

LA RILEVAZIONE TRIMESTRALE OROS SU OCCUPAZIONE E COSTO DEL LAVORO: INDICATORI E METODOLOGIE





LA RILEVAZIONE TRIMESTRALE OROS SU OCCUPAZIONE E COSTO DEL LAVORO: INDICATORI E METODOLOGIE

ISBN 978-88-458-1973-5

© 2019
Istituto nazionale di statistica
Via Cesare Balbo, 16 - Roma



Salvo diversa indicazione, tutti i contenuti pubblicati sono soggetti alla licenza Creative Commons - Attribuzione - versione 3.0. <https://creativecommons.org/licenses/by/3.0/it/>

È dunque possibile riprodurre, distribuire, trasmettere e adattare liberamente dati e analisi dell'Istituto nazionale di statistica, anche a scopi commerciali, 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.



INDICE

	Pag.
Premessa	7
Elenco sigle utilizzate nel volume	9
1. Caratteristiche generali della rilevazione Oros	11
1.1 La Rilevazione Oros nel sistema delle statistiche congiunturali sulla domanda e il costo del lavoro	11
1.2 L'esperienza Oros nei contesti di uso intensivo di dati di fonte amministrativa	13
1.3 Il modello alla base di acquisizione di input e fornitura di output di Oros	14
2. Obiettivi, fonti di dati e stime	17
2.1 Popolazione e variabili obiettivo, classificazioni e indici	17
2.2 Fonti di dati	21
2.2.1 <i>La fonte amministrativa Inps</i>	22
2.2.2 <i>Necessità di integrare con fonti statistiche e altre fonti amministrative</i>	23
2.3 Stime	25
2.3.1 <i>Stima preliminare e stima finale</i>	25
2.3.2 <i>I cambiamenti della situazione informativa e il loro impatto sulla metodologia di stima preliminare</i>	26
2.4 Le macro fasi del processo di produzione Oros	30
2.5 L'infrastruttura informatica alla base di Oros: macro fasi dell'indagine e principali tecnologie utilizzate	32
3. Le informazioni anagrafiche	37
3.1 Introduzione	37
3.2 Le fonti dati e la definizione dei criteri per la collocazione delle unità nel campo di osservazione	38
3.2.1 <i>L'Anagrafica Inps</i>	38
3.2.2 <i>L'archivio Asia e l'archivio delle partite IVA di fonte Istat</i>	40

	Pag.
3.2.3 <i>L'Anagrafe Tributaria di fonte Agenzia delle Entrate</i>	41
3.2.4 <i>L'integrazione delle fonti e la definizione dei criteri per la collocazione delle unità nel campo di osservazione</i>	41
3.2.5 <i>Il trattamento delle riclassificazioni delle variabili strutturali e anagrafica annuale</i>	43
4. Dal dato amministrativo alle informazioni statistiche. Banca dati normativa, controlli di qualità e criteri di aggregazione	47
4.1 I DM2013: copertura e variabili amministrative	47
4.2 Dalle variabili amministrative Inps alle variabili statistiche Oros: definizioni	49
4.3 La banca dati normativa Oros (BDN)	51
4.4 Le diverse fasi del trattamento preliminare	53
4.4.1 <i>Controlli sui file Ascii</i>	54
4.4.2 <i>Controlli di validità e coerenza dei codici</i>	54
4.4.3 <i>L'estrazione delle variabili statistiche</i>	55
5. Il trattamento trimestrale	59
5.1 Dai dati mensili ai dati trimestrali: attribuzione delle informazioni anagrafiche e controlli di congruenza	59
5.2 Stima delle Ula	60
5.3 Calcolo delle medie mensili e collocazione delle unità nel campo di osservazione	61
6. Editing e imputazione	65
6.1 Introduzione	65
6.2 Imputazione delle posizioni lavorative per la stima provvisoria	66
6.2.1 <i>Da un approccio macro ad un approccio micro</i>	66
6.2.2 <i>Dati "quasi completi" e stima della demografia d'impresa</i>	67
6.2.3 <i>Le mancate risposte nei dati amministrativi provvisori, caratteristiche</i>	68
6.2.4 <i>La predizione della lista delle unità attive non rispondenti</i>	71
6.2.5 <i>Ricostruzione del dato per regressione</i>	75
6.3 Stima provvisoria per la sottopopolazione delle interinali	76
6.4 Analisi dell'errore di revisione per componenti: approccio teorico	80
6.5 Imputazione nei dati finali	82
6.5.1 <i>Le mancate risposte nei dati finali</i>	82
6.5.2 <i>Il metodo di imputazione dei dati mancanti</i>	83
6.6 Editing selettivo e imputazione	84
6.6.1 <i>Quadro concettuale</i>	84
6.6.2 <i>Le funzioni di influenza: funzione finalizzata all'editing</i>	85
6.6.3 <i>Le funzioni di influenza: funzione finalizzata all'imputazione</i>	89

	Pag.
7. Integrazione con i dati di fonte indagine. L'individuazione della lista e la stima delle variabili economiche	93
7.1 Quadro di riferimento per l'integrazione con i dati di fonte indagine	93
7.2 Alcune caratteristiche dell'indagine sulle grandi imprese	95
7.3 L'individuazione della lista anagrafica Oros del panel dell'indagine sulle grandi imprese	96
7.3.1 <i>Definizione e mantenimento trimestrale</i>	96
7.3.2 <i>Definizione trimestrale per il cambio base</i>	98
7.4 La stima economica delle imprese del panel dell'indagine sulle grandi imprese	101
8. Il calcolo degli indicatori finalizzati alla diffusione nazionale	105
8.1 Introduzione	105
8.2 I dati aggregati nella banca dati Oros	106
8.2.1 <i>Le serie storiche dal 2010 e i dati correnti</i>	106
8.2.2 <i>Le serie storiche fino al 2009</i>	107
8.3 Dal lordo al netto Cig: la stima delle variabili di costo del lavoro per Ula	108
8.3.1 <i>Il quadro informativo di riferimento</i>	108
8.3.2 <i>Cenni normativi sulla Cig</i>	109
8.3.3 <i>Le fonti dati amministrative sulla Cig</i>	111
8.3.4 <i>Il file sulle ore di Cig autorizzata</i>	111
8.3.5 <i>Dinamica della Cig e impatto sugli indicatori di costo del lavoro</i>	112
8.3.6 <i>Metodo micro</i>	114
8.3.7 <i>Metodo macro</i>	116
8.3.8 <i>Calcolo degli indicatori al netto della Cig</i>	118
8.4 La stima delle posizioni lavorative totali: dalla ricostruzione delle componenti mancanti alla predisposizione delle serie storiche	120
8.4.1 <i>Il quadro informativo di riferimento</i>	120
8.4.2 <i>Le posizioni lavorative assenti per aspettative di breve durata ed altri motivi</i>	122
8.4.3 <i>Le posizioni lavorative assenti per Cig a zero ore</i>	125
8.4.4 <i>Validazione delle posizioni lavorative in serie storica: l'imputazione macro</i>	129
8.4.5 <i>Le posizioni lavorative in somministrazione</i>	130
8.4.6 <i>Ricostruzione delle serie storiche delle posizioni lavorative dipendenti totali</i>	134
8.4.7 <i>L'assegnazione delle informazioni di classificazione nella serie storica</i>	136
8.5 Gli indici Oros sul Sistema Informativo per il Trattamento degli Indicatori Congiunturali	137

	Pag.
9. La correzione per la stagionalità e gli effetti di calendario degli indicatori Oros	139
9.1 La metodologia	139
9.2 Il sistema degli indicatori Oros. Dall'approccio diretto all'approccio indiretto	140
<i>9.2.1 Configurazione concettuale dell'approccio diretto e indiretto</i>	140
<i>9.2.2 I pesi di aggregazione per l'approccio indiretto</i>	143
<i>9.2.3 Dall'approccio diretto ad un approccio "misto"</i>	146
10. I triangoli e l'analisi delle revisioni degli indicatori Oros	149
11. Gli indicatori per il regolamento LCI - Labour Cost Index	153
11.1 Il quadro di riferimento: Regolamenti europei	153
11.2 L'indice LCI: caratteristiche metodologiche	154
11.3 Le fonti	157
11.4 Il processo di produzione dell'indice LCI italiano	159
12. Gli indicatori per il regolamento STS - Short Term Business Statistics	163
12.1 Le richieste del Regolamento	163
12.2 La metodologia per la compilazione degli indicatori	165
Glossario	169
Riferimenti Bibliografici	173

PREMESSA¹

Questo documento rappresenta il primo manuale completo su fonti dati, scelte metodologiche, indicatori prodotti e tecnologie utilizzate dalla rilevazione statistica Oros, che da circa 15 anni produce indicatori trimestrali sul costo del lavoro e sulle posizioni lavorative dipendenti basati su una sistematica integrazione di dati di fonte amministrativa e dati statistici rilevati dall'Istat direttamente presso le imprese. Questa sua caratteristica l'ha resa particolarmente innovativa rispetto alle indagini tradizionali presenti fino a quel momento nel panorama nazionale, in particolar modo per le finalità di misurazione statistica delle dinamiche congiunturali che si è data fin dalla sua progettazione (Rapiti et al., 2010). Inoltre, si configura come caso esemplare anche nel contesto delle statistiche congiunturali europee (Maasing et al., 2012) dove per problemi tecnici, di qualità della fonte o anche di resistenza a modificare i processi di produzione, prevale il modello di utilizzo dei dati amministrativi come fonti ausiliarie rispetto alle indagini classiche (Costanzo, 2013).

La ricchezza informativa della principale fonte amministrativa usata - le dichiarazioni mensili che i datori di lavoro devono inviare all'Inps per i propri lavoratori dipendenti - ha rappresentato la base per la realizzazione di molteplici obiettivi informativi, con la realizzazione sia di prodotti statistici per la diffusione sia di basi di dati utili ad ulteriori processi statistici dell'Istituto. Il graduale raggiungimento degli obiettivi prefissati è stato ottenuto a seguito di uno sfruttamento crescente e sempre più consapevole di tali dati, integrati con altre fonti amministrative e statistiche progressivamente acquisite. L'uso dei dati amministrativi come fonte principale in Oros anticipa una strategia che è ora alla base del modello di produzione statistica dell'Istat, in cui non sono le fonti amministrative a integrare le indagini, ma le indagini a completamento delle fonti amministrative utilizzate quali basi dati primarie (Monducci, Luzi, 2016), in un contesto concettuale e metodologico innovativo basato sui registri statistici (Istat, 2016a).

La combinazione di fonti dati in Oros ha permesso di ampliare, nel tempo, il quadro informativo dal lato della domanda di lavoro nell'ambito delle statistiche ufficiali sul mercato del lavoro e di soddisfare diverse richieste europee obbligatorie per il Paese. Parallelamente, la predisposizione di una struttura informativa di micro e macro dati, armonizzati su variabili di classificazione e identificati con chiavi condivise, nonché di metadati organizzati ha reso il sistema Oros una fonte d'informazione ampiamente utilizzata in diversi processi statistici congiunturali e strutturali dell'Istat, alimentando con sistematicità anche approfondimenti analitici presenti nei principali prodotti informativi complessi che l'Istat produce annualmente (Rapporto annuale, Rapporto sulla competitività dei settori produttivi ecc.).

L'uso intensivo di fonti amministrative per finalità statistiche impone di dover affrontare nuove e complesse problematiche di trattamento dei dati di base (Wallgren et al., 2007; Zhang, 2012) tenendo conto, tra l'altro, del vincolo imprescindibile della tempestività, caratteristica che ha fortemente condizionato molte delle scelte analitiche e metodologiche effettuate nell'ambito della rilevazione. Tra gli aspetti da considerare, lo studio dei concetti

¹ Questa pubblicazione è stata curata da Donatella Tuzi. Hanno partecipato alla redazione del volume: Elisabetta Aquilini, Francesca Ceccato, Eleonora Cimino, Francesca Romana Pogelli, Marco Lattanzio e Donatella Tuzi. La premessa è a cura di tutti gli autori. La stesura del volume è terminata a marzo 2018. Si ringraziano Claudio Ceccarelli e Fabio Rapiti per la rilettura del volume e gli utili suggerimenti forniti.

ed il trattamento preliminare dei dati amministrativi per conferire loro significato statistico; l'armonizzazione delle variabili e unità statistiche *target*; il *matching* con altre fonti statistiche ed amministrative; la modellizzazione e la stima di mancate risposte; il *data editing* e la validazione dei micro e dei macrodati. In Oros tali aspetti vengono ampiamente affrontati, adottando scelte metodologiche in alcuni casi di tipo canonico, mentre in altri ricorrendo a soluzioni *ad hoc* tratte da *best practices* nazionali ed internazionali, avendo sempre a mente i principi della qualità dei dati prodotti e la tempestività imposta dalle diffusioni interne ed internazionali.

ELENCO DELLE PRINCIPALI SIGLE UTILIZZATE NEL VOLUME

- Asia:** Registro Statistico delle Imprese Attive
- Ateco:** Classificazione ufficiale delle attività economiche
- BDN:** Banca Dati Normativa
- BR:** Business Register
- CA:** Codice Autorizzazione
- CF:** Codice Fiscale
- Cig:** Cassa integrazione guadagni
- Cigd:** Cassa integrazione guadagni in deroga
- Cigo:** Cassa integrazione guadagni ordinaria
- Cigs:** Cassa integrazione guadagni straordinaria e solidarietà
- CSC:** Codice Statistico Contributivo
- CTC:** Codice Tipo Contributivo
- DA:** Dati Amministrativi
- Diptot:** Nel DM2013 dipendenti retribuiti
- DM:** Dichiarazioni Mensili contributive Inps
- DM10:** Dichiarazioni mensili contributive Inps a livello aziendale (esistenti fino al 2009)
- DM2013 virtuali:** Dichiarazioni mensili contributive aziendali ricostruite virtualmente dall'Inps (esistenti dal 2013)
- DMAG:** Denuncia della manodopera agricola
- E&I:** Procedura di *Editing* e Imputazione
- GI:** Indagine mensile su occupati, ore lavorate, retribuzioni e costo del lavoro nelle Grandi imprese
- GI_DA:** Grandi imprese assenti dall'indagine GI ma presenti nei dati amministrativi
- INTER:** Agenzie di somministrazione di lavoro temporaneo
- IRC:** Indagine mensile sulle Retribuzioni Contrattuali
- LCI:** Labour Cost Index
- LCS:** Labour Costs Survey
- MR:** Mancate risposte

- MRP:** Mancate risposte parziali
- MRT:** Mancate risposte totali
- Ndip:** Nel DM2013 dipendenti in forza presso l'azienda
- NO_GI:** Insieme delle unità Oros complementare all'insieme GI (PMI, GI_DA e INTER)
- Oros:** Rilevazione trimestrale su occupazione e costo del lavoro
- PMI:** Piccole e medie imprese stimate con dati amministrativi
- RACLI:** Registro Annuale sul Costo del Lavoro Individuale
- SBS:** Structural Business Statistics
- SD:** Stime finali/definitive
- SITIC:** Sistema Informativo per il Trattamento degli Indicatori Congiunturali
- SP:** Stime provvisorie/preliminari
- STS:** Short Term Business Statistics
- S13:** Lista delle istituzioni appartenenti al settore delle Amministrazioni Pubbliche
- Ula:** Unità di lavoro dipendente equivalenti a tempo pieno
- UniEmens:** Dichiarazioni mensili contributive Inps a livello individuale
- Vela:** Indagine trimestrale su posti vacanti e ore lavorate
- Vela-GI:** Sistema informativo integrato delle indagini Vela e GI

1. CARATTERISTICHE GENERALI DELLA RILEVAZIONE OROS¹

1.1 La Rilevazione Oros nel sistema delle statistiche congiunturali sulla domanda e il costo del lavoro

La rilevazione Oros² produce indicatori trimestrali su occupazione dipendente, retribuzioni, oneri sociali e costo del lavoro. Progettata alla fine degli anni '90 al fine di completare l'offerta di statistiche su tale dominio, la rilevazione si colloca attualmente all'interno di un sistema integrato di indicatori congiunturali attraverso cui l'Istituto Nazionale di Statistica diffonde informazioni su input e costo del lavoro nelle imprese ed istituzioni private dei settori dell'industria e dei servizi con l'esclusione del settore dell'amministrazione pubblica e difesa, assicurazione sociale obbligatoria. Tale sistema, i cui pilastri poggiano su tre rilevazioni, è andato gradualmente crescendo nel tempo, sia in termini di quantità e qualità delle statistiche prodotte sia nel modello generale di integrazione e progettazione. Sono parte del sistema: la rilevazione censuaria sull'occupazione, gli orari di lavoro, le retribuzioni e il costo del lavoro nelle grandi imprese (GI), che raccoglie e diffonde informazioni mensili sulle imprese con oltre 500 dipendenti (Istat, 2006); l'indagine campionaria su posti vacanti e ore lavorate (Vela), che diffonde indici relativi alle imprese con oltre 10 dipendenti³ coperti, per le imprese con oltre 500 dipendenti, dai dati raccolti dall'indagine GI (Bellisai et al., 2013); la rilevazione censuaria Oros che integra dati di fonte amministrativa (Dichiarazioni Mensili contributive Inps, cosiddetti DM) con i dati dell'indagine mensile GI, utilizza indicatori trimestrali desumibili dall'indagine Vela per stimare e diffondere indicatori su costo del lavoro e posizioni lavorative dipendenti (Figura 1.1.1).

La rilevazione Oros, sebbene di impianto non recente, rappresenta ancora oggi un caso innovativo di successo nel panorama delle statistiche europee che sfruttano dati amministrativi, in particolare per l'uso massivo che ne viene fatto a scopi congiunturali. Nata per soddisfare le richieste dei regolamenti europei, ha avuto come prima finalità il superamento di un rilevante *gap* informativo nell'ambito delle statistiche congiunturali sull'input ed il costo del lavoro (Baldi et al., 2011) in quanto, fino ad allora, solo la classe dimensionale 500 dipendenti e oltre veniva coperta attraverso l'indagine GI. Impostata sin da principio come parte di un sistema, la rilevazione Oros è stata configurata ed è andata via via sviluppandosi sulla base dei principi di minimizzazione delle incoerenze, di riduzione delle frammentazioni e di sovrapposizione tra input, fasi di processo ed output del sistema.

Nell'ambito del sistema integrato, si è andata sviluppando una complessa e fruttuosa interazione fra le varie fonti: la raccolta diretta dei dati avviene in modo coordinato, i microdati statistici Oros di fonte amministrativa vengono condivisi nell'ottica del riutilizzo, i processi di produzione, fortemente interdipendenti, vengono sempre più integrati e la politica di diffusione delle informazioni stabilita congiuntamente. A tal proposito, già a partire

1 Il presente capitolo è a cura di Donatella Tuzi.

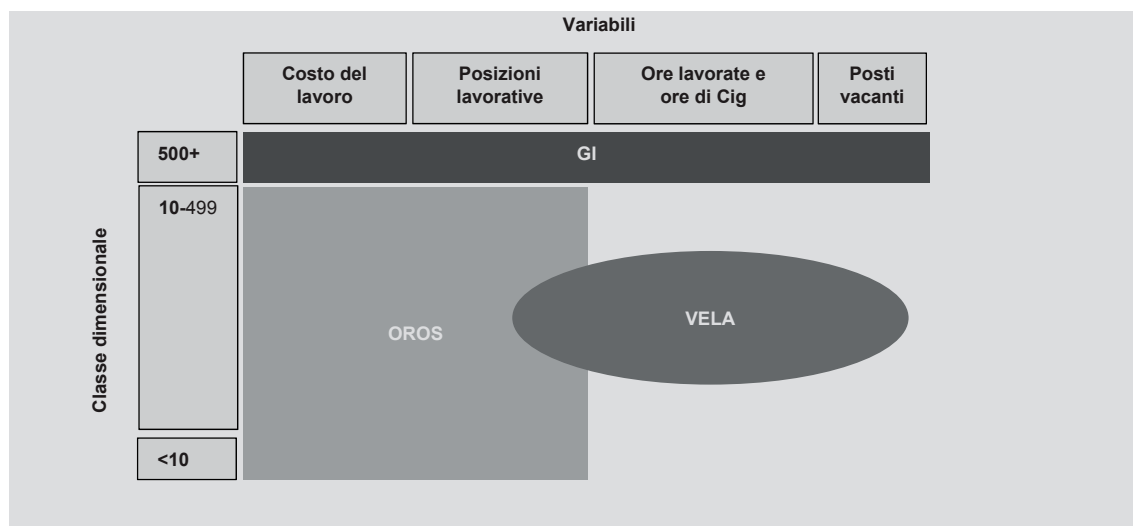
2 Oros è l'acronimo di Occupazione, Retribuzioni, Oneri Sociali.

3 A partire dai trimestri riferiti al 2016 il campione di Vela è stato ampliato per garantire la copertura delle imprese appartenenti alla classe dimensionale sotto i 10 dipendenti.

da dicembre 2013, i principali risultati statistici delle tre rilevazioni del sistema sono state accorpate in un unico comunicato stampa su “Indicatori del lavoro nelle imprese”, preferendolo alle diffusioni singole. Successivamente, a partire da settembre 2015, è stato realizzato un ulteriore passo verso una lettura congiunta delle statistiche sul mercato del lavoro con il comunicato stampa “Il Mercato del lavoro” in cui indicatori di offerta di lavoro (lato famiglie) e domanda di lavoro (lato imprese) vengono presentati insieme, allo scopo di fornire un quadro quanto più esaustivo possibile delle caratteristiche dell’occupazione e delle sue dinamiche. In tale ottica, da alcuni anni, vengono accostati nella lettura per la conferma delle dinamiche in atto anche gli indicatori trimestrali di input e costo del lavoro della Contabilità Nazionale; sebbene questi indicatori seguano definizioni e regolamenti specifici, utilizzano anche le serie storiche prodotte da Oros, Vela e GI come indicatori per la trimestralizzazione dei conti annui. Gli indicatori trimestrali prodotti dalla Contabilità Nazionale rientrano, a loro volta, nel sistema per consentire il rilascio degli indicatori sul costo orario del lavoro secondo quanto stabilito dal Regolamento Europeo su LCI (*Labour Cost Index*) e sulle posizioni lavorative totali e relativo monte retributivo secondo quanto richiesto dal Regolamento Europeo su STS (*Short Term Business Statistics*) (cfr. §2.1). Questa forte interdipendenza tra le diverse rilevazioni ha ovviamente influenzato la definizione dei tempi di rilascio dei principali output e, quindi, di processo, condizionando radicalmente anche le scelte metodologiche⁴.

Appartiene al sistema della domanda e del costo del lavoro anche l’Indagine mensile sulle retribuzioni contrattuali (IRC) (Istat, 2013a, pp.15-49) che viene utilizzata nel processo Oros solo rispetto ad alcuni macrodati sulle ore contrattualmente lavorabili, produce indicatori sulle retribuzioni contrattuali che fanno parte del comunicato congiunto sulla domanda di lavoro in quanto rappresentativi dell’evoluzione delle retribuzioni a seguito delle sole misure tabellari previste negli accordi collettivi nazionali di lavoro e non anche delle componenti di costo basate su criteri di “fatto”, e che costituiscono una base comparativa.

Figura 1.1.1 – La rilevazione Oros nel sistema integrato di statistiche congiunturali su input e costo del lavoro



⁴ Ne è un esempio la recente compressione dei tempi di produzione degli output del sistema richiesti dalla Contabilità Nazionale per rispondere alle richieste Eurostat di aumento della tempestività nel rilascio, compressione resa possibile a seguito della ridefinizione di diverse fasi dei processi di stima (Cfr. Regolamento UE n.549/2013 del Parlamento europeo e del Consiglio, relativo al Sistema europeo dei conti nazionali e regionali nell’Unione europea e successiva deroga contenuta nella decisione di esecuzione della Commissione del 26 giugno 2014).

1.2. L'esperienza Oros nei contesti di uso intensivo di dati di fonte amministrativa

La rilevazione Oros ha rappresentato il primo esempio di utilizzo di dati amministrativi a fini congiunturali in Italia (Rapiti et al., 2010) e un caso raro anche nel panorama europeo. Nell'Ue, infatti, l'uso dei dati amministrativi viene fortemente promosso e considerato un'opportunità, ma limitatamente applicato (Costanzo, 2012). In generale, il principale ostacolo allo sviluppo di statistiche basate sull'uso diretto di dati amministrativi non riguarda tanto la disponibilità di una fonte adeguata, quanto piuttosto le difficoltà legate a problemi tecnici e di qualità tra cui tempestività della fonte, comparabilità, differenze definitorie e di classificazione, nonché forme di resistenza a rimodellare i processi di produzione (Costanzo, 2013). Per questo, spesso i dati sono utilizzati in molti casi solo quali fonti ausiliarie rispetto ad indagini classiche. Rispetto agli altri paesi europei, in particolare a quelli nordici, sull'utilizzo dei dati amministrativi a fini statistici l'Istat è stato in ritardo per molti anni. Solo nella seconda metà degli anni '90, lo sviluppo dell'archivio statistico delle imprese attive (Asia), ampiamente basato su fonti amministrative, ha fatto crescere l'interesse nei confronti dell'uso di dati amministrativi anche per indagini correnti. Nel corso degli anni l'Istituto ha puntato in modo crescente sui dati amministrativi, acquisendo diverse fonti principalmente con scopi strutturali, ed usate anche per studi e/o approfondimenti specifici. Per massimizzare il vantaggio derivante dalla disponibilità di tale quantità di informazioni, nel 2005 l'Istituto si è mosso verso l'accentramento di alcune funzioni per l'acquisizione, l'archiviazione, l'integrazione e la valutazione della qualità dei dati amministrativi, creando un sistema interconnesso di dati che, da ultimo, ha portato alla realizzazione del 9° Censimento dell'industria basato per la prima volta in gran parte su dati amministrativi (Istat, 2016b).

A partire dai dati sul 2012, anche, il regolamento europeo sulle *Structural Business Statistics* (SBS) viene soddisfatto attraverso l'elaborazione dei dati del nuovo sistema informativo Frame-SBS dell'Istat, microdati di fonte amministrativa trattati statisticamente e combinati con dati delle rilevazioni dirette sulle Piccole e Medie Imprese ed esercizio di arti e professioni e sul Sistema dei Conti delle Imprese, allineati sulle variabili strutturali e di classificazione con quelle di Asia (Luzi e Monducci, 2016). Poi, dal 2013 è in funzione il nuovo Sistema integrato di microdati (SIM), finalizzato alla realizzazione dell'integrazione di tutti i microdati acquisiti dalle diverse fonti amministrative (Runci M.C. et al., 2016). Tale sistema contiene dati di diverse fonti amministrative identificati in maniera univoca e già integrati, o integrabili, in modo automatico.

Dal 2016 l'Istat ha inoltre avviato un processo di modernizzazione che ha al centro il passaggio da un sistema basato sulle rilevazioni sul campo a uno basato sui Registri realizzati a partire da fonti amministrative⁵. Il disegno prevede la realizzazione di un Sistema Integrato di Registri (SIR) composto da Registri Statistici di Base (RSB) che identificano alcune tipologie di oggetti e popolazioni "di base"⁶, i Registri Statistici Estesi (RSE) e i Registri Statistici Tematici (RST). Sul piano organizzativo, per garantire la trasversalità e la condivisione dei dati raccolti da fonti amministrative e consentire il loro utilizzo in maniera sistematica in tutti i processi produttivi, l'Istituto si è dotato di una infrastruttura operativa

5 Per una descrizione più dettagliata del progetto del sistema basato sui Registri si vedano i lavori presentati in occasione della 12° Conferenza nazionale di statistica, tenutasi a Roma il 22-24 giugno 2016: <https://www.istat.it/it/dodicesima-conferenza/programma/051-strumenti-modernizzazione>.

6 In particolare: RSB di individui, famiglie e convivenze; RSB delle unità economiche; RSB degli edifici e del territorio; RSB delle attività (relazioni lavorative, di studio, ecc.) che mette in relazione gli oggetti degli altri registri base.

su “Fonti amministrative e integrazione dei registri”, punto di riferimento di tutte le attività connesse agli archivi amministrativi, con il mandato di regolare il rapporto con gli enti, l’accesso e l’uso condiviso degli archivi amministrativi, la promozione di standard comuni di trattamento dei dati, al fine di mantenere la coerenza e la qualità complessiva della produzione statistica. A supporto delle scelte metodologiche e progettazione dei processi di produzione sottostanti, inoltre, è stata istituita l’infrastruttura “Disegno dei processi e supporto al sistema dei registri”.

In questo nuovo e più articolato scenario informativo e infrastrutturale, è prevista la reingegnerizzazione della rilevazione Oros. Per adesso, la rilevazione, essendo a regime da oltre 15 anni e dovendo garantire la diffusione di indicatori a livello nazionale ed europeo, non ha ancora raggiunto una piena integrazione. Essa dovrebbe rappresentare, comunque, un esempio a cui rifarsi, sia rispetto alle potenzialità sia rispetto alle criticità da evitare, nel delineare modelli in cui i dati amministrativi vengano utilizzati anche per finalità congiunturali. La frequenza, la puntualità con cui i dati devono essere acquisiti e controllati, la struttura dei dati e metadati sottostanti, sono specifiche caratteristiche che da sempre hanno suggerito l’adozione di controlli diretti delle varie fasi del processo Oros da parte dei tecnici della rilevazione e il mantenimento di uno stretto rapporto di collaborazione con il personale degli enti fornitori dei dati. Per esperienza, queste attività sono legate al raggiungimento degli obiettivi a breve termine e non da essi facilmente divisibili. La parziale indipendenza rispetto al sistema integrato dei registri non deve, per questo, essere necessariamente considerata un limite ma piuttosto una qualità del sistema informativo Oros, il cui sfruttamento avviene comunque in diversi processi produttivi dell’Istituto, non finalizzati unicamente alla congiuntura.

1.3 Il modello alla base di acquisizione di input e fornitura di output di Oros

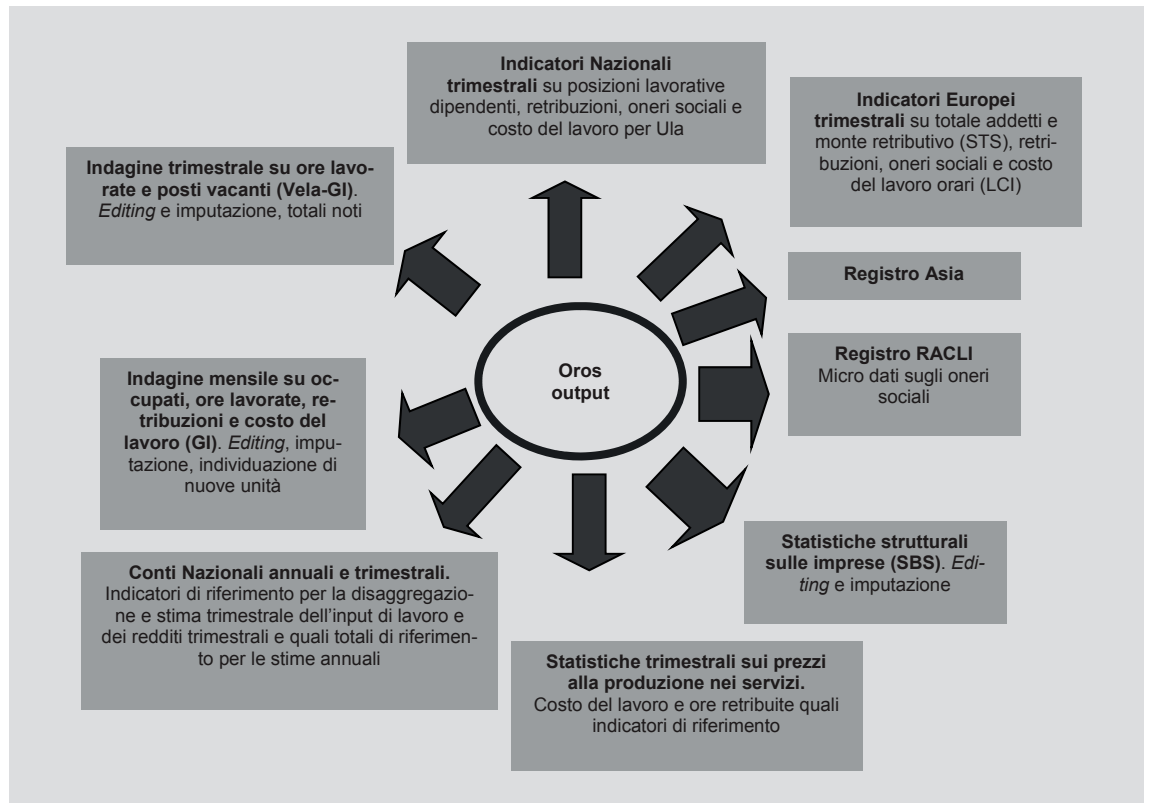
Durante l’iniziale fase di progettazione della rilevazione, al termine di un lungo periodo di studi preliminari, fu deciso che le dichiarazioni contributive che i datori di lavoro sono obbligati ad effettuare all’Inps per denunciare le retribuzioni mensili corrisposte ai dipendenti e i contributi dovuti (d’ora in avanti dichiarazioni contributive) fossero la fonte dati adeguata agli scopi dell’indagine. Tale consapevolezza superava la constatazione sul limite più rilevanti di tale fonte, ossia l’elevato livello di disaggregazione e il dettaglio delle informazioni contenute su retribuzioni ed occupazione, nonché la frammentazione di metadati utili per transcodificare i dati amministrativi in informazioni statistiche, caratteristiche che ne avrebbero reso lo sfruttamento molto complesso (Baldi et al., 2008; Rapiti et al., 2011). Inoltre, gli obiettivi di tempestività, portarono a preferire l’acquisizione dei dati originari e non riaggregati da parte dell’Inps, operazione che avrebbe impedito di raggiungere le scadenze prefissate. Fu stabilito, quindi, di acquisire l’intero *dataset* in formato grezzo, senza alcuna forma di pretrattamento (cfr. capitolo 4). Tale approccio ha implicato dover superare diverse criticità, sia in fase progettuale sia nelle fasi successive di sviluppo. Anzitutto, la necessità di implementare un complesso processo di pretrattamento, calcolo, trasformazione di informazioni amministrative (microcodici) in informazioni statistiche, controlli finalizzati ad ottenere microdati relativi alle variabili d’interesse adeguatamente identificate e aggregate; inoltre, l’esigenza di catturare ogni trimestre una considerevole mole di dati da dover in seguito gestire, mantenere (oltre 9 milioni di record per ogni mese considerato). Tuttavia, queste criticità hanno rappresentato anche un’importantissima opportunità, derivante dalla possibilità di controllare in modo diretto il processo di aggregazione e traduzione, rendendo

disponibili una quantità enorme e dettagliata di informazioni utili anche per altri processi statistici (Congia, Pacini, Tuzi 2008a; 2008b).

I microdati amministrativi dell'Inps sulle dichiarazioni contributive rappresentano una fonte ricchissima di informazioni a livello di impresa sulla composizione della forza lavoro (qualifica, tipo di contratto, ecc.), le loro retribuzioni e le varie componenti del costo del lavoro. Inizialmente, considerati i vincoli di tempo e risorse disponibili, l'Istat decise di pianificare la rilevazione Oros esclusivamente con lo scopo di raggiungere gli obiettivi specifici di compilazione di indicatori congiunturali su retribuzioni e costo del lavoro. Tra il 2000 e il 2003 Oros è stata impostata su una singola linea di produzione specializzata nella compilazione di tre tipi di indicatori congiunturali: stime trimestrali su occupazione e retribuzioni per il regolamento europeo *Short-Term Business Statistics* (regolamento STS, n. 1503/2006 del Consiglio europeo), indici di costo del lavoro orario per il regolamento europeo *Labour Cost Index* (regolamento LCI, n.450/2003 del Consiglio europeo) e, infine, indici di retribuzione e costo del lavoro per Unità di lavoro a tempo pieno (Ula) per il rilascio nazionale. Nel corso del tempo, gli sviluppi del processo sono stati principalmente finalizzati a migliorare la qualità degli indicatori prodotti. Sono state utilizzate fonti di input sempre più specializzate che si sono progressivamente rese disponibili anche per effetto di un graduale miglioramento nella tecnologia a disposizione dell'Istituto. Implementazioni *ad hoc* hanno, inoltre, consentito di rispondere alle molteplici richieste di microdati e macrodati che si sono via via presentate. D'altra parte, l'interesse nei confronti di questa fonte è progressivamente aumentato anche a seguito dei miglioramenti in termini di qualità e fruibilità raggiunti nel tempo, sollecitando il passaggio verso una strategia di "sistema" in cui le informazioni disponibili o deducibili possano essere utilizzate come base dati per molteplici obiettivi statistici (Rapiti et al., 2011). Questo cambiamento rientra in una più generale visione di sistema integrato che è ormai consolidata in Istituto, in cui sin dalla progettazione il fine sia quello di produrre statistiche nelle quali sempre più indicatori, per domini specifici, non siano più prodotti indipendentemente l'uno dall'altro ma per integrazione di *datasets* e combinando dati da diverse fonti, amministrative e d'indagine. Oros in questo contesto può essere vista come un'infrastruttura statistica comune, specializzata nell'occupazione, le retribuzioni e il costo del lavoro, integrata in un sistema comprensivo di statistiche: a) utili a compilare direttamente indicatori per scopi congiunturali; b) quale input di molti processi di Contabilità Nazionale trimestrale ed annuale; c) come fonte di variabili per alcune indagini strutturali o per i registri; d) come informazione ausiliaria per validare e imputare a livello di microdato; e) come fonte di microdati da abbinare con molte altre fonti amministrative, ecc. (Figura 1.3.1).

Pur dotato di un'apprezzabile flessibilità nella struttura informativa che ha consentito di estendere la produzione di informazioni, introdurre strutture concettuali e definitorie nuove, gli scopi congiunturali di Oros hanno limitato, nel tempo, la piena condivisione delle implementazioni necessarie per un trattamento omogeneo delle principali informazioni, limitandone le potenzialità informative. Nel nuovo contesto, i vincoli che nel passato non hanno consentito di raggiungere determinati obiettivi, potranno essere superati quando la rilevazione verrà dotata di un'adeguata infrastruttura informatica e di un costante supporto nello sviluppo delle procedure.

Figura 1.3.1 - L'output della rilevazione Oros nel sistema delle statistiche



2. OBIETTIVI, FONTI DI DATI E STIME¹

2.1 Popolazione e variabili obiettivo, classificazioni e indici

La popolazione obiettivo della rilevazione Oros sono le imprese e le istituzioni private con dipendenti, di tutte le classi dimensionali, che hanno corrisposto nel trimestre di riferimento retribuzioni imponibili a fini contributivi e svolgono la loro attività economica nei settori dell'industria (sezioni di attività economica da B ad F della classificazione Ateco 2007) e dei servizi (sezioni da G a S ad esclusione di O). L'insieme degli occupati dipendenti comprende operai, impiegati e apprendisti indipendentemente dal tipo di contratto (tempo indeterminato, determinato, stagionale, ecc.) e dal tipo di prestazione lavorativa (tempo pieno, tempo parziale). I dirigenti, esclusi dagli indici relativi al costo del lavoro, sono compresi nella stima delle posizioni lavorative dipendenti. I lavoratori in somministrazione vengono considerati dal lato delle società fornitrici e sono, quindi, inclusi nella sezione N "noleggio, agenzie di viaggio, servizi di supporto alle imprese", come esplicitato nelle raccomandazioni dei regolamenti europei (si veda in particolare il Regolamento STS).

Le principali variabili *target* sono il costo del lavoro e le sue due componenti - le retribuzioni e gli oneri sociali - e le posizioni lavorative dipendenti, sintetizzate attraverso numeri indici diffusi in forma grezza e destagionalizzata. Insieme agli indici vengono diffuse variazioni tendenziali e congiunturali. Indici e variazioni sono rilasciati a livello di sezione Ateco 2007 e sue aggregazioni, utilizzando valori arrotondati alla prima cifra decimale, coerentemente con gli standard di comunicazione e diffusione usati a livello europeo. In particolare, accanto alle singole sezioni da B a S, esclusa O, vengono pubblicati gli aggregati: industria (B-F); industria in senso stretto (B-E), servizi (G-S); servizi di mercato (G-N) di cui anche servizi alle imprese (L-N); istruzione, sanità e assistenza sociale, attività artistiche e altre attività di servizi (P-S)²; industria e servizi di mercato (B-N); industria e servizi (B-S). Solo per le posizioni lavorative dipendenti, oltre agli indici, vengono diffusi anche i rispettivi livelli accompagnati da variazioni congiunturali e variazioni tendenziali anch'esse calcolate come differenze assolute.

Per descrivere la dinamica delle variabili di costo del lavoro vengono calcolati: l'indice delle retribuzioni lorde medie per unità di lavoro equivalenti a tempo pieno (Ula), l'indice degli oneri sociali medi per Ula e l'indice del costo del lavoro medio per Ula, come sintesi dei due precedenti. Le Ula utilizzate sono una misura dell'occupazione al netto dei dipendenti posti in cassa integrazione guadagni (Cig). L'utilizzo delle Ula al netto della Cig si rende necessario per consentire una più precisa valutazione dell'input di lavoro che effettivamente concorre al processo produttivo ed è oggetto di remunerazione e contribuzione, permettendo confronti temporali più omogenei delle variabili *target* fra periodi con maggiore o minore

¹ Il paragrafo 2.1 di questo capitolo è a cura di Marco Lattanzio e Donatella Tuzi; i paragrafi 2.2, 2.3, 2.4 sono a cura di Donatella Tuzi; il paragrafo 2.5 è a cura di Francesca Ceccato.

² L'allargamento ai settori da P a S è avvenuto con la diffusione di dicembre 2013, contestualmente all'indagine Vela-GI. Tale ampliamento nel caso di Oros è stato reso possibile grazie alla definizione di criteri di individuazione della natura istituzionale dell'unità, come verrà diffusamente descritto nel capitolo 3.

ricorso alla Cig da parte delle imprese. Le Ula al netto della Cig vengono stimate come differenza tra le Ula complessive e quelle in Cig parziale nel mese (cfr. §8.3).

Le variabili di costo del lavoro per unità di lavoro sono ottenute dividendo la media mensile calcolata sul trimestre dei valori assoluti del monte delle variabili di riferimento (C per il costo, R per le retribuzioni e O per gli oneri sociali) per il corrispondente numero medio di Ula. I rispettivi numeri indice riferiti al generico trimestre t , sono *indici semplici* ottenuti rapportando la singola componente di costo del lavoro per Ula media mensile nel trimestre per la corrispondente media mensile calcolata su tutti i trimestri dell'anno base (tb). Facendo riferimento al costo del lavoro per Ula, il relativo indice, calcolato a livello settoriale, può essere espresso come:

$$I_t^{CU} = \frac{\frac{C_t}{Ula_t}}{\frac{1}{4} \sum_{tb=1}^4 \frac{C_{tb}}{Ula_{tb}}} \cdot 100 \quad (1)$$

In modo analogo si costruisce l'indice degli oneri sociali (I_t^{OU}) e l'indice delle retribuzioni (I_t^{RU}). Questa tipologia di indice ha la caratteristica di preservare la proprietà di additività che lega il costo del lavoro alle retribuzioni e agli oneri sociali, attraverso pesi che dipendono esclusivamente dai valori delle due variabili componenti all'anno base (cfr. §9.2.2). Tale caratteristica consente di ricavare gli indici del costo del lavoro anche indirettamente, come media ponderata degli indici delle retribuzioni e degli oneri sociali favorendo, tra l'altro, l'applicazione dell'approccio indiretto alla destagionalizzazione (cfr. §9.2.1). La relazione che lega gli indici delle singole sezioni agli indici dei rispettivi aggregati Ateco, invece, non è esprimibile come media ponderata ma come somma ponderata, con coefficienti dipendenti dal valore delle variabili nel trimestre di riferimento, rendendo un eventuale approccio indiretto alla destagionalizzazione su tale dimensione maggiormente complicato (cfr. §9.2.2).

Analogamente all'indagine GI, i numeri indice sulla dinamica delle variabili di costo del lavoro prodotte da Oros sono al lordo delle trattenute fiscali e previdenziali. Tali indicatori si riferiscono alle effettive erogazioni mensili corrisposte dalle imprese, secondo un criterio "di cassa" e non "di competenza". Ciò comporta ampie variazioni degli indici nei trimestri in cui vengono corrisposte mensilità aggiuntive e/o in cui si verificano circostanze di carattere episodico (corresponsione di premi, arretrati e gratifiche, slittamento di pagamenti di mensilità aggiuntive, rinnovi contrattuali, ecc.). Inoltre, tali indici si differenziano da quelli di "prezzo del lavoro" (ad esempio, l'indice delle retribuzioni contrattuali mensile) poiché, oltre a registrare l'evoluzione delle retribuzioni e del costo del lavoro di fatto, incorporano anche l'effetto dei mutamenti nella composizione dell'occupazione. I fattori che possono contribuire a tali mutamenti sono numerosi: variazioni nella proporzione di contratti di diverso tipo o di personale con anzianità diversa, mutamenti nella composizione dei dipendenti per livelli e qualifiche professionali, variazioni nella distribuzione dell'occupazione fra settori costituiscono gli esempi più rilevanti. Una delle conseguenze di questa caratteristica è che la proprietà della coerenza interna nel sistema degli indici Oros non è garantita, pertanto l'indice di un aggregato può presentare un valore che non è compreso tra il massimo e il minimo degli indici dei singoli settori che compongono l'aggregato stesso. Un effetto analogo si può presentare per i tassi di variazione calcolati su qualsiasi periodo³.

³ Il tema degli effetti di composizione e dell'assenza della proprietà della coerenza interna degli indici Oros sul costo del lavoro per Ula è approfondito nei lavori Lattanzio, 2016a e Lattanzio, 2016b in cui, accanto ad una descrizione delle

La rilevazione Oros diffonde, inoltre, indici e livelli sul numero delle posizioni lavorative dipendenti (in breve, posizioni totali), in aderenza alla definizione statistica dettata dal regolamento STS. In base a tale definizione, si definisce “posizione lavorativa dipendente” un contratto di lavoro tra una persona fisica e una unità produttiva (impresa o istituzione privata), che prevede lo svolgimento di una prestazione lavorativa a fronte di un compenso (retribuzione). Le posizioni lavorative (o “jobs”) rappresentano, quindi, il numero di posti di lavoro occupati da lavoratori dipendenti (“persone”), a tempo pieno e a tempo parziale, indipendentemente dalle ore lavorate, ad una determinata data di riferimento⁴. Come il numero di occupati anche le posizioni lavorative rappresentano pertanto una variabile di *stock* in un certo istante nel tempo. Sono inclusi anche i lavoratori che, legati all’unità produttiva da regolare contratto di lavoro, sono temporaneamente assenti per cause varie quali: ferie, permessi, maternità, cassa integrazione guadagni, ecc.

Accanto alle posizioni totali, Oros rilascia anche indici e livelli sul numero delle posizioni di lavoro in somministrazione, stimati sulla base dei dati delle agenzie di lavoro in somministrazione (ex-interinali) presenti nei dati amministrativi. Quest’ultimo indicatore rappresenta una componente della domanda di lavoro particolarmente sensibile all’andamento del ciclo economico, con caratteristiche di anticipazione dell’andamento dell’occupazione complessiva. Al fine di ottenere una misurazione statisticamente rilevante dei lavoratori in somministrazione, i dati amministrativi sono sottoposti ad una particolare procedura di trattamento, per ottenere il numero di posizioni lavorative a partire da quello dei rapporti di lavoro (missioni). Analogamente, vengono ricondotte a posizioni lavorative a tempo pieno anche i rapporti di lavoro con contratto intermittente (chiamate) (cfr. §4.4.3).

La stima delle posizioni lavorative totali è ottenuta attraverso trattamenti specifici attuati sulla variabile posizioni retribuite, già usata nel calcolo delle Ula. Tali trattamenti riguardano solo i dati delle unità derivanti dalla fonte amministrativa e non quelli delle unità rilevate dall’indagine GI. In particolare, sono due gli aspetti approfonditi: 1) l’assenza di dichiarazioni contributive ritardatarie per cui si rende necessaria l’imputazione del dato amministrativo (cfr. §6.2, §6.5, §6.6); 2) l’assenza nelle dichiarazioni contributive pervenute all’Inps delle informazioni relative ai dipendenti non retribuiti, poiché assenti per aspettativa di breve durata e collocamento in Cig per l’intero mese, inclusi nella definizione statistica della variabile. In tal caso si interviene misurando la variabile obiettivo con il supporto di informazioni ausiliarie (cfr. §8.3). Un trattamento a parte, inoltre, è riservato alle agenzie di somministrazione di lavoro, per cui sia nella ricostruzione della lista dei rispondenti ritardatari sia nell’imputazione del dato mancante, si tengono fortemente in considerazione le informazioni longitudinali disponibili sulle singole unità, la cui imputazione viene individualmente controllata e validata (cfr. §6.3). L’imputazione dei dati mancanti viene effettuata anche sulle variabili finalizzate agli indicatori di costo del lavoro. Tuttavia, considerato il più contenuto impatto che le mancate

caratteristiche metodologiche degli indici utilizzati, viene illustrata una formalizzazione utile a scindere gli effetti di composizione dell’occupazione e del costo del lavoro sull’evoluzione dell’indice complessivo fornendo, in tal modo, una chiave di lettura più chiara nell’analisi della congiuntura.

4 Per evidenziare la differenza tra il concetto di posti di lavoro - tipicamente rilevati dalle indagini economiche sulle imprese - e di persone occupate - concetto peculiare dell’indagine sulle Forze di lavoro - si riporta un esempio presente nel Manuale metodologico sulle statistiche economiche congiunturali, di cui al sopra citato regolamento STS: una persona occupata in un’impresa industriale durante il giorno e in un bar nelle ore serali nelle indagini sulle imprese viene conteggiata sia nell’attività manifatturiera, sia nei servizi mentre nell’indagine sulle Forze di lavoro, la persona viene conteggiata una sola volta, quale persona occupata. Pur rilevando anche le seconde posizioni, l’indagine sulle Forze Lavoro considera la posizioni principale, ossia quella su cui il lavoratore ha svolto più ore di lavoro o, a parità di ore, quella per cui riceve una remunerazione maggiore. Per tornare all’esempio originale si veda: <https://ec.europa.eu/eurostat/documents/1916593/1917176/KS-BG-06-001-EN.pdf>.

risposte hanno su tali variabili pro-capite, l'imputazione viene effettuata secondo criteri di selettività, ossia limitata ad un insieme ridotto di unità influenti (cfr. §6.6).

Anche per descrivere l'evoluzione delle posizioni lavorative dipendenti (D) si ricorre a *indici semplici*, calcolati per ciascun trimestre (t) dividendo, per ogni generico settore (j) le posizioni lavorative medie mensili⁵ per il corrispondente valore medio mensile calcolato sui trimestri dell'anno base (tb):

$$I_{j,t}^D = \frac{D_{j,t}}{\frac{1}{4} \sum_{tb=1}^4 D_{j,tb}} \cdot 100 \quad (2)$$

Diversamente dalle variabili di costo del lavoro, gli indici sulle posizioni lavorative per aggregazioni di Ateco possono essere ottenuti per somma ponderata degli indici relativi alle sezioni componenti, attraverso pesi fissi che dipendono dai valori della variabile all'anno base (cfr. §9.2.2) favorendo, anche in questo caso, l'applicazione dell'approccio indiretto di destagionalizzazione (cfr. §9.2.1). Accanto agli indici, come ricordato poco sopra, viene diffuso il livello delle posizioni lavorative, vale a dire il numeratore della (2), normalmente diffuso in migliaia.

Gli indici sulle retribuzioni, gli oneri sociali e il costo del lavoro per Ula nonché sulle posizioni lavorative dipendenti vengono rilasciati a livello nazionale mediante la statistica Flash "Il mercato del lavoro" e nella banca dati I.Stat, a circa 70 giorni dalla fine del trimestre di riferimento (Tavola 2.1.1). Le posizioni lavorative espresse come livelli sono, invece, rilasciate nella "Nota trimestrale congiunta sulle tendenze dell'occupazione", un prodotto congiunto tra Ministero del lavoro e delle politiche sociali, Istat, Inps, Inail e Anpal diffuso a circa 86 giorni dalla fine del trimestre di riferimento, a seguito di un accordo inter-istituzionale tra gli enti finalizzato a fornire una lettura e una valutazione coordinata dei dati sul mercato del lavoro.

Come già anticipato nel paragrafo 2.1 i dati prodotti con la rilevazione Oros vengono utilizzati per rispondere alle richieste dei regolamenti europei STS e LCI. Il primo regolamento richiede indici grezzi trimestrali sul numero totale di persone occupate (*Number of persons employed*) che, accanto alle posizioni dipendenti, include gli indipendenti⁶ (Tavola 2.1.1). La loro data di rilascio ad Eurostat è prevista entro 60 giorni dalla fine del trimestre di riferimento. Lo stesso regolamento richiede anche indici grezzi sul monte retributivo dei lavoratori dipendenti (Gross wages and salaries) da rilasciare entro 90 giorni dalla fine del trimestre di riferimento. In entrambi i casi, gli indicatori sono riferiti ad imprese ed istituzioni private nei settori Ateco 2007 da B ad N e sono rilasciati a particolari livelli di dettaglio Ateco (cfr. capitolo 12). Il regolamento LCI richiede, invece, indicatori di costo del lavoro espressi in termini di ore lavorate per le imprese e istituzioni, private e pubbliche, operanti in tutti i settori di attività economica da B a S, inclusa O, da rilasciare entro 70 giorni dalla fine del trimestre di riferimento (cfr. capitolo 11). Tali indici, che fanno parte dei PEEIs (Principal European Economic Indicators), un particolare sottoinsieme di indicatori congiunturali prescelti da Eurostat per descrivere l'evoluzione della situazione economica nell'Unione Europea⁷ vengono trimestralmente usati da Eurostat per calcolare l'aggregato europeo per l'*Hourly Labour Costs* rilasciato trimestralmente con la Newsrelease Euroindi-

5 Le posizioni lavorative medie mensili sono calcolate come media, nel trimestre, dei valori assoluti del monte mensile della variabile di riferimento.

6 Sono lavoratori indipendenti coloro che svolgono la propria attività lavorativa in una impresa, senza vincoli formali di subordinazione, con una remunerazione avente natura di reddito misto di capitale e lavoro (per dettagli sulla definizione si veda:

https://www.istat.it/it/files/2013/12/Struttura-e-dimensione-delle-imprese-09_dic_2013-Nota-metodologica.docx.pdf.

7 Per maggiori informazioni sui PEEI's si veda: <https://ec.europa.eu/eurostat/web/euro-indicators/scoreboard>.

2. Obiettivi, fonti di dati e stime

cators e pubblicati per sezione Ateco2007 sul database Euroind⁸. Questi indicatori vengono inoltre elaborati da Eurostat per ottenere una stima annua dell' *Hourly Labour Costs* in euro⁹ ottenuta estrapolando le stime quadriennali sul costo del lavoro prodotte con la *Labour Costs Survey* (LCS) mediante le tendenze annuali ricavabili dall'LCI.

Tavola 2.1.1 - Gli indicatori ufficiali rilasciati dalla rilevazione Oros

TIPO DI RILASCIO	VARIABILE	LIVELLO DI DETTAGLIO	TIPO DI INDICATORE RILASCIATO	TEMPESTIVITA'
DIFFUSIONE NAZIONALE	Retribuzioni lorde di fatto per Ula al netto della Cig Oneri sociali per Ula al netto della Cig Costo del lavoro per Ula al netto della Cig Posizioni lavorative dipendenti	Sezioni Ateco 2007 da B a S, ad esclusione di O, e relativi aggregati (1)	Indici trimestrali grezzi e destagionalizzati, variazioni tendenziali e congiunturali. Per le posizioni lavorative dipendenti anche livelli e differenze tendenziali e congiunturali	Circa 70 giorni dalla fine del trimestre di riferimento; 86 per le posizioni lavorative dipendenti espresse in livello
REGOLAMENTO STS (1165/98)	Number of persons employed Gross wages and salaries	Sezioni Ateco 2007 da B a N escluse K e L, MIGS e divisioni	Indici trimestrali grezzi	Entro 60 giorni dalla fine del trimestre di riferimento Entro 90 giorni dalla fine del trimestre di riferimento
REGOLAMENTO LCI (450/2003)	Hourly total labour costs Hourly wages and salaries Hourly employers' social security contributions plus taxes paid minus subsidies received by the employer Hourly total labour costs Non-wage costs (% of total)	Sezioni Ateco 2007 da B a S e relativi aggregati Macro aggregati per i settori da B a S (escluso O) in Ateco 2007 Aggregato B-S (escluso O) in Ateco 2007	Indici trimestrali grezzi, aggiustati per gli effetti di calendario e destagionalizzati Rapporti medi annuali grezzi % calcolata su valori medi annuali grezzi	Entro 70 giorni dalla fine del trimestre di riferimento Entro 90 giorni dalla fine del quarto trimestre dell'anno a-1

Fonte: Rilevazione Oros (occupazione, retribuzioni, oneri sociali)

(1) Le posizioni lavorative dipendenti espresse in livello relative alle sezioni B, L, M e N non vengono diffuse singolarmente, ma incluse in aggregati. Diversamente, gli indici della stessa variabile sono disponibili anche a livello delle singole sezioni citate.

(2) <http://eur-lex.europa.eu/LexUriServ/sts>

(3) <http://eur-lex.europa.eu/legal-content/LCI>

2.2 Fonti di dati

Le statistiche prodotte da Oros si basano, come già accennato, sulla combinazione di due tipologie di fonti: dati d'indagine, per la stima delle unità di grandi dimensioni (GI) e dati amministrativi (d'ora in avanti anche DA), utilizzati per la stima delle unità di piccola e media dimensione e per le unità di grandi dimensioni non coperte dall'indagine GI. Le informazioni che derivano dalle varie fonti vengono armonizzate in accordo con le definizioni e metodologie statistiche richieste e combinate attraverso tecniche di *linkage* deterministico, come verrà ampiamente approfondito nel capitolo 7.

Il perseguimento di obiettivi di qualità degli indicatori prodotti dall'indagine, le frequenti richieste di implementazione da parte dei regolamenti Europei, nonché il crescente interesse per gli output prodotti e potenzialmente disponibili da parte di utenti interni ed esterni all'Istituto ha portato, nel tempo, ad ampliare in misura considerevole l'insieme degli indi-

8 Si accede alla sezione Labour cost index a partire dal link <http://ec.europa.eu/eurostat/data/database> per accedere ai dati, e al link https://ec.europa.eu/eurostat/statistics-explained/index.php/Labour_cost_index_-_recent_trends.

9 Si veda al link: https://ec.europa.eu/eurostat/statistics-explained/index.php/Hourly_labour_costs.

catori prodotti. L'obiettivo è stato perseguito cercando di sfruttare le potenzialità dei dati disponibili nella fonte amministrativa di base descritta approfonditamente nel paragrafo 2.2.1 che segue e, là dove non possibili, fonti amministrative e statistiche ausiliarie già utilizzate o prodotte in Istituto come verrà presentato nel paragrafo 2.2.2.

2.2.1 La fonte amministrativa Inps

Come già anticipato nel paragrafo 1.3, la principale fonte amministrativa Inps che fu individuata per Oros è rappresentata dalle dichiarazioni contributive relative ai lavoratori dipendenti che i datori di lavoro devono presentare mensilmente all'Inps per denunciare le retribuzioni corrisposte e le informazioni utili al calcolo dei contributi.

Per sfruttare adeguatamente questa fonte di dati vengono parallelamente acquisite altre due fonti. In sintesi, ogni trimestre l'Istat ai fini di Oros acquisisce dall'Inps: 1) le dichiarazioni contributive mensili che i datori di lavoro devono effettuare entro l'ultimo giorno del mese successivo a quello di competenza; 2) l'anagrafica delle posizioni contributive attive alla data di scarico dei dati, contenenti anche informazioni sulle caratteristiche strutturali e particolarità contributive delle unità soggette alle dichiarazioni; 3) le ore autorizzate mensili di Cig e di solidarietà.

Attualmente i file mensili sulle dichiarazioni contributive a cui si fa riferimento sono i cosiddetti *DM2013 virtuali* (nel documento chiamati anche posizioni contributive o, più semplicemente, matricole aziendali dal loro identificativo)¹⁰. Si tratta di dichiarazioni ricostruite virtualmente dall'Inps a livello aziendale e per scopi amministrativi, a partire dai flussi delle dichiarazioni mensili contributive sui singoli lavoratori (UniEmens), allo scopo di fornire alle imprese estratti contributivi sintetici. La virtualizzazione viene effettuata attraverso procedure informatiche interne all'Inps che impiegano normalmente un paio di giorni e che hanno lo scopo di verificare anzitutto che i flussi UniEmens pervenuti rispettino vincoli aggiuntivi di quadratura oltre a quelli imposti all'immissione diretta. Eventuali dichiarazioni mensili "anomale" vengono separate e inviate a controlli particolari e, eventualmente, reintegrate nel flusso una volta rettificate da parte dell'Inps o delle aziende¹¹. I DM2013 virtuali acquisiti dall'Istat ogni trimestre riguardano:

a) un insieme *preliminare*, riferito al trimestre corrente t , di cui i primi due mesi sono scaricati a circa 36-38 giorni dalla fine del trimestre di riferimento e l'ultimo mese a 43-45 giorni. Tale modalità di acquisizione è resa necessaria dal diverso grado di riempimento dei tre mesi di competenza; infatti, considerato che la dichiarazione deve essere effettuata entro la fine del mese successivo a quello di competenza, il primo mese contiene le dichiarazioni arrivate nell'arco di circa 78 giorni dalla scadenza, il secondo 48 giorni, il terzo mese, più prossimo alla scadenza ufficiale, 18 giorni circa. Il diverso numero di dichiarazioni ritardatarie presente nei tre mesi del trimestre costituisce un problema peculiare che deve essere risolto attraverso procedure di trattamento di dati mancanti (cfr. capitolo 6). Questo insieme di dati è utilizzato per la stima preliminare del trimestre t ,

b) un insieme *finale*, riferito al trimestre $t-4$, scaricato con un ritardo di circa 1 anno dal trimestre di riferimento, utilizzato per calcolare le stime definitive di $t-4$. Tale insieme di dati è sostanzialmente una rappresentazione della popolazione totale, ossia non vi sono

¹⁰ Fino al 2010 la dichiarazione si chiamava DM10, poi si è passati al DM10 virtualizzato. Entrambe avevano caratteristiche diverse dal DM2013.

¹¹ Per una descrizione più approfondita del processo di generazione dei DM virtuali dagli Uniemens si veda: <https://www.inps.it/nuovoportaleinps/default.aspx?itemDir=43428>.

dichiarazioni mancanti. Fanno eccezione le dichiarazioni “anomale”, che non vengono virtualizzate in DM2013 per decisione dell’Inps, in quanto incomplete. Tale insieme, poco significativo dal punto di vista numerico, nel processo Oros viene tuttavia sottoposto a trattamenti specifici (cfr. §6.5).

Accanto ai file sulle dichiarazioni contributive viene, inoltre, acquisita dall’Inps l’*anagrafica* sulle singole posizioni contributive attive. Le informazioni anagrafiche devono essere fornite dall’azienda all’Inps entro il 16 del mese successivo a quello della sua attivazione o della variazione del suo stato di attività (sospensione, cessazione, ecc.). I dati sono scaricati a circa 30 giorni dalla fine del trimestre di riferimento *t*. Tale insieme di dati, contenente anche informazioni sull’inquadramento dell’azienda in termini contributivi (attività economica, codice statistico contributivo, codice autorizzazione per specifici sgravi, ecc.), viene utilizzato non solo per configurare l’unità nel campo di osservazione delle rilevazione, operazione che necessita dell’ausilio di altre fonti amministrative e statistiche, come verrà approfondito nel capitolo 3, ma anche per il calcolo di alcune componenti di costo del lavoro argomento trattato nel capitolo 3.

Infine, l’Istat acquisisce trimestralmente un set di microdati mensili sulle ore *autorizzate di Cig* per tipologia (ordinaria, straordinaria, deroga e solidarietà). Le ore a cui si riferisce l’insieme di dati sono alimentate, per quanto riguarda la cassa integrazione ordinaria industria ed edilizia dalle delibere di autorizzazione della commissione provinciale del lavoro, per la cassa integrazione straordinaria dai decreti emessi dagli uffici regionali del lavoro. E’ inoltre opportuno precisare che le ore autorizzate ogni mese non sono di competenza del mese stesso ma possono riferirsi sia a periodi precedenti il mese di autorizzazione sia a periodi successivi. Il file viene acquisito con un ritardo di circa 20 giorni dalla fine del trimestre di riferimento *t* e viene utilizzato per trattare statisticamente le Ula (cfr. §8.3) e le posizioni lavorative dipendenti (cfr. §8.4) per correggere gli effetti di distorsione che la presenza di Cig esercita sui principali indicatori rilasciati dalla rilevazione.

2.2.2 Necessità di integrare con fonti statistiche e altre fonti amministrative

Per lo sfruttamento della fonte principale, rappresentata dai dati amministrativi Inps, o per completare il quadro definitorio di alcune variabili sono necessarie altre fonti ausiliarie, sempre di fonte Inps o di altra provenienza, contenenti microdati, dati aggregati e metadati di natura legislativa o amministrativa per derivare le informazioni statistiche desiderate secondo le definizioni richieste dai diversi regolamenti. I micro dati grezzi che arrivano dall’Inps devono, anzitutto, essere transcodificati attraverso un minuzioso uso dei codici amministrativi presenti nei microdati, interpretati sulla base delle normativa vigente. Le variabili elementari ricavate dai dati grezzi devono essere successivamente aggregate o scorporate da altre variabili; infine occorre stimare e correggere variabili mancanti attraverso informazioni derivanti direttamente dall’Inps o da altre fonti (come ad esempio, Inail). Per poter effettuare questa operazione è stato necessario rintracciare, raccogliere, archiviare in una forma standardizzata e facilmente accessibile non soltanto i metadati relativi alla dichiarazione contributiva ma anche quelli indispensabili per la stima di alcune componenti del costo del lavoro non rilevate nel modello DM2013. A tale scopo è stata progettata e realizzata una banca dati normativa (BDN) per organizzare in modo sistematico e aggiornare trimestralmente i riferimenti normativi, i metodi e le procedure utilizzati (cfr. capitolo 4).

La disponibilità nei DM2013 del codice fiscale tra gli identificativi delle unità, consente di agganciare i dati Inps al Registro Asia e ad un archivio interno Istat sulle Partite Iva

ausiliario ad Asia, acquisiti annualmente, e all'Anagrafe tributaria dell'Agenzia delle Entrate acquisita trimestralmente, permettendo in tal modo di ottenere informazioni statisticamente validate o adeguatamente aggiornate sul codice di attività economica e sulla natura giuridica e istituzionale delle unità, utili per la loro collocazione nel campo di osservazione della rilevazione. Questa operazione verrà illustrata nel capitolo 2.

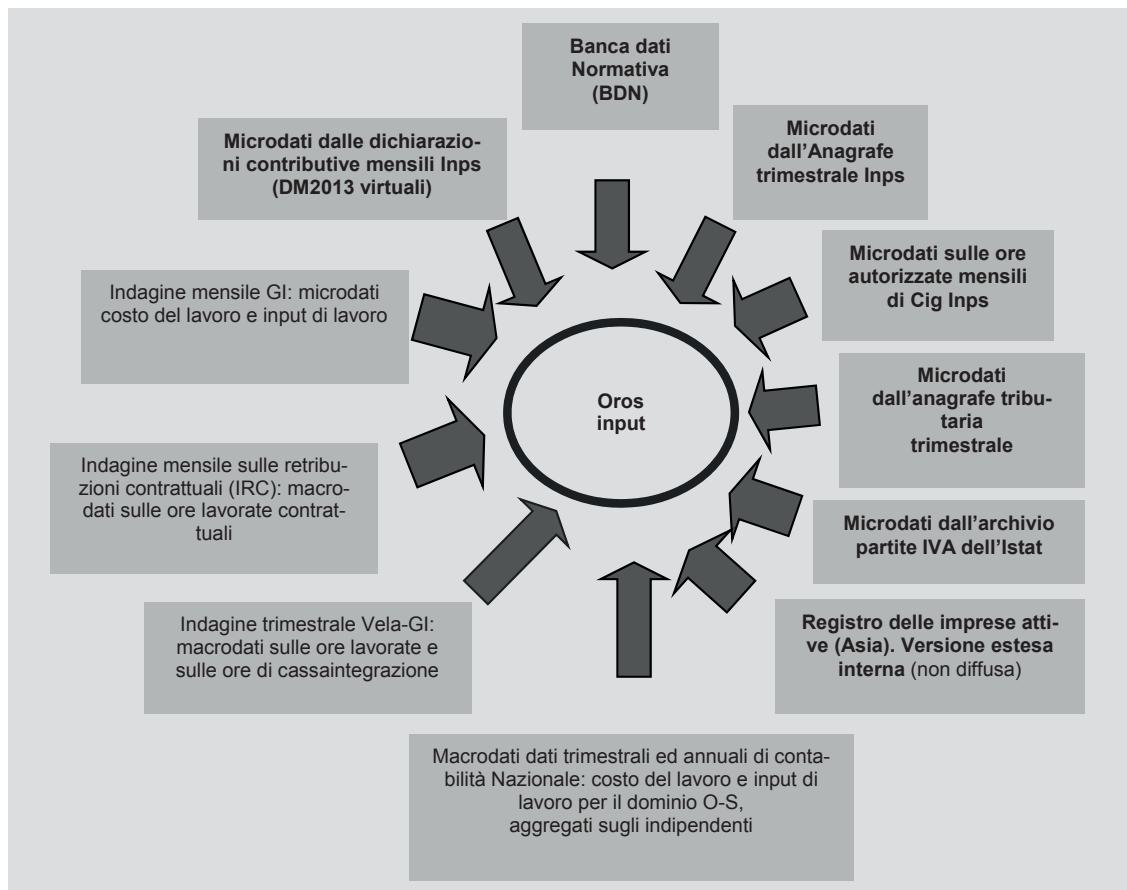
Al fine di conferire rilevanza statistica ad alcune variabili tratte dai dati amministrativi è necessario affiancare alla fonte amministrativa dati di natura statistica. In particolare, nella rilevazione Oros, alla stima dell'input di lavoro contribuiscono i macrodati dell'indagine sulle retribuzioni contrattuali sull'orario convenzionale di lavoro (cfr. §5.2, §7.4 e §8.3), nonché i macrodati tratti dall'indagine Vela-GI sulle ore fruite per Cig, utilizzate per correggere la distorsione delle ore autorizzate di fonte Inps per la stima delle Ula al netto della Cig (cfr. §8.3.7).

Si ricorre, invece, ai microdati d'indagine GI sulle unità di grandi dimensioni per la stima di tutte le variabili *target* relative alle imprese con oltre 500 dipendenti. La necessità di sostituire i dati amministrativi con i dati dell'indagine mensile GI è stata motivata in passato da problemi di copertura dei dati preliminari, fortemente incompleti sulle imprese di grandi dimensioni. A seguito della possibilità concessa alle imprese di frazionare la dichiarazione contributiva su più matricole aziendali, spesso si verificava che le dichiarazioni riferite ad un'impresa pervenissero solo con riferimento ad alcune delle matricole aziendali ad essa appartenenti. Questo fenomeno caratterizzava maggiormente le imprese di grandi dimensioni, data la complessità della loro gestione amministrativa. Ne conseguiva che il set preliminare di dati fosse incompleto sulle imprese di dimensioni più ampie spesso statisticamente autorappresentative su alcuni domini. La progressiva evoluzione della dimensione dei dati preliminari verso quella dei dati finali a metà degli anni 2000 ha radicalmente cambiato le condizioni iniziali (Grafico 2.3.2.1). Tuttavia, l'integrazione con l'indagine GI non solo è stata mantenuta ma si è nel tempo perfezionata, trovando nella migliore qualità del dato statistico il principale movente della sostituzione dei dati amministrativi con i dati statistici: il contatto diretto con i referenti delle imprese ha fornito opportunità rilevanti sia per gestire i continui e significativi cambiamenti giuridici, sia per disporre di dati di migliore qualità nel caso di risposte ritardatarie o incomplete. L'integrazione dei dati Inps con i dati dell'indagine GI, che coprono circa il 20% dell'occupazione complessiva nei settori dell'industria e dei servizi di mercato, avviene sia nella stima preliminare, sia in quella finale attraverso un minuzioso processo di individuazione di liste complementari di unità e ricostruzione dei dati economici, come verrà ampiamente argomentato nel capitolo 7.

Al fine di rispondere alle richieste dei regolamenti europei, Oros deve utilizzare fonti esterne al sistema. In particolare, la stima del numero di occupati totali per il regolamento STS, richiede la determinazione del numero trimestrale degli indipendenti. La metodologia prescelta si basa sullo sfruttamento di dati di fonte Contabilità Nazionale Annuale, utilizzando come indicatore di trimestralizzazione i dati sulle posizioni lavorative di Oros, argomento approfondito nel capitolo 12. Il regolamento LCI richiede, invece, indicatori di costo del lavoro espressi in termini di ore lavorate su tutti i settori di attività economica da B a S, incluso O. Per la compilazione di tale indicatore nei settori da B a N i dati Oros, utilizzati per la costruzione del numeratore, vengono affiancati dai dati sulle ore lavorate tratti dall'indagine Vela-GI. Per i settori da O a S, dovendo includere anche il settore pubblico, le stime dell'indice LCI vengono effettuate interamente utilizzando dati di Contabilità Nazionale. La descrizione della metodologia utilizzata per la costruzione dell'LCI è oggetto del capitolo 11.

La Figura 2.2.2.1 che segue riepiloga tutte le fonti utilizzate come input nel processo produttivo di Oros.

Figura 2.2.2.1 - Gli input di Oros



2.3 Stime

2.3.1 Stima preliminare e stima finale

Sulla base delle informazioni disponibili, ogni trimestre la rilevazione Oros rilascia una stima preliminare riferita al trimestre corrente t e una finale al trimestre $t-4$ già rilasciato come preliminare un anno prima. E' possibile rivedere anche la stima riferita ai trimestri intermedi tra $t-4$ e t qualora ciò si rendesse necessario. La stima finale differisce da quella preliminare a seguito della disponibilità di informazioni più complete ed aggiornate che si rendono disponibili nel frattempo:

- la disponibilità dell'universo delle dichiarazioni DM2013 virtuali per la produzione della stima definitiva;
- la revisione dei dati dell'indagine mensile GI;
- l'aggiornamento di informazioni di carattere strutturale sulle unità oggetto di rilevazione;
- le eventuali revisioni occasionali nella metodologia di stima degli indicatori.

I trimestri intermedi $t-1$, $t-2$ e $t-3$ possono essere influenzati, ma con impatto minore. Revisioni di maggiore entità caratterizzano normalmente la diffusione di giugno, in occasione della quale la rilevazione GI rivede i microdati relativi all'intero anno $a-1$, per incorporare

i dati pervenuti relativi ai rispondenti ritardatari in sostituzione dei dati precedentemente imputati (cfr. §7.2).

Ai fini della stima finale i dati amministrativi di fonte Inps, come accennato nel paragrafo 2.2.1, rappresentano la popolazione censuaria delle imprese non coperte dall'indagine GI, rendendo possibile l'applicazione di una metodologia estremamente semplificata basata sulla enumerazione dei dati disponibili¹². Stessa metodologia viene adottata anche per elaborare i dati censuari di fonte GI, sia nel caso delle stime finali sia di quelle preliminari. Le stime provvisorie per le imprese coperte con dati amministrativi richiedono, invece, particolari trattamenti che dipendono dalla struttura alla base dei dati amministrativi preliminari. I frequenti mutamenti amministrativi, legislativi e tecnici che si sono susseguiti nel tempo hanno radicalmente condizionato l'individuazione, la sperimentazione e, infine, la scelta dell'approccio di stima, implicando continue revisioni. Di seguito, viene illustrata l'evoluzione della situazione informativa dei dati provvisori e le scelte metodologiche che i vari cambiamenti hanno dettato, sino ad arrivare all'attuale metodologia di stima.

2.3.2 I cambiamenti della situazione informativa e il loro impatto sulla metodologia di stima preliminare

L'uso a regime di dati amministrativi per finalità statistiche comporta rischi legati soprattutto alla instabilità della fonte (Rapiti et al, 2010). Tali cambiamenti potrebbero non essere noti a priori all'ente utilizzatore che, di norma, ha scarso controllo su di essi.

Cambiamenti prevedibili possono essere legati a modifiche nella normativa che comportano variazioni nel contenuto del microdato o nel suo significato. Questa tipologia di cambiamenti può essere controllata attraverso il continuo studio della normativa specifica e il tempestivo aggiornamento dei metadati alla base della transcodifica dei dati amministrativi. Nella rilevazione Oros molti sforzi vengono prestati a questa fase, in assenza di un *framework* organico di metadati già collezionati o messi a disposizione dall'Inps.

Ben più rischiosi sono i cambiamenti che vanno ad influenzare la modalità di acquisizione, di conservazione e trasmissione dei dati, fasi che possono essere fortemente condizionate da eventi di tipo normativo, legislativo, tecnologico con conseguenze rilevanti sulla stabilità della metodologia di stima, se non adeguatamente programmati.

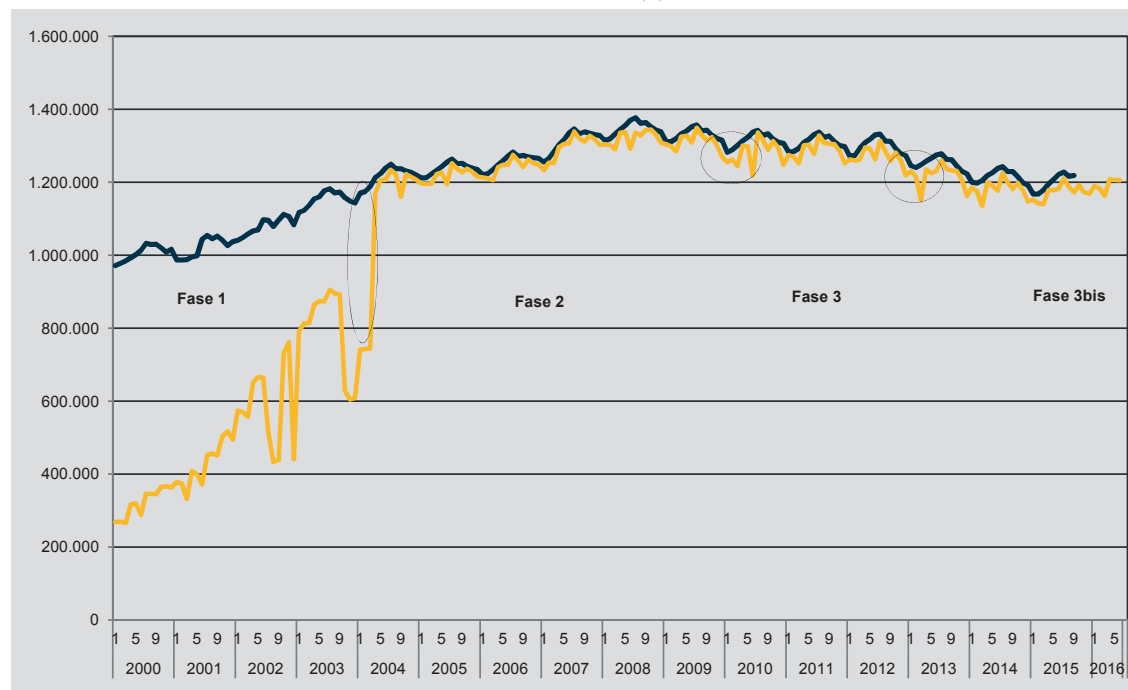
A partire dalla prima acquisizione dei dati a scopi sperimentali, alla fine degli anni '90, la fonte Inps sulle dichiarazioni contributive è stata interessata da diversi e radicali cambiamenti, che hanno avuto un impatto rilevante soprattutto sulla metodologia di stima provvisoria. Tali cambiamenti, infatti, hanno influenzato la disponibilità di microdati in tempi rapidi e, solo marginalmente, intaccato il set finale dei dati utilizzati per le stime definitive (Grafico 2.3.2.1).

La prima diffusione di dati è avvenuta a dicembre 2002, con il rilascio di serie storiche su macro aggregati. Successivamente ad aprile, luglio e ottobre 2003 la diffusione ha riguardato anche il livello di dettaglio per sezione Ateco 2007 e le prime stime provvisorie e, in particolare da ottobre, anche i primi dati destagionalizzati. La tempestività nel rilascio degli indicatori è andata gradualmente riducendosi, sino ad entrare a regime a partire da dicembre

¹² Una procedura di imputazione trimestrale di mancate risposte riguarda selezioni molto ridotte di "DM anomali/non generabili", ossia dichiarazioni contributive che pervengono all'Inps ma non sono virtualizzabili poiché caratterizzate da anomalie che non possono essere immediatamente sanate. L'Inps fornisce regolarmente all'Istat la lista di matricole caratterizzate da questa problematica distinte per mese di competenza, che consente di individuare le unità da imputare ai fini della stima delle posizioni lavorative dipendenti ricorrendo ad un metodo di stima basato su modelli di regressione (cfr. §6.5).

2. Obiettivi, fonti di dati e stime

Grafico 2.3.2.1 - Dichiarazioni contributive Inps nei dati preliminari e nei dati finali
 Mesi gennaio 2000-giugno 2016, valori assoluti(*)



Fonte: Rilevazione Oros (occupazione, retribuzioni, oneri sociali)

(*) Dichiarazioni riferite alle posizioni contributive nell'industria e servizi di mercato (settori da B a N Ateco 2007)

2003. Il più grande cambiamento risale all'inizio del 2004 quando l'invio dei DM da parte di tutte le imprese per via telematica da facoltativo è divenuto obbligatorio¹³. Tale cambiamento ha comportato il passaggio, in poco tempo, dalla disponibilità di un campione autoselezionato di unità alla disponibilità di un campione molto ampio di dati (di dimensioni prossime a quelle dell'universo) per le stime provvisorie, ovvero dal 50% circa di copertura delle unità della popolazione finale al 90-95% circa. Questo passaggio ha comportato una significativa semplificazione della procedure di stima preliminare. In particolare, prima di aprile 2004, la popolazione provvisoria era considerata un campione non-casuale, consistente in un insieme di rispondenti auto-selezionati (coloro, appunto, che adottavano su base volontaria la modalità elettronica di invio delle dichiarazioni contributive) (fase 1). Tale campione andava via via aumentando nel tempo e, in termini di ampiezza, rappresentava una buona copertura se riferita alle caratteristiche della popolazione di riferimento, incluso un adeguato numero di imprese nuove nate. In questo contesto informativo, in assenza di un *framework* di stima basato su disegno campionario, la stima preliminare avveniva mediante procedura di calibrazione sulla base di un approccio "model-assisted" (Baldi et al., 2004). Il problema della selettività del campione veniva risolto individuando i pesi all'interno di sottogruppi omogenei (*model groups*) per caratteristiche della popolazione (attività economica, dimensione, area geografica, età dell'impresa). La procedura di calibrazione forniva risultati accettabili nella stima delle variabili pro-capite, ma si rivelava scarsamente adeguata nella stima delle variabili di livello, quali le posizioni lavorative: infatti, il meccanismo non ignorabile caratterizzante i rispondenti ritardatari, i continui cambiamenti nella rappresentatività del campione, insieme

¹³ In precedenza i DM potevano essere compilati anche su supporto cartaceo e inviati per posta o consegnati manualmente presso le sedi di competenza, implicando tempi molto lunghi prima che fossero resi disponibili su supporto informatico (in questo periodo le dichiarazioni mensili prendevano il nome di DM10).

al problema della sovracopertura del registro amministrativo implicavano stime poco accurate di questa variabile, richiedendo azioni *ad hoc* supplementari sui macrodati aggregati per Ateco sulla base dell'analisi della regolarità in serie storica dell'indicatore di riferimento.

A partire da aprile 2004, rendendosi disponibile oltre il 90% di popolazione attiva per le stime preliminari, il metodo di stima preliminare è stato assimilato a quello delle stime finali, ossia per semplice somma dei dati disponibili (fase 2). Pur garantendo stime con un grado di distorsione accettabile per le variabili rapporto, secondo questo approccio rimaneva irrisolto il problema delle sottostime delle variabili di livello, le già richiamate posizioni lavorative, a causa delle dichiarazioni ritardatarie. Per la stima di tale variabile si continuava in questa fase ad utilizzare l'intervento sui macro dati. Nel corso del tempo, le correzioni basate sull'analisi delle serie storiche aggregate per Ateco è stata profondamente migliorata grazie allo sfruttamento di informazioni sull'errore di revisione registrato nelle varie occasioni di stima passate (che per alcuni aggregati mostrava una componente sistematica e predicibile) e studiando la relazione tra la dinamica dell'occupazione nei dati preliminari e finali (cfr. Istat, 2013a). Per un certo periodo di tempo la situazione informativa sembrava aver raggiunto un considerevole livello di stabilità, inducendo a sperimentare una metodologia di stima basata sulla ricostruzione dei microdati: la stabilità dei dati e la non persistenza delle mancate risposte, infatti, induceva a delineare il problema quale una questione di *wave-non response*, secondo cui il comportamento passato di ogni rispondente corrente è informativo rispetto alla sua realizzazione nella popolazione finale, ovvero l'informazione passata sui rispondenti è informativa sul dato corrente mancante.

Prima che il nuovo metodo di stima preliminare potesse essere messo a punto, a gennaio 2010 si è verificato un ulteriore grande cambiamento: il sistema di dichiarazione contributiva basato su modelli da compilare a livello d'impresa è stato sostituito dall'Inps con una nuova e più complessa modulistica (UniEmens) contenente informazioni a livello di lavoratore e alcune informazioni di sintesi a livello di impresa (fase 3). Il cambiamento nella struttura dei dati alla fonte ha comportato una modifica alla struttura dei dati acquisiti dall'Istat. L'adozione dell'UniEmens, che rappresenta una fonte di informazione ricchissima per l'Istituto anche a fini congiunturali, avrebbe richiesto un periodo di esplorazione molto lungo da impiegare principalmente nella ricostruzione dei metadati necessari per tradurre la nuova fonte in dato statistico. Per questo, ai fini di Oros, l'Istat ha potuto continuare ad acquisire i "DM10 virtuali", ovvero un'aggregazione degli UniEmens nel formato dei vecchi DM10 che l'Inps ha contestualmente iniziato a ricostruire per finalità amministrative interne. In questo contesto, sono state riavviate le vecchie sperimentazioni, che hanno dovuto tenere conto di un nuovo contesto informativo in cui, inizialmente, non solo era cresciuto il numero di rispondenti ritardatari, ma ne era mutata anche la caratterizzazione; infatti, erano le imprese più grandi a risultare più frequentemente incomplete a causa della maggiore difficoltà implicita nei nuovi modelli considerato il grande numero di dipendenti. Dal punto di vista della qualità dei microdati, invece, la nuova struttura ha mostrato sin da subito un netto miglioramento in termini di "errore di misura" dovuto in parte all'adozione di procedure più vincolanti introdotte dall'Inps nella compilazione dei nuovi modelli.

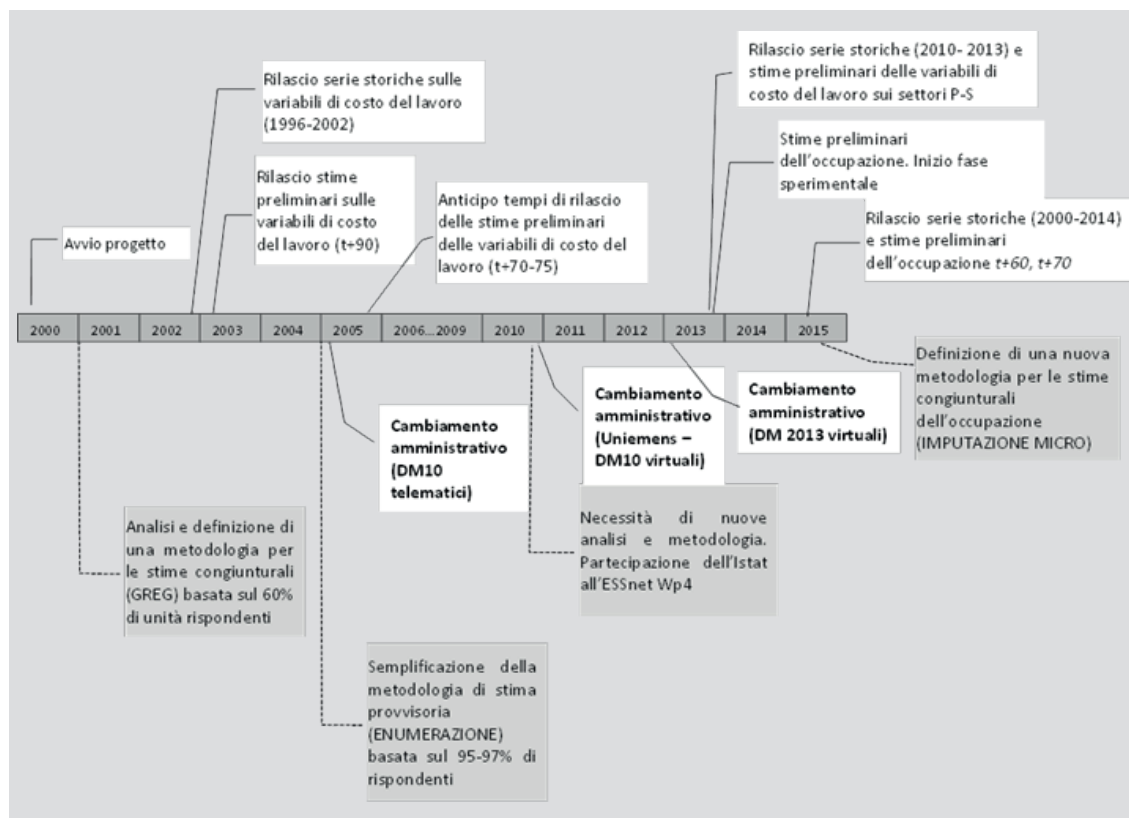
A partire da gennaio 2013 l'Inps ha introdotto un altro importante cambiamento nella ricostruzione del dato a livello d'impresa per usi amministrativi interni, che ha portato ad una struttura semplificata e più efficiente che ha dato origine ai "DM2013 virtuali" che rispetto ai precedenti DM10 virtuali sono caratterizzati dall'aggiunta di nuove variabili (fase 3bis). La ridotta sensibilità, da parte del fornitore dati, dell'impatto che tale modifica nella struttura dei dati avrebbe comportato per la rilevazione Oros ha implicato la ricerca di una soluzione rapida per adattare il processo di produzione e garantire l'uscita degli output congiunturali già program-

mati. Il team di Oros sta ancora lavorando alla semplificazione e riorganizzazione delle procedure per esplorare più efficacemente le informazioni disponibili nel nuovo DM2013 virtuale, per questo si attendono progressi continui nella produzione delle variabili d'interesse. Questa esperienza conferma che la dipendenza dei dati statistici da dati che nascono per finalità amministrative rende necessari un buon coordinamento e una collaborazione tra gli Enti coinvolti. Questa è la direzione in cui si sta lavorando per affrontare la transizione, che avverrà in tempi non brevi, all'uso delle informazioni a livello di lavoratore rese disponibili negli UniEmens.

L'esperienza Oros, che è stata la prima a livello nazionale ad unire lo sfruttamento del dato amministrativo ai fini congiunturali a quello del dato di indagine tradizionale, si protrae ormai da più di un decennio ma sin da subito ha costituito un caso di studio in ambito europeo sulla produzione di statistiche di breve termine con ridotto carico sulle imprese. A tal proposito, è utile ricordare la partecipazione di Oros al progetto ESSnet¹⁴ *“Use of Administrative Data and Accounts Data in Business Statistics” Working Package 4 - Timeliness of Administrative Sources for Monthly and Quarterly Estimate*, che è stata da stimolo all'avvio di una nuova fase di sperimentazione su un metodo di imputazione dei microdati finalizzato alla stima delle posizioni lavorative dipendenti, portando alla messa a regime e diffusione, a partire da giugno 2015, di dati sull'occupazione anche a livello nazionale.

La Figura 2.3.2.1 che segue illustra l'evoluzione della metodologia di stima degli indicatori di Oros in relazione alla mutata situazione informativa alla base, dall'avvio del progetto al rilascio delle prime serie storiche, all'ampliamento del campo di osservazione, sino ad arrivare alla più recente diffusione di dati sull'occupazione.

Figura 2.3.2.1 – Le tappe della rilevazione Oros: dall'avvio del progetto ai vari cambiamenti alla metodologia di stima alla base degli indicatori diffusi



14 Informazioni sugli ESSnet sono disponibili al link: https://ec.europa.eu/eurostat/cros/page/essnet_en.

2.4 Le macro fasi del processo di produzione Oros

Al fine di agevolare la lettura del presente manuale, nella Figura 2.4.1 che segue vengono sinteticamente illustrate le diverse fasi del processo di produzione della rilevazione Oros, dall'acquisizione dei dati e dei metadati, alla produzione dei principali output.

L'arrivo dei micro dati amministrativi viene preceduto da una fase di analisi dei più recenti aggiornamenti della normativa in materia di contribuzione obbligatoria, al fine di acquisire le informazioni necessarie per tradurre i dati amministrativi nelle informazioni statistiche alla base della compilazione ed interpretazione degli indicatori prodotti dalla rilevazione. Tali metadati, tratti principalmente da circolari e tabelle acquisite dai siti web Inps e Inail, contribuiscono ad alimentare ed aggiornare una banca dati normativa, attraverso cui è possibile attuare un trattamento preliminare dei microdati disaggregati, per poter ricavare le voci statistiche di base (fase 1).

Accanto alle fonti Inps vengono acquisiti microdati ausiliari, utili a completare le informazioni per la classificazione delle unità nel campo di osservazione della rilevazione. In particolare, si fa riferimento alla versione più aggiornata del registro Asia e dell'Archivio delle partite IVA dell'Istat e dell'Anagrafe tributaria dell'Agenzia delle entrate. Il matching tra le varie fonti avviene per codice fiscale, variabile sottoposta ad una procedura di controllo e ricostruzione nell'anagrafica Inps, finalizzata ad ampliare l'insieme dei dati abbinabili. Controlli di qualità ed operazioni di ricostruzione vengono estese anche ad altre informazioni disponibili nell'anagrafe Inps utili in varie fasi del processo Oros, quali le date degli eventi di nascita, cessazione e/o sospensione, la classificazione per attività economica delle unità amministrative etc. La disponibilità delle informazioni sulla caratterizzazione economica ed istituzionale provenienti dalle varie fonti, consente di collocare le unità nel campo di osservazione della rilevazione (fase 2).

Le informazioni anagrafiche attribuite alle unità amministrative vanno a costituire l'anagrafica Oros che viene aggiornata trimestralmente. I dati anagrafici vengono in seguito abbinati ai dati economici, resi trimestrali ed elaborati, anche mediante l'utilizzo di fonti di macrodati ausiliarie, per ottenere le principali variabili *target* utili al calcolo degli indicatori diffusi (fase 3).

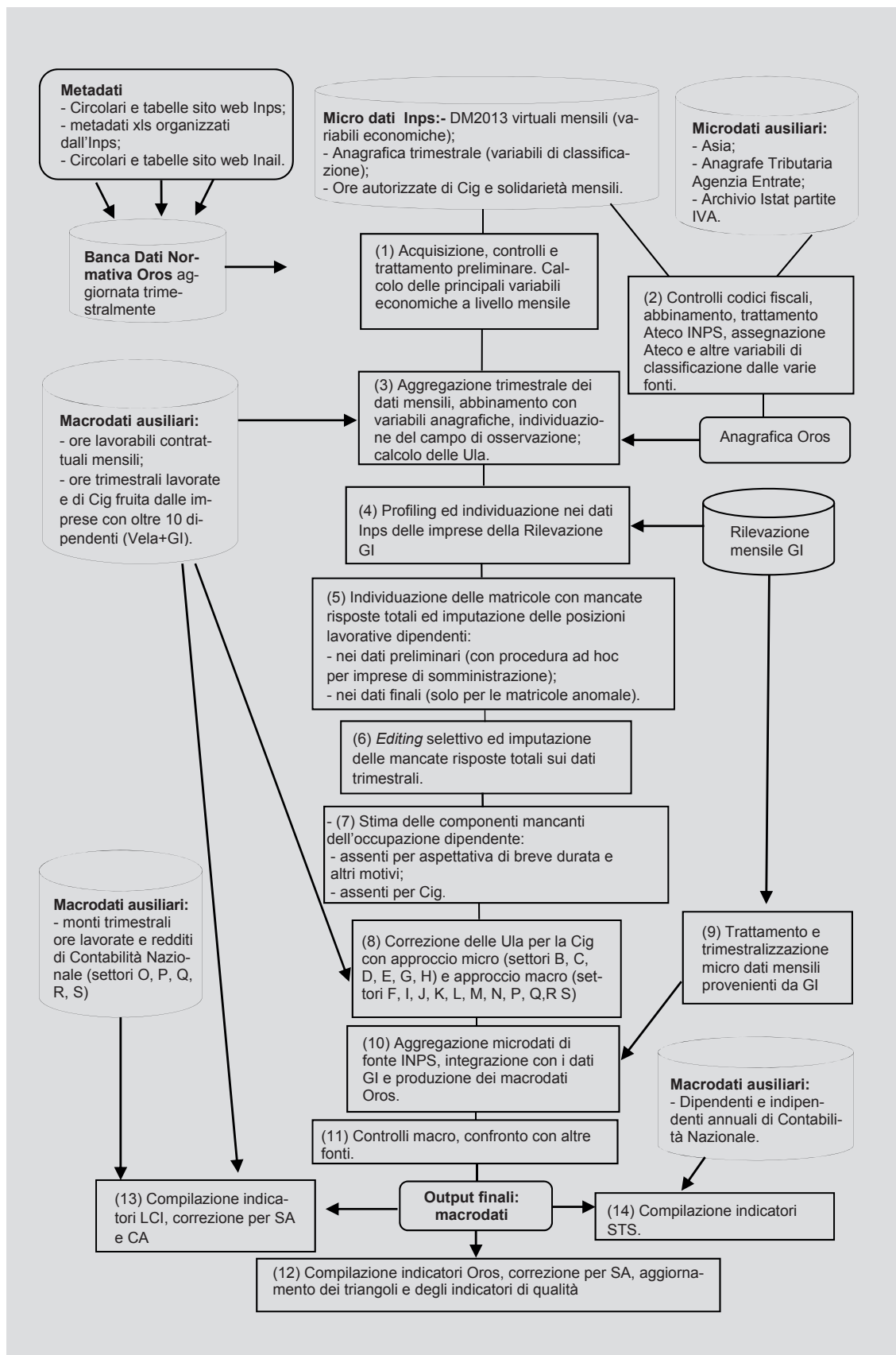
Sui microdati trimestrali, in seguito ad una complessa operazione di profiling delle imprese, vengono individuate ed estratte le unità amministrative che sono presenti anche nei dati della rilevazione mensile GI, utilizzata quale fonte principale per la stima delle variabili delle imprese con oltre 500 dipendenti (fase 4).

Sulle unità non appartenenti all'insieme delle GI vengono successivamente svolte le principali operazioni di *editing* ed imputazione (fase 5). Anzitutto, vengono individuate le unità caratterizzate da mancate risposte totali, secondo modalità differenziate nella versione preliminare o finale dei dati e, per alcune sottopopolazioni, adottando procedure *ad hoc*.

Applicando criteri di selettività, le unità non presenti nei dati GI, vengono successivamente sottoposte a controlli finalizzati a far emergere anomalie che influenzano le stime delle principali variabili. Le unità che risultano anomale vengono controllate ed eventualmente imputate (fase 6).

Prima di derivare la stima delle variabili *target* dai microdati amministrativi depurati da anomalie e mancate risposte, è necessario ricostruire alcune componenti mancanti dai dati di base. In particolare, la stima delle posizioni lavorative dipendenti richiede l'aggiunta delle posizioni assenti per cassa integrazione guadagni o in aspettativa di breve durata, non incluse nei dati finalizzati al calcolo del costo del lavoro. Tali componenti aggiuntive

Figura 2.4.1 – Il processo di produzione di Oros



vengono ricavate sfruttando alcune variabili presenti nelle dichiarazioni contributive, non utili per il calcolo del costo del lavoro, con il supporto dei dati Inps sulle ore autorizzate di Cig (fase 7). Un ulteriore trattamento è necessario per ottenere una stima delle Ula, poste al denominatore degli indicatori di Oros sul costo del lavoro, che sia depurata dal ridotto input di lavoro delle unità in Cig nel corso del mese. Tale trattamento è differenziato per sezione di attività economica e si basa sia sullo sfruttamento di informazioni sul tempo di lavoro, presenti nelle dichiarazioni mensili, e dei dati Inps sulle ore autorizzate di Cig (fase 8).

I micro e macro dati ottenuti in questa fase del processo di produzione vengono aggregati per settore di attività economica, dopo aver isolato le unità che sono presenti anche nei dati di fonte GI. Le stime su tali unità vengono sostituite con quelle ottenute utilizzando i dati di fonte GI, trattati secondo i criteri della rilevazione Oros (fasi 9 e 10).

Prima di diffondere gli indicatori prodotti, le stime delle principali variabili vengono validate anche ricorrendo a confronti con variabili *proxy* tratte da altre fonti (fase 11). Per la diffusione nazionale, segue un processo di correzione degli indici per la stagionalità e creazione dei formati di diffusione, incluso l'aggiornamento dei triangoli di revisione e calcolo degli indicatori di qualità (fase 12).

Per la compilazione degli indicatori richiesti dai regolamenti europei le stime tratte dai dati Oros devono essere integrate con dati tratti di fonti ausiliarie. In particolare, per la costruzione degli indicatori sul costo del lavoro orario vengono acquisiti macro dati dall'indagine Vela-GI e dalla Contabilità Nazionale trimestrale. Gli indici così predisposti vengono sottoposti a trattamento per la stagionalità e gli effetti di calendario (fase 13). Anche per la compilazione degli indicatori sul totale addetti per il regolamento STS richiede l'integrazione con fonti ausiliarie di macrodati. In particolare, per la stima degli indipendenti vengono tratti macrodati annuali dalla Contabilità Nazionale (fase 14).

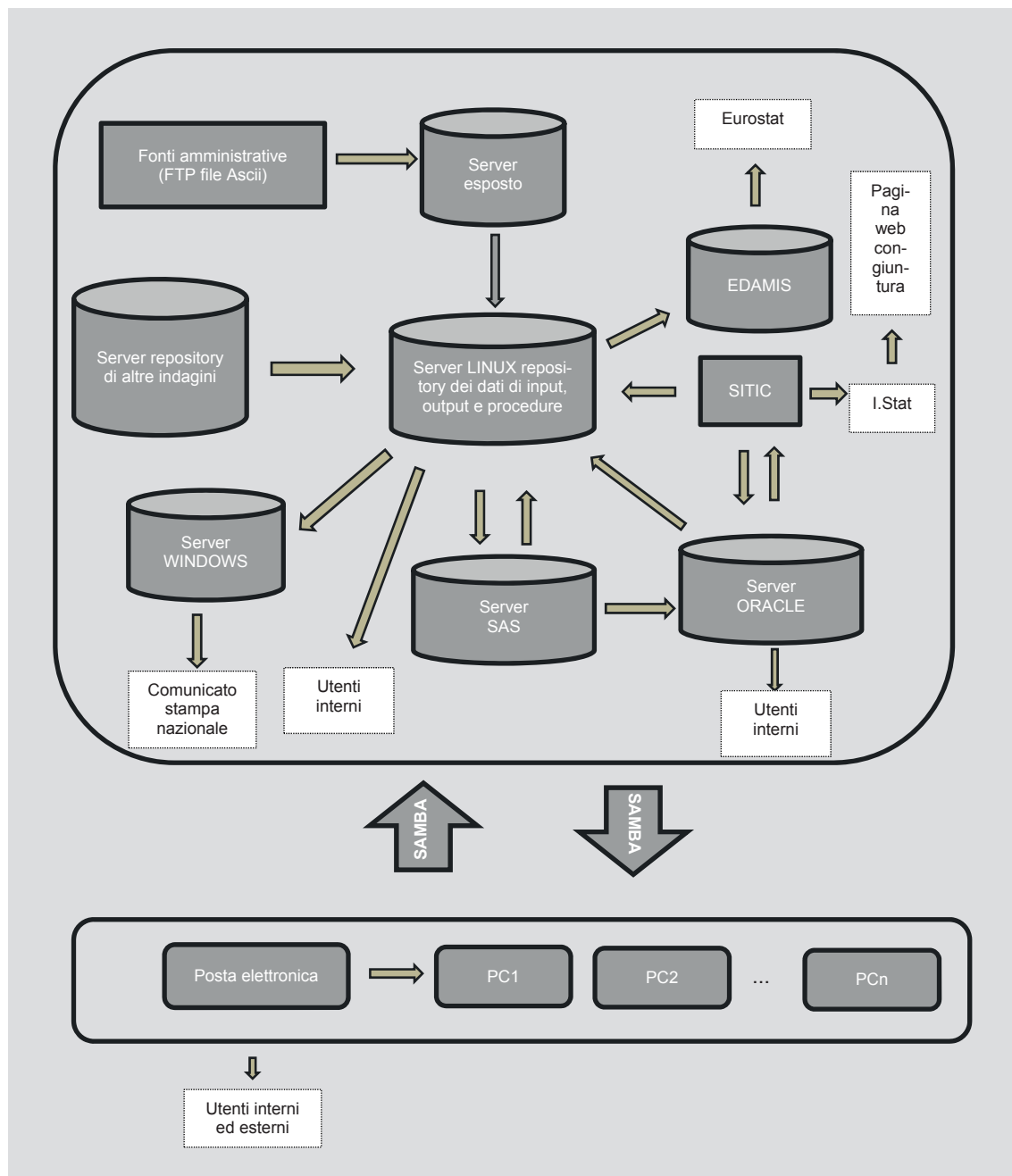
Alla base del processo di produzione di Oros vi è un'infrastruttura informatica estremamente variegata in termini di procedure e prodotti utilizzati, sviluppata nel corso del tempo a risposta dei continui mutamenti nelle basi dati, nella normativa vigente, nella disponibilità delle risorse informatiche dell'Istituto. Prima di entrare nella descrizione delle singole fasi del processo di produzione, il paragrafo successivo del manuale (§2.5) ripercorre i principali strumenti e le tecnologie alla base della piattaforma informatica della rilevazione.

2.5 L'infrastruttura informatica alla base di Oros: macro fasi dell'indagine e principali tecnologie utilizzate

L'infrastruttura informatica di Oros è nata e si è sviluppata per rispondere alle esigenze contingenti della rilevazione e alle peculiarità della base dati utilizzata. Fortemente condizionata dalla grande mole di microdati, da acquisire e trattare trimestralmente con ridotti tempi a disposizione, dalla varietà e dal continuo aggiornamento dei metadati utili alla loro transcodifica e, non da ultimo, dalle frequenti e inattese modifiche nella struttura di base dei dati amministrativi Inps, tale infrastruttura è stata disegnata e successivamente sviluppata principalmente in risposta ad esigenze di flessibilità d'implementazione, compatibili con gli standard architetturali presenti in Istituto.

L'acquisizione dei microdati dall'Inps e Sogei in formato Ascii avviene in via telematica attraverso un protocollo Secure FTP che assicura il trasferimento di dati criptati tra gli Enti secondo calendari prestabiliti. Essi vengono depositati in un Server esposto con l'esterno e da qui trasferiti tramite *script* temporizzati in un Server Linux su cui risiedono le aree di

Prospetto 2.5.1 – L'infrastruttura informatica di Oros: principali tecnologie utilizzate



lavoro dedicate all'indagine. SAS è il software principalmente usato per sviluppare le nuove procedure dell'indagine e per mantenere quelle esistenti, dall'acquisizione dei dati alla produzione degli output diffusi all'esterno e all'interno dell'Istat. Esso viene utilizzato dai Pc-Client, su cui è installato il software, connettendosi in remoto attraverso SAS-Connect ai Server Linux, su cui è a sua volta installato il software. I vantaggi di questa architettura sono lo sfruttamento delle potenzialità del processore Linux nell'elaborazione in SAS di procedure su grandi quantità di microdati che richiedono un'area temporanea di lavoro adeguatamente dimensionata (SASWORK) e la facilità di monitoraggio dei processi, nella visualizzazione di *log* e *output*, propria della versione SAS-Client. Il ricorso massiccio al SAS ha caratterizzato l'indagine sin dalla sua progettazione. Infatti, l'indagine è stata rea-

lizzata da un gruppo di lavoro costituito, fin dall'inizio, unicamente da statistici che hanno avuto come strumento di lavoro il SAS, particolarmente adatto per l'analisi, data la grande varietà di funzioni statistiche disponibili, e per l'integrazione di più fonti caratterizzate da una grande quantità di microdati. Questo aspetto ha costituito la forza e la debolezza del progetto in quanto gli obiettivi sono stati raggiunti dal gruppo con la messa a regime dell'indagine nei tempi previsti, ma la presenza nel progetto di esperti informatici, il cui coinvolgimento avrebbe ritardato i tempi iniziali di realizzazione, ma avrebbe altresì comportato nel lungo termine un incremento nell'efficienza di elaborazione e integrazione dati. Per questi motivi, SAS non è stato mai abbandonato e, tuttora, rappresenta il principale prodotto utilizzato nell'indagine per lo sviluppo di tutte le procedure, per l'interrogazione di DB ORACLE anche di altre indagini, ad esempio quelli della rilevazione GI utili all'integrazione, per la produzione di tabelle e grafici utili ai Comunicati stampa. La struttura attuale dell'indagine, in un contesto di crescente integrazione tra fonti amministrative che presentano un livello di dettaglio elevato (dati individuali in sostituzione di dati d'impresa), non è da considerarsi quella ottimale. Andrebbe riprogettata come struttura di tipo DB, basata ad esempio su procedure e tabelle relazionali ORACLE, allo scopo di alleggerire la gestione dei dati intermedi, relegando l'utilizzo di SAS solo alle analisi statistiche ausiliarie. Ad oggi, solo una delle principali fasi del processo Oros meno soggetta a cambiamenti, rappresentata dalla fase preliminare di costruzione e mantenimento delle informazioni anagrafiche, è stata sviluppata in ambiente ORACLE grazie all'unico investimento informatico fatto in passato su questa attività (cfr. capitolo 3).

Il *core* dell'architettura dell'indagine si completa con l'interazione degli output della rilevazione Oros con altri sistemi informativi presenti in Istat, tra cui il principale è rappresentato dal sistema SITIC creato per uniformare il trattamento e la conservazione degli indicatori congiunturali finalizzati ai comunicati stampa nazionali, alle diffusioni via I.Stat e alle richieste dei regolamenti europei (cfr. capitolo 8). SITIC, basato su tabelle relazionali e procedure in PL-SQL viene gestito dai tecnici dell'indagine attraverso delle maschere interattive PHP che consentono di attivare le varie funzionalità, dall'acquisizione dei dati di input al rilascio degli output. SITIC consente infatti di produrre gli indicatori nel formato di diffusione richiesto da I.Stat e/o nel formato GESMES/SDMX secondo standard concordati con Eurostat. L'aggiornamento dei dati su I.Stat consente a sua volta, attraverso una piattaforma di tipo Micro Strategy gestita all'esterno della rilevazione, l'aggiornamento dei triangoli e l'analisi delle revisioni sulla pagina web sulla congiuntura con l'ultima versione dei dati rilasciati (cfr. capitolo 10). La creazione del formato SDMX segue una procedura complessa che prevede una fase di definizione delle relazioni ("mappatura") tra le variabili presenti su SITIC e i formati SDMX idonei per quelle variabili, a cura degli informatici in collaborazione con i referenti d'indagine, attraverso il software MappingAssistant messo a disposizione da Eurostat, che risiede su server Windows dell'Istat e fornisce il codice a SITIC. Questo codice consente di produrre il file di dati nel formato richiesto da Eurostat, mentre la trasmissione dei dati ad Eurostat avviene attraverso Edamis, un applicativo web di Eurostat, che consente attraverso un concetto SEP (Single Entry Point) di trasmettere i dati in via totalmente automatica. L'aggiornamento e la diffusione degli indicatori europei è invece a cura di Eurostat.

In SITIC viene, inoltre, effettuato il trattamento delle serie storiche per gli effetti di calendario e la stagionalità utilizzando il software TRAMO-SEATS, un applicativo statistico attuale standard in Istituto. Oltre a depositare ed acquisire dati da SITIC via ORACLE, la rilevazione Oros, per motivi di coerenza interna dei dati, fa dialogare il server Linux in cui

sono presenti i principali output di rilevazione direttamente con le tabelle DB ORACLE di SITIC attraverso il SAS.

Tutta l'architettura viene gestita dai tecnici dell'indagine facendo comunicare, attraverso l'applicativo SAMBA, i propri Pc-client con i server Linux, dove risiedono dati e procedure per motivi di sicurezza (back-up continui) e di spazio disponibile (Giga byte), avendo preventivamente condiviso le aree di lavoro con questo strumento che simula l'ambiente Pc. La gestione e la conservazione dei dati di input (file Ascii) e degli elaborati intermedi (*dataset* SAS) segue una precisa politica di archiviazione: output accessori non più necessari o facilmente replicabili vengono periodicamente eliminati oppure archiviati su spazi appositamente dedicati, mentre quelli di uso corrente vengono lasciati in uso nelle specifiche aree dell'indagine. Qualsiasi tipo di trasferimento e manipolazione manuale dei file avviene attraverso protocollo SSH (Secure SHell), che prevede la gestione dei file in modalità sicura.

SAMBA viene usato anche per comunicare con i server Windows, dove avviene la stesura del comunicato stampa.

Altri strumenti, in ambiente locale, che si affiancano a quelli principali e più specifici per alcune analisi statistiche sono l'*open source* R e il pacchetto J-Demetra+per la destagionalizzazione delle serie storiche. Infine la posta elettronica rappresenta la modalità usata per comunicare ed inviare dati, perlopiù aggregati, con gli utenti esterni non essendo possibile farli accedere con viste su tabelle ORACLE o condivisione di *dataset* SAS come avviene con gli utenti interni.

3. LE INFORMAZIONI ANAGRAFICHE¹

3.1 Introduzione

Il campo di osservazione della rilevazione Oros, come già anticipato, è definito dalle imprese ed istituzioni private con dipendenti, di tutte le classi dimensionali, che svolgono la loro attività economica nei settori delle sezioni da B ad N e da P ad S della classificazione delle attività economiche Ateco 2007 (versione italiana della nomenclatura europea NACE Rev. 2).

La definizione della popolazione teorica di riferimento della rilevazione Oros corrisponde a quella del Registro statistico delle imprese attive (Asia) predisposto dall'Istat². Diffuso con 15, 24 mesi di ritardo dalla fine dei trimestri della stima provvisoria, il registro Asia, se utilizzato quale fonte primaria per l'individuazione della lista di riferimento per la popolazione Oros implicherebbe una rilevante sottocopertura di unità neo-nate nell'intervallo temporale e una sovracopertura di unità cessate. Poiché la demografia d'impresa ha un ruolo non trascurabile nella determinazione della dinamica delle variabili *target* di Oros, sin da principio si è ritenuto fondamentale dover sfruttare le potenzialità delle fonti Inps (cfr. Baldi et. al 2001, 2004, 2008), praticamente censuarie e disponibili in tempi molto rapidi ma caratterizzate da rilevanti limiti di validità statistica, sormontabili attraverso particolari trattamenti dei dati amministrativi e la loro integrazione con altre fonti amministrative e statistiche. In tal senso, ad Asia viene dato un ruolo di guida, rappresentando la principale fonte ausiliaria da cui attingere le informazioni statistiche rilevanti di classificazione.

La fonte amministrativa Inps utile per l'individuazione della lista di unità attive per le stime Oros è l'anagrafica delle posizioni contributive (d'ora in poi Anagrafica Inps), un file di microdati aggiornato tempestivamente, che contiene informazioni strutturali e relative all'inquadramento contributivo delle aziende che afferiscono al Fondo di Previdenza dei Lavoratori Dipendenti. Tale insieme di dati viene acquisito ai fini di Oros trimestralmente, con aggiornamento alla fine del trimestre *t* di riferimento della stima provvisoria. Mentre, da un lato, esso fornisce un patrimonio informativo fondamentale sia per la corretta interpretazione e l'utilizzo dei dati amministrativi contenuti nei DM, sia per la definizione di una lista molto aggiornata di imprese potenzialmente attive, dall'altro le informazioni utili alla collocazione delle unità *target* rispetto al campo di osservazione, necessitano di trattamenti di *editing* e correzione ed integrazione con altre fonti. Come verrà descritto nel dettaglio dei paragrafi che seguono, l'uso di diverse fonti da cui estrarre tali informazioni richiede la definizione di criteri di "selezione" per Ateco, per natura istituzionale e per forma giuridica. La definizione dei criteri e, di conseguenza, le fonti utilizzate per la predisposizione di un'anagrafica per Oros, sono stati oggetto di profonde revisioni nel corso del tempo anche a seguito delle crescenti richieste di statistiche sul mercato del lavoro. Ne è un esempio il

¹ Il paragrafo 3.1 di questo capitolo è a cura di Elisabetta Aquilini e Marco Lattanzio; i paragrafi 3.2.1, 3.2.2 e 3.2.3 sono a cura di Marco Lattanzio; il paragrafo 3.2.4 è a cura di Elisabetta Aquilini, Eleonora Cimino e Marco Lattanzio; il paragrafo 3.2.5 è a cura di Eleonora Cimino, Marco Lattanzio e Donatella Tuzi.

² Per informazioni più dettagliate su contenuti e metodologie del registro Asia si veda Istat, 2016b.

recente allargamento del campo di osservazione della rilevazione ai settori dei servizi personali e sociali (sezioni da P a S) (cfr. §2.3.2) che ha comportato, come verrà descritto più dettagliatamente in questo capitolo, una vera e propria riprogettazione di questa fase del processo di produzione di Oros.

In questo capitolo vengono descritte le procedure seguite per la produzione dell'anagrafica alla base delle stime di Oros, una lista completa ed aggiornata di unità affiancate da variabili strutturali statisticamente validate per la loro collocazione nel campo di osservazione.

3.2 Le fonti dati e la definizione dei criteri per la collocazione delle unità nel campo di osservazione

In questo paragrafo sono descritte le fonti amministrative e statistiche utilizzate per costruire la base dati anagrafica della rilevazione Oros, i relativi trattamenti effettuati per una corretta integrazione e, infine, le variabili e il loro utilizzo per definire i tre criteri di esclusione citati poco sopra. Le fonti a cui si fa riferimento sono l'Anagrafica Inps, l'archivio Asia, l'Anagrafe Tributaria dell'Agenzia delle Entrate, l'archivio Istat delle partite IVA.

3.2.1 L'Anagrafica Inps

Come anticipato, la principale fonte per la costruzione del registro anagrafico Oros è l'Anagrafica Inps, fonte di natura amministrativa acquisita a cadenza trimestrale (cfr. §2.2.1) che contiene informazioni anagrafiche relative alle imprese iscritte al Fondo di Previdenza dei Lavoratori Dipendenti. L'archivio, acquisito a circa 30 giorni dalla fine del trimestre di riferimento, consta di circa 4,5 milioni di *record* e include le unità che, secondo le variabili anagrafiche Inps, non risultano cessate nel trimestre oppure sono cessate e/o sospese a partire dal 2000. Tale taglio sulle unità non attive, concordato con l'Inps ai fini di Oros, comporta l'inclusione di molte unità non più attive, ma si è reso necessario a scopi cautelativi a seguito della constatazione che unità dichiarate non attive, dopo un periodo di tempo anche lungo, tornano ad inviare dichiarazioni contributive. Escluderle a priori avrebbe comportato una sottocopertura. Data la grande quantità di dati ricevuti trimestralmente, le operazioni di caricamento e gestione di questa fonte vengono effettuate in modo da ottimizzare gli spazi necessari al mantenimento e le procedure per le elaborazioni, nell'ottica di garantire l'efficienza nella gestione della base dati. In particolare, i file Ascii ricevuti vengono acquisiti e conservati secondo due diverse modalità: una prima, adottata per le variabili per le quali non si ritiene necessario mantenere memoria storica, che prevede la conservazione dell'ultima versione ricevuta che viene sovrascritta alla precedente; una seconda in cui si tiene memoria degli aggiornamenti trimestrali relativamente a un sottoinsieme specifico di variabili. Questo trattamento consente, da un lato, di monitorare gli aggiornamenti di alcune variabili amministrative utili ai fini statistici e, dall'altro, di poter riprodurre in qualsiasi momento la situazione informativa relativa ad istanti temporali precedenti. Si tenga presente, inoltre, che dopo l'abbinamento con i dati economici, operazione che decreta lo *status* di attive per le unità ricevute in anagrafica, molte di queste, non più attive da lungo termine, vengono escluse dai microdati attraverso l'applicazione di criteri *ad hoc* (cfr. §6.2.4 e §6.3).

L'unità di riferimento dell'Anagrafica Inps è la posizione contributiva, una entità di natura amministrativa che può corrispondere ad una unica impresa o a una parte di essa,

frazionata secondo criteri puramente contabili³. La chiave identificativa della posizione contributiva è una variabile a caratteri numerici di 10 cifre definita matricola aziendale, che consente di legare univocamente le informazioni anagrafiche alle informazioni economiche contenute nei DM (cfr. §4).

L'aggiornamento delle informazioni dell'Anagrafica Inps è continuo ed è in parte a carico dell'Inps stesso, in parte a carico delle imprese. In particolare, le imprese sono tenute a fornire informazioni su attivazione, variazione del tipo di attività e dello stato di attività (sospensioni, cessazioni ecc.) entro il giorno 16 del mese successivo a quello in cui l'evento è avvenuto. L'Inps aggiorna, invece, informazioni sulla caratterizzazione contributiva delle singole unità, spesso attribuite dall'Istituto stesso in seguito all'emanazione di disposizioni legislative sul mercato del lavoro e alla richiesta, da parte delle imprese che presentano le caratteristiche di legge, di potervi aderire. In riferimento agli eventi di attivazione, cessazione o sospensione dell'attività, non sono previste sanzioni dall'Istituto per soggetti inadempienti o che effettuano le comunicazioni in ritardo, causando una rilevante sovracopertura di unità inattive all'interno della lista di imprese individuate dal file di anagrafica. Tale caratteristica rappresenta uno degli inconvenienti più rilevanti di tale fonte di dati e la sua risoluzione richiede l'adozione di soluzioni metodologiche *ad hoc* basate su approfondite analisi relative alla struttura dei dati sottostanti (cfr. §6).

Una delle potenzialità più importanti di tale archivio è la sua completezza rispetto alle unità attive al trimestre di stima provvisoria di Oros, mettendo a disposizione informazioni anagrafiche e strutturali tempestive per le stime correnti.

Tra le varie informazioni presenti nell'Anagrafica Inps, vi sono alcune variabili utili ad effettuare abbinamenti con altre fonti statistiche e amministrative, come il codice fiscale dell'impresa. In questa fase del processo, il codice fiscale viene sottoposto ad operazioni di controllo e correzione, finalizzate a migliorarne la qualità e massimizzare gli abbinamenti con le fonti ausiliarie. In particolare, ne viene controllata la correttezza e la congruenza attraverso degli algoritmi appositi e, nel caso di risposta negativa al controllo, vengono effettuati dei tentativi di recupero principalmente attraverso due processi di ricostruzione: inizialmente, recuperando il codice fiscale corretto all'interno di posizioni contributive a cui afferisce la matricola in questione, dette posizioni contributive di riferimento, in alternativa si utilizza l'archivio Istat delle partite IVA per recuperare il codice fiscale nel caso in cui all'interno della variabile dell'Anagrafica sia erroneamente indicata la partita IVA dell'impresa in sostituzione del codice fiscale. Nel corso del tempo la qualità di questa variabile è andata progressivamente migliorando e la quota di codici fiscali formalmente corretti in origine è ormai quasi totalitaria. Ai codici fiscali viene attribuito un *flag* che identifica sia la correttezza o meno del valore indicato, sia la modalità di ricostruzione effettuata⁴. Gli unici codici fiscali utilizzati per le finalità di integrazione con altre fonti sono quelli che vengono indicati come corretti, anche se a seguito di recupero dalla procedura di trattamento. Negli altri casi, l'abbinamento non viene effettuato.

3 Non vi è, infatti, una regola generalizzata dettata dall'Inps per l'apertura di posizioni per la dichiarazione dei contributi rispetto all'aggregazione dei dipendenti per i quali si effettuano dichiarazioni contributive. La nascita di nuove posizioni e/o la loro cessazione o temporanea sospensione risponde piuttosto a necessità amministrative delle singole imprese (criteri territoriali, piuttosto che legati alla qualifica/inquadramento dei lavoratori ecc.).

4 Nel file di Anagrafica Inps è stata, inoltre, riscontrata la presenza di unità con codice fiscale formalmente corretto, ma ripetutamente assegnato d'ufficio dagli operatori Inps ad unità con identificativo d'impresa non comunicato. Per evitare che le informazioni assegnate mediante codice fiscale risultino scorrettamente utilizzate, è stata creata una lista di questi codici fiscali "anomali", periodicamente controllata ed eventualmente integrata. Tali unità sono in numero molto ridotto.

Tra le variabili rilevanti nell'anagrafica vi sono, inoltre, un codice di attività economica, che corrisponde all'Ateco 1991⁵ attribuito direttamente dall'Inps attraverso la transcodifica del CSC (cfr. Baldi et al., 2001, Istat, 2013a), il codice statistico contributivo, una variabile di contenuto amministrativo che ha lo scopo di identificare la matricola aziendale rispetto all'attività economica prevalente. In questa fase di processo il codice Ateco 1991 viene transcodificato in Ateco 2002 e successivamente in Ateco 2007 attraverso tabelle appositamente costruite dall'Istat (cfr. Istat, 2013a).

Una ulteriore informazione molto utile al processo Oros è il cosiddetto "codice di autorizzazione", un codice a lunghezza variabile, le cui componenti indicano l'autorizzazione della matricola ad usufruire di particolarità contributive.

Infine, sono disponibili le date di costituzione e cessazione/sospensione delle unità. Tali informazioni in Oros vengono utilizzate per definire lo status di attività delle unità in stima provvisoria. Sottoposte ad un trattamento di verifica di congruenza e, in seguito, soggette a correzioni, tali variabili si rivelano di buona qualità ma, come accennato sopra, possono risultare non aggiornate. I problemi di aggiornamento riguardano, in particolare, la registrazione delle cessazioni o sospensioni, spesso non presenti (risultano attive unità che in realtà sono cessate/sospese), ma anche di riattivazioni (risultano inattive unità che in realtà hanno ripreso la loro attività).

3.2.2 L'archivio Asia e l'archivio delle partite IVA di fonte Istat

La principale fonte statistica ufficiale da cui Oros attinge informazioni strutturali è il registro Asia imprese, versione *frozen*⁶, contenente la lista delle imprese attive all'anno t , rilasciata solitamente a $t+15$ mesi. A seguito dell'estensione del campo di osservazione ai servizi privati sociali e personali (sezioni da P ad S della classificazione Ateco 2007), con l'obiettivo di abbinare un numero maggiore di informazioni validate statisticamente ad un insieme più ampio possibile di unità della rilevazione, si è ritenuto utile fare riferimento alla versione estesa del registro (d'ora in poi Asia esteso), la quale rappresenta la base amministrativa contenente informazioni su tutte le unità giuridiche operanti sul territorio italiano, costruita tramite integrazione di varie fonti amministrative. A partire da questa base, l'archivio Asia imprese viene costruito rispettando il campo di osservazione relativo al regolamento SBS⁷, selezionando le imprese attive almeno una settimana nell'anno di riferimento, escludendo le unità classificate nelle sezioni A (agricoltura), O (amministrazioni pubbliche), T (attività di famiglie e convivenze), U (organismi extraterritoriali), le unità della divisione S-94 (attività di organizzazioni associative) e le unità classificate come istituzioni pubbliche e istituzioni private *non profit*, a prescindere dall'attività svolta⁸. La versione Estesa dell'archivio Asia consta di circa 12 milioni di record, a differenza dei 4 milioni e 500 mila unità della versione ridotta. Dal registro vengono utilizzate le variabili Ateco 2007, forma

⁵ Versione della classificazione ufficiale delle attività economiche precedente l'Ateco 2002.

⁶ Si definisce *frozen* o "congelata" la versione del registro contenente la lista delle imprese attive al tempo t , da considerarsi come una fotografia del registro scattata in un istante temporale; essa è solitamente diffusa alla fine del primo trimestre di ogni anno solare e costituisce l'universo delle imprese attive per l'anno t . Rimanendo invariata nell'anno t essa rappresenta, quindi, la popolazione di riferimento per tutte le indagini sulle imprese che hanno tale anno come periodo di riferimento. Le correzioni e gli aggiornamenti dei dati sono inseriti nel corso dell'anno nella versione *live* o *running*, ed entrano a far parte della nuova struttura delle imprese italiane a disposizione degli utilizzatori nella edizione successiva del file annuale *frozen* (Istat, 2016b).

⁷ Regolamento Ue n.295/2008 per le statistiche strutturali (SBS – Structural Business Statistics).

⁸ Si faccia riferimento al capitolo 2 di Istat 2016.

giuridica e un *flag* che identifica se l'unità è o meno classificata come istituzione pubblica (ovvero appartenente alla lista S13).

L'abbinamento tra l'archivio Inps e Asia avviene per codice fiscale e riguarda il 90% circa delle unità finalizzate alle stime provvisorie. A rimanerne fuori, una quota poco rilevante di unità per mancato abbinamento della chiave identificativa e una quota più rilevante di unità nate o riattivate tra l'anno di aggiornamento di Asia esteso e il trimestre di stima provvisoria. Per poter conferire a tali unità residuali informazioni di classificazione, si effettua un successivo abbinamento con l'Anagrafe Tributaria dell'Agenzia delle Entrate, una fonte amministrativa molto aggiornata, acquisita dall'Istat anche per altri scopi statistici. Questa fonte ausiliaria di dati viene descritta nel paragrafo che segue.

Una seconda fonte Istat, inserita nel processo Oros tra le innovazioni più recenti, è l'archivio delle partite IVA. L'uso di tale archivio, a regime da giugno 2013, ha lo scopo di migliorare la qualità del codice fiscale presente nell'Anagrafica Inps insieme alle tecniche già descritte nel paragrafo precedente.

3.2.3 L'Anagrafe Tributaria di fonte Agenzia delle Entrate

Come descritto in precedenza, il disallineamento temporale tra l'archivio Asia e l'Anagrafica Inps impedisce l'acquisizione di informazioni statistiche strutturali, validate ufficialmente, per un numero rilevante di unità neo-nate rispetto all'istante di riferimento dell'archivio. Poiché le informazioni di fonte Inps sono inadeguate a colmare tale *gap*, è stato necessario individuare una fonte ausiliaria da cui trarre le informazioni mancanti. Già utilizzata in Istituto in diversi processi produttivi e caratterizzata da tempestività nella disponibilità di informazioni aggiornate e complete, seppure di natura amministrativa, l'Anagrafe Tributaria di fonte Agenzia delle Entrate è stata prescelta quale fonte aggiuntiva di abbinamento. La fonte viene acquisita dall'Istat trimestralmente e contiene informazioni aggiornate a 20-30 giorni dall'ultimo giorno del trimestre di riferimento delle stime provvisorie di Oros. Dall'Anagrafe Tributaria vengono acquisiti l'Ateco 2007, variabile assegnata dalle Camere di Commercio in occasione dell'iscrizione e/o variazione dell'attività prevalente delle unità, e la forma giuridica. L'archivio consta di circa 12.000.000 di unità e contiene informazioni di tipo strutturale relative a tutte le unità giuridiche che sono soggette ad un regime IVA. Anche in questo caso, la variabile identificativa delle unità è il codice fiscale. Partendo da una quota pari al 10% di unità che non si abbinano con l'archivio Asia, l'abbinamento con l'anagrafe tributaria permette di recuperarne circa un 9,6%. Per il resto delle unità, pari a circa lo 0,4%, si dispone di sole informazioni di natura amministrativa Inps.

3.2.4 L'integrazione delle fonti e la definizione dei criteri per la collocazione delle unità nel campo di osservazione

Il processo di integrazione per la costruzione dell'anagrafica Oros avviene una volta acquisite le informazioni provenienti dalle varie fonti. Il primo passo consiste nell'abbinamento con la fonte partite IVA al fine di migliorare la qualità dei codici fiscali dell'Anagrafica Inps e favorire, quindi, i successivi *match* con le altre fonti utilizzate nel corso dell'intero processo della rilevazione. Successivamente, l'integrazione con le altre fonti è finalizzato a trarre informazioni utili per identificare la natura istituzionale delle unità selezionate nella lista tratta dall'Anagrafica Inps e di attribuire loro un codice Ateco validato statisticamente.

In questa fase di processo, a ciascuna unità, sono attribuiti i così detti *flag* di esclusione determinati sulla base di precisi criteri gerarchici che sfruttano in modo integrato le variabili presenti nelle fonti descritte. In generale, l'informazione proveniente dall'archivio Asia, anche se meno aggiornata rispetto alle altre, ha la precedenza in quanto informazione statisticamente validata. Le unità che presentano le caratteristiche di esclusione vengono identificate attraverso i *flag* ma non eliminate dai microdati. Tale scelta risponde al principio più generale alla base del sistema informativo di Oros, secondo il quale l'archivio è utilizzato per molteplici scopi oltre quelli della rilevazione (cfr. §1.3).

Le istituzioni pubbliche e le istituzioni private *non profit* sono fortemente presenti nelle sezioni da P a S alle quali, come sopra ricordato, l'indagine Oros è stata estesa a partire dalla diffusione di dicembre 2013. Tali unità si concentrano, in particolare, nei settori dell'istruzione (sezione P della classificazione Ateco 2007) e della Sanità e assistenza sociale (sezione Q); le organizzazioni associative, spesso di natura *non profit*, inoltre, caratterizzano in particolar modo il settore delle Altre attività di servizi (divisione S-94 della classificazione Ateco 2007). Poiché nei settori dell'industria e dei servizi di mercato (sezioni da B ad N della classificazione Ateco 2007) tali unità sono scarsamente presenti, prima dell'estensione del campo di osservazione, la loro individuazione si basava sull'utilizzo di criteri di esclusione per natura istituzionale definiti secondo regole relativamente semplificate. In particolare, anzitutto si selezionavano tutte le unità abbinate per codice fiscale con il registro Asia, nella versione ufficiale più aggiornata e, quindi, come tali definite imprese o istituzioni di natura privata. I casi residuali si gestivano attraverso informazioni tratte unicamente dall'Anagrafica Inps. In particolare, venivano escluse dal campo di osservazione unità relative alle Amministrazioni pubbliche con personale di esattorie imposte consumo nel settore Credito, assicurazione e tributi (con CSC pari a 6.03.02), gli enti pubblici e le amministrazioni statali (identificati con il primo *digit* del CSC rispettivamente paria 2 e a 3). Inoltre, sulla base degli stessi criteri adottati da Asia, si escludevano anche i *condomini* (identificati attraverso il CSC pari a 7.06.01, proprietari di fabbricato, settore terziario) la cui natura non è assimilabile a quella delle imprese. Inoltre, mentre l'individuazione delle istituzioni pubbliche veniva effettuata a livello di matricola contributiva, l'attribuzione dello status di istituzione pubblica avveniva a livello di impresa, ridefinendo istituzioni pubbliche tutte le posizioni contributive afferenti allo stesso codice fiscale.

Con l'estensione del campo di osservazione, i metodi di esclusione delle unità sono stati riesaminati e affinati per tener adeguatamente conto della caratterizzazione dei nuovi settori (sezioni da P ad S, Ateco 2007) e delle nuove informazioni tratte dall'archivio Asia Esteso. Infatti, l'acquisizione di questo nuovo archivio non solo ha permesso l'abbinamento con un numero più ampio di unità, ma ha anche reso disponibili maggiori informazioni utili alla definizione dei criteri di esclusione. In particolare, le nuove variabili di fonte Asia utilizzate nell'attuale procedura sono l'Ateco 2007, la forma giuridica e il *flag* di appartenenza alla lista S13. Dalla fonte Anagrafe Tributaria vengono, invece, tratte l'Ateco 2007 e la forma giuridica. Infine, dell'Anagrafica Inps si utilizza il CSC sia per l'individuazione delle istituzioni pubbliche (ma con criteri diversi rispetto al passato) sia per derivare l'Ateco 2007, informazione non disponibile dalla fonte (cfr. §3.2.1) ma ricavabile per transcodifica dell'Ateco 1991 e del CSC (cfr. Istat, 2013a).

L'abbinamento con gli archivi Istat e Anagrafe Tributaria avviene per codice fiscale, mentre le informazioni di fonte Inps sono disponibili a livello di matricola. Va ricordato che l'utilizzo delle variabili avviene secondo una regola di tipo gerarchico che ordina le fonti nel modo seguente: Asia, Anagrafe Tributaria e Inps; una volta effettuata la ricerca nel primo

archivio corrispondente al livello gerarchico, in caso di indisponibilità dell'informazione, essa viene ricercata nell'archivio successivo. Tenendo conto di questo principio, il caso particolare dell'attribuzione dell'Ateco prevede l'assegnazione dell'Ateco 2007 fonte Asia, l'assegnazione dell'Ateco 2007 fonte Anagrafe Tributaria alle unità residuali e, infine, l'Ateco 2007 da transcodifica di Ateco 1991 e CSC da fonte Inps. In media nei trimestri, si osserva che la provenienza dell'Ateco è per il 90% da fonte Asia, per il restante 9,6% dall'Anagrafe Tributaria e, infine, per lo 0,4% da fonte Inps. A tutte le unità afferenti alle sezioni A (agricoltura), O (amministrazioni pubbliche), T (attività di famiglie e convivenze), U (organismi extraterritoriali) e alle unità della divisione S-94 (attività di organizzazioni associative) viene assegnato un *flag* che segnala la non appartenenza al campo di osservazione di Oros.

Per l'individuazione ed esclusione delle istituzioni pubbliche, condomini e organizzazioni associative si utilizzano tre criteri relativi alla natura istituzionale, alla forma giuridica e alla natura di condominio definiti attraverso quattro variabili: la forma giuridica di fonte Asia e Anagrafe Tributaria, l'appartenenza alla lista S13 di fonte Asia e il CSC di fonte Inps. In particolare, la natura istituzionale dell'unità viene valutata utilizzando, nell'ordine, il *flag* S13 di fonte Asia e il CSC di fonte Anagrafica Inps per le unità non abbinate con Asia. Secondo quest'ultima informazione, in particolare, vengono escluse unicamente le matricole con CSC pari a 3.01.01 (Amministrazioni statali, università e scuole statali). Anche il criterio relativo alla forma giuridica utilizza in prima battuta la fonte Asia con l'esclusione dal campo di osservazione delle imprese con forma giuridica Istat pari a 1.7 (Ente privato con personalità giuridica), 1.8 (Ente privato senza personalità giuridica) e 2 (Forme disciplinate dal diritto pubblico) ad eccezione della 2.3 (Autorità indipendente)⁹; successivamente, per le unità non abbinate con Asia si considera la forma giuridica di fonte Agenzia Entrate, escludendo tutte le imprese con forma giuridica pari a 8 (Associazioni riconosciute), 9 (Fondazioni residenti), 10 (Altri enti ed istituti con personalità giuridica), 12 (Associazioni non riconosciute e comitati), 13 (Altre organizzazioni di persone o di beni senza personalità giuridica, escluse le comunioni), 15 (Enti pubblici non economici), 16 (Casse mutue e fondi di previdenza, assistenza, pensioni o simili con o senza personalità giuridica), 17 (Opere pie e società di mutuo soccorso), 18 (Enti ospedalieri), 19 (Enti ed istituti di previdenza e di assistenza sociale), 20 (Aziende autonome di cura, soggiorno e turismo), 21 (Aziende regionali, provinciali, comunali e loro consorzi), 39 (Altri enti ed istituti non residenti), 41 (Fondazioni non residenti), 42 (Opere pie e società di mutuo soccorso), 43 (Altre organizzazioni di persone e di beni), 51 (Condomini), 55 (Amministrazioni pubbliche), 56 (Fondazioni bancarie)¹⁰. Infine, si procede all'identificazione dei condomini, riconosciuti prima nel registro Asia con la forma giuridica pari a 1.8.30, poi nell'Anagrafe Tributaria con forma giuridica pari a 13 o 51 e per i casi residuali con CSC pari a 7.06.01. Anche alle unità che presentano le caratteristiche descritte viene assegnato un *flag* di non appartenenza al campo di osservazione delle stime Oros.

3.2.5 Il trattamento delle riclassificazioni delle variabili strutturali e anagrafica annuale

Con il passaggio alla nuova base 2010 è stata introdotta un'ulteriore innovazione volta a conferire agli indicatori congiunturali pubblicati una maggiore stabilità garantita attraverso la produzione di stime depurate da effetti spuri derivanti da modifiche, nel tempo, delle

⁹ Cfr. Istat 2005.

¹⁰ Cfr. Agenzia delle Entrate 2007.

variabili di classificazione. Questi effetti sono da attribuirsi principalmente a riclassificazioni delle unità rispetto alle variabili strutturali utilizzate per la loro identificazione nel campo di osservazione. I cambiamenti possono essere effettivi (trasformazioni giuridiche, modifiche nella natura istituzionale o cambiamenti nell'attività prevalente svolta) oppure apparenti (miglioramenti metodologici). In passato, al fine di fornire dati molto aggiornati sulle variabili *target* (sia micro per gli altri utilizzatori interni, sia macro per diffusione), venivano aggiornate ogni anno le informazioni strutturali derivanti dalle diverse fonti, per tutte le unità sia in stima provvisoria, sia in stima definitiva (cfr. Baldi et al, 2011). Tale operazione comportava la gestione di un numero importante di riclassificazioni derivanti, in particolare modo, dall'acquisizione, in stima definitiva, di una versione più aggiornata del registro Asia. In esso, la revisione da un anno all'altro delle variabili di classificazione delle unità può essere dovuta sia alla registrazione di modifiche nella struttura delle unità realmente verificatesi, sia alla metodologia utilizzata per la loro assegnazione. Un primo effetto sulle stime Oros era quello di provocare revisioni molto rilevanti (spesso la principale causa di revisione) e concentrate su particolari aggregati economici. Un ulteriore effetto dovuto allo spostamento da un anno all'altro di unità tra settori, più evidente nelle serie storiche degli aggregati diffusi, si manifestava in cambiamenti di livello degli indicatori *target*. Considerata la ridotta variabilità inter-settoriale dei valori pro-capite delle retribuzioni e degli oneri, tali spostamenti risultavano di trascurabile impatto sulle variabili di costo del lavoro, pubblicate a livello di sezione. Diversa l'incidenza delle variabili espresse come totali, quali le posizioni lavorative, per cui tali transizioni avevano un impatto più rilevante.

Per tali motivazioni, di concerto con i principali utilizzatori interni delle informazioni diffuse e a seguito di un'accurata analisi delle revisioni, a partire dalla diffusione di giugno 2015 si è stabilito di iniziare a produrre indicatori con variabili strutturali bloccate nel periodo di tempo che intercorre tra due cambi base¹¹. Tale operazione di sterilizzazione delle riclassificazioni riguarda ovviamente le sole unità *panel*. Per le unità neonate le variabili strutturali vengono tratte dalle tre fonti citate sopra, osservando i criteri di gerarchia descritti e, in seguito, fissando l'informazione attribuita fino al successivo cambio base. Va precisato che, data la particolarità della rilevazione (non strettamente di tipo *panel*), ed in cui vengono utilizzate informazioni tempestive e aggiornate al trimestre di riferimento, alcune delle riclassificazioni effettive vengono registrate, come nel caso di trasformazioni giuridiche che comportano nella maggior parte dei casi l'apertura di nuovi codici fiscali.

L'operazione di assegnazione delle variabili di identificazione delle unità in base all'appartenenza al campo di definizione viene effettuata ogni trimestre ma, per i criteri di blocco appena citati, riguarda solo le nuove nate, ossia le unità che per *codice fiscale* (la principale chiave di aggancio rispetto agli archivi d'impresa utilizzati per trarre le informazioni statisticamente rilevanti)¹², non risultano presenti nei trimestri precedenti.

11 Di fatto, questa è la prassi seguita nelle indagini congiunturali, in cui l'impresa viene identificata all'anno base e, a meno di modificazioni strutturali rilevanti che richiedono comunque soluzioni *ad hoc*, spesso concordate tra indagini che diffondono indicatori fra loro correlati, essa viene mantenuta fissa sino al successivo cambio base. Una trattazione su vantaggi e svantaggi sul congelamento delle informazioni strutturali nelle indagini congiunturali è contenuta nel manuale metodologico di Eurostat sulle indagini *short-term* sulle imprese (cfr. Eurostat, 2006).

12 L'abbinamento delle variabili di classificazione delle unità dell'anagrafe Oros per codice fiscale garantisce che l'unità statistica di riferimento sia l'*impresa*, l'entità presente sia in Asia, sia nell'Anagrafe Tributaria. Le unità residuali, quelle per cui le informazioni sono tratte dalla fonte Inps, rappresentano meno dell'1% dell'insieme di riferimento di Oros e sono quasi totalmente *uniposizionate*, ossia vi è una sola matricola per ciascun codice fiscale. Dato il trascurabile impatto, per le *pluriposizionate* si è stabilito di non identificare un Ateco d'impresa per evitare di ricorrere a criteri discutibili.

3. Le informazioni anagrafiche

L'aggiornamento dell'anagrafica Oros avviene secondo un criterio annuale, ossia al primo trimestre di ciascun anno si crea una lista iniziale che contiene solo le unità ricevute nello scarico del trimestre. Tale lista, nei tre trimestri successivi dell'anno, viene incrementata solo con unità nuove/riattivate, ossia non presenti nei trimestri precedenti. In tal modo, al quarto trimestre si dispone di una lista anagrafica annuale, in cui solo alle informazioni utili per l'identificazione dello *status* trimestrale dell'unità a quelle necessarie alla decodifica dei DM viene conferita una valenza trimestrale. Infine, poiché dopo un anno dalle stime preliminari viene rilasciata anche una stima finale, tale procedura implica che la lista anagrafica utilizzata per le prime stime potrà differire da quella utilizzata per le stime definitive. In particolare, mentre le informazioni anagrafiche non vengono riviste, quello che varia è il numero di unità abbinabili che, nella versione annuale, sarà maggiore, includendo anche le unità che inviano dichiarazioni contributive all'Inps prima di registrare la propria nascita. L'anagrafica annuale consente, in tal modo, di recuperare in stima definitiva un piccolo problema di sottocopertura (cfr. §6.2.4).

4. DAL DATO AMMINISTRATIVO ALLE INFORMAZIONI STATISTICHE. BANCA DATI NORMATIVA, CONTROLLI DI QUALITÀ E CRITERI DI AGGREGAZIONE¹

4.1 I DM2013: copertura e variabili amministrative

Il flusso delle denunce retributive e contributive dell'Inps, il cui sfruttamento è alla base della produzione dei principali indicatori della rilevazione Oros, riguarda la generalità dei lavoratori dipendenti del settore privato, inclusi coloro che, pur avendo copertura previdenziale e assistenziale obbligatoria (A.G.O.) presso altri Enti sostitutivi, versano all'Inps contribuzioni minori (malattia, maternità, mobilità, Cuaf, Cig, disoccupazione, ecc.). Per dare una prima quantificazione, l'uso delle dichiarazioni retributive e contributive consente di coprire il 99% circa dell'occupazione dipendente nei settori privati di riferimento di Oros (cfr. §2.1). La parte residuale dell'occupazione dipendente non rilevata attraverso il flusso informativo Inps utilizzato ai fini della rilevazione è in gran parte rappresentata dalla manodopera occupata in alcune aziende del settore agricolo, le cui dichiarazioni contributive seguono un flusso separato rispetto a quello principale sui lavoratori dipendenti (ex DMAG). La sottocopertura in questione riguarda prevalentemente operai alle dipendenze di imprese che lavorano nell'industria alimentare (divisione 10 Ateco 2007) e nel commercio all'ingrosso (divisione 46).

Nel tempo, la rilevazione Oros è stata caratterizzata da problematiche classiche delle indagini amministrative, legate ai cambiamenti nell'ambito normativo e nella raccolta dei dati, (cfr. §2.3.2). In particolare, nel 2010 il flusso delle dichiarazioni DM10 a livello di impresa e il flusso delle retribuzioni individuali Emens, sono stati integrati nel flusso UniEmens di denunce individuali su retribuzione e contribuzione. Questo cambiamento ha visto il processo Oros, tradizionalmente basato sul modulo DM10, adattarsi dapprima al nuovo DM10 virtuale (aggregazione di UniEmens per impresa, effettuata dall'Inps)² e, successivamente, al DM2013 virtuale³ attraverso un processo di reingegnerizzazione delle procedure di gestione interna. Sebbene la struttura dei dati che l'Inps invia telematicamente all'Istat sia cambiata nel tempo, il DM2013 virtuale preserva in gran parte la struttura dell'originale DM10. Nella figura 4.1.1 si riporta, per una ipotetica impresa, una esemplificazione della struttura del modello DM2013 virtuale riprodotto dall'Inps per determinare il saldo contributivo a carico dell'impresa relativo ai propri lavoratori dipendenti e ottenuto come differenza tra poste a debito e a credito.

¹ Il presente capitolo è a cura di Elisabetta Aquilini e Francesca Ceccato.

² Msg. Inps n.6323 del 3/3/2010.

³ Msg. Inps n.5281 del 28/3/2013.

Figura 4.1.1 – Esempificazione di un modulo DM2013

Dettaglio DM										
Periodo di riferimento:	Gennaio 2013			Numero di pratica:	Da Generare					
Matricola:	1111111111			Data presentazione:	22/02/2013					
Codice Fiscale:	1234567890			Totale a debito:	€ 24.904,00					
Cod. Stat. Contr.:	70108			Totale a credito:	€ 2.730,00					
Cod. Autorizzazione:				Saldo:	€ 22.174,00					
Numero dipendenti:	22									
N. dip. a tempo deter.:	1									
Forza Aziendale:	22									
*codici di quadratura										
Quadro unico										
Codice	Qualifica			Tipo Contribuzione	Tipo Lavoratore	Dipendenti	Tempo Lavoro	Imponibile	Somma a Debito	Somma a Credito
	1	2	3							
0035						3				148,31
0053	2	F				1				1296,13
2000	2	F	D	00	0	1	gg 26.00	2019,00	798,93	
2000	2	F	I	00	0	18	gg 468.00	61122,00	23330,30	
2000	2	P	I	00	0	1	hh 92.00	787,00	300,40	
200M	2	F	I	00	0	1	gg 26.00	194,00	74,05	
5100	5	F	I	30	0	1	gg 26.00	1478,00	257,92	
M053	2	F				1				1285,07
M950	2				0	5		14201,00	142,01	
Q198*									0,39	
Q298*										0,49

Fonte: Inps, Messaggio 5281 del 28/3/2013.

La nuova struttura informativa del modello DM2013 ha lo scopo di valorizzare al meglio le potenzialità e il dettaglio informativo alla base del flusso Uniemens di origine. In particolare, il nuovo modello si caratterizza per la presenza di variabili assenti nel preesistente DM10, che favoriscono un maggior dettaglio del codice, identificandone in modo più appropriato la natura contributiva⁴. Il nuovo modello, inoltre, si presenta in un quadro unico che riporta tutti gli importi senza distinguere, come nella struttura precedente, il quadro BC (voci a debito) e il quadro D (voci a credito) (Baldi et al., 2001): gli importi a credito e a debito sono riportati nelle rispettive colonne ben contraddistinte dalle aggregazioni delle codifiche che ne definiscono la natura contributiva. Gli argomenti relativi alla base informativa utilizzata per Oros verranno ampiamente trattati nei paragrafi che seguono.

Tornando al cambiamento nel flusso informativo dei dati amministrativi, si segnalano, negli anni più recenti, modifiche giuridiche che hanno influenzato la copertura dei dati Inps rispetto alla popolazione *target* della rilevazione, in particolare, la soppressione di alcuni Enti di previdenza sostitutivi e il conseguente versamento dei relativi contributi all'Inps. Tra gli eventi più recenti, si ricorda la soppressione dell'Inpdap e dell'Enpals⁵. Ad agosto 2012 è stato, infatti, introdotto il ramo per le dichiarazioni contributive e retributive dei lavoratori iscritti alla Gestione ex-Inpdap che, tuttavia, non ha avuto effetti rilevanti sul processo Oros essendo la Pubblica Amministrazione esclusa dal campo di osservazione della rilevazione. A febbraio 2013, è stato introdotto il ramo per le dichiarazioni contributive e retributive

4 Lo sfruttamento completo di queste nuove variabili (qualifica 1, qualifica 2, qualifica 3, tipo contribuzione, tipo lavoratore e tempo di lavoro), almeno per ora, non è stato inglobato nel processo produttivo Oros in quanto tale operazione avrebbe implicato la riprogettazione di vaste fasi del processo di produzione della rilevazione (nuovo studio di fattibilità, riscrittura di tutte le procedure, valutazione delle stime prodotte, revisione degli indicatori congiunturali diffusi, ecc.). Uno sfruttamento più intensivo di tali informazioni, tuttavia, consentirebbero di arricchire considerevolmente l'informazione finora tratta, sulla base di una struttura informativa semplificata rispetto a quella alla base dei tradizionali DM10. Lo sfruttamento di tali informazioni rientra, comunque, nel programma di lavoro a medio termine della rilevazione.

5 Regolamentato dall'art. 21 del decreto legge 6 dicembre 2011, n. 201, convertito, con modificazioni, dalla legge 22 dicembre 2011, n.214.

dei lavoratori iscritti alla Gestione ex-Enpals che, successivamente, è stato pienamente integrato nel ramo principale per le dichiarazioni contributive e retributive dei lavoratori dipendenti implicando, a partire dalla competenza di gennaio 2015, la disponibilità, nel flusso principale, delle informazioni complete su questa categoria di lavoratori⁶. Tale integrazione ha comportato importanti modifiche nel processo Oros legate, in particolare, all'eterogeneità dei dipendenti del settore dello sport e dello spettacolo, la cui inclusione nelle stime ha richiesto particolari trattamenti, tuttora in corso di affinamento (cfr. §6.5).

4.2 Dalle variabili amministrative Inps alle variabili statistiche Oros: definizioni

La rilevazione Oros, già dal suo impianto iniziale, si basa sull'acquisizione di dati mensili amministrativi nella loro forma integrale e grezza, cioè senza interventi o aggregazioni di sorta da parte dell'Inps. La scelta obbligata di acquisire dati amministrativi grezzi dai quali derivare variabili statistiche è da attribuirsi al breve arco temporale che intercorre tra il termine di legge previsto per l'invio mensile dei dati da parte delle imprese con dipendenti e la data di rilascio dei principali output congiunturali della rilevazione Oros (cfr. §2.2.1).

Le sezioni della dichiarazione mensile DM2013 virtuale, utili ad Oros, contengono una serie di informazioni di carattere amministrativo che permettono di identificare e caratterizzare l'impresa nonché di calcolare i debiti e crediti contributivi che l'impresa ha nei confronti dell'Inps relativamente ai lavoratori dipendenti con almeno una giornata retribuita nel mese. Le informazioni relative alle posizioni contributive, visualizzabili nella sezione dettaglio DM del modello, sono di carattere anagrafico; in particolare, riguardano la matricola assegnata dall'Inps, il codice fiscale (CF), la data di iscrizione, il codice autorizzazione (CA), il codice statistico contributivo (CSC), i dipendenti in forza (Ndip)⁷, ecc.. Le restanti informazioni, nel Quadro unico del modello, sono di carattere retributivo e contributivo; si tratta di dati molto disaggregati e non immediatamente utilizzabili per la costruzione delle variabili statistiche di interesse.

Le informazioni retributive e contributive esposte riguardano il numero delle posizioni lavorative dipendenti retribuite, delle giornate/ore retribuite per lavoratori con tempo di lavoro full-time/part-time, l'imponibile retributivo, i contributi a debito complessivi a carico del datore di lavoro e del lavoratore e i contributi a credito, derivanti da riduzioni contributive o da indennità anticipate dal datore di lavoro (malattia, maternità, assegni familiari, ecc.). Queste informazioni sono distinte rispetto al codice le cui modalità identificano diverse tipologie occupazionali e/o contributive dei lavoratori.

La struttura dei dati così progettata fa sì che i dati relativi ad ogni singola posizione contributiva siano esposti su più record, comportando, per ogni mese di competenza, un elevato numero complessivo di record da acquisire.

Il dettaglio informativo su dipendenti, retribuzioni e contribuzioni contenuto nella fonte Inps rappresenta la base per la quantificazione delle variabili statistiche retribuzioni di fatto, oneri sociali, costo del lavoro, posizioni lavorative dipendenti e unità di lavoro equivalenti a tempo pieno. È stata condotta un'analisi preliminare sul contenuto delle informazioni de-

6 Per maggiori dettagli si veda la composizione del flusso nel documento tecnico UniEmens alla sezione domanda del link <https://www.inps.it/NuovoportaleINPS/default.aspx?itemdir=50252&lang=IT>.

7 Tale informazione, che non ha finalità contributive, contiene anche i dipendenti assenti e non retribuiti. Nel processo Oros, viene utilizzata per la stima delle posizioni lavorative dipendenti totali coerenti con le definizioni statistiche (si veda §8.4).

sumibili dalla fonte Inps e valutarne la corrispondenza con le definizioni teoriche inerenti le variabili di costo lavoro stabilite dai Regolamenti Europei STS e LCI (Rapiti, 1998; Baldi et al., 2001). L'accurato confronto svolto ha portato alla definizione delle variabili attualmente prodotte nel processo Oros attraverso lo sfruttamento di tutte le informazioni presenti nei dati Inps, con l'ausilio dei metadati organizzati in una Banca Dati Normativa (BDN), e l'utilizzo di procedure di ricostruzione finalizzate alla correzione e/o imputazione di dati rispettivamente errati e mancanti.

Entrando nel dettaglio definitorio, le *retribuzioni lorde (di fatto)* della rilevazione Oros includono salari, stipendi e competenze accessorie in denaro, al lordo delle trattenute fiscali e previdenziali, corrisposte ai lavoratori dipendenti, esclusi i dirigenti⁸, direttamente e con carattere di periodicità, secondo quanto stabilito dai contratti, dagli accordi aziendali e individuali, e dalle norme in vigore. In questa definizione rientrano anche altre poste, a carico del datore di lavoro, che passano attraverso le retribuzioni imponibili Inps. In particolare:

- le indennità e eventuali integrazioni di malattia, maternità, infortunio, Cig, ecc. per la parte non a carico esclusivo delle gestioni assistenziali e previdenziali obbligatori per legge;
- le erogazioni a livello decentrato e/o legate ad incentivi di produzione (premi legati a contrattazione di II livello, premi decontribuiti fino al 2008);
- gli accantonamenti alla Cassa edile per ferie, gratifiche natalizie e riposi annui, e le retribuzioni in natura (servizi di vitto, servizi di abitazione, servizi di trasporto, indennità di trasferta e trasferimento, assegni familiari, ecc.).

Sono, invece, inclusi solo parzialmente nelle retribuzioni Oros gli incentivi all'esodo che, non rientrando nelle retribuzioni imponibili a fini contributivi, sono rilevate dall'indagine GI con riferimento alle imprese con più di 500 dipendenti.

Gli *oneri sociali* prodotti da Oros sono il complesso dei contributi a carico del datore di lavoro dichiarati agli enti di previdenza ed assistenza sociale e degli accantonamenti di fine rapporto. Come per le retribuzioni, sono parzialmente inclusi i contributi relativi agli incentivi all'esodo, ancorché ridotti, rilevati dall'indagine GI. Gran parte degli oneri sociali sono desumibili dalle dichiarazioni contributive presentate all'Inps⁹ scorporando la quota a carico del lavoratore già presente nella retribuzione di fatto. Per alcuni assicurati¹⁰, tra cui i giornalisti iscritti INPGI e i dipendenti pubblici iscritti all'ex-INPDAP, sono presenti nei dati Inps solo i contributi minori (cfr. §4.1), mentre sono completamente assenti gli accantonamenti di fine rapporto e i contributi per l'assicurazione Inail per gli infortuni sul lavoro. In questi casi, per stimare la parte mancante sono state predisposte procedure specifiche di imputazione di cui si tratterà più avanti.

Il *costo del lavoro* è dato dalla somma delle retribuzioni lorde e degli oneri sociali.

Una *posizione lavorativa dipendente* è contraddistinta da un contratto di lavoro tra una persona fisica e un'unità produttiva con sede legale sul territorio nazionale¹¹, che prevede lo

8 A partire dal 2015, l'indagine GI ha iniziato la raccolta delle informazioni sulle ore lavorate e sul costo del lavoro relative ai dirigenti, con l'obiettivo di una prossima diffusione. Questa nuova disponibilità e l'arricchimento dei dati amministrativi Inps su tale categoria di dipendenti, conseguente la soppressione dell'INPDAl (art. 42 della legge n.289 del 27 dicembre 2002 "legge finanziaria 2003"), rendono possibile uno sviluppo futuro della rilevazione Oros che riguarderà l'inclusione dei dirigenti negli indicatori su retribuzioni, oneri sociali, costo del lavoro per Ula.

9 E' utile sottolineare che gli oneri sociali Oros sono stimati sulla base dei contributi dichiarati dalle imprese a prescindere dagli effettivi adempimenti contributivi. Considerare quest'ultimo aspetto potrebbe determinare differenze anche rilevanti nella misura della stessa variabile.

10 Fino alla competenza dicembre 2014 anche per i lavoratori dello spettacolo e dello sport iscritti all'ex-Enpals le aziende esponevano solo i contributi minori nel flusso informativo Inps dei lavoratori dipendenti.

11 Sono esclusi dalla quantificazione i residenti sul territorio nazionale che lavorano per unità produttive con sede legale nei territori extra-Comunitari della Città del Vaticano o della Repubblica di San Marino.

svolgimento di una prestazione lavorativa a fronte di una retribuzione. Le posizioni lavorative rappresentano, quindi, il numero di posti di lavoro occupati da lavoratori dipendenti (a tempo pieno e a tempo parziale) indipendentemente dalle ore lavorate, ad una determinata data di riferimento. Nei dati Inps le posizioni lavorative dipendenti espresse in corrispondenza dei “codici” non comprendono i lavoratori a libro paga assenti per tutto il mese, per Cig o per altri motivi quali ferie, permessi, maternità, ecc., tali da non prevedere il pagamento di una retribuzione da parte del datore di lavoro o che percepiscono solo importi non imponibili. Ai fini della diffusione nazionale questa parte mancante dell’occupazione viene stimata con procedure specifiche basate sullo sfruttamento di relazioni tra variabili (cfr. §8.4) e sommata alle posizioni lavorative ottenute direttamente sulla base dei dati Inps per ottenere le posizioni lavorative dipendenti totali.

Come già anticipato nel §2.1, la rilevazione Oros nasce per produrre indicatori trimestrali sulla dinamica di retribuzioni, oneri sociali, costo del lavoro per *Ula*, essendo quest’ultima la variabile disponibile più idonea a quantificare l’input di lavoro che effettivamente concorre al processo produttivo. In particolare, nel processo Oros, le *Ula* rappresentano una misura del volume di lavoro calcolata riducendo in unità equivalenti a tempo pieno il valore unitario delle posizioni lavorative a tempo parziale (cfr. §5.2), di quelle dei lavoratori a chiamata (Lac) e dei lavoratori in somministrazione (ex-interinali) (cfr. §4.4.3). La variabile viene calcolata al lordo e al netto dei dirigenti, sebbene, per coerenza con la retribuzione posta al numeratore, nel calcolo dei valori unitari siano utilizzate le *Ula* al netto dei dirigenti.

4.3 La banca dati normativa Oros (BDN)

Lo sfruttamento dei microdati amministrativi nella rilevazione Oros non sarebbe possibile senza l’utilizzo di metadati completi e continuamente aggiornati. In assenza di un sistema di metadati centralizzato e standardizzato presso l’Inps, per interpretare il contenuto informativo della dichiarazione e organizzare in modo sistematico i vari aspetti che influenzano le variabili *target* di Oros è stato necessario rintracciare, raccogliere, archiviare non soltanto i metadati relativi alla dichiarazione contributiva, ma anche quelli indispensabili per la stima di alcune componenti del costo del lavoro non rilevate nel modello DM2013 e renderle disponibili in un formato facilmente accessibile. A tale scopo è stata progettata e realizzata la BDN (Baldi et al. 2001, 2008).

L’aggiornamento della BDN avviene a cadenza trimestrale e si articola in attività mirate alla gestione e documentazione dei seguenti aspetti:

- i nuovi *riferimenti normativi* sul costo del lavoro subordinato (circolari e messaggi) pubblicati nell’apposita area del sito web Inps vengono acquisiti e conservati come base informativa per ricercare nei dati eventuali nuovi codici inseriti per attuare cambiamenti normativi che si riflettono nell’andamento degli indicatori Oros;

- l’elenco e il significato dei codici (codice tipo contributivo - CTC) vengono tratti da documentazione Inps disponibile on-line (circolari e messaggi, allegato tecnico al Documento tecnico di compilazione dell’UniEmens più aggiornato, ecc.) o fornita appositamente all’Istat dall’Inps (Mappatura unica, ecc.). Tra questi, i *codici occupazione* indicano gruppi di lavoratori omogenei a fini contributivi e riportano il numero dei dipendenti, le rispettive giornate/ore retribuite, le retribuzioni imponibili e le contribuzioni di base. Il codice occupazione è costituito da 4-digit combinati tra di loro secondo regole di compatibilità: il primo rappresenta la qualifica del lavoratore, i due centrali rappresentano il tipo contri-

buzione che indica un'eventuale particolarità contributiva e, il quarto, identifica eventuali casi di retribuzione inferiore al minimale. I *codici contribuzione* segnalano, invece, delle particolarità contributive aggiuntive, a debito o a credito, rispetto ai contributi di base già registrati con i codici occupazione; in questo caso il codice è costituito da 4-digit predefiniti dall'Inps. Molti codici occupazione o contribuzione che si trovano nel DM2013 virtuale, si dice, sono "mappati", "ricostruiti", "codificati" dall'Inps a partire dal valore assunto nei vari elementi dell'UniEmens. La disponibilità di metadati sulle regole di aggregazione dei codici e il miglioramento nella comunicazione dei nuovi codici attraverso circolari e messaggi, hanno reso la BDN più esaustiva. Infine, altri codici che riguardano poste specifiche (ad es. indennità o compensazioni con l'Inps) o aventi significati diversi (ad es. codici statistici) vengono classificati come da escludere dalla ricostruzione delle variabili statistiche obiettivo della rilevazione Oros in quanto non rappresentano né occupazione né costo del lavoro. L'insieme dei codici documentati nella BDN costituiscono la base per la creazione di liste di codici ammissibili, ovvero l'insieme delle modalità formalmente corrette della variabile amministrativa¹². La tempestività nella completezza della lista dipende inoltre dall'acquisizione di nuovi codici direttamente dai dati mensili provvisori dei DM2013, prima ancora che la relativa normativa sia stata diffusa;

- l'elenco e il significato dei *codici statistici contributivi (CSC)* e dei *codici autorizzazione (CA)*, entrambi attribuiti dall'Inps a livello di impresa per determinare la caratterizzazione contributiva dell'azienda e quindi l'aliquota che deve essere applicata per versare i contributi, viene aggiornato con le informazioni derivanti dalla normativa al fine di monitorare la giusta assegnazione delle aliquote contributive Inps. Il CSC è composto da cinque caratteri numerici: il primo indica il settore di attività (ad es. industria), con i caratteri secondo e terzo è indicata la classe, che rappresenta il raggruppamento di attività della stessa natura in cui è possibile suddividere il settore (ad es. per l'industria: carta e cartotecnica; stampa ed editoria; fotografia e cinematografia, alimentare, meccanica, tessile, edilizia, ecc.); il quarto e quinto carattere indicano la categoria che individua la singola attività. Il CSC ha lo scopo di attribuire all'azienda le giuste aliquote contributive in relazione all'attività prestata ed alle assicurazioni cui è soggetta, oltre a permettere le opportune rilevazioni statistiche. Il CSC non sempre è sufficiente da solo ad individuare con esattezza l'aliquota contributiva e per questo si ricorre al CA. Il CA ha esattamente lo scopo di individuare, all'interno di aziende aventi il medesimo CSC, quelle soggette ad una particolarità contributiva o beneficiarie di sgravi e riduzioni. Il CA può anche essere attribuito a fini statistici, senza che la sua presenza influisca sul versamento dei contributi.

La BDN viene aggiornata anche a cadenza annuale su alcuni aspetti, tra cui:

- le *aliquote contributive* Inps. Pubblicate annualmente sul sito web Inps con apposito Messaggio o Circolare, contengono le aliquote con dettaglio per singola voce contributiva obbligatoria legale, a carico del datore di lavoro e del lavoratore, distinte per qualifica (operai, impiegati/viaggiatori/piazzisti, dirigenti, apprendisti, ecc.) e per tipologia contributiva (determinata da incroci tra CSC, CA, dimensione aziendale, fondi di previdenza, ecc.). Come già accennato, per calcolare correttamente il costo del lavoro Oros è neces-

¹² La lista dei codici ammissibili distingue i codici occupazione che partecipano alla ricostruzione di variabili statistiche relative a occupazione, retribuzioni e oneri sociali, dai codici contribuzione che servono solo alla ricostruzione degli oneri sociali e quindi del costo del lavoro. Non è sufficiente stabilire che i codici presenti nelle variabili amministrative siano ammissibili per farli partecipare alla ricostruzione delle variabili *target*, è prima necessario verificare la loro appartenenza alle giuste componenti dell'occupazione e del costo del lavoro. Per questo si attribuisce ad ogni codice un *flag* che ne qualifica la corretta inclusione o esclusione nella ricostruzione delle variabili *target*.

sario scorporare dagli oneri sociali la quota a carico del lavoratore in quanto già compresa nella retribuzione imponibile. Accanto all'aliquota standard a carico del lavoratore, classificata per categorie omogenee di lavoratori, identificate dal CSC dell'azienda e dalla qualifica del lavoratore, si trovano eccezioni a tali aliquote individuabili in base alla classe dimensionale e al CA dell'azienda (ad es. per l'industria: imprese editrici individuate dai CSC 1.12.03, 1.12.04, 1.12.06, 1.18.07 versano contribuzioni per l'AGO ed eventi minori in base ad una aliquota specifica che dipende anche dalla classe dimensionale dell'azienda e dalla presenza o meno dei CA 1S e 3T);

- *le aliquote Inail*. Il premio versato dal datore di lavoro e il relativo monte retributivo per attività economica vengono scaricati dalla banca dati statistica "Aziende e lavoratori" presenti sul sito web dell'Inail, solitamente con cadenza annuale e costituiscono la base per il calcolo delle aliquote medie del totale dei dipendenti. Nell'ipotesi che dirigenti ed impiegati abbiano la stessa aliquota media definita a priori, si determina, per differenza, l'aliquota media degli operai in proporzione al monte retributivo Oros distinto per qualifica.

Il passaggio dal DM10 al DM10 virtuale e, in seguito, al DM2013 virtuale ha reso più complesso l'aggiornamento della BDN anche a causa delle modifiche che hanno riguardato l'impianto dei registri Inps a livello individuale. Questo ha comportato, durante questo periodo di transizione, la perdita di riferimenti normativi diretti ed espliciti sui codici occupazione e contribuzione a livello di impresa, in quanto le informazioni a disposizione (Messaggi, Circolari, Documento e allegato tecnico dell'UniEmens) riguardavano variabili/elementi nuovi, propri della dichiarazione UniEmens a livello individuale. Il mantenimento di una BDN aggiornata è un'operazione particolarmente onerosa, in quanto si basa su competenze giuridiche, informatiche e statistiche che necessitano un costantemente aggiornamento e pertanto rendono il processo di gestione soltanto parzialmente generalizzabile. Inoltre, l'aggiornamento della BDN richiede precisione nell'individuazione delle componenti da includere nella ricostruzione delle variabili obiettivo, in modo da non provocare distorsioni nelle stime degli indicatori pubblicati. Infine, per garantire la qualità delle informazioni contenute nella BDN è fondamentale una stretta collaborazione con l'Inps, ente fornitore dei dati.

4.4 Le diverse fasi del trattamento preliminare

I metadati conservati nella BDN vengono utilizzati per la corretta trasformazione dei microdati amministrativi in microdati statistici, operazione che avviene nella cosiddetta fase di trattamento preliminare. Si tratta di una attività trimestrale che viene eseguita sui dati mensili provvisori dei DM2013 virtuali, con riferimento ai mesi di competenza del trimestre di stima preliminare t , e sui dati definitivi, per i mesi di competenza del trimestre di stima finale $t-4$. Con il passaggio dalla raccolta di microdati aggregati a livello di posizione contributiva (DM10) ai microdati raccolti a livello individuale (UniEmens-DM2013 virtuale), l'Inps ha curato molto la qualità delle informazioni raccolte inserendo vincoli di compatibilità tra codici e di coerenza tra variabili e importi, comportando un miglioramento considerevole nei dati provvisori che si discostano sempre meno da quelli definitivi, per lo stesso insieme di unità. La raccolta dei dati individuali ha implicato, però, l'aggiunta della fase di processo relativa alla "virtualizzazione" a livello di posizione contributiva che richiede all'Inps circa 10 giorni lavorativi che si aggiungono alla scadenza di legge per l'invio delle dichiarazioni contributive, implicando la disponibilità del dato relativo all'ultimo mese del trimestre a +15

giorni dalla fine di t . Questo ritardo nella disponibilità dei dati ha avuto delle ricadute sul processo produttivo Oros che per questo è stato oggetto di alcune revisioni. In particolare, a causa della riduzione del tempo a disposizione per i controlli di qualità su dati mensili, è stato necessario prevedere un anticipo sullo scarico dei primi due mesi del trimestre, inoltre, la tradizionale procedura di micro *editing* (Baldi et al., 2008) è stata sostituita con procedure di controllo e correzione basate su criteri di selettività più efficienti su dati trimestrali (cfr. §6.6).

Di seguito vengono illustrati i vari passaggi dell'attuale trattamento preliminare con particolare attenzione alla qualità del dato statistico prodotto e all'attinenza alla variabile *target*. Si tenga conto che l'intero trattamento preliminare viene effettuato su tutte le unità con dipendenti, a prescindere dalla classe dimensionale dell'impresa. Solo in fase di stima provvisoria e stima finale ai valori delle variabili *target* di fonte amministrativa vengono sostituiti i valori d'indagine per le imprese appartenenti al panel della rilevazione GI (cfr. §7.4).

4.4.1. Controlli sui file Ascii

Trimestralmente, seguendo una programmazione degli scarichi stabilita annualmente tra enti, l'Inps trasmette per via telematica all'Istat 6 file Ascii di microdati mensili per un totale di circa 0,6 GB con riferimento ai dati provvisori delle posizioni contributive con almeno un lavoratore dipendente, e altrettanti per i definitivi. Sui dati pervenuti vengono effettuati dei primi controlli sulle numerosità dei record per valutare il grado di popolamento dell'archivio DM2013 virtuale alla data programmata, informazione fondamentale soprattutto per i dati provvisori, in quanto l'Istat si riserva la possibilità di chiedere un nuovo scarico laddove non vengano rispettati determinati requisiti. Negli ultimi trimestri, ogni singolo file mensile contiene in media 1,5 milioni di posizioni contributive per un totale di circa 9,5 milioni di record, ovvero, in media, oltre 6 record distinti per codice occupazione e contribuzione relativi a una singola posizione contributiva.

I file pervenuti contengono una selezione di variabili presenti nell'archivio Inps riguardanti informazioni di carattere anagrafico, retributivo e contributivo.

La prima operazione effettuata nella fase di trattamento preliminare riguarda l'attribuzione del codice di attività economica ufficiale della rilevazione; questa informazione, assegnata a ciascun record per codice fiscale o posizione contributiva, è l'Ateco 2007 definita nell'anagrafe Oros (cfr. §3) e si utilizzerà nelle successive analisi e aggregazioni economiche previste nel trattamento preliminare.

4.4.2. Controlli di validità e coerenza dei codici

Il prerequisito fondamentale per il corretto sfruttamento dei micro dati amministrativi è la validità dei codici occupazione e contribuzione riportati in corrispondenza dei singoli record. La valutazione della loro validità si basa sulla corrispondenza con la lista dei codici ammissibili aggiornata trimestralmente in BDN. In caso di mancato abbinamento, viene effettuato un approfondimento *ad hoc* sul singolo codice finalizzato a verificare la possibilità di un eventuale aggiornamento della lista dei codici ammissibili e la conseguente corretta inclusione del codice nelle stime (cfr. §4.3).

Appurata la validità di tutti i codici, la fase successiva del trattamento prevede una serie di operazioni relative ai soli codice occupazione. In particolare, si procede all'as-

segnazione della qualifica lavorativa e del tempo di lavoro ai singoli codici attraverso la ricodifica dell'informazione contenuta nel codice stesso, finalizzata al corretto calcolo delle Ula netto dirigenti. Un trattamento particolare riguarda i codici occupazione esposti su più record in corrispondenza dei lavoratori all'estero dipendenti di unità produttive con sede legale in Italia; il trattamento consente di evitare una sovrastima dei dipendenti stessi indotta dalla modalità di esposizione delle componenti retributive convenzionali ed effettive sul DM2013 virtuale.

Un ulteriore controllo a livello di codice riguarda la verifica della completezza del contenuto delle variabili quantitative ad esso associate (numero dipendenti, numero giornate retribuite, monte retributivo, monte contributivo), finalizzata all'individuazione di possibili mancate risposte parziali. In seguito, al fine di calcolare correttamente il costo del lavoro viene assegnata a livello di codice per qualifica e di CSC l'aliquota contributiva a carico del lavoratore stabilita in BDN. Per valutare la qualità dell'aliquota assegnata, viene calcolata sui dati l'aliquota contributiva media ponderata con il numero dei dipendenti e confrontata con quella ottenuta nello stesso mese dell'anno precedente.

I dati mensili a livello di codice occupazione e contribuzione contenenti le variabili amministrative, validate dai controlli di qualità sopra esposti, costituiscono il *set* di dati provvisori "integrali" (e *set* di dati definitivi "integrali") che rappresenta la base a partire dalla quale si ricavano le variabili statistiche Oros definite nel §4.2. Per ogni posizione contributiva, il processo prevede l'aggregazione dei record relativi ai codici considerati utili ai fini degli indicatori Oros, sulla base dei metadati della BDN, sintetizzando in un solo record i dati relativi a ciascun modello DM2013 virtuale. Complessivamente costituiscono 1,5 milioni di posizioni contributive nel *set* di dati provvisori sintetizzati (e *set* di dati definitivi "sintetizzati"). Durante questa fase di aggregazione viene effettuato il calcolo delle posizioni lavorative dipendenti e delle retribuzioni di fatto per qualifica e tempo di lavoro e, infine, degli oneri sociali.

4.4.3. L'estrazione delle variabili statistiche

Come più volte detto, le variabili *target* della rilevazione Oros sono le posizioni lavorative dipendenti, le retribuzioni di fatto, gli oneri sociali e il costo del lavoro. Per una corretta quantificazione delle posizioni lavorative dipendenti sono stati considerati alcuni aspetti legati al contenuto della variabile amministrativa in relazione alle nuove tipologie contrattuali relative ai Lac e ai lavoratori in somministrazione (cfr. Istat, 2013a e §4.2). Non considerare le peculiarità di questi contratti comporterebbe la sovrastima del numero di posizioni lavorative dipendenti e, pertanto, viene adottata una procedura di "riproporzionamento" che consente di tener conto adeguatamente dell'effettivo input di lavoro che tali tipologie contrattuali apportano. In particolare, i lavoratori subordinati con contratto di lavoro a chiamata vengono esposti nella dichiarazione contributiva come lavoratori full-time e la quantità di lavoro da essi prestata viene indicata in ore retribuite. La soluzione metodologica adottata per i Lac prevede il calcolo di un fattore di riproporzionamento, ottenuto rapportando il relativo monte ore mensile al numero di ore medie contrattuali stimate dall'indagine mensile sulle retribuzioni contrattuali (IRC), attraverso il quale è possibile quantificare le posizioni lavorative *full-time equivalent* (Ula). Diverso è il trattamento che riguarda, invece, i dipendenti interinali, le cui informazioni vengono dichiarate soltanto dall'impresa somministratrice e non dall'impresa utilizzatrice. Ai fini della contabilizzazione nella dichiarazione contributiva mensile viene registrato ogni distinto rapporto di lavoro (missione) che ogni dipendente in-

terminale ha con un terzo soggetto (impresa utilizzatrice), in corrispondenza del quale vengono esposte le giornate retribuite per i lavoratori full-time, le ore retribuite per i part-time. La stima delle posizioni lavorative dipendenti in somministrazione full-time è ottenuta, a livello di singola impresa, dividendo il relativo monte giornate mensile dichiarato, per il numero di giornate lavorabili in un mese da un dipendente full-time (pari a 26 giornate convenzionalmente fissate dall'Inps come numero massimo di giornate contribuibili nel mese); le posizioni lavorative dipendenti a tempo parziale, invece, sono ottenute dividendo il relativo monte ore mensile dichiarato, per il numero di ore medie mensili lavorate dai dipendenti part-time nel loro complesso, attualmente stimate sulla base dei dati relativi all'anno base 2010 (pari a 93,3 ore). Alcuni controlli di qualità vengono effettuati sulle variabili statistiche "riproporzionate" così ottenute, per valutare la validità della metodologia nel tempo, alla luce di eventuali cambiamenti normativi e/o amministrativi.

Le retribuzioni di fatto a livello di posizione contributiva si ottengono aggregando le componenti retributive presenti nei dati mensili integrali in corrispondenza dei codici occupazione validi, mantenendo anche il dettaglio per qualifica e tempo di lavoro. A queste componenti si aggiungono i premi retributivi soggetti a decontribuzione, esposti in corrispondenza di specifici codici contribuzione, su cui non è riportato il dettaglio per qualifica e tempo di lavoro: la loro stima per qualifica avviene attraverso una ripartizione proporzionale del totale alle quote del monte retributivo relativo alle altre componenti, distinto per qualifica e tempo di lavoro.

Gli oneri sociali a livello di posizione contributiva si ottengono come saldo tra le somme delle componenti contributive a debito e a credito del datore di lavoro associate ai codici occupazione e contribuzione validi, previsti dalla lista dei codici ammissibili di BDN. Nel paragrafo 4.3 è stato spiegato che per calcolare correttamente il costo del lavoro Oros è necessario scorporare dagli oneri sociali la quota dei contributi a carico del lavoratore che, se non fosse esclusa, causerebbe una duplicazione in quanto già compresa nella retribuzione imponibile. Per consentire il calcolo di questa quota, viene assegnata, ad ogni record dell'insieme di dati in versione "integrale" con valenza contributiva, l'aliquota contributiva specifica che, applicata alla retribuzione imponibile, restituisce l'importo da sottrarre a livello di posizione contributiva al totale degli oneri sociali. All'importo sopra descritto mancano le componenti a carico del datore di lavoro riferite ai lavoratori assicurati per contribuzione obbligatoria legale ad Enti diversi dall'Inps e calcolate applicando le aliquote IVS di legge alle retribuzioni imponibili dichiarate per tali lavoratori in corrispondenza delle contribuzioni minori (cfr. §4.2). Per questo, si imputa una quota di legge pari al 22,54% della retribuzione per i giornalisti iscritti all'INPGI, e pari al 23,81% per i dipendenti iscritti all'ex-Inpdap. Quest'ultima quota è stata imputata anche per i lavoratori dello sport e dello spettacolo iscritti al soppresso fondo ex-Enpals fino ai dati di competenza dicembre 2014, dopodiché a partire da gennaio 2015, in seguito all'integrazione delle dichiarazioni contributive ad essi riferite nel flusso informativo Inps dei lavoratori dipendenti, il calcolo degli oneri sociali viene eseguito sulla base dell'effettiva contribuzione dichiarata.

Alle componenti contributive quantificate sopra si aggiungono l'assicurazione Inail contro gli infortuni sul lavoro e l'accantonamento di fine rapporto (Tfr) che non sono esposte nel DM2013 virtuale. Nel primo caso, le aliquote Inail distinte per operai ed impiegati e stabilite in BDN vengono attribuite alle posizioni contributive mensili per sezione Ateco 2007 (divisione per l'attività manifatturiera). L'aliquota Inail media degli operai e degli impiegati calcolata per settore Ateco 2007 viene moltiplicata, a livello di posizione contributiva per il monte retributivo per ottenere così il costo dell'assicurazione. Nel secondo caso, il

Tfr viene stimato come quota della retribuzione Oros (pari a 1/13,5) per singola posizione contributiva.

Prima delle operazioni conclusive di calcolo degli oneri sociali alcuni controlli di qualità vengono effettuati sulle singole componenti stimate, in particolar modo analizzando le distribuzioni percentuali e verificando la coerenza delle stesse rispetto alla componente retributiva.

In formula, per ogni posizione contributiva i e per un mese generico $m=1, 2, 3$ del trimestre corrente t , gli oneri sociali (o) a carico del datore di lavoro per le posizioni lavorative dipendenti al netto dei dirigenti (nd) stimati sulla base dei dati amministrativi comprendono:

$$o = {}_{INPS}o + {}_{ALTRI\ ENTI}o + Inail + Tfr \quad (1)$$

dove gli oneri sociali Inps che derivano dal DM2013 virtuale sono dati da:

$${}_{INPS}o = {}_{DEBITO\ DATORE}o - {}_{CREDITO\ DATORE}o - {}_{DEBITO\ LAVORATORE}o \quad (2)$$

Al termine del processo di stima di tutte le variabili statistiche utili al calcolo degli indicatori Oros, inizia una fase di verifica della qualità dei dati mensili aggregati per posizione contributiva, prima di procedere all'aggregazione trimestrale. In particolare, un controllo consiste nel verificare l'assenza di posizioni contributive duplicate nel singolo mese di competenza finalizzato alla creazione di un *set* di dati univoci per posizione contributiva. Tale duplicazione, già presente nei dati Inps, è legata alla possibilità fornita all'impresa di inviare la dichiarazione contributiva più volte ad integrazione o a sostituzione di una inviata in precedenza per la stessa competenza. Per questo, in caso di duplicazioni si verifica se i record appartenenti ad una stessa posizione contributiva sono duplicati identici o se sono duplicati diversi attraverso misure di variabilità in corrispondenza delle variabili principali del processo. Solo nel caso di duplicati diversi, ipotizzando un'integrazione della dichiarazione, i valori delle variabili si ottengono sommando i valori registrati nei duplicati. Nei casi rimanenti la procedura restituisce una sola posizione contributiva.

Con la disponibilità di un *set* di dati univoci per posizione contributiva si procede all'aggregazione trimestrale di tutte le variabili sin qui stimate mensilmente e si costruiscono *flag* di controllo sui tre mesi del trimestre di stima. Un primo *flag* valuta la presenza della posizione contributiva nei singoli mesi del trimestre ed è volto a segnalare la presenza di mancata risposta di una o più dichiarazioni contributive nel trimestre; *flag* aggiuntivi segnalano, invece, per ognuna delle singole variabili posizioni lavorative dipendenti, retribuzioni e oneri sociali la presenza di importi diversi da zero in ognuno dei tre mesi del trimestre.

L'aver sfruttato statisticamente microdati amministrativi scaricati trimestralmente, ma con dettaglio mensile e l'aver costruito *set* di microdati mensili e trimestrali per posizione contributiva ha permesso di costruire un sistema complesso di dati utilizzabile per molteplici scopi, oltre a quelli di rilevazione Oros (cfr. §1.3). Ne sono alcuni esempi l'utilizzo che ne viene fatto nell'indagine Vela-GI come popolazione di riferimento, nell'indagine GI come supporto nella procedura di *editing* e imputazione (E&I), nel registro RACLI (Registro Annuale Costo del Lavoro per Impresa) come fonte per l'integrazione di variabili mancanti da fonti trattate direttamente (cfr. Ceccarelli et al., 2017), oppure nei conti nazionali trimestrali come indicatori congiunturali di confronto e di supporto alle stime. Il continuo mutare dei dati di fonte amministrativa, che recepisce tempestivamente i cambiamenti del legislatore, è un punto di forza di una rilevazione come Oros rispetto ad una indagine di

tipo tradizionale, ma solo se il cambiamento viene gestito con tempestività all'interno del processo di produzione del dato statistico. Ed è per questo che accanto ai punti di forza legati alla grande quantità di informazione e di dettaglio tipici di un sistema complesso, si nascondono possibili punti di debolezza, come l'influenza sulla qualità dell'informazione statistica dei frequenti cambiamenti nei regolamenti amministrativi se non gestiti adeguatamente, la mancanza di metadati standardizzati per tradurre l'informazione amministrativa, la necessità di investimenti IT adeguati ed efficienti per supportare il processo produttivo caratterizzato da una grande quantità di dati grezzi. Tali controindicazioni possono essere superate creando una maggiore sinergia tra vari settori dell'Istituto (produzione, IT, privacy, ecc.) e tra l'Istat e l'Inps detentore dei dati e dei relativi metadati.

5. IL TRATTAMENTO TRIMESTRALE¹

5.1 Dai dati mensili ai dati trimestrali: attribuzione delle informazioni anagrafiche e controlli di congruenza

La fase finale del trattamento preliminare sulle dichiarazioni contributive mensili rende disponibili, come visto nel capitolo 4, variabili economiche espresse come monte trimestrale, sia per i dati di stima provvisoria, sia per quelli di stima finale, a livello di matricola aziendale. Al fine di arrivare alle stime delle variabili *target*, occorre attribuire ai microdati economici le informazioni anagrafiche rilevanti per calcolare le stime rispetto al campo di osservazione e ai domini di riferimento della rilevazione. A questo livello viene portato avanti un insieme più ristretto di variabili, quelle utili ai principali obiettivi di stima della rilevazione. Molte altre variabili vengono conservate a livello mensile per eventuali usi specifici o successivi. Attraverso un *link* per matricola aziendale, ai microdati economici trimestralizzati vengono attribuite le informazioni anagrafiche. Nel caso di stima provvisoria, l'abbinamento viene effettuato rispetto ad un'anagrafe aggiornata al trimestre *t* di riferimento, mentre nel caso di stima finale l'aggancio può avvenire rispetto all'anagrafe annuale (già disponibile sull'intero anno di stima finale) (cfr. §3.2.5), consentendo di recuperare informazioni anche su eventuali unità di sottocopertura in stima provvisoria, per le quali esisteva un dato economico ma non risultavano ancora iscritte o riattivate nell'anagrafica (si stima che tale fenomeno interessi circa lo 0,05-0,1% dell'occupazione complessiva di Oros, con un peso variabile tra settori; cfr. §6.2.3).

A questo livello viene effettuata una prima serie di controlli con l'obiettivo di verificare la presenza di errori o errate classificazioni presenti negli abbinamenti avvenuti con il registro delle informazioni anagrafiche. Le informazioni strutturali ottenute dalle varie fonti amministrative e statistiche in questa fase del processo vengono sintetizzate ed utilizzate per definire dei *flag* di appartenenza ai domini di stima e di eventuale esclusione dalle stime per motivi definitivi. Risulta, pertanto, necessario effettuare dei riscontri di congruenza delle informazioni disponibili. Vengono effettuati controlli separati per le unità abbinare e non abbinare con il registro Asia. Sul primo insieme, vengono essenzialmente effettuati conteggi di numerosità per Ateco (divisione e sezione) e calcolate distribuzioni rispetto ai totali, valori monitorati in serie storica. Inoltre, come visto nel capitolo 2, l'abbinamento effettuato con la versione estesa del registro Asia non garantisce, sebbene l'abbinamento avvenga, che tutte le variabili necessarie ai fini di Oros siano valorizzate nel registro. Poiché le informazioni non recuperate tramite il registro vengono tratte da fonti differenti, per tali unità sono necessari controlli di coerenza più stretti rispetto ad informazioni *proxy* presenti nel registro Asia.

Inoltre, controlli più attenti si rendono necessari per l'insieme delle unità non abbinare con Asia, a causa della non certificata qualità delle variabili strutturali tratte dall'anagrafe tributaria e/o dall'anagrafe Inps. Oltre a calcolare delle distribuzioni su alcuni insiemi di aggregazione, vengono effettuati dei *focus* sulle variabili utilizzate per definire ogni singolo

¹ I paragrafi 5.1 e 5.3 di questo capitolo sono a cura di Donatella Tuzi; il paragrafo 5.2 è a cura di Marco Lattanzio.

criterio di esclusione dal campo di osservazione: Ateco, Forma giuridica, Natura istituzionale, Lista condomini. Infine, vengono definite delle liste di unità di grandi dimensioni sulla base di soglie per la variabile dipendenti, le quali vengono confrontate in serie storica valutando nuove entrate ed eventuali uscite dagli insiemi così definiti.

5.2 Stima delle Ula

Nel trattamento trimestrale, una prima operazione consiste nella stima delle Unità di lavoro equivalenti a tempo pieno (Ula), variabile non disponibile nei dati amministrativi e assunta quale denominatore per il calcolo delle variabili unitarie di costo del lavoro. Le Ula rappresentano l'unità di misura del volume di lavoro prestato dalle posizioni lavorative e che concorre alla produzione del reddito a numeratore. Sono calcolate a livello di divisione Ateco 2007, per ciascun istante t e per ogni posizione contributiva i , riducendo il valore delle posizioni a tempo parziale ($d_{i,t}^p$) in equivalenti, in termini di Unità di lavoro, a quelle a tempo pieno ($d_{i,t}^f$):²

$$Ula_{i,t} = d_{i,t}^f + q_t d_{i,t}^p \quad (1)$$

Nella (1) il termine q_t è un coefficiente di riduzione trimestrale, calcolato per ogni settore, attraverso cui l'input di lavoro dei part-time viene riportato in *full-time equivalent* (traduzione in inglese del termine Ula). Non disponendo nei dati amministrativi di una variabile di tempo del lavoro per i full-time³, il coefficiente viene stimato utilizzando dati di fonti diverse, in particolare rapportando le ore retribuite medie mensili dei part-time, disponibili nei microdati Inps, ad una misura media mensile convenzionale di ore dei full-time, considerata a livello di divisione Ateco 2007 e tratta dall'indagine IRC:

$$q_t = Me(q_{i,t}) \quad \text{dove: } q_{i,t} = \frac{ore_{i,t}^p}{IRC \cdot ore_t^f} \quad (2)$$

in cui $IRC \cdot ore_t^f$ sono le ore pro-capite medie mensili dei full-time fissate per contratto al lordo di quelle retribuite e non lavorate per ferie, festività, permessi retribuiti di diversa natura (recupero festività sopresse, diritto allo studio, diritto ad assemblea). Le ore pro-capite dei part-time mediane, calcolate per divisione Ateco 2007, vengono rapportate alle ore contrattuali già disponibili a livello di divisione, ottenendo come risultato il valore del coefficiente per Ateco. Infine, per rendere disponibile questo valore a livello di singola unità, si effettua un nuovo abbinamento per divisione Ateco 2007 ai microdati Oros e conseguentemente vengono calcolate le Ula sia al lordo sia al netto dei dirigenti. L'uso della mediana come stimatore del coefficiente di aggregato ha l'obiettivo di eliminare l'influenza di eventuali *outlier* nella distribuzione della variabile ore^{4,5}. Ai fini di controllo si calcola a livello di sezione Ateco

2 Da questa operazione deriva il termine FTE – *Full Time Equivalent*, tradotto poi in Ula. L'operazione che riconduce il numero di chiamate e di missioni interinali in posizioni full-time, come visto nel §4.4.3, equivale a calcolare le Ula.

3 Nei DM2013 per i full-time si dispone, infatti, solo delle giornate retribuite.

4 In passato la qualità della variabile ore nei DM non era sempre di elevato livello. Nel corso del tempo, il campo ha iniziato ad assumere una qualità più stabile oltre che ad essere maggiormente valorizzato, consentendone un uso quale *proxy* del tempo di lavoro in diverse fasi del processo di produzione della rilevazione. Tuttavia, le peculiarità amministrative di questa variabile non rendono consigliabile il loro uso diretto a fini statistici.

5 La caratteristica di distorsione propria dello stimatore mediana viene meno in presenza di domini adeguatamente popolati, come avviene nei dati utilizzati per la stima delle Ula.

un rapporto tra le Ula stimate e i dipendenti totali, minore di 1 per definizione, e si controlla che sia stabile nel tempo:

$$\frac{Ula}{D} = \frac{D^f + qD^p}{D^f + D^p} = \frac{D^f + D^p + (q-1)D^p}{D^f + D^p} = 1 + (q-1) \frac{D^p}{D^f + D^p} = 1 + (q-1)\pi^p \quad (3)$$

Per costruzione il coefficiente q è minore di 1, per cui il termine $q-1$ è negativo e risulta pertanto minore di 1 anche il rapporto tra Ula e posizioni lavorative. L'ultimo termine della formula descrive la dipendenza del rapporto dalla quota dei part-time sul totale dipendenti (π^p) e dalla stima del coefficiente. Il rapporto tende ad 1 come caso estremo se il peso dei part-time tende a 0 oppure se il coefficiente q tende a 1, caso in cui le ore medie dei part-time coincidono con quelle dei full-time. Il rapporto tende a 0 se il peso dei part-time tende ad 1 e il coefficiente q tende a 0, ossia il contributo orario dei part-time si riduce man mano a fronte di una percentuale alta di part-time.

5.3 Calcolo delle medie mensili e collocazione delle unità nel campo di osservazione

I dati relativi alle principali variabili economiche disponibili a questo livello sono espressi come monte trimestrale. Il passaggio successivo è il calcolo di medie mensili nel trimestre, ottenute dividendo il monte trimestrale per tre (numero di mesi del trimestre). Tale procedimento implica che eventuali situazioni di assenza nel trimestre contribuiscano con un valore pari a 0 nella media mensile. Mentre nei dati finali eventuali assenze sono da attribuirsi a stati di inattività⁶, nei dati preliminari la presenza di rispondenti ritardatari, che comporta valori nulli su tutte le variabili⁷, a questo stadio non può essere discriminata da situazioni di reale inattività, a causa dell'assenza di una lista teorica di rispondenti rapidi (cfr. §6.2.4). In tale versione dei dati, quindi, la media mensile delle variabili espresse come livelli risulta una sottostima del valore finale se a una determinata assenza corrisponde non uno *status* di inattività ma un ritardo nell'invio della dichiarazione contributiva (evento non noto a priori), rendendo necessaria una correzione attraverso procedura di imputazione. Le principali variabili *target* della rilevazione, tuttavia, sono espresse come rapporto (cfr. §2.1). In tal caso, la distorsione della variabile dipende da un'eventuale periodicità tra i mesi del trimestre delle variabili che contribuiscono a numeratore e denominatore. Se vi è uniformità nei tre mesi del trimestre, la distorsione scompare con il rapporto, proprio grazie all'assenza di mancate risposte parziali nel mese. Tuttavia, le variabili trattate da Oros sono notoriamente molto periodiche (es. tredicesima mensilità a dicembre e quattordicesima a giugno se si parla di retribuzioni), producendo microdati medi mensili distorti (in sovrastima o sottostima). Di fatto, data l'esigua incidenza dei rispondenti ritardatari (cfr. §6.2.3) e la loro distribuzione molto uniforme all'interno dei domini di stima, la distorsione delle variabili pro-capite per risposta ritardataria nei dati provvisori è pressoché trascurabile negli aggregati per attività economica, rendendo un'eventuale procedura generalizzata di imputazione delle variabili pro-capite non consigliabile, poiché molto costosa in termini di implementazione e gestione, a fronte di miglioramenti della stima poco rilevanti. Si ricorre,

6 Fanno eccezione le matricole caratterizzate da DM 2013 "anomali" che vengono sottoposte ad un trattamento specifico (cfr. §6.5).

7 Nei dati relativi ai DM2013 il fenomeno delle "mancate risposte parziali" ossia assenza per omessa dichiarazione di alcune variabili del modello utilizzate da Oros è assente.

tuttavia, a una procedura di imputazione micro sulle variabili *target* per un limitato numero di mancate risposte ritenute influenti secondo criteri di selettività (cfr. §6.6). Per correggere la distorsione sulle variabili espresse in livello (posizioni lavorative) l'imputazione a livello micro viene, invece, effettuata attraverso una procedura che corregge tutte le unità potenzialmente caratterizzate da mancata risposta (cfr. §6.2 e §6.3). Nei dati finali, le mancate risposte hanno un peso irrilevante e sono circoscritte ai cosiddetti DM2013 "anomali" richiedendo, anche in questo caso, un intervento di imputazione delle posizioni lavorative, di incidenza molto contenuta (cfr. §6.5).

Attraverso aggregazione per codice fiscale⁸ viene individuata la dimensione media dell'impresa e attribuiti alle singole unità dei *flag* di appartenenza a sottopopolazioni prefissate. Tale suddivisione è utile sia a fini di controllo (*cross-section* e longitudinale), sia di trattamento dei dati. In particolare, nelle stime Oros la popolazione complessiva viene suddivisa in:

- piccole e medie imprese stimate con dati amministrativi (d'ora in avanti PMI);
- grandi imprese di rilevazione (GI). Questo insieme verrà successivamente sostituito con i dati di fonte rilevazione;
- grandi imprese che rientrano nella popolazione coperta dalla rilevazione e stimate con dati amministrativi. Sono le unità che non rientrano nell'insieme delle GI e che nell'istante in cui compaiono per la prima volta hanno più di 500 dipendenti. Tali unità vengono distinte dalle PMI poiché potenzialmente rilevanti sulle stime e oggetto di eventuali trasformazioni giuridiche che provengono dall'insieme GI (d'ora in poi GI_DA, dove DA indica dati amministrativi). Tale distinzione, tuttavia, è utile solo ai fini dei controlli. In fase di stima esse vengono accorpate e trattate allo stesso modo delle PMI;
- unità che fanno riferimento ad agenzie di somministrazione di lavoro temporaneo (ex-interinali) (d'ora in poi INTER). La necessità di isolare questa sottopopolazione è dovuta al loro rilevante peso occupazionale concentrato tra pochissime unità e la frequente e persistente presenza di risposte ritardatarie nei dati provvisori. La loro influenza sulla sezione N (Noleggio, agenzie di viaggio, servizi di supporto alle imprese) suggerisce l'applicazione di una procedura specifica di *editing* sulle variabili *target*, nonché di ricostruzione di dati mancanti sulle posizioni lavorative ai fini delle stime provvisorie (cfr. §6.3).

La sottopopolazione di appartenenza viene individuata all'anno base o all'istante di nascita del codice fiscale e lo status conservato sino al successivo cambio base. Fanno eccezione le GI e le INTER (le prime hanno priorità sulle seconde), che possono diventare tali in ogni istante di riferimento, una volta acquisite le caratteristiche. Nel processo Oros, alcune operazioni vengono effettuate trattando separatamente le sottopopolazioni elencate sopra. In altri casi, si rende necessario unicamente distinguere tra unità stimate con dati GI e restanti unità stimate con dati amministrativi ossia le PMI, le GI_DA e le INTER che, per semplicità espositiva, verranno indicate con NO_GI.

Successivamente, combinando informazioni anagrafiche viene attribuito anche lo status di appartenenza al dominio di riferimento della rilevazione. In particolare, come ampiamente descritto nel capitolo 3, vengono etichettate come escluse dalle stime le unità che non rientrano nei settori economici da B a N e da P a S (S94 esclusa) e che non sono né imprese né istituzioni private. Inoltre, a questo livello vengono individuate come non rientranti nelle stime anche quelle unità che, pur risultanti attive secondo le informazioni anagrafiche nei trimestri di stima provvisoria, non presentano dichiarazione contributiva da

⁸ Dall'aggregazione vengono esclusi i codici fiscali formalmente errati oppure "atipici" (cfr. §3.2.1). Le matricole aziendali di queste unità vengono considerate singole imprese.

almeno 12 mesi. Queste unità non rientreranno nel processo successivo d'imputazione delle mancate risposte.

Una volta definito lo status di appartenenza al dominio della rilevazione e l'eventuale *flag* di esclusione dalle stime, vengono effettuati dei controlli di coerenza tra i *flag* e le informazioni utilizzate per definirli, tramite tabulazioni e distribuzioni.

I dati relativi alle unità che rientrano nelle stime prima di essere aggregati per le variabili di classificazione d'interesse per la rilevazione e sottopopolazione di riferimento vengono sottoposti alla successive fasi di *editing* e imputazione, di cui al capitolo 6 che segue.

6. EDITING E IMPUTAZIONE¹

6.1 Introduzione

Prima di procedere alle operazioni di aggregazione per il calcolo delle variabili *target* i microdati da fonte amministrativa trimestralizzati, sia nella versione per le stime preliminari sia finali, vengono sottoposti a controlli volti ad individuare e correggere mancate risposte e/o valori anomali che indurrebbero distorsioni sulle stime.

Nei dati amministrativi le mancate risposte equivalgono a DM relativi ad imprese attive che sono assenti negli scarichi al momento dell'acquisizione dei dati da parte dell'Istat (cfr. §2.2). Vi sono almeno due ordini di motivi per tali assenze: 1) dichiarazioni contributive inviate oltre la scadenza amministrativa da parte delle aziende, in alcuni casi a seguito di particolari deroghe; 2) ritardi o mancata virtualizzazione degli Uniemens in DM2013 per motivi amministrativi o anche per problematiche informatiche in ambito Inps. Nei DM2013 sono, invece, rari i casi di mancata risposta parziale, ossia di assenza di parte delle informazioni all'interno delle dichiarazioni. Questa caratteristica è conseguenza dei molteplici controlli gestionali presenti sia nei software di "procedura paghe", utilizzati dalle aziende per generare gli Uniemens, sia nelle successive procedure *on-line* di controllo dell'Inps, per le quali le aziende devono passare prima di poter trasmettere il flusso: considerata l'elevata copertura in termini di rispondenti, l'intervento sulle mancate risposte riguarda solo dichiarazioni totalmente assenti. Per la stessa motivazione, sono rari i casi di errori nei microdati di origine: eventuali errori possono essere più facilmente generati dalle procedure di stima di Oros e su tali casi si interviene con la correzione della procedura errata e l'elaborazione viene interamente ripercorsa.

Sull'assenza di dichiarazioni contributive per ritardo d'invio si interviene con imputazione dei microdati delle principali variabili *target*. Mentre per le posizioni lavorative dipendenti l'imputazione avviene in modo estensivo (tutte le unità entrano nella procedura di trattamento), la correzione per mancata risposta delle variabili di costo del lavoro, anch'essa attuata a livello di microdati, si basa sull'individuazione selettiva di un insieme ridotto di unità da trattare: il ridotto impatto che le dichiarazioni ritardatarie hanno su tali variabili, espresse come pro-capite e la tempestività con cui gli indicatori devono essere rilasciati, suggerisce un approccio più mirato, in cui vengono controllate ed eventualmente corrette solo le unità rilevanti per le stime (cfr. §6.5).

Il processo di imputazione riguarda sia i dati della stima provvisoria sia quelli della stima finale sebbene le mancate risposte nei dati finali siano pressoché trascurabili. Mentre per le prime la ricostruzione del dato mancante richiede una preliminare predizione dello "status di attività" a causa dell'assenza di informazioni anagrafiche aggiornate che consentano di stabilire se un'assenza sia o meno mancata risposta (cfr. §6.2) per queste ultime, prodotte

¹ I paragrafi 6.1, 6.2.2, 6.2.3, 6.5.1 di questo capitolo sono a cura di Donatella Tuzi; il paragrafo 6.2.1 è a cura di Donatella Tuzi e Francesca Romana Pogelli; i paragrafi 6.2.4, 6.2.5 e 6.4 sono a cura di Francesca Romana Pogelli e Marco Lattanzio; il paragrafo 6.3 è a cura di Elisabetta Aquilini; il paragrafo 6.5.2 è a cura di Marco Lattanzio; il paragrafo 6.6 è a cura di Eleonora Cimino e Marco Lattanzio.

dalla procedura di virtualizzazione, lo *status* di attive è certo (cfr. §6.4). Un trattamento a parte è riservato all'imputazione delle mancate risposte relative alle agenzie di somministrazione di lavoro (nel documento INTER): data la loro peculiarità rispetto alle variabili considerate in Oros e il rilevante peso nel settore di attività economica in cui sono classificate, per la loro stima preliminare è stata predisposta una procedura *ad hoc* nella quale sia nella ricostruzione della lista dei rispondenti ritardatari sia nell'imputazione del dato mancante, si tengono fortemente in considerazione le informazioni longitudinali disponibili sulle singole unità, successivamente controllate e validate caso per caso (cfr. §6.3).

L'imputazione degli errori e delle mancate risposte viene condotta solo sulle unità stimate con dati da fonte amministrativa, ossia le piccole e medie imprese, le grandi imprese non presenti nei dati d'indagine e le interinali (PMI+GI_DA e INTER). I dati tratti dall'indagine GI sono acquisiti in Oros già inclusivi delle operazioni di E&I effettuate ai fini dell'indagine (Rocci e Serbassi, 2008). Eventuali errori nei dati d'indagine individuati *ex-post* vengono concordati e corretti ai fini di Oros dopo un'attenta valutazione con gli esperti dell'indagine (cfr. §7.4). Una scomposizione dell'errore di stima per cause consente, infine, di tenere sotto controllo vari aspetti del processo di stima provvisoria, indicando alcune criticità su cui intervenire al fine di correggere le stime e migliorarne il processo di produzione (cfr. §6.4).

6.2 Imputazione delle posizioni lavorative per la stima provvisoria

6.2.1 Da un approccio macro ad un approccio micro

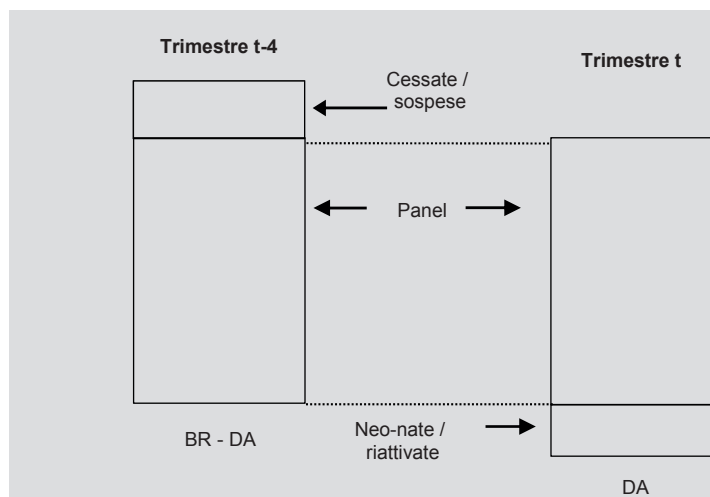
Fino a febbraio 2015 la stima provvisoria delle posizioni lavorative avveniva utilizzando un approccio macro, in cui la correzione per i dati mancanti veniva operata su dati aggregati a livello di divisione Ateco. L'approccio, di tipo empirico, era basato su analisi e valutazioni dell'indicatore in serie storica e su relazioni ricorrenti con indicatori *proxy* e sfruttando, per molti settori, la ricorrenza degli errori di revisione calcolati sui trimestri di stima finale (Istat, 2013a). In condizioni di ciclo economico stabile e con una situazione informativa consolidata, tale approccio consentiva di ottenere stime di buona qualità. In caso di variazioni improvvise del numero di rispondenti per cause di tipo "amministrativo", o in corrispondenza di punti di svolta del ciclo economico, tale approccio appariva meno robusto. Nel corso degli ultimi anni sono state condotte diverse sperimentazioni per la correzione delle stime provvisorie sull'occupazione, a fronte anche di crescenti richieste sia di dati aggregati/disaggregati ad un maggior livello di dettaglio rispetto a quello disponibile, sia di microdati completi. Queste constatazioni hanno sollecitato a muoversi verso la progettazione di un approccio di tipo micro in grado di fornire stime a livelli di classificazione più fini ma, anche, di sfruttare le informazioni più specifiche disponibili sulle singole unità oltre a consentire un controllo maggiore su particolari problemi che si presentano durante il processo di produzione dovuti, ad esempio, ad errori di misura.

Nella nuova procedura di correzione basata sull'imputazione dei microdati, tali metodi di carattere empirico continuano ad essere utilizzati solamente nella fase finale di validazione dei dati, con un impatto marginale e limitato ad alcuni settori notoriamente critici in cui il nuovo metodo ha scarse *performance* (cfr. §8.4.4).

6.2.2 Dati “quasi completi” e stima della demografia d’impresa

Alla base del nuovo metodo vi è un’importante constatazione circa le potenzialità che derivano dalla disponibilità di dati amministrativi (già indicati con DA) “quasi completi” per la stima preliminare di t (cfr. §2.2.1 e §2.3.2), ovvero la possibilità di ottenere una rappresentazione alternativa e più aggiornata della popolazione di imprese attive in quell’istante rispetto a quella che si otterrebbe utilizzando un approccio più convenzionale di individuazione della popolazione di riferimento basato sul registro statistico delle imprese (BR). Normalmente il registro è disponibile con un ritardo non irrilevante rispetto al periodo di stima². Teoricamente, in assenza di dichiarazioni ritardatarie, dato l’obbligo per tutte le imprese di inviare dichiarazioni contributive, i dati amministrativi Inps dovrebbero includere anche le imprese che hanno appena iniziato l’attività nel periodo di riferimento ed escludere quelle che hanno interrotto l’attività nel periodo immediatamente precedente. Nell’ipotesi in cui la stima provvisoria dovesse essere basata sulla lista di unità disponibili nel BR (per ipotesi a $t-4$, figura 6.2.2.1), si avrebbe una situazione in cui verrebbero incluse unità cessate o sospese tra $t-4$ e t (sovracopertura) mentre non sarebbero incluse unità nate o riattivate a t (sottocopertura). Man mano che ci si allontana dall’istante di riferimento del BR le unità caratterizzate da sovra o sottocopertura andrebbero aumentando. Va detto inoltre che, in questa ipotesi, la stima delle variabili *target* non potrebbe essere effettuata per semplice somma dei dati amministrativi disponibili per t sulle unità nel BR in quanto le assenze per sospensioni/cessazioni non sarebbero nemmeno compensate dalle nuove nate in t che, per costruzione, non sono disponibili.

Figura 6.2.2.1 - Popolazione e demografia



Data la tempistica per l’invio delle dichiarazioni da parte dei datori di lavoro all’Inps e considerati i tempi e le modalità di acquisizione da parte dell’Istat, i dati scaricati per le stime provvisorie non sono completi in termini di unità statistiche. Inoltre, in assenza di informazioni anagrafiche adeguatamente aggiornate, in particolare sulle cessazioni e sospensioni, nell’insieme delle unità mancanti non è possibile discriminare tra rispondenti ritardatari e assenti per eventi demografici. Per sfruttare al meglio le potenzialità di tale insieme informativo, il metodo individuato si basa sulla stima di uno *status* per le unità

² In Istat il registro delle imprese attive riferito all’anno a in versione finale è disponibile a marzo di $a+2$.

assenti. Successivamente, sulla lista di unità presunte attive avverrà l'imputazione dei dati mancanti.

Progressi verso la definizione del metodo nella versione attuale si sono avuti a seguito della partecipazione dell'Italia al progetto Europeo ESSnet "Use of Administrative and Accounts Data for Business Statistics" Working Package 4 – "Timeliness of Administrative Sources for Monthly and Quarterly Estimates", nel corso degli anni 2009-2013 in cui l'Istat ha riportato la propria esperienza sull'uso di dati amministrativi per scopi congiunturali illustrando e proponendo come test il caso Oros³.

La messa a regime della nuova metodologia è avvenuta a seguito di una fase di sperimentazione condotta su più mesi. Il nuovo metodo è entrato in produzione a giugno 2015, con successivi avanzamenti introdotti nelle uscite di settembre e dicembre 2015, mentre ulteriori affinamenti sono ancora in programma inducendo ad aggiornare continuamente l'ambiente di test.

Il metodo viene applicato sui dati mensili che, in seguito, sono aggregati a livello trimestrale: la disponibilità del dato mensile permette di sfruttare al meglio le informazioni passate per ricostruire, attraverso un'analisi longitudinale minuziosa, la predizione dello stato di attività delle unità che non presentano dati economici in uno o più mesi e, in modo coerente, anche nella fase d'imputazione.

6.2.3 Le mancate risposte nei dati amministrativi provvisori, caratteristiche

I dati preliminari di fonte amministrativa utilizzati per le stime provvisorie di Oros presentano un grado di riempimento che dipende strettamente dalla schedulazione dei tempi di acquisizione. In particolare, come visto nel §2.2.1 i tre mesi del trimestre vengono scaricati con tempistiche diverse:

- i mesi 1 e 2, scaricati a circa 36-38 giorni dalla fine del trimestre di riferimento, contengono le dichiarazioni pervenute rispettivamente entro 78 e 48 giorni dalla scadenza d'invio prevista per legge;
- il mese 3 a 43-45 giorni, con dichiarazioni pervenute nell'arco di 18 giorni circa.

In riferimento alle PMI e GI stimate con dati amministrativi (GI_DA), il numero di DM ritardatari rappresenta una quota che oscilla tra l'1%, normalmente rilevato nel primo mese dei vari trimestri e il 4-5% nei terzi mesi (grafico 6.2.3.1). Fa eccezione il mese di marzo 2013 in cui, a seguito dell'entrata a regime del DM2013 virtuale in sostituzione del DM10 virtuale (cfr. §2.3.2), il flusso Inps ha subito un rallentamento, comportando un'elevata presenza di rispondenti mancanti (7%). In media, l'incidenza dei ritardatari è lievemente più alta se valutata in termini di posizioni lavorative per effetto della maggiore propensione a ritardare da parte delle unità di dimensione mediamente più grande, caratterizzate da una struttura più difficile da gestire dal punto di vista amministrativo. Inoltre, si nota con chiarezza la maggior percentuale di ritardatari nei terzi mesi dei trimestri. Le più alte incidenze si osservano, inoltre, nei mesi di giugno: a giustificare il maggior ritardo, la minor presenza di personale amministrativo dedicato alla compilazione dei modelli di dichiarazione contributiva per ferie (la scadenza per il mese di competenza di giugno corrisponde alla fine di luglio).

Se analizzata per mese di competenza, la distribuzione delle incidenze dei ritardatari denota più chiaramente la periodicità legata al calendario degli scarichi dei dati: nel primo

3 La documentazione sui principali risultati del progetto è disponibile al link: http://ec.europa.eu/eurostat/cros/content/admindata-essnet-use-administrative-and-accounts-data-business-statistics_en.

mei i DM ritardatari rappresentano l'1-1,5% del totale, percentuale che sale di un punto nel secondo mese, per attestarsi tra il 3,5% e il 5% nel terzo mese (grafico 6.2.3.2).

Pur apparendo uniforme sui macro settori dell'industria, dei servizi di mercato e dei servizi sociali e personali, la distribuzione settoriale dei rispondenti ritardatari risulta più concentrata in alcune sezioni dell'Ateco 2007 (tavola 6.2.3.1). I più caratterizzati sono il settore H - Trasporto e magazzinaggio (3,8% in termini di unità e 4,3% di posizioni), il settore N - Noleggio, agenzie di viaggio, servizi di supporto alle imprese (3,1% e 3,9% rispettivamente), il settore R - Attività artistiche, sportive, di intrattenimento e divertimento (3,9% e 5,6%). Nell'ambito della sezione H è particolarmente rilevante la mancata risposta nella divisione 50 del trasporto marittimo, per effetto di una deroga amministrativa sull'invio dei moduli che prevede la proroga a 60 giorni invece che 30 per alcune categorie di lavoratori (18,9% in termini di unità e 25,2 di posizioni). La divisione J59 relativa alla produzione cinematografica e televisiva (6,5% in termini di unità e 14% di posizioni) e tutte le divisioni della sezione.

Grafico 6.2.3.1 - Incidenza delle mancate risposte in termini di numero di unità e posizioni lavorative
Mesi aprile 2012 - settembre 2015, valori percentuali

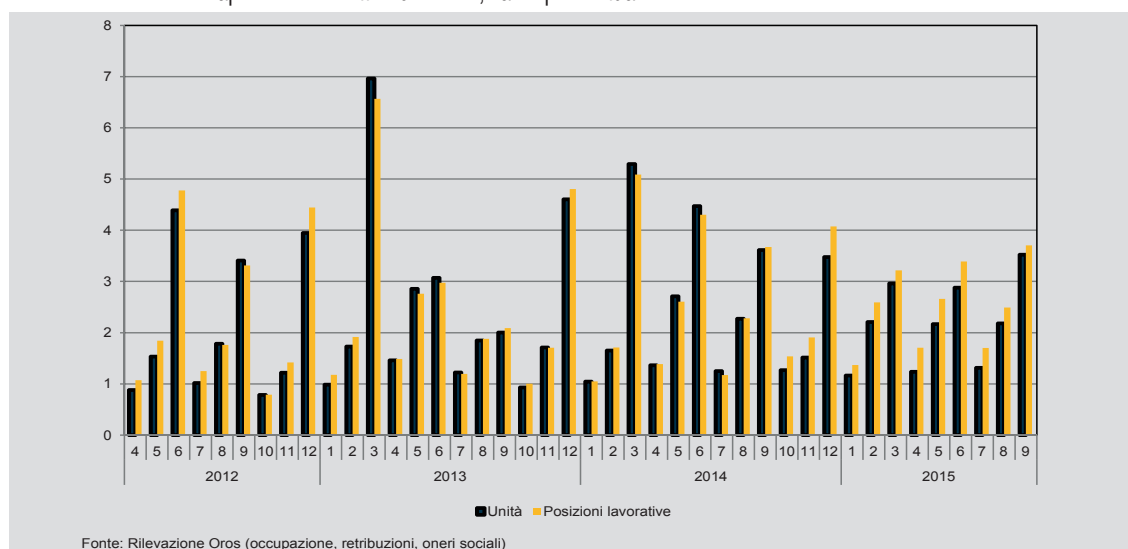
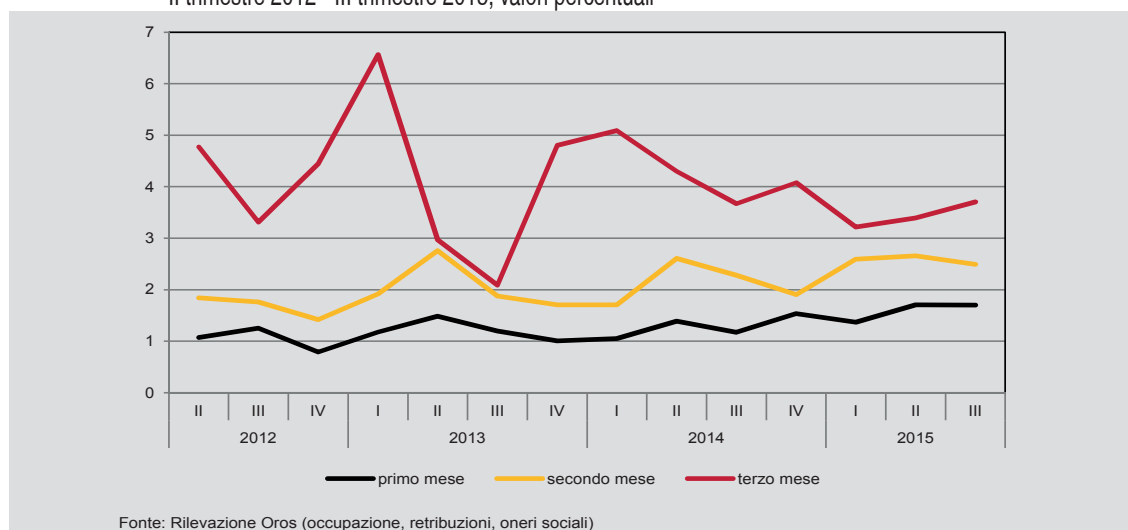


Grafico 6.2.3.2 - Incidenza dei rispondenti ritardatari nel primo, secondo e terzo mese del trimestre in termini di posizioni lavorative dipendenti
Il trimestre 2012 - III trimestre 2015, valori percentuali



Come visto sopra, la stima preliminare delle variabili *target* di Oros per le PMI e GI_DA si basa, in condizioni di stabilità del flusso informativo, su un insieme di rispondenti che, in termini di occupazione media nel trimestre, pesa per il 95-98%. Anche se soggetta ad improvvise cadute occasionali a seguito di eventi particolari, tale percentuale appare molto contenuta e non particolarmente caratterizzata a livello strutturale. Appare, invece, più caratterizzata la distribuzione per classe dimensionale, che vede la propensione a ritardare al crescere della dimensione dell'azienda. Questa evidenza può essere attribuita alla maggiore complessità nella gestione dei flussi Uniemens al crescere del numero di dipendenti per cui devono essere presentate le dichiarazioni contributive: l'incidenza delle ritardatarie passa da poco più di 2 punti per le imprese fino a 50 dipendenti a quasi 5 per le unità con oltre 500 dipendenti.

La distribuzione dei modelli ritardatari comporta distorsioni non rilevanti sugli indicatori delle variabili espresse in termini relativi, in cui sia numeratore che denominatore derivano dall'aggregazione di microdati caratterizzati da mancata risposta totale, e per questo motivo si è deciso di non operare con procedure d'imputazione micro generalizzata per la correzione dei dati mancanti, operazione limitata a poche unità rilevanti segnalate da *editing* selettivo.

L'impatto delle mancate risposte è, invece, totale sugli indicatori delle variabili espresse in termini di livello, quali ad esempio l'occupazione, rendendo indispensabile la correzione della sottostima attraverso imputazione. Nel corso degli ultimi anni sono state condotte diverse sperimentazioni per la correzione delle stime provvisorie sull'occupazione, mettendo a confronto e valutando vantaggi e svantaggi di approcci macro e micro di ricostruzione dei dati mancanti. L'esperienza ha evidenziato come gli approcci macro, basati sulla regolarità in serie storica della dinamica dell'indicatore, pur se più semplici da implementare e meno onerosi da controllare in un contesto di tempestività di produzione delle stime, rischiano di essere deboli nelle fasi di svolta del ciclo economico, limitando notevolmente l'utilizzo dei dati disponibili. Ciò a fronte di crescenti richieste sia di dati aggregati/disaggregati a diverso livello di dettaglio rispetto a quello disponibile, sia di microdati completi. Queste constatazioni hanno sollecitato a muoversi verso la progettazione di un approccio di tipo micro, in grado di fornire stime a livelli di classificazione più elevati, ma anche di sfruttare le informazioni più specifiche disponibili sulle singole unità e consentendo di tenere meglio sotto controllo anche specifici problemi che possono sorgere nel processo di produzione, ad esempio, a causa di errori di misura. Nel propendere verso un approccio micro non bisogna trascurarne i potenziali svantaggi, quali la difficoltà di modellare comportamenti in gruppi eterogenei di unità, la presenza di *missing* o *break* nei microdati che restringe l'uso di lunghe serie storiche, l'onerosità delle fasi di controllo.

6. Editing e imputazione

71

Tavola 6.2.3.1 - Incidenza delle mancate risposte dei rispondenti ritardatari in termini di unità e posizioni lavorative dipendenti per settore economico e classe dimensionale
Aprile 2012 - settembre 2015, Valori percentuali medi di periodo

SETTORI ATECO 2007 CLASSI DIMENSIONALI	Numero unità	Posizioni lavorative
PER SETTORI ATECO 2007		
INDUSTRIA E SERVIZI (B-S escluso O)	2,3	2,5
INDUSTRIA E SERVIZI DI MERCATO (B-N)	2,3	2,4
INDUSTRIA (B-F)	2,2	2
INDUSTRIA IN SENSO STRETTO (B-E)	1,9	1,8
B - Estrazione di minerali da cave e miniere	1,8	1,7
C - Attività manifatturiere	1,9	1,8
D - Fornitura di energia elettrica, gas, vapore e aria condizionata	2,1	2,6
E - Fornitura di acqua; reti fognarie, gestione dei rifiuti e risanamento	2,2	2,9
F - Costruzioni	2,6	2,4
SERVIZI DI MERCATO (G-N)	2,4	2,8
G - Commercio all'ingrosso e al dettaglio, riparazione di autoveicoli e motocicli	2,2	2
H - Trasporto e magazzinaggio	3,8	4,3
I - Attività dei servizi di alloggio e di ristorazione	2,6	2,6
J - Servizi di informazione e comunicazione	2,6	3,6
K - Attività finanziarie ed assicurative	1,9	1,9
L - Attività immobiliari	2	2,3
M - Attività professionali, scientifiche e tecniche	1,8	2,4
N - Noleggio, agenzie di viaggio, servizi di supporto alle imprese	3,1	3,9
SERVIZI SOCIALI PERSONALI (P-S)	2,4	3,1
P - Istruzione	2,6	3,2
Q - Sanità e assistenza sociale	2	2,8
R - Attività artistiche, sportive, di intrattenimento e divertimento	3,9	5,6
S - Altre attività di servizi	2,4	2,5
PER DIVISIONI ATECO 2007 CON MANCATA RISPOSTA MAGGIORE DELLA MEDIA INDUSTRIA E SERVIZI (B-S)		
Industria del tabacco (C-12)	18,5	5,9
Attività di risanamento e altri servizi di gestione dei rifiuti (E-39)	3,2	5
Trasporto marittimo e per vie d'acqua (H-50)	18,9	25,2
Trasporto aereo (H-51)	3,4	5,8
Attività di produzione cinematografica, di video e di programmi televisivi, di registrazioni musicali e sonore (J-59)	6,5	14
Attività di programmazione e trasmissione (J-60)	5,2	7,5
Attività creative, artistiche e di intrattenimento (R-90)	4,1	6,5
Attività di biblioteche, archivi, musei ed altre attività culturali (R-91)	2,7	7,4
Attività riguardanti le lotterie, le scommesse, le case da gioco (R-92)	3,8	5,5
Attività sportive, di intrattenimento e di divertimento (R-93)	3,9	5,1
PER CLASSE DIMENSIONALE		
TOTALE	2,3	2,5
0-49	2,3	2,2
50-249	2,6	2,6
250-499	3,3	3,4
500+	4,4	4,8

Fonte: Rilevazione Oros (occupazione, retribuzioni, oneri sociali)

6.2.4 La predizione della lista delle unità attive non rispondenti

L'anagrafica trimestrale dell'Inps utilizzata nelle stime Oros, come visto nel capitolo 3, è un serbatoio di matricole attive a *t* e/o cessate/sospese da gennaio 2000 e rappresenta il

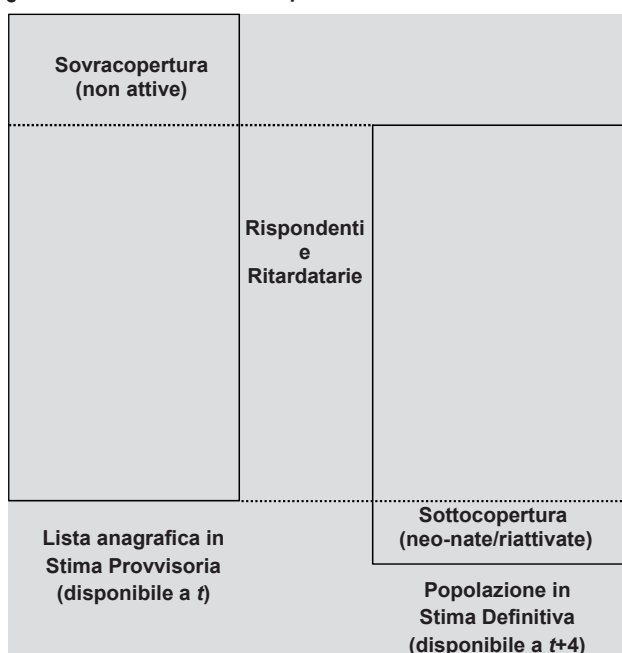


set di microdati amministrativi da cui è possibile individuare la lista di unità attive per t . Se gli eventi demografici fossero registrati tempestivamente tale *set* di dati rappresenterebbe la lista di stima per la rilevazione. Tuttavia, a causa dell'assenza di disposizioni coercitive nei confronti della comunicazione ritardataria sulla modifica dello stato di attività da parte delle aziende, tale *set* di dati è caratterizzato da errori rilevanti di copertura, in particolare riguardo la mancata registrazione di cessazioni e sospensioni, mentre le unità neonate non ancora registrate (ma che invieranno una dichiarazione contributiva) sono in numero ridotto. Al fine di limitare l'impatto di tale problematica sulle stime, si rende necessaria un'operazione di taglio ossia di selezione delle sole unità che hanno presentato almeno una dichiarazione contributiva nel corso dell'ultimo anno rispetto a t .

Confrontando la lista provvisoria definita con i criteri di cui sopra con la lista reale delle unità attive, disponibile dopo un anno con i dati finali (cfr. §2.2.1), per ogni mese m le unità della lista provvisoria definite attive possono risultare, come anche schematizzato nella figura 6.2.4.1:

- *rispondenti*, quindi attive;
- attive ma non rispondenti (*ritardatarie*);
- cessate o sospese, quindi definite impropriamente attive (*sovracopertura*);
- assenti dalla lista anagrafica ma attive nei dati finali, poiché registrate in ritardo (*sottocopertura*).

Figura 6.2.4.1 - La lista nei dati provvisori e nei dati definitivi



In un approccio micro di predizione della lista le unità di sovracopertura, le rispondenti e le ritardatarie costituiscono l'insieme di partenza da cui calcolare le stime. Le unità di sottocopertura, invece, non sono note a priori e, quindi, incideranno totalmente sull'errore di stima (sottostima), seppure in misura pressoché limitata: si calcola che tale insieme, variabile per trimestre, rappresenti lo 0,05-0,1% dell'occupazione complessiva della rilevazione.

Al contrario, malgrado il taglio rispetto alle unità attive almeno un mese tra $t-4$ e t , la sovracopertura rimane un problema molto rilevante, rappresentando oltre il 90% delle unità assenti e potenzialmente attive. Al fine di contenere l'errore di sovrastima che ne

deriverebbe, lo *status* di attività viene ulteriormente rifinito sulla base, anzitutto, delle date anagrafiche: vengono considerate non attive le unità non rispondenti nate nel mese m o successivamente e le cessate nello stesso mese, o nei mesi precedenti. Questo secondo intervento porta ad escludere dall'insieme delle potenziali attive una quota pari a circa lo 0,5% del totale delle assenti nello stesso mese perché neonate: verificando la presenza di tali unità nei dati finali, si riscontra che circa il 60% di tali unità escluse effettivamente non rispondono. Le cessate non rispondenti ad m (con data di cessazione registrata) rappresentano solo lo 0,2% delle assenti; di esse, tuttavia, il 90% sono effettivamente non rispondenti anche nei dati finali. La bassa incidenza delle cessate dichiarate fa emergere l'elevato problema di sovracopertura dell'anagrafica rendendo opportuna una successiva valutazione sulle sospensioni di attività per eventi stagionali. In mancanza di date di sospensione aggiornate, la valutazione di questi eventi, particolarmente concentrati in alcuni settori (produzione alimentare, alberghi e ristoranti ecc.) si basa sulla verifica della persistenza dell'assenza nel corso dell'anno: in particolare, si definisce stagionale un'assenza nel mese a cui corrisponde un'assenza nello stesso mese dell'anno precedente. Fa eccezione il mese di giugno, su cui per evitare sovraidentificazione di stagionalità dovuta alla particolare concentrazione di DM ritardatari nel mese⁴, l'assenza per stagionalità viene valutata anche nello stesso mese di due anni precedenti. A seguito dell'applicazione di questo criterio, le stagionali rappresentano circa il 5% del totale delle assenti. Tuttavia, sulla loro identificazione, si commette un errore di circa il 20%.

Una volta identificate le cessazioni, le sospensioni, le nascite e le stagionali, le assenze residuali vengono sottoposte ad una ulteriore valutazione, in cui lo *status* di attive viene conferito valutando la presenza dell'unità nei mesi vicini a quello di stima. In particolare, si è scelto di assegnare lo *status* di attività all'unità "presunta ritardataria", se essa ha manifestato segnali di presenza nel mese precedente.

Tale approccio, denominato anche "metodo $m-1$ ", ove $m-1$ rappresenta il mese utilizzato per stimare lo *status* di attività nel mese m del trimestre di stima provvisoria t , è stato applicato per due occasioni di stima preliminare, ossia a giugno e settembre 2015. In seguito ad alcuni approfondimenti, a partire dall'uscita di dicembre 2015, è stato lievemente rivisto per ridurre l'errore di sottostima di unità attive, inserendo una variante che prevede che la presenza in un generico mese $m-1$ (o anche la presunta attività in $m-1$) sia segnale di attività anche per il mese successivo m all'interno del trimestre t , a meno di fattori stagionali e demografici. Questo nuovo metodo è stato definito "metodo basato sul mese precedente con variante trimestrale" o in breve "*metodo $m-1$ trim*". In termini formali è possibile definire per ogni unità i , in base alla presenza (1) e/o assenza (0) dei dati della variabile occupazionale, il suo pattern di presenza nei tre mesi del trimestre t , come $p_{i,t}$ che dopo la fase di predizione, per le unità presunte assenti, assume nuove modalità e la sua stima può essere espressa come $\hat{p}_{i,t}$.

I vari pattern che si possono osservare prima e dopo la fase di predizione della lista sono riportati nella tavola 6.2.4.1 in cui è anche incluso il segnale di presenza o meno del dato occupazionale nell'ultimo mese del trimestre precedente, in quanto da tale informazione dipende l'assegnazione del *flag* d'imputazione nel primo mese del trimestre corrente t e a seguire anche eventualmente del secondo e del terzo mese.

⁴ Il mese di competenza di giugno è il più soggetto a ritardi nei tempi di consegna dei modelli a causa del periodo di ferie in cui ricade la scadenza per l'invio delle dichiarazioni contributive (fine luglio).

Tavola 6.2.4.1 – Predizione dello stato di attività nel metodo “m-1 trim”: casistiche dei pattern iniziali e predetti

Presenza del dato occupazionale nel terzo mese del trimestre t-1 (a)	Pattern iniziali del trimestre t	Pattern predetti del trimestre t
0	011	011
1	011	111
0	001	001
1	001	111
0	000	000
1	000	111
0	010	011
1	010	111
0/1	100	111
0/1	110	111
0/1	101	111
0/1	111	111

Fonte: Rilevazione Oros (occupazione, retribuzioni, oneri sociali)

(a) La presenza del dato occupazionale nel terzo mese del trimestre precedente t-1 assume il seguente significato 0=assenza, 1=presenza, 0/1=assenza o presenza

Le unità presunte ritardatarie a m possono essere classificate in base alla loro effettiva realizzazione nei dati finali disponibili dopo un anno, confrontando lo *status* predetto rispetto a quello reale (noto appunto nella popolazione finale) da cui derivano le seguenti definizioni degli *status* teorici:

1 - **attive corrette**: unità non rispondenti nei dati provvisori, definite attive secondo i criteri di predizione della lista e presenti nei dati definitivi. Si tratta di unità correttamente incluse nella lista predetta;

2 - **attive non corrette**: unità non rispondenti nei dati provvisori, definite non attive secondo i criteri della lista ma presenti nei dati definitivi. Sono unità erroneamente escluse dalla lista determinando un errore che può essere chiamato di “sottoinclusione”;

3 - **non attive corrette**: unità non rispondenti nei dati provvisori, definite non attive nella lista predetta e assenti nei dati definitivi. Si tratta di una corretta esclusione delle unità dalla lista;

4 - **non attive non corrette**: unità non rispondenti nei dati provvisori, definite attive nella lista, ma non presenti nei dati definitivi. Sono unità erroneamente incluse in fase di predizione determinando quindi un errore di “sovrainclusione”.

Nel seguente schema ogni cella rappresenta i vari status definiti sopra, risultati dal confronto delle unità predette rispetto a quelle reali:

Prospetto 6.2.4.1 - Status delle unità nella popolazione dei dati provvisori e realizzazione nei dati finali

Popolazione nei dati finali	Popolazione nei dati provvisori		
	Unità attive e Rispondenti	Non rispondenti	
		Unità presunte attive	Unità presunte non attive
Attive (rispondenti)	Attive	Attive corrette (A) (correttamente incluse)	Attive non corrette (B) (erroneamente escluse)
Non attive (cessate, sospese)	-	Non attive non corrette (D) (erroneamente incluse)	Non attive corrette (C) (correttamente escluse)

Tanto più il numero delle unità nei due gruppi delle attive non corrette (B) e delle non attive non corrette (D) si accosta, determinando un bilanciamento tra la sottocopertura e la sovracopertura delle lista, tanto più la numerosità della popolazione predetta sarà vicina

a quella *target* e la lista risulterà predetta correttamente. In ogni caso il bilanciamento deve avvenire il più possibile in termini di posizioni lavorative.

La classificazione delle unità con *status* predetto nelle casistiche appena esposte è utile per valutare *ex-post*, quando disponibili i dati finali, la tenuta del modello applicato e, analizzando l'errore per singolo sottogruppo, per affinare la metodologia applicata. Ciò si rivela particolarmente utile di fronte a cambiamenti strutturali, non rari nei dati amministrativi, che possono influenzare i comportamenti dei rispondenti ritardatari. Utilizzando tale classificazione, nel §6.4 verranno presentati alcuni risultati del metodo di individuazione della lista.

6.2.5 Ricostruzione del dato per regressione

Il secondo aspetto affrontato dal metodo d'imputazione riguarda la ricostruzione del dato mancante sulle unità presunte attive, relativo alla sottopopolazione delle PMI e GI_DA.

Partendo dai lavori svolti in ambito ESSnet, in cui erano stati presi in considerazione tre metodi di stima del dato dell'occupazione in termini di livello, di cui due basati su degli stimatori rapporto e il terzo su un modello di regressione bivariato, la scelta finale adottata è un *progress* di tali approcci.

Lo stimatore rapporto, se da un lato può risultare più intuitivo e anche più semplice in termini d'applicazione, dall'altro però non tiene conto della compresenza del dato congiunturale e tendenziale, entrambi informazioni utili per la stima del dato corrente. Ciò diventa rilevante in alcuni settori Ateco 2007 (principale variabile di classificazione) caratterizzati da un'intensa attività stagionale, come quelli connessi al turismo e alle industrie alimentari.

In un contesto di dati longitudinali in cui si sfrutta il legame dei dati tra periodi successivi e nel caso specifico di variabili occupazionali, tale legame risulta ancora più forte. Le posizioni lavorative dipendenti sono di fatto una variabile per sua natura inerziale e di conseguenza i dati passati risultano ottimi *predictor* di quelli mancanti.

Alla luce di ciò, per la stima del livello mancante sulle unità presunte ritardatarie è stato adottato un modello di regressione bivariato in cui le variabili ausiliarie sono espresse dalla stessa variabile oggetto d'analisi, ma in tempi diversi, ovvero il mese precedente ($y_{i,m-12}$) e lo stesso mese dell'anno precedente ($y_{i,m-1}$). In particolare per le unità che hanno entrambi i dati delle variabili ausiliarie, si adotta un modello bivariato, mentre nei casi in cui è disponibile solo un dato o nel mese m o nello stesso mese dell'anno precedente $m-12$, si utilizza un modello univariato.

In formule i modelli sono:

$$\begin{aligned} y_{i,m} &= \alpha_{A,m} + \beta_{1A,m} y_{i,m-1} + \beta_{2A,m} y_{i,m-12} + \varepsilon_{i,m}, & \text{(caso A)} \\ y_{i,m} &= \alpha_{B,m} + \beta_{B,m} y_{i,m-1} + \varepsilon_{i,m}, & \text{(caso B)} \\ y_{i,m} &= \alpha_{C,m} + \beta_{C,m} y_{i,m-12} + \varepsilon_{i,m} & \text{(caso C)} \end{aligned} \quad (1)$$

dove l'indice i esprime l'unità i -esima della popolazione, le variabili $y_{i,m}$ e $y_{i,m-1}$ indicano rispettivamente il numero di dipendenti nei mesi m e $m-1$, $\varepsilon_{i,m}$ è il termine di errore a media nulla. I parametri (α, β) distinti nei tre casi (A, B e C) sono stati stimati con il metodo dei minimi quadrati sui sottogruppi di unità omogenee, secondo la classificazione dell'attività economica in divisione Ateco 2007 e nel caso di divisioni di piccola entità, per alcuni loro raggruppamen-

ti, appartenenti comunque allo stesso livello di classificazione superiore (la sezione Ateco).

L'ordine della modellizzazione espressa nella (1) ha valore gerarchico ed è quello applicato sulla presenza dei rispondenti nei periodi precedenti al mese m . La presenza del dato tendenziale come variabile ausiliaria, nonostante sia più distante rispetto al dato congiunturale e , e, quindi, più soggetta a cambiamenti nei livelli, si riferisce comunque a un valore "vero" in quanto è in versione definitiva (e in teoria non affetto da errori) ed è il più indicato per rappresentare le attività stagionali.

La procedura d'imputazione micro opera in due fasi: una prima, in cui vi è la stima dei parametri dei modelli espressi nella (1) calcolata sul *set* informativo dei dati disponibili, ossia i rispondenti della stima provvisoria; una seconda, invece, consiste nella predizione del dato dell'occupazione sulle unità presunte attive non rispondenti, secondo le relazioni espresse nella (1).

Esiste, in genere, una diversa distribuzione distinta dalla tipologia di modello applicato, tra il *set* informativo delle unità rispondenti su cui vengono stimati i parametri e la realizzazione della stima provvisoria per le unità presunte ritardatarie. Tale distinzione dipende principalmente da due fattori: il primo riguarda la struttura delle mancate risposte nei dati amministrativi e il secondo rispecchia la modalità di costruzione della lista delle unità assenti presunte attive. In particolare, riguardo il primo aspetto si ha che il terzo mese del trimestre è stato sempre caratterizzato da molte più mancate risposte rispetto ai primi due mesi, a causa della stretta vicinanza tra la data di trasmissione dei dati amministrativi con la data di scadenza per l'invio delle dichiarazioni che, ricordiamo, è di circa 15 giorni, contro i 70 e i 40 giorni del primo e secondo mese rispettivamente; mentre per il secondo fattore si ha che l'assegnazione dello *status* di "unità presunta ritardataria" è riferita, a meno di fattori stagionali o di eventi demografici, a tutti e tre i mesi del trimestre di riferimento e quindi, è sufficiente che almeno in un mese precedente l'unità sia valutata come imputabile da risultare tale anche in tutti e tre i mesi del trimestre di riferimento (cfr. §6.2.4). Questo giustifica il fatto che molte unità del secondo e del terzo mese del trimestre sono imputate con il modello "tipo B e C", in quanto il dato del primo mese di tali unità è verosimilmente assente e imputabile, ma non essendo presenti contemporaneamente i due valori al lag 1 e 12 per applicare il modello "tipo A", la scelta ricade automaticamente nelle altre due casistiche.

6.3 Stima provvisoria per la sottopopolazione delle interinali

Le agenzie di fornitura di lavoro interinale (ex-interinali) sono classificate nella sezione N ("Noleggio, agenzie di viaggio, servizi di supporto alle imprese") della classificazione Ateco 2007. In termini di posizioni contributive nel 2015 ammontano a circa 200 unità; in termini occupazionali pesano in media il 22% sulla sezione N e, più nel dettaglio, poco più del 98% dell'occupazione nella divisione 78 ("Attività di ricerca, selezione e fornitura del personale"). Considerando, dunque, il potenziale rilevante impatto sulle stime di anomalie nei dati, anche di una sola unità, o eventuali assenze per ritardo nell'invio della dichiarazione, la sottopopolazione delle interinali è sottoposta a procedure di controllo e correzione specifiche rispetto al resto delle unità.

Anche l'individuazione dello stato di attività per le stime provvisorie delle interinali si basa sullo sfruttamento delle informazioni nell'anagrafe Inps. In aggiunta, l'anagrafe Inps contiene informazioni aggiornate (perché utili ai fini amministrativi) che consentono di indi-

viduare la sottopopolazione delle interinali in modo molto preciso, anche in stima provvisoria. In particolare, fino a dicembre 2010 l'Inps prevedeva che le agenzie di somministrazione di lavoro aprissero una o più posizioni contributive, tutte identificate con codice statistico contributivo (CSC) pari a 7.07.08 e codice di autorizzazione (CA) 9A. All'interno delle dichiarazioni di tali unità, il personale somministrato veniva esposto con il tipo contribuzione pari a 83 e 97 (circolare Inps 153 del 15/07/1998). A partire da gennaio 2011 l'Inps ha cambiato la modalità di identificazione (circolare Inps 149 del 24/11/2010) prevedendo l'utilizzo di due tipi di posizioni contributive: un primo tipo, relativo al personale della struttura (*staff*), identificato solo tramite CSC 7.07.08 ed un secondo, relativo ai lavoratori somministrati, contraddistinto, oltre che da CSC 7.07.08, anche dalla stringa 9A nel CA. Secondo questa nuova e più precisa modalità si ha, quindi, anche la possibilità di isolare e stimare in modo più corretto il numero di posizioni in somministrazione, variabile particolarmente rilevante tra gli indicatori prodotti dalla rilevazione (cfr. §2.1). Tuttavia, poiché il CSC 7.07.08 non identifica solo tale tipologia di imprese, la selezione delle unità per definire la lista di stima basata esclusivamente su questo criterio porterebbe ad includere molte altre unità che non svolgono attività di somministrazione di lavoro interinale. Si è stabilito, quindi, di effettuare la selezione a livello di imprese, inserendo nella lista di stima le posizioni contributive con solo personale somministrato (CSC 7.07.08 e CA 9A) e, tramite il loro codice fiscale, recuperando le matricole dello *staff* per abbinamento.

La lista così individuata viene in seguito abbinata ai dati economici provvisori. Per risolvere il problema delle mancate risposte, diversamente dalle PMI e GI_DA, per individuare lo *status* di attività delle INTER, oltre a sfruttare l'informazione anagrafica disponibile e osservare longitudinalmente il loro comportamento in termini di tempestività di risposta, si prende in considerazione anche il loro comportamento *cross-section*, ossia rispetto alle altre unità della stessa tipologia: tra le unità con dichiarazione assente, attive per data anagrafica, vengono considerate inattive quelle che nei dodici mesi precedenti non hanno presentato neanche una dichiarazione contributiva⁵. Per facilitare la generalizzazione del processo di definizione dello stato di attività sono, infine, considerate inattive in tutti i mesi di assenza del trimestre, quelle unità per cui un evento di nascita, cessazione o sospensione ricade all'interno del trimestre, a prescindere dal mese in cui l'evento viene datato⁶. Prima di procedere all'imputazione delle mancate risposte sulle singole matricole un'ultima visione d'insieme, rispetto all'impresa di riferimento, consente di verificare se l'assenza non sia in realtà da imputarsi a motivi di pura gestione amministrativa: infatti, non è insolito che i lavoratori interinali possano essere ceduti tra matricole afferenti alla stessa impresa di somministrazione, chiudendo matricole ed esponendo i relativi dipendenti sulle altre. Generalmente l'insieme delle matricole interinali interessate da almeno una mancata risposta nel trimestre supera raramente le 30 unità, pari a circa il 15% dell'insieme complessivo.

Nell'imputazione delle mancate risposte delle unità interinali, si considera l'eventualità di imputare, oltre alle posizioni lavorative, anche le variabili coinvolte nella stima del costo del lavoro. In particolare, le variabili interessate alla ricostruzione del dato sono le seguenti: dipendenti totali, dipendenti netto dirigenti distinti in part-time e full-time, monte retributivo netto dirigenti, monte oneri netto dirigenti, unità di lavoro equivalenti a tempo pieno (Ula)

5 L'errore di revisione causato da sottocopertura della lista delle unità attive a seguito di questa ipotesi è assolutamente trascurabile: le dichiarazioni ritardatarie non hanno natura inerziale nei lunghi periodi.

6 Questa ipotesi di semplificazione al trimestre, avvalorata dalla maggiore volatilità dello stato di attività di queste unità, viene successivamente verificata sulle singole unità interessate mediante ulteriori controlli sui microdati, con particolare attenzione alle unità più rilevanti in termini di peso occupazionale.

al lordo e al netto dirigenti. La ricostruzione dei dati economici per le unità caratterizzate da almeno una mancata risposta viene estesa all'intero trimestre, a prescindere dall'eventuale presenza di almeno una dichiarazione contributiva nei tre mesi che, quindi, non viene considerata nel calcolo delle stime. Questa semplificazione è avvalorata dall'evidenza sulla similitudine di comportamento tra rispondenti e mancate risposte rispetto alle variabili *target*, tali per cui la sostituzione del dato mancante con un'informazione ausiliaria tratta dai rispondenti, alla base della metodologia utilizzata, porta ad una distorsione molto ridotta.

Nella ricostruzione viene adottato un approccio deterministico, in cui si sfruttano le informazioni longitudinali, a livello trimestrale, disponibili nell'arco dell'ultimo anno. In particolare, per le variabili relative all'input di lavoro (dipendenti e Ula) e quelle relative alle retribuzioni si utilizza, valutando i trimestri precedenti da $t-1$ a $t-4$ ⁷, il primo dato medio mensile disponibile (eventualmente anche imputato) della stessa unità, a cui viene applicato un tasso di aggiornamento calcolato quale mediana dei tassi di variazione tra il trimestre in cui si selezionano le informazioni per l'imputazione ($t-x$) e il trimestre corrente (t) delle sole unità interinali rispondenti (r) nel trimestre t^8 . In formule, la ricostruzione del dato mancante per la generica unità caratterizzata da almeno una mancata risposta (mr) può essere espressa come:

$$mr\hat{y}_{i,t} = mr y_{i,t-x} \left(1 + Me(r y_{i,t,t-x}) \right) \quad (2)$$

Per la variabile oneri sociali viene attivato un trattamento a parte, ossia l'informazione mancante viene ricavata dai rispondenti del trimestre corrente, in quanto fortemente legata alla legislazione sugli obblighi contributivi, pertanto risulta importante considerare l'informazione più aggiornata. Quindi, la ricostruzione delle mancate risposte degli oneri sociali si ottiene applicando al monte retributivo () dell'unità da imputare, precedentemente stimato, l'aliquota mediana calcolata a partire dal rapporto tra monte oneri (O) e monte retribuzioni (R) relativo ai rispondenti. In formule:

$$mr\hat{o}_{i,t} = mr\hat{r}_{i,t} Me\left(\frac{rO_t}{rR_t}\right) \quad (3)$$

La considerazione del tasso di variazione interperiodale dei rispondenti nella formula di ricostruzione delle posizioni lavorative (e quindi delle Ula) e delle retribuzioni (e quindi degli oneri) consente implicitamente di tener conto sia della forte periodicità che caratterizza le variabili di costo del lavoro, sia dell'inerzia temporale delle variabili di input di lavoro.

Al fine di validare i dati imputati vengono, infine, effettuati controlli sia *cross-section* sia longitudinali sulle singole unità ricostruite. Dato il loro limitato numero, infatti, i controlli possono essere estesi a tutte le unità e non basati, a priori, su criteri di selettività, così come è necessario per il resto delle PMI e delle GI_DA (cfr. §6.6).

Un ulteriore controllo del dato stimato viene effettuato anche con i dati definitivi, dispo-

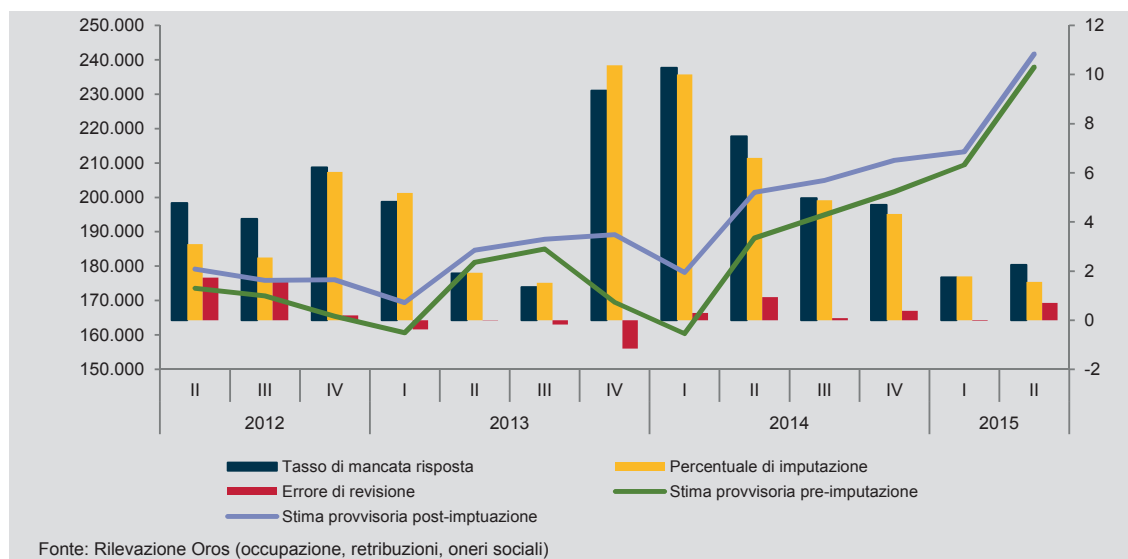
7 Si ricorda che se non è stata presentata neanche una dichiarazione contributiva tra t e $t-4$ l'unità non fa parte della lista delle attive.

8 Ancor più che per le altre sottopopolazioni, per le interinali si osserva un comportamento marcatamente inerziale nell'evoluzione delle posizioni lavorative. La variabile, più che affetta da stagionalità, è fortemente legata al ciclo economico. Il dato di $t-1$ è quello che approssima meglio il dato mancante di t . Inoltre, le retribuzioni medie di questa sottopopolazione sono molto uniformi implicando che la retribuzione di $t-j$, aggiornata con il tasso di variazione delle rispondenti rappresenti una stima opportuna per le retribuzioni medie dell'unità da imputare.

nibili dopo un anno, sia a livello di singola unità, sia di aggregato, con l'obiettivo di valutare l'errore di revisione ed individuare aspetti critici del processo di stima su cui eventualmente intervenire. A parità di trimestre, il dato di stima finale viene confrontato con il dato di stima provvisoria pre-imputazione e con quello di stima provvisoria post-imputazione. In questo modo è possibile sia effettuare una valutazione sull'errore di imputazione, sia quantificare l'errore di revisione.

Nel grafico che segue vengono messi a confronto, per le posizioni lavorative totali, i dati di stima provvisoria pre e post-imputazione e rappresentati i rispettivi errori di revisione, prima e dopo l'intervento di imputazione, per il periodo esemplificativo che va dal II trimestre 2012 al II trimestre 2015. La percentuale di imputazione si aggira, in media, intorno al 4,6% con picchi nel IV trimestre 2013 (10,4%) e nel I trimestre 2014 (10%), a causa dell'assenza di alcune grandi unità per ritardo d'invio delle dichiarazioni contributive. Un'imputazione così rilevante è la conseguenza di elevati tassi di mancata risposta che, in media, si aggirano intorno al 5%. In generale, gli errori di revisione sono molto contenuti (0,3% in media) e frequentemente di segno positivo, evidenziando la persistenza di una lieve sottostima dei dati provvisori anche dopo l'imputazione, ma confermando l'importanza di questa operazione (grafico 6.3.1).

Grafico 6.3.1 – Posizioni lavorative dipendenti delle agenzie di somministrazione: stima provvisoria pre e post imputazione, tasso di mancata risposta, percentuale di imputazione ed errore di revisione
Il trimestre 2012 – Il trimestre 2015, valori assoluti (asse sx) e valori percentuali (asse dx)



Tali analisi, svolte routinariamente anche per le altre variabili sottoposte a ricostruzione, hanno consentito nel tempo di perfezionare il metodo di stima provvisoria, introducendo piccole varianti. Inoltre, man mano che la situazione informativa è andata stabilizzandosi, dopo un periodo di transizione successivo all'avvento dell'Uniemens prima del DM2013 e successivamente, l'applicabilità della procedura è stata rivalutata. Infatti, l'analisi degli errori di revisione sulle variabili retribuzioni e oneri per Ula ha messo in evidenza che le stime post-imputazione risultavano meno accurate di quelle pre-imputazione. Tale evidenza ha portato, a partire dal rilascio di marzo 2015 (stima provvisoria del IV trimestre 2014), a calcolare l'imputazione per tutte le variabili in ogni trimestre, ma limitare l'uso del dato imputato su retribuzioni e oneri per Ula ai casi di mancata risposta particolarmente rilevante. Tali casi, nei dati di fonte Inps, non sono insoliti, per cause sia di natura amministrativa, sia

dovute a revisioni nelle procedure informatiche dell'ente fornitore. La validazione del dato imputato viene effettuata confrontando in serie storica le stime con e senza dato ricostruito e scegliendo l'alternativa che offre un profilo più stabile e/o in linea con la tendenza della stessa variabile osservata in altre sottopopolazioni.

La precisione delle stime provvisorie delle posizioni lavorative delle unità interinali viene, inoltre, valutata nell'insieme delle unità *target* della rilevazione, attraverso un approccio di scomposizione dell'errore per componenti. Tale argomento e i relativi risultati saranno esposti nel paragrafo 6.4 che segue.

6.4 Analisi dell'errore di revisione per componenti: approccio teorico

La situazione informativa alla base della rilevazione Oros consente di valutare, per ogni istante di stima provvisoria e ad un anno di distanza rispetto alla prima stima, l'errore di revisione effettivo ossia lo scostamento della stima preliminare rispetto a quella definitiva. Ciò è possibile a seguito della disponibilità di dati finali sulla popolazione di riferimento, non solo rispetto ai dati amministrativi ma anche a quelli d'indagine⁹. La possibilità di calcolare tale errore costituisce per la rilevazione Oros un'opportunità per misurare l'accostamento della stima provvisoria a quella finale e, quindi, per valutare la qualità in termini di accuratezza dei risultati dell'indagine. Tali errori possono essere calcolati a partire dallo scostamento tra micro dati consentendo di risalire alla causa che ha determinato l'errore. Un'analisi dell'errore per causa ha un potenziale informativo enorme, permettendo di intervenire sul metodo di stima provvisoria da dati amministrativi in maniera molto mirata con il fine di migliorarne le *performance*.

Considerato che la stima relativa alle posizioni lavorative dipendenti medie mensili (Y) può essere ottenuta per enumerazione dei dati stimati ottenuti da dati d'indagine per le GI e dati amministrativi rispettivamente per le PMI, GI_DA (sottopopolazioni trattate con stessa metodologia) e le INTER (per cui, come visto nel §6.3, viene adottata una metodologia specifica), si ha:

$$Y = Y_{GI} + Y_{PMI+GI_DA} + Y_{INTER} \quad (4)$$

in cui la stima preliminare (SP) e la stima finale (SD) possono essere espresse come:

$$Y^{SD} = Y_{GI}^{SD} + Y_{PMI+GI_DA}^{SD} + Y_{INTER}^{SD} \quad \text{e} \quad Y^{SP} = Y_{GI}^{SP} + Y_{PMI+GI_DA}^{SP} + Y_{INTER}^{SP} \quad (5)$$

L'errore relativo totale di revisione può essere scomposto nella parte dovuta alle tre sottopopolazioni:

$$e = \frac{(Y_{GI}^{SP} - Y_{GI}^{SD}) + (Y_{PMI+GI_DA}^{SP} - Y_{PMI+GI_DA}^{SD}) + (Y_{INTER}^{SP} - Y_{INTER}^{SD})}{Y^{SD}} \quad (6)$$

9 Nella rilevazione Oros, le revisioni vengono effettuate per incorporare negli indicatori le informazioni che si rendono disponibili successivamente alla pubblicazione della prima stima. I principali elementi considerati nel processo di revisione sono i seguenti:

- la disponibilità dell'universo delle dichiarazioni DM2013 virtuali per la produzione della stima definitiva;
- la revisione dei dati dell'indagine mensile GI;
- l'aggiornamento di informazioni di carattere strutturale sulle unità oggetto di rilevazione;
- le eventuali revisioni occasionali nella metodologia di stima degli indicatori.

e riscritta in breve risulta uguale a:

$$e = GIe' + PMI+GI_DAE' + INTERe' \quad (7)$$

in cui, per la popolazione GI è per esempio : $e'_{GI} = \frac{(GIY^{SP} - GIY^{SD})}{Y^{SD}}$

e in modo analogo vale la stessa relazione per le altre due sottopopolazioni, PMI_DA e INTER. In questo modo è possibile esprimere l'errore relativo totale come somma di componenti che rappresentano il contributo all'errore riferito a ciascuna sottopopolazione. Ogni componente, definita come e' , è a sua volta uguale al prodotto tra l'errore relativo (e) riferito alla singola sottopopolazione e il corrispondente peso occupazionale (in simboli π , che per la popolazione delle GI, è uguale a $\pi_{GI} = \frac{GIY^{SD}}{Y^{SD}}$) calcolato con dati di stima definitiva e riferito alle unità caratterizzate da questo errore. In formula si ha:

$$e = PMI+GI_DAE' + GIe' + INTERe' = PMI+GI_DAE \pi_{PMI+GI_{DA}} + GIe \pi_{GI} + INTERe \pi_{INTER} \quad (8)$$

Le successive espressioni possono ancora essere espresse in termini di componente dell'errore.

Focalizzando sulle cause di revisione delle stime ottenute utilizzando la fonte amministrativa e isolando la componente dovuta alle interinali (componente PMI e GI_DA), l'errore può essere focalizzato rispetto a tale sottopopolazione, e scomposto come:

$$PMI+GI_DAE = PMI+GI_DAE'^{Risp} + PMI+GI_DAE'^{NoRisp} + PMIe'^R \quad (9)$$

in cui gli apici delle componenti d'errore, *Risp*, *NoRisp* e *R*, indicano rispettivamente l'errore nei microdati relativi ai rispondenti nelle stime provvisorie e definitive, l'errore relativo alle mancate risposte e, infine, un errore residuale, di entità trascurabile, legato a riclassificazioni delle unità tra stima provvisoria e stima definitiva (es. i *mismatch* di vario genere tra dati amministrativi e dati d'indagine).

In particolare per le prime due componenti si ha che la prima, riferita ai rispondenti, rappresenta la differenza nel dato disponibile a parità di unità in SP e SD. Tale errore include eventuali correzioni dei dati da parte dell'Inps, eventuali rettifiche nella modalità di trattamento del dato nella rilevazione (errori di processo) o disponibilità di metadati più aggiornati che portano ad una diversa interpretazione del dato amministrativo nelle due versioni delle stime.

La seconda componente della revisione invece, riferita all'insieme delle unità assenti in SP, si può a sua volta scindere nella parte dell'errore dovuta all'individuazione della lista di rispondenti ritardatari (*NoRisp, L*) e quella causata dal processo d'imputazione (*NoRisp, Imp*), ossia:

$$PMI+GI_DAE'^{NoRisp} = PMI+GI_DAE'^{NoRisp, L} + PMI+GI_DAE'^{NoRisp, Imp} \quad (10)$$

L'ultimo termine indica l'errore dovuto all'imputazione dei microdati per le unità correttamente trattate come mancate risposte, ossia l'errore dovuto alla metodologia di ricostruzione del dato mancante, mentre l'errore che riguarda l'individuazione della lista dei rispondenti ritardatari (*L=late reporter*) può essere ancora scomposto in:

$$PMI+GI_DAE'^{NoRisp, L} = PMI+GI_DAE'^{NoRisp, Oc} + PMI+GI_DAE'^{NoRisp, Uc} \quad (11)$$

Il primo termine misura la sovra inclusione dovuta all'imputazione di unità non attive ($Oc = overcoverage$), il secondo termine esprime, invece, la sotto inclusione di rispondenti ritardatari trattati erroneamente come inattivi ($Uc = undercoverage$).

Questo approccio, che consente di scendere nel merito delle singole fonti di errore e di focalizzare meglio sulle criticità, è praticabile se è possibile disporre di informazioni dettagliate e controllabili a livello di microdato. Tali componenti d'errore sono calcolate per divisione Ateco e sue successive aggregazioni per ogni trimestre di stima definitiva, sia per creare serie storiche degli errori di revisioni e delle sue cause e/o componenti e sia come misure di sintesi dell'intero periodo attraverso delle statistiche, aggiungendo ogni volta nel periodo considerato l'ultima stima definitiva disponibile. In particolare, quali misure sintetiche degli errori di revisione sono normalmente considerate la media dei valori assoluti e dei valori in segno (che ne evidenziano ampiezza e direzione), la mediana (che non risente dei valori estremi) e la deviazione standard, insieme a revisioni minime e massime (che ne misurano la variabilità).

Le analisi effettuate nel corso della sperimentazione e i controlli routinari evidenziano una buona qualità delle stime provvisorie delle posizioni lavorative secondo la metodologia delineata in Oros con errori che, sull'aggregato B-S, da -2% circa pre-imputazione, scendono a -0,2%. In termini relativi, la discrepanza prevalente rispetto alle stime finali è concentrata nelle PMI e GI_DA, che rivestono anche il peso maggiore in termini di occupazione, ed è principalmente da attribuirsi al criterio di individuazione della lista (circa -0,15%): pur registrando un buon bilanciamento tra sovra inclusione e sotto inclusione nella lista, si osserva una leggera prevalenza della prima tipologia di errore. La caratteristica inerziale che contraddistingue l'evoluzione delle posizioni lavorative dipendenti in molti settori implica, invece, una buona rappresentazione del fenomeno attraverso la regressione utilizzata per la ricostruzione dei dati mancanti sulle PMI e GI_DA, con una lieve tendenza a sottostimare i livelli dell'occupazione finale (-0,5%). Come sintesi delle due fonti di errore, emerge una leggera prevalenza dell'errore di lista rispetto a quello dovuto alla regressione. Tale errore appare più rilevante nei settori in cui sono più frequenti eventi di demografia d'impresa (es. costruzioni). L'errore dovuto alla regressione è, invece, rilevante in settori in cui la modellizzazione è resa meno agevole dalla volatilità dell'occupazione sottostante (es. settori della produzione cinematografica e televisiva): per questi casi si continua a fare ricorso ad aggiustamenti sui macro dati basati sull'analisi in serie storica (cfr. §8.4.4). Rimane di totale impatto, anche se di entità molto ridotta, l'errore di sottostima causato dalle unità assenti dalla lista anagrafica di partenza (neo-nate che rispondono con ritardo prima di iscriversi oppure unità sospese da oltre un anno e quindi tagliate dalla lista delle potenziali attive), come già evidenziato nel §6.2.3.

6.5. Imputazione nei dati finali

6.5.1 Le mancate risposte nei dati finali

Pur acquisiti dopo un anno dalla scadenza ufficiale per l'invio delle dichiarazioni contributive da parte dei datori di lavoro (cfr. §2.2.1), anche i dati finali possono essere caratterizzati da dichiarazioni mancanti relative ad unità attive. Il fenomeno ha una rilevanza molto

ridotta ed è legato a motivazioni di gestione amministrativa da parte dell'Inps. Si tratta, nello specifico, di una quantità di DM definiti "provvisori o anomali" poiché derivanti da flussi Uniemens per cui i *check* dell'Inps sul complesso della matricola aziendale riscontrano mancata quadratura sulla somma di alcune variabili. Tali dichiarazioni mancanti sono comparse a seguito del passaggio al DM2013 virtuale, a novembre 2012. In precedenza, finché avveniva la ricostruzione del DM10 virtuale, era attiva anche una procedura di quadratura forzata avviata dopo massimo 2-3 mesi dall'arrivo delle dichiarazioni, che comportava l'eliminazione dello *status* di "provvisorietà" e l'immissione nel DataWarehouse Inps, archivio da cui vengono selezionati i microdati acquisiti per Oros. Per effetto di un successivo intervento informatico da parte dell'Inps, tali matricole sono state gradualmente riassorbite nel DataWarehouse, a seguito di controlli e successiva forzatura: a partire dalla competenza di novembre 2015 per i dati anticipati e ottobre 2014 per i dati finali, un insieme sostanziale di matricole anomale è rientrato nel flusso corrente dei dati per Oros.

Al fine di monitorarne l'impatto sulle stime finali e correggerne l'effetto di distorsione, con cadenza annuale viene richiesta all'Inps la lista di tali DM provvisori/anomali per mese di competenza. Si è osservato che, nel periodo che va dal IV trimestre 2012 – al III trimestre 2014 i DM2013 utili alle stime Oros, sono in media 1.200 per mese, senza una particolare concentrazione sui singoli mesi. Inoltre, nel corso di questi 24 mesi, nel 78% delle matricole vi è solo un DM anomalo e per un restante 11% al massimo due DM anomali ad indicare che non vi è un'inerzia da parte delle imprese a commettere errori nella compilazione della dichiarazione contributiva. Il numero di DM anomali scende drasticamente ad alcune decine di unità nei mesi successivi al III 2014, per effetto della forzatura citata sopra. Fanno eccezione il I trimestre 2015 e il mese di settembre 2015, in cui i DM anomali tornano ad essere circa un migliaio per mese, per motivi non noti. Il numero di matricole interessate è, in media, di poco inferiore a 2 mila unità.

Considerato l'esiguo numero e la scarsa sistematicità nella loro distribuzione tra le unità e per Ateco, è stato valutato trascurabile il loro impatto sulle variabili pro-capite di costo del lavoro, mentre è stata ritenuta da considerare la loro incidenza sulle stime dei livelli delle posizioni lavorative, inducendo ad applicare una procedura di ricostruzione per regressione analoga a quella impiegata nelle stime provvisorie (cfr. §6.2.5). Di seguito viene descritta la metodologia utilizzata per l'imputazione delle posizioni lavorative di queste dichiarazioni mancanti e l'impatto dei dati imputati sulle stime complessive della variabile.

6.5.2 Il metodo di imputazione dei dati mancanti

Il metodo applicato in questo contesto si basa sostanzialmente sullo stesso approccio di imputazione delle posizioni lavorative per le mancate risposte nei dati provvisori delle PMI e GI_DA descritto nel paragrafo 6.2. Come visto, esso si basa sulla risoluzione di due problematiche: 1) la predizione dello *status* di attività delle unità non rispondenti; 2) la successiva imputazione delle posizioni lavorative per le unità presunte attive.

Nella versione finale dei dati, l'imputazione dei dati mancanti dovuti a dichiarazioni provvisorie/anomale, la predizione dello *status* non è necessaria in quanto la lista della matricole interessate da questa problematica fornita dall'Inps rappresenta per definizione un insieme di unità non presenti negli scarichi trimestrali ma certamente attive e rispondenti. Per queste unità, pertanto, è necessario soltanto procedere con la fase di imputazione, sfruttando la modellistica già vista nel §6.2.5. I coefficienti del modello di regressione vengono stimati sui valori della variabile per i rispondenti nei dati finali, ricorrendo agli stessi

domini definiti per la stima provvisoria che, nel caso di dati finali, saranno caratterizzati da un numero maggiore di unità. Pur ipotizzando un'accuratezza di risultati maggiore rispetto a quella che si ottiene nell'imputazione dei dati provvisori, la valutazione dell'errore di revisione sarebbe necessaria. Tuttavia questo errore di revisione, analizzato come visto nel §6.4, non può essere calcolato, poiché le matricole anomale/provvisorie non vengono fatte rientrare dall'Inps nel flusso dei dati inviati all'Istat ai fini di Oros. Le uniche valutazioni di qualità della stima riguardano le *performance* dei modelli di regressione che, considerata la rilevante numerosità delle unità su cui sono stimati, garantiscono comunque capacità predittiva elevata.

Nel periodo che va da fine 2012 a metà 2016, le posizioni lavorative dipendenti imputate nella stima finale pesano per circa lo 0,2% in media sull'occupazione complessiva di Oros, senza alcuna caratterizzazione settoriale o di periodicità.

6.6 Editing selettivo e imputazione

6.6.1 Quadro concettuale

L'estensione del campo di osservazione della rilevazione Oros ai settori dei Servizi privati personali e sociali (sezioni Ateco 2007 da P ad S), la cui serie storica è stata diffusa per la prima volta a dicembre 2013 (cfr. §2.1), ha posto la necessità di rivedere le procedure di controllo e correzione fino ad allora a regime (Baldi et al., 2011), con il fine di monitorare in modo più mirato e consentire il mantenimento di un buon livello nella qualità dei principali indicatori prodotti, ampiamente messo a rischio dalle particolarità delle nuove unità entrate nel campo di osservazione della rilevazione. Innanzitutto, va ricordato che per questi settori non è possibile utilizzare dati dell'indagine GI, la cui copertura è attualmente limitata ai settori da B a N¹⁰, con la conseguenza che i dati relativi ad imprese di grandi dimensioni non sono sottoposti alle specifiche procedure di controllo, correzione e imputazione attuate nel processo di produzione GI. Inoltre, a partire dai dati del primo trimestre 2015, il flusso dei dipendenti iscritti alla gestione ex-Enpals è confluito in quello ordinario degli Uniemens (e quindi DM2013). In questo insieme informativo sono presenti unità con caratteristiche di costo del lavoro peculiari che si discostano da quelle delle altre imprese¹¹ del settore in cui sono classificate e anche rispetto agli altri settori. Eventi demografici o mancate risposte delle unità ex-Enpals possono produrre effetti di distorsione di entità diversa a seconda che si considerino variabili di livello, rapporto o variazioni relative. In aggiunta, il costante miglioramento della base dati amministrativa ha reso la procedura di micro *editing* selettivo utilizzata precedentemente e finalizzata alla correzione dei soli errori di misura (Baldi et al., 2008), poco utile oltre che onerosa in termini di tempo. Tale evidenza ha posto la necessità di implementare un metodo accurato, efficiente e allo stesso tempo *user-friendly* progettato per rispondere alle nuove esigenze e, in particolare, finalizzato all'individuazione e correzione di errori e mancate risposte particolarmente influenti, in un'ottica di ulteriore riduzione dei tempi di produzione e rilascio degli indicatori Oros.

¹⁰ Si ricorda che le sezioni da P ad S rientreranno nel panel dell'indagine GI a partire dal 2018, con la transizione alla base 2015.

¹¹ Ne sono un esempio i valori particolarmente elevati e variabili nel tempo delle retribuzioni pro-capite relative alle squadre di calcio, casinò, ecc. classificate nella sezione R Ateco 2007.

Le variabili *target* per la nuova procedura di *editing* selettivo sono la variazione tendenziale delle retribuzioni per Ula, degli oneri per Ula e delle posizioni lavorative dipendenti. La scelta di questa variabile è strettamente connessa alla natura stessa del processo Oros che, essendo una rilevazione di tipo congiunturale, ha come principale obiettivo la stima accurata delle variazioni nel tempo degli indicatori prodotti (anche se indici e livelli sono, comunque, statistiche di interesse). Inoltre, il controllo delle variazioni tendenziali (e non congiunturali) ha permesso una identificazione più accurata degli *outlier* che incorporano la stagionalità tipica delle variabili di costo del lavoro (si pensi ad esempio al secondo e quarto trimestre in cui vengono erogate la quattordicesima e tredicesima mensilità) contribuendo a ridurre il numero di unità influenti segnalate.

Il metodo consiste nel selezionare le unità sulla base di due funzioni di influenza applicate a ciascuna delle tre variabili *target*: una prima funzione finalizzata alla selezione di unità con dati potenzialmente errati e una seconda funzione finalizzata alla selezione di unità con mancate risposte influenti e potenzialmente imputabili che selezionano insiemi di unità non coincidenti e non necessariamente sovrapposti. La scelta delle funzioni di influenza è ricaduta su funzioni intuitive e facilmente implementabili che caratterizzano la procedura di *editing* come metodo di tipo deterministico (De Waal et al., 2011; Eurostat, 2014). Per le unità segnalate da entrambe le funzioni si effettua una prima valutazione della qualità del dato considerando, successivamente, l'eventualità del ricorso ad un'imputazione con metodo generalizzato oppure di un'imputazione *ad hoc*.

Questa procedura viene utilizzata per validare sia le stime provvisorie, sia le stime finali. Per queste ultime, tuttavia, la valutazione è limitata alle variabili di costo del lavoro e non interessa le posizioni lavorative. Come visto nei precedenti paragrafi di questo capitolo, infatti, le stime finali non sono interessate da mancate risposte, per cui le procedure di *editing* selettivo sono principalmente finalizzate a evidenziare *outlier* che, normalmente, data la qualità dei dati Inps, non interessano le posizioni lavorative.

6.6.2 Le funzioni di influenza: funzione finalizzata all'editing

La funzione finalizzata alla procedura di *editing* è rappresentata dal valore assoluto della differenza tra la variazione tendenziale della variabile relativa al settore di attività a cui l'unità in esame appartiene e la stessa variazione calcolata escludendo dall'insieme la singola unità oggetto di *editing*.

Per ogni unità k , appartenente all'aggregato S (sezione Ateco 2007) la funzione di influenza scelta è espressa tramite una funzione indicatrice nel modo seguente:

$$Infl1(k) = I_{[\tau, +\infty)} [|V(X_S) - V(X_{S-k})|] \quad (12)$$

dove S^k rappresenta l'aggregato S con l'esclusione dell'unità k -esima, $V(X)$ rappresenta la variazione tendenziale calcolata sull'aggregato specifico per la generica variabile X che può assumere il valore RU (Retribuzione media per Ula), OU (Oneri medi per Ula), D (Posizioni lavorative dipendenti). Infine, τ rappresenta la soglia di selezione delle unità influenti scelta sulla base di criteri che tengono conto del *trade-off* tra numero di unità selezionate e qualità dell'output prodotto¹². Perché una unità sia segnalata come influente dalla funzione

¹² Le evidenze tratte dall'analisi sui dati hanno portato a scegliere un valore pari a 0,2 come soglia per tutte le funzioni di influenza calcolate.

indicatrice è necessario che il valore della differenza tra variazione totale e variazione parziale, in modulo, sia maggiore o uguale alla soglia τ . La (12) può assumere le due seguenti realizzazioni:

$$\begin{aligned} \text{se } |V(X_S) - V(X_{S-k})| &\geq \tau \Rightarrow \text{Infl1}(k) = 1 && \text{l'unità } k \text{ è influente} \\ \text{se } |V(X_S) - V(X_{S-k})| &< \tau \Rightarrow \text{Infl1}(k) = 0 && \text{l'unità } k \text{ non è influente} \end{aligned} \quad (13)$$

È importante sottolineare che la popolazione di riferimento è variabile nel tempo a causa dei fenomeni di demografia d'impresa dovuti a nascite e cessazioni/sospensioni, trasformazioni giuridiche, ecc. che comportano l'entrata e l'uscita dall'insieme di unità affette da queste casistiche¹³. La funzione di influenza, implicitamente, tiene conto del contributo di ogni singola unità al valore della variazione tendenziale del settore di appartenenza, contributo che dipende dall'indicatore analizzato e dalle sue proprietà in termini di componenti caratterizzanti. In particolare, come viene mostrato in seguito, per quanto riguarda il caso più semplice di variazioni calcolate su variabili espresse nei livelli, il contributo della singola unità alla variazione del settore è funzione della micro variazione tendenziale, mentre per quanto riguarda le variazioni calcolate su variabili espresse come rapporti il contributo non dipende soltanto dalla variazione tendenziale dell'unità ma anche da un effetto di ricomposizione occupazionale non direttamente osservabile, come conseguenza di aggregati i cui indici sono calcolati come indici semplici di variabili rapporto (cfr. Lattanzio, 2016a e Lattanzio, 2016b).

Per semplicità di trattazione, la formalizzazione presentata sarà riferita solo al sottoinsieme delle unità panel della popolazione (in seguito insieme P), appartenente al generico aggregato S , per le quali si dispone dei dati sia a t sia a $t-4$.

L'indice semplice delle posizioni lavorative può essere espresso in funzione dei micro indici componenti nel modo seguente:

$$I_{P,t}^D = \sum_{i \in P} I_{i,t}^D \pi_{i,P,b}^D \quad (14)$$

ove $I_{i,t}^D$ rappresenta l'indice delle posizioni lavorative dipendenti dell' i -esima unità e $\pi_{i,P,b}^D$ rappresenta il peso in termini di posizioni lavorative riferite all'anno base b rispetto all'aggregato P delle unità panel (cfr. §9.2.2, formula (5) per la formalizzazione estesa). In base alla formulazione relativa alla scomposizione delle variazioni (Lattanzio, 2016a), per l'indice di cui sopra vale la seguente relazione:

$$V(D_P) = \sum_{i \in P} V(d_i) \pi_{i,P,t-4}^D \quad (15)$$

Dove $V(d_i)$ rappresenta la micro variazione tendenziale delle posizioni lavorative dell'unità i -esima e $\pi_{i,P,t-4}^D$ è il peso delle posizioni dell'unità i -esima (d_i) riferite al trimestre $t-4$ e calcolato rispetto all'aggregato P (D_P). A partire da questa formulazione, la differenza di cui alla funzione indicatrice (12), per la k -esima unità è esplicitabile nel modo seguente:

¹³ Nel caso particolare in cui l'unità k -esima sia nuova nata o cessata essa contribuisce al valore della variazione tendenziale relativa all'aggregato S solo con i dati rispettivamente relativi al tempo t e $t-4$.

$$\begin{aligned}
V(D_P) - V(D_{P-k}) &= \sum_{i \in P} V(d_i) \pi_{i,P,t-4}^D - \sum_{i \in P-k} V(d_i) \pi_{i,P-k,t-4}^D \\
&= V(d_k) \pi_{k,P,t-4}^D + \sum_{i \neq k} V(d_i) (\pi_{i,P,t-4}^D - \pi_{i,P-k,t-4}^D) = \\
&= V(d_k) \pi_{k,P,t-4}^D + \sum_{i \neq k} V(d_i) \left(\frac{D_{i,t-4}}{D_{P,t-4}} - \frac{D_{i,t-4}}{D_{P-k,t-4}} \right) = \\
&= V(d_k) \pi_{k,P,t-4}^D - \sum_{i \neq k} V(d_i) \frac{D_{i,t-4}}{D_{P-k,t-4}} \left(\frac{D_{P,t-4} - D_{P-k,t-4}}{D_{P,t-4}} \right) = \\
&= V(d_k) \pi_{k,P,t-4}^D - \pi_{k,P,t-4}^D \sum_{i \neq k} V(d_i) \pi_{i,P-k,t-4}^D = \\
&= \pi_{k,P,t-4}^D \left(V(d_k) - \sum_{i \neq k} V(d_i) \pi_{i,P-k,t-4}^D \right) = \\
&= \pi_{k,P,t-4}^D (V(d_k) - V(D_{P-k})) \tag{16}
\end{aligned}$$

dove P^k indica l'insieme delle unità panel con esclusione della k -esima unità. La (16) esprime la differenza tra le variazioni tendenziali delle posizioni lavorative dipendenti relative ai due aggregati P e P^k come scostamento pesato tra la variazione tendenziale dell'unità k -esima e la variazione media dell'insieme delle unità ad essa complementare. Questa formulazione, esplicitando l'effetto del contributo della singola unità alla variazione dell'aggregato, consente di identificare come influenti quelle unità le cui variazioni sono tanto più distanti dalla variazione media delle altre, fornendo una misurazione statistica di variabilità.

Per quanto riguarda il caso specifico di *editing* selettivo sulle variabili espresse in termini di rapporti (come le variabili relative al costo del lavoro per Ula), l'espressione che consente di esplicitare il contributo dell'unità alla variazione totale dell'aggregato ha una forma di interpretazione meno intuitiva, a causa della difficoltà di isolare gli effetti di ricomposizione occupazionale relativi all'unità k -esima. Nel caso particolare delle Retribuzioni per Ula, riprendendo i risultati presentati nei lavori Lattanzio, 2016a e 2016b, la scomposizione relativa alla variazione tendenziale dell'indice semplice dell'aggregato P è data dall'espressione seguente:

$$V(RU_P) = \sum_{i \in P} \left(V(ru_i) + V(\pi_{i,P}^U) + \frac{V(ru_i)V(\pi_{i,P}^U)}{100} \right) \pi_{i,P,t-4}^R \tag{17}$$

dove $\pi_{i,P}^U$ è il peso relativo delle Ula dell'unità i -esima rispetto all'aggregato P e $\pi_{i,P,t-4}^R$ è il peso relativo del monte retributivo dell'unità i -esima rispetto all'aggregato P al tempo $t-4$, indicando con:

$$\tilde{V}(i) = V(ru_i) + V(\pi_{i,P}^U) + \frac{V(ru_i)V(\pi_{i,P}^U)}{100} \tag{18}$$

i contributi micro alla variazione dell'unità i -esima. La (17) può essere espressa come segue:

$$V(RU_P) = \sum_{i \in P} \tilde{V}(i) \pi_{i,P,t-4}^R \tag{19}$$

La (19) consente di esprimere la variazione tendenziale media relativa all'aggregato P come una media ponderata di contributi micro alla variazione, che non dipendono soltanto dalla variazione tendenziale delle retribuzioni per Ula della singola unità ma anche dagli effetti di ricomposizione occupazionale citati in precedenza.

E' possibile ottenere una formalizzazione analoga alla (16), maggiormente finalizzata all'*editing*, (con riferimento al contributo dell'unità k -esima) della differenza tra variazione totale e parziale. Sempre secondo la logica di partizionamento della popolazione a cui si sta facendo riferimento, si può esprimere la variazione totale come media ponderata dei contributi relativi all'insieme parziale e alla singola unità:

$$\begin{aligned}
 V(RU_P) &= \\
 &= \left(V(RU_{P-k}) + V(\pi_{P-k,P}^U) + \frac{V(RU_{P-k})V(\pi_{P-k,P}^U)}{100} \right) \pi_{P-k,P,t-4}^R \\
 &\quad + \left(V(ru_k) + V(\pi_{k,P}^U) + \frac{V(ru_k)V(\pi_{k,P}^U)}{100} \right) \pi_{k,P,t-4}^R
 \end{aligned} \tag{20}$$

Date la (18) e la (19) segue che:

$$\begin{aligned}
 V(RU_P) - V(RU_{P-k}) &= \\
 &= \tilde{V}(k) \pi_{k,P,t-4}^R - V(RU_{P-k}) \left(1 - \pi_{P-k,P,t-4}^R \right) + \left[V(\pi_{P-k,P}^U) \left(1 + \frac{V(RU_{P-k})}{100} \right) \right] \pi_{P-k,P,t-4}^R \\
 &= [\tilde{V}(k) - V(RU_{P-k})] \pi_{k,P,t-4}^R + g(k)
 \end{aligned} \tag{21}$$

La relazione (21) così ottenuta è costituita da due addendi: il primo, similmente alla relazione (16) ottenuta per le posizioni lavorative, rappresenta uno scostamento pesato dalla media dei contributi delle altre unità, non della micro variazione ma del contributo della singola unità k -esima alla variazione totale che dipende anche dalla ricomposizione occupazionale (definiti analogamente alla (18) e alla (19)); il secondo, è un termine residuale, anch'esso funzione della ricomposizione occupazionale tra gli insiemi totale e parziale, che sarà oggetto di approfondimenti futuri finalizzati ad una più accurata definizione del valore della soglia τ .

Le unità selezionate attraverso la procedura descritta sono in numero relativamente ridotto, non superando mai le 100 unità (rispetto a 1 milione e 200 mila circa unità totali sottoposte ad analisi). Queste unità sono valutate attraverso un'analisi interattiva molto accurata allo scopo di individuare quelle a cui apportare la correzione appropriata. Il criterio generalmente seguito è quello del minimo intervento, così come raccomandato in letteratura (cfr. AA.VV., 2012), per limitare l'introduzione di valori errati (*over editing*). In particolare, i controlli effettuati trimestralmente hanno evidenziato come la maggior parte dei casi che presentano valori *outlier* nelle variabili considerate siano riconducibili a trasformazioni giuridiche o a lineamenti temporali nella corresponsione di poste retributive e, quindi, da non sottoporre ad alcun tipo di correzione. D'altra parte, come accennato in precedenza, la procedura ha consentito di selezionare le unità appartenenti al flusso ex-Enpals. Tra queste si è deciso di trattare in modo opportuno soltanto un sotto insieme caratterizzato da valori

outlier ricorrenti ma non periodici¹⁴. La correzione apportata è consistita nella sostituzione del dato medio trimestrale corrente relativo a tali unità con quello del medesimo trimestre dell'anno precedente, intervento che ha permesso di sterilizzare la variazione tendenziale media del settore di appartenenza dagli effetti spuri introdotti da queste unità.

6.6.3 Le funzioni di influenza: funzione finalizzata all'imputazione

In questo paragrafo si presenta la funzione finalizzata alla selezione di eventuali unità influenti da sottoporre ad imputazione di mancate risposte. La necessità di procedere all'imputazione (che non era tra le finalità della precedente procedura di controllo e correzione) emerge da un'attenta analisi dei diversi fattori che intervengono nel calcolo di un indice aggregato relativo a variabili espresse nei rapporti; con particolare riferimento alla variabile retribuzioni per Ula, l'indice aggregato, la cui trattazione estesa è presentata nel paragrafo 9.2.2, può essere espresso dalla seguente espressione:

$$I_{S,t}^{RU} = \sum_{j=1}^s I_{j,t}^{RU} I_{j,t}^{\pi_U} \frac{RU_{j,b} \pi_{j,b}^U}{RU_{S,b}} \quad (22)$$

La relazione evidenzia che, sebbene il calcolo di variabili pro-capite ($I_{j,t}^{RU}$) possa ridurre, in molti casi, la sensibilità delle variabili in relazione all'effetto delle mancate risposte (che riguardano sia il numeratore sia il denominatore, cfr. §6.2.3)¹⁵, d'altra parte lo stesso non può dirsi per il fattore di ricomposizione occupazionale ($I_{j,t}^{\pi_U}$) il cui effetto, influenzato dalla presenza delle mancate risposte, risulta essere sistematicamente sottostimato (Lattanzio, 2016a e 2016b).

Come premesso all'inizio di questo paragrafo, la procedura di controllo e correzione implementata prevede il calcolo di una funzione di influenza per ciascuna delle tre variabili retribuzioni pro-capite, oneri pro-capite e posizioni lavorative dipendenti. Nel caso specifico relativo alla selezione di unità influenti potenzialmente imputabili, occorre specificare che, sebbene le funzioni di influenza individuate per le tre variabili in questione siano analoghe, le finalità che esse si prefiggono sono sostanzialmente diverse; in particolare, mentre per le variabili di costo l'obiettivo finale è l'imputazione delle mancate risposte influenti, per le posizioni lavorative dipendenti, le unità selezionate (le più influenti) sono sottoposte ad un controllo interattivo finalizzato alla validazione del metodo di imputazione da modello applicato in precedenza (cfr. §6.2 e §6.3).

Per ogni unità k affetta da mancata risposta, appartenente all'aggregato S (sezione Ateco 2007), la funzione di influenza scelta per la generica variabile X (così come definita nel §6.6.2) è calcolata come differenza tra la variazione tendenziale relativa all'aggregato in cui l'unità è stata imputata ($\hat{V}(X_S)$) e la stessa variazione in cui l'unità in questione non è stata imputata ($V(X_S)$), espressa dalla seguente funzione indicatrice:

$$Infl2(k) = I_{[\psi, +\infty)}[|\hat{V}(X_S) - V(X_S)|] \quad (23)$$

¹⁴ Si tratta, nello specifico delle squadre di calcio (Divisione 93 della classificazione Ateco 2007) che presentano *outlier* nei valori retributivi per effetto dei compensi molto elevati corrisposti ai calciatori.

¹⁵ Ciò non si verifica nel caso di mancate risposte caratterizzate da periodicità infra annuale (ad esempio la tredicesima e la quattordicesima mensilità) la cui presenza può avere effetti anche sugli indici relativi ai valori pro-capite.

La scelta di questa funzione permette di valutare il contributo dell'imputazione di una mancata risposta sulla variazione tendenziale dell'aggregato considerato.

In particolare per la generica unità k , sia $X_{k,m}$ la variabile da imputare relativa al mese m (con $m=1,2,3$) del trimestre t sia $mr_{k,t}$ il numero di mancate risposte da imputare nel trimestre t , il valore imputato della variabile ($\hat{X}_{k,m}$) è stato calcolato, nel caso in cui l'unità sia rispondente in almeno un mese nel trimestre, secondo il criterio dell'imputazione con media, ossia:

$$\hat{X}_{k,m} = \bar{X}_{k,t}^R \quad \text{con} \quad m \notin R \quad (24)$$

$$\bar{X}_{k,t}^R = \frac{1}{3 - mr_{k,t}} \sum_{m \in R} X_{k,m} \quad \text{con} \quad mr_{k,t} = 1, 2$$

dove R indica l'insieme dei mesi di risposta e $\bar{X}_{k,t}^R$ è la media dei dati relativi ai soli rispondenti. Ne consegue che la media trimestrale imputata della variabile, $\hat{\bar{X}}_{k,t}$, è pari a:

$$\begin{aligned} \hat{\bar{X}}_{k,t} &= \frac{\sum_{m \in R} X_{k,m} + mr_{k,t} \left(\frac{1}{3 - mr_{k,t}} \sum_{m \in R} X_{k,m} \right)}{3} = \frac{1}{3} \frac{3(\sum_{m \in R} X_{k,m})}{3 - mr_{k,t}} \\ &= \frac{1}{3 - mr_{k,t}} \sum_{m \in R} X_{k,m} = \bar{X}_{k,t}^R \end{aligned} \quad (25)$$

La (25) mostra come il valore medio trimestrale ottenuto a seguito dell'imputazione coincida con il dato medio mensile calcolato sui mesi non affetti da mancata risposta. Va osservato, tuttavia, che il valore così imputato risente del metodo di imputazione scelto. Questa evidenza potrebbe rappresentare una criticità nella selezione di unità con mancate risposte in trimestri caratterizzati da valori *outlier* periodici (stagionali). Sebbene il metodo garantisca la selezione di unità imputate con media calcolata su almeno un mese *outlier*, ciò non è garantito per le unità in cui l'imputazione avviene attraverso la media di dati non *outlier*, per le quali sarà la variazione del fattore di ricomposizione occupazionale a determinare o meno la selezione¹⁶.

Nel caso in cui vi sia mancata risposta totale in tutti i mesi del trimestre (diverso dal caso di mancata risposta totale nei record, si ricordi che Oros non presenta casi di mancata risposta parziale nello stesso record) l'imputazione viene effettuata con il dato medio trimestrale dello stesso trimestre dell'anno precedente¹⁷:

$$\hat{\bar{X}}_{k,t} = \bar{X}_{k,t-4} \quad \text{con} \quad mr_{k,t} = 3 \quad (26)$$

Le unità con mancate risposte influenti così selezionate sono sottoposte ad analisi interattive finalizzate all'individuazione di quelle che necessitano correzioni *ad hoc*, come i casi già citati di unità con *outlier* non periodici nel trimestre. Le restanti unità sono anch'es-

¹⁶ Si sta valutando l'opportunità di affinare la metodologia per tenere adeguatamente conto di tali specificità.

¹⁷ Per completezza informativa si precisa che il dato medio trimestrale dello stesso trimestre dell'anno precedente potrebbe essere stato calcolato a sua volta secondo la (24).

se valutate caso per caso per stabilire l'opportunità di effettuare o meno un'imputazione completa (in tutti i mesi che presentano la mancata risposta) o parziale che preservi la presenza di eventuale periodicità e, in caso affermativo, sono successivamente imputate con procedure generalizzate basate sui criteri deterministici espressi dalla (24) e dalla (25).

Infine, è utile ricordare che, come accennato in precedenza (cfr. §6.6.1), essendo le due funzioni di influenza finalizzate all'*editing* e all'imputazione delle mancate risposte totali indipendenti tra loro, sono sottoposte al controllo e correzione le unità non ancora imputate per eventuali mancate risposte e viceversa. Nell'evidenza dei dati Oros, gli insiemi delle unità influenti selezionate per i due scopi sono tra loro interdipendenti; infatti, la maggior parte delle unità selezionate per l'imputazione delle mancate risposte è selezionata anche per l'*editing* (a conferma dell'indipendenza dell'influenza di queste unità da una eventuale precedente imputazione); d'altra parte, il sotto insieme di unità selezionate per l'imputazione ma non per l'*editing* include soltanto pochissime unità che presentano valori *outlier* ricorrenti ma non periodici per i quali la procedura prevede un trattamento *ad hoc* che prescinde dal loro grado d'influenza (cfr. nota 14).

7. INTEGRAZIONE CON I DATI DI FONTE INDAGINE. L'INDIVIDUAZIONE DELLA LISTA E LA STIMA DELLE VARIABILI ECONOMICHE¹

7.1 Quadro di riferimento per l'integrazione con i dati di fonte indagine

La rilevazione Oros, come introdotto nei precedenti paragrafi è frutto dell'integrazione di dati amministrativi con dati derivanti dalla rilevazione GI (Istat, 2006). Alla base della rilevazione GI vi è un panel costituito dalle imprese con almeno 500 dipendenti in media nell'anno base 2010 classificate nell'industria e nei servizi di mercato (sezioni da B a N Ateco 2007)².

Il processo di integrazione, condotto a livello di microdati, prevede la sostituzione dei dati amministrativi con dati d'indagine per quelle unità coperte dal panel GI e si basa sulla scelta di criteri di *record-linkage* deterministico e sull'armonizzazione delle variabili delle due fonti. Già dal suo primo impianto, l'integrazione dei dati amministrativi con dati d'indagine è stata una necessità a causa della scarsa rappresentatività delle imprese di grande dimensione tra i rispondenti rapidi usati per la stima preliminare di Oros (cosiddetta fase 1, §2.3.2). Con il passaggio all'invio telematico delle dichiarazioni contributive DM10 (fase 2), l'integrazione con dati d'indagine è divenuta una scelta dettata principalmente dalla qualità delle variabili raccolte e dalla loro corrispondenza ai contenuti statistici richiesti dai regolamenti. La scelta di integrare è risultata opportuna, in particolare, quando il ricorso alla Cig è divenuto un fenomeno largamente diffuso tra le imprese di grande dimensione e la disponibilità delle ore fruite di cassa integrazione (incluse quelle per i lavoratori in Cig nell'intero mese) ha permesso il calcolo di indicatori retributivi al netto della Cig di migliore qualità rispetto a quelli che si sarebbero ottenuti dai soli dati amministrativi. Successivamente, con il passaggio ai DM10 virtuali (fase 3) e DM2013 virtuali (fase 3 bis) in presenza di una crescita di mancate risposte totali nel/i mese/i appartenenti al trimestre di riferimento, l'integrazione con l'indagine GI ha permesso di mantenere alta la qualità delle stime preliminari complessive, considerato il loro impatto sulle stime totali.

Nei processi di stima congiunturale basati su dati amministrativi, il mantenimento di dati d'indagine sulle imprese di grandi dimensioni è ampiamente consigliato, anche a livello europeo. Vi sono diverse esperienze, in tal senso, che suggeriscono la sostituzione di dati amministrativi con dati d'indagine sulle grandi imprese, normalmente censuarie. La complessa struttura che normalmente caratterizza tali imprese è maggiormente controllata in una raccolta diretta di dati, garantendo una maggiore affidabilità delle stime (Maasing et al., 2012).

Nella tavola 7.1.1 si nota come nell'industria e nei servizi di mercato le posizioni lavorative dipendenti dell'indagine GI contano, in media nel 2014, il 21,8% dell'occupazione

¹ Il presente capitolo è a cura di Francesca Ceccato e Eleonora Cimino.

² A partire dal 2015, l'indagine GI ha iniziato a raccogliere i dati delle imprese appartenenti ai settori da P a S sia per contribuire alla produzione degli indicatori sui posti vacanti e sulle ore lavorate, sia per determinare i dati per l'estensione degli indicatori mensili ai settori dei servizi personali e sociali della futura nuova base di riferimento (dal 2018, anno base 2015). Complessivamente nel 2016 nei settori da B a S escluso O le imprese in rilevazione sono circa 1.670.

complessiva nello stesso aggregato. Nei settori di fornitura di energia elettrica, gas, vapore e aria condizionata (sezione D Ateco 2007) e delle attività finanziarie ed assicurative (sezione K Ateco 2007) il loro peso è pari al 66,3%.

Tavola 7.1.1 – Peso delle posizioni lavorative dipendenti nell'industria e nei servizi di mercato della sottopopolazione GI della rilevazione Oros sul totale, per settore di attività economica (a)
Anno 2014, valori medi percentuali

SETTORI ATECO 2007	Peso delle posizioni dipendenti della sottopopolazione GI sul totale
INDUSTRIA E SERVIZI DI MERCATO (B-N)	21,8
Industria (B-F)	17,3
B - Estrazione di minerali da cave e miniere	16,3
C - Attività manifatturiere	19,2
D - Fornitura di energia elettrica, gas, vapore e aria condizionata	65,0
E - Fornitura di acqua; reti fognarie, i gestione dei rifiuti e risanamento	26,8
F - Costruzioni	3,2
Servizi di mercato (G-N)	24,9
G - Commercio all'ingrosso e al dettaglio, riparazione di autoveicoli e motocicli	18,8
H - Trasporto e magazzinaggio	38,5
I - Attività dei servizi di alloggio e di ristorazione	12,6
J - Servizi di informazione e comunicazione	36,0
K - Attività finanziarie ed assicurative	66,3
L - Attività immobiliari	-
M - Attività professionali, scientifiche e tecniche	9,7
N - Noleggio, agenzie di viaggio, servizi di supporto alle imprese	19,0

Fonte: Rilevazione Oros (occupazione, retribuzioni, oneri sociali)
(a) Le posizioni lavorative includono anche la qualifica dirigenti

A seguito dell'integrazione con le imprese GI nella rilevazione Oros, circa il 76 per cento delle posizioni lavorative dipendenti totali risulta concentrato nelle PMI (tavola 7.1.2). Per quanto riguarda le imprese di grandi dimensioni, quasi tutte sono incluse nel panel della rilevazione GI (21,8% dell'industria e servizi di mercato, 20% dell'industria e servizi), quelle residuali che non fanno parte del panel GI (GI_DA) occupano poco meno dell'1 per cento dell'industria e servizi di mercato (1,7% se il peso è calcolato sul totale industria e servizi). Un peso di quasi il 2 per cento è rappresentato dai lavoratori somministrati dalle agenzie di somministrazione (dipendenti ex-interinali).

Tavola 7.1.2 – Peso delle posizioni lavorative dipendenti nell'industria e servizi di mercato (B-N) e nell'industria e servizi (B-S), per sottopopolazione di imprese nella rilevazione Oros
Anno 2014, valori medi percentuali

SOTTOPOPOLAZIONI	Peso sul totale B-N	Peso sul totale B-S
PMI Imprese di piccola e media dimensione (Fonte Oros - Inps)	75,5	76,5
GI_DA Imprese di grande dimensione assenti nel Panel dell'indagine GI (Fonte Oros - Inps)	0,8	1,7
GI Imprese di grande dimensione del Panel dell'indagine GI	21,8	20,0
INTERINALI Imprese di somministrazione (Fonte Oros - Inps)	1,9	1,8
TOTALE Totale imprese con almeno un lavoratore	100,0	100,0

Fonte: Rilevazione Oros (occupazione, retribuzioni, oneri sociali)

Dopo aver dato alcune informazioni sull'importanza dell'integrazione delle unità della sottopopolazione GI nella stima Oros, nel paragrafo 7.2 viene fatta una breve panoramica delle principali caratteristiche dell'indagine GI. Dopodiché, si affronta la metodologia seguita per l'integrazione dell'indagine GI nella rilevazione Oros, descrivendo nel paragrafo 7.3 la procedura per la corretta individuazione nella lista anagrafica Oros delle unità riconducibili alle imprese del panel GI, essenziale per evitare duplicazioni o errate esclusioni e, nel paragrafo 7.4, il processo di armonizzazione delle variabili economiche volto ad assicurare la coerenza delle stime trimestrali.

7.2 Alcune caratteristiche dell'indagine sulle grandi imprese

L'indagine GI nasce per fornire indicatori per l'analisi di breve periodo dell'andamento dell'occupazione, delle ore lavorate, delle retribuzioni di fatto e del costo del lavoro nelle imprese di grande dimensione (Istat, 2006). A partire dal 2003, diviene fonte sostitutiva di microdati sulle imprese di grandi dimensioni per la produzione di tutti gli indicatori della rilevazione Oros. Un avanzamento informativo significativo dell'indagine GI è, in seguito, ottenuto grazie all'integrazione con i dati dell'indagine trimestrale sui posti vacanti e le ore lavorate (Vela)³. In particolare, a partire dal 2012 le due indagini sono state integrate già nella fase della raccolta dati, attraverso un nuovo questionario usato per rilevare anche le variabili *target* su occupati dipendenti, ore lavorate e posti vacanti per le imprese con almeno 500 dipendenti, che rientrano nella popolazione obiettivo di entrambe le rilevazioni dando origine all'indagine Vela-GI (Bellisai et al., 2013).

La rilevazione sulle grandi imprese, condotta a cadenza mensile, si basa su un panel di imprese (panel censuario) costruito al momento della definizione dell'anno base come insieme di tutte le imprese con almeno 500 dipendenti in media annua e classificate nell'industria e nei servizi di mercato⁴. Sono escluse dal panel le attività di produzione cinematografica, di video e programmi televisivi, di registrazioni musicali e sonore (sezione J - divisione 59, Ateco2007), le attività immobiliari (sezione L – divisione 68) e le agenzie di somministrazione di lavoro interinale (sezione N – gruppo 78.2). Il panel censuario è individuato all'interno di un archivio integrato costituito in due passi: una prima versione si ottiene acquisendo informazioni dalla fonte Oros relative ad imprese che dovranno essere incluse nel panel all'anno base e da altre fonti congiunturali a disposizione (DB-sviluppo); una versione definitiva si ottiene agganciando anche informazioni provenienti da Asia relative all'anno base, disponibili successivamente rispetto alle precedenti fonti (DB-produzione).

Durante tutto il periodo di validità della base il panel resta immutato (panel chiuso) e si procede all'aggiornamento ogni 5 anni, in coincidenza con il passaggio alla base successiva. Il panel chiuso implica che nessuna unità possa entrarne o uscirne fino al suo *refresh*, anche se la dimensione dell'unità dovesse scendere sotto la soglia dimensionale scelta. Inoltre, per definizione, il panel dovrebbe seguire gli eventi giuridici e mantenere sia le nuove unità risultanti dalla scissione di unità appartenenti al panel, sia le imprese non appartenenti al panel iniziale ma incorporate da imprese che vi appartengono. Nella pratica, il primo caso viene risolto in parte attraverso la rilevazione delle variabili riferite alla nuova

3 In passato, al pari di Oros, l'indagine Vela ha utilizzato i dati mensili già raccolti dalla fonte GI per produrre le variabili trimestrali su ore lavorate, ore di straordinario e ore di cassa integrazione guadagni per le imprese di grandi dimensioni. Su tali imprese l'indagine Vela ha dovuto, quindi, raccogliere solo informazioni specifiche sui posti vacanti.

4 Vedi nota 2 per l'allargamento della rilevazione GI al settore dei servizi personali e sociali.

unità dell'impresa sulla preesistente, nel secondo la nuova impresa viene seguita solo se presenta un congruente numero di dipendenti.

Un aspetto rilevante emerso soprattutto negli ultimi anni nell'indagine GI è il *trade-off* tra la qualità e la tempestività del dato: se da un lato, la razionalizzazione delle azioni volte a garantire il rispetto dell'obbligo di risposta (solleciti e sanzioni) e l'utilizzo di nuovi strumenti IT per la raccolta standardizzata dei dati (Portale delle imprese) hanno consentito di ridurre sensibilmente il tasso di mancata risposta, dall'altro, la pressante richiesta di una riduzione della tempistica da parte dei regolamenti europei compromette la qualità del dato rendendo necessaria una revisione successiva. Pertanto, dal 2010, tutti i dati relativi all'ultimo anno, sui quali si basa la stima rilasciata a +46 giorni circa dalla fine del periodo di riferimento, vengono sottoposti ad una procedura annuale di revisione, al fine di incorporare negli indicatori le informazioni che si rendono disponibili successivamente alla pubblicazione delle prime stime (risposte tardive, rettifiche legate a eventi di trasformazione aziendale registrate con ritardo, ecc.). I dati mensili dell'anno precedente revisionati sono solitamente disponibili ad aprile di ogni anno in occasione della diffusione degli indici GI relativi al mese di gennaio.

7.3 L'individuazione della lista anagrafica Oros del panel dell'indagine sulle grandi imprese

La fase principale del processo di integrazione micro tra le due fonti è rappresentata dall'identificazione di una lista trimestrale di unità compresenti in base alla quale escludere dalla fonte Oros le unità appartenenti al panel GI. Occorre sottolineare che il *linkage* tra le due fonti si configura come un processo molto complesso e non come un semplice abbinamento tra unità, a causa dell'elevato grado di dinamicità che caratterizza la demografia delle imprese di grande dimensione (trasferimenti d'azienda, fusioni, scissioni, subentri, ecc.). Inoltre, poiché i movimenti anagrafici non sempre sono colti contestualmente dalle due fonti, si possono verificare errori di *mismatch* con conseguente impatto sulla qualità delle stime.

All'interno del processo di individuazione della lista GI, se da un lato è utile distinguere la fase della sua definizione in un dato anno base e il suo mantenimento/aggiornamento trimestrale sino al cambio base successivo, dall'altro, queste due macro fasi sono fortemente connesse. Infatti, il processo Oros che garantisce la diffusione di indicatori congiunturali nella base di riferimento data, parallelamente, imposta i lavori preparatori al cambio base successivo.

7.3.1 Definizione e mantenimento trimestrale

Dato il panel chiuso GI delle imprese con 500 e più dipendenti, il processo prevede la creazione e l'aggiornamento di una lista trimestrale delle unità Oros appartenenti al panel GI complementare a quella delle restanti unità Oros (in breve: PMI+GI_DA) (figura 7.3.1.1).

Il primo passo del processo di definizione della lista trimestrale consiste nel *matching* esatto tra le unità delle due fonti (fase automatica). Le fonti Oros e GI utilizzano come identificativi delle proprie unità di analisi, rispettivamente, la matricola aziendale Inps e il codice identificativo interno all'indagine GI. Essendo tali identificativi diversi tra loro, l'unico *matching* possibile è quello che utilizza il codice fiscale (CF) come chiave identificativa unica a livello di impresa. Il CF presenta elevati livelli di qualità in termini di completezza e correttezza in entrambe le fonti, garantiti da un trattamento preliminare cui viene sottoposto il CF di fonte amministrativa (cfr. §3.2.1) e da un'attenzione sempre crescente posta dagli

esperti d'indagine nei confronti di questa variabile, divenuta nel tempo fondamentale per collegare a livello micro i dati d'indagine con i dati di altre fonti (Rocci e Serbassi, 2008).

La lista trimestrale così individuata presenta, tuttavia, problemi residui di *mismatch* data la diversa tempistica con cui vengono registrati nelle due fonti eventuali cambiamenti di CF legati a movimenti demografici d'impresa⁵. In particolare, possono verificarsi casi di errato abbinamento, quando la stessa impresa viene registrata con CF diversi nelle due fonti, oppure, quando una trasformazione giuridica comporta la nascita di una nuova impresa a partire da un'impresa già presente nella lista ottenuta attraverso il *matching* esatto; in entrambi i casi il nuovo CF non risulta ancora presente nella lista GI. Per trattare tali casi, ogni trimestre, dopo la fase automatica, viene avviata una fase interattiva basata su una funzione di controllo che valuta la differenza tra il numero dei dipendenti nelle due fonti per lo stesso CF (fase manuale). Tuttavia, tale indicatore segnala non solo casi con probabile trasformazione giuridica, ma anche casi con presenza di mancate risposte imputate o meno. Per le imprese che superano una soglia limite prescelta, quindi, vengono isolati i casi di corretto abbinamento sui quali non sia stato effettuato un adeguato trattamento delle mancate risposte sul dato occupazionale, dai casi di errato/incompleto abbinamento tra imprese. In quest'ultimo caso, in cui la differenza nei dipendenti è attribuibile ad una vera trasformazione giuridica⁶, si inserisce il momento più delicato della fase manuale che ha come obiettivo la ricerca dei CF da aggiungere per ottenere una lista GI completa.

A tale scopo, è fondamentale l'uso congiunto delle informazioni sulle trasformazioni longitudinali delle imprese GI raccolte e aggiornate in uno specifico data base "di eventi" Istat⁷, di quelle provenienti dal Registro imprese delle Camere di Commercio (Telemaco) e dal Portale delle imprese dell'Istat. Nei casi residui in cui la trasformazione giuridica non risulti da tali informazioni, si cercano le connessioni tra imprese collegate tramite i siti web o il ricontatto diretto delle imprese.

Come è facile intuire, questa seconda fase del processo di integrazione teso a garantire la coerenza tra la lista trimestrale e i movimenti demografici delle unità appartenenti al panel chiuso, risulta essere quella che richiede il maggiore sforzo in termini di risorse e di tempo essendo difficilmente generalizzabile, ed assume rilevanza maggiore man mano che ci si allontana dall'anno base.

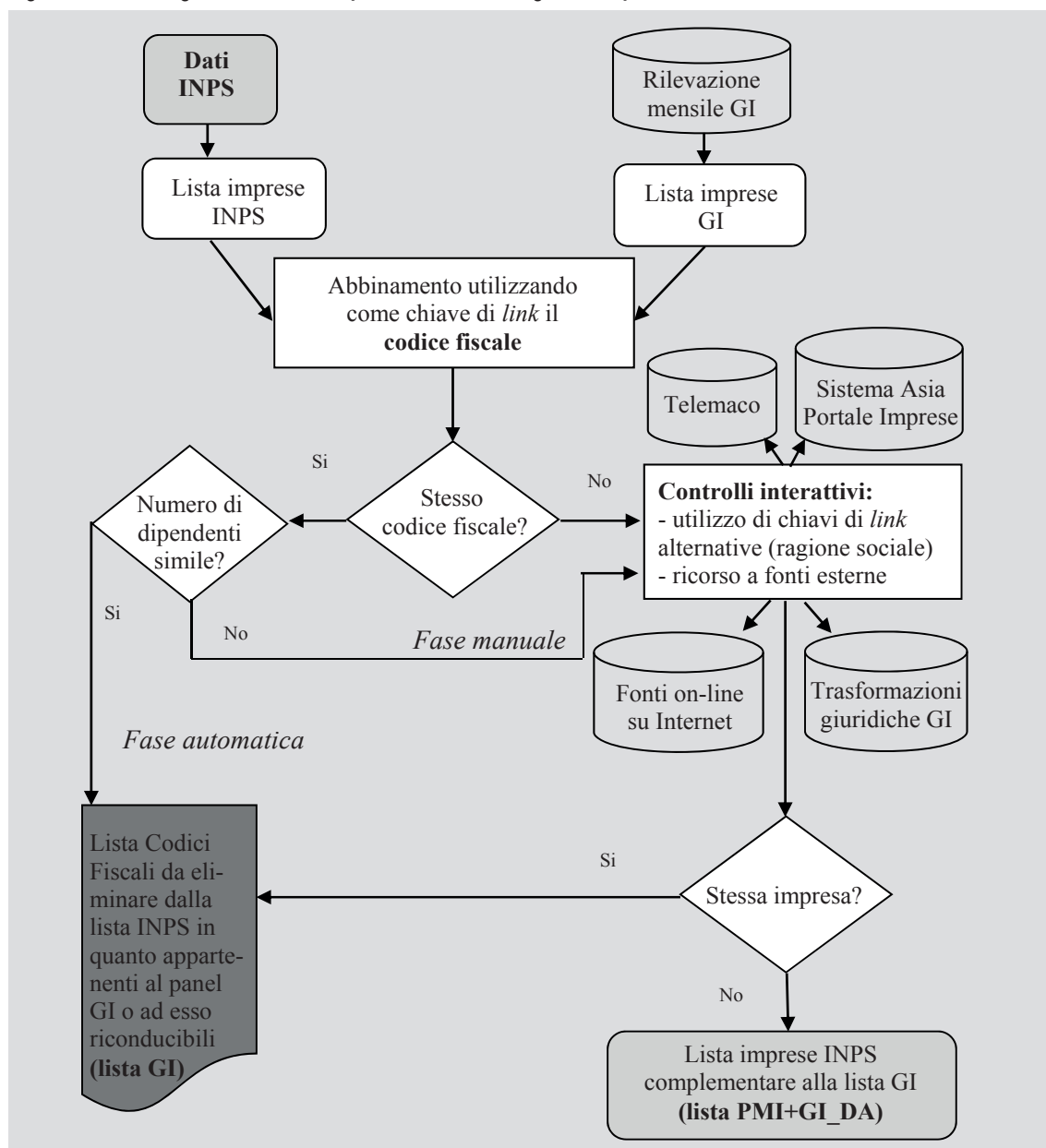
La lista trimestrale Oros delle imprese GI così definita non è solo output di un processo unidirezionale, ma è anche fonte d'informazione aggiornata che contribuisce al mantenimento della rappresentatività nel tempo del panel GI seguendo un processo ciclico di aggiornamento automatico e interattivo. Infatti, le trasformazioni giuridiche emerse vengono valutate dall'indagine GI ed eventualmente acquisite una volta l'anno, in occasione della revisione degli indicatori, per tutte le occorrenze dell'anno precedente implicando una revisione delle liste trimestrali Oros delle imprese GI. Tuttavia, per una gestione efficiente delle diverse fasi del processo Oros integrato, la revisione delle liste viene acquisita nelle stime definitive Oros un trimestre alla volta e non contemporaneamente su tutti i trimestri dell'ultimo anno completo.

5 La fonte amministrativa Inps, avendo finalità contributive contabili di legge, registra con tempestività le modifiche aziendali che comportano una nuova iscrizione anagrafica, mentre non sempre la cancellazione anagrafica viene dichiarata tempestivamente dall'impresa coinvolta, in assenza di una sanzione amministrativa.

6 Tra le trasformazioni giuridiche più frequenti si segnalano la cessione di dipendenti in impresa già esistente o in impresa di nuova iscrizione anagrafica, l'acquisizione dei dipendenti di altra impresa già esistente, ecc..

7 La necessità di tenere aggiornate le informazioni deriva dal fatto che gran parte delle imprese di grande dimensione subiscono modifiche anagrafiche e, se non si tenesse conto della loro evoluzione demografica, dopo poco tempo il panel chiuso GI non sarebbe più rappresentativo.

Figura 7.3.1.1 - Integrazione dei dati Inps con dati dell'indagine GI: il processo trimestrale



7.3.2 Definizione trimestrale per il cambio base

Accanto al processo trimestrale a regime sin qui descritto, si aggiunge ogni cinque anni l'attività straordinaria di ridefinizione delle liste trimestrali Oros per tener conto del cambio base GI che comporta la definizione di un nuovo panel d'impresе. Questa attività procede per passi: un'analisi preliminare ha inizio non appena i referenti dell'indagine GI definiscono le unità che entreranno nel nuovo panel e le rendono disponibili su un DB-sviluppo con un ritardo di circa 2 anni dalla fine dell'anno base, mentre la definizione delle nuove liste trimestrali Oros delle imprese GI avviene non appena le unità del nuovo panel vengono messe a disposizione sul DB-produzione due mesi dopo (figura 7.3.2.1).

Le informazioni estratte appositamente dal DB-sviluppo riguardano la lista dei codici impresa (identificativo dell'indagine GI) appartenenti ad almeno uno dei due panel consecutivi, il CF collegato che rappresenta la chiave di abbinamento con Oros (unico se il CF è invariato tra l'anno di riferimento della base e l'anno in cui è stato creato il DB-sviluppo, annuale in caso contrario), un *flag* stato del codice rispetto al panel precedente (*flag* stato 1 se nuovo entrato, 2 se uscito e 3 se panel) che rappresenta il legame tra i due panel consecutivi.

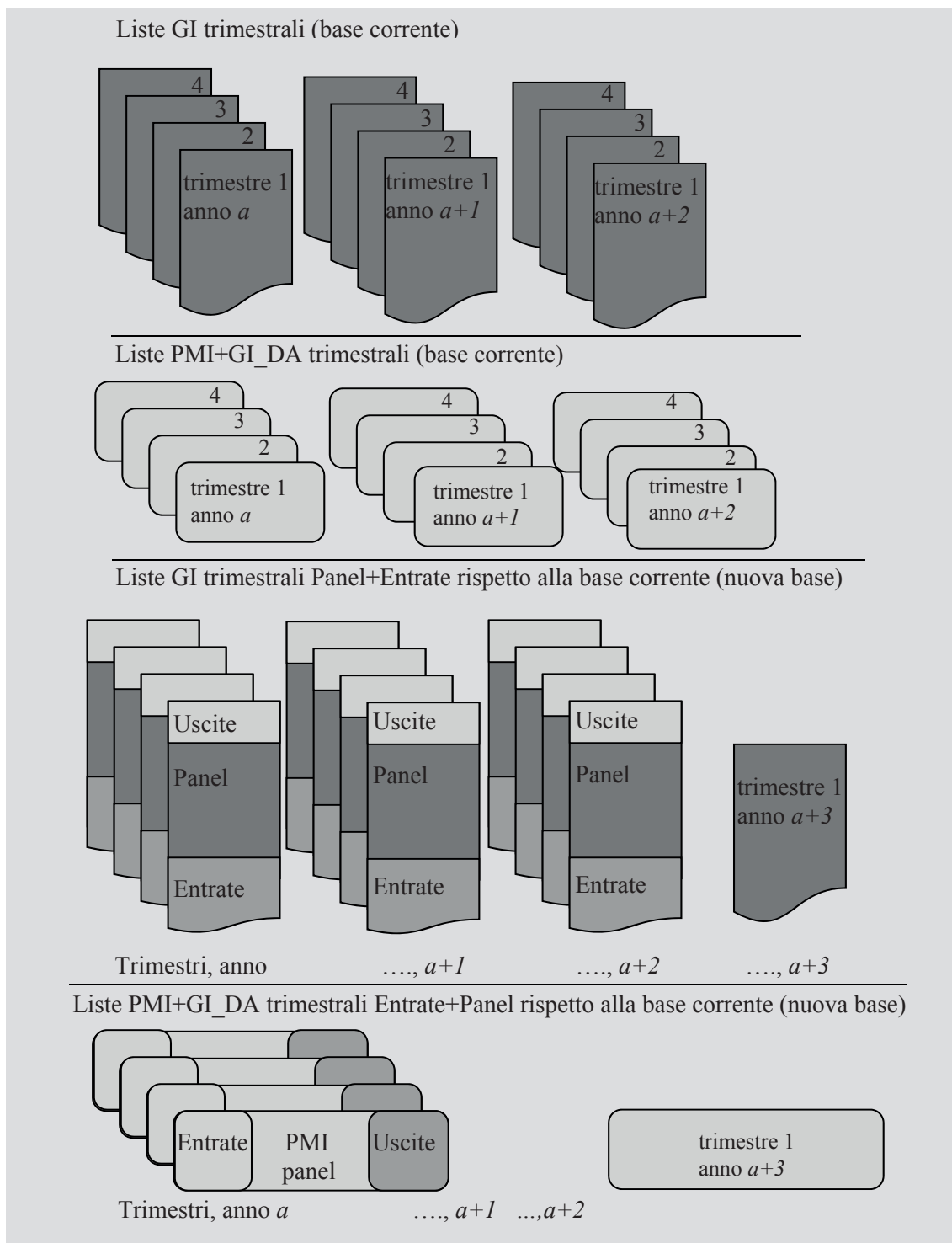
L'analisi preliminare ha come obiettivo la definizione di una lista univoca di CF determinata a partire dalla lista dei codice impresa GI, appartenenti ad almeno uno dei due panel consecutivi d'impresе, con associato il *flag* stato e la ragione sociale univoci. Nei casi non univoci, quando lo stesso CF presenta più codici, in cui almeno uno è qualificato 'panel' e, ad esempio, l'altro 'uscito' la regola definita comporta il mantenimento del CF nella lista. La lista dei CF così ottenuta si utilizza per individuare all'interno delle liste trimestrali Oros delle imprese GI del vecchio panel, il sottoinsieme di quelli ad essa complementare con l'obiettivo di assegnare anche a questi CF un *flag* stato corretto. Questo sottoinsieme include i CF che sono stati individuati durante tutte le attività trimestrali di *profiling* delle imprese portate avanti nel periodo che intercorre tra i due cambi base (CF-connessi), durante le fasi automatica e manuale di definizione delle liste trimestrali Oros delle imprese GI. Dato che nel nuovo panel GI gran parte delle imprese provengono dal vecchio panel, non è conveniente fare ex-novo il *profiling* ma si rivalutano le connessioni già individuate in precedenza per escludere connessioni non più valide. Alcuni dei CF-connessi potrebbero appartenere ad imprese che escono dal nuovo panel GI, oppure appartenere a connessioni di un passato non recente e in entrambe i casi vanno esclusi. Il criterio di esclusione usato utilizza un indicatore di probabile cessazione che tiene conto della persistenza del mancato *matching* con i micro dati trimestrali Oros (98 per la base 2010 su 534), per tutti i trimestri che vanno dal primo trimestre dell'anno base all'ultimo trimestre dell'anno base+2 e un indicatore di sovracopertura dell'anagrafe Inps per mancata cancellazione del CF, che valuta la costante assenza di dipendenti nel CF in caso di *matching* (366 codici fiscali per la base 2010 su 534).

I CF-connessi al netto degli esclusi (70 per la base 2010 su 534) possono presentare ragione sociale mancante, che viene recuperata da micro dati anagrafici precedenti all'anno base in quanto informazione utile nella fase manuale.

Al termine di questa fase si perviene a liste preliminari trimestrali Oros dei CF 'panel', 'nuovi entrati' e 'CF-connessi ai panel' per il periodo che include l'anno base e i 2 anni successivi.

Per perfezionare queste liste e renderle definitive, nella fase finale sul DB-produzione si utilizza nuovamente la funzione di controllo basata sulla differenza nell'occupazione tra le due fonti Oros e GI superiore ad una soglia prescelta.

Figura 7.3.2.1 - Integrazione dei dati Inps con dati dell'indagine GI: il processo per il cambio base



7.4 La stima economica delle imprese del panel dell'indagine sulle grandi imprese

All'aggiornamento delle liste anagrafiche con le unità del panel GI, come descritto nei paragrafi precedenti, segue il trattamento dei dati economici di fonte GI che produrrà per tali unità il dato economico da sostituire a quello di fonte Oros. A circa +46 giorni dalla fine dell'ultimo mese del trimestre di stima, l'indagine GI rende disponibili dati molto dettagliati relativi ai tre mesi del trimestre di stima con riferimento alle variabili che partecipano alla definizione del costo del lavoro consentendo, in tal modo, di pervenire alla stima delle variabili *target* Oros per questa sottopopolazione. L'integrazione garantisce una coerenza delle stime sulle stesse variabili prodotte dalle due rilevazioni, per quanto possibile e comporta, inoltre, la necessità di armonizzare alcune variabili, nonché la determinazione di variabili supplementari non presenti nella rilevazione. Il trattamento dei dati economici GI è necessario per passare da un dato mensile ad uno trimestrale e uniformare il più possibile i contenuti definitivi delle variabili stimate a partire dai dati d'indagine con quelli di fonte amministrativa. Il processo prevede anche la stima delle Ula che non rientra tra le variabili della rilevazione GI ma che, invece, è il denominatore degli indicatori retributivi di Oros.

Le variabili utili all'integrazione, rilevate mensilmente dall'indagine per ciascuna impresa del panel GI e distinte per le qualifiche impiegatizie e per quelle operaie, sono:

- il numero delle posizioni lavorative dipendenti all'inizio del mese di riferimento⁸ e alla fine del mese di riferimento (variabile di *stock*)⁹;
- il numero delle posizioni lavorative dipendenti a tempo parziale all'inizio e alla fine mese di riferimento (di cui delle variabili precedenti, variabile di *stock*);
- le ore di cassa integrazione guadagni (ordinaria e straordinaria) e di solidarietà utilizzate;
- le retribuzioni lorde totali (inclusive di componenti a carattere continuativo, saltuario e occasionale, inclusi gli incentivi all'esodo);
- gli incentivi all'esodo (diversi dal Tfr);
- i contributi sociali obbligatori legali versati all'Inps (oneri sociali a carico del datore di lavoro esclusi i conferimenti del Tfr al fondo Inps) e all'Inail.

Con riferimento al numero delle posizioni lavorative dipendenti, e non anche alle retribuzioni e contributi sociali, l'indagine GI rileva anche quelle relative ai dirigenti all'inizio e alla fine mese di riferimento. Tale informazione viene usata ai fini del calcolo degli indicatori Oros al netto dei dirigenti, attribuendola alla sola componente a tempo pieno non avendo la distinzione per tempo di lavoro.

Il trattamento mensile dei dati GI raccolti ha come obiettivo la stima delle variabili utili al calcolo degli indicatori Oros. In primo luogo, è necessario stimare le Ula al netto della Cig esclusi i dirigenti, denominatore del generico indice semplice Oros (cfr. §2.1), effettuando un opportuno riproporzionamento delle posizioni lavorative dipendenti. Nello specifico, per ogni generica unità i (identificata dal codice impresa) appartenente al panel GI, per un generico mese $m=1, 2, 3$ del trimestre corrente t e per una generica qualifica lavorativa $q=op$ (operai e apprendisti), im (impiegati, quadri e dirigenti), le posizioni lavorative dipendenti totali medie per qualifica ($\bar{d}_{i,m}^q$) sono date dalla media delle posizioni lavorative dipendenti totali alla fine del mese precedente ($d_{i,m-1}^q$) e del mese corrente ($d_{i,m}^q$)¹⁰:

8 Le posizioni lavorative dipendenti all'inizio del mese di riferimento devono corrispondere a quelle della fine del mese precedente.

9 Per completezza, si precisa che le posizioni lavorative dipendenti comprendono anche il personale temporaneamente assente per cause varie (ferie, permessi, maternità, cassa integrazione guadagni, ecc.).

10 Le posizioni lavorative dipendenti relative alla sottopopolazione del panel GI, calcolate secondo la relazione (1),

$$\bar{d}_{i,m}^q = \frac{d_{i,m-1}^q + d_{i,m}^q}{2} \quad (1)$$

con:

$$d_{i,m}^q = d_{i,m-1}^q + e_{i,m}^q - u_{i,m}^q$$

Dove: $e_{i,m}^q$ e $u_{i,m}^q$ sono, rispettivamente, le posizioni lavorative dipendenti entrate ed uscite nel mese di riferimento.

Dalla somma delle posizioni lavorative dipendenti per qualifica si ottengono le posizioni lavorative dipendenti totali medie:

$$\bar{d}_{i,m} = \sum_q \bar{d}_{i,m}^q \quad (2)$$

Ai fini del calcolo trimestrale delle Ula al netto dei dirigenti è dapprima necessario pervenire alla stima delle posizioni lavorative dipendenti per tempo di lavoro. Poiché l'indagine GI rileva, per ogni generica unità i e generico mese m , lo *stock* delle posizioni lavorative dipendenti a tempo parziale ($d_{i,m}^p$), le posizioni lavorative a tempo pieno ($d_{i,m}^f$) si ottengono per semplice differenza; in particolare, si ha:

$$d_{i,m}^f = \bar{d}_{i,m} - d_{i,m}^p \quad (3)$$

Nell'ipotesi che i dirigenti siano tutti full-time, la stima delle posizioni lavorative dipendenti full-time al netto dei dirigenti ($d_{i,m}^{f,nd}$) si ottiene come:

$$d_{i,m}^{f,nd} = d_{i,m}^f - d_{i,m}^{f,d} \quad (4)$$

dove ($d_{i,m}^{f,d}$) indica le posizioni lavorative dipendenti con qualifica dirigenziale, mentre le posizioni part-time al netto dei dirigenti sono date da:

$$d_{i,m}^{p,nd} = d_{i,m}^p \quad (5)$$

Per derivare una stima delle Ula al netto della Cig esclusi i dirigenti, per ogni generica unità i e generico mese m , alle posizioni lavorative dipendenti full-time esclusi i dirigenti si sommano le posizioni lavorative dipendenti part-time esclusi i dirigenti moltiplicate per un fattore di riduzione che tiene conto del tempo di lavoro ($q_{j,m}$) calcolato per ogni settore j (divisione Ateco 2007) nel generico mese m . Inoltre, si sottrae la stima delle Ula in Cig (${}_{cig}Ula$) ricavata rapportando le ore totali di Cig (Cig ordinaria, Cig straordinaria e solidarietà), rilevate mensilmente dall'indagine GI, al numero massimo di ore totali di Cig calcolato dall'indagine GI secondo un algoritmo che tiene conto del numero massimo di ore procapite di Cig concedibili fissato ogni anno dall'Inps (Istat, 2006 §3.1.4):

$$netto\ cig\ Ula_{i,m} = d_{i,m}^{f,nd} + q_{j,m} \cdot d_{i,m}^p - {}_{cig}Ula_{i,m} \quad \forall i \in j \quad (6)$$

rappresentano uno *stock* medio mensile che serve a misurare la variabile *target* costituita dallo *stock* medio trimestrale. Per la sottopopolazione Oros complementare al panel GI si stimano le posizioni lavorative dipendenti che abbiano prestato nel mese almeno una giornata retribuita che, pur non rappresentando uno *stock* medio (includendo i flussi degli entrati e degli usciti nel corso del mese di riferimento), sono la migliore approssimazione della variabile *target* trimestrale.

dove, come avviene per gli aggregati economici stimati con dati amministrativi, $q_{j,m} = me_{j,m} (ore_{i \in j,m}^p / IRC ore_{j,m}^f)$ rappresenta, nel generico mese m , la mediana delle Ula relative alle posizioni lavorative dipendenti part-time appartenenti al settore j , ottenute rapportando le ore pro-capite dei part-time di fonte Inps alle ore pro-capite lavorabili tratte dall'indagine IRC (cfr. §5.2).

Con riferimento al calcolo delle variabili del costo del lavoro, il metodo di calcolo, per la generica unità i e il generico mese m , prevede che il monte retributivo mensile sia dato dalla somma del monte retributivo degli operai e degli impiegati (esclusi i dirigenti):

$$r_{i,m} = \sum_q r_{i,m}^q \quad (7)$$

e che il monte dei contributi sociali sia dato dalla somma del monte contributivo degli operai e degli impiegati (esclusi i dirigenti) e dei contributi versati per il trattamento di fine rapporto (Tfr):

$$o_{i,m} = \sum_q o_{i,m}^q \quad (8)$$

dove:

$$tfr_{i,m} = (r_{i,m} - esodo_{i,m}) \cdot \left(\frac{1}{13,5}\right) \quad (9)$$

con:

$$esodo_{i,m} = \sum_q esodo_{i,m}^q$$

La (9) esprime la stima Oros degli accantonamenti per Tfr calcolata come quota delle retribuzioni lorde esclusa la componente retributiva legata all'esodo (esclusi i dirigenti).

Le componenti così calcolate a livello di microdato mensile vengono sottoposte a procedure di controllo e correzione volte a garantire la completezza e la correttezza delle informazioni anagrafiche ed economiche, valutandole anche in relazione a fenomeni di trasformazione giuridica e/o di ricorso a trattamenti di integrazione salariale (Cig). I principali controlli sono finalizzati a:

- l'individuazione automatica delle mancate risposte (MR). In particolare, il processo Oros individua gerarchicamente un insieme di cause che connotano le MR totali /MR parziali come "possibili", nel senso che giustificano la presenza della mancata risposta stessa e implicano assenza di correzione. Generalmente, questo si verifica quando un'impresa subisce una trasformazione giuridica che ne comporta la chiusura; in questi casi, infatti, nel data base dell'indagine, in corrispondenza del codice identificativo dell'impresa, risultano valorizzate soltanto le variabili relative alle posizioni lavorative dipendenti all'inizio del mese di riferimento, mentre le variabili di costo del lavoro non sono presenti. Pertanto, secondo la gerarchia adottata, si individua la causa della MR "possibile" prioritariamente attribuendola alla trasformazione giuridica. Nel caso in cui non tutte le MR parziali nel mese siano giustificate da trasformazione giuridica si procede alla verifica della presenza di ore di Cig in corrispondenza del codice. Per le unità residue si procede alla valutazione del singolo caso insieme agli esperti d'indagine. Un controllo aggiuntivo riguarda i contributi sociali, variabile che pur essendo componente del costo del lavoro non viene diffusa separatamente dall'indagine GI. In particolare, il controllo si focalizza sulle unità che presentano una

corrispondenza tra il valore dei contributi sociali e il valore del solo accantonamento per Tfr stimato come quota fissa della retribuzione lorda, segnale di contributi previdenziali mancanti. Questi casi vengono trattati insieme ai referenti d'indagine per verificare eventuali slineamenti nell'acquisizione del dato;

- l'individuazione automatica di eventuali *outlier* nei contributi sociali e nella retribuzioni lorde attraverso l'analisi delle distribuzioni delle relative variabili pro-capite e della quota dei contributi sulle retribuzioni corrispondenti (lordo e netto Cig);
- l'individuazione automatica di eventuale incoerenza nell'Ateco 2007 per lo stesso codice nei tre mesi del trimestre di stima che causerebbero problemi nell'attribuzione dell'Ateco nelle stime trimestrali. Questi casi, che si presentano raramente, vengono trattati insieme ai referenti d'indagine.

Il dato medio mensile nel trimestre a livello di singola unità si ottiene come media aritmetica dei dati mensili controllati e corretti, in modo coerente con quanto effettuato sulla sottopopolazione Oros complementare a quella del panel GI.

Come già spiegato nel §7.2, ad aprile di ogni anno, l'indagine GI sottopone a revisione le stime mensili riferite all'anno precedente per acquisire le informazioni che si rendono disponibili successivamente alla pubblicazione delle prime stime. Per tener conto delle revisioni che riguardano il dato economico GI, a giugno di ogni anno, in occasione della diffusione degli indicatori Oros relativi al primo trimestre, vengono riviste le stime Oros relative ai quattro trimestri dell'anno precedente.

8. IL CALCOLO DEGLI INDICATORI FINALIZZATI ALLA DIFFUSIONE NAZIONALE¹

8.1 Introduzione

Dopo aver effettuato gli ultimi controlli di qualità sui dati di fonte amministrativa e svolte le operazioni per l'integrazione dei micro dati amministrativi con quelli d'indagine, i valori economici tratti dalle due fonti vengono sommati per Ateco e sottopopolazione di riferimento e organizzati in una "banca dati" di stime aggregate in cui vengono aggiornate anche le rispettive serie storiche (cfr. §8.2). Mentre i dati aggregati di fonte GI non richiedono alcun trattamento aggiuntivo e sono già disponibili per integrazione e diffusione, per le stime tratte dai dati amministrativi è necessario svolgere alcune operazioni aggiuntive per derivare stime correnti e stime riviste degli indicatori *target*, in aderenza con le definizioni statistiche sottostanti. Tali trattamenti, sono in parte realizzati sui dati già aggregati per Ateco, altri richiedono il ritorno sui microdati per estrapolare componenti mancanti o distorte e correggere, successivamente, i macrodati in banca dati.

In particolare, per derivare le variabili di costo del lavoro viene effettuata un'operazione di correzione degli indicatori da fonte amministrativa per depurare le Ula stimate a partire dai microdati Inps dalla riduzione del tempo di lavoro dovuta al ricorso alla Cig (cfr. §8.3).

Parallelamente, le posizioni lavorative dipendenti vengono ricavate in seguito alla stima della componente mancante per assenza per Cig e altre motivazioni (cfr. §8.4). Per la ricostruzione delle serie storiche sulle posizioni lavorative dipendenti, unica variabile tra quelle Oros non espressa come pro-capite, in aggiunta, vengono adottati alcuni correttivi finalizzati a neutralizzare cambiamenti di livello sulle serie dovuti a riclassificazioni delle informazioni strutturali sulle singole unità (Ateco, forma giuridica ecc.) tratte dal registro statistico Asia, che implicherebbero un'errata lettura della congiuntura. Una stima *ad hoc* consente di ottenere le posizioni lavorative in somministrazione. Gli aggregati corretti da fonte amministrativa vengono infine sommati ai rispettivi aggregati da fonte indagine e ricostruite le serie storiche secondo il dettaglio utile per la diffusione nazionale e per le richieste dei regolamenti Europei.

Infine, le serie storiche degli indicatori sul costo del lavoro e sulle posizioni lavorative vengono indirizzate al sistema SITIC, sia per archiviazione secondo criteri standard, sia per la predisposizione dei formati utili per il loro trattamento per la stagionalità e successiva diffusione via I.Stat (cfr. §8.5).

¹ I paragrafi 8.1, 8.2, 8.3.1, 8.3.3, 8.4.7 e 8.5 di questo capitolo sono a cura di Donatella Tuzi; i paragrafi 8.3.2 e 8.3.4 sono a cura di Marco Lattanzio; i paragrafi 8.3.5, 8.3.6 e 8.4.3 sono a cura di Francesca Ceccato; i paragrafi 8.3.7, 8.3.8, 8.4.5 e 8.4.6 sono a cura di Elisabetta Aquilini; il paragrafo 8.4.1 è a cura di Francesca Ceccato, Eleonora Cimino e Donatella Tuzi; il paragrafo 8.4.2 è a cura di Eleonora Cimino; il paragrafo 8.4.4 è a cura di Marco Lattanzio, Francesca Romana Pogelli e Donatella Tuzi.

8.2 I dati aggregati nella banca dati Oros

8.2.1 Le serie storiche dal 2010 e i dati correnti

L'*editing* selettivo sui dati di fonte amministrativa, che tratta separatamente le variabili di costo del lavoro e le posizioni lavorative dipendenti (cfr. §6.6) rende disponibili, dopo le correzioni, i dati utili per la costruzione degli indicatori aggregati per pubblicazione.

Ai fini di successivi controlli delle variabili *target* in serie storica, le aggregazioni dei dati secondo le stime correnti (normalmente trimestri da t a $t-4$, cfr. §2.3.1) vengono effettuate e conservate per sottopopolazione di riferimento e a vari livelli di Ateco (divisione, sezione e aggregati di sezione) (Figura 8.2.1.1). I dati così aggregati vanno ad aggiornare le rispettive serie storiche disponibili allo stesso livello di dettaglio a partire dal I trimestre 2010. Mentre per i settori da P a S non si dispone di stime per i trimestri antecedenti al 2010 (cfr. §2.1), per i restanti aggregati settoriali, in conseguenza dell'ultima operazione di cambio base (da base 2005=100 a base 2010=100, avvenuta a giugno 2013), i dati risultano coerenti persino a livello micro, a seguito dell'introduzione delle novità sulle singole unità di rilevazione. La disponibilità di una lunga serie storica di microdati coerenti, che ha richiesto laboriose operazioni in occasione della transizione alla nuova base, rappresenta una potenzialità della rilevazione Oros, resa possibile dalla natura stessa dei microdati. Infatti, i dati amministrativi di fonte Inps consentono di coprire, in ciascuna occasione di rilevazione, l'intero universo delle unità *target* per cui, diversamente dalle indagini congiunturali classiche basate su dati statistici, le occasioni di cambio base rappresentano non già la circostanza per ridefinire i *panel* di riferimento, ma l'opportunità per migliorare aspetti definitivi/metodologici su tutte le unità dell'universo obiettivo. Tale configurazione consente di creare, a necessità, serie storiche sufficientemente lunghe di aggregati anche diversi da quelli di diffusione senza dover ricorrere a complesse operazioni di *back-casting* e armonizzazione di dati in serie storica.

A questo livello del processo di elaborazione, per le sottopopolazioni coperte con dati di fonte amministrativa, si rendono disponibili serie storiche relative alle variabili: retribuzioni, oneri sociali, costo del lavoro, Ula e, infine, posizioni lavorative dipendenti. Tutte le variabili sono distinte in lordo e netto dirigenti. Tuttavia, mentre le variabili di costo del lavoro sono al netto della Cig, le Ula ossia il denominatore degli indicatori di costo del lavoro, non lo sono. Allo stesso modo le posizioni lavorative stimate da dati amministrativi non includono le posizioni assenti per Cig e altre motivazioni, richieste dalla definizione statistica della variabile. Per ottenere dati rispondenti alle definizioni statistiche, per le stime tratte dai dati amministrativi è necessario svolgere alcune operazioni aggiuntive per derivare stime correnti e stime riviste degli indicatori *target*, in aderenza con le definizioni statistiche sottostanti. Tali trattamenti, sono in parte realizzati sui dati già aggregati per Ateco, altri richiedono il ritorno sui microdati per estrapolare componenti mancanti o distorte e correggere, successivamente, i macrodati in banca dati.

Nessun trattamento aggiuntivo è, invece, necessario per i dati aggregati di fonte GI che rispondono già alle definizioni statistiche sottostanti le variabili *target*.

Tali operazioni verranno descritte nei paragrafi 8.3 e 8.4.

Ai fini di controlli preliminari e per predisporre alcune forniture di dati *ad hoc*, tuttavia, per gli indicatori non ancora depurati dalle problematiche appena elencate, vengono ricostruite serie storiche dal I trimestre 2000, già inclusive della componente delle grandi

imprese di rilevazione. Si vedrà, in seguito, come una parte delle correzioni necessarie per giungere alle variabili aderenti alle definizioni statistiche vengono attuate direttamente su tali serie ricostruite, attraverso particolari accorgimenti. La ricostruzione delle serie storiche viene descritta nel paragrafo che segue.

8.2.2 Le serie storiche fino al 2009

Con il passaggio dalla base 2005 alla nuova base 2010, si è stabilito di introdurre le innovazioni previste a livello di micro dati non sull'intera serie storica delle occorrenze disponibili, ma a partire dal I trimestre 2010 e di traslare indietro le modifiche introdotte mediante raccordi in serie storica sui macrodati. Si ricorda che, tra le innovazioni introdotte, accanto all'acquisizione delle nuove informazioni della rilevazione GI, le principali novità hanno riguardato l'utilizzo di informazioni più fini sulla natura giuridica/istituzionale delle unità per consentire il successivo ampliamento del campo di osservazione della rilevazione ai settori da P a S della classificazione Ateco 2007, un miglioramento della stima del tempo di lavoro utile per il calcolo delle Ula, alcune integrazioni al metodo di stima delle Ula al netto della Cig, il perfezionamento della stima di alcune componenti degli oneri sociali, a seguito di un migliore sfruttamento di informazioni di carattere normativo (Istat, 2013b).

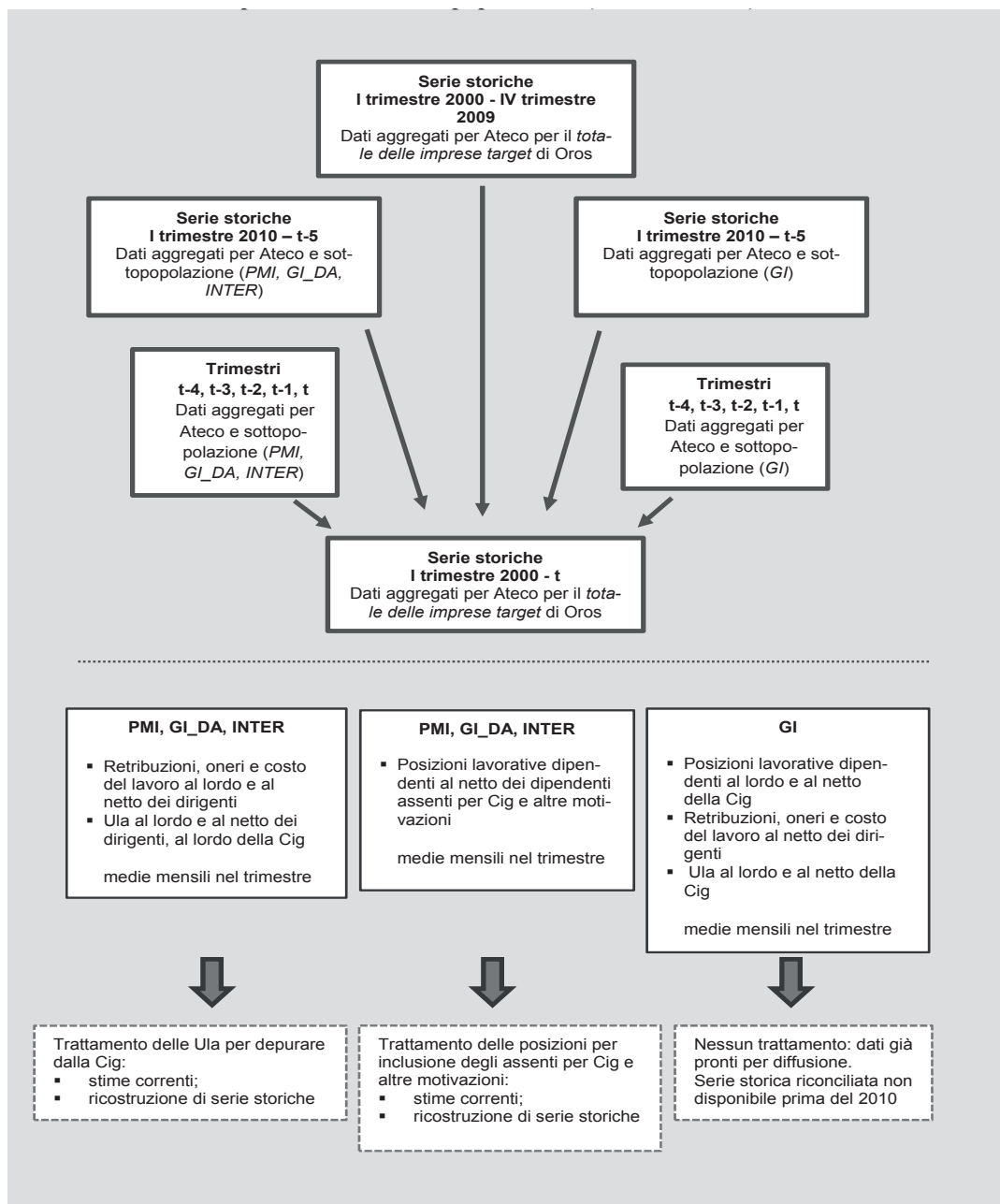
Per gli anni antecedenti al 2010 e fino al 2000, al fine di rendere disponibili dati trimestrali omogenei, si è proceduto alla ricostruzione delle serie storiche mediante raccordo, utilizzando dei coefficienti di *link* calcolati quali rapporti tra i livelli delle singole variabili che concorrono al calcolo degli indici, espressi nelle due basi di riferimento. I coefficienti sono stati calcolati in riferimento all'anno 2010, ad un livello di dettaglio di due digit (divisione) della classificazione per attività economica Ateco 2007. Tale operazione è stata effettuata sulle serie inclusive dei dati di fonte GI non disponendo, per tale sottopopolazione, di una serie storica sufficientemente lunga di dati già riconciliati nella vecchia base². L'armonizzazione delle serie storiche è stata, infine, ottenuta applicando i *link* alle medesime variabili sulle occorrenze delle serie da raccordare.

Una volta disponibili le serie storiche raccordate di tutte le variabili espresse in livello, vengono calcolate le variabili di costo del lavoro per Ula. In seguito, per tutte le variabili, dopo aver calcolato i rispettivi valori medi all'anno base, vengono derivati i rispettivi indici, secondo le relazioni esposte nel paragrafo 2.1 a cui si rimanda.

Nella Figura che segue vengono schematicamente illustrati i vari *set* di dati aggregati disponibili, per dimensione temporale, variabile e sottopopolazione. A partire da tali *set* vengono predisposte le serie storiche complessive degli indicatori *target* successivamente indirizzati alle ultime fasi di trattamento preliminari alla loro diffusione. In particolare, la correzione delle Ula per gli effetti della Cig, operazione descritta nel paragrafo 8.3 e l'inclusione nella stima delle posizioni lavorative dipendenti, degli assenti per Cig e altre motivazioni, di cui si parlerà nel paragrafo 8.4.

² Tale situazione rappresenta un notevole limite nel sistema informativo di Oros, sia per obiettivi di controllo della qualità delle stime diffuse, sia di ricostruzione di indicatori compositi che richiederebbero la necessità di scorporare dalle serie storiche delle variabili Oros la componente GI. Una problematica di questo tipo è stata affrontata nella ricostruzione delle serie storiche delle posizioni lavorative dipendenti totali, ossia inclusive degli assenti per Cig e altri motivi, come si vedrà nel dettaglio nel paragrafo 8.4.6, operazione che, per poter isolare gli aggregati derivanti da fonte indagine, ha richiesto l'imposizione di alcune ipotesi di semplificazione. Per evitare tale problematica, sono stati previsti degli accorgimenti da adottare in occasione del passaggio alla base 2015=100. Nello specifico si prevede di riconciliare la serie storica anche per la componente GI, consentendo in questo modo di disporre per ciascun aggregato prodotto da Oros, delle due componenti di stima: quella derivante da dati amministrativi, e loro trattamento, e quella ricavata dall'integrazione con l'indagine.

Figura 8.2.1.1 – Struttura dei dati aggregati nella banca dati Oros: dimensione temporale, per variabile e sottopopolazione
(base 2010=100)



8.3. Dal lordo al netto Cig: la stima delle variabili di costo del lavoro per Ula

8.3.1. Il quadro informativo di riferimento

Nell'utilizzo dei dati amministrativi per l'analisi congiunturale, una particolare attenzione va prestata alla identificazione e discriminazione dei cambiamenti economici con impatto sulle variabili *target*, dagli effetti dovuti a modificazioni nelle regole ammi-

nistrative che possono cambiare il significato del dato (Congia e Pacini, 2012). Il crescente ricorso alla Cig registrato negli ultimi anni è uno degli aspetti più rilevanti su cui è stato necessario approfondire ed intervenire, al fine di sfruttare adeguatamente il contenuto dei dati disponibili per incorporarne l'evoluzione nelle stime delle variabili Oros di costo del lavoro, significativamente influenzate, e comprenderne anche gli effetti.

Questa problematica non interessa i dati tratti dall'Indagine GI, in cui le ore di Cig e di solidarietà fruite sono dichiarate nei questionari mensili. Per quanto riguarda, invece, i dati tratti da fonte amministrativa, mentre l'assenza di lavoratori per Cig viene correttamente registrata dalle retribuzioni dichiarate all'Inps dai datori di lavoro (e di conseguenza dagli oneri sociali), l'aspetto più critico riguarda la misurazione dell'input di lavoro (le Ula) da cui devono essere escluse le ore/giornate non retribuite per Cig, variabile non utilizzata nella stima delle Ula dei full-time (cfr. §5.2). In generale, quando il ricorso alla cassa integrazione è relativamente contenuto e costante nel tempo, la stima della dinamica delle Ula non differisce in maniera sostanziale se l'occupazione ne viene definita al lordo o al netto. Per questo motivo, sino alla fine del 2008, la misura dell'occupazione utilizzata per il calcolo degli indicatori trimestrali Oros su retribuzioni lorde e costo del lavoro per Ula veniva definita al lordo dei dipendenti in cassa integrazione. Quando si registra, invece, un incremento consistente della Cig, a parità di tutte le altre condizioni, il monte retributivo erogato dalle imprese si riduce perché l'indennità di cassa integrazione è a carico dell'Inps e non rappresenta una voce di costo del lavoro per le imprese; contestualmente l'occupazione al netto Cig diminuisce mentre l'occupazione definita al lordo dei percettori del sussidio rimane stabile.

Al fine di evitare distorsioni negli indicatori retributivi Oros per gli effetti dovuti a forti variazioni nell'utilizzo della Cig e per consentire confronti temporali più omogenei fra periodi con maggiore o minore ricorso a tale strumento di politica del lavoro, il passaggio alla definizione dell'occupazione in Ula al netto della Cig si è reso indispensabile (Istat, 2013a). La correzione si basa sullo sfruttamento di fonti ausiliarie di tipo amministrativo e statistico e segue due diversi approcci che si contraddistinguono per consistenza della fonte amministrativa rispetto alla rappresentazione del fenomeno: un primo approccio, che sfrutta i microdati sulle ore autorizzate di Cig di fonte Inps combinati con le informazioni sulle ore/giornate retribuite presenti nei DM (*metodo micro*); un secondo approccio, che agisce a livello aggregato (sezione di attività economica), in cui si fa uso dei microdati sulle ore autorizzate di Cig in combinazione con dati aggregati sulle ore fruite di Cig tratti dalle indagini GI e Vela (*metodo macro*).

Di seguito, dopo aver riportato alcuni dettagli normativi sulla Cig (cfr. §8.3.2), vengono ricordate le fonti dati amministrative esistenti in Istat su tale prestazione (cfr. §8.3.3). In seguito, si entra nel merito della fonte sulle ore autorizzate di Cig dall'Inps, utilizzata per la stima gli indicatori Oros (cfr. §8.3.4) e illustrata l'evoluzione del ricorso alla Cig nell'ultimo decennio (cfr. §8.3.5). Successivamente, si entra nel merito della metodologia adottata per l'applicazione dei due diversi metodi (cfr. §8.3.6, cfr. §8.3.7) i cui risultati sono riconciliati in fase di compilazione degli indici finali (cfr. §8.3.8).

8.3.2 Cenni normativi sulla Cig

La Cig è una prestazione economica erogata dall'Inps con la funzione di sostituire od integrare la retribuzione dei lavoratori sospesi o che operano ad orario ridotto a causa di difficoltà produttive dell'azienda. Possono usufruirne gli operai, gli impiegati e i quadri mentre sono esclusi i dirigenti, gli apprendisti con contratto di apprendistato non professionalizzante e i lavoratori a domicilio³.

³ Per dettagli sulla normativa Cig si veda anche alla pagina web Inps dedicata:

L'intervento ordinario (Cigo) è rivolto alle aziende industriali non edili e alle aziende industriali ed artigiane dell'edilizia e del settore lapideo e opera in presenza di sospensioni o riduzioni temporanee e contingenti dell'attività d'impresa determinate da eventi transitori non imputabili all'imprenditore o ai lavoratori (es. mancanza di commesse, avversità atmosferiche ecc.). La durata massima è di 24 mesi nel quinquennio mobile, estesa a 30 nel settore edile.

L'intervento straordinario (Cigs) opera in caso di ristrutturazione, riorganizzazione e conversione aziendale, ovvero nei casi di crisi aziendale e di procedure concorsuali. Ne possono usufruire le imprese nei settori industriali e dell'edilizia, dell'artigianato, dei servizi di mensa e ristorazione, delle cooperative agricole e commerciali purché con oltre 15 dipendenti; inoltre le imprese commerciali con più di 200 dipendenti, le imprese editrici di giornali a prescindere dalla dimensione, le imprese di spedizioni e trasporto e le agenzie di viaggi e turismo con oltre 50 dipendenti. La durata massima della Cigs equivale a quella della Cigo.

L'intervento in deroga (Cigd), infine, è destinato ai lavoratori di imprese escluse dalla Cigs, quindi aziende artigiane e industriali con meno di 15 dipendenti o industriali con oltre 15 dipendenti compresi apprendisti, interinali e lavoratori a domicilio). Nella vigente normativa la Cigd è concessa nei casi in cui alcuni settori (tessile, abbigliamento, calzaturiero, orafo, ecc.) versino in grave crisi occupazionale. Ha durata massima di 12 mesi. Lo strumento della Cigd permette quindi, senza modificare la normativa che regola la Cig, di concedere i trattamenti straordinari anche a tipologie di aziende e lavoratori che ne sono esclusi.

Vanno, inoltre, assimilati alla Cig i contratti di solidarietà, ossia accordi stipulati tra azienda e rappresentanze sindacali che hanno come oggetto la diminuzione dell'orario di lavoro al fine di mantenere l'occupazione in caso di crisi aziendale e/o favorire nuove assunzioni. E' prevista una prima tipologia (tipo A) che rientra nel campo di applicazione della disciplina in materia di Cigs e una seconda (tipo B) per i casi che non vi rientrano.

Le modalità di pagamento della Cig sono due: 1) a *conguaglio* tramite azienda ossia l'azienda anticipa il trattamento per conto dell'Inps che, successivamente, provvede a conguagliare le somme dovute con i contributi da versare; 2) a *pagamento diretto* da parte dell'Inps che provvede, senza intermediazione, alla liquidazione degli importi d'integrazione autorizzati.

Oggetto di continui interventi finalizzati a rimodulare i vari strumenti alle esigenze congiunturali, la disciplina sulle prestazioni di Cig e solidarietà è stata recentemente interessata da una vasta operazione di riordino normativo (D.L. 148/2015) finalizzata a ridefinire i destinatari, l'ammontare e la durata della tutele, le modalità di erogazione e i termini di rimborso delle prestazioni, le condizioni di decadenza. Tra le principali novità introdotte, l'estensione della platea dei beneficiari, la semplificazione delle procedure di autorizzazione, la razionalizzazione della durata e la definizione di una nuova disciplina dell'istituto della solidarietà. Tale nuova disciplina dovrebbe anche avere risvolti positivi in termini di riordino dei flussi informativi amministrativi alla base della Cig, dalla fase di autorizzazione al suo utilizzo da parte del lavoratore.

8.3.3 Le fonti dati amministrative sulla Cig

Il complesso e variegato quadro normativo sulle prestazioni di Cig e solidarietà si riflette in una altrettanto complicata gestione amministrativa delle varie prestazioni, dalla loro autorizzazione da parte degli organi competenti, all'erogazione/conguaglio dei trattamenti di integrazione e solidarietà da parte dell'Inps. Tale struttura ha contribuito a determinare l'assenza di un sistema informativo unitario sull'evoluzione del fenomeno, in cui gli eventi fossero registrati tempestivamente e per periodo di competenza:

- nel precedente modello di dichiarazione contributiva DM10 in vigore fino a dicembre 2009 (Baldi et al., 2001), in un'apposita sezione (quadro F) venivano esposte informazioni sulla *Cig conguagliata* dalle imprese ma, a partire da tale file a parziale copertura di questo strumento, non era possibile collocare temporalmente la fruizione di tale prestazione;

- nell'Emens, modello in cui venivano comunicati i dati retributivi relativi al singolo lavoratore, era possibile individuare i dipendenti in *Cig a conguaglio* ma, almeno fino al 2010, non quelli a pagamento diretto;

- l'esistenza di un file Inps sulle ore di *Cig a pagamento diretto* per singolo individuo poteva consentire, tuttavia, di recuperare informazioni su tale tipologia di prestazione;

- dal 2010 nell'Uniemens sono contenute informazioni più dettagliate sulla Cig fruita. Sulla Cigo si hanno informazioni sulle ore conguagliate, mentre per la Cigs, accanto alle ore conguagliate, vi sono anche le *ore percepite* nel singolo mese di competenza;

- l'esistenza di un file Inps sulle ore *autorizzate di Cig* per singola matricola aziendale, di qualsiasi tipologia (a pagamento diretto o a conguaglio, ordinaria, straordinaria, deroga e solidarietà) per mese di autorizzazione. Tale file, strutturalmente molto snello e disponibile in tempi rapidi si riferisce, tuttavia, a ore di Cig che non sono di competenza ma possono riferirsi sia a periodi precedenti sia a periodi successivi il mese di autorizzazione.

Al fine di correggere gli indicatori Oros per la Cig, sarebbe stato necessario l'utilizzo congiunto di più fonti tra quelle esistenti, in parte a livello aziendale e in parte a livello individuale, richiedendo la preliminare risoluzione di problematiche di integrazione, armonizzazione tra le varie basi dati e disponibilità delle fonti nella tempistica utile per gli scopi congiunturali della rilevazione. La maggior parte di tali fonti sono state acquisite dall'Istituto solo da alcuni anni e non si rendono ancora disponibili nei tempi utili per le stime provvisorie di Oros.

Dopo un'attenta fase preliminare di studio su potenzialità informative e limitazioni delle varie fonti amministrative esistenti, considerati i vincoli di tempestività nella rilevazione Oros e in assenza di una fonte statistica esaustiva sul fenomeno⁴, la scelta metodologica adottata nel 2009, in occasione del passaggio degli indicatori alla base 2005 e in concomitanza con l'esplosione del ricorso alla cassa integrazione, fu quello di ricorrere all'uso del file sulle ore di Cig autorizzata come fonte ausiliaria di supporto rispetto ai micro dati di base contenuti nei DM.

8.3.4. Il file sulle ore di Cig autorizzata

A partire da luglio 2009 l'Istat acquisisce trimestralmente dall'Inps informazioni sulle matricole afferenti ad imprese autorizzate a fruire della Cig e solidarietà (inclusa nella Cigs) per trimestre di autorizzazione. La prima fornitura, avvenuta circa un anno prima e utilizzata ai fini esplorativi, è stata inclusiva anche degli anni 2007 e 2008 oltre che dei primi trimestri

4 L'Indagine Vela-Gi, che misura le ore complessive di cassa integrazione guadagni, ordinaria, straordinaria e in deroga e ore di solidarietà di cui le imprese hanno usufruito nel trimestre di riferimento limita, tuttavia, il proprio campo di osservazione alle imprese con almeno 10 dipendenti.

del 2009. In seguito, dopo una fase preliminare di studio, l'acquisizione è stata schedulata e, attualmente, avviene in modo regolare, con un ritardo rispetto alla fine del trimestre di riferimento delle stime preliminari t di circa 20 giorni.

La fonte, rappresentata dagli archivi amministrativi Inps che gestiscono la cassa integrazione, contiene il numero di ore autorizzate per tipologia di Cig e mese di autorizzazione. Gli archivi sono alimentati nel modo seguente: per quanto riguarda la cassa integrazione ordinaria industria ed edilizia dalle delibere di autorizzazione dell'Inps, per la cassa integrazione straordinaria e in deroga dai decreti emessi dagli uffici regionali del lavoro. I contratti di solidarietà precedenti all'entrata in vigore del D.Lgs. 148/2015 sono trattati come causale della Cigs.

Queste informazioni possono essere utilizzate non come misura diretta, ma come *proxy* per depurare dalle Ula totali l'input di lavoro non prestato a causa di Cig al fine di ottenere una misura corretta delle Ula che concorrono alla produzione del reddito misurato. L'utilizzo indiretto della variabile deriva dal fatto che essa non rappresenta l'ammontare di ore fruite di Cig ma, per ogni matricola, sono indicate le ore massime cui l'impresa è autorizzata a fruire per le quali tuttavia non è disponibile informazione riguardo l'istante o il periodo di effettivo utilizzo e nemmeno il quantitativo di ore effettivamente fruite. Inoltre, le ore di Cig in molti casi vengono autorizzate in un periodo successivo rispetto a quello in cui vengono effettivamente fruite (la maggior parte) o utilizzate successivamente. La stima Oros delle ore fruite si basa su due metodi, nei quali si combinano opportunamente le informazioni disponibili dal file di Cig autorizzata: un metodo di correzione micro che sfrutta informazioni di fonte Inps presenti nei DM2013 sulle giornate retribuite e un metodo di correzione macro basato anche sull'utilizzo delle ore fruite di Cig di fonte indagine Vela-GI. I due metodi verranno illustrati nei due paragrafi che seguono.

Oltre ad entrare nel processo di calcolo delle Ula, i microdati sulle ore autorizzate di Cig vengono usati anche nella stima delle posizioni lavorative al lordo dei beneficiari di questa prestazione, come verrà descritto in dettaglio nel paragrafo 8.4.

La disponibilità di microdati di ore di Cig autorizzata ha permesso di apportare numerosi affinamenti nelle fasi di processo che utilizzano questa fonte di dati. In *primis* la possibilità di abbinare l'informazione amministrativa con variabili statistiche strutturali a livello di microdato. Inoltre anche la possibilità, come verrà descritto in seguito, di integrare informazioni sull'utilizzo della Cig per imprese presenti nel panel GI e nei dati Oros.

La fonte è sottoposta ad un processo di valutazione della qualità tramite il confronto con i dati presenti nell'Osservatorio Statistico sulle ore di Cig autorizzata dell'Inps⁵. A causa della tempestività con cui viene fornito il set di dati si possono osservare revisioni, a volte anche importanti, tra i dati aggregati ottenuti per somma dei microdati degli scarichi trimestrali e quelli di fonte macro scaricabili nella sezione *Open Data* dell'osservatorio elaborati dall'ufficio statistico.

8.3.5. Dinamica della Cig e impatto sugli indicatori di costo del lavoro

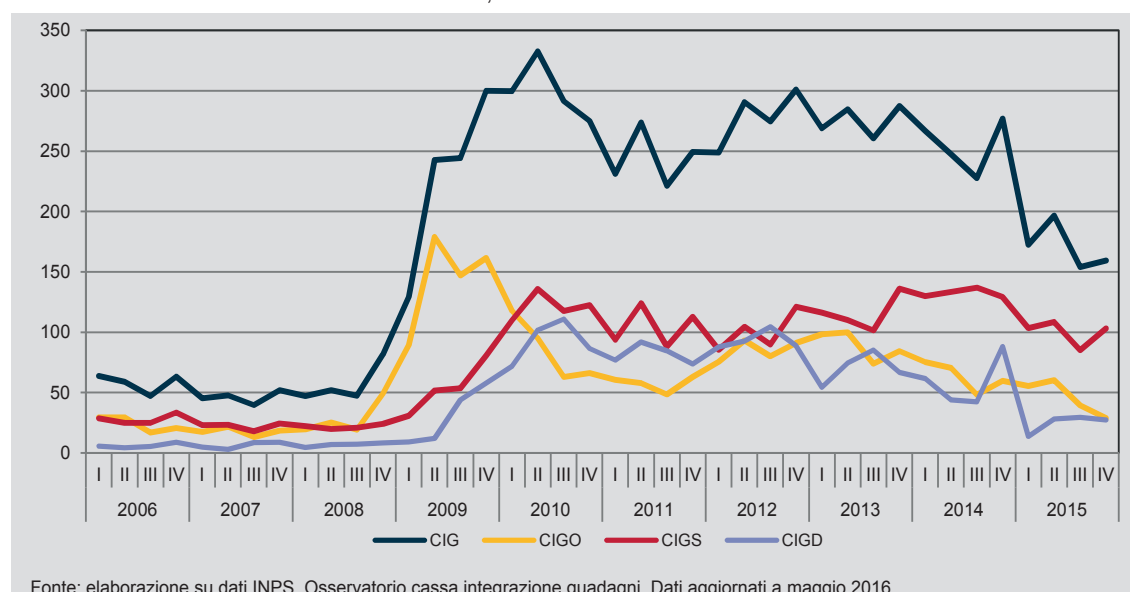
La grave crisi economica che nell'ultimo decennio ha travolto molte imprese italiane ha visto crescere in modo notevole il ricorso a strumenti di contenimento come la Cig (ordinaria, straordinaria, in deroga) o il contratto di solidarietà (difensivo, espansivo), strumenti finalizzati a sgravare le imprese in difficoltà da parte dei costi della forza lavoro temporanea-

⁵ Sezione "Banche Dati Statistiche", al link:
<https://www.inps.it/webidentity/banchedatistatistiche/menu/cig/main1.html>.

8. Il calcolo degli indicatori finalizzati alla diffusione Nazionale

mente non utilizzata. Nel corso degli anni e con l'istituzione, a fine 2008, della Cigd la platea dei beneficiari si è gradualmente ampliata, andando a coprire anche i settori produttivi non ancora considerati da tutele "ordinarie" di sostegno al reddito. Quest'ultima tipologia, di fatto, è stata spesso utilizzata con lo scopo di "prorogare" la tutela in situazioni di crisi in cui gli ammortizzatori ordinari erano stati ormai esauriti (Grafico 8.3.5.1).

Grafico 8.3.5.1 - Ore autorizzate di cassa integrazione guadagni (ordinaria, straordinaria, in deroga e totale)
I trimestre 2006 - IV trimestre 2015, valori assoluti in milioni



L'andamento delle ore di Cig autorizzate mostra come il numero massimo di ore totali concordato sia stato raggiunto nel 2010, anno in cui alla Cigo, il cui massimo si colloca a metà 2009, si è sommata la Cigd e la Cigs autorizzata per arginare la perdita di occupazione. Si ricorda che le ore di Cig autorizzate non sono quelle effettivamente fruite dalle aziende (cfr. §8.3.4) e questo pone l'accento sull'importanza di conoscere la distribuzione nel tempo del loro utilizzo, perché è quest'ultimo che, se non correttamente considerato nel calcolo dell'occupazione, può determinare una significativa distorsione nella stima degli indicatori congiunturali del costo del lavoro per Ula.

Come stimare la quantità del lavoro persa e conoscerne il periodo di competenza relativo rappresenta l'informazione cruciale. Una stima grossolana dell'effettiva competenza si può ricavare dalla percentuale di *tiraggio*, misura resa disponibile periodicamente dall'Inps e che rappresenta la quota di ore autorizzate ed effettivamente fruite nel periodo di tempo preso come riferimento (Inps, 2016). Le evidenze empiriche dedotte da precedenti studi dimostrano che solo una piccola percentuale di ore di Cig sono utilizzate nel trimestre in cui vengono autorizzate, mentre una percentuale molto più significativa viene fruita nei trimestri immediatamente successivi o precedenti (Congia e Pacini, 2012). Negli ultimi anni, però, non si è avuta stabilità nel tempo tra date di autorizzazione e competenza per cui l'evidenza è in qualche modo superata; in particolare, con il recente riordino della normativa sugli ammortizzatori sociali approvato con Decreto Lgs.148 del 14 settembre 2015 in attuazione della legge 10/12/2014 n.183 (c.d. Jobs act) ha avuto inizio una razionalizzazione necessaria sulle modalità con cui viene gestito il flusso amministrativo che sottende l'erogazione dell'integrazione salariale. La gestione del flusso

è stata innovata introducendo novità che riguardano anche la coerenza tra le denunce contributive del lavoratore e l'elaborazione dei cedolini paga e le scritture sul libro unico del lavoro⁶.

Già a partire dalla fine del 2009 in occasione della diffusione degli indicatori di costo del lavoro per Ula in base 2005, è stata proposta una metodologia per risolvere il delicato tema della "nettizzazione" di tali indicatori dal crescente utilizzo della Cig poiché il ricorso a questo strumento aveva raggiunto una tale dimensione da rendere distorta la loro interpretazione economica. L'indicatore al lordo della Cig calcolato fino ad allora sottostimava la dinamica retributiva in quanto le Ula utilizzate al denominatore non venivano depurate dalla quantità di lavoro persa per Cig e/o solidarietà. Due sono, in particolare, gli approcci individuati basati *in primis* sullo sfruttamento di tutti i micro dati disponibili, sia di fonte Inps sia di fonte indagine GI e, in seguito, di tutti i macro dati utili (tiraggio Inps, accostamento con indicatori di ore di Cig procapite fruita di fonte statistica) per derivare l'occupazione in Ula al netto della Cig. Nel corso degli ultimi anni la prima versione proposta del metodo è stata progressivamente raffinata a seguito di un migliore utilizzo delle fonti ausiliarie statistiche disponibili, nonché delle informazioni sulle ore autorizzate da parte dell'Inps. Di seguito si descrive la metodologia applicata.

8.3.6 Metodo micro

Un primo metodo, denominato *micro*, applicato a partire dal I trimestre 2010, sfrutta i microdati di fonte Inps sulle ore autorizzate di Cig combinati con le informazioni sulle ore/giornate retribuite presenti nei DM. Alla base di questo approccio vi è l'ipotesi che il tempo di lavoro retribuito sia una misura migliore dell'input di lavoro e tenga conto del contributo effettivo della forza lavoro al processo produttivo, anche dell'assenza per Cig/solidarietà.

La fornitura di micro dati mensili dell'archivio Inps sulle ore autorizzate di Cig a livello d'impresa distinte per mese di competenza, viene utilizzata per definire una *lista trimestrale* delle imprese che accedono ai vari istituti di integrazione salariale (d'ora in poi si-cig). Tale informazione, attribuita micro alle imprese con dipendenti della rilevazione Oros appartenenti all'insieme NO_GI (cfr. §5.3), permette di qualificare l'impresa per tale caratteristica, in aggiunta alle altre informazioni anagrafiche già presenti (settore di attività economica Ateco 2007, appartenenza al dominio di stima per forma giuridica) consentendo così di pervenire alla definizione di sottopopolazioni omogenee sulle quali effettuare o meno la correzione della stima per la Cig e la solidarietà.

La correzione, a livello micro, viene attuata solo sulle imprese NO_GI che ricorrono alla Cig e/o alla solidarietà e ha lo scopo di derivare una stima dell'input di lavoro nell'impresa che sia coerente con il monte retributivo pagato dalle aziende, al netto delle indennità per Cig in quanto a carico dell'Inps, operazione attuata sfruttando le informazioni sulle *giornate/ore retribuite* dichiarate nel DM. L'ipotesi alla base del metodo è che un'impresa classificata

6 L'Inps è passata da un sistema di denuncia delle informazioni sulla Cig a posteriori (soltanto nel momento successivo all'avvenuta autorizzazione) ad una denuncia corrente (anche in relazione ad eventi che precedono l'autorizzazione); da una gestione diversificata delle informazioni sulla denuncia dei periodi di riduzione o sospensione, a seconda che vi fosse anticipazione da parte dell'azienda o pagamento diretto da parte dell'Inps, ad una gestione uniforme. Tuttavia, le numerose potenzialità di questo nuovo flusso informativo presente negli archivi UniEmens dell'Inps potranno essere valutate ed eventualmente sfruttate nelle statistiche congiunturali solo a seguito di un approfondito studio di fattibilità sulla nuova fonte informativa.

nella sottopopolazione delle imprese si-cig rimanga tale dal momento della sua inclusione nel gruppo. Per questo insieme di unità deve essere derivata l'occupazione, misurata in termini di Ula, al netto della Cig che sarà il denominatore corretto per gli indicatori di costo del lavoro rilasciati da Oros a livello nazionale.

Per ogni impresa i appartenente all'insieme NO_GI e al generico settore j nel trimestre corrente t , la stima delle Ula al netto della Cig con questo metodo è ottenuta rapportando le ore totali dei full-time (ore^f) e dei part-time (ore^p) alle ore mensilmente lavorabili, nel settore j del trimestre t , in base al contratto collettivo nazionale di lavoro per un lavoratore a tempo pieno (${}_{IRC}ore$). Queste ultime, stimate dall'indagine mensile IRC dell'Istat, vengono applicate per la generica sezione j della classificazione Ateco 2007. Più in dettaglio:

$${}_{nettocig,si-cig}Ula = ({}_{si-cig}ore^f + {}_{si-cig}ore^p) / {}_{IRC}ore \quad (1)$$

Poiché, nella fonte amministrativa le imprese dichiarano le giornate retribuite in corrispondenza dei lavoratori full-time (gio^f), per ottenere le ore per questi lavoratori è necessario convertire le giornate in ore, per ogni impresa i del generico settore j nel trimestre t , moltiplicandole per il numero di ore giornaliere, a sua volta ottenuto rapportando il numero di ore mensilmente lavorabili per contratto a 26 (numero di giornate convenzionalmente fissate dall'Inps come numero massimo contribuibile nel mese):

$${}_{si-cig}ore^f = {}_{si-cig}gio^f ({}_{IRC}ore / 26) \quad (2)$$

Le ore dei lavoratori part-time sono invece riportate nel modello DM da parte delle imprese stesse.

La differenza tra le Ula al lordo Cig e quelle al netto Cig, per l'insieme delle unità NO_GI, rappresenta una stima delle Ula in Cig che per il generico settore j nel trimestre t può essere espressa come:

$${}_{cig}Ula = ({}_{lordocig}Ula - {}_{nettocig,si-cig}Ula) \quad (3)$$

dove nelle Ula al lordo della Cig sono incluse quelle delle unità che non fanno parte della lista si-cig e che, per definizione, non ne sono caratterizzate.

Per la successiva armonizzazione dei risultati ottenuti dal metodo micro per le imprese dell'insieme NO_GI con quelli che verranno ricavati dal metodo macro, di cui si forniranno dettagli nel paragrafo successivo e per riconciliare, infine, la stima ottenuta con le Ula al netto Cig da fonte GI, per il settore generico j nel trimestre t , si rende necessaria la stima delle *ore fruite* di Cig per le NO_GI che si ottengono moltiplicando le Ula in Cig di cui alla (3) per le ore "convenzionali" trimestralmente lavorabili rilasciate dall'indagine IRC (${}_{IRC}Ore$) per generica sezione j Ateco 2007 nel trimestre t .

$${}_{cig}ore\ fruite = ({}_{cig}Ula \cdot {}_{IRC}Ore) \quad (4)$$

L'occupazione in Ula che viene corretta con il metodo sopra descritto è quella dei settori dell'Ateco 2007: B - Estrazione di minerali da cave e miniere; C - Attività manifatturiere; D - Fornitura di energia elettrica, gas, vapore e aria condizionata; E - Fornitura di acqua, reti fognarie, attività di gestione dei rifiuti e risanamento; G - Commercio al dettaglio, riparazione di autoveicoli e motocicli; H - Trasporto e magazzinaggio. La motivazione principale per cui la correzione non viene attuata per i tutti i settori produttivi è data dal fatto che alcuni settori

sono da sempre caratterizzati da input di lavoro ridotto o estremamente irregolare, per i dipendenti full-time, legati ad elevato turnover, forme contrattuali atipiche di lavoro, modifiche nella composizione dell'occupazione, caratterizzazione non riconducibile al crescente ricorso alla Cig. Attuare la correzione anche su questi settori avrebbe comportato revisioni alla serie storica del costo del lavoro per Ula attribuite erroneamente al ricorso alla Cig quando in realtà derivava da un comportamento pre-crisi non riconducibile a tale strumento.

Queste motivazioni hanno indotto ad adottare, per questi settori rimanenti, un metodo diverso basato su informazioni micro sulla Cig autorizzata trattate insieme ad informazioni macro da fonti statistiche sulla Cig fruita. Un discorso a parte va fatto per i settori relativi ai servizi personali e sociali (sezioni da P a S), per i quali la prima diffusione degli indicatori Oros è avvenuta a dicembre 2013. In quell'occasione, data la particolare volatilità degli indicatori sottostanti e lo scarso impatto della Cig nei settori in oggetto, fu ritenuto preferibile non utilizzare una correzione basata sui microdati, ma ricorrere ad un approccio macro. Alla descrizione del metodo macro è dedicato il paragrafo successivo.

8.3.7 Metodo macro

Alla base dell'applicazione del metodo micro, come visto nel paragrafo precedente si fa l'ipotesi che, in caso di ricorso alla cassa integrazione, la riduzione dell'input di lavoro conseguente a Cig fruita dai lavoratori sia tempestivamente registrata dalla variabile amministrativa che riporta le giornate/ore retribuite. In alcuni settori di attività economica, tuttavia, tale ipotesi diviene troppo stringente rendendo il metodo micro non applicabile. È il caso di quei settori in cui il ricorso alla Cig non è legato soltanto alla congiuntura economica, ma anche al verificarsi di altre circostanze, quali eventi meteorologici che comportano la sospensione dell'attività produttiva o ricorso a Cig parziale nel corso della giornata (settore delle costruzioni), elevato turnover (settori degli alberghi e ristoranti, servizi di informazione e comunicazione), utilizzo di forme contrattuali atipiche (settori delle attività professionali, scientifiche e tecniche, degli altri servizi alle imprese, ecc.) e frequenti mutamenti nella composizione dell'occupazione, che possono rendere poco utilizzabile la variabile giornate/ore retribuite quale indicatore di ridotto input di lavoro per Cig (Congia, Pacini, 2012). Per questi casi è stata adottata una soluzione diversa, basata sull'utilizzo dei microdati Inps sulle ore autorizzate di Cig (cfr. §8.3.4), sfruttati a livello macro dopo essere stati opportunamente trattati per circoscriverne l'ambito di applicazione e correggere per un effetto di distorsione di cui sono caratterizzati. I settori interessati dalla stima con metodo macro sono: F: Costruzioni; I: Attività dei servizi di alloggio e di ristorazione; J: Servizi di informazione e comunicazione; K: Attività finanziarie e assicurative; L: Attività immobiliari; M: Attività professionali, scientifiche e tecniche; N: Noleggio, agenzie di viaggio, servizi di supporto alle imprese. Per queste prime sette sezioni l'approccio macro nella versione qui descritta è stato applicato a partire dal I trimestre 2008. Inoltre, P: Istruzione; Q: Sanità e assistenza sociale; R: Attività artistiche, sportive, di intrattenimento e divertimento; S: Altre attività di servizi, per le quali il metodo macro è stato applicato a partire dal I trimestre 2010, periodo da cui si dispone di stime per tale comparto dei servizi.

Un primo trattamento ai microdati sulle ore di Cig autorizzate riguarda l'esclusione dai dati originari delle unità che appartengono alla sottopopolazione delle GI di rilevazione, per cui le Ula al netto della Cig sono già disponibili a partire dai dati dell'indagine. Successivamente, i dati vengono ricondotti al dominio di riferimento di Oros (imprese, istituzioni private, in stima) e aggregati per sezione Ateco 2007, attribuita per matricola aziendale tramite abbinamento con l'anagrafica Oros.

Come più volte ricordato, trattandosi di ore *autorizzate*, tali dati sovrastimano le ore effettivamente utilizzate soprattutto in periodi di crisi quando le imprese, a scopo cautelativo, spesso richiedono l'autorizzazione di più ore di cassa integrazione di quante poi ne utilizzino realmente. Proprio per questo motivo l'Inps pubblica periodicamente anche un indicatore sull'effettivo utilizzo di questo ammortizzatore sociale, il cosiddetto "tiraggio", che misura la quota delle ore autorizzate che viene effettivamente utilizzata dalle imprese (Inps, 2016). Per ottenere una stima più adeguata delle ore fruite, l'approccio macro applica alle ore autorizzate ricavate dal precedente passaggio dei tassi di abbattimento (o, nel linguaggio Inps, di tiraggio) che sono stimati sulla base di variabili ausiliarie, quali:

- informazioni provenienti dalla dinamica delle ore fruite di Cig rilevate dall'indagine Vela-GI sulle imprese con almeno 10 dipendenti, che rappresenta una *proxy* significativa soprattutto in quei settori in cui non c'è una netta prevalenza di imprese sotto i 10 (il settore F non è trattabile con questa fonte) o sopra i 500 (escluse dai microdati considerati);

- un tasso "implicito" di tiraggio calcolato rapportando le ore fruite di Cig ottenute con metodo micro per i settori interessati da questo approccio alternativo, alle ore autorizzate. Infatti, per alcuni settori trattati con approccio macro, si riscontra una forte similitudine nel profilo delle ore autorizzate in settori trattati con approccio micro, che lascia presupporre un accostamento anche sulle ore fruite (ad esempio il settore B mostra similitudini con F, C invece con G per citare alcuni esempi).

La necessità di passare per un tasso di tiraggio introduce dei limiti molto rilevanti al livello di dettaglio Ateco con cui gli indicatori di costo del lavoro per Ula di Oros possono essere diffusi, a causa della dispendiosità temporale che questa fase richiede, nonché di un certo grado di soggettività nella scelta del valore del tiraggio, operazioni che avvengono molto a ridosso delle date di diffusione interne ed esterne. In futuro, a seguito dell'ampliamento settoriale nell'utilizzo del metodo micro, è prevista la produzione degli indicatori a livello di divisione.

Entrambe le informazioni vengono analizzate in serie storica, per arrivare a stabilire per ogni sezione Ateco un tasso di tiraggio, che corrisponde alla percentuale di ore fruite rispetto alle ore autorizzate. Tale tasso, quindi, applicato alle ore autorizzate (per definizione già al netto dei dirigenti) restituisce una misura delle ore fruite di Cig delle unità stimate con dati amministrativi (NO_GI):

$$NO_GI \text{ Ore fruite } cig_t = (NO_GI \text{ Ore autorizzate } cig_t \cdot Tiraggio) \quad (5)$$

Per rendere più snello il calcolo delle Ula netto Cig totali, piuttosto che calcolare tale indicatore solo per le NO_GI e aggregarlo successivamente alle GI, viene calcolato passando per le Ula in Cig dell'intera popolazione (GI+NO_GI), le quali si ricavano dalle ore fruite complessive. Ai fini di questo passaggio, quindi, è necessario calcolare, per le unità GI, le ore fruite di Cig aggregate per domini di pubblicazione (sezione Ateco), a partire dai dati forniti dalla rilevazione GI, quali le Ula totali complessive e le Ula in Cig. La differenza tra le due misure restituisce le Ula in Cig:

$$cig,GI \text{ Ula}_t = (GI \text{ Ula}_t - netto\ cig,GI \text{ Ula}_t) \quad (6)$$

Le ore fruite di Cig per le GI vengono ottenute moltiplicando le Ula in Cig per il numero di ore lavorabili nel trimestre, ricavate utilizzando le ore giornaliere lavorabili medie nel trimestre fornite dalla rilevazione IRC ($IRC \text{ Ore}_g$), al pari della formula (4). In formula:

$$GI \text{ Ore fruite } cig_t = (cig,GI \text{ Ula}_t \cdot IRC \text{ Ore}_g) \quad (7)$$

dove:

$${}_{IRC}Ore_t = ({}_{IRC}Ore_{g_t} \cdot Giorni\ lavorabili) \quad (8)$$

Le ore di Cig fruite totali sono ottenute sommando le ore NO_GI e GI:

$$Ore\ fruite\ cig_t = ({}_{NO_GI}Ore\ fruite\ cig_t + {}_{GI}Ore\ fruite\ cig_t) \quad (9)$$

Come ricordato poco sopra, questo approccio viene utilizzato per i trimestri di serie storica successiva al 2008. Per l'intervallo temporale precedente, a seguito della parziale disponibilità di microdati sulla Cig autorizzata, il metodo macro è stato applicato in versione semplificata, utilizzando i macrodati sulle ore autorizzate di Cig tratti dalla bancadati Inps (Inps, 2016), classificati per codice statistico contributivo (una *proxy* dell'Ateco) e riportati in Ateco attraverso dei coefficienti di ripartizione. Per questo sottoperiodo, data la ridotta diffusione della Cig, le ore autorizzate sono state utilizzate senza applicazione di tassi di tiraggio, in base all'ipotesi che tutte le ore autorizzate siano state utilizzate. L'utilizzo delle ore macro non ha consentito di isolare la quota attribuibile alla sottopopolazione delle GI, alle quali è stato, quindi, necessario estendere questo metodo e non utilizzare le ore fruite disponibili dalla fonte statistica.

8.3.8 Calcolo degli indicatori al netto della Cig

In sintesi, il quadro di metodi/fonti di derivazione delle Ula al netto della Cig può essere così sintetizzato:

Tavola 8.3.1 - Metodo di calcolo delle Ula al netto della Cig per sottopopolazione di imprese, sezione di attività economica, periodo e tipo di intervento

METODI DI CALCOLO ULA NETTO CIG	Sottopopolazione	Sezione Ateco 2007	Periodo intervento	Tipo intervento
Algoritmo basato sullo sfruttamento delle: - ore autorizzate di Cig Inps macro per CSC - tiraggio Inps	PMI+GI_DA+INTER+GI	B, C, D, E, F, G, H, I, J	Da Q1:2000 a Q4:2007	Sui macrodati
Algoritmo basato sullo sfruttamento delle: - ore autorizzate di Cig micro, selezionate sulla popolazione <i>target</i> di Oros e aggregate per Ateco - ore fruite di Cig di fonte Vela-GI	PMI+GI_DA+INTER+GI	F, I, J, K, L, M, N P, Q, R, S	Da Q1:2008 Da Q1:2010	Sui macrodati
Algoritmo basato sullo sfruttamento di: - ore/giornate retribuite nei DM - ore autorizzate di Cig	PMI+GI_DA+INTER con almeno 1 ora di Cig autorizzata	B, C, D, E, G, H	Da Q1:2010	Sui microdati
A partire dalle ore fruite di Cig di fonte GI	GI	B, C, D, E, F, G, H, I, J, K, L, M, N	Da Q1:2008	Sui microdati

Fonte: Rilevazione Oros (occupazione, retribuzioni, oneri sociali)

Una volta disponibili le ore fruite di Cig da entrambe i metodi, i quali concludono le proprie operazioni integrando i dati NO_GI con i dati GI, a livello di sezione Ateco, è possibile stimare le Ula al netto della Cig per l'intera popolazione. Con le informazioni disponibili è necessario calcolare dapprima le Ula in Cig, ottenute rapportando le ore fruite di Cig di cui alla (9) con una stima delle ore lavorabili nel trimestre tratte da fonte IRC:

$${}_{cig}Ula_t = \frac{Ore\ fruite\ cig_t}{{}_{IRC}Ore_t} \quad (10)$$

Le Ula finali al netto della Cig sono, quindi, ottenute sottraendo le Ula in Cig a alle Ula al lordo della Cig:

$${}_{\text{netto cig}} Ula_t = ({}_{\text{lordo cig}} Ula_t - {}_{\text{cig}} Ula_t) \quad (11)$$

Con il passaggio dalla base 2005=100 alla base 2010=100, le modifiche metodologiche che hanno influenzato la stima delle Ula in Cig sono state traslate alla parte precedente della serie storica attraverso un'operazione di armonizzazione della stima delle ore di Cig utilizzate dalle imprese. Tale raccordo è stato effettuato ricorrendo ad una procedura analoga a quella impiegata nei *link* delle altre variabili. Gli indicatori al netto della Cig raccordati si ottengono sottraendo le Ula in Cig raccordate alle Ula al lordo di tale componente, anch'esse raccordate.

A questo punto si dispone di tutti gli elementi per calcolare i nuovi numeri indice. Come già illustrato nel paragrafo 2 (formula (1)), il calcolo degli indici viene effettuato rapportando numeratore (costo del lavoro, retribuzioni e oneri) e denominatore (Ula al netto Cig), disponibili per sezione Ateco 2007 rispettivamente per il trimestre t e l'anno base (b), per cui si effettua una media annua sull'anno base dei quattro trimestri. Successivamente, si procede al rapporto delle due misure ottenute. Facendo riferimento al costo del lavoro, l'indice può essere espresso come:

$${}_{\text{netto cig}} I_t^{CU} = \frac{\frac{c_t}{{}_{\text{netto cig}} Ula_t}}{\frac{1}{4} \sum_{tb=1}^4 \frac{c_{tb}}{{}_{\text{netto cig}} Ula_{tb}}} \quad (12)$$

Allo stesso modo, i calcoli vengono effettuati anche per le altre due variabili (retribuzioni e oneri) e per i livelli di aggregazione Ateco superiori (macro aggregati di diffusione).

Al fine di validare l'effetto della correzione per la Cig, infine, gli indicatori Oros così ottenuti vengono confrontati, a livello di sezione Ateco, con gli indicatori calcolati per il regolamento LCI (cfr. §11), espressi in termini di ore lavorate, una misura che per definizione è depurata dalla Cig. Inoltre, le retribuzioni vengono messe a confronto con le retribuzioni contrattuali di cassa per le quali, misurando esclusivamente le componenti retributive attribuibili alla contrattazione nazionale a struttura occupazione fissa all'anno base, non sono influenzate dall'andamento della Cig (sono per definizione al netto della Cig). L'esito di tale confronto, in qualche caso, induce a ridiscutere il valore del "tiraggio" fissato nel metodo macro. Basato, infatti, su un metodo più marcatamente euristico, tale metodo diviene tanto più robusto quanto più si rendono disponibili fonti ausiliare di supporto.

Il metodo micro, che sfrutta in modo più intensivo le informazioni sulle singole realtà aziendali, viene ritenuto preferibile in termini di qualità della stima ottenuta. E' in corso un'analisi finalizzata ad estenderne l'utilizzo ad altri settori in cui la dinamica in termini di Cig si è meglio definita nel corso del tempo.

8.4. La stima delle posizioni lavorative totali: dalla ricostruzione delle componenti mancanti alla predisposizione delle serie storiche

8.4.1 Il quadro informativo di riferimento

Secondo la definizione prevista dal regolamento STS della Commissione Europea sulle statistiche congiunturali (Reg. EC 1165/98), i dipendenti totali comprendono tutti i lavoratori che hanno un contratto di lavoro subordinato con l'impresa. In particolare, sono inclusi anche quei dipendenti assenti nel corso dell'intero mese per cause varie che non implicano una retribuzione da parte dell'impresa stessa (come ad esempio il collocamento in Cig e/o in solidarietà a zero ore, la maternità, la malattia non retribuita dal datore di lavoro, le aspettative, il congedo per cariche elettive, ecc.). Va, inoltre, ricordato che, come evidenzia il Regolamento STS, i dipendenti rilevati dalle indagini sulle imprese corrispondono ai "posti" di lavoro (*jobs*) e non necessariamente coincidono con le "teste".

Come già introdotto nel capitolo 4, il modello virtuale DM2013 riporta (nella sezione relativa ai debiti e crediti contributivi) le posizioni lavorative alle quali l'azienda ha corrisposto, nel mese di competenza, una retribuzione imponibile a fini contributivi (da qui in avanti dipendenti retribuiti o Diptot). Queste posizioni (che contribuiscono a determinare il costo del lavoro sostenuto dall'impresa) rappresentano, però, una sottostima delle posizioni lavorative richieste dal regolamento poiché non includono quelle assenti nel corso dell'intero mese per cause che non implicano una retribuzione da parte dell'impresa stessa. D'altra parte, per rispondere ad esigenze contributive, queste posizioni cumulano anche eventuali modifiche nel contratto di lavoro subordinato di uno stesso dipendente intervenute nel corso del mese (ad esempio un cambio di qualifica, di tempo di lavoro, ecc. di uno stesso lavoratore implica che la sua posizione lavorativa venga conteggiata due o più volte).

Oltre alla variabile Diptot, il DM2013 riporta il numero di dipendenti in forza dichiarati dall'azienda nel periodo cui si riferisce la denuncia (da qui in avanti Ndip), cioè il numero di dipendenti con i quali essa ha instaurato un rapporto di lavoro subordinato. A differenza della variabile Diptot, tra questi dipendenti sono conteggiati anche i lavoratori che, per diversi motivi, non hanno percepito da parte dall'azienda nessun tipo di retribuzione nel corso dell'intero mese. Si tratta, in particolare, dei dipendenti in Cig a zero ore e dei dipendenti non retribuiti per altri motivi tra i quali la maternità, la malattia non retribuita dal datore di lavoro, le aspettative, il congedo per cariche elettive, ecc.⁷ (da qui in poi Dip_nor).

Anche se dal punto di vista definitorio la variabile amministrativa Ndip aderisce in misura soddisfacente alla richiesta del regolamento Europeo, la sua finalità non contributiva (che non implica l'attuazione di controlli stringenti da parte dell'Inps) può ridurne la qualità

⁷ Nello specifico, tale variabile, ricostruita nel modello virtuale DM2013 a partire dalla corrispondente variabile presente nel quadro anagrafico (aziendale) dell'Uniemens, riporta il numero dei lavoratori in forza all'azienda nel periodo cui si riferisce la denuncia, ossia l'insieme dei lavoratori per i quali sono stati denunciati imponibili contributivi e coloro che non sono retribuiti, se non sostituiti (cfr. §4.2). Nel numero devono essere indicati i lavoratori assenti per servizio militare, aspettativa non retribuita, quelli che risultano assenti per tutto il mese a seguito di prestazioni previdenziali (es. cig, maternità, allattamento, malattia ecc.). Devono essere esposti anche i lavoratori indicati con codifica particolare e non soggetti a denuncia contributiva, come ad es.: lavoratori in aspettativa per funzioni pubbliche elettive, lavoratori in aspettativa per cariche sindacali, ecc.).

Si veda "Documento tecnico per la compilazione dei flussi delle denunce retributive e contributive individuali mensili". Inps, Direzione Centrale Sistemi informativi e Telecomunicazioni.

<https://www.inps.it/nuovoportaleinps/default.aspx?itemdir=46319>.

in termini di accuratezza e completezza. Per verificare la corrispondenza tra la definizione amministrativa e il contenuto effettivo di questa variabile è stata condotta un'approfondita analisi sia *cross-section* sia longitudinale. L'analisi ha evidenziato particolari situazioni in cui la variabile presenta valori anomali⁸ ed elevata variabilità, in alcuni casi anche per l'inclusione di categorie di lavoratori non rientranti nella forza lavoro aziendale⁹. Tali evidenze hanno suggerito di non utilizzare Ndip come misura esclusiva per le posizioni lavorative totali richieste dal Regolamento.

Per i motivi appena descritti, la stima delle posizioni lavorative dipendenti totali viene effettuata seguendo una metodologia che utilizza i dipendenti retribuiti, applicando ad essi fattori correttivi calcolati sulla base dei dipendenti in forza, variabile che, includendo i dipendenti non retribuiti dall'impresa, permette l'allineamento alla definizione da regolamento. Inoltre, considerata la diversa evoluzione congiunturale dei motivi di assenza, i fattori correttivi vengono stimati con metodologie diverse per: 1) i dipendenti assenti per collocamento in Cig e/o in solidarietà a zero ore nel mese; 2) i dipendenti assenti per altri motivi che superano il mese (malattia, maternità e altro). Le metodologie utilizzate sfruttano la relazione tra Ndip e Diptot congiuntamente ad informazioni ausiliarie sulla potenziale presenza di dipendenti in Cig nelle unità in stima, informazioni ricavate dai microdati Inps sulle ore autorizzate mensili di Cig e solidarietà (cfr.§8.3.4). In particolare, a partire dai file Inps che riportano liste trimestrali di matricole aziendali per ore autorizzate di Cig, attraverso una redistribuzione temporale per tempi teorici di fruizione di Cig, le unità vengono classificate in unità potenzialmente utilizzatrici di Cig (d'ora in poi, sottopopolazione cig) o meno (sottopopolazione no-cig) per ogni trimestre di stima¹⁰, operazione necessaria per individuare con maggiore accuratezza i dipendenti non retribuiti per altri motivi di assenza evitando, in tal modo, sovrapposizioni con i dipendenti assenti perché in Cig.

Dal punto di vista formale, la variabile obiettivo posizioni lavorative totali, D , può essere definita come segue:

$$D = {}_rD + {}_{nr_altro}D + {}_{nr_cig}D \quad (13)$$

in cui i singoli addendi rappresentano:

${}_rD$ = il monte delle posizioni dipendenti retribuite, corrispondente a Diptot;
 ${}_{nr_altro}D$ = il monte delle posizioni dipendenti non retribuite assenti per altri motivi;
 ${}_{nr_cig}D$ = il monte delle posizioni dipendenti non retribuite assenti per Cig e/o solidarietà a zero ore.

Nei due paragrafi che seguono vengono descritte le soluzioni adottate nella rilevazione Oros per stimare i dipendenti associati ad entrambe le cause di assenza. E' utile sottolineare

8 Tra i valori anomali si riscontrano frequentemente casi di errori di misura caratterizzati da valori della variabile Ndip multipli con fattore pari a 100 della variabile Diptot. Questi casi, corretti da una procedura automatica che riporta il loro valore a quello di Diptot, non influiscono pertanto sulla valutazione finale relativa all'uso della variabile.

9 Ne sono un esempio i lavoratori autonomi dello spettacolo e gli sportivi professionisti i quali, per prassi Inps, devono dichiarare i propri contributi attraverso l'azienda presso la quale prestano la loro attività, non potendo esporli e pagarli autonomamente. Questi lavoratori sono conteggiati tra gli Ndip pur non rientrando nella forza lavoro aziendale (cfr. Messaggio Inps n. 1092/2017).

10 La costruzione di tale lista verrà descritta con maggior dettaglio nel successivo paragrafo 8.4.3 ed è comunque differente dalla lista individuata per la correzione delle Ula per l'effetto della Cig secondo il metodo micro, come descritto nel paragrafo 8.3.6.

are che le procedure di cui si parlerà riguardano solo i dati delle unità derivanti dalla fonte amministrativa (NO_GI) poiché per le unità GI i dipendenti non retribuiti sono dal punto di vista definitivo già inclusi nei dati rilevati.

8.4.2 Le posizioni lavorative assenti per aspettative di breve durata ed altri motivi

Come già accennato nel paragrafo precedente, la stima delle posizioni lavorative dipendenti non retribuite dal datore di lavoro assenti per altri motivi tra i quali la maternità, la malattia, le aspettative, il congedo per cariche elettive, ecc. (${}_{nr_altro}D_t$), utilizza una metodologia che riproporziona l'informazione relativa alle posizioni lavorative retribuite attraverso l'applicazione di un fattore di correzione basato sulla relazione tra i dipendenti in forza e le posizioni lavorative retribuite. In particolare, per un generico aggregato settoriale, le posizioni lavorative dipendenti non retribuite assenti per altri motivi, nel generico trimestre t , possono essere espresse nel modo seguente:

$${}_{nr_altro}D_t = {}_rD_t \times {}_{altro,no-cig}K_t \quad (14)$$

dove:

$${}_{altro,no-cig}K_t = \left(\frac{{}_{no-cig}Ndip_t}{{}_r,no-cigD_t} - 1 \right) \quad (15)$$

rappresenta il fattore correttivo delle posizioni lavorative dipendenti retribuite ed è funzione del rapporto tra le posizioni dipendenti totali (${}_{no-cig}Ndip_t$) e le posizioni dipendenti retribuite (${}_{r,no-cig}D_t$), del generico aggregato settoriale, relative alla sottopopolazione no-cig.

L'ipotesi implicita nella formula è che la differenza tra il numeratore ed il denominatore del fattore correttivo dipenda esclusivamente dalle posizioni lavorative non retribuite assenti per altri motivi. La scelta di calcolare tale fattore nella sottopopolazione complementare a quella delle unità in Cig (unità no-cig), pur rappresentando una prima misura di controllo in tal senso non è, tuttavia, sufficiente a garantire l'assenza completa di unità che siano ricorse alla cassa integrazione a zero ore (la presenza di queste unità indurrebbe un errore di sovra stima per effetto di una correzione cumulata per tutti i motivi di assenza operata dal fattore (${}_{altro,no-cig}K_t$)).

Un'ulteriore azione volta ad assicurare l'ipotesi di cui sopra, è stata quella di limitare il calcolo del correttivo ai soli anni che corrispondano ad una fase favorevole della congiuntura economica; in tali periodi, infatti, essendo il ricorso agli ammortizzatori sociali ridotto ai casi "ordinari" (es. Cigo nel settore delle costruzioni), si può ritenere che gli altri motivi di assenza rappresentino la componente mancante prevalente delle posizioni non retribuite. Per individuare con accuratezza il periodo più adeguato, è stata effettuata un'analisi in serie storica dei livelli mensili delle posizioni retribuite e totali e dei fattori di correzione, per divisione di attività economica Ateco 2007; in particolare, la scelta è ricaduta sugli anni dal 2007 al 2010. L'analisi ha evidenziato come l'andamento del fattore di correzione, sebbene calcolato nella sottopopolazione delle no-cig, sia comunque influenzato dal ricorso alla Cig mostrando valori più elevati a partire dal 2009, anno d'inizio della fase di crisi

del ciclo economico; inoltre, in molte divisioni¹¹, il fattore mostra una evidente periodicità infra annuale (a livello sia mensile sia trimestrale) seppure con alcune differenze tra gli anni in termini di regolarità.

In generale, l'analisi ha confermato valori dei coefficienti più contenuti negli anni 2007 e 2008 rispetto ai due anni successivi in cui la presenza di Cig potrebbe aver portato a una identificazione meno precisa della sottopopolazione considerata.

Le evidenze riscontrate, hanno portato a scegliere di utilizzare un fattore di correzione "costante" calcolato sui dati relativi agli anni 2007, 2008 e, solo in un caso, al 2010, da applicare alle posizioni lavorative retribuite in tutti i corrispondenti trimestri della serie storica. Nelle divisioni in cui è emersa una stagionalità regolare la scelta è ricaduta su coefficienti trimestrali. Questo approccio da un lato, consente di stimare in modo più accurato la componente mancata dell'occupazione depurandola dalla componente residuale attribuibile alla Cig (anche cogliendone, in alcune divisioni, l'andamento congiunturale) dall'altro, la scelta di un coefficiente costante assicura il controllo di eventuali effetti spuri derivanti dalla natura amministrativa della base dati. In particolare:

$$altro, no-cig \bar{K} = \left(\frac{no-cig N d i p_{\Delta}}{no-cig D_{\Delta}} - 1 \right) \quad (16)$$

rappresenta il fattore di correzione calcolato, per il generico aggregato settoriale, come rapporto medio tra il monte delle posizioni lavorative dipendenti totali e il monte delle posizioni lavorative dipendenti retribuite relativi ad un dato intervallo temporale Δ indipendente dal tempo e funzione soltanto dell'aggregato settoriale. I coefficienti, definiti dalla (16), sono calcolati a livello macro di divisione Ateco 2007 e applicati trimestralmente, sempre a livello macro, per derivare le posizioni assenti per motivi diversi dalla Cig sia nelle stime a regime sia nella serie storica (cfr. §8.4.6)¹²; inoltre, gli stessi coefficienti si utilizzano a livello micro per migliorare la stima delle posizioni in Cig (cfr. §8.4.3). Le scelte effettuate¹³, sono riassunte nella seguente Tavola 8.4.2.1:

11 Ad esempio nei settori fortemente femminilizzati, come quello dell'istruzione o delle industrie alimentari (rispettivamente divisione 10 e 85 dell'Ateco 2007), si riscontrano incrementi delle posizioni non retribuite assenti in corrispondenza dei mesi estivi in cui probabilmente è maggior il ricorso ad aspettative non retribuite da parte delle lavoratrici per dedicarsi alla cura dei figli.

12 Gli stessi coefficienti potrebbero essere applicati anche ai micro dati per derivare i dipendenti assenti per motivi diversi dalla Cig per singola unità, poiché ipotizzati costanti per settore di attività economica. Per motivi di praticità, al momento la loro applicazione viene effettuata ai dati aggregati. Tuttavia, per alcune specifiche forniture, essi vengono ricostruiti per posizione contributiva.

13 Con particolare riferimento alla divisione 93, l'assenza di una correzione deriva da una sovra copertura dei dati amministrativi sul fenomeno derivante dalla presenza in questo settore di personale che svolge attività alle dipendenze poco continuativa nel corso dell'anno (personale afferente a società sportive o d'intrattenimento).

Tavola 8.4.2.1 - Criteri per l'attribuzione del coefficiente di correzione

DIVISIONI ATECO 2007	Intervallo temporale (Δ)	Numero di coefficienti per divisione
5, 6, 7, 8, 9, 11, 12, 16, 17, 18, 19, 20, 21, 24, 26, 27, 28, 29, 30, 31, 33, 35, 36, 37, 38, 39, 45, 50, 61, 62, 64, 65, 71, 72, 75, 77, 78, 80, 87, 91, 95	Anni 2007 e 2008	1
42, 51, 58, 59, 60, 73	Anno 2008	1
92	Anno 2009	1
53	Anno 2010	1
10, 13, 14, 15, 23, 32,41, 43,46, 47, 49, 52, 56, 63, 66, 67, 68, 69, 70, 74, 79, 81, 82, 85, 86, 88, 90, 96	Trimestri 2007 e 2008	4
22, 25	Trimestri 2007	4
55	Trimestri 2008	4
93	Nessuna correzione (coefficiente=1)	0

Fonte: Rilevazione Oros (occupazione, retribuzioni, oneri sociali)

La Tavola 8.4.2.2, riporta la stima della quota media per l'anno 2010, delle posizioni dipendenti non retribuite assenti per altri motivi per sezione Ateco 2007 e fornisce una misura dell'entità della correzione. L'incidenza è calcolata sull'occupazione complessiva di Oros, inclusiva delle GI.

I coefficienti più elevati si registrano nelle sezione I (attività dei servizi di alloggio, divisione 55, e ristorazione, divisione 56) e nella sezione P (Istruzione) dell'Ateco 2007. Per entrambe le sezioni, in cui l'occupazione ha un comportamento tipicamente stagionale, il coefficiente a livello di divisione coglie la periodicità infra annuale dell'occupazione totale mostrando valori più alti proprio quando la componente retribuita scende (es. personale dei servizi di alloggio durante i mesi di flusso turistico ridotto).

Tavola 8.4.2.2 - Incidenza delle posizioni dipendenti non retribuite assenti per altri motivi sull'occupazione complessiva per sezione di attività economica e peso della sottopopolazione delle PMI in Oros
Anno 2010, valori medi percentuali

SEZIONI ATECO 2007	Incidenza delle posizioni dipendenti non retribuite assenti per altri motivi	Peso della sottopopolazione delle PMI Oros
B - Estrazione di minerali da cave e miniere	0,4	86,0
C - Attività manifatturiere	0,6	80,7
D - Fornitura di energia elettrica, gas, vapore e aria condizionata	0,2	33,1
E - Fornitura di acqua; reti fognarie, i gestione dei rifiuti e risanamento	0,5	73,0
F - Costruzioni	1,1	97,0
G - Commercio all'ingrosso e al dettaglio, riparazione di autoveicoli e motocicli	0,6	81,6
H - Trasporto e magazzinaggio	0,9	60,1
I - Attività dei servizi di alloggio e di ristorazione	3,9	86,6
J - Servizi di informazione e comunicazione	1,0	62,2
K - Attività finanziarie ed assicurative	0,4	34,0
L - Attività immobiliari	1,4	100,0
M - Attività professionali, scientifiche e tecniche	1,1	90,6
N - Noleggio, agenzie di viaggio, servizi di supporto alle imprese	1,3	61,4
P - Istruzione	2,5	100,0
Q - Sanità e assistenza sociale	1,1	100,0
R - Attività artistiche, sportive, di intrattenimento e divertimento	0,9	100,0
S - Altre attività di servizi	1,2	100,0

Fonte: Rilevazione Oros (occupazione, retribuzioni, oneri sociali)

8.4.3 Le posizioni lavorative assenti per Cig a zero ore

Al fine di completare la misurazione della variabile posizioni lavorative dipendenti totali, occorre aggiungere un'ulteriore componente di posizioni assenti dalla variabile amministrativa, i non retribuiti perché in Cig o solidarietà a zero ore nell'intero mese (in breve, in Cig). Oltre a rispondere alla necessità di integrare il contenuto definitivo della variabile, lo sviluppo e la messa in produzione della stima delle posizioni lavorative in Cig si colloca in un contesto di congiuntura economica particolarmente sfavorevole per il mercato del lavoro, in cui si è fatto ampio ricorso a questi strumenti: in tale contesto, statisticamente la stima di tale componente diviene non trascurabile non solo in termini di sottostima dei livelli ma anche di distorsione delle variazioni.

Al fine di stimare tale componente mancante è stata adottata una metodologia che opera un aggiustamento ai Diptot sfruttando, come per la stima dei dipendenti assenti per altri motivi (cfr. §8.4.2), il numero degli Ndip ma facendo ricorso ad un approccio estensivamente micro, in cui lo sfruttamento dell'informazione sui dati delle singole unità consente di cogliere in modo più fine la caratterizzazione delle singole imprese su questo aspetto.

Per scorporare da tale variabile la sola componente di non retribuiti per Cig vengono effettuati due passi:

- 1) viene stimata una lista trimestrale di unità potenzialmente caratterizzate da dipendenti in Cig a zero ore, rispetto a cui circoscrivere il calcolo degli assenti per Cig;
- 2) viene stimato un *coefficiente di correzione* da applicare ai dipendenti retribuiti che, una volta applicato, consente di quantificare i dipendenti assenti per Cig.

La stima della lista di unità con dipendenti in Cig a zero ore si basa sullo sfruttamento dei dati Inps sulle ore di Cig autorizzata per mese di autorizzazione e tipologia di strumento di integrazione salariale, fonte descritta nel paragrafo 8.3.4. Tali dati sono disponibili con la stessa tempestività dei dati sui DM2013 utilizzati per le stime provvisorie. Considerando la normativa in vigore per la richiesta di Cig, che prevede la preventiva autorizzazione di un determinato monte ore, di cui poter fruire in un arco variabile di tempo (cfr. §8.3.2), tali dati sono una rappresentazione distorta delle ore di Cig effettivamente fruita. Esse, infatti, fotografano le ore e le unità autorizzate a fruirne solo nell'istante di autorizzazione. Per contenere tale effetto di distorsione, i dati in questione vengono utilizzati con un criterio che tiene conto della durata massima della fruizione della Cig stabilita per legge. In particolare, si considerano in Cig tutte le unità che sono state autorizzate all'uso di questo strumento fino a due anni prima rispetto al trimestre di riferimento e fino ad un anno successivo (metodo lista "mobile"). Tale metodo è pienamente applicabile alla stima definitiva, per cui si dispone anche di dati successivi all'istante di riferimento. Per la stima provvisoria l'arco periodale può arrivare solo ai due anni prima, con un effetto di sottostima della lista che viene riassorbito in fase di revisione delle stime. Dalla definizione della lista sono escluse le unità dell'indagine GI, per cui le posizioni in Cig a zero ore sono già incluse nei dati rilevati.

Il secondo elemento che entra nella stima delle posizioni in Cig si basa sull'uso della variabile Ndip e del suo rapporto con i dipendenti retribuiti. L'uso della variabile, che nel processo Oros non era mai stata in precedenza necessaria per altri scopi, ha anzitutto implicato un lungo lavoro di analisi di qualità, sia in un'ottica cross-section (*editing* di errori di misura ed imputazione) che longitudinale, per validarne la continuità in serie storica. La procedura prevede la stima di un coefficiente che si basa sulla relazione tra la variabile Ndip e la variabile Diptot e viene calcolato a livello di divisione Ateco 2007 nel trimestre *t* sulla sottopopolazione delle unità cig, depurata per alcune categorie dalle unità con valori parti-

colarmente distorti della variabile $Ndip$ (le agenzie di somministrazione di lavoro interinale, le società cooperative e alcune matricole critiche con valori sistematicamente sovrastimati della variabile, squadrate). La relazione di base per il calcolo del coefficiente è la seguente:

$${}_{cig}K_t = \left(\frac{Ndip}{Diptot} - 1 \right)_t - ({}_{altro,no-cig}\bar{K} - 1) \quad (17)$$

Dal coefficiente di correzione viene esclusa la componente $({}_{altro,no-cig}\bar{K} - 1)$, indipendente dal tempo e caratterizzata per ogni generico settore j , che rappresenta la quota di assenti nel mese, tra le unità in Cig, dovuta ad aspettative di breve durata di cui alla (16).

Al fine di sfruttare al meglio le informazioni specifiche sull'incidenza della Cig per unità, il calcolo del coefficiente si basa su due diversi approcci:

1) un approccio *micro*, in cui il calcolo di (17) avviene a livello di singola unità e per ogni trimestre. Il coefficiente viene ritenuto valido quando assume valore maggiore all'unità. In caso contrario non viene attuata alcuna correzione;

2) un approccio *macro*, a livello di divisione Ateco 2007 (o raggruppamento di divisioni in caso di domini poco numerosi), in cui il coefficiente (17) viene calcolato come media ponderata dei rapporti micro, con ponderazione basata sui dipendenti retribuiti nel mese (Diptot). Dovendo essere successivamente applicato a livello di singola unità, vengono stabilite delle regole di plausibilità sui valori assunti dal coefficiente. Si esclude, anzitutto, la correzione su tutte le unità classificate nelle divisioni 36, 73 e 75 dell'Ateco 2007, che hanno evidenziato coefficienti sistematicamente bassi in un lungo periodo di analisi, tali da far presumere un ricorso alla Cig nelle imprese dei settori scarsamente significativo. Nelle restanti divisioni, il valore del coefficiente viene considerato plausibile se $< 0,97$, altrimenti posto pari a 1, portando ad assumere assenza di posizioni lavorative in Cig a zero ore.

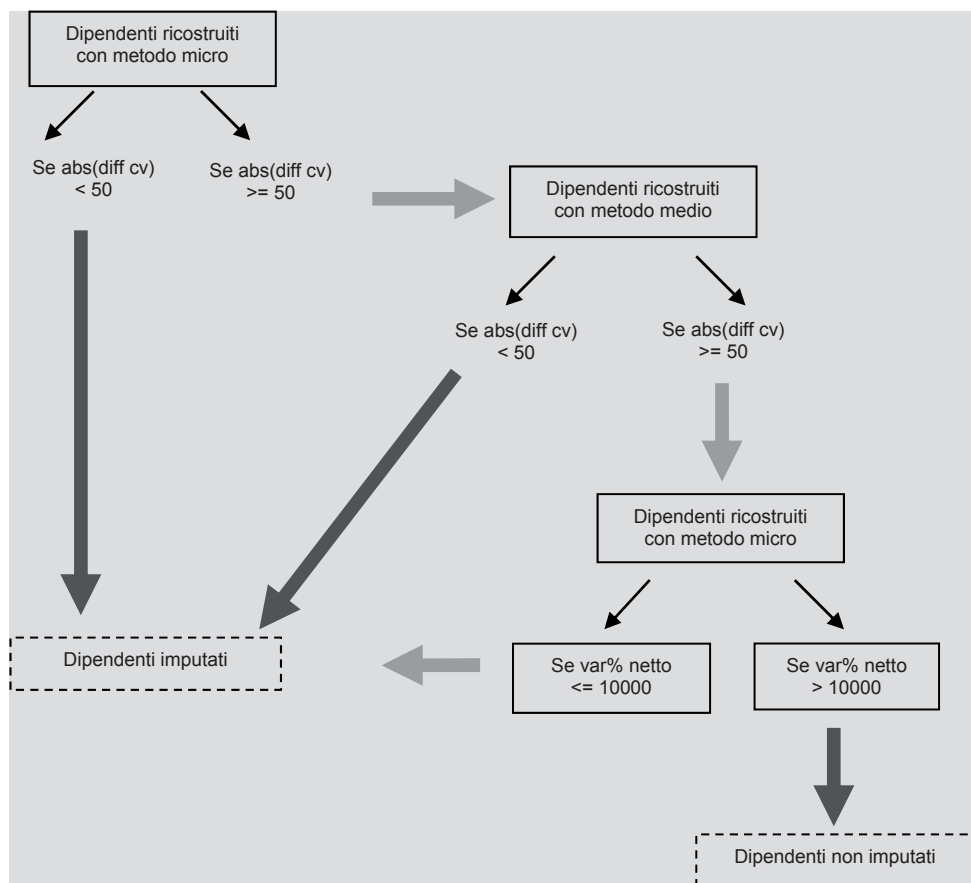
La stima delle posizioni lavorative assenti per Cig si ottiene, infine, applicando i coefficienti di correzione al numero di posizioni lavorative retribuite medie mensili nel trimestre delle unità in Cig secondo i criteri di cui sopra, sia nei microdati di stima definitiva, sia in quelli di stima provvisoria, eventualmente imputati per tener conto delle MRT (cfr. §6.5 e §6.2). Per ogni impresa i appartenente all'insieme delle NO_GI (unità da fonte amministrativa in stima secondo la rilevazione Oros e le agenzie di somministrazione di lavoro interinale, per cui si ipotizza Cig trascurabile) nel trimestre corrente t l'applicazione è data da:

$${}_{cig}d_{i,t} = {}_r d_{i,t} \times {}_{cig}k_t \quad \text{con } t \geq I:2008 \quad (18)$$

dove: ${}_r d_{i,t}$ sono i dipendenti retribuiti. L'applicazione del coefficiente (17) nella versione micro e macro segue una regola gerarchica secondo cui è preferibile il coefficiente calcolato sulla singola unità (approccio micro), più informativo sulla presenza di Cig in quell'unità e, solo in casi di risultati non plausibili, si ricorre al coefficiente calcolato a livello di divisione, a sua volta rigettato o accettato secondo soglie di ammissibilità. La plausibilità del dato ricostruito secondo i due approcci viene valutata sulla base di un criterio che analizza la stabilità dei dati al lordo della Cig in un'ottica longitudinale: sulla serie storica dei dipendenti al lordo della Cig viene calcolato un coefficiente di variazione, sottoposto a soglie di accettabilità. La regola utilizzata prevede che la differenza tra il coefficiente di variazione calcolato sugli ultimi 8 trimestri più l'ultimo dato imputato in prima istanza con il metodo micro e quello calcolato sugli ultimi 8 trimestri escluso quello di stima, non debba superare in valore assoluto la soglia di 50 (Figura 8.4.3.1). In caso contrario, il valore ottenuto con

approccio micro viene ritenuto non plausibile, per cui si passa a valutare lo scarto tra i coefficienti di variazione nell'ipotesi in cui il dato sia imputato con approccio medio, utilizzando la stessa soglia¹⁴.

Figura 8.4.3.1 – Il procedimento per l'individuazione degli *outlier* nella stima dei dipendenti in Cig a zero ore e scelta del dato finale



Nonostante la sostituzione dei dati imputati con approccio medio, rimangono dei casi residuali per cui la soglia viene superata. La plausibilità di questi casi viene valutata ricorrendo ad una regola aggiuntiva che prevede l'accettazione della ricostruzione micro solo per le unità i cui dipendenti in Cig ricostruiti non superino del 10.000% i dipendenti pre correzione (regola dell'incidenza della Cig). Anche questa soglia è stata stabilita sulla base della valutazione delle distribuzioni dei valori ricostruiti con il metodo micro e medio, inclusivi dei casi critici. La regola prevede che se i dipendenti imputati con approccio micro superano del 10.000% i dipendenti pre-imputazione allora si utilizza l'approccio medio, purché anche tali dipendenti imputati non superino la stessa soglia. Se entrambe sono al di fuori della soglia prestabilita l'imputazione non viene effettuata.

¹⁴ Tale regola viene applicata dal I trimestre 2010, a partire da cui sono disponibili 8 trimestri per ricostruire il coefficiente di variazione nel modo descritto. Per il 2009 il calcolo viene effettuato solo su 4 trimestri. Per il 2008 si utilizza una regola alternativa, che verrà di seguito descritta. La scelta della soglia di 50 è stata basata sull'analisi della distribuzione degli scarti tra coefficienti di variazione, per cui è stato selezionato il valore che porta a tagliare la soglia estrema della distribuzione (1% dei casi, molti dei quali evidenti *outlier*). Prima di arrivare alla scelta di 50 è stata effettuata una simulazione sul valore soglia di 30.

La stima dei dipendenti in Cig è disponibile a livello di micro dato dal I trimestre 2008, data a partire da cui la base di micro dati Oros risulta, dal punto di vista metodologico, delle classificazioni, ecc., coerente in serie storica (Istat, 2013). Per poter attuare la correzione anche sui trimestri precedenti, si è lavorato a livello aggregato per ogni divisione Ateco2007, applicando alla serie storica non corretta per la Cig (D_t) un coefficiente di *link* (λ) calcolato per ogni divisione Ateco 2007 solo sul I trimestre 2008 (un trimestre di Cig ancora molto contenuta) come rapporto tra la stima dei dipendenti inclusivi della Cig (${}_{\text{lordo cig}}D_{I:2008} = {}_rD_{I:2008} + {}_{\text{cig}}D_{j,I:2008}$) secondo la nuova metodologia e i dipendenti al netto Cig:

$$\lambda = \frac{{}_{\text{lordo cig}}D_{I:2008}}{{}_rD_{I:2008}} \quad (19)$$

per cui:

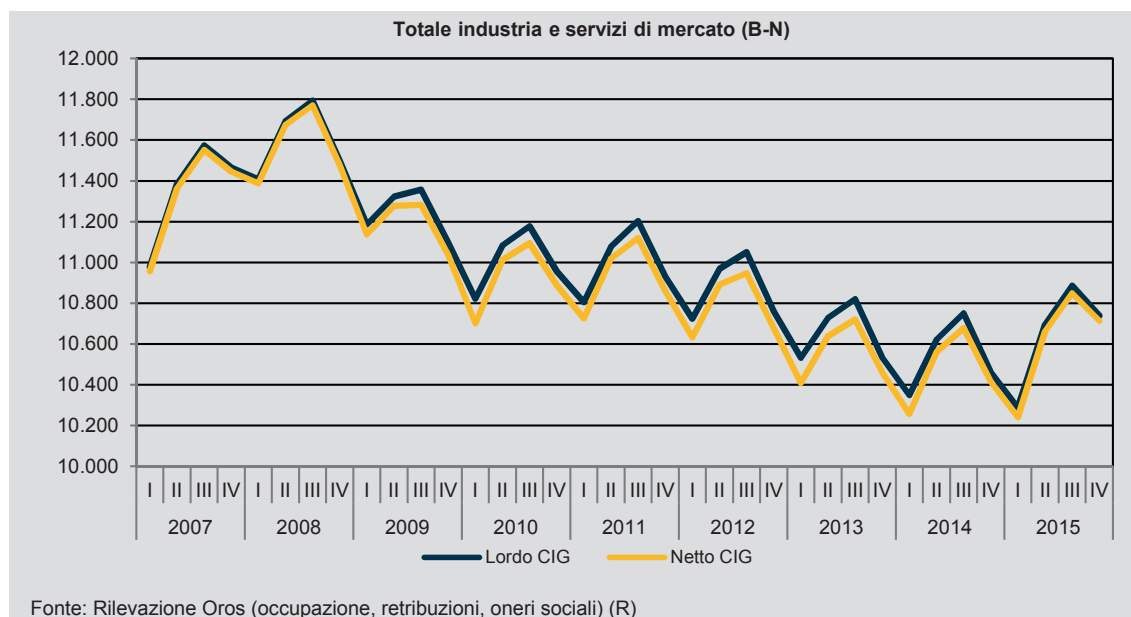
$${}_{\text{cig}}D_t = {}_rD_t \lambda \quad \text{dove } t < I:2008 \quad (20)$$

L'applicazione di tale coefficiente su tutta la serie storica I:2000-IV:2007 equivale ad assumere che l'incidenza della Cig sia stata di entità costate in tali trimestri e indipendente da fattori stagionali.

Come verrà spiegato con maggior dettaglio in seguito, inoltre, al fine di neutralizzare gli effetti dei cambiamenti di classificazione delle unità nel tempo, alle posizioni in Cig precedenti il 2012 è stato applicato un correttivo finalizzato a ricollocare i flussi di dipendenti tra i settori causati dalle riclassificazioni (cfr. §8.4.7).

Il Grafico 8.4.3.1 che segue mette a confronto la stima dei dipendenti al lordo e al netto della Cig nei settori dell'industria e dei servizi di mercato (sezioni da B a N Ateco2007) a partire dal primo trimestre 2007.

Grafico 8.4.3.1 – Posizioni lavorative dipendenti al lordo e al netto della Cig a zero ore a seguito della correzione
I trimestre 2007 - IV trimestre 2015, valori assoluti in migliaia



L'impatto della Cig, pari a 0,2% in media annua fino al 2008, cresce fino al 2013 quando, in termini di incidenza sui dipendenti totali, assume i valori maggiori (0,9%), per ridursi negli ultimi due anni di riferimento (nel 2015 pari a 0,3%) (tavola 8.4.3.1).

8. Il calcolo degli indicatori finalizzati alla diffusione Nazionale

L'impatto della correzione si diversifica notevolmente tra i settori economici, con una maggiore incidenza nell'industria (sezioni da B a F Ateco2007) e, all'interno di questo macro aggregato, nelle attività manifatturiere e nelle costruzioni e, negli ultimi 4 anni del periodo considerato, anche nell'estrazione di minerali da cave e miniere¹⁵.

Tavola 8.4.3.1 – Incidenza dipendenti in Cig a zero ore sulle posizioni lavorative dipendenti per settore di attività economica
Anni 2007-2015, valori medi percentuali

SETTORI ATECO 2007	2007	2008	2009	2010	2011	2012	2013	2014	2015
INDUSTRIA E SERVIZI (B-S escluso O)	—	—	—	0,7	0,6	0,8	0,9	0,6	0,3
INDUSTRIA E SERVIZI DI MERCATO (B-N)	0,2	0,2	0,5	0,8	0,7	0,8	0,9	0,7	0,3
INDUSTRIA (B-F)	0,4	0,4	1,0	1,5	1,2	1,5	1,7	1,2	0,7
INDUSTRIA IN SENSO STRETTO (B-E)	0,3	0,3	1,1	1,5	1,2	1,4	1,4	1,0	0,5
B - Estrazione di minerali da cave e miniere	0,5	0,3	0,5	0,8	0,7	1,6	2,2	1,7	1,0
C - Attività manifatturiere	0,3	0,3	1,2	1,6	1,3	1,5	1,5	1,1	0,5
D - Fornitura di energia elettrica, gas, vapore e aria condizionata	0,0	0,0	0,0	0,1	0,1	0,2	0,2	0,2	0,1
E - Fornitura di acqua; reti fognarie, gestione dei rifiuti e risanamento	0,2	0,1	0,1	0,3	0,3	0,3	0,5	0,2	0,1
F - Costruzioni	0,5	0,5	0,7	1,4	1,2	1,9	2,7	2,0	1,2
SERVIZI DI MERCATO (G-N)	0,0	0,0	0,1	0,2	0,2	0,3	0,3	0,3	0,1
G - Commercio all'ingrosso e al dettaglio, riparazione di autoveicoli e motocicli	0,0	0,0	0,1	0,2	0,2	0,4	0,4	0,3	0,1
H - Trasporto e magazzinaggio	0,0	0,1	0,1	0,2	0,3	0,3	0,2	0,2	0,1
I - Attività dei servizi di alloggio e di ristorazione	0,1	0,0	0,1	0,2	0,2	0,3	0,3	0,2	0,1
J - Servizi di informazione e comunicazione	0,1	0,1	0,2	0,4	0,4	0,4	0,5	0,4	0,2
K - Attività finanziarie ed assicurative	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
L - Attività immobiliari	0,1	0,1	0,1	0,3	0,6	0,5	0,7	0,7	0,6
M - Attività professionali, scientifiche e tecniche	0,0	0,0	0,1	0,3	0,3	0,3	0,4	0,3	0,1
N - Noleggio, agenzie di viaggio, servizi di supporto alle imprese	0,0	0,0	0,1	0,3	0,3	0,3	0,3	0,2	0,1
SERVIZI SOCIALI PERSONALI (P-S)	—	—	—	0,1	0,2	0,3	0,2	0,2	0,1
P - Istruzione	—	—	—	0,1	0,1	0,1	0,2	0,1	0,1
Q - Sanità e assistenza sociale	—	—	—	0,1	0,2	0,3	0,2	0,1	0,0
R - Attività artistiche, sportive, di intrattenimento e divertimento	—	—	—	0,1	0,2	0,3	0,3	0,2	0,1
S - Altre attività di servizi	—	—	—	0,2	0,2	0,4	0,3	0,3	0,1

Fonte: Rilevazione Oros (occupazione, retribuzioni, oneri sociali)

8.4.4. Validazione delle posizioni lavorative in serie storica: l'imputazione macro

A questo stadio del processo di produzione, i dati sulle posizioni lavorative totali disponibili a livello di divisione Ateco sarebbero pronti per diffusione. Tuttavia, alcuni aggiustamenti che interessano solo i trimestri di stima provvisoria sono finalizzati a completare l'intervento di correzione per le mancate risposte, attuato a livello di microdato, a seguito delle operazioni di imputazione per regressione (cfr. §6.2) e *editing* selettivo (cfr. §6.6). Tale aggiustamento, effettuato a livello aggregato (divisione Ateco), si basa su un'analisi in serie storica della variabile e di alcune relazioni ricorrenti tra indicatori legati al ritardo di risposta.

L'intervento si rende necessario per colmare alcune criticità degli approcci micro di imputazione delle posizioni lavorative, limiti che dipendono dalla regolarità nella struttura

¹⁵ Tale incidenza non è da intendersi quale impatto complessivo della Cig a zero ore nel settore, ma come entità della correzione inserita negli indicatori. Infatti, essa non include i dipendenti in Cig da fonte GI, che non necessitano di alcuna correzione e che sono già conteggiati nella componente dei dipendenti totali che deriva dall'integrazione con i dati di rilevazione.

dei dati sottostanti, sia in termini di arrivo dei DM, sia di andamento dell'occupazione: il metodo ha delle *performance* non soddisfacenti nei casi in cui vi sia mancata risposta persistente o notevole irregolarità nell'evoluzione dell'occupazione, sia all'interno dei *model groups*, sia in termini longitudinali. Ad illustrare il primo caso, la divisione 50 – trasporto marittimo e per vie d'acqua, in cui gli alti e persistenti tassi di ritardo di risposta sono il risultato di una deroga nella scadenza del pagamento dei contributi concesso alle imprese di navigazione¹⁶.

In qualche caso, l'entrata in vigore di una nuova disciplina sulla gestione dei contributi, ha comportato l'asestamento della situazione informativa con dei contraccolpi sulla disponibilità di dati per le stime provvisorie (divisioni della sezione R – attività artistiche, sportive, di intrattenimento e divertimento, settori prevalentemente afferenti alla gestione ex-Enpals, accorpata all'Inps da gennaio 2015). Il secondo caso è, invece, ben rappresentato dalla divisione 59 - attività di produzione cinematografica, di video e di programmi televisivi, di registrazioni musicali e sonore, un settore in cui l'occupazione subisce forti aumenti e drastiche riduzioni in relazione alla realizzazione di produzioni nel settore.

Per risolvere queste problematiche vengono applicati dei correttivi basati sulla regolarità dell'errore di revisione tra stima provvisoria e stima finale (caso della divisione 50), nonché di *proxy* di mancata risposta e ad altre covariate legate al livello dell'occupazione complessivo (es. dinamica dell'occupazione nella sottopopolazione delle GI di fonte Rilevazione).

L'incidenza dell'imputazione macro ha un impatto di circa 0,2% della stima provvisoria complessiva dell'occupazione, con un impatto lievemente superiore nei servizi. Nella divisione 50 il tasso di correzione si aggira intorno al 15%, e varia tra il 2% e il 15% nella divisione 59.

Allo stato attuale a tale approccio non è stata conferita una formalizzazione standardizzata, principalmente a causa del lento consolidamento della base informativa dei dati amministrativi che, come visto, sono stati interessati da frequenti cambiamenti (cfr. §2.3.2). Un possibile miglioramento, la cui applicazione è prevista nel prossimo futuro, riguarda la modellizzazione della relazione tra stima provvisoria e stima finale, una volta disponibile un adeguato numero di occorrenze temporali caratterizzate da un'adeguata stabilità della base dati su cui definire la relazione.

8.4.5 Le posizioni lavorative in somministrazione

Accanto agli indici e ai livelli sulle posizioni lavorative dipendenti per sezione Ateco 2007, la rilevazione Oros diffonde anche dati riferiti alle sole posizioni con contratto di somministrazione¹⁷. Tale componente della domanda di lavoro ha un potenziale informativo molto rilevante nella lettura della dinamica del mercato del lavoro e, più in generale, nell'analisi del ciclo economico essendo, la richiesta di posizioni in somministrazione, particolarmente sensibile alla congiuntura: le imprese, infatti, in risposta all'andamento della domanda di beni e servizi possono aggiustare con facilità e senza particolari costi il proprio input di lavoro, variando il numero di lavoratori in somministrazione richiesti alle agenzie

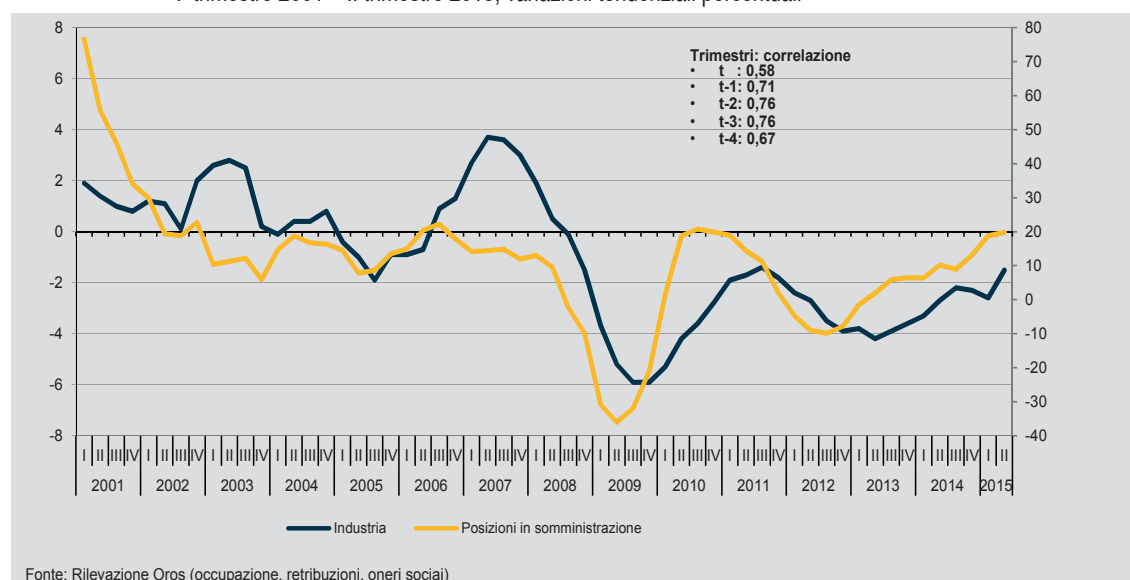
¹⁶ Si veda la Circolare Inps n. 179, 23 dicembre 2013 sulla gestione delle attività relative alla riscossione dei contributi e all'erogazione delle prestazioni per malattia, maternità, disabilità, donazione sangue per il personale assicurato ex-Ipsema e disposizioni legislative collegate.

¹⁷ Le posizioni in somministrazione non includono il personale delle agenzie fornitrici di lavoro temporaneo assunto con contratto di lavoro dipendente diverso dalla somministrazione.

8. Il calcolo degli indicatori finalizzati alla diffusione Nazionale

che forniscono questo servizio. Ciò avviene in particolare nell'industria, in cui il ricorso a forme atipiche di lavoro diverse dalla somministrazione è tradizionalmente meno diffuso che nei servizi. Tale variabile in diversi paesi viene considerata un *leading indicator* perché anticipa in maniera sistematica l'andamento dell'occupazione¹⁸. Anche in Italia la variabile ha queste caratteristiche, infatti, si osserva un'elevata correlazione con le posizioni lavorative dipendenti,¹⁹ (grafici 8.4.5.1 e 8.4.5.2) (Istat, 2015). In tali grafici è indicata anche la correlazione fra le posizioni in somministrazione e le posizioni totali, calcolata sia nel trimestre corrente (trimestre t) che nei trimestri passati fino a $t-4$. I risultati della correlazione esprimono valori massimi a due e tre trimestri di distanza nell'industria e a uno e due trimestri nei servizi e ciò è evidente anche dai punti di svolta.

Grafico 8.4.5.1 – Posizioni lavorative in somministrazione e posizioni lavorative nell'industria (B-F)
I trimestre 2001 – Il trimestre 2015, variazioni tendenziali percentuali



Nella rilevazione Oros, come indicato dai principali regolamenti sulle statistiche economiche, i lavoratori con contratto di somministrazione vengono rilevati tra i dipendenti delle agenzie di fornitura di lavoro interinale e non tra i dipendenti delle imprese utilizzatrici. Il gruppo di attività economica Ateco 2007 in cui è classificata questa tipologia di lavoro è "Attività delle agenzie di lavoro temporaneo (gruppo 78.2)" che rientra nella sezione N "Noleggiorie, agenzie di viaggio, servizi di supporto alle imprese". Le matricole relative al personale somministrato sono, nel 2015, circa 200, per poco più di 100 imprese, che detengono in media 220.000 posizioni lavorative dipendenti. Il loro peso occupazionale rappresenta approssimativamente il 93% dell'intera divisione 78 ("attività di ricerca, selezione e fornitura del personale") e il 21% della sezione N.

Nei dati Inps, le matricole relative alle agenzie ex-interinali vengono individuate attraverso dei codici assegnati per finalità amministrative. Mediante tali codici, a partire dal 2011, è possibile identificare con maggiore chiarezza il personale in somministrazione:

¹⁸ Si veda, ad esempio, il comunicato INSEE al link:

http://www.insee.fr/en/indicateurs/ind30/20150311/Crea_emplois_14T4_angl.pdf.

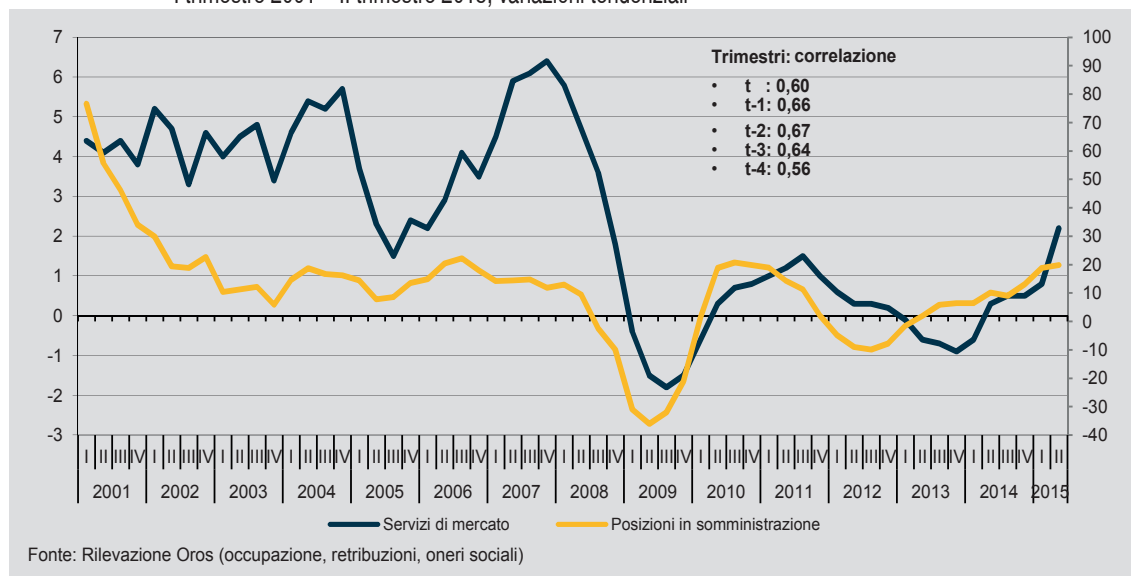
Anche CBS include tale variabile nella lista dei leading indicators. Si veda a proposito la nota metodologica:

<http://www.cbs.nl/en-GB/menu/themas/dossiers/conjunctuur/publicaties/artikelen/archief/2009/2009-10-15-f.htm>.

¹⁹ L'analisi di correlazione qui illustrata non considera i settori da P a S per la ridotta lunghezza delle serie storiche disponibili.

con circolare n. 149 del 24 novembre 2010 l'Inps ha, infatti, disposto che le agenzie di fornitura di lavoro interinale gestissero il proprio personale aprendo due distinte posizioni contributive: una destinata al personale in forza presso la struttura (il cosiddetto *staff*), caratterizzata da codice statistico contributivo pari a 7.07.08, ed una destinata ai lavoratori somministrati, caratterizzata dallo stesso codice contributivo e dal codice di autorizzazione contenente i caratteri 9A. Le matricole relative alle posizioni lavorative in somministrazione rappresentano, quindi, un sottoinsieme delle posizioni contributive riguardanti le agenzie di fornitura di lavoro interinale e per la loro dimensione d'impresa, necessitano di una procedura metodologica di riproporzionamento dei dipendenti (cfr. § 4.4.3) e d'imputazione delle mancate risposte (cfr. §6.3).

Grafico 8.4.5.2 – Posizioni lavorative in somministrazione e posizioni lavorative nei servizi di mercato (G-N)
I trimestre 2001 – II trimestre 2015, variazioni tendenziali



In occasione della prima uscita di indici sulle posizioni in somministrazione è stato necessario ricostruire l'intera serie storica dal I trimestre 2000, utilizzando dei criteri di selezione differenti nei vari periodi, a seconda dei dati disponibili. In particolare, per l'intervallo temporale I trimestre 2000 - IV trimestre 2007 gli unici dati disponibili, a livello aggregato, erano espressi in base 2000 e si riferivano all'intera forza lavoro delle agenzie di somministrazione, senza alcuna distinzione tra *staff* e personale somministrato. A partire dal I trimestre 2008 (prima uscita degli indicatori in base 2005) si è iniziato a distinguere il personale somministrato attraverso una selezione sui microdati utilizzando, in particolare, le variabili Inps relative al tipo contribuzione (cfr. § 4.2). Al fine di ottenere una stima del personale somministrato anche sulla prima parte della serie storica, si è proceduto applicando un coefficiente di raccordo, calcolato quale rapporto tra i livelli, al I trimestre 2008, delle variabili con e senza somministrati, espresse nelle due basi di riferimento. In tal modo, è stata ottenuta una serie storica per l'intervallo temporale dal I trimestre 2000 al IV trimestre 2009 in base 2005. A seguito del passaggio alla base 2010, invece, non si è reso necessario alcun raccordo in quanto la metodologia di selezione delle matricole e stima delle variabili è rimasta inalterata. Un *break* nella serie si è osservato, invece, tra il IV 2010 e il I 2011 a seguito dell'entrata in vigore del nuovo criterio di suddivisione delle matricole delle agenzie di somministrazione, dettato dall'Inps, come poco sopra ricordato. Tale modifica

8. Il calcolo degli indicatori finalizzati alla diffusione Nazionale

ha consentito di migliorare il metodo di individuazione dei lavoratori in somministrazione tra il personale delle agenzie di fornitura ma ha richiesto un intervento sull'intervallo temporale precedente, al fine di rendere la serie storica coerente. Il grafico 8.4.5.3 evidenzia un *level-shift* a partire dal I trimestre 2011 nell'andamento del rapporto tra le posizioni in somministrazioni e le posizioni totali delle agenzie di fornitura di lavoro. Per normalizzare l'andamento della serie è stato effettuato un raccordo utilizzando come coefficiente di *link* un valore pari a 0,94 ottenuto come media, per il 2011, dei rapporti trimestrali dei dipendenti somministrati sui dipendenti complessivi delle agenzie. L'adozione di una costante come *link* è motivata dalla regolarità nel tempo, a partire dal 2011, di tale rapporto.

L'effetto del *link* si può osservare nel grafico 8.4.5.3 in cui vengono messe a confronto le variazioni tendenziali delle posizioni in somministrazione raccordate e quelle del totale dipendenti delle agenzie di fornitura di lavoro interinale.

Grafico 8.4.5.3 – Posizioni lavorative totali delle agenzie di lavoro, posizioni lavorative del solo personale somministrato e loro rapporto

I trimestre 2008 – IV trimestre 2013, valori in migliaia (asse sx) e rapporti (asse dx)

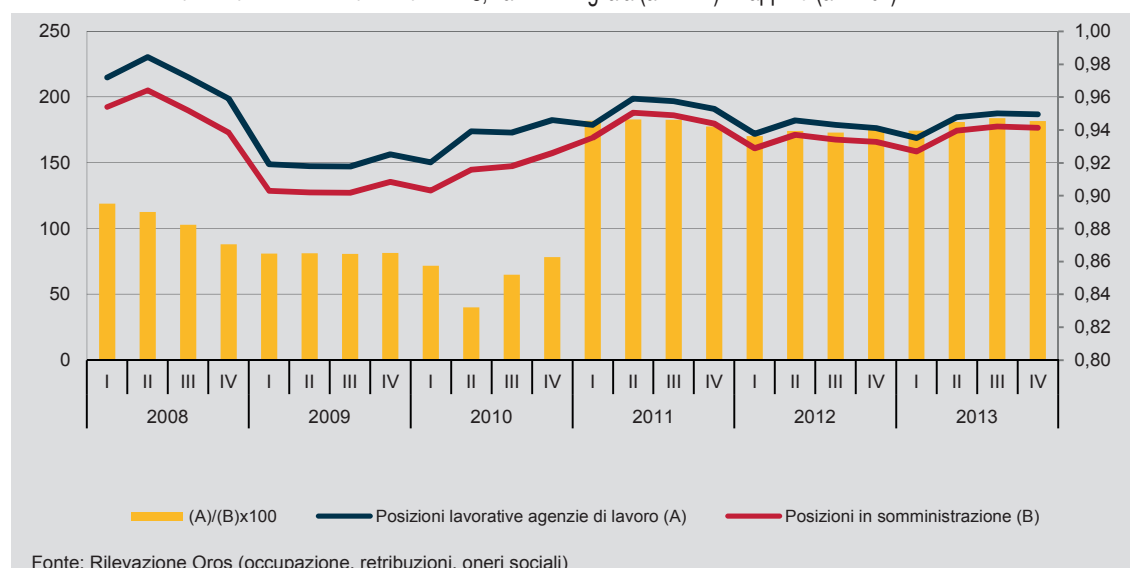
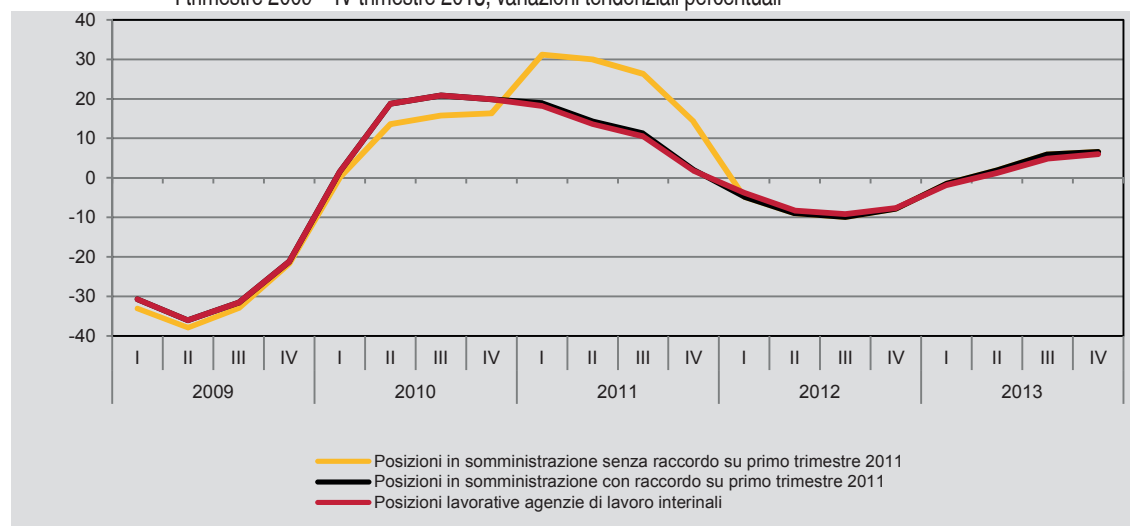


Grafico 8.4.5.4 – Posizioni lavorative totali delle agenzie di lavoro e posizioni lavorative del solo personale somministrato con e senza raccordo

I trimestre 2009 – IV trimestre 2013, variazioni tendenziali percentuali



8.4.6 Ricostruzione delle serie storiche delle posizioni lavorative dipendenti totali

Come evidenziato nel paragrafo 8.2, nelle stime a regime, le posizioni lavorative dipendenti totali (D_t) di Oros per il generico aggregato Ateco si ottengono per somma delle posizioni della sottopopolazione delle GI ($_{GI}D_t$) e della sottopopolazione delle unità stimate con dati amministrativi (NO_GI) ($_{NOGI}D_t$):

$$D_t = {}_{GI}D_t + {}_{NOGI}D_t \quad \text{per } t \geq I:2010 \quad (10)$$

dove per le NO_GI è necessario operare alcune aggiustamenti ai dati amministrativi sui dipendenti per rispondere alle definizioni statistiche, in particolare è necessario aggiungere ai dipendenti retribuiti ($_{r,NO_GI}D_t$) registrati sui DM, gli assenti nell'intero mese per Cig/solidarietà ($_{cig,NO_GI}D_t$) e/o altri motivi di assenza ($_{altro,NO_GI}D_t$):

$${}_{NOGI}D_t = {}_{r,NO_GI}D_t + {}_{altro,NO_GI}D_t + {}_{cig,NO_GI}D_t \quad \text{per } t \geq I:2010 \quad (11)$$

Le varie componenti delle relazione (10) e (11) possono essere calcolate ai livelli di dettaglio Ateco necessari, direttamente per aggregazione di microdati, a partire dal I trimestre 2010, data alla quale si ferma la ricostruzione all'indietro dei dati unitari secondo i criteri metodologici della base 2010 (cfr. §8.2.1).

Accanto alla stima a regime, come descritto nei paragrafi 8.4.2 e 8.4.3, la nuova variabile inclusiva degli elementi mancanti va ricostruita anche per i trimestri precedenti, a partire dal I trimestre 2000. Per tale intervallo temporale, tuttavia, la possibilità di utilizzare microdati per stimare tali componenti è fortemente limitata, a seguito del tipo di innovazioni metodologiche introdotte con il passaggio alle varie basi. In particolare, in occasione del passaggio alla base 2010, dati coerenti in serie storica sono disponibili ad un livello di dettaglio non inferiore alla divisione Ateco 2007 e non disaggregati per le due sottopopolazioni delle NO_GI e delle GI: le scelte metodologiche per la ricostruzione di tali componenti sono state pertanto fortemente condizionate dalla struttura di dati disponibili.

La serie storica dei dipendenti in Cig viene, anzitutto, ricostruita attraverso un procedimento *ad hoc*, già descritto nel paragrafo 8.4.3. La stima dei dipendenti assenti per altre motivazioni, come descritto nel paragrafo 8.4.2, viene effettuata attraverso un coefficiente di correzione (K_t) che si ipotizza stabile nel tempo e per divisione Ateco 2007 la quota di posizioni assenti per tali motivazioni. Tali assenti devono essere, tuttavia, scorporati da alcune categorie che non ne sono interessate: l'applicazione del coefficiente, generalizzato per divisione, anche a tali categorie porterebbe a una sovrastima delle posizioni totali. In particolare, devono essere esclusi i lavoratori dipendenti in somministrazione ($_{SOMM,NO_GI}D_t$) e i lavoratori a chiamata ($_{LAC,NO_GI}D_t$). In formule, per $t < 2010$:

$${}_{altro,NO_GI}D_t = ({}_{r,NO_GI}D_t - {}_{GI}D_t - {}_{SOMM,NO_GI}D_t - {}_{LAC,NO_GI}D_t) \times {}_{altro,no-cig}\bar{K} \quad (12)$$

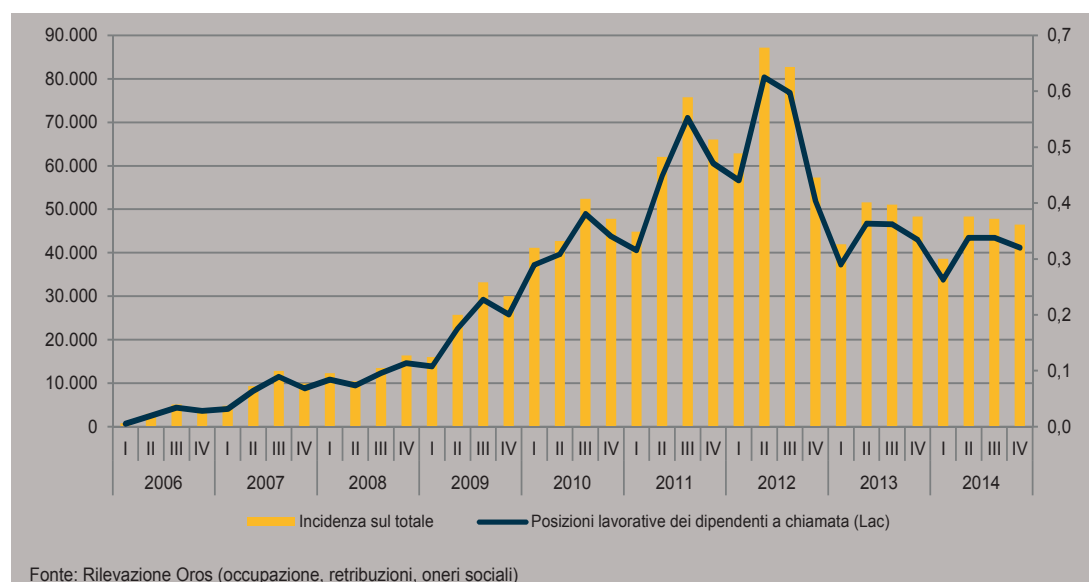
Per poter ricavare la serie storica dei dipendenti assenti per altre motivazioni è necessario ricostruire le serie storiche delle variabili ${}_{LAC,NO_GI}D_t$, ${}_{GI}D_t$ e ${}_{SOMM,NO_GI}D_t$.

La procedura utilizzata per la ricostruzione della serie storica sul numero di somministrati è stata già descritta nel paragrafo 8.4.5.

8. Il calcolo degli indicatori finalizzati alla diffusione Nazionale

Le posizioni lavorative dei dipendenti a chiamata presenti nelle stime Oros a partire dal I trimestre 2006, vengono ricondotte ad unità di lavoro a tempo pieno attraverso un'operazione di riproporzionamento (cfr. §4.4.3) che, fino al IV trimestre 2007, viene effettuata a livello di divisione. Tuttavia, per questo intervallo temporale, si dispone solo di stime calcolate secondo i criteri della base 2000, limitatamente ai settori dell'Alloggio e ristorazione, su cui il fenomeno era inizialmente concentrato. A partire dal I trimestre 2008, con il passaggio alla base 2005, il riproporzionamento è stato basato su un criterio applicato ai microdati ed esteso a tutti i settori di attività economica e (Istat, 2013a). L'introduzione di questo cambiamento ha reso necessario un primo raccordo, al fine di ricostruire serie storiche coerenti. Anche in questo caso i coefficienti per effettuare il *link* tra la base 2000 e la base 2005, sono stati calcolati a livello di divisione Ateco2007. Infine, per slittare indietro i cambiamenti introdotti con il passaggio alla base 2010, in cui i dati riproporzionati sui lavoratori a chiamata sono ancora disponibili a livello micro, è stato necessario applicare un ulteriore coefficiente di raccordo, calcolato a livello di divisione Ateco2007, rapportando i valori dell'anno 2010 in base 2005 e in base 2010. La ricostruzione della stima dei lavoratori a chiamata in base 2010, in breve, ha interessato il periodo che va dal I trimestre 2006 al IV trimestre 2009. Nel grafico 8.4.6.1 è illustrata la serie storica ricostruita dei Lac a titolo esemplificativo fino al IV trimestre 2014. Accanto ai livelli, si riporta la loro incidenza sul totale delle posizioni Oros nell'aggregato economico B-S.

Grafico 8.4.6.1 – Posizioni lavorative dei dipendenti a chiamata delle piccole e medie imprese e loro incidenza sul totale dei dipendenti
I trimestre 2006 – IV trimestre 2014, valori assoluti in migliaia (asse sx) e valori percentuali (asse dx)



Come già detto, gli assenti per altre motivazioni sono già inclusi nei dati GI, per cui, per stimare tale componente per la sola sottopopolazione delle NO_GI, è necessario scorporare dalla serie storica delle posizioni lavorative complessive (D_t) i dipendenti GI (GI_t). L'aggregato riferito alle sole GI, tuttavia, come descritto sopra è disponibile in serie storica solo a partire dal I trimestre 2010. Si rende, quindi, necessario ricostruire l'intervallo temporale precedente. Considerata la ridotta incidenza e variabilità nel tempo delle posizioni assenti per altre motivazioni, si è stabilito di adottare una soluzione

semplificata per la ricostruzione della serie storica dei dipendenti delle GI, funzionale solo al calcolo di tale componente di assenti. In particolare, è stato calcolato un coefficiente pari al rapporto, per l'anno 2010, tra i livelli delle posizioni lavorative GI e i livelli delle unità totali (GI+NO_GI). Tale valore è stato applicato alla serie storica dal I trimestre 2000 al IV trimestre 2009 delle unità nel complesso, ipotizzando implicitamente che il rapporto fra la struttura occupazione delle GI e delle unità restanti sia costante nel periodo temporale in oggetto:

$${}_{GI}D_t = \frac{{}_{GI}D_{2010}}{D_{2010}} \times D_t \quad \text{per } t \geq \text{I:2010} \quad (13)$$

Infine:

$$D_t = {}_{r,NO_GI}D_t + {}_{altro,NO_GI}D_t + {}_{cig,NO_GI}D_t \quad \text{per } t < \text{I:2010} \quad (14)$$

8.4.7 L'assegnazione delle informazioni di classificazione nella serie storica

Altro aspetto da affrontare nella ricostruzione delle serie storiche delle posizioni lavorative totali è l'assegnazione delle informazioni utili alla classificazione delle unità nel campo di osservazione della rilevazione. In questa operazione, come già ampiamente descritto nel capitolo 3, si rende necessario il blocco delle riclassificazioni delle unità tra un anno e l'altro, che potrebbero contaminare la lettura della congiuntura. Tali riclassificazioni derivano dal registro Asia, da cui vengono tratte le principali informazioni sulla natura giuridica ed istituzionale, nonché l'Ateco. Nel registro, i cambiamenti nelle informazioni strutturali delle unità possono essere dovuti a cambiamenti effettivi (trasformazioni giuridiche, modifiche nella natura istituzionale o cambiamenti nell'attività prevalente svolta) oppure apparenti (miglioramenti metodologici). Nelle serie storiche Oros tutti questi cambiamenti vengono bloccati a partire dalle occorrenze temporali del 2013, fissando le informazioni allo stesso anno 2013 nell'anagrafica della rilevazione e, quindi, a livello di microdato. Operando direttamente sull'anagrafica, questo intervento ha effetto su tutte le variabili *target* della rilevazione. Tuttavia, considerata la ridotta variabilità inter-settoriale dei valori pro-capite delle variabili di costo del lavoro, per l'intervallo temporale precedente il 2013 si è scelto di operare il blocco solo sulla stima aggregata dei livelli delle posizioni lavorative²⁰.

Per ricostruire serie storiche delle posizioni dipendenti totali depurate dagli effetti di riclassificazione dal I trimestre 2000 e fino al IV trimestre 2012 sono stati seguiti due diversi approcci, applicati a seconda della disponibilità di informazioni longitudinali dettagliate e coerenti:

1) per l'intervallo temporale 2010-2012 l'operazione è stata effettuata a partire dai microdati delle unità attive nel periodo, disponibili con caratteristiche omogenee a partire dal I trimestre 2010²¹;

2) per gli anni 2000 - 2009 l'operazione è stata realizzata a livello di macrodati, limitandosi alle sezioni Ateco 2007²².

²⁰ In breve, si è ipotizzato che le riclassificazioni passate non abbiano avuto effetto su livelli e dinamica delle variabili di costo del lavoro.

²¹ Con il passaggio alla base 2010 la coerenza dei dati in serie storica è stata realizzata partendo proprio dai micro-dati.

²² La ricostruzione delle serie storiche a livello di divisione è in programma per giugno 2018, anno in cui le stime Oros verranno rese disponibili in base 2015=100.

Nell'intervento di tipo 1) a tutte le unità presenti nel 2013 e già attive anche tra il 2010 e il 2012, sono state attribuite le informazioni rilevate nel 2013, anno di riferimento per il blocco. Alle unità attive tra il 2010 e il 2012 ma cessate prima del 2013, è stata applicata l'ultima informazione rilevata più vicina all'anno 2013. L'operazione di blocco ha interessato tutte le componenti che alimentano la stima finale dei dipendenti: al netto della Cig, in Cig e relativi alle mancate risposte dovute ai "DM anomali". Per determinare gli aggregati depurati dalle riclassificazioni sono stati stimati dei flussi, ossia differenze per divisione di attività economica tra stime originarie e stime bloccate. Tali flussi sono stati applicati a livello di divisione Ateco 2007 ai dati originari. L'impatto di questo intervento è praticamente nullo sull'aggregato B-S poiché le unità vengono spostate da un settore all'altro, prevalentemente nel campo di osservazione della rilevazione. I settori maggiormente caratterizzati sono la sezione L, dove le riclassificazioni interessano in media l'8% dell'occupazione complessiva (sottratta dalla serie originaria) e il settore S, in cui l'occupazione interessata da riclassificazione pesa per il 5% (aggiunta alla serie originaria).

Non disponendo di microdati riconciliati, l'intervento 2) applicato all'intervallo temporale 2000-2009 è stato effettuato direttamente sui macrodati omogenei in serie storica, partendo dalle serie per sezione Ateco 2007 già depurate dalle riclassificazioni secondo le operazioni di cui al punto 1). Le sezioni interessate dall'intervento sono state B, D, J, L e M. Obiettivo dell'operazione è quello di eliminare *level shifts* visibili dovuti a grandi riclassificazioni. Tali traslazioni sono state corrette e riportate indietro mettendo a confronto le dinamiche medie rilevate nei periodi di transizione tra i cambiamenti osservati (utilizzando variazioni congiunturali e tendenziali). Questa seconda operazione ha comportato una riduzione dei livelli della serie dell'aggregato B-N di circa 0,4%. I settori maggiormente interessati sono la sezione B (-13% in media nel periodo) e la sezione L (-9% in media nel periodo).

Dopo tale operazione tutti gli indicatori prodotti da Oros sono disponibili per pubblicazione, previo trattamento per gli effetti stagionalità là dove richiesto e creazione di formati standard. Tali operazioni vengono effettuate in SITIC, come verrà descritto nel paragrafo 8.5 che segue.

8.5 Gli indici Oros sul Sistema Informativo per il Trattamento degli Indicatori Congiunturali

SITIC è un applicativo Istat creato per la gestione e la conservazione degli indicatori congiunturali rilasciati dall'Istituto mediante I.Stat o attraverso i triangoli di revisione, nonché degli indicatori inviati ad Eurostat ad adempimento delle richieste dei Regolamenti²³. Questo sistema informativo è stato reso operativo a partire da metà 2009 per rispondere all'esigenza di gestire, secondo criteri di uniformità, gli output congiunturali diffusi all'esterno, dall'approccio utilizzato per il trattamento del dato per gli effetti di calendario e la stagionalità, al formato di rilascio, nonché la conservazione uniforme delle *vintage*. Gli indicatori di Oros hanno contribuito alla sperimentazione e messa a regime delle specifiche del sistema e, di fatto, vi sono entrati sin dal suo avvio.

SITIC acquisisce anzitutto gli indici grezzi puntando a tabelle ORACLE predisposte dalle singole Indagini (cfr. §2.4). Attraverso un sistema di tabelle relazionali per ogni indica-

²³ Per informazioni sul funzionamento del Sistema si veda il manuale tecnico, disponibile al link: http://sitic.istat.it/SITIC/sa/Documentazione/SITIC_ManualeUtente.pdf.

tore trattato definisce, secondo criteri standardizzati, livelli di classificazione (per attività economica, territorio ecc.), arrotondamenti a decimali, natura del dato (livello o indice), base di riferimento, approccio di destagionalizzazione seguito che, per il caso Oros, verrà dettagliatamente descritto nel capitolo 9, tipologia di aggregazione nel caso di metodo indiretto di derivazione degli aggregati, tipologia di formato da attribuire agli *output*. Inoltre, il Sistema prevede la correzione delle serie per gli effetti di calendario, applicando regressori standardizzati definiti dagli esperti di destagionalizzazione e il loro ribasamento secondo i criteri dettati dalla statistica ufficiale. A seguire, l'eventuale destagionalizzazione delle serie, utilizzando la metodologia della procedura TRAMO-SEATS. Il software per la destagionalizzazione è disponibile su un server Linux dell'Istituto in un'unica versione, concordata dagli esperti di destagionalizzazione, e lanciato in modalità batch da SITIC utilizzando i modelli definiti esternamente dall'utente. Infine, nel caso di approccio indiretto per la correzione per gli effetti di calendario e/o la destagionalizzazione e utilizzando i pesi caricati dall'utente, il Sistema deriva le serie aggregate trattate, mettendo a disposizione una molteplicità di output in formato testo e grafico. Alle serie grezze e/o trattate vengono attribuiti i formati di diffusione secondo le specifiche nazionali (formati I.Stat) ed europee (formati Gesmes e SDMX). SITIC permette di effettuare anche il caricamento delle serie predisposte per la diffusione nazionale direttamente sulla banca dati I.Stat.

Gli indicatori Oros sul costo del lavoro e sulle posizioni lavorative vengono aggiornati ogni trimestre su SITIC in forma grezza e con il livello di dettaglio necessario alla diffusione nazionale (sezioni e loro aggregazioni Ateco 2007). Gli indici vengono destagionalizzati secondo i modelli definiti e, in seguito, derivate le serie destagionalizzate indirettamente, aggregando gli indici trattati con approccio diretto attraverso pesi fissi caricati ex-ante. Mentre le serie grezze, nelle varie versioni rilasciate, vengono conservate in una tabella di base che consente di riconoscere la versione attraverso dei *flag* di ordine di revisione, le serie destagionalizzate vengono salvate in un'apposita tabella che conserva *vintage*. Gli indici grezzi e destagionalizzati vengono predisposti nel formato utile per il caricamento su I.Stat, passaggio che viene effettuato direttamente da SITIC all'istante di uscita degli indicatori su web. L'ultima versione di indici grezzi e destagionalizzati alimenta, infine, i triangoli di revisione, avviando la fase di calcolo e diffusione delle misure di revisione, argomento che verrà sviluppato nel capitolo 10.

Attraverso SITIC e secondo apposite specifiche vengono gestiti anche gli indicatori compilati da Oros per rispondere ai regolamenti LCI e STS, argomento che verrà sviluppato rispettivamente nei capitoli 11 e 12.

9. LA CORREZIONE PER LA STAGIONALITÀ E GLI EFFETTI DI CALENDARIO DEGLI INDICATORI OROS¹

9.1 La metodologia

Le serie storiche degli indicatori Oros sulle variabili di costo del lavoro e sulle posizioni lavorative vengono diffuse, oltre che come dati grezzi, anche come dati destagionalizzati, al medesimo livello di dettaglio per attività economica (sezioni e aggregati di sezioni). Non vengono, invece, diffuse le serie corrette per gli effetti di calendario, scarsamente significativi sulle variabili *target*.

Il procedimento di destagionalizzazione utilizzato è quello ufficiale adottato in Istituto che si basa sull'uso della procedura TRAMO-SEATS sviluppata da Gómez e Maravall (1996), nella versione per Linux. Alla base di TRAMO-SEATS vi è un approccio *model based* in cui la serie storica osservata viene concepita come la parte finita della realizzazione di un processo stocastico, la cui struttura probabilistica è adeguatamente descritta da un modello statistico parametrico (AAVV, 2000; AAVV, 2015). In particolare, TRAMO-SEATS si basa su modelli di tipo Arima (*AutoRegressive Integrated Moving Average*) che fanno ricorso a filtri lineari simmetrici, la cui espressione analitica dipende dal modello seguito dalla serie storica considerata e, quindi, risultano legati alle caratteristiche stocastiche della serie stessa.

In Oros, i modelli statistici utilizzati per la destagionalizzazione vengono di norma rivisti una volta all'anno, in corrispondenza della diffusione dei dati relativi al primo trimestre. Tuttavia, per permettere che l'aggiunta di una nuova informazione trimestrale possa comportare una migliore stima delle componenti non direttamente osservabili (ciclo-trend, stagionalità, componente irregolare), parametri e fattori stagionali vengono stimati ogni volta (approccio *partial concurrent*) comportando, ogni trimestre, la revisione di tutti i dati già pubblicati (Eurostat, 2015). La procedura di correzione per gli effetti di calendario, laddove significativi, viene operata con il metodo di regressione, il quale individua l'effetto del diverso numero di giorni lavorativi o di giorni specifici della settimana in essi contenuti, della presenza di festività nazionali civili e religiose, fisse e mobili (festività pasquali), nonché dell'anno bisestile attraverso l'introduzione di un regressore nel modello univariato che descrive l'andamento della serie (*regressore unico*). Tale regressore non rientra tra quelli automatici disponibili in TRAMO, che effettua la correzione della serie storica per gli effetti di calendario, ma è stato formalizzato e adottato come standard in Istituto ed è espresso come:

$$WD_t = GS_t - \frac{5}{2}GW_t \quad (1)$$

dove GS_t è il numero di giorni lavorativi (da lunedì a venerdì) nel periodo t e GW_t è il numero di sabati, domeniche e altre festività nazionali nel periodo t . Il regressore viene calcolato al di fuori della procedura TRAMO-SEATS e fatto acquisire dalla procedura stessa come regressore esterno.

¹ Il paragrafo 9.1 di questo capitolo è a cura di Francesca Ceccato e Donatella Tuzi; il paragrafo 9.2.1 e 9.2.3 sono a cura di Francesca Ceccato e Eleonora Cimino; il paragrafo 9.2.2 è a cura di Francesca Ceccato e Marco Lattanzio.

La destagionalizzazione degli indici Oros avviene attraverso un *approccio misto*, in cui si fa ricorso sia ad *approccio diretto* nel quale ciascuna serie viene corretta separatamente rispetto ai rispettivi aggregati, sia ad *approccio indiretto* in cui i totali vengono ottenuti indirettamente aggregando le serie destagionalizzate relative a serie elementari componenti. Il ricorso all'approccio indiretto è reso possibile dall'esistenza di una relazione di additività tra serie elementari e serie aggregate basata su pesi di natura non stagionale.

Pur producendo in TRAMO-SEATS le serie destagionalizzate, nella fase preliminare di identificazione dei modelli si ricorre a J-Demetra+, un software *open-source* raccomandato da Eurostat nell'ambito del Sistema Statistico Europeo per la destagionalizzazione delle serie storiche². Il software, che dispone di un'interfaccia grafica *user-friendly*, consente di effettuare in modo rapido e semplice la destagionalizzazione di numerose serie storiche contemporaneamente, integrando al suo interno i due approcci di correzione *filter based* (X-13-Arima SEATS) e *model based* (TRAMO-SEATS). Questo software si presta favorevolmente alla fase di re-identificazione di un numero non esiguo di modelli, operazione che di norma deve essere effettuata in tempi molto rapidi. In Oros J-Demetra+ viene, inoltre, periodicamente utilizzato per verificare l'assenza di stagionalità spuria sulle serie destagionalizzate con approccio indiretto, in quanto contiene alcuni test predisposti appositamente.

L'intera procedura di destagionalizzazione di Oros avviene nell'ambito di SITIC (cfr. §8.5): caricati gli indici grezzi e aggiornati i modelli di destagionalizzazione, il sistema elabora dapprima gli indici con approccio diretto e, a seguire, utilizzando i pesi caricati una tantum, deriva gli indici destagionalizzati indirettamente, operando gli arrotondamenti necessari.

9.2 Il sistema degli indicatori Oros. Dall'approccio diretto all'approccio indiretto

9.2.1 Configurazione concettuale dell'approccio diretto e indiretto

Come gran parte degli indicatori di natura economica, gli indici Oros possono essere configurati all'interno di un sistema caratterizzato da serie elementari (o *componenti*), la cui aggregazione dà luogo a serie sub-totali o totali (*marginali*). Ne sono un esempio le serie per settore di attività economica (es. sezioni Ateco 2007) e le loro aggregazioni per macro settore, nonché il totale costo del lavoro, quale aggregazione di retribuzioni e oneri. In tali contesti, la scelta della metodologia di destagionalizzazione è molto condizionata da obiettivi di consistenza interna garantiti, per definizione, dagli indici grezzi ma messi a rischio da qualsiasi operazione successiva di trattamento degli indici che non implichi una riconciliazione per costruzione quali, appunto, la destagionalizzazione. In questi contesti, le serie storiche aggregate, o marginali, possono essere trattate per gli effetti stagionali ricorrendo a due approcci, con proprietà e conseguenti risultati differenziati:

- trattando direttamente le serie aggregate (*approccio diretto*);
- aggregando le serie destagionalizzate delle serie elementari (*approccio indiretto*).

Tuttavia, la possibilità di ricorrere all'approccio indiretto è subordinata all'esistenza di una relazione di additività tra le serie componenti e la serie aggregata. Tale approccio consente di conservare la coerenza interna del sistema.

² Per maggiori dettagli su J-Demetra+ si veda al link:
https://ec.europa.eu/eurostat/cros/content/software-jdemetra_en

L'opportunità di adottare l'uno o l'altro approccio rimane tuttora una questione aperta poiché né evidenze teoriche né evidenze empiriche favoriscono uno dei due approcci. Seguendo le linee guida Eurostat l'approccio diretto è consigliabile quando le serie componenti presentano caratteristiche simili (Eurostat, 2015). Viceversa, nei casi in cui le serie mostrano caratteristiche molto diverse è preferibile ricorrere ad un approccio indiretto, particolarmente consigliato quando l'aggregato destagionalizzato contiene anche serie componenti che presentano una stagionalità debole e difficile da identificare. L'uso di un approccio o dell'altro, nel caso in cui le serie siano simili porta a risultati con discrepanze trascurabili, mentre in caso contrario le discrepanze che si riflettono nella serie aggregata destagionalizzata possono essere rilevanti. Inoltre, l'approccio diretto è preferibile per la sua trasparenza e accuratezza, garantite dalla maggiore libertà nell'identificazione dei modelli, caratteristica che si ripercuote positivamente sulla qualità delle serie trattate. Viceversa, potrebbe comportare problemi di inconsistenza tra dati aggregati e disaggregati, superabili mediante l'applicazione di opportune tecniche di riconciliazione.

Come anticipato poco sopra, l'approccio indiretto ha il vantaggio di non produrre problemi di inconsistenza grazie alla presenza di un vincolo di aggregazione tra serie componenti e serie marginali, evitando una successiva riconciliazione. Inoltre, tale approccio consente di ridurre il numero di serie da destagionalizzare e il numero di modelli da individuare, garantendo una maggiore tempestività nella produzione dei dati. D'altra parte, comporta una serie di altre attività da dover affrontare. Innanzitutto, il calcolo e il mantenimento di insiemi aggiuntivi di pesi rispetto a quelli originari. Inoltre, diviene fondamentale testare l'eventuale stagionalità residua sugli aggregati ottenuti indirettamente, in quanto potrebbero acquisire dalle serie componenti destagionalizzate effetti spuri di stagionalità. Tale aspetto viene affrontato con lo studio delle funzioni di autocorrelazione e degli spettri delle serie differenziate e destagionalizzate e delle relative componenti irregolari e altri test specifici (AAVV, 2015).

Nella pratica, è possibile usare anche una combinazione dei due approcci: ad esempio, quando la qualità dell'aggiustamento non può essere garantita al livello di dettaglio più basso, si può pensare di utilizzare un approccio diretto fino ad un certo livello e un approccio indiretto per livelli superiori di aggregazione. La scelta del livello di dettaglio individuato per l'approccio diretto dipende essenzialmente dalle necessità dell'utilizzatore più che da considerazioni statistiche (Eurostat, 2015).

La disponibilità di serie destagionalizzate sulle variabili prodotte dalla rilevazione Oros per la diffusione nazionale risale a ottobre 2003 (cfr. §2.3.2). Inizialmente rilasciate per un insieme molto ridotto di aggregati, il numero di serie destagionalizzate si è ampliato molto lentamente fino ad arrivare, solo a settembre 2015, ad un totale di 25 serie trattate, risultato della destagionalizzazione con metodo diretto di 6 aggregati economici per le 4 variabili retribuzioni, oneri, costo del lavoro e posizioni lavorative³, con l'aggiunta dell'indicatore delle posizioni lavorative in somministrazione (ex-interinali). Inizialmente, il ricorso all'approccio diretto per le serie aggregate, era motivato dalla brevità delle serie storiche e, quindi, dalla minore libertà nell'imposizione di modelli: ciascuna serie doveva essere trattata indipendentemente. In quel contesto, la destagionalizzazione con approccio diretto aveva lo svantaggio di generare un numero non trascurabile di incoerenze (dette anche *fuori range*) tra le variazioni congiunturali degli aggregati e delle rispettive componenti, più rilevante nei casi

³ Gli aggregati destagionalizzati fino al comunicato stampa di settembre 2015 sono l'industria e servizi (B-S escluso O), l'industria e servizi di mercato (B-N), l'industria (B-F), i servizi di mercato (G-N), i servizi (G-S escluso O), l'istruzione, sanità e assistenza sociale, attività artistiche e altre attività di servizi (P-S).

di aggregazione di poche componenti rispetto ad un determinato indicatore marginale. Era il caso, ad esempio, della variabile costo del lavoro rispetto alle due componenti retribuzioni e oneri sociali: per limitare la comparsa di *fuori range* non trascurabili sul totale costo del lavoro, i modelli degli oneri venivano adattati il più possibile a quelli delle retribuzioni, nei limiti della significatività dei risultati statistici, anche in considerazione del fatto che gli oneri rappresentano circa il 30% del costo del lavoro e sono correlati con le retribuzioni, seppur più volatili. Il numero di *fuori range* in tal modo veniva ridotto, ma la minore qualità dei modelli statistici utilizzati implicava revisioni più rilevanti sulle serie destagionalizzate degli oneri.

L'offerta degli indici grezzi e destagionalizzati ha avuto un notevole slancio a seguito dell'avvio della diffusione degli indicatori sulle posizioni lavorative da parte di Oros, variabile con un potenziale informativo congiunturale molto più ampio rispetto agli indicatori di costo del lavoro: a partire dall'uscita di dicembre 2015 hanno iniziato ad essere rilasciate serie destagionalizzate a livello di sezione Ateco 2007 (sezioni da B a S escluso O) per tutte le variabili prodotte dalla rilevazione, inclusi anche due nuovi aggregati⁴. Dato l'elevato numero di serie da destagionalizzare, 101 in totale tra le variabili di costo del lavoro e le posizioni lavorative, è stata fortemente presa in considerazione l'opportunità di adottare l'approccio indiretto per le serie derivanti da aggregazione di serie componenti. In particolare, è stata valutata l'aggregazione secondo la dimensione settoriale (aggregazione delle sezioni in macro aggregati economici) e secondo la dimensione delle variabili (aggregazione di retribuzione e oneri sociali in costo del lavoro). Inoltre, a partire da settembre 2016, le posizioni lavorative hanno iniziato ad essere diffuse anche in termini di livello. Essendo i dati di livello alla base del calcolo dei rispettivi indici, al fine di evitare incongruenze tra i dati diffusi, si è stabilito di utilizzare gli stessi modelli di destagionalizzazione già individuati per il trattamento degli indici, non costituendo la costruzione dell'indice motivo di modifica dell'andamento congiunturale della variabile di riferimento⁵.

Come descritto nel paragrafo 2.1, gli indici Oros relativi a un generico macro-aggregato sono costruiti come indici semplici. Per capire se fosse possibile applicare il metodo indiretto, è stato necessario verificare l'esistenza di una relazione di tipo additivo tra gli indici semplici delle componenti e il relativo macro-aggregato e che i corrispondenti pesi di aggregazione fossero, a loro volta, non stagionali. Nel paragrafo 9.2.2 che segue, si mostra la formalizzazione algebrica della relazione additiva tra l'indice del costo del lavoro per Ula e gli indici delle sue componenti (retribuzioni e oneri sociali per Ula) e tra gli indici delle posizioni lavorative⁶ di un generico aggregato economico e gli indici delle sue componenti (sezioni Ateco 2007), i cui pesi dipendono solo dall'anno base. Viceversa, si dimostra che, per le tre variabili economiche di costo del lavoro considerate, l'aggregazione secondo la dimensione settoriale non verifica la relazione additiva ricercata, ossia non caratterizzata da pesi di natura stagionale.

4 I nuovi aggregati destagionalizzati dal comunicato stampa di dicembre 2015 sono l'industria in senso stretto (B-E) e le attività immobiliari, professionali e noleggio (L-N).

5 In realtà, si possono verificare differenze di alcuni decimi di punto tra le variazioni congiunturali calcolate a partire dagli indici e a partire dai livelli. Il motivo di tali differenze è dovuto alle regole di arrotondamento che vengono utilizzate nella predisposizione degli indici e dei dati di livello nella fase precedente e successiva al trattamento per la stagionalità.

6 Per le variabili espresse come livelli la relazione di additività è valida per costruzione.

9. La correzione per la stagionalità e gli effetti di calendario degli indicatori Oros

9.2.2 I pesi di aggregazione per l'approccio indiretto

L'aggregazione di indici componenti secondo l'approccio indiretto per ottenere l'indice aggregato è resa possibile se la relazione sottostante verifica la proprietà di additività e, nel caso degli aggregati destagionalizzati, è consigliata se i pesi di aggregazione non sono a loro volta affetti da stagionalità. La formalizzazione algebrica che si presenta in questo sotto-paragrafo ha un duplice obiettivo: con essa, si vuole dimostrare che l'approccio indiretto può essere applicato per aggregare settorialmente l'indice destagionalizzato delle posizioni lavorative e per aggregare nella dimensione "Costo del lavoro" gli indici delle variabili pro-capite, mentre non è di conveniente e immediata applicazione per aggregare settorialmente le variabili pro-capite⁷; il secondo obiettivo è di ricavare la formula per calcolare i pesi di aggregazione per l'utilizzo e l'implementazione del metodo indiretto.

Sia S un aggregato (ad esempio, un macro settore Ateco) ricavato come unione di s settori componenti S_j (ad esempio, le sezioni Ateco), a loro volta visti come insiemi di unità statistiche:

$$S = \cup_{j=1}^s S_j \quad (2)$$

Si voglia calcolare indici per l'aggregato S e indici relativi ai singoli settori componenti. In prima analisi si consideri la dimensione di aggregazione settoriale della variabile posizioni lavorative. Si definisce il monte relativo al settore S_j nel modo seguente:

$$D_{j,t} = \sum_{i \in S_j} D_{i,t} \quad (3)$$

dove la lettera i varia su tutte le unità appartenenti al settore S_j al tempo t e, conseguentemente, l'ammontare relativo all'aggregato S , unione dei settori S_j , pari a:

$$D_{S,t} = \sum_{j=1}^s \sum_{i \in S_j} D_{i,t} \quad (4)$$

Le formule (3) e (4) rimangono valide anche se si considerano le altre variabili espresse come monte quali, ad esempio, le Ula (in formula, U), gli oneri sociali (O), le retribuzioni (R) e, quale somma delle due precedenti, il costo del lavoro (C).

Sia $I_{S,t}^D$ l'indice delle posizioni lavorative per l'aggregato S al tempo t con anno base b . Si dimostra che l'aggregazione settoriale nel caso delle posizioni lavorative dipendenti è scomponibile come media degli indici delle posizioni lavorative dei settori in cui è partizionato l'aggregato, ponderata con pesi fissi all'anno base b calcolati sulle posizioni lavorative del j -simo settore rispetto a quelle dell'aggregato S . In formula:

$$\begin{aligned} I_{S,t}^D &= \frac{D_{S,t}}{\frac{1}{4} \sum_{tb=1}^4 D_{S,tb}} = \frac{\sum_{j=1}^s D_{j,t}}{\sum_{j=1}^s \frac{1}{4} \sum_{tb=1}^4 D_{j,tb}} = \sum_{j=1}^s \left(\frac{D_{j,t}}{\frac{1}{4} \sum_{tb=1}^4 D_{j,tb}} \frac{\frac{1}{4} \sum_{tb=1}^4 D_{j,tb}}{\sum_{j=1}^s \frac{1}{4} \sum_{tb=1}^4 D_{j,tb}} \right) \\ &= \sum_{j=1}^s \left(I_{j,t}^D \frac{\sum_{tb=1}^4 D_{j,tb}}{\sum_{j=1}^s \sum_{tb=1}^4 D_{j,tb}} \right) = \sum_{j=1}^s I_{j,t}^D \pi_{j,b}^D \end{aligned} \quad (5)$$

La scomposizione (5) così ottenuta è una relazione additiva e i coefficienti $\pi_{j,b}^D$ costituiscono effettivamente un sistema di pesi.

⁷ Per un approfondimento metodologico sulle proprietà degli indici calcolati su variabili rapporto si veda Lattanzio, 2016a e Lattanzio, 2016b.

La (5) rimane valida anche se i settori di cui si compone l'aggregato S sono a loro volta sub-aggregazioni di settori di livello inferiore. In questo caso, il ricorso all'approccio indiretto nella destagionalizzazione dell'indice delle posizioni lavorative dell'aggregato S , a partire dagli indici destagionalizzati delle posizioni lavorative dei j -esimi settori ad esso appartenenti, presenta tutti i vantaggi di tale approccio.

Proprietà di tipo differente caratterizzano, invece, le variabili di costo del lavoro prodotte da Oros per la diffusione nazionale, espresse in termini di valori per Ula e, quindi, variabili rapporto. In particolare, mentre è possibile verificare che si osserva additività nell'aggregazione della dimensione "Costo del lavoro" con struttura di pesi indipendente da t , ciò non è verificato per l'aggregazione settoriale delle retribuzioni e degli oneri sociali per Ula, rendendo poco opportuna l'applicazione dell'approccio indiretto di destagionalizzazione.

Sia $I_{S,t}^{CU}$ l'indice semplice del costo del lavoro per Ula calcolato al tempo t con anno base b per un generico aggregato S , sia esso un singolo settore o aggregazione di settori (ad esempio, sezione o macro settore Ateco). Si vuole verificare la scomposizione in funzione degli indici di retribuzioni e oneri per Ula, verificando che:

$$\begin{aligned}
 I_{S,t}^{CU} &= \frac{C_{S,t}}{U_{S,t}} = \frac{R_{S,t} + O_{S,t}}{U_{S,t}} = \frac{\frac{R_{S,t}}{U_{S,t}} + \frac{O_{S,t}}{U_{S,t}}}{1} = \\
 &= \frac{\frac{1}{4} \sum_{tb=1}^4 \frac{C_{S,tb}}{U_{S,tb}}}{\frac{1}{4} \sum_{tb=1}^4 \frac{C_{S,tb}}{U_{S,tb}}} = \frac{\frac{1}{4} \sum_{tb=1}^4 \frac{R_{S,tb} + O_{S,tb}}{U_{S,tb}}}{\frac{1}{4} \sum_{tb=1}^4 \frac{R_{S,tb} + O_{S,tb}}{U_{S,tb}}} = \frac{\frac{1}{4} \left(\sum_{tb=1}^4 \frac{R_{S,tb}}{U_{S,tb}} + \sum_{tb=1}^4 \frac{O_{S,tb}}{U_{S,tb}} \right)}{\frac{1}{4} \left(\sum_{tb=1}^4 \frac{R_{S,tb}}{U_{S,tb}} + \sum_{tb=1}^4 \frac{O_{S,tb}}{U_{S,tb}} \right)} = \\
 &= \frac{\frac{R_{S,t}}{U_{S,t}} \frac{1}{4} \sum_{tb=1}^4 \frac{R_{S,tb}}{U_{S,tb}} + \frac{O_{S,t}}{U_{S,t}} \frac{1}{4} \sum_{tb=1}^4 \frac{O_{S,tb}}{U_{S,tb}}}{\frac{1}{4} \sum_{tb=1}^4 \frac{R_{S,tb}}{U_{S,tb}} + \frac{1}{4} \sum_{tb=1}^4 \frac{O_{S,tb}}{U_{S,tb}}} = \\
 &= \frac{I_{S,t}^{RU} \sum_{tb=1}^4 \frac{R_{S,tb}}{U_{S,tb}} + I_{S,t}^{OU} \sum_{tb=1}^4 \frac{O_{S,tb}}{U_{S,tb}}}{\sum_{tb=1}^4 \frac{R_{S,tb}}{U_{S,tb}} + \sum_{tb=1}^4 \frac{O_{S,tb}}{U_{S,tb}}} \\
 &= I_{S,t}^{RU} \pi_{S,b}^R + I_{S,t}^{OU} \pi_{S,b}^O \quad \text{con: } \pi_{S,b}^R + \pi_{S,b}^O = 1
 \end{aligned} \tag{6}$$

ovvero, l'indice del costo del lavoro per Ula al tempo t con base b per il generico aggregato S (o singolo settore s) può essere espresso come media dell'indice delle retribuzioni per Ula $I_{S,t}^{RU}$ nell'aggregato S e dell'indice degli oneri per Ula $I_{S,t}^{OU}$ nello stesso aggregato, ponderata con un sistema di pesi fissi all'anno base b e a somma 1, $\pi_{S,b}^R$ e $\pi_{S,b}^O$, che rappresentano, rispettivamente, il peso delle retribuzioni per ula medie e il peso degli oneri sociali per ula medi calcolati rispetto al costo del lavoro per ula medio.

La scomposizione (6), che garantisce l'additività degli indici nella dimensione del costo del lavoro, favorisce l'utilizzo di un approccio indiretto alla destagionalizzazione degli indici di costo del lavoro per Ula come media ponderata degli indici destagionalizzati delle retribuzioni per Ula e degli oneri sociali per Ula, sia per il generico settore s sia per tutti i suoi aggregati, in cui i pesi non sono variabili con t . La stessa relazione non è, invece, valida per le due componenti del costo del lavoro nella scomposizione per indici relativi a sotto-

9. La correzione per la stagionalità e gli effetti di calendario degli indicatori Oros

settori. Di seguito, esplicitando la relazione che lega gli indici relativi agli aggregati settoriali delle retribuzioni ai rispettivi indici componenti, estendibile anche agli oneri sociali, è possibile evidenziare tale caratteristica (Lattanzio, 2016a e Lattanzio, 2016b). Sia $I_{S,t}^{RU}$ l'indice semplice delle retribuzioni per Ula dell'aggregato S al tempo t con anno base b :

$$I_{S,t}^{RU} = \frac{\frac{R_{S,t}}{U_{S,t}}}{\frac{1}{4} \sum_{tb=1}^4 \frac{R_{S,b}}{U_{S,b}}} \quad (7)$$

Si dimostra che è possibile scomporre tale indice nell'aggregazione delle retribuzioni pro-capite calcolate per s insiemi disgiunti (le sezioni Ateco 2007) verificando tuttavia una particolare aggregazione additiva. In formula si ottiene:

$$\begin{aligned} I_{S,t}^{RU} &= \frac{\frac{\sum_{j=1}^s R_{j,t}}{\sum_{j=1}^s U_{j,t}}}{\frac{1}{4} \sum_{tb=1}^4 \frac{\sum_{j=1}^s R_{j,tb}}{\sum_{j=1}^s U_{j,tb}}} = \frac{\sum_{j=1}^s \frac{R_{j,t}}{\sum_{j=1}^s U_{j,t}}}{\frac{1}{4} \sum_{tb=1}^4 \frac{\sum_{j=1}^s R_{j,tb}}{\sum_{j=1}^s U_{j,tb}}} = \frac{\sum_{j=1}^s \left(\frac{R_{j,t}}{U_{j,t}} \frac{U_{j,t}}{\sum_{j=1}^s U_{j,t}} \right)}{\frac{1}{4} \sum_{tb=1}^4 \frac{\sum_{j=1}^s R_{j,tb}}{\sum_{j=1}^s U_{j,tb}}} \\ &= \frac{\sum_{j=1}^s \left(\frac{\frac{R_{j,t}}{U_{j,t}}}{\frac{1}{4} \sum_{tb=1}^4 \frac{R_{j,tb}}{U_{j,tb}}} \frac{1}{4} \sum_{tb=1}^4 \frac{R_{j,tb}}{U_{j,tb}} \frac{\frac{U_{j,t}}{\sum_{j=1}^s U_{j,t}}}{\frac{1}{4} \sum_{tb=1}^4 \frac{U_{j,tb}}{\sum_{j=1}^s U_{j,tb}}} \right)}{\frac{1}{4} \sum_{tb=1}^4 \frac{\sum_{j=1}^s R_{j,tb}}{\sum_{j=1}^s U_{j,tb}}} = \\ &= \frac{\sum_{j=1}^s \left(I_{j,t}^{RU} I_{j,t}^{\pi_U} \frac{1}{4} \sum_{tb=1}^4 \frac{R_{j,tb}}{U_{j,tb}} \frac{1}{4} \sum_{tb=1}^4 \frac{U_{j,tb}}{\sum_{j=1}^s U_{j,tb}} \right)}{\frac{1}{4} \sum_{tb=1}^4 \frac{\sum_{j=1}^s R_{j,tb}}{\sum_{j=1}^s U_{j,tb}}} = \\ &= \sum_{j=1}^s I_{j,t}^{RU} I_{j,t}^{\pi_U} \frac{RU_{j,b} \pi_{j,b}^U}{RU_{S,b}} \quad (8) \end{aligned}$$

ovvero, l'indice delle retribuzioni per Ula dell'aggregato S al tempo t con anno base b può essere espresso come somma ponderata del prodotto tra gli indici semplici delle retribuzioni per Ula $I_{j,t}^{RU}$ del j -esimo settore e gli indici semplici della composizione occupazionale $I_{j,t}^{\pi_U}$ del j -esimo settore rispetto all'aggregato S (che dipendono da t), con coefficienti di ponderazione fissi all'anno base b . Questi coefficienti sono funzione della retribuzione pro-capite media nel settore j all'anno base b , $RU_{j,b}$, del peso occupazionale medio nel settore j all'anno base b , $\pi_{j,b}^U$, e della retribuzione pro-capite media nell'aggregato S all'anno base b , $RU_{S,b}$.

La scomposizione (8) mette in evidenza che per ottenere indirettamente gli aggregati settoriali dell'indice delle retribuzioni pro-capite, sarebbe necessario correggere per gli effetti stagionali anche la componente $I_{j,t}^{\pi_U}$ annullando il vantaggio del ricorso a tale approccio.

9.2.3 Dall'approccio diretto ad un approccio "misto"

Come dimostrato nel paragrafo precedente, l'utilizzo dell'approccio indiretto non è conveniente per produrre gli aggregati settoriali destagionalizzati degli indici delle retribuzioni per Ula e degli oneri sociali per Ula, poiché nella relazione di aggregazione è presente una componente, nel caso specifico la composizione occupazionale, che richiederebbe la depurazione degli effetti stagionali, annullando il vantaggio del ricorso a tale approccio in termini di alleggerimento del processo. Per questo tali serie vengono trattate con approccio diretto, mentre per tutte le altre serie aggregate si adotta l'approccio indiretto. L'utilizzo combinato dei due approcci configura pertanto la destagionalizzazione degli indicatori Oros come un processo di tipo "misto" che consente di mantenere elevati livelli di accuratezza e consistenza, garantendo nel contempo l'economicità del processo e la tempistica usuale nella diffusione degli output.

Nella figura 9.2.3.1, si fornisce un quadro sintetico di tale approccio. In particolare, lo schema riporta in fiancata la dimensione settoriale e in testata la dimensione per variabile. L'approccio indiretto (Step 2), che segue in successione l'approccio diretto (Step 1), interessa la variabile costo del lavoro nella dimensione di aggregazione per variabile e la variabile posizioni lavorative dipendenti nella dimensione di aggregazione settoriale. Le scelte di aggregazione sono acquisite dal sistema SITIC, tramite cui viene effettuata la destagionalizzazione, attraverso una tabella che legge la tipologia di aggregazione che vuole essere effettuata, acquisisce i pesi per l'aggregazione (in totale, il numero di pesi per l'aggregazione indiretta è pari a 67, di cui 50 per le due componenti del costo del lavoro per Ula e 17 per le componenti settoriali delle posizioni lavorative)⁸. Come descritto nel paragrafo 9.2.2, il sistema dei pesi di aggregazione dipende dall'anno base degli indicatori e, pertanto, segue la stessa politica di aggiornamento della base degli indici.

Nella pratica, trimestralmente, gli output TRAMO-SEATS della destagionalizzazione diretta vengono controllati per verificare la bontà dei risultati ottenuti applicando i modelli di destagionalizzazione presenti sul sistema, di norma rivisti una sola volta all'anno, in occasione dell'uscita del primo trimestre (cfr. §9.1). Nel caso di risultati validi, si procede all'aggregazione indiretta delle serie destagionalizzate, mentre in caso di risultati critici si valuta la possibilità di individuare un nuovo modello di destagionalizzazione in corso d'anno. Solo nel caso in cui le revisioni degli indicatori destagionalizzati siano trascurabili, si procede alla sostituzione del nuovo modello su SITIC.

L'uso dell'approccio misto descritto, tuttavia, mantiene l'inconveniente di generare incoerenze tra le variazioni congiunturali degli aggregati settoriali destagionalizzati con approccio diretto rispetto alle componenti. Per eliminare questa inconsistenza si potrebbe ricorrere a tecniche di riconciliazione tra le componenti (sezioni o aggregati di ordine inferiore) e gli aggregati economici superiori (AAVV, 2015). Nel caso Oros tali tecniche non vengono attualmente usate sia per l'entità trascurabile dei *fuori range* sia perché nel sistema SITIC, che gestisce automaticamente l'intera operazione di destagionalizzazione, non sono state ancora implementate.

⁸ Il settore 782 delle posizioni lavorative è escluso dal calcolo indiretto degli aggregati che lo includono in quanto rappresenta, a sua volta, un di cui del settore N.

Figura 9.2.3.1 - Schema di destagionalizzazione degli indicatori Oros secondo l'approccio "misto"

Settori Ateco 2007	Variabili Oros			Posizioni lavorative dipendenti	
	Retribuzione per Ula	+	Oneri sociali per Ula		=
782 (a)	APPROCCIO DIRETTO → STEP 1	+	APPROCCIO INDIRETTO → STEP 2	=	STEP 1
B					
C					
D					
E					
F					
G					
H					
I					
J					
K					
L					
M					
N					
P					
Q					
R					
S					
B-E					STEP 2
B-F					STEP 2
G-N (b)					STEP 2
L-N (b)					STEP 2
B-N (b)					STEP 2
P-S					STEP 2
G-S (b)(c)					STEP 2
B-S (b)(c)					STEP 2

(a) Settore Ateco 2007 attribuito alle sole posizioni lavorative dipendenti in somministrazione.
 (b) La destagionalizzazione di tali aggregati non include la componente settoriale 782 in quanto già presente nella componente settore N.
 (c) L'aggregato esclude il settore dell'amministrazione pubblica e difesa; assicurazione sociale obbligatoria (sezione O Ateco 2007).

10. I TRIANGOLI E L'ANALISI DELLE REVISIONI DEGLI INDICATORI OROS¹

Tra il 2010 e il 2011 l'Istituto si è dotato di uno strumento standardizzato per la diffusione delle revisioni degli indicatori congiunturali prodotti, i cosiddetti "triangoli delle revisioni", accompagnati da una scheda riassuntiva sull'analisi delle revisioni contenuta nei triangoli. Inizialmente rilasciati in allegato ai comunicati stampa, a partire da metà 2017 la loro diffusione è stata trasportata in una specifica pagina web del sito dell'Istituto dedicata alle "Revisioni dei principali indicatori congiunturali"². Questa nuova modalità di diffusione risponde all'impegno dell'Istituto di aderire ai principi contenuti nel *Codice delle statistiche europee*, nel *Quality Assurance Framework of the European Statistical System (QAF)* e nelle *ESS Guidelines on Revision Policy for PEEIs*³ con il fine di standardizzare le informazioni diffuse dalle varie rilevazioni. A tal fine, nella pagina web vengono fornite informazioni generali su politica di revisione, principi e tipologie di revisioni. Scendendo nel merito dell'analisi delle revisioni, per ogni indagine vengono pubblicate una scheda informativa di revisione, che descrive la politica di revisione adottata per le serie grezze e per quelle destagionalizzate (ove prodotte), un elenco dei motivi delle revisioni ordinarie e straordinarie, il calendario del ciclo completo delle revisioni ordinarie, i triangoli di revisione in cui sono raccolti in forma tabellare i dati delle edizioni rilasciate nel corso del tempo, accompagnati da indicatori di qualità che sintetizzano il processo di revisione. Alcuni grafici e tabelle forniscono, infine, un quadro sintetico della revisione attuata nell'ultima uscita del dato diffuso nell'occasione immediatamente precedente.

L'analisi delle revisioni basata sui triangoli mira a quantificare, sintetizzare e valutare il processo di revisione delle stime preliminari rispetto a quelle pubblicate in periodi successivi utilizzando una particolare rappresentazione tabellare denominata, appunto, "triangolo delle revisioni" (figura 10.1). La caratteristica forma triangolare della tabella stessa deriva dal modo nel quale le diverse edizioni di serie storica (o *vintage*) sono organizzate e disposte. Le righe riportano le serie storiche dell'indice e delle rispettive variazioni rilasciate alla data del comunicato stampa, indicata in prima colonna. Le colonne successive indicano i trimestri di riferimento delle stime e forniscono una ricostruzione della storia delle stime stesse diffuse per ciascun riferimento temporale, dal primo rilascio fino all'ultimo disponibile. L'analisi delle revisioni è utile per valutare l'impatto delle informazioni aggiuntive che si rendono disponibili dopo il primo rilascio e, in riferimento al dato destagionalizzato, anche per quantificare l'effetto dovuto al processo di destagionalizzazione⁴.

Per finalizzare l'analisi, nel caso di Oros sono state prescelte solo le variabili più significative prodotte dalla rilevazione e per cui esisteva, alla data di avvio della prima pubblicazione dell'analisi delle revisioni, avvenuta a marzo 2011, un sufficiente numero di *vintage* su cui poter calcolare le revisioni. Attualmente, nella pagina web dedicata alle revisioni,

1 Il presente capitolo è a cura di Donatella Tuzi.

2 Pagina al link: <http://www4.istat.it/it/congiuntura/revisioni/indicatori-congiunturali-oggetto>.

3 Documenti disponibili rispettivamente ai link:

<http://ec.europa.eu/eurostat/documents/3859598/5921861/KS-32-11-955-EN.PDF>,

http://ec.europa.eu/eurostat/documents/64157/4392716/qaf_2012-en.pdf e

<http://ec.europa.eu/eurostat/documents/3859598/5935517/KS-RA-13-016-EN.PDF>.

4 Per ulteriori informazioni sull'analisi delle revisioni si rimanda all'approfondimento disponibile al link:

http://www.istat.it/it/files/2011/01/approfondimenti_analisi_revisioni.pdf.

vengono diffusi i triangoli del costo del lavoro, degli oneri sociali e delle retribuzioni per Ula relative all'aggregato industria e servizi di mercato (sezioni da B a N Ateco 2007). La diffusione dei triangoli, temporaneamente interrotta in occasione dell'ultima uscita degli indici Oros nella base 2005=100, ossia a marzo 2013, è stata riavviata a dicembre 2016.

Considerata la politica di revisione di Oros (cfr. §2.3), i triangoli degli indicatori prescelti assumono una forma regolare. La figura 10.1 mostra un estratto del triangolo a gradini delle variazioni tendenziali dell'indice del costo del lavoro per Ula. A seconda di come viene letto il triangolo (in senso "orizzontale", "verticale" o "diagonale") se ne trae una particolare analisi. Se letto in orizzontale si può risalire alla storia dei rilasci di una stessa informazione statistica, molto utile, ad esempio, quando si vogliono valutare modelli previsivi tipici dell'analisi delle serie storiche sfruttando l'informazione all'epoca realmente disponibile. Dalla lettura in verticale si ricava invece la storia delle revisioni delle stime riferite ad un singolo periodo temporale (mese o trimestre); considerando l'insieme dei periodi disponibili, questo tipo di analisi fornisce una misura dell'affidabilità delle stime pubblicate nel tempo, intesa come vicinanza delle stime successive rispetto a quelle precedentemente rilasciate. Infine, la lettura diagonale consente di individuare le stime caratterizzate dalla medesima epoca di rilascio: ad esempio, gli elementi appartenenti alla diagonale principale del triangolo (quella più esterna) formano la serie storica delle prime stime rilasciate per ogni periodo temporale; gli elementi appartenenti alla diagonale secondaria (quella sottostante la diagonale principale) individuano la serie delle seconde stime, ecc.

Si definisce revisione la misura:

$$R_t = (L_t - P_t) \quad (1)$$

dove P_t rappresenta le stime preliminari e L_t le stime successive. Considerata la politica di revisione degli indicatori Oros, vengono ritenuti rilevanti due indicatori di revisione che mettono a confronto la prima stima del dato relativo a t (P) con la sua revisione dopo un anno (A_1) e dopo un trimestre (R_1), ossia: $R_1 - P$ e $A_1 - P$. Inoltre, per il dato destagionalizzato vengono diffuse anche le misure $A_2 - P$ e $A_3 - P$ utili per l'analisi del processo di convergenza dei dati destagionalizzati dovuto al modello utilizzato per la correzione.

Oggetto dell'analisi delle revisioni è descrivere gli indicatori prescelti per r_t sui tassi di variazione orizzontali sia per i dati grezzi sia per i destagionalizzati e ottenerne una sintesi utilizzando alcune misure statistiche (tavola 10.1) finalizzate a far emergere le diverse peculiarità del processo di revisione. In particolare, vengono considerate quattro categorie di misure: ampiezza media, direzione, variabilità e impatto sui segni dei tassi di variazione⁵.

⁵ Per un approfondimento sulle misure di revisione si veda il documento metodologico riportato in nota 4.

Figura 10.1- Estratto del triangolo di revisione delle variazioni tendenziali del costo del lavoro per Ula

Data di pubblicazione	2011_3	2011_4	2012_1	2012_2	2012_3	2012_4	2013_1	2013_2	2013_3	2013_4	2014_1	2014_2	2014_3	2014_4	2015_1	2015_2	2015_3	2015_4	2016_1	2016_2		
Stima provvisoria P	1.5	1.7	2.2	0.7	1.7	1.7	1.6	1.8	2.4	2.4	2.3	1.8	1.1	1.6	0.9	1.5	1.0	0.5	-0.2	-1.5	0.0	
Stima rivista R1	1.6	1.8	2.2	0.7	1.7	1.7	2.1	1.8	2.4	2.4	2.4	1.8	1.2	1.7	0.9	1.5	1.0	0.5	-0.1	-1.5	0.1	
Stima rivista R2	1.6	1.8	2.3	0.7	2.4	2.4	2.1	1.9	2.4	2.6	2.4	1.8	1.2	1.7	0.9	1.5	1.0	0.6	-0.1	-1.5	0.1	
Stima rivista R3	1.6	1.8	2.3	0.9	2.4	2.4	2.1	1.9	2.5	2.6	2.4	1.8	1.3	1.6	0.9	1.5	1.1	0.6	-0.1	-1.5	0.1	
Stima rivista A1	1.4	2.0	2.2	0.6	2.0	2.0	2.3	1.8	2.3	2.5	2.6	1.8	1.0	1.3	0.9	1.3	0.6	0.1	-0.3	-0.3	-0.3	
Ultima stima U	1.5	1.8	2.2	0.7	2.0	2.0	2.3	1.8	2.3	2.5	2.5	1.8	1.0	1.3	0.9	1.3	0.6	0.1	-0.3	-1.5	0.1	
giu-11																						
set-11																						
dic-11	1.5																					
mar-12	1.6	1.7																				
giu-12	1.6	1.8	2.2																			
set-12	1.6	1.8	2.2	0.7																		
dic-12	1.4	1.8	2.3	0.7	1.7																	
mar-13	1.4	2.0	2.3	0.7	1.7	1.6																
giu-13	1.5	1.8	2.2	0.9	2.4	2.1	1.8															
set-13	1.5	1.8	2.2	0.6	2.4	2.1	1.8	2.4														
dic-13	1.5	1.8	2.2	0.7	2.0	2.1	1.9	2.4	2.4													
mar-14	1.5	1.8	2.2	0.7	2.0	2.3	1.9	2.4	2.4	2.3												
giu-14	1.5	1.8	2.2	0.7	2.0	2.3	1.8	2.5	2.6	2.4	1.8											
set-14	1.5	1.8	2.2	0.7	2.0	2.3	1.8	2.3	2.6	2.4	1.8	1.1										
dic-14	1.5	1.8	2.2	0.7	2.0	2.3	1.8	2.3	2.5	2.4	1.8	1.2	1.6									
mar-15	1.5	1.8	2.2	0.7	2.0	2.3	1.8	2.3	2.5	2.6	1.8	1.2	1.7	0.9								
giu-15	1.5	1.8	2.2	0.7	2.0	2.3	1.8	2.3	2.5	2.5	1.8	1.3	1.7	0.9	1.5							
set-15	1.5	1.8	2.2	0.7	2.0	2.3	1.8	2.3	2.5	2.5	1.8	1.0	1.6	0.9	1.5	1.0						
dic-15	1.5	1.8	2.2	0.7	2.0	2.3	1.8	2.3	2.5	2.5	1.8	1.0	1.3	0.9	1.5	1.0	0.5					
mar-16	1.5	1.8	2.2	0.7	2.0	2.3	1.8	2.3	2.5	2.5	1.8	1.0	1.3	0.9	1.5	1.0	0.5	-0.2				
giu-16	1.5	1.8	2.2	0.7	2.0	2.3	1.8	2.3	2.5	2.5	1.8	1.0	1.3	0.9	1.3	1.1	0.6	-0.1	-1.5			
set-16	1.5	1.8	2.2	0.7	2.0	2.3	1.8	2.3	2.5	2.5	1.8	1.0	1.3	0.9	1.3	0.6	-0.1	-1.5	-1.5	0.0		



Prospetto 10.1 - Misure statistiche utilizzate nei triangoli delle revisioni

MISURE STATISTICHE

Ampiezza media delle revisioni
 Revisione Media Assoluta (RMA)
 Revisione media quadratica
 Mediana delle revisioni in valore assoluto
 Revisione Media Assoluta in termini Relativi (RMAR)
 Direzione delle revisioni
 Revisione Media (RM)
 Deviazione standard della Revisione Media
 Statistica t di Student relativa alla Revisione Media
 Valori critici della distribuzione t di Student (0,1/0,05/0,01)
 Significatività statistica della Revisione Media
 Mediana delle revisioni
 Indice di asimmetria delle revisioni
 % di revisioni positive
 % di revisioni negative
 % di revisioni pari a zero
 Variabilità delle revisioni
 Deviazione standard delle revisioni (DSR)
 Revisione minima
 Revisione massima
 Campo di variazione delle revisioni
 Intervallo in cui rientra il 90% delle revisioni
 Scarto interquartile
 Impatto delle revisioni sul segno dei tassi di crescita
 Concordanza di segno tra stima successiva e stima precedente
 Misura di accelerazione per effetto delle revisioni
 Misura di decelerazione per effetto delle revisioni

Infine, ciascun triangolo di revisione è corredato da un foglio di lavoro “eventi”, contenente la descrizione degli eventi, ordinari e straordinari, che hanno causato una revisione delle stime precedentemente diffuse. La lettura incrociata delle revisioni con tali informazioni consente all’utente di conoscere i motivi sottostanti le revisioni più o meno marcate intervenute nel corso del tempo. Nel caso di revisioni straordinarie (cambi base ecc.) che rendono le *vintage* non confrontabili, indici e tassi di variazione vengono inseriti nei triangoli, ma la revisione non viene calcolata.

L’aggiornamento dei triangoli di Oros avviene con cadenza regolare e contestualmente all’uscita del comunicato stampa. Precedentemente basato su tabelle presenti in SITIC, nell’attuale versione su web la selezione dell’ultima *vintage* avviene da I.Stat e l’aggiornamento dei triangoli, anch’esso automatico, si basa su procedure disegnate in Business Architecture.

11. GLI INDICATORI PER IL REGOLAMENTO LCI - LABOUR COST INDEX¹

11.1 Il quadro di riferimento: Regolamenti europei

L'indicatore trimestrale sul costo del lavoro orario (*Labour Cost Index – LCI*) riveste un'importanza rilevante tra gli indicatori per la comprensione del processo inflazionistico e delle dinamiche del costo nell'economia. Eurostat lo inserisce tra i PEEIs (*Principal European Economic Indicators*), cioè tra quegli indicatori congiunturali che, grazie anche alla loro tempestività, consentono di misurare la dinamica economica e monetaria dell'area euro fornendo un utile supporto alla politica monetaria comune (Ciammola et al., 2009).

Il costo del lavoro rappresenta una componente importante dell'inflazione in quanto generalmente costituisce la parte principale dei costi totali sostenuti dalle imprese private nei loro processi produttivi. Queste ultime trasferiscono ai consumatori i più alti costi attraverso un aumento dei prezzi dei beni/servizi prodotti, in particolare se non controbilanciati da aumenti di produttività, generando inflazione. Pertanto disporre di un indicatore tempestivo sul costo del lavoro è di fondamentale importanza per la Banca Centrale Europea (BCE) nel monitoraggio dell'inflazione nell'area euro. Inoltre, le informazioni relative al costo del lavoro orario sono di basilare supporto nella definizione delle politiche economiche e monetarie nazionali, nella contrattazione salariale e nelle analisi economiche da parte delle istituzioni nazionali quali il Ministero del lavoro, il Ministero dell'economia, i sindacati dei lavoratori, le associazioni dei datori di lavoro, le università, ecc..

L'indicatore è disciplinato da una serie di regolamenti della Commissione Europea. Il primo è il Regolamento n. 450/2003 del 27 Febbraio 2003 con il quale viene istituita una struttura standardizzata per la produzione, trasmissione e valutazione del costo del lavoro orario dei vari Stati Comunitari. Il Regolamento n. 1216/2003 del 7 luglio 2003 spiega diffusamente le caratteristiche che l'indicatore deve garantire in termini di comparabilità, con particolare riferimento alla correzione per la stagionalità e la composizione del calendario, alla definizione di dimensioni e reportistica comuni sulla qualità, alla metodologia utilizzata per il calcolo dell'indice. Infine, il Regolamento n. 224/2007 del 1 Marzo 2007 estende il campo di osservazione di LCI ai settori della pubblica amministrazione e dei servizi personali e sociali, principalmente settori non di mercato.

L'indicatore misura su base oraria la dinamica congiunturale del costo del lavoro come somma di retribuzioni e oneri sociali riferito ai lavoratori dipendenti. Poiché riflette i cambiamenti non solo di retribuzioni e oneri sociali ma anche della quantità delle ore effettivamente lavorate², fornisce indicazioni sul peso del costo derivante dal fattore produttivo lavoro. In questo, differisce dall'indicatore Oros relativo al costo del lavoro per Ula che non tiene conto della effettiva quantità di ore lavorate dai dipendenti con contratto a tempo pieno. Le ore effettivamente lavorate rappresentano la misura più appropriata dell'input di lavoro in quanto tengono conto sia delle fluttuazioni nella quantità di lavoro prestato (ad es. straordinario), sia dei cambiamenti nella composizione occupazionale (es. aumento dei contratti part-time).

¹ Il presente capitolo è a cura di Francesca Ceccato e Eleonora Cimino.

² Sono escluse le ore pagate ma non effettivamente lavorate (ferie annuali, festività e assenze per malattia, ecc.).

11.2 L'indice LCI: caratteristiche metodologiche

Il regolamento sull'indice prevede il rilascio degli indicatori orari delle retribuzioni, oneri sociali e costo del lavoro. Le serie storiche degli indicatori sono attualmente espresse in base 2012 e sono disponibili a partire dal I trimestre 2000 in forma grezza, corretta per la composizione del calendario e destagionalizzata. Gli indicatori vengono calcolati per ogni settore di attività dell'industria e dei servizi, inclusa la pubblica amministrazione (sezioni da B a S Ateco 2007) e per gli aggregati economici richiesti dal Regolamento CE³. La stima dell'indicatore aggregato LCI richiede il calcolo degli indici elementari a livello di sezione per ciascuna delle tre variabili economiche. Per semplicità, se si considera la variabile costo del lavoro orario, l'indice semplice calcolato per la generica sezione j è dato dal rapporto tra il costo del lavoro orario nel trimestre t del generico anno a e il costo medio del lavoro orario nell'anno base b , in formula:

$$I_{j,ta,b} = \frac{w_{j,ta}}{\omega_{j,b}} \cdot 100 \quad (1)$$

dove:

$w_{j,ta}$ = costo del lavoro orario nel settore economico j , trimestre t dell'anno a ;
 $\omega_{j,b}$ = costo medio del lavoro orario nel settore economico j , anno base b .

Il regolamento CE n.1216/2003 prevede che per ciascuna variabile *target* l'indice LCI relativo agli aggregati settoriali sia definito come indice di Laspeyres concatenato al 2000 con pesi annuali calcolati a livello di sezione Ateco 2007.

Per ottenere l'indice concatenato annualmente come richiesto da Eurostat è necessario dapprima calcolare l'indice di Laspeyres come media ponderata degli indici elementari $I_{j,ta,b}$ usando come pesi il rapporto tra il costo del lavoro nelle sezioni che concorrono all'aggregato e il costo del lavoro dell'aggregato nell'anno base.

L'indice di Laspeyres sul costo del lavoro riferito al trimestre t dell'anno a , con anno base b è definito come segue:

$$LCI_{ta,b} = \frac{\sum_j w_{j,ta} h_{j,b}}{\sum_j \omega_{j,b} h_{j,b}} = \frac{\sum_j (w_{j,ta} / \omega_{j,b}) \omega_{j,b} h_{j,b}}{\sum_j W_{j,b}} = \frac{\sum_j (w_{j,ta} / \omega_{j,b}) W_{j,b}}{\sum_j W_{j,b}} \quad (2)$$

dove:

$h_{j,b}$ = ore lavorate nel settore economico j , anno base b
 $W_{j,b} = \omega_{j,b} h_{j,b}$ = costo del lavoro nel settore economico j , anno base b .

3 A partire da dicembre 2016 con il rilascio degli indicatori LCI riferiti al III trimestre 2016, sono stati diffusi i seguenti aggregati settoriali: B-E industria in s.s., B-F industria, B-N industria e servizi di mercato, B-S industria e servizi, B-S escluso O industria e servizi esclusa l'amministrazione pubblica e difesa; assicurazione sociale obbligatoria, G-J commercio all'ingrosso e al dettaglio; trasporto e magazzinaggio; attività di servizi, di alloggio e ristorazione; servizi di informazione e comunicazione, G-N servizi di mercato, G-S servizi, K-N attività finanziarie e assicurative; attività immobiliari; attività professionali scientifiche e tecniche; noleggio, agenzie di viaggio e servizi di supporto alle imprese, O-S amministrazione pubblica e difesa; assicurazione sociale obbligatoria; istruzione; sanità e assistenza sociale; attività artistiche e altre attività di servizi e P-S istruzione, sanità e assistenza sociale, attività artistiche e altre attività di servizi.

I pesi all'anno base b sono dati dal rapporto tra il costo del lavoro della sezione j e il costo del lavoro dell'aggregato:

$$\frac{W_{j,b}}{\sum_j W_{j,b}} \quad (3)$$

Per tenere conto adeguatamente dei cambiamenti nella struttura dei pesi nel corso degli anni, l'indice LCI trimestrale viene ponderato con pesi variabili annualmente e non fissi all'anno base ($LCI_{ta,a-m}$). Secondo quanto suggerito da Eurostat, i pesi usati nel calcolo dell'indice devono essere riferiti ad un periodo massimo di due anni precedenti rispetto all'anno cui si riferiscono. Per il calcolo dell'indicatore italiano questi pesi si riferiscono all'anno precedente ($m=1$). Poiché l'anno base cambia al variare degli anni, vengono costruiti dei link annuali al fine di rendere coerenti gli indici calcolati con basi di riferimento diverse e poterli unire tenuto conto dello stesso periodo di riferimento che rimane fisso ($b=0$).

I link annuali tra l'anno l e l'anno $l+1$ sono calcolati e definiti come:

$$L_{l,l+1} = \frac{\sum_j \omega_{j,l+1} h_{j,l}}{\sum_j \omega_{j,l} h_{j,l}} \quad (4)$$

dove $0 \leq l < l+1 < a$.

Questi link hanno anch'essi la formulazione di indici di Laspeyres sul costo del lavoro tra l'anno $l+1$ e l'anno precedente l .

L'indice di Laspeyres concatenato annualmente riferito al trimestre t dell'anno a con anno base b si ottiene, infine, secondo la seguente formula:

$${}_C LCI_{ta,b} = \prod_{l=b}^{a-2} L_{l,l+1} \cdot LCI_{ta,a-1} \quad \text{con } a > b \quad (5)$$

dove $\prod_{l=b}^{a-2} L_{l,l+1} = (L_{b,b+1})(L_{b+1,b+2}) \dots (L_{a-2,a-1})$

$$\text{e } {}_C LCI_{ta,b} = \prod_{l=a-1}^{b-1} (L_{l,l+1})^{-1} \cdot LCI_{ta,a-1} \quad \text{con } a \leq b \quad (6)$$

in cui $\prod_{l=a-1}^{b-1} (L_{l,l+1})^{-1} = (L_{a-1,a})^{-1} (L_{a,a+1})^{-1} \dots (L_{b-1,b})^{-1}$

Il primo anno di riferimento a è il 2000, quando il link annuale è pari a 1.

Il calcolo di LCI come un indice di Laspeyres concatenato annualmente riduce gli effetti potenziali dei cambiamenti nella composizione della forza lavoro. Tuttavia, tale tipologia di indice è caratterizzata da effetti di composizione dovuti a cambiamenti nella qualità dell'occupazione (es. *skilled/unskilled*, aumento della produttività lavorativa) e nel capitale umano. L'indice di Laspeyres concatenato così ottenuto (cfr. formule (5) e (6)) non preserva la proprietà dell'additività dell'indice di Laspeyres (cfr. formula (2)). Per superare questo in-

conveniente è possibile derivare un nuovo sistema di pesi annuali ottenuti applicando opportuni coefficienti al sistema originario dei pesi (Ciammola e Tuzi, 2010). Nel caso di $a > b$, il nuovo indice ${}_{CA}LCI_{ta,b}$ per un generico aggregato settoriale riferito all'anno base b si ottiene indirettamente come somma degli indici elementari a livello di sezione riferiti allo stesso anno base a cui si applicano i nuovi pesi. In formula:

$${}_{CA}LCI_{ta,b} = \sum_j \frac{\prod_{l=b}^{a-2} L_{l,l+1}}{m(I_{j,t,b})} \cdot \frac{W_{j,b}}{\sum_j W_{j,b}} \cdot I_{j,ta,b} \quad \text{con } a > b \quad (7)$$

dove

$$\frac{\prod_{l=b}^{a-2} L_{l,l+1}}{m(I_{j,t,b})} \cdot \frac{W_{j,b}}{\sum_j W_{j,b}}$$

rappresenta il nuovo sistema di pesi nel caso $a > b$ ottenuto applicando ai pesi iniziali un fattore correttivo che ripristina l'additività perduta con il concatenamento e:

$${}_{CA}LCI_{ta,b} = \sum_j \frac{m(I_{j,t,b})}{\prod_{l=a-1}^{b-1} L_{l,l+1}} \cdot \frac{W_{j,b}}{\sum_j W_{j,b}} \cdot I_{j,ta,b} \quad \text{con } a \leq b \quad (8)$$

in cui il fattore correttivo viene ottenuto come reciproco dell'analogo fattore usato in (7).

Quando il nuovo sistema di pesi viene applicato è necessario prestare attenzione a due proprietà:

1. solo in corrispondenza dell'anno base b , e degli anni $b-1$ e $b+1$, la somma dei pesi delle singole sezioni j equivale ad 1;
2. i pesi non possono essere usati per derivare aggregati diversi che contengono l'aggregato per cui sono stati calcolati, come semplice sommatoria, ma per ogni livello di aggregazione desiderato vanno ricalcolati i pesi ex-novo (in pratica, la non additività degli indici aggregati concatenati viene trasferita al nuovo sistema di pesi). Per superare tale inconveniente sono stati effettuati avanzamenti metodologici che verranno a breve implementati nel caso dell'indice LCI italiano (Ciammola, Tuzi, 2015), portando ad una radicale semplificazione del sistema di pesi attualmente utilizzato.

La procedura di aggregazione riportata nella (7) viene utilizzata per determinare gli indici degli aggregati in forma grezza, destagionalizzata e corretta per gli effetti di calendario a partire dai rispettivi indici semplici grezzi e corretti, configurandosi come un approccio indiretto di aggregazione. L'uso di tale approccio garantisce l'assenza di incongruenze nelle variazioni degli indici delle sezioni e degli aggregati corrispondenti. Tale vantaggio è ancor più apprezzabile per le serie destagionalizzate in quanto consente una maggiore libertà nell'identificazione dei modelli delle singole sezioni con un ritorno positivo sulla qualità delle serie trattate. Tuttavia, il passaggio all'approccio indiretto comporta una serie di attività aggiuntive da dover affrontare, in particolare il calcolo e il mantenimento di insiemi aggiuntivi di pesi rispetto a quelli originari.

Il caso dell'aggregazione settoriale, affrontato nella (7) e nella (8) con riferimento al costo del lavoro orario, è valido anche per l'aggregazione settoriale delle due variabili componenti (le retribuzioni e gli oneri sociali orari). L'aggregazione indiretta nella dimensione della variabile costo del lavoro, a partire dalle retribuzioni e dagli oneri sociali orari, rappresenta invece una media degli indici semplici delle componenti ponderata con pesi fissi all'anno base e a somma 1, dati dal monte retributivo e monte oneri sociali rispetto al monte costo del lavoro. Tali pesi, distinti per sezione, sono validi sia per l'aggregazione degli indici in forma grezza, sia in forma corretta.

11.3 Le fonti

Per il calcolo dell'indice LCI italiano si ricorre a dati provenienti da fonti diverse, la cui integrazione consente di rispondere alle richieste dei regolamenti Eurostat. Come anticipato nel paragrafo 1.1, la rilevazione Oros e la rilevazione Vela-GI rappresentano le fonti principali per derivare, rispettivamente, le variabili componenti il costo del lavoro a numeratore dell'indice LCI (retribuzioni e oneri sociali) e le ore lavorate pro-capite a denominatore dell'indicatore⁴. Queste fonti appartengono ad un sistema di processi unificati (in termini di definizioni, raccolta, fasi di processo e output prodotti) aventi l'obiettivo di produrre indicatori congiunturali armonizzati su input e costo del lavoro (Baldi et al, 2011). Tuttavia, tali fonti coprono solo la componente privata dell'occupazione e per questo vengono utilizzate solo per le stime relative ai settori dell'industria e dei servizi di mercato (da B a N dell'Ateco 2007) nei quali la componente pubblica dell'occupazione è trascurabile. In questo sistema la rilevazione Oros è usata sia come popolazione di riferimento sia come fonte censuaria per le posizioni lavorative dipendenti medie trimestrali e per le retribuzioni e gli oneri sociali, mentre la rilevazione campionaria Vela-GI rileva le posizioni lavorative dipendenti alla fine del trimestre, i posti vacanti e le ore lavorate riportandole alla popolazione di riferimento definita da Oros. Il rispetto dei tempi stabiliti per le diverse fasi del processo è il requisito fondamentale per la tenuta del sistema unificato, il cui obiettivo è la tempestività dei rilasci nazionali ed europei.

Per soddisfare la copertura dei settori da O a S dell'Ateco 2007, in cui invece la componente pubblica è rilevante, si usano i dati sul costo del lavoro e sulle ore lavorate forniti dalla Contabilità Nazionale (CN) che completano i dati prodotti dal processo unificato utilizzato per la stima dei settori da B a N, entrando a far parte a tutti gli effetti del sistema integrato sulle statistiche congiunturali su input e costo del lavoro delle imprese. L'integrazione dei dati di CN, ovviamente, ha implicato un'analisi preliminare definitoria delle variabili volta a verificarne la coerenza anche nel rispetto delle richieste europee. L'analisi ha evidenziato che le differenze principali tra le fonti usate riguardano la definizione dell'input di lavoro e il criterio di attribuzione temporale delle variabili stimate. In particolare, nella stima fornita dalla CN, che segue le definizioni ESA 2010 (Regolamento CE n.549/2013), sono inclusi anche i lavoratori irregolari e i dirigenti. Inoltre, mentre la CN utilizza il criterio di competenza per stimare le variabili economiche, e quantifica anche le retribuzioni in natura e le mance, Oros e Vela-GI usano il criterio di cassa per attribuire nel trimestre le poste retributive e contributive e coprono solo in minima parte le componenti in natura e le mance. Per ov-

⁴ Nella rilevazione Vela-GI, le ore di lavoro pro-capite sono ottenute rapportando le ore di lavoro ordinario e straordinario prestate dai dipendenti con contratto di lavoro alle posizioni lavorative dipendenti. Sono quindi escluse le ore non lavorate per ferie, malattia, cassa integrazione guadagni, solidarietà, ecc. anche se retribuite.

viare alla presenza degli irregolari nei dati di CN, ai fini della stima dell'indicatore LCI per le sezioni da O a S, si utilizzano i dati stimati dalla CN attraverso un'elaborazione *ad-hoc* che esclude tale componente occupazionale. Le altre differenze, non eliminabili e delle quali si tiene conto in fase di analisi dei risultati, non compromettono comunque l'utilizzo integrato delle fonti Istat a disposizione. In dettaglio, le stime trimestrali che la CN produce *ad-hoc* per la compilazione dell'indice LCI italiano sono ottenute tramite tecniche di disaggregazione temporale basate su indicatori trimestrali propriamente scelti applicati alle stime dei conti annuali. Le stime trimestrali così ottenute non sono confrontabili con le stime dei conti trimestrali ufficiali di CN che includono l'occupazione regolare e irregolare.

Per ogni generico settore j , con riferimento al trimestre t dell'anno a il costo del lavoro orario è dato dal rapporto tra i relativi monte costo del lavoro (C) e monte ore lavorate (Ore). La formula che segue mostra il contributo al calcolo del costo del lavoro orario trimestrale ($w_{j,ta}$) della singola indagine evidenziando l'elevato livello di integrazione che caratterizza il sistema:

$$w_{j,ta} = \frac{C_{j,ta}}{Ore_{j,ta}} = \begin{cases} \frac{OROS^{C_{j,ta}} * OROS^{D_{j,ta}}}{VELA-GI^{Ore_{j,ta}} * OROS^{D_{j,ta}}} & \text{con } j \in B, \dots, N \text{ Ateco 2007} \\ \frac{CN^{C_{j,ta}}}{CN^{Ore_{j,ta}}} & \text{con } j \in O, \dots, S \text{ Ateco 2007} \end{cases} \quad (9)$$

Dove, per i settori da B a N Ateco 2007, $OROS^{C_{j,ta}}$ è il costo del lavoro pro-capite prodotto da Oros e $VELA-GI^{Ore_{j,ta}}$ sono le ore lavorate pro-capite calcolate dalla rilevazione Vela-GI. La riconciliazione tra le fonti è garantita dalle posizioni lavorative dipendenti di Oros ($OROS^{D_{j,ta}}$) che moltiplicate per le ore lavorate pro-capite di Vela-GI restituiscono un monte ore coerente con il numeratore della formula. Per i settori da O a S, $CN^{C_{j,ta}}$ e $CN^{Ore_{j,ta}}$ sono, rispettivamente, il monte redditi e il monte ore trimestrali prodotti *ad-hoc* dalla CN per LCI.

Il costo del lavoro orario definito dalla (9) è alla base del calcolo dell'indice LCI come evidenziato nel paragrafo precedente relativo alla metodologia di calcolo.

L'uso integrato di più fonti implica che la politica di revisione degli indicatori LCI dipenda dalle politiche di revisione delle singole indagini. Il numeratore del costo del lavoro orario (cfr. formula (9)) dipende, per le sezioni da B a N, unicamente dalla rilevazione Oros che produce ogni trimestre una stima preliminare delle variabili di costo del lavoro riferita a t e una stima finale riferita a $t-4$ che tiene conto di informazioni più complete ed aggiornate rese disponibili nell'arco di un anno, con riferimento ai dati amministrativi dell'Inps (cfr. §2.3.1). In occasione della diffusione del I trimestre di ogni anno, oltre alla stima riferita a $t-4$ vengono riviste anche le stime riferite ai trimestri $t-3$, $t-2$ e $t-1$ per tener conto anche delle revisioni apportate alle variabili di costo del lavoro come prevede la revisione *standard* annuale delle stime mensili di GI sui dodici mesi dell'anno precedente che vengono integrati con i dati di fonte amministrativa. Il denominatore, dipendendo dalle posizioni lavorative Oros e dalle ore lavorate pro-capite di Vela-GI, segue la stessa politica di revisione Oros illustrata per il numeratore a cui si aggiunge la politica di revisione Vela-GI. Quest'ultima prevede, a giugno di ogni anno, la diffusione delle nuove stime delle ore lavorate pro-capite riviste per i trimestri di $a-1$, per tener conto della revisione annua dei dati mensili GI, e $a-2$, per tener conto della popolazione definitiva di riferimento trimestrale di fonte Oros acquisita da Vela-GI in questa occasione con riferimento ai trimestri di tale anno.

Per le sezioni da O a S, il numeratore e il denominatore del costo del lavoro orario dipendono principalmente dalle revisioni dei Conti trimestrali di CN, ma anche da affinamenti metodologici apportati alla procedura *ad hoc* usata. La politica dei dati grezzi di CN prodotti dalla procedura *ad hoc*, input nel processo LCI, segue lo schema sotto specificato seguendo il rilascio trimestrale:

- I trimestre di *a* (LCI diffuso a giugno di *a*), viene interamente rivisto *a-1*;
- II trimestre di *a* (LCI diffuso a settembre di *a*), sono rivisti *a-1* e *t-1*;
- III trimestre di *a* (LCI diffuso a dicembre di *a*), vengono rivisti *t-1*, *t-2* e gli ultimi quattro anni (fino ad *a-4*);
- IV trimestre di *a* (LCI diffuso a marzo di *a+1*), vengono rivisti *t-1*, *t-2*, *t-3* e gli ultimi tre anni (fino ad *a-3*).

Si sono verificate in rare occasioni revisioni non standard dovute al disallineamento temporale nella disponibilità dei dati trimestrali di fonte CN ai fini del calcolo dell'indicatore LCI, che hanno comportato revisioni non trascurabili nell'indicatore LCI posticipate rispetto a quanto avvenuto sugli indicatori trimestrali di CN.

11.4 Il processo di produzione dell'indice LCI italiano

In questo paragrafo, seguendo la metodologia descritta nel paragrafo 11.2, si evidenziano le diverse fasi del processo trimestrale che portano alla stima degli indici LCI italiani richiesti dai Regolamenti europei.

La figura 11.4.1 riporta la prima fase (step 1) in cui vengono calcolati, per la generica sezione *j* da B a S Ateco 2007, gli indici semplici per le due variabili componenti il costo del lavoro: le retribuzioni orarie ($I^R_{j,ta,b}$) e gli oneri sociali orari ($I^O_{j,ta,b}$). A partire da tali indici, calcolati come nella (1), si procede ad una seconda fase di aggregazione indiretta (step 2) per ottenere il costo del lavoro orario ($I^C_{j,ta,b}$) come sintesi degli indici prodotti nella fase precedente. Come accennato sopra, i pesi necessari per l'aggregazione indiretta del costo del lavoro dipendono solo dal peso medio annuo della singola variabile oraria rispetto al costo del lavoro orario nell'anno base dell'indice LCI, attualmente 2012. Nella terza fase (step 3) vengono calcolati i pesi delle sezioni appartenenti ai singoli aggregati richiesti⁵ e si procede all'aggregazione settoriale indiretta per le tre variabili considerate (retribuzioni orarie, oneri sociali orari e costo del lavoro orario). I pesi di aggregazione settoriale vanno ricalcolati per ogni aggregato, variano trimestralmente e seguono la metodologia ampiamente trattata nel paragrafo 11.2.

Attualmente per l'indice LCI italiano si producono 261 indicatori, risultato di:

- 3 variabili (di cui 1 con approccio indiretto);
- 18 sezioni Ateco 2007;
- 11 aggregati settoriali (tutti con approccio indiretto);
- in forma grezza, corretti per i soli effetti di calendario, corretti per gli effetti di calendario e la stagionalità.

La correzione per i soli effetti di calendario, come anche per la stagionalità, avviene seguendo la stessa politica di aggregazione indiretta già illustrata per gli indici non corretti, usando gli stessi sistemi di pesi annuali e trimestrali usati per l'aggregazione degli indici grezzi applicati alle serie corrette.

⁵ Per l'elenco degli aggregati si veda la nota 3 del capitolo.

L'approccio di destagionalizzazione usato, come per gli indicatori per la diffusione nazionale, è il *partial concurrent* (cfr. §9.1). Prima di procedere alla destagionalizzazione, una volta l'anno a giugno, in occasione del rilascio degli indicatori riferiti al I trimestre si valuta la tenuta nel tempo del modello in uso e se necessario si identificano nuovi modelli reg-Arima. Anche per l'indicatore LCI la destagionalizzazione viene effettuata su SITIC attraverso l'applicativo TRAMO-SEATS.

Figura 11.4.1 – Fasi di processo dell'indice LCI italiano

	R _{orarie}	O _{orari}	C _{orario}
B	Step 1		Step 2
C			
D	$I_{j,ta,b}^R$	$I_{j,ta,b}^O$	$I_{j,ta,b}^C$
....			
N			
O			
P			
Q			
R			
S			
B-N	Step 3 ${}_{CA}LCI_{B-N,ta,b}$		
....			
O-S	Step 3 ${}_{CA}LCI_{O-S,ta,b}$		
....			
B-S	Step 3 ${}_{CA}LCI_{B-S,ta,b}$		

Le scelte di aggregazione sono acquisite dal sistema SITIC. In particolare, sia per il costo del lavoro orario sia per l'aggregazione settoriale delle tre variabili, è necessario calcolare i pesi relativi specifici delle singole componenti, in modo che la somma di tali pesi dia 1. In totale, il numero di pesi per l'aggregazione indiretta del costo del lavoro è pari a 36, uno per ogni sezione per le 2 variabili componenti, mentre il numero dei pesi per l'aggregazione degli 11 aggregati settoriali è 285 moltiplicato per il numero di anni della serie storica meno l'ultimo anno. Come descritto nel paragrafo 11.2, infatti, il sistema dei pesi di aggregazione dipende da tutti gli anni fino ad $a-1$ e, pertanto, segue la stessa politica di revisione degli indici grezzi. Gli indici e i pesi vengono caricati trimestralmente sul sistema SITIC. Gli output di TRAMO-SEATS della destagionalizzazione diretta vengono controllati per verificare la bontà dei risultati ottenuti applicando i modelli di destagionalizzazione presenti sul sistema. Solo in caso di individuazione di un nuovo modello, si procede alla sua sostituzione attraverso l'aggiornamento su SITIC della tabella relativa ai modelli. Finita questa fase, si procede alla destagionalizzazione delle rimanenti serie che seguono l'approccio indiretto. Le serie corrette per la stagionalità così ottenute vengono sottoposte alla verifica di assenza di stagionalità residua ed, infine, vengono analizzate le revisioni nelle variazioni congiunturali delle serie prodotte per valutarne la qualità.

Tutti gli indicatori, in forma grezza e corretta, vengono prodotti in formato SDMX e inviati trimestralmente ad Eurostat attraverso EDAMIS (cfr. §2.5). Oltre agli indicatori, si forniscono anche i pesi monti relativi alle tre variabili *target* per sezione e anno della serie storica in formato SDMX anch'essi inviati via Edamis per consentire ad Eurostat di calcolare l'indicatore europeo. I dati inviati sono accompagnati da un documento dal titolo *Meta Information*, che ne raccoglie i metadati e viene inviato tramite e-mail. Annualmente, ad agosto, viene redatto un *Quality Report* relativo all'LCI dell'anno precedente che segue gli standard europei sulla qualità ESQRS, condiviso tramite l'applicativo ESS-MetadataHandler come richiesto da Eurostat e pubblicato nella sezione *National metadata* al [link](#)

http://ec.europa.eu/eurostat/cache/metadata/EN/lci_esqrs.htm.

Tale documentazione contribuisce anche alla redazione di un report biennale di Eurostat alla Commissione Europea.

Attualmente, LCI viene diffuso trimestralmente a livello europeo sul DB EUROIND di Eurostat al [link](http://ec.europa.eu/eurostat/data/database) <http://ec.europa.eu/eurostat/data/database>⁶

e concorre alla stima dell'*Annual growth in labour costs* diffusa come *News release - Euro indicators* sul sito ufficiale al [link](http://ec.europa.eu/eurostat/news/news-releases) <http://ec.europa.eu/eurostat/news/news-releases>.

Inoltre, Eurostat aggiorna trimestralmente una pagina informativa con dati e metadati al [link](http://ec.europa.eu/eurostat/statistics-explained/index.php/Labour_cost_index_-_recent_trends) http://ec.europa.eu/eurostat/statistics-explained/index.php/Labour_cost_index_-_recent_trends

L'indice LCI annualmente viene utilizzato da Eurostat quale indicatore per estrapolare il livello del costo del lavoro orario medio, a partire dagli indicatori LCS diffusi ogni 4 anni dai Paesi membri⁷. Gli *Hourly labour costs* vengono pubblicati solitamente nel mese di aprile al [link](http://ec.europa.eu/eurostat/statistics-explained/index.php/Hourly_labour_costs) http://ec.europa.eu/eurostat/statistics-explained/index.php/Hourly_labour_costs.

⁶ Alla sezione Euro indicators / PEEIs - Labour market - Labour cost index / quarterly data.

⁷ La rilevazione LCS, prodotta ogni 4 anni, fornisce una stima sulla struttura del costo del lavoro. Attualmente, l'ultimo LCS si riferisce al 2012.

12. GLI INDICATORI PER IL REGOLAMENTO STS – SHORT TERM BUSINESS STATISTICS¹

12.1 Le richieste del Regolamento

Alla rilevazione Oros è demandato anche il compito di rispondere alle richieste del regolamento STS del Parlamento Europeo e del Consiglio sulle statistiche congiunturali (n.1165/98)². In particolare, devono essere compilati indici grezzi trimestrali sul numero totale di persone occupate (*Number of persons employed*) e sul monte retributivo dei lavoratori dipendenti (*Gross wages and salaries*) nei settori Ateco 2007 da B ad N, in riferimento ad imprese ed istituzioni private.

Mentre la seconda variabile (variabile 230) viene dedotta esclusivamente dai dati Oros, per la compilazione dell'indicatore sul numero di occupati totali (variabile 210) è necessario ricorrere a fonti ausiliarie. In particolare, deve essere stimato il numero trimestrale degli *indipendenti*, assenti sia dai dati amministrativi sui DM2013, sia dai dati dell'Indagine GI. Attualmente, la fonte che rilascia dati sugli indipendenti al livello di dettaglio Ateco più fine e aderente alle richieste del regolamento è la Contabilità Nazionale che, tuttavia, su questa variabile diffonde solo dati annuali richiedendo, come verrà descritto nel paragrafo successivo, l'applicazione di una tecnica di disaggregazione trimestrale³.

Di seguito (Prospetto 12.1.1) si riporta il livello di dettaglio per Ateco richiesto dal regolamento per le due variabili. Si noti che per alcuni aggregati è necessario scendere al di sotto della divisione Ateco, livello massimo per cui sono disponibili le stime Oros e i dati sugli indipendenti di fonte Contabilità. Per ottenere questi aggregati si ricorre ad un sistema di pesi che individua la quota parte di ogni aggregato sulle divisioni di riferimento. Tale *set* viene ricavato dal sistema ponderale che Eurostat concorda con i paesi membri in occasione di ogni cambio base, pesi riferiti alla media dell'anno base e mantenuti costanti fino all'anno base successivo e utili per ricavare, a partire dagli indicatori nazionali, eventuali altri aggregati per Ateco a livello nazionale e i principali aggregati europei. Il sistema ponderale viene individuato ogni cinque anni attraverso un interscambio di informazioni tra Eurostat e i singoli paesi. L'Italia, come indicato da Eurostat, calcola i pesi a partire dalle fonti che vengono utilizzate per rispondere alle richieste del regolamento sulle statistiche strutturali (Reg. UE 295/2008 - SBS) ossia principalmente attraverso il sistema informativo Frame-SBS⁴. Per il monte retributivo, Oros calcola internamente tali pesi a partire dai

1 Il presente capitolo è a cura di Marco Lattanzio e Donatella Tuzi.

2 Per approfondimenti sulle richieste del Regolamento STS, si veda il link:
<http://ec.europa.eu/eurostat/web/short-term-business-statistics/overview/sts-in-brief>.

3 La rilevazione sulle Forze di Lavoro stima il numero di occupati indipendenti. Tuttavia, la validazione di tale dato a livello Ateco, variabile che viene comunicata dal lavoratore, avviene a livello di sezione. I dati sugli indipendenti vengono tra l'altro diffusi per aggregazioni di sezione, livello di dettaglio totalmente insufficiente per rispondere alle richieste STS.

4 Il sistema informativo Frame è una base di microdati di fonte amministrativa trattati statisticamente e combinati con quelli delle rilevazioni strutturali SBS. A partire dal 2012 l'Istat risponde alle richieste del regolamento SBS, precedentemente assolte con l'uso di dati delle due indagini strutturali SCI e PMI, esclusivamente attraverso l'elaborazione del nuovo sistema informativo. Per maggiori informazioni sul Frame si veda
<http://www.istat.it/en/files/2014/11/Nota-metodologica7.pdf>.

microdati, disponibili fino a 5-digit Ateco sia per le unità stimate con la fonte amministrativa sia con l'indagine GI. Tali pesi vengono comunicati ad Eurostat e sostituiti a quelli stimati secondo i criteri SBS per definire il sistema ponderale per l'Italia⁵. Stesso non avviene per il totale occupati, non essendo disponibili dati a livello più fine sugli indipendenti. A tale scopo, i pesi utili per determinare gli aggregati al di sotto della divisione sono quelli stimati con fonte Frame-SBS.

Prospetto 12.1.1 – Dettaglio dei settori di attività economica richiesti dal Reg. STS per le variabili “totale persone occupate” e “retribuzioni” (a)

ALLEGATI REG.STS	VARIABILI	
	Totale persone occupate (variabile 210)	Retribuzioni (variabile 230)
Allegato A - Industria	Tutte le divisioni ad esclusione di 37, 38 e 39, relative sezioni e aggregato B-E MIGs: beni intermedi, beni capitali, beni di consumo durevoli e non durevoli, energia	Tutte le divisioni ad esclusione di 37, 38 e 39, relative sezioni e aggregato B-E MIGs: beni intermedi, beni capitali, beni di consumo durevoli e non durevoli, energia
Allegato B - Costruzioni	A livello di sezione	A livello di sezione
Allegato C - Commercio al dettaglio e riparazioni	Divisioni 45,46,47 e in più 47_x_47.3 (aggregazione dei gruppi 47.1, 47.2, 47.4, 47.5, 47.6, 47.7, 47.8, 47.9)	Divisioni 45,46,47 e in più 47_x_47.3 (aggregazione dei gruppi 47.1, 47.2, 47.4, 47.5, 47.6, 47.7, 47.8, 47.9)
Allegato D - Altri servizi	Divisioni 49, 50, 51, 52, 53, 58, 59, 60, 61, 62, 63; sezione I, sezione M-STs (aggregazione di 69, 70.2, 71, 73 e 74), sezione N-STs (aggregazione di 78, 79, 80, 81.2 e 82)	Sezioni H, I, J, M-STs, N-STs

(a) Reg. CE n.1158/2005 di implementazione del Reg. CE n.1165/98 (Short-Term Statistics)

L'invio degli indicatori per STS prevede per il totale occupati una scadenza di 60 giorni dalla fine del trimestre di riferimento e per le retribuzioni di 90 giorni, con ogni mese da intendersi da 30 giorni. In tal modo, le scadenze sono fisse tra gli anni e corrispondono per le posizioni totali al 31 maggio, 31 agosto, 30 novembre e 28/29 febbraio, mentre per le retribuzioni al 30 giugno, 30 settembre, 31 dicembre, 31 marzo. Poiché le due variabili vengono elaborate contestualmente, gli indicatori vengono inviati insieme entro i 60 giorni. Gli indicatori vengono spediti via EDAMIS, un applicativo web di Eurostat che consente, attraverso un concetto SEP (*Single Entry Point*), di trasmettere i dati in via automatica. I dati vengono richiesti in formato GESMES, compilato attraverso SITIC (cfr. §8.5). Vengono inviati, in totale, 8 diversi file, uno per ogni variabile e per ogni dominio di riferimento del Regolamento (Allegato A, Allegato B, Allegato C e Allegato D) per un totale di 110 serie di dati.

Mentre il monte retributivo viene inviato ad Eurostat in formato *free*, il numero di persone occupate viene tuttora rilasciato in forma *confidential*, principalmente a causa dell'entità non ancora trascurabile delle revisioni. Tali revisioni sono in larga parte dovute alla revisione dei dati sugli indipendenti stimati da fonte Contabilità Nazionale, come verrà spiegato in maggior dettaglio nel paragrafo successivo.

Oltre ai dati, i singoli paesi devono compilare un *Metadata Template* in formato SDMX *Metadata Structure* (ESMS), nel quale vengono riportate informazioni generali sulla rilevazione, sulla metodologia utilizzata per la stima degli indicatori e sulla qualità degli indicatori

⁵ E' necessario precisare che l'utilizzo di pesi di fonte Oros per le ponderazioni utili al calcolo di indici relativi ad aggregati superiori di diffusione deriva dal fatto che la rilevazione calcola indici semplici per ogni livello. La metodologia di calcolo in questione sottintende l'utilizzo di un sistema di pesi fisso all'anno base, calcolato con le stesse variabili con cui si calcolano gli indici. Per maggiori dettagli si veda il paragrafo 9.2.2.

forniti, secondo gli standard di qualità dettati da Eurostat. Tale file, che si basa su una struttura prefissata, viene aggiornato da parte dei paesi secondo scadenze indicate da Eurostat⁶.

Di seguito si descrive la metodologia di stima adottata per la compilazione dei due indicatori, dall'acquisizione dei dati di fonte Contabilità e loro aggregazione alla componente dipendenti stimata da fonte Oros, alla ricostruzione del monte retributivo per acquisire la correzione per le mancate risposte e per l'*editing* selettivo (cfr. §6).

12.2 La metodologia per la compilazione degli indicatori

Al fine di completare la definizione di “totale persone occupate” richiesta da STS, la metodologia individuata da Oros prevede l'uso di dati annuali sulle posizioni lavorative totali definitive secondo i criteri di input di lavoro dettati dal SEC2010⁷. Tali posizioni vengono acquisite solo relativamente alla tipologia regolare e distinte in posizioni dipendenti e posizioni indipendenti.

Essendo disponibili solo a livello annuale, l'utilizzo dei dati di Contabilità Nazionale (d'ora in poi CN) richiede due passaggi: 1) ricostruzione settoriale tramite disaggregazione delle stime fornite da CN, non sempre disponibili a livello di divisione Ateco, almeno non per tutti gli anni di serie storica diffusi; 2) trimestralizzazione del dato annuale.

Le operazioni seguite per arrivare alla stima della variabile trimestrale per i domini *target* vengono di seguito descritte.

Si definiscono con:

Se_a^{cn} = totale indipendenti di fonte CN, disponibili a livello annuale per l'anno a a livello di divisione Ateco 2007;

E_{t_a} = totale dipendenti di fonte Oros, corrispondenti alla definizione utilizzata nel §8.4, disponibili per il trimestre t dell'anno a (t_a), anch'essi disponibili a livello di divisione Ateco 2007.

La stima trimestrale degli addetti Oros (\hat{P}_{t_a}) si ottiene sommando ai dipendenti trimestrali Oros una stima degli indipendenti trimestrali (\widehat{SE}_{t_a}). Quest'ultima è calcolata applicando ai dipendenti trimestrali di Oros un coefficiente ottenuto dal rapporto tra indipendenti annuali di fonte CN e la somma dei dipendenti trimestrali di fonte Oros nell'anno a e per ogni divisione Ateco 2007, come segue:

$$\hat{P}_{t_a} = (1 + \alpha_a) E_{t_a} = E_{t_a} + \widehat{SE}_{t_a} \quad (1)$$

e:

$$\widehat{SE}_{t_a} = \alpha_a \cdot E_{t_a} \quad (2)$$

dove il coefficiente, ipotizzato costante per l'intero anno, si ottiene come:

$$\alpha_a = \frac{cnSE_a}{\frac{1}{4} \sum_{t=1}^4 E_{t_a}} \quad (3)$$

6 L'ultima versione disponibile è aggiornata al 2017 e, per l'Italia, si trova al link: http://ec.europa.eu/eurostat/cache/metadata/EN/sts_inlb_m_esms_it.htm

7 Il nuovo sistema dei Conti è dettato dal Regolamento UE n.549/2013 del Parlamento europeo e del Consiglio.

Dalla relazione precedente segue che la media trimestrale degli indipendenti stimati Oros corrisponde al valore degli indipendenti di fonte CN infatti:

$$\begin{aligned} \frac{\sum_{t_a=1}^4 \widehat{SE}_{t_a}}{4} &= \frac{\sum_{t_a=1}^4 \alpha_a E_{t_a}}{4} = \frac{\sum_{t_a=1}^4 \frac{cnSE_a}{\frac{1}{4} \sum_{t_a=1}^4 E_{t_a}} E_{t_a}}{4} = \\ &= \frac{\frac{\sum_{t_a=1}^4 E_{t_a}}{\frac{1}{4} \sum_{t_a=1}^4 E_{t_a}} cnSE_a}{4} = cnSe_a \end{aligned} \quad (4)$$

La relazione (1) consente di calcolare gli addetti per i soli trimestri relativi al generico anno a per cui si dispone di dati annuali di Contabilità relativi agli indipendenti. Tali dati si rendono disponibili per la prima volta nel mese di marzo di a^*+1^8 , dove con a^* si intende l'ultimo anno disponibile della serie storica. Considerate le scadenze Eurostat (cfr. §12.1), i dati annuali relativi ad a^* possono essere utilizzati a partire dalle stime rilasciate a maggio di a^*+1 , in occasione della diffusione del primo trimestre di a^*+1 , ma unicamente per la stima dei trimestri dell'anno a^* . Per la stima dei trimestri di a^*+1 non si dispone di informazioni ausiliarie per gli indipendenti, i quali vengono stimati per previsione a partire dall'ultimo dato disponibile per a^* . La metodologia di previsione scelta consiste nell'applicare la variazione congiunturale (grezza) dei dipendenti di Oros alla stima degli indipendenti trimestrali, applicandola per la prima volta sulla stima degli indipendenti ottenuta per il quarto trimestre di a^* e via via in avanti fino al trimestre di stima corrente:

$$\widehat{SE}_{t_{a^*+1}} = \widehat{SE}_{t_{a^*+1}-1} \cdot V(E_{t_{a^*+1}}) \quad (5)$$

dove: $V(E_{t_{a^*+1}})$ è il tasso di variazione congiunturale delle posizioni lavorative dipendenti di Oros tra il generico trimestre t dell'anno a^*+1 ed il rispettivo $t-1$.

La stima relativa a questi trimestri comporta che la variazione congiunturale degli addetti sia pari alla variazione congiunturale dei dipendenti. Questo deriva dalla relazione per cui la variazione congiunturale della somma di due variabili è uguale alla media ponderata delle variazioni congiunturali delle singole variabili (cfr. Lattanzio, 2016a). Sulla dinamica tendenziale si osserva un effetto simile.

La stima degli addetti così ottenuta è soggetta a revisione, la quale dipende dalla politica di revisione di Oros (cfr §2.3.1), che influenza sia la stima della componente dei dipendenti sia degli indipendenti (come descritto più avanti), e da quella di Contabilità Nazionale, via indipendenti. Con l'uscita di marzo, in particolare, la politica di revisione della Contabilità Nazionale prevede che:

- l'anno a^* sia rilasciato per la prima volta in modalità provvisoria ad un livello di disaggregazione a 38 branche;
- l'anno a^*-1 sia rivisto per la prima volta ma considerato ancora provvisorio e rilasciato ad un livello di disaggregazione a 64 branche;
- l'anno a^*-2 venga reso definitivo e rilasciato ad un livello di disaggregazione a 88 branche coincidenti con le divisioni Ateco.

8 Una versione rivista dei dati annuali della Contabilità è inoltre disponibile nel mese di settembre di $a+1$. Tuttavia, per limitare il numero di revisioni, nelle stime per STS viene considerata solo la diffusione di marzo.

Per questo motivo il rilascio (o *vintage*) generalmente prevede tre diversi livelli di disaggregazione. Per gli anni dal 2000 fino ad a^*-2 sono disponibili dati a livello di divisione Ateco 2007.

Conseguentemente alla caratteristica della fonte, si rende pertanto necessario disaggregare i dati di a^*-1 e a^*-2 . L'operazione si effettua utilizzando dei pesi di espansione dal livello più aggregato al livello più fine calcolati sulle stime dell'anno precedente.

In particolare, per l'anno a^*-1 si applicano pesi per espandere da un livello di 64 branche alla divisione Ateco calcolati a partire dai dati dell'anno a^*-2 per i quali si hanno a disposizione entrambi i livelli di aggregazione.

A cascata, per l'anno a^* si applicano dei pesi per espandere da un livello di 38 branche alla divisione Ateco calcolati a partire da dati stimati e aggregati in 38 branche, non rilevati, relativi all'anno a^*-1 . Si sarebbero potuti utilizzare pesi di conversione calcolati su a^*-2 come si fa per l'anno a^*-1 , ma essi avrebbero risentito di una struttura economica relativa a due anni prima. Il metodo rappresenta un buon compromesso tra accuratezza delle stime e maggior vicinanza al periodo di riferimento.

Dalla metodologia di calcolo consegue una revisione connessa alle modifiche dei dati di base che influiscono anche sulle stime di disaggregazione settoriale.

Tale politica di revisione viene incorporata nella stima degli indicatori STS sugli occupati che viene quindi rivista secondo il seguente schema:

- con il rilascio del primo trimestre di ogni anno (a^*+1), a maggio, vengono rivisti tutti i trimestri all'indietro fino ad a^*-2 (per un totale di 12 trimestri);
- nelle date successive di rilascio vengono rivisti i dati relativi a tutti i trimestri dell'anno a^*+1 e dell'anno a^* . La revisione dell'intero anno a^* , che differenzia la politica di stima degli addetti da quella dei dipendenti, dipende dalle relazioni (1-3), ovvero dall'impatto che la revisione di un qualsiasi trimestre t di a esercita sull'aggregato annuo. Infatti, la revisione di ogni dato trimestrale dei dipendenti ha effetto sul denominatore della (3).

La stima del monte retributivo come richiesto da STS, non necessita del ricorso a fonti ausiliarie per cui la costruzione degli aggregati segue le fasi standard della produzione degli indicatori Oros. Così come per le posizioni lavorative dipendenti, per le quali si rilasciano stime di livello, anche per i monti retributivi è necessario tuttavia ricorrere all'imputazione per le mancate risposte. A tal fine, non viene applicato un processo d'imputazione specifico, ma la correzione è proporzionale a quella sulle posizioni (cfr. §6.2 e §8.4). Occorre, tuttavia, tener conto di una differenza definitoria che caratterizza le due variabili: i monti retributivi non includono i dirigenti mentre le posizioni totali sono comprensive di questa categoria professionale. Per poter riportare correttamente l'aggiustamento delle posizioni lavorative ai monti retributivi al netto dei dirigenti è necessario calcolare il primo indicatore escludendo i dirigenti. Per poter realizzare questo passaggio è però necessario valutare l'impatto delle mancate risposte sulla stima del numero di dirigenti. Si ipotizza che le mancate risposte non abbiano una caratterizzazione rispetto alla tipologia professionale, per cui il numero di dipendenti al netto dei dirigenti (E_t^{nd}) viene aggiustato proporzionalmente alla correzione apportata alle posizioni lavorative totali ottenendo una stima del numero di dipendenti al netto dei dirigenti imputati (\hat{E}_t^{nd}). Il termine di proporzionalità viene valutato sugli ultimi quattro trimestri di stima finale (da $t-7$ a $t-4$) ed è definito come la media della quota dei dipendenti al netto dei dirigenti rispetto ai dipendenti totali:

$$\delta^{nd} = \frac{1}{4} \frac{\sum_{t-4}^{t-7} E_t^{nd}}{E_t} \quad (6)$$

Si ipotizza che l'incidenza del numero di dirigenti sul totale delle posizioni lavorative sia invariante nel periodo considerato. Il numero di posizioni lavorative netto dirigenti corrette si ottiene come:

$$\hat{E}_t^{nd} = \delta^{nd} \cdot E_t \quad \text{per } t > t-4 \quad (7)$$

Infine, la stima Oros dei monti retributivi aggiustati per l'effetto delle mancate risposte (\hat{W}_t^{nd}) è ottenuta moltiplicando le posizioni lavorative al netto dirigenti, corrette come illustrato sopra, per l'indicatore Oros sulle retribuzioni pro-capite ritenute non distorte per effetto delle mancate risposte⁹:

$$\hat{W}_t^{nd} = \left(\frac{W_t^{nd}}{E_t^{nd}} \right) \cdot \hat{E}_t^{nd} \quad (8)$$

L'entità delle revisioni della componente degli indipendenti derivante dalle revisioni dei dati di fonte CN viene considerata ancora non trascurabile, in particolare in alcuni settori, malgrado siano stati apportati alcuni affinamenti alla metodologia nel tempo. Per tale ragione, i dati sul numero totale delle persone occupate inviato ad Eurostat mantiene ancora un *flag* di confidenzialità. Al momento, la qualità dell'indicatore continua ad essere costantemente monitorata attraverso un'accurata analisi delle revisioni per componenti di stima, al fine di valutare il momento di un eventuale rilascio in formato libero anche di questi dati.

9 Eventuali mancate risposte con effetto rilevante sulle retribuzioni pro-capite vengono corrette nella fase di *editing* selettivo (cfr. §6.6).

GLOSSARIO

Cassa integrazione guadagni (Cig)

Strumento attraverso il quale lo Stato interviene a sostegno delle imprese costrette a contrarre o sospendere la propria attività a causa di situazioni di crisi o difficoltà tipizzate dalla legge. Consiste nell'erogazione gestita dall'Inps di un'indennità sostitutiva della retribuzione in favore dei dipendenti sospesi dal lavoro o sottoposti a riduzione di orario. Si distinguono tre forme di Cig:

- ordinaria (Cigo). Si applica al settore industriale in caso di sospensione o contrazione dell'attività produttiva per situazioni aziendali dovute a eventi temporanei e non imputabili all'imprenditore o ai lavoratori o a situazioni temporanee di mercato;
- straordinaria (Cigs). Si applica alle imprese in difficoltà in caso di ristrutturazione, riorganizzazione, riconversione aziendale, crisi aziendale e nei casi di procedure concorsuali;
- in deroga (Cigd). È un sostegno economico per operai, impiegati e quadri sospesi dal lavoro che non hanno (o non hanno più) accesso alla cassa integrazione guadagni ordinaria e straordinaria. Sostiene economicamente anche apprendisti, lavoratori interinali e a domicilio di aziende in Cigo e Cigs.

Classificazione Ateco 2007

È la versione nazionale della nomenclatura europea Nace.Rev.2, pubblicata sull'Official Journal il 20 dicembre 2006 (Regolamento CE n. 1893/2006 del PE e del Consiglio del 20/12/2006) e adottata dall'Istat il 1° gennaio 2008.

Contratto di solidarietà

Accordo stipulato tra l'azienda e le rappresentanze sindacali avente ad oggetto la diminuzione dell'orario di lavoro, al fine di mantenere l'occupazione in caso di crisi aziendale (contratti di solidarietà difensivi, art. 1 legge 863/84) o favorire nuove assunzioni attraverso una contestuale e programmata riduzione dell'orario di lavoro e della retribuzione (contratti di solidarietà espansivi art. 2 legge 863/84).

Costo del lavoro

Somma delle retribuzioni lorde e degli oneri sociali.

Dati corretti per gli effetti di calendario

Dati depurati, mediante apposite tecniche statistiche, dalla variabilità attribuibile alla composizione del calendario nei singoli periodi (mesi o trimestri) dell'anno, dovuta al diverso numero di giorni lavorativi o di giorni specifici della settimana in essi contenuti e alla presenza di festività nazionali civili e religiose, fisse e mobili (festività pasquali), nonché dell'anno bisestile. Il ricorso a tale trasformazione dei dati consente di cogliere in maniera più adeguata sia le variazioni tendenziali (calcolate rispetto allo stesso periodo dell'anno precedente), sia le variazioni medie annue.

Dati destagionalizzati

Dati depurati, mediante apposite tecniche statistiche, dalle fluttuazioni attribuibili alla componente stagionale (dovute a fattori meteorologici, consuetudinari, legislativi, ecc.) e, se significativi, dagli effetti di calendario. Questa trasformazione dei dati è la più idonea a cogliere l'evoluzione congiunturale di un indicatore.

Monte ore lavorate (nelle posizioni dipendenti)

Il numero totale delle ore di lavoro ordinario e straordinario prestate dai dipendenti con contratto di lavoro.

Oneri sociali

Complesso dei contributi a carico del datore di lavoro versati agli enti di previdenza ed assistenza sociale e degli accantonamenti di fine rapporto.

Ore di cassa integrazione guadagni

Ore complessive di cassa integrazione guadagni, ordinaria, straordinaria e in deroga, e ore di solidarietà di cui le imprese hanno usufruito nel trimestre di riferimento dell'indagine.

Ore di solidarietà

Ore non lavorate a causa dell'applicazione dei contratti di solidarietà.

Ore lavorate per dipendente

Numero medio delle ore di lavoro ordinario e straordinario prestate dai dipendenti con contratto di lavoro. Sono calcolate in rapporto alle posizioni lavorative dipendenti.

Ore ordinarie

Totale delle ore lavorate, comprese quelle notturne e festive, con esclusione delle ore di straordinario, di cassa integrazione guadagni e ore non lavorate relative ad assenze per ferie, festività, permessi personali ed in genere delle ore non lavorate, anche se per esse è stata corrisposta una retribuzione.

Posizione lavorativa dipendente

E' contraddistinta da contratto di lavoro tra una persona fisica e un'unità produttiva (impresa o istituzione privata), che prevede lo svolgimento di una prestazione lavorativa a fronte di un compenso (retribuzione). Le posizioni lavorative rappresentano, quindi, il numero di posti di lavoro occupati da lavoratori dipendenti (a tempo pieno e a tempo parziale), indipendentemente dalle ore lavorate, ad una determinata data di riferimento. Come il numero di occupati anche le posizioni lavorative rappresentano pertanto una variabile di stock ad un certo istante nel tempo. Sono inclusi anche i lavoratori che, legati all'unità produttiva da regolare contratto di lavoro, sono temporaneamente assenti per cause varie quali: ferie, permessi, maternità, cassa integrazione guadagni, eccetera.

Posizione lavorativa in somministrazione (ex-interinali)

Posizione lavorativa dipendente con contratto di somministrazione. I lavoratori con contratto di somministrazione vengono rilevati tra i dipendenti delle agenzie di somministrazione di lavoro e non tra i dipendenti delle unità utilizzatrici. Le posizioni in somministrazione non includono il personale delle agenzie fornitrici di lavoro temporaneo assunto con contratto di lavoro dipendente diverso dalla somministrazione. Il gruppo di attività economica Ateco 2007 in cui è classificata questa tipologia di lavoro è “Attività delle agenzie di lavoro temporaneo (gruppo 78.2)” che rientra nella sezione N “Noleggio, agenzie di viaggio, servizi di supporto alle imprese”.

Retribuzione contrattuale di cassa

Retribuzione comprendente tutte le voci retributive considerate mensilmente nell'indice delle retribuzioni contrattuali alle quali si aggiungono eventuali arretrati e una tantum. Gli importi riferiti a ciascuna voce retributiva sono attribuiti ai mesi di effettiva erogazione. La retribuzione di cassa è calcolata per tutti i livelli di inquadramento previsti in occasione della definizione della base (che è la stessa degli indici delle retribuzioni contrattuali), al lordo delle trattenute fiscali e previdenziali. Gli aggregati superiori vengono quindi determinati secondo una struttura occupazionale costante, che consente di monitorare la dinamica retributiva al netto degli effetti dovuti a mutamenti nella struttura dell'occupazione per qualifica e livello di inquadramento.

Retribuzioni lorde “di fatto”

Salari, stipendi e competenze accessorie in denaro, al lordo delle trattenute fiscali e previdenziali, corrisposte ai lavoratori dipendenti direttamente e con carattere di periodicità, secondo quanto stabilito dai contratti, dagli accordi aziendali e individuali e dalle norme in vigore. Le retribuzioni “di fatto” si differenziano da quelle “contrattuali” perché queste ultime comprendono per definizione solo le competenze determinate dai contratti nazionali di lavoro.

Unità di lavoro dipendente equivalenti a tempo pieno (Ula)

Unità di misura del volume di lavoro prestato nelle posizioni lavorative, calcolata riducendo il valore unitario delle posizioni lavorative a tempo parziale, di quelle dei lavoratori a chiamata e dei lavoratori interinali in equivalenti a tempo pieno. Esse comprendono: quadri, impiegati, operai, commessi, apprendisti e lavoranti a domicilio. Sono esclusi i dirigenti. A differenza del numero di posizioni lavorative, le Ula sono calcolate al netto dei lavoratori in cassa integrazione e solidarietà.

Variazione congiunturale

Variazione percentuale rispetto al mese o periodo immediatamente precedente.

Variazione tendenziale

Variazione percentuale rispetto allo stesso mese o periodo dell'anno precedente.

RIFERIMENTI BIBLIOGRAFICI

- AAVV. 2000. Guida all'utilizzo di TRAMO-SEATS per la destagionalizzazione delle serie storiche. *Documenti Istat*, n.4. Roma.
<https://www.istat.it/it/archivio/219375>.
- AAVV. 2012. *Linee guida per la qualità dei processi statistici Versione 1.1*. Dicembre 2012. Istat, Roma.
<https://www.istat.it/it/archivio/86175>.
- AAVV. 2015. Relazione del team tecnico incaricato della definizione di metodi standard per la destagionalizzazione di serie storiche con metodi implementati in diversi strumenti IT (TS, X12-Arima, X13-Arima-Seats, JDemetra). Nota interna. Istat, Roma.
<https://www.istat.it/it/files/2013/12/Relazione-Gdl-ultima-vs-inviata.pdf>.
- Agenzia delle Entrate. 2007. Istruzioni per la compilazione della domanda di attribuzione del numero di codice fiscale e dichiarazione di inizio attività, variazione dati o cessazione attività ai fini IVA. https://www.agenziaentrate.gov.it/wps/content/Nsilib/Nsi/Schede/Istanze/AA7_10/?page=schedeistanze.
- Baldi C., D. Bellisai, F. Ceccato, S. Pacini, L. Serbassi, M. Sorrentino e D. Tuzi. 2011. The system of short term business statistics on labour in Italy. The challenges of data integration. Paper presentato al Workshop: ESSnet Data Integration. Madrid, 24-25 novembre.
http://www.ine.es/e/essnetdi_ws2011/ppts/Baldi_et_al.pdf.
- Baldi C., E. Cimino, F., Rapiti, P. Minicucci, R. Succi e D. Tuzi. 2001. L'utilizzo dei dati INPS per la stima trimestrale del numero dei dipendenti, le retribuzioni, il costo del lavoro e le ore lavorate. *Documenti* n.14. Istat, Roma.
<https://www.istat.it/it/archivio/219322>.
- Baldi C., F. Ceccato, E. Cimino, M.C. Congia, S. Pacini, F. Rapiti e D. Tuzi. 2004. Use of Administrative Data to produce Short Term Statistics on Employment, Wages and Labour Cost, *Essays*, 15. Istat, Roma.
<http://www.oecd.org/sdd/36232440.pdf>.
- Baldi C., F. Ceccato, E. Cimino, M.C. Congia, S. Pacini, F. Rapiti e D. Tuzi 2008. Seminario: Strategie e metodi per il controllo e la correzione dei dati nelle indagini sulle imprese: alcune esperienze nel settore delle statistiche congiunturali. *Contributi Istat*. n.13:29-61. Roma.
http://www3.istat.it/dati/pubbsci/contributi/Contributi/contr_2008/13_2008.pdf.
- Bellisai D., S. Binci, S. Gigante, A. Libratore, L. Serbassi e M. Sorrentino. 2013. Verso la modernizzazione della produzione delle statistiche congiunturali sull'input di lavoro: il processo di integrazione tra la rilevazione mensile sulle grandi imprese e quella trimestrale sui posti vacanti e le ore lavorate. *Working paper* n.11. Istat, Roma.
https://www.istat.it/it/files/2013/12/IWP_11_2013.pdf.
- Ceccarelli C., Ceccato F., Pacini S. e Rossetti F. 2017. Stima del costo del lavoro a livello individuale: un approccio da registro. *Rivista Italiana di Economia, Demografia e Statistica - Società Italiana di Economia, Demografia e Statistica (SIEDS) – Vol. LXXI, n° 3*. pp.29-40. Roma.
<http://www.sieds.it/index.php/2017/10/01/rivista-lxxi-3-luglio-settembre-2017/>.
- Ciammola A., F. Ceccato, M.C. Congia, S. Pacini, F. Rapiti e D. Tuzi. 2009. The Italian Labour Cost Index (LCI): Sources and Methods. *Contributi* n.8. Istat, Roma.
https://www.istat.it/it/files/2018/07/08_2009.pdf.
- Ciammola A., D. Tuzi. 2010. Internal Coherence in Seasonally Adjusted Chain Laspeyres Indices: an Application to the Italian Hourly Labour Cost Indicators. Paper presentato alla Conferenza: the European Conference on Quality in Official Statistics, Q2010. Helsinki 4-6 maggio.
https://q2010.stat.fi/media/presentations/session-5/ciammola_tuzi_paper.pdf.

- Ciammola A., D. Tuzi. 2015. Calendar and Seasonal adjustment of the LCI: direct versus indirect approach. The Italian experience. Paper presentato al: Workshop on Labour Cost Statistics. Roma 5-6 maggio.
http://www.istat.it/en/files/2015/04/Item-4.1-Seasonal-and-working-day-adjustment-of-the-LCI_IT.pdf.
- Congia, M.C. e S. Pacini, 2012. La stima da fonti amministrative di indicatori retributivi congiunturali al netto della cassa integrazione guadagni. *Rivista di statistica ufficiale* n. 2-3/2012. Istat, Roma.
<http://www.istat.it/it/files/2013/03/19-40.pdf>.
- Congia M.C., S. Pacini e D. Tuzi. 2008a. Quality Challenges in Processing Administrative Data to Produce Short-Term Labour Cost Statistics. Paper presentato alla Conferenza: the European Conference on Quality in Official Statistics, Q2008. Roma, 8-11 luglio.
<http://q2008.istat.it/sessions/paper/29Congia.pdf>.
- Congia M.C., S. Pacini e D. Tuzi. 2008b. The Editing Process in the Italian Short-Term Survey on Labour Cost based on Administrative Data. Paper presentato alla Conferenza: the UNECE Conference of European Statisticians, Work Session on Statistical Data Editing. Vienna, 21 – 23 aprile.
<https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2008/04/sde/wp.8.e.pdf>.
- Costanzo L. 2012. Legal barriers, quality issues: what really hampers a wider use of administrative data in business statistics? Paper presentato alla European Conference on Quality in Official Statistics (Q2012). Atene, 29 maggio – 1 giugno.
- Costanzo L. 2013. Overview of Existing Practices. Milestone 1.5/2011. Essnet Admin Data - Work Package 1: overview of MSs' Practices in the Uses of Administrative Data for Business Statistics. Report del gruppo di lavoro.
<https://ec.europa.eu/eurostat/cros/system/files/Changes%20in%20the%20use%20of%20admin%20data%20during%20the%20MEETS%20Programme.pdf>.
- De Waal T., J. Pannekoek e S. Scholtus. 2011. *Handbook of Statistical Data Editing and Imputation*. John Wiley & Sons.
- Eurostat. 2006. Methodology of short-term business statistics. Interpretation and guidelines. Luxemburg.
<http://ec.europa.eu/eurostat/documents/1916593/1917176/KS-BG-06-001-EN.pdf/973bf1c8-e598-4228-9d18-2f7704a95e45>.
- Eurostat. 2014. Handbook on Methodology of Modern Business Statistics. Luxemburg.
https://ec.europa.eu/eurostat/cros/content/handbook-methodology-modern-business-statistics_en.
- Eurostat. 2015. ESS guidelines on seasonal adjustment. *Manuals and guidelines*. Luxembourg: Publications Office of the European Union.
<http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-15-001>.
- Gómez, V., Maravall A. 1996. Programs TRAMO and SEATS, Instruction for User (Beta Version: September 1996). *Working Papers*, 9628. Banco de España.
<http://ideas.repec.org/p/bde/wpaper/9628.html>.
- Inps, Banche Dati Statistiche, Osservatorio sulle Ore autorizzate di Cassa Integrazione Guadagni, <https://www.inps.it/webidentity/banchedatistatistiche/menu/cig/main1.html>, ultima consultazione 28/04/2017.
- Inps. 2016. Cassa Integrazione Guadagni e disoccupazione. Ore autorizzate per trattamenti di integrazione salariale, domande e beneficiari di disoccupazione e mobilità. *Report mensile* novembre. Roma.
<http://www.inps.it/webidentity/banchedatistatistiche/menu/cig/main1.html>.
- Istat. 2005. Classificazione delle forme giuridiche delle unità legali. *Metodi e Norme*, n. 26. Roma.
http://www.istat.it/it/files/2011/01/mn0526classificazione_forme_giuridiche_unita_legali.pdf.
- Istat. 2006. Rilevazione mensile sull'occupazione, gli orari di lavoro e le retribuzioni nelle grandi imprese. *Metodi e Norme*, n. 29. Roma.
<https://ebiblio.istat.it/digibib/Metodi%20e%20norme/IST0049227Ed2006N29.pdf>.

- Istat. 2013a. Il sistema degli indicatori congiunturali sulla domanda di lavoro e le retribuzioni in Ateco 2007 e base 2005. Metodi. Roma.
<http://www.istat.it/it/archivio/97314>.
- Istat. 2013b. *Retribuzioni di fatto e costo del lavoro. La nuova base 2010*. Nota informativa. Roma, 14 giugno.
http://www.istat.it/it/files/2011/03/Retribuzioni-di-fatto-e-costo-del-lavoro-14_giu_2013-Nota-informativa.pdf.
- Istat. 2015. *Le tendenze recenti dell'occupazione*. Statistiche focus. Roma, 17 giugno.
<http://www.istat.it/it/files/2015/06/focus-lavoro.pdf>.
- Istat. 2016a. Il Sistema Integrato dei Registri nell'ambito del processo di modernizzazione dell'Istat. Versione 0.1 08/02/2016. Documento interno finalizzato alla diffusione di informazioni sullo stato di avanzamento del progetto di modernizzazione dell'Istituto Nazionale di Statistica in atto da 2016. Roma.
- Istat 2016b. *Atti del 9° Censimento generale dell'industria e dei servizi e Censimento delle istituzioni non profit*. Roma.
<https://www.istat.it/it/archivio/179737>.
- Lattanzio M. 2016a. Scomposizione di indici e tassi di variazione: focus su variabili rapporto e legge generale di scomposizione dei tassi di variazione (LSV). *Rivista Italiana di Economia, Demografia e Statistica - Società Italiana di Economia, Demografia e Statistica (SIEDS) – Vol. LXX, n° 4*. pp.135-146.
http://www.sieds.it/listing/RePEc/journal/2016LXX_N4rieds_sieds.pdf.
- Lattanzio M. 2016b. Indici semplici di variabili rapporto: analisi delle proprietà, scomposizione e caso di studio nella rilevazione Oros. *Documento proposto per la pubblicazione nella collana Working Paper*. Istat, Roma
- Luzi O. e R. Monducci 2016. The new statistical register “Frame SBS”: overview and perspectives. *Rivista di statistica ufficiale* n.1. Istat, Roma.
http://www.istat.it/it/files/2016/11/RSU_1_2016_Testointegrale.pdf.
- Maasing E., T. Remes, C. Baldi e P. Vlag. 2012. STS estimates based solely on administrative data: final results and recommendations. SGA 2011: Deliverable 4.1. Essnet use of administrative and accounts data in business statistics. Wp4. Timeliness of administrative sources for monthly and quarterly estimates.
https://ec.europa.eu/eurostat/cros/system/files/SGA%202011_Deliverable_4.1.pdf.
- Rapiti F. 1998. L'utilizzo dei dati INPS per la costruzione di indicatori sull'occupazione e le retribuzioni: un'analisi preliminare di fattibilità. *Documenti* n.1. Istat, Roma.
- Rapiti F.M., F. Ceccato, M.C. Congia, S. Pacini e D. Tuzi. 2010. What have we learned in almost 10-years experience in dealing with administrative data for short term employment and wages indicators? Paper presentato al seminario: Using administrative data in the production of business statistics”. Roma, 18-19 marzo.
http://www.ine.pt/filme_inst/essnet/papers/Session2/Paper2.4.pdf.
- Rocci F. e L. Serbassi. 2008. Controllo e correzione nell'indagine mensile sulle grandi imprese: metodi e prime evidenze da un'analisi retrospettiva sulla qualità. Seminario: Strategie e metodi per il controllo e la correzione dei dati nelle indagini sulle imprese: alcune esperienze nel settore delle statistiche congiunturali. *Contributi Istat*. n.13: 7-28. Roma.
http://www3.istat.it/dati/pubbsci/contributi/Contributi/contr_2008/13_2008.pdf.
- Runci M. C., G. Di Bella e L. Galiè 2016. Il sistema di integrazione dei dati amministrativi in Istat. *Rivista di statistica ufficiale* n.18. Istat, Roma.
<https://www4.istat.it/it/archivio/193056>.
- Zhang L.-C. 2012. Topics of statistical theory for register-based statistics and data integration. *Statistica Neerlandica* n.66; 41-63.
- Wallgren, A. e B. Wallgren. 2007. *Register-based Statistics - Administrative Data for Statistical Purposes*. Chichester, England, John Wiley & Sons Ltd.