



SISTEMA STATISTICO NAZIONALE
ISTITUTO NAZIONALE DI STATISTICA

**La rilevazione sulle forze di lavoro:
contenuti, metodologie,
organizzazione**



I settori

AMBIENTE E TERRITORIO		<i>Ambiente, territorio, climatologia</i>
POPOLAZIONE		<i>Popolazione, matrimoni, nascite, decessi, flussi migratori</i>
SANITÀ E PREVIDENZA		<i>Sanità, cause di morte, assistenza, previdenza sociale</i>
CULTURA		<i>Istruzione, cultura, elezioni, musei e istituzioni similari</i>
FAMIGLIA E SOCIETÀ		<i>Comportamenti delle famiglie (salute, letture, consumi, etc.)</i>
PUBBLICA AMMINISTRAZIONE		<i>Amministrazioni pubbliche, conti delle amministrazioni locali</i>
GIUSTIZIA		<i>Giustizia civile e penale, criminalità</i>
CONTI NAZIONALI		<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		<i>Agricoltura, zootecnia, foreste, caccia e pesca</i>
INDUSTRIA		<i>Industria in senso stretto, attività edilizia, opere pubbliche</i>
SERVIZI		<i>Commercio, turismo, trasporti e comunicazioni, credito</i>
COMMERCIO ESTERO		<i>Importazioni ed esportazioni per settore e Paese</i>

Alla produzione editoriale collocata nei 14 settori si affiancano le pubblicazioni periodiche dell'Istituto: Annuario statistico italiano, Bollettino mensile di statistica e Compendio statistico italiano. Il Rapporto annuale dell'Istat viene inviato a tutti gli abbonati anche ad un solo settore.



SISTEMA STATISTICO NAZIONALE
ISTITUTO NAZIONALE DI STATISTICA

**La rilevazione sulle forze di lavoro:
contenuti, metodologie,
organizzazione**

A cura di:

Saverio Gazzelloni

Per chiarimenti sul contenuto
della pubblicazione rivolgersi a:
Istat - Servizio Formazione e lavoro
Tel. 06 46734816
e-mail: infolav@istat.it

La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione

Metodi e norme n. 32 - 2006

Istituto Nazionale di Statistica
Via Cesare Balbo, 16 – Roma

Coordinamento:
Servizio produzione editoriale
Via Tuscolana, 1788 – Roma

*Prestampa e realizzazione del volume
in formato elettronico:* Nicola Piciocchi

Stampa:
C.S.R. Centro stampa e riproduzione s.r.l.
Via di Pietralata, 157 - Roma

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

Indice

	Pagina
Capitolo 1 - Introduzione	9
Capitolo 2 - I nuovi contenuti informativi	
2.1 - Premessa.....	13
2.2 - I nuovi criteri di individuazione degli occupati e dei disoccupati.....	13
2.3 - L'accrescimento del patrimonio informativo.....	16
Capitolo 3 - La rete di rilevazione	
3.1 - Premessa	23
3.2 - L'assetto organizzativo e la selezione dei rilevatori.....	23
3.3 - La formazione.....	25
Capitolo 4 - Il disegno campionario della nuova indagine e la fase di estrazione	
4.1- Premessa.....	29
4.1.1 - <i>Evoluzione storica del disegno campionario</i>	29
4.1.2 - <i>Obiettivi conoscitivi</i>	30
4.1.3 - <i>I requisiti di armonizzazione europea</i>	31
4.2 - Lo schema generale della rilevazione e del disegno campionario.....	32
4.3 - Contesto teorico di riferimento.....	33
4.3.1 - <i>Parametri di interesse dell'indagine continua</i>	34
4.3.2 - <i>Stimatori dei parametri di interesse</i>	35
4.4 - Il disegno della rilevazione.....	36
4.4.1 - <i>Lo schema di rotazione delle famiglie</i>	37
4.4.1.1 - <i>Struttura generale di schema di rotazione</i>	38
4.4.1.2 - <i>Varianza degli stimatori in presenza della rotazione</i>	38
4.4.1.3 - <i>Confronto tra gli schemi di rotazione considerati</i>	41
4.4.2 - <i>Studio della stratificazione</i>	43
4.4.2.1 - <i>Schema di campionamento nello spazio</i>	43
4.4.2.2 - <i>Schema di campionamento nel tempo</i>	46
4.4.3 - <i>Il metodo di associazione strato di comuni-settimana</i>	46
4.4.4 - <i>Confronto tra i disegni proposti</i>	48
4.5 - L'allocazione campionaria	49
4.5.1 - <i>Metodologia di allocazione</i>	49
4.5.2 - <i>Metodologia di calcolo degli errori attesi delle stime di interesse</i>	50
4.5.3 - <i>Metodologia di calcolo degli errori attesi della stima del parametro di riferimento Ue</i>	53
4.6 - Il disegno prescelto.....	54
4.7 - L'introduzione del nuovo campione.....	54
4.8 - La fase di estrazione delle famiglie.....	56
4.8.1 - <i>Il nuovo sistema di estrazione dei nominativi delle famiglie del campione</i>	56
4.8.2 - <i>Il sistema dei controlli</i>	58

Capitolo 5 - La tecnica di indagine e il monitoraggio del lavoro sul campo

5.1 - Premessa.....	61
5.2 - Il sistema Capi-Cati.....	63
5.3 - Il monitoraggio della qualità del lavoro sul campo.....	66
5.3.1 - <i>Gli standard utilizzati per la definizione degli esiti</i>	70
5.3.2 - <i>La scheda contatti per la registrazione degli esiti</i>	72
5.3.3 - <i>Gli esiti finali nell'indagine Rcfl con tecnica mista Capi-Cati</i>	73
5.3.4 - <i>I tassi calcolati per l'indagine</i>	77
5.3.5 - <i>I tassi utilizzati nell'indagine per il monitoraggio di survey</i>	77
5.3.6 - <i>I tassi utilizzati nell'indagine per il monitoraggio di field</i>	82
5.3.7 - <i>Le analisi multidimensionali</i>	84

Capitolo 6 - Il questionario elettronico

6.1 - Premessa.....	91
6.2 - Il questionario elettronico delle forze di lavoro.....	92
6.2.1 - <i>La scheda contatti</i>	93
6.2.2 - <i>I percorsi del questionario</i>	94
6.2.3 - <i>Percorsi a conferma per le successive interviste</i>	99
6.2.4 - <i>Le potenzialità del questionario elettronico: personalizzazione dei quesiti</i>	103
6.2.5 - <i>I controlli interni al questionario</i>	107
6.2.6 - <i>I controlli di range</i>	108
6.2.7 - <i>I controlli trasversali di consistenza e verosimiglianza</i>	109
6.2.8 - <i>I controlli longitudinali</i>	111
6.2.9 - <i>Le tecniche di correzione, messaggi di errore e domande a conferma</i>	111
6.2.10 - <i>Confronto tra l'indagine trimestrale forze di lavoro e la Nuova indagine continua sulle forze di lavoro</i>	112

Capitolo 7 - Il trattamento dei dati e le procedure di stima

7.1 - Premessa.....	115
7.2 - Il piano di controllo e correzione dei dati.....	117
7.2.1 - <i>L'individuazione degli errori</i>	117
7.2.2 - <i>L'imputazione</i>	119
7.2.3 - <i>Struttura e funzionalità del piano di controllo e correzione della Rcfl</i>	121
7.2.4 - <i>La matrice di incompatibilità tra professione e attività economica</i>	122
7.2.4.1 - <i>I presupposti per la realizzazione della matrice</i>	122
7.2.4.2 - <i>Le matrici: attribuzione dei giudizi di compatibilità</i>	123
7.2.4.3 - <i>L'utilizzo delle matrici di compatibilità</i>	124
7.2.5 - <i>La "procedura famiglie"</i>	126
7.2.5.1 - <i>La logica della "procedura famiglie"</i>	126
7.2.5.2 - <i>La revisione della procedura alla luce delle caratteristiche della Rcfl</i>	128
7.2.5.3 - <i>La "procedura famiglie" nella Rcfl</i>	129
7.3 - Lo stimatore e la procedura di riporto all'universo.....	132
7.3.1 - <i>Calcolo del peso base come reciproco della probabilità di inclusione</i>	134
7.3.2 - <i>La mancata risposta totale delle famiglie</i>	134
7.3.2.1 - <i>Correzione per mancata risposta per tipologia familiare</i>	138
7.3.3 - <i>I pesi finali</i>	139

	Pagina
7.3.4 - <i>I totali noti</i>	140
7.3.4.1 - <i>I totali noti della popolazione residente nelle regioni e nelle province</i>	141
7.3.4.2 - <i>I totali noti della popolazione residente nei 12 grandi comuni</i>	142
7.3.4.3 - <i>I totali noti sugli stranieri residenti</i>	143
7.3.4.4 - <i>I totali noti sul numero di famiglie</i>	144
7.3.4.5 - <i>I totali noti mensili</i>	144
7.4 - <i>La valutazione dell'errore campionario</i>	146
7.4.1 - <i>La stima della varianza campionaria dello stimatore di ponderazione vincolata</i>	147
7.4.2 - <i>Stima della varianza campionaria di una stima trimestrale di un totale</i>	149
7.4.3 - <i>Stima della varianza campionaria di una stima di media annua di un totale</i>	150
7.4.4 - <i>Stima della varianza campionaria di una stima di una differenza trimestrale</i>	152
7.4.5 - <i>Stima della varianza campionaria della stima di altri parametri della popolazione</i>	153
7.4.6 - <i>Presentazione sintetica delle stime degli errori campionari</i>	153
8 - La ricostruzione delle serie storiche dei principali indicatori del mercato del lavoro	
8.1 - <i>Premessa</i>	161
8.2 - <i>Il periodo di sovrapposizione</i>	161
8.2.1 - <i>Il processo di pre-aggiustamento dei dati</i>	162
8.3 - <i>L'approccio alla ricostruzione</i>	162
8.4 - <i>Il metodo di ricostruzione</i>	163
9 – Il sistema informativo-informatico	
9.1 - <i>Il Sistema informativo per la gestione delle indagini sulle famiglie</i>	169
9.2 - <i>L'internalizzazione del sistema Capi</i>	172
9.2.1 - <i>Introduzione</i>	172
9.2.2 - <i>L'agenda del rilevatore</i>	174
9.2.3 - <i>Il questionario elettronico</i>	177
9.2.4 - <i>Il nuovo dominio Rcfl</i>	178
9.2.5 - <i>Il sistema dei metadati</i>	179
9.3 - <i>Il raccordo tra Sigif e il sistema Capi internalizzato</i>	183
9.4 - <i>Il rinnovamento tecnologico dell'architettura informatica di sicurezza</i>	185
9.5 - <i>L'internalizzazione come investimento per il servizio di produzione</i>	188
Bibliografia	191

Capitolo 1 - Introduzione

Il Regolamento comunitario n. 577/98 prevede lo svolgimento di un'indagine armonizzata sulle forze di lavoro da effettuarsi durante tutte le 52 settimane di un anno. L'indagine è stata ulteriormente vincolata ai successivi Regolamenti n°1575/2000 (che esplicita le variabili target dell'indagine), n°1897/2000 (che definisce il concetto di persona in cerca di occupazione) e ad altri successivi regolamenti che hanno apportato alcune integrazioni alla lista delle variabili target. A questo impegno comunitario l'Istat ha risposto con la progettazione e la messa a regime di una nuova indagine, la Rilevazione continua sulle forze di lavoro (Rcfl), orientata al raggiungimento di elevati livelli di qualità.

La caratteristica maggiormente innovativa introdotta dal Regolamento comunitario è l'obbligo di svolgere le interviste nel corso di tutte le settimane dell'anno. L'innovazione ha infatti comportato una radicale riorganizzazione del disegno di campionamento, delle operazioni sul campo e delle operazioni di acquisizione, elaborazione e rilascio dei dati. La nuova indagine realizzata dall'Istat per rispondere a tali requisiti è, di fatto, un concentrato di innovazione e rappresenta una vera e propria rivoluzione nel modo di condurre le indagini. Ogni aspetto dell'indagine ha ricevuto il massimo dell'attenzione e il massimo dell'investimento per rispondere alle esigenze comunitarie e per offrire un prodotto di elevata qualità. L'innovazione si è concretizzata dunque su tutti i piani di lavoro dell'indagine, da quello statistico-metodologico a quello informatico, da quello amministrativo a quello organizzativo e tecnico. Nel presente volume si riportano sinteticamente, dopo una ricognizione delle novità introdotte anche a livello di contenuti (vedi Capitolo 2), tutti i processi attivati sul piano metodologico e su quello organizzativo per passare dalla vecchia rilevazione trimestrale sulle forze di lavoro alla nuova indagine continua prevista dal Regolamento Comunitario.

La nuova indagine è stata condotta in parallelo alla vecchia per tutto il 2003 e il primo trimestre del 2004. Dal 1° gennaio 2004 rappresenta la fonte ufficiale per la produzione di dati sul mercato del lavoro. La sovrapposizione delle due indagini è stata progettata e realizzata per costruire un modello di raccordo delle serie storiche che ha permesso di ricostruire, secondo le caratteristiche della nuova indagine, i dati pregressi prodotti dalla Rilevazione trimestrale sulle forze di lavoro (Rtfl).

Per poter attuare quanto previsto dal Regolamento Ue, nel 2000 il Consiglio d'Istituto decise di avviare una sperimentazione tesa alla costituzione di una Rete di rilevazione Istat (vedi Capitolo 3). E' stata questa una decisione storica che ha dotato l'Istituto, per la prima volta dalla sua nascita, di una rete di rilevazione autonoma di intervistatori professionisti che lavorano a diretto contatto con l'Istituto. Il numero di rilevatori necessario a coprire le aree in cui è stato suddiviso tutto il territorio nazionale è stato determinato sulla base dei seguenti elementi: a) durata della rilevazione b) numero di interviste da effettuare c) distribuzione dei comuni campione sul territorio d) durata dell'intervista e) distanza chilometrica dell'abitazione dell'intervistatore dai Comuni campione.

Gli intervistatori effettuano le interviste con un personal computer portatile. Si collegano con l'Istat tramite Internet via telematica a) per ricevere il questionario elettronico, l'agenda dei contatti, i nominativi delle famiglie da intervistare e tutte le comunicazioni necessarie alla conduzione delle interviste e, allo stesso tempo, b) per trasmettere con frequenza giornaliera i dati delle interviste complete, l'esito dei tentativi di contatto e dei contatti definitivi.

Di Saverio Gazzelloni

L'editing del volume è stato curato da Elisabetta Mancini

Per poter garantire l'alto livello professionale dei rilevatori, l'Istat ha curato con particolare attenzione la loro selezione e formazione. Il percorso formativo è stato strutturato in tre momenti: formazione d'aula, formazione continua e formazione a distanza.

La formazione d'aula mira a fornire ai partecipanti gli elementi metodologici dell'indagine, le informazioni sugli aspetti giuridici a tutela della privacy dei cittadini, sugli obiettivi e i contenuti dell'indagine, sul campione da intervistare, sul periodo di rilevazione, sulla struttura del questionario, nonché tende a sviluppare specifiche competenze comunicative e relazionali.

Le attività di formazione continua si esplicano, inoltre, quotidianamente nei rapporti regolari che si instaurano tra Servizio centrale, referenti dell'indagine e collaboratori di rete territoriali da un lato e tra quest'ultimi e rilevatori dall'altro. L'attività di formazione converge poi in debriefing che sono organizzati regolarmente con cadenza mensile nelle sedi regionali Istat.

Per far fronte al naturale turn-over dei rilevatori, a tutto ciò si è affiancata la predisposizione di strumenti per la formazione a distanza su base individuale, che prevede a) una giornata di accoglienza di un rilevatore neo-assunto nell'Ufficio regionale b) una fase di autoapprendimento (5 giorni) tramite l'utilizzo di un pacchetto multimediale (organizzato in moduli formativi e test) installato sul Pc consegnato al momento della stipula del contratto e c) altri due giorni di formazione presso gli Uffici territoriali, dove il neo-assunto viene seguito dai referenti dell'indagine e dai collaboratori di rete per la verifica e l'integrazione dell'autoformazione.

I requisiti espressi dal Regolamento 577/98 che hanno impatto sul disegno campionario (vedi Capitolo 4) riguardano la periodicità dell'indagine, la precisione delle stime, le caratteristiche generali del disegno di campionamento e la distribuzione del campione tra le settimane del trimestre.

L'Indagine è un'indagine continua, realizzata nelle 52 settimane dell'anno, che deve fornire stime attendibili a livello trimestrale e annuale.

Per popolazioni di dimensioni pari al 5 per cento della popolazione in età lavorativa, la deviazione relativa standard per la stima delle medie annuali non deve superare l'8 per cento a livello Nuts II (livello regionale); le regioni con meno di 300mila abitanti possono non soddisfare questo requisito. Per popolazioni di dimensioni pari al 5 per cento della popolazione in età lavorativa, la deviazione relativa standard per la stima delle variazioni tra due trimestri successivi, a livello nazionale, non deve superare il 2 per cento.

Dal punto di vista generale, sono state mantenute le principali caratteristiche del disegno di campionamento della Rilevazione trimestrale forze di lavoro, ovvero: due stadi di selezione con stratificazione delle unità primarie a livello provinciale con i comuni di ciascuna provincia stratificati in base alla popolazione residente e suddivisi in due sottoinsiemi (autorappresentativi - Ar e non autorappresentativi - Nar). Il disegno è stato determinato in base alla scelta di tre fattori: il numero minimo di interviste per comune; l'allocazione del campione complessivo tra i domini territoriali (regioni e province); il numero di comuni campione per strato Nar. Anche nel caso della Rcfl, in conformità a quanto viene fatto per le principali indagini sulle forze di lavoro anche in paesi extra-Ue (Current population survey americana e labour force survey canadese), è stato mantenuto uno schema di rotazione del campione (2,2,2) analogo a quello presente nella Rtfl.

Sulla base di valutazioni in termini sia di efficienza delle stime, sia di praticabilità organizzativa, è stato scelto un disegno a stratificazione mensile, per cui tutti gli strati territoriali sono rappresentati in ciascun mese del trimestre.

Poiché la quasi totalità dei Comuni dispone ormai di anagrafi informatizzate, l'Istat ha realizzato un sistema informatico che gestisce le attività connesse all'estrazione e sostituzione delle famiglie campione. Tale sistema presenta numerosi vantaggi: fornisce informazioni sulle modalità di effettuazione dell'estrazione (in particolare sui criteri di ordinamento); facilita la registrazione dei nominativi estratti e ne permette un immediato e sicuro controllo (controllo

della correttezza del passo di estrazione utilizzato, controllo della numerosità campionaria, controllo dei campi obbligatori, eccetera); facilita la trasmissione dei file dai Comuni all'Istat (vedi Capitolo 5).

L'indagine utilizza una tecnica mista per effettuare le interviste, servendosi dello stesso questionario per le differenti tecniche e per le diverse occasioni di intervista (vedi Cap. 6). In linea generale la prima intervista è realizzata dalla rete dei rilevatori sul territorio con tecnica Capi (*Computer assisted personal interview*), mentre le interviste successive vengono effettuate da una ditta con tecnica Cati (*Computer assisted telephone interview*). Esistono eccezioni a questa impostazione che fanno sì che alcune prime interviste vengano effettuate con tecnica Cati (per periodi dell'anno particolari) e interviste successive alla prima anche con tecnica Capi (per le famiglie senza telefono).

I nominativi delle famiglie del campione sono automaticamente assegnati dal sistema a ciascun rilevatore per la propria area di competenza. La prima intervista viene effettuata in una settimana specifica assegnata dal disegno di campionamento, mentre le 3 successive seguono a tre, dodici e quindici mesi di distanza dalla prima intervista. Al fine di garantire il rispetto della numerosità campionaria, al campione base vengono associati altri tre campioni suppletivi (estratti con gli stessi criteri probabilistici). Le famiglie suppletive vengono intervistate solo quando non è possibile intervistare la famiglia base. La sostituzione è gestita automaticamente in tempo reale dal questionario Capi, ed è pertanto impossibile per il rilevatore effettuare scelte arbitrarie sui nominativi a sua disposizione.

Nel corso dell'indagine l'intero processo dei contatti con le famiglie è gestito mediante procedure automatiche e tutti gli esiti dei tentativi di contatto e dei contatti sono registrati su un apposito file. In tal modo, il ricorso a tecniche di acquisizione assistite da computer (Capi e Cati) ha consentito di sviluppare uno studio accurato della qualità del processo di rilevazione dei dati. Si sono realizzati sistemi di supervisione utili sia per valutare le difficoltà incontrate dai rilevatori o dai rispondenti, sia per valutare l'andamento della rilevazione, sia per valutare la qualità complessiva del lavoro svolto.

Gli studi sulla dinamica dell'intervista hanno permesso di definirne tutti gli aspetti, individuando per ciascuna micro-fase un insieme di indicatori di qualità idonei a controllare il rispetto delle scelte metodologiche, delle procedure e dei tempi stabiliti.

Le informazioni sono continuamente diffuse via Intranet per consentire agli uffici territoriali di condividere in tempo reale con il centro di coordinamento l'analisi dell'andamento della rilevazione, fin nei minimi dettagli (gli indicatori sono infatti disaggregati per settimana, trimestre, anno, regione, provincia e intervistatore).

Oltre a controllare la rilevazione mediante indicatori ad hoc, agli uffici regionali è inoltre demandato un controllo successivo da effettuare telefonando alle famiglie immediatamente dopo lo svolgimento dell'intervista. I controlli sono di due tipi: il primo volto alla verifica dell'avvenuta intervista, il secondo volto a raccogliere informazioni sull'andamento della stessa. Questa iniziativa consente di acquisire indicazioni significative sulla qualità della rete e fornire indicazioni sull'esistenza di interviste non valide da eliminare. A tale scopo è stato predisposto un apposito questionario condiviso in modalità client-server sulla Intranet dell'Istat per avere centralizzati in tempo reale gli esiti dei controlli telefonici sulle famiglie.

Il questionario elettronico utilizzato per la rilevazione (vedi Capitolo 7) è stato organizzato e costruito in modo da ridurre al minimo i problemi legati alla gestione del susseguirsi delle domande e alla gestione delle schermate. Tutti i tentativi di contatto con le famiglie campione sono schedulati e storicizzati.

Le classificazioni delle professioni e delle attività economiche avvengono tramite motore di ricerca durante l'intervista o, qualora ciò non sia possibile, al termine della stessa utilizzando gli

alberi gerarchici delle classificazioni CP2001 per le professioni e Ateco 2002 per le attività economiche.

Sono state definite, nell'ambito dei percorsi del questionario, regole *hard* (quelle che inibiscono la prosecuzione dell'intervista finché non è stata corretta la risposta errata) e regole *soft* (quelle che lasciano la possibilità di chiarire i dubbi con il rispondente e quindi di valutare se effettuare o meno la correzione).

Rispetto agli errori, quindi, l'utilizzo di tecniche C.a.i. (*Computer Assisted Interviewing*) ha consentito di inserire già nella fase di raccolta delle informazioni statistiche tutte le regole che vengono normalmente applicate nelle successive fasi di revisione e correzione dei dati rilevati.

Nel caso di interviste successive alla prima, sia per le interviste Capi che le interviste Cati, il software permette di utilizzare le informazioni raccolte durante l'intervista precedente, permettendo così all'intervistatore di chiedere all'intervistato solamente una conferma o una disconferma della precedente situazione.

Ulteriori innovazioni sono state introdotte sul piano del trattamento dei dati e delle procedure di stima (vedi Capitolo 8). Particolare attenzione è stata rivolta ai piani di correzione delle informazioni raccolte. Infatti, pur in presenza della notevolissima riduzione del numero di errori (dovuti al rispondente o al rilevatore) ottenuta con l'utilizzo di tecniche *computer assisted*, la complessità del questionario elettronico ha comportato la necessità di un piano di correzione molto articolato e complesso finalizzato a rendere logicamente compatibili tutte le informazioni raccolte.

Di particolare rilievo, sempre nell'ambito dei complessi piani di check attivati successivamente alla raccolta dei dati (che hanno però positivi e continui *feed-back* anche sulla messa a punto dello strumento di rilevazione), sono i due processi attivati per il controllo e la correzione delle classificazioni delle professioni e delle attività economiche da un lato e, dall'altro, delle relazioni di parentela nell'ambito della famiglia.

Anche nelle strategie di calcolo dei pesi campionari per il riporto all'universo e nella valutazione dell'errore campionario sono state introdotte novità di rilievo che assicurano un maggior livello di qualità dei risultati finali.

Nel penultimo capitolo del presente volume (Capitolo 9) si riportano i processi attivati e le metodologie applicate per ricostruire le serie storiche dei principali aggregati del mercato del lavoro a seguito dell'introduzioni della nuova indagine. Il cambiamento introdotto, infatti, ha riguardato praticamente tutti gli elementi portanti dell'indagine e le stime prodotte hanno risentito ovviamente di tutto ciò. Al fine di rendere immediatamente disponibili le serie per eseguire analisi di lungo periodo e rendere coerenti i processi di destagionalizzazione necessari all'analisi congiunturale, si è perciò lavorato intensamente, e in parallelo all'implementazione della nuova indagine, per rendere disponibili le nuove serie ricostruite contestualmente alla diffusione dei primi dati prodotti con la nuova rilevazione continua.

Il capitolo 10, in chiusura di volume, illustra sinteticamente tutte le caratteristiche del sistema informativo-informatico che ha reso possibile la realizzazione della nuova indagine, dalla gestione della rete, all'estrazione del campione, dalle trasmissioni delle interviste al trattamento dei dati, al monitoraggio della qualità del lavoro sul campo. Di particolare rilievo, inoltre, è il processo attivato nell'ambito dell'assetto informatico per procedere all'internalizzazione della componente Capi del sistema, messo in atto successivamente all'avvio dell'indagine e realizzato per entrare a regime per effettuare le prime interviste del 2006.

Capitolo 2 - I nuovi contenuti informativi

2.1 Premessa

L'indagine sulle forze di lavoro ricopre nel nostro paese un ruolo di primo piano nella documentazione statistica sul mercato del lavoro. Chiamata a rispondere all'esigenza centrale di misurare l'occupazione e la disoccupazione, l'indagine approfondisce inoltre modi e gradi della partecipazione al mercato del lavoro. Tale fabbisogno conoscitivo si è progressivamente ampliato nel corso degli ultimi anni sollecitando un potenziamento dell'impianto contenutistico, oggi soddisfatto dalla nuova indagine sulle forze di lavoro.

Come per la Rtfi, il principale obiettivo della nuova indagine è rappresentato dalla produzione delle stime ufficiali degli occupati e delle persone in cerca di occupazione. A tal fine è necessario ripartire la popolazione in età lavorativa (15 anni e oltre) in tre gruppi esaustivi e mutuamente esclusivi: occupati, disoccupati, inattivi. La suddivisione avviene secondo criteri classificatori basati sulle definizioni ispirate dall'International labour office e recepite dai Regolamenti comunitari. Nell'applicazione di questi criteri viene seguito un principio gerarchico: prima si identificano gli occupati, successivamente - tra tutti i non occupati - le persone in cerca di occupazione e, infine, le persone inattive, quelle non incluse tra gli occupati o i disoccupati.

Lo strumento utilizzato per rendere operativi i prescritti criteri classificatori è il questionario elettronico, la cui tecnica di somministrazione è illustrata successivamente. In questa sede conta sottolineare il compito delicato e complesso rappresentato dalla stesura del questionario. Determinata in misura molto vasta dai quesiti indicati dai Regolamenti comunitari, la stesura del questionario ha posto particolare attenzione all'ordine delle domande, ai numerosi svincoli (o domande filtro), alla chiarezza del linguaggio nella formulazione dei quesiti. A tale ultimo riguardo, in una molteplicità di casi, i quesiti vengono rivolti in modo differente a individui che si trovano in situazioni diverse. La dilatazione degli obiettivi conoscitivi e la specificità del contesto nazionale hanno peraltro motivato quesiti aggiuntivi stabiliti dall'Istat. Nel seguito l'esame dei contenuti informativi della nuova indagine farà pressoché esclusivo riferimento al nucleo centrale del questionario, composto da otto sezioni ognuna caratterizzata da uno specifico obiettivo conoscitivo, che rileva le informazioni sugli individui intervistati¹.

2.2 - I nuovi criteri di individuazione degli occupati e dei disoccupati

Nella condizione di occupato si classificano le persone (con almeno 15 anni) che nella settimana di riferimento, quella che in generale precede la settimana in cui viene condotta l'intervista, hanno svolto almeno un'ora di lavoro in una qualsiasi attività che preveda un corrispettivo monetario². Nella condizione di occupato vengono inoltre classificati coloro che, sempre nella settimana di riferimento, hanno svolto almeno un'ora di lavoro anche non retribuito nella ditta di un familiare nella quale collaborano abitualmente.

Di Mario Albisinni

¹ Il questionario individuale è preceduto da una "Scheda generale" in cui vengono registrate le informazioni di base sulla famiglia e sugli individui che la compongono. Più in particolare, si rilevano le caratteristiche socio-demografiche di tutti i componenti della famiglia e si individuano gli effettivi componenti da intervistare. Per maggiori ragguagli si vedano le prime pagine del questionario, oggi diffuso in una versione semplificata rispetto a quello informatico effettivamente utilizzato.

² Il corrispettivo dell'attività lavorativa svolta può anche essere determinato in natura o in una fornitura di servizi.

La domanda sullo svolgimento di un'attività, in termini di una o più ore, apre la sezione B del questionario rivolta a distinguere gli occupati dai non occupati. Essa intende cogliere un fenomeno oggettivo, determinante per l'esito classificatorio nella condizione di occupato. Basata sui principi stabiliti in sede comunitaria, la strategia di interrogazione seguita dalla Rcfl è mutata radicalmente rispetto alla Rtf³. La rilevazione trimestrale si apriva difatti con un quesito sulla condizione dichiarata, ossia sulla percezione che l'intervistato ha della propria condizione. Scontando fortemente gli atteggiamenti soggettivi, l'autopercezione era decisiva per l'identificazione degli occupati. Nella sequenza dei quesiti, quello sullo svolgimento di una o più ore lavorative era collocato in seconda battuta e, rivolto ai soggetti che dichiaravano una condizione diversa da quella di occupato, consentiva esclusivamente il calcolo dei cosiddetti occupati non dichiarati, un aggregato di entità storicamente modesta nelle serie storiche della Rtf⁴.

La condizione di occupato nella nuova indagine sulle forze di lavoro è dunque completamente svincolata dall'opinione che il soggetto intervistato ha del proprio status. In base ai criteri adottati un individuo può essere classificato come occupato ma percepire in modo differente la propria condizione. D'altro canto, un individuo può non rientrare nei criteri per essere classificato come occupato nonostante possa ritenersi tale. L'innovazione introdotta dalla RCFL con la modifica del quesito di apertura del questionario individuale è di gran lunga la più importante per la costruzione della condizione occupazionale. Ciononostante, la Rcfl ha mantenuto in una specifica parte del questionario, collocata dopo le sezioni necessarie alla classificazione della posizione lavorativa dell'intervistato, il tradizionale quesito sulla condizione autopercepita. La scelta fatta, che non risponde ad alcun vincolo comunitario, consente di confrontare i risultati ottenuti con i nuovi criteri con quelli connessi alle valutazioni soggettive sulla partecipazione al mercato del lavoro.

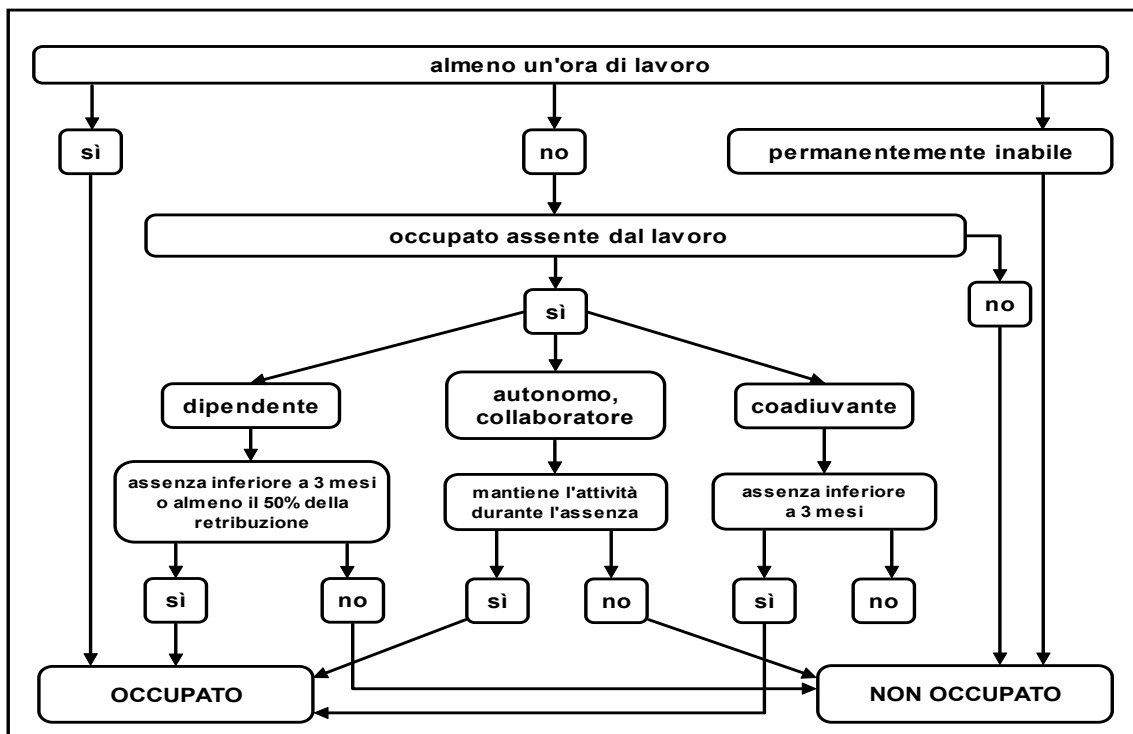
Con riguardo alla stima dell'occupazione, i principi formulati in sede comunitaria sono alla base di un'ulteriore caratteristica della Rcfl. Si tratta delle condizioni che devono essere verificate per classificare come occupato un individuo assente dal lavoro. In particolare, se l'intervistato non ha svolto ore di lavoro, è necessario verificare l'esistenza di due requisiti: la ripresa dell'attività dopo il periodo di assenza e la durata non superiore a tre mesi dell'assenza stessa. Per i lavoratori dipendenti la durata dell'assenza può essere superiore a tre mesi se la retribuzione percepita è almeno pari al 50 per cento di quella antecedente l'inizio del periodo di assenza; per i lavoratori autonomi se l'attività viene comunque mantenuta. Non rilevata dalla Rtf¹, l'introduzione della variabile "assenza dal lavoro" amplia il solco dovuto al passaggio da un criterio classificatorio fondato sull'autopercezione ad uno centrato sul verificarsi di specifiche condizioni oggettive. Nella precedente indagine difatti la sola dichiarazione del proprio stato di occupato consentiva anche a un soggetto assente dal lavoro di rientrare tra gli occupati, senza l'accertamento di alcuna condizione e vincolo di prossimità temporale.

Lo schema sottostante riporta in estrema sintesi il percorso seguito dalla Rcfl per l'individuazione degli occupati.

³ Si veda in particolare quanto riportato in "Labour force survey: principles for the formulation of the questions on labour status" del Regolamento 1897/2000.

⁴ Nella Rtf¹ un soggetto che si dichiarava non occupato, e che rispondeva di avere effettuato una o più ore di lavoro, poteva trovarsi in una delle seguenti condizioni: disoccupato alla ricerca di nuova occupazione, in cerca di prima occupazione, con un lavoro che inizierà in futuro, casalingo/a, studente, ritirato/a dal lavoro, altra condizione. Nella media degli anni tra il 1993 e il 2003 l'incidenza degli occupati non dichiarati rimane intorno all'1 per cento del totale degli occupati.

Figura 2.1 - Individuazione degli occupati



L'individuazione delle persone in cerca di occupazione si fonda su requisiti classificatori molto dettagliati, validi per le persone di età compresa tra 15 e 74 anni⁵. Il primo requisito è risultare non occupato; il secondo è di essere disponibile a lavorare (o ad avviare un'attività autonoma) entro le due settimane successive il momento dell'intervista; il terzo è avere fatto almeno un'azione di ricerca di lavoro, tra quelle previste, nelle quattro settimane precedenti l'intervista⁶. Tale ultimo criterio non viene applicato agli individui non occupati che dichiarano di avere trovato un lavoro che inizierà entro tre mesi dalla data dell'intervista. Permane comunque anche per questo gruppo l'osservanza del requisito sulla disponibilità entro le due settimane qualora fosse possibile anticipare l'inizio del lavoro.

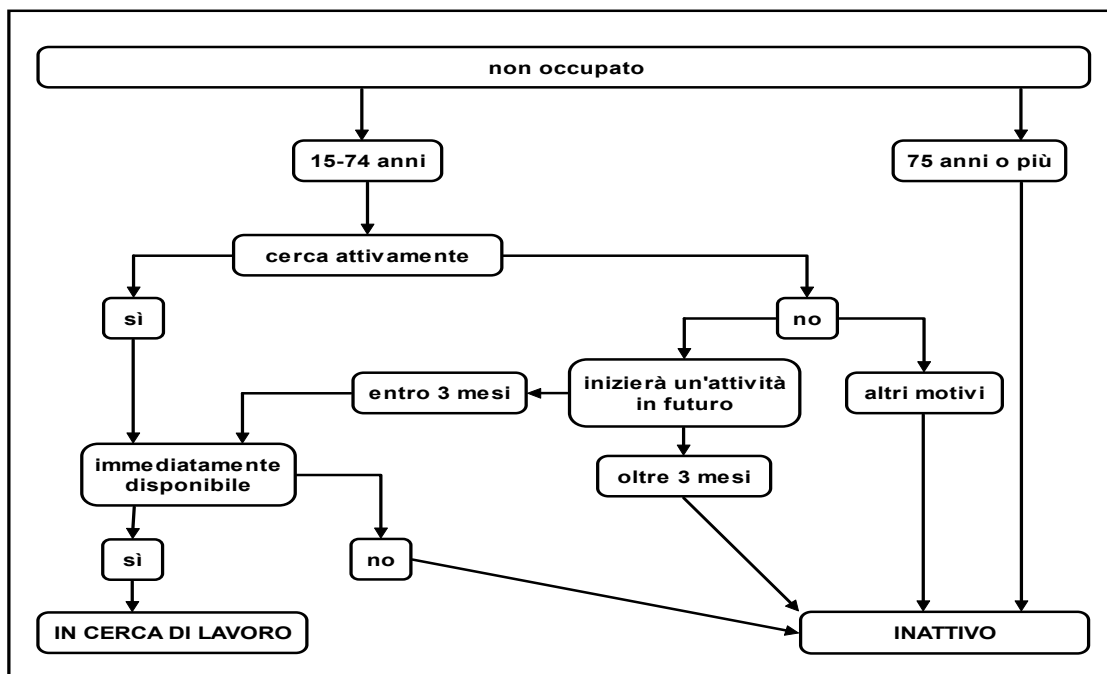
In confronto alle convenzioni definitorie seguite nella Rtf1 le modifiche introdotte dalla Rcf1 interessano aspetti "secondari". Si tratta, nel dettaglio, dell'introduzione del limite di età superiore a cui riferire la stima delle persone in cerca di occupazione, non fissato nella precedente indagine; della verifica del limite temporale e della disponibilità per coloro che inizieranno un'attività in futuro, entrambe escluse nella Rtf1.

Lo schema sottostante individua il percorso per la classificazione delle persone in cerca di occupazione.

⁵ Si veda in particolare quanto riportato in *Commission Regulation (EC) No 1897/2000 of 7 September 2000 implementing Council Regulation (EC) No 577/98 on the organisation of a labour force sample survey in the community concerning the operational definition of unemployment.*

⁶ Le azioni di ricerca cosiddette attive alle quali si fa riferimento sono riportate nel questionario dal quesito F16 al quesito F28.

Figura 2.2 - Individuazione delle persone in cerca di occupazione



In definitiva, pur continuando a rappresentare il fondamentale contenuto informativo dell'indagine sulle forze di lavoro, le stime degli occupati e delle persone in cerca di occupazione prodotte dalla Rcf1 presentano importanti elementi di diversità nelle modalità di costruzione e negli aspetti definitori rispetto alla precedente rilevazione. Ma al di là di tali aspetti, ulteriori differenze nelle stime sulla partecipazione al mercato del lavoro sono attese derivare dalle più radicali innovazioni metodologiche introdotte con la Rcf1: dal passaggio da una rilevazione concentrata in una singola settimana ogni tre mesi ad una ripartita nell'arco delle tredici settimane di un trimestre, all'utilizzo di una nuova rete di rilevazione, all'impiego di tecniche di indagine *computer assisted*.

2.3 - L'accrescimento del patrimonio informativo

Con un'articolazione del questionario rivolta a cogliere in modo più nitido le multiformi caratteristiche del mercato del lavoro, la Rcf1 accresce sensibilmente il patrimonio informativo. In particolare, l'accrescimento riguarda importanti aspetti dell'attività lavorativa, della disoccupazione, dell'istruzione e formazione e delle relazioni familiari.

Colmando un vuoto della precedente indagine, la Rcf1 soddisfa il fabbisogno conoscitivo sulle prestazioni lavorative svolte con modalità di collaborazione coordinata e continuativa o di prestazione d'opera occasionale. Nella Rcf1, centrata sulla tradizionale dicotomia tra lavoro dipendente e autonomo, anche i soggetti coinvolti dalle appena richiamate tipologie di lavoro "non standard" potevano esclusivamente e alternativamente dichiararsi ed essere classificati lavoratori dipendenti o indipendenti. Nella Rcf1 la posizione nella professione viene invece rilevata tenendo distinte le collaborazioni e le prestazioni occasionali sia dal lavoro alle

dipendenze sia da quello autonomo⁷. Ciò consente agli intervistati di collocarsi correttamente nella griglia prevista, superando le precedenti ambiguità e difficoltà nella scelta della posizione professionale.

Ne consegue un primo allargamento dell'offerta informativa, ulteriormente accresciuto da un insieme di quesiti volti a determinare l'effettivo grado di autonomia nello svolgimento del lavoro di collaborazione coordinata e continuativa o di prestazione occasionale. Più in particolare si tratta dell'individuazione dei principali connotati di erogazione dell'attività: a favore di uno o più datori di lavoro, nel luogo di pertinenza del committente o altrove, secondo predeterminati schemi di orario o senza alcun vincolo⁸. Sebbene il paradigma classificatorio di ultima istanza collochi le collaborazioni coordinate e continuative e le prestazioni d'opera occasionali nel lavoro autonomo, le informazioni sulle modalità con cui queste figure professionali svolgono realmente il proprio lavoro forniscono importanti elementi di contesto per una valutazione del loro stadio di affiancamento all'area del lavoro subordinato piuttosto che a quella del lavoro autonomo.

In ogni caso sembra opportuno sottolineare che, come per l'insieme dei quesiti sulla condizione lavorativa, l'esistenza di un rapporto di impiego nella forma di collaborazione coordinata e continuativa o di prestazione occasionale riguarda un periodo temporale rigidamente delimitato, in generale la settimana precedente l'intervista. Con riferimento poi alle collaborazioni coordinate e continuative esercitate come attività principale, la nuova indagine rileva quelle per cui esiste l'obbligo dei versamenti contributivi da parte del committente⁹. Sono pertanto escluse le collaborazioni svolte dai professionisti, vincolati al pagamento diretto dei contributi stabiliti. Tale esclusione evita indebite sovrapposizioni tra la figura del collaboratore/professionista e quella del libero professionista, categoria già prevista all'interno del lavoro autonomo. Essa inoltre consente all'intervistato di riconoscere univocamente la figura del collaboratore coordinato e continuativo¹⁰.

Infine la Rcfi non tiene conto delle recenti innovazioni nella regolamentazione del rapporto di collaborazione coordinata e continuativa¹¹. Come è noto, le nuove norme prescrivono nella generalità dei casi l'individuazione di un progetto, per definizione a termine, come modalità di esecuzione della fattispecie in esame¹². La Rcfi invece non distingue tra collaborazioni a progetto e altre tipologie di collaborazione coordinata e continuativa, rilevandole attualmente in un unico aggregato. La scelta fatta risponde essenzialmente ad un criterio generale sulla qualità delle risposte acquisite tramite un'indagine campionaria sulle famiglie. L'esigenza di affidabilità statistica delle stime ha consigliato di non rilevare le nuove tipologie contrattuali previste dal quadro normativo in attesa che queste raggiungano un adeguato livello di diffusione. Allo stato

⁷ Nel questionario la posizione nella professione viene rilevata nella sezione B, per gli individui occupati assenti dal lavoro; nella sezione C, per l'attività lavorativa principale dei soggetti che hanno svolto almeno un'ora di lavoro; nella sezione D, per l'attività lavorativa secondaria degli occupati; nella sezione E, per le precedenti esperienze di lavoro dei non occupati. La struttura del quesito è identica nelle quattro sezioni e la risposta dell'intervistato deve essere fornita dopo avere ascoltato la lettura dei casi previsti.

⁸ I relativi quesiti dal C5 al C7 si trovano nella sezione sull'attività lavorativa principale.

⁹ In base alla modalità di versamento dei contributi previdenziali, la gestione separata presso l'Inps, istituita dalla legge 335 del 1995, distingue due categorie: quella dei collaboratori, i cui contributi vanno versati dal committente nel mese successivo a quello del pagamento del compenso e quella dei professionisti, tenuti a versare direttamente i contributi in occasione della dichiarazione dei redditi.

¹⁰ Nelle indagini pilota, condotte prima dell'avvio della Rcfi, era emersa la difficoltà degli intervistati di definire la propria posizione di collaboratore coordinato e continuativo confusa con quella di lavoratore autonomo o dipendente.

¹¹ Si fa riferimento al Decreto Legislativo 10 settembre 2003, n. 276.

¹² Le norme sulle collaborazioni a progetto, oltre a non trovare applicazione per le pubbliche amministrazioni e per il loro personale, escludono taluni soggetti (principalmente i pensionati di vecchiaia e gli iscritti ad albi professionali) e le prestazioni inferiori ai 30 giorni e sino a 5mila euro annui di corrispettivo.

attuale, i soggetti intervistati potrebbero difatti trovare forti difficoltà a distinguere e a riconoscersi in una delle fattispecie normative previste.

Un ulteriore significativo arricchimento dell'offerta informativa prodotta dalla Rcfl è rappresentato dalla rilevazione del lavoro interinale¹³. Le nuove informazioni consentiranno di integrare quelle attualmente disponibili dalle fonti amministrative, in particolare dell'Inps e dell'Inail, che soffrono in primo luogo di ritardi nell'aggiornamento degli archivi. L'archivio dell'Inps non consente peraltro di stimare il numero dei lavoratori interinali. Cogliendo le contribuzioni mensili di ogni singola missione pure se effettuata dallo stesso individuo, l'archivio è difatti soggetto a possibili duplicazioni¹⁴. Si è pertanto più volte fatto ricorso a successive elaborazioni dei dati di base per una stima dei "lavoratori equivalenti a tempo pieno", un aggregato comunque non coincidente con quello della nuova indagine sulle forze di lavoro¹⁵. D'altro canto, le informazioni ricavabili dai dati dell'Inail, sebbene forniscono una stima dell'occupazione interinale, non appaiono esaustive dell'insieme dei settori di attività economica¹⁶.

La Rcfl fornisce inoltre nuovi contributi conoscitivi su altri rilevanti aspetti del mercato del lavoro. In risposta alle sollecitazioni di ampliamento del quadro informativo sulle motivazioni del ricorso al tempo parziale, con particolare riguardo a quelle orientate a supportare l'analisi sulla conciliazione tra tempi di lavoro e impegni familiari, la Rcfl misura quella parte di part-time volontario potenzialmente disponibile a lavorare a tempo pieno. Tale scelta viene collegata all'eventuale accessibilità di adeguati servizi per la cura dei figli e/o dei familiari¹⁷. Integrando la base informativa già presente nella precedente indagine sul luogo di svolgimento dell'attività principale, la Rcfl rileva poi sia il numero sia l'articolazione degli spostamenti legati a motivi di lavoro¹⁸. Alla misurazione dell'ammontare complessivo del fenomeno si accompagna difatti la possibilità di cogliere le caratteristiche dei trasferimenti originati da esigenze lavorative, distinguendo quelli a breve (all'interno dei confini provinciali o regionali) da quelli a lungo raggio (interregionali). Infine, la Rcfl allarga l'orizzonte conoscitivo sul delicato tema dell'inserimento lavorativo attraverso la rilevazione della data di inizio della prima occupazione¹⁹.

¹³ Il riscontro dell'esistenza di un rapporto contrattuale con un'agenzia di fornitura di lavoro interinale viene fornito, nella sezione sull'attività principale degli occupati, dalla risposta al quesito C24 del questionario. Al riguardo, si ricorda che il lavoro interinale è caratterizzato da due distinti contratti: il contratto di fornitura e il contratto di lavoro temporaneo. Nel contratto di fornitura, l'impresa fornitrice mette a disposizione di un'impresa utilizzatrice uno o più lavoratori. Nel contratto di lavoro temporaneo, l'impresa fornitrice assume il lavoratore che metterà a disposizione dell'impresa utilizzatrice. L'impresa fornitrice può assumere il lavoratore a tempo determinato per la durata della cosiddetta missione presso l'impresa utilizzatrice o a tempo indeterminato (in questo caso, l'impresa fornitrice paga un'indennità di disponibilità quando il lavoratore non si trova in missione).

¹⁴ I dati dell'Inps sono desunti dai versamenti contributivi che le agenzie di lavoro temporaneo sono tenute ad effettuare mensilmente in quanto titolari del rapporto di lavoro alla base della missione interinale. Nello stesso mese un individuo può essere impiegato in più imprese utilizzatrici ed essere quindi interessato da più contratti di fornitura. In tale caso la fonte previdenziale conta questo individuo un numero di volte pari a quello dei contratti in cui è coinvolto.

¹⁵ Tra le pubblicazioni che riportano una stima dei "lavoratori equivalenti a tempo pieno", si veda Istat, *Rapporto annuale: la situazione del Paese nel 2001*, Roma, 2002; Ministero del Lavoro e delle Politiche sociali, *Rapporto di monitoraggio sulle politiche occupazionali e del lavoro*, 2003. Una valutazione dei "lavoratori equivalenti a tempo pieno" è stata inoltre proposta in più occasioni a partire da campioni di dati forniti dalle Agenzie di lavoro temporaneo.

¹⁶ A fini di prevenzione degli infortuni, l'impresa utilizzatrice è tenuta a comunicare all'Inail il proprio codice fiscale, quello del lavoratore, la data di inizio e di fine utilizzo del lavoratore interinale. Depurando dalle duplicazioni, costituite dai rapporti aperti e chiusi dallo stesso lavoratore nell'arco di tempo osservato, è possibile calcolare il numero di lavoratori interinali. L'universo di riferimento dell'Inail include le imprese dell'industria e dei servizi. All'interno del terziario è tuttavia escluso il lavoro interinale prestato nelle Pubbliche Amministrazioni, ad eccezione di quello svolto nella sanità, nelle amministrazioni locali e negli Enti pubblici non economici.

¹⁷ Si veda il percorso proposto dai quesiti dal C27 al C30 nella parte del questionario relativa all'orario di lavoro abituale.

¹⁸ Si fa riferimento ai quesiti C53 e C54 sull'attività lavorativa principale.

¹⁹ Si vedano nel questionario i quesiti dal C60 al C62.

Un ampliamento dei contenuti informativi viene fornito anche con riguardo ai quesiti retrospettivi sulla condizione lavorativa e la residenza, ad un anno di distanza dall'intervista. Rilevati nella precedente indagine solo in relazione alla settimana di riferimento di aprile, tali quesiti vengono somministrati continuativamente nella Rcfl²⁰. La disponibilità di una base informativa non più limitata ad uno specifico periodo temporale diviene così il presupposto per una più estesa valutazione, per ogni singolo trimestre di riferimento dei dati prodotti dalla Rcfl, sia dei cambiamenti percepiti dal soggetto intervistato nella condizione professionale sia dei fenomeni di mobilità territoriale²¹.

I cambiamenti che interessano la rilevazione delle persone in cerca di lavoro sono di portata più limitata rispetto a quelli relativi all'occupazione, soprattutto per quanto riguarda l'ampliamento dei contenuti. A tale proposito, l'innovazione più rilevante, in linea con quanto avviene per il part-time, riguarda la possibile stima dell'effetto della carenza di servizi di assistenza. La Rcfl prevede difatti un apposito quesito, rivolto a coloro che dichiarano di non cercare lavoro per prendersi cura dei figli e/o di altri familiari. L'arricchimento dei contenuti deriva inoltre dall'ampliamento del ventaglio delle modalità di risposta per le domande relative alle motivazioni della mancata partecipazione al mercato del lavoro²². In confronto alla Rtfl, l'accrescimento è ottenuto sia suddividendo i motivi familiari e personali sia isolando la voce relativa alla maternità. Un ulteriore ampliamento contenutistico deriva infine dalla verifica del limite temporale entro cui inizierà l'attività lavorativa chi è in procinto di avviarne una²³.

Appaiono comunque consistenti le innovazioni che riguardano la sequenza e la formulazione dei quesiti utilizzati per raccogliere le informazioni relative alla disoccupazione. Il cambiamento di portata più ampia concerne la rilevazione delle azioni di effettuate dai non occupati che si dichiarano alla ricerca di un lavoro. Per ciascuna azione, infatti, è previsto uno specifico quesito, in sostituzione dell'unica domanda a risposta multipla utilizzata nella precedente indagine²⁴.

²⁰ Conformandosi al Regolamento comunitario all'epoca vigente, che prescriveva la conduzione di un'indagine sulle forze di lavoro nella primavera di ogni anno e dettagliava l'elenco dei quesiti, la Rtfl introdusse nell'ottobre 1992 un nuovo questionario. In quella occasione si scelse di limitare all'indagine di aprile la richiesta delle nuove informazioni sulla situazione nello stesso mese dell'anno precedente. La restrizione è proseguita fino ai nostri giorni. Nella Rcfl i quesiti in esame sono riportati nella sezione I.

²¹ Rispetto alla condizione professionale, un aspetto di ordine generale, valido tanto per la Rtfl quanto per la Rcfl, attiene la cautela con cui interpretare dati retrospettivi che possono scontare un "effetto ricordo" da parte del soggetto intervistato. Nella Rcfl, la sezione I del questionario riporta la medesima articolazione delle modalità di risposta per la condizione autopercepita al momento dell'intervista e per quella un anno prima. Utilizzando i criteri classificatori della Rcfl, e non l'autopercezione, i cambiamenti di condizione professionale possono essere appropriatamente esaminati con i dati di flusso. Come per la Rtfl, anche la nuova indagine incorpora difatti una struttura longitudinale derivante dal sistema di rotazione delle famiglie. Le informazioni raccolte sugli stessi individui in diversi momenti temporali rappresentano pertanto la componente longitudinale dell'indagine. Questa consente di individuare sia il numero delle transizioni in entrata e in uscita dai diversi stati sia le caratteristiche degli individui coinvolti in tali transizioni.

²² Si tratta del quesito che rileva i motivi per cui non si è cercato lavoro pur non avendo un'occupazione (quesito F10), e di quelli che rilevano i motivi della eventuale indisponibilità ad iniziare a lavorare entro due settimane dall'intervista (quesiti F14, F34, F47).

²³ Come è stato richiamato in precedenza, tale innovazione risponde alla necessità di adeguare la definizione di disoccupazione a quella sancita dalla normativa comunitaria.

²⁴ Nel questionario i quesiti sulle azioni di ricerca attiva sono quelli da F16 a F28. Il quesito a risposta multipla è ancora utilizzato per la ricerca di lavoro da parte degli occupati (quesito F6). Tale disparità di trattamento deriva dalla differente rilevanza che l'informazione sull'aver effettuato azioni di ricerca attiva ricopre per gli occupati e per i non occupati. Nel caso degli occupati tale informazione rappresenta un aspetto conoscitivo di contesto, senza peraltro contribuire alla determinazione dell'occupazione. Nel caso dei non occupati costituisce invece un elemento fondamentale per la distinzione tra disoccupati e inattivi. Pertanto, per gli occupati si è scelta la soluzione che comporta la minore molestia statistica.

Ciascun quesito contiene un esplicito riferimento al periodo in cui le azioni devono essere state effettuate²⁵. Inoltre la Rcfl separa e colloca in fondo alla lista i quesiti relativi alle azioni passive, che non prefigurano un'attiva partecipazione al mercato del lavoro e che la Rtfl inseriva indistintamente in un più generale elenco di azioni attive e passive. Seguendo l'approccio generale del questionario, alcune domande vengono rivolte in modo differente a individui che si trovano in situazioni diverse. Tale è il caso delle persone che cercano un lavoro e di coloro che inizieranno un'attività in futuro: ai primi vengono rivolti quesiti relativi alle caratteristiche del lavoro che stanno cercando, mentre dai secondi si rilevano informazioni sull'attività che stanno per iniziare²⁶.

Un accrescimento del patrimonio informativo si delinea poi sul tema dell'istruzione e della formazione. Strettamente legata alla recente rivisitazione operata in sede comunitaria dei concetti e delle variabili di interesse per l'indagine sulle forze di lavoro, la prima innovazione apportata dalla Rcfl riguarda la più evidente separazione tra educazione di tipo formale e attività di formazione²⁷. L'apprendimento al di fuori del sistema di istruzione viene peraltro esteso nella Rcfl ad un insieme di attività: dai seminari alle lezioni private, dalle conferenze all'Università della terza età²⁸. Alla possibilità di indagare i processi di apprendimento e formazione dei soggetti intervistati si accompagna dunque un'ampia disponibilità informativa sulle modalità intraprese per accrescere il bagaglio di conoscenze, distinguendo tra motivi professionali e interessi personali. In tale contesto, la successiva registrazione dell'argomento del corso di formazione e/o di altra attività, secondo la classificazione concordata a livello comunitario, rappresenta un'ulteriore implementazione della Rcfl in grado di monitorare i contenuti dei percorsi formativi.

L'esame dei nuovi contenuti della Rcfl ha finora fatto esclusivo riferimento alle informazioni raccolte sugli individui. In realtà, l'indagine amplia l'offerta informativa anche con riguardo ai dati rilevati sulle relazioni tra i componenti la famiglia intervistata. Pregiudiziale all'esposizione delle peculiarità di tale ampliamento è tuttavia la definizione di famiglia adottata dalla Rcfl. L'unità di rilevazione dell'indagine è definita come famiglia di fatto e non anagrafica. La famiglia di fatto è rappresentata da un insieme di persone coabitanti, legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o da vincoli affettivi. Gli elementi discriminanti per l'individuazione della famiglia di fatto sono la convivenza abituale nonché il legame di parentela e/o affettivo che lega i componenti della famiglia²⁹. In tale contesto, l'ampio dettaglio

²⁵ La ratio di tali innovazioni risiede nel superamento delle inevitabili imprecisioni derivanti da una domanda a risposta multipla. Infatti, sebbene la ricerca di lavoro comporti solitamente una pluralità di azioni, il quesito a risposta multipla può indurre l'intervistato a dichiararne una sola, che può essere la più recente o la più rilevante, con conseguente sottostima delle azioni compiute. Una sovrastima deriva invece dalla possibilità che l'intervistato perda di vista il riferimento temporale, posto solo nella formulazione del quesito, e indichi azioni in realtà svolte in precedenza.

²⁶ All'interno della sezione F del questionario, anche il quesito sulla volontà di lavorare (F12) presenta una doppia formulazione: per chi non cerca lavoro e per chi lo cerca ma non ha fatto alcuna azione nelle quattro settimane precedenti l'intervista.

²⁷ Si veda il Regolamento n. 2104/2002 (Official Journal of the European Community del 29.11.2002). Il questionario della Rcfl riporta i quesiti sull'istruzione e la formazione nella sezione H. Il riferimento che in questa sezione viene fatto alle quattro settimane precedenti l'intervista, già presente nella RTFL e stabilito dalla normativa comunitaria, tende a minimizzare le distorsioni informative dovute a problemi di ricordo da parte dell'intervistato.

²⁸ La Rtfl si limitava a rilevare la partecipazione a corsi di formazione professionale o ad altri corsi.

²⁹ Ciò significa che il legame di parentela non si pone come vincolo esclusivo per l'individuazione della famiglia di fatto. Non sono considerati componenti della famiglia gli ospiti, i collaboratori domestici, gli affittuari di parte dell'abitazione. Attraverso le risposte dei familiari presenti, l'indagine rileva le informazioni dei componenti della famiglia temporaneamente assenti. L'assenza temporanea è definita nei casi di occupazione stagionale o temporanea in altro Comune o all'estero; di servizio di leva o civile sostitutivo; di noviziato religioso; di ricovero in istituti di cura; di detenzione in attesa di giudizio; di affari, turismo, breve cura; di servizio statale all'estero; di missioni di lavoro, compresa la frequenza di corsi di qualificazione o aggiornamento professionale; di imbarco su navi della marina militare o mercantile. Non vengono invece rilevate le informazioni dei componenti che hanno lasciato definitivamente la famiglia, anche se non hanno ancora cambiato la residenza anagrafica.

previsto dalla nuova indagine per la registrazione delle relazioni di parentela rappresenta il più rilevante aspetto innovativo della Rcfl. In confronto alle sei possibilità di classificazione delle relazioni di parentela offerte dalla Rtfl, la nuova indagine ne individua diciassette³⁰. Per tutti gli individui con stato civile di coniugato/a la Rcfl rileva inoltre l'anno di matrimonio, informazione assente nella precedente indagine. Diviene pertanto possibile classificare, come per le indagini Multiscopo, la relazione tra l'intestatario del foglio di famiglia (persona di riferimento) e gli altri componenti, con la possibilità di ricostruire famiglie e nuclei con le relative tipologie³¹. Ne discende un accrescimento della base statistica disponibile per lo studio dei rapporti che intercorrono tra famiglie e mercato del lavoro, informazioni integrabili con quelle di tipo socio-demografico provenienti da altre indagini sulle famiglie.

³⁰ Le relazioni di parentela, registrate nella "Scheda generale" che precede il questionario individuale, sono riportate con la relativa codifica nell'allegato 1 del questionario.

³¹ Il criterio seguito per la classificazione delle diverse tipologie familiari si basa sull'individuazione della presenza o assenza di un (o più) nucleo (i) familiare (i), inteso come legame di parentela e/o affettivo tra due o più persone. Secondo tale criterio, la Rcfl rende possibile la classificazione di 41 tipologie familiari.

Capitolo 3 - La rete di rilevazione

3.1 – Premessa

A partire dalla “prova generale” della nuova indagine, svoltasi tra ottobre e dicembre 2002, si è attivata una rete di rilevatori professionisti direttamente gestita e monitorata dall’Istituto per la realizzazione delle interviste faccia a faccia con tecnica Capi. Tale scelta fortemente innovativa nella storia dell’Istituto è stata motivata dall’esigenza di rafforzare la qualità della rilevazione così come dal bisogno di supportare il processo con un continuo lavoro di monitoraggio e assistenza. Peraltro, la costituzione della rete, la più grande presente in Italia, ha comportato, e comporta per il suo mantenimento, un notevole impegno da parte di diverse strutture a livello centrale e territoriale. La copertura nazionale è garantita da una suddivisione del territorio in 312 aree. In ciascuna area opera un rilevatore. L’ampiezza delle aree, che nella maggior parte dei casi comprende più comuni, comporta che gli intervistatori si spostino all’interno di una zona con diametro massimo di 100 chilometri. Le aree possono comprendere comuni di province diverse della stessa regione¹.

In questo capitolo si riporta una descrizione sintetica dei principali aspetti della nuova rete di rilevazione. Per maggiori approfondimenti si rimanda al volume Istat. *La rete di rilevazione Capi dell’Istat per la conduzione dell’indagine continua sulle Forze di Lavoro*, Roma: Istat 2005 (Metodi e Norme, n. 24)

3.2 - L’assetto organizzativo e la selezione dei rilevatori

Il disegno organizzativo della rete si basa su quattro attori che interagiscono tra loro: il centro di coordinamento, gli Uffici territoriali, i Comuni, gli intervistatori. Il centro di coordinamento è responsabile della progettazione e della gestione complessiva delle attività tecnico-organizzative, giuridico-amministrative e del sistema informatico a supporto delle precedenti attività². Gli Uffici territoriali costituiscono il punto di forza dell’intera organizzazione. In ciascun Ufficio lavorano per l’indagine un referente e uno o più coordinatori. Gli Uffici territoriali, oltre ad interagire con il centro nella selezione, nella gestione dei contratti e nel monitoraggio, supportano quotidianamente gli intervistatori, gestiscono momenti autonomi di formazione nell’ambito di *debriefing* periodici; curano l’assegnazione dei carichi di lavoro; mantengono i rapporti con i Comuni nella fase di estrazione e trasmissione dei nominativi delle famiglie del campione. I Comuni, già coinvolti nella fase di estrazione del campione, aiutano se necessario gli intervistatori durante la loro attività fornendo indicazioni per il reperimento delle famiglie e intervenendo nei casi in cui si riscontrano difficoltà nel farsi concedere l’intervista³.

Nel processo di produzione gli intervistatori svolgono un ruolo centrale. Le attività assegnate comprendono: la partecipazione alle riunioni di coordinamento mensili presso gli Uffici territoriali; il collegamento telematico settimanale attraverso numero verde per la

Di Gianlorenzo Bagatta

¹ Nei comuni metropolitani le aree ricoprono porzioni del territorio comunale. Solo per il Piemonte e la Valle d’Aosta possono essere assegnate al rilevatore interviste che comprendono comuni di entrambe le regioni.

² Nel centro di coordinamento sono rappresentati il Servizio formazione e lavoro, gli Uffici del personale e della ragioneria, le strutture informatiche.

³ I problemi sorgono dall’ampiezza delle aree che porta gli intervistatori ad operare frequentemente in comuni a loro sconosciuti.

ricezione del questionario elettronico, dell'agenda dei contatti, dei nominativi delle famiglie da intervistare e di tutte le comunicazioni necessarie alla conduzione delle interviste; i contatti preliminari con le famiglie e la successiva effettuazione delle interviste; il collegamento telematico giornaliero attraverso numero verde per la trasmissione degli esiti dei contatti e dei dati relativi alle interviste.

Il processo di selezione dei rilevatori è stato rivolto a garantire l'uniformità sia nelle modalità di reclutamento sia nelle professionalità individuate⁴. La valutazione dei candidati, oltre che su alcuni pre-requisiti, si basa anche su altre caratteristiche preferenziali. Come pre-requisiti si tiene conto dell'età, della capacità nell'utilizzo del personal computer, della disponibilità alla mobilità. Titoli preferenziali sono: diploma di scuola media secondaria superiore; precedenti esperienze di interviste realizzate con compilazione di questionario cartaceo e/o elettronico; precedenti esperienze in indagini statistiche. La scelta tiene successivamente conto delle capacità comunicative-relazionali, delle potenzialità al *problem-solving*, dell'attitudine gestionale-organizzativa. Per minimizzare il rischio di abbandono, si valutano inoltre la motivazione verso il tipo di attività, la disponibilità di tempo per poter svolgere le interviste con continuità e soprattutto nelle ore serali, la possibilità di lavorare tutto l'anno⁵.

Dal punto di vista operativo la selezione è stata, e viene tuttora realizzata, in quattro fasi: la pubblicizzazione del reclutamento; la preselezione attraverso lo *screening* dei curricula pervenuti; il precolloquio individuale insieme a test informatici e psicoattitudinali; la selezione finale con una presentazione dettagliata delle modalità del lavoro e un ultimo colloquio. Sulla base delle indicazioni fornite dall'Istat, una società di reclutamento e selezione svolge, nelle singole filiali distribuite sul territorio, le prime tre fasi. Nella fase finale, invece, vi è il diretto intervento del personale dell'Istituto. Le modalità di selezione sono state peraltro testate dall'Istat e modificate nel corso del tempo. Conclusa la costituzione della rete, il mantenimento è caratterizzato da azioni di reclutamento mirate a gestire il naturale turn-over dei rilevatori. L'eventualità che un elevato numero di intervistatori abbandonino la rete in tempi ristretti metterebbe a rischio la stabilità della rilevazione e comporterebbe costi elevati per la selezione, la formazione, la gestione amministrativa dei nuovi rilevatori. Nel primo anno e mezzo di vita la rete ha visto rinnovarsi circa un terzo dei suoi componenti. Il tasso di turn-over complessivo risulta pari a circa il 69 per cento; quello di uscita si attesta nell'intero periodo intorno al 34 per cento e nella media mensile al 2 per cento⁶. Una concentrazione di abbandoni si è verificata nel periodo iniziale di vita della rete ed è stata indotta da un processo di selezione volto a sostituire con rapidità soprattutto le persone che non garantivano continuità nel lavoro. Con il passare del tempo, si sono andate affinando le procedure di selezione, prestando maggiore attenzione agli aspetti motivazionali e caratteriali del candidato, più che a requisiti di tipo oggettivo come l'età o il titolo di studio. Tale attività si è riflessa in una riduzione dei tassi di uscita (che si sono attestati intorno al 14 per cento annuo) rispetto a quelli registrati nella fase iniziale⁷.

⁴ Il processo di selezione deve d'altro canto essere tempestivo nel ricambio dei rilevatori.

⁵ In generale, la permanenza nella rete è più elevata quando il candidato non presenta un profilo troppo elevato con forti ambizioni di affermazione professionale e si trova in una situazione personale e/o familiare compatibile con gli impegni richiesti dall'attività di intervistatore.

⁶ Il tasso di turn-over è calcolato come rapporto percentuale fra la somma degli intervistatori usciti e entrati dal 1° novembre 2002 al 31 marzo 2004 e il numero medio di intervistatori attivi nello stesso periodo. Il tasso di uscita è calcolato come rapporto percentuale tra il numero di intervistatori che hanno cessato l'attività fra il 1° novembre 2002 e il 31 marzo 2004 e il numero medio di intervistatori attivi nello stesso periodo.

⁷ Sebbene le uscite si siano progressivamente ridotto, esse presentano due picchi in corrispondenza dei mesi di scadenza dei contratti. La scadenza del contratto è un periodo in cui si concentrano le interruzioni dei rapporti di lavoro, sia che queste avvengano per volontà dell'Istituto sia che derivino da una scelta del rilevatore. Nel primo caso l'Istituto, qualora non pienamente soddisfatto dei risultati del lavoro svolto, preferisce non rinnovare il contratto; nel secondo, il rilevatore in generale rassegna le dimissioni contestualmente alla naturale scadenza del contratto.

Dalla tavola 2.1, che riporta le principali caratteristiche degli intervistatori, emerge in primo luogo la maggiore presenza tra i rilevatori della componente femminile, in particolare nel Centro. Il 55 per cento circa ha un'età compresa tra i 30 e i 44 anni, mentre nel Nord-est l'età media è più alta. Il 61 per cento dei rilevatori è in possesso di un diploma, ma nel Mezzogiorno l'incidenza dei laureati è significativamente più elevata della media nazionale. Al momento dell'assunzione, la maggioranza dei rilevatori era alla ricerca di un'occupazione pur avendo avuto precedenti esperienze. Una parte della rete è comunque impegnata in altre attività, lavorative o di studio.

Nel corso del tempo si è tenuta sotto controllo la possibilità di conciliare il lavoro di intervistatore con altri impegni, anche alla luce degli indicatori di qualità emersi⁸.

Prospetto 3.1 - Principali caratteristiche degli intervistatori (composizioni percentuali)

CARATTERISTICHE (a)	Ripartizione geografica					
	Nord Ovest	Nord Est	Centro	Sud	Isole	Italia
Sesso						
Femmine	57,3	53,2	68,0	59,5	54,1	58,4
Maschi	42,7	46,8	32,0	40,5	45,9	41,6
Classe di età						
18-24	4,9	9,7	4,0	7,6	8,1	6,8
25-29	22,0	11,3	28,0	27,8	40,6	24,5
30-34	25,6	19,4	16,0	36,7	29,7	26,1
35-44	31,6	25,8	36,0	26,6	21,6	28,7
45-oltre	15,9	33,8	16,0	1,3	.	13,9
Titolo di studio						
Licenza media	6,1	8,1	4,0	-	-	3,9
Diploma	61,0	64,5	70,0	55,7	54,1	61,0
Laurea	32,9	27,4	26,0	44,3	45,9	35,1
Condizione (b)						
Già occupato	37,8	29,0	22,0	24,1	16,2	27,4
In cerca con esperienza	36,6	37,1	48,0	41,7	43,3	40,6
In cerca senza esperienza	-	1,6	8,0	12,7	10,8	6,1
Casalinga	4,9	11,3	6,0	2,5	-	5,2
Studente	12,2	9,7	12,0	7,6	16,2	11,0
Ritirato dal lavoro	1,2	6,5	-	-	-	1,6
Altro	7,3	4,8	4,0	11,4	13,5	8,1
Totale	100,0	100,0	100,0	100,0	100,0	100,0

(a) Al 31 marzo 2004; (b) Condizione prevalente dichiarata nella fase di selezione.

3.3 - La formazione

La qualità dei dati raccolti è strettamente legata al processo di armonizzazione delle conoscenze, competenze e capacità relazionali della rete⁹. Per questo motivo si è attuato un progetto formativo che coinvolge l'intera struttura che gestisce l'indagine, al cui interno i rilevatori costituiscono una parte assolutamente rilevante ma non esaustiva. In questa struttura

⁸ Aspetti peculiari del lavoro del rilevatore sono la flessibilità e l'autonomia nella gestione dei tempi, che lo rendono conciliabile con altri tipi di impegni. Va sottolineato che i responsabili della selezione hanno comunque attentamente valutato ogni singolo curriculum, scegliendo fra i candidati già occupati solo chi svolgeva un'attività compatibile con quella di rilevatore.

⁹ Le indagini Istat sulla vittimizzazione dei cittadini del 1997, sulle condizioni di salute del 1999-2000 e sull'uso del tempo del 2002-2003 avevano sostanzialmente delineato i contenuti e i metodi degli interventi formativi nonché i principali aspetti di un ideale profilo professionale del rilevatore. Le esperienze acquisite da queste indagini sono state particolarmente utili per la selezione e la formazione degli intervistatori della Rcf.

reticolare i flussi formativi non sono unidirezionali - dal centro alla periferia seguendo una logica piramidale - e concentrati in una sola occasione temporale. Essi si caratterizzano sia per l'alto livello di circolarità assicurato dai continui *feed-back* tra Servizio formazione e lavoro, Uffici territoriali e rilevatori, sia per la continuità delle occasioni formative. Alla formazione centralizzata e decentralizzata dei rilevatori si accompagna difatti quella dell'insieme della rete. La progettazione di un modulo di formazione a distanza affronta poi il reiterato e oneroso aspetto della formazione dei rilevatori che sostituiscono quelli che abbandonano la rete.

Inizialmente si è partiti con una formazione in aula di tipo centralizzato, rivolta a fornire ai rilevatori neo-assunti gli strumenti necessari per iniziare ad effettuare le interviste. La scelta è derivata dalla necessità di investire al meglio e in modo più possibile uniforme sui 311 rilevatori appena reclutati¹⁰. L'aspetto caratterizzante è la durata del corso: un'intera settimana lavorativa. In questo periodo vengono illustrati gli obiettivi e i contenuti dell'indagine, la struttura del questionario, gli aspetti metodologici, la normativa a tutela della privacy. Di particolare interesse è la parte del modulo formativo, gestita direttamente da psicologi, mirata a sviluppare specifiche competenze comunicative. Vengono inoltre svolte esercitazioni pratiche che simulano le situazioni di intervista e permettono ai rilevatori di prendere confidenza con la dotazione informatica hardware e software messa a loro disposizione.

La formazione in aula di tipo decentralizzato si è svolta e si svolge attraverso *debriefing* organizzati con cadenza pressoché mensile dagli Uffici territoriali. In queste occasioni vengono sistematicamente affrontati gli eventuali problemi relativi all'organizzazione del lavoro (gestione dei carichi, durata delle interviste, reperibilità delle famiglie, rapporto con i Comuni), alle tecniche di approccio con le famiglie e alle strategie comunicative, al contenimento delle cadute/rifiuti, ad aspetti problematici del questionario. In realtà, i *debriefing* rappresentano il momento di sintesi di un quotidiano percorso formativo¹¹. In questo quadro, va peraltro sottolineato che anche il Servizio formazione e lavoro collabora alla gestione di problemi la cui soluzione, comportando riflessioni e decisioni a livello metodologico, il più delle volte si configura come ulteriore input formativo per il Servizio stesso, per i colleghi degli Uffici territoriali e per i rilevatori.

Un modulo specifico di formazione a distanza è stato progettato e realizzato per garantire la formazione dei nuovi rilevatori assunti a livello territoriale. Modellato sull'architettura a tre livelli dell'intera rete, il modulo di formazione a distanza si caratterizza come esperienza ibrida di *e-learning*, nell'ambito della quale convivono momenti di formazione faccia a faccia, utilizzo di elementi multimediali e ricorso alle trasmissioni via internet. Scendendo nel dettaglio, per la formazione a distanza è previsto che all'intervistatore neo-assunto, durante una prima giornata di formazione faccia a faccia svolta nell'Ufficio territoriale di competenza, venga consegnato un personal computer con una presentazione multimediale dell'indagine e dei contenuti del questionario. Il corso, da seguire a casa nell'arco di circa cinque giorni, è organizzato per moduli consecutivi. Ogni modulo è propedeutico al successivo e richiede per abilitarlo una serie di risposte a un test di chiusura. I risultati del test vengono trasmessi via internet al referente territoriale e al Servizio formazione e lavoro. In tal modo si realizza un tutoraggio continuo e condiviso del nuovo rilevatore. Terminata l'autoformazione, è prevista un'ultima fase di almeno due giorni di formazione faccia a faccia presso l'Ufficio territoriale. In questa circostanza il neo-assunto viene seguito dal referente e dal collaboratore di rete per la verifica e l'integrazione della formazione acquisita tramite il pacchetto multimediale.

¹⁰ Tale formazione ha interessato successivamente i gruppi dei nuovi rilevatori che sostituiscono quelli che abbandonano la rete. Al riguardo, la formazione a distanza ha successivamente sostituito quella svolta in aula.

¹¹ Tutti i contatti telefonici e le mail tra rilevatori e gli Uffici territoriali si configurano come veri e propri interventi formativi in itinere in quanto azioni utili ad ottimizzare le performance di intervista e, quindi, la qualità dei dati raccolti.

In sintesi, la formazione è mirata ad aumentare le competenze professionali in termini di capacità di raccogliere informazioni qualitativamente ineccepibili. Per ottenere questo risultato è necessario, come visto, operare su più dimensioni per far sì che gli intervistatori sappiano gestire efficacemente il *setting* dell'intervista. Il lavoro svolto mostra che è necessario investire parallelamente sulle dimensioni cognitive, metodologiche, psicologiche e comportamentali. Solo con un ricco bagaglio di strumenti il rilevatore professionista è in grado di mediare tra le rigorose indicazioni ricevute e le concrete situazioni che affronta quotidianamente nel corso del suo lavoro. In altri termini, il rilevatore non solo deve conoscere i contenuti dell'indagine e gli aspetti giuridici connessi alla rilevazione ma deve anche sapere gestire le strategie organizzative, la strumentazione informatica e le relazioni interpersonali nei rapporti con le famiglie¹². Tutto ciò ha comportato che nel percorso formativo fossero coinvolti metodologi, statistici, esperti del mercato del lavoro, informatici, amministrativi e psicologi. In definitiva, l'investimento nella formazione continua dei rilevatori, sperimentato in parte nell'ambito delle indagini Multiscopo, ha raggiunto oggi livelli qualitativi particolarmente elevati.

Nell'ambito del progetto formativo non è stata trascurata l'importanza della formazione dei formatori. Trasmettere le informazioni necessarie per effettuare la rilevazione richiede una consolidata esperienza nello svolgere le operazioni sul campo, abilità tecnica e conoscenze specifiche in campo didattico. Per questo motivo si è cercato di costruire nuove professionalità nel campo della formazione. In particolare, i referenti di indagine e i coordinatori di rete sono chiamati a gestire la formazione continua, la formazione d'aula e i momenti integrativi della formazione a distanza. È evidente, quindi, la necessità di avere competenze sia su quello che deve fare il rilevatore (approccio alla famiglia, gestione dell'intervista, controllo emotivo, organizzazione del tempo) sia su specifiche tecniche da utilizzare in quanto formatori (gestione di aula, dinamiche comunicative, gestione dei gruppi).

La formazione dei formatori riguarda anche il personale del Servizio formazione e lavoro. Tutti i *feed-back* che si hanno dai momenti di formazione rappresentano spunti per l'affinamento delle tecniche di indagine e per la soluzione di aspetti metodologici¹³. In questo senso è corretto dire che la formazione diviene un luogo privilegiato di unione tra metodologia statistica e metodologia di indagine¹⁴.

Il progetto formativo non può dirsi concluso. Accanto all'ottimizzazione degli strumenti finora utilizzati, sono in fase di progettazione ulteriori iniziative volte a delineare in modo ancora più preciso il profilo professionale dell'intervistatore, a migliorare il processo di selezione, ad investire ulteriormente sulla formazione dei formatori.

¹² I rilevatori devono saper gestire sia i momenti relativamente semplici del contatto e della presentazione degli obiettivi dell'indagine sia le situazioni difficili che possono verificarsi con famiglie particolari. Inoltre i rilevatori devono possedere un adeguato autocontrollo oltre che un buon livello di autostima, che può essere messo in crisi in momenti particolari del lavoro.

¹³ Il primo investimento che si fa sul rilevatore è quello di convincerlo che se le soluzioni individuali, spesso metodologicamente scorrette, adottate per far fronte a situazioni reali non previste in fase di progettazione vengono condivise come patrimonio comune, si attiva un processo virtuoso in cui il metodologo, il ricercatore e il rilevatore, trovando una soluzione corretta al problema specifico, contribuiscono al progressivo miglioramento della qualità del dato.

¹⁴ Gli interventi formativi sinteticamente descritti si sono avvalsi del sistema informativo-informatico a supporto dell'indagine, sfruttando le potenzialità offerte dalla rete telematica per condividere le informazioni utili.

Capitolo 4 - Il disegno campionario della nuova indagine e la fase di estrazione

4.1 - Premessa

4.1.1 - Evoluzione storica del disegno campionario

L'indagine sulle forze di lavoro è stata la prima indagine campionaria effettuata dall'Istituto Nazionale di Statistica. La prima rilevazione fu condotta nel mese di settembre del 1952 sulla base di uno studio del Prof. Pompilj, anche se solo nell'aprile del 1959 essa ha assunto quel carattere di sistematicità che ancora oggi la distingue e che prevede una rilevazione a *cadenza trimestrale* con rotazione del campione di famiglie. A partire da tale data l'indagine ha subito molteplici modifiche che ne hanno mutato profondamente le caratteristiche. Alcune variazioni hanno interessato solamente aspetti specifici (quali ad esempio il questionario oppure la procedura di riporto dei dati all'universo) altre, invece, costituiscono veri e propri ridisegni in quanto hanno mutato completamente l'impianto complessivo della rilevazione.

L'indagine del 1959 era condotta mediante intervista diretta e la struttura originaria del campione era già disegnata a due stadi (comuni-famiglie), con stratificazione dei comuni. La stratificazione iniziale del campione prevedeva, a differenza del disegno attuale, la ripartizione dei comuni in due grandi gruppi, i comuni capoluogo di provincia e altri comuni con almeno 20.000 abitanti (gruppo A) e i comuni con popolazione inferiore a 20.000 abitanti (gruppo B). I comuni del gruppo B erano ripartiti per zona altimetrica (montagna, collina e pianura) e per attività economica prevalente (agricoltura, industria e altre attività). I comuni del gruppo A erano inclusi con certezza nel campione; i comuni del gruppo B, invece, venivano selezionati con probabilità proporzionali al loro peso demografico. Circa un terzo dei comuni del gruppo B venivano ruotati all'inizio di ogni ciclo di rilevazione (ossia il momento di selezione delle famiglie dalle anagrafi comunali), che coincideva con la rilevazione di luglio. Il campione di famiglie veniva scelto mediante selezione sistematica dalle anagrafi e seguiva lo schema di rotazione trimestrale del tipo (2-2-2), così come previsto anche dal disegno attuale; tale schema di rotazione sarà illustrato dettagliatamente nel paragrafo 4.4.1. I principali inconvenienti che questo disegno comportava riguardavano una maggiore variabilità nei valori delle stime prodotte in occasione della rilevazione di luglio di ciascun anno, causata dalla rotazione dei comuni; inoltre, la stratificazione dei comuni era poco efficiente in termini di riduzione della variabilità campionaria e poteva essere modificata solo con cadenza decennale in base ai risultati aggiornati del nuovo censimento generale della popolazione.

Fino agli inizi degli anni ottanta la numerosità campionaria trimestrale era di circa 1400 comuni e 90 mila famiglie. A partire da quest'ultima data e fino al 1990 molte regioni, che in quel periodo avevano istituito degli *osservatori del lavoro*, iniziarono a richiedere all'Istat ampliamenti dei campioni regionali per poter ottenere stime attendibili a livello provinciale per i più importanti aggregati di interesse. Questi ampliamenti portarono la dimensione del campione del 1990 a circa 2 mila comuni e 140 mila famiglie a trimestre. Differenti studi hanno mostrato che l'ampliamento generalizzato del campione, a fronte di una riduzione della variabilità delle stime di interesse, ha prodotto nel tempo un aumento degli errori non campionari a causa dell'effettiva impossibilità di tenere sotto controllo, da parte dell'Istat, un numero così elevato di

Di Claudia De Vitiis, Loredana Di Consiglio, Stefano Falorsi (parr.4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7), Rita Ranaldi (parr.4.8)

comuni, comportando per questa via un incremento dell'*errore totale*. Il disegno campionario introdotto nel luglio 1990 era basato sulla medesima struttura a due stadi, anche se furono introdotte rilevanti innovazioni ancora oggi presenti nel disegno d'indagine.

La prima innovazione riguardò la definizione della numerosità campionaria e la sua allocazione tra i domini territoriali pianificati (province, regioni e ripartizioni) scelta in base al criterio del rispetto di prefissati livelli attesi negli errori delle stime principali riferiti ai differenti domini territoriali. In particolare, la nuova numerosità campionaria, di primo e di secondo stadio, fu fissata a 1.237 comuni campione e 70 mila famiglie. Dall'aprile 1992 il campione ha subito un'ulteriore modifica che ha portato alla numerosità di 1.327 comuni e 73 mila famiglie (successivamente innalzata a 76 mila nel 1995 in occasione dell'introduzione delle nuove province).

Un'altra importante innovazione introdotta con il ridisegno del 1990 ha riguardato l'introduzione di un criterio di stratificazione dei comuni sulla base della dimensione demografica degli stessi. Tale modifica, oltre a introdurre un miglioramento dell'efficienza delle stime, consentiva di aggiornare annualmente la stratificazione delle unità di primo stadio in base ai nuovi dati comunali anagrafici. In particolare, con questo disegno venne introdotta la struttura di stratificazione del campione di comuni ancora oggi utilizzata. Occorre ricordare, infine, che con il ridisegno del 1990 venne abolita la rotazione di un terzo dei comuni campione in occasione del mese di luglio, per eliminare il salto nella serie storica delle stime trimestrali, causata dalla rotazione dei comuni.

Altre importanti modifiche hanno riguardato lo stimatore utilizzato. A partire dalla rilevazione di gennaio 1991 venne introdotto uno stimatore del rapporto combinato e post-stratificato per sesso e classi di età in sostituzione dello stimatore del rapporto separato post-stratificato per sesso che era stato adottato fino a ottobre 1990. Infine, a partire dall'anno 1997, è stato adottato uno stimatore di ponderazione vincolata che utilizzava la stessa definizione dei post-strati dello stimatore del rapporto combinato ma che consentiva di produrre stime coerenti riferite alle famiglie e agli individui a esse appartenenti.

L'indagine corrente produce i dati a partire dal primo trimestre dell'anno 2004 ed è il frutto di un percorso di studio iniziato nel 1999 e conclusosi con l'introduzione graduale dei gruppi di rotazione del campione dal primo trimestre 2003. Quest'ultimo ridisegno dell'indagine ha comportato una profonda modifica di importanti aspetti del disegno campionario per rispettare i requisiti imposti dal regolamento europeo.

Nel paragrafo seguente si richiamano i principali obiettivi conoscitivi che costituiscono il contesto di base per la definizione del disegno campionario dell'indagine forze lavoro; si riportano successivamente gli aspetti salienti della normativa europea che hanno costituito un elemento vincolante per la caratterizzazione del nuovo campione.

4.1.2 - Obiettivi conoscitivi

L'indagine ha la finalità principale di ottenere informazioni sulla situazione lavorativa, sulla ricerca di lavoro e sugli atteggiamenti verso il mercato del lavoro della popolazione in età lavorativa.

La *popolazione di interesse* è costituita da tutti i componenti delle famiglie residenti in Italia, anche se temporaneamente emigrati all'estero, mentre esclude i membri permanenti delle *convivenze* (ospizi, istituti religiosi, caserme, eccetera). La famiglia è intesa come *famiglia di fatto*, ossia come un insieme di persone legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o da vincoli affettivi, coabitanti e aventi dimora abituale nello stesso comune; nel caso in cui la famiglia selezionata coabiti con altre famiglie viene intervistata soltanto quella estratta.

I parametri di interesse dell'indagine sono espressi in termini di numero di individui che possiedono un dato attributo z e riguardano i seguenti aggregati di popolazione: forze di lavoro¹, occupati, persone in cerca di occupazione.

Relativamente a ciascuno di tali aggregati, i principali parametri oggetto di stima sono il numero di individui che presentano l'attributo z nel trimestre t e nell'anno a (ottenuto come media dei quattro trimestri dell'anno a) e la differenza tra il numero di individui che presentano l'attributo z nel trimestre t e il numero di individui che presentano il medesimo attributo in un trimestre precedente t' .

I *domini territoriali di studio* sono l'intero territorio nazionale, le ripartizioni geografiche e le regioni; a partire dal 1993, vengono considerate anche le province unicamente per la costruzione di stime medie annue.

Le stime dell'indagine vengono pubblicate anche a livello di sottopopolazioni differenti dai domini territoriali pianificati, quali, ad esempio, il sesso e le classi di età o il titolo di studio. La numerosità campionaria di questi domini non pianificati non può essere fissata a priori e risulta pertanto aleatoria.

4.1.3 - I requisiti di armonizzazione europea

Gli obiettivi che hanno guidato la revisione dell'indagine derivano sostanzialmente dall'esigenza di uniformarla ai requisiti dell'Unione europea (contenuti nel regolamento n.577/98 del 9 marzo 1998 pubblicato nella Gazzetta ufficiale della Unione europea). Ciò ha comportato una profonda rivisitazione del disegno di rilevazione per quanto riguarda sia la definizione del questionario sia la strategia di campionamento.

I due principali aspetti del Regolamento che hanno avuto delle implicazioni sulla progettazione del disegno campionario dell'indagine sono vincoli sulla periodicità della rilevazione e sulla precisione delle stime. Per la periodicità dell'indagine, il regolamento stabilisce quanto segue :

- “L'indagine è un'indagine continua che fornisce risultati trimestrali e annuali”;
- “Le informazioni raccolte nell'indagine si riferiscono generalmente alla situazione nel corso della settimana di calendario (...) che precede l'intervista, detta *settimana di riferimento*”
- “le settimane di riferimento sono ripartite uniformemente su tutto l'anno”;
- “tra la settimana di riferimento e la data dell'intervista non possono intercorrere più di cinque settimane”;
- “i trimestri e gli anni di riferimento sono rispettivamente insiemi di 13 e 52 settimane consecutive”.

Per quanto riguarda invece la precisione delle stime i vincoli imposti sono i seguenti:

- “Per un gruppo di disoccupati (...) che rappresenta il 5 per cento della popolazione in età lavorativa, la *deviazione relativa standard* per la stima delle medie annuali (...) non deve superare l'8 per cento a livello Nuts II², della popolazione parziale in questione”. Per le regioni con meno di 300.000 abitanti non è necessario che questo requisito sia soddisfatto.

¹ L'indagine, riferendosi a tutti i componenti delle famiglie estratte, consente di rilevare altresì le *non forze lavoro* che comprendono – oltre ai bambini e ragazzi fino a tredici anni, le persone di quattordici anni e più che dichiarano di trovarsi in una delle seguenti condizioni non professionali: casalinga, studente, ritirato dal lavoro, inabile, servizio di leva, altra condizione non professionale, e che non lavorano o non cercano lavoro.

² Per l'Italia con Nuts II si intende il livello regionale.

- “per popolazioni di dimensioni pari al 5% della popolazione in età lavorativa, la deviazione relativa standard per la stima delle variazioni tra due trimestri successivi, a livello nazionale, non deve superare il 2% della popolazione parziale in questione”.

Si fa presente che la valutazione degli errori campionari attesi delle stime prodotte dall'indagine ha evidenziato che la numerosità campionaria che veniva utilizzata dall'indagine precedente era già sufficiente a soddisfare i vincoli europei sulla precisione campionaria delle stime.

Poiché dal punto di vista della periodicità la struttura dell'indagine precedente non rispettava i requisiti del Regolamento, gli aspetti più rilevanti del ridisegno sono stati quelli connessi all'introduzione di una rilevazione di tipo continuo. La precedente indagine era, infatti, un'indagine ripetuta a data fissa che veniva svolta in una sola settimana del trimestre e tutto l'impianto organizzativo era strutturato per operare in un intervallo di tempo limitato e concentrato. L'esigenza di passare a un processo d'indagine continuo nel tempo ha comportato, pertanto, una profonda ristrutturazione della strategia campionaria.

A parte i vincoli imposti dal Regolamento, inoltre, il ridisegno ha costituito un'opportunità per ridefinire alcuni degli obiettivi dell'indagine in termini di precisione delle stime d'interesse riferite ai diversi domini territoriali (si veda paragrafo 4.6).

4.2 - Lo schema generale della rilevazione e del disegno campionario

Con riferimento a ciascun trimestre la rilevazione è di tipo continuo e viene condotta tutte le settimane. Ogni famiglia campione viene intervistata una sola volta in una specifica settimana. Questo tipo di articolazione temporale dell'indagine costituisce una rilevante innovazione rispetto al disegno di indagine precedente, in base al quale ciascuna rilevazione trimestrale veniva svolta in una sola settimana del trimestre (le quattro settimane di rilevazione di ogni anno cadevano, rispettivamente, nei mesi di gennaio, aprile, luglio e ottobre).

I campioni relativi a trimestri differenti sono parzialmente sovrapposti in base a uno schema di rotazione secondo cui una famiglia è inclusa nel campione per due rilevazioni successive e, dopo una pausa di due trimestri, viene reinserita nel campione per altre due rilevazioni. Ciò comporta una sovrapposizione del campione teorico del 50 per cento a un trimestre di distanza, del 25% a tre trimestri, del 50 per cento a quattro trimestri e del 25 per cento a cinque trimestri.

La sovrapposizione dei campioni trimestrali risponde all'esigenza di ridurre le fluttuazioni delle stime di livello riferite a trimestri differenti. Inoltre, la specifica tecnica di rotazione permette di ridurre la stima delle variazioni nette tra trimestri consecutivi e tra trimestri a un anno di distanza.

Un approfondimento della tecnica di rotazione è riportato nel paragrafo 4.4.1.

La prima intervista di ogni famiglia campione è di tipo diretto, condotta mediante tecnica Capi (intervista faccia a faccia assistita da computer) mentre le tre successive sono da effettuarsi, in linea teorica, mediante intervista telefonica con tecnica Cati (per le eccezioni alla regola generale vedi Capitolo 6). Nonostante le interviste successive alla prima siano di tipo telefonico, poiché la prima intervista è di tipo diretto, si è reso comunque necessario mantenere un certo grado di clusterizzazione territoriale del campione per contenere i costi legati all'organizzazione sul campo della rilevazione e in particolare quelli connessi agli spostamenti dei rilevatori tra i comuni campione.

Il disegno di campionamento adottato per ciascuna rilevazione trimestrale si basa su una stratificazione temporale di tipo mensile; infatti, il campione trimestrale di unità finali di

campionamento (le famiglie) viene suddiviso in tre gruppi distinti ciascuno dei quali viene assegnato casualmente a un mese del trimestre in modo tale che ciascuno dei tre gruppi costituisca un campione rappresentativo della popolazione di riferimento nel mese considerato.

Per un approfondimento sulla scelta della stratificazione temporale entro il trimestre si rimanda al paragrafo 4.4.3.

Il disegno di campionamento adottato in ciascun trimestre è un disegno a due stadi con stratificazione delle unità di primo stadio; le unità di primo stadio sono i comuni e le unità di secondo stadio sono le famiglie. All'interno di ciascuna provincia i comuni sono suddivisi in due sottoinsiemi: i comuni la cui dimensione demografica è superiore a una prefissata soglia detti comuni *auto rappresentativi* (Ar); i rimanenti comuni detti *non auto rappresentativi* (Nar). Ciascun comune Ar costituisce strato a sé stante e viene incluso con certezza nel campione; i comuni Nar, invece, vengono stratificati sulla base della dimensione demografica e da ogni strato così definito viene estratto un comune con probabilità proporzionale alla dimensione demografica. Dalla lista anagrafica di ogni comune campione viene selezionato, mediante scelta sistematica, un campione di famiglie; tutti gli individui appartenenti alle famiglie estratte vengono intervistati.

A differenza delle famiglie, che ruotano secondo lo schema sopra introdotto, i comuni campione rimangono sempre gli stessi nel tempo.

Complessivamente, in ciascuna rilevazione trimestrale sono coinvolti circa 1.246 comuni campione di cui 346 Ar; le famiglie previste sono circa 77 mila per un totale di circa 200 mila individui.

4.3 - Contesto teorico di riferimento

Consideriamo una popolazione \mathcal{P} composta da N individui, che per semplicità si suppone costante nel tempo e definiamo, con riferimento alla generica caratteristica oggetto di indagine z , la variabile dicotomica y che indica, per ogni settimana s , se il generico individuo della popolazione presenta o meno tale caratteristica; pertanto y_{qs} risulta pari a 1 se il q -esimo individuo della popolazione possiede la caratteristica z nella settimana s e pari a 0 altrimenti. Ad esempio, se la caratteristica d'interesse z si riferisce alla condizione "occupato" o "non occupato", la variabile dicotomica riferita all'individuo q nella settimana s sarà pari a 1 nel caso in cui in tale settimana egli risulta "occupato", mentre sarà pari a 0 altrimenti.

Il seguente quadro simbolico rappresenta l'insieme delle informazioni teoricamente acquisibili nel trimestre su tutti gli elementi della popolazione:

Prospetto 4.1. Schema della popolazione d'interesse nel tempo

	1	...	s	...	S_t	
1	y_{11}	...	y_{1s}	...	y_{1S_t}	$Y_{1.}$
...
q	y_{q1}	...	y_{qs}	...	y_{qS_t}	$Y_{q.}$
...
N	y_{N1}	...	y_{Ns}	...	y_{NS_t}	$Y_{N.}$
	$Y_{.1}$...	$Y_{.s}$...	$Y_{.S_t}$	Y_t

In tale prospetto si è indicato con t l'indice di trimestre e con S_t il numero (sempre pari a 13) di settimane del trimestre; inoltre, $Y_{.s}$ (per semplicità nel seguito Y_s) denota il numero di individui che presentano la caratteristica d'interesse nella settimana s e Y_q il numero di settimane in cui l'individuo q presenta la caratteristica z .

4.3.1 - Parametri di interesse dell'indagine continua

Sulla popolazione \mathcal{P} i principali parametri di livello riferiti al generico dominio territoriale di interesse (provincia, regione, ripartizione o intero territorio nazionale) sono:

- la frequenza assoluta e relativa di individui che presentano la caratteristica z nella settimana s ($s = 1, \dots, S_t$):

$$Y_s = \sum_{q=1}^N y_{qs} \qquad \bar{Y}_s = \frac{1}{N} \sum_{q=1}^N y_{qs};$$

(1)

- la frequenza media trimestrale, assoluta e relativa, di individui che presentano la caratteristica z :

$$Y_t = \frac{1}{S_t} \sum_{s=1}^{S_t} Y_s \qquad \bar{Y}_t = \frac{1}{S_t \cdot N} \sum_{s=1}^{S_t} Y_s;$$

(2)

- la frequenza media annuale, assoluta e relativa, di individui che presentano la caratteristica z :

$$Y_a = \frac{1}{4} \sum_{t=1}^4 Y_t \qquad \bar{Y}_a = \frac{1}{4} \sum_{t=1}^4 \bar{Y}_t.$$

(3)

Altri parametri d'interesse sono rappresentati dalle variazioni nel tempo, quali la differenza (o variazione netta) tra frequenze medie trimestrali tra due trimestri consecutivi

$${}_1\Delta_t = Y_t - Y_{t-1},$$

(4)

o tra due trimestri a distanza di un anno:

$${}_4\Delta_t = Y_t - Y_{t-4}.$$

(5)

Inoltre, sono di interesse anche parametri riferiti al mese, quali:

- la frequenza media mensile, assoluta e relativa, di individui che presentano la caratteristica z :

$$Y_m = \frac{1}{S_m} \sum_{s=1}^{S_m} Y_s \qquad \bar{Y}_m = \frac{1}{S_m \cdot N} \sum_{s=1}^{S_m} Y_s,$$

(6)

avendo indicato con m l'indice di mese e con S_m il numero di settimane del mese³ m ;

³ Nel caso in cui non sia possibile assegnare univocamente una settimana al mese m , poiché alcuni giorni della settimana appartengono a un mese e i rimanenti appartengono a un mese contiguo, l'assegnazione viene definita sulla base di un criterio di prevalenza dei giorni ($S_m=5$ o a 4 a seconda del calendario).

- la differenza tra il numero medio di individui che presentano la caratteristica z nel mese m e il numero medio di individui che presentano la medesima caratteristica nel mese $m-1$:

$${}_1\Delta_m = Y_m - Y_{m-1};$$

(7)

- la differenza tra il numero medio di individui che presentano la caratteristica z nel mese m e il numero medio di individui che presentano la medesima caratteristica nel mese $m-12$ (stesso mese dell'anno precedente):

$${}_{12}\Delta_m = Y_m - Y_{m-12} .$$

(8)

Come si può notare, nel passaggio a un'indagine continua svolta tutte le settimane del trimestre il parametro Y_t è esprimibile come media dei parametri relativi alle singole settimane a esso appartenenti, a differenza dell'indagine precedente nella quale la settimana di riferimento, una per trimestre, era rappresentativa dell'intero trimestre e pertanto i parametri trimestrali coincidevano di fatto con quelli relativi alla settimana prefissata.

L'obiettivo principale dell'indagine è fornire stime attendibili dei parametri (2), (3), (4) e (5) con riferimento ai domini territoriali d'interesse; tuttavia, poiché essa deve soddisfare anche i vincoli introdotti dal Regolamento UE sulle stime dei parametri di livello e di variazione relativi al 5 per cento della popolazione in età lavorativa, è utile riportare l'espressione formale anche per questi ultimi. In particolare, si ha che il primo parametro, relativo alla media annua, è dato dall'espressione

$$\tau_a = 0,05 \frac{1}{4} \sum_{t=1}^4 X_t ,$$

mentre il secondo parametro, riferito alla differenza tra due trimestri, è dato da

$$\tau_{\Delta} = 0,05 (X_t - X_{(t-1)}),$$

avendo indicato con X_t il numero di persone in età lavorativa nel trimestre t .

4.3.2 - Stimatori dei parametri di interesse

Si supponga di aver estratto un campione casuale rappresentativo della popolazione oggetto di indagine per ciascuna settimana dell'anno utilizzando il disegno di campionamento a due stadi di selezione con stratificazione delle unità primarie descritto successivamente nel paragrafo 4.4.2. Indichiamo con π_{sj} la probabilità di inclusione assegnata alla generica famiglia j e a tutti gli individui a essa appartenenti per la settimana di rilevazione s . Indichiamo, inoltre, con g_s il numero di famiglie rilevate nella generica settimana s , con N_{sj} il numero di individui appartenenti alla famiglia j della settimana s , con w_{sjq} , il coefficiente di riporto finale dell'individuo q appartenente alla famiglia j della settimana s , espresso da:

$$w_{sjq} = \frac{1}{\pi_{sj}} \delta_{sjq} ,$$

in cui δ_{sjq} rappresenta un fattore correttivo del peso diretto che garantisce l'uguaglianza tra i totali noti della popolazione per sesso e classi di età e le corrispondenti stime campionare, Y_{sjq} è

la variabile indicatrice pari a 1 se il q -esimo individuo della famiglia j rilevata nella settimana possiede la caratteristica z e pari a 0 altrimenti.

La stima del numero di persone che possiedono la caratteristica z con riferimento alla settimana s è data da

$$\hat{Y}_s = \sum_{j=1}^{g_s} \sum_{q=1}^{N_{sj}} Y_{sjq} w_{sjq} .$$

(9)

Sulla base della (9) si possono stimare i parametri introdotti nel precedente paragrafo:

a) stime di livello riferite ai diversi intervalli temporali

$$\hat{Y}_t = \frac{1}{S_t} \sum_{s=1}^{S_t} \hat{Y}_s ,$$

(10)

$$\hat{Y}_m = \frac{1}{S_m} \sum_{s=1}^{S_m} \hat{Y}_s$$

(11)

$$\hat{Y}_a = \frac{1}{S_a} \sum_{s=1}^{S_a} \hat{Y}_s ,$$

(12)

b) stime di variazioni nette

$${}_1\hat{\Delta}_m = \hat{Y}_m - \hat{Y}_{m-1}$$

(13)

$${}_{12}\hat{\Delta}_m = \hat{Y}_m - \hat{Y}_{m-12}$$

(14)

$${}_1\hat{\Delta}_t = \hat{Y}_t - \hat{Y}_{t-1}$$

(15)

$${}_4\hat{\Delta}_t = \hat{Y}_t - \hat{Y}_{t-4} .$$

(16)

4.4 - Il disegno della rilevazione

Il disegno di rilevazione delle forze di lavoro prevede una rotazione delle famiglie campione da un trimestre all'altro e una distribuzione del campione di ciascun trimestre tra le settimane che lo costituiscono, mentre non è prevista la reintervista di una famiglia all'interno di uno stesso trimestre.

In particolare, l'articolazione tra trimestri si basa su un *campionamento ruotato*, che prevede che i campioni di unità da osservare nelle varie occasioni siano formati a partire dal campione dell'occasione precedente, sostituendo a rotazione alcune unità statistiche con altre selezionate per la prima volta. Il campionamento ruotato può essere realizzato secondo differenti meccanismi di rinnovo delle unità, detti *schemi di rotazione*. Per la definizione dello schema di rotazione per la nuova indagine sulle forze di lavoro è stato condotto uno studio comparativo tra schemi alternativi, tra i quali alcuni prevedono la rotazione delle unità anche tra i mesi di uno stesso trimestre.

Relativamente alla distribuzione del campione all'interno di ciascun trimestre, occorre considerare che per soddisfare il requisito di continuità dell'indagine è stato necessario introdurre nel disegno campionario trimestrale anche la dimensione temporale, non presente nelle rilevazioni condotte in passato. Infatti, rispetto ai disegni precedenti in cui la rilevazione era effettuata solo in una settimana del trimestre, l'indagine attuale, richiedendo la distribuzione del campione su tutte le settimane, ha posto il problema della stratificazione temporale in aggiunta alla *usuale* stratificazione territoriale.

Per quanto riguarda la distribuzione del campione tra le settimane che costituiscono il trimestre si possono adottare differenti schemi di stratificazione. Infatti, con riferimento a ciascuno strato territoriale, definito dal disegno campionario a due stadi comuni-famiglie, si può selezionare un campione rappresentativo della popolazione della settimana *s*, oppure si possono selezionare campioni rappresentativi della popolazione mensile o trimestrale.

La scelta dello schema è scaturita da un'analisi comparativa tra differenti schemi proposti di articolazione temporale entro il trimestre.

In questo paragrafo si descrive sia lo schema di articolazione del campione tra trimestri differenti, sia l'articolazione dell'indagine con riferimento alle differenti settimane che compongono un dato trimestre di indagine.

4.4.1 - Lo schema di rotazione delle famiglie

L'indagine sulle forze di lavoro segue uno schema di rotazione trimestrale in cui le famiglie vengono intervistate per due trimestri consecutivi, escluse per due trimestri e successivamente re-intervistate per altri due trimestri⁴. Nei quattro trimestri di rilevazione in cui ciascuna famiglia campione viene intervistata, la settimana di rilevazione è sempre la medesima, ad esempio se una famiglia viene intervistata per la prima volta nella seconda settimana del trimestre, anche nei successivi trimestri in cui è coinvolta verrà intervistata nella seconda settimana.

Lo schema di rotazione è lo stesso che veniva utilizzato per l'indagine precedente; tuttavia, nella fase di ridisegno dell'indagine per passare a una rilevazione continua nel tempo, si è ritenuto utile riconsiderare lo schema di rotazione delle famiglie, conducendo un'analisi comparativa di possibili schemi di rotazione alternativi. Nella scelta degli schemi si è tenuto in considerazione il fatto che, nonostante la principale finalità dell'indagine sia quella di produrre stime trimestrali e annue, sono di interesse anche le stime mensili. Per garantire la rappresentatività mensile del campione è possibile ricorrere a due disegni di rilevazione differenti: il primo adotta uno schema di rotazione mensile, che prevede la sovrapposizione del campione anche tra mesi differenti di uno stesso trimestre; il secondo si basa su una rotazione trimestrale e una distribuzione del campione tra le settimane che costituiscono il trimestre che garantisce una stratificazione di tipo mensile.

Il presente paragrafo ha la finalità di presentare lo schema di rotazione utilizzato per l'indagine e di derivare le espressioni della varianza campionaria degli stimatori dei parametri di interesse, tenendo conto dell'impatto dovuto alla rotazione. A tale scopo ci si basa su una struttura generale di schema di rotazione, utile anche per inquadrare gli schemi considerati nel confronto.

⁴ Tenendo conto che lo schema trimestrale di rotazione delle famiglie è del tipo (2-2-2) e del fatto che nei quattro trimestri di rilevazione in cui viene intervistata ciascuna famiglia campione la settimana di rilevazione è sempre la medesima (ad esempio la famiglia viene intervistata sempre nella seconda settimana di ogni trimestre), le distanze effettive tra la prima intervista e le successive sono rispettivamente di 1, 4 e 5 trimestri (intendendo per trimestre un periodo di 13 settimane consecutive).

4.4.1.1 - Struttura generale di schema di rotazione

Al fine di introdurre una rappresentazione generalizzata degli schemi di rotazione, si considerano schemi che prevedono due sequenze di interviste delle famiglie in successive occasioni di indagine, separate da un periodo di interruzione. In tale contesto, i possibili schemi vengono indicati mediante la terna $(E_\Omega, \bar{E}_\Omega, E_\Omega)$, in cui E_Ω indica il numero di occasioni d'indagine consecutive in cui le famiglie campione sono intervistate, mentre \bar{E}_Ω indica il numero di occasioni in cui le famiglie escono dalla rilevazione prima di rientrarvi per altre E_Ω occasioni. L'indice Ω è pari a M per gli schemi mensili, in cui l'occasione d'indagine è il mese, ed è pari a T per gli schemi trimestrali. Ad esempio, utilizzando tale simbologia lo schema adottato attualmente per l'indagine viene indicato come $(2_T, 2_T, 2_T)$.

Con riferimento al generico schema $(E_\Omega, \bar{E}_\Omega, E_\Omega)$ il campione complessivo rilevato in una certa occasione sarà costituito, pertanto, da $2E_\Omega$ sottogruppi di unità di uguale numerosità: le unità che entrano per la prima volta nel campione, le unità che vi sono entrate nell'occasione precedente, fino alle unità presenti per l'ultima volta. Tali sottocampioni prendono il nome di *gruppi di rotazione*. Utilizzando la medesima notazione è possibile definire schemi di rotazione del tipo $(E_\Omega, -, -)$ in cui un'unità dopo essere rimasta nel campione per E_Ω occasioni, ne esce definitivamente; in tal caso il *numero di gruppi di rotazione* che vengono intervistati in ogni occasione è pari a E_Ω .

Dal momento che l'indagine sulle forze di lavoro è basata su una rilevazione di tipo continuo svolta in tutte le settimane, il campione assegnato a ciascuna occasione di indagine (mese o trimestre) viene suddiviso in campioni settimanali. Inoltre, per quanto riguarda gli schemi trimestrali qui considerati, la distribuzione del campione trimestrale tra le settimane viene effettuata in modo da realizzare una stratificazione temporale di tipo mensile che garantisce la possibilità di costruire stime mensili. Per tale ragione il campione trimestrale si suppone suddiviso in tre sottocampioni indipendenti di uguale numerosità, ciascuno assegnato a un mese del trimestre.

4.4.1.2 - Varianza degli stimatori in presenza della rotazione

La varianza campionaria dello stimatore $\hat{\theta}$ di un generico parametro θ basato su informazioni riferite a più occasioni di indagine (ad esempio una media annua di parametri di livello, oppure una differenza tra i parametri relativi a due occasioni d'indagine), può essere espressa mediante la seguente formula generale

$$Var(\hat{\theta}) = Var_{ccs}(\tilde{\theta}) \text{ deff}(\tilde{\theta}) \text{ eff}(\hat{\theta}) \text{ reff}(\hat{\theta}) \text{ ,}$$

in cui, relativamente a un dato schema di rotazione, $\text{reff}(\hat{\theta})$ rappresenta l'*effetto della rotazione*, ossia l'impatto sulla varianza dovuto alla sovrapposizione dei campioni relativi alle diverse occasioni di indagine, rispetto a una situazione di riferimento di assenza di sovrapposizione, in cui i campioni delle diverse occasioni sono selezionati in modo indipendente. La quantità $Var_{ccs}(\tilde{\theta})$ è la varianza di uno stimatore ottenuto con una strategia campionaria di riferimento, ottenuta utilizzando un ipotetico campione casuale semplice di pari numerosità, in termini di unità finali, a quella del disegno effettivo e lo stimatore espansione $\tilde{\theta}$. Le quantità $\text{deff}(\tilde{\theta})$ ed

$eff(\hat{\theta})$ sono statistiche note rispettivamente come *effetto del disegno* ed *effetto stimatore* e misurano complessivamente l'impatto sulla varianza campionaria dello stimatore $\hat{\theta}$ dovuto all'utilizzo della strategia campionaria complessa adottata (il significato di tali quantità verrà approfondito nel paragrafo 4.5).

Con riferimento al generico schema di rotazione, sia r l'indice di *gruppo di rotazione*, R il numero dei *gruppi di rotazione* presenti in un mese; \hat{Y}_{mr} lo stimatore del parametro mensile Y_m ottenuto in base alle unità campione del gruppo di rotazione r rilevate nel mese m . Negli schemi di rotazione trimestrali ciascun gruppo di rotazione r , che è riferito al trimestre, è suddiviso in tre sottogruppi mensili, ciascuno di numerosità campionaria pari a un terzo della numerosità complessiva del gruppo. Pertanto, in tal caso, lo stimatore \hat{Y}_{mr} è applicato al sottoinsieme di unità del gruppo di rotazione r rilevate nel mese m .

Per ciascuno dei parametri (2)-(8), si derivano di seguito le corrispondenti espressioni degli stimatori e delle varianze campionarie, in termini di effetto della rotazione, in funzione degli stimatori \hat{Y}_{mr} riferiti ai singoli gruppi di rotazione:

- stimatore del parametro Y_m e relativa varianza di campionamento

$$\hat{Y}_m = \frac{1}{R} \sum_{r=1}^R \hat{Y}_{mr} ;$$

(17)

$$Var(\hat{Y}_m) = \frac{1}{R^2} \left[\sum_{r=1}^R Var(\hat{Y}_{mr}) \right];$$

(18)

- stimatore del parametro Y_t e relativa varianza

$$\hat{Y}_t = \frac{1}{3} \frac{1}{R} \sum_{m=1}^3 \sum_{r=1}^R \hat{Y}_{mr} ;$$

(19)

$$Var(\hat{Y}_t) = \frac{1}{3^2} \frac{1}{R^2} \left[\sum_{m=1}^3 \sum_{r=1}^R Var(\hat{Y}_{mr}) + \sum_{m=1}^3 \sum_{m' \neq m}^3 \sum_{r=1}^R \sum_{r'=1}^R Cov(\hat{Y}_{mr}, \hat{Y}_{m'r'}) \right];$$

(20)

- stimatore del parametro Y_a e relativa varianza:

$$\hat{Y}_a = \frac{1}{12} \frac{1}{R} \sum_{m=1}^{12} \sum_{r=1}^R \hat{Y}_{mr} ;$$

(21)

$$Var(\hat{Y}_a) = \frac{1}{12^2} \frac{1}{R^2} \left[\sum_{m=1}^{12} \sum_{r=1}^R Var(\hat{Y}_{mr}) + \sum_{m=1}^{12} \sum_{m' \neq m}^{12} \sum_{r=1}^R \sum_{r'=1}^R Cov(\hat{Y}_{mr}, \hat{Y}_{m'r'}) \right]; \quad (22)$$

- stimatore del parametro ${}_k \Delta_m$ (per $k=1$ o 12) e relativa varianza:

$${}_k \hat{\Delta}_m = \frac{1}{R} \left[\sum_{r=1}^R \hat{Y}_{mr} - \sum_{r=1}^R \hat{Y}_{m-k,r} \right]; \quad (23)$$

$$Var({}_k \hat{\Delta}_m) = \frac{1}{R^2} \left[\sum_{r=1}^R Var(\hat{Y}_{mr}) + \sum_{r=1}^R Var(\hat{Y}_{m-k,r}) - 2 \sum_{r=1}^R \sum_{r'=1}^R Cov(\hat{Y}_{mr}, \hat{Y}_{m-k,r'}) \right]; \quad (24)$$

- stimatore del parametro ${}_k \Delta_t$ (per $k=1$ o 4) e relativa varianza:

$${}_k \hat{\Delta}_t = \frac{1}{3} \frac{1}{R} \left[\sum_{m=1}^3 \sum_{r=1}^R \hat{Y}_{mr} - \sum_{m'=1}^3 \sum_{r'=1}^R \hat{Y}_{m'r'} \right]; \quad (27)$$

$$Var({}_k \hat{\Delta}_t) = Var(\hat{Y}_t) + Var(\hat{Y}_{t-k}) - \frac{1}{3^2} \frac{1}{R^2} \left[2 \sum_{m=1}^3 \sum_{m' \neq m}^3 \sum_{r=1}^R \sum_{r'=1}^R Cov(\hat{Y}_{mr}, \hat{Y}_{m'r'}) \right] \quad (28)$$

in cui $m \in t, m' \in t-k$.

Per derivare le espressioni esplicite delle varianze sopra riportate, si possono introdurre le ipotesi semplificatrici di omoschedasticità e di costanza nel tempo della struttura di autocovarianza, ossia:

- 1) $Var(\hat{Y}_{mr}) = 3 R Var(\hat{Y}_t) = 3 R Var(\hat{Y}_t), \quad \forall m, r;$
- 2) $Cov(\hat{Y}_{mr}, \hat{Y}_{(m-k)r}) = \rho_k \sqrt{Var(\hat{Y}_{mr}) Var(\hat{Y}_{(m-k)r})} = 3 R \rho_k Var(\hat{Y}_t),$

essendo $Var(\hat{Y}_t)$ la varianza dello stimatore del parametro Y_t riferito a un generico trimestre basato sull'intero campione di un trimestre di indagine e ρ_k il coefficiente di correlazione della variabile y a k mesi di distanza.

Considerando ora lo schema $(2_T, 2_T, 2_T)$ effettivamente utilizzato per l'indagine, nel prospetto 4.2 si riporta l'articolazione rispetto al tempo e ai gruppi di rotazione.

Prospetto 4.2 - Lo schema $(2_T, 2_T, 2_T)$

Trimestre	Gruppi di rotazione									
I trimestre anno a	A4	B3			E2	F1				
II trimestre anno a		B4	C3			F2	G1			
III trimestre anno a			C4	D3			G2	H1		
IV trimestre anno a				D4	E3			H2	I1	
I trimestre anno a+1					E4	F3			I2	J1
II trimestre anno a+1						F4	G3			J2 K1

Le espressioni effettive ottenute a partire dalle formule generali delle varianze degli stimatori dei parametri di interesse per lo schema $(2_T, 2_T, 2_T)$ sono di seguito riportate:

$$Var(\hat{Y}_m) = 3Var(\hat{Y}.)$$

$$Var(\hat{Y}_t) = Var(\hat{Y}.)$$

$$Var(\hat{Y}_a) = \frac{1}{4}Var(\hat{Y}.) \left[1 + \frac{3}{4}\rho_3 + \frac{1}{8}\rho_9 \right]$$

$$Var({}_1\hat{\Delta}_m) = 6Var(\hat{Y}.)$$

$$Var({}_{12}\hat{\Delta}_m) = 6Var(\hat{Y}.) \left[1 - \frac{1}{2}\rho_{12} \right]$$

$$Var({}_1\hat{\Delta}_t) = 2Var(\hat{Y}.) \left[1 - \frac{1}{2}\rho_3 \right]$$

$$Var({}_4\hat{\Delta}_t) = 2Var(\hat{Y}.) \left[1 - \frac{1}{2}\rho_{12} \right]$$

Nelle precedenti espressioni, nel membro di destra il termine tra parentesi rappresenta l'effetto della rotazione, mentre il primo fattore è la varianza dello stimatore in assenza di rotazione.

4.4.1.3 - Confronto tra gli schemi di rotazione considerati

La scelta degli schemi da porre a confronto è ricaduta su schemi che prevedono un numero limitato di reinterviste delle famiglie campione, al fine di ridurre il carico statistico sulle unità campionarie. Inoltre sono stati considerati schemi che garantiscono la sovrapposizione del campione sia a un trimestre che a un anno di distanza al fine di stabilizzare le stime delle variazioni nette tra due trimestri consecutivi e tra due trimestri a distanza di un anno. Tuttavia, la sovrapposizione tra le unità appartenenti a campioni di diverse occasioni di indagine determina un peggioramento nell'efficienza delle stime medie annue (e per gli schemi di

rotazione mensili anche per le stime trimestrali), pertanto nella scelta degli schemi si è cercato un compromesso tra queste esigenze contrastanti.

Gli schemi di rotazione presi in considerazione sono:

- R1=($2_M, 10_M, 2_M$);
- R2=($3_M, 9_M, 3_M$);
- R3=($2_T, 2_T, 2_T$);
- R4=($5_T, -, -$).

Tra i possibili schemi mensili non è stato preso in considerazione lo schema ($4_M, 8_M, 4_M$) utilizzato per l'indagine statunitense Current Population Survey, che prevede otto interviste per ciascuna famiglia campione; è stato inoltre escluso dall'analisi lo schema ($6_M, -, -$) utilizzato per l'indagine Labour Force Survey (Statistics Canada), perché non garantisce la sovrapposizione a un anno di distanza.

Sulla base delle formule sopra derivate e delle ipotesi 1) e 2) è possibile ottenere le espressioni delle varianze e dei corrispondenti effetti della rotazione per i diversi schemi di rotazione considerati, con riferimento a ciascun parametro di interesse; le suddette espressioni, vengono riportate prospetto 4.3

Prospetto 4.3 - Varianze di campionamento e effetto della rotazione per gli schemi R1-R2-R3-R4 con riferimento a ciascuno stimatore

Stime	Var camp. in assenza di rotaz.	Schema R1	Schema R2	Schema R3	Schema R4
\hat{Y}_m	$3Var(\hat{Y}.)$	1	1	1	1
\hat{Y}_t	$Var(\hat{Y}.)$	$\left[1 + \frac{2}{3}\rho_1\right]$	$\left[1 + \frac{8}{9}\rho_1 + \frac{2}{9}\rho_2\right]$	1	1
\hat{Y}_a	$\frac{1}{4}Var(\hat{Y}.)$	$\left[1 + \frac{11}{12}\rho_1 + \frac{1}{24}\rho_{11}\right]$	$\left[1 + \frac{88}{72}\rho_1 + \frac{40}{72}\rho_2 + \frac{4}{72}\rho_{10} + \frac{4}{72}\rho_{11}\right]$	$\left[1 + \frac{3}{4}\rho_3 + \frac{1}{8}\rho_9\right]$	$\left[1 + \frac{6}{5}\rho_3 + \frac{3}{5}\rho_6 + \frac{1}{5}\rho_9\right]$
${}_1\hat{\Delta}_m$	$6Var(\hat{Y}.)$	$\left[1 - \frac{1}{2}\rho_1\right]$	$\left[1 - \frac{2}{3}\rho_1\right]$	1	1
${}_{12}\hat{\Delta}_m$	$6Var(\hat{Y}.)$	$\left[1 - \frac{1}{2}\rho_{12}\right]$	$\left[1 - \frac{1}{2}\rho_{12}\right]$	$\left[1 - \frac{1}{2}\rho_{12}\right]$	$\left[1 - \frac{1}{5}\rho_{12}\right]$
${}_1\hat{\Delta}_t$	$2Var(\hat{Y}.)$	$\left[1 + \frac{1}{2}\rho_1\right]$	$\left[1 + \frac{12}{18}\rho_1\right]$	$\left[1 - \frac{1}{2}\rho_3\right]$	$\left[1 - \frac{4}{5}\rho_3\right]$
${}_4\hat{\Delta}_t$	$2Var(\hat{Y}.)$	$\left[1 + \frac{2}{3}\rho_1 - \frac{2}{12}\rho_{11} - \frac{6}{12}\rho_{12} - \frac{2}{12}\rho_{13}\right]$	$\left[1 + \frac{16}{18}\rho_1 + \frac{4}{18}\rho_2 - \frac{9}{18}\rho_{12} - \frac{4}{18}\rho_{13} - \frac{1}{18}\rho_{14}\right]$	$\left[1 - \frac{1}{2}\rho_{12}\right]$	$\left[1 - \frac{1}{5}\rho_{12}\right]$

Per effettuare la scelta dello schema di rotazione, è stata condotta un'analisi empirica per mettere a confronto gli schemi considerati in termini dell'effetto della rotazione sulla varianza delle stime. Relativamente alle correlazioni fra trimestri, sono stati utilizzati i valori delle correlazioni fra trimestri stimate in base ai dati dell'indagine sulle forze di lavoro degli ultimi anni; per quanto riguarda, invece, le correlazioni tra mesi, è stata effettuata un'interpolazione sulla base dei ρ trimestrali.

Nel prospetto 4.1 sono riportate le valutazioni numeriche ottenute, dalle quali emerge che per quanto riguarda gli schemi di tipo mensile risulta migliore lo schema A, mentre tra i due trimestrali risulta migliore lo schema C.

Prospetto 4.4 - Effetti della rotazione rispetto alla varianza in assenza di rotazione negli schemi considerati

STIME	Schemi mensili		Schemi trimestrali	
	$(2_M, 10_M, 2_M)$	$(3_M, 9_M, 3_M)$	$(2_T, 2_T, 2_T)$	$(5_T, -, -)$
Mensile	1,00	1,00	1,00	1,00
Trimestrale	1,51	1,85	1,00	1,00
Annua	1,73	2,43	1,65	2,46
Diff. 1 mese	0,62	0,49	1,00	1,00
Diff. 12 mesi	0,69	0,69	0,69	0,87
Diff. 1 trim.	1,38	1,51	0,63	0,40
Diff. 4 trim.	0,99	1,37	0,69	0,87

Per ragioni di tipo prevalentemente organizzativo legate al controllo delle operazioni sul campo e ai tempi di trasferimento dei dati relativi a due occasioni di indagine consecutive, è stato stabilito di utilizzare uno schema trimestrale e di conseguenza è stato scelto lo schema $(2_T, 2_T, 2_T)$.

4.4.2 - Studio della stratificazione

I disegni di campionamento studiati per la definizione dell'articolazione del campione trimestrale tra le settimane sono noti in letteratura come *disegni di campionamento nello spazio e nel tempo* (Vos, 1963). In tali disegni si seleziona un campione dalle N unità della popolazione nei T tempi differenti in cui è suddiviso il periodo di riferimento di un'indagine (il trimestre per l'indagine forze lavoro), al fine di ottenere un campione rappresentativo utile alla stima dei parametri relativi all'intero di periodo di riferimento.

Il disegno di campionamento finale scaturisce, ovviamente, dalla definizione congiunta dello *schema di campionamento nello spazio* (selezione delle unità) e dello *schema di campionamento nel tempo* (selezione del periodo di rilevazione per ciascuna unità campionaria).

4.4.2.1 - Schema di campionamento nello spazio

Il disegno di campionamento nello spazio è di tipo complesso a più stadi, con stratificazione delle unità di primo stadio - i comuni - e suddivisione di questi, in base alla dimensione demografica, in Ar e Nar ; all'interno di ciascun comune selezionato viene estratto dalla lista anagrafica un campione casuale di famiglie; tutti gli individui appartenenti alle famiglie campione vengono intervistati.

La stratificazione dei comuni viene effettuata nell'ambito di ciascuno dei *domini territoriali pianificati*⁵ minimi, che per l'indagine sulle forze di lavoro sono le provincie.

Una descrizione dettagliata delle caratteristiche metodologiche di questo piano di campionamento, comunemente adottato per le principali indagini Istat sulle famiglie, è contenuta nel lavoro di Falorsi *et al.* (1996); tuttavia, ne vengono richiamati i principali aspetti teorici e la conseguente applicazione all'indagine sulle forze di lavoro.

La definizione del campione, che prevede la formazione degli strati di comuni, la definizione del numero di comuni campione e del numero di famiglie campione per comune, dipende dalla scelta del numero minimo di famiglie campione per comune \bar{g} , dalla frazione di campionamento f_d assegnata a ciascun dominio pianificato d e dal numero di comuni campione per strato, \bar{c} .

In particolare, per il generico dominio pianificato d , in base alla scelta di \bar{g} e f_d , si determina una soglia di popolazione λ_d in base alla quale si definisce la suddivisione dei comuni universo in AR e NAR:

$$\lambda_d = \frac{\bar{g} \cdot \bar{\varphi}_d}{f_d},$$

essendo $\bar{\varphi}_d$ l'ampiezza media delle famiglie nel dominio d .

I comuni di maggiore dimensione demografica, che hanno popolazione residente maggiore o uguale a λ_d , costituiscono ciascuno strato a sé stante, e vengono definiti, pertanto, AR. Per i rimanenti comuni NAR si formano strati di comuni approssimativamente di uguale ampiezza pari a $\lambda_d \cdot \bar{c}$.

Nell'ambito di ciascun dominio territoriale d , si adotta un disegno *autoponderante*, nel senso che a tutte le famiglie del dominio d si assegna uguale probabilità di inclusione nel campione π_{hij} pari alla frazione di campionamento del dominio

$$\pi_{hij} = f_d,$$

in cui hij , indica la generica famiglia j del comune i dello strato h , essendo la probabilità di inclusione delle famiglie campione espressa da

$$\pi_{hij} = \bar{c} \frac{N_{hi} g_{hi}}{N_h G_{hi}} \quad (29)$$

dove N_h e N_{hi} sono rispettivamente i totali delle popolazioni residenti nello strato h e nel comune hi , g_{hi} e G_{hi} sono rispettivamente il numero di famiglie campione e residenti nel comune hi .

Tenendo conto del criterio di formazione degli strati di comuni, è facile verificare che, in base alla (29), per il campione di famiglie g_{hi} del generico comune i dello strato h , si ha

- $g_{hi} = \frac{1}{\bar{c}} \frac{N_h}{\varphi_d} f_d \cong \bar{g}$, nel caso in cui il comune hi è NAR;
- $g_{hi} = \frac{N_{hi}}{\varphi_d} f_d \geq \bar{g}$, nel caso in cui il comune hi è AR.

⁵ Per dominio pianificato si intende una sottopopolazione per la quale viene determinata la numerosità campionaria che consente di ottenere un prefissato livello dell'errore di campionamento delle stime.

Passando adesso a descrivere le modalità di scelta dei fattori \bar{g} , f_d e \bar{c} , è importante ricordare che lo *schema di rotazione* delle famiglie prescelto è lo schema trimestrale $(2_T, 2_T, 2_T)$, per il quale non è prevista la sovrapposizione del campione all'interno del trimestre; ha senso, quindi, riferire i tre fattori al trimestre.

Una volta definiti i precedenti fattori, la definizione complessiva del disegno di campionamento nello spazio e nel tempo dipende dalla modalità di suddivisione del campione trimestrale tra le settimane e dallo schema probabilistico di selezione delle settimane campione (par. 4.4.3).

Scelta delle frazioni di campionamento

La scelta delle frazioni di campionamento è legata all'allocazione del campione trimestrale di famiglie nei diversi domini territoriali; nelle indagini Istat sulle famiglie, in cui sono sempre presenti più variabili di interesse e più domini territoriali di studio, l'allocazione viene studiata, in generale, mediante i metodi di *allocazione multivariata*.

Scelta del numero minimo di famiglie campione per comune

La definizione del numero minimo, \bar{g} , deriva in generale da considerazioni legate sia all'efficienza delle stime che a esigenze di tipo organizzativo⁶. La scelta di \bar{g} è legata sia alla distribuzione del campione trimestrale nelle diverse settimane, e in particolare al numero di volte nel trimestre che il comune campione è coinvolto nella rilevazione (\bar{s}_{AR} per i comuni Ar e \bar{s}_{NAR} per i comuni Nar), sia allo schema di rotazione delle famiglie $(2_T, 2_T, 2_T)$ che definisce il numero di gruppi di rotazione in cui è suddiviso il campione settimanale.

Pertanto, avendo definito una numerosità campionaria minima settimanale per comune \bar{g}_s , \bar{g} il numero minimo per comune è dato dal prodotto del numero minimo di interviste per settimane e il numero di settimane in cui il comune è coinvolto nella rilevazione trimestrale

$$\bar{g} = \bar{g}_s \bar{s}_{NAR}, \quad (30)$$

sotto la condizione che il numero di volte in cui i comuni Ar sono coinvolti nella rilevazione trimestrale sia maggiore o uguale al numero di volte in cui vi sono coinvolti i comuni Nar ($\bar{s}_{AR} \geq \bar{s}_{NAR}$).

A sua volta il valore di \bar{g}_s dipende dallo schema di rotazione e dalla tecnica di rilevazione. Tenendo conto che il campione settimanale è suddiviso in 4 gruppi di rotazione e che solo le famiglie del gruppo *entrante*, vengono intervistate mediante intervista diretta (Capi), mentre le famiglie degli altri tre gruppi di rotazione vengono intervistate mediante intervista telefonica (CATI), vale la condizione $\bar{g}_s = 4 \cdot \bar{\bar{g}}_s$, dove $\bar{\bar{g}}_s$ è il numero minimo di interviste dirette settimanali.

⁶ Per quanto riguarda l'efficienza delle stime, è noto che al crescere del numero minimo di famiglie per comune segue un aumento della variabilità delle stime, mentre le esigenze di tipo organizzativo richiedono che tale numero non sia troppo esiguo per l'organizzazione delle interviste nel comune e per il controllo del numero di comuni campione, essendo quest'ultimo inversamente legato a detto numero minimo.

Scelta del numero di comuni campione per strato

Per la definizione del numero di comuni campione per strato \bar{c} si adotta in genere il criterio di sceglierlo il più piccolo possibile. La scelta $\bar{c}=1$ è la più opportuna dal punto di vista dell'efficienza delle stime, in quanto porta alla costruzione di un maggiore numero di strati e quindi a un migliore effetto della stratificazione di primo stadio; tuttavia in tale caso occorre aggregare gli strati per il calcolo delle varianze campionarie. La definizione di $\bar{c} > 1$ non comporta problemi in fase di calcolo delle varianze, ma può risultare meno efficiente dal punto di vista dell'effetto della stratificazione. Tuttavia, alcune valutazioni empiriche hanno mostrato che la riduzione dell'efficienza delle stime al crescere di \bar{c} è molto contenuta.

La definizione di \bar{c} deve tener conto anche della scelta di \bar{s}_{AR} e \bar{s}_{NAR} ; infatti, come verrà chiarito meglio nel seguito, dalla definizione di queste tre quantità dipende il tipo di stratificazione temporale (mensile o trimestrale) adottata.

4.4.2.2 -Schema di campionamento nel tempo

Nella definizione degli obiettivi dell'indagine solo il trimestre costituisce obbligatoriamente un dominio di stima temporale ma non la settimana, non è necessario, pertanto, che nel campione di ciascuna settimana siano rappresentati tutti gli strati territoriali; è necessario, invece, che la rilevazione sia effettuata tutte le settimane, associando, ad esempio, un campione casuale di settimane ai comuni campione dello strato. In base allo schema di campionamento nello spazio e allo schema di selezione casuale delle settimane si possono ottenere disegni a *stratificazione settimanale, mensile o trimestrale*, a seconda che tutti gli strati territoriali di comuni siano rappresentati in ciascuna settimana, mese o trimestre dell'anno.

Per i disegni a stratificazione mensile e trimestrale, inoltre, è necessario definire i criteri di selezione delle settimane da associare agli strati territoriali.

Lo studio per la scelta del disegno campionario ha considerato i differenti criteri di stratificazione temporale e diversi metodi di associazione casuale settimane-comuni per i disegni a stratificazione mensile e/o trimestrale.

Il disegno prescelto si basa su una stratificazione mensile in quanto anche stime mensili sono di un certo interesse per la rilevazione forze lavoro. Inoltre, lo schema prescelto risponde anche a esigenze operative e di efficienza.

4.4.3 - Il metodo di associazione strato di comuni-settimana

Dal momento che non è possibile effettuare la rilevazione tutte le settimane in ciascuno strato di comuni, si è pensato di raggruppare gli strati di comuni in *superstrati* e far sì che in ciascun superstrato la rilevazione sia effettuata tutte le settimane del trimestre, e assegnare poi casualmente le diverse settimane del trimestre agli strati da cui è costituito ogni superstrato.

Poiché le settimane del trimestre sono tredici sarebbe necessario raggruppare in un superstrato 13 strati di comuni, una condizione considerata troppo vincolante; pertanto, in alternativa a questo schema sono stati formati superstrati costituiti da quattro strati di comuni, ciascuno dei quali effettua la rilevazione in tre settimane del trimestre (a distanza di 4/5 settimane l'una dall'altra) in modo tale che il superstrato sia rilevato per dodici settimane, essendo le settimane assegnate a ciascuno strato disgiunte da quelle assegnate agli altri strati del superstrato. La settimana del trimestre in cui il superstrato non è coinvolto nella rilevazione è scelta in modo casuale. L'*uniformità* del campione complessivo tra le 13 settimane viene garantita dalla scelta casuale della settimana di rilevazione mancante che differisce da superstrato a superstrato.

Nell'ambito di gruppi di 13 superstrati (52 strati) si selezionano senza reimmissione le 13 settimane mancanti, ciascuna delle quali da associare a uno specifico superstrato.

Nel prospetto 4.5 è riportato un esempio di schema di associazione strato-settimana: le colonne S1, S2 e S3 indicano rispettivamente la prima, la seconda e la terza settimana di rilevazione assegnata a ciascuno strato; le righe rappresentano gli strati, raggruppati in superstrati di quattro strati; nelle intersezioni righe-colonne sono contenuti numeri da 1 a 13 che indicano le settimane effettive di rilevazione; ad esempio per lo strato 7 del superstrato 2, le tre settimane di rilevazione assegnate sono la 3^a, la 7^a e la 12^a.

Come è possibile notare, nei superstrati 1, 2 e 13 sono rispettivamente mancanti la 2^a, l'8^a e la 10^a settimana del trimestre.

Prospetto 4.5 - Esempio di associazione strato-settimana

		S1	S2	S3
Superstrato 1	Strato 1	1	6	10
	Strato 2	3	7	11
	Strato 3	4	8	12
	Strato 4	5	9	13
<hr/>				
Superstrato 2	Strato 5	1	5	10
	Strato 6	2	6	11
	Strato 7	3	7	12
	Strato 8	4	9	13
<hr/>				
.....				
Superstrato 13	Strato 49	1	5	9
	Strato 50	2	6	11
	Strato 51	3	7	12
	Strato 52	4	8	13

I superstrati sono formati all'interno delle province unendo strati contigui costituiti, pertanto, da comuni di dimensione media il più possibile simile. L'assegnazione casuale tra strati e settimane è stata effettuata a livello Italia, avendo effettuato un ordinamento casuale delle province all'interno delle regioni e degli strati all'interno dei superstrati.

E' importante far presente che in questo schema casuale di selezione delle settimane si ottiene sostanzialmente una stratificazione mensile del campione, infatti, le tre settimane di rilevazione associate a ciascuno strato sono separate da un intervallo temporale di quattro o cinque settimane; ciò garantisce la presenza di ciascuno strato di comuni in ognuno dei tre mesi del trimestre⁷. Inoltre, considerando complessivamente i quattro strati di ciascun superstrato si ottiene una copertura quasi completa delle settimane di ciascun mese.

Infine, si osservi che lo schema è applicabile nei disegni per i quali gli strati di comuni sono coinvolti in tre settimane del trimestre; ciò può essere realizzato sia quando il numero di comuni campione per strato è pari a uno e le settimane di rilevazione sono pari a tre ($\bar{c}=1$, $\bar{s}_{AR}=3$, $\bar{s}_{NAR}=3$), sia per un disegno di campionamento per il quale in ciascun strato NAR si selezionano tre comuni campione coinvolti ciascuno in una sola settimana di rilevazione; ovviamente le settimane di rilevazione per ciascun comune AR devono comunque essere pari a 3 ($\bar{c}=3$, $\bar{s}_{AR}=3$, $\bar{s}_{NAR}=1$).

Il disegno $\bar{c}=1$, $\bar{s}_{AR}=3$, $\bar{s}_{NAR}=3$ è quello effettivamente adottato per la selezione del campione dell'indagine forze lavoro. Nella formazione degli strati di comuni all'interno di ciascuna provincia, si è imposta la condizione che il numero complessivo di strati sia un multiplo

⁷ A seconda del calendario ogni trimestre sarà formato da due mesi di quattro settimane e un mese di cinque. Le settimane a cavallo tra due mesi vengono assegnate a uno dei due mesi secondo una logica di prevalenza dei giorni.

di quattro per poter formare almeno un superstrato all'interno di ciascuna provincia e poter applicare il metodo sopra descritto di associazione tra strato e settimana del trimestre.

Tenendo conto della (30), avendo prescelto un numero $\bar{g}_s = 4$ di interviste dirette, il valore di \bar{g} per il disegno considerato risulta pari a 48. Nel seguente schema si descrive la ripartizione del campione trimestrale di famiglie tra i gruppi di rotazione (si veda 4.4.1) e tra le settimane del trimestre (si veda prospetto 4.5) per quanto riguarda i comuni NAR.

Prospetto 4.6 - Ripartizione del campione trimestrale di famiglie in base al piano di campionamento attuale con $\bar{g} = 48$

	Gruppo 1	Gruppo 2	Gruppo 3	Gruppo 4	Totale Mese
Camp. Mese 1	4	4	4	4	16
Camp. Mese 2	4	4	4	4	16
Camp. Mese 3	4	4	4	4	16
Totale Gruppo	12	12	12	12	48

Per i comuni AR il numero di interviste dirette deve risultare maggiore o uguale a 4, inoltre, data la struttura di rotazione e di articolazione tra le settimane, il numero complessivo di famiglie campione per comune deve risultare maggiore di 48 e multiplo di 12.

4.4.4 - Confronto tra i disegni proposti

La scelta del disegno descritto è stata effettuata confrontando differenti disegni a stratificazione settimanale, mensile o trimestrale.

Le principali caratteristiche dei disegni presi in considerazione possono essere riassunte nel prospetto seguente.

Prospetto 4.7 - Principali caratteristiche dei disegni di campionamento analizzati

Disegno	\bar{C}	\bar{s}_{AR}	\bar{s}_{NAR}	\bar{g}	Dirette AR a settimana	Dirette NAR a settimana	Comuni Campione	Comuni Campione AR	Stratificazione temporale
0	13	13	1	52	≥ 1	13	1499	179	Settimanale
1	1	3	3	48	≥ 4	4	1219	354	Mensile
2	1	1	1	48	≥ 12	12	1189	340	Trimestrale
3	3	3	1	48	≥ 4	12	1334	199	Mensile
4	1	Variabile	1	48	[8, 18]	12	1189	340	Trimestrale

Un primo confronto tra i disegni ha riguardato l'equiripartizione del campione di famiglie tra le settimane del trimestre. A tale scopo è stato effettuato uno studio tramite simulazione: per ciascun disegno, una volta effettuata la stratificazione e la selezione dei comuni campione, è stata replicata 100 volte la fase di associazione tra comuni e settimane.

Sono stati calcolati due indicatori della distanza dall'equiripartizione: il primo è la media dei 100 coefficienti di variazione tra le numerosità dei 13 campioni settimanali regionali ottenuti nelle singole replicazioni; il secondo è il coefficiente di variazione delle 13 dimensioni campionarie settimanali medie ottenute sulla base delle 100 replicazioni.

Dall'esame dei risultati è emerso che, sia a livello nazionale che a livello regionale, i due disegni che meno si allontanano dall'equiripartizione sono il disegno 1 e il disegno 4. I due

disegni hanno differenti implicazioni di tipo organizzativo: il disegno 1 prevede che tutti i comuni, anche i NAR, effettuino tre rilevazioni in ciascun trimestre; il disegno 4, invece, prevede che i comuni NAR effettuino una sola rilevazione a trimestre, mentre i comuni AR effettuano la rilevazione un numero variabile di settimane.

I disegni sono stati messi a confronto anche in termini di efficienza campionaria delle stime prodotte. Per ciascun disegno sono stati selezionati 500 campioni indipendenti da 4 differenti pseudo-popolazioni generate a partire dai dati censuari e dell'indagine Forze Lavoro.

Alla luce dei risultati della simulazione, si può affermare che il disegno a stratificazione settimanale è quello meno efficiente, mentre il disegno 1 risulta generalmente più efficiente per le regioni del centro-nord e il disegno 4 risulta migliore nelle regioni del centro-sud.

In conclusione, avendo osservato una sostanziale equivalenza in termini di efficienza tra i due disegni 1 e 4, si è scelto il disegno 1 che garantisce una rappresentatività a livello mensile e, quindi, la possibilità di ottenere stime mensili.

I dettagli delle analisi svolte per la scelta del disegno e dello schema di associazione comuni-settimana sono riportati in De Vitiis *et al.* (2005).

4.5 - L'allocazione campionaria

La dimensione e l'allocazione del campione della precedente indagine sulle forze di lavoro erano state definite per produrre stime con errore di campionamento sotto controllo a livello nazionale e regionale. La dimensione complessiva del campione era stata ottenuta avendo imposto che la stima del numero delle persone in cerca di occupazione avesse un errore relativo a livello nazionale pari a 1,3 per cento. L'allocazione regionale era stata ottenuta in modo proporzionale alla dimensione demografica con il vincolo che l'errore relativo della stima delle persone in cerca di occupazione fosse inferiore al 10 per cento. Infine, la numerosità campionaria a livello provinciale era stata determinata in modo proporzionale all'interno delle regioni, purché tale numerosità fosse non inferiore a un minimo fissato atto a garantire una minima affidabilità delle stime provinciali.

Nello studio della nuova indagine sono stati considerati differenti disegni campionari che, pur rispettando i requisiti UE sull'errore di campionamento, si differenziano tra di loro sia per la scelta del tipo di *dominio territoriale pianificato*, sia per l'allocazione del campione tra i domini. In particolare, sono stati considerati tre disegni per i quali la provincia è il dominio territoriale più disaggregato, il primo con la stessa numerosità complessiva in termini di unità finali e la stessa allocazione a livello provinciale del disegno dell'indagine precedente, gli altri due, pur mantenendo la numerosità complessiva inalterata, considerano due differenti allocazioni definite imponendo vincoli differenti sugli errori a livello regionale o provinciale.

Infine, è stato studiato un ultimo disegno che ha come dominio pianificato la regione. Questo disegno è stato considerato per valutare quanto la rinuncia a tenere sotto controllo gli errori provinciali possa portare a una riduzione della numerosità campionaria complessiva.

Il disegno prescelto è il disegno con dominio provinciale in cui sono fissati vincoli più stringenti a livello provinciale.

Nel paragrafo successivo viene descritta la metodologia utilizzata per la determinazione della numerosità complessiva e dell'allocazione tra i domini.

4.5.1 - Metodologia di allocazione

Al fine di valutare i disegni alternativi, sia in termini di errori campionari attesi, sia relativamente alle implicazioni di carattere operativo, è necessario, per ciascun disegno,

espletare tutte le fasi di definizione dello schema campionario: allocazione tra i domini di stima, suddivisione dei comuni in Ar e Nar, stratificazione dei comuni Nar e definizione delle numerosità campionarie per comune.

Le fasi di definizione del disegno e di valutazione degli errori sono tra loro strettamente legate, in quanto per la determinazione stessa della numerosità campionaria e dell'allocazione tra i domini è necessario disporre di informazioni sulla variabilità dei fenomeni investigati e sugli *effetti del disegno, dello stimatore e della rotazione*. In particolare l'effetto del disegno è una statistica che a sua volta dipende dalla definizione della numerosità campionaria. Tenendo costanti gli effetti dello stimatore e della rotazione, la determinazione della numerosità campionaria e dell'allocazione viene effettuata per mezzo di una procedura iterativa.

La metodologia utilizzata per la determinazione della numerosità campionaria e per l'allocazione tra i domini è basata su una generalizzazione del metodo di allocazione multivariata proposta da Bethel (1989), che consente di calcolare le numerosità campionarie dei domini di stima imponendo dei vincoli sugli errori relativi attesi di prefissate stime di interesse per un disegno a uno stadio stratificato e per un solo livello di dominio di studio (ad esempio le regioni o le province). La metodologia adottata consiste in un'estensione del metodo di Bethel al caso di campionamento a più stadi in cui sono presenti diversi livelli di dominio di stima (vedi Falorsi *et al.*, 1998). L'estensione al caso del disegno a più stadi viene ottenuta inflazionando la varianza di campionamento delle stime riferite a ciascun dominio con i corrispondenti effetti del disegno a due stadi. La corretta applicazione dell'estensione della metodologia di Bethel comporta la ripetizione della procedura di allocazione fino a quando l'effetto del disegno (e quindi la variabilità delle stime) si stabilizza.

Per la definizione delle numerosità campionarie alla prima iterazione sono utilizzate le informazioni desunte dall'indagine precedente sugli errori campionari e sugli effetti del disegno. Al passo successivo, sulla base di tali numerosità per dominio e delle caratteristiche del disegno è possibile ottenere una nuova valutazione della variabilità, che è poi inserita nella procedura per ottenere una nuova allocazione. Ci si arresta quando si verifica che la variabilità valutata alla fine di questo passo risulta approssimativamente uguale a quella calcolata all'iterazione precedente.

4.5.2 - Metodologia di calcolo degli errori attesi delle stime di interesse

I disegni studiati producono stime con errori campionari attesi differenti rispetto a quelli ottenuti con il disegno precedente in conseguenza delle modifiche introdotte a livello di allocazione, di costruzione degli strati (numero minimo di interviste per comune e dimensione degli strati) e quindi di parametri del disegno a due stadi (dimensione dei grappoli).

La fase di valutazione degli errori attesi dei disegni alternativi si basa sul calcolo degli effetti sulla variabilità delle stime dovuti alle modifiche introdotte.

La varianza di campionamento della generica stima \hat{Y}_{dt} , riferita alla generica provincia d , può essere espressa mediante la seguente formula

$$Var(\hat{Y}_{dt}) = Var_s(\tilde{Y}_{dt}) deff(\tilde{Y}_{dt}) eff(\hat{Y}_{dt}) \quad (31)$$

essendo

$$Var_s(\tilde{Y}_{dt}) \cong N_d^2 \frac{S_{dy}^2}{n_d}$$

in cui $Var_s(\tilde{Y}_{dt})$, a meno del fattore di correzione per popolazioni finite (approssimativamente uguale a uno), è la varianza di campionamento di una strategia di campionamento di riferimento,

ottenuta utilizzando un ipotetico campione casuale semplice di pari numerosità, in termini di unità finali, a quella del disegno effettivo e lo stimatore espansione, \tilde{Y}_{dt} , essendo S_{dy}^2 la varianza di popolazione, N_d e n_d rispettivamente il numero di individui della popolazione e del campione.

L'effetto del disegno $deff(\tilde{Y}_{dt})$ si ottiene come rapporto tra la varianza campionaria $Var(\tilde{Y}_{dt})$ dello stimatore espansione, riferita al disegno a due stadi effettivamente utilizzato e la varianza campionaria $Var_s(\tilde{Y}_{dt})$ dello stimatore espansione riferita al campione casuale semplice. Mentre l'effetto dello stimatore $eff(\hat{Y}_{dt})$ è dato dal rapporto tra la varianza campionaria $Var(\hat{Y}_{dt})$ del campionamento a due stadi dello stimatore effettivamente utilizzato e la varianza campionaria della strategia definita dal disegno campionario complesso utilizzato e lo stimatore espansione.

Sotto l'ipotesi di costanza della varianza di popolazione e dei coefficienti di correlazione intraclassa e considerando che le quantità caratterizzanti il disegno di campionamento (dimensione medie campionarie dei cluster di primo e secondo stadio) sono approssimativamente costanti entro la provincia, l'effetto del disegno si può scomporre in un effetto $deff_{str}(\tilde{Y}_{dt})$ dovuto alla stratificazione all'interno della provincia, in un effetto, $deff_{2st}(\tilde{Y}_{dt})$, dovuto allo schema di campionamento a due stadi (o effetto di clusterizzazione)

$$deff(\tilde{Y}_{dt}) = \frac{Var(\tilde{Y}_{dt})}{Var_s(\tilde{Y}_{dt})} = \frac{Var_{str}(\tilde{Y}_{dt})}{Var_s(\tilde{Y}_{dt})} \frac{Var(\tilde{Y}_{dt})}{Var_{str}(\tilde{Y}_{dt})} = deff_{str}(\tilde{Y}_{dt}) deff_{2st}(\tilde{Y}_{dt}) \quad (32)$$

dove $Var_{str}(\tilde{Y}_{dt})$ è la varianza di un ipotetico campione a uno stadio stratificato, avente la stessa stratificazione e le stesse numerosità campionarie negli strati, in termini di unità finali, del disegno a due stadi effettivamente utilizzato, in cui si selezionano gli individui al primo stadio di selezione.

Per la valutazione dell'effetto di clusterizzazione si può utilizzare la seguente espressione (vedi Falorsi e Falorsi, 1996):

$$deff_{2st}(\tilde{Y}_{dt}) = \frac{n_d}{N_d^2} \left\{ \frac{N_{d,AR}^2}{n_{d,AR}} {}_y deff_{d,AR} + \frac{N_{d,NAR}^2}{n_{d,NAR}} {}_y deff_{d,NAR} \right\}, \quad (33)$$

dove

$${}_y deff_{d,AR} = (1 + {}_y \rho_{d,AR} (\bar{b}_{d,AR} - 1)) \quad (34)$$

$${}_y deff_{d,NAR} = (1 + {}_y \rho_{d,NAR} (\bar{b}_{d,NAR} - 1)) \quad (35)$$

in cui $N_{d,AR}$, $N_{d,NAR}$, $n_{d,AR}$ e $n_{d,NAR}$ sono rispettivamente le numerosità della popolazione e del campione di parte AR e NAR, \bar{b}_{AR} è la numerosità media della famiglia nella parte AR, \bar{b}_{NAR} è il numero medio di individui campione nei comuni NAR, infine, ${}_y \rho_{d,AR}$ e ${}_y \rho_{d,NAR}$ sono i coefficienti di correlazione intraclassa all'interno dell'unità primaria (famiglia per la parte AR,

comune per la parte NAR) della variabile considerata (ad esempio, persone in cerca di occupazione).

Per la valutazione degli errori attesi in relazione ai differenti disegni campionari proposti, tutte le quantità $N_{d,AR}$, $N_{d,NAR}$, $n_{d,AR}$, $n_{d,NAR}$, \bar{b}_{AR} e \bar{b}_{NAR} sono facilmente desumibili una volta definito il disegno di campionamento e la sua allocazione tra i domini. I coefficienti di correlazione intraclasse, invece, presenti nella (33), sono un parametro incognito caratterizzante la popolazione. Tali coefficienti sono stati stimati sulla base dell'indagine precedente dal momento che è ragionevole supporre che questi si mantengano stabili al variare del disegno.

L'effetto della stratificazione all'interno dei domini $deff_{str}(\tilde{Y}_{dt})$ viene definito, come già detto sopra, come rapporto tra la varianza del campionamento stratificato casuale semplice e la varianza del campionamento casuale semplice. Per la valutazione di tale effetto sono stati utilizzati i dati del censimento 1991 (disponibili, però, solo a livello comunale) sul numero di occupati, di disoccupati e di persone in cerca di prima occupazione.

Per tutti i disegni considerati l'effetto stratificazione è risultato pressoché irrilevante (ossia prossimo all'unità). Questo è dovuto principalmente al tipo di stime considerate, ossia stime di frequenze per le quali le varianze di popolazione non variano molto in un ampio intervallo del valore del parametro oggetto di stima (vedi Cicchitelli *et al.* 1997). E', comunque, vero che si sarebbe forse potuto ottenere un effetto maggiore avendo a disposizione dati non a livello di comune ma di individui, anche se è noto che la stratificazione nelle indagini sulle famiglie non ha un grande impatto sull'efficienza delle stime (vedi Zannella, 1991).

Una volta calcolati gli errori attesi delle stime trimestrali, per ottenere una valutazione degli errori di campionamento delle stime annuali e differenza tra due trimestri è necessario tenere conto degli effetti della parziale sovrapposizione dei campioni trimestrali che deriva dalla rotazione delle unità.

Ipotizzando la varianza di popolazione costante nell'anno, la varianza di campionamento della stima annuale, ${}_a\hat{Y}_d$, può essere espressa mediante la seguente formula (vedi par. 4.4.1)

$$Var({}_a\hat{Y}_d) \cong \frac{1}{4} N_d^2 \frac{S_{dy}^2}{n_d} deff({}_a\tilde{Y}_d) eff({}_a\hat{Y}_d) reff({}_a\hat{Y}_d) ,$$

(36)

in cui $deff({}_a\tilde{Y}_d)$ e $eff({}_a\hat{Y}_d)$ sono gli analoghi delle quantità precedenti per stimatori di quantità medie annue ipotizzando un disegno campionario per il quale i campioni trimestrali sono indipendenti e $reff({}_a\hat{Y}_d)$ rappresenta l'effetto della rotazione sulla stima media annua (vedi par.4.4.1).

Analogamente la varianza della stima differenza tra due trimestri consecutivi, ${}_1\hat{\Delta}_t$, è espressa da

$$Var({}_1\hat{\Delta}_t) \cong 2 N_d^2 \frac{S_{dy}^2}{n_d} deff({}_1\tilde{\Delta}_{dt}) eff({}_1\hat{\Delta}_{dt}) reff({}_1\hat{\Delta}_{dt}) ,$$

(37)

in cui $reff({}_1\hat{\Delta}_{dt})$ rappresenta l'effetto della rotazione sulla stima di variazione netta tra due trimestri consecutivi.

Poiché nelle diverse province vengono selezionati campioni indipendenti, le varianze delle stime riferite ad altri domini territoriali, quali le regioni, le ripartizioni e l'intero territorio nazionale, si ottengono come somma delle varianze provinciali.

La fase di progettazione è stata realizzata facendo riferimento allo stimatore effettivamente adottato per l'indagine precedente, pertanto le differenti strategie si diversificano solo per l'effetto dovuto al disegno di campionamento e non per l'effetto connesso allo stimatore.

4.5.3 - Metodologia di calcolo degli errori attesi della stima del parametro di riferimento UE

Per quanto riguarda il parametro τ , relativo ai vincoli del regolamento europeo sulla precisione delle stime, i coefficienti di correlazione intraclasse della parte AR e NAR necessari per la valutazione dell'effetto del disegno non possono essere stimati direttamente sulla base dei dati dell'indagine, poiché τ è un parametro fittizio cui non corrisponde una variabile rilevabile.

Si è fatto pertanto ricorso all'uso di modelli regressivi che mettono in relazione il valore delle stime e il valore dei corrispondenti errori campionari. A partire dal modello, adattato su una nuvola consistente di punti che copre il campo di variabilità delle stime dell'indagine forze di lavoro, è stato predetto il valore dell'errore relativo $\hat{\varepsilon}(\tilde{\tau})$ della stima $\tilde{\tau}$ del parametro τ .

Infine è stato possibile determinare un valore per i coefficienti di correlazione intraclasse regionali della variabile 5% della popolazione in età lavorativa, ${}_{\tau}\rho_{AR}$ e ${}_{\tau}\rho_{NAR}$, tenendo conto della relazione che lega questi con l'effetto del disegno.

Il modello regressivo utilizzato per la determinazione della relazione tra valore della stima e valore del corrispondente errore relativo è quello usuale adottato per la presentazione sintetica degli errori campionari di stime di frequenze assolute e relative:

$$\log(\hat{\varepsilon}^2(\tilde{Y}_t)) = a + b \log(\tilde{Y}_t), \quad (38)$$

dove $\hat{\varepsilon}^2(\tilde{Y}_t)$ è il quadrato dell'errore relativo della stima \tilde{Y}_t .

I parametri della (38) sono stati stimati a livello regionale, separatamente per la parte AR e la parte NAR, sulla base di una nuvola di punti costituita da trenta stime e dai corrispondenti errori relativi.

Sulla base dei modelli adattati per la parte AR e la parte NAR sono stati valutati i corrispondenti errori relativi $\hat{\varepsilon}_{AR}(\tilde{\tau})$, $\hat{\varepsilon}_{NAR}(\tilde{\tau})$. A partire da questi ultimi, tramite la (32), è stato possibile valutare l'effetto del disegno che, una volta depurato dalla componente di stratificazione (vedi par. 4.5.2), mediante la (34) e la (35), permette di ricavare i coefficienti di correlazione intraclasse

$${}_{\tau}\rho_{AR} = \frac{deff_{AR}(\tilde{Y}_t) - 1}{(\bar{b}_{AR} - 1)} \quad (39)$$

$${}_{\tau}\rho_{NAR} = \frac{deff_{NAR}(\tilde{Y}_t) - 1}{(\bar{b}_{NAR} - 1)}. \quad (40)$$

Infine, i valori dei coefficienti di correlazione intraclasse ${}_{\tau}\rho_{AR}$ e ${}_{\tau}\rho_{NAR}$ sono stati applicati per la valutazione degli effetti dei disegni, e quindi dei corrispondenti errori, secondo la metodologia utilizzata per le variabili osservabili (par. 4.5.2).

4.6 - Il disegno prescelto

Nella tabella 4.2 sono riportati gli estremi superiori degli errori relativi attesi percentuali delle stime trimestrali di totali nei differenti domini di stima (province, regioni, Italia) per il disegno prescelto.

Per la stima relativa al 5% della popolazione in età lavorativa sono stati imposti i vincoli richiesti dal regolamento europeo per i domini regionale e nazionale opportunamente adattati a vincoli su stime trimestrali tenendo conto degli effetti della rotazione sulle stime annuali e differenza (par. 4.4.1).

Inoltre, si è imposto un vincolo sulla stima regionale delle persone in cerca di occupazione pari al 12%, mentre per la medesima stima a livello provinciale il vincolo è stato posto pari al 25% (per la stima trimestrale, il che equivale a un errore circa pari a 17 per cento per la corrispondente stima annua). Infine, si è posto un vincolo pari a 8% a livello provinciale per la stima annuale delle forze di lavoro.

Prospetto 4.8 - vincoli sugli errori relativi percentuali imposti a livello di provincia, regione e Italia

	Forze di lavoro	Occupati	Persone in cerca di occupazione	Persone in cerca di occupazione	Disoccupati	Altre persone in cerca di occupazione	5% popolazione in età lavorativa
Provincia	8%	-	25%	-	-	-	-
Regione	-	-	12%	-	-	-	11,5%
Italia	0,5%	0,5%	1,96%	2,7%	2,7%	3,5%	1,65%

4.7 - L'introduzione del nuovo campione

Il nuovo campione Forze Lavoro è stato introdotto gradualmente sulla base delle seguenti considerazioni.

Il campione dell'indagine sulle forze lavoro è formato da quattro sottocampioni di uguale numerosità, detti gruppi di rotazione identificati rispettivamente con i numeri da 1 a 4. L'indagine a regime prevede che in ogni occasione di indagine un gruppo di rotazione entra per la prima volta, un gruppo è intervistato per la seconda volta e così via.

Sulla base di evidenze emerse da verifiche effettuate in indagini simili, in Italia e all'estero, è ragionevole attendersi che esista una differenza sistematica tra le risposte fornite, in una medesima occasione d'indagine, dagli individui appartenenti a gruppi di rotazione differenti e quindi intervistati per la prima, seconda, terza o quarta. Ad esempio, è stato riscontrato nell'indagine Lfs canadese che i tassi di disoccupazione sono più elevati tra gli individui intervistati per la prima volta. Tale differenza sistematica produce un effetto distorsivo sulle stime che è sempre presente in indagini di questo tipo ed è noto come *rotation bias*. Nel periodo di introduzione della nuova indagine la composizione del campione, in termini di gruppi di rotazione, varia da trimestre a trimestre fino al raggiungimento della composizione teorica; durante tale periodo, pertanto, si possono riscontrare delle discontinuità nella serie storica, attribuibili alla differente composizione del campione da un trimestre all'altro. Di conseguenza, fintanto che il campione non raggiunge la composizione teorica, è sconsigliabile utilizzare i dati della nuova indagine per produrre stime ufficiali.

Inoltre, è possibile che l'utilizzo di nuovi intervistatori e di una nuova tecnica d'indagine introduca delle differenze sistematiche nelle prime occasioni d'indagine. Pertanto, l'introduzione

graduale del nuovo campione consente ai nuovi rilevatori di acquisire la necessaria esperienza durante i trimestri in cui le stime sono ancora basate sulla precedente indagine. L'introduzione graduale del campione è stata anche opportuna per poter valutare l'impatto organizzativo della nuova indagine. Infine, un adeguato periodo di sovrapposizione tra le due indagini è stato utile anche al fine di porre a confronto le stime ottenute con la precedente e la nuova indagine.

Schema teorico di introduzione del campione

Al fine di produrre, a partire dal primo trimestre di un ipotetico anno *a*, stime basate sulla nuova indagine che non presentino le distorsioni causate dai problemi sopra menzionati, sarebbe opportuno che il campione fosse introdotto parallelamente al campione dell'indagine precedente a partire da ottobre dell'anno *a-2* secondo lo schema 1. In tale schema i gruppi indicati con le lettere da A-I appartengono alla nuova indagine, mentre i gruppi N-Z appartengono alla precedente indagine.

Si sottolinea il fatto che lo schema sopra riportato è l'unico che permette di introdurre il nuovo campione in modo corretto e tale da rispettare completamente lo schema di rotazione.

Prospetto 4.9 - Schema teorico di introduzione del nuovo campione

TRIMESTRE	Gruppi di rotazione														
IV trimestre anno a-2	A1					N4	O3			R2	S1				
I trimestre anno a-1	A2	B1					O4	P3			S2	T1			
II trimestre anno a-1		B2	C1					P4	Q3			T2	U1		
III trimestre anno a-1			C2	D1					Q4	R3			U2	V1	
IV trimestre anno a-1	A3			D2	E1					R4	S3			V2	Z1
I trimestre anno a	A4	B3			E2	F1									
II trimestre anno a		B4	C3			F2	G1								
III trimestre anno a			C4	D3			G2	H1							
IV trimestre anno a				D4	E3			H2	I1						

4.8 – La fase di estrazione delle famiglie

L'estrazione delle unità di secondo stadio effettuata dai Comuni del campione deve seguire alcune rigide regole metodologiche, dall'utilizzo di uno specifico passo di estrazione ai criteri di ordinamento delle liste anagrafiche. Una particolare attenzione va quindi rivolta ai problemi che emergono in questa fase dell'indagine al fine di evitare gli effetti distorsivi legati a procedure di estrazione metodologicamente non corrette.

La fase di estrazione delle famiglie ha l'obiettivo non solo di produrre l'elenco teorico delle unità campione da intervistare e, quindi, di fornire al rilevatore tutte le informazioni necessarie al reperimento delle famiglie estratte sul territorio, ma anche quello di raccogliere dati anagrafici utili ad effettuare dei controlli a posteriori sulla qualità dei dati successivamente rilevati⁸. Ad esempio, l'informazione sulla cittadinanza⁹ dell'intestatario della scheda di famiglia consente di valutare la qualità delle stime prodotte relativamente alla popolazione straniera e quindi di considerare eventuali distorsioni determinate, ad esempio, da una sottostima delle famiglie con intestatario straniero. La tecnica di rilevazione Capi-Cati implica peraltro che i nominativi delle famiglie estratte vengano trasmessi dal Centro di coordinamento ai rilevatori. È quindi indispensabile gestire direttamente le procedure necessarie al trattamento delle liste. Questa necessità si traduce nell'ambito della Rcfl in un ulteriore investimento in qualità.

4.8.1 - Il nuovo sistema di estrazione dei nominativi delle famiglie del campione

Il sistema di estrazione delle famiglie chiama in causa, come attori principali, il Centro di coordinamento, gli Uffici territoriali e, ovviamente, i Comuni del campione che forniscono le liste dei nominativi. Tale sistema si basa su un software prodotto dall'Istat e sulle potenzialità delle trasmissioni telematiche. L'attivazione del sistema è stata preceduta dalla ricognizione sul grado di informatizzazione dei Comuni del campione¹⁰. Da tale verifica è risultato che la stragrande maggioranza di essi dispone di anagrafi informatizzate, o quantomeno di postazioni di lavoro informatiche il più delle volte con collegamento di rete. Il processo di acquisizione dei nominativi delle famiglie del campione è stato progettato al fine di sfruttare le potenzialità dei Comuni che hanno l'anagrafe informatizzata; di aiutare i Comuni senza anagrafe informatizzata a registrare i nominativi del campione in modo uniforme; di ridurre i tempi di acquisizione del campione; di monitorare l'intero processo. Lo schema di seguito riportato illustra il percorso seguito per l'estrazione delle famiglie del campione.

In primo luogo, è necessario sottolineare che la comunicazione indirizzata ai Comuni (lettera circolare) riporta: le regole da seguire per effettuare un'estrazione metodologicamente corretta; le modalità di trasmissione dei nominativi (in base al grado di informatizzazione del Comune); i tempi da rispettare. Alla comunicazione è allegato un particolare software applicativo prodotto dall'Istituto, denominato Gefi (Gestione elenchi famiglie per le indagini), per la registrazione controllata dei nominativi e per la produzione del file da inviare all'Istat¹¹.

⁸ Si può analizzare quanto il campione effettivo si discosti dal campione teorico, ad esempio quanto il numero delle famiglie di anziani risultante dalla rilevazione è sottostimato o sovrastimato rispetto a quello estratto.

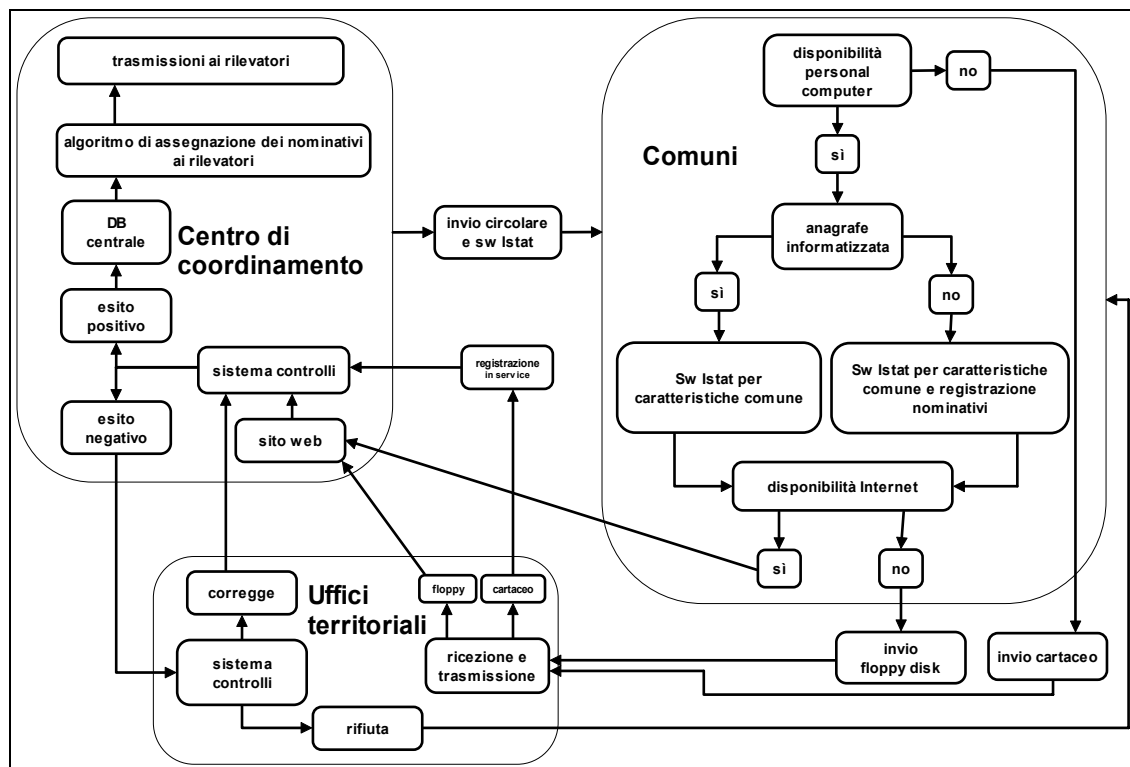
⁹ Tale informazione è rilevata a partire dall'estrazione del 2004.

¹⁰ Il grado di informatizzazione considera congiuntamente il possesso di almeno un personal computer, la possibilità di creare un file di nominativi secondo le indicazioni fornite dall'Istat, la presenza di una connessione Internet.

¹¹ Con riferimento al grado di informatizzazione, i Comuni che non possiedono personal computer effettuano l'estrazione manualmente e inviano le liste su supporto cartaceo agli Uffici territoriali. I Comuni con anagrafe informatizzata producono autonomamente la lista delle famiglie campione e utilizzano il software Istat solo per la registrazione delle caratteristiche del Comune e delle modalità di estrazione, mentre i Comuni con anagrafe non informatizzata, oltre a ciò, utilizzano Gefi anche per registrare i dati delle famiglie estratte. La trasmissione delle informazioni al Centro si avvale in parte della rete Internet. Qualora il Comune non disponga di collegamento in rete l'invio avviene per posta all'Ufficio territoriale.

Dallo schema sottostante si evince un doppio sistema di controllo delle informazioni, uno presso il Servizio formazione e lavoro e l'altro presso gli Uffici territoriali¹². Questi ultimi hanno il fondamentale compito di gestire i rapporti con i Comuni nella fase preliminare, nel corso dell'estrazione e dopo l'effettuazione dei controlli¹³. Il Servizio formazione e lavoro, oltre a supportare il lavoro degli Uffici territoriali e a verificare la correttezza delle informazioni raccolte, governa il complesso algoritmo di assegnazione dei nominativi ai rilevatori. In base a tale algoritmo, costruito nel rispetto delle caratteristiche del disegno campionario, i nominativi raccolti vengono associati ad ogni singolo rilevatore assegnando anche la specifica settimana di calendario in cui intervistare la famiglia.

Sistema di estrazione dei nominativi delle famiglie del campione



Il nuovo sistema di estrazione qui sinteticamente descritto offre numerosi vantaggi. In primo luogo, permette una verifica della correttezza delle procedure seguite nel corso dell'estrazione. Facilita inoltre la registrazione dei nominativi estratti e ne permette un immediato controllo; riduce i tempi del processo lavorativo, in quanto i dati trasmessi dai Comuni o dagli Uffici territoriali vengono immediatamente caricati sulla base dati; riduce i problemi di reperibilità

¹² I controlli effettuati presso il Centro riguardano l'individuazione in modo informatizzato di eventuali informazioni errate o mancanti, mentre quelli gestiti dagli Uffici territoriali comportano un'integrazione delle informazioni o il rifiuto del file trasmesso dal Comune.

¹³ Gli Uffici territoriali devono gestire eventuali sostituzioni dei Comuni, registrare le presenze/assenze dei Comuni del campione alle riunioni di istruzione, integrare e correggere laddove possibile le informazioni ricevute, inoltrare richieste di nuove estrazioni ai Comuni qualora vengano riscontrate lacune nella qualità dei dati ricevuti, effettuare il monitoraggio dei tempi stabiliti per l'estrazione e per la ricezione degli elenchi, controllare gli elenchi cartacei, monitorare l'andamento dell'estrazione per tenere costantemente sotto controllo la situazione.

delle famiglie che derivano da eventuali errori nelle liste anagrafiche e consente, dunque, una minore caduta campionaria.

4.8.2 - Il sistema dei controlli

Al fine di illustrare gli strumenti a disposizione per il monitoraggio della qualità dei dati raccolti nel corso dell'estrazione, a titolo esemplificativo si riportano nella tavola seguente i risultati più significativi relativi all'esperienza maturata durante l'acquisizione dei nominativi delle famiglie del campione avvenuta nel 2004¹⁴.

Prospetto 5.1 - Comuni del campione, informatizzazione dell'anagrafe, modalità di estrazione e criteri di ordinamento delle famiglie per regione - Anno 2004 (valori assoluti e composizioni per 100 Comuni della stessa regione)

REGIONI	N. Comuni	Anagrafe informatizzata	Estrazione informatizzata (b)	Criteri di ordinamento delle famiglie			
				Alfabetico di via	Cognome intestatario	Data di iscrizione in anagrafe	Altro
Piemonte/Valle d'Aosta(a)	143	99,3	77,5	88,1	2,1	3,5	6,3
Lombardia	173	100,0	64,7	87,3	1,7	4,0	7,0
Provincia di Bolzano	25	100,0	100,0	80,0	8,0	8,0	4,0
Provincia di Trento	48	97,9	91,5	75,0	8,3	8,3	8,4
Veneto	72	100,0	62,5	88,9	2,8	2,8	5,5
Friuli-Venezia Giulia	37	100,0	97,3	83,8	5,4	5,4	5,4
Liguria	25	100,0	88,0	84,0	-	8,0	8,0
Emilia-Romagna	69	100,0	72,5	88,4	2,9	4,3	4,4
Toscana	77	100,0	88,3	80,5	5,2	3,9	10,4
Umbria	20	100,0	90,0	100,0	-	-	-
Marche	32	100,0	81,3	84,4	6,3	3,1	6,2
Lazio	53	100,0	73,6	77,4	11,3	5,7	5,6
Abruzzo	32	100,0	93,8	75,0	12,5	9,4	3,1
Molise	28	100,0	42,9	82,1	7,1	7,1	3,7
Campania	89	97,8	79,3	88,8	2,2	2,2	6,8
Puglia	72	100,0	66,7	76,4	6,9	6,9	9,8
Basilicata	36	91,7	60,6	88,9	-	5,6	5,5
Calabria	76	96,1	69,9	82,9	6,6	6,6	3,9
Sicilia	103	97,1	68,0	80,6	4,9	3,9	10,6
Sardegna	36	100,0	69,4	86,1	5,6	5,6	2,7
Italia	1.246	99,0	74,4	84,3	4,4	4,7	6,6

(a) I dati relativi alle due regioni sono unificati in quanto l'Ufficio regionale Istat di Torino gestisce entrambe le regioni; (b) per 100 Comuni con anagrafe informatizzata.

La quasi totalità dei Comuni del campione dispone di un servizio anagrafico informatizzato. Tuttavia, circa un quarto dei Comuni, per l'indisponibilità di software adeguati o per la difficoltà nel trasferire a livello informatico la procedura prevista dall'Istat per l'estrazione, ha effettuato un'estrazione di tipo manuale. Grazie soprattutto all'opera di convincimento svolta dagli Uffici territoriali, l'utilizzo del software messo a disposizione dall'Istituto è stato ampio: solamente 10

¹⁴ Questi nominativi costituiscono il serbatoio utilizzato per lo svolgimento delle prime interviste da luglio 2004 a tutto il mese di giugno 2005.

Comuni, pari allo 0,8% del campione, hanno inviato all'Istat un elenco cartaceo dei nominativi delle famiglie campione¹⁵.

Nella circolare per l'estrazione delle famiglie spedita ai Comuni si specifica che al momento dell'estrazione le anagrafi devono essere ordinate per alfabetico di via. Circa otto Comuni ogni dieci hanno predisposto la lista anagrafica secondo tale criterio di ordinamento, mentre percentuali residue di Comuni hanno ordinato l'anagrafe per cognome dell'intestatario della scheda di famiglia, per data di iscrizione all'anagrafe o in base ad altri criteri.

Complessivamente dei 49.598 controlli effettuati durante la fase di estrazione del 2004 sui Comuni che hanno inviato elenchi informatizzati, il 12,6 per cento, corrispondente all'83,6 per cento dei file trasmessi dai Comuni, ha segnalato la presenza di errori¹⁶. Gli errori maggiormente commessi hanno riguardato in ordine decrescente i dati relativi al coniuge dell'intestatario della scheda di famiglia, la correttezza del tracciato record, la cittadinanza dell'intestatario della scheda di famiglia¹⁷ e il Cap. A motivo dei controlli effettuati il 26,9 per cento dei Comuni ha dovuto apportare correzioni inviando più di un file. Gli errori presenti nei file trasmessi dai Comuni sono distribuiti agli Uffici territoriali affinché vengano sensibilizzati i Comuni stessi per non ripetere nell'estrazione successiva gli errori commessi.

¹⁵ Una buona proxy delle modalità d'uso dell'applicativo è rappresentata dalla categoria di informatizzazione cui appartiene il Comune. Il 35,6 per cento dei Comuni non è stato in grado di produrre il file dei nominativi con le specifiche Istat, registrandoli con il programma sviluppato dall'Istat; il 63,6 per cento ha estratto i nominativi dalle anagrafi informatizzate producendo un file secondo le specifiche Istat, utilizzando Gefi solo per registrare le informazioni sulle procedure di estrazione e generare il file da inviare all'Istat.

¹⁶ Le percentuali di errore risultano molto più elevate rispetto a quelle registrate nel corso della precedente estrazione. Ciò è da attribuire al fatto che, a partire dal 2004, è stata distribuita ai Comuni campione una nuova versione del software applicativo Gefi che prevede un nuovo tracciato record a seguito dell'introduzione di ulteriori informazioni da registrare.

¹⁷ Si tratta di nuove tipologie di errore rispetto a quelle tradizionalmente commesse dovute alle innovazioni introdotte nel 2004. Nel corso del 2003, infatti, i principali errori commessi hanno riguardato il Cap, la data di nascita e il sesso dell'intestatario della scheda di famiglia, il numero di quartine e famiglie fornite.

Capitolo 5 - La tecnica di indagine e il monitoraggio del lavoro sul campo

5.1 – Premessa

Allorché si cerca di definire la nozione di qualità, in ambito statistico, ci si rende immediatamente conto che si ha a che fare con un concetto ancora estremamente articolato. In letteratura ci si riferisce essenzialmente alla qualità del prodotto o, meglio, alla qualità dell'informazione statistica prodotta. Le dimensioni considerate, generalmente, si esprimono in termini di attendibilità ed accuratezza, cui fa espressamente riferimento Eurostat allorché offre indicazioni ai paesi membri in termini di qualità dell'indagine sulle forze di lavoro (Doc. D1/Empl/15/2004). L'attendibilità dei dati esprime un livello di qualità sempre misurabile mediante l'errore medio di stima. L'accuratezza, generalmente intesa, esprime il grado di corrispondenza tra la stima ottenuta dall'indagine e il vero valore della caratteristica in oggetto della popolazione obiettivo. In tutte le indagini campionarie l'accuratezza è funzione di due tipologie di errore, gli errori campionari e gli errori non campionari presenti.

Gli errori non campionari sono particolarmente legati alle procedure di misurazione del fenomeno oggetto di studio e possono insorgere ad ogni passo del processo di produzione delle informazioni statistiche. Essi dipendono, quindi, dal processo di rilevazione, dai soggetti che vi prendono parte, dalle tecniche di rilevazione e di trattamento dei dati e sono determinati da tutte le irregolarità e disattenzioni che generano uno scostamento tra ciò che è stato progettato e ciò che invece si verifica nella pratica delle indagini. Sebbene siano generalmente evitabili condizionano immancabilmente, in tutto o in parte, i risultati dell'indagine. Tenendo conto dei vari passi di un'indagine, rientrano tra gli errori di natura non campionaria gli errori di mancata risposta dovuti alla difficoltà di contattare le unità di indagine per cause accidentali o al rifiuto dell'intervistato di cooperare all'indagine o di fornire alcune delle informazioni richieste o all'impossibilità di partecipare all'indagine. In alcuni casi, la mancata risposta totale può essere associata a carichi di lavoro troppo pesanti, ad esempio nel caso della nuova indagine continua sulle forze di lavoro, numero eccessivo di interviste settimanali assegnate, aree territoriali di competenza del rilevatore troppo vaste, eccessiva durata delle interviste, eccetera.

Tra gli errori non campionari vanno poi considerati gli errori di misurazione: in questo caso l'informazione è disponibile ma non è corretta. Gli errori di misurazione possono dipendere dall'intervistatore o dalla reticenza del rispondente, da una codifica sbagliata eccetera..

La visione della qualità dei dati è stata ampliata considerando ulteriori e non meno importanti dimensioni della stessa: la rilevanza, la tempestività, l'accessibilità e trasparenza, la comparabilità, la coerenza e completezza (Eurostat Doc. D1/Empl/15/2004). La rilevanza o pertinenza indica la capacità dei dati di soddisfare gli obiettivi conoscitivi dell'indagine e la loro concreta utilizzabilità da parte delle statistiche ufficiali e di quanti altri vogliono servirsene. La tempestività si riferisce alla possibilità di disporre di dati aggiornati in tempi brevi rispetto al momento di riferimento degli stessi. Le informazioni statistiche possono essere considerate tempestive se la diffusione avviene in tempi congrui in rapporto al fenomeno osservato ed alle esigenze degli utilizzatori. Minore è il tempo, maggiore è la validità e l'utilità delle informazioni. L'accessibilità riguarda la possibilità per gli utilizzatori di entrare in possesso o di utilizzare i dati in relazione alle proprie finalità. La trasparenza riguarda la disponibilità di documentazione adeguata a descrivere le caratteristiche dell'indagine e le fasi del processo produttivo.

La comparabilità riguarda la possibilità di effettuare confronti omogenei nel tempo e nello spazio riguardanti il fenomeno di interesse. Il grado di comparabilità è influenzato da tutti i cambiamenti che possono avvenire, nel tempo, nello spazio, nonché relativamente alle definizioni e alle scelte operative connesse al processo produttivo. La coerenza fa riferimento alla possibilità di ricorrere all'uso di più fonti statistiche anche con diversa frequenza senza che ciò comporti informazioni contraddittorie.

In questa ottica la qualità dei dati statistici non può essere misurata solo in termini di qualità del prodotto e dell'errore totale ma deve essere valutata anche in termini di qualità del processo analizzando tutte le fasi che generano i dati stessi.

È da considerare inoltre che molteplici sono state in questi anni le trasformazioni che hanno interessato e continuano ad interessare il processo di produzione di informazioni statistiche. Tali trasformazioni hanno riguardato, come precedentemente anticipato, in particolare lo sviluppo di banche dati e di tecniche informatiche come pure il ricorso a tecniche di campionamento e di raccolta dei dati sempre più complesse.

Il processo di produzione consiste, quindi, sempre più spesso in una molteplicità di scelte operative che possono influire sui risultati finali e la qualità è progressivamente diventata tema centrale nella pianificazione di una indagine.

Un elemento che attualmente accomuna le logiche e gli strumenti adottati per valutare la qualità riguarda, infatti, l'orientamento ad una attenzione sempre maggiore per i singoli sottoprocessi di una indagine considerando per ciascuno di essi l'organizzazione, i soggetti coinvolti, i singoli output eterogenei e parziali. A partire da queste considerazioni, la nuova indagine sulle forze di lavoro è stata progettata analizzandone il processo produttivo, in termini di attività combinate ed evidenziandone le criticità. Ciò ha reso possibile la valutazione di tutte le potenziali fonti di errore mettendo a punto un complesso sistema di controlli per ciascuna fase dell'indagine al fine di valutare, misurare e prevenire tutti i possibili errori dovuti alle scelte metodologiche, agli aspetti operativi o, ancora, alla performance di tutti gli operatori coinvolti.

Il sistema di controlli di qualità rispetto ai modi, ai tempi e agli obiettivi è diviso principalmente in tre fasi:

- la fase dei “controlli preventivi”, intesa come un insieme di azioni che è possibile attuare prima della raccolta dei dati per evitare l'insorgere dell'errore (ad esempio controlli sull'effettivo invio della lettera di preavviso per la sensibilizzazione dei rispondenti o i controlli sulle modalità di estrazione del campione);
- la fase dei “controlli in corso d'opera”, intesa come un insieme di azioni predisposte al fine di individuare gli errori che insorgono durante il processo di produzione. Esse, generalmente, riguardano il rispetto della metodologia, delle procedure e dei tempi stabiliti per l'esecuzione delle interviste con interventi correttivi volti a risolvere i problemi e a minimizzare le distorsioni (ad esempio l'help desk per i rilevatori, il controllo delle trasmissioni, la valutazione degli indicatori di field);
- la fase dei “controlli successivi”, intesa come un insieme di azioni predisposte per valutare il livello di errore non campionario (ad esempio determinazione ed analisi degli indicatori di survey o ancora controlli successivi sulla qualità delle interviste e sulla effettiva realizzazione delle stesse tramite interviste di controllo).

La presenza di caratteristiche peculiari quali la raccolta dei dati nel corso di tutte le settimane dell'anno, i ritorni sulla stessa unità di rilevazione per più occasioni di intervista, una tecnica mista di rilevazione (*computer assisted*), i tempi ristretti di elaborazione e trattamento dei dati, rende complessa la definizione di misure della qualità del processo di produzione della rilevazione continua sulle forze di lavoro.

Il controllo di qualità della nuova rilevazione sulle forze di lavoro si basa, quindi, su una dettagliata documentazione di ogni operazione del processo di produzione dei dati e sulla predisposizione, per ciascuna di esse, di indicatori di tipo quantitativo coerente con gli standard internazionali per le indagini ufficiali (Aapor standard definitions). Il sistema di indicatori consente permanentemente di controllare le diverse fasi del processo stesso, di verificare le scelte attuate dal punto di vista metodologico, di standardizzare le procedure nelle varie fasi della rilevazione e di individuare i punti critici del processo. Gli indicatori sono costantemente e approfonditamente analizzati a livello territoriale (ripartizioni geografiche, regioni e comuni campione), cosicché qualsiasi cambiamento apportato nel processo di produzione possa essere immediatamente valutato e nel caso di situazioni problematiche possano essere tempestivamente definite misure correttive idonee a minimizzare gli errori che si possono incontrare in itinere.

A tale proposito l'Istat, sia per il Capi che per il Cati, ha fissato una serie di regole relativamente alla gestione, trasmissione, ed esecuzione delle interviste, che devono essere rigorosamente osservate. L'intero sistema dell'indagine a partire dall'acquisizione dei nominativi delle famiglie campione fino alla predisposizione dei files di output è gestito tramite procedure automatiche che ne facilitano il tempestivo e costante controllo. Questo approccio offre l'indubbio vantaggio di identificare responsabilità di processo o di segmenti dello stesso, in alternativa o congiuntamente alle classiche responsabilità di struttura, consente di sviluppare capacità di analisi dei fenomeni gestionali in un'ottica di tipo trasversale, superando le barriere poste dagli stessi assetti organizzativi per garantire il miglior risultato possibile. Esso permette non solo di riconoscere le attività critiche che potrebbero inficiare i risultati e rendere il processo di produzione delle informazioni statistiche lento, complicato, ridondante ma anche di ridisegnarlo in una logica di miglioramento incrementale e continuo.

5.2 - Il sistema Capi-Cati

La Rcf1 prevede che ogni unità campionaria sia soggetta ad un ciclo di quattro interviste con l'utilizzo di una tecnica mista di rilevazione *computer assisted*: Capi-Cati.

Il disegno organizzativo dell'indagine prevede che la prima intervista venga effettuata faccia a faccia da un rilevatore Istat presso il domicilio della famiglia con tecnica Capi, cioè con l'ausilio di un personal computer che gestisce il questionario elettronico. Le successive interviste vengono realizzate con intervista telefonica assistita da computer (tecnica Cati) da rilevatori di una società esterna, tranne nel caso di famiglie senza telefono che vengono reintervistate tramite Capi¹. Il ricorso a tecniche di rilevazione *computer assisted* è sembrata la risposta migliore per far fronte alle innovazioni metodologiche ed organizzative connesse ad un'indagine continua. L'uso combinato delle due tecniche ha permesso di sfruttare i vantaggi di entrambe. Nel corso della prima intervista, con tecnica Capi, il contatto diretto che si instaura tra intervistatore e intervistato favorisce la collaborazione del rispondente, limitando le mancate risposte totali, e prepara il terreno per le successive interviste. Conquistata la disponibilità della famiglia, l'intervista telefonica con tecnica Cati comporta sia la limitazione della molestia statistica sull'unità campionaria sia una notevole riduzione dei costi economici.

¹ L'invio di una lettera del Presidente dell'Istat informa le famiglie dell'imminente intervista faccia a faccia o telefonica.

Al criterio generale, prima intervista Capi e successive interviste Cati, si affiancano ulteriori regole al fine di gestire particolari situazioni.

L'attività di rilevazione si interrompe in alcune settimane dell'anno in cui la reperibilità delle famiglie è piuttosto scarsa. In particolare, l'interruzione delle interviste è prevista per le prime tre settimane di agosto e per il periodo tra Natale e capodanno². Le interviste da svolgersi durante l'interruzione vengono recuperate in un periodo successivo. In tale frangente, per evitare il sovraccarico della rete di rilevazione Capi, che altrimenti si troverebbe a gestire un numero troppo elevato di interviste con il rischio di un aumento del tasso di caduta, tutte le famiglie in prima intervista che dispongono di telefono vengono recuperate con tecnica Cati.

Il ricorso al Cati avviene anche in altre situazioni. Può accadere difatti che alcune interviste Capi non vengano assegnate dagli Uffici territoriali perché occorre smaltire carichi di lavoro pregressi o per assenze temporanee di uno o più rilevatori. In questo caso, l'utilizzo del Cati permette di alleggerire la pressione sulla rete dei rilevatori garantendo un'articolazione più equilibrata dell'intero processo.

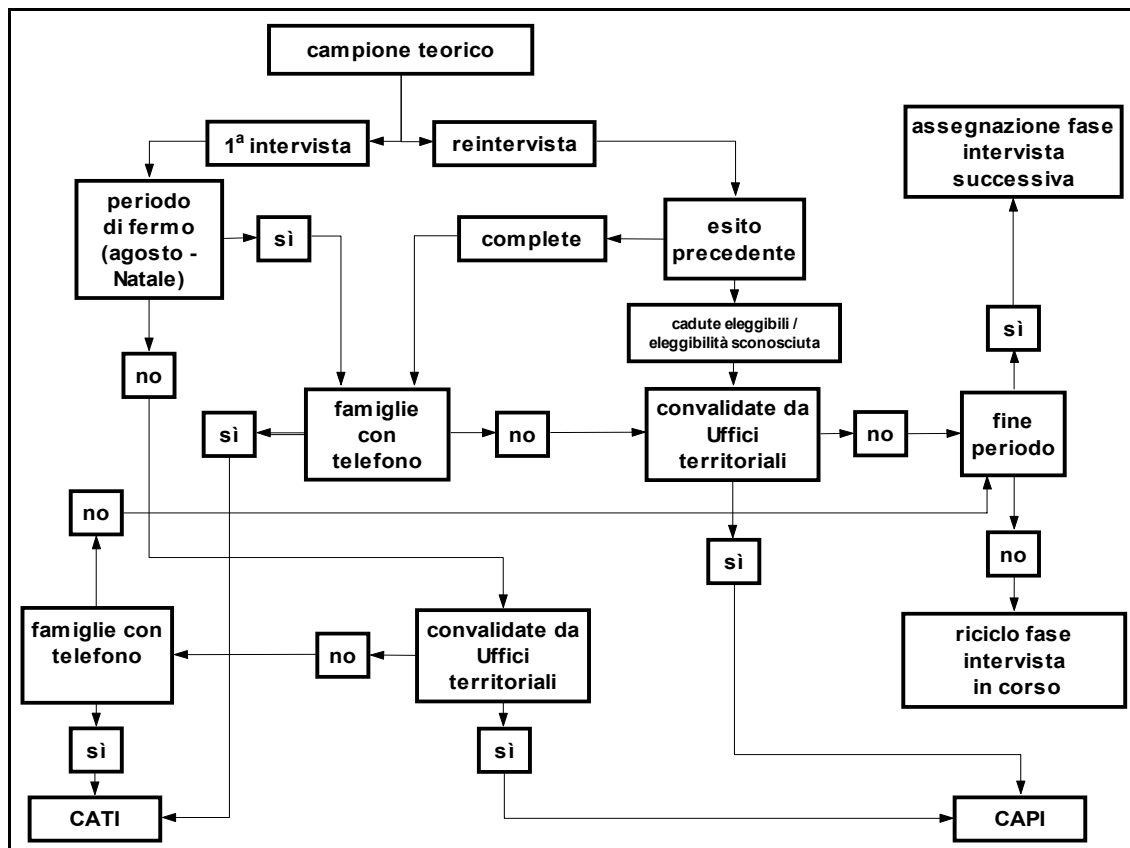
In caso di caduta della famiglia, la gestione delle sostituzioni è diversa per le due tecniche. In Capi, sia in prima sia nelle successive interviste, la famiglia caduta è sostituita automaticamente dalla famiglia successiva della quartina. In Cati, invece, la sostituzione è ammessa solo in caso di prima intervista. Per tale motivo, un'ultima regola è stata prevista in caso di caduta di un'intervista Cati. In tale occasione, la successiva intervista viene condotta con tecnica Capi per recuperare l'unità campionaria. Al fine di contenere le sostituzioni, le famiglie cadute sono state divise in tre grandi gruppi: eleggibili, non eleggibili, con eleggibilità sconosciuta. Le famiglie eleggibili sono quelle che costituiscono l'universo di riferimento dell'indagine. Le famiglie non eleggibili in fase di reintervista sono quelle che hanno perso l'eleggibilità (ad esempio, per trasferimento all'estero, per decesso). Le famiglie con eleggibilità sconosciuta rappresentano un gruppo residuale per le quali non è stato possibile accertare l'eleggibilità, ad esempio per irreperibilità sul territorio³. Quando si verifica la caduta di una famiglia eleggibile o con eleggibilità sconosciuta viene ricontattata la stessa famiglia; quando si tratta invece di una famiglia non eleggibile viene contattata direttamente la famiglia sostitutiva.

A partire dalle famiglie del campione teorico, lo schema seguente riporta i criteri di individuazione della tecnica di rilevazione nel caso della prima e delle successive interviste.

² Nei periodi indicati i rilevatori Capi generalmente fruiscono delle ferie.

³ Con specifico riferimento alla caduta di una famiglia eleggibile, l'unica eccezione in cui si contatta direttamente la famiglia sostituita è rappresentata dal trasferimento in un altro Comune di quella eleggibile.

Figura 5.1 - Individuazione della tecnica di rilevazione



Acquisiti dalla banca dati i nominativi delle famiglie da intervistare, il sistema Capi effettua una prima assegnazione teorica tra i rilevatori attivi. Prima che i nominativi vengano inviati ai rilevatori, nell'arco dei cinque giorni successivi gli Uffici territoriali verificano l'effettiva situazione sul territorio e convalidano le assegnazioni. Con frequenza settimanale, il sistema inoltra ai rilevatori i nominativi⁴.

Ricevuti i nominativi, gli intervistatori hanno a disposizione sei settimane per la loro gestione: la prima per fissare gli appuntamenti, la seconda per effettuare la gran parte delle interviste, le rimanenti (cosiddette settimane di coda) per completare le interviste assegnate e non ancora realizzate⁵. Un'ulteriore settimana è infine dedicata all'invio dei dati. Il sistema a questo punto provvede a chiudere automaticamente il ciclo relativo ad una settimana di riferimento. Inizialmente si era scelto un numero di settimane di coda più basso, ma le sperimentazioni hanno mostrato una riduzione significativa del tasso di completezza.

⁴ Gli Uffici territoriali danno priorità alle interviste in scadenza, cioè a quelle che provengono dai periodi di interruzione dell'attività Capi o a quelle non distribuite per vari motivi nelle settimane precedenti. Gli Uffici possono inoltre ridistribuire i carichi di lavoro riassegnando parzialmente o totalmente dei nominativi precedentemente attribuiti. Il sistema provvede a trasferire ai rilevatori di nuova assegnazione i nominativi eliminandoli dal/dai rilevatore/i a cui erano precedentemente assegnati. Le famiglie distribuite ai rilevatori rappresentano la base degli indicatori di *field* prodotti per effettuare il monitoraggio della performance degli intervistatori.

⁵ Per le ultime 2 settimane del trimestre, invece, il periodo di intervista previsto è di due settimane. La scelta di differenziare il periodo di rilevazione nasce dall'esigenza di ridurre i tempi di diffusione delle stime trimestrali.

Data la centralizzazione del database, le operazioni di assegnazione verso il sistema Cati sono più semplici. Le interviste vengono infatti assegnate ad un unico rilevatore virtuale, il servente Cati della ditta in service che acquisisce, in maniera continuativa attraverso un collegamento telematico con il Centro di coordinamento, i nominativi delle famiglie del campione e i dati delle eventuali precedenti interviste. Per la rilevazione Cati non è prevista la settimana per fissare gli appuntamenti. Ciò significa che le interviste iniziano la settimana successiva a quella di riferimento dei dati. Le famiglie rimangono quindi in rilevazione per cinque settimane⁶.

5.3 - Il monitoraggio della qualità del lavoro sul campo

Il monitoraggio globale della Rilevazione continua forze di lavoro è molto articolato in quanto complessa è la struttura organizzativa dell'indagine. Nel processo produttivo dei dati sono infatti coinvolti una pluralità di soggetti ed enti: il servizio centrale dell'Istat, preposto alla progettazione ed impostazione metodologica dell'indagine e degli strumenti di rilevazione, che coordina l'intero processo di produzione, i 1246 Comuni che provvedono all'estrazione dei nominativi delle famiglie campione dalle liste delle famiglie residenti nel Comune, le 103 camere di commercio con compiti essenzialmente organizzativi, gli Uffici territoriali dell'Istat (dove lavorano per l'indagine un referente e i coordinatori di rete) cui è affidata la sensibilizzazione e l'assistenza tecnica ai Comuni e alle famiglie campione, la formazione e l'assistenza ai rilevatori, la supervisione delle operazioni sul campo, la società di service che effettua le interviste telefoniche, i 312 rilevatori della rete di rilevazione Capi, le 75 mila famiglie che effettuano l'intervista ogni anno, e gli utenti finali che utilizzano i dati elaborati.

Considerando esclusivamente la fase di rilevazione sul campo da monitorare, i flussi di informazioni che si instaurano tra i vari soggetti coinvolti sono molteplici. Una volta ricevuti dai Comuni campione i nominativi delle famiglie selezionate, che rappresentano il campione teorico delle famiglie da intervistare, il sistema informativo-informatico denominato Sigif (Sistema gestione indagini famiglie) provvede ad assegnare la tecnica di rilevazione (Capi o Cati) alle famiglie del campione teorico. Le famiglie alle quali è stata assegnata la tecnica di indagine Cati vengono distribuite dal sistema Sigif alla società in service che si occupa di effettuare le interviste telefoniche alle famiglie; da qui, poi, parte il flusso informativo in direzione opposta contenente le interviste complete e non complete Cati. Più complessa è la distribuzione delle interviste alle quali è stata assegnata la tecnica di rilevazione Capi. Infatti, tali famiglie vengono in primo luogo "inviati" dal sistema Sigif agli Uffici Regionali Istat, i quali si preoccupano di convalidare/assegnare le varie famiglie ai rilevatori di competenza; successivamente il sistema Sigif, ricevute le assegnazioni delle interviste da parte degli Uffici Regionali, si preoccupa di trasmettere/distribuire i nominativi ai vari rilevatori Capi, i quali a loro volta effettuano le interviste alle famiglie e ritrasmettono direttamente al sistema Sigif le interviste realizzate. Un sistema di monitoraggio dell'intera fase di raccolta dati non può non tener conto di tutte le interrelazioni che intercorrono tra questi soggetti.

Lo sforzo maggiore compiuto nella costruzione di un sistema di indicatori utili a monitorare tutta la fase di rilevazione nel suo complesso è stato quello di integrare gli esiti provenienti dalla tecnica di rilevazione Capi con quelli provenienti dalla tecnica di rilevazione Cati, il tutto coerentemente con gli standard ufficiali nazionali ed internazionali per le indagini face to face e telefoniche, cioè con il Sistema Informativo di Documentazione delle Indagini (Sidi) a livello

⁶ L'assegnazione delle interviste ai rilevatori Cati è automatizzata e segue criteri di casualità. La rilevazione viene effettuata tutti i giorni, con diversi turni di lavoro, a partire dalle 12.30 fino alle ore 21.00. I due turni principali, che prevedono il maggior numero di rilevatori in contemporanea, sono quelli che vanno dalle 12.30 alle 15.30 e dalle 18.30 alle 21.00. Precedenti sperimentazioni hanno evidenziato che queste due fasce orarie sono quelle in cui è massima la reperibilità delle famiglie.

nazionale e gli standard definiti da *The American Association for Public Opinion Research* (Aapor) a livello internazionale. Infatti la letteratura internazionale riferisce di proposte e sforzi effettuati per la standardizzazione degli esiti definitivi (*survey outcome codes*), ma separatamente per indagini face to face e indagini telefoniche, mentre è molto scarsa relativamente a tentativi di standardizzazione per indagini con tecnica mista.

Per illustrare il percorso seguito nella concettualizzazione del sistema di monitoraggio, si è sintetizzato nello schema gerarchico riportato in figura 5.2 tutti i vari passaggi del campione teorico dal momento successivo alla sua acquisizione fino all'intervista o meno delle famiglie campione. A tal proposito è utile dare la definizione di quartina. I Comuni campione, che si occupano dell'estrazione delle famiglie da intervistare, forniscono all'Istat delle quartine di famiglie, ove la prima è denominata famiglia base e le tre successive sono dette famiglie suppletive. Il rilevatore deve cercare di intervistare la famiglia base, qualora ciò non fosse possibile, il rilevatore deve procedere alla sua sostituzione e può procedere alla sua intervista, e così via fino all'esaurimento di tutte le famiglie della quartina. Tutte le quartine del campione teorico vengono sottoposte al processo di assegnazione della tecnica di rilevazione da parte del sistema Sigif. Il criterio generale per l'assegnazione della tecnica di indagine è quello di ricorrere alla tecnica Capi nel caso di prima intervista e alla tecnica Cati nel caso di reintervista⁷.

Una volta assegnata la tecnica di rilevazione, le quartine seguono strade diverse, più articolata nel caso di intervista Capi, più semplice nel caso di intervista Cati.

Il sistema Capi acquisisce settimanalmente i nominativi delle famiglie (e relative sostitute) da intervistare e provvede a suddividerli tra i rilevatori attivi. Prima che sia effettuato l'invio definitivo ai rilevatori, il sistema avverte i referenti di indagine presso gli Uffici territoriali dell'Istat quali, a loro volta, provvedono alla convalida delle quartine verificando con i rilevatori la fattibilità delle interviste in ogni settimana specifica. Alcune quartine possono non essere convalidate per problemi di carico di lavoro dei rilevatori.

Al livello successivo del percorso Capi, le quartine convalidate dai referenti degli Uffici Regionali vengono distribuite ai rilevatori. Può accadere che qualche quartina non venga distribuita per problemi nella trasmissione via modem.

Le quartine distribuite ai rilevatori rappresentano il punto di partenza per il monitoraggio di *field* come performance degli intervistatori.

Le quartine distribuite Capi si possono distinguere a loro volta tra "quartine toccate" e "quartine non toccate", ove per "quartina toccata" si intende una quartina per la quale il rilevatore ha effettuato almeno un'azione minima che la riguarda (scaturita ad esempio dal semplice click del mouse, oppure dalla ricerca del numero di telefono).

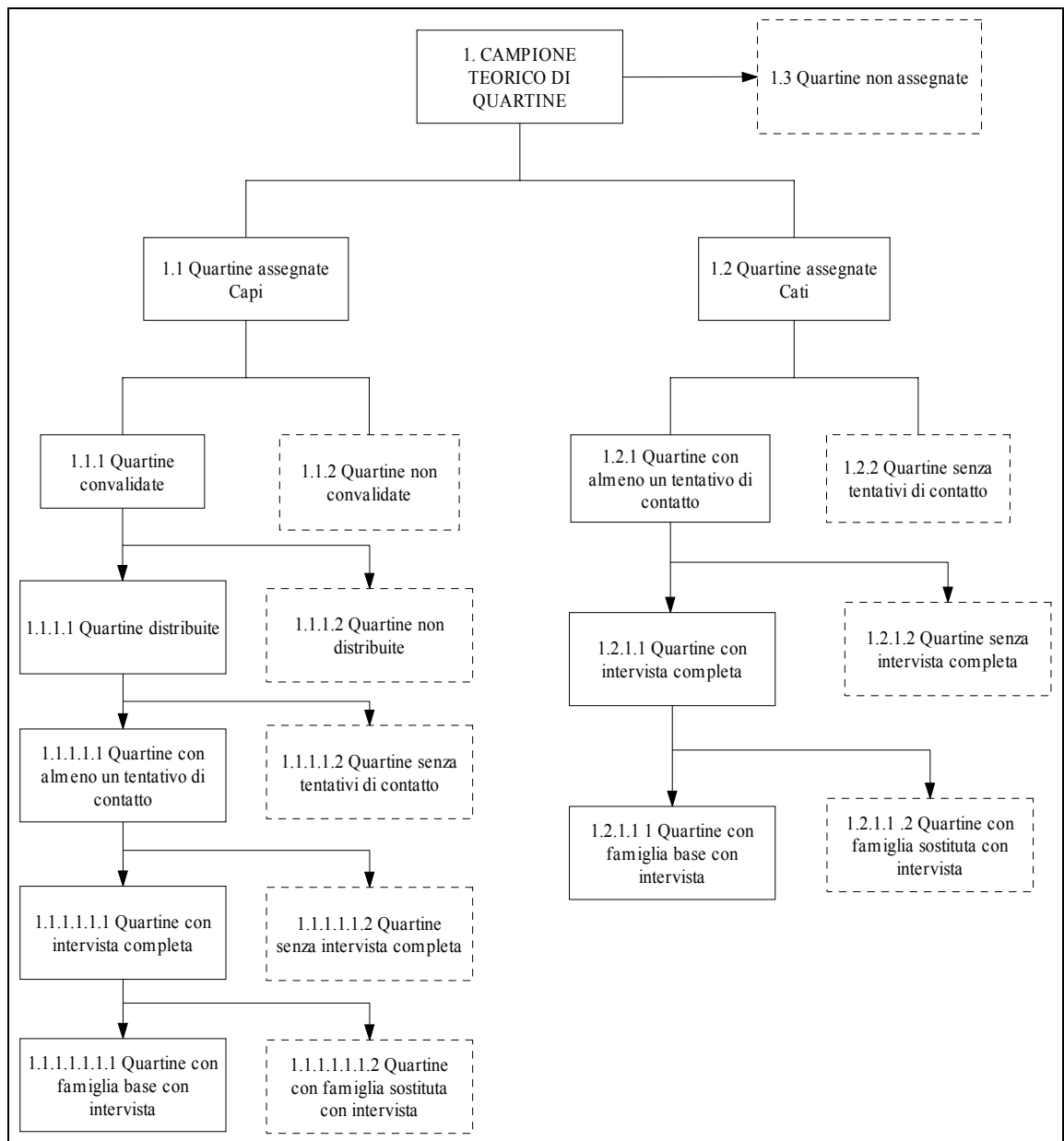
All'ultimo livello del percorso Capi, tra le "quartine toccate" possiamo distinguere le quartine con intervista completa dalle quartine che invece non hanno prodotto alcuna intervista completa, dando luogo ad una caduta della famiglia che compone la quartina stessa.

⁷ In alcuni casi il sistema non può più assegnare la tecnica di rilevazione: si tratta delle cosiddette "quartine bucate". Le quartine bucate sono quelle quartine per le quali risultano cadute tutte le famiglie che la compongono: il rilevatore ha provato ad intervistare senza successo sia la famiglia base, sia le famiglie suppletive comportando l'inutilizzabilità dell'intera quartina e conseguentemente la riduzione della dimensione campionaria.

In sintesi, le quartine possono essere raggruppate in:

- **Quartine assegnate**, cioè la parte del campione teorico che deve essere intervistato nella specifica occorrenza temporale. È il numero di quartine da intervistare (inteso come l'insieme di quattro famiglie di cui una base e tre sostitutive) che viene assegnato rispettivamente agli intervistatori del Capi e del Cati.
- **Quartine bucate**, cioè la parte del campione teorico non intervistabile nella specifica occorrenza temporale. E' il numero di quartine (inteso come l'insieme di quattro famiglie di cui una base e tre sostitutive) per le quali non sono più disponibili ulteriori famiglie sostitutive poiché, alla precedente occasione di intervista, si sono chiuse tutte con un esito definitivo di caduta.
- **Quartine convalidate**, cioè la parte delle quartine assegnate agli intervistatori del Capi nella specifica occorrenza temporale. E' il numero di quartine (inteso come l'insieme di quattro famiglie di cui una base e tre sostitutive) che viene sottoposto all'approvazione degli Uu.rr. nella fase di convalida dei carichi di lavoro degli intervistatori Capi.
- **Quartine distribuite**, cioè la parte delle quartine convalidate dagli Uu.rr. per gli intervistatori del Capi nella specifica occorrenza temporale. E' il numero di quartine (inteso come l'insieme di quattro famiglie di cui una base e tre sostitutive) che vengono messe in trasmissione agli intervistatori Capi
- **Quartine toccate**, cioè la parte delle quartine trasmesse agli intervistatori del Capi nella specifica occorrenza temporale. È il numero di quartine (inteso come l'insieme di quattro famiglie di cui una base e tre sostitutive) che l'intervistatore inizia a lavorare a partire dalla semplice consultazione delle informazioni sulla famiglia fino alla registrazione dell'esito del contatto.

Figura 5.2 - Scomposizione del campione teorico



Molto meno articolato risulta essere il percorso Cati. Infatti, le quartine, alle quali è stata assegnata la tecnica di rilevazione Cati, vengono tutte trasmesse alla società in service che si occupa di effettuare le interviste telefoniche. Quindi i due percorsi, Capi e Cati, risultano coincidenti nella parte finale, per cui anche nel percorso Cati possono risultare delle “quartine toccate” e delle “quartine non toccate” e delle quartine con intervista completa e delle quartine senza intervista completa.

Sempre nella fase di concettualizzazione del sistema di indicatori per il monitoraggio, va tenuto presente che, da un punto di vista teorico, ogni quartina partecipa a quattro occasioni di

indagine, in ciascuna delle quali deve dar luogo a quattro interviste complete nel corso di 15 mesi prima di uscire definitivamente dal campione.

La *wave di quartina* individua la fase del “ciclo di vita” della quartina, cioè l’occasione di indagine in cui si trova la quartina: una wave di quartina pari a 3 significa che la quartina sta svolgendo la terza occasione di intervista e che quindi rimane ancora un’occasione di indagine prima di concludere il suo ciclo. *La wave di famiglia*, invece, è un attributo esclusivo della famiglia e rappresenta il numero progressivo di intervista della famiglia: una wave di famiglia pari a 2 significa che una certa famiglia sta effettuando la seconda intervista che la riguarda.

Se non ci sono sostituzioni all’interno della quartina, la wave di famiglia coincide con la wave di quartina. Se, invece, ci sono sostituzioni, allora la wave di famiglia è minore o uguale della wave di quartina, ad esempio se la famiglia base di una quartina effettua la prima e seconda intervista e viene sostituita dalla prima famiglia suppletiva alla terza occasione di indagine, la wave di quartina sarà 1, 2, 3 e 4 rispettivamente per ciascuna occasione di indagine; la wave di famiglia per la famiglia base sarà 1 e 2 rispettivamente per la prima e seconda occasione di indagine, mentre la wave di famiglia per la prima famiglia suppletiva nella terza e quarta occasione di indagine sarà rispettivamente 1 e 2, essendo rispettivamente la prima e la seconda intervista per la famiglia suppletiva. Il ciclo di vita della famiglia è legato al ciclo di vita della quartina, nel senso che una volta che la quartina raggiunge la wave di quartina 4, anche la famiglia in quell’occasione finisce il suo ciclo di vita indipendentemente dalla wave di famiglia cui è giunta.

5.3.1 - Gli standard utilizzati per la definizione degli esiti

Come anticipato, per il calcolo degli indicatori sono stati definiti in primo luogo gli esiti di indagine (*survey outcome codes*) considerando gli standard dettati da Aapor, che descrivono in modo esaustivo e semplice gli esiti definitivi dell’indagine e permettono di calcolarne facilmente i tassi finali.

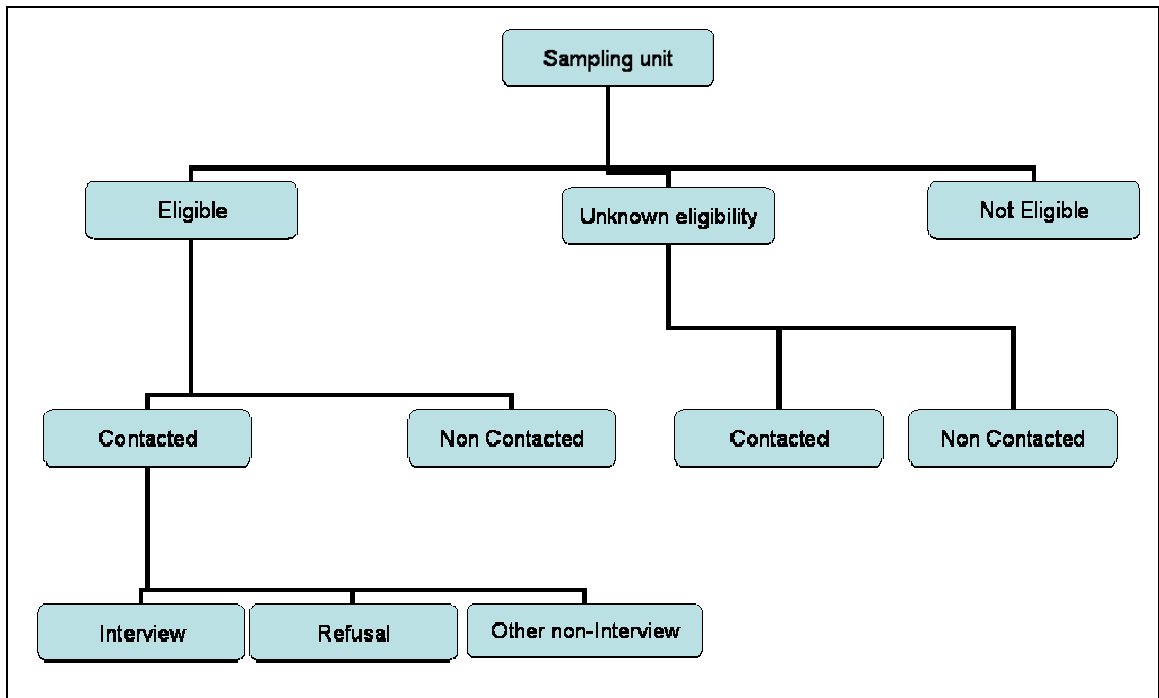
Tuttavia le indicazioni Aapor non sono state direttamente applicate, per tre motivi fondamentali:

- la tecnica di indagine utilizzata è una tecnica mista Cati-Capi. Tra le differenti *waves* la tecnica di intervista è diversa (per la prima *wave* l’intervista è Capi, ma in momenti di recupero può anche essere Cati, viceversa per le *waves* successive l’intervista è Cati, ma per le famiglie senza numero di telefono viene svolta di persona). Aapor indica separatamente le tipologie di esito per le differenti tipologie di indagine;
- tutti i componenti della famiglia di fatto devono essere intervistati (e non solo uno di essi all’interno della famiglia estratta, come previsto da Aapor);
- il campione di base può essere modificato dalle sostituzioni.

Per questi motivi gli standards Aapor sono stati adattati alle esigenze dell’indagine, esigenze che sono in realtà comuni a molte indagini effettuate dall’Istat e ad altri Istituti nazionali di statistica europei.

Come già proposto dall’Istituto Statistico Inglese, gli esiti finali dei contatti con le famiglie sono stati organizzati in maniera gerarchica (Figura 5.3):

Figura 5.3



Il primo livello gerarchico divide le famiglie a seconda dell'eleggibilità:

- Eleggibili (*Eligible*)
- Eleggibilità incerta (*Unknown eligibility*)
- Non eleggibili (*Not eligible*)

Il secondo livello divide le famiglie contattate da quelle non contattate:

- Eleggibili (*Eligible*)
 - Contattati (*Contacted*)
 - Non Contattati (*Non Contacted*)
- Eleggibilità incerta (*Unknown eligibility*)
 - Contattati (*Contacted*)
 - Non Contattati (*Non Contacted*)

Il terzo riguarda i singoli esiti relativi alle famiglie eleggibili e contattate. A partire da tale livello è possibile avere dei differenti livelli di dettaglio negli esiti a seconda della tecnica di rilevazione utilizzata.

Durante la prima fase di contatto con la famiglia l'eleggibilità è sconosciuta qualora non si riesca a raccogliere elementi utili ai fini della corretta classificazione. Un nominativo inserito nel campione è eleggibile quando si è accertato, con la persona interessata o con altre persone che forniscono informazioni certe, che quel nominativo corrisponde alla famiglia selezionata. Nelle *wave* successive per le famiglie con eleggibilità certa, bisogna assicurarsi della permanenza dello status di eleggibilità, in quanto potrebbero essere intervenuti mutamenti della famiglia (uscita del PR e del coniuge, trasferimento all'estero) tali da rendere non più eleggibile la stessa.

5.3.2 - La scheda contatti per la registrazione degli esiti

Recepite e riviste le raccomandazioni di Aapor, un problema che si è posto è stato quello di disporre di uno strumento che consentisse di assegnare gli esiti definitivi a seguito dei contatti con le unità campionarie. Considerati i vincoli spaziali e temporali, nonché la complessità della rilevazione, dovuti alla presenza di una tecnica di rilevazione mista, si è reso necessario dotarsi di uno strumento che, sfruttando la tecnica computer assisted, permettesse di gestire i tentativi di contatto con le famiglie, determinare l'eleggibilità delle stesse, attribuire gli esiti definitivi delle unità campionarie e gestire le sostituzioni. È stato perciò sviluppato un modello di una scheda contatti adeguata sia nel caso di intervista *face to face* sia telefonica, in modo tale che potessero essere raccolte tutte le informazioni necessarie per ricostruire lo schema riportato in figura 5.2.

Sono state poi predisposte concretamente quattro schede contatti: una per la prima intervista Capi, una per le successive reinterviste Capi, una per la prima intervista Cati e una per le reinterviste Cati. La necessità di ricorrere a tale molteplicità di schede è discesa dal fatto che risultano differenti, a seconda della tecnica di rilevazione e dell'occasione di intervista per la famiglia, le modalità di approccio con le famiglie stesse, le regole che determinano l'assegnazione degli esiti definitivi alle famiglie e la gestione dei contatti con le famiglie.

La scheda contatti si presenta come una guida strutturata per il rilevatore utile per gestire il suo rapporto con le famiglie, dalla presentazione del rilevatore e dell'indagine alla conclusione dell'intervista. Tale scheda consente di determinare l'eleggibilità della famiglia campione, attraverso una serie di domande chiave che permettono di stabilire se la famiglia intervistata coincide con la famiglia estratta dall'anagrafe comunale.

Con la scheda contatti vengono gestiti tutti i tentativi di contatto con la famiglia, sia telefonici sia a domicilio, registrandone gli esiti: "occupato", "non risponde", "fax", "numero inesistente", "segreteria telefonica" e "risponde qualcuno" nel caso di tentativo di contatto telefonico; "non c'è nessuno in casa", "indirizzo errato o inesistente" e "c'è qualcuno in casa" nel caso di visita a domicilio. Una volta instaurato un contatto voce con qualcuno, viene registrato in automatico, in base all'andamento della conversazione, l'esito del contatto.

La scheda contatti gestisce inoltre le sostituzioni delle famiglie cadute. Quando si verifica la caduta di una famiglia, ad esempio per rifiuto a partecipare all'indagine o perché non eleggibile, un messaggio di warning avvisa il rilevatore che la famiglia verrà sostituita; immediatamente gli viene proposta la rispettiva famiglia sostituita con la quale iniziare nuovi tentativi di contatto. Il rilevatore, effettuati tutti i tentativi di contatto previsti a livello metodologico, ha la possibilità di scegliere se sostituire o meno la famiglia in funzione dei forti vincoli temporali della rilevazione, valutando se rischiare di perdere l'intervista per continuare a cercare di contattare la famiglia del campione base o, al contrario, abbassare la qualità del dato intervistando subito una famiglia sostitutiva evitando però una mancata risposta totale. La scheda contatti, infine, svolge anche un'importante funzione di agenda: permette di gestire gli appuntamenti con le unità campionarie e le eventuali interruzioni temporanee delle interviste a seguito di un impegno improvviso della famiglia; avvisa con dei messaggi di warning qualora siano stati fissati più appuntamenti in uno stesso momento oppure qualora sia stato fissato un appuntamento troppo in là nel tempo o successivamente la chiusura del periodo di rilevazione.

La scheda contatti è stata ampiamente sotto osservazione durante tutta la fase sperimentale dell'indagine, attraverso uno stretto contatto con i rilevatori sia Capi che Cati, i quali hanno segnalato il verificarsi di eventuali situazioni non contemplate dallo strumento in esame.

5.3.3 - Gli esiti finali nella Rilevazione Rcfl con tecnica mista Capi-Cati

Alla fine del periodo dedicato alle interviste, devono avere un esito definitivo sia le famiglie base del campione (prima famiglia della quartina), sia tutte quelle utilizzate per sostituire le cadute durante la rilevazione. Per ogni quartina dunque vi sarà almeno una famiglia con esito definitivo, ma solo una famiglia può avere l'esito definitivo "intervista effettuata". L'assegnazione dell'esito può avvenire, sia per la tecnica Capi che per quella Cati, immediatamente prima del periodo dedicato alle interviste (come visto per le famiglie base delle quartine che non sono distribuite), sia durante (come esito derivante dai tentativi di contatto), sia dopo (con la chiusura degli esiti provvisori, e delle famiglie inserite nel processo di rilevazione senza tentativi di contatto).

Si analizzano di seguito i quattro gruppi principali di esiti familiari Aapor, descrivendo in dettaglio il contenuto di ciascuno di essi con riferimento all'indagine Rcfl.

Interviste

L'esito definitivo intervista (EF1), sia per la tecnica Capi che per la tecnica Cati, si divide in due gruppi:

Interviste (EF1)

Capi	Cati
1.1 intervista completa	1.1 intervista completa
1.2 intervista parziale	1.2 intervista parziale

In fase di progettazione dell'indagine è stato stabilito in quali casi accettare un'intervista incompleta. In particolare si considera valida un'intervista parziale quando l'intervistato ha risposto a tutto il questionario lasciando in sospeso solo alcune codifiche da fare.

Eleggibili non intervistate

Le famiglie eleggibili per le quali non è stata ottenuta un'intervista (EF2) vengono distinte in *contattate e non contattate*.

Le famiglie contattate non intervistate sono classificate come segue:

Eleggibili Contattate non intervistate (EF2.1)

Capi	Cati
2.1.1 rifiuto <ul style="list-style-type: none">• Rifiuto telefonico e al domicilio pre-lettera• Rifiuto telefonico e al domicilio post-lettera• Rifiuto solo a domicilio pre-lettera• Rifiuto solo a domicilio post-lettera	2.1.1 rifiuto <ul style="list-style-type: none">• Rifiuto telefonico pre-lettera• Rifiuto telefonico post-lettera
2.1.2 interruzione definitiva	2.1.2 interruzione definitiva
2.1.3 contatore esaurito con contatto <ul style="list-style-type: none">• CE per eccesso di tentativi con almeno un contatto (telefonico o a domicilio) per ottenere il 1° appuntamento o iniziare l'intervista• CE per eccesso di appuntamenti (telefonici o a domicilio) dopo il 1° appuntamento/inizio intervista• CE per eccesso di tentativi (telefonico o a domicilio) dopo il 1° appuntamento/inizio intervista	2.1.3 contatore esaurito con contatto⁸ <ul style="list-style-type: none">• CE per eccesso di tentativi con almeno un contatto per ottenere il 1° appuntamento o iniziare l'intervista• CE per eccesso di appuntamenti dopo il 1° appuntamento/inizio intervista• CE per eccesso di tentativi dopo il 1° appuntamento/inizio intervista
2.1.4 fine periodo di rilevazione con contatto <ul style="list-style-type: none">• FP con contatto ma senza appuntamento• FP con appuntamento• FP iniziata intervista	2.1.4 fine periodo di rilevazione con contatto <ul style="list-style-type: none">• FP con contatto ma senza appuntamento• FP con appuntamento• FP iniziata intervista
2.1.5 altro motivo non intervista <ul style="list-style-type: none">• Assenza prolungata• Malato grave	2.1.5 altro motivo non intervista <ul style="list-style-type: none">• Assenza prolungata• Malato grave

Per il Capi i “*rifiuti*” comprendono sia quei casi in cui c’è stato un preliminare contatto telefonico per prendere l’appuntamento per l’intervista e si abbia avuto un rifiuto, sia i casi di rifiuto direttamente a domicilio. È bene sottolineare che al rifiuto ottenuto per telefono deve comunque seguire un tentativo di contatto a domicilio.

Per il Cati i rifiuti comprendono solo quei casi in cui c’è stato un contatto telefonico con la famiglia, ma questa non ha accettato di concedere l’intervista. I rifiuti, in entrambe le tecniche di rilevazione, vengono ulteriormente scomposti in base al momento in cui avviene il rifiuto stesso: nella fase iniziale di contatto prima ancora che venga chiesto se è stata ricevuta la lettera di presentazione dell’indagine o dopo tale quesito. La lettera inviata dal Presidente dell’Istat rappresenta, infatti, un forte strumento di sensibilizzazione delle famiglie.

L’esito “contatore esaurito con contatto”, in entrambe le tecniche di indagine riguarda quei casi in cui ci sono stati contatti con la famiglia, ma alla fine sono stati effettuati il numero massimo dei tentativi previsti dalle regole di chiamata senza concludere l’intervista. Si ha contatore esaurito quando: una volta contattata la famiglia si sono susseguiti una serie di contatti senza nemmeno prendere il primo appuntamento per l’intervista; sono stati presi troppi appuntamenti per l’intervista; si è superato il numero massimo di tentativi (occupato, non risponde eccetera) dopo il primo appuntamento. È da sottolineare che se si tratta di una *wave di quartina* successiva alla prima e la tecnica utilizzata è quella Cati non sono previste regole per

⁸ Questo esito non è presente per le *wave di quartina* successive alla prima

determinare l'eccesso di tentativi ed i tentativi di contatto devono continuare fino al termine del periodo utile alle interviste. Di conseguenza in Cati l'esito "Contatore esaurito con contatto" non sarà presente ma la chiusura dei tentativi avviene imputando l'esito "Fine periodo di rilevazione con contatto".

Come anticipato le famiglie devono essere contattate in un determinato periodo temporale. Se alla fine di questo periodo la famiglia assegnata non ha nessun esito definitivo, ma è stata contattata, le verrà assegnato l'esito "Fine periodo di rilevazione con contatto".

Gli esiti che rientrano in "altro motivo non intervista" riguardano quelle situazioni in cui il rispondente è impossibilitato a rilasciare l'intervista. Essi includono il caso in cui risponde al telefono una colf o un'altra persona non eleggibile ma la famiglia non è reperibile nel periodo dedicato alle interviste e quelli in cui la mancata intervista è dovuta all'impossibilità fisica o mentale (malattia grave) del rispondente/i.

Le famiglie eleggibili non contattate e non intervistate sono classificate come segue:

Eleggibili non Contattate non Intervistate (EF2.2)

Capi	Cati
• 2.2.1 la famiglia vive nel Comune, ma non si sa dove	• 2.2.1 la famiglia vive nel Comune, ma non si sa il numero di telefono
• 2.2.2 la famiglia vive in un altro Comune italiano ⁹	• 2.2.2 la famiglia vive in un altro Comune italiano

In entrambi i casi sopraindicati, l'eleggibilità della famiglia è accertata, ma non può essere fatta l'intervista.

Eleggibilità sconosciuta

I casi di eleggibilità sconosciuta riguardano quelle situazioni in cui non si è riusciti ad appurare se il nominativo corrisponda realmente alla famiglia da intervistare. Questa classe include:

⁹ Nel caso di trasferimento in un altro Comune, la famiglia non può essere considerata un fuori lista, in quanto il campione dell'indagine è a livello nazionale.

Eleggibilità incerta (EF3)

Capi	Cati
3.1 contatore esaurito senza contatto <ul style="list-style-type: none">• CE non c'è nessuno in casa e non si hanno informazioni• CE non c'è nessuno in casa e provati tentativi telefonici senza successo	3.1 contatore esaurito senza contatto <ul style="list-style-type: none">• CE tutti non risponde• CE tutti occupato<ul style="list-style-type: none">• CE segreteria telefonica senza messaggio• CE tentativi misti senza contatto
3.2 fine periodo di rilevazione senza contatto <ul style="list-style-type: none">• FP nuovo senza tentativi• FP iniziati tentativi senza contatto	3.2 fine periodo di rilevazione senza contatto <ul style="list-style-type: none">• FP nuovo senza tentativi• FP iniziati tentativi senza contatto
3.3 irreperibili per nome/indirizzo errato	3.3 irreperibili per telefono errato¹⁰
3.4 non convalidate	3.4 numero errato¹¹
3.5 non distribuite	

In entrambe le tecniche l'esito "contatore esaurito" riguarda i casi in cui sono stati effettuati più tentativi di quelli previsti, senza mai riuscire a contattare la famiglia. Anche in questo caso nel Cati per le *waves* successive alla prima tale esito non è previsto.

L'esito "fine periodo" è identico nelle due differenti tecniche di rilevazione e può riguardare nominativi vergini (senza nessun tentativo) o nominativi con qualche tentativo ma senza contatto.

Sono con eleggibilità incerta anche le famiglie irreperibili per nome o indirizzo errato.

Nel Capi sono considerate con eleggibilità incerta le famiglie base delle quartine mai arrivate agli intervistatori per essere lavorate (quartine non convalidate e non distribuite).

Non eleggibili

I casi di non eleggibilità derivano dal fatto che il campione include nominativi che non corrispondono alle famiglie selezionate. Essi comprendono:

¹⁰ Questo esito è presente solo nelle *wave di famiglia* successiva alla prima per le famiglie provenienti dal Capi

¹¹ Nel caso dell'indagine sulle Forze di lavoro si ricorda che i nominativi vengono estratti dalle anagrafi dei Comuni campione, e che il recapito telefonico è inserito dal rilevatore in occasione della prima intervista. Nelle interviste successive alla prima, le famiglie con esito "telefono errato" sono da considerare con eleggibilità sconosciuta, in quanto, pur essendo la famiglia già stata intervistata nella *wave* precedente, non è comunque possibile essere certi che la famiglia selezionata viva ancora presso il recapito contattato.

Non Eleggibili (EF4)

Capi	Cati
4.1 La famiglia vive all'estero	4.1 La famiglia vive all'estero
4.2 Famiglia è deceduta	4.2 Famiglia è deceduta
4.3 La famiglia vive permanentemente in istituzione/convivenza	4.3 La famiglia vive permanentemente in istituzione/convivenza
4.4 Insieme di individui	4.4 Insieme di individui
4.5 Negozio, Impresa, Azienda, Ufficio	4.5 Negozio, Impresa, Azienda, Ufficio
4.6 Seconda casa	4.6 Seconda casa

5.3.4 - I tassi calcolati per l'indagine

Una volta standardizzati gli esiti, si è provveduto a progettare il piano di monitoraggio dell'indagine e più in particolare i tassi da calcolare, distinguendo tra monitoraggio di *surve* ed monitoraggio di *field*.

Per il piano di *survey*, sono stati calcolati i seguenti tassi:

- i tassi di risposta (ed i complementari tassi di non risposta);
- i tassi di cooperazione;
- i tassi di contatto;
- i tassi di non risposta degli eleggibili;
- i tassi di rifiuto;
- i tassi di fuori target;
- i tassi di sostituzione.

In ogni classe sono inclusi più tassi con differente significato per la valutazione complessiva dell'indagine e differente contributo al monitoraggio della fase di rilevazione sul campo.

Nel paragrafo successivo sono spiegati in dettaglio i tassi che corrispondono a quelli Aapor.

5.3.5 - I tassi utilizzati nell'indagine per il monitoraggio di survey

I tassi di risposta

Il tasso di risposta ha l'obiettivo di misurare la validità del campione, ovvero quanta parte del campione depurato dai casi non eleggibili è stato intervistato.

In letteratura il tasso di risposta (*Response Rate* - RR) viene definito come il numero di interviste effettuate diviso il numero di famiglie eleggibili del campione (Frankel, 1983; Groves, 1989; Hidiroglou, et al., 1993; Kviz, 1977; Lessler and Kalsbeek, 1992; Massey, 1995). La definizione è abbastanza ampia, perché i termini di intervista e famiglie eleggibili possono essere interpretati in maniera differente e, in base all'accezione considerata, generare differenti modalità di calcolo.

Aapor specifica che il tasso può essere calcolato in sei differenti modi:

- $RR1 = [EF1.1/(EF1+EF2+EF3)]$
- $RR2 = [EF1.1+EF1.2/(EF1+EF2+EF3)]$
- $RR3 = [EF1.1/(EF1+EF2+\alpha(EF3))]$
- $RR4 = [EF1.1+EF1.2/(EF1+EF2+\alpha(EF3))]$
- $RR5 = [EF1.1/(EF1+EF2)]$
- $RR6 = [EF1.1+EF1.2/(EF1+EF2)]$

I sei tassi differiscono in due aspetti:

- il primo aspetto riguarda il numeratore, ovvero la possibilità di considerare intervistate solo le unità con intervista completa oppure oltre queste anche quelle con intervista parziale (le cui risposte mancanti sono comunque fonte di errore);
- il secondo aspetto riguarda il denominatore, ovvero la possibilità di considerare come unità eleggibili tutte le unità del campione escludendo solo quelle non eleggibili, oppure escludendo anche una percentuale di unità (stimata in base all'esperienza) con eleggibilità sconosciuta, oppure infine escludendo oltre alle unità non eleggibili anche tutte quelle con eleggibilità sconosciuta, lasciando quindi solo quelle con eleggibilità certa¹².

Per l'indagine sono calcolati soltanto i tassi di risposta Aapor Rr1, Rr2, Rr5, Rr6 essendo impossibile stimare la parte di casi eleggibili inclusi in quelli con eleggibilità sconosciuta. Solo l'utilizzo e l'analisi congiunta di questi quattro tassi ha permesso contemporaneamente di valutare la validità del campione utilizzando confrontandola con altre indagini anche internazionali (con il tasso Rr1), l'incidenza sulla validità dello stesso delle unità con eleggibilità sconosciuta (con il tasso Rr5) e l'effetto delle risposte parziali (con i tassi Rr3 e Rr6).

I tassi di cooperazione

In letteratura con il termine di tasso di cooperazione (*Co-operation Rate - Coop*) viene indicato il rapporto tra i casi intervistati e tutte le unità eleggibili contattate. Ci sono due livelli di cooperazione: il livello familiare e quello individuale. In questo caso è considerato solo il livello familiare. Il tasso di cooperazione potrebbe essere calcolato anche considerando solo i rifiuti rispetto ai contattati sicuramente eleggibili.

Tale tasso indica quindi la disponibilità delle unità del campione a concedere l'intervista, ovvero quanta parte delle unità contattate, e quindi in grado di rispondere, ha risposto.

¹² In particolare, *Response Rate 1* (RR1) è il numero di interviste complete diviso il numero di interviste (complete più parziali) più il numero di famiglie contattate ma non intervistate (rifiuti più interruzioni definitive più contatore esaurito per appuntamenti più altro), più le famiglie non contattate ma con eleggibilità certa, più tutti i casi di eleggibilità incerta. Questo tasso tra tutti è il più rigoroso nella determinazione della validità del campione, avendo al numeratore soltanto le interviste complete (escludendo quelle parziali perché affette da errore) e al denominatore tutte le unità del campione escludendo solo i non eleggibili. *Response Rate 2* (RR2) ha al numeratore sia le interviste complete che quelle parziali. Il denominatore è lo stesso del tasso RR1. Nella valutazione della validità del campione è meno rigoroso in quanto include anche le interviste con risposte mancanti. *Response Rate 3* (RR3), è il numero di interviste complete diviso il numero di interviste (complete più parziali) più il numero di famiglie contattate ma non intervistate (rifiuti più interruzioni definitive più contatore esaurito per appuntamenti più altro), più le famiglie non contattate ma con eleggibilità certa, più una percentuale (α), definita in base a precedenti esperienze, di casi di eleggibilità sconosciuta che possono considerarsi eleggibili. *Response Rate 4* (RR4), ha al numeratore sia le interviste complete che quelle parziali. *Response Rate 5* (RR5), è il numero di interviste complete diviso il numero di interviste (complete più parziali) più il numero di famiglie contattate ma non intervistate (rifiuti più interruzioni definitive più contatore esaurito per appuntamenti più altro), più le famiglie non contattate ma con eleggibilità certa. Tale tasso è un caso particolare di RR3 in quanto considera tutti i casi di eleggibilità sconosciuta come non eleggibili. *Response Rate 6* (RR6) ha al numeratore sia le interviste complete che quelle parziali e corrisponde al valore massimo che può assumere il tasso di risposta calcolato nei differenti modi. E' bene calcolare i tassi RR5 ed RR6 solo nel caso in cui si attribuisca alle unità con eleggibilità sconosciuta una probabilità prossima ad uno di essere non eleggibili.

Aapor specifica che il tasso può essere calcolato in quattro differenti modi:

- $COOP1 = [EF1.1/(EF1+EF2.1)]$
- $COOP2 = [EF1.1+EF1.2/(EF1+EF2.1)]$
- $COOP3 = [EF1.1/(EF1+EF2.1.1+EF2.1.2)]$
- $COOP4 = [EF1.1+EF1.2/(EF1+EF2.1.1+EF2.1.2)]$

I quattro tassi differiscono in due aspetti:

- il primo aspetto riguarda il numeratore, ovvero la possibilità di considerare intervistate solo le unità con intervista completa oppure oltre queste anche quelle con intervista parziale (dove la disponibilità è parziale);
- il secondo aspetto riguarda il denominatore, ovvero la possibilità di considerare come unità in grado di rispondere tutte le unità eleggibili contattate, oppure solo quelle con facoltà di decidere, escludendo quindi le unità la cui scelta di rifiutare non è stata volontaria, ma è stata dettata da una causa di forza maggiore (malati gravi, esclusi per eccesso di tentativi, per fine periodo)¹³.

I tassi di rifiuto

In letteratura con il termine tasso di rifiuto (*Refusal Rate* - REF) solitamente si indica il rapporto tra i casi che hanno rifiutato o interrotto definitivamente l'intervista e tutte le unità eleggibili.

Tale tasso esprime l'incidenza dei rifiuti.

AAPOR specifica che il tasso può essere calcolato in tre differenti modi:

- $REF1 = [(EF2.1.1+EF2.1.2)/(EF1+EF2+EF3)]$
- $REF2 = [(EF2.1.1+EF2.1.2)/(EF1+ EF2+ \alpha(EF3))]$
- $REF3 = [(EF2.1.1+EF2.1.2)/(EF1+ EF2)]$

I tre tassi differiscono per il denominatore. Nel primo (Ref1) sono considerate come unità eleggibili tutte le unità del campione escludendo solo quelle non eleggibili, nel secondo (Ref2) è esclusa anche una percentuale di unità (stimata in base all'esperienza) con eleggibilità sconosciuta, nel terzo infine sono escluse tutte le unità non eleggibili e quelle con eleggibilità sconosciuta, lasciando quindi solo quelle con eleggibilità certa¹⁴.

¹³ *Co-operation Rate* 1 (COOP1) è il numero di interviste complete diviso il numero di interviste (complete più parziali) più il numero di famiglie contattate ma non intervistate con almeno un componente eleggibile (rifiuti più interruzioni definitive più contatore esaurito per appuntamenti) più i malati gravi. Questo tasso di cooperazione è il più rigoroso e quello con il valore più basso. *Co-operation Rate* 2 (COOP2) ha al numeratore sia le interviste complete che quelle parziali. *Co-operation Rate* 3 (COOP3) definisce i rispondenti che hanno la facoltà di decidere spontaneamente (per questo il contatore esaurito ed i malati gravi sono esclusi dal denominatore). *Co-operation Rate* 4 (COOP4) ha al numeratore sia le interviste complete che quelle parziali. Tale tasso è il più grande tra i quattro, ma in termini di utilità è il più significativo, esprimendo il reale interesse delle unità intervistate a collaborare ai fini della realizzazione dell'indagine.

¹⁴ *Refusal Rate* 1 (Ref1) è il numero di rifiuti diviso il numero di interviste (complete più parziali), più il numero di famiglie contattate ma non intervistate (rifiuti più interruzioni definitive più contatore esaurito per appuntamenti più altro), più le famiglie non contattate ma con eleggibilità certa, più tutti i casi di eleggibilità incerta. *Refusal Rate* 2 (Ref2) è il numero di rifiuti diviso il numero di interviste (complete più parziali) più il numero di famiglie contattate ma non intervistate (rifiuti più interruzioni definitive più contatore esaurito per appuntamenti più altro), più le famiglie non contattate ma con eleggibilità certa, più una percentuale □ di casi con eleggibilità incerta come nel caso di RR3 e RR4. *Refusal Rate* 3 (Ref3), analogamente al tasso RR6, è il numero di rifiuti diviso il numero di interviste (complete più parziali) più il numero di famiglie contattate ma non intervistate (rifiuti più interruzioni definitive più contatore esaurito per appuntamenti più altro), più le famiglie non contattate ma con eleggibilità certa. Tale tasso assume il valore più alto tra i tre.

Per l'indagine sono calcolati soltanto i tassi di rifiuto Aapor RF1, RF3, essendo impossibile stimare la parte di casi eleggibili inclusi in quelli con eleggibilità sconosciuta. Inoltre, in aggiunta ai tassi di rifiuti Aapor che hanno consentito una confrontabilità con le altre indagini, sono stati calcolati anche altri tassi di rifiuto.

Per valutare il *field* è stato calcolato il *tasso netto di rifiuto* per individuare i rifiuti rispetto al totale delle persone con la possibilità di rilasciare l'intervista (interviste complete, interruzioni definitive e rifiuti):

- Tasso netto di rifiuto = $[EF2.1.1+EF2.1.2/(EF1+ EF2.1.1+EF2.1.2)]$

Gli altri indicatori calcolati per il monitoraggio di *survey*

Per monitorare al meglio l'indagine anche al di là dei tassi calcolati così come consigliato dall'Aapor sono stati calcolati una serie di indicatori relativi al campione considerato come insieme di quartine.

Il piano delle tavole per il monitoraggio dell'indagine riporta, oltre ai tassi di cui si è parlato nei paragrafi precedenti, tutte le dimensioni assolute presenti nello schema riportato in figura 5.2, distintamente e congiuntamente per le due tecniche di rilevazione (Prospetto 5.1).

Prospetto 5.1 - Indicatori di Survey

	INDICATORI	Calcolo (vedi Figura 5.1)
1.1	% di Quartine assegnate su campione teorico	1.1+1.2/1
1.2	% di Quartine bucate su campione teorico	1.3/1
2.1	% di Quartine assegnate Capi su quartine assegnate	1.1/(1.1+1.2)
2.2	% di Quartine assegnate Capi su campione teorico	1.1/1
3.1	% di Quartine convalidate Capi su quartine assegnate	1.1.1/1.1
3.2	% di Quartine convalidate Capi su campione teorico	1.1.1/1
4.1	% di Quartine distribuite Capi su quartine assegnate	1.1.1.1/1.1.1
4.2	% di Quartine distribuite Capi su quartine convalidate	1.1.1.1/1.1
4.3	% di Quartine distribuite Capi su campione teorico	1.1.1.1/1
5.1	% di Quartine toccate Capi su quartine distribuite	1.1.1.1.1/1.1.1.1
5.2	% di Quartine toccate Capi su campione teorico	1.1.1.1.1/1
6.1	% di Quartine assegnate Cati su quartine assegnate	1.2/(1.1+1.2)
6.2	% di Quartine assegnate Cati su campione teorico	1.2/1
7.1	% di Quartine toccate Cati su quartine assegnate	1.2.1/1.2
7.2	% di Quartine toccate Cati su campione teorico	1.2.1/1
8.1	% di Complete Capi su quartine distribuite	1.1.1.1.1.1/1.1.1.1.1
8.2	% di Complete Capi base su quartine distribuite	1.1.1.1.1.1.1/1.1.1.1.1
8.3	% di Complete Capi sostitute su quartine distribuite	1.1.1.1.1.1.2/1.1.1.1.1
9.1	% di Complete Cati su quartine assegnate	1.2.1.1/1.2
9.2	% di Complete Cati base su quartine assegnate	1.2.1.1.1/1.2
9.3	% di Complete Cati sostitute su quartine assegnate	1.2.1.1.2/1.2
10.1	% di Complete Capi e Cati su campione teorico (tasso completezza)	(1.1.1.1.1.1+1.2.1.1)/1
10.2	% di Complete Capi e Cati base su campione teorico (tasso di fedeltà)	(1.1.1.1.1.1.1+1.2.1.1.1)/1
10.3	% di Complete Capi e Cati sostitute su campione teorico (tasso di infedeltà)	(1.1.1.1.1.1.2+1.2.1.1.2)/1

La percentuale di quartine assegnate Capi/Cati sul campione teorico (1.1) e il suo complemento (1.2) forniscono una misura del numero di quartine per le quali è stato possibile assegnare/non assegnare la tecnica di rilevazione, in particolare il secondo indica l'incidenza delle "quartine bucate", le quali, come più volte detto, producono una riduzione della dimensione campionaria e quindi un effetto sull'errore campionario.

La percentuale di quartine assegnate con tecnica di rilevazione Capi sul totale delle quartine con tecnica assegnata (2.2) indica in che misura incide la tecnica di indagine Capi in tutta la fase di rilevazione. Analogamente dicasi per lo stesso indicatore calcolato in riferimento alla tecnica Cati (7.2).

Tutti gli altri indicatori vengono calcolati a cascata separatamente per i due percorsi Capi e Cati.

Di una certa rilevanza risulta essere l'indicatore ottenuto rapportando il numero di quartine convalidate Capi sul totale delle quartine assegnate (3.1) ed il suo complemento che fornisce una misura della mancata risposta totale dovuta a mancata convalida da parte degli Uffici territoriali Istat, mentre la percentuale di quartine distribuite Capi sul totale delle quartine assegnate (4.1) ed il complemento forniscono una misura della mancata risposta totale dovuta a problemi tecnici di trasmissione via modem. La percentuale di quartine toccate Capi sul totale delle quartine distribuite (5.1) Capi e il suo complemento danno un'indicazione della "pigrizia" del rilevatore

Capi nell'effettuare una pur minima azione sulla quartina, cioè offre una misura della difficoltà dell'intervistatore a far fronte all'intero carico di lavoro settimanalmente trasmesso.

La percentuale di quartine con intervista completa Capi sul totale delle quartine distribuite Capi (8.1) e la percentuale di quartine con intervista completa Cati sul totale delle quartine assegnate Cati (9.1) indicano quante quartine effettivamente distribuite hanno dato luogo ad un'intervista completa. Questi stessi indicatori vengono calcolati separatamente a seconda che la famiglia della quartina con intervista completa sia una famiglia base o una famiglia suppletiva.

Congiuntamente per la tecnica di rilevazione Capi e per la tecnica di rilevazione Cati viene calcolato il tasso di completezza ottenuto come rapporto tra il numero di quartine con intervista completa e il numero di quartine del campione teorico (10.1), il quale indica quanta parte del campione teorico ha dato luogo ad un'intervista completa e quindi dà una misura di quanto l'obiettivo di intervistare tutto il campione teorico sia stato raggiunto o meno. Tanto più tale indicatore è prossimo al 100 per cento, tanto più può ritenersi raggiunto l'obiettivo dell'indagine. Il rapporto tra il numero di quartine in cui la famiglia che ha fornito l'intervista completa è la famiglia base e il numero totale delle quartine del campione teorico (tasso di fedeltà) (10.2) fornisce una misura dell'aderenza del campione effettivo al campione teorico. Invece, il rapporto tra il numero di quartine in cui la famiglia che ha fornito l'intervista completa è una famiglia suppletiva e il numero totale delle quartine del campione teorico (tasso di infedeltà) (10.3) offre un'indicazione di quanto si discosti il campione effettivo dal campione teorico, essendo stato necessario il ricorso alle famiglie suppletive delle quartine per ottenere delle interviste complete.

Un grande impegno è stato speso nella individuazione delle variabili fondamentali di incrocio da utilizzare. Le più semplici sono sicuramente le variabili territoriali e temporali. La prima è rappresentata dalle 20 Regioni in cui è suddiviso il territorio italiano, utile per ciascun Ufficio territoriale Istat per monitorare l'andamento dell'intera fase di rilevazione della rispettiva Regione di appartenenza; la seconda è rappresentata dalla settimana di riferimento dei dati, in tal modo per ogni trimestre è possibile scendere nel dettaglio degli indicatori sopra illustrati per ciascuna delle 13 settimane che lo compongono. Nell'analisi degli indicatori, però, ci si è resi conto che talvolta è necessario un dettaglio maggiore tanto da dover prendere in considerazione ulteriori dimensioni, quali la wave di quartina e la wave di famiglia.

Nel sistema di tavole l'incrocio con la wave di quartina è utile per analizzare l'andamento dei vari indicatori a seconda della fase del ciclo di vita delle varie quartine; è utile per individuare in modo univoco il gruppo di rotazione di appartenenza della quartina.

Nello stesso tempo, però, è molto importante l'incrocio con la wave di famiglia, in quanto i vari tassi possono assumere valori differenti a seconda se la famiglia è alla prima intervista o ad interviste successive. Infatti, generalmente si ottengono delle performance migliori se la famiglia è ad un'intervista successiva alla prima, per il fatto di aver già acquisito una certa familiarità con l'indagine e il rilevatore, rispetto ad una prima intervista per la quale il rilevatore deve effettuare un'opera maggiore di convincimento. In modo analogo, si ottengono performance peggiori per famiglie che si trovano alla quarta intervista, per le quali entra in gioco un fattore stanchezza.

5.3.6 - I tassi utilizzati nell'indagine per il monitoraggio di field

Oltre ad analizzare e tenere sotto controllo ciò che accade al campione, è importante analizzare più nel dettaglio ciò che accade sul *field*. Infatti il rilevatore può rappresentare una fonte non indifferente di errore non campionario se risulta impreparato sul questionario di rilevazione, sulle modalità e le regole di conduzione dell'intervista e scarsamente motivato nello svolgere la sua attività. Come già evidenziato, tutta la rete di rilevazione è sottoposta ad un processo di formazione strutturato in formazione d'aula, formazione continua e formazione a distanza. Ma la sola formazione non è sufficiente a garantire la qualità del processo di

rilevazione, ed è necessario controllare come la formazione ricevuta si concretizza in comportamenti adeguati durante la fase di rilevazione per contenere al massimo l'errore non campionario proveniente dall'operato dei rilevatori.

I tassi di risposta, di cooperazione e di rifiuto proposti da Aapor, nonché gli altri indicatori più specifici per il monitoraggio di *survey*, possono essere calcolati per ciascun rilevatore, ma a tal proposito è scaturita una profonda riflessione circa le informazioni e i relativi indicatori che meglio caratterizzano il *field*.

Oltre ai dati assoluti, le prime indicazioni, come già menzionato, provengono dal rapporto tra le famiglie che vengono intervistate e quelle che bisogna intervistare (tasso di completezza). A questo proposito, è stato necessario ripensare alle grandezze da usare come denominatore. Se l'interesse è quello di verificare l'operato sul campo, ossia se un intervistatore ha svolto correttamente il compito che gli è stato affidato, non è opportuno riportare il numero di interviste complete a quelle teoriche: si dovrà, invece, riportare le interviste complete a quelle che sono state effettivamente distribuite all'intervistatore. Questo perché, se si è interessati alle performance dell'intervistatore, non si può far scontare a questi il non perfetto funzionamento del sistema a monte del suo lavoro.

Analogamente a quanto fatto per il monitoraggio di *survey*, un altro indicatore che fornisce informazioni utili a comprendere le performance del *field* è quello che si è definito tasso di fedeltà al campione. Similmente a quanto detto per il tasso di completezza è utile riportare il numero di quartine in cui la famiglia che ha fornito l'intervista completa è la famiglia base al numero delle quartine del campione teorico effettivamente distribuite, indicatore che permette di analizzare lo scostamento dal campione teorico assegnato per ogni singolo rilevatore.

Per una più completa valutazione del lavoro del *field* è utile disporre della distribuzione delle mancate risposte totali (*eligible and not eligible total non-response*) per motivo della non risposta. Tale distribuzione aiuta il supervisore a comprendere su quale aspetto si può intervenire per migliorare le performance del rilevatore. Ad esempio una situazione in cui la percentuale delle cadute dovute al rifiuto della famiglia a partecipare all'indagine è particolarmente alta, indica che la capacità persuasiva dell'intervistatore per ottenere l'intervista è piuttosto bassa e che quindi sono necessari degli interventi tempestivi tesi a migliorare le capacità di conquista e mantenimento dell'intervista.

Durante la formazione dei rilevatori sono state date regole ben precise circa le possibilità di ricorso all'intervista *proxy*. Il tasso di *proxy*, ottenuto riportando il numero di famiglie con almeno un componente per il quale è stata effettuata un'intervista *proxy* al numero di famiglie intervistate, oppure l'indicatore ottenuto riportando il numero di interviste *proxy* al numero dei rispondenti, permettono una prima valutazione sull'accuratezza con cui vengono rilevati i dati.

È possibile valutare il lavoro del *field* non soltanto facendo riferimento agli esiti definitivi delle interviste, ma anche combinando tra loro esiti definitivi e provvisori, ad esempio conteggiando il numero di contatti che si sono avuti, gli appuntamenti che sono stati presi e le interviste che sono state portate a termine nel corso di un giorno o di una settimana di calendario: una settimana in cui un intervistatore ha preso molti appuntamenti, ha avuto molti contatti con le famiglie ma non ha portato a termine nessuna intervista potrebbe essere il segnale di qualcosa che non funziona correttamente e che potrebbe non dipendere direttamente dall'intervistatore ma da altre situazioni contingenti su cui magari si può provare ad intervenire per cercare di rimuoverle. Per questa ragione una delle cose che si monitora è, ad esempio, il numero di contatti con il Comune per verificare se l'eccessivo numero di contatti non andati a buon fine possa in qualche modo essere dovuto a liste anagrafiche poco aggiornate ad esempio per la presenza di indirizzi relativi a case per le vacanze. Un esempio di indicatore utile per valutare le performance del *field* utilizzando le informazioni relative ai contatti è il cosiddetto "tasso di

pigrizia” (quartine con nessuna azione / quartine assegnate) che permette di valutare eventuali eccessi di carico di lavoro rispetto alle capacità organizzative di ogni singolo intervistatore.

Un ulteriore strumento per monitorare il funzionamento del *field* è data dalla reportistica sulle trasmissioni. Durante la formazione è stato più volte ribadito ai rilevatori di connettersi via modem spesso, possibilmente una volta al giorno, al fine di ricevere i nominativi da intervistare e trasmettere le interviste effettuate. Dal numero di tentativi di trasmissione e di quelli andati a buon fine si può valutare se un’eventuale mancanza di interviste da parte di uno specifico intervistatore si possa attribuire a problemi di trasmissione o ad un effettivo non raggiungimento, in tempo utile, dell’intervista completa.

Molte delle informazioni che vengono analizzate per valutare la performance di *field* sono sfruttate anche per costruire indicatori sull’operato dei rilevatori in una specifica settimana di calendario. Gli indicatori legati alla settimana di riferimento sono infatti disponibili solo alla chiusura di tutte le interviste relative a quella specifica settimana dell’anno, ovvero dopo un periodo lungo circa 6 settimane, intervallo di tempo troppo lungo se si vogliono utilizzare le informazioni raccolte per intervenire e modificare eventuali processi critici. Per questo motivo i principali tassi ed indicatori sono calcolati per il *field* sia per settimana di riferimento, sia per settimana di calendario.

Gli indicatori per settimana di calendario vengono costruiti utilizzando una logica trasversale che considera simultaneamente tutti gli eventi relativi alla specifica settimana, e cioè il carico di lavoro complessivo (dato dalle nuove interviste da effettuare più tutte quelle non effettuate per settimane di riferimento precedenti), i contatti e gli appuntamenti relativi alla settimana di calendario (e quindi svincolati dalla settimana di riferimento a cui sono associate le famiglie del campione), il numero di interviste completate in quella specifica settimana (anche in questo caso considerando insieme le interviste da effettuare in quella settimana e quelle che avrebbero dovuto essere realizzate prima ma sono state gestite nei periodi di coda previsti). La possibilità di avere simili informazioni su una precisa settimana di calendario permette di monitorare in tempi strettissimi l’operato dei rilevatori in quanto non bisogna attendere la fine di un ciclo di interviste relative ad una settimana specifica inclusa nel disegno campionario. La combinazione dei dati relativi al carico residuo, ai contatti, agli appuntamenti presi e alle interviste completate ha permesso anche di elaborare indicatori di attività, efficacia ed efficienza per ogni singolo rilevatore nell’ambito di una singola settimana.

Il lavoro degli intervistatori è controllato anche analizzando tutti gli esiti dei tentativi di contatto per giorno della settimana e fascia di orario. Tale analisi, effettuata per ogni intervistatore, consente di valutare la distribuzione del lavoro nell’arco della settimana e il numero e la tipologia di contatti effettuati per ottenere un esito definitivo.

5.3.7 - Le analisi multidimensionali

Al fine di comprendere il fenomeno “rilevazione” in sé molto complesso, gli indicatori di *field* sono stati analizzati anche con tecniche di analisi multidimensionale.

In particolare, si è voluto rappresentare il lavoro svolto dalle 20 Regioni italiane come sintesi degli n rilevatori in esse presenti, osservando p esiti di intervista in t intervalli temporali. A tal fine è stato applicato il metodo Statis che, partendo dai principali indicatori di qualità, ha permesso di stimare i fattori latenti che misurano le diversità strutturali, nonché la loro dinamica nel tempo.

Statis (Structuration des tables A trois indices de la statistique) è un metodo di analisi multivariata dinamica che permette di analizzare fenomeni multidimensionali (multiway), esprimibili in forma di matrici a tre vie: variabili-spazio-tempo. La metodologia è stata elaborata da Escoufier (1985) ed ha trovato numerose applicazioni anche in Italia D’Ambra, (1986);

Fachin-Vichi (1993); Tassinari-Vichi (1994)). Questa tecnica di analisi esplorativa si basa sullo studio di una matrice di dati a tre vie ottenuta dalla successione temporale di T matrici di dati X_t (una per ciascun periodo od occasione) aventi uguali dimensione tX_j , dove i sono le unità statistiche o individui, j le variabili e t è il periodo di rilevazione.

Il metodo Statis si articola in tre fasi: *interstruttura*, *compromesso ed intrastruttura*. Obiettivo della fase dell'Interstruttura è l'individuazione di un idoneo spazio vettoriale dove poter rappresentare le T occasioni. La matrice dell'interstruttura **IS** esprime le relazioni esistenti fra le diverse matrici e ha come elemento generico il coefficiente di correlazione vettoriale di Escoufier:

$$Rv_{u_t} = tr(\Theta_t \Theta_t') / \sqrt{tr(\Theta_t)^2 tr(\Theta_t')^2}$$

con $\Theta_t = \frac{1}{n}(X_t)' X_t$ matrice di varianza e covarianza di X_t

Nella fase *Compromesso* si vogliono riassumere nel miglior modo possibile le informazioni contenute nelle T matrici delle varianze e covarianze rappresentando gli individui o le variabili. La matrice compromesso **CO** è ottenuta come combinazione lineare delle Θ_j :

$$CO = \sum_{t=1}^T u_t \Theta_t$$

dove u_t è il t -esimo elemento dell'autovettore associato al più grande autovalore ottenuto dalla diagonalizzazione di **IS**. Nell'ultima fase, l'*Intrastruttura*, l'obiettivo è trovare lo spazio vettoriale dove rappresentare le traiettorie compiute dai singoli individui e dalle singole variabili.

Con tale metodo è possibile evidenziare e confrontare, fornendone una rappresentazione compatta, il comportamento degli indicatori caratteristici (*i tassi di indagine* nel caso specifico) relativi alle unità (*i singoli intervistatori o le Regioni*) nei differenti tempi (*settimane di rilevazione, trimestri oppure anni*) in sottospazi fattoriali, di dimensione ridotta rispetto alle dimensioni originali del fenomeno. Il contributo che tale metodo fornisce permette di studiare le modificazioni avvenute nel corso del tempo tenuto conto della struttura dei risultati prodotti dalle Regioni, senza far necessariamente ricorso ad un modello.

L'analisi che si riporta nel successivo paragrafo ha come unità le singole Regioni e come tempi di riferimento le 13 settimane di intervista del secondo trimestre del 2003. Tale analisi può essere ripetuta ogni trimestre anche per ogni singola Regione, considerando come unità statistiche i rilevatori.

I risultati

Per ragioni di sintesi si presentano qui di seguito soltanto i risultati dell'analisi svolta sulle 20 Regioni italiane e non su tutti i rilevatori italiani. La matrice analizzata è dunque una matrice a tre vie con i seguenti *indicatori di field* (Figura 5.4):

- Tasso di risposta
- Tasso di rifiuto
- % di interviste interrotte definitivamente
- % di Contatore Esaurito con contatto
- % di Contatore Esaurito senza contatto
- % Assenza prolungata
- % Malato grave
- % Famiglia irreperibile

- % Fuori Target
- % Fine Periodo di rilevazione

Figura 5.4 - I dati dell'analisi multidimensionale della rete di rilevazione

	Tasso di risposta	Tasso di rifiuto	Fuori Target
$X_{t_1}(n_1, p)$	I_1					
	I_2					
					
	I_{20}					
$X_{t_2}(n_2, p)$	I_1					
	I_2					
					
	I_{20}					
$X_{t_n}(n_n, p)$	I_1					
	I_2					
					
	I_{20}					

La figura 5.5 mostra il primo piano fattoriale della diagonalizzazione dell'IS. Il primo asse fattoriale che spiega il 40% della variabilità totale evidenzia forti differenze strutturali tra le settimane centrali del trimestre (sett9 , sett7) e quelle finali (sett11, sett12, sett13). Il secondo asse che spiega il 12% del fenomeno, contrappone prevalentemente le strutture della prima metà del trimestre a quelle della seconda metà. Dall'analisi del piano emerge la similitudine nella struttura di tassi della settimana 1 (sett1) con le settimane finali del trimestre. Ciò è spiegabile se, come si vedrà successivamente, si analizzano le relazioni tra i differenti indicatori.

Passando ad analizzare nella figura 5.6 il piano generato dalla diagonalizzazione della matrice CO, si nota come il primo fattore ($\lambda_1= 39,5$ per cento) sia correlato positivamente con il tasso di risposta e negativamente con il tasso di rifiuto ed i "fine periodo". Tale fattore può essere interpretato come un indicatore della performance. Il secondo asse ($\lambda_2= 12,5$ per cento) contrappone i differenti motivi di caduta (in alto i fuori target e gli irreperibili, in basso i fine periodo e gli impossibilitati a rispondere). La disposizione degli esiti sul piano CO consente di spiegare il motivo della similitudine tra la prima settimana e le ultime del trimestre. La prima settimana del trimestre di solito è caratterizzata da una forte motivazione dei rilevatori che ha come effetto buoni rendimenti in termini di tasso di risposta ma motivi di caduta principalmente dovuti ai rifiuti o all'eccesso di tentativi. Nelle settimane successive (sett2, sett3, sett4), la motivazione all'intervista è ancora forte e inoltre si impara a gestire le obiezioni dei rispondenti contenendo i rifiuti. E' proprio nella settimana 4 che si ottengono infatti i migliori risultati. Dalla 5 settimana alla 10, inizia ad incidere il fattore stanchezza, e ciò comporta una diminuzione del tasso di risposta, ed un aumento del motivo di caduta per fine periodo. Nelle ultime settimane del

trimestre (sett11, sett12, sett13) i rilevatori ricevono una spinta motivazionale forte a chiudere il trimestre con buoni risultati, così il tasso di risposta aumenta di nuovo.

Le elaborazioni effettuate nei trimestri successivi hanno confermato la presenza di questo fattore ciclico nel trimestre, anche se molto attenuato grazie agli interventi formativi che hanno mirato a coinvolgere costantemente i rilevatori in tutto l'arco del trimestre.

Per analizzare il comportamento delle Regioni durante il trimestre è stato poi analizzato il piano CO formato dalle unità. Tale piano che spiega il 68,9% della variabilità totale dovuta ai differenti risultati prodotti dalle Regioni, mostra in modo chiaro sul primo asse ($\lambda_1 = 59,1$ per cento) il divario tra le piccole Regioni italiane (il Molise, le Marche, la Basilicata) con elevata performance e le grandi Regioni (il Piemonte, il Lazio) con bassa performance (Figura 5.7). Il secondo asse che contribuisce per il 9,21 per cento alla varianza totale, separa le Regioni in cui è più facile contattare le famiglie (oltre al Molise e alla Basilicata, il Veneto e la Toscana) da quelle dove il contatto e la persuasione all'intervista è più difficile (la Sardegna, la Liguria, l'Umbria, Le Marche).

Figura 5.5 - Il piano dell'Interstruttura (primo e secondo asse)

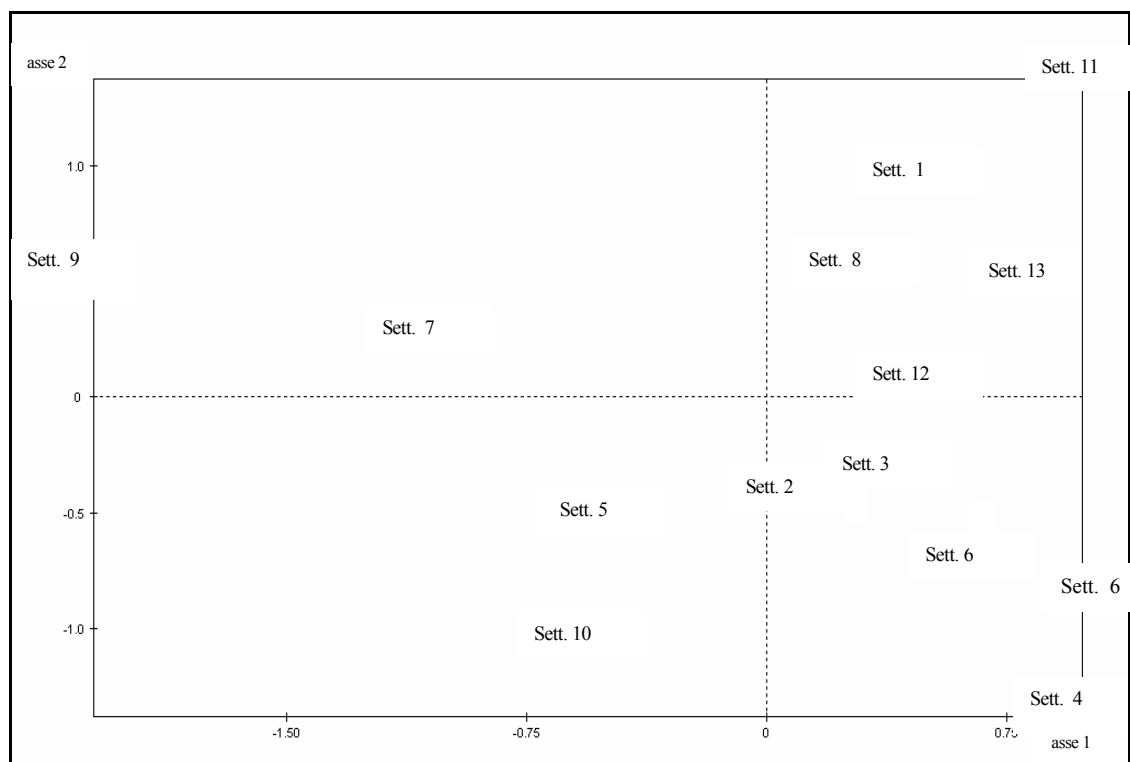


Figura 5.6 - Il piano del Compromesso delle variabili (primo e secondo asse)

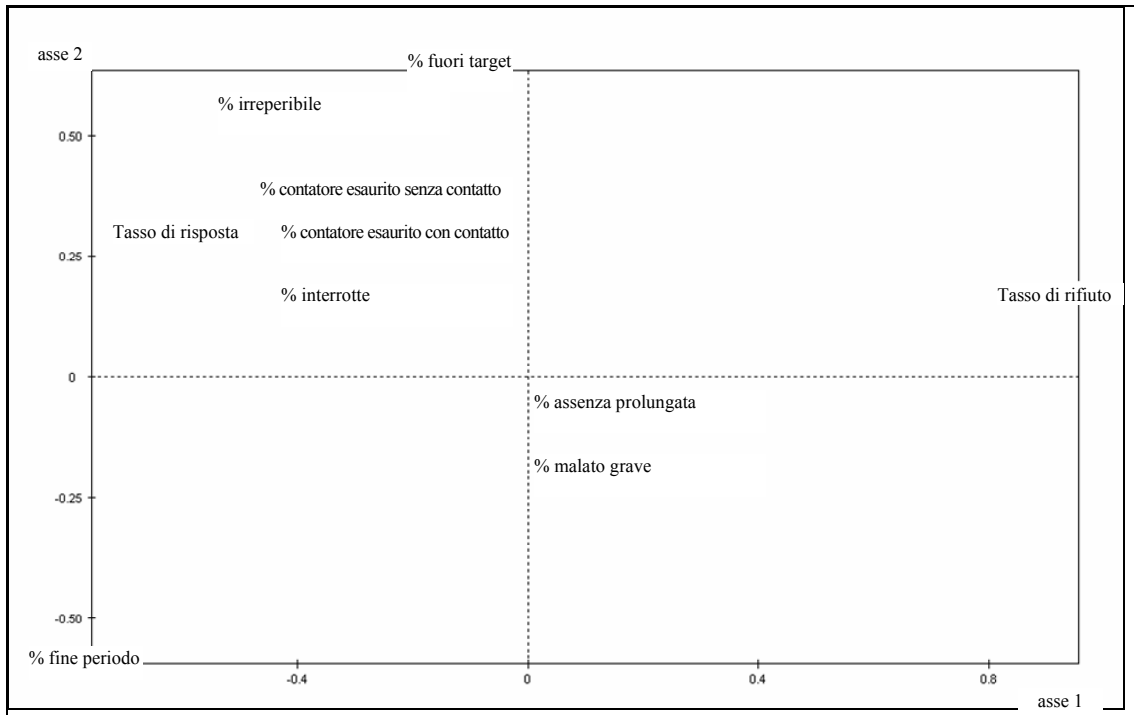
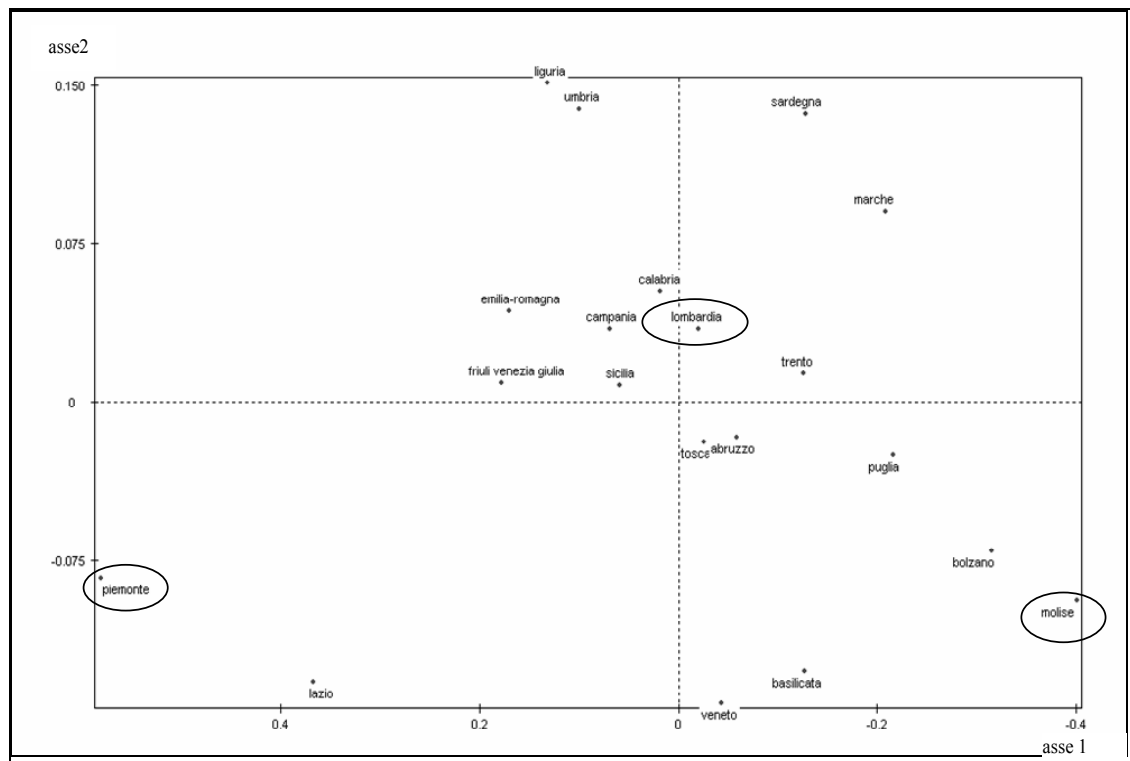
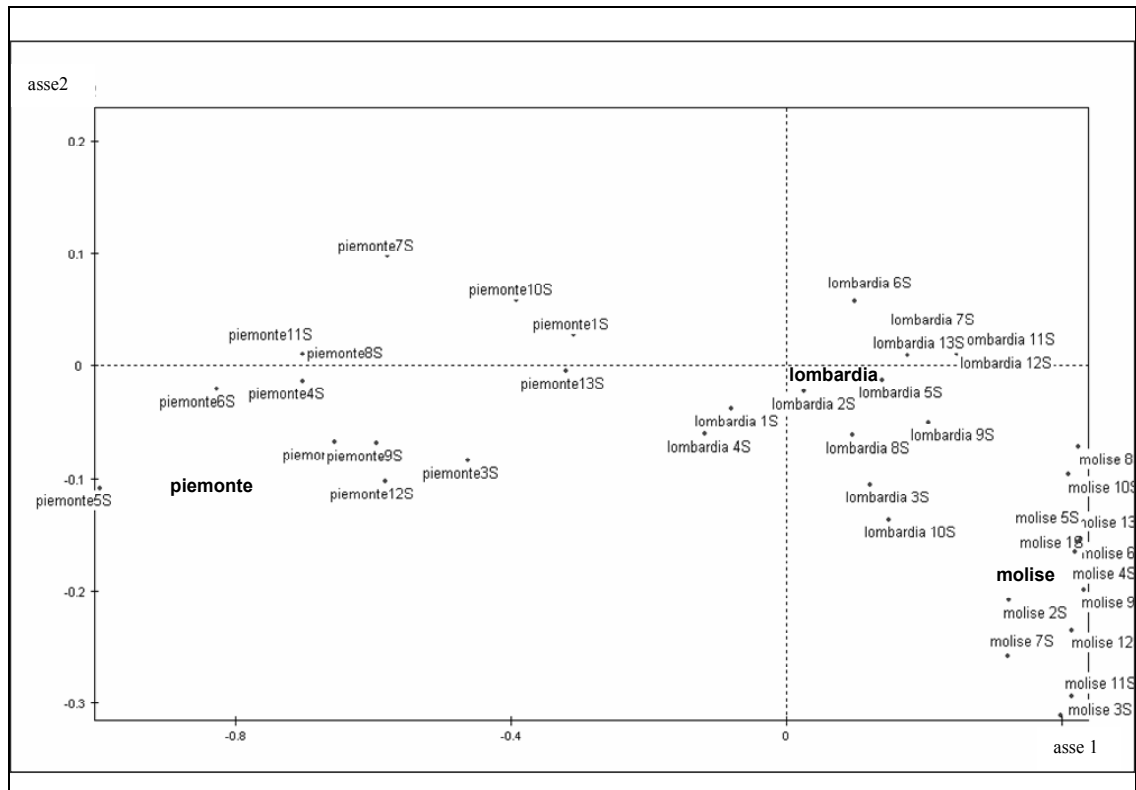


Figura 5.7 - Il piano del Compromesso degli individui (primo e secondo asse)



Per comprendere, infine, l'evoluzione temporale delle Regioni, evidenziata in figura 5.7, sono stati proiettati sul piano compromesso i valori associati alle Regioni nelle 13 settimane.

Figura 5.8 - Il piano del Compromesso degli individui (primo e secondo asse)



Per semplicità sono state considerate solo tre Regioni per i risultati ottenuti nel secondo trimestre del 2003: il Molise, che può essere considerata la migliore in termini di risultati ottenuti, il Piemonte con i risultati peggiori e la Lombardia con risultati medi.

Le traiettorie individuate confermano, anche se con proporzioni differenti e con qualche specificità individuale, il trend naturale del trimestre.

Come anticipato, questa analisi può essere ripetuta alla fine di ogni trimestre e consente di analizzare in modo veloce ed immediato il rendimento della rete di rilevazione ed il suo andamento nel tempo. Inoltre consente di analizzare gli effetti sulla rete di rilevazione delle iniziative poste in essere durante il trimestre (*de-briefing, incentivi...*), distinguendoli dagli effetti “naturalisti” che comportano, come visto, delle differenze tra le settimane del trimestre (maggiore stanchezza al centro del trimestre, maggiore motivazione all’inizio e alla fine del trimestre).

Capitolo 6 - Il questionario elettronico

6.1 - Premessa

L'impiego di un questionario elettronico e, soprattutto, di una rete di rilevatori professionali dell'Istat permette un maggior controllo sul processo di produzione del dato migliorando decisamente la qualità dell'informazione raccolta.

Considerato che una parte rilevante dell'errore non campionario si verifica durante la rilevazione e la registrazione delle informazioni, il questionario elettronico gioca un ruolo di primo piano nel prevenirlo e ridurlo. In questa ottica, l'indagine definitiva è stata preceduta da indagini pilota svolte nel corso del 2001 e del 2002 che hanno permesso di testare diverse versioni del questionario. La versione definitiva è nata dopo un attento studio della letteratura dedicata all'argomento, dei regolamenti Eurostat e delle indagini svolte in altri paesi, nonché da un intenso monitoraggio sul campo.

I debriefing periodici organizzati con i rilevatori Capi da parte dell'Istat centrale, il contatto giornaliero dei rilevatori Capi con i referenti degli uffici regionali Istat, e il monitoraggio in sala delle interviste Cati, con postazioni dedicate che consentono l'ascolto dell'intervista e la visualizzazione di ciò che viene registrato, hanno generato un continuo feedback tra i ricercatori e la rete di rilevazione. Gli stimoli e le domande provenienti dalla rete di rilevazione hanno consentito di specificare meglio alcuni concetti inizialmente considerati chiari e non problematici¹. Il feedback con la rete di rilevazione ha permesso di migliorare il wording delle domande per renderle più chiare agli intervistati, di affinare i percorsi spostando l'ordine di alcuni quesiti, di predisporre quando necessario alcuni elementi di ausilio alla conduzione dell'intervista (quali help in linea per l'intervistato e motori di ricerca), nonché di inserire ulteriori controlli in linea di coerenza fra successive risposte.

Nella fase successiva all'avvio dell'indagine questo stretto rapporto tra la rete di rilevazione e i ricercatori ha permesso di tenere sempre sotto controllo il processo di rilevazione. Ad esempio, in alcuni casi si è proceduto ad una integrazione della formazione risolvendo nuovi problemi verificatisi nella somministrazione dell'intervista per la difficoltà di comprensione di alcuni quesiti.

La strategia di fondo messa in atto per la conduzione della Rcfl si esplica nel passaggio, nella misura più ampia possibile, dalla correzione degli errori *offline*, successivi alla rilevazione dei dati, alla prevenzione e correzione degli errori *online*, ossia durante la rilevazione. A tal fine, la costruzione delle regole inserite nel questionario di tipo *hard* (quelle che prevedono una

Di Luciana Quattrociochi, con la partecipazione di Gabriella Grassia, Federica Pintaldi

¹ Un buon esempio è la domanda nella quale si chiede l'anno di inizio dell'attuale lavoro. Questa informazione apparentemente chiara si è rilevata non priva di ambiguità. I lavoratori dipendenti spesso chiedevano se la domanda si riferiva all'anno di inizio dell'attività lavorativa con l'attuale datore di lavoro o all'anno di inizio dell'attuale mansione e/o qualifica (nel caso di cambiamenti nel tempo). I dipendenti con contratto a termine, i quali lavorano da molti anni per lo stesso datore di lavoro ma hanno stipulato più contratti di lavoro, chiedevano se l'anno di inizio lavoro doveva riferirsi all'anno di inizio dell'ultimo contratto a termine o all'anno di inizio del rapporto di lavoro con quel datore di lavoro. I lavoratori autonomi chiedevano di specificare meglio se per anno di inizio lavoro dovesse intendersi l'inizio dell'attività autonoma o l'inizio dell'ultima commessa. Per tale ragione, si è deciso di differenziare la formulazione della domanda per i lavoratori dipendenti e per i lavoratori autonomi. Ai primi si chiede l'anno di inizio lavoro con l'attuale datore di lavoro, senza tenere conto di eventuali cambiamenti di qualifica e/o di mansione e considerando la continuità di fatto del lavoro. Ai lavoratori autonomi, invece, si chiede l'anno di inizio di quella particolare attività lavorativa a prescindere dalla singola commessa. La formulazione specifica su questo quesito e la presenza di help in linea agevolano il rilevatore nei casi più difficili.

soluzione obbligatoria da trovare insieme al rispondente nel contesto dell'intervista) e di tipo *soft* (quelle che fanno messaggi di *warning* sullo schermo del Pc del rilevatore) ha seguito la stessa strategia di approccio alla qualità del dato. Le regole sono state definite per fasi successive. Nella fase di sperimentazione, durata oltre un anno, la maggior parte delle regole erano definite come *soft* e i *range* delle variabili erano molto ampi o non presenti. L'analisi dei dati ha consentito di evidenziare i casi di apparente incongruenza tra le informazioni. I *debriefing* con i rilevatori, l'approfondimento sull'evoluzione del mercato del lavoro e, in alcuni casi, l'inserimento di domande di controllo hanno permesso di distinguere i casi di effettiva incongruenza tra le informazioni dai casi dove l'apparente incongruenza in realtà identificava un profilo particolare.

Nei prossimi paragrafi per prima cosa si descriveranno la scheda contatti e il nuovo questionario sulle Forze di Lavoro, evidenziando le potenzialità offerte dalla tecnica *computer assisted* nella semplificazione dei percorsi, nell'uso di informazioni rilevate precedentemente e nell'utilizzo di strumenti di ausilio all'intervista. In secondo luogo, si descriverà la strategia di costruzione delle regole di incompatibilità all'interno del questionario per permettere la correzione di eventuali errori *online*. Infine, nell'ultima parte del lavoro si procederà ad un confronto tra il piano di controllo *online* della nuova indagine continua e il piano di controllo *offline* della vecchia indagine al fine di valutare i miglioramenti raggiunti nella qualità dei dati prodotti.

6.2 - Il questionario elettronico delle Forze di lavoro

Nelle indagini statistiche il questionario è lo schema di un'intervista strutturata, la cui funzione è quella di raccogliere le informazioni rilevanti sulle variabili quantitative e qualitative oggetto dell'indagine (Buratta, 1989). E' importante ricordare che il questionario è una fonte di errori non campionari ed è pertanto necessario che sia progettato secondo criteri che ne permettano la prevenzione e la misurazione. Con l'introduzione dei sistemi di rilevazione assistita, il questionario cambia in numerosi aspetti (Prospetto 6.1).

Prospetto 6.1 - I principali cambiamenti nel questionario per i metodi Computer Assisted Survey Information Collection (CASIC)

Paper&Pencil	CASIC
<ul style="list-style-type: none"> • Modello di rilevazione • Lettura delle domande dal modello di rilevazione • Scrittura delle risposte sul modello • Percorso del questionario controllato dal rispondente/rilevatore • Codifica a freddo • Controlli di range a freddo • Controlli di coerenza delle risposte a freddo • Istruzioni per il rilevatore su documenti cartacei • Domande impersonali • Impossibilità di accedere a dati derivanti da fonti esterne 	<ul style="list-style-type: none"> • Questionario elettronico • Lettura delle domande dal video • Scrittura sulla tastiera o touch screen • Percorso delle domande gestito in automatico dal computer • Codifica a caldo • Controllo di range a caldo • Controlli di coerenza delle risposte a caldo • Help in linea delle domande • Formulazione personalizzata delle domande • Possibilità di accedere a dati derivanti da fonti esterne

Nei sistemi *computer intensive*, le fasi di *caricamento*, *codifica* e *controllo e correzione* dei dati sono contemporanee alla rilevazione dei dati stessi, per cui una buona progettazione del questionario diviene fondamentale per la qualità dei dati (Keller, 1995). La piena automazione del processo di rilevazione, controllo e immissione dei dati implica che in un solo programma devono essere inglobate le varie funzioni in modo integrato secondo un piano coordinato.

La progettazione del questionario elettronico della RcfI ha avuto come obiettivo quello di sfruttare al massimo le potenzialità offerte dai sistemi *computer intensive*, in primo luogo partendo dal dettagliare tutte le funzioni del questionario elettronico e, in secondo luogo, realizzando il questionario in modo iterativo testando e riprogettando eventuali punti critici.

In sintesi, gli elementi caratteristici del questionario della Rcfl realizzati grazie alle potenzialità offerte dal questionario elettronico sono i seguenti:

- interfaccia verso altri software;
- cura dell'aspetto grafico (colori, carattere, dimensione, ecc.);
- gestione del contatto iniziale con le famiglie in modalità computer assisted;
- ricostruzione di tutti gli esiti di contatto e monitoraggio dell'indagine;
- gestione dei salti condizionati tra un quesito e successivi sollevando il rilevatore dalla possibilità di generare errori;
- personalizzazione dei percorsi per prima e per successiva intervista;
- ricostruzione longitudinale della famiglia e rilevazione di eventuali cambiamenti;
- personalizzazione nella formulazione delle domande per profilo del rispondente;
- presentazione delle modalità di risposta per profilo del rispondente;
- particolare gestione della modalità di risposta "non sa" per le interviste *proxy*;
- gestione delle codifiche mediante motori di ricerca *online* oppure a posteriori dopo la chiusura dell'intervista;
- *help online* con la visualizzazione di schede di istruzione per specifici quesiti;
- ricostruzione in automatico di alcune variabili;
- controlli di *range*;
- controlli di consistenza tra le risposte fornite ad un quesito e quelle fornite in quesiti precedenti;
- controlli di coerenza tra le risposte fornite ad un quesito e quelle fornite allo stesso quesito in precedenti *waves*;
- inserimento di domande di controllo al fine di prevenire errori nella fase di acquisizione dell'informazione;
- memorizzazione delle risposte errate per consentire di valutare le modalità di risposta che hanno generato percorsi errati;
- contatore dell'attivazione delle regole di compatibilità (per le regole *hard*) e contatore di quante volte l'attivazione della regola ha determinato un cambiamento della risposta (per le regole *soft*).

6.2.1 - La scheda contatti

La scheda contatti ha la funzione di gestire in automatico i contatti con la famiglia tramite la predisposizione di percorsi guidati e la registrazione dell'esito del tentativo di contatto. Inoltre, in caso di caduta della famiglia del campione base, la scheda contatti permette la sostituzione automatica della famiglia. La scheda si compone di due parti: 1) un'agenda iniziale nella quale sono visualizzate tutte le famiglie e il calendario; 2) un percorso in forma di questionario che guida il contatto iniziale con la famiglia.

Nell'agenda iniziale sono presenti le informazioni derivanti dallo scheda di famiglia fornita dal Comune (nome, cognome e data di nascita dell'intestatario della scheda di famiglia, eventuale coniuge, il numero di componenti della famiglia, l'indirizzo dell'abitazione), la settimana di riferimento e il periodo di tempo fissato per svolgere l'intervista. L'utilizzo del questionario elettronico permette di gestire in automatico il periodo a disposizione per svolgere l'intervista, terminato il quale il rilevatore non potrà più accedere alla famiglia che scomparirà dall'agenda.

Scelta una famiglia dalla schermata iniziale vengono visualizzate le informazioni sulla stessa in dettaglio, lo stato dell'intervista (ad esempio: "nessun contatto", "appuntamento per intervista", "intervista iniziata", ecc.) e la storia dei contatti precedenti avuti con la famiglia con i relativi esiti (tipo di contatto, data del contatto e esito del contatto). Per le waves successive sono anche presenti le principali informazioni sui componenti della famiglia intervistati la wave precedente: il nome, l'età, il sesso, la relazione di parentela con la persona di riferimento, la condizione occupazionale, intervista personale o *proxy*. Una parte dell'agenda, inoltre, è dedicata alla visualizzazione dei recapiti a disposizione per la famiglia con la possibilità di aggiornare le informazioni.

Per effettuare un tentativo di contatto occorre entrare nel percorso guidato e scegliere tra l'effettuazione di un contatto a domicilio o di un contatto telefonico. La sequenza dei quesiti si differenzia a seconda che si tratti del primo contatto con la famiglia o di successivo contatto; in questo ultimo caso si tiene conto delle informazioni registrate in precedenza. I percorsi previsti permettono di ridurre gli errori di sovracopertura e, in alcuni casi, di correggere le informazioni sbagliate. Le domande iniziali, infatti, sono volte ad accertare se la famiglia estratta in anagrafe risponda alle caratteristiche richieste dal disegno campionario, ossia è presente l'intestatario o il coniuge della scheda di famiglia e si tratta dell'abitazione principale, nel caso di prima intervista, oppure è presente almeno un componente di 15 anni intervistato la wave precedente, in successiva intervista. In caso di indirizzo errato o inesistente, si abilita il contatto con il Comune al quale l'intervistatore dovrà rivolgersi per accertare se in anagrafe sono presenti informazioni più aggiornate che permettano di reperire la famiglia.

Gli esiti dei tentativi di contatto con la famiglia possono essere provvisori ("telefono occupato", "in casa non c'è nessuno", "appuntamento per intervista", eccetera), nel qual caso si procederà ad un successivo contatto con la famiglia, oppure definitivi di intervista completa o di caduta della famiglia². La famiglia caduta durante il periodo di rilevazione viene sostituita in automatico con la famiglia successiva della quartina, fino ad arrivare alla quarta famiglia disponibile. La gestione della sostituzione della famiglia in base all'esito del tentativo di contatto, senza la possibilità per il rilevatore di scegliere la famiglia sostituita né di visualizzare le informazioni prima di effettuare la sostituzione, permette di rispettare i criteri di causalità delle sostituzioni eliminando i problemi di discrezionalità nella gestione delle sostituzioni da parte del rilevatore.

La registrazione di tutti gli esiti di tentativo di contatto e la trasmissione degli stessi per via telematica al successivo collegamento con l'Istat da parte del rilevatore consente altresì di costruire con tempestività gli indicatori di monitoraggio dell'indagine. Nel corso della rilevazione, dunque, è possibile tenere sotto controllo l'andamento dell'indagine studiando in modo sistematico l'incidenza dei diversi motivi di caduta delle famiglie, nonché le strategie di contatto più efficaci per riuscire ad ottenere l'intervista. Inoltre viene monitorato il lavoro svolto dai singoli rilevatori considerando più indicatori contemporaneamente (numero interviste complete, tasso di rifiuto, percentuale di interviste *proxy*, eccetera). Ciò consente di predisporre interventi mirati a seconda del verificarsi delle diverse cause di caduta, tenendo sotto controllo la loro concentrazione territoriale e il singolo rilevatore.

² La caduta della famiglia deriva da: a) *errori di sovracopertura*, per decesso, trasferimento in altro comune o all'estero, casa per le vacanze o famiglia che vive in istituzione; b) *esplicito rifiuto* della famiglia a partecipare all'indagine o a terminare un'intervista iniziata; d) *irreperibilità della famiglia*, per assenza della famiglia per tutto il periodo di rilevazione, per eccesso di tentativi di contatto, per indirizzo errato o inesistente; e) *fine del periodo di rilevazione*.

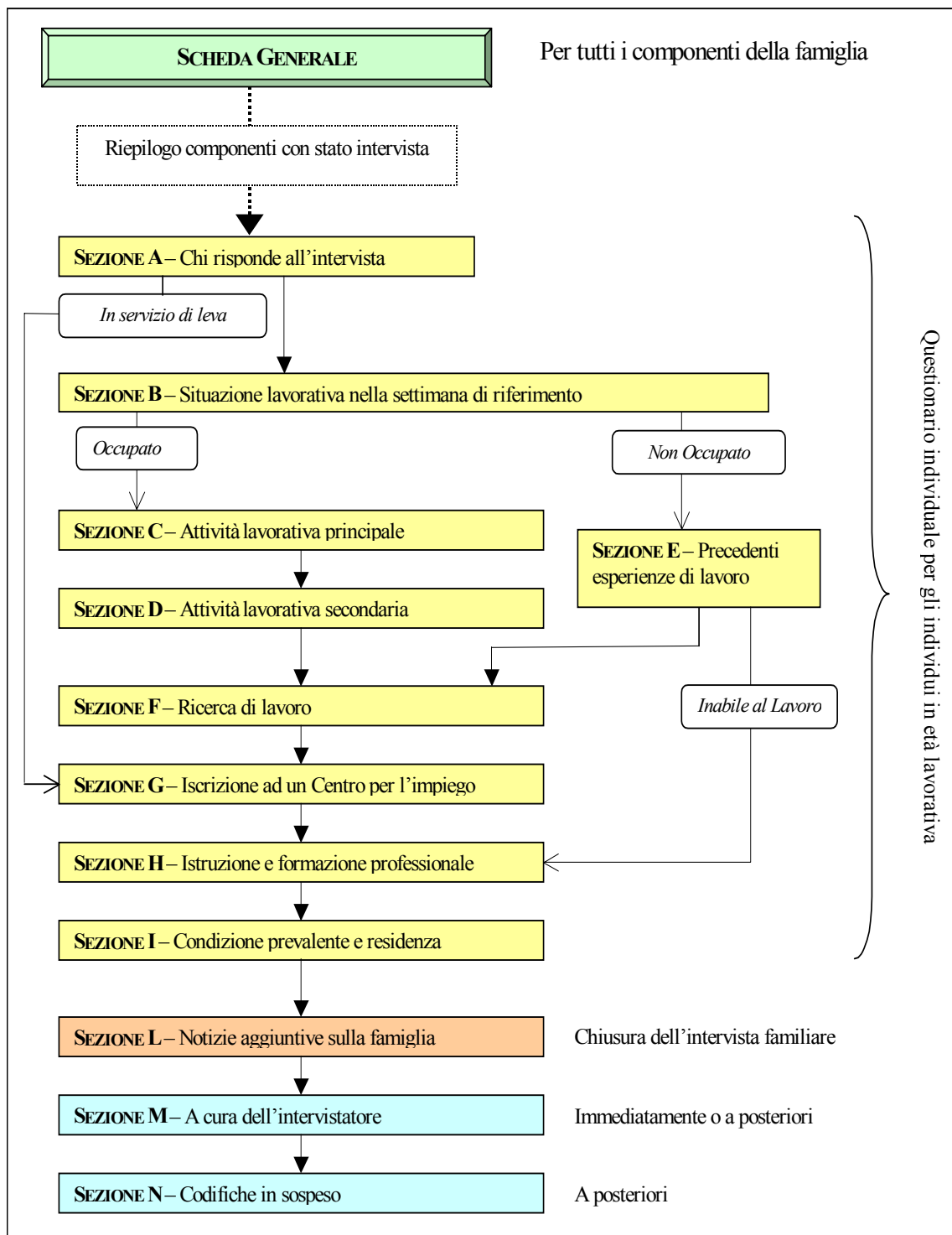
6.2.2 - I percorsi del questionario

Il primo aspetto affrontato nella costruzione del questionario è stato quello della gestione automatica dei percorsi. L'utilizzo della modalità *computer assisted* consente di gestire al meglio la complessità insita nel questionario garantendo una elevata qualità dei dati. Inoltre, nella definizione del questionario elettronico è stata posta una cura particolare all'aspetto grafico. Si è cercato di utilizzare degli standard di visualizzazione che facilitassero il più possibile il lavoro degli intervistatori, compreso il ricorso a differenti colori in relazione alla funzione svolta dal testo.

Il questionario si compone di un scheda generale per la rilevazione delle notizie anagrafiche sulla famiglia, di un questionario individuale di 9 sezioni da ripetere per ogni componente in età lavorativa (con almeno 15 anni), di due sezioni di chiusura dell'intervista familiare e di una sezione per gestire eventuali codifiche in sospeso. Il flusso tra i differenti blocchi è riportato nella figura 6.1.

La modalità scelta per la visualizzazione del questionario e per la gestione del percorso è quella *item based*, con la somministrazione prevalentemente di una domanda per schermata con sequenza fissa e controllata interamente dal computer. Per la somministrazione della prima intervista il flusso è più complesso in quanto tutte le informazioni vengono rilevate per la prima volta.

Figura 6.1 – Struttura del nuovo questionario sulle forze di lavoro



Nella scheda generale si registrano le informazioni anagrafiche, lo stato civile e la cittadinanza di tutti i componenti della famiglia e il titolo di studio dei componenti con almeno 15 anni³. La prima informazione registrata è il numero di componenti della famiglia di fatto. Nella successiva schermata inizia la registrazione delle informazioni sul primo componente. Terminati i quesiti da porre al primo componente il programma passa automaticamente al secondo e così via fino ad arrivare all'ultimo componente della famiglia. Ovviamente, il numero di componenti per i quali si rilevano le informazioni inserite nella scheda generale è legato al numero di componenti della famiglia di fatto e non a quello relativo alla famiglia anagrafica. Alla scheda generale risponde un componente della famiglia maggiorenne per tutti.

Conclusa la scheda generale si passa ad una schermata di riepilogo dove vengono visualizzati il nome, l'età e lo stato dell'intervista individuale di ciascun componente della famiglia. Inizialmente tutti i componenti della famiglia in età lavorativa presentano lo stato dell'intervista "da iniziare", mentre per i componenti con meno di 15 anni l'intervista è conclusa senza rilevare ulteriori informazioni (stato dell'intervista "completa"). Selezionando un individuo si entra nel questionario individuale.

Nella sezione A si registra la data di inizio dell'intervista individuale, chi sta rispondendo ai quesiti sull'individuo selezionato (intervista personale o intervista *proxy*), il motivo dell'eventuale intervista *proxy*, il familiare che risponde in caso di intervista *proxy*. Come si vedrà, oltre a monitorare il numero e la qualità delle interviste *proxy*, le informazioni registrate nella sezione A consentono di gestire la formulazione delle domande e l'apparizione a video di alcune modalità di risposta. Conclusa la sezione A i militari in servizio di leva obbligatorio (per definizione non forze di lavoro) passano direttamente alla sezione G (centro per l'impiego).

In tutti gli altri casi si passa alla sezione B nella quale si rileva la situazione lavorativa nella settimana di riferimento. Le domande previste in questa sezione permettono di accertare se l'intervistato possieda i requisiti stabiliti in sede comunitaria per essere considerato occupato. Gli intervistati classificati come occupati passano alla sezione C e poi alla sezione D, ossia alle sezioni dedicate rispettivamente all'attività lavorativa principale e secondaria; i non occupati, invece, giungono direttamente alla sezione E relativa alle precedenti esperienze di lavoro.

Trattandosi della Rilevazione sulle forze di lavoro, la sezione C dedicata all'attività lavorativa principale costituisce la sezione più articolata del questionario individuale, che si compone complessivamente di sessantatre quesiti; di questi venti sono rivolti a tutti gli occupati: la posizione nella professione, la professione, l'attività economica, l'orario di lavoro abituale e nella settimana di riferimento, l'orario desiderato, le domande relative al lavoro svolto in orari disagiati, il luogo di lavoro e l'anno di inizio lavoro. Le altre domande, somministrate solo in alcuni casi, mirano ad approfondire particolari situazioni lavorative. Ad esempio, sono previsti alcuni quesiti specifici per chi svolge un lavoro alle dipendenze (carattere dell'occupazione, tipo di contratto, eccetera); alcune domande sono rivolte solo a chi svolge un lavoro part-time (motivo del lavoro part-time, desiderio di un lavoro a tempo pieno, eccetera.); altre approfondiscono il motivo della differenza tra l'orario di lavoro dichiarato nella settimana di riferimento e l'orario abituale. Grazie all'utilizzo del questionario elettronico la diramazione dei percorsi occupazionali viene gestita in automatico senza ostacolare la conduzione dell'intervista.

La sezione D, molto più breve rispetto alla sezione precedente (complessivamente di 9 quesiti), rileva la presenza di un'eventuale seconda attività lavorativa e le caratteristiche principali della stessa.

³ La scelta di inserire il titolo di studio più elevato conseguito nella scheda generale nasce dall'esigenza di disporre di questa informazione prima di compilare le sezioni dedicate all'attività lavorativa, per poter inserire dei controlli tra il lavoro svolto e il titolo di studio conseguito.

La sezione E è rivolta agli intervistati che dalla sezione B risultano non occupati. Per prima cosa si accerta se l'intervistato abbia mai svolto un lavoro nel corso della vita. Per chi ha lavorato in passato si rileva l'anno di conclusione dell'ultima attività lavorativa e per chi ha concluso un lavoro negli ultimi sette anni, si rilevano alcune caratteristiche relative all'ultimo lavoro svolto. Gli intervistati che non hanno mai lavorato nella vita perché permanentemente inabili al lavoro saltano le due sezioni successive passando direttamente alla sezione H.

Il percorso degli occupati e dei non occupati, ad eccezione dei permanentemente inabili al lavoro, si ricongiunge nella sezione F dedicata alla ricerca del lavoro. All'interno di questa sezione i percorsi cambiano parzialmente a seconda della condizione occupazionale. Per i non occupati il percorso è più articolato, in quanto questa sezione contiene le informazioni necessarie per individuare le persone in cerca di occupazione. Per gli occupati, invece, il percorso è più breve.

La sezione G è dedicata al rapporto che ha l'intervistato con i centri pubblici per l'impiego, la percezione di un sussidio di disoccupazione, nonché eventuali rapporti con agenzie private di collocamento. A questa sezione rispondono anche i militari in servizio di leva obbligatorio.

La sezione H è dedicata alle attività formative. In questa sezione si rileva l'iscrizione ad un corso scolastico o universitario, la partecipazione nelle ultime quattro settimane ad altri corsi di formazione e il conseguimento di un titolo rilasciato a conclusione di un corso di formazione professionale riconosciuto dalla regione.

La sezione I rileva la condizione occupazionale prevalente percepita dall'intervistato nella settimana di riferimento e l'anno precedente, l'eventuale cambiamento di residenza un anno e due anni precedenti l'indagine. A questo punto l'intervista individuale è conclusa. Se sono presenti altri componenti della famiglia da intervistare si passa alla schermata di riepilogo e si ripete il percorso, dalla sezione A alla sezione I, per il successivo componente. Terminati i componenti da intervistare si passa alla sezione L per la chiusura dell'intervista familiare.

La sezione L, posta all'ultimo componente della famiglia intervistato, registra alcune informazioni sulla famiglia utili ad agevolare il successivo contatto con la famiglia: ulteriori numeri di telefono, nuovi recapiti, e l'orario preferito dalla famiglia per la successiva intervista.

La sezione M a cura dell'intervistatore fornisce alcune informazioni sull'andamento dell'intervista. In questa sezione il rilevatore registra la disponibilità della famiglia a fornire risposte nel corso dell'intervista e la disponibilità ad effettuare una successiva intervista, eventuali difficoltà emerse nel corso dell'intervista e il luogo dove si è svolta la stessa. Considerata la delicatezza dei quesiti, soprattutto nelle situazioni problematiche, il rilevatore può decidere se compilare subito la sezione oppure chiudere l'intervista e compilare la sezione M successivamente.

Infine la sezione N serve a registrare eventuali codifiche della professione e dell'attività economica lasciate in sospeso nel corso dell'intervista utilizzando un differente criterio di codifica. All'inizio della sezione viene proposta una tabella riepilogativa delle codifiche rimaste in sospeso per tutti i componenti della famiglia. Scegliendo il nome di un componente si procede alla codifica delle variabili ad esso relative.

In qualsiasi punto del questionario è possibile interrompere l'intervista, temporaneamente o in modo definitivo. Nel primo caso, il rilevatore fisserà un nuovo appuntamento con la famiglia per completare l'intervista; nel successivo contatto dopo le domande iniziali della scheda contatti l'intervista riprenderà dal punto dove si è interrotta la volta precedente. Nel caso di interruzione definitiva dell'intervista da parte della famiglia si passa alla compilazione della scheda rifiuto e chiusa l'intervista la famiglia scomparirà dall'agenda.

La qualità dell'intervista migliora anche per la possibilità di registrare in automatico alcune variabili, riducendo il numero di domande da sottoporre all'intervistato. Innanzitutto, le variabili relative a data e orari dell'intervista, sia familiare sia individuale, sono registrate in automatico.

Altre variabili sono ricostruibili in automatico a partire dalle informazioni rilevate in precedenti domande. Un esempio sono le tre variabili che, per ciascun componente, rilevano la presenza in famiglia del padre, della madre e del partner (coniuge o convivente). In più del 95 per cento dei casi queste informazioni sono ricostruibili a partire dalla relazione di parentela dell'intervistato con la persona di riferimento (PR) della famiglia. L'implementazione di un insieme di regole di corrispondenza tra la relazione di parentela con la PR e queste variabili permette, nella stragrande maggioranza dei casi, di registrare in automatico l'informazione. In primo luogo per la PR le informazioni sono sempre ricostruibili, controllando la presenza in famiglia del coniuge o del convivente, e dei genitori. Per gli altri componenti l'informazione è spesso comunque ricostruibile⁴.

Trattandosi di un'indagine longitudinale, all'interno di ogni sezione il percorso è molto articolato anche in virtù della scelta di realizzare un unico questionario da utilizzare sia per la prima *waves* di intervista, sia per le successive. L'utilizzo delle informazioni raccolte nelle fasi di intervista precedenti, con la possibilità di gestire le domande *a conferma* nelle *waves* successive e di registrare i cambiamenti dei nuclei familiari verificatisi in successive interviste, semplifica la conduzione delle successive interviste e permette un controllo longitudinale delle informazioni.

6.2.3 - Percorsi a conferma per le successive interviste

Come anticipato, le famiglie sono intervistate quattro volte a distanza di tre, dodici e quindici mesi dalla somministrazione della prima intervista. La possibilità di ricostruire la storia lavorativa dell'intervistato per più di un anno fornisce preziose informazioni di carattere longitudinale sullo stato occupazionale dello stesso. D'altro canto in molti casi le informazioni registrate alla prima intervista non cambiano o mutano solo parzialmente. Per tale ragione il regolamento Eurostat 1897/00 indica la possibilità di abbreviare l'intervista nelle successive *waves* per quei soggetti che mostrano una condizione occupazionale presumibilmente stabile nel tempo, tenendo conto delle informazioni registrate precedentemente. La semplificazione nella raccolta dell'informazione riduce il disturbo statistico per l'intervistato diminuendo il rischio di un suo rifiuto a partecipare all'indagine.

A tal fine il nuovo questionario Rcfl prevede dei percorsi differenti a seconda che si tratti di prima o di successiva intervista. Nel secondo caso si tiene conto dei dati raccolti nelle precedenti *waves*, il che abbrevia l'intervista permettendo allo stesso tempo di controllare la qualità dei dati raccolti in precedenza. In alcuni casi la singola domanda viene posta a conferma, in altri una domanda a conferma permette di registrare più informazioni contemporaneamente, in altri ancora la domanda a conferma ha la funzione di domanda filtro che può comportare o meno la compilazione in automatico anche di un'intera sezione del questionario. Nelle indagini pilota la buona performance del questionario a conferma ha avvalorato la scelta iniziale⁵.

Al fine di uniformare la raccolta delle informazioni per tutte e quattro le *waves* di intervista, così da velocizzare i tempi di elaborazione dei dati e di implementazione di eventuali modifiche

⁴ Soltanto nei casi in cui l'informazione non è ricostruibile il quesito viene somministrato all'intervistato, il che può accadere per le famiglie dove sono presenti più nuclei familiari (ad esempio, se in famiglia sono presenti due fratelli di PR e un nipote di PR, in automatico non è ricostruibile chi dei due fratelli è il genitore del nipote di PR).

⁵ Dal monitoraggio e dai debriefing con i rilevatori è emersa la reticenza degli intervistati a concedere un'ulteriore intervista della stessa complessità e lunghezza, in particolar modo quando la situazione lavorativa risultava invariata rispetto alla precedente intervista; una delle obiezioni più frequenti recitava: "Non è cambiato niente, tutto come la scorsa intervista, perché mi fa sempre le stesse domande?". L'utilizzo di un questionario con percorsi a conferma che velocizzano la raccolta dell'informazione, invece, ha comportato una maggiore disponibilità a partecipare all'indagine. Tra le altre cose, l'intervistato è motivato a rispondere perché percepisce un buon utilizzo e la correttezza delle informazioni fornite nel corso della *wave* precedente.

del questionario, si è costruito un questionario unico per la prima e per le successive interviste. Il questionario unico consente di registrare le situazioni di modifica parziale del nucleo familiare, registrando l'uscita di alcuni componenti e/o la presenza di nuovi componenti in *waves* successive, prevede la possibilità di cambiamenti parziali nel percorso di intervista di un individuo presente in più occasioni di indagine, associa a ciascun individuo un codice univoco che rimane invariato nel corso delle successive *waves*. Questo ultimo aspetto riveste particolare importanza per la costruzione dei file longitudinali permettendo l'accoppiamento degli individui intervistati in più occasioni di indagine in modo automatico e senza rischio di errore⁶.

Tutto ciò ha comportato un lungo lavoro per progettare e scrivere le regole che permettessero di far funzionare correttamente il questionario nelle diverse situazioni. Illustriamo brevemente la logica seguita e le parti del questionario che presentano quesiti specifici per la reintervista.

La maggior parte dei quesiti a conferma si trovano nella scheda generale, poiché si riferiscono a informazioni stabili nel tempo (il sesso, la data e il luogo di nascita, eccetera.), o a informazioni che cambiano raramente (quali lo stato civile o il titolo di studio).

In generale, il rilevatore confermerà o meno la risposta registrata nella *wave* precedente ed è sempre prevista la possibilità di non confermare. Il quesito a conferma è inserito subito prima del quesito al quale si riferiscono le informazioni presentando la stessa numerazione del quesito non a conferma con il suffisso c. Ad esempio, per la rilevazione del titolo di studio più elevato conseguito dall'intervistato il quesito a conferma e quello non a conferma presentano il seguente ordine di apparizione:

Per le waves successive:

Per il programma: porre il quesito solo se SG.24 precedente intervista diverso da missing o non sa o non risponde; altrimenti passare a SG.24

SG24c Mi conferma che "NOME" ha "DESCRITTIVO TITOLO DI STUDIO SG24 PRECEDENTE INTERVISTA" esempio: " <i>Mi conferma che John ha la licenza media?</i> "	
Conferma	1 _ porre SG.24=SG.24 precedente intervista e seguire i filtri di SG.24
Non conferma	2 _ passare a SG.24

Per la prima intervista:

SG.24 Qual è il titolo di studio più elevato che "NOME" ha conseguito? esempio: " <i>Qual è il titolo di studio più elevato che John ha conseguito?</i> "	
Nessun titolo	1 _ passare a SG.28
Licenza elementare	2 _ passare a SG.26
Licenza media	3 _ passare a SG.26
.....

Il quesito a conferma viene posto solo nel caso il quesito non a conferma relativo all'intervista precedente abbia un valore diverso da "*missing*" o da "non sa" o da "non risponde". In tutti gli altri casi si passa direttamente al quesito non a conferma. Questa istruzione di programma, per filtrare la somministrazione del quesito a conferma, presenta il vantaggio di risultare semplice da implementare consentendo il passaggio al quesito non a conferma in situazioni molto diverse: a) nella prima intervista (*missing*); b) per un nuovo componente entrato a far parte della famiglia in una successiva intervista (*missing*); c) per un componente presente alla precedente intervista che ha cambiato percorso del questionario, ad esempio da non

⁶ Nella Rtf1, al contrario, uno dei problemi principali per la costruzione dei files longitudinali degli individui era la mancanza di un codice univoco per gli individui. Infatti, utilizzando un questionario cartaceo, nelle successive interviste i componenti della famiglia potevano essere registrati seguendo un ordine diverso rispetto all'intervista precedente. L'accoppiamento degli individui intervistati in successive *waves* veniva realizzato utilizzando un insieme di variabili (data di nascita, nome, titolo di studio, etc.) e tramite dei modelli probabilistici nel caso di discordanza di alcune informazioni.

occupato è diventato occupato (*missing*); d) nella *wave* precedente l'intervistato non ha fornito una risposta (non sa o non risponde).

Se l'informazione viene confermata dall'intervistato automaticamente si registra la risposta precedentemente raccolta nel relativo quesito non a conferma e si seguono i filtri corrispondenti allo stesso. Al contrario, se il dato rilevato precedentemente non viene confermato – perché si tratta di una nuova condizione rilevata o di un errore nell'intervista precedente – al rispondente si porrà nuovamente il quesito non a conferma per registrare il cambiamento di stato.

In alcuni casi il quesito a conferma contiene il contenuto di più domande semplificando ulteriormente la raccolta dell'informazione. Ad esempio, questo è il caso degli intervistati con cittadinanza straniera:

Per il programma: se SG.16 prima intervista=1 in "CITTADINANZA" visualizzare "italiana"; se SG.16 prima intervista=2 in "CITTADINANZA" visualizzare il descrittivo di SG.17 precedente intervista

SG.16c Mi conferma che "NOME" ha la cittadinanza "CITTADINANZA"?

Si 1 (porre SG16= SG.16 precedente intervista, SG.17=SG.17 precedente intervista, e passare a SG.18c)

No 2 (se SG.16 precedente intervista=2 passare a SG.16; se SG.16 precedente intervista=1 passare a SG.17)

SG.16 "NOME" ha la cittadinanza italiana?

Si 1 (passare a SG.19)

No 2

SG.17 Di quale stato estero ha la cittadinanza?

Descrizione _____ Codice

Nella progettazione del questionario unico la parte più complessa da progettare è quella della scheda generale per le situazioni in cui nella successiva intervista si verifica un cambiamento del nucleo familiare.

Come visto, nel corso della prima *wave* il numero di componenti per i quali si rilevano le informazioni inserite nella scheda generale corrisponde al numero di componenti della famiglia di fatto. Diversamente, nella *wave* successiva, registrato il numero di componenti della famiglia di fatto, prima di tutto si accerta la presenza degli individui intervistati nella *wave* precedente. Per i componenti ancora presenti si registrano le informazioni anagrafiche chiedendo conferma di quelle già registrate la *wave* scorsa. Per i componenti precedentemente rilevati che non fanno più parte della famiglia al momento della nuova intervista si rilevano alcune informazioni volte ad individuare il motivo principale per il quale l'individuo non fa più parte della famiglia, l'eventuale nuova provincia di residenza e la condizione occupazionale. Queste informazioni sono molto utili per studiare la mobilità occupazionale e/o territoriale degli individui⁷.

Registrate le informazioni su tutti i componenti intervistati alla precedente *wave*, il software procede al calcolo della differenza tra il numero di componenti della famiglia di fatto nella *wave* attuale e il numero di componenti intervistati nella precedente rilevazione e ancora presenti. Nel caso la differenza sia uguale a zero (nessun cambiamento nel nucleo familiare) o negativa (uscita di alcuni componenti rilevati la precedente *wave*) si conclude la scheda generale.

Contrariamente, se la differenza è positiva appare a video un messaggio che avverte il rilevatore che sono entrati a far parte "n" nuovi componenti, con n uguale alla differenza tra il numero di componenti della famiglia di fatto nella *wave* attuale e il numero di componenti

⁷ Si tenga presente, infatti, che il disegno campionario non prevede di inseguire gli individui che escono dal nucleo familiare, ad esempio come nel caso del figlio che si sposa e va a costituire un nuovo nucleo familiare.

ancora presenti in famiglia dalla *wave* precedente. Il messaggio cambia a seconda che sia entrato a far parte della famiglia un nuovo componente o più componenti:

SCRIPT SG.1:

Per il programma: se [(numero componenti famiglia di fatto) – (somma dei componenti presenti alla precedente wave e ancora presenti in famiglia) =1] visualizzare il seguente messaggio:

Rispetto alla rilevazione precedente, è entrato a far parte della famiglia un nuovo componente. Le prossime domande si riferiscono a lui
--

Per il programma: se [(numero componenti famiglia di fatto) – (somma dei componenti presenti alla precedente wave e ancora presenti in famiglia) >1] visualizzare il seguente messaggio:

Rispetto alla rilevazione precedente, sono entrati a far parte della famiglia “[numero componenti famiglia di fatto) – (somma dei componenti presenti alla precedente wave e ancora presenti in famiglia)]” nuovi componenti. Le prossime domande si riferiscono a loro.

Dopodiché, in automatico il software attiverà il percorso di prima intervista (tutti i quesiti non a conferma) per registrare le informazioni relative ai dati anagrafici dei componenti entrati a far parte della famiglia al momento della nuova intervista. Per i nuovi componenti il codice identificativo individuale viene assegnato in automatico assumendo un valore progressivo a partire dal valore successivo a quello più elevato registrato nelle *waves* precedenti.

Ulteriori specifiche si sono rese necessarie per differenziare l'identificazione del PR in prima e in successive interviste. Nella prima *wave*, come detto, il questionario include le informazioni derivanti dalla scheda di famiglia fornita dal Comune di residenza dell'intervistato. L'intestatario della scheda di famiglia è il primo componente sul quale si registrano le informazioni. Se l'intestatario è presente in famiglia questo diviene la persona di riferimento della famiglia, ossia quella in base alla quale vengono registrate le relazioni di parentela. Se l'intestatario non è più presente in famiglia, il coniuge o il convivente assume il ruolo di persona di riferimento.

Nelle *waves* successive, invece, la prima persona sulla quale si chiedono le informazioni è la PR alla precedente intervista. Quando la PR è ancora presente si registrano/confermano le informazioni anagrafiche e poi, se presente, si passa al componente successivo. Se la persona di riferimento non fa più parte della famiglia, occorre stabilire una nuova PR. Nel programma è stata inserita una regola di selezione dei componenti presenti alla precedente intervista seguendo un criterio di parentela e di età. Il software propone la persona che risponde a determinati requisiti per diventare la nuova PR che sarà il coniuge, se presente, altrimenti il primo componente che soddisfa determinate caratteristiche e nel caso in cui siano più di uno a soddisfarle, l'individuo che tra questi possiede l'età più elevata. Logicamente quando cambia la persona di riferimento per i restanti componenti le informazioni relative alla relazione di parentela con la persona di riferimento non vengono poste a conferma⁸.

Nel questionario individuale i percorsi a conferma sono previsti in particolar modo nella sezione C, e per alcuni quesiti nelle sezioni D, H e I.

Nella sezione C i quesiti a conferma riguardano le caratteristiche principali del lavoro svolto: la posizione della professione, la professione, l'attività economica, la durata del lavoro, il motivo del lavoro a termine, la distinzione tra lavoro a tempo pieno o part-time, il motivo del lavoro part-time, il luogo di lavoro e l'anno di inizio lavoro. Si tenga presente, però, che se l'intervistato non conferma le prime informazioni – posizione nella professione, professione e attività economica – i restanti quesiti della sezione C verranno posti non a conferma. Questo per

⁸ La scheda generale presenta una portata più ampia in confronto al questionario individuale legato ai contenuti dell'indagine. Per cui la strategia proposta è estendibile a tutte le indagini longitudinali per le quali è prevista una scheda anagrafica su tutti i componenti della famiglia prima di iniziare le interviste individuali specifiche sui contenuti dell'indagine.

il fatto che, presumibilmente, nel tempo intercorso tra le due rilevazioni l'intervistato ha cambiato lavoro. Tale condizione aggiunge un'ulteriore istruzione di programma per porre i quesiti a conferma in questa sezione.

Nella sezione D l'unico quesito a conferma riguarda l'attività economica relativa al secondo lavoro.

Nella sezione H si chiede conferma del corso scolastico o universitario seguito e dell'eventuale titolo rilasciato da un corso di formazione professionale riconosciuto dalla regione. In questo ultimo caso, se si tratta dello stesso corso rilevato nella *wave* precedente, le domande relative alle caratteristiche del corso (titolo di studio necessario per accedere al corso, argomento del corso, durata del corso, anno di conclusione del corso) non vengono poste all'intervistato, ricavando l'informazione direttamente dalla *wave* precedente.

Nella sezione I le domande a conferma riguardano le informazioni relative ad un anno precedente: condizione professionale percepita, le caratteristiche del lavoro svolto un anno prima e la residenza. A differenza degli altri quesiti a conferma presenti nel questionario, per i quali si fa riferimento alla *wave* di intervista precedente, in questo caso le informazioni sono ricavate dalla *wave* di intervista realizzata un anno prima. Per cui i quesiti a conferma sono presenti nelle *wave* 3 e 4, in relazione alle informazioni rilevate rispettivamente alle *wave* 1 e 2.

Un'ultima particolarità del questionario riguarda la sezione E, relativa alle precedenti esperienze di lavoro dei non occupati, dove è previsto un percorso breve in fase di reintervista. In questo caso, infatti, se l'intervistato non ha iniziato e terminato un nuovo lavoro nel periodo intercorso tra le due interviste le informazioni rilevate in precedenza rimangono invariate.

All'inizio della sezione è presente la seguente domanda filtro:

Per wave successiva: porre il quesito solo se E1 precedente intervista diverso missing, cioè non era occupato ed aveva risposto alla sezione E; altrimenti passare a E1

E1c. Negli ultimi "N-MESI" "NOME" ha svolto un lavoro per il quale ha ricevuto un guadagno?

PER L'INTERVISTATORE: gli ultimi "N-MESI" si concludono "DOMENICA SETTIMANA DI RIFERIMENTO"

- | | |
|---------------------------------------|---|
| Sì | 1 _ |
| No | 2 _ (porre da E1 a E18 come E1-E18 precedente wave e passare a sezione F) |
| No, permanentemente inabile al lavoro | 3 _ (porre da E1 a E18 come E1-E18 precedente wave e passare a sezione H) |

Nel testo della domanda, al posto di "n-mesi", comparirà il numero di mesi trascorso tra le due successive interviste. Se l'intervistato risponde affermativamente verrà nuovamente posta l'intera sezione E. Se, invece, l'intervistato risponde di non aver lavorato nel periodo considerato il programma procede a registrare in automatico tutte le informazioni della sezione E rilevate alla precedente intervista.

6.2.4 - Le potenzialità del questionario elettronico: personalizzazione dei quesiti

Oltre alla gestione automatica dei filtri tra successive domande il questionario della RcfI si è avvalso di altri elementi d'ausilio alla conduzione dell'intervista. La progettazione del questionario ha previsto specifiche regole per personalizzare la formulazione dei quesiti, nei casi in cui ciò comportava un miglioramento nella qualità dell'informazione rilevata.

Innanzitutto per i quesiti riferiti ad un determinato periodo temporale legato alla settimana di riferimento viene visualizzata in automatico la data variando la formulazione a seconda dei casi. Ad esempio, nella sezione B le prime due domande presentano la seguente formulazione:

B1. “LA SCORSA SETTIMANA” “NOME” ha svolto almeno un’ora di lavoro? Consideri il lavoro da cui ha ricavato o ricaverà un guadagno o il lavoro non pagato solo se effettuato abitualmente presso la ditta di un familiare

- Sì 1|_| (passare a sezione C)
No 2|_|
Permanentemente inabile al lavoro 3|_| (passare a sezione E)

B2. Sempre nella settimana che va “DA LUNEDI’ ... A DOMENICA...” “NOME” aveva comunque un lavoro che non ha svolto, ad esempio: per ridotta attività, per malattia, per vacanza, per cassa integrazione guadagni, etc.?

IN CASO DI NECESSITÀ LEGGERE: Consideri il lavoro da cui ha ricavato o ricaverà un guadagno o il lavoro non pagato solo se effettuato abitualmente presso la ditta di un familiare

- Sì 1|_|
No 2|_| (passare a sezione E)

Nel quesito B1 la prima parte della domanda varierà in base alla data nella quale si sta somministrando l’intervista. Se l’intervista viene svolta la settimana successiva alla settimana di riferimento comparirà “*la settimana scorsa*”, se svolta due settimane dopo la settimana di riferimento comparirà “*due settimane fa*”, e così via fino ad arrivare al tempo massimo per effettuare l’intervista di cinque settimane. Nel quesito B2, invece, si visualizzerà la data relativa alla settimana di riferimento associata a quella famiglia (ad esempio: “*da lunedì 22 a domenica 28 marzo*”). Considerato l’andamento stagionale di molti fenomeni legati al mercato del lavoro, la possibilità di inserire il periodo di riferimento dati nel testo della domanda assume particolare rilevanza.

Un’altra personalizzazione nella formulazione delle domande è prevista per le interviste *proxy*. Quando risponde l’interessato la domanda è formulata in seconda persona: “*Lei ha svolto almeno un’ora di lavoro...*”. Nel caso di intervista *proxy*, invece, la domanda viene posta in terza persona e nel testo della domanda appare il nome del componente della famiglia al quale si riferiscono le informazioni; ad esempio: “*Mario ha svolto almeno un’ora di lavoro...*”. In questo modo si migliora la fluidità dell’intervista facilitando l’intervistatore nel porre le domande, nonché si evita il rischio di confondere nella stessa intervista informazioni appartenenti a individui diversi.

Cambiamenti nella formulazione della domanda sono previsti anche in relazione a particolari caratteristiche dell’intervistato, rilevate nel corso dell’intervista. Ad esempio, l’attività economica è un’informazione che assume un significato parzialmente diverso per i lavori dipendenti e per i lavoratori autonomi. Nel primo caso l’attività economica si riferisce all’attività principale svolta dall’azienda nella quale lavora l’intervistato, oppure, se lavora in un’azienda con più sedi, all’attività economica svolta nella sede in cui lavora. Per i lavoratori autonomi, invece, l’attività economica coincide con il lavoro svolto e/o i principali beni o servizi prodotti. Per tale ragione, il quesito relativo all’attività economica prevede tre diverse formulazioni:

Formulazione per i dipendenti che lavorano in una azienda con una sola sede

C15. Cosa fa l’Ente o l’Azienda presso la quale “NOME” lavora? Indichi i principali beni e/o servizi prodotti

Formulazione per i dipendenti che lavorano in una azienda con più sedi

C15. Quale attività viene svolta nella sede in cui “NOME” lavora? Indichi i principali beni e/o servizi prodotti

Formulazione per i lavoratori autonomi

C15. Qual è l’attività principale che “NOME” svolge e quali sono i beni e/o servizi prodotti?

Il questionario elettronico permette altresì di visualizzare all’occorrenza solo le modalità di risposta attinenti ad un particolare profilo, evitando di registrare un’informazione errata o di dovere implementare una regola di incompatibilità tra successive risposte.

In primo luogo, particolarmente efficace è stata la scelta di gestire la modalità “*non sa*” a seconda di chi risponde alle domande del questionario individuale. È noto, infatti, che la presenza di questa modalità tra le alternative di risposta può indurre il rispondente ad uno scarso impegno a collaborare per i quesiti che richiedono uno sforzo di memoria. Tuttavia, quando la risposta non è fornita direttamente dall’interessato, qualora il rispondente non fosse in grado di rispondere la mancanza di questa modalità di risposta impedirebbe il proseguimento dell’intervista o indurrebbe il rilevatore a selezionare una risposta fittizia falsando i dati. Grazie al questionario elettronico è stato possibile utilizzare criteri diversi per la gestione delle mancate risposte parziali a seconda se risponde direttamente l’interessato o un altro componente della famiglia (risposta *proxy*). Per molte domande la modalità di risposta “*non sa*” viene visualizzata solo in caso di intervista *proxy*. Questa strategia ha permesso di contenere le mancate risposte parziali e le interruzioni definitive.

In altri casi, le modalità di risposta visualizzate cambiano in relazione al profilo dell’intervistato. Ad esempio, nel quesito dove si rileva il motivo principale per il quale la settimana di riferimento l’intervistato ha effettuato un numero di ore maggiori rispetto all’orario abituale le modalità di risposta visualizzate cambiano parzialmente per i lavoratori dipendenti e per gli autonomi:

C33. Qual è il motivo principale per cui “NOME” ha lavorato più del solito?

per i lavori dipendenti visualizzare le modalità 1, 2, 3, 996

per i lavoratori autonomi visualizzare la modalità 1, 4, 996

Orario variabile, flessibile	1 _
Straordinario retribuito	2 _
Straordinario non retribuito (ore in più non recuperabili)	3 _
Maggiori impegni di lavoro	4 _
Altro (specificare)(C33_a)_____	996 _

Di particolare interesse è l’utilizzo di motori di ricerca per la codifica di variabili fondamentali dell’indagine quali, ad esempio, l’attività lavorativa e la professione. In questi ultimi casi, data la complessità dell’informazione rilevata e la rilevanza della stessa per l’indagine, si è previsto un doppio criterio di codifica. Il rilevatore può scegliere se codificare subito tramite motore di ricerca o a posteriori nella sezione N seguendo un diverso criterio. Ad esempio, per la codifica della professione nella sezione C sono previste tre domande:

C11. Può dirmi il nome della sua professione?

C12. Codifica della professione tramite motore di ricerca (visualizzazione di 6.300 nomi di professioni)

1.1.1.1. Membro del consiglio dei ministri

1.1.1.1. Membro del senato

.....

1.1.1.2 Membro del consiglio regionale

.....

9.9.9.9. Codifica non effettuata

Per il programma: Se C12 diverso da 9.9.9.9., cioè riesce a codificare, registrare la C12_d (nome professione) e la C12_c (codice professione a 4 digit) secondo la tabella delle professioni e passare a C14; altrimenti, se C12=9.9.9.9., cioè non riesce a codificare, passare a C13

C13. Può descrivermi con precisione in che consiste il suo lavoro?

In primo luogo, una domanda aperta registra il nome della professione dichiarato dall'intervistato (C11). Nella schermata successiva, viene visualizzato l'elenco delle 6.300 voci professionali, ossia esempi di professioni, le quali corrispondono a 519 categorie della classificazione delle professioni⁹. Tramite una parola chiave il rilevatore attiverà il motore di ricerca per l'individuazione della voce corrispondente alla professione dell'intervistato. Il rilevatore può effettuare più tentativi inserendo altre parole chiave. Se, dopo successivi tentativi non riesce a trovare la voce corrispondente, sia per motivi di tempo sia per la mancanza dell'esempio nella lista, assegnerà il codice 9.9.9.9 che indica "codifica non effettuata". In questo ultimo caso, passerà ad una domanda aperta (C13) nella quale dovrà descrivere nel dettaglio il lavoro svolto dall'intervistato e la codifica verrà effettuata a fine intervista, nella sezione N.

Nella sezione N la codifica viene effettuata per mezzo di un programma a finestre che consente di scorrere facilmente l'intera classificazione rispettando la logica gerarchica della stessa¹⁰. Nella prima schermata appariranno i 9 Grandi Gruppi. Selezionando uno di essi si visualizzano i Gruppi appartenenti al Grande gruppo; con un'ulteriore selezione le classi corrispondenti al Gruppo selezionato, e proseguendo in modo analogo le categorie. Una volta individuata la categoria professionale, la codifica è effettuata selezionando il codice corrispondente.

In automatico vengono visualizzate le selezioni effettuate nelle schermate precedenti e alcune informazioni di ausilio alla codifica della professione (il titolo di studio conseguito

⁹ La classificazione delle professioni presenta una struttura gerarchica ad albero a 4 digit, secondo la quale via via si specificano le caratteristiche del tipo di lavoro svolto fino ad arrivare alla codifica della professione. Progressivamente si scende di dettaglio: il primo livello identifica i Grandi gruppi (9), il secondo i diversi Gruppi (37), il terzo le classi (121) e il quarto le categorie professionali (519). Per ciascuna categoria è previsto un certo numero di voci professionali che identificano esempi di professioni associati a una determinata categoria e, per tale ragione, più voci presentano lo stesso codice a 4 digit. Le voci professionali (nel complesso 6.300) sono esempi di professioni che appartengono ad una determinata categoria, ma non esauriscono le professioni appartenenti alla categoria. Sulla logica della classificazione delle professioni vedi "Classificazione delle professioni", Istat, 2001. La classificazione italiana è armonizzata alla classificazione internazionale delle professioni Isco-88, edita dall'Ilo (1990).

¹⁰ Vedi nota precedente.

dall'intervistato, il nome della professione, la descrizione del lavoro svolto, e l'attività dell'azienda).

Il questionario elettronico consente altresì di inserire degli *help on line* da attivare su richiesta del rilevatore per chiarire all'intervistato il significato della domanda o di una particolare modalità di risposta. Per le domande più complesse accanto a ciascuna modalità di risposta è posto un pulsante con la scritta "help", selezionando il quale il rilevatore apre una schermata dove è spiegato nel dettaglio il significato della modalità di risposta.

L'utilizzo di *help online* facilita la conduzione dell'intervista tramite tecniche di *probing* assistite, facilitando l'intervistatore nel fornire eventuali spiegazioni all'intervistato e, di conseguenza, nello scegliere la risposta corretta.

6.2.5 - I controlli interni al questionario

La rilevazione *computer assisted*, a differenza di quella basata sul questionario cartaceo, offre la possibilità di effettuare controlli sull'ammissibilità di una risposta prima che sia posta la successiva domanda (Baker, Bradburn e Johnson, 1995). L'eventuale incompatibilità della risposta rispetto ad altri dati registrati nel sistema permette, nella maggior parte dei casi, di ottenere informazioni aggiuntive per risolvere subito l'incompatibilità. Il beneficio derivante dall'utilizzo di tecniche *computer assisted* piuttosto che del questionario cartaceo, in cui il controllo può essere svolto solo dopo la registrazione dei dati e senza una reale possibilità di confronto con il rispondente, è dunque notevole. La possibilità di effettuare i controlli già in fase di rilevazione dei dati (controlli a caldo) si traduce in una più elevata qualità dei dati raccolti.

I controlli a caldo permettono di prevenire il realizzarsi di diverse tipologie di errore (Fabbris, Bassi, 1997), causati o dal rispondente o dal rilevatore:

- **errori di range** sul dominio delle variabili;
- **errori di rotta** riguardanti il percorso da seguire nella somministrazione del questionario;
- **errori di consistenza o di compatibilità**, ossia di coerenza tra quanto viene risposto ad una domanda rispetto a quanto è stato dichiarato in una o più domande precedenti;
- **errori di verosimiglianza**, ossia di coerenza tra quanto viene risposto ad una domanda rispetto alla probabilità del verificarsi di quella modalità, essendo nota la distribuzione della variabile;
- **errori longitudinali**, sia di coerenza tra quanto affermato in *waves* successive di intervista su fenomeni osservati ripetutamente nel tempo, sia di mancato abbinamento dei record individuali causato da cambiamenti nei codici identificativi degli individui tra una *waves* e l'altra o da errata trascrizione dei codici identificativi degli individui.

La prevenzione dell'inserimento di un dato errato avviene attraverso messaggi esplicativi dell'errore che compaiono sullo schermo appena viene inserita una risposta che contrasti con le *regole di correttezza*.

Allo scopo di decidere l'ammissibilità di una risposta tramite controlli di compatibilità e verosimiglianza, si utilizzano, in genere, le seguenti procedure: a) la conciliazione tra le risposte incoerenti; b) il *probing* della posizione del rispondente.

Nel primo caso, si tenta di stabilire, insieme al rispondente, quale tra le informazioni poste a confronto, oppure utilizzate per determinare la verosimiglianza dell'ultima risposta, sia

attendibile. In altre parole, si rende evidente al rispondente che la risposta data non è coerente con quanto ha affermato prima o in tempi differenti, oppure con informazioni acquisite esternamente. L'interpellato, posto in contraddizione, fornisce una risposta che concilia l'incoerenza registrata. La modalità conciliante può essere l'ultima, una delle precedenti, o una nuova modalità (Bates, 1996).

Il *probing* si attua con una o più domande che "prendono alla larga" l'argomento, al fine di aiutare il rispondente a formarsi un'idea più precisa di ciò che viene rilevato nell'indagine ed a trovare la risposta nella propria memoria, allo scopo di ottenere informazioni corrette. Il *probing* può essere utile anche per eliminare mancate risposte o risposte elusive.

Esistono diversi modi di trattare le informazioni incoerenti. Quando l'intervista non può proseguire se non viene corretto il dato errato, l'errore viene trattato in modalità **Hard**. L'*Hard check* è un errore giudicato rigidamente inammissibile in fase di progettazione del questionario, e per questa ragione si ritiene che debba essere necessariamente risolto. Quando invece è lasciata all'operatore la scelta di effettuare o meno la correzione, l'errore viene trattato in modalità **Soft**. Il *Soft check* è piuttosto un *warning (attenzione)*: se l'intervistatore giudica il dato corretto, può non modificarlo.

In fase di progettazione del questionario della Rcfl, per sfruttare appieno le potenzialità del computer, sono stati inseriti i seguenti controlli:

- controlli di range;
- controlli di rotta;
- controlli trasversali di consistenza e verosimiglianza;
- controlli longitudinali.

La strategia utilizzata per la costruzione delle regole è stata incentrata sull'esperienza.

Potendo sfruttare un periodo di due anni di prova, tutte le regole inserite sono state testate.

In alcuni casi le regole sono risultate non pertinenti, in altri, da *Soft* sono diventate *Hard*.

Nei primi test del questionario quasi tutti i *range* sono stati impostati ampi e in alcuni casi addirittura non sono stati fissati, in modo da studiare la distribuzione delle variabili e la presenza di valori anomali. Inoltre le regole di compatibilità e di coerenza sono state prevalentemente impostate come regole *Soft*.

Le regole definitive sono state inserite solo successivamente all'analisi dei dati riguardanti le distribuzioni delle variabili e dei suggerimenti dei rilevatori raccolti durante i *debriefing*. Ciò ha permesso di ricavare utili indicazioni per valutare la validità delle regole stesse, la loro aderenza con la realtà indagata e la correttezza e la comprensibilità dei quesiti posti.

Per tutti i tipi di regole è previsto un contatore in grado di sommare, a livello individuale, il numero di volte in cui ciascuna regola viene attivata. È stato sviluppato, inoltre, un contatore per segnalare il numero di volte in cui viene corretta ciascuna delle variabili coinvolte nella regola attivata.

6.2.6 - I controlli di range

I controlli di *range* sono stati inseriti principalmente per evitare gli errori di digitazione.

Complessivamente 27, i controlli di *range* per le variabili quantitative sono distribuiti in quasi tutte le sezioni.

Nella scheda generale vi sono i controlli sul numero di componenti (*SG19_ANNO*), sulla data di nascita (*SG19_ANNO*) o, in sua mancanza, sull'età dichiarata (*SG20_ANNO*), sull'anno di matrimonio.

Nel questionario individuale il numero maggiore di controlli di *range* si trova nella sezione C (dedicata agli occupati) e riguardano: il numero di addetti (*C19*), la durata del lavoro a termine (*C21*), le ore di lavoro abituali (*C31*), le ore di straordinario (*C36*), le ore di lavoro nella settimana (*C37*), le ore di lavoro desiderate (*C39*), l'anno di inizio lavoro (*C55*) o l'età di inizio lavoro (*C56*), l'anno di inizio del primo lavoro (*SG61*) o l'età di inizio del primo lavoro (*SG62*), e il reddito (*SG63*).

Nella sezione D vi è un solo controllo sulle ore di lavoro per la seconda attività (*D2*).

Nella sezione F i controlli riguardano i mesi trascorsi dall'ultima azione di ricerca (*F31*), i mesi di ricerca (*F31*), il reddito minimo richiesto (*F42*) e i mesi di ricerca attività del lavoro che sta per iniziare (*F43*).

Come anticipato, in fase di test, alcuni *range* non sono stati fissati con l'obiettivo di studiare la distribuzione delle variabili. Tale strategia ha comportato, diversamente dall'uso comune, la previsione di controlli misti *Hard-Soft*, definendo, all'interno dell'intervallo dei valori validi, dei sotto-intervalli da segnalare come *attenzione*, e lasciando la possibilità all'intervistato di conciliare le informazioni fornite oppure no.

Ad esempio, per la variabile *C31* (ore di lavoro abituali), la distribuzione dei dati ha portato a fissare il limite superiore del *range* a 105 ore, con attivazione del controllo *Soft* per valori superiori a 84.

Anche per le domande del questionario *C55* (anno di inizio lavoro) e *C56* (età di inizio lavoro) si è seguita la medesima logica. In quest'ultimo caso, il limite superiore è stato fissato pari all'età attuale del rispondente, mentre il limite inferiore pari a 6 anni, con attivazione del controllo *Soft* per valori compresi tra i 6 e i 13 anni.

6.2.7 - I controlli trasversali di consistenza e verosimiglianza

I controlli trasversali di coerenza e compatibilità sono stati progettati in funzione di ciò che si rileva in ogni singola *wave*, sfruttando le conoscenze a priori sulle variabili e sulle relazioni esistenti tra le modalità di risposta.

Tali controlli, in complesso 118 divisi in *Hard e Soft*, riguardano sia i controlli di coerenza tra le notizie anagrafiche dei componenti della famiglia (regole familiari), sia i controlli di coerenza tra le risposte fornite dall'intervistato durante la rilevazione (regole trasversali). Ciascuna di queste regole individua una possibile incompatibilità tra due o più modalità di variabili presenti nel questionario.

Per studiare il fenomeno, in fase di test il numero di regole *hard* è stato limitato al fine di non forzare l'intervista chiedendo una risoluzione obbligata dell'incompatibilità, se non in casi veramente certi.

In fase di test solo le regole familiari sono state prevalentemente impostate come *Hard*. In appena l'0,8 per cento dei casi esaminati in fase di test si è attivata almeno una regola *Hard*. La regola che, tra quelle familiari, si è attivata più di frequente (0,4% dei casi) ha riguardato le variabili "Relazione di parentela" (*SG12*) e "Stato civile" (*SG22*), per le quali era previsto che la persona di riferimento ed il coniuge dovessero avere entrambi stato civile coniugato e uguale data di matrimonio. Le correzioni non sono mai state sulla relazione di parentela ma sempre sulla variabile "Stato civile".

Nella scheda generale dopo aver rilevato le informazioni su tutti i componenti del nucleo familiare e le rispettive relazioni di parentela con PR, per ciascun componente si rilevava la presenza in famiglia del padre, della madre e del partner. Dall'analisi dei risultati è emersa una

possibile incoerenza tra lo stato civile “coniugato” e la contemporanea assenza del coniuge in famiglia. Dai *debriefing* con i rilevatori, però, è emerso che nella maggior parte dei casi ciò era da imputare alla presenza di coppie a distanza piuttosto che a separazioni di fatto. Per tale ragione nella stesura definitiva tale regola è stata gestita con la tecnica del *probing*, ponendo le seguenti domande:

SG.32 In famiglia è presente il coniuge di “NOME”?

- Non presente in famiglia 0|_
- “NOME 1° componente”, “ETÀ” anni 1|_
- “NOME 2° componente”, “ETÀ” anni 2|_
- “NOME 3° componente”, “ETÀ” anni 3|_
- “NOME 4° componente”, “ETÀ” anni 4|_
- “NOME 5° componente”, “ETÀ” anni 5|_

Se non è presente il coniuge, ovvero non avviene alcuna correzione, allora:

SG.32bis Mi può dire per quale motivo il coniuge di “NOME” non è presente in famiglia?

- Separazione di fatto 1|_
- Coppia a distanza 2|_

Per quanto riguarda le regole *trasversali*, il loro principale obiettivo è di identificare nella vita dell’intervistato una sequenza di eventi coerente. Particolare attenzione è stata quindi posta nell’individuazione del percorso personale e lavorativo dell’intervistato. Ad esempio, si può supporre di scandire la vita personale e lavorativa di un individuo nella sequenza logica dei seguenti fatti:



La mancata successione di questi semplici eventi costituisce un campanello di allarme per attivare controlli e verifiche. Per individuare i percorsi dell’intervistato sono state inserite nel questionario numerose date (o l’età corrispondente): la data di nascita dell’intervistato, la data di conseguimento del titolo di studio, la data di matrimonio, la data di inizio dell’attuale lavoro (per gli occupati), la data di inizio del primo lavoro, la data di fine dell’attività lavorativa, la data dell’ultima ricerca di lavoro, la data di inizio del corso di formazione seguito, la data di fine del corso di formazione seguito, la data di iscrizione al centro per l’impiego, la data di ultimo contatto con il centro di lavoro.

Molte delle regole trasversali servono a controllare la coerenza tra le differenti date dichiarate.

Altre regole individuano l’incoerenza di una risposta confrontandola con una data precedentemente indicata. Ad esempio, il conseguimento di un determinato titolo di studio è legato all’età dell’intervistato.

In fase di test le regole trasversali di coerenza sono state impostate prevalentemente come *Soft*. Tali regole si sono attivate nel 6,4 per cento degli individui intervistati, nel 93,3 per cento una sola volta, nel 6,1 per cento due volte e nello 0,5 per cento tre volte.

Le analisi effettuate sulle attivazioni delle regole e sulle rispettive conciliazioni hanno permesso di affinare considerevolmente l'insieme dei controlli in linea, correggendo eventuali errori nelle regole stesse, migliorando i messaggi di errore, cambiando le formulazioni dei quesiti, e, soprattutto, definendo quali regole lasciare *Soft* e quali trasformare in *Hard*.

Ad esempio la regola di consistenza tra la domanda *B1* (in cui si chiede all'intervistato se ha lavorato nella settimana di riferimento almeno un'ora) e la domanda *C37* (numero di ore lavorate nella settimana di riferimento), che si attiva se $B1=SI$ e $C37=0$, è rimasta *Soft*, ma si è aggiunta una regola *Hard* alla *D1* (ha un secondo lavoro). Infatti, se l'intervistato dichiara di aver lavorato la settimana di riferimento deve aver lavorato almeno 1 ora nel primo o nel secondo lavoro. Tale regola, infatti, si è attivata in fase di test nello 0,2 per cento dei casi di individui che hanno più di un lavoro e che dunque si riferiscono a lavori diversi nelle due differenti sezioni del questionario, ed è stata conciliata nella quasi totalità dei casi.

Al contrario, la regola tra la domanda *C27* (lavoro a tempo pieno o lavoro a tempo parziale) e la *C31* (orario di lavoro abituale), che si attiva se $C27=lavoro\ a\ tempo\ parziale$ e $C31>32$ ore (soglia superiore prevista per la definizione di contratto part-time), è stata lasciata *Soft*. Tale scelta è scaturita dall'analisi delle attivazioni che, avvenute nello 0,6 per cento degli intervistati (pari al 7,7 per cento degli abilitati a rispondere), non sono state conciliate nel 27 per cento dei casi. Dai *debriefing* è emerso che la non conciliazione è avvenuta per motivi plausibili, anche se non frequenti, di part-time verticale e contratti privatistici di lavoro che possono derogare al limite superiore di 32 ore settimanali.

Complessivamente le regole di consistenza *Hard* attualmente presenti nel questionario sono 40.

6.2.8 - I controlli longitudinali

I controlli longitudinali sono stati inseriti per correggere le variabili in funzione dei valori che le stesse hanno assunto nelle precedenti occasioni.

Come anticipato, è stato implementato un unico questionario valido per tutte le quattro *waves*. Tale scelta ha consentito di ottenere un identificatore personale (ID) univoco, risolvendo tutti i problemi di abbinamento delle informazioni in *waves* differenti, con la possibilità di mantenere a video le informazioni rilevate in tempi precedenti e di porre perciò le domande a conferma. Il primo controllo attivato è quello sulla composizione familiare nelle differenti *waves*. Se non sono intervenuti cambiamenti si chiedono una serie di informazioni a conferma. Ad esempio, la domanda sul *titolo di studio* è posta a conferma ed in caso di variazione è inserito il vincolo di una "non decrescita". Ciò permette di verificare la coerenza nel tempo delle risposte fornite tra una *wave* e l'altra.

Per le domande relative all'occupazione, si è tenuto conto non solo dell'esigenza che i percorsi delineati debbano necessariamente seguire delle strutture logiche, evidenziate dai controlli di coerenza e verosimiglianza, ma anche del fatto che alcuni "stati", risultanti dalle classificazioni ufficiali sull'occupazione, sono irripetibili e si verificano una sola volta durante l'*iter* lavorativo individuale. Per esempio, la condizione di essere "in cerca di prima occupazione" è irripetibile, per cui alla domanda *I1* del questionario la condizione "in cerca di prima occupazione" è considerata incoerente rispetto alla modalità "occupato" e "disoccupato in cerca di una nuova occupazione" dichiarata nelle occasioni precedenti.

6.2.9 - Le tecniche di correzione, messaggi di errore e domande a conferma

La procedura di correzione *a caldo* prevalentemente utilizzata per decidere l'ammissibilità di una risposta è stata la conciliazione. La tecnica del *probing*, se pur utilizzata in qualche caso

per la correzione, è servita maggiormente a contenere le mancate risposte. Ad esempio il *probing* è stato utilizzato per tutte le domande che si riferiscono ad una data. Nei casi in cui l'intervistato dichiara di non ricordare l'anno di accadimento di un particolare evento (ad esempio: anno di conseguimento del titolo di studio, anno di inizio del primo lavoro, eccetera) si passa ad una successiva domanda di rilancio nella quale si chiede quanti anni aveva. Tale strategia ha ridotto in modo significativo le mancate risposte parziali, in particolar modo per gli intervistati anziani e nei casi di intervista *proxy*.

I messaggi di errore sono stati progettati con attenzione per infastidire il meno possibile l'intervistato. Ad esempio, se l'intervistato ha dichiarato alla domanda C1 di svolgere un lavoro come imprenditore e alla domanda C4 sul numero di dipendenti afferma di non averne, si attiva il seguente messaggio: "Il nostro codice civile definisce imprenditore colui che organizza il lavoro di altre persone e ha almeno un dipendente". L'accento nel messaggio di errore è posto sul numero di dipendenti e non sull'attività svolta per non sminuire la dichiarazione dell'intervistato. Se viene confermata l'assenza di dipendenti, il sistema torna alla domanda sulla posizione nella professione per correggere l'informazione.

Infine, sono state inserite all'interno del questionario alcune domande di riepilogo per consentire il controllo e l'eventuale modifica veloce delle informazioni fornite.

Nella scheda generale, ad esempio, dopo aver rilevato tutte le informazioni sul singolo, si attiva una domanda di controllo che riporta il riepilogo delle informazioni registrate:

SG.29	Quindi, riepilogando, "NOME" ha "ETA" anni è "SESSO" ed è "STATO CIVILE" "ANNO MATRIMONIO" "RELPAR" di "NOME PR" "TITOLO DI STUDIO" "ANNO DI CONSEGUIMENTO". È corretto?
-------	--

Se qualche informazione non viene confermata, il sistema consente di cambiarla, attivando però eventuali nuovi controlli che da tale variazione potrebbero scaturire.

6.2.10 Confronto tra l'indagine trimestrale forze di lavoro e la Nuova indagine continua sulle Forze di lavoro

Per valutare il miglioramento nella qualità dei dati raccolti con il questionario elettronico si è effettuato un confronto tra il piano di controllo *off-line* della Rilevazione Trimestrale sulle Forze di Lavoro (Rtfl) e il piano di controllo *on-line* illustrato nei precedenti paragrafi.

Nella Rtfl il controllo sui dati raccolti tramite questionario cartaceo avveniva a partire dalla fase di registrazione dei dati, evidenziando i possibili errori di *range* e di percorso. I controlli relativi alla coerenza dei dati e all'individuazione/correzione delle mancate risposte parziali venivano effettuati successivamente attraverso l'interazione di diversi programmi che implementavano due differenti metodologie di correzione: a) deterministica; b) probabilistica.

I confronti tra le due indagini possono essere effettuati solo distinguendo tra i controlli di coerenza trasversali e quelli longitudinali.

Per i controlli di coerenza trasversali, la procedura di imputazione usata per la correzione del file dell'indagine trimestrale delle Forze di Lavoro consente di intervenire in particolare su:

- l'omessa compilazione di alcuni quesiti;
- le risposte ridondanti, ovvero la compilazione non dovuta di alcuni quesiti;
- i valori fuori *range*;
- l'incompatibilità tra le risposte fornite a diversi quesiti.

Nel 2003 i record con almeno una correzione sono stati circa il 44 per cento (ovviamente con una variabilità trimestrale), con un numero medio di correzioni per singolo record del 2,5 per cento.

Solo il 6,3 per cento dei record subisce una correzione a causa di incompatibilità tra successive risposte. Ne deriva immediatamente che, grazie alla gestione in automatico dei *range* e del percorso all'interno del questionario, vi è un notevole guadagno in termini di qualità dei dati (38 per cento di record corretti in meno).

Nelle due indagini la percentuale di correzioni a causa di risposte non compatibili risulta piuttosto simile (6,3 per cento per la Rtf1 e 7,0 per cento per la Rcfl). Tuttavia tale percentuale assume un significato notevolmente diverso. In primo luogo, le regole della "nuova" indagine sono più numerose anche per il maggior numero di domande. In secondo luogo, le correzioni hanno consentito di risolvere l'incongruenza in sede di intervista, conciliando i dati con l'intervistato, piuttosto che imputarli *a-posteriori*.

Per quanto riguarda i controlli longitudinali la differenza sostanziale tra le due indagini risiede nel codice identificativo del componente intervistato. Nella Rtf1 sorgevano problemi di abbinamento degli identificativi: confrontando le principali variabili anagrafiche, solo l'80 per cento dei record dopo sei mesi sono abbinati correttamente e tale percentuale si riduce al 75 per cento a distanza di 12 mesi.

Nella Rcfl tale problema è totalmente risolto grazie alla creazione di un identificativo unico. Ad esempio, se nella prima intervista la famiglia era composta da due componenti, il PR con codice identificativo "1" e il coniuge con codice identificativo "2", nella successiva intervista per i nuovi componenti il codice identificativo assumerà un valore progressivo a partire da "3". Il codice assegnato ai nuovi componenti risulterà sempre superiore a quello registrato per i componenti alla precedente intervista anche quando nella *wave* attuale alcuni di questi ultimi non fanno più parte della famiglia. Riprendendo il nostro esempio, il nuovo componente avrà il codice "3" anche se nella seconda intervista il coniuge di PR (codice identificativo "2") non è più presente in famiglia. In questo modo è possibile seguire la storia del nucleo familiare per tutte e quattro le *waves* di intervista, registrando gli eventuali cambiamenti e rendendo automatico l'accoppiamento degli individui presenti in più occasioni di indagine. Ciò avvalorava ulteriormente il guadagno in termini di qualità dei dati raccolti.

Capitolo 7 - Il trattamento dei dati e le procedure di stima

7.1 - Premessa

La rilevazione continua è stata progettata e realizzata con l'obiettivo di ottenere un elevato standard qualitativo delle informazioni raccolte, sin dal momento della rilevazione. Ciò incide in modo determinante sulle scelte metodologiche effettuate per la progettazione e realizzazione del disegno campionario, per la definizione delle strategie per il trattamento dei dati e per la definizione delle procedure di stima.

Le innovazioni che hanno portato alla realizzazione della RcfI hanno comportato mutamenti importanti sulle metodologie scelte e, soprattutto, sull'intera filosofia di lavoro. Il monitoraggio continuo e in "tempo reale" di ogni operazione garantisce la possibilità di migliorare il processo in corso d'opera: il tutto si concretizza in una maggiore qualità del dato raccolto.

In un contesto così importante, il presente capitolo pone l'accento su quelle fasi di processo che compongono il trattamento dei dati: nella fattispecie, oltre ad una descrizione dell'intero processo che porta dal dato rilevato alla produzione delle stime, saranno analizzate in modo approfondito le scelte metodologiche adottate per la fase di controllo e correzione dei dati e per la costruzione delle stime.

Come già ampiamente illustrato, la nuova rilevazione sulle forze di lavoro prende vita da quanto stabilito dal regolamento 577/98 del Consiglio dell'Unione europea. Nell'esigenza di armonizzare e meglio cogliere le caratteristiche del mercato del lavoro è stato introdotto un elemento che, da solo, ha cambiato radicalmente l'approccio a tutto il processo rispetto al passato ovvero la continuità dell'indagine.

Le nuove tecniche introdotte sin dalla fase di estrazione delle liste delle famiglie campione, e in particolare, la scelta di condurre le interviste con tecnica assistita da computer hanno comportato sensibili miglioramenti nella qualità dei dati raccolti che nell'ottica del trattamento dei dati si traduce in una drastica riduzione delle correzioni a posteriori¹.

Le innovazioni introdotte in termini di tecnica di rilevazione e l'importante contenuto informatico a supporto hanno indotto a progettare, e realizzare, una strategia di controllo e correzione dati diversa rispetto a quella usata per le indagini tradizionali. In particolare, il dato "grezzo" ha un livello qualitativo estremamente più elevato rispetto alle indagini eseguite con tecniche non assistite da computer. Il dato grezzo presenta, infatti, un numero estremamente ridotto di incongruenze tra le informazioni raccolte, in quanto queste ultime vengono sottoposte ad un piano di verifica implementato nel questionario elettronico che agisce sin dal momento della rilevazione. Tale piano si basa sul controllo dei domini delle variabili, dei percorsi del questionario e delle incongruenze logiche tra le informazioni raccolte.

Il buon livello qualitativo che i dati grezzi raggiungono ha fatto sì che l'impatto dei piani di check in termini di correzione fosse piuttosto trascurabile, cosicché il check è stato immaginato uno strumento di monitoraggio del questionario e della fase di rilevazione piuttosto che uno strumento di correzione a posteriori.

In tale contesto, gli strumenti che potevano essere utilizzati potevano essere molteplici in quanto la fase di controllo era gestibile in qualsiasi linguaggio di programmazione. Si è scelto di

Di Claudio Ceccarelli, con la partecipazione di Simona Rosati (parr. 7.2.1, 7.2.2, 7.2.3), Davide Di Laurea (par. 7.2.4), Raffaella Cascioli (par. 7.2.5), Antonio Rinaldo Discenza (par. 7.3), Silvia Loriga (par. 7.4)

¹ A tal proposito, nella Rilevazione trimestrale sulle forze di lavoro, circa l'80 per cento delle correzioni a posteriori era dovuto a errori nei percorsi, con le tecniche *Computer Assisted* tale tipologia di errore viene annullata.

implementare i controlli principalmente in Scia² in modo tale da avere a disposizione anche il modulo di correzione nei casi in cui emerga la necessità di correggere a posteriori le eventuali incongruenze riscontrate.

La strategia adottata tratta essenzialmente due tipologie di errori: gli errori strutturali e quelli logico-formali; i primi derivano dai percorsi strutturali del questionario mentre i secondi dalle incongruenze logico formali delle informazioni raccolte³. Nel questionario elettronico, ovviamente, sono implementate tutte le regole (denominate *hard*) che determinano i percorsi e il flusso del questionario stesso e gran parte di quelle logico-formali (denominate *soft*)⁴. La scelta di non introdurre tutte le possibili regole *soft* deriva dalla volontà di non appesantire il carico informatico del questionario elettronico e l'impatto negativo che questo può avere nella conduzione dell'intervista⁵. In assenza di errori strutturali del questionario, quindi, le sole tipologie di correzione che vengono effettuate riguardano le incongruenze logico-formali di derivanti da forzature di regole *soft* del questionario o di incongruenze non rilevate.

Altro importante elemento di innovazione riguarda la matrice di compatibilità delle informazioni su attività economica e professione degli individui classificati come occupati.

In estrema sintesi, è stata messa a punto una matrice per classificare ogni combinazione delle informazioni rilevate per professione e attività economica come congruente, incongruente o indeterminata per livello di dettaglio della codifica. La matrice è organizzata in modo da consentire il controllo delle congruenze tra tutti i livelli (a due, a tre e a quattro digit) delle informazioni rilevate per la professione e per l'attività economica. L'area dell'indeterminatezza tra le due variabili è oggetto di verifica e analisi, con l'obiettivo di procedere nel corso del tempo ad una eventuale revisione della matrice.

Le informazioni contenute nella matrice hanno rappresentato la base per la costituzione di un set di regole *soft* inserite nel questionario elettronico. Nel caso di una incongruenza segnalata durante l'intervista, il rilevatore chiede maggiori dettagli nella descrizione della professione e attività economica in modo da poter codificare, in tempo reale o alla fine dell'intervista, in modo adeguato.

Altro elemento di innovazione è rappresentato dalla "procedura famiglie". Tale procedura permette, dopo aver controllato e corretto le notizie anagrafiche dei componenti, di costruire i nuclei presenti nella famiglia di fatto intervistata, e classificarla secondo un set di tipologie familiari. In estrema sintesi, la correzione avviene dopo aver diviso le famiglie monocomponente da quelle pluricomponente. Per le prime si riscontra solo la congruenza tra le informazioni rilevate sulla singola persona. Per le seconde, sono individuate le coppie possibili all'interno di ogni famiglia e sono corrette le caratteristiche dei componenti (relazione di parentela con la persona di riferimento, stato civile e anno di matrimonio) in base alle possibili relazioni con tutte le coppie individuate, tenendo conto delle forme familiari più frequenti.

Infine le procedure di stima, come ogni altro aspetto legato all'armonizzazione, hanno in primo luogo l'obiettivo di garantire il rispetto dei vincoli comunitari.

² Com'è noto, SCIA (Sistema per il Controllo e l'Imputazione Automatica) richiama la metodologia illustrata da Fellegi e Holt in "A Systematic Approach to Automatic Edit and Imputation", Journal of the American Statistical Association, n.71, 1976

³ In Scia, il controllo degli errori strutturali avviene mediante la creazione di *edit* strutturali mentre il controllo degli errori logico-formali danno luogo agli edit logico-formali.

⁴ A differenza della regola *hard*, quella *soft* può essere disattivata dal rilevatore in maniera forzata qualora non riesca a rilevare informazioni coerenti.

⁵ Questo accade, ad esempio, quando si vuole verificare la congruenza di informazioni rilevate nella fase iniziale e in quella finale del questionario. In tali casi, infatti, sarebbe molto dispendioso e aumenterebbe la molestia statistica sulla persona intervistata dover chiedere di nuovo una serie di informazioni solo perché è stata rilevata un'incongruenza con la parte iniziale del questionario. La gestione di casi del genere lasciata al rilevatore porterebbe ad una "forzatura" quasi certa della regola, con effetti ben più gravi dal punto di vista della qualità del dato.

Come la precedente indagine, il disegno di campionamento della Rcfl è in grado di garantire stime a livello regionale con cadenza trimestrale, come richiesto da regolamento comunitario, e sono rivolte a soddisfare altre esigenze di tipo informativo, in particolare, quelle di fornire stime provinciali in media d'anno e indicatori a livello familiare.

Il raggiungimento simultaneo di tali obiettivi è possibile attraverso l'utilizzo della metodologia basata sugli stimatori calibrati, incorporata in un software generalizzato progettato e realizzato dall'Istituto⁶. L'aggancio di tali stimatori ad opportuni totali noti consente di produrre stime regionali a livello trimestrale, provinciali in media d'anno e un unico coefficiente di riporto all'universo per famiglie e per individui.

La realizzazione del coefficiente finale di riporto all'universo si articola in tre fasi distinte. La prima fase prevede il calcolo del coefficiente base ottenuto come reciproco della probabilità che ogni famiglia ha di far parte del campione⁷. Successivamente, per tenere conto delle interviste non realizzate in ciascun Comune, si determina un elemento correttivo del coefficiente base detto coefficiente per mancata risposta totale⁸. L'ultima fase riguarda il vincolo delle stime ottenute con il coefficiente base corretto per mancata risposta a "totali noti" desunti da fonti anagrafiche⁹.

L'atto conclusivo della procedura di stima è rappresentato dal calcolo della stima degli errori di campionamento. Come è noto, l'errore campionario rappresenta una misura del grado di attendibilità delle stime e consente all'utente finale di valutarne la bontà. Con l'ausilio delle metodologie implementate nello stesso software che calcola gli stimatori calibrati, si determina l'errore campionario a differenti livelli territoriali delle stime dei principali aggregati prodotti dall'indagine. Successivamente, si determina un modello statistico che, basato sul legame esistente tra l'entità della stima l'errore campionario che ne deriva, offre la possibilità di calcolare l'errore campionario associato a qualsiasi stima prodotta dall'indagine a differenti livelli territoriali. La stima dell'errore campionario è realizzata per ogni trimestre e in media d'anno.

7.2 - Il piano di controllo e correzione dei dati

7.2.1 - L'individuazione degli errori

Al termine del processo di rilevazione l'individuazione degli errori avviene mediante l'attivazione di un insieme di regole di incompatibilità (*edit*), cioè un insieme di asserzioni sulla non ammissibilità di codici (modalità) per la singola variabile o di combinazioni di codici relativi a più variabili. Per errori si intendono le incoerenze logiche tra le informazioni rilevate, le mancate risposte ad uno o più quesiti ed eventuali valori non ammissibili delle singole variabili. Occorre precisare che le risposte mancanti ad uno o più quesiti, pur non essendo

⁶ Il software, denominato Genesees, richiama la metodologia illustrata da Deville, J.C. e Särndal, C.E. in *Calibration Estimation in Survey Sampling*, Journal of the American Statistical Association, Vol. 87, n.418, 1992.

⁷ Per la determinazione del coefficiente base, si tiene conto della probabilità di estrarre un generico Comune dallo strato di appartenenza, e della probabilità di estrarre le famiglie campione dalla lista anagrafica di tale Comune. Ogni famiglia del campione, appartenente dello stesso Comune, ha pertanto lo stesso coefficiente base.

⁸ In genere, il numero di famiglie intervistate per ciascun Comune del campione è inferiore al numero teorico previsto dal disegno campionario. Al fine di calcolare l'effettiva proporzione delle correzioni per mancata risposta totale. Attraverso un processo iterativo di ottimizzazione matematica che tiene conto dei totali noti, si determina infine il coefficiente finale di riporto all'universo.

⁹ La RCFL utilizza come totali noti: la popolazione residente per regione, sesso e quattordici classi d'età (0-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75 e più), e la popolazione residente per provincia, sesso e tre classi d'età (0-14, 15-64, 65 e più).

ammesse dal questionario elettronico, talvolta possono presentarsi a causa di problemi tecnici incorsi durante l'intervista e in ogni caso non controllabili a priori. Analogamente, eventuali errori di percorso del questionario possono essere attribuiti ad eventuali mal funzionamenti di componenti *software* o *hardware* dello strumento di rilevazione.

Le regole definite dagli esperti dell'indagine costituiscono l'insieme delle *regole esplicite*. L'attivazione di una o più regole di incompatibilità in corrispondenza di un dato *record* indica che il *record* in esame è errato rispetto all'insieme di regole definite. Al contrario, un *record* che non provochi l'attivazione di nessuna regola risulta esatto rispetto all'insieme delle regole e quindi non necessita di essere corretto.

Le regole esplicite relative a variabili qualitative (cioè non soggette a metriche specifiche) sono state implementate in Scia (Sistema per il Controllo e l'Imputazione Automatica), un sistema di correzione dei dati realizzato dall'Istat (Barcaroli e Venturi, 1997), che si basa interamente sulla metodologia di Fellegi e Holt (1976). Più precisamente, è stata utilizzata la versione di SCIA contenuta in CONCORD (Controllo e CORrezioni Dati), un sistema progettato per ambienti operativi Windows (Riccini, Margarucci e Floris, 2000).

Le rimanenti regole, relative a variabili quantitative, sono state tradotte in regole del tipo Se-Allora all'interno di un programma SAS (SAS Institute Inc., 1999).

L'esigenza di utilizzare due differenti *software* per implementare le regole è dovuta al fatto che, in generale, non si dispone di *software* generalizzati che trattino simultaneamente variabili qualitative e quantitative. Infatti, pur prevedendo il trattamento delle variabili quantitative, la metodologia di Fellegi and Holt è stata sviluppata essenzialmente per quelle qualitative. D'altra parte, il controllo automatico dei dati, effettuato mediante il metodo di Fellegi e Holt, è particolarmente adatto per individuare i *record* errati in modo tempestivo e senza ulteriori impieghi di risorse umane.

Con riferimento a una singola variabile, distinguiamo due tipi di errore: l'errore *casuale* e quello *sistematico*. Gli errori casuali possono essere assimilati ai normali errori di misurazione, hanno la stessa probabilità di verificarsi in variabili differenti e non sono correlati agli errori di altre variabili. Gli errori sistematici difettano di una definizione propria, ma possono essere definiti per negazione come errori non probabilistici, dovuti generalmente alle imperfezioni nella struttura di rilevazione (per esempio, l'errore causato da un quesito erroneamente specificato). Tale distinzione si riflette, come vedremo in seguito, sul tipo di correzione. Infatti, la correzione deterministica è più adatta a trattare errori di tipo sistematico, mentre quella probabilistica risulta più efficiente per quanto riguarda gli errori casuali.

Nel Prospetto 7.1 si riporta il numero di regole esplicite utilizzate nella Rcf. Con riferimento al sistema Scia, osserviamo che l'intero insieme delle regole è stato suddiviso in tre sottoinsiemi a causa dell'ampia numerosità delle regole. In altri termini, la complessità computazionale del metodo di Fellegi e Holt è tale che per indagini su larga scala, con centinaia di *edit* espliciti e centinaia di variabili, spesso è necessario dividere l'insieme delle regole in due o più parti indipendenti (due insiemi di *edit* si dicono totalmente indipendenti se non condividono variabili). Tuttavia, essendo molto raro il caso di totale indipendenza tra due o più insiemi di *edit*, è opportuno minimizzare il numero di variabili in comune.

Il prospetto 7.2 riporta la percentuale di *record* errati secondo il tipo di errore riscontrato al termine della fase di controllo dei dati. Ricordiamo che un *record* è considerato errato se provoca l'attivazione di almeno una regola, mentre l'attribuzione della natura dell'errore dipende sostanzialmente sia dal tipo di regola, o meglio dal tipo di incompatibilità che esprime la regola, sia dal numero di attivazioni della regola stessa.

Prospetto 7. 1 - Numero di *edit* espliciti

Edit	Numero di edit
SCIA	2.483
- <i>Insieme A</i>	961
- <i>Insieme B</i>	1.397
- <i>Insieme C</i>	125
SAS	142

Prospetto 7.2 - Record errati per tipo di errore

Errore	Record (%)
Sistematico	2.5
Casuale	2.8

7.2.2 - L'imputazione

L'imputazione dipende essenzialmente dalla natura dell'errore, sebbene alcune considerazioni di carattere pratico non dovrebbero essere trascurate durante l'implementazione di una procedura di imputazione.

Nel caso di dati *Missing Completely at Random* – *MAR* (Rubin, 1976; Little, 1982) si può dimostrare che i metodi deterministici introducono elementi di distorsione nelle distribuzioni delle variabili e tendono ad attenuare la varianza delle stesse, mentre i metodi probabilistici non alterano la proprietà di correttezza dei parametri di una distribuzione (Kalton and Kasprzyk, 1982). Tuttavia, si può ragionevolmente assumere che i metodi deterministici siano più adatti per correggere gli errori di tipo sistematico, mentre i metodi probabilistici siano maggiormente indicati per gli errori provenienti da un modello di generazione casuale.

L'imputazione deterministica consiste nell'assegnare un solo valore predeterminato e potrà riguardare o meno le variabili coinvolte nelle regole di incompatibilità; può dipendere, inoltre, dai valori assunti da altre variabili rilevate. Al contrario, il metodo probabilistico non prevede la possibilità (o la necessità) di definire a priori, per ogni situazione di errore, l'elenco delle azioni da intraprendere per eliminare gli errori, ma assegna un valore in base a un modello stocastico (per esempio, il modello di regressione) oppure utilizzando i valori "esatti" di un'unità ritenuta simile a quella in cui è stata riscontrata l'incompatibilità (*metodo del donatore*). Generalmente la similarità tra le due unità è definita in base ad una prefissata funzione di distanza.

I metodi di imputazione proposti da Fellegi e Holt sono di tipo *hot-deck*, cioè i valori delle variabili errate vengono sostituiti con i valori "esatti" che le stesse variabili assumono in una o più unità del campione nelle quali non è stato riscontrato alcun errore (*record* esatti).

In particolare le strategie di imputazione proposte sono:

- *imputazione sequenziale*, che consiste nel correggere separatamente le variabili errate, individuando se necessario tanti *record* donatori quante sono le variabili errate;
- *imputazione congiunta*, in cui le variabili errate vengono corrette simultaneamente attribuendo loro i valori che le stesse variabili assumono nel *record* donatore.

Il primo metodo presenta un inconveniente dovuto proprio all'imputazione sequenziale delle variabili errate, che assicura il solo mantenimento delle distribuzioni marginali delle stesse nell'insieme dei dati esatti (eccetto il caso di indipendenza probabilistica, in cui verrebbero preservate anche le distribuzioni congiunte). Al contrario, il metodo di imputazione congiunta, apportando la correzione congiuntamente su tutte le variabili errate, preserva le distribuzioni marginali e congiunte delle stesse nell'insieme dei dati esatti.

Per entrambi i metodi, qualora non fosse possibile individuare un donatore, si può ricorrere all'imputazione "forzata" di ciascuna variabile. Tale metodo consiste nell'attribuire, alla variabile errata, un qualunque valore estratto casualmente dalla distribuzione marginale semplice della variabile stessa nell'insieme dei dati esatti.

Al fine di individuare per ciascun *record* errato quali variabili modificare per riportare il *record* stesso a una situazione di correttezza, il metodo di Fellegi e Holt prevede che vengano considerati anche gli *edit* impliciti, cioè quelli logicamente derivati dagli *edit* espliciti. Gli *edit* espliciti e quelli impliciti costituiscono l'*insieme completo* degli *edit*, indispensabile per garantire la correttezza finale di un *record*. Infatti, pur essendo sufficienti ad individuare la presenza di errori all'interno dei *record*, le regole esplicite non sono sufficienti a determinare quali variabili correggere per riportare il *record* ad una situazione globale di correttezza, e tanto meno il minor numero di variabili da modificare (*principio del minimo cambiamento*). Infatti, se le variabili errate venissero selezionate considerando solo le regole esplicite attivate, potrebbe avvenire che, una volta modificato il valore di tali variabili, si attivino altre regole esplicite, mantenendo quindi una situazione di errore nel *record*.

Nella RcfI la procedura di correzione è costituita dall'uso combinato del metodo selettivo e di quello automatico. Riteniamo infatti, in accordo anche con altri autori (Barcaroli *et al.*, 1998), che questa strategia di imputazione sia la più adatta in quanto consente una maggiore efficienza del processo di correzione e un migliore impiego delle risorse disponibili. Il metodo selettivo consiste nell'individuare tutte quelle unità che contengono errori potenzialmente influenti sui risultati finali, mentre il metodo automatico è nella fattispecie il metodo di Fellegi e Holt. Nel caso in esame le unità influenti vengono individuate sulla base dei risultati forniti dalla fase di controllo dei dati mediante il metodo di Fellegi e Holt.

In pratica, dopo aver individuato i *record* errati, come descritto nel paragrafo precedente, questi vengono suddivisi in *record* "critici" e *record* "non critici" (Ceccarelli e Rosati, 2005). I primi sono considerati tali perché possono influire significativamente sul valore finale delle stime, mentre quelli non critici hanno un impatto trascurabile sulle stime. I *record* critici che risultano errati a causa di errori sistematici vengono corretti mediante un algoritmo deterministico di imputazione, mentre i *record* errati a causa di errori casuali vengono corretti in base al metodo probabilistico di Fellegi e Holt.

I rimanenti *record*, cioè quelli non critici, anch'essi vengono corretti applicando il metodo probabilistico di Fellegi e Holt. In questo caso, l'uso del metodo probabilistico è giustificato dall'aver riscontrato sia un esiguo numero di attivazioni delle regole sia una vasta molteplicità di incompatibilità. In altri termini, se un'incompatibilità di tipo sistematico si verificasse per un numero ristretto di *record* (supponiamo non più di 10 *record*) e la correzione venisse effettuata in modo probabilistico, ciò non arreherebbe significativi cambiamenti al valore delle stime prodotte. D'altra parte, l'utilizzo del metodo deterministico per gli errori sistematici riscontrati in numero esiguo rappresenterebbe un uso spropositato della procedura di correzione, non giustificato da un significativo miglioramento della qualità dei dati.

In complesso l'impatto dell'intero processo di imputazione sui dati della RcfI è decisamente contenuto (prospetto 7.3). Infatti, circa il 95 per cento dei *record* non subisce alcuna correzione, mentre quasi il rimanente 5 per cento di *record* risulta aver subito al più tre correzioni.

Infine, nel prospetto 7.4 sono indicati i valori percentuali di *record* errati per tipo di imputazione nell'ambito della metodologia di Fellegi e Holt.

Prospetto 7.3 - Record per numero di imputazioni

Numero di imputazioni	Record (%)
0	95.5
1	3.5
2	0.7
3	0.3
4 or more	0.0

Prospetto 7.4 - Record errati per tipo di imputazione

Imputazione	Record (%)
Congiunta	41.1
Sequenziale	58.9

7.2.3 - Struttura e funzionalità del piano di controllo e correzione della Rcfl

Il piano di controllo e correzione della Rcfl agisce a livello di singola unità (*record*) per individuare e correggere le mancate risposte parziali. Come già accennato in precedenza queste comprendono le mancate risposte ad uno o più quesiti, i valori non ammissibili delle singole variabili e le incoerenze logiche tra le informazioni rilevate.

Il primo passo consiste nell'eliminare tutti quei *record* che, per qualsiasi motivo, presentino intere sezioni del questionario non compilate o per i quali non sia possibile ricavare alcune informazioni fondamentali, ma mancanti, sulla base di quelle disponibili (per esempio, un *record* per il quale la sezione B sia completamente vuota, viene corretto mediante un algoritmo deterministico utilizzando le informazioni disponibili della sezione C oppure della sezione E, cioè se $C1 \neq \text{blank}$ allora $B1 = \text{Sì}$, mentre se $E1 \neq \text{blank}$ allora $B1 = \text{No}$). In questo modo vengono eliminati i *record* che hanno un numero troppo elevato di informazioni mancanti. Ciò equivale ad assumere che l'individuo in esame non abbia portato a compimento l'intervista e di conseguenza, in accordo con le scelte effettuate in sede di rilevazione, che l'intera famiglia, a cui appartiene l'individuo stesso, anch'essa non abbia collaborato all'intervista. Per questo motivo oltre al *record* in esame vengono eliminati anche tutti i *record* individuali della famiglia corrispondente.

Successivamente viene avviata la fase di controllo dei dati, svolta tramite Scia per le variabili qualitative e mediante programma Sas per quelle quantitative, che consente di individuare i *record* critici e i tipi di errore ad essi associati. Se l'errore riscontrato è di tipo sistematico, la correzione dei *record* critici avviene mediante un algoritmo deterministico implementato in Sas, che prevede per ogni condizione di errore la corrispondente azione di correzione, mentre gli errori considerati casuali vengono corretti tramite SCIA. In entrambe i casi è necessario tener conto anche delle regole di percorso delle variabili coinvolte nella correzione al fine di garantire la correttezza finale dei *record*. Occorre sottolineare che la correzione probabilistica delle variabili quantitative, per quanto detto in precedenza, può essere effettuata in SCIA solo mediante un'opportuna trasformazione delle regole quantitative in corrispondenti regole qualitative, operazione che inevitabilmente comporta un rilevante aumento della complessità computazionale del metodo di Fellegi e Holt.

Al termine di questa fase alcune incoerenze di natura sistematica tra professione e attività economica (due informazioni particolarmente delicate sotto il profilo della codifica), individuate

per mezzo della matrice Ateco-Profes (Paragrafo 7.1.5), di cui si parlerà ampiamente nel seguito, vengono corrette mediante un algoritmo deterministico implementato in Sas.

Infine, i *record* non critici che presentano varie incompatibilità e tipologie di errore, ma in misura tale da non influire significativamente sul valore finale delle stime, vengono corretti con una procedura probabilistica iterativa. Tale procedura rappresenta un metodo semplice ed efficace per superare l'inconveniente dovuto all'impossibilità di generare l'insieme completo delle regole necessario per riportare i *record* errati a una situazione di correttezza. Inizialmente dall'insieme delle regole esplicite implementate in Scia vengono estratti gli *edit* che segnalano la presenza di incompatibilità nei *record* non critici. A partire dal sottoinsieme di *edit* così ottenuto il metodo di Fellegi e Holt è in grado di generare l'insieme completo degli *edit* e quindi di fornire una prima soluzione di correzione per i *record* in esame. Una volta corretti, tali *record* vengono di nuovo sottoposti a controllo completo in quanto la correzione effettuata, non tenendo conto dell'insieme totale delle regole esplicite, elimina sicuramente le incompatibilità riscontrate, ma può generarne delle altre. Se in questa fase non vengono segnalate nuove incompatibilità, allora si può ritenere concluso il processo di correzione, altrimenti è necessario avviare una seconda iterazione del processo in cui i nuovi *edit* vengono aggiunti a quelli precedentemente selezionati. In quest'ultimo caso si procederà con una nuova correzione e così via finché i *record* corretti non presentino più alcun errore.

Il *file* corretto così ottenuto sarà sottoposto al controllo della "procedura famiglie" (paragrafo 7.1.6) che, oltre a fornire indicazioni sulle varie tipologie di nuclei familiari desumibili dai dati, corregge anche eventuali errori riguardanti le variabili strutturali di ciascun nucleo.

7.2.4 - La matrice di incompatibilità tra professione e attività economica

7.2.4.1 - I presupposti per la realizzazione della matrice

Il passaggio dalla rilevazione trimestrale a quella continua ha permesso di apportare delle modifiche sostanziali alle modalità di codifica delle informazioni riguardanti la professione svolta ed il settore di attività economica di pertinenza. Innanzitutto è stato migliorato il dettaglio della codifica: il settore di attività economica viene ora codificato a 4 cifre anziché a 2 come nella rilevazione trimestrale¹⁰. Gli altri aspetti rilevanti sono costituiti dal questionario elettronico e dalla gestione diretta della rete di rilevatori, in quanto gli ausili forniti dal questionario elettronico (al cui interno è presente un motore di ricerca per le voci professionali e per le attività economiche, con la possibilità di codificare in corso di intervista o a posteriori) e la formazione specifica ricevuta dai rilevatori lasciavano infatti precludere ad un notevole miglioramento qualitativo dei dati raccolti.

In una rilevazione di tipo tradizionale condotta con metodo "Paper And Pencil Interview" (Papi) la sola possibilità di intervento è costituita dal piano di controllo e correzione. Un intervento, quindi, esclusivamente a posteriori, con il rischio potenziale di una correzione pesantemente invasiva. La combinazione delle innovazioni introdotte con la rilevazione continua ha consentito un ampliamento dell'insieme dei possibili interventi. Si è, difatti, aggiunta la possibilità di un monitoraggio *in itinere* che permette di intervenire sia direttamente in fase di rilevazione e registrazione del dato, sia in sede di correzione a posteriori. Difatti, la combinazione di queste innovazioni consente un monitoraggio capillare e una valutazione puntuale dell'attività di codifica. Questo processo costituisce un ausilio per il piano di controllo e correzione a posteriori; ma potrebbe avere effetti positivi sulla qualità anche in fase di raccolta e registrazione dei dati, tramite la predisposizione di un set di regole in linea che permettono la

¹⁰ Le professioni venivano già in precedenza codificate al massimo livello di dettaglio (4 cifre).

correzione di informazioni reciprocamente incompatibili già in sede di intervista e attraverso la possibilità di effettuare specifici interventi sui rilevatori.

Per rendere effettivamente praticabile un processo di monitoraggio così strutturato, si è resa necessaria la predisposizione di un ulteriore strumento che consenta una “misurazione” della affidabilità dei processi di codifica. A tal fine sono state realizzate le matrici¹¹ di compatibilità Ateco–Profes, della cui realizzazione e perfezionamento si da conto qui di seguito.

7.2.4.2 - *Le matrici: attribuzione dei giudizi di compatibilità*

Nella progettazione delle matrici la valutazione congiunta delle informazioni concernenti la professione degli occupati e il settore di attività economica presso cui si svolge il lavoro è stata assunta come punto di partenza. Per ogni coppia di codici ATECO e PROFES il lavoro è consistito nello stabilire se valutarla come una combinazione compatibile, poco plausibile o incompatibile. In altri termini, si è stabilita una convenzione per l’attribuzione dei giudizi di reciproca compatibilità.

Operativamente, si è scelto di assegnare i giudizi di plausibilità o compatibilità piuttosto che quelli di scarsa plausibilità o incompatibilità. Si è trattato, quindi, di enucleare il ventaglio di figure professionali necessarie a portare avanti una determinata attività economica e, viceversa, determinare l’insieme di attività che si avvalgono di una determinata figura professionale. I giudizi di scarsa plausibilità o di incompatibilità sono stati assegnati in maniera residuale. Ossia, una volta stabilite le combinazioni da considerare compatibili, il loro complemento rispetto all’insieme complessivo delle possibili coppie di codifiche è stato identificato come il sottoinsieme di combinazioni scarsamente plausibili.

Dato questo approccio, si possono commettere 2 tipi di errore:

- assegnare la piena compatibilità quando in realtà la combinazione è poco plausibile;
- assegnare un giudizio di scarsa plausibilità quando la coppia di codici è, in realtà, compatibile.

La procedura scelta - di assegnare i giudizi di compatibilità e derivare in maniera residuale quelli di scarsa plausibilità - mira a minimizzare gli errori del primo tipo.

L’attribuzione è stata effettuata a vari livelli di dettaglio. In particolare, per le professioni si è fatto ricorso alle categorie professionali (codice a 4 cifre), salvo i casi in cui le classi professionali (prime 3 cifre) fossero già sufficienti a far ravvisare la compatibilità con alcuni codici Ateco. Per quanto attiene alle attività economiche, si è fatto ricorso soprattutto alle classi di attività (codici a 4 cifre), salvo i casi in cui si potesse assegnare un giudizio di compatibilità già a livello di divisioni (prime 2 cifre) o di gruppi (prime 3 cifre).

Alcuni esempi possono aiutare a comprendere la logica operativa appena presentata:

- per gli “imprenditori e amministratori di aziende private nelle costruzioni (edilizia)” (codice CP2001: 1.2.1.3) è possibile identificare una compatibilità con la divisione delle costruzioni (codice Ateco: 45).
- per la classe professionale dei “membri di organismi di governo e di assemblee con potestà legislativa e regolamentare” (codice CP2001: 1.1.1) è possibile identificare una compatibilità con la classe Ateco “Attività generali di Amministrazione pubblica” (codice Ateco: 75.11).

¹¹ In questi paragrafi si parla, in generale, di matrici di compatibilità. Questo perché sono state predisposte due versioni della matrice in virtù della diversa combinazione di classificazioni usate dalla rilevazione continua. Difatti nel 2003, anno della messa a regime della RcfI, le classificazioni adottate erano rispettivamente la Ateco1991 e la CP2001. La Ateco 2002 è stata adottata subito dopo il suo rilascio, a partire dal 2004, e ciò ha comportato la costruzione della seconda matrice di compatibilità.

Considerato l'ordinamento lessicografico che informa entrambe le classificazioni, quando assegnati ad un livello di dettaglio di ordine inferiore (codici delle professioni a 3 cifre; codici delle attività economiche a 2 o 3 cifre) i giudizi di compatibilità vengono estesi ai codici di maggior livello di dettaglio. Ciò significa, rispetto al primo esempio, che la piena compatibilità CP2001: 1.2.1.3 vs Ateco: 45 viene applicata a tutte i gruppi e le classi comprese nella divisione delle costruzioni. Similmente nel secondo esempio, tutte le 4 categorie professionali comprese nella classe 1.1.1. sono considerate pienamente compatibili con Ateco: 75.11.

Una volta conclusa la prima versione della matrice di compatibilità, si è proceduto ad una fase di raffinamento, revisione, correzione ed integrazione delle relazioni di reciproca compatibilità già identificate. A tal fine, è risultato rilevante il controllo delle singole voci professionali sul versante delle professioni, mentre per le attività economiche sono state di fondamentale importanza le note esplicative della classificazione Ateco.

Il ricorso ad alcuni casi pratici può esemplificare il lavoro di questa fase di revisione:

- si prenda in considerazione la categoria degli “elettricisti nelle costruzioni civili ed assimilati” (codice CP2001: 6.1.3.7). Tra le voci professionali di pertinenza la classificazione riporta la voce relativa a “Elettricista di ripresa cine e tv”. Questo porta ad ampliare l'insieme dei codici Ateco rispetto ai quali considerare la categoria professionale pienamente compatibile, dovendo includere anche le “attività radiotelevisive” (codice Ateco: 92.20) (*revisione tramite il controllo delle voci professionali*);
- nella prima fase si era stabilita una relazione di compatibilità fra gli “istruttori di guida” (codice CP2001: 3.4.3.1) e le “autoscuole, scuole di pilotaggio e nautiche” (codice Ateco: 80.41). Difatti nelle voci professionali degli “istruttori di guida” vengono compresi gli istruttori di scuola di volo, gli istruttori di scuola guida e i piloti istruttori aeronautici. Dalle note esplicative del codice Ateco 80.41 si ricava che la relazione di compatibilità va estesa alle “attività delle scuole di pilotaggio abilitate al rilascio di brevetti di pilota di linee commerciali” (codice Ateco: 63.23), trattate separatamente dalle scuole di volo non abilitate al rilascio di tali brevetti (*revisione tramite il controllo delle note esplicative della classificazione Ateco*).

La considerazione del fenomeno dell'integrazione fra attività economiche ha condotto ad un'ulteriore revisione delle relazioni di compatibilità: alcune figure professionali possono essere impiegate in attività secondarie o ausiliare, non rilevabili dalla codifica dell'attività economica principale.

Ultimata questa fase, la versione definitiva della matrice di compatibilità (basata sulle classificazioni Ateco1991 e CP2001) è stata interfacciata con i dati Rcfl del 2003. Simultaneamente, si è provveduto a realizzare una versione della matrice basata sulla classificazioni Ateco2002 e CP2001.

7.2.4.3 - L'utilizzo delle matrici di compatibilità

Come accennato in precedenza, lo scopo delle matrici era quello di fornire uno strumento che permettesse il monitoraggio e la valutazione dell'attività di codifica. Questo, a sua volta, era funzionale alla progettazione di interventi sia in sede di rilevazione, sia durante il piano di controllo e correzione.

Tra gli interventi effettuati in sede di rilevazione è possibile annoverare i seguenti:

- elaborazione di un set regole di incompatibilità incluse nel software del questionario elettronico;
- affinamento delle voci comprese nei database dei motori di ricerca interni al questionario elettronico;

- note tecniche ad uso dei rilevatori.

Dall'analisi dei dati elaborati con le matrici è stato possibile identificare alcuni errori sistematici di codifica. A questi si è posto immediato rimedio tramite gli interventi appena citati, nonché in fase di correzione.

Ad esempio, nel corso del primo anno di sperimentazione della nuova indagine è stato rilevato un errore sistematico che riguardava la divisione della "Pubblica Amministrazione" (codice Ateco: 75). Si è notato che, nel corso del 2003, la RcfI soffriva di una sovrastima¹² degli occupati nella "Pubblica Amministrazione". Incrociando tali dati con le informazioni sulle professioni, si notava una incidenza elevata di categorie professionali (maestri di scuola elementare, docenti di scuola media inferiore o superiore, braccianti, imprenditori agricoli, solo a titolo di esempio) non riconducibili a questo settore di attività economica.

Alla risoluzione di questo errore sistematico si è provveduto con diversi interventi: innanzitutto, sono state rielaborate le voci del database che potevano indurre questo tipo di errore. I rilevatori sono stati poi istruiti tramite l'invio di una nota tecnica appositamente predisposta. Infine si è provveduto ad inserire questi casi fra le regole di incompatibilità in linea.

L'insieme di queste regole è stato completato includendo altri casi di incompatibilità, selezionati a partire dai risultati delle matrici sulla base della loro rilevanza, nel senso più avanti specificato.

Con riferimento al piano di controllo e correzioni occorrono alcune considerazioni preliminari.

Come visto in precedenza, la procedura di assegnazione delle relazioni di compatibilità o plausibilità punta a minimizzare l'errore del primo tipo, ossia quello derivante dal considerare reciprocamente compatibile una combinazione che in realtà è poco plausibile. La relazione complementare non è quindi necessariamente una relazione di incompatibilità, ma più precisamente una relazione di scarsa plausibilità.

Da ciò se ne deduce che l'insieme delle relazioni di compatibilità fra professioni e attività economiche può essere utilizzato nella sua interezza per la determinazione dei valori ammissibili quando si tratti di imputare una informazione mancante, in presenza dell'altra. Ma non vale il contrario: la relazione di scarsa plausibilità ha degli attributi che la rendono applicabile in misura più ampia rispetto ad una relazione di incompatibilità in senso stretto. In altri termini, le combinazioni incompatibili sono un sottoinsieme proprio di quelle scarsamente plausibili.

Esiste un secondo ordine di considerazioni che sconsiglia l'uso della matrice come strumento da applicare in maniera automatica per l'individuazione delle incompatibilità. Si riprenda il caso della classe professionale dei "membri di organismi di governo e di assemblee con potestà legislativa e regolamentare" (codice CP2001: 1.1.1) compatibile solo ed unicamente con la classe "Attività generali di Amministrazione pubblica" Ateco 75.11. Questo implica che è stato assegnato un giudizio di incompatibilità tra il codice CP2001 1.1.1 (e tutte le categorie professionali sottostanti) e tutti i codici Atecodiversi da 75.11. Evidentemente, però, questa relazione di incompatibilità non ha nessun potere informativo sulla "gravità" dell'incompatibilità occorsa né, di conseguenza, su come questa si ripercuote sulla qualità dei dati. Ossia, questa relazione non discrimina fra l'incompatibilità dovuta al fatto che un record con classe professionale 1.1.1 sia associato ad codice Ateco 75.12, piuttosto che 75.23 o 01.11. È invece evidente che la "gravità" relativa a questi tre casi di incompatibilità ha una diversa graduazione (crescente nell'esemplificazione in esame). Quindi, un uso automatico e non ragionato delle relazioni di incompatibilità che si possono derivare dalla matrice, anche prescindendo dalla

¹² Che si trattasse di una sovrastima lo si è dedotto dal confronto effettuato con altre fonti, prima fra tutte la stima degli occupati della Contabilità Nazionale. Questa fonte, difatti, pur essendo basata sulla distribuzione degli occupati per attività economica delle rilevazioni delle forze di lavoro, proprio sulla divisione relativa alla "Pubblica Amministrazione" provvede ad integrare i dati del servizio FOL con altre fonti di tipo amministrativo.

considerazione circa la non coincidenza della relazione di scarsa plausibilità con quella di incompatibilità, comporterebbe il rischio di un intervento di correzione dei dati pesantemente invasivo (in termini di numero di record interessati) senza peraltro garantire necessariamente un significativo miglioramento della qualità degli stessi.

Infine, c'è da osservare che in presenza di un giudizio di scarsa plausibilità, nulla si può arguire su quale delle due informazioni in gioco sia la più attendibile e quale meno.

Una volta svolte e tenute in debito conto tutte queste considerazioni, si è provveduto a selezionare le regole di incompatibilità da inserire nel piano di correzione. Il criterio di selezione adottato è stato quello della rilevanza dell'errore di codifica nei primi 5 trimestri di effettuazione della Rcfl. La rilevanza è stata valutata in funzione sia della frequenza con cui gli errori si riscontravano nei dati, sia della "gravità" della incompatibilità nel senso sopra descritto.

In particolare, si è provveduto ad una correzione di tipo deterministico sui casi di incompatibilità strettamente riconducibili all'errore sistematico di codifica, non senza avere opportunamente verificata l'affidabilità della codifica della professione, facendo ricorso al campo descrittivo compilato dai rilevatori.

Con particolare riguardo alla distribuzione dell'occupazione per settori di attività economica, ultimata la correzione sono state condotte approfondite analisi di confronto e validazione, che hanno offerto ottimi risultati, con i dati più recenti del Censimento industria e servizi e della contabilità Nazionale.

7.2.5 - La "procedura famiglie"

La possibilità di analizzare la situazione occupazionale nell'ambito della famiglia permette di cogliere in un più ampio contesto sociale e relazionale le modalità di partecipazione al mercato del lavoro, le problematiche connesse alla conciliazione tra impegni lavorativi e responsabilità familiari nonché le situazioni familiari connesse ad una maggiore o minore capacità di inserimento nel mondo del lavoro.

Le trasformazioni del mercato del lavoro e le sue caratteristiche vanno inserite in un ambito di mutamenti e peculiarità dei contesti familiari che permettono di accrescere fortemente le potenzialità di analisi del fenomeno occupazionale, anche e soprattutto in un periodo in cui la velocità di queste trasformazioni richiedono sicuramente strategie e adattamenti che non coinvolgono esclusivamente il singolo individuo ma anche i soggetti legati ad esso da vincoli familiari.

Il passaggio dalla rilevazione trimestrale a quella continua ha permesso di accrescere considerevolmente il patrimonio informativo anche relativamente alle relazioni familiari.

Al fine di ottenere un dato statistico utilizzabile non solo a livello individuale ma, come appena detto, anche a livello familiare, è stata rivista ed applicata nella Rcfl la "procedura famiglie" sviluppata ed applicata nell'ambito delle Indagini Sociali Multiscopo. Il risultato è stato quello di impiegare uno strumento ormai consolidato che, fornendo delle classificazioni analoghe a quelle utilizzate nelle diverse indagini sulle famiglie, permette di effettuare più facilmente analisi integrate delle informazioni raccolte dalla Rcfl con le altre indagini sociali realizzate dall'Istat.

7.2.5.1 - La logica della "procedura famiglie"

Come sottolineato in precedenza il prodotto finale della procedura famiglie è la classificazione di queste all'interno di differenti tipologie familiari. E' importante tuttavia

evidenziare che questo risultato si concretizza attraverso un passo preliminare e fondamentale della procedura che è quello di controllare ed eventualmente correggere la coerenza di alcune variabili strutturali per uno stesso individuo e tra individui di una stessa famiglia. Questa fase si articola nei seguenti passi:

- la procedura si basa su due variabili “pilastro”, Sesso ed Età, le quali vengono considerate corrette a monte della procedura stessa. Le altre variabili considerate dalla procedura sono lo stato civile, la relazione di parentela e l’eventuale anno di matrimonio;
- le famiglie vengono divise in due blocchi: le famiglie mono-componente e le famiglie pluri-componente. Per le famiglie pluricomponente il criterio seguito è quello di individuare le coppie possibili all’interno della famiglia e correggere i componenti in base alla coppia riscontrata. Le coppie possibili vengono costruite con i componenti di una famiglia che hanno le variabili sesso ed età compatibili. Ad ogni coppia viene poi assegnato un punteggio che si basa sul grado di compatibilità di tutte le variabili. Le coppie vengono suddivise, in base al punteggio ottenuto, in tre categorie: coppie coniugate, coppie conviventi, non coppie;
- l’assegnazione del punteggio per la determinazione delle coppie all’interno delle famiglie avviene secondo un criterio per il quale ciascuna delle cinque variabili considerate nella procedura ha un punteggio diverso a seconda della combinazione con la quale questa si presenta in una coppia di individui esaminati, per cui:

per rientrare tra le coppie (coniugate o conviventi) occorre avere necessariamente sesso diverso ed età >di 14 anni (>36 per le coppie di nonni). Le coppie hanno poi un punteggio diverso a seconda delle differenze di età più o meno marcate, delle combinazioni degli stati civili, delle date di matrimonio uguali, diverse o mancanti, e delle compatibilità tra le relazioni di parentela. I punteggi più alti, da 7 a 11 identificano le coppie di coniugi, il punteggio pari a 6 identifica le coppie di conviventi e infine i punteggi inferiori a 6 non identificano alcuna coppia;

alcune regole di correzione del punteggio parziale modificano successivamente i risultati ottenuti spostando a volte una coppia da uno stato di coniuge ad uno di convivente o di non coppia e viceversa;

i punteggi assegnati a ciascuna delle combinazioni delle 5 variabili vanno da un minimo ad un massimo che permette in qualche modo di dare un peso diverso alle variabili stesse. Per esempio, nel caso dello stato civile si passa da 0 a 3 punti ed il valore massimo si assume nel caso di entrambi i componenti coniugati; per la relazione di parentela il punteggio va invece da -2 a 2 con il valore massimo nel caso di due componenti con relazione di parentela compatibile con lo stato di coniugato. Anche senza entrare ancora di più nello specifico, l’attribuzione di questi punteggi fa già dedurre che le dichiarazioni sugli stati civili hanno un peso maggiore di quelle sulle relazioni di parentela. La qualità di risposta sulla relazione di parentela è in effetti probabilmente più scadente, anche considerando il numero di modalità di risposta;

- successivamente alla fase di costituzione delle coppie si effettuano i controlli di coerenza tra tutti gli individui all’interno della famiglia per vedere la correttezza delle relazioni. L’individuazione delle incoerenze avviene attraverso “incompatibilità” stabilite a priori;
- le correzioni di possibili incoerenze avverranno considerando sempre la veridicità della coppia individuata nel passo di assegnazione dei punteggi.

la correzione delle famiglie effettuata dalla procedura è principalmente deterministica e solo in parte probabilistica. L’attribuzione delle relazioni di parentela è vincolata alla compatibilità

complessiva con tutti i componenti della famiglia e tiene conto delle forme familiari più frequenti. La correzione probabilistica avviene solo per la variabile “stato civile” per la quale, a volte, è opportuno attribuire casualmente uno stato scegliendo un valore possibile secondo una legge di probabilità che può derivare dalla distribuzione osservata degli stati civili per sesso e classi di età.

7.2.5.2 - La revisione della procedura alla luce delle caratteristiche della RCFL

Il primo sviluppo della “procedura famiglie” fu effettuato nel corso dei primi anni '90 per coprire le esigenze di specifiche indagini sociali caratterizzate da determinate tecniche di rilevazione. È dunque intuibile che la procedura ha richiesto particolari adattamenti e/o affinamenti per l'attuale Rcfl.

Da questo punto di vista, la prima considerazione ha riguardato le variabili relative al sesso e all'età, sulle quali la procedura non effettua modifiche. Le prime indagini per le quali la procedura è stata sviluppata rilevavano più volte queste variabili in modo da avere una maggiore garanzia di correttezza del dato: il controllo e l'eventuale correzione avveniva comunque a monte dell'applicazione della procedura famiglie. L'indagine Rcfl rileva una sola volta la variabile sesso e l'età viene proposta a conferma dopo la dichiarazione della data di nascita. Al termine della raccolta dei dati anagrafici dei componenti si effettua comunque un riepilogo delle informazioni raccolte volto ad avere una ulteriore conferma, da parte del rispondente, di quanto registrato: ciò tutela sufficientemente sulla correttezza dell'informazione raccolta sul sesso e sull'età.

La seconda riflessione ha riguardato la fase di assegnazione dei punteggi, basilare per la definizione delle coppie, coniugate e conviventi, all'interno di una famiglia. Questo passo della procedura è fondamentale poiché, successivamente alla fase di costituzione delle coppie, si effettuano i controlli di coerenza tra tutti gli individui all'interno della famiglia per vedere la correttezza delle relazioni: le correzioni di possibili incoerenze avvengono considerando sempre la veridicità della coppia individuata. Considerando tutti i punteggi attribuibili a ciascuna delle 5 variabili anagrafiche si arriva ad una molteplicità di combinazioni a ciascuna delle quali corrisponde un punteggio complessivo che permette di caratterizzare la coppia. Nella Rcfl, a differenza di alcune delle indagini sociali per le quali la procedura si è sviluppata, alcune delle combinazioni dei punteggi non sono mai presenti. Ciò accade perché la rilevazione è effettuata con tecnica mista Capi/Cati e dunque l'acquisizione dei dati è di migliore qualità grazie al controllo che avviene in linea: l'analisi di questa fase della procedura ha quindi fatto riferimento ai soli casi che possono accadere nella Rcfl.

Nel corso dell'analisi preliminare della procedura è emerso, per alcune combinazioni, un criterio di attribuzione della tipologia di “coppia” non sempre chiaro nella fase di differenziazione tra coppia e non coppia e tra coppie coniugate e conviventi. In particolare, la distinzione tra coppia coniugata e coppia convivente viene effettuata, in alcuni specifici casi, in base alle sole differenze di età della coppia, o solo a seconda che le date di matrimonio siano diverse invece che mancanti. A volte si osserva una discriminazione tra coppia coniugata, convivente o non coppia in base a delle particolari combinazioni di relazioni di parentela che non appaiono invece, a tal fine, fortemente discriminanti.

È importante tuttavia segnalare che questi casi riguardano poche e particolari combinazioni di valori delle variabili coinvolte. Inoltre, l'impatto sui dati delle correzioni che seguono a queste particolari situazioni di attribuzione di punteggio, ha mostrato che sono rari i casi per i quali risultava meno convincente l'attribuzione dello stato di coppia o non coppia. In pratica, le segnalazioni di cui sopra non influiscono sulla complessiva efficacia della procedura. Si è pertanto deciso di lasciare invariata la parte della procedura che assegna alle coppie i punteggi suddetti, mantenendo sotto controllo i particolari esiti della procedura qui considerati.

Il terzo punto di approfondimento ha riguardato le regole di incompatibilità che la procedura utilizza per l'individuazione delle incoerenze tra le variabili familiari rilevate e le correzioni che la procedura attua in seguito all'individuazione di incompatibilità. A questo fine si è ritenuto utile effettuare, attraverso un test della procedura sui primi dati raccolti con la Rilevazione Continua, un'analisi dell'impatto che le regole stesse hanno sui dati ed un esame delle correzioni delle incoerenze apportate dalla procedura. Lo studio simultaneo dei due aspetti, incompatibilità e correzione, ha permesso di avere una visione più completa di come la procedura lavora in questa fase. L'esempio descritto di seguito indica uno dei principali risultati emersi da questa analisi.

Le regole di incompatibilità attivate il maggior numero di volte sono riferite ad una situazione in cui per un soggetto "coniugato" non risulta presente in famiglia un coniuge: la procedura non accetta un individuo "coniugato" senza la presenza in famiglia del coniuge. Queste situazioni possono essere determinate, in effetti, o da un'errata risposta per la valorizzazione della variabile "stato civile" (dovuta alla presenza di una 'separazione di fatto', che, non essendo ancora formalizzata, può indurre il rispondente a percepirsi ancora come coniugato) o anche da una situazione in cui coppie effettivamente coniugate vivono abitualmente in differenti comuni. Questa segnalazione di errore, molto frequente nei risultati analizzati, ha fatto emergere la necessità di rilevare anche le cosiddette 'coppie a distanza': per questi individui non si dovrebbe porre alcuna correzione e la procedura dovrebbe accettare la presenza di un "coniugato" senza la presenza in famiglia del coniuge. L'introduzione nel questionario, nel corso di una rilevazione successiva a quella utilizzata per le prime analisi, di una domanda sul motivo per cui il coniuge non è presente in famiglia ha evidenziato infatti che oltre l'80 per cento degli individui che hanno dichiarato stato civile coniugato e per i quali non è presente il coniuge si trova di fatto in una situazione di "coppia a distanza". L'utilizzo delle risposte al quesito introdotto permetterebbe di risolvere a monte, cioè in fase di rilevazione, la specifica incoerenza nel caso di separati di fatto e di accettare come corretto lo stato di coniugato di coloro che dichiarato di essere coppie a distanza. Le regole coinvolte potrebbero dunque non essere più nell'insieme dei controlli effettuati dalla procedura famiglie, con la conseguenza che il dato rilevato subirebbe meno interventi correttivi e soprattutto si introdurrebbe una nuova tipologia familiare, fino ad oggi non rilevata, che nel tempo potrebbe assumere un interesse crescente.

L'analisi dell'applicabilità alla Rcfl e le valutazioni emerse in seguito all'analisi delle performance della procedura famiglie hanno messo in luce l'opportunità di alcuni interventi, come quello appena illustrato per l'individuazione delle "coppie a distanza", che potrebbero effettuarsi sulla stessa: questi riuscirebbero a ridurre ulteriormente l'impatto, tuttora piuttosto contenuto, delle correzioni effettuate. Tali azioni, tuttavia, potranno più opportunamente intraprendersi in concerto con una revisione della stessa procedura nell'ambito delle diverse indagini sociali che la utilizzano.

In generale, la procedura famiglie applicata ai dati della Rcfl ha mostrato che l'impatto delle correzioni è piuttosto ridotto. Ciò è sicuramente dovuto alla nuova tecnica di rilevazione dei dati che permette una revisione dell'informazione al momento dell'acquisizione mediante il controllo dei domini delle variabili e soprattutto attraverso l'implementazione di regole di incompatibilità in linea che verificano la coerenza tra le risposte fornite. Il passaggio della procedura famiglie sui dati Rcfl ha segnalato che gli individui con almeno un errore in una delle variabili anagrafiche e di relazione familiare sono circa l'1 per cento del totale.

7.2.5.3 - La procedura famiglie nella Rcfl

L'unità di rilevazione della Rcfl è la famiglia di fatto e non quella anagrafica. La famiglia di fatto è costituita dall'insieme delle persone coabitanti legate da vincoli di matrimonio o parentela, affinità, adozione, tutela o da vincoli affettivi. I criteri secondo il quale un individuo

appartiene alla famiglia di fatto sono dunque quelli di convivenza abituale all'interno della famiglia e di legame di parentela e/o affettivo con gli altri componenti della famiglia stessa. Un aspetto innovativo della Rcfl rispetto alla Rtfl riguarda la registrazione dell'informazione sulla relazione di parentela con l'intestatario del foglio di famiglia (la persona di riferimento) ad un forte dettaglio: si individuano 17 diverse relazioni di parentela contro le 6 possibilità di classificazione presenti nella Rtfl. Questo ampliamento dell'informazione raccolta permette di applicare alla Rcfl la procedura di ricostruzione delle famiglie, individuandone i nuclei esistenti e classificando le stesse in 41 diverse tipologie familiari. I nuclei sono definiti dai legami di coppia e genitori/figli e sono di 4 tipi: coppia con figli, coppia senza figli, monogenitore maschio, monogenitore femmina. Una famiglia può coincidere con un nucleo, può essere formata da un nucleo più altri membri aggregati, da più nuclei (con o senza membri aggregati), o da nessun nucleo (persone sole, famiglie composte da due sorelle, da un genitore con figli separato, divorziato o vedovo, eccetera). Dunque, nella "procedura famiglia", una delle caratteristiche salienti che distingue i diversi tipi di famiglia è la presenza o l'assenza, all'interno della famiglia, di nuclei, ma anche, nel caso questi esistano, il numero di nuclei ed il tipo di nucleo. Le tipologie familiari si differenziano inoltre in base al sesso e stato civile dei genitori, alle relazioni di parentela tra i componenti di differenti nuclei e alla presenza di altri membri isolati.

La "procedura famiglie" interviene sul file dei dati grezzi, cioè dei dati ottenuti nella fase di rilevazione, con il controllo e l'eventuale correzione delle notizie anagrafiche e di relazione tra i componenti; l'insieme dei dati individuali, con l'esclusione delle informazioni anagrafiche, vengono invece sottoposti al generale piano di controllo e correzione di cui si è già parlato precedentemente. Queste due procedure di correzione lavorano in parallelo e, alla fine del processo, i dati elaborati da ciascuna vengono integrati in un unico file finale di dati puliti. In tale file si ritrovano anche le 7 nuove variabili, prodotte dalla procedura famiglie, che classificano, come detto precedentemente, le famiglie in nuclei e tipologie familiari.

In particolare vengono inserite sulfile successivamente messo a disposizione dell'utente:

- la tipologia familiare (**tf**)
- tre variabili che identificano il gruppo nucleo:
 - numero nucleo 1 (**nn1**)
 - tipo nucleo 1 (**tn1**)
 - relazione di parentela nel nucleo1 (**rpn1**)
- altre 3 variabili che identificano il gruppo nucleo secondo una logica a volte differente dalla precedente:
 - numero nucleo 2 (**nn2**)
 - tipo nucleo 2 (**tn2**)
 - relazione di parentela nel nucleo 2 (**rpn2**).

Il gruppo nucleo con indice 1 riporta i figli come sempre appartenenti al nucleo dei genitori (eccetto i figli coniugati), mentre il gruppo nucleo con indice 2 riporta nel nucleo dei genitori solo i figli minori e i figli celibi o nubili, mentre i figli separati, divorziati o vedovi che risultano conviventi con i genitori risultano membri isolati. Ovviamente anche la relazione di parentela del nucleo cambia in funzione dell'appartenenza al nucleo. E' importante sottolineare che, sebbene vengano individuati i nuclei con entrambi i criteri suddetti, la tipologia familiare che viene attribuita nel caso in cui nn1 si differenzia da nn2 è quella relativa ad nn2.

Il "numero nucleo" è un progressivo da 0 (persona singola) a n. dove n è un numero intero che identifica tutti i componenti del medesimo nucleo.

Il “tipo nucleo” è un codice volto a differenziare alcuni particolari tipologie di nuclei, e cioè:

- Persona isolata
- Coppia con figli
- Coppia senza figli
- Monogenitore maschio
- Monogenitore femmina.

La “relazione di parentela nel nucleo” privilegia, ove possibile, la figura femminile attribuendogli la qualifica di capo-nucleo e specifica, all’interno di un nucleo, il rapporto genitori-figli nel modo seguente:

- Persona singola
- Capo nucleo
- Coniuge o convivente del capo nucleo
- Figlio.

La “tipologia familiare” è la variabile prodotta dalla procedura che individua sinteticamente 41 tipologie familiari, e cioè.

- famiglie senza nuclei
 - persona sola
 - genitore con figli non celibi o nubili
 - insieme di parenti
 - insieme di parenti più altri
 - insieme di persone non parenti
- famiglie con un solo nucleo senza persone isolate
 - coppia coniugate senza figli
 - coppia non coniugate senza figli
 - coppia coniugate con figli
 - coppia non coniugate con figli
 - genitore maschio celibe solo con figli
 - genitore maschio coniugato non convivente solo con figli
 - genitore maschio separato solo con figli
 - genitore maschio divorziato solo con figli
 - genitore maschio vedovo solo con figli
 - genitore femmina nubile solo con figli
 - genitore femmina coniugata non convivente solo con figli
 - genitore femmina separata solo con figli
 - genitore femmina divorziata solo con figli
 - genitore femmina vedova solo con figli
- famiglie con un solo nucleo con persone isolate
 - coppia coniugate senza figli
 - coppia non coniugate senza figli
 - coppia coniugate con figli
 - coppia non coniugate con figli
 - genitore maschio celibe solo con figli
 - genitore maschio coniugato non convivente solo con figli
 - genitore maschio separato solo con figli
 - genitore maschio divorziato solo con figli
 - genitore maschio vedovo solo con figli

- genitore femmina nubile solo con figli
- genitore femmina coniugata non convivente solo con figli
- genitore femmina separata solo con figli
- genitore femmina divorziata solo con figli
- genitore femmina vedova solo con figli
- famiglie con due nuclei senza altre persone
 - con due generazioni
 - di tipo fraterno
 - binucleare di altro tipo
- famiglie con due nuclei con altre persone
 - con due generazioni
 - di tipo fraterno
 - binucleare di altro tipo
- famiglie con tre o più nuclei
 - senza altre persone
 - con altre persone.

7.3 - Lo stimatore e la procedura di riporto all'universo

Le stime prodotte dalla rilevazione continua sulle forze di lavoro sono stime di frequenze assolute e relative, riferite a tutti i componenti delle famiglie residenti in Italia, anche se temporaneamente all'estero. Sono escluse le famiglie che vivono abitualmente all'estero e i membri permanenti delle convivenze (istituti religiosi, caserme eccetera).

Tali stime fanno riferimento con diversa cadenza temporale ai seguenti domini territoriali¹³:

- territorio nazionale (con cadenza trimestrale e annuale¹⁴);
- ripartizioni geografiche (con cadenza trimestrale e annuale);
- regioni (con cadenza trimestrale e annuale);
- province (solo stime annuali, o calcolate come media degli ultimi 4 trimestri consecutivi).

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

Lo stimatore scelto per la Rcfl è uno stimatore di ponderazione vincolata¹⁵, cioè uno stimatore in cui i pesi finali consentono di ottenere nell'ambito di diversi domini territoriali (regioni, province autonome di Trento e Bolzano, province, grandi comuni) stime di popolazione residente (per sesso e classi di età) uguali ai corrispondenti totali noti di fonte anagrafica. I pesi sono calcolati a livello familiare e ciò implica che a ciascun individuo appartenente alla stessa

¹³ Si sta valutando la possibilità di produrre stime dei principali aggregati in media annua, riferite ai 13 comuni con popolazione superiore ai 250mila abitanti (grandi comuni). Di questo obiettivo si è tenuto già conto nella procedura di riporto all'universo.

¹⁴ E' in fase di studio un progetto per la produzione di stime mensili dei principali aggregati a livello nazionale.

¹⁵ Nella letteratura in lingua inglese tale stimatore è noto con il nome di *calibration estimator*; per ulteriori informazioni si veda Deville & Sarndal (1992).

famiglia venga attribuito un medesimo peso di riporto all'universo; l'unicità del peso a livello familiare assicura la coerenza delle stime familiari con le stime individuali.

Il regolamento comunitario 577/98 stabilisce alcune condizioni per la procedura di riporto all'universo dei dati campionari e in particolare impone che i pesi finali debbano essere calcolati tenendo conto:

- della probabilità di inclusione delle unità nel campione ;
- della distribuzione della popolazione a livello regionale per sesso e classi di età quinquennali nota da fonti esterne all'indagine.

Lo stimatore della Rcf1, che soddisfa i requisiti suddetti, è stato studiato per soddisfare molteplici esigenze informative (attuali o future) e per cercare di dare maggiore accuratezza alle stime provinciali in media annua; è stato inoltre predisposto per verificare la possibilità di ottenere stime affidabili a livello dei grandi comuni, a livello mensile, per i cittadini stranieri, e a livello di sistemi locali del lavoro (utilizzando stimatori composti).

Lo stimatore del totale di una generica variabile Y assume la seguente forma:

$$\tilde{Y} = \sum_{j \in s} y_j w_j \quad (1)$$

dove con riferimento al campione s si indica con:

y_j il valore della variabile Y osservato sulla j -ma unità campionaria;

w_j è il peso di riporto all'universo (o peso finale) associato alla j -ma unità campionaria.

I pesi di riporto all'universo w_j vengono determinati attraverso tre fasi successive:

- calcolo del peso base k_j , ottenuto come reciproco della probabilità di inclusione della j -ma unità campionaria;
- calcolo del fattore di correzione δ_j per mancata risposta totale da applicare al peso base, ottenuto come l'inverso del tasso di risposta. Questo fattore è necessario per tenere conto delle interviste non realizzate e quindi riequilibrare l'allocatione del campione effettivo tra i vari strati previsti dal disegno campionario. Nella Rcf1 questo fattore tiene conto, inoltre, dei tassi di risposta specifici di alcune tipologie di famiglie in modo da riequilibrare la distribuzione per tipologia familiare delle famiglie intervistate rispetto a quella del campione teorico;
- calcolo del fattore di correzione γ_j da applicare al peso base corretto $d_j = k_j \delta_j$ che consente di soddisfare la condizione di uguaglianza tra i totali noti della popolazione e le corrispondenti stime campionarie.

Lo stimatore del totale (1) può quindi essere scritto come segue

$$\tilde{Y} = \sum_{j \in s} y_j w_j = \sum_{j \in s} y_j k_j \delta_j \gamma_j = \sum_{j \in s} y_j d_j \gamma_j \quad (2)$$

7.3.1 - Calcolo del peso base come reciproco della probabilità di inclusione

Il peso base k_j assegnato a ciascuna famiglia è definito come il reciproco della probabilità di inclusione $k_j = 1/\pi_j$. Per la natura del disegno campionario della Rcfl (a 2 stadi, con stratificazione delle unità di primo stadio (comuni), con selezione di tutti i comuni autorappresentativi (Ar), con estrazione di un solo comune campione per ciascuno strato non autorappresentativo (Nar), con estrazione di un campione casuale di famiglie da ciascun comune campione), la probabilità di inclusione, e dunque il peso base, sono gli stessi per tutte le famiglie appartenenti allo stesso comune (o strato); il peso base può essere calcolato come segue

$$k_{hi} = \frac{1}{\pi_{hi}} = \frac{P_h}{P_{hi}} \frac{M_{hi}}{m_{hi}} \quad (3)$$

dove :

h denota l'indice di strato;

i è l'indice di comune;

P_h indica il totale della popolazione residente nello strato h ;

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

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

m_{hi} indica il numero di famiglie campione nel comune i dello strato h ¹⁶.

I pesi base vengono calcolati sulla base dei dati anagrafici più aggiornati relativi alla popolazione e alle famiglie residente nei comuni italiani.

7.3.2 - La mancata risposta totale delle famiglie

Come accade per la maggior parte delle indagini campionarie, anche per la Rcfl il numero di famiglie realmente intervistate è inferiore a quello previsto dal disegno campionario a causa dell'impossibilità di intervistare alcune famiglie estratte¹⁷. Questo fenomeno che prende il nome di mancata risposta totale familiare (d'ora in poi Mrtf) può manifestarsi con diversa intensità nei diversi comuni di una stessa provincia. In generale l'intensità della Mrtf risulta più alta nelle grandi città e nelle città capoluogo e quindi è necessario intervenire per ripristinare il corretto peso delle unità campione nei diversi strati dei domini territoriali pianificati.

Se la Mrtf si manifesta in maniera casuale, cioè non è correlata alle caratteristiche della famiglia e degli individui, si può calcolare un fattore di correzione per mancata risposta totale m_{hi}/m_{hi}^* , ottenuto come l'inverso del tasso di risposta familiare per ciascuno comune, dove m_{hi}^* è il numero di famiglie campione effettivamente intervistate nel comune i dello strato h .

¹⁶ Si noti che la precedente espressione diventa $k_{hi} = M_{hi}/m_{hi}$ nel caso dei comuni Ar che costituiscono uno strato a se.

¹⁷ Le ragioni possono essere molteplici: impossibilità da parte del rilevatore di effettuare il numero di interviste a lui affidate, impossibilità di reperire le famiglie (base e sostitute), rifiuto da parte delle famiglie (base e sostitute) a partecipare all'intervista.

Il peso base corretto per mancata risposta totale diventa perciò

$$k_{hi}^* = \frac{1}{\pi_{hi}} = \frac{P_h}{P_{hi}} \frac{M_{hi}}{m_{hi}^*} \quad (4).$$

In effetti, però, la Rcfl risente di un'auto-selezione delle famiglie campione dovuta alla diversa probabilità di risposta delle diverse tipologie di famiglie; realizzando meno interviste del dovuto per alcune tipologie familiari, e realizzandone di più per altre, si determina una distorsione del campione effettivo rispetto a quello teorico che potrebbe generare una certa distorsione nelle stime finali.

Questo problema è il risultato del concorrere di due cause. In primo luogo, il meccanismo di sostituzione delle famiglie "cadute" all'interno della quartina (costruita sulla base del disegno in modo casuale mediante un passo di estrazione sistematico), meccanismo che fornisce automaticamente ai rilevatori i nominativi delle famiglie sostituite, di fatto garantisce la casualità della sostituzione ma non garantisce che le famiglie sostituite siano il più possibile simili a quelle base sostituite. E' ormai noto, invece, che alcune tipologie di famiglie difficilmente reperibili da parte degli intervistatori (single giovani, giovani coppie, ecc) vengano di fatto sostituite con altre aventi diverse caratteristiche e più disponibili a partecipare all'indagine (coppie anziane, famiglie numerose, ecc)¹⁸.

In secondo luogo, in caso di irreperibilità o rifiuto delle famiglie da reintervistare (alla 2^a, 3^a e 4^a intervista) la procedura attualmente utilizzata non prevede alcuna sostituzione nel caso di intervista telefonica¹⁹ generando un sottodimensionamento del campione e ottenendo, in particolari situazioni, una distribuzione di famiglie campione diversa dalla teorica²⁰.

Lo schema seguente mostra un ventaglio di situazioni che si possono verificare in un generico trimestre durante la fase di intervista e che determinano l'ottenimento di un campione effettivo diverso da quello teorico. In grigio sono evidenziate le famiglie che devono essere intervistate nel trimestre. Generalmente, queste dovrebbero essere le famiglie base, tranne i casi in cui le famiglie base sono state già sostituite in precedenti occasioni di indagine (è il caso delle quartine 4,10, m-1 nell'esempio). In effetti buona parte delle famiglie che devono essere intervistate partecipano effettivamente all'indagine (segnalate con "i" nello schema); altre rifiutano di partecipare o non vengono raggiunte e quindi non vengono intervistate (segnalate con "ni"). Considerando solo queste ultime, alcune di queste non devono essere sostituite (se Cati) o non si riesce a sostituirle per fine del periodo di rilevazione ammesso (è il caso delle quartine 2 e m-1); per altre, invece, si cerca di intervistare con più o meno successo quelle che seguono nella quartina (è il caso delle quartine 4, 7 e 11).

¹⁸ Questo problema non è affatto imputabile al cattivo comportamento dei rilevatori bensì alla diversa probabilità di contatto e alla diversa disponibilità a partecipare all'indagine di determinate tipologie di famiglie.

¹⁹ Per un approfondimento si veda il paragrafo sulle tecniche di intervista

²⁰ Ad esempio, durante i periodi destinati alle vacanze può accadere di avere una rappresentazione carente di tipologie di famiglie che più delle altre possono permettersi periodi lunghi fuori dal comune di residenza.

Prospetto 7. 5 - Schema di sostituzione delle famiglie campione all'interno delle quartine

	Quartina 1	Quartina 2	Quartina 3	Quartina 4	Quartina 5	Quartina 6	Quartina 7	Quartina 8	Quartina 9	Quartina 10	Quartina 11	Quartina	Quartina	Quartina m-2	Quartina m-1	Quartina m
Famiglia base	i	ni	i		i	i	ni	i	i		ni	i	i	i		i
Prima famiglia sostitutiva				ni			ni				i				ni	
Seconda famiglia sostitutiva				i			ni			i						
Terza famiglia sostitutiva							ni									

Per verificare l'esistenza di auto-selezione e monitorare costantemente il suo effetto si è fatto uso delle informazioni anagrafiche presenti nella lista delle famiglie campione (base e sostitutive) compilata da tutti i comuni che partecipano all'indagine. La lista contiene informazioni sul numero di componenti della famiglia, età e sesso del capofamiglia, esistenza del coniuge; dall'estrazione del 2004 viene rilevata anche la cittadinanza del capofamiglia, e ciò consentirà di monitorare la partecipazione all'indagine degli stranieri residenti²¹. La lista delle famiglie del campione teorico, ulteriormente depurata dalle famiglie che vengono dichiarate non eleggibili dai rilevatori in fase di contatto²², può essere dunque efficacemente utilizzata per ottenere una buona stima della distribuzione delle famiglie residenti, per provincia, per numero di componenti e per tipologia familiare.

Nel grafico che segue si mettono a confronto le distribuzioni delle famiglie per tipologia rispettivamente:

- utilizzando tutte le famiglie base che risultano eleggibili; tale distribuzione fornisce una buona stima della distribuzione delle famiglie anagrafiche residenti per tipologia familiare;
- utilizzando solo le famiglie del campione effettivo²³ (che comprende, naturalmente, sia famiglie base che sostitute).

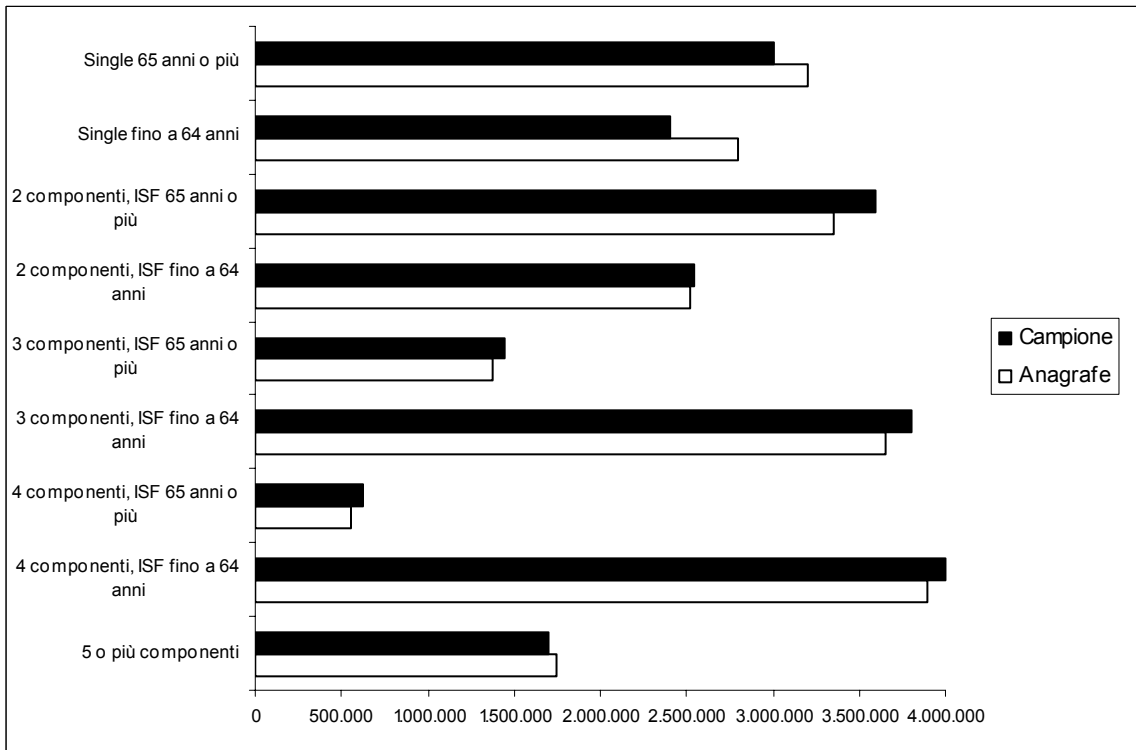
In assenza di auto-selezione le due distribuzioni dovrebbero coincidere.

²¹ Al momento la quota di stranieri presenti nel campione effettivamente intervistato è più bassa di quella risultante dalle fonti anagrafiche; l'informazione sulla cittadinanza del capofamiglia sull'intera lista delle famiglie campione aiuterà a comprendere se ciò sia dovuto al ritardo nell'aggiornamento delle liste delle famiglie campione o ai più bassi tassi di risposta per gli stranieri. Questa informazione consentirà di applicare, se necessario, uno specifico fattore di correzione della mancata risposta per le famiglie di stranieri.

²² Si veda il capitolo 6.

²³ Sebbene il disegno preveda di intervistare tutti i componenti della famiglia di fatto, per effettuare in modo corretto il confronto, in questa fase la tipologia familiare viene ricostruita sulla base dei dati risultanti in anagrafe anche per le famiglie effettivamente intervistate.

Figura 7.1 - Distribuzione delle famiglie campione e dell'anagrafe per tipologia familiare. I Trimestre 2005

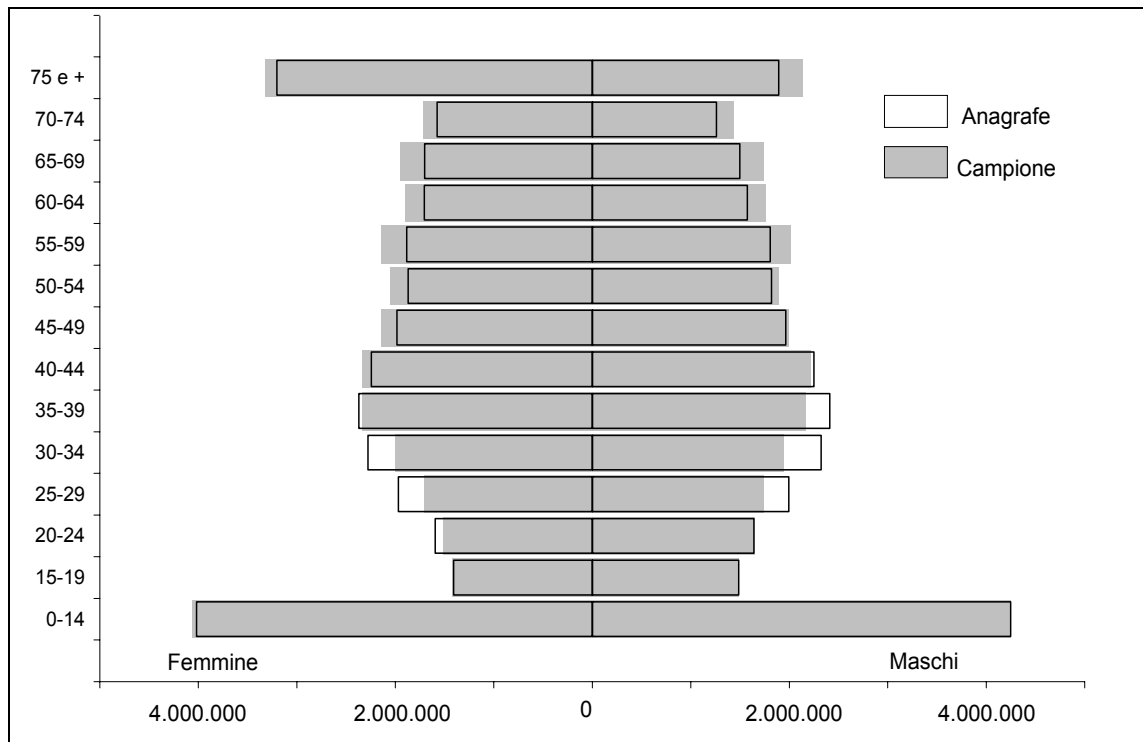


Si osserva che il campione effettivo non rappresenta adeguatamente alcune tipologie familiari (single di entrambi i sessi e in particolare quelli con meno di 65 anni) mentre ne sovrarappresenta altre (famiglie di 2 componenti con intestatario della scheda anagrafica con più di 64 anni, famiglie di 3 e 4 componenti).

L'auto-selezione a livello familiare determina un effetto distorsivo anche sulla distribuzione riferita agli individui. Come prima, al fine di verificare se esistono delle differenze tra la distribuzione teorica e quella risultante dal campione, nel grafico che segue si confrontano due distribuzioni della popolazione residente in famiglia per sesso e classi di età:

- la prima è calcolata sulla base dei dati forniti mensilmente da tutte le anagrafi comunali al servizio demografico dell'Istat e costituisce il totale noto di riferimento della procedura di riporto all'universo (rappresentato nel grafico come linea);
- la seconda è stimata utilizzando solo le famiglie del campione effettivo (che comprende, naturalmente, sia famiglie base che sostitute); in assenza di auto-selezione dovrebbe coincidere con la precedente;

Figura 7.2 - Distribuzione degli individui del campione e della popolazione per sesso e classi di età



Anche a livello individuale si osserva che il campione effettivo non rappresenta adeguatamente alcuni segmenti della popolazione (giovani di entrambi i sessi e in particolare quelli tra 25-34 anni) mentre risulta sovra-rappresentarne altri (in particolare ultra cinquantacinquenni di entrambi i sessi)²⁴.

7.3.2.1 - Correzione per mancata risposta per tipologia familiare

A causa dell' autoselezione del campione a livello familiare e individuale, prima di procedere al calcolo del peso finale mediante post-stratificazione, è necessario effettuare una correzione della mancata risposta familiare che tenga conto dei diversi tassi di risposta riscontrati per le diverse tipologie.

Il correttore per mancata risposta, calcolato per la generica regione r e la tipologia familiare f è

$$\delta_{rf} = \frac{F_{rf}^*}{\tilde{F}_{rf}} = \frac{(F_{rf}^{esiti} / F_r^{esiti}) \cdot F_r^*}{\tilde{F}_{rf}} \quad (5)$$

dove:

²⁴ Le differenze, in questo caso, sono sicuramente amplificate dal fatto che il dato anagrafico che costituisce il totale noto è aggiornato a tre mesi prima l'inizio del trimestre, mentre il campione del trimestre è stato estratto da una lista di famiglie molto meno aggiornata (per alcuni trimestri alcuni gruppi di rotazione sono stati estratti fino a 2 anni prima).

F_{rf}^* è il totale delle famiglie della tipologia f nella regione r stimato utilizzando le informazioni anagrafiche del campione teorico e i dati del bilancio demografico annuale;

\tilde{F}_{rf} è il totale delle famiglie della tipologia f nella regione r stimate sul campione effettivo applicando il peso base corretto k_{hi} ; $\tilde{F}_{rf} = \sum_{h,i \in r} y_{rhif} k_{rhi}$ dove y_{rhif} è una variabile indicatrice che assume valore 1 se la famiglia è della tipologia f ;

F_{rf}^{esiti} è il totale delle famiglie della tipologia f nella regione r stimato utilizzando le informazioni anagrafiche di tutte le famiglie eleggibili del campione teorico;

F_r^{esiti} è il totale delle famiglie nella regione r stimato utilizzando le informazioni anagrafiche di tutte le famiglie eleggibili del campione teorico;

F_r^* è il totale delle famiglie della regione r calcolato sui dati del bilancio demografico annuale

Ne consegue che il peso base corretto per tipologia familiare diventa :

$$d_{rhif} = k_{rhi} \delta_{rf} \quad (6)$$

ed è tale da rispettare il vincolo

$$\sum_{h,i \in r} y_{rhif} d_{rhif} = F_{rf}^* \quad (7)$$

Il prodotto tra il peso base corretto e il fattore di correzione che tiene conto della Mrt per tipologie familiari determina il peso base corretto per tipologia familiare, o più semplicemente peso diretto d_{rhif} , da utilizzare come input per il calcolo dei pesi finali di riporto all'universo mediante la post-stratificazione.

7.3.3 - I pesi finali

Per il calcolo dei pesi finali di riporto all'universo si utilizza una procedura generalizzata sviluppata dall'Istat che prende il nome di Genesees²⁵ già adottata per la Rtfl, così come per numerose altre indagini condotte dall'Istat, e largamente utilizzata da altri Istituti Nazionali di Statistica. Questa procedura consente di ottenere i pesi finali della famiglia degli stimatori di ponderazione vincolata (in inglese, *calibration estimator*), calcolando per la Rcfl un peso unico per famiglia e individui.

I pesi finali si ottengono risolvendo un problema di minimo vincolato, in cui la funzione da minimizzare è una funzione di distanza tra i pesi diretti d_j e i pesi finali w_j e i vincoli sono

²⁵ Per ulteriori dettagli si veda: Genesees: Funzione riponderazione (2005)

condizioni di uguaglianza delle stime campionarie di alcune variabili ausiliarie con i rispettivi totali noti desunti da fonti esterne all'indagine;

$$\begin{cases} \min \left\{ \sum_{j \in s} \text{dist}(d_j, w_j) \right\} \\ \sum_{j \in s} \mathbf{x}_j w_j = \mathbf{t} \end{cases} \quad (8)$$

dove \mathbf{t} è il vettore dei totali noti e \mathbf{x}_j è il vettore delle variabili ausiliarie osservate sulla j -ma unità campionaria.

“La funzione di distanza logaritmica troncata garantisce che i pesi finali w_j siano positivi e che i fattori correttivi γ_j siano compresi in un predeterminato intervallo, escludendo l'applicazione di fattori correttivi troppo grandi (con pesi finali molto vicini a quelli dei pesi diretti). La funzione di distanza e la sua derivata prima rispetto a w_j devono soddisfare alcune condizione di regolarità (Deville e Särndal, 1992).

7.3.4 - I totali noti

I totali noti introdotti come vincoli nel calcolo dei pesi finali, consentono di migliorare l'accuratezza delle stime: quanto più le variabili ausiliarie considerate sono correlate con le variabili oggetto d'indagine, tanto più si riduce la distorsione delle stime. L'utilizzo dello stimatore di ponderazione vincolata è subordinato alla conoscenza periodica (trimestrale per l'indagine in oggetto) di variabili ausiliarie riferite all'intera popolazione che siano correlate con le variabili oggetto d'indagine. Un notevole impulso al miglioramento della procedura è sicuramente dovuto alla disponibilità all'interno dell'Istituto di nuove, migliori e più tempestive informazioni da utilizzare come variabili ausiliarie. Non meno importante è stata la disponibilità di elaboratori elettronici sempre più potenti e veloci in grado di gestire la mole di calcoli richiesti dalla procedura.

Attualmente la procedura di riporto all'universo vincola le stime campionarie trimestrali ai seguenti totali noti:

- popolazione residente in famiglia a livello regionale per sesso e 14 classi di età;
- popolazione residente in famiglia a livello provinciale per sesso e 5 classi di età;
- popolazione residente in famiglia nei 12 grandi comuni per sesso 5 classi di età;
- cittadini stranieri residenti in famiglia per regione, sesso e nazionalità;
- numero di famiglie residenti per regione.
- popolazione residente in famiglia per ciascun mese.

La conoscenza di questi totali noti e il loro inserimento nella procedura di calcolo dei pesi finali di riporto all'universo consente di soddisfare ulteriori esigenze informative (attuali e future), oltre alla tradizionale produzione di stime a livello regionale (trimestrali) o provinciale (in media annua), quali:

- stime a livello sub-regionale: stime provinciali trimestrali e stime per grandi comuni;

- stime riguardanti gli stranieri residenti;
- stime a livello familiare;
- stime mensili;

I totali noti trimestrali sono ottenuti combinando le statistiche relative alle seguenti fonti disponibili in Istat:

- Popolazione residente per età, sesso e stato civile al 1° Gennaio²⁶; fornisce i risultati della Rilevazione sulla “Popolazione residente comunale per sesso, anno di nascita e stato civile” condotta dall’Istat presso le Anagrafi di tutti comuni italiani (è disponibile con circa tredici mesi di ritardo rispetto al periodo di riferimento);
- Bilancio demografico della popolazione residente: fornisce i dati relativi alla popolazione, alle famiglie e alle convivenze di tutti comuni italiani riferiti al 31 dicembre di ciascun anno (è disponibile con circa sei mesi di ritardo rispetto al periodo di riferimento);
- Bilancio demografico della popolazione straniera residente: fornisce i dati relativi ai cittadini stranieri per sesso e cittadinanza rilevati su tutti comuni italiani e riferiti al 31 dicembre di ciascun anno (è disponibile con circa otto mesi di ritardo rispetto al periodo di riferimento);
- Stima rapida della popolazione per sesso, età e provincia di residenza: fornisce anticipatamente la struttura “provvisoria” della popolazione riferita al 1 gennaio, 1 aprile, 1 luglio e 1 ottobre;
- Censimento della popolazione 2001: fornisce l’ammontare della popolazione complessiva e dei cittadini stranieri che risiedono nelle convivenze.

7.3.4.1 - I totali noti della popolazione residente nelle regioni e nelle province

In passato, poiché non erano disponibili con frequenza trimestrale i totali noti a livello provinciale, la procedura di riporto all’universo per la Rilevazione trimestrale sulle forze di lavoro era articolata in due passi successivi: nel primo, si correggeva il peso base corretto per mancata risposta totale imponendo i vincoli relativi alla popolazione residente per provincia e sesso; nel secondo, a partire dal peso calcolato al primo passo si giungeva al peso finale imponendo i vincoli relativi alla popolazione residente per regione, sesso e 14 classi di età (0-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75 e oltre). L’applicazione della procedura in due passi presentava il limite di non assicurare al termine del secondo passo il rispetto integrale di tutti i vincoli introdotti al primo passo²⁷.

Nell’attuale procedura di calcolo dei pesi di riporto finali della Rilevazione continua sulle forze di lavoro, tutti i vincoli vengono imposti contemporaneamente. I principali vincoli riguardanti la popolazione residente in famiglia sono i seguenti:

- popolazione per regione²⁸, sesso e 14 classi d’età (0-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75 e oltre);
- popolazione per provincia, sesso e 5 classi d’età (0-14; 15-29; 30-49; 50-64; 65 e più).

²⁶ Per informazioni e dati riguardanti questo punto e i tre successivi si veda <http://demo.istat.it>

²⁷ Uno dei problemi più frequenti riguardava le stime sulla popolazione provinciale per classe di età, che non essendo vincolata poteva risultare molto diversa dai dati anagrafici.

²⁸ Per il Trentino Alto Adige questi stessi vincoli, nonché tutti gli altri che verranno presentati nei prossimi paragrafi, sono posti separatamente per le province autonome di Trento e Bolzano.

I totali noti trimestrali sono desunti dalle distribuzioni della popolazione per provincia, sesso ed età al 1 gennaio, 1 aprile, 1 luglio e 1 ottobre (note anche come “stime rapide”). Queste distribuzioni sono calcolate dal Servizio demografico dell’Istat sulla base dei dati anagrafici trasmessi mensilmente dai comuni. Dalla popolazione residente complessiva vengono esclusi gli individui che risultano residenti nelle convivenze.

$$P_{rse}^* = P_{rse}^{(sr)} - C_{rse} \quad (9)$$

$$P_{pse}^* = P_{pse}^{(sr)} - C_{pse} \quad (10)$$

P_{rse}^* , P_{pse}^* sono i totali noti rispettivamente per la regione r e la provincia p distintamente per sesso s e per le rispettive classi di età e ;

$P_{rse}^{(sr)}$, $P_{pse}^{(sr)}$ sono le popolazioni residenti rispettivamente nella regione r e nella provincia p di sesso s e classe di età e desunte dalla fonte “Stima rapida;

C_{rse} , C_{pse} è la popolazione residente nelle convivenze di sesso s e classe di età e ²⁹

7.3.4.2 - I totali noti della popolazione residente nei 12 grandi comuni

La domanda sempre crescente di statistiche sul mercato del lavoro a livello sub-provinciale, nonché la specificità della popolazione residente nei comuni più grandi ha suggerito l’introduzione nella procedura di calcolo dei pesi di riporto del vincolo sulla popolazione residente nei 13 grandi comuni con popolazione superiore alle 250mila unità (Roma, Milano, Napoli, Firenze, Torino, Bologna, Genova, Venezia, Verona, Bari, Palermo, Catania e Messina) per sesso e 5 classi d’età (0-14; 15-29; 30-49; 50-64; 65 e più).

I dati relativi al 1° gennaio sono desunti dalla Popolazione dei comuni; poiché tali dati sono disponibili diverso tempo dopo il periodo di riferimento (circa 13 mesi) le stime trimestrali P_{cse}^* riferite al 1° gennaio, 1° aprile, 1° luglio e 1° ottobre vengono aggiornate dal servizio Formazione e Lavoro dell’Istat, assumendo che l’incidenza della popolazione del grande comune sulla popolazione dell’intera provincia di appartenenza si mantenga costante nel corso dell’anno, come indicato nella formula che segue:

$$P_{cse}^* = \frac{P_{cse}^{(bd)}}{P_{pse}^{(bd)}} P_{pse}^{(sr)} - C_{cse} \quad (11)$$

dove:

$P_{cse}^{(bd)}$ è la popolazione residente nel grande comune c di sesso s e classe di età e desunta dalla fonte “Bilancio demografico”;

²⁹ I dati sulla popolazione residente nelle convivenze sono disponibili alla data del Censimento della popolazione del 2001 e successivamente sono disponibili con cadenza annuale dal Bilancio demografico.

$P_{pse}^{(bd)}$ è la popolazione residente nella provincia p di sesso s e classe di età e desunta dalla fonte “Bilancio demografico”;

$P_{pse}^{(sr)}$ è la popolazione residente nella provincia p di sesso s e classe di età e desunta dalla fonte “Stima rapida”

C_{cse} è la popolazione residente nelle convivenze nel grande comune c di sesso s e classe di età e .

7.3.4.3 - I totali noti sugli stranieri residenti

La presenza di stranieri residenti in Italia ha raggiunto negli ultimi anni livelli molto significativi. La produzione di stime relative agli stranieri residenti tramite la rilevazione sulle forze di lavoro è un obiettivo ritenuto prioritario sia a livello nazionale sia a livello internazionale. Finora le stime riguardanti gli stranieri non sono state diffuse in quanto il tasso di risposta dei cittadini stranieri era notevolmente più basso di quello degli italiani, e dunque, le stime della popolazione straniera residente risultavano molto più basse dei relativi dati demografici. Con il passaggio alla rilevazione continua, e a seguito del miglioramento dei tassi di risposta dovuti al lavoro svolto dalla nuova rete di rilevazione, si è deciso di introdurre nella procedura di calcolo dei pesi finali di riporto anche alcuni vincoli sulla popolazione straniera residente per regione, distintamente per maschi, femmine, cittadini dell’Unione europea (a 25 Paesi), con altra cittadinanza.

I dati sulla popolazione straniera residente riferiti al 1 gennaio sono desunti dal “Bilancio demografico sugli stranieri”; le stime trimestrali riferite al 1 aprile, 1 luglio e 1 ottobre vengono aggiornate dal servizio Formazione e Lavoro assumendo che l’incidenza della popolazione straniera sulla popolazione complessiva della provincia di appartenenza si mantenga costante rispetto all’inizio dell’anno³⁰.

I totali noti sugli stranieri S_{rs}^* e S_{ps}^* (rispettivamente a livello regionale e provinciale, per sesso) vengono perciò determinati come segue

$$S_{rs}^* = \sum_{p \in r} S_{ps}^* = \sum_{p \in r} \left(\frac{S_{ps}^{(is)}}{P_{ps}^{(bd)}} P_{ps}^{(sr)} - CS_{ps} \right) \quad (12)$$

dove

$S_{ps}^{(is)}$ è la popolazione straniera residente nella provincia p di sesso s desunta dalla fonte “Bilancio demografico sugli stranieri”;

$P_{ps}^{(bd)}$ è la popolazione residente nella provincia p di sesso s desunta dalla fonte “Bilancio demografico” ;

³⁰ È in fase di studio la possibilità di ottenere stime più accurate della popolazione straniera riferite ai tre momenti dell’anno per cui non si hanno informazioni mediante l’uso di ulteriori fonti esterne all’indagine.

$P_{ps}^{(sr)}$ è la popolazione residente nella provincia p di sesso s desunta dalla fonte “Stima rapida” ;

CS_{ps} sono gli stranieri residenti nelle convivenze della provincia p di sesso s .

7.3.4.4 - I totali noti sul numero di famiglie

Negli ultimi anni è cresciuto l’interesse nei confronti dell’analisi del mercato del lavoro effettuata anche secondo una prospettiva familiare³¹. Ne deriva l’esigenza di produrre stime trimestrali non distorte anche del numero di famiglie residenti. A tale scopo, e tenendo conto del fatto che ogni gruppo di rotazione dovrebbe rappresentare esattamente un quarto della popolazione complessiva, è stato introdotto nella procedura di calcolo dei pesi finali di riporto all’universo un vincolo sul numero di famiglie residenti per regione e gruppo di rotazione.

I totali noti sulle famiglie residenti al 1° gennaio sono diffusi con il “Bilancio demografico”, quelle riferite al 1° aprile, 1° luglio e 1° ottobre vengono aggiornate dal servizio Formazione e Lavoro dell’Istat assumendo che il numero di componenti medio delle famiglie in ciascuna provincia rimanga invariato rispetto al “Bilancio Demografico” di inizio anno³².

I totali noti sulle famiglie F_r^* sono pertanto determinati come segue

$$F_r^* = \sum_{p \in r} \frac{F_p^{(bd)}}{P_p^{(bd)}} P_p^{(sr)} \quad (13)$$

dove:

$F_p^{(bd)}$ è il numero di famiglie della provincia p desunte dal “Bilancio demografico”;

$P_p^{(bd)}$ è la popolazione residente nella provincia p desunta dal “Bilancio demografico” ;

$P_p^{(sr)}$ è la popolazione residente nella provincia p desunta dalla fonte “Stima rapida” .

7.3.4.5 - I totali noti mensili

Uno dei cambiamenti più rilevanti tra la Rftl e la Rcfl è la ripartizione del campione trimestrale di famiglie nelle 13 settimane del trimestre. Può accadere, comunque, che in determinati periodi dell’anno (alcune settimane in estate , Natale, eccetera) i tassi di risposta risultino più bassi rispetto a quelli del resto del trimestre, distorcendo di fatto la distribuzione temporale del campione. Questa distorsione del campione può tradursi in una distorsione delle stime finali trimestrali nella misura in cui si manifestano mensilmente variazioni consistenti sul mercato del lavoro. La necessità di ristabilire l’equidistribuzione temporale del campione e la

³¹ Analisi e stime su famiglie sono contenute nel Rapporto annuale dell’Istat degli anni 2003, 2004, 2005 e sono contenute nel volume di media annua delle forze di lavoro del 2004.

³² La scelta di stimare per tre trimestri il numero di famiglie è stata ritenuta preferibile rispetto a lasciare invariato il numero di famiglie per quattro trimestri consecutivi; in quanto a fronte di una crescita della popolazione aumenterebbe anche il numero di componenti medio, in palese controtendenza con i dati demografici relativi agli anni più recenti.

possibilità di produrre in futuro stime mensili hanno condotto alla decisione di introdurre nella procedura di calcolo dei pesi finali di riporto vincoli anche sui mesi.

A livello mensile si impone che ciascun mese rappresenti una quota prefissata della popolazione residente a livello regionale per sesso. Uno dei mesi del trimestre (quello con 5 settimane) rappresenterà i 5/13 della popolazione complessiva; i restanti due mesi (con 4 settimane ciascuno) rappresenteranno entrambi i 4/13 della popolazione complessiva.

I totali noti mensili P_{rsm}^* vengono determinati secondo la formula:

$$P_{rsm}^* = \frac{P_{rs}^*}{13} z_m \quad (14)$$

dove:

P_{rs}^* è il totale noto per regione e sesso (vedi par. 7.2.4.1);

z_m è il numero di settimane nel mese (4 o 5).

7.4 - La valutazione dell'errore campionario

Il calcolo delle stime prodotte da un'indagine campionaria dovrebbe essere sempre accompagnato da una valutazione dell'accuratezza delle stime stesse, cioè da una misura della dispersione delle stime attorno al valore vero del parametro della popolazione che si intende stimare.

A tale proposito è opportuno distinguere due diverse fonti di errore che generano delle differenze tra le stime prodotte e il parametro oggetto di stima:

- l'errore campionario, proprio di ogni indagine campionaria, che deriva dall'aver osservato la variabile di interesse solo su una parte (campione) della popolazione;
- l'errore non campionario che deriva essenzialmente da:
 - errori nelle liste della popolazione utilizzate per selezionare le unità del campione (sotto-copertura o sovra-copertura, liste incomplete o doppie registrazioni);
 - mancate risposte totali;
 - mancate risposte parziali dovute a risposte mancanti o non ammissibili a causa di errori di rilevazione o di registrazione (risposte fuori dominio o risposte tra loro incompatibili sulla base di regole di coerenza tra variabili opportunamente definite);
 - tutto ciò che ha a che fare con le tecniche di indagine utilizzate e i comportamenti dei rilevatori.

Le tecniche adottate per il contenimento, la correzione e la valutazione dell'impatto sulle stime degli errori non campionari sono state affrontate nei capitoli 6 e 7 e nel paragrafo 7.2.

In questo paragrafo si descrivono le metodologie e le tecniche utilizzate per la valutazione dell'errore campionario associato alle stime prodotte dall'indagine sulle forze di lavoro³³.

Indicando con Y il parametro di interesse (ad esempio il totale della variabile Y nella popolazione), una misura dell'errore campionario di uno stimatore \hat{Y} di Y è l'errore quadratico medio (in inglese *mean square error*) così definito:

$$MSE(\hat{Y}) = E(\hat{Y} - Y)^2 = \sum_{s \in S} [\hat{Y}(s) - Y]^2 \cdot p(s)$$

dato dal valore atteso delle differenze al quadrato tra le stime campionarie $\hat{Y}(s)$ e il valore vero del parametro Y ; il valore atteso è calcolato su tutti i campioni s appartenenti allo spazio dei campioni S (dato il disegno campionario che è stato scelto) e ciascun campione ha un peso $p(s)$ pari alla sua probabilità di essere estratto.

L'errore quadratico medio può anche essere espresso nella forma seguente:

$$MSE(\hat{Y}) = E(\hat{Y} - Y)^2 = Var(\hat{Y}) + [E(\hat{Y}) - Y]^2 = Var(\hat{Y}) + [Bias(\hat{Y})]^2$$

distinguendo le due parti che lo compongono: la varianza campionaria dello stimatore e il quadrato della sua distorsione.

Come si può osservare dalla precedente espressione, nel caso di stimatori non distorti (o asintoticamente non distorti) l'errore quadratico medio coincide (o coincide approssimativamente al crescere della dimensione del campione) con la varianza campionaria; per tali stimatori, pertanto, il problema della valutazione dell'errore campionario si riduce al problema di stima della varianza campionaria. Oltre alla varianza campionaria sono comunemente utilizzate anche le seguenti due quantità:

³³ Negli sviluppi che seguono si assume l'assenza di errori non campionari.

l'errore assoluto (in inglese *standard error*): $\sigma(\hat{Y}) = \sqrt{\text{Var}(\hat{Y})}$

e l'errore relativo o coefficiente di variazione: $\varepsilon(\hat{Y}) = \frac{\sigma(\hat{Y})}{E(\hat{Y})}$

che spesso viene espresso in termini percentuali.

Conoscendo la stima \hat{Y} di un parametro Y della popolazione e la stima dell'errore assoluto $\hat{\sigma}(\hat{Y})$ ad essa associato, è possibile costruire un intervallo di confidenza che, con livello di fiducia P , contiene al suo interno il valore del parametro Y oggetto di stima³⁴; tale intervallo è:

$$\left\{ \hat{Y} - k_p \hat{\sigma}(\hat{Y}) \leq Y \leq \hat{Y} + k_p \hat{\sigma}(\hat{Y}) \right\}$$

dove il valore di k_p dipende dalla forma della distribuzione campionaria dello stimatore e dal valore scelto per il livello di confidenza P ; per grandi campioni (quali sono quelli dell'indagine sulle forze di lavoro) si fa comunemente riferimento alla distribuzione normale e si ha ad esempio, per $P=0,95$ $k_p=1,96$.

7.4.1 - La stima della varianza campionaria dello stimatore di ponderazione vincolata

Per alcuni stimatori, in particolare per quelli lineari (funzioni lineari delle osservazioni campionarie), è nota l'espressione dello stimatore della varianza campionaria; utilizzando le osservazioni campionarie, pertanto, è possibile ottenere una stima del parametro della popolazione e contestualmente una stima della sua precisione.

Lo stimatore utilizzato per la Rcfl è uno stimatore di ponderazione vincolata, le cui principali caratteristiche sono state illustrate nel paragrafo 7.2.

Lo stimatore di ponderazione vincolata è uno stimatore distorto, funzione non lineare delle osservazioni campionarie, del quale non è nota l'espressione dello stimatore della varianza campionaria.

Si dimostra che, al crescere della dimensione del campione, lo stimatore di ponderazione vincolata converge asintoticamente allo stimatore di regressione generalizzata (Deville e Särndal, 1992)³⁵. Ne consegue che per campioni sufficientemente grandi (quali sono quelli dell'indagine sulle forze di lavoro), è possibile affermare che lo stimatore di ponderazione vincolata abbia approssimativamente le stesse proprietà dello stimatore di regressione generalizzata (correttezza, consistenza) e che abbiano la stessa varianza campionaria.

Pertanto, il problema della stima della varianza campionaria dello stimatore di ponderazione vincolata viene ricondotto al più semplice problema di stima della varianza dello stimatore di regressione generalizzata.

Anche lo stimatore di regressione generalizzata è uno stimatore non lineare. Per calcolarne la varianza campionaria si può però ricorrere al metodo proposto da Woodruff (Woodruff,

³⁴ Una frazione pari a P degli intervalli di confidenza associati a tutti i campioni nello spazio dei campioni S contiene il valore vero del parametro Y .

³⁵ Più precisamente, è possibile definire una classe di stimatori di ponderazione vincolata, contenente tutti gli stimatori che scaturiscono dalla risoluzione di un problema di minimo vincolato, e che si differenziano per la misura di distanza adottata. All'interno di tale classe riveste un ruolo centrale lo stimatore di regressione generalizzata che si ottiene utilizzando la funzione di distanza euclidea; è possibile dimostrare che tutti gli stimatori di ponderazione vincolata convergono asintoticamente, all'aumentare della dimensione campionaria, allo stimatore di regressione generalizzata (Deville e Särndal, 1992).

1971); applicando tale metodo è possibile calcolare la varianza approssimata di uno stimatore non lineare (che sia però funzione regolare di totali) calcolando la varianza dell'espressione linearizzata dello stimatore³⁶:

$$\hat{Y} \cong \hat{Z} = \sum_{h=1}^H \hat{Z}_h \quad \text{da cui: } \hat{V}ar(\hat{Y}) \cong \hat{V}ar(\hat{Z}) = \sum_{h=1}^H \hat{V}ar(\hat{Z}_h)$$

$$\text{in cui } \hat{Z}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} z_{hij} w_{hij}, \quad h=1, \dots, H.$$

Nella precedente espressione e negli sviluppi che seguono si assume la seguente notazione:

- h identifica lo strato ($h=1, \dots, H$);
- i identifica il comune; n_h è il numero di comuni campione nello strato h (a meno di alcune eccezioni (in alcuni casi si modificano i comuni campione se quelli estratti non contengono un numero sufficiente di famiglie) è sempre pari a 1); N_h è il numero di comuni complessivamente contenuti nello strato h ;
- j identifica la famiglia; m_{hi} è il numero di famiglie campione nel comune i dello strato h ; M_{hi} è il numero di famiglie complessivamente presenti nel comune i dello strato h ;
- k identifica l'individuo; P_{hi} è il numero di individui presenti nel comune i dello strato h ; P_{hij} è il numero di componenti della famiglia j nel comune i dello strato h ;
- $z_{hij} = y_{hij} - \mathbf{x}'_{hij} \hat{\beta}$ è la variabile linearizzata in cui $\mathbf{x}_{hij} = (x_{1hij}, \dots, x_{khij}, \dots, x_{Khij})'$ è il vettore contenente i valori delle K variabili ausiliarie osservate per la famiglia j , $\hat{\beta}$ è il vettore dei coefficienti di regressione del modello lineare che lega la variabile di interesse Y alle K variabili ausiliarie X e y_{hij} è il valore della variabile Y osservato nella famiglia j ;
- w_{hij} è il peso di riporto all'universo associato alla famiglia j .

Prima di descrivere la procedura di stima della varianza campionaria delle stime trimestrali, di media annua e delle differenze tra trimestri dell'indagine sulle forze di lavoro, è opportuno richiamare sinteticamente alcune caratteristiche del disegno campionario. Come è stato ampiamente descritto nel capitolo 4, il disegno campionario dell'indagine sulle forze di lavoro è un disegno che prevede due stadi di selezione: i comuni rappresentano le unità di primo stadio e le famiglie sono le unità di secondo stadio. Il disegno prevede una stratificazione al primo stadio dei comuni in strati omogenei in base all'ampiezza demografica; gli strati si distinguono in strati auto-rappresentativi (Ar), costituiti da un solo comune che viene estratto nel campione con probabilità pari a uno, e strati non auto-rappresentativi (Nar), contenenti un certo numero di comuni, tra i quali ne viene estratto uno. Da ciascun comune campione viene estratto un campione casuale semplice di famiglie. Le famiglie estratte vengono intervistate quattro volte in un periodo di 15 mesi e seguono uno schema di rotazione del tipo 2-2-2 (ciascuna famiglia viene intervistata per due trimestri consecutivi, dopo i quali esce temporaneamente dal campione per due trimestri, per poi rientrarvi per due ulteriori trimestri, dopo i quali esce definitivamente dal campione); applicando questo schema di rotazione si ha una sovrapposizione del 50 per cento del campione tra due trimestri consecutivi e del 50 per cento tra lo stesso trimestre in due anni successivi.

³⁶ Per ulteriori dettagli si veda: Woodruff (1971) o Istat (2005).

7.4.2 - Stima della varianza campionaria di una stima trimestrale di un totale

Sia Y_d il totale trimestrale della variabile Y nel dominio di stima d costituito dall'unione di più strati e sia \hat{Y}_d il corrispondente stimatore di ponderazione vincolata:

$$Y_d = \sum_{h=1}^{H_d} \sum_{i=1}^{N_h} \sum_{j=1}^{M_{hi}} \sum_{k=1}^{P_{hij}} Y_{hijk}$$

$$\hat{Y}_d = \sum_{h=1}^{H_d} \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} \sum_{k=1}^{P_{hij}} y_{hijk} w_{hij}.$$

Poiché l'estrazione del campione avviene in modo indipendente nei vari strati, la stima della varianza campionaria di \hat{Y}_d si ottiene sommando le stime della varianza calcolate in ciascuno strato appartenente al dominio d :

$$\hat{var}(\hat{Y}_d) \cong \hat{var}(\hat{Z}_d) = \sum_{h=1}^{H_d} \hat{var}(\hat{Z}_h).$$

La stima della varianza viene effettuata in modo differente negli strati Ar e negli strati Nar.

Negli strati Ar la varianza dipende dalla variabilità tra le famiglie nello strato ed è stimata mediante la seguente espressione (si omette per semplicità l'indice i del comune, che negli strati AR coincide con lo strato)³⁷:

$$\sum_{h=1}^{H_{AR}} \hat{var}(\hat{Z}_h) = \sum_{h=1}^{H_{AR}} M_h \frac{(M_h - m_h)}{m_h(m_h - 1)} \sum_{j=1}^{m_h} (z_{hj} - \bar{z}_h)^2$$

dove $\bar{z}_h = \frac{1}{m_h} \sum_{j=1}^{m_h} z_{hj}$.

Negli strati Nar invece la varianza dipende essenzialmente dalla variabilità tra i comuni nello strato³⁸. Poiché l'attuale disegno campionario prevede l'estrazione di un solo comune campione da ciascuno strato ($n_h=1$), il calcolo della varianza campionaria non può essere effettuato all'interno del singolo strato Nar; il calcolo avviene pertanto aggregando coppie di strati ("simili" all'interno della provincia in termini di ampiezza demografica dei comuni che contengono) e viene effettuato mediante la seguente espressione (dove H_{NAR} identifica ora l'unione di due strati Nar):

$$\sum_{h=1}^{H_{NAR}} \hat{var}(\hat{Z}_h) = \sum_{h=1}^{H_{NAR}} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\hat{z}_{hi} - \frac{\hat{z}_h}{n_h} \right)^2$$

³⁷ Negli strati Ar il campione è equivalente ad un campione stratificato ad uno stadio, poiché viene meno il primo stadio di selezione, quello dei comuni campione.

³⁸ Lo stimatore riportato di seguito per la varianza negli strati Nar presenta una distorsione positiva; l'uso di questa espressione è giustificato dalla difficoltà di calcolo delle probabilità di inclusione del secondo ordine delle unità di primo stadio richieste per definire uno stimatore approssimativamente corretto (Istat, 2005).

$$\text{in cui si ha } \hat{z}_{hi} = \sum_{j=1}^{m_{hi}} z_{hij} w_{hij} \text{ e } \hat{z}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} z_{hij} w_{hij} .$$

Possiamo quindi riscrivere l'espressione dello stimatore della varianza campionaria come segue:

$$\hat{V}ar(\hat{Y}_d) \cong \hat{V}ar(\hat{Z}_d) = \sum_{h=1}^{H_{AR}} \hat{V}ar(\hat{Z}_h) + \sum_{h=1}^{H_{NAR}} \hat{V}ar(\hat{Z}_h)$$

in cui H_{AR} e H_{NAR} indicano rispettivamente il numero di strati Ar e Nar (raggruppati a due a due) in cui è suddiviso il dominio d .

7.4.3 - Stima della varianza campionaria di una stima di media annua di un totale

La stima di media annua del parametro ${}_a Y_d$ (il totale della variabile Y nel dominio di stima d riferito ad un certo anno a) viene calcolata effettuando la media aritmetica semplice delle quattro stime trimestrali ${}_t \hat{Y}_d$:

$${}_a \hat{Y}_d = \frac{1}{4} \sum_{t=1}^4 {}_t \hat{Y}_d .$$

Nel calcolare la stima della varianza campionaria di una stima di media annua è necessario tenere conto della correlazione esistente tra le stime trimestrali (dovuta alla parziale sovrapposizione del campione tra i quattro trimestri dell'anno). Pertanto, la stima della varianza di una stima di media annua è data da (si omette per semplicità l'indicazione del dominio di riferimento d):

$$\hat{V}ar({}_a \hat{Y}) = \hat{V}ar\left(\frac{1}{4} \sum_{t=1}^4 {}_t \hat{Y}\right) = \frac{1}{16} \left(\sum_{t=1}^4 \hat{V}ar({}_t \hat{Y}) + \sum_{t \neq t'} \hat{C}ov({}_t \hat{Y}, {}_{t'} \hat{Y}) \right)$$

in cui ${}_t \hat{Y}$ e ${}_{t'} \hat{Y}$ sono le stime trimestrali riferite a due generici trimestri t e t' , e $\hat{C}ov({}_t \hat{Y}, {}_{t'} \hat{Y})$ è la stima della covarianza tra ${}_t \hat{Y}$ e ${}_{t'} \hat{Y}$.

Tenendo conto dello schema di rotazione delle famiglie adottato dall'indagine sulle forze di lavoro e sostituendo $\hat{C}ov({}_t \hat{Y}, {}_{t'} \hat{Y}) = \hat{\rho}_{t,t'} \cdot \sqrt{\hat{V}ar({}_t \hat{Y}) \cdot \hat{V}ar({}_{t'} \hat{Y})}$ si ottiene³⁹:

$$\hat{V}ar({}_a \hat{Y}) = \frac{1}{16^2} \left\{ 16 \sum_{t=1}^4 \hat{V}ar({}_t \hat{Y}) + 2 \left[8 \left(\hat{\rho}_{1,2} \sqrt{\hat{V}ar({}_1 \hat{Y}) \hat{V}ar({}_2 \hat{Y})} + \hat{\rho}_{2,3} \sqrt{\hat{V}ar({}_2 \hat{Y}) \hat{V}ar({}_3 \hat{Y})} + \hat{\rho}_{3,4} \sqrt{\hat{V}ar({}_3 \hat{Y}) \hat{V}ar({}_4 \hat{Y})} \right) + 4 \hat{\rho}_{1,4} \sqrt{\hat{V}ar({}_1 \hat{Y}) \hat{V}ar({}_4 \hat{Y})} \right] \right\} .$$

³⁹ Per ulteriori dettagli di veda: De Vitiis, Di Consiglio, Falorsi (2005).

Per stimare la precedente espressione è necessario disporre, per ogni variabile di interesse Y , delle stime delle varianze delle stime trimestrali per i quattro trimestri dell'anno e delle stime dei coefficienti di autocorrelazione tra le stime riferite a trimestri successivi e tra il primo e il quarto trimestre. Per procedere più agevolmente, si introducono le seguenti ipotesi semplificatrici:

- si assume che la struttura di autocovarianza sia costante nell'anno: si indicano con $\hat{\rho}_1$ e $\hat{\rho}_3$ le stime dei coefficienti di correlazione rispettivamente ad un trimestre e a tre trimestri di distanza. I coefficienti $\hat{\rho}_1$ e $\hat{\rho}_3$ vengono stimati ogni anno distintamente per tre gruppi di variabili, quelle relative agli occupati, alle persone in cerca di occupazione e alle non forze di lavoro, a livello regionale;
- si assume che le varianze campionarie $Var({}_t\hat{Y})$ e i totali nella popolazione ${}_tY$ siano pressoché costanti nei quattro trimestri dell'anno: si pone quindi ${}_t\hat{Y} = \hat{Y}$, $\hat{Var}({}_t\hat{Y}) = \hat{Var}(\hat{Y})$, per $t=1, \dots, 4$.

Con semplici sostituzioni, è possibile riscrivere l'espressione precedente nella forma semplificata:

$$\hat{var}_a(\hat{Y}) = \frac{1}{4} \hat{var}(\hat{Y}) \cdot \left(1 + \frac{3}{4} \hat{\rho}_1 + \frac{1}{8} \hat{\rho}_3 \right) = \hat{var}(\hat{Y}) \cdot \text{effrot}_a.$$

In questo modo la varianza campionaria di una stima di media annua viene stimata moltiplicando una stima della varianza trimestrale media nei quattro trimestri dell'anno per un coefficiente, detto effetto rotazione (effrot_a), che dipende dalle correlazioni tra le stime rispettivamente ad un trimestre e a tre trimestri di distanza.

Nel prospetto seguente sono riportati gli effetti rotazione stimati per il 2004.

Prospetto 7.5 – Effetti rotazione stimati per occupati, persone in cerca di occupazione e non forze di lavoro, per regione – Anno 2004

REGIONE	Effetti rotazione		
	Occupati	Persone in cerca di occupazione	Non forze lavoro
Piemonte	0,673	0,581	0,672
Valle d'Aosta	0,672	0,560	0,670
Lombardia	0,673	0,575	0,673
Trentino Alto-Adige	0,668	0,555	0,668
Veneto	0,671	0,560	0,671
Friuli-Venezia Giulia	0,673	0,573	0,673
Liguria	0,673	0,576	0,671
Emilia-Romagna	0,672	0,573	0,672
Toscana	0,670	0,563	0,668
Umbria	0,671	0,569	0,670
Marche	0,671	0,588	0,671
Lazio	0,670	0,575	0,667
Abruzzo	0,668	0,569	0,663
Molise	0,667	0,585	0,664
Campania	0,662	0,582	0,656
Puglia	0,662	0,590	0,659
Basilicata	0,663	0,580	0,656
Calabria	0,663	0,581	0,656
Sicilia	0,664	0,582	0,655
Sardegna	0,664	0,593	0,661
Italia	0,669	0,580	0,666

7.4.4 - Stima della varianza campionaria di una stima di una differenza trimestrale

In modo analogo a quanto precedente descritto per la stima di una media annua, anche per la stima della varianza campionaria di una stima di una differenza trimestrale è necessario tenere conto della correlazione esistente tra le stime riferite a due successivi trimestri (tra i quali si realizza una sovrapposizione del 50 per cento del campione).

Indicando con ${}_t\Delta_1 = Y_t - Y_{t-1}$ il parametro da stimare, la varianza campionaria dello stimatore ad essa associato assume la seguente espressione:

$$Var({}_t\hat{\Delta}_1) = Var({}_t\hat{Y}) + Var({}_{t-1}\hat{Y}) - 2 Cov({}_t\hat{Y}, {}_{t-1}\hat{Y}).$$

Si introduce per semplicità l'ipotesi di omoschedasticità:

$$Var({}_t\hat{Y}) = Var({}_{t-1}\hat{Y}) = Var({}_\bullet\hat{Y});$$

sostituendo nella precedente espressione

$$Cov({}_t\hat{Y}, {}_{t-1}\hat{Y}) = \rho_1 Var({}_\bullet\hat{Y})$$

e tenendo conto che la correlazione è nulla su 50% del campione che non si sovrappone, si ha:

$$\hat{V}ar(\hat{\Delta}_1) = 2 \hat{V}ar(\hat{Y}) - 2 \frac{1}{2} \hat{\rho}_1 \hat{V}ar(\hat{Y}) = \hat{V}ar(\hat{Y})(2 - \hat{\rho}_1).$$

In questo modo la varianza campionaria di una stima di una differenza trimestrale viene stimata moltiplicando la stima della varianza trimestrale media per un coefficiente che dipende dalla stima della correlazione tra le due stime, calcolata sulla parte di campione che si sovrappone nei due trimestri.

7.4.5 - Stima della varianza campionaria della stima di altri parametri della popolazione

La metodologia e gli sviluppi fin qui descritti con riferimento alla stima della varianza campionaria di una stima del totale di una variabile quantitativa restano analogamente validi anche per la stima di una frequenza assoluta di una variabile qualitativa (che corrisponde al totale della variabile espressa in forma dicotomica) oltre che per la stima di una frequenza relativa (media della variabile espressa in forma dicotomica) riferita all'intera popolazione, o a un dominio pianificato, o anche ad un sottogruppo della popolazione il cui totale sia noto e sia stato vincolato nella procedura di calcolo dei pesi di riporto all'universo.

Se, invece, si vuole stimare la varianza di una stima di una frequenza relativa riferita ad una sottopopolazione differente (ad esempio la popolazione di coloro che presentano una certa modalità di una variabile di interesse) è necessario ricorrere ad un'approssimazione. Infatti la stima di una frequenza relativa riferita ad un sottogruppo della popolazione, è ottenibile come rapporto tra due quantità entrambe stimate:

$$\hat{R}_d = \frac{\hat{N}_d}{\hat{D}_d}.$$

Una valutazione approssimata dell'errore campionario relativo di \hat{R}_d si può calcolare come:⁴⁰

$$\hat{\varepsilon}(\hat{R}_d) = \sqrt{\hat{\varepsilon}^2(\hat{N}_d) - \hat{\varepsilon}^2(\hat{D}_d)}.$$

7.4.6 - Presentazione sintetica delle stime degli errori campionari

Ad ogni stima \hat{Y}_d corrisponde una stima dell'errore campionario relativo $\hat{\varepsilon}(\hat{Y}_d)$ che consente di valutarne la precisione; pertanto, per consentire una corretta interpretazione delle stime prodotte, sarebbe necessario presentare contestualmente a ciascuna stima anche il corrispondente errore campionario stimato. Ciò, tuttavia, non è possibile quando le stime prodotte sono in numero molto elevato (quale è il caso della produzione corrente di stime

⁴⁰ L'approssimazione è valida sotto l'ipotesi di incorrelazione tra \hat{R}_d e \hat{D}_d .

trimestrali e di media annua dell'indagine sulle forze di lavoro), sia per il consistente carico di lavoro che ne deriverebbe, sia perché le tavole pubblicate risulterebbero complesse e di non facile consultazione. Inoltre, procedendo in questo modo, non sarebbero comunque disponibili le stime degli errori campionari per le stime non pubblicate dall'Istat, che l'utente può ricavare in modo autonomo, effettuando elaborazioni sui dati elementari.

Per questi motivi si ricorre frequentemente ad una presentazione sintetica delle stime degli errori campionari, basata sul metodo dei modelli regressivi. Questo metodo si basa sulla determinazione di una semplice funzione matematica che mette in relazione ciascuna stima con il proprio errore campionario relativo stimato.

Il modello utilizzato per le stime di frequenze assolute e relative riferite agli individui è del tipo seguente⁴¹:

$$\log(\hat{\varepsilon}^2(\hat{Y}_d)) = a + b \log(\hat{Y}_d)$$

dove i parametri a e b vengono stimati secondo il metodo dei minimi quadrati.

Nei prospetti 7.6, 7.7 e 7.8 sono riportati i valori dei coefficienti a e b e dell'indice di determinazione R^2 dei modelli stimati per l'interpolazione degli errori campionari relativi di stime trimestrali e di media annua di frequenze assolute e relative, rispettivamente per il totale Italia, ripartizioni geografiche, regioni e province (solo per le stime di media annua).

Utilizzando gli opportuni coefficienti è possibile calcolare una stima dell'errore campionario relativo di una determinata stima di una frequenza assoluta \hat{Y}_d applicando la seguente formula:

$$\hat{\varepsilon}(\hat{Y}_d) = \sqrt{\exp(a + b \log(\hat{Y}_d))}.$$

Infine, i prospetti 7.9, 7.10 e 7.11 hanno lo scopo di rendere più agevole e immediata la valutazione degli errori campionari. Le tabelle hanno la seguente struttura:

- in testata sono elencati valori crescenti di stima (1.000, 5.000, 10.000, ...) di frequenze assolute;
- in fiancata sono riportati i domini di riferimento delle stime: rispettivamente il totale Italia e le ripartizioni, le regioni e le province;
- le celle interne contengono gli errori campionari relativi percentuali stimati mediante la formula precedente.

Consultando queste tavole è possibile disporre di una valutazione immediata (anche se meno precisa rispetto all'applicazione della formula precedente), dell'errore campionario di una generica stima di una frequenza assoluta (o relativa), cercando nella testata il valore che più si avvicina alla stima di interesse e in fiancata il dominio di riferimento.

⁴¹ Le stime di frequenze assolute e relative di variabili qualitative rappresentano la quasi totalità delle stime prodotte dall'indagine sulle forze di lavoro. Per poche variabili quantitative rilevate dall'indagine (quali ad esempio le ore lavorate) viene applicato un diverso modello per la stima degli errori campionari.

Prospetto 7.6 – Valori dei coefficienti a, b e R² delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime trimestrali e di media annua per ripartizione geografica – I trimestre 2005 e Anno 2004

RIPARTIZIONI GEOGRAFICHE	Stime trimestrali: I trimestre 2005			Stime di media annua: Anno 2004		
	a	b	R ²	a	b	R ²
Italia	7,090376	-1,093615	96,77	6,229296	-1,093166	97,40
Settentrione	7,279997	-1,115075	96,18	6,123194	-1,090509	96,38
<i>Nord-Ovest</i>	7,369694	-1,127757	96,13	6,209234	-1,100126	96,37
<i>Nord-Est</i>	6,834968	-1,087693	94,98	6,091737	-1,098788	95,85
Centro	7,254659	-1,107285	96,30	6,624524	-1,125174	96,66
Mezzogiorno	5,999308	-1,022109	96,48	5,784840	-1,076111	97,55

Prospetto 7.7 – Valori dei coefficienti a, b e R² delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime trimestrali e di media annua per regione – I trimestre 2005 e Anno 2004

REGIONI	Stime trimestrali: I trimestre 2005			Stime di media annua: Anno 2004		
	a	b	R ²	a	b	R ²
Piemonte	7,244358	-1,146728	95,66	5,815637	-1,091773	95,43
Valle d'Aosta	4,613139	-1,176841	94,68	3,562220	-1,142107	95,58
Lombardia	7,251426	-1,113728	95,22	6,602257	-1,129564	96,47
Trentino-Alto Adige	5,460679	-1,131530	95,16	4,324546	-1,102727	95,99
Veneto	7,178152	-1,111419	94,68	6,230127	-1,102631	95,43
Friuli-Venezia Giulia	6,618122	-1,124697	93,44	5,525993	-1,095743	95,13
Liguria	7,109753	-1,162943	95,46	5,918286	-1,121329	95,73
Emilia-Romagna	6,389271	-1,052536	93,17	6,024737	-1,101009	95,16
Toscana	6,753427	-1,086844	93,63	6,330792	-1,122462	95,64
Umbria	6,787924	-1,169301	93,64	5,478888	-1,121161	95,61
Marche	7,023267	-1,137685	93,51	6,100002	-1,132804	95,30
Lazio	7,094630	-1,081980	95,03	6,650507	-1,117732	96,45
Abruzzo	6,694334	-1,124438	93,66	5,751732	-1,114243	95,19
Molise	5,264734	-1,169658	95,64	3,947684	-1,105994	96,71
Campania	6,592351	-1,058255	96,44	5,742372	-1,061462	96,91
Puglia	6,231055	-1,043600	96,10	5,785077	-1,080675	96,61
Basilicata	5,443596	-1,112175	95,54	4,413726	-1,087645	96,58
Calabria	5,472514	-1,020103	94,13	4,863090	-1,043123	96,52
Sicilia	5,937216	-1,028514	95,97	5,421745	-1,055974	96,31
Sardegna	6,142338	-1,038193	92,47	5,480101	-1,072391	95,77

Prospetto 7.8 – Valori dei coefficienti a, b e R² delle funzioni utilizzate per l'interpolazione degli errori campionari delle stime di media annua per provincia – Anno 2004

PROVINCE	a	b	R ²	PROVINCE	a	b	R ²
Torino	6,490849	-1,145593	96,53	Perugia	5,462589	-1,137227	95,65
Vercelli	4,991248	-1,176478	94,67	Terni	5,707110	-1,141036	91,79
Novara	5,363369	-1,149876	94,69	Pesaro e Urbino	6,275741	-1,184072	94,55
Cuneo	6,032509	-1,128380	94,22	Ancona	6,030689	-1,149460	93,17
Asti	4,826930	-1,160597	95,26	Macerata	5,708624	-1,163788	94,01
Alessandria	6,433572	-1,165453	93,19	Ascoli Piceno	5,904164	-1,105308	92,50
Biella	4,767291	-1,179218	94,96	Viterbo	6,521190	-1,184660	86,99
Verbania	4,520957	-1,128587	94,28	Rieti	4,206765	-1,130871	94,85
Aosta	3,562220	-1,142107	95,58	Roma	6,875403	-1,129123	96,47
Varese	7,044568	-1,180326	94,60	Latina	5,889955	-1,103476	93,64
Como	6,154304	-1,172093	94,79	Frosinone	5,653064	-1,111221	94,13
Sondrio	5,007455	-1,153579	92,91	Aquila	5,223886	-1,097129	94,36
Milano	6,874581	-1,145363	96,29	Teramo	5,323087	-1,141270	94,21
Bergamo	6,368762	-1,129190	95,60	Pescara	6,338930	-1,166902	92,95
Brescia	6,338454	-1,137546	94,42	Chieti	6,052470	-1,161224	93,32
Pavia	5,798784	-1,145964	94,80	Campobasso	3,929492	-1,107273	96,29
Cremona	5,437044	-1,165335	94,84	Isernia	4,123374	-1,161011	94,43
Mantova	5,095781	-1,164671	95,65	Caserta	5,659402	-1,104467	94,85
Lecco	4,839797	-1,158895	95,59	Benevento	5,157734	-1,121754	95,15
Lodi	7,124399	-1,286843	90,45	Napoli	5,691644	-1,051142	96,50
Bolzano	4,644783	-1,119444	95,54	Avellino	6,413452	-1,183026	93,53
Trento	4,156592	-1,134776	96,32	Salerno	5,883229	-1,079693	94,61
Verona	6,363704	-1,167512	95,01	Foggia	5,741146	-1,114575	94,76
Vicenza	6,426511	-1,162196	94,18	Bari	6,012609	-1,097599	95,47
Belluno	4,993020	-1,157408	94,19	Taranto	6,233067	-1,121195	94,50
Treviso	6,893530	-1,165533	93,33	Brindisi	5,623122	-1,105643	94,73
Venezia	6,218257	-1,113275	94,70	Lecce	5,407681	-1,107040	96,53
Padova	6,594391	-1,146179	94,39	Potenza	4,367466	-1,079840	96,31
Rovigo	5,756851	-1,135304	92,28	Matera	4,248483	-1,094244	95,91
Udine	5,911727	-1,119843	93,92	Cosenza	5,321347	-1,080692	94,73
Gorizia	4,811929	-1,125229	94,13	Catanzaro	5,161095	-1,120644	95,85
Trieste	5,685429	-1,130339	94,34	Reggio Calabria	5,336793	-1,084172	95,10
Pordenone	5,512650	-1,169273	94,59	Crotone	4,708338	-1,098524	95,52
Imperia	5,831078	-1,136361	91,76	Vibo Valentia	3,481708	-1,066827	96,55
Savona	5,813018	-1,119681	92,94	Trapani	5,100984	-1,082588	95,72
Genova	5,836831	-1,147774	95,76	Palermo	5,519664	-1,084078	96,11
La Spezia	6,629500	-1,191168	93,00	Messina	5,469004	-1,098306	95,18
Piacenza	5,869587	-1,206401	93,76	Agrigento	6,608057	-1,151183	93,08
Parma	5,122901	-1,172885	95,61	Caltanissetta	5,849945	-1,138657	92,24
Reggio Emilia	6,088495	-1,179766	94,40	Enna	4,477531	-1,106996	96,62
Modena	6,421597	-1,153354	94,28	Catania	4,922921	-1,054152	96,11
Bologna	6,847927	-1,157705	93,55	Ragusa	7,252989	-1,219536	88,52
Ferrara	6,934555	-1,207680	92,55	Siracusa	5,402504	-1,058526	95,39
Ravenna	6,220282	-1,176684	92,17	Sassari	5,876045	-1,104002	94,09
Forlì	5,948176	-1,134659	91,12	Nuoro	5,146569	-1,114513	94,11
Rimini	5,570087	-1,142050	94,15	Cagliari	5,931895	-1,124316	94,80
Massa Carrara	6,410165	-1,227352	94,73	Oristano	5,173018	-1,141756	91,28
Lucca	6,829712	-1,175924	89,00				
Pistoia	5,738128	-1,121508	93,04				
Firenze	6,766529	-1,149425	95,00				
Livorno	6,239514	-1,127432	93,83				
Pisa	5,711427	-1,098058	92,92				
Arezzo	4,854224	-1,142559	95,40				
Siena	5,118375	-1,141793	93,47				
Grosseto	5,339083	-1,130150	93,14				
Prato	6,301404	-1,186605	93,80				

Prospetto 7.9 – Valori interpolati degli errori relativi percentuali delle stime trimestrali e di media annua per ripartizione geografica – I trimestre 2005 e Anno 2004

RIPARTIZIONI GEOGRAFICHE	Stime trimestrali: I trimestre 2005										
	5.000	7.500	10.000	25.000	50.000	100.000	250.000	500.000	750.000	1.000.000	5.000.000
Italia	32,9	26,3	22,5	13,6	9,3	6,4	3,9	2,7	2,1	1,8	0,8
Settentrione	33,0	26,3	22,4	13,5	9,1	6,2	3,7	2,5	2,0	1,7	0,7
<i>Nord-ovest</i>	32,7	26,0	22,1	13,2	8,9	6,0	3,6	2,4	1,9	1,6	0,7
<i>Nord-est</i>	29,7	23,8	20,4	12,4	8,5	5,8	3,5	2,4	1,9	1,7	0,7
Centro	33,7	26,9	22,9	13,8	9,4	6,4	3,9	2,6	2,1	1,8	0,7
Mezzogiorno	25,8	21,0	18,1	11,4	8,0	5,6	3,5	2,5	2,0	1,7	0,8

RIPARTIZIONI GEOGRAFICHE	Stime di media annua: Anno 2004										
	5.000	7.500	10.000	25.000	50.000	100.000	250.000	500.000	750.000	1.000.000	5.000.000
Italia	21,4	17,2	14,7	8,9	6,1	4,2	2,5	1,7	1,4	1,2	0,5
Settentrione	20,5	16,5	14,1	8,5	5,9	4,0	2,4	1,7	1,3	1,1	0,5
<i>Nord-ovest</i>	20,6	16,5	14,1	8,5	5,8	4,0	2,4	1,6	1,3	1,1	0,5
<i>Nord-est</i>	19,5	15,6	13,3	8,1	5,5	3,8	2,3	1,6	1,2	1,1	0,4
Centro	22,8	18,1	15,4	9,2	6,2	4,2	2,5	1,7	1,4	1,2	0,5
Mezzogiorno	18,4	14,8	12,7	7,8	5,3	3,7	2,2	1,5	1,2	1,1	0,4

Prospetto 7.10 – Valori interpolati degli errori relativi percentuali delle stime trimestrali e di media annua per regione – I trimestre 2005 e Anno 2004

REGIONI	Stime trimestrali: I trimestre 2005										
	1.000	5.000	10.000	25.000	50.000	75.000	100.000	250.000	500.000	750.000	1.000.000
Piemonte	71,3	28,3	19,0	11,3	7,6	6,0	5,1	3,0	2,0	1,6	1,4
Valle d'Aosta	17,2	6,7	4,4	2,6	1,7	1,4	1,1	0,7	0,4	0,4	0,3
Lombardia	80,2	32,7	22,2	13,4	9,1	7,2	6,2	3,7	2,5	2,0	1,7
Trentino-Alto Adige	30,8	12,4	8,4	5,0	3,4	2,7	2,3	1,4	0,9	0,7	0,6
Veneto	77,9	31,9	21,7	13,0	8,9	7,1	6,0	3,6	2,5	2,0	1,7
Friuli-Venezia Giulia	56,2	22,8	15,4	9,2	6,2	5,0	4,2	2,5	1,7	1,4	1,2
Liguria	63,0	24,7	16,5	9,7	6,5	5,1	4,3	2,5	1,7	1,3	1,1
Emilia-Romagna	64,4	27,6	19,2	11,8	8,2	6,6	5,7	3,5	2,4	2,0	1,7
Toscana	68,6	28,6	19,6	11,9	8,2	6,6	5,6	3,4	2,3	1,9	1,6
Umbria	52,5	20,5	13,7	8,0	5,3	4,2	3,6	2,1	1,4	1,1	0,9
Marche	65,9	26,4	17,8	10,6	7,1	5,6	4,8	2,8	1,9	1,5	1,3
Lazio	82,7	34,6	23,8	14,5	10,0	8,0	6,8	4,2	2,9	2,3	2,0
Abruzzo	58,5	23,7	16,0	9,6	6,5	5,2	4,4	2,6	1,8	1,4	1,2
Molise	24,5	9,5	6,4	3,7	2,5	2,0	1,7	1,0	0,6	0,5	0,4
Campania	69,8	29,8	20,7	12,7	8,8	7,1	6,1	3,8	2,6	2,1	1,8
Puglia	61,3	26,5	18,4	11,4	8,0	6,4	5,5	3,4	2,4	1,9	1,7
Basilicata	32,6	13,3	9,1	5,5	3,7	3,0	2,5	1,5	1,0	0,8	0,7
Calabria	45,5	20,0	14,1	8,8	6,2	5,0	4,3	2,7	1,9	1,6	1,3
Sicilia	55,8	24,4	17,1	10,7	7,5	6,1	5,2	3,3	2,3	1,9	1,6
Sardegna	59,8	25,9	18,1	11,2	7,8	6,4	5,5	3,4	2,4	1,9	1,7

REGIONI	Stime di media annua: Anno 2004										
	1.000	5.000	10.000	25.000	50.000	75.000	100.000	250.000	500.000	750.000	1.000.000
Piemonte	42,2	17,5	12,0	7,3	5,0	4,0	3,4	2,1	1,4	1,1	1,0
Valle d'Aosta	11,5	4,6	3,1	1,8	1,2	1,0	0,8	0,5	0,3	0,3	0,2
Lombardia	54,9	22,1	14,9	8,9	6,0	4,8	4,1	2,4	1,6	1,3	1,1
Trentino- Alto Adige	19,3	7,9	5,4	3,3	2,2	1,8	1,5	0,9	0,6	0,5	0,4
Veneto	50,0	20,6	14,0	8,5	5,8	4,6	3,9	2,4	1,6	1,3	1,1
Friuli-Venezia Giulia	36,0	14,9	10,2	6,2	4,2	3,4	2,9	1,7	1,2	1,0	0,8
Liguria	40,1	16,3	11,0	6,6	4,5	3,6	3,0	1,8	1,2	1,0	0,8
Emilia- Romagna	45,4	18,7	12,8	7,7	5,3	4,2	3,6	2,2	1,5	1,2	1,0
Toscana	49,1	19,9	13,5	8,1	5,5	4,4	3,7	2,2	1,5	1,2	1,0
Umbria	32,2	13,1	8,9	5,3	3,6	2,9	2,4	1,5	1,0	0,8	0,7
Marche	42,2	17,0	11,5	6,8	4,6	3,7	3,1	1,9	1,2	1,0	0,8
Lazio	58,6	23,8	16,2	9,7	6,6	5,2	4,5	2,7	1,8	1,4	1,2
Abruzzo	37,8	15,4	10,5	6,3	4,3	3,4	2,9	1,7	1,2	0,9	0,8
Molise	15,8	6,5	4,4	2,7	1,8	1,4	1,2	0,7	0,5	0,4	0,3
Campania	45,2	19,2	13,3	8,2	5,7	4,6	3,9	2,4	1,7	1,3	1,2
Puglia	43,2	18,1	12,4	7,6	5,2	4,2	3,6	2,2	1,5	1,2	1,0
Basilicata	21,2	8,8	6,1	3,7	2,5	2,0	1,7	1,1	0,7	0,6	0,5
Calabria	31,0	13,4	9,3	5,8	4,0	3,3	2,8	1,7	1,2	1,0	0,8
Sicilia	39,2	16,8	11,6	7,2	5,0	4,0	3,4	2,1	1,5	1,2	1,0
Sardegna	38,1	16,1	11,1	6,8	4,7	3,8	3,2	2,0	1,4	1,1	0,9

Prospetto 7.11 – Valori interpolati degli errori relativi percentuali delle stime di media annua per provincia – Anno 2004

PROVINCE	Stime di media annua: Anno 2004										
	1.000	5.000	10.000	25.000	50.000	75.000	100.000	250.000	500.000	750.000	1.000.000
Torino	49,1	19,5	13,1	7,8	5,2	4,1	3,5	2,1	1,4	1,1	0,9
Vercelli	20,9	8,1	5,4	3,1	2,1	1,6	1,4	0,8	0,5	0,4	0,4
Novara	27,5	10,9	7,3	4,3	2,9	2,3	1,9	1,2	0,8	0,6	0,5
Cuneo	41,4	16,7	11,3	6,7	4,6	3,6	3,1	1,8	1,2	1,0	0,8
Asti	20,3	8,0	5,3	3,1	2,1	1,7	1,4	0,8	0,6	0,4	0,4
Alessandria	44,6	17,4	11,6	6,8	4,6	3,6	3,0	1,8	1,2	0,9	0,8
Biella	18,5	7,1	4,8	2,8	1,8	1,4	1,2	0,7	0,5	0,4	0,3
Verbania	19,4	7,8	5,3	3,2	2,1	1,7	1,4	0,9	0,6	0,5	0,4
Aosta	11,5	4,6	3,1	1,8	1,2	1,0	0,8	0,5	0,3	0,3	0,2
Varese	57,4	22,2	14,8	8,6	5,7	4,5	3,8	2,2	1,5	1,2	1,0
Como	37,9	14,7	9,8	5,7	3,8	3,0	2,5	1,5	1,0	0,8	0,7
Sondrio	22,8	9,0	6,0	3,6	2,4	1,9	1,6	0,9	0,6	0,5	0,4
Milano	59,5	23,7	15,9	9,4	6,3	5,0	4,3	2,5	1,7	1,3	1,1
Bergamo	48,9	19,7	13,3	7,9	5,4	4,3	3,6	2,2	1,5	1,2	1,0
Brescia	46,8	18,7	12,6	7,5	5,1	4,0	3,4	2,0	1,4	1,1	0,9
Pavia	34,7	13,8	9,3	5,5	3,7	2,9	2,5	1,5	1,0	0,8	0,7
Cremona	27,1	10,6	7,1	4,2	2,8	2,2	1,9	1,1	0,7	0,6	0,5
Mantova	22,9	9,0	6,0	3,5	2,3	1,9	1,6	0,9	0,6	0,5	0,4
Lecco	20,5	8,1	5,4	3,2	2,1	1,7	1,4	0,8	0,6	0,4	0,4
Lodi	41,4	14,7	9,4	5,2	3,3	2,6	2,1	1,2	0,8	0,6	0,5
Bolzano	21,4	8,7	5,9	3,5	2,4	1,9	1,6	1,0	0,7	0,5	0,4
Trento	15,9	6,4	4,3	2,6	1,7	1,4	1,2	0,7	0,5	0,4	0,3
Verona	42,7	16,7	11,1	6,5	4,4	3,4	2,9	1,7	1,1	0,9	0,8
Vicenza	44,9	17,6	11,8	6,9	4,6	3,7	3,1	1,8	1,2	1,0	0,8
Belluno	22,3	8,8	5,9	3,5	2,3	1,8	1,6	0,9	0,6	0,5	0,4
Treviso	56,1	21,9	14,7	8,6	5,7	4,5	3,8	2,2	1,5	1,2	1,0
Venezia	47,9	19,6	13,3	8,0	5,4	4,3	3,7	2,2	1,5	1,2	1,0
Padova	51,6	20,5	13,8	8,2	5,5	4,3	3,7	2,2	1,5	1,2	1,0
Rovigo	35,2	14,1	9,5	5,7	3,8	3,0	2,6	1,5	1,0	0,8	0,7
Udine	40,2	16,3	11,1	6,6	4,5	3,6	3,0	1,8	1,2	1,0	0,8
Gorizia	22,8	9,2	6,2	3,7	2,5	2,0	1,7	1,0	0,7	0,5	0,5
Trieste	34,6	13,9	9,4	5,6	3,8	3,0	2,6	1,5	1,0	0,8	0,7
Pordenone	27,7	10,8	7,2	4,2	2,8	2,2	1,9	1,1	0,7	0,6	0,5
Imperia	36,4	14,6	9,9	5,9	3,9	3,1	2,7	1,6	1,1	0,8	0,7
Savona	38,3	15,5	10,5	6,3	4,3	3,4	2,9	1,7	1,2	0,9	0,8
Genova	35,1	14,0	9,4	5,5	3,7	2,9	2,5	1,5	1,0	0,8	0,7
La Spezia	45,0	17,2	11,4	6,6	4,4	3,4	2,9	1,7	1,1	0,9	0,7
Piacenza	29,2	11,0	7,3	4,2	2,8	2,2	1,8	1,0	0,7	0,5	0,5
Parma	22,5	8,8	5,8	3,4	2,3	1,8	1,5	0,9	0,6	0,5	0,4
Reggio Emilia	35,7	13,8	9,2	5,3	3,6	2,8	2,4	1,4	0,9	0,7	0,6
Modena	46,2	18,3	12,2	7,2	4,8	3,8	3,2	1,9	1,3	1,0	0,9
Bologna	56,3	22,2	14,8	8,7	5,8	4,6	3,9	2,3	1,5	1,2	1,0
Ferrara	49,5	18,7	12,3	7,1	4,7	3,6	3,1	1,8	1,2	0,9	0,8
Ravenna	38,5	14,9	9,9	5,8	3,9	3,0	2,6	1,5	1,0	0,8	0,7
Forlì	38,9	15,6	10,5	6,3	4,2	3,4	2,9	1,7	1,1	0,9	0,8
Rimini	31,4	12,5	8,4	5,0	3,4	2,7	2,3	1,3	0,9	0,7	0,6
Massa Carrara	35,6	13,2	8,7	4,9	3,2	2,5	2,1	1,2	0,8	0,6	0,5
Lucca	52,4	20,3	13,5	7,9	5,3	4,1	3,5	2,0	1,4	1,1	0,9
Pistoia	36,6	14,9	10,1	6,0	4,1	3,3	2,8	1,7	1,1	0,9	0,8
Firenze	55,6	22,1	14,8	8,7	5,9	4,7	3,9	2,3	1,6	1,2	1,0
Livorno	46,1	18,6	12,6	7,5	5,1	4,0	3,4	2,1	1,4	1,1	0,9
Pisa	39,2	16,2	11,1	6,7	4,6	3,7	3,1	1,9	1,3	1,0	0,9
Arezzo	21,9	8,7	5,9	3,5	2,3	1,9	1,6	0,9	0,6	0,5	0,4
Siena	25,0	10,0	6,7	4,0	2,7	2,1	1,8	1,1	0,7	0,6	0,5
Grosseto	29,1	11,7	7,9	4,7	3,2	2,5	2,2	1,3	0,9	0,7	0,6
Prato	38,8	14,9	9,9	5,7	3,8	3,0	2,5	1,5	1,0	0,8	0,6

Prospetto 7.11 segue – Valori interpolati degli errori relativi percentuali delle stime di media annua per provincia – Anno 2004

PROVINCE	Stime di media annua: Anno 2004										
	1.000	5.000	10.000	25.000	50.000	75.000	100.000	250.000	500.000	750.000	1.000.000
Perugia	30,2	12,1	8,2	4,8	3,3	2,6	2,2	1,3	0,9	0,7	0,6
Terni	33,7	13,5	9,1	5,4	3,6	2,9	2,4	1,4	1,0	0,8	0,7
Pesaro e Urbino	38,6	14,9	9,9	5,7	3,8	3,0	2,5	1,5	1,0	0,8	0,6
Ancona	38,5	15,3	10,2	6,1	4,1	3,2	2,7	1,6	1,1	0,9	0,7
Macerata	31,2	12,2	8,2	4,8	3,2	2,5	2,1	1,3	0,8	0,7	0,6
Ascoli Piceno	42,1	17,3	11,8	7,1	4,8	3,9	3,3	2,0	1,4	1,1	0,9
Viterbo	43,6	16,8	11,1	6,5	4,3	3,4	2,8	1,7	1,1	0,9	0,7
Rieti	16,5	6,6	4,5	2,7	1,8	1,4	1,2	0,7	0,5	0,4	0,3
Roma	63,0	25,4	17,2	10,2	6,9	5,5	4,7	2,8	1,9	1,5	1,3
Latina	42,1	17,3	11,8	7,1	4,9	3,9	3,3	2,0	1,4	1,1	0,9
Frosinone	36,4	14,9	10,1	6,1	4,1	3,3	2,8	1,7	1,2	0,9	0,8
Aquila	30,8	12,7	8,7	5,3	3,6	2,9	2,5	1,5	1,0	0,8	0,7
Teramo	27,8	11,1	7,5	4,4	3,0	2,4	2,0	1,2	0,8	0,6	0,5
Pescara	42,3	16,5	11	6,5	4,3	3,4	2,9	1,7	1,1	0,9	0,8
Chieti	37,4	14,7	9,8	5,8	3,9	3,0	2,6	1,5	1,0	0,8	0,7
Campobasso	15,6	6,4	4,4	2,6	1,8	1,4	1,2	0,7	0,5	0,4	0,3
Isernia	14,3	5,6	3,7	2,2	1,5	1,2	1	0,6	0,4	0,3	0,3
Caserta	37,3	15,4	10,5	6,3	4,3	3,4	2,9	1,8	1,2	1,0	0,8
Benevento	27,4	11,1	7,5	4,5	3,1	2,4	2,1	1,2	0,8	0,7	0,6
Napoli	45,6	19,6	13,6	8,4	5,8	4,7	4,1	2,5	1,7	1,4	1,2
Avellino	41,5	16,0	10,6	6,2	4,1	3,2	2,7	1,6	1,1	0,8	0,7
Salerno	45,5	19,1	13,1	8,0	5,5	4,4	3,8	2,3	1,6	1,3	1,1
Foggia	37,6	15,3	10,4	6,2	4,2	3,4	2,9	1,7	1,2	0,9	0,8
Bari	45,6	18,9	12,9	7,8	5,3	4,3	3,6	2,2	1,5	1,2	1,0
Taranto	47,0	19,0	12,9	7,7	5,2	4,2	3,6	2,1	1,4	1,1	1,0
Brindisi	36,5	15,0	10,2	6,2	4,2	3,4	2,9	1,7	1,2	0,9	0,8
Lecce	32,6	13,4	9,1	5,5	3,7	3,0	2,6	1,5	1,0	0,8	0,7
Potenza	21,3	8,9	6,1	3,7	2,6	2,1	1,8	1,1	0,7	0,6	0,5
Matera	19,1	7,9	5,4	3,3	2,2	1,8	1,5	0,9	0,6	0,5	0,4
Cosenza	34,2	14,3	9,9	6,0	4,1	3,3	2,8	1,7	1,2	1,0	0,8
Catanzaro	27,5	11,2	7,6	4,5	3,1	2,4	2,1	1,2	0,8	0,7	0,6
Reggio Calabria	34,1	14,2	9,8	6,0	4,1	3,3	2,8	1,7	1,2	0,9	0,8
Crotone	23,7	9,8	6,7	4,0	2,8	2,2	1,9	1,1	0,8	0,6	0,5
Vibo Valentia	14,3	6,1	4,2	2,6	1,8	1,4	1,2	0,8	0,5	0,4	0,4
Trapani	30,5	12,7	8,8	5,3	3,7	2,9	2,5	1,5	1,1	0,8	0,7
Palermo	37,4	15,6	10,7	6,5	4,5	3,6	3,1	1,9	1,3	1,0	0,9
Messina	34,7	14,3	9,8	5,9	4,0	3,2	2,8	1,7	1,1	0,9	0,8
Agrigento	51,1	20,2	13,6	8,0	5,4	4,3	3,6	2,1	1,4	1,1	1,0
Caltanissetta	36,5	14,6	9,8	5,8	3,9	3,1	2,7	1,6	1,1	0,8	0,7
Enna	20,5	8,4	5,7	3,5	2,4	1,9	1,6	1,0	0,7	0,5	0,4
Catania	30,7	13,2	9,1	5,6	3,9	3,2	2,7	1,7	1,2	0,9	0,8
Ragusa	55,7	20,9	13,7	7,8	5,1	4,0	3,4	1,9	1,3	1,0	0,8
Siracusa	38,5	16,4	11,4	7,0	4,9	3,9	3,4	2,1	1,4	1,2	1,0
Sassari	41,7	17,1	11,7	7,1	4,8	3,8	3,3	2	1,3	1,1	0,9
Nuoro	27,9	11,4	7,7	4,6	3,2	2,5	2,1	1,3	0,9	0,7	0,6
Cagliari	40,0	16,2	11,0	6,5	4,4	3,5	3,0	1,8	1,2	1,0	0,8
Oristano	25,7	10,3	6,9	4,1	2,8	2,2	1,9	1,1	0,7	0,6	0,5

Capitolo 8 - La ricostruzione delle serie storiche dei principali indicatori del mercato del lavoro

8.1 - Premessa

Contestualmente alla pubblicazione dei dati si è diffusa nel 2004 una ricostruzione dei principali indicatori del mercato del lavoro per ottenere serie temporalmente coerenti, a partire dal quarto trimestre del 1992. La ricostruzione all'indietro di una serie è una stima della serie stessa coerente con i risultati ottenuti con una nuova metodologia di rilevazione diversa da quella utilizzata nel passato.

Tre ordini di motivi rendono questo raccordo importante per gli utenti:

- per eseguire una corretta analisi di lungo periodo del mercato del lavoro servono serie storiche per i principali indicatori coerenti temporalmente;
- sia per una corretta analisi tendenziale sia per quella congiunturale il confronto deve essere effettuato su dati coerenti con la nuova metodologia, inoltre i dati devono essere disponibili con tempestività dato il valore di immediatezza di tale tipo di analisi;
- da qualche anno vengono pubblicate ogni trimestre, per alcuni indicatori chiave, le serie di dati destagionalizzati. Queste serie sono in realtà stime che utilizzano come dati di input le serie dei dati ufficiali non destagionalizzate. Se, cambiando la metodologia dell'indagine, non si avessero avute immediatamente a disposizione le serie raccordate con il passato, non sarebbe stato possibile stimare gli effetti stagionali finché le serie "nuove" non fossero state lunghe abbastanza da rendere queste stime significative. In base alla passata esperienza questa lunghezza minima non avrebbe potuto essere inferiore ai sei anni.

Il metodo presentato in questo capitolo ha consentito la pubblicazione delle serie ricostruite dei principali indicatori del mercato del lavoro contestualmente alla prima uscita dei dati raccolti con la nuova metodologia d'indagine. Strumento chiave per la ricostruzione è stato il periodo di sovrapposizione tra la vecchia e la nuova indagine, presentato nel prossimo paragrafo. L'approccio alla ricostruzione è spiegato nel paragrafo 9.3, mentre la metodologia sviluppata per la ricostruzione stessa è spiegata nell'ultimo paragrafo.

8.2 - Il periodo di sovrapposizione

Le informazioni principali a disposizione per il processo di revisione delle serie storiche vengono dall'insieme dei dati riferiti al periodo che va dal primo trimestre del 2003 al primo trimestre del 2004. Durante questo periodo le due indagini, la Rtf1 e la Rcf1, sono state condotte in parallelo. Di qui in avanti ci si riferirà a questi cinque trimestri come al "periodo di sovrapposizione". Oltre a raccogliere le informazioni fondamentali per la ricostruzione, la sovrapposizione di indagini aveva lo scopo di introdurre gradualmente la nuova indagine per testarne tutte le fasi e per portare il campione a regime. Una volta a regime il campione è costituito da un quarto di individui intervistati per la prima volta, un quarto di individui intervistati per la seconda volta, un quarto di individui intervistati per la terza volta e infine un quarto di individui intervistati per la quarta e ultima volta. Questa struttura (considerando i due trimestri di sosta tra la seconda e la terza intervista) impiega sei trimestri a entrare in regime. Per questo è stata adottata una procedura di entrata a regime accelerata, già descritta nel capitolo 4.

Di Riccardo Gatto

In ogni caso i dati del periodo di sovrapposizione non sono stati ottenuti da un campione a regime e quindi si è resa necessario un pre-aggiustamento dei dati prima che essi potessero essere utilizzati nella ricostruzione.

8.2.1 - Il processo di pre-aggiustamento dei dati

I dati della Rcfl hanno richiesto un pre-aggiustamento per omogeneizzare la struttura campionaria provvisoria con quella a regime.

Come visto la struttura campionaria della Rcfl nel periodo di sovrapposizione, data l'ingresso accelerato delle unità nel campione, è così composta:

- 2003-I campione interamente alla prima intervista;
- 2003-II / 2003-III mezzo campione alla prima intervista e mezzo alla seconda intervista;
- 2003-IV un terzo di campione alla prima intervista e due terzi alla seconda;
- 2004-I mezzo campione alla terza intervista, un quarto alla seconda e l'ultimo quarto alla prima intervista.

Uno studio preliminare ha dimostrato che i comportamenti dei secondi, terzi e quarti rispondenti possono essere assimilati tra loro mentre i primi rispondenti mostrano modalità di risposta diverse. Il pre-aggiustamento è stato necessario per correggere la compresenza nella struttura del campione di primi e altri rispondenti.

Fatta questa semplificazione il campione del primo trimestre 2004 risultava a regime mentre i dati dei trimestri dal secondo al quarto trimestre del 2003 sono stati corretti con una tecnica di post-stratificazione dei dati individuali. Il campione del primo trimestre 2003, prevedendo solo prime interviste, è stato corretto nei dati aggregati. Su un numero insieme di indicatori è stato calcolato, per il primo trimestre 2004, un vettore di parametri di aggiustamento che considerasse le differenze tra rispondenti alle prima e alle altre interviste. Questo vettore è stato poi applicato allo stesso insieme di indicatori relativo al primo trimestre 2003.

8.3 - L'approccio alla ricostruzione

Come si è visto nei capitoli precedenti, le innovazioni introdotte dalla nuova metodologia d'indagine riguardano tutti i passi del processo produttivo dei dati. Dalle definizioni al questionario, dal calendario alle tecniche e i supporti di raccolta delle informazioni, dai piani di compatibilità al campionamento. L'entità del cambiamento non ha consentito un approccio microfondato alla procedura di raccordo. Nessuna elaborazione dei microdati può consentire una omogeneizzazione dei risultati. In compenso, alla data della prima pubblicazione, le due metodologie d'indagine avevano convissuto per oltre un anno, offrendo una sovrapposizione di serie che ha consentito il confronto tra le diverse procedure. L'approccio proposto quindi non può che essere macro ed operare, quindi, ricostruendo direttamente i dati aggregati.

Le informazioni a disposizione per la ricostruzione sono quindi le serie prodotte dalla Rtfi, dal 1992-IV trimestre al 2004-I trimestre (che per ogni serie costituirà il "ramo vecchio"), e i primi cinque dati delle serie prodotte con la nuova metodologia d'indagine, dal 2003-I trimestre al 2004-I trimestre (il "ramo nuovo"). Il ramo nuovo risulta "corto" ma interamente sovrapposto agli ultimi dati del ramo vecchio. E' stata valutata la possibilità di utilizzare informazioni sul

mercato del lavoro provenienti da fonti esterne alla rilevazione, come le serie di origine amministrativa.

Vista però la passata esperienza, si è ritenuto difficile poter utilizzare dati di questa natura: ritardi nella disponibilità dei dati, parzialità e non omogeneità di trattamento dell'informazione offerta hanno suggerito di non seguire questa ulteriore possibilità.

Le considerazioni viste finora hanno fatto propendere per un approccio *macrofondato*, *model based* e *per componenti*. Il significato di approccio *macrofondato* è già stato chiarito, mentre l'approccio *model based* si basa sulla stima di un modello statistico che rappresenta la relazione tra due insiemi di dati, nel nostro caso le serie secondo la metodologia vecchia e nuova. L'approccio *model based* presenta il vantaggio di esplicitare in modo semplice e completo tutte le ipotesi formulate (a vantaggio della chiarezza e trasparenza delle procedure statistiche). Inoltre consente un'elaborazione rapida e in automatico di un gran numero di serie.

Di fatto, però, al momento della elaborazione si hanno pochi dati a disposizione (appena cinque dati di sovrapposizione), con poco tempo per l'elaborazione dei dati (ponendo fermo l'obiettivo di presentare le serie ricostruite contestualmente alla prima uscita del dato prodotto con la nuova metodologia d'indagine), con un gran numero di serie da ricostruire e un insieme complesso di vincoli di aggregazione che lega le serie tra loro. Per questi motivi, e per le ovvie esigenze di trasparenza, si è deciso di utilizzare una procedura semplice ed efficace, facilmente meccanizzabile, escludendo approcci teoricamente più complessi.

Nell'approccio *per componenti* si ricostruiscono per ogni singola serie separatamente le tre componenti temporali: il ciclo-trend, che ne costituisce la dinamica di medio-lungo periodo, la stagionalità, che ne rappresenta il pattern infra-annuale, e la componente erratica, di brevissimo periodo.

L'approccio *per componenti* è stato suggerito da diverse osservazioni: le serie degli indicatori del mercato del lavoro sono caratterizzate da una forte componente stagionale, che è la componente più interessata, nell'ampiezza delle sue oscillazioni, da cambiamenti d'indagine. Quindi si ritiene che le nuove serie non saranno caratterizzate solo da uno slittamento di livello rispetto le attuali, ma anche da un diverso pattern stagionale. In generale si ritiene che le numerose cause di differenza tra le serie producano effetti che possono manifestarsi in modo diverso sulle diverse componenti.

8.4 - Il metodo di ricostruzione

Per ogni serie il ramo vecchio ha una lunghezza sufficiente tale da non presentare problemi di identificazione e stima del modello Autoregressive Integrated Mobile Average (Arima) che lo rappresenti e non presenta quindi problemi per la scomposizione in ciclo-trend, componente stagionale e componente erratica. Per questo scopo è stata impiegata la procedura Tramo-Seats, implementata nel software Demetra (version 2 SP1)¹.

Il ramo nuovo di ogni serie presenta invece solo cinque osservazioni e non può venire scomposto con i metodi tradizionali. Il metodo proposto ha previsto di usare le informazioni provenienti dal ramo vecchio per operare una scomposizione anche sul ramo nuovo, e con i cinque valori per ogni serie calcolare delle *proxy* per i parametri del modello di relazione tra ramo vecchio e nuovo della serie. Applicare queste *proxy* a tutti i valori del ramo vecchio ha prodotto come risultato quello di "portare" all'indietro il punto iniziale del ramo nuovo al 1992-IV trimestre. Si parla di *proxy* e non di stime, perché non nascono propriamente da una procedura d'inferenza statistica.

¹ Demetra è un software con il copyright delle Comunità europee, sviluppato per Eurostat a cura di J. Dossè e S. Hoffmann.

Formalizzando: si chiami Y_t il ramo “vecchio” e y_t il ramo “nuovo”, dove t è il consueto indice temporale.

L’ipotesi base è che esista una relazione del tipo

$$y_t = f(Y_t)$$

Per quanto detto le serie possono essere decomposte come segue:

$$Y_t = T_t + S_t + E_t$$

$$y_t = t_t + s_t + e_t$$

dove T è la componente di ciclo-trend, S la componente stagionale ed E la componente erratica. Si ricordi che t_t , s_t ed e_t non appartengono al set informativo iniziale.

L’approccio per componenti prevede che ogni componente del ramo nuovo sia funzione della stessa componente del ramo vecchio:

$$t_t = f_1(T_t)$$

$$s_t = f_2(S_t)$$

$$e_t = f_3(E_t)$$

da cui segue che

$$y_t = f_1(T_t) + f_2(S_t) + f_3(E_t)$$

Resta da proporre una specificazione delle f_1 , f_2 e f_3 :

Il ciclo-trend del ramo nuovo della serie sia proporzionale a quello del ramo vecchio, quindi la f_1 diventa:

$$t_t = \alpha * T_t$$

Per la componente stagionale si ipotizzi una forma moltiplicativa di relazione, con un parametro diverso per ogni trimestre, quindi la f_2 diventa:

$$s_t = \beta_i * S_t$$

con $i = \{I, II, III, IV\}$, l’indice che tiene conto della stagionalità.

La componente erratica del ramo nuovo è composta di una parte funzione della componente erratica del ramo vecchio e di una parte “idiosincratICA”, propria esclusivamente del ramo nuovo, che sarà indicata con u_t :

$$e_t = \gamma * E_t + u_t$$

Riassumendo la forma completa diventa:

$$y_t = \alpha * T_t + \delta_i * \beta_i * S_t + \gamma * E_t + u_t$$

dove di nuovo δ_i è una variabile indicatrice uguale a 1 quando t corrisponde al trimestre i e uguale a zero altrimenti.

Si scriva l'intero sistema delle informazioni a disposizione:

sia $t = 1$ per 2003-I,
 $t = 2$ per 2003-II,
 $t = 3$ per 2003-III,
 $t = 4$ per 2003-IV e
 $t = 5$ per 2004-I;

si raddoppino inoltre gli indici per S_{it} , dove S_{it} è uguale al valore della componente stagionale del ramo vecchio se i corrisponde al trimestre in t e uguale a zero altrimenti, allora:

$$\begin{aligned} y_1 &= \alpha * T_1 + \beta_1 * S_{11} + \beta_2 * S_{21} + \beta_3 * S_{31} + \beta_4 * S_{41} + \gamma * E_1 + u_1 \\ y_2 &= \alpha * T_2 + \beta_1 * S_{12} + \beta_2 * S_{22} + \beta_3 * S_{32} + \beta_4 * S_{42} + \gamma * E_2 + u_2 \\ y_3 &= \alpha * T_3 + \beta_1 * S_{13} + \beta_2 * S_{23} + \beta_3 * S_{33} + \beta_4 * S_{43} + \gamma * E_3 + u_3 \\ y_4 &= \alpha * T_4 + \beta_1 * S_{14} + \beta_2 * S_{24} + \beta_3 * S_{34} + \beta_4 * S_{44} + \gamma * E_4 + u_4 \\ y_5 &= \alpha * T_4 + \beta_1 * S_{15} + \beta_2 * S_{25} + \beta_3 * S_{35} + \beta_4 * S_{45} + \gamma * E_5 + u_5 \end{aligned}$$

Un metodo per ottenere dei valori *proxy* per i parametri è quello di minimizzare la componente erratica pura del ramo nuovo, per esempio trovando l'insieme di valori che renda minima la somma dei quadrati degli u_t . Per fare questo si trasformi il sistema che diventa, una volta annullati i valori pari a zero:

$$\begin{aligned} y_1 &= \alpha * T_1 + \beta_1 * S_{11} + 0 + 0 + 0 + \gamma * E_1 + u_1 \\ y_2 &= \alpha * T_2 + 0 + \beta_2 * S_{22} + 0 + 0 + \gamma * E_2 + u_2 \\ y_3 &= \alpha * T_3 + 0 + 0 + \beta_3 * S_{33} + 0 + \gamma * E_3 + u_3 \\ y_4 &= \alpha * T_4 + 0 + 0 + 0 + \beta_4 * S_{44} + \gamma * E_4 + u_4 \\ y_5 &= \alpha * T_4 + \beta_1 * S_{15} + 0 + 0 + 0 + \gamma * E_5 + u_5 \end{aligned}$$

Un vincolo si dovrà imporre per impedire che nella componente stagionale si inserisca una regolarità di lungo periodo: se la somma delle componenti stagionali nel corso di un anno è diversa da zero allora queste “porterebbero” nella serie una variazione nel ciclo-trend. Questo vincolo, di solito non imposto in maniera stringente negli algoritmi di destagionalizzazione, in questo caso, avendo a disposizione dati che coprono esclusivamente cinque trimestri, è giustificato, inserendo una semplificazione accettabile nel sistema.

Il vincolo prenderà la forma:

$$\beta_1 * S_{11} + \beta_2 * S_{22} + \beta_3 * S_{33} + \beta_4 * S_{44} = 0$$

quindi:

$$\beta_4 * S_{44} = -(\beta_1 * S_{11} + \beta_2 * S_{22} + \beta_3 * S_{33})$$

il sistema ulteriormente semplificato diventa:

$$\begin{aligned}
 y_1 &= \alpha * T_1 + \beta_1 * S_{11} + 0 + 0 + \gamma * E_1 + u_1 \\
 y_2 &= \alpha * T_2 + 0 + \beta_2 * S_{22} + 0 + \gamma * E_2 + u_2 \\
 y_3 &= \alpha * T_3 + 0 + 0 + \beta_3 * S_{33} + \gamma * E_3 + u_3 \\
 y_4 &= \alpha * T_4 - \beta_1 * S_{11} - \beta_2 * S_{22} - \beta_3 * S_{33} + \gamma * E_4 + u_4 \\
 y_5 &= \alpha * T_4 + \beta_1 * S_{15} + 0 + 0 + \gamma * E_5 + u_5
 \end{aligned}$$

portando a cinque il numero di parametri presenti nel sistema.

Se y è il vettore (5x1) delle y_t , A la matrice (5x5) delle componenti stagionali del ramo vecchio, h il vettore (5x1) dei parametri da individuare e u il vettore (5x1) delle u_t , il sistema si potrà scrivere:

$$y = A * h + u$$

con:

$$A = \begin{bmatrix} T_1 & S_{11} & 0 & 0 & E_1 \\ T_2 & 0 & S_{22} & 0 & E_2 \\ T_3 & 0 & 0 & S_{33} & E_3 \\ T_4 & -S_{11} & -S_{22} & -S_{33} & E_4 \\ T_5 & S_{15} & 0 & 0 & E_5 \end{bmatrix}$$

e

$$h = (\alpha, \beta_1, \beta_2, \beta_3, \gamma)$$

Il punto cui si è arrivati assomiglia molto a una regressione lineare: ora individuare il vettore h che renda minima la somma dei quadrati delle u_t è facile:

$$\operatorname{argmin} u * u' = \operatorname{inv}(A'A) * A'y$$

dove l'apice rappresenta l'usuale operatore di trasposizione e $\operatorname{inv}(\cdot)$ è l'operatore di inversione di una matrice.

A questo si dovrà aggiungere la determinazione di β_4 :

$$\beta_4 = -\beta_1 * S_{11} / S_{44} - \beta_2 * S_{22} / S_{44} - \beta_3 * S_{33} / S_{44}$$

Il metodo descritto dà ottimi risultati in termini di scomposizione del ramo nuovo delle serie ma a volte fallisce in termini di modello di ricostruzione. Ciò accade a causa della specificazione

moltiplicativa della componente stagionale in presenza di realizzazioni molto piccole della componente stessa, in questi casi il parametro β_i assume facilmente valori troppo grandi. In questi casi, fatta salva la scomposizione, è stato applicato per la ricostruzione un modello a stagionalità additiva.

Aggregando le serie ricostruite sorge un problema di aggregazione. La procedura sin qui descritta è monovariata e permette di trattare ogni serie singolarmente senza tenere conto dei vincoli di aggregazione che le legano tra loro. Per le serie a un livello di aggregazione più alto è possibile quindi ottenere il riallineamento attraverso un metodo diretto o uno indiretto, riaggregando le serie riallineate elementari. Il problema sorge a causa della sostanziale non-linearità della procedura: sia $RP(*)$ la procedura di ricostruzione descritta e siano $x_t = x_{1t} + x_{2t}$ rispettivamente una serie aggregata e le due serie addendi. In generale non si verificherà $RP(x_t) = RP(x_{1t}) + RP(x_{2t})$. Questo è un risultato atteso vista la ben nota non-linearità della procedura di scomposizione delle serie storiche. Preso un certo livello di aggregazione, sufficientemente alto da essere rappresentativo, come benchmark, le serie di livello superiore si sono ottenute semplicemente in modo indiretto aggregando queste. Alle serie di livello inferiore è stata applicata una procedura di data mining che ha come input le serie addendi ricostruite e come output le stesse serie ma coerenti con le serie di benchmark. Il risultato è un dataset di serie storiche tutte perfettamente coerenti tra loro, a tutti i livelli di aggregazione.

Nel prospetto 8.1 sono riportati i valori di alcuni indicatori utili per valutare gli effetti del raccordo; in particolare:

$$\text{Media aritmetica dello scostamento } \bar{v} = \frac{1}{n} \sum_{t=1}^n (r_t - p_t)$$

$$\text{Media aritmetica dello scostamento in valore assoluto } \bar{v}' = \frac{1}{n} \sum_{t=1}^n |r_t - p_t|$$

$$\text{Media aritmetica dello scostamento relativo } \bar{w} = \frac{1}{n} \sum_{t=1}^n \frac{r_t - p_t}{p_t}$$

$$\text{Media aritmetica dello scostamento relativo in valore assoluto } \bar{w}' = \frac{1}{n} \sum_{t=1}^n \left| \frac{r_t - p_t}{p_t} \right|$$

$$\text{Coefficiente di correlazione } \rho = \frac{1}{n} \sum_{t=1}^n (r_t - \bar{r})(p_t - \bar{p}) / \sqrt{\sum_{t=1}^n (r_t - \bar{r})^2 \sum_{t=1}^n (p_t - \bar{p})^2},$$

dove p_t indica la serie vecchia e r_t la serie ricostruita. L'entità dello scostamento, relativamente al trimestre t , è dato da: $v_t = r_t - p_t$.

Prospetto 8.1 - Indicatori sintetici di revisione per alcune significative serie ordinate per la Media aritmetica dello scostamento relativo in valore assoluto

SERIE	Scostamento (dati in migliaia)		Scostamento relativo (dati in percentuale)		Coefficiente di correlazione
	Medio	Medio assoluto	Medio	Medio assoluto	
Forze di lavoro Italia	84	91	0.004	0.004	0.990
Occupati mezzogiorno	-26	31	-0.004	0.005	0.992
Occupati maschi	-109	110	-0.008	0.008	0.970
Occupati Italia	177	177	0.009	0.009	0.996
Forze di lavoro Nord	104	104	0.009	0.009	0.990
Forze di lavoro Mezzogiorno	-83	84	-0.011	0.011	0.982
Forze di lavoro maschi	-167	167	-0.012	0.012	0.883
Occupati Nord	125	125	0.012	0.012	0.996
Forze di lavoro Centro	63	63	0.014	0.014	0.983
Occupati Centro	78	78	0.019	0.019	0.992
Occupati Industria	-167	167	-0.024	0.024	0.963
Forze di lavoro femmine	251	251	0.029	0.029	0.996
Persone in cerca femmine	-35	39	-0.027	0.030	0.988
Persone in cerca Nord	-22	22	-0.036	0.036	0.999
Occupati altre attività	457	457	0.036	0.036	0.997
Persone in cerca Centro	-14	14	-0.036	0.037	0.998
Persone in cerca Italia	-93	93	-0.038	0.038	0.997
Occupati femmine	287	287	0.038	0.038	0.998
Persone in cerca Mezzogiorno	-57	57	-0.039	0.039	0.997
Persone in cerca maschi	-58	58	-0.049	0.049	0.998

Gli scostamenti medi mostrano sia segni positivi sia segni negativi, la procedura di ricostruzione ha quindi agito sia aumentando sia diminuendo i livelli medi delle serie secondo i casi. Probabilmente questo ha a che fare più con l'impatto dell'introduzione della nuova indagine sui diversi aggregati che con le performance della procedura di ricostruzione. In ogni caso quest'ultima sembra ben cogliere tale impatto.

Elevati livelli del coefficiente di correlazione, quasi mai inferiore allo 0,9, mostrano come la dinamica delle serie ricostruite ricalchi quella delle serie precedentemente pubblicate. Questo anche a fronte di scostamenti medi assoluti (che in qualche modo misurano la distanza tra le serie) elevati. È il caso della serie degli occupati nei servizi che in media mostra uno scostamento di 457 mila unità dopo la ricostruzione ma registra un coefficiente di correlazione pari a 0,997. Gli scostamenti relativi non superano mai lo 0,049 per cento, e ciò significa che nessuna serie, almeno tra le principali, risulta rivoluzionata dalla ricostruzione, anche se l'effetto del passaggio alla nuova indagine è spesso sensibile.

Capitolo 9 – Il sistema informativo-informatico

9.1 - Il Sistema informativo per la gestione delle indagini sulle famiglie

Tutte le fasi del processo produttivo sono supportate dal sistema informativo-informatico denominato Sigif (Sistema informativo per la gestione delle indagini sulle famiglie). Il sistema risponde al “modello a rete” dell’indagine, in cui alla condivisione degli obiettivi si affianca l’opera di comunicazione, sensibilizzazione e mutua assunzione di impegni tra i diversi protagonisti. Le metodologie e le Ict utilizzate sono rivolte ad assicurare la riservatezza, la sicurezza e l’integrità delle informazioni. Esse sono cruciali non solo per la conduzione efficiente della nuova indagine ma anche per il controllo e monitoraggio delle attività.

In quest’ottica l’Istat ha sperimentato l’applicabilità della Action Workflow Analysis (Awa), una metodologia basata sulla teoria degli atti linguistici. Permettendo la rappresentazione di tutte le dimensioni del processo (organizzazione, rete di impegni fra gli attori, tecnologie), l’utilizzo di tale metodologia ha rappresentato per l’Istituto un fattore di forte innovazione. Nel caso della nuova indagine sulle forze di lavoro, l’Awa ha contribuito a disegnare, con una notevole razionalizzazione dei costi, il progetto complessivo, considerando congiuntamente le componenti organizzative, quelle specifiche funzionali e quelle tecnologiche¹.

I vincoli tecnologici rappresentano i fondamentali fattori di cui si è tenuto conto per la costruzione del sistema informativo-informatico. Essi fanno riferimento agli standard informatici (hardware, software di base e d’ambiente, software applicativo, basi di dati) e alla rete telematica dell’Istituto. Dati i vincoli tecnologici e l’architettura organizzativa su tre livelli (centro di coordinamento, Uffici territoriali, rilevatori) dell’indagine, il sistema garantisce il raggiungimento di differenti obiettivi: dal raccordo delle tecniche di indagine Capi-Cati, all’assegnazione ai rilevatori delle informazioni necessarie per l’effettuazione delle interviste, alla raccolta dei feed-back del lavoro sul campo. Per il funzionamento del sistema è stato necessario costruire una rete virtuale sicura, “blindata” dal punto di vista dell’intrusione di virus, basata sulle rete telefonica fissa nazionale. Viene peraltro garantito per le interviste Capi un livello di servizio con non più di un giorno di fermo nell’arco di un bimestre.

Il sistema integra otto sottosistemi modulari. Tale integrazione funzionale e tecnologica è realizzata tramite un database (centrale dal punto di vista concettuale e distribuito dal punto di vista fisico) dove tutti i flussi informativi ed i dati della Rcfl sono organizzati e memorizzati².

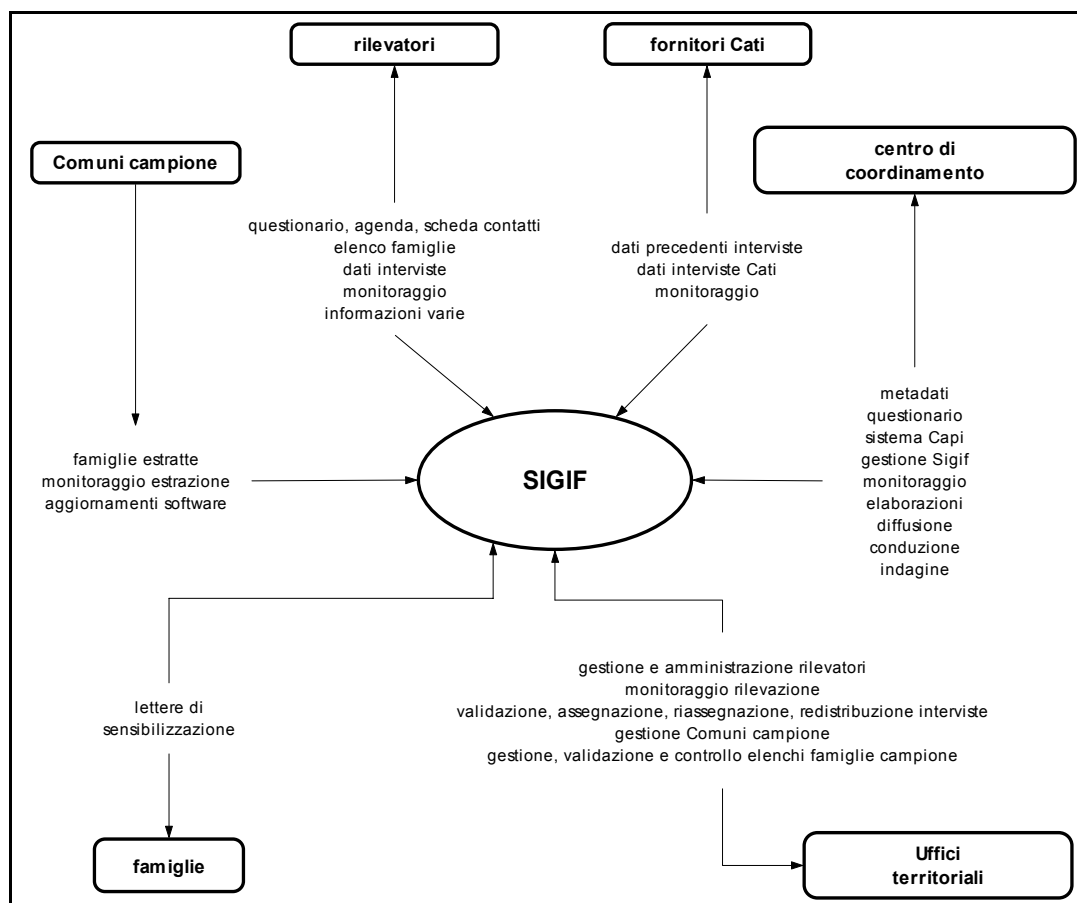
Lo schema sottostante rappresenta i principali flussi delle informazioni.

Di Antonio Toma (par. 9.1), Gabriele Budano, Paolo Floris, Marco Scuderi (parr. 9.2.1, 9.2.2, 9.2.3, 9.2.4), Sandro Baldanza, Claudio Kolenc, Marco Iudicone (par. 9.2.5), Stefania Bergamasco, Antonella Filacchione, Fabian Mancuso, Enrico Manna, Luca Rossi par..9.3), Gabriele Budano, Antonio Toma (par. 9.4), Vincenzo Triolo (par. 9.5)

¹ In base alla teoria dell’Action Workflow Analysis, il processo produttivo non è altro che una “conversazione per fare” tra un richiedente A (cliente) e un rispondente B (fornitore) il cui svolgersi è rappresentato da un diagramma di stati e di azioni che evidenziano il significato di processo come workflow. Tale modello è coerente con la definizione di processo come “mutua assunzione di impegni tra attori che intendono raggiungere in forma cooperante un obiettivo” (vedi Winograd-Flores, “Conversation for action”). La scelta di utilizzare tale modello risulta in linea con le indicazioni del Centro Nazionale per l’Informatica nella Pubblica Amministrazione e la strategia nazionale di e-government. Si vedano a tal proposito i siti www.cnipa.it e www.italia.gov.it.

² Il sistema Sigif dialoga inoltre con il Sirc (Sistema informativo per la rete di rilevazione comunale) e supporta alcune delle funzionalità necessarie alla conduzione dell’indagine EU-SILC e di alcune indagini Multiscopo. Uno degli obiettivi del sistema è difatti capitalizzare, in termini di costi e di *best practices*, i metodi, le tecniche e gli strumenti informatici messi a punto con la Rcfl per un più vasto sistema integrato di gestione delle indagini sulle famiglie.

Figura 9.1 - Il sistema informativo-informatico



I sottosistemi riguardano:

- la gestione dei rilevatori, per gli aspetti amministrativi, organizzativi e per il supporto alla selezione, assunzione, formazione, ai compensi e per la gestione delle aree geografiche di rilevazione;
- l'estrazione e la predisposizione del campione dei Comuni e delle famiglie (ci si riferisce qui alle funzionalità necessarie per l'individuazione dei Comuni e delle famiglie del campione, al controllo e alla correzione ex-ante degli eventuali errori di lista);
- la sensibilizzazione delle famiglie;
- la gestione delle interviste Capi;
- il raccordo tra queste e quelle Cati;
- il monitoraggio del processo (lo scopo di questo sottosistema è permettere la valutazione delle operazioni relative alla selezione degli intervistatori, all'estrazione del campione, alle operazioni sul campo, alla sensibilizzazione delle famiglie, alla predisposizione dei file, alla pubblicazione dei risultati, nonché controllare il rispetto delle scelte metodologiche, delle procedure e dei tempi stabiliti);
- la definizione dei metadati e il trattamento, controllo, correzione, validazione, elaborazione e diffusione dei dati;
- la complessiva infrastruttura tecnologica.

In particolare, il sistema Capi-Cati è composto da due ambienti principali fisicamente separati ma concettualmente in collegamento tra loro: il server Capi e il server della ditta di rilevazione Cati. Attraverso l'utilizzo di procedure di validazione, gli Uffici territoriali leggono settimanalmente le attribuzioni automatiche delle interviste ai rilevatori e ne confermano la correttezza, aggiornando il database. Il sistema Capi gestisce poi la trasmissione delle famiglie ai rilevatori e la ricezione dei contatti e delle interviste effettuate.

Nello stesso tempo viene creato l'input Cati, sempre attraverso un doppio processo di lettura/aggiornamento. Il file estratto popola il database della ditta Cati: gli operatori telefonici ricevono automaticamente sul loro *client* i dati della famiglia da contattare, attraverso una procedura di schedulazione dei nominativi, e trasmettono al server le informazioni relative ai contatti e alle interviste effettuate. In caso di reintervista, le informazioni individuali, sia per il Capi sia per il Cati, comprendono anche quelle rilevate nelle precedenti occasioni di indagine.

A conclusione delle operazioni descritte, il sistema inserisce in Sigif gli esiti dei contatti con le famiglie e le informazioni raccolte con il questionario elettronico. Da Sigif vengono poi generati i dati finali di output per ambedue le componenti Capi e Cati (interviste individuali, tentativi ed esiti definitivi di contatto, indicatori di qualità, contatori delle regole, informazioni per le interviste di controllo).

L'ambiente tecnologico standardizzato ed integrato si è basato fino alla fine del 2005 su:

- un sistema di acquisizione unico di dati provenienti da enti esterni (Comuni, Ditte di rilevazione Cati, etc.) implementato con tecnologie secure-Web e secondo le regole contenute nel codice sulla privacy.
- un sistema Capi formato da 312 postazioni client dei rilevatori (Pc portatili) e dalle postazioni desktop dei supervisor degli uffici territoriali e del personale coinvolto nell'indagine, da un server centrale ad alta affidabilità, da un sistema di accesso remoto e dai dispositivi hardware e software per la gestione della sicurezza di rete³. Il server centrale è stato collocato all'interno della Intranet Istat ed i rilevatori hanno inviato le informazioni accedendovi in modo sicuro e trasferendo le stesse in maniera criptata rispettando tutte le regole tecniche contenute nel codice sulla privacy. Il sistema Software Capi è stato implementato tramite la concessione in licenza ad uso a tempo determinato (fino alla fine del 2005) dei pacchetti Capi forniti da un Rti (Raggruppamento temporaneo d'impresa). Tale Rti ha fornito, inoltre, i servizi ausiliari per la gestione dell'indagine quali lo sviluppo e la manutenzione del questionario elettronico, i servizi di assistenza sul campo, di help-desk di primo e secondo livello nonché i P.c. portatili dei rilevatori corredati di tutto il software di base e d'ambiente. Il sistema Capi è stato realizzato tramite due componenti:

componente software Capi client. Tale componente, eseguita sui Pc portatili dei rilevatori, è costituita dall'apposito modulo runtime che permette tramite un "unico cruscotto di gestione" l'implementazione dell'agenda, dei contatti e del questionario elettronico (per le funzionalità necessarie ai rilevatori);

³ Il sistema di Capi management per la componente di gestione delle interviste è stato realizzato da un raggruppamento temporaneo di imprese (RTI) tramite la personalizzazione del pacchetto Capi commercializzato dalla società francese Conversoft. Tale personalizzazione è stata realizzata mediante l'implementazione di apposito software applicativo realizzato in ambiente MS-SQL-SERVER e Borland Delphi. La componente di gestione e monitoraggio delle trasmissioni è stata realizzata in maniera "quasi-nativa" tramite il software Conversoft. Sono state previste, comunque, diverse procedure batch (realizzate in PL/SQL e SAS) per la gestione dei flussi tra componente Capi ed i diversi sottosistemi SIGIF-RCFL. Le operazioni di gestione delle interviste Capi sono a carico degli uffici territoriali, i quali, tramite opportune maschere sviluppate all'interno dell'applicativo Delphi procedono, tenuto conto della situazione del field, alle assegnazioni, rassegnazioni e ridistribuzioni delle interviste. Tale sistema è stato attivo dalla fine del 2002 alla fine del 2005.

Componente software Capi server o Capi management, per supportare le operazioni più complesse, sia dal punto di vista metodologico che tecnico, ed in particolare la gestione delle interviste Capi⁴, la gestione e il monitoraggio delle trasmissioni⁵;

- un sistema basato su tecnologia Oracle distribuito sugli Uurr e su Bolzano e Trento tramite la Intranet Sicura dell'Istat;
- un sistema di elaborazione dei dati basato su tecnologie Oracle, Ssa, procedure di elaborazione in C++ e software generalizzati;
- un sistema di archiviazione ed organizzazione dell'informazione in Oracle ed in ambiente Aix sui nodi dell'elaboratore SP ad architettura SMP;
- un sistema di posta elettronica dedicato ad uso esclusivo della rete di rilevatori Istat.

9.2 - L'internalizzazione del sistema Capi

9.2.1 - Introduzione

Il sistema Capi appena descritto ha supportato l'indagine tra il mese di ottobre 2002 e il mese di dicembre 2005. Nel corso di questo periodo, pur riconoscendo la funzionalità e la qualità del sistema messo a disposizione dal Rti esterno all'Istat, si è presa coscienza della possibilità e della necessità di apportare miglioramenti all'intero sistema Capi. Si è così deciso di attivare un gruppo di lavoro che procedesse all'internalizzazione del sistema Capi, con l'obiettivo di gestire direttamente all'interno dell'Istat tutti i processi e i prodotti necessari al regolare svolgimento dell'indagine. Nel corso del suddetto periodo, infatti, sono emersi alcuni punti critici del sistema Capi che, nel lungo periodo, avrebbero potuto avere anche effetti negativi sulla qualità dei dati raccolti. Ci si riferisce in particolare a:

- il rischio di discontinuità nella gestione Capi, in quanto la durata degli appalti a ditte private comporta necessariamente dei limiti temporali superati i quali si corre il rischio che l'intero sistema possa all'improvviso cambiare gestore a seguito di un cambio di ditta fornitrice (con tutti i problemi e gli aggravii di lavoro derivanti dalla necessità di fornire nuove specifiche hardware e software, specifiche di tipo funzionale, tracciati record, questionari, con notevoli rischi di impatto negativo sullo svolgimento dell'indagine continua);
- la mancanza di un controllo completo di tutte le fasi del processo, poiché alcune macro funzioni e alcuni prodotti rimangono di proprietà della ditta appaltatrice che le inserisce nei processi di produzione senza fornire documentazione dettagliata e le ritira al termine dell'appalto (con conseguente necessità di rimettere in produzione funzioni già sviluppate);
- il controllo indiretto delle modifiche da apportare in corso d'opera ai prodotti sviluppati dalla ditta, con conseguente allungamento dei tempi di produzione, impatto sui carichi di lavoro dell'Istat, schedulazione dei lavori con la ditta appaltatrice, effettuazione di controlli e test aggiuntivi;
- alla difficoltà di gestire l'obsolescenza dell'hardware con Rti chiamati ad occuparsi dell'intero sistema e non solo della qualità dell'hardware.

⁴ Per le operazioni di assegnazione, riassegnazione e redistribuzione delle interviste.

⁵ Per le operazioni di schedulazione, preparazione, gestione e monitoraggio delle trasmissioni (invio una tantum per le differenti versioni del questionario e per il materiale formativo ed informativo, invio settimanale degli indirizzari delle famiglie e/o dei dati delle precedenti interviste per reinterviste Capi, ricezione giornaliera delle interviste realizzate).

Alla luce di queste considerazioni, nel corso del 2004 e del 2005 si è provveduto alla progettazione ed alla realizzazione di un sistema Capi totalmente gestito dall'Istat, caratterizzato da standard qualitativi elevati e funzionale alla risoluzione dei problemi appena citati.

Il sistema è stato progettato e messo in esercizio nell'ottica del miglioramento continuo del processo, sia dal punto di vista della gestione e sviluppo informatico che da quello di razionalizzazione dello stesso processo di produzione. In tale prospettiva l'Istat ha internalizzato il sistema Capi realizzandolo:

- con il software Blaise (prodotto dall'Istituto di Statistica Olandese ed ormai diventato uno standard all'interno dell'Istat) per lo sviluppo del questionario (comprensivo della scheda contatti con le famiglie) e con il software Visual Basic per lo sviluppo dell'agenda del rilevatore;
- con lo sviluppo di applicazioni ad hoc a supporto di funzionalità non implementate da Blaise;
- con lo sviluppo della componente di gestione e monitoraggio delle trasmissioni basata su una console di gestione capace di costruire il flusso trasmissivo tramite protocolli standard (ftp);
- con l'integrazione delle funzionalità di gestione delle interviste nelle componenti Sigif sviluppate in Oracle.

Contestualmente si è deciso di consolidare l'architettura attuale apportando l'opportuno rinnovamento all'infrastruttura tecnologica, in particolar modo per le tecnologie a supporto della sicurezza di rete e degli accessi, coerentemente con il nuovo assetto dell'architettura di elaborazione dell'Istituto.

Il nuovo sistema Capi presenta numerosi aspetti innovativi ed assicura uno svolgimento più efficace ed efficiente dei processi necessari a garantire la rilevazione. In primo luogo, la gestione interna del questionario permette controlli ed interventi molto più rapidi sullo stesso ed alleggerisce l'Istituto dall'onere di dover contattare e controllare la ditta esterna fornitrice per ogni minimo mutamento del questionario. A ciò si aggiunga che il nuovo sistema si presenta estremamente più efficiente sul piano delle trasmissioni (utilizzo di protocolli iper-collaudati), con conseguente notevole risparmio di costi per l'Istituto. Il nuovo sistema, completamente realizzato all'interno e con pacchetti software standard di costo contenuto e/o open source, viene gestito da risorse interne rendendo così l'Istituto indipendente da fornitori esterni.

L'internalizzazione del sistema Capi comporta anche l'internalizzazione di alcune attività informatiche specialistiche attualmente demandate al RTI. In particolare, oltre alla normali attività di sviluppo e manutenzione delle componenti software, è stato necessario organizzare un pool di risorse addette alla produzione e manutenzione del questionario elettronico. Questo nuovo approccio facilita notevolmente la crescita professionale delle stesse risorse umane.

La progettazione di un sistema di acquisizione dati con tecnica Capi comporta essenzialmente l'ideazione di un ambiente che contenga il questionario elettronico e l'agenda degli appuntamenti (sul lato client), si occupi della spedizione dei dati basandosi su un protocollo di trasmissione sicuro e si basi su un'architettura hardware e software sul lato server in grado di ricevere e trasmettere le informazioni al client.

E' stato dunque necessario progettare e realizzare un'architettura di rete sicura ed affidabile per garantire le trasmissioni e per innalzare ulteriormente il livello di sicurezza, dotando il sistema di un apparato di crittografia valido sia per i dati residenti sul personal computer che per i dati che devono viaggiare sulla rete.

Alla luce di queste considerazioni, le novità introdotte hanno riguardato la scelta dell'hardware⁶, sia per la parte client che per la parte server, l'architettura di rete, un sistema di sicurezza, il software per la costruzione del questionario e dell'agenda, il protocollo di trasmissione.

Il nuovo sistema Capi, dopo una fase di test in parallelo con il vecchio sistema, è operativo a partire dal mese di gennaio 2006.

9.2.2 - L'agenda del rilevatore

Il rilevatore, in apertura del sistema, deve poter visualizzare l'elenco completo delle famiglie da intervistare. Per ogni famiglia il rilevatore deve poter visualizzare una serie di informazioni relative alla famiglia, alla quartina, agli eventuali appuntamenti già fissati, alle scadenze per il completamento dell'intervista. Il rilevatore deve poter ordinare le famiglie per ciascuna delle informazioni che saranno a video. Il rilevatore deve poter indicare, tra le altre informazioni, per ogni tentativo di contatto: il tipo, il motivo, l'esito, data, ora e minuto del tentativo di contatto, l'esito dell'eventuale contatto, il motivo dell'eventuale rifiuto, il tipo e il motivo dell'eventuale interruzione, l'eventuale data di appuntamento. Il sistema deve consentire la gestione coerente degli appuntamenti ed essere di supporto all'attività di schedulazione e pianificazione del rilevatore (avvisi di sovrapposizione di appuntamenti, appuntamenti troppo ravvicinati, appuntamenti oltre il periodo di rilevazione) Il rilevatore dovrà gestire i tentativi di contatto e i contatti con le famiglie in base alle regole definite di cui il sistema deve tenere traccia. Qualora si giunga, per esaurimento dei tentativi, alla caduta della famiglia da intervistare, il sistema, avendo registrato l'iter dei tentativi di contatto, dovrà avvertire il rilevatore della necessità di procedere alla sostituzione della famiglia. In caso di conferma, da parte del rilevatore, di sostituzione della famiglia, il sistema deve gestire la chiusura del rapporto con questa, registrando sia lo stato della famiglia sostituita sia lo stato della famiglia sostitutiva nonché lo stato del ciclo e della quartina di riferimento. Il sistema, al momento della chiusura di ogni tentativo di contatto, deve aggiornare lo stato della famiglia, della quartina, del ciclo.

L'agenda del rilevatore prevede 5 funzioni per a) la gestione vera e propria dell'agenda, b) la trasmissione delle interviste, c) l'invio di e-mail all'Istat, d) la chiamata dell'help desk ed e) la consultazione del manuale per il rilevatore.

La prima funzione attiva la parte gestionale dell'agenda che pone all'attenzione del rilevatore tutta una serie di strumenti necessari a contattare e a intervistare le famiglie in carico.

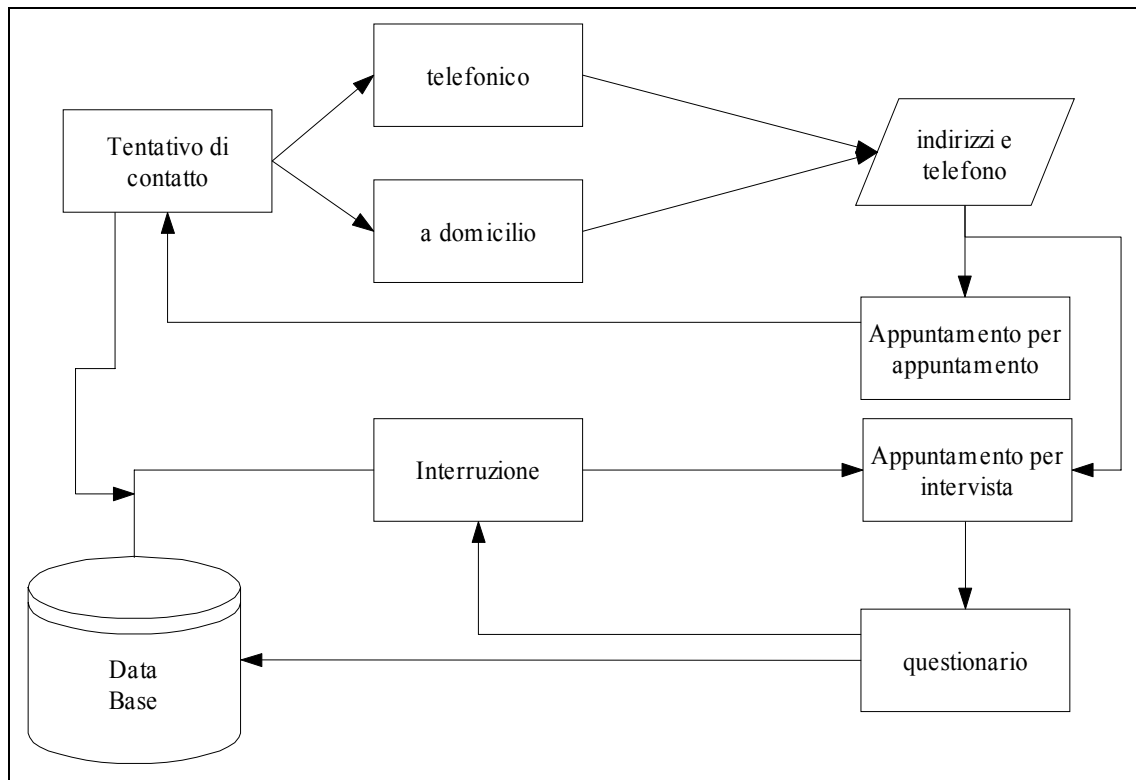
Attraverso questa prima schermata il rilevatore sarà in grado di controllare i nominativi delle famiglie da intervistare, verificarne i dati anagrafici, avere un quadro d'insieme degli appuntamenti divisi per giorno e per settimana.

Controllato il quadro d'insieme inizia il lavoro del rilevatore, consistente nel contare le famiglie da intervistare. Si inizia normalmente con una telefonata per presentarsi e preannunciare la propria visita e si termina con l'intervista. Tra la prima telefonata e l'intervista finale potranno esserci uno o più passaggi (altre telefonate, appuntamenti rinviati o andati a vuoto), famiglie sostituite per varie ragioni. Nella figura seguente è riportato il percorso logico per lo svolgimento

⁶ E' stata bandita ed assegnata un'apposita gara per l'acquisizione del nuovo hardware, consistente in 370 personal computer portatili con Cpu intel centrino, sistema operativo Windows XP, 512Mb di Ram, 80 Gb di hard disk e modem a 56Kbs e di circa 40 personal computer desktop con Cpu intel pentium 4 da utilizzare presso gli uffici regionali Istat e presso il servizio FoL. Sono stati inoltre acquisiti due server, dotati di forte capacità di calcolo e di una disponibilità di memoria di massa adeguata, che lavorano in cluster, utilizzano Windows 2003 Server come sistema operativo e che sono dotati di 8 Dischi da 120 GB SCSI, di cui 2 con sistema Raid 1 e 6 con sistema Raid 5, di 2 alimentatori ridondanti e di un sistema di bilanciamento flusso Dati e Rete

delle interviste. La procedura guida il rilevatore in questo percorso, permettendogli di attivare le varie funzioni in modo semplice ed intuitivo.

Figura 9.2 – Flow chart dei contatti



Ottenuto un appuntamento per effettuare l'intervista, il rilevatore si recherà presso il domicilio della stessa, registrerà le informazioni fornitegli e trasmetterà i dati al server Istat.

Terminate le interviste il rilevatore si avvale della funzione per la trasmissione dei dati al server che, come già detto, avviene utilizzando il protocollo Ftp con un sistema di crittografia a chiave pubblica. Naturalmente queste scelte sono trasparenti per l'utente, infatti una configurazione interna prevede già la presenza di una user e di una password personalizzata per quello specifico client. All'atto dell'attivazione della trasmissione viene creata una connessione sicura ed inizia il trasferimento dei dati, prima dal server al client e successivamente dal client al server. Al termine della trasmissione la connessione verrà chiusa in modo automatico.

Più in particolare, nel corso della trasmissione vengono prima caricate le famiglie da intervistare, eventuali aggiornamenti software e solo in seguito il software si occupa di trasmettere in formato ascii al server tutto il data base residente sul client (formato dalle interviste e dai tentativi di contatto).

All'avvio del processo di trasmissione viene presentata una schermata con delle informazioni preliminari per il rilevatore e sul lato sinistro della schermata viene visualizzata una barra verticale che durante l'intero processo di trasmissione indica lo stato di avanzamento della stessa.

Uno snodo fondamentale per lo sviluppo delle applicazioni necessarie alla realizzazione delle interviste Capi è quello del dialogo tra l'agenda del rilevatore da un lato e, dall'altro, la registrazione dei contatti con le famiglie da intervistare e il questionario vero e proprio.

La scelta del software di base da utilizzare riguardava essenzialmente lo sviluppo del questionario elettronico e dell'agenda. Fatta un'analisi di mercato, per la realizzazione della scheda contatti e del questionario la scelta si è orientata su di un prodotto sviluppato dall'Istituto nazionale di statistica olandese (CBS) chiamato Blaise, già utilizzato in Istat con successo per altre rilevazioni Cati. Le altre componenti software relative all'agenda sono state sviluppate in Visual Basic ed in linguaggi del mondo open source.

Il prodotto realizzato per personal computer con sistema operativo Windows, è un sistema Client Server per:

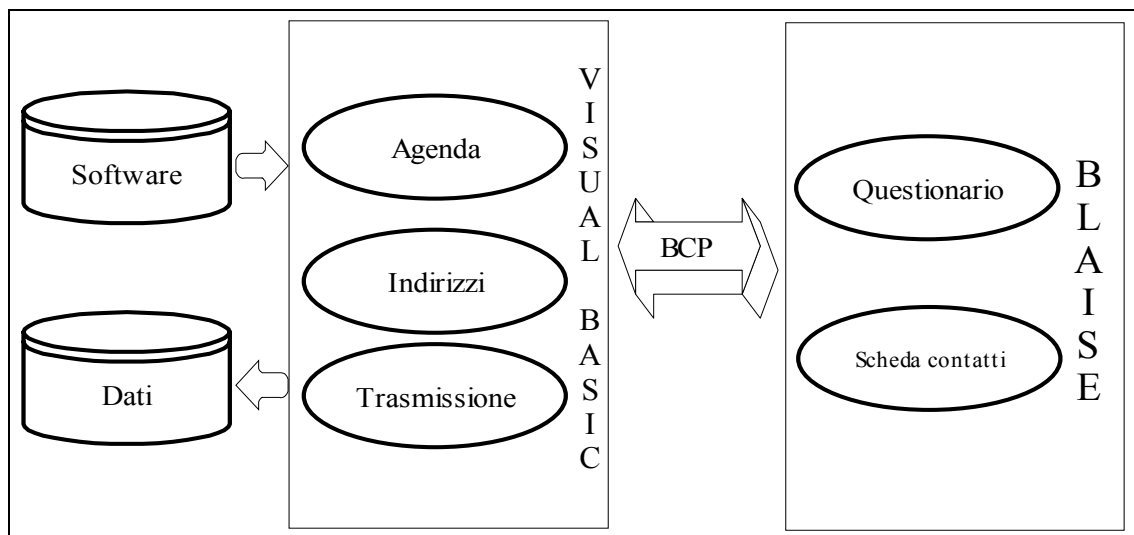
- la rilevazione dei dati
- la trasmissione dei dati in Istat e la loro elaborazione.

Il sistema si compone di due parti. La parte client, che opera su sistema operativo Windows XP professional SP1, necessita del sistema Blaise 4.6 con la componente BCP (Blaise Component Pack 2.0) e utilizza moduli esterni realizzati in Visual Basic 6. La parte server opera su sistema operativo Windows 2003 server, utilizza moduli esterni realizzati in visual basic 6 ed un data base interno realizzato in Access.

Nella figura seguente viene mostrato lo schema del flusso dei dati che vanno dall'agenda alla scheda contatti ed al questionario e viceversa con l'utilizzo del modulo BCP.

Come si può vedere scopo del Bcp è instaurare un colloquio tra la componente Blaise e la componente Visual Basic. Tale colloquio è naturalmente trasparente per l'utente finale che passerà da un'applicazione all'altra con semplici click del mouse.

Figura 9.3 – Schema logico del flusso agenda-questionario



Ricapitolando i compiti del sistema sono:

- gestire i contatti e il questionario;
- gestire l'agenda degli appuntamenti;

- garantire la trasmissione dei dati e l'aggiornamento del software e dei parametri d'indagine;
- estrarre e approntare dati nelle rispettive cartelle di lavoro per rilevatore;
- caricare i dati elaborati e periodicamente trasmessi dal rilevatore sul Server e caricarli sul data base di gestione;
- catturare i file di log generati durante le fasi di trasmissione necessari a verificare la qualità dei dati ricevuti e trasmessi.

9.2.3 - Il questionario elettronico

Per lo sviluppo del questionario è stato utilizzato, come appena ricordato, il software Blaise 4.6, prodotto che, pur non essendo un linguaggio di programmazione orientato alla soluzione di problemi informatici, è estremamente funzionale alla costruzione di questionari complessi, quale è quello dell'indagine sulle forze di lavoro.

Il questionario è stato sviluppato rispettando tutte le caratteristiche e le indicazioni metodologiche già delineate nel capitolo 6. Le funzionalità presenti nel linguaggio Blaise hanno permesso di migliorare notevolmente alcune caratteristiche del questionario che possono risultare estremamente utili ad una conduzione efficiente ed efficace dell'intervista.

Per ogni campo imputato a video, ad esempio, si eseguono e si controllano sempre tutte le regole dell'intero programma, permettendo all'intervistatore di tornare sempre indietro e cambiare le risposte (se le regole lo consentono).

Il valore di un campo imputato fa scattare una regola ad esso concatenata, in qualsiasi punto la regola si trovi. Pertanto un'incompatibilità fra campi scatterà sempre proponendo i valori in conflitto, permettendo all'intervistatore di cambiare le variabili chiamate in causa dalla regola di incompatibilità.

Se il valore di un campo viene cambiato, automaticamente verranno presentate le domande relative al cambiamento e verranno mutati i percorsi del questionario logicamente conseguenti al cambio introdotto.

Ad esempio se il sesso di un intervistato di 20 anni è stato erroneamente imputato come "maschio" e nel corso dell'intervista ci si accorge dell'errore di imputazione, si può tornare indietro per cambiare il sesso in "femmina" ed automaticamente tutte le risposte relative al "servizio militare" precedentemente date verranno congelate. All'atto della registrazione finale che chiude l'intervista, i dati congelati non verranno presi in considerazione nel processo di scrittura del record finale.

La schermata del questionario è divisa dall'alto in basso in tre parti: la parte superiore riporta la domanda da porre all'intervistato, la seconda le modalità di risposta ed infine la parte inferiore riporta le risposte date alle domande precedenti, la risposta data alla domanda in corso ed infine le successive domande da porre, aiutando in tal modo l'intervistatore a contestualizzare la risposta fornita in un ambito allargato di risposte.

E' naturalmente possibile sospendere l'intervista in un qualsiasi momento e riprenderla quando necessario riportando il focus dell'intervista sul componente della famiglia che ha interrotto l'intervista e sulla domanda rimasta in sospeso.

Per alcune risposte codificate, il questionario fa uso di tabelle esterne che permettono al rilevatore di ricercare in modo veloce i valori da imputare, ad esempio i codici relativi alle province, ai comuni, titoli di studio, attività economiche e professionali.

9.2.4 - Il nuovo dominio Rcfl

Per la gestione del sistema Capi internalizzato è stato predisposto un apposito dominio di rete. Un dominio è costituito da un insieme di computer connessi in rete aventi una gerarchia definita.

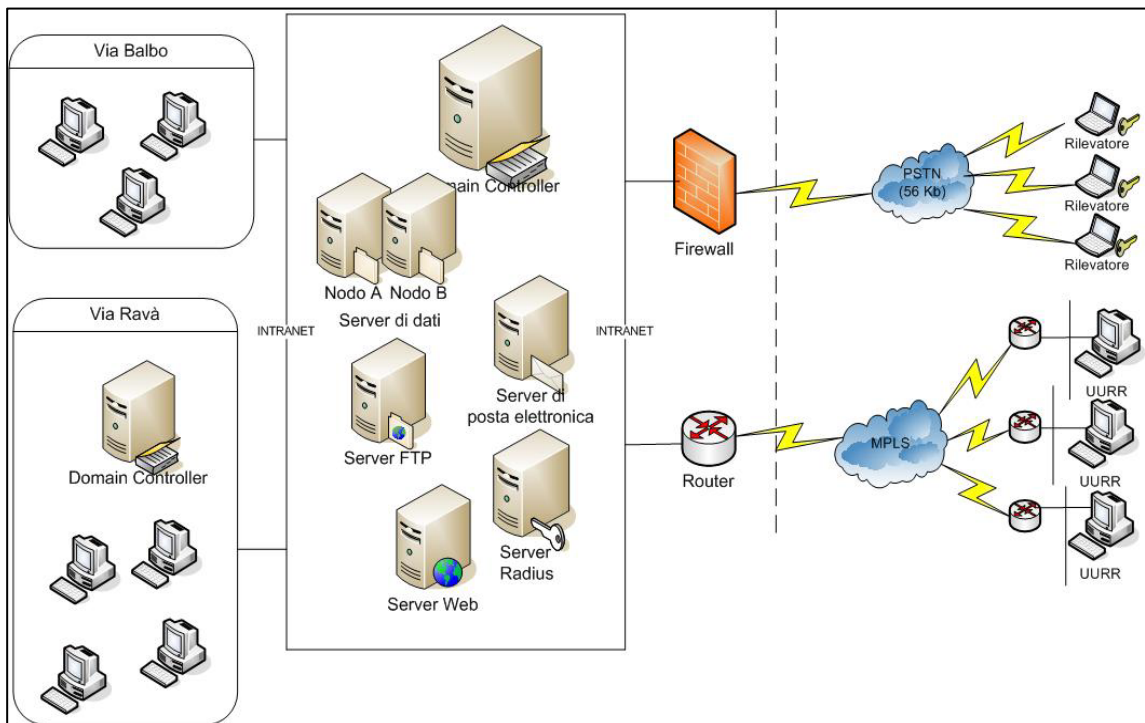
Utenti e computer appartenenti ad un dominio condividono uno stesso database che gestisce tutti i servizi. Il database e tutte le risorse sulla rete sono amministrati da uno o più domain controller sul quale sono impostati anche i criteri di protezione unificati.

Su di esso sono inoltre impostati tutti gli account utente e vi sono centralizzate tutte le operazioni di configurazione. Gli utenti sui computer client possono accedere alle risorse del dominio previa autenticazione da un server al momento dell'accesso e ad ogni utente è associata un'unica password.

Nello specifico, i servizi necessari al dominio Rcfl, ossia i server necessari a svolgere tutti i compiti richiesti sono, oltre a quello di raccolta dati, il server di posta elettronica, il server ftp, il server web ed il server radius necessario al controllo degli utenti.

Nella figura seguente viene riportato l'intero schema del dominio Rcfl.

Figura 9.4 – Il dominio rcfl



9.2.5 - Il sistema dei metadati

Metalav è il sistema informativo che tratta i metadati dell'indagine continua sulle forze di lavoro. Tale sistema è stato progettato e sviluppato all'interno del servizio Fol per documentare elettronicamente i questionari dell'indagine. In seguito all'internalizzazione del sistema Capi il

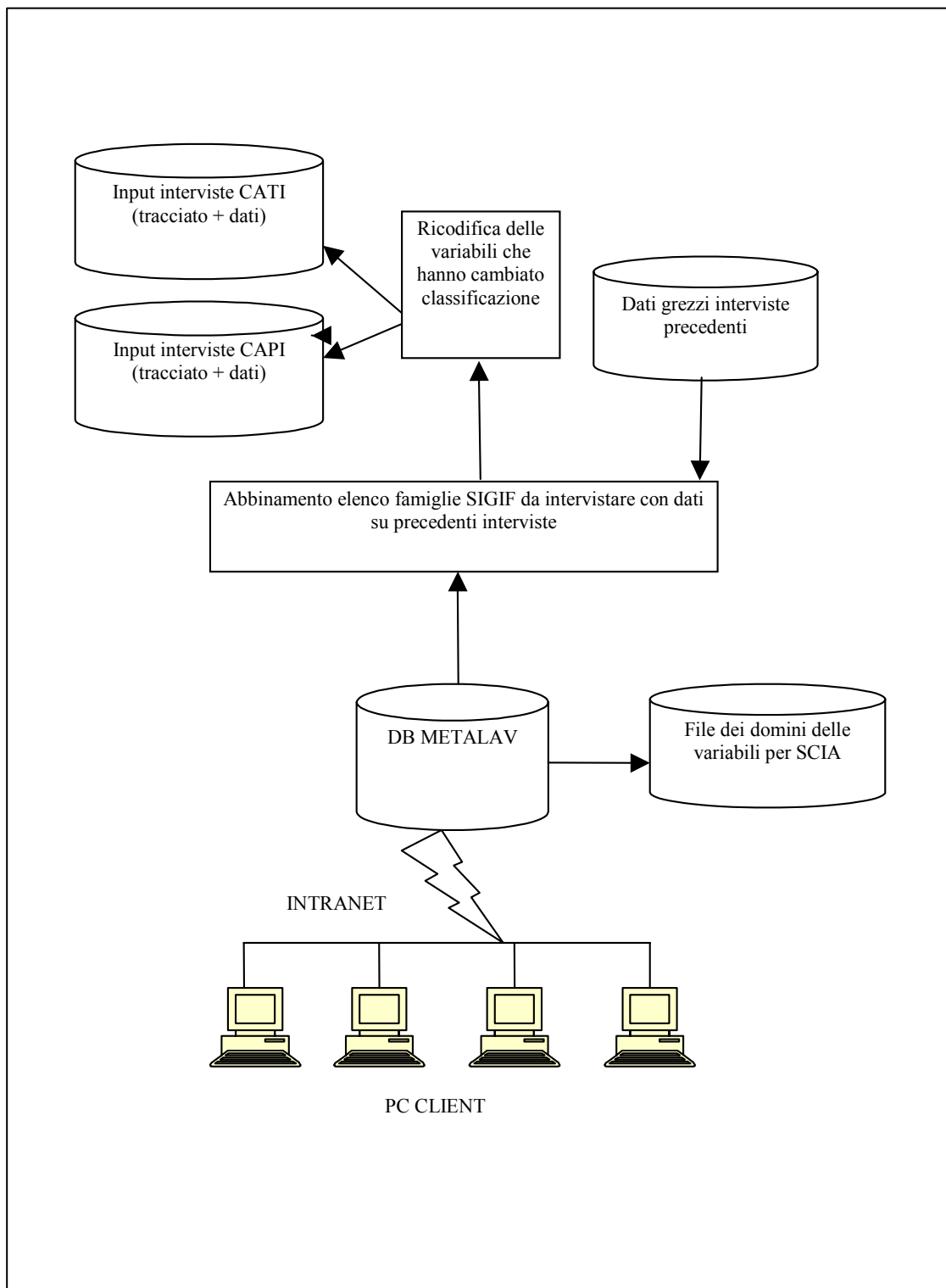
database è stato integrato con gli applicativi che ora servono per la produzione e diffusione del dato statistico ed è estremamente funzionale alla gestione del nuovo sistema direttamente gestito dall'Istat.

Il sistema ha due macro-obiettivi: il primo è documentativo ed il secondo è quello di offrire supporto alla produzione e all'analisi dei dati. Esso rappresenta uno strumento utile per chi vuole lavorare con i dati dell'indagine⁷.

Dal punto di vista della documentazione del processo di indagine, la storicizzazione delle versioni dei questionari ed il loro legame con le variabili che ne fanno parte è un aspetto importante di Metalav. Il sistema è in grado di archiviare secondo il modello logico relazionale i contenuti dei questionari (per esempio le variabili, le classificazioni, le modalità, i salti, i filtri) ed i loro legami logici (Figura 9.5).

⁷ Ad esempio, contiene i programmi di generazione automatica di codice per la lettura dei file delle forze di lavoro e per la produzione dei formati applicati alle variabili.

Figura 9.5 – Utilizzo del DB METALAV per la produzione



Per quanto riguarda la produzione, il sistema consente la produzione dei file, dei tracciati e l'elenco delle modalità relativi a: microdati, sistan e standard; la creazione dei file dei domini per SCIA; la preparazione dei formati Sas per la elaborazione dei dati e la tabulazione. Anche i file ed i tracciati per le interviste Capi e Cati vengono generati a partire dalle informazioni presenti in Metalav.

Nel corso del tempo, nell'ambito dell'indagine si registrano continui cambiamenti che riguardano i contenuti dell'indagine. Ad esempio, l'introduzione di una nuova variabile nel questionario è un evento da documentare per la successiva creazione dei file standard per gli utenti e dei file per il Sistan; inoltre si deve aggiornare il file dei domini per il check e il file dei formati per la tabulazione. Se una variabile cambia classificazione è necessario ricodificare la stessa variabile sul file dei grezzi di un trimestre precedente, al fine di somministrare un nuovo questionario a conferma. Metalav tiene conto di tutto ciò per gestire più agevolmente non solo tutti i processi di gestione dell'indagine e di trattamento dei dati, ma anche tutte le fasi successive di elaborazione legate all'indagine stessa.

Se si desidera accoppiare le informazioni anagrafiche estratte da Sigif con quelle prodotte dal sistema Capi o dal sistema Cati occorre gestire la diversa nomenclatura e tradurre in un linguaggio comune le classificazioni utilizzate dai diversi sistemi. Metalav gestisce le differenze appena citate che si evidenziano tra le diverse versioni di questionario e tra le diverse fasi elaborative. La documentazione cartacea è stata necessariamente ampliata ed organizzata in un DB elettronico in modo tale da completare la base informativa sui metadati dell'indagine e renderne il contenuto accessibile tramite programmi automatici e non procedure manuali.

Le variabili hanno una serie di attributi che possono cambiare a seconda del contesto; essi sono il nome, la posizione all'interno del tracciato, il default, il nome relativo alla procedura di estrazione delle informazioni dal sistema sigif e la lunghezza. Esse cambiano a seconda della tecnica di rilevazione, della versione di questionario e della fase di intervista.

Se ad esempio una variabile cambia *label* da un questionario ad un altro, è possibile identificarla nei diversi questionari grazie alle chiavi primarie del DB.

Se la denominazione di una variabile cambia a seconda della fase elaborativa diventa indispensabile avere una tabella per l'identificazione della variabile presente nelle varie fasi elaborative in cui la variabile è trattata.

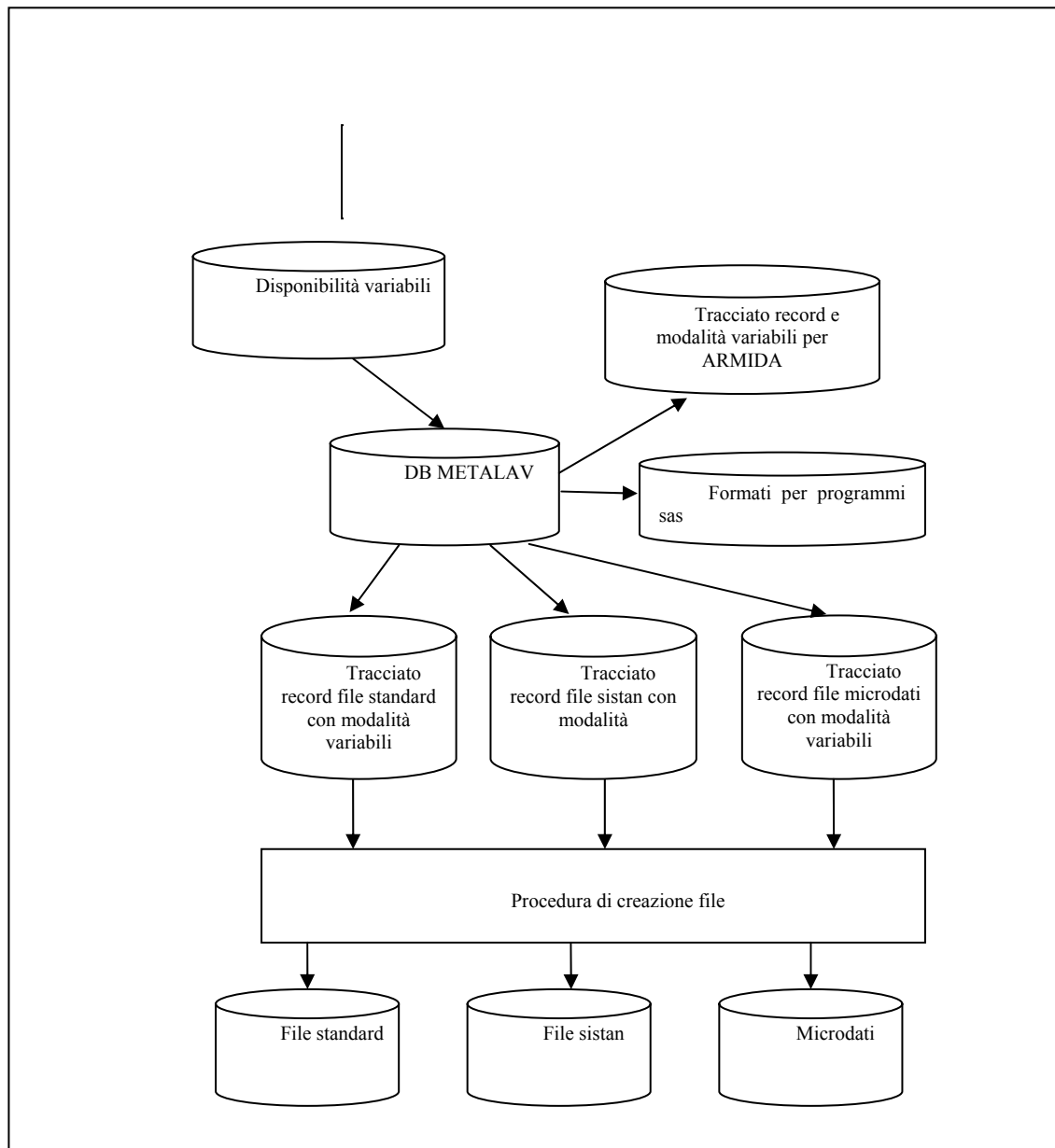
Il valore di default di una variabile cambia a seconda si tratti di prima intervista o reintervista oppure a seconda che si tratti di tecnica Capi o Cati. Anche la valorizzazione dei diversi default è memorizzata nel DB.

I dati grezzi delle precedenti interviste vengono forniti per le reinterviste al Capi o al Cati e suddivisi per settimana in base alla tipologia di tracciato e nel caso del Capi anche in base al rilevatore. La ricodifica delle vecchie classificazioni in quelle adottate dall'ultima versione di questionario è possibile grazie a Metalav che storicizza appunto tali cambiamenti: questa caratteristica del sistema evita di dare informazioni con valori fuori dominio rispetto all'ultima versione di questionario.

La gestione dei domini delle variabili è parte fondamentale anche per la diffusione che ne utilizza il contenuto per i piani di tabulazione (Figura 9.6).

Metalav, in sintesi, è l'ambiente che permette il confronto delle informazioni presenti nelle diverse fasi e spesso con nomi diversi.

Figura 9.6 – Utilizzo del DB METALAV per la diffusione



Da ultimo, va ricordato che l'accesso a Metalav è possibile dal sito Intranet del servizio Fol, dove l'utente (del Servizio Fol o di qualsiasi altro servizio dell'Istituto) può ricavare informazioni sui metadati dell'indagine e sul cambiamento di questi nel tempo attivando una serie di filtri appositamente predisposti.

L'alimentazione del DB avviene attraverso procedure automatiche che l'utente del Servizio Fol attiva interagendo con una interfaccia grafica a maschere. La crescente richiesta di ulteriori utilizzi del Db ha consigliato la migrazione dei metadati del Db dall'ambiente Windows a Unix.

In quest'ottica è stata avviata la migrazione di Metalav su Oracle che vedrà così garantite la sicurezza e l'integrità che solo un RDBMS (Relational database management System) professionale può dare.

9.3 - Il raccordo tra Sigif e il sistema Capi internalizzato

Nell'ambito dell'internalizzazione del sistema Capi si è dovuto procedere sul lato server alla trasformazione del sistema predisposto per gestire l'indagine tra il centro e gli uffici territoriali e per rendere operativi i rilevatori sul territorio. Qui di seguito si riportano le specifiche del sistema Sigif rivisto alla luce delle modifiche apportate al sistema a seguito dell'internalizzazione e le scelte operative adottate per una realizzazione efficiente ed efficace del sistema stesso.

Il sistema Capi dovrà in modo automatico e a scadenza prefissata, preparare la pianificazione delle interviste analizzando i nominativi delle famiglie da intervistare. I nominativi saranno organizzati in gruppi di quattro famiglie (quartine), di cui la prima rappresenta la famiglia di base e le altre tre le riserve (con priorità di assegnazione). All'interno di ogni quartina ci saranno una serie di indicazioni, tra cui l'area territoriale di appartenenza e lo stato delle singole famiglie. Insieme alle quartine il sistema dovrà acquisire i dati relativi ad eventuali interviste precedenti.

Le funzionalità devono prevedere l'intera gestione, da parte del supervisore regionale e del centro, del parco rilevatori: ogni transazione dovrà agire sullo "stato" del rilevatore, che risulterà opportunamente storicizzato. Il supervisore deve poter visualizzare opportune informazioni sugli stati dei rilevatori e gestirne la variazione di stato in base agli input provenienti dal field o dal centro Istat attraverso funzionalità, quali:

- attivazione: successiva alla sottoscrizione del contratto di collaborazione o al termine della causa di interruzione temporanea (l'attivazione di un rilevatore deve prevedere la possibilità di indicare la data di inizio attività lavorativa oltre alla data di sottoscrizione del contratto);
- interruzione: definitiva (il rilevatore non verrà mai più attivato) o temporanea (il rilevatore non risulta disponibile per un periodo di tempo limitato e sarà attivabile in futuro);
- rifiuto: il rilevatore non accetta l'incarico proposto;
- sostituzioni dei rilevatori: il supervisore deve poter convalidare, modificare o rifiutare le richieste di sostituzione provenienti dai rilevatori o procedere alla sostituzione senza richiesta del rilevatore. Per ogni sostituzione il supervisore indicherà i rilevatori sostituiti.

Tutti i dati disponibili nel sistema Capi relativi alla gestione del field devono essere fruibili dal centro e dai supervisori, per la parte di competenza.

In base alle aree di appartenenza, il sistema Capi dovrà suddividere i nominativi, secondo modalità parametrizzabili, tra i rilevatori che saranno attivi nelle aree in questione ed effettuare l'assegnazione agli stessi. Il sistema dovrà prendere in considerazione, per l'assegnazione dei nominativi, solo i rilevatori risultanti attivi alla data dell'esecuzione delle interviste relative ai nominativi in questione. Al termine dell'operazione di assegnazione, prima che sia effettuato l'invio ai rilevatori, il sistema dovrà avvertire i supervisori di competenza rimanendo in attesa di convalida.

Effettuata l'assegnazione automatica dei nominativi ai rilevatori, il sistema dovrà permetterne il controllo e l'eventuale modifica delle assegnazioni da parte dei supervisori

regionali. Nel caso di mancata convalida, da parte dei supervisori, il sistema in modo automatico ed a scadenza prefissata avvertirà il centro del ritardo. Le variazioni nelle assegnazioni e le mancate convalide dovranno essere storicizzate. Il centro dovrà settimanalmente visionare lo stato delle convalide.

Il supervisore deve inoltre poter redistribuire il carico di lavoro di un rilevatore ad uno o più rilevatori afferenti alla medesima area o ad aree ad essa contigue, procedendo alla riassegnazione parziale o totale dei nominativi precedentemente attribuiti al rilevatore in questione. L'attribuzione dovrà essere effettuata su base puntuale o su un range di nominativi. In base a quanto definito dal supervisore il sistema provvederà, nei successivi collegamenti, a riassegnare automaticamente i nominativi, trasferendoli ai rilevatori di nuova assegnazione ed eliminandoli dal P.c. del rilevatore a cui erano precedentemente assegnati. Le variazioni nelle assegnazioni dovranno essere storicizzate. Il Centro dovrà essere in grado di visionare lo stato delle assegnazioni e le variazioni. Tale funzionalità deve offrire al supervisore anche la possibilità di visionare la situazione del rilevatore in questione (di interviste già effettuate, interviste da effettuare, altre informazioni) al fine di ottimizzare le valutazioni.

Il sistema di gestione delle interviste realizza le funzionalità di gestione del field, assegnazione, rassegna e redistribuzione interviste tramite un'architettura basata sulla base dati Sigif e le interazioni tra utenti finale e sistema avvengono tramite form realizzate in ambiente Oracle.

L'implementazione delle funzioni preliminari, principalmente di natura algoritmica, alla verifica ed eventuale modifica delle assegnazione teorica delle interviste avviene automaticamente ed in maniera parametrizzabile tramite procedure PL/Sql. Altre funzioni fondamentali e previe all'assegnazione sono quelle di gestione del field e di dettaglio carico. Tramite esse il sistema è in grado di fotografare la situazione del field e permettere quindi agli Uffici territoriali e al centro di adattare il carico alla situazione operativa della rete di rilevazione.

Esso, inoltre, implementa le funzioni di:

- preparazione dei file di dati da spedire ai rilevatori Capi per le prime interviste;
- preparazione dei file di dati da spedire ai rilevatori Capi per le interviste successive;
- archiviazione settimanale dei dati grezzi di intervista suddivisi per tecnica sul server SP/Unix;
- archiviazione trimestrale dei dati raggruppati (solo le interviste utili) sul server SP/Unix;
- aggiornamento della base dati Sigif sugli esiti delle interviste

In questa fase le informazioni circa le famiglie da intervistare sono prelevate dal Data Base SIGIF da una procedura Php che crea un file ascii unico delle quartine con il relativo tracciato record. Questo file viene messo a disposizione per essere successivamente ripreso da una procedura SAS che crea un file per ciascuno dei rilevatore attivi. L'insieme dei file resta a disposizione su SP per la successiva trasmissione ai rilevatore. L'ambiente di archiviazione è strutturato tramite un insieme di aree di lavoro organizzate come archivio/&anno/&tecnica/&settimana dove &anno è l'anno gestionale, &tecnica è la tecnica di indagine (Capi o Cati), &settimana indica sia la settimana di riferimento che quella gestionale (sr20050103_g20050201). Siccome il numero di invii settimanali può essere superiore a 1, il nome di file da spedire contiene necessariamente anche la data e ora di creazione. Il file è di tipo Ascii a campi variabili delimitato. L'ordine delle variabili viene definito in modo da minimizzare la grandezza del file e in modo da utilizzare un tracciato unico per interviste e reinterviste.

Questa fase è simile a quella del punto 1 (ed in effetti si integra con essa) e realizza:

- l'integrazione delle informazioni estratte dalla base dati SIGIF con una serie di informazioni derivanti dalle precedenti interviste e dall'intervista effettuata l'anno precedente (solo per le wave 3 e 4) per permettere l'esecuzione dell'intervista a conferma;
- l'adattamento delle variabili da ripassare alla versione di questionario in corso di somministrazione.

L'insieme delle operazioni di questa fase viene gestito tramite una procedura Sas e PL/SQL che integra e completa quella indicata al punto 1. In questo modo il file che viene creato per ciascuno dei rilevatori conterrà sia le informazioni sulle famiglie in prima intervista sia quelle delle famiglie in seconda, terza e quarta intervista.

Per ogni trasmissione questi file possono contenere interviste appartenenti a settimane diverse e possono contenere anche dati già inviati in precedenza. E' stata implementata una procedura capace di raggruppare i file trasmessi da tutti i rilevatori e di scomporre il file completo per settimana archiviando i file risultanti nelle relative aree di lavoro (le stesse aree indicate al punto 1).

Periodicamente dai file settimanali ottenuti con entrambe le tecniche vengono estratte le interviste utili che vengono di seguito cumulate per ottenere due file trimestrali in formato sia SAS (per un uso più agevole) sia Ascii (come archivio per futura memoria): il primo sarà un file trimestrale di record familiari; il secondo un file trimestrale di record individuali. Questi file vengono archiviati sui nodi SP in un percorso, diverso da quello dei punti 1 e 3, come di seguito indicato: `grezzi/&anno/&trim/` dove `&anno` è l'anno di riferimento (non quello gestionale come nei punti precedenti), `&trim` indica il trimestre di riferimento (non quello gestionale come nei punti precedenti).

Viene poi attivata una procedura sviluppata in Visual Basic che compatta i file in arrivo dai rilevatori e produce un file settimanale cumulato delle interviste realizzate evitando le duplicazioni dei record.

Acquisiti i file dei rilevatori tramite le procedure di interscambio tra Capi server e il sistema di gestione delle interviste, una procedura realizzata in PL SQL legge i file di input, estrae gli esiti e li registra sul Database ed assegna la prossima tecnica di intervista.

L'ambiente operativo attuale è residente sui nodi SP e sistema operativo Unix/Aix. Tale ambiente deve essere comunque migrato sui nuovi sistemi di elaborazione, in fase di acquisizione, per il rinnovamento degli ambienti di elaborazione costituiti da affidabili sistemi server a memoria di massa condivisa (Storage Area Network) operanti con sistema operativo Linux.

9.4 - Il rinnovamento tecnologico dell'architettura informatica di sicurezza

Il nuovo sistema Capi, oltre a soddisfare i requisiti funzionali a fondamento dell'architettura di riferimento, deve rispondere ai requisiti di sicurezza informatica. Il sistema attualmente in esercizio basa la propria sicurezza di rete, e quindi la sicurezza delle informazioni scambiate tra rilevatori ed Istat, su un'architettura proprietaria basata su Ipsec e TripleDes. Tale soluzione, già in piedi da tre anni, presenta tutti i problemi di "obsolescenza tecnologica" (ad esempio, il software client di cifratura e autenticazione non è più supportato per Windows XP).

L'architettura viene rinnovata anche per soddisfare un requisito emergente da parte della rete di rilevazione. Molti rilevatori, infatti, soprattutto nelle regioni del Nord Italia, non sono più provvisti di linee telefoniche analogiche native, ma di altre tipologia (nuove linee Adsl, sistemi integrati forniti da diverse società di telecomunicazione) per cui il concentratore Vpn Sll ed il sistema firewall devono essere implementati per permettere la modalità accesso sicura via Internet.

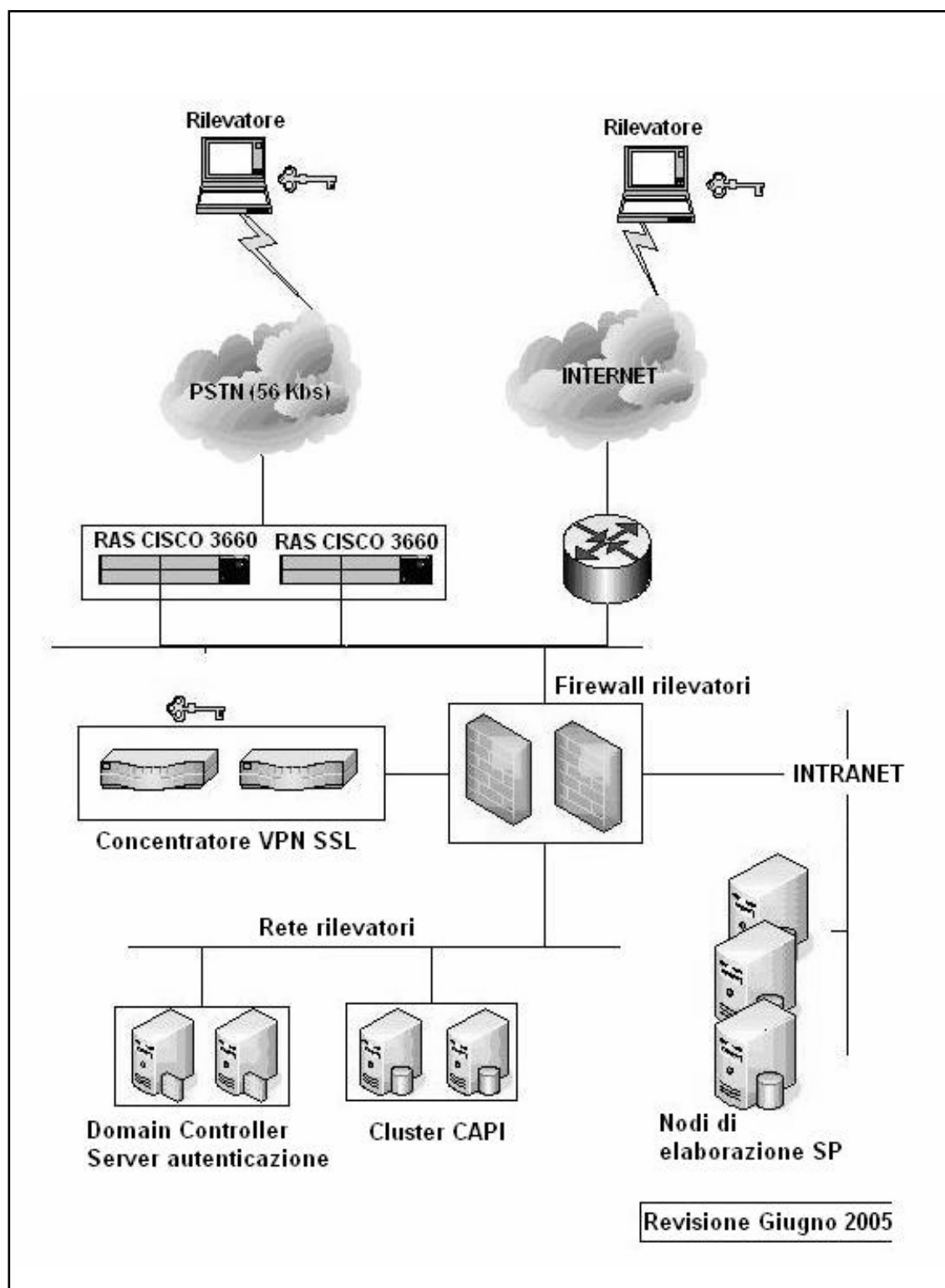
La nuova soluzione tecnologica attiva dal 2006 per la componente di sicurezza informatica è basata sulla costruzione di una Spvlan (Secure private virtual lan) realizzata tramite le architetture ssl e vpn. La soluzione prevede l'introduzione presso la sede centrale dell'Istat, a valle del sistema di accesso remoto già esistente, di un concentratore vpn ssl (socket secure layer) ridondato che garantirà l'autenticazione forte dei rilevatori per l'accesso al server Capi e la cifratura di tutti i flussi trasmessi e ricevuti. A valle del concentratore VPN sarà introdotto un firewall ridondante che garantirà il controllo degli accessi e l'ispezione dei protocolli applicativi.

Per quanto riguarda i Pc portatili utilizzati dai rilevatori (con sistema operativo Windows XP), la soluzione prevede l'introduzione di un software che inibisce l'accesso al portatile se non si è in possesso di un token crittografico (realizzato su smart-key usb) e della relativa password di accesso. Inoltre sul portatile sarà presente il software antivirus e Personal Firewall in uso presso l'Istituto. All'atto della connessione con la rete Istat, il concentratore vpn garantisce che sul PC portatile sia attivato il software antivirus e che sia stato aggiornato.

La gestione dell'ambiente operativo delle postazioni client dei rilevatori è realizzata tramite un apposito server di autenticazione integrato nel sistema di domain controller. Tale sistema (Domain controller + Server di autenticazione) è deputato alla gestione unica controllata degli ambienti operativi client, alla autenticazione applicativa completamente integrata nel sistema di accesso sicuro (concentratore ssl vpn) ed alla gestione del dominio esclusivamente dedicato ai rilevatori.

Nella figura seguente è rappresentato lo schema dell'architettura di riferimento.

Figura 9.7 – Schema architettura tecnologica sicurezza di rete



9.5 - L'internalizzazione come investimento per il servizio di produzione

Le caratteristiche intrinseche dell'indagine continua sulle forze di lavoro, in particolare l'alternanza delle tecniche di rilevazione su un panel ruotato, fanno sì che possano coesistere contemporaneamente numerose versioni del questionario elettronico utilizzate per la somministrazione delle interviste sulle unità afferenti allo stesso periodo di osservazione. Ipotizzando infatti di utilizzare ditte esterne per il supporto alla realizzazione delle interviste, si possono avere versioni del questionario diversificate per la tecnica Cati e per la tecnica Capi. Si configura poi l'esigenza di utilizzare questionari parametrizzati in base ad ogni singola fase di intervista. C'è inoltre l'esigenza di sviluppare questionari in lingua tedesca per la minoranza linguistica presente nella provincia autonoma di Bolzano. Un altro motivo che determina la proliferazione delle versioni del questionario risiede nel vincolo dovuto alla somministrazione, un trimestre all'anno, del modulo aggiuntivo (modulo ad-hoc) previsto dal Regolamento comunitario. Si somma a quanto detto la necessità di procedere a sperimentazioni di nuovi quesiti, di modifiche nel wording di quesiti già esistenti, di inserimento di modalità non previste precedentemente. Per tutti questi motivi si stima che il numero di diverse versioni del questionario elettronico utilizzate contemporaneamente in un trimestre possa arrivare anche a 20. Naturalmente ciò comporta notevoli problemi di gestione e difficoltà nella comparazione dell'uniformità dei dati rilevati.

Nell'ottica del contenimento di tali ostacoli si è proceduto, in fase progettuale, ad uniformare il più possibile i documenti di analisi che forniscono le specifiche per lo sviluppo dei software di acquisizione delle interviste. Lo sforzo per l'unificazione ha riguardato principalmente i seguenti aspetti:

- utilizzo di uno standard unico per le istruzioni di sviluppo (gestione di percorsi e salti, assegnazioni automatiche di variabili, visualizzazioni parametrizzate e form);
- utilizzo di uno standard unico per la definizione della nomenclatura delle variabili (utilizzato anche dal sistema Metalav precedentemente descritto);
- progettazione di un unico questionario parametrizzato in base alla fase di intervista.

Tenuto conto di tutta la complessità appena descritta relativa alla gestione dei questionari, le maggiori difficoltà in sede di esternalizzazione della gestione del sistema Capi e del sistema Cati hanno riguardato essenzialmente:

- il coordinamento e il controllo delle attività svolte dai diversi soggetti coinvolti;
- la tempestività delle azioni di manutenzione adattativa e correttiva del sistema.

Infatti, in assenza di gestione diretta del processo, si configura normalmente un rapporto con ditte esterne diverse per ogni tecnica d'indagine. Inoltre, trattandosi di un'indagine che non ha soluzione di continuità, potenzialmente si possono sovrapporre ad ogni scadenza di contratto due ditte diverse per ogni tecnica. Ciò porterebbe quindi ad avere quattro soggetti esterni all'Istituto che lavorano contemporaneamente nei diversi aspetti dell'indagine in corso. Simili situazioni possono generare problematiche logistiche di notevole complessità. Ad esempio, sono estreme le difficoltà organizzative che si attuano per la realizzazione dei test sul software, richiedendo spesso:

- una complessa ed estremamente onerosa allocazione delle risorse umane addette all'effettuazione dei controlli;
- notevoli spostamenti sul territorio tra una ditta e l'altra;
- il possesso da parte del servizio di produzione di una formazione tecnica diversificata a seconda dell'ambiente di sviluppo utilizzato dalle diverse ditte.

La compresenza di ditte diverse (e quindi anche di strumenti di rilevazione diversi) possono avere ricadute negative sul field di rilevazione e di conseguenza sull'andamento dell'indagine, dovute principalmente al possibile utilizzo contemporaneo di due diversi questionari elettronici e a due possibili diversi sistemi di trasmissione. Ciò può determinare la necessità di procedere spesso a sessioni di formazione e training per gli intervistatori Capi e Cati.

Un altro fondamentale problema consiste nella estrema complessità di raccordo tra i flussi di scambio dei dati tra le diverse tecniche e, trasversalmente alle tecniche, tra i diversi soggetti.

Considerato tutto ciò, a partire dalle scelte generali precedentemente esplicitate l'internalizzazione del sistema Capi ha seguito nel concreto diverse fasi di sviluppo: l'intero sistema in tutti i suoi aspetti è stato sperimentato nel corso dell'anno 2005 ed è entrato in esercizio all'inizio dell'anno 2006. Sono state previste inoltre le sperimentazioni per la parziale internalizzazione della componente Cati (sviluppo interno del questionario e del sistema di schedulazione con gestione degli stessi da parte di ditta esterna), che entrerà in esercizio entro l'anno 2007.

La concentrazione in un unico ambiente di sviluppo fa sì che l'organizzazione delle risorse risulti ottimizzata per il servizio di produzione. Le operazioni di test e controllo vengono effettuate in un unico ambito e risultano per questo meno gravose. Nello stesso tempo possono essere contenute le criticità dovute al raccordo dei flussi di scambio tra le diverse tecniche di indagine.

Di conseguenza, per il servizio di produzione sono numerosi i vantaggi introdotti con l'internalizzazione, tra cui si possono elencare:

- il superamento delle difficoltà legate al grado di conoscenza dell'ambiente di sviluppo utilizzato dalla ditta esterna, specie per quanto riguarda i percorsi complessi. Spesso infatti, quando la gestione è affidata ad una ditta esterna, deve essere concordato con la ditta stessa un *escamotage* tecnico che permetta di risolvere eventuali limiti tecnici dichiarati del prodotto utilizzato, senza poter offrire soluzioni alternative;
- la possibilità di controllare già durante la fase di test il contenuto di variabili e contatori registrati. Tutti i software che gestiscono l'implementazione di questionari elettronici si poggiano per l'archiviazione dei dati su una piattaforma proprietaria; per questo motivo, in presenza della ditta esterna che fa da intermediario non è immediato l'accesso a tali dati e quasi sempre è necessario richiedere alla ditta lo sviluppo di ulteriori procedure per la loro lettura;
- gli interventi di manutenzione adattativa e correttiva del software sono tempestivi e a completo controllo del servizio di produzione. In presenza della ditta esterna la tempistica di realizzazione degli stessi interventi e il loro impatto sul software sono valutati unicamente dalla ditta stessa e devono quindi essere concordati con questa;
- una quota non trascurabile di risorse umane del servizio di produzione hanno avuto una possibilità di crescita professionale e di valorizzazione all'interno del servizio stesso. Tutte le attività di manutenzione adattativa ed evolutiva del questionario elettronico sono infatti attualmente di competenza del servizio e, in prospettiva, tutto il controllo del sistema sarà nelle mani del servizio.

Con particolare riferimento all'internalizzazione del sistema Capi, è doveroso fare un breve accenno anche ai vantaggi ottenuti relativamente al sistema di trasmissione. La gestione in-house delle problematiche tecniche permette infatti:

- di lavorare costantemente sull'ottimizzazione dei protocolli di trasmissione (in termini di affidabilità, sicurezza e velocità);
- una maggiore efficacia del monitoraggio delle trasmissioni;
- una maggiore tempestività ed incisività degli interventi di manutenzione;
- la possibilità di tenere direttamente sotto controllo le componenti client da parte dei gestori del server (per interventi diretti sui pc dei rilevatori, per la verifica di eventuali contestazioni sulle interviste, per gli aggiornamenti generalizzati delle componenti software necessarie al funzionamento del client).

La complessità dell'indagine determina, come già illustrato, un sistema informativo-informatico di supporto molto articolato. Ciò ha comportato un notevole sforzo nel coordinamento delle azioni da intraprendere per mettere in comunicazione i vari sottosistemi nell'ambito del processo di internalizzazione della componente Capi del sistema.

A tale scopo, come già detto nel paragrafo sul raccordo tra il Sigif e il sistema Capi internalizzato, sono stati utilizzati diversi ambienti e linguaggi, che si interfacciano tra di loro attraverso flussi di comunicazione di dati e parametri. Per tali flussi si utilizza lo standard Ascii. Normalmente i flussi di dati sono accompagnati da un altro file Ascii che ne definisce il tracciato. Questo standard, facilmente utilizzabile in ogni ambiente, ha permesso il colloquio parametrico tra le diverse componenti che entrano in gioco.

Nello stesso tempo è stata consolidata nel servizio la gestione dell'intero processo proprio a seguito della ricognizione dei raccordi tra le varie parti del sistema. Anche questo processo ha comportato un'analisi che ha a sua volta contribuito alla crescita professionale di risorse interne al servizio, esplorando in larga misura la possibilità di riconversione delle varie professionalità interne al servizio.

Per i motivi sopra esposti si arriverà, al momento della conclusione del processo di internalizzazione di tutto il sistema Cati-Capi come qui è stato descritto, ad un altissimo livello di controllo da parte dell'Istituto dell'intero processo di produzione. Più precisamente, la supervisione delle fasi essenziali del processo sarà di esclusiva competenza del servizio di produzione. Le risorse addette alla supervisione di queste fasi inseriscono nel processo, proprio per la loro raggiunta competenza e professionalità, standard elevati di capacità di sviluppo e controllo. L'investimento in tale direzione è stato generalizzato ed ha riguardato profili e competenze di diversa natura.

Il consolidamento all'interno del servizio dei ruoli professionali destinati al controllo del processo fa sì che gli interventi correttivi ed evolutivi possano essere attuati in maniera estremamente più tempestiva rispetto al passato. Inoltre, il passaggio dal controllo del prodotto al controllo totale del processo determina un notevole miglioramento della qualità dell'indagine e quindi, in conclusione, della qualità del dato rilevato.

Bibliografia

- Aa.vv. *Guida all'utilizzo di TRAMO-SEATS per la destagionalizzazione delle serie storiche*, Roma:Istat, 2000. (Documenti, n.4).
- Abbate, Corrado. "La completezza delle informazioni e l'imputazione da Donatore con Distanza Mista Minima". Il prodotto Rida, Istat,1996.(Documento interno).
- Amendola, A., F.E., Caroleo, G., Coppola. *Differenziali territoriali nel mercato del lavoro e sviluppo in Italia*. Celpe, Università di Salerno 1997 .(Discussion Paper n. 36)
- Appel, M. V., II W.L., Nicholls, W. Nicholls. *New CASIC Techonology at the U.S. Census Bureau*. ASA, U.S. Census Bureau, vol.2 pp.1079-1084, 1993.
- Appel, M. V., W. L., Nicholls II, Nicholls, W. *New Casic Technology at the U.S. Census Bureau*. ASA, U.S. Census Bureau, vol. 2, pp. 1079-1084, 1993.
- Baffigi, A. "I differenziali territoriali nella struttura dell'occupazione e della disoccupazione". Un'analisi con dati a livello provinciale (1981-1995) in *Struttura della contrattazione differenziali salariali e occupazione in ambiti regionali*, di Biagioli, M., Caroleo, F.E., Destefanis, F.E. , 1999.
- Baker, R.P., N.M.,Bradburn, R.A Johnson.. *Computer-assisted personal interviewing: an experimental evaluation of data quality and cost*. 1995. (Journal of Official Statistics vol.11 n.4, pp. 413-431) .
- Balbi, S.e S Balzano. *Controlli logici per indagini ripetute: l'indagine sugli sbocchi occupazionali dei laureati di Economia a Napoli ("Federico II")*., Padova: Captor, 2000.
- Balbi, S., N.C., Lauro, G., Scepi,. *Multidimensional Data Analysis and Experimental Design, Proceedings of the 49-th ISI Session*. 1993. (Contributed Papers, 1, 67-68).
- Balbi, S., N.C., Lauro, G., Scepi,. *Multiway Data Analysis Technique for Comparing Surveys, Methodologica*. 3, 79-90,1994.
- Balbi, S., R.,Verde."Structuring Questionnaires as Symbolic Objects: a New Tool for Improving Data Quality in Surveys". In *Third International Meeting on New Techniques and Technologies for Statistics* . Sorrento : 1998.
- Ballin M., P.D., Falorsi, G., Brancato, T.,Capponi, A.,Rubino e P.,Vicari. *Le strategie adottate per la prevenzione ed il trattamento delle mancate risposte totali nell'indagine Long Form2000*.
- M. Ballin, Falorsi,P.D., Falorsi ,S. e Pallata, S. (2000) *Il trattamento delle mancate risposte totali nelle indagini Istat sulle famiglie e sulle imprese: soluzioni attuali e linee di ricerca*.
- Barcaroli, G. *Un approccio logico formale al problema del controllo e della correzione dei dati statistici*.. Istat, 1993. (Quaderni di Ricerca, n.9)
- Barcaroli, G., O., Luzi, A.,Mannari, "Metodi e software per il controllo e la correzione dei dati". Istat, 1996. (Documento interno).
- Barcaroli, G., M.,Venturi, DAISY (Design, Analysis and Imputation System): *Structure, Methodology, and First Applications*. In: J. Kovar and L. Granquist, (eds.) *Statistical Data Editing, Volume II*, U.N. 1997 (Economic Commission for Europe, 40-51).

- Bethel, J.. *Sample Allocation in Multivariate Survey*. 1989(Survey Methodology, 15, pp 47-57) .
- Box, G. E. P., G. M ,Jenkins. *Time Series Analysis: Forecasting and Control*. Holden Day, San Francisco: 1970.
- Bradburn, N.M., et al. *A comparison of computer assisted personal interviews with personal interviews in the National longitudinal survey of labour market behavior youth cohort*. Proceedings of the Annual Research Conference, Washington, US Bureau of the Census, pp. 389-397, 1991.
- Buratta, V., L.L., Sabbadini, E., Fortunato. *Manuale di tecniche di indagine, il questionario: progettazione, redazione e verifica*. Istat, 1989 (Note e Relazioni) .
- Camporese, R., S., Gazzelloni, P, Piergentili. *Process quality control to prevent non sampling errors in Italian Multipurpose system of social surveys*. NTTS Conference, June 2001.
- Calcioli, R. *Il controllo dei dati nell'indagine continua delle forze di lavoro: analisi delle regole di incompatibilita' implementate nel software di acquisizione dei dati..* Istat, 2003.(Documento interno)
- Ceccarelli, C., A.R., Discenza, S., Loriga.. *The Impact of the new Labour Force Survey on the employed classification, Classification and data analysis 2005 – Book of short papers*, Meeting of the Classification and data analysis group of the Italian Statistical Society. 65:68, MUP, Parma: 2005.
- Chen, C., Liu, L., Joint. *Estimation of Model Parameters and Outlier Effects in Time Series*, 1993 (Journal of the American Statistical Association. 88, 284-297).
- Cicchitelli, G., A,Herzel , G.E., Montanari. *Il Campionamento Statistico*. Il Mulino, 1992.
- Cochran, W.,G.. *Sampling Technique*. 3rd ed., Wiley, New York:1977.
- Commission Regulation (EC) no 1575/2000 of 9 July 2000 implementing Council Regulation (EC) no 577/98 on the organisation of a labour force sample survey in the Community concerning the codification to be used for data transmission from 2001 onwards. (Official Journal of the European Communities del 20 luglio 2000).
- Commission Regulation (EC) No 1897/2000 of 7 September 2000 implementing Council Regulation (EC) No 577/98 on the organisation of a labour force sample survey in the community concerning the operational definition of unemployment.
- Commission Regulation (EC) no 2104/2002 of 28 November 2002 adapting Council Regulation (EC) no 577/98 on the organisation of a labour force sample survey in the Community and Commission Regulation (EC) no 1575/2000 implementing Council Regulation (EC) no 577/98 as far as the list of education and training variables and their codification to be used for data transmission from 2003 onwards are concerned (Official Journal of the European Communities del 29 novembre 2002).
- Commission Regulation (EC) no 246/2003 of 10 February 2003 adopting the programme of ad hoc modules, covering the years 2004 to 2006, to the labour force sample survey provided by Council Regulation (EC) no 577/98 (Official Journal of the European Communities del 11 febbraio 2003).

Commission Regulation (EC) no 430/2005 of 15 March 2005 implementing Council Regulation (EC) no 577/98 on the organisation of a labour force sample survey in the Community concerning the codification to be used for data transmission from 2006 onwards and the use of a sub-sample for the collection of data on structural variables (Official Journal of the European Communities del 17 marzo 2005).

Council Regulation (EC) no 577/98 of 9 March 1998 on the organisation of a labour force sample survey in the Community(fficial Journal of the European Communities del 14 marzo 1998).

Couper, M.P.. *Changes in Interview Setting Under Capi*. JOS, 1996. (Statistics Sweden, vol.12, n.3 pp.301-316).

D'Ambra, L.. *Alcune estensioni dell'analisi in componenti principali per lo studio di sistemi evolutivi*. Uno studio sul commercio internazionale dell'elettronica. (Ricerche Economiche, n.2).

Dagum ,E., B.. *The Effects of Asymmetric Filters on seasonal Factor Revisions*. 1982.(Journal of the American Statistical Association, 77, 732:738) 2.

De Vitiis, C., L.,Di Consiglio, S.,Falorsi . *Studio del disegno campionario per la nuova rilevazione continua sulle Forze di Lavoro*. 2005.(Contributi Istat, n. 6).

De Vitiis, C., S., Falorsi,. *Indagine Forze di Lavoro: Studio della nuova strategia di campionamento*.1998. (Documento interno al progetto interarea).

De Vitiis, C., L., Di Consiglio, S.,Falorsi.*Indagine Forze di Lavoro: valutazione degli errori di campionamento nell'indagine attuale*. marzo 1999. (Documento interno al progetto interarea).

De Vitiis C., L., Di Consiglio, S., Falorsi. e D., Pagliuca. 2000 *Indagine sulle forze di lavoro: disegni campionari per l'indagine continua 2000*.

Decressing, J., A.,Fatàs, *Regional Labor Market Dynamics in Europe*.1995 (European Economic Review, n. 3, 1627-1655).

Den Butter, F.A.G., Fase, M.M.G.. *Seasonal Adjustment as a Practical Problem*. North Holland, Amsterdam:1991.

Deville, J.C., C. E.,Särndal, *Calibration Estimators in Survey Sampling*. 1992.(Journal of the American Statistical Association, vol. 87, pp. 376-382).

Deville J. C., C. E, Särndal. *Calibration Estimator, in Survey Sampling* 1992 (Journal of the American Statistical Association, vol. 87, pp.376-382).

Di Laurea, D., R., Gatto, A., Righi, A., Spizzichino. "Ricostruzione/riallineamento delle serie del mercato del lavoro a livello regionale 1977-2003". *Acts of the VI meeting Giornate di Studio sulla Popolazione*, SIS, Italian Society of Statistics, Padova :16-18 febbraio 2005.

Dippo, C., A.,Polivka, D., Kostanich. *Effects of Methodological Change in the Current Population Survey*.1994. (BLS Working Papers) .

Escoufier, Y., et al.. *Statistique et analyse des données*. (Bulletin des Statisticiens Universitaires, vol.10, n.1).

Estevao V., M. A., Hidiroglou C. E, Särndal. *Methodological Principles for a Generalized Estimation System at Statistics Canada*. C. E 1995 (Journal of Official Statistics, Vol.11 No. 2, pp.181-20)4.

- European Communities. *The European Union labour force survey, Methods and definitions – 2001*. 2003
- Eurostat. *Standard Quality Report*. May 2002.
- Eurostat. *Indagine sulle forze di lavoro (metodi e definizioni)* 1992
- Fabbris, L.. *Abbinamenti tra fonti d'errore nella formazione dei dati e misure dell'effetto degli errori sulle stime*. 1991. (Bollettino SIS n. 22).
- Fabbris L., F., Bassi. *On-line likelihood controls in Computer-Assisted Interviewing.*, Book 1, 51° sessione ISI, Istanbul: 1997.
- Fabbris L. *Rilevazione di dati assistita da computer o per via telematica nelle indagini longitudinali prospettive su famiglie e imprese*.2001. (Rivista di statistica ufficiale n. 2) .
- Fabbris, L.. *L'indagine campionaria, Metodi, Disegni e tecniche di campionamento*. NIS, 1989.
- Falorsi, et al. 2001.(Quaderno ric.) .
- Falorsi, P.,D., M., Ballin, C., De Vitiis, G., Scepi., *Principi e metodi del software generalizzato per la definizione del disegno di campionamento nelle indagini sulle imprese condotte dall'Istat*.1998 (Statistica Applicata ,Vol. 10, n.2).
- Falorsi, P.,D., S.,Falorsi. “Indagine sulle forze di lavoro: descrizione della strategia di campionamento e valutazione dell’errore campionario dei principali indicatori provinciali del mercato del lavoro”. 1996.(Documenti Istat).
- Falorsi, P. D. e S.,Falorsi. “Un Metodo di Stima Generalizzato per le Indagini sulle Imprese e sulle Famiglie” . 2,1996. (Documenti Istat).
- Falorsi, S. e C., Rinaldelli. *Un software generalizzato per il calcolo delle stime e degli errori di campionamento*. 1998 (Statistica Applicata, Vol. 10, N° 2, pp.217-234).
- Fellegi, I.,P., D., Holt. *A Systematic Approach to Automatic Edit and Imputation*. 1976(Journal of the American Statistical Association, 71, 17-35).
- Filippucci, C. *La rilevazione dei dati assistita da computer: acquisizioni e tendenze della metodologia statistica e informatica.*, Sorrento: XXXIX Riunione Scientifica della S.I.S, 1998.
- Futterman, M.. *Cati Instrument logical structures: an analysis with applications*. JOS Statistics Sweden, vol.4, n.4 pp.333-348, 1988.
- Futterman, M.. *Cati Instrument logical structures: an analysis with applications*, JOS 1988 (Statistics Sweden, vol. 4, no. 4, pp. 333-348).
- Gatto, R. *Report on the experiment on the proposed methodology for the resion of time series*.ISTAT, mimeo,1 2004.
- Gatto, R.. *Sulla ricostruzione delle serie dei principali indicatori del mercato del lavoro*. PhD Thesis, Dipartimento di Scienze Economiche, University of Rome “La Sapienza”, 2 2004.
- Gatto, R., P., Gennari, N.,Massarelli. “La ricostruzione e il riallineamento delle serie storiche delle forze di lavoro 1984 – 1992” In *Acts of the meeting occupazione e disoccupazione in Italia: misura e analisi dei comportamenti*. Murst (Ministry of University and Research), Bressanone :15-16 gennaio 2001.

- Giusti, A., M., Pratesi. *An application of new technologies in data capture: an evaluation of Cati survey*. NTTS-95, Bonn: 1995.
- Giusti, A., M., Pratesi. *Nuove tecnologie nel processo di produzione dei dati statistici*. Rimini: XXXVIII Riunione Scientifica della S.I.S., 1997.
- Golinelli, R., M., Monterastelli. *Un metodo di ricostruzione di serie storiche compatibili con la nuova contabilità nazionale. (1951-89)*. 1990. (Nota di lavoro di Prometeia, n. 9001).
- Gomez, V., A., Maravall. *Guide for Using the Programs TRAMO and SEATS*. Beta Version.
- Gomez, V., A., Maravall. *Programs TRAMO and SEATS: Instructions for the User*. Bank of Spain, 1997.
- Grassia, M. G., A., Martini. "Il sistema informativo del monitoraggio e le innovazioni di qualità nel trattamento di esiti, tassi ed indicatori". In *Il monitoraggio del processo e la stima dell'errore nelle indagini telefoniche*, Muratore, M. G., Sabbadini, L. L., Signore, M. Applicazioni all'indagine sulla Sicurezza dei cittadini, ISTAT, (Metodi e norme). Being printed
- Grassia, M.G., M., G., Muratore. *The contribution of symbolic objects theory to errors prevention in Cati questionnaires*. IV International Seminar on New Techniques and Technologies – NTTS, Creta: 2001.
- Groves, R., M., et al. *Telephone Survey Methodology*. New York: John Wiley, 1988.
- Hamilton, J.D.. *Time Series Analysis*. Princeton: Princeton University Press, 1994.
- Harvey, A.C.. *Time Series Models*. Hemel Hempstead: Harvester Wheatsheaf. Cambridge University Press, 1993.
- Hendry, D., F.. *Dynamic Econometrics*. Oxford: Oxford University Press.
- Iannucci, L., L., Quattrociochi, S., Vialetti.. *A quality control approach to C.A.T.I. operations in Safety of citizen survey: the non-response and substitution rates monitoring*. NTTS Conference, Novembre, 1998.
- Istat., *Forze di lavoro: disegno dell'indagine e analisi strutturali*. 1991 (Annali di statistica, serie IX – Vol. 11).
- Istat. *La nuova rilevazione sulle forze di lavoro: Contenuti, metodologie, organizzazione*. Roma, 3 giugno 2004, http://www.istat.it/dati/pubbsci/contributi/Contr_anno2005.htm/2005_16.pdf
- Pagliuca, D." cur." *Genesees ,v. 3.0 . Funzione Riponderazione*. Roma: Istat, 2005. (Tecniche e Strumenti)
- Kalton, G., D., Kasprzyk.. *Proceedings of the Section on Survey Research Methods*. Imputing for Missing Survey Responses .American Statistical Association, 22-31, 1982.
- Keller, W.J.. *Changes in Statistical Techonology*. JOS, 1995. (Statistics Sweden, vol.11).
- Kish L. *Sampling with Unequal Probabilities*, 2002.
- Little, R.J.A.. *Models for Nonresponse in Sample Surveys*, 1982. (Journal of the American Statistical Association, 77, 237-250).
- Marbach, G.. *Le ricerche di mercato*. UTET, 2000.
- Masselli, M.. *Manuale di tecniche di indagine: il sistema di controllo della qualità dei dati* Roma ISTAT, 1989 . (Note e Relazioni n. 1).

- Mauro, P., A., Spilimbergo. *How Do the Skilled and the Unskilled Respond to Regional Shocks?* IMF, 1-1999. (Staff Papers, vol.46).
- Mockovak, W.P.. *Issues and steps involved in designing a questionnaire for computer-assisted interviewing*. International Conference on "Computer-assisted survey information collection", S. Antonio, Texas: 1996.
- Muratore, M. G., L. Quattrociochi. *A new quality approach for telephone survey: from interviewers training to daily monitoring system to prevent and to measure non response*. Online publication on the Joint Program in Survey Methodology site www.jpsm.umd.edu/icsn99, 1999.
- Muratore, M. G., L. Quattrociochi. "L'Indagine sulla Sicurezza dei Cittadini" in. *Indagini sociali telefoniche: metodologie ed esperienze della statistica ufficiale (Telephone social surveys: methodologies and experiences of official statistics)*, Muratore, M. G., L., Quattrociochi, L.L., Sabbadini, Istat, (Metodi e norme). (being printed).
- Muratore, M. G., L.L., Sabbadini. "Molestie e violenze sessuali nelle indagini dell'ISTAT: metodologia, organizzazione, principali risultati (Sexual molestation and rapes in ISTAT surveys: methodology, organisation, chief results)". In *Proceedings of the Conference, Le famiglie interrogano le politiche sociali*, Bologna: 29-31 March 1999, published on CD-ROM.
- Nicholls, W.L., R.P., Baker, J., Martin. "The effects of new data collection technologies on survey data quality". In *Survey Measurement and Process Quality*, Lyberg, L. et al., (eds) J.Wiley & Sons, New York: 1997.
- Nicholls, W. L., R.P., Baker, J., Martin. "The effects of new data collection technologies on survey data quality". In *Survey Measurement and Process Quality*, Lyberg, L., et al. (eds), J. Wiley & Sons, New York: 1997.
- Overman, H. G., D., Puga. *Unemployment Clusters Across Europea Regions and Countries*. Working Papers UT-ECIPA-DPUGA-99-03, 1999.
- Pench, L.R., P., Sestito, E., E., Frontini. *Some Unpleasant Arithmetics of Regional Unemployment in the EU. Are there Any Lessons for EMU*. Unione Europea, DG XII, Brussel, 1999.
- Piazza, T. *New methodological possibilities offered by computer assisted interviewing*, Book 1, 51° sessione ISI, Istanbul: 1997.
- Planas, C.. *Applied Time Series Analysis: Modelling, Forecasting, Unobserved Components Analysis and the Wiener-Kolmogorov Filter*. Eurostat, 1997. (Statistical document).
- Polivka, E., Anne, Miller, M., Stephen. *The CPS after the Redesign: Refocusing the Economic Lens*. in Labor Statistics Measurement Issues, John Haltiwanger, Marilyn E. Manser and Robert Topel, editors. NBER Chicago: The University of Chicago Press, 1998.
- Polivka, Anne, E.. *Comparisons Of Labor Force Estimates From the Parallel Survey And the CPS During 1993: Major Labor Force Estimates*. CPS Overlap Analysis Team Technical Report 1. BLS, 1994.
- Quintano C. e R., Castellano. *Strategie di trattamento delle mancate risposte per la qualità in alcune indagini dell'ISTAT*, 2000.
- Presidenza Del Consiglio Dei Ministri, Commissione per la Garanzia Dell'informazione Statistica. *Gli impegni connessi alle indicazioni comunitarie nell'area delle statistiche del lavoro, Rapporto finale*.

Presidenza Del Consiglio Dei Ministri, Commissione per la Garanzia Dell'informazione Statistica *Analisi delle procedure di correzione/imputazione utilizzate dall'Istat nelle principali indagini sulle famiglie 2000*.

Regulation (EC) no 1991/2002 of the European Parliament and of the Council of 8 October 2002 amending Council Regulation (EC) no 577/98 on the organisation of a labour force sample survey in the Community (Official Journal of the European Communities del 9 novembre 2002)

Regulation (EC) no 2257/2003 of the European Parliament and of the Council of 25 November 2003 amending Council Regulation (EC) no 577/98 on the organisation of a labour force sample survey in the Community to adapt the list of the survey characteristics (Official Journal of the European Communities del 23 dicembre 2003)

Rinaldelli C. *Indagine sulle Forze di Lavoro – procedura generalizzata di stima per il calcolo dei pesi per individui e famiglie*. 1998 (Documento interno).

Riccini , E., Margarucci, P., Floris. *Controllo e correzione dati*. Istat, Roma: 2000(Manuale utente)

Rubin, D.B. *Inference and Missing Data*. 1976.(Biometrika, 63, 581-592).

Saris, W.E., *Computer- Assisted Interviewing*. Newbury Park: 1991.

SAS Institute Inc. SAS OnlineDoc®, Version 8, Cary, NC: SAS Institute Inc. 1999.

Sautory O. *Calibration on known marginal counts for sample surveys: practical experiences at INSEE*. Paper presentato al Workshop on Uses of Auxiliary Information in Surveys, Orebro: 1992

Singh A. C. , C. A., Mohl, *Understanding Calibration Estimators in Survey Sampling*. (Survey Methodology, vol. 22, No 2, pp. 107-115). 1996

SIS. *Contributi della Commissione Scientifica sulla qualità dei dati* 1991. (Bollettino n. 22)

Sperry, S., et al.. *Computer-assisted personal interviewing on the current beneficiary survey*. Conference of the American Association of Public Opinion Research, Phoenix, 1991.

Statistics Canada, *Quality Guidelines*, fourth edition, October 2003.

The American Association For Public Opinion Research. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Ann Arbor, Michigan: AAPOR.2000.

Venturi, M. *La ricostruzione dei microdati della “vecchia” indagine sulle forze di lavoro secondo la nuova impostazione metodologica. Sperimentazione sui dati di luglio 1992*. ISTAT, mimeo, 1993.

Weeks, M.F.. *Computer Assisted Survey Information Collection: A rievolution of CASIC Methods and their Implication for Survey Operations*. JOS 1992. (Statistics Sweden, vol.8 n.4, pp. 445-465).

Willenborg, L.C.R.J., *Computational Aspects of Survey Data Processing*, CWI Tract. 1988.

Willenborg, L.C.R.J., *Contribution to a Methodology of Computer aided Questionnaire Design*. Netherlands Central Bureau of Statistics, 1986.

Willenborg, L.C.R.J.. *Recognition of Records with a Correct Routing Structure*. Internal Report, Netherlands Central Bureau of statistics, 1985.

Willenborg. L.C.R.J.. *Two Characterizations of the Routing Structure in a Questionnaire: Balance and Complexity*. Internal report, Netherlands Central Bureau of Statistics, Herleen, 1985.

Woodruff ,R.S. *A Simple Method for Approximating the Variance of a Complicated Estimate*. (Journal of the American Statistical Association, vol. 66, pp. 411-414).

Zannella, F. “Metodologia, programmi e sperimentazioni relativi alla progettazione di una procedura generalizzata per la stratificazione dei comuni”. *In Atti della giornata di studio sul campionamento statistico*. Istat,1991 (Annali di statistica ,serie IX).



Stampato da consorzio R.T.I.
C.S.R. S.r.l. - System Graphic S.r.l
Settembre 2006 – copie 750

Metodi e Norme - Nuova serie - Volumi pubblicati

Anno 2000

6. *L'indice del costo della vita valevole ai fini dell'applicazione della scala mobile delle retribuzioni. Dalle origini alla cessazione (1945-97)*
7. *Le nuove stime dei consumi finali delle famiglie secondo il Sistema Europeo dei Conti SEC95*

Anno 2001

8. *La nuova indagine sulle cause di morte. La codifica automatica, il bridge coding e altri elementi innovativi*
9. *Il settore delle costruzioni in contabilità nazionale. I nuovi standard europei dettati dal SEC95*
10. *Indagini sociali telefoniche. Metodologia ed esperienze della statistica ufficiale*
11. *Elenco dei comuni al 31 maggio 2001* 
12. *Classificazione delle professioni* 

Anno 2002

13. *Le statistiche culturali in Europa*
14. *Gli investimenti lordi di contabilità nazionale dopo la revisione: nota metodologica*
15. *Panel Europeo sulle famiglie*


Anno 2003

16. *Metodi statistici per il record linkage*
17. *Metodologia e organizzazione dell'indagine multiscopo sulla domanda turistica "Viaggi e vacanze"*
18. *Classificazione delle attività economiche. Ateco 2002*


Anno 2004


19. *Inventario sulle fonti e metodi di calcolo per le valutazioni a prezzi costanti - Italia*
20. *Metodologia e tecniche di tutela della riservatezza nel rilascio di informazione statistica*
21. *Metodologia di stima degli aggregati di contabilità nazionale a prezzi correnti*
22. *Numeri indici dei prezzi alla produzione dei prodotti industriali venduti sul mercato interno - Base 2000=100*


Anno 2005

23. *I conti economici nazionali per settore istituzionale: le nuove stime secondo il Sec 95* 
24. *La rete di intervistatori Capi dell'Istat per la conduzione dell'indagine continua sulle Forze di Lavoro*
25. *Il monitoraggio del processo e la stima dell'errore nelle indagini telefoniche*
26. *Classificazione delle forme giuridiche delle unità legali*

Anno 2006

27. *Gli stranieri nella rilevazione continua sulle forze di lavoro*
28. *L'indagine campionaria sulle nascite: obiettivi, metodologia e organizzazione*
29. *Rilevazione mensile sull'occupazione, gli orari di lavoro e le retribuzioni nelle grandi imprese*
30. *La classificazione Istat dei titoli di studio italiani. Anno 2003* 
31. *Il sistema di indagini sociali multiscopo. Contenuti e metodologia delle indagini*
32. *La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione*

 dati forniti su floppy disk

 dati forniti su cd-rom



***Produzione editoriale
&
Altri servizi***

La produzione editoriale

LE PUBBLICAZIONI A CARATTERE GENERALE

Annuario statistico italiano 2005
pp. XXIV-824+1 cd-rom; € 50,00
ISBN 88-458-0895-5

Bollettino mensile di statistica
pp. 164 circa; € 15,00
ISSN 0021-3136

Compendio statistico italiano 2005
pp. 388; € 10,00
ISBN 88-458-1309-6

Genesees V. 3.0
Funzione Riponderazione
Tecniche e strumenti, n. 2, ed. 2005
pp. 220+1 cd-rom; € 27,00
ISBN 88-458-0870-X

Genesees V. 3.0
Funzione Stime ed Errori
Tecniche e strumenti, n. 3, ed. 2005
pp. 252+1 cd-rom; € 27,00
ISBN 88-458-0896-3

L'innovazione di frontiera
Relazioni
Quaderni del MIPA, n. 3, ed. 2005
pp. 212; € 15,00
ISBN 88-458-0877-7

Politiche per l'internazionalizzazione e competitività del sistema economico italiano
Sportello Italia
Quaderni del MIPA, n. 4, ed. 2005
pp. 300; € 22,00
ISBN 88-458-0891-2

Rapporto annuale
La situazione del Paese nel 2004
pp. XXXII-420; € 25,00
ISBN 88-458-0887-4
ISSN 1594-3135

Rapporto annuale
La situazione del Paese nel 2004
pp. XXXII-420+1 cd-rom; € 30,00
ISBN 88-458-0888-2
ISSN 1594-3135

LE NOVITÀ EDITORIALI A CARATTERE TEMATICO

AMBIENTE E TERRITORIO

Statistiche ambientali
Annuari, n. 8, edizione 2005
pp. 548+1 cd-rom; € 39,00
ISBN 88-458-1293-6

POPOLAZIONE

Decessi: caratteristiche demografiche e sociali (*)
anno 2001
Annuari, n. 10, edizione 2005
pp. 144; € 11,00
ISBN 88-458-0872-6

Matrimoni, separazioni e divorzi (*)
anno 2001
Annuari, n. 14, edizione 2005
pp. 160; € 13,00
ISBN 88-458-0876-9

Movimento migratorio della popolazione residente: iscrizioni e cancellazioni anagrafiche
anno 2000
Annuari, n. 13, edizione 2005
pp. 228+1 cd-rom; € 25,00
ISBN 88-458-0875-0

Popolazione e movimento anagrafico dei comuni
anno 2003
Annuari, n. 16, edizione 2005
pp. 264+1 cd-rom; € 27,00
ISBN 88-458-1300-2

SANITÀ E PREVIDENZA

Gli assicurati alle gestioni pensionistiche invalidità, vecchiaia e superstiti (*)
anno 2003
Informazioni, n. 30, edizione 2005
pp. 72; € 9,00
ISBN 88-458-1304-5

L'assistenza residenziale in Italia: regioni a confronto
anno 2001
Informazioni, n. 18, edizione 2005
pp. 96+1 cd-rom; € 14,00
ISBN 88-458-0898-X

Cause di morte
anno 2001
Annuari, n. 17, edizione 2005
pp. 428; € 28,00
ISBN 88-458-0886-6

Decessi: caratteristiche demografiche e sociali (*)
anno 2001
Annuari, n. 10, edizione 2005
pp. 144; € 11,00
ISBN 88-458-0872-6

L'interruzione volontaria di gravidanza in Italia
anno 2002
Informazioni, n. 16, edizione 2005
pp. 114; € 12,00
ISBN 88-458-0889-0

Le notizie di malattie infettive in Italia
anno 2002
Informazioni, n. 2, edizione 2005
pp. 128; € 22,00
ISBN 88-458-0863-7

Nuove evidenze nell'evoluzione della mortalità per tumori in Italia
anni 1970-1999
Indicatori statistici, n. 5, edizione 2005
pp. 118; € 11,00
ISBN 88-458-0904-8

Le organizzazioni di volontariato in Italia (*)
anno 2001
Informazioni, n. 27, edizione 2005
pp. 250; € 22,00
ISBN 88-458-1298-7

Le pubblicazioni con (*) sono riportate in più settori editoriali



Sistema sanitario e salute della popolazione

Indicatori regionali - Anni 2001-2002
Informazioni, n. 14, edizione 2005
pp. 268+1 cd-rom; € 27,00
ISBN 88-458-0883-1

Statistiche della previdenza e dell'assistenza sociale (*)

II - I beneficiari delle prestazioni pensionistiche
anno 2002
Annuari, n. 4, edizione 2005
pp. 156+1 cd-rom; € 17,50
ISBN 88-458-0864-5

Statistiche della previdenza e dell'assistenza sociale (*)

I - I trattamenti pensionistici
anno 2003
Annuari, n. 5, edizione 2005
pp. 112+1 cd-rom; € 16,00
ISBN 88-458-1292-8

Struttura e attività degli istituti di cura

anno 2002
Informazioni, n. 17, edizione 2005
pp. 222; € 22,00
ISBN 88-458-0897-1

CULTURA

Cultura, socialità e tempo libero (*)

anno 2003
Informazioni, n. 12, edizione 2005
pp. 112+1 cd-rom; € 17,00
ISBN 88-458-0881-5

I diplomati universitari e il mercato del lavoro (*)

Inserimento professionale dei diplomati universitari - Indagine 2002
Informazioni, n. 7, edizione 2005
pp. 292+1 cd-rom; € 25,00
ISBN 88-458-0871-8

La produzione libraria nel 2003

Dati definitivi
Informazioni, n. 20, edizione 2005
pp. 92; € 9,00
ISBN 88-458-0902-1

Lo sport che cambia (*)

I comportamenti emergenti e le nuove tendenze della pratica sportiva in Italia
Argomenti, n. 29, edizione 2005
pp. 292; € 20,00
ISBN 88-458-0905-6

Statistiche culturali

anni 2002-2003
Annuari, n. 43, edizione 2005
pp. 236; € 22,00
ISBN 88-458-0892-0

FAMIGLIA E SOCIETÀ

I consumi delle famiglie

anno 2003
Annuari, n. 10, edizione 2005
pp. 168+1 cd-rom; € 19,00
ISBN 88-458-0900-5

Cultura, socialità e tempo libero (*)

anno 2003
Informazioni, n. 12, edizione 2005
pp. 112+1 cd-rom; € 17,00
ISBN 88-458-0881-5

Famiglia, abitazione e zona in cui si vive

anno 2003
Informazioni, n. 19, edizione 2005
pp. 112+1 cd-rom; € 17,00
ISBN 88-458-0901-3

Il monitoraggio del processo e la stima dell'errore

nelle indagini telefoniche
Applicazione all'indagine sulla sicurezza dei cittadini
Metodi e norme, n. 25, edizione 2005
pp. 124; € 12,00
ISBN 88-458-1299-5

La rete di rilevazione Capi dell'Istat per la conduzione dell'indagine continua sulle Forze di Lavoro (*)

Metodi e norme, n. 24, edizione 2005
pp. 104; € 11,00
ISBN 88-458-0894-7

I servizi pubblici e di pubblica utilità: utilizzo e soddisfazione

anno 2003
Informazioni, n. 26, edizione 2005
pp. 164+1 cd-rom; € 19,00
ISBN 88-458-1297-9

Lo sport che cambia (*)

I comportamenti emergenti e le nuove tendenze della pratica sportiva in Italia
Argomenti, n. 29, edizione 2005
pp. 292; € 20,00
ISBN 88-458-0905-6

Stili di vita e condizioni di salute

anno 2003
Informazioni, n. 25, edizione 2005
pp. 120+1 cd-rom; € 16,00
ISBN 88-458-1291-X

I viaggi in Italia e all'estero nel 2003 (*)

Informazioni, n. 1, edizione 2005
pp. 96+1 cd-rom; € 12,50
ISBN 88-458-0861-0

PUBBLICA AMMINISTRAZIONE

Gli assicurati alle gestioni pensionistiche

invalidità, vecchiaia e superstiti (*)
anno 2003
Informazioni, n. 30, edizione 2005
pp. 72; € 9,00
ISBN 88-458-1304-5

I bilanci consuntivi delle amministrazioni provinciali

anno 2002
Informazioni, n. 9, edizione 2005
pp. 48+1 cd-rom; € 14,00
ISBN 88-458-0874-2

I bilanci consuntivi delle regioni e delle province autonome

anno 2001
Informazioni, n. 4, edizione 2005
pp. 32+1 cd-rom; € 14,00
ISBN 88-458-0866-1

I bilanci consuntivi e i servizi delle Camere di commercio, industria, artigianato e agricoltura

anno 2001
Informazioni, n. 5, edizione 2005
pp. 52+1 cd-rom; € 14,00
ISBN 88-458-0867-X

Le organizzazioni di volontariato in Italia (*)

anno 2001
Informazioni, n. 27, edizione 2005
pp. 250; € 22,00
ISBN 88-458-1298-7

Statistiche della previdenza e dell'assistenza sociale (*)

II - I beneficiari delle prestazioni pensionistiche - Anno 2002
Annuari, n. 4, edizione 2005
pp. 156+1 cd-rom; € 17,50
ISBN 88-458-0864-5

Statistiche della previdenza e dell'assistenza sociale (*)

I - I trattamenti pensionistici
anno 2003
Annuari, n. 5, edizione 2005
pp. 112+1 cd-rom; € 16,00
ISBN 88-458-1292-8

GIUSTIZIA

Matrimoni, separazioni e divorzi (*)

anno 2001
Annuari, n. 14, edizione 2005
pp. 160; € 13,00
ISBN 88-458-0876-9



Statistiche giudiziarie civili
anno 2003
Annuari, n. 12, edizione 2005
pp. 260; € 22,00
ISBN 88-458-1295-2

Statistiche giudiziarie penali
anno 2003
Annuari, n. 12, edizione 2005
pp. 344; € 32,00
ISBN 88-458-1289-8

▶ **CONTI NAZIONALI**

Contabilità nazionale
Tomo 1 - Conti economici nazionali - Anni 1992-2003
Annuari, n. 9, edizione 2005
pp. 236; € 22,00
ISBN 88-458-0862-9

I conti economici nazionali per settore istituzionale: le nuove stime secondo il Sec95
Metodi e norme, n. 23, edizione 2005
pp. 336; € 32,00
ISBN 88-458-0884-X

Valore aggiunto ai prezzi di base dell'agricoltura per regione anni 1999-2004
Informazioni, n. 23, edizione 2005
pp. 134+1 cd-rom; € 17,00
ISBN 88-458-1288-X

▶ **LAVORO**

I diplomati universitari e il mercato del lavoro (*)
Inserimento professionale dei diplomati universitari - Indagine 2002
Informazioni, n. 7, edizione 2005
pp. 292+1 cd-rom; € 25,00
ISBN 88-458-0871-8

Lavoro e retribuzioni
anno 2001
Annuari, n. 7, edizione 2005
pp. 236+1 cd-rom; € 25,00
ISBN 88-458-0879-3

La rete di rilevazione Capi dell'Istat per la conduzione dell'indagine continua sulle Forze di Lavoro (*)
Metodi e norme, n. 24, edizione 2005
pp. 104; € 11,00
ISBN 88-458-0894-7

Le retribuzioni contrattuali annue di competenza
base dicembre 1995=100
anni 1996-2001
Informazioni, n. 15, edizione 2005
pp. 56+1 cd-rom; € 14,00
ISBN 88-458-0885-8

Statistiche della previdenza e dell'assistenza sociale (*)
II - I beneficiari delle prestazioni pensionistiche - Anno 2002
Annuari, n. 4, edizione 2005
pp. 156+1 cd-rom; € 17,50
ISBN 88-458-0864-5

Statistiche della previdenza e dell'assistenza sociale (*)
I - I trattamenti pensionistici anno 2003
Annuari, n. 5, edizione 2005
pp. 112+1 cd-rom; € 16,00
ISBN 88-458-1292-8

▶ **PREZZI**

Il valore della moneta in Italia dal 1861 al 2004
Informazioni, n. 24, edizione 2005
pp. 168; € 14,00
ISBN 88-458-1290-1

▶ **AGRICOLTURA**

Statistiche dell'agricoltura
anno 2000
Annuari, n. 48, edizione 2005
pp. 356; € 27,00
ISBN 88-458-0893-9

▶ **INDUSTRIA**

Classificazione delle forme giuridiche delle unità legali (*)
Metodi e norme, n. 26, edizione 2005
pp. 140; € 10,00
ISBN 88-458-1307-X

I consumi energetici delle imprese industriali
anno 2002
Informazioni, n. 13, edizione 2005
pp. 44+1 cd-rom; € 14,00
ISBN 88-458-0882-3

Conti economici delle imprese (*)
anno 2001
Informazioni, n. 28, edizione 2005
pp. 136+1 cd-rom; € 17,00
ISBN 88-458-1301-0

La produzione dell'industria dei prodotti chimici e delle fibre sintetiche e artificiali
Statistica per trimestri - Anno 2003
Informazioni, n. 22, edizione 2005
pp. 52+1 cd-rom; € 14,00
ISBN 88-458-0906-4

La produzione dell'industria siderurgica
anno 2003
Informazioni, n. 11, edizione 2005
pp. 52+1 cd-rom; € 14,00
ISBN 88-458-0880-7

La produzione dell'industria tessile e dell'abbigliamento
Statistica per trimestri anno 2003
Informazioni, n. 21, edizione 2005
pp. 56+1 cd-rom; € 14,00
ISBN 88-458-0903-X

▶ **SERVIZI**

Classificazione delle forme giuridiche delle unità legali (*)
Metodi e norme, n. 26, edizione 2005
pp. 140; € 10,00
ISBN 88-458-1307-X

Conti economici delle imprese (*)
anno 2001
Informazioni, n. 28, edizione 2005
pp. 136+1 cd-rom; € 17,00
ISBN 88-458-1301-0

Trasporto merci su strada
anno 2003
Informazioni, n. 10, edizione 2005
pp. 56+1 cd-rom; € 14,00
ISBN 88-458-0878-5

Statistiche dei trasporti
anni 2002-2003
Annuari, n. 4, edizione 2005
pp. 284; € 22,00
ISBN 88-458-0890-4

Statistiche del trasporto aereo
anno 2002
Informazioni, n. 3, edizione 2005
pp. 44+1 cd-rom; € 13,00
ISBN 88-458-0865-3

I viaggi in Italia e all'estero nel 2003 (*)
Informazioni, n. 1, edizione 2005
pp. 96+1 cd-rom; € 12,50
ISBN 88-458-0861-0

▶ **COMMERCIO ESTERO**

Commercio estero e attività internazionali delle imprese 2004
1. Merci, servizi, investimenti diretti
2. Paesi, settori, regioni
+ *L'Italia nell'economia internazionale*
Rapporto ICE 2004-2005 + 1 cd-rom
+ Sintesi del Rapporto ICE
Annuari, n. 7, edizione 2005
pp. 368 + 432 + 456 + 44
€ 100,00 (in cofanetto)
ISBN 88-458-0899-8

PRODOTTI CENSUARI

5° CENSIMENTO GENERALE DELL'AGRICOLTURA 22 OTTOBRE 2000

Caratteristiche strutturali delle aziende agricole

Fascicolo nazionale; € 25,00
Fascicoli regionali; € 22,00
Fascicoli provinciali; € 22,00

Caratteristiche tipologiche delle aziende agricole

Fascicolo nazionale; € 20,00
Fascicoli regionali; € 14,00

VOLUMI TEMATICI

La coltivazione della vite in Italia

Volume I - Caratteristiche generali
pp. 300; € 26,50; ISBN 88-458-1280-4
Volume II - Vitigni
pp. 248; € 22,00; ISBN 88-458-1281-2

La donna in agricoltura

pp. 316; € 14,00; ISBN 88-458-1284-7

Le imprese agricole

pp. 338; € 22,00; ISBN 88-458-1283-9

Le infrastrutture delle aziende agricole

pp. 150; € 11,50; ISBN 88-458-1279-0

La zootecnia in Italia

pp. 380; € 26,50; ISBN 88-458-1282-0

14° CENSIMENTO GENERALE DELLA POPOLAZIONE E DELLE ABITAZIONI - 21 OTTOBRE 2001

Primi risultati

pp. 300+1 cd-rom; € 25,00; ISBN 88-458-0689-8

Popolazione legale

pp. 312+1 cd-rom; € 27,00; ISBN 88-458-1069-0

Struttura demografica e familiare della popolazione residente - Italia

pp. 294+1 cd-rom; € 32,00; ISBN 88-458-1388-6

Popolazione residente e abitazioni nelle province italiane

Fascicoli provinciali; € 19,00/27,00

8° CENSIMENTO GENERALE DELL'INDUSTRIA E DEI SERVIZI - 22 OTTOBRE 2001

Imprese, istituzioni e unità locali

Fascicolo nazionale; € 31,50
Fascicoli regionali; € 22,00
Fascicoli provinciali; € 14,00

Per gli utenti che acquistano oltre 10 volumi dei **PRODOTTI CENSUARI** è previsto uno sconto del 25%.

Altri prodotti e servizi

ABBONAMENTI 2006

L'abbonamento consente di disporre di tutte le informazioni relative al settore tematico prescelto, diffuse attraverso le pubblicazioni edite nel 2006, accompagnate, ove previsto, da supporto informatico (floppy disk, cd-rom). Gli abbonati riceveranno per posta i prodotti che saranno via via pubblicati nel/i settore/i prescelto/i, editi nell'anno di sottoscrizione dell'abbonamento, ad esclusione dei volumi inseriti nelle collane: Tecniche e strumenti, Essays, Quaderni del Mipa e Censimenti. Oltre all'abbonamento ai singoli settori editoriali è prevista la modalità di abbonamento "Tutti i settori", che comprende tutta la produzione editoriale dell'Istituto edita nel 2006 compresi i volumi del Commercio estero. L'abbonamento all'area "Generale", infine, comprende 11 numeri del *Bollettino mensile di statistica* e l'*Annuario statistico italiano*. Tutti coloro che sottoscriveranno un abbonamento anche ad un solo settore riceveranno, gratuitamente, una copia del *Rapporto annuale*. Per meglio comprendere il sistema degli abbonamenti è possibile visionare, sul sito www.istat.it, l'elenco 2004 e l'elenco 2005 delle pubblicazioni inviate agli abbonati alle edizioni 2004 e 2005.

Per sottoscrivere gli abbonamenti si può utilizzare il modulo riportato nella pagina seguente.

WWW.ISTAT.IT

Nel sito Internet è possibile informarsi sulla produzione editoriale più recente, richiedere prodotti e servizi offerti dall'Istat, leggere e prelevare i comunicati stampa, accedere alle Banche Dati, collegarsi con altri siti nazionali e internazionali. Inoltre, è possibile consultare il catalogo della produzione editoriale on line, dove ci sono tutte le informazioni relative ai prodotti a partire dalle edizioni 2000.

Ulteriori informazioni possono essere richieste a:

ISTAT - Direzione centrale per la diffusione dell'informazione statistica e il supporto alla produzione editoriale
Via Cesare Balbo, 16 - 00184 ROMA - Tel. 0646733278/80 - Fax 0646733477 - e-mail: marketing@istat.it

Abbonamenti 2006

Inviare questo modulo via fax al numero 0646733477 oppure spedire in busta chiusa a:
Istituto nazionale di statistica - DCDE - Commercializzazione e Marketing - Via Cesare Balbo, 16 - 00184 Roma
Per ulteriori informazioni telefonare ai numeri 0646733278/79/80

Desidero sottoscrivere i seguenti abbonamenti per l'anno 2006:

TIPOLOGIE DI ABBONAMENTO	PREZZI	
	ITALIA Euro	ESTERO Euro
Generale (Bollettino mensile di statistica, Annuario statistico italiano)	[] 150,00	[] 170,00
Ambiente e territorio	[] 50,00	[] 55,00
Popolazione (escluso censimenti)	[] 80,00	[] 90,00
Sanità e previdenza	[] 180,00	[] 200,00
Cultura	[] 80,00	[] 90,00
Famiglia e società	[] 120,00	[] 130,00
Pubblica amministrazione	[] 100,00	[] 110,00
Giustizia	[] 80,00	[] 90,00
Conti nazionali	[] 100,00	[] 110,00
Lavoro	[] 120,00	[] 130,00
Prezzi	[] 50,00	[] 55,00
Agricoltura (escluso censimenti)	[] 50,00	[] 55,00
Industria (escluso censimenti)	[] 120,00	[] 130,00
Servizi	[] 120,00	[] 130,00
Commercio estero	[] 80,00	[] 90,00
Tutti i settori (escluso censimenti)	[] 900,00	[] 1.000,00
Per un totale di	_____	_____
Eventuale sconto (a)	_____	_____
Importo da pagare	_____	_____

Qualunque abbonamento, anche a un solo settore, comprende l'invio di una copia del Rapporto annuale.

(a) Sconti e agevolazioni: il Sistan, gli Enti pubblici e le Università usufruiscono di uno sconto del **20% solo** se sottoscrivono l'abbonamento direttamente con l'Istat.

RICHIEDENTE ABBONAMENTO

Cognome _____ Nome _____

Ente _____ Qualifica _____

Codice fiscale/ P. IVA | | | | | | | | | | | | | | | | | | | | | |

Indirizzo _____ CAP _____ Città _____

Prov. _____ Tel. _____ Fax _____

E-mail _____

Data _____ Firma _____

DESTINATARIO DEI PRODOTTI (SE DIVERSO DAL RICHIEDENTE)

Cognome _____ Nome _____

Ente _____ Indirizzo _____

CAP _____ Città _____ Tel. _____ Fax _____

MODALITÀ DI PAGAMENTO:
 Gli importi dovranno essere versati dall'acquirente, **dopo il ricevimento della fattura**, sul c/c postale n. 619007, oppure con bonifico bancario c/o la Banca Nazionale del Lavoro, **indicando con chiarezza il numero, la data della fattura e il codice cliente**. Per i versamenti tramite bonifico bancario le coordinate sono: c/c n. 218050, ABI 01005.8, CAB 03382.9; via swift: B.N.L.I. IT RR, codice CIN K, codice anagrafico 63999228/j.

INFORMATIVA - I dati da lei forniti saranno utilizzati per l'esecuzione dell'ordine e per l'invio, da parte dell'Istat, di promozioni commerciali, senza alcun impegno da parte sua. Il trattamento dei dati avverrà nell'assoluto rispetto della disciplina dettata dal d.lgs. 196/2003; essi non verranno utilizzati per finalità diverse da quelle indicate e saranno trattati esclusivamente dai dipendenti dell'Istituto incaricati. Il titolare dei dati è l'Istituto nazionale di statistica, Via Cesare Balbo, 16 - 00184 Roma - tel. 064673.3266/68, fax 0646733477; responsabile del trattamento è il Direttore centrale per la diffusione dell'informazione statistica e il supporto alla produzione editoriale, anche per quanto riguarda l'esercizio dei diritti dell'interessato di cui all'articolo 7 del d.lgs. 196/2003.

PIÙ INFORMAZIONI. PIÙ VICINE A VOI.

I Centri d'Informazione Statistica

Per darvi più servizi e per esservi più vicino l'Istat ha aperto al pubblico una rete di Centri d'Informazione Statistica che copre l'intero territorio nazionale. Oltre alla vendita di prodotti informatici e pubblicazioni, i Centri rilasciano certificati sull'indice dei prezzi, offrono informazioni tramite collegamenti con le banche dati del Sistema statistico nazionale (Sistan) e dell'Eurostat (Ufficio di statistica della Comunità europea), forniscono elaborazioni statistiche "su misura" ed assistono i laureandi nella ricerca e selezione dei dati.

Presso i Centri d'Informazione Statistica, semplici cittadini, studenti, ricercatori, imprese e operatori della pubblica amministrazione troveranno assistenza qualificata e un facile accesso ai dati di cui hanno bisogno. D'ora in poi sarà più facile conoscere l'Istat e sarà più facile per tutti gli italiani conoscere l'Italia. Per gli orari di apertura al pubblico consultare il sito www.istat.it nella pagina "Prodotti e servizi".

ANCONA Corso Garibaldi, 78
Telefono 071/5013090-1 Fax 071/5013085

BARI Piazza Aldo Moro, 61
Telefono 080/5789317 Fax 080/5789335

BOLOGNA Galleria Cavour, 9
Telefono 051/6566111 Fax 051/6566182

BOLZANO Viale Duca d'Aosta, 59
Telefono 0471/414000 Fax 0471/414008

CAGLIARI Via Firenze, 17
Telefono 070/34998700-1 Fax 070/34998732-3

CAMPOBASSO Via G. Mazzini, 129
Telefono 0874/604854-8 Fax 0874/604885-6

CATANZARO Viale Pio X, 116
Telefono 0961/507629 Fax 0961/741240

FIRENZE Via Santo Spirito, 14
Telefono 055/2393311 Fax 055/2393335

GENOVA Via San Vincenzo, 4
Telefono 010/58497201 Fax 010/542351

MILANO Via Fieno, 3
Telefono 02/806132214 Fax 02/806132205

NAPOLI Via G. Verdi, 18
Telefono 081/4930190 Fax 081/4930185

PALERMO Via Empedocle Restivo, 102
Telefono 091/7290915 Fax 091/521426

PERUGIA Via Cesare Balbo, 1
Telefono 075/5826411 Fax 075/5826485

PESCARA Via Caduta del Forte, 34
Telefono 085/44120511-2 Fax 085/4216516

POTENZA Via del Popolo, 4
Telefono 0971/377211 Fax 0971/36866

ROMA Via Cesare Balbo, 11/a
Telefono 06/46733102 Fax 06/46733101

TORINO Via Alessandro Volta, 3
Telefono 011/5166711 Fax 011/539412

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.

E a disposizione dell'utenza una sala di consultazione al secondo piano

ROMA Via Cesare Balbo, 16 Telefono 06/4673.2380 Fax 06/4673.2617

E-mail: biblio@istat.it

Orario: Piano secondo

da lunedì a venerdì 9.00 - 18.00

 *Lavoro*

La rilevazione sulle forze di lavoro: contenuti, metodologie, organizzazione

Il volume presenta un panorama completo delle innovazioni di carattere metodologico, contenutistico e organizzativo relative alla nuova indagine continua sulle forze di lavoro, modificata secondo quanto richiesto dal regolamento comunitario 577/98, che ne definisce le procedure metodologiche e le variabili da rilevare. Nei vari capitoli sono illustrate le modifiche introdotte nei contenuti, nelle definizioni e nelle classificazioni dell'indagine. Sul versante della metodologia statistica, dei metodi e delle tecniche di indagine e dell'organizzazione del processo vengono presentate le procedure innovative messe a regime. Completano il quadro informativo offerto la descrizione della rete di rilevatori direttamente gestita dall'Istat, delle innovazioni introdotte per il monitoraggio della qualità del lavoro sul campo e del sistema informativo-informatico che ha reso possibile l'avvio e la messa a regime della nuova indagine.

€ 22,00

ISBN 88-458-1357-6



1M012006032000000