti europei.....	82
Bibliografia.....	85

Capitolo 1 - Metodologie di stima dei redditi lordi

1.1 - Introduzione

L'indagine Eu-Silc (European Statistics on Income and Living Conditions) è la prima indagine a livello europeo ad avere come obiettivo la produzione di statistiche sul reddito individuale e familiare sia al netto sia al lordo dell'imposizione fiscale e contributiva.

All'avvio del progetto Eu-Silc, i paesi, fra cui l'Italia, in cui le informazioni sulle imposte e sui contributi sociali non venivano rilevate con l'indagine né venivano utilizzati gli archivi amministrativi, dovevano avviare studi di fattibilità per la conversione dei redditi campionari, preferibilmente attraverso tecniche di modellizzazione piuttosto che di semplice imputazione statistica.

L'Istat ha avviato uno studio di fattibilità nel 2004 analizzando le diverse alternative possibili. Per misurare l'imposizione fiscale e contributiva a livello individuale è infatti possibile: *i*) stimare le imposte e i contributi sociali attraverso un modello di microsimulazione; *ii*) utilizzare gli archivi fiscali disponibili; *iii*) rilevare le imposte e i contributi sociali attraverso l'indagine campionaria.

La strategia di rilevazione adottata dall'Istat per l'indagine Eu-Silc aveva escluso la raccolta diretta dei redditi lordi nel corso dell'intervista alle famiglie e le strade percorribili rimanevano il ricorso a un modello di microsimulazione oppure l'utilizzo degli archivi amministrativi. L'impiego dei dati fiscali o delle tecniche di microsimulazione andava attentamente valutato sulla base dei requisiti richiesti dalle variabili target dell'indagine.

Come noto, i dati derivanti dalle dichiarazioni dei redditi dei contribuenti non contengono informazioni su alcune componenti di reddito (redditi esenti, redditi a tassazione separata o soggetti a imposta sostitutiva) e possono avere problemi di copertura rispetto agli individui presenti nel campione dell'indagine. I dati campionari, a loro volta, possono essere affetti da reticenza, *under-reporting* (sottodichiarazione) o da insufficiente rappresentatività di alcune tipologie di reddito o di percettori.

Ne consegue che le informazioni desumibili dalle dichiarazioni dei redditi dei contribuenti o dei sostituti di imposta devono comunque essere integrate da altre fonti amministrative, ma non sempre si ottiene tutta l'informazione necessaria per la costruzione delle variabili obiettivo. Parimenti nel caso di utilizzo di un modulo di conversione dei redditi di un modello di microsimulazione va considerato che qualunque sottostima o sovrastima del dato netto campionario si traduce in una sottostima o sovrastima dei redditi lordi, con evidenti effetti negativi sulla qualità delle stime prodotte.

All'avvio dell'indagine Eu-Silc, Eurostat ha incaricato un gruppo di ricerca dell'Università di Siena di predisporre una procedura di conversione dei redditi che consentisse di utilizzare uno strumento armonizzato per la produzione delle variabili di reddito lorde e rappresentasse un valido sostegno soprattutto per i paesi che avrebbero dovuto costruire un modello di microsimulazione.¹ La Commissione europea ha adottato il modello di microsimulazione SM2

Il capitolo è a cura di Gabriella Donatiello

¹ Si veda: 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". Preparato da V. Verma e G. Betti e co-ricercatori. EU-SILC 133/04, Luxembourg, 2004.

dell'Università di Siena come procedura raccomandata per la costruzione delle variabili lorde dell'indagine Eu-Silc.

Prima del progetto Eu-Silc, per la conversione dei redditi campionari sono stati sviluppati modelli di microsimulazione "separati", basati su indagini nazionali non armonizzate. Il progetto Eu-Silc ha offerto un'occasione importante per disegnare un modello di microsimulazione flessibile e adattabile alle diverse realtà nazionali, con il vantaggio di potere utilizzare una base armonizzata di dati di input.

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 per l'anno di imposta 2003 e utilizzato sperimentalmente sui microdati dell'indagine Echip (European Community Household Panel) 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) e anche diverse strategie di raccolta dei dati (da fonte campionaria e/o amministrativa).

Il modulo di conversione netto-lordo del modello SM2 tende a 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 per alcuni aspetti meno dettagliato rispetto a modelli altamente specializzati.

Per la produzione dei redditi lordi Eu-Silc, prevista in base al regolamento dell'indagine dall'anno 2007, l'Istat ha deciso di testare il modello SM2 e di sperimentare una metodologia più complessa che rappresenta un'importante innovazione delle procedure di stima dei redditi lordi. Considerando la complessità della normativa fiscale e la disponibilità dei dati di fonte amministrativa adoperati per la costruzione dei redditi netti dell'indagine, è stata sperimentata l'integrazione delle informazioni tributarie nel modello di microsimulazione. L'utilizzo congiunto e innovativo del modello di microsimulazione e degli archivi amministrativi consentiva non solo di superare alcuni dei problemi menzionati precedentemente ma potenziava i vantaggi ottenibili dall'impiego esclusivo degli archivi fiscali da una parte e delle tecniche di microsimulazione dall'altra.

Il processo di produzione dei redditi lordi Eu-Silc è stato caratterizzato da una fase iniziale di studio del modello SM2 dell'Università di Siena e di successiva implementazione della versione preliminare del modello SM2-Eu-Silc per gli anni di indagine 2004, 2005 e 2006. A questa prima fase di sperimentazione è seguito poi lo sviluppo della versione definitiva del modello SM2-Eu-Silc 2007, 2008 e 2009, con l'utilizzo congiunto della microsimulazione e del *record linkage* per la costruzione delle variabili obiettivo di reddito lordo individuale e familiare.

L'implementazione da parte dell'Istat del modello SM2-Eu-Silc ha richiesto la transizione dalla versione preliminare applicata ai dati Echip a quella dei dati Eu-Silc e la costruzione del file di input e delle variabili ausiliarie sulla base delle informazioni rilevate con la nuova indagine. È stato, quindi, necessario modificare alcune *routines* di conversione netti-lordi, in particolare per la stima del reddito da lavoro autonomo e dell'imposta regionale sulle attività produttive (Irap). Sono state, inoltre, aggiornate le regole fiscali per includere il secondo modulo della riforma dell'Irpef del 2005 e le successive variazioni della normativa introdotte dalla legge finanziaria del 2007.

L'indagine Eu-Silc utilizza attualmente cinque archivi amministrativi (Agenzia delle entrate e Inps) che consentono di ricostruire i redditi netti e lordi per tipologia di fonte, le ritenute, le imposte pagate e gli oneri deducibili e detraibili. In particolare per la stima dei redditi lordi, le fonti amministrative sono state utilizzate, assieme ai dati campionari, come input del modello di microsimulazione e come *benchmark* dei redditi lordi microsimulati. La disponibilità di dati fiscali e dei redditi microsimulati ha consentito una comparazione e

validazione incrociata dei risultati molto utile per la costruzione delle variabili obiettivo. In presenza di dati sia campionari sia amministrativi, l'utilizzo del modello si è rivelato prezioso per individuare alcune incoerenze nei dati amministrativi e nello stesso tempo il modello è stato validato utilizzando i dati fiscali disponibili.

Il modello di microsimulazione ha consentito, inoltre, di stimare il carico fiscale e contributivo degli individui non presenti negli archivi amministrativi perché componenti di fatto delle famiglie (individui presenti al momento dell'intervista ma non inclusi nelle liste campionarie), oppure perché le informazioni anagrafiche necessarie alla procedura di *record linkage* risultavano incomplete. Il modello SM2-Eu-Silc ha fornito anche la stima dei contributi sociali posti a carico dei lavoratori e dei datori di lavoro.

Il volume presenta la metodologia di stima dei redditi lordi dell'indagine Eu-Silc attraverso capitoli dedicati alle innovazioni metodologiche introdotte e alle fasi più rilevanti di costruzione dei redditi individuali e familiari al lordo dell'imposizione fiscale e contributiva.

Il primo capitolo descrive le caratteristiche principali delle tecniche di microsimulazione e le procedure di lordizzazione delle variabili di reddito utilizzate solitamente nelle indagini campionarie. Il capitolo 2 è dedicato alla descrizione della struttura del modello di microsimulazione dell'Università di Siena (SM2), costruito per la conversione dei redditi dell'indagine Eu-Silc e adoperato dall'Istat assieme alle procedure di *record linkage* per la costruzione dei redditi lordi. Il capitolo 3 presenta l'esperienza Istat di integrazione delle fonti campionarie e amministrative per la rilevazione dei redditi Eu-Silc, finalizzata a ridurre la sottostima delle variabili di reddito. Il capitolo 4 evidenzia la strategia di costruzione dei redditi lordi e la scelta dell'utilizzo congiunto delle tecniche di microsimulazione e di integrazione dei dati campionari e amministrativi. Il capitolo 5 fornisce, infine, una valutazione del livello di precisione delle stime dell'indagine Eu-Silc.

1.2 - I modelli di microsimulazione: una sintesi

La tecnica comunemente utilizzata per la conversione dei redditi campionari è il modello di microsimulazione che imputa le imposte e i contributi sociali in base al regime fiscale relativo al periodo di riferimento del reddito. I modelli di microsimulazione sono tradizionalmente adoperati per analizzare la distribuzione dei redditi e della povertà e simulare l'impatto di misure alternative di politica economica e fiscale. Tali modelli sono in grado di simulare l'effetto redistributivo delle politiche sulla base di un campione rappresentativo di individui e famiglie.

Le tecniche di microsimulazione hanno cominciato a diffondersi all'inizio degli anni Cinquanta dello scorso secolo in Europa, Stati Uniti, Canada e Australia, stimolate dalle idee di Guy Orcutt e dalle critiche crescenti ai modelli macroeconomici.² Tuttavia è negli anni Ottanta che le microsimulazioni cominciano a essere utilizzate intensamente come strumento di decisione politica, soprattutto nei paesi anglosassoni.³

Da un'iniziale descrizione degli effetti redistributivi dei sistemi di tassazione e dei trasferimenti sociali, i modelli di microsimulazione diventano progressivamente uno strumento complesso di valutazione finalizzato alle modifiche dei sistemi di welfare. Recentemente le tecniche di microsimulazione hanno raggiunto anche i paesi dell'Ex Unione Sovietica e l'Africa e sono utilizzate in nuovi campi di applicazione come l'analisi per piccole aree, la sanità,

² Orcutt, G.H. "A New Type of Socio-Economic System". *Review of Economics and Statistics*, 80: 1081-1100, 1957.

³ Martini A e U. Trivellato. *The Role of Survey Data in Microsimulation Models for Social Policy Analysis*. LABOUR, 11(1), 83 – 112, 1997; Atkinson A.B. "EUROMOD and the Development of EU Social Policy". DIW Berlin, Discussion Papers 467, 2005.

l'offerta di lavoro e le risposte comportamentali degli agenti economici. La maggior parte dei modelli di microsimulazione possono essere divisi sostanzialmente in due grandi categorie: i modelli statici e quelli dinamici.

I modelli statici rappresentano la tipologia più diffusa per analizzare la disuguaglianza e la povertà: tali modelli sono definiti statici in quanto misurano l'impatto dovuto a misure fiscali e di politica sociale senza tenere conto dei cambiamenti nella struttura della popolazione o nei comportamenti degli individui. Ciò significa che gli eventi che cambiano le caratteristiche individuali nel tempo sono mantenuti al valore osservato in uno specifico momento. In questi modelli, tuttavia, vengono utilizzate alcune tecniche di aggiustamento dei pesi campionari e delle variabili monetarie (*static ageing techniques*) per aggiornare i microdati e garantire nel tempo la rappresentatività del campione rispetto alla popolazione.

In definitiva, i modelli statici confrontano una popolazione iniziale con uno scenario a legislazione vigente e una popolazione finale con scenari modificati da interventi fiscali, previdenziali o assistenziali al fine di calcolare i cambiamenti nei livelli di reddito e individuare "chi guadagna e chi perde" a seguito delle modifiche ipotizzate.

I modelli dinamici sono invece generalmente adoperati per analizzare l'evoluzione della struttura demografica della popolazione al fine di valutarne l'impatto sui conti pubblici e soprattutto sui sistemi pensionistici. Tali modelli consentono di comparare gli effetti di politiche alternative nel medio e nel lungo periodo, proiettando nel futuro le unità campionarie attraverso la simulazione dei principali eventi della vita (nascita, matrimonio, separazione, disoccupazione, pensionamento). La probabilità di questi eventi è assegnata a ciascuna unità per ciascun periodo attraverso numeri casuali (metodo Monte Carlo). L'evoluzione dinamica della popolazione, definita *dynamic aging*, viene analizzata sia in base alla legislazione vigente (scenario base) sia in base a cambiamenti delle regole fiscali e/o delle possibili risposte comportamentali degli individui (scenari alternativi).

La scelta fra l'utilizzo di un modello statico piuttosto che dinamico dipende dagli obiettivi di analisi e soprattutto dalla disponibilità e qualità dei microdati utilizzati.⁴

Fra i modelli statici più noti a livello internazionale vi sono TAXBEN realizzato fin dal 1983 dall'Istituto per gli studi fiscali nel Regno Unito e POLIMOD sviluppato nel 1993 all'Università di Essex (Regno Unito) per analizzare l'effetto redistributivo della tassazione e dei trasferimenti sociali. STINMOD è invece il modello di microsimulazione costruito dal Centro nazionale per la modellizzazione sociale ed economica (Natsem, Università di Canberra) dell'Australia. Negli Stati Uniti, fin dal 1973, è stato utilizzato il modello TRIM (Transfer Income Model) dell'Urban Institute di Washington e in Canada il modello SPSD/M (Social Policy Simulation Database and Model) è stato implementato da Statistics Canada come strumento di analisi del sistema fiscale nazionale.

I modelli di microsimulazione dinamici si sono sviluppati soprattutto dagli anni Novanta, ma all'inizio degli anni Settanta DYNASIM, voluto da Guy Orcutt all'Urban Institute di Washington, rappresenta il primo modello pionieristico degli Stati Uniti e DYNASIM3 è la versione odierna aggiornata. In Australia, presso Natsem, sono utilizzati DYNAMOD e DYNAMOD-2 e in Canada nel 1998 fu completato il modello di microsimulazione dinamica (DYNACAN) per analizzare la sostenibilità finanziaria del sistema previdenziale. Nel Regno Unito, nel 1992, fu realizzato PENSIM per stimare la distribuzione futura dei redditi da

⁴ Mitton L., H.Sutherland e M. Weeks (Eds). *Microsimulation Modelling for Policy Analysis. Challenges and Innovations*. United Kingdom: Cambridge University Press, 2000.

pensione e in Francia, alla fine degli anni Novanta, fu costruito DESTINIE per fornire previsioni di spesa sul sistema pensionistico.⁵

A livello europeo, il bisogno crescente di analisi comparate fra i paesi aderenti all'Unione europea ha portato nel 1998 allo sviluppo di EUROMOD attraverso un consorzio di esperti provenienti da 15 paesi guidati dall'Università di Essex. EUROMOD è un modello di microsimulazione statico in grado di stimare gli effetti delle politiche sociali e fiscali sui redditi familiari nei paesi membri dell'Unione. Attualmente EUROMOD è esteso a quattro nuovi paesi membri dell'Unione europea.

Prima dell'introduzione dell'indagine Eu-Silc, la base dati di EUROMOD era costituita da differenti fonti di microdati nazionali e ciò richiedeva un enorme dispendio di energie per l'armonizzazione delle diverse fonti e la trasformazione dei dati nel formato richiesto dal modello.⁶ La nuova indagine Eu-Silc ha sostituito le fonti nazionali e l'utilizzo di un'unica indagine disponibile per l'intera Europa a 27 paesi rappresenta un indubbio vantaggio soprattutto in termini di qualità dei dati e comparabilità internazionale. Inoltre, la disponibilità di variabili di reddito nette e lorde rappresenta un utile *benchmark* per i risultati generati dalle procedure di conversione dei redditi nazionali utilizzate in EUROMOD.

In Italia, è solo dalla metà degli anni Ottanta che cominciano a essere utilizzati i primi modelli di microsimulazione statici. La base dati di riferimento è rappresentata dall'indagine sui bilanci delle famiglie italiane della Banca d'Italia.

Fra i primi modelli vi sono ITAXMOD dell'Ispe in grado di simulare dapprima l'impatto delle imposte sul reddito e poi anche gli effetti redistributivi dei trasferimenti e dei contributi sociali.⁷ Successivamente sono stati sviluppati il modello TBM (Tax-Benefit Model) dalle Università di Venezia e Roma,⁸ il modello DIRIMOD da parte dell'Istituto Prometeia assieme alle Università di Modena e Bologna e anche il modello MASTRICT (Modello di analisi e simulazione delle imposte, dei contributi e dei trasferimenti) dell'Istat.⁹

MASTRICK rappresenta un'evoluzione del modello TBM finalizzato a valutare l'impatto dei provvedimenti fiscali e sociali sulla distribuzione del reddito familiare, mediante la stima delle imposte, dei trasferimenti e del reddito disponibile degli individui e delle famiglie. Il modello MASTRICT ricostruisce i redditi lordi a partire dai redditi netti rilevati dalla Banca d'Italia.

Successivamente sono stati implementati il modello MARTA del Centro Europa ricerche e il modello MAPP del Centro analisi delle politiche pubbliche (Capp) dell'Università di Modena in grado di simulare sia i trasferimenti sia le imposte dirette e indirette utilizzando

⁵ Una rassegna dei principali modelli statici è disponibile in: Atkinson A.B. e H. Sutherland H "Tax Benefits Models". STICERD Occasional Paper, n. 10, London School of Economics, London; Sutherland H. "Static microsimulation models in Europe: a survey, University of Cambridge, Microsimulation Unit". Discussion Paper, 1995. Per i modelli dinamici si veda: Harding A. "Dynamic microsimulation models: problems and prospects". Discussion Paper WSP/48, London, STICERD, London School of Economics, 1990; Zaidi A. e K. Rake. "Dynamic Microsimulation Models: A Review and Some Lessons for SAG". SAGE Discussion Paper n. 2, SAGEDP/02, 2001. Una lista completa dei modelli di microsimulazione più importanti è disponibile sul sito web dell'International Microsimulation Association all'indirizzo: <http://www.microsimulation.org/IMA/Population-based.htm>.

⁶ Figari F., H. Levy e H. Sutherland. "Using the EU-SILC for policy simulation: prospects, some limitations and some suggestions". In EUROSTAT. *Comparative EU statistics on Income and Living Conditions: Issues and Challenges - Proceedings of the EU-SILC conference*. Helsinki, 6-8 November 2006, 345-373, 2007.

⁷ Si veda: Lugaresi S. *ITAXMOD*. Ispe, 1989; Di Biase R., M. Di Marco, F. Di Nicola e G. Proto, *ITAXMOD: A Microsimulation Model of the Italian Personal Income Tax and of Social Security Contributions*. Ispe, 1995 (Documenti di Lavoro, n. 16).

⁸ Si veda: Rizzi D. "Un modello integrato per lo studio degli effetti delle politiche economiche sulla povertà". *Politica Economica*, n. 2, 1990; Rossi N. *La crescita ineguale. Primo rapporto sulla distribuzione e redistribuzione*. Bologna: Il Mulino, 1993.

⁹ Proto G. "Il modello di microsimulazione MASTRICT: struttura e risultati". *Rivista di Statistica Ufficiale*, n. 3, settembre-dicembre, pp. 25-55, 1999.

contemporaneamente l'indagine della Banca d'Italia e l'indagine sui consumi delle famiglie dell'Istat.¹⁰

In termini generali, il programma di calcolo della tassazione e dei trasferimenti sociali di un modello di microsimulazione permette di replicare le caratteristiche rilevanti di un sistema fiscale sulla base di alcune assunzioni, che semplificano necessariamente la complessità della realtà osservata. Vengono quindi modellizzate alcune relazioni basilari a livello microeconomico in grado di spiegare le variabili endogene Y (ad esempio il reddito disponibile, la partecipazione al mercato del lavoro eccetera) come funzione di variabili esogene X (ad esempio le caratteristiche socioeconomiche dell'unità di analisi) e di alcuni parametri di sistema P (aliquote fiscali, scaglioni di reddito, settori di attività economica eccetera).

I modelli di microsimulazione statici utilizzati per l'analisi delle politiche redistributive sono essenzialmente costituiti da: (i) una base di microdati contenente informazioni sociali, demografiche ed economiche sull'unità di analisi (individui, famiglie); (ii) un insieme di regole fiscali parametrizzate in grado di riprodurre per ogni unità campionaria lo schema di tassazione vigente nell'anno di riferimento del reddito (Figura 1.1).

I modelli di microsimulazione dinamici presentano, inoltre, un terzo elemento distintivo costituito dall'insieme delle risposte comportamentali finalizzato a riprodurre gli eventi demografici (nascita, nuzialità, divorzio, morte, migrazione), economici (mercato del lavoro, livello di reddito) e le reazioni comportamentali generate da eventi esterni, soprattutto da modifiche delle politiche fiscali (Figura 1.2).

La struttura portante di un qualsiasi modello di microsimulazione, sia esso statico che dinamico, è rappresentata dalla base di microdati. Nella maggior parte dei casi, il campione rappresentativo della popolazione deriva da indagini trasversali e longitudinali sulle famiglie, da fonti censuarie e dagli archivi amministrativi. I dati disponibili devono essere in genere completati da variabili ausiliarie necessarie per la simulazione delle diverse politiche fiscali e contributive, si pensi ad esempio alle spese per consumo, al risparmio, alla ricchezza o ad altre informazioni non rilevate direttamente dalle indagini le quali devono essere imputate facendo ricorso a tecniche econometriche o a procedure di *record linkage* e *statistical matching* di diverse indagini e/o archivi amministrativi.¹¹

Per la parametrizzazione del modello possono essere utilizzati diversi linguaggi di programmazione (ad esempio Sas, Stata, C++) e in particolare per i modelli dinamici risultano essenziali le procedure di aggiornamento stocastiche che utilizzano le probabilità di transizione tra stati per simulare l'evoluzione nel tempo della popolazione di riferimento in termini di dimensione, struttura e composizione. Tali modelli consentono di valutare gli effetti distributivi intergenerazionali e intragenerazionali delle politiche sociali, utilizzando diverse fonti di dati sia di tipo longitudinale, se disponibile, sia di tipo sezionale per coorte.

In definitiva, l'output di un modello di microsimulazione consente di analizzare gli effetti distributivi, di gettito e i costi finanziari totali delle riforme economiche proposte al fine di valutare i vantaggi e gli svantaggi delle alternative possibili, confrontando la popolazione nello scenario base con uno scenario finale generato dalla modifica delle regole fiscali e contributive.

Per la costruzione di un modello di microsimulazione risulta essenziale la qualità dei microdati di input, mentre il grado di complessità, in particolare per i modelli di tipo statico, deriva essenzialmente dal livello di dettaglio della parametrizzazione delle regole fiscali, dalla necessità di un aggiornamento annuale della base dati e/o della normativa e dall'eventuale presenza di differenti periodi di riferimento per il reddito, la tassazione, o per i saldi fiscali.

¹⁰ Viene utilizzata un'unica base dati derivante dal *matching* statistico delle due indagini. Si veda: Baldini M. *MAPP98: un Modello di Analisi delle Politiche Pubbliche*. Materiali di Discussione, CAPP, 2001.

¹¹ Siveda: Martini A. e U. Trivellato. *The Role of Survey Data in Microsimulation Models for Social Policy Analysis*. LABOUR, 11(1), 83 – 112, 1997.

Figura 1.1 - Struttura semplificata di un modello di microsimulazione statico

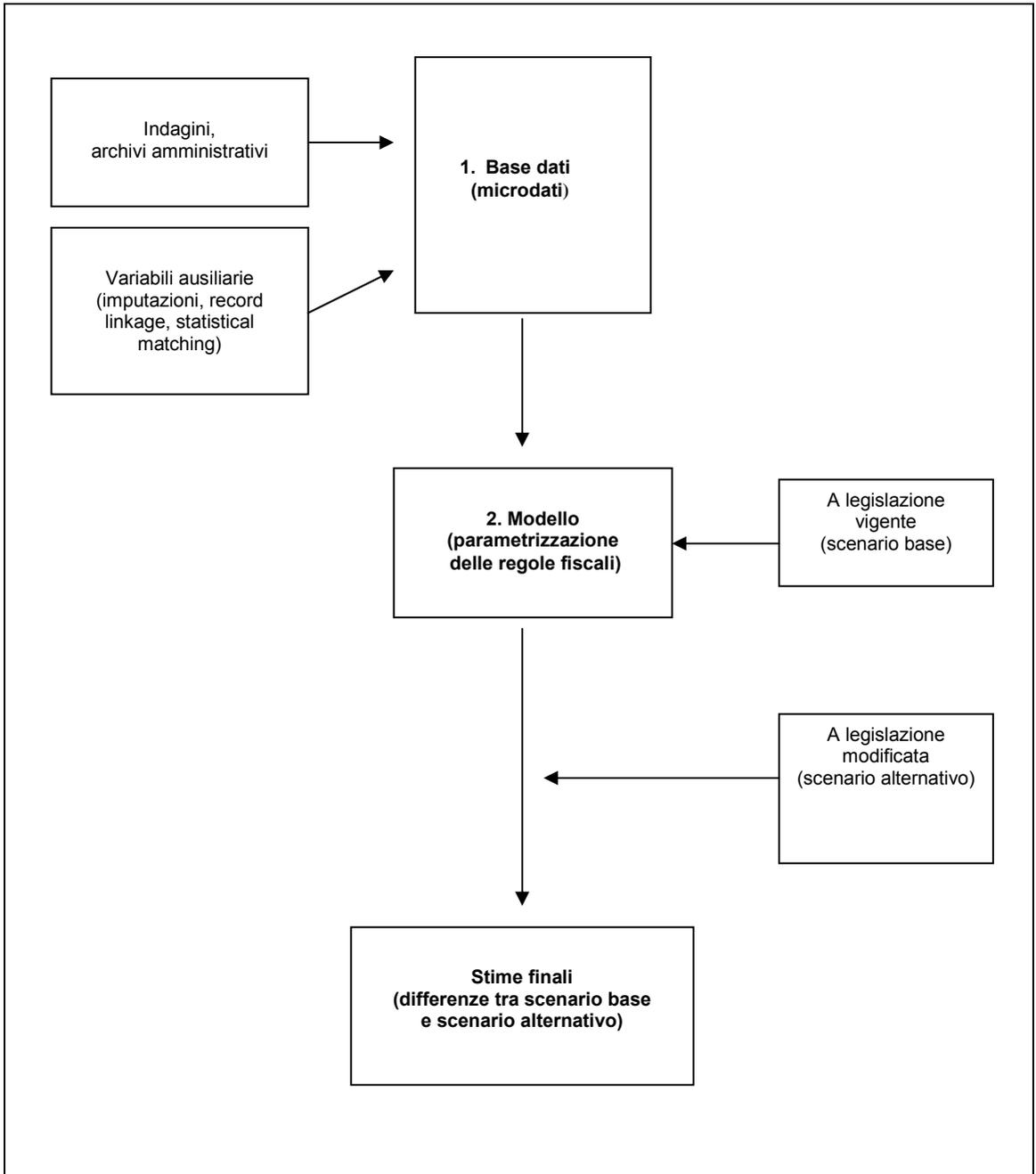
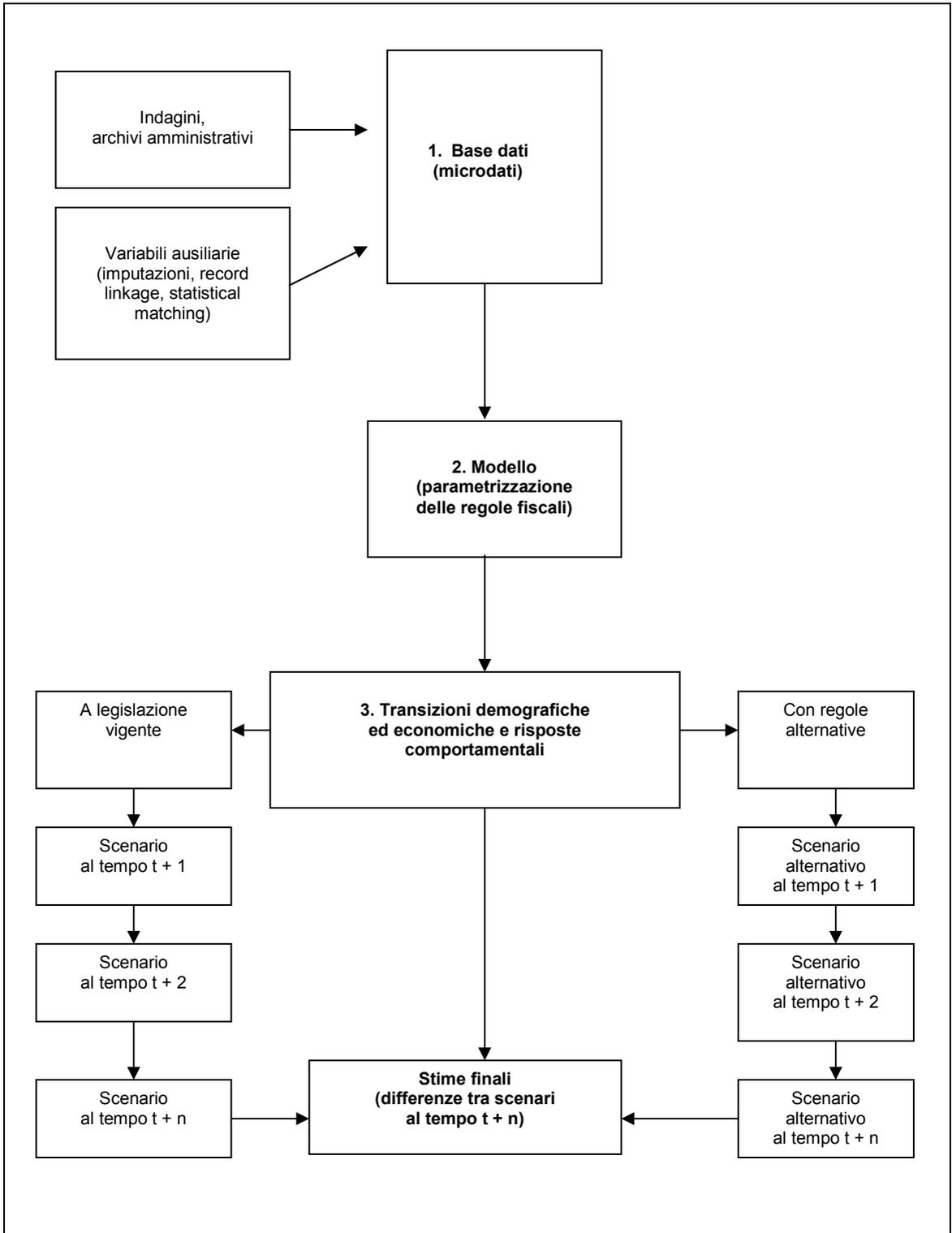


Figura 1.2 - Struttura semplificata di un modello di microsimulazione dinamica



1.3 - Le procedure di lordizzazione delle variabili di reddito nelle indagini campionarie

Le indagini campionarie di solito rilevano le informazioni sui redditi al netto dell'imposizione fiscale e contributiva per le difficoltà intrinseche connesse alla raccolta di tali dati. La modellistica di microsimulazione completa le informazioni derivanti dalle indagini statistiche attraverso la stima delle imposte, dei contributi sociali, dei beneficiari dei trasferimenti sociali e dei sussidi, poiché si tratta di dati essenziali sia per i *policy makers* sia per coloro che analizzano gli impatti distributivi, di gettito e gli incentivi e i disincentivi delle proposte di riforma delle politiche sociali.

Le procedure di conversione dei redditi netti campionari in lordi rispondono quindi all'esigenza di fornire una soluzione a un problema complesso, soprattutto quando i dati sui redditi lordi e sulle imposte pagate dagli individui e dalle famiglie non sono rilevate né direttamente dalle indagini né risultano disponibili dalle fonti amministrative.

Vi sono diverse procedure di lordizzazione dei redditi: da quelle relativamente semplici a procedure piuttosto complesse. Uno dei metodi più semplici è quello di utilizzare un modello statistico per determinare empiricamente un coefficiente di conversione netto-lordo come funzione del livello del reddito, dell'ampiezza familiare, della composizione della famiglia e di altre caratteristiche rilevate dalle indagini. Il coefficiente di conversione può essere calcolato sulla base di dati aggregati, come le statistiche fiscali o i dati di Contabilità nazionale.¹²

Nel caso in cui le informazioni sul reddito netto e lordo risultano entrambe disponibili, è possibile stimare un coefficiente di conversione per una singola componente di reddito e applicare poi tale coefficiente a tutte le altre componenti di reddito. È questo il metodo applicato da Eurostat per l'European Community Household Panel (Echp), in cui il reddito da lavoro dipendente era rilevato sia al netto che al lordo della tassazione. In questo caso, il coefficiente di conversione risulta funzione del solo livello del reddito.

Il metodo statistico, tuttavia, non consente di tenere conto delle caratteristiche individuali che modificano, anche in modo sostanziale, l'applicazione delle regole fiscali e contributive e soprattutto il modello statistico risulta in genere specificato su poche osservazioni tali da ridurre la capacità esplicativa. Un secondo limite importante è dovuto solitamente all'utilizzo di un coefficiente di conversione per unità familiare (come nel caso dell'indagine Echp), che non permette di differenziare i singoli individui all'interno della famiglia. In questo caso vengono ignorate tutte le differenze di imposizione dovute ad esempio alle diverse tipologie di reddito percepito e alle condizioni individuali dei singoli membri della famiglia (incapienza, carichi familiari eccetera).

Per superare i limiti dell'approccio statistico è possibile utilizzare un metodo più accurato che consente di calcolare un reddito lordo per ogni reddito netto, attraverso un'inversione analitica delle regole fiscali e contributive relative all'anno di riferimento del reddito. Si tratta di una procedura piuttosto complessa da un punto di vista computazionale e soprattutto molto dispendiosa in termini di tempo e risorse richieste per modellizzare e cambiare le procedure in caso di modifica della normativa fiscale.

Alcuni modelli di microsimulazione europei utilizzano un approccio iterativo per la conversione dei redditi campionari che consente di calcolare differenti livelli di reddito lordo per ogni individuo del campione di cui è noto un reddito netto. Applicando la normativa fiscale con le deduzioni e detrazioni previste dal sistema di tassazione, dal "reddito lordo stimato" si ricava un "reddito netto stimato" e lo si confronta con il reddito netto campionario fino a

¹² Betti G. e V. Verma. "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.

quando i due valori coincidono. A questo punto il reddito lordo stimato si ritiene, con un buon grado di approssimazione, corretto.

Un esempio di algoritmo iterativo è quello utilizzato in EUROMOD per 15 paesi dell'Unione europea, la cui procedura può essere descritta come segue.¹³ Nella fase iniziale, l'algoritmo utilizza il reddito netto campionario (SampleY) per calcolare il reddito lordo (GY_0) e nell'iterazione $x=0$ il fattore di conversione netto-lordo K_0 è posto uguale a 1 o a qualsiasi altro valore che sia ritenuto plausibile:

$$(1) K_x = 1 \quad x=0$$

$$(2) GY_x = K_x * \text{SampleY}$$

Applicando le regole fiscali previste nell'anno di riferimento del reddito e implementate nel modello, si può ricavare dal reddito lordo un reddito netto simulato (SimY_0)

$$(3) \text{SimY}_x = (1-t_x) * GY_x$$

dove t_x è l'aliquota media effettiva di tassazione.

Si pone quindi a confronto il reddito netto campionario e quello simulato con una procedura di controllo.

$$(4) |(\text{SampleY} - \text{SimY}_x) / \text{SampleY}| < |\delta|$$

Data la complessità e non linearità dei sistemi fiscali è teoricamente possibile che l'algoritmo non converga. Dopo un certo numero di iterazioni, fissato arbitrariamente, la procedura ricomincia con un valore iniziale differente (e casuale) di reddito lordo GY_0 .

Se il valore di reddito stimato (SimY_0) non è abbastanza vicino al reddito netto campionario (SampleY), viene stimato un nuovo reddito lordo (GY_1) mediante l'equazione 2 che utilizza un nuovo coefficiente di conversione K_1 , ossia:

$$(5) K_{x+1} = K_x (\text{SampleY} / \text{SimY}_x)$$

Dal nuovo reddito lordo GY_1 viene ricavato un nuovo reddito netto (equazione 3) e così via fino a quando il valore del reddito netto simulato (SimY_x) converge fino al valore del reddito netto campionario (SampleY).

Rispetto ad altre metodologie, l'algoritmo iterativo è in grado di applicare anche regole fiscali complesse con risultati più accurati rispetto ai modelli statistici che stimano un coefficiente di conversione dei redditi netti in lordi utilizzando poche variabili esplicative. L'approccio iterativo consente, inoltre, di differenziare la tassazione dei singoli individui all'interno dell'unità fiscale familiare, anche in caso di tassazione congiunta dei coniugi, e la tassazione per fonte di reddito. Si tratta di un indubbio vantaggio considerando che all'interno della stessa famiglia vi possono essere differenze rilevanti in termini di carichi fiscali dovute alle diverse tipologie di reddito (redditi totalmente o parzialmente esenti, redditi sottoposti a imposta sostitutiva eccetera) e ad altre condizioni individuali del contribuente (ad esempio incapienza, carichi familiari). In definitiva rispetto ad altre tecniche di conversione dei redditi

¹³ Immervoll H. e C. O'Donoghue. "Imputation of Gross Amounts from Net Incomes in Household Surveys. An Application using EUROMOD". EUROMOD Working Paper n. EM1/01, June 2001.

campionari, il metodo iterativo permette di ottenere una maggiore precisione delle stime e una migliore qualità del dato.

1.3.1 - La conversione dei redditi dell'indagine Eu-Silc

Dall'anno di indagine 2007, l'Istat ha reso disponibili i redditi individuali e familiari al netto e al lordo dell'imposizione fiscale e contributiva, così come previsto dal regolamento dell'Unione europea relativo alla produzione di statistiche sul reddito e le condizioni di vita (Eu-Silc). Il livello di dettaglio delle variabili richieste è piuttosto ampio: il reddito è disaggregato per tipologia (lavoro dipendente, lavoro autonomo, capitale reale, capitale finanziario, trasferimenti pensionistici e altri sussidi) e l'imposizione fiscale e contributiva è misurata per fonte di reddito.

Per la costruzione delle variabili target, i paesi aderenti al progetto Eu-Silc possono utilizzare sia le indagini campionarie sia gli archivi amministrativi, ma in caso di assenza di informazioni sui redditi lordi e sulle imposte pagate derivanti dalle indagini o dalle fonti fiscali, è necessario ricorrere alla modellistica di microsimulazione per la conversione dei redditi campionari.

All'avvio del progetto Eu-Silc, Eurostat ha incaricato l'Università di Siena di predisporre una procedura di conversione dei redditi che consentisse ai paesi di utilizzare uno strumento armonizzato per la produzione delle variabili di reddito lorde e rappresentasse un valido sostegno soprattutto per i paesi che avrebbero dovuto costruire un modello di microsimulazione ad hoc.

Prima del progetto Eu-Silc, diversi paesi hanno sviluppato modelli di microsimulazione basati su indagini nazionali non armonizzate finalizzati alla stima degli impatti redistributivi delle misure di politica economica e fiscale. La maggior parte di questi modelli incorpora moduli specifici di conversione dei redditi netti campionari in lordi o viceversa. Tuttavia le procedure di conversione dei modelli di microsimulazione "nazionali" complessi, pur essendo generalmente più precise, non risultano sufficientemente flessibili e "trasportabili" ad altri paesi, mentre le procedure più semplici, che considerano soltanto gli aspetti basilari del regime fiscale e contributivo, non garantiscono necessariamente la comparabilità internazionale.

Il progetto Eu-Silc ha rappresentato un'occasione importante per disegnare un modello di microsimulazione flessibile e adattabile alle diverse realtà nazionali, con il vantaggio di utilizzare una base armonizzata di dati di input.

Il modello di microsimulazione SM2 predisposto dall'Università di Siena è stato adottato dalla Commissione europea come procedura raccomandata per la costruzione delle variabili lorde del progetto Eu-Silc ed 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 per l'anno di imposta 2003 e utilizzato sperimentalmente sui microdati dell'indagine Echip 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) e anche diverse strategie di raccolta dei dati (da fonte campionaria e/o amministrativa).

Il modulo di conversione netto-lordo del modello SM2 è basato su un algoritmo iterativo e tende a 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 a un modello altamente specializzato come EUROMOD. Tale modello è costruito per rappresentare nel dettaglio e stimare gli effetti delle politiche sociali e fiscali sui redditi familiari nei paesi membri dell'Unione. EUROMOD consente infatti di simulare i trasferimenti sociali, completando le

informazioni raccolte con le indagini campionarie, e utilizza come base 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 “sottoprodotto” l’eventuale stima di dati dettagliati sui singoli benefici e sulle singole imposte pagate.

Con l’avvio della nuova indagine Eu-Silc, l’Istat nel 2004 ha deciso di testare il modello SM2 per costruire le variabili di reddito lordo individuale e familiare e di utilizzare una procedura più complessa che tenesse conto della disponibilità dei dati amministrativi adoperati sin dalla fase di costruzione dei redditi netti. L’utilizzo congiunto e innovativo del modello di microsimulazione e dei dati fiscali ha potenziato i vantaggi ottenibili dall’impiego esclusivo degli archivi amministrativi da una parte e delle tecniche di microsimulazione dall’altra, migliorando la qualità delle statistiche prodotte.

Nei capitoli successivi viene descritta dettagliatamente la struttura e le caratteristiche del modello di microsimulazione SM2 e la strategia di costruzione dei redditi lordi dell’indagine Eu-Silc.

Capitolo 2 - Il modello di microsimulazione dell'Università di Siena (SM2) per la conversione dei redditi dell'Indagine Eu-Silc

2.1 - Introduzione

La realizzazione del Siena Microsimulation Model (SM2) per la conversione dei redditi campionari dell'indagine Eu-Silc è stata condotta dal gruppo di ricerca dell'Università di Siena nell'ambito del progetto "Development of Appropriate modelling or imputation to Construct the EU-SILC Target Income Variables for each EU Member States" commissionato dall'Istituto europeo di statistica (Eurostat). Il modello è stato inizialmente predisposto e testato sulla base dei dati dell'indagine Echp (European Community Household Panel) per tre paesi dell'Unione: Francia, Italia e Spagna.

Nelle indagini campionarie che raccolgono informazioni sui redditi individuali, di solito gli intervistati possono dichiarare i redditi sia al lordo sia al netto della tassazione o di altre deduzioni previste dalla normativa fiscale. Tali informazioni devono poi essere rese omogenee soprattutto se si intende utilizzarle per analisi comparate tra diversi gruppi di popolazione o tra diversi paesi.

Il modello di microsimulazione SM2 è stato concepito come uno strumento flessibile in grado di trasformare le informazioni sul reddito derivanti da un'indagine campionaria o da altri fonti in un formato standardizzato utile per diversi tipi di analisi, incluso la simulazione di imposte e trasferimenti sociali.¹

La tecnica di microsimulazione utilizzata in SM2 presenta alcune caratteristiche peculiari. SM2 è un modello di tipo statico, incorpora quindi il regime fiscale riferito a un determinato periodo di tempo, tuttavia a differenza di altri modelli non necessita di dati di input in forma omogenea ed è in grado di generare come output redditi sia al lordo sia al netto dell'imposizione fiscale e contributiva. In effetti il modello SM2 è stato costruito per il contesto specifico dell'indagine Eu-Silc, con l'obiettivo di produrre informazioni comparabili sui redditi netti e lordi delle famiglie e degli individui, ed è stato sviluppato per essere utilizzato in paesi che presentano differenze rilevanti nei sistemi di tassazione. Sm2 è in grado, infatti, di gestire i diversi sistemi di tassazione attraverso un set standardizzato di procedure basate su una logica comune.

Partendo dai dati di reddito individuali e familiare, il modello SM2 è in grado di stimare tutte le informazioni per componente di reddito e di disaggregare gli importi lordi in imposte, contributi sociali e in reddito netto e reddito disponibile.²

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 nazionale, in grado di convertire i dati di input in una forma standardizzata. Queste *routines*

Il capitolo è a cura di Gianni Betti, Vijay Verma e Gabriella Donatiello

¹ Si veda Verma V., G. Betti, F. Ballini, M. Natilli e S. Galgani. "Personal income in the gross and net forms: applications of the Siena Micro-Simulation Model (SM2)". Working Paper n. 54, Università degli Studi di Padova: Dipartimento di Scienze Statistiche, 2003.

² Si veda: 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". Preparato da Verma V., G. Betti e co-ricercatori. EU-SILC 133/04, Luxembourg, 2004.

nazionali sono richiamate dal modulo centrale a cui forniscono i dati di input, tuttavia nel modello rimangono chiaramente distinte la struttura centrale, comune ai diversi regimi fiscali, e le *routines* specifiche per i diversi contesti 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: variabili raccolte con 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 modello SM2 è in grado di fornire una procedura robusta di conversione dei redditi netti e lordi delle famiglie e rispetto ai tradizionali modelli di microsimulazione SM2 tende a un livello intermedio di complessità, ossia è meno dettagliato per alcuni aspetti ma più completo per altri. Caratteristica rilevante di SM2 è che le peculiarità dei differenti sistemi di tassazione nazionale possono essere gestite dalla struttura generale del modello mediante la definizione di particolari tipi di "deduzioni" o di "detrazioni" per alcune specifiche componenti di reddito.

Nei paragrafi seguenti vengono descritte la struttura e le principali caratteristiche del modello di microsimulazione dell'Università di Siena. Sono quindi presentate le assunzioni basilari del modello e le relazioni tra reddito, imposte e contributi sociali per introdurre prima il modulo di conversione lordo-netto utilizzato in SM2 e poi la procedura iterativa per la produzione dei redditi lordi, partendo dalle informazioni disponibili di reddito netto. Infine, vengono riportati alcuni risultati della sperimentazione di SM2 con i dati di Echip per l'Italia, la Francia e la Spagna e anche alcuni dati sull'implementazione di SM2 con i dati della nuova indagine Eu-Silc (Italia e Turchia).

2.2 - La struttura del Siena Microsimulation Model (SM2)

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.

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.

Nel modello SM2, considerando la situazione in cui una persona riceve reddito da un'unica fonte ed è tassata separatamente come singola "unità fiscale"³ si hanno le relazioni riportate nel Prospetto 2.1.

Prospetto 2.1 - Terminologia e relazioni tra le componenti di reddito

1 Reddito lordo	$GG = G + SS(G)$
2 Contributi sociali dei lavoratori	$S = S(G)$
3 Reddito lordo al netto dei contributi sociali	$H = G - S$
4 Ritenuta alla fonte delle imposte e contributi	$XS = H$
	$XST = H - T(H)$
	$XT = H + S(G) - T(H)$
5 Deduzioni	$D = D(H)$
6 Reddito imponibile	$Y = H - D$
7 Imposta lorda	$W = W(Y)$
8 Detrazioni	$C = C(Y)$
9 Imposta netta	$X = W - C$
10 Reddito netto	$N = H - X$
SS:	Contributi sociali del datore di lavoro
XS	Ritenuta alla fonte dei contributi sociali
XST	Ritenuta alla fonte delle imposte e dei contributi
XT	Ritenuta alla fonte delle imposte

Il reddito lordo di un individuo o di una famiglia è uguale al reddito ricevuto prima di qualsiasi ritenuta alla fonte delle imposte 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 (S). Come noto, i contributi sociali gravano sui redditi da lavoro e comprendono i contributi dei datori di lavoro, dei lavoratori dipendenti e degli autonomi (incluso i lavoratori parasubordinati).

In SM2, il reddito lordo 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 a una riduzione del reddito lordo al netto dei contributi sociali.

Il reddito imponibile (Y) è ricavato dal reddito lordo al netto dei contributi sociali meno le deduzioni. L'imposta lorda è calcolata inizialmente come funzione del reddito imponibile. Le detrazioni d'imposta, che corrispondono a 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 è quindi uguale all'imposta lorda meno le detrazioni e infine il reddito netto (N) è uguale al reddito lordo al netto dei contributi sociali meno l'imposta pagata.

2.3 - L'algoritmo di conversione lordi-netti in SM2

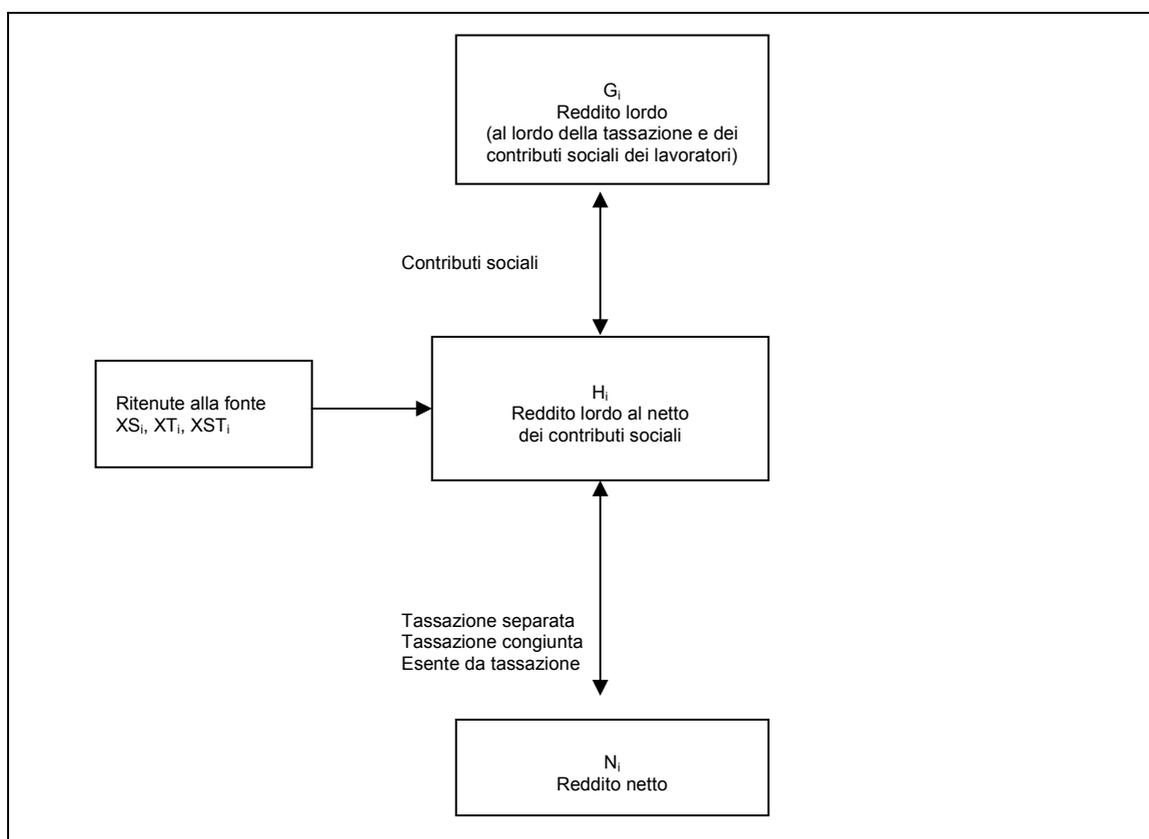
Se si considera una situazione più realistica in cui vi sono più fonti di reddito o più di un individuo nell'unità fiscale, la connessione tra redditi netti e lordi è esemplificata nella figura 2.1. In questo caso occorre distinguere essenzialmente tra le componenti di reddito che vanno sommate per tipologia e successivamente aggregate per individui nell'unità fiscale e le componenti che vanno trattate separatamente ai fini dell'imposizione fiscale e contributiva.

³ Con unità fiscale si intende l'individuo o la famiglia i cui redditi costituiscono l'imponibile per la tassazione e l'imposizione contributiva.

In SM2, la relazione tra reddito lordo al netto dei contributi sociali di una singola componente di reddito (H_i) e le altre quantità come il reddito lordo (G_i) e il reddito dopo le ritenute alle fonte (XST_i) è comunque semplice e indipendente dalle altre componenti di reddito o dagli altri individui nell'unità fiscale. Questo vale anche per la relazione tra reddito lordo al netto dei contributi sociali (H_i) e reddito netto (N_i) per le componenti che sono tassate separatamente a un'aliquota costante o per le componenti di reddito che sono esenti da tassazione.

Per determinare l'ammontare delle imposte dovute, i redditi imponibili sono in genere aggregati per tipologia e per individui nell'unità fiscale. In questo caso, la relazione tra il reddito lordo al netto dei contributi sociali (H_i) e il reddito netto (N_i) è più complessa. Ad ogni modo il procedimento che parte dai redditi lordi disponibili per calcolare i redditi netti è relativamente semplice, in quanto è possibile utilizzare l'aliquota di tassazione con una singola iterazione per ogni componente di reddito.

Figura 2.1 - Dai redditi lordi ai redditi netti



L'algoritmo di conversione lordo-netto descritto nel prospetto 2.2 rappresenta la struttura generale del modello che consente di gestire anche situazioni specifiche molto diversificate. L'algoritmo è stato testato in quattro paesi europei in cui il sistema fiscale è stato esaminato in

dettaglio (Italia, Spagna, Francia e Grecia).⁴ Le relazioni funzionali descritte richiedono che le variabili non abbiano effettivamente valori negativi.

Prospetto 2.2 - L'algoritmo di conversione lordo-netto

CALCOLO DEL REDDITO		Totale	Per componente di reddito (a)
1	REDDITO LORDO (b)	$G = \sum G_i \leftarrow$	G_i
2	Contributi sociali dei lavoratori		$S_i = S_i(G_i)$
3	REDDITO LORDO AL NETTO DEI CONTRIBUTI SOCIALI	$H = \sum H_i \leftarrow$	$H_i = G_i - S_i$
4	Deduzioni specifiche per componente		$D_i = D_i(H_i)$
AGGREGAZIONE PER COMPONENTI DI REDDITO E INDIVIDUI NELL'UNITÀ DI TASSAZIONE			
5	REDDITO IMPONIBILE	$Y = \sum Y_i \leftarrow$	$Y_i = H_i - D_i$
6	Deduzioni comuni	$D_0 = D_0(H)$	
7	Reddito imponibile al netto delle deduzioni comuni	$Y_0 = Y - D_0$	
8	Imposta lorda	$W_0 = W_0(Y_0)$	
9	Detrazioni d'imposta comuni	$C_0 = C_0(Y_0)$	
10	IMPOSTA LORDA al netto delle detrazioni comuni	$W = W_0 - C_0$	
11	Detrazioni d'imposta specifiche	$C = \sum C_i \leftarrow$	$C_i = C_i(Y_i)$
12	IMPOSTA NETTA	$X = W - C$	
13	REDDITO NETTO TOTALE	$N = H - X$	
14	Aliquota di tassazione	$R_0 = X/H$	
15	ALIQUOTA = IMPOSTA LORDA/REDDITO IMPONIBILE	$R = W/Y$	
DISAGGREGAZIONE – REDDITO INDIVIDUALE PER COMPONENTE			
16	Imposta netta per componente		$X_i = R \cdot Y_i - C_i$
17	REDDITO NETTO PER COMPONENTE		$N_i = H_i - X_i$

(a) La relazione funzionale di questa colonna potrebbe essere più complessa o diversa.

(b) Il reddito lordo inclusivo dei contributi sociali del datore di lavoro (SS) è uguale a: $GG = G + SS(G_i)$.

I contributi sociali dovuti sul reddito da lavoro (S_i) sono una funzione del reddito lordo (G_i) e nel caso dei redditi da lavoro dipendente escludono i contributi a carico del datore di lavoro (si veda la nota (b) del prospetto 2.2). La relazione funzionale $S_i(G_i)$ dipende dalla tipologia di reddito e in SM2, per calcolare i contributi sociali dei lavoratori dipendenti e autonomi sono utilizzate delle *subroutines* specifiche separate dal programma principale descritto nel prospetto 2.2.

Il reddito imponibile (Y_0 , rigo 7 del prospetto 2.2) si ottiene sottraendo dal reddito lordo le deduzioni che rappresentano la parte di reddito esente da tassazione. Le deduzioni sono funzione del reddito lordo e nel modello sono distinte due tipologie: (i) le “deduzioni

⁴ Si veda: Betti G., Donatiello G. e V. Verma. “The Siena Micro Simulation Model (SM2) for net-gross conversion of EU-SILC income variables”. *International Journal of Microsimulation*, 3(2), 2010.

specifiche” relative a particolari componenti di reddito (D_i) e (ii) le “deduzioni comuni” (D_0) che si applicano al totale del reddito imponibile.

Dopo la rimozione delle deduzioni specifiche è necessario sommare le componenti di reddito sottoposte a tassazione e anche i redditi degli individui presenti nella *tax unit*. Il modello consente comunque di escludere i redditi esenti da imposta o sottoposti a tassazione separata (la procedura è descritta al paragrafo 2.4.2).

L’imposta lorda è inizialmente calcolata come funzione del reddito imponibile (rigo 8) e dipende dallo schema di tassazione nazionale, mentre le imposte dovute sono generalmente ridotte dalle detrazioni d’imposta. Ancora una volta SM2 distingue fra le “detrazioni comuni” (rigo 9) che non riguardano una particolare componente di reddito e le “detrazioni specifiche” relative a una specifica componente di reddito (rigo 11). L’imposta netta (rigo 12) deriva quindi dall’imposta lorda meno le detrazioni, sia specifiche che comuni, e il reddito netto totale è uguale al reddito lordo meno l’imposta netta (rigo 13). Va sottolineato che sia l’imposta netta che il reddito netto (righi 12 e 13) sono riferiti, in questa fase, al reddito totale e non alle singole componenti di reddito.

L’aliquota di tassazione (R) rappresenta l’aliquota effettiva applicata a tutte le componenti di reddito e in SM2 è definita in due forme: (i) come una misura descrittiva del peso totale della tassazione (rigo 14), ossia è il rapporto tra l’imposta netta e il reddito lordo; (ii) come l’aliquota della tassazione dovuta prima di applicare qualsiasi detrazione d’imposta, ossia è il rapporto tra imposta lorda e il reddito imponibile (rigo 15) al netto delle deduzioni specifiche per componente di reddito (rigo 5).

L’aliquota R ha quindi due importanti funzioni: è lo strumento che consente di disaggregare l’imposta netta e il reddito netto per componente di reddito e inoltre è un parametro essenziale del procedimento iterativo per la conversione dei redditi netti in lordi.

La disaggregazione dell’imposta (rigo 16) e dei redditi netti per componente (rigo 17) è resa possibile in quanto R rappresenta l’aliquota di tassazione della singola componente di reddito e non una semplice aliquota media a livello del reddito totale. Tale disaggregazione è estremamente rilevante se si pensa, ad esempio, alla possibilità di avere in questo modo informazioni sui redditi sia al netto sia al lordo dei trasferimenti sociali.

2.4 - La procedura iterativa per la conversione dei redditi netti in lordi

La disponibilità di dati sui redditi individuali disaggregati per tipologia può variare in modo sostanziale da un paese a un altro e perfino tra individui e famiglie di uno stesso paese. Tuttavia le variazioni più rilevanti riguardano essenzialmente due aspetti: *i*) se una componente di reddito sia soggetta o meno all’imposizione fiscale o contributiva e *ii*) la tipologia di dati disponibili per la costruzione dei redditi.

Nei sistemi nazionali, l’imposizione fiscale di solito differisce in base alla fonte di reddito percepito. Alcune componenti di reddito possono essere a tassazione separata o soggette a imposta sostitutiva con un’aliquota costante. Altre componenti sono esenti da tassazione, come ad esempio i trasferimenti sociali, mentre vi sono casi in cui i redditi possono essere esenti fino a una certa soglia e tassati oltre tale soglia. In altri casi è possibile applicare una doppia tassazione. Vi sono anche situazioni che presentano ulteriori complessità e in alcuni sistemi fiscali dei paesi aderenti al progetto Eu-Silc gli individui possono scegliere tra più opzioni di tassazione.

La tipologia di dati disponibili per la costruzione delle variabili target dell’indagine Eu-Silc può quindi differire in modo sostanziale nei diversi paesi europei. I valori possono essere al lordo o al netto dei contributi sociali e/o della tassazione e in caso di ritenuta alla fonte delle

imposte, tale ritenuta può anche variare in base ad accordi fra il datore di lavoro e il lavoratore e può corrispondere o meno all'imposta netta finale. Questa ultima distinzione è particolarmente rilevante: è necessario infatti distinguere esattamente tra "ritenuta alla fonte delle imposte" e "ritenuta finale delle imposte" che, ad esempio, nel sistema fiscale italiano corrisponde all'imposta netta. In diversi sistemi fiscali, la ritenuta alla fonte può divergere anche significativamente dall'imposta finale e in alcuni paesi al contribuente è data la facoltà di scegliere l'aliquota da applicare alle ritenute. In Spagna, ad esempio, esiste un sistema di deduzioni alla fonte che consente di far coincidere l'imposta finale con le ritenute.

La struttura centrale del modello SM2 è in grado di gestire le diverse modalità di raccolta dei dati sintetizzati nel prospetto 2.3. Inoltre definendo in modo appropriato alcune deduzioni e detrazioni d'imposta (specifiche per tipologia di reddito) è possibile incorporare nella procedura standard del modello variazioni più complesse nella modalità di raccolta dei dati.

Prospetto 2.3 - Modalità di raccolta dei dati per le diverse componenti di reddito (a)

MODALITÀ (X _i) DI RACCOLTA DEI DATI	
G _i	Reddito lordo (al lordo della tassazione e dei contributi sociali del lavoratore, se applicabile)
H _i	Reddito lordo (al lordo della tassazione e al netto dei contributi sociali del lavoratore, se applicabile)
N _i	Reddito netto (al netto dell'imposta finale e dei contributi sociali, ossia ammontare finale ricevuto)
REDDITO PERCEPITO DOPO LA RITENUTA ALLA FONTE	
X _{Ti}	Ritenuta delle imposte (escluso i contributi sociali); T _i = ritenuta alla fonte delle imposte
X _{S_i}	Ritenuta dei contributi sociali (ma non delle imposte); S _i = ritenuta alla fonte dei contributi sociali
X _{T_{S_i}}	Ritenuta delle imposte e dei contributi sociali; T _i +S _i

(a) Componente di reddito (i) sottoposta a tassazione e ai contributi sociali.

Nel modello SM2, con ritenuta alla fonte si intende "l'ammontare delle imposte dovute per fonte di reddito", senza considerare l'apporto di altre eventuali componenti di reddito e senza tenere conto delle caratteristiche personali degli individui della *tax unit*. In molti casi il calcolo della ritenuta alla fonte è relativamente semplice, se basata su regole standard, e può essere espresso come:

$$T_i = (H_i - X_{ST_i}) = T_i(H_i)$$

dove la ritenuta alla fonte (T) è la differenza tra il reddito lordo al netto dei contributi sociali (H) e l'ammontare di reddito ricevuto dopo la ritenuta di imposte e contributi sociali (X_{T_{S_i}}), ossia una funzione nota del reddito lordo H per una specifica componente di reddito. Quindi X_{T_{S_i}} è direttamente riportabile ad H_i senza tener conto delle altre componenti di reddito.

La relazionale funzionale tra H_i e il reddito netto finale N_i è invece molto più complessa e dipende dal totale delle fonti di reddito dell'unità di tassazione. Le difficoltà maggiori si incontrano, tuttavia, quando la normativa sulle ritenute alla fonte non segue regole standardizzate, oppure le regole non sono applicate uniformemente o perfino non "esistono" nel senso che il contribuente può scegliere o può negoziare l'aliquota da applicare alle ritenute. In queste circostanze, la costruzione dei redditi lordi partendo dai redditi netti richiede il reperimento di informazioni aggiuntive sulle ritenute subite o sulla normativa applicata nei singoli casi.

A differenza delle ritenute alla fonte, l'imposta finale calcolata da SM2 corrisponde all'imposta netta dovuta sulla base del reddito totale (considerando quindi tutte le componenti di reddito) e sulla base delle caratteristiche dell'unità di tassazione. In questo caso il calcolo è più complesso in quanto deve necessariamente tenere conto degli elementi distintivi della *tax unit* (individui, famiglie o da altro tipo), delle circostanze particolari che possono influire sull'imposizione fiscale (incapienza, carichi familiari, eccetera) e del reddito totale derivante da tutte le fonti disponibili. Si suppone inoltre che, a differenza delle ritenute alla fonte, la normativa per il calcolo dell'imposta finale segua regole standard e che non possa variare sulla base di accordi individuali.

2.4.1 - Le routines di conversione di SM2

Il prospetto 2.4 descrive la procedura di conversione utilizzata nel modello SM2 che consente di trasformare le diverse tipologie di dati rilevati in un formato standard. Le componenti di reddito possono essere divise sostanzialmente in due gruppi N e H, ossia gli ammontari rilevati rappresentano il "netto finale" (N_i), oppure una qualsiasi altra tipologia di reddito ($G_i, X_{S_i}, X_{T_i}, X_{TS_i}, H_i$) direttamente trasformabile in reddito lordo al netto dei contributi sociali (H_i). Per tutte le tipologie di reddito, diverse dal reddito netto finale N_i , conviene mantenere il reddito lordo H_i , come obiettivo standard della procedura di conversione:

$$[G_i, H_i, X_{S_i}, X_{TS_i}, X_{T_i}] \Rightarrow H_i.$$

Questo tipo di procedura di conversione coinvolge le seguenti relazioni funzionali per componente di reddito: per i contributi sociali $S_i = S_i(G_i)$ e per la ritenuta alla fonte delle imposte $T_i = T_i(H_i)$. Come evidenziato precedentemente le ritenute alla fonte possono essere basate sia su regole generali sia su accordi specifici tra le parti.

Nella maggior parte dei casi, H_i può essere calcolato direttamente dall'ammontare di reddito disponibile: ad esempio dal reddito lordo (G_i) rilevato per la componente di reddito i sottoposta all'imposizione contributiva. In questo caso si ha: $H_i = G_i - S_i(G_i)$. In altri casi è necessaria una procedura iterativa, piuttosto semplice che di solito converge velocemente poiché viene applicata separatamente per ogni componente di reddito e non vi sono altri parametri da stimare. L'iterazione numerica è difatti necessaria soltanto perché la quantità non nota da determinare (H_i) appare in un'equazione implicita.

Prospetto 2.4 - Calcolo di H_i in base alla componente di reddito disponibile

Set H (a)			
Valore dato	X_{S_i}	$H_i = X_{S_i}$	Una semplice iterazione, in genere una per ogni componente di reddito
	G_i	$H_i = G_i - S_i(G_i)$	
$P_i =$	X_{T_i}	$H_i = G_i - S_i(G_i)$ dove $G_i = X_{T_i} + T_i(H_i)$	
	X_{TS_i}	$H_i = X_{TS_i} + T_i(H_i)$	
Set N (b)			
Valore dato	N_i	$H_i = Y_i + D_i(H_i)$	Doppia iterazione (i) con valore definito di R, per ciascuna componente di reddito (ii) per determinare R, aliquota comune a tutte le componenti di reddito
$P_i =$		dove $Y_i = [H_i - N_i + C_i(Y_i)] / R$	

(a) Tutte le componenti di reddito al lordo della tassazione o raccolte in forma diversa dal "netto finale".

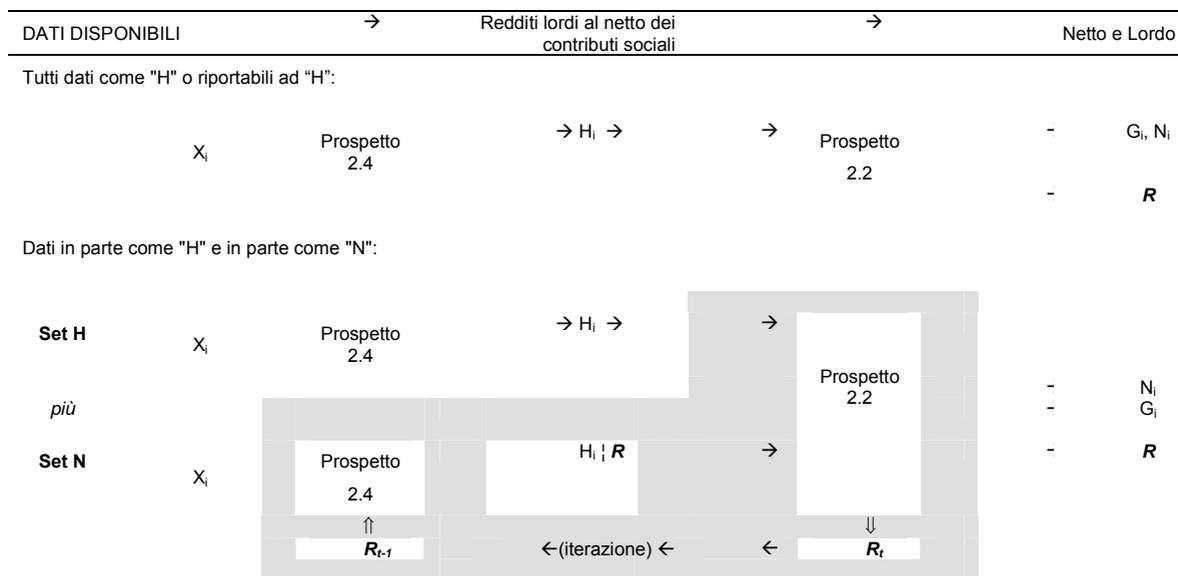
(b) Componenti di reddito sottoposte alla tassazione (con o senza contribuzione sociale) e per le quali "l'ammontare netto finale" ($X_i = N_i$) risulta specificato nei dati rilevati.

Se i redditi rilevati con l'indagine campionaria o altra fonte sono al netto dell'imposizione fiscale e contributiva, il procedimento inverso che va dai redditi netti N_i ai redditi lordi H_i è notevolmente più complesso, in quanto implica un doppio *loop* iterativo. Il primo *loop* interno applica un valore definito del parametro "aliquota di tassazione" (R , come definito nel prospetto 2.2) a ogni componente di reddito e per tutti gli individui della stessa unità fiscale. Successivamente un secondo *loop* iterativo ottiene un valore convergente del parametro R comune a tutte le componenti di reddito.

La procedura iterativa di SM2 è sintetizzata nel prospetto 2.5. La procedura distingue tra l'insieme dei redditi lordi H e redditi netti N così come sono stati definiti nel prospetto 2.4. Le quantità richieste H_i per il set H sono calcolate una sola volta utilizzando le equazioni del prospetto 2.4; successivamente le quantità calcolate H_i costituiscono i valori di input nel ciclo iterativo per il parametro R richiesto per il calcolo del set N . Il parametro R viene così stimato utilizzando tutte le informazioni disponibili derivanti da entrambi i set H e N .

La procedura iterativa dai redditi netti ai lordi può essere tuttavia affetta da due problemi comuni alle tecniche di microsimulazione: ossia la non convergenza e la convergenza multipla dei valori. Per non convergenza si intende che partendo da un valore dato di reddito netto non si ottiene nessun valore di reddito lordo principalmente per due motivi: (i) non esiste un valore lordo a causa di alcune caratteristiche specifiche dei dati, delle regole applicate o di errori nel calcolo delle deduzioni e detrazioni in genere stimate sulla base di informazioni esogene rispetto ai dati disponibili; (ii) può accadere che anche se esiste una soluzione, le *routines* di conversione del modello di microsimulazione non riescono a convergere con un numero "accettabile" (predefinito) di iterazioni.

Prospetto 2.5 - Struttura generale della procedura iterativa



In caso di non convergenza, la procedura iterativa di SM2 calcola comunque un valore di reddito lordo e restituisce un valore di reddito netto stimato il più vicino possibile al valore netto rilevato. Si può così utilizzare il coefficiente del reddito netto (rilevato/stimato) per modificare il reddito lordo stimato in modo proporzionale per ogni componente di reddito. L'aggiustamento richiesto è di solito molto contenuto.

Per individuare casi di convergenza multipla, le *routines* di conversione di SM2 introducono una piccola perturbazione stocastica nell'aliquota di tassazione R , in modo da verificare se vi sia o meno una "convergenza locale", ossia se vi siano più valori di reddito lordo a cui corrisponde un valore di reddito netto stimato identico a quello rilevato. In presenza di convergenza multipla va scelto un criterio "ragionevole" per scegliere la soluzione più adeguata.

2.4.2 - Le deduzioni e detrazioni specifiche: uno strumento per gestire la complessità

Una delle caratteristiche più rilevanti di SM2 è che con la definizione di alcune "deduzioni specifiche" e "detrazioni specifiche d'imposta", molte caratteristiche peculiari dei diversi regimi fiscali possono essere incluse nelle procedure standard descritte precedentemente senza alterare in alcun modo la struttura generale.

Diversamente dalle "comuni" deduzioni e detrazioni d'imposta, le deduzioni e detrazioni specifiche riguardano alcune componenti di reddito e consentono di includere in SM2 le diverse modalità di tassazione nazionali. Nel prospetto 2.6 è elencata una casistica di deduzioni e detrazioni specifiche adatte ad alcuni sistemi fiscali europei e in particolare all'Italia, Francia e Spagna.

Si consideri una situazione molto comune in cui una particolare componente di reddito è esente da tassazione (ad esempio i benefici familiari) e le restanti componenti di reddito sono sommate e sottoposte a tassazione (rigo 1 del prospetto 2.6). In questo caso è sufficiente inserire una deduzione specifica per la componente esente da tassazione pari al reddito lordo al netto dei contributi sociali ossia: $D_i = H_i$; in questo modo non si considera il contributo di questa componente nel totale del reddito sottoposto a tassazione e soprattutto non si deve separare tale componente dagli altri redditi.

Allo stesso modo se una componente di reddito è tassata con un'aliquota costante (f_i) a differenza delle restanti componenti di reddito (rigo 2), si introduce una deduzione specifica pari al reddito lordo al netto dei contributi sociali ($D_i = H_i$) e una detrazione specifica definita come una quantità negativa ossia: $C_i = -f_i \times H_i$. In tal modo la componente tassata con aliquota costante (ad esempio il reddito da attività finanziaria) non contribuisce all'imponibile sottoposto a tassazione corrente, ma l'imposta finale è automaticamente incrementata dell'ammontare appropriato.

Prospetto 2.6 - Alcuni esempi di deduzioni e detrazioni specifiche del modello SM2 (a)

TIPO DI TASSAZIONE DELLA COMPONENTE i (a)	Deduzione specifica	Detrazione specifica
1 Esente da imposta	$D_i = H_i$	-
2 Tassazione ad aliquota costante f_i	$D_i = H_i$	$C_i = -f_i \cdot H_i$
3 Esente da imposta ad aliquota costante f_i	-	$C_i = +f_i \cdot H_i$
4 Oneri deducibili	+ deduzioni comuni	-
5 Oneri detraibili	-	+ detrazioni comuni
6 Tassazione speciale non relative al reddito	-	- detrazioni comuni
7 Doppia tassazione ad aliquota costante f_i	-	$C_i = -f_i \cdot H_i$
8 Parte ΔS_i dei contributi sociali sottoposti a tassazione	$\sim \Delta S_i$	-

(a) Differenti forme di tassazione possono essere previste per i casi 3, 4 e 7: essendo l'aliquota di tassazione una funzione più generale dell'ammontare di reddito.

Gli oneri deducibili e detraibili per spese sostenute dal contribuente sono considerati nel modello come deduzioni comuni (rigo 4) e detrazioni comuni (rigo 5) applicabili al totale dei

redditi imponibili. I casi di doppia tassazione con aliquota costante, come ad esempio nel sistema fiscale italiano in cui il reddito da lavoro autonomo è sottoposto all'Imposta regionale sulle attività produttive (Irap) e all'Irpef, sono trattati nel modello con l'introduzione di una detrazione d'imposta specifica negativa (rigo 7).

Infine, l'ultimo esempio (rigo 8) rappresenta un caso particolare previsto nel sistema fiscale francese in cui una parte dei contributi sociali è sottoposta a tassazione. In SM2 la parte sottoposta a tassazione è considerata come una deduzione negativa che quindi effettivamente si aggiunge al reddito lordo al netto dei contributi sociali ($H_i = G_i - S_i$). In tal modo il totale dei redditi imponibili è incrementato per la sola parte dei contributi sociali sottoposta a tassazione, ossia: S_i ; $Y_i = H_i + S_i$.

2.5 - La sperimentazione e l'implementazione del modello SM2-Eu-Silc

Il modello SM2 è stato inizialmente predisposto e testato sulla base dei dati dell'indagine Echp disponibile per tre paesi dell'Unione europea: Francia, Italia e Spagna. In SM2 il file di input e le variabili ausiliarie sono costruite sui dati Echp-Udb-Pdb⁵ dell'anno 1998 e identificano per i tre paesi a cui è stato applicato il set di variabili ausiliarie, in base alla normativa fiscale nazionale, necessarie alla costruzione delle variabili target di reddito individuale e familiare previste dal progetto Eu-Silc.

I principali risultati della sperimentazione sono sintetizzati nei prospetti seguenti che riportano la bontà dell'adattamento delle stime ottenute con il modello SM2 ai dati di Contabilità nazionale forniti dagli istituti nazionali di statistica di Francia, Italia e Spagna.

In particolare, il prospetto 2.7 mostra il confronto tra le stime ottenute con il modello SM2 applicato ai dati francesi dell'Echp e le stime fornite dall'Istituto nazionale di statistica francese (Insee).

Prospetto 2.7 - Confronto tra le stime del modello SM2 applicato all'indagine Echp in Francia e i risultati pubblicati dall'Insee - Anno 2003

COMPONENTI DI REDDITO	Composizione del reddito lordo		Comparazione	
	Media	Valori percentuali	SM2	Insee
	(1)	(2)	(4)	(5)
[1] Netto+imposta familiare	67.858	67,2		
[1a] netto		66,4	86,6	85,7
[1b] imposta familiare		0,8 ←	1,1 ←	1,1
[2] Imposta personale sul reddito	4.086	4,0	5,3	5,9
[3] Contributi previdenziali CSG,CRDS, SI sul reddito da capitale	5.431	5,4	7,0	7,3
Subtotale [1]+[2]+[3] ('revenu avant imports')	77.375	76,7	100,0	100,0
[4]Altri contributi sociali individuali	6.180	6,1	Media/famiglia (euro)	
[5] Contributi sociali del datore di lavoro	17.397	17,2	30.200	29.856
Totale dei contributi sociali	29.008	28,7		
Totale lordo	100.952	100,0		

Fonte: Insee: Insee Premiere n. 916, Agosto 2003. "Des ménages modestes aux ménages aisés: des sources de revenus différentes".
SM2: Francia Echo Onda 6 SM2 Model

⁵ Echp Users' Data Base (Udb) e Production Data base (Pdb).

Il confronto è stato possibile non sull'intero reddito lordo, ma sul reddito imponibile, inclusi alcuni contributi previdenziali a carico del contribuente (Csg, Crds, contributi su reddito da capitale) considerati come vere e proprie imposte nel sistema fiscale francese.

Posto pari a 100 il reddito imponibile, il prospetto 2.7 mostra un ottimo adattamento della distribuzione del reddito (e delle sue componenti principali) stimato con SM2 a quello proveniente dalla Contabilità nazionale francese.

Per quanto riguarda l'Italia, SM2 applica ai dati Echp del 1998 il sistema di tassazione dell'anno 2003, in modo da includere il primo modulo della 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 e delle addizionali regionali e comunali.

Nel prospetto 2.8, il reddito lordo è disaggregato in contributi sociali, imposte e reddito netto. Il confronto con i dati di Contabilità nazionale mostra un buon adattamento dei risultati di SM2: i contributi sociali dei lavoratori dipendenti appaiono, tuttavia, un po' sovrastimati rispetto ai dati Istat e anche le imposte risultano sovrastimate dell'1,3 per cento.

Prospetto 2.8 - Confronto tra le stime del modello SM2 applicato all'indagine Echp in Italia e i risultati pubblicati dall'Istat - Anno 2003

COMPONENTI DI REDDITO	SM2 (dati: Echp 1998 tassazione: 2003)		Istat	Errore (%)
	Euro	Valori percentuali		
Reddito al lordo dei contributi sociali dei lavoratori	8.880	100,0	100,0	
- Contributi sociali dei lavoratori dipendenti	416	4,7	3,6	1,1
- Contributi sociali dei lavoratori autonomi	202	2,3	1,8	0,5
Reddito lordo al netto dei contributi sociali dei lavoratori	8.261	93,0	94,6	(1,6)
Imposta personale del reddito e sulle attività finanziarie	1.044	11,8	13,0	(1,3)
Reddito netto	7.217	81,3	81,6	(0,3)

Fonte: Istat, Conti nazionali (1998); SM2: Italia Echp Onda 6

Nel caso della Spagna, la validazione delle stime microsimulate mediante il confronto con i dati pubblicati dall'Oecd mostra un buon risultato (Prospetto 2.9).

Prospetto 2.9 - Confronto tra le stime del modello SM2 applicato all'indagine Echp in Spagna e i risultati pubblicati dall'Oecd - Anni 1997 e 1999

COMPONENTI DI REDDITO	SM2 (Echp 1999)		Oecd (1997)		
	Media (a)	Valori percentuali	(1) Quota presunta di Reddito individuale/Pil 0.67	(2) 0.69	(3) 0.71
Reddito al lordo dei contributi sociali	1.377.492	100,0	100,0	100,0	100,0
Contributi sociali dei datori di lavoro	157.072	11,4	12,4	12,0	11,7
Contributi sociali dei lavoratori dipendenti	47.287	3,4	2,8	2,7	2,7
Contributi sociali dei lavoratori autonomi	30.450	2,2	2,4	2,3	2,3
Reddito lordo al netto dei contributi sociali dei lavoratori	1.142.683	83,0	82,4	83,0	83,3
Imposta personale sul reddito	145.647	10,6	11,0	10,7	10,4
Reddito netto	997.040	72,4	71,4	72,2	73,0

Fonte: Oecd: Revenue Statistics, Parigi (1999); SM2: Spagna Echp Onda 7 (a) Pesetas 1999 pro capite.

Il modello SM2 è stato reso disponibile da Eurostat per tutti i paesi dell'Unione europea e per quei paesi extra-Ue nei quali è condotta l'indagine Eu-Silc. La logica che sta alla base del modello SM2 è stata adottata dagli istituti nazionali di statistica di Spagna e Portogallo. L'implementazione integrale della modellistica è stata invece adottata in Italia, Grecia e Turchia. La Macedonia ha appena iniziato la messa a punto del modello sulla base della prima onda dell'indagine Eu-Silc rilevata nel maggio 2010.

Di seguito nei prospetti 2.10 e 2.11 sono riportati i risultati delle stime microsimulate per l'Italia e la Turchia utilizzando come base dati del modello SM2 la nuova indagine Eu-Silc. Per quanto riguarda l'Italia, lo sviluppo del modello SM2-Eu-Silc e il processo di produzione dei redditi lordi basato sull'utilizzo congiunto delle tecniche di microsimulazione e dell'utilizzo degli archivi amministrativi è spiegato dettagliatamente nel Capitolo 4.

Prospetto 2.10 - Confronto tra le stime del modello SM2 applicato all'indagine Silc in Italia e i risultati pubblicati dall'Istat - Anno 2003

COMPONENTI DI REDDITO	SM2 (Dati Eu-Silc 2003)	Istat	Errore (%)
Reddito al lordo dei contributi sociali	100,0	100,0	
Contributi sociali	15,9	15,7	0,2
- Dei datori di lavoro	9,9	11,4	-1,5
- Dei lavoratori dipendenti	3,5	2,8	0,7
- Dei lavoratori autonomi	2,5	1,6	0,9
Reddito lordo al netto dei contributi sociali dei lavoratori	84,1	84,3	-0,2
Imposta personale sul reddito e sulle attività finanziarie	12,8	12,0	0,8
Reddito netto	71,3	72,2	-0,9

Fonte : Istat, Conti nazionali (2003); SM2: Italia Eu-Silc Onda 1

Prospetto 2.11 - Confronto tra le stime del modello SM2 applicato all'indagine Silc in Turchia e i risultati pubblicati dal Turkstat - Anno 2006

COMPONENTI DI REDDITO	Conti nazionali	Silc
GG - Reddito lordo	100.000	100.000
Totale dei contributi sociali	8.801	8.849
SS - Contributi sociali dei datori di lavoro		4.757
S - Contributi sociali dei lavoratori dipendenti e autonomi		4.092
G - Reddito lordo al netto dei contributi sociali dei datori di lavoro	91.199	91.151
Totale imposte	6.869	7.451
Imposta sul reddito da lavoro dipendente		6.086
Imposta sul reddito da lavoro autonomo	0,384	0,442
Imposte dul reddito da capitale finanziario		0,689
Imposte dul reddito da capitale reale		0,001
Damga		0,232
N - Reddito netto disponibile	84.331	83.699

Fonte: Ballini et al. (2009)

Capitolo 3 - Strategie di rilevazione dei redditi in Eu-Silc: l'uso congiunto di fonti amministrative e campionaria

3.1 - Introduzione

L'esperienza italiana sul progetto Eu-Silc include nel processo di rilevazione e misurazione delle variabili di reddito elementi di originalità che la contraddistinguono rispetto ai restanti istituti statistici europei. La peculiarità del progetto italiano consiste nell'utilizzo combinato di informazioni amministrative e campionarie sui redditi, attraverso una strategia che sfrutta la *record linkage* tra dati individuali di reddito rilevati, rispettivamente, da un'indagine campionaria Papi (Paper and pencil face-to-face interview)¹ e da fonti di natura amministrativa (Agenzia delle entrate e Inps). I vantaggi che si conseguono, applicando una simile strategia di raccolta dati, sono principalmente due: migliore copertura delle informazioni e maggiore accuratezza nella misurazione del dato.

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 del reddito dell'indagine), l'effetto di interazione intervistato-intervistatore, la reticenza a dichiarare il vero, costituiscono alcuni degli esempi dei potenziali errori insiti nelle indagini campionarie che utilizzano come unico strumento di rilevazione l'intervista. D'altra parte, un'indagine come quella Eu-Silc che ammetta il ricorso all'uso di risposte *proxy*, non può che perdere in qualità dell'informazione quando a rispondere è un soggetto diverso rispetto all'unità di osservazione. A ciò si aggiungono tipologie di errori di natura non campionaria, come quelli imputabili a registrazione, codifica, autoselezione dei rispondenti eccetera.

In questo contesto, la possibilità di integrare basi di dati campione e amministrative rappresenta un importante disegno per migliorare la qualità dell'informazione statistica. Questo tipo di approccio, dapprima sperimentato su alcune componenti di reddito, è stato in seguito applicato con successo nell'ambito delle prime quattro *waves* dell'indagine italiana Eu-Silc. A fronte di un guadagno in termini di qualità del dato, l'approccio in questione richiede, tuttavia, il soddisfacimento di alcuni importanti requisiti, tipici dell'integrazione delle informazioni da molteplici fonti.² In particolare, occorre preliminarmente garantire che: le unità statistiche siano definite allo stesso modo nelle varie fonti del dato; tutte le fonti facciano riferimento allo stesso collettivo statistico (copertura); il contenuto informativo delle variabili della fonte "donatrice" sia identico a quello delle variabili "obiettivo" dell'indagine Eu-Silc (armonizzazione delle variabili e delle classificazioni).

In questo capitolo verrà illustrato il processo di costruzione del dato relativo ai redditi dell'indagine italiana Eu-Silc, focalizzando l'attenzione sulle soluzioni metodologiche adottate per superare i vincoli relativi all'integrazione delle fonti campionaria e amministrativa, nonché i problemi connessi alle fasi di *editing* e imputazione.

Il capitolo è a cura di Paolo Consolini; il paragrafo 3.5.2 è a cura di Paolo Consolini, Clodia Delle Fratte, Davide Di Laurea e Roberta Ricci

¹ Nel 2011 è prevista la transizione alla tecnica di rilevazione Capi (Computer Assisted Personal Interview).

² Si veda: van der Laan P. "Integrating administrative registers and household surveys". *Netherlands Official Statistics*, Vol. 15, Summer 2000.

3.2 - Le fasi operative del processo di integrazione dei dati di fonti diversa

L'impianto metodologico su cui poggia il processo di integrazione tra i dati della rilevazione Papi e i dati di fonte amministrativa in Eu-Silc può essere riassunto nelle seguenti sei fasi operative (si veda anche la figura 3.1):

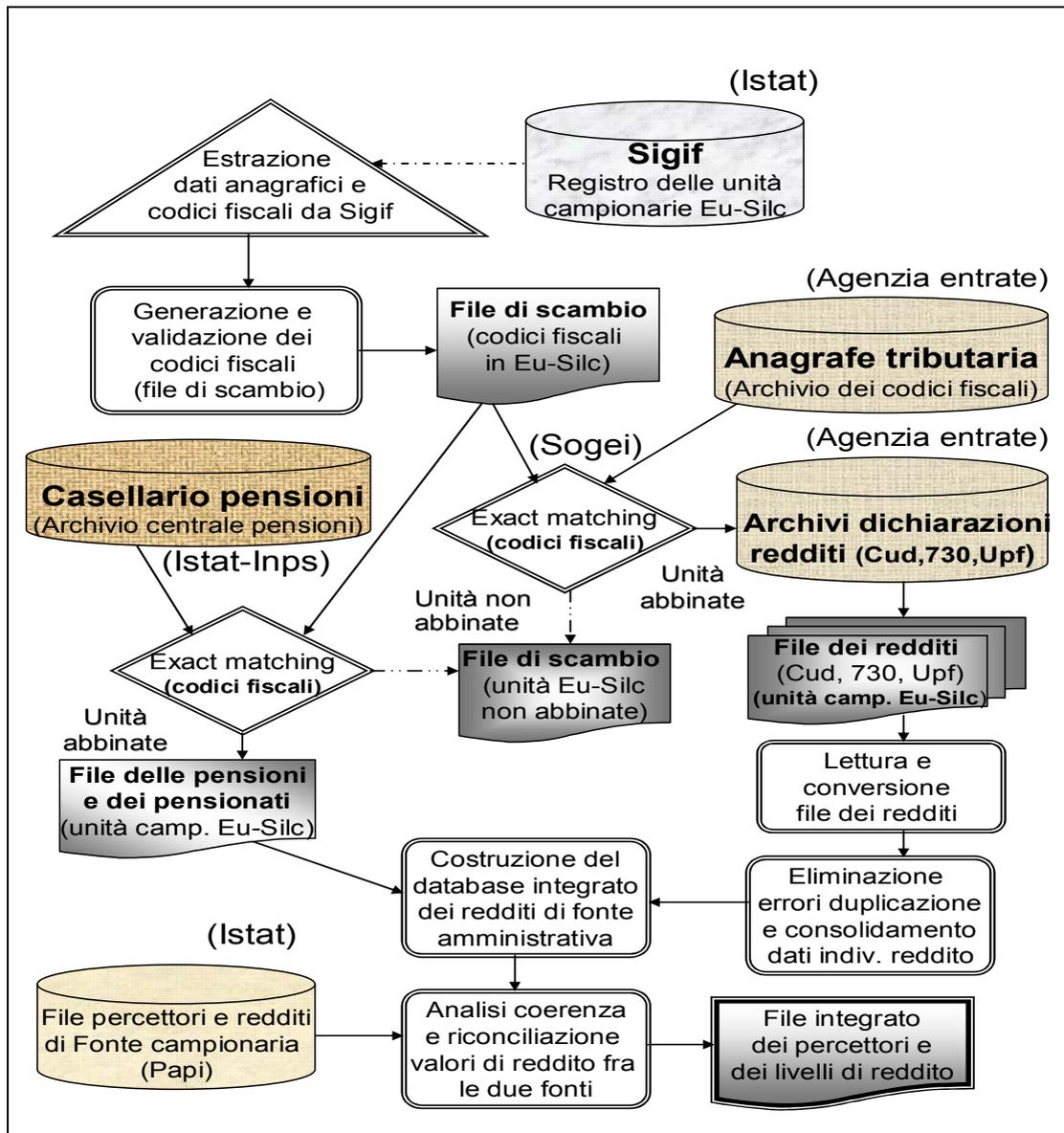
- 1) *Input dei dati campionari*: la fase inizia con il recupero delle informazioni individuali dal registro familiare e individuale Sigif (Sistema integrato di gestione dell'indagine sulle famiglie)³ per il calcolo del codice fiscale della persona fisica (*matching key*) presunto⁴ o rilevato; prosegue con un'analisi della correttezza e compatibilità dei codici fiscali rilevati in Sigif con i codici presunti; e si conclude con la preparazione del file standard per lo scambio e la trasmissione dei dati con la Sogei.⁵ Sempre in questa fase ricade l'attività di caricamento delle informazioni individuali sul reddito rilevate dal questionario;
- 2) *Input dei dati amministrativi*: in questa fase si predispongono e si aggiornano il programma di lettura e conversione dei database fiscali di Sogei precedentemente definiti (selezione delle unità e delle variabili di interesse). Successivamente si implementa un piano di identificazione e correzione dei record duplicati, con relativo consolidamento delle informazioni relative allo stesso individuo riportate su più record;
- 3) *Armonizzazione delle definizioni e classificazione dei redditi di fonte amministrativa*: per poter confrontare le informazioni riportate nelle fonti amministrative con quelle campionarie è necessario che le prime vengano armonizzate e classificate in base ai principi con cui sono costruite le seconde, cioè secondo le definizioni Eurostat di variabile target;
- 4) *Record linkage tra dati campionari e amministrativi*: in questa fase le unità che compongono il database amministrativo, definito al passo precedente, vengono abbinate alle unità che hanno preso parte all'indagine campionaria tramite la chiave identificativa del codice fiscale;
- 5) *Individuazione e soluzione delle incoerenze sulla tipologia di percettori fra le due fonti*: una volta stabilito un abbinamento tra le unità di fonte campionaria e amministrativa può accadere che una stessa unità possieda redditi di diversa natura a seconda della fonte, ovvero che i dati, riportati nei due giacimenti informativi, presentino delle incoerenze. Per risolvere tali incoerenze si è definito un sistema di regole che stabilisce come attribuire le componenti di reddito a livello di singola unità;
- 6) *Riconciliazione dei valori difformi di reddito fra le due fonti*: pur in presenza di una struttura di reddito "coerente" fra le fonti del dato è possibile che una stessa unità statistica presenti valori di reddito "difformi" tra le due fonti. In quest'ultima circostanza si tenta di riconciliare i valori facendo opportune assunzioni in merito alla reticenza degli intervistati a dichiarare il vero, agli errori di misurazione, alla sottostima del reddito fiscale dichiarato (elusione, evasione).

³ 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).

⁴ Per codice fiscale "presunto" si intende il codice desumibile dalle informazioni relative a: nome, cognome, sesso, data e luogo di nascita del soggetto. Quest'ultimo di norma coincide con il codice fiscale effettivo attribuito a quel soggetto nell'Anagrafe tributaria. Tuttavia in casi rari (stimabili intorno al 2 per mille) può accadere che due o più soggetti possiedano informazioni anagrafiche tali da generare un medesimo codice fiscale (cosiddetti "omocodici") per cui il sistema dell'Anagrafe tributaria attribuirà loro un codice fiscale distinto rispetto al presunto.

⁵ La Sogei (Società generale di informatica) è parte integrante del sistema "Fisco", per il quale svolge la funzione di ente strumentale preposto al settore dell'Information and Communication Technology (ICT) con la missione di sviluppare e gestire il sistema informativo della fiscalità.

Figura 3.1 - Le fasi del processo di integrazione delle fonti campionaria e amministrativa in Eu-Silc



3.3 - Generazione e validazione dei codici fiscali (presunti)

Per legare insieme dati individuali provenienti da diverse fonti occorre predisporre una procedura, deterministica o probabilistica, che consenta di identificare quali coppie di record fra due o più basi di dati afferiscono alle stessa unità.⁶ Nell'approccio deterministico una coppia di record si definisce "abbinata" se i due record concordano esattamente su ciascun elemento

⁶ Cfr. Newcombe H.B. *Handbook of Record Linkage: Methods for Health and Statistical Studies, Administration, and Business*. Oxford (UK): Oxford University Press, 1988.

costitutivo all'interno di un collettivo di identificativi denominati *match key*.⁷ Questo tipo di approccio è quello generalmente seguito quando si dispone di identificativi univoci e accurati, come ad esempio i codici fiscali. La mancanza di identificativi unici o poco affidabili fa preferire, invece, approcci di tipo probabilistico che rimandano all'individuazione di una regola ottimale che tiene sotto controllo il livello degli errori generati da una determinata scelta di abbinamento tra record.⁸ La disponibilità, nell'ambito dell'indagine Eu-Silc, di codici fiscali (persone fisiche) rilevati presso le anagrafi dei comuni campione, oppure ricavabili dalle caratteristiche demografiche degli individui registrati (codici fiscali presunti) ha orientato la scelta verso una strategia deterministica di *record linkage*. Sebbene l'approccio probabilistico possa essere applicato, in modo residuale, al sottoinsieme di unità non abbinate tramite chiave esatta (strategia mista), esso in realtà non è adottato in quanto la Sogei utilizza esclusivamente i codici fiscali (o in via residuale: nominativo-residenza), quale chiave esatta di *matching* fra gli archivi tributari e il *frame* di Eu-Silc. I risultati conseguiti nel tempo sul numero di unità del campione che si accoppiano con successo ai codici all'Anagrafe tributaria (collettivo degli identificativi) mostrano la validità dell'approccio utilizzato.

Il calcolo del codice fiscale ha previsto una fase preliminare di lettura, correzione e imputazione dell'informazioni di base contenute in Sigif (Sistema gestione indagini famiglie). Il sistema informatico-informativo denominato Sigif, istituito presso la Direzione centrale per le indagini su condizione e qualità della vita (Dccv) dell'Istat, gestisce e monitora i flussi informativi tra il sistema centrale, i comuni e i rilevatori. Una volta ricevuti dai comuni campione i nominativi delle famiglie selezionate (campione teorico) il sistema Sigif acquisisce una serie di dati anagrafici sulle persone appartenenti a tali nuclei che sono alla base della creazione dei rispettivi codici. Per generare questi ultimi è tuttavia necessario disporre di informazioni corrette. Tale condizione in pratica non è sempre soddisfatta, per cui il codice presunto che si ottiene, applicando una qualunque procedura automatica a partire dai dati elementari inesatti presenti in Sirc (Sistema informativo della rete di rilevazione comunale), non può che essere mancante o errato (*garbage in, garbage out*). Per altro verso, nulla assicura il perfetto allineamento tra codice fiscale rilevato nel sistema Sirc e codice presunto generabile sulla base dei dati elementari. In presenza di valori difformi si preferisce il codice rilevato al presunto, sempre che il primo non assuma valori inammissibili. Infine, nel caso in cui ambedue le modalità, rilevazione diretta e procedura automatica, non siano in grado di ottenere codici fiscali formalmente validi si demanda alla Sogei (Società generale dell'informatica) il compito di recuperare il dato attraverso altre chiavi di *matching* (ad esempio: nominativo, comune di residenza, indirizzo dell'abitazione).

La procedura di generazione del codice fiscale è stata perfezionata nel corso del tempo. In particolare, nella prima edizione dell'indagine Eu-Silc (anno 2004) il codice fiscale è stato desunto dalle sole informazioni correnti contenute nel sistema Sigif. Mentre per le *wave* successive è stata utilizzata, in aggiunta, la lista dei codici fiscali (persone fisiche) validati nell'anno precedente ($T-1$) dalla Sogei.⁹ In essa sono compresi i soggetti con codici fiscali validati al tempo $T-1$ che dovranno essere reintervistati all'anno corrente T .

L'assegnazione dei codici fiscali per l'indagine Eu-Silc 2007 è stata quindi sviluppata separatamente su quattro segmenti del campione (Figura 3.2):

a) l'insieme di record riferiti a soggetti con codici fiscali validati al tempo $T-1$ (40.594 casi);

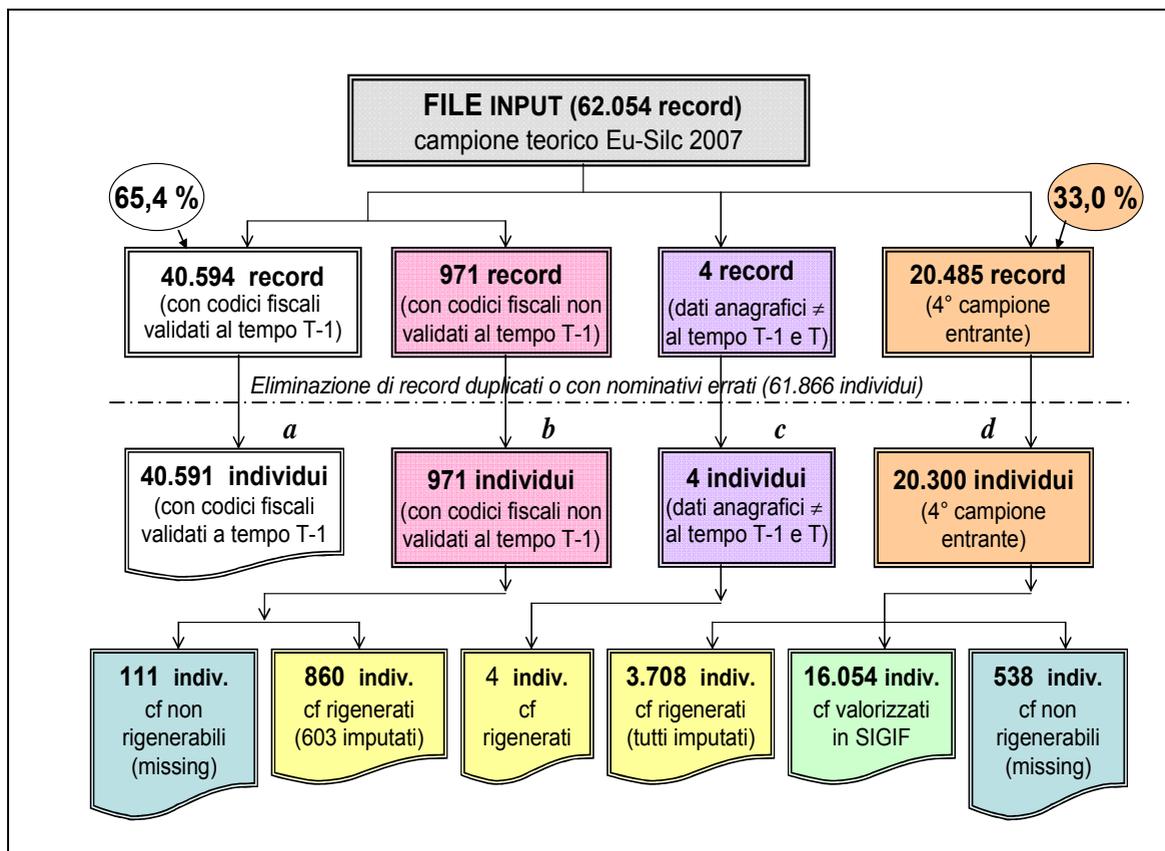
⁷ Herzog T.N., F.J. Scheuren e W.E. Winkler. *Data quality and record linkage techniques*. New York, 2007.

⁸ Cfr. Fellegi I.P. e A.B. Sunter "A theory of record linkage". *Journal of the American Statistical Association*, vol. 64, pp.1183-1210, 1969; Scanu M. *Metodi statistici per il record linkage*. Roma: Istat, 2003. (Metodi e Norme, n. 16).

⁹ Questa scelta si è resa necessaria in quanto Sigif non prevedeva l'aggiornamento dei codici fiscali sulla scorta delle informazioni raccolte e validate negli anni passati.

- b) l'insieme di record associati a soggetti con codici fiscali non validati al tempo $T-1$ con identica struttura anagrafica al tempo T (971 casi);
- c) l'insieme di record associati a soggetti con codici fiscali non validati al tempo $T-1$ che cambiano struttura anagrafica al tempo T (4 casi);
- d) l'insieme di record riguardanti i soggetti del 4° campione entrante di Eu-Silc (20.485 casi).

Figura 3.2 - Struttura del processo di assegnazione dei codici fiscali a partire dal campione teorico dell'indagine Eu-Silc 2007



Per il raggruppamento *a* sono stati semplicemente attribuiti i codici fiscali riconosciuti come validi nella precedente indagine, eliminando i possibili duplicati (3 record). Ai fini dell'assegnazione dei codici per gli altri tre segmenti si è analizzato *ex ante* il contenuto informativo dei relativi record. Nell'edizione 2007 la base dati inerente ai gruppi *b*, *c* e *d* era costituita da 21.275 record, ma solo 20.626 disponevano di informazioni sufficienti a generare un codice fiscale.

Prospetto 3.1 - Profilo degli errori e piano di controllo/correzione in fase di generazione dei codici fiscali (indagine Eu-Silc 2007)

CARATTERISTICHE ARCHIVIO SIGIF-EUSILC_07	Numero di record	Numero record imputabili da Codice fiscale	Note
Data-set in ingresso	62.054	-	
- Dati con codici fiscali validati al tempo T-1 (1° output)	40.594	-	Da includere direttamente nel file di output
- Dati con codici fiscali non validati al tempo T-1	971	-	Da rigenerare nuovamente
- Dati con anagrafica al tempo T differente rispetto al tempo T-1	4	-	Da rigenerare nuovamente
- Dati relativi al 4° di campione entrante	20.485	-	- 4.554 codici fiscali mancanti - 15.251 codici fiscali valorizzati
Data-set in ingresso per la generazione del codice fiscale	21.460	-	Input della procedura di generazione del codice
Duplicati rispetto alla chiave: "codice_fiscale-cognome-nome-data_nascita"	(-) 0	-	Eliminazione
Dati con Cognome e/o Nome <i>mancanti</i> o <i>non meglio specificati</i>	(-)185	-	Eliminazione
Data-set potenzialmente idoneo a valorizzare un codice fiscale	21.275	-	
Dati con valori <i>mancanti</i> nel: Codice_Istat_comune/Stato estero	906	272	Imputazione tramite il codice fiscale di Sigif
Dati con valori <i>mancanti</i> in Codice_Istat comune/Stato estero e codice territoriale non assegnato	(-)634	-	Eliminazione
Dati con valori <i>errati</i> in Codice Codice_Istat_comune/Stato estero	158	150	Imputazione codice territoriale tramite codice fiscale di Sigif
Dati con valori <i>errati</i> in Codice Codice_Istat_comune/Stato estero e codice territoriale non assegnato	(-)8	-	Eliminazione
Dati con valori <i>mancanti</i> in: "sesso dell'individuo"	1.285	1.278	Imputazione tramite "codice fiscale" e "Nome" di Sigif
Dati con valori <i>mancanti</i> nel codice fiscale e sesso	(-)7	-	Eliminazione
Dataset con informazioni tali da generare un codice fiscale pieno	20.626	-	

Il prospetto 3.1 descrive le azioni intraprese nel piano di *editing* e imputazione sui dati elementari di Sigif per l'anno 2007 in presenza di errori, di duplicati o di valori mancanti. All'interno della base dati utilizzata per la procedura di generazione sono stati imputati 4.568 codici fiscali, dei quali 603 afferenti al segmento *b* "soggetti con codici fiscali non validati al tempo T-1 e dati anagrafici immutati nei due anni" e altri 3.708 al segmento *d* "4° campione entrante". La stessa procedura ha, inoltre, generato 16.315 codici fiscali (presunti) riferiti a soggetti il cui codice era già presente in Sigif; 16.058 dei quali appartenenti al 4° campione entrante e altri 257 riferiti a persone con codici non validati nell'anno precedente. In 508 casi sono stati creati codici fiscali diversi rispetto a quelli rilevati dai comuni, mentre in altri 8 la

procedura non è stata in grado di produrre valori a fronte di codici già presenti in Sigif.¹⁰ In caso di incoerenza fra codice fiscale presunto (generato automaticamente) e codice rilevato dal comune (Sirc) si è preferito il secondo, sempre che il suo valore fosse compatibile con i dati elementari (caratteristiche individuali: età, sesso, nome e cognome) e ammissibile rispetto alla lista dei codici territoriali.¹¹ Infine, in appena 649 casi non è stato possibile attribuire alcun codice fiscale alla persona fisica per assenza di informazioni di base rilevate dai comuni. In definitiva, il file di scambio inviato alla Sogei per la validazione dei codici fiscali relativi alle unità del campione teorico Eu-Silc 2007 comprendeva 61.866 record (distinti in base alla chiave “progressivo componente”).

Nei quattro anni di indagine 2004-2007 si è potuto osservare un costante miglioramento nell’accuratezza e nella copertura delle informazioni relative ai codici fiscali ottenuti tramite le procedure di calcolo automatico o rilevati direttamente dai comuni (Prospetto 3.2). I risultati del processo di validazione della Sogei sui codici fiscali trasmessi dall’Istat mostrano che la percentuale di codici coerenti (codificati con “0”) cresce costantemente nel tempo, passando dal 92,1 per cento per l’anno 2004 al 95,7 per cento dell’anno 2005, sino a giungere al 96,2 per cento e il 96,7 per cento rispettivamente negli anni 2006 e 2007. Parallelamente nel corso degli stessi anni, risulta più che dimezzata la percentuale dei codici fiscali “errati o mancanti” non altrimenti recuperabili: dal 6,0 per cento di inizio indagine al 2,4 per cento dell’edizione 2007. In altri termini, la percentuale di unità del campione che nel periodo 2004-2007 trovano un abbinamento esatto (univoco) con i codici fiscali dell’Anagrafe tributaria sale di circa 3,5 punti percentuali, passando dal 94,0 per cento al 97,6 per cento.

Prospetto 3.2 - Risultati del processo di validazione dei codici fiscali da parte della Sogei relativi all’indagine Eu-Silc - Anni 2004-2007

RISULTATI	Anni dell’indagine			
	2004 (% di casi)	2005 (% di casi)	2006 (% di casi)	2007 (% di casi)
Coerenti (“0”)	92,1	95,7	96,2	96,7
Errati ma recuperati dalla Sogei (“1”)	0,3	0,3	0,2	0,1
Mancanti ma imputati dalla Sogei (“2”)	1,6	0,8	1,1	0,8
Errati o mancanti irrecuperabili (“9”)	6,0	3,2	2,6	2,4
Totale codici fiscali	100,00	100,00	100,00	100,00

3.4 - Variabili di reddito: armonizzazione del contenuto e delle definizioni tra le fonti del dato

3.4.1 - Redditi da lavoro autonomo¹²

Le difficoltà insite alla rilevazione e misurazione del reddito da lavoro autonomo tramite indagine campionaria sulle famiglie sono bene espresse nel manuale di Camberra dalle due seguenti citazioni:

¹⁰ Dal momento che questi codici risultavano non validati al tempo *T-I* ne è stato rimosso il relativo valore (*missing value*).

¹¹ Il codice territoriale è associato al luogo di nascita dell’individuo, sia esso comune italiano o stato estero, ed è formato da una quaterna di caratteri (in posizione 12^a, 13^a, 14^a, 15^a) composta da una lettera dell’alfabeto seguita da tre cifre. I valori di tale quaterna variano, dunque, in funzione del comune italiano o stato estero di nascita e sono tabulati nel volume “Codice dei Comuni d’Italia” o nel volume “Codice degli Stati esteri”, redatti a cura della Direzione generale del catasto e dei servizi tecnici erariali.

¹² Il paragrafo si articola in due parti, in una prima parte vengono sinteticamente illustrati i metodi, i concetti e le definizioni che trovano trattazione più approfondita nell’articolo di: Consolini P., M. Di Marco, R. Ricci e S. Vitaletti. *Administrative and Survey Microdata on Self-Employment: the Italian Experience with the EU-SILC project*. Iarw 29th general conference, Joensuu, Finland, 20-26 August 2006. Nella seconda viene descritto, per sommi capi, lo schema di costruzione dei redditi autonomi da fonte amministrativa. Per maggiori dettagli si rimanda al lavoro più analitico di: Consolini P. *Integrazione di dati campionari Eu-Silc con dati di fonte amministrativa*. Roma: Istat, 2009. (Metodi e Norme, n. 38).

“...Income data for the self-employed are also generally regarded as unreliable as a guide to living standards...”¹³

“...Household surveys are notoriously bad at measuring income from capital and self-employment income ...”¹⁴.

In generale, tutte le fonti del dato utilizzabili ai fini della rilevazione del reddito autonomo mostrano varie lacune per ciò che attiene al concetto di reddito disponibile. Le indagini campionarie che ricorrono allo strumento dell'intervista diretta sono tipicamente affette da errori di “*under-reporting*”, cioè dalla tendenza dell'intervistato a dichiarare meno del vero; mentre quelle che utilizzano i dati di fonte amministrativa (*register-based survey*) possono ignorare componenti di reddito non dichiarate al fisco (evasione) oppure includere deduzioni implicite non sempre individuabili nei documenti fiscali (erosione).¹⁵ D'altra parte, se si desidera recuperare l'informazione sui redditi autonomi dai libri contabili si ottiene il reddito dell'impresa al netto delle tasse, intesa come un'unica entità giuridica, e quindi senza il dettaglio sulla tassazione alle persone fisiche relativa ai profitti degli imprenditori. Si ricorda, inoltre, che il regime fiscale italiano prevede una serie di misure, quali ad esempio la deduzione dei profitti che sono trattenuti e reinvestiti nell'attività dell'impresa, la deduzione delle perdite di esercizio degli anni precedenti, i regimi speciali di tassazione molto favorevoli per le imprese piccole o di nuova costituzione o altri ancora mirati a talune categorie di imprese o di lavoratori autonomi che aderiscono agli Studi di settore,¹⁶ che di fatto rendono distante la definizione di reddito tassabile per l'attività autonoma dal concetto di reddito disponibile.

Il “questionario dell'indagine italiana Eu-Silc” è strutturato in modo tale da stimolare il ricordo e la quantificazione del reddito autonomo da parte del rispondente, indirizzandolo nella giusta direzione. La domanda sull'ammontare di reddito autonomo viene infatti preceduta da una serie di possibili utilizzi di questa fonte di guadagno (consumi, risparmi, trasferimenti a favore di altre famiglie eccetera). Tale sequenza ha lo scopo di suggerire all'intervistato un'interpretazione più vicina possibile al concetto di “denaro prelevato dall'impresa per scopi personali” così come è stabilito da Eurostat. In effetti la traduzione letterale del quesito suggerito nel manuale Eu-Silc (*Have you drawn out money from business for your personal use?*) potrebbe evocare al rispondente un concetto sbagliato, facendogli intendere che si tratta di soldi utilizzati ai fini personali sottratti illegalmente all'attività e pertanto non documentati nei libri contabili. Di contro, le perdite derivanti dall'attività autonoma vanno intese come “*money put into*”, cioè come denaro immesso in quella attività dal lavoratore indipendente e, dunque, sottratto ad altre destinazioni (consumi e risparmi). In questo ambito è bene distinguere i flussi di denaro utilizzati per gli investimenti nell'impresa da quelli utilizzati per coprire il disavanzo prodotto in un esercizio. I primi vanno considerati come un incremento del valore dell'impresa (incremento del patrimonio aziendale) mentre i secondi sono, a tutti gli effetti, perdite. Durante lo svolgimento delle indagini pilota, precedenti al lancio della nuova indagine sui redditi Eu-Silc, si è potuto osservare che gli intervistati mostravano maggiore prontezza a rispondere a domande poste in uno schema semplice, rispetto a uno schema analitico che mutua i concetti tipici della contabilità d'impresa o della tassazione fiscale.

¹³ Canberra Group, 2001, p.54.

¹⁴ Canberra Group, 2001, p.62.

¹⁵ Si ricorda infatti che il reddito imponibile ai fini Irpef degli autonomi è calcolato al netto dei costi di produzione calcolati o in maniera analitica, o forfettaria/semplificata.

¹⁶ L'uso di soglie minori di tolleranza, implicite nel calcolo degli “Studi di settore”, di fatto lascia mediamente fuori dalla tassazione le quote di reddito che un operatore percepisce al di sopra di quanto ritenuto sufficiente dal fisco (cfr. Di Nicola e Polin, 2006).

Prospetto 3.3 - Redditi da lavoro autonomo: percettori, voci economiche e fonti del dato - Anno 2007

VARIABILE TARGET EU-SILC	Definizioni	Tipologie di percettori	Voci economiche	Fonti del dato (periodo d'imposta 2006)			
				Agenzia delle entrate			Istat Indagine Eu-Silc
				Modello Cud/770	Modello 730	Modello Upf	
PY050N/G Guadagni o perdite in denaro derivanti dal lavoro autonomo	Insieme dei guadagni ricevuti, durante il periodo di riferimento dei redditi, come risultato dell'attività corrente o pregressa di lavoro autonomo	Collaboratori coordinati e continuativi	- Remunerazione delle attività prestate in collaborazioni a progetto e coordinate e continuative	Parte B e C	Quadro C (a)	Quadro RC (a)	
		Soci di cooperative	- Remunerazione dell'attività lavorativa prestata in cooperative di produzione e lavoro, cooperative di servizi, cooperative agricole e di prima trasformazione	Parte B	Quadro C (a)	Quadro RC (a)	
		Sacerdoti della Chiesa cattolica e ministri del culto	- Remunerazione del servizio prestato nelle diocesi e congrue corri-sposte sui Fondi del culto, beneficenza e religione	Parte B	Quadro C (a)	Quadro RC (a)	
		Personale medico ospedaliero, docenti universitari e operatori sanitari	- Remunerazione dell'attività medico/sanitaria intramuraria	Parte B	Quadro C (a) (b)	Quadro RC (a) (b)	
		Agenti e rappresentanti di commercio che non possiedono partita Iva	- Provvigioni inerenti a rapporti di commissione, agenzia, mediazione, rappresentanza di commercio e procacciatori d'affari	Mod. ord. 770 (b): Sezione autonomi			✓
		Utilizzatori delle proprie opere d'ingegno	- Utilizzazione ai fini economici di opere d'ingegno e royalties	Mod. ord. 770: Sezione autonomi	Quadro D	Quadro RL	
		Associati in partecipazione (artt. 2549-2554 c.c.)	- Partecipazione agli utili di impresa come corrispettivo dell'apporto esclusivamente di lavoro	(c)	Quadro D	Quadro RL	
		Lavoratori autonomi occasionali e venditori a domicilio (art.2222 c.c.)	- Redditi diversi (art. 67, comma 1, lett. l), Tuir)	(c)	Quadro D	Quadro RL	
		Liberi professionisti e artisti	- Proventi derivanti dall'esercizio di libera professione e attività artistica			Quadro RE	
		Titolari d'impresa	- Remunerazione dell'attività imprenditoriale			Quadri RF/RG/RD	
Coadiuvanti di impresa familiare o coniugale, soci di società di persone	- Remunerazione dell'attività di collaborazione nella conduzione dell'impresa familiare o in società di persone			Quadro RH			

(a) La voce corrispondente è rilevata solo se il percettore presenta una dichiarazione del modello 730 o del modello Unico persone fisiche.

(b) La rilevazione di questa voce è ancora in fase progettuale.

(c) Se il compenso supera la soglia di 5.000 euro allora sorge l'obbligo da parte del committente di dichiarare i dati retributivi ai fini previdenziali per la Gestione separata Inps, Sezione 4, Parte C del Cud/770.

A partire dall'edizione Eu-Silc 2006 è stata poi rivisitata e modificata la struttura del questionario pertinente ai redditi autonomi, con la creazione di una nuova sezione specifica per i compensi derivanti da collaborazioni coordinate e continuative (co.co.co.). Si rammenta che questa fonte di guadagno in Eu-Silc è classificata tra le componenti di reddito autonomo. Tale innovazione si era resa necessaria in quanto si era appurato che una quota significativa di co.co.co. tendeva a dichiararsi come percettore di reddito da lavoro dipendente invece che autonomo.

La “rilevazione del reddito da lavoro autonomo a partire dalle fonti fiscali” costituisce un processo che presenta elevati livelli di complessità. Il reddito in questione è, infatti, associabile a figure professionali che trovano differente collocazione nell’ambito della disciplina giuridico-fiscale. Così ad esempio, i redditi percepiti dai soci di cooperative di produzione e lavoro, delle cooperative di servizi, delle cooperative agricole e di prima trasformazione, sono classificati come componenti autonome ma vengono assimilati fiscalmente ai redditi dal lavoro dipendente e pertanto registrati nei modelli Cud/770. Similmente i redditi derivanti dalle collaborazioni coordinate e continuative, pur essendo classificati ai fini statistici come redditi da lavoro autonomo, sono anch’essi assimilati fiscalmente ai redditi da lavoro dipendente e, dunque, inquadrati nelle dichiarazioni Cud/770. Altre voci economiche, quali i proventi derivanti dall’utilizzazione economica di opere di ingegno, sono poi rilevabili dal modello 730 e dal modello Unico persone fisiche. Infine, tutti i guadagni derivanti da attività professionale e di impresa, in qualità di titolare o partecipante (coadiuvante/socio), sono esclusivamente registrate nell’archivio delle dichiarazioni Unico persone fisiche. Nel prospetto 3.3 è riportato un quadro sinottico di queste voci.

La rilevazione e misurazione del reddito netto da lavoro autonomo tramite fonti amministrative è stata rivista e migliorata nel corso del tempo¹⁷ Per il calcolo delle componenti di reddito autonomo assimilabili fiscalmente ai redditi da lavoro dipendente (redditi dei co.co.co., dei soci di cooperative eccetera) è stata adottata una definizione di “reddito netto” in coerenza con le definizioni Eu-Silc: “reddito al netto delle ritenute alla fonte di tipo contributivo e fiscale”. Ciò alla stregua dei redditi da lavoro dipendente. Mentre per i redditi da lavoro autonomo desumibili dal modello Unico persone fisiche si è preferito rivedere il criterio di calcolo, in quanto sono stati osservati notevoli scostamenti tra i valori relativi agli acconti e alle ritenute, versate in anticipo dall’autonomo rispetto alla sua dichiarazione finale (cioè le cosiddette. ritenute alla fonte), e i valori dell’imposta definitiva, ciò soprattutto in riferimento ai redditi medio-alti da lavoro autonomo. Tale situazione ha suggerito l’abbandono del criterio che prevedeva l’impiego delle ritenute alla fonte, come componente negativa al calcolo del reddito netto, a favore dell’utilizzo dell’imposta (netta) effettivamente pagata. Il calcolo del reddito netto autonomo in assenza di altre componenti economiche è immediato. Dalla seguente figura 3.3 si evince che esso è ottenuto dalla differenza tra il reddito complessivo (rigo RN1 col. 2) e l’imposta effettiva, a sua volta ricavata dalla somma delle componenti afferenti alle ritenute totali (RN21 col. 2), agli acconti versati (RN27 col. 2) e al saldo d’imposta ((+)RN29/(-)RN30).

In presenza di altre componenti economiche il reddito netto da lavoro autonomo si ottiene attribuendo *pro quota* l’imposta effettiva, in ragione del peso che questi ultimi (Prospetto 3.3 e figura 3.4) hanno rispetto a tutte le fonti di reddito percepite. La ripartizione *pro quota* dell’imposta definitiva rispetto alle relative varie componenti economiche (lorde) si ottiene con la seguente formula: $T^A = T \cdot (Y_L^A / Y_L)$;

dove T e Y_L indicano rispettivamente il totale delle “imposte pagate” e il “reddito complessivo”, mentre T^A e Y_L^A denotano le stesse grandezze riferite alla sola componente da lavoro autonomo.

¹⁷ Ciò ha comportato il rilascio di una nuova versione del reddito da lavoro autonomo per l’indagine Eu-Silc 2004 (file standard di microdati) che teneva conto della nuova definizione.

Figura 3.3 - Modello della dichiarazione dei reddito Unico persone fisiche - Quadro RN "Calcolo dell'irpef" anno 2006

		PERIODO D'IMPOSTA 2005																											
		CODICE FISCALE <input style="width: 100%; height: 20px;" type="text"/>																											
REDDITI QUADRO RN Determinazione dell'IRPEF		EURO 																											
MODELLO GRATUITO	QUADRO RN IRPEF	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 60%;">RN1 REDDITO COMPLESSIVO (sommare tutti i redditi Irpef)</td> <td style="width: 20%; text-align: right;">Credito per fondi comuni di cui ai quadri RF, RG e RH</td> <td style="width: 20%;"></td> </tr> <tr> <td></td> <td style="text-align: right;">1</td> <td style="text-align: right;">2</td> </tr> <tr> <td></td> <td style="text-align: right;">,00</td> <td style="text-align: right;">,00</td> </tr> <tr> <td>RN2 Deduzione per abitazione principale</td> <td></td> <td style="text-align: right;">,00</td> </tr> <tr> <td>RN3 Oneri deducibili (riportare l'importo di rigo RP26)</td> <td></td> <td style="text-align: right;">,00</td> </tr> <tr> <td>RN4 Deduzione per la progressività dell'imposizione (art. 11 del Tuir)</td> <td></td> <td style="text-align: right;">,00</td> </tr> <tr> <td>RN5 Deduzione per oneri di famiglia (art. 12 del Tuir)</td> <td></td> <td style="text-align: right;">,00</td> </tr> <tr> <td>RN6 REDDITO IMPONIBILE (RN1 col. 1 + RN 1 col. 2 - RN2 - RN3 - RN4 - RN5; indicare zero se il risultato è negativo)</td> <td></td> <td style="text-align: right;">,00</td> </tr> <tr> <td>RN7 IMPOSTA LORDA</td> <td></td> <td style="text-align: right;">,00</td> </tr> </table>	RN1 REDDITO COMPLESSIVO (sommare tutti i redditi Irpef)	Credito per fondi comuni di cui ai quadri RF, RG e RH			1	2		,00	,00	RN2 Deduzione per abitazione principale		,00	RN3 Oneri deducibili (riportare l'importo di rigo RP26)		,00	RN4 Deduzione per la progressività dell'imposizione (art. 11 del Tuir)		,00	RN5 Deduzione per oneri di famiglia (art. 12 del Tuir)		,00	RN6 REDDITO IMPONIBILE (RN1 col. 1 + RN 1 col. 2 - RN2 - RN3 - RN4 - RN5; indicare zero se il risultato è negativo)		,00	RN7 IMPOSTA LORDA		,00
	RN1 REDDITO COMPLESSIVO (sommare tutti i redditi Irpef)	Credito per fondi comuni di cui ai quadri RF, RG e RH																											
		1	2																										
		,00	,00																										
	RN2 Deduzione per abitazione principale		,00																										
	RN3 Oneri deducibili (riportare l'importo di rigo RP26)		,00																										
	RN4 Deduzione per la progressività dell'imposizione (art. 11 del Tuir)		,00																										
	RN5 Deduzione per oneri di famiglia (art. 12 del Tuir)		,00																										
RN6 REDDITO IMPONIBILE (RN1 col. 1 + RN 1 col. 2 - RN2 - RN3 - RN4 - RN5; indicare zero se il risultato è negativo)		,00																											
RN7 IMPOSTA LORDA		,00																											
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 60%;">RN21 RITENUTE TOTALI (riportare la somma delle ritenute indicate nei singoli quadri) (di cui sospese ¹ ,00)</td> <td style="width: 20%;"></td> <td style="width: 20%; text-align: right;">2</td> </tr> <tr> <td>RN22 ALTRE RITENUTE SUBITE</td> <td></td> <td></td> </tr> <tr> <td>RN23 DIFFERENZA (RN17 o RN18 col. 7 - RN19 col. 2 - RN20 col.1 - RN20 col.2 - RN20 col.3 - RN21 col. 2 - RN22) se tale importo è negativo vedere istruzioni</td> <td></td> <td></td> </tr> <tr> <td>RN24 Crediti d'imposta per le imprese e i lavoratori autonomi</td> <td style="text-align: right;">,00</td> <td></td> </tr> <tr> <td>RN25 ECCEDEZZA D'IMPOSTA RISULTANTE DALLA PRECEDENTE DICHIARAZIONE</td> <td></td> <td></td> </tr> <tr> <td>RN26 ECCEDEZZA D'IMPOSTA RISULTANTE DALLA PRECEDENTE DICHIARAZIONE COMPENSATA NEL MOD. F24</td> <td></td> <td></td> </tr> <tr> <td>RN27 ACCONTI VERSATI (di cui sospesi ¹ ,00)</td> <td></td> <td style="text-align: right;">2</td> </tr> <tr> <td>RN28 Irpef da versare o da rimborsare risultante dal Mod. 730/2006</td> <td style="text-align: right;">1</td> <td style="text-align: right;">2</td> </tr> <tr> <td></td> <td style="text-align: right;">,00</td> <td style="text-align: right;">,00</td> </tr> </table>		RN21 RITENUTE TOTALI (riportare la somma delle ritenute indicate nei singoli quadri) (di cui sospese ¹ ,00)		2	RN22 ALTRE RITENUTE SUBITE			RN23 DIFFERENZA (RN17 o RN18 col. 7 - RN19 col. 2 - RN20 col.1 - RN20 col.2 - RN20 col.3 - RN21 col. 2 - RN22) se tale importo è negativo vedere istruzioni			RN24 Crediti d'imposta per le imprese e i lavoratori autonomi	,00		RN25 ECCEDEZZA D'IMPOSTA RISULTANTE DALLA PRECEDENTE DICHIARAZIONE			RN26 ECCEDEZZA D'IMPOSTA RISULTANTE DALLA PRECEDENTE DICHIARAZIONE COMPENSATA NEL MOD. F24			RN27 ACCONTI VERSATI (di cui sospesi ¹ ,00)		2	RN28 Irpef da versare o da rimborsare risultante dal Mod. 730/2006	1	2		,00	,00	
RN21 RITENUTE TOTALI (riportare la somma delle ritenute indicate nei singoli quadri) (di cui sospese ¹ ,00)		2																											
RN22 ALTRE RITENUTE SUBITE																													
RN23 DIFFERENZA (RN17 o RN18 col. 7 - RN19 col. 2 - RN20 col.1 - RN20 col.2 - RN20 col.3 - RN21 col. 2 - RN22) se tale importo è negativo vedere istruzioni																													
RN24 Crediti d'imposta per le imprese e i lavoratori autonomi	,00																												
RN25 ECCEDEZZA D'IMPOSTA RISULTANTE DALLA PRECEDENTE DICHIARAZIONE																													
RN26 ECCEDEZZA D'IMPOSTA RISULTANTE DALLA PRECEDENTE DICHIARAZIONE COMPENSATA NEL MOD. F24																													
RN27 ACCONTI VERSATI (di cui sospesi ¹ ,00)		2																											
RN28 Irpef da versare o da rimborsare risultante dal Mod. 730/2006	1	2																											
	,00	,00																											
Determinazione dell'imposta	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 60%;">RN29 IMPOSTA A DEBITO</td> <td style="width: 20%;"></td> <td style="width: 20%;"></td> </tr> <tr> <td>RN30 IMPOSTA A CREDITO</td> <td></td> <td></td> </tr> <tr> <td>RN31 RESIDUO DELLE DETRAZIONI E DEI CREDITI D'IMPOSTA RELATIVO AI RIGHI RN13, RN14 E RN15</td> <td style="text-align: right;">Residuo RN13 ¹</td> <td style="text-align: right;">Residuo RN14 ²</td> </tr> <tr> <td></td> <td style="text-align: right;">,00</td> <td style="text-align: right;">,00</td> </tr> <tr> <td></td> <td></td> <td style="text-align: right;">Residuo RN15 ³</td> </tr> </table>		RN29 IMPOSTA A DEBITO			RN30 IMPOSTA A CREDITO			RN31 RESIDUO DELLE DETRAZIONI E DEI CREDITI D'IMPOSTA RELATIVO AI RIGHI RN13, RN14 E RN15	Residuo RN13 ¹	Residuo RN14 ²		,00	,00			Residuo RN15 ³												
RN29 IMPOSTA A DEBITO																													
RN30 IMPOSTA A CREDITO																													
RN31 RESIDUO DELLE DETRAZIONI E DEI CREDITI D'IMPOSTA RELATIVO AI RIGHI RN13, RN14 E RN15	Residuo RN13 ¹	Residuo RN14 ²																											
	,00	,00																											
		Residuo RN15 ³																											

Figura 3.4 - Schema riepilogativo dei redditi e ritenute incluse nel Quadro RN “Calcolo dell’Irpef” Unico persone fisiche anno 2006

IRPEF – SCHEMA RIEPILOGATIVO DEI REDDITI E DELLE RITENUTE		REDDITI	RITENUTE
	TIPO DI REDDITO		
Dominicali	(Quadro RA, rigo RA11 col. 9)		
Agrari	(Quadro RA, rigo RA11 col. 10)		
Fabbricati	(Quadro RB, rigo RB11)		
Lavoro dipendente	(Quadro RC - Sez. I, rigo RC5 col. 2) vedere istruzioni		
	(Quadro RC - Sez. II, rigo RC9)		
	(Quadro RC - Sez. III, rigo RC10)		
	(Quadro RC - Sez. IV, rigo RC13) vedere istruzioni		
Lavoro autonomo	(Quadro RE, rigo RE21 e rigo RE23)		
	(Quadro RE, rigo RE22 col. 2 e rigo RE23)		
Impresa	(Quadro RF, rigo RF54 e rigo RF55 col. 5)		
	(Quadro RG, rigo RG32 e rigo RG33 col. 5)		
Partecipazione	(Quadro RH, rigo RH16 e rigo RH17)		
Plusvalenze di natura finanziaria	(Quadro RT, rigo RT16 + rigo RT22 e rigo RT27)		
Altri redditi	(Quadro RL - Sez. I, rigo RL3 col. 2 e col. 3)		
	(Quadro RL - Sez. II-A, rigo RL18 col. 1 e RL19 col. 1)		
	(Quadro RL - Sez. II-B, rigo RL21 col. 2 e rigo RL22 col. 2)		
	(Quadro RL - Sez. III, rigo RL29 col. 1 e rigo RL30 col. 1)		
Allevamento	(Quadro RD, rigo RD16 e rigo RD17)		
Tassazione separata (solo in caso di opzione per la tassazione ordinaria)	(Quadro RM, rigo RM15, col. 1 e col. 2)		
REDDITO COMPLESSIVO E TOTALE RITENUTE (riportare rispettivamente al rigo RN1 colonna 2 e al rigo RN21, colonna 2)			

La divergenza rispetto alla definizione di reddito netto da lavoro autonomo adottata da Eurostat (*PY050N*), con la conseguente sostituzione dell’imposta definitiva al posto delle ritenute alla fonte, ha implicato un aggiustamento della componente residuale “saldo d’imposta”; utilizzata in Eu-Silc come posta correttiva nel calcolo del reddito disponibile familiare (*HY145*). Tale operazione si è resa necessaria per evitare possibili doppi conteggi sulla quota parte di saldo già inclusa nel reddito da lavoro autonomo. Pertanto, si è stabilito di detrarre dalla posta correttiva *HY145* la quota parte imputabile all’autonomo quale saldo tra il totale delle ritenute alla fonte e l’imposta definitiva pagata da questi.

Per quanto concerne la comparabilità tra la definizione di reddito autonomo da questionario e quella di fonte fiscale occorre dire che esse, seppure facciano riferimento alla stessa nozione di reddito, si approssimano nella misura in cui: l’intervistato sappia effettivamente individuare e classificare le proprie fonti di guadagno compilando la corretta sezione di redditi del questionario Eu-Silc;¹⁸ l’intervistato non commetta errori durante il processo di ricordo (ad esempio assenza effetti telescopici); in presenza di risposte *proxy* l’intervistato conosca a fondo i redditi degli altri familiari; i redditi desumibili dalla fonte fiscale riflettano le effettive entrate da quella fonte (il che non sempre è garantito quando si applicano regimi speciali di tassazione).

¹⁸ Secondo Moore (si veda Moore J.C., L.L.Stinson e E.J. Welniak. “Income Measurement Error in Surveys: A Review”. *Journal of Official Statistics*, Vol. 16, Issue 4, 2000) la rilevazione campionaria dei redditi è generalmente un processo articolato in due stadi, in cui si richiede all’intervistato di riportare, dapprima, le proprie fonti di guadagno e, successivamente, l’ammontare di reddito percepito per ciascuna fonte. Un errore di riposta può presentarsi in ambedue gli stadi. Pertanto può accadere che una intera fonte di entrata non venga dichiarata affatto (*item non response*) ovvero venga classificata diversamente, generando un “falso positivo” cioè una componente di reddito nella realtà mai percepita.

3.4.2 - Redditi da lavoro dipendente

Per la rilevazione dei redditi da lavoro dipendente tramite questionario è stata prevista una batteria di domande aventi una struttura che parte dalla richiesta di informazioni di carattere generale e procede con la diagnosi di aspetti sempre più particolareggiati. Tale struttura aveva dato i migliori esiti nei test cognitivi, in fase di messa a punto del questionario sui redditi da lavoro dipendente, durante le tre indagini pilota di Eu-Silc. Lo schema prevede un paio di quesiti generali sulle “mensilità retribuite nell’anno” e sul valore della “retribuzione mensile netta percepita in un dato anno, senza tredicesima o arretrati”, e la successiva elencazione di voci specifiche che formano il salario accessorio. Queste ultime voci sono a loro volta suddivise in “compensi aggiuntivi”,¹⁹ aventi carattere ricorrente, e in “compensi accessori”,²⁰ pagati con cadenza ultramensile. In relazione alle componenti accessorie è richiesto all’intervistato di specificare se esse sono o meno comprese nel valore della retribuzione mensile fornita in precedenza; in caso negativo ne è richiesta la quantificazione. Una voce a sé è assegnata alle mensilità aggiuntive, quali tredicesime e quattordicesime; mentre gli arretrati da lavoro dipendente, dapprima inclusi nella sezione redditi da lavoro dipendente (sino all’indagine Eu-Silc 2005), sono stati spostati nella sezione “Altri redditi”. Ciò per consentire la loro rilevazione anche in mancanza di retribuzioni percepite nell’anno di riferimento dei redditi.

La ricostruzione del “reddito netto da lavoro dipendente di fonte fiscale” è un processo di notevole complessità in quanto richiede, da una parte, lo studio analitico della materia fiscale – ciò al fine di identificare le fonti di rilevazione, nonché definire il contenuto e il significato per le singole voci economiche – e dall’altra l’implementazione di soluzioni metodologiche ottimali per l’integrazione delle informazioni tra le varie fonti e la classificazione delle relative poste. Il primo passo nella costruzione del dato di interesse è di individuare le fonti del dato che lo rilevano. A tale scopo sono stati identificati come giacimenti informativi il modello Cud/770, il modello 730 e il modello Unico persone fisiche (Upf). Tra questi il più rilevante in termini di copertura del dato e di contenuto informativo è l’archivio dei modelli Cud/770. Se si considera, ad esempio, il modello Cud, si nota come al punto 1 della parte B (Sezione dati fiscali) esso riporti i “redditi per i quali è possibile fruire della deduzione di cui all’art. 11 e delle detrazioni di cui all’art. 14 del Tuir”. L’aggregato comprende stipendi, pensioni, mensilità aggiuntive, indennità di trasferta, borse di studio, compensi derivanti da rapporti di collaborazione coordinata e continuativa, compensi ai soci di cooperative di produzione e lavoro eccetera. Si tratta, in sostanza, di componenti di reddito eterogenee e classificabili secondo lo schema riportato nella seguente figura 3.5.

Ai fini, dunque, di una corretta attribuzione degli importi individuali di reddito da lavoro dipendente è necessario scorporare quest’ultima componente di reddito (evidenziata in grassetto) dalle restanti tipologie di reddito²¹ Come vedremo in seguito, per realizzare la separazione tra le varie componenti economiche si è proceduto a scorporare, dapprima, il reddito da lavoro dipendente dai trasferimenti pensionistici. Independentemente dal criterio di

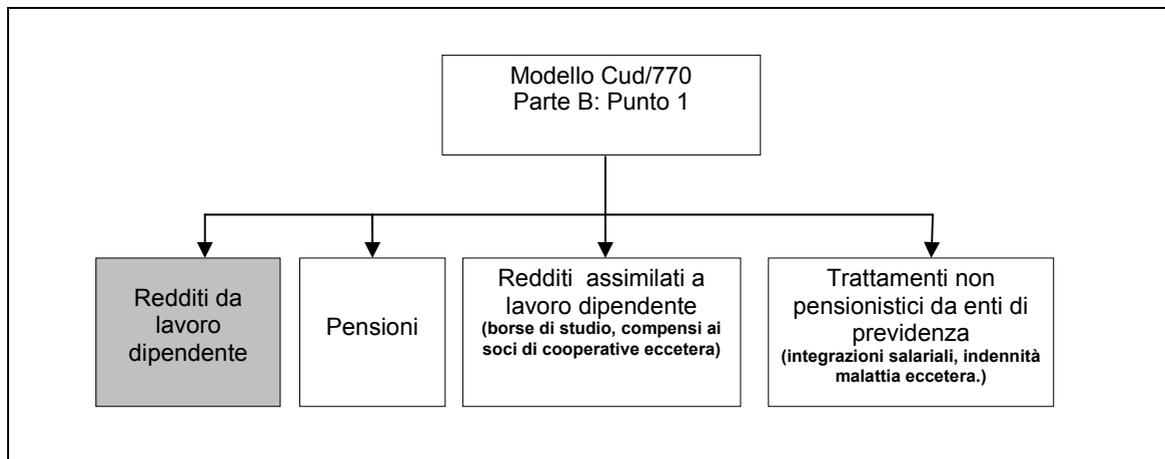
¹⁹ Tra essi figurano gli straordinari, le indennità per lavoro serale o notturno o per periodi festivi, le indennità di rischio o di responsabilità per le funzioni esercitate o per la reperibilità, le indennità di trasferimento per sede disagiata, le mance, e così via.

²⁰ Questi comprendono i primi e gli incentivi di produttività o di risultato, le indennità per missioni, premi legati alla partecipazione agli utili aziendali, quote di azioni e/o obbligazione o diritti di opzione su esse.

²¹ Il calcolo del reddito netto da lavoro dipendente risulta più complesso nel caso di presenza di redditi al punto 2 della parte B del modello Cud. Al punto 2 sono, infatti, riportate alcune componenti non tutte collocabili nella sfera dei redditi da lavoro dipendente, ma pertinenti anche alla tipologia del lavoro autonomo. Tra queste ultime vi sono i compensi per l’attività libero professionale intramuraria del personale dipendente del Servizio sanitario nazionale e i compensi corrisposti ai membri delle commissioni tributarie. Tra le componenti che fanno parte del reddito da lavoro dipendente vi sono i compensi/indennità dati a membri dei corpi legislativi e di governo, dirigenti amministrativi e giudiziari della pubblica amministrazione e di organizzazioni di interesse nazionale e sopranazionale. L’inclusione di queste ultime componenti è stata implementata nell’edizione Eu-Silc 2006 e ha previsto la revisione della variabile target per le precedenti forniture di microdati.

estrazione della componente di reddito da lavoro dipendente, il punto cruciale ai fini dell'integrazione delle informazioni è quello di definire esattamente il contenuto della variabile fiscale che deve essere armonizzata con la definizione di reddito Eu-Silc.

Figura 3.5 – Classificazione delle componenti di reddito di cui al punto 1, Parte B, modelli Cud/770



In base alle normative fiscali (d.p.r. n. 917/1986 e successive disposizioni) gli elementi costitutivi del reddito da lavoro dipendente (base imponibile), esposti al punto 1 della parte B, includono tanto gli stipendi, emolumenti, premi di produttività, mensilità aggiuntive quanto una serie di benefici economici concessi al lavoratore. Tra i benefici che determinano la base imponibile si possono annoverare le indennità per ferie non godute, le indennità di trasferta o diaria fuori dalla comune sede di lavoro per la parte eccedente i limiti di legge (ad esempio per la parte eccedente 46,48 euro delle trasferte in Italia e 77,47 euro all'estero in caso di rimborso delle sole spese di viaggio), le indennità di prima sistemazione per una quota del 50 per cento, ed una lista di *fringe benefits*. Questi ultimi comprendono a loro volta: “erogazioni liberali” (tassati per la parte eccedente 258,23 euro); “prestazioni sostitutive della mensa” quali buoni pasto (per la parte eccedente i 5,29 euro); “prestiti agevolati” (costituisce base imponibile il 50 per cento della differenza tra il tasso di interesse applicato dal datore e il tasso ufficiale di riferimento); “auto aziendali” (il 30 per cento della tariffa Aci corrispondente a una percorrenza convenzionale di 15 mila chilometri); “indennità di trasporto” alla singola persona (quelle collettive non concorrono a determinare la base imponibile). Non tutti questi benefici economici rientrano nella definizione Eurostat di “reddito in denaro da lavoro dipendente” corrispondente alla variabile target *PY010N/G*. Alcuni di essi sono ricompresi nella complementare definizione Eurostat di “retribuzione in natura *PY020N/G*”; come ad esempio l’auto aziendale, i buoni pasto, le erogazioni liberali, che tuttavia ai fini fiscali sono calcolate in modo convenzionale e in misura inferiore all’effettivo valore. Altri ancora, non fanno parte della componente del reddito da lavoro dipendente (in denaro o in natura), quali le indennità di trasferta (costi di produzione) e gli alloggi gratuiti o a prezzo inferiore di mercato offerti dal datore di lavoro (affitti imputati *HY030G*). La figura 3.6 sintetizza, in maniera esauriente, l’insieme delle differenze o corrispondenze tra le poste rilevate al punto 1 (parte B modello Cud/770) e le voci che compongono la variabile target *PY010N*.

Figura 3.6 - Elenco delle corrispondenze/differenze tra voci economiche del reddito dipendente da fonte Eu-Silc (PY010N) e da fonte fiscale (punto 1 parte B modello Cud/770)

<ul style="list-style-type: none"> - Stipendi e salari pagati in denaro per il periodo effettivamente lavorato; - Remunerazione del periodo non lavorato (ferie retribuite); - Incrementi stipendiali nel tempo; - Compensi ai direttori di azienda; - Pagamenti a cottimo; - Buoni pasto per la parte eccedente € 5,29; - Provvigioni, gratifiche per la parte eccedente € 258,23; - Mance dei croupiers (in misura del 25 per cento); - Pagamenti supplementari (compensi per lavoro straordinario, indennità di cassa, mensilità aggiuntive eccetera); - Partecipazioni, titoli, azioni corrisposte al dipendente per la parte eccedente € 2065,83; - Premi di produttività; - Indennità di trasferimento in sedi disagiate (in misura del 50 per cento); - Prestazioni di servizi/indennità di trasporto individuali (casa-sede). 	<ul style="list-style-type: none"> - Mance (escluso croupier); - Gratifiche per i primi € 258,23; - Buoni pasto giornalieri per i primi € 5,29; - Partecipazione utili aziendali per i primi € 2065,83; - Indennità di trasferta per il 50 per cento; - Indennità di trasporto collettivo; - Contributi asilo nido. - Arretrati - Compensi membri cariche elettive (punto 2) 	PY010N
<ul style="list-style-type: none"> - Retribuzioni in natura (PY020N); - Alloggi gratuiti o a prezzi ridotti a dipendenti; - Prestiti agevolati ai dipendenti (50 per cento della differenza tra tasso ufficiale e agevolato); - Indennità di trasferta per la parte eccedente i limiti di legge. 	<ul style="list-style-type: none"> <input type="checkbox"/> Poste comuni a Cud e PY010N <input checked="" type="checkbox"/> Poste del Cud fuori da PY010N <input checked="" type="checkbox"/> Poste di PY010N fuori dal Cud 	

Punto 1 – Parte B – Mod. Cud/770

Dalla figura si evince che la definizione di “reddito dipendente cui fa riferimento il fisco” non comprende esattamente tutte le voci economiche previste nella variabile target Eu-Silc PY010N (errori di sotto-copertura), anche se occorre puntualizzare che le poste trascurate sono marginali rispetto alle restanti voci. Per altro verso, il dato fiscale include al suo interno componenti che non fanno parte delle retribuzioni in denaro (errori di sovra-copertura). A sua volta, il questionario Eu-Silc tralascia, sino all’edizione 2006, alcune poste (facoltative) di reddito da lavoro dipendente (*fringe benefits*), solo in seguito divenute obbligatorie ai fini della rilevazione (Prospetto 3.4). Per quanto concerne il tema della comparabilità, la definizione di reddito dipendente implicita nel questionario Eu-Silc non si attaglia perfettamente a quella di fonte fiscale; in quanto una parte del salario accessorio del dipendente è parzialmente dedotto ai fini del calcolo della base imponibile fiscale, mentre altri elementi costitutivi della stessa base imponibile (seppure marginali) non fanno parte del reddito dipendente Eu-Silc.²² A queste differenze si sovrappongono possibili scarti dovuti ad errori di misurazione tipici soprattutto delle rilevazioni tramite questionario.²³

Allo scopo di confrontare il contenuto delle due fonti, il reddito individuale *mensile* da lavoro dipendente di fonte Eu-Silc deve essere, dapprima, scisso nelle componenti esenti o

²² Un discorso a parte merita la componente “arretrati da lavoro dipendente”, rilevata sia da una sezione apposita del Cud/770 (Compensi relativi agli anni precedenti – punti 57-58 Parte B - modello anno 2006) che da uno specifico quesito nel questionario Eu-Silc. Per essa, almeno in teoria, non ci sono problemi di comparabilità tra le due fonti; anche se in pratica si assiste a notevoli deficit informativi da parte della fonte campionaria, soprattutto in relazione a importi di minore entità (effetto memoria).

²³ Per ulteriori approfondimenti si rimanda a Moore *op. cit.*; Sirken M.G. et al. *Cognition and Survey Research*. John Wiley & Sons, 1999.

meno da tassazione e successivamente, per la sola componente tassabile, moltiplicato per il numero di mensilità retribuite nell'anno. La "variabile derivata" così ottenuta rappresenta l'importo *annuo* della componente tassabile da lavoro dipendente che è l'aggregato direttamente comparabile con il dato di fonte fiscale (al netto delle ritenute contributive e fiscali).

Prospetto 3.4 - Le meta-informazioni sul reddito da lavoro dipendente: definizioni, percettori, voci economiche e fonti del dato

Variabile target Eu-Silc	Definizione	Percettori	Voci economiche	Fonte del dato				
				Agenzia delle entrate			Inps	Istat
				Modello Cud/770	Modello 730	Modello Upf	Casellario pensioni	Rilevazione Eu-Silc
PY010N Reddito netto in denaro da lavoro dipendente	Insieme delle componenti monetarie pagate dal datore di lavoro ai dipendenti a remunerazione dell'attività prestata al netto del prelievo contributivo e fiscale	Lavoratori dipendenti del settore privato e pubblico	- Stipendi e salari pagati in denaro per il periodo effettivamente lavorato	✓	✓	✓		✓
			- Remunerazione del periodo non lavorato (ferie retribuite)	✓	✓	✓		✓
			- Incrementi stipendiali nel tempo	✓	✓	✓		✓
			- Compensi ai direttori di azienda	✓	✓	✓		✓
			- Pagamenti a cottimo	✓	✓	✓		✓
			- Rimborsi spese per l'accudimento dei figli (asili nido e colonie climatiche)					✓ ⁰⁷
			- Provvigioni, gratifiche, mance	P	P	P		✓
			- Pagamenti supplementari (compensi per lavoro straordinario, indennità di cassa, mensilità aggiuntive eccetera)	✓	✓	✓		✓
			- Partecipazioni, titoli, azioni corrisposte al dipendente	P	P	P		✓
			- Premi di produttività	✓	✓	✓		✓
			- Indennità di trasferimento in sedi disagiate	P	P	P		✓
			- Prestazioni di servizi/indennità di trasporto, rimborsi viaggio (casa-sede)	P	P	P		✓ ⁰⁷
PY020N Reddito netto in natura da lavoro dipendente	Insieme delle componenti non-monetarie pagate dal datore di lavoro ai dipendenti a remunerazione dell'attività prestata al netto del prelievo contributivo e fiscale	Lavoratori dipendenti del settore privato e pubblico	- Concessione al lavoratore di veicoli aziendali in uso privato e sgravi dei relativi costi (assicurazione, bolli, rifornimenti eccetera)	P	P	P		✓
			- Mensa aziendale o prestazioni sostitutive (buoni pasto o ticket restaurant)	P	P	P		✓ ⁰⁷
			- Altri beni o servizi forniti a titolo gratuito o a prezzo ridotto	P	P	P		✓ ⁰⁷

Legenda: ✓ = completa rilevazione
 ✓⁰⁷ = completa rilevazione dal 2007
 P = parziale rilevazione

3.4.3 - Redditi da pensione

La sezione del questionario Eu-Silc dedicata alle pensioni è stata disegnata in modo tale da evocare alla mente dell'intervistato i principali eventi o condizioni che danno luogo all'erogazione di uno o più trattamenti pensionistici e, al tempo stesso, fornire a questi uno schema di classificazione come guida per la corretta attribuzione del trasferimento pensionistico alla relativa variabile target.²⁴ Le domande di questa sezione sono strutturate in tre punti, al primo si chiede all'intervistato di ricordare se egli abbia ricevuto nell'anno di riferimento una pensione del tipo specificato, al secondo si richiede l'informazione sul numero di mensilità percepite, e al terzo infine si invita a fornire un valore dell'importo medio mensile (al netto del prelievo fiscale) per quella tipologia di pensione. I trattamenti pensionistici erogati dal sistema previdenziale e assistenziale italiano, tuttavia, non esauriscono l'insieme dei trasferimenti sociali legati alle variabili target Eu-Silc: "Vecchiaia (PY100N/G)", "Superstiti (PY110N/G)" e "Invalidità (PY130N/G)", ma rappresentano tuttavia una quota preponderante. Infatti, alla formazione dei redditi compresi nelle variabili target Eu-Silc concorrono tanto le prestazioni di natura pensionistica quanto quelle non pensionistiche.²⁵ Si precisa, inoltre, che il risparmio previdenziale volontario individuale, considerato il "terzo pilastro"²⁶ dei sistemi previdenziali, attualmente esula dalla definizione di reddito Eu-Silc, come pure dal sistema della protezione sociale.

La rilevazione dei "redditi netti da pensione da fonte amministrativa" è un processo che parte da una prima fase di "armonizzazione" del contenuto delle variabili registrate rispettivamente nel Casellario centrale dei pensionati dell'Inps²⁷ e nelle fonti fiscali. Ciò al fine di ottenere una base omogenea per il confronto dei dati tra le due fonti. Dal primo archivio si costruisce la variabile composta "importo lordo annuale per tipologia di pensione", intesa come prodotto tra l'importo mensile rilevato al 31 dicembre e il numero di mensilità corrispondenti,²⁸ da confrontare con il dato fiscale. In una fase successiva, il processo applica una classificazione delle pensioni da Casellario in base al regime fiscale, cioè a seconda che siano o meno soggette a imposizione. In ultimo, si stabilisce una connessione tra i due archivi, attraverso la chiave di *matching* data dal codice fiscale, e si confrontano gli importi annui lordi delle pensioni soggette

²⁴ Le diverse tipologie di trattamenti pensionistici vigenti nel nostro Paese sono state classificate secondo le seguenti variabili target Eu-Silc: "Vecchiaia (PY100N/PY100G)", "Superstiti (PY110N/PY110G)" e "Invalidità (PY130N/PY130G)". Nella prima categoria rientrano tutte le pensioni che coprono il rischio connesso al sopraggiungere dell'età anziana; di esse fanno parte sia le pensioni di vecchiaia e anzianità che le pensioni sociali. Nella seconda categoria rientrano tutti i trattamenti temporanei corrisposti a persone che hanno perduto il coniuge o altro parente che generalmente costituiva la fonte primaria di sostentamento. Infine nella terza ricadono le prestazioni sociali che garantiscono un reddito a persone la cui capacità di lavorare è ridotta a causa di invalidità fisica o mentale (comprende le pensioni di inabilità, infortuni sul lavoro e malattie professionali).

²⁵ Per una disamina di questo argomento si rinvia a: Consolini P. *Le prestazioni sociali monetarie non pensionistiche: aspetti istituzionali e classificazioni statistiche*. Roma: Istat, 2000. (Documenti Istat, n. 2); Consolini P. e R. De Carli. *Le prestazioni sociali monetarie non pensionistiche: unità d'analisi, fonti e rappresentazione statistica dei dati*. Roma: Istat, 2002. (Documenti Istat, n. 1); Guerrucci D. *Confronto tra i sistemi di classificazione dell'assistenza sociale*, contributo al Gruppo di lavoro Istat-Usci. Roma 1996.

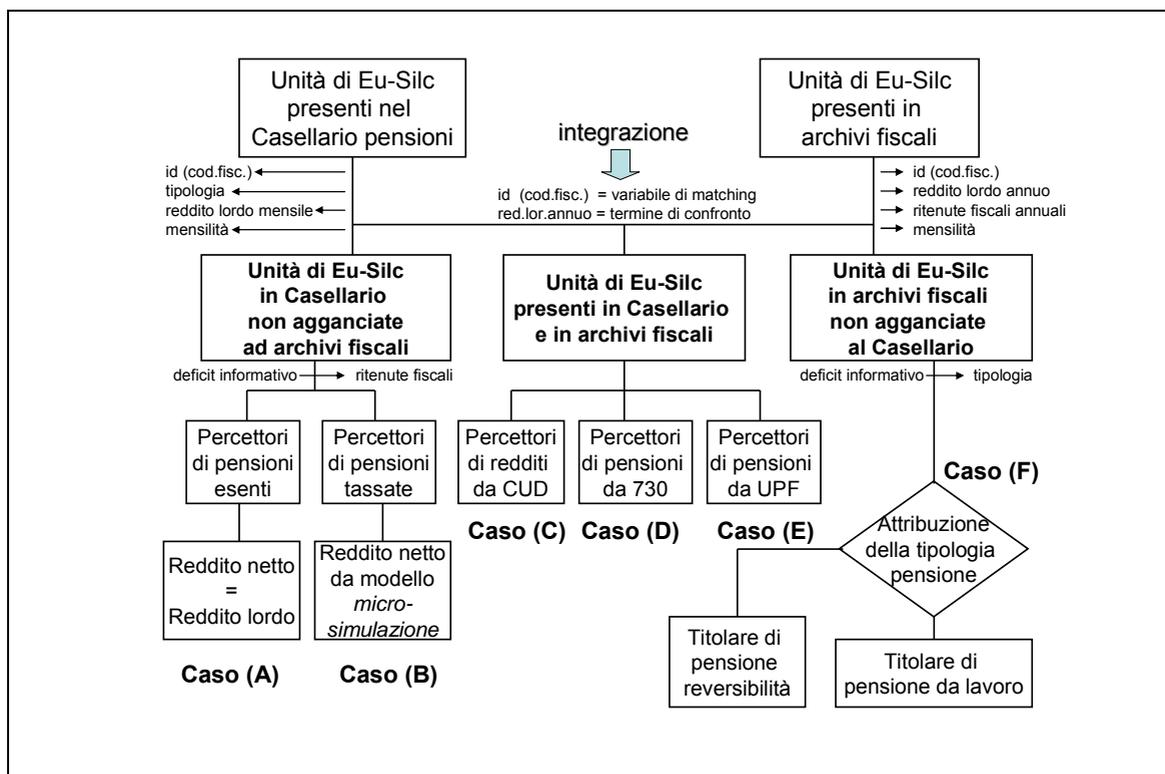
²⁶ Rientrano in questo pilastro i contratti sottoscritti da individui o imprese con fornitori di servizi assicurativi, nonché le forme di investimento individuale e collettivo in titoli trasferibili gestiti da fondi comuni di investimento. Per approfondimenti sull'informazione e classificazione statistica degli strumenti del terzo pilastro si rimanda a: Mancini A. *Previdenza integrativa e fondi pensione: linee per lo sviluppo dell'informazione statistica*. Roma: Istat, 2001. (Documenti Istat, n. 7).

²⁷ Il Casellario centrale dei pensionati, istituito presso l'Inps con il d.p.r. del 31 dicembre 2001 e successive modificazioni dal d.l. del 23 febbraio 1995, raccoglie, conserva e gestisce i dati relativi all'universo dei trattamenti pensionistici (e relativi titolari) gestiti dal nostro sistema previdenziale. Per maggiori dettagli sul contenuto informativo del Casellario pensionistico si confronti: Nazzaro O. *Statistiche sulle pensioni e sui percettori di prestazioni pensionistiche: progetto di realizzazione delle pubblicazioni per la collana Informazioni*. Roma: Istat, 1999. (Documenti Istat, n. 16).

²⁸ In particolare, per i trattamenti del Casellario con decorrenza antecedente l'anno di riferimento dei redditi si è proceduto, per ciascuna tipologia di pensione, a moltiplicare l'importo mensile per il numero massimo di mensilità erogabili in un anno. Mentre, per le pensioni decorrenti nello stesso anno di riferimento dei redditi si è proceduto a calcolare l'importo annuo in due fasi: dapprima moltiplicando l'importo mensile per il numero di mensilità intercorrenti tra il mese di decorrenza e la fine dell'anno (31 dicembre) e, successivamente, applicando gli eventuali ratei di tredicesima.

a tassazione da Casellario con gli importi registrati presso gli archivi fiscali per le stesse unità statistiche (Figura 3.7). Per le pensioni non soggette al prelievo fiscale si utilizza semplicemente il dato del Casellario pensionistico, in quanto per definizione la fonte fiscale non rileva componenti esenti; nella fattispecie vi è la perfetta coincidenza tra reddito netto e lordo da pensione (caso A nella figura 3.7). Mentre per le pensioni soggette a imposta, rilevate dalla sola fonte previdenziale, si ricorre a un modello di microsimulazione per determinare una stima dell'imposta e del reddito netto (caso B). Nel caso opposto, cioè quando è la fonte fiscale a rivelare esclusivamente il dato, si presenta il problema di "attribuire la relativa tipologia" di pensione. In questo caso la scelta verte sulle principali tipologie soggette a tassazione: pensioni da lavoro e di reversibilità (caso F). In presenza di un abbinamento tra le unità delle due fonti (casi C, D e E), si analizza la coerenza dei dati e si formulano opportune ipotesi per riconciliare i valori eventualmente difformi (ad esempio correggendo il numero di mensilità da Casellario con quelle da fisco). Ciò che si ottiene dall'integrazione delle fonti amministrative è la variabile derivata²⁹ "importo annuo netto/lordo da pensione per tipologia" che fornisce l'argomento da attribuire alla corrispondente variabile target Eu-Silc: *PY100N/G*, *PY110N/G*, *PY130N/G*.

Figura 3.7 - Struttura del processo di rilevazione/integrazione dei redditi da pensione tra fonti amministrative e fiscali (Eu-Silc 2006)



Sebbene le fonti amministrative e campionaria facciano riferimento alla stessa definizione di reddito netto da pensione secondo una data tipologia, e quindi risultino armonizzate nei

²⁹ Questa particolare "variabile derivata" è definita da Wallgren and Wallgren (cfr: Wallgren A., B. Wallgren. *Register-based Statistics. Administrative Data for Statistical Purposes*. John Wiley & Sons, 2007) come "variable derived adjoining", cioè una variabile ottenuta "accostando" diverse variabili rilevate presso molteplici fonti del dato, attraverso abbinamenti di tipo "uno-a-uno" o "uno-a-molti" tra le rispettive unità statistiche.

contenuti, nella realtà esse divergono per una pluralità di fattori, già menzionati nei due precedenti paragrafi. L'evidenza empirica mostra che l'intervistato incontra difficoltà a classificare correttamente la tipologia di pensione percepita. Questo problema si determina ogniqualvolta l'intervistato non sia a conoscenza dei termini propri della disciplina previdenziale. Tuttavia pur essendo edotto, risulta del pari difficile distinguere una tipologia e da un'altra quando questi è precettore di più trattamenti pensionistici di natura differente. Infatti, molto spesso il pensionato non è al corrente dell'esatta ripartizione dell'assegno pensionistico, corrisposto mensilmente in un unico versamento a cura dello stesso ente previdenziale, tra le varie voci funzionali (trattamenti).

Lo strumento delle "differenze relative" in termini di valori di reddito rilevati sulle stesse unità statistiche da diverse fonti del dato (previdenziali e fiscali) rappresenta il perno della struttura decisionale utilizzata in fase di definizione dei livelli delle pensioni e, più in generale, in fase di attribuzione delle componenti di reddito in compresenza di informazioni da Casellario e fonti fiscali (confronta casi C, D ed E di figura 3.7).³⁰

A rigore, per effettuare correttamente un confronto al livello di singola variabile target tra gli importi lordi annui delle pensioni del Casellario e quelli degli archivi fiscali occorre fissare a priori quali siano le tipologie pensionistiche da Casellario soggette o meno a tassazione (armonizzazione delle definizioni). Ciò in quanto le componenti pensionistiche esenti da tassazione, seppure rilevate nel Casellario, per definizione esulano dal contesto fiscale. Non costituiscono un problema per l'analisi comparativa le pensioni da lavoro (vecchiaia e anzianità) e quelle di reversibilità (con l'eccezione delle rendite Inail erogate ai superstiti), essendo teoricamente³¹ sempre gravate da imposta. Del pari semplice è il caso delle pensioni sociali e delle indennità di accompagnamento, dove però non si applica alcuna tassazione (redditi esenti). I trattamenti pensionistici inclusi nella variabile target "trasferimenti monetari per la funzione invalidità (PY130N/G)" sono invece parzialmente³² tassati e rappresentano una casistica di difficile trattazione.

La procedura di integrazione e riconciliazione dei valori pensionistici delle diverse fonti (amministrative) è articolata in più fasi sequenziali e prevede l'utilizzo a cascata di una batteria di differenze relative (prese in valore assoluto) tra i dati rilevati presso i vari archivi.

A titolo illustrativo, la "prima" tipologia di differenze relative, indicata con *DREL1_PEN*, fornisce informazioni sullo scostamento relativo degli importi annui per i trattamenti pensionistici teoricamente soggetti a tassazione (pensioni da lavoro e reversibilità) rilevati rispettivamente da fonte Casellario e fonte Cud:

$$DREL1_PEN = |CPLAV_E_13 + CPREV_E_13 - DB_PEN_RLOR| / DB_PEN_RLOR$$

dove *CPLAV_E_13* indica l'importo annuale lordo della prestazione individuale "pensione da lavoro" ottenuta come prodotto dell'importo mensile rilevato al 31 dicembre dell'anno di riferimento per il numero di mensilità corrispondenti,³³ *CPREV_T_13* denota a sua volta l'importo annuale lordo della prestazione individuale "pensione di reversibilità", infine *DB_PEN_RLOR* rappresenta l'importo delle pensioni tassabili di fonte Cud (di cui al punto 1 parte B). La struttura e l'elenco completo delle 11 batterie di differenze relative utilizzate in

³⁰ Con riguardo all'edizione italiana Eu-Silc del 2008, la percentuale di casi in cui il reddito da pensione è contestualmente rilevato da fonte Casellario e da fonte Cud è pari al 90,2 per cento del complesso dei titolari di pensione rilevati in una delle predette fonti (Caso C).

³¹ In realtà queste due tipologie di pensioni sono "effettivamente" gravate da imposta quando il loro importo lordo supera la soglia della cosiddetta "no tax area" che il legislatore ha fissato nella misura di € 7.000 per i redditi da pensione riferiti all'anno 2005.

³² In particolare, l'esenzione si applica alle pensioni erogate agli invalidi civili, ai non vedenti e ai non udenti, le rendite (dirette) per infortunio sul lavoro e per le malattie professionali, le pensioni (dirette) di guerra, per assegno di benemeranza ai perseguitati politici. Sono, al contrario, soggette a tassazione l'assegno di invalidità o le pensioni di inabilità pagate dall'Inps in favore dei lavoratori dipendenti e autonomi.

³³ Le mensilità pensionistiche si deducono a loro volta dalla data di decorrenza della pensione (mese e anno).

questo ambito è illustrato nella figura 3.8. Il “secondo” gruppo di differenze relative, denotato con $DREL2_PEN$, fornisce le stesse informazioni del caso precedente, quando però nell’argomento da confrontare³⁴ si includono anche le pensioni di invalidità di fonte Casellario: $CPINV_E_13$ (parzialmente soggette a imposizione).

La “terza” categoria di differenze relative ($DREL3_PEN$) confronta a sua volta l’importo annuo da Casellario delle pensioni di invalidità ($CPINV_E_MR=CPINV_E*MESI_RAT$), delle pensioni da lavoro ($CPLAV_E_MR=CPLAV_E*MESI_RAT$) e di reversibilità ($CPREV_T_MR=CPREV_T*MESI_RAT$), calcolate secondo un diverso criterio di “annualizzazione” (importi mensili del Casellario per il numero di mensilità dedotte dalle sole informazioni presenti nelle fonti fiscali), con l’importo lordo delle pensioni di fonte Cud (DB_PEN_RLOR). La quarta categoria di differenze relative “ $DREL4A_PEN$ ” paragona, invece, l’importo annuo da Casellario delle pensioni di invalidità “ $CPINV_E_13$ ” delle pensioni da lavoro ($CPLAV_E_13$) e di reversibilità ($CPREV_T_13$) con la somma degli importi lordi annui dei redditi da pensione (DB_PEN_RLOR) e da lavoro dipendente (DB_LAV_RLOR) da fonte Cud. In questa circostanza si ipotizza la presenza di errori di attribuzione/scomposizione delle componenti reddituali “reddito da lavoro dipendente e da pensione” di fonte Cud; per cui risultano impropriamente escluse dai redditi da pensione componenti erroneamente attribuite ai redditi da lavoro dipendente.

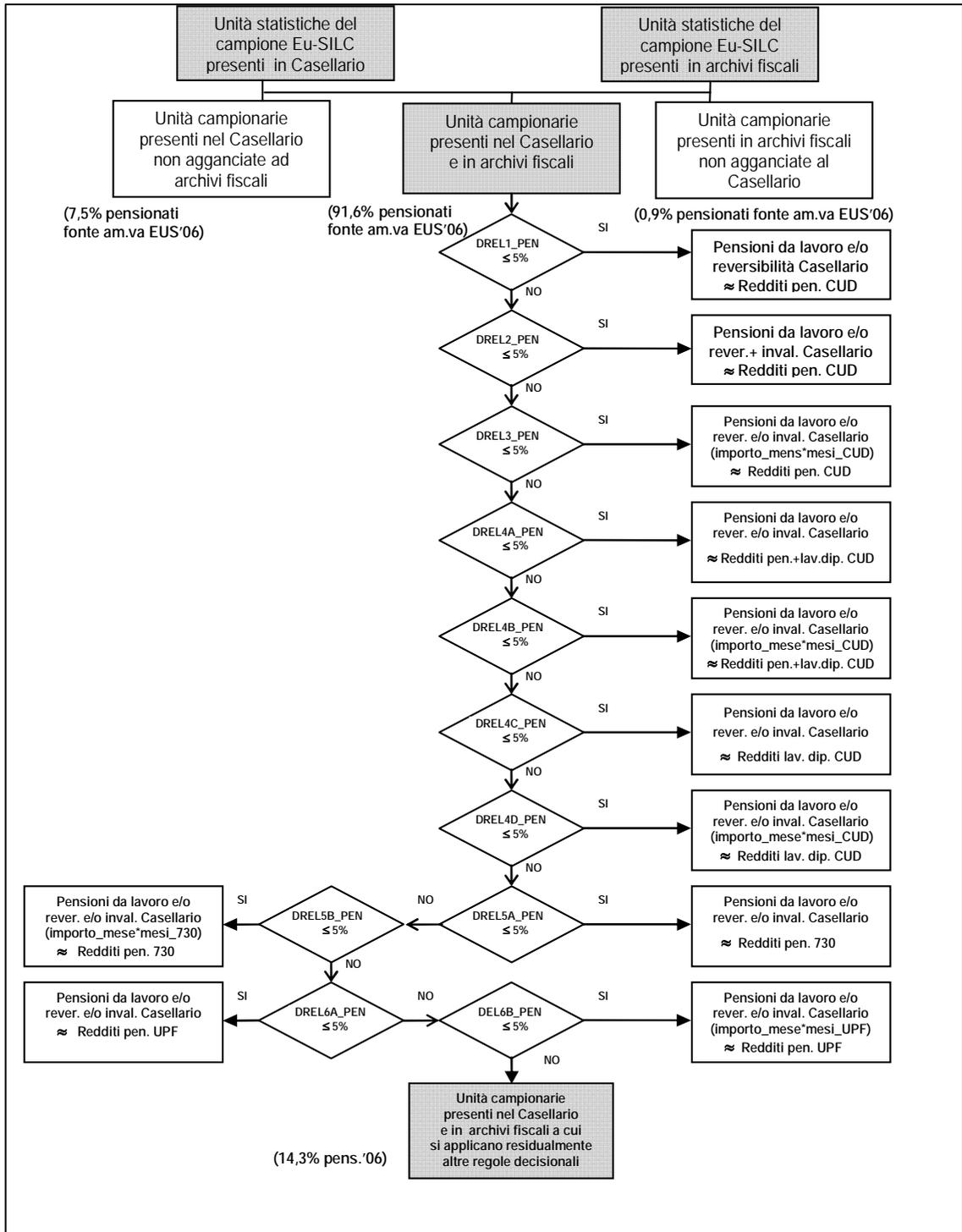
I ragionamenti fatti in considerazione delle prime quattro tipologie di differenze relative si applicano, *mutatis mutandis*, alle successive sette. La struttura e l’elenco completo delle 11 batterie di differenze relative utilizzate in questo ambito è illustrato nella figura 3.8.

Una volta definito l’argomento da confrontare (tipologia di “differenza relativa”) il passo successivo consiste nel fissare una soglia di tolleranza sugli scostamenti relativi tra i dati di fonte amministrativa (Casellario) e le restanti fonti fiscali che definisca l’accettabilità dei valori di reddito assunti in una delle due fonti. Ai fini l’integrazione dei dati pensionistici di diversa fonte si è stabilito di assegnare una soglia pari al 5 per cento; di conseguenza si attribuisce un valore alla variabile target quando la differenza relativa in valore assoluto, tra gli importi pensionistici osservati nelle due fonti, è contenuta nell’intervallo 0-5 per cento. In particolare, si assume come valore finale il dato sul reddito lordo/netto pensionistico rilevato presso la fonte del Casellario. Con riguardo al complesso delle unità statistiche che nell’anno 2005 presentano dati pensionistici contestualmente nel Casellario e nella fonte Cud/770, si osserva che ben l’84,1 per cento di esse possiede un valore di soglia inferiore al 5 per cento in almeno una delle prime nove batterie di differenze relative considerate.³⁵ Mentre se si considera la batteria completa delle 11 differenze relative si nota che il vincolo del 5 per cento viene soddisfatto dal 85,7 per cento degli stessi soggetti. Dalla figura 3.8 emerge che nel restante 14,3 per cento si devono applicare altre regole per la riconciliazione dei valori delle due fonti.

³⁴ Relativamente alle unità statistiche dell’intersezione tra Casellario e Cud, ovvero per i casi “C”, “D” ed “E” della figura 3.7.

³⁵ Se lo stesso valore assoluto lo si rapporta al complesso delle unità statistiche Eu-Silc aventi almeno un reddito pensionistico in una delle fonti (campionaria e/o amministrativa) citate si perviene a una percentuale del 76 per cento circa. In altri termini, più di tre quarti delle unità campionarie di Eu-Silc 2006, a cui è associata una pensione da fonte amministrativa e/o fiscale, possiede valori di reddito lordo annuo pensionistico convergenti tra il Casellario e il Cud.

Figura 3.8 - Albero decisionale in uso per l'integrazione dei dati pensionistici da fonte Casellario e fonti fiscali: differenze relative tra importi annui delle due tipologie di fonti Eu-SILC 2006



Le ipotesi di lavoro utilizzate per risolvere le divergenze non spiegate in precedenza considerano segnatamente la natura delle pensioni e il sistema di tassazione. Tra queste assume un certo rilievo l'ipotesi sull' "esenzione dal prelievo fiscale" delle pensioni di invalidità e quella sull'applicazione della *no tax area*, anche in relazione al cumulo di più pensioni. In particolare, si osserva come gran parte dei casi residuali³⁶ trovano spiegazione nella presenza di importi pensionistici (per trattamenti di invalidità, reversibilità e vecchiaia/anzianità, distinti o cumulati) inferiori alla soglia della *no tax area* (7 mila euro). In questi casi è lecito assumere che vi siano scarti anche consistenti tra gli importi lordi da archivi fiscali e da Casellario e che i dati reddituali di quest'ultima fonte (sistematicamente più alti dei primi) si accostino di più al concetto di reddito lordo. Ricade in quest'ultima casistica il 5,7 per cento delle unità campionarie Eu-Silc 2006 a cui è associato un trasferimento pensionistico di fonte amministrativa e/o fiscale. Per altro verso, le differenze tra gli importi lordi annui delle due fonti del dato sono riconducibili al cumulo di pensioni tutte di invalidità, in parte esenti (3,1 per cento). Anche in questo caso è lecito assumere come prossimo al dato vero (lordo) il valore degli importi lordi di fonte Casellario. Ugualmente rilevante è l'assunzione che la *no tax area* trovi applicazione anche nei confronti di quei redditi pensionistici (pensioni da lavoro e/o reversibilità) di poco superiori alla soglia dei 7 mila euro e comunque inferiori agli 8 mila euro (3,3 per cento). Infine, per riconciliare parte delle osservazioni contenenti dati pensionistici delle due fonti, si è deciso di rilassare la condizione relativa alla soglia del 5 per cento sulla batteria delle 11 differenze relative, per includere anche i casi compresi nell'intervallo 5-10 per cento (0,9 per cento). Il restanti casi in cui le fonti di discostano sottintendono errori legati all'attribuzione dei redditi nelle fonti fiscali e alla presenza di segnali di errore di duplicazione nel Cud non individuabili tramite altre fonti fiscali (1,3 per cento).

3.5 - Analisi dell'incoerenza e riconciliazione delle componenti di reddito tra le fonti del dato

Una volta completata la ricostruzione delle diverse componenti di reddito (variabili target dell'indagine di Eu-Silc) a partire dai giacimenti informativi di natura fiscale, o più in generale di natura amministrativa, il passo seguente consiste nell'individuare, dapprima, un "sistema di regole" che permettano la soluzione ai problemi di incoerenza tra le informazioni rilevate su una stessa unità da fonte campionaria e amministrativa (integrata), e successivamente nel "riconciliare" il contenuto difforme delle variabili tra le diverse fonti amministrative (ottenendo così una sintesi dei dati degli archivi amministrativi). Le strategie adottate variano a seconda della variabile target considerata, ovvero della corrispondente componente di reddito. Nel presente lavoro verranno analizzate tre distinte strategie, impiegate rispettivamente nell'attribuzione finale dei redditi da pensione, dei redditi da lavoro dipendente e da lavoro autonomo.

3.5.1 - Riconciliazione del reddito da pensione tra Casellario, fonti fiscali e fonte campionaria

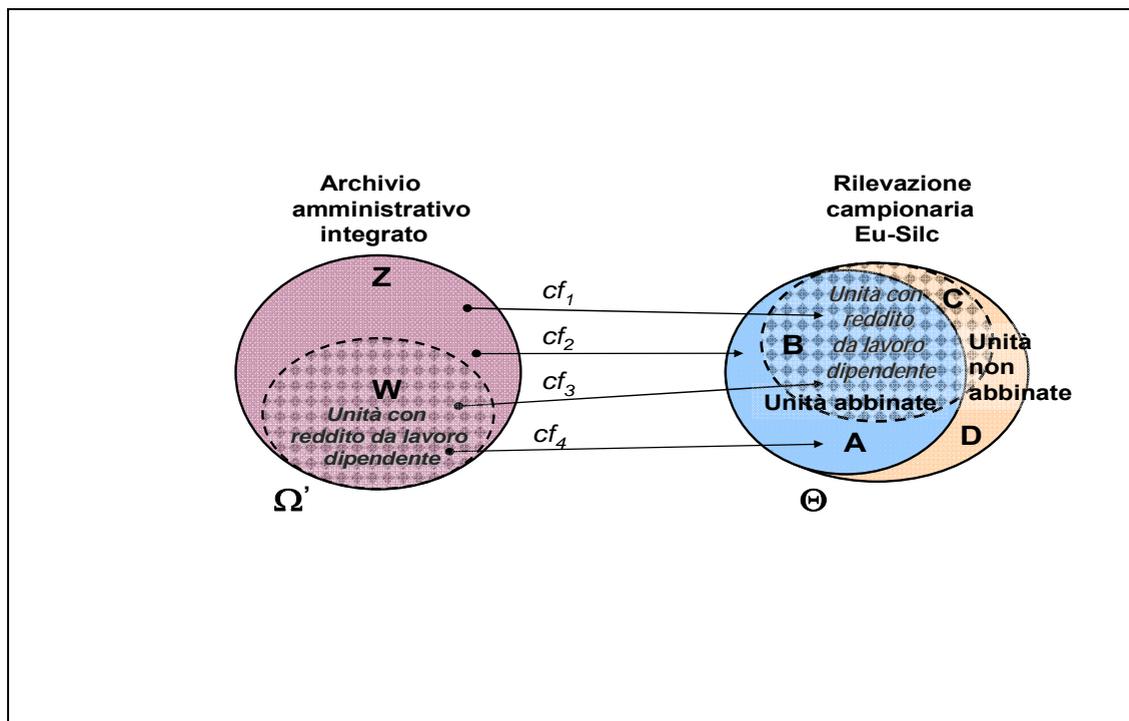
I problemi di incoerenza tra dati pensionistici di fonte campionaria e amministrativa sono risolti a monte, cioè privilegiando il contenuto del secondo giacimento, in quanto più accurato e di maggiore contenuto informativo (copertura). Viceversa, nei casi in cui non è possibile stabilire alcun abbinamento tra unità campionarie e fonti amministrative (*units not matched*), si utilizza il valore delle pensioni (importi netti) desumibile esclusivamente dalle interviste (4,7 per cento dei casi in Eu-Silc 2006).

³⁶ Relativi, cioè, al caso in cui tutte le 11 tipologie di differenze relative non soddisfano la soglia del 5 per cento.

3.5.2 - Riconciliazione del reddito da lavoro tra fonti fiscali e campionaria

L'identificazione dei percettori di reddito da lavoro dipendente, tramite l'archivio amministrativo integrato, si realizza quasi interamente sulla base dei dati riportati nella fonte fiscale Cud/770 (99,2 per cento in Eu-Silc 2006) e solo marginalmente tramite le altre fonti fiscali 730 e UPF (0,85 per cento). Si deve precisare che non sempre la fonte Cud/770 contiene informazioni corrette, complete e coerenti ai fini dell'attribuzione dei livelli di reddito da lavoro percepiti. Infatti, se si considerasse esclusivamente il contenuto informativo del Cud/770 si rischierebbe di attribuire impropriamente alla componente "reddito da lavoro dipendente" elementi reddituali facenti in realtà parte dei trasferimenti pensionistici; per i quali si è trovata traccia nel Casellario pensionistico (1,1 per cento dei pensionati rilevati nella precedente fase). Per altro verso, si osserva che nella stessa edizione 2006 soltanto lo 0,7 per cento dei percettori finali di reddito da lavoro dipendente era in realtà erroneamente classificato in prima istanza (in base alle sole informazioni del Cud/770) come titolare di pensione. L'identificazione dei percettori di reddito da lavoro (e la relativa attribuzione dei valori delle retribuzioni) tramite fonte amministrativa (integrata) è dunque preceduta logicamente dalla fase di assegnazione del reddito da pensione. In altri termini, questa ultima fase vincola i risultati della prima. Una volta identificati i redditi da lavoro dipendente e i rispettivi percettori il passo seguente è di effettuare un'analisi della coerenza dei valori riportati tra le varie fonti amministrative. I risultati sperimentali ottenuti in questa fase mostrano che la fonte Cud/770 dispone di informazioni molto attendibili sui redditi dei dipendenti e sulle mensilità retribuite.³⁷

Figura 3.9 - Tipologie di incoerenza sui dati del reddito da lavoro dipendente rilevati da più fonti



³⁷ Per approfondimenti si rinvia a: Consolini P. *Integrazione di dati campionari Eu-Silc con dati di fonte amministrativa*. Roma: Istat, 2009. (Metodi e Norme, n. 38).

Prima di esaminare i casi di incoerenza tra dati reddituali di natura “campionaria” e “amministrativa” (integrati) occorre innanzitutto fare chiarezza su ciò che attiene la possibilità di stabilire o meno un abbinamento tra le fonti in questione. Infatti, generalizzando il discorso, si può parlare di incoerenza solo quando due o più fonti del dato presentano valori difformi su una stessa unità di analisi. Resta implicita in questa definizione che l’unità sia osservabile su più fonti del dato (almeno due) e che dunque vi sia un abbinamento tra unità statistiche delle varie fonti. Se si prende a titolo illustrativo la figura 3.9, si nota che il nostro campo di osservazione è ristretto alle sole unità campionarie che trovano un accoppiamento con i dati dell’archivio amministrativo integrato (Ω'): sottoinsiemi **A** (unità abbinate con redditi da lavoro dipendente) e **B** (unità abbinate con redditi da lavoro dipendente).

Limitandoci ad esaminare i casi in cui è possibile stabilire un abbinamento tra le fonti ($\Theta' = \{\omega \in \Theta, \vartheta \in \Omega' : \omega = \vartheta\}$, dove $\Theta' = \mathbf{A} \cup \mathbf{B}$), emergono due tipologie di incoerenze: indicate a titolo illustrativo con i codici **cf1** e **cf4** in figura 3.9. Formalizzando il discorso, dato che ogni unità è posta in corrispondenza biunivoca con il suo codice fiscale, la *prima tipologia di incoerenza* è associata alle unità del sottoinsieme $\mathbf{B}' = \{b \in \mathbf{B}, z \in \mathbf{Z} : b = z\}$, formato dalle unità del campione che dichiarano nell’intervista di aver percepito un reddito da lavoro dipendente e che sono abbinate ad unità dell’archivio amministrativo integrato non aventi questa tipologia di reddito nelle rispettive dichiarazioni fiscali; la *seconda tipologia di incoerenza* è a sua volta legata alle unità del sottoinsieme $\mathbf{A}' = \{a \in \mathbf{A}, w \in \mathbf{W} : a = w\}$, rappresentate dalle unità campionarie che dichiarano di non possedere redditi da lavoro dipendente e che tuttavia sono abbinate a unità dell’archivio amministrativo con redditi di questa tipologia nelle rispettive dichiarazioni fiscali. Non si può invece parlare di incoerenza in relazione alle unità campionarie titolari di reddito da lavoro dipendente Θ_L ($\Theta_L = \mathbf{B} \cup \mathbf{C}$) dove non sia possibile stabilire alcun legame con le unità di fonte amministrativa: sottoinsieme $\mathbf{C} = \{c \in \Theta_L, \omega \in \Omega' : c \neq \omega\}$. Come vedremo in seguito, la procedura utilizzata dirime prima i casi di incoerenza tra i percettori di reddito da lavoro dipendente e subordinatamente quelli tra i percettori di reddito autonomo. Le assunzioni fatte per risolvere le incoerenze utilizzano, tra l’altro, le dichiarazioni degli intervistati circa la loro condizione lavorativa attuale e pregressa e il calendario lavorativo retrospettivo (anno di riferimento dei redditi).

In sintesi, il confronto tra i redditi da lavoro dipendente delle due tipologie di fonti si realizza attraverso l’identificazione dei possibili casi di incoerenza o di coerenza (Prospetto 3.5).

Prospetto 3.5 - Analisi delle incoerenze tra redditi da lavoro dipendente di fonte campionaria e amministrativa

FONTE AMMINISTRATIVA INTEGRATA	Fonte campionaria	
	Percettori di reddito da lavoro dipendente	Non percettori di reddito da lavoro dipendente
Percettori di reddito da lavoro dipendente	Coerenza (1)	Incoerenza (2)
Non percettori di reddito da lavoro dipendente	Incoerenza (1)	Coerenza (2)

Come si osserva dal prospetto 3.5 i nodi critici sono rappresentati dalle incoerenze del primo e del secondo tipo, ovvero dalle unità che secondo la fonte amministrativa percepiscono redditi da lavoro dipendente mentre non li percepiscono in base alla fonte campionaria e viceversa. La soluzione ai problemi di incoerenza, come vedremo in seguito, si ottiene tenendo conto della struttura dei redditi rispettivamente di fonte campionaria e amministrativa. La stessa analisi delle incoerenze è applicata al caso dei redditi da lavoro autonomo. Viceversa, nei casi di “coerenza del primo tipo”, cioè quando ambedue le fonti segnalano la presenza di redditi da lavoro dipendente per la stessa unità statistica, si procede ad analizzare i livelli di reddito (variabile derivata) rilevati rispettivamente dalle due fonti e a esaminare eventuali scostamenti. Le ipotesi che vengono utilizzate per “riconciliare” le difformità tra i livelli di reddito rilevati nelle fonti citate differiscono a seconda della tipologia di reddito analizzata.

Le strategie utilizzate per risolvere l’ “incoerenza del primo tipo” (intervistati che dichiarano all’intervista di aver percepito reddito da lavoro dipendente ma che sono abbinati a unità dell’archivio amministrativo integrato che non possiedono tale retribuzione) hanno comportato: la riclassificazione di questa tipologia di reddito in altre componenti economiche certificate dalla fonte fiscale, il ricalcolo o ancora l’integrazione del dato con altre fonti di guadagno rilevabili esclusivamente dal fisco. Le informazioni utilizzate ai fini delle azioni da intraprendere hanno riguardato: la condizione professionale svolta nei 12 mesi dell’anno dei redditi come da questionario (calendario retrospettivo), i livelli e la composizione del reddito presente nelle rispettive fonti del dato. Nell’indagine Eu-Silc 2006 tali incoerenze hanno coinvolto un totale di 822 individui. Le azioni intraprese nei riguardi delle suddette incoerenze hanno portato alla:

- 1) eliminazione della componente di reddito da lavoro dipendente con sostituzione della componente reddito da co.co.co. di fonte amministrativa³⁸ (112 casi);
- 2) rimozione del reddito da lavoro dipendente campionario e subentro del reddito da lavoro autonomo di provenienza fiscale (130 casi);
- 3) rettifica della struttura dei redditi campionaria (permane il reddito da lavoro dipendente) con integrazione della componente di reddito da collaborazione coordinata e continuativa di sola fonte fiscale³⁹ (43 casi);
- 4) rettifica della struttura dei redditi campionaria (permane il reddito da lavoro dipendente) con inserimento del reddito da lavoro autonomo (113 casi);
- 5) trasformazione dello status di percettore di reddito dipendente campionario in favore di borsista, percettore di sussidi di disoccupazione eccetera (16 casi);
- 6) conservazione dello status originario di percettore di reddito da lavoro dipendente e della relativa struttura di reddito di natura campionaria (408 casi).

Le “incoerenze del secondo tipo”, contrassegnate dall’assenza di dati individuali sul reddito da lavoro dipendente da questionario e dalla contestuale presenza di questo tipo di reddito nelle dichiarazioni fiscali, hanno interessato un totale di 1.841 soggetti per l’indagine It-Silc 2006. L’ipotesi di lavoro adottata per la soluzione di quest’altra forma di incoerenza è che il reddito da lavoro dipendente di fonte fiscale debba integrare e completare le informazioni campinarie. Esso dunque va inserito nella struttura dei redditi complessivi, a meno di segnali di misclassificazione della tipologia di reddito fra le due fonti. In presenza di elementi probatori, colti in sede di intervista, tali da far reputare come estremamente inverosimile la presenza di un reddito dipendente si procede alla relativa riallocazione. Dal punto di vista operativo si è

³⁸ Si applica ai soggetti che nel calendario retrospettivo inerente alla condizione professionale da questionario dichiarano di non aver mai assunto la condizione di lavoratore dipendente, ovvero dichiarano all’indagine un reddito netto complessivo paragonabile (cioè con uno scarto compreso tra +/- 20 per cento) al reddito da collaborazione coordinata e continuativa della fonte fiscale.

³⁹ Si assume corretto il livello di reddito da co.co.co. rilevato al fisco e si ricava il valore del reddito da lavoro dipendente come differenza tra il dato campionario (reddito dipendente) e l’importo fiscale.

proceduto a verificare la presenza di una serie di segnali del questionario discordanti rispetto all'assegnazione di parte amministrativa. Tra i segnali del questionario che depongono per la riallocazione del reddito da lavoro dipendente verso altre forme sono stati considerati: *i*) assenza di una condizione professionale da lavoratore dipendente in tutti i 12 mesi dell'anno di riferimento dei redditi (calendario retrospettivo); *ii*) posizione professionale da lavoratore autonomo e nessun cambiamento di attività nei 12 mesi precedenti; *iii*) posizione professionale dell'attività svolta in passato diversa da lavoro dipendente. In altri termini, in assenza di segnali di lavoro dipendente e in presenza di segnali di lavoro di tipo autonomo nel periodo temporale di riferimento dei redditi si reputa che il reddito da lavoro di fonte fiscale sia più verosimilmente riferibile al lavoro autonomo di cui il questionario porta evidenza. Di conseguenza, il reddito di fonte fiscale viene concordemente riclassificato.

Volendo esemplificare i passi della procedura, si può immaginare di suddividere il problema della riallocazione del reddito in due passi: a seconda della presenza o meno nel questionario di componenti di reddito diverse dal lavoro dipendente. In assenza di altre forme di reddito campionario, si accetta la componente di reddito da lavoro unicamente rilevata nella fonte fiscale⁴⁰ (1.154 casi). Nel caso opposto, cioè in presenza di altre tipologie di reddito campionario (687 casi), si procede all'analisi dei segnali provenienti dalla fonte campionario. In ben 217 casi, la procedura applicata all'indagine Eu-Silc 2006 riclassifica i redditi da lavoro dipendente di provenienza fiscale verso altre tipologie (coerentemente con l'informazione campionario).

Quello che si intende evitare, attraverso la rimozione delle incoerenze del primo e del secondo tipo, è l'inclusione di tipologie di reddito che, rispetto alle finalità dell'indagine, andrebbero a ragione classificate diversamente; ovvero la presenza di errori di duplicazione nei redditi, o relative sottovoci, in quanto già dichiarati in sede d'indagine ma classificati in maniera differente.

L'ipotesi sottostante la "riconciliazione" dei livelli retributivi individuali tra le due fonti del dato (coerenza del primo tipo) è che il "valore vero" sia contenuto nella fonte fiscale, sempre che il reddito rilevato nel questionario "non risulti superiore" a quello di origine fiscale. In quest'ultima circostanza si è tentato di riconciliare le discrepanze attraverso un sistema di regole applicate selettivamente in funzione della distanza dei livelli di reddito tra le due fonti.

In riferimento a differenze relative tra il dato campionario e amministrativo comprese nell'intervallo (0,1), cioè quando il valore campionario è al più due volte il valore fiscale, sono stati applicati alcuni criteri decisionali legati a: *i*) esistenza di elementi retributivi da questionario totalmente o parzialmente esenti da tassazione; *ii*) attendibilità delle risposte fornite dall'intervistato. La presenza nel questionario di taluni compensi parzialmente o totalmente esenti da tassazione ha fatto propendere nella correttezza dell'informazione di natura campionario.⁴¹ Viceversa in assenza di questi compensi si è privilegiato il dato amministrativo a patto che la differenza relativa tra valori delle due fonti risultasse comunque contenuta in una soglia del 10 per cento. In mancanza di queste due condizioni, si è proceduto a valutare l'attendibilità della risposta dell'intervistato relativamente al reddito dichiarato, ricorrendo al giudizio del rilevatore o all'identificazione del rispondente: persona interessata/altro familiare (*proxy*). Il dato campionario viene convalidato solo se il giudizio del rilevatore è positivo (risposte abbastanza/molto verosimili) e se, al tempo stesso, la persona rispondente coincide con la persona interessata. L'ipotesi sottostante la convalida del dato campionario è che il soggetto

⁴⁰ Vi è da osservare che tale regola non è stata applicata in soli 9 casi in quanto le altre informazioni di natura amministrativa portavano a riclassificare questi redditi verso altre tipologie: compensi a soci di cooperativa, borse di studio eccetera.

⁴¹ Di conseguenza si è assunto che la differenza nei livelli delle due fonti è ascrivibile a tali elementi retributivi.

abbia percepito compensi fuori busta non risultanti dalle dichiarazioni dei redditi. Al contrario, se vengono meno i requisiti di affidabilità, si considera corretta l'informazione amministrativa.

In corrispondenza, poi, di forti discrepanze nei valori delle retribuzioni tra le fonti, rappresentate da differenze relative superiori al 100 per cento tra il dato campionario rispetto all'amministrativo, si è stabilito di considerare la "distanza" tra i "valori osservati nelle due fonti" e il "valore imputato" dal *package* statistico IVEware⁴² sulla stessa variabile reddito dipendente una volta eliminato il valore assunto dalla corrispondente unità campionaria. Tra le variabili utilizzate nel modello di regressione multipla incluso nelle procedure di IVEware sono stati considerati diversi elementi correlati con il tenore di vita del soggetto (possesso di beni durevoli, titolo di godimento e caratteristiche dell'abitazione principale, indicatori di deprivazione eccetera) e quindi indirettamente con il suo livello di reddito.

3.5.3 - Riconciliazione del reddito autonomo tra fonti fiscali e campionaria

La soluzione al problema delle incoerenze del primo e secondo tipo sui redditi da lavoro autonomo e parasubordinato (co.co.co.) è stata elaborata in modo tale da non contraddire i risultati della precedente attribuzione di reddito da lavoro dipendente. In questa fase sono state quindi considerate esclusivamente le possibili incoerenze derivanti dalla diversa allocazione tra fonte campionaria e fonte fiscale di queste due sole componenti economiche.

Indirizzando la questione sull' "incoerenza del primo tipo" della componente da lavoro autonomo⁴³ (in senso stretto) si possono configurare quattro diverse combinazioni (esiti) nella struttura dei dati (Prospetto 3.6):

1) presenza di reddito da lavoro autonomo e da collaborazione coordinata e continuativa di fonte campionaria ed esistenza della sola componente co.co.co di fonte amministrativa;

2) presenza del solo reddito autonomo (in senso stretto) da questionario Eu-Silc ed esistenza della sola componente co.co.co nel dato amministrativo;

3) presenza di ambedue le tipologie di reddito da fonte campionaria e assenza di questi stessi redditi di natura amministrativa;

4) presenza del solo reddito da lavoro autonomo (in senso stretto) campionario e assenza di redditi di autonomi e parasubordinati da fonte amministrativa. In relazione a ciascuna di queste quattro combinazioni si procede a individuare e correggere la possibile causa di incoerenza.

⁴² Sviluppato dal *Survey Research Center Institute for Social Research dell'University of Michigan*, il package si compone di diverse applicativi in linguaggio SAS e Fortran che realizzano l'imputazione multivariata dei valori mancanti tramite una sequenza di modelli di regressione multipla univariata (approccio "variable by variable"), partendo dalla variabile con minore tasso di non risposta ed utilizzando come predittori le variabili senza valori mancanti. Ciascun modello mette in relazione tra la variabile da imputare con tutte le altre variabili e utilizza sia i valori osservati sia i valori imputati nei passi precedenti. Il singolo modello è selezionato separatamente senza bisogno di definire un modello multivariato per l'intero insieme dei dati di cui, però, si assume l'esistenza. A seconda della natura della variabile da imputare è selezionato un differente modello di regressione tra quelli disponibili. Così, ad esempio per le variabili continue fa riferimento al modello di regressione lineare normale (con eventuale trasformazione della variabile se non è normale nella scala originaria), viceversa per le variabili binarie utilizza il modello di regressione logistico, infine per le variabili categoriali considera i modelli di tipo log-lineare. Per approfondimenti si rinvia a: Manzari A., *Valutazione comparativa di alcuni metodi di imputazione singola delle mancate risposte parziali per dati quantitativi*. Roma: Istat, 2004. (Documenti Istat, n. 18); Vitaletti S. "Correzioni e imputazioni delle informazioni sui redditi". In *L'indagine europea sui redditi e le condizioni di vita delle famiglie (Eu-Silc)*, par. 4.4. Roma: Istat. 2008. (Metodi e Norme, n. 37); Raughunatan E. et al. "A Multivariate technique for multiply imputing missing values using a sequence of regression models". *Survey methodology*, Vol. 27, n. 1, pp. 85-95, 2001.

⁴³ Essa attiene alle unità del campione che dichiarano nell'intervista di aver percepito un reddito da lavoro autonomo e che sono abbinate ad unità dell'archivio amministrativo integrato non aventi questa tipologia.

Prospetto 3.6 - Analisi delle incoerenze tra redditi da lavoro autonomo di fonte campionaria e amministrativa

FONTE AMMINISTRATIVA INTEGRATA	Fonte campionaria Eu-Silc	
	Presenza di reddito da lavoro autonomo (in senso stretto)	
	Presenza di reddito da co.co.co.	Assenza di reddito da co.co.co.
Assenza di reddito da lavoro autonomo (in senso stretto)		
<i>Presenza di reddito da co.co.co.</i>	Esito 1 (3 casi)	Esito 2 (187 casi)
<i>Assenza di reddito da co.co.co.</i>	Esito 3 (13 casi)	Esito 4 (485 casi)

In merito al punto 1, l'incoerenza è risolta tenendo conto dei livelli di reddito assunti nelle due fonti e ricorrendo a due distinte assunzioni: 1) presenza di errori di "duplicazione" nelle due componenti di reddito da lavoro autonomo e parasubordinato campionarie (replicazione delle informazioni sul questionario); 2) "evasione totale" del reddito da lavoro autonomo dichiarato al fisco ma riportato sul questionario.

In relazione al punto 2 la soluzione all'incoerenza è a sua volta stata articolata in tre punti a seconda che: *i*) l'attività lavorativa principale autodichiarata nell'intervista includa le voci: "collaborazione coordinata e continuativa" o "prestazione d'opera occasionale"; *ii*) la differenza relativa tra il reddito da lavoro autonomo campionario e il reddito da co.co.co. di fonte fiscale sia inferiore o uguale al 20 per cento; *iii*) l'attività lavorativa principale autodichiarata sia diversa da quella del punto *i* e la differenza relativa tra le componenti di cui al punto *ii* risulti maggiore del 20 per cento (domina il livello di reddito campionario).

Nelle situazioni riferite ai punti *i* e *ii* si ipotizza la presenza di errori di misclassificazione tra le due componenti da parte degli intervistati e di conseguenza si interviene sostituendo la tipologia di reddito co.co.co. desunta dal fisco al reddito autonomo campionario. Nel caso *iii* si assume, invece, la compresenza delle due componenti di reddito (autonomo e parasubordinato) e si calcola il livello di reddito da lavoro autonomo (in senso stretto) come differenza tra il dato campionario (autonomo) e il dato fiscale (co.co.co.).

Infine, per le due restanti combinazioni di reddito, di cui a punti 3 e 4, si ipotizza che gli archivi fiscali siano affetti da deficit informativi e si assume valida la struttura di reddito di provenienza campionaria.

Se si analizza l' "incoerenza del primo tipo", relativamente all'altra componente reddituale da collaborazione coordinata e continuativa,⁴⁴ si perviene anche in questa circostanza a quattro diverse combinazioni. Vi è da osservare che una di queste coincide esattamente con la combinazione al punto 4 appena descritta (esito 4 del prospetto 3.6). Mentre un'altra, in analogia con la casistica 3 sopra citata, non presenta problematicità e pertanto non interviene sulla struttura del reddito campionario. Le situazioni più complesse riguardano:

- 1) la contestuale presenza di reddito autonomo e parasubordinato di fonte campionaria a fronte dell'esistenza della sola componente autonoma (in senso stretto) di fonte amministrativa (15 casi);
- 2) la discordanza nella classificazione dei redditi tra le fonti; ciò che nel dato campionario è classificato come reddito da co.co.co. si presenta nel dato fiscale come reddito autonomo (46 casi).

⁴⁴ Essa riguarda le unità del campione che dichiarano nell'intervista di aver percepito un reddito da co.co.co. e che sono abbinate ad unità dell'archivio amministrativo integrato non aventi questa tipologia.

Nella prima circostanza si ipotizza valida la struttura del reddito di natura fiscale e pertanto si impone l'intervento della sola componente di reddito autonomo. Parallelamente, si controlla/corregge la presenza di eventuali errori di duplicazione, segnalati da importi comparabili⁴⁵ nelle due componenti di reddito campionarie.

Nel secondo caso l'incoerenza è spiegata dalla presenza di errori di misclassificazione, per cui l'intervistato interpreta la propria attività come collaborazione coordinata e continuativa (occasionale) anziché come prestazione professionale (e quindi non occasionale), coerentemente con la relativa documentazione fiscale. In questo caso si interviene riallocando il reddito da co.co.co. in reddito da lavoro autonomo.

Se si passa a esaminare l' "incoerenza del secondo tipo" sui redditi da lavoro autonomo,⁴⁶ si giunge identicamente a quattro distinte combinazioni nella struttura del reddito. Tra queste una coincide con la seconda casistica associata all'incoerenza del primo tipo sui redditi co.co.co., mentre le altre due segnalano la presenza di redditi esclusivamente da fonte amministrativa. Pertanto, l'unica combinazione rilevante ai fini dell'analisi è quella che vede il concorso di ambedue le tipologie di reddito da fonte amministrativa e la presenza della sola componente di reddito da co.co.co. campionaria (47 casi). In tale circostanza si ipotizza corretta la struttura dei redditi da fisco e si aggiunge quindi alla struttura campionaria un'ulteriore voce di entrata: il reddito autonomo (in senso stretto). Il valore di quest'ultima componente è identico al dato di fonte fiscale, mentre l'importo del reddito da co.co.co. è pari: *i*) al suo valore campionario, quando quest'ultimo è inferiore alla somma delle due componenti di reddito rilevate dal fisco; *ii*) alla "differenza" tra il suo "valore campionario" e il "livello di reddito autonomo di fonte fiscale", quando viceversa il primo risulta superiore alla somma degli importi fiscali.

Infine, per quanto concerne l' "incoerenza del secondo tipo" sui redditi da co.co.co.⁴⁷ si perviene, *mutatis mutandis*, a un risultato analogo a quello applicato alla stessa incoerenza sui redditi autonomi. In questo caso l'unica combinazione rilevante è quella che vede il concorso di entrambe le componenti di reddito fiscali e la sola rilevazione dei redditi da lavoro autonomo di natura campionaria (144 casi). Anche in questo caso si impone la struttura di reddito fiscale e si aggiunge però come componente nuova di reddito quella da co.co.co.. Per il calcolo degli importi relativi alle due componenti valgono le stesse regole applicate al passo precedente con componenti di reddito invertite.

In conclusione, la riconciliazione tra i livelli di reddito autonomo di fonte campionaria e amministrativa è ottenuta assegnando il massimo dei due valori nell'ipotesi che in questo modo riduce al minimo il grado di sottostima del reddito vero. 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} . Escludendo la possibilità di over-reporting, cioè che il rispondente dichiarerà all'intervista un reddito netto superiore a quello effettivamente percepito, la scelta del maggiore fra il dato campionario e 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.

⁴⁵ Nel senso che la differenza relativa tra i valori delle rispettive componenti campionarie ricade nell'intervallo compreso tra il +/- 20 per cento.

⁴⁶ Essa coinvolge le unità del campione che dichiarano nell'intervista di non aver percepito alcun reddito da lavoro autonomo e che tuttavia sono abbinate a unità dell'archivio amministrativo integrato percettrici di questa forma di reddito.

⁴⁷ Essa coinvolge le unità del campione che dichiarano nell'intervista di non aver percepito alcun reddito da lavoro autonomo e che tuttavia sono abbinate a unità dell'archivio amministrativo integrato percettrici di questa forma di reddito.

Capitolo 4 - La strategia di costruzione dei redditi lordi Eu-Silc: microsimulazione e integrazione dei dati campionari e amministrativi

4.1 - Introduzione

Il progetto Eu-Silc è stato avviato nel 2003 sulla base di un “*gentlemen agreement*” in sei paesi dell’Unione europea (Austria, Belgio, Danimarca, Grecia, Irlanda, Lussemburgo) oltre che in Norvegia. Nel 2004 l’indagine entra ufficialmente a regime in 12 paesi, fra cui l’Italia, con una base normativa dettata dal regolamento del Parlamento europeo e del Consiglio (n. 1177/2003). Dal 2005 l’indagine Eu-Silc è estesa ai 25 paesi dell’Unione, oltre che in Norvegia e Islanda. L’approccio innovativo voluto con la nuova indagine europea consente ai paesi aderenti al progetto di avere un “quadro di riferimento comune” più che “un’indagine comune” sul reddito, come era nel caso del precedente panel europeo Echip.¹ Da questo punto di vista, il regolamento dell’indagine Eu-Silc definisce una lista di variabili primarie e secondarie armonizzate, le procedure, le linee guida comuni e una classificazione delle variabili che consenta una buona comparabilità delle statistiche prodotte.

Fin dall’inizio del progetto Eu-Silc è stato quindi previsto un periodo di transizione, fino all’anno di indagine 2007, che consentisse ai paesi aderenti di sperimentare gradualmente le diverse metodologie e di armonizzare le tecniche di costruzione per alcune componenti di reddito, fra cui i fitti imputati delle abitazioni, le retribuzioni in natura dei lavoratori dipendenti (*fringe benefits*), gli autoconsumi, i redditi lordi e i contributi sociali posti a carico del datore di lavoro.

In particolare per la costruzione dei redditi lordi, diversi paesi, fra cui l’Italia, in cui le informazioni sull’imposizione fiscale e contributiva non venivano rilevate con l’indagine né venivano utilizzati gli archivi amministrativi, dovevano avviare studi di fattibilità per la conversione dei redditi netti campionari in lordi, preferibilmente attraverso tecniche di modellizzazione piuttosto che di semplice imputazione statistica.

L’Istat ha avviato uno studio di fattibilità nel 2004 per la costruzione dei redditi lordi analizzando le diverse alternative possibili. Per misurare l’imposizione fiscale e contributiva a livello individuale è infatti possibile: *i*) stimare le imposte e i contributi sociali attraverso un modello di microsimulazione; *ii*) utilizzare gli archivi fiscali disponibili; *iii*) rilevare le imposte e i contributi sociali attraverso l’indagine campionaria.

La strategia di rilevazione adottata dall’Istat per l’indagine Eu-Silc aveva escluso a priori la raccolta diretta dei redditi lordi nel corso dell’intervista alle famiglie, per le difficoltà di misurazione e di conoscenza da parte dell’intervistato dei dati sulle fonti e sull’ammontare, per la reticenza o diffidenza nel caso fosse stato chiesto al rispondente di consultare i documenti fiscali, motivi che avrebbero potuto causare elevati tassi di mancata risposta oltre a implicare un carico eccessivo per il rispondente.

Le strade percorribili erano dunque il ricorso a un modello di microsimulazione per la conversione dei redditi netti campionari in lordi oppure l’utilizzo degli archivi amministrativi. In entrambi i casi occorreva valutare i vantaggi e gli eventuali svantaggi di ciascuna tecnica.

Il capitolo è a cura di Gabriella Donatiello

¹ European Community Household Panel.

Le variabili richieste dal progetto Eu-Silc intendono misurare il benessere economico corrente delle famiglie e degli individui. Principale obiettivo dell'indagine è la produzione di statistiche annuali e longitudinali comparabili sul reddito e il livello di povertà ed esclusione sociale nell'Unione europea. La povertà a livello europeo è misurata convenzionalmente in base al reddito disponibile e per stimare l'impatto redistributivo delle misure di politica economica e fiscale è necessario disporre di redditi lordi disaggregati per tipologia.

La definizione di reddito accolta in Eu-Silc è quella internazionalmente accettata e basata sulla definizione economica di "reddito monetario disponibile", inteso come somma dei consumi e dei risparmi effettuati in un dato periodo di tempo. Le diverse fonti che in Eu-Silc concorrono alla formazione del reddito disponibile sono:

- i redditi da lavoro (dipendente, autonomo);
- i redditi da capitale (reale e finanziario);
- i trasferimenti pubblici (pensioni e altri trasferimenti);
- i trasferimenti privati (aiuti in denaro da altre famiglie o istituzioni).

Per stimare il reddito lordo occorre aggiungere al reddito disponibile le informazioni relative al prelievo fiscale e contributivo obbligatorio.

Secondo il Regolamento Eu-Silc, il reddito lordo totale è pari alla somma dei redditi netti, delle imposte personali sui redditi, delle imposte patrimoniali e dei contributi sociali posti a carico dei lavoratori e dei datori di lavoro. Per stimare le imposte dirette è necessario disporre anche di informazioni dettagliate sugli oneri deducibili e detraibili, oltre che sulle imposte anticipate e sui saldi d'imposta.

A tal fine, lo studio di fattibilità avviato dall'Istat nel 2004 per la costruzione dei redditi lordi ha analizzato due differenti approcci, ossia la possibilità di utilizzare: (i) le imposte e i contributi sociali derivanti dagli archivi amministrativi disponibili; (ii) i redditi netti campionari lordizzati attraverso tecniche di modellizzazione del sistema fiscale nazionale.

L'utilizzo esclusivo degli archivi amministrativi o della tecniche di microsimulazione è stato attentamente valutato sulla base dei requisiti richiesti dalle variabili target dell'indagine.

Quando si utilizzano gli archivi fiscali per costruire le statistiche sui redditi individuali va considerato che possono insorgere diversi problemi dovuti a:

- definizione delle variabili di reddito: la definizione di reddito tassabile può essere diversa da quella utilizzata per lo studio della distribuzione dei redditi e la definizione stessa di unità tassabile può divergere;
- classificazione dei percettori di reddito: vi possono essere differenze rilevanti nelle classificazioni, ad esempio i lavoratori parasubordinati e i soci di cooperativa sono per il fisco assimilati al lavoro dipendente mentre sono autonomi in Eu-Silc;
- copertura incompleta rispetto ai redditi rilevati nell'indagine e/o agli individui del campione.

Per quanto riguarda la copertura dell'informazione disponibile, gli archivi fiscali non contengono alcun dato su alcune categorie di reddito: vi possono essere infatti redditi totalmente esenti da imposta (le pensioni sociali, le indennità familiari, alcune pensioni di invalidità, i redditi al di sotto della soglia di esenzione), redditi soggetti a imposta sostitutiva (le attività finanziarie), redditi a tassazione separata (gli arretrati da lavoro e da pensione) e vanno considerate inoltre le deduzioni legali dall'imponibile (l'erosione fiscale che consente anche riduzioni consistenti dell'imposta lorda) e quelle illegali (evasione fiscale).

I redditi esenti, i redditi soggetti a imposta sostitutiva o a tassazione separata rientrano tuttavia nella definizione di reddito disponibile di Eu-Silc, mentre l'evasione fiscale, al pari dell'*under-reporting* (sottodichiarazione), non è richiamata nel regolamento Eu-Silc per le difficoltà oggettive a utilizzare una metodologia condivisa di stima. I redditi netti derivanti dalle

fonti fiscali escludono quindi una parte del reddito effettivamente percepito e i redditi imponibili delle dichiarazioni fiscali non sono distinti per componente (da lavoro, da pensioni eccetera).

Rispetto all'informazione contenuta nei dati amministrativi, Eu-Silc rileva anche i lavoratori del sommerso, che sono sconosciuti all'anagrafe tributaria, e gli individui che non è stato possibile agganciare con i dati dell'Agenzia delle Entrate perché non erano presenti nelle liste campionarie (i cosiddetti componenti di fatto delle famiglie presenti al momento dell'intervista), oppure perché le informazioni anagrafiche necessarie alla procedura di *record linkage* risultavano incomplete.

Ne consegue che le informazioni desumibili dalle fonti fiscali ossia dalle dichiarazioni dei redditi dei contribuenti o dei sostituti di imposta devono comunque essere integrate da altre fonti amministrative (ad esempio con il Casellario pensioni dell'Inps), ma non sempre si ottiene tutta l'informazione necessaria per costruire le variabili obiettivo. Gli archivi attualmente disponibili non contengono, ad esempio, alcuna informazione sui contributi sociali a carico dei datori di lavoro e dei lavoratori del settore privato e coprono solo parzialmente i contributi versati dai lavoratori autonomi.

Vi sono tuttavia diversi problemi da affrontare anche quando si decide di utilizzare un modulo di conversione dei redditi campionari di un modello di microsimulazione. Va considerato innanzitutto che qualunque sottostima o sovrastima del dato netto campionario si traduce in una sottostima o sovrastima dei redditi lordi e in quest'ultimo caso l'aumento dell'imposta progressiva risulta più che proporzionale, con evidenti effetti negativi sulla qualità delle stime prodotte.

Si deve quindi ricorrere all'utilizzo di alcune metodologie che consentono di ridurre la sottostima o la sovrastima dei redditi lordi, ad esempio attraverso la minimizzazione dell'*under-reporting* dei redditi campionari che riguarda soprattutto alcune tipologie di reddito (lavoro autonomo, benefici assistenziali). In altri casi si tenta di misurare la differenza fra redditi netti campionari e redditi netti fiscali, ricorrendo di solito a ipotesi arbitrarie o a imputazioni dei dati da fonti esterne. Le differenze tra i redditi campionari e i fiscali sono dovute in realtà a informazioni che difficilmente possono essere rilevate attraverso le indagini sui redditi e che riguardano le deduzioni legali e illegali dall'imponibile, le differenze fra redditi effettivi e redditi figurativi (ad esempio i redditi catastali, gli studi di settore per gli autonomi) e tutte le spese per consumo che rappresentano oneri deducibili e detraibili per il contribuente. Il ricorso a tali metodologie di "aggiustamento" del dato netto campionario prima della microsimulazione risulta, inoltre, particolarmente dispendioso in termini di risorse e di tempo richiesti.

Per la costruzione dei redditi lordi dell'indagine Eu-Silc, fin dal 2001 Eurostat aveva incaricato un gruppo di ricerca dell'Università di Siena di predisporre un modello di microsimulazione per la conversione dei redditi campionari.² L'Istat ha deciso di testare il modello SM2 e di sperimentare una metodologia più complessa del consueto utilizzo di un modello di microsimulazione per la conversione di redditi campionari, considerando la complessità della normativa fiscale e la disponibilità dei dati di fonte amministrativa.

Nonostante i problemi connessi all'utilizzo delle fonti amministrative, le informazioni contenute nelle dichiarazioni fiscali sui redditi netti e lordi, sugli oneri deducibili e detraibili, sui contributi sociali, sulle imposte, sugli acconti versati e sui saldi finali risultano comunque estremamente utili. I dati dell'Agenzia delle entrate sono, inoltre, la fonte più attendibile per quanto riguarda le imposte pagate dai contribuenti. Utilizzando i dati fiscali era inoltre possibile sperimentare l'integrazione delle informazioni tributarie in un modello di microsimulazione.

² Si veda il Capitolo 2 dedicato al modello di microsimulazione SM2 dell'Università di Siena.

In definitiva, l'utilizzo congiunto e innovativo del modello di microsimulazione e degli archivi amministrativi consentiva non solo di superare alcuni dei problemi menzionati precedentemente ma potenziava i vantaggi ottenibili dall'impiego esclusivo degli archivi fiscali da una parte e delle tecniche di microsimulazione dall'altra.

4.2 - La versione preliminare del modello SM2-Eu-Silc

Il complesso processo di produzione dei redditi lordi Eu-Silc, iniziato nel 2004, è stato caratterizzato da una fase iniziale di studio del modello SM2 dell'Università di Siena e di successiva implementazione di una versione preliminare del modello SM2-Eu-Silc per gli anni di indagine 2004, 2005 e 2006. A questa prima fase di sperimentazione è seguito poi lo sviluppo della versione definitiva del modello SM2-Eu-Silc 2007, 2008 e 2009, con l'utilizzo congiunto delle microsimulazione e del *record linkage* per la costruzione delle variabili obiettivo di reddito lordo individuali e familiari.

Il modello SM2 è stato costruito nel 2003 con i dati disponibili dell'indagine Echip Udb Pdb³ dell'anno 1998 e con regole fiscali riferite all'anno 2003 per l'Italia, Francia e Spagna, paesi a cui è stata applicata la procedura di conversione sviluppata dall'Università di Siena. Nella prima fase di lavoro, l'Istat ha eseguito uno studio approfondito della struttura del modello SM2, dell'algoritmo di conversione dei redditi campionari e delle *routines* utilizzate per il calcolo della tassazione e dei contributi sociali per potere innanzitutto adattare il modello SM2 ai dati della nuova indagine Eu-Silc.

L'implementazione della versione preliminare del modello SM2-Eu-Silc ha infatti richiesto la transizione dalla precedente indagine Echip alla nuova indagine 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:

1. preparazione della base dati del modello:
 - sostituzione delle variabili Echip con le nuove variabili Eu-Silc;
 - costruzione delle variabili ausiliarie;
2. modifica di alcune procedure di conversione netti-lordi di SM2 per stimare:
 - il reddito da lavoro autonomo;
 - l'Irap;
 - i contributi sociali dei lavoratori parasubordinati.

La base dati di un modello di microsimulazione è il risultato finale di diversi programmi di calcolo necessari alla predisposizione dei microdati e delle variabili ausiliarie. In SM2 le variabili ausiliarie identificano, per i tre paesi a cui è stato applicato, un set di informazioni non rilevate direttamente dall'indagine ma necessarie per modellizzare la normativa fiscale nazionale.

La base dati del modello SM2-Eu-Silc è stata preparata utilizzando la prima *wave* dell'indagine Eu-Silc (anno 2004, con periodo di riferimento dei redditi anno 2003) ed effettuando una sostituzione diretta, dove possibile, delle variabili target Eu-Silc alle corrispondenti variabili Echip. Successivamente sono state individuate e costruite tutte le variabili ausiliarie necessarie alla procedura di conversione dei dati campionari.

Fra le numerose variabili ausiliarie necessarie al modello, particolare attenzione è stata riservata alla costruzione delle *tax units* all'interno delle famiglie e all'individuazione dei settori di attività economica essenziali per stimare i contributi sociali del lavoro dipendente e autonomo.

³ Echip Users' Data Base (Udb) e Production Data base (Pdb).

Come è noto, il sistema fiscale nazionale è basato su un'imposizione a livello individuale e anche nel caso di dichiarazione congiunta dei coniugi la scelta tra tassazione separata o congiunta riguarda in definitiva solo la fase finale del pagamento delle imposte. Poiché le imposte personali sui redditi sono dovute sulla somma dei redditi tassabili è necessario ricostruire la tassazione sulle singole componenti di reddito e individuare tutte le condizioni personali in grado di modificare l'imposizione fiscale, come ad esempio l'incapienza o i carichi familiari.

Per quanto riguarda le *tax units*, in base al nostro ordinamento vi possono essere essenzialmente due tipi di "unità fiscali" a livello familiare:

- "la famiglia" che include il capofamiglia e le persone a carico; il reddito sottoposto a tassazione è soltanto quello del capofamiglia mentre i redditi delle persone a carico sono esenti da tassazione: vi è quindi un'unica *tax unit*;
- "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 a una detrazione se il loro reddito non supera la soglia di 2.840,51 euro all'anno. È stato quindi necessario ricostruire le famiglie rilevate dall'indagine per individuare:

- le famiglie nucleari con capofamiglia e i fiscalmente a carico;⁴
- i "membri isolati" che sono parte della famiglia ma che dichiarano un proprio reddito.

L'individuazione delle *tax units* in SM2-Eu-Silc si è avvalsa del software "Procedura famiglie" sviluppato dall'Istat per le indagini sociali, il quale consente di ricostruire le diverse relazioni di parentela nei nuclei familiari.⁵ Una volta ricostruiti i nuclei familiari è stato elaborato un programma specifico per individuare i familiari fiscalmente a carico e costruire le *tax units* a cui applicare le regole di tassazione.

Per stimare i contributi sociali del lavoro dipendente e autonomo è necessario distinguere i diversi regimi contributivi e individuare con precisione il settore di attività economica per numero di addetti sulla base dei raggruppamenti Nace. Sono state quindi utilizzate le maggiori informazioni disponibili in Eu-Silc, rispetto a Echnp, riguardo alla posizione lavorativa dei rispondenti, al settore di appartenenza, al calendario retrospettivo della condizione professionale (numero di mesi) e al numero di addetti dell'unità locale dell'impresa, al fine di assegnare le aliquote contributive previste per l'industria e il settore pubblico.

Tra le variabili ausiliare, vi sono anche alcune informazioni dell'indagine campionaria che necessitano di modifiche per essere utilizzate nel modello di microsimulazione. Fra queste vi sono, ad esempio, le variabili rilevate solo a livello familiare, come i trasferimenti monetari non pensionistici alle famiglie, i trasferimenti privati tra le famiglie, o i redditi da capitale reale. Per calcolare l'imponibile individuale è necessario assegnare tali importi agli individui della famiglia. Nel file di input del modello, gli importi dichiarati sono stati attribuiti per convenzione al maggiore percettore di reddito all'interno della famiglia.

Solitamente nei modelli di microsimulazione alcuni redditi, come i redditi esenti, i redditi sottoposti a imposta sostitutiva o a tassazione separata, devono essere sottratti preliminarmente dal reddito imponibile Irpef in quanto non concorrono a determinare il valore dell'imposta. Il

⁴ Possono essere fiscalmente a carico:

- il coniuge non legalmente ed effettivamente separato;
- i figli, anche se naturali riconosciuti, adottivi, affidati o affiliati;
- i seguenti altri familiari:

- il coniuge anche legalmente ed effettivamente separato 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 da lui assegni alimentari non risultanti da provvedimenti dell'autorità giudiziaria.

⁵ Per un approfondimento si veda: Istat. *La misurazione delle tipologie familiari nelle indagini di popolazione*. Roma: Istat, 2010. (Metodi e Norme n. 46).

modello SM2 consente di superare tale limite in quanto, per le componenti di reddito che devono essere escluse dall'imponibile contributivo, basta inserire una deduzione specifica pari al reddito della componente che consente di sottrarre questa componente al totale del reddito sottoposto a tassazione, senza dovere separare tale componente dagli altri redditi. Allo stesso modo se una componente di reddito è tassata con un'aliquota costante (ad esempio l'Irap o i redditi da attività finanziaria), si introduce una deduzione pari al valore della componente e una detrazione specifica definita come una quantità negativa. In tal modo l'imposta finale è automaticamente incrementata dell'ammontare appropriato.⁶

Nella versione preliminare del modello Sm2-Eu-Silc, la transizione dai dati Echp ai dati Eu-Silc ha richiesto, oltre allo sviluppo di nuove procedure per la costruzione della base dati, anche la modifica di alcune *routines* di conversione netti-lordi, in particolare per la stima del reddito da lavoro autonomo e il calcolo dell'imposta regionale sulle attività produttive (Irap). Nell'indagine Echp, infatti, il reddito da lavoro autonomo era rilevato al lordo dell'imposta; in Eu-Silc invece i redditi autonomi sono al netto dell'imposizione fiscale. È stata modificata, inoltre, anche la procedura di stima dei contributi sociali previsti in caso di reddito derivante da attività di collaborazioni coordinate e continuative dei lavoratori parasubordinati.

4.2.1 - La parametrizzazione del sistema fiscale nazionale in SM2

Dopo la preparazione della base dati è stata implementata una versione preliminare del modello SM2-Eu-Silc introducendo le variabili dell'indagine Eu-Silc 2004 come prima modifica rilevante rispetto alle assunzioni e alle procedure previste in SM2.

Nel modello SM2 vi sono quattro *routines* specifiche per l'Italia, richiamate dal programma principale attraverso macro di Sas, due delle quali riguardano i contributi sociali dei lavoratori dipendenti e degli autonomi e due *routines* sono dedicate al calcolo della tassazione diretta e dell'imposta sulle attività produttive (Irap) dovuta dai lavoratori autonomi. Assieme al calcolo dell'imposta Irpef, il modello stima anche le addizionali regionali e comunali.

Il regime di tassazione dell'anno 2003 previsto in SM2 include il cosiddetto "primo modulo della riforma dell'Irpef" che aveva operato una revisione delle aliquote e degli scaglioni di reddito con l'introduzione della cosiddetta *no tax area*, ossia il meccanismo di sconto della base imponibile per i redditi da lavoro e da pensione. La riforma aveva sostanzialmente modificato le detrazioni da lavoro e aveva previsto una clausola di salvaguardia in modo che, a parità di condizioni, il nuovo regime non risultasse più sfavorevole del precedente.

In Italia, la principale imposta sul reddito (Irpef) è calcolata applicando un'aliquota marginale progressiva a scaglioni di reddito crescenti. In SM2, le diverse fonti di reddito sono raggruppate per tipologia e sottoposte alla tassazione sul reddito (Prospetto 4.1).

I redditi da lavoro (dipendente e autonomo) sono soggetti al prelievo contributivo obbligatorio e per i lavoratori autonomi vi è una doppia tassazione: la tassazione sul reddito e l'imposta regionale sulle attività produttive (Irap). In SM2, la doppia tassazione degli autonomi è trattata considerando l'Irap come una detrazione negativa.

Il reddito da attività finanziarie (depositi bancari, titoli, obbligazioni, fondi comuni, gestioni patrimoniali, azioni, eccetera) è sottoposto ad imposta sostitutiva con un'aliquota costante che varia in base all'ampiezza del capitale, ma indipendentemente dal reddito del contribuente. Tali redditi non contribuiscono all'imponibile totale e in SM2 sono trattati come una detrazione specifica negativa.

Il reddito da capitale reale (affitti, rendite di terreni e fabbricati), a differenza del reddito da attività finanziarie, è sottoposto all'imposta Irpef. I trasferimenti sociali alle famiglie (assegni

⁶ Per una descrizione dettagliata delle caratteristiche del modello SM2 si rimanda al Capitolo 2.

familiari), alcune pensioni di invalidità (erogare dall'Inail), i contributi per la casa sono generalmente esenti da imposta e quindi nel modello non contribuiscono all'imponibile totale. Le borse di studio e le indennità di disoccupazione, al pari del reddito da lavoro dipendente, sono sottoposti invece a tassazione. I pagamenti differiti (Tfr), come le liquidazioni da lavoro e da pensione, e gli arretrati non sono inclusi in SM2.

Prospetto 4.1 - Principali componenti di reddito, imposte e contributi sociali del sistema fiscale italiano inclusi in SM2 - Anno 2003

COMPONENTE DI REDDITO	Contributi sociali (Si)	Imposta	Incluso nell'imponibile	Deduzione specifica per componente di reddito (D _i)	Detrazione specifica per componente di reddito (C _i) (a)
Reddito da lavoro dipendente	Datori di lavoro S ₀ (G ₁) Lavoratori S ₁ (G ₁)	Irpef (b)	X	D ₁ (Y ₁)	
Reddito da lavoro autonomo	S ₂ (G ₂)	Irpef	X	D ₂ (Y ₂)	-f ₂ (H ₂) "Irap" (c)
Pensione		Irpef	X	D ₃ (Y ₃)	
Reddito da capitale non finanziario		Irpef	X		
Reddito da capitale reale		Irpef (a)	X		
Reddito da attività finanziarie		Imposta sostitutiva (aliquota costante K ₆)		H ₆	- K ₆ *H ₆
Borse di studio, indennità di disoccupazione		Irpef	X		
Trasferimenti alle famiglie, pensioni di invalidità, contributi per l'abitazione, altri trasferimenti (d)		Esente		H ₇	
Patrimonio					
Valore degli immobili (rendita catastale)		Ici			-f ₉ (valore)

Fonte: Betti et al. (2010)

(a) Sul totale della rendita catastale e per l'85 per cento del reddito da affitto di terreni e fabbricati.

(b) Sopra un certo limite.

(c) Ulteriore imposta sul reddito da lavoro autonomo (Irap, Imposta regionale sulle attività produttive). f(.) sta per "funzione di".

(d) Alcune pensioni possono essere tassate.

I contributi previdenziali e assistenziali obbligatori non sono sottoposti a tassazione Irpef, a differenza dei contributi previdenziali volontari. Il modello SM2 include i contributi sociali obbligatori per i lavoratori dipendenti, gli apprendisti, e per i lavoratori autonomi e parasubordinati. Per i lavoratori dipendenti e per i datori di lavoro, l'aliquota contributiva varia in base al settore di appartenenza (individuato con la classificazione delle attività economiche Nace), al numero di addetti e alla posizione lavorativa. In SM2 sono calcolati i contributi sociali per il settore pubblico e per le aziende private con una disaggregazione fino a 15 lavoratori, da 15 a 50 e con più di 50 lavoratori e con una qualifica professionale distinta in operai, impiegati e dirigenti. I regimi contributivi agevolati previsti per particolari zone geografiche o per le caratteristiche del lavoratore o dei contratti di lavoro non sono inclusi nel modello per mancanza

di informazioni di base dettagliate. Parimenti non è possibile distinguere tra aziende industriali e aziende artigiane per cui anche per le imprese artigiane sono utilizzate le aliquote dell'industria. Per quanto riguarda i lavoratori autonomi, il modello calcola la contribuzione prevista per i commercianti, gli artigiani, i lavoratori agricoli e i liberi professionisti.

Il modello SM2 include anche i contributi sociali dei lavoratori parasubordinati, i collaboratori coordinati e continuativi (co.co.co.) e i collaboratori a progetto (co.co.pro.), che rappresentano una categoria particolare e intermedia fra il lavoro autonomo e il lavoro dipendente. Da un punto di vista fiscale, i redditi dei co.co.co. sono assimilati ai redditi da lavoro dipendente, tuttavia sono sottoposti a una disciplina specifica con aliquote contributive, da parte dei committenti e degli stessi collaboratori, inferiori rispetto ai lavoratori dipendenti. Secondo il regolamento dell'indagine, il reddito dei lavoratori parasubordinati è tuttavia classificato come reddito da lavoro autonomo. L'indagine Echp non rilevava i lavoratori parasubordinati e in SM2 questa categoria di lavoratori era imputata da fonti esterne.

Le informazioni relative agli oneri deducibili e detraibili, che solitamente non sono raccolte con un'indagine campionaria sui redditi, sono imputate in SM2 mediante il ricorso a fonti esterne. In particolare, le deduzioni dall'imponibile previste dalla normativa fiscale sono calcolate sulla base dei dati contenuti nel Rapporto di Euromod per l'Italia,⁷ ossia come percentuali di spesa derivata dai dati di Contabilità nazionale per scaglioni di reddito.

Allo stesso modo 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 dall'Indagine sui consumi delle famiglie dell'Istat e dall'indagine sui bilanci delle famiglie della Banca d'Italia.

Una volta richiamate le *routines* specifiche per il calcolo della tassazione e dei contributi sociali, la *routine* principale di SM2 produce i dati aggregati dalle variabili necessarie alla procedura di conversione dei redditi netti in lordi e genera due output finali contenenti i microdati delle variabili obiettivo nette e lorde di Eu-Silc e i valori medi delle stesse.

4.2.2 - Le innovazioni del modello SM2-Eu-Silc

La complessità di un modello di microsimulazione deriva sostanzialmente dal livello di dettaglio con cui è inserita la normativa fiscale e dalla necessità di un aggiornamento annuale dei microdati e del regime fiscale e contributivo. La qualità delle stime microsimulate e la loro comparabilità a livello internazionale dipendono tuttavia dalla qualità dei dati di input. I dati rilevati dalla nuova indagine Eu-Silc rappresentano un importante cambiamento per il modello SM2, evidente soprattutto nella fase di costruzione della base dati e delle variabili ausiliarie necessarie alla microsimulazione.

Il modello SM2-Eu-Silc, rispetto al precedente SM2, è caratterizzato oltre che da una nuova base dati soprattutto dalle innovazioni di processo introdotte con le procedure di *record linkage* con i dati amministrativi. In definitiva la versione preliminare del modello SM2-Eu-Silc presenta:

1. miglioramento qualitativo della base dati:

- passaggio da Echp a Eu-Silc;

2. innovazione di processo:

- utilizzo del *record linkage* dei dati campionari e amministrativi.

La maggiore disponibilità di informazioni dell'indagine Eu-Silc rispetto alla precedente Echp consente di migliorare sensibilmente la qualità delle stime microsimulate dei redditi lordi individuali e familiari. Il contributo fornito dall'indagine Eu-Silc risulta particolarmente significativo per il calcolo dei contributi sociali dei lavoratori parasubordinati. L'indagine Echp

⁷ Euromod. *Euromod Country Report Italy (2001 Tax-Benefit System)*. September 2004.

non rilevava tale categoria di lavoratori, per cui in SM2 era stata utilizzata una stima dei parasubordinati rispetto alla popolazione di lavoratori dipendenti derivata da fonti esterne. Allo stesso modo, anche il calcolo dei contributi sociali dei lavoratori dipendenti e autonomi beneficia delle informazioni più dettagliate di Eu-Silc, utili all'individuazione del settore di appartenenza per l'applicazione delle aliquote contributive previste (Prospetto 4.2).

In Eu-Silc sono presenti, inoltre, i dati 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 per l'indennità di maternità e il congedo parentale delle lavoratrici oltre alle 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) assente, nell'indagine Echp. Le informazioni di Eu-Silc sulle liquidazioni da lavoro e da pensione e sugli arretrati da lavoro dipendente permettono, a differenza di SM2, di includere tali informazioni nella base dati e nelle *routines* di conversione di SM2-Eu-Silc.

L'utilizzo congiunto dei dati campionari e amministrativi per la costruzione delle variabili di reddito dell'indagine Eu-Silc⁸ ha consentito di introdurre in SM2-Eu-Silc alcune importanti innovazioni metodologiche.

L'innovazione di processo più rilevante riguarda l'utilizzo dei dati amministrativi. La costruzione di un data set integrato e coerente di tutte le informazioni campionarie e amministrative disponibili ha consentito, nella versione preliminare del modello SM2 Eu-Silc, di utilizzare i dati di fonte fiscale come input del modello di microsimulazione per il calcolo degli oneri deducibili e detraibili piuttosto che stimarli con tecniche di regressione sulla base di fonti esterne.

Le fonti amministrative utilizzate in SM2-Eu-Silc derivano sostanzialmente da cinque archivi: le dichiarazioni dei redditi dei contribuenti, modello 730 e Unico persone fisiche (Upf), le dichiarazioni dei sostituti di imposta, certificazioni Cud e modello 770 (quadro SA dei redditi e delle ritenute per coloro che non devono presentare la dichiarazione) e il Casellario pensionistico elaborato congiuntamente dall'Istat e dall'Inps.

In particolare, per calcolare gli oneri deducibili⁹ e detraibili¹⁰ dei contribuenti sono stati utilizzati il prospetto di liquidazione dei modelli 730 e il quadro RN dei modelli UPF del dichiarante e del coniuge.¹¹

⁸ Si veda il Capitolo 3 sull'uso congiunto delle fonti amministrative e campionarie.

⁹ Deduzione per abitazione principale, crediti di imposta sui dividendi, oneri deducibili (contributi previdenziali e assistenziali, contributi per addetti ai servizi domestici e familiari, erogazioni liberali a favore di istituzioni religiose, spese mediche e di assistenza dei portatori di handicap, assegno periodico corrisposto al coniuge con esclusione della quota di mantenimento dei figli, previdenza complementare e altri oneri).

¹⁰ Detrazioni per oneri e spese Sezione I e III del quadro E (spese sanitarie, interessi per mutui ipotecari, assicurazione sulla vita e contro gli infortuni, spese di istruzione, spese funebri, altri oneri e spese per interventi di recupero del patrimonio edilizio per le quali spetta la detrazione del 41 per cento o del 36 per cento) e altre detrazioni Sez IV del quadro E (detrazioni per gli inquilini di alloggi locati con contratti in regime convenzionale, detrazioni per canone di locazione dei lavoratori dipendenti che trasferiscono la residenza per motivi di lavoro, detrazioni per spese di mantenimento dei cani guida e altre detrazioni).

¹¹ Per i contribuenti in cui erano presenti entrambe le dichiarazioni 730 e Upf è stato utilizzato l'Upf in quanto contiene generalmente informazioni integrative e successive rispetto al 730.

Prospetto 4.2 - Aliquote contributive del settore privato incluse nel modello SM2-Eu-Silc per numero di dipendenti, settore economico e qualifica professionale - Anno 2008

SETTORE ECONOMICO	Qualifica	Numero di dipendenti							
		Fino a 15		Da 16 a 50		Più di 50		Fino a 50	
		Datore di lavoro	Lavoratore	Datore di lavoro	Lavoratore	Datore di lavoro	Lavoratore	Datore di lavoro	Lavoratore
Industria in genere	Operai	30,88	9,19	31,78	9,49	32,08	9,49
	Impiegati	28,66	9,19	29,56	9,49	29,86	9,49
	Dirigenti	26,96	9,19	27,26	9,19	27,26	9,19
Industria edile	Operai	34,98	9,19	35,58	9,49	35,58	9,49
	Impiegati	27,46	9,19	30,06	9,49	30,36	9,49
	Dirigenti	26,96	9,19	26,96	9,19	26,96	9,19
Industria lapidei	Operai	32,68	9,19	33,58	9,49	33,58	9,49
	Impiegati	28,66	9,19	29,56	9,49	29,86	9,49
	Dirigenti	26,96	9,19	27,26	9,19	27,26	9,19
Credito e Assicurazioni (a)	Operai	29,31	9,19						
	Impiegati	26,76	9,19						
	Dirigenti	26,76	9,19						
Commercio ingrosso e dettaglio	Operai					29,88	9,49	28,98	9,19
	Impiegati					29,88	9,49	28,98	9,19
	Dirigenti					26,84	9,19	26,54	9,19
Terziario (a)	Operai	28,98	9,19						
	Impiegati	28,98	9,19						
	Dirigenti	26,54	9,19						
Agricoltura (a)	Impiegati	25,63	8,84						
	Dirigenti	24,13	8,84						

Fonte: Inps

(a) Aliquote non differenziate per numero di dipendenti.

Nella versione preliminare del modello SM2-Eu-Silc, le informazioni fiscali hanno fornito i dati relativi agli oneri e alle spese per consumo in modo da completare il file di input del modello e ottenere la microsimulazione dei redditi lordi dell'indagine 2004. I dati contenuti negli archivi fiscali sono stati poi utilizzati come *benchmark* dei redditi lordi microsimulati consentendo di comparare e validare il modello di microsimulazione.

Il modello preliminare è stato successivamente aggiornato per gli anni di indagine 2005 e 2006, in quanto è stato necessario modificare la parametrizzazione della normativa fiscale a seguito dell'approvazione del cosiddetto "secondo modulo della riforma dell'Irpef" del 2005. La riforma prevedeva, infatti, una ulteriore revisione delle aliquote e degli scaglioni di reddito, trasformava le detrazioni familiari in deduzioni e azzerava le precedenti detrazioni relative ai redditi da lavoro dipendente, autonomo e da pensione.

La fase di sperimentazione dell'integrazione delle microsimulazioni con i dati amministrativi ha fornito buoni risultati e la versione preliminare del modello SM2-Eu-Silc è

stata presentata con un lavoro congiunto con l'Università di Siena alla prima conferenza internazionale delle microsimulazioni a Vienna nel 2007.¹²

Rispetto all'utilizzo dei dati di sola fonte amministrativa, il modello di microsimulazione SM2-Eu-Silc presentava numerosi vantaggi. Consentiva, infatti, la microsimulazione delle imposte e dei contributi sociali per gli individui campionari non presenti negli archivi a causa dei mancati "agganci" con i dati fiscali; permetteva di stimare i contributi sociali a carico dei datori di lavoro, informazione non presente negli archivi disponibili (variabile obiettivo PY030G), e infine forniva una stima della tassazione sul reddito e del prelievo contributivo delle famiglie residenti in Italia (variabile target HY140G).

4.3 - La versione definitiva del modello Sm2-Eu-Silc e l'uso congiunto di record linkage e microsimulazione per la costruzione dei redditi lordi

Lo sviluppo della versione definitiva del modello SM2-Eu-Silc e l'utilizzo congiunto delle tecniche di microsimulazione e di *record linkage* ha consentito all'Istat, con l'indagine Eu-Silc 2007, di produrre per la prima volta in Italia i redditi individuali e familiari sia al netto sia al lordo dell'imposizione fiscale e contributiva.

Il processo di produzione dei redditi lordi Eu-Silc può essere sintetizzato in tre fasi rilevanti:

1. implementazione della versione definitiva del modello di microsimulazione SM2-Eu-Silc;
2. utilizzo dell'integrazione dei dati campionari e amministrativi per le microsimulazioni e per il calcolo delle imposte e ritenute;
3. validazione integrata dei due passi precedenti e costruzione delle variabili obiettivo di reddito lordo individuale e familiare.

Il modello SM2-Eu-Silc, a differenza della versione preliminare, è caratterizzato da un forte impiego degli archivi amministrativi utilizzati non solo per stimare gli oneri deducibili e detraibili ma soprattutto per costruire la base di microdati del modello. I cinque archivi disponibili dell'Agenzia delle entrate e dell'Inps hanno consentito di costruire i redditi netti e lordi per tipologia di fonte, gli oneri deducibili e detraibili, le ritenute e le imposte pagate in termini di Irpef e addizionali regionali e comunali. Per il calcolo dei contributi sociali effettivi, sono state analizzate le sezioni dei dati previdenziali e assistenziali delle certificazioni Cud dell'Inps e dell'Inpdap relative ai contributi trattenuti al lavoratore del settore privato e al lavoratore parasubordinato e il totale dei contributi versati dal lavoratore e dal datore di lavoro del settore pubblico. Per i lavoratori autonomi sono state esaminate attentamente le informazioni contenute nel modello Upf. La parzialità delle informazioni relative soprattutto ai contributi versati dai lavoratori autonomi e la mancanza di informazione sui contributi a carico dei datori di lavoro del settore privato hanno tuttavia limitato l'utilizzo di tali fonti.

Dopo la raccolta di tutte le informazioni disponibili (campionarie e amministrative), il database integrato è stato sottoposto a una complessa procedura di controllo di coerenza e correzione dei dati netti, lordi, delle ritenute e imposte di fonte amministrativa. In particolare, sono state eliminate alcune anomalie tra ritenute, contributi sociali e redditi corrispondenti. Non sono state utilizzate le ritenute o le imposte pagate di fonte fiscale quando non era presente alcun dato di reddito, oppure quando i valori risultavano incoerenti. Sono stati anche verificati e corretti sulla base del prospetto di liquidazione del modello 730 o del quadro RN del modello

¹² Si veda: Betti G., G. Donatiello e V. Verma. *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.

Upf alcuni valori errati, ad esempio delle detrazioni o delle addizionali regionali e comunali. Per rendere assolutamente non rintracciabile l'informazione desunta dagli archivi amministrativi è stata inoltre aggiunta una componente stocastica alle ritenute e imposte utilizzate.

Il database integrato è stato, quindi, utilizzato come file di input del modello di microsimulazione e le stime microsimulate sono state comparate con i dati fiscali disponibili. In presenza di dati amministrativi, le imposte microsimulate sono state poi sostituite con le imposte risultanti dalle dichiarazioni fiscali in quanto corrispondenti al "dato vero", ossia alle imposte effettivamente versate dai contribuenti.

La disponibilità di dati fiscali e dei redditi microsimulati ha consentito una comparazione e validazione incrociata dei risultati molto utile per la costruzione delle variabili obiettivo di reddito lordo individuale e familiare.

In presenza di dati sia campionari sia amministrativi, l'utilizzo del modello si è rivelato prezioso per discriminare alcune incoerenze nei dati amministrativi e nello stesso tempo il modello è stato validato utilizzando come *benchmark* i dati fiscali di reddito lordo per tipologia e soprattutto i valori dell'Irpef e delle addizionali regionali e comunali.

L'utilizzo del modello di microsimulazione ha consentito in particolare di stimare il carico fiscale e contributivo degli individui non presenti negli archivi amministrativi, perché componenti di fatto delle famiglie (ossia individui presenti al momento dell'intervista ma che non erano inclusi nella scheda anagrafica della famiglia al momento dell'estrazione del campione teorico di Eu-silc) o per l'impossibilità di stabilire un abbinamento dei codici fiscali (non è possibile, infatti, agganciare gli individui che riportano nella scheda informazioni incomplete sui codici fiscali e sui dati anagrafici). Il modello SM2-Eu-Silc ha fornito, inoltre, la stima dei contributi sociali posti a carico dei lavoratori e dei datori di lavoro, informazione solo parzialmente presente negli archivi amministrativi attualmente utilizzati e ha permesso di stimare l'imposta regionale sulle attività produttive (Irap) dei lavoratori autonomi.

Oltre al consueto aggiornamento annuale, il modello SM2-Eu-Silc è stato nuovamente modificato nella versione 2008 e 2009, a seguito delle variazioni previste dalla legge finanziaria del 2007 che ha apportato ulteriori correzioni alla precedente riforma fiscale, con cambiamenti delle aliquote e degli scaglioni di reddito, l'azzeramento delle *no tax area* e delle deduzioni da lavoro e per carichi di famiglia e la loro trasformazione in detrazioni d'imposta.

4.4 - La costruzione delle variabili obiettivo di reddito lordo familiare e individuale

L'integrazione delle microsimulazioni con i dati amministrativi e le tecniche di imputazione dei dati mancanti rappresenta un avanzamento metodologico delle procedure di conversione dei redditi netti campionari in lordi.

Il database finale Eu-Silc dei redditi individuali e familiari al lordo dell'imposizione fiscale e contributiva è in definitiva costruito come somma dei redditi netti (campionari o imputati) e delle imposte e ritenute di fonte amministrativa, se disponibili, oppure come somma dei redditi netti e delle imposte microsimulate. Include, inoltre, la stima dei contributi sociali dei lavoratori e dei datori di lavoro (Prospetto 4.3).

Prospetto 4.3 - Variabili target dell'indagine Eu-Silc

REDDITO	Netto	Lordo
Reddito lordo familiare		HY010
Reddito netto familiare disponibile	HY020	
Reddito netto familiare prima dei trasferimenti non pensionistici	HY022	
Reddito netto familiare prima dei trasferimenti sociali	HY023	
Componenti di reddito familiare		
Fitto imputato		HY030G
Reddito da capitale reale	HY040N	HY040G
Indennità familiari	HY050N	HY050G
Altre indennità	HY060N	HY060G
Sussidi per l'abitazione	HY070N	HY070G
Trasferimenti interfamiliari ricevuti	HY080N	HY080G
Reddito da capitale finanziario	HY090N	HY090G
Interessi per mutui e prestiti	HY100N	HY100G
Reddito dei minori di 16 anni	HY110N	HY110G
Imposte sul patrimonio	HY120N	HY120G
Trasferimenti interfamiliari versati	HY130N	HY130G
Saldo d'imposta	HY145N	
Imposte e contributi sociali		HY140G
Componenti di reddito individuale		
Reddito monetario da lavoro dipendente	PY010N	PY010G
Reddito in natura da lavoro dipendente	PY020N	PY020G
Auto aziendale	PY021N	PY021G
Contributi sociali dei datori di lavoro		PY030G
Contributi a pensioni private	PY035N	PY035G
Reddito da lavoro autonomo	PY050N	PY050G
Valori dei beni prodotti per autoconsumo	PY070N	PY070G
Pensioni private	PY080N	PY080G
Indennità di disoccupazione	PY090N	PY090G
Pensioni di anzianità e vecchiaia	PY100N	PY100G
Pensioni di reversibilità	PY110N	PY110G
Pensione di invalidità	PY130N	PY130G
Borse di studio	PY140N	PY140G
Retribuzione mensile lorda dei lavoratori dipendenti		PY200G

Per quanto riguarda i redditi lordi da lavoro dipendente, da pensione e da disoccupazione l'imposta calcolata corrisponde alla ritenuta alla fonte (secondo la definizione Eurostat di reddito lordo). Per il reddito lordo da lavoro autonomo, diversamente da quanto previsto dal regolamento, la tassazione corrisponde all'imposta netta poiché gli acconti e le ritenute sul lavoro autonomo, a differenza dell'imposta anticipata sul reddito da lavoro dipendente e da pensione, possono divergere anche significativamente dall'imposta effettivamente pagata. L'imposta sul lavoro autonomo include anche una parte dell'Irap. L'imposta regionale sulle attività produttive è stimata applicando l'aliquota Irap alla parte della base imponibile del

contribuente autonomo che corrisponde al reddito imponibile Irpef. Essendo calcolata sull'imponibile Irpef, l'Irap è, per questa parte, assimilabile a una imposta diretta.¹³

La stima dei contributi sociali riguarda solo i contributi effettivi, ossia i contributi obbligatori (e quelli volontari se previsti dai contratti collettivi di lavoro) per il conseguimento delle prestazioni previdenziali e assistenziali (malattia, invalidità, malattie professionali o infortuni sul lavoro, vecchiaia, maternità), posti a carico del lavoratore dipendente, autonomo, del lavoratore parasubordinato (collaboratori coordinati e continuativi e dei collaboratori a progetto) e del datore di lavoro. I contributi figurativi non sono, infatti, rilevati nell'indagine Eu-Silc.

Secondo il regolamento Eu-Silc, i contributi sociali dovuti sul reddito da lavoro autonomo includono sia i contributi sociali dei lavoratori autonomi sia i contributi dei lavoratori parasubordinati. I contributi sociali a carico del datore di lavoro sono la somma dei contributi effettivi del datore di lavoro, dell'onere per l'accantonamento al trattamento di fine rapporto (Tfr) e anche dei contributi posti a carico dei committenti dei lavoratori parasubordinati (variabile target PY030G).

I redditi lordi da lavoro dipendente (variabile PY010G) corrispondono alla retribuzione lorda, ossia sono stimati come somma del reddito netto da lavoro dipendente, dell'imposta personale sul reddito e dei contributi sociali posti a carico del lavoratore. Nello stesso modo è calcolato il reddito lordo da lavoro autonomo (variabile PY050G).

L'importo delle variabili target lorde dei redditi esenti da imposta, come ad esempio le indennità familiari (HY050G), i trasferimenti non pensionistici (HY060G), i contributi per la casa (HY070G), i contributi per l'istruzione (PY140G), i beni prodotti per l'autoconsumo (PY070G) e i trasferimenti interfamiliari (variabili HY080G e HY130G), sono per definizione uguali all'importo delle variabili nette.

Nel file di microdati Eu-Silc, rilasciato dall'Istat per finalità di studio e ricerca (file standard), le variabili nazionali di reddito lordo forniscono un'informazione più dettagliata rispetto alle corrispondenti variabili europee. Le variabili nazionali sono ovviamente coerenti con le definizioni e i criteri adottati in sede europea, ma consentono di calcolare distintamente le aliquote fiscali e contributive, che risultano invece aggregate nelle variabili Eurostat.

Nel file standard sono quindi disponibili tre variabili distinte (Csdi, Csa, Csda) per i contributi sociali posti a carico dei lavoratori dipendenti e autonomi e dei datori di lavoro, a fronte di un'unica variabile Eurostat prevista solo per i contributi sociali dei datori di lavoro. È stata, inoltre, costruita una variabile, non presente fra le variabili target Eurostat, relativa all'imposta complessiva sul reddito individuale (Ytax) che corrisponde alla somma delle imposte dirette (Irpef e addizionali regionali e comunali) e dell'imposta sostitutiva sulle attività finanziarie.

Il reddito da lavoro dipendente e autonomo (YdipL e YautL) è quindi calcolato al lordo della tassazione ma al netto dei contributi sociali, in modo che si possa stimare disgiuntamente il peso dell'imposizione fiscale e contributiva sul lavoro.¹⁴ Nel prospetto 4.4 è fornita una comparazione tra le variabili di reddito lordo individuale europee e quelle nazionali che consente di ricostruire i diversi aggregati.

In definitiva nel file nazionale Eu-Silc, il reddito lordo individuale da lavoro, da pensione (YpenG), da trasferimenti non pensionistici (YtdisoG), quali disoccupazione, Cig, borsa lavoro, trattamento di fine rapporto dei lavoratori dipendenti a tempo determinato, corrisponde alla somma del reddito netto e dell'imposta sul reddito per tipologia di fonte.

¹³ Per una descrizione dettagliata delle variabili target si veda Eurostat. Eu-Silc 065 (2008 operation). *Description of Target variables: Cross-sectional and Longitudinal*. Luxembourg 2008 operation.

¹⁴ Si veda: Donatiello G. e P. Consolini. *La distribuzione del carico fiscale e contributivo in Italia. Anno 2007*. Roma: Istat, 27 aprile 2010. (Approfondimenti e Nota metodologica).

Prospetto 4.4 - Raccordo tra variabili individuali di reddito lordo europee e italiane

VARIABILI	Nome	Raccordo
Variabili Eurostat		
Retribuzione lorda da lavoro dipendente	PY010G	
Retribuzione lorda non monetaria da lavoro dipendente	PY020G	
Reddito lordo da lavoro autonomo	PY050G	
Reddito lordo da disoccupazione	PY090G	
Reddito lordo da pensione di anzianità e vecchiaia	PY100G	
Reddito lordo da pensione di reversibilità	PY110G	
Reddito lordo da pensione di invalidità	PY130G	
Contributi sociali a carico del datore di lavoro	PY030G	
Variabili nazionali		
Contributi sociali a carico del lavoratore dipendente	Csdi	
Contributi sociali a carico del lavoratore autonomo	Csa	
Contributi sociali a carico del datore di lavoro	Csda	= PY030G
Imposta personale complessiva sul reddito	Ytax	
Reddito lordo da lavoro dipendente, al netto dei contributi sociali	YdipL	YdipL + Csdi = PY010G+PY020G
Reddito lordo da lavoro autonomo, al netto dei contributi sociali	YautL	YautL + Csa = PY050G
Reddito lordo da disoccupazione	YtdisoG	= PY090G
Reddito lordo da pensione	YpenG	= PY100G+PY110G+PY130G

Capitolo 5 - La valutazione della precisione delle stime

5.1 - Introduzione

Lo scopo principale delle indagini campionarie eseguite dall'Istat è quello di stimare e diffondere parametri di popolazioni e sub-popolazioni di interesse.

È consuetudine accompagnare la diffusione delle stime con un'adeguata documentazione contenente informazioni e indicatori per: a) mettere a conoscenza l'utente delle metodologie utilizzate nella rilevazione e nel processo di elaborazione dei dati; b) consentire la valutazione delle stime prodotte ai fini di una loro corretta utilizzazione.

Con riferimento a quest'ultimo punto, gli errori di campionamento rivestono una particolare importanza.

La loro misura è infatti essenziale per una adeguata utilizzazione delle stime prodotte, tenendo presente che esistono diverse categorie di utenti che si servono degli errori campionari per scopi diversificati, dalla valutazione dell'attendibilità delle stime fino alla progettazione di nuove indagini. Per questo motivo, l'Istat generalmente diffonde le principali stime insieme a indicatori quali l'errore relativo di campionamento, l'errore assoluto di campionamento, l'intervallo di confidenza.

A questo scopo, in Istat sono stati sviluppati alcuni software (tra questi Genesees, Generalised Software for Sampling Errors in Surveys), basati sulla metodologia standard¹ di stima della varianza campionaria, che consentono il calcolo della varianza campionaria (e gli indicatori a essa associati) delle stime più diffuse dall'Istituto quali frequenze, medie, totali.

Si ricorda che lo sviluppo di software autonomo e generalizzato è una realtà praticata in numerosi istituti nazionali di statistica; esclusivamente a titolo di esempio, citiamo il software Ges (Generalised Estimation System) di Statistics Canada e il software Poulpe (Programme optimal et universal pour la livraison de la précision des enquêtes) dell'Insee.

Oltre alle stime consuete sopra citate, negli ultimi anni si è spostata sempre di più l'attenzione verso la produzione di statistiche complesse – ossia espresse da funzioni non lineari – per le quali non si può applicare direttamente la metodologia standard di stima della varianza campionaria.

Un primo caso rilevante si è presentato per la stima di statistiche sulla povertà relativa a livello regionale a partire dall'indagine sui consumi delle famiglie. La valutazione della varianza campionaria di queste stime è stato un interessante obiettivo metodologico affrontato dall'Istat nel periodo 2002-2003.

A seguire, l'indagine europea su reddito e condizioni di vita (Eu-Silc) ha proposto un ulteriore e importante caso di statistiche complesse. Infatti, l'indagine in questione stima non solo i redditi e le condizioni di vita, ma anche una serie di misure complesse di povertà e disuguaglianza,² dette indicatori europei di povertà e disuguaglianza o indicatori di Laeken, che sono trasmesse ad Eurostat insieme ai corrispondenti errori di campionamento.

Il capitolo è a cura di Claudia Rinaldelli

¹ Per metodologia standard si intende metodologia notoriamente diffusa e applicata; la letteratura sulla teoria del campionamento da popolazioni finite, fornisce, per i più importanti disegni di campionamento, le formule per il calcolo della varianza degli stimatori frequentemente utilizzati nelle indagini.

² La linea di povertà relativa, l'incidenza di povertà relativa, l'indice di disuguaglianza della distribuzione del reddito (indice di Gini), il *gender pay gap*, l'intensità di povertà relativa, il rapporto interquintile.

Anche per queste misure complesse non è possibile applicare direttamente la metodologia standard di stima della varianza campionaria.

Gli studi e le sperimentazioni eseguiti in Istat hanno permesso di individuare e proporre una soluzione metodologica per stimare la varianza campionaria degli indicatori di Laeken mediante l'uso di variabili linearizzate e la tecnica di ricampionamento delle replicazioni bilanciate ripetute (Brr). Al fine di rendere applicabili le soluzioni metodologiche sopra citate, è stato necessario sviluppare nuove procedure informatiche per implementare la tecnica di ricampionamento del Brr e per gestire le variabili linearizzate individuate.

Dato il maggiore onere computazionale della tecnica Brr rispetto alla linearizzazione, la coerenza dei risultati ottenuti applicando le due differenti metodologie, la raccomandazione di Eurostat sull'argomento, l'Istat ha adottato ufficialmente l'approccio della linearizzazione per il calcolo degli errori campionari degli indicatori di Laeken che sono quindi calcolati correntemente dal 2005 con la procedura software sviluppata per l'approccio di linearizzazione.

Sintetizzando quanto sopra esposto, in linea generale, l'indagine Eu-Silc fornisce le seguenti stime:

1. medie dei redditi, ad esempio il reddito medio familiare netto;
2. frequenze sulle condizioni di vita, ad esempio il numero di famiglie per caratteristiche e titolo di godimento dell'abitazione;
3. gli indicatori di povertà relativa e disuguaglianza sociale (indicatori di Laeken).

La precisione degli indicatori di povertà relativa e disuguaglianza sociale è valutata impiegando metodologie più complesse sopra citate mentre gli errori di campionamento delle medie dei redditi e delle frequenze sono calcolati mediante la metodologia standard di stima della varianza campionaria.

Poiché il presente volume è dedicato alla metodologia di stima dei redditi lordi, in questo capitolo si farà riferimento alla stima della varianza campionaria degli aggregati di reddito mentre per la descrizione degli studi sperimentali e delle procedure utilizzate per stimare la varianza campionaria degli indicatori di Laeken, si rimanda al volume *L'indagine europea sui redditi e le condizioni di vita delle famiglie (Eu-Silc)*. Istat, 2008 (Metodi e Norme n. 37).

5.2 - La valutazione della precisione delle stime dei redditi

Come descritto nel precedente paragrafo, i principali indicatori 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 , ad esempio l'ammontare del reddito familiare netto nella regione Lazio, la stima dell'errore assoluto di campionamento è espressa da:

$$\hat{\sigma}(\hat{Y}_d) = \sqrt{\hat{V}\hat{a}r(\hat{Y}_d)} \quad (1)$$

e l'errore di campionamento relativo è espresso da:

$$\hat{\varepsilon}(\hat{Y}_d) = \frac{\hat{\sigma}(\hat{Y}_d)}{\hat{Y}_d} \quad (2)$$

Nella (1) $V\hat{a}r(\hat{Y}_d)$ denota la stima della varianza campionaria di \hat{Y}_d .

L'errore di campionamento assoluto permette inoltre di costruire un intervallo di confidenza, nel quale con una certa probabilità si trova il parametro oggetto di stima:

$$\Pr\{\hat{Y}_d - k\hat{\sigma}(\hat{Y}_d) \leq Y_d \leq \hat{Y}_d + k\hat{\sigma}(\hat{Y}_d)\} = P \quad (3)$$

Nella (3) il valore di k dipende dal valore fissato per la probabilità P ; ad esempio, per $P=0,95$ si ha $k=1,96$.

Il calcolo della varianza campionaria $V\hat{a}r(\hat{Y}_d)$ e dei conseguenti indicatori è eseguito mediante la procedura generalizzata correntemente utilizzata in Istat (software Genesees).

Prospetto 5.1 - Reddito familiare netto (esclusi i fitti imputati), errore di campionamento relativo percentuale e assoluto, intervallo di confidenza per regione e ripartizione geografica - Anno 2004

REGIONI	Reddito familiare netto (media in euro)	Errore di campionamento		Intervallo di confidenza (p=0,95)	
		Relativo %	Assoluto		
Piemonte	27.522	2,6	719	26.113	28.931
Valle d'Aosta/Vallée d'Aoste	27.474	3,7	1.01	25.495	29.453
Lombardia	30.095	1,8	537	29.042	31.148
Liguria	24.355	2,4	575	23.228	25.482
Trentino-Alto Adige/Südtirol	29.353	2,6	757	27.869	30.837
<i>Bolzano/Bozen</i>	<i>31.116</i>	<i>3,8</i>	<i>1.169</i>	<i>28.825</i>	<i>33.407</i>
<i>Trento</i>	<i>27.774</i>	<i>3,5</i>	<i>982</i>	<i>25.850</i>	<i>29.698</i>
Veneto	28.044	1,7	488	27.087	29.001
Friuli-Venezia Giulia	28.025	2,6	734	26.586	29.464
Emilia-Romagna	29.895	1,6	477	28.960	30.830
Toscana	29.190	1,6	465	28.279	30.101
Umbria	27.340	2,0	553	26.257	28.423
Marche	27.908	2,4	656	26.622	29.194
Lazio	28.157	3,1	878	26.437	29.877
Abruzzo	25.303	2,7	679	23.973	26.633
Molise	23.603	4,1	961	21.719	25.487
Campania	23.011	2,5	575	21.883	24.139
Puglia	22.382	3,2	714	20.982	23.782
Basilicata	20.778	3,1	647	19.510	22.046
Calabria	19.835	3,5	696	18.472	21.198
Sicilia	19.847	2,5	494	18.879	20.815
Sardegna	24.984	2,6	656	23.698	26.270
Nord-ovest	28.682	1,3	384	27.929	29.435
Nord-est	28.879	1,0	293	28.305	29.453
Centro	28.405	1,6	447	27.529	29.281
Sud	22.510	1,5	333	21.857	23.163
Isole	21.127	1,9	405	20.333	21.921
ITALIA	26.521	0,7	173	26.181	26.861

Il prospetto 5.1 costituisce un classico esempio di modalità di diffusione delle stime e dei corrispondenti errori di campionamento, adottato dall'Istat. In particolare, il prospetto riporta l'errore campionario relativo, assoluto e l'intervallo di confidenza del reddito medio familiare netto (esclusi i fitti imputati) per regione, ripartizione geografica e Italia per la rilevazione Eu-Silc dell'anno 2004.

5.3 - La documentazione sulla qualità secondo i regolamenti europei

È stata finora descritta la modalità di documentazione della precisione delle stime adottata dall'Istat. Tuttavia occorre ricordare che, essendo l'indagine Eu-Silc disciplinata da normativa europea, i singoli istituti nazionali di statistica predispongono per Eurostat una dettagliata documentazione sulla qualità delle stime prodotte.

Il regolamento europeo n. 1177/2003 noto come "Regolamento quadro (Framework Regulation)" regola aspetti quali gli scopi dell'indagine, i riferimenti temporali (data di inizio della rilevazione, periodo di riferimento per la rilevazione del reddito), 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*), mentre il regolamento attuativo n. 28/2004³ disciplina la produzione e i contenuti delle relazioni sulla qualità (*quality report*) e sull'attuazione dell'indagine Eu-Silc.

Il presente paragrafo è dedicato alla descrizione del suddetto regolamento che 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 a Eurostat.

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 5.2 riassume la tipologia dei *quality report* e i relativi tempi di consegna con riferimento alla rilevazione eseguita nell'anno N.

Prospetto 5.2 - 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).

³ Legge 5/42, pubblicata sull' Official Journal of the European Union del 9 gennaio.2004.

Prospetto 5.3 - Media ed errore campionario assoluto per componente di reddito, dimensione della famiglia, classe di età e sesso - Anno 2004

COMPONENTE DI REDDITO	Media (in euro)	Errore campionario assoluto (in euro)
Reddito netto familiare disponibile	24.950,77	147,21
Reddito netto familiare prima dei trasferimenti non pensionistici	23.925,83	148,34
Reddito netto familiare prima dei trasferimenti sociali	16.744,65	147,35
Componenti di reddito netto familiare		
Reddito da capitale reale	4.812,87	213,69
Indennità familiari	1.142,84	26,63
Altre indennità	9.023,80	1.005,25
Sussidi per l'abitazione	506,57	58,26
Trasferimenti interfamiliari ricevuti	4.677,23	239,94
Reddito da capitale finanziario	936,23	39,32
Interessi per mutui e prestiti	2.456,22	72,70
Reddito dei minori di 16 anni	2.078,20	316,72
Imposte sul patrimonio	358,58	4,66
Trasferimenti interfamiliari versati	3.417,16	164,02
Saldo d'imposta	245,79	25,59
Componenti di reddito netto individuale		
Reddito monetario da lavoro dipendente	14.263,78	129,80
Reddito in natura da lavoro dipendente	1.447,10	97,10
Contributi a pensioni private	1.743,78	51,96
Reddito da lavoro autonomo	15.788,86	373,63
Pensioni private	5.946,06	638,01
Indennità di disoccupazione	3.109,02	171,00
Pensioni di anzianità e vecchiaia	11.612,40	83,61
Pensioni di reversibilità	6.437,21	302,55
Pensione di invalidità	5.873,61	206,75
Borse di studio	2.969,11	273,75
Retribuzione mensile lorda dei lavoratori dipendenti	1.608,46	14,64
Per dimensione della famiglia		
1 componente	14.048,44	243,04
2 componenti	16.020,85	237,11
3 componenti	16.093,13	283,24
4 componenti e più	13.607,86	216,36
Per classe di età		
<25 anni	13.251,73	133,84
25-34 anni	15.410,91	155,65
35-44 anni	14.959,91	167,55
45-54 anni	16.016,93	189,00
55-64 anni	16.843,09	183,59
65+ anni	14.219,65	113,76
Per sesso		
Maschi	15.235,21	94,46
Femmine	14.420,52	93,73

Il regolamento n. 28/2004 stabilisce in maniera dettagliata il contenuto dei *quality report*; poiché la precisione delle stime dei redditi è documentata obbligatoriamente nei *quality report* finali, si fa particolare riferimento a questi.

I *quality report* finali devono contenere:

- 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;
- 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);
- la descrizione della procedura utilizzata per calcolare il fitto figurativo se diversa dal metodo proposto da Eurostat.

Il prospetto 5.3 costituisce un esempio di modalità di diffusione delle stime e dei corrispondenti errori di campionamento predisposto dall'Istat per Eurostat nell'ambito dei *quality report* finali.

In particolare, il prospetto riporta, secondo quanto stabilito dal regolamento attuativo n. 28/2004, per ogni reddito medio elencato il corrispondente errore campionario assoluto con riferimento alla componente trasversale dell'indagine Eu-Silc dell'anno 2004.

Il regolamento attuativo n. 28/2004 stabilisce inoltre 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.

Bibliografia

Atkinson A.B. "EUROMOD and the Development of EU Social Policy". DIW Berlin, Discussion Papers 467, 2005.

Atkinson A.B. e H. Sutherland. "Tax Benefits Models". STICERD Occasional Paper, n. 10, London School of Economics, London, 1988.

Baldini M. *MAPP98: un Modello di Analisi delle Politiche Pubbliche*. Materiali di Discussione, CAPP, 2001.

Ballini F., G. Betti, G. Çalik, Z. Daşkiran e M. Karakaş. "The Siena Micro-Simulation Model (SM2) for the tax-benefit system in Turkey". Working Paper n. 81, Dipartimento di Metodi Quantitativi, Università di Siena, 2009.

Betti G., G. Donatiello e V. Verma. "The Siena Micro Simulation Model (SM2) for net-gross conversion of EU-SILC income variables". *International Journal of Microsimulation*, 3(2), 2010.

Betti G., G. Donatiello e V. Verma. *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.

Betti G. e V. Verma. "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., V. Verma, F. Ballini, M. Natilli e S. Galgani. "Statistical Imputation in Conjunction with Micro-Simulation of Income Data". *Rivista Italiana di Economia, Demografia e Statistica*, 58(3), 35-43, 2003.

Binder D.A. e M.S. Kovacevic. "Estimating some measures of income inequality from survey data: an application of the estimating equation approach". *Survey Methodology*, 21, pp.137-145, 1994.

Bourguignon F. e A. Spadaro. "Microsimulation as a Tool for Evaluating Redistribution Policies". Working Paper n. 2005 – 02, PSE-Paris-Jourdan Sciences Economiques, Paris 2005.

Bravi R., D. Moretti e C. Rinaldelli. *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 Recommendatio*. Ottawa, 2001.

Consolini P. *Le prestazioni sociali monetarie non pensionistiche: aspetti istituzionali e classificazioni statistiche*. Roma: Istat, 2000. (Documenti Istat, n. 2).

Consolini P. e R. De Carli. *Le prestazioni sociali monetarie non pensionistiche: unità d'analisi, fonti e rappresentazione statistica dei dati*. Roma: Istat, 2002. (Documenti Istat, n. 1).

Consolini P., M. Di Marco, R. Ricci e S. Vitaletti. "Administrative and Survey Microdata on Self-Employment: the Italian Experience with the EU-SILC project". Iariw 29th general conference Joensuu, Finland, 20-26 August 2006.

Consolini P. *Integrazione di dati campionari Eu-Silc con dati di fonte amministrativa*. Roma: Istat, 2009. (Metodi e Norme, n. 38).

Coccia G., N. Pannuzi e C. Rinaldelli. *Poor and non poor households: the estimation from sample surveys*. Book of Short papers, CLADAG 2005. MUP editore, pp.69-72, June 2005.

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*.

Deville J.C. "Variance Estimation for complex statistics and estimators: linearization and residual techniques". *Survey Methodology*, vol. 25 n.2, pp.193-203, 1999.

De Vitiis C., L. Di Consiglio, S. Falorsi, C. Pauselli e C. Rinaldelli. "La valutazione dell'errore di campionamento delle stime di povertà relativa". Convegno Istat "Povertà Regionale ed Esclusione Sociale", Roma 17 dicembre 2003.

Di Biase R., M. Di Marco, F. Di Nicola e G. Proto. *ITAXMOD: A Microsimulation Model of the Italian Personal Income Tax and of Social Security Contributions*. Ispe, 1995 (Documenti di Lavoro, n. 16).

Di Nicola F. e V. Polin. "Gli studi di settore: fondamenti, risultati e prospettive". In. *Rapporto Finanza pubblica e redistribuzione*. Isae, Novembre 2006.

Donatiello G. e P. Consolini. *La distribuzione del carico fiscale e contributivo in Italia. Anno 2007*. Istat, 27 aprile 2010. (Approfondimenti e nota metodologica).

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*, 2003.

Euromod. "Euromod: an integrated European benefits-tax model". Final Report, edited by Sutherland, H., Euromod Working Paper n. EM9/01, University of Essex: ISER 2001.

Euromod. "Euromod Country Report Italy (2001 Tax-Benefit System)". September 2004 <http://www.iser.essex.ac.uk/files/msu/emod/documentation/countries/italy/cr01it0904.pdf>

Eurostat. "Technical document on intermediate and final quality reports, Statistics on Income and Living Conditions (EU-SILC)". Doc. EU-SILC 132/04, 2004.

Eurostat. “Commission Regulation concerning EU-SILC as regards the list of primary target variable”. EU-SILC 107/02, Luxembourg, 2002.

Eurostat. Doc.Pan n. 167/2003-12 “Echp Udb, Construction of variables. From ECHP questions to UDB variables”. Luxembourg: Eurostat, 2003.

Eurostat. Doc.Pan n. 168/2003-12 “Echp Udb Manual. European Community Household Panel, Longitudinal Users' Database. Waves 1 to 8 - Survey years 1994 to 2001”. Luxembourg: Eurostat, 2003.

Eurostat. Doc.Pan n. 164/2002-12 “Imputation of Income in the Echp”. Luxembourg: Eurostat, 2002.

Eurostat. Eu-Silc 065 (2008 operation). *Description of Target variables: Cross-sectional and Longitudinal*. 2008 operation. Luxembourg: Eurostat, 2008.

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”. Preparato da V. Verma, G. Betti e co-ricercatori. EU-SILC 133/04, Luxembourg, 2004.

Falorsi S. e C. Rinaldelli. *Un software generalizzato per il calcolo delle stime e degli errori di campionamento*. Statistica Applicata, 10, 2, pp. 217-234, 1998.

Fellegi I.P. e A.B. Sunter. “A theory of record linkage”. *Journal of the American Statistical Association*, vol. 64, pp.1183-1210, 1969.

Guerrucci D. *Confronto tra i sistemi di classificazione dell'assistenza sociale, contributo al Gruppo di lavoro Istat-USCI*. Roma, 1996.

Figari F., H. Levy e H. Sutherland. “Using the EU-SILC for policy simulation: prospects, some limitations and some suggestions”. In Eurostat. *Comparative EU statistics on Income and Living Conditions: Issues and Challenges - Proceedings of the EU-SILC conference*, Helsinki, 6-8 November 2006, 345-373, 2007.

Harding A. “Dynamic microsimulation models: problems and prospects”. Discussion Paper WSP/48, London, STICERD, London School of Economics, 1990.

Herzog T.N., F.J. Scheuren. e W.E. Winkler. *Data quality and record linkage techniques*. New York: Springer ed 2007.

Immervoll H. e C. O'Donoghue. “Imputation of Gross Amounts from Net Incomes in Household Surveys. An Application using EUROMOD”. EUROMOD Working Paper n. EM1/01, June 2001.

Immervoll H. e C. O'Donoghue. “Towards a Multi-Purpose Framework for Tax-Benefit Microsimulation”. EUROMOD Working Paper n. EM2/01, December 2001.

Istat. *GENESEES v3.0 Funzione stime ed errori. Manuale Utente e Aspetti Metodologici*. 2005.

Istat. *La misurazione delle tipologie familiari nelle indagini di popolazione*. Roma: Istat, 2010. (Metodi e Norme, n. 46).

Istat. *L'indagine europea sui redditi e le condizioni di vita delle famiglie (Eu-Silc)*. Roma: Istat, 2008. (Metodi e Norme, n. 37).

Kovacevic M.S. e D.A. Binder. "Variance estimation for measures of income inequality and polarization – the estimating equations approach". *Journal Official Statistics*, Vol. 13 n. 1, pp. 41-58, 1997.

Lugaresi S. *ITAXMOD*. Ispe, 1989.

Mancini A. *Previdenza integrativa e fondi pensione: linee per lo sviluppo dell'informazione statistica*. Roma: Istat, 2001. (Documenti Istat, n. 7).

Manzari A. *Valutazione comparativa di alcuni metodi di imputazione singola delle mancate risposte parziali per dati quantitativi*. Roma: Istat, 2004. (Documenti Istat, n. 18).

Martini A. e U. Trivellato. *The Role of Survey Data in Microsimulation Models for Social Policy Analysis*. LABOUR, 11(1), 83 – 112, 1997.

McCarthy P.J."Pseudoreplication: Further evaluation and application of the balanced Half-Sample Technique". *Vital and Health Statistics*, Series 2 n. 31. Washington, D.C: National Center for Health Statistics, Public Health Service, 1969.

Mitton L., H. Sutherland e M. Weeks (Eds). *Microsimulation Modelling for Policy Analysis. Challenges and Innovations*. United Kingdom: Cambridge University Press, 2000.

Moore J.C., L.L. Stinson e E.J. Welniak."Income Measurement Error in Surveys: A Review". *Journal of Official Statistics*, Vol. 16, Issue 4, 2000.

Moretti D., C. Pauselli e C. Rinaldelli. "Il 'Relative Poverty Gap' e il 'Quintile Share Ratio': la stima dell'errore di campionamento secondo le Replicazioni Bilanciate Ripetute". *Statistica Applicata*, 17, 2, pp.179-190, 2005.

Moretti D., C. Pauselli e C.Rinaldelli. "La stima della varianza campionaria di indicatori complessi di povertà e disuguaglianza". *Statistica Applicata*, 17, 4, pp. 529-550, 2005.

Moretti D.e C. Rinaldelli. "Variance estimation for relative poverty measures and inequality indicators from complex sample surveys". Atti del Quarto Convegno S.Co.2005, September 2005, pp. 67-72. Cluep editrice Padova. 2005.

National Statistical Service of Greece. "Statistics on income and living conditions model: net gross income conversion". *Report to Eurostat*. Pireus, 2007.

Nazzaro O. *Statistiche sulle pensioni e sui percettori di prestazioni pensionistiche: progetto di realizzazione delle pubblicazioni per la collana Informazioni*. Roma: Istat, 1999. (Documenti Istat, n. 16).

Newcombe H.B. *Handbook of Record Linkage: Methods for Health and Statistical Studies, Administration, and Business*. Oxford (UK): Oxford University Press, 1988.

Orcutt G. H. "A New Type of Socio-Economic System". *Review of Economics and Statistics*, 80: 1081-1100, 1957.

Pauselli C. e C. Rinaldelli. "La valutazione dell'errore di campionamento delle stime di povertà relativa secondo la tecnica Replicazioni Bilanciate Ripetute". *Rivista di Statistica Ufficiale*, 2/2003, pp.7-22, 2003.

Proto G. "Il modello di microsimulazione MASTRICT: struttura e risultati". *Rivista di Statistica Ufficiale*, n. 3, settembre-dicembre, pp. 25-55, 1999.

Raughunatan E. et al. "A Multivariate technique for multiply imputing missing values using a sequence of regression model". *Survey methodology*, vol. 27, n. 1, pp. 85-95, 2001.

Rizzi D. "Un modello integrato per lo studio degli effetti delle politiche economiche sulla povertà". *Politica Economica*, n. 2, 1990.

Rossi N. *La crescita ineguale, Primo rapporto sulla distribuzione e redistribuzione*. Bologna: Il Mulino, 1993.

Rao J.N.K. e C.F.J. Wu. "Resampling inference with complex survey data". *American Statistical Association*, 83, 401, pp. 231-241, 1988.

Rao N.K. e J. Shao J. "On balanced half-sample variance estimation in stratified random sampling". *Journal of American Statistical Association*, vol. 91, n. 433, 1996.

Rinaldelli C. "Statistiche complesse e software". *Statistica & Società*. Anno III, n. 2, 01.2005, pp.27-29, 2005.

Rinaldelli C. "Experiences of variance estimation for relative poverty measures and inequality indicators in official sample surveys". *Proceedings in Computational Statistics*, COMPSTAT 2006, Italy, August 2006, pp. 1465-1472, 2006.

Rodrigues C. F. "Income in EU-SILC – Net/Gross Techniques for Building and Using EU-SILC Databases". In *Eurostat and Statistics Finland "Comparative EU Statistics on Income and Living Conditions: Issues and Challenges. Proceedings of the EU-SILC conference, Helsinki, 6-8 November 2006"*. 2007.

Särndal C.E., B. Swensson e J. Wretman. *Model assisted survey sampling*. New York: Springer-Verlag, 1992.

Scanu M. *Metodi statistici per il record linkage*. Roma: Istat, 2003. (Metodi e Norme, n. 16).

Sirken M.G. et al. *Cognition and Survey Research*. John Wiley & Sons, 1999.

Sutherland H., F. Figari, O. Lelkes, H. Levy, C. Lietz, D. Mantovani, e P. Alari. “Improving the Capacity and Usability of Euromod – Final Report”. EUROMOD Working Paper n. EM4/08, 2008.

Sutherland H. “Static microsimulation models in Europe: a survey. University of Cambridge, Microsimulation Unit”. Discussion Paper, 1995.

van der Laan P. “Integrating administrative registers and household surveys”. *Netherlands Official Statistics*, Vol. 15, Summer 2000.

Verma V. e G. Betti. *Sampling error for measures of inequality and povert. Book of Short papers*. CLADAG 2005, MUP editore, pp.175-178, June 2005.

Verma V., G. Betti, F. Ballini, M. Natilli e S. Galgani. “Personal income in the gross and net forms: applications of the Siena Micro-Simulation Model (SM2)”. Working Paper n. 54, Università degli Studi di Padova: Dipartimento di Scienze Statistiche, 2003.

Vitaletti S. “Correzioni e imputazioni delle informazioni sui redditi”. In *L’indagine europea sui redditi e le condizioni di vita delle famiglie (Eu-Silc)*, par. 4.4. Roma: Istat, 2008. (Metodi e Norme, n. 37).

Wallgren A. e B. Wallgren. *Register-based Statistics. Administrative Data for Statistical Purposes*. John Wiley & Sons, 2007.

Wolter K.M. *Introduction to variance estimation*. New York: Spriger-Verlag, 1985.

Zaidi A. e K. Rake. “Dynamic Microsimulation Models: A Review and Some Lessons for SAGE”. SAGE Discussion Paper n. 2, SAGEDP/02, 2001.

Metodi e Norme - Nuova serie - Volumi pubblicati

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 dell'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)*

Anno 2009

38. *Integrazione di dati campionari Eu-Silc con dati di fonte amministrativa*
39. *La misura della povertà assoluta*
40. *Classificazione delle attività economiche. Ateco 2007 - Derivata dalla Nace Rev. 2*
41. *L'ambiente di codifica automatica dell'Ateco 2007 - Esperienze effettuate e prospettive*
42. *L'indagine sulle professioni - Anno 2007 - Contenuti, Metodologia e Organizzazione*

Anno 2010

43. *Istruzioni integrative per l'applicazione dell'Icd-10 nella codifica delle cause di morte - Seconda edizione con aggiornamenti Oms fino al 2009* ○
44. *Regional Gva Inventory - Italy (in corso di stampa)*
45. *Guida all'utilizzo della nuova ateco versione 2007*
46. *La misurazione delle tipologie familiari nelle indagini di popolazione*
47. *Il test cognitivo per migliorare la qualità del dato - Alcune esperienze di utilizzo nelle indagini sociali*
48. *Guida alla vigilanza anagrafica*

Anno 2011

49. *La metodologia di stima dei redditi lordi nell'indagine Eu-Silc - Indagine europea sui redditi e le condizioni di vita delle famiglie*

I Centri di informazione statistica

PIÙ INFORMAZIONI. PIÙ VICINE A VOI.

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 "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 051/6566111-152
Fax 051/6566185-182

BOLZANO Via Canonico M. 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 Lungarno C. Colombo, 54
Telefono 055/6237711
Fax 055/6237735

GENOVA Via San Vincenzo, 4
Telefono 010/5849718
Fax 010/5849735

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 Pretoria, 342
Telefono 0971/377211
Fax 0971/36866

ROMA Via Cesare Balbo, 11/a
Telefono 06/46733102-6
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 - secondo piano - Telefono 06/4673.2380 Fax 06/4673.2617

<https://contact.istat.it>

Orario: da lunedì a giovedì 9.00 - 16.00 venerdì 9.00 - 14.00

La metodologia di stima dei redditi lordi nell'indagine Eu-Silc

Indagine europea sui redditi e le condizioni di vita delle famiglie

Il volume presenta la metodologia di stima dei redditi lordi dell'indagine Eu-Silc attraverso capitoli dedicati alle innovazioni metodologiche e alle fasi più rilevanti di costruzione dei redditi individuali e familiari al lordo dell'imposizione fiscale e contributiva. Per la produzione dei redditi lordi Eu-Silc, prevista in Italia dall'anno di indagine 2007, l'Istat ha deciso di utilizzare il modello di microsimulazione SM2 dell'Università di Siena e di sperimentare una metodologia più complessa che rappresenta un'importante innovazione nelle procedure di conversione dei redditi netti campionari. Sulla base dell'esperienza Eu-Silc di integrazione dei dati campionari e amministrativi è stata sperimentata l'integrazione delle informazioni tributarie nel modello di microsimulazione. L'utilizzo congiunto e innovativo delle microsimulazioni e delle fonti amministrative ha consentito di potenziare i vantaggi ottenibili dall'impiego esclusivo degli archivi fiscali e delle tecniche di microsimulazione e di migliorare la qualità delle stime prodotte.

The methodology of construction of Eu-Silc gross income variables

European survey on income and living conditions

This book presents the methodology used to estimate Eu-Silc gross income variables. Chapters are devoted to the methodological improvements and the most important stages in reconstructing individual and household incomes before taxes and social insurance contributions. According to Eu Regulation, in Italy the construction of gross income variables became mandatory starting from survey year 2007 and ISTAT opted to start with the SM2 micro-simulation model from the University of Siena and experiment with a more complex methodology, which reflects a major innovation in the procedures used to convert a sample of net incomes. The Eu-Silc experience with the integration of survey and administrative data is what provides the basis for experimenting with a conjunction of fiscal data and the micro-simulation model. This innovative strategy of combining micro-simulations and administrative sources succeeded in improving the quality of estimates that were produced, thus expanding on the benefits of using fiscal records or micro-simulation techniques in isolation.

ISBN 978-88-458-1683-3



1M012011049000000

€ 12,00