

Le pagine seguenti sono un estratto del volume "Indagine Plus. Il mondo del lavoro tra forma e sostanza. Terza annualità"



167
I LIBRI DEL
FONDO SOCIALE EUROPEO

ISSN: 1590-0002

L'ISFOL, ente nazionale di ricerca, opera nel campo della formazione, delle politiche sociali e del lavoro al fine di contribuire alla crescita dell'occupazione, al miglioramento delle risorse umane, all'inclusione sociale e allo sviluppo locale ed è sottoposto alla vigilanza del Ministero del Lavoro e delle Politiche sociali. L'ISFOL svolge e promuove attività di studio, ricerca, sperimentazione, documentazione e informazione fornendo un supporto tecnico-scientifico al Parlamento, al Ministero del Lavoro e ad altri Dicasteri, alle Regioni, agli Enti locali ed alle Istituzioni sulle politiche e sui sistemi della formazione ed apprendimento lungo tutto l'arco della vita, del mercato del lavoro e dell'inclusione sociale. Fa parte del Sistema Statistico Nazionale, e collabora con le Istituzioni comunitarie. Svolge inoltre il ruolo di assistenza metodologica e scientifica per le azioni di sistema del Fondo sociale europeo, è Agenzia nazionale Lifelong Learning Programme - Programma settoriale Leonardo da Vinci.

Commissario Straordinario: Matilde Mancini
Direttore Generale: Aviana Bulgarelli

Riferimenti
Corso d'Italia, 33
00198 Roma
Tel. + 39 06854471
Web: www.isfol.it

La Collana
I Libri del Fondo sociale europeo
raccolge e valorizza i risultati tecnico-scientifici conseguiti nei Piani di attività Isfol per la programmazione di FSE 2007-2013 Obiettivo Convergenza PON "Governance e Azioni di sistema" e Obiettivo Competitività regionale e occupazione PON "Azioni di sistema"

I testi pubblicati in Collana sono valutati dal Comitato tecnico scientifico interno ISFOL

La Collana I Libri del Fondo sociale europeo è curata da Isabella Pitoni responsabile del Servizio per la comunicazione e la divulgazione scientifica Isfol. Coordinamento editoriale: Valeria Cioccolo, Paola Piras



Unione europea
Fondo sociale europeo



MINISTERO DEL LAVORO
E DELLE POLITICHE SOCIALI
DIREZIONE GENERALE PER LE POLITICHE
ATTIVE E PASSIVE DEL LAVORO
DIREZIONE GENERALE PER LE POLITICHE
DEI SERVIZI PER IL LAVORO

fse per il tuo futuro
Programmi operativi nazionali
per la formazione e l'occupazione

ISFOL

INDAGINE PLUS IL MONDO DEL LAVORO TRA FORMA E SOSTANZA TERZA ANNUALITÀ

ISFOL
ISTITUTO PER LO SVILUPPO DELLA
FORMAZIONE PROFESSIONALE
DEI LAVORATORI

Il volume è stato finanziato dal Fondo sociale europeo nell'ambito dei Programmi operativi nazionali a titolarità del Ministero del lavoro e delle politiche sociali "Azioni di sistema" (Ob. Convergenza) Asse Occupabilità, Obiettivo specifico 2.1, Progetto 4 "Analisi di target specifici del mercato del lavoro", Attività 1 e Governance e Azioni di sistema (Ob. Competitività regionale e occupazione) Asse Occupabilità, Obiettivo specifico 2.1, Progetto 4 "Analisi di target specifici del mercato del lavoro", Attività 1.

Il volume raccoglie i principali risultati della terza edizione dell'indagine Isfol Plus (Participation, Labour, Unemployment, Survey).

Il libro è a cura di *Emiliano Mandrone* e *Debora Radicchia*.

Sono autori dei testi:

Francesca Bergamante (par. 10.6), *Pietro Checcucci* (par. 8.1), *Gianni Corsetti* (cap. 12), *Roberta Fefè* (par. 8.3), *Monja Ferretti* (parr. 6.1, 6.2), *Michele Giammatteo* (capp. 3, 12, par. 8.3), *Roberto Landi* (parr. 4.2.4.4, 4.3 4.6), *Andrea Laudadio* (par.11.1,11.2), *Federica Laudisa* (cap. 6-approfondimento), *Cristina Lolli* (parr. 6.1, 6.2), *Emiliano Mandrone* (Introduzione, capp. 1, 3, 7, 9, parr. 2.1, 2.3, 2.4, parr 4.1, 4.4, 4.5, 4.6, 6.2, 10.5, 11.1, 11.2,11.3,11.4,11.6,11.7,11.8), *Sante Marchetti* (par. 8.3), *Valeria Perotti* (par. 11.5), *Debora Radicchia* (Introduzione, cap. 5, parr. 2.2, 4.2, 4.3, 10.1, 10.2, 10.3, 10.4), *Antonella Scarangella* (par. 7.7), *Giuliana Scarpetti* (par. 8.2), *Simona Tenaglia* (par. 11.1,11.2,11.3).

Editing del volume: Anna Nardone.

Questo testo è stato sottoposto con esito favorevole al processo di peer review interna curato dal Comitato tecnico scientifico interno ISFOL.

Testo chiuso: ottobre 2010.

Copyright (C) [2012] [ISFOL]

Quest'opera è rilasciata sotto i termini della licenza Creative Commons Attribuzione – Non Commerciale – Condividi allo stesso modo 3.0 Italia License.

(<http://creativecommons.org/licenses/by-nc-sa/3.0/deed.it>)



ISBN: 978-88-543-0076-7

12 Metodologia dell'Indagine campionaria longitudinale PLUS*

L'indagine ISFOL PLUS (Participation Labour Unemployment Survey) è una rilevazione nazionale campionaria sull'offerta di lavoro alla sua quarta annualità (sulle 5 originariamente previste) ed è presente nel Piano Statistico Nazionale dal 2006.

L'obiettivo primario dell'indagine PLUS è quello di fornire stime statisticamente affidabili di fenomeni rari o solo marginalmente esplorati dalle maggiori rilevazioni sul mercato del lavoro italiano (ISTAT, INPS). Infatti, se la Rilevazione Continua sulle Forze di Lavoro (RCFL) dell'ISTAT fornisce regolarmente gli aggregati e gli indicatori ufficiali sul mercato del lavoro (tassi di disoccupazione, occupazione, attività, ecc.), l'indagine PLUS è principalmente rivolta all'approfondimento di specifici aspetti, quali la distribuzione dei contratti (dipendente, autonomo, informale, ecc.), la ricerca di lavoro, la partecipazione lavorativa delle donne, dei giovani e delle persone con più di cinquanta anni, le scelte di pensionamento, istruzione e formazione, le dinamiche intergenerazionali.

L'indagine è stata ideata per analizzare il mercato del lavoro attuale, contraddistinto da una sempre più veloce trasformazione delle modalità di ricerca e svolgimento del lavoro (rimaste pressoché invariate nei decenni passati) in nuove e molteplici forme, in cui i concetti di occupazione e disoccupazione sfuggono oggi sempre più alle categorie tradizionali alle quali eravamo stati abituati. Per affrontare le sfide epistemologiche appena descritte, l'indagine ISFOL PLUS si distingue dalle usuali indagini su base familiare per l'assenza di rispondenti proxy e per la capacità di integrare diversi aspetti del mercato del lavoro, spesso analizzati in maniera disgiunta. Collegando le dinamiche occupazionali e reddituali con la realtà personale e familiare essa ricostruisce un quadro di legami e relazioni invisibili ad analisi monotematiche. In questa ottica l'indagine disegna la reale fisionomia del mercato del lavoro così come viene percepita dai diretti interessati, con particolare riferimento alle problematiche dei giovani, delle donne e degli ultra cinquantenni. La visione dinamica di questi fenomeni è garantita dal disegno longitudinale che consiste in una quota d'interviste ripetute ai medesimi soggetti (panel) che ne tracciano con precisione la traiettoria nel mondo del lavoro al passare del tempo.

* Di Gianni Corsetti e Michele Giammatteo (ISFOL).

L'azione dei ricercatori che si sono occupati dell'Indagine ISFOL PLUS è stata, fin dal primo giorno, improntata al confronto sulle metodologie, all'ampia partecipazione nella stesura del questionario, alla trasparenza nelle fasi di somministrazione e del trattamento dei dati, alla comunicazione dei risultati e alla diffusione dei microdati.

Numerosi e continui sono stati i momenti di confronto e molteplici le pubblicazioni riguardanti l'indagine PLUS, sia sugli aspetti metodologici¹ sia relativi al disegno campionario della rilevazione, sia inerenti ai risultati che alle analisi condotte sui dati. Ampia è stata la diffusione dei risultati e la distribuzione dei microdati², con un circolo "virtuoso" che ha migliorato l'indagine e reso condivise (e quindi più autorevoli) le scelte metodologiche e definitorie del gruppo che cura la rilevazione.

Prima di approfondire nel dettaglio gli aspetti concernenti il disegno campionario, vogliamo di seguito evidenziare le principali caratteristiche dell'indagine PLUS, mostrandone le potenzialità e il rigore concettuale e metodologico che la caratterizzano. Il primo aspetto che ci sembra importante affrontare riguarda la definizione di alcuni aggregati caratterizzanti le analisi standard sul mercato del lavoro. I dati PLUS sull'occupazione sono ispirati a un criterio classificatorio diverso rispetto ai dati RCFL. Infatti, mentre la rilevazione PLUS definisce come occupati e in cerca di lavoro le persone che si *auto-definiscono* tali, RCFL segue un percorso che identifica la condizione in base ad alcune informazioni "oggettive" che riguardano: per gli occupati l'aver lavorato almeno un'ora nella settimana di riferimento dell'intervista; per le persone in cerca l'aver compiuto almeno un'azione di ricerca entro i 30 giorni che precedono l'intervista ed essere immediatamente disponibili a lavorare. Da questa dovuta precisazione ne segue come l'impianto ISTAT (EUROSTAT) sottenda nel proprio meccanismo contatore una certa sovrastima degli occupati e sottostima delle persone in cerca. L'idea generale dell'indagine PLUS di registrare, nel modo più accurato possibile, la condizione *auto percepita* dai soggetti intervistati fa sì che anche la distinzione tra *persone in cerca* ed *inattivi* sia differente da quanto adottato in RCFL. In particolare:

- a. si considerano *persone in cerca*, e quindi attive, alcune tipologie di individui che per l'ISTAT sono da considerare *inattivi*;
- b. non si considerano occupati quei soggetti che svolgono una attività lavorativa che non è, in termini economici e secondo la propria percezione, tale da giustificare la loro inclusione in tale categoria (studenti, pensionati da lavoro e casalinghe - lavoratrici/ori saltuari), considerandoli *occupati non prevalenti*.

1. Per una descrizione esaustiva delle metodologie adottate si veda Mandrone e Radicchia (2006) e Giammatteo (2009).

2. I dati ISFOL PLUS vengono messi a disposizione della comunità scientifica entro 12 mesi dalla rilevazione attraverso la sottoscrizione del protocollo comunicazione dati dell'Istituto.

Perché questa scelta? Essendo il fine di PLUS quello di verificare la qualità dell'attuale occupazione è stato necessario riferirsi allo status percepito dall'intervistato, poiché utilizzando uno schema analogo a quello dell'ISTAT, ovvero seguendo le indicazioni dei regolamenti comunitari, si correva il rischio di includere tra gli *occupati* quelli con *condizione non prevalente* e tra gli inattivi i *disoccupati che non rientrano nella classificazione ISTAT*. Queste sottopopolazioni rappresentano proprio gli ambiti di maggior interesse per lo studio e le politiche per l'attivazione stabile e continuativa. Ovviamente da PLUS è possibile ricostruire gli occupati nelle definizioni ISTAT-EUROSTAT essendo stati somministrati i quesiti necessari alla loro individuazione. A tale scopo, si è deciso di vincolare i dati PLUS ad alcuni aggregati ufficiali di fonte RCFL, precisando che gli *occupati non prevalenti* e gli inattivi che si dichiarano in cerca sono stati considerati nelle condizioni cui si attribuivano autonomamente.

La prima conseguenza è che le caratteristiche dell'occupazione sono al netto della componente degli *occupati non prevalenti* (ovvero gli individui considerati occupati secondo la definizione ISTAT ma con un'attività economicamente non tale da farli annoverare tra gli occupati *tout court*) e al lordo di alcune categorie di disoccupati (o in cerca di lavoro) che invece non rientrano nella definizione ISTAT (*inattivi che cercano lavoro*). Questo rende possibile la somministrazione di quesiti estremamente dettagliati sulla natura e le caratteristiche del lavoro, dell'istruzione e della condizione familiare, consentendo - attraverso moduli dedicati - di fornire stime attendibili anche per aggregati molto piccoli o temi specifici. L'impianto dell'indagine consente, infatti, di produrre stime statisticamente significative di aggregati anche relativamente poco numerosi nella popolazione (70.000-100.000 individui), pari a circa lo 0,5% dell'occupazione, con una probabilità del 95% che l'intervallo compreso tra $\pm 5\%$ del valore stimato comprenda il valore vero corrispondente.

Nella tab. 12.1 è riportato un confronto tra le stime RCFL e PLUS (anno 2008). Premettiamo che secondo le classificazioni ISTAT si intendono per collaboratori i soli *collaboratori puri*, cioè coloro che pur essendo lavoratori autonomi non sono professionisti e svolgono la loro attività prevalentemente attraverso forme di lavoro di collaborazione coordinata e continuativa o a progetto.

Tabella 12.1 Confronto tra le stime ISTAT-RCFL e ISFOL PLUS

ISTAT - RCFL		ISFOL PLUS	
Definizione	Numerosità	Numerosità	Definizione
Dipendenti permanenti (a tempo indeterminato)	15.038.430	14.693.754	Dipendente a tempo indeterminato
		329.662	Altro Dipendente**
Dipendenti a termine (a tempo determinato)	2.295.215	1.241.182	Dipendente a tempo determinato
		1.069.047	Altre forme lavoro dipendente a termine*
Totale dipendenti	17.333.646	17.333.646	
Collaboratori e consulenti	442.422	925.707	"Finte" Collaborazioni Coord. e Continuative e a Progetto
		380.051	Collaborazioni (<i>proprie</i>) e Collaborazioni Occasionali
Autonomi	5.193.835	3.765.908	Imprenditori e professionisti
		276.192	"Finte" partite IVA
		288.400	Altro Autonomo**
Totale autonomi	5.636.258	5.636.258	
Totale occupati	22.969.904	22.969.904	

* CFL, Apprendistato, Contratto d'inserimento, Lavoro interinale o a somministrazione, Job sharing o lavoro ripartito, Lavoro intermittente o a chiamata, Alternanza scuola-lavoro, Stage, Pratica professionale, Tirocinio.

** Lavoro con "contratto informale" - "non sa o non ricorda": riclassificati in lavoro dipendente o autonomo attraverso l'utilizzo di controlli previsti nel questionario e, per una componente residua, su base probabilistica.
Fonte: elaborazione su microdati ISTAT-RCFL e ISFOL PLUS 2008

L'indagine PLUS identifica anche i finti collaboratori (definiti *parasubordinati*), ovvero quegli occupati con forme di lavoro autonome che svolgono lavori con modalità tipiche del lavoro dipendente. Una voce a sé stante è stata prevista per le finte Partite IVA e le collaborazioni occasionali, che sono risultate essere delle forme di lavoro fortemente subordinate.

Si rammenta che il confronto tra le popolazioni di riferimento è possibile soltanto prendendo in considerazione la sottopopolazione RCFL definita dai domini di studio PLUS, descritti nel paragrafo successivo³.

12.1 Il disegno dell'indagine

L'indagine PLUS raccoglie, in ogni annualità, informazioni su circa 38.000 individui intervistati telefonicamente attraverso un sistema CATI ed in assenza di rispondenti proxy. Il questionario, di circa 200 domande complessive, è sottoposto ad un campione di persone residenti con età compresa tra 18 e 64 anni⁴. Nell'ottica della riduzione del fastidio statistico dell'indagine, il questionario è organizzato in sezioni appositamente studiate per specifici target di popolazione (occupati, disoccupati, giovani, donne, ecc.). Inoltre, a partire dalla sua seconda annualità (2006) una consistente quota di interviste longitudinali (panel), effettuate in anni consecutivi agli stessi individui, è inclusa nel campione e include circa il 60% delle interviste per ogni annualità⁵.

La pianificazione delle interviste da effettuare è stata realizzata sulla base di un *campionamento stratificato per quote* con definizione di domini di studio parzialmente sovrapposti⁶. Il campione è suddiviso in nove target fondamentali (o domini di interesse) costituiti da:

1. *giovani occupati, in età compresa tra 18 e 29 anni;*
2. *giovani studenti, in età compresa tra 18 e 29 anni;*
3. *giovani altra condizione, in età compresa tra 18 e 29 anni;*
4. *donne attive, in età compresa tra 18 e 39 anni;*
5. *donne inattive, in età compresa tra 18 e 39 anni;*
6. *anziani attivi, in età compresa tra 50 e 64 anni;*
7. *anziani inattivi (pensionati da lavoro), in età compresa tra 50 e 64 anni;*
8. *in cerca (definizione estesa);*
9. *occupati.*

3. La scelta di focalizzare l'attenzione su una particolare - seppure molto significativa - quota della popolazione italiana è coerente con l'obiettivo prioritario dell'indagine PLUS: fornire stime attendibili di fenomeni rari e marginali, ossia analizzare nel dettaglio le composizioni degli stock di riferimento forniti dall'ISTAT tramite la Rilevazione sulle Forze di lavoro (RCFL).

4. Nel 2005 e 2006 la fascia di età degli intervistati era compresa tra i 15 e i 64 anni.

5. La presenza di interviste panel contribuisce ad "alleggerire" ulteriormente il questionario, evitando di effettuare domande su fenomeni invariati nel tempo a persone intervistate l'anno precedente (esempio: primo anno di inizio dell'attività lavorativa, anno di conseguimento e caratteristiche del titolo di studio posseduto - se invariato rispetto all'intervista precedente - ecc.).

6. La scelta del campionamento per quote è stata motivata dall'esigenza di ridurre notevolmente la numerosità campionaria necessaria alla produzione di stime statisticamente significative per piccole sottopopolazioni di interesse. Come scelta alternativa si sarebbero potute utilizzare strategie classiche di campionamento a due stadi (es: comuni e famiglie) che però, oltre a richiedere rilevazioni sul campo molto più onerose, avrebbero comportato la rinuncia ad una delle principali caratteristiche dell'indagine PLUS, quella dell'assenza di rispondenti proxy.

Allo scopo di poter fornire stime attendibili anche per sottopopolazioni di questi 9 domini (ad esempio, limitatamente a ciascuna delle 20 regioni italiane) si è proceduto alla pianificazione di un *campionamento stratificato*, dove gli strati - definiti dall'incrocio delle variabili riportate in tab. 12.2 - costituiscono una partizione del campione e (per sottoinsiemi di strati) degli stessi domini di studio.

Tabella 12.2 Variabili di stratificazione campionaria, PLUS 2008

Variabili	Modalità
Regione	Piemonte e Valle d'Aosta, Lombardia, Trentino A.A., Veneto, Friuli V.G., Liguria, Emilia Romagna, Toscana, Umbria, Marche, Lazio, Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sicilia, Sardegna
Tipo di comune	Comune metropolitano, Comune non Metropolitano
Sesso	Maschi, Femmine
Età in classi	18-24, 25-29, 30-39, 40-49, 50-64
Condizione occupazionale	Occupato, In cerca di occupazione, Studente, Pensionato da lavoro, Altro inattivo (casalinga)

12.1.1 Piano di campionamento e riporto all'universo

Il numero di interviste da realizzare per ciascuno degli strati è stato determinato in modo da fornire stime attendibili per l'intera popolazione di riferimento e per particolari sottoinsiemi d'interesse, attraverso l'implementazione di una procedura di allocazione multi-dominio (*multi-domain allocation*), basata sulla risoluzione di un problema di minimizzazione vincolata. Più precisamente, sono stati fissati a priori livelli di varianza minima per i domini di interesse elencati sopra e le loro disaggregazioni territoriali per regione e tipo di comune (metropolitano, non metropolitano). Formalmente, possiamo rappresentare la numerosità della popolazione in ogni dominio (o sotto dominio) come semplice somma di sottopopolazioni di strato,

$$N_d = \sum_h N_h I_{h,d}$$

dove $I_{h,d}$ è una variabile indicatrice che assume valore 1 (0) se lo strato h è (non è) incluso nel dominio d . Indicando con p_d la generica stima della proporzione $P_d = N_d / N$ della popolazione nel dominio d , la procedura è sintetizzata dal seguente problema di minimo vincolato,

$$\begin{aligned} \sum_h n_h &= \min \\ \text{s.t. } V(p_d) &\leq V_d^* \quad \forall d \end{aligned}$$

dove il limite V_a^* varia opportunamente per diversi sotto domini. La criticità tipica dei campionamenti per quote, rappresentata dalla non conoscenza a priori delle probabilità di inclusione delle unità di rilevazione (ovvero, delle numerosità N_h), è stata superata considerando come popolazione di riferimento quella ottenuta tramite stima dalla rilevazione RCFL dell'ISTAT.

La fase successiva, che si vuole di seguito descrivere, è relativa alla scelta del tipo di stimatore di ponderazione vincolata da adottare ai fini del calcolo del coefficiente di riporto all'universo. In particolare, si è ricorsi all'implementazione di metodologie basate sull'utilizzo dello stimatore di regressione generalizzato (GREG estimator). Esso garantisce che le stime delle frequenze assolute delle variabili ausiliarie utilizzate come regressori siano coincidenti con i totali noti osservati nella popolazione ed imposti come vincoli di calibrazione. In primo luogo ciò implica che la composizione demografica e occupazionale della popolazione reale di riferimento sia, per costruzione, riprodotta in sede di analisi. Inoltre, esso permette di correggere eventuali distorsioni causate da fattori non completamente controllabili nella fase di progettazione dell'indagine e connessi, invece, alla fase di rilevazione, come l'autoselezione del campione dovuta alla maggiore propensione media alla risposta telefonica di talune categorie di soggetti. Consideriamo una generica variabile di interesse Y e definiamo il suo totale sulla popolazione di riferimento P come,

$$Y = \sum_{k \in P} y_k$$

dove k è una generica unità appartenente a P . Seguendo un approccio di tipo totalmente predittivo⁷, lo stimatore \hat{Y} di Y è formalmente derivabile attraverso la somma di due semplici componenti:

$$\hat{Y} = Y_S + \tilde{Y}_{\bar{S}} = \sum_{k \in S} y_k + \sum_{k \in \bar{S}} \tilde{y}_k$$

dove Y_S è la parte osservata sul campione e $\tilde{Y}_{\bar{S}}$ è la somma dei valori predetti \tilde{y}_k (con $S \cup \bar{S} = P$). Il passo fondamentale è quello di ipotizzare l'esistenza di una relazione (lineare) tra la generica variabile di studio Y e un opportuno insieme di variabili esplicative $x=(x_1, \dots, x_K)$ tale che,

$$\tilde{y}_k = B' x_k + \varepsilon_k \quad (1)$$

dove la componente errore ε_k soddisfa le ipotesi standard di omoschedasticità - $E(\varepsilon_k) = 0$, $E(\varepsilon_k)^2 = V(\varepsilon_k) = \sigma^2$ - e x è definito come il vettore di variabili ausiliarie. Come già anticipato, è stato implementato un approccio basato sullo stimatore di

7. Vedi Dorfman A.H., Royall R.M., Valliant R. (2000).

regressione generalizzato (GREG), nel caso particolare dei modelli con variabili strumentali⁸. Data la probabilità di inclusione π_k per ogni unità k del campione, i pesi base $d_k = 1/\pi_k$, e il conseguente stimatore standard di Horwitz-Thompson $\hat{Y}_{HT} = \sum_s d_k y_k$, è stato possibile definire il seguente stimatore finale;

$$\hat{Y} = \sum_{k \in S} w_k y_k \quad (2)$$

dove,

$$w_k = d_k (1 + \gamma' z_k)$$

$$\gamma' = \left(\sum_{k \in U} x_k - \sum_{k \in S} d_k x_k \right) \left(\sum_{k \in S} d_k z_k x_k' \right)^{-1}$$

$$z_k = x_k$$

L'utilizzo dello stimatore (2) - ossia, del fattore di correzione γ - assicura il rispetto della fondamentale equazione di calibrazione $\sum_{k \in S} w_k x_k = \sum_{k \in U} x_k$.

La procedura sopra descritta consente di: i) garantire la coerenza tra le stime prodotte tramite i dati PLUS e RCFL, sia in termini delle più generali distribuzioni demografiche che riguardo i principali indicatori del mercato del lavoro (occupati dipendenti e autonomi, persone in cerca di lavoro e inattivi) con una disaggregazione territoriale a livello regionale e dei principali comuni metropolitani italiani; ii) migliorare l'efficienza (o precisione) delle stime; iii) contribuire al controllo della distorsione dei dati campionari dovuti al ben noto fenomeno dell'attrito da selezione (dovuto principalmente alla strategia campionaria per quote, al metodo CATI di effettuazione delle interviste, nonché alla presenza di una consistente quota panel nel campione).

In generale, i risultati di una metodologia di calibrazione sono tanto più robusti quanto più "solida" è la mole di informazione ausiliaria implementata attraverso l'imposizione dei totali noti. Tra i principali prerequisiti desiderabili al fine di ottenere stimatori GERG con varianza piccola, i seguenti quattro aspetti sono stati presi in considerazione nella progettazione PLUS:

- definizione di un affidabile e plausibile modello di regressione di Y su x ;
- scelta di variabili ausiliari altamente correlate con le principali variabili di studio;
- inclusione di variabili ausiliarie già presenti tra le variabili di stratificazione;
- elevata affidabilità (buona reputazione) della fonte di dati da cui derivare i totali noti da implementare nella regressione.

In merito a questo ultimo punto, la RCFL dell'ISTAT rappresenta la fonte ufficiale di derivazione di tutti i principali indicatori del mercato del lavoro. Essa è basata su

8. Vedi Deville e Särndal (1992), Särndal e Lundström (2005).

una robusta metodologia che garantisce un'elevata precisione delle stime sull'occupazione, sulla disoccupazione e l'inattività oltre che per fondamentali distribuzioni demografiche (sesso, età) a livello nazionale, regionale e provinciale. Per questo, i totali noti imposti nella procedura PLUS sono stati scelti al fine di garantire coerenza in due fondamentali ambiti:

- la composizione socio-demografica della popolazione italiana (per regione, sesso, età e titolo di studio);
- la distribuzione degli individui tra le diverse condizioni occupazionali (occupati, attivi e inattivi) e tipologie di lavori (part-time, full-time, dipendenti, autonomi).

La tab. 12.3 riassume in un semplice schema le variabili categoriali (e le corrispondenti modalità) considerate nella definizione dei totali noti $\sum_{k \in U} x_k$. Il vettore finale di informazione ausiliaria è derivabile attraverso l'incrocio delle variabili della colonna I con quelle della colonna II. Complessivamente sono stati presi in considerazione 129 totali noti, 16 dei quali definiti - solo per le unità panel - dalla distribuzione per condizione occupazionale, tipo di lavoro, sesso ed età osservate nel 2006.

Tabella 12.3 Variabili di calibrazione cross-section, PLUS 2008

Colonna I	Colonna II	N. di totali noti
Condizione occupazionale (<i>Dipendenti full-time, Dipendenti part-time, Autonomi, Disoccupati, Inattivi</i>)	Titolo di studio (<i>Elementare, Licenza media, Diploma, Laurea</i>); Area geografica (<i>Nord-ovest, Nord-est, Centro, Sud e Isole</i>); Sesso; Età (<i>18-24, 25-29, 30-39, 40-49, 50-64 anni</i>)	75
Sesso	Età, Area geografica	18
Età	Area geografica	20
Condizione occupazionale nel 2006 (<i>Dipendenti, Autonomi, Disoccupati, Inattivi</i>)	Sesso Età nel 2006 (<i>18-39, 40-64 anni</i>)	16

12.1.2 Il peso longitudinale

Come già ricordato, l'indagine PLUS è una delle poche fonti informative sul mercato del lavoro italiano in grado di consentire analisi di tipo panel. Per ogni coppia di anni (oltre che per il periodo 2005-2006-2008) sono stati definiti i pesi di riporto all'universo longitudinale, ottenuti sulla base della stessa metodologia illustrata per il caso cross-section e con l'aggiunta di poche specifiche variazioni.

Il trattamento statistico implementato si basa sullo studio del sottostante modello di mancata risposta panel e di alcuni flussi tra condizioni di interesse primario per l'indagine, quali quello tra disoccupazione e occupazione, da lavoratore "standard" a "non standard", da part-time a full-time, ecc. Attraverso l'utilizzo di algoritmi di classificazione non parametrica - Classification tree analysis (C&RT) - si individuano delle sottopopolazioni omogenee rispetto ad alcune caratteristiche individuali, rispetto alle quali specificare i totali noti da imporre come vincoli di calibrazione longitudinale. Il peso panel finale è ottenuto a partire dal peso sezionale corrispondente all'anno di inizio periodo t_0 e applicando un correttore di calibrazione che da una parte tiene conto degli effetti di attrito, dall'altra sintetizza molta dell'informazione utile ai fini della produzione delle principali stime longitudinali. È interessante notare come per il periodo 2006-2008 la selezione endogena (da modello) delle variabili utili ai fini della spiegazione della mancata risposta panel siano rimaste pressoché invariate rispetto al 2005-2006. Per quanto riguarda, invece, lo studio delle transizioni notevoli si è preferito svincolare gli algoritmi di classificazione dai risultati precedenti, garantendo "omogeneità metodologica" piuttosto che una più stringente stabilità nella definizione delle sottopopolazioni di riferimento.

Il panel triennale 2005-2008 è composto da 12.957 unità campionarie, corrispondenti al 32,1% del campione PLUS 2005, al lordo degli individui fuori target nel 2006 e 2008⁹. Il coefficiente di riporto all'universo triennale è stato definito attraverso i seguenti passi:

- a partire dal panel 2005-2006 è stata identificata una prima popolazione di riferimento (totali noti su variabili 2005) eliminando le unità campionarie sicuramente non eleggibili nel 2008, per la derivazione di un peso intermedio calibrato ($w_{05-06corr}$);
- successivamente, a partire dal panel 2006-2008 è stata identificata la seconda popolazione di riferimento (totali noti su variabili 2006), eliminando le unità che sicuramente non potevano essere osservate nel 2005, nonché quelle non eleggibili nel 2008;
- il peso $w_{05-06corr}$ è stato utilizzato come peso base nell'ultima calibrazione che assume come totali noti gli aggregati ottenuti da più di 40 stime derivabili dalla popolazione definita al punto 2.

Sul versante della riduzione degli errori non campionari, a partire dal 2009, è stato implementato un sistema integrato di monitoraggio delle interviste con un duplice obiettivo: i) monitorare in tempo reale le interviste effettuate, con il fine ultimo di limitare e, per quanto possibile, correggere in corso d'opera le principali fonti di errori non campionari; ii) rendere disponibili tutte le informazioni relative ai tentativi di

9. Uomini e donne tra 62 e 64 anni nel 2005, più il collettivo delle casalinghe tra 40 e 49 anni presenti come target d'indagine solo nel 2005.

chiamata e sistematizzare le stesse ai fini del calcolo di alcuni importanti indicatori sulla qualità dell'indagine.

Occorre prima di tutto evidenziare che, essendo l'intervista effettuata solo a rispondenti diretti, la raccolta di informazioni su aspetti molto dettagliati del tipo di lavoro e delle sue modalità di svolgimento, della ricerca di lavoro e, più in generale, della partecipazione attiva degli individui, è avvenuta con una bassa incidenza di errori di misura, che sono invece tipici delle indagini telefoniche basate su rispondenti proxy.

Con riferimento alla qualità dell'intervista assumono un'importanza notevole i seguenti fattori: i criteri adottati di selezione degli intervistatori, la loro esperienza lavorativa pregressa, il livello di istruzione posseduto e, soprattutto, la formazione specifica all'indagine in oggetto. A tale proposito è opportuno sottolineare che ogni intervistatore PLUS è stato sottoposto ad un training formativo di otto ore di teoria (su obiettivi, definizioni, classificazioni, ecc.) alternate ad altrettante ore di pratica svolta in sala direttamente sul sistema CATI. La presenza di un responsabile ISFOL (oltre a quella del personale dipendente della società appaltatrice) è stata garantita per ogni gruppo di *briefing* (10 gruppi in totale) per almeno sette giorni successivi all'inizio delle interviste, con ulteriori interventi di *debriefing* generali o per gruppi specifici di intervistatori. Il monitoraggio, e conseguente produzione in tempo reale¹⁰ di indicatori sintetici sull'andamento delle interviste, ha permesso di effettuare interventi mirati di formazione specifica dovuta all'insorgere di particolari (ricorrenti e non isolate) "anomalie" o "difficoltà" nello svolgimento dell'intervista, sempre garantendo l'anonimato del personale coinvolto. I ricercatori ISFOL sono stati così in grado di risolvere in itinere problemi di esecuzione della fase di rilevazione che davano origine a durate eccessive, perdite di nominativi utili (dovute ad interviste impropriamente interrotte), scarsa percentuale di contatti non finalizzati, ecc.

L'implementazione di tale sistema di monitoraggio integrato delle interviste è stato reso possibile grazie alla fornitura giornaliera di tutte le informazioni richieste dal gruppo di lavoro PLUS e sintetizzabili nelle seguenti due tipologie di file:

- quello contenente informazioni circa i tentativi di contatto (esito temporaneo o finale, durata, giorno e ora, numero identificativo del gruppo di *briefing* dell'intervistatore, identificativo - criptato - dell'intervistatore, variabili di stratificazione);
- un secondo file contenente i record finali delle interviste giornaliere concluse con esito positivo, corredati del codice univoco intervistatore con il quale procedere al collegamento diretto con il file precedente.

Il sistema di monitoraggio dell'indagine PLUS ha permesso di limitare notevolmente l'insieme degli errori riscontrabili sul database definitivo, che in passato hanno richiesto un dispendio notevole di lavoro da dedicare alle fasi di controllo e correzione *ex-post*.

10. Si intende per ogni turno lavorativo giornaliero, in genere mattutino-pomeridiano e serale.

Esso garantisce oltretutto la possibilità di effettuare analisi sulla qualità dell'indagine secondo gli standard consolidati a livello internazionale (indicatori AAPOR), svolgere una selezione mirata degli intervistatori, o proporre analisi sull'effetto intervistatore attraverso l'adozione di modelli multilivello. Questo ultimo strumento è stato già implementato per l'indagine PLUS, ma soltanto ai fini di una opportuna gestione e controllo del processo di rilevazione¹¹.

11. Vedi Corsetti *et al.* (2010) per un approfondimento sulle metodologie e gli strumenti adottati a tale scopo.