

istat working papers

N.19
2015

Soluzioni metodologiche per l'utilizzo integrato delle fonti statistiche per le stime dell'occupazione

AA.VV.

istat working papers

N.19
2015

**Soluzioni metodologiche per l'utilizzo
integrato delle fonti statistiche
per le stime dell'occupazione**

AA.VV.

Comitato scientifico

Giorgio Alleva
Tommaso Di Fonzo
Fabrizio Onida

Emanuele Baldacci
Andrea Mancini
Linda Laura Sabbadini

Francesco Billari
Roberto Monducci
Antonio Schizzerotto

Comitato di redazione

Alessandro Brunetti
Romina Fraboni
Maria Pia Sorvillo

Patrizia Cacioli
Stefania Rossetti

Marco Fortini
Daniela Rossi

Segreteria tecnica

Daniela De Luca Laura Peci Marinella Pepe Gilda Sonetti

Istat Working Papers

**Soluzioni metodologiche per l'utilizzo integrato delle fonti statistiche
per le stime dell'occupazione**

N. 19/2015

ISBN 978-88-458-1868-4

© 2015

Istituto nazionale di statistica
Via Cesare Balbo, 16 – Roma

Salvo diversa indicazione la riproduzione è libera,
a condizione che venga citata la fonte.

Immagini, loghi (compreso il logo dell'Istat),
marchi registrati e altri contenuti di proprietà di terzi
appartengono ai rispettivi proprietari e
non possono essere riprodotti senza il loro consenso.

Soluzioni metodologiche per l'utilizzo integrato delle fonti statistiche per le stime dell'occupazione

AA.VV.¹

Sommario

Nel paper viene illustrata la metodologia utilizzata per l'integrazione statistica dei microdati del campione della rilevazione delle forze di lavoro con quelli delle fonti amministrative che in maniera pressoché esaustiva coprono l'intero spettro delle posizioni lavorative regolari. Gli anni di riferimento sono il 2010 e il 2011. Viene in particolare mostrato come il processo di integrazione abbia consentito di individuare, misurare e riconciliare le discrepanze emerse tra l'indagine e le fonti amministrative. È stata inoltre messa a punto una procedura di stima individuale dell'input di lavoro non regolare e delle ore effettivamente lavorate associate alle posizioni lavorative regolari e non regolari, basata sulla misurazione e correzione degli errori di misura della rilevazione attraverso con il supporto di variabili ausiliarie. L'output finale è rappresentato da una base dati costituita da tutte le posizioni lavorative generate dagli occupati nel campione dell'indagine stimati attraverso il processo di integrazione, con indicazione della natura regolare o meno della singola posizione e con le principali caratteristiche di ciascuna posizione: ordine, regolarità, attività economica, ore lavorate, tipo di posizione e datore di lavoro.

Parole chiave: Rilevazione delle forze di lavoro, Fonti amministrative, Integrazione delle fonti, Occupazione, Posizioni lavorative, Ore effettivamente lavorate, Input di lavoro non regolare

Abstract

This paper describes the methodology adopted for the statistical integration of the microdata of the labour force survey with those extracted from the administrative sources which almost exhaustively trace regular jobs in Italy. The reference years are 2010 and 2011. It is shown how statistical integration has brought to the identification, measurement and reconciliation of the incoherences between survey and administrative signals. Through the integrated sample it has been possible to provide individual estimates of regular and irregular employment and jobs. Weekly hours worked per job have also been estimated, with the correction through auxiliary variables of survey measurement errors. The final output is an integrated database where each job performed by those who are estimated as being in employment according to the integrated sample is described by a vector of integrated statistical information: inter alia, order, regularity, economic activity, actually worked hours, type, characteristics of the employer.

Keywords: Labour force survey, Administrative data, Data integration, Employment, Jobs, Hours actually worked, Non regular labour input

¹ Il paper è un estratto del report finale del Gruppo di lavoro Istat dedicato allo "Sviluppo di soluzioni metodologiche per l'utilizzo integrato delle fonti statistiche per le stime dell'occupazione". L'obiettivo del Gruppo di lavoro è stato quello di mettere a punto la nuova metodologia di stima dell'input di lavoro per il benchmark 2011 della contabilità nazionale, con particolare riferimento alla stima della componente non regolare. Ne hanno fatto parte, e vanno dunque considerati come autori di questo lavoro, Federica Battellini e Carlo De Gregorio (coordinatori), Patrizia Altieri, Ciro Baldi, Andrea Bendandi, Danilo Birardi, Cecilia Casagrande, Claudio Ceccarelli, Andrea Cutillo, Simone De Angelis, Francesca Della Ratta Rinaldi, Davide Di Laurea, Antonio Discenza, Danila Filipponi, Gerardo Gallo, Annelisa Giordano, Silvia Loriga, Filippo Orpalo, Alessandro Orteni, Silvia Pacini, Federica Pintaldi, Maria Giovanna Piras, Rita Ranaldi, Roberta Ricci, Roberta Rizzi, Laura Serbassi, Giampiero Siesto, Marina Sorrentino, Mattia Spaziani, Andrea Spizzichino, Francesca Tartamella, Mauro Tibaldi, Caterina Viviano. La stesura del report finale e del paper è stata curata da Carlo De Gregorio e Annelisa Giordano. Tutti i dati riportati nel testo, nei grafici e nelle tavole sono l'esito di procedure sperimentali, e in quanto tali non sono confrontabili direttamente con analoghe stime diffuse dall'istituto. La bibliografia finale riporta i testi raccolti nel corso dell'attività del GdL su vari argomenti, essenzialmente riguardanti aspetti metodologici e occupazione non regolare: solo una parte di tali testi è effettivamente citata nel testo. Nel paper sono inoltre presenti riferimenti a documenti e presentazioni messe a punto nel corso delle attività del GdL e in buona parte non ancora pubblicati. Le opinioni espresse impegnano esclusivamente gli autori e non implicano alcuna responsabilità da parte dell'Istat.

Indice

| | Pag. |
|--|------|
| Introduzione e sintesi..... | 7 |
| 1. Il modello di stima dell'input di lavoro: aspetti introduttivi e visione d'insieme..... | 9 |
| 2. La stima degli occupati | 27 |
| 3. La stima delle posizioni lavorative | 54 |
| 4. La stima delle ore effettivamente lavorate | 73 |
| Conclusioni: processo produttivo, rischi, sviluppi | 111 |
| Riferimenti bibliografici | 115 |

Introduzione e sintesi

Il presente working paper è un estratto del report finale del Gruppo di lavoro Istat dedicato allo “Sviluppo di soluzioni metodologiche per l'utilizzo integrato delle fonti statistiche per le stime dell'occupazione” (d'ora in avanti indicato con GdL-OCC o più semplicemente GdL), il quale ha avuto “il compito di individuare e mettere a punto soluzioni metodologiche per la stima annuale dell'input di lavoro per la contabilità nazionale a partire dall'anno di benchmark 2011 (...) con particolare riferimento alle ore lavorate e all'occupazione non regolare (occupati, posizioni e ore lavorate, distinti in regolari, grigie e non regolari) e assicurando un uso integrato delle basi informative presenti in Istituto”. Il GdL ha iniziato le proprie attività nell'ultimo trimestre del 2012 e le ha terminate a dicembre 2013. Il report finale è stato consegnato nel mese di aprile 2014, preceduto da un report intermedio all'inizio del 2013.

L'obiettivo del GdL è stato quello di fornire un nuovo sistema di stime relative all'input di lavoro – e più specificamente numero di occupati, posizioni lavorative e ore effettivamente lavorate - a supporto delle attività relative al nuovo benchmark dei conti nazionali: se per un verso questo obiettivo includeva senz'altro la stima della componente regolare, dall'altro al GdL è stato in particolare richiesto di mettere a punto una nuova procedura per la stima su base annuale della componente non regolare dell'input di lavoro. Il GdL ha conseguentemente sviluppato un sistema integrato di stime per gli anni 2010 e 2011 degli occupati nel campo di osservazione della rilevazione sulle forze di lavoro (LFS) accompagnato da un impianto teorico che accoglie anche le stime relative al resto del campo di osservazione (convivenze e stranieri non residenti), non sviluppate dal GdL (descritto nel par. 3 della Sez.1).

Nel modello teorico generale finalizzato al benchmark (Sez.1 par.3) le stime delle posizioni lavorative regolari convergono su una popolazione di riferimento determinata dalle posizioni lavorative dei registri statistici delle unità economiche: ciò allo scopo di sfruttare appieno le informazioni delle fonti amministrative, che in termini di principio offrono una copertura esaustiva delle posizioni regolari. La stima della componente regolare è pertanto indipendente dalle stime della popolazione utilizzate per il riporto all'universo di LFS: l'ipotesi alla base di questo approccio è che le posizioni lavorative stimate attraverso i registri statistici dell'Istat (che hanno come campo di osservazione una popolazione presente) siano compatibili con le stime relative alla popolazione residente utilizzata per il riporto all'universo dell'indagine. La verifica di questa ipotesi ha avuto esito positivo, in particolare con riferimento al confronto con le stime post-censuarie della popolazione che incorporano la valutazione della sottocopertura censuaria (le cosiddette stime PES).

Le posizioni e gli occupati non regolari nel campo di osservazione LFS dipendono invece dal sistema di ponderazione dell'indagine.

Il sistema integrato di stime si basa sull'aggancio tramite codice fiscale dei microdati LFS con l'insieme delle informazioni individuali tratte dalle fonti amministrative sugli occupati (indicate d'ora in poi genericamente con ADMIN), prevalentemente di natura previdenziale, che descrivono in maniera esaustiva l'occupazione regolare. Le informazioni ADMIN sono trattate come una sezione aggiuntiva del questionario LFS. È stato così creato il campione integrato LFS-ADMIN che aggiunge alle variabili rilevate dall'indagine informazioni dettagliate sulle posizioni lavorative tracciate in ADMIN e sui corrispondenti datori di lavoro presenti nei registri statistici (Sez.1 par.2 e Sez.2 par.1).

Il confronto individuale fra informazioni LFS e ADMIN è stato utilizzato per individuare, misurare e riconciliare le incoerenze fra le fonti. In particolare la determinazione della condizione di occupazione e delle sue caratteristiche è il risultato del metodo di risoluzione dei casi di incoerenza fra segnali di occupazione LFS e ADMIN. Questo metodo (Sez.2 par.2-4) è alla base della nuova stima delle tre componenti dell'input di lavoro messa a punto dal GdL: l'occupazione (Sez.2), le posizioni lavorative (Sez.3) e le ore lavorate (Sez.4).

Le incoerenze tra le informazioni contenute in ADMIN e LFS derivano essenzialmente da due tipologie di cause: la sovrastima della condizione di occupato da parte di ADMIN e la sottostima da parte di LFS (Sez.2 par.3). Esse si manifestano principalmente attraverso la non trascurabile incidenza di casi di non occupati LFS con coperture previdenziali attive nella settimana di riferimento dell'intervista LFS.

A questo proposito, è opportuno evidenziare la diversa natura e finalità delle due fonti. Da un lato la fonte ADMIN, ottenuta mediante integrazione di diverse fonti amministrative, prevalentemente di natura previdenziale; la fornitura delle informazioni riportate in tali fonti è generalmente effettuata dalle imprese o dai lavoratori autonomi a fini di certificazione amministrativa e copre in modo pressoché esaustivo l'occupazione regolare. Dall'altro lato la fonte LFS, un'indagine campionaria basata su interviste effettuate presso le famiglie; l'obiettivo principale della rilevazione è classificare la popolazione in base alla condizione occupazionale, secondo le definizioni definite nei regolamenti Eurostat coerentemente con le definizioni ILO, sulla base di quesiti rivolti agli intervistati.

Dal lato ADMIN vanno dunque messe in conto situazioni di disallineamento tra posizione contributiva ed effettivo svolgimento dell'attività lavorativa, che per vari motivi si possono presentare, ad esempio nel caso in cui si effettuino versamenti di contributi previdenziali a fini pensionistici non legati all'effettivo svolgimento di un'attività lavorativa, oppure in caso di disallineamento temporale tra lavoro svolto e versamento dei contributi o infine a seguito della difficoltà di ricondurre il dato ADMIN alla settimana di riferimento LFS.

Dal lato LFS vanno invece messe in conto situazioni in cui per vari motivi l'intervistato non riporta in sede di intervista l'effettivo svolgimento di un'attività lavorativa nella settimana di riferimento (ad esempio a causa di un "effetto memoria" in relazione alla distanza che intercorre tra la settimana in cui si conduce l'intervista e la settimana di riferimento, oppure per limitare il carico statistico legato all'intervista), oppure tale "omissione" è dovuta alla conduzione di un'intervista proxy (in cui cioè non è il diretto interessato a rispondere ai quesiti, ma un suo familiare), oppure ancora la mancata registrazione dell'attività lavorativa avviene a causa di un comportamento irregolare da parte dell'intervistatore.

Inoltre, trattandosi di un'analisi che si basa sul *record linkage* tra dati ADMIN e dati LFS, vanno messe in conto imperfezioni nelle operazioni di aggancio delle informazioni o nelle chiavi di aggancio (il codice fiscale). Infine, va tenuto presente che in una minoranza di casi l'assenza del codice fiscale in LFS è stata risolta mediante imputazione. In tali casi è naturale che si verifichi l'incoerenza tra le informazioni nelle due fonti.

Tutte queste precisazioni sono necessarie per poter correttamente interpretare le discrepanze tra le due fonti. Queste incoerenze del resto sono state approfonditamente studiate dal GdL al fine di identificare gli opportuni metodi di correzione al fine di giungere alle stime dell'input di lavoro di per il benchmark della contabilità nazionale, in termini di occupati, posizioni lavorative e ore lavorate (che costituivano peraltro l'obiettivo finale del GdL stesso).

Nel prosieguo del lavoro tali incoerenze vengono approfonditamente analizzate e opportunamente trattate, talvolta ci si riferirà ad esse anche con termini quali "distorsioni" o "discrepanze", senza per questo voler esprimere un giudizio sulla "bontà" delle rispettive fonti. Le considerazioni fin qui fatte aiuteranno a interpretare correttamente le valutazioni che verranno illustrate nel seguito.

L'utilizzo di un modello probabilistico (Sez.2 par.3) ha consentito di produrre per ciascun individuo un indicatore di validazione del segnale di occupazione ADMIN: un segnale ADMIN validato indica per ipotesi una posizione lavorativa regolare a prescindere dal segnale LFS. In maniera complementare, gli occupati LFS privi di segnali ADMIN validati sono classificati fra gli occupati non regolari.

È stato di conseguenza stimato per dominio un tasso di sottocopertura dell'occupazione regolare da parte di LFS. Nell'ipotesi che questo tasso sia indipendente dalla natura regolare o irregolare della posizione lavorativa, esso è utilizzato per stimare per strato un numero aggiuntivo di occupati non regolari da estrarre con criteri probabilistici dai non occupati LFS privi di segnali validati.

Il numero di occupati del campione integrato LFS-ADMIN è perciò dato dalla somma degli occupati regolari (titolari di posizioni lavorative in ADMIN validate a prescindere dalla condizione LFS) e degli occupati non regolari (occupati LFS senza coperture ADMIN validate e non occupati LFS recuperati dalle stime della sottocopertura).

Vengono così stimati circa 1,2 milioni di occupati in più rispetto a LFS: di questi, oltre un milione sono occupati con almeno una posizione ADMIN validata; altri 150 mila circa sono privi di posizioni regolari (Sez.2 par.4).

La stima delle posizioni lavorative (Sez.3) ha avuto come obiettivo la creazione di un archivio delle

posizioni lavorative regolari e non regolari associate a tutti gli intervistati LFS (archivio LFS-ADMIN_JOBS). Per ogni posizione lavorativa è stato prodotto un vettore con le sue caratteristiche (gerarchia, regolarità, tipologia, settore, territorio, classe dimensionale, forma giuridica del datore, ecc.).

La singola posizione regolare è identificata dalla coppia individuo-datore di lavoro che nella settimana di riferimento dell'intervista LFS ha un segnale amministrativo validato. La sola considerazione delle posizioni regolari evidenzia la consistente sottodichiarazione a LFS delle posizioni secondarie e ha reso possibile la stima di una misura di questa sottocopertura per dominio.

Il numero delle posizioni lavorative non regolari viene derivato dal confronto condotto a livello individuale fra il numero di posizioni regolari identificate in ADMIN e il numero di posizioni rilevate da LFS: è corretto al rialzo con un modello probabilistico nell'ipotesi che la sottocopertura delle posizioni non regolari sia di intensità uguale a quella delle posizioni regolari.

Il metodo stima nel 2010 circa 26 milioni di posizioni lavorative di cui il 10,4% non regolari e il 9,6% secondarie (Sez.3 par.6). Nell'agricoltura una posizione su cinque non è regolare, seguono costruzioni commercio e servizi alle famiglie. Nel 2011 si ha una contrazione delle posizioni lavorative non regolari e una tenuta delle regolari. Se si confronta con LFS, la nuova procedura stima oltre 3 milioni di posizioni in più: per due terzi si tratta di posizioni secondarie.

Con la stima delle ore lavorate (Sez.4) viene associato a ogni record di LFS-ADMIN_JOBS il numero di ore effettivamente lavorato per ciascuna posizione lavorativa nella settimana di riferimento. È il risultato di un lavoro di imputazione statistica – prevalentemente tramite donatore con una tecnica hot deck sequenziale - che sfrutta le variabili LFS sulle ore lavorate e fa leva sia sulle informazioni strutturali relative all'intervistato LFS sia sull'informazione relativa alla posizione lavorativa reperita anche attraverso ADMIN.

Preliminarmente sono stati stimati e corretti alcuni errori di misura che caratterizzano le variabili LFS sulle ore lavorate: quelli legati alla misura delle giornate non lavorate a causa di ferie e festività (effetto questo risultato in realtà non significativo), alle assenze per congedi e malattie, alle interviste proxy, alla propensione a cumulare sulla posizione principale le ore lavorate nelle secondarie.

Il modello ha stimato 42 miliardi di ore lavorate nel 2010 di cui un po' meno di 4 miliardi non regolari (Sez.4 par.8). Il tasso di irregolarità stimato in termini di ore lavorate è più basso di circa un punto rispetto a quello stimato in base alle posizioni lavorative, a causa di un minore intensità oraria delle posizioni lavorative non regolari.

Sono molte e estremamente rilevanti le aree di contaminazione e le potenzialità che derivano dallo sviluppo delle attività di integrazione del tipo sperimentato con LFS-ADMIN: per esempio, le stime relative alla componente grigia dell'input di lavoro, il supporto alle attività innovative relative ai processi di LFS e di ADMIN, gli approfondimenti possibili per accrescere la conoscenza e le caratteristiche degli occupati non regolari, la possibilità di qualificare ulteriormente le analisi di tipo longitudinale condotte sia sui dati campionari sia sui record delle fonti amministrative, la possibilità di estendere agevolmente l'approccio ad altre indagini sulle famiglie, le prospettive che si aprono attraverso il ponte che viene gettato fra le statistiche sulle imprese e quelle sulle famiglie, a cominciare dalla possibilità di contribuire efficacemente al dibattito relativo alla produttività del lavoro.

1. Il modello di stima dell'input di lavoro: aspetti introduttivi e visione d'insieme

1.0 Sintesi del paragrafo

Il modello di stima dell'input di lavoro è un processo di produzione annuale, al momento utilizzabile per la stima a circa 24 mesi dalla fine dell'anno di riferimento ($t+24$). Utilizza tutte le informazioni disponibili in Istat con un approccio di integrazione e senza costi diretti aggiuntivi; produce esternalità positive su altri settori della produzione statistica; è coerente con i processi innovativi e le strategie di lungo periodo di Istat ed Eurostat; è coerente con le competenze metodologiche dell'Istat; è compatibile con le risorse dedicate alle stime della Contabilità nazionale; è ispirato a principi di trasparenza e contendibilità.

Il GdL ha sviluppato e messo in produzione il sistema di stime dell'input di lavoro (occupati,

posizioni lavorative e ore lavorate) fondandolo sulla definizione del campione integrato LFS-ADMIN, ottenuto agganciando attraverso i codici fiscali i microdati LFS con le fonti amministrative individuali sull'occupazione (ADMIN).

Lo sviluppo delle stime riferite alla porzione del campo di osservazione del benchmark della contabilità nazionale (CN) non osservato tramite LFS (cittadini stranieri non residenti e cittadini residenti presso le convivenze) non è stato sviluppato dal GdL ma è stato curato direttamente dalla contabilità nazionale con una metodologia analoga a quella messa a punto dal GdL ma che evidentemente non ha potuto contare sulla ricca base informativa fornita da LFS.

Il GdL ha prodotto il quadro teorico-metodologico unitario che assicura la coerenza delle stime relative a queste due componenti del campo di osservazione e del processo di stima.

Le informazioni ADMIN agganciate ai singoli individui intervistati tramite LFS vengono trattate come una sorta di sezione aggiuntiva del questionario. Il confronto e l'analisi della coerenza fra le informazioni LFS e le informazioni ADMIN sono utilizzate per stimare e riconciliare le incoerenze delle due fonti e per stimare le caratteristiche dell'input di lavoro, e in particolare delle componenti regolari e non regolari.

Il modello si articola in tre fasi, organizzate in sequenza: stima degli occupati (sez.2), delle posizioni lavorative (sez.3) e delle ore effettivamente lavorate (sez.4).

La procedura di stima è costituita da un articolato processo di imputazione delle unità rilevate con il campione LFS basato sull'utilizzo integrato delle informazioni ausiliare ADMIN e di quelle di indagine, e volto a individuare e correggere le incoerenze delle due fonti.

Il processo di imputazione è basato sull'utilizzo in sequenza di modelli di regressione logistica e di metodi probabilistici da donatore (prevalentemente di tipo hot deck sequenziale gerarchico) o deterministici (prevalentemente imputazione mediante un valore medio o modale).

Le stime lorde delle posizioni lavorative regolari, al lordo delle sovracoperture delle fonti ADMIN, convergono sulle posizioni lavorative dei registri statistici. La stima della componente regolare delle posizioni lavorative è perciò indipendente dalle stime della popolazione. L'ipotesi di fondo è che l'occupazione stimata attraverso i registri statistici (che origina da una popolazione presente) sia compatibile con la stima della popolazione residente. Questa ipotesi appare verificata dalle stime della popolazione che incorporano la misurazione sottocopertura censuaria (le cosiddette stime PES: cfr. il paragrafo 1.1.3).

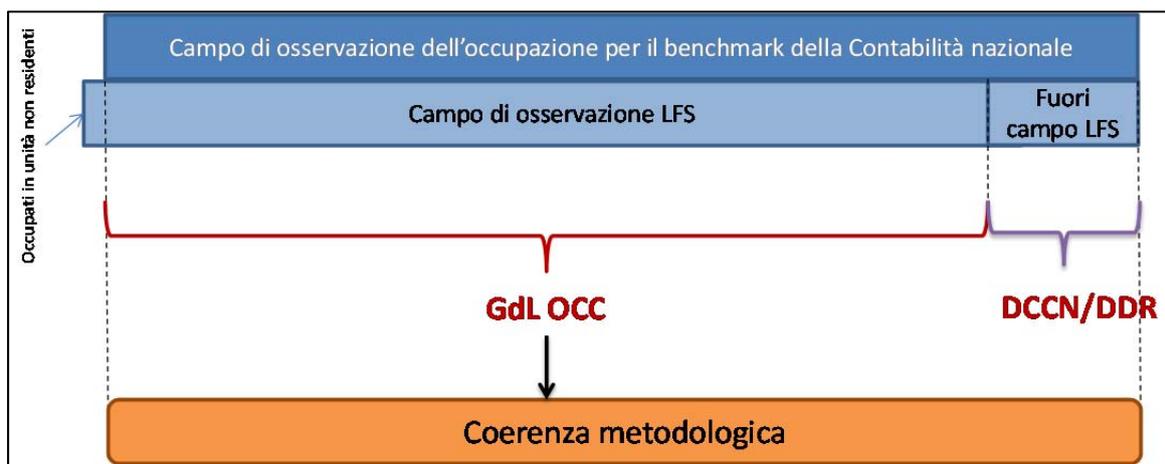
Le posizioni e gli occupati non regolari nel campo di osservazione LFS provenendo dal riporto all'universo della popolazione di riferimento delle informazioni tratte da questa indagine dipendono invece dal sistema di ponderazione dell'indagine stessa.

1.1 Aspetti introduttivi: approccio, definizioni, quadro teorico e operativo

1.1.1. *Approccio*

Il nuovo approccio alla stima delle diverse misure di occupazione proposto dal GdL è basato sull'utilizzo integrato dei microdati relativi agli individui, rilevati attraverso l'indagine continua delle forze di lavoro (LFS) e le diverse fonti amministrative sugli individui e i loro datori di lavoro (abbreviate con ADMIN). Il campo di osservazione del campione integrato LFS-ADMIN coincide dunque con quello LFS ed è pertanto costituito dagli individui residenti in famiglia mentre esclude i residenti presso le convivenze e i cittadini stranieri non residenti. È su questo segmento che si è concentrata l'attività del GdL. Il completamento della copertura del campo di osservazione ai residenti presso le convivenze e ai cittadini stranieri non residenti è stato parallelamente sviluppato dalla Contabilità nazionale utilizzando i microdati del Censimento della popolazione (CP), delle liste anagrafiche comunali (LAC) e delle fonti amministrative sui permessi di soggiorno (PDS) integrati con le fonti ADMIN, con un approccio metodologico coerente con quello sviluppato nell'ambito del GdL e avente per oggetto il campione integrato LFS-ADMIN. Questo capitolo fornisce una descrizione del quadro complessivo proposto dal GdL e attraverso il quale sono state rese coerenti le stime relative a queste due componenti (Figura 1).

Figura 1. Componenti del modello di stima dell'input di lavoro nel campo di osservazione della Contabilità nazionale



L'approccio può essere sintetizzato come segue:

- ✓ Le stime sulle diverse misure di occupazione sono ottenute utilizzando il campione dei dati individuali LFS e i relativi coefficienti di riporto all'universo per gli occupati residenti in famiglia, i dati del censimento popolazione per gli occupati in convivenza e i dati sui titolari dei permessi di soggiorno non iscritti in anagrafe per i non residenti. Dal computo sono esclusi i residenti occupati presso unità produttive non residenti;
- ✓ Le fonti amministrative integrate (ADMIN) descrivono le singole relazioni fra individui, residenti e non residenti da un lato, e datori di lavoro residenti dall'altro; i registri statistici, inoltre, forniscono informazioni sui datori di lavoro, la classe dimensionale, la forma giuridica e la localizzazione territoriale (fino alla provincia) consentendo anche la costruzione dei raggruppamenti per settore istituzionale;
- ✓ L'approccio alla stima sfrutta in generale la possibilità di aggancio di microdati individuali su individui e datori di lavoro. Per la componente più rilevante, quella dei residenti in famiglia, viene sfruttata la possibilità di integrare questi microdati con quelli di LFS: le fonti amministrative forniscono infatti "segnali di copertura" (contributiva e/o assistenziale) mentre LFS fornisce "segnali di occupazione". Nessun segnale diretto sull'occupazione, proveniente cioè da una rilevazione, è invece disponibile per i residenti presso le convivenze (a parte l'occasione censuaria e, in prospettiva, il censimento permanente), per i titolari di permessi di soggiorno e per gli altri cittadini stranieri presenti ma non residenti sul territorio nazionale;
- ✓ Le informazioni ADMIN agganciate ai singoli individui intervistati tramite LFS vengono trattate come una sorta di sezione aggiuntiva del questionario che riporta le tracce lasciate da ciascun individuo nelle banche dati che descrivono i rapporti di lavoro formalizzati;
- ✓ Il confronto e l'analisi della coerenza fra le informazioni LFS e le informazioni individuali di ADMIN sono utilizzate per imputare e stimare nuove variabili e per correggere le incoerenze fra le due fonti al fine di ottenere la stima del livello e delle caratteristiche dell'input di lavoro per il benchmark della contabilità nazionale;
- ✓ Le evidenze che emergono dall'integrazione LFS-ADMIN non solo affiancano le fonti ADMIN ma forniscono anche l'occasione di ulteriori miglioramenti metodologici per un loro corretto utilizzo, anche nella stima dell'input di lavoro riconducibile alla popolazione posta al di fuori del campo di osservazione di LFS.

Il modello di stima dell'input di lavoro è un processo di produzione annuale, al momento utilizzabile per la stima disponibile in $t+24$ mesi dei valori relativi all'anno t . Una volta a regime potrà essere messo ulteriormente a punto onde anticipare la stima e adottare periodi di riferimento infra-

annuali, fermo restando che esso è certamente assoggettabile a ulteriori miglioramenti, sia di carattere metodologico e sia relativi al processo di produzione. I principi che hanno improntato la definizione dell'architettura e il respiro strategico del modello sono i seguenti: il modello deve ...

- ✓ Essere replicabile annualmente ed essere predisposto ad accogliere ulteriori processi innovativi;
- ✓ Utilizzare tutte le informazioni disponibili in ISTAT con un approccio di integrazione e senza costi diretti aggiuntivi, e produrre possibilmente esternalità positive su altri settori della produzione statistica;
- ✓ Inserirsi nel quadro dei processi innovativi e le strategie di lungo periodo di ISTAT ed EUROSTAT;
- ✓ Essere collegato e collegabile con le competenze metodologiche dell'ISTAT e con gli indirizzi che queste intraprendono;
- ✓ Essere compatibile con le forze e le competenze disponibili per la gestione corrente, nel quadro delle risorse, degli obiettivi e delle strategie nella produzione delle stime della Contabilità nazionale;
- ✓ Essere sviluppato in un linguaggio di programmazione diffuso in istituto (nella fattispecie il SAS) ed essere interamente gestibile, anche nella manutenzione ordinaria, dai ricercatori della Contabilità nazionale;
- ✓ Essere ben documentato e contendibile;
- ✓ Privilegiare soluzioni fondate, robuste nei risultati, adatte alle competenze, il più possibile semplici da controllare, gestire e implementare.

Questa innovazione metodologica trova la sua ragione di essere e prende le mosse dal notevole investimento effettuato dall'ISTAT nell'acquisizione e nell'utilizzo delle fonti amministrative. In particolare, l'elemento fondante è la possibilità di collegare in maniera esaustiva lavoratori e datori di lavoro presenti negli archivi amministrativi e, pertanto, di ricostruire con un elevato grado di precisione i percorsi individuali legati alle forme regolari di occupazione. Come si vedrà, questa nuova prospettiva determina anche la necessità di riconsiderare da un punto di vista concettuale le categorie di lavoro regolare e non regolare, a causa della accresciuta mole di informazioni ausiliarie di carattere amministrativo disponibili (paragrafo 1).

D'altra parte questo approccio appare in linea con quello strategico suggerito da EUROSTAT nella cosiddetta *vision*²: favorire la ricerca di economie di scala e di scopo attraverso lo sfruttamento delle opportunità offerte dall'integrazione fra le fonti, siano esse di indagine o amministrative, e fra le competenze disponibili all'interno degli istituti di statistica. Di fatto la collaborazione fra i vari dipartimenti di produzione statistica dell'ISTAT operata attraverso il GdL è stata determinante per lo sviluppo di questa innovazione: pur sotto la guida della Contabilità nazionale, la condivisione di saperi ed esperienze è stata estremamente efficace in particolare con i colleghi impegnati negli altri settori della produzione statistica. Estremamente importanti, soprattutto se considerati in prospettiva, si sono inoltre rivelati, su alcuni temi specifici, i suggerimenti forniti dai metodologi dell'ISTAT i quali tuttavia non erano stati coinvolti direttamente nelle attività del GdL.

L'obiettivo di questa innovazione è stato quello di supportare le stime per il benchmark della Contabilità nazionale relative alle componenti regolari e non regolari dell'occupazione, delle posizioni lavorative e delle ore lavorate, e di aumentare contestualmente la precisione delle stime con particolare riferimento all'attività economica e alle classi dimensionali. A questo obiettivo di fondo se ne sono affiancati in corso d'opera altri legati alle numerose esternalità sui processi produttivi sia della Contabilità nazionale, sia di altri settori della produzione statistica. È il caso:

- ✓ delle rilevazioni sulle famiglie. Le informazioni ADMIN consentono ad esempio di valutare taluni errori non campionari, misurare distorsioni, ecc.;

² European Commission (2009).

- ✓ dei registri statistici. Le informazioni delle indagini sulle famiglie permettono ad esempio di migliorare la scelta dei criteri definitivi, di stabilire la loro capacità di misurare l'occupazione regolare, ecc.;
- ✓ delle rilevazioni sulle imprese. Le informazioni lato famiglia e ADMIN consentono di correggere le distorsioni dovute al tasso di mancata risposta, come ad esempio nel caso della stima delle ore lavorate, o di introdurre innovazioni, quali la stima dell'input di lavoro "grigio" e delle forme parziali di evasione.

Nel prosieguo del paragrafo vengono riepilogate le caratteristiche e i concetti di base del nuovo metodo di stima. Nei paragrafi successivi vengono dapprima descritte le basi informative; successivamente viene fornita una visione d'insieme del modello di stima, seguita da tre paragrafi dedicati separatamente alle stime degli occupati, delle posizioni lavorative e delle ore lavorate. L'attenzione verrà posta in particolare sulle stime ottenute tramite l'utilizzo della base dati integrata LFS-ADMIN e relative dunque al campo di osservazione di LFS: sono queste le stime che hanno infatti rappresentato l'oggetto specifico dell'attività del GdL.

1.1.2. Quadro teorico e operativo

Il nuovo sistema di stime dell'input di lavoro si fonda sull'utilizzo integrato dei microdati LFS e delle fonti ADMIN e si basa sui seguenti criteri:

- ✓ Le stime vengono effettuate utilizzando il campione LFS e i suoi coefficienti di riporto all'universo. Pertanto esse riguardano il campo di osservazione di LFS, ossia la popolazione residente al netto delle convivenze;
- ✓ Le informazioni ADMIN agganciate ai singoli individui intervistati tramite LFS vengono trattate come parti di informazione aggiunte a ciascun questionario LFS: una sorta di sezione aggiuntiva del questionario che riporta le tracce lasciate da ciascun individuo nelle banche dati che descrivono i rapporti di lavoro formalizzati;
- ✓ Il confronto e l'analisi della coerenza fra le informazioni LFS e le informazioni ADMIN sono utilizzate per stimare e correggere le incoerenze fra le due fonti;
- ✓ La stima delle caratteristiche dell'input di lavoro, e in particolare delle componenti regolari e non regolari, viene basata sull'analisi congiunta dei microdati LFS e delle tracce individuali ADMIN.

Il modello si articola in tre fasi, organizzate in sequenza: stima degli occupati, delle posizioni lavorative e delle ore lavorate. Il nuovo sistema è perciò costituito da una sequenza di altrettante procedure di stima a partire dal campione integrato LFS-ADMIN:

1. occupazione;
2. posizioni lavorative;
3. ore lavorate.

Si tratta di procedure che forniscono stime annuali, con un buon dettaglio territoriale e settoriale, relative alla componente regolare e a quella non regolare oltre che alle varie caratteristiche dell'input di lavoro (dipendente o indipendente, genere, classe di età, ecc.). Sono state al momento prodotte le stime relative agli anni 2010 e 2011.

Fondamentalmente la procedura di stima è costituita da un articolato processo di imputazione delle unità rilevate con il campione LFS basato sull'utilizzo integrato delle informazioni ausiliare ADMIN e di quelle di indagine, e volto alla individuazione e riconciliazione di alcuni elementi di incoerenza chiaramente presenti nelle due fonti. Il processo di imputazione è a sua volta fondato sull'utilizzo in sequenza di modelli di regressione logistica e di metodi di imputazione probabilistici tramite donatore (prevalentemente di tipo *hot deck* sequenziale gerarchico) o deterministici (prevalentemente imputazione mediante un valore medio o modale).

Si tratta di scelte che sono state discusse all'interno del GdL e che – com'è auspicabile - continueranno a suscitare ulteriori approfondimenti e migliorie di cui è senz'altro suscettibile il modello

di stima e che potranno riguardare anche altri temi, come ad esempio le tecniche di record linkage basate sui codici fiscali.

I modelli logistici, in particolare, possono essere:

- ✓ Migliorati, operando sulla specificazione dei modelli;
- ✓ Messi in discussione, sulla base del confronto con i risultati ottenibili con metodi alternativi, per esempio basati sull'analisi delle variabili latenti o su altre tecniche di analisi multivariata.

È evidente che si tratta di approfondimenti importanti e che assumono per il futuro una rilevanza strategica che va ben oltre le necessità della Contabilità nazionale. La scelta del modello logistico è in effetti stata valutata da GdL ed effettuata sulla base di considerazioni che riguardano:

- ✓ L'esito di alcune sperimentazioni condotte utilizzando analisi fattoriale e cluster;
- ✓ L'accessibilità e la relativa semplicità del metodo;
- ✓ La ragionevolezza, la qualità e la robustezza dei risultati prodotti;
- ✓ La disponibilità di stabili, adeguate e integrabili procedure SAS.

Quanto all'utilizzo di metodi di imputazione le soluzioni adottate sono anch'esse suscettibili di ottimizzazioni e miglioramenti. La scelta delle procedure hot deck è originata dalle seguenti motivazioni³:

- ✓ È stata adottata quando è stato ritenuto necessario non alterare le caratteristiche della distribuzione delle variabili oggetto dell'imputazione;
- ✓ È di gestione informatica relativamente agevole e richiede tempi di elaborazione ridotti.

Non tutte queste motivazioni sono irrinunciabili e dunque su questi metodi sarà possibile intervenire a vari livelli.

1.1.3. *Aspetti definitivi e concettuali*

Il concetto di "popolazione di riferimento" utilizzato per queste stime corrisponde a quelli definiti nei regolamenti, e in particolare nel capitolo 11 del nuovo SEC⁴ anche se per taluni di essi qualche distinguo è comunque necessario. Si possono individuare in particolare i seguenti sottoinsiemi, a loro volta eventualmente articolabili con ulteriore dettaglio:

- A. il campo di osservazione di LFS: si tratta della componente largamente maggioritaria, che rappresenta ben oltre il 95% della popolazione complessiva;
- B. la popolazione residente fuori del campo di osservazione LFS (i residenti nelle convivenze), una componente piuttosto esigua caratterizzata peraltro da un tasso di attività assai modesto;
- C. gli stranieri non residenti, a loro volta distinguibili in comunitari e non comunitari: è la componente più difficile da trattare poiché scarsamente tracciata nelle fonti amministrative.

Le stime dell'input di lavoro ottenute con la base dati integrata LFS-ADMIN dipendono dai coefficienti di riporto all'universo di LFS e dunque dalla stima della popolazione nel campo di osservazione dell'indagine. Se da un lato gli aspetti definitivi delle rispettive popolazioni di riferimento sono chiari e al più necessitano di opportune soluzioni operative, dall'altro è stata influente ai fini delle attività del GdL l'identificazione della stima della popolazione residente da utilizzare per il calcolo dei coefficienti di riporto all'universo di LFS.

Il periodo in cui il GdL ha espletato la sua attività si è infatti parzialmente e inevitabilmente sovrapposto al processo di ricostruzione della popolazione residente a seguito delle risultanze censuarie e al processo di ricostruzione dei dati della LFS coerentemente con la popolazione così ricostruita. Più precisamente, in quel periodo venivano resi disponibili dall'Istituto: a) i risultati del

³ Grande et al. 2003; Andridge et al. 2010.

⁴ Regulation (EU) no. 549/2013 of the European Parliament and of the Council of 21 May 2013 on the European System of national and regional accounts in the European Union (OJ L 174/2, 26.6.2013).

Censimento della popolazione 2011, con il rilascio della popolazione legale al 9 ottobre 2011, che ha visto una sensibile riduzione del dato di popolazione residente rispetto alla popolazione calcolata; b) i risultati della Post-census Enumeration survey (PES) o indagine di copertura, normata dalla Commissione Europea (Commission Regulation n. 1151/2010), una indagine campionaria indipendente dal Censimento della popolazione avente l'obiettivo di misurare la sotto-copertura e la sovracopertura del Censimento, che forniva una stima finale della sottocopertura censuaria; c) i risultati ancora parziali a quella data della revisione in corso delle anagrafi, generati dal consueto confronto censimento-anagrafe, effettuato dai comuni mediante il sistema SIREA (Sistema di Revisione dell'Anagrafe), che ha permesso all'Istituto e alle Prefetture di controllare e monitorare quotidianamente le operazioni e di calcolare in modo uniforme le rettifiche da apportare al calcolo della popolazione, dovute al processo di revisione anagrafica; tali risultanze implicano, come già accaduto in occasione dei precedenti censimenti, un incremento della popolazione nel periodo post-censuario essenzialmente non dipendente dalle dinamiche naturali o migratorie.

D'altro canto, al fine di garantire la qualità delle stime LFS in particolare in un'ottica di analisi intertemporale del mercato del lavoro, vi era l'esigenza di disporre di serie di dati di popolazione che non presentassero break in corrispondenza del Censimento della popolazione e con un andamento "regolare" anche nel periodo post-censuario, la cui dinamica fosse determinata esclusivamente dai movimenti "reali" cioè legati ai flussi naturali e migratori della popolazione, piuttosto che da poste "amministrative" legate alle revisioni anagrafiche.

Per le ragioni fin qui descritte, la scelta dell'Istituto, fatta nel periodo in cui erano in corso le attività del GdL, è stata quella di produrre una "popolazione statistica", ottenuta utilizzando le suddette fonti.

La popolazione "statistica" allinea pertanto le popolazioni intercensuarie del periodo 2001-2011 al dato del Censimento inclusivo della correzione PES al 9 ottobre 2011 e prosegue nel periodo post-censuario 2011-2013 allineandosi con la popolazione calcolata dell'Istat al 1 gennaio 2014, che incorpora le revisioni anagrafiche di SIREA. Tale popolazione soddisfa pertanto il requisito della coerenza con le suddette fonti e della regolarità nel tempo e si allinea con la popolazione calcolata dell'Istat al 1° gennaio 2014, elemento essenziale al fine di assicurare continuità anche nel periodo successivo⁵.

Nel periodo in cui erano in corso le attività del GdL non tutti gli elementi suddetti erano già del tutto noti; in particolare, si disponeva dei risultati del Censimento della popolazione, di una stima provvisoria della PES e delle risultanze parziali di SIREA. Conseguentemente l'impostazione delle procedure di stima adottate dal GdL si è basata su stime provvisorie di tale popolazione "statistica". I pesi della LFS sono stati pertanto ricalcolati utilizzando tale popolazione "statistica" provvisoria. Occorre tuttavia rilevare come i risultati definitivi della popolazione "statistica" sono risultati comunque molto vicini a quelli provvisori e così anche le stime LFS ricalcolate in via provvisoria e definitiva. Si può affermare senza ombra di dubbio che questo elemento di approssimazione non ha generato nessun impatto ai fini della stima degli aggregati di Contabilità Nazionale.

Riassumendo:

- ✓ l'ipotesi su cui si basa la soluzione proposta dal GdL è che la stima della popolazione sia coerente con la stima degli occupati regolari proveniente dai registri statistici. Con questa ipotesi il GdL si è proposto di anticipare quello che appare come il naturale sviluppo del forte investimento compiuto dall'istituto nell'utilizzo e nell'integrazione delle fonti amministrative sugli individui e sulle altre unità statistiche;
- ✓ data questa ipotesi, per la stima dell'occupazione regolare vengono utilizzate le informazioni – formalmente esaustive – provenienti dai registri statistici, combinate con le informazioni campionarie deducibili dalla base dati integrata LFS-ADMIN;
- ✓ per la componente degli occupati regolari, in questa maniera viene adottato un criterio analogo

⁵ ISTAT. 2015. Ricostruzione statistica delle serie regionali di popolazione del periodo 1/1/2002-1/1/2014. Nota informativa - 14 gennaio 2015. Roma: ISTAT.

- per i sottoinsiemi A, B e C e viene sfruttata tutta l'informazione esaustiva dei registri;
 - ✓ per la stima della componente non regolare degli occupati di cui al sottoinsieme A si può solamente contare sul campione integrato LFS-ADMIN e sul sistema di riporto all'universo delle stime;
 - ✓ per le componenti B e C si devono usare e integrare le stime sulla popolazione con le stime sulla componente sommersa fornite dal campione.
- I vantaggi di questo approccio sono i seguenti:
- ✓ data l'ipotesi di partenza, le stime della popolazione influenzano direttamente la sola componente non regolare e hanno dunque un impatto relativamente contenuto. La distanza fra le stime ottenibili con i pesi censuari non corretti dalle stime PES e i pesi pre-censuari è intorno ai 100 mila occupati non regolari. Poiché si tratta di due soluzioni limite, appare ragionevole un posizionamento intermedio con un campo di variazione delle possibili stime piuttosto contenuto (prevedibilmente intorno alle 10-20 mila unità);
 - ✓ le stime delle componenti dell'occupazione regolare, dal momento che sono agganciate a quelle dei registri statistici, sono per costruzione coerenti oltre che con i registri anche con le fonti amministrative utilizzate per la stima dei conti economici delle imprese. Si tratta di una coerenza che ha rilievo da un punto di vista formale e che dovrà in una fase successiva tenere conto del fatto che le fonti ADMIN producono in realtà una sovrastima dell'input di lavoro effettivo la cui entità ha potuto essere misurata e corretta attraverso il campione integrato LFS-ADMIN;
 - ✓ questa soluzione getta un ponte fra stime della popolazione, registri statistici e fonti ADMIN sull'occupazione, inserendosi in una strategia sulla quale l'istituto ha cominciato a investire, e che appare coerente con la visione strategica della statistica pubblica promossa da Eurostat.
 - ✓ i recenti risultati della rilevazione PES indicano come la stima della sottocopertura censuaria porta a una correzione delle stime della popolazione residente pienamente compatibile con le posizioni lavorative misurate dai registri statistici (tenuto ovviamente conto delle diverse tipologie di popolazione di riferimento delle due fonti). Si tratta di un risultato intermedio rispetto a quello misurato attraverso i sistemi di ponderazione CENS e DEM.

Le stime degli occupati ottenute con la nuova procedura sono relative allo stesso campo di osservazione di LFS e riguardano perciò la sola popolazione residente al netto delle convivenze (il sottoinsieme A definito in precedenza). Formalmente la definizione SEC2010 di occupato coincide con quella adottata in LFS, con una sola eccezione: gli occupati dipendenti in condizione non lavorativa temporanea (il caso dei cassintegrati). A questi individui LFS può attribuire la condizione di occupato o non occupato sulla base di un criterio basato sulla durata dell'assenza dal lavoro e sulla percentuale del reddito percepito. Secondo il SEC2010, invece, la totalità di questi soggetti è da includersi fra gli occupati ma al tempo stesso ad essi non vengono fatte corrispondere posizioni lavorative.

L'utilizzo integrato di LFS e ADMIN permette la correzione di alcune incoerenze fra le fonti. In particolare:

- ✓ è stato possibile individuare e proporre una correzione della sottocopertura della condizione di occupato che caratterizza LFS. Se da un lato per questa fonte si ha una perfetta corrispondenza della definizione della condizione di occupato con quella definita nel SEC, dall'altro è stato possibile stimare la componente di occupazione che gli intervistati a non rivelano all'indagine. Le motivazioni di questo comportamento vanno evidentemente approfondite e possono essere verosimilmente ricondotte sotto un ampio ventaglio di cause, non tutte dipendenti dalla volontà dell'intervistato;
- ✓ è stato inoltre possibile individuare e correggere la distorsione indotta dalla sovracopertura della condizione di occupazione rilevata dalle fonti ADMIN. In questo caso il problema è essenzialmente legato alla mancata corrispondenza della posizione lavorativa tracciata in ADMIN con il quadro regolamentare e con le definizioni di occupazione sottostanti il SEC.

Anche per le posizioni lavorative la corrispondenza con la nozione SEC appare elevata: nella

nuova procedura di stima la singola posizione lavorativa viene fatta corrispondere alla singola relazione fra individuo e datore di lavoro secondo la regola “*a ogni coppia distinta di codici fiscali lavoratore-impresa corrisponde una posizione lavorativa*”. La natura della singola posizione lavorativa, se sia cioè dipendente o indipendente, come pure la gerarchia delle posizioni lavorative vengono stimate combinando le informazioni ADMIN con quelle LFS. Le caratteristiche della posizione principale vengono poi assegnate al singolo occupato.

L’eventualità di reperire eventualmente tracce di una singola coppia di codici fiscali su più fonti amministrative distinte viene in sostanza interpretata come una rappresentazione plurima di una medesima relazione, a meno che le fonti non si caratterizzino per una diversa tipologia dell’occupazione sottostante (dipendente o indipendente)⁶.

Occorre inoltre richiamare il fatto che, coerentemente con il SEC, la condizione di cassaintegrazione non dà luogo a una posizione lavorativa.

Con riferimento alle ore effettivamente lavorate, la grandezza stimata con la nuova metodologia corrisponde alla definizione SEC. Com’è noto, le uniche fonti che forniscono informazioni dirette su questa variabile sono le rilevazioni campionarie sulle imprese e sulle famiglie. La disponibilità tramite le fonti amministrative di variabili ausiliarie correlate con le ore effettivamente lavorate, in particolare con le ore retribuite, consente peraltro di correggere alcune incoerenze presenti nelle fonti di indagine e di dare un significato alla differenza fra le misure rilevate dal lato delle imprese e quelle che emergono dal lato delle famiglie.

Il GdL ha inoltre sviluppato un prototipo di stima per piccole aree per consentire il raffronto sia fra le ore effettivamente lavorate stimate con informazioni raccolte dal lato impresa e le ore retribuite sia fra queste due grandezze e le ore effettivamente lavorate stimate dal lato famiglia (sez. 4 par. 10). Questa procedura porta alla stima della componente non regolare dell’input di lavoro riconducibile alle posizioni lavorative regolari – il cosiddetto “grigio”⁷ – e costituirà l’innesco per ulteriori approfondimenti legati alla stima del corrispondente circuito dei redditi.

Per quanto attiene alle stime riguardanti l’economia non osservata (NOE), l’occupazione in attività economiche informali o sommerse (pur non illegali) viene in termini di principio colta all’interno delle definizioni adottate in queste stime. Più ambigua e controversa è la questione se si stia includendo o meno nelle stime una parte degli occupati in attività illegali. Di fatto, è come se si fosse ipotizzato che costoro non rivelano mai la loro attività lavorativa illegale se intervistati nell’ambito di LFS. È importante notare, inoltre, che le due categorie di riferimento utilizzate nel SEC, le attività informali e quelle sommerse in senso stretto, non sono distinguibili sulla base delle informazioni LFS.

Dal punto di vista logico, si parte dall’identificazione delle posizioni regolari come quelle posizioni lavorative cui corrisponde un segnale amministrativo valido: indichiamo perciò con questo termine le posizioni lavorative osservabili nelle fonti amministrative. Viceversa, una posizione non regolare è una posizione non tracciata in ADMIN. Alle posizioni regolari possono corrispondere tuttavia prestazioni lavorative in parte non osservabili, nel senso che l’effettivo input di lavoro (misurato con le ore effettivamente lavorate) pur retribuito, potrebbe essere solo parzialmente regolare, mentre un’altra parte, non tracciabile, potrebbe ricadere in attività economiche sommerse.

Come per la tipologia occupazionale, la natura regolare o non regolare attribuita a un occupato dipende dalle caratteristiche della sua posizione lavorativa principale. Si possono così avere occupati regolari con posizioni secondarie non regolari, o con posizioni tutte regolari ma con una parte delle ore lavorate non regolari; oppure anche occupati con posizione principale non regolare ma con almeno una posizione secondaria regolare.

⁶ Solo in questo caso la coppia di codici fiscali può dare luogo a due posizioni lavorative distinte.

⁷ Nel rapporto finale della c.d. Commissione Giovannini (MEF 2011) viene rilevato che “*rimangono escluse dalla stima (dell’economia non osservata, ndr) tutte le diverse forme d’irregolarità parziale (il cosiddetto lavoro grigio), in genere connesse al ridotto pagamento dei contributi, alla pratica della retribuzione fuori busta, all’utilizzo irregolare di contratti di prestazione d’opera.*” (pag.10). Inoltre “*l’attuale metodo di stima potrebbe essere esteso con la collaborazione del Ministero del Lavoro e degli enti da questo vigilati, per considerare il “lavoro grigio”, che ha un impatto diretto sull’evasione contributiva*” pag.149.

Sono state inoltre poste, come già anticipato, le condizioni per produrre una stima per dominio della componente di ore non regolari effettivamente lavorate nell'ambito di posizioni regolari.

1.2. Le basi informative

1.2.1. I codici fiscali degli individui nel campione LFS

La base informativa integrata annuale utilizzata per le stime è stata ottenuta integrando i microdati della rilevazione LFS con quelli relativi alle fonti amministrative sull'occupazione. Prendendo in considerazione la popolazione con almeno 15 anni, il campione annuo LFS consta di circa 570 mila interviste: nell'85% circa dei casi si tratta di interviste dirette. Data la particolare natura longitudinale del campione LFS⁸, a questo campione corrispondono circa 320 mila individui distinti i cui codici fiscali sono stati ricostruiti affiancando due metodi diversi:

- ✓ da un lato ricorrendo al registro delle persone fisiche curato dall'ISTAT⁹;
- ✓ dall'altro attraverso l'aggancio dell'intestatario del foglio di famiglia con le liste anagrafiche comunali (LAC).

Sono stati così ricostruiti i codici fiscali del 95% del campione LFS: i casi di mancata ricostruzione sono dovuti da un lato alla impossibilità di ricostruire correttamente il codice attraverso i dati anagrafici degli intervistati (doppi nomi, donne coniugate, stranieri), dall'altro alla sfasatura temporale delle LAC. Dall'anno di riferimento 2012 tuttavia l'estrazione del campione LFS avviene a partire dalle LAC e ciascun individuo estratto sarà pertanto corredato di codice fiscale.

Data la natura non casuale dei casi di mancata ricostruzione dei codici fiscali, questi sono stati risolti imputando a tali individui il codice fiscale di un donatore¹⁰. In questo modo è stato attribuito un codice fiscale al 100% degli individui del campione: ciò ha reso possibile l'utilizzo dei medesimi coefficienti di riporto all'universo adottati nell'indagine senza doverli pertanto stimare nuovamente.

Tavola 1. Osservazioni LFS e ricostruzione dei codici fiscali - Anni 2010 e 2011

| Indicatori | 2010 | 2011 | var. % |
|---------------------------------|---------|---------|--------|
| Con 15 anni o più | 573,984 | 570,490 | -0.6 |
| Codici Fiscali distinti | 321,477 | 319,703 | -0.6 |
| Interviste per CF | 1.785 | 1.784 | -0.1 |
| Con Codice Fiscale ricostruito | 541,621 | 539,213 | -0.4 |
| Con Codice Fiscale imputato | 32,363 | 31,277 | -3.4 |
| % | 5.6 | 5.5 | |
| Osservazioni per Codice Fiscale | 1.685 | 1.687 | 0.1 |

Si indichi con N l'ammontare della popolazione in età di lavoro nel campo di osservazione LFS. A ciascun individuo-trimestre¹¹ i nel campione LFS (con $i=1, \dots, n$) è associato un peso campionario

⁸ Ciascuna famiglia intervistata viene contattata quattro volte: per due trimestri consecutivi e per gli stessi due trimestri dell'anno successivo, dopo di che esce definitivamente dal campione: ne segue che in un anno solare un individuo può apparire nel campione al massimo due volte, e una volta sola se la prima intervista è nel quarto trimestre o se la quarta intervista è nel primo trimestre.

⁹ Si ringraziano in particolare Giuseppe Garofalo e Carla Runci.

¹⁰ La scelta del donatore è stata effettuata dal DISA tramite la procedura RIDA. Come variabili di stratificazione sono state usate: il trimestre, la condizione professionale, il genere, la cittadinanza e l'età. Per il matching sono state usate una ventina di variabili, rappresentative di territorio, famiglia, redditi, istruzione, posizione, anzianità, attività economica, motivo non lavoro, altro.

¹¹ Il singolo record LFS è identificato da una chiave individuale e dal trimestre (poiché ogni chiave individuale può apparire in due trimestri). Per semplicità nel prosieguo, quando non vi è rischio di ambiguità, verrà usato il termine "individuo" per indicare il singolo record.

w_i tale che $\sum_i w_i = N$. L'indice i è in sostanza la chiave di aggancio con questo archivio. Ad essa corrisponde un vettore R_i di variabili raccolte attraverso la rilevazione, fra cui la chiave individuale (CHI), il trimestre ($TRIM$), lo status occupazionale (OCC), il peso campionario w_i e il codice fiscale (CF) dell'intervistato. Si avrà in particolare:

$$R_i = \{CHI_i, CF_i, TRIM_i, w_i, OCC_i^R, \dots\} \quad <1>$$

$1 \times m(r)$

dove $m(r)$ è il numero di variabili rilevate da LFS. Indichiamo infine con $R_{n \times m(r)}$ l'intera matrice dei dati raccolti tramite LFS.

1.2.2. Le fonti amministrative sugli occupati

I codici fiscali degli individui intervistati tramite LFS sono stati cercati presso $K = 9$ basi dati amministrative riguardanti gli occupati. Si tratta in particolare di cinque fonti tipiche del lavoro dipendente:

- ✓ Emens (EMEN¹²), che raccoglie i dipendenti delle imprese dell'industria e dei servizi, informazioni dettagliate sulle singole settimane dell'anno e con altre informazioni di rilievo riguardanti gli imponibili, l'orario, il contratto, ecc.;
- ✓ Inpdap (INPD) che raccoglie i dipendenti delle Amministrazioni pubbliche;
- ✓ Enpals (ENPA), che raccoglie i dipendenti delle imprese attive nei settori dell'intrattenimento;
- ✓ Colf e Badanti (DOME) che raccoglie badanti e collaboratori domestici;
- ✓ DMAG, che raccoglie i dipendenti delle imprese agricole.

Le restanti quattro fonti riguardano le posizioni lavorative indipendenti:

- ✓ SILO_I (INDI), l'archivio dell'ISTAT sugli indipendenti dell'industria e dei servizi, che a sua volta sintetizza informazioni contenute principalmente negli archivi dell'INPS, delle Camere di commercio e dell'Agenzia delle entrate;
- ✓ Autonomi agricoli (AUAG);
- ✓ Gestione separata INPS sui parasubordinati (PARA);
- ✓ Gestione separata INPS sui collaboratori professionali (PROF)

A queste fonti si aggiungono le informazioni INAIL sulle coperture assicurative dei lavoratori. In particolare:

- ✓ Archivio DNA (INAD), relativo ai lavoratori dipendenti;
- ✓ Archivio parasubordinati (INAP);
- ✓ Archivio interinali (INAI).

Da ogni archivio sono state estratte le informazioni eventualmente presenti relative a ciascun individuo nel campione LFS. Sono state pertanto raccolte informazioni su:

- ✓ Validità della copertura della fonte nella settimana di riferimento dell'intervista LFS, basata sulla coerenza delle date riportate nella fonte;
- ✓ Numero di giorni di copertura contrattuale nella settimana di riferimento, nel trimestre

¹² D'ora in avanti, per semplicità, le fonti amministrative vengono indicate con nomi di quattro lettere.

- di riferimento e nell'anno;
- ✓ Analoghe informazioni sul reddito imponibile;
- ✓ Il codice fiscale del committente;
- ✓ Altre informazioni sulla tipologia del rapporto di lavoro (contratto, durata, orario e altro).

Non tutte le fonti forniscono queste informazioni né le forniscono con la stessa precisione e affidabilità, in particolare con riferimento alle variabili che recano le informazioni relative alla settimana di riferimento dell'intervista. Per alcune fonti, come EMEN, la fonte stessa arriva a descrivere le singole settimane dell'anno (con la sfasatura di un giorno rispetto a LFS) e dunque il livello di precisione è assai elevato. Per altre (es. PARA), la fonte specifica una data di inizio e una di fine del contratto, una informazione comunque importante anche se meno precisa rispetto a EMENS. In altre fonti ancora (es. PROF o AUAG) il segnale è annuale (copertura nell'anno) e dunque è estremamente impreciso.

Tavola 2. Riepilogo delle basi dati amministrative

| Fonte | Archivio | Sigla | Campo | Date copertura | Precisione | Altre info | Importi |
|--------|-------------------|-------|--|---------------------------------------|--------------|-----------------------------------|---------|
| ISTAT | Silo indipendenti | INDI | Indipendenti imprese industria e servizi | Anno e stato di attività dell'impresa | Debole | Archivi complementari | |
| INPS | Emens | EMEN | Dipendenti imprese industria e servizi | Settimana INPS | Forte | Ampie e dettagliate | si |
| | Inpdap | INPD | Dipendenti PA | Periodo da...a | Media | Scarse | |
| | Gestione separata | PARA | Parasubordinati | Periodo da...a | Media | Ampie sul tipo di rapporto | si |
| | Autonomi agricoli | AUAG | Coltivatori diretti, imprenditori agricoli | Anno | Debole | Scarse | |
| | Dmag | DMAG | Dipendenti agricoli | Mese | Media-debole | Scarse | si |
| | Domestici | DOME | Colf e badanti | Periodo da...a e settimana retribuita | Forte | Tipo di rapporto e ore retribuite | si |
| | Gestione separata | PROF | Collab. profess. | Anno e stato di attività dell'impresa | Debole | Scarse | si |
| ENPALS | Enpals | ENPA | Lavoratori dello spettacolo | Periodo da...a | Media | Tipo di rapporto | si |
| INAIL | DNA | INAD | Dipendenti | Periodo da...a | Media | Scarse | |
| | Parasubordinati | INAP | Parasubordinati | Periodo da...a | Media | Scarse | |
| | Interinali | INAI | Interinali | Periodo da...a | Media | Scarse | |

Per ogni fonte k si avrà dunque un generico record identificato dalla coppia ij , dove j identifica il datore di lavoro.

Le informazioni relative ai datori di lavoro, identificati tramite codice fiscale nelle fonti ADMIN, sono state reperite attraverso l'aggancio con i registri statistici dell'ISTAT: ASIA, ASIA Agricoltura, S13, ISP, ecc. Queste informazioni riguardano in particolare:

- ✓ Il codice di attività economica;
- ✓ La classe dimensionale;
- ✓ Il settore istituzionale;
- ✓ La forma giuridica.

Per il j -esimo datore di lavoro dell'individuo i è dunque disponibile un vettore B_{ij}^k che include queste e altre informazioni specifiche del registro di provenienza.

Le informazioni complessivamente derivabili dalla fonte k con riferimento alla singola posizione ij sono sintetizzabili in un singolo vettore come segue:

$$A_{ij}^k = (C_{ij}^k, Y_{ij}^k, T_{ij}^k, B_{ij}^k) \quad \langle 2 \rangle$$

dove C è una variabile dicotomica che assume valore pari a '1' quando le date di copertura del contratto con il datore di lavoro j e presenti nella fonte k (con $k=1, \dots, K$) sono compatibili con la settimana dell'intervista LFS dell'individuo i ; Y identifica genericamente tutte le variabili della fonte amministrativa inerenti agli imponibili e/o alle giornate retribuite; T identifica invece tutte le variabili descrittive della tipologia dell'occupazione.

A ogni individuo i è pertanto associata in corrispondenza di ciascuna fonte k una matrice A_i^k con un numero di righe pari al numero di datori di lavoro distinti con cui ha intrattenuto rapporti di lavoro registrati dalla fonte. Detto $m(k)$ il numero di variabili della fonte k e J_i^k il numero di posizioni assunte da i nella fonte, la matrice A_i^k avrà dimensioni $J_i^k \times m(k)$.

Considerando tutto l'insieme delle interviste LFS si ottiene la matrice A^k - di dimensioni $\left(\sum_i J_i^k\right) m(k)$ - che costituisce la matrice delle posizioni lavorative degli intervistati LFS presso la fonte k . Questa matrice è agganciabile "molti a uno" con l'archivio R (cfr. formula $\langle 1 \rangle$) usando i come chiave di aggancio.

Dopo avere ordinato le posizioni di ciascun individuo i per valori decrescenti di C e Y , per ogni i possiamo trattare l'indice j come un ordinamento decrescente dell'importanza della singola posizione nella fonte. A partire dalla matrice A_i^k si costruisce pertanto il vettore a_i^k che riporta le informazioni relative alla posizione principale dell'individuo i nella fonte e al riepilogo delle informazioni relative alle eventuali posizioni secondarie:

$$a_i^k = (A_{i1}^k; \sum_{j>1} C_{ij}^k, \sum_{j>1} Y_{ij}^k) \quad \langle 3 \rangle$$

Da questo vettore, considerando tutti gli intervistati LFS, si giunge alla matrice a^k (di dimensioni $n \times m(k)$) agganciabile 1 a 1 con l'archivio R usando i come chiave di aggancio. Questa matrice costituisce l'"archivio dei record", ed è interpretabile a sua volta come una estensione del questionario dell'indagine ottenuta aggiungendo a LFS le informazioni ADMIN.

1.3. Il modello per la stima annuale dell'input di lavoro: una visione d'insieme

1.3.1. Premessa

Il modello per la stima dell'input di lavoro ha l'obiettivo di stimare, con un adeguato livello di dettaglio, l'ammontare di tre variabili riferite al campo di osservazione dei conti nazionali: le posizioni lavorative (P), gli occupati (L) e le ore lavorate (H)¹³. Per comodità espositiva e metodologica

¹³ Riferimento ai regolamenti. È importante sottolineare che nel caso degli occupati e delle posizioni lavorative queste variabili esprimono una media annuale calcolata su base settimanale: un occupato equivale dunque a 52 settimane lavorate indipendentemente dal numero di posizioni

si esplicitano due ulteriori dimensioni: la residenza degli occupati e la regolarità della loro prestazione lavorativa.

Per descrivere la residenza si utilizza un indicatore dicotomico (residente o non residente in Italia); in un secondo momento sarà utile introdurre ulteriori modalità per tenere conto della cittadinanza e della natura - regolare o non regolare - della presenza sul territorio.

Si utilizza una variabile dicotomica anche per descrivere la natura regolare o non regolare dell'input di lavoro. Questa scelta può essere efficacemente messa in discussione sotto il profilo semantico e teorico. La regolarità infatti viene qui "certificata" dai registri statistici e amministrativi sull'occupazione, ed è pertanto da intendersi più precisamente come esito dell'associazione coerente dell'input di lavoro con gli adempimenti formali previsti dalle normative che disciplinano il mercato del lavoro: una prestazione lavorativa è regolare se tracciata nei record amministrativi¹⁴. In realtà, la cresciuta (e crescente) quantità e qualità di microdati integrati sull'occupazione rende possibile una articolazione dei profili legati alla regolarità (o alla non regolarità): questa circostanza rende opportuno - in fase di interpretazione ma ancor prima in fase di impostazione metodologica - considerare la regolarità come una variabile continua o eventualmente discreta ma articolata in più modalità. Infine, diversamente dalla residenza - che etichetta univocamente tutte le prestazioni del singolo individuo - la regolarità ha invece caratteristiche di trasversalità: infatti, ore non regolari possono essere associate a prestazioni lavorative regolari, e occupati regolari possono assumere anche posizioni lavorative secondarie non regolari¹⁵.

Si indichi con $r=1,2$ la caratteristica rispettivamente regolare o non regolare di ciascuna misura del lavoro e con $s=1,2$ quella rispettivamente di residente o non residente di ciascun occupato. La Figura 2 illustra il campo di osservazione degli obiettivi della stima¹⁶. In particolare, data una partizione J del campo di osservazione¹⁷ e prendendo come riferimento le posizioni lavorative (ma la stessa relazione vale per occupati e ore lavorate), l'obiettivo della stima sarà dato da:

$$P = \sum_{r,s,j \in J} P_{rsj} = \sum_{j \in J} (P_{11j} + P_{12j} + P_{21j} + P_{22j}) \quad <4>$$

dove le componenti residenti ($P_{11j} + P_{21j}$) rappresentano il 99% circa delle posizioni regolari e una quota comunque largamente maggioritaria (poco al di sotto del 90%) delle posizioni non regolari.

e dal numero di ore lavorate; una posizione è una relazione lavoratore-datore di 52 settimane indipendentemente dal numero di ore lavorate. Le ore lavorate, invece, costituiscono un ammontare complessivo su base annua. Si tornerà più oltre su questo punto.

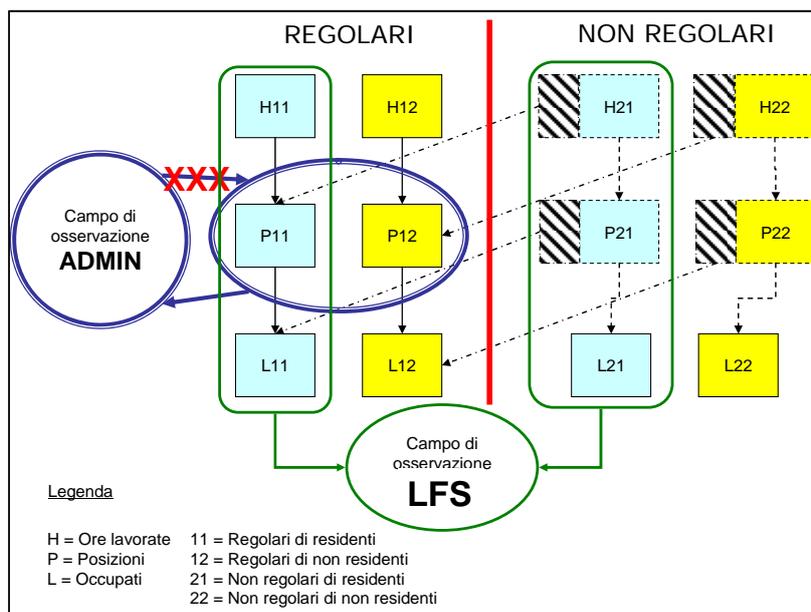
¹⁴ Se una attività lavorativa non dovesse prevedere adempimenti formali ricadrebbe inevitabilmente - in questa accezione - fra le "non regolari" (cioè non tracciata) senza per questo violare alcuna norma amministrativa.

¹⁵ Cfr. la Figura 1. Un occupato regolare è tale in quanto la sua posizione lavorativa principale è una posizione regolare. Nel prosieguo dell'esposizione, per semplicità espositiva, si identificherà l'occupato regolare come un occupato con almeno una posizione regolare.

¹⁶ Nella figura 2 le aree contrassegnate dalle righe orizzontali corrispondono a componenti non regolari di posizioni e ore associate rispettivamente a occupati e posizioni regolari. In particolare, nel caso delle ore si tratta della componente grigia, che ha senso misurare solo con riferimento alle posizioni regolari dipendenti.

¹⁷ Ottenuta per esempio sulla base di una combinazione di criteri riferibili alla natura della posizione, al territorio, al settore istituzionale, alla forma giuridica, all'attività economica, alla classe dimensionale dei datori di lavoro, ecc..

Figura 2. Input di lavoro e fonti amministrative



Per questa stima vengono utilizzate e integrate due tipologie di fonti: l’insieme degli archivi amministrativi sull’occupazione, che per brevità sono indicati con ADMIN, e che includono esaurientemente le posizioni lavorative regolari degli occupati residenti e non residenti¹⁸; i microdati LFS i quali forniscono una rappresentazione campionaria – coerente con il quadro regolamentare - degli occupati residenti, regolari e non regolari.

1.3.2. La stima delle posizioni lavorative

La fonte ADMIN descrive un insieme di relazioni individuo-datore che contiene i sottoinsiemi P_{1sj} : ogni posizione regolare è infatti inclusa per definizione in ADMIN. Non è invece necessariamente valido il contrario: non tutte le relazioni ADMIN identificano posizioni regolari¹⁹. Introduciamo pertanto il termine “posizioni lavorative potenziali” (indicate con Π) per identificare le relazioni contenute in ADMIN. Le “posizioni potenziali” sono dunque l’universo delle relazioni di lavoro tracciate negli archivi amministrativi ed espresse in media annua. Per essere considerate posizioni lavorative regolari la potenza e la coerenza del segnale amministrativo di ciascuna posizione potenziale deve essere validata: il campione integrato LFS-ADMIN assolve – fra l’altro - a questo compito fornendo la possibilità di stimare l’effettiva consistenza delle posizioni regolari nel novero di quelle potenziali.

Dal momento che le fonti amministrative consentono di individuare le posizioni potenziali degli individui residenti (Π_1) e non residenti (Π_2)²⁰, si avrà in ogni dominio j :

$$\Pi_{sj} \geq P_{1sj} \quad <5>$$

¹⁸ ADMIN è costituito dall’integrazione delle fonti sull’occupazione rilevata dal lato dei datori di lavoro. Sono fonti di tipo statistico (censimenti, registri, indagini) e amministrativo (archivi sui versamenti contributivi, sulle partite IVA, sulle persone d’impresa e sui soci delle società), che descrivono le singole relazioni fra individui e datori di lavoro e forniscono informazioni sui datori di lavoro quali il codice di attività economica, la classe dimensionale, la forma giuridica e la localizzazione territoriale (fino alla provincia) consentendo anche la costruzione dei raggruppamenti per settore istituzionale.

¹⁹ Come si vedrà più oltre, questo è uno dei risultati più importanti ottenuti con la base dati integrata LFS-ADMIN

²⁰ La possibilità di isolare le componenti residenti e non residenti è già alla portata delle basi informative presenti in istituto ed è ottenibile attraverso l’aggancio dei microdati ADMIN con le liste anagrafiche comunali (LAC). Nel prosieguo si assumerà tuttavia che sia acquisito l’aggancio “esatto” dei microdati ADMIN con le LAC.

I microdati della fonte campionaria LFS sono stati agganciati attraverso il codice fiscale con la fonte ADMIN (LFS-ADMIN): si è trattato di un'imputazione nel questionario LFS di nuove variabili dedotte da ADMIN e riferite alle informazioni sulle posizioni potenziali. L'analisi congiunta dei microdati di indagine e amministrativi ha inoltre portato alla correzione di alcune distorsioni presenti nella fonte campionaria e in ADMIN. In particolare, attraverso i coefficienti di riporto all'universo, sono state stimate per dominio j le seguenti variabili:

$$\begin{aligned} & \hat{L}_{11j}, \hat{L}_{21j} \text{ (gli occupati regolari e non regolari residenti)} \\ & \hat{\Pi}_{1j} \text{ (le posizioni regolari potenziali degli occupati residenti)}^{21)} \\ & \hat{P}_{11j} \text{ (le posizioni lavorative regolari degli occupati regolari residenti)}^{22)} \\ & \hat{P}_{21j} \text{ (le posizioni lavorative non regolari degli occupati residenti)} \quad <6> \\ & \hat{H}_{11j}, \hat{H}_{21j} \text{ (le ore lavorate regolari e non regolari degli occupati residenti)} \end{aligned}$$

La stima del totale delle posizioni regolari potenziali dei residenti ottenuta con il campione integrato LFS-ADMIN ($\hat{\Pi}_{1j}$) è direttamente confrontabile con l'analogo totale della base dati ADMIN (Π_1):

$$\Pi_1 = \sum_j \Pi_{1j} \cong \hat{\Pi}_1 = \sum_j \hat{\Pi}_{1j} \quad <7>$$

dove l'errore campionario nella stima dei livelli, data la numerosità campionaria piuttosto elevata, è pressoché trascurabile^{23 24}. Ciononostante è evidente che l'adozione di una partizione J molto fine²⁵ si accompagna a errori campionari più marcati sul singolo dominio. È pertanto ragionevole sfruttare appieno le informazioni esaustive disponibili sulle posizioni potenziali onde pervenire a una stima più accurata e robusta delle posizioni lavorative regolari. In particolare, attraverso il campione LFS-ADMIN si stima per dominio il coefficiente λ_j che esprime la quota di posizioni lavorative regolari sul totale delle posizioni potenziali²⁶:

$$\lambda_j = f(\hat{P}_{11j}, \hat{\Pi}_{1j}) \quad <8>$$

Il coefficiente stimato è un indicatore della misura della sovrastima delle fonti amministrative e può essere utilizzato quindi per correggere le posizioni potenziali di fonte ADMIN²⁷:

$$\hat{P}_{11j} = \hat{\lambda}_j \Pi_{1j} \quad <9>$$

²¹ L'indice r relativo alla regolarità della posizione viene escluso perché ridondante.

²² Le stime sulle posizioni per dominio j attraverso LFS-ADMIN sono inclusive anche della stima dell'ordine delle posizioni e in particolare del numero delle posizioni principali e di quelle secondarie.

²³ Il peso medio del singolo individuo del campione annuo LFS è vicino a 100. La bontà dell'accostamento dipende semmai dalla corretta specificazione della popolazione di riferimento di LFS. Si assume nel prosieguo che la popolazione di riferimento sia quella "giusta".

²⁴ La tenuta del confronto a livelli più disaggregati appare del resto soddisfacente: il campione LFS-ADMIN restituisce infatti, a parità di campo di osservazione, una struttura dell'occupazione dipendente praticamente identica a quella del registro ASIA. Il registro ASIA è quello costruito con i criteri più vicini e compatibili con gli standard regolamentari cui fa riferimento LFS e le stime di CN, soprattutto con riferimento all'occupazione dipendente.

²⁵ Ad esempio, tre digit Nace, sette classi di addetti e regione.

²⁶ Per la stima del coefficiente λ si è adottato uno stimatore diretto con risultati soddisfacenti: un'alternativa possibile e più sofisticata è uno stimatore *area-level* sintetico o per piccole aree.

²⁷ Nel testo si utilizza la convenzione di indicare con l'hat ^ le stime ottenute tramite il campione integrato LFS-ADMIN e con il doppio hat le stime ottenute utilizzando le informazioni ausiliarie.

Si ottiene così una redistribuzione delle posizioni regolari \hat{P}_{11j} stimate tramite il campione integrato correggendo la parte di errore campionario che riguarda le posizioni potenziali. Date le condizioni di partenza, e in assenza di significativi effetti di composizione assicurata dall'ampiezza del campione di indagine, si ha infatti:

$$\hat{P}_{11} \cong \hat{\hat{P}}_{11} \quad <10>$$

Con riferimento alle posizioni afferenti agli occupati non residenti occorre distinguere il caso delle posizioni regolari da quelle non regolari.

Per le posizioni regolari dei non residenti la stima è analoga a quella adottata per i residenti. Tramite LFS-ADMIN si stima il coefficiente λ_{2j} e si deriva²⁸:

$$\hat{P}_{12j} = \hat{\lambda}_{2j} \Pi_{2j} \quad <11>$$

Quanto alle posizioni non regolari dei non residenti occorre distinguere ulteriormente due insiemi: le posizioni afferenti ai non residenti regolarmente presenti sul territorio (A) e quelle relative alle presenze non regolari (B). La componente A viene stimata attraverso un rapporto di coesistenza δ_{2j}^A fra posizioni regolari e non regolari ottenuto attraverso LFS-ADMIN²⁹:

$$\hat{P}_{22j}^A = \delta_{2j}^A \hat{P}_{12j} \quad <12>$$

Infine, la componente (B) viene posta convenzionalmente pari al numero degli occupati stimato a parte sulla base delle informazioni disponibili:

$$\hat{P}_{22j}^B = \hat{L}_{12j}^B \quad <13>$$

Aggregando le stime indicate nelle espressioni <6> e <9> - <13> si ottiene la stima delle posizioni lavorative:

$$\hat{P} = \sum_{r,s,j \in J} \hat{P}_{rsj} = \sum_{j \in J} \left[\hat{P}_{11j} + \hat{P}_{12j} + \hat{P}_{21j} + \left(\hat{P}_{22j}^A + \hat{P}_{22j}^B \right) \right] \quad <14>$$

²⁸ Per la stima di λ si utilizzano le informazioni sulla cittadinanza dei residenti nel campione integrato.

²⁹ In sostanza si utilizza il campione LFS relativo agli stranieri residenti e si stima il coefficiente tenendo conto delle aree geografiche di provenienza.

1.3.3. La stima degli occupati

Le posizioni lavorative regolari e non regolari sono state derivate dall'utilizzo delle fonti ADMIN e della fonte campionaria integrata LFS-ADMIN e sono state in parte ottenute sulla base di una stima preliminare degli occupati: in particolare, attraverso quest'ultima si è preliminarmente stimato il numero degli occupati residenti ($\hat{L}_{11j}, \hat{L}_{21j}$).

Per il segmento degli occupati residenti, in seguito all'aggiustamento della stima LFS-ADMIN delle posizioni per strato ottenuto attraverso l'universo delle posizioni potenziali, occorre aggiustare conseguentemente il numero degli occupati per dominio j . L'ipotesi è che l'incidenza delle posizioni principali in ciascun dominio sia correttamente misurata da LFS-ADMIN. Dunque:

$$\hat{L}_{11j} = \left(\frac{\hat{L}_{11j}}{\hat{P}_{11j}} \right) \hat{P}_{11j} = \hat{\gamma}_{11j} \hat{P}_{11j} \quad <15>$$

con $\hat{L}_{11} \cong \hat{L}_{11}$. Per la porzione non residente, il numero di occupati regolari è ottenuto analogamente attraverso la stima di un coefficiente γ :

$$\hat{L}_{12j} = \hat{\gamma}_{12j} \hat{P}_{12j} \quad <16>$$

Per quanto riguarda gli occupati non regolari, la stima della componente non residente viene derivata dalla stima delle posizioni per la componente A (non residenti regolarmente presenti sul territorio):

$$\hat{L}_{22j}^A = \hat{\gamma}_{22j}^A \hat{P}_{22j}^A \quad <17>$$

Per la componente B (non residenti irregolarmente presenti) la stima costituisce un dato esogeno. La stima definitiva degli occupati è dunque data da:

$$\hat{L} = \sum_{r,s,j \in J} \hat{L}_{rsj} = \sum_{j \in J} \left[\hat{L}_{11j} + \hat{L}_{12j} + \hat{L}_{21j} + \left(\hat{L}_{22j}^A + \hat{L}_{22j}^B \right) \right] \quad <18>$$

1.3.4. La stima delle ore lavorate

Per la stima delle ore effettivamente lavorate viene utilizzata la base dati integrata LFS-ADMIN e in particolare le informazioni sulle ore lavorate nella posizione principale e nell'insieme delle posizioni secondarie da parte degli intervistati LFS in condizione di occupazione (sez.4).

Si parte dunque dalla stima per dominio del totale delle ore annuali lavorate dai residenti nelle due componenti regolari e non regolari, nelle prime (indicate con l'apice ') e nelle seconde posizio-

ni (doppio apice “) ($\hat{H}'_{11j}, \hat{H}'_{21j}, \hat{H}''_{11j}, \hat{H}''_{21j}$).

Per le posizioni regolari dei residenti, queste stime vanno rimodulate sulla base della successiva stima ottenuta adattando le posizioni potenziali LFS-ADMIN all’universo delle posizioni potenziali ADMIN. In particolare, le ore totali del dominio j vengono poste pari al pro capite delle posizioni stimato con LFS-ADMIN moltiplicato per una stima delle prime posizioni finali e tenuto conto dell’ordine delle posizioni:

$$\hat{H}_{11j} = \left[\left(\frac{\hat{H}'_{11j}}{\hat{P}'_{11j}} \right) \left(\frac{\hat{P}'_{11j}}{\hat{P}_{11j}} \right) + \left(\frac{\hat{H}''_{11j}}{\hat{P}''_{11j}} \right) \left(\frac{\hat{P}''_{11j}}{\hat{P}_{11j}} \right) \right] \hat{P}_{11j} \quad <19>$$

Per le posizioni regolari dei non residenti si procede all’imputazione di un pro capite orario opportunamente stimato per dominio tramite LFS-ADMIN:

$$\hat{H}_{12j} = \left[\hat{h}'_{12j} \hat{P}'_{12} + \hat{h}''_{12j} \hat{P}''_{12} \right] \quad <20>$$

Si procede analogamente per i non regolari irregolarmente presenti:

$$\hat{H}_{22j} = \hat{h}_{22j} \hat{P}_{22} \quad <21>$$

In definitiva:

$$\hat{H} = \sum_{r,s,j \in J} \hat{H}_{rsj} = \sum_{j \in J} \left[\hat{H}_{11j} + \hat{H}_{12j} + \hat{H}_{21j} + \hat{H}_{22j} \right] \quad <22>$$

La stima della componente “grigia” delle ore lavorate nelle posizioni dipendenti regolari viene infine basata sul confronto fra questa stima (limitatamente al dominio corrispondente) e la stima delle ore lavorate misurata dal lato delle statistiche sulle imprese (sez.4 par.10).

2. La stima degli occupati

2.0 Sintesi

Le singole fonti amministrative contengono segnali di copertura di primo livello, basati sulla compatibilità fra le date relative alla copertura previdenziale presenti in ciascuna fonte e la settimana di riferimento dell’intervista LFS.

L’eterogeneità dell’incidenza dei segnali di copertura dipende da almeno due aspetti concomitanti: da un lato, dalle caratteristiche dello specifico rapporto di lavoro descritto da ciascuna fonte e dalle eterogeneità dei vari segmenti del mercato del lavoro (dipendente, autonomo, atipico, agricolo, ecc.); dall’altro dalla precisione e accuratezza con cui la copertura della fonte può essere ricondotta alla settimana di riferimento dell’intervista LFS.

L'incidenza dei casi di incoerenza fra il segnale di copertura di primo livello delle fonti ADMIN e il segnale di occupazione LFS appare fortemente eterogenea tra le fonti. Sono all'opera effetti di sovrastima dell'occupazione da parte di ADMIN legati sia alla difficoltà di ricondurre il dato alla settimana di riferimento LFS sia alla non necessaria corrispondenza fra traccia ADMIN ed effettivo svolgimento dell'attività lavorativa. In maniera speculare alcune incoerenze sono invece riconducibili alla sottodichiarazione di svolgimento di attività lavorative degli intervistati LFS. Gli effetti delle interviste proxy nel determinare le incoerenze sono invece trascurabili.

Le fonti colgono dunque in maniera non omogenea i segnali reali di occupazione, appaiono "specializzate" su particolari segmenti del mercato del lavoro e questo insieme di elementi le rende poco confrontabili fra loro. Esse mostrano in particolare una tendenza a sovra determinare lo stato di occupazione in ragione di due tipi di effetti che interagiscono fra loro: quelli dovuti alla scarsa precisione sulla copertura temporale delle fonti; quelli dovuti alla natura del rapporto di lavoro: più questo è intrinsecamente discontinuo più viene indebolita la potenza stessa del segnale di copertura.

La stima degli occupati si basa su alcune ipotesi: il segnale di occupazione di LFS non è soggetto ad errore mentre il segnale di non occupazione di LFS può esserlo; l'assenza di coperture ADMIN di primo livello non è soggetta ad errore mentre una copertura di primo livello non è sufficiente a garantire la validità del segnale di occupazione.

La sovracopertura ADMIN è stata corretta attraverso un modello logistico (un modello per ciascuna fonte) che stima la probabilità di essere un occupato per LFS date le caratteristiche del segnale di copertura della fonte e le caratteristiche dell'individuo. Sono stati selezionati fra gli occupati di CN gli individui con copertura ADMIN e una probabilità predetta superiore a una soglia prefissata.

La sottocopertura dell'occupazione da parte di LFS è stata corretta assumendo da parte degli occupati non regolari la stessa propensione a occultare la propria condizione lavorativa registrata dagli occupati regolari. La selezione degli individui è avvenuta per strato sulla base di un modello logistico.

La mancata validazione dei segnali di copertura ha una maggiore incidenza presso le fonti caratterizzate da una maggiore incoerenza fra segnali ADMIN e segnali LFS. Spiccano i casi delle fonti sugli indipendenti e sui lavoratori agricoli. Nei due anni considerati l'incidenza delle correzioni del segnale ADMIN è pressoché identica.

L'incidenza delle correzioni è inoltre diversa a seconda della posizione lavorativa registrata in LFS: è molto bassa per gli occupati LFS, intermedia per disoccupati e inattivi "grigi", molto alta per gli altri inattivi. L'incidenza delle mancate validazioni è appena più alta presso gli inattivi grigi rispetto ai disoccupati.

Con questo metodo vengono stimati circa 1,2 milioni di occupati in più rispetto a LFS, molto simile peraltro alla differenza solitamente stimata nei conti nazionali: oltre un milione sono occupati con almeno una posizione regolare; altri 150 mila circa sono privi di posizioni regolari.

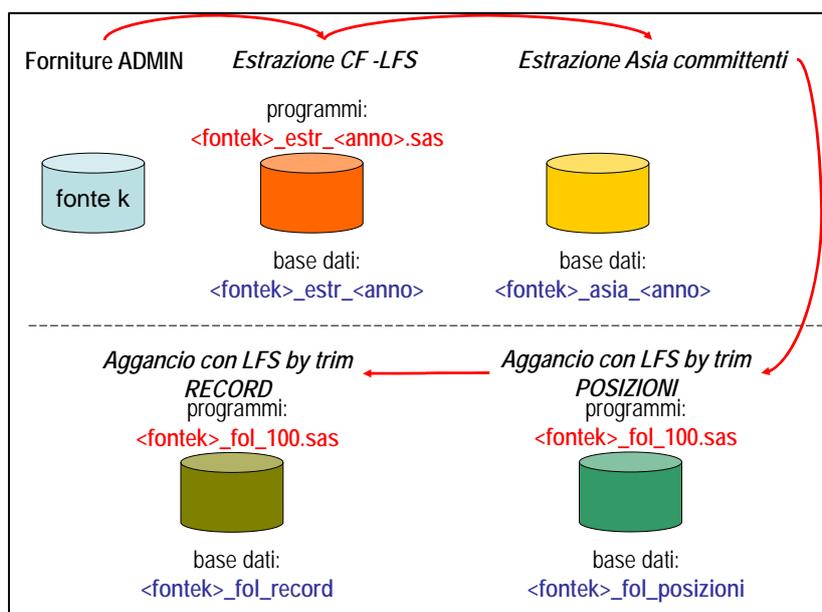
Nel complesso, gli occupati privi di posizioni regolari sono pari all'11% circa del totale.

2.1. Coerenza fra le fonti e scelte operative

2.1.1. Organizzazione dei dati

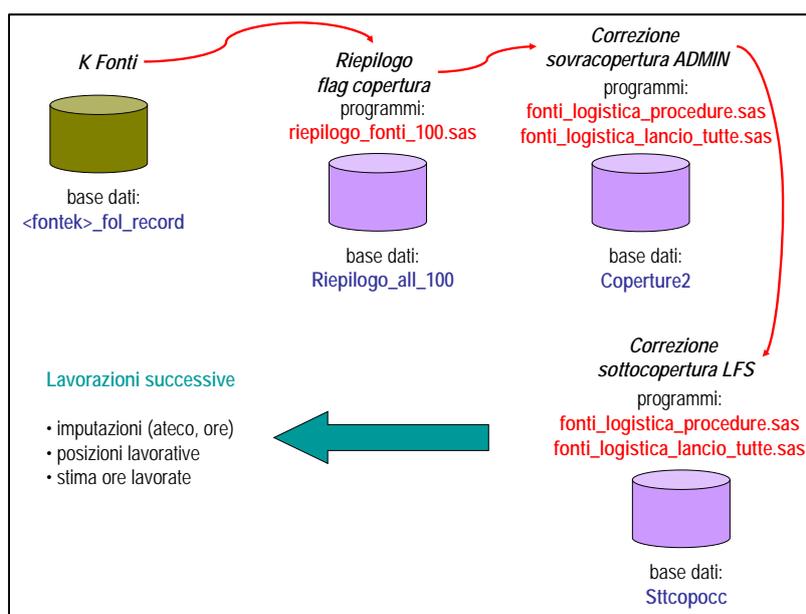
La Figura 3 sintetizza le modalità di organizzazione e trattamento dei dati di base. Dai microdati relativi alla fornitura di ciascuna fonte k sono stati estratti i record relativi ai codici fiscali degli individui intervistati tramite LFS con procedure SAS dedicate. I codici fiscali dei datori di lavoro associati a questi record individuali sono stati a loro volta agganciati con le informazioni contenute nei registri statistici sulle unità economiche. A seguire, i record della fonte k sono stati agganciati con i record individuali LFS dando luogo a un archivio delle posizioni lavorative regolari potenziali nella fonte k , identificando attraverso le informazioni sul trimestre e sulla settimana di riferimento dell'intervista LFS la effettiva copertura della fonte amministrativa. Infine, a partire dall'archivio delle posizioni lavorative è stato prodotto un archivio dei record della fonte k , nel quale in presenza di più posizioni per un medesimo individuo l'informazione sulle posizioni secondarie è collasata su più variabili in modo da garantire una corrispondenza uno a uno con i record LFS.

Figura 3. Estrazione e aggancio dei record ADMIN



Questa prima fase di lavorazione ha portato alla creazione di basi dati separate per ciascuna fonte relative alle posizioni lavorative e ai loro riepiloghi individuali per ciascun intervistato LFS. Nella fase successiva queste fonti sono state elaborate con l'obiettivo di pervenire a una stima degli occupati nel campo di osservazione LFS. A tal fine (cfr. la Figura 4), per ciascun individuo sono stati riepilogati tutti i segnali di copertura provenienti dalle varie fonti ADMIN in una base dati integrata che è stata chiamata RIEPILOGO_ALL_<anno>. A partire da questa base dati sono state condotte le stime sulla sovracopertura delle posizioni lavorative causata dalle fonti ADMIN (base dati COPERTURE2_<anno>) e della sottocopertura degli occupati causata da LFS (base dati SOTTOCOP_OCC_<anno>). Queste due basi dati, una volta associate a RIEPILOGO_ALL, costituiscono la base informativa sugli occupati, regolari e non regolari. Il paragrafo successivo entra nel merito delle procedure descritte nella Figura 4.

Figura 4. Riepilogo delle fonti e stima delle sovra e sottocoperture



2.1.2. Evidenze empiriche

Le K fonti amministrative che costituiscono la componente ADMIN forniscono in realtà segnali e informazioni che non sono immediatamente fra loro confrontabili e che necessitano pertanto di un trattamento statistico allo scopo di “pesare” opportunamente ciascuna fonte e rendere così possibile un uso congiunto del ventaglio dei segnali ADMIN disponibili. Alcune evidenze a sostegno di questa affermazione emergono chiaramente dall'analisi della base dati integrata LFS-ADMIN. Per comodità espositiva è opportuno definire con il termine “*copertura di primo livello*” di una fonte amministrativa k per l'individuo i (indicata con C_{il}^k) l'evento che segnala la presenza nella settimana di riferimento dell'intervista LFS di una traccia amministrativa in k valida in quella settimana, ossia una traccia le cui date di validità individuano un intervallo di tempo che include almeno parte della settimana di riferimento dell'intervista. In sintesi, le analisi condotte sui segnali di copertura C_{il}^k e sul segnale di occupazione di LFS rivelano quanto segue:

- ✓ Le fonti mostrano una diversa incidenza degli individui con coperture di primo livello ($C_{il}^k = 1$) (Tavola 3);
- ✓ Vi è una significativa presenza di segnali non coerenti provenienti da LFS e da ADMIN, e in particolare di segnali di copertura di primo livello ADMIN associati a individui che non risultano occupati secondo LFS (Tavola 4);
- ✓ Al tempo stesso le fonti evidenziano una notevole eterogeneità per quanto riguarda l'incidenza dei casi di incoerenza.

Tavola 3. Incidenza (a) degli individui con copertura di primo livello ($C_{il}^k = 1$) per fonte. Anni 2010 e 2011

| Fonti | 2010 | | | 2011 | | |
|-------------------|--------|---------|-------|--------|---------|-------|
| | Totale | Dirette | Proxy | Totale | Dirette | Proxy |
| Dirette | | | | | | |
| EMEN | 82.5 | 82.7 | 81.5 | 82.4 | 82.8 | 80.8 |
| INDI | 81.1 | 80.9 | 81.7 | 81.0 | 80.8 | 81.9 |
| INPD | 93.0 | 93.1 | 92.8 | 92.3 | 92.5 | 91.1 |
| PARA | 59.8 | 60.2 | 58.0 | 60.2 | 60.8 | 57.0 |
| PROF | 98.3 | 98.3 | 98.3 | 98.6 | 98.6 | 98.3 |
| ENPA | 51.9 | 52.9 | 48.3 | 52.5 | 53.3 | 49.3 |
| DMAG | 48.2 | 47.5 | 51.5 | 49.6 | 49.2 | 51.6 |
| AUAG | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| DOME | 79.3 | 79.7 | 76.6 | 75.8 | 76.4 | 72.7 |
| Ausiliarie | | | | | | |
| INAD | 87.0 | 87.3 | 85.6 | 86.8 | 87.2 | 84.4 |
| INAI | 57.6 | 58.2 | 55.0 | 57.5 | 58.0 | 55.0 |
| INAP | 33.3 | 33.3 | 33.7 | 34.9 | 34.3 | 37.5 |

Note: (a) Percentuale degli individui con copertura sul totale degli individui agganciati dalla fonte

Fonte: LFS-ADMIN, dati provvisori, sistema di ponderazione censuario

Tavola 4. Non occupati LFS con copertura di primo livello ($C_{il}^k = 1$) per fonte. Anni 2010 e 2011

| Fonti | 2010 | | | | | 2011 | | | | |
|-------------------|-------|------------|------------------|-------------|----------------|-------|------------|------------------|-------------|----------------|
| | % (b) | % proxy(c) | Composizione (d) | | | % (b) | % proxy(c) | Composizione (d) | | |
| | | | Disocc. | Zona grigia | Altri inattivi | | | Disocc. | Zona grigia | Altri inattivi |
| <i>Dirette</i> | | | | | | | | | | |
| EMEN | 4.6 | 5.8 | 20.2 | 31.4 | 48.4 | 4.4 | 6.0 | 21.2 | 33.3 | 45.6 |
| INDI | 25.6 | 23.1 | 11.3 | 14.4 | 74.3 | 26.7 | 24.8 | 11.7 | 15.3 | 73.0 |
| INPD | 1.7 | 2.5 | 9.0 | 25.3 | 65.7 | 1.6 | 2.3 | 10.4 | 29.8 | 59.7 |
| PARA | 16.4 | 16.7 | 12.8 | 16.6 | 70.6 | 17.6 | 17.5 | 11.5 | 15.4 | 73.1 |
| PROF | 10.4 | 8.1 | 19.6 | 19.0 | 61.4 | 11.0 | 15.3 | 18.4 | 23.3 | 58.3 |
| ENPA | 17.1 | 21.7 | 25.0 | 22.0 | 53.0 | 15.9 | 19.5 | 24.9 | 27.8 | 47.3 |
| DMAG | 24.7 | 20.8 | 13.3 | 41.7 | 45.0 | 22.0 | 21.9 | 14.1 | 43.5 | 42.4 |
| AUAG | 22.9 | 19.7 | 3.3 | 7.5 | 89.2 | 22.6 | 17.3 | 1.3 | 6.8 | 91.9 |
| DOME | 10.6 | 10.5 | 15.1 | 26.3 | 58.6 | 9.3 | 10.9 | 16.5 | 21.4 | 62.1 |
| <i>Ausiliarie</i> | | | | | | | | | | |
| INAD | 12.2 | 11.3 | 13.9 | 20.9 | 65.3 | 12.6 | 12.6 | 13.7 | 21.0 | 65.3 |
| INAI | 27.7 | 29.6 | 17.9 | 18.3 | 63.8 | 26.4 | 28.1 | 15.0 | 20.5 | 64.4 |
| INAP | 15.0 | 16.3 | 35.7 | 24.5 | 39.7 | 16.1 | 20.4 | 43.4 | 29.1 | 27.5 |

Note: (b) Percentuale dei non occupati LFS sul totale degli individui con copertura C=1; (c) Come per (b) ma riferito alle sole interviste proxy; (d) Percentuale sul totale dei non occupati LFS con copertura C=1;
 Fonte: LFS-ADMIN sistema di ponderazione censuario

2.1.3. La diversa incidenza dei segnali di copertura

L'eterogeneità dell'incidenza dei segnali di copertura (Tavola 3) dipende da almeno due aspetti concomitanti:

- ✓ Da un lato, dalle caratteristiche dello specifico rapporto di lavoro descritto da ciascuna fonte e dalle eterogeneità dei vari segmenti del mercato del lavoro (dipendente, autonomo, atipico, agricolo, ecc.);
- ✓ Dall'altro dalla precisione e accuratezza con cui la copertura della fonte può essere ricondotta alla settimana di riferimento dell'intervista LFS.

Il primo effetto è chiaramente visibile nel caso di fonti come DMAG, PARA, INAI o ENPA: in breve, le fonti che riguardano forme contrattuali più discontinue evidenziano una quota più bassa di individui che, pur tracciati dalla fonte nell'anno di riferimento, non risultano però effettivamente coperti dalla fonte nella settimana di riferimento dell'intervista LFS.

Il secondo effetto è dovuto alla poca precisione, in taluni casi, dei segnali relativi alle date di copertura della fonte. Per esempio, nel caso di AUAG il segnale di presenza è esclusivamente annuale: ne consegue che un segnale di copertura di primo livello $C_{il}^{AUAG} = 1$ viene di fatto assegnato a tutti gli individui rintracciati nella fonte. Qualcosa di analogo avviene nella fonte PROF, dove manca un riferimento temporale puntuale dell'attività lavorativa effettivamente svolta e dove l'unico segnale assimilabile è fornito dalle date di inizio e cessazione della partita Iva. Vi è dunque un'evidente possibilità di sovrastima della copertura offerta da queste fonti. Ciò riguarda anche, sia pure in misura inferiore, il segnale di copertura INDI.

Da segnalare infine che le interviste proxy non sembrano giocare un ruolo apprezzabile nel determinare questo risultato (Tavola 3).

2.1.4. L'incoerenza fra i segnali LFS e ADMIN

L'incidenza dei casi di incoerenza fra il segnale di copertura di primo livello delle fonti ADMIN e il segnale di occupazione LFS appare fortemente eterogenea tra le fonti (Tavola 4). Questa eterogeneità sembra avere caratteri di stabilità nei due anni considerati. Le ragioni che sottostanno a

queste evidenze possono essere di vario genere:

- ✓ Possono dipendere, in generale, dalla diversa finalità delle fonti, e in particolare dal fatto che le fonti ADMIN non necessariamente misurano l'occupazione con criteri compatibili con quelli fissati nei regolamenti della statistica ufficiale. In questo caso sono dunque all'opera effetti di sovrastima dell'occupazione da parte di ADMIN legati sia alla già citata difficoltà di ricondurre il dato alla settimana di riferimento LFS sia alla non necessaria corrispondenza fra traccia ADMIN ed effettivo svolgimento dell'attività lavorativa;
- ✓ Possono inoltre dipendere dalla propensione degli intervistati LFS a omettere in qualche caso la segnalazione dello svolgimento di attività lavorative nella settimana di riferimento. I motivi in questo caso possono essere vari - fra cui tentativi di ridurre il disturbo statistico, imprecisioni dovute alle interviste proxy, desiderio di occultare le proprie attività - e possono dipendere anche dal comportamento del rilevatore. In ogni caso sono all'opera effetti che portano alla sottostima dell'occupazione da parte di LFS.

La diversa incidenza delle incoerenze nelle varie fonti conferma senz'altro la possibilità di una sovrastima da parte di ADMIN. È infatti evidente che:

- ✓ Le fonti meno precise sulle date di copertura sono anche quelle con una maggiore incidenza delle incoerenze con LFS (come nel caso di INDI e AUAG);
- ✓ Le fonti più corpose sul lavoro dipendente (EMEN e INPD), che peraltro identificano dei rapporti di lavoro piuttosto standardizzati e "tipici", sono per contro quelle meno incoerenti con LFS;
- ✓ Le fonti caratterizzate da una maggiore discontinuità dei rapporti di lavoro (DMAG, PARA, ENPA) mostrano anch'esse una elevata incidenza di incoerenze;
- ✓ Le fonti si differenziano anche sulla tipologia prevalente dei segnali di incoerenza con LFS. In alcune fonti essi caratterizzano più frequentemente i disoccupati (ENPA, PROF, EMEN), in altre la zona grigia degli inattivi (DMAG, DOME, INPD), in altre ancora gli inattivi *tout court* (AUAG e INDI);
- ✓ Gli effetti delle interviste proxy nel determinare le incoerenze sono però trascurabili, leggermente più evidenti per le fonti sul lavoro agricolo.

Le fonti colgono dunque in maniera non omogenea i segnali reali di occupazione, appaiono "specializzate" su particolari segmenti del mercato del lavoro, e questo insieme di elementi le rende poco confrontabili fra loro. Esse mostrano in particolare una tendenza a sovra determinare lo stato di occupazione in ragione di due tipi di effetti che interagiscono fra loro:

- ✓ Quelli dovuti alla scarsa precisione del periodo di riferimento;
- ✓ Quelli dovuti alla natura del rapporto di lavoro che esse descrivono: più questo è intrinsecamente discontinuo più viene indebolita la potenza stessa del segnale di copertura.

Per altro verso, la presenza di effetti di sottostima da parte di LFS trova chiare conferme. Infatti:

- ✓ Anche EMEN e INPD mostrano una quota, modesta ma non trascurabile, di incoerenze (dato anche il peso assoluto di queste fonti), segno del fatto che comunque esiste un effetto di occultamento della condizione lavorativa in LFS e dunque di sottocopertura delle posizioni lavorative;
- ✓ Questo effetto, come si vedrà, è peraltro confermato dalla notevole sottostima delle seconde posizioni da parte di LFS, misurabile anche solo su una fonte relativamente sicura come EMEN;
- ✓ L'effetto delle interviste proxy - nelle quali colui che risponde per l'intervistato può rischiare di classificarlo come non occupato - è in realtà assai modesto e non spiega se non in minima parte la presenza delle incoerenze.

Ulteriori analisi di queste incoerenze condotte usando in tandem l'analisi delle corrispondenze multiple e l'analisi cluster hanno peraltro evidenziato piuttosto chiaramente come nel determinare le incoerenze siano all'opera distorsioni causate sia da ADMIN sia da LFS. Esistono infatti situazioni individuali per le quali la non compatibilità dei segnali LFS e ADMIN appare ragionevolmente attribuibile a una sovracopertura da parte di ADMIN: si tratta di cluster di individui con segnali ADMIN di copertura di primo livello molto deboli (economici, temporali, contrattuali) associati a un segnale di non occupazione LFS. All'opposto, esistono situazioni ADMIN decisamente stabili e consolidate corrispondenti a non occupati LFS.

2.1.5. Scelte operative e definizione di "posizione regolare"

L'analisi ha condotto alle seguenti conclusioni:

- ✓ occorre rendere omogenei e accrescere la precisione dei segnali di copertura delle fonti ADMIN eliminando gli effetti di sovracopertura;
- ✓ occorre stimare e correggere gli effetti dovuti alla sottocopertura degli occupati da parte di LFS.

Con riferimento al primo obiettivo, il problema affrontato dal GdL è stato quello di individuare i criteri con cui selezionare le osservazioni con segnali di copertura di primo livello troppo deboli e tali dunque da essere considerati insufficienti a giustificare uno stato di occupazione. In termini generali, ciò ha significato studiare separatamente per ciascuna fonte l'associazione fra lo stato di occupazione LFS e la copertura di primo livello ADMIN sulla base delle numerose covariate provenienti dalle due fonti. Per questa finalità è stato utilizzato un modello logistico attraverso il quale sono state discriminate le osservazioni a seconda dell'efficacia del segnale di copertura. In sostanza, sulla base delle predizioni stimate dal modello sono stati validati i segnali di copertura amministrativa (coperture di secondo livello).

Ai segnali validati è stata associata l'etichetta di posizione lavorativa "regolare", intendendo con questo termine il fatto che:

- ✓ la posizione lavorativa è tracciata in ADMIN;
- ✓ a essa corrisponde una effettiva attività lavorativa nella settimana di riferimento LFS.

Il carattere di regolarità è dunque qui riferito alla corrispondenza fra lo svolgimento dell'attività lavorativa e una copertura amministrativa validata, ma non esclude:

- ✓ la copertura amministrativa possa essere formalmente non regolare (p.es., una collaborazione a progetto che nasconde un lavoro subordinato, o a causa di un sottoinquadramento);
- ✓ l'input di lavoro associato alla prestazione lavorativa sia in parte svolto in condizioni di irregolarità (p.es. perché ADMIN segnala un part-time mentre la posizione effettiva è full-time, o perché vi sono degli straordinari "fuori busta").

Quest'ultimo caso si presta più agevolmente a misurazione diretta della componente non regolare basata sul confronto fra orario effettivo e teorico, almeno per quanto riguarda l'occupazione dipendente. La quantificazione invece del primo effetto appare più problematica sia perché richiede la disponibilità di informazioni ulteriori sia perché richiede un metodo di stima indiretto fondato sulla differenza qualitativa fra la posizione lavorativa tracciata in ADMIN e quella effettivamente ricoperta. È senz'altro a portata di mano invece l'analisi di casistiche particolari, come quelle legate ai cosiddetti *falsi part-time*: un Istat working paper è stato recentemente dedicato a questo tema, basato proprio sui microdati del campione integrato LFS-ADMIN³⁰.

³⁰ De Gregorio – Giordano (2014).

Con riferimento al secondo obiettivo (la correzione della sottocopertura dell'occupazione misurata da LFS), si parte dalla constatazione del fatto che tramite la correzione dei segnali di copertura ADMIN e il segnale di secondo livello si ottiene una stima puntuale degli occupati con posizioni lavorative definite "regolari". Da un lato ciò porta a una consistente riduzione della eterogeneità fra le fonti in merito all'incidenza dei casi di non coerenza con LFS, dal momento che come si vedrà l'incidenza dei segnali non validati è molto più elevata proprio nelle fonti dove l'incoerenza è più forte; dall'altro quelle incoerenze – benché ridotte – comunque persistono anche con i segnali di secondo livello e questo residuo viene pertanto interpretato come misura della sottostima degli occupati regolari da parte di LFS. Per il segmento degli occupati regolari, questa sottostima viene pertanto corretta automaticamente su ciascun individuo del campione LFS sulla base del segnale di copertura ADMIN di secondo livello: tutti gli intervistati LFS con copertura di secondo livello sono definiti "occupati regolari", e una parte di essi appartengono al segmento dei non occupati LFS. La loro incidenza sul totale degli occupati regolari rappresenta la base per la valutazione della sottocopertura LFS relativa al segmento complementare, quello degli occupati privi di posizioni regolari.

2.2. Impostazione della stima

2.2.1. Aspetti preliminari

Per la stima degli occupati vengono utilizzate le informazioni relative ai singoli individui LFS raccolte nella base dati integrata O la cui riga generica è

$$O_i = \{R_i; a_i^1, \dots, a_i^K\} \quad <1>$$

A ogni individuo i del campione sono associate le variabili LFS, tutte le variabili delle $K=9$ fonti ADMIN e le $m(b)$ variabili contenute nei registri statistici sui datori di lavoro (relativi alle prime due posizioni per ciascuna fonte k).

Due insiemi di variabili assumono un rilievo importante in questa fase della stima. Anzitutto, tramite le informazioni contenute in R , ciascun individuo i viene classificato secondo una variabile dicotomica OCC_i^R a seconda che esso risulti occupato (1) o non occupato (0) nella settimana di riferimento dell'intervista LFS.

Inoltre, da ciascuna fonte k viene ricavato un "indicatore di primo livello", e cioè il flag binario $C_i^k = C_{i1}^k$ che indica la presenza (1) o l'assenza di copertura (0) nella posizione principale assunta dall'individuo nella fonte k . Si tratta di un indicatore asimmetrico, nel senso che si assume che $C_i^k = 0$ indichi con "certezza" la mancanza di copertura mentre $C_j^k = 1$ è un segnale la cui efficacia dipende dalla precisione con cui la fonte fornisce indicazioni relative alla settimana di riferimento dell'intervista LFS³¹.

Partendo da questo indicatore di primo livello relativo a ciascuna fonte, si perviene a un indicatore complessivo individuale di primo livello C_i che segnala la copertura previdenziale dell'individuo i in almeno una fonte:

$$\text{se } \exists k : C_i^k = 1 \Rightarrow C_i = 1 \quad <2>$$

³¹ L'ipotesi che $C=0$ indichi con "certezza" mancanza di segnale rappresenta ovviamente una semplificazione della realtà. Gli archivi amministrativi possono infatti: 1) essere non aggiornati e la mancanza di copertura potrebbe essere dovuta a una non tempestiva comunicazione o registrazione di occupazione; 2) recare informazioni sulla durata dei contratti non allineate con la durata effettiva dell'occupazione. Nel primo caso l'ipotesi $C=0$ equivale pertanto ad assumere la completezza degli archivi, ipotesi ragionevole per l'anno oggetto di analisi ma da valutare comunque bene per i periodi futuri; nel secondo caso l'ipotesi $C=0$ equivale ad assegnare a lavoro non regolare le prestazioni effettuate senza la copertura contrattuale.

Considerando i segnali OCC_i^R e C_i , è possibile – tramite i coefficienti di riporto all’universo di LFS - classificare la popolazione N come segue:

Figura 5. Segnale di copertura primario

| | | C | | |
|---------|--------|----------|----------|----------|
| | | 1 | 0 | Totale |
| OCC^R | 1 | X_{11} | X_{12} | X_{10} |
| | 0 | X_{21} | X_{22} | X_{20} |
| | Totale | X_{01} | X_{02} | N |

L’obiettivo è pervenire alla stima del numero complessivo di occupati regolari e non regolari che rientrano nel campo di osservazione della contabilità nazionale (OCC^C). A partire dalla base dati integrata O , occorre stimare la classificazione della popolazione in età di lavoro nel campo di osservazione LFS in base alla natura della effettiva condizione lavorativa:

Figura 6. Stime CN dell’occupazione

| OCC^C | | | |
|-------------------|-----------------------|--------------|--------|
| Occupati Regolari | Occupati non regolari | Non occupati | Totale |
| Y_1 | Y_2 | Y_3 | N |

Come descritto nei successivi paragrafi, tale riclassificazione viene effettuata;

- ✓ sulla base di alcune ipotesi a priori;
- ✓ attraverso l’applicazione in sequenza di due modelli statistici finalizzati a individuare e correggere, da un lato, i casi di sovracopertura delle fonti amministrative e, dall’altro, i casi di sottocopertura degli occupati LFS senza segnale di copertura previdenziale nelle fonti.

2.2.2. Le ipotesi

Il modello con il quale vengono analizzate le incoerenze fra le fonti è fondato su quattro ipotesi.

Ipotesi 1. Il segnale OCC_i^R di occupazione LFS, se uguale a 1, non è soggetto a errore. Si assume in sostanza che gli occupati LFS siano anche occupati CN:

$$\text{se } i \in X_{10} \Rightarrow i \in Y_1 \cup Y_2 \quad <3>$$

o in maniera equivalente

$$\text{se } OCC_i^R = 1 \Rightarrow OCC_i^C = 1 \quad <4>$$

Questa ipotesi si poggia sulla natura selettiva del quesito oggettivo LFS (almeno un’ora lavorata nella settimana di riferimento) e sull’assunzione che l’intervistato non abbia interesse a celare la

condizione di non occupazione³².

Ipotesi 2. Il segnale di copertura primario, se uguale a 0, indica con probabilità pari a 1 l'effettiva assenza di copertura previdenziale nella settimana di riferimento. Ne segue che, tenuto conto dell'ipotesi 1, la condizione di occupato LFS senza copertura primaria identifica un occupato non regolare. Detta P_{iscy} la probabilità che l'individuo i proveniente dall'insieme X_{sc} venga classificato nell'insieme obiettivo Y_y , si assume che l'assenza di segnale di copertura primario indichi con certezza la mancata copertura previdenziale: unita all'ipotesi 1, questa assunzione implica che la probabilità che risulti un occupato non regolare è pari ad uno:

$$P_{i122} = \Pr[i \in Y_2 \mid i \in X_{12}] = 1 \quad <5>$$

Ipotesi 3. Il segnale di copertura primario, se uguale a 1, indica con probabilità minore o uguale a uno l'effettiva presenza di una copertura previdenziale nella fonte k . Questa ipotesi, unita all'ipotesi 1, implica che un individuo con copertura primaria in almeno una fonte ha una probabilità compresa fra 0 e 1 di essere un occupato regolare:

$$P_{i111} = \Pr[i \in Y_1 \mid i \in X_{11}] \in [0,1] \quad <6>$$

mentre il complemento

$$P_{i112} = \Pr[i \in Y_2 \mid i \in X_{11}] = 1 - P_{i111} \quad <7>$$

è la probabilità di essere classificato fra gli occupati non regolari. Infine, è pari a zero la probabilità di non essere occupato (per l'ipotesi 1):

$$P_{i113} = \Pr[i \in Y_3 \mid i \in X_{11}] = 0 \quad <8>$$

Ipotesi 4. Il segnale OCC_i^R di occupazione LFS, se uguale a zero, può essere soggetto a errore di misura. L'ipotesi è che l'intervistato possa occultare più o meno volontariamente, anche a causa di una intervista proxy, la propria condizione di occupato. Questa ipotesi è supportata da un lato dai risultati di altre rilevazioni sulle famiglie e dall'altro dal peso stesso dei non occupati dotati di coperture previdenziali. Un individuo classificato fra i non occupati privi di segnali di copertura avrà pertanto una certa probabilità di essere un occupato non regolare:

$$P_{i222} = \Pr[i \in Y_2 \mid i \in X_{22}] = [0,1] \quad <9>$$

³² Questa assunzione vale anche per le interviste proxy. Si tratta evidentemente anche in questo caso di una semplificazione, la quale poggia sia sulla non significatività che il tipo di intervista mostra nello spiegare i casi di incongruenza fra fonti di indagine e fonti amministrative sia sulla considerazione che, essendo la variabile "condizione professionale" una variabile core dell'indagine, la sua rilevazione tramite intervista proxy non ne dovrebbe ridurre l'affidabilità.

mentre, a complemento di questa, ha una probabilità

$$P_{i223} = \Pr[i \in Y_3 | i \in X_{22}] = 1 - P_{i222} \quad <10>$$

di essere effettivamente un non occupato. Ovviamente, per l'ipotesi 2

$$P_{i221} = \Pr[i \in Y_1 | i \in X_{22}] = 0 \quad <11>$$

Con riferimento agli individui classificati tra i non occupati con segnali di copertura, la loro effettiva collocazione dipende dalla potenza del segnale (ossia dalla probabilità che esso li faccia corrispondere a occupati regolari) e dalla probabilità di "occultamento" della condizione di occupato. In sostanza, la probabilità di essere un occupato regolare è:

$$P_{i211} = \Pr[i \in Y_1 | i \in X_{21}] = [0,1] \quad <12>$$

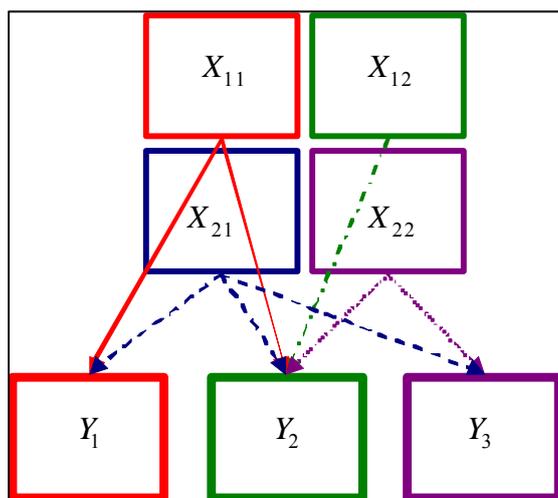
quella di essere un occupato non regolare è pari a

$$P_{i212} = \Pr[i \in Y_2 | i \in X_{21}] = [0,1] \quad <13>$$

mentre quella di essere un non occupato è il complemento a uno delle prime due:

$$P_{i213} = \Pr[i \in Y_3 | i \in X_{21}] = 1 - P_{i211} - P_{i212} \quad <14>$$

Figura 7



2.3. Modelli per la correzione delle fonti

Vengono utilizzati due modelli con l'obiettivo di correggere:

1. la sovracopertura del segnale amministrativo;
2. la sottocopertura degli occupati non regolari da parte di LFS.

2.3.1. La correzione della sovracopertura dei segnali amministrativi

In particolare, le singole fonti amministrative possono essere affette, in misura variabile, da:

- ✓ scarsa precisione dell'indicatore di copertura di primo livello. L'utilizzo dell'intervallo fra inizio e cessazione dell'attività è una convenzione che non assicura l'effettivo svolgimento dell'attività lavorativa. Può dunque generare false coperture;
- ✓ a prescindere dalla copertura temporale, la copertura della fonte non assicura necessariamente l'effettivo svolgimento dell'attività lavorativa nella settimana di riferimento dell'intervista LFS. Ciò può generare false coperture ma anche false mancanze di coperture.

Per la correzione dell'indicatore di primo livello viene stimata, per ciascun individuo e per ciascuna fonte, la probabilità che il segnale amministrativo sia sufficiente a garantire la natura regolare della condizione di occupato. Si agisce pertanto sul sottoinsieme delle osservazioni con segnale di copertura $C=I$, ossia nel sottoinsieme $X_{01} = X_{11} \cup X_{21}$. L'ipotesi di lavoro è quella di sfruttare il segnale di occupazione LFS per individuare le variabili delle fonti amministrative che influenzano maggiormente la coerenza dei segnali di copertura con la condizione di occupato. Si ipotizza che P_{i01}^k la probabilità per l'individuo i di risultare occupato a LFS (variabile risposta), sia funzione dell'insieme delle variabili contenute nella fonte k (a_i^k) e dell'insieme delle variabili strutturali LFS (R_i):

$$P_{i01}^k = \hat{\text{Pr}}[\text{OCC}_i^R = 1 \mid i \in X_{01}, R_i, a_i^k] \quad <15>$$

Per la stima di questa relazione viene utilizzato un modello logistico:

$$\text{logit}(P_{i01}^k) = \beta_0 + \sum_{v=1}^r \beta_v^R R_{vi} + \sum_{v=1}^{m(k)} \beta_v^{a^k} a_{vi}^k + \sum_{v=1}^m \beta_v^k M_{vi}^k \quad <16>$$

dove:

- ✓ R_v sono r variabili categoriche strutturali della rilevazione LFS: territorio (regione o ripartizione), genere, classe di età, cittadinanza, stato civile;
- ✓ a_v^k sono $m(k)$ variabili categoriche e quantitative contenute nella fonte amministrativa k . Indicano, a seconda dei casi, aspetti strutturali (il settore di attività e la forma giuridica del datore, il numero di posizioni lavorative), l'intensità del rapporto di lavoro in termini di copertura contrattuale nella settimana di riferimento, giornate retribuite e valori imponibili (nel trimestre e nell'anno di riferimento), altre caratteristiche del rapporto di lavoro (durata, regime orario, tipologie contrattuali, ecc.), la presenza di coperture in altre fonti;

- ✓ M_v^k è una selezione di m effetti misti ottenuti combinando variabili dentro e fra gli insiemi a e R .

Il modello è stato stimato sulle sole osservazioni corrispondenti a interviste dirette (escluse dunque le interviste proxy)³³. I parametri del modello sono stati successivamente applicati su tutte le osservazioni per ottenere una stima individuale della probabilità di occupazione \hat{P}_{i01}^k .

Da queste stime viene derivato un “indicatore di copertura di secondo livello”, D_i^k , che traduce in una dimensione binaria la variabile continua stimata \hat{P}_{i01}^k :

$$D_i^k = \begin{cases} 0 & C_i^k \neq 1 \\ 0 & \Leftarrow \hat{P}_{i01}^k < \alpha^k \\ 1 & \Leftarrow \hat{P}_{i01}^k \geq \alpha^k \end{cases} \quad <17>$$

dove $0 < \alpha^k < 1$ è una soglia prefissata. Analogamente a quanto fatto per il segnale primario C , è possibile definire un indicatore di secondo livello complessivo binario D_i che indica la copertura previdenziale dell'individuo i in almeno una fonte:

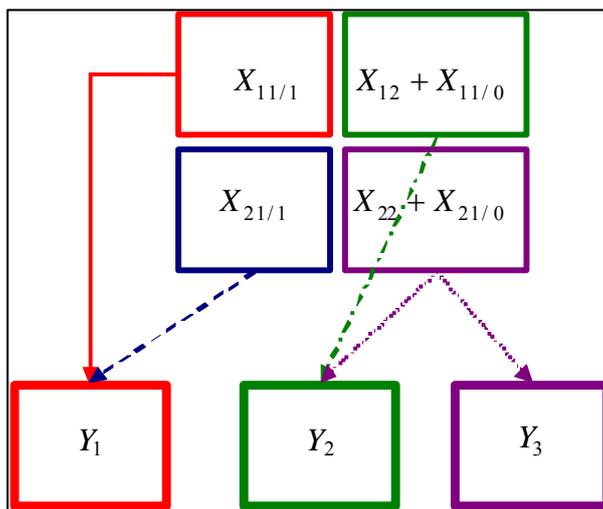
$$\text{se } \exists k : D_i^k = 1 \Rightarrow D_i = 1 \quad <18>$$

Tramite l'indicatore di secondo livello, dagli insiemi X_{11} e X_{21} - costituiti dagli individui con copertura previdenziale - vengono scorporati due sottoinsiemi relativi alle unità prive di segnale di secondo livello. Si tratta rispettivamente dei sottoinsiemi $X_{11/0}$ e $X_{21/0}$: il primo va ad alimentare il bacino degli occupati senza copertura e il secondo si colloca nel bacino dei non occupati senza copertura (Figura 8).

³³ Nella versione che è stata testata in una prima elaborazione sono stati anche esclusi gli individui che dichiarano di aver avuto esperienze lavorative. L'assenza di esperienze lavorative riportata in LFS è stata associata infatti a una condizione di non occupazione, e ciò a prescindere dalla presenza o meno di segnali di copertura. Si tratta di una ipotesi comunque da valutare ed eventualmente applicare limitatamente al segmento degli intervistati non *proxy*.

Figura 8. Segnale di copertura secondario

| | | | | |
|---------|---|------------|---------------------|----------|
| | | D | | |
| | | 1 | 0 | |
| OCC^R | 1 | $X_{11/1}$ | $X_{12} + X_{11/0}$ | X_{10} |
| | 0 | $X_{21/1}$ | $X_{22} + X_{21/0}$ | X_{20} |
| | | $X_{01/1}$ | $X_{02} + X_{01/0}$ | N |



2.3.2. La correzione della sottocopertura degli occupati non regolari da parte di LFS

L'obiettivo è quello di stimare la porzione degli individui non occupati secondo LFS e privi di segnali di copertura che va riclassificata nel segmento degli occupati non regolari. La possibilità di una sottocopertura degli occupati non regolari è segnalata dal fatto che vi è senz'altro sottocopertura nel segmento degli occupati regolari: la correzione del segnale amministrativo e la creazione del flag di copertura D identificano infatti un sottoinsieme $X_{21/1}$ costituito da occupati secondo le fonti amministrative ($D=1$) che non risultano occupati secondo LFS. Questo sottoinsieme consente pertanto una stima della sottocopertura dell'occupazione regolare tramite il rapporto di coesistenza:

$$h = X_{21/1} / X_{11/1} \quad <19>$$

il quale indica il numero medio di individui con copertura nelle fonti amministrative ma che non sono occupati secondo LFS per ogni individuo con copertura e occupato LFS. Questo numero viene in sostanza interpretato come un indicatore della propensione media a sottodichiarare lo stato di occupazione da parte degli individui con coperture di secondo livello.

Per la stima della sottocopertura degli occupati nel segmento complementare degli individui LFS privi di segnali ($D=0$) si parte dall'ipotesi che il rapporto di coesistenza stimato per $D=1$ non sia distorto e che sia dunque applicabile anche al segmento complementare³⁴. La mancata copertura

³⁴ La scelta di utilizzare un rapporto di coesistenza in luogo di un rapporto di composizione dipende dal fatto che i sottoinsiemi dei non occupati con copertura e privi di copertura non sono fra loro omogenei. Dunque si è preferito riportare l'indicatore al corrispondente segmento degli occupati.

degli occupati spiegata dal coefficiente h può essere imputabile:

- ✓ a un eventuale errore di rilevazione, dovuto per esempio alla negligenza e opportunismo del rilevatore e/o dell'intervistato;
- ✓ all'occultamento volontario da parte dell'intervistato della propria condizione di occupato, dovuto per esempio al timore di perdere diritti acquisiti o di esporsi a ulteriori verifiche fiscali o previdenziali. In questo caso potrebbero anche celarsi posizioni lavorative non rilevate anche solo parzialmente emerse.

Motivazioni analoghe dovrebbero sottostare al comportamento del segmento $D=0$, per il quale i motivi di occultamento della condizione di occupato potrebbero risultare rafforzate dal momento che celerebbero posizioni lavorative interamente non regolari. È dunque verosimile attendersi che la stima sul segmento $D=0$ - ottenuta utilizzando le evidenze emerse nel segmento $D=1$ - rappresenti una stima di minima e comunque prudentiale.

Si è proceduto pertanto come segue:

- ✓ si adotta una stratificazione G basata su alcune variabili strutturali LFS (territorio, età, genere, cittadinanza, tipo di intervista, trimestre);
- ✓ per ogni strato $g \in G$ viene prodotta una stima del rapporto di coesistenza:

$$\hat{h}_g = \frac{X_{g21/1}}{X_{g11/1}} \quad \langle 20 \rangle$$

- ✓ per ogni strato g si stima il numero di occupati non regolari inclusi fra in non occupati LFS privi di coperture ($D=0$) utilizzando il rapporto di coesistenza stimato;

$$\hat{X}_{g22/1} = \hat{h}_g (X_{g22} + X_{g21/0}) \quad \langle 21 \rangle$$

- ✓ si modella per il segmento $D=0$ la probabilità di essere occupato LFS:

$$P_{i02} = \hat{\Pr}[OCC_i^R = 1 | i \in X_{02} \cup X_{01/0}, R_i, a_i] \quad \langle 22 \rangle$$

- ✓ utilizzando il modello logistico

$$\logit(P_{i02}) = \beta_0 + \sum_{v=1}^r \beta_v^R R_{vi} + \sum_{v=1}^{m(k)} \beta_v^a a_{vi} + \sum_{v=1}^m \beta_v M_{vi} \quad \langle 23 \rangle$$

dove le R_v sono le variabili categoriche strutturali LFS (territorio, genere, classe di età, cittadinanza, stato civile), a_v sono le variabili che qualificano l'eventuale presenza o copertura primaria nelle fonti amministrative, M_v^k sono m effetti misti;

- ✓ vengono stimate le probabilità individuali di occupazione irregolare \hat{P}_{i02} ;

- ✓ fra i non occupati con $D=0$ in ciascun segmento g vengono selezionati come occupati non regolari gli intervistati con la probabilità più elevata di occupazione fino all'occorrenza del totale stimato $\hat{X}_{g22/1}$.

In particolare, detta \hat{P}_{mg02} la successione ordinata in ordine decrescente delle \hat{P}_{i02} tale che $\hat{P}_{mg02} > \hat{P}_{m+1g02}$ si avrà:

$$m \in g, D_m = 1, OCC_m^R = 0, \sum_{u=1}^m w_u \leq X_{g22/1} \Rightarrow m \in X_{22/1} \quad \langle 24 \rangle$$

dove w è il coefficiente di ponderazione LFS.

Questa stima completa lo schema concettuale che definisce il quadro delle stime dell'occupazione di CN a partire dall'utilizzo combinato di LFS e delle fonti amministrative sull'occupazione (figure 9 e 10).

Figura 9. Correzione della sottocopertura LFS degli occupati non regolari

| | | D | | |
|-----|---|------------|--------------------------------|--------------------------|
| | | 1 | 0 | |
| OCC | 1 | $X_{11/1}$ | $X_{12} + X_{11/0} + X_{22/1}$ | $X_{10} + X_{22/1}$ + |
| | 0 | $X_{21/1}$ | $X_{22} + X_{21/0} - X_{22/1}$ | $X_{20} - X_{22/1}$ |
| | | $X_{01/1}$ | $X_{02} + X_{01/0}$ | N |

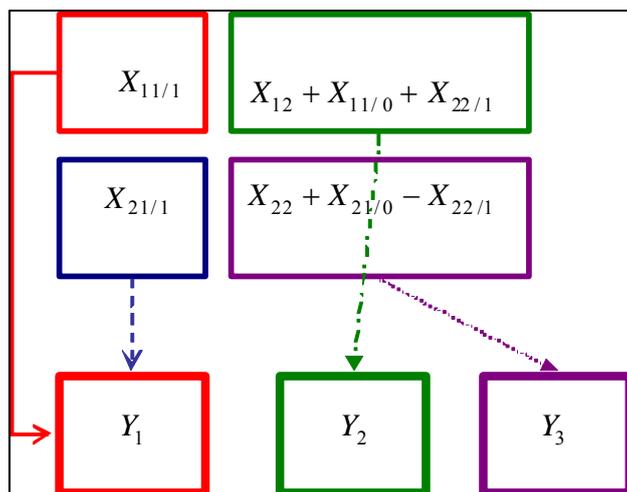


Figura 10. Stime CN

| OCC^C | | | |
|-----------------------|--------------------------------|--------------------------------|--------|
| Occupati regolari | Occupati non regolari | Non occupati | Totale |
| $X_{11/1} + X_{21/1}$ | $X_{12} + X_{11/0} + X_{22/1}$ | $X_{22} + X_{21/0} - X_{22/1}$ | N |

Si noti comunque che l'etichetta di occupato regolare viene attribuita, in questa fase del processo di stima, agli individui che abbiano almeno una posizione regolare ossia:

- ✓ tracciata nelle fonti,
- ✓ relativa a periodi temporali compatibili con la settimana di riferimento dell'intervista e
- ✓ validata con modello probabilistico.

Ciò vale a prescindere dall'ordine della posizione lavorativa regolare. Una volta portata a termine la stima delle posizioni lavorative, e una volta attribuito perciò un ordine a tutte le posizioni afferenti a un medesimo individuo, saranno più propriamente classificati fra i regolari solo gli occupati con la posizione principale regolare.

2.4. I principali risultati

2.4.1. La validazione del segnale di copertura

La procedura di correzione dei segnali di copertura di primo livello provenienti dalle singole fonti ha operato in maniera piuttosto differenziata (Tavola 5). In particolare:

- ✓ La mancata validazione dei segnali di copertura ($C_i^k = 1$ e $D_i^k = 0$) ha un maggiore incidenza presso le fonti caratterizzate da una maggiore incoerenza fra segnali ADMIN e segnali LFS;
- ✓ Spicca evidentemente il caso di INDI, dove fra un quarto e un quinto delle coperture di primo livello sono state annullate. Le fonti agricole sono ambedue fortemente soggette alla procedura di correzione;
- ✓ Seguono ENPA e PARA, con oltre il 10% delle coperture non validate; PROF e DOME si collocano vicino 5%; EMEN e INPD hanno una incidenza delle correzioni percentualmente irrisoria;
- ✓ Nei due anni considerati l'incidenza delle correzioni del segnale ADMIN è pressoché identica. Nel 2011 si nota infatti solo una sua contrazione in DMAG e DOME;
- ✓ Il medesimo indicatore riferito alle interviste proxy non sembra evidenziare distorsioni significative.

Tavola 5. Incidenza delle coperture di primo livello non validate (a), per anno, fonte e tipo di intervista

| Fonte | 2010 | | | 2011 | | |
|-------|--------|-------------------|-------|--------|-------------------|-------|
| | Totale | di cui Interviste | | Totale | di cui Interviste | |
| | | Dirette | Proxy | | Dirette | Proxy |
| EMEN | 1.5 | 1.5 | 2.0 | 1.5 | 1.4 | 2.0 |
| INDI | 22.7 | 23.2 | 20.1 | 24.2 | 24.4 | 23.0 |
| INPD | 0.3 | 0.3 | 0.5 | 0.4 | 0.4 | 0.4 |
| PARA | 13.6 | 13.5 | 13.9 | 14.1 | 14.3 | 13.0 |
| PROF | 6.6 | 6.5 | 7.2 | 6.1 | 6.1 | 6.0 |
| ENPA | 13.7 | 13.6 | 14.2 | 14.0 | 12.5 | 20.9 |
| DMAG | 18.0 | 19.5 | 11.1 | 13.7 | 14.0 | 12.8 |
| AUAG | 17.0 | 17.2 | 15.4 | 17.3 | 17.7 | 14.0 |
| DOME | 6.8 | 6.4 | 9.5 | 4.6 | 4.4 | 5.7 |

Note: (a) Incidenza percentuale degli individui con copertura validata sul totale degli individui con copertura di primo livello nella fonte
Fonte: LFS-ADMIN sistema di ponderazione censuario

L'incidenza delle correzioni varia a seconda della posizione lavorativa registrata in LFSb (Tavola 6). Infatti:

- ✓ L'incidenza è molto bassa per gli occupati LFS, intermedia per disoccupati e inattivi "grigi", molto alta per gli altri inattivi;
- ✓ EMEN e INPD vedono validata la copertura a oltre il 99% degli occupati. Solo nelle fonti agricole si scende sotto al 95%;
- ✓ L'incidenza delle mancate validazioni è appena più alta presso gli inattivi grigi rispetto ai disoccupati: fanno eccezione PARA e AUAG. Una notevole distanza fra disoccupati e inattivi grigi si ha in ENPA e PROF;
- ✓ Con riferimento agli altri inattivi, spicca la mancata riconferma di circa tre quarti delle coperture di primo livello INDI, ENPA e PARA;
- ✓ I risultati dei due anni sono molto simili.

Tavola 6. Incidenza delle coperture di primo livello non validate (a), per anno, fonte e condizione LFS

| Fonte | 2010 | | | | 2011 | | | |
|-------|----------|-------------|----------------|----------------|----------|-------------|----------------|----------------|
| | Occupati | Disoccupati | Inattivi grigi | Altri inattivi | Occupati | Disoccupati | Inattivi grigi | Altri inattivi |
| EMEN | 0.6 | 12.9 | 16.9 | 28.0 | 0.5 | 15.0 | 14.2 | 32.8 |
| INDI | 4.2 | 50.1 | 54.5 | 84.6 | 4.6 | 52.4 | 58.2 | 86.0 |
| INPD | 0.1 | 9.8 | 11.1 | 16.1 | 0.1 | 22.1 | 21.1 | 9.8 |
| PARA | 3.9 | 46.0 | 40.6 | 71.2 | 3.7 | 36.1 | 32.4 | 73.6 |
| PROF | 2.3 | 19.8 | 36.1 | 53.6 | 1.8 | 15.9 | 31.2 | 52.4 |
| ENPA | 2.8 | 45.4 | 71.0 | 77.5 | 4.1 | 66.5 | 53.6 | 74.2 |
| DMAG | 8.1 | 32.2 | 36.1 | 63.5 | 5.7 | 39.5 | 33.5 | 52.4 |
| AUAG | 5.9 | 34.9 | 23.9 | 57.9 | 5.5 | 29.6 | 27.0 | 60.3 |
| DOME | 2.8 | 32.6 | 34.7 | 44.9 | 2.1 | 11.2 | 26.5 | 34.7 |

Note: (a) Incidenza percentuale degli individui con copertura non validata sul totale degli individui con copertura di primo livello nella fonte
Fonte: LFS-ADMIN sistema di ponderazione censuario

La correzione del segnale di copertura di primo livello porta inoltre alla riduzione dei casi di incoerenza fra le fonti (si confrontino le Tavole 4 e 7 e si veda la Figura 11). Si può notare in particolare che:

- ✓ Le riduzioni più forti si hanno per le fonti dove l'incidenza di questi casi era più elevata;
- ✓ La particolarità della fonte INDI, per la quale le incoerenze scendono da oltre il 25% a meno del 10%;
- ✓ La maggiore omogeneità fra le fonti relativamente alle coperture di secondo livello;
- ✓ Le fonti agricole continuano a mostrare incidenze significative delle incoerenze;

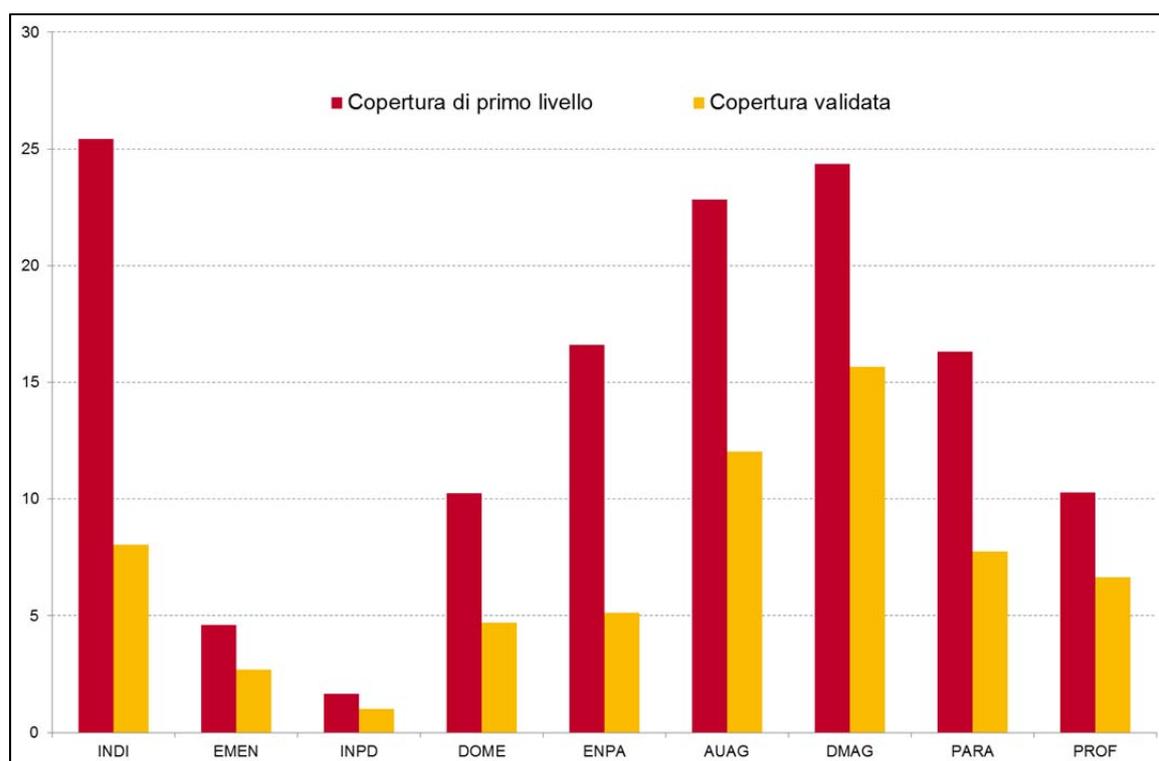
- ✓ La composizione per condizione LFS dei casi di incoerenza continua a riflettere specificità settoriali e normative, come nel caso di AUAG e INDI;
- ✓ I dati relativi al 2010 e al 2011 portano a risultati molto simili;
- ✓ L'effetto delle interviste proxy è assai modesto.

Tavola 7. Incidenza dei non occupati LFS sul totale degli individui con copertura validata (a), per anno, fonte e condizione LFS

| Fonte | 2010 | | | | | 2011 | | | | |
|-------|--------|-------|--------------|----------------|----------------|--------|-------|--------------|----------------|----------------|
| | Totale | Proxy | Composizione | | | Totale | Proxy | Composizione | | |
| | | | Disocc. | Inattivi grigi | Altri inattivi | | | Disocc. | Inattivi grigi | Altri inattivi |
| EMEN | 3.7 | 4.7 | 22.5 | 33.3 | 44.2 | 3.4 | 4.9 | 23.3 | 37.0 | 39.7 |
| INDI | 7.8 | 8.8 | 23.8 | 27.7 | 48.5 | 7.8 | 8.5 | 25.1 | 28.8 | 46.1 |
| INPD | 1.4 | 2.1 | 9.5 | 26.3 | 64.2 | 1.4 | 2.1 | 9.5 | 27.5 | 63.0 |
| PARA | 7.1 | 9.8 | 18.6 | 26.6 | 54.8 | 7.6 | 9.2 | 19.9 | 28.0 | 52.1 |
| PROF | 6.3 | 6.5 | 27.9 | 21.5 | 50.6 | 7.0 | 12.7 | 26.1 | 27.1 | 46.8 |
| ENPA | 6.3 | 12.6 | 43.5 | 20.3 | 36.2 | 6.2 | 11.0 | 24.9 | 38.6 | 36.5 |
| DMAG | 15.7 | 16.3 | 17.3 | 51.2 | 31.6 | 14.7 | 16.2 | 14.8 | 50.2 | 35.0 |
| AUAG | 12.5 | 10.5 | 4.7 | 12.5 | 82.7 | 11.6 | 8.5 | 2.2 | 11.8 | 86.0 |
| DOME | 6.8 | 8.1 | 17.0 | 28.8 | 54.2 | 7.0 | 8.8 | 20.7 | 22.1 | 57.2 |

Note: (a) Incidenza percentuale degli individui con copertura validata sul totale degli individui con copertura di primo livello nella fonte
 Fonte: LFS-ADMIN sistema di ponderazione censuario

Figura 11. Incidenza dei non occupati LFS sul totale degli individui con copertura amministrativa, per tipo di copertura e fonte. Anno 2010



2.4.2. Ulteriori aspetti della validazione dei segnali di copertura

Attraverso la correzione dell'indicatore di copertura di primo livello tramite l'indicatore di secondo livello *D* si determina una riduzione dei casi di incoerenza fra la condizione lavorativa rilevata da LFS e la copertura delle fonti amministrative. La loro incidenza si riduce di due terzi e il

loro numero assoluto si riduce drasticamente. La contrazione più vistosa riguarda gli inattivi in senso stretto: se sulla base dell'indicatore di primo livello questo segmento dava luogo a oltre l'80% delle incoerenze, con quello di secondo livello supera di poco il 50% (Tavola 8).

Tavola 8. Popolazione in età di lavoro per condizione lavorativa e indicatore copertura (solo record LFS con codice fiscale)

| Condizione | Indicatore di copertura | | | | |
|-----------------|-------------------------|-------|------------|-------|---------------|
| | C=1 | | D=1 | | |
| | Stime (a) | % | Stime (b) | % | diff. % (b-a) |
| OCC | 18,952,654 | 87.9 | 18,693,017 | 95.6 | -1.4 |
| DIS | 340,014 | 1.6 | 191,291 | 1.0 | -43.7 |
| NFL zona grigia | 483,173 | 2.2 | 230,638 | 1.2 | -52.3 |
| NFL altri | 1,779,631 | 8.3 | 444,806 | 2.3 | -75.0 |
| Totale | 21,555,471 | 100.0 | 19,559,752 | 100.0 | -9.3 |
| Totale non OCC | 2,602,817 | 12.1 | 866,735 | 4.4 | -66.7 |

La stima della sovracopertura agisce in maniera differenziata a seconda della fonte. Nel caso di INPD e EMEN l'abbattimento del segnale di primo livello riguarda una parte esigua delle osservazioni (Tavola 9). Ben diverso, al contrario, è il caso di INDI che vede abbattuto oltre un quinto dei segnali di copertura di primo livello. In posizione intermedia, le fonti PARA, AUAG, DMAG e ENPA mostrano un'incidenza degli abbattimenti superiore al 10%³⁵.

Tavola 9. Individui stimati con LFS, per presenza e coperture di primo e secondo livello per fonte. Anno 2010

| | di cui: | | | | | | |
|---------------|------------|---------------|---------------|-------|---------------|-------|-------------|
| | Presenza | Solo presenza | Copertura C=1 | % (a) | Copertura D=1 | % (a) | diff. % (b) |
| Fonti dirette | | | | | | | |
| EMEN | 13,173,478 | 2,258,566 | 10,914,912 | 82.9 | 10,688,699 | 81.1 | -2.1 |
| INDI | 9,082,378 | 1,710,179 | 7,372,199 | 81.2 | 5,764,450 | 63.5 | -21.8 |
| INPD | 3,294,550 | 234,488 | 3,060,062 | 92.9 | 3,039,342 | 92.3 | -0.7 |
| PARA | 1,285,319 | 513,982 | 771,336 | 60.0 | 675,392 | 52.5 | -12.4 |
| AUAG | 470,896 | 0 | 470,896 | 100.0 | 400,147 | 85.0 | -15.0 |
| DMAG | 847,663 | 434,638 | 413,025 | 48.7 | 358,909 | 42.3 | -13.1 |
| DOME | 578,707 | 122,745 | 455,962 | 78.8 | 425,586 | 73.5 | -6.7 |
| PROF | 255,537 | 4,216 | 251,321 | 98.4 | 238,136 | 93.2 | -5.2 |
| ENPA | 256,195 | 121,239 | 134,956 | 52.7 | 117,008 | 45.7 | -13.3 |
| Inail | | | | | | | |
| INAD | 16,395,332 | 2,128,660 | 14,266,672 | 87.0 | | | |
| INAP | 1,167,847 | 490,783 | 677,064 | 58.0 | | | |
| INAI | 436,600 | 288,602 | 147,997 | 33.9 | | | |

Note: (a) Percentuale rispetto al totale delle presenze; (b) diff.% rispetto alla copertura C=1

Nel complesso, l'indicatore di copertura di secondo livello annulla poco meno del 10% dei segnali di copertura di primo livello (Tavola 10). Questa azione si espleta tuttavia in misura variabile rispetto alla condizione lavorativa oggettiva misurata da LFS. Sugli occupati LFS l'abbattimento

³⁵ I segnali Inail non sono stati corretti poiché sono stati utilizzati come variabili ausiliarie per la correzione del segnale C=1.

delle coperture di primo livello incide relativamente molto poco (1,4%), pur riguardando comunque circa 260 mila unità. Cresce però di intensità man mano che ci si allontana dalla condizione di occupato: fra i disoccupati con esperienza lavorativa (DIS1 e DIS2) l'abbattimento del segnale di primo livello si aggira infatti intorno al 30% dei casi, è intorno al 50% per gli inattivi della zona grigia (NFL1 e NFL3) e supera abbondantemente il 60% per gli altri inattivi³⁶.

Tavola 10. Individui per condizione professionale per copertura nelle fonti amministrative

| Condizione | Presenza nelle fonti dirette | Con copertura di primo livello | | Con copertura di secondo livello | | |
|---------------|------------------------------|--------------------------------|-------------|----------------------------------|---------------|-------------|
| | | Totale (a) | % Inail (a) | Totale (b) | diff. % (b-a) | % Inail (a) |
| OCC | 19,893,238 | 18,952,654 | 63.2 | 18,693,017 | -1.4 | 63.7 |
| DIS1 | 585,201 | 198,380 | 35.5 | 135,913 | -31.5 | 46.8 |
| DIS2 | 216,259 | 76,868 | 42.3 | 55,378 | -28.0 | 51.5 |
| DIS3 | 164,621 | 64,766 | 41.2 | 0 | -100.0 | |
| NFL1 | 515,955 | 232,007 | 45.7 | 114,063 | -50.8 | 57.3 |
| NFL2 | 110,579 | 43,153 | 42.5 | 21,960 | -49.1 | 50.7 |
| NFL3 | 425,112 | 208,013 | 44.4 | 94,615 | -54.5 | 55.5 |
| NFL4 | 2,005,215 | 978,600 | 33.6 | 349,907 | -64.2 | 45.8 |
| NFL5 | 1,140,845 | 801,031 | 7.7 | 94,900 | -88.2 | 18.8 |
| Totale | 25,057,025 | 21,555,471 | 59.0 | 19,559,752 | -9.3 | 62.9 |
| OCC | 19,893,238 | 18,952,654 | 63.2 | 18,693,017 | -1.4 | 63.7 |
| DIS | 966,081 | 340,014 | 38.2 | 191,291 | -43.7 | 48.2 |
| NFL | 4,197,706 | 2,262,803 | 26.8 | 675,444 | -70.2 | 45.4 |

Note: (a) Percentuale rispetto al rispettivo totale delle coperture

Legenda: DIS=disoccupati

DIS1=DIS con esperienze ex occupati

DIS2=DIS con esperienze ex inattivi

DIS3=DIS senza esperienze

NFL1=NFL con 15-64, ricerca non attiva ma disponibili

NFL2=NFL con 15-64, cercano ma non disponibili

NFL3=NFL con 15-64, non cercano ma disponibili

NFL4=NFL con 15-64, non cercano non disponibili

NFL5= con 64 anni o più

La stessa analisi, condotta sulla condizione percepita, rivela abbattimenti del segnale di copertura più intensi – superiori al 70% - per casalinghe, studenti e pensionati, mentre alla categoria “Altro” corrisponde una quota di abbattimento sensibilmente più bassa rispetto al resto degli inattivi (Tavola 11).

³⁶ I segnali di copertura dei 65 mila disoccupati senza esperienze lavorative sono completamente annullati, ciò anche a causa del vincolo posto sulle persone che dichiarano di essere prive di esperienze lavorative. Inoltre, l'incidenza delle coperture Inail associate all'indicatore di secondo livello è in media più elevata rispetto alle coperture di primo livello (Tavola 10).

Tavola 11. Individui per condizione professionale percepita e copertura nelle fonti amministrative

| Condizione | Presenza nelle fonti dirette | Con Copertura C=1 | | Con Copertura D=1 | | |
|------------------------|------------------------------|-------------------|-------------|-------------------|------------------------|-------------|
| | | Totale | % Inail (a) | Totale | Diff. % rispetto a C=1 | % Inail (a) |
| OCC | 19,599,581 | 18,732,519 | 63.2 | 18,492,948 | -1.3 | 63.7 |
| DIS cerca nuova occup. | 1,622,744 | 655,811 | 42.0 | 448,686 | -31.6 | 53.2 |
| DIS cerca prima occup. | 339,870 | 165,638 | 49.1 | 14,706 | -91.1 | 61.2 |
| Casalinga | 1,004,896 | 577,386 | 26.6 | 161,697 | -72.0 | 0.0 |
| Studente | 624,546 | 216,451 | 54.4 | 59,438 | -72.5 | 68.2 |
| Ritirato | 1,513,945 | 979,802 | 14.1 | 239,800 | -75.5 | 30.7 |
| Inabile | 88,620 | 65,447 | 15.8 | 17,872 | -72.7 | 29.8 |
| Altro | 262,822 | 162,416 | 54.4 | 124,605 | -23.3 | 64.9 |
| Totale | 25,057,025 | 21,555,471 | 59.0 | 19,559,752 | -9.3 | 62.9 |
| OCC | 19,599,581 | 18,732,519 | 63.2 | 18,492,948 | -1.3 | 63.7 |
| DIS | 2,967,510 | 1,398,835 | 36.5 | 625,089 | -55.3 | 51.6 |
| NFL | 2,489,934 | 1,424,116 | 24.9 | 441,715 | -69.0 | 45.4 |

Note: (a) Percentuale rispetto al rispettivo totale delle coperture.

La composizione dei casi di incoerenza fra condizione lavorativa e indicatore di copertura D è per più versi dissimile da quella risultante dal primo livello, dal momento che la conferma del segnale di copertura ha agito in maniera differenziata (Tavola 12). È più elevata la porzione delle incoerenze riconducibili ai disoccupati con esperienze lavorative, la cui incidenza è molto alta nel caso di DMAG, ENPA, EMEN e INDI. La quota delle casalinghe con coperture scende in tutte le fonti (tranne PROF): nella fonte DOME continua comunque a dare conto di oltre metà delle incoerenze. La quota degli studenti si riduce vistosamente, quella dei pensionati cresce solo in EMEN e INPD (mentre ha una forte contrazione in INDI).

2.4.3. L'impatto della stima della sottocopertura degli occupati

Sulla base dell'indicatore di copertura di secondo livello D, la popolazione di riferimento LFS viene classificata rispetto alla condizione lavorativa come segue (la Tavola 13 è riferita al 2010: per l'anno 2011 i risultati sono analoghi).

Con riferimento al sottoinsieme degli intervistati LFS con coperture di secondo livello (D=1), viene stimato il rapporto di coesistenza dei non occupati rispetto agli occupati. A tal fine, gli occupati LFS sono individuati come l'insieme unione degli individui occupati secondo la condizione lavorativa oggettiva e percepita. La stima del coefficiente viene effettuata per strati, definiti dall'incrocio delle seguenti variabili strutturali LFS:

- ✓ TRIM. Il trimestre;
- ✓ RIP5. Le cinque ripartizioni territoriali;
- ✓ CLETA2. Una riclassificazione in tre classi di età: meno di 35, 35-64, 65 e oltre;
- ✓ SG11. Il genere;
- ✓ CITTAD. La cittadinanza: italiana, Ue, extra-Ue;
- ✓ PROXY. Il tipo di intervista: proxy, diretta.

Tavola 12. Individui non occupati stimati da LFS, per livello di copertura, condizione percepita e fonte amministrativa

| Condizione percepita | Livelli di copertura | | | | | | | | | |
|------------------------|----------------------|-------|-------------|-------|-------------|-------|-------------|-------|-------------|-------|
| | C=1 | D=1 | C=1 | D=1 | C=1 | D=1 | C=1 | D=1 | C=1 | D=1 |
| | <i>EMEN</i> | | <i>INDI</i> | | <i>INPD</i> | | <i>PARA</i> | | <i>AUAG</i> | |
| OCC | 5.3 | 9.2 | 0.2 | 0.6 | 12.5 | 20.6 | 0.2 | 0.2 | 0.1 | 0.2 |
| DIS cerca nuova occup. | 32.4 | 48.9 | 19.5 | 42.5 | 23.5 | 33.9 | 19.1 | 33.9 | 5.6 | 9.3 |
| DIS cerca prima occup. | 13.9 | 2.0 | 4.3 | 0.9 | 5.8 | 0.3 | 8.6 | 1.4 | 3.0 | 0.9 |
| Casalinga | 15.0 | 11.8 | 21.4 | 15.7 | 28.3 | 11.4 | 10.5 | 14.2 | 39.7 | 34.9 |
| Studente | 15.9 | 4.3 | 3.8 | 2.0 | 5.1 | 0.1 | 19.5 | 6.7 | 4.1 | 2.7 |
| Ritirato | 10.7 | 14.9 | 45.2 | 31.1 | 14.1 | 21.3 | 39.1 | 40.1 | 44.3 | 48.6 |
| Inabile | 1.5 | 1.4 | 3.0 | 2.9 | 2.7 | 2.0 | 0.9 | 0.5 | 1.4 | 1.9 |
| Altro | 5.3 | 7.5 | 2.6 | 4.3 | 8.0 | 10.5 | 2.1 | 3.0 | 1.7 | 1.4 |
| Totale | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | <i>DMAG</i> | | <i>DOME</i> | | <i>PROF</i> | | <i>ENPA</i> | | | |
| OCC | 0.4 | 0.4 | 0.0 | 0.0 | 2.2 | 3.5 | 2.9 | 10.6 | | |
| DIS cerca nuova occup. | 43.3 | 63.2 | 22.3 | 38.9 | 29.2 | 36.4 | 30.2 | 59.6 | | |
| DIS cerca prima occup. | 5.4 | 0.6 | 5.2 | 0.0 | 7.7 | 0.1 | 12.4 | 2.1 | | |
| Casalinga | 29.4 | 22.3 | 62.1 | 52.0 | 6.4 | 8.1 | 4.3 | 7.9 | | |
| Studente | 4.6 | 1.8 | 2.7 | 0.4 | 12.0 | 3.4 | 40.4 | 14.6 | | |
| Ritirato | 13.7 | 8.2 | 5.7 | 6.8 | 40.5 | 45.2 | 6.6 | 4.8 | | |
| Inabile | 1.6 | 1.5 | 0.6 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | | |
| Altro | 1.6 | 1.9 | 1.4 | 1.9 | 2.0 | 3.3 | 3.2 | 0.4 | | |
| Totale | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 | | |

Tavola 13. Individui con copertura per condizione rilevata con LFS. Anno 2010

| | | D | | |
|-----|--------|------------|------------|------------|
| | | 1 | 0 | Totale |
| LFS | OCC | 18,693,017 | 3,026,052 | 21,719,069 |
| | NOCC | 866,735 | 26,035,480 | 26,902,215 |
| | Totale | 19,559,752 | 29,061,532 | 48,621,284 |

La stima del rapporto di coesistenza rivela la presenza di circa 4,5 individui non occupati con copertura di secondo livello ogni 100 occupati con copertura. In particolare, le variabili di stratificazione indicano che:

- ✓ il trimestre ha un effetto trascurabile;
- ✓ il mezzogiorno ha un rapporto di coesistenza decisamente più elevato;
- ✓ la classe di età più anziana mostra un rapporto dieci volte più elevato della media;
- ✓ il genere non mostra effetti apprezzabili;
- ✓ la cittadinanza evidenzia un valore più elevato per i cittadini extra-Ue;
- ✓ l'effetto di sottocopertura delle interviste proxy è apprezzabile.

È da notare che, data la numerosità delle osservazioni, tutte le stime hanno un intervallo di confidenza strettissimo.

Tavola 14. Popolazione con copertura di secondo livello per condizione lavorativa RCFL (ammontare e rapporto di coesistenza)

| Variabili | OCC | NOCC | Rapp.coes. |
|----------------|------------|---------|------------|
| | a | b | b/a% |
| Totale | 18,725,869 | 833,883 | 4.5 |
| TRIMESTRE | | | |
| T1 | 18,444,800 | 808,882 | 4.4 |
| T2 | 18,753,980 | 825,705 | 4.4 |
| T3 | 18,729,019 | 850,802 | 4.5 |
| T4 | 18,975,676 | 850,144 | 4.5 |
| RIPARTIZIONE | | | |
| NO | 5,672,788 | 200,213 | 3.5 |
| NE | 4,377,055 | 151,530 | 3.5 |
| CE | 3,935,316 | 152,036 | 3.9 |
| SU | 3,173,885 | 238,660 | 7.5 |
| IS | 1,566,825 | 91,443 | 5.8 |
| CLASSE DI ETA' | | | |
| fino a 34 | 4,727,564 | 208,716 | 4.4 |
| 35-64 | 13,764,860 | 530,489 | 3.9 |
| 65 e oltre | 233,445 | 94,678 | 40.6 |
| GENERE | | | |
| Maschi | 11,230,024 | 502,588 | 4.5 |
| Femmine | 7,495,845 | 331,295 | 4.4 |
| CITTADINANZA | | | |
| Italiani | 17,294,996 | 763,737 | 4.4 |
| Ue | 471,060 | 19,763 | 4.2 |
| Extra-Ue | 959,814 | 50,384 | 5.2 |
| INTERVISTA | | | |
| diretta | 15,905,551 | 694,973 | 4.4 |
| proxy | 2,820,318 | 138,910 | 4.9 |

Note: (a) Include i circa 30 mila individui occupati secondo la condizione percepita

Agendo sul sottoinsieme privo di segnale di copertura di secondo livello ($D=0$), si stima per ciascun individuo la probabilità di risultare occupato (non regolare) in funzione di alcune variabili strutturali LFS (variabili R), di alcune variabili relative alla presenza e alla copertura di primo livello nelle fonti amministrative (variabili A) e di alcune interazioni (variabili M). Il modello adottato è quello logistico:

$$\log it(P_{j02}) = \beta_0 + \sum_{v=1}^r \beta_v^R R_{vj} + \sum_{v=1}^a \beta_v^A A_{vj} + \sum_{v=1}^m \beta_v M_{vj} \quad <25>$$

Si adotta come variabile risposta la variabile OCC, creata ponendone il valore uguale a uno se l'individuo risulta occupato secondo la condizione oggettiva ($COND3=1$) o percepita ($I1=1$):

$$OCC_j = \begin{cases} 1 \Leftarrow j \in \{j: COND3_j = 1\} \cup \{j: I1 = 1\} \\ 2 \Leftarrow j \notin \{j: COND3_j = 1\} \cup \{j: I1 = 1\} \end{cases}$$

Le variabili strutturali LFS utilizzate nel modello sono le seguenti:

- ✓ CITTAD. Cittadinanza (tre modalità);
- ✓ SG11. Genere (due modalità);
- ✓ REG. Regione (20 modalità);
- ✓ STACIV. Stato civile (cinque modalità);
- ✓ I5. Condizione lavorativa percepita l'anno precedente (7 modalità);
- ✓ CLETAD. Classe di età (7 modalità);
- ✓ ESPLAV. Esperienza lavorativa (due modalità: pari a '1' se occupato).

Le variabili ricavate dalle fonti amministrative sono invece le seguenti:

- ✓ PRESI, PRESI. Rispettivamente presenza/assenza nelle nove fonti dirette e in quelle Inail;
- ✓ COP1, COPI. Rispettivamente copertura/non copertura nelle fonti dirette e in quelle Inail;
- ✓ INAD, INAP, INAI. Rispettivamente copertura di primo livello (1), presenza (0) o mancata presenza (99) nelle singole fonti Inail;
- ✓ EMEN, INDI, DMAG, PARA, PROF, ENPA, INPD, DOME. Copertura di primo livello (1), presenza (0) o mancata presenza (99) nelle singole fonti dirette.

Sono inoltre state considerate alcune interazioni. In particolare:

- ✓ Interazioni fra variabili R: CITTAD*SG11*STACIV*CLETAD, CITTAD*SG11, CITTAD*I5, CITTAD*STACIV, I5*CLETAD, SG11*CLETAD, SG11*I5, SG11*STACIV, STACIV*CLETAD, I5*RIP5;
- ✓ Interazioni fra variabili A: COP1*COPI;
- ✓ Interazioni fra variabili A e variabili R: COP1*CLETAD, COPI*CLETAD, COPI*RIP5, I5*COPI, PRESI*CLETAD, PRESI*CLETAD, SG11*COPI.

L'attribuzione della condizione di occupato non regolare, sulla base delle stime per strato del rapporto di coesistenza:

- ✓ viene stimata per strato il totale dei non occupati privi di segnali che confluiscono negli occupati irregolari;
- ✓ l'attribuzione a questa categoria delle singole osservazioni per strato viene effettuata sfruttando l'ordinamento delle probabilità di occupazione individuali stimate attraverso la regressione logistica.

Sono state così stimate 134 mila transizioni verso la condizione di occupato, corrispondenti allo 0,5% della popolazione non occupata priva di segnali amministrativi (Tavola 15). Questa incidenza è più elevata:

- ✓ presso i cittadini stranieri,
- ✓ nelle classi di età intermedie,
- ✓ fra i disoccupati con esperienze lavorative e fra gli inattivi dell'area grigia,
- ✓ fra i maschi,
- ✓ fra quanti si percepiscono disoccupati e fra quanti si percepivano occupati l'anno precedente,
- ✓ nelle regioni del meridione.

Tavola 15. Incidenza delle transizioni verso la condizione di occupato sul totale dei non occupati privi di segnali di copertura.

| Variabile | Incidenza | Variabile | Incidenza | Variabile | Incidenza |
|-----------|-----------|-----------|-----------|-----------|-----------|
| CITTAD | | GENERE | | REG | |
| Italiana | 0.5 | M | 0.8 | PIE | 0.5 |
| Ue | 1.1 | F | 0.3 | VDA | 0.7 |
| Extra-Ue | 0.8 | I1 | | LOM | 0.4 |
| CLETAD | | DIS nuova | 3.4 | TAA | 0.6 |
| cletad=2 | 0.2 | DIS prima | 0.2 | VEN | 0.1 |
| cletad=3 | 1.7 | Casal | 0.1 | FVG | 0.4 |
| cletad=4 | 1.5 | Stud | 0.0 | LIG | 0.3 |
| cletad=5 | 1.3 | Ritirati | 0.3 | EMR | 0.4 |
| cletad=6 | 0.1 | Inabili | 0.0 | TOS | 0.3 |
| cletad=7 | 0.3 | Altro | 1.4 | MAR | 0.4 |
| cletad=8 | 0.1 | I5 | | UMB | 0.3 |
| COND 10 | | OCC | 13.4 | LAZ | 0.5 |
| DIS1 | 4.7 | DIS nuova | 0.1 | ABR | 0.5 |
| DIS2 | 1.9 | DIS prima | 0.2 | MOL | 0.6 |
| DIS3 | 0.0 | Casal | 0.0 | CAM | 0.9 |
| NFL1 | 2.0 | Stud | 0.0 | PUG | 0.8 |
| NFL2 | 1.8 | Ritirati | 0.0 | BAS | 0.5 |
| NFL3 | 1.0 | Inabili | 0.0 | CAL | 1.2 |
| NFL4 | 0.2 | Altro | 0.2 | SIC | 0.6 |
| NFL5 | 0.2 | | | SAR | 0.5 |
| ESPLAV | | | | | |
| SI | 0.8 | | | | |
| NO | 0.0 | | | | |

2.4.4. Le stime dell'occupazione di CN: descrizione dei risultati principali

La Tavola 16 riporta relativamente a ciascun segmento di occupazione il numero di osservazioni campionarie e le corrispondenti stime separatamente per il 2010 e il 2011, nonché la variazione percentuale fra i due anni del numero di osservazioni e delle stime. Nell'ultima colonna viene inoltre riportata la tipologia di occupato corrispondente a ciascun segmento, ovvero se regolare o parzialmente irregolare o interamente irregolare: si tratta di un indicatore qualitativo attribuito dal GdL stesso a scopo meramente descrittivo.

Tavola 16. Occupati di CN per segmento. Anni 2010 e 2011 (dati provvisori; numero di osservazioni LFS e stime)

| Segmenti | Anno 2010 | | Anno 2011 | | Var. % | | Tipo |
|---|-----------|------------|-----------|------------|--------|-------|--------------|
| | Oss. | Stime | Oss. | Stime | Oss. | Stime | |
| 1 Occupati LFS con copertura nelle fonti dirette, di cui: | 205.796 | 19.773.462 | 201.060 | 19.932.273 | -2,3 | 0,8 | Regol. |
| 1,1 con copertura Inail | 125.082 | 12.131.768 | 119.768 | 11.969.440 | -4,2 | -1,3 | Regol. |
| 1,2 privi di copertura Inail | 70.837 | 6.677.950 | 71.871 | 7.023.858 | 1,5 | 5,2 | Regol. |
| 1,9 con CF imputati | 9.877 | 963.744 | 9.421 | 938.975 | -4,6 | -2,6 | Regol. |
| 2 Occupati LFS senza copertura, di cui: | 25.047 | 2.432.114 | 24.318 | 2.369.300 | -2,9 | -2,6 | N/Regol. |
| 2,1 con copertura non validata | 4.172 | 343.196 | 4.250 | 355.235 | 1,9 | 3,5 | Parz. |
| 2,2 con sola copertura Inail | 5.462 | 530.282 | 5.632 | 554.608 | 3,1 | 4,6 | Parz. |
| 2,3 con presenza Inail + fonti dirette | 3.067 | 303.034 | 2.896 | 278.064 | -5,6 | -8,2 | Inter. |
| 2,4 con sola presenza Inail | 132 | 11.899 | 114 | 10.226 | -13,6 | -14,1 | Inter. |
| 2,5 con presenza solo nelle fonti dirette | 2.321 | 237.081 | 2.384 | 252.104 | 2,7 | 6,3 | Inter. |
| 2,6 non presenti in alcuna fonte | 8.570 | 870.813 | 7.852 | 801.854 | -8,4 | -7,9 | Inter. |
| 2,9 con CF imputati | 1.323 | 135.810 | 1.190 | 117.209 | -10,1 | -13,7 | Parz./Inter. |
| 3 Non occupati LFS con copertura fonti dirette, di cui: | 12.187 | 1.098.400 | 11.641 | 1.077.773 | -4,5 | -1,9 | Regol. |
| 3,1 con copertura Inail | 5.817 | 543.601 | 5.498 | 519.862 | -5,5 | -4,4 | Regol. |
| 3,2 privi di copertura Inail | 5.601 | 482.261 | 5.454 | 493.260 | -2,6 | 2,3 | Regol. |
| 3,9 con CF imputati | 769 | 72.538 | 689 | 64.651 | -10,4 | -10,9 | Regol. |
| 4 Non occupati LFS occupati non regolari, di cui: | 1.736 | 155.370 | 1.731 | 142.288 | -0,3 | -8,4 | N/Regol. |
| 4,1 con copertura non validata | 456 | 39.956 | 434 | 34.200 | -4,8 | -14,4 | Parz. |
| 4,2 con sola copertura Inail | 423 | 36.341 | 441 | 38.698 | 4,3 | 6,5 | Parz. |
| 4,3 con presenza Inail + fonti dirette | 160 | 13.933 | 157 | 11.338 | -1,9 | -18,6 | Inter. |
| 4,4 con sola presenza Inail | 2 | 199 | 3 | 488 | 50 | 145,8 | Inter. |
| 4,5 con presenza solo nelle fonti dirette | 143 | 12.816 | 200 | 17.293 | 39,9 | 34,9 | Inter. |
| 4,6 non presenti in alcuna fonte | 443 | 42.112 | 396 | 31.344 | -10,6 | -25,6 | Inter. |
| 4,9 con CF imputati | 109 | 10.014 | 100 | 8.926 | -8,3 | -10,9 | Parz./Inter. |
| 5 Occupati CN (1 + 2 + 3 + 4), di cui: | 244.766 | 23.459.345 | 238.750 | 23.521.634 | -2,5 | 0,3 | |
| 5,1 occupati regolari (1 + 3), di cui: | 217.983 | 20.871.861 | 212.701 | 21.010.046 | -2,4 | 0,7 | Regol. |
| con copertura Inail | 130.899 | 12.675.369 | 125.266 | 12.489.302 | -4,3 | -1,5 | |
| 5,2 occupati non regolari (2 + 4), di cui: | 26.783 | 2.587.484 | 26.049 | 2.511.588 | -2,7 | -2,9 | N/Regol. |
| con copertura Inail | 5.885 | 566.623 | 6.073 | 593.306 | 3,2 | 4,7 | |
| 6 Occupati LFS (1 + 2) | 230.843 | 22.205.575 | 225.378 | 22.301.573 | -2,4 | 0,4 | |
| 7 Residuo (5 - 6) | 13.923 | 1.253.770 | 13.372 | 1.220.061 | -4 | -2,7 | |
| Quota non regolari (5.2 / 5) | 10,9% | 11,0% | 10,9% | 10,7% | | | |

Nota: Le stime corrispondono ai pesi censuari al lordo delle correzioni PES

Si può notare anzitutto che con il metodo fin qui proposto vengono stimati in ciascuno dei due anni circa 1,2 milioni di occupati in più rispetto a LFS: di questi, oltre un milione sono occupati con almeno una posizione lavorativa regolare e corrispondono agli intervistati con un segnale di copertura regolare che risultano però non occupati a LFS; altri 150 mila circa sono occupati di CN privi di posizioni regolari. Anch'essi risultano non occupati secondo LFS ma sono stati stimati e imputati sulla base della procedura di correzione della sottocopertura LFS dell'occupazione non regolare.

Gli occupati privi di posizioni regolari sono 2,5 milioni, pari all' 11% circa del totale degli occu-

pati di CN. Si tratta di 2,4 milioni di occupati LFS privi di posizioni regolari, di cui circa 350-400 mila con copertura nelle fonti dirette non validate e poco meno di un milione assolutamente privi di alcun segnale amministrativo. A questi si aggiungono i 150 mila occupati non regolari stimati con la correzione della sottocopertura. Gli occupati "non regolari" privi di segnali amministrativi sono dunque circa un milione; quelli con copertura di primo livello non validate sono circa 400 mila; quelli con copertura Inail sono circa 600 mila.

3. La stima delle posizioni lavorative

3.0 Sintesi

La stima delle posizioni lavorative ha avuto come obiettivo la creazione, a partire dal campione integrato LFS-ADMIN, di un archivio delle posizioni lavorative regolari e non regolari associate agli intervistati LFS (archivio LFS-ADMIN_JOBS)

Per ogni posizione lavorativa di ciascun individuo del campione integrato è stato prodotto un vettore di informazioni contenente la gerarchia della posizione lavorativa (principale, secondaria, ecc.), la natura regolare o non regolare, la natura dipendente o indipendente, il codice di attività economica, la regione in cui la posizione lavorativa viene ricoperta, la classe dimensionale del committente, il settore istituzionale del committente.

La singola posizione regolare viene in particolare identificata dalla coppia individuo-datore di lavoro che nella settimana di riferimento dell'intervista LFS ha un segnale amministrativo validato. Il numero complessivo degli occupati e delle posizioni regolari nel campione integrato LFS-ADMIN è pertanto un dato, come del resto costituisce un dato la natura regolare o non regolare di ogni singolo occupato.

Per descrivere la posizione lavorativa regolare sono state agganciate le informazioni di tipo amministrativo a essa relative: codice Ateco, tipo di posizione, classe di addetti, forma giuridica e settore istituzionale del datore. La regione di lavoro è invece stata dedotta da LFS.

Il numero delle posizioni lavorative non regolari è oggetto di stima e viene derivato dal confronto condotto a livello individuale fra il numero di posizioni regolari e il numero di posizioni rilevate nell'intervista LFS, correggendo la sottodichiarazione delle posizioni lavorative secondarie da parte degli intervistati LFS.

La correzione della sottostima della componente non regolare è basata sull'ipotesi che la propensione a nascondere una seconda posizione regolare sia uguale alla propensione a non dichiarare a LFS una seconda posizione non regolare.

Le caratteristiche delle posizioni non regolari (codice ateco, tipo di posizione e regione) sono derivate principalmente da LFS, sebbene siano state utilizzate anche le informazioni relative alle coperture di primo livello non validate. In mancanza di tali informazioni, le caratteristiche sono state imputate tramite donatore.

Il metodo di attribuzione dell'ordine della posizione differisce a seconda del numero delle posizioni regolari dell'individuo e del numero di posizioni che l'individuo stesso ricopre secondo la rilevazione LFS. In generale è attribuito in base a un indicatore di similarità con la posizione principale dichiarata all'indagine.

Nel complesso il metodo stima nel 2010 circa 26 milioni di posizioni lavorative di cui il 10.4% non regolari e il 9.6% secondarie. Il tasso di irregolarità delle posizioni secondarie è più elevato di circa 4-5 punti percentuali. Nell'agricoltura una posizione su cinque non è regolare, seguono costruzioni commercio e servizi alle famiglie. Nel 2011 la componente non regolare evidenzia una contrazione delle posizioni lavorative a fronte di una sostanziale tenuta della componente regolare.

Se si confronta con la stima delle posizioni lavorative di LFS, la nuova procedura ne stima oltre 3 milioni in più, equivalenti al 15% in più: due terzi della differenza sono dovuti alle posizioni secondarie il cui numero è più elevato di 6-7 volte rispetto alle posizioni conteggiabili con LFS. Agricoltura e costruzioni sono i settori dove più ha inciso il differenziale sulle seconde posizioni. Con riferimento alle posizioni principali la differenza con LFS è solo del 5% ed è molto ampia in

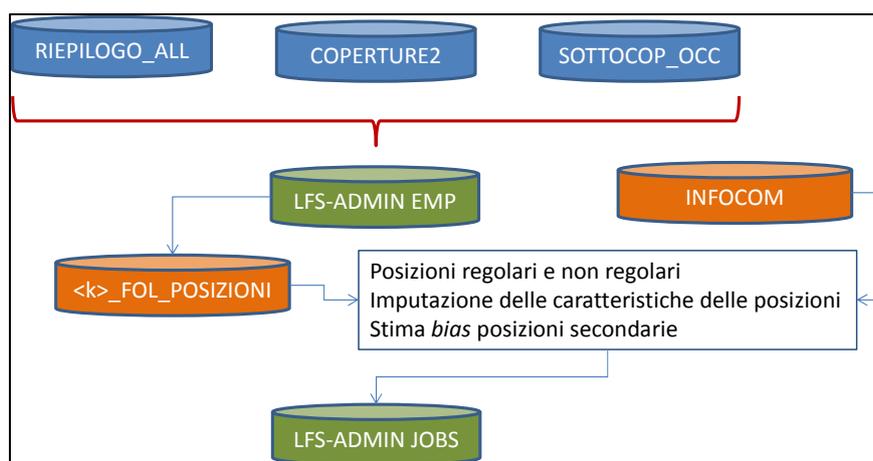
agricoltura (18%).

3.1. Aspetti generali

La stima delle posizioni lavorative è fondata sull'utilizzo della base dati integrata LFS-ADMIN³⁷ e in particolare utilizza i seguenti dataset (cfr. Figura 12):

- ✓ i tre dataset prodotti dalla procedura di stima dell'occupazione e descritti nella sezione 2, e in particolare quello che riepiloga le coperture amministrative di primo livello (RIEPILOGO_ALL), quello che reca gli individui con coperture validate (COPERTURE2) e il dataset che contiene gli individui classificati come occupati non regolari e recuperati attraverso la procedura di stima della sottodichiarazione LFS della condizione di occupato (SOTTOCOP_OCC). Ogni record corrisponde alla singola intervista LFS;
- ✓ il dataset LFS-ADMIN_EMP ottenuto dall'aggancio di questi tre data set, i cui record sono rappresentati dal singolo individuo del campione LFS classificato come occupato. Si tratta perciò di un dataset di dati individuali;
- ✓ i nove dataset relativi alle posizioni lavorative degli intervistati LFS tracciate nelle fonti amministrative (<fonte>_FOL_POSIZIONI);
- ✓ il dataset sui datori di lavoro ricavato dai registri statistici su imprese e istituzioni (INFOCOM);
- ✓ un dataset che costituisce l'output finale di questa procedura di stima (LFS-ADMIN JOBS) in cui il singolo record è costituito da una posizione lavorativa e che dunque si collega "uno a molti" con LFS-ADMIN_EMP.

Figura 12. Le basi dati del processo di stima delle posizioni lavorative



L'obiettivo è produrre un vettore di informazioni associate a ciascuna delle posizioni lavorative assunte dagli occupati di Contabilità nazionale (CN). In particolare, si vuole individuare:

- ✓ il numero delle posizioni lavorative afferenti a ciascun individuo del campione integrato LFS-ADMIN_EMP;
- ✓ la gerarchia delle posizioni lavorative individuali (principale, secondaria, ecc.);
- ✓ la natura regolare o non regolare della posizione lavorativa;
- ✓ la natura dipendente o indipendente della posizione;
- ✓ il codice di attività economica;
- ✓ la regione in cui la posizione lavorativa viene ricoperta;
- ✓ la classe dimensionale dell'unità presso cui la posizione è attiva;

³⁷ La metodologia può essere applicata con qualsiasi struttura di ponderazione: i dati che vengono via via presentati sono in particolare basati sulla popolazione censuaria.

- ✓ il settore istituzionale a cui l'unità appartiene.

Il lavoro descritto nella sezione 2 è il punto di partenza per la stima delle posizioni lavorative. Quest'ultima riguarda infatti tutti gli intervistati LFS classificati fra gli occupati nella procedura di stima dell'occupazione (che indicheremo per brevità occupazione di CN): si tratta pertanto sia degli occupati LFS sia dei non occupati LFS, riclassificati come occupati per le stime della contabilità nazionale. Inoltre, la eventuale qualifica di regolarità viene attribuita a tutte e sole le posizioni lavorative con un segnale di copertura validato nella settimana di riferimento della intervista LFS. Il numero complessivo e le stime individuali ricavate nella sezione 2 relativamente agli occupati (regolari e non) e al numero delle posizioni lavorative regolari nel campione integrato LFS-ADMIN costituiscono pertanto per questa fase successiva di stima un dato³⁸. Per contro, il numero effettivo delle posizioni lavorative non regolari è oggetto di stima e viene derivato dal confronto condotto a livello individuale fra il numero di posizioni regolari e il numero di posizioni rilevate nell'intervista LFS: occorre infatti misurare e tenere nel debito conto gli effetti generati dalla evidente sottodichiarazione delle posizioni lavorative secondarie da parte degli intervistati LFS.

L'insieme degli occupati di CN, precedentemente stimato, è stato pertanto suddiviso in due sottoinsiemi disgiunti: l'insieme O_1 degli occupati con almeno una posizione regolare, ai quali corrispondono solo posizioni regolari o, per parte di essi, anche delle posizioni lavorative non regolari; e l'insieme O_2 degli occupati privi di posizioni regolari, ai quali corrispondono esclusivamente posizioni lavorative non regolari.

3.2. La stima delle posizioni lavorative regolari

3.2.1. L'identificazione delle posizioni lavorative regolari

Si considera il solo sottoinsieme O_1 , relativo agli occupati di CN con almeno una posizione regolare in una delle fonti (indicando con k la generica fonte, questo insieme è identificato dalla condizione che $D_k = 1$ per almeno un k)³⁹. La singola posizione regolare viene in particolare identificata dalla coppia individuo-datore di lavoro che nella settimana di riferimento dell'intervista LFS ha un segnale amministrativo validato.

Le fonti sulle posizioni lavorative sono state distinte in due categorie:

- ✓ fonti sui dipendenti (EMEN, ENPA, DOME, DMAG, INPD) e
- ✓ fonti sugli indipendenti (INDI, AUAG, PROF, PARA).

Sono state così considerate tutte le posizioni lavorative distinte assunte dagli intervistati LFS nell'insieme delle fonti. In particolare, nel caso di agganci di una stessa relazione individuo-datore su più fonti di una stessa categoria (p.es.: EMEN e ENPA, oppure INDI e PARA) questa viene collassata in una singola posizione. Se l'aggancio riguarda però fonti di natura diversa (es.: INDI e EMEN) la relazione individuo-datore viene spezzata in due posizioni (l'obiettivo è lasciare traccia di questa promiscuità per poterne tenere conto nella stima delle ore lavorate).

3.2.2. Le informazioni disponibili

A ogni posizione lavorativa è stato pertanto assegnato:

- ✓ un vettore \mathbf{A} di variabili descrittive del rapporto di lavoro, desunto dalle nove fonti di-

³⁸ In questa particolare accezione, stiamo provvisoriamente definendo come "occupato regolare" un occupato con almeno una posizione lavorativa regolare a prescindere dal fatto se questa sia o meno la posizione principale. In termini di principio, non si tratta di occupati regolari in senso stretto, dal momento che per essere tali deve avere caratteristiche di regolarità la posizione principale. Come si vedrà, il numero di occupati con posizione principale non regolare e posizioni secondarie regolari è decisamente esiguo.

³⁹ Il flag D_k è stato definito nel quadro della procedura per la stima degli occupati regolari e non regolari di CN.

- rette a cui sono stati agganciati i microdati LFS;
- ✓ un vettore \mathbf{C} di variabili relative al datore di lavoro, contenenti queste ultime il codice di attività economica, il numero degli addetti, la forma giuridica, la provincia. Le informazioni sul datore di lavoro sono state reperite ricorrendo al contenuto dei registri statistici dell'Istat: Asia Imprese, Asia Agricoltura, S13, Inpdap, ISP, Istituzioni pubbliche non S13 e potenziali S13, archivio Silo_I;
- ✓ un vettore \mathbf{R} di variabili relative all'individuo raccolte tramite la rilevazione LFS.

Indicando con i un generico occupato con posizione regolare ($i \in O_1$), con j il generico datore di lavoro e con P_{ij} il vettore di variabili che descrivono la posizione lavorativa dell'occupato i presso il datore j , si avrà

$$P_{ij} = [R_i; A_{ij}; C_j] \quad \langle 1 \rangle$$

Le variabili contenute nel vettore R_i e utilizzate nella stima sono le seguenti (si omettono i suffissi i e j):

- ✓ il numero di posizioni lavorative (si tratta della variabile D1 di LFS, tradotta nella variabile $p^R = 0,1,2,> 2$);
- ✓ il codice Ateco della posizione principale e dell'eventuale secondaria (ATE^{R1} e ATE^{R2});
- ✓ la tipologia (dipendente/indipendente) della posizione principale e dell'eventuale secondaria ($DIPIND^{R1}$ e $DIPIND^{R2}$);
- ✓ il comune in cui lavora, da cui è stata estratta la regione ($REGIO^R$);
- ✓ le rimanenti variabili del questionario LFS.

Ovviamente, per gli individui che non risultano occupati a LFS le informazioni sulle posizioni lavorative rese a LFS sono mancanti.

Le variabili contenute nel vettore A_{ij} variano a seconda delle fonti agganciate con la posizione P_{ij} e riguardano:

- ✓ il numero di posizioni lavorative regolari ($p^A = 0,1,2,3,\dots$);
- ✓ la tipologia dell'occupazione ($DIPIND^A$) ricavata sulla base della fonte amministrativa di provenienza sulla base del criterio esplicitato in precedenza;
- ✓ l'intensità in termini di durata e copertura contrattuale del rapporto di lavoro nella settimana di riferimento, nel trimestre di riferimento e nell'anno;
- ✓ gli importi imponibili nel trimestre e nell'anno;
- ✓ altre caratteristiche, quali le tipologie contrattuali, part-time/full-time, ecc.

Le variabili contenute nel vettore C_j riguardano il datore di lavoro e in particolare:

- ✓ il codice Ateco (ATE^C);
- ✓ la classe dimensionale in termini di addetti ($CLAD^C$);
- ✓ la forma giuridica (FG^C);

- ✓ il settore istituzionale ($SIST^C$)⁴⁰;
- ✓ altre variabili contenute nei registri statistici.

3.2.3. L'ordinamento delle posizioni regolari

Il grosso delle informazioni contenute in A_{ij} viene utilizzato per determinare la gerarchia delle posizioni regolari. Sia $P_i = \{P_{ij}, j = 1, \dots, a_i\}$ l'insieme delle a_i posizioni lavorative regolari assunte dall'individuo i : per gran parte degli intervistati LFS $a_i = 1$, tuttavia un po' meno del 10% di essi presenta due o più posizioni regolari. Per gli individui con posizioni plurime è necessario pertanto produrre un ordinamento di queste allo scopo di individuare in un secondo momento le posizioni principali e quelle secondarie. A tale scopo sono stati adottati due criteri di ordinamento:

1. ordinamento nelle fonti. Se sono presenti più posizioni registrate in una medesima fonte (ad esempio EMENS) queste sono state ordinate sulla base di una gerarchia di variabili:
 - a. i giorni di copertura nella settimana di riferimento;
 - b. gli imponibili nel trimestre di riferimento;
 - c. le giornate di copertura nel trimestre di riferimento;
 - d. gli imponibili nell'anno;
 - e. le giornate di copertura nell'anno.
2. ordinamento fra le fonti. Se sono presenti più posizioni registrate in fonti diverse (ad esempio EMENS e INPDAP), queste sono state ordinate sulla base delle seguenti gerarchie di fonti:
 - a. INPD, EMEN, DMAG, DOME, ENPA per quanto riguarda le fonti sui dipendenti;
 - b. INDI, PROF, PARA, AUAG per le fonti sugli indipendenti.

Ovviamente il criterio 2 prevale sul criterio 1. Senza modificare la notazione, si assumerà per semplicità che l'indice j indichi anche la gerarchia delle posizioni regolari.

3.2.4. L'assegnazione delle posizioni principali e secondarie

A ogni posizione lavorativa regolare è stato attribuito il connotato di posizione principale oppure secondaria nell'ambito del complesso delle posizioni, regolari e non, ricoperte da ciascun occupato di CN. Per questa operazione di stima sono state usate congiuntamente tutte le informazioni contenute nei vettori P_{ij} . Conviene tuttavia tenere distinti i casi degli individui con una o con più posizioni regolari.

In generale, per descrivere la posizione lavorativa regolare sono state utilizzate le informazioni di tipo amministrativo a essa relative. In particolare:

$$\left\{ \begin{array}{l} ATE = ATE^C \\ DIPIND = DIPIND^C \\ CLAD = CLAD^C \\ FG = FG^C \\ SIST = SIST^C \\ REGIO = REGIO^R \end{array} \right. \quad \langle 2 \rangle$$

⁴⁰ Per l'assegnazione di settore istituzionale, codice Ateco, forma giuridica, classe di addetti e regione è stata creata una scala di priorità nelle fonti sui datori di lavoro: S13, Istituzioni pubbliche non S13 candidate e non S13 tout court, ISP, altri registri.

Nell'attribuzione dell'ordine della posizione il criterio differisce a seconda del numero delle posizioni regolari dell'individuo e del numero di posizioni che l'individuo stesso ricopre secondo la rilevazione LFS. Negli altri casi (e in particolare in presenza di non occupati LFS), le informazioni descrittive della posizione sono state imputate con altri metodi (si veda più oltre).

Nel caso delle posizioni lavorative regolari associate a individui con una sola posizione regolare⁴¹, alla posizione regolare viene generalmente attribuita la qualifica di posizione principale, identificata dalla variabile

$$ORD = 1 \quad <3>$$

Questa soluzione generale viene derogata in alcuni casi:

- ✓ Nei (pochi) casi in cui non è disponibile il codice Ateco del datore di lavoro e dalla rilevazione LFS l'individuo risulta avere una sola posizione lavorativa ($p^R = 1$) il codice stesso viene imputato:

$$ATE = ATE^{R1} \quad <4>$$

- ✓ Per gli individui con almeno due posizioni LFS ($p^R > 1$), vengono calcolati i due indicatori di similarità (s^{R1}, s^{R2}) fra la posizione regolare e le due posizioni LFS che misurano il numero di digit coincidenti dei codici Ateco e la coincidenza (o meno) della tipologia delle posizioni (dipendente/indipendente). L'ordine della posizione regolare viene scelto come quello con un indice di similarità più elevato. In caso di eguaglianza si opta per la posizione principale:

$$\begin{cases} s^{R1} \geq s^{R2} \Rightarrow ORD = 1 \\ s^{R1} < s^{R2} \Rightarrow ORD = 2 \end{cases} \quad <5>$$

Nei casi in cui non è disponibile il codice Ateco del datore di lavoro, quest'ultimo viene imputato sulla base dei codici rilevati tramite LFS:

$$\begin{aligned} se \quad ORD = 1 &\Rightarrow ATECO = ATECO^{R1} \\ se \quad ORD = 2 &\Rightarrow ATECO = ATECO^{R2} \end{aligned} \quad <6>$$

- ✓ Se l'individuo non è occupato secondo LFS e non è disponibile il codice Ateco del datore di lavoro quest'ultimo viene imputato tramite donatore, selezionato casualmente all'interno di domini definiti dall'incrocio fra tipologia di occupazione regolare, fonti

⁴¹ Cfr. la procedura AR1 del programma SAS che esegue la stima.

agganciate, genere, ripartizione geografica e tre classi di età⁴²:

$$ATECO = ATECO^{ST} \quad <7>$$

- ✓ Se la forma giuridica del datore di lavoro è mancante, questa viene imputata tramite donatore selezionato casualmente all'interno di domini definiti dall'incrocio fra tipologia di occupazione (dipendente o indipendente), tipo di codice fiscale del datore (alfanumerico o numerico) e tre classi dimensionali del committente⁴³.

$$FG = FG^{ST} \quad <8>$$

- ✓ Per gli individui non occupati LFS, la variabile REGIO è stata imputata con la regione di residenza.

Lo schema generale di questa sequenza del processo di imputazione è schematizzata nella Tavola 17.

Tavola 17. Schema dell'imputazione delle informazioni sulla posizione lavorativa per gli individui con una sola posizione regolare

| Disponibilità Ateco Admin | Posizioni LFS | Stima Ateco | Ordine posizione |
|------------------------------|------------------|-----------------------------|---|
| SI | 0 | Imputazione Ateco Admin | Principale |
| SI | 1 | Imputazione Ateco Admin | Principale |
| SI | >1 | Imputazione Ateco Admin | Principale o secondaria: criterio di prossimità |
| NO | 0 | Ateco due-digit da donatore | Principale |
| NO | 1 | Imputazione Ateco LFS | Principale |
| NO | >1 | Imputazione Ateco LFS | Principale o secondaria: criterio di prossimità |

Si consideri adesso il caso delle posizioni regolari associate a individui con più di una posizione regolare⁴⁴. Per ognuna delle $a_i > 1$ posizioni regolari viene calcolato un indice di distanza con tutte le posizioni rilevate tramite LFS. La procedura è identica a quella esposta in precedenza e prende in considerazione i codici Ateco, la tipologia occupazionale e la gerarchia delle posizioni. Si attribui-

⁴² La scelta del donatore viene effettuata tramite una procedura *hot-deck* sequenziale. I domini così definiti sono articolati gerarchicamente verso partizioni via via meno complesse onde assicurare che il donatore venga scelto all'interno di un numero di potenziali donatori superiore a una soglia prefissata (almeno 20 unità). Per ciascuna imputazione viene adottata la partizione in domini più fine che assicura il superamento della soglia.

⁴³ La procedura di selezione del donatore e dei domini è identica a quella utilizzata per l'imputazione del codice di attività economica.

⁴⁴ Cfr. la procedura AR2 del programma SAS che esegue la stima.

sce alla posizione principale quella con un indice di similarità più elevato con la posizione principale LFS:

$$j' | s_{j'}^{R1} = \max_j (s_j^{R1}) \Rightarrow ORD_{j'} = 1 \quad <9>$$

Si attribuisce inoltre alla posizione secondaria la posizione j , diversa da j' , con l'indice di similarità più elevato con la posizione secondaria LFS:

$$j'' | s_{j''}^{R1} = \max_{j \neq j'} (s_j^{R1}) \Rightarrow ORD_{j''} = 2 \quad <10>$$

Alle restanti posizioni si attribuiscono i numeri di ordine successivi sulla base della gerarchia delle posizioni regolari definita dall'indice j .

L'attribuzione delle caratteristiche di ciascuna posizione è stata effettuata utilizzando le informazioni contenute nei registri statistici:

$$\left\{ \begin{array}{l} ATE = ATE^C \\ DIPIND = DIPIND^C \\ CLAD = CLAD^C \\ FG = FG^C \\ SIST = SIST^C \\ REGIO = REGIO^R \end{array} \right. \quad <11>$$

Le deroghe a questa soluzione hanno riguardato casi specifici:

- ✓ Il codice Ateco del registro era mancante. In questo caso per le prime due posizioni si è utilizzato il codice Ateco della corrispondenti posizioni principale o secondaria LFS

$$ATE_{j'} = ATE^{R1} \text{ o } ATE_{j''} = ATE^{R2} \quad <12>$$

- ✓ Per le posizioni successive e per i non occupati LFS è stato imputato il codice ateco tramite donatore sulla base della procedura esposta in precedenza:

$$ATE = ATE^{ST} \quad <13>$$

- ✓ Se la forma giuridica del datore di lavoro è mancante, questa è stata imputata tramite donatore sulla base della procedura esposta in precedenza:

$$FG = FG^{ST} \quad <14>$$

- ✓ Per gli individui non occupati LFS, la variabile REGIO è stata imputata con la regione di residenza desunta dal questionario LFS.

Sulla base di questo sistema, le posizioni regolari di ciascun individuo vengono numerate in ordine di importanza e assegnate come segue (Tavola 18):

Tavola 18. Schema dell'imputazione delle informazioni sulla posizione lavorativa per gli individui con più posizioni regolari

| Disponibilità Ateco Admin | Occupato LFS | Stima Ateco | Ordine posizione |
|------------------------------|-----------------|-----------------------------|--------------------------------|
| SI | NO | Imputazione Ateco Admin | Ordinamento posizioni regolari |
| | 1 POS. | Imputazione Ateco Admin | Criterio di prossimità |
| SI | 2 o più POS. | Imputazione Ateco Admin | Criterio di prossimità |
| NO | NO | Ateco due-digit da donatore | Ordinamento posizioni regolari |
| NO | 1 POS. | Imputazione Ateco LFS | Criterio di prossimità |
| NO | 2 o più POS. | Imputazione Ateco LFS | Criterio di prossimità |

3.3. Posizioni regolari e posizioni LFS: alcuni confronti

3.3.1. La sottostima LFS delle posizioni lavorative

Nella Tavola 19 sono riportate, con riferimento al solo 2010, le posizioni regolari misurate attraverso la base integrata LFS-ADMIN: esse corrispondono a 20,8 milioni di individui, di cui oltre un milione sono relative a non occupati LFS.

Vi è una diversità di fondo fra le due misurazioni possibili delle posizioni lavorative attraverso la base dati integrata LFS-ADMIN, a seconda che esse siano condotte sulla base delle informazioni LFS o sulla base di quelle ADMIN. Infatti, se si considera il solo sottoinsieme degli occupati LFS con almeno una posizione ADMIN, l'incidenza degli individui con più di una posizione è dell'8,8% se misurata tramite il conteggio delle posizioni ADMIN, mentre è dell'1,5% se misurata attraverso il conteggio delle posizioni LFS. Si tratta di un divario significativo, la cui entità si acuisce ulteriormente se si considera il fatto che la misurazione LFS è inclusiva delle posizioni non regolari.

Le ragioni di questo divario vanno evidentemente approfondite ulteriormente. Occorre tuttavia chiarire che le possibili giustificazioni che facciano leva sulla sottostima delle posizioni secondarie dovuta alle interviste proxy arrivano a spiegare un porzione marginale di quella differenza: basti pensare che, con riferimento alle sole interviste dirette, la percentuale di individui che indicano a LFS di avere più posizioni sale appena all'1,6%. Un effetto proxy certamente esiste, dal momento che fra questi intervistati la medesima incidenza è appena dell'1,1%, e va dunque tenuto in debita considerazione: tuttavia, la distanza nella stima delle posizioni va spiegata altrimenti. È piuttosto ragionevole concludere che LFS sottostima le posizioni lavorative. Se si parte dall'ipotesi che le posizioni lavorative validate con il dato amministrativo siano "vere", automaticamente emerge una sottostima da indagine con riferimento anche solo al sottoinsieme delle posizioni lavorative regolari. In particolare, secondo ADMIN vi sono 1,7 milioni di occupati LFS che hanno almeno due posizioni lavorative regolari, mentre secondo LFS il loro numero supererebbe di poco le 300 mila unità.

Tradotto in termini di posizioni lavorative, le sole posizioni regolari identificate attraverso la fonte amministrativa ammontano a 22,9 milioni, circa 2,8 milioni in più rispetto a quelle misurate attraverso la sola LFS, le quali includono però anche le posizioni non regolari. Assumere dunque la bontà delle posizioni regolari misurate attraverso le fonti amministrative da un lato implica l'ammissione di una significativa sottostima delle posizioni lavorative da parte di LFS e dall'altro suggerisce la necessità di misurare opportunamente la corrispondente sottostima delle posizioni secondarie non regolari da parte dell'indagine. Appare dunque necessario introdurre dei correttivi a

questa sottocopertura, i quali sono descritti nei paragrafi successivi.

Tavola 19. Intervistati LFS per numero di posizioni regolari e di posizioni LFS. Anno 2010

| Numero posizioni LFS | Numero posizioni regolari (ADMIN validate) | | | Totale |
|----------------------|--|-----------|---------------------------------|------------|
| | Una | Due | Tre o più | |
| | | | <i>Individui</i> | |
| Una | 17,882,186 | 1,434,087 | 156,992 | 19,473,264 |
| Due | 146,438 | 125,261 | 20,640 | 292,339 |
| Tre o più | 6,855 | 2,364 | 1,600 | 10,820 |
| Nessuna | 946,170 | 81,840 | 10,341 | 1,038,351 |
| Totale | 18,981,649 | 1,643,553 | 189,573 | 20,814,775 |
| | | | <i>Posizioni ADMIN validate</i> | |
| Una | 17,882,186 | 2,868,173 | 529,430 | 21,279,789 |
| Due | 146,438 | 250,523 | 64,807 | 461,767 |
| Tre o più | 6,855 | 4,729 | 5,499 | 17,083 |
| Nessuna | 946,170 | 163,681 | 35,087 | 1,144,938 |
| Totale | 18,981,649 | 3,287,106 | 634,823 | 22,903,578 |
| | | | <i>Posizioni LFS</i> | |
| Una | 17,882,186 | 1,434,087 | 156,992 | 19,473,264 |
| Due | 292,876 | 250,523 | 41,280 | 584,678 |
| Tre o più | 20,565 | 7,093 | 4,801 | 32,459 |
| Nessuna | | | | |
| Totale | 18,195,627 | 1,691,703 | 203,072 | 20,090,402 |

3.3.2. Il confronto fra le informazioni sull'attività economica

I risultati delle regole di imputazione adottate per le caratteristiche delle posizioni lavorative suggeriscono le seguenti osservazioni (Tavole 20 e 21):

- ✓ In due terzi dei casi il codice ateco LFS coincide con quello ADMIN al secondo digit. Si arriva al 75% se si considera la sezione;
- ✓ Per gli individui con una sola posizione regolare, i casi in cui questa venga identificata come una posizione secondaria sono molto esigui;
- ✓ L'imputazione del codice Ateco tramite LFS o tramite donatore ha una incidenza irrisoria;
- ✓ L'adattamento delle imputazioni sulle seconde posizioni ai codici di attività dichiarato a LFS è meno soddisfacente.

Tavola 20. Assegnazione del codice Ateco e della gerarchia della posizione regolare per gli individui con una sola posizione ADMIN (Anno 2010)

| Posizioni LFS | Imputazione Ateco | | Prima posizione | Seconda posizione | Totale | |
|-------------------|-------------------|--------------|-------------------|-------------------|-------------------|--------------|
| | Origine (a) | Cfr. A-R (b) | | | n | % |
| Una posizione LFS | A | 2 digit | 12,117,196 | | 12,117,196 | 63.8 |
| | A | sezione | 1,924,453 | | 1,924,453 | 10.1 |
| | A | tipologia | 3,578,438 | | 3,578,438 | 18.8 |
| | A | nessuna | 238,180 | | 238,180 | 1.3 |
| | R | | 26,106 | | 26,106 | 0.1 |
| Due posizioni LFS | A | 2 digit | 90,001 | 14,051 | 104,052 | 0.5 |
| | A | sezione | 9,130 | 1,007 | 10,137 | 0.1 |
| | A | tipologia | 28,704 | 1,797 | 30,501 | 0.2 |
| | A | nessuna | 1,283 | | 1,283 | 0.0 |
| | R | | 465 | | 465 | 0.0 |
| Tre posizioni LFS | A | 2 digit | 4,140 | 168 | 4,308 | 0.0 |
| | A | sezione | 301 | | 301 | 0.0 |
| | A | tipologia | 2,011 | | 2,011 | 0.0 |
| | A | nessuna | 149 | | 149 | 0.0 |
| | R | | 86 | | 86 | 0.0 |
| Non occupato LFS | A | | 937,916 | | 937,916 | 4.9 |
| | I | | 8,255 | | 8,255 | 0.0 |
| TOTALE | | | 18,966,813 | 17,024 | 18,983,836 | 100.0 |
| | | 2 digit | 12,211,337 | 14,220 | 12,225,557 | 64.4 |
| | A | sezione | 1,933,884 | 1,007 | 1,934,891 | 10.2 |
| | | tipologia | 3,609,153 | 1,797 | 3,610,949 | 19.0 |
| | | altri | 1,177,527 | 0 | 1,177,527 | 6.2 |
| | R | | 26,658 | 0 | 26,658 | 0.1 |
| | I | | 8,255 | 0 | 8,255 | 0.0 |

Note: (a) A=ADMIN, R=LFS, I=Imputazione ateco ADMIN per strato; (b) Livello di coincidenza fra ateco ADMIN e ateco LFS. Per tipologia si intende Dipendente/indipendente

Tavola 21. Assegnazione del codice Ateco e della gerarchia della posizione regolare per gli individui con più di una posizione ADMIN (Anno 2010)

| Posizioni LFS | Imputazione Ateco | | Prima posizione | | Seconda posizione | Ulteriori posizioni | Totale | |
|-----------------------|-------------------|--------------|-----------------|-------|-------------------|---------------------|-----------|-------|
| | Origine (a) | Cfr. A-R (b) | n | % | | | n | % |
| Una posizione LFS | A | 2 digit | 1,111,802 | 60.7 | 398,477 | 76,453 | 1,586,793 | 40.5 |
| | A | sezione | 142,885 | 7.8 | 106,290 | 12,052 | 261,234 | 6.7 |
| | A | tipologia | 314,255 | 17.2 | 493,221 | 79,329 | 886,822 | 22.6 |
| | A | nessuna | 15,707 | 0.9 | 546,712 | 38,022 | 600,442 | 15.3 |
| | R | | 5,336 | 0.3 | 45,284 | 9,591 | 60,211 | 1.5 |
| Due posizioni LFS (c) | A | 2 digit | 107,499 | 5.9 | 93,317 | 5,976 | 206,798 | 5.3 |
| | A | sezione | 10,219 | 0.6 | 10,211 | 725 | 21,155 | 0.5 |
| | A | tipologia | 26,891 | 1.5 | 33,204 | 9,354 | 69,450 | 1.8 |
| | A | nessuna | 726 | 0.0 | 8,910 | 6,037 | 15,672 | 0.4 |
| | R | | 566 | 0.0 | 260 | 1,435 | 2,262 | 0.1 |
| Tre posizioni LFS (c) | A | 2 digit | 3,419 | 0.2 | 2,260 | 1,012 | 6,691 | 0.2 |
| | A | sezione | | 0.0 | 32 | | 32 | 0.0 |
| | A | tipologia | 513 | 0.0 | 1,673 | 886 | 3,072 | 0.1 |
| | A | nessuna | | 0.0 | | 401 | 401 | 0.0 |
| | R | | 32 | 0.0 | | | 32 | 0.0 |
| Non occupato LFS | A | | 91,602 | 5.0 | 89,086 | 13,714 | 194,407 | 5.0 |
| | I | | 579 | 0.0 | 3,095 | 692 | 4,366 | 0.1 |
| TOTALE | | | 1,832,031 | 100.0 | 1,832,032 | 255,678 | 3,919,841 | 100.0 |
| | A | 2 digit | 1,222,720 | 66.7 | 494,054 | 83,441 | 1,800,282 | 45.9 |
| | | sezione | 153,103 | 8.4 | 116,532 | 12,778 | 282,421 | 7.2 |
| | | tipologia | 341,658 | 18.6 | 528,098 | 89,569 | 959,344 | 24.5 |
| | | altri | 108,035 | 5.9 | 644,708 | 58,173 | 810,922 | 20.7 |
| | R | | 5,934 | 0.3 | 45,544 | 11,026 | 62,505 | 1.6 |
| | I | | 579 | 0.0 | 3,095 | 692 | 4,366 | 0.1 |

Note: (a) A=ADMIN, R=LFS, I=Imputazione ateco ADMIN per strato; (b) Livello di coincidenza fra ateco ADMIN e ateco LFS. Per tipologia si intende Dipendente/indipendente; (c) La coerenza dei codici ateco viene verificata rispetto alla seconda posizione LFS

3.4. La stima delle posizioni non regolari degli occupati CN con una o più posizioni regolari

3.4.1. La stima della sottocopertura delle seconde posizioni non regolari

Gli occupati con almeno una posizione regolare (insieme O_1) possono ovviamente essere occupati anche in posizioni lavorative non regolari: tali posizioni possono eventualmente essere posizioni principali. Il metodo adottato per l'individuazione di queste posizioni non regolari poggia sul confronto effettuato a livello di singolo individuo fra le posizioni lavorative rilevate attraverso la rilevazione LFS (le quali in termini di principio possono essere sia regolari sia non regolari) e quelle risultanti dalle fonti amministrative (in questo caso si tratta di sole posizioni regolari). In generale, se il numero di posizioni LFS è più elevato si è assunto che ciò sia dovuto alla presenza di posizioni non regolari.

Tuttavia, come si è visto in precedenza, il numero di posizioni regolari misurate attraverso la base integrata LFS-ADMIN è più elevato rispetto a quello delle posizioni misurate attraverso la rilevazione: ciò lascia presumere la sottostima delle posizioni lavorative da parte di LFS con riferimento sia alle posizioni regolari sia a quelle non regolari. La correzione della sottostima della componente regolare è resa possibile proprio dalla disponibilità delle fonti amministrative: l'ipotesi dunque è che la componente ADMIN fornisca una stima corretta delle posizioni regolari. Sulla base di questa assunzione si può pertanto misurare l'entità della sottostima dell'indagine con riferimento alla componente regolare delle seconde posizioni lavorative. Per la componente non regolare

non sono invece disponibili fonti ausiliarie: si assumerà semplicemente che sia soggetta allo stesso tipo di sottostima, ossia che la propensione a nascondere una seconda posizione non regolare sia uguale alla propensione a nascondere una seconda posizione regolare.

3.4.2. Le stime individuali delle posizioni non regolari

Dal solo conteggio delle posizioni lavorative regolari attraverso la base dati integrata LFS-ADMIN emerge dunque una sottostima delle posizioni lavorative da parte dell'indagine. Di fatto la reale numerosità delle posizioni lavorative dell'individuo i è data dalla somma delle sue posizioni regolari e non regolari:

$$P_i = P_i^{reg} + P_i^{nreg} \quad <15>$$

dove $i \in O_1$

L'ipotesi di lavoro è che le fonti amministrative descrivano in maniera non distorta e sufficientemente precisa, la componente regolare:

$$P_i^{reg} = a_i \quad <16>$$

Le informazioni provenienti dalla rilevazione LFS indicano, per i soli occupati risultanti dall'indagine, una stima delle posizioni lavorative complessive di cui è noto il totale (p) ma non la ripartizione in regolari e non regolari:

$$p_i = p_i^{reg} + p_i^{nreg} \quad <17>$$

Una stima – probabilmente “prudenziale” - della componente non regolare di p_i è basata sull'assegnazione a questa componente delle posizioni generate dall'eccedenza del totale delle posizioni LFS rispetto alle posizioni regolari:

$$\hat{p}_i^{nreg} = \max(0; p_i - a_i) \quad <18>$$

Per una stima per strato (più esattamente per regione) delle posizioni non regolari effettive si assume che LFS sottostimi le posizioni non regolari con la medesima intensità con cui sottostima quelle regolari:

$$\hat{P}_{ST}^{nreg} = \frac{\sum_i a_i}{\sum_i (p_i - \hat{p}_i^{nreg})} \hat{p}_i^{nreg} \quad <19>$$

Per gli occupati di CN che non sono occupati secondo LFS viene stimato un rapporto di coesistenza per strato (regione), misurato sugli occupati LFS, che esprime il numero di posizioni non regolari per ogni posizione regolare

$$d_{ST} = \frac{\sum_{i \in ST} \hat{P}_i^{nreg}}{\sum_{i \in ST} \hat{P}_i^{reg}} \quad \langle 20 \rangle$$

con i tale che $p_i > 0$ (cioè occupato LFS), e lo si applica al totale delle posizioni regolari dello strato corrispondente dei non occupati LFS:

$$\hat{P}_{ST}^{nreg} = d_{ST} P_{ST}^{reg} \quad \langle 21 \rangle$$

L'assegnazione delle posizioni non regolari supplementari di strato ai singoli individui viene effettuata ricorrendo alla probabilità individuale di avere una posizione secondaria stimata attraverso una regressione logistica che utilizza le principali variabili LFS e le informazioni relative alle posizioni regolari⁴⁵.

3.4.3. Le caratteristiche delle posizioni non regolari

In generale, la soluzione adottata per imputare le caratteristiche di ciascuna posizione non regolare è basata sull'ordine della posizione. Se si tratta di una posizione principale si avrà:

$$\left\{ \begin{array}{l} ATE = ATE^{R1} \\ DIPIND = DIPIND^{R1} \\ CLAD = n.d. \\ FG = n.d. \\ SIST = "imprese" \\ REGIO = REGIO^R \end{array} \right. \quad \langle 22 \rangle$$

Per le posizioni secondarie (seconda, terza, ecc.), simmetricamente:

⁴⁵ Le variabili utilizzate per la stima della probabilità di avere attività secondarie derivano sia da LFS sia da ADMIN. Da LFS provengono la classe di età, il genere, la regione, la cittadinanza, lo stato civile, il trimestre. Da ADMIN provengono la sezione Nace del committente e i segnali di validazione e copertura nelle singole fonti. Gli effetti misti riguardano il genere e la sezione Nace, il genere e la regione, la sezione Nace e la ripartizione, la classe di età e la ripartizione.

$$\left\{ \begin{array}{l} ATE = ATE^{R2} \\ DIPIND = DIPIND^{R2} \\ CLAD = n.d. \\ FG = n.d. \\ SIST = "imprese" \\ REGIO = REGIO^R \end{array} \right. \quad \langle 23 \rangle$$

Nel caso delle posizioni non regolari imputate come effetto della correzione della sottocopertura non si hanno a disposizione informazioni LFS sulle posizioni secondarie e si utilizzano informazioni imputate tramite donatore per strato, con quest'ultimo definito in base al tipo di occupazione regolare (dipendente o indipendente), alle fonti delle posizioni regolari, al genere, alla ripartizione e alla classe dimensionale del datore di lavoro regolare⁴⁶:

$$\left\{ \begin{array}{l} ATE = ATE^{ST} \\ DIPIND = DIPIND^{ST} \\ CLAD = n.d. \\ FG = n.d. \\ SIST = "imprese" \\ REGIO = REGIO^R \end{array} \right. \quad \langle 24 \rangle$$

3.5. La stima delle posizioni non regolari degli occupati CN privi di posizioni regolari

Agli occupati che non ricoprono posizioni regolari, ossia a tutti gli occupati appartenenti all'insieme O_2 , sono state associate tutte le informazioni eventualmente disponibili e relative alle posizioni tracciate nelle fonti amministrative. Per questi occupati il grado di copertura delle fonti potrà infatti essere:

- ✓ di primo livello ($C_k = 1$), che assicura dunque una copertura nella settimana di riferimento della fonte k dell'intervista LFS che non è stata validata e non costituisce pertanto un segnale di occupazione regolare. Si tratta dell'insieme O_{21} : per questi casi è anche disponibile la probabilità stimata di essere un occupato regolare (che sarà inferiore alla soglia prefissata, il 50%);
- ✓ di sola presenza nell'anno ($C_k = 0$), con la settimana di riferimento esterna all'intervallo di copertura della fonte (insieme O_{22});
- ✓ assente, ovvero non vi è traccia dell'individuo nelle fonti amministrative (insieme O_{23}).

I metodi di imputazione delle caratteristiche delle posizioni non regolari è differenziato a seconda del sottoinsieme a cui appartengono gli occupati.

Per gli occupati con copertura di primo livello vengono individuate fino a due posizioni lavora-

⁴⁶ La procedura adottata è di tipo hot deck sequenziale applicata in maniera gerarchica su partizioni via via meno fini. L'impostazione è identica a quella descritta in precedenza con riferimento all'imputazione di attività economica e forma giuridica delle posizioni regolari.

tive distinte con un ordine gerarchico che dipende dalla probabilità stimata di costituire una posizione regolare. Per gli occupati con sola presenza nelle fonti, vengono registrate fino a due posizioni lavorative distinte secondo una gerarchia prefissata fra le fonti⁴⁷. In entrambe i casi, attraverso l'aggancio con i due principali committenti vengono così recuperate informazioni in merito al codice di attività economica e al tipo di posizione (dipendente o indipendente). Queste informazioni verranno poi utilizzate per la stima e l'imputazione di queste variabili alle posizioni non regolari.

L'attribuzione del numero di posizioni lavorative a ciascun occupato non regolare viene effettuata correggendo la sottocopertura di LFS. Nell'indagine all'intervistato vengono infatti attribuite fino a tre posizioni lavorative non regolari. La correzione viene apportata ipotizzando che l'incidenza delle doppie posizioni fra gli individui che ne dichiarano all'indagine una sola sia identica, per strato, a quella registrata nell'insieme O_1 degli occupati CN con almeno una posizione regolare. In particolare, si stima il coefficiente di strato:

$$g_{ST}^{O_{11}} = \frac{O_{11}^{pluri}}{O_{11}} \quad \langle 25 \rangle$$

dove O_{11} è l'insieme degli occupati CN con almeno una posizione regolare e che dichiarano a LFS di avere una sola posizione lavorativa, mentre O_{11}^{pluri} è la parte di questo insieme che risulta invece avere almeno due posizioni lavorative. Questo coefficiente viene utilizzato per stimare il numero effettivo di occupati con più posizioni nell'insieme O_2 :

$$O_{2,ST}^{pluri} = g_{ST}^{O_{11}} O_{2,ST} \quad \langle 26 \rangle$$

Per l'attribuzione della tipologia di posizione viene data priorità a LFS. L'attribuzione del codice Ateco viene invece effettuata sfruttando l'informazione di fonte LFS (disponibile per i soli occupati LFS) e quella disponibile dalla fonte amministrativa: il principio di fondo è quello di sostituire all'informazione LFS quella amministrativa solo qualora vi siano elementi sufficienti per assicurare con buona probabilità che le due fonti si riferiscano alla stessa posizione.

Per il trattamento delle singole posizioni, e in particolare per l'attribuzione del codice Ateco e della tipologia della posizione, gli occupati sono stati classificati in base al tipo di informazioni disponibili sulle posizioni:

- ✓ Occupati LFS con una sola posizione non regolare e una sola posizione lavorativa agganciata. In questo caso, se l'occupato ha una posizione amministrativa con copertura di primo livello e questa posizione coincide almeno per quanto riguarda il tipo di posizione con l'indicazione LFS, viene imputata l'Ateco della posizione amministrativa. Se invece l'occupato ha solo una presenza nelle fonti, le caratteristiche di queste ultima vengono usate nell'imputazione solo se il suo codice Ateco coincide almeno al secondo digit con quello dichiarato a LFS. Se nessuno di questi criteri viene rispettato, all'intervistato viene attribuito il codice Ateco dichiarato a LFS;
- ✓ Occupati LFS con più posizioni non regolari e una sola posizione lavorativa agganciata. Si procede con lo stesso trattamento differenziato sulla base della presenza o meno di

⁴⁷ La gerarchia è la seguente: DOME, EMEN, INDI, PROF, PARA, DMAG, ENPA, AUAG, INPD. Il numero dei casi in oggetto è tuttavia decisamente marginale.

una copertura di primo livello. In particolare, si verifica l'adattamento della posizione amministrativa anzitutto sulla prima posizione LFS e poi sulla seconda. Alla posizione che rimane scoperta, o eventualmente a entrambe le posizioni, viene infine attribuito il corrispondente codice Ateco LFS. In presenza di una terza posizione LFS, per la quale non ci sono però indicazioni di sorta dall'indagine, a questa vengono attribuite le medesime caratteristiche della seconda posizione o eventualmente le caratteristiche della posizione amministrativa qualora non fosse stata utilizzata per le prime due posizioni.

- ✓ Occupati LFS con una sola posizione non regolare e due posizioni lavorative agganciate. Viene imputato il codice Ateco della posizione amministrativa più vicina al dato LFS sulla base del medesimo criterio differenziato. Se la posizione rimane scoperta si utilizza l'informazione LFS.
- ✓ Occupati LFS con più posizioni non regolari e due posizioni lavorative agganciate. Vengono imputati a ciascuna posizione i codici di attività economica delle posizioni amministrative più vicine sulla base del criterio differenziato. Se non sono disponibili si usano le informazioni LFS. Per la eventuale terza posizione si usa il committente amministrativo principale se non utilizzato per le prime due posizioni o si duplica la seconda posizione.
- ✓ Occupati LFS con una o più posizioni e privi di presenze amministrative. Viene utilizzata l'informazione LFS sulle prime due posizioni mentre per la eventuale terza posizione viene duplicata la seconda.
- ✓ Non occupati LFS in cassa integrazione (CIG) e privi di posizione amministrativa. Questo segmento viene assegnato all'occupazione regolare: vengono imputate Ateco, forma giuridica, addetti e tipo di posizione sulla base della moda di strato calcolata sugli individui LFS in CIG già classificati fra gli occupati regolari. Gli strati sono ottenuti incrociando genere e ripartizione. La posizione assegnata è una sola.
- ✓ Non occupati LFS. Imputazione di Ateco e tipo di posizione tramite selezione casuale di un donatore per strato. Gli strati sono definiti incrociando genere, ripartizione, cittadinanza e classe di età. La posizione assegnata è una sola.

3.6. Qualche risultato

Nel complesso il metodo stima nel 2010 circa 26 milioni di posizioni lavorative di cui il 10.4% non regolari e il 9.6% secondarie (Tavola 22). Il tasso di irregolarità delle posizioni secondarie è più elevato di circa 4-5 punti percentuali rispetto alle posizioni regolari. Nell'agricoltura una posizione su cinque non è regolare, seguono costruzioni commercio e servizi alle famiglie. Nel 2011 il quadro strutturale è pressoché analogo, tuttavia è interessante notare come la componente non regolare evidenzia una contrazione delle posizioni lavorative a fronte di una sostanziale tenuta della componente regolare: a perdere terreno sono soprattutto le posizioni non regolari delle costruzioni e nei servizi.

Se si confronta con la stima delle posizioni lavorative di LFS, la nuova procedura ne stima oltre 3 milioni in più, equivalenti al 15% in più (Tavola 23): due terzi della differenza sono dovuti alle posizioni secondarie il cui numero è più elevato di 6,7 volte rispetto alle posizioni conteggiabili con LFS. Agricoltura e costruzioni sono i settori dove più ha inciso il differenziale sulle seconde posizioni. Con riferimento alle posizioni principali la differenza con LFS è solo del 5% ed è molto sostenuta in agricoltura (18%).

Tavola 22. Posizioni medie annue (.000) per ordine della posizione, regolarità e settore. Anni 2010 e 2011

| Settori | Principali | | Secondarie | | Totale | | Tasso irregolarità % | | | Incid. % Se-cond. |
|--|------------|------------|------------|------------|---------|------------|----------------------|------------|--------|-------------------|
| | Regol.i | Non regol. | Regol.i | Non regol. | Regol.i | Non regol. | Principali | Secondarie | Totale | |
| 2010 | | | | | | | | | | |
| Agricoltura | 740 | 204 | 351 | 52 | 1,090 | 257 | 21.7 | 13.0 | 19.1 | 29.9 |
| Industria in s.s. | 4,184 | 288 | 125 | 15 | 4,310 | 303 | 6.4 | 10.8 | 6.6 | 3.0 |
| Costruzioni | 1,704 | 259 | 132 | 10 | 1,837 | 270 | 13.2 | 7.3 | 12.8 | 6.8 |
| Commercio e alberghi | 4,246 | 503 | 402 | 45 | 4,647 | 548 | 10.6 | 10.0 | 10.5 | 8.6 |
| Servizi alle imprese | 4,599 | 389 | 550 | 68 | 5,149 | 457 | 7.8 | 11.0 | 8.2 | 11.0 |
| Servizi alle famiglie | 5,568 | 672 | 550 | 176 | 6,118 | 848 | 10.8 | 24.2 | 12.2 | 10.4 |
| Totale | 21,042 | 2,315 | 2,110 | 367 | 23,151 | 2,682 | 9.9 | 14.8 | 10.4 | 9.6 |
| 2011 | | | | | | | | | | |
| Agricoltura | 746 | 203 | 351 | 56 | 1,097 | 259 | 21.4 | 13.7 | 19.1 | 30.0 |
| Industria in s.s. | 4,173 | 284 | 150 | 14 | 4,323 | 297 | 6.4 | 8.3 | 6.4 | 3.5 |
| Costruzioni | 1,633 | 229 | 167 | 13 | 1,801 | 242 | 12.3 | 7.3 | 11.8 | 8.8 |
| Commercio e alberghi | 4,233 | 486 | 477 | 42 | 4,711 | 529 | 10.3 | 8.1 | 10.1 | 9.9 |
| Servizi alle imprese | 4,710 | 371 | 608 | 61 | 5,318 | 432 | 7.3 | 9.1 | 7.5 | 11.6 |
| Servizi alle famiglie | 5,683 | 664 | 567 | 168 | 6,250 | 833 | 10.5 | 22.9 | 11.8 | 10.4 |
| Totale | 21,178 | 2,237 | 2,321 | 354 | 23,499 | 2,591 | 9.6 | 13.2 | 9.9 | 10.3 |
| Variazioni % (in p.p. sulle percentuali) | | | | | | | | | | |
| Agricoltura | 0.8 | -0.7 | 0.2 | 6.6 | 0.6 | 0.8 | -0.3 | 0.7 | 0.0 | 0.1 |
| Industria in s.s. | -0.3 | -1.4 | 19.7 | -10.4 | 0.3 | -1.8 | -0.1 | -2.5 | -0.1 | 0.5 |
| Costruzioni | -4.2 | -11.8 | 26.5 | 27.2 | -2.0 | -10.3 | -0.9 | 0.0 | -1.0 | 2.1 |
| Commercio e alberghi | -0.3 | -3.3 | 18.8 | -5.1 | 1.4 | -3.5 | -0.3 | -1.9 | -0.5 | 1.3 |
| Servizi alle imprese | 2.4 | -4.6 | 10.7 | -10.9 | 3.3 | -5.5 | -0.5 | -2.0 | -0.6 | 0.6 |
| Servizi alle famiglie | 2.1 | -1.1 | 3.1 | -4.4 | 2.2 | -1.8 | -0.3 | -1.4 | -0.4 | 0.0 |
| Totale | 0.6 | -3.4 | 10.0 | -3.5 | 1.5 | -3.4 | -0.4 | -1.6 | -0.5 | 0.7 |

Tavola 23. Confronto fra posizioni medie annue degli occupati di CN (a) e stimate da LFS, per ordine della posizione e settore. Anno 2010

| Settori | Posizioni principali | | | Posizioni secondarie | | | Totale posizioni | | |
|-----------------------|----------------------|--------------|--------|----------------------|--------------|--------|------------------|--------------|--------|
| | Occupati CN | Occupati LFS | Indice | Occupati CN | Occupati LFS | Indice | Occupati CN | Occupati LFS | Indice |
| Agricoltura | 943,997 | 799,983 | 118 | 402,934 | 22,311 | 1806 | 1,346,932 | 822,295 | 164 |
| Industria in s.s. | 4,471,943 | 4,319,894 | 104 | 140,558 | 15,846 | 887 | 4,612,501 | 4,335,740 | 106 |
| Costruzioni | 1,963,904 | 1,848,361 | 106 | 142,551 | 11,350 | 1256 | 2,106,455 | 1,859,711 | 113 |
| Commercio e alberghi | 4,748,646 | 4,438,456 | 107 | 446,264 | 59,915 | 745 | 5,194,910 | 4,498,371 | 115 |
| Servizi alle imprese | 4,988,440 | 4,714,714 | 106 | 617,490 | 92,491 | 668 | 5,605,930 | 4,807,205 | 117 |
| Servizi alle famiglie | 6,239,866 | 6,018,675 | 104 | 726,272 | 168,707 | 430 | 6,966,138 | 6,187,382 | 113 |
| Totale | 23,356,796 | 22,140,083 | 105 | 2,476,069 | 370,620 | 668 | 25,832,866 | 22,510,704 | 115 |

Nota: (a) il riferimento alla CN è da intendersi nel senso più generico di "stima con la nuova procedura di supporto alle stime CN".

4. La stima delle ore effettivamente lavorate

4.0 Sintesi

Il punto di partenza per la stima delle ore è costituito dalla stima delle posizioni lavorative, e in particolare dal campione integrato LFS-ADMIN_JOBS. L'obiettivo è associare a ogni posizione lavorativa una quantificazione delle ore effettivamente lavorate nella settimana di riferimento dell'intervista: si tratta in sostanza di aggiungere a LFS-ADMIN_JOBS una o più variabili relative al numero di ore effettivamente lavorate (LFS-ADMIN_JOBSh).

La base informativa è costituita dalle informazioni rilevate tramite LFS sulle ore lavorate nella settimana di riferimento dell'intervista. Si tratta in particolare del quesito C37, il quale riporta le ore lavorate nella posizione principale, e del quesito D2 relativo alle ore lavorate nell'insieme delle eventuali posizioni secondarie. L'informazione sulla posizione principale è disponibile ovviamente solo per gli occupati LFS.

La stima delle ore lavorate è il risultato di un lavoro di verifica e correzione delle informazioni raccolte tramite LFS e di imputazione statistica per gli occupati e le posizioni lavorative per i quali l'informazione LFS è mancante. Questo lavoro è stato condotto sul campione integrato LFS-ADMIN_JOBS e fa leva perciò sia sulle informazioni strutturali relative all'intervistato LFS sia sull'informazione relativa alla posizione lavorativa reperita anche attraverso ADMIN.

Preliminarmente sono state corrette alcune distorsioni che caratterizzano le variabili sulle ore lavorate raccolte tramite LFS. Il GdL ha considerato tre classi di cause di distorsione: la sottostima delle giornate non lavorate a causa di ferie e festività (risultata non significativa); la sottostima delle assenze per congedi e malattie; la sovrastima delle ore lavorate generata dalle interviste proxy.

Per un numero nel complesso limitato di unità, le ore lavorate nella posizione principale che risultavano mancanti nei dati dell'intervista sono state imputate sulla base dell'informazione disponibile sulle ore abitualmente lavorate.

L'evidente squilibrio fra le informazioni disponibili e quelle da imputare sulla prima e sulla seconda posizione ha costretto a modulare di conseguenza i metodi di imputazione, principalmente basati su tecniche di donazione hot deck. Complessivamente, occorre infatti imputare le ore lavorate nella prima posizione di circa 14 mila unità osservate avendo a disposizione oltre 200 mila osservazioni, mentre per le posizioni secondarie occorre imputare oltre 20 mila unità con un bacino di rispondenti di poco superiore a 3 mila osservazioni.

È stata misurata e corretta la sovrastima di ore lavorate sulle posizioni principali da parte degli individui con posizioni secondarie non dichiarate a LFS, ipotizzando che abbiano risposto erroneamente al quesito C37 concentrando sulla posizione principale le ore lavorate su più posizioni.

Gli effetti più rilevanti sulla stima delle ore lavorate sono prodotti nell'ambito delle procedure di imputazione sulle prime e soprattutto sulle seconde posizioni. Le correzioni per gli eventi di malattia portano alla riduzione di 0.5 ore dell'orario pro capite settimanale. La correzione delle proxy riduce di 0.2 ore il pro capite settimanale. L'imputazione in base alle ore abituali lascia invariato il pro capite mentre accresce dello 0.2% il monte ore, dal momento che recupera le misure orarie altrimenti mancanti.

L'imputazione delle ore lavorate sulle prime posizioni produce una lieve contrazione del pro capite ampiamente compensato dal netto incremento – rispetto a LFS - delle posizioni lavorative: il monte ore delle prime posizioni aumenta del 4.6%. L'effetto sulle seconde posizioni è più marcato: il monte ore corrispondente rispetto a LFS cresce di oltre sei volte, con un aumento di oltre 1.2 miliardi del monte ore annuale: l'effetto delle imputazioni sul monte ore complessivo sale così oltre il 7%.

Considerando l'insieme di tutte le posizioni, gli effetti dell'imputazione sono stati più marcati nel terzo trimestre, fra le donne, nel Mezzogiorno, fra gli stranieri extra-Ue e nelle classi di età estreme. I pro capite (per posizione lavorativa) sono generalmente diminuiti: l'aumento del monte ore è determinato essenzialmente dal maggiore numero di posizioni lavorative.

4.1. Introduzione

Il punto di partenza di questo lavoro è costituito dalla stima delle posizioni lavorative derivata dal campione integrato LFS-ADMIN, e in particolare dal dataset LFS-ADMIN_JOBS i cui record sono costituiti dalle singole posizioni lavorative ricoperte dagli individui osservati tramite il campione LFS¹. Dal momento che a ciascun individuo nel campione possono corrispondere più posizioni lavorative, ognuna di esse è contraddistinta da un numero d'ordine che identifica la posizione principale, quella secondaria, ecc. A ogni posizione lavorativa inoltre sono associate altre variabili che ne descrivono le caratteristiche: l'attività economica, il tipo di occupazione (dipendente o indipendente), la natura della posizione (regolare o non regolare), le caratteristiche del datore di lavoro (forma giuridica e classe di addetti), il settore istituzionale di competenza (imprese, famiglie, istituzioni pubbliche, istituzioni sociali private, ecc.). L'obiettivo della stima delle ore lavorate è dunque quello di associare a ogni posizione lavorativa una quantificazione delle ore effettivamente lavorate nella settimana di riferimento dell'intervista: si tratta in sostanza di aggiungere a LFS-ADMIN_JOBS una o più variabili relative al numero di ore effettivamente lavorate.

A tal fine vengono utilizzate le informazioni rilevate tramite LFS sulle ore lavorate nella settimana di riferimento dell'intervista. Si tratta in particolare del quesito C37, il quale riporta le ore lavorate nella posizione principale, e del quesito D2 relativo alle ore lavorate nell'insieme delle eventuali posizioni secondarie. L'informazione sulla posizione principale è disponibile ovviamente solo per gli occupati LFS, ossia per gli intervistati che all'indagine risultano aver lavorato almeno un'ora nella settimana di riferimento (o di essere stati assenti da un'occupazione). L'informazione sulle posizioni secondarie è invece disponibile per i soli occupati LFS che hanno dichiarato di avere più posizioni lavorative nella settimana di riferimento.

Benché buona parte delle posizioni lavorative stimate tramite LFS-ADMIN afferiscano a occupati LFS, una porzione di esse riguarda comunque anche non occupati LFS. Si tratta principalmente di non occupati LFS con posizioni lavorative principali regolari, mentre solo una parte minoritaria ha posizioni lavorative principali non regolari. Per tutti questi individui ovviamente non è disponibile l'informazione sulle ore lavorate né nella posizione principale né in quella secondaria: in questi casi occorre dunque ricorrere a procedure di imputazione statistica. Analogamente, vanno imputate le ore lavorate nelle posizioni lavorative secondarie stimate tramite LFS-ADMIN e afferenti a occupati LFS che non dichiarano di averne e per i quali dunque la variabile D2 non è valorizzata. Infine, ulteriori imputazioni vanno effettuate nei casi – per la verità relativamente pochi – di occupati LFS con valori mancanti sulle ore lavorate.

Questo lavoro di imputazione delle ore lavorate, condotto sul campione integrato LFS-ADMIN, completa solo in parte la stima delle ore lavorate. Se da un lato infatti le ore lavorate nella posizione principale sono univocamente associate alla posizione stessa, le ore cumulate nelle posizioni secondarie (la variabile D2, appunto) vanno distribuite sulle posizioni secondarie, che in alcuni casi sono più di una².

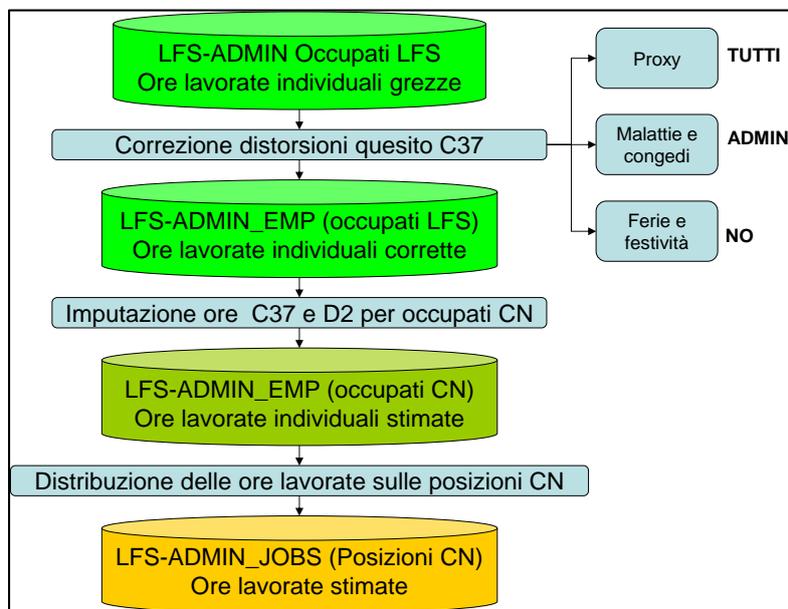
Limitatamente al campo di osservazione LFS, la stima delle ore lavorate è di fatto il risultato di un lavoro di imputazione statistica compiuta sul campione integrato LFS-ADMIN_JOBS sfruttando le variabili LFS sulle ore lavorate e facendo leva sia sulle informazioni strutturali relative all'intervistato LFS sia sull'informazione relativa alla posizione lavorativa reperita anche attraverso ADMIN. Preliminarmente tuttavia è necessario considerare e trattare alcune possibili distorsioni che caratterizzano proprio le variabili sulle ore lavorate raccolte tramite LFS. Nel corso dell'attività del GdL sono state in particolare considerate tre classi di cause di distorsione: la sottostima delle giornate non lavorate a causa di ferie e festività; l'analoga sottostima a causa di assenze per conge-

¹ Si continua a usare per comodità espositiva il termine individuo per identificare la singola intervista LFS, ossia la combinazione fra chiave individuale e trimestre di rilevazione.

² Occorre poi considerare il campo di osservazione esterno a LFS-ADMIN, costituito dalla posizioni lavorative degli occupati non residenti: questa parte della stima – che non è trattata in questo documento – dovrà prevedibilmente fare leva su un processo di imputazione statistica basato su LFS, che potrà contare per la parte regolare di questa occupazione sulle informazioni relative al rapporto di lavoro reperibili dalle fonti ADMIN.

di e malattie; la sovrastima delle ore lavorate generata dalle interviste proxy. Il trattamento e la correzione di questi effetti è di fatto preliminare alla stima attraverso imputazione statistica. La Figura 13 illustra i passaggi del processo di stima delle ore lavorate. Nel paragrafo 2 vengono descritti gli interventi effettuati per controllare queste distorsioni, mentre nei paragrafi successivi viene descritto nei dettagli il processo di imputazione statistica delle ore lavorate.

Figura 13. Dalle ore lavorate individuali rilevate da LFS alle ore lavorate nelle posizioni lavorative di CN



4.2. Correzione delle distorsioni LFS

4.2.1 La sottostima delle giornate non lavorate per ferie o festività

La presenza di questo effetto è spesso citata in letteratura, sebbene non sia di fatto accompagnata da evidenze empiriche. Fa eccezione il lavoro svolto dall'Istat³ nel quale viene analizzato il comportamento dei rispondenti in occasione di alcune festività occasionali (*bank holidays*). In particolare si evidenzia l'assenza di effetti memoria dovuti all'ampiezza del *lag* fra settimana dell'intervista e settimana di riferimento, la possibile presenza di effetti dovuti al tipo di intervista (le interviste CAPI favorirebbero una sottostima delle assenze), l'assenza di effetti dovuti alle interviste proxy. Gli autori riportano in particolare i risultati di una indagine pilota effettuata in occasione della rilevazione condotta durante il breve periodo feriale che ha luogo all'inizio di novembre (2012). L'introduzione di una semplice avvertenza (d'ora in avanti *warning*) per richiamare alla memoria dell'intervistato l'evento festivo ha infatti portato a un incremento nella quota di intervistati che hanno riconosciuto di avere lavorato di meno.

La principale novità su questo tema è stata determinata proprio dalla introduzione di questo *warning* in maniera definitiva nella LFS relativa al 2013. I dati del 2013 - esaminati dal GdL nelle riunioni del 21 novembre, 4 dicembre e 19 dicembre del 2013 - non sembrano evidenziare la presenza di chiari effetti di riduzione della stima delle ore lavorate indotte dal *warning*. In realtà si registra una forte contrazione delle ore lavorate a partire dal 2012 e dunque indipendentemente da

³ Loriga S., Spizzichino A. 2013. *Working hours: analysis of Italian LFS results versus administrative data and business survey*. Eurostat, riunione WG.. Loriga S., Spizzichino A. 2012. *Analisi sulle ore non lavorate per festività o ferie*. Presentazione alla riunione del GdL OCC del 21 novembre 2012.

qualsiasi effetto attribuibile al warning⁴. Si è pertanto dedotto che:

- ✓ Il warning non ha rivelato l'esistenza di effetti di sottostima delle ore non lavorate per ferie e festività;
- ✓ Non ci sono elementi sufficienti a supporto dell'introduzione di una procedura di abbattimento delle ore lavorate per tenere conto di tali effetti.

4.2.2 La sottostima delle giornate non lavorate per congedi o malattia

Per una misura della sottostima delle ore non lavorate per congedi o malattia possono essere sfruttate le informazioni presenti nello stesso campione integrato LFS-ADMIN e che riguardano in particolare la quasi totalità delle posizioni dipendenti regolari, ossia le posizioni lavorative con segnale Emens o Inpdap.

Per quanto riguarda gli occupati con posizione lavorativa principale tracciata in Emens⁵, è stato possibile fare una serie di correzioni sulla base degli eventi di assenza registrati nella fonte amministrativa relativi a malattie e infortuni di durata non inferiore a sette giorni, congedi per maternità/paternità e parentali, donazioni di sangue e riposi giornalieri per figli di età inferiore all'anno⁶.

Il criterio adottato per la correzione è il seguente:

- ✓ se l'intervistato ha dichiarato di aver lavorato meno del solito nella settimana, si assume che egli abbia tenuto conto di tali eventi nella sua risposta sulle ore effettivamente lavorate e, pertanto, non viene apportata alcuna correzione. In caso contrario, si devono distinguere due eventualità:
- ✓ in caso di donazione di sangue o di riposi giornalieri per figli di età inferiore all'anno, sono state corrette le ore effettivamente lavorate decurtandole dell'equivalente di una giornata lavorativa media, posta pari al 20% delle ore abitualmente lavorate (dichiarate nel quesito C31), assumendo perciò una settimana lavorativa di cinque giorni⁷;
- ✓ in corrispondenza delle altre tre tipologie di eventi di assenza (malattie e infortuni di durata non inferiore ai sette giorni nonché congedi per maternità/paternità e congedi parentali), invece, il numero di ore effettivamente lavorate viene azzerato, assumendo la perfetta coincidenza della settimana LFS e dell'evento registrato in Emens.

L'eventuale correzione per eccesso che può derivare da questo approccio può trovare una qualche compensazione per il fatto che non è invece possibile apportare correzioni per gli eventi di malattia/infortunio di durata inferiore ai sette giorni in quanto tali eventi non sono per ora registrati nella fonte amministrativa⁸.

Per gli occupati con posizione principale Inpdap, in maniera analoga, è stata sfruttata la variabile dicotomica relativa alla fruizione di congedi e aspettative nella settimana di riferimento dell'intervista. Il criterio adottato per la correzione è lo stesso sopra descritto per la decurtazione dell'equivalente di una giornata lavorativa. Per gli eventi di maternità, invece, sono state annullate le ore effettivamente lavorate.

E' importante sottolineare che a partire dal mese di riferimento di febbraio 2012 si potrà disporre del dato INPS di ottima qualità relativo a tutti gli eventi di malattia dei lavoratori dipendenti, a

⁴ Si veda su questo tema il lavoro di De Gregorio e Giordano: *Un esercizio sulla valutazione dell'effetto del warning sulle ore effettivamente lavorate nella posizione lavorativa principale*. Si tratta di un approfondimento reso necessario dalla necessità di pervenire a una soluzione per la stima di eventuali effetti dovuti a ferie o festività.

⁵ De Gregorio C., Giordano A. 2012. *Utilizzo combinato di LFS e Emens per una stima delle ore non lavorate per malattia*. Presentazione alla riunione del GdL OCC del 21 novembre 2012.

⁶ Più in particolare, gli eventi di assenza considerati sono i seguenti: "malattia, per eventi di durata non inferiore a sette giorni"; "periodi di congedo di maternità e paternità ex artt. 16, 17, 20 e 28, D.Lgs. n. 151/2001"; "periodi di congedo parentale disciplinati dall'art. 35, comma1, D.Lgs. n.151/2001, (6 mesi entro i 3 anni di vita del bambino)"; "infortunio, per eventi di durata non inferiore a sette giorni"; "assenza per donazione di sangue (art.13 della Legge 04/05/1990 n.107)"; "riposi giornalieri fino al primo anno di vita del bambino, disciplinati dagli artt.39 e 40 del D.Lgs. n. 151/2001 ex permessi per allattamento".

⁷ Se il dato relativo alla variabile C31 è mancante, le ore effettivamente lavorate vengono corrette con lo stesso criterio ma la giornata media viene calcolata sulla base di quanto dichiarato nel quesito C37 anziché C31.

⁸ De Gregorio C., Giordano A. 2013. *Stima delle ore non lavorate per malattia*.

prescindere dalla durata degli eventi stessi⁹. I dati attualmente disponibili, infatti, vengono acquisiti al fine del calcolo dei contributi figurativi e, quindi, contemplano solo le malattie di durata non inferiore a sette giorni: da quella data, invece, essendo anche entrate in vigore le sanzioni per il mancato invio in forma elettronica dei certificati di malattia di qualsiasi durata, è verosimile ritenere che tali informazioni siano esaustive rispetto alla totalità delle malattie.

La soluzione adottata si differenzia a seconda che le posizioni lavorative afferiscano a occupati LFS o a non occupati LFS. Con riferimento agli occupati, attraverso le informazioni desumibili da EMENS e INPDAP è possibile stimare una propensione a non dichiarare a LFS l'assenza dal lavoro e le corrispondenti ore non lavorate:

$$\mu_s = \frac{\sum_{i \in Q} w_i^s}{\sum_{i \in K} w_i^s} \quad \langle 1 \rangle$$

dove w sono i coefficienti di riporto all'universo di LFS, s identifica uno strato di una partizione S individuata dall'incrocio fra trimestre, genere, regime orario, classe di età e ripartizione geografica, Q identifica l'insieme degli intervistati non proxy con eventi di malattia registrati in EMENS o in INPDAP e K identifica tutti gli intervistati non proxy con posizioni lavorative in EMENS o INPDAP e con segnali di malattia registrati in LFS. In sostanza μ individua una relazione che traduce per strato il numero di eventi registrati da LFS in eventi effettivi misurati nelle fonti. Questo indicatore ha il vantaggio di non introdurre forzature sulla propensione ad avere eventi di malattia, il cui differenziarsi fra i vari segmenti di popolazione e occupazione si presume sia colto da LFS. Il numero di individui supplementari cui attribuire un evento di malattia è allora determinato da:

$$m_s = \mu_s \sum_{i \in G} w_i^s \quad \langle 2 \rangle$$

dove G è l'insieme degli occupati LFS nello strato s con eventi di malattia rilevati all'indagine. Per il segmento dei non occupati LFS, per i quali non si dispone dell'informazione sugli eventi di malattia, la stima del numero di eventi effettivi viene effettuata assumendo una identica incidenza per strato:

$$\eta_s = \frac{\sum_{i \in L} w_i^s}{\sum_{i \in N} w_i^s} \quad \langle 3 \rangle$$

dove L è l'insieme dei non occupati LFS con segnali di malattia registrati in EMENS o in INPDAP e N è il numero di non occupati LFS con coperture nelle due fonti. Gli strati s sono analoghi a quelli utilizzati per gli occupati, con la cittadinanza al posto del regime orario. Analogamente a quanto visto in precedenza, il numero di individui supplementari cui attribuire un evento di malattia è allo-

⁹ L'obbligo di invio telematico dei certificati medici all'INPS decorre da giugno 2010; l'Istat ha richiesto le basi dati a partire da gennaio 2011 ma è da febbraio 2012 che sono in vigore le sanzioni.

ra determinato da:

$$n_s = \eta_s \sum_{i \in M} w_i^s \quad \langle 4 \rangle$$

dove M è l'insieme dei non occupati LFS non presenti nelle due fonti amministrative.

Per entrambi i segmenti (occupati e non occupati LFS) la stima delle unità aggiuntive cui attribuire eventi di malattia è effettuata separatamente per gli eventi brevi e per quelli lunghi nonché per le posizioni principali e quelle secondarie. Infine, per ciascuno strato, la selezione delle unità cui effettivamente imputare gli eventi di malattia è avvenuta con un criterio casuale.

4.2.3 La correzione per dell'effetto delle interviste proxy

Un'ulteriore correzione è stata apportata sulla variabile C37 rilevata tramite interviste *proxy*, tema sul quale esiste un'ampia letteratura. Benché l'entità di un effetto di sovrastima delle ore lavorate dovuto alle interviste *proxy* risulti nel complesso limitato¹⁰, di esso si è tuttavia tenuto conto: alle ore rilevate tramite interviste *proxy* è stato applicato un coefficiente correttivo stimato a livello di strato pari al rapporto tra il pro capite orario rilevato con interviste dirette e quello rilevato tramite interviste *proxy*¹¹.

Infine sono state imputate zero ore agli individui in CIG secondo ADMIN e non occupati secondo le definizioni LFS. Questi individui sono stati opportunamente etichettati dal momento che, secondo il regolamento SEC 2010, la loro posizione di cassintegrato (prevedibilmente a zero ore) non costituisce una posizione lavorativa.

4.3. L'utilizzo delle ore abitualmente lavorate

4.3.1. Aspetti generali

A seguito della correzione della distorsione indotta dalle interviste *proxy* abbiamo dunque a disposizione la base informativa LFS corretta che verrà utilizzata per imputare le ore lavorate di LFS-ADMIN_JOBS. Dopo queste correzioni preliminari, si passa infatti all'imputazione delle ore sulle prime e sulle seconde posizioni degli occupati di CN della base dati LFS-ADMIN laddove l'informazione è mancante. A questo scopo, le posizioni lavorative di LFS-ADMIN_JOBS sono state divise in sei insiemi distinti ed esaustivi:

1. Posizioni di individui monoposizione con ore lavorate rilevate;
2. Posizioni di individui con più di una posizione e con ore lavorate rilevate su tutte le posizioni;
3. Posizioni di individui con più di una posizione ma con ore lavorate rilevate solo sulla posizione principale;
4. Posizioni di individui monoposizione con ore lavorate non rilevate;
5. Posizioni di individui con più di una posizione e con ore lavorate non rilevate;
6. Posizioni di individui con più di una posizione ma con ore lavorate rilevate solo sulle posizioni secondarie.

La Tavola 24 sintetizza e quantifica le sei casistiche. I primi due insiemi riguardano gli occupati

¹⁰ De Gregorio C., Giordano A. 2012, *La misura dell'effetto delle interviste proxy sulla stima delle ore lavorate*. Documento presentato alla riunione del GdL OCC del 12 Settembre. Il documento contiene una breve rassegna della letteratura esistente sul tema dell'effetto delle interviste proxy sulla stima delle ore lavorate.

¹¹ Tale correzione non è stata apportata ai lavoratori a chiamata (LAC), sia a tempo determinato sia a tempo indeterminato: tale scelta è dovuta alla constatazione che l'orario dei LAC può essere molto diverso da quello dei lavoratori con contratti "standard" e che l'esigua numerosità delle interviste proxy relative ai LAC non consente di inserire anche tale informazione (LAC/non LAC) come variabile di stratificazione.

per i quali vi è una disponibilità completa di informazioni sulle ore lavorate. Si tratta nel complesso di circa l'86% del totale degli occupati di CN, la maggior parte dei quali con una sola posizione lavorativa: l'insieme 1 raccoglie infatti l'85% circa degli occupati. Viceversa, l'insieme 2 è relativamente piccolo dal momento che raccoglie meno dell'1,5% dell'occupazione complessiva stimata da CN (un po' più di 3 mila osservazioni campionarie).

Vi è un evidente squilibrio fra le informazioni disponibili e quelle da imputare sulla prima e sulla seconda posizione, il che costringe a modulare conseguentemente i metodi di imputazione. Complessivamente, occorre infatti imputare le ore lavorate nella prima posizione di circa 14 mila osservazioni campionarie avendo a disposizione informazioni rilevate su oltre 200 mila osservazioni, mentre per le posizioni secondarie occorre imputare oltre 20 mila unità con un bacino di rispondenti di poco superiore a 3 mila osservazioni.

Tavola 24. Distribuzione degli occupati di CN, per numero di posizioni stimate da CN, informazioni LFS sulle ore lavorate (variabili C37 e D2) e bacino dei donatori. Anni 2010 e 2011 (osservazioni e stime)

| Insieme | N. posizioni CN | Informazione sulle ore lavorate | | Bacino donatori | Anno 2010 | | | | Anno 2011 | | | |
|---------------|-----------------|---------------------------------|----------|--------------------------|----------------|------------|-------------------|------------|----------------|------------|-------------------|------------|
| | | C37 | D2 | | Osserv. | | Stime | | Osserv. | | Stime | |
| | | | | | n | % | n | % | n | % | n | % |
| 1 | Una | presente | - | - | 206,739 | 84.8 | 19,981,296 | 85.5 | 200,545 | 84.4 | 19,938,994 | 85.2 |
| 2 | Più di una | presente | presente | - | 3,480 | 1.4 | 330,719 | 1.4 | 3,172 | 1.3 | 304,444 | 1.3 |
| 3 | Più di una | presente | missing | Insieme 2 (per D2) | 19,494 | 8.0 | 1,789,633 | 7.7 | 20,438 | 8.6 | 1,948,447 | 8.3 |
| 4 | Una | missing | - | Insieme 1 (per C37) | 12,788 | 5.2 | 1,145,068 | 4.9 | 12,141 | 5.1 | 1,091,807 | 4.7 |
| 5 | Più di una | missing | missing | Insieme 2 (per C37 e D2) | 1,269 | 0.5 | 109,827 | 0.5 | 1,439 | 0.6 | 131,125 | 0.6 |
| 6 | Più di una | missing | presente | Insieme 2 (per C37) | 6 | 0.0 | 254 | 0.0 | 3 | 0.0 | 450 | 0.0 |
| Totale | | | | | 243,776 | 100 | 23,356,797 | 100 | 237,738 | 100 | 23,415,267 | 100 |

Fonte: Base dati LFS-ADMIN

Gli insiemi più corposi da imputare sono gli insiemi 3 e 4. L'insieme 3 è costituito da occupati con più di una posizione lavorativa stimata da CN ma per i quali sono note solamente le ore lavorate nella prima posizione (l'unica dichiarata a LFS). Si tratta nel complesso di circa 20 mila osservazioni. L'imputazione delle ore sulla seconda posizione usa le informazioni dell'insieme 2 e, come si vedrà, è stata effettuata tramite donatore con una procedura di tipo hot deck, modificata per superare ad alcuni inconvenienti dovuti alla esiguità dell'insieme dei donatori.

L'insieme 4 riguarda essenzialmente non occupati LFS: in massima parte si tratta di occupati regolari, per i quali cioè è stata reperita una posizione lavorativa validata nelle fonti amministrative; in misura inferiore, si trovano in questo sottoinsieme occupati di CN non regolari recuperati tramite la stima della sottocopertura LFS. In pochi casi appartengono a questo sottoinsieme quanti, pur risultando occupati in LFS, non sono stati in grado di rispondere al quesito C37. Per questo insieme il processo di imputazione utilizza l'informazione dell'insieme 1: la dimensione del bacino dei donatori è notevole e si è pertanto adottato un metodo di imputazione di tipo hot deck sequenziale gerarchico con un livello di stratificazione piuttosto fine.

La scelta del metodo di imputazione hot deck è finalizzata a non alterare la distribuzione della variabile oggetto d'imputazione, assumendo che il meccanismo di mancata risposta sia di tipo *missing at random* (MAR). Da un punto di vista operativo, quindi, dopo aver ordinato tutti i record (individuo-trimestre) secondo un numero casuale generato con distribuzione uniforme, il dato mancante viene sostituito con il valore assunto nel record immediatamente precedente: tale operazione viene fatta nell'ambito di classi omogenee, ossia all'interno delle quali è possibile ipotizzare che gli intervistati abbiano caratteristiche simili rispetto alle ore effettivamente lavorate. La numerosità minima delle classi di imputazione viene impostata pari a un valore soglia (generalmente pari a 20 osservazioni): laddove esistano classi di numerosità inferiore, si passa a una partizione meno fine; il punto iniziale di ciascuna classe, utile per imputare il primo record di strato nel caso sia *missing*, viene posto pari alla media ponderata per classe delle ore effettivamente lavorate, calcolata al netto delle interviste *proxy*.

Gli ultimi due insiemi sono decisamente più esigui. L'insieme 5 riguarda occupati CN con più di una posizione ma che non sono occupati LFS. L'insieme 6, alquanto marginale, è composto da occupati LFS per i quali è valorizzato solo il numero di ore sulle seconde posizioni ma è mancante il dato sulla prima. In ambedue i casi l'imputazione è effettuata utilizzando i dati dell'insieme 2.

4.3.2. L'utilizzo dell'informazione sulle ore abitualmente lavorate

Per un numero nel complesso esiguo di unità per le quali non si dispone delle ore effettivamente lavorate nella posizione principale si può tuttavia contare sulla presenza della variabile relative alle ore abitualmente lavorate. Si tratta di osservazioni che appartengono agli insiemi 4, 5 e 6. Il criterio adottato per l'imputazione è il seguente:

- ✓ Si seleziona casualmente un donatore con lo stesso numero di ore abituali;
- ✓ Si escludono dai donatori le unità con eventi di malattia;
- ✓ Si utilizza una stratificazione basata su alcune variabili: motivo per cui si è lavorato di meno, orario abituale, regime orario, trimestre, genere, tipo di posizione, settore di attività;
- ✓ Gli strati sono organizzati gerarchicamente, da una partizione più fine a una partizione con pochi strati;
- ✓ L'insieme 4 viene imputato usando le osservazioni dell'insieme 1;
- ✓ Gli insiemi 5 e 6 usando le informazioni dell'insieme 2;
- ✓ La scelta del donatore avviene tramite una procedura hot deck sequenziale, iterata sulle partizioni gerarchiche con un vincolo sulla numerosità minima dei donatori per strato.

4.4. Stima delle ore lavorate sulle seconde posizioni – Insieme 3

4.4.1. Aspetti problematici

Si tratta della parte più problematica del processo di imputazione. L'informazione sulle ore la-

vorate sulle seconde posizioni di fonte LFS infatti non è esaustiva a causa della notevole sottostima delle seconde posizioni da parte dell'indagine che emerge in modo piuttosto chiaro dal confronto con ADMIN. Si tratta di un problema noto in letteratura. D'altra parte, anche gli stessi dati LFS in parte lo confermano: non tutti coloro che hanno dichiarato di lavorare *part-time* perché svolgono un secondo lavoro hanno poi effettivamente dichiarato posizioni secondarie¹.

La tavola 25 evidenzia come il 90% circa degli occupati nell'insieme 3 sia costituito da individui con presenza validata nelle fonti amministrative. Si tratta di una proporzione pressoché identica a quella osservata nell'insieme 2, nel quale tuttavia l'incidenza delle interviste *proxy* è molto più bassa. Si può notare inoltre come l'insieme 2 rappresenta fra il 15% e il 20% dell'insieme 3, con una incidenza più elevata fra gli individui con posizione ADMIN e più bassa nel 2011.

Tavola 25. Occupati di CN negli insiemi 2 e 3, per presenza nelle fonti, tipo di intervista e anno (dati assoluti e %)

| Tipo di occupati di CN | Anno 2010 | | | | Anno 2011 | | | |
|------------------------------|---------------|------------|------------------|------------|---------------|------------|------------------|------------|
| | Oss. | % | Stime | % | Oss. | % | Stime | % |
| <i>Totale Insieme 3</i> | | | | | | | | |
| Con segnale ADMIN, di cui: | 17,418 | 89.4 | 1,599,691 | 89.4 | 18,262 | 89.4 | 1,749,543 | 89.8 |
| - intervista diretta | 14,584 | 74.8 | 1,338,142 | 74.8 | 15,228 | 74.5 | 1,464,361 | 75.2 |
| - intervista proxy | 2,834 | 14.5 | 261,549 | 14.6 | 3,034 | 14.8 | 285,182 | 14.6 |
| Senza segnale ADMIN, di cui: | 2,076 | 10.6 | 189,943 | 10.6 | 2,176 | 10.6 | 198,905 | 10.2 |
| - intervista diretta | 1,793 | 9.2 | 162,724 | 9.1 | 1,886 | 9.2 | 170,542 | 8.8 |
| - intervista proxy | 283 | 1.5 | 27,219 | 1.5 | 290 | 1.4 | 28,362 | 1.5 |
| Totale | 19,494 | 100 | 1,789,633 | 100 | 20,438 | 100 | 1,948,447 | 100 |
| <i>Totale Insieme 2</i> | | | | | | | | |
| Con segnale ADMIN, di cui: | 3,208 | 92.2 | 300,895 | 91.0 | 2,910 | 91.7 | 277,086 | 91.0 |
| - intervista diretta | 2,872 | 82.5 | 270,421 | 81.8 | 2,602 | 82.0 | 248,084 | 81.5 |
| - intervista proxy | 336 | 9.7 | 30,474 | 9.2 | 308 | 9.7 | 29,002 | 9.5 |
| Senza segnale ADMIN, di cui: | 272 | 7.8 | 29,824 | 9.0 | 262 | 8.3 | 27,358 | 9.0 |
| - intervista diretta | 239 | 6.9 | 25,879 | 7.8 | 234 | 7.4 | 24,577 | 8.1 |
| - intervista proxy | 33 | 0.9 | 3,945 | 1.2 | 28 | 0.9 | 2,780 | 0.9 |
| Totale | 3,480 | 100 | 330,719 | 100 | 3,172 | 100 | 304,444 | 100 |
| <i>% Insieme 2</i> | | | | | | | | |
| Con segnale ADMIN | 18.4 | | 18.8 | | 15.9 | | 15.8 | |
| Senza segnale ADMIN | 13.1 | | 15.7 | | 12.0 | | 13.8 | |
| Totale | 17.9 | | 18.5 | | 15.5 | | 15.6 | |

4.4.2. Due tipologie di dati mancanti

Le stime ottenute tramite la base dati integrata LFS-ADMIN hanno rivelato l'esistenza di numerose posizioni lavorative secondarie associate a individui che hanno indicato all'indagine una sola posizione lavorativa – quella principale - e le corrispondenti ore lavorate (quesito C37). La maggior

¹ Si tratta di coloro che hanno risposto al quesito C29 ("Per quale tra questi motivi lavora part-time?") con la modalità '004' ("Svolge un secondo lavoro") e al successivo quesito D1 sul secondo lavoro (*Oltre al lavoro di cui abbiamo parlato finora, "LA SETTIMANA SCORSA", cioè "DA LUNEDÌ... A DOMENICA..."*, Lei aveva anche altri lavori? Consideri qualsiasi lavoro (ad esempio, lezioni private, lavori svolti nel fine settimana) da cui ricava un guadagno o il lavoro non pagato solo se effettuato abitualmente presso la ditta di un familiare) hanno risposto "No".

parte di queste seconde posizioni sono state dedotte dalle fonti amministrative: costituiscono pertanto rapporti di lavoro attivi nella settimana di riferimento dell'intervista che hanno superato il processo di validazione statistica. Questa componente è perciò costituita da posizioni lavorative regolari. Una parte minoritaria di posizioni secondarie è invece costituita da posizioni non regolari.

Sulle ragioni della sottostima dei "secondi lavori" da parte di LFS si possono fare varie congetture: l'intervistato può mostrarsi reticente per timore di controlli da parte delle autorità fiscali o amministrative, o per evitare un maggiore disturbo statistico atteso; l'intervistatore stesso potrebbe avere intrattenuto comportamenti fraudolenti; può esservi un'ambiguità difficilmente evitabile nell'identificazione delle posizioni che porta l'intervistato con più attività a indicarle in una singola posizione e a collassarne le ore lavorate.

D'altra parte, dal punto di vista della stima del numero delle posizioni lavorative - una volta accettati il criterio di identificazione basato sull'indicatore composito "*chiave individuo-datore-trimestre*" e il metodo di assegnazione delle posizioni secondarie non regolari - non restano ulteriori margini di aggiustamento. Non altrettanto si può dire invece per la stima delle ore lavorate e in particolare per l'imputazione delle ore lavorate nelle seconde posizioni (quesito D2), dal momento che è ragionevole sospettare che le ore effettivamente lavorate indicate in C37, con riferimento al lavoro principale, contabilizzino in qualche caso anche le ore lavorate nelle posizioni secondarie: taluni individui potrebbero in pratica avere omesso l'indicazione di posizioni secondarie ma avere al tempo stesso tenuto conto delle ore in esse lavorate cumulandole di fatto all'interno del quesito C37. Indichiamo questi individui come insieme 3B, che costituisce dunque una parte dell'insieme degli occupati LFS cui imputare il numero di ore nelle posizioni secondarie. Per l'imputazione delle ore lavorate nelle posizioni secondarie vanno dunque tenuti distinti gli individui dell'insieme 3B, dagli individui che hanno omesso le seconde posizioni anche dal monte ore (l'insieme complementare 3A).

Il problema è dunque quello di stimare la consistenza di queste due componenti dell'insieme 3 e di identificare stocasticamente gli individui a essi appartenenti.

4.4.3. Alcune evidenze

Confrontando l'insieme 2 e l'insieme 3, l'esame delle ore lavorate nelle prime posizioni rivela alcune cose interessanti (Tavola 26):

- ✓ Fra gli occupati full time, quelli dell'insieme 2 (ossia del bacino dei donatori con ore nelle posizioni secondarie rilevate) hanno una media più bassa di ore lavorate nella posizione principale (C37), del 10% circa rispetto all'insieme 3 (3.2 ore). Questa differenza appare statisticamente significativa. Sembra dunque confermata la presenza di posizioni plurime non rivelate ma contabilizzate nelle ore lavorate;
- ✓ Sempre per i full time, nell'insieme 2 la media delle ore lavorate complessivamente (C37+D2) supera di circa undici ore la media di C37 nell'insieme 3. Fra un quarto e un terzo delle ore da imputare alle posizioni secondarie dell'insieme 3 è verosimilmente già contabilizzato sulla posizione principale;
- ✓ Con riferimento agli occupati full time nella posizione principale, la componente afferente all'insieme 3B e da trasferire dunque da C37 è più elevata fra quanti hanno una posizione principale da indipendente delle imprese o irregolare; non è invece significativa fra i titolari di posizioni EMEN;
- ✓ Fra i part time la componente dell'insieme 3B sembrerebbe praticamente assente, se si eccettua qualche segnale debole fra i lavoratori domestici.

Tavola 26. Media delle ore lavorate dagli occupati, per regime orario e insieme. Anno 2010

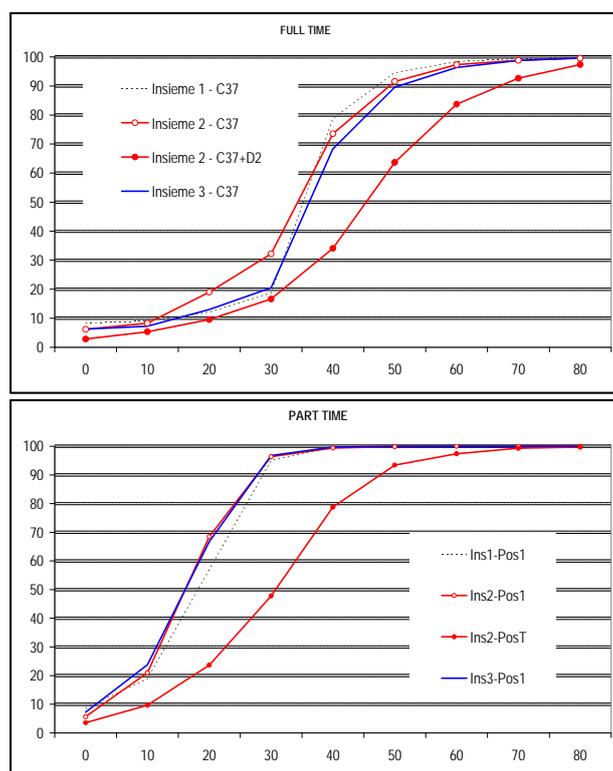
| Insiemi | stima | scost. da insieme 2 | Intervallo di confidenza | |
|--------------------|-------|---------------------|--------------------------|-------------|
| | | | Limite inf. | Limite sup. |
| <i>Full time</i> | | | | |
| Insieme 1 (C37) | 36.7 | 36.7 | 36.6 | 36.8 |
| Insieme 2 (C37) | 35.4 | 35.4 | 34.7 | 36.1 |
| Insieme 2 (C37+D2) | 46.4 | 46.4 | 45.6 | 47.2 |
| Insieme 3 (C37) | 38.6 | 38.6 | 38.4 | 38.8 |
| <i>Part time</i> | | | | |
| Insieme 1 (C37) | 19.6 | 19.6 | 19.5 | 19.7 |
| Insieme 2 (C37) | 18.2 | 18.2 | 17.7 | 18.6 |
| Insieme 2 (C37+D2) | 31.5 | 31.5 | 30.7 | 32.2 |
| Insieme 3 (C37) | 18.0 | 18.0 | 17.7 | 18.3 |

Note: (a) Nella terza colonna lo scostamento è calcolato rispetto alla variabile C37; (b) L'intervallo di confidenza è stato calcolato nell'ipotesi di simple random sampling all'interno dell'insieme 1.

L'andamento delle distribuzioni cumulate pare confermare queste deduzioni e rivela altri aspetti interessanti (Figura 14):

- ✓ Dal confronto sulla C37 fra insieme 1 e insieme 3 si conferma come le differenze siano concentrate nei segmenti orari più elevati, i quali dunque sembrano caratterizzare in misura significativa l'insieme 3;
- ✓ La distribuzione dell'insieme 3 è intermedia alle due distribuzioni dell'insieme 2 relative alla prima posizione (C37) e al monte ore individuale (C37+D2);
- ✓ La distribuzione dell'insieme 3 è più vicina a quella della C37;
- ✓ La distribuzione della C37 fra i part time dell'insieme 3 è identica a quella dei part time dell'insieme 2.

Figura 14. Distribuzione cumulata delle ore lavorate dagli occupati per regime orario e insieme. Anno 2010



4.4.4. Il metodo di imputazione

In considerazione di queste evidenze il GdL ha adottato un approccio differenziato all'imputazione delle ore lavorate per i due sottoinsiemi dell'insieme 3, 3A e 3B. In particolare si tratta di:

- Determinare un criterio per la stima della numerosità per strato dell'insieme 3B, composto come detto da individui per i quali le ore imputate sulle posizioni secondarie vanno sottratte a quelle dichiarate sulla posizione principale;
- Individuare le unità di ciascuno strato che appartengono a 3B;
- Imputare a queste unità le ore lavorate sulla seconda posizione e, per differenza con la variabile C37 originale, quelle lavorate nella posizione principale;
- Applicare all'insieme complementare 3A un criterio di imputazione delle ore nella posizione secondaria.

Si assume – sulla base delle evidenze empiriche espone in precedenza – che l'insieme 3B sia costituito unicamente da individui occupati a tempo pieno. Per stimarne la consistenza, sui full time dell'insieme 3 viene introdotta una stratificazione per classe di ore lavorate (*clh*) nella prima posizione² e tipo di fonte (*fon*)³. La loro distribuzione per classi di ore viene confrontata con quella analoga dell'insieme 2: prendendo come riferimento quest'ultima, sono state stimate le frequenze teoriche dell'insieme 3 e sono state determinate le numerosità per strato delle unità da rimuovere per rendere identiche le due distribuzioni. In formule (i suffissi 2 e 3 identificano gli insiemi e *t* i valori teorici):

² Sette classi ottenute con blocchi di dieci ore chiusi a destra, con classi dedicate ai casi con zero e 40 ore e con un'ultima classe definita oltre le 60 ore.

³ Sono stati tenuti distinti gli occupati non regolari, quelli con prima e seconda posizione INDI, gli altri con sola prima posizione INDI, e il resto degli occupati.

$$n_{3t}(clh, fon) = n_3(., fon) \left[n_2(clh, fon) / n_2(., fon) \right] \quad <5>$$

$$n_{3B}(clh, fon) = \max[0; n_{3t}(clh, fon) - n_3(clh, fon)] \quad <6>$$

Le unità rimosse – prevalentemente afferenti alle classi di ore più elevate - sono selezionate attraverso un numero casuale e sono assegnate all'insieme 3B, mentre le restanti fanno parte dell'insieme 3A. Sull'insieme 3B è stata condotta una imputazione *hot deck* sequenziale gerarchica usando una partizione più fine per ore complessivamente lavorate (variabile discreta) e fonte di provenienza, e partizioni via via meno fini ottenute eliminando la fonte e usando una divisione in classi delle ore. I donatori provengono dall'insieme 2, con un rapporto di 1.2 donazioni per donatore.

Per l'imputazione del resto dell'insieme 3 il rapporto sale evidentemente a 6.7 fra i full time mentre è relativamente contenuto fra i part time. Per utilizzare l'hot deck anche in questo segmento e per ovviare alla prevedibile forte incidenza delle imputazioni tramite media o ripetute, è stato gonfiato l'insieme dei donatori attraverso una clonazione ripetuta e un ordinamento casuale. In particolare, i dati che vengono presentati sono stati derivati decuplicando l'insieme 2 (Tavola 27).

Tavola 27. Osservazioni, per insieme e regime orario. Anno 2010

| Regime | Insieme 2 | Insieme 3 | Insieme 3A | Insieme 3B |
|---|-----------|-----------|------------|------------|
| <i>Osservazioni</i> | | | | |
| Full time | 2,039 | 15,946 | 13,581 | 2,365 |
| Part time | 1,441 | 3,548 | 3,548 | 0 |
| Totale | 3,480 | 19,494 | 17,129 | 2,365 |
| <i>Per ogni osservazione nell'insieme 2</i> | | | | |
| Full time | | 7.8 | 6.7 | 1.2 |
| Part time | | 2.5 | 2.5 | 0.0 |
| Totale | | 5.6 | 4.9 | 0.7 |

La partizione più fine ha utilizzato otto variabili (nell'ordine gerarchico: regime orario, tipologia della posizione principale, classe di ora lavorata nella prima posizione, trimestre, genere, aggregazione di fonti ADMIN, settore prima posizione, settore seconda posizione) via via collassate per ottenere partizioni più aggregate (Tavola 28).

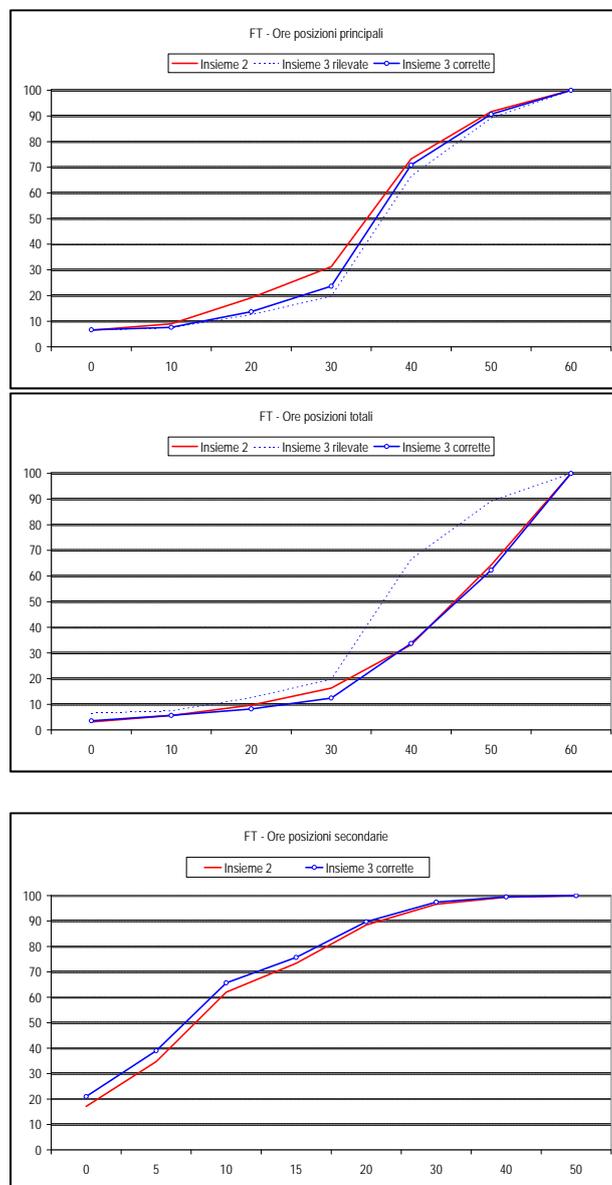
Tavola 28. Numero di osservazioni imputate tramite hot deck dell'insieme 3A per livello della partizione

| Variabili della partizione | Osservazioni imputate | % |
|----------------------------|-----------------------|------|
| 7 | 4,242 | 24.8 |
| 6 | 6,529 | 38.1 |
| 5 | 4,737 | 27.7 |
| 4 | 1,409 | 8.2 |
| 3 | 202 | 1.2 |
| 2 | 10 | 0.1 |
| Totale | 17,129 | 100 |

Le distribuzioni cumulate delle ore rilevate e imputate sono molto simili: in quelle dell'insieme 2 vi è un apprezzabile maggiore addensamento nelle modalità comprese fra 20 e 40 ore. Le posizioni totali arrivano quasi a coincidere. Anche le stime delle posizioni secondarie convergono rapi-

damente oltre i valori modali della distribuzione, mentre l'insieme 3 appare più concentrato sulle modalità orarie più basse (Figura 15).

Figura 15. Distribuzioni cumulate delle ore rilevate e imputate



4.4.5. La soluzione adottata

La soluzione effettivamente adottata per le stime prevede la stima separata delle ore lavorate dagli occupati negli insiemi 3B e 3A basata sulla tecnica di donazione hot deck per cella. In particolare:

- Si stimano con il metodo basato sugli scostamenti delle distribuzioni gli individui dell'insieme 3B, costituito dagli occupati full time le cui ore sulle posizioni secondarie sono incorporate nella risposta al quesito C37;
- Per questi individui viene stimata la variabile D2 scorporando alcune ore dalla loro risposta al quesito C37 sulla base di un metodo di donazione hot deck. Il metodo produce ovviamente per questi soggetti una nuova stima delle ore sulla prima posizione, mentre il loro monte ore resta invariato.

- Anche per gli individui dell'insieme 3A si adotta una imputazione hot deck inflazionando l'insieme dei donatori: viene così stimato un valore per D2 che fa corrispondentemente crescere il monte ore individuale.

La Tavola 29 illustra le numerosità per strato delle osservazioni relative ai donatori e ai dati mancanti: l'anno di riferimento è il 2010 (il 2011 restituisce risultati del tutto analoghi).

Tavola 29. Ore lavorate sulle posizioni secondarie degli occupati CN, per insiemi, modalità delle principali variabili di imputazione e tecnica di imputazione - Anno 2010

| Variabili | Osservazioni | | | | | Pro capite ore lavorate II posizioni | | | Scostamenti (3A-2) | | | |
|------------------------------------|--------------|-------|------------|-------|----------|--------------------------------------|------------|----------|--------------------|--------|----------|-------|
| | Insieme 2 | | Insieme 3A | | (3A-2) % | Insieme 2 | Insieme 3A | | Coeff. | | Hot-deck | |
| | n | % | n | % | | | Coeff. | Hot-deck | Ass. | % | Ass. | % |
| Totale | 3,480 | 100.0 | 17,129 | 100.0 | 4.9 | 11.8 | 10.9 | 11.0 | -0.9 | -7.9 | -0.7 | -6.3 |
| Regime orario Posizione principale | | | | | | | | | | | | |
| Full-Time | 2,067 | 59.4 | 13,578 | 79.3 | 6.6 | 10.8 | 10.3 | 10.5 | -0.5 | -4.2 | -0.3 | -2.3 |
| Part-time | 1,413 | 40.6 | 3,548 | 20.7 | 2.5 | 13.2 | 12.8 | 12.9 | -0.4 | -3.0 | -0.2 | -1.8 |
| Tipo Posizione principale | | | | | | | | | | | | |
| Dipendente | 2,192 | 63.0 | 13,578 | 79.3 | 6.2 | 11.1 | 10.3 | 10.5 | -0.7 | -6.7 | -0.5 | -4.9 |
| Indipendente | 1,288 | 37.0 | 3,548 | 20.7 | 2.8 | 13.0 | 12.8 | 12.9 | -0.3 | -2.2 | -0.1 | -1.0 |
| Fonte1 | | | | | | | | | | | | |
| D0 | 514 | 14.8 | 1,958 | 11.4 | 3.8 | 11.2 | 11.9 | 12.3 | 0.7 | 6.6 | 1.1 | 9.5 |
| auag | 12 | 0.3 | 6 | 0.0 | 0.5 | 17.3 | 14.2 | 11.5 | -3.1 | -18.0 | -5.8 | -33.3 |
| dmag | 26 | 0.7 | 390 | 2.3 | 15.0 | 13.3 | 10.5 | 10.0 | -2.8 | -21.3 | -3.4 | -25.2 |
| dome | 232 | 6.7 | 459 | 2.7 | | 11.1 | 11.7 | 11.0 | | | | |
| emen | 1,196 | 34.4 | 6,907 | 40.3 | 5.8 | 11.2 | 9.8 | 10.3 | -1.4 | -12.5 | -0.9 | -7.8 |
| enpa | 1 | 0.0 | 8 | 0.0 | 8.0 | 8.0 | 9.8 | 6.3 | 1.8 | 22.8 | -1.7 | -21.7 |
| indi | 963 | 27.7 | 5,719 | 33.4 | | 13.1 | 12.1 | 12.0 | | | | |
| inpd | 461 | 13.2 | 1,531 | 8.9 | 3.3 | 11.2 | 8.9 | 9.2 | -2.3 | -20.6 | -2.0 | -18.1 |
| para | 73 | 2.1 | 151 | 0.9 | 2.1 | 13.6 | 13.6 | 14.2 | 0.0 | 0.1 | 0.6 | 4.3 |
| prof | 2 | 0.1 | - | - | - | 3.7 | - | 1.0 | -3.7 | -100.0 | -2.7 | -73.0 |
| Regolarità Posizione principale | | | | | | | | | | | | |
| Regolare | 2,966 | 85.2 | 15,171 | 88.6 | 5.1 | 11.2 | 10.7 | 10.9 | -0.5 | -4.2 | -0.3 | -2.7 |
| Irregolare | 514 | 14.8 | 1,958 | 11.4 | 3.8 | 11.9 | 11.9 | 12.3 | 0.0 | 0.2 | 0.4 | 3.0 |
| Tipo Posizione secondaria | | | | | | | | | | | | |
| Dipendente | 1,345 | 38.6 | 1,958 | 11.4 | 1.5 | 12.1 | 11.9 | 12.3 | -0.2 | -1.7 | 0.1 | 1.0 |
| Indipendente | 2,134 | 61.3 | 15,171 | 88.6 | 7.1 | 11.6 | 10.7 | 10.9 | -0.8 | -7.3 | -0.7 | -5.8 |
| Fonte2 | | | | | | | | | | | | |
| D0 | 1,616 | 46.4 | 2,132 | 12.4 | 1.3 | 10.5 | 12.1 | 12.5 | 1.6 | 15.0 | 2.0 | 18.7 |
| auag | 7 | 0.2 | 19 | 0.1 | 2.7 | 23.3 | 8.3 | 13.7 | -15.0 | -64.3 | -9.6 | -41.1 |
| dmag | 10 | 0.3 | 98 | 0.6 | 9.8 | 18.0 | 11.8 | 9.8 | -6.2 | -34.5 | -8.3 | -45.8 |
| dome | 119 | 3.4 | 461 | 2.7 | 3.9 | 11.8 | 11.1 | 11.2 | -0.8 | -6.4 | -0.6 | -5.1 |
| emen | 532 | 15.3 | 2,558 | 14.9 | 4.8 | 13.0 | 9.7 | 10.3 | -3.3 | -25.4 | -2.7 | -20.6 |
| enpa | - | - | 48 | 0.3 | | - | 12.2 | 16.9 | 12.2 | | 16.9 | |
| indi | 998 | 28.7 | 10,990 | 64.2 | 11.0 | 13.2 | 10.9 | 10.9 | -2.3 | -17.3 | -2.3 | -17.2 |
| inpd | 82 | 2.4 | 315 | 1.8 | 3.8 | 14.2 | 9.2 | 9.3 | -5.1 | -35.5 | -4.9 | -34.6 |
| para | 116 | 3.3 | 505 | 2.9 | 4.4 | 10.2 | 10.3 | 11.0 | 0.2 | 1.7 | 0.8 | 8.1 |
| prof | - | - | 3 | 0.0 | | - | 9.6 | 2.3 | 9.6 | | 2.3 | |

Tavola 29. Segue Ore lavorate sulle posizioni secondarie degli occupati CN, per insiemi, modalità delle principali variabili di imputazione e tecnica di imputazione - Anno 2010

| Variabili | Osservazioni | | | | | Pro capite ore lavorate II posizioni | | | Scostamenti (3A-2) | | | |
|---------------------------------|--------------|-------|------------|-------|----------|--------------------------------------|------------|----------|--------------------|-------|----------|-------|
| | Insieme 2 | | Insieme 3A | | (3A-2) % | Insieme 2 | Insieme 3A | | Coeff. | | Hot-deck | |
| | n | % | n | % | | | Coeff. | Hot-deck | Ass. | % | Ass. | % |
| Totale | 3,480 | 100.0 | 17,129 | 100.0 | 4.9 | 11.8 | 10.9 | 11.0 | -0.9 | -7.9 | -0.7 | -6.3 |
| Regolarità Posizione secondaria | | | | | | | | | | | | |
| Regolare | 1,864 | 53.6 | 14,997 | 87.6 | 8.0 | 10.5 | 10.7 | 10.8 | 0.1 | 1.2 | 0.3 | 2.8 |
| Irregolare | 1,616 | 46.4 | 2,132 | 12.4 | 1.3 | 13.0 | 12.1 | 12.5 | -0.8 | -6.5 | -0.4 | -3.5 |
| Trimestre | | | | | | | | | | | | |
| Q1 | 985 | 28.3 | 4,271 | 24.9 | 4.3 | 11.4 | 10.4 | 10.8 | -1.0 | -9.1 | -0.6 | -5.0 |
| Q2 | 932 | 26.8 | 4,483 | 26.2 | 4.8 | 12.1 | 11.3 | 11.3 | -0.8 | -6.7 | -0.8 | -6.6 |
| Q3 | 712 | 20.5 | 4,054 | 23.7 | 5.7 | 12.1 | 10.7 | 11.5 | -1.4 | -11.3 | -0.5 | -4.5 |
| Q4 | 851 | 24.5 | 4,321 | 25.2 | 5.1 | 11.6 | 11.0 | 10.5 | -0.6 | -5.4 | -1.1 | -9.6 |
| Genere | | | | | | | | | | | | |
| M | 2,062 | 59.3 | 11,012 | 64.3 | 5.3 | 12.6 | 11.4 | 11.7 | -1.2 | -9.7 | -0.9 | -7.1 |
| F | 1,418 | 40.7 | 6,117 | 35.7 | 4.3 | 10.6 | 9.9 | 9.8 | -0.7 | -6.7 | -0.8 | -7.1 |
| Settore Posizione principale | | | | | | | | | | | | |
| A | 184 | 5.3 | 1,871 | 10.9 | 10.2 | 182.0 | 12.1 | 12.4 | -169.9 | -93.4 | -169.6 | -93.2 |
| B-E | 334 | 9.6 | 2,039 | 11.9 | 6.1 | 8.0 | 11.2 | 11.5 | 3.2 | 40.0 | 3.5 | 43.2 |
| F | 129 | 3.7 | 1,306 | 7.6 | 10.1 | 294.0 | 10.8 | 10.6 | -283.2 | -96.3 | -283.4 | -96.4 |
| G;I | 498 | 14.3 | 2,961 | 17.3 | 5.9 | 9.0 | 11.5 | 10.9 | 2.5 | 28.2 | 1.9 | 21.5 |
| H;J-N | 813 | 23.4 | 3,325 | 19.4 | 4.1 | 21.0 | 11.2 | 11.6 | -9.8 | -46.8 | -9.4 | -44.6 |
| O-S | 1,233 | 35.4 | 4,906 | 28.6 | 4.0 | 128.0 | 9.6 | 10.2 | -118.4 | -92.5 | -117.8 | -92.0 |
| T-U | 289 | 8.3 | 721 | 4.2 | 2.5 | 310.0 | 11.4 | 10.7 | -298.6 | -96.3 | -299.3 | -96.6 |

4.5. Imputazione delle ore lavorate sulla prima posizione – Insieme 4

4.5.1. Imputazione relativa agli occupati di CN con una sola posizione e con segnale ADMIN valido

Per quanto riguarda l'imputazione delle ore lavorate sulla prima posizione, si considera in primo luogo il sottoinsieme degli occupati monoposizione (l'insieme 4 definito in precedenza).

Si procede dapprima con l'imputazione relativa agli occupati di CN con segnale amministrativo validato, a prescindere quindi dallo stato di occupazione rilevato attraverso LFS⁴ (Tavola 30). L'imputazione della variabile C37 viene condotta separatamente per ciascuna fonte al fine di sfruttare al massimo le specificità informative di ciascuna di esse.

⁴ Questi due insiemi possono essere separati: nel 2010 ci sono 511 rk relativi a OCC LFS. Si avrebbe un guadagno in accuratezza solo per il sottoinsieme che non ha segnali in EMEN, INPD e DMAG.

Tavola 30. Occupati di CN dell'insieme 4 e con variabile C37 da imputare, per segnale ADMIN, tipo di occupazione LFS. Anni 2010 e 2011

| Tipo di occupati di CN | Anno 2010 | | | | Anno 2011 | | | |
|------------------------------|-----------|------|-----------|------|-----------|------|-----------|------|
| | Oss. | % | Stime | % | Oss. | % | Stime | % |
| Con segnale ADMIN, di cui: | 10,725 | 83.9 | 966,389 | 84.4 | 10,135 | 83.5 | 931,860 | 85.4 |
| - Occupati RCFL | 511 | 4.0 | 47,177 | 4.1 | 564 | 4.6 | 50,958 | 4.7 |
| - Non occupati RCFL | 10,214 | 79.9 | 919,212 | 80.3 | 9,571 | 78.8 | 880,902 | 80.7 |
| Senza segnale ADMIN, di cui: | 2,063 | 16.1 | 178,679 | 15.6 | 2,006 | 16.5 | 159,948 | 14.6 |
| - Occupati RCFL | 176 | 1.4 | 17,516 | 1.5 | 159 | 1.3 | 12,957 | 1.2 |
| - Non occupati RCFL | 1,887 | 14.8 | 161,163 | 14.1 | 1,847 | 15.2 | 146,990 | 13.5 |
| Totale | 12,788 | 100 | 1,145,068 | 100 | 12,141 | 100 | 1,091,807 | 100 |

Nella procedura di imputazione hot deck fanno parte del gruppo dei donatori tutti i record con la variabile C37 valorizzata, una sola posizione e rilevati attraverso intervista diretta.

La definizione degli strati dipende dalla fonte in modo da sfruttare al meglio tutte le informazioni contenute in ciascuna di esse: nella Tavola 31 sono riportate le variabili utilizzate per la definizione delle celle d'imputazione per ciascuna fonte.

Tavola 31. Variabili identificative degli strati per fonte ADMIN

| Fonte ADMIN | Variabili di stratificazione |
|-------------|--|
| EMEN | Regime orario; trimestre; genere; attività economica |
| INPD | Regime orario; trimestre; genere; natura giuridica dell'istituzione |
| DMAG | Regime orario; trimestre; genere; regione di residenza; attività economica; forma giuridica |
| INDI | Trimestre; genere; attività economica; classe dimensionale impresa; forma giuridica |
| PARA | Trimestre; genere; attività economica; classe dimensionale impresa; forma giuridica |
| PROF | Trimestre; genere; attività economica; classe dimensionale impresa; forma giuridica |
| ENPA | Trimestre; genere; attività economica; classe dimensionale impresa; forma giuridica |
| AUAG | Trimestre; genere; imprendit. agric. professionale (IAP)/coltivat.diretto-caponucleo/familiare; regione di residenza |
| DOME1 | Trimestre; tipo di contratto (colf/badante); classe di ore contrattuali; cittadinanza |
| DOME2 | Trimestre; classe di ore contrattuali; cittadinanza |

Si tratta dunque di variabili strutturali legate all'individuo o all'impresa o alle caratteristiche del contratto di lavoro: la variabile regime orario ha due sole modalità (full-time/part-time)⁵; per quanto riguarda l'attività economica si utilizza la divisione NACE del datore di lavoro; per la forma giuridica le modalità previste sono "società di persone", società di capitali" e "altro"; come misura della classe dimensionale dell'impresa si utilizza la media settimanale del numero di dipendenti di ASIA, articolata in tre classi (0, 1-2, ≥ 3). Per quanto riguarda i rapporti di lavoro domestico, la variabile numero di ore contrattuali è articolata in tre classi: <20; [20, 30); ≥ 30 . In corrispondenza dei

⁵ Per quanto riguarda le fonti EMEN e DMAG, la variabile regime orario viene dedotta dall'indagine LFS se valorizzata, altrimenti dalla fonte amministrativa. Il motivo di tale scelta risiede nel fatto che, se l'informazione proveniente dalle due fonti non coincide, si privilegia il dato d'indagine in quanto ritenuto più affidabile: infatti, da un'analisi condotta sulla concordanza della variabile regime orario proveniente da EMEN e da LFS è emerso che, in caso di discordanza, si riscontra un'elevata incidenza di lavoratori che, pur dichiarando a LFS di svolgere un lavoro *full-time*, nel dato amministrativo risultano svolgere un lavoro *part-time*: tale circostanza fa propendere a favore del dato di indagine, facendo presumere che dietro la modalità assunta dal dato amministrativo si celi un lavoratore full-time parzialmente in nero. Per la fonte INPD, data la natura della fonte stessa, si privilegia il dato amministrativo (De Gregorio – Giordano, "Nero a metà": i dipendenti part-time delle imprese italiane nelle fonti amministrative e nei dati di indagine).

record per i quali risulta mancante anche la variabile “tipo di contratto”, si utilizza un’articolazione in 7 classi: <10; [10, 20); [20, 25); [25, 30); [30, 40); [40, 50); ≥50.

Dai risultati riportati nella Tavola 32 emerge che, in corrispondenza di alcune fonti ADMIN, l’imputazione è stata effettuata in corrispondenza delle partizioni meno fini: il motivo di tale risultato risiede nella esigua numerosità di strato dei corrispondenti donatori e, al contempo, nella prefissata numerosità minima di strato.

Tavola 32. Occupati CN nell’insieme 4 con segnale ADMIN valido, per livello della partizione di imputazione e fonte ADMIN. Anni 2010 e 2011

| Livello della partizione di imputazione | Anno 2010 | | Anno 2011 | |
|--|-----------|---------|-----------|---------|
| | Oss. | Stime | Oss. | Stime |
| EMEN | | | | |
| 4 Regime orario; trim; genere; sezione NACE | 3,990 | 395,980 | 3,624 | 363,987 |
| 3 Regime orario; trim; genere | 70 | 6,698 | 81 | 7,692 |
| Totale EMEN | 4,060 | 402,677 | 3,705 | 371,679 |
| INPD | | | | |
| 4 Regime orario; trim; genere; natura giuridica dell’istituzione | 464 | 35,519 | 427 | 36,451 |
| 3 Regime orario; trim; genere | 36 | 2,629 | - | - |
| 2 Regime orario; trim | - | - | - | - |
| 1 Regime orario | - | - | - | - |
| Totale INPD | 500 | 38,147 | 427 | 36,451 |
| DMAG | | | | |
| 6 Regime orario; trim; genere; regione di residenza; sezione NACE; forma giuridica | 105 | 6,835 | 207 | 13,762 |
| 5 Regime orario; trim; genere; regione di residenza; sezione NACE | 125 | 8,555 | 81 | 5,536 |
| 4 Regime orario; trim; genere; regione di residenza; | 205 | 16,360 | 166 | 11,930 |
| 3 Regime orario; trim; genere | 197 | 16,046 | 210 | 17,744 |
| 2 Regime orario; trim | 2 | 37 | - | - |
| Totale DMAG | 634 | 47,832 | 664 | 48,971 |
| INDI | | | | |
| 5 Trim; genere; sezione NACE; dimensione impresa; forma giuridica | 3,975 | 347,670 | 3,538 | 311,446 |
| 4 Trim; genere; sezione NACE; dimensione impresa | 289 | 27,494 | 346 | 31,233 |
| 3 Trim; genere; sezione NACE | 122 | 9,803 | 364 | 38,133 |
| 2 Trim; genere | 65 | 5,230 | 48 | 4,956 |
| 1 Trim; | - | - | - | - |
| Totale INDI | 4,451 | 390,197 | 4,296 | 385,769 |
| PARA | | | | |
| 5 Trim; genere; sezione NACE; dimensione impresa; forma giuridica | 50 | 4,449 | 79 | 7,010 |
| 4 Trim; genere; sezione NACE; dimensione impresa | 29 | 2,533 | 20 | 2,128 |
| 3 Trim; genere; sezione NACE | 94 | 8,930 | 105 | 8,868 |
| 2 Trim; genere | 197 | 18,239 | 203 | 19,495 |
| 1 Trim; | - | - | - | - |
| Totale PARA | 370 | 34,151 | 407 | 37,500 |
| PROF | | | | |
| 5 Trim; genere; sezione NACE; dimensione impresa; forma giuridica | - | - | - | - |
| 4 Trim; genere; sezione NACE; dimensione impresa | - | - | - | - |
| 3 Trim; genere; sezione NACE | - | - | - | - |
| 2 Trim; genere | - | - | - | - |
| 1 Trim; | 33 | 3,061 | 40 | 3,920 |
| Totale PROF | 33 | 3,061 | 40 | 3,920 |

Tavola 32. Segue Occupati CN nell'insieme 4 con segnale ADMIN valido, per livello della partizione di imputazione e fonte ADMIN. Anni 2010 e 2011

| Livello della partizione di imputazione | Anno 2010 | | Anno 2011 | |
|---|-----------|---------|-----------|---------|
| | Oss. | Stime | Oss. | Stime |
| ENPA | | | | |
| 5 Trim; genere; sezione NACE; dimensione impresa; forma giuridica | - | - | - | - |
| 4 Trim; genere; sezione NACE; dimensione impresa | - | - | - | - |
| 3 Trim; genere; sezione NACE | - | - | - | - |
| 2 Trim; genere | - | - | - | - |
| 1 Trim; | 13 | 1,116 | 21 | 1,985 |
| Totale ENPA | 13 | 1,116 | 21 | 1,985 |
| AUAG | | | | |
| 4 Trim; genere; IAP/coltivat.diretto-caponucleo/familiare; regione di residenza | 69 | 2,880 | 27 | 1,224 |
| 3 Trim; genere; IAP/coltivat.diretto-caponucleo/familiare | 162 | 13,203 | 156 | 11,082 |
| 2 Trim; genere | 126 | 7,609 | 87 | 6,174 |
| 1 Trim; | - | - | - | - |
| Totale AUAG | 357 | 23,692 | 270 | 18,480 |
| DOME1 | | | | |
| 4 Trimestre; colf/badante; classe di ore contrattuali (3); cittadinanza | 266 | 22,625 | 273 | 24,340 |
| 3 Trimestre; colf/badante; classe di ore contrattuali (3) | 19 | 1,293 | 22 | 1,586 |
| 2 Trimestre; colf/badante | 15 | 1,055 | 10 | 1,178 |
| Totale DOME1 | 300 | 24,973 | 305 | 27,104 |
| DOME2 | | | | |
| 4 Trimestre; classe di ore contrattuali (7); cittadinanza | - | - | - | - |
| 3 Trimestre; classe di ore contrattuali (7) | - | - | - | - |
| 2 Trimestre; classe di ore contrattuali (3); | - | - | - | - |
| 1 Trimestre | 7 | 543 | - | - |
| Totale DOME2 | 7 | 543 | - | - |
| TOTALE | 10,725 | 966,389 | 10,135 | 931,860 |

4.5.2. Imputazione relativa agli occupati di CN con una sola posizione e senza segnale ADMIN valido

Rientrano in questa classe gli occupati non regolari di CN con una sola posizione e privi dunque di segnali ADMIN validi. Il metodo di imputazione prescelto è sempre l'*hot-deck* gerarchico sequenziale, dove le celle d'imputazione sono definite attraverso le modalità assunte dalle variabili riportate nella Tavola 33:

Tavola 33. Variabili identificative degli strati per tipologia di occupazione di CN

| |
|--|
| Occupati RCFL |
| Regime orario; trimestre; genere; tipo di posizione (dip. / indip.); sezione NACE (7 cl.); sezione NACE |
| Non occupati RCFL |
| Trimestre; genere; tipo di posizione (dip./indip); sezione NACE (7 cl.); sezione NACE; ripartizione geografica |

L'insieme dei donatori è costituito, per entrambi i gruppi, dagli occupati LFS con una sola posizione, senza copertura amministrativa, caratterizzati dall'aver la variabile C37 valorizzata e rilevata tramite intervista diretta.

Dalla Tavola 34 emerge come nella maggior parte dei casi la procedura d'imputazione venga effettuata al livello di dettaglio più fine.

Tavola 34. Occupati di CN non regolari, per livello della partizione usata per l'imputazione e condizione lavorativa LFS. Anni 2010 e 2011

| Partizione | Regime orario | Trimestre | Genere | Tipo pos. | Settori (7) | Sezione | Ripartizione | Anno 2010 | | Anno 2011 | |
|-------------------|---------------|-----------|--------|-----------|-------------|---------|--------------|-----------|---------|-----------|---------|
| | | | | | | | | Oss. | Stime | Oss. | Stime |
| OCCUPATI RCFL | | | | | | | | | | | |
| 6 | x | x | x | x | x | x | | 146 | 14,806 | 131 | 10,619 |
| 5 | x | x | x | x | x | | | 30 | 2,710 | 28 | 2,338 |
| Totale | | | | | | | | 176 | 17,516 | 159 | 12,957 |
| NON OCCUPATI RCFL | | | | | | | | | | | |
| 6 | | x | x | x | x | x | x | 1,239 | 103,969 | 1,192 | 93,109 |
| 5 | | x | x | x | x | x | | 529 | 45,965 | 554 | 45,599 |
| 4 | | x | x | x | x | | | 97 | 9,138 | 83 | 6,893 |
| 3 | | x | x | x | | | | 22 | 2,091 | 18 | 1,390 |
| Totale | | | | | | | | 1,887 | 161,163 | 1,847 | 146,990 |

4.6. Imputazione delle ore lavorate sulla prima e sulla seconda posizione – Insieme 5

Rientrano nell'insieme 5 gli occupati di CN con più posizioni lavorative in corrispondenza dei quali non si conosce il numero di ore lavorate né sulla prima né sulle seconde posizioni: nel biennio considerato si tratta di circa lo 0.5% di tutti gli occupati di CN.

Al fine di mantenere inalterata la relazione tra le due variabili da imputare, l'imputazione di C37 e di D2 è fatta contestualmente nel senso che, per ciascuna coppia di dati mancanti, il donatore è lo stesso. L'insieme dei donatori prescelto è, pertanto, l'insieme 2 al netto delle interviste *proxy*; il metodo è sempre l'*hot-deck* gerarchico sequenziale.

Anche per questo gruppo si è sfruttata al massimo tutta l'informazione disponibile. Sono state fatti due tipi di imputazioni a seconda della disponibilità della variabile sul regime orario: tale variabile, disponibile per circa un terzo delle osservazioni da imputare, origina da LFS o, in subordine, dalle fonti EMEN, INPD e DMAG. Per le rimanenti osservazioni - circa i due terzi di questo insieme - nella specificazione degli strati si è ommesso il regime orario.

Le variabili usate per la definizione delle celle d'imputazione di questo insieme coincidono con quelle individuate per l'insieme 3, vale a dire: regime orario sulla prima posizione (quando disponibile), trimestre, genere, tipo di posizione (dipendente/indipendente con riferimento alla prima posizione), combinazione del tipo di posizioni (dipendente/indipendente sulla prima e sulla seconda); regolarità o meno della prima posizione, regolarità o meno di entrambe le posizioni e attività economica (due livelli: sezione Nace e aggregazione di sezioni). La Tavola 35 riepiloga lo schema dell'imputazione.

Tavola 35. Occupati di CN dell'insieme 5, per tipologia e partizione usata per l'imputazione. Anni 2010 e 2011

| Partizione | Variabili usate per le partizioni | | | | | | | | Anno 2010 | | Anno 2011 | | |
|---------------------------------|-----------------------------------|-----------|--------|------------|----------------|------------------|----------------------|-------------|-----------|-------|-----------|-------|---------|
| | Regime orario | Trimestre | Genere | Tipo pos.1 | Tipo pos.1 e 2 | Regolarità pos.1 | Regolarità pos.1 e 2 | Settori (7) | Sezioni | Oss. | Stime | Oss. | Stime |
| Insieme 5 - Con regime orario | | | | | | | | | | | | | |
| 9 | X | X | X | X | X | X | X | X | X | - | - | 1 | 85 |
| 8 | X | X | X | X | X | X | X | X | | 19 | 1,669 | 4 | 418 |
| 7 | X | X | X | X | X | X | X | | | 77 | 6,757 | 121 | 10,716 |
| 6 | X | X | X | X | X | X | | | | 88 | 7,671 | 67 | 6,099 |
| 5 | X | X | X | X | X | | | | | 32 | 2,808 | 38 | 3,338 |
| 4 | X | X | X | X | | | | | | 241 | 22,779 | 267 | 24,560 |
| Totale | | | | | | | | | | 457 | 41,684 | 498 | 45,216 |
| Insieme 5 - Senza regime orario | | | | | | | | | | | | | |
| 8 | | X | X | X | X | X | X | X | X | 28 | 2,778 | 34 | 3,115 |
| 7 | | X | X | X | X | X | X | X | | 81 | 6,959 | 49 | 5,384 |
| 6 | | X | X | X | X | X | X | | | 390 | 31,900 | 514 | 46,822 |
| 5 | | X | X | X | X | X | | | | 302 | 24,987 | 305 | 27,455 |
| 4 | | X | X | X | X | | | | | 11 | 1,520 | 19 | 1,638 |
| 3 | | X | X | X | | | | | | - | - | 20 | 1,495 |
| Totale | | | | | | | | | | 812 | 68,143 | 941 | 85,909 |
| Totale Insieme 5 | | | | | | | | | | 1,269 | 109,827 | 1,439 | 131,125 |

4.7. Imputazione delle ore lavorate sulla prima posizione degli occupati di CN con ore sulle seconde posizioni – Insieme 6

Analogamente a quanto si fa nell'insieme 3 per l'imputazione delle ore lavorate nelle seconde posizioni conoscendo le ore lavorate nella prima posizione, nell'insieme 6 si stimano le ore sulle prime posizioni sulla base del valore assunto da D2 e dell'imputazione, tramite *hot-deck* gerarchico sequenziale, del rapporto di coesistenza precedentemente descritto.

Le classi d'imputazione sono definite sulla base delle stesse variabili definite per gli insiemi 3 e 5, prevedendo anche in questo caso la possibilità che la variabile regime orario non sia valorizzata (Tavola 36).

Tavola 36. Occupati di CN dell'insieme 6 per partizione usata per l'imputazione. Anni 2010 e 2011

| Partizione | Variabili delle partizioni | | | | | | | Anno 2010 | | Anno 2011 | |
|------------|----------------------------|-----------|--------|------------|----------------|------------------|----------------------|-----------|-------|-----------|-------|
| | Regime orario | Trimestre | Genere | Tipo pos.1 | Tipo pos.1 e 2 | Regolarità pos.1 | Regolarità pos.1 e 2 | Oss. | Stime | Oss. | Stime |
| 7 | X | X | X | X | X | X | X | 1 | 62 | 1 | 70 |
| 6 | X | X | X | X | X | X | | 4 | 166 | 1 | 154 |
| 5 | X | X | X | X | X | | | 1 | 26 | 1 | 226 |
| Totale | | | | | | | | 6 | 254 | 3 | 450 |

4.8. Effetti delle procedure di correzione e di imputazione

Gli effetti più rilevanti sulla stima delle ore lavorate sono prodotti nell'ambito delle procedure di imputazione sulle prime e soprattutto sulle secondo posizioni. In particolare, con riferimento alle posizioni principali le correzioni per gli eventi di malattia portano alla riduzione di 0.5 ore dell'orario pro capite settimanale e dell'1% del monte ore annuale, variazione che incorpora anche uno 0.4% in più di posizioni principali (Tavola 37). Sulle posizioni secondarie l'effetto sul monte ore è ancora meno marcato a causa del più netto incremento delle posizioni lavorative.

L'effetto della correzione delle proxy è anch'esso piuttosto modesto, trattandosi di 0.2 ore sul pro capite e di una riduzione dello 0.4% del monte ore. La forzatura a zero ore dei non occupati LFS in CIG non ha invece effetti apprezzabili sul monte ore.

L'imputazione in base alle ore abituali lascia invece logicamente invariato il pro capite mentre accresce dello 0.2% il monte ore, dal momento che recupera la misure orarie altrimenti mancanti.

Più importante è stato invece l'effetto dell'imputazione delle ore lavorate sulle prime posizioni (insieme 4): a fronte di una lieve contrazione del pro capite il più netto incremento delle posizioni trascina il monte ore delle prime posizioni in aumento del 4.6%. Decisamente vistoso infine l'effetto sulle seconde posizioni, il cui monte ore cresce di oltre sei volte, con un aumento di oltre 1.2 miliardi del monte ore annuale: l'effetto sul monte ore complessivo sale così oltre il 7%.

Gli effetti dell'imputazione (Tavola 38) sono stati più marcati nel terzo trimestre, fra le donne, nel Mezzogiorno, fra gli stranieri Extra-Ue e nelle classi di età estreme. I pro capite sono generalmente diminuiti e l'aumento del monte ore è stato determinato essenzialmente dal maggiori numero di posizioni lavorative.

Tavola 37. Indicatori relativi agli step della procedura di stima delle ore lavorate per tipo di posizione e anno - Anno 2010

| step | Occupati CN | | monte ore (mln) | | procapite | | variazioni % | |
|-----------------------------|-------------|-------------|-----------------|--------|-----------|-------|--------------|-------|
| | Oss. | Stime (000) | sett | anno | sett | anno | Occupati | Monte |
| Sulle posizioni principali | | | | | | | | |
| 0. rilevato RCFL | 228,988 | 22,031 | 755.0 | 39,260 | 34.3 | 1,782 | | |
| 1. correzioni per malattia | 229,829 | 22,111 | 747.5 | 38,872 | 33.8 | 1,758 | 0.4 | -1.0 |
| 2. correzione per proxy | 229,829 | 22,111 | 744.4 | 38,709 | 33.7 | 1,751 | 0.0 | -0.4 |
| 3. imputazione cig 0 ore | 230,251 | 22,153 | 744.4 | 38,709 | 33.6 | 1,747 | 0.2 | 0.0 |
| 4. imputazione ore abituali | 230,627 | 22,189 | 745.7 | 38,775 | 33.6 | 1,748 | 0.2 | 0.2 |
| 5. imputazione ore finale | 243,776 | 23,357 | 780.6 | 40,589 | 33.4 | 1,738 | 5.3 | 4.6 |
| Sulle posizioni secondarie | | | | | | | | |
| 0. rilevato RCFL | 3,497 | 332 | 3.9 | 204 | 11.8 | 614 | | |
| 1. correzioni per malattia | 3,604 | 342 | 3.9 | 203 | 11.4 | 593 | 3.0 | -0.4 |
| 5. imputazione ore finale | 24,260 | 2,232 | 25.9 | 1,348 | 11.6 | 604 | 569.0 | 561.5 |
| Sul totale delle posizioni | | | | | | | | |
| 0. rilevato RCFL | 228,988 | 22,031 | 759 | 39,464 | 34.4 | 1,791 | | |
| 1. correzioni per malattia | 229,829 | 22,111 | 751 | 39,075 | 34.0 | 1,767 | 0.4 | -1.0 |
| 2. correzione per proxy | 229,829 | 22,111 | 748 | 38,912 | 33.8 | 1,760 | 0.0 | -0.4 |
| 3. imputazione cig 0 ore | 230,251 | 22,153 | 748 | 38,912 | 33.8 | 1,757 | 0.2 | 0.0 |
| 4. imputazione ore abituali | 230,627 | 22,189 | 750 | 38,978 | 33.8 | 1,757 | 0.2 | 0.2 |
| 5. imputazione ore finale | 243,776 | 23,357 | 806 | 41,937 | 34.5 | 1,795 | 5.3 | 7.5 |

Anno 2011

| step | Occupati CN | | monte ore (mln) | | procapite | | variazioni % | |
|-----------------------------|-------------|-------------|-----------------|--------|-----------|-------|--------------|-------|
| | Oss. | Stime (000) | sett | anno | sett | anno | Occupati | Monte |
| Sulle posizioni principali | | | | | | | | |
| 0. rilevato RCFL | 223,411 | 22,123 | 756.4 | 39,335 | 34.2 | 1,778 | | |
| 1. correzioni per malattia | 224,132 | 22,190 | 749.2 | 38,959 | 33.8 | 1,756 | 0.3 | -1.0 |
| 2. correzione per proxy | 224,132 | 22,190 | 745.2 | 38,752 | 33.6 | 1,746 | 0.0 | -0.5 |
| 3. imputazione cig 0 ore | 224,477 | 22,223 | 745.2 | 38,752 | 33.5 | 1,744 | 0.2 | 0.0 |
| 4. imputazione ore abituali | 224,917 | 22,263 | 746.6 | 38,822 | 33.5 | 1,744 | 0.2 | 0.2 |
| 5. imputazione ore finale | 237,737 | 23,415 | 779.4 | 40,530 | 33.3 | 1,731 | 5.2 | 4.4 |
| Sulle posizioni secondarie | | | | | | | | |
| 0. rilevato RCFL | 3,187 | 306 | 3.5 | 181 | 11.4 | 592 | | |
| 1. correzioni per malattia | 3,291 | 316 | 3.5 | 181 | 11.0 | 573 | 2.9 | -0.2 |
| 5. imputazione ore finale | 25,063 | 2,386 | 26.7 | 1,391 | 11.2 | 583 | 623.4 | 593.6 |
| Sul totale delle posizioni | | | | | | | | |
| 0. rilevato RCFL | 223,411 | 22,123 | 760 | 39,516 | 34.4 | 1,786 | | |
| 1. correzioni per malattia | 224,132 | 22,190 | 753 | 39,140 | 33.9 | 1,764 | 0.3 | -1.0 |
| 2. correzione per proxy | 224,132 | 22,190 | 749 | 38,933 | 33.7 | 1,755 | 0.0 | -0.5 |
| 3. imputazione cig 0 ore | 224,477 | 22,223 | 749 | 38,933 | 33.7 | 1,752 | 0.2 | 0.0 |
| 4. imputazione ore abituali | 224,917 | 22,263 | 750 | 39,003 | 33.7 | 1,752 | 0.2 | 0.2 |
| 5. imputazione ore finale | 237,737 | 23,415 | 806 | 41,921 | 34.4 | 1,790 | 5.2 | 7.4 |

Tavola 38. Effetti dell'imputazione per segmenti di occupazione. Anno 2010

| | Posizioni principali | | Posizioni secondarie | | Totale posizioni | | Variaz. % step5-step0 | | |
|---------------------|----------------------|---------------------------|----------------------|---------------------------|------------------|---------------------------|-----------------------|-----------|-------------|
| | 0. rilevato RCFL | 5. imputazione ore finale | 0. rilevato RCFL | 5. imputazione ore finale | 0. rilevato RCFL | 5. imputazione ore finale | Pos. princ. | Pos. sec. | Totale pos. |
| TRIMESTRE | | | | | | | | | |
| | <i>Monte ore</i> | | | | | | | | |
| 1 | 193.8 | 199.2 | 1.1 | 6.0 | 194.9 | 205.2 | 2.8 | 473 | 5.3 |
| 2 | 198.6 | 204.6 | 1.1 | 6.9 | 199.7 | 211.5 | 3.0 | 528 | 5.9 |
| 3 | 172.5 | 180.2 | 0.8 | 6.4 | 173.3 | 186.7 | 4.5 | 688 | 7.7 |
| 4 | 190.2 | 196.5 | 1.0 | 6.4 | 191.1 | 202.9 | 3.3 | 572 | 6.2 |
| | <i>Occupati</i> | | | | | | | | |
| 1 | 5,479 | 5,798 | 92 | 554 | 5,479 | 5,798 | 5.8 | 501 | 5.8 |
| 2 | 5,541 | 5,869 | 90 | 583 | 5,541 | 5,869 | 5.9 | 546 | 5.9 |
| 3 | 5,489 | 5,836 | 68 | 535 | 5,489 | 5,836 | 6.3 | 692 | 6.3 |
| 4 | 5,522 | 5,853 | 82 | 560 | 5,522 | 5,853 | 6.0 | 582 | 6.0 |
| | <i>Pro capite</i> | | | | | | | | |
| 1 | 35.4 | 34.4 | 11.4 | 10.9 | 35.6 | 35.4 | -2.9 | -4.6 | -0.5 |
| 2 | 35.8 | 34.9 | 12.1 | 11.8 | 36.0 | 36.0 | -2.7 | -2.8 | 0.0 |
| 3 | 31.4 | 30.9 | 12.1 | 12.0 | 31.6 | 32.0 | -1.7 | -0.6 | 1.3 |
| 4 | 34.4 | 33.6 | 11.6 | 11.5 | 34.6 | 34.7 | -2.5 | -1.4 | 0.2 |
| GENERE | | | | | | | | | |
| | <i>Monte ore</i> | | | | | | | | |
| 1 | 488.8 | 503.1 | 2.5 | 17.6 | 491.3 | 520.7 | 2.9 | 619 | 6.0 |
| 3 | 266.2 | 277.5 | 1.5 | 8.1 | 267.6 | 285.6 | 4.2 | 454 | 6.7 |
| | <i>Occupati</i> | | | | | | | | |
| 1 | 13,051 | 13,758 | 194 | 1,419 | 13,051 | 13,758 | 5.4 | 631 | 5.4 |
| 3 | 8,980 | 9,599 | 138 | 812 | 8,980 | 9,599 | 6.9 | 489 | 6.9 |
| | <i>Pro capite</i> | | | | | | | | |
| 1 | 37.5 | 36.6 | 12.6 | 12.4 | 37.6 | 37.8 | -2.4 | -1.7 | 0.5 |
| 3 | 29.6 | 28.9 | 10.6 | 10.0 | 29.8 | 29.8 | -2.5 | -5.9 | -0.2 |
| RIPARTIZIONE | | | | | | | | | |
| | <i>Monte ore</i> | | | | | | | | |
| 1 | 222.4 | 225.1 | 1.3 | 7.5 | 223.7 | 232.6 | 1.2 | 492 | 4.0 |
| 2 | 166.5 | 168.4 | 0.8 | 5.5 | 167.3 | 173.9 | 1.1 | 608 | 3.9 |
| 3 | 157.7 | 161.4 | 0.9 | 5.2 | 158.6 | 166.6 | 2.3 | 475 | 5.0 |
| 4 | 141.0 | 154.0 | 0.6 | 4.9 | 141.6 | 158.9 | 9.2 | 725 | 12.2 |
| 5 | 67.4 | 71.8 | 0.4 | 2.6 | 67.7 | 74.4 | 6.6 | 609 | 9.8 |
| | <i>Occupati</i> | | | | | | | | |
| 1 | 6,518 | 6,784 | 107 | 612 | 6,518 | 6,784 | 4.1 | 473 | 4.1 |
| 2 | 4,862 | 5,049 | 68 | 481 | 4,862 | 5,049 | 3.9 | 611 | 3.9 |
| 3 | 4,642 | 4,857 | 80 | 476 | 4,642 | 4,857 | 4.6 | 493 | 4.6 |
| 4 | 4,037 | 4,520 | 48 | 446 | 4,037 | 4,520 | 12.0 | 839 | 12.0 |
| 5 | 1,972 | 2,146 | 30 | 217 | 1,972 | 2,146 | 8.9 | 626 | 8.9 |
| | <i>Pro capite</i> | | | | | | | | |
| 1 | 34.1 | 33.2 | 11.9 | 12.3 | 34.3 | 34.3 | -2.8 | 3.3 | -0.1 |
| 2 | 34.2 | 33.3 | 11.5 | 11.5 | 34.4 | 34.4 | -2.6 | -0.5 | 0.1 |
| 3 | 34.0 | 33.2 | 11.3 | 10.9 | 34.2 | 34.3 | -2.2 | -3.1 | 0.4 |
| 4 | 34.9 | 34.1 | 12.6 | 11.1 | 35.1 | 35.2 | -2.5 | -12.1 | 0.2 |
| 5 | 34.2 | 33.4 | 12.2 | 11.9 | 34.3 | 34.6 | -2.1 | -2.3 | 0.9 |

Tavola 38. Segue Effetti dell'imputazione per segmenti di occupazione. Anno 2010

| | Posizioni principali | | Posizioni secondarie | | Totale posizioni | | Variaz. % step5-step0 | | |
|-----------------------|----------------------|---------------------------|----------------------|---------------------------|------------------|---------------------------|-----------------------|-----------|-------------|
| | 0. rilevato RCFL | 5. imputazione ore finale | 0. rilevato RCFL | 5. imputazione ore finale | 0. rilevato RCFL | 5. imputazione ore finale | Pos. princ. | Pos. sec. | Totale pos. |
| CITTADINANZA | | | | | | | | | |
| | <i>Monte ore</i> | | | | | | | | |
| 1 | 695.9 | 718.4 | 3.4 | 23.8 | 699.2 | 742.2 | 3.2 | 607 | 6.1 |
| 2 | 19.1 | 19.8 | 0.2 | 0.5 | 19.3 | 20.2 | 3.6 | 144 | 4.9 |
| 3 | 40.0 | 42.3 | 0.4 | 1.5 | 40.4 | 43.8 | 5.6 | 311 | 8.5 |
| | <i>Occupati</i> | | | | | | | | |
| 1 | 20,282 | 21,486 | 285 | 2,053 | 20,282 | 21,486 | 5.9 | 620 | 5.9 |
| 2 | 564 | 597 | 15 | 45 | 564 | 597 | 5.7 | 196 | 5.7 |
| 3 | 1,185 | 1,274 | 32 | 133 | 1,185 | 1,274 | 7.5 | 320 | 7.5 |
| | <i>Pro capite</i> | | | | | | | | |
| 1 | 34.3 | 33.4 | 11.8 | 11.6 | 34.5 | 34.5 | -2.5 | -1.7 | 0.2 |
| 2 | 33.9 | 33.2 | 12.1 | 10.0 | 34.2 | 33.9 | -2.0 | -17.4 | -0.8 |
| 3 | 33.8 | 33.2 | 11.9 | 11.6 | 34.1 | 34.4 | -1.7 | -2.1 | 0.9 |
| CLASSE DI ETA' | | | | | | | | | |
| | <i>Monte ore</i> | | | | | | | | |
| 2 | 39.6 | 43.1 | 0.1 | 0.7 | 39.7 | 43.8 | 8.9 | 436 | 10.2 |
| 3 | 162.4 | 169.8 | 0.8 | 4.9 | 163.2 | 174.7 | 4.5 | 540 | 7.0 |
| 4 | 240.8 | 245.3 | 1.4 | 8.3 | 242.2 | 253.6 | 1.9 | 481 | 4.7 |
| 5 | 210.5 | 214.1 | 1.1 | 7.4 | 211.5 | 221.5 | 1.7 | 588 | 4.7 |
| 6 | 89.7 | 93.9 | 0.5 | 3.6 | 90.1 | 97.5 | 4.7 | 675 | 8.2 |
| 7 | 10.5 | 12.4 | 0.0 | 0.6 | 10.6 | 13.0 | 18.0 | 1,624 | 23.6 |
| 8 | 1.6 | 2.0 | 0.0 | 0.2 | 1.6 | 2.1 | 23.9 | 2,464 | 33.2 |
| | <i>Occupati</i> | | | | | | | | |
| 2 | 1,174 | 1,315 | 10 | 56 | 1,174 | 1,315 | 11.9 | 443 | 11.9 |
| 3 | 4,714 | 5,077 | 66 | 436 | 4,714 | 5,077 | 7.7 | 556 | 7.7 |
| 4 | 7,011 | 7,328 | 125 | 737 | 7,011 | 7,328 | 4.5 | 489 | 4.5 |
| 5 | 6,111 | 6,368 | 89 | 638 | 6,111 | 6,368 | 4.2 | 618 | 4.2 |
| 6 | 2,652 | 2,831 | 38 | 300 | 2,652 | 2,831 | 6.8 | 698 | 6.8 |
| 7 | 316 | 374 | 3 | 52 | 316 | 374 | 18.2 | 1,560 | 18.2 |
| 8 | 53 | 64 | 0 | 11 | 53 | 64 | 20.8 | 2,154 | 20.8 |
| | <i>Pro capite</i> | | | | | | | | |
| 2 | 33.7 | 32.8 | 12.2 | 12.0 | 33.8 | 33.3 | -2.7 | -1.3 | -1.5 |
| 3 | 34.5 | 33.4 | 11.6 | 11.3 | 34.6 | 34.4 | -3.0 | -2.5 | -0.6 |
| 4 | 34.3 | 33.5 | 11.4 | 11.3 | 34.5 | 34.6 | -2.5 | -1.5 | 0.2 |
| 5 | 34.4 | 33.6 | 12.2 | 11.7 | 34.6 | 34.8 | -2.4 | -4.2 | 0.5 |
| 6 | 33.8 | 33.2 | 12.3 | 12.0 | 34.0 | 34.4 | -1.9 | -3.0 | 1.3 |
| 7 | 33.3 | 33.2 | 11.7 | 12.1 | 33.4 | 34.9 | -0.2 | 3.9 | 4.6 |
| 8 | 29.9 | 30.7 | 12.5 | 14.3 | 30.1 | 33.2 | 2.5 | 13.8 | 10.3 |

4.9. Primi commenti sulla stima dell'input di lavoro: posizioni lavorative e ore lavorate regolari e non regolari

Premessa

A titolo esemplificativo e con l'obiettivo di mostrare e verificare le potenzialità investigative del campione integrato LFS-ADMIN vengono qui brevemente commentati alcuni risultati che emergono dalla stima delle posizioni lavorative e delle ore lavorate ottenuti attraverso la base dati integrata LFS-ADMIN_JOBSh.

Si tratta di database in cui il singolo record è costituito da una posizione lavorativa: a ciascuna posizione sono associate numerose variabili che riguardano sia il singolo occupato cui la posizione afferisce sia alcune caratteristiche qualitative e quantitative della posizione. I dati che commentiamo sono riportati all'universo con il sistema di ponderazione basato sui pesi censuari prima della correzione post-censuaria ottenuta tramite la rilevazione PES. Si tratta perciò di risultati sperimentali e pertanto non ha senso il confronto con altre stime già diffuse dall'istituto.

L'obiettivo del commento è quello di esplicitare il quadro delle potenzialità che la nuova metodologia offre soprattutto del segmento dell'occupazione non regolare, con particolare riferimento al numero delle posizioni e alla stima delle ore lavorate: esso ha pertanto anche finalità di controllo. Si tratta di risultati di analisi univariate condotte sulle principali variabili strutturali: il genere, il trimestre, l'età, il territorio, il settore di attività economica, la cittadinanza, il tipo di posizione, lo stato civile, il titolo di studio, la composizione della famiglia, l'ordine della posizione. La Tavola 39 in coda al paragrafo riporta alcuni dati.

Aspetti generali

Nel 2011 la nuova procedura prodotta dal GdL stima un input di lavoro non regolare pari a circa 3,62 miliardi di ore lavorate riconducibili a circa 2,6 milioni di posizioni lavorative non regolari. Per il 2010 le ore lavorate sono 3,78 miliardi, originate da 2,7 milioni di posizioni lavorative.

Rispetto al totale delle ore lavorate, quelle non regolari rappresentano circa il 9% del totale (un po' meno nel 2011): in ambedue gli anni l'incidenza delle ore lavorate è di circa un punto percentuale inferiore a quella delle posizioni lavorative. Questa circostanza è determinata dal fatto che i pro capite orari delle posizioni non regolari (o più propriamente le ore medie lavorate per posizione) sono sensibilmente più bassi rispetto a quelli delle posizioni regolari: l'entità di questa differenza è pari al 15% circa, corrispondenti grossomodo a 4 ore settimanali e oltre 200 ore annuali.

Fra il 2010 e il 2011, le stime evidenziano una riduzione delle ore lavorate non regolari del 4,3% di quasi un punto percentuale superiore alla riduzione delle corrispondenti posizioni lavorative (-3,4%). La riduzione dell'input di lavoro non regolare è di circa 162 milioni di ore ed è dovuta in massima parte alla riduzione delle posizioni lavorative mentre la riduzione del pro capite (di 13 ore su base annua) incide per un quinto della riduzione complessiva.

Nel caso delle posizioni regolari si assiste a un fenomeno almeno in parte analogo: in questo caso le ore lavorate crescono, ma la loro crescita è l'effetto combinato di un aumento delle posizioni lavorative regolari e di una riduzione del corrispondente pro capite orario (-17 ore).

Il genere

La componente maschile assorbe circa il 65% delle ore lavorate complessive. Nel caso di quelle non regolari questa quota però si riduce al 61% (pari a 2,3 miliardi di ore). Ciò è dovuto al maggiore tasso di irregolarità delle donne.

In particolare presso la componente femminile l'incidenza delle posizioni non regolari supera di oltre due punti percentuali l'analogo indicatore misurato su quella maschile: questo effetto viene in parte mitigato dal più basso pro capite femminile, che su base annua è di oltre 200 ore inferiore a quello degli uomini.

La riduzione di 162 milioni di ore dell'input non regolare misurata fra il 2010 e il 2011 è imputabile principalmente agli uomini (-134 milioni di ore, pari al -5,8% contro il -1,9% femminile). In particolare la componente maschile dà conto dell'80% circa della riduzione. In ambedue i casi è la riduzione nel numero delle posizioni non regolari a determinare l'andamento: mentre per gli uomini esso è accompagnata anche da una riduzione del pro capite di circa 30 ore, per le donne il pro capite orario delle posizioni non regolari aumenta leggermente.

L'età

Il grosso dell'input di lavoro non regolare si concentra nelle stesse classi di età in cui si accumula quello regolare: di fatto oltre l'81% delle ore regolari sono lavorate da individui in età compresa

fra 25 e 54 anni. Tuttavia, nella componente non regolare sono maggiormente rappresentate le classi di età estreme. Fra i giovani al di sotto dei 25 anni ad esempio si colloca più del 10% delle posizioni non regolari mentre l'incidenza di quelle regolari non raggiunge il 5%. Un altro 5% di posizioni non regolari sono originate da occupati con almeno 65 anni, una concentrazione assai maggiore rispetto a quella registrata per le posizioni regolari (meno del 2%).

Ovvio corollario di queste evidenze è il fatto che il tasso di irregolarità delle classi di età estreme schizza al 20% o oltre: questo tasso ha un chiaro andamento "a U" con un minimo fra i 35 e i 54 anni.

Da notare tuttavia che sono proprio le classi di età estreme a evidenziare pro capite orari più sostenuti per le posizioni non regolari. Fra i più giovani questi pro capite, misurati su base annua, sono superiori di oltre 100 ore rispetto alle classi di età centrali: questa circostanza, peraltro, si presenta anche nel segmento regolare, sia pure con minore evidenza. Fra i lavoratori più anziani, se da un lato è vero che il pro capite non regolare è relativamente alto, dall'altro la stessa cosa non accade con riferimento all'input di lavoro regolare.

La riduzione delle ore lavorate non regolari registrata nel 2011 è quasi completamente imputabile alle classi di età giovanili e in particolare agli occupati con meno di 35 anni: questi segmenti del resto riportano anche una riduzione ancorché più modesta della parte regolare dell'input di lavoro. Sia per gli anziani che per i più giovani la riduzione è stata determinata essenzialmente dal venir meno delle posizioni lavorative più che dalla riduzione del pro capite orario.

Il territorio

Nel Mezzogiorno è concentrato poco più di un quarto delle ore lavorate regolari e quasi la metà delle ore lavorate non regolari. Tutte le regioni del Mezzogiorno e il Lazio hanno almeno il 10% delle ore lavorate di tipo non regolare, con punte che sfiorano il 20% in Campania e Calabria. Nel resto del Centro e del Nord il fenomeno appare più contenuto e omogeneo. La Lombardia, date le dimensioni, rappresenta pur sempre oltre il 12% delle posizioni (seconda solo alla Campania) e l'11% delle ore non regolari (seconda alla Campania e al Lazio) sebbene abbia un tasso di irregolarità inferiore al 6%.

I pro capite orari sono in genere più elevati nel Mezzogiorno, e ciò avviene soprattutto nel caso delle posizioni lavorative non regolari: in tutte le regioni essi superano le 1.400 ore annuali, circa 100 ore in media in più rispetto al Centro-Nord. Nel segmento regolare vi è invece una notevole omogeneità.

Oltre tre quarti della riduzione delle ore non regolari registrata nel 2011 proviene da tre regioni: Lazio, Campania e Sicilia, mentre nel complesso delle regioni del Nord il fenomeno è rimasto abbastanza stabile. La contrazione del Mezzogiorno è essenzialmente determinata dalla contrazione delle posizioni non regolari, mentre la riduzione dei pro capite ha inciso per meno del 20% nella riduzione dell'input non regolare.

La cittadinanza

I cittadini italiani esprimono nel 2011 il 92% delle ore lavorate nelle posizioni regolari e l'82% delle ore lavorate in quelle non regolari. Nel complesso l'input di lavoro non regolare degli stranieri residenti è di poco inferiore ai 700 milioni di ore a fronte di circa 3 miliardi di ore regolari (nel 2011): il tasso di irregolarità supera il 20% fra i cittadini Ue e è appena al di sotto di tale soglia per quelli extra-Ue⁶.

La caratteristica delle ore lavorate dai cittadini stranieri riguarda proprio il segmento non regolare, il quale evidenzia pro capite non troppo dissimili da quelli del segmento regolare (100 ore in meno su base annua) ma soprattutto decisamente superiori a quelli dei cittadini italiani (di 100-150

⁶ Tali risultati si riferiscono esclusivamente al campione LFS-ADMIN, al netto quindi delle integrazioni degli stranieri basate sui permessi di soggiorno e altre stime.

ore). Perciò per gli stranieri non sembra sussistere un divario così ampio fra pro capite orario regolare e non regolare.

Ulteriori differenze fra italiani e stranieri riguardano la dinamica 2010-2011. Il già citato calo dell'input di lavoro non regolare è unicamente dovuto alla componente italiana, mentre anzi quella straniera ha registrato un incremento. Se da un lato la componente italiana ha visto contrarsi quasi esclusivamente le posizioni non regolari, quella straniera evidenzia il fenomeno opposto: sono aumentate leggermente le posizioni lavorative non regolari e sono un po' diminuiti invece i pro capite. L'andamento relativo all'input di lavoro regolare, del resto, è stato pressoché analogo.

Il tipo di occupazione

Le posizioni dipendenti rappresentano oltre due terzi delle posizioni lavorative e delle ore lavorate regolari e una quota inferiore, pari al 60% circa, di quelle non regolari. Il tasso di irregolarità è infatti maggiore fra le posizioni indipendenti (intorno all'11%) rispetto a quelle dipendenti (8% circa).

I pro capite orari non si differenziano significativamente e per entrambe le tipologie di posizione rivelano un livello più basso associato alle posizioni non regolari.

E' la dinamica 2010-2011 a differenziare invece le due tipologie. Due terzi della riduzione delle ore lavorate non regolari provengono infatti dagli indipendenti e sono state determinate quasi esclusivamente dalla riduzione di posizioni lavorative a parità di pro capite. Per i dipendenti, viceversa, la riduzione del 2,5% delle ore lavorate non regolari è in parte dovuta anche alla riduzione del pro capite orario.

Il settore di attività economica

Agricoltura, costruzioni, pubblici esercizi, commercio e manifattura esprimono circa la metà delle posizioni non regolari e oltre il 55% delle ore lavorate non regolari: i primi tre anche con un notevole divario rispetto alle analoghe quote misurate sull'input regolare.

I tassi di irregolarità più elevati, intorno o oltre il 20%, si riscontrano nell'agricoltura, nell'intrattenimento, nei servizi alla persona e in quelli domestici. Tassi oltre il 10% si riscontrano inoltre nelle costruzioni e nei pubblici esercizi.

Se si escludono istruzione (P), informazione (J) e servizi domestici (T), in tutti gli altri settori le ore lavorate non regolari sono diminuite fra il 2010 e il 2011, con particolare evidenza nelle costruzioni e nel commercio. Le diminuzioni sono state quasi sempre determinate da un minor numero di posizioni lavorative. Nei servizi domestici, viceversa, è proprio il numero delle posizioni lavorative ad avere generato l'aumento delle ore lavorate non regolari (286 milioni nel 2011), dal momento che i pro capite orari sono rimasti pressoché invariati.

Il trimestre

A livello trimestrale non si notano peculiarità particolari se non con riferimento alla dinamica 2010-2011. Il terzo trimestre è infatti l'unico a non registrare riduzioni dell'input non regolare e ciò a causa di un incremento dei pro capite orari che ha compensato la riduzione di posizioni lavorative. D'altra parte anche gli altri trimestri evidenziano una certa eterogeneità delle componenti che hanno determinato il calo del 2011: nel primo trimestre sia il pro capite sia le posizioni, nel secondo prevalentemente le posizioni, nel quarto soprattutto le posizioni.

Lo stato civile

Lo stato civile interagisce con l'età ma rivela altri aspetti, a cominciare dalla forte presenza di ore lavorate non regolari riconducibili a celibi/nubili, i cui tassi di irregolarità superano il 10%. Appare evocativa la quota particolarmente elevata di ore lavorate non regolari riconducibili ai separati, che evidenziano anche un elevatissimo tasso di crescita.

L'ordine della posizione

Le posizioni secondarie sono più del 13% del totale delle posizioni non regolari, tuttavia rappresentano solo il 5.5% delle ore non regolari. La loro incidenza è comunque maggiore nel segmento non regolare.

Il tasso di irregolarità delle posizioni secondarie è pari al 15%, circa il doppio rispetto alle posizioni principali.

Il numero di componenti della famiglia di fatto

Si nota una maggiore concentrazione di ore e posizioni non regolari presso unità familiari mononucleari o, al contrario, con molti componenti. In questi segmenti i tassi di irregolarità superano il 10% e il pro capite di ore lavorate è superiore di 150-200 ore rispetto alle altre tipologie di nuclei familiari, diversamente da quanto avviene per l'input regolare.

Dal punto di vista della dinamica le famiglie mononucleari si differenziano perché il loro input di lavoro, regolare e non regolare, cresce fra il 2010 e il 2011 in misura significativa; viceversa, sembra evidenziarsi un fenomeno inverso al crescere del numero dei componenti, soprattutto con riferimento alle posizioni non regolari.

Il titolo di studio

Il titolo di studio evidenzia un addensarsi dei non regolari negli strati inferiori, tuttavia la metà di posizioni e ore non regolari è attribuibile a diplomati. Da notare sempre che negli strati inferiori i pro capite orari sono decisamente più elevati.

Tavola 39. Tassi di irregolarità, ore lavorate pro capite annuali e variazioni 2010-2011, per modalità di alcune variabili strutturali - Anni 2010 e 2011

| | | Quota Non regolari | | | | Ore lavorate pro capite | | | | Variazioni % 2010-2011 | | | | | |
|---------------|----------------------|--------------------|------|--------------|------|-------------------------|-------|--------------|-------|------------------------|------|----------|--------------|------|----------|
| | | Posizioni | | Ore lavorate | | Regolari | | Non regolari | | Posizioni | | | Ore lavorate | | |
| | | 2010 | 2011 | 2010 | 2011 | 2010 | 2011 | 2010 | 2011 | Totali | Reg. | Non reg. | Totali | Reg. | Non reg. |
| TOTALE | | 10,4 | 9,9 | 9,0 | 8,6 | 1.652 | 1.635 | 1.410 | 1.397 | 1,0 | 1,5 | -3,4 | 0,0 | 0,4 | -4,3 |
| Genere | M | 9,5 | 9,0 | 8,5 | 8,0 | 1.790 | 1.767 | 1.580 | 1.553 | 0,7 | 1,3 | -4,2 | -0,5 | 0,0 | -5,8 |
| | F | 11,7 | 11,2 | 9,9 | 9,7 | 1.447 | 1.438 | 1.208 | 1.214 | 1,4 | 1,9 | -2,4 | 0,9 | 1,2 | -1,9 |
| Classi di età | 15-24 | 22,0 | 20,0 | 19,9 | 18,0 | 1.706 | 1.710 | 1.503 | 1.504 | -4,2 | -1,8 | -12,8 | -3,8 | -1,6 | -12,7 |
| | 25-34 | 12,0 | 11,8 | 10,8 | 10,3 | 1.664 | 1.642 | 1.474 | 1.413 | -2,0 | -1,8 | -3,6 | -3,6 | -3,1 | -7,6 |
| | 35-44 | 8,5 | 8,0 | 7,1 | 6,8 | 1.649 | 1.635 | 1.346 | 1.380 | -0,3 | 0,2 | -6,2 | -0,8 | -0,6 | -3,8 |
| | 45-54 | 8,1 | 7,9 | 6,8 | 6,6 | 1.652 | 1.637 | 1.372 | 1.355 | 3,7 | 3,9 | 1,0 | 2,8 | 3,0 | -0,2 |
| | 55-64 | 10,0 | 9,8 | 8,7 | 8,6 | 1.628 | 1.600 | 1.387 | 1.374 | 5,8 | 6,0 | 4,0 | 4,0 | 4,1 | 2,9 |
| | 65-74 | 23,4 | 22,5 | 21,4 | 21,4 | 1.621 | 1.556 | 1.446 | 1.456 | 2,7 | 3,9 | -1,2 | -0,3 | -0,3 | -0,5 |
| | 75 + | 36,7 | 36,2 | 35,9 | 34,3 | 1.455 | 1.506 | 1.404 | 1.390 | -7,8 | -7,0 | -9,2 | -6,0 | -3,7 | -10,1 |
| Cittadinanza | ITA | 9,6 | 9,0 | 8,2 | 7,7 | 1.655 | 1.635 | 1.386 | 1.373 | 0,2 | 0,9 | -6,1 | -0,9 | -0,4 | -6,9 |
| | UE | 22,8 | 22,4 | 21,7 | 21,0 | 1.650 | 1.659 | 1.554 | 1.522 | 14,8 | 15,3 | 13,1 | 14,9 | 16,0 | 10,8 |
| | Extra-UE | 17,9 | 18,4 | 17,4 | 17,4 | 1.601 | 1.609 | 1.545 | 1.501 | 8,2 | 7,5 | 11,2 | 8,1 | 8,1 | 8,1 |
| Stato civile | Celibe/nub. | 13,8 | 13,0 | 12,2 | 11,4 | 1.698 | 1.687 | 1.476 | 1.456 | 1,4 | 2,3 | -4,4 | 0,7 | 1,6 | -5,7 |
| | Coniug. | 8,2 | 7,8 | 6,9 | 6,6 | 1.639 | 1.619 | 1.358 | 1.353 | 0,0 | 0,5 | -5,6 | -1,1 | -0,7 | -6,0 |
| | Sep. di fatto | 16,4 | 18,2 | 15,8 | 17,1 | 1.650 | 1.634 | 1.573 | 1.515 | 9,0 | 6,7 | 20,8 | 7,4 | 5,7 | 16,4 |
| | Sep. legalm. | 11,8 | 12,3 | 10,6 | 10,7 | 1.591 | 1.556 | 1.411 | 1.331 | 4,0 | 3,5 | 8,1 | 1,3 | 1,2 | 2,0 |
| | Divorz. | 10,1 | 9,9 | 8,3 | 8,0 | 1.601 | 1.574 | 1.292 | 1.257 | 9,7 | 9,9 | 7,4 | 7,8 | 8,1 | 4,5 |
| | Vedovo/a | 15,7 | 15,7 | 13,6 | 14,6 | 1.563 | 1.528 | 1.321 | 1.403 | 2,1 | 2,1 | 2,3 | 1,0 | -0,2 | 8,6 |
| | Comp. famiglia fatto | 1 | 14,4 | 14,3 | 13,5 | 13,3 | 1.690 | 1.674 | 1.577 | 1.546 | 7,0 | 7,1 | 6,4 | 5,8 | 6,1 |
| | 2 | 9,9 | 9,8 | 8,2 | 8,1 | 1.652 | 1.628 | 1.328 | 1.330 | 1,1 | 1,3 | -0,6 | -0,2 | -0,2 | -0,4 |
| | 3 | 8,9 | 8,7 | 7,4 | 7,1 | 1.637 | 1.628 | 1.343 | 1.314 | 0,6 | 0,8 | -1,5 | 0,0 | 0,3 | -3,7 |
| | 4 | 9,6 | 8,7 | 8,3 | 7,6 | 1.647 | 1.625 | 1.393 | 1.391 | 0,0 | 1,0 | -9,3 | -1,1 | -0,4 | -9,5 |
| | 5 + | 13,6 | 12,5 | 12,3 | 11,3 | 1.676 | 1.653 | 1.492 | 1.480 | -1,9 | -0,6 | -9,8 | -3,0 | -2,0 | -10,5 |

Tavola 39. Segue Tassi di irregolarità, ore lavorate pro capite annuali e variazioni 2010-2011, per modalità di alcune variabili strutturali - Anni 2010 e 2011

| | | Quota Non regolari | | | | Ore lavorate pro capite | | | | Variazioni % 2010-2011 | | | | | |
|------------------|----------------------|--------------------|------|--------------|------|-------------------------|-------|--------------|-------|------------------------|-------|----------|--------------|-------|----------|
| | | Posizioni | | Ore lavorate | | Regolari | | Non regolari | | Posizioni | | | Ore lavorate | | |
| | | 2010 | 2011 | 2010 | 2011 | 2010 | 2011 | 2010 | 2011 | Totali | Reg. | Non reg. | Totali | Reg. | Non reg. |
| TOTALE | | 10,4 | 9,9 | 9,0 | 8,6 | 1.652 | 1.635 | 1.410 | 1.397 | 1,0 | 1,5 | -3,4 | 0,0 | 0,4 | -4,3 |
| Titolo di studio | Nessun titolo | 29,2 | 23,4 | 28,4 | 23,5 | 1.732 | 1.689 | 1.660 | 1.696 | 12,0 | 21,2 | -10,4 | 10,6 | 18,2 | -8,5 |
| | Licenza elem. | 18,3 | 19,0 | 16,2 | 16,7 | 1.674 | 1.656 | 1.441 | 1.416 | -7,1 | -7,9 | -3,4 | -8,3 | -8,9 | -5,0 |
| | Licenza media | 11,3 | 11,1 | 10,0 | 9,6 | 1.711 | 1.696 | 1.498 | 1.450 | 0,0 | 0,3 | -2,1 | -1,1 | -0,6 | -5,2 |
| | Dipl. 2-3 anni | 7,6 | 7,4 | 6,2 | 6,2 | 1.691 | 1.675 | 1.355 | 1.390 | 3,2 | 3,4 | 0,2 | 2,5 | 2,4 | 2,8 |
| | Dipl. 4-5 anni | 8,8 | 8,2 | 7,5 | 7,0 | 1.662 | 1.643 | 1.392 | 1.378 | 1,9 | 2,5 | -4,4 | 0,9 | 1,4 | -5,3 |
| | Accademie | 16,8 | 17,5 | 13,5 | 14,3 | 1.286 | 1.200 | 993 | 944 | -14,3 | -14,9 | -10,9 | -19,9 | -20,6 | -15,2 |
| | Dipl. univ | 8,8 | 8,4 | 7,8 | 7,0 | 1.488 | 1.492 | 1.295 | 1.226 | -16,4 | -16,0 | -20,0 | -16,5 | -15,8 | -24,3 |
| | Laurea triennale | 14,2 | 11,6 | 11,4 | 9,7 | 1.582 | 1.527 | 1.230 | 1.256 | 8,4 | 11,7 | -11,8 | 5,8 | 7,8 | -9,9 |
| | Laurea e post-laurea | 9,7 | 9,3 | 8,2 | 8,2 | 1.513 | 1.497 | 1.264 | 1.309 | 2,8 | 3,2 | -1,5 | 2,1 | 2,1 | 2,1 |
| Trimestre | 1 | 10,7 | 10,2 | 9,2 | 8,7 | 1.698 | 1.675 | 1.426 | 1.393 | 1,6 | 2,2 | -3,0 | 0,2 | 0,8 | -5,3 |
| | 2 | 10,3 | 9,9 | 9,1 | 8,6 | 1.716 | 1.703 | 1.482 | 1.456 | 1,0 | 1,5 | -3,0 | 0,3 | 0,7 | -4,7 |
| | 3 | 10,4 | 10,1 | 9,1 | 9,1 | 1.535 | 1.514 | 1.320 | 1.342 | 0,9 | 1,2 | -1,9 | -0,2 | -0,2 | -0,3 |
| | 4 | 10,1 | 9,5 | 8,7 | 8,2 | 1.659 | 1.645 | 1.413 | 1.398 | 0,5 | 1,2 | -5,6 | -0,3 | 0,3 | -6,6 |
| Posizione | Principale | 9,9 | 9,6 | 8,8 | 8,4 | 1.764 | 1.758 | 1.546 | 1.530 | 0,3 | 0,6 | -3,2 | -0,1 | 0,3 | -4,2 |
| | Secondarie | 14,9 | 13,2 | 15,7 | 14,3 | 534 | 511 | 566 | 558 | 8,0 | 10,2 | -4,4 | 3,7 | 5,5 | -5,7 |
| Settore istituz. | ISP | 0,0 | 0,0 | 0,0 | 0,0 | 1.446 | 1.417 | | | 3,2 | 3,2 | | 1,2 | 1,2 | |
| | S13 | 0,0 | 0,0 | 0,0 | 0,0 | 1.455 | 1.449 | | | 0,2 | 0,2 | | -0,2 | -0,2 | |
| | Altro | 12,5 | 12,0 | 10,6 | 10,2 | 1.700 | 1.680 | 1.410 | 1.397 | 1,0 | 1,7 | -3,4 | 0,0 | 0,5 | -4,3 |
| Tipo posizione | Dip | 9,1 | 9,0 | 8,0 | 7,8 | 1.633 | 1.630 | 1.414 | 1.395 | 0,6 | 0,8 | -1,1 | 0,3 | 0,6 | -2,5 |
| | Indip | 13,2 | 12,0 | 11,1 | 10,4 | 1.697 | 1.646 | 1.405 | 1.400 | 1,9 | 3,2 | -6,8 | -0,7 | 0,1 | -7,0 |

Tavola 39. Segue Tassi di irregolarità, ore lavorate pro capite annuali e variazioni 2010-2011, per modalità di alcune variabili strutturali - Anni 2010 e 2011

| | | Quota Non regolari | | | | Ore lavorate pro capite | | | | Variazioni % 2010-2011 | | | | | |
|---------|-----|--------------------|------|--------------|------|-------------------------|-------|--------------|-------|------------------------|------|----------|--------------|------|----------|
| | | Posizioni | | Ore lavorate | | Regolari | | Non regolari | | Posizioni | | | Ore lavorate | | |
| | | 2010 | 2011 | 2010 | 2011 | 2010 | 2011 | 2010 | 2011 | Totali | Reg. | Non reg. | Totali | Reg. | Non reg. |
| TOTALE | | 10,4 | 9,9 | 9,0 | 8,6 | 1.652 | 1.635 | 1.410 | 1.397 | 1,0 | 1,5 | -3,4 | 0,0 | 0,4 | -4,3 |
| Regione | PIE | 7,5 | 7,0 | 5,9 | 5,4 | 1.654 | 1.632 | 1.262 | 1.237 | 1,6 | 2,2 | -6,2 | 0,3 | 0,9 | -8,0 |
| | VDA | 8,3 | 8,5 | 7,8 | 7,4 | 1.680 | 1.680 | 1.573 | 1.456 | 1,2 | 1,0 | 2,6 | 0,6 | 1,0 | -5,0 |
| | LOM | 7,4 | 7,2 | 5,7 | 5,7 | 1.653 | 1.651 | 1.252 | 1.268 | 0,8 | 0,9 | -1,0 | 0,7 | 0,8 | 0,3 |
| | TAA | 6,3 | 6,2 | 5,3 | 5,2 | 1.597 | 1.592 | 1.323 | 1.319 | 0,0 | 0,1 | -1,7 | -0,3 | -0,2 | -2,0 |
| | VEN | 6,4 | 6,0 | 5,0 | 4,9 | 1.651 | 1.640 | 1.265 | 1.323 | 1,3 | 1,7 | -4,7 | 0,9 | 1,0 | -0,3 |
| | FVG | 6,8 | 6,5 | 5,7 | 5,3 | 1.670 | 1.611 | 1.396 | 1.310 | 0,6 | 0,9 | -4,1 | -3,0 | -2,6 | -9,9 |
| | LIG | 7,6 | 7,9 | 5,5 | 6,5 | 1.631 | 1.624 | 1.165 | 1.309 | 0,0 | -0,3 | 4,5 | 0,2 | -0,8 | 17,4 |
| | EMR | 6,7 | 7,0 | 5,3 | 5,7 | 1.645 | 1.618 | 1.300 | 1.310 | 2,4 | 2,1 | 7,1 | 0,8 | 0,4 | 7,9 |
| | TOS | 8,7 | 8,1 | 7,1 | 6,7 | 1.640 | 1.608 | 1.323 | 1.316 | 2,1 | 2,7 | -4,7 | 0,3 | 0,7 | -5,1 |
| | MAR | 9,1 | 9,8 | 7,1 | 7,1 | 1.618 | 1.606 | 1.226 | 1.139 | 3,2 | 2,4 | 10,5 | 1,7 | 1,6 | 2,7 |
| | UMS | 8,8 | 8,0 | 7,1 | 6,4 | 1.631 | 1.611 | 1.300 | 1.265 | 0,3 | 1,1 | -9,0 | -0,9 | -0,1 | -11,4 |
| | LAZ | 12,7 | 11,9 | 11,3 | 10,2 | 1.643 | 1.622 | 1.433 | 1.363 | 1,3 | 2,2 | -5,2 | -0,3 | 0,9 | -9,8 |
| | ABR | 11,8 | 11,5 | 10,6 | 10,1 | 1.642 | 1.643 | 1.462 | 1.428 | 2,4 | 2,8 | -0,4 | 2,2 | 2,8 | -2,8 |
| | MOL | 11,5 | 11,9 | 10,5 | 10,3 | 1.718 | 1.669 | 1.561 | 1.428 | 1,8 | 1,4 | 5,3 | -1,8 | -1,5 | -3,6 |
| | CAM | 20,2 | 18,7 | 19,5 | 18,3 | 1.682 | 1.669 | 1.610 | 1.622 | -1,9 | -0,1 | -9,2 | -2,4 | -0,9 | -8,5 |
| | PUG | 14,0 | 14,0 | 12,6 | 12,4 | 1.707 | 1.680 | 1.510 | 1.459 | 0,6 | 0,7 | 0,5 | -1,2 | -0,9 | -2,9 |
| | BAS | 11,7 | 10,6 | 10,4 | 9,5 | 1.706 | 1.624 | 1.503 | 1.436 | 1,0 | 2,3 | -8,3 | -3,7 | -2,7 | -12,4 |
| | CAL | 20,6 | 18,9 | 19,5 | 18,3 | 1.624 | 1.614 | 1.513 | 1.544 | 2,0 | 4,2 | -6,3 | 2,0 | 3,5 | -4,3 |
| | SIC | 15,5 | 14,7 | 14,8 | 14,1 | 1.663 | 1.635 | 1.575 | 1.553 | -0,2 | 0,8 | -5,2 | -1,8 | -0,9 | -6,5 |
| | SAR | 12,9 | 12,9 | 10,5 | 10,5 | 1.622 | 1.621 | 1.286 | 1.285 | 2,7 | 2,6 | 3,0 | 2,6 | 2,6 | 2,8 |

Tavola 39. Segue Tassi di irregolarità, ore lavorate pro capite annuali e variazioni 2010-2011, per modalità di alcune variabili strutturali - Anni 2010 e 2011

| | | Quota Non regolari | | | | Ore lavorate pro capite | | | | Variazioni % 2010-2011 | | | | | |
|---------------|---------------|--------------------|------|--------------|------|-------------------------|-------|--------------|-------|------------------------|------|----------|--------------|------|----------|
| | | Posizioni | | Ore lavorate | | Regolari | | Non regolari | | Posizioni | | | Ore lavorate | | |
| | | 2010 | 2011 | 2010 | 2011 | 2010 | 2011 | 2010 | 2011 | Totali | Reg. | Non reg. | Totali | Reg. | Non reg. |
| TOTALE | | 10,4 | 9,9 | 9,0 | 8,6 | 1.652 | 1.635 | 1.410 | 1.397 | 1,0 | 1,5 | -3,4 | 0,0 | 0,4 | -4,3 |
| Sezione | AGR. | 19,1 | 19,1 | 18,5 | 18,5 | 1.577 | 1.572 | 1.514 | 1.514 | 0,7 | 0,6 | 0,8 | 0,4 | 0,3 | 0,8 |
| | ESTRAZ. | 9,6 | 9,1 | 8,9 | 7,8 | 1.778 | 1.705 | 1.639 | 1.457 | 5,3 | 5,9 | -0,4 | 0,5 | 1,6 | -11,5 |
| | MANIFATT. | 6,6 | 6,6 | 5,1 | 5,0 | 1.764 | 1.774 | 1.352 | 1.327 | 0,0 | 0,0 | 0,5 | 0,4 | 0,5 | -1,4 |
| | ENERGIA | 4,4 | 4,7 | 4,2 | 4,3 | 1.812 | 1.822 | 1.701 | 1.653 | 0,2 | -0,1 | 6,0 | 0,6 | 0,5 | 3,0 |
| | ACQUA | 7,4 | 3,6 | 6,3 | 3,0 | 1.736 | 1.738 | 1.451 | 1.420 | 3,1 | 7,4 | -49,8 | 3,9 | 7,6 | -50,8 |
| | COSTRUZ. | 12,8 | 11,8 | 12,5 | 11,5 | 1.743 | 1.730 | 1.689 | 1.668 | -3,0 | -2,0 | -10,3 | -3,8 | -2,7 | -11,4 |
| | COMM. | 8,6 | 8,1 | 8,0 | 7,6 | 1.813 | 1.773 | 1.681 | 1.662 | -0,4 | 0,1 | -6,1 | -2,5 | -2,1 | -7,1 |
| | TRASP. | 7,3 | 6,6 | 6,8 | 6,0 | 1.817 | 1.798 | 1.687 | 1.625 | 1,1 | 1,8 | -8,1 | -0,1 | 0,7 | -11,4 |
| | HORECA | 16,3 | 15,7 | 14,5 | 13,7 | 1.733 | 1.707 | 1.515 | 1.457 | 4,6 | 5,3 | 0,8 | 2,8 | 3,8 | -3,0 |
| | INFORM. | 7,7 | 8,8 | 6,0 | 7,1 | 1.714 | 1.683 | 1.318 | 1.340 | 4,3 | 3,1 | 19,2 | 2,4 | 1,2 | 21,2 |
| | FINANZA | 6,4 | 5,1 | 5,5 | 4,4 | 1.743 | 1.698 | 1.477 | 1.449 | -2,3 | -1,0 | -21,9 | -4,7 | -3,6 | -23,4 |
| | IMMOB. | 9,7 | 7,9 | 10,3 | 8,7 | 1.217 | 1.206 | 1.307 | 1.340 | 6,2 | 8,3 | -13,0 | 5,5 | 7,3 | -10,8 |
| | PROFESS. | 9,2 | 8,7 | 7,9 | 7,3 | 1.659 | 1.642 | 1.406 | 1.369 | 2,2 | 2,9 | -4,0 | 1,2 | 1,9 | -6,5 |
| | SUPP. IMPR. | 8,4 | 7,3 | 7,4 | 6,3 | 1.486 | 1.488 | 1.290 | 1.281 | 5,4 | 6,7 | -8,6 | 5,6 | 6,8 | -9,2 |
| | PA | 0,1 | 0,1 | 0,1 | 0,0 | 1.552 | 1.551 | 652 | 494 | 0,0 | 0,0 | -8,6 | -0,1 | -0,1 | -30,8 |
| | ISTRUZ. | 11,1 | 10,5 | 8,7 | 9,1 | 1.173 | 1.162 | 890 | 990 | -0,7 | 0,0 | -6,0 | -0,5 | -0,9 | 4,6 |
| | SANITA' | 9,1 | 7,8 | 8,2 | 7,2 | 1.568 | 1.524 | 1.389 | 1.383 | 2,3 | 3,8 | -12,3 | -0,2 | 0,9 | -12,7 |
| | INTRATT. | 30,7 | 28,1 | 25,7 | 22,4 | 1.526 | 1.495 | 1.189 | 1.104 | 3,2 | 7,0 | -5,5 | 0,4 | 4,9 | -12,3 |
| | ALTRI SERVIZI | 22,6 | 19,4 | 19,3 | 17,8 | 1.677 | 1.653 | 1.373 | 1.485 | -2,1 | 2,0 | -15,9 | -1,4 | 0,5 | -9,0 |
| | FAM./CONVIV. | 29,2 | 31,9 | 25,4 | 28,8 | 1.271 | 1.256 | 1.049 | 1.081 | 12,4 | 8,0 | 22,9 | 11,8 | 6,7 | 26,7 |
| | EXTRATERR. | 31,6 | 31,6 | 29,7 | 30,0 | 1.769 | 1.804 | 1.620 | 1.676 | -3,0 | -3,0 | -3,1 | -0,7 | -1,1 | 0,2 |

4.10. Stime dell'input di lavoro grigio

4.10.1. Fonti d'indagine sulle ore lavorate lato impresa e lato famiglia

Nel corso della prima fase di attività del GdL è stato avviato un programma di lavoro sull'analisi del confronto fra le ore lavorate dai dipendenti dell'industria e dei servizi collocati nel campo di osservazione del registro ASIA. Si è in particolare deciso di approfondire le cause della persistente differenza che è possibile riscontrare fra le ore effettivamente lavorate rilevate attraverso le indagini statistiche sulle imprese – sia strutturali che congiunturali - e l'analogo indicatore stimato attraverso LFS.

Le stime dal lato delle famiglie evidenziano per un verso il medesimo profilo stagionale delle stime dal lato delle imprese. Tuttavia, il livello delle stime è sensibilmente diverso: i pro capite stimati dal lato famiglia sono decisamente più elevati. Il GdL ha dunque ritenuto necessario individuare la quota parte di questo divario che dipende dalle distorsioni delle fonti di indagine e di isolare l'eventuale differenziale orario che può essere interpretato come orario effettivo di lavoro ulteriore rispetto a quello retribuito.

Questa attività è proseguita nella seconda parte delle attività del GdL e ha seguito due linee di approfondimento parallele:

- ✓ correzione delle distorsioni dal lato delle famiglie;
- ✓ correzione delle distorsioni dal lato delle imprese;
- ✓ messa a punto di una metodologia di stima del differenziale fra le fonti dovuto alla presenza di ore lavorate in eccesso rispetto a quelle retribuite "in busta".

La correzione delle distorsioni dal lato LFS è stata esposta in questa sezione, e ha portato alla creazione del campione integrato LFS-ADMIN_JOBSh, nel quale sono evidenziabili tutte le posizioni lavorative dipendenti presso imprese presenti in ASIA, con informazioni sulle ore effettivamente lavorate al netto delle correzioni per la sottostima delle ore non lavorate per malattia o causate dalle interviste proxy.

Un'attività aggiuntiva è stata condotta con riferimento al tema della correzione delle distorsioni dal lato impresa e all'aumento del dettaglio delle stime. I primi risultati di questa attività, relativi al 2010, sono stati peraltro presentati nel corso delle "Giornate metodologiche in Istat" e sono stati riportati nel lavoro "Combining business surveys and administrative data to estimate hours actually worked: a small area approach"¹. Nel mese di dicembre una versione bayesiana della stima per piccole aree - "Joint use of survey and administrative sources to estimate the hours actually worked in the business sector"² - è stato presentato al convegno di Barcellona.

L'approccio è stato basato sui seguenti passaggi:

- ✓ integrazione delle imprese rispondenti alle rilevazioni congiunturali GI e VELA e alle rilevazioni annuali SCI e PMI e creazione di un campione unico con informazioni su base annuale;
- ✓ creazione di un database esaustivo delle imprese con dipendenti dell'archivio ASIA con le informazioni annualizzate desumibili dall'archivio EMENS, tra cui una stima delle ore retribuite;
- ✓ sulle imprese agganciate nei due archivi è stata stimata la relazione fra ore effettivamente lavorate e ore retribuite, ed è stata sfruttata l'informazione ausiliaria disponibile sulle ore retribuite sull'universo delle imprese per approfondire il dettaglio delle stime. Il metodo prescelto è basato sulle procedure di stima per piccole aree.

I risultati ottenuti hanno consentito da un lato di verificare la sovrastima delle ore lavorate misurate dalle rilevazioni annuali sulle imprese e causate dal *non response bias* di SCI e PMI; dall'altro

¹ Baldi, De Gregorio, Giordano, Pacini e Sorrentino.

² Baldi, De Gregorio, Giordano, Pacini, Solari e Sorrentino.

hanno reso possibile la creazione di un prototipo di archivio di imprese con dipendenti con una stima individuale delle ore effettivamente lavorate dichiarate dalle imprese.

I risultati di questo archivio sono stati a questo punto confrontati per dominio con i dati di LFS-ADMIN_JOBSh relativi alle posizioni dipendenti presso imprese di ASIA: i primi risultati (Tavola 40 e Figure 16 e 17) evidenziano come il differenziale delle ore lavorate lato famiglia sia pari a circa il 12% nella manifattura e al 9% nel commercio. Si evidenzia inoltre come l'elemento che maggiormente caratterizza le imprese sia costituito dal differenziale fra ore lavorate e retribuite. Sono infatti le ore retribuite a caratterizzare maggiormente le imprese minori e quelle del meridione, dove questo pro capite è decisamente più basso. In particolare nelle microimprese e dalla Campania alla Sicilia le ore retribuite sono inferiori a quelle effettivamente lavorate dichiarate all'indagine.

Questo approccio configura peraltro una serie di ulteriori attività di sviluppo per il necessario consolidamento metodologico di queste stime, in parte già avviate:

- ✓ replica della stima sul 2011;
- ✓ ingegnerizzazione dei processi di creazione delle basi dati e di stima;
- ✓ affinamento metodologico delle procedure di stima;
- ✓ estensione del metodo di stima per piccole aree al lato famiglia, sfruttando appieno le basi informative individuali di fonte EMENS per dare potenza al campione LFS.

4.10.2. Discrepanze fra regime orario contrattuale ed effettivo

Il GdL ha individuato nella discrepanza fra part-time contrattuali ed effettivi una delle possibili modalità di svolgimento di prestazioni lavorative grigie. La base informativa LFS-ADMIN_JOBSh consente in effetti di raffrontare il regime orario dichiarato a LFS con quello contrattuale, per singolo individuo agganciato con posizioni lavorative tracciate in EMENS. Questo lavoro è stato sviluppato nei mesi successivi alla chiusura del GdL ed evidenzia come il 20% circa dei contratti di lavoro part-time corrispondano di fatto a posizioni lavorative effettive full-time. È stato possibile giungere a prime stime dell'entità in termini di ore lavorate dell'input di lavoro grigio derivante da questi rapporti di lavoro e la quota di imponibile EMENS non dichiarata dalle imprese.

Tavola 40. Prime stime delle ore lavorate lato famiglia e lato impresa ottenuto utilizzando LFS-ADMIN_JOBSh e le stime per piccole aree sui rispondenti di VELA, GI, SCI e PMI. Anno 2010

| Regione | Ore lavorate | | | Dipendenti | | Imprese | | Indici (denom.=100) | | |
|--------------------|---------------------|----------------------|-----------------------|------------|----------|---------|----------|---------------------|-----------|-----------|
| | Ore retribuite a | Lato impresa b | Lato famiglia c | Asia | Campione | Asia | Campione | b/a * 100 | c/a * 100 | c/b * 100 |
| | | | | | | | | | | |
| <i>MANIFATTURA</i> | | | | | | | | | | |
| Totale | 1,810 | 1,552 | 1,739 | 3,477,326 | 33,605 | 264,778 | 12,417 | 85.7 | 96.1 | 112.1 |
| NO | 1,846 | 1,563 | 1,749 | 2,442,751 | 23,681 | 151,189 | 8,173 | 84.6 | 94.7 | 112 |
| CE | 1,799 | 1,565 | 1,743 | 558,491 | 5,559 | 52,788 | 2,309 | 87 | 96.9 | 111.3 |
| SU | 1,669 | 1,499 | 1,698 | 476,084 | 4,365 | 60,801 | 1,935 | 89.8 | 101.7 | 113.3 |
| <10 | 1,644 | 1,529 | 1,686 | 573,980 | 5,015 | 193,350 | 2,914 | 93 | 102.6 | 110.3 |
| 10-19 | 1,794 | 1,585 | 1,736 | 513,124 | 4,942 | 40,996 | 2,639 | 88.4 | 96.8 | 109.5 |
| 20-99 | 1,829 | 1,593 | 1,748 | 1,005,013 | 9,675 | 26,095 | 4,409 | 87.1 | 95.6 | 109.7 |
| 100-249 | 1,870 | 1,582 | 1,743 | 460,709 | 4,651 | 3,050 | 1,426 | 84.6 | 93.2 | 110.1 |
| 250 e oltre | 1,871 | 1,487 | 1,763 | 924,501 | 9,322 | 1,287 | 1,029 | 79.4 | 94.2 | 118.6 |
| <i>COMMERCIO</i> | | | | | | | | | | |
| Totale | 1,727 | 1,573 | 1,713 | 2,043,151 | 17,847 | 393,170 | 7,890 | 91.1 | 99.2 | 108.9 |
| NO | 1,763 | 1,576 | 1,701 | 1,199,877 | 10,741 | 178,460 | 4,234 | 89.4 | 96.5 | 108 |
| CE | 1,742 | 1,573 | 1,729 | 367,157 | 2,800 | 79,386 | 1,344 | 90.3 | 99.3 | 109.9 |
| SU | 1,631 | 1,568 | 1,733 | 476,118 | 4,306 | 135,324 | 2,312 | 96.1 | 106.3 | 110.5 |
| <10 | 1,618 | 1,587 | 1,686 | 768,696 | 6,643 | 356,593 | 3,946 | 98.1 | 104.2 | 106.2 |
| 10-19 | 1,818 | 1,629 | 1,776 | 303,059 | 2,813 | 24,986 | 1,566 | 89.6 | 97.7 | 109 |
| 20-99 | 1,838 | 1,613 | 1,796 | 376,107 | 3,380 | 10,324 | 1,623 | 87.7 | 97.7 | 111.4 |
| 100-249 | 1,846 | 1,614 | 1,821 | 125,865 | 1,075 | 813 | 384 | 87.5 | 98.7 | 112.8 |
| 250 e oltre | 1,723 | 1,470 | 1,622 | 469,424 | 3,936 | 454 | 371 | 85.3 | 94.1 | 110.3 |

Figura 16. Ore retribuite ed effettivamente lavorate (dal lato delle imprese e delle famiglie) per regione. Anno 2010

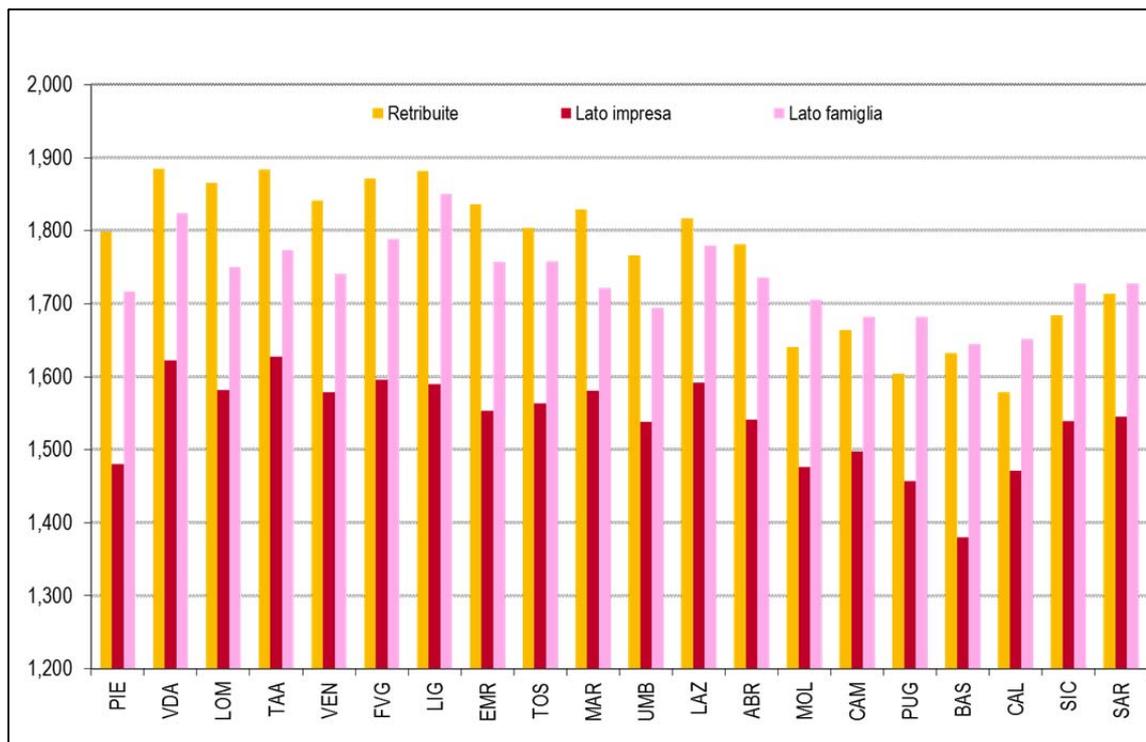
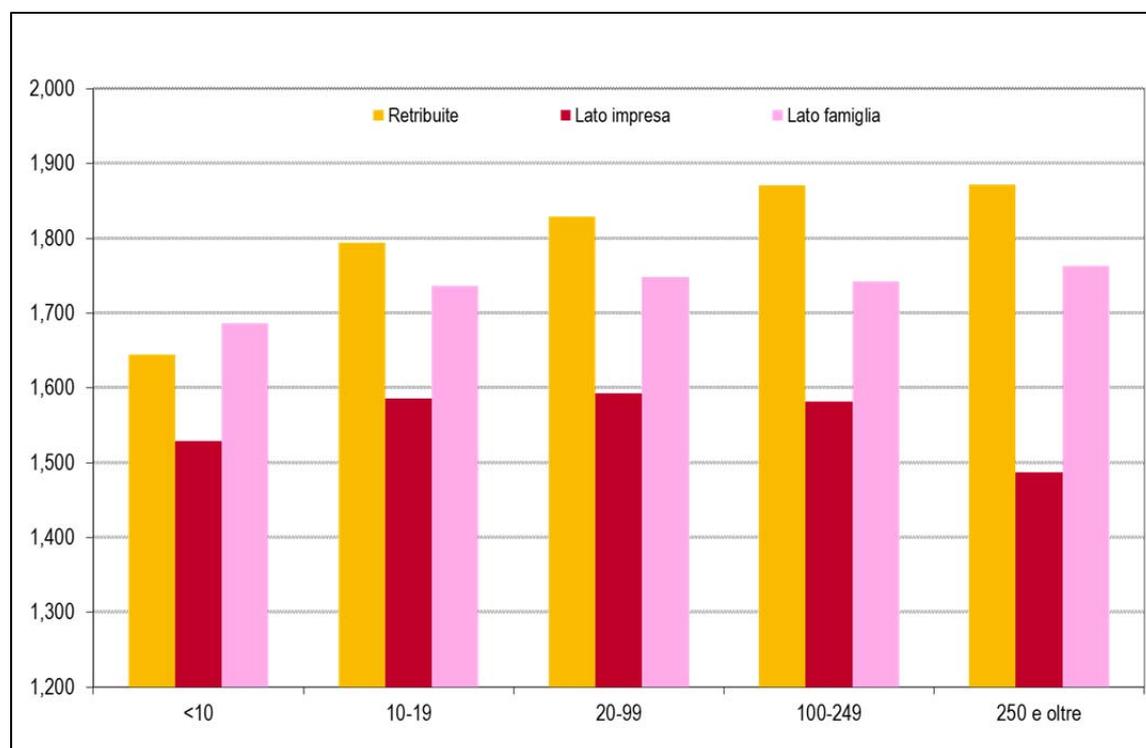


Figura 17. Ore retribuite ed effettivamente lavorate (dal lato delle imprese e delle famiglie) per classe dimensionale delle imprese. Anno 2010



Conclusioni: processo produttivo, rischi, sviluppi

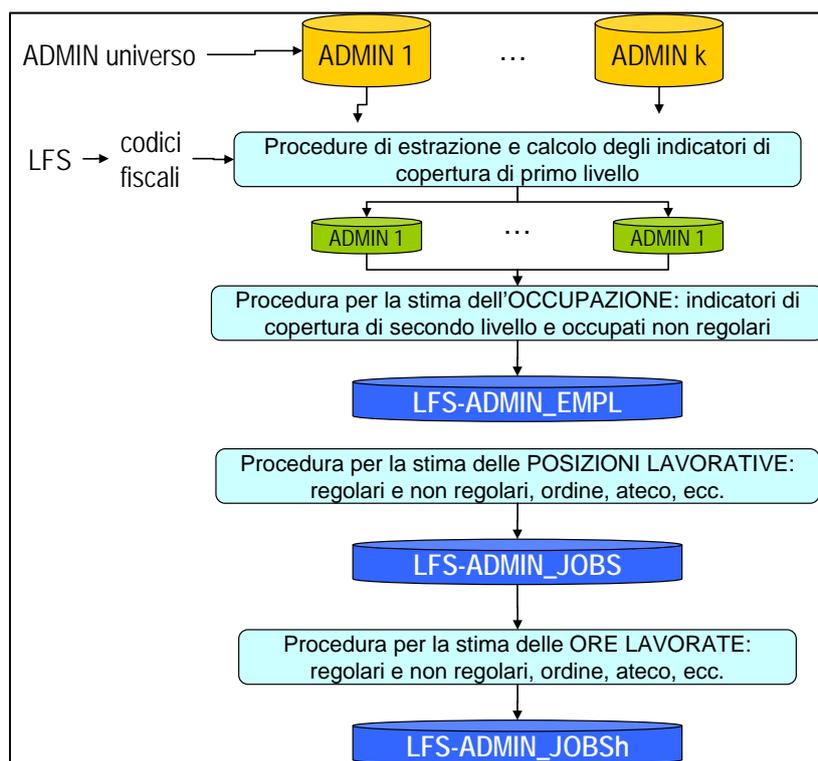
Il processo produttivo delle stime dell'input di lavoro messo a punto dal GdL è stato suddiviso in sequenze ordinate dal punto di vista logico che definiscono un percorso che va dalla acquisizione delle fonti alla produzione di una base informativa costituita dal campione integrato LFS-ADMIN. Si tratta tuttavia di un prodotto il quale, sebbene sia da subito utilizzabile ai fini della produzione corrente, assume le caratteristiche di un prototipo il cui assetto può, e anzi deve, essere sottoposto a ulteriori miglioramenti e ottimizzazioni, sia nella gestione delle basi dati sia nel contenuto e nelle soluzioni adottate nelle procedure per il trattamento dei dati.

La procedura di stima annuale basata sul campione integrato LFS-ADMIN si compone di due macro fasi distinte:

- ✓ La estrazione e prima lavorazione delle basi dati ADMIN;
- ✓ Le fasi di stima dell'input di lavoro.

Si tratta di due macro fasi collocate in un ordine sequenziale e la cui onerosità dipende essenzialmente dalla stabilità delle forniture ADMIN. Attualmente il processo produttivo è stato pensato e organizzato come una sequenza di procedure SAS. Nella prima macro fase vengono eseguite un blocco di procedure dedicate per ciascuna fonte ADMIN, che estraggono e trattano i record degli individui presenti nel campione LFS, portando alla creazione di una base dati unica. A partire da questa base dati vengono lanciate tre procedure disposte in sequenza che producono le stime su occupazione, posizioni lavorative e ore lavorate relative agli individui nel campione.

Figura 18



Dall'anno di riferimento 2013 la ricostruzione dei codici fiscali dovrebbe essere associata direttamente ai microdati del campione LFS attraverso le Liste anagrafiche comunali (LAC). Per l'anno di riferimento 2012 dovranno invece essere replicate:

- ✓ le procedure di aggancio con il Registro delle persone fisiche e con le LAC,
- ✓ l'imputazione del codice fiscale tramite donatore.

Non si tratta di procedure onerose, ma richiedono un coordinamento fra diverse strutture. Ci sono inoltre ampi margini di miglioramento della qualità degli agganci e del record linkage. Questo

tema è stato infatti trascurato poiché è stata data priorità alla messa a punto dell'intera procedura di stima, ma è evidente che la misurazione degli aspetti legati alla qualità delle stime è di estrema importanza per la loro validazione. Al netto di questa attività supplementare, nel complesso la ricostruzione dei codici fiscali dovrebbe ragionevolmente richiedere pochi giorni di lavoro.

I tempi di acquisizione delle forniture sono relativamente certi dal momento che corrispondono a un calendario prefissato concordato con i fornitori istituzionali. I programmi per l'estrazione delle informazioni relative agli individui nel campione sono già disponibili e, se non intervengono variazioni significative, le procedure non hanno tempi di esecuzione significativi.

Si tratta di procedure separate per ciascuna fonte, dal momento che i formati di acquisizione sono diversi: eseguono semplicemente la lettura, l'estrazione dei codici fiscali LFS e l'archiviazione in un data set SAS.

Cambiamenti nelle variabili e soprattutto l'utilizzo di nuove fonti possono invece rendere più laboriosa la procedura di estrazione.

Qualche incognita la riserva in particolare il trattamento delle informazioni contenute nei nuovi registri degli occupati curati dall'ISTAT, la cui struttura si stabilizzerà prevedibilmente nei prossimi anni, soprattutto con riferimento al segmento degli indipendenti.

Per il trattamento di ADMIN e calcolo dei flag di copertura di primo livello viene usato un blocco di procedure, una per ciascuna fonte. Ciascuna procedura provvede a elencare per ciascun individuo le posizioni lavorative assunte nell'anno di riferimento dell'intervista LFS e provvede a estrarre le informazioni necessarie, individuando se la settimana dell'intervista ricade all'interno delle date di validità di ciascuna posizione lavorativa.

Questo blocco di procedure inoltre calcola altri indicatori riepilogativi per ciascuna posizione, riferiti alla settimana dell'intervista, ai quattro trimestri e all'intero anno: giornate di copertura contrattuale, imponibili, giornate retribuite, eventi (malattia, cassa integrazione, ecc.). A ogni posizione è inoltre associato il codice fiscale del datore di lavoro.

Si crea così, per ogni fonte, un archivio delle posizioni in cui ciascun intervistato è presente nella fonte nell'anno dell'intervista: si tratta di un archivio delle posizioni nella fonte. A partire da questo archivio viene generato un archivio riepilogativo per ciascun intervista LFS che ordina le posizioni lavorative assunte da ciascun individuo nella fonte in base alla loro rilevanza nella settimana dell'intervista: un record per ogni intervistato dunque, con informazioni dettagliate sulla posizione principale nella fonte e notizie di riepilogo sulle altre posizioni lavorative nella fonte. Riepilogando, per ogni fonte si generano due archivi, quello delle posizioni e quello dei record: nel complesso dunque, date le k fonti, $2k$ archivi.

Questo blocco di procedure dipende dalle variabili presenti nelle fonti e dall'organizzazione dei record: è dunque sensibile a cambiamenti nelle fonti, sia nel nome delle variabili sia nel loro numero. Se ciò avviene occorre dunque intervenire manualmente sul codice. Vi sono peraltro ampi spazi di miglioramento per generalizzare le procedure e rendere minimo l'impatto di questi cambiamenti. Nel complesso, se non vi sono cambiamenti, le procedure richiedono pochissimo tempo.

Dai k archivi delle posizioni si estraggono i codici fiscali dei datori di lavoro per agganciarne le caratteristiche nei registri statistici esistenti: codice di attività economica, numero di addetti dipendenti e indipendenti, regione, forma giuridica. Queste informazioni vengono agganciate ai $2k$ archivi delle posizioni e dei record.

Il trattamento e l'estrazione delle informazioni contenute nei registri statistici non presenta difficoltà di rilievo, almeno fin tanto che ci si limita alla considerazione di ASIA. Più problematico appare invece il ricorso a prodotti non standardizzati (come il cosiddetto "ASIA-12milioni") o di esecuzione problematica, come le S13 e le loro unità locali, con particolare riferimento agli istituti scolastici, o come le istituzioni sociali private e non profit. Le informazioni nei registri inoltre non sempre sono disponibili nell'anno di riferimento coerente, almeno nelle fasi iniziali di stima. Ciò impone dunque l'utilizzo di archivi riferiti a $t-1$.

Per le procedure di stima vale in buona parte quanto detto per le procedure di acquisizione dei dati. I tre programmi che producono in sequenza le stime sono già generalizzati in modo da essere indipendenti dall'anno di riferimento. La procedura delle stime dell'occupazione esegue dapprima, una fonte per volta, delle regressioni logistiche attraverso le quali sottopone a verifica i flag di co-

pertura ai fini della loro validazione: è la parte di procedura che stima la sovracopertura delle fonti ADMIN. La seconda parte della procedura provvede a stimare per dominio la sottocopertura dell'occupazione da parte di LFS. L'output di questa procedura è un dataset (LFS-ADMIN_EMPL) che contiene un flag relativo alla validazione della copertura di primo livello (copertura di secondo livello) e un flag che identifica i non occupati LFS trasformati in occupati non regolari.

In assenza di variazioni nelle basi dati le procedure di stima richiedono un tempo assai limitato. Più tempo in realtà dovrebbe tuttavia essere dedicato all'analisi delle stime e dei modelli, aumentando il valore aggiunto metodologico della procedura.

La procedura che stima le posizioni lavorative utilizza LFS-ADMIN_EMPL insieme a tutti i k archivi delle posizioni lavorative per fonte, calcola le posizioni lavorative individuali e ne stima i dati eventualmente mancanti. Viene così creato l'archivio LFS-ADMIN_JOBS.

Su questo archivio opera l'ultima procedura, quella per la stima delle ore lavorate, che vi aggiunge le variabili relative alle ore effettive. Il dataset finale si chiama LFS-ADMIN_JOBSh.

Se vi sono cambiamenti in variabili rilevanti delle fonti ADMIN occorrerà intervenire manualmente anche sui programmi di stima e gestirne l'impatto sulle stime. Questa attività può riguardare in misura particolare la gestione dei modelli logistici nella stima dell'occupazione e dei modelli di imputazione per la stima delle ore lavorate.

Più complessa si presenterà invece la gestione di nuove fonti, le quali implicano l'aggiunta di procedure ex novo, lo studio delle nuove basi dati, delle nuove procedure di estrazione, e l'aggiunta di processi di stima interamente nuovi anche se chiaramente mutuati da quelli adottati per le altre fonti. In questo caso le modifiche possono essere laboriose, ma comunque gestibili con un numero limitato di giornate-uomo.

I rischi principali, in parte già richiamati, sono dovuti a:

- ✓ Modifiche nei tracciati record ADMIN. Può in sostanza accadere che appaiano nuove variabili nelle forniture ADMIN e/o che venga meno la disponibilità di alcune variabili. In ambedue i casi occorre gestire le fasi di acquisizione, assicurare la continuità dei processi di validazione delle coperture; modificare le procedure di stima e/o stratificazione nelle varie fasi della stima. Si tratta di rischi che appaiono gestibili anche con la dotazione standard;
- ✓ Mancata disponibilità di forniture in uso. Si tratta di eventualità poco probabili e per certi aspetti poco verosimili. In questo caso occorrerà capire la reale gravità dell'evento (per esempio più fonti che collassano in una unica fornitura): il venir meno di una fonte può causare un danno grave che può essere risolto solamente con procedure di stima che facciano affidamento sulle edizioni precedenti;
- ✓ Nuove forniture. È già un evento pressoché certo, dal momento che l'istituto ha proseguito nella strategia di acquisizione delle fonti. Le nuove forniture da un lato migliorano la qualità del prodotto finale, ma dall'altro possono comportare modifiche anche molto rilevanti nella strategia complessiva della stima e nelle procedure. Questo tipo di novità va gestito con grande cura;
- ✓ Modifiche nelle variabili della fonte LFS. È un evento poco probabile ma che può comportare effetti significativi se ad essere coinvolte sono le variabili maggiormente utilizzate nel nuovo processo di stima;
- ✓ Modifiche nelle tecniche di rilevazione di LFS. Sono eventi più verosimili che impattano sul processo di stima qualora abbiano una influenza apprezzabile sulle variabili più importanti del nuovo processo di stima;
- ✓ Rischi informatici. Se da un lato tali rischi possono essere ridotti attraverso una opportuna razionalizzazione delle basi dati e della loro collocazione, dall'altro andrebbero gestiti e controllati i rischi legati alla disponibilità del software usato per le stime (il SAS);
- ✓ Rischi di coordinamento. Sono i rischi di coerenza sequenziali con le attività svolte in altre strutture dell'istituto, come per esempio quelle legate alla gestione di LFS o, più verosimilmente, quelle legate alla produzione dei registri statistici sulle unità economiche, e in particolar modo quelle extra-ASIA.

Sono molte le aree di contaminazione dell'innovazione introdotta per la stima dell'input di lavoro

ro e basata sull'utilizzo integrato di LFS e ADMIN. Ne elenchiamo qui alcune di quelle caratterizzate da orizzonti di breve-medio periodo:

- ✓ Ore lavorate grigie. Sulla stima del lavoro grigio il GdL ha già sviluppato un prototipo per la stima di tre componenti delle ore lavorate dai dipendenti dell'industria e dei servizi: le ore retribuite, quelle effettivamente lavorate lato impresa e le ore effettivamente lavorate lato famiglia. Appare pertanto strategica la prosecuzione degli esperimenti di integrazione fra fonti ADMIN esaustive (nella fattispecie EMENS), fonti di indagine lato impresa di natura mensile o trimestrale (VELA e GI) e annuale (SCI e PMI), e campione integrati LFS-ADMIN. Il consolidamento di queste stime può portare sia alla quantificazione delle ore lavorate grigie sia alla quantificazione dei flussi monetari associati a questo input di lavoro;
- ✓ Sempre sul tema delle stime del lavoro grigio, è possibile attraverso LFS-ADMIN studiare la relazione fra regimi orari contrattuali ed effettivi, con particolare riferimento all'utilizzo dei contratti part-time. Un lavoro in tal senso è stato già avviato;
- ✓ Analisi specifiche potranno riguardare alcuni segmenti del mercato del lavoro e in particolare la componente non regolare. Il riferimento va in particolare alla quantificazione delle retribuzioni e più in generale dei flussi monetari associati ai non regolari;
- ✓ L'aggancio LFS-ADMIN consente l'uso integrato di dati lato impresa e lato famiglia per l'analisi degli *skill* e più in generale delle caratteristiche dell'occupazione associate a determinati segmenti di imprese;
- ✓ È inoltre possibile valutare il *substitution bias* di LFS, agganciando gli individui del campione teorico e studiandone le caratteristiche. Questa attività è stata predisposta e richiede solo una pianificazione delle attività e l'estrazione dei record dalle basi dati già disponibili. Si può inoltre valutare la possibilità di sfruttare questi risultati ai fini della correzione delle mancate risposte totali nella LFS;
- ✓ Analisi longitudinali per anticipare alcune stime basate solo su LFS. Dal momento che la stima della componente non regolare è effettuata su base annuale ma a t-2, si tratta di mettere a punto modelli di stima per predire al meglio le singole componenti (fra cui quelle regolari e non regolari) dell'input di lavoro potendo contare sui dati correnti - mensili o trimestrali - di LFS e di alcune fonti ADMIN;
- ✓ Attraverso l'utilizzo integrato di ASIA Unità Locali si potrà valutare ancora meglio la validità del quesito sull'attività economica posto con LFS. Anche per questa attività le informazioni di base sono già disponibili;
- ✓ Estensioni dell'esperimento verso altri settori dell'istituto o indagini per scopi di validazione e innovazione. Di particolare importanza appare la replica dell'esperimento sulla rilevazione sull'uso del tempo (TUS 2013), investimento cruciale per l'ulteriore sviluppo delle stime delle ore lavorate;
- ✓ Estensione del ventaglio delle fonti amministrative utilizzate negli agganci ad ambiti non direttamente collegati all'attività lavorativa, come ad esempio le pensioni, i redditi, i trattamenti non pensionistici.

Riferimenti bibliografici

Istituzioni internazionali

- Consiglio UE. 2003. Council resolution on transforming undeclared work into regular Employment(2003/C 260/01)
- EC. 1996. *Council regulation (EC) No 2223/96 of 25 June 1996 on the European system of national and regional accounts in the Community*. Bruxelles: EC
- EC. 2009. Communication from the Commission to the European parliament and the Council on the production method of EU statistics: a vision for the next decade. COM(2009) 404 final. Bruxelles: EC
- EC. 2010. *ESA 2010 Manual*. Draft 20 December 2010. Bruxelles: EC.
- European Commission. 2009. *Communication from the Commission to the European Parliament and the Council on the production method of EU statistics: a vision for the next decade*. Bruxelles, COM(2009) 404 final
- ILO. 1962. *Resolution concerning statistics of hours of work, adopted by the Tenth International Conference of Labour Statisticians*. October. Geneva: ILO
- ILO. 2008. *Measurement of working time*. Report II. 18th International Conference of Labour Statisticians. Geneva, 24 November–5 December
- ILO. 2013. Labour Inspection and Undeclared Work in the EU. *Working Document n. 29*. Geneva.
- OECD. 2002. *Measuring the non-Observed Economy. A Handbook*. Paris: OECD
- OECD. 2004. *OECD measures of total hours worked*. The OECD Productivity Database. Paris. OECD
- OECD. 2007. Factors explaining differences in hours worked across OECD countries. OECD (ECO/CPE/WP1(2007)11)
- OECD. 2009. Comparability of labour input measures for productivity analysis. OECD (STD/CSTAT/WPNA(2009)11)
- UN. 2009. *System of National Accounts 2008*. New York: 2008

Aspetti metodologici

- Agresti A. 2007. *An Introduction to Categorical Data Analysis*. Second Edition. JohnWiley & Sons, Inc., Hoboken, New Jersey
- Andridge R.R., Little R.J.A. 2010. A Review of Hot Deck Imputation for Survey Non-response. *International Statistical Review*, 78, 1, 40–64
- Boehm L.E. (1989). Reliability of proxy response in the current population survey. In: *Proceedings of the American statistical association*, Survey research methods section, 486-489.
- Bohem L.R. 1989. Reliability of Proxy Response in the Current Population Survey. U.S. Bureau of Labor Statistics
- Cascioli R. 2006. Integrazione dei dati micro dalla Rilevazione delle Forze di Lavoro e dagli archivi amministrativi INPS: risultati di una sperimentazione sui dati campione di 4 province. *Contributi Istat*, no. 6.
- Dawe F., Knight I. (1997). A study of proxy response in the Labour Force Survey. In: ONS (2003). *Report on proxy response study based on LFS questions*. In: Labour Force Survey User Guide. Volume 1: Background & Methodology. Capitolo 11, 50-59. Lavoro molto importante e molto citato.
- Falorsi P.D., S. Falorsi, A. Russo. 1994. Empirical Comparison of Small Area Estimation Methods for the Italian Labour Force Survey, *Survey Methodology*, 20, 171-176
- Fosen J. e Li-Chun Zhang. 2011. *Quality evaluation of employment status in register-based census*. Proc. 58th World Statistical Congress, Dublin (Session STS024)

- Grande E., Luzi O. 2003. Metodologie per l'imputazione delle mancate risposte parziali: analisi critica e soluzioni disponibili in Istat. *Contributi Istat*, no. 6.
- Kuijvenhoven L. e S. Scholtus. 2010. Estimating accuracy for statistics based on register and survey data. *Discussion paper* (10007). Statistics Netherlands, The Hague/Heerlen
- Kuijvenhoven L. e S. Scholtus. 2011. Bootstrapping Combined Estimators based on Register and Sample Survey Data. *Discussion paper* (201123). Statistics Netherlands, The Hague/Heerlen
- Istat. 2015. Ricostruzione statistica delle serie regionali di popolazione del periodo 1/1/2002-1/1/2014. *Nota informativa*, 14 gennaio. Roma: ISTAT
- ISTAT, CBS, GUS, INE, SSB, SFSO, EUROSTAT. 2012. *ESSnet on Data Integration. Report on WP2. Methodological developments*. EUROSTAT.
- ISTAT, CBS, GUS, INE, SSB, SFSO, EUROSTAT. 2012. *ESSnet on Data Integration. Report on WP4. Case studies*. EUROSTAT.
- Kleven Ø. Lagerstroem B.O., Thomsen I. (2008). A simple model for studying the effects of proxy interviewing: the norwegian labour force survey as a case study. *Statistics Norway Documents*, n.2
- Krieg S., V. Blaess, M. Smeets. 2012. Small area estimation of turnover of the Structural Business Survey. *CBS Discussion paper*. Den Haag: CBS
- Max A.H.Y., Roberts E., Elliott R.F., Bell D., Scott A., (2006). Comparing the New Earnings Survey (NES) and the Labour Force Survey (LFS): An Analysis of the Differences between the Data Sets and their Implications for the Pattern of Geographical Pay in the UK. *Regional Studies*.
- Mood C. 2010. Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review* vol. 26 no.1, 67–82
- Mukhopadhyay P.K., McDowell A. 2011. Small Area Estimation for Survey Data Analysis Using SAS Software. *SAS Global Forum 2011*
- Rao J.N.K. 2003. *Small Area Estimation*. Wiley series in survey methodology. 327
- The EURAREA Consortium. 2004. *Enhancing Small Area Estimation Techniques to meet European Needs*. Project reference volumes 1-3.
- Thomsen I., Villund O. (2011). Using Register Data to Evaluate the Effects of Proxy Interviews in the Norwegian Labour Force Survey. *Journal of Official Statistics* 27: 87-98.
- Törmälehto V.-M. 2008. *Social statistics - integrated use of survey and administrative data at Statistics Finland*. International Association for Official Statistics Conference on Reshaping Official Statistics, Shanghai, 14-16 October
- Yung W., S. Rubin-Bleuer, S. Landry. 2010. Small Area Estimation for Business Surveys. *Survey Research Methods – JSM 2010*

Ore lavorate

- Ahmad N., F. Lequiller, P. Marianna, D. Pilat, P. Schreyer, A. Wöfl. 2003. *Comparing labour productivity growth in the OECD area - the role of measurement*. Paper. OECD National Accounts Experts Meeting. Paris: OECD
- Alesina A.F., E.L. Glaeser e B. Sacerdote. 2006. Work and Leisure in the U.S. and Europe: Why So Different? *NBER Macroeconomics Annual* 2005, Volume 20
- Baker J., N. von Sanden. 2006. *Estimating Average Annual Hours Worked*. Research Paper. Canberra: ABS
- Barham C. e J. Leonard. 2002. Trends and sources of data on sickness absence. *National Statistics feature*, ONS
- Barham C. e N. Begum. 2005. Sickness absence from work in the UK. *National Statistics feature*, April, ONS

- Barmby T.A., Ercolani M.G. e J.G. Treble. 2002. Sickness absence: an international comparison. *The Economic Journal*, 112 (June), F315–F331
- Barmby T.A., M.G Ercolani. e J.G. Treble. 2004. Sickness absence in the UK 1984-2002. *Swedish Economic policy Review*, 11, 65-88
- Bell D., P. Elias. 2003. *The definition, classification and measurement of working time arrangements: A survey of issues with examples from the practices in four countries*. Conditions of Work and Employment Programme. Geneva: ILO
- Bell L.A. e R.B. Freeman. 2001. The incentive for working hard: explaining hours worked differences in the US and Germany. *Labour Economics* n.8, 181–202
- Bird D., O.Black e P. Hopwood. 2004. Estimating hours worked in the context of the Labour Costs Index. Paris Group, 29th September-1st October, Session 2 ‘Measuring Hours Actually Worked’
- BLS. 2010. Illness-related work absences during flu season. *Issues in Labor Statistics*, July
- Bolling S. 2004. Hours of Absence, Overtime and Hours Actually Worked. Statistics Sweden
- Bruyere M., O. Chagny. 2002. *The fragility of international comparisons of employment and hours worked an attempt to reduce data heterogeneity*. July: 5
- Causa O. 2008. Explaining Differences in Hours Worked among OECD Countries. *Economics department working papers*: 596. Paris: OECD
- De Paola M., V. Pupo e V. Scoppa. 2009. Absenteeism in the Italian Public Sector: The Effects of Changes in Sick Leave Compensation. *Working Paper* n. 16, Dipartimento di Economia e Statistica, Università della Calabria
- Evans J.M., Lippoldt D.C., P. Marianna. 2001. *Trends in Working Hours in OECD Countries*. OECD Labour Market and Social Policy Occasional Papers: 45. Paris: OECD
- Fleck S. E. 2009. International comparisons of hours worked: an assessment of the statistics. *Monthly Labor Review*: May: 3-31
- Frazis H., J. Stewart. 2010. *Why Do BLS Hours Series Tell Different Stories About Trends in Hours Worked?* In: Abraham K.G., J.R. Spletzer, M. Harper (eds.). *Labor in the New Economy*. Chicago: University of Chicago Press
- Fursman L. 2006. Working Long Hours in New Zealand: A profile of long hours workers using data from the 2006 Census. Department of Labour, New Zealand
- Galarneau D., Maynard J.P. e J. Lee. 2005. Whither the workweek? *Perspectives* (June), Statistics Canada
- Greenwood A.M. 2001. *The hours that we work: the data we need, the data we get*. ILO Bulletin of labour statistics: 1. Geneva: ILO
- Hansson J., C.-H. Gustafson. 2006. Working time – a complicated variable in measuring labour productivity. *Yearbook on Productivity*. Statistics Sweden
- Harper P. 2006. *Implementing New Estimates of Hours Worked into the Australian National Accounts*. Information Paper. Canberra: ABS
- Heisz A., S. LaRochelle-Côté. 2003. *Working hours in Canada and the United States*. Analytical Studies Branch research paper series. Statistics Canada
- Hoffman E., Mata A. 1999. *Measuring working time: an alternative approach to classifying time use*. Geneva: ILO
- Hogarth T., W.W. Daniel, A.P. Dickerson, D. Campbell, M. Wintherbotham e D. Vivian. 2003. The business context to long hours working. Employment relations research series n.23. University of Warwick, Institute for employment research
- Isgut A., L. Bialas, J. Milway. 2006. *Explaining Canada-US Differences in Hours Worked: Where does the time go?* Canadian Economics Association 40th Annual Meeting. Concordia University, Montréal, Québec

- Istat. 2004. *Background paper on the estimation of hours actually worked*. Istat working group on hours worked. Intermediate Report. Roma: Istat
- Istat. 2006. *Tempi di lavoro e valorizzazione delle competenze. Rapporto annuale sulla situazione economica del paese*. Ch. 4. Roma.
- Istat. 2011. *I tempi del lavoro*. Argomenti n. 40. Roma: Istat
- Keil U.-K. 2004. Working Hours and Total Hours Worked in Germany - Measurement by the FSO. Paris Group Meeting, 29 September – 1 October
- Kirkland K. 2000. On the decline in average weekly hours worked. *Monthly Labor Review*, July
- Köerner T. 2012. *Quality issues regarding the number of hours actually worked in the German LFS*. 7th Workshop on labour force survey methodology - Data processing and data quality. Madrid, Spain, 10 – 11 May
- Langenberg H. 2004. Reflections on hours of work. Paper for the Paris Group Meeting, Lisbon, 29 September - 1 October
- Livanos I. e A. Zangelidis. 2010. *Sickness Absence: a Pan-European Study*. MPRA Paper No. 22627,
- Loriga S. e A. Spizzichino. 2013. Working hours: analysis of Italian LFS results versus administrative data and business survey.
- Lusiniyan L. e L. BONATO. 2007. Work Absence in Europe. *IMF Staff Papers* Vol. 54, No. 3
- Ma A.H.Y., E.Roberts, R.F. Elliott, D. Bell, A. Scott. 2007. Comparing the New Earnings Survey (NES) and the Labour Force Survey (LFS): An Analysis of the differences between the data sets and their implications for the pattern of geographical pay in the UK. *Regional Studies*, 40:6, 645-665
- Maynard J.-P, L. Chung, D. Sunter. 2004. *Measuring Hours Actually Worked*. Session 2, Paris Group, September 29-October 1
- Maynard J.-P. 2005. *Annual Measure of the Volume of Work Consistent with the SNA: the Canadian Experience*. Economic analysis methodology paper series. Statistics Canada
- Mizunoya T. 2001. *An International Comparison of Unpaid Overtime Work Among Industrialized Countries*. Mimeo
- Naur M., Solbjergghøj L. 2004. *User demands and their consequences for the measurement of working time*. Statistics Denmark
- Rodgers W.L., C. Brown, G.J. Duncan. 1993. Errors in Survey Reports of Earnings, Hours Worked, and Hourly Wages. *Journal of the American Statistical Association*, Vol. 88, No. 424 (Dec., 1993), pp. 1208-1218
- Skoglund T. 2001. *Employment in the Norwegian National Accounts*. Statistics Norway Department of Economic Statistics
- Villund O. (2009). *Measuring working hours in the Norwegian Labour Force Survey A pilot study of data quality using administrative registers*. Statistics Norway Reports n.3.
- Walling A. (2007). *Understanding statistics on full-time/part-time employment*. Economic & Labour market review, vol.1 n. 2 (February)
- Williams R.D. (2002). *Hours worked: a comparison of estimates from the Labour Force and New Earnings Survey*. Labour Market Division, Office for National Statistics.
- Williams R.D. (2004). *Investigating hours worked measurements*. Labour Market Division, Office for National Statistic.
- Williams R.D. 2002. Hours worked: a comparison of estimates from the Labour Force and New Earnings Survey. *Labour Market trends* (August): 429-442
- Williams R.D. 2004. Investigating hours worked measurements. *Labour Market trends* (February): 71-79
- Ypma G., B. van Ark. 2006. *Employment and Hours Worked in National Accounts: A Producer's View on Methods and a User's View on Applicability*. EU KLEMS working paper series: 10

Lavoro irregolare

- Baccini A., L. Castellucci e M. Vasta. 2003. Lavoro sommerso: il caso della Toscana. *Studi e note di economia* n. 1.
- Baldassarini A. 2001. Non regular foreign input of labour in the new national accounts estimates. OECD Meeting of national accounts experts, Paris, 9-12 October
- Barthelemy P., F. Miguelez, E. Mingone, A. Pahl e E.A. Wenig. 1990. Underground Economy And Irregular Forms Of Employment (Travail Au Noir), Commissione Dal Dg V
- Bellardi L.. 1989. Intervento dello Stato e Relazioni Industriali: Un'iniziativa In Materia di Lavoro Sommerso, In Bellardi L. - Brunetta R. (A Cura di), *Innovazione e Lavoro Nel Mezzogiorno : Soggetti e Meccanismi di Regolazione Dello Sviluppo*, Bari : Edizioni Dedalo
- Bianco G. 2004. Politiche economiche per l'emersione del lavoro
- Bignami R., G. Casale e M. Fasani. 2013. *Labour inspection and employment relationship*. ILO Working Document n. 28, Geneva.
- Biscourp P. 2004. Les rythmes de travail entre 1995 et 2001: faible progression de l'irrégularité. *Insée première* n. 994 (Dicembre)
- Brandolini A. e E. Viviano. 2013. Behind and beyond the (headcount) employment rate. *Statistical working paper*, Eurostat
- Camarero L. e J. Oliva. 2004. Las trabajadoras invisibles de las áreas rurales: un ejercicio estadístico de estimación. Departamento de Sociología (Universidad Pública de Navarra)
- Cappariello R. and R. Zizza. 2009. Dropping the books and working off the books. *Temì di discussione*. Banca d'Italia,
- Cappariello R. e R. Zizza. 2009. Istruzione ed economia sommersa. In: Banca d'Italia, *Mezzogiorno e politiche regionali*. Seminari e convegni n. 2, novembre, p. 191-214.
- Cnel. 2008. Gli immigrati nel mercato del lavoro italiano. Rapporto realizzato dal CRELI dell'Università Cattolica di Milano
- Commissione Europea. 1998. *Paper On Unclared Works, Dg V*, Bruxelles
- Coniglio N. 2009. Diseconomie ambientali. Discussione. In: Banca d'Italia, *Mezzogiorno e politiche regionali*. Seminari e convegni n. 2, p. 215-222.
- Damiani M.R.. 1998. *Un'indagine Sul Lavoro Non Dichiarato Attraverso Gli Uffici Periferici Del Ministero Del Lavoro*, In *Critica Sociologica*, 1998, Pp. 58
- De Gregorio C. e A. Giordano. 2014. "Nero a metà": contratti part-time e posizioni full-time fra i dipendenti delle imprese italiane. Istat working paper n.3.
- De Gregorio C. e A. Giordano. 2015. *The heterogeneity of irregular employment in Italy: some evidence from the Labour force survey integrated with administrative data*. Istat working paper n.1.
- EMN Italia. 2010. Mercato occupazionale e immigrazione. Terzo rapporto
- Eurobarometer. 2014. *Undeclared work in the European Union. Report*. Special Eurobarometer 402. Survey requested by the European Commission, Directorate-General for Employment, Social Affairs and Inclusion and coordinated by Directorate-General for Communication
- Eurofound. 2010. *Fifth European Working Conditions Survey*, Publications Office of the European Union, Luxembourg.
- European Union Agency for Fundamental Rights. 2011. *Migrants in an irregular situation employed in domestic work: Fundamental rights challenges for the European Union and its Member States*. Luxembourg: Publications Office of the European Union
- Fondazione Censis. 2003. *Promuovere regolarità e trasparenze nel mercato del lavoro*. Ricerca sul sommerso e manuale a supporto dei Servizi per l'impiego.
- Fondazione Censis. 2006. Affidamento di attività' di ricerca e analisi sul fenomeno del lavoro irregolare e sommerso nell'ambito territoriale della regione marche. Report finale

- Frey L.. 1990. Il Mercato Del Lavoro Sommerso : Aspetti di Continuità e Nuove Tendenze. In: Frey L. e M. Capparucci (a cura di), *Progresso tecnico, flessibilità, occupazione, con Particolare riguardo all'Italia*, Angeli
- Frey L.. 1991. Il Mercato Del Lavoro Sommerso : Aspetti di Continuità e Nuove Tendenze, In Maruani M. - Reynard E. (A Cura Di), *La Flessibilità Del Lavoro In Italia*, Angeli
- Gagliardi F. 2006. Il lavoro sommerso in Provincia di Arezzo: stima delle ULA mediante un metodo misto. *Working Paper* n. 61, June
- Galantino L.. 1998. *Una Componente Strutturale Del Mercato Del Lavoro Italiano: L'economia Sommersa*, In *Dir. Lav.*, 1998, N. 1-2, Pp. 67 Ss.
- Garibaldi P. 2006. Labour force participation, unemployment and shadow activity. Università di Torino e Collegio Carlo Alberto, Settembre
- Giannelli G.C. e L. Mangiavacchi, Piccoli L. 2009. Size and Value of Unpaid Family Work in Europe. *Working paper*
- ILO. 2013. *Labour inspection and undeclared work in the EU*. Working document, No. 29. Geneva: ILO
- IRES. 2009. *Emersione e legalità per un lavoro sicuro. Quali fattori di sviluppo per l'impresa*. Report di ricerca nazionale
- Isfol. 2005. Emersione del lavoro irregolare. Il punto su...
- Isfol. 2007. Dimensione di genere e lavoro sommerso. Indagine sulla partecipazione femminile al lavoro nero e irregolare. *Monografico* n. 13 dell' Osservatorio Istituzionale e Normativo
- Isfol. 2007. *La terziarizzazione del sommerso dimensioni e caratteristiche del lavoro nero e irregolare nel settore dei servizi*. I libri del Fondo sociale europeo, n.154
- Isfol. 2007. Le politiche per l'emersione del lavoro nero e irregolare. Quadro istituzionale e normativo. *Monografico* n. 11 dell' Osservatorio Istituzionale e Normativo
- Isfol. 2011. *Dimensioni e caratteristiche del lavoro sommerso/irregolare in agricoltura*. Area "Sistemi Locali e Integrazione delle Politiche".
- Jandl M., C. Hollomey e A. Stepien. 2009. Migration and Irregular Work in Austria: Results of a Delphi-Study. *International migration papers* n.90
- Kedir A.M., M.D. Fethi e C.C. Williams. 2011. *Evaluating tax evasion in the European Union: a case study of the prevalence and character of 'envelope wage' payments*. Working Paper No. 11/33, Department of Economics, University of Leicester, June.
- Kikilias E. 2009. *An Inquiry into the Correlates of Informal Economy and Undeclared Work*. National Center for Social Research Institute of Social Policy (INSPO), Social Cohesion Bulletin n. 1.
- Klatzmann R. 1989. *Le Travail Noir*, P.U.F.
- Krumplytė J. 2010. *Analysis of undeclared work: the Lithuanian case*. 6th International Scientific Conference Vilnius May 13–14, 2010, Business and management selected papers.
- Kupiszewski M. e H. Mattila. 2008. *Addressing the irregular employment of immigrants in the European Union: between sanctions and rights*. International Organization for Migration (IOM) Regional Mission for Central and South Eastern Europe
- Lemieux T., B. Fortin - P. Frechette. 1994. The Effect Of Taxes On Labor Supply In The Underground Economy, *The American Economic Review*, N. 3, P. 232;
- MEF. 2011. *Economia non osservata e flussi finanziari*. Rapporto finale sull'attività del gruppo di lavoro sull'"Economia Sommersa e il Flussi Finanziari" istituito dal Ministero dell'Economia e delle Finanze.
- Meldolesi L. 1998. *L'economia Sommersa nel Mezzogiorno*, In *St. Mer.*, P. 325

- Merikull J. e K. Staehr. 2010. Unreported Employment and Envelope Wages in Mid-Transition: Comparing Development and Causes in the Baltic Countries. *Comparative Economic Studies*, n. 52, p. 637–670.
- Pacifico M. 2004. Le politiche per l'emersione del lavoro non regolare sono davvero fallite? Laboratorio di Politiche
- Reyneri E. 2001. Migrants involvement in irregular employment in the Mediterranean countries of the European Union. *International Migration Papers* (IMP), ILO
- Rossi N. 1997. Disoccupazione, Offerta di Lavoro e Lavoro Sommerso, In: De Nardis -Galli (a cura di), *La Disoccupazione Italiana*, Il Mulino, P. 82;
- Sivananthiran A. e C.S. Venkata Ratnam. 2005. *Informal economy : The Growing Challenge For Labour administration*. ILO-IIRA
- TNS Infratest Sozialforschung, Munich, Germany, The Rockwool Foundation Research Unit, Copenhagen, Denmark Regioplan Beleidsonderzoek BV; Amsterdam, Netherlands. 2006. *Feasibility Study on a direct survey about undeclared work*. Final report, financed by and prepared for the use of the European Commission, Directorate-General for Employment and Social Affairs
- Undocumented Worker Transitions. 2007. *Migration and Irregular Work in Europe. Literature Review*. Work package 3. EU Sixth Framework programme
- Williams C. e J. Windbank. 1995. Black market work in the European Community: peripheral work for peripheral localities? *International Journal of Urban and Regional Research*, vol. 19, p. 23 ss.
- Williams C.C. 2009(b) The prevalence of envelope wages in Europe. *Employee Relations*, n.31, p. 412-2.
- Williams C.C. 2009. Evaluating the extent and nature of envelope wages in the European Union: a geographical analysis. *European Spatial Resource Policy*, n. 16, p. 115-29.
- Williams C.C. 2010. Beyond the formal/informal jobs divide: evaluating the prevalence of hybrid 'under-declared' employment in South- Eastern Europe. *International Journal of Human Resources and Management*, n. 21, p.2529-46.
- Williams C.C. 2013. Evaluating Cross-National Variations in Under-Declared Wages. *The Open Area Studies Journal*, n. 5, p. 12-21.
- Williams C.C. e J. Padmore. 2013. Evaluating the prevalence and distribution of quasi-formal employment in Europe. *Industrial Relations*, n. 68, p. 71-95.
- Williams C.C. e S. Nadin. 2012. *Tackling undeclared work in the European union*. CESifo Forum 2.